Creation of novel insulin receptor constructs from fragments containing binding sites 1 or 2

Mateusz Kozak

Master of Science by Research

University of York

Chemistry

December 2015

ABSTRACT

Insulin signalling regulates crucial processes of metabolism, growth and differentiation, and insulin is an essential medicine for treatment of diabetes mellitus. The insulin receptor (IR) is an $(\alpha\beta)_2$ dimeric transmembrane protein responsible for mediating the effects of insulin and, to a degree, insulin-like growth factors. It is over 1000 amino acids long, multi-domain, highly glycosylated and stabilized by multiple disulphide bridges. These considerations have limited our structural understanding of the receptor and its interactions with insulin despite its great medical importance. Insulin binding is complex: IR binding sites are contained entirely within the extracellular α chain, and each monomeric chain contains two distinct sub-sites that bind insulin to form a high-affinity crosslink. Despite being a symmetrical dimer, the full-length receptor binds only one insulin molecule with high affinity, presumably because an asymmetric fit induced by ligand forming one sub-site crosslink prevents formation of the other.

This work seeks to produce and characterize novel, high-affinity heterodimeric constructs of IR α chain modified to only contain one possible site 1-2 crosslink. To this end, expression vectors were engineered based on plasmids obtained from Novo Nordisk A/S coding for fusion proteins of IR1-593 and IR310-601.678-719/731 receptor fragments connected with poly(Gly-Ser) linkers containing protease recognition sites. Expression levels were found to be comparable to those of established IR constructs like IR(1-310). Two alternative signal peptides were tested, but did not improve protein expression or secretion. Two constructs were expressed and purified on a preparative scale and their oligomerization state and cleavage was assessed. Purification of a construct with a 3C protease site linker showed that producing two receptor fragments from a single polypeptide chain is a viable approach. If a site 1 – site 2' limited heterodimers could be reconstituted, it would open an exciting new approach to structural studies of insulin/IR complexes.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGEMENTS

First and foremost I would like to thank Professor Marek Brzozowski for being my thesis advisor and for teaching me basics of crystallography, but also for strong support out of the laboratory and being a fundamentally good person in addition to being a good boss. I thank Professor Gideon Davies for being my Independent Panel Member.

I would also like to thank Drs Jakob Brandt and Gerd Schluckebier of Novo Nordisk for creating the opportunity for me to enter insulin research. I am extremely grateful to them for sharing their accumulated knowledge, extremely unique expertise and a look into research at one of the top biotechnology companies in the world. Parvaneh Sadeghi also provided invaluable technical support, know-how and a welcoming presence during my visits to Copenhagen.

I am thankful to Drs Cristina Viola and Tim Ganderton for teaching me everything I know about tissue culture, for presenting and challenging scientific ideas and for being all-around great co-workers. For their friendship, all the good times in York and many discussions scientific and otherwise, many thanks must go to Catherine, Łukasz and Dr Dan.

Finally, I would like to thank all of the above for patience and guidance at pivotal moments.

AUTHOR'S DECLARATION

All the experiments described were performed solely by the author. This text was written entirely by the author and none of the work presented has been used for fulfilment of a research degree, nor previously published. Ideas and materials from outside sources are credited in the text.

1 INTRODUCTION

1.1 INSULIN

1.1.1 The discovery of insulin

Insulin is a 51 amino acids long peptide hormone synthesised in the β-cells of the islets of Langerhans in the pancreas. It is crucial to the regulation of metabolism and growth, and dysfunctions in insulin signalling are the cause of diabetes mellitus. As a relatively simple protein with crucial medical applications, it has a special place in the history of biochemistry and biotechnology for being a number of 'firsts': it was the first protein to have its primary sequence determined by Frederic Sanger [1–4], one of the first proteins to have its 3D structure determined by means of X-ray crystallography by Dorothy Hodgkin and colleagues [5] and the first biopharmaceutical to be produced using recombinant DNA technology. Its discovery is attributed to Frederick Banting, who first described a process of isolating it from bovine pancreas in 1922 [6] along with Charles Best and James Collip (a more complete story of insulin discovery can be found in ref. 7). Shortly after the announcement of its discovery, insulin production patents were secured by the American pharmaceutical company Eli Lilly which started commercial distribution under the brand name Iletin in 1923. At the same time, Danish Nobel Prize winner August Krogh founded Nordisk Insulin Laboratory in Copenhagen, a non-profit organisation that would later become Novo Nordisk A/S and make Denmark the second biggest producer of insulin after the United States.

1.1.2 The basics of insulin signalling

Insulin and the tightly related Insulin-like Growth Factors (IGFs) are ancestral peptide hormones that play a crucial role in metazoan processes of metabolism, growth and differentiation. Insulin-like peptides are highly evolutionarily conserved and can be found in invertebrates [8], plants [9] and even unicellular eukaryotes [10] where they play a developmental role. The metabolic control exerted by insulin in vertebrates appears to be a recent evolutionary development coinciding with IGFs becoming a major mitogenic hormone [11]. Physiological effects of insulin are well-established and include increase of nutrient uptake, primarily glucose, by muscle, fat and liver cells, increase in glycogen, fatty acid and protein synthesis and decreased catabolism and autophagy. Maintaining homeostasis by regulation of blood glucose levels is also a key consequence of insulin's ability to stimulate nutrient uptake.

Insulin action is mediated primarily through two pathways: the PI3K/Akt pathway and the Ras (MAPK) pathway [12,13]. The dimeric insulin receptor has two cytosolic tyrosine kinase domains that trans-autophosphorylate upon insulin binding and subsequently phosphorylate their intracellular substrates (IRS, insulin receptor substrate proteins). Phosphorylated IRS proteins are recognized by intracellular effectors that contain Src-homology-2 (SH2) domains, including the p85 subunit of phosphoinositide 3-kinase and Grb2 that activates the Ras pathway. Activation of the PI3K pathway and protein kinase B (Akt/PKB) has a plethora of effects including glucose uptake (through regulation of Glut-4 glucose transporter trafficking), glycogen synthesis (through glycogen synthase kinase 3) or inhibition of apoptosis (through the Beclin protein family). Through the PI3K pathway, insulin can also enhance mTor signalling that regulates various processes of growth and the cell cycle. The insulin signalling network is complex and beyond the scope of this work; please see ref. 13 for a review.

1.1.3 Insulin biosynthesis

The product of the insulin gene is a 110 amino acid precursor peptide with a 24 amino acid N-terminal signal peptide that directs it to the secretion pathway and is cleaved off in the rough endoplasmic reticulum [14,15]. The remaining amino acids form proinsulin comprising three segments: the N-terminal B chain, the C-terminal A chain and a connecting C chain [16,17]. This precursor undergoes further proteolytic processing with the subtilisin-like prohormone convertases PC2 and PC3 to remove the intervening C-peptide [18,19], and carboxypeptidase H to remove dibasic residues at the B chain C-terminus [20]. The liberated C peptide has been reported to bind to membranes of several cell types, in particular renal tubular cells, and saturate its receptor at physiological concentrations [21]. As such, administration of the C-peptide has no effect on healthy patients and its physiological function has long been elusive; in diabetic patients it appears to limit nerve and renal damage associated with the disease [22].

1.1.4 The structure of insulin

The mature form of insulin is a 5808 Da heterodimer comprising of two chains, 21 amino acid chain A and 30 amino acid long chain B. The two chains are disulphide bonded through cysteines A7-B7 and A20-B19. The third cysteine bridge, A6-A11, stabilizes the parallel structure of two A chain α-helices (A1-8 and A12-18).

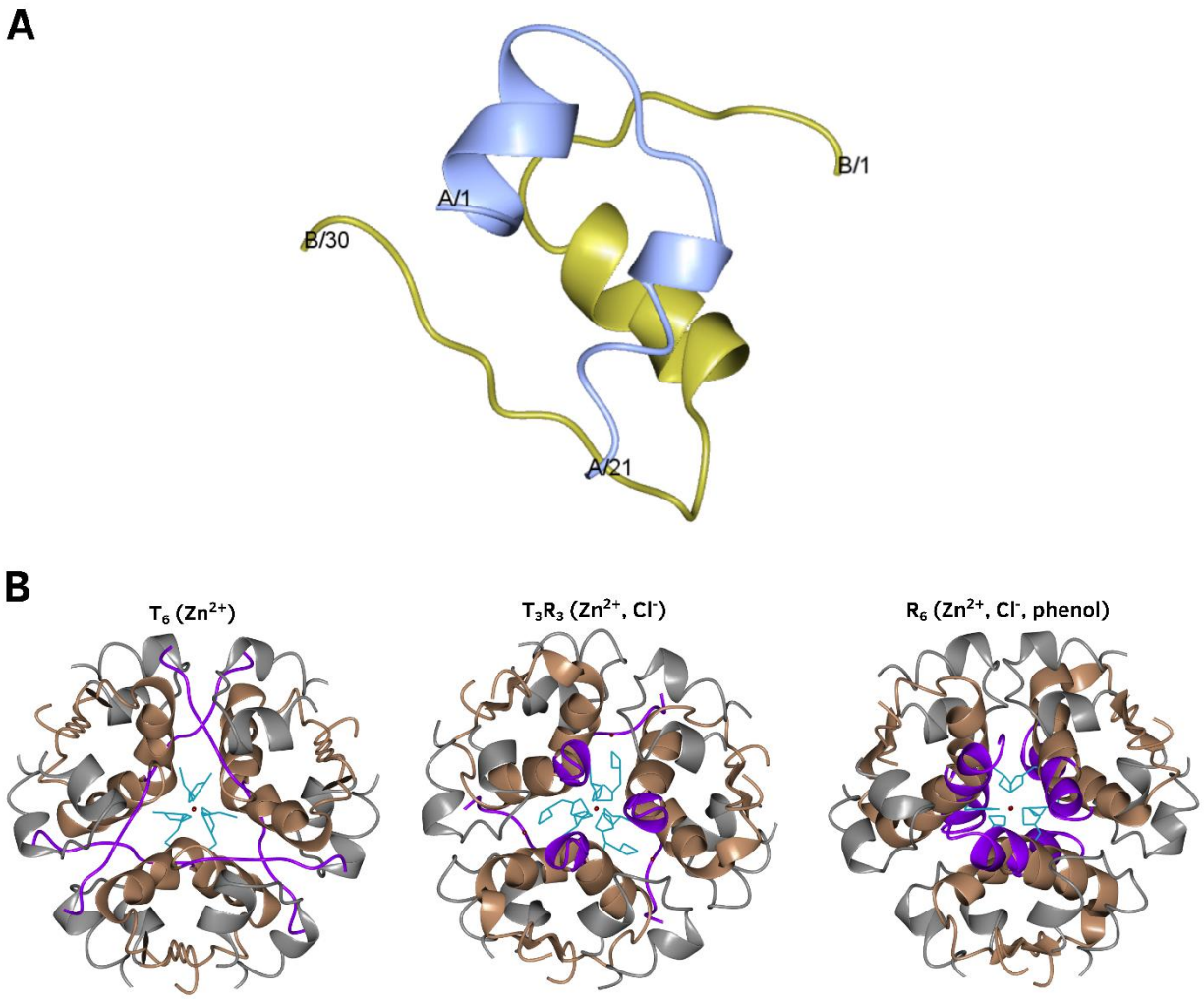

Figure 1.1| The structure of insulin
A| Porcine insulin molecule 1 from 2-zinc insulin crystals (PDB 4INS). **B|** T→R transition in insulin hexamers. B chain residues 1-8 can adopt a fully extended conformation called the T-state or an α-helical conformation called the R-state. Hexamers can exist as T_6 (left, 4INS), T_3R_3 (middle, 1ZNI) or R_6 (right, 1ZNJ) states depending on the crystallization conditions. B chain residues 1-8 are shown in purple, HisB10 side chains in cyan, A chains in grey and B chain residues 9-30 in brown. Made with ccp4mg [23].

The structure of insulin (Figure 1.1A) and its numerous analogues has been studied extensively by both X-ray crystallography and NMR methods. Insulin hexamers form rhombohedral crystals in the presence of divalent metal cations, particularly Zn^{2+} [5]; hexameric crystals are also the pancreatic storage form of insulin. In the absence of zinc, insulin dimers can form cubic crystals, particularly under alkaline conditions and in the presence of trace amounts of organic solvents [24]. The A chain of insulin comprises of two antiparallel α-helices (A1-8 and A12-18) connected by a non-canonical turn. The B chain has one central α helix (B9-19) surrounded by type II' (B7-10) and type I (B20-23) β-turns. Whereas the C-terminal portion of the B chain forms a flexible β-strand, the N-terminus (B1-

B8) adopts two distinct conformations: an extended T-state or an α-helical R-state that forms a part of the central helix [25].

The T→R transition (Figure 1.1B) represents a major variation in crystal structures of insulin [26]. In the classical 2-Zn structure [5], three symmetrical dimers surround two axial Zn^{2+} ions coordinated by three water molecules and three B10 histidyl residues, and all molecules adopt the extended T_6 conformation. With the addition of chloride ions, a transitional crystalline form T_3R_3 is adopted [25,27] and the twofold symmetry of the dimer is lost: the B chain N-terminus of molecule 1 forms an extension of the central α-helix while molecule 2 retains the conformation found in the T_6 hexamer. An R_6 hexamer in which a B1-19 helix is present in all molecules can be stabilized by small cyclic alcohols, notably phenol, which bind in a pocket created by the B1-8 helix and several A-chain residues, particularly cysteines A6 and A11 [28,29]. NMR studies suggest that insulin monomer adopts a T-like fold in solution and has a similar structure to 2-Zn insulin [30].

1.2 THE INSULIN RECEPTOR

Insulin and IGFs are recognized by highly similar transmembrane receptors with intracellular tyrosine kinase effector domains and a dimeric $(\alpha\beta)_2$ subunit structure.

The insulin receptor is a multi-domain, over 300 kDa extracellular membrane protein. The receptor monomer is expressed as a single polypeptide chain that is proteolytically cleaved into an extracellular α subunit and a transmembrane β subunit containing the intracellular tyrosine kinase domain. These subunits are covalently bound by a single disulphide bridge to form a monomer of the $(\alpha\beta)_2$ receptor; monomer-monomer interactions, as well as the intra-subunit structure, are further stabilized by a number of disulphide bridges .

1.2.1 Primary structure of the IR

The human insulin receptor cDNA was first isolated and described in 1985 [31,32]. The gene, located on chromosome 19, consists of 22 exons, 11 for each subunit, spread over 120 kb [33]. The IR gene mRNA undergoes alternative splicing in a tissue-specific and developmentally regulated manner [34], giving rise to two isoforms differing by the presence of a 36 bp exon 11. The IR has multiple domains (see Figure 1.2 in section 1.2.4) and most domain boundaries coincide with exon boundaries [35]: exon 1 encodes the signal peptide, exon 2 the leucine-rich domain 1 (L1), exon 3 the cysteine-rich (CR) domain and exons 4-6 the second leucine-rich domain (L2). These are followed by three fibronectin type III domains FnIII-1 (exons 7-8), FnIII-2 (exon 9 and 3' region of exon 12) and FnIII-3 (exons 13-14). FnIII-2 spans across both subunits and is interrupted by the insert domain (ID, exons 10, 11 and 5' region of exon 12); furin cleavage site is located within the ID, between exons 11 and 12.

The β subunit consists of ID, FnIII-2 and FnIII-3 on the extracellular side, a single-pass transmembrane domain (exon 15) and the tyrosine kinase (exons 16-22) on the cytosolic side.

1.2.2 IR post-translational modifications

The $(\alpha\beta)_2$ subunit composition of the receptor has been known since 1980 [36], and the primary sequence provided evidence that the protein is expressed as a single chain, 1382 amino acid long precursor. The precursor chain is synthesised in the ER and needs to undergo both partial glycosylation [37] and disulphide pairing/dimerization [38] before being processed in the *trans*-Golgi network and inserted into the membrane. The main pro-IR processing enzyme is the subtilisin-like proprotein convertase furin, which recognises a consensus site RKRR↓ [39]. The mature, glycosylated receptor consists of a ~135 kDa α subunit and a ~90 kDa β subunit, connected with a single cysteine bridge in each monomer.

The IR primary sequence contains 37 cysteines in the α subunit, 4 cysteines in the β subunit extracellular region, and 6 in the β subunit cytosolic region. A small number of disulphide bonds that stabilize the IR homodimer and are highly susceptible to reducing agents were termed class I bonds; by contrast, the α-β bond (class II) can tolerate DTT concentrations as high as 50 mM [40]. The only α-β bond, connecting aligned domains FnIII-2 and FnIII-3, was determined to be C647-C872 by mass spectrometry and N-terminal sequencing of fragments obtained by CNBr and enzymatic digestions [41]; a later X-ray crystal structure [42] presented a conflicting view where C872 was unpaired and C647-C860 was the α-β bond. The main class I α-α bond is C524-C524 [43] but C524 mutants retain their dimeric structure despite lowered tyrosine kinase activity [44,45]. The other α-α bond appears to be formed by C682, as only C524S,C682S double mutants express predominantly as αβ monomers [46]; the other cysteines of the C682, C683, C685 triplet are suspected to form small intra-chain loops or additional dimer bonds [41]. N-ethylmaleimide alkylation studies suggest there is one free thiol group per αβ monomer located in the β subunit [47], and the other cysteines are expected to form stabilizing intra-chain disulphides [48].

The IR glycosylation accounts for ~50 kDa of its apparent molecular weight [37,49]. The β subunit has six O-linked mucin-type glycans, all found at residues within the ID/FnIII-2 region: T744, T749, S757, S758, T759, and T763 [50]. N-linked glycans can be found in both subunits and include complex (endoglycosidase F sensitive) and high-mannose (endoglycosidase H sensitive) species at residues 16, 25, 111, 215, 255, 295, 418, 514, 606, 624, 671, 742, 755, and 893 [51].

1.2.3 Alternative splicing produces IR-A and IR-B isoforms

Alternative splicing of exon 11 [34] gives rise to the two isoforms of the insulin receptor, IR-A (exon 11 minus) and IR-B (exon 11 plus). Exon 11 is 36 bp long and encodes 12 amino acids located at the binding-critical C-terminus of the α subunit (residues 717-728). The main functional consequence of IR alternative splicing is IR-A affinity for IGF-II. Comparative competition studies with radiolabeled insulin [52] reveal a slight difference between IR-A and -B in affinity for insulin (EC_{50} of 0.9 and 1.0 nM, respectively) but significantly different affinities for IGF-II (EC_{50} of 2.5 and >20 nM); similar results were obtained when comparing insulin- or IGF-II induced autophosphorylation. The same study found that IR-A affinity for IGF-II is very similar to that of IGF-I receptor (IGF-IR), which is considered the physiological receptor of both IGF-I and IGF-II, and that neither IR isoform binds IGF-I with high affinity. IR-A also preferentially stimulates mitogenic or metabolic pathways based on activation with IGF-II or insulin: IGF-II is better at activating the Shc/ERK pathway than insulin, which is more effective at activating the PI3K/Akt pathway [53], and the differences in signalling effects were observed at gene expression levels as well [54]. IGF-II, but not IGF-I, can stimulate proliferation of $IGF1R^{-}/IR$ cells [55]. In mouse embryos, a double $Igf1r^{-}/Ir^{-}$ knockout produces a phenotype with 30% of wild-type weight, as does a double $Igf1r^{-}/Igf2^{-}$ knockout [56]. An $Igf1r^{-}/Igf1^{-}$ mutant shows less severe growth retardation at 45% of wild-type weight, and an $Igf1r^{-}/Igf2r^{-}$ mutant which produces an excess of IGF-II due to abrogation of IGF2R mediated turnover is normal birth weight. This $Igf2r^{-}$ rescue effect is abolished in $Igf1r^{-}/Igf2r^{-}/Ir^{-}$ triple mutants which exhibit the 30% phenotype. Together, biochemical and *in vivo* data suggest that IR-A is an IGF-II receptor important for mediating its mitogenic effects. Consistently, IR-A is ubiquitous but predominantly expressed during prenatal development and in cancerous cells [52], whereas IR-B is located in tissues that are targets for metabolic actions of insulin like liver, kidneys, muscle and adipocytes [57]. As both isoforms are expressed by all insulin-sensitive cells to some extent, IR-A/IR-B heterodimeric hybrid receptors also exist, which further complicates the picture of IR activation by different ligands. However, it has been suggested that exon 11 acts as a sorting signal to direct IR isoforms to different plasma membrane domains and to limit heterodimerization [58].

1.2.4 Structure of the IR

The earliest IR-related structure to be solved was the tyrosine kinase of the β subunit in its basal (un-phosphorylated) state [59], and structures of active (phosphorylated) TK domain followed [60,61]. The first of these (PDB ID 1IRK) comprised 306 residues (IR978-1283) organised into a two-lobe structure similar to serine kinases like the cyclic-AMP-dependent protein kinase, with an N-terminal lobe made of an α-helix and a β-sheet of 5 antiparallel β strands, and a C-terminal lobe

composed of 8 α-helices and 4 β-strands. The placement of an activation loop where three tyrosine phosphorylation sites (Y1158, Y1162 and Y1163) are located revealed a novel autoinhibitory mechanism where Tyr1162 is bound in the *cis* active site in a way that prevents simultaneous binding of ATP. The second structure (1IR3) shows a major displacement of the activation loop upon phosphorylation of the three tyrosine residues that allows for ATP- and substrate binding. The third structure (1P14) shows that the juxtamembrane region residue Tyr984, which was mutated to phenylalanine in previous studies, represses IRK activity in the basal state by interacting with the N-terminal lobe α-helix and preventing a hinge movement necessary for ATP binding.

Structures of the IR ectodomain relevant to this work were solved later, and are far less numerous than those of the cytosolic region. Post-translational modifications, including a large number of disulphide bonds and heavy glycosylation, combined with generally low expression yields constitute a significant barrier to producing protein of adequate purity, homogeneity and concentration for crystallographic studies. Nevertheless, the Melbourne-based group of Professor Mike Lawrence and Dr Colin Ward published three structures of the *apo* ectodomain [42,62,63], as well as several structures of minimized constructs in complex with insulin and its analogues [64–66]. There is a certain level of redundancy in the *apo* structures: 2HR7 [62] shows L1-CR-L2 domains (non-insulin binding construct IR.1-485), whereas 3LOH [63] represents a re-refinement of 2DTG [42] and both cover the same dimeric IRΔβ ectodomain construct (IR1-917 with a Δ735-753 deletion that removed all O-glycosylation sites) in complex with four Fab molecules, a pair from 83-7 and 83-14 antibodies [67] each, as well as 20 N-terminal residues of an insulin mimetic peptide S519 [68]. The first, 3.8 Å resolution, ectodomain structure revealed an unanticipated antiparallel 'folded over' domain organisation resembling an inverted V (Figure 1.2). This finding is contrary to previous models of IR and IR-insulin interactions based on scanning transmission electron microscopy data which proposed a crossing of the monomers and an insulin-binding pocket created by adjacent apical L1 domains [69,70]. Also surprising is also re-evaluation of the α-β connecting disulphide bond, which appears to be between C647 and C860 in the structure despite strong previous evidence for a C647-C872 cystine [41]. This structure provided a point of reference for further theories of the structural changes occurring upon insulin binding, as well as the model of binding itself: it became apparent how insulin could crosslink a binding surface of the L1 domain of one monomer (binding site 1) and loops in the FnIII-1 domain (site 2) of the other. Neither the mimetic peptide nor most of the insert domain residues (IR.656-754) were resolved in the initial structure 2DTG.

