# Insulin Analogues for Insulin Receptor Studies and Medical Applications

**Christopher John Watson**

**PhD**

**University of York**

**Department of Chemistry**

**September 2012**

# Abstract

The structure of insulin molecule was determined by Dorothy Hodgkin in 1969. Subsequently, it has been established that insulin must rearrange upon binding to its receptor (Insulin Receptor – IR). However, all known structures of the hormone depict its storage or inactive form. It has been shown that some residues, key for IR binding, are buried inside the insulin molecule and must be exposed for an efficient insulin-IR complex formation. It has been postulated that the C-terminal region of the B-chain (~$B^{20}$-$B^{30}$) is dynamic in this process, and that the detachment of the $B^{20}$-$B^{30}$ β-strand leads to the activation of insulin. However, the understanding of the molecular basis of the insulin regulatory role is hindered by the lack of the structure of the insulin-IR complex; only 3-D description of the apo-form of the IR ectodomain is known. The very complex molecular biology behind expression and production of IR fragments also hampers progress in this field.

In order to facilitate progress towards determination of the insulin-IR complex crystal structure this work delivered: (i) structural characterisation of highly-active insulin analogues for stable hormone-IR complexes, (ii) development of various attempts for an alternative production of L1 domain of human IR, (iii) structural characterisation of the role of residues $B^{24}$ and $B^{26}$ for insulin function, (iv) clarification of individual contributions of hydrogen bonds stabilising the insulin dimer, (v) understanding of the structural basis of different functionality of click-chemistry based novel insulin analogues.

This work established that:

(i) the structural signature of the highly active insulin analogues is new -turn at the C-terminus of the B-chain (the B26 turn) achieved by *trans*-to-*cis* isomerisation of the Phe$^{B25}$-Tyr$^{B26}$ peptide bond. This conformational change exposes residues responsible for IR binding, (ii) the production of the L1 domain in *E. coli*, instead of the usual mammalian expression system, is not feasible, (iii) the structural invariance of the Phe$^{B24}$ is fundamental to the formation of the insulin-IR complex. It acts as an anchoring and side-chain pivot for the B26 turn, (iv) removal of the NH$^{B25}$-CO$^{A19}$ dimer interface hydrogen bond is sufficient for a complete disruption of the dimer, whilst the other four hydrogen bonds had a less marked effect in this process, (v) the formation of the B26 turn can be efficiently mimicked by click-chemistry based intra-B-chain crosslinks.

These results provide a wealth of information about the active form of insulin, they provide important tools towards the first insulin-IR complex, and deliver novel insulin analogues.

# List of Contents

# Table of Figures

# Table of Tables

# Acknowledgments

## Author's Declaration

This thesis required input from collaborators to be successful, theses inputs are summarised below:

- The insulin analogues were semi-synthesised in Prague where their binding affinity was also measured.
- The insulin receptor constructs used for crystallisation of insulin:IR complexes at the late stage of this work (hence omitted in the abstract) were designed and expressed in Melbourne.
- The binding affinity of the insulin analogues and the insulin receptor was measured by ITC in Melbourne.
- DILP5 and IBP were expressed and complexed in Denmark.

I declare that all the rest of the work presented in this thesis is my own.

Christopher Watson

September 2012

## Summary of Aims

The main aims for the project are:

- To determine the crystal structures of series of highly-active insulin analogues in order to delineate the active conformation of the hormone
- To develop alternative methods of expression of the L1 domain of human IR
- To identify clear structural and functional roles of $Phe^{B24}$ and $Phe^{B26}$: two key invariant amino acids in vertebrate insulins, important for IR binding
- To clarify the individual contributions of hydrogen bonds that stabilise the insulin dimer
- To explore the application of click-chemistry in making new generations of insulin analogues

As this thesis consists of multidirectional, and in some cases independent lines of research, with structural biology of insulin as the only common feature, an additional summary is also presented at the beginning of every appropriate chapter.

This work also included attempts in the production of Insulin-like Growth Factor-1 (IGF-1) core for future semi-syntheses. Crystallisation and preliminary structural work on complex of *Drosophila* DLP-5 insulin-like peptide with its IMPL-2 insulin binding protein have also been undertaken within this project. However, as these subjects are side- (or late) tracks of the main aspects of this thesis they are described in the appendix.

# Chapter 1. Introduction

## 1.1. Summary of Insulin Molecular Data

Insulin is a 51 amino acid protein hormone, which plays an important role in many key cellular processes. It exerts its regulatory function (e.g. uptake of glucose into cells) through a signalling pathway initiated by binding of hormone in its monomeric form to its dimeric, tyrosine-kinase type membrane receptor (insulin receptor – IR (see chapter 1.4)). Insulin governs the intermediary metabolism of glucose along with the regulation of carbon sources. It is also responsible for stimulating lipogenesis, glycogen and protein synthesis whilst inhibiting their reverse conversions (Figure 1)[1]. Moreover, insulin is also involved in the stimulation of cell growth and differentiation.
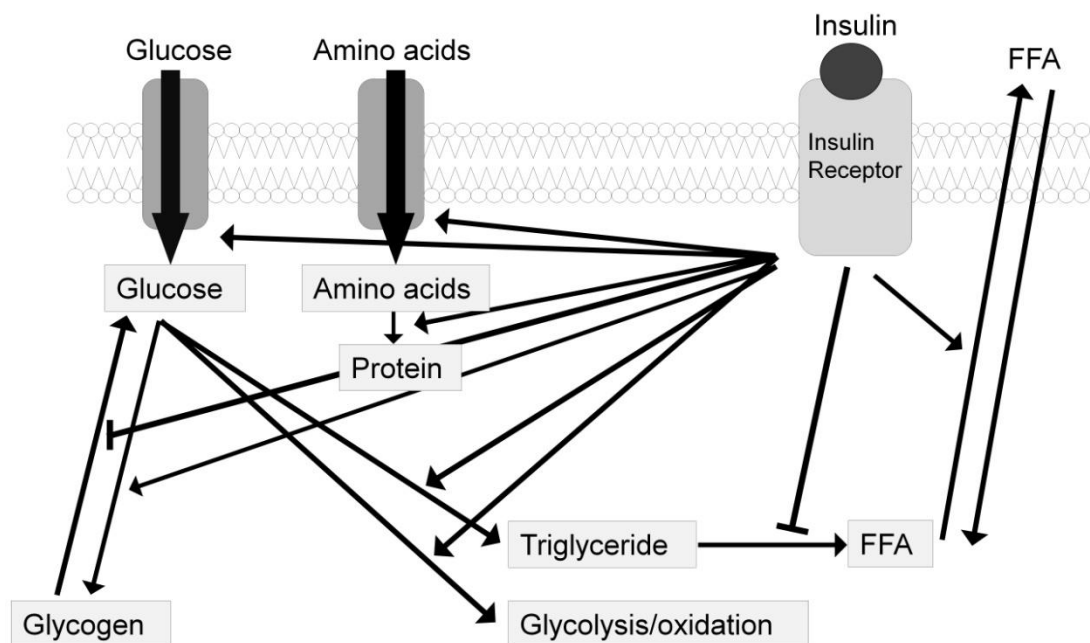


***Figure 1. The main pathways of insulin-dependent regulation of metabolism.***
Insulin is the most potent anabolic hormone, promoting the synthesis and storage of carbohydrates, lipids and protein whilst inhibiting their catabolism. Insulin stimulates the uptake of glucose, amino acids and fatty acids into cells, whilst also increasing the expression or activity of enzymes that are responsible for these processes. Insulin inhibits the activity or expression of those that are responsible for catalysing degradation[1].

The mature form of human insulin consists of 51 amino acids arranged into an A-chain ($Gly^{A1}$-$Asn^{A21}$) and a B-chain ($Phe^{B1}$-$Thr^{B30}$) of total molecular mass of 5808 Da (

Figure 2). The molecule is stabilised by two inter-( $A^6$-$A^{11}$, $A^7$-$B^7$) and one intra-chain disulphide bonds ($A^{20}$-$B^{19}$).

```
                    1           6           11          16          21          26
B-Chain   F V N Q H   L C G S H   L V E A L   Y L V C G   E R G F F   Y T P K T


A-Chain   G I V E Q   C C T S I   C S L Y Q   L E N Y C   N
```

*Figure 2. The amino acid sequence of mature human insulin.*

The first insulin crystal structure was determined in 1969 by Dorothy Hodgkin[2] who studied hexameric crystals of pig insulin, which differs from the human hormone by only the B-chain C-terminal residue ($Thr^{B30} \rightarrow Ala^{B30}$)[2-4]. It has been discovered that in the A-chain residues $A^2$-$A^8$ and $A^{13}$-$A^{19}$ form α-helices, whilst the B-chain is organised by $B^9$-$B^{19}$ α-helix, $B^{19}$-$B^{22}$ β-turn and $B^{23}$-$B^{30}$ that form a β-sheet (Figure 3). Subsequently, it has been discovered that the B-chain N-terminus ($B^1$-$B^8$) can exist in two different conformations: R and T[5]. The so-called R-state of the hormone corresponds to more 'relaxed' and allosterically more active form of this protein, in which an additional α-helical region is created by the $B^1$-$B^{19}$ residues (i.e. $B^9$-$B^{19}$ α-helix is extended). The 'tense' T-state state is less active, with residues $B^1$-$B^{19}$ residues not forming a regular secondary structure and folding away from the core of the hormone. The overall arrangement of the A and B chains results in the A-chain N-terminus and B-chain C-terminus being brought into close proximity. This forms a hydrophobic core buried within the molecule, consisting of $Ile^{A2}$, $Val^{A3}$, $Cys^{A11}$, $Leu^{A16}$, $Leu^{B11}$ and $Leu^{B15}$ [3,6].
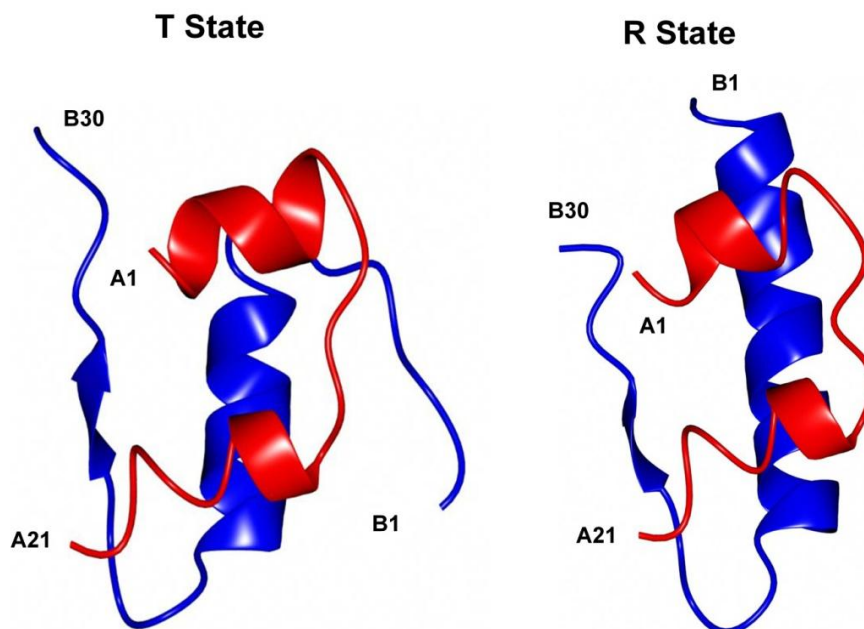


**T State**          **R State**

*Figure 3. The T- and R-states of insulin monomer.*

The A-chain is shown in red, whilst the B-chain is in blue. (pdb used: 1mso [T-state] and 1znj [R-state]); the monomers showed here are derived from hexamer oligomers present in these crystals.

Insulin can exist in three oligomeric forms: hexameric, dimeric and monomeric. Hexameric insulin is the main storage form of the hormone with $Zn^{2+}$ ion located at its centre of the hexamer, and coordinated by $His^{B10}$. Upon the dilution of hexamer containing crystals into the local environment (e.g. during excretion of insulin into the bloodstream) the hexamer disassociates into three dimers, which subsequently split into monomers (Figure 4). The insulin hexamer (Figure 5) contains two or four (even number) $Zn^{2+}$ ions coordinated at its centre [3]. In the 2-$Zn^{2+}$ structure, all six monomers are in the T-state (denoted as $(T_6)$)[2]. In the 4-$Zn^{2+}$ hexamer (stabilised by a high $Cl^-$ concentration), three of the monomers are in the T-state and three are in the R-state $(R_3T_3)$[7]. A similar state to $R_3T_3$ has been found [8] termed $T_3R_3^f$, where in the $R^f$ conformer the $B^4$-$B^{19}$ residues are R-like, α-helical, but the $B^1$-$B^3$ are in an extended form. A fully $R_6$ form exists in phenol-(or phenol-like compounds) containing crystals [5] There are currently over 100 structures of insulin in the Protein Data Bank, though there are none of monomeric wild type (WT) insulin, nor a complex between monomeric insulin and IR.
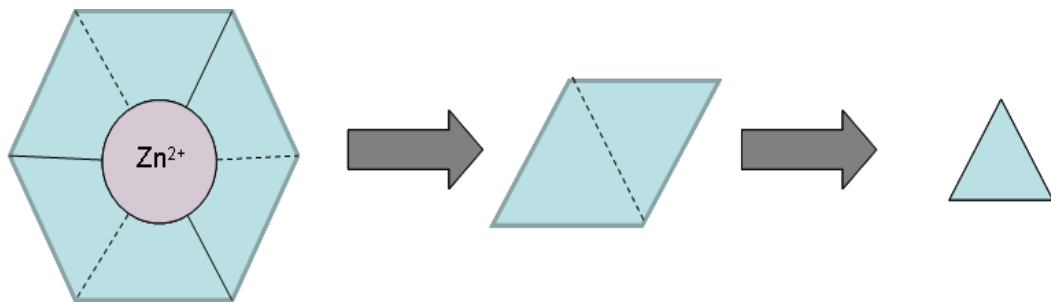


***Figure 4. The hexamer-to-monomer disassociation of insulin.***

The hexameric form, upon the loss of the $Zn^{2+}$ ions at its centre breaks down into a monomer via a dimeric state.
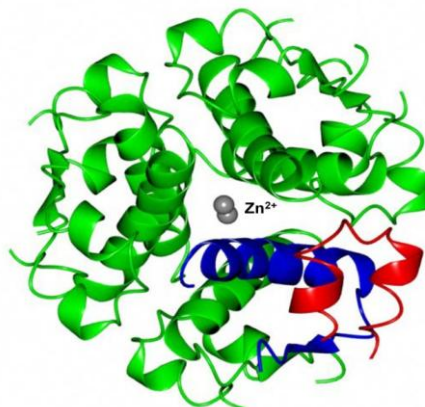


***Figure 5. The general organisation of the insulin hexamer.***

The insulin hexamer comprises three distinct dimers. One of the monomers has its chains coloured as per Figure 3, with the rest shown in green. The $Zn^{2+}$ ions are located at the centre of the hexamer. (pdb used: 1ZNJ).

The insulin hexamer is formed by three dimers that are arranged in a 32 point symmetry (Figure 5). The dimer interface is stabilised by strong interactions, mostly hydrogen bonds, between the two anti-parallel β-strands at the C-terminus of the B-chain provided by each of the monomers (~$B^{21}$-$B^{27}$) (Figure 6).
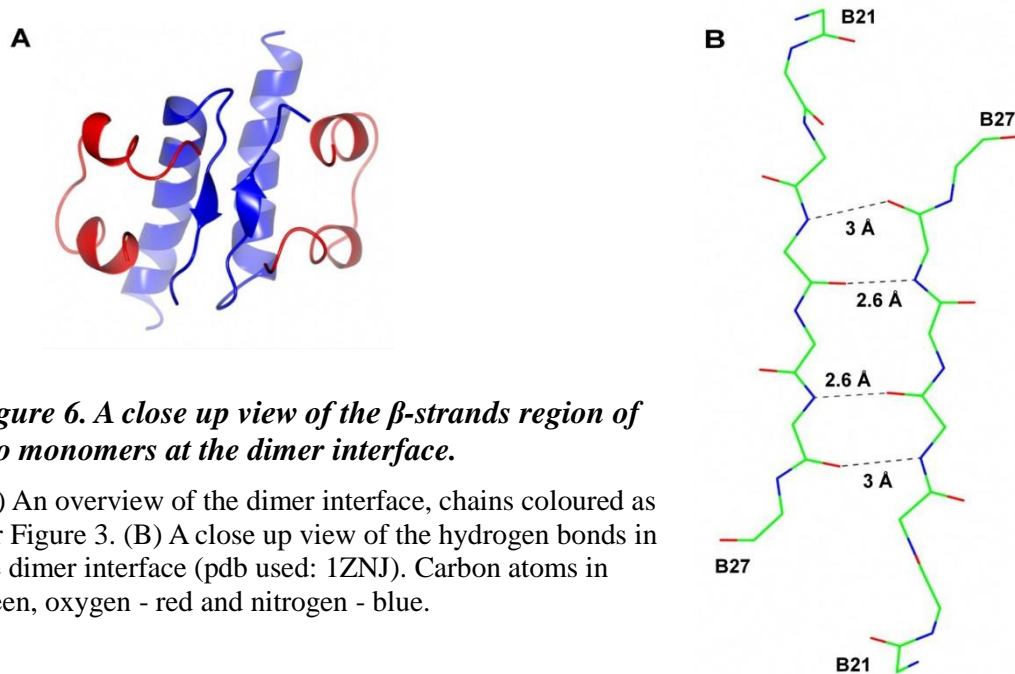


***Figure 6. A close up view of the β-strands region of two monomers at the dimer interface.***

(A) An overview of the dimer interface, chains coloured as per Figure 3. (B) A close up view of the hydrogen bonds in the dimer interface (pdb used: 1ZNJ). Carbon atoms in green, oxygen - red and nitrogen - blue.

The Insulin-like Growth Factors 1 and 2 (IGF-I and IGF-II) are insulin-related small single-chain hormones with molecular masses of 7.6 kDa and 7.5 kDa respectively[9,10](Figure 7). IGFs are major growth factors in the embryonic-puberty phase of development[11]. The IGFs are also involved in the induction of differentiation of whole cells, and processes within cells[12]. The actions of IGF are tightly controlled by growth factors[13].

| B-Chain | 1 | 6 | 11 | 16 | 21 | 26 |
|---|---|---|---|---|---|---|
| Insulin | F V N Q H | L C G S H | L V E A L | Y L V C G | E R G F F | Y T P K T |
| IGF-I | G P E T | L C G A E | L V D A L | Q F V C G | D R G F Y | F N K P T |
| IGF-II | A Y R P S E T | L C G G E | L V D T L | Q F V C G | D R G F Y | F S R P A |

| A-Chain | 1 | 6 | 11 | 16 | 21 | 26 |
|---|---|---|---|---|---|---|
| Insulin | G I V E Q | C C T S I | C S L Y Q | L E N Y C | N | |
| IGF-I | G I V D E | C C F R S | C D L R R | L E M Y C | A P L K P | A K S A |
| IGF-II | G I V E E | C C F R S | C D L A L | L E T Y C | A   T P | A K S A |

***Figure 7. Amino acid sequence of the mature human insulin, IGF-I and IGF-II hormones.***

## 1.2. Discovery and Genetic Engineering of Insulin

In 1869, Paul Langerhans identified some previously undiscovered tissue clumps scattered throughout the pancreas [14]. Whilst the functions of these clumps were initially unknown, they later became known as the Islets of Langerhans. The name insulin has its origins in Insel, the German word for islet/island. Edouard Laguesse subsequently suggested that these cells might be involved in the production of secretions that could have a role in digestion[15].

Oscar Minkowski, in collaboration with Joseph von Mering, established the relationship between the pancreas and diabetes [14,16]. Soon after in 1901, Eugene Opie showed the link between the Islets of Langerhans and diabetes, stating "Diabetes Mellitus … is caused by destruction of the Islets of Langerhans and occurs only when these bodies are in part or wholly destroyed" [17].

During the next two decades various attempts were made to isolate the secretion of the Islets for application in the treatment of diabetes. Zuelzer was first to apply a pancreatic extract, and was partially successful in the treatment of dogs with diabetes in 1906[14]. Insulin (called "pancrein") was isolated for the first time by Paulesco in 1921[18]. The final breakthrough in early insulin research and clinical applications was provided by Frederick Banting, J. J. R. Macleod, Charles Best, James Collip and Clark Noble in 1922 when the first injection of the purified insulin on January 11th 1922 was given to a 14 year old diabetic patient[16]. They scaled up their biochemistry with the involvement of Eli Lilly & Co, and in November 1922 the first large amounts of highly refined, pure insulin were available for clinical and widespread treatments of diabetes[15].

Further understanding of the insulin molecule was provided by Frederick Sanger who determined the primary structure of insulin in 1951 [19]. This was the first determination of any protein primary structure, an achievement for which Frederick Sanger was awarded the 1958 Nobel Prize in Chemistry. This work lead to the first full synthesis of insulin by Panayotis Katsoyannis[20] and Helmut Zahn[21] in the 1960s. Finally, Dorothy Hodgkin made the pioneering structural breakthrough in 1969, when after decades of work she determined the crystal structure of insulin utilising X-ray diffraction methods[4], for which she was awarded the Nobel prize in Chemistry in 1964.

Sanger's discovery, along with advancements in molecular biology, allowed Herbert Boyer to insert the human insulin gene into genetically engineered *E.coli,* and produce this hormone in large and pure amounts in 1978[22]. Genetech (founded by Boyer), along with Eli Lilly and Co, delivered the first commercial biosynthetic human insulin called Humulin in 1982[22]. Currently, there are there are 1.4 million people in the UK alone who rely upon daily injections of recombinant insulin to control their diabetes.

## 1.3. Insulin Biosynthesis and Bioavailability

There are three types of stimulus that are responsible for the release of insulin. The first is the detection of an increase in blood sugar levels. Insulin secretion can also be stimulated in response to an increased blood concentration of alternative carbon sources such as amino acids and fatty acids, and neural stimuli such as the sight and taste of food.

A prolonged increase in blood sugar levels is detected by both the hypothalamus and pancreatic beta-cells[23], which triggers the translation of insulin mRNA in the Islets of Langerhans, the beta-cells in the pancreas. The insulin mRNA is translated as a single chain molecule called preproinsulin, which consists of 110 amino acids. The preproinsulin is transported to the rough endoplasmic reticulum, where its signal peptide is removed during insertion into the cisternal space, forming proinsulin (Figure 8).
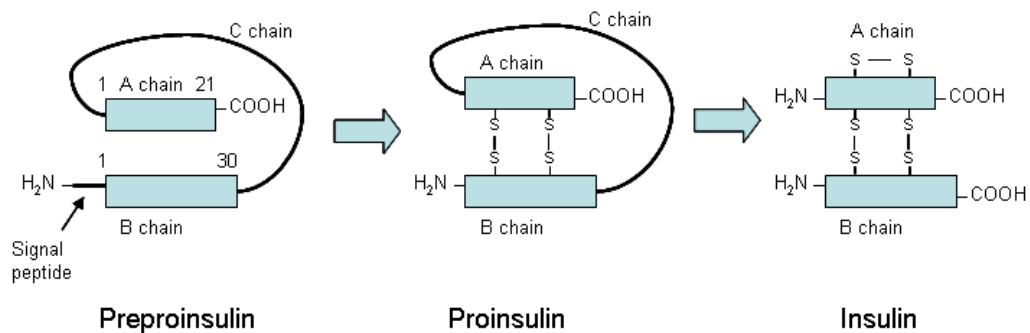


*Figure 8. The process of insulin biosynthesis and maturation.*

The processing of preproinsulin is an important step in the formation of the mature hormone. The signal peptide removed from the preproinsulin is packaged where it is accumulated prior to being recycled in the cytoplasm. Proinsulin consists of three domains; an amino-terminal B-chain, a carboxy-teminal A chain, and B-A connecting peptide known as the C-peptide. The endoplasmic reticulum contains several specific endopeptidases, which excise the 31 amino acid C-peptide chain from the proinsulin to create the mature insulin molecule. Insulin is prevented from possibly becoming misfolded, by formation of a hexameric assembly with $Zn^{2+}$ ions. The insulin hexamer is stabilised further by microcrystallisation[24]. The C-peptide and mature insulin are packaged into secretory granules, which accumulate in the cytoplasm. Once the beta-cells are stimulated (Figure 9), the insulin is secreted from the cell by exocytosis of the secretory granules; the dilution-caused leakage of $Zn^{2+}$ ions from hormone hexamers/dimers leaves an insulin zinc-free monomer[25] in the bloodstream[26]. The C-peptide is also secreted (Figure 9), but its biological activity has not yet been identified.

**The Beta Cell**

20-40 min

Constitutive Pathway

Energy Dependent
Transfers

Rough
Endoplasmic
Reticulum

Lysosome

Insulin
mRNA

Golgi

Insulin
Gene

Nucleus

Insulin +
C-peptide

Regulated Secretion

Mature
Granules

Pores

Signal
Peptidase

Proinsulin

Preproinsulin
Synthesis
(S-S Bonds)

Clathrin Coat

Plasma Membrane

Membrane Recycling

40 min      10-15 min      30-40 min      2-4 hours      1-2 days

*Figure 9. Insulin biosynthesis within beta-cells of the pancreas.*

The major steps involved in insulin biosynthesis, from the expression of preproinsulin to the secretion of mature insulin. Preproinsulin is converted to proinsulin in the rough endoplasmic reticulum, which is then followed by the folding of the molecule and formation of the correct pattern of disulphide bonds. Proinsulin is then transported to the Golgi, from which it is secreted in granules, in which the conversion of the pro-hormone to the mature insulin takes place. Fusion of the granules with the plasma membrane leads to the eventual secretion of mature insulin from the beta-cell. A small proportion of unprocessed proinsulin is also secreted via an unregulated pathway[27].

The pathway by which insulin is secreted is known, but not fully understood. Glucose enters the beta-cells by facilitated diffusion utilising a glucose transporter, GLUT2. The glucose molecule is then phosphorylated by the rate-limiting enzyme glucokinase. Simultaneously, insulin inhibits glucose-6-phosphatase to prevent the phosphates from being removed. This causes the glucose molecule to become trapped within the beta-cells, where its metabolism increases the ATP:ADP ratio, causing the closure of the ATP-gated cell membrane potassium channels. This prevents the escape of potassium ions from the cell, and its subsequent depolarisation, which then activates the voltage gated calcium channels. Subsequently, a rapid increase in intracellular calcium triggers the exocytosis of insulin-containing granules[28].

Release of insulin stimulates the uptake of glucose into cells by the sequestration of intracellular vesicles containing GLUT4 from intracellular stores to the cell membrane[29]. The exact mechanism by which a GLUT4-vesicle reaches the membrane is unknown, but it is believed that this vesicle is attached to an intracellular site that is linked to the membrane by a microtubule network[30]. The GLUT4-vesicles are untethered from their intracellular site and are transported along the microtubules towards the membrane in response to insulin:IR

binding and the resultant IR-initiated phosphorylation cascade. The docking and fusion of the vesicles with the membrane makes the GLUT4 available for glucose uptake.  It has been postulated that insulin may also regulate the docking and fusion of the GLUT4 vesicle with the cell membrane by a direct regulation of the v/t-SNARE proteins that are required for an efficient fusion of the vesicle and the cell membrane[29]. The time-dependent decrease in the insulin blood level results in a smaller number of stimulated IRs, which leads to the reverse endocytosis of GLUT4 transporters from the membrane back into vesicles, and their return to the cytoplasm[31].

Stimulation of insulin release is readily observed in animals and humans. The normal fasting blood glucose concentration in humans is tightly controlled at 80-90 mg/100 ml (~5 mM)[1] (Figure 10), and is associated with very low levels of insulin secretion (Figure 11) and its short (4-6 min) half-life in the blood [32].  The serum levels of insulin increase dramatically almost immediately after the meal. This initial increase is due to secretion of preformed/stored insulin, which is quickly depleted. A secondary rise in insulin is assured by newly synthesised insulin that is immediately released [27].
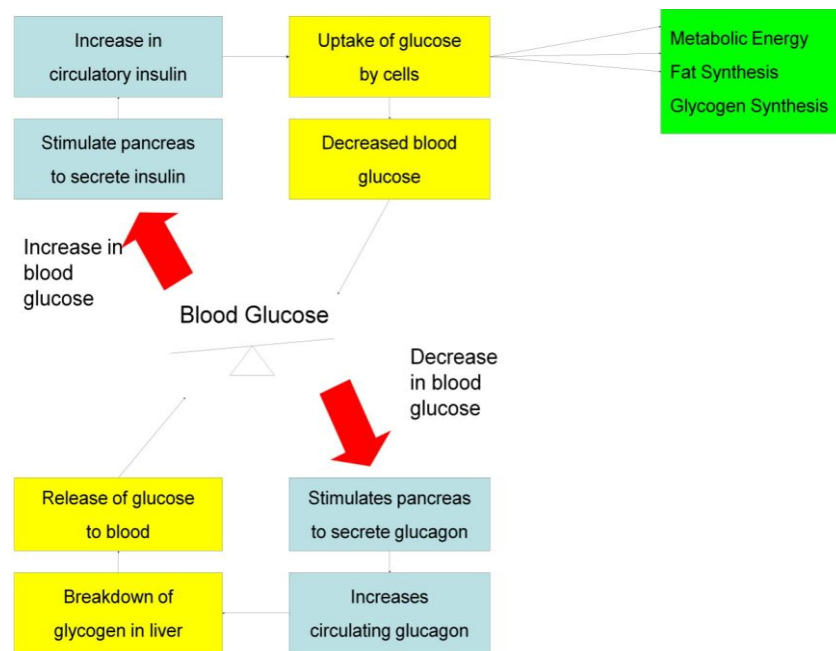


*Figure 10. The general network of regulation of blood glucose levels.*

The increase of the blood sugar level leads to secretion of insulin by the pancreas. The body-circulation of this hormone, and its binding to the IR results in the uptake of glucose into cells, causing the blood sugar level to decrease.  Consequently, glucagon  - a hormone that works in tandem with insulin reversing its actions - is excreted by the pancreas to breakdown glycogen into glucose in the liver.  This glucose is then released into the bloodstream to raise the blood sugar level.
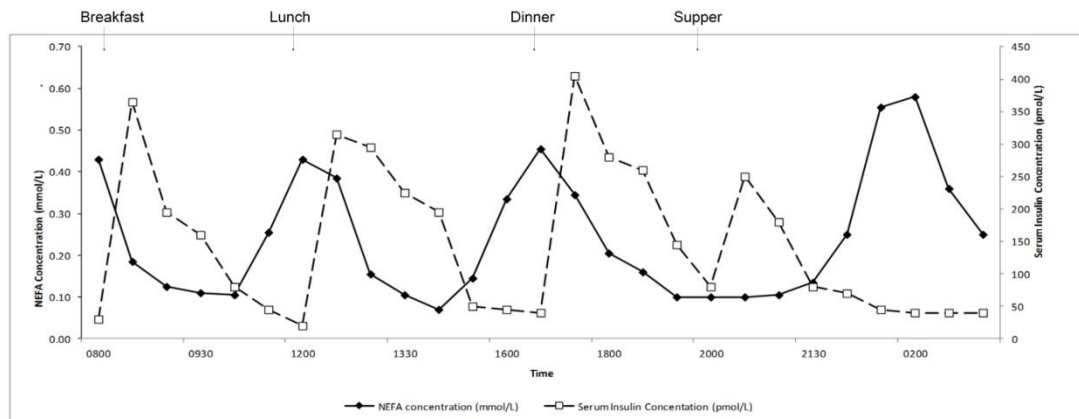
*Figure 11. Daily profile of NEFA (non-esterified fatty acid) concentrations and serum insulin concentrations.*

The NEFA levels are an indicator of blood glucose levels. Adapted from Daly[33].

After initiation of the intracellular signalling the insulin:IR complex is endocytosed[34]. The complex is transported along the endocytic pathway, where is either recycled back to the cell surface, or targeted to the lysosome for degradation via the late endosomes [35]. Internalisation of the IR depends on the IR autophosphorylation state, and is followed by downstream effects through the clathrin-mediated pathway[31]. There is also evidence for other, alternative pathways of IR internalisation [34,36].

The overall maintenance of the plasma glucose level is achieved by balancing the glucose intake into the plasma from the intestine and its uptake to peripheral tissue. A drop of the blood glucose level below ~2.5 mM threshold level results in the stoppage of glycogen synthesis in the liver, and simultaneous activation of its breakdown. Insulin has a versatile role in the maintenance of blood glucose levels[1] as it is also responsible for the inhibition of hepatic glucose production. Moreover, the metabolic effects resulting from the insulin actions depend on the type of the tissue, e.g. some tissues take up glucose independently of insulin. For example, the brain relies upon glucose as its sole energy source[37] and utilises a non insulin-dependent glucose transporter. The skeletal muscle is the primary tissue responsible for insulin-regulated glucose uptake, while only a small fraction of glucose is taken up by adipose tissue[38].

## 1.4. Molecular Organisation of the Insulin Receptor

The first indication that insulin does not play an 'enzymatic' role in carbohydrate metabolism, but rather it interacts with the cell membrane was postulated by Levine in 1949, who proposed that insulin facilitates the uptake of hexoses into cells[39]. Subsequently House and Weidemann showed binding of insulin to liver cells using radio-labelled insulin[40]. Only in 1980-1981 was it demonstrated that the IR is a dimer with an apparent molecular weight of ~350 kDa and comprised of two α-subunits (~120-130 kDa) and two β-subunits (~90 kDa) interconnected by disulphide bonds[41,42].

Soon, it was discovered that the intracellular parts of the β-subunits contain tyrosine kinase (TK) that is activated upon insulin binding[43,44]. This led to the subsequent identification of an intracellular substrate for this kinase - IR substrate 1, IRS1[45,46].

Two laboratories determined the cDNA of the human IR independently in 1985[47,48] revealing two different isoforms (IR-A, IR-B) of the receptor. They arise from alternative splicing, which results in different (12 amino acids) lengths of the IR at the C-terminus of the α-subunit. In IR-A exon 11, consisting of 36 base pairs, is excised whilst being preserved in the IR-B. Further studies led to the discovery of 18 potential N-linked glycosylation sites[49], with a further six O-linked glycosylation sites[50]. The structure of the unphosphorylated human TK was determined in 1994 by Hubbard *et al.[51]*, elucidating its activated, phosphorylated structure. This revealed the auto-phosphorylation of three tyrosines ($Tyr^{1158}$, $Tyr^{1162}$, $Tyr^{1163}$) in the activation loop (A-loop) leads to a dramatic change of the IR conformation[52]. The crystal structure of the IR ectodomain dimer was solved in 2006, a major achievement in insulin research[53]. The structure solution was facilitated by co-crystallisation with monoclonal antibodies (Fabs)[54,55].

IRs share very similar domain organisation to the Epidermal Growth Factor receptors (EGFR). In both families of these receptors an aberrant signalling has been implicated in different types of cancers. Study of the EGFRs can lead to important information in the study of the IR[56].

The IR (Figure 12) is a member of a sub-family of receptors that contain a TK, whose other members include the Insulin-Like Growth Factor (IGF)-I receptor (IGF-1R) and insulin receptor-related receptor (IRR)[57]. These receptors are dimeric receptors which contain two α- and two β-subunits that function as allosteric enzymes, where the apo-form of an α-subunit inhibits the TK activity of the β-subunit [1]. Binding of insulin to the IR abolishes the repression of TK activity, resulting in cross-phosphorylation of the β-subunit that induces further conformational changes and an increase of its activity [58]. The similar structure of the substrates of the IR, IGF-I and IRR receptors, results in their functional

hybrids, hence a mutation in one type of the receptor can have a functional consequences for the other types [57]. The hormone binding sites of the receptor are located within the fully-extracellular α-subunits.

The IR ectodomain consists of the complete α-chain and 194 residues ($Ser^{724}$-$Lys^{917}$) of the β-chain. Each IR monomer contains several structural domains (Figure 12) including a leucine-rich repeat domain (L1 residues 1–157), a cysteine-rich region (CR, residues 158–310), a second leucine-rich repeat domain (L2, residues 311–470), and three fibronectin type-III domains (FnIII-1 residues 471–595, FnIII-2 residues 596–808, and FnIII-3 residues 809–906). The FnIII-2 region includes a 118-residue insert domain (ID, residues 638–756) that contains the αβ cleavage site at residues 720–723, which creates the α- and β-chains. The segments of the ID that lie within the α- and β-chains are identified as IDα and IDβ, respectively. The C-terminus of the α-subunits consists of residues 704-719 (CT peptide). The C-terminus of the FnIII-3 domains consists of a single transmembrane helix, followed by a ~40 residue intercellular juxtamembrane region (JM), a tyrosine kinase (TK) catalytic domain, and a ~100-residue C-terminal tail (Figure 13) [59-61]. The IR receptor assembles into an inverted "V" shape[53,61], with the L1-CR-L2 domains forming one side of the V, whilst the three FnIII domains form the other. When the receptor dimerises, a two-fold rotation about the inverted V axis is formed by the L1-CR-L2 side packing against the three FnIII domains in the other monomer[59]. The IGF-1R has a very similar domain organisation.
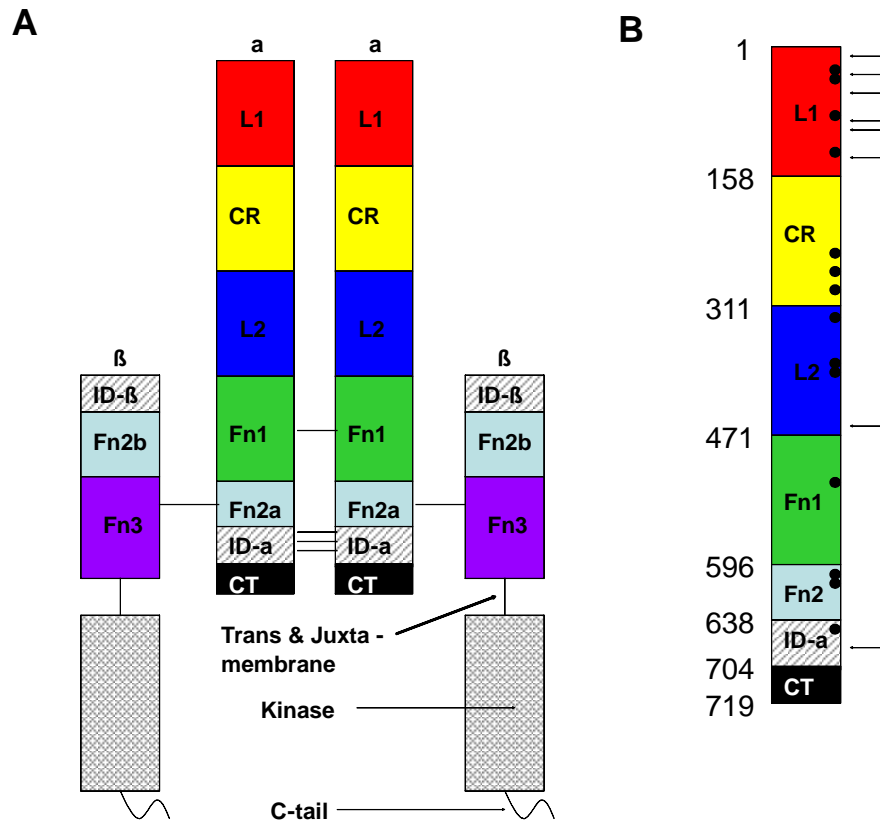
*Figure 12. (A) An overview of the organisation of the IR. (B) A close up of the α-subunit of the IR.*

In (B) The arrows indicate insulin binding locations, as determined by single amino acid-directed mutagenesis[60]. The majority of these binding "hotspots" are located in the L1 region. The black spots are sites of *N*-glycosylation.
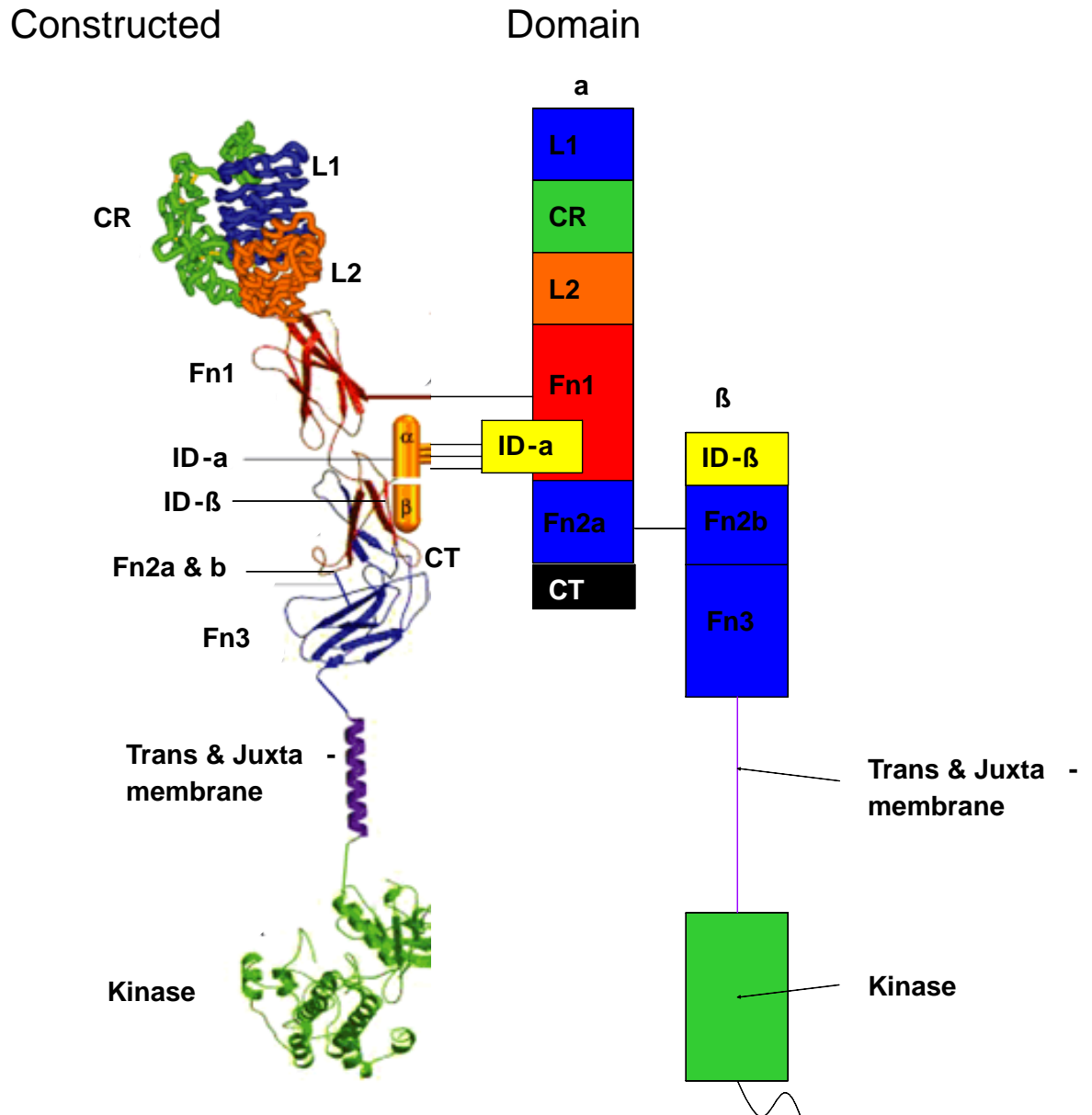
*Figure 13. Modular structure of the IR aligned against known structures of its domains.*

The supra-molecular domain organisation of the IR. The structures shown do not imply the actual structure of the IR, but are shown to indicate the domain structures known. Adapted from De Meyts and Whittaker[62].

The IR L1 domain contains the majority of residues involved in insulin binding and consists of a series of β-sheets [55]. Mutations in these residues lead to insulin binding defects[63], making it the key domain of binding[60] (Figure 14).
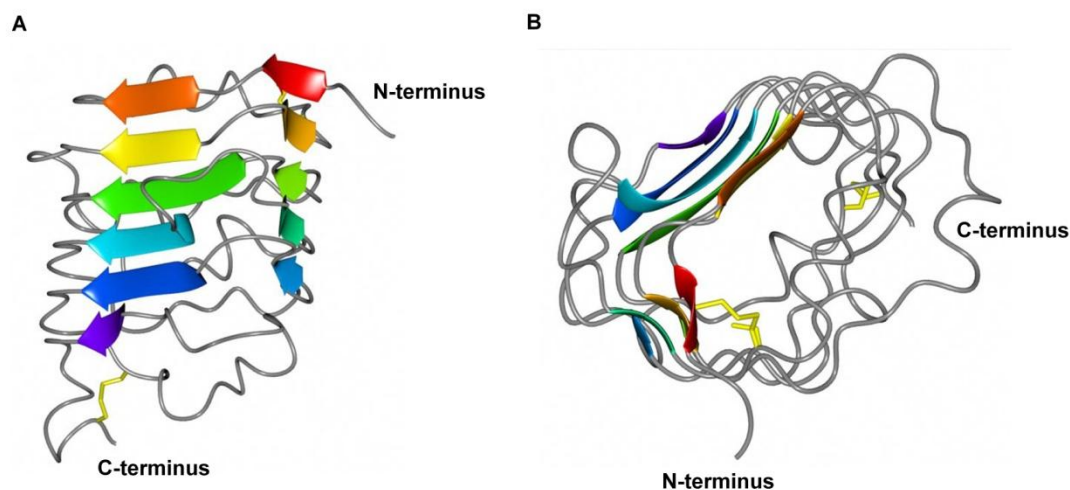
*Figure 14 Main chain fold of the IR L1 region observed in apo-IR ectodomain crystal structure.*

Two different views of the L1 domain of the human apo-IR ectodomain crystal structure. The β-strands are colour through in rainbow style (pdb used : 2hr7).

There have been many unsuccessful attempts to achieve a stable insulin:IR complex for structural studies. Interestingly, the CT peptide has been shown to be vital for effective binding of insulin to the IR[64] .  Despite its location at the end of the α-subunit it interacts with the L1 domain and insulin molecule, although their structural 3-D arrangement is unknown[65].

The isoforms of the IR have been identified in various human cells[66], as well as in the cells of rats, rhesus monkeys and sheep [67]. The expression of IR-A and IR-B is tissue-dependent.  The liver almost entirely expresses IR-B, whilst muscle, isolated adipocytes and cultured fibroblasts preferentially express IR-B over IR-A[67].  IR-A is found predominantly in leukocytes, whilst similar amounts of both isoforms are found in placenta, skeletal muscle and adipose tissue[68].  The steady-state binding of insulin is significantly higher in IR-A than in IR-B[69].  Importantly, the ability of IGF-I to compete with insulin for binding is significantly higher for IR-A than IR-B [67], and IGF-II binds preferentially to the A isoform over the B form with a very high nanomolar affinity (Figure 15, Table 1). The effects of IGF-II via IR-A are mitogenic compared to the metabolic effects of insulin [70]. Upon activation of the IR by insulin and phosphorylation of the IRS-1, this IR substrate acts as a docking centre where other enzymes and substrates can be activated to transduce further the effects of insulin[45,46].
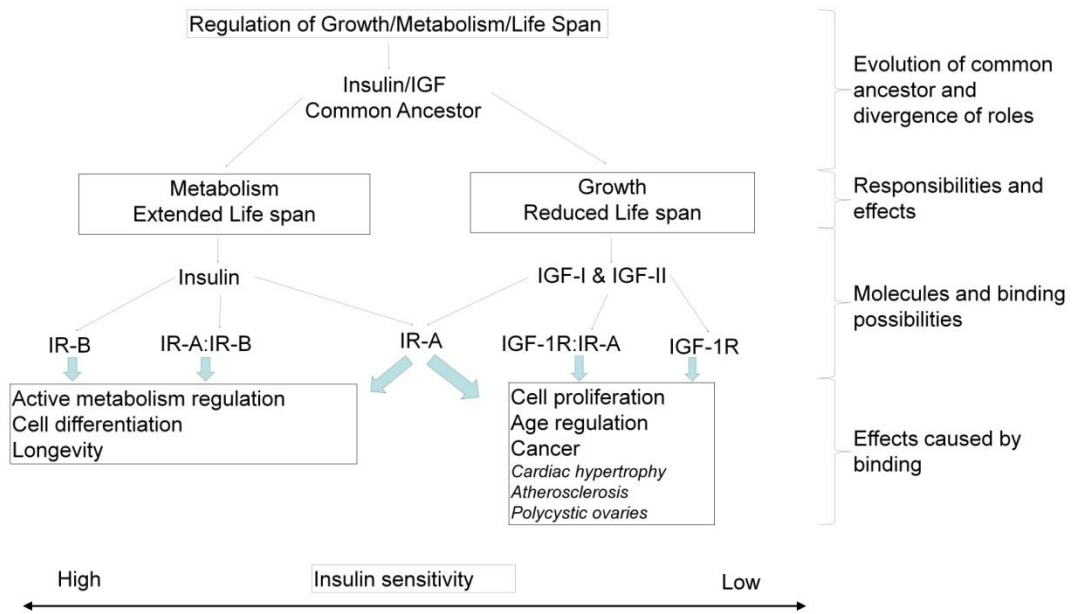
*Figure 15. The divergence of IGF and insulin, and their various effects.*


*Table 1. The binding affinity of insulin, IGF-I and IGF-II towards IR-A, IR-B and IGF-1R.*

$K_D$ is measured in nM[71].

|        | Insulin   | IGF-II  | IGF-I      |
|--------|-----------|---------|------------|
| IR-A   | 0.2-1.0   | 1.0-6.6 | 9.0->30.0  |
| IR-B   | 0.5-1.6   | 36.3    | >30.0      |
| IGF-1R | >30.0     | 0.5-4.4 | 0.2-2.5    |

## 1.5. Molecular Action of Insulin

There is a general consensus within the literature that the conformation of the active form of insulin monomer, i.e. its structure in the complex with IR must be different from the currently known structures of the hormone. This is evident as the residues that are crucial for the interaction with the IR are hidden under the B-chain C-terminus (Figure 16), in which conformational flexibility has also been indicated[4,72].

This observation results from a plethora of structural and functional data that focused on the structure-function relationship within the insulin molecule[73]. They have shown that the $B^{24}$-$B^{26}$ part of the β-strand region, the central part of the B-chain α-helix, and the N-terminal α-helix of the A-chain are important for high affinity binding between insulin and its receptor[74]. As mentioned before, insulin exists in either the T- and R-state[8], though it is possible that neither of these states are adopted upon receptor binding[75]. Insulin NMR studies have shown that the C-terminus of the B-chain is flexible[4], but the chemical crosslinking of $Gly^{A1}$ and $Lys^{B29}$ resulted in fully inactive hormone with a structure that is similar to the known storage form[3].

This indicates that upon binding of insulin to the IR, the C-terminus of the B-chain must move away exposing the hydrophobic core of the hormone (Figure 13) [3,6]. The importance of the movement of the B-chain C-terminus is highlighted as its removal can increase the binding affinity of an analogue[76]. The ability of the C-terminal region of the B-chain to move has led to a so-called induced-fit hypothesis in which the B-chain is required to move during binding to the IR in order to expose the hydrophobic core beneath[46,47]. It has also been shown that residues $B^{24}$-$B^{27}$ have an alternating pattern of interaction with the IR (L1 and ID-α)[77]. This alternating pattern is caused by the side chains in the β-strand projecting outwards in opposite directions. This indicates that for binding to occur, following the detachment, the C-terminus of the B-chain becomes inserted between the L1 and ID-α domains[77].

The 'active' hydrophobic core buried within the molecule consists of $Ile^{A2}$, $Val^{A3}$, $Cys^{A11}$, $Leu^{A16}$, $Leu^{B11}$ and $Leu^{B15}$ [3,6] (Figure 16). This hydrophobic core may represent the IR binding epitope; the flexibility of the C-terminus of the B-chain[4] would allow this core to be exposed.
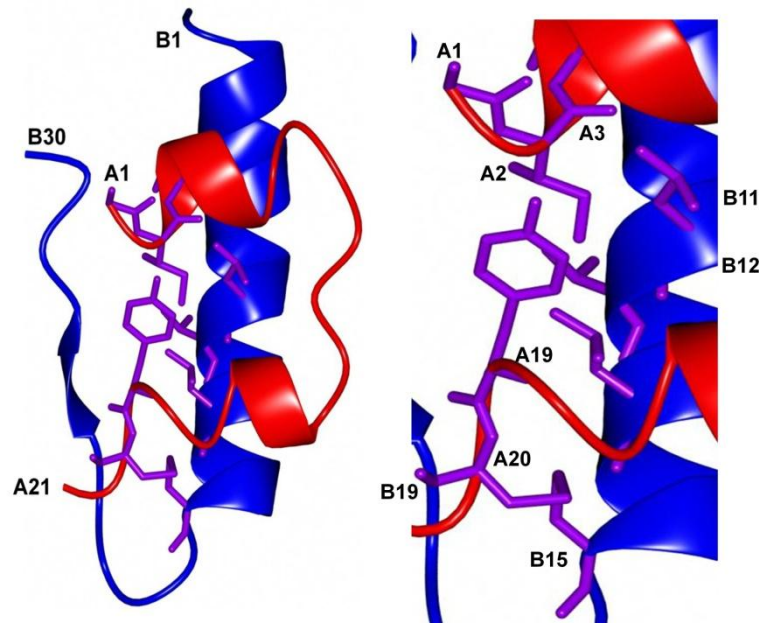
***Figure 16. The end of the B-chain hides the insulin binding epitope.***

Residues $B^{24}$-$B^{30}$ hide the insulin binding epitope, preventing any interaction with the insulin receptor. Residues $A^{1-3}$, $A^{19-20}$, $B^{11-12}$, $B^{15}$ and $B^{19}$ (shown in magenta)form the classical binding epitope[65].

Alanine-scanning mutagenesis helped to identify the receptor-binding epitopes within insulin molecule[78]. The crucial epitope consists of $Ile^{A2}$, $Val^{A3}$, $Tyr^{A19}$, $Gly^{B23}$, $Phe^{B24}$, $Val^{B12}$, and $Tyr^{B16}$ [78]. However, residues $Leu^{B6}$, $Gly^{B8}$, $Leu^{B11}$, $Glu^{B13}$ and $Phe^{B25}$, though not believed initially to be part of the binding site, were also identified as important for IR binding[78].

However, the residues predicted to be the binding epitope differ to some extent depending upon the method employed in the analysis. Mutagenesis and modifications[44,51] has identified $Gly^{A1}$, $Gln^{A5}$, $Tyr^{A19}$, $Asn^{A21}$, $Val^{B12}$, $Tyr^{B16}$, $Gly^{B23}$, $Phe^{B24}$, $Phe^{B25}$ and $Tyr^{B26}$ as the binding epitope. This is in some contradiction to the crystal structure derived model[74] which involves $Gly^{A1}$, $Glu^{A4}$, $Gln^{A5}$, $Tyr^{A19}$, $Asn^{A21}$ $Gly^{B23}$ and $Tyr^{B26}$, whilst alanine-scanning alone has shown $Ile^{A2}$, $Val^{A3}$, $Tyr^{A19}$, $Gly^{B23}$ and $Phe^{B24}$ [78] as the IR epitope. It is clear that the correct composition of the IR binding epitope within insulin cannot be confirmed until the crystal structure of the insulin:receptor complex is attained.

It has been shown that particular insulin analogues with substitutions at positions $A^{13}$ and $B^{17}$ – not considered as important for IR binding by several studies - exhibit an unusual correlation between binding affinity and biological potency[79]. Therefore, it was proposed by Schäffer *et al*[79] that these observed differences in binding affinities indicate a second binding epitope. This so-called site 2 must employ the above residues in addition to those in the classical binding site 1. A model of high affinity binding is proposed where the insulin molecule site 1 binds to the IR site 1, whilst the insulin site 2 of the same monomer

binds to site 2 on the other receptor α-subunit (Figure 17)[60]. It is believed that the binding of the insulin molecule to the IR causes a conformational change within the receptor. This change activates the kinases' activity in the β-subunits, resulting in their autophosphorylation[79].
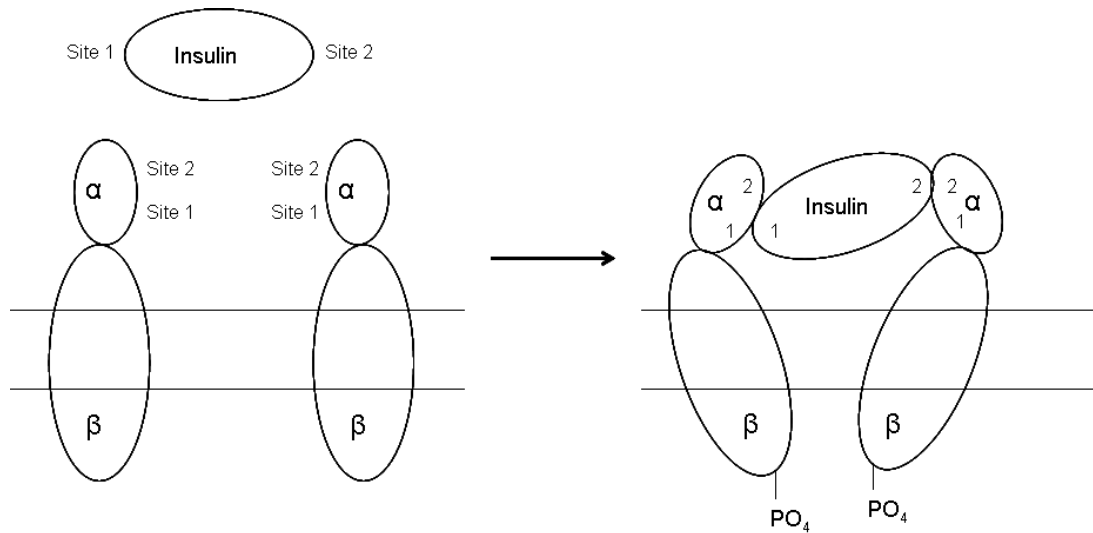


***Figure 17. A representation of the two receptor-binding domains on the insulin molecule and two binding sites on each α-subunit of the IR, demonstrating high affinity binding between the two.***

Insulin has been shown to form stable dimers and hexamers in the presence of $Zn^{2+}$. The dimer interface is formed of $Gly^{B8}$, $Ser^{B9}$, $Val^{B12}$, $Tyr^{B16}$, $Gly^{B23}$, $Phe^{B24}$, $Phe^{B25}$, $Tyr^{B26}$, $Thr^{B27}$ and $Pro^{B28}$, whilst $Phe^{B1}$, $Val^{B2}$, $Gln^{B4}$, $Glu^{B13}$, $Ala^{B14}$, $Leu^{B17}$, $Val^{B18}$, $Gly^{B20}$, $Leu^{A13}$, $Tyr^{A14}$ and $Glu^{A17}$ are engaged in hexamer formation. However, the most important residue for integrity of the hexamer is $His^{B10}$, the key side chain for coordination of the $Zn^{2+}$ ions[80].

## 1.6. The Causes of and Treatment of Diabetes Mellitus

Diabetes Mellitus affects 194 million people worldwide[81], and can be characterised by a marked decrease in the efficiency with which blood glucose levels are controlled.  Diabetes mellitus can be categorised into two classes, idiopathic and secondary. Idiopathic diabetes is further divided into two sub-classes, type I and type II[82], although this definition has not always been a clear division[83].

Idiopathic diabetes type I, or insulin dependent diabetes mellitus (IDDM) normally appears in childhood (juvenile-onset diabetes) as a result of the pancreatic beta-cells being destroyed in an autoimmune response.  This autoimmune response results in an insufficient amount of insulin being present within the body; this can be rectified by regular injection of the hormone.  The root of this disorder is believed to be genetic, although the actual onset can be caused by a viral infection.

Idiopathic diabetes type II, or non-insulin dependent diabetes mellitus (NIDDM) is characterised by constant high levels of blood sugar, or hyperglycaemia despite usually near-normal insulin secretion levels. NIDDM usually appears in patients over the age of 40 (adult-onset diabetes).  Hyperglycaemia in these cases is caused by insulin resistance, although a normal rate of glucose uptake can be achieved by artificially high levels of insulin.  The achievable maximum rate of glucose uptake is reduced.  There are usually two main reasons for the onset of NIDDM, insulin resistance, usually caused by a genetic defect or obesity, or insulin deficiency[84].

Secondary diabetes comprises a broad range of different health problems, all of which individually produce the symptoms of diabetes.  These health problems damage, injure or cause destruction of the pancreas.  The diabetes symptoms are a secondary condition produced as a result of the main illness, and if treated, they may disappear [85].

## 1.7. Chemical Synthesis of Insulin

The study of the interaction between insulin and the IR required the development of insulin analogues, and a synthetic method for their production was developed [73]. The solution based peptide approach of full synthesis of insulin uses successive steps of protection and deprotection. This was the standard method used to produce insulin analogues in the early research into insulin:IR[73], in a variety of labs[86,87]. This method required several steps to produce insulin or a single chain, which meant that the yield was severely limited.

Solution based peptide synthesis was improved upon by the use of solid-phase polypeptide synthesis. A polymer gel support is used to brace a growing polypeptide chain, which is added to by ligating a series of protected amino acids. This method has a greater efficiency that the solution based peptide synthesis[88].

Advancements in the methodology of chemical synthesis led to a simpler method for the condensation of modified polypeptide fragments into longer polypeptides, which is known as native chemical ligation[89]. This method is beneficial because native peptide bonds are formed. An advantage of this method is that non-natural amino acids can be incorporated[90].

Insulin analogues can also be prepared by a semi-synthetic method which is much more efficient than a total synthesis method, allowing larger quantities to be produced in a more rapid manner. The total synthesis of insulin involves full synthesis of a single chain followed by disulphide bond formation with the other chain [73,86,87] The semi-synthesis is based on the unique feature of insulin that the $B^{23}$-$B^{30}$ region (that contain side chains crucial for binding to the IR) can be proteolytically and selectively removed, and a modified synthetic peptide can be enzymatically ligated back onto the hormone[73,91,92] (Figure 18). A significant advantage of the semi-synthesis is that all insulin disulphide bonds are maintained in this process hence formation of non-native -S-S- isomers that tamper full synthesis of the hormone is avoided. This problem can be also overcome by the use of ester insulin intermediates[93]. Peptide excision is limited to areas that are outside the core of the insulin molecule, as defined by the disulphide bonds. The use of enzymes results in high yields, which can then be utilised by the enzyme for the reverse reaction[92,94].
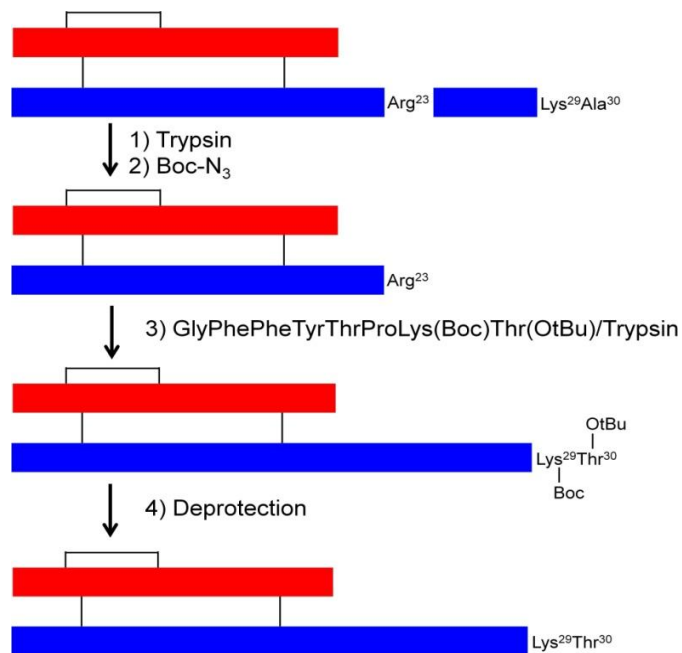
*Figure 18. Semi-synthesis of human insulin from porcine.*

The A-chain is coloured red, the B-chain is coloured blue. Reaction 1: Trypsin is used to cleave the B-chain C-terminal octapeptide at pH 9-9.5, yielding desoctapeptideinsulin (DOI). Reaction 2: Both the A-chain and B-chain N-termini are protected using Boc-N$_3$. Reaction 3: (Boc)$_2$-DOI is coupled to a protected octapeptide sequence resembling that of the human insulin B-chain N-terminal octapeptide. Reaction 4: Product is deprotected using TFA and anisole to yield native human insulin[91] (modified from Mayer 2007)[73].

## 1.8. Insulin Analogues

The purification of insulin was greatly advanced by the discovery that it could be crystallised [95] and that insulin formulations could be greatly improved [96] by the addition of zinc.

Insulin-based treatments have to balance the ability to mimic normal glucose profiles (Figure 11), avoiding hypo- and hyperglycaemia, whilst causing minimal disruption to patients.

Generally, there is a need for two types of insulin; a rapid acting insulin which would mimic an insulin peak that occurs after a meal, and a long acting insulin that is maintained at low level concentrations, for extended periods of time, e.g. very important in overnight glucose control (Figure 11)[97-99].

Insulin analogues are modified versions of human insulin[100] by the substitution of amino acids, and thus have for example, a different absorption period or length of duration.

New pharmaceutical preparations require knowledge of factors such as temperature, site of injection, exercise, massage, dosage levels and insulin formulation [98,101]. The initial rate of absorption of insulin is low as the injection is followed by an initial lag phase; this can be reduced by decreasing insulin concentration and injection volume. Diffusion into capillaries also limits the rate of absorption. Insulin is absorbed across the capillary barrier after the insulin hexamer has disassociated into dimers or monomers. For insulin to start taking effect, the zinc ions have to be removed and the injection volume diluted.

The development of recombinant DNA techniques led to the shift from animal sources to a more reliable and safe source of insulin. The first biosynthesis of human insulin was described by Goeddel *et al.*[102], using an *E. coli* expression system to produce separate A- and B-chains, followed by an exchange reaction between S-sulphonates and thiols [20,103] to form the correct disulphide bonds. Yeast has also been used in the production of insulin as it has a plasmid that codes for a single polypeptide chain, the insulin precursor, where the A- and B-chain are linked by a few amino acids. Due to the complex machinery present in yeast, the correct disulphide bonds are formed during fermentation. The linking amino acids were then cleaved to produce the mature insulin molecule[104].

Pharmaceutical companies use $Zn^{2+}$-stabilised hexamers in their formulations to delay release[98]. The treatment of diabetes is becoming increasingly dependent on analogues which utilise altered pharmacokinetics[105]. Rapid acting analogues have been designed to limit self-assembly [106] or to accelerate disassembly from their storage form, whilst long term acting analogues have been designed to promote self-assembly or slow disassembly[107].

Insulin is soluble at a wide range of pH values, >7 and <5, although between these

values it undergoes isoelectric precipitation[108]. At physiological pH insulin is less soluble in subcutaneous tissue, whilst still soluble in the acidic pH used in the pharmaceutical preparation. Insulin Glargine was designed to exploit the pH effect to become a long acting analogue [107]. The B-chain of the analogue is extended by the addition of paired Arg residues, whose positively charged side chains cause the pI to move 7.0 (insulin pI ~5.4)[107]. Glargine is formulated at pH 4.0, and upon injection undergoes pH-dependent precipitation, forming a deposit for slow release[107].

Dimer formation is a prerequisite to hexamer formation, residues involved in monomer-monomer interactions are key in dimer formation. There are several methods to make insulin monomeric at pharmaceutical concentrations[109]; an introduction of steric hindrance, abolition of metal binding sites, disruption of hydrophobic binding surfaces by the insertion of hydrophilic residues or the introduction of charge repulsion to residues in close proximity in the dimer interface.

Protamine, surfen and zinc[99,108] are all used to keep insulin in hexameric form. Protamines form complexes with insulin molecules as they are strongly basic proteins, consisting mainly of arginines. Surfen is a urea derivative that is produced synthetically. Zinc is at the centre of the insulin hexamer, but it has also been shown[98] that in a neutral pH solution lacking citrate or phosphate, zinc can slow down insulin uptake.

Human insulin is available in many forms. Short- and intermediate-acting, called regular and Neutral Protaime Hagedorn (NPH) respectively. The rapid-acting insulin analogues available currently are Lispro (Eli Lilly), Aspart (Novo Nordisk) and Glulisine (Sanofi-Aventis), whilst Glargine (Sanofi-Aventis) and Detemir (Novo Nordisk) are long-acting insulin analogues. The actions of insulins used for the treatment of diabetes are shown in Figure 19-A, whilst the profiles are shown in part B.

In regular insulin, the formulation is clear at neutral pH, and has 0.4% zinc added to allow the insulin monomers to associate and form hexamers. Phenol or m-cresol is added to the formulation to prevent micro-organism growth. NPH (Hagedorn Laboratory, Denmark), has protamine added to regular insulin in a 1:6 ratio. The positively charged protamine binds to the negatively charged insulin, in a neutral solution buffered with phosphate. NPH also has zinc and m-cresol added [98]. These additions in solution at neutral pH buffered with acetate forms insoluble insulin-zinc complexes [110], and it is this that the Lente insulin formulation is based upon. Insulin Detemir is a long-acting insulin analogue due to the addition of a fatty acid chain. The combinations of human insulins and insulin analogues are shown in Figure 19[98].

**A**

| Insulin Preparation | Onset of Action | Peak Action | Effective Duration of Action |
|---|---|---|---|
| **Rapid-Acting Insulin Analogues** | | | |
| Insulin lispro | 5-15 min | 30-90 min | 3-5 h |
| Insulin aspart | 5-15 min | 30-90 min | 3-5 h |
| Insulin glutisine | 5-15 min | 30-90 min | 3-5 h |
| **Short-Acting Insulin** | | | |
| Regular | 30-60 min | 2-3 h | 5-8 h |
| **Intermediate-Acting Insulin** | | | |
| NPH | 2-4 h | 4-10 h | 10-16 h |
| Lente | 3-4 h | 4-12 h | 12-18 h |
| **Long-Acting Insulin** | | | |
| Ultralente | 6-10 h | 10-16 h | 18-24 h |
| Insulin glargine | 2-4 h | Peakless | 20-24 h |
| Insulin determir | 2-4 h | 6-14 h | 16-20 h |
| **Insulin Mixtures** | | | |
| 70/30 human mix (70% NPH, 30% regular) | 30-60 min | Dual | 10-16 h |
| 75/25 lispro analogue mix (75% intermediate, 25% lispro) | 5-15 min | Dual | 10-16 h |
| 70/30 aspart analogue mix 70% intermediate, 30% aspart) | 5-15 min | Dual | 10-16 h |
| 50/50 lispro analogue (50% intermediate, 50% lispro) | 5-15 min | Dual | 10-16 h |
| 50/50 human mix (50% NPH, 50% regular) | 30-60 min | Dual | 10-16 h |

**B**
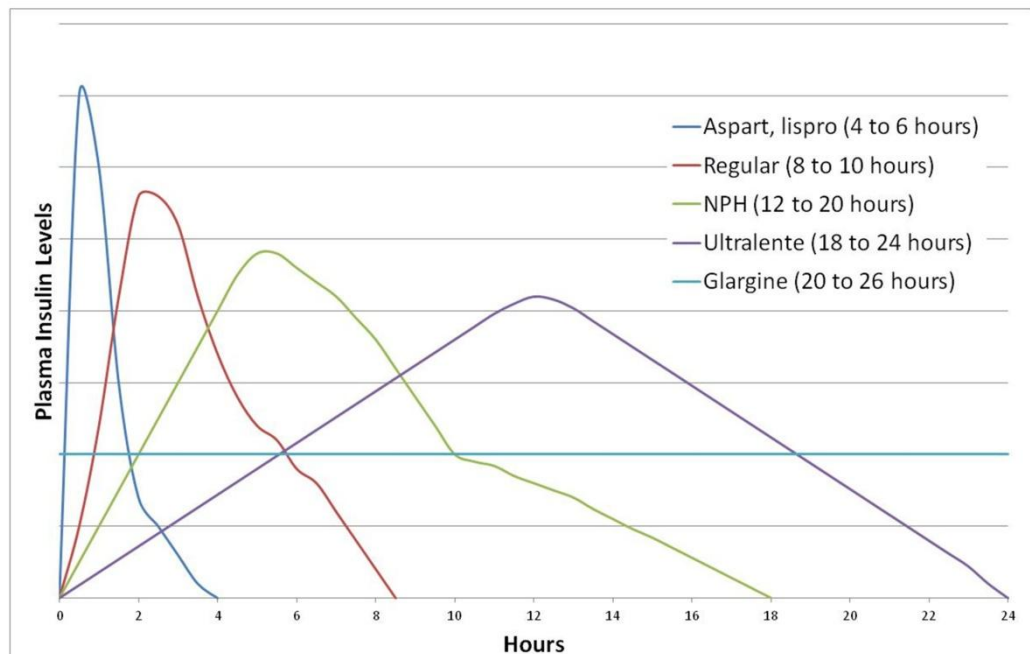


*Figure 19.  Human insulins and analogues available with their course of action.*

(A) A table of various human insulin preparations and their time course.  (B) Profiles of human insulins and analogues[98].

Since the introduction of insulin analogues, there have been improvements in glycemic control, a lowered risk of hypoglycaemia, lower glucose variability and for certain analogues, less weight gain[111]. Further developments in insulin analogues can further reduce these risks.

Like almost all protein based medications, insulin cannot be taken orally, as it would be broken down into fragments during the digestive process, causing the loss of activity. This means that insulin has to be administered subcutaneously, or inhaled. Currently insulin is delivered by injection, or directly by an insulin pump. Both of these methods allow the user to deliver a specific amount of insulin. Injecting insulin requires the use of needles, creating wounds. An insulin pump requires a permanent "link" into the patient, removing the need for needles to inject insulin, and eliminates the wounds that they produce. Alternative methods for insulin delivery are being developed, such as oral and nasal. For oral delivery, insulin is encapsulated in nanoparticles made with a blend of biodegradable polyesters[112]. Nasal delivery is based upon a similar technique to that of the oral system, where insulin is complexed with chitosan, to form insulin-chitosan nanoparticles[113]. When the management of diabetes is under control, there are few side-effects of taking insulin, though skin conditions from repeated injections are common[114].

Type II diabetes is characterised by a resistance to the effects of insulin. This resistance is normally restricted to the metabolic pathway, where the signalling pathway that affects cell proliferation and migration are uninduced[115]. It has been shown that the IR is over expressed in cancer cells[115,116], which means that circulating insulin can affect the growth of cancerous cells. With high blood glucose levels, and therefore raised insulin levels, in type II diabetes, the mitogenic effects of insulin can be increased. As described previously, cross-signalling is possible between insulin and IGF. A raised level of insulin is therefore more likely to produce uncontrolled mitogenic effects via IGF-R.

## 1.9. Importance of Insulin-Like Growth Factor in Insulin Biology

The Insulin-like Growth Factors 1 and 2 (IGF-I and IGF-II) are single-chain polypeptides with molecular masses of 7.6 kDa and 7.5 kDa respectively[9,10]. The names of the IGFs originate from their structural similarity to proinsulin and their similar metabolic effects, such as causing hypoglycemia[117]. This hypoglycemic effect is seen through activation of the IR rather than the IGF-1R. At high concentrations IGFs can induce effects through binding to the IR[117]. In a variety of tissues throughout the body, IGFs are synthesized *de novo*, and their mitotic effects can be seen in the local environment through intracrine, paracrine, and autocrine methods[11]. The IGFs are also involved in the induction of differentiation of whole cells, and processes within cells[12].

IGFs can bind to two types of membrane bound, TK receptors (IGF-1R and IGF-2R)[63]. IGF-1R comprises two α- subunits and two β-subunits; the β-subunit contains a transmembrane region[65]. IGF-1R preferentially binds IGF-I[65], although it has been shown that at high concentrations insulin is also able to bind due to the structural similarity shared with IGF[11] (Table 1). The preferential binding affinity of IGF-1R is IGF-I≥IGF-II>>Insulin. IGF-2R comparatively only binds IGFs with its preference being IGF-II>>IGF-I (Table 2), though this is not a signalling receptor with its functions still not fully understood. For example, its structure is identical to the mannose-6-phosphate receptor [118]. Most of the effects of both IGF-I and IGF-II are through IGF-1R, whilst in non-mammalian species the ability of IGF-II to bind to the mannose-6-phosphate receptor seems to have been lost[119].

The action of IGF-I, produced mainly in the liver, are growth hormone (GH) dependent[120] and its production can be slowed down by GH insensitivity, lack of hormone receptor, malnutrition or a failure of the downstream signalling pathway. In contrast, the action of IGF-II is not GH dependent[120]. Although IGF-II has no growth promoting effects in hypophysectomised rats[121], it has been shown in rodents and other species to function as a foetal growth factor. Deletion of the gene encoding IGF-II resulted in slowed intrauterine and perinatal growth[122], though its role in humans remains unclear. IGF-II is involved in cell growth, proliferation and differentiation, and is important in embryonic growth and prenatal development[122]. Mice embryos which lack the IR are only slightly affected, though severe growth problems occur if IGFR-I is missing[123]. IGF signalling may affect the risk of cancer with increased levels of IGF having been implicated as a cause of lung cancer[124].

**Table 2. Binding affinities between insulin and IGFs, and their receptors[57,70,119]**

|  | Type I receptor | Type II receptor | Insulin receptor |
| --- | --- | --- | --- |
| IGF-I | High | Very low | Very low |
| IGF-II | High or intermediate | High | Low |
| Insulin | Very low | None | High |

A characteristic of IGFs is that unlike other growth factors, their actions are regulated by a separate group of proteins known as Insulin-like Growth Factor Binding Proteins (IGFBPs), whose molecular weights vary from 17 to 43 kDa[11]. IGFBPs show high affinity to IGFs, and can compete with the IGF receptor for binding of these molecules. The actions of IGFBPs have been shown to prevent the hypoglycemic response caused by IGFs[125], along with prolonging the time taken to clear the IGFs[126], the transportation of IGFs to the site of action and to take effect upon the action of the IGFs[127]. Six IGFBPs have been identified and characterised (IGFBP-1 to IGFBP-6) (Table 3).

**Table 3. The effects caused upon binding of IGFBPs to IGFs.**

|  | Effect |
| --- | --- |
| IGFBP-1 | Regulate changes in serum IGFs and moderate IGFs actions at a cellular level |
| IGFBP-2 | Inhibit IGF (particularly IGF-II action upon particular cells) |
| IGFBP-3 | Major IGF carrier protein |
| IGFBP-4 | Control of growth in bone and cartilage |
| IGFBP-5 | Adheres tightly to fibroblast extracellular matrix (ECM). Control of growth in bone and cartilage |
| IGFBP-6 | Decreases steroidogenesis in response to IGF-I |

IGFBP-1 can cross the capillary barrier[128] and regulate the actions of IGF at a cellular level[128], whilst also regulating serum levels of IGF[11,129]. IGFBP-2 is believed to inhibit all IGF action[130], although it is also thought to stimulate IGF-I in certain cell types[11]. IGFs are carried in the circulation by IGFBP-3, which is the main IGF carrier. Large complexes of 150 kDa in size have been seen consisting of IGFBP-3, IGF and the acid labile subunit[131,132], which are unable to cross the capillary barrier[132]. IGFBP-4 and IGFBP-5 are involved in controlling bone and cartilage growth[133,134], whilst IGFBP-5 also has the ability to tightly adhere to fibroblasts extracellular matrix[135]. A decrease in steroidogenesis is stimulated upon binding of IGFBP-6 with FSH and LH following the signalling of IGF-I[11].

## 1.10. Crystallography, X-Ray Diffraction and Current Approaches to X-Ray Crystallography

The process from protein crystallisation to the final validation of its structure is a multifaceted and complex process. In this section, the main steps required for solution of the structure of a protein by X-ray crystallography will be described, and the specific methods applied in this thesis are covered in more detail (Figure 20).
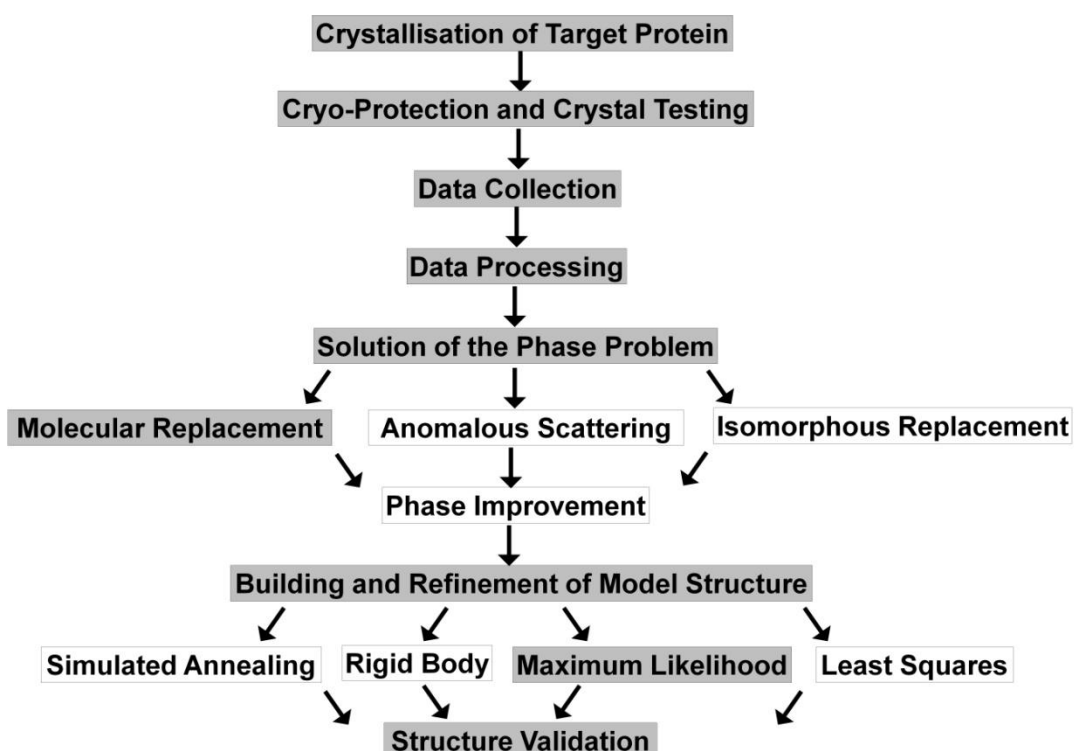


***Figure 20. Flowchart of the steps from crystallisation to structure validation.***
The methods used in this project are highlighted in grey.

### 1.10.1. Crystallography

The procedures by which biological macromolecules are crystallised are complex and poorly understood, despite the registration of the first X-ray diffraction pattern from protein crystal (pepsin) in 1934[136]. This leads to a seemingly unscientific approach of trial and error, i.e. screening the target protein with a large array of conditions, followed by optimisation of any resulting suitable conditions.

The formation of crystals relies upon changes of protein solubility that allow protein molecules to move from the solution- to solid- phase. There is a solubility limit at

an equilibrium at which the number of molecules entering solution equal those that are being lost from solution to solid[137]. Below the solubility limit the solute is said to be an undersaturated solution, and at the solubility limit it is a saturated solution (Figure 21). Crystals cannot grow in either of these states as they requires a system where thermodynamic forces are driving molecules from solution into the solid state, a supersaturated solution[137]. This transition comes when an activation energy barrier has been overcome[137], which could occur with the formation of nuclei by a number of aggregated protein molecules, or by the appearance of an ordered region. The more supersaturated the solution can become, the more likely the nuclei will form.
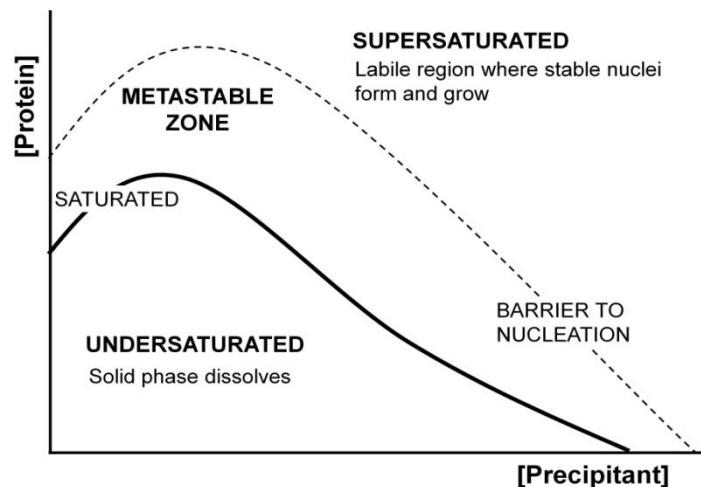


*Figure 21. A simplified phase diagram for protein crystal growth.*

At supersaturation there is high probability that critical nuclei will form, this is the labile region. At lower values of supersaturation, the metastable region, nuclei are not formed, but if already formed will continue to grow. Crossing the equilibrium line of saturation into undersaturation causes the solid phase to dissolve and molecules will be monomeric and disperse.

The population of protein molecules must be in the labile region in order for crystal growth to be initiated, by the formation of nuclei. Obtaining fewer but larger crystals is more desirable, which requires the crystal systems to move to the metastable region. The selection of crystallisation conditions is crucial in obtaining a balance between the two. If very high supersaturation is achieved, protein precipitation may occur, or a large number of microcrystals may form. Once crystals start growing in the metastable zone, supersaturation will decrease as the concentration of protein in solution drops meaning large crystals can be formed.

A supersaturated system can be achieved by several methods, though the system must be altered from its undersaturated state to supersaturated gradually to avoid precipitation[137]. The most commonly used methods include increasing salt concentration,

mixing the protein with a precipitant, and altering the pH or temperature of the system. Increasing salt concentrations causes competition between ions and protein molecules for water molecules to form hydration shells. If the salt concentration is increased, the water will be displaced from the protein, which promotes the protein molecules to self-associate to fulfil their electrostatic potential, causing solubility to be reduced[137]. Precipitating agents such as polyethylene glycols (PEGs) and organic solvents can be used in a similar way[138]. PEGs dehydrate a protein causing the protein molecules to push together and become excluded from the solution.

## 1.10.2. X-Ray Diffraction

When a protein crystal is exposed to an intense X-ray beam, the incident X-rays are diffracted by the 3D array of atoms that form the crystal, producing a diffraction pattern. Incoming X-rays are scattered by the electrons in each atom of the crystal. X-rays are electromagnetic waves that interact with the electrons and the same coherent frequency radiation is emitted to that of the incoming wave[139].
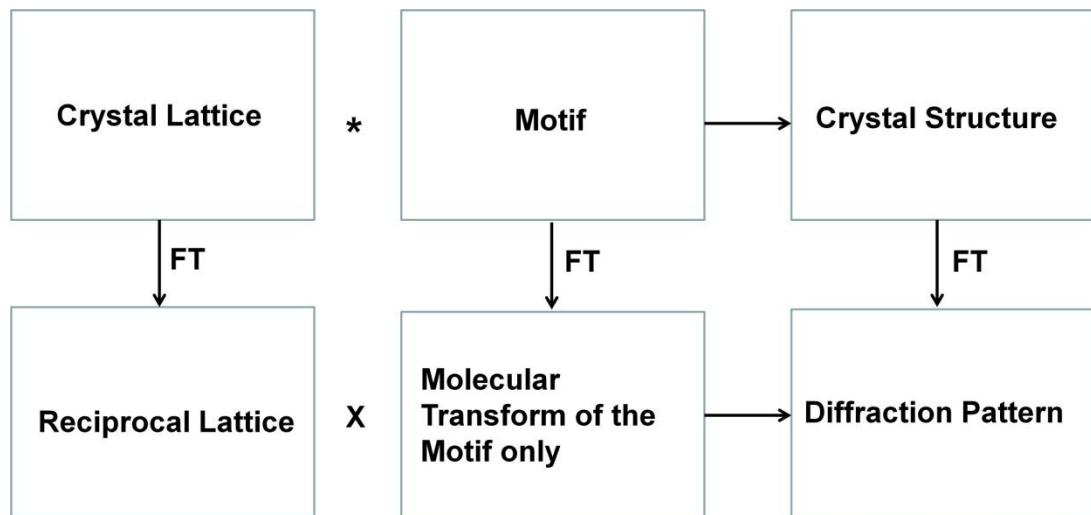


***Figure 22. The schematic representation of relationship between some main crystallographic terms.***
The terms used in this figure are X-product, *-convolution and FT-Fourier transform

This so-called reciprocal lattice is related to the crystal lattice and the diffraction pattern (Figure 22). This lattice is imaginary, yet is a helpful concept for determining the direction of the diffracted beam[139]. The unit cell dimensions in the reciprocal lattice are approximately reciprocal to the dimensions of the real unit cell, and correspond to the positions of the measured reflections in the diffraction pattern. Using the position of the

spots on the diffraction image, the repeating distances between the reflecting planes in the crystal lattice can be determined.  Bragg's Law relates the wavelength of the incident radiation to the angle of reflection and distances between the reflecting planes (Figure 23).
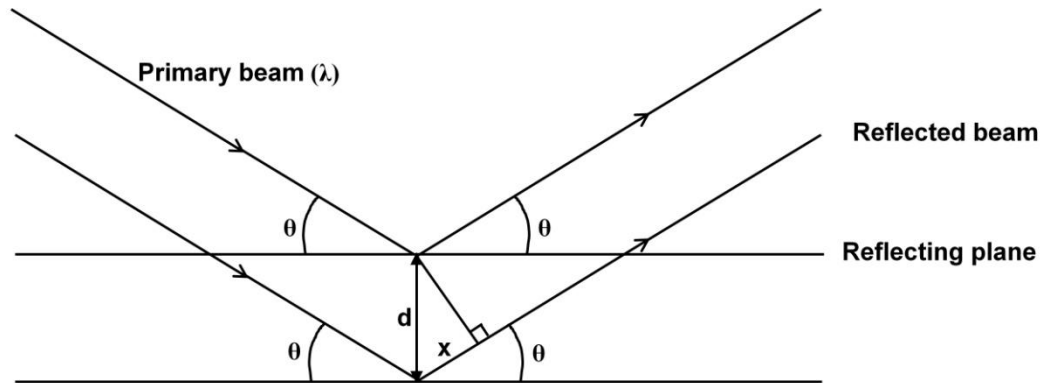


**BRAGGS LAW: $n\lambda = 2d\sin\theta = x =$** Difference in path between the reflected beam from 2 similar planes

*Figure 23. Representation of Bragg's Law.*

The incoming X-ray beam is reflected from the lattice planes.  The distance between the planes (d) can be determined from the wavelength ($\lambda$) of the beam and the angle of reflection ($\theta$).  The order of diffraction is (n).  Adapted from J. Drenth, Principles of Protein X-ray Crystallography[139].

