Identifying Small Molecule Modulators of IR-IGF1R Hybrid Formation

Samuel James Turvey

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

Leeds Institute of Cardiovascular and Metabolic Medicine

Faculty of Medicine and Health

September 2023

The candidate confirms that the work submitted is their own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 1 contains work described in: Recent developments in the structural characterisation of the IR and IGF1R: implications for the design of IR-IGF1R hybrid receptor modulators; S.J. Turvey, M.J. McPhillie, M.T. Kearney, S.P. Muench, K.J. Simmons, C.W.G Fishwick. *RSC Med. Chem*. **2022**, *13*(4):360-374.

Chapter 3 contains work described in: A small molecule reveals role of insulin receptor-insulin like growth factor-1 receptor heterodimers;
K.J. Simmons, C.G. Myers, H. Viswambharan, N. J Haywood, K. Bridge, S.J. Turvey, T. Armstrong, L. Lunn, P.J Meakin, E. Clavane, D. J. Beech, R. M. Cubbon, S. B Wheatcroft, M. J. McPhillie, C.W.G Fishwick, M. T. Kearney, *BioRχiv*, **2021**.

This thesis contains work carried out solely by S. J. Turvey with the following exceptions:

1. Compounds **11**, **12**, **13** and **14** were synthesised by Dr Martin McPhillie.

2. Compounds **57**, **58**, **59** and **60** were synthesised by Hannah McCarrick.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Samuel James Turvey to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

© The University of Leeds and Samuel James Turvey

Acknowledgements

Abstract

Type 2 diabetes is characterised by the disruption of insulin and insulin-like growth factor (IGF) signalling. The key hubs of these signalling cascades - the Insulin receptor (IR) and Insulin-like growth factor 1 receptor (IGF1R) – are known to form functional IR-IGF1R hybrid receptors which are insulin resistant. However, the mechanisms underpinning IR-IGF1R hybrid formation are not fully understood. Identifying the means to modulate hybrid formation will inform future therapies targeting this receptor. In this work, we aimed to develop improved structural models of the IR-IGF1R hybrid receptors and evaluate these through a combination of structural studies and structure-based drug design. Initially, two homology models of the major heterodimer interface contained in the hybrid receptors were generated from the published crystal structures of the IR and IGF1R homodimers. To pinpoint suitable sites for intervention, computational hotspot prediction was utilised to identify promising epitopes for targeting for point mutagenesis and virtual screening. Specific IGF1R point mutations F450A, R391A and D555A show reduced affinity of the hybrid receptor in a BRET based donor-saturation assay, confirming hybrid formation could be modulated at this interface. Additionally, a virtual screening campaign was directed against epitopes identified in the two hybrid homology models. Small molecules prioritised by screening were evaluated for their ability to modulate hybrid formation in a bioluminescence resonance energy transfer (BRET) assay. This led to the identification of a small molecule which was able to increase hybrid formation with an EC_{50} of 196 μM, which was confirmed in an orthogonal immunoprecipitation-based assay. RT qPCR experiments determined that this compound was likely to be promoting hybrid formation through promoting the expression of the INSR and IGF1R genes. Finally, we attempted to generate IR-IGF1R hybrid protein perform structural studies on the hybrid receptor, and further refine our structural models. These experiments provide the basis for rational design of more effective hybrid receptor modulators, supporting the prospect of identifying a small molecule that specifically interacts with this target.

Table of Contents

Table of Figures

Table of Tables

List of Abbreviations

Common Amino Acids

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

Abbreviations

αC	α-helix C
αCT	Alpha C-terminal helix
ADME	Absorption, distribution, metabolism, and excretion
AGE (products)	Advanced glycation end (products)
Akt	Protein kinase B
AMPK	5' adenosine monophosphate-activated protein kinase
ATP	Adenosine triphosphate
BRET	Bioluminescence resonance energy transfer
CASP	Critical assessment of methods of protein structure prediction
CR	Cysteine rich domain
CVD	Cardiovascular disease
DM	Diabetes Mellitus
DNA	Deoxyribonucleic acid
DPPIV	Dipeptidyl-peptidase 4
ECD	Ectodomain
EM	Electron microscopy
eNOS	Endothelial nitric oxide synthase

ERK	Extracellular-signal regulated kinase
FnIII-1	First fibronectin type III domain
FnIII-2	Second fibronectin type III domain
FnIII-3	Third fibronectin type III domain
GDP	Guanosine diphosphate
GLP1	Glucagon-like peptide 1
GRB2	Growth factor receptor bound protein 2
GTP	Guanosine triphosphate
HGMD	Human gene mutation database
HMG-CoA reductase	β-Hydroxy β-methylglutaryl-CoA reductase
HPLC	High-performance liquid chromatography
HUVEC	Human umbilical vein endothelial cell
ID	Insert domain
IGF	Insulin-like growth factor
IGF1	Insulin-like growth factor 1
IGF1R	Insulin-like growth factor 1 receptor
IGF2R	Insulin-like growth factor 2 receptor
IRS	Insulin receptor substrate
IRS1	Insulin receptor substrate 1
K_{ATP} (channel)	ATP-sensitive potassium (channel)
KFC	Knowledge-based FADE and contacts
L-NAME	N omega-Nitro-L-arginine methyl ester hydrochloride
L1	First leucine rich domain
L2	Second leucine rich domain
LCMS	Liquid chromatography – mass spectrometry
LDL	Low-Density Lipoprotein
MAPK	Mitogen-activated protein kinase
MCCB	Medicinal chemistry chemical biology
MEK	Mitogen-activated protein kinase/ extracellular signal-regulated kinase kinase
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NMR	Nuclear magnetic resonance
NO	Nitric oxide
PAE	Predicted aligned error
PAIN	Pan-assay interference
PDB	Protein Data Bank
PEI	Polyethylenimine
PI3K	Phosphoinositide 3-kinase
PIP_2	Phospholipid phosphatidylinositol (4,5)-bisphosphate

PIP3	Phosphatidylinositol (3,4,5)-triphosphate
pLDDT	Predicted local distance difference test
PPAR-ɣ	Peroxisome-proliferator activated receptor ɣ
PPI	Protein-protein interaction
Ras	Ras GTPase
RAF (kinase)	Rapidly accelerated fibrosarcoma (kinase)
RAGE	Receptor for advanced glycation end products
RET	Resonance energy transfer
RNA	Ribonucleic acid
ROS	Reactive Oxygen Species
RT qPCR	Reverse transcription quantitative polymerase chain-reaction
SH2 (domain)	SRC homology 2 (domain)
SH3 (domain)	SRC homology 3 (domain)
SOS	Son of Sevenless
Structure-activity relationship	SAR
TACOS	Template-based assembly of complex structures
TKD	Tyrosine kinase domain
TMD	Transmembrane domain
TNF-α	Tumour necrosis factor α
vHTS	Virtual high-throughput screening
WT	Wild-type
XRCD	X-ray crystal diffraction
YPET	Yellow fluorescent protein

1 Introduction

1.1 Cardiovascular Disease and Diabetes

1.1.1 Cardiovascular Disease

Cardiovascular diseases (CVDs) are the leading cause of death globally, representing a major public health challenge[2]. These illnesses affect the heart and vasculature, encompassing a wide range of conditions including heart disease, stroke, and peripheral arteriole disease, each with unique causes and risk factors. The impact of CVDs on quality of life is significant, leading to substantial morbidity, disability, and healthcare expenses. Moreover, the global prevalence of CVDs is increasing, driven by factors such as aging populations, sedentary lifestyles and unhealthy diets[3]. Despite significant progress in the treatment of cardiovascular diseases[3, 4], these diseases remain incurable. The growing healthcare burden of CVDs highlights the urgent need for more effective prevention, diagnosis, and treatment of these conditions as a public health priority.

1.1.2 Atherosclerosis

A common pathology of cardiovascular diseases is the development of atherosclerosis. This inflammatory disorder of the vasculature manifests in its latter stages as lesions of the arterial wall and the formation of atherosclerotic plaques[1]. The rupturing of such plaques can directly lead to acute life-threatening thrombotic events such as myocardial infarction, where the blood supply to the heart is compromised. Additionally, reduced function of atherosclerotic blood vessels perpetuates several chronic conditions such as coronary heart disease in which reduced blood flow affects the function of the heart[1]. Severe atherosclerosis can

lead to lifelong disability and a host of clinical sequalae dependant on the organ affected by ischaemia[1].
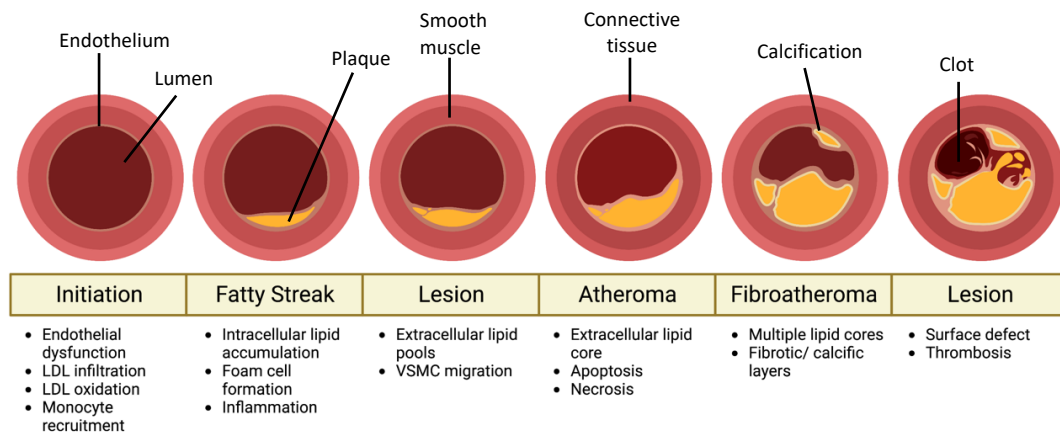
Figure 1 - Schematic overview depicting the progression of atherosclerosis. Each diagram depicts an arteriole cross section with atherosclerosis progressing from left to right. Created with BioRender.com.

The causes of atherosclerosis are multifaceted and not currently wholly understood. Atherosclerosis is an inflammatory response by the endothelial cells of blood vessel walls in response to retained low-density lipoprotein particles (LDL) (Figure 1)[5]. Endothelial cells are responsible for the transcytosis of LDL from the bloodstream to the subendothelial space. Accumulation and subsequent oxidation of LDL leads to an inflammatory response in the blood vessels[6]. It is not clear whether this accumulation of LDL is the cause or the result of such inflammation. Lipid engulfment by recruited macrophages produces foam cells, which subsequently undergo apoptosis and form the lipid core of an atherosclerotic lesion. Atherosclerotic lesions calcify over time, increasing their propensity to rupture. The subsequent release of proinflammatory cytokines leads to further local inflammation and augments the plaque progressions. The major pro-atherosclerotic stimuli are oxidised LDL, proinflammatory cytokines and perturbed blood flow[7]. As such, major atherosclerotic risk factors include hypercholesterolemia, hypertension, and chronic inflammation. Additionally, diseases such as diabetes mellitus promote atherosclerosis by instigating endothelial cell dysfunction[8].

Endothelial cell dysfunction is both a symptom and an early-stage driver of atherosclerosis[8-10]. The endothelial cells lining the blood vessels comprise the vascular endothelium. As the interface between the bloodstream and underlying tissue, the vascular endothelium performs a variety of biochemical signalling roles

to maintain vascular homeostasis[5]. The vascular endothelium regulates vascular tone and permeability, coordinates the vascular inflammatory response, and induces angiogenesis. These processes are mediated by complex crosstalk with endothelial cells and the surrounding cell types, including smooth muscle cells, monocytes, and macrophages. Intuitively, dysregulation of endothelial cell signalling leads to vascular dysfunction and several pathologies. The accumulation of LDL particles and increased recruitment of immune cells characteristic of atherosclerosis appears to depend on preceding endothelial cell dysfunction[8-10].

Of the signalling pathways mediated by the vascular endothelium, the release of nitric oxide (NO) is perhaps the most important in promoting robust vascular function[11]. This small, permeable signalling molecule diffuses across cell membranes to modulate surrounding endothelial cells, smooth muscle cells and leukocytes. NO signalling has been shown to supress endothelial cell inflammation and regulate vascular tone through the activation of soluble guanylate cyclase[6, 11]. NO production is promoted by laminar flow conditions[12], which helps to rationalise the detrimental effects of turbulent or high-pressure blood flow on the vasculature. Additionally, NO can be released in response to upstream signalling by various hormones.

Currently, there are no treatments able to consistently reverse the formation of atherosclerotic plaques, and medical intervention focusses on placating the development of atherosclerosis or treatment of the symptoms. The most widely used group of medications used to treat atherosclerosis are statins, which reduce blood cholesterol and stabilise the atherosclerotic plaque by inhibiting the rate limiting enzyme in the biosynthesis of cholesterol: β-Hydroxy β-methylglutaryl-CoA reductase (HMG-CoA reductase)[13]. Additionally, atherosclerosis can be treated with medication to reduce blood pressure or medications that reduce clotting. Highly

advanced atherosclerosis may require surgical treatment such as angioplasty to reopen the blood-vessel and re-establish blood flow[14].

1.1.3   Diabetes

Diabetes is a metabolic disorder characterised by deficiencies in insulin secretion or insulin activity in the body. Diabetes is categorised into type 1 and type 2: type 1 diabetes results from an autoimmune disorder targeting the pancreatic β cells which impairs their production of insulin[15]; type 2 diabetes results from a decline in whole body insulin sensitivity or a reduction in β-cell function, usually in response to chronic hyperinsulinemia[16] (Figure 2). Both type 1 and type 2 diabetes manifest as persistent hyperglycaemia if unmanaged, which progressively leads to chronic vascular and neuropathic complications. The increasing worldwide prevalence of type 2 diabetes is of particular concern, as this necessitates long term management of blood glucose to avoid resulting cardiovascular complications, resulting in a rising global healthcare burden.
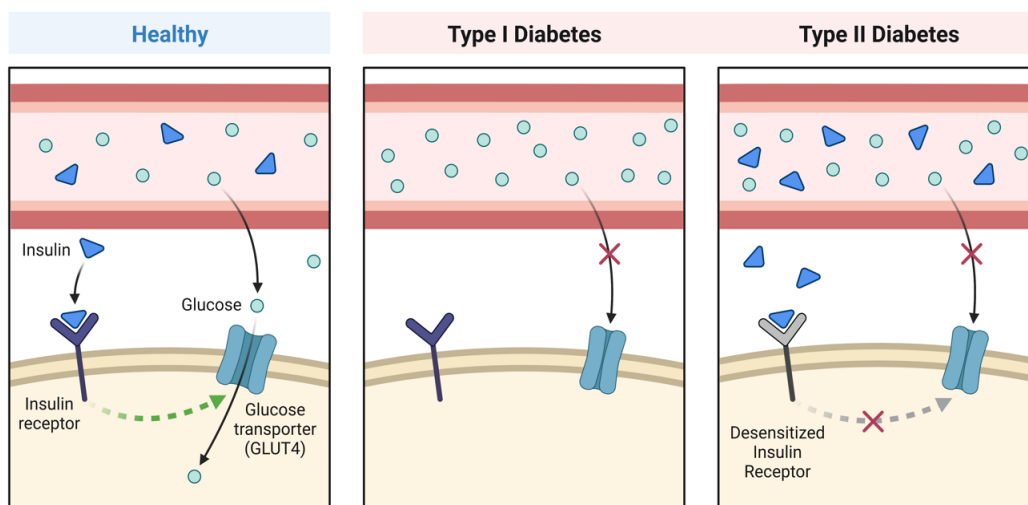
Figure 2 - Schematic representation of Type I vs Type II diabetes. In a healthy individual, binding of insulin to the insulin receptor in adipose and skeletal muscle tissue results in the facilitated diffusion of glucose from the bloodstream into cells via the GLUT4 transporter. Glucose is then utilised for glycolysis or stored as glycogen. In type 1 diabetes, a lack of insulin in the bloodstream prevents activation of the GLUT4 transporter. In type 2 diabetes, reduced cell insulin sensitivity prevents activation of the GLUT4 transporter. Therefore, both pathologies result in a reduction of facilitated diffusion glucose into the cell via the GLUT4 transporter, which can manifest as chronic hyperglycaemia. Created with BioRender.com.

A key vascular consequence of diabetes is the promotion of atherosclerosis[17]. The link between diabetes and atherosclerosis is well-established but not wholly

understood. Known mechanisms connecting diabetes and atherosclerosis include dyslipidaemia, increased oxidative stress, inflammation, and the production of advanced glycation end products (Figure 3)[17].
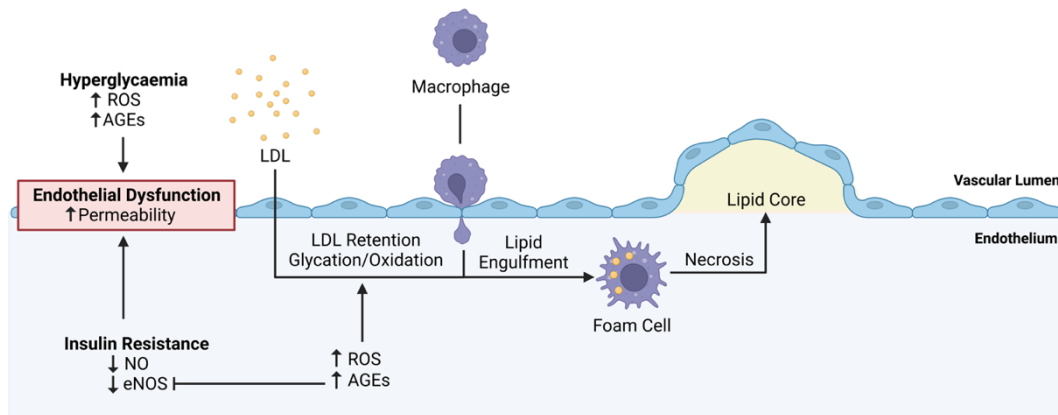
Figure 3 - The mechanistic links between Type 2 diabetes and atherosclerosis, detailing the species and signalling implicated in plaque formation. Chronic hyperglycaemia promotes the formation of ROS and AGEs, whilst insulin resistance reduces NO bioavailability. The resulting endothelial dysfunction leads to an increase in endothelial permeability, promoting LDL retention in the vascular endothelium. The resulting inflammation leads to the recruitment of macrophages, which subsequently engulf oxidised LDL to form foam cells. Foam cells undergo necrosis to form the bulk of a lipid core in an atherosclerotic lesion [1]. Created with BioRender.com.

Dyslipidaemia relates to the altered processing and accumulation of LDL particles that affects atherosclerosis. It is proposed that several modifications of LDL such as glycation and oxidation are promoted in diabetes[18]. Modified LDL has increased retention in the vascular wall and reduced affinity for the LDL receptor. This promotes LDL uptake by lesion cells through non-specific phagocytosis and subsequent incorrect intracellular processing, which promotes the formation of foam cells[19]. LDL extracted from type 1 diabetic patients and injected into the arteriole wall of diabetic mice had four times the retention time of LDL extracted from healthy individuals, highlighting the effects of diabetic LDL modifications on LDL retention[20]. Similarly, patients with type two diabetes have prolonged intravascular LDL residence time , which can be reduced by insulin treatment[21].

Diabetes is also associated with increase oxidative stress on the vascular endothelium. Oxidative stress is caused by the imbalance of oxidative and reducing processes inside the cell, resulting in the accumulation of reactive oxygen species (ROS). The term ROS typically refers to any of superoxide radicals, hydroxide radicals, singlet oxygen species and hydrogen peroxide, all of which are generated as metabolic by-products[22]. The oxidation of LDL by accumulated ROS represents a

distinct pathological pathway of LDL modification from the pathways found in dyslipidaemia. It is well established that chronic hyperglycaemia results in increased oxidative stress on the vascular endothelium[23].