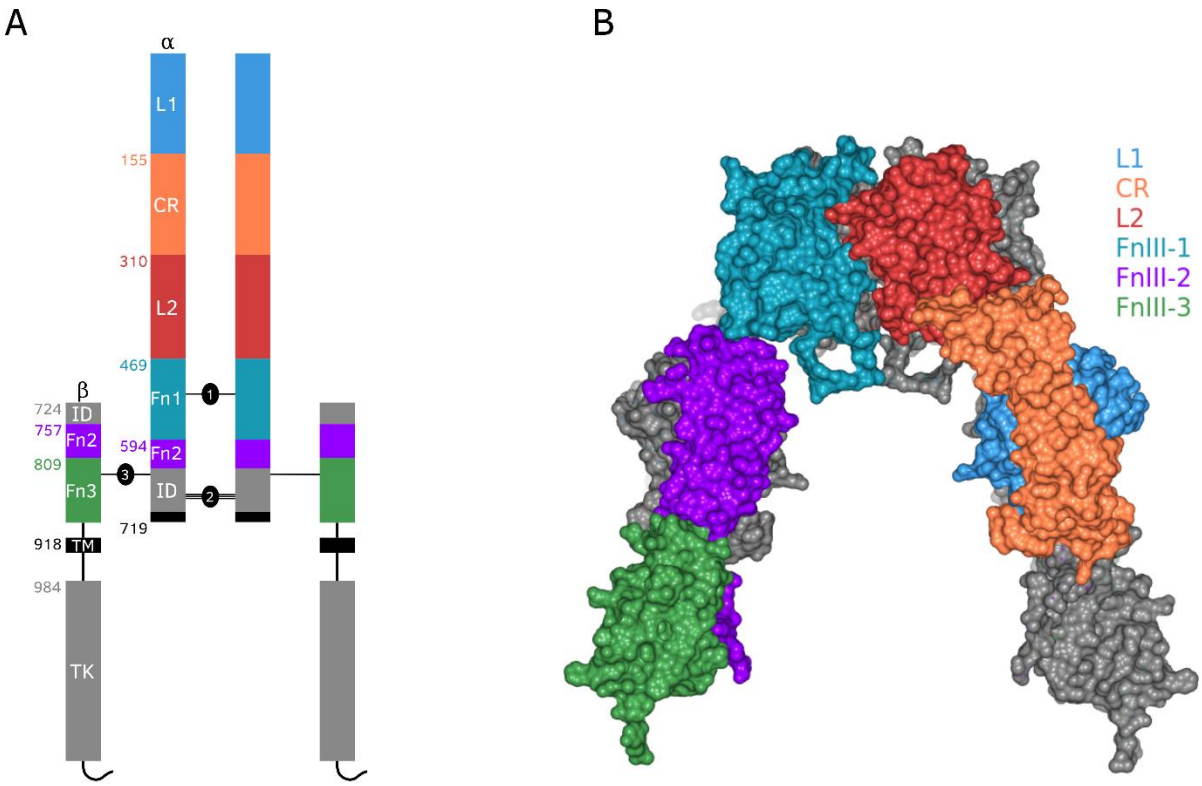

Figure 1.2| Domains and structure of the human insulin receptor.
Panel A| Schematic drawing (domain length to scale) of the insulin receptor dimer. L1 – first leucine-rich domain, CR – cysteine-rich, L2 – second leucine-rich domain, Fn1, 2 and 3 – fibronectin type III domains; ID – insert domain, TM – transmembrane domain, TK – tyrosine kinase. 1, 2 and 3 represent inter-chain disulphide bonds. **Panel B|** Crystal structure of insulin receptor extracellular domain (PDB ID: 2DTG, representation created using ccp4mg [23]). One monomer is represented in grey and the other in colours corresponding to fig. A. Non-coloured parts of panel A (ID, α C-terminus and β subunit beyond FnIII-3) are not visible in the structure.

The re-refined structure 3LOH includes residues 693-710, corresponding to a part of the binding-critical C-terminus of the α subunit (αCT), situated lying across the central β-sheet of L1 domain. These two elements constitute the so-called tandem binding element of site 1, which explains how neighbouring residues of insulin B chain can interact with distal N- and C-terminal parts of IR α subunit. The IR-insulin structures are covered later in the text.

1.3 INSULIN-IR INTERACTIONS

1.3.1 Insulin binding by the IR exhibits complex kinetic properties

Insulin-IR interactions have been studied extensively with radioligand binding analyses and mutagenesis of both IR and insulin [71]. The curvilinear Scatchard plots of ^{125}I-insulin binding were initially thought to represent two sites of varying affinity [72]. In 1973, an alternative explanation was suggested based on the fact that cold insulin accelerates dissociation of ^{125}I-insulin: the two binding sites of the IR exhibit negative cooperativity [73]. The existence of two binding sites in the IR

has also been confirmed by photoaffinity labelling [74] and the fact that purified IR binds only one molecule of insulin with full, sub-nanomolar affinity has been established with crosslinking experiments [75]. Negative cooperativity of the two sites depends on the integrity of IR structure: soluble, non-anchored ectodomain constructs bind two molecules with equal nanomolar affinity and produce linear Scatchard plots [76] while αβ monomers bind one molecule of insulin, also with nanomolar affinity [77]. This has led to the development of a model in which each α subunit has two binding sub-sites (sites 1 and 2) and bound insulin cross-links two sub-sites (site 1 and 2' or sites 1' and 2) from each subunit [78,79]. The site 1 – site 2' crosslink has an apparent affinity of 0.2 nM, an order higher than the affinity of its constituents (6 nM and 400 nM for sites 1 and 2 respectively). This cross-linking would induce an asymmetric conformation and prevent two insulin molecules from simultaneous bridging of a site 1-2' pair. However, binding of a second molecule to the vacant site 1' in insulin/IR complex could still occur and the receptor could oscillate between site 1-2' and 1'-2 cross-linked states. Depending on structural details of the active conformation, a third molecule could bind to the vacant site 2 at high enough concentrations and a 3-insulin state would be unable to switch between active conformations of a 2-insulin state. A mathematical description of possible states of IR-insulin complex termed the 'harmonic oscillator' model has been developed to explain the binding kinetics of insulin, as well as IGF-I and possibly other allosteric receptors activated by ligand cross-linking [80].

1.3.2 Residues involved in insulin-IR binding revealed by biochemical studies

Insights from mutations found in some diabetic patients, as well as alanine scanning mutagenesis, accurately determined which insulin residues were important for IR binding. The classical binding site was identified first and corresponds to residues of the B chain, many of which are involved in insulin dimerization, such as GlyB8, LeuB11, ValB12, TyrB16, PheB24, PheB25, TyrB26, as well as additional contacts in the A-chain: GlyA1, IleA2, ValA3, GlnA5, TyrA19 and AsnA21 [81]. These residues have been termed site 1 and shown to interact with site 1 of IR, which consists of residues of L1 domain [82,83] from one α subunit, and the C-terminus of the other α subunit [84,85]. The second binding site involves residues of the hexamer-forming surface of insulin SerA12, LeuA13, GluA17, HisB10, GluB13, and LeuB17 [79,86] and loops at the C-terminus of the first FnIII domain [42,87,88].

Structure-function studies of insulin analogues provide more insights into the conformational changes in insulin necessary for IR binding. The fact that the AlaA1 – LysB29 cross-linked analogue is completely inactive despite adopting the same structure as wild-type insulin [89] led to the development of a detachment hypothesis stating that residues B22-30 need to rotate away from the

insulin core and expose several nonpolar residues implicated in site 1 binding that are buried in the crystal structures of wild-type insulin, such as IleA2, ValA3 and TyrA19. Consistently, a GlyB24 mutant (of the naturally invariant PheB24 residue) that is a complete IR agonist was found by NMR to have a disordered B20-30 region [30]. A more limited detachment downstream of PheB24 was proposed with the discovery of super-active (~200–400% of wild type insulin affinity), B27-30 truncated analogues that contain an N-methylated B25-26 peptide bond or D-amino acids in the B26 position resulting in a type two β-turn between B24-B26 [90]. Full-length insulin analogues with natural amino-acid substitutions that adopt a B26 turn-like conformation were also described [91]. While their structures show that a B26-30 detachment is plausible in full-length insulin, they do not possess the super-active properties of previous chemically modified analogues (40-80% wild-type insulin activity). Interestingly, one of them (AsnB26 insulin) is more active than wild-type insulin in respect to the IR-B isoform (142%), but not IR-A (83%). The relevance of the T/R structural transition of the B chain N-terminus is less clear. GlyB8 is an invariant residue that is implicated in the switch due to its small side chain and lack of chiral restrictions. L-amino acid substitutions at this position impair folding and disulphide pairing but not necessarily activity as the L-SerB8 mutant is 90% as active as wild-type insulin [92,93]. On the other hand D-amino acid substitutions can stabilize a T-like fold but greatly impair activity [93]. A recent systematic analysis of T/R-like locked analogues substituted at residues B3, B5 and B8 revealed that the R-state likely does not meet requirements for IR binding [94]. At this time it is apparent that the N-terminal flexibility of the B-chain is important for insulin-IR interactions, and that T state represents an inactive storage form, but neither of the conformations adopted by the hexamer can account for the bound state [26,94].

1.3.3 Structure of site 1 complex was solved by X-Ray crystallography

Three papers published recently by the Melbourne group describe the site 1 interactions of insulin and insulin analogues with two IR constructs as well as a hybrid IR/IGF-1R construct [64–66]. The previous structure of the ectodomain dimer (PDB 3LOH) [63] revealed that the C-terminal residues of the α subunit (residues 693-710 in the structure) form an α helix in contact with the central β sheet of the L1 domain, and both the L1 domain and the αCT were previously implicated in IR binding site 1 [82–85]. Four breakthrough structures from 2013 [64] included (i) the 'microreceptor' construct, consisting of IR.1-310 (L1-CR domains) with exogenous αCT(704-719), in complex with human insulin and a Fab fragment of 83-7 antibody [67], PDB 3W11/4OGA (ii) the same μIR in complex with 83-7 Fab and high affinity B26-truncated insulin analogue [D- ProB26]-DTI-NH_2 [90], PDB 3W12 (iii) a μIR consisting of IR.1-310 and a longer αCT(697-719) in complex with 83-7 Fab and [D- ProB26]-DTI-NH_2, PDB 3W13, and (iv) a IR.1-593.704-719 construct (L1-CR-L2-FnIII-1 domains fused to αCT

fragment) in complex with bovine insulin and a Fab fragment of 83-14 antibody [67], PDB ID 3W14. The structures were solved to resolutions between 3.9-4.4 Å and presented a consistent view of the interactions between insulin A chain, B chain central helix, αCT and L1 domain (Figure 1.3). The αCT is displaced on the surface of the L1 domain in respect to the *apo* structure and facilitates crucial insulin-IR contacts. His710 is a key residue that fits into a pocket formed by ValA3, GlyB8, SerB9, and ValB12 while Phe714 fits a pocket of GlyA1, IleA2, TyrA19, LeuB11, ValB12, and LeuB15. L1 residues provide contacts for the αCT peptide as well as B chain residues ValB12 and TyrB16.

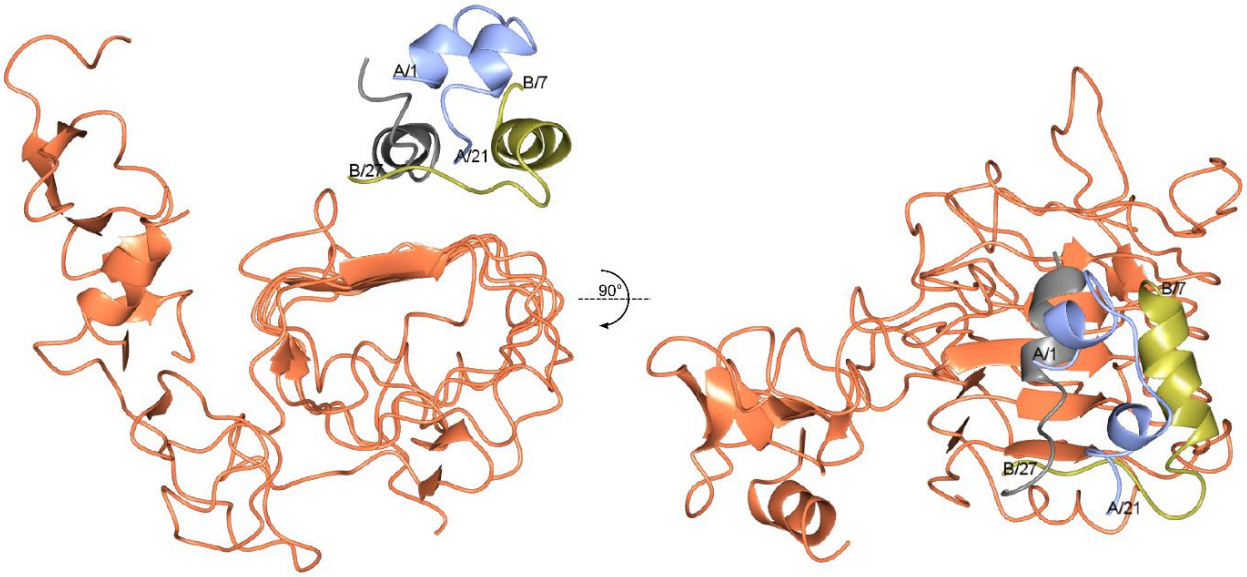

Figure 1.3| Structure of insulin/μIR complex
The structure of insulin in complex with microreceptor construct based on PDB file 4OGA [65]. IR1-310 is drawn in coral, αCT peptide in grey, insulin chain A in light blue and chain B in yellow. Chain A interacts with αCT positioned on the central β sheet of L1 domain and chain B interacts with both L1 and the αCT. B chain C-terminus is detached in respect to inactive insulin conformation (see Figure 1.1) to avoid steric clash with αCT and allow access to A chain site 1 residues. Site 2 interactions are notably absent from this and other constructs. Made with ccp4mg [23].

Neither B chain termini were visible in the four complex structures, but a follow-up paper [65] described a re-refinement of the insulin-μIR complex structure which resolved the B22-B27 segment. A partial detachment hypothesis was confirmed: the B20-27 segment is rotated by about 10° and a more dramatic change can be seen downstream of B24 as the B24-27 segment is rotated by about 50°. The detachment of the C-terminal segment allows for the aforementioned contacts of the IR with A2, A3, and B12 while B20-B27 residues contact both the L1 domain and the αCT. PheB24 contacts mostly the L1 domain (Asn15, Leu37, Phe39, and Phe714), PheB25 inserts between Val715 and Pro718, and the TyrB26 side chain contacts only L1 residues Asp12 and Arg14.

Finally, a 2015 publication presents structures that add coherency and address issues relevant to the previous set [66]. 4XST is a ligand-less structure of the μIR with exogenous αCT(697-719) crystallized in the absence of 83-7 Fab and solved at 3.0 Å resolution. It shows an arrangement of αCT

similar to the one seen in the Fab-stabilized *apo* ectodomain structure 3LOH, in the same position on the L1 β-sheet but with an intermediary rotation in respect to structures of insulin complexes and the *apo* ectodomain. This finding validates the αCT rearrangement seen from 3LOH to 3W11 as a consequence of insulin binding and not construct minimization or antibody influence. The second structure 4XSS is of a hybrid μIR composed of IR1-310 and IGF1R.αCT(691-706) in complex with IGF-I solved to 3.0 Å resolution. IR αβ halves can form heterodimeric hybrid receptors with analogous IGF-1R fragments [95] and these hybrid receptors have higher affinity for IGF-I than insulin [96]. The IGF1RαCT occupies the same position on the IR as the insulin-bound IRαCT, with similar contacts in convergent sequence elements. IGF-1 residues 6-25 (B domain) and 42-52 (A domain) were resolved, equivalent to insulin central B helix and C-terminal portion and the A chain respectively. Both show a similar arrangement as the insulin structure, with a partial detachment of the B domain C terminus into a crevice formed by L1 and IGF1RαCT residues. IGF-1 connecting C domain (unlike insulin IGFs are single chain) is however invisible, and the authors speculate that amino acids 31-40 might loop around the αCT to create an interesting threaded topology.

The known structures of IR ectodomain constructs are summarized in Table 1.1.

Table 1.1| Summary of known IR structures

Construct	Ligand	Fab	Resolution [Å]	PDB	Ref.
IR1-485	-	-	2.3	2HR7	[62]
IR1-917Δ735-753	-	83-7 and 83-14	3.8 3.8	2DTG 3LOH	[42] [63]
IR1-310+αCT (704-719)	human insulin	83-7	3.9 3.3	3W11 4OGA	[64] [65]
IR1-310+αCT (704-719)	[D-ProB26]-DTI-NH2 [90]	83-7	4.3	3W12	[64]
IR1-310+αCT (697-719)	[D-ProB26]-DTI-NH2 [90]	83-7	4.3	3W13	[64]
IR1-593.704-719	bovine insulin	83-14	4.4	3W14	[64]
IR1-310+IGF1RαCT (691-706)	IGF-I	-	3.0	4XSS	[66]
IR1-310+αCT (697-719)	-	-	3.0	4XST	[66]

Known structures of insulin receptor constructs. Construct – amino acid coverage of the employed construct. '+αCT' means the αCT peptide was added exogenously and not part of the main IR construct chain. Fab denotes the antibody binding fragments in the complex, both described in ref. 67.

1.3.4 Site 2 structure and model of activation still a point of contention

The recent IR/insulin complex structures revealed the details of site 1 interactions between insulin, the IR L1 domain and the αCT segment, rationalized a great amount of biochemical data and confirmed key concepts in insulin research such as the detachment of the B chain C-terminus hypothesis. They also determined the two key directions of future IR structural research.

The first of these directions is the recapitulation of the lower affinity binding site 2, and obtaining a comprehensive picture of the site 1 – site 2 crosslink. While mutagenesis studies and the recent structures identified key residues involved in site 1 binding, a significant effort was also undertaken to create IR constructs that reconstitute the high affinity insulin binding. The μIR construct used for nearly all structural studies has an affinity of ~30 nM [64], several orders of magnitude lower than the affinity of the hIR estimated at 5-20 pM [35,79]. Minimized IR constructs published so far include the already mentioned μIR, a soluble ectodomain construct (sIR) with a nanomolar affinity [97], and a monomeric 'minireceptor' (mIR) consisting of residues IR1-468 fused to αCT(704-719). A few constructs include the FnIII-1 domain, most notably the Melbourne group's construct of IR1-593.αCT(704-719) described as 'complex D' in the 2013 paper [64] and Novo Nordisk's mIR.Fn0/Ex10 construct (IR residues 1-601 fused to 650-719) which recapitulates the picomolar affinity of hIR [35]. IR1-593.αCT is the only one of these to be crystallized and have its structure solved, but not only is it lacking several of purported site 2 residues such as Ile602, Lys616, Asp620, and Pro621 [88], the construct also forms a non-physiological tetrameric assembly in the crystal wherein αCT segments of a IR593.αCT homodimer form tandem binding sites with L1 domains of another IR593.αCT homodimer (see supplementary information for [64] and discussion in the comments of [98]).

The second key direction is the 'big picture' of ligand-induced transmembrane activation. Several contradictory models of activation have been postulated recently. Based on their interpretation of *apo* ectodomain and IR593.αCT/insulin structures, as well as FRET experiments on IGF1R, Kavran et al. [98] proposed that upon insulin binding a rotation of L1-CR about a hinge between CR and L2 disrupts the interface between L1 and FnIII2'-3' that forces the transmembrane domain separation at 120 Å apart. This disruption would then result in TM and TK domain proximity and transphosphorylation. An opposite view is presented by Lee et al. [99] based on their very interesting finding that TM-mimicking peptides can activate the IR: they propose that TM domains are held close in the inactive state and their separation upon insulin binding drives IR activation. The authors do not address how this model fits within the known IR structures, especially the ectodomain structures which show a clear separation of the inverted V 'legs' even if the distance between TM domains is less

in the full-length receptor, and instead mention the older cryo-EM structures referenced before [69]. Finally, in a review of novel structural insights from their four insulin/μIR complexes [100], Ward et al. postulate that a yet unknown conformational change abrogates the interaction of juxtamembrane region with TK domains (see [61]) and releases them to autophosphorylate; this is a so-called 'yo-yo' model.

This work seeks mostly to answer the question of full affinity binding, and aims to create novel heterodimeric IR constructs that would be useful to study a site 1-site 2' crosslink structure. This is based on Novo Nordisk's unpublished observations that such constructs' affinity is comparable to that of hIR and their mIR.Fn0/Ex10, the structure of which could not be solved.

2 PROJECT GOALS

The goal of this work was to express and purify a heterodimeric, high-affinity IR construct based on previously unpublished work by Novo Nordisk A/S, containing one copy of binding sites 1 and 2 (Figure 2.1). The purpose of this construct would be primarily to be used in structural studies by means of X-ray crystallography in an attempt to further the understanding of insulin-IR binding to sites 1 and 2' and the asymmetric ligand-induced state of the receptor.

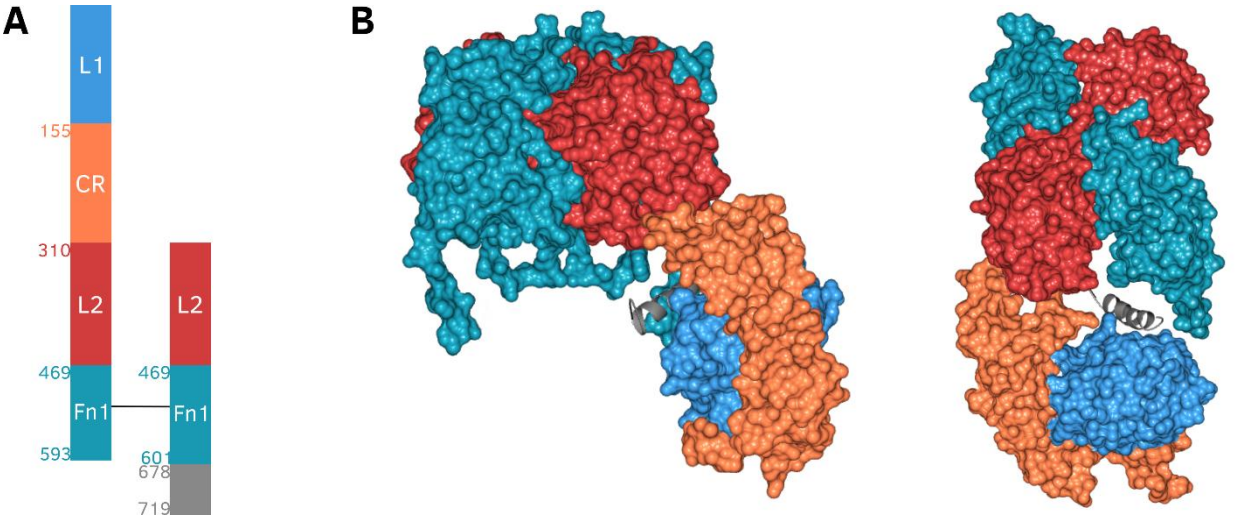

Figure 2.1| Design of the heterodimeric single-site IR constructs
A| Schematic representation of designed heterodimeric IR constructs composed of fragments IR1-593 and IR310-601.678-719 connected with a single cysteine bridge at Cys524. **B|** Projected structure based on 3LOH [63] obtained by deleting fragments of the *apo* ectodomain in front and top views. Colour code corresponds to panel A. Beyond FnIII-1 domain only residues 693-710 of the αCT are visible, represented by a grey α-helix.

To this end, the approach of a single-chain fusion protein with a cleavable linker was chosen. Constructs created on the basis of Novo Nordisk's IM1242 single-chain heterodimeric IR(1-593)-$(GGGS)_5$GGG-(310-601)-(678-731)C682,683,685S-FLAG construct could be cleaved either *in vitro* (using 3C rhinovirus protease) or in the secretory pathway during expression (utilizing a furin cleavage sequence) and purified with common peptide tags 6×His or StrepII.

Figure 2.2| Schematic representation of pIM1242-derived heterodimeric IR constructs
Site-directed mutagenesis was performed on pIM1242 plasmid obtained from Novo Nordisk A/S to yield vectors coding for a single polypeptide chain containing amino acids 1-593 of the human IR, three different linker sequence variants followed by IR amino acids 310-601 and 678-719, followed by two different variants of C-terminal peptide tags and/or IR exon 11 (IR720-731). Disulphide bridge forming cysteines 682,683 and 685 have been mutated to serines. Disulphide-forming Cys524 was left intact in both copies of the FnIII-1 domain.

The initial goal was to create a library of DNA plasmids based on the IM1242 template (Figure 2.2). These would then be assessed in small-scale expression tests and selected constructs expressed on preparative scale and purified. Further goals include polypeptide cleavage, heterodimer formation, activity studies and ultimately structural studies.

3 MATERIALS AND METHODS

3.1 MATERIALS

3.1.1 List of reagents

Reagent	Supplier
Ammonium persulfate	Sigma
30% Acrylamide/Bis Solution, 37.5:1	Fisher
Agar	Lab M Ltd.
Agarose	Melford
Agarose, low melting point	Sigma
Ampicilin, sodium salt	Sigma
Bromophenol Blue	Sigma
Coomassie Brilliant Blue R250	Fisher
Ethanol	Fisher
Glycerol	Fisher
Glycine, electrophoresis grade	Fisons
Imidazole, low UV	Sigma
Isopropanol	Fisher
Methanol	Fisher
Sodium chloride	Melford
Sodium Dodecyl Sulfate (SDS), granulated	Melford
Tris base	Melford
Enzymes and molecular biology reaction buffers	
Q5® buffer, 5X	NEB
ClaI	Promega
CutSmart® buffer	NEB
Dimethyl sulfoxide (DMSO)	NEB
dNTPs, 10 mM ea.	Thermo Fisher
Q5® Polymerase	NEB
SacII	Promega
SYBR® Safe DNA Gel Stain	Thermo Fisher
T4 Ligase	NEB
T4 ligase buffer, 10X	NEB
T4 Polynucleotide kinase	Promega

3.1.2 Antibodies

Mouse monoclonal anti-polyhistidine peroxidase conjugate antibody was supplied by Sigma (catalogue number A7058) and StrepMAB Classic antibody by IBA (2-1502-0010). Primary mouse

83-7 anti-IR antibody was supplied by Merck (MAB1138) and secondary goat anti-mouse IgG peroxidase conjugate antibody by Sigma (A4416).