When the crystal is rotated in the X-ray beam, the reciprocal lattice also rotates. This allows other points on the lattice to produce diffraction spots, which are represented by the Ewald sphere (Figure 24).
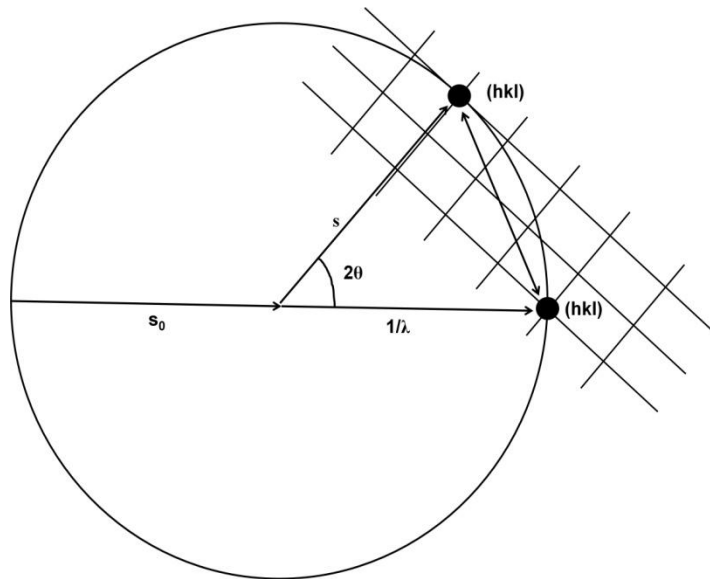
*Figure 24. The Ewald Sphere.*

The reciprocal lattice points will only produce diffraction when the points contact the surface of the Ewald sphere. The direction of the incoming beam is denoted by $\mathbf{s_0}$, the reflected beam by $\mathbf{s}$, and the sphere has a radius of $\mathbf{1/\lambda}$. Adapted from J. Drenth, Principles of Protein X-ray Crystallography[139].

It can be seen from the Ewald sphere that not all reciprocal lattice points can produce diffracting beams at the same time. The reciprocal lattice points, denoted *hkl*, must be in contact with the surface of the sphere to produce a reflection[139]. The reflecting planes giving rise to the observed diffraction are crystal lattice planes. These (imaginary) lattice planes divide up the **a, b** and **c** axes of the unit cell into a number of equal planes. These planes are denoted by their indices, *h, k* and *l*. Therefore, each reflection that is recorded on the diffraction images, also has coordinates or indices of *h, k* and *l*.

Determination of the arrangement of the molecules within the unit cell (crystal motive) that can be expressed by coordinates $(x_j, y_j, z_j)$ of the j-atom, utilises not only the directional information of reflections (e.g. $\theta$ angle) but the intensities of the reflections as well. They are necessary for calculation of the electron density of the crystal structure in later stages of crystal structure analysis[139].

X-rays interact with all electrons in the protein molecule, producing scattered waves. Every wave produced by every atom in every unit cell in the crystal adds up to obtain the scattering produced by the crystal. This is called the Structure Factor (F), because it relies upon the structure of the crystal forming motif of the atoms in all unit cells[139]. A factor that affects the intensity of the scattered X-rays is the temperature–related vibration of atoms within the crystal structure. Atoms vibrate around an equilibrium point, which means that the X-rays do not meet the equivalent atoms in the same position within each unit cell. This can reduce the intensity of the scattered X-rays. To account for this

effect, the atomic scattering factor of the atoms must be multiplied by a temperature-dependent factor (B)[139]. This results in four unknown parameters per atom in a structure; its coordinates $(x_j, y_j, z_j)$ and B.

The intensity I*(hkl)* of the diffracted reflection is proportional to the square of the amplitude of the structure factor *F(hkl)*, so therefore its modulus |F| can be calculated directly from the collected X-ray data (Equation 1).

$$I(hkl) \cong F(hkl)^2 \quad \rightarrow \quad \sqrt{I} \cong |F|$$

*Equation 1.*

The structure factor is related to the electron density, as it is a function of the electron density distributed over the unit cell (from the atomic scattering of X-rays). In principle, once *F(hkl)* has been reconstructed, the electron density ρ at every position x, y, z in the unit cell can be calculated. This is done by a Fourier transform, as *F(hkl)* is the Fourier transform of ρ(xyz) of the entire unit cell. The reverse is also true, so ρ(xyz) can be written as a function of the *F(hkl)*[139]. Using the experimentally derived information (Equation 1), this equation can be re-written (Equation 2).

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l F(hkl) \exp[-2\pi i\ (hx + ky + lz)]$$

*Equation 2.*

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \exp\left[-2\pi i\ (hx_j + ky_j + lz_j) + i\alpha(hkl)\right]$$

| Volume of Cell | Structure Factor Amplitudes | Position of Atoms j | Phases |

*Equation 3.*

A new equation (Equation 3) allows the electron density ρ(xyz) at every position within the unit cell to be calculated. However, one key problem remains, and that is the unknown phase angle of the diffracted X-rays. These values cannot be derived from the diffraction pattern, giving rise to the phase problem, which must be solved for a successful determination of atomic coordinates $((x_j, y_j, z_j))$ of each atom.

### 1.10.3. Data Processing

The collected X-ray data needs to be processed to provide working, main X-ray data set. The HKL2000[140]programme, which uses DENZO and SCALEPACK, was used to process the data in this thesis. Initially the unit cell dimensions, indexing of the data and determination of its Laue point group symmetry must be determined first. Indexing comprises searching for peaks on a single diffraction image, followed by comparing the angles and distances of the observed diffraction spots to match the relationships characteristic for each of the seven crystal systems. The crystal system which has the highest symmetry and best fit between the observed and predicted diffraction patterns is chosen[141]. Following indexing, and refinement of the cell parameters the images are integrated to give measured intensities of the X-ray reflections. The correct integration includes several additional corrections, for example reflection positional refinement, accurate assessment of the background, absorption and anisotropy resulting from crystal shape, size and properties, etc[141]. After integration, the data are scaled and if there is more than one data set, merged as well. Scaling puts all images and symmetry equivalent reflections on the same scale. Differences can occur here due to unequal exposures in high and low resolution, beam decay and crystal decay[141]. The quality of the data is usually assessed by checking the ratio of intensity to error of intensity $<I/\sigma(I)>$, and the agreement between symmetry related reflections, $R_{merge}$[141] (Equation 4). $R_{pim}$ (precision indicating merging R-factor) is an alternative, and more accurate, measure of data quality as that it takes account of multiplicity in the data[142].

$$R_{merge} = \frac{\left[\sum_{hkl}\sum_{i}\left|I_{i(hkl)} < I_{(hkl)} >\right|\right]}{\left[\sum_{hkl}\sum_{i}I_{(hkl)}\right]}$$

***Equation 4.***

The exact type of the space group (e.g. taking into account occurrence of some systematic absences) can be adjusted at the last stage of data processing[141].

## 1.10.4. Solution of the Phase Problem

### 1.10.4.1. Molecular Replacement

If a suitable model of a studied protein already exists (with minimum sequence homology/similarity ca. 30%), molecular replacement is the fastest and most convenient method to solve the phase problem for an unknown structure. An appropriate model is chosen based upon its sequence homology or belonging to the same fold family as the target. The non-crystallographic symmetry of the model can also be considered to account for a particular oligomeric state of the protein. The placement of the model molecule in the new unit cell comprises of two separate steps: rotational and translational searches[141]. Computational methods are used to search all possible orientations and positions of the model in the new unit cell. MOLREP[143] is an example of one of these programs and was used in this thesis. The rotation step is carried out first, using Patterson maps of the known model and unknown protein. A Patterson map is a vector map, where vectors between atoms in the structure are represented as vectors from the origin to peaks in the Patterson map[139]. As all phases are set to zero, no prior phase knowledge is required. The rotation function calculates the relationship between Patterson vectors of the model and unknown structure, rotating the model until the two different maps overlap giving the best superposition[139]. The subsequent translational search requires calculation when the two molecules to overlap in real space. The simplest way to do this is trial and error approach, where the model is moved through the whole asymmetric unit of the unknown protein, with the comparison of the fits of $F_{calc}$ and $F_{obs}$ structure factors monitored by calculation of the $R$ factors[144] (Equation 5).

$$R = \frac{\sum_{hkl} ||F_{obs}| - |F_{calc}||}{\sum_{hkl} |F_{obs}|}$$

***Equation 5.***

### 1.10.4.2.   Isomorphous Replacement and Anomalous Scattering

If a suitable molecular replacement model is not available or molecular replacement fails even with an appropriate model, then another method of solving the phase problem is required. Isomorphous replacement is a technique that involves attaching heavy atoms such as Hg, Pt or Au (or their compounds) to the protein by either soaking native protein crystals in solutions of complexes of such metals, or co-crystallisation of a protein solution with heavy atom compounds. Data sets are collected for both the native crystal and the heavy atom ones. Since heavy atoms have a large number of electrons, there are differences in the diffraction data that can lead to identification of heavy atom sites in the unknown structure. Their coordinates facilitate preliminary estimation of the phases.

Anomalous scattering-based approach relies upon utilisation of so called anomalous scattering that is especially pronounced at wavelengths similar atom absorption range. Single wavelength anomalous dispersion (SAD) and multi-wavelength anomalous dispersion (MAD) are two techniques used. MAD uses the same crystal with anomalous diffraction collected at different wavelengths. SAD uses two crystals collected at a single wavelength, with one crystal used to collect a native data, and another with a heavy atom Since C, N and O atoms do not give strong anomalous scattering, a selenomethionine derivative is often used when native Met side chains are replaced by SeMet amino acids on the protein expression level. Alternatively, soaking of atoms known for their anomalous diffraction properties (e.g. Sm) can be used here as well. The SeMet approach has a great advantage over any soaking method as it alleviate potential changes of unit cell dimensions (i.e lack of so-called isomorphism) upon soaking of the heavy atoms.

### 1.10.5. Refinement of the Model using Maximum Likelihood Method

Following the solution of the phase problem, the process of model building and refinement of its structure can follow.  When the first model has been produced, the calculated structure factors generally are in poor agreement with the observed structure factors[139].  Refinement improves and completes the model closing the gap between the calculated and observed structure factors.  The method of refinement used in this thesis was maximum likelihood, implemented in REFMAC[145].  Maximum likelihood is a statistical method for fitting a mathematical model to some data, by stating that the best set of

parameters on the evidence of the data is the one that explains the observations with the highest probability[137]. In crystallography, this means the likelihood for a reflection is the probability of the observed structure factor amplitude, given the current model[137]. The correlation between the data and the model is optimised by adjustment of the model to the observed electron density. This involves changing the positional parameters and temperature factors of all non-hydrogen atoms in the structure[139]. Stereochemical information on bond lengths, angles and Van der Waals contacts can be applied to make an important contribution to refinement. This most commonly involves restrained refinement where these parameters are allowed to vary around standard values, allowing easy movement of small parts of the model[139]. Molecular graphics programmes, such as Coot[138] used in this work, allow manual manipulation and intervention into the model to achieve better fit to the electron density map. Two types of calculated maps are output from refinement, usually 2mFo-DFc map, which represents the content of the crystal, and mFo-DFc which represents the difference between the contents of the crystal and the current atomic model[145]. Refinement is monitored by calculation of the $R$ factor and $R_{free}$ factors. These values measure the fit of the calculated structure factors to the observed data, where the values will decrease as the observed and calculated structure factors agree more closely[144].

The $R_{free}$ factor is a more reliable measure of the refinement quality than the $R$ factor, as it is calculated using a small subset of reflections (usually 5%) that are randomly removed from the data before refinement is started. If the $R_{free}$ value is decreasing steadily then it can be regarded as an unbiased improvement of the model[144]; its increase may suggest over-fitting of the maps. The $R_{free}$ will always be slightly larger than the $R$ factor, and at the end of refinement they should ideally be within a few percent of each other.

### 1.10.6. Structure Validation

Once a structure has been solved and refined, it is likely that it contains some errors that have been introduced through fitting of the model to the electron density maps. These errors have to be corrected before the structure can be published or deposited in the Protein Data Bank (PDB). One measure of the reliability and quality of the model is the Ramachandran Plot, where the stereochemistry of the protein main chain is investigated. The dihedral angles of $\varphi$ and $\Psi$ for each residue are plotted against each other, and are clustered into regions of preferred, allowed and outlier values[139]. Other measures for the removal of errors of the model (implemented in programmes such as MolProbity[146]) is

checking residues with unusually high B factors, close Van der Waals contacts, unfavourable or unlikely hydrogen bonds, and unusual side chain rotamers.

## 1.10.7. Current Approaches to X-Ray Crystallography

Modern synchrotrons allow more data set to be collected for smaller and worse quality crystals. Diffraction images at synchrotron are collected charge couple device (CCD) detectors. These detectors produce a charge every time photon hits the detector. The total charge is calculated at the end of the image. CCD detectors are being replaced with PILATUS detectors. PILATUS (**pi**xel appar**atu**s for the **S**LS) has many advantages over CCD detectors as it uses single photon counting instead of an accumulative method in CCD detectors. When a photon hits the PILATUS detector, it immediately produces a charge, which is recorded if it exceeds a predetermined threshold. This results in no read-out noise associated with the accumulative method used in CCD detectors. Due to lack of background noise, a superior signal-to-noise ratio is achieved, and a readout time of in a range of 5 ms (compared to 1-120 s in CCD detectors) allows data sets to be collected in a formidable short periods of time[147].

New generation synchrotrons produce light with such a low divergence that it gives micro-beams smaller than 10 μm[148]. It has a great impact on data collection strategy as the intra-crystal variations within one single mounted crystal can be explored. A mesh-scan probes the diffraction quality across the entire crystal and produces a diffraction contour map allowing the best data to be collected[149].

Future developments at synchrotron beam-lines are challenging existing technological setups. For example, crystals are currently vitrified following cryoprotection in liquid (or gaseous) nitrogen. There are plans to locate the entire environment of an X-ray experiment in vacuum, which should reduce the air-related absorption of X-rays and result in a more intense and clearer image at longer wavelengths. This would require a new way of mounting crystals. The advancements of the PILATUS-like detectors leads to development of semi-spherical detectors that will be able detect X-rays scattered at large diffracting angles expanding the resolution of the data even further.

# Chapter 2.  High Affinity Insulin Analogues as Tools for Insulin:Insulin Receptor Studies

## 2.1. Summary

The current crystal structures of insulin describe the inactive, storage form of the hormone. The hydrophobic core of insulin contains the residues $A^{1-3}$, $A^{19-20}$, $B^{11-12}$, $B^{15}$ and $B^{19}$ that are involved in binding the IR[3,6]. Various studies have shown that the B-chain C-terminus is mobile[4,72] and must move to allow the binding epitope to become exposed[3,6].

This chapter describes efforts that probed whether a change or modification of an amino acid that would help induce a turn in the B20-B30 chain would expose the binding epitope, and increase the chance of finding the active form of insulin. Subsequently novel mutations can lead to improved insulin binding affinity[100]. To achieve these series of unique semi-synthetic insulin analogues were created to expose the binding epitope; previously described semisynthetic methods have been employed in making of these analogues[73,92,150].

As the result high affinity (~200-500%) analogues truncated at $B^{26}$ were produced to explore the nature of the transition of the B-chain C-terminus. The high affinity analogues presented in this chapter show a structural convergence in the form of a new β-turn in the B-chain at position $B^{26}$, the B26 turn. The formation of the B26 turn at $B^{24}$-$B^{26}$ is associated with the *trans*-to-*cis* isomerisation of the $B^{25}$-$B^{26}$ peptide bond, which is initiated by the *N*-methylation of $B^{26}$. This turn and the resulting high affinity associated with these analogues is in agreement with the literature. With this I propose that the B26 turn is the active conformation of insulin upon binding to the IR.

Subsequently, click-chemistry was used to create a cross-linked analogue that is permanently fixed in an active conformation. The 1,2,3-triazole ring chemically links side chains in an aim to mimic the B26 turn. A longer cross-link causes structural rigidity, which obscures the binding epitope and causes a greatly reduced binding affinity (0.13%). Shortening the cross-link reduces the rigidity and results in a binding affinity close to WT insulin (101.5%).

The resulting conformation arising from the presence of the B26 turn causes the previously buried amino acids, important in IR binding to become unmasked. High affinity analogues provide key structural details vital for a new approach in the design of future insulin analogues.

## 2.2. Crystallisation, Structure Determination and Refinement of Insulin Analogues to Probe the Active Conformation of Insulin

The insulin analogues crystallised and studied in this chapter are displayed in Table 4 with their sequence change from WT insulin illustrated along with their binding affinity.

***Table 4. A comparison of insulin analogues structure and their relative receptor binding affinity.***

Changes in the B-chain sequence are indicated in bold. [$N$MeXaaB$^n$]: methylation of the $N$-peptide atom preceding Xaa B$^n$ amino acid; -NH$_2$: C-terminal carboxy amide.

| B-Chain Position | 24 | 25 | 26 | 27 | 28 | 29 | 30 | Affinity |
|---|---|---|---|---|---|---|---|---|
| **WT:** | F | F | Y | T | P | K | T | 100 |
| [$N$MeAla$^{B26}$]-DTI-NH$_2$ | F | F | **Me A - *CONH*$_2$** | | | | | 465 |
| [$N$MeAla$^{B26}$]-insulin | F | F | **Me A** | T | P | K | T | 21 |
| [$N$MeTyr$^{B26}$]-insulin | F | F | **Me Y** | T | P | K | T | 21 |
| [D-Pro$^{B26}$]-DTI-NH$_2$ | F | F | ***D-P - CONH*$_2$** | | | | | 356 |
| [NMeHis$^{B26}$]-DTI-NH$_2$ | F | F | **Me H - *CONH*$_2$** | | | | | 214 |

Insulin analogue [$N$MeTyr$^{B26}$]-insulin was dissolved in 20 mM HCl to a concentration of 10 mg/ml. The best crystal identified from the screening was obtained from drops equilibrated 1:3 (protein:well) in well solution consisting of 0.055 M Na$_2$SO$_4$ pH 3.0 (0.5 ml well volume), and grown at 21°C. The crystal was flash-cooled in liquid nitrogen without any cryoprotection. The X-ray data were collected at ESRF on station ID29, and processed using HKL2000[140]. The crystal structure of [$N$MeTyr$^{B26}$]-insulin was solved by molecular replacement using default settings in Molrep[143]. The search model consisted of a truncated model of monomeric insulin 1mso.pdb[151], which involved the removal of residues B$^1$-B$^5$ and B$^{25}$-B$^{30}$. Subsequent manual model building and refinement with Coot[138] and 10 cycles of REFMAC5[145], resulted in a final model which was validated by MolProbity[146]. Data collection, reduction and refinement statistics are detailed Table 5. Crystal conditions are documented in Table 6.

The other insulin analogues were crystallised (in monomeric form), and had their structure determined and refined as per [$N$MeTyr$^{B26}$]-insulin. The differences in the crystal conditions and data collection, reduction and refinement conditions are detailed in Table 6 and Table 5 respectively.

*Table 5. Data collection, reduction and refinement statistics for monomeric insulin analogues.*

| | [NMeAla$^{B26}$]-DTI-NH$_2$ | [NMeAla$^{B26}$]-insulin pH 3.0 | [D-Pro$^{B26}$]-DTI-NH$_2$ | [NMeTyr$^{B26}$]-insulin monomer |
|---|---|---|---|---|
| PDB code | 2wru | 2wrx | 2wrw | 2ws1 |
| Data collection Beamline | ESRF, ID23-1 | ESRF, ID14-1 | ESRF, ID14-2 | ESRF, ID29 |
| Wavelength (Å) | 1.071 | 1.0723 | 0.9330 | 0.9762 |
| Space group | I4$_1$22 | P4$_1$2$_1$2 | I4$_1$22 | I4$_1$22 |
| Cell dimensions | | | | |
|   *a, b, c* (Å) | 38.7 38.7 123.1 | 39.2 39.2 124.9 | 39.4 39.4 123.9 | 38.2 38.2 124.0 |
|   α, β, γ (°) | 90.0 90.0 90.0 | 90.0 90.0 90.0 | 90.0 90.0 90.0 | 90.0 90.0 90.0 |
| Resolution (Å) | 50–1.57(1.63–1.57) | 30.0–1.50(1.55–1.50) | 50.0–2.41(2.45–2.41) | 30.0–1.5(1.53–1.50) |
| $R_{sym}$ | 0.057(0.162) | 0.063(0.469) | 0.055(0.217) | 0.051(0.424) |
| $I / \sigma I$ | 23.0(11.0) | 9.2(5.5) | 14.6(9.2) | 19.2(9.5) |
| Completeness (%) | 95.0(61.6) | 94.6(60.5) | 98.5(98.1) | 95.1(52.2) |
| Redundancy | 10.6(2.0) | 13.4(1.5) | 6.0(5.0) | 9.1(5.5) |
| Refinement | | | | |
| Resolution (Å) | 37.0–1.57 | 37.4–1.50 | 25–2.41 | 30–1.6 |
| No. reflections | 6339 | 14801 | 1977 | 6145 |
| $R_{work} / R_{free}$ | 0.179/0.255 | 0.180/0.250 | 0.236/0.300 | 0.184/0.247 |
| No. atoms | 438 | 862 | 385 | 445 |
|   Protein | 376 | 760 | 259 | 404 |
|   Ligand/ion | 8 | 1 | - | - |
|   Water | 61 | 101 | 26 | 41 |
| *B*-factors | | | | |
|   Protein | 21.1 | 35.2 | 23.8 | 21.5 |
|   Ligand/ion | 34.7 | 35.4 | - | - |
|   Water | 18.8 | 30.2 | 22.5 | 28.3 |
| R.m.s. deviations | | | | |
|   Bond lengths (Å) | 0.022 | 0.023 | 0.017 | 0.022 |
|   Bond angles (°) | 1.744 | 1.980 | 1.632 | 1.707 |
| Ramachandran statistics (%) | | | | |
| Preferred/Allowed/Outliers | 97.2/2.8/0 | 94.0/4.8/0 | 95.24/4.76/0 | 100.0/0/0 |

| | [NMeHis[B26]]-DTI-NH$_2$ | Cyclo[G-PrgFYTPK(N$_3$)T]-insulin | Cyclo[GFFY-Pent(N$_3$)-P-G-Prg]-insulin |
|---|---|---|---|
| PDB code | 2wrv | | |
| Data collection Beamline | ESRF, ID14-2 | Diamond, i03 | ESRF |
| Wavelength (Å) | 0.9330 | 0.9763 | 0.8726 |
| Space group | I4$_1$22 | P2$_1$2$_1$2 | P3$_1$ |
| Cell dimensions | | | |
| $a$, $b$, $c$ (Å) | 39.2 39.2 123.6 | 55.4 56.9 60.0 | 61.0 61.0 82.0 |
| α, β, γ (°) | 90.0 90.0 90.0 | 90.0 90.0 90.0 | 90.0 90.0 120.0 |
| Resolution (Å) | 27.0–2.15(2.19–2.15) | 27.72-1.57 (1.61-1.57) | 50.00-1.54 (1.57-1.54) |
| $R_{sym}$ | 0.067(0.269) | 0.079 (0.309) | 0.068 (0.488) |
| $I / \sigma I$ | 14.5(6.8) | 11.9 (2.3) | 7.5 (3.6) |
| Completeness (%) | 97.8(97.4) | 82.4 (40.1) | 100.0 (100.0) |
| Redundancy | 7.2(7.0) | 5.9 (3.2) | 5.7 (5.7) |
| Refinement | | | |
| Resolution (Å) | 37.34–2.15 | 27.72-1.57 | 52.82-1.54 |
| No. reflections | 2686 | 21129 | 48163 |
| $R_{work} / R_{free}$ | 0.199/0.277 | 0.203/0.265 | 0.167/0.207 |
| No. atoms | 413 | 1778 | 2968 |
| Protein | 367 | 1491 | 2492 |
| Ligand/ion | 1 | 15 | 95 |
| Water | 45 | 272 | 381 |
| $B$-factors | | | |
| Protein | 20.6 | 18.759 | 18.690 |
| Ligand/ion | 30.0 | 31.611 | 12.603 |
| Water | 21.1 | 30.671 | 31.727 |
| R.m.s. deviations | | | |
| Bond lengths (Å) | 0.021 | 0.025 | 0.033 |
| Bond angles (°) | 1.677 | 2.141 | 3.081 |
| Ramachandran statistics (%) | | | |
| Preferred/Allowed/Outliers | 90.48/9.52/0 | 96.9/3.1/0 | 98.7/0.4/0.9 |

## Table 6. Crystal conditions for monomeric insulin analogues.

| | [NMeAla[B26]]-DTI-NH$_2$ | [NMeAla[B26]]-insulin | [NMeTyr[B26]]-insulin | [D-Pro[B26]]-DTI-NH$_2$ | [NMeHis[B26]]-DTI-NH$_2$ |
|---|---|---|---|---|---|
| Well Condition | 0.32 M Na$_2$SO$_4$, pH 3.0 | 0.18 M Li$_2$SO$_4$, 0.1 M NaAc pH 3.0 | 0.055M Na$_2$SO$_4$ pH 3.0 | 0.015M Cs$_2$SO$_4$ pH 3.0 | 0.1M NaCit, 0.3M Tris pH8.2, 0.6 mM ZnAc, 0.06% phenol |
| Drop Ratio (Protein:Well) | 1:3 | 1:3 | 1:3 | 1:3 | 1:3 |
| Temperature | 21°C | 21°C | 21°C | 21°C | 21°C |
| Cryoprotectant | None | None | None | None | None |

## 2.3. The Formation of a Unique β-Turn in High Affinity Analogues

The most potent (465%) insulin analogue [$N$MeAla$^{B26}$]-DTI-NH$_2$ which contains a $N$-methylation at position B$^{26}$ with the B-chain truncated after the Ala$^{B26}$, was crystallised as a monomer revealing a new conformation in the B$^{24}$-B$^{26}$ position. This feature resembles a type II β-turn at Phe$^{B24}$-$N$MeAla$^{B26}$, referred to here as the B26 turn (Figure 25). Within this turn a *cis* conformation of the Phe$^{B25}$-$N$MeAla$^{B26}$ peptide bond is formed, causing the B-chain residues B$^{22}$-B$^{30}$ to move away from its usual position seen in WT dimers and hexamers, exposing the insulin binding epitope. The B26 turn in stabilised by the presence of a 2.8 Å hydrogen bond that is formed between the CO group of the B$^{24}$ residue and the NH$_2$ terminus of the B$^{26}$ position, which mimics the main chain NH. The hydrogen bond that is formed between the í and í+3 residues results in the truncated B-chain C-terminus pointing away from the main receptor binding epitope. This confirms previous work which showed that the flexibility of the B-chain C-terminus[4] is required during the induced-fit hypothesis upon binding to the IR[3,6] to expose the hydrophobic core[73,77].

The B26 turn causes a structural re-arrangement of the B-chain. A rotation of the Phe$^{B25}$ side chain occurs, which results in an overhang of the A-chain part of site one (~7.7 Å from Ile$^{A2}$) (Figure 26). This rotation is in contrast to Phe$^{B24}$, which remains in place. Along with the B26 turn the A-chain N-terminal helix rotates by ~40°, causing a displacement of ~3 Å. Tyr$^{A19}$ is shifted by 2.9 Å, enabling a hydrogen bond to be formed with Gln$^{A5}$, which results in the non-polar receptor binding surface being extended (Figure 26). The crystal structure of [$N$MeHis$^{B26}$]-DTI-NH$_2$, another highly active insulin analogue (214%), contains the same geometries (Cβ of both are only 0.3 Å different from each other) and have the same hydrogen bond network.
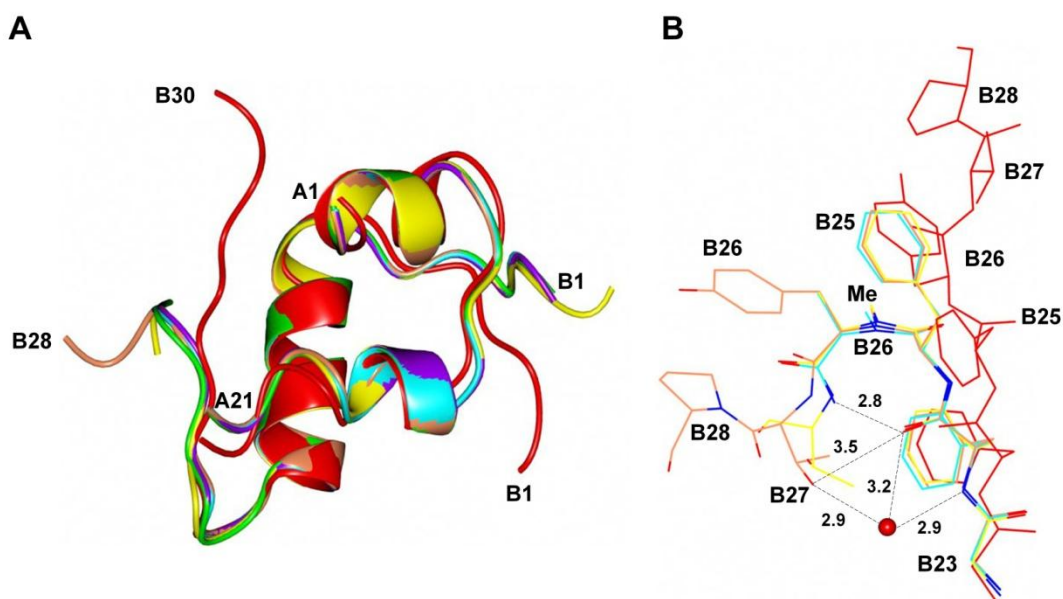
**Figure 25.  Structural features of crystallised insulin analogues.**

(A) An overlay of the general fold of crystal structures of the highly-active and other insulins: Red: wild-type insulin, Cyan: [NMeAla^B26]-DTI-NH_2, Purple: [NMeHis^B26]-DTI-NH_2, Green: [D-Pro^B26]-DTI-NH_2, Coral: [NMeTyr^B26]-insulin, and Yellow: [NMeAla^B26]-Insulin. (B) A close-up view of the B26 turn in the most representative analogues: Green: wild-type, Cyan: [NMeAla^B26]-DTI-NH_2, Yellow: [NMeAla^B26]-insulin, and Coral [NMeTyr^B26]-insulin. Oxygen atoms in red and nitrogen in blue. W – water molecule, hydrogen bonds (dashed lines) lengths in Å, and Me – methylation.
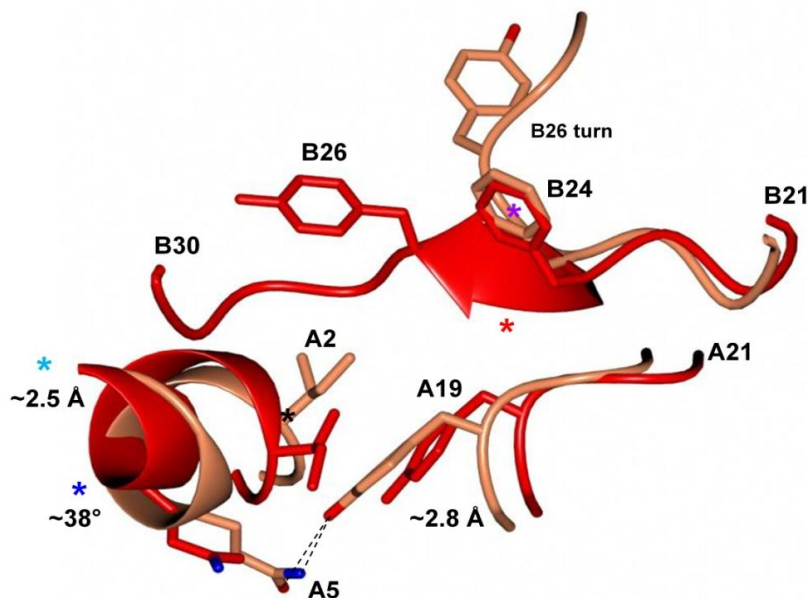


**Figure 26. The main structural changes in insulin upon activation.**

Human insulin (red[151]) and [NMeTyr^B26]-insulin (coral), which is isomorphous to the highly-active [NMeAla^B26]-DTI-NH_2.  Magenta Star-Phe^B24 side-chain structural pivot point, Red Star – Phe^B25 main chain structural pivot, Black Star – Ile^42 Cα main chain pivot point, Green Star – Tyr ^A19 shift and formation of a hydrogen bond (dashed lines) with Gln^A5, Blue Star – rotation of A1 helix, Cyan Star – translation of the A1 helix[75]. Oxygen atoms in red and nitrogen in blue.

The analogue $[D-Pro^{B26}]$-DTI-NH$_2$ is also a high affinity (359%) insulin analogue that possesses the B26 turn (Figure 27). The D-Pro mimics the *cis* Phe$^{B25}$-Xaa$^{B26}$ *N*-methylated peptide bond in the B26 turn. The i+3 β-turn hydrogen bond is not formed here (CO$^i$-NH$^{i+3}$ distance ~4.00 Å), although the D-Pro causes the B-chain C-terminal CONH$_2$ to adopt a conformation and position similar to the analogues with B26 turn. The C-terminal NH$_2$ group is displaced ~1.1 Å from hydrogen-bonded equivalents in other highly active analogues.

The structures presented here suggest the conservation of specific structural features that may be present in WT insulin upon binding to the insulin receptor. The high affinity insulin analogues may convey an active structural conformation that allows the buried biologically important residues to be exposed, allowing insulin:IR binding. Although the structures presented here are not WT insulin, the information extracted from these structures gives indirect information about the active motif of insulin.

The formation of the i+3 hydrogen bond in the high affinity analogues is one of the most important events to occur during activation of the insulin molecule (Figure 25). This formation causes other hydrogen bonds to be created (Figure 25), which results in a strong hydrogen bonding network, further stabilising the B26 turn.

The most striking piece of information from the structures is the B26 turn, which is caused by the *trans*-to-*cis* isomerisation of the Phe$^{B25}$-Tyr$^{B26}$ peptide bond. The presence of the strong 2.8 Å hydrogen bond between Phe$^{B24}$(CO)-Thr$^{B27}$(NH) indicates that the formation of the B26, or a turn similar to that seen here, is more than likely the preferred direction for the movement of the C-terminus of the B-chain following detachment of the Tyr$^{B26}$-Thr$^{B30}$ region. Chemical crosslinking of Gly$^{A1}$ and Lys$^{B29}$ produces a crystal structure and displays assembly properties very similar to that of WT insulin, despite these similarities the cross linked insulin was biologically inactive[3]. This implies that upon binding of insulin to the IR, the C-terminus of the B-chain must move away, exposing the hydrophobic core[3,6]. This research suggests that it is the requirement for the C-terminus of the B-chain to be rearranged, as per the B26 turn, for IR binding to occur.
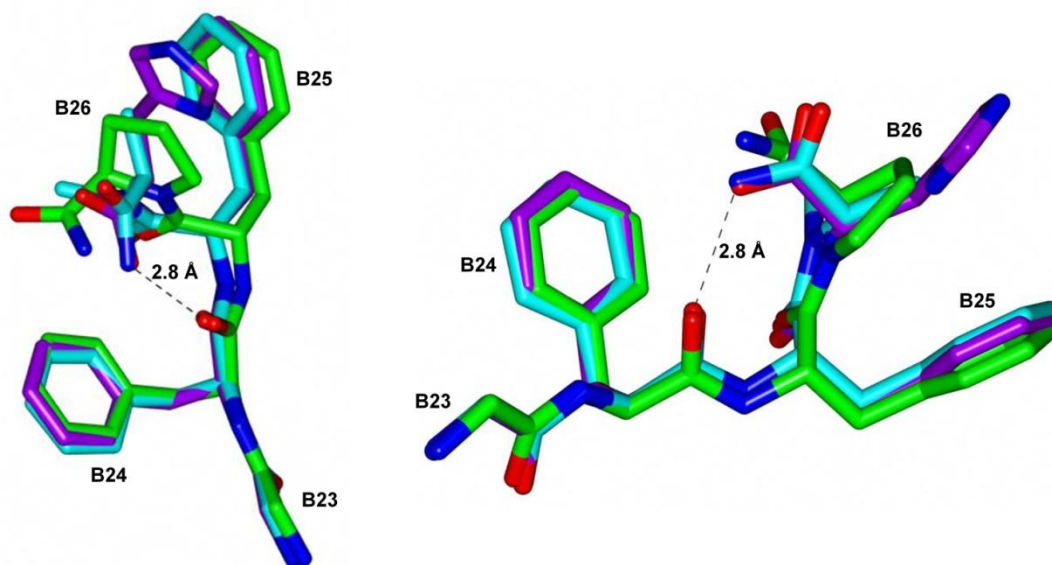
*Figure 27. The B26 turn and B26-like turn in truncated analogues.*

Super positioning of the $B^{23}$-$B^{26}$ region in [NMeAla$^{B26}$]-DTI-NH$_2$ (carbons in cyan), [D-Pro$^{B26}$]-DTI-NH$_2$ (carbons in green) and [NMeHis$^{B26}$]-DTI-NH$_2$ (carbons in purple). The dashed line indicates the COB$^{24}$-NHB$^{26}$ hydrogen bond in [NMeAla$^{B26}$]-DTI-NH$_2$ which is characteristic of the B26 turn. Oxygen atoms in red and nitrogen in blue.

## 2.4. The Appearance of the B26 Turn in Full Length Insulin Analogues

The two insulin analogues described previously are truncated in the B-chain after the $B^{26}$ position, whilst the capabilities of full length analogues have not been investigated. Both [$N$MeTyr$^{B26}$]-insulin and [$N$MeAla$^{B26}$]-insulin have a binding affinity of 21%, despite the possession of the B26 turn and the structural characteristics seen in the truncated analogues (Figure 28). The B26 turn in the full length analogues contains the same í and í+3 hydrogen bond between the CO in Phe$^{B24}$ and NH in Thr$^{B27}$ with a bonding distance (~3.0-3.3 Å [$N$MeTyr$^{B26}$]-insulin). In the full length analogues the Thr$^{B27}$ peptide bond NH groups are in an equivalent spatial position to that of the $NH_2$ group in the $CONH_2$ terminus, making hydrogen bond formation possible. In a typical type II β-turn, the $C^{\alpha i}$-$C^{\alpha i+3}$ distance is ~7 Å, in the full-length analogues the $C^{\alpha i}$-$C^{\alpha i+3}$ distance is ~6 Å, which makes hydrogen bonds more likely to form. The C-terminal region of the B-chain is well defined (Lys$^{B29}$-Thr$^{B30}$ are disordered), following the direction forced by the B26 turn, allowing the full exposure of binding site one in insulin (Figure 28). The B26 turn is further stabilised in this conformation by the formation of additional hydrogen bonds between Phe$^{B24}$ CO with the –OH in the side chain of Thr$^{B27}$, and water molecules. [$N$MeTyr$^{B26}$]-insulin was also crystallised in hexameric form, showing that despite the presence of the $N$Me group on residue B$^{26}$, it is still capable of forming a hexamer. The full length analogue [$N$MeAla$^{B26}$]-insulin was unable to be crystallised as a hexamer, only as a monomer, indicating that the change of the amino acid removes an important interaction in the dimer interface.
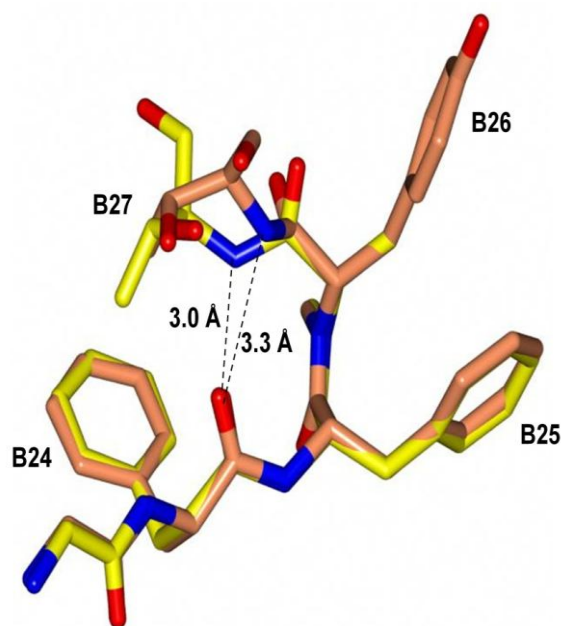
***Figure 28. The B26 turn in full length insulin analogues.***

A view of the B[24]-B[27] region of two insulin analogues forming the B26 turn. [*N*MeTyr[B26]]-insulin (C-atoms in coral) and [*N*MeAla[B26]]-insulin form the B26 turn with an í+3 hydrogen bond (3.0Å and 3.3Å), which is comparable to the 2.8Å seen in the highly active analogue [*N*MeAla[B26]]-DTI-NH$_2$. Oxygen atoms in red and nitrogen in blue.

## 2.5. The Structural Signatures of Active Insulin Analogues

Both truncated and full length analogues, showing high affinity, possess a novel conformation in the N-terminus of the B-chain. Working from C to N-terminus, the main chain follows the path taken by the T-state of insulin, until the $C\alpha$ in $His^{B5}$, at which point it deviates. A pseudo type II $\beta$-turn (Figure 29) is formed, which is stabilised as in the C-terminus, by a $i+3$ hydrogen bond (3.0 Å) between the NH in the main chain of $His^{B5}$ ($i+3$) and the side chain of $Asn^{B3}$. Further stabilisation is achieved by a hydrogen bond (3.0 Å) with the $N\delta2$ in $Asn^{B3}$. The importance of the $B^5$ side chain in this unique formation is further reinforced by the close proximity (~3.3 Å) of the $N\delta1$ of $His^{B5}$ to the main chain C in $Thr^{A8}$. This deviates from the T-state B-chain N-terminus causing the network of interactions between A- and B-chain to become almost entirely destroyed, leaving only the $NHB^6 - COA^6$ bond (2.8 Å).
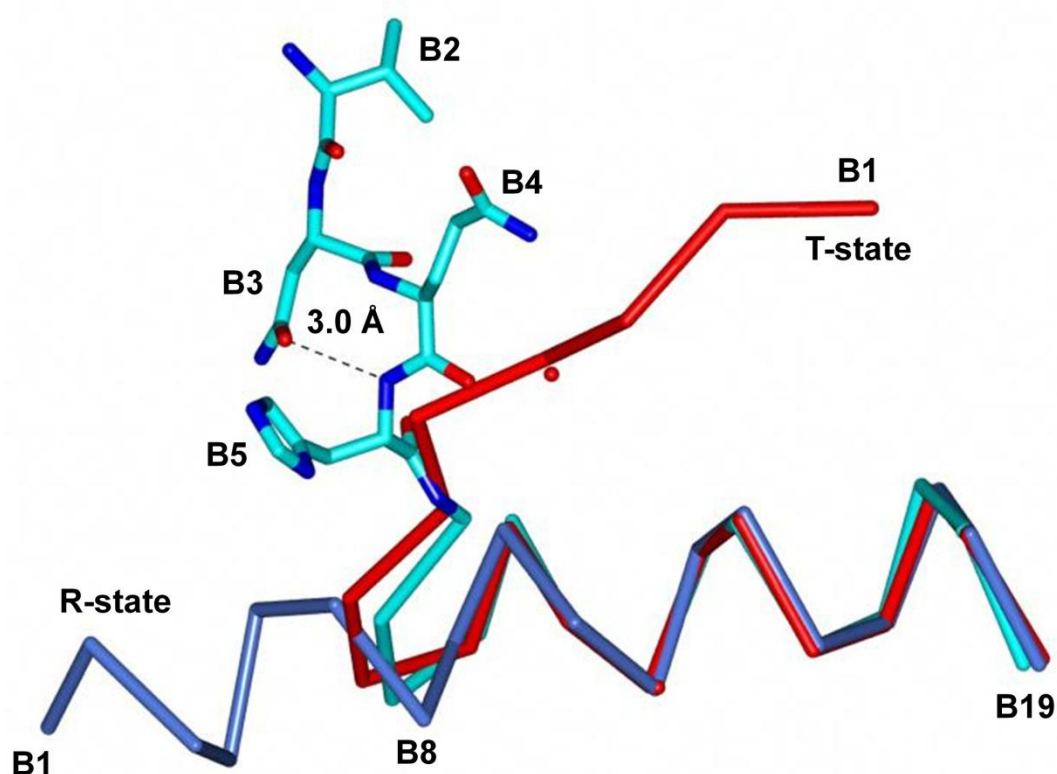


***Figure 29. The unique conformation of high affinity insulin analogues at the B-chain N-terminus.***

A representation of the pseudo $\beta$-turn at the B-chain N-terminus from [$N$MeAla$^{B26}$]-DTI-NH$_2$ (C-atoms shown in cyan). All atoms are shown for $B^2$-$B^5$, and a $C\alpha$ trace for $B^6$-$B^{19}$. The $C\alpha$ trace of insulin in the T-(red trace) and R-states (blue trace) are derived from 1mso.pdb and 1znj.pdb respectively. Oxygen atoms in red and nitrogen in blue.

The three-dimensional structural traits presented here show the conservation and convergence of key structural elements across multiple analogues. The elements discussed here may be those present when insulin becomes active upon IR binding, or during an activation process. The structures present a conformation of what I believe to be the active/activated insulin conformation, although the evidence is indirect. The B26 turn is associated with the *trans*-to-*cis* isomerisation of the Phe[B25]-Tyr[B26] peptide bond, which results in the most striking structural feature of the active/activated form of insulin. Another critical structural feature of the B26 turn is the formation of a hydrogen bond between Phe[B24] (CO) and Thr[B27] (NH). This hydrogen bond is likely to be critical to the stability of the turn, as when the $B^{24}$ CO group is reduced to $CH_2$, the insulin molecule loses its activity[152]. NMR shows that the potent [D-Ala[B26]]-DTI-$NH_2$ analogue[153] has mobility in the $B^{26}$ region, indicating that the formation of the B26 turn can easily be achieved.

These structures and observations led to the proposal that for binding of insulin to the IR to occur, the C-terminus of the B-chain (Tyr[B26]-Thr[B30]) must be detached allowing the B26 turn to be formed along with the *trans*-to-*cis* isomerisation of the Phe[B25]-Tyr[B26] peptide bond. This theory is consistent with the induced-fit theory and peptide bond isomerisation[61,154]. Detachment of the C-terminus exposes the hydrophobic core of insulin, so the protein environment supplied by the IR must be involved in the stabilisation of the insulin molecule. Due to this detachment and isomerisation, structures of high affinity analogues are more representative of the conformation that insulin adopts upon binding to the IR, than of the native hormone in its storage form. If these structural changes and re-arrangements are correct, this indicates that the IR acts as a chaperone in the activation of insulin[61], especially for the *trans*-to-*cis* isomerisation.

A striking structural feature is the invariance seen in the Cα in all of these structures. The re-arrangement of the B-chain and the concerted changes seen in the N-terminal A-chain helix cause a large hydrophobic area (consisting of Gly[A1], Ile[A2], Val[A3] and Tyr[A19]) to be exposed for direct interactions with the IR (Figure 26). In receptor binding, the importance of Ile[A2] is illustrated by it being completely exposed on the large hydrophobic area. The structural conservation of the Cα in Ile[A2] suggests that it may act as a pivot point for the A-chain during activation. The A-chain IR-binding surface is further stabilised by the movement of Tyr[A19] to form a hydrogen bond with Gln[A5]. The movement of Tyr[A10] causes a transformation of the Gly[A1]-Ile[A2]-Val[A3] hydrophobic surface into a binding epitope.

It has been shown that Phe[B24] and Phe[B25] play key structural roles in insulin activity[73], which is confirmed by the structures of the high affinity insulin analogues presented here. In WT insulin and all other insulin analogues the aromatic ring of Phe[B24] is spatially invariant, which is in contrast to the high affinity analogues that possess a ~90° rotation of Phe[B25]. This causes an overhang of ~7 Å of the A-chain over the binding epitope. Although there is

considerable movement of the side chain, the Phe$^{B25}$ Cα position is well conserved in a variety of analogues. This observation further reinforces the need of Phe$^{B24}$ to assist the detachment of the C-terminus of the B-chain in a stereo-specific manner[77]. The importance and role played by the Phe$^{B24}$ side chain leads to it being described as an important pivot point during the activation of the insulin molecule (Figure 26). The ability of Phe$^{B25}$ to be the initiation point for the formation of the B26 turn relies upon the structural stability of the Phe$^{B24}$ side chain. The Phe$^{B24}$ side chain is anchored by a hydrophobic cavity, lined by residues Leu$^{B15}$, Val$^{B12}$, Tyr$^{B16}$ (whose side chains seal off the cavity by collapsing into it during activation) and Cys$^{B19}$. This is then further complemented by the spatial stability of Phe$^{B25}$ Cα, which allows this and Phe$^{B24}$ to act as an anchor, enabling the pivot to cause the formation of the B26 turn, and therefore IR binding.

Along with the structural convergence seen in the B$^{24}$-B$^{26}$ region of the highly-active analogues, there is also a consistent presence of a unique B$^{3}$-B$^{5}$ pseudo β-turn in the high affinity analogues. This causes the N-terminus of the B-chain to be in a conformational state between that of the T- and R-states.

The high affinity analogues described here present an interesting insight into structure-function relationship, not only of the whole active conformation, but also of the local structures that can be involved in binding of insulin to the IR. When insulin becomes activated, the B26 turn, along with Phe$^{B25}$ causes a hydrophobic knob to form along with the A-chain epitope. This forms a highly hydrophobic IR-binding cleft lined with Leu$^{B15}$, Leu$^{B11}$ and Val$^{B12}$ side chains. The shape of the cleft is reminiscent of a hydrophobic hand that ensures a good grip on the receptor, the thumb (Phe$^{B25}$ and the B26 turn), a hydrophobic palm (Leu$^{B15}$, Leu$^{B11}$, Val$^{B12}$) and hydrophobic fingers (Tyr$^{A19}$, Gly$^{A1}$-Ile$^{A2}$ and Val$^{A3}$).

The high affinity insulin analogues presented here show unique molecular features and structural elements in the conformational landscape of the insulin molecule. These give appreciable insight into the activation and the active form of insulin.

## 2.6. Application of Click Chemistry in the Development of Novel Insulin Analogues

A structural re-arrangement of the C-terminus of the B-chain[61,75,77] exposes the binding epitope of insulin, allowing effective binding of the IR. Cyclo[G-PrgFYTPK(N$_3$)T]-insulin was designed to mimic the conformation seen in high affinity analogues such as [NMeTyr[B26]]-insulin, by fixing the B26 turn-like conformation by chemically cross-linking amino acid from the B20-B30 region[75].

The Cyclo[G-PrgFYTPK(N$_3$)T]-insulin was crystallised as a tetramer of two dimers, which are not related along the normal dimer interface.. The low binding affinity (0.13%) of Cyclo[G-PrgFYTPK(N$_3$)T]-insulin can be explained by the structure at the B-chain C-terminus of the analogue. The B26 turn is not present in Cyclo[G-PrgFYTPK(N$_3$)T]-insulin (Figure 30). The chemical cross-link causes conformational and structural rigidity, which results in the conversion of the B$^{25}$-B$^{26}$ peptide bond from *trans*-to-*cis* being impossible.
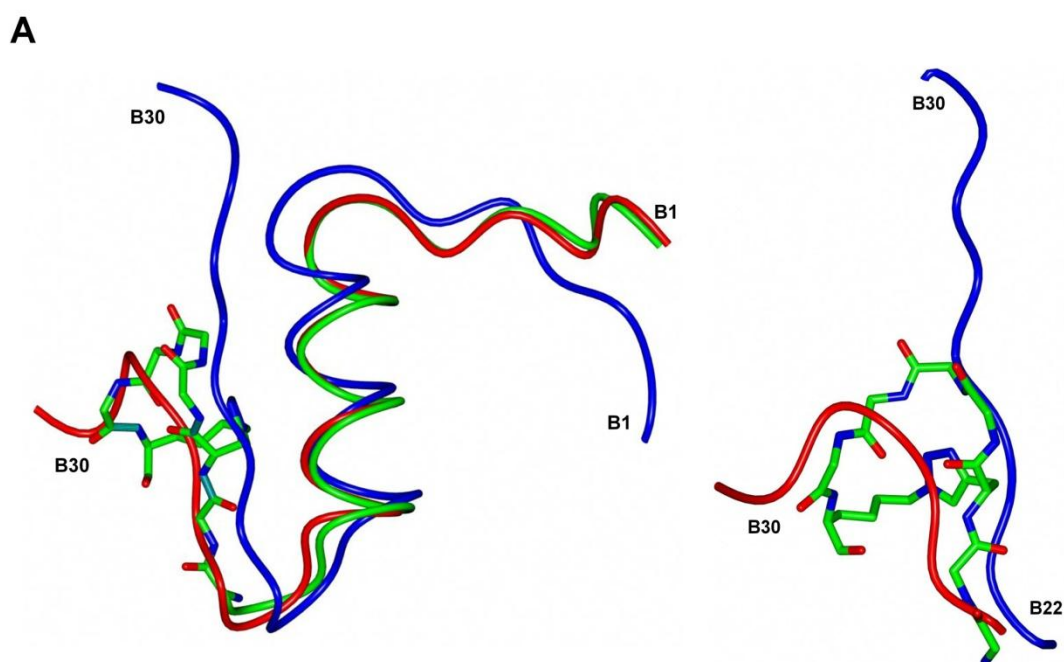


***Figure 30. An overview of the structure of Cyclo[G-PrgFYTPK(N₃)T]-insulin.***

Left: An overview of Cyclo[G-PrgFYTPK(N$_3$)T]-insulin, against T-state insulin (shown in blue, PDB used was 1mso.pdb) and [NMeTyr[B26]]-insulin (shown in red). Right: The C-terminus of the B-chain Cyclo[G-PrgFYTPK(N$_3$)T]-insulin, against T-state insulin (shown in blue, PDB used was 1mso.pdb) and [NMeTyr[B26]]-insulin (shown in red). Oxygen atoms in red and nitrogen in blue.

The conformation adopted by $B^{25}$ and $B^{26}$ protrudes towards the insulin binding epitope. This structure results in the limited exposure of the binding epitope to the IR, which is reflected in the reduced binding affinity.

It is possible that tethering residues in the $B^{24}$-$B^{30}$ region may cause conformational rigidity in insulin, which reduces its ability to bind to the IR. Photo cross-linking studies have found that the active form of the insulin B-chain C-terminus may be an extended β-strand which inserts itself between different subunits of the IR[77]. Conformational locking of this section of insulin eliminates the possibility of such a transition occuring.

Cyclo[GFFY-Pent($N_3$)-P-G-Prg]-insulin was crystallised as a hexamer, similar to the R-state of WT insulin (Figure 31), which explains the almost native binding affinity, 101.5%. The chemical cross-link at the B-chain C-terminus causes a small nodule to be formed. This structure was stabilised by a weak hydrogen bond between $COB^{27}$ and $NHB^{30}$.
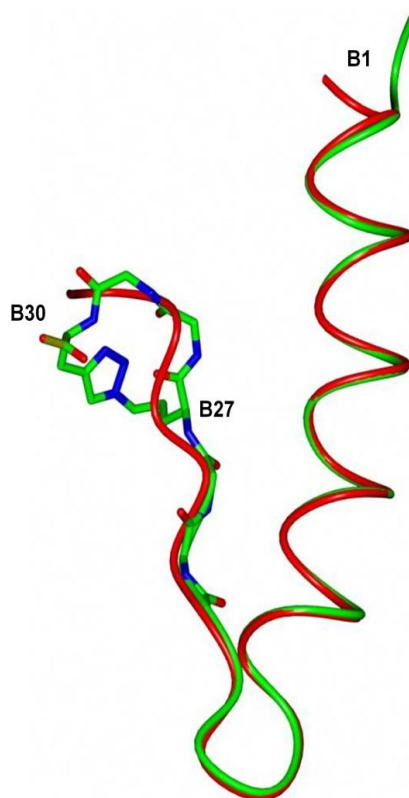


***Figure 31. The cross-link of Cyclo[GFFY-Pent($N_3$)-P-G-Prg]-insulin.***
Cyclo[GFFY-Pent($N_3$)-P-G-Prg]-insulin is shown in green and R-state insulin is shown in red. PDB used was 1znj.pdb. Oxygen atoms in red and nitrogen in blue.

The structural rigidity induced by the cross-link in Cyclo[G-PrgFYTPK($N_3$)T]-insulin caused a vastly reduced binding affinity (0.13%). The same structural rigidity caused by the chemical cross-link in Cyclo[GFFY-Pent($N_3$)-P-G-Prg]-insulin produced a slightly raised binding affinity (101.5%), as it does not interfere with the structural re-arrangement of the C-terminus of the B-chain required for binding to the IR.

## 2.7. Conclusion and Future Use of Click-Chemistry in the Development of Insulin Analogues

The aim behind creating chemically cross-linked insulin analogues was to artificially maintain the structural conformation seen in the high affinity insulin analogues such as [*N*MeTyr$^{B26}$]-insulin.  The conformational rigidity seen in Cyclo[G-PrgFYTPK(N$_3$)T]-insulin is a result of the chemical cross-link between residues B$^{24}$ and B$^{29}$, preventing the formation of the B26 turn.  This causes the binding epitope to be partially masked, impairing IR binding. These two factors combined result in the low binding affinity.  These structural features act upon binding affinity along with the chemical effect caused by the positive charge of the triazole ring, which is in contrast to the hydrophobic environment present in WT insulin. The 1,2,3-triazole ring at the centre of the chemical cross-link in Cyclo[G-PrgFYTPK(N$_3$)T]-insulin introduces a positive charge, where in WT insulin Phe$^{B24}$ has a neutral charge and forms part of the binding epitope[61].  It is believed that B$^{24}$ interacts with the leucine-rich repeat of the IR[61], and is part of a hydrophobic surface.  The chemical difference observed by the IR upon binding may be the reason that the binding affinity is reduced.

Cyclo[GFFY-Pent(N$_3$)-P-G-Prg]-insulin  shows very high structural symmetry with the R-state seen in WT insulin. The cross-linked C-terminus of the B-chain causes a nodule to form that is further stabilised by a weak hydrogen bond. Due to the location of this nodule, away from the B$^{26}$ position, the B26 turn can still form and allow binding to the IR[75].

The chemical cross-link for Cyclo[G-PrgFYTPK(N$_3$)T]-insulin created negative effects such as structural rigidity and a partial obscuring of the binding epitope. The location of the cross-link in Cyclo[GFFY-Pent(N$_3$)-P-G-Prg]-insulin allows the B26 turn[75] to form upon binding with the IR, meaning that the binding epitope would not be obscured. Structural rigidity in the nodule caused by the cross-link reduces the overal flexibility in the C-terminus of the B-chain. The nodule acts as a handle for the IR to "grab onto" during binding to negate any negative effect and effectively restore binding affinity.

The ability to fix the B-chain C-terminus into a conformation resembling the B26 turn would require further design to reduce the negative effects seen in Cyclo[G-PrgFYTPK(N$_3$)T]-insulin and to a lesser effect Cyclo[GFFY-Pent(N$_3$)-P-G-Prg]-insulin. Structural flexibility in the C-terminal region is important for the re-arrangement of the region upon binding to the IR, therefore inserting a cross-link would require the formation of a structured turn similar to the B26 turn, whilst still maintaining structural flexibility in this region.

# Chapter 3.  The Role of Residues B24 and B26 in Insulin Dimer Formation and Hormone Activity

## 3.1. Summary

Insulin residues $B^{24}$-$B^{26}$ of the B-chain C-terminus play key roles in the interactions with the IR, as well as in insulin self-association into dimers.

The role of Phe$^{B24}$, Phe$^{B25}$ and Tyr$^{B26}$ in dimer formation have been studied previously[155], although the individual contribution of each amide hydrogen has not been fully established.  The work described in this chapter addresses two outstanding issues concerning the role of B24 and B26 sites: (i) the role of $B^{24}$/$B^{26}$ main chains in the stability of the dimer interface, and (ii) the role of the Phe$^{B24}$ Tyr$^{B26}$ side chains in insulin activity.

The C-terminus of the B-chain of insulin, along with its role in IR binding[156,157] is also responsible for the formation and stabilisation of insulin dimers. The dimerisation of insulin is biologically important as it is vital step in the storage and bloodstream delivery of the hormone[24,73].  Dimerisation occurs by the formation of an anti-parallel β-sheet between the B-chain C-terminal segments of two insulin monomers[2,80,151]. These secondary structures are stabilised four reciprocal but complementary hydrogen bonds that are formed between the carbonyl oxygen of Tyr$^{B26}$ (Phe$^{B24}$) of the first molecule, to the amide hydrogen of Phe$^{B24}$ (Tyr$^{B26}$) in the opposite molecule (Figure 32).  The amide hydrogen of Phe$^{B25}$ forms a hydrogen bond with the carbonyl Tyr$^{A19}$. Both molecules are linked by two-fold symmetry.
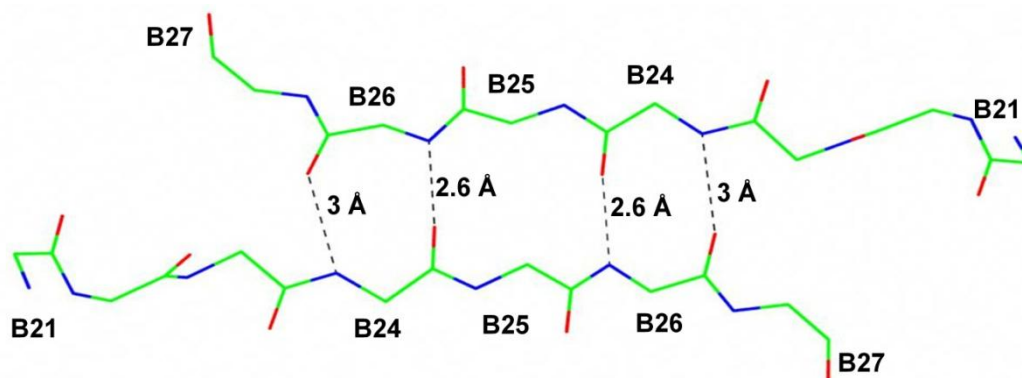


***Figure 32. The hydrogen bonding network at the insulin dimer interface.***
Carbon atoms in green, oxygen - red and nitrogen - blue.

The second part of this chapter focuses on the two amino acids Phe$^{B24}$ and Tyr$^{B26}$ and the role of these side chains and their conformations.  The aromatic residues Phe$^{B24}$ and Tyr$^{B26}$ are invariant and believed to be involved in the first contact upon binding between insulin and the IR, with each having its own contribution in insulin:IR interactions[100,72].  The

replacement of Phe[B24] or Tyr[B26] produces analogues with a wide range of activities. When Phe[B24] is replaced by an L-amino acid (except Gly), the analogues possess low binding affinities and biological activity[100]. Whilst the incorporation of any D-amino acid (side chain independent), yields enhanced analogues with increased binding affinity in most cases[150,153], except D-Pro. The replacement of Tyr[B26] displays less extreme effects on their binding affinity, with Trp[B26]-insulin[72] displaying 18%, and Glu[B26]-insulin showing 125%[106] (compared to WT insulin).

## 3.2. Crystallisation, Structure Determination and Refinement of Insulin Analogues to Probe the Role of Residues B24 and B26

The insulin analogues crystallised in this chapter are displayed in Table 7 with their sequence change from WT insulin illustrated along with their binding affinity.

*Table 7. A comparison of insulin analogues structure used in this chapter and their relative receptor binding affinity.*

Changes in the B-chain sequence are indicated in bold. [$N$MeXaaB$^n$]: methylation of the $N$-peptide atom preceding Xaa B$^n$ amino acid; -NH$_2$ : C-terminal carboxy amide.

| B-Chain Position | 24 | 25 | 26 | 27 | 28 | 29 | 30 | Affinity |
|---|---|---|---|---|---|---|---|---|
| **WT:** | F | F | Y | T | P | K | T | 100 |
| [$N$MePhe$^{B25}$]-insulin | F | **Me** F | Y | T | P | K | T | 0.2 |
| [$N$MeTyr$^{B26}$]-insulin | F | F | **Me** Y | T | P | K | T | 21 |
| [Tyr$^{B26}$,$N$MePhe$^{B26}$,Lys$^{B28}$,Pro$^{B29}$]-insulin | F | **Y** | **Me F** | T | **K** | **P** | T | 4 |
| [Phe$^{B26}$]-insulin | F | F | **F** | T | P | K | T | 46 |
| [His$^{B24}$]-insulin | **H** | F | Y | T | P | K | T | 1.5 |
| [Asn$^{B26}$]-insulin | F | F | **N** | T | P | K | T | 83 |

Insulin analogue [$N$MePhe$^{B26}$]-insulin was dissolved in 20 mM HCl to a concentration of 10 mg/ml. The best crystal identified from the screening was obtained from drops equilibrated 1:3 (protein:well) in well solution consisting of 0.6 M Na$_2$SO$_4$, 0.3 M Tris pH 8.2, 0.6mM ZnAc, 0.06% phenol (0.5 ml well volume), and grown at 21°C. The crystal was flash-cooled in liquid nitrogen without any cryoprotection. The X-ray data were collected in house on a Rigaku RUH3R with a MAR345 detector, and processed using HKL2000[140]. The crystal structure of [$N$MePhe$^{B26}$]-insulin was solved by molecular replacement using default settings in Molrep[143]. The search model consisted of a truncated model of hexameric insulin 1znj.pdb, which involved the removal of residues B$^1$-B$^5$ and B$^{25}$-B$^{30}$. Subsequent manual model building and refinement with Coot[138] and 10 cycles of REFMAC5[145], resulted in a final model which was validated by MolProbity[146]. Data-collection, reduction and refinement statistics are detailed Table 8. Crystal conditions are documented in Table 9.

The other insulin analogues that were crystallised as hexamers had their structure determined and refined as per [$N$MePhe$^{B26}$]-insulin. For crystals of monomeric insulins the molecular replacement was conducted using a truncated (removal of residues B$^1$-B$^5$ and B$^{25}$-B$^{30}$) model of monomeric insulin 1mso.pdb[151]. The differences in the crystal conditions and

data collection, reduction and refinement conditions are details in Table 9 and Table 8 respectively.

***Table 8. X-ray date and refinement statistics for the insulin analogues used in this chapter.***

| | [NMeTyr$^{B26}$]-insulin hexamer | [Tyr$^{B25}$,NMePhe$^{B26}$,Lys$^{B28}$,Pro$^{B29}$]-insulin | [NMePhe$^{B25}$]-insulin |
|---|---|---|---|
| PDB code | 2ws6 | 3zs2 | 3zqr |
| Data collection Beamline | ESRF, ID14-1 | ESRF, ID23-1, ADSC Q315 detector | In-house MAR345, Rigaku RUH3R |
| Wavelength (Å) | 0.9340 | 1.0712 | 1.54178 |
| Space group | P2$_1$ | P2$_1$ | P2$_1$ |
| Cell dimensions | | | |
| a, b, c (Å) | 58.7 61.9 47.3 | 46.7 62.2 57.6 | 46.6 61.8 58.2 |
| α, β, γ (°) | 90.0 111.9 90.0 | 90.0 111.3 90.0 | 90.0 111.2 90.0 |
| Resolution (Å) | 40.0–1.50(1.55–1.50) | 30.0-1.97 (2.04-1.97) | 20.0-1.9 (1.93-1.90) |
| R$_{sym}$ | 0.048(0.479) | 0.109 (0.205) | 0.047 (0.164) |
| <I / σI> | 11.7(5.0) | 13.5 (8.5) | 16.3 (10.0) |
| Completeness (%) | 99.8(99.9) | 90.2 (61.7) | 95.3 (74.7) |
| Redundancy | 3.9(3.0) | 2.3 (1.5) | 3.2 (2.0) |
| Refinement | | | |
| Resolution (Å) | 40.0 - 1.50 | 50.0 – 2.60 | 20.0 – 19.0 |
| No. reflections | 47779 | 18615 | 22062 |
| R$_{work}$ / R$_{free}$ | 0.150/0.199 | 0.188/0.240 | 0.223/0.281 |
| No. atoms | 2775 | 2440 | 2460 |
| Protein | 2523 | 2244 | 2250 |
| Ligand/ion | 6 | 46 | 46 |
| Water | 246 | 150 | 164 |
| B-factors | | | |
| Protein | 22.3 | 36.4 | 38.17 |
| Ligand/ion | 31.9 | 32.1 | 29.64 |
| Water | 30.2 | 46.6 | 39.10 |
| R.m.s. deviations | | | |
| Bond lengths (Å) | 0.025 | 0.019 | 0.022 |
| Bond angles (°) | 2.173 | 1.72 | 1.838 |
| Ramachandran statistics (%) | | | |
| Preferred/Allowed/Outliers | 98.6/1.4/0 | 99.2/0.8/0 | 95.6/4.4/0 |

| | [His$^{B24}$]-insulin | [Asn$^{B26}$]-insulin | [Phe$^{B26}$]-insulin monomer |
|---|---|---|---|
| PDB code | | | |
| Data collection Beamline | Diamond i24 | Diamond, i02 | ESRF id232 |
| Wavelength (Å) | 0.9686 | 0.9795 | 0.8726 |
| Space group | I422 | P4$_1$2$_1$2 | P2$_1$2$_1$2$_1$ |
| Cell dimensions | | | |
| $a$, $b$, $c$ (Å) | 57.5 57.5 56.2 | 45.6 45.6 90.0 | 43.9 45.8 51.3 |
| α, β, γ (°) | 90.0 90.0 90.0 | 90.0 90.0 90.0 | 90.0 90.0 90.0 |
| Resolution (Å) | 40.62-1.70 (1.76-1.70) | 45.63-1.81 (1.85-1.81) | 50.00-1.60 (1.63-1.60) |
| $R_{sym}$ | 0.157 (0.559) | 0.11 (0.817) | 0.092 (0.467) |
| $<I / \sigma I>$ | 10.2 (4.1) | 19.1 (4.2) | 5.2 (2.9) |
| Completeness (%) | 99.6 (95.0) | 99.0 (99.6) | 100 (100) |
| Redundancy | 6.5 (6.0) | 14.2 (15.6) | 7.1 (7.2) |
| Refinement | | | |
| Resolution (Å) | 40.62-1.70 | 36.06-1.81 | 34.15-1.59 |
| No. reflections | 5182 | 11459 | 13606 |
| $R_{work}$ / $R_{free}$ | 0.179/0.217 | 0.185/0.238 | 0.163/0.202 |
| No. atoms | 340 | 936 | 998 |
| Protein | 300 | 813 | 802 |
| Ligand/ion | 5 | 5 | 15 |
| Water | 35 | 118 | 181 |
| $B$-factors | | | |
| Protein | 32.562 | 19.241 | 10.556 |
| Ligand/ion | 41.190 | 12.582 | 37.221 |
| Water | 50.511 | 28.677 | 23.131 |
| R.m.s. deviations | | | |
| Bond lengths (Å) | 0.025 | 0.028 | 0.024 |
| Bond angles (°) | 2.543 | 2.081 | 2.388 |
| Ramachandran statistics (%) | | | |
| Preferred/Allowed/Outliers | 94.4/5.6/0 | 98.6/1.4/0 | 98.9/1.1/0 |

***Table 9. Crystal conditions for the insulin analogues used in this chapter.***

|  | [*N*MeTyr$^{B26}$]- insulin hexamer | [Tyr$^{B25}$,*N*MePhe$^{B26}$,Lys$^{B28}$,Pro$^{B29}$]- insulin | [*N*MePhe$^{B25}$]- insulin |
| --- | --- | --- | --- |
| Well Condition | 0.3 M Tris pH 8.2, 0.1 M NaCit, 0.02% ZnAc, 0.06% phenol | 0.1M NaCit, 0.3M Tris pH 8.2, 0.6mM ZnAc, 0.06% phenol | 0.6M Na$_2$SO$_4$, 0.3M Tris pH 8.2, 0.6mM ZnAc, 0.06% phenol |
| Drop Ratio (Protein:Well) | 1:3 | 1:2 | 1:2 |
| Temperature | 21°C | 21°C | 21°C |
| Cryoprotectant | None | None | None |

|  | [His$^{B24}$]-insulin | [Asn$^{B26}$]-insulin | [Phe$^{B26}$]-insulin monomer |
| --- | --- | --- | --- |
| Well Condition | 2M NaCl, 3M Urea, 0.1M phosphate buffer pH 6.5, 6mM ZnAc, 0.5% resorcinol | 0.15M (NH$_4^+$)$_2$SO$_4$ pH 3.0 | 0.15M Na$_2$SO$_4$ pH 3.0 |
| Drop Ratio (Protein:Well) | 1:3 | 1:3 | 1:3 |
| Temperature | 21°C | 21°C | 21°C |
| Cryoprotectant | None | None | None |

## 3.3. The Physical-Chemical Properties of Insulin in Various Association States

The role of the amino acids at positions $B^{24}$, $B^{25}$ and $B^{26}$ in dimer formation was elucidated here by the characterisation of several novel full-length insulin analogues that each has *N*-methylated peptide bonds at these positions. The dimerisation capability of each insulin analogue was determined was determined in collaborators' laboratory in Institute of Organic Chemistry and Biochemistry (IOCB), in order to identify contribution of amide nitrogen of each of these amino acid towards insulin association. WT insulin (capable of forming dimers) and des($B^{23}$-$B^{30}$) octapeptide (DOI, incapable of dimer formation[158]) were used as references for the dimeric and monomeric forms of insulin respectively.

The ability of an insulin analogue to dimerise was measured by ITC dilution measurements that detect heat energy changes upon the dissociation of insulin dimers (Table 10)[159]. Human/porcine insulin and DOI were used as references for normal and non-dimerisable insulin respectively.

***Table 10. ITC analysis of dimerisation capabilities of insulin and insulin analogues.***

The number of repetitions of a measurement is indicated in brackets. The experimental values are taken from Antolikova *et al.*[159].

| Insulin | $K_d$ (μM) |
|---|---|
| Porcine insulin (n = 2) | $9.03 \pm 0.50$ |
| Human insulin (n = 8) | $8.81 \pm 1.05$ |
| DOI (n = 2) | No heat effect |
| [*N*MePhe$^{B24}$]-insulin (n = 2) | $587 \pm 99$ |
| [*N*MePhe$^{B25}$]-insulin (n = 2) | No heat effect |
| [*N*MeTyr$^{B26}$]-insulin (n = 2) | $142 \pm 30$ |
| [*N*MePhe$^{B26}$]-insulin (n = 2) | $1240 \pm 10$ |

ITC experiments show that the $K_d$ of human and porcine insulin (Table 10) are in close agreement, with previously reported values (8.8 μM and 9 μM respectively)[160]. Strazza *et al.*[161] used concentration spectroscopy to determine the $K_d$ of porcine insulin at pH 2 ($K_d$ = 15μM) and pH 7 ($K_d$ = 18 μM). This reliability therefore allows a systematic look at the individual hydrogen bonds role in the $B^{24}$-$B^{26}$ dimer stability, evaluating their roles. No heat effect could be detected for DOI, confirming that insulin dimerisation is caused by interactions at the B-chain C-terminus of insulin. The analogue [*N*MeTyr$^{B26}$]-

insulin has a $K_d$ of 142 μM, which is 15-fold weaker than that of insulin. The removal of the phenolic hydroxyl ([$N$MePhe$^{B26}$]-insulin) further suppresses the dimerisation ability to $K_d$ = 1.24 mM. [$N$MePhe$^{B24}$]-insulin caused a 65-fold higher $K_d$ compared to insulin of 587 μM, but the greatest observed effect was for [$N$MePhe$^{B25}$]-insulin, which gave no heat effect, like DOI. pH has little effect upon insulin dimerisation experiments[161], so all ITC experiments were performed at a pH of 2.5 to avoid reduced solubility of insulin at a more physiological pH[160].

## 3.4. The Structural Properties of Insulin in Various Association States

The crystal structure of [$N$MePhe$^{B25}$]-insulin was obtained in the hexameric form (R$_6$). [$N$MeTyr$^{B26}$]-insulin was obtained in both monomeric and hexameric (R$_6$) forms, whilst [Tyr$^{B25}$,$N$MePhe$^{B26}$]-insulin was in hexameric (R$_6$) form[75,162].

[$N$MePhe$^{B25}$]-insulin adopted the R$_6$ hexamer conformation (such as 1znj.pdb), with six phenol ligands located at the type I site, with two Zn$^{2+}$ cations and two Cl$^-$ anions as their axial ligands. The type I site is located at the dimer interface, with the phenol forming hydrogen bonds with CO$^{A6}$ and NH$^{A11}$ as well as loose Van der Waals contacts (~3.7 Å) with the imidazole of His$^{B5}$ [163]. The hydrogen bonds in the dimer interface are not significantly disrupted (Figure 33-A) by the $N$-methylation of the Phe$^{B25}$ peptide bond. All four hydrogen bonds (B$^{24}$-B$^{26}$) were preserved, with just an increase of ~0.2 Å of the CO$^{B24}$-NH$^{D26}$. The position of the $N$-methylation is outside the dimer interface, which maintains these hydrogen bonds. The $N$Me$^{B25}$ is accommodated in the A$^{19}$-A$^{21}$ region.

The most dramatic disruption of the ability of insulin to dimerise, based upon ITC data, is caused by the $N$-methylation of the B$^{25}$ amide in [$N$MePhe$^{B25}$]-insulin. The dimer interface is very similar to WT insulin (Figure 33-A) with the $N$-methylations easily accommodated, by pointing outwards from the dimer interface. The lack of an amide hydrogen bond at B$^{25}$ resulted in no heat effect, the same as the negative control. This hydrogen bond is usually donated to CO$^{A19}$ (Figure 34), which when released from the C-terminal region of the B-chain will result in the higher mobility of this region. This increased mobility can be observed in the side chain of Phe$^{B25}$ as only one can be seen. The monomeric nature of [NMePhe$^{B25}$]-insulin can be attributed to the detachment of Phe$^{B25}$ from the A chain (Tyr$^{A19}$)[159].
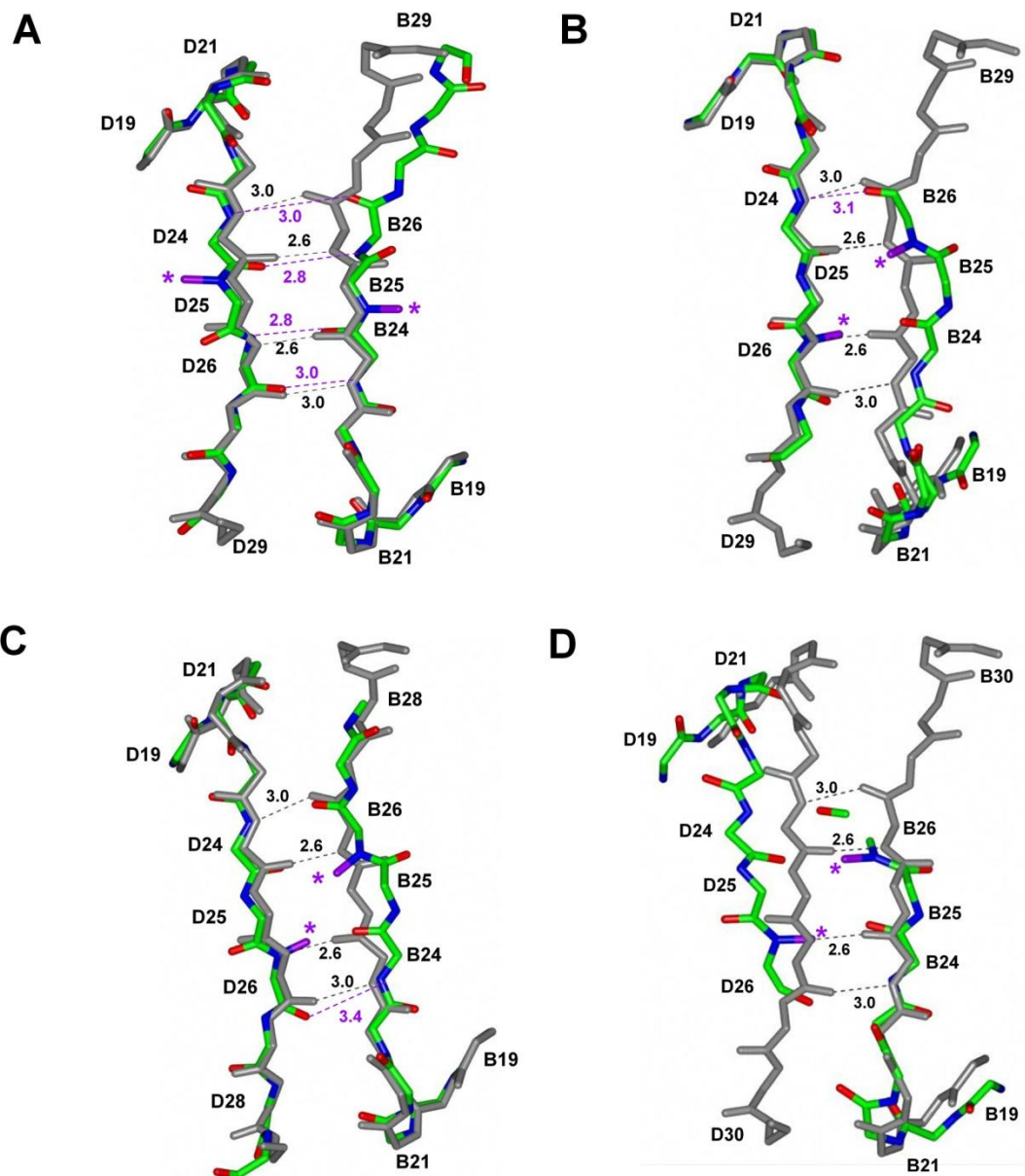
***Figure 33. The effect upon the dimer interface by the introduction of an N-methylation.***

Reference wild type insulin is in grey (1znj.pdb) with hydrogen bonds indicated by dashed black lines, distances are in Å. (A) Is the dimer interface of [NMePhe$^{B25}$]-insulin. (B) Is the dimer interface of [NMeTyr$^{B26}$]-insulin. (C) Is the dimer interface of [Tyr$^{B25}$,NMePhe$^{B26}$]-insulin . (D) Is the dimer interface of [Tyr$^{B25}$,NMePhe$^{B26}$,Lys$^{B28}$,Pro$^{B29}$]-insulin. Carbon atoms in green, oxygen - red and nitrogen - blue. The methyl groups on N-methylated atoms are coloured in purple and are indicated by a purple asterisk.
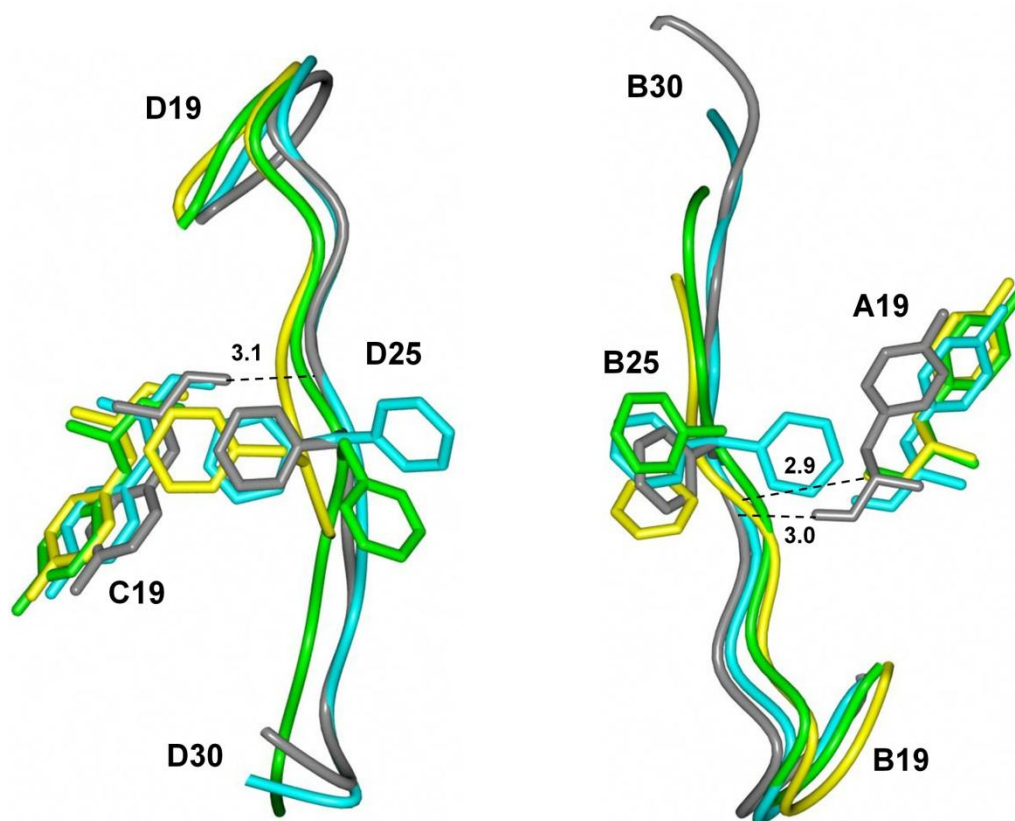
*Figure 34. The effect of the intrachain coupling of the $NH^{B25}$-$CO^{A19}$ hydrogen bond.*

Wild type insulin ($T_6$) 1mso.pdb is shown in grey, wild type insulin ($R_6$) 1znj.pdb is shown in cyan, [$N$MePhe$^{B25}$]-insulin is shown in green and [$N$MeTyr$^{B26}$]-insulin is shown in yellow. Dashed lines indicate hydrogen bonds (distances given in Å).

[$N$MeTyr$^{B26}$]-insulin was obtained in both the $R_6$ hexameric and monomeric forms, under the appropriate condition. The hexamer crystals appeared very similar to those of [$N$MePhe$^{B25}$]-insulin (as well as the crystallisation conditions). The $N$-methylation of Tyr$^{B26}$ caused a much more disturbed dimer interface, where both groups point towards the dimer interface, causing only one of the original four hydrogen bonds to remain intact, $CO^{B26}$-$NH^{D24}$ (Figure 33-B). The disturbance caused by the $N$-methylation is not uniform in the two β-strand; one follows the path of WT insulin, whilst the other β-strands experience a bulge in the $B^{24}$-$B^{26}$ region, to allow the $N$-methyl group to be accommodated. This bulging of the dimer interface causes structural stress that is seen in the higher mobility of $B^{28}$-$B^{30}$. The bulge in the β-sheet to accommodate the N-methylations causes the NHB$^{25}$ peptide to move closer to $CO^{A19}$, resulting in a unique 2.9 Å hydrogen bond, unseen in $R_6$ hexamers (Figure 34). It could be this hydrogen bond that is responsible for the lower dimerisation capability.

[Tyr$^{B25}$,$N$MePhe$^{B26}$]-insulin was crystallised as an $R_6$ hexamer form similar to those of [$N$MeTyr$^{B26}$]-insulin and [$N$MePhe$^{B25}$]-insulin, though the bulge at the dimer interface of

this analogue is apparent in both β-sheets (Figure 33-C). The structural consequences of the *N*-methylation are the same as those discussed for [*N*MeTyr[B26]]-insulin, whilst the side chain substitutions of residues B[25] and B[26] has little effect.

[Tyr[B25],*N*MePhe[B26],Lys[B28],Pro[B29]]-insulin was crystallised as an $R_6$ hexamer. The multiple changes to the analogue from insulin WT allow extensive probing and characterisation of the dimer interface and its hydrogen bonding network. The B[25]-B[26] Tyr↔Phe and B[28]-B[29] Lys↔Pro substitutions along with the *N*-methylation of B[26] cause the dimer interface to become fully disrupted (Figure 33-D), meaning that $R_1R_1$ dimer formation here is entirely dependent upon phenol induced interactions. The disruption of the dimer interface is similar to that seen in [*N*MeTyr[B26]]-insulin, where the two β-sheets aren't uniformly displaced. One of the β-strands follows the conformation shown by WT insulin, whereas the counter β-strand departs from the dimer interface dramatically from D[23] onwards and becomes fully disordered crystallographically after D[27].

Analogues that have peptide bonds at the B[24]-B[25] or B[25]-B[26] position reduced $CH_2$-NH have extremely low binding affinities. Their binding affinities can be attributed to the loss of the carbonyl oxygens that form hydrogen bonds with the IR. [*N*MeTyr[B26]]-insulin[75] shows a reduced binding affinity (21%) to WT insulin. This is reduced less than in analogues that are *N*-methylated in the B[24] and B[25] positions (3 and 0.17% respectively). This indicates that the NH of B[26] (and its hydrogen bond) are less important in the binding of insulin to the IR than those in positions B[24] and B[25], or has a different role. It has been shown[75] that a *N*-methylation of Tyr[B26] causes the induction of a novel type II β-turn at positions B[24]-B[26] (the B26 turn), which causes the departure of the C-terminus of the B-chain (the B[22]-B[30] β-strand) from the usual hexamer/dimer conformation. There are several reasons to explain the reduced binding affinity of [*N*MeTyr[B26]]-insulin. The first is that the arrangement of B[27]-B[30] caused by the *N*-methylation may be non-optimal. The reduced affinity may also be due to the loss of the NH[B26] hydrogen bonds that are required for binding to the IR.

The analogue [Phe[B26]]-insulin shows the importance of the phenolic character of Tyr[B26] to binding affinity, with reduced binding affinity (46%)[71]. An analogue that combines the loss of the phenolic character of the side chain along with a *N*-methylation of this peptide bond, [*N*MePhe[B26]]-insulin, results in a binding affinity of 4%. This accumulation of multiple unfavourable modifications is further demonstrated by [*N*MePhe[B25],*N*MePhe[B26]]-insulin, which has a binding affinity of 0.44% due to the two *N*-methylations and the loss of the phenolic nature of the B[26] side chain.

In summary, the evidences resulting from the dimerisation capabilities of the [*N*MePhe$^{B24}$]-insulin and [*N*MePhe$^{B26}$]-insulin analogues and their crystal structures, indicate that the loss of the NHB$^{24}$-COB$^{26}$ hydrogen bond has the greatest impact on dimer formation. This is due to its flanking position in the dimer interface, allowing solvent to more easily penetrate the β-sheet, resulting in the solvation/breaking of the two remaining central hydrogen bonds[164]. Unzipping of the two remaining hydrogen bonds could also be due to this lack of the flanking hydrogen bonds combined with the highly mobile B$^{27}$-B$^{30}$ region. The insulin analogue [*N*MePhe$^{B25}$]-insulin displays a dimerisation capability identical to that of DOI. This *N*-methylation results in the loss of the hydrogen bond with the carbonyl Tyr$^{A19}$. Consequently the C-terminus of the B-chain is no longer held in place, and is allowed to move freely in solution.

## 3.5. The Importance of the B24 and B26 sites in Insulin Activity

### 3.5.1. The Steric Importance of the B24 Site

The role played by Phe$^{B24}$ in insulin activation was probed by a combination of crystal and NMR structures (all NMR structures have been determined at IOCB). Phe$^{B24}$ is an invariant residue in all vertebrate insulins. In the storage forms of the hormone the side chain of Phe$^{B24}$ points towards the hydrophobic core of the protein. The main chain CO and NH of Phe$^{B24}$ are crucial for the activity of insulin and for its ability to dimerise[152,159].

[His$^{B24}$]-insulin analogue is almost inactive (1.5%) and was crystallised as a monomer with the C-terminus of the B-chain being completely disordered (B$^{22}$-B$^{30}$) (Figure 35-A). The side chain of Tyr$^{B16}$ is flipped from its usual position into the B$^{24}$ pocket, which is stabilised by van der Waals contacts with the Val$^{B12}$ side chain. This indicates that the Phe$^{B24}$ pocket needs to be occupied in the native hormone.

The NMR structure of [His$^{B24}$]-insulin at pH 2 shows great disorder at the C-terminus of the B-chain, similar to the crystal structure (Figure 35-B). This may be attributed to the disruptive effect of double protonated His side chain[165] that cannot be maintained in Phe$^{B24}$ hydrophobic cavity.