Additionally, the formation of advanced glycation end (AGE) products in diabetes is thought to promote atherosclerosis. AGE compounds are derived from glucose and oxidised fatty acids which tend to accumulate both within endothelial cells and in the sub-endothelial space in response to hyperglycaemia. The modification of intracellular proteins by AGE products leads to altered protein function. Additionally, the modification of extracellular matrix proteins by AGE products promotes the interaction of these products with the receptor for advanced glycation end products (RAGE) on macrophages and endothelial cells. These interactions promote pro-inflammatory effects and the generation of intracellular ROS[24, 25].

1.1.4   Treatments for Diabetes

Treatment approaches differ according to the type of diabetes. Type 1 diabetes treatment focusses on insulin replacement therapy, with patients typically utilize a combination of fast-acting and long-acting insulins to regulate glucose levels and achieve glucose homeostasis. Additionally, medications such as aspirin or

cholesterol-lowering drugs such as statins may be prescribed to protect against the cardiovascular complications associated with type 1 diabetes[15].

Treatment of type 2 diabetes focusses on regulating blood sugar levels through various strategies. Initially, efforts are directed towards controlling blood sugar levels through modifications in exercise and dietary regimens. However, if lifestyle changes do not lead to adequate regulation in blood glucose concentrations, pharmacological assistance may be required. As patients typically still produce insulin, insulin replacement therapy is not usually necessary. Instead, pharmacological treatments involve enhancing insulin secretion, improving insulin sensitivity of the target tissues, and reducing glucose production. Medications used to treat type 2 diabetes aim to achieve these goals and help individuals maintain better glycaemic control (Figure 4)[26].
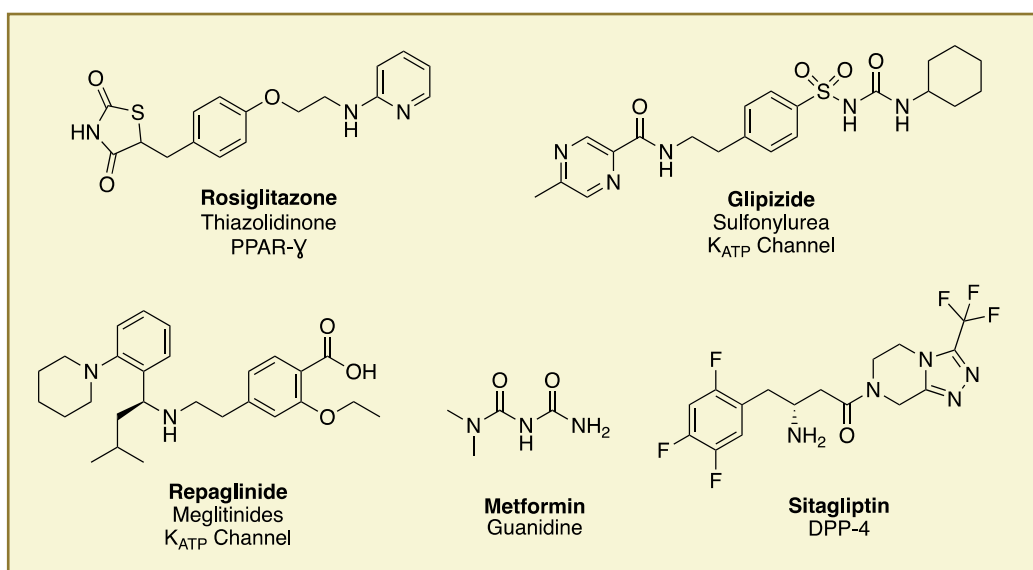
Figure 4 - Selected drugs licensed for the treatment of Type-2 diabetes, with the drug class and molecular target indicated.

The first line treatment for type 2 diabetes is metformin (Figure 4), which exerts its effects through several mechanisms. This orally bioavailable biguanide reduces gluconeogenesis in the liver, increases the insulin sensitivity of many tissues and supresses appetite. While many guanidine-containing compounds display anti hyperglycaemic affects, metformin uniquely decouples efficacy from toxicity[27]. Despite its wide usage, metformin's mechanism of action is poorly understood and is unlikely to depend on a specific protein target. Proposed mechanisms include uncoupling of mitochondrial electron transport[28] or the activation of AMPK (5'

adenosine monophosphate-activated protein kinase), a crucial cellular energy sensor[29]. Unfortunately, metformin shows reduced efficacy over time and cannot halt the progressive decline of pancreatic β-cell function in type 2 diabetes, meaning it cannot reverse the progress of the disease[26].

Additional insulin secretagogues include the sulfonylureas and meglitinides (Figure 4). Sulfonylureas simulate the release of insulin by inhibiting the ATP-sensitive potassium (K_{ATP}) channel in the pancreatic β-cells[30]. Inhibition of K_{ATP} channels increases intracellular Ca^{2+} concentrations, promoting membrane fusion and release of insulin granules. Due to their direct action on insulin secretion, sulfonylureas pose a higher risk of hypoglycaemic events compared to metformin[31]. Meglitinides are a newer class of insulin secretagogues, which similarly bind to K_{ATP} despite being structurally discrete to sulfonylureas[32]. Inhibition of K_{ATP} is faster with meglitinides relative to sulfonylureas, as is the reversal of inhibition. This shorter duration of action reduces the risk of hypoglycaemia compared to sulfonylureas, although this is still relatively high compared to other classes of treatment. As with sulfonylureas and metformin, meglitinides do not prevent the progressive decline of β-cell function associated with type 2 diabetes[26].

Thiazolidines (Figure 4) represent an alternative class of insulin senstizers[33] and are capable of reversing β-cell deterioration in type 2 diabetes[34]. These drugs work by binding to peroxisome-proliferator activated receptor ɣ (PPAR-ɣ). This promotes a gene expression pattern resulting in increased fatty acid storage in adipocytes. The net effect of this is to both increase cellular glucose utilization and decrease glucose production. Additionally, insulin sensitivity is improved, likely through the altered secretion of hormones from adipocytes. Similarly mediated through PPAR-ɣ, thiazolidines increase the production of insulin enhancing factors such as adiponectin, whilst reducing the secretion of substances that impair insulin function such as tumour necrosis factor α TNF-α[35]. However, thiazolidines have fallen out of favour due to associations with increased risk of myocardial infarction[36] and bladder cancer[37].

Finally, glucagon-like peptide (GLP1) modulation has been targeted as a means of treatment for type-2 diabetes. GLP1 is secreted from the intestine after eating, and

results in glucose dependant insulin secretion from pancreatic beta cells. Dipeptidyl-peptidase 4 (DPPIV) is a cell surface protease responsible for the degradation of GLP1[38]. Blockade of DPPIV prolongs the half-life of GLP1, promoting insulin secretion. However, a recent meta-analysis found no favourable effects of DPPIV treatment on all-cause or cardiovascular mortality compared to placebo[39]. Peptide GLP-1 agonists, which require subcutaneous injection, showed favourable outcomes in all-cause mortality and cardiovascular events relative to placebo[39]. Promising small-molecule alternatives are currently in development[40], but have not yet reached the market.

In summary, of the current medications for type 2 diabetes, metformin is viewed as the safest and most efficacious. The management of type 2 diabetes is typically monitored through assessment of blood glucose levels, and most classes of medications treating type 2 diabetes aim to improve insulin sensitivity, increase insulin production, or reduce glucose synthesis. The main problem associated with current medications is owed to the chronic and progressive nature of type 2 diabetes; medications tend to become less effective over time, particularly as the disease progresses, and most do not prevent β-cell deterioration. There is a need for novel therapies that blunt disease progression to provide durable disease control. Such medications should target the pathogenic disturbances that progress the disease and its associated complications, particularly insulin sensitivity and endothelial dysfunction. Consequently, the identification and validation of novel therapeutic targets able to persistently restore insulin sensitivity and rectify endothelial function is of high priority.

1.2 The Insulin Receptor and IGF1 Receptor

1.2.1 Insulin Signalling

Type 2 Diabetes is effectively the result of dysregulated insulin signalling. The hormone insulin is released from pancreatic β cells in response to elevated blood glucose levels. Insulin acts primarily through binding to the insulin receptor (IR), a receptor tyrosine kinase (RTK) present in the cell membrane of various tissues. The canonical tissues targeted by insulin are the skeletal muscle, adipose tissue, and liver. Insulin signalling promotes glucose uptake, protein synthesis and lipid

synthesis in these tissues. Increasingly in the context of cardiovascular disease, the direct action of insulin on the endothelial cells of the vascular endothelium is acknowledged to play a role in robust endothelial function.
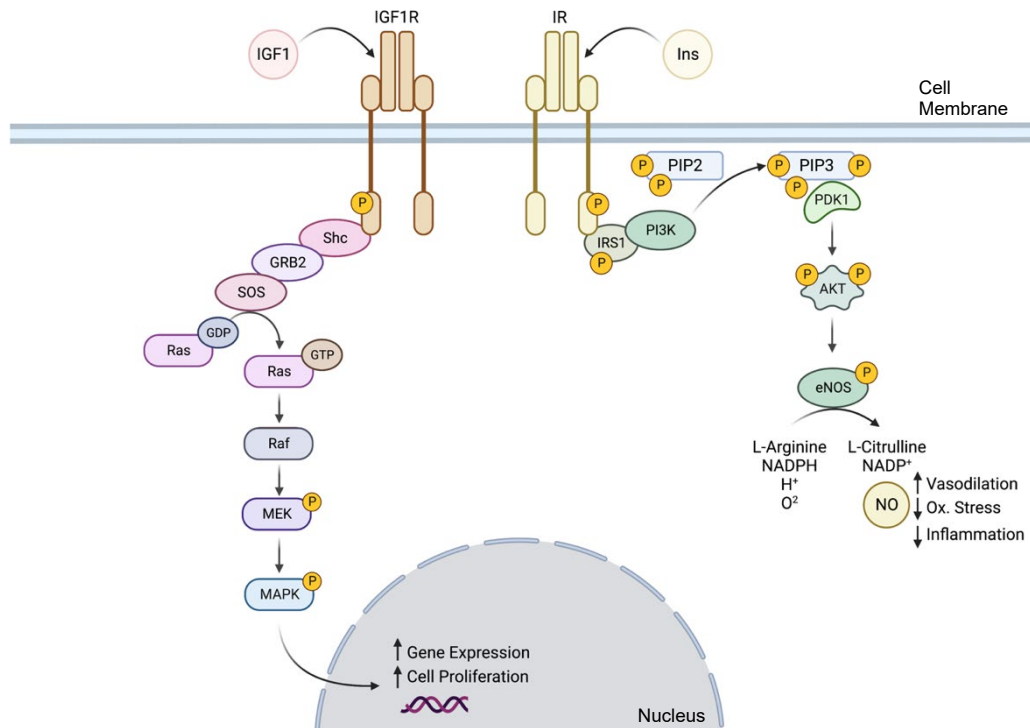
Figure 5 - Schematic overview of insulin and IGF1 intracellular signalling in the vascular endothelium. The AKT/PI3K (right) and Ras/MAPK (left) pathways are shown. Yellow circles indicate phosphorylation sites. Created with BioRender.com.

Upon binding to the insulin receptor, insulin signals primarily through the protein kinase B (Akt) / phosphoinositide 3-kinase (PI3K) pathway (Figure 5). The binding of insulin induces a conformational change in the IR and the activation of two intracellular receptor tyrosine kinase domains. The receptor is then able to bind to and phosphorylate multiple intracellular adaptor proteins such as the insulin receptor substrates (IRSs). Phosphorylated insulin receptor substrate 1 (IRS1) has multiple binding sites for proteins containing the SRC homology 2 (SH2) domain, such as PI3K. The binding of pIRS1 leads to increased catalytic activity of PI3K, which converts the membrane phospholipid phosphatidylinositol (4,5)-bisphosphate (PIP$_2$) to phosphatidylinositol (3,4,5)-triphosphate (PIP$_3$). Phosphorylated PIP$_3$ is recognised by Akt and PDK1, resulting in the phosphorylation of Akt by PDK-1. In the vascular endothelium, this culminates in the activation of endothelial nitric oxide synthase (eNOS) by AKT. Phosphorylated eNOS synthesises NO from L-

arginine and reduced nicotinamide adenine dinucleotide phosphate (NADPH)[41].
Several studies have demonstrated that insulin resistance results in reduced
bioavailability of NO[42-44] and promotes endothelial dysfunction[45].

1.2.2   IGF Signalling

In addition to insulin, the signalling of the related insulin-like growth factors (IGFs)
is important in modulating endothelial cell function. This signalling system is
comprised of the hormones IGF1 and IGF2, the IGF1 receptor (IGF1R) and IGF2
receptor (IGF2R) as well as six IGF binding proteins[46]. IGFs play an essential role in
cell-growth and proliferation[47]. Additionally, IGFs are believed to have
atheroprotective affects. In particular, low serum levels of IGF1 are associated with
increased risk of ischaemic heart disease[48].

IGF1 is a major growth factor in human adults and is released by the liver in
response to the action of growth hormone. IGF1 binds to the IGF1R with high
affinity. Upon ligand binding, the activated IGF1R signals primarily via the Ras (Ras
GTPase)/ mitogen-activated protein kinase (MAPK) pathway (Figure 5). The IGF1R
intracellular region interacts with Shc, which in turn is bound by growth factor
receptor bound protein 2 (GRB2) through its SH2 domain. GRB2 binds and activates
guanine nucleotide exchange factor Son of Sevenless (SOS) through two SRC
homology 3 (SH3) domains. Activated SOS then removes guanosine diphosphate
(GDP) from Ras, which can then bind guanosine triphosphate (GTP) and become
activated. This initiates a kinase cascade, in which rapidly accelerated fibrosarcoma
(RAF) kinase, MAPK/ERK kinases (MEKs) and MAPK are sequentially activated.
Ultimately, this leads to the phosphorylation of transcriptional activators by MAPK,
inducing the long-term mitogenic results of IGFI signalling[47].

As the name suggests, the IGF system displays significant crosstalk with the insulin
signalling system, despite exerting distinct physiological effects. The downstream
signalling of the IR and IGF1R overlaps significantly, with both activated receptors
able to signal through IRS and Shc. Mutagenesis suggests that a single residue in the
intracellular region of the receptors is a key determinant of these specificities[49].
Extracellular signalling is unlikely to significantly contribute to receptor crosstalk.
Insulin and IGF1 bind to each other's receptors with 100- to 1000-fold lower affinity

than their own receptor. Due to the physiological concentrations of each ligand, it is likely they only activate their own cognate receptor *in vivo*[50].

IGF1 signalling appears to have a beneficial effect on endothelial cell function through anti-inflammatory action[51] and stimulation of angiogenesis[52]. Additionally, IGF1 stimulation on the endothelium has been shown to elicit NO production. Stimulation of Human umbilical vein endothelial cell (HUVEC) cells with IGF1 gave rise to NO production at physiological concentrations (9-30 nM), a concentration at which IGF1 negligibly activates the IR. IGF1 stimulated NO production is inhibited by treatment with N omega-Nitro-L-arginine methyl ester hydrochloride (L-NAME), an eNOS competitive inhibitor, αIR-3, an IGF1R antibody, and wortmannin, a PI3K inhibitor[53]. This confirms that IGF1 signalling through PI3K and eNOS results in the release of the vasoprotective factor NO. However, it does appear that the IR is a more potent stimulator of NO production. Maximal stimulation of HUVEC cells with IGF1 shows only 40% NO production relative to the maximal response on the same cells with insulin[53]. This is presumably due to the preference for the IR to signal through the PI3K/AKT pathway relative to the IGF1R.

1.2.3   Genetics and Biosynthesis of the IR and IGF1R

The IR and IGF1R share similar structure. Each protein is encoded by the INSR and IGF1R genes of 1382 and 1367 amino acids respectively. The insulin receptor has two alternative splice variants: the canonical IR-B which includes the 12 amino acids of exon 11 upstream of a tetrabasic RKRR furin proteolytic cleavage site and the shorter IR-A isoform which excludes exon 11[54, 55]. Additionally, both contain a 30-residue signal peptide which is removed during translocation to the cell membrane. Both the IR and IGF1R are synthesised as single chain pro-receptors, and subsequently undergo significant posttranslational processing including co-translational cleavage, glycosylation, and dimerisation in the endoplasmic reticulum[55, 56]. Subsequently, the receptors are transported to the Golgi-apparatus
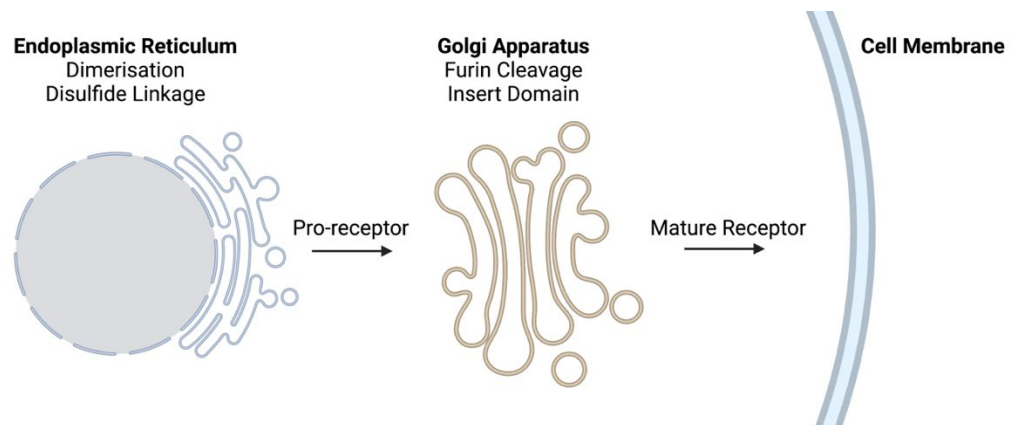
Figure 6 - The sequence and locations of receptor post-translational processing in the biosynthesis of the IR and IGF1R. The IR and IGF1R are initially dimerised as a single chain pro-receptor in the endoplasmic reticulum, before furin cleavage and addition of the insert domain in the golgi apparatus and subsequent translocation of the mature receptor to the cell membrane. Created with BioRender.com.

where they are cleaved by furin into their characteristic α and β chains, with a none-globular insert domain added between the α and β subunits prior to membrane insertion (Figure 6)[55].

Unless otherwise stated, all numbering and naming in this thesis refers to the canonical IR-B including signal sequence.
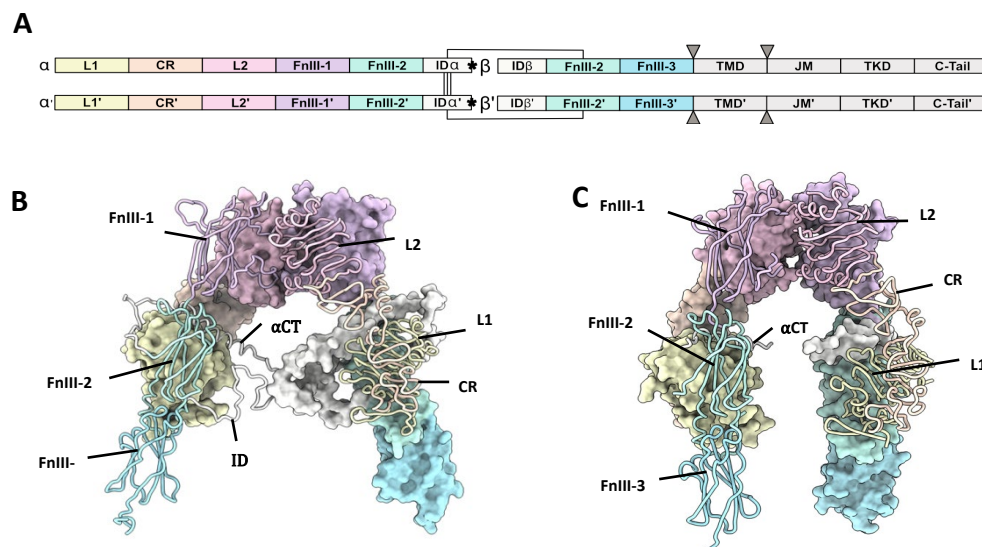
Figure 7 - (A) Domain organisation of the IR and IGF1R. Monomers have been denoted as monomer and monomer' for clarity. The position of the αCT helix is denoted with asterixes (*). The position of the cell membrane has been indicated with triangles (▼); (B) XRCD apo-structure of the IR ectodomain (PDB:4ZXB). One monomer is shown with a space filling representation and the other with a ribbon representation; (C) XRCD apo-structure of the IGF1R ectodomain (PDB:5U8R). One monomer is shown with a space filling representation and the other with a ribbon representation. Note - the ID is not resolved in the IGF1R structure. Reproduced from Ref. 59 with permission from the Royal Society of Chemistry.