3.1.3 Bacterial strains and mammalian cell lines

E. coli XL-10 Gold (Stratagene) of phenotype *endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tetR F'[proAB lacIqZΔM15 Tn10(TetR Amy CmR)]* was used for cloning.

Two mammalian cell lines were used for expression: Chinese Hamster Ovary (CHO) K1 cells were obtained from Sigma (catalogue no. 85051005) and Human Embryonic Kidney (HEK) 293T cells were a gift from Dr Jared Cartwright (Biology Technology Facility, University of York).

3.1.4 Growth media

Bacterial media

Bacterial cultures were grown in Miller Luria-Bertani Broth containing 10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl and 1.5g/l Tris (pH 7.2) prepared from pre-made granulated capsules (Melford Laboratories). Solid cultures were grown on 2% agar-LB plates with appropriate antibiotics.

Tissue culture growth media and reagents

Gibco™ reagents were used for all mammalian tissue culture work. Cells were cultured in DMEM/F12 medium with HEPES, L-Glutamine and Phenol Red (catalogue number 31330095) supplemented with 10% Fetal Bovine Serum. 1X Phosphate Buffered Saline pH 7.4 (10010056) and 10X (0.5%) trypsin-EDTA (15400-054) were also used.

3.1.5 Plasmid vectors

The pIM1242 vector used as the initial template for all constructs created with site-directed mutagenesis was obtained from Dr Jakob Brandt, Novo Nordisk A/S.

The pOPINE vector [101] coding for L1-CR-CD4 IR construct was obtained from Dr Tim Ganderton at York Structural Biology laboratory, as was pcDNA3-YFP vector coding for Yellow Fluorescent Protein.

3.1.6 Oligonucleotides

All oligonucleotides were synthesised by Eurofins Genomics. Lyophilized oligonucleotides were dissolved in TE buffer (10 mM Tris pH 8.5, 1 mM EDTA) to stock concentration of 100 μM and stored at -20 °C. A list of oligonucleotides used in this work can be found in Table 3.1.

Table 3.1| List of oligonucleotides used in this study

Non-overlapping primers for site-directed mutagenesis. Melting temperatures for PCR primers calculated using NEB Tm calculator/Q5 Polymerase for template-annealing fragments (uppercase) only.

Name	Sequence (5'->3')	Description	Length (bp)	Tm (°C)	GC (%)
mk5F	TAGTCTAGAGGATCTGGGGTGGCATCC	Remove FLAG tag (5F/5Ra) or FLAG tag and exon 11 (5F/5Rb)	27	71	56
mk5Ra	AGATGGTCTAGGGACGAAAACCACGTTGTGC		31	74	52
mk5Rb	AGATGGTCTAGGGTCCTCGGCACC		24	73	63
mk7F	cctcagttcgagaagggtggcctggaggttctgttccaggggccaCACC TGTACCCCGGAGAGGTGTGTCC	Add N-terminal HLYPGE-6His-GG-StrepII-GG-3C	71	75	63
mk7R	gtggctccagccaccgtggtggtggtgatgatgctctccggggtacagg tgGCCCGCGGCGCCCAGTAGCAG		72	80	69
mk8Fa	cgcaggGGAGGTGGCAAGGTGTGCCAC	Replace part of poly-GS linker with furin site: FnIII-GGG-RKRR-GGG-L2	27	72	70
mk8Ra	tttccgGCCACCACCGGTGGCATCTGT		27	74	63
mk9F	caccaccacTAGTCTAGAGGATCTGGGGTGGCATCC	Add 6×His tag to IR-A (9F/9Ra) or IR-B (9F/9Rb) C-terminus	36	71	58
mk9Ra	gtgatgatgAGATGGTCTAGGGACGAAAACCACG		34	69	50
mk9Rb	gtgatgatgAGATGGTCTAGGGTCCTCGGCACC		33	73	58
mk10F	cagttcgagaagTAGTCTAGAGGATCTGGGGTGGCATCC	Add StrepII tag to IR-A (10F/10Ra) or IR-B (10F/10Rb) C-terminus	39	71	54
mk10Ra	agggtggctccaAGATGGTCTAGGGACGAAAACCACG		37	69	57
mk10Rb	agggtggctccaAGATGGTCTAGGGTCCTCGGCACC		36	73	64
mk11F	ttccaggggccaGGAGGTGGCAAGGTGTGC	Replace part of poly-GS linker with 3C site: FnIII-GGG-3C-GGG-L2	30	68	67
mk11R	cagaacctccagGCCACCACCGGTGGCATC		30	71	67
mk13F	cgcaggtcccttGGAGGTGGCAAGGTGTGC	Replace part of linker with furin site: FnIII-GGG-PSRKRRSL-GGG-L2	30	68	67
mk13R	tttccgagatggGCCACCACCGGTGGCATC		30	71	63
tgF	gaagttctgtttcagggcccgCACCTGTACCCCGGAGAGGTG	Add 6×His-3C sequence to the N-terminus	42	70	62
tgR	caggtgatggtgatgGTGATGGCCCGCGGCGCCCAGTAGCAG		42	80	67

Table 3.1 continued: sequencing primers.

Name	Sequence (5'->3')	Description	Length (bp)	GC (%)
mks1F	CAACCCCTCTGTGCCCCTG	Sequencing primer (5'->3') for C-terminus and 3' flanking sequence (anneals to sequence for IR amino acids 594-599)	19	68
mks1R	CAGGGGCACAGAGGGGTTG	Sequencing primer (3'->5') for linker and second FnIII-1&L2 domains (anneals to sequence for IR amino acids 594-599)	19	68
mks2F	ATGCCTGGGTCCCTGTC	Sequencing primer (5'->3') for first FnIII-1&L2 domains (anneals to sequence for IR amino acids 303-309)	17	65
mks3F	CTAACTGACACACATTCC	Sequencing primer (5'->3') for N-terminus and 5' flanking sequence (anneals to SV40 promoter fragment)	18	44
cv3	CCATGGGCACCGGGGG	Sequencing primer (5'->3') for L1 and CR domains (anneals to start of signal peptide sequence)	16	81
cv14	CCTTCGATGACAGAGCAATTCTCCAG	Sequencing primer (3'->5') for N-terminus and 5' flanking sequence (anneals to bases coding for N-terminal AA in L1)	26	50
M13F	TGTAAAACGACGGCCAGT	Sequencing primer (3'->5') for C-terminus and 3' flanking sequence (anneals to lacZ gene fragment)	18	50

3.2 MICROBIOLOGY AND DNA ANALYSIS

3.2.1 Preparation of competent *E. coli* cells

Based on the method of Inoue et al. [102] *E. coli* strain XL-10 Gold (Stratagene) was streaked on LB-Agar plates with appropriate antibiotics and incubated overnight at 37 °C. One colony was transferred to 3 ml LB medium with appropriate antibiotics and cultured for 12 h at 37 °C. 150 µl of culture was transferred to 250 ml LB with appropriate antibiotics and cultured with shaking at 18 °C and 220 RPM.

The culture was placed on ice for 10 min. when the OD_{600} reached 0.6 and then harvested in sterile centrifuge bottles at 3000 RPM and 4 °C for 10 min. The supernatant was discarded and the cell pellet was gently resuspended by swirling in 80 ml of 55 mM $MnCl_2$, 15 mM $CaCl_2$ and 250 mM KCl solution in 10 mM PIPES-KOH buffer (pH 6.7). The cell suspension was placed on ice for 10 min., centrifuged at 3000 RPM and 4 °C for 10 min. and the pellet was again gently resuspended in 20 ml of the same solution. 1.4 ml of sterile DMSO was added to 7% final concentration and the suspension was placed on ice for 10 min. After incubation on ice, 200 µl aliquots were frozen in liquid N_2 and stored at -80 °C.

3.2.2 Plasmid transformation of *E. coli* cells

Based on the method of Inoue et al. [102] a microcentrifuge tube of competent XL-10 Gold cells was thawed on ice. Approx. 1 ng of plasmid was added to 50 µl of cells and the suspension was gently mixed by flipping. The tube was incubated on ice for 30 min, transferred to a 42 °C water bath for 45 s and incubated on ice for 2 min. 950 µl of sterile LB medium was added and 100 µl was plated on LB-agar plates with relevant antibiotics (for ampicillin, outgrowth phase was generally not performed).

3.2.3 Agarose gel electrophoresis

Agarose was dissolved to a desired concentration in 50 ml of 1x TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA) by heating in microwave. SYBR® Safe DNA Gel Stain was added at a dilution of 1:100 000, and the solution was mixed and poured into a casting tray with comb. The gel was left to solidify and placed in a horizontal electrophoresis cell.

DNA samples were mixed with 6X Gel Loading Dye, Purple (NEB) at a 5:1 ratio, and were loaded onto a gel and separated for 45 min. at a constant voltage of 100 V. Developed gels were visualized under UV light.

3.2.4 Plasmid propagation

Bacterial cells were transformed and grown on LB-Agar plates with 100 µg/ml ampicillin (3.2.2). Chosen colonies were transferred to disposable glass culture tubes containing 5 ml of liquid LB and 100 µg/ml ampicillin and incubated with shaking at 250 RPM for 16 hours at 37 °C. Cultures were stopped and centrifuged for 10 minutes at 4 °C at 5000 RPM. The supernatant was discarded and cellular plasmid DNA was isolated using a QIAprep Spin Miniprep Kit (QIAGEN) following the manufacturer's instructions.

Plasmids to be used for mammalian cell transfection were additionally produced on the midiprep scale. 400 ml of LB medium with 100 µg/ml ampicillin was inoculated with a single colony off a freshly transformed plate and grown overnight at 37 °C and 250 RPM. Cells were harvested by centrifugation at 5000×g for 10 min. at 4 °C. The supernatant was discarded and plasmid DNA was isolated using NucleoBond® Xtra Midi kit (Macherey-Nagel) according to the manufacturer's instructions.

3.3 SITE-DIRECTED MUTAGENESIS

Site directed mutagenesis was used to generate deletions and insertions introducing affinity tags and protease cleavage sites. Q5® High-Fidelity DNA Polymerase (NEB) was used to amplify whole plasmids by PCR with mutagenic primers, followed by enzymatic phosphorylation and blunt-ended ligation to re-close linearized plasmids. PCR primers were designed to be non-overlapping and contain 5' overhangs introducing desired insertions (see Table 3.1 for a list of primers used).

PCR

25 µl polymerase chain reactions were pre-mixed on ice according to Table 3.2 (typically prepared using a multi-reaction master mix). The reactions were transferred to a pre-heated thermocycler and a two-step cycling program was used for whole-plasmid amplification (Table 3.3). 5 µl samples of PCR products were analysed by electrophoresis on a 0.7% agarose gel. If non-specific products appeared (commonly, an ~800 bp secondary product for primers annealing within the polypeptide linker sequence, i.e. mk8, mk11, mk13) the reactions were repeated in 60 µl volume, separated on a 0.7% low melting point agarose gel and the desired bands were excised and purified using a QIAquick Gel Extraction Kit (QIAGEN) according to manufacturer's instructions.

After PCR, 1 µl (10 U) of DpnI (NEB) was added to the reaction mix and the template plasmid was subjected to overnight digestion at 37 °C. The remaining product was purified with a PCR Clean-up Kit (QIAGEN) according to the manufacturer's instructions. The products were quantified by measuring UV light absorbance at 260 nm.

Table 3.2| 25 µl PCR reaction mix for whole-plasmid amplification with Q5 polymerase

	c	v [µl]	c_{final}
Q5 buffer	5x	5	1x
template	~1 ng/µl	1.25	~0.05 ng/µl
dNTP mix	10 mM	0.5	2 mM
Q5® Polymerase	2000 U/ml	0.25	20 U/ml
Forward primer	10 µM	1.25	0.5 µM
Reverse primer	10 µM	1.25	0.5 µM
DMSO	100%	0.75	3%
MilliQ H_2O		14.75	
total		25	

Table 3.3| PCR thermal cycling scheme for two-step whole-plasmid amplification with Q5 polymerase

Temperature	Time	Cycle repetitions	Step
98 °C	pause		Thermal cycler pre-heating
98 °C	1 min.		Initial denaturation
98 °C	15 s	30x	Denaturation
72 °C	4 min.		Extension
72 °C	10 min.		Final extension
4 °C	pause		

Phosphorylation and ligation

1 µl of each purified reaction product was added to 3 µl of MilliQ water, 0.5 µl of 10X T4 ligase buffer (NEB) and 1 µl of T4 Polynucleotide Kinase (Promega). Control reactions were also set up in which the enzyme was replaced with equal volume of MilliQ water. Samples were incubated at 37 °C for 30 minutes. After incubation, 4 µl of MilliQ water, 0.5 µl of T4 ligase buffer and 0.5 µl of T4 ligase (NEB) were added to each reaction, and MilliQ water was added instead of T4 ligase to control reactions. Reactions were incubated at room temperature for 60 minutes.

5 µl of each reaction and control was used to transform 50 µl salt-competent XL-10 Gold cells as previously described. Cells were plated on LB-Agar with 100 µg/ml ampicillin and incubated overnight at 37 °C.

3.4 MAMMALIAN CELL LINE MAINTENANCE

CHO K1 and HEK 293T cells were initially obtained from ATCC and stocks were stored under liquid N_2. Working cell stocks were maintained in Corning CellBIND® T75 flasks in 10 ml DMEM/F-12 medium supplemented with 10% FBS and passaged twice a week at regular intervals. All tissue culture plates, flasks and roller bottles were incubated at 37 °C under 5% CO_2 atmosphere. All media were warmed up to 37 °C before being used in cell culture. All work was performed in a laminar flow hood, sterilized with 10 min. irradiation with UV light and 70% ethanol wipe prior to use, and all equipment and consumables entering the flow hood was sprayed with 70% ethanol.

To passage the cells, media was aspirated and the monolayer was gently washed twice with 10 ml PBS. 2 ml of trypsin-EDTA was added and incubated for approx. 2 minutes (until detachment). 8 ml of DMEM/F-12 with 10% FBS was added and the solution was aspirated and centrifuged for 5 min. at 200×g. The supernatant was discarded and cells were resuspended in 10 ml DMEM/F-12 supplemented with 10% FBS. 1 ml of cell suspension was added to 9 ml of DMEM/F-12 supplemented with 10% FBS in a fresh, sterile T75 flask for a 1:10 split.

3.5 SMALL SCALE MAMMALIAN EXPRESSION

PEI transfection protocols were based on the method of Aricescu et al. [103], mammalian expression guidelines developed in York Structural Biology Laboratory by Dr. Tim Ganderton and optimization experiments performed during the course of this work (see section 4.2).

The generated constructs were used to transiently transfect either CHO K1 or HEK 293T mammalian cells to check for expression. For small scale expression tests, each well of a 6-well plate was seeded with 250 000 cells and grown for 24 h at 37°C/5% CO_2 in 2 ml of growth media supplemented with 10% Fetal Bovine Serum, either DMEM for HEK 293T or DMEM/F10 for CHO K1. Cells were allowed to reach 50-60% confluency (typically after ~24h of growth) before being transfected. Polyethylenimine (PEI) was used as a transfection reagent at varying DNA:PEI ratios: PEI MAX MW 4,000 and linear PEI 25,000 (Polysciences, Inc.) were used with CHO K1 and HEK293T cell lines, respectively.

Transfection

Immediately before transfection, 1 ml of cultured media was discarded and 1 ml of fresh medium not containing FBS was added, lowering the FBS concentration to 5%. In a microcentrifuge tube, an appropriate amount of PEI was added to 200 µl of relevant pre-warmed growth medium and incubated for 10 min. at room temperature. 2 µg of midiprep DNA was added to the medium-PEI mixture, incubated for 10 min. at room temperature and the mixture was added to a single well of a

6-well plate. Valproic acid was added to a concentration of 500 µg/ml. 16h post-transfection, the growth medium was discarded and 2 ml of fresh FBS-free medium was added. The cells were allowed to grow for additional 2 days, after which the cells and media were collected for protein analysis.

Sample preparation

Cells were dislodged with a cell scraper and transferred to a microcentrifuge tube with the cultured media. The samples were centrifuged for 5 min. at 700×g and the supernatant was transferred to a clean microcentrifuge tube. 1 ml of Lysis Buffer (1% CHAPS, 0.1% SDS in 1x TBS pH 7.4) was added to the cell pellets which were briefly vortexed to facilitate lysis. 80 µl of media and lysate samples were added to 20 µl of the appropriate 5x Sample Buffer (reducing, non-reducing, or native) and applied directly to gel or frozen at -20 °C.

3.6 LARGE-SCALE PROTEIN EXPRESSION

Leading up to large-scale expression experiment, cells from working stock cultures were expanded into multiple T175 flasks. T75 working stocks were passaged as per the usual protocol (section 3.4), except 9 ml of final cell re-suspension were evenly divided into four T175 tissue culture flasks (Corning) and 30 ml of DMEM-F12 supplemented with 10% FBS was added. The cells were cultured for 72 h, and expanded in the same way into twenty T175 flasks (1:5 split). After 96h of growth, cells were detached from flasks as previously described, divided evenly into six 2125 cm^2 tissue culture roller bottles (Greiner) with 300 ml DMEM/F-12 supplemented with 10% FBS. Assuming 10^5 cells/cm^2 at confluency, it would translate to a seeding density of 2.7×10^4 per cm^2 of roller bottles (5.8×10^7 cells per bottle).

The cells were grown under standard conditions on a rolling platform, rotating at ~0.7 RPM, and transfected after 48 h. Prior to transfection, FBS concentration was lowered to 5% by exchanging 150 ml media for fresh, FBS-free DMEM/F-12. Per bottle, 2 mg of PEI MAX was incubated at room temperature for 10 min. in 40 ml of FBS-free DMEM/F-12. Subsequently, 500 µg of plasmid DNA was added and incubated for 30 min., and the mix was added to the cells which were then incubated for further 12 h. The media was then exchanged for 300 ml of fresh serum-free DMEM/F-12 and cells were cultured for additional 96 h before harvesting.

3.7 PROTEIN PURIFICATION FROM CULTURED MAMMALIAN MEDIA

Sample preparation

In this study, two constructs were purified from cultured media: mk11 (section 4.4) and mk13 (section 4.5), both of them bearing a C-terminal 6×His tag. Both constructs were purified using a metal affinity chromatography step followed by anion exchange and gel filtration.

Cultured media (300 ml per roller bottle) was harvested and centrifuged at 5000×g for 10 min. The media was concentrated down to ~80 ml and further diafiltrated against 2 l of IMAC binding buffer (25 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole) using a KrosFlo® Research IIi Tangential Flow Filtration System with a 30 kDa cut-off hollow-fibre mPES filter (Spectrum Labs) running at 0.8 ml·min.$^{-1}$ and 8 psi of transmembrane pressure.

Ion Metal Affinity Chromatography

A 1 ml HisTrap FF (GE Healthcare) column was connected to an ÄKTA Pure FPLC system (GE Healthcare) and equilibrated with IMAC binding buffer (25 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole) over 20 CV.

Concentrated and buffer exchanged media was loaded onto the column using a sample pump at a flow rate of 1 ml·min.$^{-1}$ After sample application, the column was washed with 20 CV of buffer. Bound proteins were eluted either with a linear or step gradient to elution buffer (25 mM Tris pH 8.0, 500 mM NaCl, 500 mM imidazole) and the eluate was collected in 1 ml fractions which were analysed with SDS-PAGE and Western Blotting.

Anion exchange chromatography

Protein-containing eluate fractions were diluted 30-fold in 25 mM Tris pH 8.0 buffer, with a final NaCl concentration of 16 mM. A 1 ml HiTrapQ HP (GE healthcare) column was connected to an ÄKTA Pure FPLC system and equilibrated with 25 mM Tris pH 8.0 buffer over 20 CV.

Concentrated and buffer exchanged media was loaded onto the column using a sample pump at a flow rate of 1 ml·min.$^{-1}$ After sample application, the column was washed with 20 CV of equilibration buffer. Bound proteins were eluted either with a linear or step gradient to elution buffer (25 mM Tris pH 8.0, 1 M NaCl) and the eluate was collected in 1 ml fractions which were analysed with SDS-PAGE and Western Blotting. Chosen fractions were pooled and concentrated in a 30 kDa cut-off Vivaspin centrifugal filter unit (GE).

Gel filtration

Gel filtration was performed on a Superdex S200 10/300 GL chromatography column (GE Healthcare) connected to an ÄKTA Pure FPLC system in a 25 mM Tris pH 8.0, 200 mM NaCl buffer at 0.5 ml·min^{-1} flow rate.

3.8 PROTEIN ANALYSIS (SDS-PAGE AND WESTERN BLOT)

Protein samples were subject to analysis on SDS-PAGE under reducing and non-reducing conditions, differing by the presence of 5% β-mercaptoethanol in the sample buffer. Samples were mixed with 5X sample buffer in 4:1 ratio, boiled for 5 min. at 95 °C and typically 20 µl was loaded on discontinuous 4/10% or 4/7.5% gels. The proteins were separated under 200 V for approx. 60 minutes in 1X running buffer. After electrophoresis, gels were stained by incubation with Coomassie-based instant dye (60 mg/L of Coomassie Brilliant Blue G-250 in 35 mM HCl) and de-stained in deionised water. Native PAGE, without SDS in sample buffer or gel formulations, was also routinely used. In these cases, the samples were not boiled and the gels used were continuous 7.5% acrylamide, ran under 100V for ~110 min. Size markers were not used for native PAGE. Gel recipes are presented in Table 3.4.

Table 3.4| SDS-PAGE gel formulation

SDS-PAGE resolving gel			
	5%	**7.5%**	**10%**
H_2O	5.7 ml	4.9 ml	4.0 ml
30% acrylamide/bis-, 37.5:1	1.7 ml	2.5 ml	3.3 ml
Resolving buffer (1.5 M Tris pH 8.8, 0.4% SDS)	2.5 ml		
10% APS	50 µl		
TEMED	8 µl		
Total	10 ml		
SDS-PAGE stacking gel			
H_2O	3.2 ml		
30% acrylamide/bis-, 37.5:1	0.5 ml		
Stacking buffer (0.5 M Tris pH 6.8, 0.4% SDS)	1.3 ml		
10% APS	25 µl		
TEMED	8 µl		
Total	10 ml		
native PAGE continuous 7.5% gel			
H_2O	9.8 ml		
30% acrylamide/bis-, 37.5:1	5 ml		
Resolving buffer (1.5 M Tris pH 8.8)	5 ml		
10% APS	100 µl		
TEMED	8 µl		
Total	10 ml		

Table shows the ingredient volumes for preparation of 2 gels each.

For Western Blotting, proteins were transferred to a PVDF membrane using a BioRad semi-dry blotting system in 1X Towbin buffer for 45 min. under 25V. The membrane was blocked overnight in blocking buffer (5% Bovine Serum Albumin in TBS-Tween).

The membranes were probed either with an IR-specific antibody 83-7 (Life Technologies) or tag-specific antibodies StrepMAB-Classic HRP conjugate (IBA GmbH) and Monoclonal Anti-polyHistidine–Peroxidase antibody (SIGMA). Peroxidase conjugate antibodies were incubated with the membrane for 1 h with shaking at room temperature in 0.5% BSA/TBS-Tween (StrepMAB 1:10 000, anti-His 1:5000), followed by three 15 min. washes in TBS-Tween. 83-7 anti-IR antibody was incubated with the membrane for 1 h with shaking at room temperature in 0.5% BSA/TBS-Tween at 1:2000 dilution, followed by three 15 min. washes in TBS-Tween and incubation for 1h with a secondary goat anti-mouse IgG HRP conjugate antibody (SIGMA) at 1:2000 dilution in TBS-Tween.

Antibodies were detected using the luminol-based Amersham ECL reagent: 0.75 ml of substrates A and B each were mixed, applied to membrane and incubated for 5 min. The chemiluminescent signal was detected using a SynGene G:BOX XT16 imager at fully open aperture and 2 min. exposure time.

5x native PAGE Sample Buffer

Tris-HCl, pH 6.8	250 mM
Glycerol	50% (v/v)
Bromophenol blue	0.25% (w/v)

5x SDS-PAGE Sample Buffer (non-reducing)

Tris-HCl, pH 6.8	250 mM
SDS	10% (w/v)
Glycerol	50% (v/v)
Bromophenol blue	0.25% (w/v)

5x SDS-PAGE Sample Buffer (reducing)

Tris-HCl, pH 6.8	250 mM
SDS	10% (w/v)
Glycerol	50% (v/v)
Bromophenol blue	0.25% (w/v)
β-mercaptoethanol	5% (v/v)

1X SDS-PAGE running buffer

Tris base	25 mM
Glycine	250 mM
SDS	0.1% (w/v)

1X native PAGE running buffer

Tris base	25 mM
Glycine	250 mM

1X Towbin buffer

Tris base	25 mM
Glycine	192 mM
methanol	5% (v/v)

4 RESULTS

4.1 CREATION OF SINGLE-CHAIN HETERODIMERIC IR CONSTRUCTS BY SITE-DIRECTED MUTAGENESIS

A plasmid based on the pZem vector [104], containing the sequence of a single-chain heterodimeric IR construct with a non-cleavable 25 AA linker pIM1242 [signal sequence-(1-593)-$(GGGS)_5$GGG-(310-601)-(678-731)C682,683,685S-FLAG], was obtained from Dr Jakob Brandt, Novo Nordisk A/S (Figure 4.1).