In contrast the NMR structure of [His$^{B24}$]-insulin at pH 8 is well ordered (Figure 35-C). The His side chain is accommodated in the B$^{24}$ pocket. The N$\varepsilon$ is protonated and forms a 2.4 Å hydrogen bond with CO$^{B12}$, whilst N$\delta$ is deprotonated and forms a 2.9 Å hydrogen bond with NH$^{B26}$. The NMR structure indicates that the dimer interface had been disrupted, which explains its inability to dimerise. Moreover, the monomeric character of this analogue may results from quite visible increased compactness of the hole insulin molecule; the B20-B30 chain is closer to the core to the hormone than in WT insulin. In addition, the side chain of Tyr$^{B26}$ is shifted, and forms a $\pi$ stacking interaction with Pro$^{B28}$, whilst the side chain hydroxyl forms hydrogen bonds with the main chain CO$^{B8}$ (2.8 Å) and/or NH$^{B8}$ (3.3 Å).
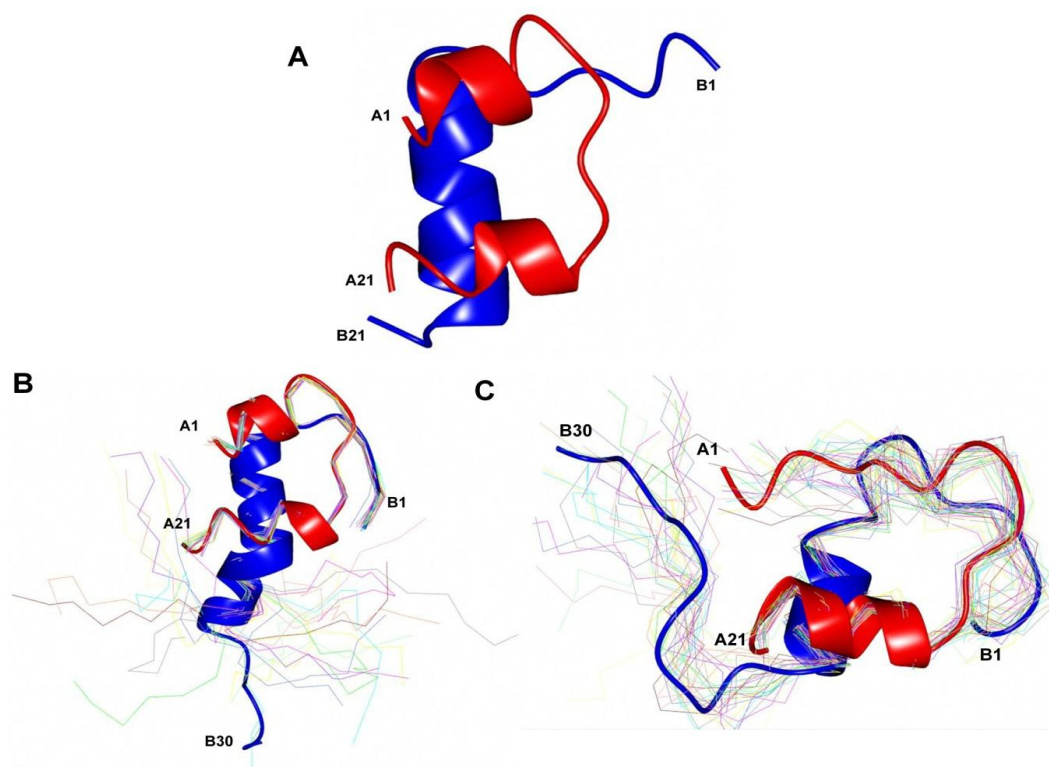
*Figure 35. The crystal and NMR structures of [His$^{B24}$]-insulin at different pHs.*
(A) The crystal structure of [His$^{B24}$]-insulin at pH 3. (B) NMR structure of [His$^{B24}$]-insulin at pH 2. (C) NMR structure of [His$^{B24}$]-insulin at pH 8. The A-chain is coloured red in all part, and the B-chain blue. NMR structures were produced by collaborators in Prague.

The [D-His$^{B24}$]-insulin analogue is a high affinity insulin (217%) that is monomeric in NMR structures at pH 2 and pH 8. At pH 2 the NMR structure shows disorder at the C-terminus of the B-chain (B$^{22}$-B$^{30}$) as seen in [His$^{B24}$]-insulin (Figure 36-A). The pH 8 NMR structure of [D-His$^{B24}$]-insulin is more ordered in the C-terminus of the B-chain than L-His or WT insulin (Figure 36-B). The D-His side chain is oriented out of the insulin core, exposed to the solvent. As a result of this transition, the side chain of Phe$^{B25}$ has turned and downshifted towards the Phe$^{B24}$ pocket, partially taking its place. This clearly indicates the importance of the maintenance of the integrity of the B$^{24}$ pocket. The B$^{21}$-B$^{23}$ turn is more open that that seen in WT insulin. The presence of D-His, a non-naturally occurring amino acid may result in an artificial structure, however the enhanced flexibility of the B$^{25}$-B$^{30}$ due to the lower chance of hydrogen bonds forming, results in a higher affinity of this analogue.
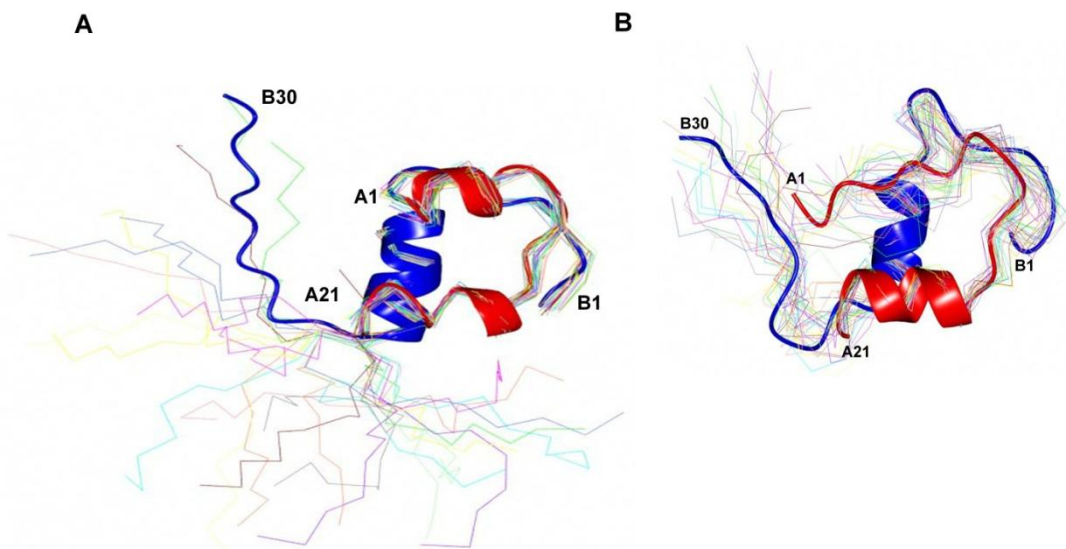
*Figure 36. NMR structures of [D-His^{B24}]-insulin at different pHs.*

(A) NMR structure of [D-His$^{B24}$]-insulin at pH 2. (B) NMR structure of [D-His$^{B24}$]-insulin at pH 8. The A-chain is coloured red in all part, and the B-chain blue. NMR structures were produced by collaborators in Prague.

### 3.5.2. The Steric Importance of the B26 Site

Tyrosine at position $B^{26}$ is invariant in many species, although its role in the interaction with the IR is unclear. Replacing the Tyr with another amino acid can lead to analogues with a range of affinities between less than 1%, up to 4-5 times higher than that of WT insulin[150].

A reduction in the binding affinity of an analogue can be accounted for by an increase in the size of the amino acid in the $B^{26}$ position in shortened analogues ([$N$MePhe$^{B26}$]-DTI-NH$_2$ [150], [$N$MeTyr$^{B26}$]-DTI-NH$_2$) or in full length analogues ([$N$MeTyr$^{B26}$]-insulin and [$N$MeAla$^{B26}$]-insulin)[150]. The increase in the size of the amino acid at the $B^{26}$ position results in binding affinities of 21-72%.

The full length analogues [$N$MeTyr$^{B26}$]-insulin and [$N$MeAla$^{B26}$]-insulin both possess the same structural features around the B26 turn[75] as with the high affinity analogues, despite their low affinity (21 % for both analogues). The structures show that the $N$-methylated group is located on the outside of the $B^{26}$ position, causing the B26 turn. Residues $B^{27}$-$B^{30}$ are stabilised by hydrogen bonds between amino acids located at í and í+3 positions.

The structural importance of the $B^{26}$ position is chiral dependent as indicated by its replacement with [D-Pro$^{B26}$]-DTI-NH$_2$ (359%). The chirality of the D-Pro is important at the Cα of the amino acid., as it mimics the *cis* Phe$^{B25}$-Xaa$^{B26}$ $N$-methylated peptide bond in the high affinity analogues such as [$N$MeAla$^{B26}$]-DTI-NH$_2$.

Some $B^{26}$ truncated analogues, [$N$MeTyr$^{B26}$]-DTI-NH$_2$ and [$N$MePhe$^{B26}$]-DTI-NH$_2$ show much lower affinities than their higher affinity relatives (72% and 36% respectively). In these analogues the presence of a bulky side group Phe$^{B26}$/Tyr$^{B26}$ or the $N$-methylation of the $B^{26}$ peptide bond in [$N$MeTyr$^{B26}$]-DTI-NH$_2$ could result in a detrimental entropic effect. This causes a destabilising effect on the formation of the B26 turn. The lack of a crystal structure suggests increased flexibility, which has been observed using NMR in a related truncated insulin analogue[166].

In full length WT insulin, substitution of Tyr$^{B26}$ with Ala$^{B26}$ results in reduction of binding affinity to ~36%[78]. The crystal structure of monomeric [$N$MeAla$^{B26}$]-insulin was obtained under both monomeric and hexameric conditions, whilst [$N$MeTyr$^{B26}$]-insulin was obtained as a monomer and hexamer, under relevant monomeric and hexameric conditions. The presence of Ala infers a preference in state towards the monomer, whereas Tyr doesn't preferentially crystallise in a particular form. The low potencies of [$N$MeTyr$^{B26}$]-insulin and [$N$MeAla$^{B26}$]-insulin (~21%) confirm the trend seen that full length analogues are generally less active than their truncated homologues[73]. This could result from the $B^{27}$-$B^{30}$ region

causing detrimental effects to the structure, such as steric and entropic factors. A reason for this could be that in order to stabilise the $B^{27}$-$B^{30}$ segment, the IR could be required to provide significant contacts. Also, it is possible that the structural restraints enforced by formation of the B26 turn may not be the optimal conformation for the $B^{27}$-$B^{30}$ residues, or may not even be the final conformation.

### Table 11. The $K_d$ and relative IR binding affinity of human insulin and insulin analogues.

The $K_d$ is a representation of the dissociation constant of binding of insulin/insulin analogue to the IR. Each value shown indicates the mean ± the standard error. The number in parentheses indicates the number of replications. The relative binding affinity is defined as ($K_d$ of human insulin/$K_d$ of analogue) x 100. The experimental values are taken from Antolikova et al.[159].

|  | $K_d$ ± SEM [nM] (n) | Binding Affinity [%] |
|---|---|---|
| Human Insulin | 0.39 ± 0.01 (6) | 100 |
| [Gln$^{B26}$]-insulin | 0.51 ± 0.02 (4) | 76 |
| [Asp$^{B26}$]-insulin | 0.44 ± 0.01 (4) | 89 |
| [Asn$^{B26}$]-insulin | 0.47 ± 0.01 (4) | 83 |

The ability of insulin, [Phe$^{B26}$]-insulin and [Asn$^{B26}$]-insulin to dimerise are all different. The ability of [Asn$^{B26}$]-insulin, and [Phe$^{B26}$]-insulin to dimerise is reduced approximately 86 and 3 fold respectively, compared to insulin, as determined by ITC (Table 12).

### Table 12. ITC analysis of the dimerization capabilities of insulin and insulin analogues.

Each value shown indicates the mean ± the standard error. The number in parentheses indicates the number of replications. The experimental values are taken from Antolikova et al.[159].

|  | $K_d$ |
|---|---|
| Human Insulin (n = 8) | 8.81 ± 1.05 |
| [Asn$^{B26}$]-insulin (n = 3) | 865.3 ± 308.6 |
| [Phe$^{B26}$]-insulin (n = 3) | 27.2 ± 7.4 |

Both [Phe$^{B26}$]-insulin and [Asn$^{B26}$]-insulin were crystallized under dimeric conditions, which gave rise to an unusual monomer-monomer crystal organisation (Figure 37). In [Phe$^{B26}$]-insulin and [Asn$^{B26}$]-insulin, an unusual pseudo-dimer is formed by rotation of the α-helix in the B-chain by 50°, where the only close contact between the monomers is between the –OH of serine A$^{12}$ of one monomer, and the –NH$_2$ of glutamine B$^4$ in the second. These form a hydrogen bond of ~2.6 Å (Figure 37). Whilst the spatial arrangements of [Asn$^{B26}$]-insulin and [Phe$^{B26}$]-insulin are similar, there are also some different features (Figure 38).
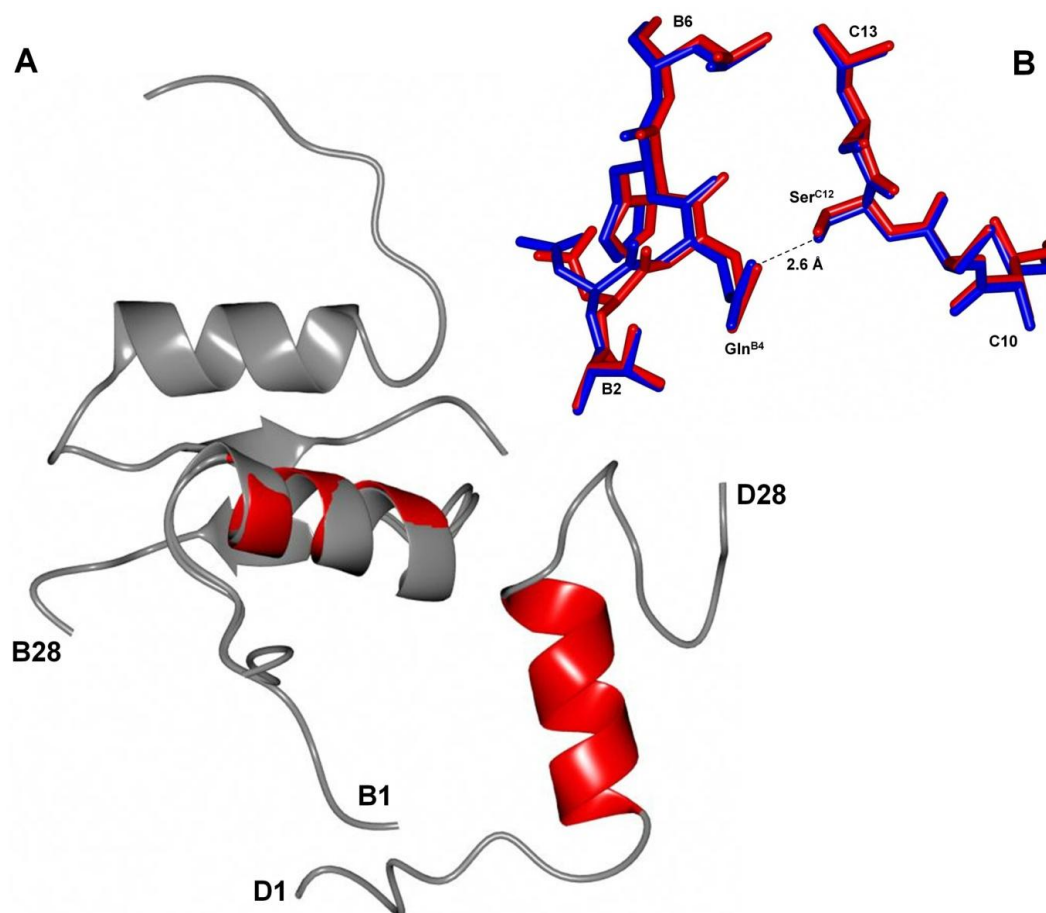


*Figure 37. A representation of the unusual dimer interface in [Asn$^{B26}$]-insulin.*

(A) The main chain representation (chains B and D only) of the dimer interface in human insulin (1znj.pdb in grey) and of [Asn$^{B26}$]-insulin (red). (B), The close contact between two molecules in the unusual dimer interface of [Asn$^{B26}$]-insulin (chains B and C, in red) and [Phe$^{B26}$]-insulin (Chains B and C, in blue). Dashed lines indicate hydrogen bonds.
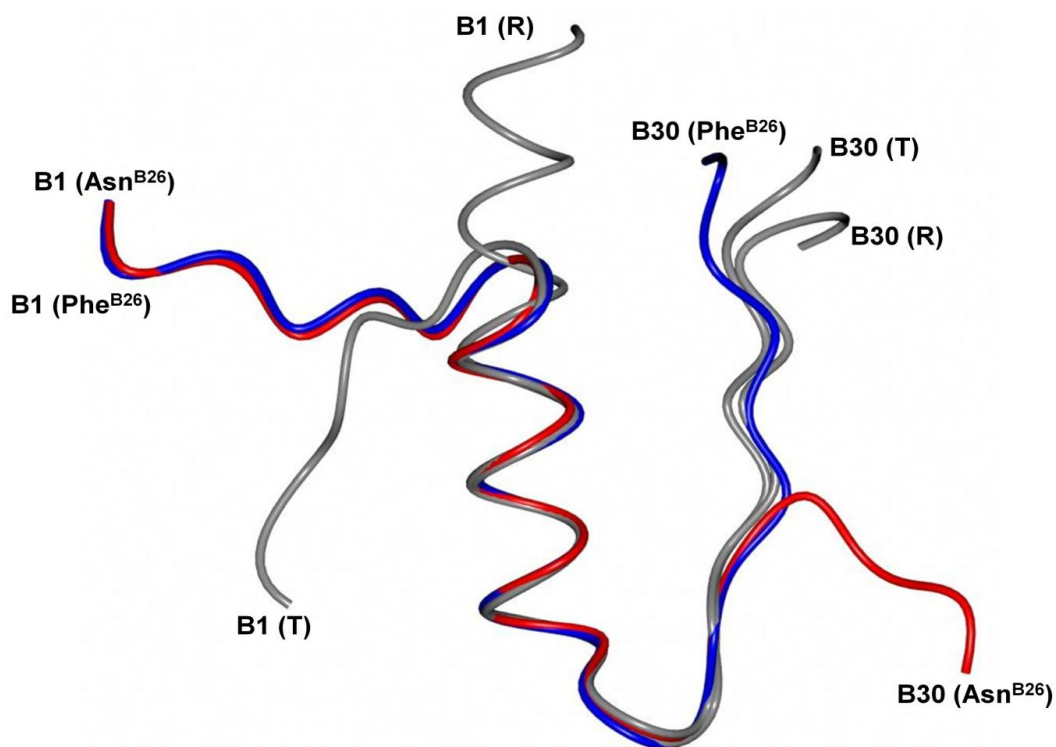
*Figure 38. An overlay of the main chain of T- and R-state insulin, [Asn^{B26}]-insulin and [Phe^{B26}]-insulin.*

T-state insulin (1znj.pdb in grey), R-state (1mso.pdb in grey), [Asn$^{B26}$]-insulin (in red) and [Phe$^{B26}$]-insulin (in blue).

Both substitutions at the B$^{26}$ position have a similar effect on the N-terminus of the B-chain, which occupies a conformation similar to that in [*N*MeAla$^{B26}$]-DTI-NH$_2$, which is between the T- and R-states. This conformation is formed by a type II β-turn, which is stabilised by a pseudo *i*+3 hydrogen bond between the main chain NH of His$^{B5}$ and the side chain of Asn$^{B3}$ [75]. The hydrogen bond formed by the side chain hydroxyl of Tyr$^{B26}$ via a bridging water molecule with Gly$^{B8}$ is lost in both analogues. This causes the disruption of an apolar area consisting of Gly$^{B8}$, Val$^{B12}$, Tyr$^{B16}$, Phe$^{B24}$ and Tyr$^{B26}$ (Figure 39), and therefore a switch of conformation to the intermediate state. This intermediate state was previously described in several insulin analogues[75,167,168], however without any thorough discussions of its relevance.

In contrast, the C-terminus of the B-chain of [Asn$^{B26}$]-insulin is in a different conformation. In [Phe$^{B26}$]-insulin, the C-terminus of the B-chain occupies a similar conformation to that seen in WT insulin. The phenyl ring of Phe$^{B26}$ occupies the same position as that of the tyrosine in WT insulin, though the Phe$^{B25}$NH-Tyr$^{A19}$CO hydrogen bond, which is crucial to insulin dimerisation[159] is missing in both monomers with the distance now 4.65 and 4.20 Å.
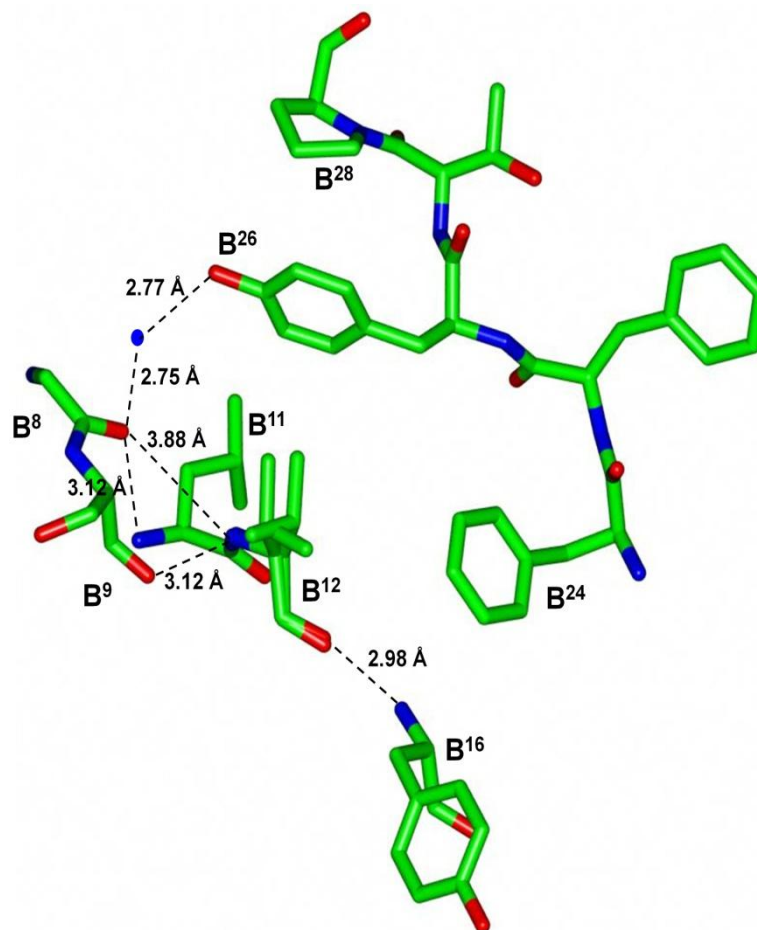
***Figure 39. A view of the interactions of Tyr^B26 and its surrounding residues in the T-state of human insulin.***

Carbon atoms in green, oxygen - red and nitrogen - blue. Dashed lines indicate hydrogen bonds. PDB used: 1znj.pdb.

[Asn^B26]-insulin has a novel conformation in the C-terminus of the B-chain compared to [Phe^B26]-insulin. The polar residue asparagine is forced out of the hydrophobic pocket, causing a new C-terminal conformation to form, which causes new intramolecular interactions to be created. The C-terminal turn is stabilised by close contacts with Asn^A21, where three hydrogen bonds are created with several C-terminal residues (Figure 40). The hydrogen bond between Gly^B23CO and Asn^A21NH in WT insulin is 3.0 Å, whilst in [Asn^B26]-insulin it is decreased to 2.8 Å. The side chain of Asn^A21 forms two hydrogen bonds that are not seen in WT insulin. The first is a bifureated hydrogen bond, between the carbonyl of the Asn^A21 side chain and the main chain amides of Phe^B25 (3.3 Å) and Asn^B26 (2.7 Å). The second hydrogen bond is between the amide of the Asn^A21 side chain, and the Asn^B26 main chain carbonyl (2.8 Å) (Figure 40). The Phe^B25NH-Tyr^A19CO hydrogen bond that is crucial to insulin dimerisation[159] is missing in both monomers, with the distance now 5.22 and 4.68 Å.
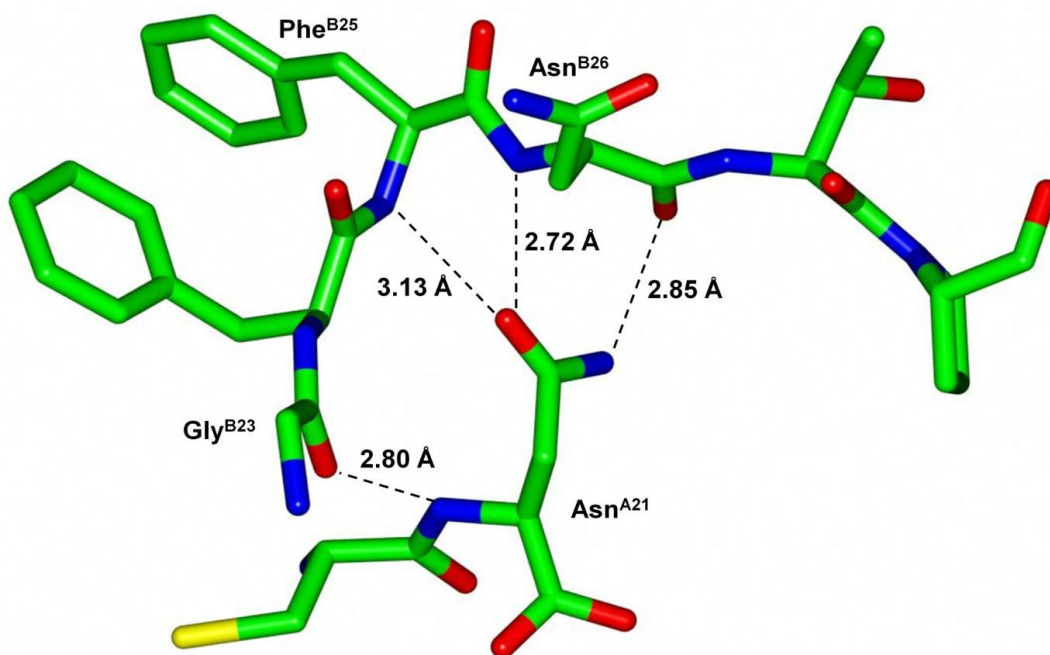
*Figure 40. The bent conformation of the C-terminus of the B-chain of [Asn^{B26}]-insulin, stabilised by interactions with Asn^{A21}.*

Carbon atoms in green, oxygen - red and nitrogen - blue. Dashed lines indicate hydrogen bonds.

## 3.6. Conclusion

All the N-methylated full-length analogues showed a significant reduction in binding affinity, which confirmed the role of the amide hydrogens direct interaction/involvement with the IR. The systematic *N*-methylation of the individual residues caused a reduction in dimerisation ability.

The dramatic loss in dimerisation capability was caused by the *N*-methylation of $B^{25}$, which disrupted the $NHB^{25}$-$COA^{19}$ hydrogen bond that links the β-strand to the core of insulin. Greater mobility of the C-terminus of the B-chain causes insulin to tend towards the monomeric state.

The analogue [*N*MePhe$^{B24}$]-insulin produced the greatest loss of dimerisation ability from the loss of a hydrogen bond in the dimer interface. This analogue lost the flanking hydrogen bonds, allowing the solvent easy penetration into the dimer β-sheet. The increased accessibility of the dimer interface allows easier break of the two remaining hydrogen bonds[164].

These analogues helped to determine the importance and contribution of the $NHB^{24}$-$COB^{26}$, $NHB^{26}$-$COB^{24}$ and $NHB^{25}$-$COA^{19}$ hydrogen bonds in the formation and stabilisation of insulin dimers. Positions $B^{24}$ and $B^{26}$ in the dimer interface have been shown integral to stability of the insulin dimer[159]. Substitutions of each of these residues has created analogues of both high and low affinities[75,100,150].

The functional and structural observations lead to a greater understanding of the association-dissociation pathways of insulin. This may allow the design of more controlled insulin drugs for the treatment of diabetes[159].

This work also provides a more systematic investigation of the role of individual $B^{24}$ and $B^{26}$ side chain for insulin:IR binding.

Substitution of $B^{24}$ with His causes structures that are pH and chirality dependent. In both isomers at low pH the C-terminus of the B-chain is disordered, whilst at pH 8, the region becomes structured. The low pH in both the crystal and NMR structures causes the side chains in [His$^{B24}$]-insulin, [D-His$^{B24}$]-insulin and surrounding residues to become protonated. This results in hydrogen bonds not being formed, preventing the $B^{24}$ position from becoming an anchor point at the C-terminus of the B-chain, resulting in the disorder seen in the structures.

The $B^{24}$ side chain of [His$^{B24}$]-insulin sits in a similar position to that observed in WT insulin, pointing towards the hydrophobic core. Due to the presence of N atoms in [His$^{B24}$]-insulin, hydrogen bonds are able to be formed with neighbouring amino acids that would not be possible in WT insulin. This causes the $B^{24}$ pocket to be more tightly bound. During

the activation of insulin, the C-terminus of the B-chain becomes detached from the core of the insulin molecule, though in [His[B24]]-insulin this is greatly reduced due to the tighter binding in the $B^{24}$ pocket.

The NMR structure of [D-His[B24]]-insulin at pH 8 shows an arrangement that is more flexible than [His[B24]]-insulin yet ordered. Contradictory to that seen in insulin and [His[B24]]-insulin, the $B^{24}$ side chain of [D-His[B24]]-insulin points away from the hydrophobic core. This increases the ability of the C-terminus of the B-chain to move, which is shown by the increased binding affinity, 217%. This increase in flexibility has ramifications on the ability of insulin to form dimers[159], so changing the residue would have other detrimental effects on insulins method of action. This indicates that Phe is the optimum residue at $B^{24}$, as it does not bind too tightly or weakly.

The Glu, Asn and Asp $B^{26}$ analogues were designed based upon the work of Brange[106,169]. The binding affinities of these analogues are close to that of WT insulin (Table 11). They indicate that a substitution by a polar residue is better incorporated at the B26 site[78].

The main chain and side chain of Tyr[B26] play roles in monomer:monomer interactions[80]. The side chain of Tyr[B26] is part of a polar region consisting of Gly[B8], Val[B12], Tyr[B16] and Phe[B24], that is involved in the stabilisation of the structure of insulin[170]. A small change in the amino acid at $B^{26}$ could disrupt these interactions, and cause a change in the insulin structure.

ITC data (Table 12) supports the inability to form dimer or hexamer crystals for [Asn[B26]]-insulin and [Phe[B26]]-insulin. [Asn[B26]]-insulin shows a bent structure in the C-terminus of the B-chain, caused by the repulsion of the Asn[B26] side chain from the hydrophobic pocket. The loss of the Tyr[B26]OH-Gly[B8]CO hydrogen bond as described previously, led to the intermediate conformation seen at the B-chain N-terminus which has been seen previously[75,168]. This may play a role in insulin:IR interaction.

In [Phe[B26]]-insulin the Phe[B26] occupies the same space as the tyrosine ring in WT insulin, meaning that the hydrophobic core is maintained. The loss of the Tyr[B26]OH-water-Gly[B8]CO hydrogen bond leads to the intermediate conformation seen at the N-terminus of the B-chain. The loss of this hydrogen bond has a knock-on effect on the structurally important Phe[B25]NH-Tyr[A19]CO hydrogen bond, which is lost[159], causing an enhanced mobility of the C-terminal β-strand of the B-chain.

[Asn[B26]]-insulin forms a new C-terminal turn in the B-chain similar to the B26 turn[75], and retains near full binding affinity. This new turn may be closer to the actual conformation of WT insulin complexed with the IR, than seen in analogues with $B^{26}$ *N*-methylation due to the lack of a non biological addition.

The susceptibility of the C-terminus of the B-chain to form a turn indicates that this

plays an important part in IR binding. Initial binding may induce detachment of the Tyr$^{B26}$ side chain from the hydrophobic pocket, causing the formation of the B-chain C-terminus turn, and the loss of the previously described hydrogen bonds, causing the formation of the intermediate conformation of the N-terminus. These conformational changes allow both N- and C-terminal residues to interact with the IR. The B$^{24}$ pocket acts as an anchoring point allowing the B-chain C-terminus turn to form. The local structures at either termini of the B-chain possess the structural signatures of activated insulin, which imply an activation scheme (Figure 41). The formation of the active conformation of the insulin molecule and binding to the IR is simultaneous.
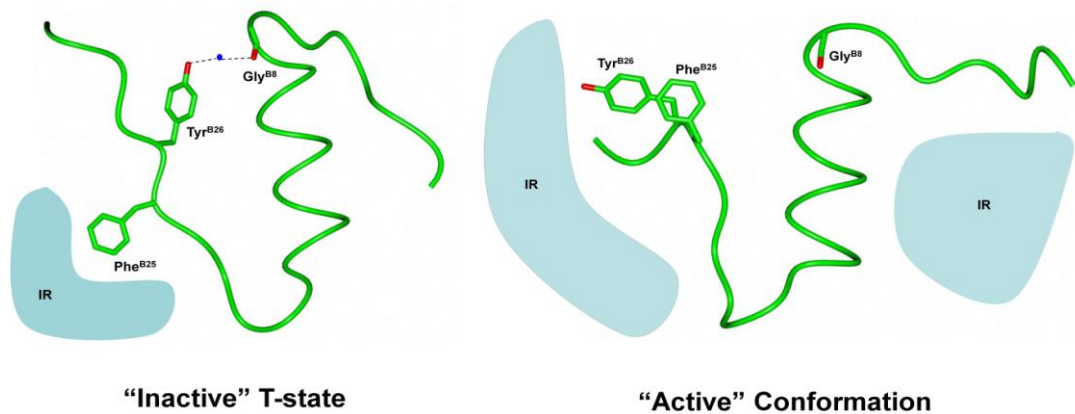


*Figure 41. A hypothetical scheme of insulin-IR binding.*

# Chapter 4.  Insulin Receptor Expression

## 4.1. Summary

The IR has been expressed many times previously[63,50,171,172] using mammalian cell cultures. The IR L1 domain contains the majority of residues involved in insulin binding[55,63,60]. This makes it a key target for expression with the aim of achieving a 'workable' and compact insulin:IR complex. The IR consists of a series of well-defined domains, linked by flexible linker regions (Figure 13). Removal of these flexible regions would increase the likelihood of obtaining a crystal structure of the complex thus facilitating more systematic studies on this hormone:receptor system. Therefore the work in this chapter focused on cloning and expression of soluble L1 domain of IR for structural studies.

Initial protein expression was attempted in *E. coli* using Rosetta-Gami 2, to aid the correct formation of disulphide bond but an insoluble protein expression was produced here. As the expression in TUNER cells produced high levels of protein expression, *in vitro* refolding was attempted. Following on-column and redox refolding of the L1 domain soluble protein was achieved. Analysis of the disulphide bonds, indicated that they had formed in an intermolecular manner, rather than intramolecular. This meant that the L1 domain had to be expressed in a higher organism. However, expression in *S. cerevisiae* and *P. pastoris* failed to produce soluble, correctly formed protein.

The L1 domain had not been previously expressed alone, but instead as part of a larger IR construct. There may be interactions between the L1 domain and other domains that are important for protein stability that are absent by removing these other domains. Mammalian cells contain complex post-translational machinery that aid disulphide bond formation, and other post-translational modification hence they may be ultimately needed for more successful expression of the l1 domain.

## 4.2. The L1 Region and Previous Expression

The L1 region of the IR comprises the amino acid sequence shown in Figure 42, which contains four disulphide bonds $Cys^8$-$Cys^{26}$, $Cys^{126}$-$Cys^{155}$, $Cys^{159}$-$Cys^{182}$ and $Cys^{169}$-$Cys^{188}$.

```
                10                  20                  30
H L Y P G E V C P G   M D I R N N L T R L   H E L E N C S V I E
                40                  50                  60
G H L Q I L L M F K   T R P E D F R D L S   F P K L I M I T D I
                70                  80                  90
L L L F R V Y G L E   S L K D L F P N L T   V I R G S R L F F N
               100                 110                 120
Y A L V I F E M V H   L K E L G L Y N L M   N I T R G S V R I E
               130                 140                 150
K N N E L C Y L A T   I D W S R I L D S V   E D N H I V L N K D
               160                 170                 180
D N E E C G D I C P   G T A K G K T N C P   A T V I N G Q F V E
               190
R C W T H S H C Q K
```

*Figure 42. The amino acid sequence of the L1 region in the human IR.*

The L1 region of the IR has been expressed previously as part of a larger construct. Using mammalian expression systems, Lou *et al.*[63] expressed the L1-CR-L2 regions in Lec8 cells, which are Chinese Hamster Ovary (CHO) cells that have limited protein glycosylation abilities. Sparrow *et al.*[50] produced the entire ectodomain of the IR in CHO-K1 (CHO cells that require proline in the medium for growth) and Lec8 cells. Kristensen *et al.*[171] expressed the L1-CR-L2 regions with various lengths of C-terminal sequences in Baby Hamster Kidney cells (BHK), which give stable transformations. Huang *et al.*[172] expressed the L1-CR-L2-FnIII-1 and C-terminal insert domain-derived segment in 293H cells, which give good growth and transfection efficiency.

Two constructs were considered: full length of L1 (as Figure 42) and a short construct. The short construct, residues 5- 157 (Figure 42), has an apparent molecular weight of 17819.6 Da and a theoretical pI of 4.98, whilst the full length construct (entire L1 region) has an apparent molecular weight of 21928.4 Da, and a theoretical pI of 5.80. The constructs were designed after analysing the available structure of L1 (pdb 2hr7.pdb). The structure of the full length L1 region (Figure 43-A and B) is a barrel consisting of a series of β-sheets. The main binding residues are located on the larger β-sheet side, and are responsible for insulin binding[55,60,63]. The C-terminus lacks secondary structure, apart from small regions of β-strands (Figure 43-A). The short construct (Figure 43-B) was designed to remove this region, along with the two disulphide bonds closest to the C-terminus. The

benefit of removing these in an *E. coli* expression system is that the likelihood of the correct formation of the remaining disulphide bonds is increased. The disulphide bonds are at opposing termini, which increases the probability of the bonds forming correctly during translation of the L1 mRNA.
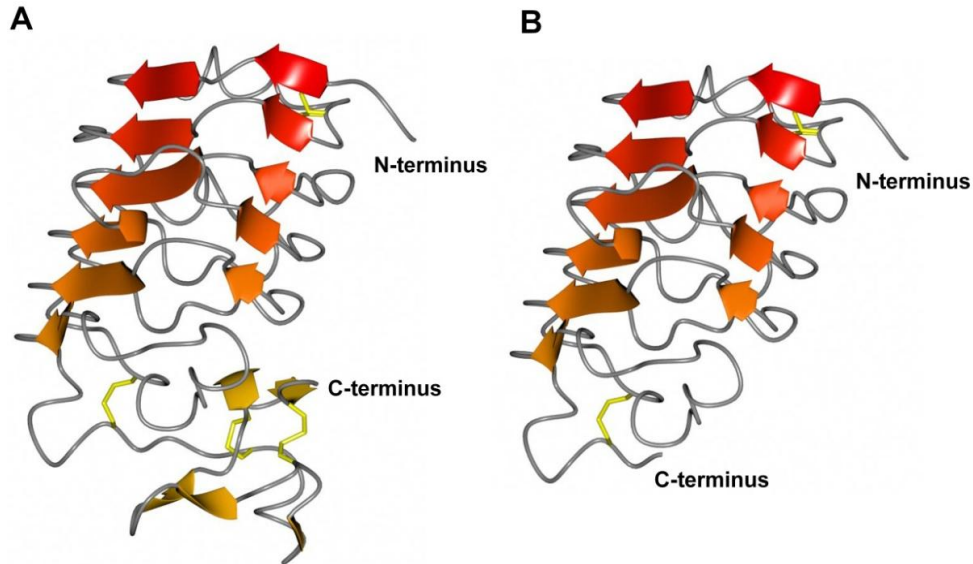


***Figure 43. Main chain comparison of the full length and short construct of the L1 region.***

(A) Residues 4-190 from Figure 42. (B) Residues 4-157 from Figure 42. (pdb used is 2hr7.pdb). Disulphide bonds are coloured in yellow.

The binding between insulin and the IR is dependent upon the CT peptide[64]. Various constructs of the IR were investigated by Kristensen *et al*., where it was shown that the addition of the CT peptide restored binding. I will aim to repeat this approach later on by adding the CT peptide to the expressed L1 region.

The most appropriate *E. coli* cell strains for expression of the L1 region would aid in the formation of the disulphide bonds. Rosetta-Gami 2 combines the advantages of two different strains; Rosetta 2 and Origami 2. Rosetta2 is designed to enhance expression of eukaryotic proteins that contain rare codons, whilst Origami 2 have mutations in thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes, which greatly enhance disulphide bond formation in the cytoplasm by making it a more reducing environment. ArcticExpress cells express the chaperone proteins Cpn10 and Cpn60 from the psychrophilic bacterium *Oleispira Antarctica,* which aid protein folding between 4-12 °C. BL21 cells are an all-purpose strain for high-level protein expression and easy induction. Lemo21 has the basic features of BL21 whilst also allowing tuneable expression. The expression level can be controlled by varying the level of lysozyme (lysY), which is the natural inhibitor of T7 RNA polymerase. The level of lysozyme is controlled by the addition of L-rhamnose to the expression culture at levels up to 2000 µM. TUNER cells have a

deletion of *lacZY* which means that an adjustable level of protein expression is possible. The *lacY* allows a uniform entry of IPTG into all cells in culture, producing a concentration-dependent level of induction.

## 4.3. Methods

### 4.3.1.  Ligation-Independent Cloning of the L1 Region

The L1 protein was expressed using a ligation-independent cloning (LIC) expression vector (pET-YSBLIC3C).  The LIC site contains a N-terminal His$_6$ tag and a 3C protease cleavage site.  Primers were used to amplify the target gene by PCR, leaving the LIC additions as an overhang. The full method, primers and LIC-specific sequence are given in the appendix. After purification, LIC-specific overhangs were created by incubating the insert with a T4 DNA polymerase (T4 pol) in the presence of dATP.  The vector was cleaved and linearised using BseRI and 5'-single-stranded overhangs were generated with T4 pol and dTTP.  The vector and target gene overhangs, which were designed to be complementary to each other, were allowed to anneal and then directly transformed into competent *E. coli* cells, where nicks can be repaired [173].

### 4.3.2.  Creation of L1 Region Constructs

Production of the L1 region was attempted using various pET and pMAL vectors. Vector specific primers were used to amplify the target gene (see appendix).  The plasmid was linearised using restriction enzymes, which left specific sticky ends (see appendix for details). Complementary sticky ends were also produced on the target gene.  Ligation and transformation was performed. Full details of this method and primers are given in the appendix.

### 4.3.3.  On-Column Refolding of Expressed L1 Protein

The target insoluble protein (L1 region) was solubilised using a denaturing solution consisting of 6 M guanidine hydrochloride, 20 mM Tris pH 8.0 and 10 mM β-mercaptoethanol (buffer A). Ni-NTA resin was equilibrated with buffer A before binding with the solubilised protein overnight, whilst being continually rocked in a sealed container. The column was washed with buffer A, followed by buffer A with 20 mM imidazole to remove any non-specific column-bound proteins.  The column was then washed with a detergent solution of 0.1% Triton X-100, 20 mM Tris pH 7.5 and 0.5 M NaCl to remove any contaminating protein molecules bound to the target protein or column.  The detergent was removed with 5 mM β-cyclodextrin, 20 mM tris pH 7.5 and 0.1 M NaCl.  The column was then washed with a solution of 0.1 M NaCl, 20 mM Tris pH 7.5 before the target protein

was eluted over an increasing imidazole gradient (0 – 1M) in 20 mM Tris pH 7.5, 0.1 M NaCl[174].

### 4.3.4.  Redox Refolding of L1 Expressed Protein

The L1 protein was buffer exchanged from the final on-column refolding buffer into 0.1 M NaCl, 0.1 M Tris (pH 7.5).  The refolding of L1 was performed by dilution with 20 mM $Na_2HPO_4$ (pH 7.3), 1 M L-arginine, 20% v/v glycerol, 10 mM reduced glutathione and 1 mM oxidised glutathione, to an L1 concentration of approximately 10 mg/ml, followed by dialysis against 20 mM $Na_2HPO_4$ (pH 7.3), 0.5 M L-arginine and 10% v/v glycerol for 12 hours at 4°C.  Further dialysis against 20 mM $Na_2HPO_4$ (pH 7.3), 0.2 M L-arginine and 5% v/v glycerol for 12 hours at 4°C was then followed by two dialysis steps against 20 mM $Na_2HPO_4$ (pH 7.3) for 12 hours at 4°C [175].

### 4.3.5.  Cloning, Transformation and Antibiotic Selection in *S. cerevisiae*

The L1 protein was cloned into the pYES2/CT vector. Sequence specific primers were used to amplify the L1 gene (see appendix).  The pYES2/CT plasmid was linearised using restriction enzymes which left sticky ends specific to the L1 gene.  Specific sticky ends were also produced on the L1 gene.  The plasmid and L1 gene sticky ends were complementary to each other so they were ligated and then directly transformed into competent *E. coli* cells, where nicks can be repaired.  The plasmid was isolated from the *E. coli* cells and then used to transform *S. cerevisiae* cells.  Yeast extract peptone dextrose medium (YPD) was inoculated with a single *S. cerevisiae* colony and left to grow at 30°C overnight.  The overnight culture was used to further inoculate a larger volume of YPD, which was then incubated for another 2-4 hours at 30°C.  Following incubation, the cells were pelleted and re-suspended in 1X TE (1 mM EDTA and 10 mM Tris, pH 7.5), and then re-pelleted and re-suspended in a reduced volume of 1X LiAc (pH 7.5) / 0.5X TE, before being incubated at room temperature for 30 min.  Plasmid DNA, 1 mg, was mixed with 100 µg thermally denatured salmon sperm DNA and 100 µl of the *S. cerevisiae* suspension.  1X LiAc / 40% v/v PEG3350 / 1X TE was added (700µl) and left at 30°C for 30 min.  DMSO was added (to a final concentration of ~10 %) before the cells were heat shocked at 42°C for 7 min.  The cells were pelleted immediately and re-suspended in 1X TE before being re-pelleted and re-suspended in a small volume of 1X TE.  The cell suspension was spread on

SC uracil selective plates containing glucose (SC is synthetic minimal defined medium for yeast) before being incubated at 30$^o$C to allow colonies to grow.

## 4.3.6. Cloning, Transformation and Antibiotic Selection in *P. pastoris*

The L1 protein was cloned into the pPICZαA vector. Creation of L1 constructs, ligation and transformation were performed as described previously (Full methods and primers can be found in the appendix). The plasmid was isolated from the *E. coli* cells and used to transform *P. pastoris* X-33 cells. YPD medium was inoculated with a single *P. pastoris* X-33 colony, which was left to grow at 30$^o$C overnight. This overnight culture was then used to inoculate a larger volume of YPD and left to grow again at 30$^o$C overnight. The overnight culture was centrifuged (1000 x g, 5 min), after which the cells were re-suspended in 50 ml sterile water and re-centrifuged (1000 x g, 5 min). The cells were then re-suspended in 20 ml sterile water, and re-centrifuged (1000 x g, 5 min) before being re-suspended again in 5 ml ice cold 1 M sorbitol. Following a final centrifugation the cells were re-suspended in 300 μl 1 M sorbitol and stored on ice.

The L1-pPICZαA plasmid was linearised by incubating 10 μg of the plasmid with 10 μl *SacI*. The linearisation reaction was stopped by heating the sample to 65$^o$C for 15 min. The linearised DNA was then purified using a PCR clean up kit.

An aliquot of cell suspension, 80 μL, and the linearised DNA were mixed gently and left to incubate in an electroporation cuvette on ice for 5 min. The cells were electroporated (30 s, 25 μF, 200 Ω, 2000 V) and then, after the addition of 1 ml ice-cold sorbitol the cells were left to incubate at 30$^o$C for 2 hours. The cells were then plated out on YPDS plates containing 100 μg/ml Zeocin and incubated at 30$^o$C until colonies appeared.

## 4.4. Results and Discussion

### 4.4.1. Expression of the L1 Region in *E. coli*

The L1 region of the IR is one of the main parts of the IR involved in insulin binding. Rosetta-Gami 2 cells were used initially to produce the L1 protein and were grown and induced as described in the appendix.
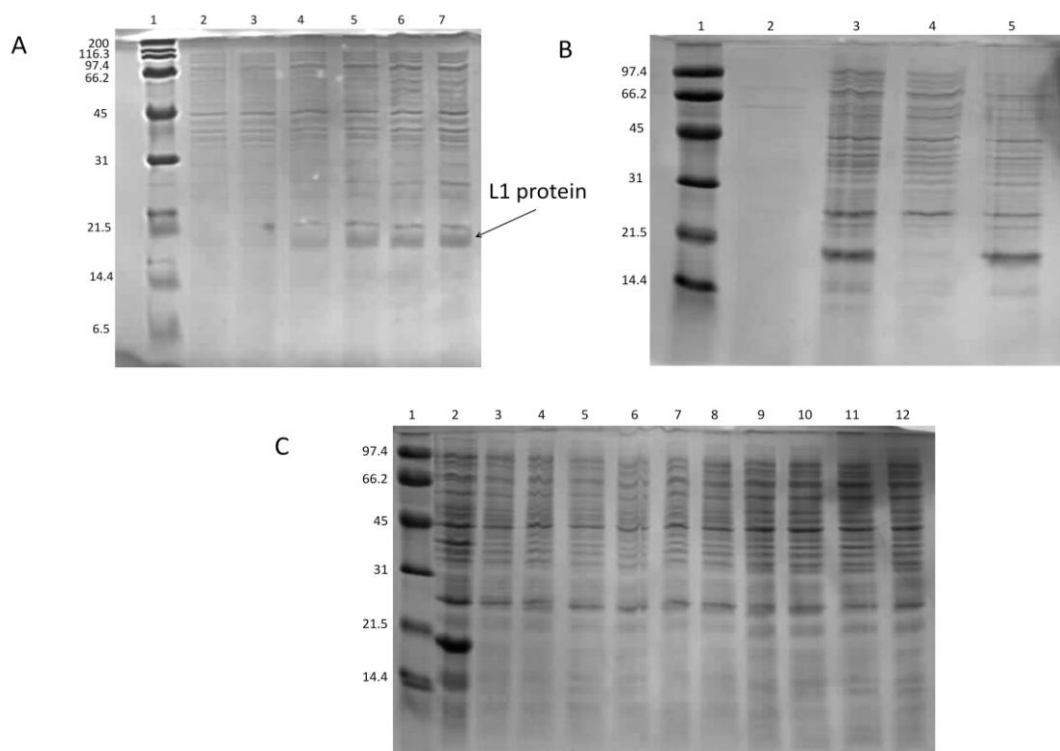


***Figure 44. A time course expression trial and solubility test of L1 expressed in Rosetta-Gami 2.***

(A) A 15% reduced SDS-PAGE of a time course expression test of L1 in Rosetta-Gami 2. Lane 1: Broad range molecular weight marker. Lane 2: 0.5 h post-induction sample. Lane 3: 1 h post-induction sample. Lane 4: 2 h post-induction sample. Lane 5: 4 h post-induction sample. Lane 6: 6 h post-induction sample. Lane 7: overnight sample. (B) A 15% SDS-PAGE of a solubility test of L1 produced in Rosetta-Gami 2. Lane 1: Low range molecular weight marker. Lane 2: Media sample. Lane 3: Total sample. Lane 4: Soluble fraction. Lane 5: Insoluble fraction. (C) A 15% SDS-PAGE of a solubility screen using L1 produced in Rosetta-Gami 2. Lane 1: Low range molecular weight marker. Lane 2: Control sample. Lane 3-12: Soluble fractions from buffers used in buffer trial.

The expression of the L1 protein was monitored at various intervals throughout an overnight expression, and analysed by SDS-PAGE (Figure 44-A). The L1 protein with the His$_6$ tag, has a molecular weight of 20.191 kDa. The presence of bands on the gel around the expected molecular weight indicates that the protein is being produced post induction, although the weakness of the bands suggests that the protein is only being expressed in low

levels. The target protein was not present in the soluble fraction (Figure 44-B). This could be a result of an unsuitable re-suspension buffer. A buffer screen has been designed, by Lindwall et al., that utilises a range of reagents and conditions in an attempt to improve protein solubility [176]. Out of the 30 conditions tried, none resulted in the L1 region becoming soluble (Figure 44-C). Due to the difficulties in expressing soluble L1 protein in Rosetta-Gami 2 cells, alternative methods to improve solubility were investigated.

There were various methods that would increase protein solubility. Three methods were utilised within this study. The first technique, the Eco construct, utilised the pMAL-p2 plasmid to fuse the L1 protein to a maltose binding protein. The soluble maltose binding protein should result in a more soluble L1 construct. The maltose binding protein is purified using an amylose column and is engineered to be cleavable (this construct has a molecular weight of ~64 kDa). Two more constructs of the short L1 region construct, termed Nco and Xho, utilised the pET22 plasmid to add a pelB signal sequence to the L1 protein (molecular weight of 21.5 and 20.0 kDa respectively). pelB is responsible for translocation of the protein to the periplasmic space, where the environment is much more reducing and so more favourable to the formation of disulphide bonds. The L1 protein is still being produced in small amounts and is present in the insoluble fraction (Figure 45-A).
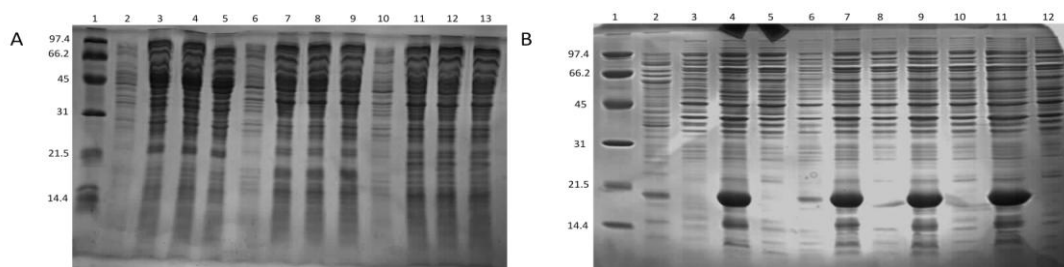


***Figure 45. A solubility trial of the L1 region of various constructs and different cell strains.***

(A) A 15% reduced SDS-PAGE of a solubility check of Eco, Nco and Xho constructs. Lane 1: Low range molecular weight marker. Lanes 2-13 contain soluble samples. Lane 2: Uninduced Eco sample. Lane 3: Eco induced with 0.2 mM IPTG. Lane 4: Eco induced with 0.5 mM IPTG. Lane 5: Eco induced with 1 mM IPTG. Lane 6: Uninduced Nco sample. Lane 7: Nco induced with 0.2 mM IPTG. Lane 8: Nco induced with 0.5 mM IPTG. Lane 9: Nco induced with 1 mM IPTG. Lane 10: Uninduced Xho sample. Lane 11: Xho induced with 0.2 mM IPTG. Lane 12: Xho induced with 0.5 mM IPTG. Lane 13: Xho induced with 1 mM IPTG. (B) A 15% SDS-PAGE of a solubility test of L1 produced in TUNER and BL21 cells. Lane 1: Low range molecular weight marker. Lane 2: Reference sample. Lane 3: BL21 uninduced. Lane 4: BL21 total sample. Lane 5: BL21 soluble sample. Lane 6: TUNER uninduced. Lane 7: Total sample of TUNER induced with 1 mM IPTG. Lane 8: Soluble sample of TUNER induced with 1 mM IPTG. Lane 9: Total sample of TUNER induced with 0.5 mM IPTG. Lane 10: Soluble sample of TUNER induced with 0.5 mM IPTG. Lane 11: Total sample of TUNER induced with 0.2 mM IPTG. Lane 12: Soluble sample of TUNER induced with 0.2 mM IPTG.

TUNER cells are a strain of *E. coli* that has been modified to be "tuneable" with respect to the amount of target protein produced in response to induction by IPTG. The wide bands seen for the TUNER cells indicate that markedly more protein has been expressed (Figure 45-B). An increased level of over-expression was achieved, although the protein was still insoluble.  Despite this, the ability to produce L1 in increased amounts meant that an *in vitro* method of solubilisation could be investigated.

It is possible to purify and re-fold target proteins using a nickel affinity column[174]. This method binds denatured protein to the Ni column using the target's $His_6$ tag.  While non-target proteins can also bind to the target protein via non-specific interactions, various solutions can be applied to the column to remove these unwanted proteins. Varying salt solutions are used to refold the protein, before elution over an imidazole gradient (Figure 46-A). The results of the SDS-PAGE of fractions across the elution peak indicate that the target protein has been purified to a relatively high level and is also soluble (Figure 46-B).
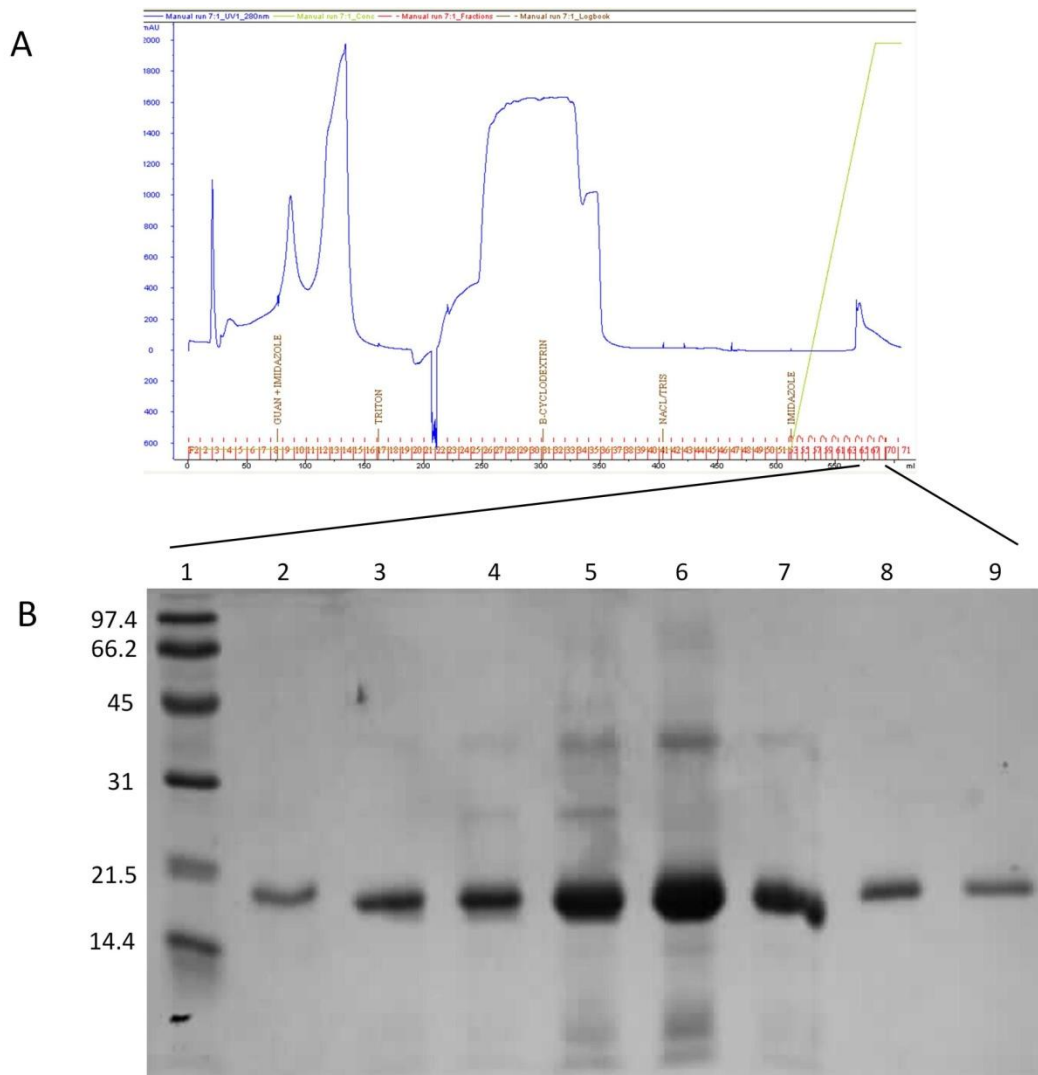
***Figure 46. Analysis of on-column (Ni²⁺ affinity) refolding purification of the L1 region.***

(A) The blue trace indicates a UV 280nm reading and the green trace shows the percentage imidazole concentration of the eluting solution. The labels on the x-axis show the points at which the new solutions were added. (B) A 15% reduced SDS-PAGE of the analysis of the fractions from the elution peak in Figure 46-A. Lane 1: Low range molecular weight marker. Lanes 2-9: Fractions across the elution peak in Figure 46-A.
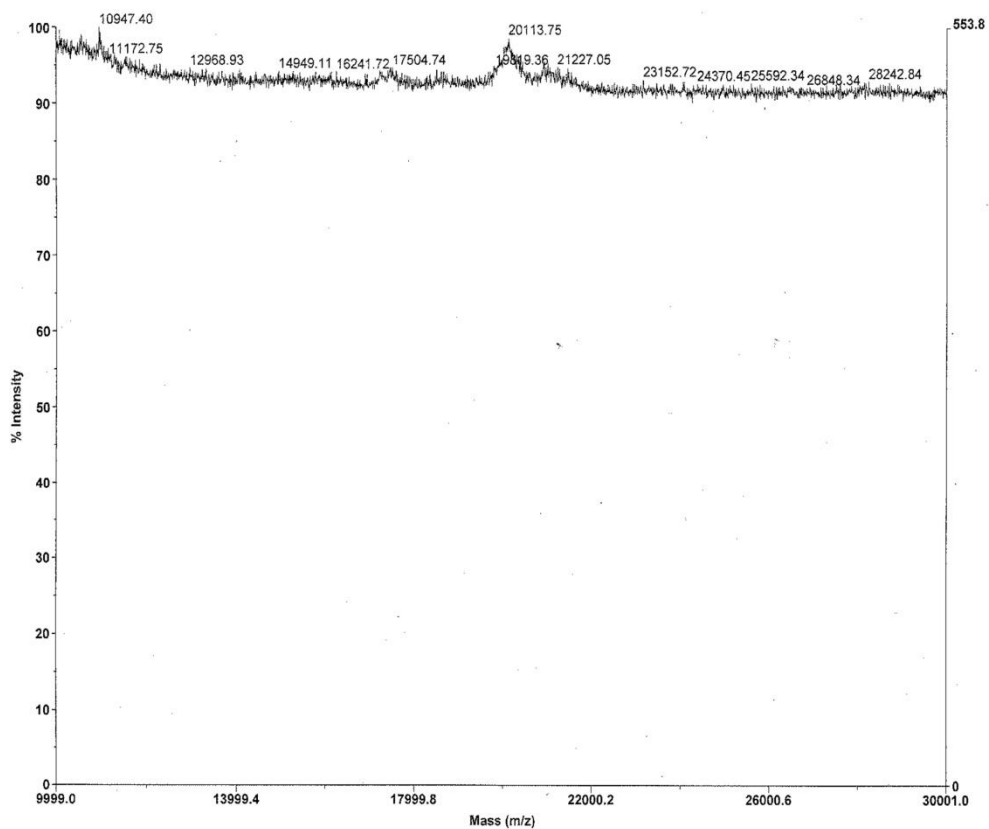
*Figure 47. MALDI of purified L1 protein.*

A mass spectrometry (MALDI) analysis of the L1 sample was performed to check that the protein produced was correct (Figure 47). The mass of the L1 region expressed is 20.191 kDa, where the observed mass from MALDI is 20.113 kDa; a difference of 80 Da which is not an issue as the L1 mass is theoretical and does not account for pH effects of the buffer or the harsh purification method used.

The L1 protein was expressed to form complexes with a range of insulin analogues, and allow determination of the crystal structure. To test binding, L1 was mixed with various insulin analogues and WT insulin in a 1:5 ratio of L1:insulin, and after being incubated at room temperature for 3 hours, the samples were run on a native gel (Figure 48).
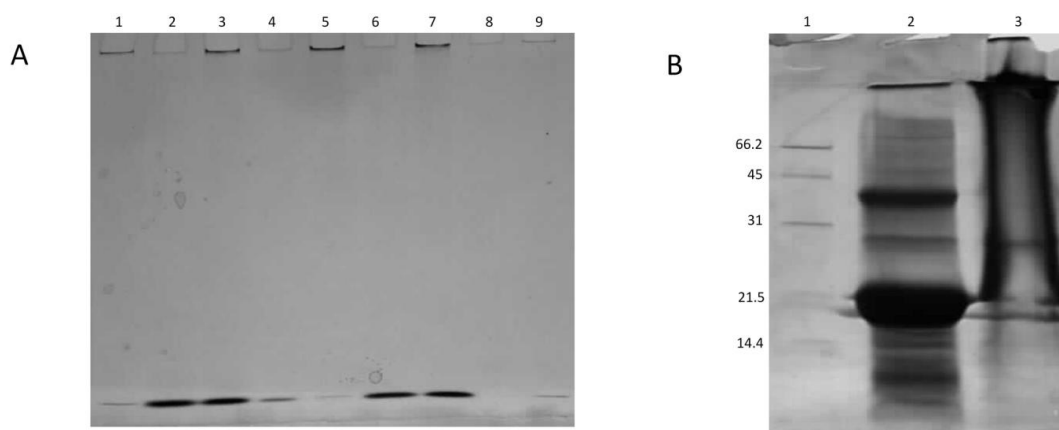
***Figure 48. Complexation trials of various insulin analogues and WT insulin each with the L1 region expressed in TUNER cells, and an investigation of disulphide bond formation in the L1 region.***

(A) A 7.5% native gel of the complexations each of LZ30, 33 and 45 insulin analogues and WT insulin with the TUNER L1 receptor.  Lane 1: IR.  Lane 2:  LZ30.  Lane 3: LZ30 and IR.  Lane 4: LZ33.  Lane 5: LZ33 and IR.  Lane 6: LZ45.  Lane 7: LZ45 and IR  Lane 8: WT insulin.  Lane 9: WT insulin and IR. (B) A 15% SDS-PAGE of the TUNER L1 disulphide formation.  Lane 1: Low range molecular weight marker.  Lane 2: Redox refolded material in a reducing buffer.  Lane 3: Redox refolded material in a non-reducing buffer.

The L1 protein precipitated out of solution and was unable to enter the gel (Figure 48-A). To check the protein, an SDS-PAGE was run with a sample of the protein in both reducing and non-reducing sample buffers.  The protein only appears as a normal band in the presence of the reducing sample buffer (Figure 48-B).  The smearing of the sample in the non-reducing sample buffer suggests that the disulphide bonds formed incorrectly and had cross-linked with other protein molecules.

Although the disulphide bonds were incorrectly formed due to the relatively large amount of soluble L1 protein, a re-folding experiment could be attempted.  A redox method of re-folding other proteins that is based on redox potentials has been described by Walter *et al.*[175].  A sample of the on-column refolded material was subjected to this method (Figure 49-B).
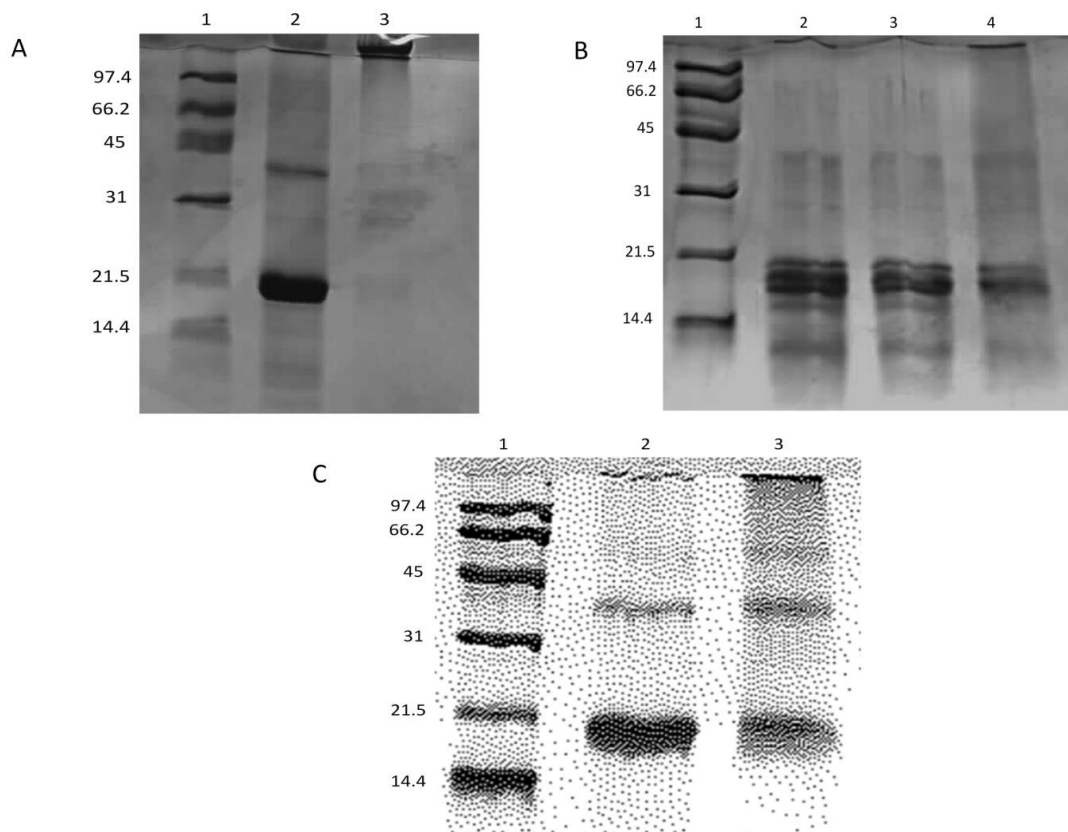
***Figure 49. Native gel electrophoresis investigating the correct disulphide bond formation in redox refolded L1 protein.***

(A) A 15% SDS-PAGE of redox refolded L1 material disulphide. Lane 1: Low range molecular weight marker. Lane 2: Redox refolded material in a reducing buffer. Lane 3: Redox refolded material in a non-reducing buffer. (B) A 15% SDS PAGE of 100X redox refolded L1 material disulphide. Lane 1: Low range molecular weight marker. Lane 2: Control sample prior to redox re-folding in reducing sample buffer. Lane 3: Sample of L1 in a reducing sample buffer. Lane 4: Sample in a non-reducing buffer. (C) A 15% SDS-PAGE of a disulphide check of fresh TUNER L1 material. Lane 1: Low range molecular weight marker. Lane 2: Sample in a reducing sample buffer. Lane 3: Sample in a non-reducing buffer.

The refolded material was not formed correctly and thus unable to enter the gel (Figure 49-A). A small band is present in the third lane, at approximately 30 kDa, indicating that a small fraction of the material was correctly refolded. A higher level of dilution (100x) was attempted, which resulted in the majority of the sample successfully refolding (Figure 49-B). As the solubilisation and refolding of the L1 protein took place over several weeks, a fresh preparation of L1 protein was made. This preparation followed the on-column refolding protocol before being passed through a gel filtration column, followed by the redox protocol. There appears to be a large proportion of protein in the non-reducing sample buffer compared to that in the reducing buffer (Figure 49-C). The preparation also seems to be relatively free from contamination. A mass spectrometry

analysis of this sample indicated that the protein matched the predicted molecular weight of the L1 protein. When the protein was used to complex with insulin analogues, it was unable to enter the native gel.

The over-expression of a protein can lead to it being insoluble and targeted towards inclusion bodies. It has been shown that if protein synthesis is halted, the equilibrium between protein solubilisation and aggregation can be shifted towards protein solubilisation. Protein synthesis can be halted by the addition of chloramphenicol[177]. This was attempted using BL21 and TUNER cells (Figure 50), although no soluble protein was obtained.
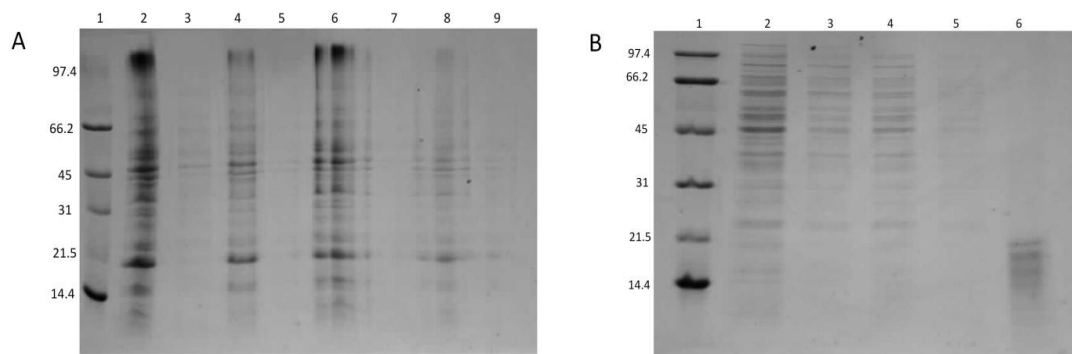


***Figure 50. SDS-PAGE of chloramphenicol shock-based expression method to induce re-folding of the L1 protein.***

(A) A 15% reducing SDS-PAGE of the chloramphenicol shock upon BL21 and TUNER cells. Lane 1: Low range molecular weight marker. Lane 2: BL21 control total sample. Lane 3: BL21 control soluble sample. Lane 4: BL21 induced total sample. Lane 5: BL21 induced soluble sample. Lane 6: TUNER control total sample. Lane 7: TUNER control soluble sample. Lane 8: TUNER induced soluble sample. Lane 9: TUNER induced soluble sample. (B) A 15% SDS-PAGE of the chloramphenicol shock upon BL21 and TUNER cells. Lane 1: Low range molecular weight marker. Lane 2: BL21 control soluble sample. Lane 3: BL21 induced soluble sample. Lane 4: TUNER control soluble sample. Lane 5: TUNER induced soluble sample. Lane 6: Control sample from *E. coli*.

Lemo21 and ArcticExpress were used to try and express soluble L1 protein with the correct disulphide bonds. It is apparent that for Lemo2,1 there is an over-expressed soluble protein of a similar weight to the expected molecular weight of the L1 protein. ArcticExpress appears to have one clone that has over-expressed a correctly sized insoluble protein, whilst another clone has also over-expressed a correctly sized partially-soluble protein (Figure 51).
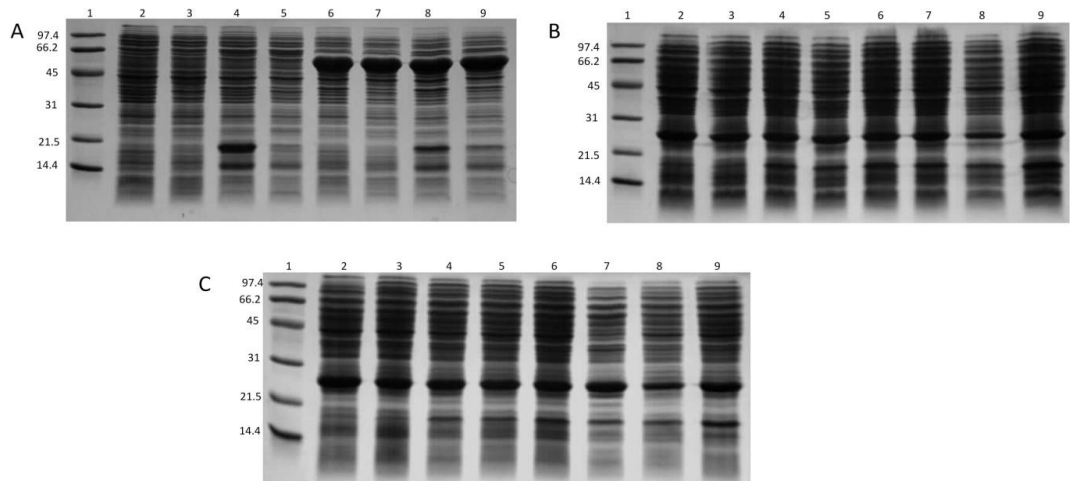
***Figure 51. Solubility trial of the L1 protein expressed in ArcticExpress and Lemo21 cells.***

A 15 % reducing SDS-PAGE of the solubility trail undertaken in ArcticExpress and Lemo21. Gel (A) ArcticExpress, Gel (B) Lemo21 total samples and Gel (C) Lemo21 soluble samples. Gel (A) Lane 1: Low range molecular weight marker. Lane 2: S1 uninduced total sample. Lane 3: S1 uninduced soluble sample. Lane 4: S1 induced total sample. Lane 5: S1 induced soluble sample. Lane 6: L1 uninduced total sample. Lane 7: L1 uninduced soluble sample. Lane 8: L1 induced total sample. Lane 9: induced soluble sample. Gel (B) Lane 1: Low range molecular weight marker. Lane 2: uninduced sample. Lane3: 0 μM Rhamnose. Lane 4: 100 μM Rhamnose. Lane 5: 250 μM Rhamnose. Lane 6: 500 μM Rhamnose. Lane 7: 750 μM Rhamnose. Lane 8: 1000 μM Rhamnose. Lane 9: 2000 μM Rhamnose. Gel (C) Lane 1: Low range molecular weight marker. Lane 2: uninduced sample. Lane 3: 0 μM Rhamnose. Lane 4: 100 μM Rhamnose. Lane 5: 250 μM Rhamnose. Lane 6: 500 μM Rhamnose. Lane 7: 750 μM Rhamnose. Lane 8: 1000 μM Rhamnose. Lane 9: 2000 μM Rhamnose.
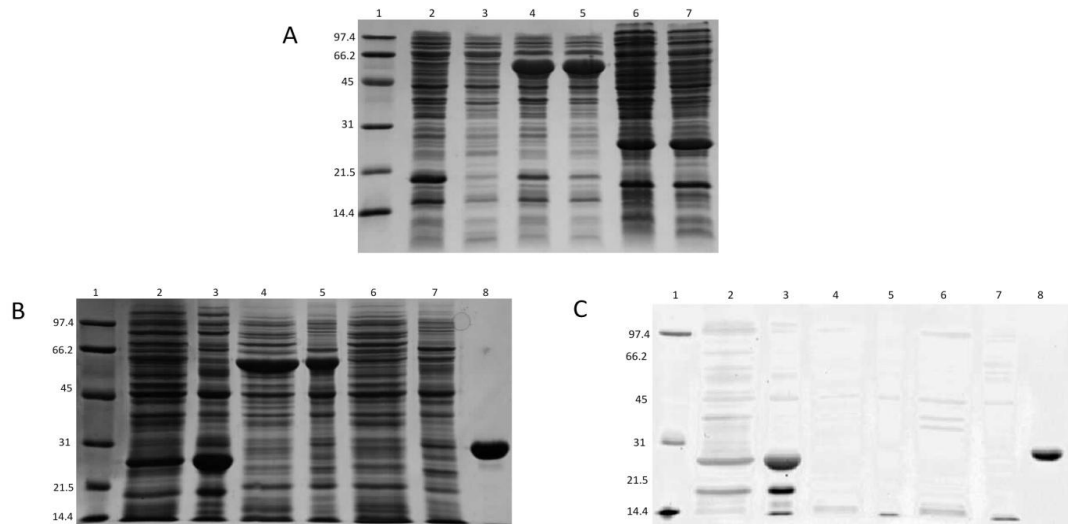
***Figure 52. Comparison of the best results from the solubility of L1 domain expressed in ArcticExpress and Lemo21, using comassie and His$_6$-tag stains.***

(A) A 15% reducing SDS-PAGE of the solubility of ArcticExpress and Lemo21. Lane 1: Low range molecular weight marker. Lane 2: ArcticExpress S1 total sample. Lane 3: ArcticExpress S1 soluble sample. Lane 4: ArcticExpress L1 total sample. Lane 5: ArcticExpress L1 soluble sample. Lane 6: Lemo21 2000 μM Rhamnose total sample. Lane 7: Lemo21 2000 μM Rhamnose soluble sample. (B and C) A 12% reducing SDS PAGE of the best results from the solubility of ArcticExpress and Lemo21 expressed before and after Ni$^{2+}$ column. Gel (B) is comassie stained, and Gel (C) is His$_6$-tag stained, both gels have the same layout. Lane 1: Low range molecular weight marker. Lane 2: Lemo21 soluble load. Lane 3: Lemo21 elution fraction. Lane 4: ArcticExpress L1 soluble load. Lane 5: ArcticExpress L1 elution fraction. Lane 6: ArcticExpress S1 soluble load. Lane 7: ArcticExpress S1 elution fraction. Lane 8: 25 kDa positive control.

The total and soluble fractions of both the Lemo21 and ArcticExpress cells were assessed (Figure 52-A) to confirm the presence of a His$_6$-tag on the target protein using a His$_6$-tag stain. It appears that both of the bands in Lemo21 give a positive response (Figure 52-B&C).

Both Lemo21 colonies, along with one from ArcticExpress were grown in 50 ml volumes and subjected to a Ni$^{2+}$ affinity chromatography purification. The cells were lysed in 125 mM NaCl, 100 mM HEPES pH 7.4, and only the soluble fraction loaded onto the column and purified (Figure 53).
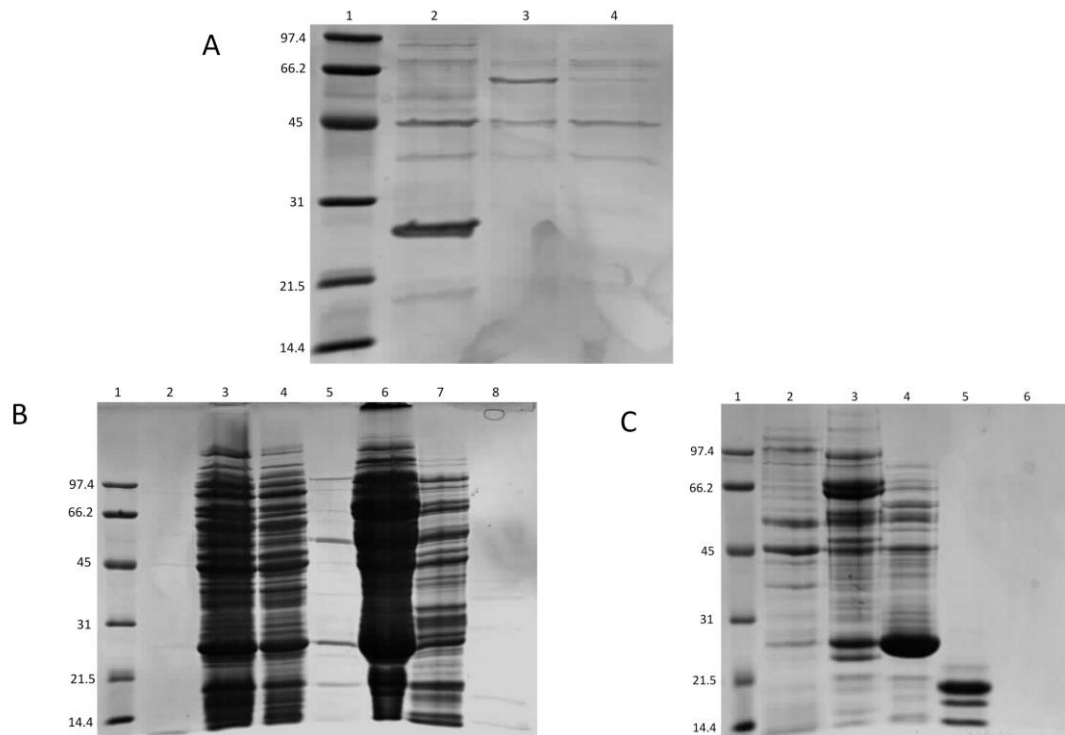
***Figure 53. Elution fractions from small scale Ni²⁺ purification from ArcticExpress and Lemo21 cells expression trials.***

(A) A 12% reducing SDS-PAGE of the elution fractions from ArcticExpress and Lemo21 $Ni^{2+}$ purification. Lane 1: Low range molecular weight marker. Lane 2: Lemo21 elution fraction. Lane 3: ArcticExpress L1 elution fraction. Lane 4: ArcticExpress S1 elution fraction. (B and C) A 12% reducing SDS PAGE of the purification of L1 from Lemo21. (B) $Ni^{2+}$ purification of L1 from Lemo21. Lane 1: Low range molecular weight marker. Lane 2: Pre-induction total sample. Lane 3: Soluble load. Lanes 4-8: elution fractions. (C) Gel filtration purification of S1 from Lemo21. Lane 1: Low range molecular weight marker. Lanes 2-7: elution fractions.

The best approach to obtain soluble protein would be Lemo21, as neither of the ArcticExpress clones have the correct molecular weight for L1 protein (Figure 53-A). The major protein to have bound to the column is at 25 kDa (lane 2), whilst the band at approximately 20 kDa has bound at a much lower level. The Lemo21 construct was grown in a 2 L culture, then purified by $Ni^{2+}$ affinity followed by gel filtration (Figure 53-B&C).

Both proteins that potentially correspond to L1 as identified by the $His_6$-tag stain bound to the $Ni^{2+}$ column and were separated by gel filtration. A sample of each was mixed with the high affinity analogues LZ29 and LZ30 to ascertain whether a complex could be formed. From the native gel (Figure 54-A) it appears that the larger protein binds to the analogue. To check that this result was not an artefact of the high affinity analogue, a second complexation was set up using WT insulin (Figure 54-B), to which the protein bound. The two bands were sent for protein identification (trypsin digest of the protein sample followed by LC-MS with the results searched against a database) and the results

112

showed that the larger protein was a chloramphenicol acetyltransferase mutant, whilst the smaller protein was lysozyme.
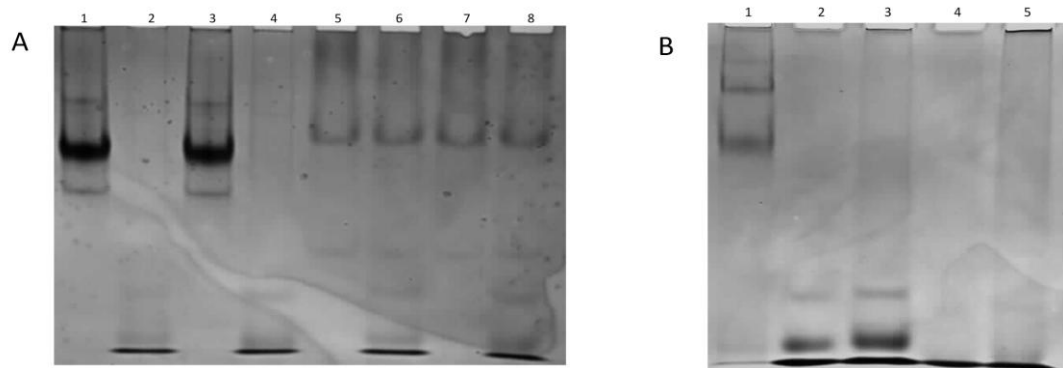


***Figure 54.  Native gel of insulin-complex formation trials of Lemo21 produced L1 material and various insulin analogues.***

(A) A 10 % native gel of the complexation of both Lemo21 peaks with LZ29 and LZ30. Lane 1: Higher peak.  Lane 2: Higher peak and LZ29.  Lane 3: Higher peak.  Lane 4: Higher peak and LZ30.  Lane 5: Lower peak.  Lane 6: Lower peak and LZ29.  Lane 7: Lower peak.  Lane 8: Lower peak and LZ30. (B) A 10% native gel of the complexation between the higher Lemo21 peak with WT insulin and LZ30.  Lane 1: Higher peak.  Lane 2: WT insulin.  Lane 3: Higher peak and WT insulin.  Lane 4: LZ30.  Lane 5: Higher peak and LZ30.

The clone of ArcticExpress, S1, that gave the larger expression of the protein at approximately 20 kDa was grown and subjected to a 30 condition solubility screen [176] (Table 13 Figure 55).

***Table 13. Thirty Reagents for Solubilisation of Over-Expressed Proteins[176]***

| Condition | Contents |
|---|---|
| 1. | 100 mM Tris, 10% glycerol, pH 7.6 |
| 2. | 100 mM Tris, 50 mM LiCl, pH 7.6 |
| 3. | 100 mM HEPES, 50 mM $(NH_4)_2SO_4$, 10% glycerol, pH 7.0 |
| 4. | 100 mM HEPES, 100 mM KCl, pH 7.0 |
| 5. | 100 mM Tris, 50 mM NaCl, 10% isopropanol, pH 8.2 |
| 6. | 100 mM $K_2HPO_4/KH_2PO_4$, 50 mM $(NH_4)_2SO_4$, 1% Triton X-100, pH 6.0 |
| 7. | 100 mM triethanolamine, 100 mM KCl, 10 mM DTT, pH 8.5 |
| 8. | 100 mM Tris, 100 mM sodium glutamate, 10 mM DTT, pH 8.2 |
| 9. | 250 mM $KH_2PO_4/K_2HPO_4$, 0.1% CHAPS, pH 6.0 |
| 10. | 100 mM triethanolamine, 50 mM LiCl, 5 mM EDTA, pH 8.5 |
| 11. | 100 mM sodium acetate, 100 mM glutamine, 10 mM DTT, pH 5.5 |
| 12. | 100 mM sodium acetate, 100 mM KCl, 0.1% $n$-octyl-β-D-glucoside, pH 5.5 |
| 13. | 100 mM HEPES, 1 M $MgSO_4$, pH 7.0 |
| 14. | 100 mM HEPES, 50 mM LiCl, 0.1% CHAPS, pH 7.0 |
| 15. | 100 mM $KH_2PO_4/K_2HPO_4$, 2.5 mM $ZnCl_2$, pH 4.3 |
| 16. | 100 mM Tris, 50 mM NaCl, 5 mM calcium acetate, pH 7.6 |
| 17. | 100 mM triethanolamine, 50 mM $(NH_4)_2SO_4$, 10 mM $MgSO_4$, pH 8.5 |
| 18. | 100 mM Tris, 100 mM KCl, 2 mM EDTA, 1% Triton X-100, pH 8.2 |
| 19. | 100 mM sodium acetate, 1M $MgSO_4$, pH 5.5 |
| 20. | 100 mM Tris, 2M NaCl, 0.1% $n$-octyl-β-D-glucoside, pH 7.6 |
| 21. | 100 mM Tris, 1 M $(NH_4)_2SO_4$, 10 mM DTT, pH 8.2 |
| 22. | 100 mM sodium acetate, 50 mM LiCl, 5 mM calcium acetate, pH 5.5 |
| 23. | 100 mM HEPES, 100 mM sodium glutamate, 5 mM DTT, pH 7.0 |
| 24. | 100 mM triethanolamine, 100 mM sodium glutamate, 0.02% $n$-octyl-β-D-glucoside, 10% glycerol, pH 8.5 |
| 25. | 100 mM Tris, 50 mM NaCl, 100 mM urea, pH 8.2 |
| 26. | 100 mM triethanolamine, 100 mM KCl, 0.05% dextran sulfate, pH 8.5 |
| 27. | 100 mM $KH_2PO_4/K_2HPO_4$, 50 mM $(NH_4)_2SO_4$, 0.05% dextran sulfate, pH 6.0 |
| 28. | 100 mM HEPES, 50 mM LiCl, 0.1% deoxycholate, pH 7.0 |
| 29. | 100 mM Tris, 100 mM KCl, 0.1% deoxycholate, 25% glycerol, pH 7.6 |
| 30. | 100 mM potassium acetate, 50 mM NaCl, 0.05% dextran sulfate, 0.1% CHAPS, pH 5.5 |
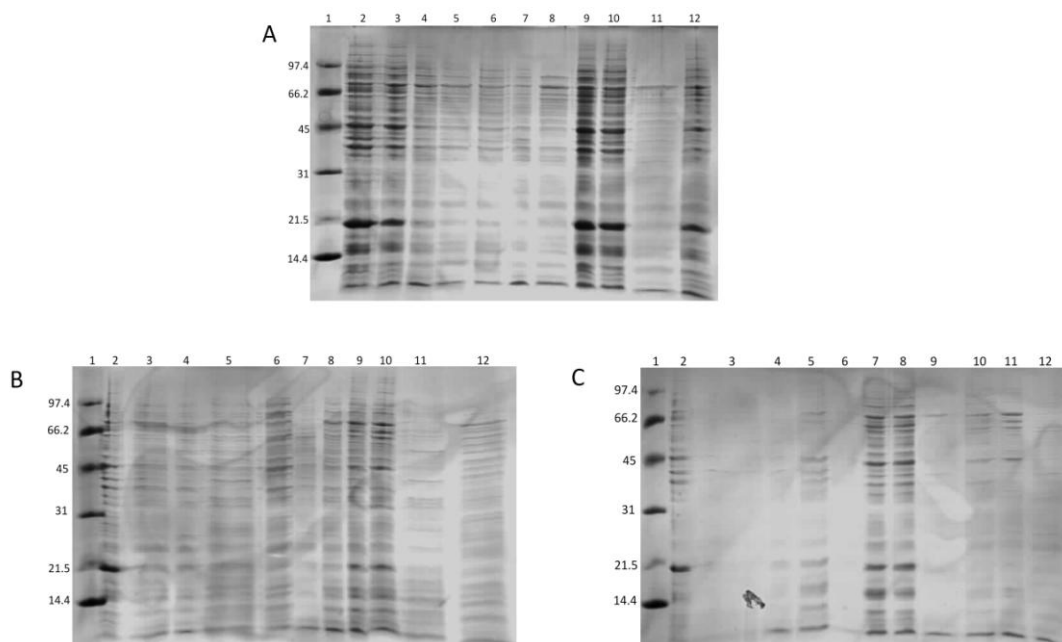
***Figure 55. Solubility screen for the L1 domain expressed in ArcticExpress cells.***

A 15 % reducing SDS-PAGE of the ArcticExpress solubility screen. Gel (A) Lane 1: Low range molecular weight marker. Lane 2: Total ArcticExpress sample. Lanes 3-12: conditions 1-10. Lane 1: Low range molecular weight marker. Lane 2: Total ArcticExpress sample. Lanes 3-12: conditions 1-10. Gel (B) Lane 1: Low range molecular weight marker. Lane 2: Total ArcticExpress sample. Lanes 3-12: conditions 11-20. Gel (C) Lane 1: Low range molecular weight marker. Lane 2: Total ArcticExpress sample. Lanes 3-12: conditions 21-30.