The IR and IGF1R show similar domain organisation (Figure 7). The extracellular portion of each receptor monomer contains two leucine rich domains (L1 and L2), a cysteine rich domain (CR), and three fibronectin type III domains (FnIII-1, FnIII-2,

and FnIII-3). The FnIII-2 domain contains the insert domain (ID) spanning the α and β subunits. The ID contains a key ligand binding region at the C-terminal end of the α-subunit termed the α C-terminal helix (αCT). The remainder of the β subunit contains a single-pass helical transmembrane domain (TMD), and an intracellular tyrosine kinase domain (TKD) flanked by regulatory juxta membrane and C-tail regions. Three disulfide bonds linking the αβ monomers are contained in the ID, whilst the single disulfide linking the α and β subunits is in the FnIII-2 domain (C534, C682, C683 and C685 respectively).

Dimerisation appears to be pivotal in the robust processing of the IR. By mutating the cysteine amino acids responsible for interchain disulfide bonds, monomeric insulin receptors have been generated[57]. However, these monomeric insulin receptors were transported to the membrane with approximately half the efficiency of the wild-type dimeric receptors. Surprisingly, the mutant receptor could still form non-covalent dimers, both with and without the addition of insulin, indicating that covalent linkage is not required for dimerisation. Notably, when the same cysteine residues were mutated to arginine or aspartic acid, which introduces repulsive charge between each monomer, expression at the membrane reduced to approximately 10% relative to the wild-type receptor[58].

1.3 Structures of the IR and IGF1R

1.3.1 Structures of the apo-IR and apo-IGF1R Ectodomain

The structures of the IR and IGF1R have been characterised by a combination of X-ray crystal diffraction (XRCD) and cryo-electron microscopy (cryo-EM), revealing the structural basis of receptor function[59]. Apo-structures for both the IR[60] (PDB: 4ZXB, Figure 7A) and IGF1R[61] (PDB: 5U8R, Figure 7B) ectodomain have been determined by X-ray crystallography to 3.3 Å and 3.0 Å respectively. Both show a characteristic inverted V-shape with each αβ monomer arranged in an antiparallel fashion. The first leg of the V is formed by the L1, CR and L2 domains, which pack against the second leg formed by the FnIII-1', FnIII-2' and FnIII-3' domains (the dash denoting the domains arise from the alternate monomer). The L2 and FnIII-1 domains comprise the apex of the V. This arrangement places the FnIII-3 domains distal to the membrane. Major intermonomer interfaces occur between the L2 and FnIII-1',

as well as the L1 and FnIII-2' domains. The structures determine that the αCT helix lies across the second β-sheet of the L1', forming the major hormone binding epitope.

1.3.2 Ligand Binding

1.3.2.1 Insulin and IGF1

Similarly, the ligands IGF1 and insulin also show considerable structural homology (Figure 8). The primary difference between the hormones is that IGF1 is composed of a single polypeptide chain (Figure 8 A, C), whilst insulin comprises a two-chain polypeptide connected by disulfide bonds. The structure of IGF1 is divided into four domains: B, C, A and D. Alternatively, insulin is portioned into A and B chains (Figure 8 A,B), corresponding to the A and B domains of IGF1.
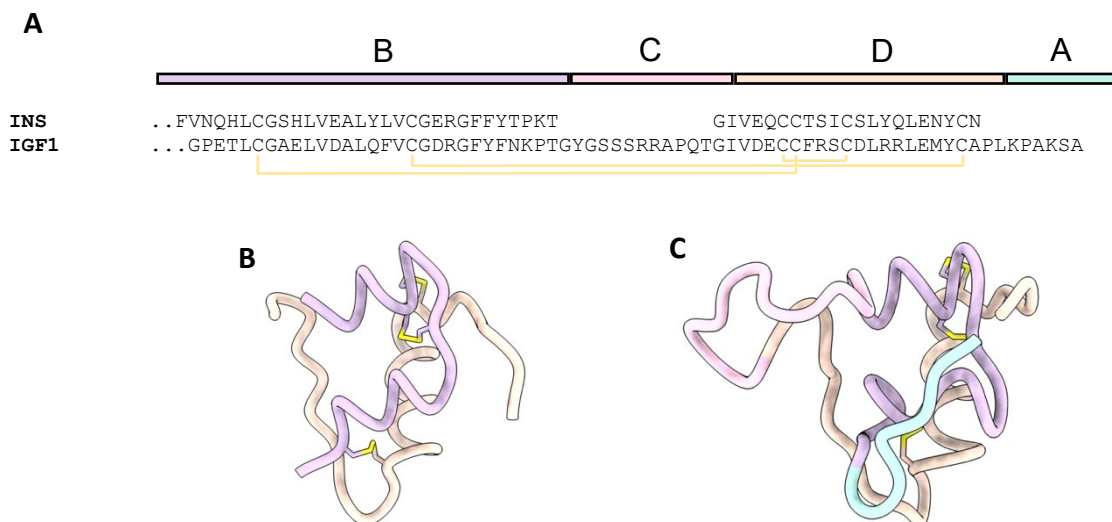
Figure 8 - The structures of free insulin and IGF1. B, C, A and D domains are depicted in purple, pink, orange and cyan respectively. Disulfide bonds are depicted in yellow; A) Sequence alignment of Insulin and IGF1 showing the locations of their relevant domains; B) The structure of insulin (PDB:4INS); C) The structure of IGF1 (PDB:2GF1). Reproduced from Ref. 59 with permission from the Royal Society of Chemistry.

The primary structural features of the hormones are three conserved α-helices: two antiparallel α-helices in the A domains connected by a short linker, and a single α-helix in the B domain. The C-domain is a disordered strand which links the B and A chains of IGF1 but is absent in insulin. The short, disordered D-domain extends from the A-domain C-terminus in IGF1. The tertiary structure of the hormones is

preserved by three disulfide bonds: two interchain disulfides between the A and B domains and a single intrachain disulfide in the A domain.

The ligand bound structures of the IR and IGF1R show that the hormone binding modes are largely conserved between the two receptors. These structures determine that the receptor's primary binding epitope is comprised of residues distributed across the αCT, L1' and FnIII-1', which had largely been predicted with prior mutagenesis studies[62-66] (Figure 9). Insulin and IGF1 bind to these epitopes such that the αCT inserts into a hydrophobic crevice between the A and B chains, or A and B domains of IGF1. To reveal this hydrophobic cavity, the insulin residues B24-B30 or analogous residues in the IGF1 C-domains detach from the hormone core upon binding to the receptor. Simultaneously, residues on the opposite side of the hormones contact the adjacent loops of the FnIII-1', such that both monomers of the receptor are contacted simultaneously. This mode of binding had effectively been predicted from kinetic and mutational data in the 'crosslinking model'[67]. In line with previous naming conventions, the receptor epitope comprising the αCT and L1' will hereby be referred to as site 1a, whilst the epitope at the FnIII-1 domain termed site 1b'.
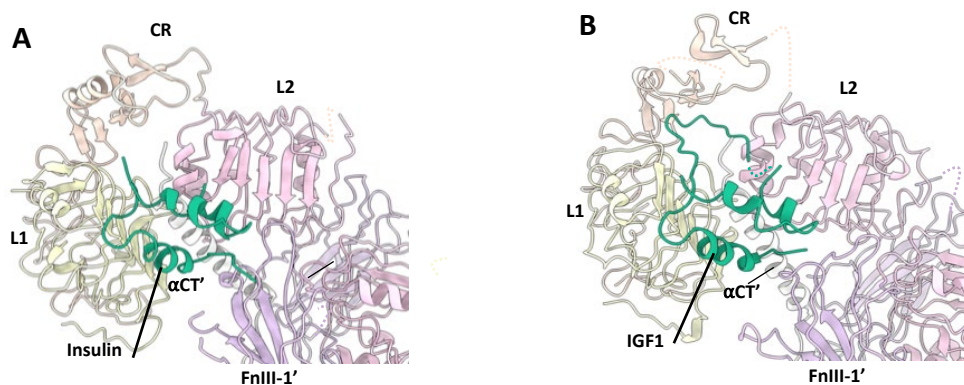
Figure 9 - A comparison of insulin and IGF1 bound to site 1 of the IR and IGF1R. (A) Insulin (green) bound to site 1 of the IR (PDB:6HN5); (B) IGF1 (green) bound to site 1 of the IGF1R (PDB:6PYH). Unresolved regions of the receptors are indicated by dotted lines. Reproduced from Ref. 59 with permission from the Royal Society of Chemistry.

1.3.3 IR Ligand Bound Structures

1.3.3.1 Γ-Shaped Structure

A high-resolution XRCD structure of the IR ECD determined that the receptor adopts a Γ-shape when bound to a single insulin ligand (PDB: 6HN4 & 6HN5, Figure 10B).[68] The construct used to obtain this structure utilises a leucine zipper fused to the
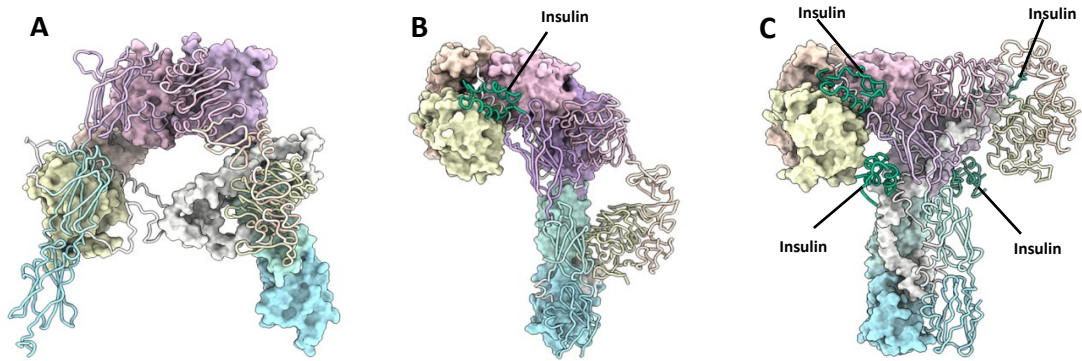
Figure 10 - (A) XRCD structure of the apo-IR ectodomain (PDB:4ZXB); (B) Cryo-EM structure of the Γ-Shaped IR ectodomain bound to a single insulin (PDB:6HN4 and 6HN5); (C) Cryo-EM structure of the T-shaped IR ectodomain bound to four insulin (PDB:6SOF). One monomer in each structure is depicted with a molecular surface and the other in cartoon format. Insulin is depicted in green. Reproduced from Ref. 59 with permission from the Royal Society of Chemistry.

FnIII-3 domains to restore negative cooperativity to the soluble IR ECD, which may have been essential in achieving a singularly bound structure. Low resolution negative-stain electron microscopy images of full-length IRs embedded into lipid nanodiscs support that this conformation is relevant in a membrane bound context.[69]

Comparison of the Γ-shaped structure to the V-shaped apo-structure reveals ligand binding affects large scale conformational rearrangements in the receptor.[68] The L1/FnIII-2' interface on the ligand bound monomer is completely disrupted. Consequently, the L1, CR, L2 and αCT' domains are displaced upwards towards the top of the receptor, whilst the FnIII domains come together, forming the head region and the perpendicular stalk of the Γ-shape respectively. These rearrangements can be described by two rigid-body rotations, with the L1-CR-L2 domains rotating approximately 35° with respect to the L2-FnIII-1 interdomain linker, and a second a rotation of the L1-CR domains by 55° with respect to the CR-L2 linker.

1.3.3.2 T-Shaped Structures

Several structures detail the IR bound to two or four insulin ligands adopting a T-shaped structure. The first evidence of the IR adopting this topology was found using negative stain EM, imaging the full length IR reconstituted into lipid nanodiscs.[69] Whilst this study did not provide atomic level detail, it confirms the relevance of the V-shaped, Γ-shaped and T-shaped IR structures in a membrane

bound context. A subsequent cryo-EM structure shows the IR-ECD bound to two insulin ligands adopting this conformation, which after applying C_2 symmetry during refinement, achieved a resolution of 4.3 Å (PDB: 6CE9).[70] Whilst large parts of the FnIII domains are not resolved in this structure, the modelled domains adopt a T-shape, with the ligand binding regions well resolved. Two further cryo-EM structures detail the receptor displaying a T-shaped structure, despite being bound to four insulin ligands. The first structure achieved a global resolution of 4.3 Å (PDB: 6SOF, Figure 10C),[69] whilst the second enforced C_2 symmetry to produce a higher resolution of 3.2 Å (PDB: 6PXW), at the expense of masking asymmetries exhibited in the T-shaped dimer.[71]

Analysis of the T-shaped receptor structures show the conformational changes with respect to the apo-receptor are equivalent to those seen in the Γ-shaped structure but occur in both monomers. Therefore, the head of the T-shaped receptor is comprised of the L1, CR, L2 and αCT domains of both monomers, whilst the FnIII domains associate and extend perpendicular to the centre of the head region.

It is notable that all T-shaped IR structure samples were prepared under supraphysiological insulin concentrations (28 μM,[70] 40 μM,[72] 4:1 molar ratio

insulin: IR [71]). Therefore, it is possible the IR may not bind two or more ligands at physiological insulin concentrations of up to 5 nM[73].

1.3.4   IGF1R Ligand Bound Structures

1.3.4.1   Γ-Shaped Structures

Six ligand bound structures of IGF1R ectodomain currently exist. Of these, five adopt a Γ-shaped conformation with a single bound ligand. The first Γ-shaped structure is a cryo-EM structure of full-length mouse IGF1R bound to IGF1 at 4.3 Å resolution[74] (PDB:6PYH, Figure 11B). Due to poor resolution of the TMDs and kinase domains, atomic models were only built for the receptor ectodomain. A second cryo-EM structure has been determined of human holo-IGF1R bound to IGF1, albeit to a lower resolution of 7.7 Å.[75, 76] The low resolution of the map prohibited 3D model building of the complex. The same authors also report a cryo-EM structure of human holo-IGF1R bound to insulin at 4.7 Å resolution[75] (PDB:6JK8). Once again, limited resolution of the transmembrane and kinase domains meant only the receptor ectodomain and ligand could be modelled.
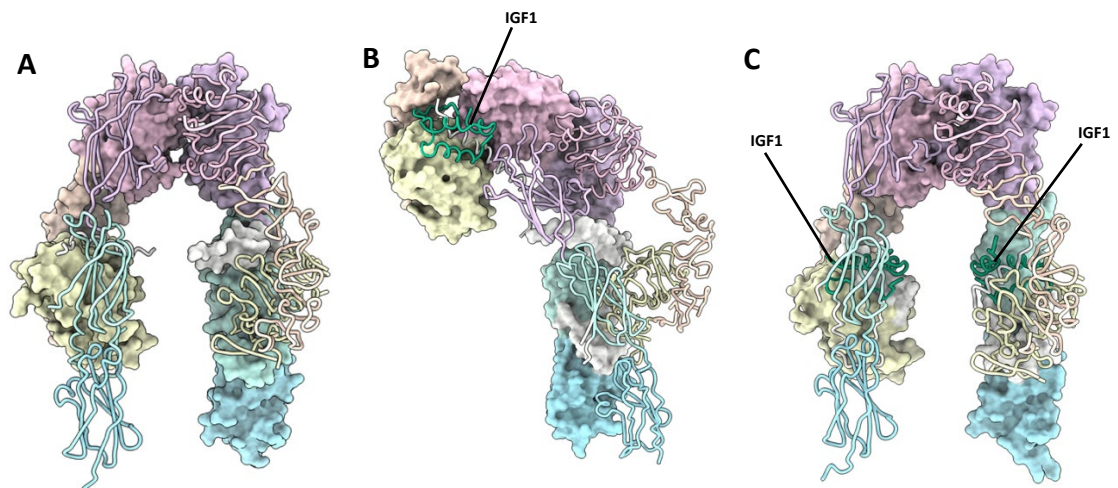
Figure 11 - (A) XRCD structure of the apo-IGF1R ectodomain (PDB: 5U8R); (B) Crystal structure of the Γ-shaped IGF1R ectodomain bound to a single IGF1 (PDB:6PYH); (C) Crystal structure of the IGF1R ectodomain bound to two IGF1 (PDB:5U8Q). Note – this structure is unlikely to depict a physiologically relevant conformation of the receptor ectodomain. One monomer in each structure is depicted with a molecular surface and the other in cartoon format. IGF1 is depicted in green. Reproduced from Ref. 59 with permission from the Royal Society of Chemistry.

The final ligand bound structure is an XRCD structure of IGF1R bound to two IGF1 ligands determined to 3.3 Å (PDB:5U8Q, Figure 11C).[61] This structure was produced by soaking the crystals used in the IGF1R apo-structure (5U8R) with IGF1. It appears the constraints of the crystal lattice prevented large scale rearrangement upon the

binding of IGF1, with the receptor adopting a conformation intermediate of a V and T-shaped structure. Therefore, whilst this structure isn't likely to represent a physiologically relevant conformation of the IGF1R, it provides a key intermediate conformation as part of a growing portfolio of IGF1R structures.

1.3.4.2 Ligand binding: Site 2

One of the most notable features of the T-shaped IR structures is the presence of four insulin bound to the receptor[71, 72]. Whilst two of the insulin are bound to the classical 1a/1b' and 1a'/1b, two further insulin are bound to symmetry related sites on the FnIII-1 domains. These sites will hereby be referred to as site 2 and 2'. Insulin residues that contact the receptor at site 2 on the opposite side of the hormone to those contacting site 1. As site 2 is more accessible, it may represent the preliminary insulin binding site, prior to insulin transitioning to the higher affinity site 1. This explains why insulin bound to site 2 are only observed with saturating insulin conditions. Insulin bound to this site interact with the L1 domain, and therefore potentially disrupt the L1-FnIII-2' interface present in the apo receptor, supporting a model of induced fit for hormone binding.

No analogous site has yet been observed directly for IGF1R. However, the existence of the IGF1R structure bound to two IGF1[61] (5U8Q, Figure 11C) implies that IGF1 binds to the IGF1R through a mechanism of induced fit. This structure was obtained by crystal soaking of the IGF1R in the apo-conformation, preventing IGF1R from sampling alternative conformations when confined by the crystal lattice. As site 1 is blocked by the FnIII-1 domain in the apo-conformation of the receptors, IGF1 would first have to bind to an accessible site on the receptor, which implies the presence of an analogous site 2.