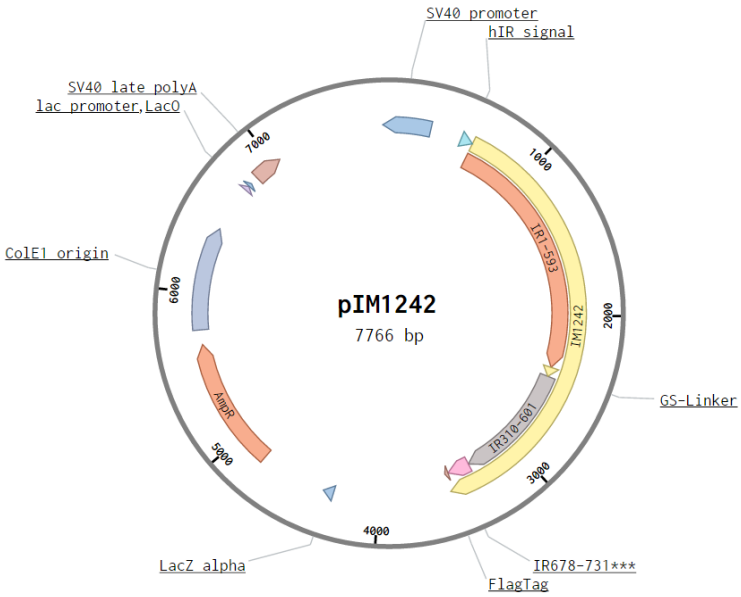

Figure 4.1| Map of pIM1242 plasmid
pIM1242 contains the single-chain, heterodimeric IR construct IM1242 sequence cloned into Novo Nordisk's pZem vector under the control of SV40 promoter sequence. The construct contains a native human IR signal sequence, a non-cleavable $(GGGS)_5$GGG linker and a C-terminal FLAG tag.

Site-directed mutagenesis, by whole-plasmid PCR followed by phosphorylation and ligation of the linear product (section 3.3) was undertaken to introduce deletions, insertions and substitutions. After site directed mutagenesis, plasmid DNA was isolated from bacterial colonies and endonuclease digestions were used to select clones that behaved consistently with the expected size and location of restriction sites (Figure 4.2). Afterwards, chosen clones were sequenced at GATC Biotech, covering the mutagenesis region, to select the correct mutants. Due to the number of constructs and length of sequence (2991 bp for IM1242), only a limited number of constructs have been sequenced in full (specifically mk11-14 and mk35).

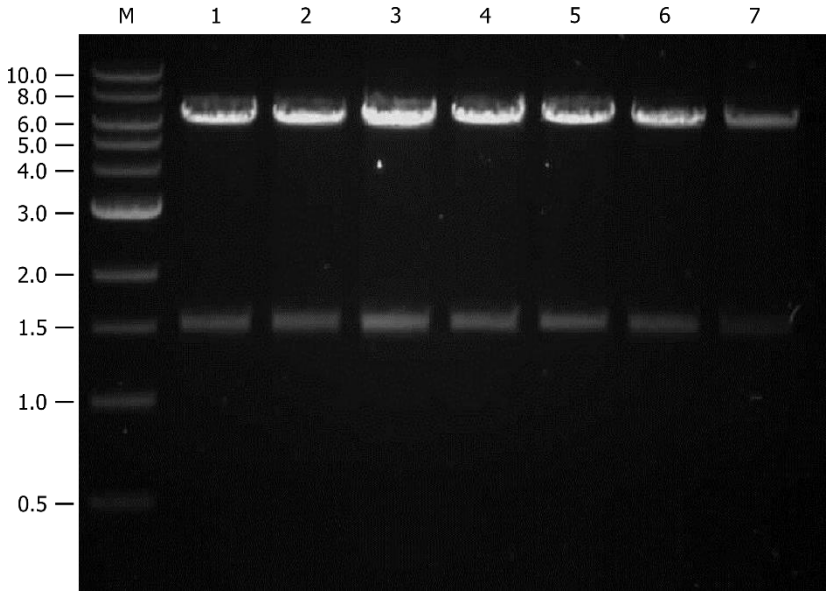

Figure 4.2| Example digests of plasmids isolated after mutagenesis with restriction endonucleases ClaI and SacII
Three clones of pZem-mk1 (lanes 1-3) and pZem-mk2 (lanes 4-6) each were isolated from single bacterial colonies after site-directed mutagenesis and subjected to a 3-hour double digest using SacII and ClaI restriction endonucleases. Lane 7 shows parent vector pIM1242 cleaved with the same enzymes. M – NEB 1kb ladder, sizes in kilobases. Expected fragment sizes after digestion: 6180 and 1526 bp for mk1, 6216 and 1526 bp for mk2, 6240 and 1526 bp for pIM1242. First clones of both mk1 and mk2 (lanes 1 and 4) were sequenced in the mutagenesis region and confirmed to have the desired mutation.

Series of mutagenic primers annealing to pIM1242 (Table 3.1) were designed to:

a) remove the C-terminal FLAG tag,
b) create an isoform A construct lacking exon 11,
c) introduce a 6×His or a StrepII peptide tag on the C-terminus,
d) substitute part of the poly-GS linker for a 3C rhinovirus protease or a furin cleavage site,
e) introduce a cleavable N-terminal 6×His tag,
f) substitute the native signal peptide with heterologous sequences from human interleukin-2 and *Gaussia princeps* luciferase.

A combinatorial approach to mutagenesis was taken (Figure 4.3), generating constructs presented in Table 4.1.

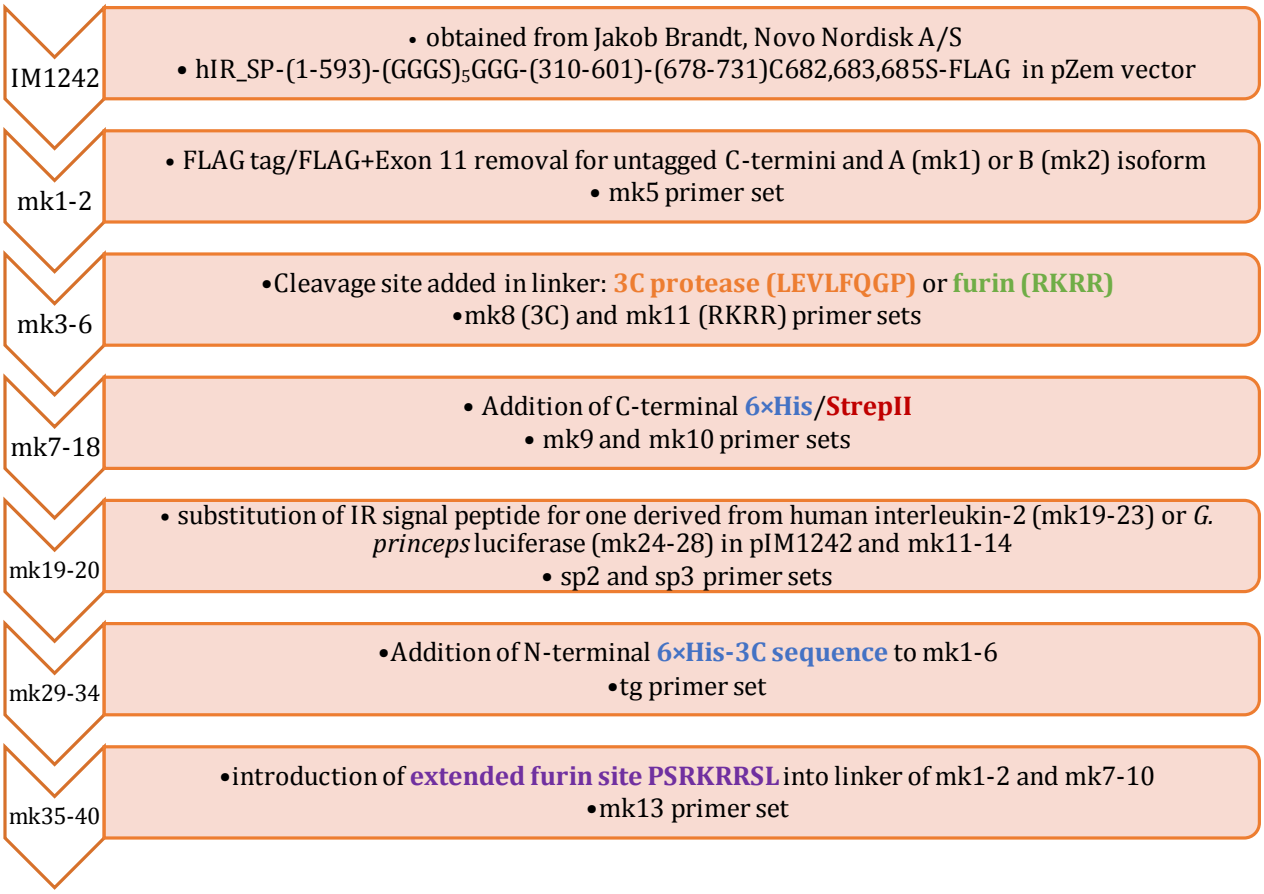

Figure 4.3| Summary of mutagenesis process
An overview of the mutagenesis steps taken to generate given constructs from Table 4.1 is given, with a summary of added features and primers (see Table 3.1) used for each mutagenesis reaction.

Table 4.1| List of constructs generated on the basis of pIM1242

Name	Signal	Linker	CT isoform	NT tag	CT tag	Sequenced with
mk1	hIR	$(GGGS)_5GGG$	A			M13F
mk2	hIR	$(GGGS)_5GGG$	B			M13F
mk3	hIR	GGG-RKRR-GGG	A			mks1R
mk4	hIR	GGG-RKRR-GGG	B			mks1R
mk5	hIR	GGG-3C-GGG	A			mks1R
mk6	hIR	GGG-3C-GGG	B			mks1R
mk7	hIR	$(GGGS)_5GGG$	A		His	M13F
mk8	hIR	$(GGGS)_5GGG$	A		StrepII	M13F
mk9	hIR	$(GGGS)_5GGG$	B		His	M13F
mk10	hIR	$(GGGS)_5GGG$	B		StrepII	M13, mks1R
mk11	hIR	GGG-RKRR-GGG	A		His	M13F, mks1R, cv14, cv3, mks2F
mk12	hIR	GGG-RKRR-GGG	A		StrepII	M13F, mks1R, cv14, cv3, mks2F

Table continued from page 39

mk13	hIR	GGG-3C-GGG	A		His	M13F, mks1R, cv14, cv3, mks2F
mk14	hIR	GGG-3C-GGG	A		StrepII	M13F, mks1R, cv14, cv3, mks2F
mk15	hIR	GGG-RKRR-GGG	B		His	M13, mks1R
mk16	hIR	GGG-RKRR-GGG	B		StrepII	M13, mks1R
mk17	hIR	GGG-3C-GGG	B		His	M13, mks1R
mk18	hIR	GGG-3C-GGG	B		StrepII	M13, mks1R
mk19	IL-2	GGG-RKRR-GGG	A		His	mks3F
mk20	IL-2	GGG-RKRR-GGG	A		StrepII	mks3F
mk21	IL-2	GGG-3C-GGG	A		His	mks3F
mk22	IL-2	GGG-3C-GGG	A		StrepII	mks3F
mk23	IL-2	$(GGGS)_5GGG$	A		FLAG	mks3F
mk24	luc	GGG-RKRR-GGG	A		His	mks3F
mk25	luc	GGG-RKRR-GGG	A		StrepII	mks3F
mk26	luc	GGG-3C-GGG	A		His	mks3F
mk27	luc	GGG-3C-GGG	A		StrepII	mks3F
mk28	luc	$(GGGS)_5GGG$	A		FLAG	mks3F
mk29	hIR	$(GGGS)_5GGG$	A	His-3C		mks3F
mk30	hIR	$(GGGS)_5GGG$	B	His-3C		mks3F
mk31	hIR	GGG-RKRR-GGG	A	His-3C		mks3F
mk32	hIR	GGG-RKRR-GGG	B	His-3C		mks3F
mk33	hIR	GGG-3C-GGG	A	His-3C		mks3F
mk34	hIR	GGG-3C-GGG	B	His-3C		mks3F
mk35	hIR	GGG-PSRKRRSL-GGG	A			M13F, mks1R, mks3F, cv3, mks2F
mk36	hIR	GGG-PSRKRRSL-GGG	B			mks1R
mk37	hIR	GGG-PSRKRRSL-GGG	A		His	mks1R
mk38	hIR	GGG-PSRKRRSL-GGG	A		StrepII	mks1R
mk39	hIR	GGG-PSRKRRSL-GGG	B		His	mks1R
mk40	hIR	GGG-PSRKRRSL-GGG	B		StrepII	mks1R

Table includes all constructs generated during the course of this work. Name – construct name in mkxx format; signal – N-terminal signal sequence: hIR – native, IL-2 – human interleukin-2, luc – *G. princeps* luciferase (see main text for details); linker – linker sequence between FnII-1 (first repeat) and L2 (second repeat) domains, 3C – 3C protease site LEVLFQGP; CT isoform indicates presence (B) or absence (A) of exon 11 at the C-terminus; NT/CT tag – affinity tags present at either terminus. Last column indicates the extent of sequencing for each construct by listing primers used in sequencing reactions (see Table 3.1 for details on primers).

First constructs: mk1-mk18

First, the FLAG-tag was removed from pIM1242 yielding mk1 (untagged IR-A C-terminus) and mk2 (IR-B) to allow for use of more economically viable affinity resins in purification.

Subsequently, 3C protease (LEVLFQGP [105]) or furin (RKRR present between α and β subunits of native IR [39]) cleavage sites were added to yield mk3-mk6. C-terminal 6×His or StrepII [106] peptide tags were added to mk1-2 to create non-cleavable linker constructs mk7-10, and to mk3-6, to yield constructs mk11-18, which contained both cleavable linkers as well as C-terminal peptide tags.

Alternative signal peptides

Subsequently, the native IR signal sequence was replaced in pIM1242 and mk11-14. N-terminal signal peptides direct transmembrane and extracellular proteins to the secretory pathway where they are proteolytically removed [107]. These peptides can influence protein expression and/or secretion levels [108,109] and therefore two alternatives to the native signal sequence were tested:

a) the signal peptide derived from human interleukin-2 precursor (UniprotKB P60568), a sequence commonly used in industrial production of recombinant biopharmaceuticals [108], amino acid sequence `MYRMQLLSCIALSLALVTNS`;
b) the signal peptide derived from marine copepod *Gaussia princeps* luciferase precursor (UniProtKB Q9BLZ2), which has been shown to increase protein secretion levels compared to albumin signal peptide for recombinant proteins like human endostatin [110], amino acid sequence `MGVKVLFALICIAVAEA`.

N-terminal tagging

An attempt was made to introduce an N-terminal HLYPGE-HHHHHH-StrepII-3C sequence with the rationale of creating N-terminally double-tagged constructs. Based on unpublished observations from Novo Nordisk that introducing a FLAG-tag between the signal peptide and the first 6 N-terminal amino acids can significantly lower expression levels, it was also attempted to duplicate these amino acids before the tags using the mk7 primer set. The PCR could not be optimized for these primers, likely owing to their propensity to create both hetero- and homodimers. A gradient of 3-6-9% DMSO concentrations and 62-65-68-72 °C annealing temperatures, as well as one-primer PCR followed by a two-primer PCR strategy were all attempted and failed to yield any product.

Using the 'tg' primer set, a simpler N-terminal tag was introduced instead with a HHHHHH-LEVLFQGP sequence added to the N-terminus of mk1-6 after the signal peptide (constructs mk29-34).

Extended furin cleavage sequence

Furin is a major subtilisin-like proprotein convertase ubiquitous in the Golgi apparatus and secretion pathway [111]. Furin cleaves after consensus site R-X-K/R-R, and RKRR site is present in the human insulin receptor between α and β subunits [39] and was hence introduced into a number of previous constructs as an attempt to engineer an IR heterodimer processed in the secretion pathway. It has been reported that the presence of this site is however not sufficient for efficient furin cleavage which requires two polar regions outside of the core sequence [112]. PiTou, a bioinformatic tool for prediction of furin cleavage sites which analyses a 20 amino acid motif [113], was used to score possible variants of furin site in the context of the poly-glycine linker in the heterodimer construct (Table 4.2). It was found that an addition of 2 amino acids from each side of native hIR furin cleavage sequence to the basic RKRR motif results in PiTou score as high as the native cleavage site between hIR subunits. Thus, the PSRKRRSL site was introduced into the linker of mk1-2 and mk7-10 to create mk35-40, six constructs with the same linker and alternative C-termini (A/B isoforms untagged/StrepII-tagged/His-tagged).

Table 4.2| PiTou scoring of furin linker variants

Sequence	Motif-P14-P1	Motif-P1'-P6'	PiTou Score
native IR α/β site	HNVVFVPRPSRKRR	SLGDVG	14,0595
Fn1-GGG-RKRR-GGG-L2	YVQTDATGGGRKRR	GGGKVC	6,98861
Fn1-GGGSRKRRSGGG-L2	VQTDATGGGSRKRR	SGGGKV	10,2293
Fn1-GGGPSRKRRSGGG-L2	QTDATGGGPSRKRR	SGGGKV	10,8181
Fn1-GGGSRKRRSLGGG-L2	VQTDATGGGSRKRR	SLGGGK	13,7215
Fn1-GGGPSRKRRSLGGG-L2	QTDATGGGPSRKRR	SLGGGK	14,3103

Amino acids flanking native IR furin site (purple) were added around RKRR site in the sequence of heterodimer construct mk11 and the altered sequences were scored with PiTou furin site prediction software [113] until a sequence with a score at least equal to wild-type hIR was found.

4.2 EVALUATION OF EXPRESSION AND DETECTION CONDITIONS

Detection

mk11 was the first construct to be expressed in 6-well plate cultures of CHO K1 cell line using 1:1, 3.5:1 and 6:1 PEI:DNA ratios (w/w). mk11 is present in cultured media and when not reduced, it can be detected by the 83-7 monoclonal antibody [67] which recognizes an epitope within the CR domain (Figure 4.4). Because the antibody appears to be more sensitive when the proteins are first separated with native rather than denaturing PAGE (compare panel B, where faint signal appears only for one transfection sample, to panel C and Figure 4.6 below), this method has been used for all other analyses of expression media containing IR constructs. Anti-His antibody produced no signal, raising questions about the integrity and accessibility of His-tagged C-terminus.

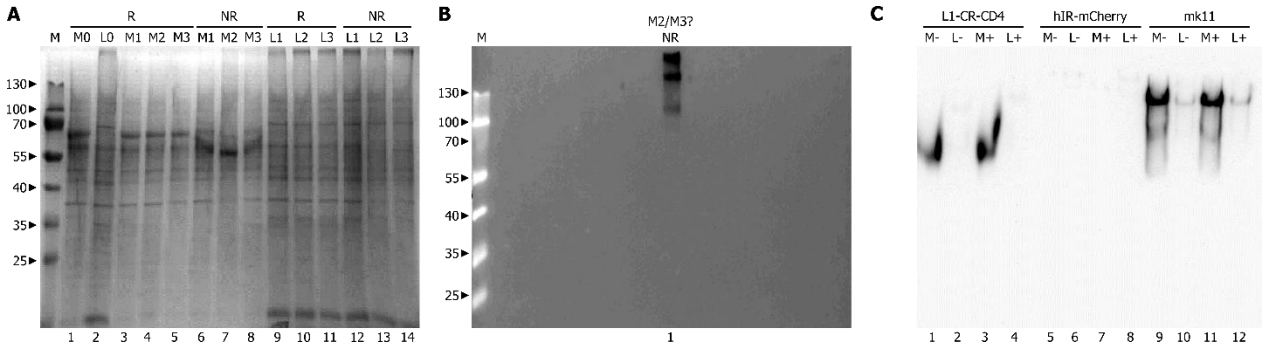

Figure 4.4| Detection of mk11 in CHO K1 media and lysate samples

Panel A| CHO K1 cells were cultured in a 6-well dish and transfected with 2 µg of pZem-mk11 (section 3.5). Cultured media (lanes 1 and 3-8) and lysate (lanes 2 and 9-14) samples were analysed on 10% SDS-PAGE under reducing (lanes 1-5 ad 9-11) and non-reducing (lanes 6-8 and 12-14) conditions. Three PEI:DNA ratios were used (1:1 in lanes 3, 6, 9, 12; 3.5:1 in lanes 4, 7, 10, 13; 6:1 in lanes 5, 8, 11, 14). Lanes 1 and 2 come from untransfected cells. **Panel B|** Equivalent gel was blotted onto a PVDF membrane and developed using the 87-3 anti-IR antibody. Only one sample produced any signal, most likely equivalent of gel lane 7 or 8; this signal is very weak, resulting in a high background when compared to panel C. This experiment was repeated in HEK 293T cells with the same result (not shown). Repeat blots were probed against with anti-His antibody, which did not produce any signal. **Panel C|** CHO K1 cells were transfected with pZem-mk11 (lanes 9-12) at 4:1 PEI:DNA ratio. L1-CR-CD4 construct (lanes 1-4) and full-length hIR with C-terminal mCherry fusion protein (lanes 5-8) expression vectors were obtained from Dr Tim Ganderton and used as controls; YFP coding vector was used for visual control of transfection efficiency. Cultured media (odd numbered lanes) and lysate (even numbered lanes) samples were applied to 7.5% native PAGE and probed with the 83-7 antibody. Cells were optionally treated (+/- lanes) with valproic acid after transfection and no change in protein levels was observed.

Transfection and expression

Initially, PEI MAX MW 4,000 (catalog number 24885) from Polysciences Inc. was used to transfect CHO K1 cells and linear PEI MW 25,000 (catalog number 23966) was used to transfect HEK 293T cells according to established lab practice. A small-scale transfection of both cells with a pcDNA3-YFP vector coding for Yellow Fluorescent Protein showed similar transfection efficiencies for both reagents in CHO K1 cells. However, higher levels of cell detachment and death were observed in HEK 293T cells at typical harvest time using the higher MW linear PEI reagent (Figure 4.5). Thus, PEI MAX was used for further experiments in both cell lines.

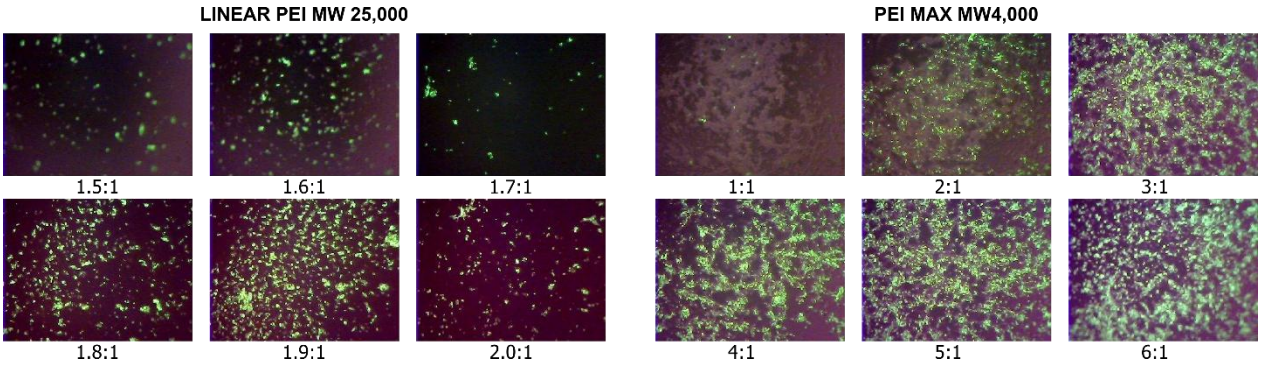

Figure 4.5| HEK 293T cells 72h post-transfection with a pcDNA3-YFP vector with two different PEI reagents
HEK293T cells were cultured in two 6-well plates and transfected according to small-scale expression protocol (section 3.5) with 2 µg of pcDNA3-YFP plasmid. Two different reagents were used (PEI MAX MW 4,000 and linear PEI MW 25,000 from Polysciences, Inc.) in concentrations routine for these reagents, expressed in PEI:DNA (w/w) ratios under each well photo. Photos taken at 10x magnification. More cells expressing YFP can be seen using PEI MAX as transfection reagent, and greater cell detachment and cell death can be seen using the linear PEI MW 25,000 (compare mostly un-transfected, but almost confluent culture when 2 µg PEI MAX is used, to mostly transfected but more sparsely populated cultures when linear PEI is used).