Samples with the highest solubility were seen in conditions 1, 7, 8 and 10 (Table 13), although conditions 7 and 8 contain DTT which would reduce the disulphide bonds present in the protein. These samples were re-tested in the buffer without DTT, but also with EDTA, due to its presence in condition 10.
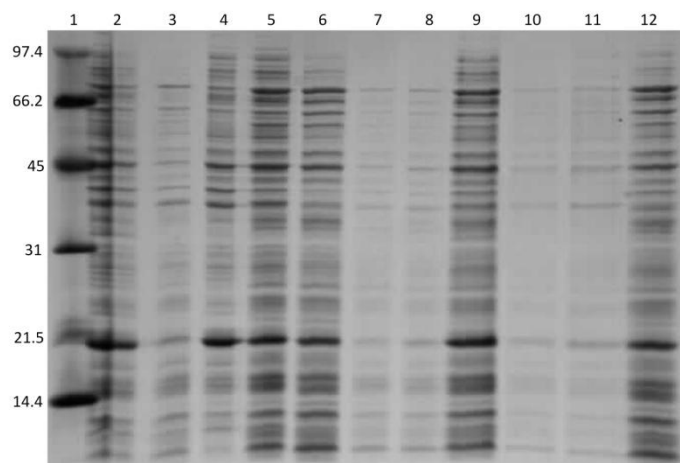
***Figure 56. Comparison of the best conditions achieved for expression of soluble L1 protein in ArcticExpress, and the test for DTT requirment in conditions 7 and 8.***

A 15% reducing SDS-PAGE exploring the solubility of L1 in condition 1, 7, 8 and 10. Lane 1: Low range molecular weight marker. Lane 2: Total sample. Lane 3: soluble sample. Lane 4: insoluble sample. Lane 5: condition 1 soluble fraction. Lane 6: condition 7 soluble fraction. Lane 7: condition 7 soluble fraction without DTT. Lane 8: condition 7 soluble fraction without DTT. Lane 9: condition 8 soluble fraction without DTT with EDTA. Lane 10: condition 8 soluble fraction without DTT. Lane 11: condition 8 soluble fraction without DTT with EDTA. Lane 12: condition 10 soluble fraction.

It appears that DTT is required in conditions 7 and 8 for soluble protein, however conditions 1 and 10 give comparable results to one another based upon solubility levels (Figure 56). Protein identification showed that the expressed protein was not L1.

## 4.4.2. Expression of the L1 Region in *S. cerevisiae*

Due to the problems encountered in expression of the L1 region in *E. coli*, *S. cerevisiae* was used to express L1 protein. The L1 gene was successfully cloned into the pYES2/CT plasmid (MW = ~25 kDa). Small scale (15 ml) culture volumes were grown with the cells lysed by digesting the cell walls, followed by homogenisation (see appendix). The soluble fraction was then passed over a Ni column (Figure 57-A).
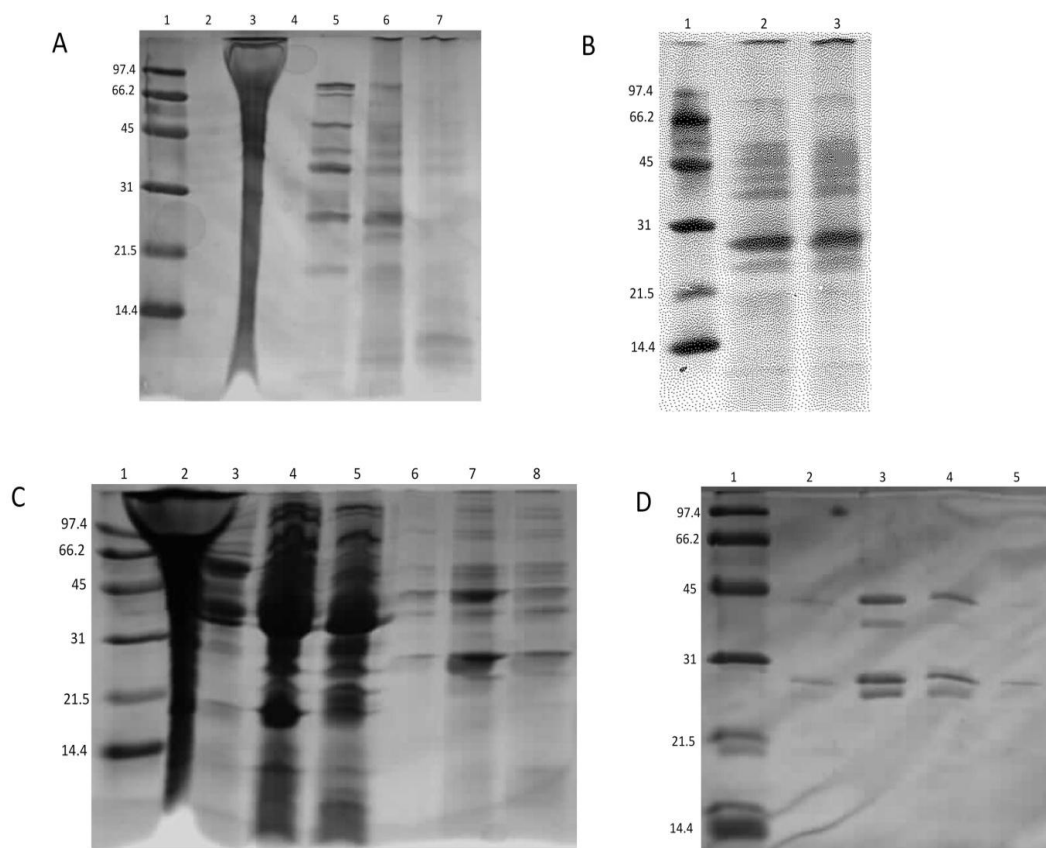


*Figure 57. Purification of the L1 region produced in S. cerevisiae.*

(A) A 15% reducing SDS-PAGE of the Ni$^{2+}$ purification of L1 in *S. cerevisiae*. Lane 1: Low range molecular weight marker. Lane 3: Run through sample. Lanes 5-7: Samples from the Ni column elution peak. (B) A 15% reducing SDS-PAGE of the disulphide check of the L1 region produced in *S. cerevisiae*. Lane 1: Low range molecular weight marker. Lane 2: Sample in a reducing sample buffer. Lane 3: Sample in a non-reducing buffer. (C) A 15% reducing SDS-PAGE of the Ni column purification of the L1 region produced in *S. cerevisiae*. Lane 1: Low range molecular weight marker. Lane 2: Run through sample. Lanes 3-5: Samples of non-specific binding washed off by wash buffer. Lanes 6-8: Samples from the Ni column elution peak. (D) A 15% reducing SDS-PAGE of gel filtration purification of the L1 region produced in *S. cerevisiae*. Lane 1: Low range molecular weight marker. Lanes 2-5: Samples from the gel filtration column elution peak.

There is a band present at ~25 kDa which could correspond to the predicted molecular weight of the L1 protein. Lanes 5 and 6 show a high level of contamination meaning further purification was required. The correct formation of the disulphide bonds was assessed by using both reducing and non-reducing sample buffer (Figure 57-B). The disulphide bonds may be correctly formed, as both the reducing and non-reducing samples ran identically on an SDS gel (Figure 57-B). A larger preparation of L1 expressed in *S. cerevisiae* was then prepared and purified by a $Ni^{2+}$ column (Figure 57-C) followed by gel filtration (Figure 57-D). There is a distinct double band present on the gel (Figure 57-D). A mass spectrometry analysis of this sample indicated the presence of a single species present at 25 kDa, which matched the predicted molecular mass of the L1 protein. When the protein was used in an attempt to form a complex with insulin analogues, it was unable to enter the gel. A fresh preparation of the protein was used for protein identification of these bands. The results indicated that neither corresponded to the L1 protein.

### 4.4.3. Expression of the L1 Region in *P. pastoris*

Following the lack of success of the *S. cerevisiae* studies, *P. pastoris* was used to express the L1 protein. The vector, pPICZαA, has a signal sequence that targets the protein for secretion to the media and has a C-terminal $His_6$ tag. A 3C protease site was added using the reverse primer to allow the removal of the $His_6$ tag. A long construct was trialled along with the short construct, (see appendix for primers and constructs). The plasmid was transformed into *P. pastoris* X-33 cells as per the method described in the appendix. Colonies were picked and grown in a 48 well plate containing BMGY media (Figure 58). The cultures were then used to produce induction/expression cultures of $OD_{600}=2$ at induction in BMMY.

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | L3 | L3 | S1 | L3 | L3 | S1 |
| 2 | L3 | L3 | S1 | L3 | L3 | S1 |
| 3 | L3 | L1 | S1 | L3 | L1 | S1 |
| 4 | L3 | L1 | S1 | L3 | L1 | S1 |
| 5 | L3 | S1 | Empty | L3 | S1 | Empty |
| 6 | L3 | S1 | Empty | L3 | S1 | Empty |
| 7 | L3 | S1 | Empty | L3 | S1 | Empty |
| 8 | L3 | S1 | Empty | L3 | S1 | Empty |

*Figure 58. P. pastoris 48 well expression layout.*

The wells in columns A, B and C were used to produce starter cultures from single colonies. The starter cultures were used to inoculate the induction media in the corresponding well in columns D, E and F to an $OD_{600}=2$.

Following 96 hours of induction, the cultures were spun down (12,000 x g, 5 min) and a sample of the supernatant was used for a dot blot experiment to test for the presence of a $His_6$ tag (Figure 59).
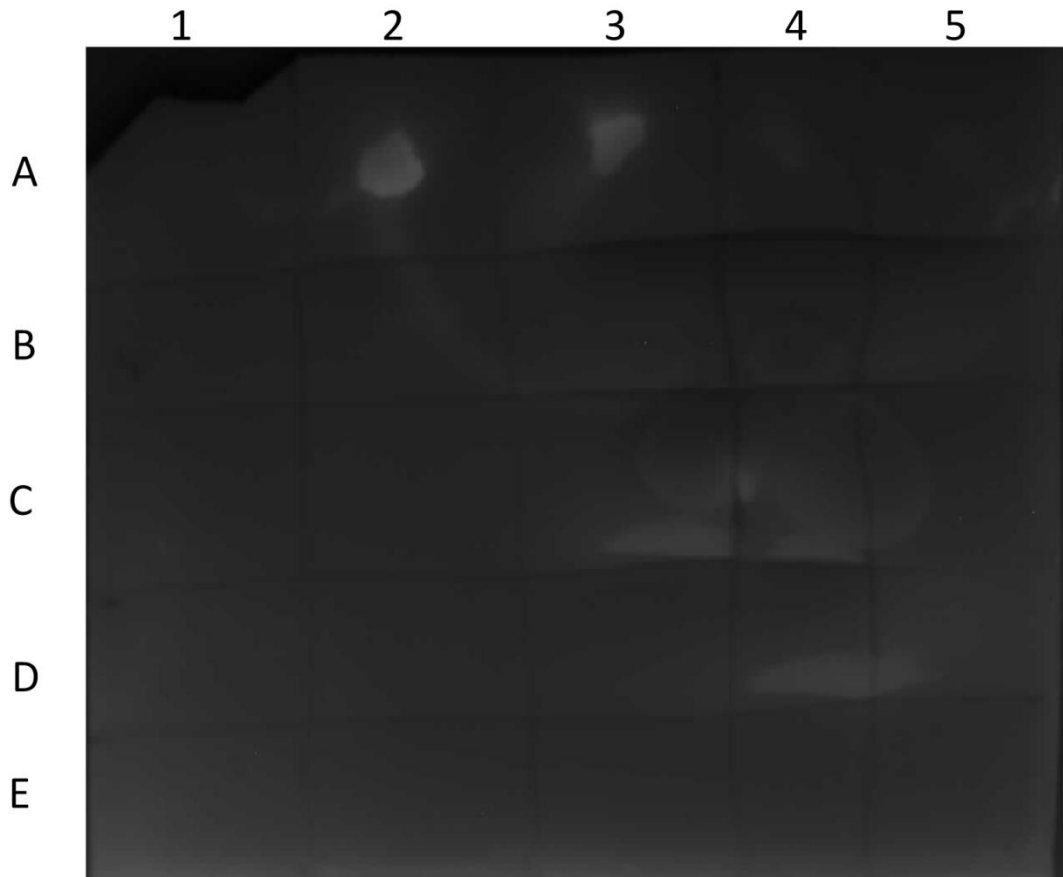
*Figure 59. A dot blot experiment to probe for expression of the L1 region in P. pastoris.*

A dot blot experiment of the expression trial for L1. Square A1: Uninduced Media. Square A2: 100 ng/μl $His_6$-tagged protein. Square A3: 50 ng/μl $His_6$-tagged protein. Square A4: 10 ng/μl $His_6$-tagged protein. Square A5: 1 ng/μl $His_6$-tagged protein. Square B1-E5 contain samples from Figure 58. Squares B1-B5: Samples D1-D5. Squares C1-C5: Samples D6-E2. Squares D1-D5: E3-E7. Squares E1-E5: E8-F4.

Positive results were seen for squares E1, E2, E7 and E8, with the strongest result being that of E7 so this clone was taken forward (Figure 59). A small scale expression trial (150 ml) was carried out on this clone (E7-L1) to produce, identify and purify the L1 protein.
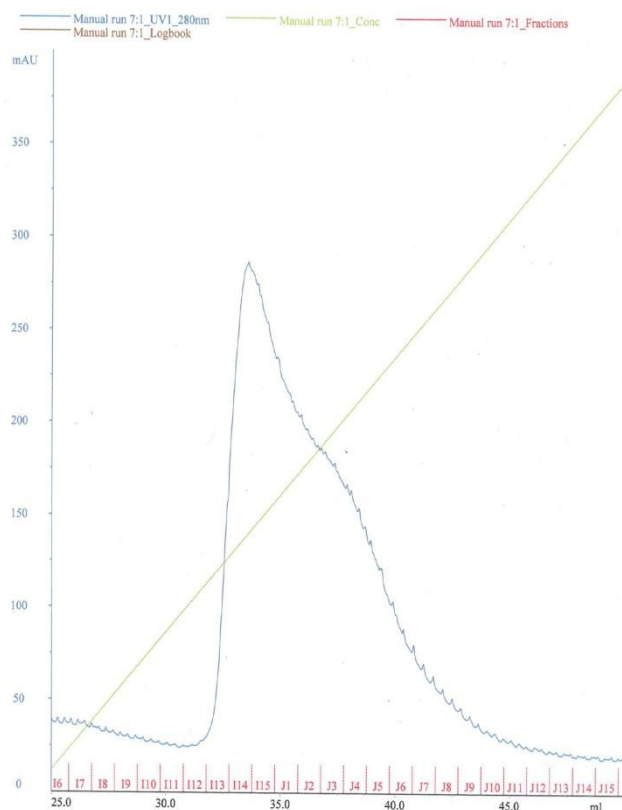
***Figure 60.  The Ni²⁺ purification trace for the small scale production of L1 from E7-L1.***

The blue trace is the $A_{280}$ reading and the green line is the percentage of elution buffer from 0% to 100%.


The expressed L1 protein bound to the Ni column and eluted upon application of the imidazole gradient (Figure 60).  Samples were run on an SDS-PAGE gel to ascertain whether the protein purified (gel not shown).  No bands were present, which could be due to the low levels of protein produced owing to the small volume of culture.  To increase the amount of material produced a 3 L expression was performed, followed by a Ni column purification.

The fractions across the elution peak (chromatograph not shown) were pooled and run on an SDS-PAGE gel to assess the success of the expression.  There were no bands present, which indicated that the protein was either at a concentration too low to be detected or was not present.  To increase sensitivity, a copper stain was applied to the gel and revealed bands of approximately 37 and 42 kDa (gel not shown). The bands could be the L1 protein, but this cannot be stated with confidence.  A dot blot of samples taken from various points across the purification was conducted to indicate if the bands detected were those of the L1 protein.
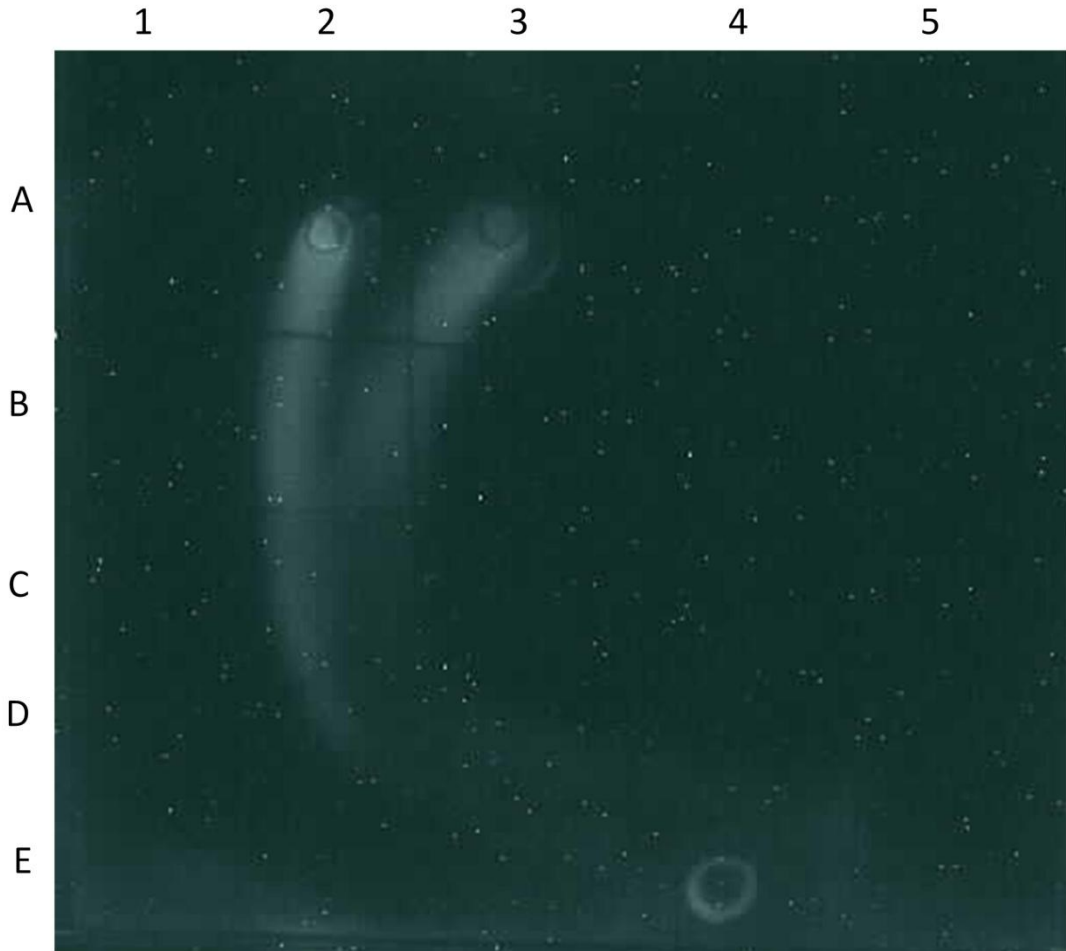
***Figure 61.  Dot blot analysis of L1 purification after expression in P. pastoris.***

A dot blot for the elution of L1 from a Ni column and gel filtration.  Square A1: $H_2O$.
Square A2: 100 ng/µl $His_6$-tagged protein.  Square A3: 10 ng/µl $His_6$-tagged protein.  Square
A4 and A5: Elution peak sample.  Square B1-B5: Elution peak sample. Square C1-C5:
Elution peak sample. Square D1-D4: Elution peak sample.  Square D5: Concentrated
elution peak 1.  Square E1: Concentrated elution peak 2.  Square E2: Concentrated gel
filtration peak 1.  Square E3: Concentrated gel filtration peak 2.  Square E4: Media sample.

The dot blot shows positive results for 1B and 1C which correspond to the positive
control, whilst also giving a positive result for E4, the media sample (Figure 61).  This
indicates that the L1 protein was produced although it did not bind to the column.  The
media sample was analysed by coomassie, copper and silver staining methods (Figure 62),
showing inconclusive results.
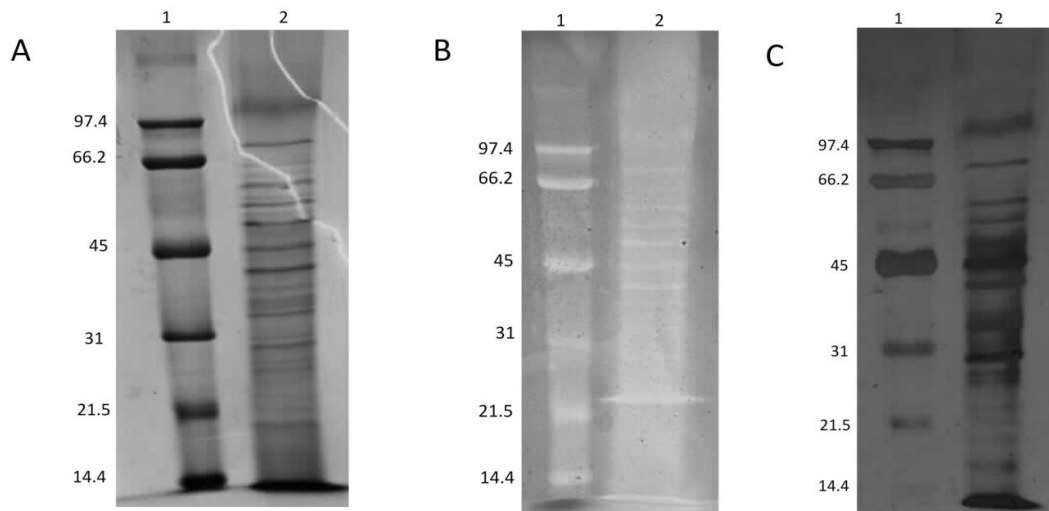
***Figure 62. A SDS-PAGE staining analysis of L1 protein expressed in P. pastoris.***

A 12 % reducing SDS-PAGE of the *P. pastoris* media following expression. Gel (A), stained with Coomassie, gel (B) copper stained and gel (C) silver stained. All gels; lane 1: Low range molecular weight marker. Lane 2: Media sample.

To ascertain whether the positive signal from the dot blot was of the L1 protein, a western blot was run.
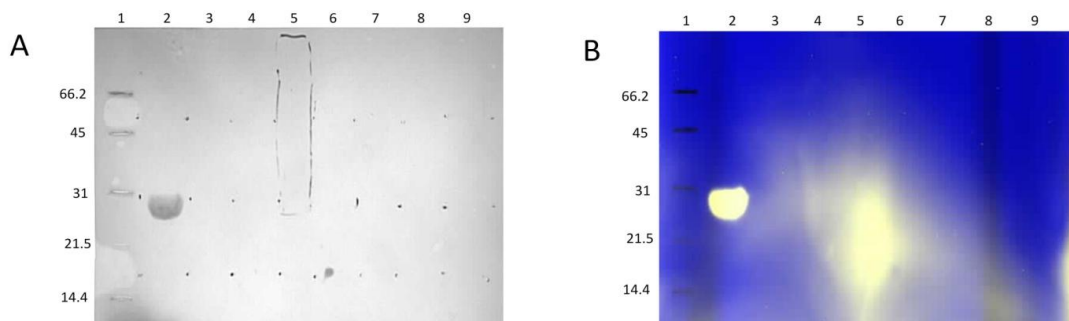


***Figure 63. A western blot of the purification of the L1 protein expressed in P. pastoris.***

A 12 % reducing SDS-PAGE western blot of the purification of L1. Membrane (A) shows the result following transfer and staining with Ponceau red, (B) shows the western blot results. (A) lane 1: Low range molecular weight marker. Lane 2: Positive control sample. Lane 3: Sample buffer only. Lane 4: Empty. Lane 5: Media sample. Lane 6: Media sample. Lane 7: Empty. Lane 8: Elution peak concentrated sample. Lane 9: Elution peak concentrated sample. (B) lane 1: Low range molecular weight marker. Lane 2: Positive control sample. Lane 3: Sample buffer only. Lane 4: Empty. Lane 5: Media sample. Lane 6: Media sample. Lane 7: Empty. Lane 8: Elution peak concentrated sample. Lane 9: Elution peak concentrated sample.

The western blot shows that there is a positive signal to the anti-His$_6$ tag antibody in the region of 18-29 kDa (Figure 63). The band is difficult to distinguish on the western blot due to the high intensity of the band and the smearing present. The predicted molecular weight of the L1 protein expressed in *P. pastoris* is ~23 kDa, which corresponds to the smeared band present. It was considered that the media or a short loading time could be the cause of weak binding to the Ni column. A 3 L culture was split four ways; one part was loaded onto a column overnight, whilst the others were dialysed overnight against 500 mM NaCl + 50 mM NaPO$_4$ pH 7.4, 500 mM NaCl + 100 mM HEPES pH 7.4 and 125 mM NaCl + 100 mM HEPES pH 7.4, before loading onto a Ni-affinity column.

There are peaks present for all samples on the chromatogram upon elution with imidazole (chromatograph not shown). Samples from all four peaks were concentrated ~10-fold and run on an SDS-PAGE gel.
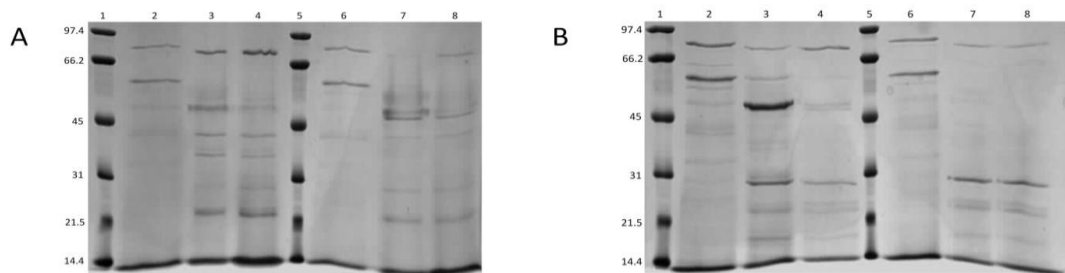


***Figure 64. Analysis of media loading change of L1 protein expressed in P. pastoris.***

A 12 % reducing SDS-PAGE of the purification trial of the L1 protein. Gel (A) lanes 2-4 samples from overnight load, and lanes 6-8 samples from 500 mM NaCl + 50 mM NaPO$_4$ pH 7.4 Gel (B) lanes 2-4 samples from 500 mM NaCl + 100 mM HEPES pH 7.4, and lanes 6-8 samples from 125 mM NaCl + 100 mM HEPES pH 7.4. Gel (A), lane 1: Low range molecular weight marker. Lane 2: Run through. Lane 3 and 4: Samples from elution peak. Lane 5: Low range molecular weight marker. Lane 6: Run through. Lane 7 and 8: Samples from elution peak. Gel (B), lane 1: Low range molecular weight marker. Lane 2: Run through. Lane 3 and 4: Samples from elution peak. Lane 5: Low range molecular weight marker. Lane 6: Run through. Lane 7 and 8: Samples from elution peak.

There is possibly L1 protein in each of the purifications, although it cannot be identified due to the presence of other bands (Figure 64).
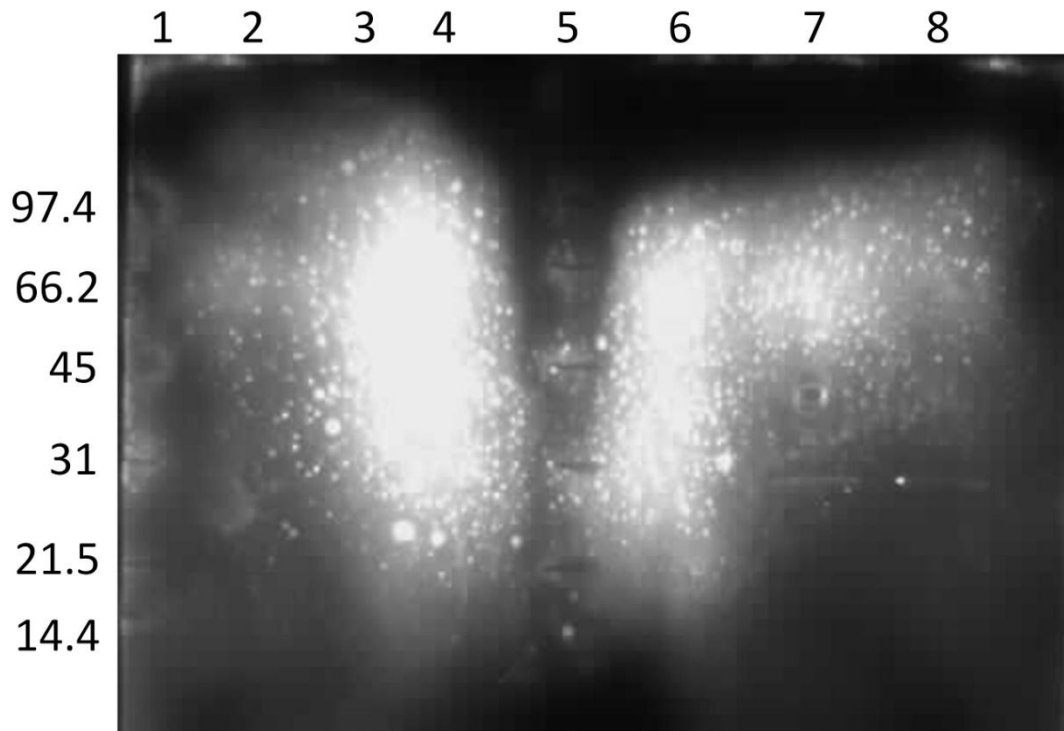
***Figure 65. Western blot analysis of media loading change following $Ni^{2+}$ purification of L1 expressed in P. pastoris.***

A 12 % reducing SDS-PAGE western blots of the purification trial of the L1 protein. Western blot lanes 2-4 samples from overnight load, and lanes 6-8 samples from 500 mM NaCl + 50 mM $NaPO_4$ pH 7.4. Lane 1: Low range molecular weight marker. Lane 2: Run through. Lane 3 and 4: Samples from elution peak. Lane 5: Low range molecular weight marker. Lane 6: Run through. Lane 7 and 8: Samples from elution peak.

The western blots show two bands present in lanes 7 and 8 (Figure 65) at approximately 30 kDa. Using Figure 65, the strong bands present in Figure 64 can be implied to be the L1 protein. A 3 L scale fermentation was followed by concentration of the media and dialysed against 125 mM NaCl and 100 mM HEPES pH 7.5, followed by an overnight load onto the Ni column to maximise binding. An SDS-PAGE and western blot of the purification were run (Figure 66).
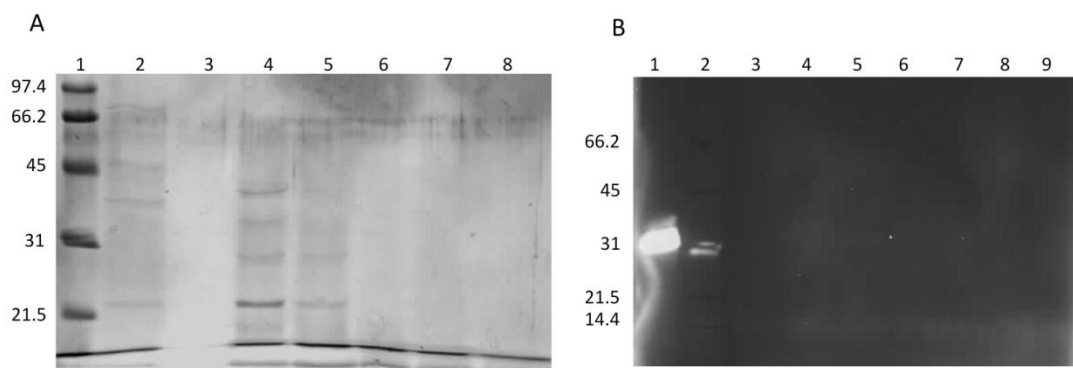
***Figure 66. Analysis of E7 L1 purification from P. pastoris following media loading change.***

A 15% reducing SDS PAGE of the purification and western blot of the L1 protein. Gel (A), lane 1: Low range molecular weight marker. Lane 2: Run through. Lanes 3-8: Samples from elution peak. Gel (B), lane 1: Positive control sample. Lane 2: Low range molecular weight marker. Lanes 4-9: Samples from elution peak.

The bands appear consistent in molecular weight to previous gels, and did not need to be concentrated to appear on the gel. The samples that gave a positive result on the western blot were sent for protein ID. The results identified beta-lactamase as a strong signal, whilst also having strong spectra that were unassigned. Beta-lactamase is an enzyme attributed to the breakdown of the antibiotic Zeocin, upon which the electroporated *P. Pastoris* cells were selected.

## 4.5. Conclusion

The expression of the L1 region in *E. coli* led to a good level of protein expression although it was insoluble. Following various attempts to express soluble protein, on-column refolding resulted in soluble protein. The soluble protein had incorrectly formed disulphide bonds, with intermolecular bonds rather than intramolecular ones. Attempts to express the protein in *S. cerevisae* and *P. pastoris* were unsuccessful.

The presence of *N*-glycosylations is important in ensuring protein solubility[178,179]. The L1 region has four predicted *N*-glycosylation sites at positions Asn[16], Asn [25], Asn[78] and Asn [111]. These *N*-glycosylations would not be present after expression in *E. coli*. All eukaryotic proteins are glycosylated to a certain level, with the process occurring through a complex series of steps, originating in the endoplasmic reticulum[180]. The solubility of the expressed L1 in *E. coli* following on-column refolding brings the requirement of glycosylation for soluble protein into question. The refolded material was soluble in a solution that comprised 20 mM Tris pH 7.5 and 0.1 M NaCl following on-column refolding, chemicals that are not shown to increase solubility. Expression in *S. cerevisae* and *P. pastoris* yielded no soluble protein indicating that the presence of *N*-glycans in this instance yielded no appreciable results.

Previous expression of the IR[50,63,171,172] was undertaken by using a variety of mammalian expression systems. This is preferred as the protein processing machinery possessed by these cells closely resembles that of human cells[181], as the particular glycoform displayed on the protein surface can influence its effect[181]. This method of producing the L1 region could result in soluble protein, though the production of soluble protein from on-column refolding questions this assumption.

The structure of L1 (Figure 43) shows a barrel with a series of β-sheets. At both the N- and C-terminus of the barrel there is a disulphide bond, which fixes the termini to the barrel. The correct formation of these disulphide bonds is therefore critical to a correctly folded and soluble protein. The re-solubilised expressed L1 protein in *E. coli* was shown to have incorrectly formed disulphide bonds, despite being soluble. It would be assumed that moving to an organism such as *S. cerevisae, P. pastoris,* or a mammalian system would solve this issue. It has been shown that the pathways and proteins employed in the correct disulphide formation show remarkable similarity between prokaryotes and eukaryotes[182]. This would indicate that if protein expression had been possible in either of the yeast systems, a similar result would be seen to that in *E. coli*. The high expression levels seen in *E. coli* may have caused an overload in these pathways, hence the lower expression levels in yeast and mammalian cell culture. This, coupled with the glycosylation effects discussed

previously, may result in soluble protein that is correctly formed.

The expression of the IR described in the literature contains many different constructs, from L1-CR-L2 to the entire ectodomain[50,63,171,172], however the L1 domain has not been expressed by itself. Expression of the "core" L1 region is a sensible strategy in obtaining a complex between insulin and the IR, because removing the flexible regions between the domains of the IR increases the possibility that crystals will form. Examining the crystal structure of L1-CR-L2 from 2hr7.pdb[63], it can be seen that there are hydrogen bonds formed between the L1 and CR regions (Figure 67).
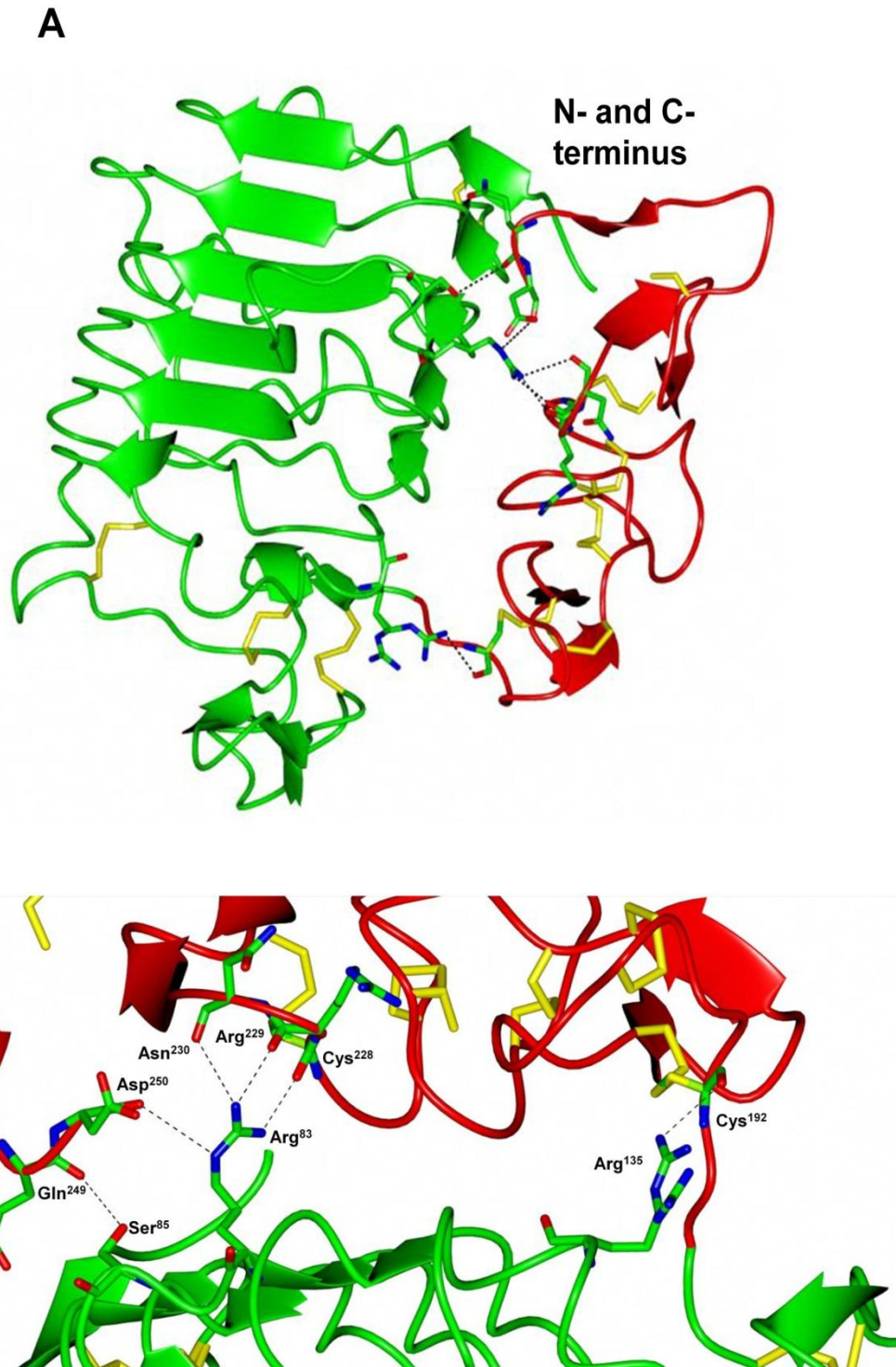
***Figure 67. Interaction between the L1 and CR regions in the IR.***

(A) An overview of the L1-CR contacts. (B) A close up of the L1-CR interactions. The L1 region is coloured green, and the CR in red. (pdb used: 2hr7.pdb). The dashed line indicates a hydrogen bond. Carbon atoms in green, oxygen – red and nitrogen – blue. Disulphide bonds are in yellow.

The side chains of both domains form six hydrogen bonds between them. These hydrogen bonds could help stabilise the regions that lack secondary structure, (the areas between the series of β-sheets), and could be important in the formation of a soluble protein.

The glycosylation of the L1 region, its interactions with the CR, the choice of expression systems and construct design are inconsequential to the correct formation of the disulphide bonds. As demonstrated in *E. coli* soluble protein can be achieved, but it is the formation of the disulphide bonds that prevented the future use of this protein. Solving the disulphide bond formation issue should result in the production of large amounts of the L1 region that are correctly formed.

# Chapter 5. Insulin:Insulin Receptor Complex Formation

## 5.1. Summary

The ultimate aim of this project was to obtain the crystal structure of insulin-IR complex.  Successful complexation and structure derivation could lead to a new rationale for the design of new insulin analogues and drugs for the treatment of diabetes. This can be achieved by analysis of the insulin- IR interface and the nature of conformational changes occurring in this hormone and its receptor upon binding.  Insulin analogues that expose the main IR binding epitopes have the highest affinity for IR and are therefore most likely to form complexes, thus making them the most promising candidates for structure solution of the complex[4,75].

The selection of the best IR construct for the formation of an effective complex with insulin is not clear. Various constructs were investigated by Kristensen *et al.*[64], all had insulin binding affinities in a range of 9.3-11 nM (Figure 68). All three constructs investigated that lacked the 704-719 CT segment in the receptor construct showed no insulin binding, but the addition of 10 μM of the CT peptide, resurrected normal levels of insulin binding. This indicates that the presence of the CT peptide in solution (or its inclusion in the IR construct) is vital for an effective engagement of insulin and IR. The structural location of the CT peptide in the complex is believed to be on the binding face of the L1-domain[53]. Two constructs, which were supplied by WEHI collaborators, IRB17dB and IRB13 (Figure 68), were used in attempts to make complexes with various insulin analogues (Table 14). Fabs were associated with the IR constructs and were added for crystallisation to facilitate crystal contacts.  The CT peptide consisting of 709-719 residues of the IR α-subunit was synthesised and delivered by IOCB Prague collaborators.
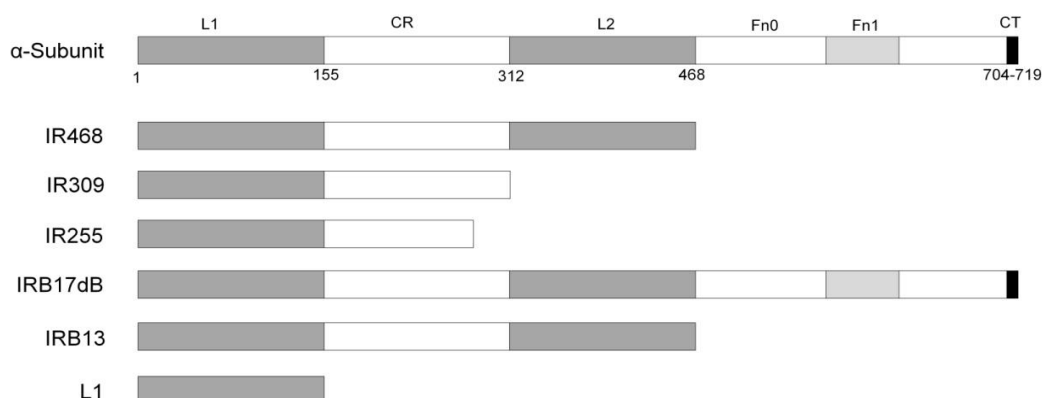
***Figure 68. Schematic of previously investigated IR constructs.***

The receptor constructs IR468, IR309 and IR255 were with and without their C-terminal CT peptide by Kristensen *et al.*[64]. The L1 region expression was described in Chapter 4. IRB17dB and IRB13 were supplied by Australian collaborators.

The mechanism behind the formation of a complex between insulin:IR:CT peptide is unknown.

Various attempts were also undertaken to obtain a direct, binary complex between insulin and the CT peptide to obtain any insight into the nature of interactions of these proteins. These trials were unsuccessful, which indicates that insulin and the CT peptide do not interact directly or interaction between the two is dependent upon the presence of domains of the IR[183].

The IR-binding abilities of various insulin analogues to constructs IRB17dB and IRB13 were assessed by a native gel electrophoresis. Several high affinity analogues were identified as being able to form a IR-complex and were used in crystal trials. Diffraction-quality crystals of some complexes were obtained using the short IR construct IRB13. These crystals lead to the first crystal structure of the insulin:IR complex.

## 5.2. Methods Used for Trials for a Direct, Binary Insulin:CT Peptide Complex Formation and Insulin:IR:CT Complexes

The formation of insulin analogue and the CT peptide putative binary complexes were studied by 1:5 (insulin:CT) molar ratio mixing of the proteins BUFFER?. They were incubated overnight at $37^{\circ}$C.

The binary complexes of the IR constructs IRB17dB and IRB13 with insulin were formed by mixing the IR and insulin analogue in a 1:1 molar ratio in TBSA (25 mM Tris pH 8.0, 25 mM NaCl and 2 mM KCl). A tertiary complex of insulin analogues, IR and the CT peptide was formed by mixing the IR and CT peptide, then incubating at RT for 1 hour before adding the insulin analogue and leaving for a further 3-4 hours at RT. The reaction was buffered against TBSA, with the proteins mixed at final ratio of 1:1.5:3 (IR, CT peptide and insulin respectively).

The insulin analogues used to form complexes in this chapter are shown in Table 14.

*Table 14. A comparison of insulin analogue structures and their relative receptor binding affinities used in this chapter.*

Changes in the B-chain sequence are indicated in bold. [$N$MeXaaB$^{n}$]: methylation of the N-peptide atom preceding Xaa B$^{n}$ amino acid; -NH$_2$ : C-terminal carboxy amide.

| | | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | Affinity |
|---|---|---|---|---|---|---|---|---|---|---|
| | WT: | G | F | F | Y | T | P | K | T | 100 |
| [$N$MeHis$^{B26}$]-DTI-NH$_2$ | LZ29 | G | F | F | **Me H -** *CONH$_2$* | | | | | 214 |
| [$N$MeAla$^{B26}$]-DTI-NH$_2$ | LZ30 | G | F | F | **Me A -** *CONH$_2$* | | | | | 465 |
| [$N$MeAla$^{B26}$]-insulin | LZ33 | G | F | F | **Me A** | T | P | K | T | 21 |
| [$N$MePhe$^{B26}$]-DTI-NH$_2$ | LZ34 | G | F | F | **Me F -** *CONH$_2$* | | | | | 36 |
| [Gly$^{B26}$]-insulin | LZ36 | G | F | F | **G** | T | P | K | T | 69 |
| [$N$MeGly$^{B26}$]-insulin | LZ38 | G | F | F | **Me G** | T | P | K | T | 92 |
| [$N$MeTyr$^{B26}$]-insulin | LZ45 | G | F | F | **Me** Y | T | P | K | T | 21 |
| [D-Ala$^{B26}$]-DTI-NH$_2$ | LZ53 | G | F | F | **D-A-***CONH$_2$* | | | | | 381 |
| [Pro$^{B26}$]-DTI-NH$_2$ | LZ57 | G | F | F | **P** | T - *CONH$_2$* | | | | 20 |
| [D-Pro$^{B26}$]-DTI-NH$_2$ | LZ59 | G | F | F | **D-P -** *CONH$_2$* | | | | | 356 |

## 5.3. Study on Binary, Direct Insulin:CT Peptide Complexes

### 5.3.1. Results and Discussion

If the CT peptide is absent from a complex between insulin and the IR, binding cannot occur[64]. Therefore an 'intermediate' step in insulin-IR interaction was studied here by following a putative binary insulin:CT interaction monitored by the native gel electrophoresis (Figure 69).
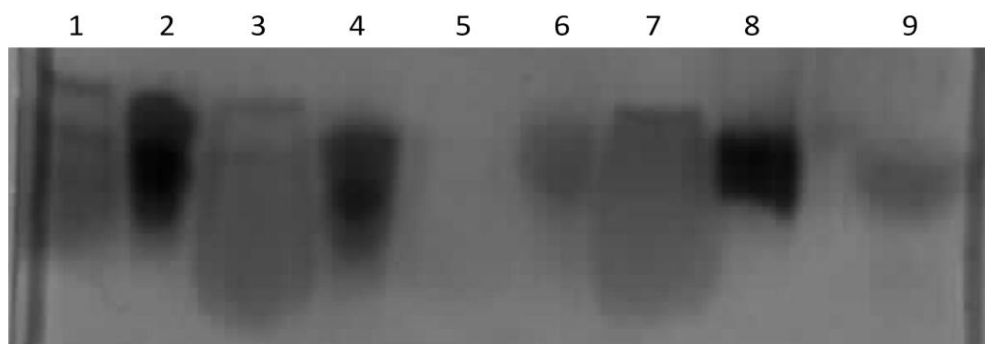


***Figure 69. Initial screening of binary insulin:CT peptide complex formation.***
A 17.5% native gel analysing complexation between insulin analogues and the CT peptide. Lane 1: WT insulin. Lane 2: WT insulin and CT peptide. Lane 3: LZ30. Lane 4: LZ30 and CT peptide. Lane 5: LZ33. Lane 6: LZ33 and CT peptide. Lane 7: LZ59. Lane 8: LZ59 and CT peptide. Lane 9: CT peptide.

The absence of new bands present in the lanes containing both an analogue and the CT peptide indicates that a complexes have not been formed (Figure 69). As insulin analogues were dissolved in 20 mM HCl, whilst the CT peptide was in 50 mM Tris pH 9.0, the non-optimum pH resulting from the mixing of these buffers could be the reason for a lack of detectable complexes. Therefore Dynamic light scattering (DLS) was employed to gain more physic-chemical information about high affinity insulin analogue LZ30 and the CT peptide both independently and together, was performed. The hydrodynamic size, indication of possible interactions and aggregation state of the proteins were obtained. The molecular weights of insulin and the CT peptide are 5.6 and 1.8 kDa respectively, and thus the theoretical weight of a 1:1 complex between insulin and the CT peptide is 7.4 kDa. However, the only stable/interpretable DLS result for measurements of the insulin:CT solutions was approximately three orders of magnitude larger than the size of monomeric insulin (Table 15).

***Table 15.  DLS data of complexation between LZ30 and CT peptide under varying conditions.***

A table showing the results of DLS data for LZ30 conducted with and without the CT peptide.  - indicates no readings were able to be recorded.

|  |  | MW (kDa) | Rh (nm) | Pol | Baseline |
|---|---|---|---|---|---|
| LZ30 | RT | - | - | - | - |
| pH 3.0 | 37$^o$C | - | - | - | - |
| CT | RT | 5.95 x 10$^3$ | 24.1 | 0.40 | 1.000 |
| pH 9.0 | 37$^o$C | - | - | - | - |
| LZ30 + CT | RT | - | - | - | - |
| pH 7.6 | 37$^o$C | - | - | - | - |
| LZ30 + CT | RT | - | - | - | - |
| ~pH 2.0 | 37$^o$C | - | - | - | - |

The failure to obtain any results in the expected range (~7.4 kDa) could indicate aggregation of insulin in the solution. To prevent this, small amounts of denaturants (urea and guanidinium-hydrochloride [Guan-HCl]) were added to the same conditions. This resulted in the DLS data now falling within the expected range (Table 16).

*Table 16. DLS data of complexation between LZ30 and CT peptide with various denaturants.*

A table showing the results of DLS assessments conducted in solutions containing the insulin analogue LZ30 and the CT peptide, both independently and together, in the presence of a denaturing agent. "-" indicates no readings were able to be recorded.

| Condition | Name | | MW (kDa) | Rh (nm) | Pol | Baseline |
|---|---|---|---|---|---|---|
| LZ30 pH 3.0<br><br>50 mM Urea | 1A | I Peak 2 | 18.36 | 0.17 | 0.46 | 1.001 |
| | | M Peak 2 | 17.07 | 0.22 | 0.46 | 1.001 |
| CT pH 9.0<br><br>50 mM Urea | 1B | | $1.16 \times 10^5$ | 55 | 0.52 | 0.999 |
| LZ30 + CT pH 7.6<br><br>50 mM Urea | 1C | - | - | - | - | - |
| LZ30 + CT ~pH 2.0<br><br>50 mM Urea | 1D | - | - | - | - | - |
| LZ30 pH 3.0<br><br>50 mM Guan-HCl | 2A | - | - | - | - | - |
| CT pH 9.0<br><br>50 mM Guan-HCl | 2B | - | - | - | - | - |
| LZ30 + CT pH 7.6<br><br>50 mM Guan-HCl | 2C | - | - | - | - | - |
| LZ30 + CT ~pH 2.0<br><br>50 mM Guan-HCl | 2D | I Peak 2 | 8.12 | 1.46 | 0.48 | 1.000 |
| | | M Peak 2 | 7.09 | 1.38 | 0.48 | 1.000 |

The molecular weights indicated in condition 2D: 8.12 and 7.09 kDa indicate that a complex may have been formed between LZ30 (insulin analogue) and the CT peptide. Upon running this sample on a native gel (Figure 70), the indication that a complex had been formed was further supported by the absence of the LZ30 band and the presence of a new band.
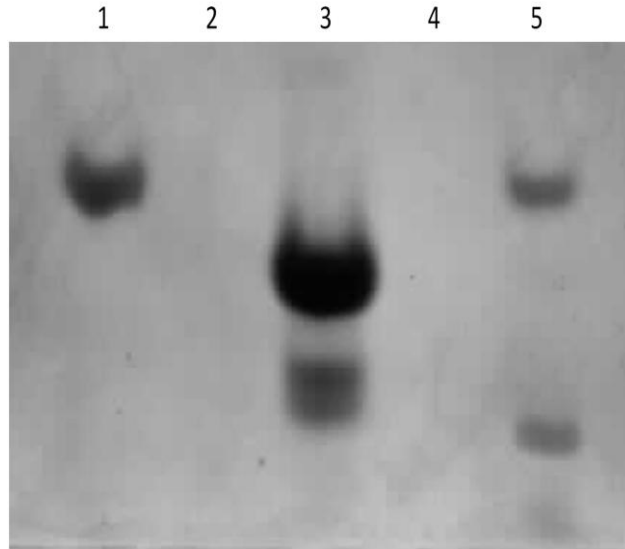
***Figure 70. The complexation of LZ30 and the CT peptide in condition 2D of Table 16 (presence of 50 mM Guan-HCl***

A 12% native gel of the complexation of LZ30 and the CT peptide in 2D conditions. Lane 1: CT peptide. Lane 3: LZ30. Lane 5: LZ30 and the CT peptide in solution under condition 2D from Table 16.

The next step undertaken was to investigate abilities of other, high- and low affinity insulins analogues to form a direct complexes with the CT peptide.

Numerous bands present in lanes with the insulin analogue and CT peptide indicate some types of oligomerisation (Figure 71-A and –B). In a native or SDS-PAGE gel, the bands are usually blue in colour due to the Coomassie Blue stain, however the bands circled are those that were purple, which corresponds to the CT peptide. There are also purple bands present in LZ30, 53 and 45, which could indicate an analogue:CT peptide complex being present. The true importance and meaning of the purple bands is unknown and have not been studied here.
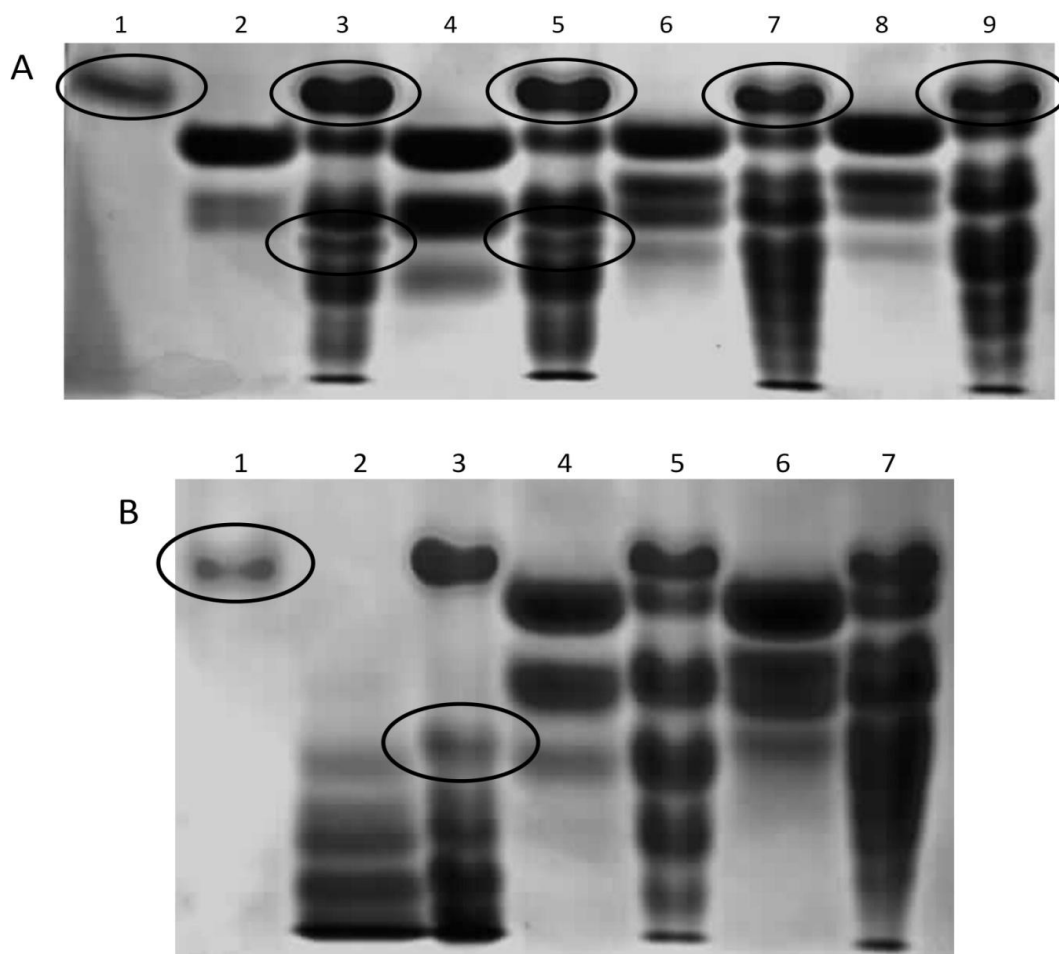
**Figure 71. A range of insulin analogues and CT peptide complexed in 2D conditions from Table 16.**

(A) A 14% native gel of a range of insulin analogues complexed with the CT peptide in 2D conditions . Lane 1: CT peptide. Lane 2: LZ30. Lane 3: LZ30 and CT peptide. Lane 4: LZ53. Lane 5: LZ53 and CT peptide. Lane 6: LZ57. Lane 7: LZ57 and CT peptide. Lane 8: LZ33. Lane 9: LZ33 and CT peptide. (B) A 14% native gel of a range of insulin analogues and CT peptide complexed in 2D conditions. Lane 1: CT peptide. Lane 2: LZ45. Lane 3: LZ45 and CT peptide. Lane 4: LZ36. Lane 5: LZ36 and CT peptide. Lane 6: LZ38. Lane 7: LZ38 and CT peptide. The bands circled are purple in colour.

A native gel was used to check whether these additional bands were the consequence of either a complex being formed, or as a by-product of the conditions (Figure 72). Due to the presence of the same bands in lanes 2 and 3 it can been assumed that the extra bands were formed as a by-product of the conditions present in the samples and not those of a new complex formation.
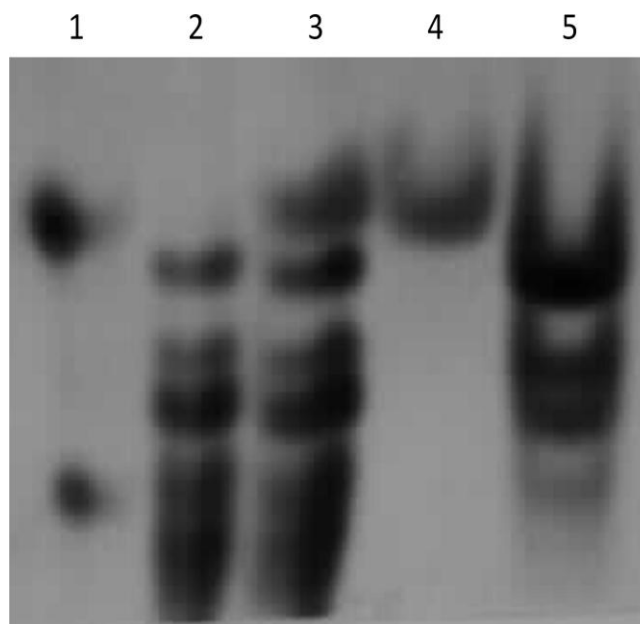
***Figure 72. The complexation of LZ57 and the CT peptide in 2D conditions of Table 16.***

A 14% native gel of the complexation of LZ57 and the CT peptide in 2D conditions. Lane 1: CT peptide in 2D conditions. Lane 2: LZ57 in 2D conditions. Lane 3: LZ57 and CT peptide complexed in condition 2D. Lane 4: the CT peptide in native conditions. Lane 5: LZ57 in native conditions.

Insulin is soluble at a pH $<5$ and $>7$[108], so it was decided to investigate dissolving insulin at a higher pH, closer to pH 9.0 in which the CT peptide can only be dissolved. The solubility of WT insulin was investigated in Tris buffers at a range of pH values (Table 17). The samples were then run on a native gel to assess their ability to enter the gel (Figure 73).

***Table 17. WT insulin solubility trial.***

A trial of various conditions to investigate the solubility properties of WT insulin in alkali conditions.

| Tris Buffer Concentration | pH | Soluble | Notes |
|---|---|---|---|
| 1 M | 10 | Yes | Clear |
| 1 M | 9.6 | Yes | Clear |
| 1 M | 9 | Yes | Minor precipitation, goes clear with time |
| 1 M | 8.6 | No/Yes | Some precipitation, goes clear with time. |
| 1 M | 8.4 | No/Yes | Some precipitation, goes clear with time. |
| 1 M | 8.2 | No/Yes | Some precipitation, goes clear with time. |
| 1 M | 7.5 | No | Precipitation |
| 0.5 M | 10 | Yes | Clear |
| 0.25 M | 10 | No | Still some precipitation with 2x dilution |
| 50 mM | 9 | No | Insoluble |
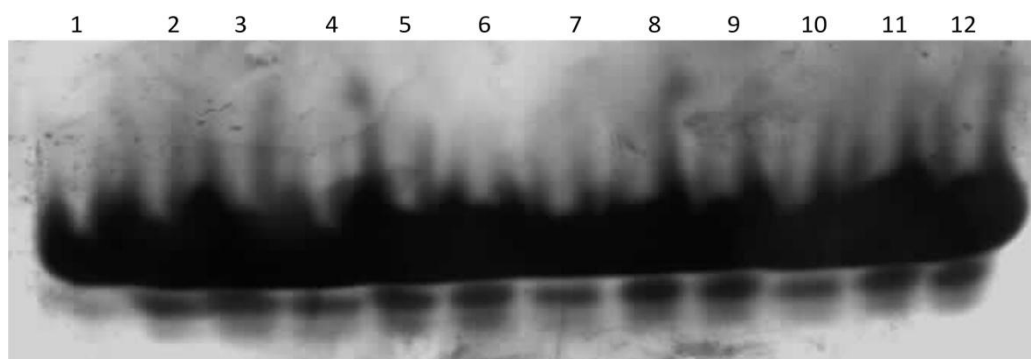


***Figure 73. A native gel assessment of soluble WT insulin at various pHs and Tris concentrations.***

A 12% native gel of WT insulin dissolved in a range of pHs and Tris concentrations. All lanes contain WT insulin. Lane 1: 1 M Tris pH 10. Lane 2: 0.5 M Tris pH 10. Lane 3: 1 M Tris pH 9.6. Lane 4: 0.5 M Tris pH 9.6. Lane 5: 1 M Tris pH 9. Lane 6: 0.5 M Tris pH 9. Lane 7: 1 M Tris pH 8.6. Lane 8: 0.5 M Tris pH 8.6. Lane 9:1 M Tris pH 8.4. Lane 10: 0.5 M Tris pH 8.4. Lane 11: 1 M Tris pH 8.2. Lane 12: 0.5 M Tris pH 8.2.

WT insulin is not able to fully dissolve at a range of pH values (Table 17). At pHs and buffer concentrations in which insulin can fully dissolve, it is also able to enter a native gel (Figure 73). The ability to dissolve insulin at an alkaline pH (closer to pH 9 of the soluble CT peptide) could facilitate more likely formation of insulin:CT peptide.

The initial complexation trial between LZ30 and LZ34 in an alkaline solution and WT insulin with the CT peptide (Figure 74-A) shows no obvious gel shift indicative of

complexation, although there is perhaps a difference between the lanes of LZ30 (lanes 4 and 5) and LZ34 (lanes 6 and 7) before and after the addition of the CT peptide.
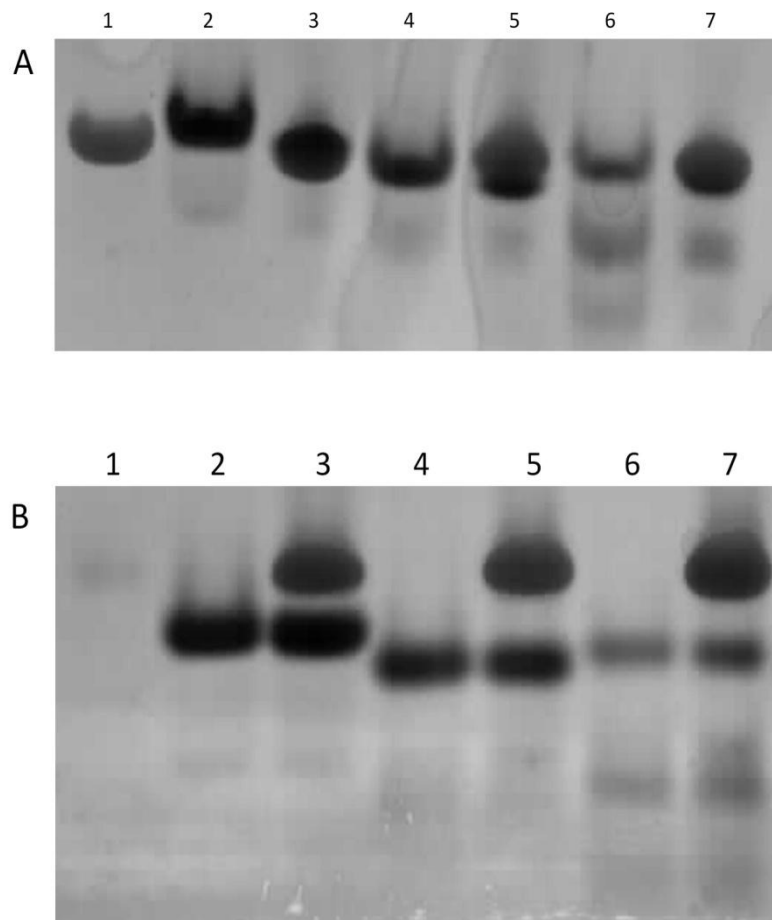


***Figure 74. An initial complexation trial between WT insulin in an alkaline solution and the LZ30 and LZ34 insulin analogues, with CT peptide.***

(A) A 12% native gel of an initial complexation trial between WT insulin in an alkaline solution, LZ30 and LZ34 insulin analogues and CT peptide. Lane 1: CT peptide. Lane 2: WT insulin. Lane 3: WT insulin and CT peptide. Lane 4: LZ30. Lane 5: LZ30 and CT peptide. Lane 6: LZ34. Lane 7: LZ34 and CT peptide. (B) A 12% native gel with a stacking gel of an initial complexation trial between alkali WT insulin with a stacking gel, LZ30 and LZ34 insulin analogues and CT peptide. Lane 1: CT peptide. Lane 2: WT insulin. Lane 3: WT insulin and CT peptide. Lane 4: LZ30. Lane 5: LZ30 and CT peptide. Lane 6: LZ34. Lane 7: LZ34 and CT peptide.

The differences between the bands in Figure 74-A were hard to distinguish, so another gel was run using a stacking gel to increase band separation to determine whether a new species had been created (Figure 74-B). This indicated that the potential shift in Figure 74-A was caused by smearing between the two bands, and not the creation of a new species (Figure 74-B).

A 1:5 ratio between WT insulin or insulin analogues and the CT peptide has been used to investigate the complexation between these species. This approach did not yield

any results.  The next step was an investigation of whether this ratio is appropriate.  Using the most potent insulin analogue, LZ30, a range of ratios from 1:7 to 7:1 (LZ30:CT peptide) were assessed.
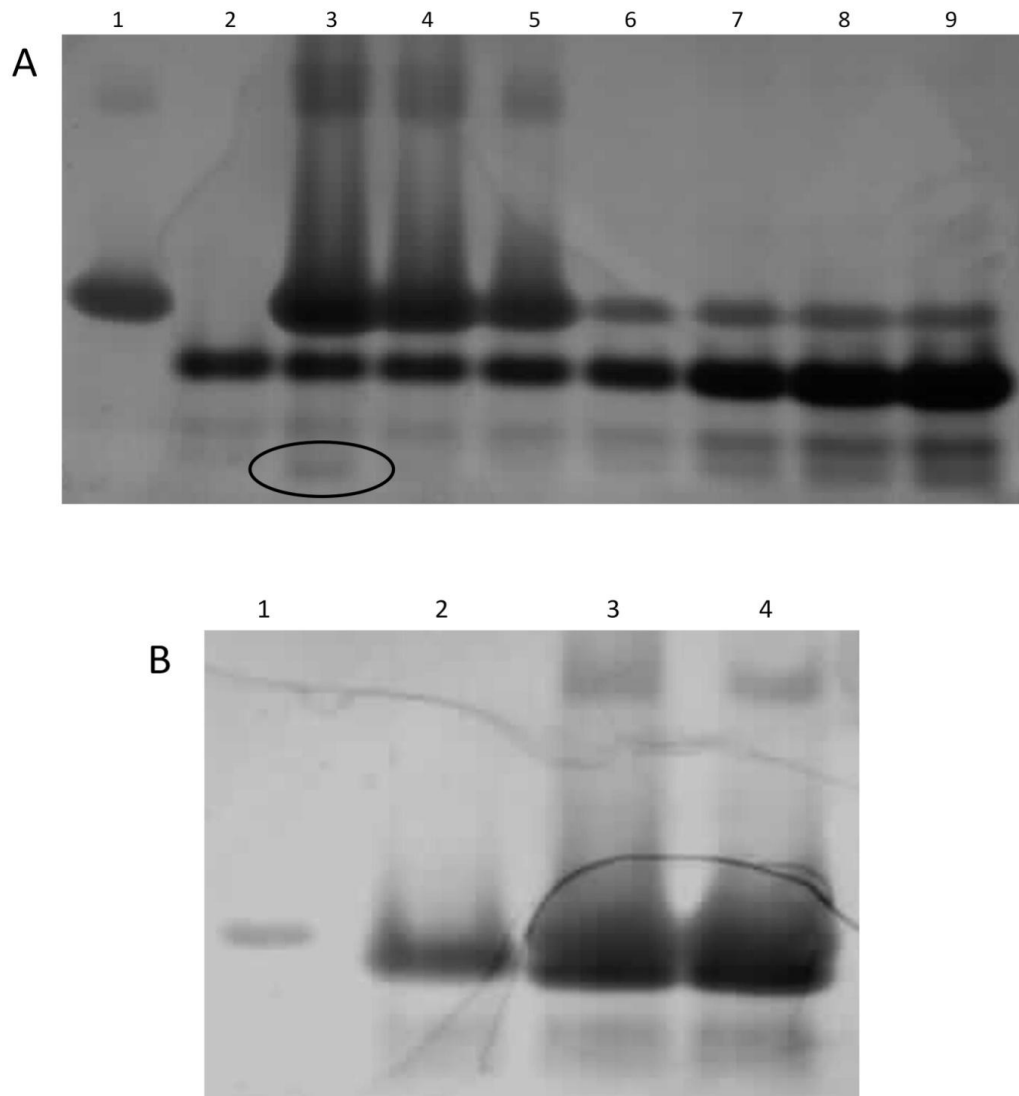


**Figure 75.  A complexation trial between LZ30 (alkali) and CT peptide at different molar ratios.**

(A) A 12% native gel of a complexation trial between LZ30 (alkali) and CT peptide at a variety of ratios.  Lane 1: CT.  Lane 2: LZ30.  Lane 3: LZ30 and CT peptide at 1:7 ratio.  Lane 4: LZ30 and CT peptide at 1:5 ratio.  Lane 5: LZ30 and CT peptide at 1:3 ratio.  Lane 6: LZ30 and CT peptide at 1:1 ratio.  Lane 7: LZ30 and CT peptide at 3:1 ratio.  Lane 8: LZ30 and CT peptide at 5:1 ratio.  Lane 9: LZ30 and CT peptide at 7:1 ratio. (B) A 17% native gel of a complexation trial between LZ30 (alkali) and CT peptide at a 1:7 ratio respectively at two temperatures. Lane 1: CT peptide. Lane 2: LZ30.  Lane 3: LZ30 and CT peptide incubated at 37$^{\circ}$C.  Lane 4: LZ30 and CT peptide incubated at RT.

No new major species appear to be present, although the circled band (lane 3) does appear in only one condition (Figure 75-A). This sample was re-run on a higher percentage gel, where it is no longer visible, indicating that it must have been an artefact of the gel (Figure 75-B).

The change in pH of the insulin molecule could impede its ability to bind to the IR. The insulin molecule dissolved in 1.0 and 0.5 M Tris and was complexed with IR468, then run on a native gel (Figure 76). A gel shift occurred, which meant that despite a change in pH, binding between insulin and the IR still occurred.
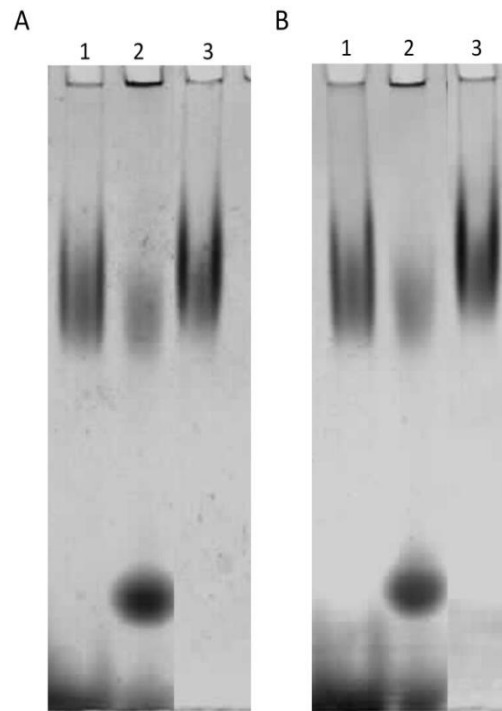


***Figure 76.  Native gel electrophoresis of the initial complexation trials of the IR and WT insulin +/- CT peptide.***

A 7.5% native gel of the initial complexation between IR468 and WT insulin dissolved in 1 M Tris pH 9.0 (gel A) and 0.5 M Tris pH 9.0 (gel B). Gel (A) Lane 1: WT insulin and IR468.  Lane 2: WT insulin, IR468 and CT peptide.  Lane 3: IR468. Gel (B) Lane 1: WT insulin and IR468.  Lane 2: WT insulin, IR468 and CT peptide.  Lane 3: IR468.

### 5.3.2. Conclusion

The ability to form a complex between insulin and the IR is dependent upon the presence of the CT peptide[64], which is believed to be located on the binding surface of the L1 region of the IR[53]. The nature of the interactions between insulin, the IR and the CT peptide are unknown; therefore a crystal structure of this complex would provide vital evidence.

The inability of a complex to be formed between insulin and the CT peptide could be due to many reasons. For example, binding of the CT peptide to the IR could be required prior to insulin binding. The low resolution structure of the L1 domain of the IR[53] shows electron density of the CT peptide binding to the face of the L1 region that is α-helical. IR could have a chaperone effect on folding of this peptide inducing its α-helical conformation. This would allow its binding to the insulin molecule, meaning that the L1 region is necessary for the binding=active structure of the CT peptide and insulin. Another reason for the lack of complex formation may be that the insulin molecule and the CT peptide do not actually interact with each other, and that the CT peptide is involved in other interactions.

## 5.4. Formation of Stable Insulin:IR Complexes With and Without CT Peptide (Collaborators IR Constructs)

### 5.4.1.  Results and Discussion

The ability of IRB17dB and IRB13 to form a binary IR-insulin complex was tested on a native gel. For IRB17dB it can be seen that a visible shift has occurred in lanes 2 (LZ29), 4 (LZ33) and 6 (LZ45) (Figure 77-A). Of these, LZ29 (lane 2) is a highly active analogue (~214%), whilst LZ33 and LZ45 (lanes 4 and 6 respectively) are low affinity analogues (both ~20%). All analogues display a B26- or B26-like turn. It appears that reducing the size of the side chain in position $B^{26}$, reduces the ability of the analogue to form a complex with IRB17dB, contrary to the binding affinity in cases. A full length analogue restores binding capability despite the reduced size of the side chain (LZ30 and LZ33). In all of the lanes with IRB17dB, a degree of oligmerisation seems to be present, and parts of some of the samples failed to enter the gel as well.  This oligomerisation could also explain why some of the analogues failed to bind to the receptor due to its not optimum conformation.

There is a distinct shift for the high affinity analogues LZ29 (~214%, lane 2) and LZ59 (~360%, lane 7) (Figure 77-B).  A fairly high degree of smearing has occurred, possibly due to glycosylation, making the analysis more difficult. However the shifts seen for LZ29 and LZ59 indicating receptor binding is clear.
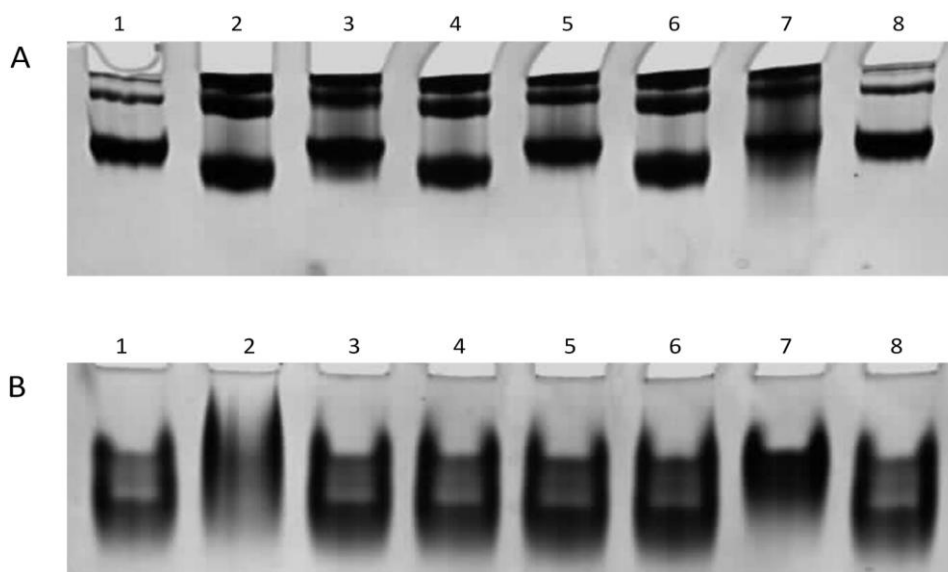
***Figure 77. Native gel electrophoresis of the initial binary complexation of IRB17dB (A) and IRB13 (B) with insulin analogues.***

A 7.5% native gel of the initial complexation between IR-B17dB (A) and IR-B13 (B) with insulin analogues. Gel (A), all lanes contain IRB17dB. Lane 1: Receptor only. Lane 2: LZ29. Lane 3: LZ30. Lane 4: LZ33. Lane 5: LZ36. Lane 6: LZ45. Lane 7: LZ59. Lane 8: Receptor only. Gel (B), all lanes contain IRB13. Lane 1: Receptor only. Lane 2: LZ29. Lane 3: LZ30. Lane 4: LZ33. Lane 5: LZ36. Lane 6: LZ45. Lane 7: LZ59. Lane 8: Receptor only.

Subsequently, the complexation of IRB17dB with insulin analogues and WT insulin was conducted comparing IRB17dB alone, or in the presence of the CT peptide (Figure 78). From the gel shifts it appears that LZ29 (gel A, lanes 4 and 5) and LZ30 (gel A, lanes 6 and 7) are able to bind to the receptor with and without the presence of the CT peptide (Figure 78-A). The LZ30 analogue had been unable previously to bind to the IR, though for this gel a fresh preparation was used, indicating that the age of the sample may play an important role. LZ45 seems to bind to the receptor (Figure 78-B). LZ59 (gel B, lanes 4 and 5) alone does not appear to be able to bind to the receptor, although upon addition of the CT peptide, a complex was formed. All +/- CT peptide lanes are smeared, which could indicate the dynamic character of the complexes. LZ33 appears to bind to the receptor (Figure 78-C, lanes 6 and 7). LZ36 binds upon the addition of IRB17dB, whilst upon the addition of the CT peptide there appears to be dissociation of the complex (gel C, lanes 4 and 5). The band present corresponds to a transient state, indicating that the binding of LZ36 to the IR is dynamic and dependent upon the CT peptide. In all gels WT insulin appears to bind to the receptor, with the gel shifts indicating binding very similar to that of the insulin analogues (Figure 78).
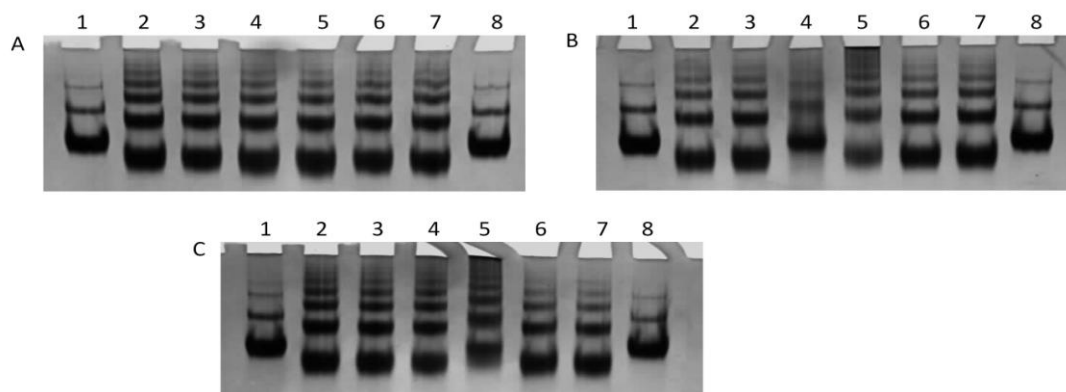
***Figure 78. Native gel electrophoresis of the IRB17dB – insulin complexes in the presence or absence of the CT peptide.***

A 5% native gel of the complexation of IRB17dB with insulin analogues, investigating the effect of the CT peptide. All lanes on all three gels contain IRB17dB. Gel (A), lane 1: Receptor only. Lane 2: WT insulin. Lane 3: WT insulin with CT peptide. Lane 4: LZ29. Lane 5: LZ29 with CT peptide. Lane 6: LZ30. Lane 7: LZ30 with CT peptide. Lane 8: Receptor only. Gel (B), lane 1: Receptor only. Lane 2: WT insulin. Lane 3: WT insulin with CT peptide. Lane 4: LZ59. Lane 5: LZ59 with CT peptide. Lane 6: LZ45. Lane 7: LZ45 with CT peptide. Lane 8: Receptor only. Gel (C), lane 1: Receptor only. Lane 2: WT insulin. Lane 3: WT insulin with CT peptide. Lane 4: LZ36. Lane 5: LZ36 with CT peptide. Lane 6: LZ33. Lane 7: LZ33 with CT peptide. Lane 8: Receptor only.

Native gels were run to investigate the complexation of IRB13 with insulin analogues and WT insulin comparing IRB13 alone, to IRB13 in the presence of the CT peptide (Figure 79). For LZ30 (gel A, lanes 4 and 5) and LZ33 (gel B, lanes 6 and 7) (Figure 79) some upward smearing in the bands indicates that some binding has occurred. The addition of CT peptide does not show special effect upon IRB13 binding to any of the analogues.
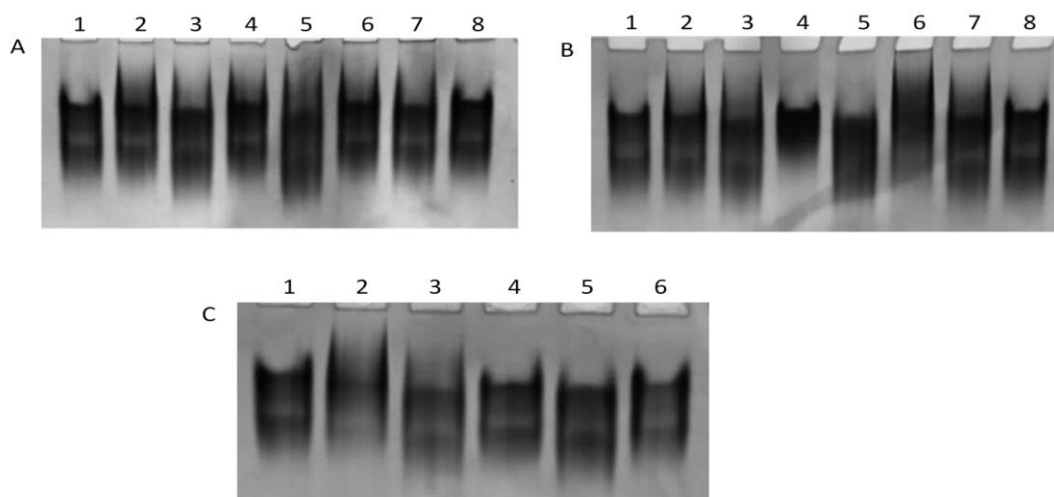
***Figure 79. Native gel electrophoresis of the complexation of IRB13 with insulin analogues, investigating the effect of the CT peptide.***

A 7.5% native gel of the complexation of IRB13 with insulin analogues, investigating the effect of the CT peptide. All lanes on all three gels contain IRB13.  Gel (A), lane 1: Receptor only.  Lane 2: WT insulin.  Lane 3: WT insulin with CT peptide.  Lane 4: LZ30. Lane 5: LZ30 with CT peptide.  Lane 6: LZ45.  Lane 7: LZ45 with CT peptide.  Lane 8: Receptor only. Gel (B), lane 1: Receptor only.  Lane 2: WT insulin.  Lane 3: WT insulin with CT peptide.  Lane 4: LZ59.  Lane 5: LZ59 with CT peptide.  Lane 6: LZ33.  Lane 7: LZ33 with CT peptide.  Lane 8: Receptor only. Gel (C), lane 1: Receptor only.  Lane 2: WT insulin.  Lane 3: WT insulin with CT peptide.  Lane 4: LZ36.  Lane 5: LZ36 with CT peptide.  Lane 6: Receptor only.

The addition of certain organic compounds, notably TFEtOH and phenol, can promote α-helical formation of disordered proteins. It appears that low pH (pH 4.0) promotes an addition of the CT peptide to the analogue:receptor complex (Figure 80).  The promotion of the α-helix seems to be present with the addition of phenol because of the stabilisation of the complex, whilst TFEtOH appears to destabilise the complex.
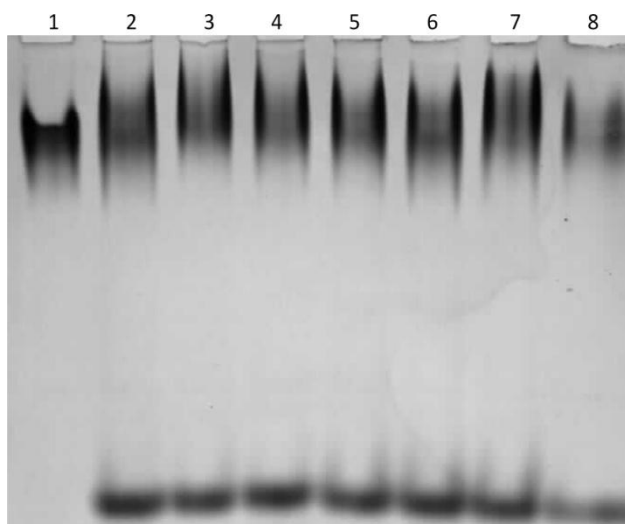
***Figure 80.  Native gel electrophoresis of the complexation between IRB13 with LZ29 and the CT peptide.***

A 7.5% native gel of the complexation of IRB13 with LZ29 and the CT peptide, investigating the effects pH and additives. All lanes contain IR-B13.  Gel (A), lane 1: Receptor only.  Lane 2: LZ29 and CT peptide.  Lane 3: LZ29 and CT peptide, pH 4.0.  Lane 4: LZ29 and CT peptide, pH 6.5.  Lane 5: LZ29 and CT peptide, pH 8.5.  Lane 6: LZ29 and CT peptide with 0.1 % TFEtOH.  Lane 7: LZ29 and CT peptide with 0.1 % phenol.  Lane 8: LZ29.