1.3.5 Structures of the IR and IGF1R Transmembrane and Intracellular Regions
1.3.5.1 Transmembrane Domains

The structure of the IR transmembrane region has been shown by NMR to consist of a single α helix in detergent micelles[77] (PDB: 2MFR, Figure 12A). This is in line with the structure predicted by molecular dynamics simulations of the IR-TM region in a phospholipid membrane,[78] indicating that the structure has not been altered
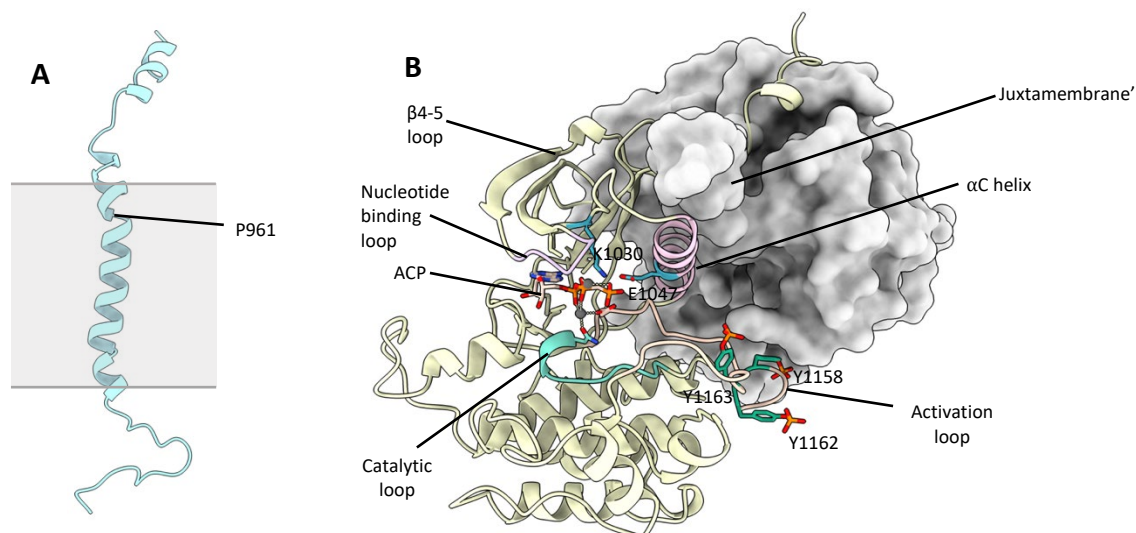
Figure 12 - (A) NMR structure of the IR-TMD (PDB:2MFR). Points of membrane entry are indicated in grey; (B) Crystal structure of the IR-TKD activated dimer (PDB:4XLV). One of the TKDs is depicted in cartoon format and the other with a molecular surface representation. The αC helix (pink), nucleotide binding loop (purple), activation loop (salmon), catalytic loop (cyan) are highlighted. Key residues K1030 and E1047 (blue), as well as phosphorylation sites Y1158, Y1162 and Y1163 (green) are shown. Reproduced from Ref. 59 with permission from the Royal Society of Chemistry.

through incorporation into lipid vesicles. The α-helix contains a single kink at Gly960, owed to the presence of Pro961.

Currently no structure exists of the IGF1R TM domains. However, molecular dynamics simulations have modelled the conformation of the TM regions in a phospholipid bilayer.[78] These simulations show that the IGF1R TM regions adopt a single α-helix with a kink at Pro941, comparable to the IR TMD.

Several lines of evidence point to the IR and IGF1R TMDs dimerising during receptor activation, as is implied by the convergence of the membrane proximal FnIII-3 domains in the ligand bound IR and IGF1R structures.[79] Whilst the IR-TMDs were found to be monomeric in detergent micelles, crosslinking studies indicated that they could dimerise within the detergent micelles depending on the protein/detergent ratio[77]. This is supported by NMR experiments of isolated IR-TMDs, which determine the IR TMD can adopt dimeric conformations in membrane-like environments.[80] Similarly, molecular dynamics simulations and FRET analysis indicate that the IGF1R-TMD regions dimerise upon IGF1R activation.[81]

1.3.5.2 Kinase Domains

Structures for the inactive and phosphorylated (active) states of the IR TKD have been determined by XRCD to 2.1 Å (PDB:1IRK, Figure 13A)[82] and 1.9 Å (PDB:1IR3,
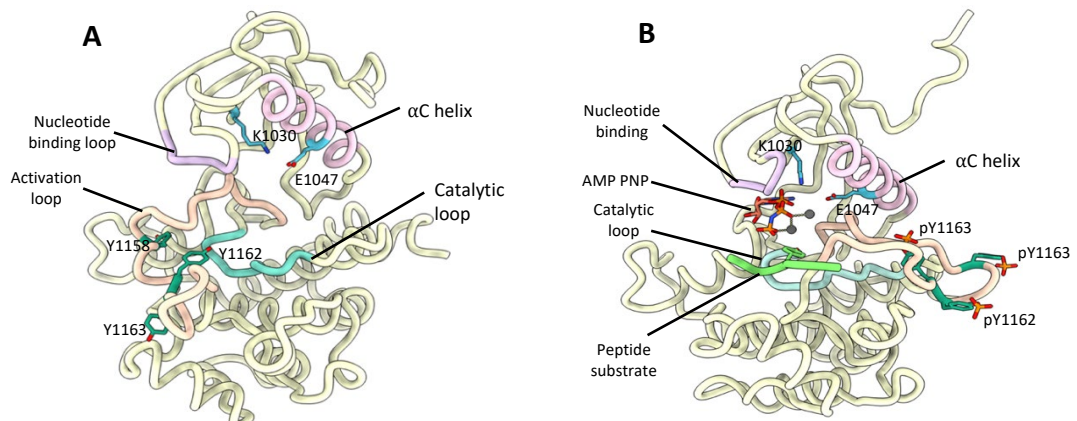
Figure 13 - (A) Crystal structure of the inactivated IR kinase domain (PDB:1IRK) with the αC helix (pink), nucleotide binding loop (purple), activation loop (salmon), catalytic loop (cyan) highlighted. Key residues K1030 and E1047 (blue), as well as phosphorylation sites Y1158, Y1162 and Y1163 (green) are shown; (B) Crystal structure of the phosphorylated (activated) IR kinase domain (PDB:1IR3) with the αC helix (pink), nucleotide binding loop (purple), activation loop (salmon), catalytic loop (cyan) and bound peptide substrate (lime) highlighted. Key residues K1030 and E1047 (blue), as well as phosphorylated residues pY1158, pY1162 and pY1163 (green) are shown. Reproduced from Ref. 59 with permission from the Royal Society of Chemistry.

Figure 13B)[83] respectively. The latter is complexed with a peptide substrate and adenosine triphosphate (ATP) analogue. Similarly, a structure of the IGF1R TKD in its active state (PDB:1K3A)[84] has been determined to 2.1 Å, whilst two structures of the inactive state (PDB:1M7N,[85] 1P40[86]) have been elucidated to 2.7 Å and 1.5 Å respectively. Several structures of the IR and IGF1R kinase domains bound to small-molecule inhibitors have since been published.[87-92]

The IR and IGF1R TKDs show typical kinase topology; comprised of a small N-terminal and larger C-terminal lobe connected by a single linker. The C-terminal lobe encompasses the active site and activation loop, the latter containing three IR tyrosine phosphorylation sites Tyr1158, Tyr1162 and Tyr1163 (Tyr1131, Tyr1135 and Tyr1136 in IGF1R).[84, 93, 94]The activation loop itself traverses the catalytic cleft between the two lobes. In the basal state, Tyr1162 and Tyr1135 occupy their respective receptors active site, causing the activation loop to block access to the ATP binding site. This prevents cis phosphorylation of the catalytic tyrosine residues in the basal state conformation (Figure 13A).

Upon receptor activation, the TKDs affect trans-autophosphorylation of the three catalytic tyrosine residues. To stabilise the phosphorylated tyrosine groups, the activation loop is significantly displaced from its basal state conformation. This

provides unrestricted access to the kinase active site and ATP binding site (Figure 13B).

Additionally, the TKDs of the IR and IGF1R associate in the active state, functioning as a dimeric unit to promote optimal substrate phosphorylation and promote subsequent downstream signalling. This has been shown with a crystal structure of the phosphorylated IR TKDs and preceding juxta membrane region to 2.3 Å (PDB:4XLV), in which the TKDs adopt a dimeric conformation (Figure 12B). Subsequent biochemical studies determined the IGF1R TKD forms a similar activated dimer in the structure, the juxta membrane region binds to a cavity formed between the β-sheet and α-helix C (αC). The αC helix is known to be an important regulatory element in protein kinases, typically forming a glutamate-lysine ion pair which coordinates the α and β phosphates of ATP in the active kinase.[95] The dimeric conformation of the IR and IGF1R kinase domains promotes the formation the salt bridge between Lys1030 and Glu1047 (Lys1003 and Glu1020 in IGF1R respectively), which in turn stimulates binding of ATP and substrate phosphorylation.

1.4   The IR-IGF1R Hybrid Receptor

Due to their high homology, the IR and IGF1R are known to heterodimerise and form functional hybrid receptors in tissues where they are co-expressed. Hybrid receptors were originally identified by two separate research groups, observing that IR specific antibodies could bind a subpopulation of IGF1Rs in several cell types[96] and tissues[97]. Subsequently, it was determined that hybrid receptors bind IGF1 with ~50-fold higher affinity than insulin[98-100]. Hybrids are expressed in several mammalian tissues, including tissues relevant to insulin and IGFI signalling, and are estimated to represent between 40 - 90% of total IGF1R[97, 101].

Details concerning the biosynthesis of IR-IGF1R hybrid receptors are not currently known. The prevailing mechanism of hybrid receptor formation is that hybrid receptors form from mixed pools of IR and IGF1R pro-receptor upon synthesis in the endoplasmic reticulum. The proportion of hybrid receptors in tissue correlates with the relative expression of IR and IGF1R[101, 102]. Consistent with this, increased IGF1R expression appears to drive hybrid formation, with both transient expression

in cultures fibroblasts[103] and stable expression in mice[104]. Additionally hybrid receptors spontaneously form from pools of purified IR and IGF1R monomer when exposed to IGF1 & insulin or Mg^{2+} and ATP[102] *in vitro*. This indicates that hybrid receptors can form spontaneously do not require specific cellular machinery.

The physiological role, signalling properties and mechanisms regulating the formation of IR-IGF1R hybrid receptors are currently poorly understood. However, it is believed that hybrid receptors confer insulin resistance by sequestering IR protein and reducing the available insulin binding sites on the cell membrane[104, 105]. There is an association between hybrid receptor expression and insulin resistance[106], obesity[107], hyperinsulinemia[108], hyperglycaemia[109], and type 2 diabetes[110]. Additionally, it appears that manipulating IR/IGF1R stoichiometry through genetic means can modulate insulin sensitivity of a tissue, potentially through modulating hybrid levels: transgenic mice with an endothelial specific over-expression of the IGF1R show reduced endothelial insulin sensitivity, as evidenced by increased NO production relative to wild-type mice[105]. Conversely, mice produced by crossing a whole-body IR haplo-insufficient strain with mice featuring endothelial specific knockdown of the IGF1R showed enhanced insulin-stimulated NO release in endothelial cells compared to the haplo-insufficient model[104]. This implies that reducing IGF1R expression in IR deficient mice results in restoration of endothelial insulin sensitivity. Taken together, these data suggest that the IR-IGF1R hybrid receptors may form part of a positive feedback loop in several metabolic pathologies: downregulation of the IR in these diseases may result in additional incorporation of IR protein into hybrid receptors, which do not respond to insulin.

This highlights the promise of modulating IR-IGF1R hybrid receptors formation as a potential therapy for type 2 diabetes.

1.4.1 Structures of the IR-IGF1R Hybrid Receptor

Structures of the IR-IGF1R hybrid receptors have emerged relatively recently, presumably due to the difficulties in expressing and purifying these heterodimers. The first structure of an IR-IGF1R receptor construct was a 3.0 Å crystal structure of IGFI complexed with an IGF1R αCT and an IR L1-CR fragments (Figure 14A, PDB: 4XSS). This structure represents IGFI bound to the primary hybrid ligand binding site[111], confirming that IGF1 binds in an analogous manner to ligands interacting with the IR and IGF1R. In 2021, a cryo-EM structure of IGF1 complexed to a leucine-zipper fused IR-B-IGF1R ectodomain (Figure 14B, PDB: 7S0Q, 7S8V) was determined to a resolution of 3.7 Å [112]. This cryo-EM structure revealed a congruent architecture with the structures of the single-liganded homodimeric receptors. Notably, IGFI bound preferentially to the side of the heterodimer containing the IR αCT and IGF1R L1 CR domains, in contrast to the prior crystal structure. Interestingly, no structure of the apo-hybrid ectodomain was reported, presumably
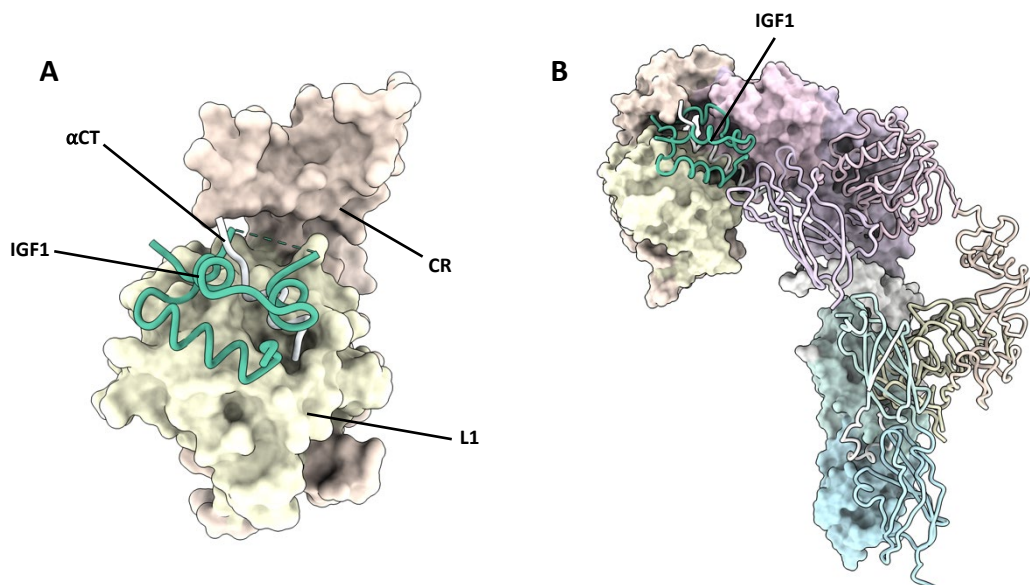
Figure 14 - Structures of the IR-IGF1R hybrid receptor; A) Crystal structure of IGF1 complexed with an IGF1R αCT and IR L1-CR fragment (PDB: 4XSS); B) Cryo-EM structure of IGF1 complexed with leucine-zipper fused IR-IGF1R hybrid (note leucine zipper was not resolved). The IR monomer is depicted with ribbons, and IGF1R depicted with space filling representations (PDB: 7SOQ, 7S8V). Reproduced from Ref. 59 with permission from the Royal Society of Chemistry.

due to the inherent structural flexibility of the receptor in the absence of ligand stabilisation, which prevented obtaining maps of sufficient resolution.

1.4.2   Small Molecule Modulators of the IR-IGF1R Hybrid Receptors Identified via Virtual High-Throughput Screening.

Despite the promise of targeting the IR-IGF1R hybrid receptors for the treatment of metabolic disease, selectivity issues and lack of structural information mean the hybrid receptors remain a formidable drug target. Conventional targeting of the kinase domains offers no scope for selectivity, as these are regions are fully conserved in the IR and IGF1R homodimers. The most effective method of selectively targeting the hybrid receptors appears to be modulating heterodimer formation by targeting the interaction between the IR and IGF1R monomers. This is exemplified by the discovery of the structurally related small molecules hybrid inhibitor 1 (**HI1)** and hybrid inhibitor 2 (**HI2)** by Simmons et. al. [113]. These compounds are reported to selectively reduce hybrid formation by targeting the IR-IGF1R protein-protein interaction (PPI, Figure 15)[113]. These quinoline containing
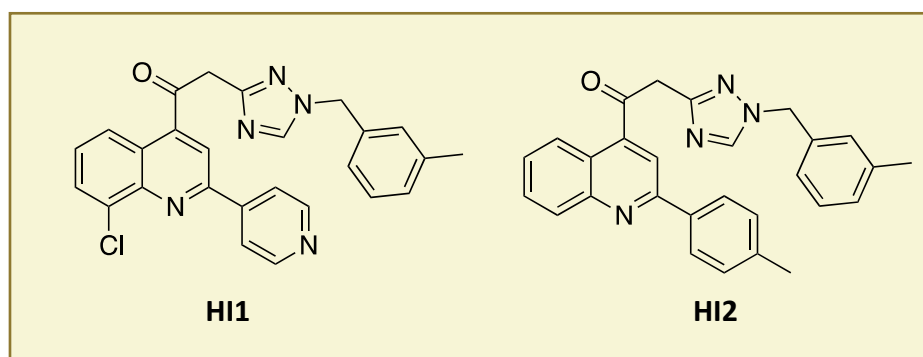
*Figure 15 - Chemical structures of hybrid inhibitors **HI1** and **HI2***

small molecules were identified from a virtual high-throughput screening (vHTS) campaign directed against an apo-IR-IGF1R hybrid homology model. The homology model produced in this study utilised the published structure of the apo-IR homodimer[60], modelling the IGF1R monomer with the iTASSER server[114-116] as no structure of the apo-IGF1R homodimer was published at the time. Computational evaluation of the generated homology model identified regions of amino acid residues which formed key interactions at the IR-IGF1R dimerisation interface, termed hotspots. These epitopes were then targeted for small molecule

intervention, leading to the identification of the HI series. As such, the HI series is predicted to inhibit hybrid formation by binding to IR monomers, thereby preventing dimersation of the IR and IGF1R. Presumably, this must occur prior to disulfide linkage of the receptor monomers in the endoplasmic reticulum. The specific region implicated in this interaction covers IR residues 400-570, which homology modelling predicts to lie at the IR-IGF1R interface formed by the receptor L2 and FnIII-1 domains. However, no structural studies have been conducted to confirm the proposed mechanism of action.
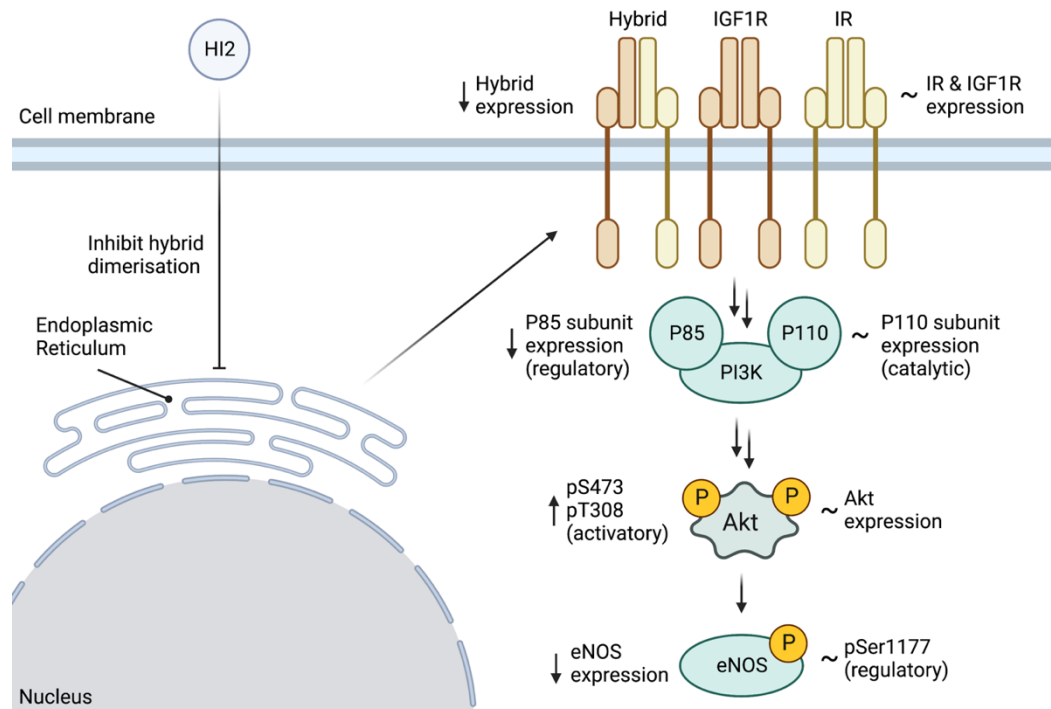
*Figure 16 - Summary of the biological effects of hybrid inhibitor **HI2** on the PI3K/Akt pathway detailing the effects of treating HUVECs with 100 μM **HI2**. **HI2** selectively inhibits hybrid formation, showing no effect on IR or IGF1R expression. This leads to an increase in the expression of the PI3K P85 regulatory subunit, an increase in phosphorylation of activatory Akt S473 and T308 phosphosites, and a decrease in eNOS expression. The expression of the PI3K P110 catalytic subunit, total Akt expression and phosphorylation of the eNOS regulatory S1177 site are unchanged upon HI2 treatment. This indicates that hybrid receptors signal through the PI3K/Akt pathway and that the downstream effects are distinct from those of the IR or IGF1R. Created with BioRender.com.*

The most potent of these molecules, **HI2** has been employed as a tool compound to investigate the downstream signalling capabilities of the IR-IGF1R hybrid receptors (Figure 16). In HUVECs treated with media containing 100 μM of **HI2** for 24 hours, hybrid formation was reduced by more than 50%, whilst no effect on IR or IGF1R expression at the protein or RNA level was observed. This specificity allowed, for the first time, specific evaluation of the IR-IGF1R hybrid signalling pathways.