With mk11 as the reference construct, a range of 1:1-1:6 DNA:PEI MAX ratios was investigated: CHO K1 and HEK 293T cells grown in 6-well plates were transfected using different ratios of DNA to PEI and samples of cultured media were analysed after 72h by native PAGE followed by Western Blotting using the 83-7 anti-IR antibody (Figure 4.6, panel A). For both cell lines the signal appears stronger at PEI:DNA ratio greater than 3:1. A ratio of 4:1 was chosen for further experiments with these constructs to limit PEI-related cytotoxicity as increased PEI amount do not seem to increase protein in media. Another factor that can affect transfection efficiency and protein yield is confluency at the time of transfection, with various sources stating between 50 and 90% as the optimal confluency [103,114]. This factor was investigated indirectly by transfecting CHO K1 and HEK 293T cells with mk11 plasmid 24 hours after seeding at different cell densities: 2, 3, 4, 5, 6 or 7 $\times 10^4$ cells per cm^2 of a well in a 6-well culture plate. Samples of cultured media were analyzed after 72 h by native PAGE followed by Western Blotting using the 83-7 anti-IR antibody (Figure 4.6, panel B). Increases in protein levels in media can be seen up to a seeding density of 5×10^4 cells$\cdot cm^{-2}$ which was used for further experiments.

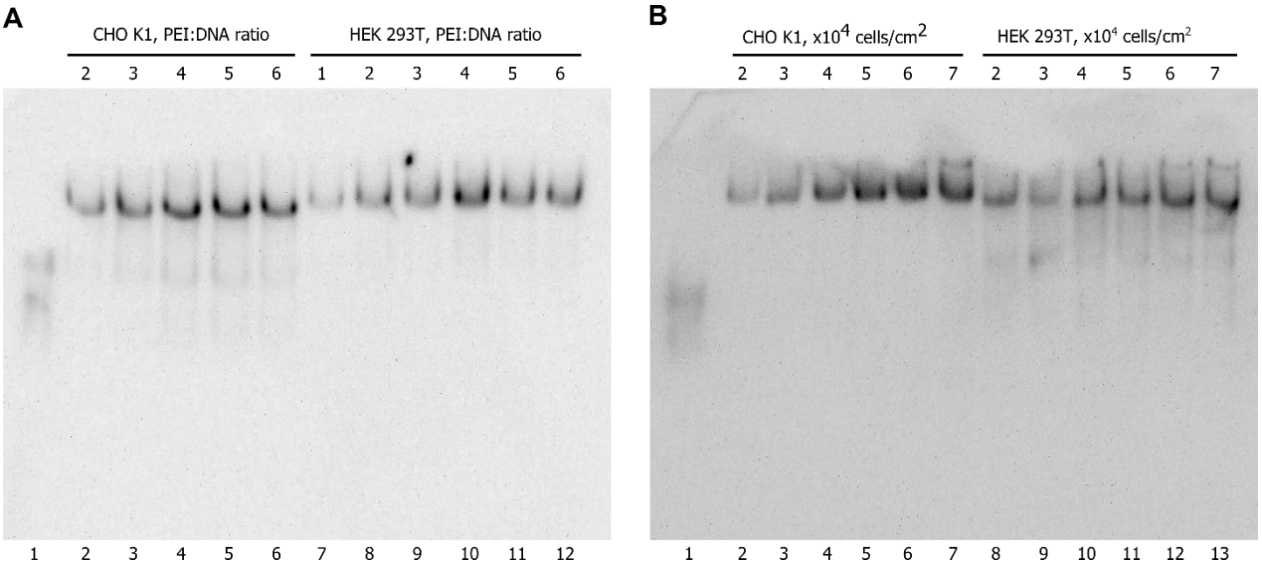

Figure 4.6| Influence of PEI:DNA ratio and cell seeding density on mk11 levels in cultured media

CHO K1 and HEK 293T cells were cultured in a 6-well dish and transfected with 2 µg of pZem-mk11. Samples of 72 h cultured media were subjected to native PAGE and Western Blotting with the 83-7 antibody. Sample of cultured media from CHO K1 cells expressing L1-CR-CD4 construct obtained from Dr Tim Ganderton (YSBL) was also included as a positive control. **Panel A|** The influence of PEI:DNA ratio on Western Blot signal. CHO K1 and HEK 293T cells were transfected using increasing amounts of PEI in each well (expressed in PEI:DNA ratio at 1:1, 2:1, 3:1, 4:1, 5:1 and 6:1). Lane 1 – L1-CR-CD4 cultured media, lanes 2-6: mk11/CHO K1 cultured media, lanes 7-12: HEK 293T/mk11 cultured media. CHO K1 1:1 PEI:DNA sample was lost due to a bacterial infection. **Panel B|** The influence of cell seeding density. Both cell lines were seeded at increasing cell densities per well (2, 3, 4, 5, 6 or 7 $\times 10^4$ cells per cm^2) 24 h before transfection. Lane 1 – L1-CR-CD4 cultured media, lanes 2-6: CHO K1/mk11 cultured media, lanes 7-12: HEK 293T/mk11 cultured media.

4.3 CONSTRUCT EVALUATION

4.3.1 mk11-mk14, expression and detection

mk11, mk12, mk13 and mk14 were the first constructs to be expressed on a small scale (Figure 4.7). All of these lack exon 11, have either a 6×His (mk11 and mk13) or a StrepII (mk12 and mk14) tag on the C-terminus and contain either a furin site RKRR (mk11 and mk12) or a 3C protease site LEVLFQGP (mk13 and mk14) in the linker sequence.

The constructs are present in the expression media in amounts comparable to the L1-CR-CD4 construct expressed from a pOPINE vector [101]. L1-CR-CD4 is based on the IR.1-310 construct used previously in structural studies [64–66] and became a reference standard in our lab.

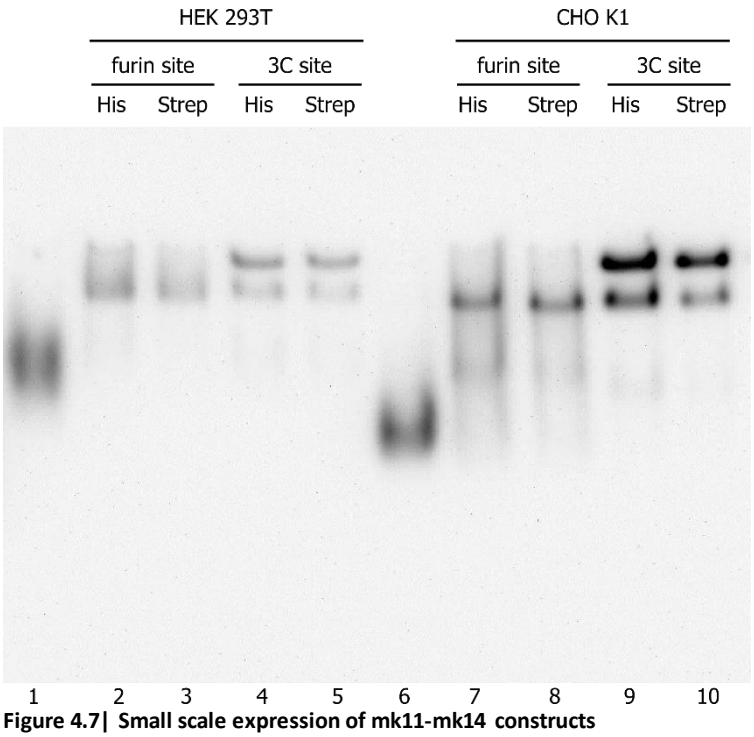

Figure 4.7| Small scale expression of mk11-mk14 constructs
L1-CR-CD4 (lanes 1 and 6), mk11 (lanes 2 and 7), mk12 (lanes 3 and 8), mk13 (lanes 4 and 9) and mk14 (lanes 5 and 10) were expressed in HEK 293T and CHO K1 cells cultured in 6-well plates (section 3.5). The media samples were applied to a 7.5% non-reducing, non-denaturing PAGE, transferred to a PVDF membrane and probed with the 83-7 IR-specific antibody. L1-CR-CD4 is included as a positive control for 83-7 and size reference. Furin-linker constructs appear as one band with significant smearing, whereas 3C-linker constructs appear as two bands. Repeat blots were probed with anti-His (SIGMA) and StrepMAB Classic (IBA) HRP-conjugate antibodies, but produced no signal.

The constructs could be detected in media only with the 83-7 monoclonal antibody that recognizes a conformational epitope in the CR domain [67]. Neither the anti-His nor StrepMAB Classic HRP conjugate antibodies produced any signal. This could be caused by C-terminal degradation, inaccessibility of the C-terminal tag to antibodies, or simply lower sensitivity of tag-specific, HRP-conjugated antibodies than the 83-7 primary/secondary system.

This failure to detect signals from tag-specific antibodies presents a major limitation of Western Blots in studying these constructs in cultured media samples. All IM1242-derived constructs have two copies of the FnIII-1 domain fragment containing a free cysteine that forms an inter-subunit disulphide bond in native IR (Figure 2.2), as an attempt to create a heterodimeric IR construct connected by these domain fragments. Previously, anti-His antibodies did not produce signals on blots from SDS-PAGE, with or without a reducing agent in the sample buffer. In this experiment, tag-specific antibodies produced no signal on blots from native PAGE; the experiment was also repeated using PAGE without SDS but with β-Me in the sample buffer and none of the three antibodies (anti-His, StrepMAB Classic or 83-7) produced any signal. This made it impossible to compare the behaviour of these constructs under reducing and non-reducing conditions, which could have revealed what kind of homo- and heterodimers are formed before choosing constructs for large-scale expression.

Interestingly, on native PAGE mk11 and mk12 constructs behave differently to mk13 and mk14: mk11 and mk12 run as a single band with more smearing, whereas mk13 and mk14 run as two distinct bands, possibly indicating higher order oligomers owing either to differences in processing (due to the furin site present in mk11/mk12), or a linker with greater length or flexibility more accommodating of oligomer formation.

4.3.2 Alternative signal peptides

Five constructs: pIM1242 and mk11-14 were mutated to contain alternative signal peptides derived from human interleukin-2 and *G. princeps* luciferase (section 4.1), giving constructs mk19-mk28 (Table 4.1). The expression of these constructs was assessed with small-scale transfections and native PAGE followed by Western Blotting of cultured media (Figure 4.8).

Prior to this experiment, samples of cultured media from BHK cells stably expressing IM1242 and CHO Lec8 cells stably expressing IR 'midireceptor' construct IM1010 (IR1-601.678-731.C682S.C683S.C685S-FLAG) were obtained from Novo Nordisk A/S. They were included in each expression blot for concentration and size comparisons.

Neither of the two tested signal peptides improved levels of secreted protein in the media for any construct. Comparison of signal from these expression screens to IM1242 and IM1010 reference samples suggests that higher levels of secreted proteins were achieved in the HEK 293T cell line (Figure 4.8: compare signal in lanes 4, 7, 10 vs. lane 1 and lanes 16, 19, 22 vs. lane 15 on both blot pairs).

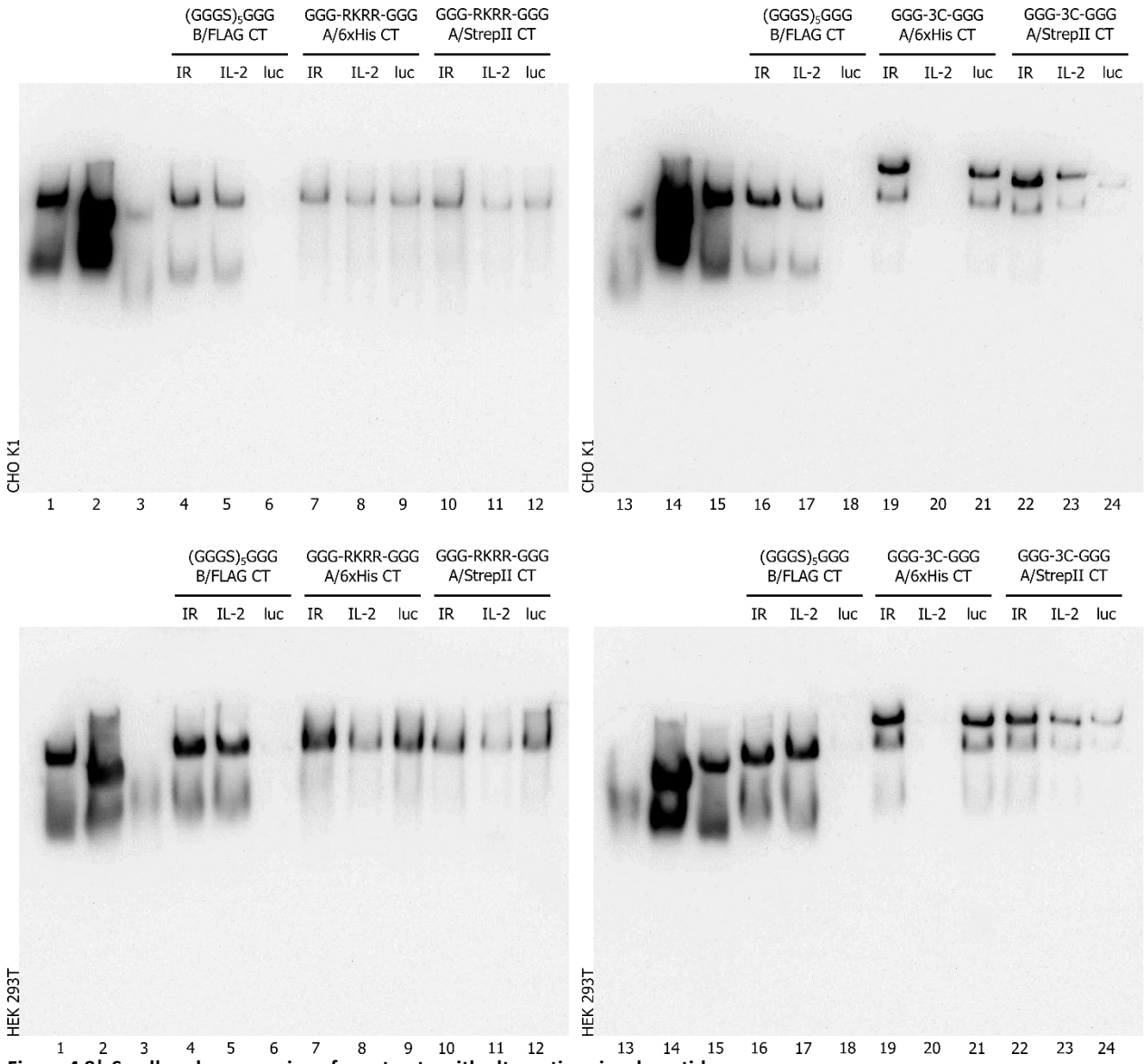

Figure 4.8| Small-scale expression of constructs with alternative signal peptides
IM1242, mk11, mk12, mk13 and mk14 with signal peptides derived from native IR sequence (*IR*), human interleukin-2 (*IL-2*) or *G. princeps* luciferase (*luc*) were expressed in 6-well plate cultures of CHO K1 (above) and HEK 293T (below) cell lines, separated on a 7.5% native PAGE and blotted with the 83-7 IR-specific antibody. Every blot includes L1-CR-CD4 expressed during the course of the same experiment as an internal positive control (lanes 3 and 13) as well as cultured media from BHK cells stably expressing IM1242 (lanes 1 and 15) and CHO Lec8 cells stably expressing IM1010 homodimeric 'midireceptor' construct (lanes 2 and 14) obtained from Novo Nordisk A/S as external reference for size and concentration. IM1242-derived constructs appear twice (lanes 4-6 and 16-18). Lanes 7-9 are mk11 and its derivatives (mk19 and mk24), 10-12 mk12 and derivatives (mk20 and mk25), 19-21 mk13 and derivatives (mk21 and mk26) and 22-24 mk14 and derivatives (mk22 and mk27). Descriptions above lanes refer to signal peptide, linker sequence, isoform and C-terminal tag of the construct.

Constructs that include IR Cys524, for example a mIR.Fn0 construct which comprises of IR residues 1-601 followed by residues 704-719, have been shown to exist as a mixture of dimeric and monomeric forms [35] and both IM1010 and IM1242 exhibit the same double band behavior. The single band observed previously for mk11 and mk12 (Figure 4.7) has the same mobility as the dimer band of IM1242 (Figure 4.8: compare lanes 7, 10 vs. lane 4) whereas mk13 and mk14 are recognized

by the antibody as three bands: fastest migrating band similar to IM1242 monomer; middle band similar to IM1242 dimer; and the slowest migrating band, presumably of higher oligomeric species or aggregate (Figure 4.8: compare lanes 19, 22 vs. lane 16).

4.3.3 N-terminal tags and extended furin site constructs

Two later sets of constructs include the N-terminally His-tagged constructs mk29-mk34 as well as constructs that include an extended furin cleavage site PSRKRRSL.

The N-terminal 6×His-3C tag in constructs mk31 and mk33 does not affect the amount of protein in media compared to their C-terminally tagged equivalents mk11 and mk13 (Figure 4.9). Mk35, an untagged A-isoform construct that includes the extended furin cleavage site PSRKRRSL in its linker sequence, behaves drastically different to IM1242 or mk11 and does not appear as a single band but rather a smear running similarly to a IM1242 monomer; this would suggest that this construct undergoes a greater degree of processing (possibly indicating no cleavage of RKRR-site constructs and close to full cleavage for PSRKRRSL-site constructs) resulting in a greater heterogeneity of the oligomeric species or a monomer of greater flexibility in the native environment.

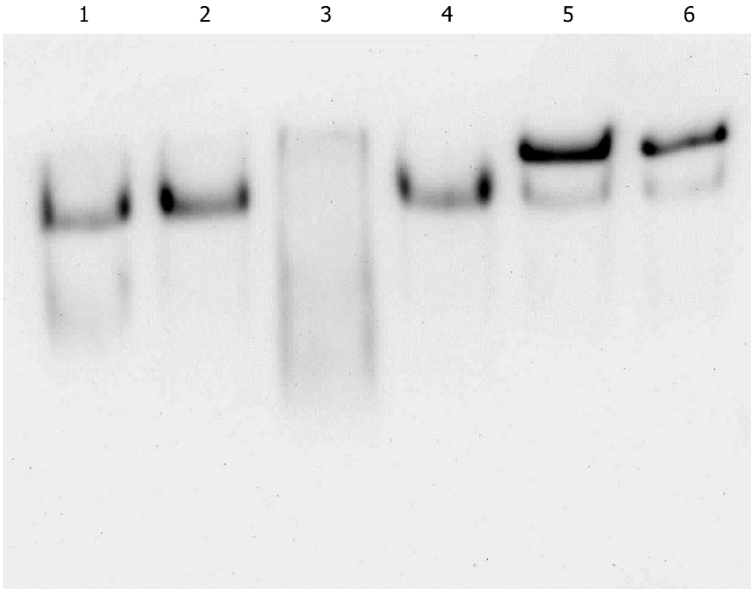

Figure 4.9| Small scale expression of constructs with N-terminal tags and extended furin cleavage site
IM1242 (lane 1), mk11 (lane 2), mk35 (lane 3), mk31 (lane 4), mk13 (lane 5) and mk33 (lane 6) were expressed in a 6-well plate culture of HEK 293T cells and the cultured media samples were analysed by native PAGE followed by blotting with the 83-7 antibody. mk31 and mk33 are N-terminally tagged equivalents of mk11 and mk13, respectively, and mk35 is an untagged A-isoform construct with extended furin site in linker (GGG-PSRKRRSL-GGG).

4.4 MK11: LARGE SCALE EXPRESSION AND PURIFICATION OF A CONSTRUCT WITH A FURIN SITE LINKER

mk11 is an A-isoform construct containing a C-terminal 6×His tag and a linker sequence with furin cleavage site RKRR (Figure 4.10) in a pZem expression vector.

mk11		**mk11 cleaved at RKRR↓**			
		N-terminal		C-terminal	
length [aa]	943	length [aa]	600	length [aa]	343
MW [Da]	107975	MW [Da]	68787	MW [Da]	39206
pI	6.69	pI	7.02	pI	6.34
$\varepsilon_{0.1\%}$ [(mg/mL)$^{-1}$cm^{-1}]	1.40	$\varepsilon_{0.1\%}$ [(mg/mL)$^{-1}$cm^{-1}]	1.35	$\varepsilon_{0.1\%}$ [(mg/mL)$^{-1}$cm^{-1}]	1.48

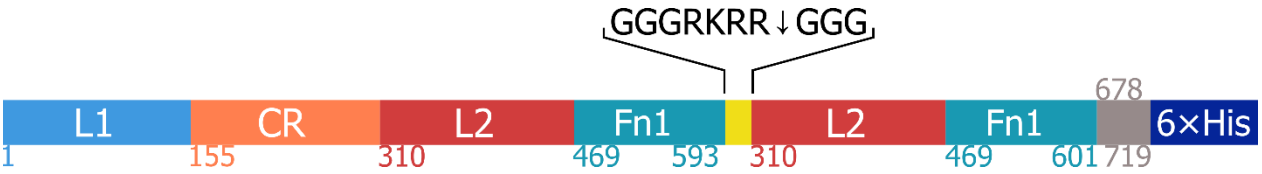

Figure 4.10| mk11 construct map and properties
mk11 is an IR(1-593)-GGGRKRRGGG-IR(310-601)-IR(678-719)-6×His heterodimeric construct with C682S, C683S, C685S mutations. MW – molecular weight of polypeptide chain (excluding glycans), pI – isoelectric point, $\varepsilon_{0.1\%}$ - extinction coefficient at 280 nm [(mg/mL)$^{-1}$cm^{-1}]. Parameters were calculated using ProtParam [115].

mk11 was expressed according to section 3.6, but two roller bottles were lost to bacterial infection leaving 1.2 l of viable cultured media. The sample was loaded onto a 1 ml HisTrap FF column using an ÄKTA Pure FPLC system, washed with binding buffer over 20 CV and eluted with a linear gradient of 20-500 mM imidazole over 20 CV (Figure 4.11). Collected fractions were analysed by native and denaturing PAGE as well as Western Blotting.

SDS-PAGE of eluate fractions shows a very heterogeneous protein mixture, with no discernible band likely to represent mk11 (Figure 4.12, panel A). Signals produced on a Western Blot by anti-His antibodies was similarly non-specific and inconclusive, and based on these as well as previous results the use of anti-His antibodies for these constructs is not recommended. However, the 83-7 antibody confirmed the presence of IR epitope in eluted fractions separated under native conditions (Figure 4.12, panel B).

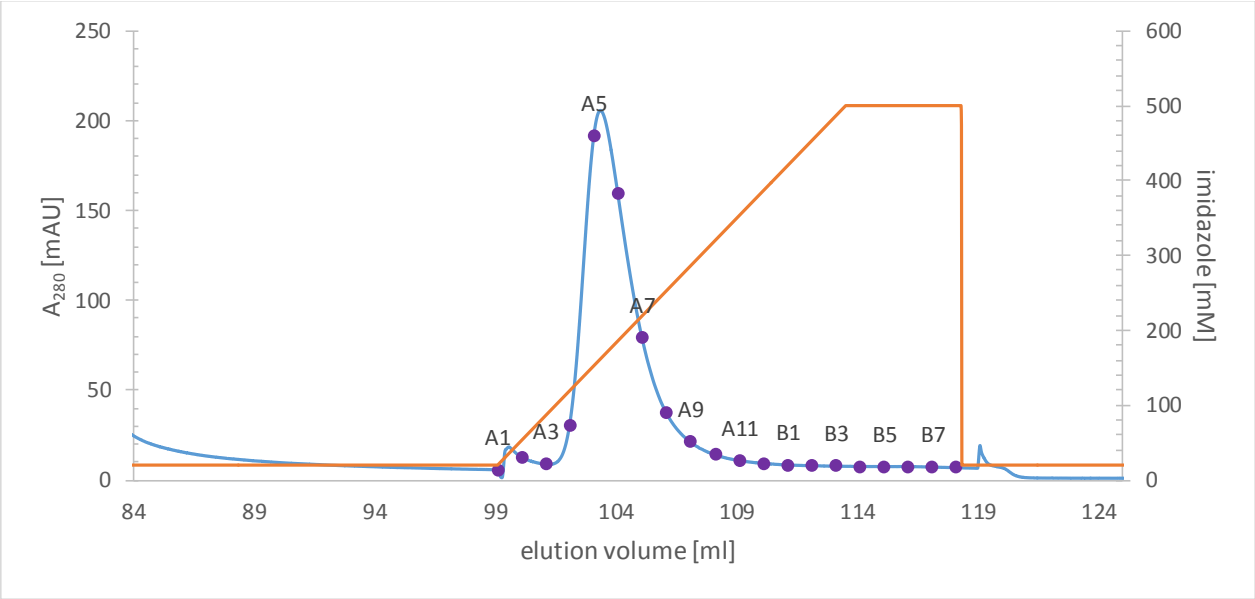

Figure 4.11| Elution of mk11 from a nickel affinity resin
mk11 was loaded onto a 1 ml HisTrap FF column, washed with 20 mM imidazole over 20 CV and eluted with a linear gradient of 20-500 mM imidazole over 20CV. 1 ml fractions were collected (shown by purple points with labels above every other fraction) and subjected to gel electrophoresis and western blotting.