A systematic approach was taken to analyse the effect of addition of phenol during the complexation of insulin, insulin analogues and IRB13 (Figure 81).  The only clear difference upon the addition of phenol occurs when it is added to WT insulin and CT peptide (gel A, lanes 4 and 6).  This could indicate a good route upon which to try and form a ternary complex, with insulin, the IR and CT peptide.
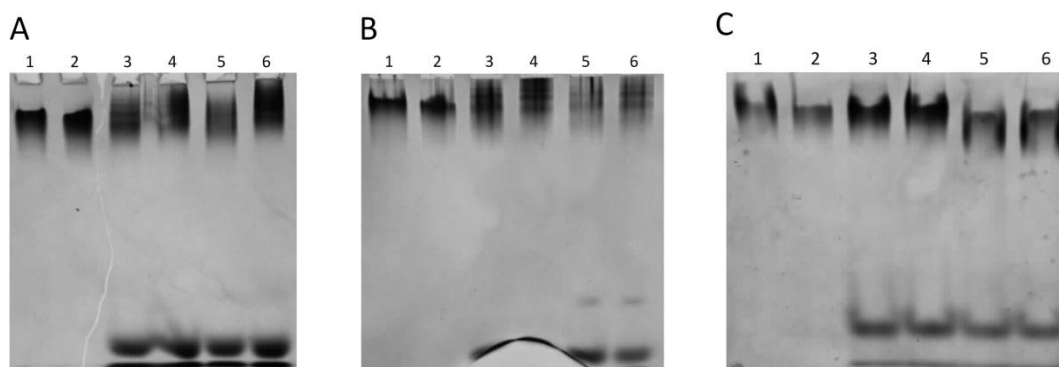
***Figure 81. Native gel electrophoresis of the complexation between IRB13 with various insulins, investigating the effects of the CT peptide, pH and additives.***

A 7.5% native gel of the complexation of IR-B13 with various insulins, investigating the effects the CT peptide, pH and additives. All lanes on all three gels contain IR-B13. Gel (A), lane 1: 0.15 % phenol. Lane 2: Receptor only. Lane 3: WT insulin. Lane 4: WT insulin and 0.15 % phenol. Lane 5: WT insulin and CT peptide. Lane 6: WT insulin and CT peptide with 0.15 % phenol. Gel (B), lane 1: 0.15 % phenol. Lane 2: Receptor only. Lane 3: LZ29. Lane 4: LZ29 and 0.15 % phenol. Lane 5: LZ29 and CT peptide. Lane 6: LZ29 and CT peptide with 0.15 % phenol. Gel (C), lane 1: 0.15 % phenol. Lane 2: Receptor only. Lane 3: LZ30. Lane 4: LZ30 and 0.15 % phenol. Lane 5: LZ30 and CT peptide. Lane 6: LZ30 and CT peptide with 0.15 % phenol.

To gain physical information on binding between the IR and insulin analogues, the ITC measurements[184] were conducted on LZ30. They were performed by collaborators at WEHI in Melbourne, Australia. LZ30 was injected into TBSA, which gave no heat of dilution, indicating a good negative control. When LZ30 was injected into IR-B13, it also gave no heat of dilution. The addition of LZ30 into IRB13 in the presence of the CT peptide, however gave a heat of dilution, with a $K_d = 4.85$ nM. WT insulin in this system has a $K_d = 16.6$ nM. This indicates that the binding of the LZ30 analogue to IRB13 is CT peptide dependent. The data of analogue binding previously published[150] show different binding affinities, although still in the low nM range. The proportions between LZ30 and WT binding affinities remain consistent.

The IRB13:insulin:CT peptide complex was buffered in TBSA. The best crystal identified from the screening was obtained from drops equilibrated 1:1 (protein:well) in well solution consisting of 1M tri-sodium citrate pH 8 (1 ml well volume), and grown at 21°C. The crystal was flash-cooled in liquid nitrogen using paraffin oil as cryoprotectant. The X-ray data were collected at Diamond at station I24, and processed using XDS[185]. The data collection statistics are shown for three crystals (Table 18).

***Table 18. X-ray data processing and refinement of the insulin:IR complex.***

|  | Data set 1 | Data set 2 | Data set 3 |
|---|---|---|---|
| Receptor / Fab | IR310.T/Fab83-7 | IR310.T/Fab83-7 | IR310.T/Fab83-7 |
| Insulin | Human | [$D$-Pro$^{B26}$]-DTI-NH$_2$ | [$D$-Pro$^{B26}$]-DTI-NH$_2$ |
| $\alpha$CT peptide | 704-719 | 704-719 | 697-719 |
| X-ray source | Diamond Light Source | Diamond Light Source | Diamond Light Source |
| No. of frames | 180 | 400 | 200 |
| Oscillation range (°) | 0.2 | 0.2 | 0.2 |
| Exposure / frame (s) | 0.2 | 0.2 | 0.2 |
| Unit cell dimensions | $a$ = 168.91 Å | $a$ = 169.49 Å | $a$ = 169.22 Å |
| Space group | P23 | P23 | P23 |
| Solvent content (%) | 80 | 80 | 80 |
| Data completeness (%) | 98.4 (99.6) | 99.7 (100.0) | 98.5 (98.1) |
| Resolution (Å) | 46.8 – 3.9 (4.0 – 3.9) | 29.6 – 4.3 (4.53 - 4.3) | 29.5 – 4.3 (4.54 – 4.3) |
| $R_{merge}$ | 0.089 (1.32) | 0.188 (2.69) | 0.101 (1.09) |
| $R_{pim}$ | 0.089 (0.83) | 0.067 (0.97) | 0.053 (0.59) |
| $<I/\sigma(I)>$ | 8.95 (1.0) | 7.8 (1.0) | 8.2 (1.4) |
| Redundancy | 4.1 (4.2) | 8.8 (8.8) | 4.4 (4.3) |

[1] Numbers in parentheses refer to the statistic in the highest resolution shell.

[2] $R_{merge}$ = | $I_i(hkl)$ - $<I(hkl)>$ | / $I_i(hkl)$

[3] $R_{pim}$ = $[N/(N-1)]^{1/2}$ | $I_i(hkl)$ - $<I(hkl)>$ | / $I_i(hkl)$

The data was processed and revealed the structures of complexes between insulin, IRB13 and the CT peptide.

### 5.4.2. Conclusion

The sample of the IRB17dB construct showed some degree of aggregation even prior to formation of the insulin complexes. This could be, for example, due to the CT peptides becoming dislodged from their own receptor and associating transiently with neighbouring homodimers, resulting in their oligomerisation. Domain swapping cannot be excluded here as well. Obtaining a homogeneous mixture of IRB17dB would be required for more successful crystallisation attempts.

Consistently, the LZ29 and LZ30 have shown successfully IR complex formation on native-PAGE analysis. These insulins correspond to some of the highest affinity analogues (214% and 465% respectively). It is therefore understandable that LZ59, another high affinity analogue (356%) formed a stable complex with IRB13 as well. LZ59 displays the same structural characteristics associated with the high affinity analogues seen in LZ29 and LZ30.

The resolutions obtained for the crystals of the complex between the IR, insulin (WT and LZ59) and the CT peptide are low, around 4 Å. This resolution is not high enough to produce detailed interactions between side chains, but it does show how and where the three components come together and form one of the most sought after complexes in structural biology.

# Chapter 6. Conclusions and Future Perspectives

## 6.1. Exploitation of the Conformation Associated with High Affinity Insulin Analogues

For many years the 3-D structures of insulin were derived from its relatively similar storage forms. However, wealth of interdisciplinary data indicated reorganisation on the C-terminus of the B-chain, needed for the exposure of the hydrophobic core of the insulin molecule upon binding to the IR.

The development and structural characterisation of a wide range of high affinity insulin analogues led to the discovery of their structural signature - B26 turn[75]. The B26 turn fulfils the requirements of the active conformation of insulin. The C-terminus of the B-chain is detached from its position in proximity to the N-terminus of the A-chain exposing hydrophobic core of the hormone. The side chains around the B25-B26 sites alternate their positions, forming pattern of interactions that facilitate effective interactions with L1 and ID-αCT regions of the IR upon hormone-receptor binding[77].

Future exploitations of in the B26 turn will involve rational modifications of the B24-B30 C-terminal residues. For example, improvement of IR-binding affinity could require either increase of the stability of the B26 turn (e.g. by optimisation of the click-chemistry based crosslinking) or improvement of specific insulin side chains interactions with the IR.

## 6.2. Further Investigation of the Dimer Interface

The dimer interface is formed by a series of four parallel hydrogen bonds between two anti-parallel β-sheets at the C-terminus of the B-chain. Residues $Phe^{B24}$ and $Tyr^{B26}$ form two hydrogen bonds each from their carbonyl and amide groups[2,80,151]. However, Chapter 3 shows that the most important interaction is that formed by the main chain amide from $Phe^{B25}$ to $Tyr^{A19}$, loss of which causes a dramatic reduction in dimerisation capability[159].

From all insulin aggregation forms (hexameric, dimeric and monomeric) it is the monomer of the hormone that binds to the IR. Current insulin analogues in diabetes-linked clinical applications have tailored characteristics for a specific (i.e. long- or rapid-aciting) use. An increasing of the strength of the dimer interface would lead to new insulin analogues that with more prolonged effects and low-level insulin release. This can be achieved by either increasing the amount of hydrogen bonds formed between the anti-parallel main chain atoms, forming new interactions between side chains of the residues in the interface, or by forming new interactions with the core of the insulin molecule, such as

that seen between Phe$^{B25}$ and Tyr$^{A19}$. A reverse process would help to create a more rapid acting analogue.

## 6.3. Improved Receptor Binding

Insulin binds to the IR in its monomeric form following conformational changes of its storage form. The B26 turn at the C-terminus of the B-chain is characteristic of high affinity analogues, in which the N-terminus of the B-chain exhibits also an intermediate state between the R- and T-states. This type II-like pseudo turn could be of special importance for activation of insulin as well. A range of new analogues addressing the importance of the residues in the B1-B8 N-terminal region would lead to a greater understanding of the active form of insulin upon binding to the IR. Improvements in binding caused by the analogues probing the N-terminal turn could lead to analogues of increased binding affinity similar to those seen at the C-terminus of the B-chain. Coupling of the improvements to both termini of the B-chain would result in a super affinity analogue that would assist high resolution studies of the insulin:IR complex and may provide novel 3-D scaffold for insulin organic mimics.

## 6.4. Improvements in the Crystals of Insulin:IR

A high resolution structure of the complex between Insulin:IR was a key aim of this project. Attempts to express a correctly formed L1 construct were unsuccessful, but through collaboration with the group of Prof. Mike Lawrence (WEHI) a certain insulin:IR complex was crystallised that diffracted to 3.9-4.3 Å.

The construct that produced these crystals consisted of L1-CR-L2, with each domain consisting of a stable individual element linked by a flexible linker region. These flexible regions could be disrupting the crystals ability to diffract to a high resolution. Using a construct that maintains the binding regions important in insulin:IR binding, whilst minimising the flexibility experienced (such as just the L1 region) could improve crystal quality, and therefore diffraction.

## 6.5. Insulin and Insulin-Like Growth Factor Relationship

Insulin and IGF-I and -II are closely related protein hormones that have different biological functions. Insulin is a key metabolism controller, whilst IGF-I and -II are major growth factors. When released into the bloodstream they bind tightly and specifically to their homologous receptors.

The ability of insulin and IGF to cross-bind to their receptors[117] means that insulin can be implied in the regulation of proliferation and growth. IGF-I and -II have been identified as cancer specific growth factors[124]. The structural factors that signal for metabolic or cell growth need to be identified.

The efficient production of an insulin-like core from the full-length IGFs facilitates semi-synthesis of novel IGFs and insulin-IGF hybrids for characterisation of their molecular and cellular specificities as well as IGF-1R and IR binding. This will allow the creation of designer insulin and IGF analogues with specific effects, for example an insulin analogue which has only metabolic effects, and is an antagonist of IGF-I and -II.

# Appendix 1. General Materials and Methods

## A.1. General Materials and methods

### A.1.1. General Materials, Chemicals, and Bacterial and Fungal Strains

Chemicals purchased for use in this project were of the highest purity available. Synthetic genes were purchased from GenScript. PCR primers were ordered from MWG-Eurofins (Germany). KOD hot start DNA polymerase was obtained from Novagen. The pET-YSBLIC3C vector was obtained from the Structural Biology Labratory (University of York). Restriction enzymes *EcoRI, HindIII, NcoI, NotI, SacI* and *XhoI* were purchased from New England Bio Labs. PCR clean-up and mini-prep kits were purchased from Qiagen, and a gel extraction kit from Sigma Aldrich. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Melford Labs. Methanol came from Fisher. Yeast nitrogen base was purchased from Sigma.

The following bacterial and fungal cultures were acquired from collections within the University of York: *E. coli*; XL-10 gold, BL21, Tuner, Rosetta-Gami 2, lemo21 and ArcticExpress, and fungal; *S. cerevisiae* INVSc1, and *P. pastoris* X-33.

### A.1.2. SDS-PAGE Analysis of Proteins

Cell extract samples were diluted in a 4x loading buffer (0.5 M Tris, pH 6.8) containing: glycerol 20 %, SDS 4 %, bromophenol blue 0.1 %, and β-mercaptoethanol 5 %. Resolving gels (15 or 17.5 % acrylamide) were set by adding tetramethylethtlenediamine (TEMED) to a resolving buffer (1.5M Tris, 0.4 % SDS, pH 8.8) containing acrylamide and ammonium persulphate (APS). A stacking gel was poured from a 0.5 M Tris buffer (0.4 % SDS, pH 6.8) containing acrylamide, APS, and TEMED. A BioRad low molecular mass marker mixture was used that contained phosphorylase B 97.4 kDa, serum albumin 66.2 kDa, ovalbumin 45 kDa, carbonic anhydrase 31 kDa, Soybean trypsin inhibitor 21.5 kDa and lysozyme 14.4 kDa. Following this gels were stained with Coomassie Blue dye, and de-stained using a 6:3:1 mixture of water : propan-2-ol : glacial acetic acid. Samples and the molecular mass marker were boiled at 95 $^{\circ}$C for 5 minutes prior to loading. This method was based upon that described by Laemmli[186].

### A.1.3. Native Gel Analysis of Proteins

Cell extracts samples were diluted in a 4x loading buffer (0.5 M Tris, pH 6.8) containing: glycerol 20% and bromophenol blue 0.01%. Resolving gels (5 or 7.5% acrylamide) were set by adding tetramethylethtlenediamine (TEMED) to a resolving buffer (1.5M Tris, pH 8.8) containing acrylamide and ammonium persulphate (APS). Gels were then stained with Comassie Blue dye and de-stained using a 6:3:1 mixture of water : propan-2-ol : glacial acetic acid. This method was based upon that described by Laemmli[186].

### A.1.4. Copper Staining of Acrylamide Gels

A gel was run as described previously. Following electrophoresis, the gel was momentarily washed in milliQ water. The gel was then transferred to a 0.3 M $CaCl_2$ solution for 5 min. The bands were then able to be visualised. Based upon the method described by Lee *et al*.[187].

### A.1.5. Silver Staining of Acrylamide Gels

A gel was run as described previously. Following electrophoresis, the gel was washed twice for 5 min each time in ultra pure water. The gel was fixed by immersion in a solution of 30% ethanol: 10% acetic acid for 15 min twice. The fixant was washed away by two 5 min washes in 10% ethanol, followed by two washes for 5 min in ultra pure water. Sensitiser was applied to the gel for 1 min, and then washed for 1 min twice with water. The gel was stained for 30 min in the dark using the stain solution. Following staining, the gel was washed twice for 20 sec with ultra pure water, then immersed in developer solution for 2-3 min until bands appear. The reaction was stopped with a 5% acetic acid solution for 10 min. The kit used was Silver SNAP Stain Kit II.

### A.1.6. Western Blotting

A gel was run, as described previously. Layers of thick absorbent paper and nitrocellulose membrane were soaked in transfer buffer (25 mM Tris, 193 mM glycine, 20% v/v methanol, pH 8.3) for 10 min. The transfer block was prepared from the anode consisting of thick absorbent paper, nitrocellulose membrane, polyacrylamide gel and thick

absorbent paper. A transfer was performed for 25 min at 25 V, after which Ponceau Red was used to visualise transferred proteins. The marker was marked using a pencil. The membrane was washed twice for 10 min each time with 15 mL 1X TBS (150 mM NaCl, 10 mM Tris, pH 7.5). The membrane was then blocked for 1 h with 20 mL blocking solution prior to being washed twice for 10 min each time with 20 mL 1X TBSTT (500 mM NaCl, 20 mM Tris, 0.2% v/v Triton X-100, 0.05% v/v Tween 20, pH 7.5). A 10 min wash followed with 15 mL 1X TBS, after which an anti-His antibody, diluted 1:1000 in blocking solution, was applied to the nitrocellulose for 1 h at room temperature. The membrane was washed again twice for 10 min each time with 20 mL 1X TBSTT, and then for a further 10 min with 15 mL 1X TBS. Excess solution was then drained away. The substrate solution was prepared immediately before use by composing a solution comprising equal parts of 2X Luminol/Enhancer and 2X peroxide solution. The substrate solution was applied to the membrane and allowed to develop for 1-2 min, with maximum luminescence occurring after 5 min, before taking white light and bioluminescence images, and combining them to form a composite image.

### A.1.7. Dot Blot

The nitrocellulose was marked into a grid using a pencil, after which it was allowed to soak in methanol, causing the nitrocellulose to become permeable. Samples were then applied to the nitrocellulose directly and allowed to absorb into the surface. The membrane was washed twice for 10 min each time with 15 mL 1X TBS (150 mM NaCl, 10 mM Tris, pH 7.5). The membrane was blocked for 1 h with 20 mL blocking solution prior to being washed twice for 10 min each time with 20 mL 1X TBSTT (500 mM NaCl, 20 mM Tris, 0.2% v/v Triton X-100, 0.05% v/v Tween 20, pH 7.5). A 10 min wash followed with 15 mL 1X TBS, after which an anti-His antibody diluted 1:1000 in blocking solution was applied to the nitrocellulose for 1 h at room temperature. The membrane was again washed twice for 10 min each time with 20 mL 1X TBSTT, and then for 10 min with 15 mL 1X TBS. Excess solution was then drained away. The substrate solution was prepared immediately before use by composing a solution comprising equal parts of 2X Luminol/Enhancer and 2X peroxide solution. The substrate solution was applied to the membrane and allowed to develop for 1-2 min, with maximum luminescence occurring after 5 min, before taking white light and bioluminescence images, and combining them to form a composite image.

### A.1.8. Molecular Biology Techniques

### A.1.8.1. Ligation Independent Cloning (LIC)

### A.1.8.1.1. Preparation of LIC Vector using PCR Method

Cleavable His-tagged pET-YSBLIC vector, 1 μg, was transformed into competent DH5α cells and plated onto Luria Bertani (LB) agar + kanamycin (30 μg/mL) before being incubated at 37 $^o$C overnight. The cells were harvested before a mini-prep was carried out to extract the vector DNA (20 μg), which was used as the template for PCR. Each 50 μL PCR volume contained 0.4 μM of forward and reverse primers (ordered from MWG), 0.2 mM dNTPs, 1 μL KOD hot start DNA polymerase (1 unit/μL), 1 mM MgSO$_4$, 5 μL 10X buffer, 50 ng template DNA, 5% v/v DMSO, and water to 50 μL. Thermocycling conditions were as follows: initial denature 3 min at 94 $^o$C, followed by 35 cycles of: denature 30 s at 94$^o$C, annealing 30 s at 60 $^o$C, and extension 2 min at 72 $^o$C, with a final extension step 3 min at 72 $^o$C. *DpnI* was added to the DNA and incubated at 37 $^o$C for 2 h to digest the methylated DNA. The reaction mixture was run on a preparative agarose gel and the bands corresponding to linear vector DNA were cut from the gel and extracted (7.5 μg).

### A.1.8.1.2. Preparation of LIC Vector using BseRI Method

1 μL LIC vector was transformed into competent NovaBlue cells, plated out onto LB + kanamycin (30 μg/mL) agar, and incubated at 37$^o$C overnight. A single colony was picked to inoculate 2 x 50 mL LB + kanamycin (30 μg/mL) and incubated 37$^o$C overnight. All 100 mL culture was harvested before a mini-prep was carried out to extract the vector DNA. A restriction digest comprising 66 μL vector DNA (7.9 μg), 8 μL *BseRI*, 16 μL 10xbuffer (NEB2) and 70 μL water was incubated at 37$^o$C for 1 h 50 min. The sample was run on a preparative agarose gel. The bands were cut out and purified to give linear LIC vector DNA, which was concentrated to 20 ng/μL.

### A.1.8.1.3. Vector LIC T4pol Reaction

A mixture comprising the following: 3.75 μg linearised vector DNA, 15 μL T4pol buffer (10X), 15 μL dTTP (25 mM stock), 7.5 μL DTT (100 mM stock), 3 μL T4 polymerase (2.5 U/μL LIC qualified – Novagen) and H$_2$O to 150 μL, was incubated at 22$^o$C for 30 min, and then at 75$^o$C for 20 min.

### A.1.8.1.4. Insert LIC T4pol Reaction

A mixture comprising the following: 0.2 pmol PCR product (insert), 2 μL T4pol buffer (10x), 2 μL dATP (25 mM stock), 1 μL DTT (100 mM stock), 0.4 μL T4 polymerase (2.5 U/μL LIC qualified – Novagen) and $H_2O$ to 20 μL, was incubated at $22^oC$ for 30 min, and then at $75^oC$ for 20 min.

### A.1.8.1.5. LIC Annealing Reaction

The annealing reaction comprised 1.5 μL of T4 treated vector (50 ng/μL) added to 2.5 μL of the insert T4 reaction, and incubated at room temperature for 10 min. The reaction then received an addition of 1.5 μL EDTA (25 mM stock) and was incubated for a further 10 min at r/t. A negative control using water was also set up.

### A.1.8.1.6. Transformation of Competent *E. coli* Cells

A transformation was performed by adding 5.5 μL of the LIC annealing reaction to 50 μL of competent *E. coli* XL-10 gold cells, and was then incubated on ice for 5 min. The cells were heat-shocked at $42\ ^oC$ for 30 s, and then incubated on ice for a further 5 min. 0.2 mL of LB medium was added and the cells were incubated at $37\ ^oC$ for 1 h. The cells were plated out onto LB agar + kanamycin (30 μg/mL) and incubated overnight at $37\ ^oC$.

### A.1.8.1.7. NdeI/NcoI Reaction Digest

Colonies from the transformation of XL-10 gold cells were picked to inoculate 5 mL cultures of LB + kanamycin (30 μg/mL) and were incubated overnight at $37^oC$ before carrying out a mini-prep. Extracted DNA was then digested in a solution comprising: 1 μg DNA, 1 μL NE buffer 2 (10x), 1 μL *NcoI*, 1 μL *NdeI*, and water up to 10 μL, before being incubated at $37^oC$ for 3 h. The mixture was then run out on an agarose gel.

## A.1.8.2. Traditional Molecular Biology
## A.1.8.2.1. Amplification of the Target Gene by PCR

Each 50 μL PCR volume contained 0.4 μM of forward and reverse primers (ordered from MWG), 2 mM dNTPs, 1 μL KOD hot start DNA polymerase (1 unit/μL, 1 mM MgSO$_4$, 5 μL 10X buffer, 10 ng template DNA, and water to 50 μL. Thermocycling conditions were as follows: initial denature 3 min at 94$^o$C, followed by 35 cycles of: denature 30 s at 94$^o$C, annealing 30 s at 60$^o$C, and extension 2 min at 72$^o$C, with a final extension step 3 min at 72$^o$C. The amplified target gene was cleaned using a PCR clean up kit.

## A.1.8.2.2. Preparation of the Plasmid

1 μL vector was transformed into competent Nova Blue cells, plated onto LB + antibiotic agar and incubated at 37˚C overnight. A single colony was picked to inoculate 10 mL LB + antibiotic and incubated at 37˚C overnight. All 10 mL culture was harvested before a mini-prep was carried out to extract the vector DNA. A restriction digest comprising 10 μL vector DNA, 2.5 μL 10X buffer, 1 μL each restriction enzyme, made up to 25 μL with water, and incubated at 37˚C for 2 hours. 1 μL of phosphatase was added and incubated at 37˚C for 2 hours before heat activating the enzymes at 65˚C for 15 min before being stored at -20˚C.

## A.1.8.2.3. Preparation of the Insert

20 μL of the target gene from PCR was incubated at 37˚C for 8.5 hours before being stored at -20°C.

## A.1.8.2.4. Ligation Reaction

A ligation was setup with a solution comprising 5 μL digested vector, 2 μL digested insert, 1.5 μL 10X buffer, and 1.5 μL ligase, made up to 15 μL with water.

### A.1.8.2.5. Transformation of Competent *E. coli* Cells

A transformation was performed by adding 4 μL of the ligated vector to 50 μL of competent *E. coli* XL-10 gold cells, and then incubated on ice for 5 mins. The cells were heat-shocked at 42$^o$C for 30 s and then incubated on ice for a further 5 min. 0.2 mL of LB medium was added to the reaction and the cells were incubated at 37$^o$C for 1 h. The cells were then plated out onto LB agar + antibiotic and incubated overnight at 37$^o$C

### A.1.8.2.6. Restriction Digest

Colonies from the transformation of XL-10 gold cells were picked to inoculate 5 mL cultures of LB + antibiotic and incubated overnight at 37$^o$C before carrying out a mini-prep. Extracted DNA was then digested in a solution comprising: 1 μg DNA, 1 μL 10X buffer, 1 μL restriction enzyme 1, 1 μL restriction enzyme 2, and water up to 10 μL, before being incubated at 37$^o$C for 3 h. The mixture was then run out on an agarose gel.

### A.1.9. DNA Sequencing

Sanger sequencing of cloned genes and PCR products was carried out by the Technology Facility (University of York).

### A.1.10. Purification of Plasmid DNA Using a Mini-Prep Kit

A single colony was picked from an antibiotic selective plate and used to inoculate a 5 mL overnight culture. The 5 mL culture was harvested using a microcentrifuge for 1 min at 13,000 x g, and the cell pellet resuspended in 250 μL of P1 (resuspension buffer). 250 μL of P2 (lysis buffer) was added to the suspension and mixed thoroughly before the addition of 350 μL N3 (neutralisation buffer). The solution was centrifuged for 10 min at 13,000 x g and the supernatant applied to a QIAprep spin column. Following centrifugation for 1 min at 13,000 x g, the spin column was washed with 0.5 mL PB (binding buffer) and 0.75 mL PE (wash buffer), before eluting the DNA with 50 μL water after 1 min incubation.

### A.1.11. Agarose Gel Analysis of DNA

An agarose gel was created by heating 0.5 g of agarose in a 50 mL solution of Tris-acetate buffer pH 8.0 containing EDTA, using a microwave until dissolved to create a 1% gel.  Upon cooling and prior to the gel setting, 2 μL of the DNA stain Sybr safe® was added. 8 μL samples were mixed with 2 μL of loading buffer (New England Biolabs) prior to loading onto the gel in addition to a DNA ladder (New England Biolabs), before the gel was run at 100 V and 400 mA for 1 h.

### A.1.12. Extraction of DNA from Agarose Gels

DNA bands were cut from agarose gels with a scalpel and extracted using a gel extraction kit (Sigma).  The gel slice was then dissolved using a gel solubilisation solution (300 μL solution / 100 mg gel) at $55^{o}$C for 10 min before precipitating the DNA with isopropanol (100 μL / 100 mg gel).  The DNA binding column was prepared by adding 0.5 mL of preparation solution and centrifuging (13,000 x g, 1 min) before binding the DNA and further centrifuging for 1 min.  The column was washed and the DNA eluted in 50 μL of distilled water at $65^{o}$C before the concentration was determined using a nanodrop spectrophotometer.

### A.1.13. Preparation of Competent Cells

A single colony was picked from an antibiotic selective plate and used to inoculate an overnight culture.  This was used to create a new 20 mL culture following a 1:100 dilution and incubation for 2 hours at $37^{o}$C.  Following incubation on ice for 15 min the cells were harvested (4,000 x g, 10 min, $4^{o}$C).  The cells were resuspended in 1 mL ice-cold TSS (10% PEG 3.35K, 50 mM $MgCl_2$, 5% DMSO, pH 6.5) before being aliquoted into 50 μL fractions, following which they were snap frozen in liquid $N_2$ and stored at $-80^{o}$C.

### A.1.14. Expression of Cloned Genes in *E. coli*

A 4 μL sample of the LIC construct containing the insert was added to 50 μL of competent *E. coli* cells (BL21, TUNER or Rosetta-Gami 2), and then incubated on ice for 5 min. The cells were heat-shocked at $42^{o}$C for 30 s and then incubated on ice for a further 5 min.  0.2 mL LB medium was added to the cells, which were then incubated at $37^{o}$C for 1 h.

The cells were plated out onto LB agar + kanamycin (30 μg/mL) (also + chloramphenicol 30 μg/mL when using Rosetta-Gami 2 cells) or ampicillin (100 μg/mL), and incubated overnight at 37°C. Expression tests were carried out by picking colonies from these plates to inoculate 5 mL overnight cultures of LB + kanamycin (+ chloramphenicol). Cultures were set up with a 1:100 overnight culture:media dilution, and incubated at 37°C until reaching an optical density of $A_{600}$ = 0.7 when they were induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) (0.1-1 mM). Cultures were then incubated overnight at a range of temperatures between 16 and 37°C.

### A.1.15. Lysis of *S. cerevisiae*

A *S. Cerevisiae* cell pellet was re-suspended (to a $OD_{600}$ 50-100) in a breaking buffer consisting of 50 mM sodium phosphate, pH 7.4, 5% glycerol and 1 mM PMSF. An equal volume of acid-washed glass beads was added, before vortexing the mixture for 30 sec, followed by 30 sec on ice. This was repeated four times to lyse the cells by shear force.

### A.1.16. Purification Techniques
#### A.1.16.1. Immobilised Metal Affinity Chromatography

The purification was carried out using Hi Trap Chelating HP columns with a 5 mL volume from GE Healthcare charged with $Ni^{2+}$. The column was equilibrated with 5 column volumes of starting buffer, followed by 5 column volumes of the starting buffer with 20mM imidazole. The bound protein was eluted by an increasing gradient of the starting buffer with 20mM imidazole and the starting buffer with 1 M imidazole, over 20 column volumes.

#### A.1.16.2. Size Exclusion Chromatography

The gel filtration column was pre-incubated in 150 mM NaCl prior to the protein sample being injected onto the column. The column was run at a constant rate until all proteins had been eluted.

### A.1.16.3. Ion Exchange Chromatography

The purification was carried out using Hi Trap Q HP columns with a 1 mL volume from GE Healthcare.  Prior to use, the column was equilibrated with 5 column volumes of buffer (20 mM $Na_2HPO_4$, pH 7.3), followed by 5 column volumes of buffer B (20 mM $Na_2HPO_4$, pH 7.3, 1 M NaCl) followed by 5 column volumes buffer A.  The protein sample was loaded onto the column.  The column was washed with 5 column volumes prior to the protein being eluted with an increasing gradient to buffer B over 20 column volumes.

## A.1.17. Protein Analytical Techniques

### A.1.17.1. Matrix-Assisted Laser Desorption/Ionisation (MALDI)

Samples were sent for MALDI mass spectrometry analysis at Central Science Laboratory (York).

### A.1.17.2. Electrospray Mass Spectrometry

Samples were sent for Electrospray mass spectrometry analysis at the Technology Facility (University of York).

### A.1.17.3. Protein Identification

Protein samples were analysed using protein identification (protein fingerprinting) by the Technology Facility (University of York).

## A.1.18. Crystal Cryoprotection

Crystals were cryoprotected by soaking sequentially in cryoprotectants (glycerol-mother liquor solutions) or straight into various oils, prior to being vitrified in liquid nitrogen.  Alternatively, crystals were vitrified without a cryoprotectant.

## B.1. Molecular Biology

### B.1.1. L1 Expression in *E. coli*

#### B.1.1.1. Ligation Independent Cloning

```
                  10                      20                      30
H L Y P G E V C P G   M D I R N N L T R L   H E L E N C S V I E
                  40                      50                      60
G H L Q I L L M F K   T R P E D F R D L S   F P K L I M I T D I
                  70                      80                      90
L L L F R V Y G L E   S L K D L F P N L T   V I R G S R L F F N
                 100                     110                     120
Y A L V I F E M V H   L K E L G L Y N L M   N I T R G S V R I E
                 130                     140                     150
K N N E L C Y L A T   I D W S R I L D S V   E D N H I V L N K D
                 160                     170                     180
D N E E C G D I C P   G T A K G K T N C P   A T V I N G Q F V E
                 190
R C W T H S H C Q K
```

***Figure 82. The amino acid sequence of the L1 region of the human IR.***

The expressed region consists of residues Gly$^5$-Ile$^{158}$.

TACGGGGTCATTCCCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGG
CAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGAAGTTCTGTTCCAGGGACCAGC
A<span style="color:red">GGTGAAGTTTGTCCGGGTATGG</span>ATATTCGTAATAATCTGACCCGTCTGCATGAACTGGAA
AATTGCAGCGTGATTGAAGGTCATCTGCAAATTCTGCTGATGTTTAAAACCCGTCCGGAA
GATTTTCGTGATCTGAGCTTTCCGAAACTGATTATGATTACCGATATTCTGCTGCTGTTTCG
TGTTTATGGTCTGGAAAGCCTGAAAGACCTGTTTCCGAATCTGACCGTTATTCGTGGTAGC
CGTCTGTTTTTTAATTATGCCCTGGTGATTTTTGAAATGGTGCATCTGAAAGAACTGGGTC
TGTATAATCTGATGAATATTACCCGTGGTAGCGTGCGTATTGAAAAAAATAATGAACTGTG
CTATCTGGCAACCATTGATTGGAGCCGTATTCTGGATTCTGTGGAAGATAATCATATT<span style="color:red">GTGC</span>
<span style="color:red">TGAATAAAGATGATAATGAAGAATGCGGTGATTAA</span>CGCGCCTTCTCCTCACATATGGCTAG
CATGACTGGTGGACAGCAAATGGGTCGCGGATCCGAATTCGAGCTCCGTCGACAAGCTT
GCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCG
AAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGG
CCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGA
ATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCG
TGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCTTCCTTTCTC
GCCACGTTCGCGGCTTTCCCCGTCAGCTCTAATCGGGGCTCCTTAGGTTCCGATTTATGCT
TACGCACTCGACCCAAAAACTTGATAGTGATGTCACGTATGGCATCGTTGATGACCGATTT
CGCCTTGAGCTGGATCAGTCTATATGACTGTCACTGACACCTACCTATTCCGGCAA

*Figure 83. The base pair sequence of the L1 region.*

Primer regions are indicated in red.


Forward CCAGGGACCAGCAGGTGAAGTTTGTCCGGGTATGG

Reverse GAGGAGAAGGCGCGTTAATCACCGCATTCTTCATTATCATCTTTATTCAGCAC

*Figure 84. Primers for LIC.*

## B.1.1.2. Molecular Biology

```
              10                    20                    30
H L Y P G E V C P G   M D I R N N L T R L   H E L E N C S V I E
              40                    50                    60
G H L Q I L L M F K   T R P E D F R D L S   F P K L I M I T D I
              70                    80                    90
L L L F R V Y G L E   S L K D L F P N L T   V I R G S R L F F N
             100                   110                   120
Y A L V I F E M V H   L K E L G L Y N L M   N I T R G S V R I E
             130                   140                   150
K N N E L C Y L A T   I D W S R I L D S V   E D N H I V L N K D
             160                   170                   180
D N E E C G D I C P   G T A K G K T N C P   A T V I N G Q F V E
             190
R C W T H S H C Q K
```

*Figure 85. The amino acid sequence of the L1 region of the human IR.*

Nco    ATCTGTCCATGGGTGAAGTTTGTCCGGGTATGG

Eco    CATCTGGAATTCGGTGAAGTTTGTCCGGGTATGG

xho    ACCCGGCTCGAGATCACCGCATTCTTCATTATCATCTTTATTCAG

Hind    CTGTACAAGCTTATCAATCACCGCATTCTTCATTATCATCTTTATTCAG

*Figure 86. Primers used with the pET22 and pMAL vectors.*

Primer pairs were Eco-Hind, Nco-Hind and Nco-Xho.

### B.1.2. L1 *S. cerevisiae*
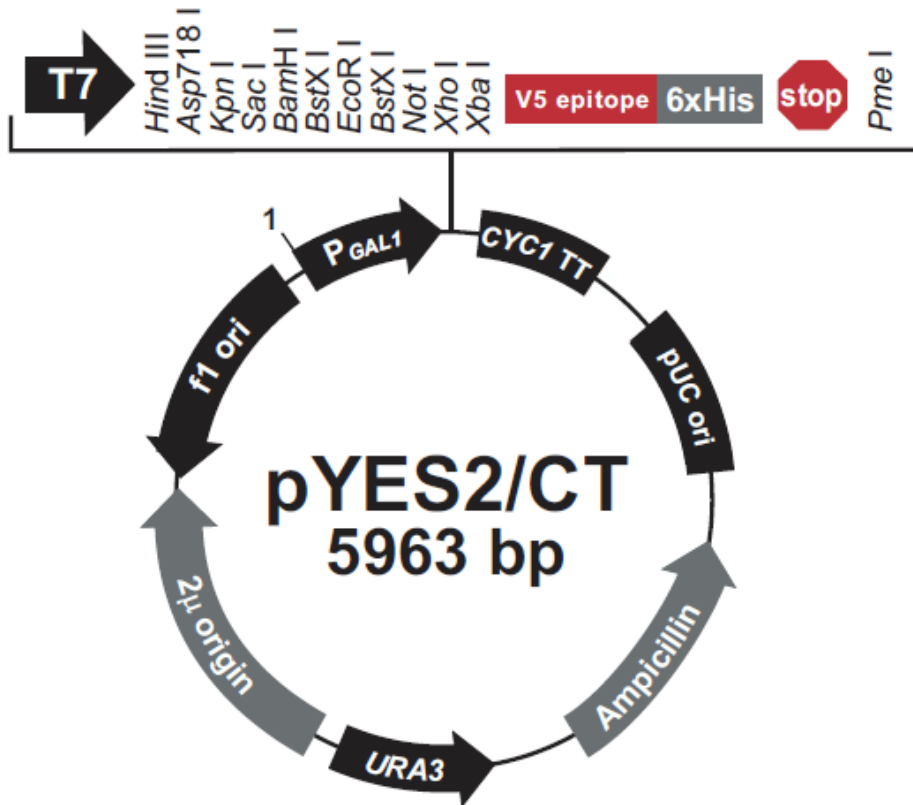


*Figure 87. Vector map for L1 expression in S. cerevisiae.*

CATTTGTATCCAGGTGAAGTTTGTCCTGGTATGGATATCAGAAACAACTTA
ACAAGATTGCATGAATTAGAAAACTGTTCTGTTATTGAAGGTCATTTGCAA
ATCTTGTTGATGTTCAAGACTAGACCAGAAGATTTCAGAGATTTGTTTTTC
CTAAGTTGATCATGATCACAGATATCTTGTTGTTGTTTAGAGTTTACGGTTT
GGAATCCTTAAAGGATTTGTTCCCAAATTTGACCGTTATTAGAGGTAGTAG
ATTATTTTTCAACTACGCTTTGGTTATTTTCGAAATGGTTCATTTGAAGGAA
TTGGGTTTGTACAATTTGATGAACATTACCAGAGGTTCCGTTAGAATCGAA
AAGAATAACGAATTGTGTTACTTGGCAACTATCGATTGGTCTAGAATTTTA
GATTCAGTTGAAGATAACCATATTGTTTTGAATAAGGATGATAACGAAGAA
TGTGGTGACATTTGTCCAGGTACTGCTAAAGGTAAAACAAATTGTCCTGC
AACCGTTATTAACGGTCAATTCGTTGAAAGATGTTGGACTCATTCACATTG
TCAAAAGTAA

*Figure 88. Optimised base pair sequence of the L1 region for expression in S. cerevisiae.*

Forward          TTTTAAGCTTAAAATGTCTGGTGAAGTTTGTCCTGGTATGG

Reverse          TTTTCTCGAGGTCACCACATTCTTCGTTATCATCC

*Figure 89. Primers used for S. cerevisiae L1 sequence.*

### B.1.3. L1 *P. pastoris*
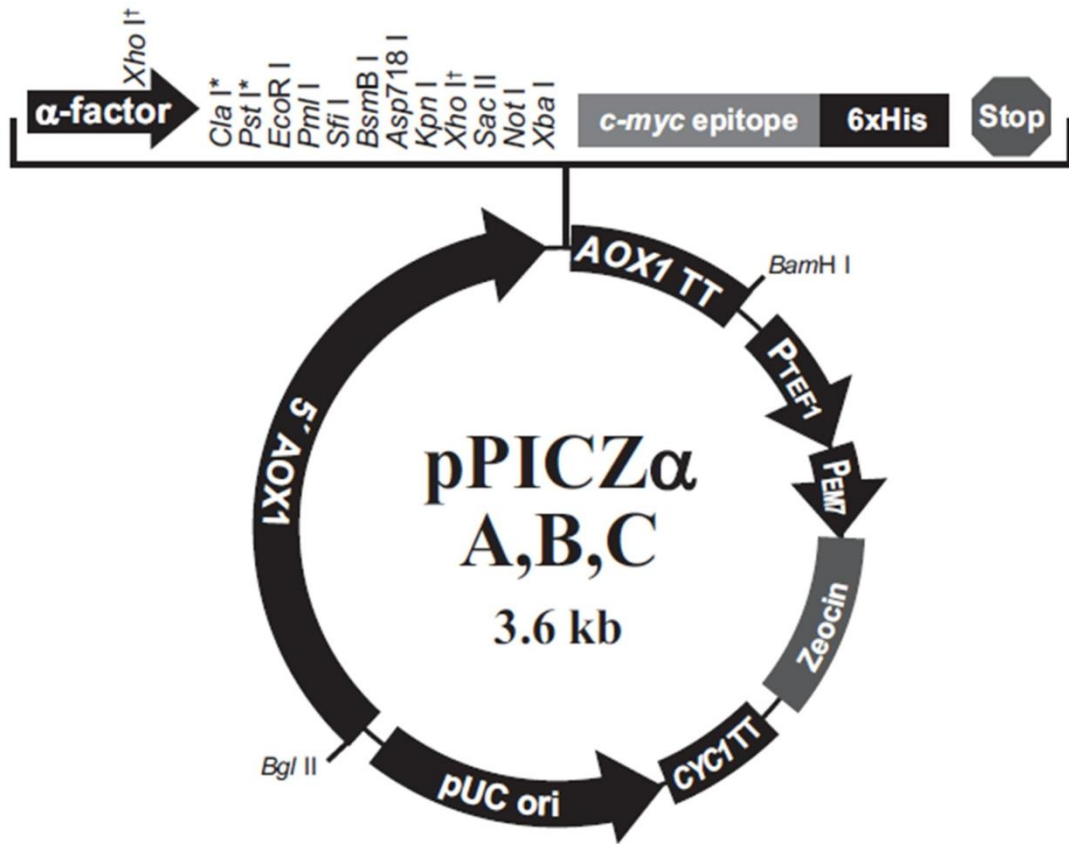


*Figure 90. Vector map for L1 expression in P. pastoris.*

```
          10                    20                    30
H L Y P G E V C P G   M D I R N N L T R L   H E L E N C S V I E
          40                    50                    60
G H L Q I L L M F K   T R P E D F R D L S   F P K L I M I T D I
          70                    80                    90
L L L F R V Y G L E   S L K D L F P N L T   V I R G S R L F F N
         100                   110                   120
Y A L V I F E M V H   L K E L G L Y N L M   N I T R G S V R I E
         130                   140                   150
K N N E L C Y L A T   I D W S R I L D S V   E D N H I V L N K D
         160                   170                   180
D N E E C G D I C P   G T A K G K T N C P   A T V I N G Q F V E
         190
R C W T H S H C Q K
```

*Figure 91. Gene sequence of the L1 region in human IR.*

The expressed region consists of residues Gly[5]-Ile[158].

CATTTGTATCCAGGTGAAGTTTGTCCTGGTATGGATATCAGAAACAACTTAACAA
GATTGCATGAATTAGAAAACTGTTCTGTTATTGAAGGTCATTTGCAAATCTTGTTG
ATGTTCAAGACTAGACCAGAAGATTTCAGAGATTTGTCTTTTCCTAAGTTGATCAT
GATCACAGATATCTTGTTGTTGTTTAGAGTTTACGGTTTGGAATCCTTAAAGGATT
TGTTCCCAAATTTGACCGTTATTAGAGGTAGTAGATTATTTTTCAACTACGCTTTG
GTTATTTTCGAAATGGTTCATTTGAAGGAATTGGGTTTGTACAATTTGATGAACAT
TACCAGAGGTTCCGTTAGAATCGAAAAGAATAACGAATTGTGTTACTTGGCAACT
ATCGATTGGTCTAGAATTTTAGATTCAGTTGAAGATAACCATATTGTTTTGAATAA
GGATGATAACGAAGAATGTGGTGACATTTGTCCAGGTACTGCTAAAGGTAAAAC
AAATTGTCCTGCAACCGTTATTAACGGTCAATTCGTTGAAAGATGTTGGACTCAT
TCACATTGTCAAAAGTAA

*Figure 92. Gene sequence of the L1 region for expression in P. pastoris.*

Primer locations are marked in red.


Forward        CTGAATTCGGTGAAGTTTGTCCTGGTATGG

Reverse        GTTCTAGAATGGTCCCTGGAACAAAACCTCCAATCCGTCACCA

                CATTCTTCGTTATCATCC

*Figure 93. Primers used for L1 expression in P. pastoris.*

## C.1. Insulin Screens

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| A | $Na_2SO_4$ 0.05M pH 3.0 | $Na_2SO_4$ 0.10M pH 3.0 | $Na_2SO_4$ 0.15M pH 3.0 | $Na_2SO_4$ 0.20M pH 3.0 | $Na_2SO_4$ 0.25M pH 3.0 | $Na_2SO_4$ 0.35M pH 3.0 |
| B | $Am_2SO_4$ 0.05M pH 3.0 | $Am_2SO_4$ 0.10M pH 3.0 | $Am_2SO_4$ 0.15M pH 3.0 | $Am_2SO_4$ 0.20M pH 3.0 | $Am_2SO_4$ 0.25M pH 3.0 | $Am_2SO_4$ 0.35M pH 3.0 |
| C | $Li_2SO_4$ 0.010M pH 3.0 | $Li_2SO_4$ 0.015M pH 3.0 | $Li_2SO_4$ 0.020M pH 3.0 | $Cs_2SO_4$ 0.01M pH 3.0 | $Cs_2SO_4$ 0.015M pH 3.0 | $Cs_2SO_4$ 0.020M pH 3.0 |
| D | Hexameric Condition A 0.1M Na Citrate, 0.3M Tris pH 8.2, 0.6mM ZnAc, 0.06% Phenol | Hexameric Condition B 0.6M $Na_2SO_4$, 0.3M Tris pH 8.2, 0.6mM ZnAc, 0.06% Phenol | Hexameric Condition C 0.1M Na Citrate, 0.3M Tris pH 7.5, 0.6mM ZnAc, 0.06% Phenol | Hexameric Condition D 0.6M $Na_2SO_4$, 0.3M Tris pH 7.5, 0.6mM ZnAc, 0.06% Phenol | Hexameric Condition E 5mM ZnAc, 35mM Na Citrate, 0.7% Phenol, 0.7M NaCl, 0.3M Tris pH 7.5 | Hexameric Condition F 5mM ZnAc, 35mM Na Citrate, 0.7% Phenol, 1.1M NaCl, 0.3M Tris pH 7.5 |

*Figure 94. Insulin screen 1*

An insulin screen developed to form monomeric and hexameric insulin molecules.

|   | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| A | 0.01M Mg Formate 0.1M NaAc pH 3.0 | 0.1M Na Formate 0.1M NaAc pH 3.0 | 0.1M Na Citrate 0.1M NaAc pH 3.0 | 0.01M Li Formate 0.1M NaAc pH 3.0 | 0.01M Rb$_2$SO$_4$ 0.1M NaAc pH 3.0 | 0.01M K$_2$SO$_4$ 0.1M NaAc pH 3.0 |
| B | 0.02M Mg Formate 0.1M NaAc pH 3.0 | 0.2M Na Formate 0.1M NaAc pH 3.0 | 0.2M Na Citrate 0.1M NaAc pH 3.0 | 0.02M Li Formate 0.1M NaAc pH 3.0 | 0.05M Rb$_2$SO$_4$ 0.1M NaAc pH 3.0 | 0.05M K$_2$SO$_4$ 0.1M NaAc pH 3.0 |
| C | 0.06M Mg Formate 0.1M NaAc pH 3.0 | 0.5M Na Formate 0.1M NaAc pH 3.0 | 0.5M Na Citrate 0.1M NaAc pH 3.0 | 0.06M Li Formate 0.1M NaAc pH 3.0 | 0.1M Rb$_2$SO$_4$ 0.1M NaAc pH 3.0 | 0.1M K$_2$SO$_4$ 0.1M NaAc pH 3.0 |
| D | 0.1M Mg Formate 0.1M NaAc pH 3.0 | 0.8M Na Formate 0.1M NaAc pH 3.0 | 0.8M Na Citrate 0.1M NaAc pH 3.0 | 0.1M Li Formate 0.1M NaAc pH 3.0 | 0.3M Rb$_2$SO$_4$ 0.1M NaAc pH 3.0 | 0.3M K$_2$SO$_4$ 0.1M NaAc pH 3.0 |

*Figure 95. Insulin screen 2.*

An insulin screen developed to form monomeric insulin molecules.

|   | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| A | 20% PEG 3350 0.2M NaCl +10%1M Tris.HCl pH 8.5 | 20% PEG 3350 0.2M NaCl 3mM $ZnCl_2$ +10%1M Tris.HCl pH 8.5 | 20% PEG 3350 0.2M NaCl 3mM $ZnCl_2$ 0.7% Phenol +10%1M Tris.HCl pH 8.5 | 40% MPD 0.2M NaCl +10%1M Tris.HCl pH 8.5 | 40% MPD 0.2M NaCl 3mM $ZnCl_2$ +10%1M Tris.HCl pH 8.5 | 40% MPD 0.2M NaCl 3mM $ZnCl_2$ 0.7% Phenol +10%1M Tris.HCl pH 8.5 |
| B | 20% PEG 3350 0.2M NaBr +10%1M Tris.HCl pH 8.5 | 20% PEG 3350 0.2M NaBr 3mM $ZnCl_2$ +10%1M Tris.HCl pH 8.5 | 20% PEG 3350 0.2M NaBr 3mM $ZnCl_2$ 0.7% Phenol +10%1M Tris.HCl pH 8.5 | 40% MPD 0.2M NaBr +10%1M Tris.HCl pH 8.5 | 40% MPD 0.2M NaBr 3mM $ZnCl_2$ +10%1M Tris.HCl pH 8.5 | 40% MPD 0.2M NaBr 3mM $ZnCl_2$ 0.7% Phenol +10%1M Tris.HCl pH 8.5 |
| C | 20% PEG 3350 0.2M NaCitrate +10%1M Tris.HCl pH 8.5 | 20% PEG 3350 0.2M NaCitrate 3mM $ZnCl_2$ +10%1M Tris.HCl pH 8.5 | 20% PEG 3350 0.2M NaCitrate 3mM $ZnCl_2$ 0.7% Phenol +10%1M Tris.HCl pH 8.5 | 40% MPD 0.2M NaCitrate +10%1M Tris.HCl pH 8.5 | 40% MPD 0.2M NaCitrate 3mM $ZnCl_2$ +10%1M Tris.HCl pH 8.5 | 40% MPD 0.2M NaCitrate 3mM $ZnCl_2$ 0.7% Phenol +10%1M Tris.HCl pH 8.5 |
| D | 20% PEG 3350 0.2M $NaNO_3$ +10%1M Tris.HCl pH 8.5 | 20% PEG 3350 0.2M $NaNO_3$ 3mM $ZnCl_2$ +10%1M Tris.HCl pH 8.5 | 20% PEG 3350 0.2M $NaNO_3$ 3mM $ZnCl_2$ 0.7% Phenol +10%1M Tris.HCl pH 8.5 | 40% MPD 0.2M $NaNO_3$ +10%1M Tris.HCl pH 8.5 | 40% MPD 0.2M $NaNO_3$ 3mM $ZnCl_2$ +10%1M Tris.HCl pH 8.5 | 40% MPD 0.2M $NaNO_3$ 3mM $ZnCl_2$ 0.7% Phenol +10%1M Tris.HCl pH 8.5 |
| E | 20% PEG 3350 0.2M NaCl +10%1M Glycine.HCl pH 3.0 | 20% PEG 3350 0.2M NaCl 3mM $ZnCl_2$ +10%1M Glycine.HCl pH 3.0 | 20% PEG 3350 0.2M NaCl 3mM $ZnCl_2$ 0.7% Phenol +10%1M Glycine.HCl pH 3.0 | 40% MPD 0.2M NaCl +10%1M Glycine.HCl pH 3.0 | 40% MPD 0.2M NaCl 3mM $ZnCl_2$ +10%1M Glycine.HCl pH 3.0 | 40% MPD 0.2M NaCl 3mM $ZnCl_2$ 0.7% Phenol +10%1M Glycine.HCl pH 3.0 |
| F | 20% PEG 3350 0.2M NaBr +10%1M Glycine.HCl pH 3.0 | 20% PEG 3350 0.2M NaBr 3mM $ZnCl_2$ +10%1M Glycine.HCl pH 3.0 | 20% PEG 3350 0.2M NaBr 3mM $ZnCl_2$ 0.7% Phenol +10%1M Glycine.HCl pH 3.0 | 40% MPD 0.2M NaBr +10%1M Glycine.HCl pH 3.0 | 40% MPD 0.2M NaBr 3mM $ZnCl_2$ +10%1M Glycine.HCl pH 3.0 | 40% MPD 0.2M NaBr 3mM $ZnCl_2$ 0.7% Phenol +10%1M Glycine.HCl pH 3.0 |
| G | 20% PEG 3350 0.2M NaCitrate +10%1M Glycine.HCl pH 3.0 | 20% PEG 3350 0.2M NaCitrate 3mM $ZnCl_2$ +10%1M Glycine.HCl pH 3.0 | 20% PEG 3350 0.2M NaCitrate 3mM $ZnCl_2$ 0.7% Phenol +10%1M Glycine.HCl pH 3.0 | 40% MPD 0.2M NaCitrate +10%1M Glycine.HCl pH 3.0 | 40% MPD 0.2M NaCitrate 3mM $ZnCl_2$ +10%1M Glycine.HCl pH 3.0 | 40% MPD 0.2M NaCitrate 3mM $ZnCl_2$ 0.7% Phenol +10%1M Glycine.HCl pH 3.0 |
| H | 20% PEG 3350 0.2M $NaNO_3$ +10%1M Glycine.HCl pH 3.0 | 20% PEG 3350 0.2M $NaNO_3$ 3mM $ZnCl_2$ +10%1M Glycine.HCl pH 3.0 | 20% PEG 3350 0.2M $NaNO_3$ 3mM $ZnCl_2$ 0.7% Phenol +10%1M Glycine.HCl pH 3.0 | 40% MPD 0.2M $NaNO_3$ +10%1M Glycine.HCl pH 3.0 | 40% MPD 0.2M $NaNO_3$ 3mM $ZnCl_2$ +10%1M Glycine.HCl pH 3.0 | 40% MPD 0.2M $NaNO_3$ 3mM $ZnCl_2$ 0.7% Phenol +10%1M Glycine.HCl pH 3.0 |

*Figure 96. Insulin screen 3.*

An insulin screen developed to form monomeric and hexameric insulin molecules.

# Appendix 2. The Use of Insulin-Like Growth Factor as a Basis for New Novel Insulin Analogues

## A.2.1. Summary

    IGF-I and IGF-II are single-chain polypeptides with molecular masses of 7.6 kDa and 7.5 kDa respectively [9,10]. There is a cross-over between insulin and IGF, so it is important to investigate the relationship between these two molecules (Figure 97). Creating an insulin-like core from IGF, could lead to an understanding of this cross signalling.
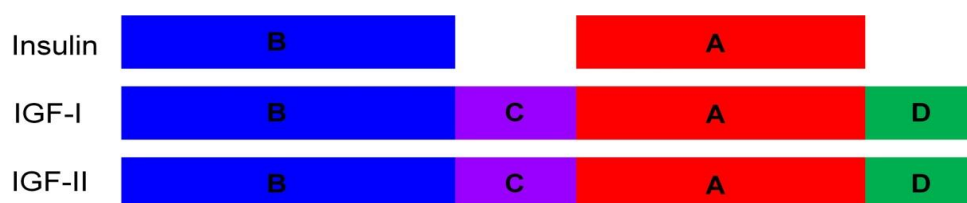


***Figure 97.  The domain layout of insulin, IGF-I and IGF-II.***

Regions labelled A and B are based upon the chain labels of insulin. Region C is the linking peptide of insulin that is proteolytically removed but remains in IGF, and region D is the signal peptide of insulin that is removed.

    The cleavage sites of trypsin in IGF (Figure 98) are spread throughout the molecule and it is possible to cleave IGF into an insulin-like core region using trypsin. This is because the disulphide bonds in insulin ($A^6$-$A^{11}$, $A^7$-$B^7$ and $A^{20}$-$B^{19}$) correspond almost identically to those seen in IGF (Figure 98). Following trypsin digest, part of chains C and D (Figure 97) are be removed leaving chains A and B.
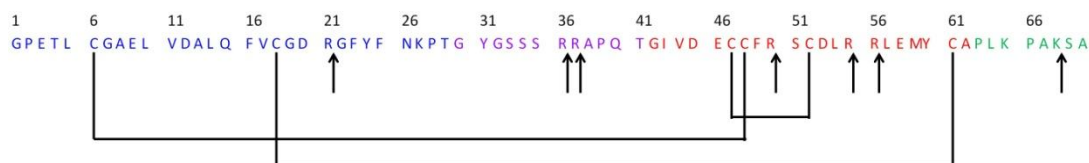


***Figure 98.  IGF disulphide bonds and trypsin cleavage sites.***

The arrows indicate trypsin cleavage sites, and disulphide bonds are indicated by the joined cysteine residues. The letter colours correspond to the regions in Figure 97.

## A.2.2. Methods

A IGF:trypsin digest was set up in a 50:1 molecular ratio respectively in 0.3 M Tris pH 9.2. The samples were incubated for a specified period of time at $37^{\circ}$C.  The digestion was halted using PMSF at a 0.2% w/v concentration.

## A.2.3. Results and Discussion

To ensure the starting material was at the expected molecular weight, a sample of IGF was analysed by electrospray mass spectrometry. The mass of IGF was confirmed to be 7649 Da (Figure 99).
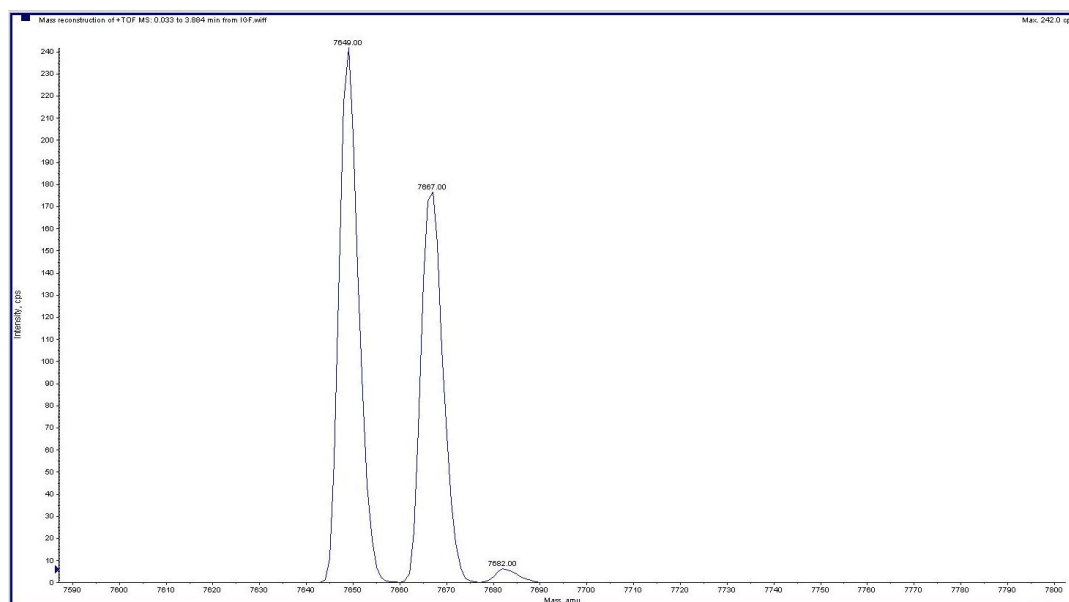


***Figure 99. Electrospray mass spectrometry of IGF.***

To promote the proteolytic cleavage, urea and guanidine hydrochloride were added at various concentrations to aid the availability of the tryptic cleavage sites. The initial digestions were run on SDS-PAGE in a non-reducing sample buffer to ascertain the completeness of digestion (Figure 100 and Figure 101).
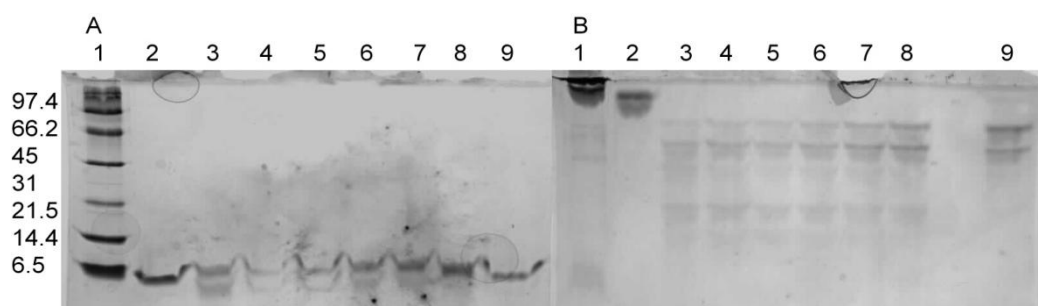


***Figure 100. IGF digestion with various concentrations of urea.***

A 17.5% SDS-PAGE gel IGF digestion with various concentrations of urea with SDS in the running buffer (A) and without (B). Gels (A) and (B) lanes 2-9 contain IGF. Gel (A), Lane 1: Broad range molecular weight marker. Lane 2: IGF only. Lane 3: no urea. Lane 4: 50 mM urea. Lane 5: 100 mM urea. Lane 6: 200 mM urea. Lane 7: 500 mM urea. Lane 8: 1 M urea. Lane 9: 2 M urea. Gel (B), Lane 1: Broad range molecular weight marker. Lane 2: IGF only. Lane 3: no urea. Lane 4: 50 mM urea. Lane 5: 100 mM urea. Lane 6: 200 mM urea. Lane 7: 500 mM urea. Lane 8: 1 M urea. Lane 9: 2 M urea.
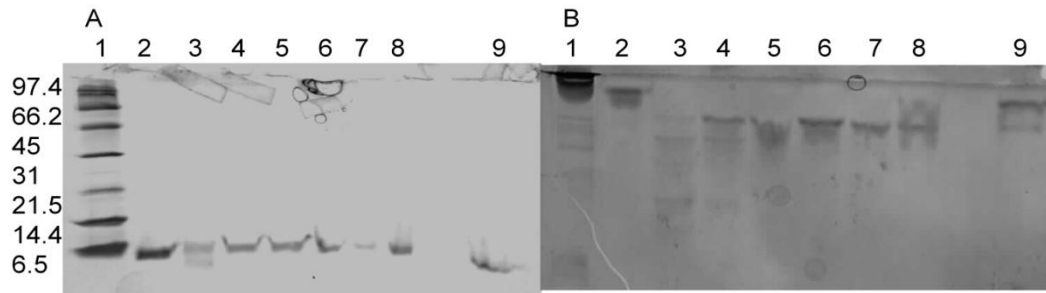
*Figure 101. IGF digestion with various concentrations of guanidine hydrochloride.*

A 17.5% SDS-PAGE gel showing the digestion of IGF with various concentrations of guanidine hydrochloride with SDS in the running buffer (A) and without (B). Gels (A) and (B) lanes 2-9 contain IGF. Gel (A), Lane 1: Broad range molecular weight ladder. Lane 2: IGF only. Lane 3: no guanidine hydrochloride. Lane 4: 50 mM guanidine hydrochloride. Lane 5: 100 mM guanidine hydrochloride. Lane 6: 200 mM guanidine hydrochloride. Lane 7: 500 mM guanidine hydrochloride. Lane 8: 1 M guanidine hydrochloride. Lane 9: 2 M guanidine hydrochloride. Gel (B), Lane 1: Broad range molecular weight marker. Lane 2: IGF only. Lane 3: no guanidine hydrochloride. Lane 4: 50 mM guanidine hydrochloride. Lane 5: 100 mM guanidine hydrochloride. Lane 6: 200 mM guanidine hydrochloride. Lane 7: 500 mM guanidine hydrochloride. Lane 8: 1 M guanidine hydrochloride. Lane 9: 2 M guanidine hydrochloride.

Digestion occurs at all concentrations of urea, although further digestion also seems to occur past the desired level, with multiple breakdown products present in the sample. The samples containing guanidine hydrochloride seem to allow the trypsin to digest the IGF completely. The guanidine hydrochloride samples were investigated to determine at which point in a time course, the optimum digestion occurs (Figure 102).
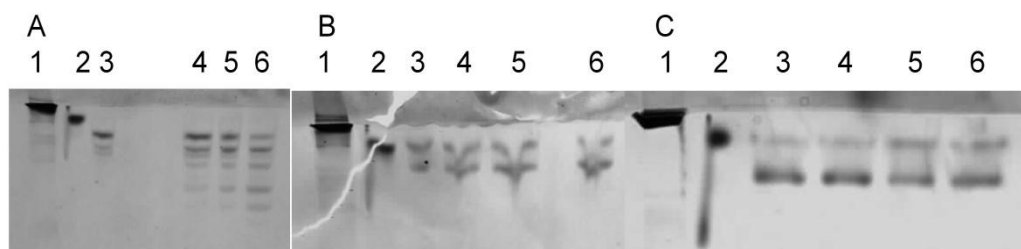
***Figure 102. Time course digestion of IGF with the addition of various concentrations of guanidine hydrochloride.***

A 15% SDS-PAGE gel showing the digestion of a IGF time course in various concentrations of guanidine hydrochloride. Gels (A), (B) and (C) lanes 2-6 contain IGF, gels were run without SDS in the running buffer. Gel (A) contains no guanidine hydrochloride, Lane 1: Broad range molecular weight marker.  Lane 2: IGF only.  Lane 3: 15 min digestion.  Lane 4: 30 min digestion.  Lane 5: 60 min digestion.  Lane 6: 120 min digestion. Gel (B) contains 1 M guanidine hydrochloride, Lane 1: Broad range molecular weight marker.  Lane 2: IGF only.  Lane 3: 15 min digestion.  Lane 4: 30 min digestion. Lane 5: 60 min digestion.  Lane 6: 120 min digestion. Gel (C) contains 2 M guanidine hydrochloride, Lane 1: Broad range molecular weight marker.  Lane 2: IGF only.  Lane 3: 15 min digestion.  Lane 4: 30 min digestion.  Lane 5: 60 min digestion.  Lane 6: 120 min digestion.

It appears that the best digestion occurs at 2 M guanidine hydrochloride for 15 min, so this was used and then analysed by MALDI. Multiple peaks could be seen at 5589, 5745, 5901 and 6057 Da (Figure 103).   The difference in the peaks was averaged to 156 Da.  This difference can be attributed to adducts of de-fluorinated PMSF.  The IGF fragment has multiple PMS adducts, leading to the observation of multiple peaks (Figure 103-A). The PMS adducts were removed by repeated washing steps with 2 mM Tris (Figure 103-B).
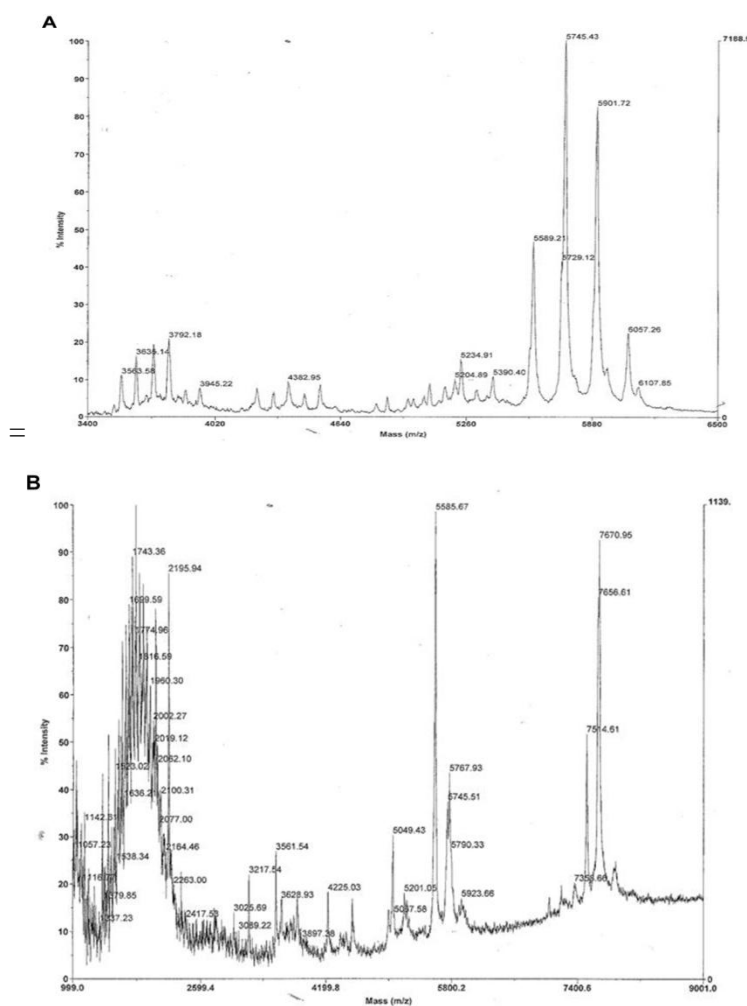
***Figure 103. MALDI sample of IGF digested in the presence of 2 M guanidine-hydrochloride for 15 minutes.***

(A) MALDI sample of IGF digested in 2 M guanidine hydrochloride for 15 min. (B) MALDI sample of IGF digested in 2 M guanidine hydrochloride for 15 min following washing, and containing an IGF internal calibrant.

Trypsin digestion of the IGF molecule can lead to various length peptides dependent on where the trypsin cuts (Figure 98, Table 19). The major species present following digestion had a mass of 5585 Da (Figure 103-B), and were analysed to investigate which core fragments remained following cleavage (Table 20).

***Table 19. Lengths and weights of all possible fragments of IGF.***

| Fragment/Identity | 1-21 | 22-36 | 37 | 38-50 | 51-55 | 56 |
|---|---|---|---|---|---|---|
| Molecular Weight (Da) | 2193.4 | 1667.8 | 174.2 | 1438.6 | 592.6 | 174.2 |

| Fragment/Identity | 57-68 | 69-70 | 38-55 | 57-70 | Sodium | Hydroxylation |
|---|---|---|---|---|---|---|
| Molecular Weight (Da) | 1363.7 | 176.2 | 2013.2 | 1521.8 | 23.0 | 16.0 |

*Table 20. Analysis of IGF MALDI digest data.*

| MALDI Result (Da) | Fragment/Identity | Molecular Weight (Da) |
|---|---|---|
| | 1-21 | 2193.4 |
| | 38-55 | 2013.20 |
| 5585 | 57-68 | 1363.70 |
| | Hydroxylation | 16.00 |
| | Total | 5586.3 |

Following digestion, residues 1-21, 38-55 and 57-68 remain connected (Table 20). In comparison to the insulin molecule (Figure 97), where regions C and D have been partially removed from IGF leaving a core similar to insulin.

### A.2.4. Conclusion

The cross signalling of insulin and IGF requires more understanding. This knowledge could be gained by the creation of IGF-insulin hybrids. The cleavage of IGF with trypsin left an insulin-like core that comprised residues 1-21, 38-55 and 57-68 of IGF. The structural overlay of this core (representation) and insulin (Figure 104) show structures of high similarity, a good start for the study of the insulin-IGF cross signalling. The method has caused intra-chain cleavages to occur, so a crystal structure is required to confirm the actual structure of the IGF insulin-like core.



***Figure 104. Super positioning of insulin and the digested IGF.***

The insulin analogue (1mso.pdb) is coloured in green, and a representation of the digested IGF molecule (1imx.pdb) is superimposed in red. Disulphide bonds are shown in yellow.

Using trypsin to digest away the C and D regions causes several problems for further applications due to the method used. Trypsin cleavage results in broken chains, for example between residues Arg$^{50}$ and Ser$^{51}$. There are also missing residues that would form secondary structures such as Arg$^{56}$ (Figure 104). These cleavages and loss of residues could result in the loss of secondary structure, which could cause problems with receptor binding. The location of Arg$^{56}$ is in the α-helix C-terminal region equivalent to the insulin A-chain, so its loss would cause the α-helix to be destroyed. Nicks in the peptide backbone could be repaired, such as between Arg$^{50}$ and Ser$^{51}$, by a condensation reaction using trypsin in an

organic solvent[92]. The ability of this method to retain/regenerate secondary structure is unknown.

The C-peptide is cleaved from proinsulin whilst it is in hexameric form[24]. The C-peptide is located on the outside of the spherical proinsulin hexamer, and is readily available to be cleaved by proteases. Due to the similarity between proinsulin and IGF, this could be repeated protecting the A- and B-chain equivalent regions from cleavage, allowing the removal of the C-peptide. However the ability of IGF to form a hexamer is doubted[188].

Following creation of the insulin-like IGF core there are various fragments present in the reaction mixture that need to be removed from the IGF. Following the creation of semi-synthetic insulin analogues as described by Zakova *et al.*[92] the analogues are purified by RP-HPLC, a method which can be used here. Similarly gel filtration could also purify the IGF core.

The method of producing semi-synthetic insulin analogues as described by Zakova *et al.*[92] uses trypsin to remove the C-terminal octapeptide from the B-chain of the insulin. Changing the buffer to an organic solvent allows trypsin to attach a synthetic peptide to the C-terminus. The IGF insulin-like core was left in an equivalent state to insulin following cleavage using trypsin. The availability of IGF and trypsin, and the simplicity of the method used means that the IGF insulin-like core can be produced in large enough amounts to allow many custom peptides to be added to the B-chain C-terminus. This permits the investigation into the relationship between insulin binding[75] and IGF binding, and the cross signalling relationship between them.

# Appendix 3. *Drosophila melanogaster* Insulin-Like Peptide 5 and Insect Insulin Binding Protein

## A.3.1. Summary

The genome of *D. melanogaster* contains seven insulin-like genes, *dilp1-7*, that are expressed in a highly tissue- and stage-specific pattern[189]. The gene *dilp5* has 27.8 % identity to the human insulin gene[189].

*Drosophila melanogaster* insulin-like peptide 5 (DILP5) has a fold similar to other proteins in the larger insulin family such as human insulin, human relaxin and bombyxin II from *Bombyx mori*[190]. There is a greater degree of flexibility, and therefore disorder, in the C-terminus of the B-chain in DILP5 compared to insulin (Figure 105-A). The β-turn seen in insulin is in an extended conformation in DILP5 ($B^{20}$-$B^{23}$), which causes the C-terminus to be orientated differently from insulin (Figure 105-B). The B-chain N-terminus in DILP5 exhibits a fold similar to that seen in the T-state of insulin. The A-chain of DILP5 is a close mimic of insulin[190].
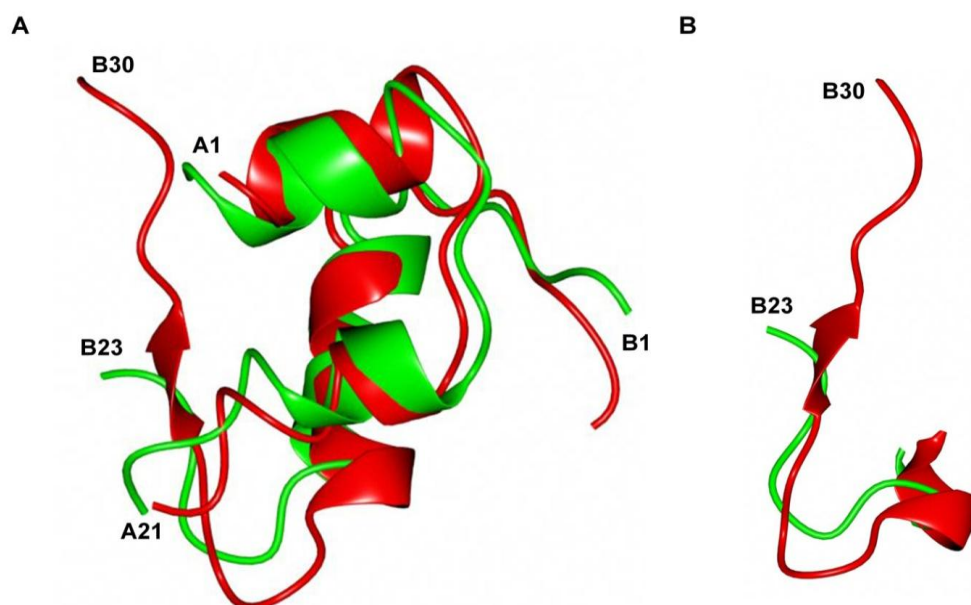


*Figure 105. A structural comparison of insulin and DILP5.*
Human insulin is shown in red (pdb used: 1mso), whilst DILP5 is in green (pdb used: 2wfv).

DILP5 is able to form dimers along a crystallographic 2-fold axis, similar to insulin. The dimerisation differs between the two. Insulin dimerises along the anti-parallel β-sheet at the B-chain C-terminus, whereas in DILP5 this occurs between the anti-parallel β-sheet present at the N-terminus of the B-chain[190].

DILP5 is able to bind to insect insulin-binding proteins (IMPL-2) (Figure 106). It has been proposed that in vertebrates IMPLs are required for metabolism and growth control[191]. The binding of DILP5 to IMPL-2 from the fall armyworm *S. frugiperda* has been proven through assays and chemical cross-linking experiments, showing specific binding[190,191].
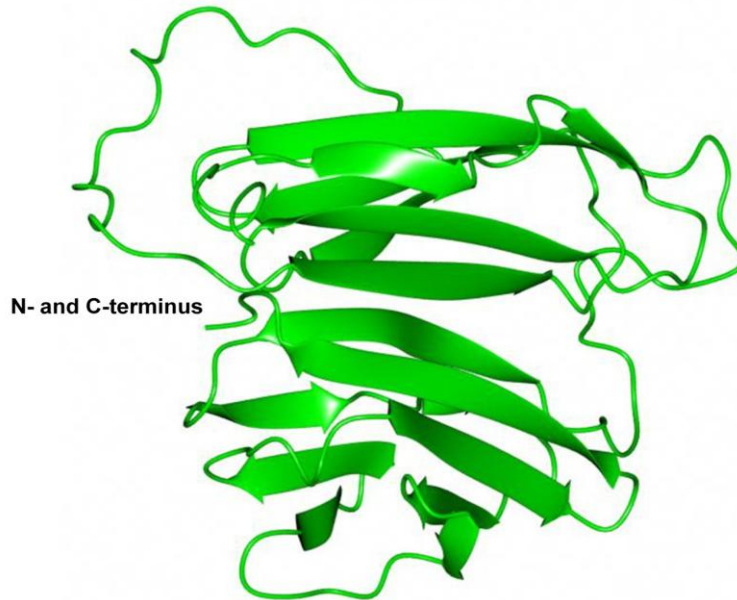


*Figure 106. Structure of IMPL-2.*

The monomeric form of the IMPL-2 from unpublished data.

## A.3.2. Crystallisation, Structure Determination and Refinement

The DILP5:IMPL-2 complex was buffered in 20 mM HEPES pH 7.5 to a concentration of 9.3 mg/ml. The best crystal identified from the screening was obtained from drops equilibrated 1:1 (protein:well) in well solution consisting of 0.2M $MgCl_2$, 8% PEG20k, 8% PEG550MME and 0.1M NaAc pH 4.5 (0.1 ml well volume), and grown at 4°C following equilibration at 21°C. The crystal was flash-cooled in liquid nitrogen using paraffin oil as cryoprotection. The X-ray data were collected at Diamond at station i24, and processed using XDS[185]. The crystal structure of DILP5:IMPL-2 complex was solved by molecular replacement using default settings in Molrep[143]. The search model consisted of DILP5 2wfv.pdb[190] and an unpublished structure for IMPL-2. Subsequent manual model building and refinement with Coot[138] and 10 cycles of REFMAC5[145], resulted in a final model which was validated by MolProbity[146]. Data collection, reduction and refinement statistics are detailed Table 21.

.

Table 21. Data collection, reduction and refinement statistics for the DILP5:IMPL-2 complex.

| | DILP5 + IMPL-2 |
|---|---|
| PDB code | |
| Data collection Beamline | Diamond i24 |
| Wavelength (Å) | 0.9150 |
| Space group | $P4_12_12$ |
| Cell dimensions | |
| $a, b, c$ (Å) | 150.8 150.8 125.3 |
| $\alpha, \beta, \gamma$ (°) | 90.0 90.0 90.0 |
| Resolution (Å) | 20.00-3.50 |
| $R_{sym}$ | 0.162 (1.235) |
| $I / \sigma I$ | 8.2 (1.4) |
| Completeness (%) | 99.5 (99.8) |
| Redundancy | 4.3 (4.4) |
| Refinement | |
| Resolution (Å) | 29.60-3.48 |
| No. reflections | 17857 |
| $R_{work}$ / $R_{free}$ | 0.273/0.315 |
| No. atoms | 7387 |
| Protein | 7387 |
| Ligand/ion | - |
| Water | - |
| $B$-factors | |
| Protein | 60.321 |
| Ligand/ion | - |
| Water | - |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.011 |
| Bond angles (°) | 1.754 |
| Ramachandran statistics (%) | |
| Preferred/Allowed/Outliers | 86.4/6.6/7.0 |

### A.3.3. Results and Discussion

The DILP5:IMPL-2 complex was obtained from our collaborators, where DILP5 and IMPL-2 were expressed separately and were purified to a high level prior to the formation of the complex. Prior to crystallisation the purity of the complex was ascertained (Figure 107-A), with there being multiple bands present that equate to IMPL-2.
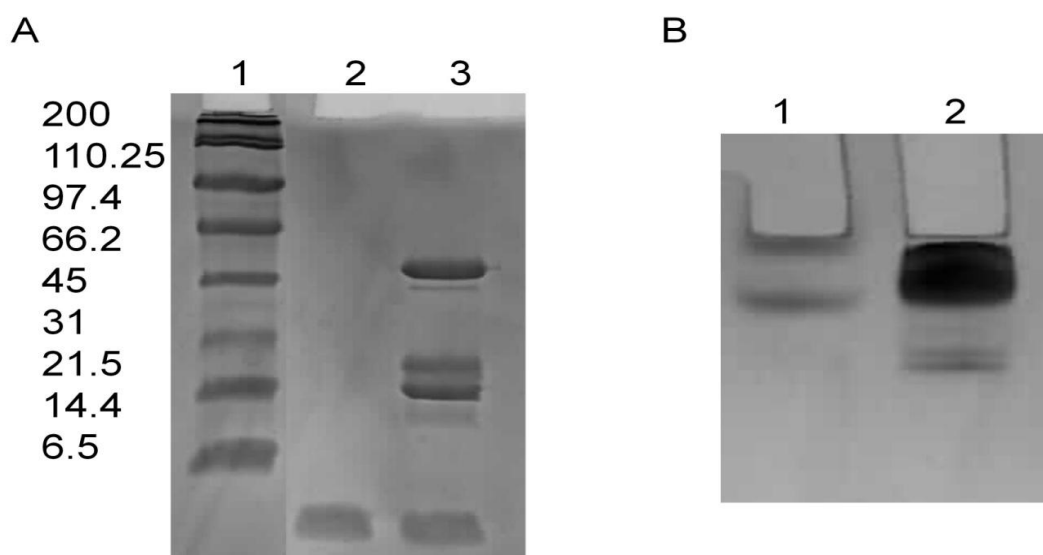


***Figure 107. Physical characterisation of the DILP5:IMPL-2 complex.***
(A) A 20% non-reduced SDS-PAGE of samples of DILP5 and DILP5:IMPL-2 complex. Lane 1: Broad range molecular weight marker. Lane 2: DILP5. Lane 3: DILP5 and IMPL-2. (B) A 20% native PAGE of sample of DILP5 and DILP5:IMPL-2 complex. Lane 1: DILP5. Lane 2: DILP5 and IMPL-2.

The molecular weight of IMPL-2 is ~23 kDa, with a band approximately twice this size present, which indicates due to the non-reducing gel that some disulphide bonds are being formed intermolecularly rather than intramolecularly. This could result in there being multiple species present in the complex. The DILP5 sample has a band that has entered the gel and another that has just entered the gel (Figure 107-B). This band could indicate a high oligomeric state present in the sample. The native gel of the DILP5:IMPL-2 complex indicates one dominant species present, though this is contaminated with a secondary band above the main band. Tertiary bands are present on the gel at lower intensities. Due to the apparent formation of the complex and its prominence in solution, crystallography trials were undertaken.

The complex crystallised in $P4_12_12$ and diffracted to 3.5 Å with four molecules of the DILP5:IMPL-2 complex in the asymmetric unit (Figure 108). This complex shows a screwed 2-fold rotation axis around the central point in the unit cell. The DILP5 binds to the IMPL-2 in a groove formed by a series of β-sheets (Figure 109).
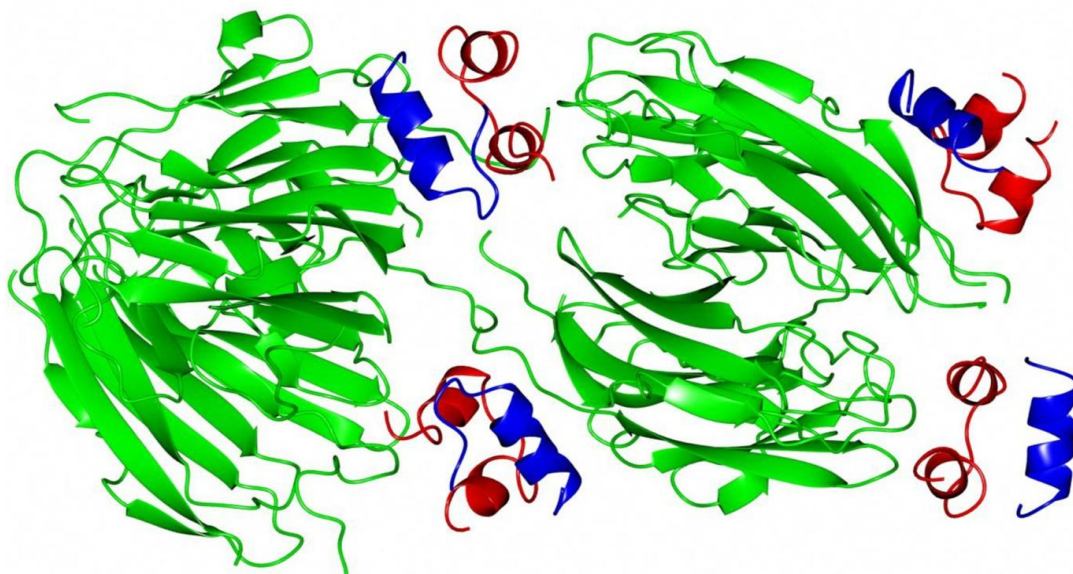
***Figure 108. The complex of DILP5 and IMPL-2***

IMPL-2 is coloured green, DILP5 A-chains are red and DILP5 B-chains are blue.
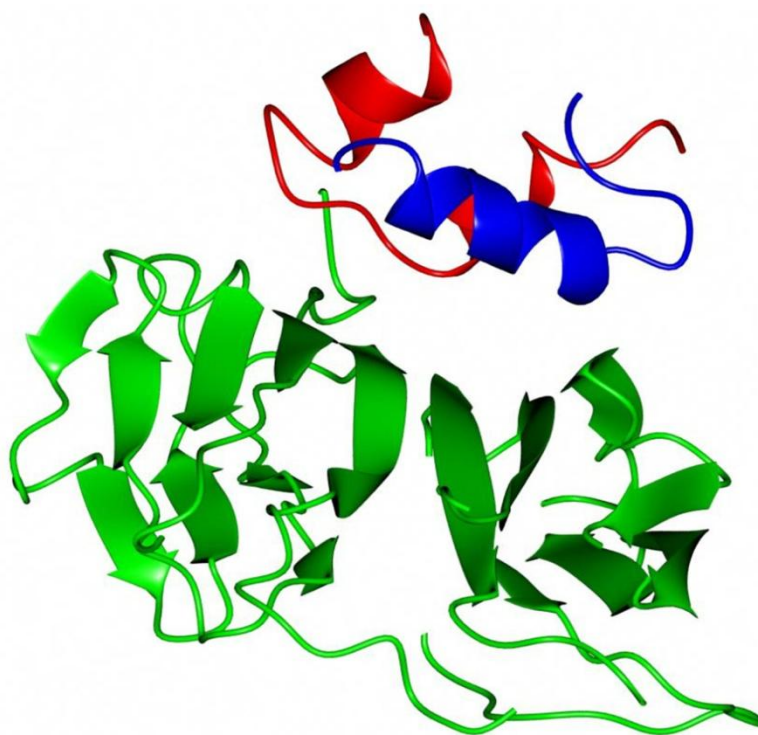


***Figure 109. A view of the binding groove in IMPL-2 cause by the binding of DILP5.***

IMPL-2 is coloured green, DILP5 A-chain is red and DILP5 B-chain in blue.