Treatment of HUVEC cells with media containing 100 µM of **HI2** for 24 hours resulted in reduced expression of the PI3K regulatory p85 subunit but no effect of the expression of the catalytic p110 subunit. Total Akt expression was unaffected, but phosphorylation of activatory sites S473 and T308 was increased, indicative of Akt activation. Whilst total eNOS expression was reduced, activatory phosphorylation of the Ser1177 eNOS regulatory site was unchanged. Additionally, treatment with **HI2** enhanced insulin-induced phosphorylation of S473 pAkt at 24 hr timepoints. These data indicate that inhibition of hybrid formation increases basal and insulin induced activation of the Akt/PI3K pathway. Although direct measurements of NO were not performed, HUVECs treated with 100 µM **HI2** for 24 hrs showed increased endothelial sprouting in an endothelial bead sprouting assay, consistent with increased NO bioavailability.[117] These results indicate hybrid signalling is not simply combination of the signalling properties of IR and IGF1R homodimers and is distinct from the effect of altering IR or IGF1R gene expression.

The HI series faces some major limitations. **HI1** exhibited cytotoxic activity in mammalian cells, reducing cells to *ca.* 70% viability when treated at 100 µM for 24 hrs. Alternatively, **HI2** has a low aqueous solubility of 1.8 µM when measured at pH 7.4 in PBS, which hinders its practical utility. Furthermore, primary ADME studies established that **HI2** possesses a short half-life of 6.97 mins and 1.03 mins in human and mouse liver microsomes respectively. These characteristics emphasize the need to enhance the solubility and metabolic stability of this series to enable effective analysis of hybrid function *in vivo.* However, the reported SAR of the HI series is relatively flat, with seemingly minor modifications resulting in a complete loss of potency. Therefore, it is unclear if the solubility and metabolic stability of this series can be improved whilst retaining potency, and it may be necessary to identify alternative chemical scaffolds with similar activity.

1.5   Project Rationale

Despite growing understanding of the mechanisms driving type 2 diabetes, current treatments remain suboptimal at reversing the molecular pathologies and chronic degeneration of the disease. The association of IR-IGF1R hybrid receptors with insulin resistance makes them a promising novel target for the treatment of

metabolic disease. Current knowledge of hybrid biology indicates that reduction of hybrid formation results in sensitisation to insulin, activation of the Akt/PI3K signalling and may result in the release of the vasoprotective signalling factor NO. This raises the possibility of targeting IR-IGF1R hybrids to restore robust endothelial function in the context of type 2 diabetes. However, hybrid modulation through targeting the kinase domains or ligand binding regions is challenging due to selectivity issues with the IR or IGF1R. A feasible route to specifically modulating hybrid signalling involves altering the propensity for IR-IGF1R heterodimerisation. This comprises targeting of the protein-protein interaction between the IR and IGF1R monomers, offering a chance of selectivity as the structure of the hybrid dimerisation epitopes differs from those of the IR and IGF1R homodimers. Assessing the effectiveness of this approach through chemical or biochemical means would be of great utility in the discovery of hybrid modulators and significantly improve our understanding of the mechanisms affecting hybrid formation.

The HI series of small molecules has set some precedent for this approach. However, the exact mode of action of this series has not been validated, and the low solubility of these compounds in aqueous solution makes structural studies particularly challenging. Therefore, identification of molecules with improved solubility that can modulate hybrid formation would be highly valuable. Such molecules could be utilised for structural studies, allowing the rationale design of improved inhibitors, and shedding light on the structural factors governing IR-IGF1R hybrid formation.

Information to facilitate structure-based design of small molecules targeting these interfaces would greatly augment the search for hybrid modulators. Whilst a high-resolution experimental structure of the apo-hybrid receptor represents the ideal basis for structure-based drug design, the hybrid receptors remain a formidable structural target due to their flexibility, membrane-bound nature and the difficulties in purifying them from IR and IGF1R homodimers. Homology modelling may provide a valuable alternative, as high-quality structures of the IR and IGF1R homodimers can be utilised as templates. Recent breakthroughs, including the

publication of the IGF1R ectodomain dimer structure[61] and significant improvements in computational protein prediction[118] mean there is significant scope to refine the IR-IGF1R hybrid homology model utilised in the discovery of **HI2**. Generation of an improved hybrid model can serve as the basis for enhanced screening to identify hybrid modulators with improved potency and physicochemical properties. Complementary mutagenesis can be employed to validate the significance of the identified interfaces in hybrid formation and function, as well as allowing targeting of specific hotspot epitopes.

1.6 Project Goals

This thesis aims to evaluate the molecular mechanisms affecting IR-IGF1R hybrid formation. Specifically, it aims to develop improved structural models of the IR-IGF1R hybrid receptors and evaluate these through a combination of structure-based drug design and point mutagenesis studies. It aims to rationally develop methods of modulating IR-IGF1R hybrid formation, with the expectation that these methods will be relevant in the treatment and understanding of type 2 diabetes.

Hypothesis: Understanding the structure of IR-IGF1R hybrids and how to modulate their formation will further our understanding of their mechanism of action.

1.6.1 Aims

1. To improve the current hybrid homology model based on recent structural studies.
2. To use virtual high-throughput screening to identify novel small molecule modulators of hybrid formation.
3. To evaluate these small molecules using established screening cascades and structural studies.
4. To produce protein for structural studies on the IR-IGF1R complex.

2 Computational Modelling and Virtual Screening of the IR-IGF1R Hybrid Receptor

2.1 Virtual High-Throughput Screening

Virtual High-Throughput Screening provides a cost and time effective alternative to traditional high-throughput screening (HTS) for identifying potential modulators of a protein target. vHTS allows millions of commercially available compounds to be computationally evaluated for their predicted ability to bind to a protein and is utilised to focus larger chemical libraries into prioritised subsets of molecules for lower-throughput biological screening. Such screens have resulted in hit rates of up to 30%, compared to 1% for a typical HTS screen[119].

Molecular docking is among the most popular vHTS techniques, due to its ability to rapidly predict the binding conformation of a small molecule into a target protein structure. Recent advances in computational hardware and screening algorithms mean that the molecular docking of several hundred million compounds is now viable. A docking screen requires both a library of compounds to be screened and a structure of the targeted protein binding site, which can be experimentally determined using structural biology techniques or derived from a homology model. Typically, docking algorithms hold the protein receptor static whilst the ligand remains flexible. This approximation is necessary as the protein macromolecule is typically much larger, with many degrees of freedom, and this conformational flexibility makes the docking calculations computationally expensive. Whilst proteins are known to be dynamic, they are typically more rigid at binding sites and undergo relatively small conformational changes upon ligand binding[120].

Docking methodologies can be split into several stages. Initially, multiple conformations of the ligand are generated to allow sampling of its possible binding modes. Most docking algorithms require 3D molecular models of ligands as starting points, and purchasable compound databases typically only contain 2D structures.

Therefore, a distinct ligand preparation step is often performed to enumerate the low energy conformation, ionisation states and tautomers a ligand can sample.

Similarly, the protein structure is prepared for docking by adding hydrogens, optimizing hydrogen bonds, and removing atomic clashes. These steps are necessary as they are not typically performed during the refinement of x-ray crystal structures. Additionally, any missing regions of the experimental structure, such as sidechains or flexible loops may be modelled and added at this stage. Next, a grid is generated around the target protein, which describes the forcefield around the receptor protein with a degree of granularity to reduce computational burden. The prepared ligands are then placed into the grid to determine their optimal binding conformation in a process known as searching. This generates a set of predicted conformations between the ligand and receptor known as poses. Finally, each pose is assigned a score representing an estimate of its binding affinity, allowing the discrimination of favourable poses from unfavourable ones. Scoring algorithms typically account for factors such as Van der Waals forces, hydrogen bonds, electrostatic interactions and desolvation penalties occurring between the ligand and protein. The ideal scoring function would correctly predict the binding free energy between a ligand and its target. However, restrictions in computational methods and resources mean that scoring functions are not completely accurate measures of binding energetics. The simplified forcefields commonly employed in scoring calculations do not always capture the complexity of ligand binding events. Therefore, docking scores are typically used to rank compounds relative to each other, rather than using the score as an absolute measure of binding affinity. As such, ligand docking and scoring has proved an efficient way of distinguishing molecules with biological activity from larger compound libraries[121], provided that results are interpreted prudently and complemented with experimental techniques.

2.2   Glide

Among the most popular docking algorithms is Glide[122-125], an add-on to the Schrödinger Maestro package[126]. Glide features 3 separate docking modes, which allow the user flexibility between speed and computational demand: Glide HTVS (high-throughput virtual screening), which is intended for extremely large

databases and has restricted conformational sampling; Glide SP (standard precision); and Glide XP (extra-precision), which utilises more extensive sampling and harsher scoring penalties for the shape-complementarity of ligand and receptor. Glide XP is specifically designed for use on high scoring ligand poses from Glide SP to remove false positives. This creates a workflow of increasingly rigorous screening procedures, where only the top 10-20% the highest scoring poses from the previous round are screened with computationally expensive procedures.

Unlike other popular docking packages such as GOLD[127], AutoDock[128] and eHiTS[129], Glide SP approximates a complete systematic search of the positional, conformational and orientational space of the docked ligand. A series of hierarchical filters mean that the most computationally expensive calculations are reserved for the most promising fraction of docking solutions for each molecule: initial rough positioning is followed by torsionally flexible optimisation. The best candidates are refined via Monte Carlo sampling of the predicted pose, allowing Glide to provide a suitable balance between speed and accuracy when predicting docked poses. The Glide SP scoring function includes terms assessing Van der Waals energy, coulombic energy, lipophilic interactions, H-bonds, rotatable bonds, and polar interactions within the receptor site. Additionally, a series of rewards and penalties are applied for more specific features, such as buried polar groups, hydrophobic enclosure, and amide twists. By comparison, Glide XP is guided by an anchor and growth strategy, where anchor fragments, typically rings, are chosen from the SP poses and the molecule regrown bond by bond from these anchor points. Energy minimization and scoring are carried out on an ensemble of poses generated by this sampling strategy, meaning more extensive sampling is performed for each ligand relative to Glide SP. Furthermore, the XP scoring function includes additional terms, with more comprehensive treatment of several terms included in the SP scoring function. The XP scoring function specifically rewards the occupancy of hydrophobic pockets, which is typically undervalued in the SP scoring function. Solvation is also modelled by docking water molecules into poses that have otherwise good scores, and penalties applied to structures where groups are inadequately solvated. The improved sampling method utilised by Glide XP means

substantially harsher penalties can be applied for serious violations of physical principles relative to Glide SP.

Further benefits of Glide include its integrated suite of tools, which include SiteMap[130] for the detection of small molecule binding sites. This is of particular utility in the screening of protein-protein interaction inhibitors, as such targets are unlikely to contain defined regions that are evolved to bind small molecules. SiteMap can be utilised to detect binding sites that able to bind small molecules at a protein-protein interface. SiteMap uses the scoring function SiteScore, to evaluate potential small molecule binding regions on a protein structure. SiteScore is based on the weighted sum of several calculated properties, including exposure, hydrophobicity/hydrophilic balance, H-bond donor/acceptor character and the site volume. The SiteScore is calibrated such that the average SiteScore for 157 separate sub-micromolar sites is 1.0, with a SiteScore of 0.80 the benchmark to distinguish ligand binding and none-ligand binding sites[131].

2.3   Homology Modelling of Insulin Receptor/ IGF1 Receptor Structure

As several related structures of the unliganded IR and IGF1R had been determined[60, 61] (section 1.3.1), homology modelling provides a viable route for building a structural model of the unliganded IR-IGF1R hybrid receptor. Homology modelling[132, 133] is a computational structure prediction method which calculates protein structure from the amino acid sequence and the structure of related proteins. This uses the principle that amino acid sequences sharing a high degree of sequence similarity are likely to have similar structure and function. Therefore, a template protein for which the 3D structure is known can be used to predict a structural model of the target protein. Typically, the homology modelling process involves first identifying template proteins by target-template sequence alignment. This step dramatically influences the resulting model quality; accurate models require templates with high homology and sequence coverage. Additionally, high resolution templates produce better models, as they have lower numbers of errors or missing residues. The most common databases used for identifying templates are the protein data bank (PDB)[134, 135] and SWISS-MODEL[136-138]. At this stage, distinctions can be made between alignment-based homology modelling and

threading-based homology modelling. Alignment-based homology modelling such as I-TASSER[114-116] or SWISS-MODEL[136, 138] refers to prediction using sequence alignments with at least domain level sequence coverage, whilst threading methodologies such as RaptorX[139, 140] uses templates with only fold-level homology to the target. In practice, there is no strict boundary between the two techniques, and hybrid algorithms can utilise a combination of both methods.

Increasingly, homology modelling utilises multiple templates in predicting a model, as this can result in improved accuracy. Once templates are identified, the model is built by inserting the target sequence into the template and adjusting the structure to fit the target sequence. Target regions that are not aligned to a template are filled in by a process known as loop modelling. Due to the lack of sequence coverage, these regions are the most susceptible to modelling errors. The resulting model is then refined by energy minimization.

As homology models are based on assumptions that the target protein shares significant structural similarity with the template proteins, it is important to assess the quality of any generated model[141]. This can be performed *in silico,* by assessing metrics such as stereochemical quality and overall energy, or by comparison of the model to the experimentally determined structures of close homologues. If possible, the model should also be experimentally validated, normally through functional analysis on active sites, ligand binding sites or interaction interfaces. Validation allows the identification of artefacts or inconsistencies produced from modelling, and the discrimination between areas of high- and low-quality prediction.

2.3.1 Homology Modelling Using the TACOS Methodology
2.3.2 TACOS: Model Building

While several homology modelling packages and servers are publicly available, not all are suited to every protein. Many homology modelling methods are limited to predicting the structures of single polypeptide chains. To produce a high-quality

model of the IR-IGF1R hybrid structure, it was necessary for the homology modelling methodology to accurately predict protein complexes.

The TACOS[142] (Template-based Assembly of COmplex Structures) server fit this criterion, as it is designed to model the structures of protein-protein complexes. With two input sequences, this method identifies protein complex structure templates from the PDB using multi-chain threading algorithms. With templates identified, they are broken into fragments and used to reassemble the complex structure with replica-exchange Monte Carlo simulations[143]. Models are selected by structure clustering and then refined into full-atomic models.

Due to the high computational cost of this pipeline, only sequences of up to 1500 amino acid residues could be predicted. This precluded the modelling of the entire hybrid receptor structure with the TACOS methodology; it would instead be necessary to prioritise regions of the receptor that were most promising for modelling and subsequent screening. As our design rationale was focussed on targeting the IR-IGF1R PPI, the published IR and IGF1R apo-receptor structures were examined to determine the major interfaces forming between the two receptor monomers. In both receptors, three separate interfaces form: between the L1: FnIII-2' domains, the L2: FnIII-1' domains and between the αCT: L1' (Figure 7). With these interfaces present in both the IR and IGF1R structures, it was reasoned that the same interfaces were likely to form in the IR-IGF1R hybrid receptor. Of these, the L2: FnIII-1 interface was chosen due to this interface containing the lowest levels of conservation between the IR and IGF1R monomers, meaning that selectivity was more likely to be achievable at this interface. Additionally, whilst the binding mode of **HI2** has not been experimentally elucidated, this is the same

region that was targeted for the discovery of the HI series[113], providing some precedent in the successful targeting of this interface.

Sequence alignments were performed using MUSTER[144] on residues 332-619 of the IR-B (UniProt ID: P06213-1) and 331-608 of the IGF1R (UniProt ID: P08069) sequences downloaded from Uniprot[145, 146]. These regions both share 100% similarity with the related regions of the IR crystal structure (PDB:4ZXB, chain E)[60] and IGF1R crystal structure (PDB: 5U8R chain A)[61], despite both experimental structures utilising mutated and truncated constructs. TACOS identified the IR crystal structure (PDB:4ZXB) as the best template for the global protein complex, with 68% identity and 86% coverage. The normalised z-score for this alignment was 4.4, clearing the benchmark of 2.5 for a good alignment. Overall, this indicates that the templates utilised were of good quality.
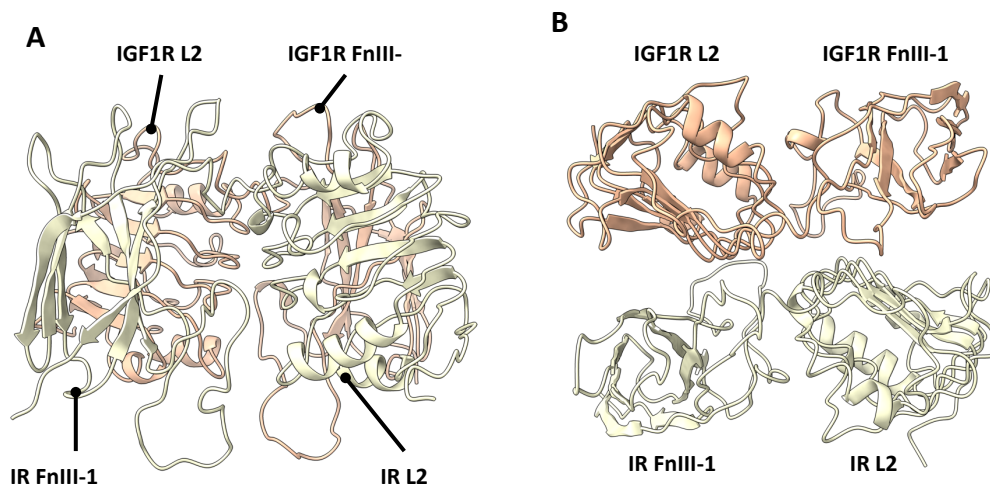
Figure 17 - Homology models of the IR-IGF1R L2: FnIII-1 interface. The IR monomer is depicted in yellow and the IGF1R monomer is orange A) Face-on view of the TACOS homology model method B) Top-down view of the TACOS homology model.