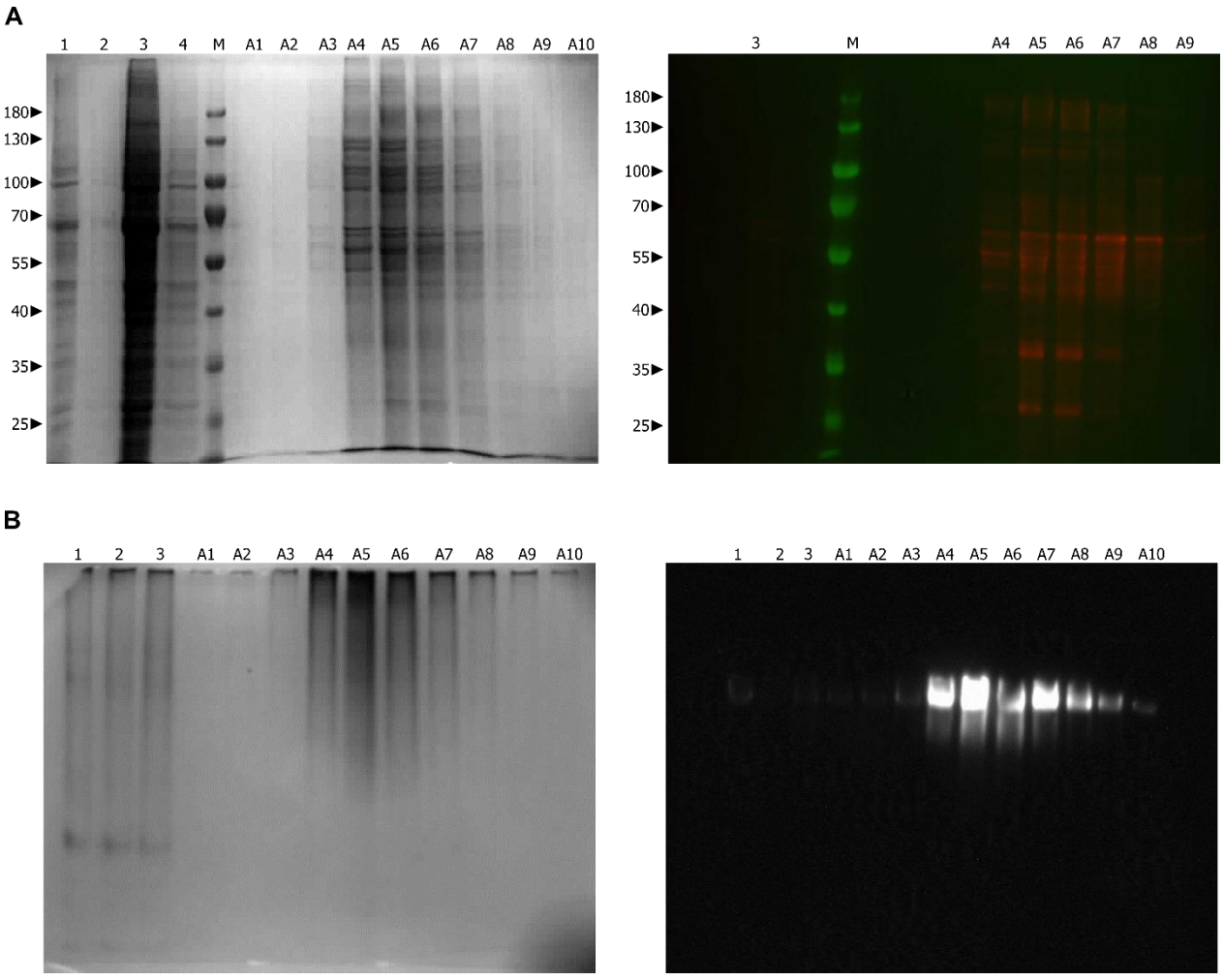

Figure 4.12| Electrophoretic and Western Blot analysis of mk11 His-tag purification
Samples from mk11 purification steps were subjected to PAGE (left) and Western Blot (right) analysis. **Panel A|** Purification samples were separated by SDS-PAGE (left) on a 10% acrylamide gel under reducing conditions (5% β-Me in sample buffer). The samples were transferred to a PVDF membrane and probed with monoclonal anti-His HRP conjugate antibody (right). 1 – expression medium, 2 – tangential flow filtration permeate, 3 – column flow-through, 4 – column wash, A1-A10 – elution fractions (see Figure 4.11). Green channel – visible light signal, red – chemiluminescence signal. **Panel B|** Purification samples were separated by native, non-reducing PAGE (left), transferred to a PVDF membrane and probed with 83-7 (anti-CR) antibody (right). 1 – expression medium, 2 – column flow-through, 3 – column wash, A1-A10 – elution fractions (see Figure 4.11).

Eluate fractions A4-A8 were pooled and diluted in 150 ml of 25 mM Tris pH 8.0 buffer (to a final NaCl concentration of 16 mM). The sample was loaded onto a 1 ml HiTrapQ HP anion exchange column, washed with 20 CV of 25 mM Tris pH 8.0 buffer and eluted with a linear 0-500 mM NaCl gradient over 20 CV (Figure 4.13).

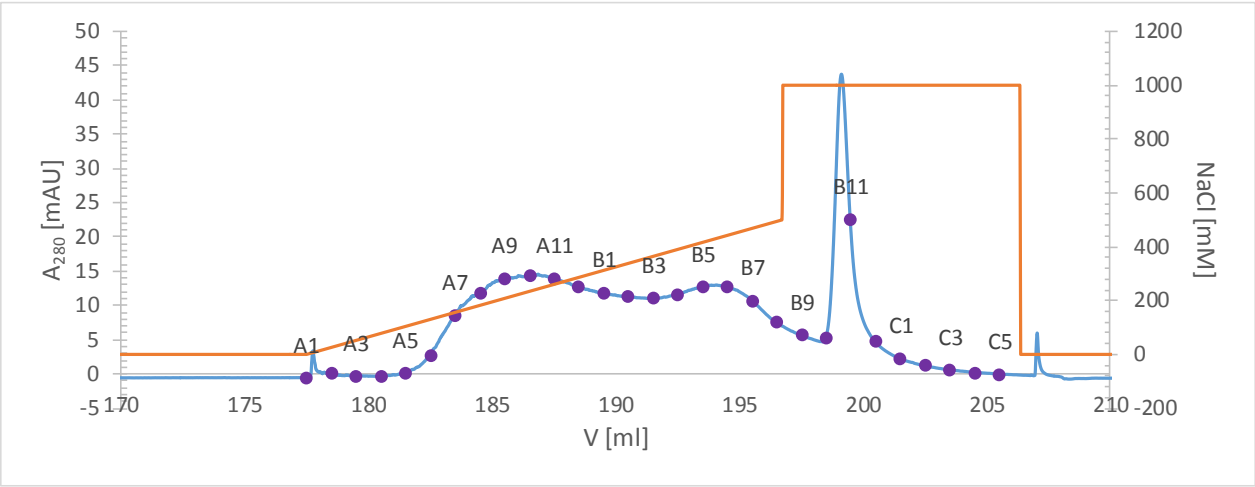

Figure 4.13| Elution of mk11 from an anion exchange resin
Combined fractions from IMAC purification applied to a 1 ml HiTrap Q HP, washed with equilibration buffer over 20 CV and eluted with a linear gradient of 0-500 mM NaCl over 20CV. 1 ml fractions were collected (shown by purple points with labels above every other fraction) and subjected to gel electrophoresis and western blotting (Figure 4.14).

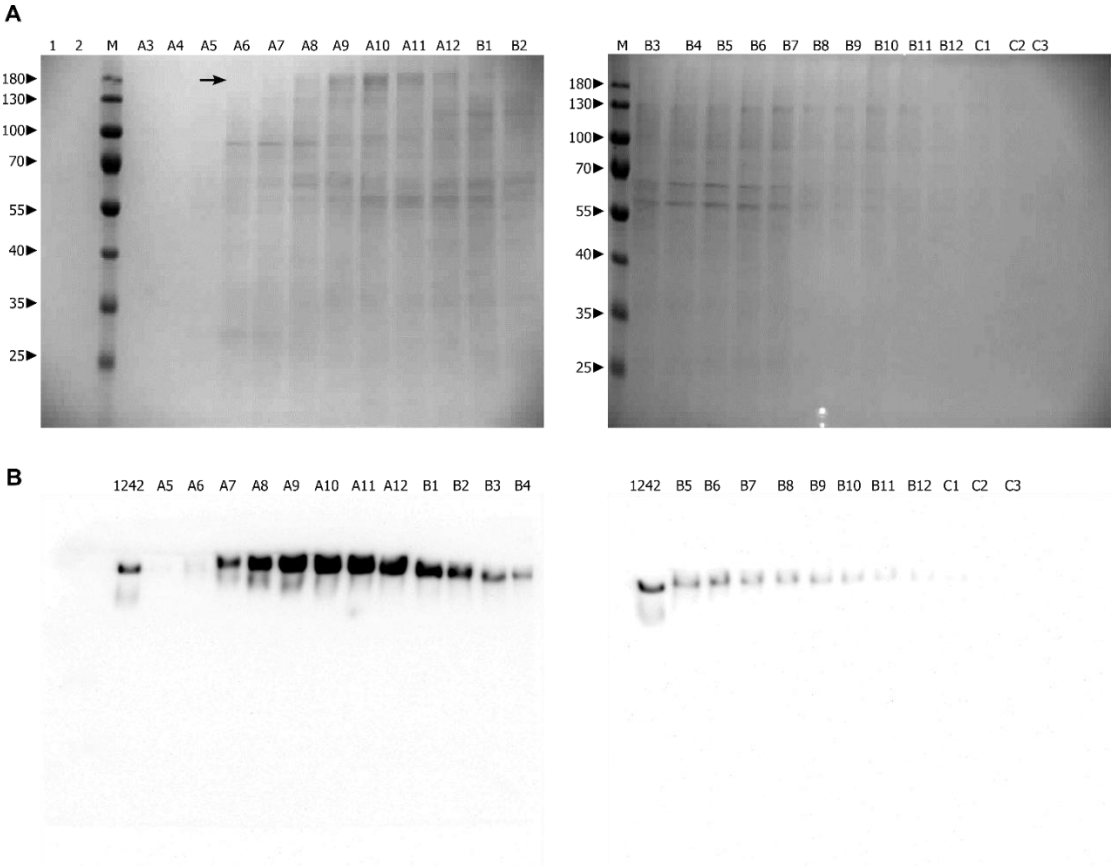

Figure 4.14| SDS-PAGE and Western blot analysis of mk11 purification using anion exchange column
After His-tag purification, mk11 was further purified using a 1ml HiTrap Q anion exchange column and purification samples were analyzed using SDS-PAGE and Western Blot analysis. **Panel A|** Purification samples were separated by SDS-PAGE on a 10% acrylamide gel under reducing conditions (5% β-Me in sample buffer). 1 – column flow-through, 2 – column wash, M – PageRuler prestained protein ladder, A3-C3 – elution fractions (see Figure 4.13). Suspected mk11 band marked with an arrow. **Panel B|** Purification samples were separated by native, non-reducing PAGE (gels not shown), transferred to a PVDF membrane and probed with 83-7 (anti-CR) antibody. 1242 – sample of expression medium containing IM1242 (from small scale screen, 4.3.2), A5-C3 – elution fractions (see Figure 4.13).

Proteins bound to the anion exchange elute in two broad, overlapping peaks, one corresponding to approximately 200 mM NaCl and another around 400 mM NaCl. Western Blots show that mk11 is present primarily in the first peak (fractions A7-B1, Figure 4.14, panel B) and comparing its abundance in each fraction to SDS-PAGE gel, it is likely that the band of nearly 180 kDa marked with an arrow in Figure 4.13 represents mk11 (with glycosylation accounting for apparent mobility being lower than expected from calculated molecular mass alone).

Fractions A7-B1 were pooled and concentrated to approximately 100 µl in a 30 kDa cut-off Amicon concentrator; A_{280} was measured at 0.720 giving a yield of approximately 72 µg total protein, assuming that 1 AU = 1 mg/ml for a low purity sample. This sample was applied to both reducing and non-reducing SDS-PAGE in an attempt to determine whether *in vivo* furin cleavage was taking place (Figure 4.15). The rest of the sample was applied to a Superdex 200 10/300 GL gel filtration column in a non-reducing buffer (25 mM Tris pH 8.0, 200 mM NaCl).

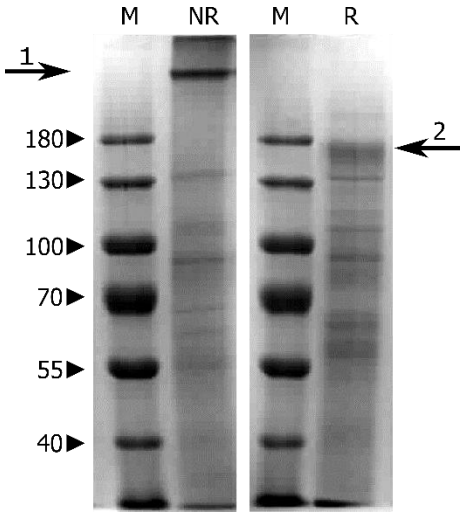

Figure 4.15| mk11 after purification using IMAC and anion exchange resins under reducing and non-reducing conditions. After anion exchange purification step, concentrated sample was separated with SDS-PAGE under reducing (R) and non-reducing (NR) conditions. Pictured are fragments of two 7.5% acrylamide gels differing by addition of 5% β-mercaptoethanol in sample buffer. Arrows 1 and 2 mark bands suspected to be the mk11 construct. M – PageRuler Prestained Protein Ladder.

In both reducing and non-reducing conditions, mk11 appears to be a single band with mobility similar to un-cleaved mk13 (analogous construct containing a 3C-cleavable linker, see Figure 4.24). A band that runs well above the 180 kDa marker band under non-reducing conditions disappears upon addition of 5% β-mercaptoethanol in SDS-PAGE sample buffer and a less homogeneous band just below 180 kDa appears.

In gel filtration using the Sephadex S200 GL10/300 column under non-reducing conditions, a similar high molecular weight protein appears in a peak at 11.3 ml, but a number of impurities persist even after gel filtration (Figure 4.16, Figure 4.17).

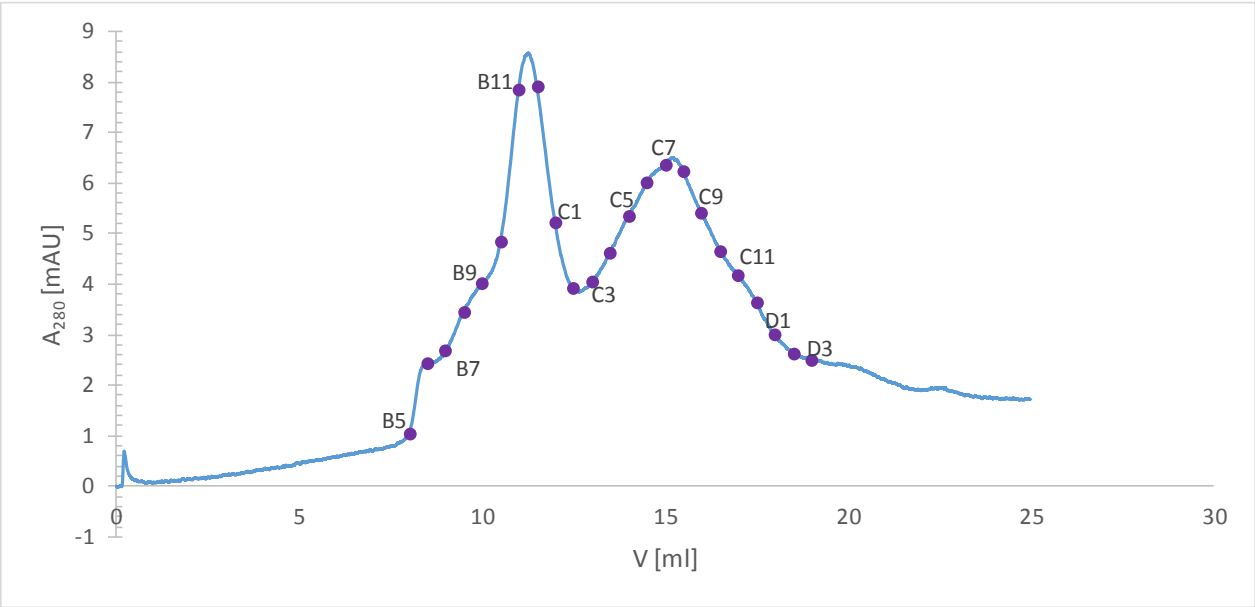

Figure 4.16| Elution profile of mk11 from a Superdex 200 10/300 GL column
After IMAC and anion exchange purifications, mk11 sample was applied to a Superdex 200 10/300 GL gel filtration column. 0.5 ml fractions were collected (marked by purple points with labels above every other fraction).

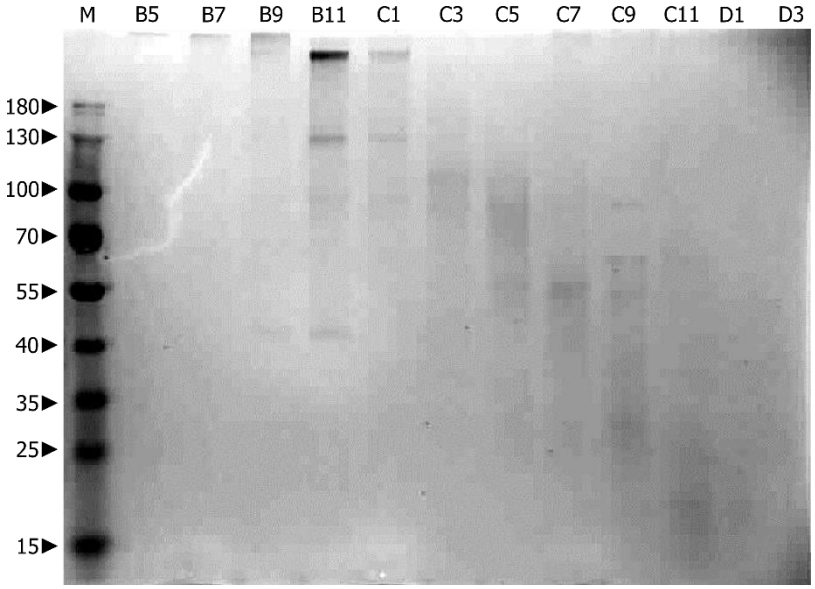

Figure 4.17| SDS-PAGE analysis of mk11 gel filtration
After IMAC and anion exchange purifications, mk11 sample was applied to a Superdex 200 10/300 GL gel filtration column and 0.5 ml fractions were collected (Figure 4.16). Every other fraction from B5-D3 was applied to a 10%% SDS-PAGE gel under non-reducing conditions.

4.6 MK13: LARGE SCALE EXPRESSION AND PURIFICATION OF A 3C-CLEAVABLE CONSTRUCT

mk13 is an A-isoform construct containing a C-terminal 6×His tag and a linker sequence with 3C cleavage site LEVLFQGP (Figure 4.18) in a pZem expression vector.

mk13		mk13 cleaved at LEVLFQ↓GP			
		N-terminal		C-terminal	
length [aa]	947	length [aa]	602	length [aa]	345
MW [Da]	108262	MW [Da]	68920	MW [Da]	39360
pI	6.34	pI	6.34	pI	6.34
$\varepsilon_{0.1\%}$ [(mg/mL)$^{-1}$cm^{-1}]	1.40	$\varepsilon_{0.1\%}$ [(mg/mL)$^{-1}$cm^{-1}]	1.35	$\varepsilon_{0.1\%}$ [(mg/mL)$^{-1}$cm^{-1}]	1.48

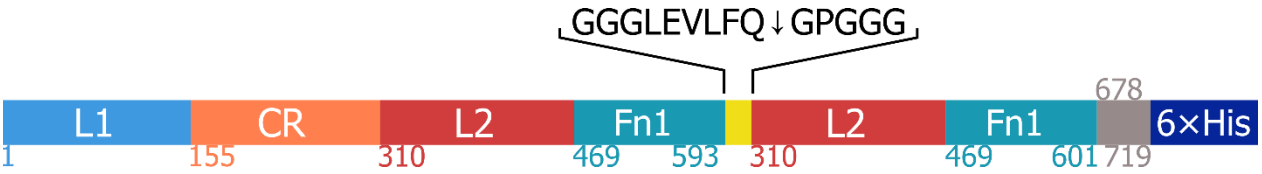

Figure 4.18| mk13 construct map and properties
mk13 is an IR(1-593)-GGGLEVLFQGPGGG-IR(310-601)-IR(678-719)-6×His heterodimeric construct with C682S, C683S, C685S mutations. MW – molecular weight of polypeptide chain (excluding glycans), pI – isoelectric point, $\varepsilon_{0.1\%}$ - extinction coefficient at 280 nm [(mg/mL)$^{-1}$cm^{-1}]. Parameters were calculated using ProtParam [115].

mk13 was expressed in HEK 293T cells according to section 3.6 and purified from 1.8 l of cultured media according to section 3.6. In the IMAC step, the imidazole concentration in the wash buffer was increased to 32 mM (2.5% elution buffer) and the protein was eluted with 500 mM imidazole instead of a linear gradient (Figure 4.19). At this step, it is unclear whether this improved the purity of mk13 in eluates compared to mk11 preparation (section 4.4). Collected fractions were analysed by non-reducing SDS-PAGE followed by Western Blotting with the 83-7 antibody (Figure 4.20) which proved sufficiently sensitive to analyse purification samples. Similarly to blots from native PAGE of expression media, the mk13 protein is present in 3 distinct oligomeric forms, albeit poorly resolved on the 10% gel.

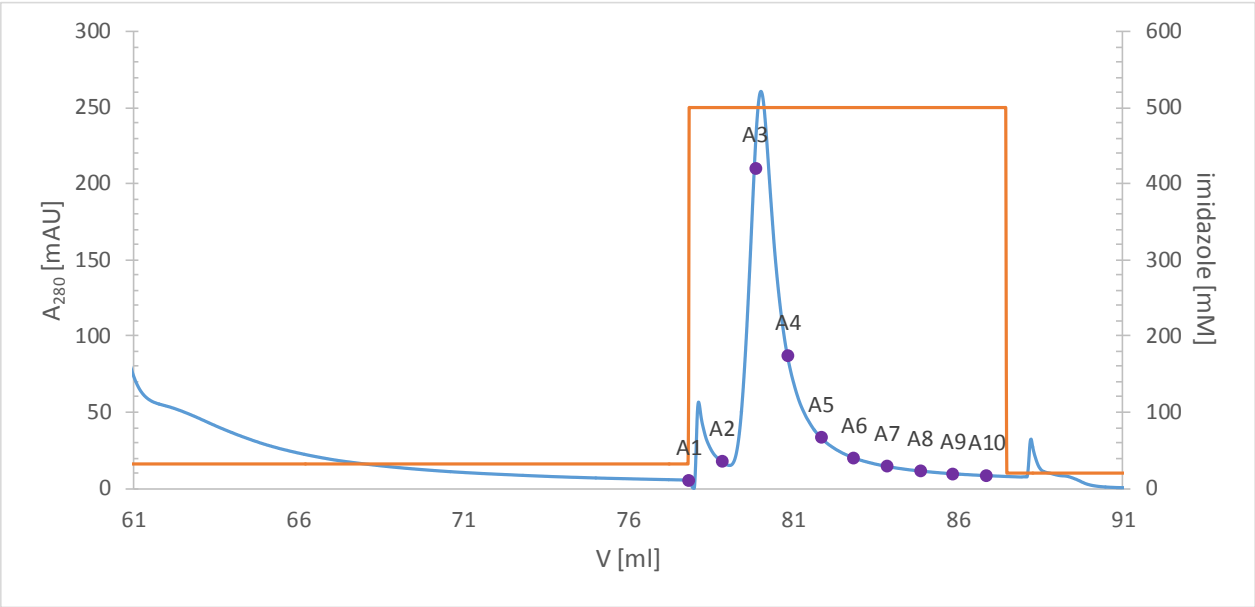

Figure 4.19| Elution of mk13 from a nickel affinity resin
After concentration and buffer exchange of mk13 expression media with a Tangential Flow Filtration system, sample was loaded onto a 1 ml HisTrap FF column, washed with 32 mM imidazole over 20 CV and eluted with 500 mM imidazole. 1 ml fractions were collected (shown by purple points) and subjected to gel electrophoresis and western blotting.