The low resolution of the structure causes many of the side chains to lack density entirely, or only having density available for $C\gamma$. The result of this is that the interactions between DILP5 and IMPL-2 in the groove are missing. The distances between the main chain atoms of the B-chain in DILP5 and IMPL-2 range between 7.5 and 9.1 Å (Figure

110). This is a size that could accommodate side chains easily and form hydrogen bonds. There is one weak hydrogen bond, 3.8 Å, present in the interface between DILP5 and IMPL-2, located between $Val^{237}$ in IBP and $Cys^{A14}$ in DILP5. These residues are located away from the 7.5 Å gap in the binding groove. This indicates that an increased resolution should lead to future elucidation of interactions further along the groove.
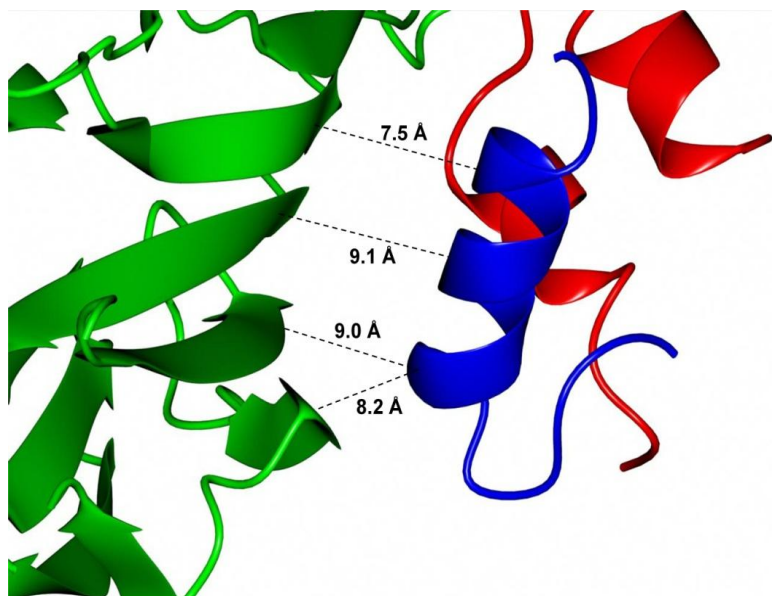


***Figure 110. The space available in the binding groove between IMPL-2 and DILP5.***

IMPL-2 is coloured green, DILP5 A-chain is red and DILP5 B-chain in blue. The dashed lines indicate the main chain to main chain distance.
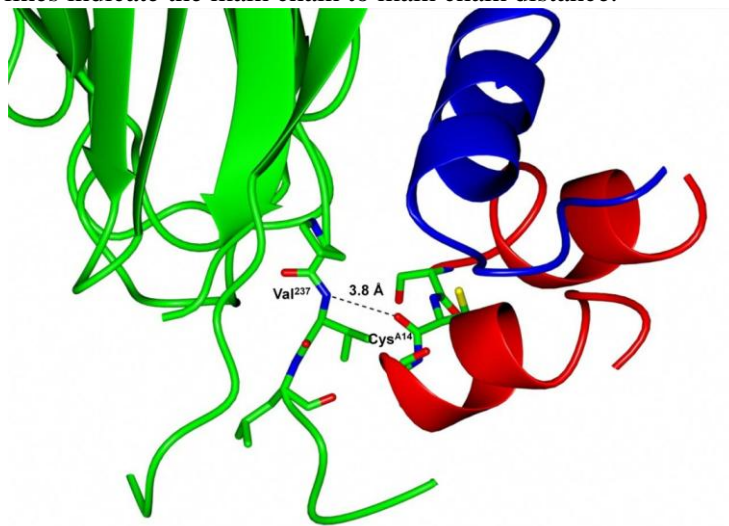


***Figure 111. The hydrogen bond in the binding groove between IMPL-2 and DILP5.***

IMPL-2 is coloured green, DILP5 A-chain is red and DILP5 B-chain in blue. The dashed line indicates a hydrogen bond. Carbon atoms in green, oxygen – red, nitrogen – blue, and sulphur – yellow.

## A.3.4. Conclusion

The binding of DILP5 to IMPL-2 has been implicated in the regulation of metabolism and growth in vertebrates[191]. This binding has been proven in various assays and cross-linking experiments[190,191], but the molecular basis of this interaction had not been determined.

The structure of the complex shows that the binding of DILP5 to IMPL-2 occurs in the centre of a β-sheet. The binding causes a groove to be created, increasing the surface area in which stabilising interactions can be formed. There are residues present in both DILP5 and IMPL-2 that can form hydrogen bonds between either side chain or main chain atoms. Higher resolution crystallographic data would elucidate the location of these hydrogen bonds.

# Abbreviations

| | |
|---|---|
| Å | Ångström |
| $A^6$ | Position 6 in the A-chain |
| BMGY | Buffered Glycerol-complex Medium |
| BMMY | Buffered Methanol-complex Medium |
| CR | Cysteine rich region of the insulin receptor |
| CT peptide | The 18 amino acid region at the C-terminus of the α-subunit of the insulin receptor |
| dATP | Deoxy Adenosine Triphosphate |
| dTTP | Deoxy Thymidine Triphosphate |
| Da | Dalton, the mass of 1/12 of a carbon atom |
| DLS | Dynamic light scattering |
| DNA | Deoxyribonucleic acid |
| DOI | Des-octa insulin |
| DTI | Des-tetra insulin |
| FnIII | Fibronectin type III region of the insulin receptor |
| ID | Insert domain |
| IGF | Insulin-Like Growth Factor |
| IGFBP | Insulin-Like Growth Factor Binding Protein |
| IGF-1R | Insulin-Like Growth Factor Type 1 Receptor |
| IPTG | Isopropylthiogalactoside |
| IR | Insulin receptor |
| IR-A | Insulin receptor isoform A |
| IR-B | Insulin receptor isoform B |
| IRR | Insulin receptor-related receptor |
| ITC | Isothermal Titration Calorimetry |
| JM | Juxtamembrane region |
| $K_d$ | Dissociation constant |
| LIC | Ligation Independent Cloning |
| L1 | Leucine region 1 of the insulin receptor |
| L2 | Leucine region 2 of the insulin receptor |
| MALDI | Matrix-Assisted Laser Desorption/Ionisation |
| mg | Milligrams |
| ml | Millilitres |
| mM | Millimolar |

| | |
|---|---|
| mRNA | Messenger RNA |
| NEFA | Non-esterified fatty acid |
| Ni-NTA | Nickel-nitriloacetic acid |
| $N$Me | A methylation on the main chain nitrogen |
| NMR | Nuclear magnetic resonance |
| PCR | Polymerase chain reaction |
| pET-YSBLIC | YSBL Ligation Independent Cloning vector |
| pH | Concentration of $H^+$ ions |
| $Phe^{B1}$ | Phenylalanine in position 1 in the B-chain |
| pI | Isoelectric point |
| RNA | Ribonucleic acid |
| $R_6$ | The R-state of hexameric insulin |
| $R_1R_1$ | A dimer of insulin consisting of one R-state insulin and one T-state |
| SDS-PAGE | Sodium dodecyl sulphate polyacrylamide gel electrophoresis |
| TK | Tyrosine kinase |
| UV | Ultraviolet |
| WT | Wild type |
| YPD | Yeast Extract Peptone Dextrose Medium |
| YPDS | Yeast Extract Peptone Dextrose Medium with Sorbitol |
| 3C | Protease |

# References

1       Saltiel, A. R. & Kahn, C. R. Insulin Signalling and the Regulation of Glucose and Lipid Metabolism. *Nature* **414**, 799-806 (2001).

2       Adams, M. J. *et al.* Structure of Rhombohedral 2 Zinc Insulin Crystals. *Nature* **224**, 491-495 (1969).

3       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**, 425-433 (1991).

4       Zoete, V., Meuwly, M. & Karplus, M. A Comparison of the Dynamic Behavior of Monomeric and Dimeric Insulin Shows Structural Rearrangements in the Active Monomer. *Journal of Molecular Biology* **342**, 913-929 (2004).

5       Derewenda, U. *et al.* Phenol Stabilizes More Helix in a New Symmetrical Zinc Insulin Hexamer. *Nature* **338**, 594-596 (1989).

6       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).

7       Bentley, G., Dodson, E., Dodson, G. U. Y., Hodgkin, D. & Mercola, D. A. N. Structure of Insulin in 4-Zinc Insulin. *Nature* **261**, 166-168 (1976).

8       Smith, G. D., Ciszak, E. & Pangborn, W. A Novel Complex of a Phenolic Derivative with Insulin: Structural Features Related to the T → R Transition. *Protein Science* **5**, 1502-1511 (1996).

9       Rinderknecht, E. & Humbel, R. E. The Amino Acid Sequence of Human Insulin-Like Growth Factor I and Its Structural Homology with Proinsulin. *Journal of Biological Chemistry* **253**, 2769-2776 (1978).

10      Rinderknecht, E. & Humbel, R. E. Primary Structure of Human Insulin-Like Growth Factor Ii. *FEBS Letters* **89**, 283-286 (1978).

11      Wang, H. S. & Chard, T. Igfs and Igf-Binding Proteins in the Regulation of Human Ovarian and Endometrial Function. *Journal of Endocrinology* **161**, 1-13 (1999).

12      Florini, J. R. *et al.* "Spontaneous" Differentiation of Skeletal Myoblasts Is Dependent Upon Autocrine Secretion of Insulin-Like Growth Factor-Ii. *Journal of Biological Chemistry* **266**, 15917-15923 (1991).

13      Nilsson, A. *et al.* Regulation by Growth Hormone of Number of Chondrocytes Containing Igf-I in Rat Growth Plate. *Science* **233**, 571-574 (1986).

14      Bliss, M. Who Discovered Insulin? *News Physiology Scientist* **1**, 31-36 (1986).

15      Rosenfeld, L. Insulin: Discovery and Controversy. *Clinical Chemistry* **48**, 2270-2288 (2002).

16      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**, 141-146 (1922).

17      Opie, E. L. On the Relation of Chronic Interstitial Pancreatitis to the Islands of Langerhans and to Diabetes Melutus. *The Journal of Experimental Medicine* **5**, 397-428 (1901).

18      Murray, I. Paulesco and the Isolation of Insulin. *Journal of the History of Medicine and Allied Sciences* **26**, 150-157 (1971).

19      Stretton, A. O. The First Sequence. Fred Sanger and Insulin. *Genetics* **162**, 527-532 (2002).

20      Katsoyannis, P. G., Tometsko, A., Zalut, C., Johnson, S. & Trakatellis, A. C. Studies on the Synthesis of Insulin from Natural and Synthetic a and B Chains. I. Splitting of Insulin and Isolation of the S-Sulfonated Derivatives of the a and B Chains. *Biochemistry* **6**, 2635-2642 (1967).

21      Zahn, H. My Journey from Wool Research to Insulin. *Journal of Peptide Science : an Official Publication of the European Peptide Society* **6**, 1-10 (2000).

22      Johnson, I. S. Human Insulin from Recombinant DNA Technology. *Science* **219**, 632-637 (1983).

23      Schuit, F. C., Huypens, P., Heimberg, H. & Pipeleers, D. G. Glucose Sensing in Pancreatic B-Cells: A Model for the Study of Other Glucose-Regulated Cells in Gut, Pancreas, and Hypothalamus. *Diabetes* **50**, 1-11, (2001).

24      Dodson, G. & Steiner, D. The Role of Assembly in Insulin's Biosynthesis. *Current Opinions in Structural Biology* **8**, 189-194 (1998).

25      Phillips, N. B. *et al.* Supramolecular Protein Engineering: Design of Zinc-Stapled Insulin Hexamers as a Long Acting Depot. *Journal of Biological Chemistry* **285**, 11755-11759 (2010).

26      Malaisse, W. J., Sener, A., Herchuelz, A. & Hutton, J. C. Insulin Release: The Fuel Hypothesis. *Metabolism* **28**, 373-386 (1979).

27      Hartley, T., Brumell, J. & Volchuk, A. Emerging Roles for the Ubiquitin-Proteasome System and Autophagy in Pancreatic Beta-Cells. *American Journal of Physiology, Endocrinology and Metabolism* **296**, E1-10 (2009).

28      Gembal, M., Gilon, P. & Henquin, J. C. Evidence That Glucose Can Control Insulin Release Independently from Its Action on Atp-Sensitive K+ Channels in Mouse B Cells. *Journal of Clinical Investigation* **89**, 1288-1295 (1992).

29      Pessin, J. E., Thurmond, D. C., Elmendorf, J. S., Coker, K. J. & Okada, S. Molecular Basis of Insulin-Stimulated Glut4 Vesicle Trafficking. Location! Location! Location! *Journal of Biological Chemistry* **274**, 2593-2596 (1999).

30      Guilherme, A. *et al.* Perinuclear Localization and Insulin Responsiveness of Glut4 Requires Cytoskeletal Integrity in 3t3-L1 Adipocytes. *Journal of Biological Chemistry* **275**, 38151-38159 (2000).

31      Carpentier, J. L. *et al.* Two Steps of Insulin Receptor Internalization Depend on Different Domains of the Beta-Subunit. *Journal Cell Biology* **122**, 1243-1252 (1993).

32      Duckworth, W. C., Bennett, R. G. & Hamel, F. G. Insulin Degradation: Progress and Potential. *Endocrine Reviews* **19**, 608-624 (1998).

33      Daly, M. E. *et al.* Acute Effects on Insulin Sensitivity and Diurnal Metabolic Profiles of a High-Sucrose Compared with a High-Starch Diet. *American Journal of Clinical Nutrition* **67**, 1186-1196 (1998).

34      Jozic, I. B., G. Barbieri, A. Inhibition of Rab5 Activation During Insulin Receptor-Mediated Endocytosis. *Current Cellular Biochemistry* **1**, 20-32 (2011).

35      Fucini, R. V., Okada, S. & Pessin, J. E. Insulin-Induced Desensitization of Extracellular Signal-Regulated Kinase Activation Results from an Inhibition of Raf Activity Independent of Ras Activation and Dissociation of the Grb2-Sos Complex. *Journal of Biological Chemistry* **274**, 18651-18658 (1999).

36      Fan, J. Y. *et al.* Receptor-Mediated Endocytosis of Insulin: Role of Microvilli, Coated Pits, and Coated Vesicles. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 7788-7791 (1982).

37      Shepherd, P. R. & Kahn, B. B. Glucose Transporters and Insulin Action--Implications for Insulin Resistance and Diabetes Mellitus. *The New England Journal of Medicine* **341**, 248-257 (1999).

38      Goodyear, L. J., Hirshman, M. F., Smith, R. J. & Horton, E. S. Glucose Transporter Number, Activity, and Isoform Content in Plasma Membranes of Red and White Skeletal Muscle. *American Journal of Physiology* **261**, E556-561 (1991).

39      Levine, R., Goldstein, M. & et al. The Action of Insulin on the Distribution of Galactose in Eviscerated Nephrectomized Dogs. *Journal of Biological Chemistry* **179**, 985 (1949).

40      House, P. D. & Weidemann, M. J. Characterization of an [125 I]-Insulin Binding Plasma Membrane Fraction from Rat Liver. *Biochemical and Biophysical Research Communications* **41**, 541-548 (1970).

41      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**, 7137-7141 (1980).

42      Sparrow, L. G. *et al.* The Disulfide Bonds in the C-Terminal Domains of the Human Insulin Receptor Ectodomain. *Journal of Biological Chemistry* **272**, 29460-29467 (1997).

43      Kasuga, M. *et al.* Insulin Stimulation of Phosphorylation of the Beta Subunit of the Insulin Receptor. Formation of Both Phosphoserine and Phosphotyrosine. *Journal of Biological Chemistry* **257**, 9891-9894 (1982).

44      Kasuga, M., Zick, Y., Blithe, D. L., Crettaz, M. & Kahn, C. R. Insulin Stimulates Tyrosine Phosphorylation of the Insulin Receptor in a Cell-Free System. *Nature* **298**, 667-669 (1982).

45      White, M. F., Maron, R. & Kahn, C. R. Insulin Rapidly Stimulates Tyrosine Phosphorylation of a Mr-185,000 Protein in Intact Cells. *Nature* **318**, 183-186 (1985).

46      Sun, X. J. *et al.* Structure of the Insulin Receptor Substrate Irs-1 Defines a Unique Signal Transduction Protein. *Nature* **352**, 73-77, (1991).

47      Ebina, Y. *et al.* The Human Insulin Receptor Cdna: The Structural Basis for Hormone-Activated Transmembrane Signalling. *Cell* **40**, 747-758 (1985).

48      Ullrich, A. *et al.* Human Insulin Receptor and Its Relationship to the Tyrosine Kinase Family of Oncogenes. *Nature* **313**, 756-761 (1985).

49      Sparrow, L. G. *et al.* N-Linked Glycans of the Human Insulin Receptor and Their Distribution over the Crystal Structure. *Proteins: Structure, Function, and Bioinformatics* **71**, 426-439, (2008).

50      Sparrow, L. G. *et al.* The Location and Characterisation of the O-Linked Glycans of the Human Insulin Receptor. *Proteins* **66**, 261-265, (2007).

51      Hubbard, S. R., Wei, L., Ellis, L. & Hendrickson, W. A. Crystal Structure of the Tyrosine Kinase Domain of the Human Insulin Receptor. *Nature* **372**, 746-754, (1994).

52      Hubbard, S. R. Crystal Structure of the Activated Insulin Receptor Tyrosine Kinase in Complex with Peptide Substrate and Atp Analog. *EMBO Journal* **16**, 5572-5581, (1997).

53      McKern, N. M. *et al.* Structure of the Insulin Receptor Ectodomain Reveals a Folded-over Conformation. *Nature* **443**, 218-221 (2006).

54      Soos, M. A. *et al.* Monoclonal Antibodies Reacting with Multiple Epitopes on the Human Insulin Receptor. *Biochemistry Journal* **235**, 199-208 (1986).

55      Ward, C. W. & Lawrence, M. C. Landmarks in Insulin Research. *Frontiers in Endocrinology* **2**, 76 (2011).

56      Ward, C. W. & Lawrence, M. C. Similar but Different: Ligand-Induced Activation of the Insulin and Epidermal Growth Factor Receptor Families. *Current Opinions in Structural Biology* **22**, 360-366, (2012).

57      Fernandez, A. M. *et al.* Functional Inactivation of the Igf-I and Insulin Receptors in Skeletal Muscle Causes Type 2 Diabetes. *Genes and Development* **15**, 1926-1934 (2001).

58      Ullrich, A. & Schlessinger, J. Signal Transduction by Receptors with Tyrosine Kinase Activity. *Cell* **61**, 203-212 (1990).

59      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**, 6771-6776 (2010).

60      De Meyts, P. The Insulin Receptor: A Prototype for Dimeric, Allosteric Membrane Receptors? *Trends in Biochemical Sciences* **33**, 376-384 (2008).

61      Ward, C. W. & Lawrence, M. C. Ligand-Induced Activation of the Insulin Receptor: A Multi-Step Process Involving Structural Changes in Both the Ligand and the Receptor. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* **31**, 422-434 (2009).

62     De Meyts, P. & Whittaker, J. Structural Biology of Insulin and Igf1 Receptors: Implications for Drug Design. *Nature Reviews Drug Discovery* **1**, 769-783 (2002).

63     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**, 12429-12434 (2006).

64     Kristensen, C., Andersen, A. S., Ostergaard, S., Hansen, P. H. & Brandt, J. Functional Reconstitution of Insulin Receptor Binding Site from Non-Binding Receptor Fragments. *Journal of Biological Chemistry* **277**, 18340-18345 (2002).

65     Lawrence, M. C., McKern, N. M. & Ward, C. W. Insulin Receptor Structure and Its Implications for the Igf-1 Receptor. *Current Opinion in Structural Biology* **17**, 699-705 (2007).

66     Seino, S. & Bell, G. I. Alternative Splicing of Human Insulin Receptor Messenger Rna. *Biochemical and Biophysical Research Communications* **159**, 312-316 (1989).

67     Neuvians, T. P., Pfaffl, M. W., Berisha, B. & Schams, D. The Mrna Expression of Insulin Receptor Isoforms (Ir-a and Ir-B) and Igfr-2 in the Bovine Corpus Luteum During the Estrous Cycle, Pregnancy, and Induced Luteolysis. *Endocrine* **22**, 93-100 (2003).

68     Benecke, H., Flier, J. S. & Moller, D. E. Alternatively Spliced Variants of the Insulin Receptor Protein. Expression in Normal and Diabetic Human Tissues. *Journal of Clinical Investigation* **89**, 2066-2070 (1992).

69     Mosthaf, L. *et al.* Functionally Distinct Insulin Receptors Generated by Tissue-Specific Alternative Splicing. *EMBO Journal* **9**, 2409-2413 (1990).

70     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).

71     Gauguin, L. *et al.* Structural Basis for the Lower Affinity of the Insulin-Like Growth Factors for the Insulin Receptor. *Journal of Biological Chemistry* **283**, 2604-2613, (2008).

72     Pittman, I. t. & Tager, H. S. A Spectroscopic Investigation of the Conformational Dynamics of Insulin in Solution. *Biochemistry* **34**, 10578-10590 (1995).

73     Mayer, J. P., Zhang, F. & DiMarchi, R. D. Insulin Structure and Function. *Biopolymers* **88**, 687-713 (2007).

74     Pullen, R. A. *et al.* Receptor-Binding Region of Insulin. *Nature* **259**, 369-373 (1976).

75     Jiracek, 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**, 1966-1970 (2010).

76     Nakagawa, S. H. & Tager, H. S. Role of the Phenylalanine B25 Side Chain in Directing Insulin Interaction with Its Receptor. Steric and Conformational Effects. *Journal of Biological Chemistry* **261**, 7332-7341 (1986).

77     Xu, B. *et al.* Decoding the Cryptic Active Conformation of a Protein by Synthetic Photoscanning: Insulin Inserts a Detachable Arm between Receptor Domains. *Journal of Biological Chemistry* **284**, 14597-14608 (2009).

78     Kristensen, C. *et al.* Alanine Scanning Mutagenesis of Insulin. *Journal of Biological Chemistry* **272**, 12978-12983 (1997).

79     Schaffer, L. A Model for Insulin Binding to the Insulin Receptor. *European Journal of Biochemistry* **221**, 1127-1132 (1994).

80     Baker, E. N. *et al.* The Structure of 2zn Pig Insulin Crystals at 1.5 a Resolution. *Philoophicals Transactions of the Royal Society of London B: Biological Sciences* **319**, 369-456 (1988).

81     Simoens, S. Biosimilar Medicines and Cost-Effectiveness. *ClinicoEconomics and Outcomes Research* **2011:3**, 29-36 (2011).

82     Alberti, K. G. & Zimmet, P. Z. Definition, Diagnosis and Classification of Diabetes

Mellitus and Its Complications. Part 1: Diagnosis and Classification of Diabetes Mellitus Provisional Report of a Who Consultation. *Diabetic Medicine : a Journal of the British Diabetic Association* **15**, 539-553 (1998).

83      Alberti, K. G. & Zimmet, P. Z. New Diagnostic Criteria and Classification of Diabetes--Again? *Diabetic Medicine : a Journal of the British Diabetic Association* **15**, 535-536 (1998).

84      Association, A. D. Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* **34 Suppl 1**, S62-69 (2011).

85      Trubo, R. Researchers Investigate Factors Linked to Development of Secondary Diabetes. *JAMA : the Journal of the American Medical Association* **294**, 668-670 (2005).

86      Dixon, G. H. & Wardlaw, A. C. Regeneration of Insulin Activity from the Separated and Inactive a and B Chains. *Nature* **188**, 721-724 (1960).

87      Katsoyannis, P. G., Fukuda, K. & Tometsko, A. Insulin Peptides .9. Synthesis of a-Chain of Insulin and Its Combination with Natural B-Chain to Generate Insulin Activity. *Journal of the American Chemical Society* **85**, 2863-& (1963).

88      Marglin, A. & Merrifield, R. B. The Synthesis of Bovine Insulin by the Solid Phase Method. *Journal of the American Chemical Society* **88**, 5051-5052 (1966).

89      Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of Proteins by Native Chemical Ligation. *Science* **266**, 776-779 (1994).

90      Dawson, P. E. & Kent, S. B. Synthesis of Native Proteins by Chemical Ligation. *Annual Review of Biochemistry* **69**, 923-960 (2000).

91      Inouye, K. *et al.* Enzyme-Assisted Semisynthesis of Human Insulin. *Journal of the American Chemical Society* **101**, 751-752 (1979).

92      Zakova, L. *et al.* The Use of Fmoc-Lys(Pac)-Oh and Penicillin G Acylase in the Preparation of Novel Semisynthetic Insulin Analogs. *The Journal of Peptide Science* **13**, 334-341 (2007).

93      Sohma, Y., Hua, Q. X., Whittaker, J., Weiss, M. A. & Kent, S. B. H. Design and Folding of [Glu(A4)(O(Beta)Thr(B30))]Insulin ("Ester Insulin"): A Minimal Proinsulin Surrogate That Can Be Chemically Converted into Human Insulin. *Angewandte Chemie-International Edition* **49**, 5489-5493 (2010).

94      Young, J. D. & Carpenter, F. H. Isolation and Characterization of Products Formed by Action of Trypsin on Insulin. *Journal of Biological Chemistry* **236**, 743-& (1961).

95      Abel, J. J. Crystalline Insulin. *Proceedings of the National Academy of Sciences of the United States of America* **12**, 132-136 (1926).

96      Romans, R. G., Scott, D. A. & Fisher, A. M. Preparation of Crystalline Insulin. *Industrial & Engineering Chemistry* **32**, 908-910, (1940).

97      Skyler, J. S. Diabetes Mellitus: Pathogenesis and Treatment Strategies. *Journal of Medicinal Chemistry* **47**, 4113-4117 (2004).

98      Joshi, S. R., Parikh, R. M. & Das, A. K. Insulin--History, Biochemistry, Physiology and Pharmacology. *The Journal of the Association of Physicians of India* **55 Suppl**, 19-25 (2007).

99      Gualandi-Signorini, A. M. & Giorgi, G. Insulin Formulations--a Review. *European Review for Medical and Pharmacological Sciences* **5**, 73-83 (2001).

100     Du, H., Shi, J., Cui, D. & Zhang, Y. Insulin Analogs with B24 or B25 Phenylalanine Replaced by Biphenylalanine. *Acta Biochimica et Biophysica Sinica (Shanghai)* **40**, 133-139 (2008).

101     Kurtzhals, P. Engineering Predictability and Protraction in a Basal Insulin Analogue: The Pharmacology of Insulin Detemir. *International Journal of Obesity and Related Metabolic Disorders : Journal of the International Association for the Study of Obesity* **28 Suppl 2**, S23-28 (2004).

102     Goeddel, D. V. *et al.* Expression in Escherichia Coli of Chemically Synthesized Genes for Human Insulin. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 106-110 (1979).

103    Katsoyannis, P. G. & Tometsko, A. Insulin Synthesis by Recombination of a and B Chains: A Highly Efficient Method. *Proceedings of the National Academy of Sciences of the United States of America* **55**, 1554-1561 (1966).

104    Kjeldsen, T. Yeast Secretory Expression of Insulin Precursors. *Applied Microbiology and Biotechnology* **54**, 277-286 (2000).

105    Hirsch, I. B. Insulin Analogues. *The New England Journal of Medicine* **352**, 174-183 (2005).

106    Brange, J. *et al.* Monomeric Insulins Obtained by Protein Engineering and Their Medical Implications. *Nature* **333**, 679-682 (1988).

107    Campbell, R. K., White, J. R., Levien, T. & Baker, D. Insulin Glargine. *Clinical Therapeutics* **23**, 1938-1957; discussion 1923 (2001).

108    Brange, J. *Galenics of Insulin : The Physico-Chemical and Pharmaceutical Aspects of Insulin and Insulin Preparations*. (Springer-Verlag, 1987).

109    Brange, J. The New Era of Biotech Insulin Analogues. *Diabetologia* **40 Suppl 2**, S48-53 (1997).

110    Hallas, M., Oller, K., Petersen, K. & Schlichtkrull, J. Crystalline and Amorphous Insulin-Zinc Compounds with Prolonged Action. *Science* **116**, 394-398 (1952).

111    Freeman, J. S. Are Analogue Insulins Superior to Human Insulin in Clinical Practice? *Current Diabetes Reports* **10**, 176-183 (2010).

112    Damgé, C., Maincent, P. & Ubrich, N. Oral Delivery of Insulin Associated to Polymeric Nanoparticles in Diabetic Rats. *Journal of Controlled Release* **117**, 163-170 (2007).

113    Dyer, A. M. *et al.* Nasal Delivery of Insulin Using Novel Chitosan Based Formulations: A Comparative Study in Two Animal Models between Simple Chitosan Formulations and Chitosan Nanoparticles. *Pharmaceutical Research* **19**, 998-1008 (2002).

114    Chantelau, E., Spraul, M., Mühlhauser, I., Gause, R. & Berger, M. Long-Term Safety, Efficacy and Side-Effects of Continuous Subcutaneous Insulin Infusion Treatment for Type 1 (Insulin-Dependent) Diabetes Mellitus: A One Centre Experience. *Diabetologia* **32**, 421-426 (1989).

115    Belfiore, A. & Malaguarnera, R. Insulin Receptor and Cancer. *Endocrine-Related Cancer* **18**, R125-R147 (2011).

116    Sciacca, L. *et al.* Insulin Receptor Activation by Igf-Ii in Breast Cancers: Evidence for a New Autocrine/Paracrine Mechanism. *Oncogene* **18**, 2471-2479 (1999).

117    Froesch, E. R. & Zapf, J. Insulin-Like Growth Factors and Insulin: Comparative Aspects. *Diabetologia* **28**, 485-493 (1985).

118    Morgan, D. O. *et al.* Insulin-Like Growth Factor Ii Receptor as a Multifunctional Binding Protein. *Nature* **329**, 301-307 (1987).

119    Clairmont, K. B. & Czech, M. P. Chicken and Xenopus Mannose 6-Phosphate Receptors Fail to Bind Insulin-Like Growth Factor Ii. *Journal of Biological Chemistry* **264**, 16390-16392 (1989).

120    Zapf, J., Walter, H. & Froesch, E. R. Radioimmunological Determination of Insulinlike Growth Factors I and Ii in Normal Subjects and in Patients with Growth Disorders and Extrapancreatic Tumor Hypoglycemia. *Journal of Clinical Investigation* **68**, 1321-1330 (1981).

121    Brown, A. L. *et al.* Developmental Regulation of Insulin-Like Growth Factor Ii Mrna in Different Rat Tissues. *Journal of Biological Chemistry* **261**, 13144-13150 (1986).

122    DeChiara, T. M., Efstratiadis, A. & Robertson, E. J. A Growth-Deficiency Phenotype in Heterozygous Mice Carrying an Insulin-Like Growth Factor Ii Gene Disrupted by Targeting. *Nature* **345**, 78-80 (1990).

123    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).

124    Velcheti, V. & Govindan, R. Insulin-Like Growth Factor and Lung Cancer. *The*

*Journal of Thoracic Oncology* **1**, 607-610 (2006).

125    Katz, L. E., DeLeon, D. D., Zhao, H. & Jawad, A. F. Free and Total Insulin-Like Growth Factor (Igf)-I Levels Decline During Fasting: Relationships with Insulin and Igf-Binding Protein-1. *Journal of Clinical Endocrinology Metabolism* **87**, 2978-2983 (2002).

126    Guler, H. P., Zapf, J., Schmid, C. & Froesch, E. R. Insulin-Like Growth Factors I and Ii in Healthy Man. Estimations of Half-Lives and Production Rates. *Acta Endocrinologica (Copenhagen)* **121**, 753-758 (1989).

127    Busby, W. H., Jr., Klapper, D. G. & Clemmons, D. R. Purification of a 31,000-Dalton Insulin-Like Growth Factor Binding Protein from Human Amniotic Fluid. Isolation of Two Forms with Different Biologic Actions. *Journal of Biological Chemistry* **263**, 14203-14210 (1988).

128    Signorello, L. B. *et al.* Insulin-Like Growth Factor-Binding Protein-1 and Prostate Cancer. *Journal of the National Cancer Institute* **91**, 1965-1967 (1999).

129    Lee, P. D. *et al.* Kinetics of Insulin-Like Growth Factor (Igf) and Igf-Binding Protein Responses to a Single Dose of Growth Hormone. *Journal of Clinical Endocrinology and Metabolism* **82**, 2266-2274 (1997).

130    Duan, C., Ding, J., Li, Q., Tsai, W. & Pozios, K. Insulin-Like Growth Factor Binding Protein 2 Is a Growth Inhibitory Protein Conserved in Zebrafish. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 15274-15279 (1999).

131    Yakar, S. *et al.* Circulating Levels of Igf-1 Directly Regulate Bone Growth and Density. *Journal of Clinical Investigation* **110**, 771-781 (2002).

132    Yakar, S., Wu, Y., Setser, J. & Rosen, C. J. The Role of Circulating Igf-I: Lessons from Human and Animal Models. *Endocrine* **19**, 239-248 (2002).

133    Mohan, S., Bautista, C. M., Wergedal, J. & Baylink, D. J. Isolation of an Inhibitory Insulin-Like Growth Factor (Igf) Binding Protein from Bone Cell-Conditioned Medium: A Potential Local Regulator of Igf Action. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 8338-8342 (1989).

134    Kiefer, M. C. *et al.* Characterization of Recombinant Human Insulin-Like Growth Factor Binding Proteins 4, 5, and 6 Produced in Yeast. *Journal of Biological Chemistry* **267**, 12692-12699 (1992).

135    Jones, J. I., Gockerman, A., Busby, W. H., Jr., Camacho-Hubner, C. & Clemmons, D. R. Extracellular Matrix Contains Insulin-Like Growth Factor Binding Protein-5: Potentiation of the Effects of Igf-I. *Journal of Cellular Biology* **121**, 679-687 (1993).

136    Bernal, J. D. & Crowfoot, D. X-Ray Photographs of Crystalline Pepsin. *Nature* **133**, 794-795 (1934).

137    McCoy, A. Liking Likelihood. *Acta Crystallographica Section D* **60**, 2169-2183 (2004).

138    Emsley, P. & Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallographica Section D* **60**, 2126-2132 (2004).

139    Drenth, J. *Principles of Protein X-Ray Crystallography*. (Springer-Verlag, 1994).

140    Otwinowski, Z. & Minor, W. Processing of X-Ray Diffraction Data Collected in Oscillation Mode. *Macromolecular Crystallography, Part A* **276**, 307-326 (1997).

141    Geworth, D. *Hkl2000 Manual*. Vol. 2.6 (2007).

142    Weiss, M. S. Global Indicators of X-Ray Data Quality. *Journal of Applied Crystallography* **34**, 130-135 (2001).

143    Vagin, A. & Teplyakov, A. Molrep: An Automated Program for Molecular Replacement. *Journal of Applied Crystallography* **30**, 1022-1025 (1997).

144    Blow, D. M. *Outline of Crystallography for Biologists*. (Oxford University Press, 2002).

145    Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D* **53**, 240-255 (1997).

146 Chen, V. B. *et al.* Molprobity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallographica Section D* **66**, 12-21 (2010).

147 Kraft, P. *et al.* Performance of Single-Photon-Counting Pilatus Detector Modules. *Journal of Synchrotron Radiation* **16**, 368-375 (2009).

148 Evans, G., Axford, D., Waterman, D. & Owen, R. L. Macromolecular Microcrystallography. *Crystallography Reviews* **17**, 105-142 (2011).

149 Bowler, M. W. *et al.* Diffraction Cartography: Applying Microbeams to Macromolecular Crystallography Sample Evaluation and Data Collection. *Acta Crystallographica. Section D, Biological Crystallography* **66**, 855-864 (2010).

150 Zakova, L. *et al.* Insulin Analogues with Modifications at Position B26. Divergence of Binding Affinity and Biological Activity. *Biochemistry* **47**, 5858-5868 (2008).

151 Smith, G. D., Pangborn, W. A. & Blessing, R. H. The Structure of T6 Human Insulin at 1.0 a Resolution. *Acta Crystallographica Section D* **59**, 474-482 (2003).

152 Nakagawa, S. H., Johansen, N. L., Madsen, K., Schwartz, T. W. & Tager, H. S. Implications of Replacing Peptide Bonds in the Cooh-Terminal B Chain Domain of Insulin by the Ψ(Ch2-Nh) Linker. *International Journal of Peptide and Protein Research* **42**, 578-584 (1993).

153 Kurapkat, G. *et al.* The Solution Structure of a Superpotent B-Chain-Shortened Single-Replacement Insulin Analogue. *Protein Science* **8**, 499-508 (1999).

154 Pal, D. & Chakrabarti, P. Cis Peptide Bonds in Proteins: Residues Involved, Their Conformations, Interactions and Locations. *Journal of Molecular Biology* **294**, 271-288, (1999).

155 Ciszak, E. *et al.* Role of C-Terminal B-Chain Residues in Insulin Assembly: The Structure of Hexameric Lysb28prob29-Human Insulin. *Structure* **3**, 615-622, (1995).

156 De Meyts, P., Van Obberghen, E. & Roth, J. Mapping of the Residues Responsible for the Negative Cooperativity of the Receptor-Binding Region of Insulin. *Nature* **273**, 504-509 (1978).

157 Keefer, L. M. *et al.* Impaired Negative Cooperativity of the Semisynthetic Analogues Human [Leub24]- and [Leub25]-Insulins. *Biochemical and Biophysical Research Communications* **100**, 1229-1236, (1981).

158 Goldman, J. & Carpenter, F. H. Zinc Binding, Circular Dichroism, and Equilibrium Sedimentation Studies on Insulin (Bovine) and Several of Its Derivatives. *Biochemistry* **13**, 4566-4574, (1974).

159 Antolikova, E. *et al.* Non-Equivalent Role of Inter- and Intramolecular Hydrogen Bonds in the Insulin Dimer Interface. *Journal of Biological Chemistry* **286**, 36968-36977, (2011).

160 Lovatt, M., Cooper, A. & Camiller, P. Energetics of Cyclodextrin-Induced Dissociation of Insulin. *Journal of Inclusion Phenomena and Macrocyclic Chemistry* **25**, 169-172, (1996).

161 Strazza, S., Hunter, R., Walker, E. & Darnall, D. W. The Thermodynamics of Bovine and Porcine Insulin and Proinsulin Association Determined by Concentration Difference Spectroscopy. *Archives of Biochemistry and Biophysics* **238**, 30-42, (1985).

162 Žáková, L. *et al.* Toward the Insulin−Igf-I Intermediate Structures: Functional and Structural Properties of the [Tyrb25nmepheb26] Insulin Mutant†. *Biochemistry* **43**, 16293-16300, (2004).

163 Whittingham, J. L., Edwards, D. J., Antson, A. A., Clarkson, J. M. & Dodson, G. G. Interactions of Phenol and M-Cresol in the Insulin Hexamer, and Their Effect on the Association Properties of B28 Pro → Asp Insulin Analogues‡. *Biochemistry* **37**, 11516-11523, (1998).

164 Ganim, Z., Jones, K. C. & Tokmakoff, A. Insulin Dimer Dissociation and Unfolding Revealed by Amide I Two-Dimensional Infrared Spectroscopy. *Physical Chemistry Chemical Physics* **12**, 3579-3588 (2010).

165 Ludvigsen, S., Olsen, H. B. & Kaarsholm, N. C. A Structural Switch in a Mutant

Insulin Exposes Key Residues for Receptor Binding. *Journal of Molecular Biology* **279**, 1-7, (1998).

166    Hua, Q. X., Ladbury, J. E. & Weiss, M. A. Dynamics of a Monomeric Insulin Analog: Testing the Molten-Globule Hypothesis. *Biochemistry* **32**, 1433-1442, (1993).

167    Yao, Z. P. *et al.* Structure of an Insulin Dimer in an Orthorhombic Crystal: The Structure Analysis of a Human Insulin Mutant (B9 Ser -> Glu). *Acta Crystallographica Section D-Biological Crystallography* **55**, 1524-1532 (1999).

168    Whittingham, J. L. *et al.* Insulin at Ph 2: Structural Analysis of the Conditions Promoting Insulin Fibre Formation. *Journal of Molecular Biology* **318**, 479-490, (2002).

169    Brange, J., Owens, D. R., Kang, S. & Vølund, A. Monomeric Insulins and Their Experimental and Clinical Implications. *Diabetes Care* **13**, 923-954, (1990).

170    Blundell, T. L., Dodson, G. G., Hodgkin, D. C. & Mercola, D. A. Insulin: The Structure in the Crystal and Its Reflection in Chemistry and Biology. *Advances in Protein Chemistry* **26**, 279-402 (1972).

171    Kristensen, C., Wiberg, F. C., Schaffer, L. & Andersen, A. S. Expression and Characterization of a 70-Kda Fragment of the Insulin Receptor That Binds Insulin - Minimizing Ligand Binding Domain of the Insulin Receptor. *Journal of Biological Chemistry* **273**, 17780-17786 (1998).

172    Huang, K. *et al.* The a-Chain of Insulin Contacts the Insert Domain of the Insulin Receptor: Photo-Cross-Linking and Mutagenesis of a Diabetes-Related Crevice. *Journal of Biological Chemistry* **282**, 35337-35349, (2007).

173    Fogg, M. J. & Wilkinson, A. J. Higher-Throughput Approaches to Crystallization and Crystal Structure Determination. *Biochemical Society Transactions* **36**, 771-775 (2008).

174    Oganesyan, N., Kim, S. H. & Kim, R. On-Column Chemical Refolding of Proteins. *PharmaGenomics* **4**, 5 (2004).

175    Walter, T. S. *et al.* Crystallization and Preliminary X-Ray Analysis of Mouse Rank and Its Complex with Rankl. *Acta Crystallographica Section F: Structural Biology and Crystalization Communications* **65**, 597-600 (2009).

176    Lindwall, G., Chau, M., Gardner, S. R. & Kohlstaedt, L. A. A Sparse Matrix Approach to the Solubilization of Overexpressed Proteins. *Protein Engineering* **13**, 67-71 (2000).

177    Carrio, M. M. & Villaverde, A. Protein Aggregation as Bacterial Inclusion Bodies Is Reversible. *FEBS Letters* **489**, 29-33 (2001).

178    Tams, J. W. & Welinder, K. G. Mild Chemical Deglycosylation of Horseradish Peroxidase Yields a Fully Active, Homogeneous Enzyme. *Analytical Biochemistry* **228**, 48-55, (1995).

179    Tams, J. W., Vind, J. & Welinder, K. G. Adapting Protein Solubility by Glycosylation.: N-Glycosylation Mutants of Coprinus Cinereus Peroxidase in Salt and Organic Solutions. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1432**, 214-221, (1999).

180    Barnes, G., Hansen, W. J., Holcomb, C. L. & Rine, J. Asparagine-Linked Glycosylation in Saccharomyces Cerevisiae: Genetic Analysis of an Early Step. *Molecular and Cellular Biology* **4**, 2381-2388 (1984).

181    Hossler, P., Khattak, S. F. & Li, Z. J. Optimal and Consistent Protein Glycosylation in Mammalian Cell Culture. *Glycobiology* **19**, 936-949, (2009).

182    Sevier, C. S. & Kaiser, C. A. Formation and Transfer of Disulphide Bonds in Living Cells. *Nature Reviews Molecular Cell Biology* **3**, 836-847 (2002).

183    Ward, C. W. & Lawrence, M. C. Ligand-Induced Activation of the Insulin Receptor: A Multi-Step Process Involving Structural Changes in Both the Ligand and the Receptor. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* **31**, 422-434, (2009).

184    Menting, J. G., Ward, C. W., Margetts, M. B. & Lawrence, M. C. A Thermodynamic

Study of Ligand Binding to the First Three Domains of the Human Insulin Receptor: Relationship between the Receptor Alpha-Chain C-Terminal Peptide and the Site 1 Insulin Mimetic Peptides. *Biochemistry* **48**, 5492-5500, doi:10.1021/bi900261q (2009).

185    Kabsch, W. Xds. *Acta Crystallographica. Section D, Biological Crystallography* **66**, 125-132, (2010).

186    Laemmli, U. K. Cleavage of Structural Proteins During the Assembly of the Head of Bacteriophage T4. *Nature* **227**, 680-685 (1970).

187    Lee, C., Levin, A. & Branton, D. Copper Staining: A Five-Minute Protein Stain for Sodium Dodecyl Sulfate-Polyacrylamide Gels. *Analytical Biochemistry* **166**, 308-312, (1987).

188    Blundell, T. L., Bedarkar, S., Rinderknecht, E. & Humbel, R. E. Insulin-Like Growth Factor: A Model for Tertiary Structure Accounting for Immunoreactivity and Receptor Binding. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 180-184 (1978).

189    Brogiolo, W. *et al.* An Evolutionarily Conserved Function of the Drosophila Insulin Receptor and Insulin-Like Peptides in Growth Control. *Current Biology : CB* **11**, 213-221 (2001).

190    Sajid, W. *et al.* Structural and Biological Properties of the Drosophila Insulin-Like Peptide 5 Show Evolutionary Conservation. *Journal of Biological Chemistry* **286**, 661-673, (2011).

191    Arquier, N. *et al.* Drosophila Als Regulates Growth and Metabolism through Functional Interaction with Insulin-Like Peptides. *Cell Metabolism* **7**, 333-338, (2008).

# Bibliography

The publications arising from the work presented in this thesis are included here.

Jiracek, J. *et al.* Implications for the active form of human insulin based on the structural convergence of highly active hormone analogues. *Proceeding of the National Acadamy of Sciences of America* **107**, 1966-1970 (2010).

Antolikova, E. *et al.* Non-equivalent role of inter- and intramolecular hydrogen bonds in the insulin dimer interface. *The Journal of Biological Chemistry* **286**, 36968-36977, (2011).

Menting, J.G. and Whittaker, J. *et al.* How insulin engages its primary binding site on the insulin receptor. *Nature* **493**, 241-245, (2013).

Zakova, L. *et al*. Structural Integrity of the B24 Site in Human Insulin Is Important for Hormone Functionality. *The Journal of Biological Chemistry* **288**, in press, (2013)

# Implications for the active form of human insulin based on the structural convergence of highly active hormone analogues

Jiří Jiráček[a], Lenka Žáková[a], Emília Antolíková[a], Christopher J. Watson[b], Johan P. Turkenburg[b], Guy G. Dodson[b], and Andrzej M. Brzozowski[b,1]

[a]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v.v.i., Flemingovo nám. 2, 166 10 Prague 6, Czech Republic and [b]York Structural Biology Laboratory, Department of Chemistry, The University of York, Heslington, York, YO10 5YW, United Kingdom

Insulin is a key protein hormone that regulates blood glucose levels and, thus, has widespread impact on lipid and protein metabolism. Insulin action is manifested through binding of its monomeric form to the Insulin Receptor (IR). At present, however, our knowledge about the structural behavior of insulin is based upon inactive, multimeric, and storage-like states. The active monomeric structure, when in complex with the receptor, must be different as the residues crucial for the interactions are buried within the multimeric forms. Although the exact nature of the insulin's induced-fit is unknown, there is strong evidence that the C-terminal part of the B-chain is a dynamic element in insulin activation and receptor binding. Here, we present the design and analysis of highly active (200–500%) insulin analogues that are truncated at residue 26 of the B-chain (B$^{26}$). They show a structural convergence in the form of a new β-turn at B$^{24}$-B$^{26}$. We propose that the key element in insulin's transition, from an inactive to an active state, may be the formation of the β-turn at B$^{24}$-B$^{26}$ associated with a *trans* to *cis* isomerisation at the B$^{25}$-B$^{26}$ peptide bond. Here, this turn is achieved with N-methylated L-amino acids adjacent to the *trans* to *cis* switch at the B$^{25}$-B$^{26}$ peptide bond or by the insertion of certain D-amino acids at B$^{26}$. The resultant conformational changes unmask previously buried amino acids that are implicated in IR binding and provide structural details for new approaches in rational design of ligands effective in combating diabetes.

β-turn | diabetes | peptide bond isomerisation | protein | structure

The peptide hormone insulin regulates blood glucose levels with a widespread impact on lipid and protein metabolism. It is a molecule of major therapeutic importance in the treatment of diabetes. The mature form of insulin is formed by two chains "A" and "B" with a B chain running from Phe$^{B1}$-Thr$^{B30}$ and an A chain Gly$^{A1}$-Asn$^{A21}$, stabilized by two inter and one intra chain disulphide bonds. Insulin's metabolic actions are expressed through binding as a monomer to the insulin receptor (IR). The structure of insulin, which has been known for four decades (1), has not provided insight into the mode of receptor binding and hormone activation. This is because detailed three-dimensional knowledge of insulin's complex structural behavior is limited to its inactive storage (hexameric, dimeric) states (2–4). The NMR structures of the monomeric form of insulin facilitated by mutations (5), applications of organic co-solvents (6) or truncation of the B-chain (7) merely confirm the conformations known from the inactive forms but also indicate intrinsic mobility of the N- and C termini of the B-chain. It has also been found that the N terminus of insulin can exist in so called T (extended) or R (helical) conformations, however, their role for insulin activation is still ambiguous (3, 4).

It is widely acknowledged that insulin must therefore undergo induced-fit structural changes upon binding to the IR because residues crucial for receptor interactions are hidden in the known structures of the native hormone (8–11). Although the exact nature of the changes that expose these residues to the receptor surface is unknown, there is strong evidence that the C-terminal part of the B-chain is a dynamic element in the transition that leads to IR binding (10, 12, 13). The structure of the insulin-IR complex is not known but extensive insulin sequence and mutagenesis studies have identified the likely functional sites responsible for key interactions with the receptor (4, 14–18). The main IR-binding site one consists of Leu$^{B11}$, Val$^{B12}$, Leu$^{B15}$, Gly$^{B23}$, Phe$^{B24}$, Phe$^{B25}$, and Tyr$^{B26}$ on chain B and Gly$^{A1}$, Ile$^{A2}$, Val$^{A3}$, Gln$^{A5}$, Tyr$^{A19}$, and Asn$^{A21}$ on chain A (19, 20). The shielding of the A-chain residues in site one by the ∼B$^{25}$-B$^{30}$ residues of B-chain C terminus in native insulin structures led to a widely accepted induced-fit hypothesis for insulin-IR binding, requiring this part of the B chain to be displaced (3, 4, 10).

As the lack of structural detail of the active form of insulin bound to the receptor restrains the progress in combating diabetes, we have tried to characterise the structural signatures of the active conformation of this hormone employing partial chemical synthesis/modification of highly-active (i.e., 200–500%) insulins, IR-binding studies and x-ray crystallography. Two types of "semisynthetic" insulins were prepared: one in which the nitrogen atom in the Phe$^{B25}$-Xaa$^{B26}$ peptide bond was methylated (NMe$^{B26}$), and others in which the chirality at the B$^{26}$ Cα was changed to the D-enantiomer. The nine insulin crystal structures reported here provide persuasive structural evidence concerning the nature of the induced-fit changes upon receptor binding.

## Results

**Highly active B$^{26}$-Shortened/N-Methylated Insulins.** Chemical changes were systematically introduced in the insulin molecule to achieve high-activity. Semisynthesis (21) of several unique and some previously reported (22, 23) insulin analogues with an N-methylated (NMe) Phe$^{B25}$-Xaa$^{B26}$ peptide bond or with a D-enantiomer at position B$^{26}$ yielded insulins with various binding activities (Table 1, Figs. S1–S3). These modifications indicate a clear pattern that correlates with high binding activity of the hormone: (*i*) shortening of its B-chain by deletion of residues B27-B30, (*ii*) the use of a carboxyamide (CONH$_2$) C terminus at position B$^{26}$, and (*iii*) a concurrent incorporation of NMe$^{B26}$ (22), D-amino acid (23) or D/L-Pro at position B$^{26}$. In contrast, an increase of the bulkiness of the B$^{26}$ amino acid in shortened analogues ([NMePhe$^{B26}$]-DTI-NH$_2$ (22), [NMeTyr$^{B26}$]-DTI-NH$_2$) or similar modifications in the full-length analogues [[NMeTyr$^{B26}$]-

**Table 1. Comparison of relative receptor binding affinity of some insulin analogues**

| B-chain position: | 24 | 25 | 26 | 27 | 28 | 29 | 30 | Affinity, % |
|---|---|---|---|---|---|---|---|---|
| Human insulin: | F | F | Y | T | P | K | T | 100 |
| [$N$MeAla$^{B26}$]-DTI-NH$_2$[†] | F | F | **Me A – CONH$_2$** | — | — | — | — | **465 (22)** |
| [$D$ – Pro$^{B26}$]-DTI-NH$_2$[†] | F | F | **$D$ – P – CONH$_2$** | — | — | — | — | **359‡** |
| [$N$MeHis$^{B26}$]-DTI-NH$_2$[†] | F | F | **Me H – CONH$_2$** | — | — | — | — | **214 (22)** |
| [$N$MeAla$^{B26}$]-insulin[†] | F | F | **Me A** | T | P | K | T | 21 (22) |
| [$N$MeTyr$^{B26}$]-insulin[†§] | F | F | **Me Y** | T | P | K | T | 21‡ |
| [ProB26]-DTI[†§] | F | F | **P – COOH** | — | — | — | — | **81‡** |
| [Pro$^{B26}$]-DTI-NH$_2$ | F | F | **P – CONH$_2$** | — | — | — | — | **359‡** |
| [$N$MePhe$^{B26}$]-DTI-NH$_2$ | F | F | **Me F – CONH$_2$** | — | — | — | — | 36 (22) |
| [$N$MeTyr$^{B26}$]-DTI-NH$_2$ | F | F | **Me Y – CONH$_2$** | — | — | — | — | 72‡ |
| [$D$ – Ala$^{B26}$]-DTI-NH$_2$ | F | F | **$D$ – A – CONH$_2$** | — | — | — | — | **~400(1252)¶** |

Relative receptor binding affinity is defined as (IC$_{50}$ of human insulin/IC$_{50}$ of analogue) ×100.

[†]Crystal structures presented in this report,

‡data presented in this study (the values of IC$_{50}$s are given in legends for Figs. S1–S3). Changes in B-chain sequence, high activities, and nonstandard chemical groups in bold;

§analogue for which both monomer and hexamer crystal structures have been determined, DTI: *des*-tetrapeptide; [$N$MeXaa$^n$]: methylation of the peptide bond N-atom preceding Xaa B$^n$ amino acid; -NH$_2$: *C*-terminal carboxyamide.

¶Only one analogue: [$D$-Ala$^{B26}$]-DTI-NH$_2$, 1,252% from ref. 23, was reevaluated here as it showed IR affinity within the range of other analogues semisynthesised and presented in this work.

insulin and [$N$MeAla$^{B26}$]-insulin (22)] resulted in low or moderate (21–72%) binding activities (Table 1).

**B-Chain of the Highly Active Insulins Possess a Unique β-Turn.** The apparent correlation between chemical composition and binding activity within the present insulin analogues prompted their crystallographic characterization (Tables S1–S4) in order to identify the underpinning structural features and to get further insight into the structural origins of the active conformation of insulin.

The most active (465%) insulin presented in this study: $N^{B26}$-methylated and B$^{26}$-truncated analogue [$N$MeAla$^{B26}$]-DTI-NH$_2$ (22), was crystallized as a monomer with a unique conformation at B$^{24}$-B$^{26}$. Its most striking feature is a type II β-turn at Phe$^{B24}$-$N$MeAla$^{B26}$ referred to here as the B26 turn (Fig. 1, Fig. S4$A$). This turn is associated with a *cis* conformation of the Phe$^{B25}$-$N$MeAla$^{B26}$ peptide bond. The B$^{24}$ CO group forms a 2.8 Å hydrogen bond with NH$_2$ of the terminal Ala$^{B26}$ carboxyamide, in which the NH$_2$ group mimics the peptide NH of the $i + 3$ (B$^{27}$) amino acid.

The most important consequence of the B26 turn is the departure of the B-chain B$^{22}$-B$^{30}$ β-strand from its typical hexamer/dimer like conformation (Fig. 1). The truncated B-chain *C*-termi-

nal segment now points away from IR site one making the latter accessible for a direct interaction with the IR. Another structural consequence of the B26 turn is a rotation of the Phe$^{B25}$ side chain that now overhangs the A-chain part of site one (∼7.7 Å from nearest Ile$^{A2}$) (Fig. 2, 4). In contrast, Phe$^{B24}$ remains in place. These B-chain changes are complemented by large movements of the A-chain, most notably the $N$-terminal helix is displaced by ∼3 Å and rotated by ∼40°, whereas Tyr$^{A19}$ shifts 2.9 Å away to form a hydrogen bond with Gln$^{A5}$ extending the non-polar receptor binding surface (Fig. 2). These "high-activity motifs" of the insulin structure were confirmed by the crystal structure of an another highly-active (214%) truncated analogue ([$N$MeHis$^{B26}$]-DTI-NH$_2$) (22, Fig. S5). These both highly-active insulins possess B26 turns with identical geometries (e.g., the corresponding Cβ atoms of His$^{B26}$ and Ala$^{B26}$ are only *ca.* 0.3 Å from each other) and hydrogen bond networks.

**$D$-Pro$^{B26}$ Mimics β-turn Resulting in High Activity.** To explain the structural role of $D$-amino acid chirality at position B$^{26}$ in achieving highly active (359%) insulin analogue (Table 1), the crystal structure of the truncated [$D$-Pro$^{B26}$]-DTI-NH$_2$ analogue was determined. The change of the chirality at the B$^{26}$ Cα center in this
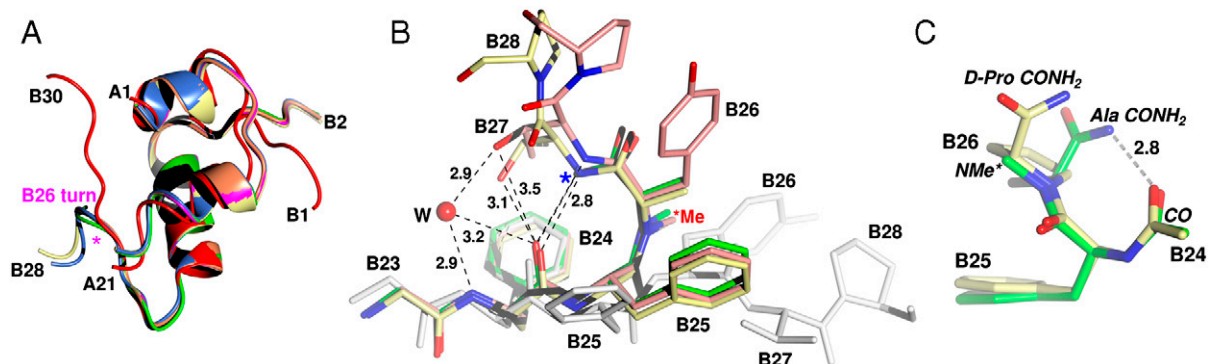


**Fig. 1.** Structural features of the highly-active insulin analogues. (*A*) An overlay of the general fold of crystal structures of the highly-active and other insulins reported here: *Magenta*: [$N$MeAla$^{B26}$]-DTI-NH$_2$, *Coral*: [$N$MeHis$^{B26}$]-DTI-NH$_2$, *green*: [$D$-Pro$^{B26}$]-DTI-NH$_2$, *Blue*: [$N$MeAla$^{B26}$]-insulin, *Yellow*: [$N$MeTyr$^{B26}$]-insulin (monomer structure), *Red*: wild type human insulin (32); the location of the B26 turn is marked by the *Star* in *Magenta*. (*B*) A close-up view of the B26 turn in the most representative analogues. N, O atoms in *Blue* and *Red*, resp. [except for normal human insulin (32) that is in *White*], C atoms color code: *Green* — [$N$MeAla$^{B26}$]-DTI-NH$_2$, *Yellow* — [$N$MeAla$^{B26}$]-insulin, *Pink* — [$N$MeTyr$^{B26}$]-insulin, *W* — water molecule, hydrogen bonds (*Dashed Lines*) lengths in Å. *Me (Red*) — $N$Me$^{B26}$, *Blue* * — indicates the three-dimensional convergence of N atoms from the B$^{26}$CONH$_2$ and B$^{26}$NH peptide groups. (*C*) Superposition (on B$^{23}$-B$^{25}$ Cα) of B26 turns in [$N$MeAla$^{B26}$]-DTI-NH$_2$ and [$D$-Pro$^{B26}$]-DTI-NH$_2$ showing a conformational convergence of peptide methylation ($N$Me$^{B26}$) and $D$-Pro$^{B26}$ chirality.
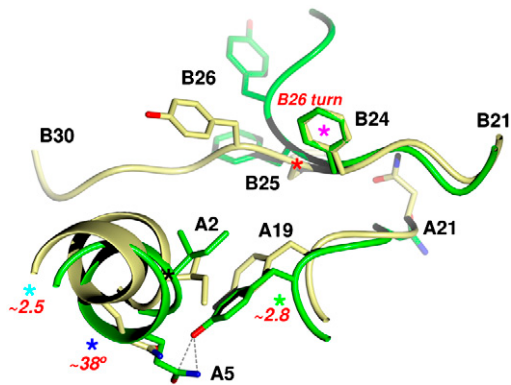
**Fig. 2.** Main structural changes occurring in insulin during a transition from ground-state to a proposed activated/active form based on the structures of some highly-active analogues; *Yellow* — human insulin (32); *Green* — [*N*MeTyr$^{B26}$]-insulin (isomorphous with the highly-active truncated [*N*MeAla$^{B26}$]-DTI-NH$_2$ but with the almost full B-chain structural definition), *Magenta Star* — Phe$^{B24}$ side-chain structural pivot, *Red Star* — Phe$^{B25}$ main chain structural pivot, *Black Star* — Ile$^{A2}$ Cα main chain pivot, *Green Star* — Tyr$^{A19}$ shift and formation of a hydrogen bond (*Dashed Lines*) with Gln$^{A5}$, *Blue Star* — rotation of the A1 helix (in degrees), *Cyan Star* — translational move of the A1 helix (in Å).

analogue caused by the introduction of D-Pro generates a local structure similar to that seen in the *cis* Phe$^{B25}$-Xaa$^{B26}$ *N*-methylated peptide bond high-activity insulin analogues. The absence of the *cis*-peptide bond in [D-Pro$^{B26}$]-DTI-NH$_2$ is compensated by the D-Pro in attaining a conformation similar to the B26 turn (Fig. 1C, Fig. S5). Although the $i + 3$ β-turn hydrogen bond is not formed here (CO$^i$-NH$^{i+3}$ distance of ~4.0 Å), the D-Pro residue positions the B-chain C-terminal CONH$_2$ moiety in a manner similar to that, observed in insulins containing a B26 turn; the NH$_2$-terminal group is only ~1.1 Å from its hydrogen-bonded equivalents in two other highly-active analogues. Hence, the conformations of the B-chain C-termini are associated with a β-turn in all three highly-active insulins and are correlated with their similar high affinity receptor binding.

As a very similar high activity (~360%) was also observed with the truncated L-Pro containing analogue with the carboxyamide group at the B-chain C terminus ([L-Pro$^{B26}$]-DTI-NH$_2$, Table 1), the shortened L-Pro analogues were also investigated by x-ray crystallography. However, only the monomeric and hexameric structures of C terminus-amide lacking truncated L-Pro homologue ([L-Pro$^{B26}$]-DTI, 81% affinity) could be crystallized and determined. They show that the B$^{20}$-B$^{26}$ segments are severely disordered in both crystals forms.

**Full-length Insulin Analogues Can Adopt B26 Turn.** The capacity of full-length insulin to adopt the B26 turn was investigated further in the crystal structures of monomers of the full-length analogues: [*N*MeTyr$^{B26}$]-insulin and [*N*MeAla$^{B26}$]-insulin (22). These variants also possess the B26 turn and all other general structural features characteristic for the most active truncated insulin molecules including the exposure of site one (Fig. 1A and B, Fig. 2, Fig. S4B). The B26 β-turn is stabilized in these analogues by a typical $i + 3$ β-turn hydrogen bond between Phe$^{B24}$ and Thr$^{B27}$ (CO$^{B24}$-NH$^{B27}$ distances of ~2.9–3.3 Å). As expected, the Thr$^{B27}$ peptide bond NH groups of the full-length analogues are spatially equivalent to CONH$_2$ terminus NH$_2$ group of the truncated insulins. The Cα$^i$-Cα$^{i+3}$ distances of ~6 Å are well within the ~7 Å limit typical for a type II β-turn. The section of the C-terminal segment that is well defined, Thr$^{B27}$-Pro$^{B28}$ (Lys$^{B29}$-Thr$^{B30}$ is disordered) follows the direction enforced by the B26 turn (identical to its conformation found in truncated insulins), and departs further from site one (Fig. 1A and B, Fig. 2)

that is, thus, fully revealed. The B26 β-turn is further stabilized here by additional hydrogen bonds between Phe$^{B24}$ CO with the side-chain OH of Thr$^{B27}$ and water molecules. As the crystallization experiments were carried out in parallel in the monomer- and hexamer-like conditions (pH 3.0 and pH 7.5–8.2 respectively) (Fig. S6), the hexamer crystals of the full-length [*N*MeTyr$^{B26}$]-insulin have also been obtained. Despite the presence of *N*Me group at position B26, this insulin analogue is still capable of forming a typical hexamer, in which the dimer interface adapts to the *N*Me-induced disruption as described before (24). Interestingly, crystallizations of the full-length [*N*MeAla$^{B26}$]-insulin, even under typical hexamer conditions, always yield the monomer crystal form. It has to be noted that some crystal structures of analogues described here are unique crystallographic descriptions of the full-length monomeric form of insulin.

**New Pseudo β-Turn Conformation of B-Chain *N* Terminus.** All monomers of the highly-active insulin analogues reported here also show a structural convergence of their *N* termini that adopts a unique conformation not previously described. Here, the main chain of B-chain *N* terminus follows the direction typical for the T-state till the Cα atom of the Phe$^{B5}$ at which it deviates from the T-fold forming, instead, a pseudo type II β-turn (Fig. 3). The turn is stabilized by the $i + 3$ hydrogen bond (3.0 Å) between main chain NH of the His$^{B5}$ ($i + 3$) and the side chain of the Asn$^{B3}$ that mimics the $i$- main chain CO group. This turn is stabilized further by an additional side chain Asn$^{B3}$- side chain (Nδ2) His$^{B5}$ hydrogen bond (3.0 Å). The Nδ1 atom of the His$^{B5}$ is also close to the main chain CO of the Thr$^{A8}$ (~3.3 Å) reflecting the importance of B$^5$ side chain in this unique formation. Only the main chain hydrogen bond: NH B$^6$-CO A$^6$ (2.8 Å) remains here from the network of B-chain—A-chain interactions that contribute to the stabilization of a typical T-state B chain *N* terminus.

## Discussion

**Structural Signatures of the Activated/Active Form of Insulin.** The conservation and three-dimensional convergence of specific structural features that lead to the exposure of buried, biologically active residues in chemically modified, highly-active insulin analogues discussed here suggest key structural traits that may be present in insulin's active conformation on the receptor or during its activation process. Although the evidence for the conformational motifs of the activated/active insulin conformation is indirect, the convergence of the structural signatures of highly-active hormone analogues observed here is to our mind quite persuasive.
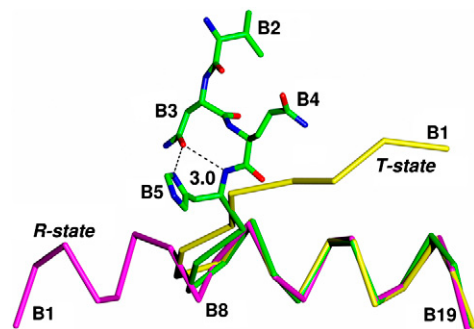


**Fig. 3.** Unique conformation of B-chain *N* terminus in the highly-active insulin analogues. One representative pseudo β-turn from the [*N*MeAla$^{B26}$]-DTI-NH$_2$ analogue is given in *Green* (only for the B$^2$-B$^5$ residues all atoms are given; Cαs trace for the B$^6$-B$^{19}$ region, hydrogen bonds in *Dash Lines*, distances in Å). Cα trace of the B$^1$-B$^{19}$ region typical for the T-state form of insulin (1 mso.pdb) in *Yellow*; the same region representative for the R-state of the hormone — derived from the hexamer form of the [L-Pro$^{B26}$]-DTI) — in *Magenta*.

It can be envisaged that the most striking element of the activated/active form of insulin is the formation of B26 turn associated with *trans*-to-*cis* isomerisation of the Phe[B25]-Tyr[B26] peptide bond. The Phe[B24](CO)-Thr[B27](NH) hydrogen bond—crucial for the stabilization of B26 turn—is probably also critical to the productive receptor binding in wild-type insulin because the inhibition of its formation by selective reduction of the B[24] CO group to $\psi$(CH$_2$-NH) in the native hormone molecule abolishes its activity (25). The B26 turn conformation at B[26] is found in all three truncated, highly-active insulins described here. Although the structure of the highly-active [L-Pro[B26]]-DTI-NH$_2$ insulin is not known the loss of potency in carboxyamide-lacking [L-Pro[B26]]-DTI correlates well with a loss of B26 turn. Furthermore, the NMR structures of another very potent [D-Ala[B26]]-DTI-NH$_2$ insulin (23) (reevaluated here) indicated mobility of the B[26] region. It is to be expected that there is conformational freedom in this structural element and that the formation of B26 turn can be readily achieved.

Therefore, we propose that these structural changes (formation of B26 turn and *trans-cis* isomerisation of the Phe[B25]-Tyr[B26] peptide bond) follow the detachment of the Tyr[B26]-Thr[B30] segment from the insulin molecule during, or on, binding to the receptor. These observations are consistent with the induced-fit theory of insulin-IR binding and with *trans/cis* peptide bond isomerisation analyses (26–29). The protein environment supplied by IR is important in the unmasking and stabilisation of the active surfaces of insulin. We suggest that the analogues with higher affinity described here have an active surface more representative of the conformation of insulin bound to the receptor than do the structures of the storage forms of the native hormone. The scale and nature of the structural rearrangements leading to highly active analogues underlines a "chaperone-like" role of the IR in insulin activation (30), especially if *trans-cis* peptide isomerisation is one of the key steps.

The restructuring of the B-chain and the concerted changes to the *N*-terminal helix of the A-chain reveal a large hydrophobic area made up of Gly[A1], Ile[A2], Val[A3], and Tyr[A19] priming them for direct interactions with the IR (Fig. 2, 4). The importance of Ile[A2] to receptor binding is indicated by its complete exposure on the IR-binding surface and by the striking spatial invariance of its Cα atom in these transformations. The structural conservation of the A[2] Cα suggests its role as the A-chain main chain pivot in insulin activation. The shift of Tyr[A19] into hydrogen bond contact with Gln[A5] further stabilises the A-chain IR-binding surface with the edge of the Tyr[A19] ring transforming the Gly[A1]-Ile[A2]-Val[A3] hydrophobic surface into a three-rim (Tyr[A19], Gly[A1]-Ile[A2], and Val[A3]) evenly spaced "epitope" (Fig. 4).

The key structural roles of Phe[B24] and Phe[B25] in insulin activity, well known from several studies (4, 15), are firmly confirmed by structures of the highly-active insulins described here. The aromatic ring of Phe[B24] is spatially invariant in native hormone and all other insulin analogues. This is in contrast to the ~90° rotation of Phe[B25] side chain that subsequently overhangs (by ~7 Å) the A-chain part of site one (Tyr[A19]-Gly[A1]-Ile[A2]-Val[A3]) "epitope" in the highly active insulins. However, in spite of its considerable side-chain movement the position of the Phe[B25] Cα atom is well conserved in different forms of insulin, confirming further the role of Phe[B24] in stereo-specific detachment of the B-chain β-strand (10). Therefore, it may be suggested that the Phe[B24] is a B-chain side-chain pivot during insulin activation (Fig. 2). Phe[B24] is structurally invariant side chain that, being anchored into the hydrophobic cavity lined up by Leu[B15], Val[B12], Tyr[B16] (its side chain collapses on this cavity in the activation process sealing it off) and Cys[B19], provides structural stability during rearrangement of the *C*-terminal segment of the B-chain. Phe[B25] complements the role of Phe[B24] in this process acting as the B-chain main chain pivot with the B[25] Cα as the first B-main chain turning point in the B26 turn and, subsequently, initiation of IR binding (Fig. S7).

In addition to the structural convergence of the B[24]-B[26] region in the highly active insulin analogues the consistent similarity of the unique B[3]-B[5] pseudo β-turn of B-chain *N* termini observed there should be also noted. This turn locates the *N* terminus conformation in-between the T- and R-forms. However, the physiological relevance of this unique state of the hormone (e.g., as a snapshot of the T-R transition or as its more stable high-activity related form) is unclear and requires further studies.

### Lower Activity Insulin Analogues Fit the B26 Turn-Related Activation Model.

Interestingly, some of the B[26]-truncated analogues: [*N*MeTyr[B26]]-DTI-NH$_2$ and [*N*MePhe[B26]]-DTI-NH$_2$ (22), show lower activities, 72% and 36% respectively (Table 1). At present, there is no straightforward structural explanation for this behavior. In the native, full-length insulin the replacement of Tyr[B26] to Ala (without any *N*-methylation) results in a lowering of its activity to ~36% (14). However, the presence of a more bulky aromatic side chains (Phe[B26]/Tyr[B26]) in the B[26]-truncated (and *N*-methylated) analogues may result, for example, in a detrimental, entropic effect leading to the destabilisation of the short B26 turn. It is interesting that these analogues have not crystallized, suggesting an increased flexibility that has also been observed by NMR for a related truncated insulin analogue (31).

The crystal structures of the full-length analogues result in intriguing observations. Firstly, only the monomer structures of [*N*MeAla[B26]]-insulin have been obtained under both "monomer" and "hexamer" crystallization conditions whilst both monomer and hexamer crystals of [*N*MeTyr[B26]]-insulin were grown in the relevant monomer and hexamer crystallization environments. This indicates that the monomer conformation is preferentially stabilized by Ala in the B[26] position, consistent with the high
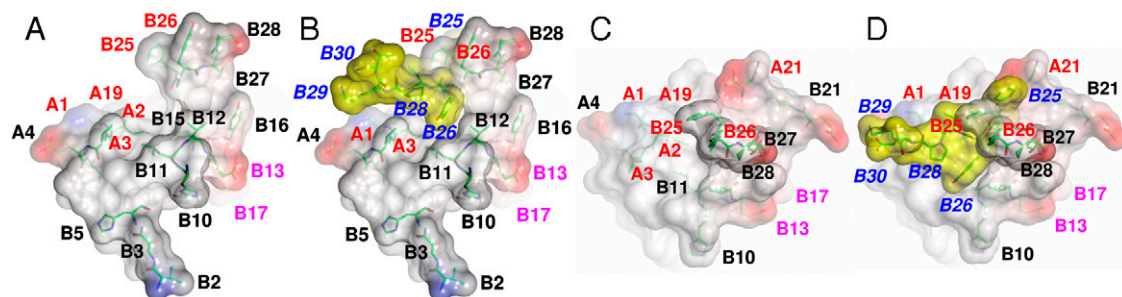


**Fig. 4.** Molecular shape of a putative active form of insulin based on the electrostatic surface of the [*N*MeTyr[B26]]-insulin analogue as the representative of B26 turn-containing insulins. (*A*) A "side" view showing the formation of the hydrophobic IR-binding hand-like insulin shape composed of site one residues (*Red Numbering*), and lined by Leu[B11], Leu[B15]; labels for some residues of site two are in *Magenta*. (*B*) As (*A*) but with B25-B30 chain of the wild type human insulin modeled in (32) (*Yellow Surface, Blue-Italic Numbering*) showing the obstruction of site one prior to the formation of the B26 turn. (*C-D*) "Top" views of insulin surfaces equivalent to the (*A* and *B*) representations after 90° rotation.

potency of the B[26] truncated insulin. The reduced potencies of [$N$MeTyr[B26]]-insulin and [$N$MeAla[B26]]-insulin analogues (~21%, Table 1) confirm the trend that the full-length analogues are generally less active than their truncated homologues (4) and suggest that this may result from several disadvantageous steric and entropic factors originating in the B[27]-B[30] region. For example, (*i*) the B[27]-B[30] segment may require a significant contribution from the IR to be induced to a fully active conformer of insulin, (*ii*) the structural constrain of the B26 turn may not impose the optimal conformation for the B[27]-B[30] residues, and (*iii*) the B26 turn may not represent the final and the most optimum active insulin conformation (Fig. S7).

**Molecular Shape of Active Insulin.** The structure-function relation-ships of the highly-active insulin analogues described here not only give an insight into detailed definition of the putative local structural signatures of an active form of insulin, but shed also light on the possible molecular shape of the hormone's active conformer. In the process of insulin activation the Phe[B25] and the B26 turn form a hydrophobic knob that together with the A-chain "epitope" defines a unique, highly hydrophobic IR-bind-ing cleft, lined by Leu[B15], Leu[B11], and Val[B12] side chains [Fig. 4; it should be mentioned that the normal insulin B[25]-B[30] substructure modeled in Fig. 4B is actually in more intimate contacts with other A- and B-chain residues in the ground-state insulin mole-cule due to a different relative orientation of these chains in highly-active insulin analogues and its ground form (Fig. 2)]. The shape of this cleft is reminiscent of a macromolecular hand that ensures the grip on the receptor by a hydrophobic thumb (Phe[B25] and B26 turn), a hydrophobic palm (Leu[B15]/Leu[B11]/Val[B12]), and hydrophobic fingers (Tyr[A19], Gly[A1]-Ile[A2], and Val[A3]).

The molecular features of several highly-active, and related, insulin analogues discussed here provide unique structural elements for the conformational landscape of the hormone. They give considerable insight into the activation of human insulin and should help in development of unique types of hormone ana-logues for the treatment of diabetes mellitus.

## Material and Methods

**Semisynthesis of Insulin Analogues and Binding Assays.** In general, solid-phase synthesis of peptides, enzymatic semisynthesis of analogues, purification of analogues, preparation of rat adipose membranes, and binding assays were performed as described previously (21, 22) (*SI Text*).

**Crystallization, Structure Determination, and Refinement.** Crystallization con-ditions (Table S1) were found initially by an in-house developed screen based on the available insulin crystallization protocols and they were refined indi-vidually by the hanging drop method. X-ray data (Tables S2–S4) were pro-cessed by HKL2000 (33) and all computational crystallography was performed by the CCP4 suite of programs (34). Structures were determined by molecular replacement with differently truncated 1mso.pdb-derived monomeric insulins as the model (32). Crystallization details, x-ray data, re-finement statistics, and model building protocols are provided in *SI Text*.

1. Adams MJ, et al. (1969) Structure of rhombohedral 2 zinc insulin crystals. *Nature*, 224:491–495.
2. Derewenda U, et al. (1991) X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule. a completely inactive analogue. *J Mol Biol*, 220:425–433.
3. Weiss MA (2009) *Insulin and IGFs*, ed Litwack G (Elsevier Academic, Inc, San Diego), pp 33–49.
4. Mayer JP, Zhang F, DiMarchi RD (2007) Insulin structure and function. *Biopolymers*, 88:687–713.
5. Ludvigsen S, Roy M, Thogersen H, Kaarsholm NC (1994) High-resolution structure of an engineered biologically potent insulin monomer, B16 Tyr − − > His, as determined by nuclear magnetic resonance spectroscopy. *Biochemistry*, 33:7998–8006.
6. Bocian W, et al. (2008) Structure of human insulin monomer in water/acetonitrile solution. *J Biomol NMR*, 40:55–64.
7. Hua QX, Weiss MA (1991) Comparative 2D NMR studies of human insulin and des-pentapeptide insulin: Sequential resonance assignment and implications for protein dynamics and receptor recognition. *Biochemistry*, 30:5505–5515.
8. Hua QX, Shoelson SE, Kochoyan M, Weiss MA (1991) Receptor binding redefined by a structural switch in a mutant human insulin. *Nature*, 354:238–241.
9. Ludvigsen S, Olsen HB, Kaarsholm NC (1998) A structural switch in a mutant insulin exposes key residues for receptor binding. *J Mol Biol*, 279:1–7.
10. Xu B, et al. (2009) Decoding the cryptic active conformation of a protein by synthetic photoscanning insulin inserts a detachable arm between receptor domains. *J Biol Chem*, 284:14597–14608.
11. McKern NM, et al. (2006) Structure of the insulin receptor ectodomain reveals a folded-over conformation. *Nature*, 443:218–221.
12. Zoete V, Meuwly M, Karplus M (2004) A comparison of the dynamic behavior of mono-meric and dimeric insulin shows structural rearrangements in the active monomer. *J Mol Biol*, 342:913–929.
13. Pittman I, Tager HS (1995) A spectroscopic investigation of the conformational dynamics of insulin in solution. *Biochemistry*, 34:10578–10590.
14. Kristensen C, et al. (1997) Alanine scanning mutagenesis of insulin. *J Biol Chem*, 272:12978–12983.
15. Mirmira RG, Nakagawa SH, Tager HS (1991) Importance of the character and config-uration of residues B24, B25, and B26 in insulin-receptor interactions. *J Biol Chem*, 266:1428–1436.
16. Brange J, Owens DR, Kang S, Volund A (1990) Monomeric insulins and their experi-mental and clinical implications. *Diabetes Care*, 13:923–954.
17. Haneda M, Chan SJ, Kwok SC, Rubenstein AH, Steiner DF (1983) Studies on mutant human insulin genes: Identification and sequence analysis of a gene encoding [SerB24]insulin. *P Natl Acad Sci USA*, 80:6366–6370.
18. Xu B, et al. (2002) Chiral mutagenesis of insulin's hidden receptor-binding surface: Structure of an allo-isoleucine(A2) analogue. *J Mol Biol*, 316:435–441.
19. Pullen RA, et al. (1976) Receptor-binding region of insulin. *Nature*, 259:369–373.
20. De Meyts P, Whittaker J (2002) Structural biology of insulin and IGF1 receptors: Implications for drug design. *Nat Rev Drug Discov*, 1:769–783.
21. Zakova L, et al. (2007) The use of Fmoc-Lys(Pac)-OH and penicillin G acylase in the preparation of novel semisynthetic insulin analogs. *J Pept Sci*, 13:334–341.
22. Zakova L, et al. (2008) Insulin analogues with modifications at position B26. Divergence of binding affinity and biological activity. *Biochemistry*, 47:5858–5868.
23. Kurapkat G, et al. (1999) The solution structure of a superpotent B-chain-shortened single- replacement insulin analogue. *Protein Sci*, 8:499–508.
24. Zakova L, et al. (2004) Toward the insulin-IGF-I intermediate structures: Functional and structural properties of the [Tyr(B25)NMePhe(B26)] insulin mutant. *Biochemistry*, 43:16293–16300.
25. Nakagawa SH, Johansen NL, Madsen K, Schwartz TW, Tager HS (1993) Implications of replacing peptide bonds in the COOH-terminal B chain domain of insulin by the $\psi(CH_2-NH)$ linker. *Int J Pept Protein Res*, 42:578–584.
26. Pal D, Chakrabarti P (1999) Cis peptide bonds in proteins: residues involved, their conformations, interactions and locations. *J Mol Biol*, 294:271–288.
27. Sykora D, Zakova L, Budesinsky M (2007) High-performance liquid chromatography and nuclear magnetic resonance study of linear tetrapeptides and octapeptides containing *N*-methylated amino acid residues. *J Chromatogr A*, 1160:128–136.
28. Fischer G (2000) Chemical aspects of peptide bond isomerisation. *Chem Soc Rev*, 29:119–127.
29. Harrison RK, Stein RL (1992) Mechanistic studies of enzymatic and nonenzymic prolyl cis-trans isomerization. *J Am Chem Soc*, 114:3464–3471.
30. Ward CW, Lawrence MC (2009) Ligand-induced activation of the insulin receptor: A multi-step process involving structural changes in both the ligand and the receptor. *Bioessays*, 31:422–434.
31. Hua QX, Ladbury JE, Weiss MA (1993) Dynamics of a monomeric insulin analogue: Testing the molten-globule hypothesis. *Biochemistry*, 32:1433–1442.
32. Smith GD, Pangborn WA, Blessing RH (2003) The structure of T-6 human insulin at 1.0 angstrom resolution. *Acta Crystallogr D*, 59:474–482.
33. Otwinowski Z, Minor W (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol*, 276A:307–326.
34. Bailey S (1994) Collaborative computational project, number 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D*, 50:760–763.

# Non-equivalent Role of Inter- and Intramolecular Hydrogen Bonds in the Insulin Dimer Interface*[S]

Emília Antolíková[‡], Lenka Žáková[‡], Johan P. Turkenburg[§], Christopher J. Watson[§], Ivona Hančlová[‡],
Miloslav Šanda[‡], Alan Cooper[¶], Tomáš Kraus[‡], A. Marek Brzozowski[§1], and Jiří Jiráček[‡2]

From the [‡]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v.v.i., Flemingovo nám. 2, 166 10 Prague 6, Czech Republic, the [§]York Structural Biology Laboratory, Department of Chemistry, The University of York, Heslington, York YO10 5YW, United Kingdom, and the [¶]School of Chemistry, Glasgow University, College of Science and Engineering, Glasgow G12 8QQ, Scotland, United Kingdom

Apart from its role in insulin receptor (IR) activation, the C terminus of the B-chain of insulin is also responsible for the formation of insulin dimers. The dimerization of insulin plays an important role in the endogenous delivery of the hormone and in the administration of insulin to patients. Here, we investigated insulin analogues with selective *N*-methylations of peptide bond amides at positions B24, B25, or B26 to delineate their structural and functional contribution to the dimer interface. All *N*-methylated analogues showed impaired binding affinities to IR, which suggests a direct IR-interacting role for the respective amide hydrogens. The dimerization capabilities of analogues were investigated by isothermal microcalorimetry. Selective *N*-methylations of B24, B25, or B26 amides resulted in reduced dimerization abilities compared with native insulin ($K_d = 8.8$ $\mu$M). Interestingly, although the *N*-methylation in [*N*MeTyrB26]-insulin or [*N*MePheB24]-insulin resulted in $K_d$ values of 142 and 587 $\mu$M, respectively, the [*N*MePheB25]-insulin did not form dimers even at high concentrations. This effect may be attributed to the loss of intramolecular hydrogen bonding between NHB25 and COA19, which connects the B-chain $\beta$-strand to the core of the molecule. The release of the B-chain $\beta$-strand from this hydrogen bond lock may result in its higher mobility, thereby shifting solution equilibrium toward the monomeric state of the hormone. The study was complemented by analyses of two novel analogue crystal structures. All examined analogues crystallized only in the most stable R$_6$ form of insulin oligomers (even if the dimer interface was totally disrupted), confirming the role of R$_6$-specific intra/intermolecular interactions for hexamer stability.

Insulin is an important polypeptide hormone that controls a wide range of cellular processes such as the regulation of blood glucose uptake and has a large impact on protein and lipid metabolism. However, despite decades of intensive research, many questions about the structure of insulin and its mechanism of action remain. The solid state-based structural insight into the insulin molecule is limited to inactive dimeric or hexameric storage forms (1–3), whereas the insulin monomer represents the active form of the hormone when binding to the insulin receptor (IR).[3] It is also widely accepted that insulin undergoes a profound structural change during this process (4–6), a hypothesis supported by a plethora of highly dynamic hormone conformers identified by NMR studies (7–13). Attempts to determine the structure of the insulin-IR complex have been unsuccessful so far. However, the regions of the insulin molecule responsible for the interaction with the IR (3, 14) or for its dimerization and hexamerization (15, 16) have been functionally and structurally identified in a number of insulin analogues.

The insulin molecule consists of two peptide chains, a 21-amino acid A-chain and a 30-amino acid B-chain, interconnected by two interchain and one intrachain disulfide bridges. The C terminus of the B-chain of insulin, particularly residues B24–B26, plays a substantial role in the initial contact with the receptor. It is believed that the C terminus of the B-chain of insulin must be detached away from the central B-chain $\alpha$-helix of insulin (2, 6). One of the main signatures of this so-called "active form" of insulin should be the exposure of the previously hidden amino acids Gly-A1, Ile-A2, and Val-A3, which are important for the interaction with IR (3). Recently, we described crystal structures of several shortened and full-length insulin analogues with modifications at the B26 position (17). The structural convergence of some of these highly active analogues (200–400%) enabled us to postulate that the active form of human insulin is characterized by a formation of a new type II $\beta$-turn at positions B24–B26.

Besides its role in IR activation and IR negative cooperativity (18, 19), the C terminus of the B-chain is also responsible for the formation and stabilization of the insulin dimers that are the building blocks of storage hexamers. In addition, the dimeriza-

---

[3] The abbreviations used are: IR, insulin receptor; DOI, des(B23–B30)-octapeptide insulin; DTI, des(B27–B30)tetrapeptide insulin; Fmoc, fluorenylmethoxycarbonyl; ITC, isothermal titration microcalorimetry.

tion of insulin plays an important role in the endogenous delivery of the hormone from the pancreas to the circulatory system and in the administration of insulin to diabetic patients (3, 20–22). The structures of insulin dimers and hexamers have been determined from several crystal structures (*e.g.* Refs. 15, 23, 24). They revealed that the key interactions between insulin molecules I and II of the dimer are complementary and involve reciprocal hydrogen bonds between the carbonyl oxygen of Tyr-B26 (Phe-B24) of one molecule and the amide hydrogen of Phe-B24 (Tyr-B26) of the complementary second molecule of the dimer. Hence, antiparallel β-strands of the C termini of B-chains are connected by four hydrogen bonds. It has also been found that the amide hydrogen of Phe-B25 of molecule II forms an intramolecular hydrogen bond with the carbonyl of Tyr-A19. The side chains of Phe-B24, Tyr-B26, Tyr-B16, and Pro-B28 reinforce further the stabilization of the dimer by a network of Van der Waals contacts (15).

Our previous studies on the structural and functional relevance of the B26 (17, 25, 26) and B25 (27) positions of insulin focused on modulation of hormone conformation by the incorporation of an *N*-methyl group on the amide nitrogen of the peptide bond in shortened and full-length insulin analogues. Subsequently, we investigated the selective application of this *N*-methylation within the B24–B26 region of insulin by the systematic elimination of the inter/intramolecular hydrogen bonds, which are important for the stability and folding of the insulin monomer, dimer formation, and interaction with IR. Therefore, the approach reported here represents an application of a very specific, chemically controlled molecular tool for the functional assessment of the hydrogen bond network related to the dimer interface.

Here, we present the synthesis and characterization of several novel full-length insulin analogues with the selective *N*-methylation of the peptide bond at position B24, B25, or B26 to delineate the role of each of these particular peptide units in insulin dimer/hexamer assembly. In this process, we correlate the binding affinities of the analogues to IR (determined in rat adipose membranes) using isothermal titration microcalorimetry (ITC) dilution experiments, performed to assess the thermodynamic contributions toward hormone dissociation. The wild-type insulin (capable of dimer formation) and des(B23-B30)octapeptide insulin (DOI, not able to dimerize (28, 29)) were used as references for the dimeric and monomeric forms, respectively. This study is also complemented by the analysis of two novel crystal structures and two previously published analogue crystal structures that have not been hitherto discussed in the context presented here. This approach allowed us to shed light on the contributions of individual peptide bond amides to the dimerization and binding affinity of insulin.

## MATERIALS AND METHODS

*Materials*—2-Chlorotrityl chloride resin, protected amino acids, and reagents for solid phase synthesis of peptides were purchased from Novabiochem Merck (Laufelfingen, Switzerland). Fmoc-Lys(phenylacetyl)-OH was prepared as described previously (30). Tosyl phenylalanyl chloromethyl ketone-treated trypsin was purchased from Sigma, and penicillin G acylase was from Fluka. Human [125I]monoiodotyrosylA14-in-

sulin was purchased from PerkinElmer Life Sciences. Human and porcine insulins were purchased from Sigma. All other chemicals and solvents were obtained from Sigma-Aldrich.

*Peptide Synthesis and Enzymatic Semisynthesis*—The syntheses of peptides and the semisyntheses of analogues were performed according to Zakova *et al.* (30). The identity of peptides and analogues was confirmed with a Fourier transform mass spectrometer LTQ Orbitrap XL (Thermo Fisher).

*Receptor Binding Studies*—Receptor binding studies with plasma membranes prepared from epididymal adipose tissue of adult male Wistar rats were performed according to Zakova *et al.* (26).

*Isothermal Titration Microcalorimetry*—Human and porcine insulin (Sigma) were used without further purification. DOI was prepared in our laboratory from porcine insulin by tryptic cleavage (31). Purification was performed using reverse-phase HPLC, and identification was confirmed with electrospray ionization mass spectrometry. The analogues were prepared as described above.

Peptide solutions (~1 mM) were prepared in 0.1 M glycine/HCl buffer, pH 2.5 (total volume of 1 ml). Peptide solutions were dialyzed in dialysis cassettes (Slide-A-Lyzer cut-off, molecular weight 3,500, Pierce) against 600 ml of the buffer at 4 °C. The buffer was changed three times (after 3, 20, and 26 h). Following dialysis, the concentrations of peptides were determined from UV absorbance at 280 nm (insulin, $\epsilon = 5,840$ M$^{-1}$ cm$^{-1}$; DOI, $\epsilon = 4,560$ M$^{-1}$ cm$^{-1}$). Concentrations were confirmed by amino acid analysis (± 10%). The final dialysis buffer and peptide solutions were briefly degassed prior to ITC dilution titration.