The predicted structure of the hybrid L2: FnIII-1 interface is depicted in Figure 17A, B. The interface between the IR and IGF1R is predicted to form between the β-sheet of the L2 domain and the second (C-terminal) β-sheet and loop regions of the FnIII-1 domains of each monomer respectively. However, due to the differences in sequence homology of the IR and IGF1R chains in this region, the interactions are not truly conserved at the contacts between IR L2 and the IGF1R FnIII-1, and vice versa. The predicted L2:FnIII-1 interface is extremely similar to the analogous

L2:FnIII-1 interfaces of the IR[60] and IGF1R[61] unliganded ectodomain structures, with RMSD values of 1.149 Å and 0.832 Å relative to the relevant regions of the IGF1R (5U8R) and IR (4ZXB) respectively. Analysis of the TACOS model using PDBePISA[147, 148] determined that the hybrid L2: FnIII-1 interface covers an area of 1491.1 Å^2, with a ΔG (Gibbs free energy) of -5.4 kcalmol^{-1}. Typical PPIs exhibit interface areas ranging from 1300 to 2500 Å^2 and free energies between -6 and -12 kcal mol^{-1} [149], meaning the hybrid L2:FnIII-1 interface is both smaller and lower affinity than a typical PPI. This may be explained by the fact that the hybrid receptor monomers are held together by disulfide bonds. Therefore, a tightly bound interface is not required for robust receptor function. For subsequent computational hotspot analysis and virtual high-throughput screening, this model is will be referred to as the TACOS model.

2.3.3   TACOS: Model Validation

One limitation of the TACOS modelling methodology is the absence of integrated tools for assessing the quality of the generated models. Therefore, the SWISS-MODEL structure assessment suite[136] was utilised to evaluate the TACOS model.

QMEANDisCO[150] score was used as the primary metric of assessment and generates a score for both global and local quality. The QMEANDisCO score evaluates the consistency of pairwise Cα- Cα distances in the model by comparing them against restraints extracted from homologous structures. Effectively, this can be seen as a score of how comparable the model is to an ensemble of experimental structures.
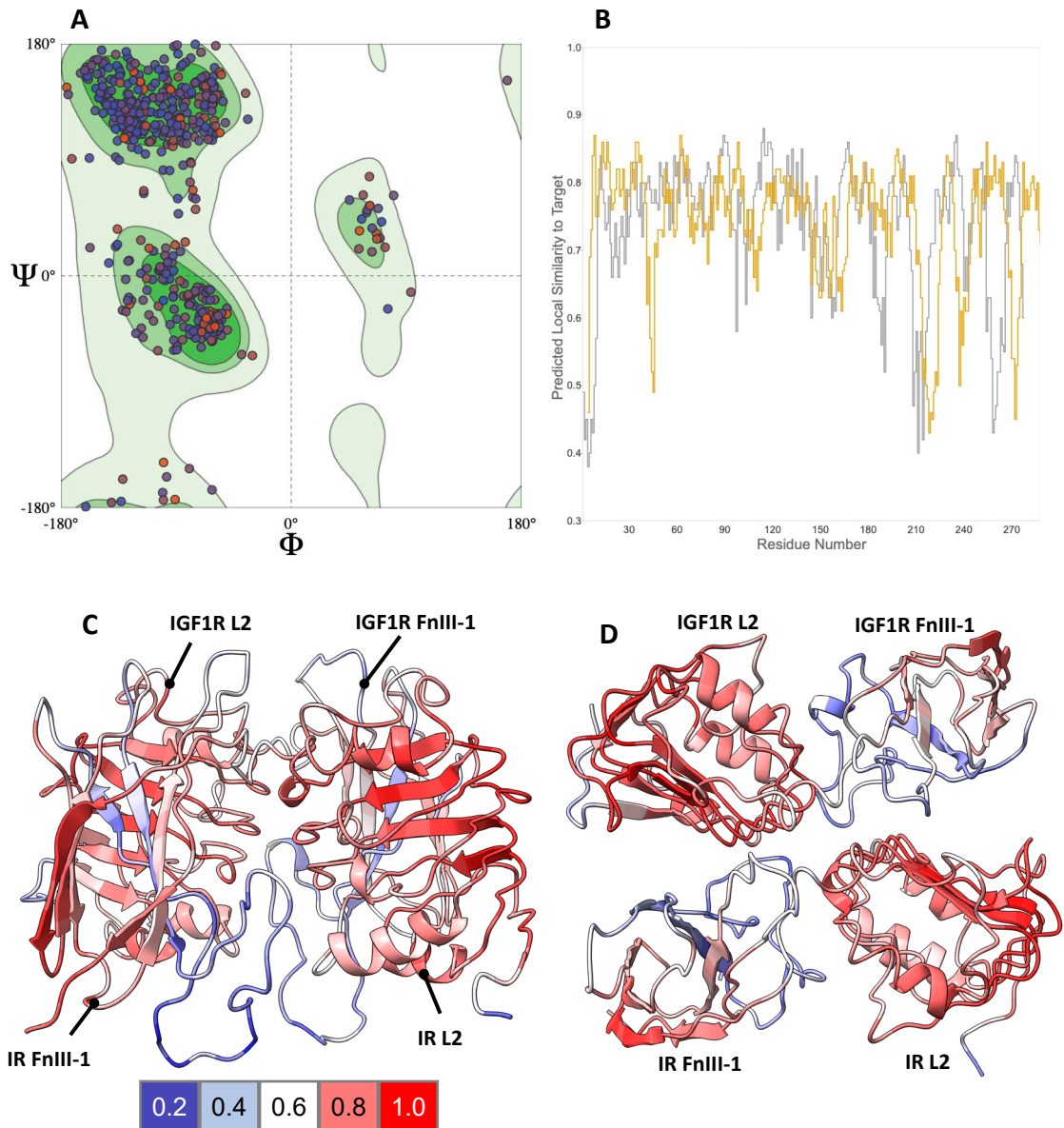
Figure 18 - TACOS homology model validation with SWISS-MODEL structure assessment; A) Ramachandran plot of TACOS homology model. φ = C-N-Cα-C angle, Ψ = N-Cα-C-N angle of amino acid backbones. Contours of favoured regions are extracted from 12,521 non-redundant experiment structures (pairwise identity cut-off 30%, resolution cut-off 2.5 Å) such that 99.7% φ/Ψ are contained in the first contour, 95.0% in the second contour and 80.0% in the third; B) Plot of the per residue QMEANDisCO score. The IR monomer (grey) and IGF1R monomer (orange) are both shown. Residue numbers refer to model numbering; C) Face-on view of the TACOS homology model coloured by QMEANDisCO score according to the key (colour key at bottom); D) Top-down view of the TACOS homology model coloured by QMEANDisCO score (colour key at bottom).

The score is calibrated between 1 and 0, with below 0.6 indicating areas of lower quality.

The TACOS global QMEANDisCO score was calculated as 0.74 ± 0.05. Additionally, QMEANDisCO per residue analysis was utilised (Figure 18B-D) assess the local quality relative to the overall model quality. Areas of significantly reduced local quality were identified at IR residues 533-554 and 581-594 (model numbering:201-

222 and 249-262), as well as in IGF1R residues 544-566 and 599-614 (model numbering: 213-235 and 268-283). These regions correspond to the β5 strand, β6 strand and connecting loop in the FnIII-1 domains. The lower quality of these regions is unsurprising, as these regions are relatively low resolution in both the template crystal structures. Additionally, a Ramachandran plot (Figure 18A) determined that 92.11% residues exhibited favoured dihedral angles, with 0.72% outliers. Overall, assessment metrics for the model indicated it was of good quality[136, 141, 150], with specific regions in the FnIII-1 domain predicted to be of lower quality.

2.3.4 TACOS: Hotspot Prediction

Whilst PPI interfaces will usually comprise a large area compared to a typical small-molecule binding pocket in a protein , it is accepted that a relatively small-number of amino acid residues contribute a large percentage of the total ΔG^f of the PPI interface[151]. Clusters of such residues are termed 'hotspots'. Drug molecules typically only need to bind to hotspot regions to disrupt a PPI interface. Therefore, computational hotspot prediction was utilised to identify hotspot residues at the L2: FnIII-1 interface which we could target in subsequent vHTS screening campaigns.

The KFC2 server[152-154] was chosen to computationally predict hotspot residues at the L2: FnIII-1 interface of the TACOS model. The KFC server (Knowledge-based FADE and Contacts) predicts hotspots at protein-protein interaction interfaces by recognising structural features indicative of key binding contacts[152]. The KFC2 server provides two alternative methods of hotspot prediction (KFC2a & KFC2b), which predict hotspots largely on features related to solvent accessible surface and local plasticity. The two methods share only two of eight predictive features, providing two largely independent methods, with KFC2a offering better sensitivity but lower specificity for predicting hotspot residues when compared with KFC2b. The KFC2 methodologies supply a score for each residue, which is normalised such that a score greater than zero indicates a hotspot residue. Scores for hotspot

residues typically range between 0-1.5, with higher scores indicating greater confidence in the hotspot prediction.

Chain	Amino acid	Residue	KFC2a	KFC2a Conf	KFC2b	KFC2b conf
IR	GLN	433	Hotspot	0.48	-------	-0.07
IR	PHE	454	-------	-0.03	Hotspot	0.02
IR	HIS	456	Hotspot	0.71	Hotspot	0.10
IR	TYR	457	Hotspot	0.47	-------	-0.31
IR	PRO	459	Hotspot	0.92	-------	-0.37
IR	LYS	487	Hotspot	1.59	-------	-0.02
IR	THR	488	Hotspot	1.18	-------	-0.29
IR	ASP	491	Hotspot	1.88	-------	-0.12
IR	GLN	492	Hotspot	1.58	Hotspot	0.05
IR	THR	598	Hotspot	0.99	-------	-0.50
IR	SER	600	Hotspot	0.37	-------	-0.61
IGF1R	PHE	450	Hotspot	0.62	Hotspot	0.04
IGF1R	PRO	452	Hotspot	0.77	-------	-0.42
IGF1R	LYS	453	Hotspot	0.04	-------	-0.33
IGF1R	ASN	478	Hotspot	0.23	-------	-0.42
IGF1R	ARG	480	Hotspot	1.13	Hotspot	0.04
IGF1R	ASN	481	Hotspot	1.15	-------	-0.06
IGF1R	GLU	484	Hotspot	0.79	-------	-0.26
IGF1R	ARG	485	Hotspot	1.40	Hotspot	0.28
IGF1R	ARG	518	Hotspot	0.42	-------	-0.18
IGF1R	LEU	586	Hotspot	1.09	-------	-0.03

Table 1 - Hotspot residues predicted by KFC2 for the TACOS homology model. Residues also predicted by KFC2a in the AlphaFold homology model are highlighted in blue.

KFC2 hotspot analysis of the Hybrid L2: FnIII-1 interface classified 21 residues as hotspots using the KFC2a method, and six residues were classified as hotspots using KFC2b (Table 1). Of the residues classified as hotspots using KFC2b, only IR F454 was not similarly classified by KFC2a, albeit both methods gave F454 a confidence score close to zero indicating low confidence in this prediction. The hotspot residues clustered at the top of the L2 domains of both chains (Figure 19A, B), with residues on these domains comprising 15 of 20 hotspot residues predicted by KFC2a and four of six predicted by KFC2b. As such, the hotspot interactions are
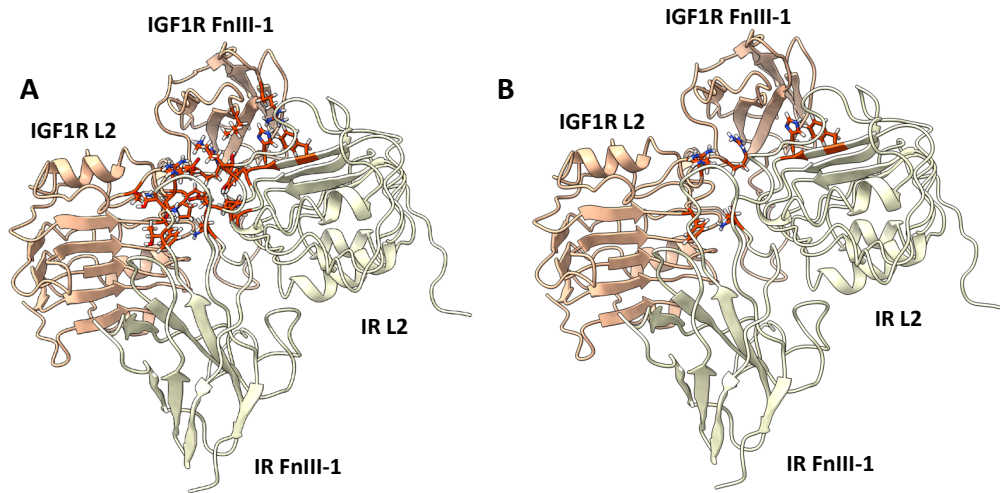
Figure 19 - KFC2 hotspot prediction for the TACOS homology model at the L2: FnIII-1 hybrid interface. The IR monomer is depicted in yellow, the IGF1R monomer depicted in orange, and the hotspot residues depicted in red. A) Hotspot residues predicted by KFC2a on the TACOS homology model; B) Hotspot residues predicted by KFC2b on the TACOS homology model.

notably asymmetric; whilst the IR and IGF1R monomers contribute 11 and 10 hotspot residues respectfully, only IR P459 (IGF1R P450) is conserved between two chains. The hotspot analysis indicates that interface formation is largely governed by hydrophobic interactions, with no hydrogen bonds occurring between hotspot residues from either chain. In the mature hybrid receptor, the IR and IGF1R are held together by intermolecular disulfide bonds, and the dominance of hydrophobic interactions at the L2:FnII-1 interface is consistent with that of many obligate PPI's[155]. Specifically, residues IR Y457 and IGF1R F450 insert into hydrophobic cavities on the opposite monomer, forming critical reciprocal interactions (Figure 20A). A further key cluster of contacts occurs where residues IR K487, T488, D491 and Q492 pack against IGF1R residues N478, R480, E484 and R485 (Figure 20B). The number and relative proximities of the predicted hotpot residues was encouraging, as they formed compact, well-defined regions that could potentially be targeted by a small molecule.

To investigate their potential functional significance, hotspot residues identified by KFC2 were compared with the Human Gene Mutation Database[156] (HGMD), to determine if these residues had previously been identified as disease-causing mutations. Notably, the two point mutations IR K487E[157], T488P[158] have been documented as natural disease-causing variants associated with Donohue
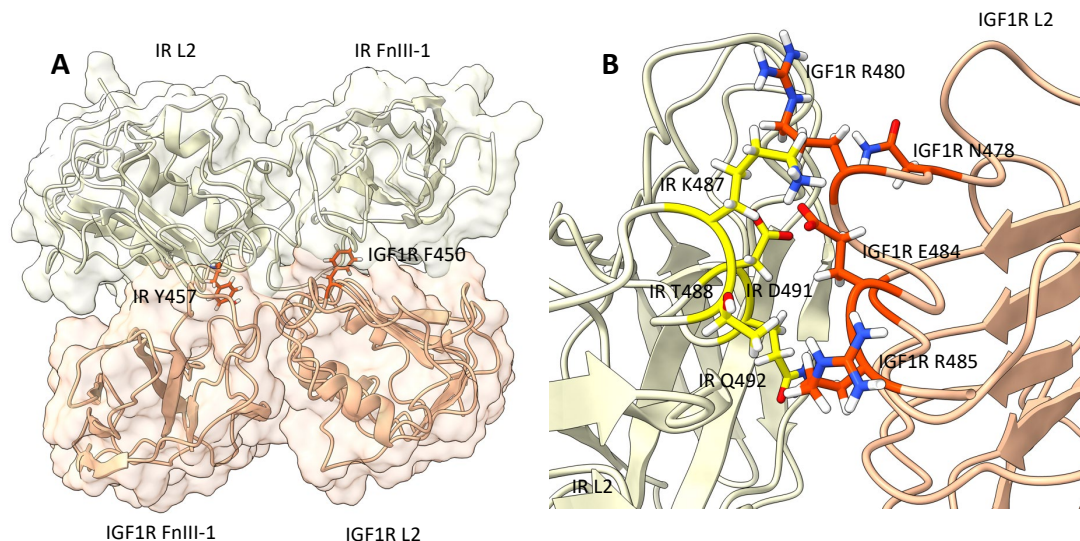
Figure 20 - Key hotspot interactions at the TACOS hybrid interface; A) The locations of IRY457 (red) and IGF1R F450 (red). The IR (yellow) and IGF1R (orange) are shown with a surface and cartoon representation; B) Close up of IR (light yellow) K487, T488, D491 and Q492 (dark yellow) packing against IGF1R (orange) residues N478, R480, E484 and R485 (red).

Syndrome. This disease is characterised by severe insulin resistance and significant growth restriction in humans, implying these residues have significant functional importance. However, none of the IGF1R mutations identified by KFC2 had previously been implicated in disease.

2.3.5   TACOS: Virtual High-Throughput Screening

With the identification of well-defined hotspot epitopes, docking based vHTS using the TACOS IR-IGF1R hybrid homology model appeared a promising route to identifying modulators of the IR-IGF1R PPI.

Two libraries were chosen for this virtual screening campaign, specifically the Enamine PPI library and our in-house MCCB library. The Enamine PPI library is a commercially available library of 40,640 diverse small-molecules, with compounds chosen to mimic protein-structural features such as β-sheets. The medicinal chemistry/chemical biology (MCCB) group chemical library features a diverse set of 27,000 compounds designed to cover a wide chemical space, ensuring representation of different molecular scaffolds, sizes, shapes, and physicochemical properties. Both libraries were subjected to pan-assay interference (PAIN) compound checks and flagged compounds removed. PAINs compounds[159] are classes of molecules known to cause interference in biological assays, thereby frequently causing false positives. These scaffolds are associated with promiscuous

binding, aggregation, or reactivity. Therefore, it is preferable to remove such scaffolds to minimise the risk of misleading assay results.

This screening campaign was performed using the Schrödinger small-molecule discovery platform[126] mounted on the ARC3 advanced research computing cluster at the University of Leeds. As such, the preparation of energy minimized 3D ligand structures was performed with LigPrep[160], binding site detection performed with SiteMap[130, 131, 161], protein preparation with the Maestro Protein Preparation wizard[162, 163], and receptor grid generation and flexible docking performed with Glide[122-125]. A schematic representation of the docking workflow is shown in Figure 21.

Glide SP 67,000	Glide XP 10,000	Clustering 2,000	Visual Inspection 500	Purchase 42
• Fast • Torsionally flexible	• Anchor and grow • Solvation • Lipid	• Tanimoto similarity	• Steric clashes • H-bonds • Hydrophobic interactions	
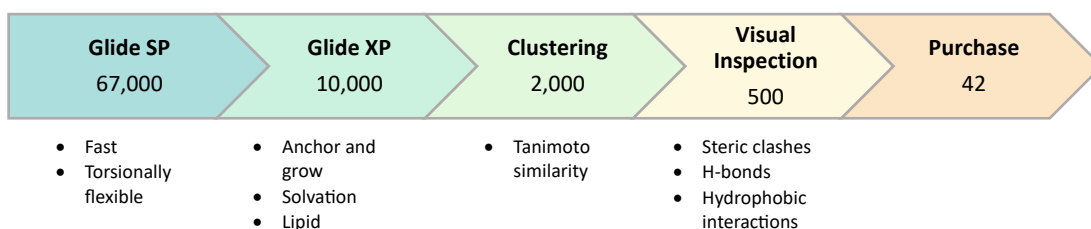
Figure 21 - Schematic of the vHTS workflow detailing the number of ligands selected at each stage.