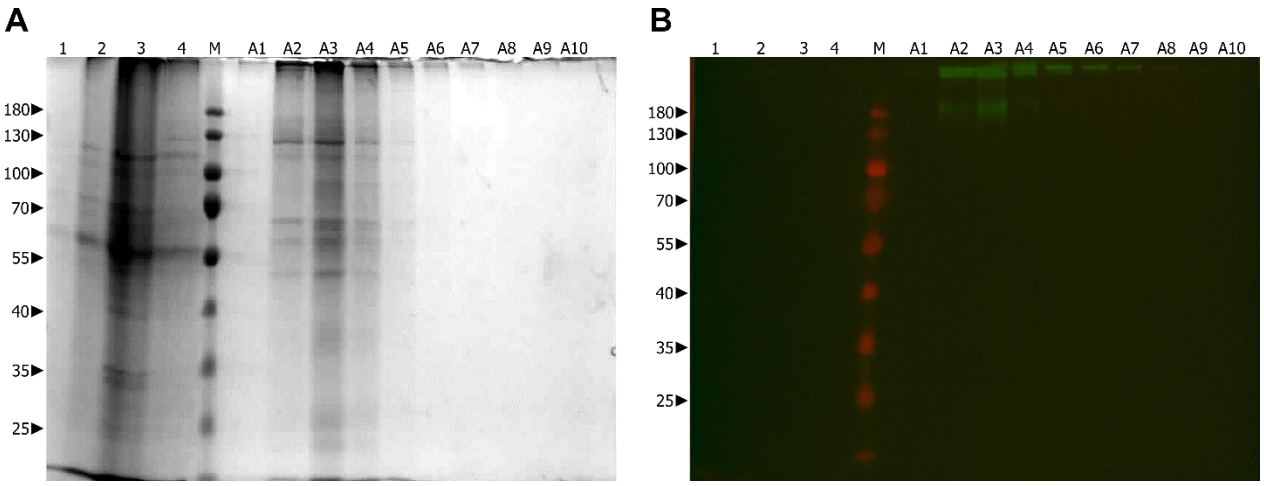

Figure 4.20| SDS-PAGE and Western Blot analysis of mk13 His-tag purification
Samples from mk13 purification steps were subjected to SDS-PAGE and Western Blot analysis. **Panel A|** Purification samples were separated by SDS-PAGE on a 10% acrylamide gel under non-reducing conditions (5% β-Me in sample buffer). **Panel B|** The samples were transferred to a PVDF membrane and probed with monoclonal 83-7 (anti-CR) antibody. 1 – IM1242 (for mobility reference) expression media from small-scale tests, 2 – mk13 expression media, 3 – column flow-through, 4 – column wash, A1-A10 – elution fractions (see Figure 4.19). Green channel – visible light signal, red – chemiluminescence signal.

Fractions A2-A6 were pooled, diluted and loaded onto a 1 ml HiTrapQ HP anion exchange column. Instead of the linear gradient used in mk11 purification, mk13 was eluted with a stepwise 50-200-500 mM NaCl gradient, each step over 5 CV and a final column flush with 1 M NaCl (Figure 4.21).

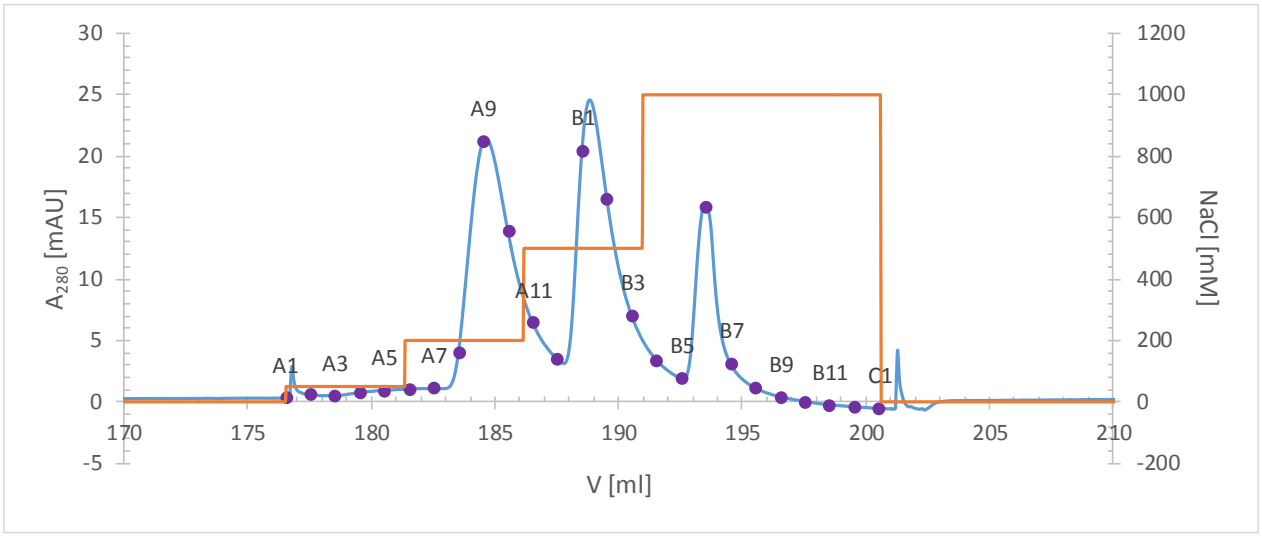

Figure 4.21| Elution of mk13 from an anion exchange resin
Combined fractions from IMAC purification were diluted 30-fold in salt-free buffer and loaded onto a 1 ml HiTrap Q HP column using an ÄKTApure liquid chromatography system, washed with equilibration buffer over 20 CV and eluted with a stepwise gradient of 50-200-500 mM NaCl over 5CV each. 1 ml fractions were collected (shown by purple points with labels above every other fraction) and subjected to gel electrophoresis and western blotting (Figure 4.22).

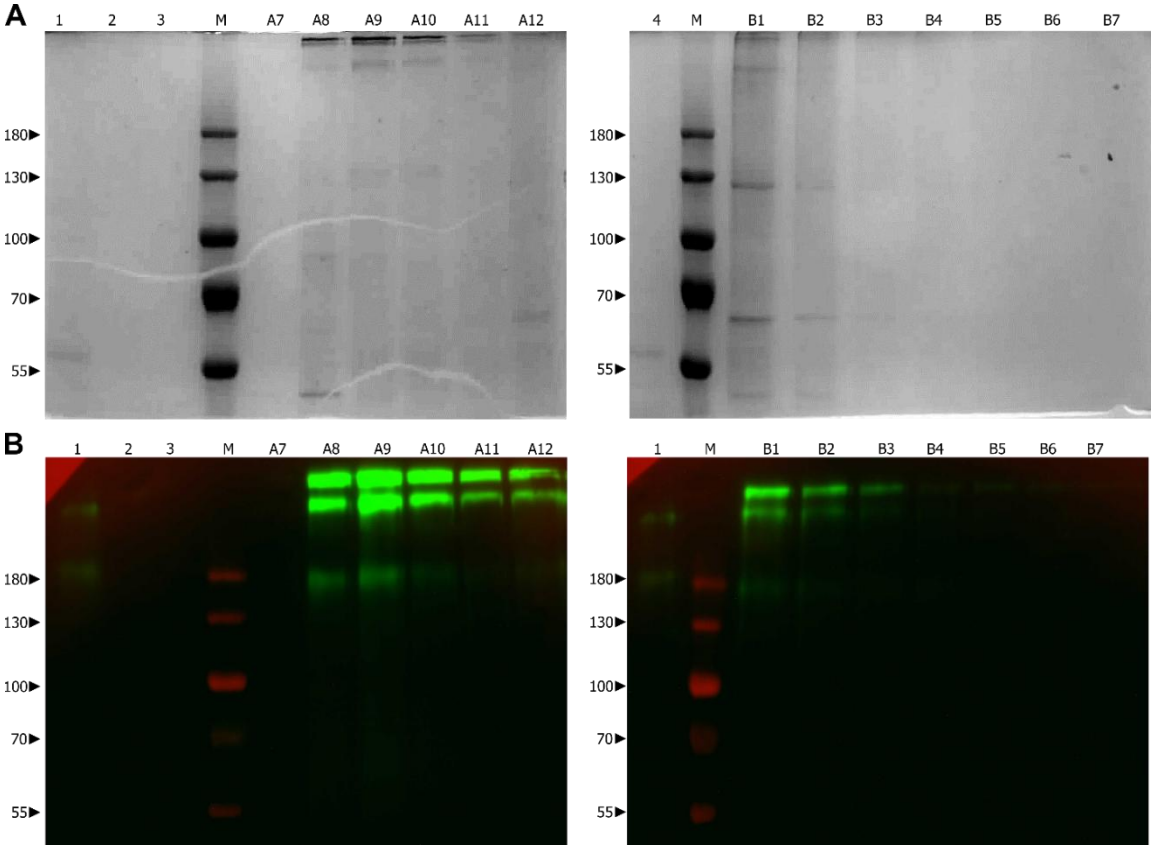

Figure 4.22| SDS-PAGE and Western blot analysis of mk13 purification using anion exchange column
After His-tag purification, mk13 was further purified using a 1ml HiTrap Q anion exchange column and purification samples were analyzed using SDS-PAGE and Western Blot analysis. Purification samples were separated by SDS-PAGE on a 7.5% acrylamide gel under non-reducing conditions (**panel A**), transferred to a PVDF membrane and probed with 83-7 (anti-CR) antibody (**panel B**). 1 – sample of expression medium containing IM1242 (from expression screening), 2 – column flow-through, 3 – column wash, M – PageRuler prestained protein ladder, A7-B7 – elution fractions (see Figure 4.21).

After anion exchange purification, fractions A8-A12 were collected and concentrated to approx. 100 µl. Absorbance at 280 nm was measured to be 1.260, corresponding to 126 µg of total protein assuming $\varepsilon_{280}=1\ (mg/mL)^{-1}cm^{-1}$ for a heterogeneous protein mixture. The purity achieved in these two steps, while not sufficient for structural studies, represents a significant improvement over the mk11 purification (compare lanes 1 and 2 on Figure 4.24, panel B) likely owing to the utilization of more aggressive washing and step elutions.

After this step, the sample was divided and 70 µl was used for cleavage testing while 30 µl was used to determine elution volume of un-cleaved mk13 from a Superdex S200 column.

4.6.1 Cleavage using 3C rhinovirus protease

Cleavage in non-reducing conditions

On non-reducing gels, both native and denaturing, mk13 runs as three distinct bands, likely representing a monomer, a dimer and a higher oligomeric form, and the two high molecular weight forms are only clearly resolved by low percentage gels. mk13 also contains a 3C rhinovirus protease cleavage site between two differently truncated IR sequences. mk13 cleavage was tested in non-reducing conditions (25 mM Tris pH 8.0, 200 mM NaCl buffer) to determine whether a) the site was undergoing cleavage and b) the cleavage affected the oligomerization state (Figure 4.23).

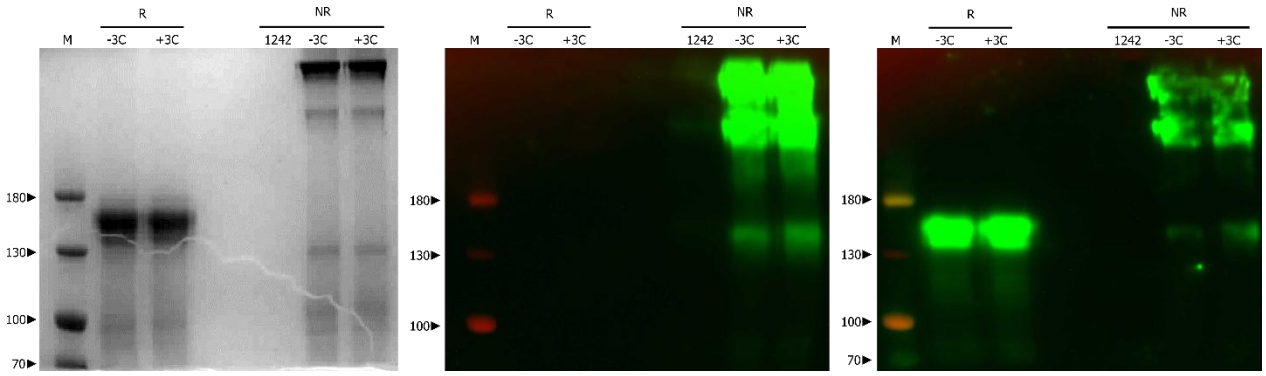

Figure 4.23| Cleavage of mk13 3C protease site under non-reducing conditions
mk13 purified by nickel affinity and anion exchange was visualized by SDS-PAGE under reducing (R) and non-reducing (NR) conditions. Same protein sample incubated for 24h was also applied to the gel under reducing and non-reducing conditions. The samples were run on a 5% acrylamide gel (left panel), transferred to a PVDF membrane and developed using the 83-7 anti-CR monoclonal antibody (middle panel), and then the same blot was washed and developed using the monoclonal anti-His HRP conjugate antibody (right panel). M – PageRuler Prestained Protein Ladder, -3C/+3C – mk13 incubated for 24h in a non-reducing buffer with and without 3C protease, 1242 – expression media containing IM1242 for size reference.

Incubation of mk13 sample with 3C protease does not lead to any significant cleavage under non-reducing conditions. As the 3C protease is a cysteine protease, it is possible that its thiol group is becoming oxidized in the non-reducing environment, inhibiting the cleavage process. Another possibility is that the cleavage site is occluded in the oligomeric states seen in non-reducing buffer

conditions. There is also a clear signal from the anti-His antibody, suggesting that the C-terminus is intact and previous attempts to optimize this antibody for His-tagged constructs (e.g. Figure 4.12) failed due to low concentrations of target protein in cultured media and initial purification steps.

Cleavage in the presence of a reducing agent

Cleavage was further investigated in the same buffer with an addition of a gradient of DTT concentrations (Figure 4.24). Two high molecular weight oligomers present in mk13 sample in non-reducing buffer persist in 0.05 mM DTT, even though they are mostly disrupted and majority of the sample is cleaved in this buffer. 0.1 mM DTT is enough to facilitate near-complete cleavage as well as complete disruption of oligomeric species presumably mediated by disulphide bond between FnIII-1 cysteines (Figure 4.24, lane 5 in both panels).

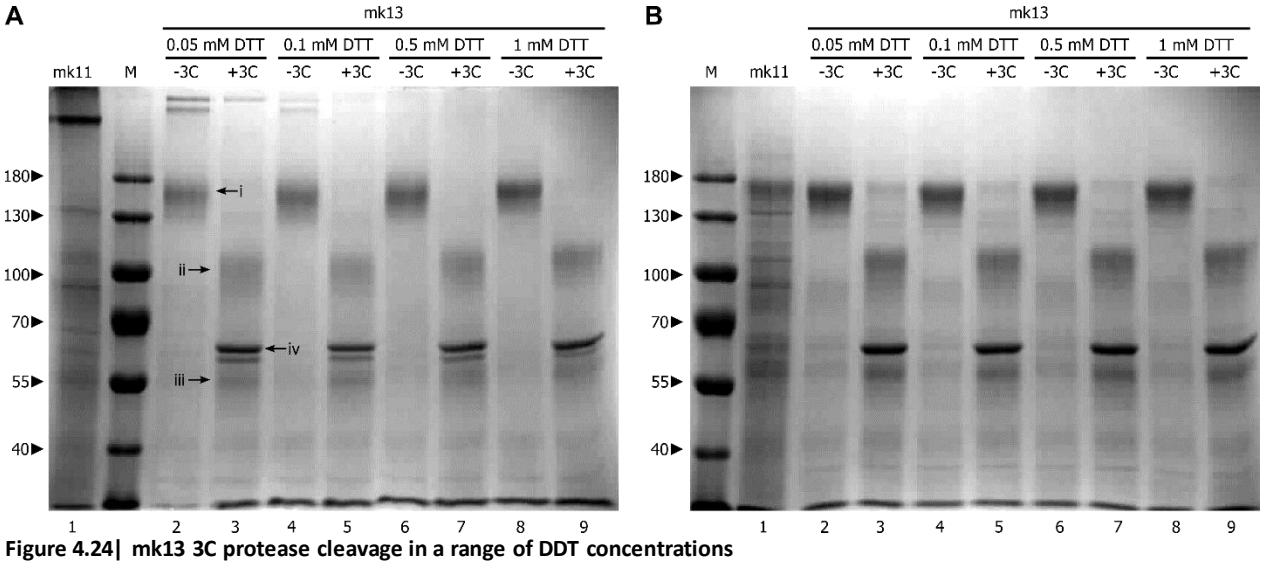

Figure 4.24| mk13 3C protease cleavage in a range of DDT concentrations
IMAC and anion exchange purified mk13 sample was subjected to overnight cleavage with 3C protease. The samples were treated with non-reducing (**panel A**) or reducing (**panel B**) sample buffer and analysed on 7.5% SDS-PAGE to determine oligomeric state in cleavage buffer (non-reducing gel) and cleavage completeness (reducing gel). M – PageRuler prestained protein ladder, lane 1 – sample of mk11 purification (IMAC+IEC, section 4.4), lanes 2-9: samples of mk13 incubated overnight with or without 3C protease in a range of DTT concentrations (0.05, 0.1, 0.5 and 1 mM). Three arrows mark likely mk13 bands: un-cleaved monomer (i), cleaved N-terminal portion (ii) and cleaved C-terminal portion (iii). Band iv represents MBP-3C fusion protein (69 kDa).

Gel filtration in cleavage buffer

The leftover protein sample (30 µl) was applied to a Superdex S200 10/300 GL column in 0.1 mM DTT, 25mM Tris pH 8.0, 200 mM NaCl buffer. In this buffer mk13 eluted in a single peak at 11 ml, similarly to mk11 dimers in non-reducing buffer (Figure 4.25).

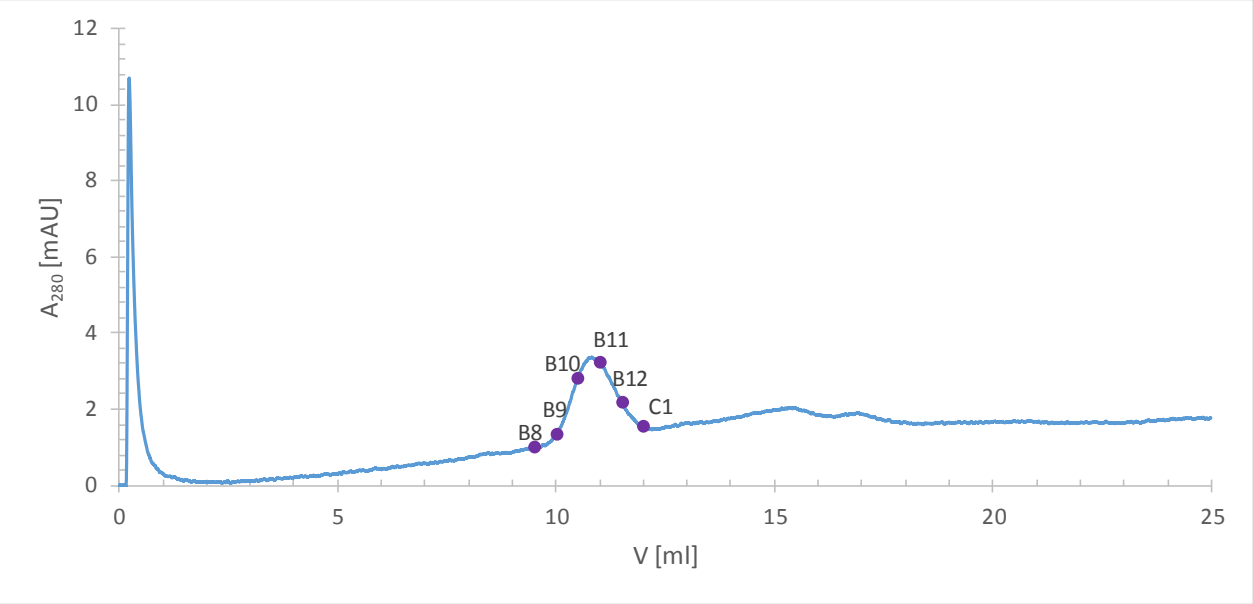

Figure 4.25| Elution profile of mk13 from a Superdex 200 10/300 GL column
After IMAC and anion exchange purifications, 30 µl of mk13 sample was applied to a Superdex 200 10/300 GL gel filtration column. 0.5 ml fractions were collected (peak fractions marked by purple points). Fractions B8-C1 corresponding to the peak were concentrated and analysed on SDS-PAGE (see below).

Fractions B8-C1 corresponding to the peak from gel filtration were pooled and concentrated back to 30 µl in a Vivaspin™ 500 30 kDa cut-off centrifugal filter unit. Protein concentration was measured by absorbance at 280 nm to be 0.14 mg/ml (for a total of 4.3 µg) using mk13 calculated extinction coefficient.

The concentrated sample was divided into three 10 µl fractions: the first one was mixed with 5×SDS-PAGE sample buffer (non-reducing) and incubated overnight at 4 °C. The second fraction was incubated overnight with 3C protease at 1:10 (w/w) ratio at 4 °C and treated with 5× sample buffer (non-reducing). The third fraction was treated with 3C protease the same as second fraction, and subsequently buffer-exchanged by diluting to 500 µl in a 25 mM Tris pH 8.0, 200 mM NaCl buffer, concentrating to 10 µl and repeating the process for a total of five times before mixing with non-reducing sample buffer. All three fractions were applied to a 7.5% SDS-PAGE (Figure 4.26).

The untreated sample appears to be highly pure mk13, but surprisingly, contains all three oligomeric species despite eluting in a single peak from the gel filtration column. As DTT is highly unstable [116], it is possible that the eluted protein was primarily monomeric but during the course of gel filtration and 3C treatment of the rest of the sample (2 days in total), the thiol groups underwent re-oxidation. It is also possible that in 0.1 mM DTT the mk13 monomer is unfolded and elutes with higher oligomers. 3C-treated sample contains a trace of the higher oligomer and similar amounts of monomer and two cleavage products. This could represent incomplete cleavage, or cleavage products forming a bond between their Cys524 residues and both of these events would be consistent with

DTT degradation in buffer. This was not explored due to insufficient amount of sample at that time point. The third fraction (3C-treated and buffer-exchanged) did not appear on the gel presumably due to losses in the buffer exchange process.

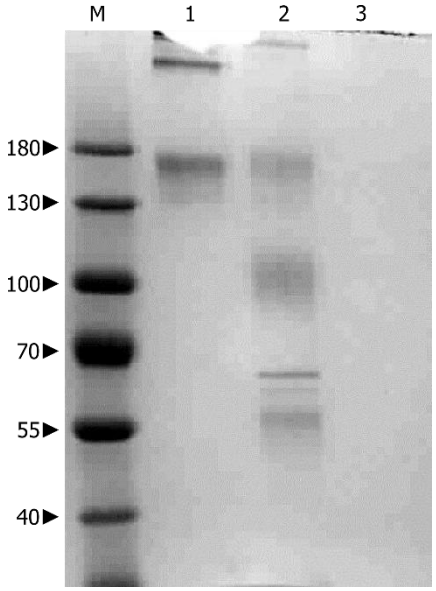

Figure 4.26| mk13 size exclusion in 0.1 mM DTT followed by 3C treatment and buffer exchange
Elution fractions B8-C1 from size exclusion chromatography (Figure 4.25) were pooled, concentrated and divided into three fractions. Fraction 1 was applied to gel as-is (lane 1), fraction 2 was treated overnight with 3C protease (lane 2) and fraction 3 had DTT removed by a series of dilution and concentration procedures (lane 3). M – PageRuler prestained protein ladder.

5 DISCUSSION

The creation of a library of novel heterodimeric insulin receptor constructs forms the core of this work. If successfully crystallized, these constructs could provide insights into the missing structure of site 1 – site 2' crosslinks [71] as well as the ongoing debate about models of signal transduction in the IR and related receptors [98–100]. Guidelines for detection and expression have also been established and levels of proteins produced in transiently transfected cells approach those of established stable cell lines (section 4.3.2). However, obtaining amounts of pure protein sufficient for crystallographic studies has not been accomplished and represents a bottleneck to overcome in future studies of these constructs.

Two constructs, mk11 and mk13, have been produced on preparative scales in HEK293T cells. A purification procedure utilizing immobilized metal affinity chromatography followed by anion exchange chromatography and gel filtration has been optimized to yield pure, albeit not homogeneous, samples of mk13 (see Figure 4.26). As StrepII-tagged constructs have also been made, a change to streptavidin-based affinity chromatography as the initial purification step presents an opportunity to improve the process that has not been explored.

Three major questions were raised in the course of analysing protein samples obtained during this work: what is their pre-cleavage oligomeric state of the single-chain constructs, can they be cleaved, and can a heterodimer consisting of two halves of the cleaved polypeptide be reformed?