Calorimetric experiments were performed using a VP-ITC® apparatus (MicroCal, Inc.) operated in dilution/dissociation mode, with the reaction vessel (1.417 ml) held at 25 °C. The instrument was temperature-equilibrated prior to the start of injection. The sample contents were stirred at a speed of 307 rpm over the duration of the titration. A typical experiment comprised of 20 injections of 12.5 μl (first injection of 2 μl) of insulin solution into the reaction vessel, initially loaded with dialysis buffer. The duration of each injection was 25 s, with a time interval between injections of 300 s.

Integrated heat pulse data were corrected for small injection/mixing effects from controls that were performed separately and analyzed under identical conditions by omitting the first injection and using MicroCal LLC Origin software (version 7). The software employed an updated and corrected (June 2008) version of the dissociation analysis procedure and was validated by comparison with earlier analysis methods (32, 33). For human insulin, the experimental values (± S.E.) were determined from eight parallel experiments. For porcine insulin, DOI, and insulin analogues, the experimental values (± range) were determined from two parallel experiments. The initial concentrations in the injection syringe ranged from 0.208 to 1.220 mM (supplemental Table S1).

*X-ray Studies*—Crystallization of the analogues ([*N*MePheB24]-insulin, [*N*MePheB25]-insulin, and [TyrB25,*N*MePheB26,LysB28,ProB29]-insulin) were performed with the in-house insulin crystallization screens covering most of the reported crystal growth parameters of insulin. Crystallization

**TABLE 1**

Values of $IC_{50}$ and relative receptor binding affinities of human insulin and insulin analogues

| Peptide | $IC_{50} \pm S.E.^a$ | Potency[b] |
|---|---|---|
| | $n_M$ (n) | % |
| Human insulin | 0.89 ± 0.06 (3) | 100[c] |
| [NMePheB24]-insulin | 30.1 ± 2.2 (3) | 2.96[c] |
| [NMePheB25]-insulin | 515 ± 32 (3) | 0.17[c] |
| [NMeTyrB26]-insulin | 4.30 ± 0.46 (3) | 20.7[d] |
| [PheB26]-insulin | 1.94 ± 0.31 (3) | 45.9[c] |
| [NMePheB26]-insulin | 21.3 ± 2.7 (3) | 4.17[c] |
| [NMePheB25,NMePheB26]-insulin | 201 ± 39 (3) | 0.44[c] |

[a] $IC_{50}$ values represent concentrations of insulin or the analogue that cause half-maximal inhibition of binding of human [$^{125}$I]monoiodotyrosylA14-insulin to IR. Each value represents the mean ± S.E. of multiple determinations (n).

[b] Relative receptor binding affinity is defined as ($IC_{50}$ of human insulin/$IC_{50}$ of analogue) × 100.

[c] This study.

[d] From Ref. 17.

conditions, data collection, and refinement and model statistics are given in supplemental Table S2. X-ray data were processed by HKL2000 (34), and model building and refinement were performed by the CCP4 suite of programs (35) and COOT (36). Crystal structures were solved by Molrep (37) with B1–B6 and B23–B30-truncated insulin hexamer or monomer as a model (Protein Data Bank code 1MSO) (24) and refined by Refmac5.6 (38). All figures were made by CCP4mg (35). For structural comparisons of the dimer interfaces, the best defined (*i.e.* with the best electron density for typically very mobile Phe-B25 side chain and B28–B30 amino acids) dimers (referred here as chains AB and CD) from the relevant structures were superimposed on the B9–B19 and D9–D19 Cα atoms by the LSQKAB program from the CCP4 suite. Both reported crystal structures contain one full hexamer in the asymmetric unit. As the nature of changes along the dimer interface observed here was virtually identical within each of the three dimers related by hexamer symmetry, only one dimer was selected for comparisons as a fully representative structural entity.

## RESULTS

*Binding Affinities*—Insulin analogues were prepared by trypsin-catalyzed semisynthesis, and their binding affinities to the insulin receptor in membranes from rat adipose tissue were determined (Table 1 and Fig. 1). Four of these analogues are novel molecules, with [PheB26]-insulin having been already reported by Gauguin *et al.* (39). The crystal structure and binding affinity of [NMeTyrB26]-insulin was recently reported by us elsewhere (17). The first three analogues differ in the position of N-methylation of the peptide bond amide at the B24, B25, or B26 positions of human insulin. Additionally, two insulin analogues were also produced in which Tyr at the B26 position was replaced with either Phe or N-MePhe. The last analogue contains two mutations, *i.e.* N-MePhe residue in both B25 and B26 positions.

All analogues showed diminished binding affinity compared with human insulin. The most drastic reduction in binding was observed in the case of analogues N-methylated at the B25 position (<1%). The N-methylation of the B24 amide also had a highly negative effect on the binding affinity of the respective analogue (3%). In contrast, the B26 position was much more tolerant to the modifications, with [PheB26]-insulin displaying
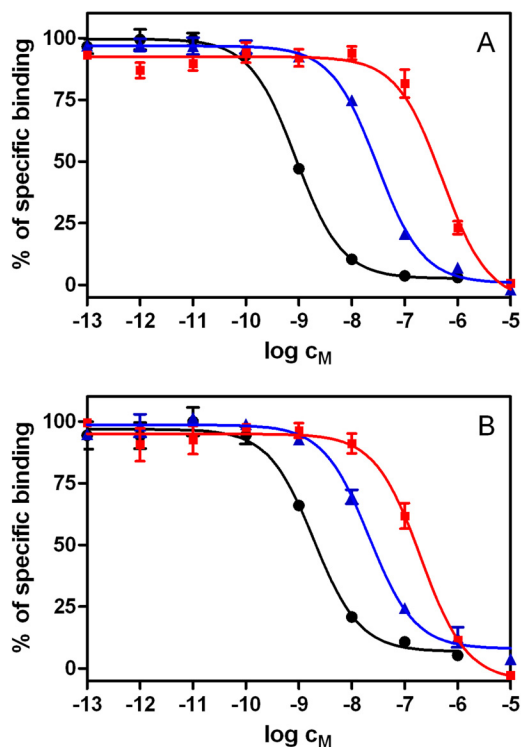


FIGURE 1. **Inhibition of binding of human [$^{125}$I]-insulin to adipose plasma membranes by insulin and insulin analogues.** *A*, human insulin (*black circles*), [NMePheB24]-insulin (*blue triangles*), and [NMePheB25]-insulin (*red squares*). *B*, [PheB26]-insulin (*black circles*), [NMePheB26]-insulin (*blue triangles*), and [NMePheB25,NMePheB26]-insulin (*red squares*). Quantitative information is provided in Table 1. See "Materials and Methods" for details.

only moderate reduction of the affinity (45.9%) (very similar to its binding affinity (47.6%), determined using an IM-9 cells binding assay (39)). N-Methylation of Tyr-B26 reduces the binding affinity of the analogue to 20.7%, and the combination of both N-methylation and the substitution of Phe for Tyr at B26 results in a very weak binding of this insulin analogue (4.2%).

*ITC Measurements*—The key part of our study was devoted to the investigation of the dimerization ability of insulin analogues by ITC dilution measurements that detect heat energy changes upon the dissociation of insulin dimers. The dimerization properties of the analogues were set against human/porcine insulin and DOI as benchmarks of regular and non-dimerization hormone behavior. The results of titration experiments are summarized in Table 2, and the complete data from all titration experiments used for the calculation of kinetic and thermodynamic parameters are shown in supplemental Table S1. The dilution experiments with human or porcine insulin yielded sequences of endothermic heat pulses characteristic of molecular dissociation as described previously (32). As the insulin concentration in the calorimeter cell increased with successive injections, dissociation decreased, and the magnitude of the heat uptake diminished accordingly, giving heat typical of dilution curves for human insulin, as shown in Fig. 2*A*. The example of the dilution process giving no heat effect is shown in Fig. 2*B* (dilution curve of [NMePheB25]-insulin). Dilution curves for porcine insulin, DOI, [NMePheB24]-insulin, [NMeTyrB26]-insulin, and [NMePheB26]-insulin are

**TABLE 2**

**ITC analyses of dimerization capabilities of insulin and insulin analogues**

The experimental data were analyzed without first injection using LLC ITC Origin 7 software. The experimental values for human insulin are means of eight parallel experiments ± S.E. ($n = 8$). The experimental values for porcine insulin, DOI, and insulin analogues were determined from two parallel experiments and are provided as means ± range ($n = 2$) with the exception of DOI and [*N*MePheB25]-insulin, where no heat effect was observed. The ITC experiments were performed at injection syringe concentrations of insulins ranging from 0.213 mM to 1.22 mM. The complete ITC data are shown in supplemental Table S1.

| Insulin | $K_d$ | $\Delta H^o_d$ | $\Delta G^o_d$ | $\Delta S^o_d$ |
|---|---|---|---|---|
| | μM | kJ/mol | kJ/mol | J/K·mol |
| Porcine insulin ($n = 2$) | 9.03 ± 0.50 | 41.96 ± 0.45 | 28.78 ± 0.14 | 44.22 ± 1.06 |
| Human insulin ($n = 8$) | 8.81 ± 1.05 | 56.93 ± 2.56 | 28.95 ± 0.28 | 93.90 ± 8.15 |
| DOI ($n = 2$) | No heat effect | | | |
| [*N*MePheB24]-insulin ($n = 2$) | 587 ± 99 | 10.82 ± 0.46 | 18.47 ± 0.42 | −25.68 ± 2.97 |
| [*N*MePheB25]-insulin ($n = 2$) | No heat effect | | | |
| [*N*MeTyrB26]-insulin ($n = 2$) | 142 ± 30 | 70.85 ± 15.50 | 22.02 ± 0.53 | 163.86 ± 53.75 |
| [*N*MePheB26]-insulin ($n = 2$) | 1240 ± 10 | 37.84 ± 8.34 | 16.58 ± 0.02 | 71.34 ± 28.07 |
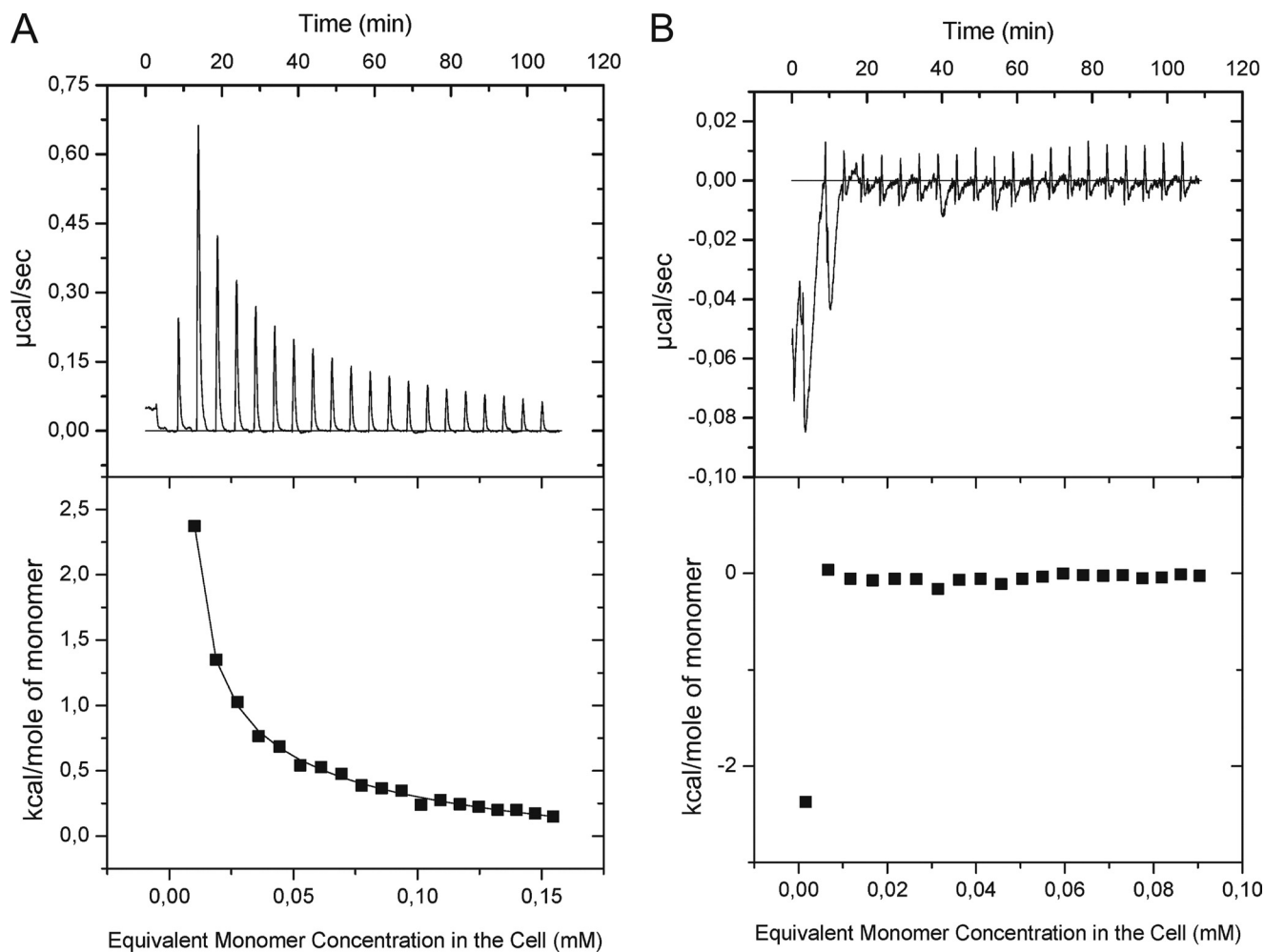


FIGURE 2. **Representative dilution ITC curves for human insulin (*A*) and [*N*MePheB25]-insulin (*B*).**

shown in supplemental Fig. S2. The curves can be fit in terms of a monomer-dimer equilibrium model (Ins$_2$ ↔ 2Ins; $\Delta H^o_d$; $K_d$ = [Ins$_{tot}$]$^2$/[Ins$_2$], where [Ins$_{tot}$] represents the total free insulin monomer concentration) yielding the apparent equilibrium constant of insulin dissociation $K_d$ and enthalpy of dissociation $\Delta H^o_d$ (per mol of dimer). The values of the thermodynamic parameters $\Delta G^o_d$ and $\Delta S^o_d$ were calculated from standard expressions: $\Delta G^o_d = -RT\ln K_d$ and $\Delta G^o_d = \Delta H^o_d - T\Delta S^o_d$, respectively.

The titration experiments with human insulin resulted in an apparent dissociation constant $K_d$ of ~8.8 μM. This value is in a very close agreement with $K_d$ of ~9 μM, which we determined for porcine insulin (Table 2) and with the previous determinations of the $K_d$ under similar conditions (32). Human and porcine insulin represent molecules with native and full dimerization potency. In contrast, the dilutions of DOI did not yield any heat effect, indicating that this (B23–B30)-truncated insulin analogue does not form dimers within the range of concentrations tested here.

All investigated analogues showed considerably reduced dimerization capabilities. The *N*-methylation of the B26 position resulted in an analogue ([*N*MeTyrB26]-insulin) with $K_d$ of

~142 $\mu$M, which was >15-fold higher than the $K_d$ of native insulin. Moreover, the combination of *N*-methylation and substitution of Phe for Tyr at the B26 position ([*N*MePheB26]-insulin) suppresses the dimerization capacity even further to a $K_d$ of ~1.24 mM. The *N*-methylation of B24 amide nitrogen yielded the analogue ([*N*MePheB24]-insulin) with a low dimerization capability ($K_d$ of ~587 $\mu$M, which is 65-fold higher than the $K_d$ of native insulin). However, the most drastic effect of the *N*-methylation on insulin dimerization was observed in the case of the modification at the B25 position. The [*N*MePheB25]-insulin analogue, similarly to DOI, gave no heat effect and behaved solely as a monomer, even at millimolar concentrations (Fig. 2*B*).

*Structural Characterization of Analogues*—The analogues [*N*MePheB24]-insulin and [*N*MePheB25]-insulin were subjected to intensive crystallization screening under a plethora of monomeric, dimeric, and hexameric conditions. Only $R_6$ hexameric crystals of [*N*MePheB25]-insulin were obtained. The monomeric and hexameric ($R_6$) crystals of [*N*MeTyrB26]-insulin and hexameric ($R_6$) of [TyrB25,*N*MePheB26]-insulin, also discussed in this work, were already reported by us (Refs. 17 and 27, respectively). However, their oligomeric organization was not described there in detail, as these studies were focused on the monomeric form of the hormone. In addition, the [TyrB25,*N*MePheB26,LysB28,ProB29]-insulin analogue, not investigated here by the ITC due to its Lys ↔ Pro sequence swap and, thus, likely monomeric properties, was also crystallized in the hexameric ($R_6$) form; we reported the synthesis and binding affinity (4%) of this analogue elsewhere (30).

*Crystal Structure of [NMePheB25]-Insulin*—This analogue yielded crystals only under hexameric conditions in the presence of $Zn^{2+}$ and phenol (see supplemental Table S2). The analogue oligomer adopted a typical $R_6$ conformation with six phenol ligands at the so-called type I site (located at the dimer interface, with hydrogen bonds to COA6 and NHA11, loose van der Waals contacts (~3.7 Å) with imidazole of His-B5 (40)) and two $Zn^{2+}$ cations and two $Cl^-$ anions as their axial ligands. The overall structure of this hexamer followed general fold and most of the structural details of a typical $R_6$ hexamer such as that of the insulin structure of Protein Data Bank code 1ZNJ.

In the [*N*MePheB25]-insulin $R_6$ hexamer, the *N*-methylation of the NHB25 did not result in significant disruption of the dimer interface hydrogen bonds motifs (Fig. 3*A*). All four symmetrical B24–B26 hydrogen bonds were preserved there despite a small weakening of the B24CO-NHD26 interaction by ~0.2 Å. The conservation of these interactions is likely due to an "outward" orientation of the *N*MeB25 groups, which point out away from the dimer interface (Fig. 3*A*) and are thus easily accommodated in the vicinity of the A19–A21 region. However, the introduction of two *N*MeB25 groups resulted in much greater mobility of the Phe-B25 side chains; only one B25 phenyl moiety (with poor definition in the electron density maps) could be located. Both C-terminal threonines B30 are not visible in the structure.

*Crystal Structure of [NMeTyrB26]-Insulin Hexamer*—Although we already reported this structure elsewhere (17), the nature of the dimer interface and its $R_6$ hexamer structure was not discussed in that report. Remarkably, it can be crystallized
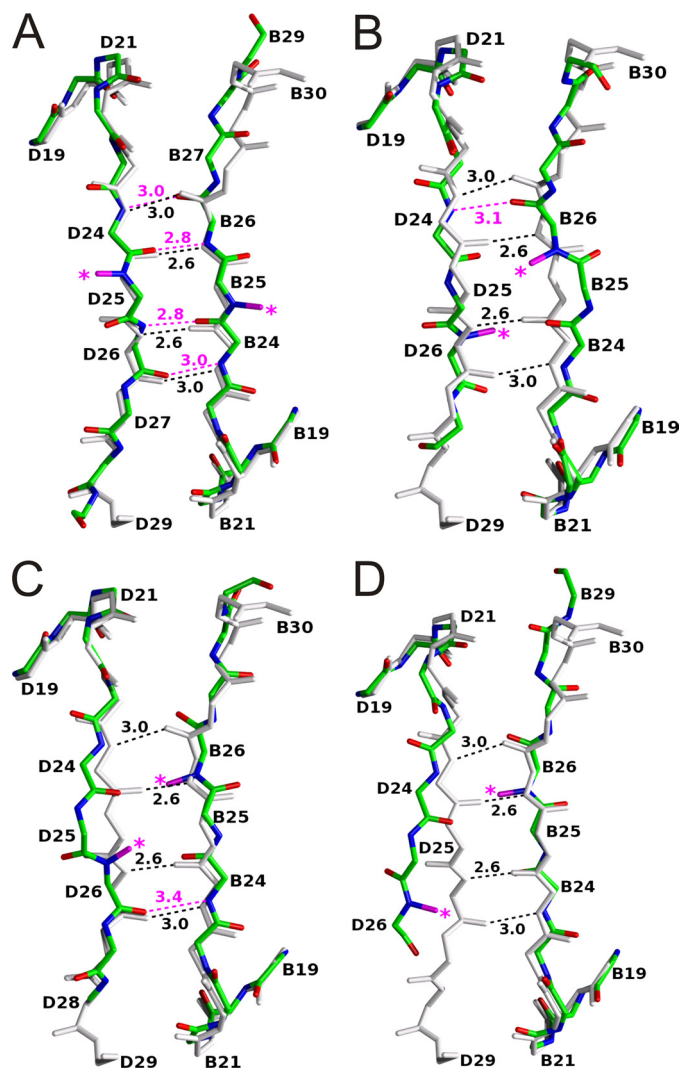


FIGURE 3. **Main chain representation of the dimer (chains AB in molecule I and chains CD in molecule II) interface in [*N*MePheB25]-insulin (Protein Data Bank code 3ZQR) (*A*), [*N*MeTyrB26]-insulin (Protein Data Bank code 2WS6) (*B*), [TyrB25,*N*MePheB26]-insulin (Protein Data Bank code 1W8P) (*C*), and [TyrB25,*N*MePheB26,LysB28,ProB29]-insulin (Protein Data Bank code 3ZS2) (*D*) analogues.** The reference wild-type insulin (Protein Data Bank code 1ZNJ) is in white with its dimer hydrogen bonds as *black dashed lines*; hydrogen bonds of the analogues are colored *magenta*. The methyl groups at *N*-methylated atoms are indicted by *magenta asterisks*; hydrogen bond distances are in Å.

both as a monomer and $R_6$ hexamer under appropriate conditions. The hexamer crystallization protocol, similar to conditions for [*N*MePheB25]-insulin (see supplemental Table S2), yielded crystals of the $R_6$ oligomer very similar to that of [*N*MePheB25]-insulin. However, the methylation of NHB26 resulted in a much more disturbed dimer interface than in the [*N*MePheB25]-insulin structure. As both B26 *N*-methyl groups pointed into the dimer interface, its quartet of the hydrogen bonds was almost lost; only one of them, COB26-NHD24, remained in place (Fig. 3*B*). The disturbance created by *N*-MeB26 groups had a non-symmetrical character. Only one β-strand of the two-strand dimer β-sheet bulged out between B24–B26 to accommodate the more bulky substitutions of the NHB26 hydrogen atoms; the other D21–D27 β-strand followed the typical fold of the wild-type insulin. However, it still exerted

some signs of induced structural stress, reflected in its higher mobility, with B28–B30 residues being fully disordered. Interestingly, the β-strand bulge in the B-chain brought the B25NH peptide unit closer to the COA19 group, resulting in the unusual, for the $R_6$ hexamer, B25NH-COA19 hydrogen (2.9 Å) bond.

*Crystal Structure of [TyrB25,NMePheB26]-Insulin*—The structure of our already reported (27) analogue with NHB26 N-methylation and the Tyr ↔ Phe B25–B26 swap was also analyzed here in more detail. The main focus of the previous discussion of this structure was the impact of the IGF-like swap of B25–B26 side chains, combined with simultaneous modification of the main chain, on insulin and IGF-I structure-function relationships. Here, we tried to extract the relevant structural features in the context of the dimer interface. [TyrB25, NMePheB26]-insulin formed an $R_6$ hexamer that is quite similar to [NMeTyrB26]-insulin and [NMePheB25]-insulin oligomers. Although the bulging of the B24–B26 chain was "symmetrical" to that observed in the [NMeTyrB26]-insulin (Fig. 3C), its nature and structural meaning was practically identical to the [NMeTyrB26]-insulin change, considering the 32 symmetry of the hexamer and the relativity of its chain nomenclature.

*Crystal Structure of [TyrB25,NMePheB26,LysB28,ProB29]-Insulin*—The crystal structure of [TyrB25,NMePheB26,LysB28, ProB29]-insulin was obtained under hexameric conditions (supplemental Table S2) in the pursuit of a very extensive characterization of the dimer interface hydrogen bond network of the B25–B26 side chains for the stability of this region. The $R_6$ hexamer was also very similar to all other $R_6$ oligomers described here. However, the B28–B29 Lys ↔ Pro and B25–B26 Tyr ↔ Phe swaps, combined with N-MeB26, fully disrupted the dimer interface, with none of its hydrogen bonds being preserved (Fig. 3D). Hence, the remarkable stability of these $R_1R_1$ dimers depended entirely on the phenol-induced and R-form specific inter- and intramolecular interactions (see "Discussion"). The disruption of the dimer interface was also very asymmetrical, in a fashion similar to that observed in [NMeTyrB26]-insulin; one β-strand followed the wild-type conformation, whereas its counterpart systematically departed from the dimer interface from D23 to become fully disordered from D27 onwards.

## DISCUSSION

*Impact of Modifications on Binding Affinity of Analogues*—Analogues N-methylated at the B24 position ([NMePheB24]-insulin, 3%) or B25 ([NMePheB25]-insulin, 0.17%) showed drastically reduced binding potency. To our knowledge, [NMePheB24]-insulin represents the first insulin analogue with a backbone modification at the B23–B24 peptide bond.

The low binding activity of [NMePheB25]-insulin is in agreement with findings of Wollmer et al. (41) for an almost inactive (3–4%) insulin analogue ([depsi(B24–B25)]-insulin) in which the B24–B25 CO-NH peptide bond was replaced by a CO-O ester bond. It seems that the very low binding affinity of [NMePheB24]-insulin and [NMePheB25]-insulin results from the lack of the respective peptide bond amides that may be crucial for the direct interaction of insulin with IR. These obser-

vations are also supported by findings of Nakagawa et al. (42) concerning inactive insulin analogues in which peptide bonds at the B24–B25 or B25–B26 positions were replaced by reduced $CH_2$-NH groups. They attributed the extremely low activity of their analogues to the loss of respective carbonyl oxygens involved in H-bonding with IR. Another possible reason for the low binding affinity of [NMePheB24]-insulin and [NMePheB25]-insulin, N-methylation-induced conformational changes in the structures of their B24–B30 strands, will be discussed below.

Interestingly, the [NMeTyrB26]-insulin (17) also displayed lower binding potency toward IR (21%, Table 1) than human insulin, but significantly less reduced than that of analogues N-methylated at the B24 and B25 positions (3 and 0.17%, respectively). This may indicate that the NH of B26 is less important for IR binding than its counterparts at the B24 and B25 positions and/or has a different function. Our recent study (17) demonstrated that N-methylation of Tyr-B26 induces a new type II of β-turn at positions B24–B26 (B26 turn), which results in the departure of B-chain B22–B30 β-strand from its typical hexamer/dimer conformation. The B26 turn is stabilized by a typical $i+3$ β-turn hydrogen bond between CO of Phe-B24 and NH of Thr-B27. These large conformational changes unmask previously buried amino acids Gly-A1, Ile-A2, and Val-A3, which are implicated in IR binding. We proposed (17) that similar conformational changes may occur in the native insulin molecule upon binding to IR. Hence, it is possible that the reduced binding affinity of [NMeTyrB26]-insulin (21%) may be caused by a non-optimal arrangement of residues B27–B30. However, the reduced affinity of this analogue may also result from the loss of NHB26 hydrogen bonding capabilities, required in IR interactions, or by a combination of these factors. The impact of some distinct (e.g. long range) N-methylation-induced conformational changes in the structures of B24–B30 strands in [NMePheB24]-insulin and [NMePheB25]-insulin cannot be excluded as well. However, their existence still awaits structural proof as [NMePheB24]-insulin eluded all crystallization trials so far, and [NMePheB25]-insulin has been crystallized only in the form of a typical $R_6$ hexamer (see below).

The last three analogues in Table 1 represent molecules with single, double, or multiple modifications compared with native human insulin, respectively. The [PheB26]-insulin analogue was included in this study to examine the importance of the Tyr-B26 hydroxyl group for insulin affinity, as the next analogue studied here, [NMePheB26]-insulin, featured two combined modifications: the loss of phenolic character of the B26 side chain with N-methylation of the B26NH. The [PheB26]-insulin showed 46% binding affinity that was in full agreement with data (47%) already reported for this analogue (39), clearly indicating that the loss of Tyr-B26 hydroxyl group is responsible for the reduced binding affinity of this analogue. Therefore, an even more drastic (actually to 4%) loss of affinity of the [NMePheB26]-insulin was expected, as it results from a double (side and main chain) modification of the hormone at this position. Subsequently, the practically total loss of affinity (0.44%) of the [NMePheB25,NMePheB26]-insulin analogue is not surprising, as it fosters the cumulative effect of three unfavorable

modifications (two *N*-methylations and loss of a phenolic B26 side chain).

*ITC Measurements and Dimerization Capabilities of Analogues*—ITC measurements of insulin dimer dissociation require relatively high amounts of protein material. Their optimum required concentration should be in the 0.2–1 mM range, but concentrations closer to the high end yield more reliable data (32). Hence, a typical ITC experiment in a 1.4-ml reaction vessel requires ~8 mg of insulin, which presents some experimental challenges due to the relatively low yields of semisynthetic reactions (30). Also, the recovery of analogues from reaction mixtures after ITC experiments is not very efficient. Therefore, we determined the dimerization capabilities of analogues only at two replicates at different protein concentrations. However, the complete ITC data presented in supplemental Table S1 indicate that duplicate measurements performed here produced very consistent and reliable results.

As some reports (43, 44) have shown, pH has a rather small effect on the outcomes of insulin-dimerization experiments; thus, all ITC experiments were performed here also at acidic pH 2.5 to avoid reduced insulin solubility at the physiological pH (32). As expected, our ITC experiments with human and porcine insulin gave almost identical dissociation constants of dimerization ($K_d$) of ~9 $\mu$M, and the changes in Gibbs free energies of dissociation ($\Delta G^o{}_d$) of ~29 kJ/mol (Table 2). These results are close to the $K_d$ and $\Delta G^o{}_d$ of ~12 $\mu$M and 28 kJ/mol, respectively, measured by Lovatt *et al.* (32) for bovine insulin under similar ITC conditions. Strazza *et al.* (43) used concentration difference spectroscopy to determine the monomer-dimer association constants of porcine insulin, finding only minor differences in $K_d$ at pH 2 (15 $\mu$M) and at pH 7 (18 $\mu$M). Equilibrium sedimentation, an alternative assessment of insulin dimerization, provided $K_d$ of ~4.5 $\mu$M for bovine insulin (at pH 8.0) (28), whereas this approach for porcine insulin gave $K_d$ of ~7 $\mu$M (at pH 7.0) (44). Altogether, the dimerization dissociation constants for human and porcine insulin determined in this study fall well into the range of $K_d$ values determined experimentally by other groups and different techniques. The reliability of the approach presented here therefore allows a more systematic insight into the evaluation of the roles of individual hydrogen bonds involving B24–B26 peptide units in dimer stability. They are discussed below in the context of structural changes caused by their alternations.

*Structure-Function Relationships in Insulin Analogue Dimers*—The resultant minor differences in equilibrium constants and thermodynamic parameters of insulin dimerization at different pHs led to the conclusion that insulin association is mainly driven by the nonpolar interactions (44). However, Strazza *et al.* (43) also indicated that insulin dimerization is under enthalpic control and that the formation of four hydrogen bonds in the apolar protein environment is the main driving force in the assembly of insulin dimer. However, molecular dynamics simulations suggested that insulin dimerization results mainly from nonpolar interactions (contributed mostly by residues B24–B26) (45) and that the role of hydrogen bonds between monomers is to provide a correct directional and spatial guidance to prevailing non-polar interactions. Our experimental data do not exclude any of these hypotheses but indicate more
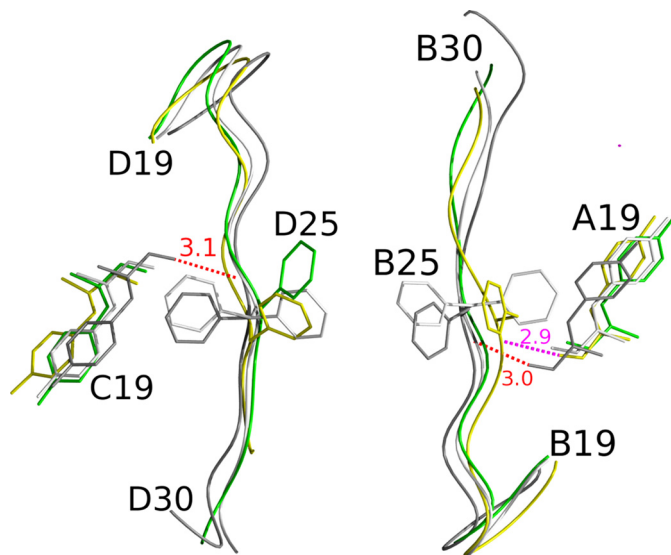


FIGURE 4. **The B-A (*right panel*) and C-D (*left panel*) insulin intrachain coupling effect of the B25NH-COA19 hydrogen bond in the wild-type insulin T$_6$ dimer (Protein Data Bank code 1MSO) (in *dark gray*), wild-type R$_6$ dimer (Protein Data Bank code 1ZNJ) (*white*), [*N*MePheB25]-insulin (Protein Data Bank code 3ZQR) (*green*), and [*N*MeTyrB26]-insulin (Protein Data Bank code 2WS6) (*yellow*).** The T state characteristic B25NH-COA19 hydrogen bond in the T$_6$ dimer is in *red* on both panels (distances in Å); this AB monomer-only hydrogen bond in the [*N*MeTyrB26]-insulin analogue is in *magenta*. Phe-B25 side chains are disordered in [*N*MePheB25]-insulin and are not represented here.

clearly the important individual contributions of B24–B26 hydrogen bonds toward the stability of the dimer.

The individual knock-outs of the NHB24 ([*N*MePheB24]-insulin) and NHB26 ([*N*MeTyrB26]-insulin) hydrogen bond donors increased $K_d$ from 9 $\mu$M to 587 and 142 $\mu$M, respectively (Table 2). As the role of the B24–B26 side chains can, to some extent, be decoupled in this process due to the 2-fold symmetry of these regions, the B24NH–COB26 hydrogen bond can be seen as a more important polar stabilizer of the insulin dimer. Its loss causes a much larger (~4-fold) dimer disruptive effect than abolishing the B26NH-COB24 interaction. This can result from its flanking positions on the dimer interface, closer to the solvent exposed termini of the dimer $\beta$-sheet (Fig. 3, *A* and *B*). The lack of B24NH-COB26 interactions, linked with the mobility-prone character of the B27–B30 terminal residues (observed in many insulin crystal and NMR structures; see, for example, Fig. 3, *B–D*), can lead to easier solvent penetration through the dimer interface and to faster unzipping of the $\beta$-sheet via solvation/breaking of the remaining central pair of B26NH–COB24 hydrogen bonds (46). The lower $K_d$ of the [*N*MeTyrB26]-insulin analogue, its higher dimeric stability than the [*N*MePheB24]-insulin, can also result from the nontypical (for the R-state) B25NH–COA19 hydrogen bond in one of its monomers. It seems that it is a side effect of the $\beta$-strand bulge that brings NHB25 into the proximity of the Tyr-A19 main chain, allowing the formation of an intramolecular hydrogen bond (2.9 Å, Fig. 4), which may be important for insulin dimerization (see "Discussion"). However, as this interaction is not symmetrical ($\beta$-strand bulge occurs only in one monomer), it cannot fully compensate for the instability of the interface in this analogue. Therefore, it is possible that the lack of positive crystallization results of the [*N*MePheB24]-insulin analogue

(significant but rather indirect evidence) may result from the rapid monomer-dimer kinetics that prevents a stable crystal nucleation process. Moreover, it is interesting that the higher stability of the [*N*MeTyrB26]-insulin dimer is also reflected in its remarkable availability to form monomeric and hexameric crystal structures (17); appropriate physicochemical solution parameters can shift the monomer-oligomer equilibrium toward one of the stable (crystallisable) forms of this analogue, especially if the most stable (>8 h (47)) phenol-induced R-form of the hexamer is assured. Although crystallization trials cannot be considered as an unambiguous experimental proof, the ~15-fold higher $K_d$ of [*N*MeTyrB26]-insulin compared with native insulin (Table 2) also agrees with the lack of its dimer crystals, even under extensive dimer crystallization conditions (17).

Interestingly, the most dramatic/disruptive impact on insulin dimerization properties, evident in the ITC data, is not provided by the knock-out of each of the intermonomer β-sheet hydrogen bonds pairs, but by the *N*-methylation of the B25 amide unit in the [*N*MePheB25]-insulin analogue. Despite the almost wild-type nature of the dimer interface here (Fig. 3*A*) and an easy spatial accommodation of the outward (from the dimer β-sheet interface) pointing B25 *N*-Me groups, the lack of intramolecular hydrogen bond donor properties of B25NH group results in no-heat effects in the ITC measurements of this analogue, in a fashion similar to the monomeric-benchmark DOI insulin. The only direct hydrogen bond-related impact of the loss of the B25NH proton is the lack of the B25NH-COA19 hydrogen bond that connects B-chain β-strand with the core of the molecule in the T-forms (dominant insulin state in the metal/ligand-free ITC experiments) of the hormone (Fig. 4). The release of the C-terminal β-strand residues from this intramolecular hydrogen bond lock may result in its higher mobility shifting solution equilibrium toward the monomeric state. Additionally, increased dynamics of this region can contribute to, or be amplified further by, the Phe-B25 side chain mobility. Indeed, the B25 phenyl ring can be located with difficulty (poor electron density map definition) only for one of the B25 residues. Therefore, it is possible that the monomeric character of [*N*MePheB25]-insulin can be associated with the detachment of Phe-B25 (and surrounding B-chain) from the A-chain.

The β-strand detachment-driven monomeric behavior in the solution and formation of hexameric crystals of [*N*MePheB25]-insulin are somehow reflected in the crystal structure (and solution properties) of the [LysB28,ProB29]-insulin, an analogue also with a very substantially diminished (~300×) dimerization capacity (48). Its monomeric behavior did not also prohibit its crystallization as the $T_3R_3^f$ hexamer in the presence of zinc and phenol (16). The [*N*MePheB26]-insulin was an additional step in probing of the resistance of the dimer interface by prevention of the "central" B26NH-COB24 hydrogen bond of the dimer interface, combined with the loss of B26 phenolic character. It is surprising that, although the "peripheral" dimer-spanning B24NH-COD26 hydrogen bonds are preserved here (like in the [*N*MeTyrB26]-insulin analogue), the loss of the central pair (B26NH-COD24) of these interactions, combined with a simultaneous loss of a Tyr-B26 OH group, cannot prevent the further dramatic increase of the $K_d$ into the mM range (1.240 mM). It seems then that the phenolic

indirect hydrogen bonds of the Tyr-B26 side chain (*e.g.* to Tyr-D16) and its multiple van der Waals interactions (*e.g.* Phe-B24, Pro-B28, Ile-A2, Val-A3, Ile-B11, and Val-B12), multiple combined contributions, are rather important for monomer (and dimer) stability. This also indicates that the dimer-stabilizing role of B24–B26 hydrogen bonds cannot be unambiguously decoupled from the individual contributions of the associated individual side chains.

*Role of Hexameric Form in Dimer Stability*—It has to be stressed that all dimer-related comparisons and correlations of structural information with the solution (ITC) studies presented here are based on different quaternary forms of insulin. Structural data were derived exclusively from the $R_6$ hexamers (two $Zn^{2+}$, six phenol ligands), and none of the analogues discussed in this report attained the conformation of $T_3R_3$ ($T_3R_3^f$) or $T_6$ hexamers. In contrast, the ITC solution measurements were performed in a metal/ligand-free environment, with the T state being the likely form of the hormones.

Considering the dimer β-sheet disruptive character of the chemical changes in analogues studied here, it is not surprising that the most stable form of the hexamer ($R_6$) (47, 49, 50) was found to be adopted by all modified insulin. The dominance of the $R_6$ hexameric form likely originates from the additional reinforcement of dimer stability provided the B1–B6 helical extension. It not only delivers typical phenol site I-generated interactions (40) but also a few new, strong (~2.7 Å) intermolecular dimer-spanning protein-protein hydrogen bonds such as Tyr-B16NH-COHis-D5. It seems that these interactions must also compensate for a loss of the B25NH-COA19 intramolecular hydrogen bond, which occurs only in the T state of insulin. Hence, the changes in dimer interfaces in various analogues are displayed here against virtually the same reference-like core of the $R_6$ oligomer.

It is interesting that even very disruptive effects on the dimer interface (*i*) resulting from the interference with the B24-B26 intermolecular hydrogen bonds, (*ii*) amplified further by some mutations in this region, and (*iii*) leading, for example, to a full detachment of dimer β-strands (*i.e.* [TyrB25,*N*MePheB26,LysB28,ProB29]-insulin, Fig. 3*D*), are still compensated and accommodated in a stable $R_6$ hexamer. Energy-driven superiority of the $R_6$ oligomer is especially visible in the already reported (17) crystal structure of the truncated [ProB26]-DTI analogue, in which the dimer interface is totally disrupted beyond B/D21–22 residues (Fig. 5). The maintenance of the hexameric form there, despite the lack of the entire dimer β-strand interface, is somehow compatible with the association of DOI into a $R_6$ hexamer observed in the presence of $Zn^{2+}$ and cyclohexanol (29). Although the $T_6 \rightarrow T_3R_3 \rightarrow R_6$ dynamic transitions are rather well described by the SMB model (51–56), the conformational events on the monomer → dimer → hexamer pathway are much less understood. Our results presented here provide further evidence that the nature of the changes on the insulin dimer interfaces (and associated other parts of the insulin molecule) is quite asymmetrical (Fig. 3, *B–D*); thus, they fit and support well the occurrence of structural asymmetries that are important features of a SMB model for the allosteric behavior of insulin (57).
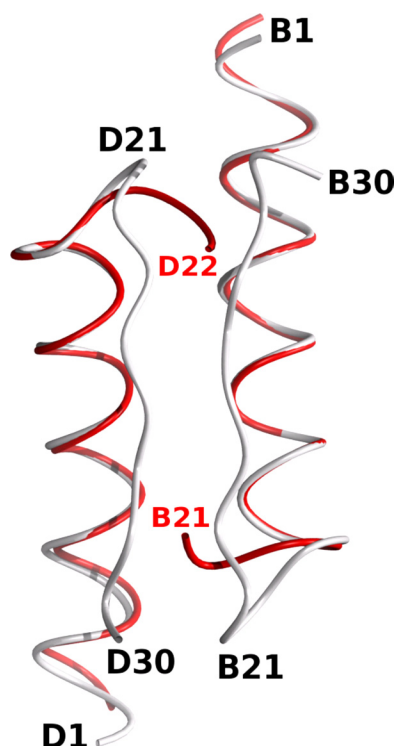
FIGURE 5. **General fold of the B-chain, including the dimer interface, in the [ProB26]-DTI dimer interface-deprived analogue (Protein Data Bank code 2WS7) (in *red*) and in the wild-type insulin (Protein Data Bank code 1ZNJ) (in *white*).**

Undoubtedly, the formation and dissociation of insulin hexamers and dimers into active monomers plays an essential role in the endogenous delivery of the hormone (20, 22). First, the release of insulin hexamers from secretory granules (pH 5.5) into the bloodstream (pH 7.4) results in the loss of zinc (due to the higher pH) and in the dissociation of hexamers to insulin dimers (21). The process of dimer dissociation is likely concentration-driven ($K_d$ of insulin dimerization is $\sim$9 $\mu$M). Therefore, the dissociation of insulin dimers into monomers in the bloodstream is a dilution-driven phenomenon. The clinically used and fast acting [LysB28,ProB29]-insulin (Humalog®) (58) or [AspB28]-insulin (Novolog®) (59, 60) are not able to form dimers, but still associate into hexamers (16), providing evidence that the disruption of insulin dimerization has a fundamental impact on the endogenous action of this hormone.

## CONCLUSIONS

The C terminus of the B-chain, namely, residues B24–B26 and B28–B29, is a crucial part of insulin for its interaction with IR and for its self-association into dimers. The importance of Phe-B24, Phe-B25, and Tyr-B26 residues in the formation of dimers is well established; however, the contribution of their respective individual amide hydrogens for the stability of this aggregates has not been characterized. Here, we delineated their roles by preparation of a series of insulin analogues with selective *N*-methylations of peptide bond amides at positions B24, B25, or B26. All of these *N*-methylated insulin analogues showed significantly impaired binding affinities to the receptor, confirming also the role of respective amide hydrogens in the direct interaction/involvement with insulin receptor. System-

atic, individual *N*-methylations of B24, B25, and B26 peptide bonds in studied analogue amides resulted in a wide spectrum of reduced dimerization abilities of the analogues. Surprisingly, the most dramatic impact on insulin dimerization capability was exerted by the *N*-methylation at the B25 position, likely due to a loss of the intramolecular $\beta$-strand-hormone-core connecting NHB25-COA19 hydrogen bond. The subsequent higher mobility of the $\beta$-strand in this analogue severely shifts the solution equilibrium toward the monomeric state. All examined analogues were crystallized only as $R_6$ hexamers, which is the most stable form of insulin oligomers and the only oligomeric form capable to accommodate/withstand disrupted insulin $\beta$-strands. This study helped to understand the importance and contribution of the B24NH-COB26, B26NH-COB24, and, especially, B25NH-COA19 hydrogen bonds in the formation and stabilization of insulin dimers. The behavioral, functional, and structural correlations presented here provide much evidence that clearly supports and further validates the SMB model of insulin allostery. A better understanding of the structure-function relationships in the association-dissociation pathways of this hormone is important for the understanding of its physiology but may also facilitate development of better, more controlled insulin for clinical applications.

## REFERENCES

1. Derewenda, U., Derewenda, Z., Dodson, E. J., Dodson, G. G., Bing, X., and Markussen, J. (1991) *J. Mol. Biol.* **220,** 425–433
2. Weiss, M. A. (2009) in *Insulin and IGFs* (Litwack, G., Ed.) pp. 33–49, Elsevier Academic Press, Inc., San Diego, CA
3. Mayer, J. P., Zhang, F., and DiMarchi, R. D. (2007) *Biopolymers* **88,** 687–713
4. Hua, Q. X., Shoelson, S. E., Kochoyan, M., and Weiss, M. A. (1991) *Nature* **354,** 238–241
5. Ludvigsen, S., Olsen, H. B., and Kaarsholm, N. C. (1998) *J. Mol. Biol.* **279,** 1–7
6. Xu, B., Huang, K., Chu, Y. C., Hu, S. Q., Nakagawa, S., Wang, S., Wang, R. Y., Whittaker, J., Katsoyannis, P. G., and Weiss, M. A. (2009) *J. Biol. Chem.* **284,** 14597–14608
7. Kaarsholm, N. C., and Ludvigsen, S. (1995) *Receptor* **5,** 1–8
8. Keller, D., Clausen, R., Josefsen, K., and Led, J. J. (2001) *Biochemistry* **40,** 10732–10740
9. Ludvigsen, S., Roy, M., Thøgersen, H., and Kaarsholm, N. C. (1994) *Biochemistry* **33,** 7998–8006
10. Olsen, H. B., Ludvigsen, S., and Kaarsholm, N. C. (1996) *Biochemistry* **35,** 8836–8845
11. Hua, Q. X., and Weiss, M. A. (1991) *Biochemistry* **30,** 5505–5515
12. Huang, K., Xu, B., Hu, S. Q., Chu, Y. C., Hua, Q. X., Qu, Y., Li, B., Wang, S., Wang, R. Y., Nakagawa, S. H., Theede, A. M., Whittaker, J., De Meyts, P., Katsoyannis, P. G., and Weiss, M. A. (2004) *J. Mol. Biol.* **341,** 529–550
13. Hua, Q. X., Xu, B., Huang, K., Hu, S. Q., Nakagawa, S., Jia, W., Wang, S., Whittaker, J., Katsoyannis, P. G., and Weiss, M. A. (2009) *J. Biol. Chem.* **284,** 14586–14596
14. Kristensen, C., Kjeldsen, T., Wiberg, F. C., Schäffer, L., Hach, M., Havelund, S., Bass, J., Steiner, D. F., and Andersen, A. S. (1997) *J. Biol. Chem.* **272,** 12978–12983
15. Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. M., Hubbard, R. E., Isaacs, N. W., Reynolds, C. D., Sakabe, K., Sakabe, N., and Vijayan, N. M. (1988) *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **319,** 369–456
16. Ciszak, E., Beals, J. M., Frank, B. H., Baker, J. C., Carter, N. D., and Smith, G. D. (1995) *Structure* **3,** 615–622
17. Jirácek, J., Záková, L., Antolíková, E., Watson, C. J., Turkenburg, J. P., Dodson, G. G., and Brzozowski, A. M. (2010) *Proc. Natl. Acad. Sci. U.S.A.* **107,** 1966–1970

18. De Meyts, P., Van Obberghen, E., and Roth, J. (1978) *Nature* **273,** 504–509
19. Keefer, L. M., Piron, M. A., De Meyts, P., Gattner, H. G., Diaconescu, C., Saunders, D., and Brandenburg, D. (1981) *Biochem. Biophys. Res. Commun.* **100,** 1229–1236
20. DeFelippis, M. R., Chance, R. E., and Frank, B. H. (2001) *Crit. Rev. Ther. Drug Carrier Syst.* **18,** 201–264
21. Dodson, G., and Steiner, D. (1998) *Curr. Opin. Struct. Biol.* **8,** 189–194
22. Brange, J., Owens, D. R., Kang, S., and Vølund, A. (1990) *Diabetes Care* **13,** 923–954
23. Adams, M. J., Blundell, T. L., Dodson, E. J., Dodson, G. G., Vijayan, M., Baker, E. N., Harding, M. M., Hodgkin, D. C., Rimmer, B., and Sheat, S. (1969) *Nature* **224,** 491–495
24. Smith, G. D., Pangborn, W. A., and Blessing, R. H. (2003) *Acta Crystallogr. D* **59,** 474–482
25. Záková, L., Barth, T., Jirácek, J., Barthová, J., and Zórad, S. (2004) *Biochemistry* **43,** 2323–2331
26. Záková, L., Kazdová, L., Hanclová, I., Protivínská, E., Sanda, M., Budesínský, M., and Jirácek, J. (2008) *Biochemistry* **47,** 5858–5868
27. Záková, L., Brynda, J., Au-Alvarez, O., Dodson, E. J., Dodson, G. G., Whittingham, J. L., and Brzozowski, A. M. (2004) *Biochemistry* **43,** 16293–16300
28. Goldman, J., and Carpenter, F. H. (1974) *Biochemistry* **13,** 4566–4574
29. Pittman, I., 4th, and Tager, H. S. (1995) *Biochemistry* **34,** 10578–10590
30. Záková, L., Zyka, D., Jezek, J., Hanclová, I., Sanda, M., Brzozowski, A. M., and Jirácek, J. (2007) *J. Pept. Sci.* **13,** 334–341
31. Bromer, W. W., and Chance, R. E. (1967) *Biochim. Biophys. Acta* **133,** 219–223
32. Lovatt, A., Cooper, A., and Camilleri, P. (1996) *Eur. Biophys. J.* **24,** 354–357
33. McPhail, D., and Cooper, A. (1997) *J. Chem. Soc. Faraday Trans.* **93,** 2283–2289
34. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
35. Bailey, S. (1994) *Acta Crystallogr. D* **50,** 760–763
36. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D* **60,** 2126–2132
37. Vagin, A., and Teplyakov, A. (1997) *J. Appl. Crystallogr.* **30,** 1022–1025
38. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D* **53,** 240–255
39. Gauguin, L., Klaproth, B., Sajid, W., Andersen, A. S., McNeil, K. A., Forbes, B. E., and De Meyts, P. (2008) *J. Biol. Chem.* **283,** 2604–2613
40. Whittingham, J. L., Edwards, D. J., Antson, A. A., Clarkson, J. M., and Dodson, G. G. (1998) *Biochemistry* **37,** 11516–11523
41. Wollmer, A., Gilge, G., Brandenburg, D., and Gattner, H. G. (1994) *Biol. Chem. Hoppe-Seyler* **375,** 219–222
42. Nakagawa, S. H., Johansen, N. L., Madsen, K., Schwartz, T. W., and Tager, H. S. (1993) *Int. J. Pept. Protein Res.* **42,** 578–584
43. Strazza, S., Hunter, R., Walker, E., and Darnall, D. W. (1985) *Arch. Biochem. Biophys.* **238,** 30–42
44. Pekar, A. H., and Frank, B. H. (1972) *Biochemistry* **11,** 4013–4016
45. Zoete, V., Meuwly, M., and Karplus, M. (2004) *J. Mol. Biol.* **342,** 913–929
46. Ganim, Z., Jones, K. C., and Tokmakoff, A. (2010) *Phys. Chem. Chem. Phys.* **12,** 3579–3588
47. Hassiepen, U., Federwisch, M., Mülders, T., and Wollmer, A. (1999) *Biophys. J.* **77,** 1638–1654
48. Brems, D. N., Alter, L. A., Beckage, M. J., Chance, R. E., DiMarchi, R. D., Green, L. K., Long, H. B., Pekar, A. H., Shields, J. E., and Frank, B. H. (1992) *Protein Eng.* **5,** 527–533
49. Rahuel-Clermont, S., French, C. A., Kaarsholm, N. C., Dunn, M. F., and Chou, C. I. (1997) *Biochemistry* **36,** 5837–5845
50. Huus, K., Havelund, S., Olsen, H. B., Sigurskjold, B. W., van de Weert, M., and Frokjaer, S. (2006) *Biochemistry* **45,** 4014–4024
51. Seydoux, F., Malhotra, O. P., and Bernhard, S. A. (1974) *CRC Crit. Rev. Biochem. Mol. Biol.* **2,** 227–257
52. Roy, M., Brader, M. L., Lee, R. W., Kaarsholm, N. C., Hansen, J. F., and Dunn, M. F. (1989) *J. Biol. Chem.* **264,** 19081–19085
53. Brader, M. L., Kaarsholm, N. C., Lee, R. W., and Dunn, M. F. (1991) *Biochemistry* **30,** 6636–6645
54. Choi, W. E., Brader, M. L., Aguilar, V., Kaarsholm, N. C., and Dunn, M. F. (1993) *Biochemistry* **32,** 11638–11645
55. Bloom, C. R., Choi, W. E., Brzovic, P. S., Ha, J. J., Huang, S. T., Kaarsholm, N. C., and Dunn, M. F. (1995) *J. Mol. Biol.* **245,** 324–330
56. Dunn, M. F. (2005) *Biometals* **18,** 295–303
57. Brzovic, P. S., Choi, W. E., Borchardt, D., Kaarsholm, N. C., and Dunn, M. F. (1994) *Biochemistry* **23,** 13057–13069
58. DiMarchi, R. D., Chance, R. E., Long, H. B., Shields, J. E., and Slieker, L. J. (1994) *Horm. Res.* **41,** 93–96
59. Hedman, C. A., Lindström, T., and Arnqvist, H. J. (2001) *Diabetes Care* **24,** 1120–1121
60. Homko, C., Deluzio, A., Jimenez, C., Kolaczynski, J. W., and Boden, G. (2003) *Diabetes Care* **26,** 2027–2031

# LETTER

# How insulin engages its primary binding site on the insulin receptor

John G. Menting[1]*, Jonathan Whittaker[2]*, Mai B. Margetts[1], Linda J. Whittaker[2], Geoffrey K.‑W. Kong[1], Brian J. Smith[1,3], Christopher J. Watson[4], Lenka Žáková[5], Emília Kletvíková[5], Jiří Jiráček[5], Shu Jin Chan[6], Donald F. Steiner[6], Guy G. Dodson[4]‡, Andrzej M. Brzozowski[4], Michael A. Weiss[2], Colin W. Ward[1] & Michael C. Lawrence[1,7]

**Insulin receptor signalling has a central role in mammalian biology, regulating cellular metabolism, growth, division, differentiation and survival[1,2]. Insulin resistance contributes to the pathogenesis of type 2 diabetes mellitus and the onset of Alzheimer's disease[3]; aberrant signalling occurs in diverse cancers, exacerbated by cross-talk with the homologous type 1 insulin-like growth factor receptor (IGF1R)[4]. Despite more than three decades of investigation, the three-dimensional structure of the insulin–insulin receptor complex has proved elusive, confounded by the complexity of producing the receptor protein. Here we present the first view, to our knowledge, of the interaction of insulin with its primary binding site on the insulin receptor, on the basis of four crystal structures of insulin bound to truncated insulin receptor constructs. The direct interaction of insulin with the first leucine-rich-repeat domain (L1) of insulin receptor is seen to be sparse, the hormone instead engaging the insulin receptor carboxy-terminal α-chain (αCT) segment, which is itself remodelled on the face of L1 upon insulin binding. Contact between insulin and L1 is restricted to insulin B-chain residues. The αCT segment displaces the B-chain C-terminal β-strand away from the hormone core, revealing the mechanism of a long-proposed conformational switch in insulin upon receptor engagement. This mode of hormone–receptor recognition is novel within the broader family of receptor tyrosine kinases[5]. We support these findings by photo-crosslinking data that place the suggested interactions into the context of the holoreceptor and by isothermal titration calorimetry data that dissect the hormone–insulin receptor interface. Together, our findings provide an explanation for a wealth of biochemical data from the insulin receptor and IGF1R systems relevant to the design of therapeutic insulin analogues.**

Insulin comprises two chains (A and B) containing three α-helices (residues A1–A8, A12–A18 and B9–B19) constrained by one intra- and two interchain disulphide bonds[6] (Fig. 1a). The insulin receptor is a disulphide-linked $(\alpha\beta)_2$ homodimer; the extracellular portion of each αβ protomer contains six domains (L1, CR, L2, FnIII-1, FnIII-2 and FnIII-3) and an insert domain (ID) within FnIII-2[7] (Fig. 1b). The α-chain component of the ID is terminated by a segment termed αCT, spanning residues 704–719 (using the numbering of the insulin receptor exon 11− isoform[8]). Two surfaces of insulin are understood to interact with the insulin receptor[9,10]. The first consists predominantly of hormone-dimerizing residues and contacts the primary binding site on the receptor (site 1; dissociation constant ($K_d$) ~6.4 nM) comprising the αCT segment from one insulin receptor α-chain and the central β-sheet of L1 (L1–$\beta_2$) of the other α-chain within the insulin receptor dimer[11–14] (Supplementary Fig. 1a). The second consists predominantly of hormone-hexamerizing residues and is proposed to interact with a secondary insulin receptor site (site 2; $K_d$ ~400 nM) at the junction of FnIII-1 and FnIII-2 of the insulin

receptor α-chain opposite to that contributing L1 to site 1 (Supplementary Fig. 1a)[7,9–11,13–15]. Structural study of the insulin receptor and IGF1R ectodomains is difficult; their large size, flexible multidomain morphology and extensive glycosylation and disulphide bonding posing challenges in production and crystallization. Only three ectodomain structures exist: those of the *apo*-insulin receptor[7,12] and the L1–CR–L2 fragments of insulin receptor[16] and IGF1R[17] (3.7 Å, 2.3 Å and 2.6 Å resolution, respectively); none is in complex with ligand.

We present here the first view of the insulin–site 1 interaction on the basis of four crystal structures (complexes A, B, C and D) varyingly using the L1–CR construct IR310.T (Fig. 1c) and the L1–CR–L2–(FnIII-1)–αCT(704–719) homodimeric construct IR593.αCT (Fig. 1d), which has αCT(704–719) linked to FnIII-1. Complex A (3.9 Å resolution) contains human-insulin-bound IR310.T in complex with a Fab fragment from monoclonal antibody 83-7 (ref. 7) and exogenous αCT(704–719) peptide. Complex B (4.3 Å resolution) is the same as complex A but uses the high-affinity, truncated insulin analogue [D-Pro$^{B26}$]-DTI-NH$_2$[18]. Complex C (4.4 Å resolution) is the same as complex B but uses the longer peptide αCT(697–719). Complex D (4.4 Å resolution) contains bovine insulin bound to IR593.αCT in complex with a Fab fragment from monoclonal antibody 83-14 (ref. 7).

The limited diffraction of these crystals necessitated the use of the following: (1) weak data at the diffraction limit (a process validated by the half-data set Pearson correlation coefficient $CC_{1/2}$ criterion[19,20] and ultimately by the evident quality (see below) of the resultant electron-density difference maps (EDDMs)); (2) B-factor-sharpened EDDMs in model building[21]; and (3) dynamic 'reference-structure' and 'jelly-body' restraints during individual atomic-coordinate refinement whereby we exploited the known higher-resolution structures of the underlying fragments[22,23]. Molecular replacement located the insulin receptor and Fab domains and generated a clear solution for the insulin three-helix 'core' structure within each complex, positioning the core identically on the respective L1–$\beta_2$ surfaces. The strongest feature in the resultant EDDMs was a common, helical-like structure lying adjacent to the insulin B-helix on the L1–$\beta_2$ surface that, upon B-factor sharpening, exhibited side-chain density of distinctive sizes and at spacing consistent with an α-helix. In complex D, this feature was connected to the C-terminal strand of FnIII-1, allowing its unambiguous interpretation as αCT and restricting its prima facie sequence register to an ambiguity of at most one residue unit, with one alternative strongly favoured. We confirmed our register assignment objectively by devising a broad-window analysis of sequence fit both to the higher-resolution EDDM of complex A and to the profiled residue environments within that complex. Crystallographic processing and refinement statistics are provided in Table 1 and Supplementary Table 1, with further detail in Methods.

[1]Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3052, Australia. [2]Department of Biochemistry, School of Medicine, Case Western Reserve University, Cleveland, Ohio 44106, USA. [3]Department of Chemistry, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria 3086, Australia. [4]York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, UK. [5]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v.v.i., 16610 Prague, Czech Republic. [6]Department of Medicine, University of Chicago, Chicago, Illinois 60637, USA. [7]Department of Medical Biology, University of Melbourne, Royal Parade, Parkville, Victoria 3010, Australia.
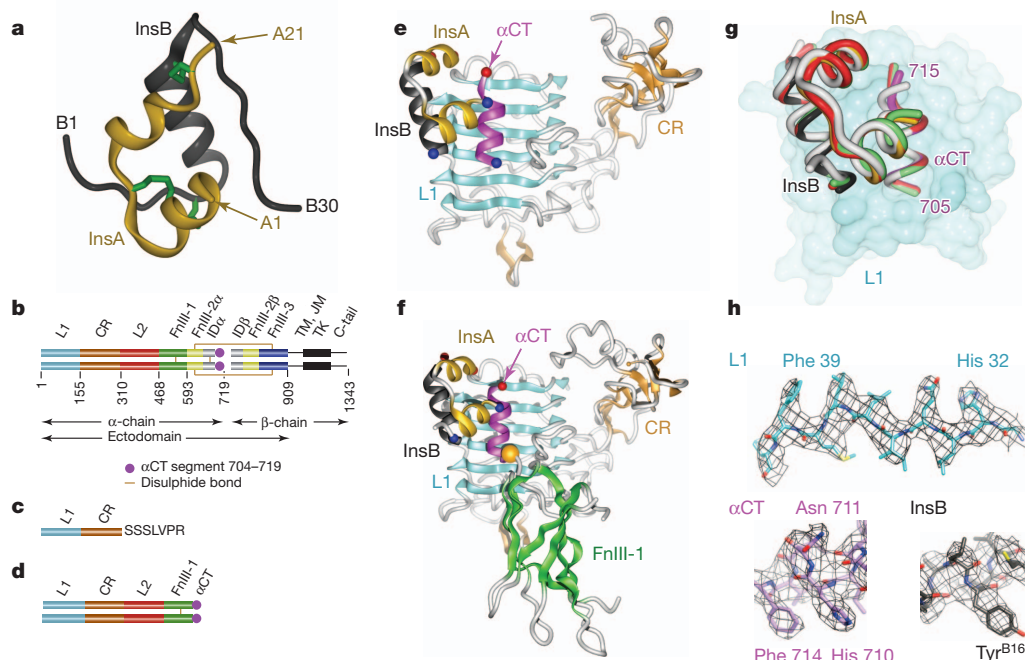*These authors contributed equally to this work.
‡Deceased.

**Figure 1 | Structure of insulin, insulin receptor and the site 1 complexes.**
**a**, Insulin. InsA, A chain; InsB, B chain. **b**, Insulin receptor. CR, Cys-rich domain; FnIII-1, FnIII-2, FnIII-3, first, second and third fibronectin type III domains; ID, insert domain; L1, L2, first and second leucine-rich-repeat domains; TK, tyrosine kinase; TM, JM, transmembrane and juxtamembrane segments. **c**, **d**, IR310.T and IR593.αCT domain structure, respectively.

**e**, **f**, Insulin-bound site 1 in complexes A and D, respectively. Blue, red spheres, observed chain termini; orange sphere, FnIII-1–αCT junction. **g**, Overlay of insulin-bound site 1 in complexes A (coloured as in **e**), B (red), C (green) and D (white). **h**, Sample $2F_{obs} - F_{calc}$ map volumes ($B_{sharp} = -160$ Å$^2$; contours = 1.1–1.5σ) for complex A.

The structures of the insulin-bound site 1 assembly within complexes A and D are shown in Fig. 1e, f, respectively; those within complexes B and C are effectively isomorphous with complex A (Fig. 1g). The Fab-complexed forms of the IR310.T and IR593.αCT monomers are shown in Supplementary Fig. 2a, b, respectively. As prepared and crystallized, complex D contains a dimer of IR593.αCT homodimers crosslinked by four insulins; further description of this assembly is provided in Supplementary Fig. 2c–e. Despite the non-native connection of αCT to FnIII-1 in IR593.αCT (Fig. 1d), the insulin-bound site 1 assembly within complex D is also superimposable on that of complex A (Fig. 1g). Refinement of these complexes provided EDDMs that revealed side-chain bulk for all components within the observed site 1 interface (Fig. 1h). The mode of insulin–site 1 interaction emerging from these structures is as follows. The insulin B-helix (B7–B21) engages the C-terminal end of the L1–β₂ strands; the insulin A-chain has no interaction with L1 (Fig. 1e, f). Both chains interact

extensively with αCT. Notably, the αCT helix is, with respect to its *apo*-insulin receptor counterpart, both displaced on the L1–β₂ surface and C-terminally extended to include residues 711–715 (see Fig. 2a and further discussion below). Residues B22–B30 are not resolved in these structures, but, critically, the αCT helix occupies volume that would otherwise contain insulin residues B26–B30 if the latter retained their receptor-free conformation (see Fig. 2b and Supplementary Fig. 3a). Displacement of residues B26–B30 from the insulin core confirms a long-speculated[24] induced fit upon insulin receptor binding (Supplementary Fig. 1b). In contrast, the helical core of insulin seems, as far as can be discerned at the described resolution, closely similar to that of uncomplexed hormone (Supplementary Fig. 3b). The models suggest that the two most critical hormone-engaging residues in αCT are (1) His 710, which inserts into a pocket formed by invariant[10] insulin residues Val$^{A3}$, Gly$^{B8}$, Ser$^{B9}$ and Val$^{B12}$; and (2) Phe 714, which occupies a hydrophobic crevice formed by invariant insulin residues

**Table 1 | X-ray diffraction data collection statistics**

| | Complex A, data set 1 | Complex A, data set 2 | Complex B | Complex C | Complex D |
|---|---|---|---|---|---|
| | Human insulin + IR310.T–83-7 + αCT(704–719) | Human insulin + IR310.T–83-7 + αCT(704–719) | [D-Pro$^{B26}$]-DTI-NH₂ + IR310.T–83-7 + αCT(704–719) | [D-Pro$^{B26}$]-DTI-NH₂ + IR310.T–83-7 + αCT(697–719) | Bovine insulin + IR593.αCT/83-14 |
| Space group | P23 | P23 | P23 | P23 | I2 |
| Cell dimensions | | | | | |
| $a, b, c$ (Å) | 168.15, 168.15, 168.15 | 168.91, 168.91, 168.91 | 169.49, 169.49, 169.49 | 169.23, 169.23, 169.23 | 118.15, 140.10, 190.02 |
| $\alpha, \beta, \gamma$ (°) | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 | 90, 95.04, 90 |
| Completeness (%) | 99.8 (99.8)* | 98.3 (99.5) | 99.7 (100.0) | 98.5 (98.1) | 87.7 (84.9)† |
| Resolution (Å) | 59.5–4.0 (4.1–4.0) | 46.8–3.9 (4.0–3.9) | 29.6–4.3 (4.53–4.3) | 29.5–4.30 (4.54–4.3) | 56.3–4.4 (4.5–4.4) |
| $R_{merge}$ | 0.156 (2.36) | 0.089 (1.57) | 0.188 (2.69) | 0.101 (1.09) | 0.206 (1.31) |
| ‡$R_{pim}$ | 0.039 (0.57) | 0.048 (0.83) | 0.067 (0.97) | 0.053 (0.59) | 0.143 (0.89) |
| $I/\sigma I$ | 11.5 (1.6) | 9.0 (1.0) | 7.8 (1.0) | 8.2 (1.4) | 2.8 (0.8) |
| Redundancy | 17.3 (17.4) | 4.1 (4.2) | 8.8 (8.8) | 4.4 (4.3) | 2.9 (2.8) |
| §CC$_{1/2}$ | 0.998 (0.580) | 0.998 (0.342) | 0.996 (0.330) | 0.997 (0.438) | 0.980 (0.232) |

Each data set was collected from a single crystal.
* Highest-resolution shell statistics are shown in parentheses.
† Data completeness was compromised by radiation damage.
‡ $R_{pim}$, precision indicating merging R-factor[30].
§ CC$_{1/2}$, Pearson correlation coefficient between independently merged halves of the data set[19]. Highest-resolution shell CC$_{1/2}$ values are significant at at least $P = 0.001$.
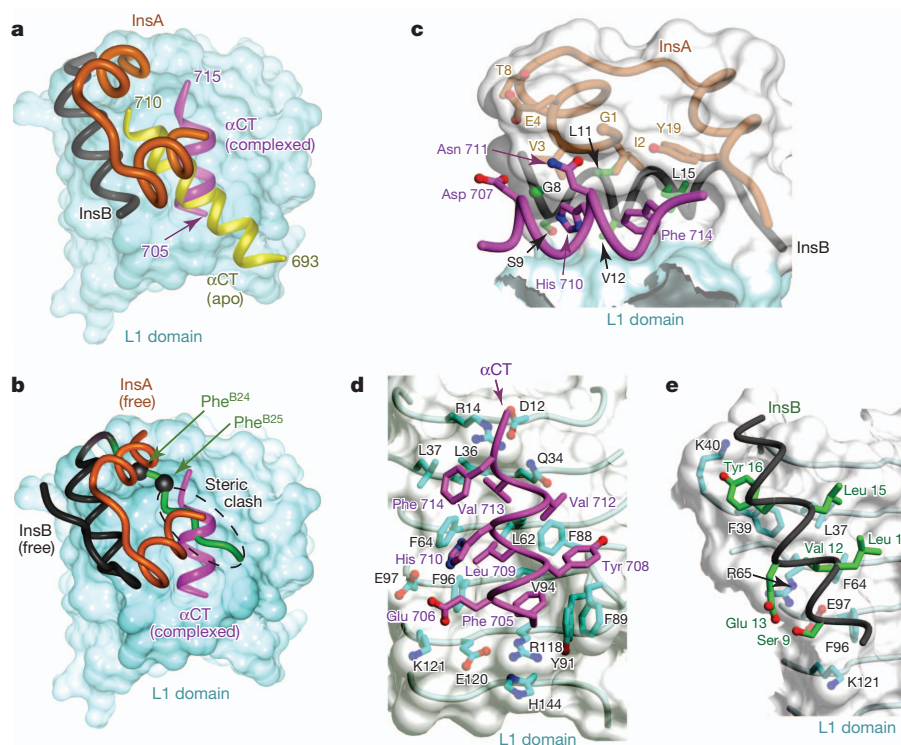
**Figure 2 | The insulin–site 1 interaction. a**, Altered disposition of αCT with respect to that in *apo*-insulin receptor[12]. **b**, Superposition (via the A- and B-chain helices) of receptor-free insulin onto the insulin–site 1 complex, indicating steric clash of the B-chain C-terminal segment (green) with αCT. See also Supplementary Fig. 3a. **c**, Interaction (at the bulk side-chain level of detail) between αCT(704–719) and insulin. View direction is parallel to L1–β2 surface (cyan); white shading shows insulin surface. **d**, **e**, Interaction (at the bulk side-chain level of detail) between L1 and αCT and between L1 and insulin B-chain helix, respectively. Panels based on complex A.

Gly$^{A1}$, Ile$^{A2}$, Tyr$^{A19}$, Leu$^{B11}$, Val$^{B12}$ and Leu$^{B15}$ (Fig. 2c). Asn 711 is directed towards Gly$^{A1}$, Val$^{A3}$ and Glu$^{A4}$. The hydrophobic face of αCT comprising Phe 705, Tyr 708, Leu 709, Val 712 and Val 713 engages a non-polar groove on L1-β$_2$ formed by Leu 36, Leu 37, Leu 62, Phe 64, Phe 88, Phe 89, Val 94 and Phe 96 (Fig. 2d). Direct interaction between the insulin B-chain B9–B19 helix and L1 seems to be limited: Val$^{B12}$ is positioned between Phe 39, Phe 64 and Arg 65, whereas Tyr$^{B16}$ adjoins Phe 39 (Fig. 2e). Detailed rotameric and atomic level detail of the above interactions is precluded by the limited diffraction.

Relocation of αCT upon hormone binding is supported by recent studies indicating that residues buried beneath αCT in the *apo*-ectodomain may directly contact insulin[25], and that peptide substitutions in the αCT segment upstream of Phe 705 that increase the affinity of the segment for L1 concomitantly impair insulin binding[12]. To investigate further, we used complex C to test whether the relocation observed in complexes A, B and D is a consequence of αCT truncation, as in the *apo*-insulin receptor the visible interaction of αCT with L1-β$_2$ begins at Glu 697. It is not—the relocation persists in complex C (Fig. 1g) with no detectable ordering of residues 697–704.

We validated our interpretation of the αCT engagement with L1–β$_2$ and insulin further by isothermal titration calorimetry (ITC). First, Ala-substituted αCT(704–719) peptides were titrated against the L1–CR–L2 construct IR485 (ref. 16). Second, insulin was titrated against IR485 pre-saturated with Ala-substituted αCT(704–719) (Supplementary Tables 2 and 3 and Supplementary Fig. 4). Individual F705A and L709A substitutions significantly impaired αCT binding to IR485, supporting their modelled engagement with L1–β$_2$. In contrast, individual H710A, N711A and F714A substitutions significantly impaired insulin binding, supporting their modelled engagement with insulin. These data are summarized schematically in Fig. 3a. An apparent exception is V713A, which affects insulin but not αCT affinity, despite Val 713 being modelled as engaged with L1–β$_2$. However, this
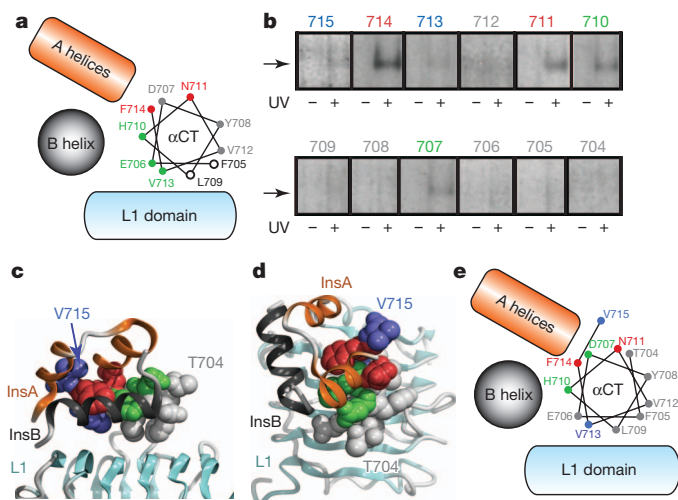


**Figure 3 | Insulin interactions in L1–CR–L2 mini-insulin receptor and *holo*-insulin receptor. a**, Helical wheel representation of ITC-derived insulin affinities for IR485 in the presence of Ala-substituted αCT(704–719) (red, >100× reduction upon Ala substitution; green, >10× reduction; grey, <10× reduction; open circle, not determined). **b**, Reducing gel autoradiograms obtained from *holo*-insulin receptor after photo-crosslinking of αCT helix to bound $^{125}$I-[Tyr$^{A14}$]-insulin. Arrowed band indicates crosslinked insulin receptor α-chain–insulin A chain. Colours indicate crosslinking efficiency (red, strong; green, medium; blue, weak; grey, none). **c**, **d**, Qualitative crosslinking efficiency from **b** mapped onto αCT segment within the site 1 complexes. **e**, Helical wheel representation of crosslinking data presented in **b**–**d**.

observation aligns with the disorder of Val 713 in the *apo*-insulin receptor ectodomain structure[12]; the interaction of Val 713 with L1–$\beta_2$ may thus instead be required to stabilize the critical[11] insulin-engaging residue Phe 714 at the terminus of the $\alpha$CT helix.

The insulin–site 1 interactions described here are further supported by a 'photo-scan' of the *holo*-insulin receptor $\alpha$CT segment, wherein individual $\alpha$CT residues were successively substituted by the photo-activatable residue *para*-azido-Phe (Pap) and then exposed to ultra-violet light to probe covalent interaction with [$^{125}$I]-Tyr$^{A14}$ insulin[25]. Strong crosslinking was observed for photo-probes at sites 711 and 714, medium at sites 707 and 710, and weak at sites 713 and 715; no crosslinking was observed at sites 705, 706, 708, 709 or 712 (Fig. 3b). This photo-crosslinking efficiency pattern in *holo*-insulin receptor exhibits a marked correlation with and supports the insulin–site 1 juxtaposition described above (Fig. 3c–e).

Complexes B and C probed whether the stable $\beta$-turn at B24–B26 in receptor-free [D-Pro$^{B26}$]-DTI-NH$_2$[18] might result in ordering of residues B22–B26 upon binding site 1—it did not, as these residues remained unresolved. We investigated the insulin B-chain C-terminal strand further by seeking to photo-crosslink individual Pap probes at B24 and B25 to the truncated L1–CR–L2 construct IR468[26] in the presence of an $\alpha$CT peptide. In both *holo*- and ectodomain insulin receptor complexes, these probes respectively crosslink to L1 and $\alpha$CT[27,28]. We observed photo-crosslinking for Pap$^{A3}$, Pap$^{A8}$ and Pap$^{B16}$ (as in *holo*-insulin receptor[27]) but not for Pap$^{B24}$ or Pap$^{B25}$ (Supplementary Fig. 5). We thus propose that full stabilization of B22–B30 upon detachment from the hormone core requires the presence of the FnIII domains.

Superposition of the site 1 complexed assembly onto the L1 domain in the *apo*-insulin receptor ectodomain structure (Supplementary Fig. 6) shows that the insulin surface proposed[10] to bind site 2 overlays the FnIII-1–FnIII-2 junction of the alternative $\alpha\beta$ protomer, that is, the proposed site 2 itself. However, significant steric clash is present, indicating that insulin bridging of sites 1 and 2 requires relative displacement of the FnIII-1 and FnIII-2 domains away from the L1–CR module. Such change would alter the relative location of the membrane-proximal FnIII-3 domains, potentially initiating signal transduction[13].

Concordance of our structures with the extensive biochemical literature of this system is presented in Supplementary Discussion. Briefly, the structures are consistent with alanine scanning mutagenesis of the insulin receptor L1 domain, revealing now that of the 'hot-spot' residues, the set Gln 34, Leu 36, Leu 37, Phe 64, Phe 88, Phe 89 and Glu 120 contact the $\alpha$CT segment whereas only Phe 39 contacts insulin. The structural engagement of insulin receptor residues His 710 and Phe 714 with insulin concords with (1) alanine scanning mutagenesis of the insulin receptor ectodomain; (2) hydrogen/deuterium exchange studies of the closely related type 1 insulin-like growth factor receptor (IGF1R); and (3) the lack of hormone binding by insulin receptor chimaeras and IGF1R chimaeras containing the $\alpha$CT segment from the insulin-receptor-related receptor, which has threonine at the residue equivalent to insulin receptor Phe 714 (IGF1R Phe 701). Our structures are also consistent with mutational analysis of insulin, showing now that the hotspot residues Gly$^{A1}$, Ile$^{A2}$, Val$^{A3}$, Tyr$^{A19}$, Gly$^{B8}$ and Leu$^{B11}$ make contact with the key $\alpha$CT residues His 710 and Phe 714 (Fig. 2c). Although residues Gly$^{B23}$, Phe$^{B24}$ and Phe$^{B25}$ are unresolved here, evidence is presented for the likely repositioning of the B-chain C terminus into the volume between the $\alpha$CT segment, the L1–$\beta_2$ surface and the adjacent CR domain; such repositioning rationalizes the enhanced affinity of analogues containing D-amino-acid substitutions at B24.

Our structural and biochemical findings provide groundbreaking insight into the primary insulin–insulin receptor interaction. Particularly important in binding is the $\alpha$CT segment, which is without known counterpart in other mammalian receptor tyrosine kinases[5], although it is conserved among metazoan insulin receptor and IGF1R. The concomitant structural re-arrangements of hormone and receptor in our model rationalize long-standing anomalies in clinical insulin variants associated with diabetes[29] and define new structural targets for novel therapeutics targeting these receptors.

## METHODS SUMMARY

IR310.T was produced by proteolysis of an engineered mini-receptor precursor cIR485 expressed from a stable Chinese hamster ovary Lec8 cell line and purified by ion-exchange and size-exclusion chromatography (SEC), followed by combination with 83-7, re-purification by SEC and finally combination with excess $\alpha$CT peptide and insulin (or analogue). Crystallization conditions for IR310.T-derived complexes A, B and C were 0.9–1.1 M tri-sodium citrate, 0.1 M imidazole-HCl (pH 8.0) + 0.02% sodium azide (NaN$_3$). IR593.$\alpha$CT was expressed from a stable Chinese hamster ovary Lec8 cell line followed by insulin affinity chromatography and then SEC of the resultant bovine insulin complex. Samples were combined with 83-14 and re-purified by SEC. The crystallization condition for the IR593.$\alpha$CT-derived complex D was 9% polyethylene glycol 3350, 200 mM proline, 100 mM HEPES-NaOH (pH 7.5). Diffraction data were collected at the Australian Synchrotron (beamline MX2) and Diamond Light Source (DLS; beamline I24). Structures were solved by molecular replacement. Introduction of Pap residues into the $\alpha$CT segment used orthogonal tRNA/amber suppression technology. Mutant holoreceptors were expressed transiently using co-transfected 293PEAK rapid cells in Pap-containing medium. Cell detergent lysates were then subjected to wheat-germ agglutinin chromatography; eluates containing the receptor were concentrated before photo-crosslinking to $^{125}$I-[Tyr$^{A14}$] insulin. Photo-crosslinked products were resolved by gel electrophoresis and auto-radiographed. IR468 was produced by transient expression in HEK293 cells. $^{125}$I-photo-activatable insulin analogues were incubated with IR468 in the presence of $10^{-5}$ M $\alpha$CT(703–719)–Myc and then photo-crosslinked. Samples were resolved by gel electrophoresis and auto-radiographed. For ITC, Ala-substituted $\alpha$CT peptides (80 μM concentration) were titrated into a sample cell containing IR485 (10 μM concentration); and zinc-free porcine insulin (32–48 μM concentration) was titrated into a sample cell containing IR485 (4–6 μM concentration) pre-incubated with Ala-substituted $\alpha$CT peptide (10× molar concentration). All samples were in Tris-buffered saline (pH 8.0) plus azide (TBSA).

**Full Methods** and any associated references are available in the online version of the paper.

1. Taniguchi, C. M., Emanuelli, B. & Kahn, C. R. Critical nodes in signalling pathways: insights into insulin action. *Nature Rev. Mol. Cell Biol.* **7,** 85–96 (2006).
2. Cohen, P. The twentieth century struggle to decipher insulin signalling. *Nature Rev. Mol. Cell Biol.* **7,** 867–873 (2006).
3. Talbot, K. *et al.* Demonstrated brain insulin resistance in Alzheimer's disease patients is associated with IGF-1 resistance, IRS-1 dysregulation, and cognitive decline. *J. Clin. Invest.* **122,** 1316–1338 (2012).
4. Pollak, M. The insulin and insulin-like growth factor receptor family in neoplasia: an update. *Nature Rev. Cancer* **12,** 159–169 (2012).
5. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141,** 1117–1134 (2010).
6. Adams, M. J. *et al.* Structure of rhombohedral 2 zinc insulin crystals. *Nature* **224,** 491–495 (1969).
7. McKern, N. M. *et al.* Structure of the insulin receptor ectodomain reveals a folded-over conformation. *Nature* **443,** 218–221 (2006).
8. Seino, S. & Bell, G. I. Alternative splicing of human insulin receptor messenger RNA. *Biochem. Biophys. Res. Commun.* **159,** 312–316 (1989).
9. Schäffer, L. A model for insulin binding to the insulin receptor. *Eur. J. Biochem.* **221,** 1127–1132 (1994).
10. De Meyts, P. Insulin and its receptor: structure, function and evolution. *Bioessays* **26,** 1351–1362 (2004).
11. Ward, C. W. & Lawrence, M. C. Ligand-induced activation of the insulin receptor: a multi-step process involving structural changes in both the ligand and the receptor. *Bioessays* **31,** 422–434 (2009).
12. 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. *Proc. Natl Acad. Sci. USA* **107,** 6771–6776 (2010).
13. Ward, C. W. & Lawrence, M. C. Similar but different: ligand-induced activation of the insulin and epidermal growth factor receptor families. *Curr. Opin. Struct. Biol.* **22,** 360–366 (2012).
14. Kiselyov, V. V., Versteyhe, S., Gauguin, L. & De Meyts, P. Harmonic oscillator model of the insulin and IGF1 receptors' allosteric binding and activation. *Mol. Syst. Biol.* **5,** 243 (2009).
15. Whittaker, L., Hao, C., Fu, W. & Whittaker, J. High-affinity insulin binding: insulin interacts with two receptor ligand binding sites. *Biochemistry* **47,** 12900–12909 (2008).

16. 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. *Proc. Natl Acad. Sci. USA* **103,** 12429–12434 (2006).
17. Garrett, T. P. *et al.* Crystal structure of the first three domains of the type-1 insulin-like growth factor receptor. *Nature* **394,** 395–399 (1998).
18. Jiráček, J. *et al.* Implications for the active form of human insulin based on the structural convergence of highly active hormone analogues. *Proc. Natl Acad. Sci. USA* **107,** 1966–1970 (2010).
19. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336,** 1030–1033 (2012).
20. Evans, P. Resolving some old problems in protein crystallography. *Science* **336,** 986–987 (2012).
21. Brünger, A. T., DeLaBarre, B., Davies, J. M. & Weis, W. I. X-ray structure determination at low resolution. *Acta Crystallogr. D* **65,** 128–133 (2009).
22. Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D* **68,** 368–380 (2012).
23. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67,** 355–367 (2011).
24. 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).
25. Whittaker, J. *et al.* α-Helical element at the hormone-binding surface of the insulin receptor functions as a signaling element to activate its tyrosine kinase. *Proc. Natl Acad. Sci. USA* **109,** 11166–11171 (2012).
26. Kristensen, C., Wiberg, F. C. & Andersen, A. S. Specificity of insulin and insulin-like growth factor I receptors investigated using chimeric mini-receptors. Role of C-terminal of receptor α subunit. *J. Biol. Chem.* **274,** 37351–37356 (1999).
27. Xu, B. *et al.* Diabetes-associated mutations in insulin: consecutive residues in the B chain contact distinct domains of the insulin receptor. *Biochemistry* **43,** 8356–8372 (2004).
28. Xu, B. *et al.* Decoding the cryptic active conformation of a protein by synthetic photo-scanning. Insulin inserts a detachable arm between receptor domains. *J. Biol. Chem.* **284,** 14597–14608 (2009).
29. Xu, B. *et al.* Diabetes-associated mutations in insulin identify invariant receptor contacts. *Diabetes* **53,** 1599–1602 (2004).
30. Weiss, M. S. Global indicators of X-ray data quality. *J. Appl. Cryst.* **34,** 130–135 (2001).

**Supplementary Information** is available in the online version of the paper.

**Author Contributions** J.G.M. and J.W. contributed equally to the paper. J.G.M. purified and crystallized samples, collected data and performed the ITC study; J.W. and L.J.W. performed receptor photo-crosslinking experiments; M.B.M. performed molecular biology, cell culture and crystallization experiments; S.J.C. performed insulin photo-crosslinking experiments; G.K.-W.K. and C.J.W. performed crystallography experiments; B.J.S. performed calculations; E.K., L.Z. and J.J. prepared insulin analogues; C.W.W., M.A.W., J.W., D.F.S., S.J.C., J.G.M. and M.C.L. designed the experiments and analysed data. C.W.W., M.A.W., A.M.B., G.G.D. and M.C.L. wrote the paper. All authors discussed the results and commented on the manuscript.

## METHODS

**Cloning, production and purification of IR310.T.** High yield and correct folding of IR310.T was achieved by design of a thrombin-cleavable version of IR485[16] (termed cIR485) for use as precursor, followed by production and purification protocols for cIR485 similar to those described for IR485[16]. A 1,741-nucleotide base fragment of insulin receptor was synthesized (DNA2.0) and inserted into the HindIII–XbaI sites of the vector pEE14 (Lonza). The fragment encoded a protein cIR485 ('cleavable' IR485) which consists, in order, of (1) the 27-residue insulin receptor native signal sequence; (2) residues 1–310 of the insulin receptor α-chain (the L1–CR module); (3) a thrombin cleavage site SSSLVPRGSSS; (4) residues 311–485 of the insulin receptor α-chain (the L2 domain); (5) an enterokinase cleavage site DDDDK; and (6) a c-Myc purification tag EQKLISEEDLN[31] (see Supplementary Fig. 7). A 106-nucleotide-base non-coding fragment 5′-GTCG ACGGTACCCCGGGGAATTAATTCCGGGGGCCGCCTCGGAGCATGACC CCCGCGGGCCAGCGCCGCGCGCTCTGATCCGAGGAGACCCCGCGCTC CCGCAGCC-3′ was included between the HindIII site and the region expressing the signal peptide; bases 29–106 of this fragment correspond to those immediately upstream of the human insulin receptor coding region. CHO Lec8 cells (CRL-1737; ATCC)[32] were transfected with the cIR485-containing expression vector using Lipofectamine 2000 (Life Technologies). Selection was undertaken in the presence of methionine sulphoximine (MSX), with cIR485 production monitored via western blot with monoclonal antibodies 83-7[33] and 9E10[34]. Production scale-up, performed under contract by CSIRO, was in spinner flasks (New Brunswick Scientific) containing DMEM/F12 medium (Life Technologies) supplemented with 10% fetal bovine serum (FBS; Life Technologies) plus 25 μM MSX. The medium was supplemented with 0.8 mM butyric acid (Sigma-Aldrich) during the last week of the production phase. A total of 50 l of culture supernatant was collected; harvests were filtered (0.2 μm cutoff) and then concentrated 10-fold (30 kDa cutoff) after addition of 0.1 mM phenylmethanesulphonylfluoride (PMSF; Sigma-Aldrich) and 0.02% sodium azide (NaN₃). Accumulated supernatant harvests were adjusted to pH 8 by addition of 3 M Tris-HCl (pH 8.5). Initial purification of cIR485 was by 9E10 antibody affinity chromatography and size-exclusion chromatography (SEC) using procedures effectively identical to those described for purification of IR485[16]. Purified cIR485 in Tris-buffered saline (24.8 mM Tris-HCl pH 8.0, 137 mM NaCl, 2.7 mM KCl) plus 0.02% NaN₃ (TBSA) was then incubated overnight at 37 °C with 0.25 units human thrombin (Roche) per mg of cIR485 in the presence of 10 mM CaCl₂. Completion of proteolysis was assessed by SDS–PAGE, which revealed bands corresponding to the estimated molecular mass of the IR310.T fragment and the insulin receptor L2 domain (51 kDa and 30 kDa, respectively; Supplementary Fig. 8a). Western blotting with monoclonal antibody 83-7 confirmed that the CR domain was contained in the upper band alone. The sample was diluted 8-fold in buffer A (10 mM ethanolamine-HCl, pH 9.6 + 0.02% NaN₃), centrifuged for 5 min at 17,000g to remove particulates and then loaded onto a Mono Q 5/50 GL column (GE Healthcare). The sample was eluted with a 60 column-volume gradient of buffer A to buffer B (10 mM ethanolamine-HCl, 400 mM NaCl, pH 9.6 + 0.02% NaN₃; Supplementary Fig. 8b) and the fractions assessed by SDS–PAGE. Fractions containing IR310.T were pooled, concentrated and re-run in TBSA on a Superdex 200 10/300 column (GE Healthcare). The chromatogram exhibited three overlapping peaks, probably arising from multimerization (Supplementary Fig. 8c). SDS–PAGE of fractions revealed the presence of three closely spaced bands, which we attributed to varying glycosylation (Supplementary Fig. 8d). The final yield of IR301.T was ~0.5 mg ml⁻¹ of cell culture.