2.3.6   TACOS: Binding Site Prediction & Grid Generation

SiteMap was utilised to detect regions of the protein that were suitable for binding small molecules. SiteMap identifies ligand binding regions by first identifying areas on the outside of the protein which have sufficient van der Waals contacts and enclosure to function as site points. Contiguous site points within a set proximity are then grouped and evaluated according to a scoring function which accounts for the size, enclosure and hydrophobicity of the identified region[131]. The IGF1R monomer was chosen for this calculation as it contained most hotspot residues, and the IR monomer was removed from the structure to allow sites at the PPI interface to be detected. SiteMap detected five separate sites, of which only one was in proximity to the hotspot residues detected by KFC2. This site was located directly adjacent to the hotspot epitope formed by IGF1R residues N478 - R485 and extended backwards into the groove formed between the IGF1R L2: FnIII-1 domains (Figure 22). The corresponding site score was 0.997, and total surface area was 111
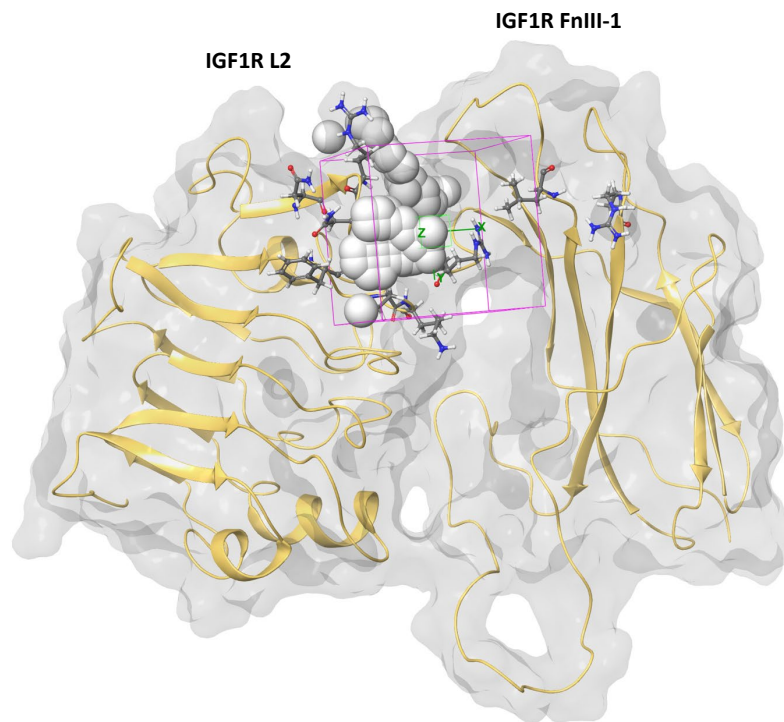
Figure 22 - IGF1R monomer from the TACOS homology model showing the ligand binding site detected by SiteMap (white spheres) and the location of the grid generated for docking (pink box). Hotspot residues are depicted in grey.

$Å^2$. This was encouraging as site score is calibrated such that a score > 0.8 indicates a drug-binding region, with a site score > 1.0 deemed particularly promising.

This site was utilised directly for receptor grid generation with Glide with a grid size of 10x10x10 Å. Both the Enamine PPI library and MCCB library were prepared with LigPrep[160]. This performs several functions: most importantly, generating several 3D low energy conformers for each 2D input structure, but additionally adding hydrogen atoms, desalting, optimising tautomers and generating ionization states at pH 7. Subsequently, each library was docked into the receptor in Glide SP, before the top 10,000 scoring poses, ranked by Glide Gscore, were redocked utilising Glide XP. The top 2,000 poses were clustered using fingerprint analysis to identity highly represented poses. The Tanimoto coefficient[164] (cut off > 0.5) was used as the

chemical similarity metric to ensure highly populated poses were well represented in prioritised compounds.

2.3.7 Triaging of Screening Results from the TACOS Homology Model

Post the initial docking workflow, poses were ranked by predicted binding affinity to the IGF1R monomer (Glide XP score). The best scoring molecules had XP scores of -6.582 kcal mol^{-1} and -6.574 kcal mol^{-1} for the Enamine PPI and MCCB libraries respectively. The magnitude of docking scores varies substantially with the nature of the receptor, but good scores for enclosed binding pockets can range from -9 kcal mol^{-1} to -11 kcal mol^{-1}. The scores observed above are typical of a relatively shallow binding pocket, as ligands have fewer opportunities to form interactions with the receptor. In the absence of alternative molecules known to bind to this pocket, it is impossible to perform benchmarking studies to further evaluate the magnitude of the docking scores. It was noteworthy that the Enamine PPI library did not produce lower scoring poses than the MCCB library, as the Enamine PPI library is based around common PPI inhibitor scaffolds. Clustering separated the top 1000 scoring molecules into 30 separate groups, indicated good diversity amongst the top 1000 scoring ligands.

The top scoring poses from the 30 groups were visually evaluated to inspect their docking poses. The binding poses of top scoring molecules in each group were visually inspected to identify predicted hydrogen bonds, π-stacking and electrostatic interactions, as well as and steric clashes with the IGF1R. Additionally, the IR monomer was superimposed onto the docked structures and those that formed steric clashes with further hotspot residues on the IR chain also prioritised. Visual analysis of the docked poses identified several key interactions. By far the most common interactions between protein and ligand were H-bonds or electrostatic interactions formed with the hotspot residue E484 (Figure 23A, B). Many high scoring poses featured ligands capable of extending down the binding site to access a hydrophobic region in the IGF1R adjacent to F450 and P452 (Figure 23A, B). Aromatic rings were commonly present in this space, which in some cases allowed a pi-stack with hotspot F450 (Figure 23B). When the structure of the IR
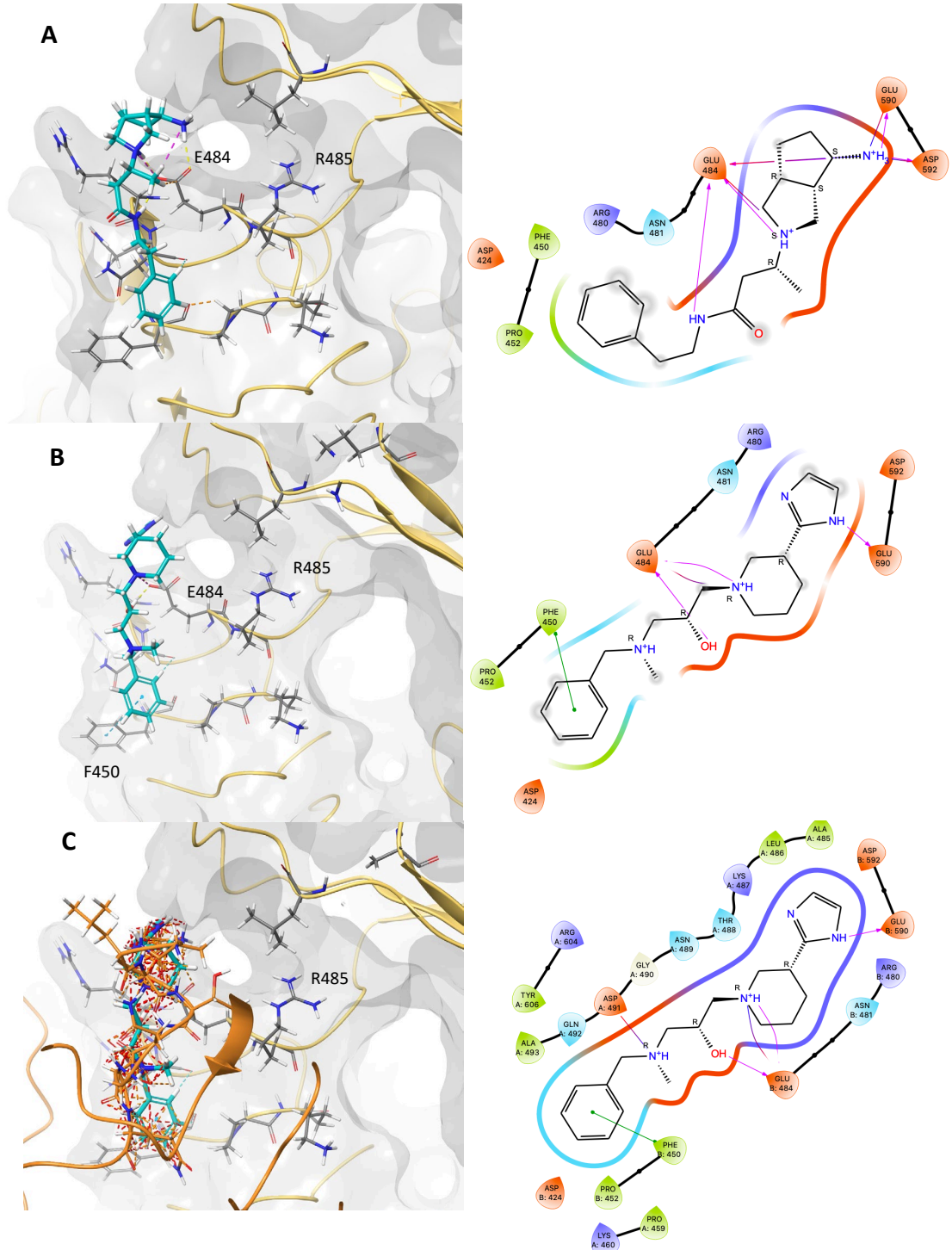
Figure 23 - Docked poses of ligands bound to the IR: 3D images (left) and 2D interaction diagrams (right) are shown for each ligand. In the 2D diagrams IR residues are labelled chain A and IGF1R labelled chain B. The IGF1R (gold ribbons, grey surface), hotspot residues (grey residues) and ligand (cyan) are depicted with H-bonds (yellow), π-stack (sky blue) and electrostatic interactions (purple) shown. For E & F the IR monomer (orange) is superimposed, with clashes shown (orange and red lines): A) Ligand Z1491395228; B) Ligand Z1491395185; C) As in B, but with the IR monomer superimposed.

monomer was superimposed onto the docked poses, significant steric clashes were

observed with the docked ligand and IR residues 485-493, which includes hotspot

residues T488, D491 and Q492 (Figure 23C). Therefore, it was predicted that such poses would inhibit the formation of the IR: IGF1R hybrids.

2.4   Creating a Model of the IR-IGF1R Hybrid Receptor using AlphaFold

2.4.1   AlphaFold: Model Building

The recent development of the AlphaFold 2[165] algorithm for protein structure prediction provided an opportunity to further improve the homology model of the IR-IGF1R hybrid receptor. The AlphaFold system is a deep learning based algorithm that uses artificial intelligence and machine learning techniques to predict protein structures; it is viewed as a significant advancement in the field of homology modelling[166]. AlphaFold's proficiency was demonstrated at the critical assessment of methods of protein structure prediction - round 14 (CASP14), a biennial competition where protein prediction methodologies are tested against proteins for which the structures have been experimentally determined. AlphaFold 2 was able to predict models that rivalled experimental accuracy for two thirds of the targets[118], substantially outperforming every other methodology tested. It was decided to leverage this new technique to provide an improved model of the IR-IGF1R hybrid.

Whilst the AlphaFold 2 system is trained on individual protein chains, a separate system known as AlphaFold multimer[167] has been developed for the prediction of multimeric protein complex structures. To allow direct comparison to the TACOS
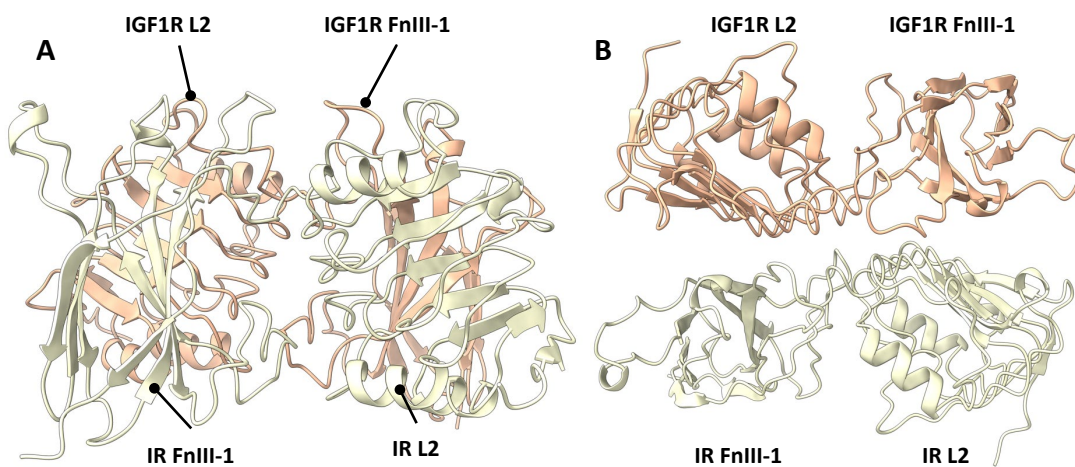
Figure 24 - AlphaFold homology models of the L2: FnIII-1 interface. IR monomer (blue) and IGF1R monomer (grey) are shown; A) Face on view of homology model built with AlphaFold method; B) Top-down method of homology model built with AlphaFold method.

model, the sequences corresponding to residues 332-619 of the IR-B and 331-608 of the IGF1R sequences were submitted to AlphaFold multimer to predict the amino acid sequence corresponding to the L2: FnIII-1 interface of the hybrid receptor (Figure 24A,B). The model predicted by AlphaFold showed good agreement to the TACOS model with an RMSD of 0.862 Å. Specifically, the secondary structure elements comprising the FnIII-1: L2 interface were highly concurrent, whilst the areas of lower agreement were typically located on the flexible loop regions of the FnIII-1 β-sheets. As the loop regions of the FnIII-1 β-sheets are lower resolution in the template crystal structures and likely highly flexible, discrepancies in modelling of these regions are not particularly concerning. Such regions do not form a major proportion of the L2: FnIII-1 interface and will therefore be largely avoided when targeting the interface with small molecules.

2.4.2   AlphaFold: Model Validation

The AlphaFold methodology provides some additional metrics indicating the quality of regions of the predicted model. Firstly, AlphaFold calculates the predicted local distance difference test (pLDDT) which evaluates how well the environment in a reference structure is reproduced in a protein model through comparison of Cα local difference distance tests; it is effectively a metric assessing the local quality of a prediction and conceptually related to the QMEANDisCO score. Regions of the
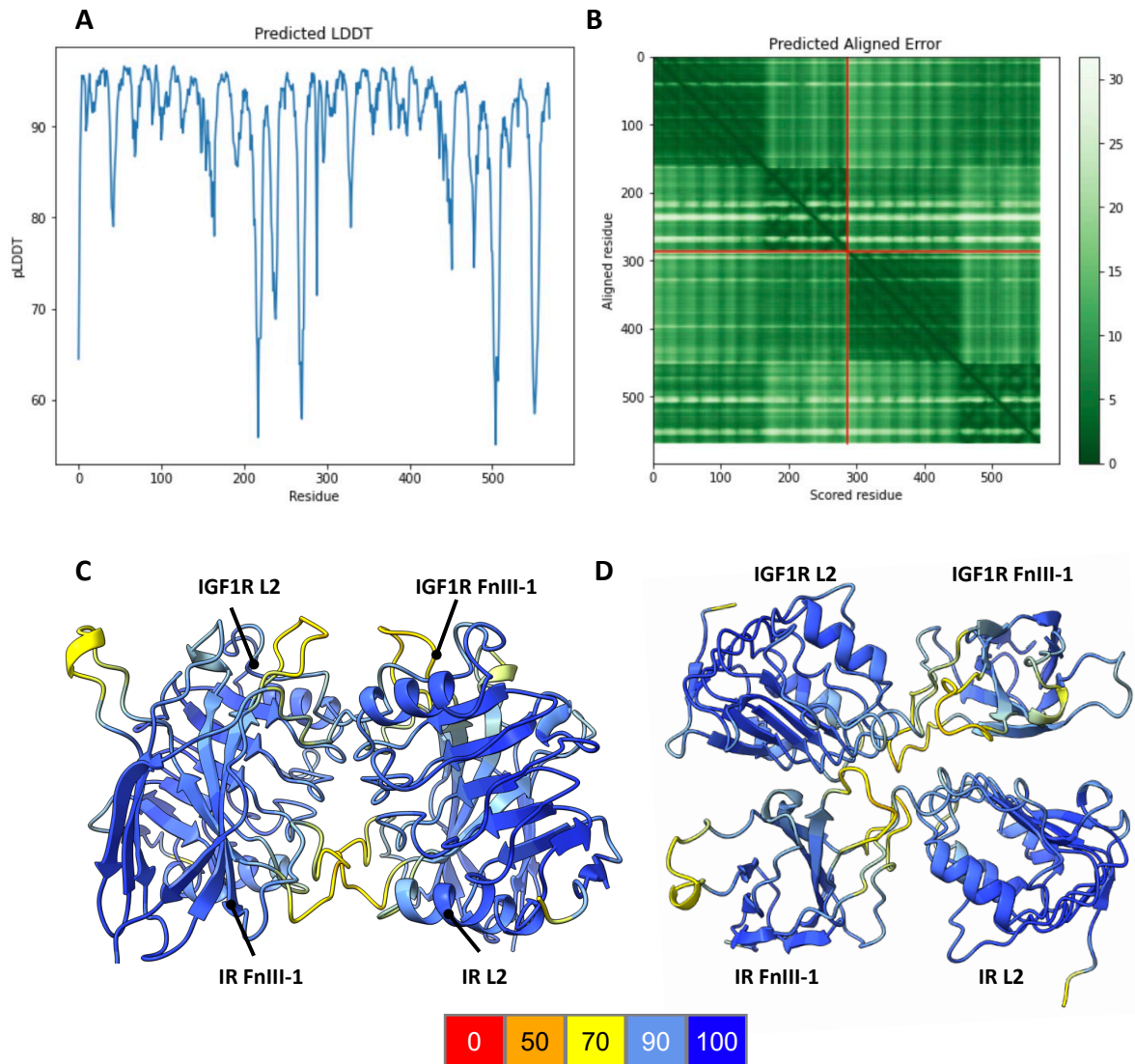
Figure 25 - AlphaFold homology model of the L2: FnIII-1 interface. A) Plot of the per residue pLDDT score; B) Plot of the PAE for the homology model. C) Face on view of the Alpha fold homology model with the ribbon coloured by pLDDT (colour key at bottom) D) Top-down view of the AlphaFold homology with ribbon coloured by pLDDT (colour key at bottom).

structure with pLDDT higher than 70 are expected to be modelled well, whilst regions with pLDDT higher than 90 should be modelled with very high accuracy.

The pLDDT for the AlphaFold model is generally high (Figure 25A), with score less than 70 isolated to the interdomain linker regions (IR 491-494, IGF1R 480-484; model numbering 159-162, IGF1R 154-158) and flexible loops on the edges of the FnIII-1 domain (IR 543-555, 599-605, IGF1R 538-548, 586-595; model numbering: IR 211-223, 267-273, IGF1R 212-222, 260-269)(Figure 25C,D). AlphaFold also produces a predicted aligned error (PAE); it generates a 2D plot where the colour at a given point represents the expected positional error in residue X if the predicted and true structures are aligned on residue Y. This is a measure of the confidence in the

relative positioning and orientation of domains in AlphaFold prediction. The PAE for the AlphaFold model is generally lower than 10 Å (Figure 25B), indicating the confidence in the relative positions of the domains is high. This is reassuring, as the L2: FnIII-1 PPI is a multidomain interaction and therefore accurate modelling is reliant on the correct positioning of both domains.