Surprisingly, seemingly small modifications in the linker between two fusion partners greatly affect the potential for oligomerisation of the pre-cleavage chain. Compared to IM1242, a construct with a 25 amino acid Gly-Ser linker which exists as a mixture of dimeric and monomeric protein, mk11 exists solely as a dimer and mk13 as a monomer, a dimer, and an unspecified higher oligomeric form. In the course of investigating mk13 cleavage conditions, it was established that a relatively small amount of reducing agent (0.1 mM DTT) can disrupt the oligomeric species. It is unclear if other disulphide bonds necessary for proper folding and activity of these constructs are preserved, but it would not be surprising as (αβ) hemireceptors have been reported to bind insulin in the presence of 1 mM DTT [40].

The construct containing a furin site in the linker sequence, mk11, is not cleaved *in vivo*. A series of constructs with an extended furin site incorporating additional amino acids flanking the RKRR motif in full-length IR have been made. One of those constructs has been expressed in a small-scale screen, and it behaviour is intriguingly different than that of mk11 in Western Blot experiments (see Figure 4.9), but its precise state has not been determined.

A construct with a 3C protease site in the linker sequence, mk13, is cleavable in mildly reducing conditions (at least 0.1 mM DTT) which also disrupts the oligomers formed by this protein. This shows that the approach to producing IR fragments as a single polypeptide chain to simplify purification and stoichiometry (compared to co-transfection or polycistronic vectors) is a viable one. It is currently unclear whether cleavage products would form the desired heterodimer after removal of the reducing agent; this question represents the most immediate point of further inquiry. Alternatively, the 3C site could be replaced with a different recognition site, for example thrombin. As a serine protease thrombin does not require reducing conditions to work, although the recognition site might still be rendered inaccessible by oligomer formation.

As the construct design implies only one binding site comprising of both sub-sites (Figure 2.1), unlike any other with a solved structure so far (see Table 1.1), further attempts to reconstitute and characterize it are a worthwhile pursuit.

REFERENCES

1. Sanger, F. & Thompson, E. O. P. The amino-acid sequence in the glycyl chain of insulin. I. The identification of lower peptides from partial hydrolysates. *The Biochemical Journal* **53,** 353–66 (1953).
2. Sanger, F. & Thomson, E. O. P. The amino-acid sequence in the glycyl chain of insulin. II. The investigation of peptides from enzymic hydrolysates. *The Biochemical Journal* **53,** 366–74 (1953).
3. Sanger, F. & Tuppy, H. The amino-acid sequence in the phenylalanyl chain of insulin. I. The identification of lower peptides from partial hydrolysates. *The Biochemical Journal* **49,** 463–81 (1951).
4. Sanger, F. & Tuppy, H. The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates. *The Biochemical Journal* **49,** 481–90 (1951).
5. Adams, M. J. *et al.* Structure of Rhombohedral 2 Zinc Insulin Crystals. *Nature* **224,** 491–495 (1969).
6. Banting, F. G., Best, C. H., Collip, J. B., Campbell, W. R. & Fletcher, A. A. Pancreatic Extracts in the Treatment of Diabetes Mellitus. *Canadian Medical Association Journal* **12,** 141–6 (1922).
7. Rosenfeld, L. Insulin: discovery and controversy. *Clinical Chemistry* **48,** 2270–88 (2002).
8. Brogiolo, W. *et al.* An evolutionarily conserved function of the Drosophila insulin receptor and insulin-like peptides in growth control. *Current Biology* **11,** 213–221 (2001).
9. Rodríguez-López, C. D., Rodríguez-Romero, A., Aguilar, R. & de Jiménez, E. S. Biochemical characterization of a new maize (Zea mays L.) peptide growth factor. *Protein and Peptide Letters* **18,** 84–91 (2011).
10. Le Roith, D., Shiloach, J., Roth, J. & Lesniak, M. a. Evolutionary origins of vertebrate hormones: substances similar to mammalian insulins are native to unicellular eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **77,** 6184–8 (1980).
11. Jin Chan, S. & Steiner, D. F. Insulin Through the Ages: Phylogeny of a Growth Promoting and Metabolic Regulatory Hormone. *American Zoologist* **40,** 213–222 (2000).
12. Belfiore, A., Frasca, F., Pandini, G., Sciacca, L. & Vigneri, R. Insulin receptor isoforms and insulin receptor/insulin-like growth factor receptor hybrids in physiology and disease. *Endocrine Reviews* **30,** 586–623 (2009).
13. Taniguchi, C. M., Emanuelli, B. & Kahn, C. R. Critical nodes in signalling pathways: insights into

insulin action. *Nature Reviews Molecular Cell Biology* **7,** 85–96 (2006).

14. Bell, G. I. *et al.* Nucleotide sequence of a cDNA clone encoding human preproinsulin. *Nature* **282,** 525–527 (1979).
15. Shields, D. & Blobel, G. Cell-free synthesis of fish preproinsulin, and processing by heterologous mammalian microsomal membranes. *Proceedings of the National Academy of Sciences of the United States of America* **74,** 2059–63 (1977).
16. Steiner, D. F., Cunningham, D., Spigelman, L. & Aten, B. Insulin biosynthesis: evidence for a precursor. *Science* **157,** 697–700 (1967).
17. Steiner, D. F. & Oyer, P. E. The biosynthesis of insulin and a probable precursor of insulin by a human islet cell adenoma. *Proceedings of the National Academy of Sciences of the United States of America* **57,** 473–80 (1967).
18. Smeekens, S. P. & Steiner, D. F. Identification of a human insulinoma cDNA encoding a novel mammalian protein structurally related to the yeast dibasic processing protease Kex2. *The Journal of Biological Chemistry* **265,** 2997–3000 (1990).
19. Smeekens, S. P., Avruch, a S., LaMendola, J., Chan, S. J. & Steiner, D. F. Identification of a cDNA encoding a second putative prohormone convertase related to PC2 in AtT20 cells and islets of Langerhans. *Proceedings of the National Academy of Sciences of the United States of America* **88,** 340–344 (1991).
20. Davidson, H. W. & Hutton, J. C. The insulin-secretory-granule carboxypeptidase H. Purification and demonstration of involvement in proinsulin processing. *Biochemical Journal* **245,** 575–582 (1987).
21. Rigler, R. *et al.* Specific binding of proinsulin C-peptide to human cell membranes. *Proceedings of the National Academy of Sciences* **96,** 13318–13323 (1999).
22. Hills, C. E. & Brunskill, N. J. Cellular and physiological effects of C-peptide. *Clinical Science* **116,** 565–574 (2009).
23. McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. Presenting your structures: the CCP 4 mg molecular-graphics software. *Acta Crystallographica Section D Biological Crystallography* **67,** 386–394 (2011).
24. Dodson, E. J., Dodson, G. G., Lewitova, A. & Sabesan, M. Zinc-free cubic pig insulin: crystallization and structure determination. *Journal of Molecular Biology* **125,** 387–96 (1978).
25. Kaarsholm, N. C., Ko, H. C. & Dunn, M. F. Comparison of solution structural flexibility and zinc binding domains for insulin, proinsulin, and miniproinsulin. *Biochemistry* **28,** 4427–4435 (1989).

26. Weiss, M. A. The structure and function of insulin: decoding the TR transition. *Vitamins and Hormones* **80,** 33–49 (2009).
27. Smith, G. D., Swenson, D. C., Dodson, E. J., Dodson, G. G. & Reynolds, C. D. Structural stability in the 4-zinc human insulin hexamer. *Proceedings of the National Academy of Sciences of the United States of America* **81,** 7093–7 (1984).
28. Derewenda, U. *et al.* Phenol stabilizes more helix in a new symmetrical zinc insulin hexamer. *Nature* **338,** 594–596 (1989).
29. Smith, G. D. & Dodson, G. G. Structure of a rhombohedral R6 insulin/phenol complex. *Proteins* **14,** 401–8 (1992).
30. Hua, Q. X., Shoelson, S. E., Kochoyan, M. & Weiss, M. A. Receptor binding redefined by a structural switch in a mutant human insulin. *Nature* **354,** 238–241 (1991).
31. Ebina, Y. *et al.* The human insulin receptor cDNA: The structural basis for hormone-activated transmembrane signalling. *Cell* **40,** 747–758 (1985).
32. Ullrich, A. *et al.* Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. *Nature* **313,** 756–761 (1985).
33. Seino, S., Seino, M., Nishi, S. & Bell, G. I. Structure of the human insulin receptor gene and characterization of its promoter. *Proceedings of the National Academy of Sciences of the United States of America* **86,** 114–8 (1989).
34. Seino, S. & Bell, G. I. Alternative splicing of human insulin receptor messenger RNA. *Biochemical and Biophysical Research Communications* **159,** 312–6 (1989).
35. Brandt, J., Andersen, A. S. & Kristensen, C. Dimeric fragment of the insulin receptor alpha-subunit binds insulin with full holoreceptor affinity. *The Journal of Biological Chemistry* **276,** 12378–84 (2001).
36. Massague, J., Pilch, P. F. & Czech, M. P. Electrophoretic resolution of three major insulin receptor structures with unique subunit stoichiometries. *Proceedings of the National Academy of Sciences of the United States of America* **77,** 7137–41 (1980).
37. Ronnett, G. V, Knutson, V. P., Kohanski, R. a, Simpson, T. L. & Lane, M. D. Role of glycosylation in the processing of newly translated insulin proreceptor in 3T3-L1 adipocytes. *The Journal of Biological Chemistry* **259,** 4566–75 (1984).
38. Wu, J. J. & Guidotti, G. Proreceptor Dimerization Is Required for Insulin Receptor Post-translational Processing. *Journal of Biological Chemistry* **279,** 25765–25773 (2004).
39. Bass, J., Turck, C., Rouard, M. & Steiner, D. F. Furin-mediated processing in the early secretory pathway: sequential cleavage and degradation of misfolded insulin receptors. *Proceedings of*

the National Academy of Sciences of the United States of America **97,** 11905–9 (2000).

40. Czech, M. P., Massague, J. & Pilch, P. F. The insulin receptor: structural features. *Trends in Biochemical Sciences* **6,** 222–225 (1981).
41. Sparrow, L. G. *et al.* The disulfide bonds in the C-terminal domains of the human insulin receptor ectodomain. *The Journal of Biological Chemistry* **272,** 29460–7 (1997).
42. McKern, N. M. *et al.* Structure of the insulin receptor ectodomain reveals a folded-over conformation. *Nature* **443,** 218–21 (2006).
43. Schäffer, L. & Ljungqvist, L. Identification of a disulfide bridge connecting the alpha-subunits of the extracellular domain of the insulin receptor. *Biochemical and Biophysical Research Communications* **189,** 650–3 (1992).
44. Macaulay, S. L., Polites, M., Hewish, D. R. & Ward, C. W. Cysteine-524 is not the only residue involved in the formation of disulphide-bonded dimers of the insulin receptor. *The Biochemical Journal* **303 (Pt 2,** 575–81 (1994).
45. Bilan, P. J. & Yip, C. C. Unusual Insulin Binding to Cells Expressing an Insulin Receptor Mutated at Cysteine 524. *Biochemical and Biophysical Research Communications* **205,** 1891–1898 (1994).
46. Lu, K. & Guidotti, G. Identification of the cysteine residues involved in the class I disulfide bonds of the human insulin receptor: properties of insulin receptor monomers. *Molecular Biology of the Cell* **7,** 679–91 (1996).
47. Finn, F. M., Ridge, K. D. & Hofmann, K. Labile disulfide bonds in human placental insulin receptor. *Proceedings of the National Academy of Sciences of the United States of America* **87,** 419–23 (1990).
48. Ward, C. W., Hoyne, P. a & Flegg, R. H. Insulin and epidermal growth factor receptors contain the cysteine repeat motif found in the tumor necrosis factor receptor. *Proteins* **22,** 141–53 (1995).
49. Herzberg, V. L., Grigorescu, F., Edge, A. S. B., Spiro, R. G. & Kahn, C. R. Characterization of insulin receptor carbohydrate by comparison of chemical and enzymatic deglycosylation. *Biochemical and Biophysical Research Communications* **129,** 789–796 (1985).
50. Sparrow, L. G. *et al.* The location and characterisation of the O-linked glycans of the human insulin receptor. *Proteins: Structure, Function, and Bioinformatics* **66,** 261–265 (2006).
51. Sparrow, L. G. *et al.* N-linked glycans of the human insulin receptor and their distribution over the crystal structure. *Proteins* **71,** 426–39 (2008).
52. Frasca, F. *et al.* Insulin receptor isoform A, a newly recognized, high-affinity insulin-like growth

factor II receptor in fetal and cancer cells. *Molecular and Cellular Biology* **19,** 3278–3288 (1999).

53. Sciacca, L. *et al.* In IGF-I receptor-deficient leiomyosarcoma cells autocrine IGF-II induces cell invasion and protection from apoptosis via the insulin receptor isoform A. *Oncogene* **21,** 8240–8250 (2002).
54. Pandini, G. *et al.* Differential Gene Expression Induced by Insulin and Insulin-like Growth Factor-II through the Insulin Receptor Isoform A. *Journal of Biological Chemistry* **278,** 42178–42189 (2003).
55. Morrione, A. *et al.* Insulin-like growth factor II stimulates cell proliferation through the insulin receptor. *Proceedings of the National Academy of Sciences of the United States of America* **94,** 3777–3782 (1997).
56. Louvi, A., Accili, D. & Efstratiadis, A. Growth-promoting interaction of IGF-II with the insulin receptor during mouse embryonic development. *Developmental Biology* **189,** 33–48 (1997).
57. Moller, D. E., Yokota, A., Caro, J. F. & Flier, J. S. Tissue-Specific Expression of Two Alternatively Spliced Insulin Receptor mRNAs in Man. *Molecular Endocrinology* **3,** 1263–1269 (1989).
58. Uhles, S., Moede, T., Leibiger, B., Berggren, P.-O. & Leibiger, I. B. Isoform-specific insulin receptor signaling involves different plasma membrane domains. *The Journal of Cell Biology* **163,** 1327–37 (2003).
59. Hubbard, S. R., Wei, L., Ellis, L. & Hendrickson, W. a. Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* **372,** 746–54 (1994).
60. Hubbard, S. R. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. *The EMBO Journal* **16,** 5572–81 (1997).
61. Li, S., Covino, N. D., Stein, E. G., Till, J. H. & Hubbard, S. R. Structural and biochemical evidence for an autoinhibitory role for tyrosine 984 in the juxtamembrane region of the insulin receptor. *The Journal of Biological Chemistry* **278,** 26007–14 (2003).
62. Lou, M. *et al.* The first three domains of the insulin receptor differ structurally from the insulin-like growth factor 1 receptor in the regions governing ligand specificity. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 12429–34 (2006).
63. Smith, B. J. *et al.* Structural resolution of a tandem hormone-binding element in the insulin receptor and its implications for design of peptide agonists. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 6771–6 (2010).
64. Menting, J. G. *et al.* How insulin engages its primary binding site on the insulin receptor. *Nature* **493,** 241–5 (2013).

65. Menting, J. G. *et al.* Protective hinge in insulin opens to enable its receptor engagement. *Proceedings of the National Academy of Sciences of the United States of America* **111,** E3395–404 (2014).

66. Menting, J. G. *et al.* Structural Congruency of Ligand Binding to the Insulin and Insulin/Type 1 Insulin-like Growth Factor Hybrid Receptors. *Structure* 1–12 (2015). doi:10.1016/j.str.2015.04.016

67. Soos, M. a *et al.* Monoclonal antibodies reacting with multiple epitopes on the human insulin receptor. *Biochemical Journal* **235,** 199–208 (1986).

68. Schaffer, L. *et al.* Assembly of high-affinity insulin receptor agonists and antagonists from peptide building blocks. *Proceedings of the National Academy of Sciences* **100,** 4435–4439 (2003).

69. Luo, R., Beniac, D., Fernandes, A., Yip, C. & Ottensmeyer, F. Quaternary Structure of the Insulin-Insulin Receptor Complex. *Science* **285,** 1077–1080 (1999).

70. Yip, C. C. & Ottensmeyer, P. Three-dimensional structural interactions of insulin and its receptor. *The Journal of Biological Chemistry* **278,** 27329–32 (2003).

71. De Meyts, P. Insulin/receptor binding: The last piece of the puzzle? *BioEssays* **37,** 389–397 (2015).

72. Gavin, James R., I. *et al.* Characteristics of the Human Lymphocyte Insulin Receptor. *J Biol Chem* **248,** 2202–2207 (1973).

73. de Meyts, P., Roth, J., Neville, D. M., Gavin, J. R. & Lesniak, M. A. Insulin interactions with its receptors: experimental evidence for negative cooperativity. *Biochemical and Biophysical Research Communications* **55,** 154–61 (1973).

74. Yip, C. C. & Jack, E. Insulin receptors are bivalent as demonstrated by photoaffinity labeling. *J Biol Chem* **267,** 13131–13134 (1992).

75. Pang, D. T. & Shafer, J. a. Evidence that insulin receptor from human placenta has a high affinity for only one molecule of insulin. *The Journal of Biological Chemistry* **259,** 8589–96 (1984).

76. Markussen, J., Halstrom, J., Wiberg, F. C. & Schaffer, L. Immobilized insulin for high capacity affinity chromatography of insulin receptors. *Journal of Biological Chemistry* **266,** 18814–18818 (1991).

77. Sweet, L. J., Morrison, B. D. & Pessin, J. E. Isolation of functional alpha beta heterodimers from the purified human placental alpha 2 beta 2 heterotetrameric insulin receptor complex. A structural basis for insulin binding heterogeneity. *The Journal of Biological Chemistry* **262,** 6939–42 (1987).

78. De Meyts, P. The structural basis of insulin and insulin-like growth factor-I receptor binding and negative co-operativity, and its relevance to mitogenic versus metabolic signalling. *Diabetologia* **37 Suppl 2,** S135–48 (1994).
79. Schäffer, L. A model for insulin binding to the insulin receptor. *European Journal of Biochemistry / FEBS* **221,** 1127–1132 (1994).
80. Kiselyov, V. V, Versteyhe, S., Gauguin, L. & De Meyts, P. Harmonic oscillator model of the insulin and IGF1 receptors' allosteric binding and activation. *Molecular Systems Biology* **5,** 243 (2009).
81. Pullen, R. A. *et al.* Receptor-binding region of insulin. *Nature* **259,** 369–73 (1976).
82. Huang, K. *et al.* How insulin binds: the B-chain alpha-helix contacts the L1 beta-helix of the insulin receptor. *Journal of Molecular Biology* **341,** 529–50 (2004).
83. Williams, P. F., Mynarcik, D. C., Yu, G. Q. & Whittaker, J. Mapping of an NH2-terminal ligand binding site of the insulin receptor by alanine scanning mutagenesis. *The Journal of Biological Chemistry* **270,** 3012–6 (1995).
84. Mynarcik, D. C., Yu, G. Q. & Whittaker, J. Alanine-scanning mutagenesis of a C-terminal ligand binding domain of the insulin receptor alpha subunit. *The Journal of Biological Chemistry* **271,** 2439–42 (1996).
85. Chan, S. J., Nakagawa, S. & Steiner, D. F. Complementation analysis demonstrates that insulin cross-links both alpha subunits in a truncated insulin receptor dimer. *The Journal of Biological Chemistry* **282,** 13754–8 (2007).
86. Kristensen, C. *et al.* Alanine scanning mutagenesis of insulin. *The Journal of Biological Chemistry* **272,** 12978–83 (1997).
87. Hao, C., Whittaker, L. & Whittaker, J. Characterization of a second ligand binding site of the insulin receptor. *Biochemical and Biophysical Research Communications* **347,** 334–9 (2006).
88. Whittaker, L., Hao, C., Fu, W. & Whittaker, J. High-affinity insulin binding: insulin interacts with two receptor ligand binding sites. *Biochemistry* **47,** 12900–9 (2008).
89. Derewenda, U. *et al.* X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule. A completely inactive analogue. *Journal of Molecular Biology* **220,** 425–33 (1991).
90. Jirácek, J. *et al.* Implications for the active form of human insulin based on the structural convergence of highly active hormone analogues. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 1966–70 (2010).
91. Žáková, L. *et al.* Human insulin analogues modified at the B26 site reveal a hormone conformation that is undetected in the receptor complex. *Acta Crystallographica Section D Biological Crystallography* **70,** 2765–2774 (2014).

92. Guo, Z.-Y., Zhang, Z., Jia, X.-Y., Tang, Y.-H. & Feng, Y.-M. Mutational Analysis of the Absolutely Conserved B8Gly: Consequence on Foldability and Activity of Insulin. *Acta Biochimica et Biophysica Sinica* **37,** 673–679 (2005).
93. Nakagawa, S. H. *et al.* Chiral mutagenesis of insulin. Foldability and function are inversely regulated by a stereospecific switch in the B chain. *Biochemistry* **44,** 4984–99 (2005).
94. Kosinová, L. *et al.* Insight into the Structural and Biological Relevance of the T/R Transition of the N-Terminus of the B-Chain in Human Insulin. *Biochemistry* **53,** 3392–3402 (2014).
95. Soos, M. A. & Siddle, K. Immunological relationships between receptors for insulin and insulin-like growth factor I. Evidence for structural heterogeneity of insulin-like growth factor I receptors involving hybrids with insulin receptors. *Biochem J* **263,** 553–563 (1989).
96. Soos, M. A., Field, C. E. & Siddle, K. Purified hybrid insulin/insulin-like growth factor-I receptors bind insulin-like growth factor-I, but not insulin, with high affinity. *The Biochemical Journal* **290 (Pt 2,** 419–26 (1993).
97. Andersen, A. S. *et al.* Changing the insulin receptor to possess insulin-like growth factor I ligand specificity. *Biochemistry* **29,** 7363–7366 (1990).
98. Kavran, J. M. *et al.* How IGF-1 activates its receptor. *eLife* **3,** e03772 (2014).
99. Lee, J., Miyazaki, M., Romeo, G. R. & Shoelson, S. E. Insulin receptor activation with transmembrane domain ligands. *The Journal of Biological Chemistry* **289,** 19769–77 (2014).
100. Ward, C. W., Menting, J. G. & Lawrence, M. C. The insulin receptor changes conformation in unforeseen ways on ligand binding: Sharpening the picture of insulin receptor activation. *BioEssays* **35,** 945–954 (2013).
101. Berrow, N. S. *et al.* A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research* **35,** e45 (2007).
102. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of Escherichia coli with plasmids. *Gene* **96,** 23–28 (1990).
103. Aricescu, A. R., Lu, W. & Jones, E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallographica Section D, Biological Crystallography* **62,** 1243–50 (2006).
104. Andersen, A. S. *et al.* Identification of determinants that confer ligand specificity on the insulin receptor. *The Journal of Biological Chemistry* **267,** 13681–6 (1992).
105. Waugh, D. S. An overview of enzymatic reagents for the removal of affinity tags. *Protein Expression and Purification* **80,** 283–293 (2011).
106. Schmidt, T. G., Koepke, J., Frank, R. & Skerra, A. Molecular interaction between the Strep-tag

affinity peptide and its cognate target, streptavidin. *Journal of Molecular Biology* **255,** 753–766 (1996).

107. Blobel, G. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *The Journal of Cell Biology* **67,** 835–851 (1975).
108. Zhang, L., Leng, Q. & Mixson, a. J. Alteration in the IL-2 signal peptide affects secretion of proteins in vitro and in vivo. *Journal of Gene Medicine* **7,** 354–365 (2005).
109. Kober, L., Zehe, C. & Bode, J. Optimized signal peptides for the development of high expressing CHO cell lines. *Biotechnology and Bioengineering* **110,** 1164–1173 (2013).
110. Knappskog, S. *et al.* The level of synthesis and secretion of Gaussia princeps luciferase in transfected CHO cells is heavily dependent on the choice of signal peptide. *Journal of Biotechnology* **128,** 705–715 (2007).
111. Nakayama, K. Furin: a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. *Biochemical Journal* **327,** 625–635 (1997).
112. Tian, S. A 20 Residues Motif Delineates the Furin Cleavage Site and its Physical Properties May Influence Viral Fusion. *Biochemistry Insights* **2,** 9–20 (2009).
113. Tian, S., Huajun, W. & Wu, J. Computational prediction of furin cleavage sites by a hybrid method and understanding mechanism underlying diseases. *Scientific Reports* **2,** 261 (2012).
114. Nettleship, J. E., Rahman-Huq, N. & Owens, R. J. The production of glycoproteins by transient expression in Mammalian cells. *Methods in Molecular Biology* **498,** 245–63 (2009).
115. Gasteiger, E. *et al.* in *The Proteomics Protocols Handbook* (Walker, J. M.) 571–607 (Humana Press, 2005). doi:10.1385/1-59259-890-0:571
116. Stevens, R., Stevens, L. & Price, N. The stabilities of various thiol compounds used in protein purifications. *Biochemical Education* **11,** 70 (1983).