**Large-scale production and purification of Fab 83-7.** A hybridomal cell line expressing monoclonal antibody 83-7[33] was a gift from K. Siddle. Large-scale production and purification of monoclonal antibody 83-7 and preparation of Fab 83-7 from monoclonal antibody 83-7 were based on protocols described previously[7]. In particular, monoclonal antibody 83-7 was proteolysed with dithiothreitol-activated papain (Sigma-Aldrich) at 37 °C. Proteolysis was stopped by adding iodoacetamide (IAA; Sigma-Aldrich) and the reaction mixture then passed down a Superdex 200 26/60 column (GE Healthcare). Fractions containing 83-7 F(ab')₂ were isolated and reduced with mercaptoethylamine (Sigma-Aldrich) and then alkylated with IAA, followed by further SEC and cation-exchange chromatography on Mono S (GE Healthcare).

**Production of Fab 83-7-complexed IR310.T.** IR310.T was mixed with a slight molar excess of Fab 83-7 and the complex purified by SEC using a Superdex S200 10/300 column (GE Healthcare) in TBSA buffer. Fractions containing the complex of IR310.T/Fab 83-7 were then pooled, concentrated and exchanged into 10 mM HEPES-NaOH (pH 7.5) + 0.02% NaN₃ for crystallization screens.

**Cloning, production and purification of IR593.αCT.** CHO Lec8 cells stably expressing IR593.αCT were obtained from CSIRO. These cells are derived by co-transfection with both the linearized expression plasmid pEE14 (Lonza) and a pEF puromycin plasmid (Springer Labs). The pEE14 expression plasmid

contained, inserted between its HindIII–XbaI sites, a fragment encoding (in order) the 27-amino-acid residue insulin receptor signal sequence, insulin receptor residues 1–593 and insulin receptor residues 704–719 (see Supplementary Fig. 9). The 106-nucleotide-base non-coding fragment described above was also included here between the HindIII site and the start codon. Selection was undertaken in the presence of puromycin, with IR593.αCT production monitored via europium-labelled insulin binding assays[35]. Production scale-up, performed under contract by CSIRO, was in spinner flasks (New Brunswick Scientific) containing DMEM/F12 medium (Life Technologies) supplemented with 10% FBS plus 10 μg ml⁻¹ puromycin. Accumulated supernatant batches were adjusted to pH 8 by addition of 3 M Tris-HCl (pH 8.5). PMSF was added to 0.1 mM concentration and NaN₃ to 0.02%. The culture supernatant was filtered (0.2 μm cutoff), concentrated 10-fold (10 kDa cutoff) and pumped through a Sephacryl S100 column (GE Healthcare) onto an in-house manufactured bis-Boc-insulin MiniLeak medium resin column (Kem-En-Tec)[36]. The latter was eluted with bovine insulin in TBSA. PMSF was again added to 0.1 mM concentration and the eluates concentrated for Superdex S200 26/60 or S200 10/300 (GE Healthcare) SEC. The SEC profiles showed a single major peak corresponding to the insulin-complexed IR593.αCT homodimer and a further peak corresponding to excess bovine insulin (Supplementary Fig. 8g). At higher concentrations, an additional peak appeared corresponding to a dimer of IR593.αCT homodimers (Supplementary Fig. 8h). Contents of fractions were analysed by SDS–PAGE (data not shown). The final yield of IR593.αCT was ~0.4 mg l⁻¹ of cell culture supernatant.

**Large-scale production and purification of Fab 83-14.** A hybridomal cell line expressing Mab 83-14[33] was a gift from K. Siddle. Cells were grown in 950 cm² roller bottles at 37 °C in 5% CO₂ (balanced with air) at 10 r.p.m., using approximately 300 ml H-SFM medium supplemented with 5% fetal calf serum (FCS). Typically, 2–3 l of cell culture at a viable cell density of 2–3 × 10⁶ cells ml⁻¹ was achieved. Cells were then pelleted (5 min, 350g) and re-suspended in serum-free medium. The culture was collected when cell viability dropped below 30% (after 5–7 days). Monoclonal antibody 83-14 from collected supernatant was captured using a ProSep-vA column (Millipore). Monoclonal antibody 83-14 was proteolysed with dithiothreitol-activated papain (Sigma-Aldrich) at 37 °C. Proteolysis was stopped by adding IAA and the reaction mixture containing Fab 83-14 was then passed down a Superdex 200 26/60 column. Final purification of Fab 83-14 was by cation-exchange chromatography on a MonoS column (GE Healthcare).

**Purification of the IR593.αCT + Fab 83-14 + insulin complex.** The IR593.αCT protein was combined with an excess of bovine insulin and Fab 83-14 and then re-subjected to SEC. The SEC profiles (Supplementary Fig. 8i) showed a major peak with a molecular mass corresponding to that of a dimer of Fab-complexed IR593.αCT homodimers and a shoulder of molecular weight corresponding to that of an Fab-complexed IR593.αCT homodimer, as well as peaks corresponding to excess Fab 83-14 and insulin. Fractions corresponding to the higher molecular mass component of the major peak were pooled and then concentrated in a 10 mM HEPES-NaOH (pH 7.5) + 0.02% NaN₃ buffer for crystallization.

**Production of IR468-containing media.** A DNA fragment encoding the signal peptide plus residues 1–468 of human insulin receptor was inserted into the HindIII–XbaI sites of the vector pcDNA3.1zeo(+) (Life Technologies) and then transiently transfected into HEK293 cells. Cells were grown in Dulbecco's modified Eagle's medium (Life Technologies) supplemented with 10% FCS for 72 h after transfection; media containing IR468 were then collected and stored at −20 °C until used.

**Crystallization and diffraction data collection of complexes A to D.** Complexes A, B, C and D are as defined in Table 1. Complexes A, B and C were prepared for crystallization by combining the Fab 83-7 bound IR310.T (final concentration 3.5 mg ml⁻¹) with a 1.5× molar ratio of αCT peptide (Genscript) and 3× molar ratio of human insulin (Sigma-Aldrich) or [D-Pro^B26]-DTI-NH₂ in 10 mM HEPES-NaOH (pH 7.5) + 0.02% sodium azide (NaN₃). [D-Pro^B26]-DTI-NH₂ was prepared as described previously[37]. A single crystallization condition for complex A was detected using a sparse-matrix 792-condition screen (CSIRO Collaborative Crystallization Centre) and optimized manually to 0.9–1.1 M tri-sodium citrate, 0.1 M imidazole-HCl (pH 8.0) + 0.02% NaN₃ for all complexes A, B and C. An initial crystallization condition for complex D at 2.0 mg ml⁻¹ was detected in a similar 792-condition screen and optimized manually to 9% PEG3350, 200 mM proline, 100 mM HEPES-NaOH (pH 7.5). For data collection, the complex A and B crystals were cryo-protected by transfer to paraffin oil HR403 (Hampton Research), the complex C crystal by progressive increase of the precipitant concentration to saturated tri-sodium citrate, and complex D crystal by addition of 20% glycerol to the crystallant solution. All crystals were cryo-cooled by plunging into liquid nitrogen and maintained at ~100 K during data collection. Data set 1 for the complex A crystal and the data set for the complex D crystal were collected at beamline MX2 at the Australian Synchrotron[38] ($\lambda = 0.95369$ Å). The remaining data sets were collected at beamline I24 at the

Diamond Light Source ($\lambda = 0.96860$ Å). Diffraction data were processed with XDS[39] and CCP4[40]. Resolution limits were assessed on the basis of the significance of $CC_{1/2}$ at the $P = 0.001$ level[19]. Data processing statistics are provided in Table 1.

**Structure solution and refinement of complex A.** PHASER[41] located within the asymmetric unit a single copy each of the L1–CR module and the variable module of 83-7 (using data set 1). The 83-7 constant module was not located and ultimately presumed disordered. TLS parameters and individual restrained isotropic B-factors were then refined using autoBUSTER[42], followed by atomic coordinate-only refinement, yielding $R^{xpct}_{work}/R^{xpct}_{free} = 0.368/0.363$ (where $R^{xpct}$ is the crystallographic R-factor reported by autoBUSTER, computed using the expectation value of $F_{calc}$ instead of the value itself)[43]. The difference density maps contained four helix-like features encompassing the first, third, fourth, fifth and seventh highest peaks in the map (the second highest peak being glycan conjugated to Asn 111; Supplementary Fig. 10a). These features lay adjacent to the L1–$\beta_2$ surface, and PHASER then positioned the three insulin helices as a rigid entity onto three of these (TFZ = 11.4). Correctness of this solution was verified by exhaustive six-dimensional real-space search using ESSENS[44] (Supplementary Fig. 10b). The structure was then refined further, yielding $R^{xpct}_{work}/R^{xpct}_{free} = 0.320/0.339$. The remaining helix-like feature in the above difference map (the first and fifth highest peaks) revealed clear side-chain protrusions upon negative B-factor map sharpening[21], spaced consistent with an underlying α-helix (Supplementary Fig. 10c). We concluded that it arose from the αCT peptide as there were no further sources of peptide in the crystallization sample. A ten-residue poly-alanine α-helix was then docked into this feature using ESSENS; the best fit was adequately discriminated from lower-scoring fits and visual inspection confirmed the direction of the helix as judged by protruding side-chain density (Supplementary Fig. 10c). A procedure was then devised to assign, in an objective fashion, the sequence register to this 10-residue segment, based on compatibility of individual residue side chains both with sharpened difference electron density and with their surrounding protein environment. (1) Fit to difference density used a method designed not only to assess the fit of atoms within the density, but also to penalize the existence of volumes of positive difference density into which no atoms had been placed. Briefly, voxels corresponding to the αCT feature were excised from the B-factor-sharpened[21] map using CHIMERA[45] (cutoff level = 0.16 e$^-$ Å$^{-3}$) and placed within a rectangular grid volume large enough to allow a >8 Å buffer zone around the feature. All grid points outside of the feature were set to 0 e$^-$ Å$^{-3}$, with the resultant map being termed $M_{obs}$. Coordinates for residues 1, 2 and 10 of the fitted poly-alanine helix were then deleted from the model, as the density associated with them displayed poorer α-helical geometry. Ten alternatives, corresponding to all possible alignments of sub-sequences of αCT(704–719) with the heptameric poly-alanine structure, were then assessed. Rotamers for each residue within each of these ten models were then assigned manually using COOT[46] based on visual inspection of the fit of the trial rotamer to the density at the corresponding site in $M_{obs}$. The ten individual models were then 'real-space-refined' within COOT to achieve optimal fit to $M_{obs}$, maintaining tight helical restraints. An electron density map $M_{calc}$, on the same grid as $M_{obs}$, was then generated (using SFALL within CCP4) for each heptamer model in isolation, with the B-factors of all main-chain atoms being set to 10 Å$^2$ and of all side-chain atoms set to 20 Å$^2$ to allow for subsequent comparison with a sharpened map. All grid points within $M_{calc}$ with $\rho < 0.45$ e$^-$ Å$^{-3}$ were set to 0 e$^-$ Å$^{-3}$. Correlation coefficients $CC = (\langle xy \rangle - \langle x \rangle \langle y \rangle)/\sqrt{[(\langle x^2 \rangle - \langle x \rangle^2)(\langle y^2 \rangle - \langle y \rangle^2)]}$ between $M_{obs}$ and each $M_{calc}$ were then calculated using MAPMAN[47]. Each CC was then 'normalized' by dividing its value by the CC value calculated for the underlying poly-alanine heptamer, and the quotient termed the trial sequence's 'density score'. The heptamer with trial register 705–711 was seen to have the highest density score (Supplementary Table 4). (2) Compatibility of residue side-chain environment within the L1–αCT–insulin interface was assessed as follows. For each of the ten trial heptamer (real-space refined) models described above, an 'environment score' was generated using VERIFY3D[48] to assess compatibility with the surrounding L1–$\beta_2$ and insulin surfaces. Trial register 705–711 again scored highest (Supplementary Table 4). A 'combined score' was then computed as the product (density score) × (environment score) to assist assessment of lower-ranked trial registers (Supplementary Table 4). The next highest combined score was for trial register 709–715, which is related to trial register 705–711 by a one-turn translation along the helix, effectively maintaining hydrophobic-to-hydrophobic docking with the L1 surface. However, register 709–715 was judged to be most unlikely, as it would bring Pro 716 and Pro 718 into the remaining C-terminal region of the helix. We concluded that register assignment 705–711 was correct, given that it ranked highest on all criteria. The assignment aligned with that assessed directly from the EDDMs for complex D, wherein the αCT segment is directly attached to the C-terminus of FnIII-1 and the ambiguity of assignment strictly limited. Nine sugar residues were also included at N-linked sites 16, 25, 111, 225 and 255[16,49]. Final refinement was against data set 2, using local structural

similarity restraints (LSSRs)[22] against Protein Data Bank (PDB) structures 2HR7-A (L1-CR), 1IL1-B (83-7 light chain) and 1FNS-H (83-7 heavy chain), yielding $R^{xpct}_{work}/R^{xpct}_{free} = 0.264/0.292$ (see Supplementary Table 1). Ramachandran statistics are: favoured region 91%, acceptable region 5% and disallowed region 4%.

**Structure solution and refinement of complexes B and C.** These structures were solved independently. MOLREP[50] was used to locate the L1–CR/83-7 fragment and the structure then refined with REFMAC5[23]. [D-Pro$^{B26}$]-DTI-NH$_2$ (PDB entry 2WRW) was then located using MOLREP, its position being effectively identical to that in the native insulin complex (complex A). The αCT segment from the wild-type insulin complex was fitted in the final electron density maps using COOT. Final refinement gave $R_{work}/R_{free} = 0.289/0.349$ and $R_{work}/R_{free} = 0.289/0.335$, respectively (Supplementary Table 1). Ramachandran statistics for both complexes are: favoured region 89%, acceptable region 7% and disallowed region 4%.

**Structure solution and refinement of complex D.** Molecular replacement with PHASER located sequentially two copies each of a module consisting of Fab 83-14 in complex with the FnIII-1 domain, an L1–CR module and an L2 domain, using search models derived from PDB entry 3LOH[12]. The assembly was subject to rigid-body and TLS refinement using autoBUSTER, followed by individual restrained isotropic B-factor and coordinate refinement, subject both to LSSRs to PDB entries 2HR7-A (L1-CR-L2), 3LOH-E (FnIII-1), 1PLG-H (83-14 heavy chain) and 2VXU-M (83-14 light chain) and to non-crystallographic symmetry (NCS) restraints, yielding $R^{xpct}_{work}/R^{xpct}_{free} = 0.308/0.299$. Difference electron density analysis revealed: (1) two helix-like features of electron density, each connected to a respective insulin receptor residue Thr 593 and lying on the L1–$\beta_2$ surface of an adjacent monomer, thus identifiable as the αCT segment of IR593.αCT; and (2) two clusters of three helix-like features, each cluster lying adjacent to a respective one of the above αCT segments and putatively identifiable as insulin. Insulin was then docked into each of the above putative insulin–αCT clusters of features by means of an exhaustive six-dimensional ESSENS search. The highest scoring solution in each instance overlaid insulin accurately onto the respective three-helix cluster (Supplementary Fig. 10d) and was well discriminated from lower scoring solutions; the pose of insulin and αCT with respect to L1–$\beta_2$ was effectively identical at the two NCS related sites. The two insulins were then included in the model followed by further refinement. COOT was then used to position poly-alanine α-helices into the respective density features corresponding to the αCT segment (sharpened by the application of a negative B-factor), followed by mutation of the sequence to that of αCT residues 704–715. Assignment of the αCT register presented no more than a single residue prima facie ambiguity (depending on the way in which residues at the FnIII-1–αCT peptide junction were built). One alternative was strongly favoured, aligning independently with that determined for complexes A, B and C above. Residues B5–B7 of each insulin could be built into electron density following a path similar to that seen in certain high-affinity insulin analogues[18]. Further crystallographic refinement was then undertaken with autoBUSTER, yielding final $R^{xpct}_{work}/R^{xpct}_{free} = 0.244/0.260$ (Supplementary Table 1). Ramachandran statistics are favoured region 88%, acceptable region 7% and disallowed region 5%.

**Holoreceptor photo-crosslinking.** Production and photo-crosslinking of derivatized holoreceptors were performed as described previously[25].

**Photo-crosslinking of Pap-derivatized insulins to IR468.** Photo-activatable insulin derivatives Pap$^{A3}$-Ins, Pap$^{A8}$-DKP-Ins, Pap$^{B16}$-DKP-Ins, Pap$^{B24}$-DKP-Ins, [Pap-$^{B25}$]-DP-Ins were synthesized as described[28,51] and iodinated using the chloramine-T oxidation method. αCT(703–719)-Myc peptide was synthesized in-house. $^{125}$I-photo-activatable insulin analogues were incubated overnight at 4 °C with IR468 and $10^{-5}$ M αCT(703–719)-Myc in 100 mM HEPES (pH 7.5), 100 mM NaCl, 10 mM MgCl$_2$, 0.05% bovine serum albumin, 0.025% Triton X-100. One aliquot was exposed to short-wavelength ultraviolet light (Mineralight UVG-54) for 30 s at 1 cm and a second aliquot treated with 0.2 mM disuccinimidyl suberate (DSS; Pierce). Samples were diluted with Laemmli sample buffer, resolved on 10% Tricine SDS-PAGE and auto-radiographed.

**Isothermal titration calorimetry (ITC).** ITC analysis of insulin binding to IR310.T and its Fab 83-7 complex followed protocols identical to those described previously[52]. The affinity of zinc-free human insulin for IR310.T in the presence of a 10× molar ratio of αCT(704–719) was determined to be $K_d = 30$ nM (Supplementary Fig. 8e) and the affinity of zinc-free human insulin for IR310.T–Fab 83-7 in the presence of a 10× molar ratio of αCT(704–719) to be $K_d = 48$ nM (Supplementary Fig. 8f). ITC analysis of mutant αCT peptides binding to IR485 and of insulin binding to IR485 in complex with mutant αCT peptide were also performed as described previously[52], with samples prepared in TBSA (25 mM Tris-HCl, 137 mM sodium chloride, 2.7 mM potassium chloride, pH 8.0 + 0.02% NaN$_3$).

31. Hoogenboom, H. R. *et al.* Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res.* **19,** 4133–4137 (1991).

32. Stanley, P. Chinese hamster ovary cell mutants with multiple glycosylation defects for production of glycoproteins with minimal carbohydrate heterogeneity. *Mol. Cell. Biol.* **9,** 377–383 (1989).
33. Soos, M. A. *et al.* Monoclonal antibodies reacting with multiple epitopes on the human insulin receptor. *Biochem. J.* **235,** 199–208 (1986).
34. Evan, G. I., Lewis, G. K., Ramsay, G. & Bishop, J. M. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol. Cell. Biol.* **5,** 3610–3616 (1985).
35. Denley, A. *et al.* Structural determinants for high-affinity binding of insulin-like growth factor II to insulin receptor (IR)-A, the exon 11 minus isoform of the IR. *Mol. Endocrinol.* **18,** 2502–2512 (2004).
36. Markussen, J., Halstrom, J., Wiberg, F. C. & Schäffer, L. Immobilized insulin for high capacity affinity chromatography of insulin receptors. *J. Biol. Chem.* **266,** 18814–18818 (1991).
37. Žáková, L. *et al.* Insulin analogues with modifications at position B26. Divergence of binding affinity and biological activity. *Biochemistry* **47,** 5858–5868 (2008).
38. McPhillips, T. M. *et al.* Blu-ice and the distributed control system: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J. Synchrotron Radiat.* **9,** 401–406 (2002).
39. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* **66,** 133–144 (2010).
40. Collaborative Computing Project, No The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
41. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63,** 32–41 (2007).
42. Bricogne, G. *et al.* BUSTER version 2.10 (Global Phasing Ltd, 2011).
43. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60,** 2210–2221 (2004).
44. Kleywegt, G. J. & Jones, T. A. Template convolution to enhance or detect structural features in macromolecular electron-density maps. *Acta Crystallogr. D* **53,** 179–185 (1997).
45. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comp. Chem.* **25,** 1605–1612 (2004).
46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
47. Jones, T. A. & Thirup, S. Using known substructures in protein model building and crystallography. *EMBO J.* **5,** 819–822 (1986).
48. Lüthy, R., Bowie, J. U. & Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **356,** 83–85 (1992).
49. Sparrow, L. G. *et al.* N-linked glycans of the human insulin receptor and their distribution over the crystal structure. *Proteins Struct. Funct. Bioinform.* **71,** 426–439 (2008).
50. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66,** 22–25 (2010).
51. Hua, Q. X. *et al.* Enhancing the activity of a protein by stereospecific unfolding. The conformational life cycle of insulin and its evolutionary origins. *J. Biol. Chem.* **248,** 14586–14596 (2009).
52. Menting, J. G., Ward, C. W., Margetts, M. B. & Lawrence, M. C. A thermodynamic study of ligand binding to the first three domains of the human insulin receptor: relationship between the receptor α-chain C-terminal peptide and the site 1 insulin mimetic peptides. *Biochemistry* **48,** 5492–5500 (2009).

# Structural Integrity of the B24 Site in Human Insulin Is Important for Hormone Functionality*[S]

**Lenka Žáková**[‡], **Emília Kletvíková**[‡], **Václav Veverka**[‡], **Martin Lepšík**[‡], **Christopher J. Watson**[§], **Johan P. Turkenburg**[§], **Jiří Jiráček**[‡1], and **Andrzej M. Brzozowski**[§2]

*From the [‡]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v.v.i., Flemingovo nám. 2, 166 10 Prague 6, Czech Republic and the [§]York Structural Biology Laboratory, Department of Chemistry, The University of York, Heslington, York, YO10 5DD, United Kingdom*

**Background:** The structure of the C-terminal B21-B30 chain of insulin bound to the insulin receptor remains undetermined.
**Results:** The structures of B24-modified insulins were determined.
**Conclusion:** The integrity of Phe[B24] but flexibility of B25-B30 insulin residues are important for receptor binding.
**Significance:** The knowledge of the receptor-bound structure of insulin is important for the design of new insulin receptor agonists.

Despite the recent first insight into the insulin-insulin receptor complex, the structural role of the C terminus of the B-chain of insulin in this assembly remains unresolved. Previous studies have suggested that this part of insulin must rearrange to reveal amino acids crucial for interaction with the receptor. The role of the invariant Phe[B24], one of the key residues of the hormone, in this process remains unclear. For example, the B24 site functionally tolerates substitutions to D-amino acids but not to L-amino acids. Here, we prepared and characterized a series of B24-modified insulin analogues, also determining the structures of [D-His[B24]]-insulin and [His[B24]]-insulin. The inactive [His[B24]]-insulin molecule is remarkably rigid due to a tight accommodation of the L-His side chain in the B24 binding pocket that results in the stronger tethering of B25-B28 residues to the protein core. In contrast, the highly active [D-His[B24]]-insulin is more flexible, and the reverse chirality of the B24C[α] atom swayed the D-His[B24] side chain into the solvent. Furthermore, the pocket vacated by Phe[B24] is filled by Phe[B25], which mimics the Phe[B24] side and main chains. The B25→B24 downshift results in a subsequent downshift of Tyr[B26] into the B25 site and the departure of B26-B30 residues away from the insulin core. Our data indicate the importance of the aromatic L-amino acid at the B24 site and the structural invariance/integrity of this position for an effective binding of insulin to its receptor. More-

over, they also suggest limited, B25-B30 only, unfolding of the C terminus of the B-chain upon insulin activation.

---

Human insulin is a 51-amino acid hormone that regulates blood glucose levels and affects lipid/protein metabolism and, to some extent, the life span. Insulin exerts its actions through binding as a monomer to the $(\alpha\beta)_2$-homodimeric, tyrosine kinase-type insulin receptor (IR) (1).

Insulin is the final product of an extensive processing from single-chain preproinsulin (2) and consists of an A-chain (Gly[A1]-Asn[A21]) and B-chain (Phe[B1]-Thr[B30]) that are linked by two interchain disulfide bridges (Cys[A7]-Cys[B7] and Cys[A20]-Cys[B19]). In addition, the A chain includes an intrachain Cys[A6]-Cys[A11] disulfide bond. Insulin is capable of dimerization at low micromolar concentrations (3), and the presence of divalent metal ions leads to the formation of insulin hexamers that represent the storage form of this hormone (2). There is overwhelming evidence that monomeric insulin, which dissociates from its storage forms, must undergo structural changes upon binding to IR. Mainly, the C-terminal ~B22-B30 part of the B-chain has to rearrange to reveal amino acids that are crucial for the effective formation of the hormone-receptor interface (Leu[B11], Val[B12], and Leu[B15] on the B chain and Gly[A1], Ile[A2], Val[A3], and Gln[A5] on the A chain) (4, 5) and that are obstructed by the C-terminal $\beta$-strand in the dimeric and hexameric forms of the hormone (6, 7). The importance of the B-chain C-terminal residues in IR binding has also been demonstrated by their critical role in negative cooperativity (8). The changes at the C terminus are likely accompanied by the reorganization of the N terminus of the B-chain (B1-B8), which adopts one of the conformations ranging from the so-called T-state (extended) to the R-state ($\alpha$-helical) folds (9). The current understanding of the three-dimensional insulin structure upon binding to the IR is based mostly on inactive storage (hexamers, dimers) states of the hormone (9–11). Our recent structural and functional study of the insulin-IR complex (5) provided the first insight into the mode of insulin binding to its receptor, although the

---

ZSI

AQ: J

## Structural Integrity of the B24 Site in Insulin

structural features of the B21-B30 residues (and the B-chain N terminus) have not been determined. Nevertheless, the obtained structural data suggest that the ~B25-B30 chain in the IR-complex must acquire a different conformation from that regularly found in a receptor-free state.

The NMR structures of the monomeric form of insulin, facilitated by mutations (12–14) or truncations of the B-chain (15, 16), have confirmed both the conformations known from the inactive forms of the hormone and the intrinsic mobility of the N and C termini of the B-chain. Systematic photo-cross-linking mapping of the photo-activated surface of insulins in complexes with the IR have provided further insight into the role of the individual hormone side chains for stability of the insulin-IR interface (17, 18). These experiments indicated direct contacts of B24/B26 and A1-A8 side chains with the IR, suggesting the detachment model of insulin activation (18). This model postulates a radical departure of the ~B22-B30 chain from the core of the hormone and its insertion between the IR domains.

However, our recent studies on highly active (~200 – 450%) insulin analogues led to the identification of their structural and functional signatures, which suggest that less dramatic rearrangements of the B25-B30 residues may be required for an "active" (*i.e.* as in IR complex) form of the hormone (19). The main feature of these high affinity analogues is a so-called "B26-turn," which is stabilized by a formation of the B24CO-NHB27 hydrogen bond and *trans-cis* isomerization of the B25-B26 peptide bond. Contrary to the detachment model, in which the whole B22-B30 chain is fully separated from the insulin core, the $Phe^{B24}$ side chain in the B26-turn-activation model is invariantly anchored in the hydrophobic cavity (as in the wild-type insulin) lined up by $Val^{B12}$, $Leu^{B15}$, and $Cys^{B19}$, with only B26-B30 residues departing from an "inactive," *i.e.* β-strand-like, conformation.

The invariant $Phe^{B24}$ is one of the key residues in the insulin molecule. Its peculiar role in the "activation process" of the hormone is further supported by lack of tolerance at this site for any L-amino acid substitutions, whose introduction leads usually to a complete or very significant loss of the binding affinity (20). Only $Ser^{B24}$ (7–15%) (21, 22) and $Leu^{B24}$ (~10–50%) (23–29) insulin analogues retained some activity. Interestingly, such intolerance for L-amino acid mutations is even more striking, as incorporation of D-amino acids at the B24 site (except D-Pro (30) and, to a lesser extent, D-Phe(p-ethyl) (31)) results in an increase of insulin affinity to the IR (30). Moreover, the loss of chirality at the B24 $C^\alpha$ atom, resulting from "racemic"-like mutations by an achiral side chain-containing amino acid (α-aminoisobutyric acid), also diminishes insulin activity (30). However, a complete loss of the B24 side chain combined with the lack of B24$C^\alpha$ chirality is relatively well tolerated in the $Gly^{B24}$ analog (~22–78% affinity) (21, 30, 32). Interestingly, whereas D-Phe at the B24 site in full-length insulin enhances the affinity (140–180%) (21, 30), the same substitution in the truncated analog *des*-pentapeptide(B26-B30)-insulin-B25-amide (itself equipotent with human insulin) significantly reduces binding affinity (8%) (33). Similarly, a distinct negative effect (1% binding affinity) was found for $Gly^{B24}$ in *des*-pentapeptide(B26-B30)-insulin-B25-amide (32).

The insulin molecule can be considered the result of complex functional and evolutionary development. The small size of this hormone, the proximity of its key side chains, and the compactness of the long range (allosteric) intramolecular interactions combined with the t complex behavior of this protein (challenging bio-synthesis, processing, storage, activation of the monomer, "induced fit" etc.), gives each of its amino acids a multi-faceted and convoluted function. Therefore, the role of the invariant $Phe^{B24}$ should be considered in a wider context of its molecular contributions, for example, toward conformational switching during hormone activation, facilitation of proper disulfide pairing, and hormone self-assembly (17).

Here, we provide an explanation of the role for the $Phe^{B24}$ site in the context of insulin structure and function by studying the effects of its substitutions by a combination of functional, structural (NMR, x-ray), and computational molecular dynamics (MD) approaches. Systematic substitutions of chiral (L/D) amino acids and structural main chain constraints inducing (Pro and Sar[3]) amino acids resulted in four novel B24-substituted analogues, whose functional and structural relationships have been established. Furthermore, the possible cooperative character of B24 and B26 sites in insulin-IR interactions and their impact on insulin affinity was investigated in three double (D-His$^{B24}$, Glu$^{B26}$) insulin mutants.

### EXPERIMENTAL PROCEDURES

*Materials*—2-Chlorotrityl chloride resin, protected amino acids, and reagents for solid-phase synthesis of peptides were purchased from Novabiochem Merck (Laufelfingen, Switzerland). Fmoc-Lys(Pac)-OH (fluorenylmethoxycarbonyl-Lys-(phenylacetyl)-OH) was prepared as described previously (34). TPCK (L-1-tosylamido-2-phenylethyl chloromethyl ketone)-treated trypsin was purchased from Sigma, and penicillin G acylase was from Fluka. Human [$^{125}$I]monoiodotyrosylA14-insulin was purchased from Perkin Elmer Life Sciences. Human insulin was purchased from Sigma. All other chemicals and solvents were obtained from Sigma.

*Peptide Synthesis and Enzymatic Semisynthesis*—The syntheses of peptides and the semisyntheses of analogues were performed according to Záková *et al.* (34). The identity of the peptides and analogues was confirmed with an FTMS mass spectrometer LTQ-orbitrap XL (Thermo Fisher).

*Receptor Binding Studies*—Receptor binding studies with plasma membranes prepared from epididymal adipose tissue of adult male Wistar rats were performed according to Záková *et al.* (35). Competitive binding curves were plotted using GraphPad (San Diego, CA) Prism 5, comparing the best fits for single-binding site models. The half-maximal inhibition values of binding of $^{125}$I-insulin to the receptor ($IC_{50}$) were obtained from nonlinear regression analysis. The only exception was [ProB24]-insulin, which was tested for binding to the insulin receptor in membranes in IM-9 lymphocytes according to the procedure that we described recently in Morcavallo *et al.* (36). The binding data for [ProB24]-insulin were analyzed with Excel software using a one-site fitting program developed in the lab-

---

[3] The abbreviations used are: Sar, sarcosine; r.m.s.d., root-mean-square deviations.

oratory of Dr. Pierre De Meyts (A. V. Groth and R. M. Shymko, Hagedorn Research Institute, Denmark, a kind gift of Pierre De Meyts) and GraphPad Prism 5, and the dissociation constant was determined. The dissociation constant of human ^{125}I-insulin was set up to 0.3 nM. Receptor binding assays were performed under conditions excluding the depletion of free ligand.

*X-ray Studies*—Crystallization of [HisB24]-insulin was performed using in-house insulin crystallization screens covering most of the reported crystal insulin growth parameters. Crystallization conditions, data collections, and refinement and models statistics are given in supplemental Table S1. X-ray data were processed by HKL2000 (37) and XDS, and model building and refinement were performed using the CCP4 suite of programs (38) and COOT (39). Crystal structures were solved using Molrep (40) with B1-B6- and B23-B30-truncated insulin hexamer or hexamer-derived monomer as a model (PDB code 1mso) (41) and refined by Refmac 5.6 (42). The figures were made using CCP4 mg (38). For structural comparisons of the dimer interfaces, the best defined dimers (referred here as monomers A-B and C-D) from the relevant structures were superimposed on the B9-B19 and D9-D19 $C^\alpha$ atoms using the LSQKAB program from the CCP4 suite.

*NMR Spectroscopy and Structure Calculations*—NMR spectra were acquired from 0.35-ml samples of 0.2–0.5 mM insulin analogues in a 20% $d_3$-acetic acid (pH 1.9) or 25 mM deuterated-Tris buffer (pH 8.0) containing 5% $D_2O$, 95% $H_2O$. All the NMR data were collected at 25 °C on a 600 MHz Bruker Avance spectrometer equipped with a triple-resonance ($^{13}C,^1H,^{15}N$) cryoprobe. A series of homonuclear spectra were recorded to determine the sequence-specific resonance assignments for [HisB24]-insulin and [D-HisB24]-insulin, in particular, two-dimensional TOCSY (two-dimensional total correlation spectroscopy) spectra with a 55-ms mixing time, two-dimensional double quantum filtered COSY spectra, and two-dimensional NOESY spectra, which were acquired with an NOE mixing time of 200 ms.

The family of converged structures for [HisB24]-insulin and [D-HisB24]-insulin was initially calculated using Cyana 2.1 (43). The combined automated NOE assignment and structure determination protocol was used to automatically assign the NOE cross-peaks identified in two-dimensional NOESY spectra and to produce preliminary structures. Subsequently, five cycles of simulated annealing combined with redundant dihedral angle constraints (Redac) (44) were used to produce the sets of converged structures with no significant restraint violations (distance and van der Waals violations <0.2 Å), which were further refined with two cycles of restrained molecular dynamics simulated annealing using AMBER (45). Initial energy minimization (2000 steps) was followed by 15 ps of simulated annealing in a vacuum and 20 ps of simulated annealing using a generalized Born solvent model (46). The 30 structures with the lowest AMBER energy and with no distance constraint violation greater than 0.2 Å were selected. Analysis of the family of structures obtained was carried out using the programs Molmol (47) and PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4, Schrödinger, LLC).

*Molecular Dynamics Simulations*—Two starting structures (A-B and C-D monomers) of the wild-type human insulin for simulation runs were derived from the high resolution 1mso crystal structure (41). The MD simulations of [HisB24]-insulin and [D-HisB24]-insulin analogues were based on pH 8.0 NMR structures of the analogues obtained in this work. Details of the structure preparation before MD simulation runs are shown in the supplemental information. The insulin monomers were solvated in a box of TIP3P water molecules (48) extending 11 Å from the protein. Two $Na^+$ counterions were added to neutralize the charge. The systems were relaxed in several steps before MD (for details, see the supplemental information). The MD was run in periodic boundary conditions using AMBER (45) and comprised three stages: warming, equilibration, and production runs. In the first stage, the temperature was raised gradually from 10 to 300 K over 25 ps. The equilibration stage was run for 200 ps, and a production of 50 ns length ensued at 300 K and 1 atm. Three trajectories for each case were simulated using different starting conditions (random number generator seed for assigning the initial velocities), and only the most structurally stable ones according to the root mean square deviations (r.m.s.d.) were chosen for further analyses. All the bond lengths were constrained using the SHAKE algorithm (49), which allowed a time step of 2 fs. The frames were saved every 5 ps. The analyses of the MD trajectories included measurements of the positional fluctuations of insulin residues, selected interatomic distances, and backbone r.m.s.d. with respect to the starting structure.

## RESULTS

### Binding Affinities of Analogues

Seven new insulin analogues have been prepared by enzymatic semisynthesis. Their binding affinities, presented here in two groups for single-B24 and double-B24/B26-substituted analogues, are shown in Table 1, with the corresponding binding curves provided in supplemental Fig. S1. We consider here the [D-HisB24, GluB26, LysB28, ProB29]-insulin as a "double-mutant" as the $Lys^{B28} \leftrightarrow Pro^{B29}$ swap does not have any impact on binding affinity (50). This swap was introduced into this analog to promote its monomeric behavior (51) for solution NMR studies.

In the group of B24-single substituted analogues, [D-HisB24]-insulin displays the highest binding affinity of ~212% compared with wild-type insulin. In contrast, its L-amino acid analog, [HisB24]-insulin, has an only very low IR affinity of 1.5%. Mutations with an expected impact on conformation of the main chain yielded a [SarB24]-insulin analog with highly reduced affinity (6.3%), whereas [ProB24]-insulin was inactive (0.16%).

In the second group of analogues, two "positive" hormone mutations (D-His$^{B24}$ and Glu$^{B26}$, each individually with higher IR affinity) were combined to elucidate their simultaneous effect on binding affinity. The high biological activity of the [GluB26]-insulin was already reported (52), but the actual binding affinity data were not available. Therefore, the [GluB26, LysB28, ProB29]-insulin was prepared as a reference molecule, and it confirmed the high affinity (193%) of this analog. Surprisingly, the simultaneous combination of the positive mutations of the B24 and B26 sites, D-His$^{B24}$ and Glu$^{B26}$, respectively,

## Structural Integrity of the B24 Site in Insulin

**TABLE 1**

Values of $IC_{50}$, $K_d$, and the relative binding affinities of human insulin and insulin analogues

| Analogue | $IC_{50} \pm$ S.E.[a], ($n$) | $K_d \pm$ S.E., ($n$) | Potency[b] |
|---|---|---|---|
| | $n$M | $n$M | % |
| **Single B24-substituted analogues** | | | |
| Human insulin | 0.89 ± 0.06 (3) | ND[c] | 100 ± 7 |
| Human insulin | ND | 0.39 ± 0.01 (6) | 100 ± 3 |
| [HisB24]-insulin | 59.5 ± 6.2 (3) | ND | 1.5 ± 0.16 |
| [D-HisB24]-insulin | 0.42 ± 0.04 (3) | ND | 212 ± 20 |
| [SarB24]-insulin | 14.1 ± 2.4 (5) | ND | 6.3 ± 1.1 |
| [ProB24]-insulin | ND | 240 ± 7 (5) | 0.16 ± 0.005 |
| **Double B24/B26-substituted analogues** | | | |
| [D-HisB24,GluB26]-insulin | 31.5 ± 2.9 (4) | ND | 2.8 ± 0.26 |
| [D-HisB24,GluB26,LysB28,ProB29]-insulin | 49.8 ± 3.9 (4) | ND | 1.8 ± 0.14 |
| [GluB26,LysB28,ProB29]-insulin | 0.46 ± 0.06 (3) | ND | 193.0 ± 25 |

[a] $IC_{50}$ values represent the concentrations of insulin or the analogues that cause half-maximal inhibition of binding of human [^{125}I]monoiodotyrosylA14-insulin to IR. Each value represents the mean ± S.E. of multiple determinations ($n$).

[b] Relative receptor binding affinity (potency) is defined as ($IC_{50}$ or $K_d$ of human insulin/$IC_{50}$ or $K_d$ of analogue) × 100.

[c] ND, Not determined.

abolished the individual affinity gains, having a detrimental effect on the affinities of both double analogues ([D-HisB24, GluB26]-insulin with 2.8% affinity and [D-HisB24, GluB26, LysB28, ProB29]-insulin with 1.8% binding affinity).

### X-ray Structure of [HisB24]-insulin

This analog gave crystals only under monomeric conditions (pH 3.0, see supplemental Table S1) with one molecule in the asymmetric unit. Despite relatively high resolution data, the monomer of this analog displayed a high degree of disorder (supplemental Fig. S2). Only the Val[B2]-Glu[B21] part of the B-chain was defined in the electron density maps with the Glu[B21] side chain being fully disordered. The position of the A-chain C-terminal Asn[A21] also could not be allocated. The B2-B6 B-chain N terminus adopts a T-R "intermediate" conformation observed in other monomeric x-ray structures of insulin analogues (19) and some of its dimeric forms (53) as well. The total disorder of B22-B30 chain (hence the lack of B25NH-COA19 hydrogen bond) results in the conformation of the A-chain observed in some other monomeric analogues (19, 54) (movement from the wild-type insulin T-state conformation, *i.e.* away from the B24-B30 chain, and ~38° rotation of the A1-A9 helix around A2C[α] axis). In addition, the lack of the B23CO-NHA21 hydrogen bond did not allow clear localization of the A-chain C-terminal Asn[A21], which is also missing in the model. The Tyr[B16] is flipped toward the "empty" B24 site and is stabilized by a close (~3.55 Å) van der Waals contact with the Val[B12] side chain. The remains of the B19-B22 turn (B19-B21 main chain) adopt a conformation not observed in other insulins (close toward Tyr[B16] side chain). However, it may represent one of the few alternative directions of the B20-B30 chain that cannot be traced and refined in this analog.

### NMR Structures of [HisB24]-insulin and [D-HisB24]-insulin Analogues

NMR spectra for both His[B24] and D-His[B24] analogues were acquired at pH 1.9 and pH 8.0 to address the issues related to the $pK_a$ of the insulin side chains and potential disruptive influence of the buffering systems (55).

Both analogues provided good quality NMR spectra under acidic conditions (20% $d_3$-acetic acid, pH 1.9), which is indicative of a monomeric behavior for the modified insulin molecules. The essentially complete sequence-specific assignment of ^{1}H NMR resonances was obtained for [HisB24]-insulin and [D-HisB24]-insulin using a combination of two-dimensional homonuclear experiments. The detailed resonance assignments are shown in supplemental Tables S2 and S3. The chemical shift values for H[ε2] from the side chains of three histidines (at positions B5, B10, and B24) in the acidic pH were indicative of their double-protonated (charged) state (56). In addition, a significant line-broadening of the [1]H[N] NMR signals of several residues (Thr[A8], Ser[A9], Cys[A11], Leu[B6], and Gly[B8]) was observed, suggesting the presence of chemical exchange due to the isomerization of the adjacent disulfide bonds in the NMR time scale. Because of limited solubility and susceptibility to aggregation, it was not possible to obtain reasonable quality NMR data for L- and D-His[B24] analogues in aqueous solution at pH 7.

However, the analogues were fully soluble and predominantly monomeric at pH 8.0 up to a concentration of 0.5 mM. Nevertheless, a significant proportion of the H[N] signals were not present in the NMR spectra due to a rapid exchange of the amide hydrogens with the solvent at basic pH. Consequently, the number of assigned resonances for both analogues was significantly lower at this pH. In particular, 91.0% (95.1% excluding H[N]) resonances could be assigned for [HisB24]-insulin and 86.5% (93.0% excluding H[N]) resonances for [D-HisB24]-insulin. The assignments obtained for analogues at pH 8.0 are shown in supplemental Tables S4 and S5. Because of the extensive overlaps of the NMR signals in certain spectral regions, the number of resonances that could not be unambiguously assigned was higher in the [D-HisB24]-insulin analog. The chemical shift values of the B24 histidine side chains for both analogues suggest their mono-protonated (neutral) state under basic conditions.

The ^{1}H resonance assignments obtained for [HisB24]-insulin and [D-HisB24]-insulin under acidic and basic conditions allowed automated assignment of the NOEs identified in two-dimensional NOESY spectra using protocols described under "Experimental Procedures." The numbers of observed NOE
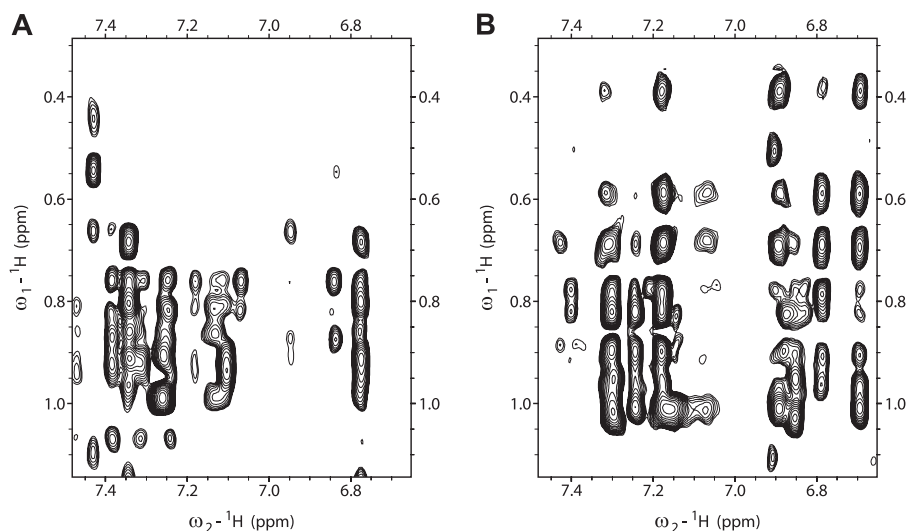
FIGURE 1. **The side chains of the aromatic residues from the C-terminal loop of B-chain re-establish the interaction with the methyl groups from the protein core under basic conditions.** *Panel A* shows a region from the two-dimensional NOESY spectrum of [HisB24]-insulin acquired at pH 1.9 showing NOE cross-peaks between aromatic and methyl hydrogens. The limited number of NOEs originates mainly from the contacts involving TyrA14, TyrA19, HisB5, and HisB10 with the adjacent methyl side chains. *Panel B* shows the data obtained for [HisB24]-insulin at pH 8.0. The additional NOE cross-peaks are from the contacts between the aromatic rings (His[B24] and Tyr[B25]) and methyl groups (Leu[B11] and Leu[B15]).

peaks, distance constraints, and structural statistics for obtained structures are shown in supplemental Table S6.

The $^1$H chemical shifts values for both analogues display a high degree of similarity for the parts of the molecule further from the introduced modification at the position B24, as illustrated by the calculated chemical shift differences for H$^N$ and H$^\alpha$ at pH 1.9 and 8.0 shown in supplemental Table S7. The significant differences were found for the H$^N$ chemical shift values from residues at positions B20-B26 at pH 1.9 and B23-B27 at pH 8.0. Interestingly, only a very limited number of the NOE contacts between the side chains of aromatic residues (His[B24], Phe[B25], Tyr[B26]) from the C-terminal part of the B-chain and hydrophobic core of insulin were observed in the NOESY spectra acquired at pH 1.9. However, an extensive network of these contacts re-emerged in the NOESY spectra obtained at pH 8 (Fig. 1).

The solution structures calculated from the NMR data for [HisB24]-insulin and [D-HisB24]-insulin at pH 1.9 and pH 8.0 are shown in Fig. 2. Detailed analysis revealed that A-chain of both analogues adopts a characteristic helix-turn-helix conformation independently of the pH. As expected, the N-terminal $\alpha$-helix of A-chain (A1-A6) is not very well defined at pH 8.0 due to an insufficient number of NOE contacts involving backbone amide hydrogens, which are key for stabilization of the local helical conformation in proteins. The effect is more evident for [D-HisB24]-insulin (Fig. 2D). In addition, the overall appearance of the [D-HisB24]-insulin structure was affected by a significantly lower number of NOE peaks observed in the NOESY spectrum, which resulted in a reduced number of the non-redundant distance constraints (supplemental Table S7).

Under acidic conditions (pH 1.9), the C-terminal parts of the B-chain (B21-B30) were found disordered in both analogues despite a well defined central $\alpha$-helix (B9-B20). Hence, medium to long range NOE contacts were not identified between residues B21 and B30 and the rest of the molecule (Fig. 2, *A* and *C*).

At pH 8.0, the NOE pattern indicating the reestablishment of the contacts between the core hydrophobic residues and the C-terminal loop of the B-chain was identified in the NOESY spectra of both analogues (Fig. 1) before the structure calculation. However, the overall organization of the pH 8.0 NMR structures (Fig. 2, *B* and *D*) of these analogues revealed significant differences in the conformations of their respective B-chains (Fig. 3).

The overall feature of the [HisB24]-insulin pH 8.0 NMR structure is its well ordered and defined character. Detailed analysis of the [HisB24]-insulin structure revealed that the histidine side chain is very well accommodated and stabilized in the B24 pocket not only by typical, wild-type-like nonpolar interactions (with Val[B12] and Leu[B15]) but by new hydrogen bonds as well (Fig. 3*A*). The protonated B24 imidazole N$^{\epsilon2}$ atom forms a strong (2.9- 3.2 Å) hydrogen bond (calculated between heteroatoms) with Val[B12] CO, whereas the deprotonated imidazole N$^{\delta1}$ atom forms a weaker and transient hydrogen bond (3.4- 3.9 Å) with Tyr[B26] NH. Subsequently, the Tyr[B26] side chain is shifted closer to the core of the molecule compared with its position in human insulin, and subsequently its OH group is engaged in hydrogen bonds with GlyB8 CO and/or GlyB8 NH (3.8 – 8.8 and 3.3–9.9 Å, respectively). However, the observed ranges suggest that these hydrogen bonds may be transient or mediated by water molecules, as observed in human insulin (57). The shift of Tyr[B26] side chain also enables its C-H π-like stacking contact with Pro[B28] side chain. All new interactions that involve B24-B28 residues result in a very firm stabilization of the B21-B30 $\beta$-strand, its tethering into the analog core, and the increased overall compactness of this protein.

In contrast to the ordered and compacted structure of the [His]-insulin analog, parts of the [D-His]-insulin structure at pH 8.0 are not fully defined. This is despite similar and clear reestablishment of contacts between the insulin hydrophobic core and the C-terminal loop of B-chain at pH 8.0, which were also

ZSI
ZSI
ZSI

## Structural Integrity of the B24 Site in Insulin
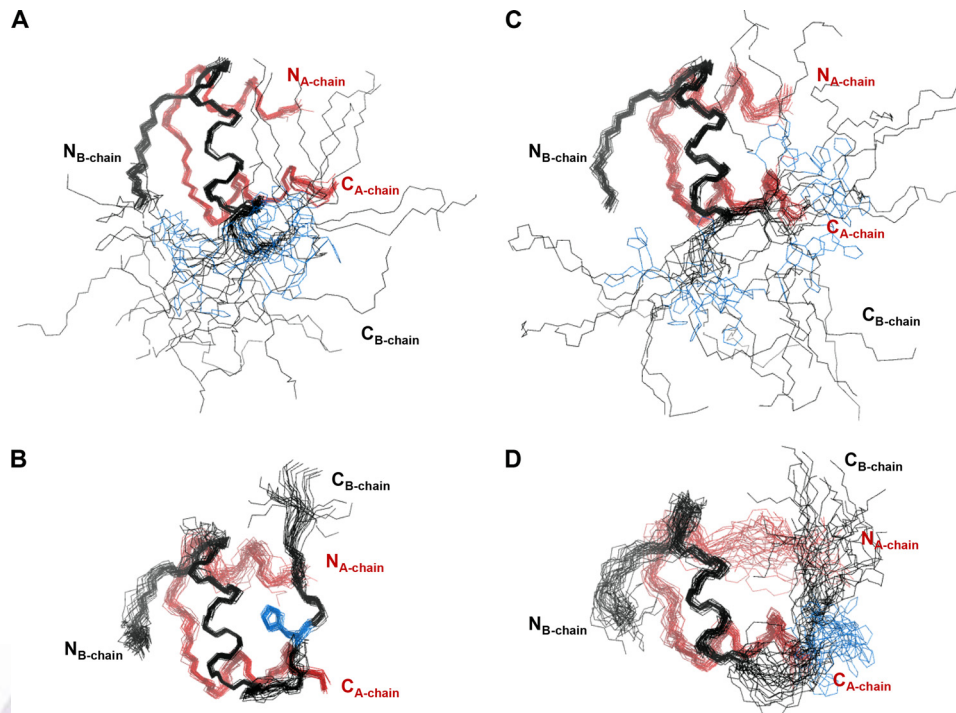


**AQ: I**  FIGURE 2. **The solution structures of the [HisB24]-insulin and [D-HisB24]-insulin analogues.** Structures are represented by best-fit superpositions of the protein backbone for 30 converged structures. The A-chain is colored in *red*, and the B-chain is in *black* in all structures. *Panel A* shows [HisB24]-insulin at pH 1.9, whereas *panel B* contains a set of structures obtained for the same analog at pH 8.0. [D-HisB24]-insulin structure at pH 1.9 is shown in *panel C* and at pH 8.0 in *panel D*. The positions of histidine side chains are colored in *blue*. The N and C termini of the insulin A and B chains are marked as $N_{A-chain}$, $C_{A-chain}$ (in *red*) and $N_{B-chain}$, $C_{B-chain}$ (in *black*), respectively. The C-terminal part of the B-chain (residues B21-B30), which is found flexible in both analogues under acidic conditions, reestablished extensive contacts with the protein core in the basic environment.

undefined in this analog at pH 1.9. The incorporation of the D-form of the His residue at the B24 site had a profound impact on the organization of the main chain and side chains in this region (Fig. 3*B*). This substitution changed the local B24-backbone topology, where the D-chirality of the B24C$^{\alpha}$ swayed the D-His$^{B24}$ away from the insulin core. The uprooting of the His side chain from the Phe$^{B24}$ pocket had, in turn, profound consequences on the B24 down- and upstream parts of the B-chain. Most remarkably, the cavity of the His-depleted Phe$^{B24}$ hydrophobic pocket was filled again by the Phe$^{B25}$, which mimics the "missing" D-His side chain. This retractile change of the "register" along the B23-B30 strand results in subsequent structural downshift of the Tyr$^{B26}$ into the B25 site. Here, the Tyr$^{B26}$ follows the Phe$^{B25}$-typical outward conformation. In addition, in contrast to the [HisB24] analog pH 8.0 structure, the B27-B30 residues depart from the insulin molecule core revealing previously fully buried Val$^{B12}$ and Leu$^{B15}$ (supplemental Fig. S3). The N-terminal arms of the B-chains (B1-B6) were found in the T-like states in both analogues, as in solution structures of monomeric insulins (7).

### MD of the Wild-type Insulin and [HisB24]- and [D-HisB24]-insulin Analogues

*Overall Structural Stability*—The 50-ns MD trajectories of the A-B and C-D insulin monomers derived from the 1mso dimer of human insulin (41) became stable after 20 ns with a r.m.s.d. oscillation ~2.5 or 3 Å from the starting x-ray structure (Fig. 4*A*).

The backbone r.m.s.d. of the [HisB24]-insulin 50-ns trajectory was more stable with r.m.s.d. oscillating ~2 Å from the first

NMR structure. The hydrogen bond pattern of [HisB24]-insulin observed in the NMR structure of this analog (N$^{\epsilon2}$-Val$^{B12}$ CO of 2.9–3.2 Å and N$^{\delta1}$-Tyr$^{B26}$ HN of 3.4-3.9 Å) was fully preserved throughout the whole simulation run. In contrast, the [D-HisB24]-insulin trajectory departed very significantly from the starting NMR structure (up to r.m.s.d. of ~5-6 Å) and appeared stable only after 40 ns at a higher plateau of 2.6 Å (Fig. 4*B*).

*Segment Movements*—To understand the movements of the C terminus of the B-chain (residues B24-B28) in relation to the insulin rigid core (B7-B19 central $\alpha$-helix), five representative distances (B24C$^{\alpha}$-B15C$^{\alpha}$, B25C$^{\alpha}$-B15C$^{\alpha}$, B26C$^{\alpha}$-B12C$^{\alpha}$, B27C$^{\alpha}$-B11C$^{\alpha}$, and B28C$^{\alpha}$-B8C$^{\alpha}$) between these structural elements were measured in [HisB24]-insulin, [D-HisB24]-insulin, and in A-B and C-D monomers of human insulin in the course of the MD production runs. The C$^{\alpha}$ atoms of residues B26-B28 in the wild-type A-B insulin monomer (supplemental Fig. S4*A*) displayed significant movements; however, the B24 and B25 C$^{\alpha}$ atoms remained relatively stable. Smaller shifts in the initial part of the simulation were observed in the case of wild-type C-D monomers (supplemental Fig. S4*B*); however, they tend to converge to the A-B-like r.m.s.d. after 20 ns. All selected distances in the [D-HisB24]-analog displayed relatively high flexibility in the course of MD runs (Fig. 5*B*). In contrast, the B24-B28 chain in [HisB24]-insulin was very rigid (Fig. 5*A*) without any significant movements (*cf.* Fig. 4*B*).

*Residue Flexibility*—Positional fluctuations of the wild-type insulin and analog residues over the 50-ns MD simulation runs are shown in supplemental Fig. S5. The largest flexibility
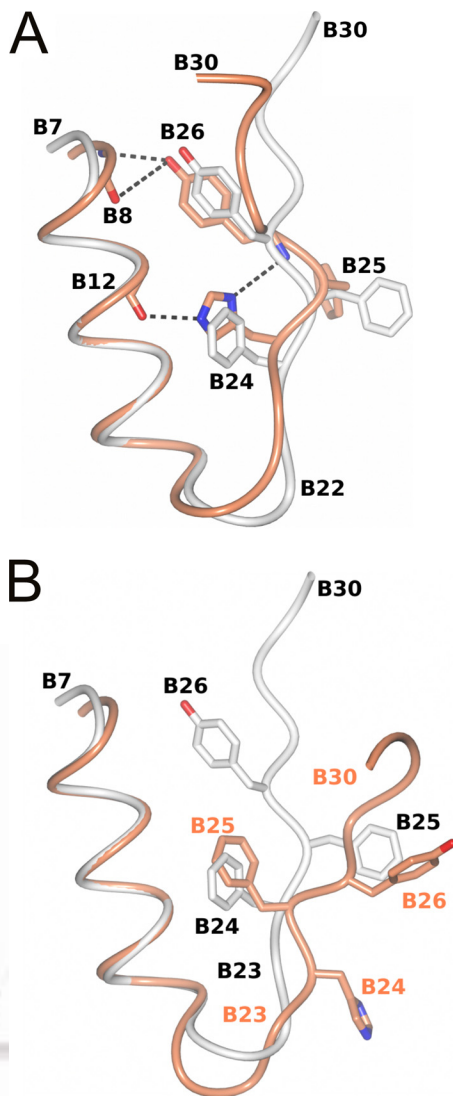
*Structural Integrity of the B24 Site in Insulin*



FIGURE 3. **An overlay of the B-chain of human insulin (PDB code 1mso (41), in *gray*) with the B-chain of [HisB24]-insulin (*panel A*, in *coral*) or [D-HisB24]-insulin (*panel B*, in *coral*) analogues (representative NMR structures at pH 8. 0).** Only the positions of Phe/His[B24], Phe[B25], and Tyr[B26] side chains are shown. *Dashed lines* in *panel A* indicate selected hydrogen bonds stabilizing the structure of [HisB24]-insulin.



FIGURE 4. **Backbone r.m.s.d. of human insulin monomers A-B (in *red*) and C-D (in *blue*) derived from 1mso dimer (41) (*panel A*) and [D-HisB24]-insulin (in *red*) and [HisB24]-insulin (in *blue*) (*panel B*) derived from their pH 8.0 NMR structures (this work) during 50 ns MD simulation runs.**
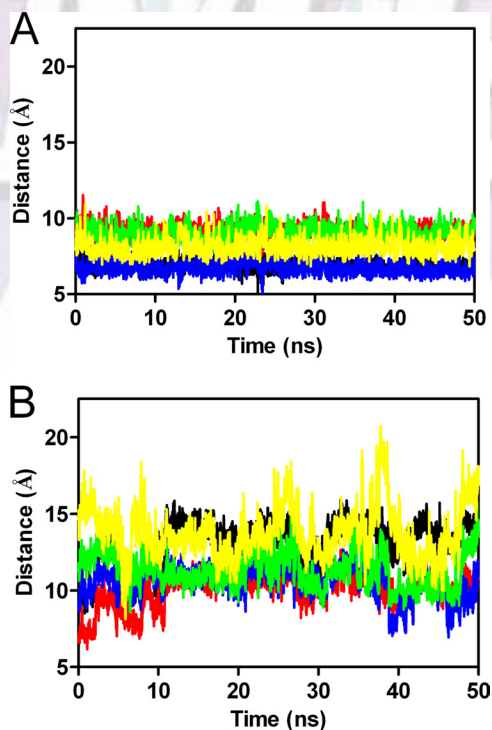


FIGURE 5. **The changes in five representative distances (B24Cα-B15Cα in *black*, B25Cα-B15Cα in *red*, B26Cα-B12Cα in *blue*, B27Cα-B11Cα in *green*, and B28Cα-B8Cα in *yellow*) were measured in [HisB24]-insulin (*panel A*) and [D-HisB24]-insulin (*panel B*) in the course of 50-ns MD runs.**

(>500) is concentrated on the N and C termini of the B-chain in both the A-B and C-D monomers of human insulin. The [HisB24]-insulin yielded the most stable structures (supplemental Fig. S5*A*). This trend, increased residue flexibility of the human insulin and [D-His]-insulin (supplemental Fig. S5*B*) and lower flexibility of [His]-insulin, was observed throughout the entirety of the insulin molecule.

## DISCUSSION

The three aromatic amino acids Phe[B24], Phe[B25], and Tyr[B26] from the C-terminal part of the B-chain are some of the key residues identified in human insulin. These residues are important for both formation of the dimer interface of the hormone (3) and its interactions with IR (18, 20). Considering the prolonged absence of structural description of the B22-B30-chain-IR interface (5), we focused here on a more systematic probing of the role of Phe[B24] aromatic side chain and chirality of its Cα atom for the structure and function of the hormone. Extensive and complex data emerging from other studies on the B24 site are summarized in supplemental Table S8. Focusing on

## Structural Integrity of the B24 Site in Insulin

the B24 site, we have tested its functional and structural response to (i) an alternative polar aromatic side chain (His), (ii) inverse $C^\alpha$ chirality (D-His), and (iii) structural main chain restrains (Pro, Sar) on hormone structure and function. The histidine side chain was selected for its partially aromatic character, similar size to phenylalanine side chain, and its imidazole hydrogen-bond formation potential. The possible functional cooperativity between B24 and B26 sites in insulin-IR interactions was further investigated in a double [D-HisB24, GluB26]-insulin mutant and its variant ([D-HisB24, GluB26, LysB28, ProB29]-insulin) (Table 1).

AQ: B

The replacement of $Phe^{B24}$ by any L-amino acid disrupts insulin biological activity and affinity (supplemental Table S8). The binding affinities of non-aromatic $Leu^{B24}$, $Ser^{B24}$, and $Ala^{B24}$ insulin analogues do not exceed 20% in most cases. Interestingly, the simultaneous loss of the side chain and chirality at the B24 position (*i.e.* $Gly^{B24}$]-insulin) is relatively well tolerated, yielding affinities in the range of 22–78%. In contrast, the replacement of $Phe^{B24}$ by D-amino acids enhances binding affinity in most of the analogues (*i.e.* [D-AlaB24]-insulin, 150%; [D-PheB2⁴]-insulin, 140–180%), with an exception for the D-$Pro^{B24}$ analog. Double B24/B26 mutants followed this trend as the [D-TyrB24, PheB26]-insulin possesses 163% affinity ([PheB26]-insulin itself displays 45% binding affinity (3, 58)), whereas its L-amino acid variant [TyrB24, PheB26]-insulin is inactive (2%). Interestingly, the incorporation of an achiral amino acid derivative with a pair of methyl groups attached to the $C^\alpha$ carbon atom, $\alpha$-aminoisobutyric acid, at the B24 site results in an $\alpha$-aminoisobutyric acid B24-insulin analog with no potency (0.3%).

Therefore, the high affinity (212%) of the [D-HisB24]-insulin and weak binding of [HisB24]-insulin (1.5% affinity) obtained here were in agreement with trends observed in available data. However, they did not provide immediate explanation of the binding pattern observed in B24-site insulin mutants. Therefore, a comprehensive structural characterization of [HisB24]- and [D-HisB24]-insulins was undertaken in this study. NMR structures were determined in parallel at acidic and basic pH values (*i.e.* pH ~1.9 *versus* pH 8.0) to avoid possible bias resulting from the pH and buffer effects on insulin conformation (55). Additional insights were obtained from theoretical MD simulations.

Indeed, the crystal and NMR structures of [HisB24]-insulin and [D-HisB24]-insulin analogues obtained here provide rational explanation of the importance of an aromatic invariance of the B24 site and underline the conservation of the structural integrity of the $Phe^{B24}$ and its pocket for an effective hormone functionality.

The striking feature of both the crystal and NMR structures of [HisB24]- and [D-HisB24]-insulins is the widespread disorder of B20-B30 chains at pH 1.9 compared with their well- and reasonably ordered conformations at pH 8.0 (Fig. 2). A similar disorder of the B20-B30 part of insulin has been observed in acidic pH NMR structures of other B24 mutants such as [GlyB24]-insulin (7, 59), [GlyB24,GluB16, *des*B30]-insulin (55) and in [D-AlaB24]-insulin (17). Hence, it is possible that this phenomenon arises from the structurally deleterious effects of low pH and/or acetic acid on some insulin analogues (55). How-

ever, it cannot be excluded that disorder of [HisB24]-insulin and [D-HisB24]-insulin analogues at pH 1.9 arise from the repulsive effect of the doubly protonated $His^{B24}$ that cannot be accommodated in the B24 hydrophobic cavity. Therefore, we focused here on analysis of more physiological pH 8.0-derived insulin structures and their defined conformations.

The NMR spectra at pH 8.0 yielded a well ordered and compact structure of [HisB24]-insulin (Fig. 2B). Its main feature is a stable accommodation of the $N^{\epsilon 2}$-monoprotonated imidazole ring inserted in the B24 hydrophobic cavity. Interestingly, the histidine side chain is accommodated there even more firmly than the phenyl moiety of $Phe^{B24}$ as it thrusts into the cavity ~1.2 Å deeper than the $Phe^{B24}$ phenyl ring of the wild-type insulin. This $His^{B24}$-related compactness of the structure is a likely result of the combination of the combined effects of the still-present aromatic character of the B24 imidazole with its hydrogen bond-forming capability. In addition to close van der Waals imidazole B24-$Leu^{B15}$ contacts, the side chain of $His^{B24}$ is anchored firmly by strong $B24N^{\epsilon 2}$-CO$Val^{B12}$ and weaker $B24N^{\delta 1}$-NH$Tyr^{B26}$ hydrogen bonds (Fig. 3A). This tight engagement of $His^{B24}$ within the B24 cavity has a precipitous effect on the remaining part of the B25-B30 chain, which is also pulled back close to the core of the hormone. This insulin-wrapping conformation of the B-chain is further stabilized by $Gly^{B8}$CO/NH-OH$Tyr^{B26}$ direct hydrogen bonds and wedging of the $Tyr^{B26}$ phenyl ring between $Pro^{B28}$ and $Ile^{A2}$. This "immobilization" and tethering of the B24-B30 chain into the [HisB24]-insulin core at pH 8.0 are likely reasons for the low affinity of this analog and of its monomeric behavior.

The "over-conservation" of the B24-B30-fold observed in [HisB24]-insulin is in contrast with the radical rearrangement of the B20-B30 chain in the [D-HisB24]-insulin at the same, high pH 8.0 (Fig. 2D). Such rearrangement is a result of the pull-out of the histidine side chain from the Phe B24 pocket by the inverse chirality of the D-$His^{B24}$ $C^\alpha$ atom and subsequent/simultaneous filling-up of the emerging empty cavity by another available aromatic side chain that is provided here by $Phe^{B25}$. The locations of $Phe^{B24}$ and $Phe^{B25}$ aromatic rings in the wild-type and [D-HisB24]-insulins, respectively, are very similar, with a remarkable conservation of the position of the $C^\alpha$ atoms that differ by only ~0.55 Å in both analogues (Fig. 3B).

The compensatory capability of the $Phe^{B25}$ to act as $Phe^{B24}$ structural substitute has also been observed in the pH 8.0 NMR structure of [GlyB24, GluB16, *des*B30]-insulin in which both the side chain of the B24 residue and the chirality of its $C^\alpha$ atom have been removed (55). In contrast, the NMR structure of [D-AlaB24]-insulin (17) indicated disorder of the B20-B30 part of the molecule. However, due to a lack of some NMR experimental detail there (*e.g.* pH at which the NMR spectra were measured), we suppose that the significant mobility of this chain in [D-AlaB24]-insulin arise from the use of low pH (deuterioacetic acid) in these NMR experiments (17). The "down-register" shift of the B24-B30 chain in [D-HisB24]-insulin is structurally propagated there on both sides of the B24 site. For example, the B19-B22 $\beta$-turn is lost in favor of a larger bulge, $Tyr^{B26}$ occupies $Phe^{B25}$ position, and the B26-B30 chain departs from the insulin core. Importantly, these large scale

structural rearrangements are correlated with the high affinity of this analog. Furthermore, it is not excluded that the B26→B25 downshift observed in [D-HisB24]-insulin contributes also to the high affinity of this analog because Glendorf *et al.* (60) reported [TyrB25]-insulin with a high binding affinity (285%).

The NMR/x-ray experimental structural evidence about [HisB24]- and [D-HisB24]-insulins was confirmed by MD simulations. First, it was established that the overall mobility and dynamics of the wild-type insulin does not depend on slightly different starting models (*i.e.* A-B *versus* C-D monomers derived from the human insulin dimer, PDB code 1mso) (41). Although the C terminus of the B-chain detached sooner from the insulin core in the A-B monomer compared with the C-D one (*cf.* the overall r.m.s.d. in Fig. 4*A*), both simulations converged to stable plateaus of similar r.m.s.d. after an ~20-ns-long run. The reason for this delayed detachment in the C-D monomer is possibly due to the salt bridge between the Gly[A1] N and Thr[B30] O atoms present in the x-ray structure only in this monomer and persisting in the MD up to four ns. In addition, our simulations agreed with previous insulin MD studies (61, 62); all of the simulations indicate the same range of insulin mobility and partial, approximately B25-B30, detachment of the B-chain from the hormone core. Second, the MD simulations also fully agreed with the distinct NMR conformational stabilities of [HisB24]- and [D-HisB24]-insulins. The r.m.s.d. for [HisB24]-insulin are strikingly invariant and steady, in contrast with the higher r.m.s.d. of [D-HisB24]-insulin that indicates its increased mobility that exceeds the dynamics of the wild-type insulin (Fig. 4*B*). Third, the analysis of selected distances between the key regions of the hormone indicates the detachment of the B25-B30 B-chain in the wild-type and [D-HisB24]-insulin but conservation of the B24 C[α] region (Fig. 5).

Direct correlation of the affinities of [HisB24]- and [D-HisB24]-insulin with their contrasting structural features observed here, *i.e.* "gluing" of the B24-B30 chain to insulin core *versus* "liberation" of the B25-B30 chain, respectively, confirm the established consensus about the detachment of the B-chain needed for insulin activation upon its binding to the IR. However, the scale of the detachment of the B20-B30 chain in this process is still not determined. It is debatable whether a full unfolding of the approximately B22-B30 chain is needed for insulin activation (18) or whether the effective engagement of the hormone with IR can be achieved by a more modest reshaping of the B25-B30 residues in a form similar to that observed in B26-turn containing high affinity analogues (19). The high affinity of the [D-HisB24]-insulin and its structural persistence for conservation and occupation of the Phe[B24] cavity by an aromatic moiety suggest that the Phe[B24] side chain may remain in place upon insulin activation. The maintenance of the B24 "aromatic" cavity in a very mobile and reshaped, but still highly active, analog is in concordance with the strict preservation of the B24 side chain in high affinity analogues in which it was considered as one of the crucial invariant side chain pivots (19). In addition, it appears that [D-HisB24]-insulin, with its less rigid and more flexible structure, is better primed than native insulin for the detachment and activation of the approximately B25-B30 C-terminal β-strand upon binding to the IR.

The need for the structural invariance of the B24 C[α] atoms is supported further by the low affinities of [ProB24]-insulin and [SarB24]-insulin analogues (Table 1 and supplemental Table S8) that contain the main chain constraining amino acids (63). The Pro[B24] and Sar[B24] substitutions not only result in the absence of aromatic side chains, but their specific dihedral angles impose restrictions on moderate flexibility at the B24 site, likely excluding any B25→B24-like compensatory structural downshift observed in highly active [D-HisB24]-insulin. In contrast, the dihedral $\varphi/\phi$ angles of glycine and D-amino acids at the B24 site allow for a rearrangement of insulin molecule that leads to an effective IR binding. The structural vulnerability of the main chain at the B24 site and the sensitivity of its C[α] chirality are, therefore, expressed in both aromatic pulldown (B25→B24) structural rearrangements and maintenance of the position of this atom. These underline the need for an overall conservation of the phenylalanine-like character of this site. However, it cannot be excluded that the low affinities of [ProB24]-insulin and [SarB24]-insulin analogues may result from the loss of B24 amide hydrogens, which may be important for IR binding (3).

Despite the progress in understanding the roles of individual substitutions at the B24 site, the negative-cooperative effects of double B24/B26 mutations in [D-HisB24, GluB26]- and [D-HisB24, GluB26, LysB28,ProB29]-insulins, which abolish affinity gains of single site replacements at the B24/B26 sites, are less obvious. The only one available NMR structure of [D-HisB24, GluB26, LysB28, ProB29]-insulin determined at pH 1.9 that shows large scale disorder of the B-chain (data not shown) observed in "acidic" structures of [HisB24]- and [D-HisB24]-insulins, and the structural effect of double [D-HisB24,GluB26] mutation remains unclear. However, the low affinity of the [D-HisB24,GluB26]-mutants may result from the unfavorable replacement of Phe[B25] aromatic ring by glutamic acid side chain if the B25 → B24 downshift takes place in this analog as well as in [D-HisB24]-insulin. If such rearrangement does occur there, the Glu[B26]-side chain should occupy the position of the B25 phenyl ring in wild-type insulin, hence mimicking the Phe[B25]→ Glu[B25] mutation, which is most likely fully detrimental as observed in the case of similar Phe[B25]→ Asp[B25] substitution (64). In contrast, the B25→B24 downshift in high affinity [D-HisB24]-insulin is compensated by the move of another aromatic Tyr[B26] side chain into the B25 position. It is also possible that the Glu[B26] side chain follows the Tyr[B26] solvent-pointing conformation observed in [D-HisB24]-insulin and, therefore, is capable of forming hydrogen bonds to also solvent-exposed D-His[B24] and neighboring Asn[A21] or Thr[B30]. These interactions altogether can have a negative effect on IR binding due to the over-stabilization of this region.

In summary, our data rationalize the positive effects of D-amino acid substitutions at the B24 site. They indicate the importance of the aromatic character of the B24 side chain for insulin activity, which is demonstrated by the B25→B24 "rescue" downshift of Phe[B25] into the B24 site. They also underline the invariance of the position of the B24C[α] atom for effective IR binding. Moreover, they also provide supporting evidence for a limited, B25 onward, unfolding of the C terminus of the B-chain during hormone activation in disfavor of its full (B22-B30)

## Structural Integrity of the B24 Site in Insulin

detachment. However, the actual nature of Phe[B24] engagement with IR requires further insulin-IR complex studies.

### REFERENCES

1. Ward, C. W., and Lawrence, M. C. (2009) Ligand-induced activation of the insulin receptor. A multi-step process involving structural changes in both the ligand and the receptor. *Bioessays* **31,** 422–434
2. Steiner, D. F. (2011) Adventures with insulin in the islets of Langerhans. *J. Biol. Chem.* **286,** 17399–17421
3. Antolíková, E., Žáková, L., Turkenburg, J. P., Watson, C. J., Hančlová, I., Šanda, M., Cooper, A., Kraus, T., Brzozowski, A. M., and Jiráček, J. (2011) Non-equivalent role of inter- and intramolecular hydrogen bonds in the insulin dimer interface. *J. Biol. Chem.* **286,** 36968–36977
4. Pullen, R. A., Lindsay, D. G., Wood, S. P., Tickle, I. J., Blundell, T. L., Wollmer, A., Krail, G., Brandenburg, D., Zahn, H., Gliemann, J., and Gammeltoft, S. (1976) Receptor binding region of insulin. *Nature* **259,** 369–373
5. Menting, J. G., Whittaker, J., Margetts, M. B., Whittaker, L. J., Kong, G. K., Smith, B. J., Watson, C. J., Záková, L., Kletvíková, E., Jiráček, J., Chan, S. J., Steiner, D. F., Dodson, G. G., Brzozowski, A. M., Weiss, M. A., Ward, C. W., and Lawrence, M. C. (2013) How insulin engages its primary binding site on the insulin receptor. *Nature* **493,** 241–245
6. Dodson, E. J., Dodson, G. G., Hubbard, R. E., and Reynolds, C. D. (1983) Insulin's structural behavior and its relation to activity. *Biopolymers* **22,** 281–291
7. Hua, Q. X., Shoelson, S. E., Kochoyan, M., and Weiss, M. A. (1991) Receptor binding redefined by a structural switch in a mutant human insulin. *Nature* **354,** 238–241
8. De Meyts, P., Van Obberghen, E., and Roth, J. (1978) Mapping of the residues responsible for the negative cooperativity of the receptor binding region of insulin. *Nature* **273,** 504–509
9. Weiss, M. A. (2009) in *Insulin and IGFs*, Elsevier Academic Press, Inc., San Diego, CA
10. Derewenda, U., Derewenda, Z., Dodson, E. J., Dodson, G. G., Bing, X., and Markussen, J. (1991) X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule. A completely inactive analogue. *J. Mol. Biol.* **220,** 425–433
11. Mayer, J. P., Zhang, F., and DiMarchi, R. D. (2007) Insulin structure and function. *Biopolymers* **88,** 687–713
12. Ludvigsen, S., Roy, M., Thøgersen, H., and Kaarsholm, N. C. (1994) High resolution structure of an engineered biologically potent insulin monomer, B16 Tyr→His, as determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **33,** 7998–8006
13. Hua, Q. X., Hu, S. Q., Frank, B. H., Jia, W., Chu, Y. C., Wang, S. H., Burke, G. T., Katsoyannis, P. G., and Weiss, M. A. (1996) Mapping the functional surface of insulin by design. Structure and function of a novel A-chain analogue. *J. Mol. Biol.* **264,** 390–403
14. Weiss, M. A., Hua, Q. X., Lynch, C. S., Frank, B. H., and Shoelson, S. E. (1991) Heteronuclear two-dimensional NMR studies of an engineered insulin monomer. Assignment and characterization of the receptor binding surface by selective ^{2}H and ^{13}C labeling with application to protein design. *Biochemistry* **30,** 7373–7389
15. Hua, Q. X., and Weiss, M. A. (1991) Comparative two-dimensional NMR studies of human insulin and *des*-pentapeptide insulin. Sequential resonance assignment and implications for protein dynamics and receptor recognition. *Biochemistry* **30,** 5505–5515
16. Kurapkat, G., Siedentop, M., Gattner, H. G., Hagelstein, M., Brandenburg, D., Grötzinger, J., and Wollmer, A. (1999) The solution structure of a superpotent B chain-shortened single-replacement insulin analogue. *Protein Sci.* **8,** 499–508
17. Hua, Q. X., Xu, B., Huang, K., Hu, S. Q., Nakagawa, S., Jia, W., Wang, S., Whittaker, J., Katsoyannis, P. G., and Weiss, M. A. (2009) Enhancing the activity of a protein by stereospecific unfolding. Conformational life cycle of insulin and its evolutionary origins. *J. Biol. Chem.* **284,** 14586–14596
18. Xu, B., Huang, K., Chu, Y. C., Hu, S. Q., Nakagawa, S., Wang, S., Wang, R. Y., Whittaker, J., Katsoyannis, P. G., and Weiss, M. A. (2009) Decoding the cryptic active conformation of a protein by synthetic photoscanning. Insulin inserts a detachable arm between receptor domains. *J. Biol. Chem.* **284,** 14597–14608
19. Jirácek, J., Záková, L., Antolíková, E., Watson, C. J., Turkenburg, J. P., Dodson, G. G., and Brzozowski, A. M. (2010) Implications for the active form of human insulin based on the structural convergence of highly active hormone analogues. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 1966–1970
20. Mirmira, R. G., Nakagawa, S. H., and Tager, H. S. (1991) Importance of the character and configuration of residues B24, B25, and B26 in insulin-receptor interactions. *J. Biol. Chem.* **266,** 1428–1436
21. Shoelson, S. E., Lu, Z. X., Parlautan, L., Lynch, C. S., and Weiss, M. A. (1992) Mutations at the dimer, hexamer, and receptor binding surfaces of insulin independently affect insulin-insulin and insulin-receptor interactions. *Biochemistry* **31,** 1757–1767
22. Shoelson, S., Fickova, M., Haneda, M., Nahum, A., Musso, G., Kaiser, E. T., Rubenstein, A. H., and Tager, H. (1983) Identification of a mutant human insulin predicted to contain a serine-for-phenylalanine substitution. *Proc. Natl. Acad. Sci. U.S.A.* **80,** 7390–7394
23. Inouye, K., Watanabe, K., Tochino, Y., Kanaya, T., Kobayashi, M., and Shigeta, Y. (1981) Semisynthesis and biological properties of the [B24-leucine], [B25-leucine] and [B24-leucine, B25-leucine] analogues of human insulin. *Experientia* **37,** 811–813
24. Wollmer, A., Strassburger, W., Glatter, U., Dodson, G. G., McCall, M., Gattner, H. G., Danho, W., Brandenburg, D., and Rittel, W. (1981) Two mutant forms of human insulin. Structural consequences of the substitution of invariant B24- or B25-phenylalanine by leucine. *Hoppe Seyler's Z. Physiol. Chem.* **362,** 581–591
25. Tager, H., Thomas, N., Assoian, R., Rubenstein, A., Saekow, M., Olefsky, J., and Kaiser, E. T. (1980) Semisynthesis and biological activity of porcine [LeuB24]insulin and [LeuB25]insulin. *Proc. Natl. Acad. Sci. U.S.A.* **77,** 3181–3185
26. Gattner, H. G., Danho, W., Behn, C., and Zahn, H. (1980) The preparation of two mutant forms of human insulin, containing leucine in position B24 or B25, by enzyme-assisted synthesis. *Hoppe Seyler's Z. Physiol. Chem.* **361,** 1135–1138
27. Inouye, K., Watanabe, K., Tochino, Y., Kobayashi, M., and Shigeta, Y. (1981) Semisynthesis and properties of some insulin analogs. *Biopolymers* **20,** 1845–1858
28. Keefer, L. M., Piron, M. A., De Meyts, P., Gattner, H. G., Diaconescu, C., Saunders, D., and Brandenburg, D. (1981) Impaired negative cooperativity of the semisynthetic analogues human [LeuB24] and [LeuB25] insulins. *Biochem. Biophys. Res. Commun.* **100,** 1229–1236
29. Assoian, R. K., Thomas, N. E., Kaiser, E. T., and Tager, H. S. (1982) [LeuB24]insulin and [AlaB24]insulin- altered structures and cellular processing of B24-substituted insulin analogs. *Proc. Natl. Acad. Sci. U.S.A.* **79,** 5147–5151
30. Mirmira, R. G., and Tager, H. S. (1989) Role of the phenylalanine B24 side chain in directing insulin interaction with its receptor. Importance of main chain conformation. *J. Biol. Chem.* **264,** 6349–6354
31. Svoboda, I., Brandenburg, D., Barth, T., Gattner, H. G., Jirácek, J., Velek, J., Bláha, I., Ubik, K., Kasicka, V., and Pospísek, J. (1994) Semisynthetic insulin analogues modified in positions B24, B25, and B29. *Biol. Chem. Hoppe Seyler* **375,** 373–378
32. Nakagawa, S. H., and Tager, H. S. (1993) Importance of main-chain flexibility and the insulin fold in insulin-receptor interactions. *Biochemistry* **32,** 7237–7243
33. Casaretto, M., Spoden, M., Diaconescu, C., Gattner, H. G., Zahn, H., Brandenburg, D., and Wollmer, A. (1987) Shortened insulin with enhanced *in vitro* potency. *Biol. Chem. Hoppe Seyler* **368,** 709–716
34. Záková, L., Zyka, D., Jezek, J., Hanclová, I., Sanda, M., Brzozowski, A. M., and Jirácek, J. (2007) The use of Fmoc-Lys(Pac)-OH and penicillin G acylase in the preparation of novel semisynthetic insulin analogs. *J. Pept. Sci.* **13,** 334–341
35. Záková, L., Kazdová, L., Hanclová, I., Protivínská, E., Sanda, M., Budesínský, M., and Jirácek, J. (2008) Insulin analogues with modifications at position B26. Divergence of binding affinity and biological activity. *Biochemistry* **47,** 5858–5868
36. Morcavallo, A., Genua, M., Palummo, A., Kletvikova, E., Jiracek, J., Brzozowski, A. M., Iozzo, R. V., Belfiore, A., and Morrione, A. (2012) Insulin

tapraid4/zbc-bc/zbc-bc/zbc01913/zbc4610-13z    ZSUBMIT    xppws S=5    7/3/13    9:24    **4/Color Figure(s) F2,3,4,5**    ARTNO: M112.448050

*Structural Integrity of the B24 Site in Insulin*

and insulin-like growth factor II differentially regulate endocytic sorting and stability of insulin receptor isoform A. *J. Biol. Chem.* **287,** 11422–11436

37. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326

38. Collaborative Computational Project, Number 4 (1994) The Ccp4 Suite. Programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763

39. Emsley, P., and Cowtan, K. (2004) Coot. Model-building tools for molecular graphics. *Acta Crystallogr. D.* **60,** 2126–2132

40. Vagin, A., and Teplyakov, A. (1997) MOLREP. An automated program for molecular replacement. *J. Appl. Crystallogr.* **30,** 1022–1025

41. Smith, G. D., Pangborn, W. A., and Blessing, R. H. (2003) The structure of T-6 human insulin at 1.0 angstrom resolution. *Acta Crystallogr. D.* **59,** 474–482

42. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D.* **53,** 240–255

43. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319,** 209–227

44. Güntert, P., and Wüthrich, K. (1991) Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints. *J. Biomol. NMR* **1,** 447–456

45. Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.* **26,** 1668–1688

46. Tsui, V., and Case, D. A. (2000) Molecular dynamics simulations of nucleic acids with a generalized born solvation model. *J. Am. Chem. Soc.* **122,** 2489–2498

47. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL. A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14,** 51–55

48. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935

49. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690

50. Kurtzhals, P., Schäffer, L., Sørensen, A., Kristensen, C., Jonassen, I., Schmid, C., and Trüb, T. (2000) Correlations of receptor binding and metabolic and mitogenic potencies of insulin analogs designed for clinical use. *Diabetes* **49,** 999–1005

51. Ciszak, E., Beals, J. M., Frank, B. H., Baker, J. C., Carter, N. D., and Smith, G. D. (1995) Role of C-terminal B-chain residues in insulin assembly. The structure of hexameric Lys[B28]Pro[B29]-human insulin. *Structure* **3,** 615–622

52. Brange, J., Ribel, U., Hansen, J. F., Dodson, G., Hansen, M. T., Havelund, S., Melberg, S. G., Norris, F., Norris, K., and Snel, L. (1988) Monomeric insulins obtained by protein engineering and their medical implications. *Nature* **333,** 679–682

53. Yao, Z. P., Zeng, Z. H., Li, H. M., Zhang, Y., Feng, Y. M., and Wang, D. C. (1999) Structure of an insulin dimer in an orthorhombic crystal. The structure analysis of a human insulin mutant (B9 Ser→Glu). *Acta Crystallogr. D.* **55,** 1524–1532

54. Whittingham, J. L., Scott, D. J., Chance, K., Wilson, A., Finch, J., Brange, J., and Guy Dodson, G. (2002) Insulin at pH 2. structural analysis of the conditions promoting insulin fibre formation. *J. Mol. Biol.* **318,** 479–490

55. Ludvigsen, S., Olsen, H. B., and Kaarsholm, N. C. (1998) A structural switch in a mutant insulin exposes key residues for receptor binding. *J. Mol. Biol.* **279,** 1–7

56. Cosgrove, M. S., Loh, S. N., Ha, J. H., and Levy, H. R. (2002) The catalytic mechanism of glucose-6-phosphate dehydrogenases. Assignment and H-1 NMR spectroscopy pH titration of the catalytic histidine residue in the 109-kDa *Leuconostoc mesenteroides* enzyme. *Biochemistry* **41,** 6939–6945

57. Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. M., Hubbard, R. E., Isaacs, N. W., and Reynolds, C. D. (1988) The structure of 2Zn pig insulin crystals at 1.5 Å resolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **319,** 369–456

58. Gauguin, L., Klaproth, B., Sajid, W., Andersen, A. S., McNeil, K. A., Forbes, B. E., and De Meyts, P. (2008) Structural basis for the lower affinity of the insulin-like growth factors for the insulin receptor. *J. Biol. Chem.* **283,** 2604–2613

59. Hua, Q. X., Shoelson, S. E., and Weiss, M. A. (1992) Nonlocal structural perturbations in a mutant human insulin. Sequential resonance assignment and ^{13}C-isotope-aided two-dimensional NMR studies of [PheB24→Gly]insulin with implications for receptor recognition. *Biochemistry* **31,** 11940–11951

60. Glendorf, T., Stidsen, C. E., Norrman, M., Nishimura, E., Sorensen, A. R., and Kjeldsen, T. (2011) Engineering of insulin receptor isoform-selective insulin analogues. *PLoS ONE* **6,** 1–6

61. Zoete, V., Meuwly, M., and Karplus, M. (2004) A comparison of the dynamic behavior of monomeric and dimeric insulin shows structural rearrangements in the active monomer. *J. Mol. Biol.* **342,** 913–929

62. Zoete, V., and Meuwly, M. (2006) Importance of individual side chains for the stability of a protein fold. Computational alanine scanning of the insulin monomer. *J. Comput. Chem.* **27,** 1843–1857

63. Tran, T. T., Treutlein, H., and Burgess, A. W. (2006) Designing amino acid residues with single conformations. *Protein Eng. Des. Sel.* **19,** 401–408

64. Drejer, K., Kruse, V., Larsen, U. D., Hougaard, P., Bjørn, S., and Gammeltoft, S. (1991) Receptor binding and tyrosine kinase activation by insulin analogs with extreme affinities studied in human hepatoma HepG2 cells. *Diabetes* **40,** 1488–1495

AQ: E

AQ: F

AQ: G

AQ: H