To allow direct comparison with the TACOS model, the AlphaFold model was also evaluated with the SWISS-MODEL structure assessment suite[136]. The AlphaFold model demonstrated a global QMEANDisCO score for the model was 0.78 ± 0.05, indicating a slight improvement compared to the TACOS model. Consistent with the
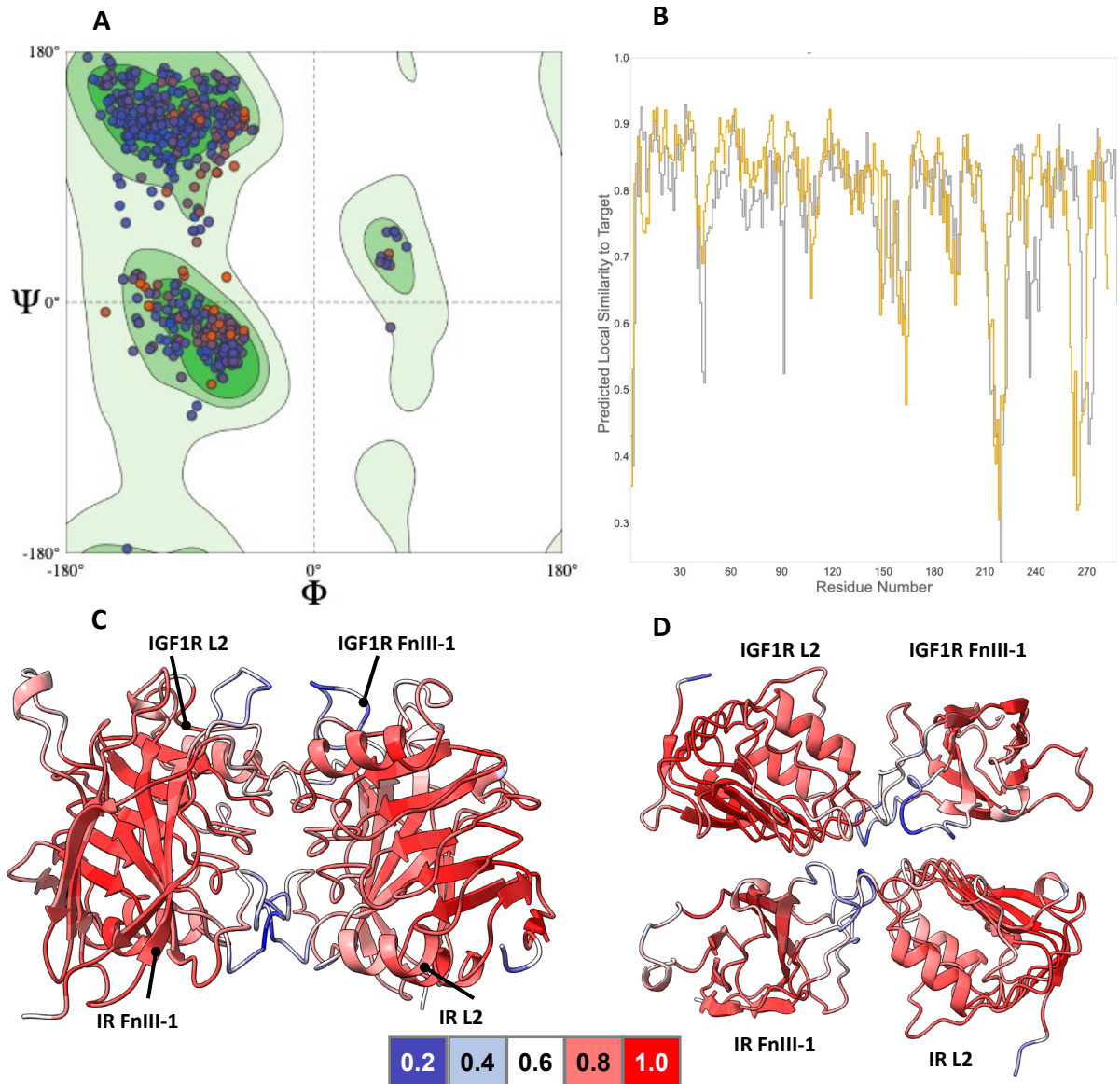
Figure 26 - AlphaFold homology model validation with SWISS-MODEL structure assessment; A) Ramachandran plot of AlphaFold homology model; B) Plot of the per residue QMEANDisCO score. The IR monomer (grey) and IGF1R monomer (orange) are both shown; C) Face on view of the AlphaFold homology model coloured by QMEANDisCO (colour key at bottom); D) Top-down view of the AlphaFold homology model coloured by QMEANDisCO (colour key at bottom).

pLDDT score, per residue QMEANDisCO analysis identified lower quality regions of the FnIII-1 domains in both the IR and IGF1R domains (Figure 26B-D). Importantly, these regions were confined to the same loop regions identified by the pLDDT score. This implies that the β5 and β6 strands comprising the core of FnIII-1 domain are predicted with significantly higher confidence in the AlphaFold model relative to the corresponding region of the TACOS model.

Ramachandran analysis of the AlphaFold model determined 96.29% residues occupied favoured conformations, with 0.53% outliers (Figure 26A). In comparison,

the corresponding values for the TACOS model were 92.11% and 0.72% respectively. Overall, assessment of the AlphaFold and TACOS model indicates that the AlphaFold shows a modest improvement in structural quality, particularly in the β5 and β6 strands of the FnIII-1 domains.

2.4.3 AlphaFold: Hotspot Prediction

KFC2 hotspot prediction server was utilised to predict hotspot residues of the AlphaFold homology model. The KFC2a analysis predicted nine IR and nine IGF1R residues as hotspots respectively, with KFC2b predicting the six and five hotspot residues for the IR and IGF1R monomers respectively (Table 2). These residues are

Chain	Amino acid	Residue	KFC2a	KFC2a Conf	KFC2b	KFC2b Conf
IR	ARG	398	Hotspot	0.47	Hotspot	0.23
IR	ARG	399	Hotspot	1.11	Hotspot	0.15
IR	TYR	428	Hotspot	0.17	Hotspot	0.14
IR	LEU	430	Hotspot	1.43	Hotspot	0.29
IR	TYR	457	Hotspot	0.99	Hotspot	0.11
IR	LYS	487	Hotspot	0.53	-------	-0.52
IR	GLN	492	Hotspot	0.33	-------	-0.35
IR	LEU	528	Hotspot	0.57	-------	-0.13
IR	PHE	530	--------	-0.52	Hotspot	0.04
IR	LEU	596	Hotspot	0.85	-------	-0.02
IGF1R	ARG	391	Hotspot	0.46	Hotspot	0.13
IGF1R	HIS	392	Hotspot	0.85	Hotspot	0.17
IGF1R	LEU	423	Hotspot	0.93	Hotspot	0.24
IGF1R	PHE	450	Hotspot	1.16	Hotspot	0.23
IGF1R	ASN	478	Hotspot	0.13	-------	-0.5
IGF1R	ARG	485	Hotspot	0.43	-------	-0.18
IGF1R	ILE	521	Hotspot	0.1	-------	-0.29
IGF1R	SER	522	Hotspot	0.42	-------	-0.42
IGF1R	ASP	555	Hotspot	0.62	Hotspot	0.05

Table 2 - Hotspot residues predicted by KFC2 for the AlphaFold homology model. Residues also predicted by KFC2a in the TACOS homology model are highlighted in blue.

clustered towards the top of L2 and FnIII-1 domains (Figure 27A, B), forming to relatively localised epitopes.
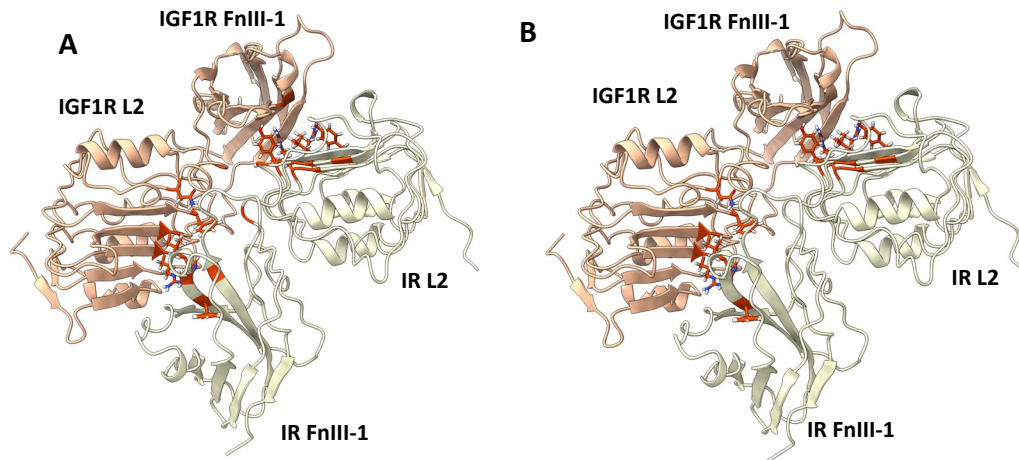
Figure 27 - KFC2 hotspot prediction for the AlphaFold homology models of the L2: FnIII-1 hybrid interface. The IR monomer is depicted in yellow, the IGF1R monomer depicted in orange and the hotspot residues depicted in red. A) Hotspot residues predicted by KFC2a on the AlphaFold homology model; B) Hotspot residues predicted by KFC2b on the AlphaFold homology model.

The interactions mediated by the hotspot residues exhibits a degree of symmetry between the two receptor chains, although specific interactions are not always conserved. For instance, IR Y457 and IGF1R F450 establishes reciprocal interactions by inserting hydrophobic regions between the L2 and FnIII-1 domains (Figure 28A). Similarly, both IR R398 and R399 (Figure 28B), as well as IGF1R R391 and H392 (Figure 28C) insert into charged grooves on the opposite FnIII-1 monomer to form
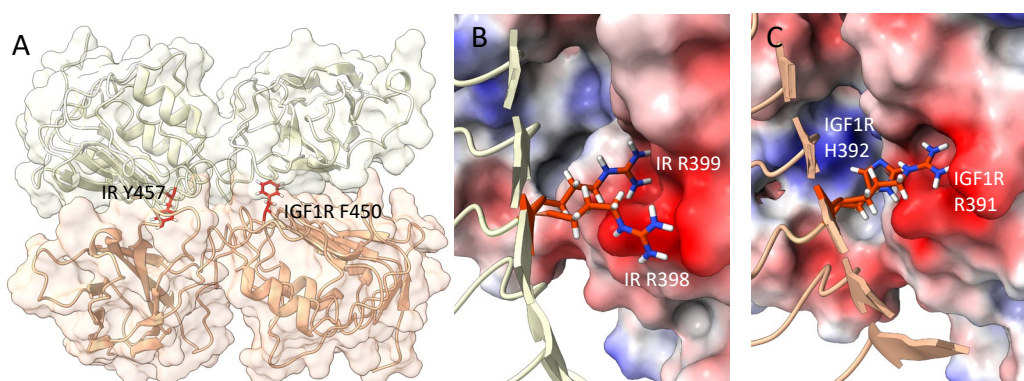
Figure 28 - Key hotspot interactions at the AlphaFold hybrid interface; A) The positions of IR (yellow) Y457 (red) and IGF1R (yellow) F450 (red), showing their reciprocal interactions from each monomer; B) Close up view of hotspot residues IR R399 and R398. The IR monomer is shown with a yellow ribbon, and the IGF1R monomer is shown with an electrostatic surface; C) Close up view of hotspot residues IGF1R R391 and H392. The IGF1R monomer is shown with an orange ribbon, and the IR monomer is shown with an electrostatic surface.

electrostatic interactions. Additionally, both IR L196 and IGF1R I295 were involved in packing against hydrophobic regions on the opposite monomer.

Three hotspot residues from each chain were predicted by KFC2 for both the TACOS and AlphaFold model. These residues include IR Y457, IR K487, IR Q492, IGF1R F450, IGF1R N478 and IGF1R R485, which are all contained in the L2 domain. It is notable that there are discrepancies in the hotspot residues predicted within the FnIII-1 domains. This disparity can likely be attributed to reduced quality of this region in the TACOS model, indicating that the FnIII-1 residues predicted by the AlphaFold model should be prioritised over those from the TACOS model. The contrasting predictions of hotspot residues between the AlphaFold and TACOS models highlights the variations in local structural details despite the high RMSD indicating good overall agreement between the two models. This reinforces the importance of considering multiple modelling approaches and prioritising regions of higher confidence when utilising results obtained from homology models.

Querying of the predicted AlphaFold hotspot residues with HGMD[156] determined only the IR point mutation K487E was associated with disease. This was similarly identified from the TACOS model and is implicated in Donohue Syndrome[157].

2.4.4   AlphaFold: Virtual High-Throughput Screening

Similar to the TACOS model, the AlphaFold model was used for docking based-virtual high throughput screening using the same methodology. Again, the Schrodinger suite was utilised: protein preparation performed by the Maestro

Protein Preparation wizard[162, 163]; SiteMap[161] utilised to identify a suitable binding pocket and Glide[123-125] for grid preparation and docking.

2.4.5 AlphaFold: Binding Site Detection & Grid Generation

IR L2   IR FnIII-
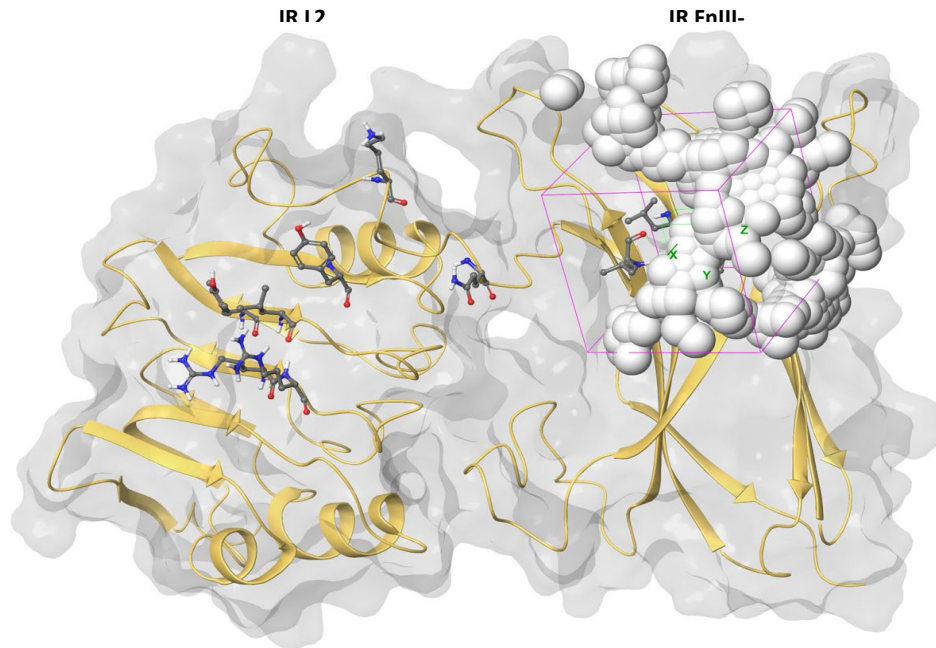
Figure 29 - The IR monomer from the AlphaFold homology model (gold, grey surface) showing the ligand binding site detected by SiteMap (white spheres) and the location of the grid generated for docking (pink box). Hotspot residues are shown in grey stick representation.

Binding site detection using SiteMap did not detect an analogous groove to that observed on the IGF1R L2 domain of the TACOS model. Instead, the highest scoring site on the in the AlphaFold model detected in the groove between the IGF1R L2 and FnIII-1 domains, with a site score of 0.962. However, this site was based in an area pLDDT scores were lower (<90%), and it was anticipated that the linking regions between the domains were likely to be flexible and dynamic. Therefore, this site was avoided as it was unlikely that this region was accurately modelled. Instead, a region was identified on the IR FnIII-1 domain, which comprised of several pockets and groves dimpled on the domain surface, linking together into a larger binding region (Figure 29A, B). The SiteScore for this region was 0.852 over 89.5 $Å^3$, which is deemed acceptable as a score of 0.80 represents a druggable site. Whilst only 3 of the IR hotspot residues were in contact with this detected site, several of the IGF1R hotspot residues bound this region with reciprocal interactions, meaning binding to this region would likely disrupt the IR-IGF1R PPI. A

20x20x20 Å grid was generated directly from the site points comprising this binding site, and this was subsequently utilised for vHTS.

2.4.6 AlphaFold: Screening Triage

Screening of the AlphaFold homology model utilized the same vHTS methodology as performed with the TACOS model. The lowest scoring poses were -5.505 and -6.156 for the Enamine PPI library and MCCB library respectively. Markedly, the diversity library tended to score more favourably than the focussed PPI library. Additionally, it was noted that the scores are higher than those obtained in the TACOS screen, suggesting the identified poses are less favourable. This was somewhat anticipated, as the epitope identified by SiteScore for the AlphaFold
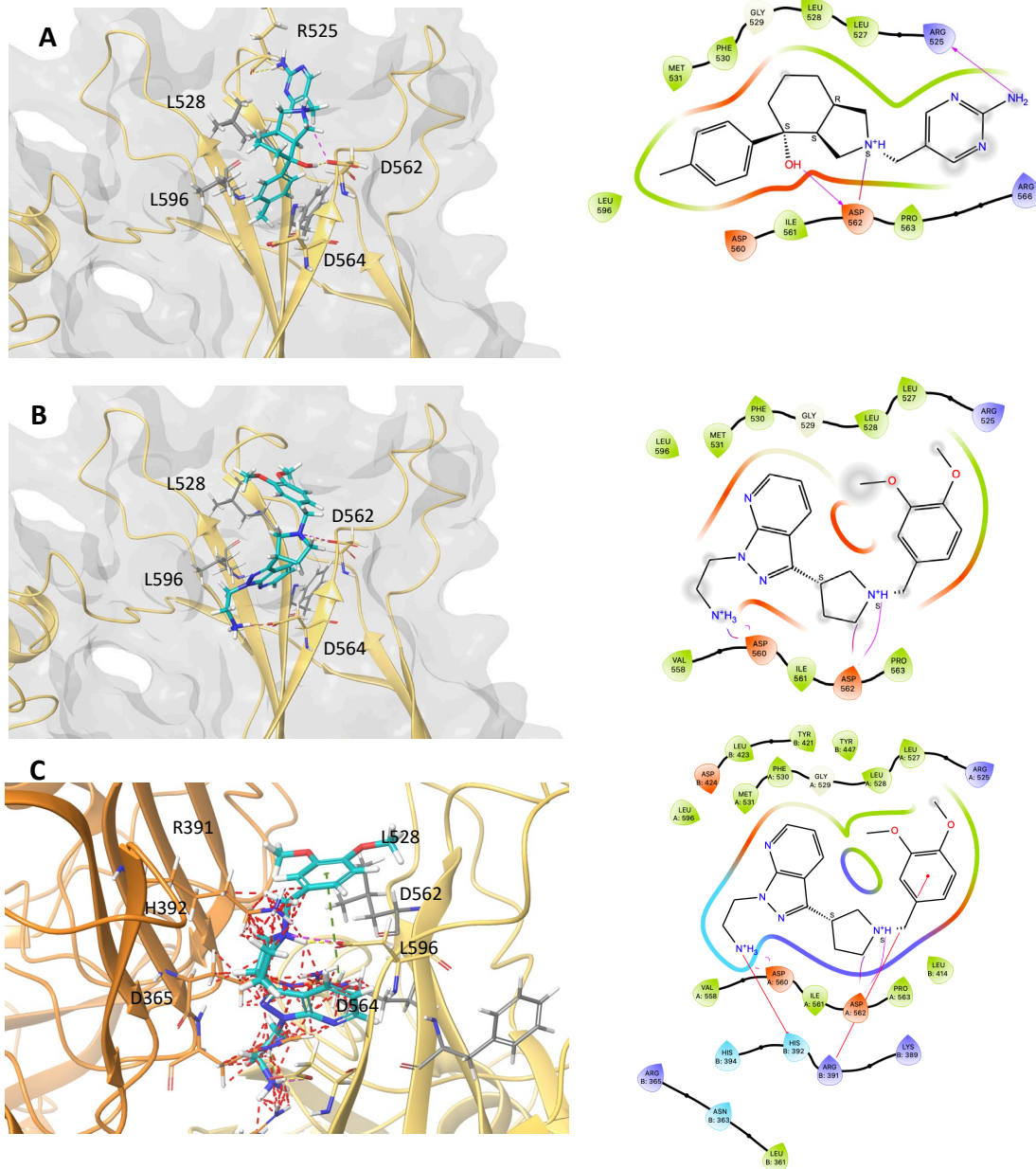
Figure 30 - Docked poses of ligands bound the IR. 3D images (left) and 2D representations (right) are shown. In 2D diagrams, chain A refers to the IR and chain B refers to the IGF1R. The IR (gold ribbons, grey surface), hotspot residues (grey residues) and ligand (cyan) are depicted with H-bonds (yellow), π-stack (sky blue) and electrostatic interactions (purple) shown. For E & F the IGF1R monomer (orange) is superimposed, with clashes shown (orange and red lines): A) Ligand AEM14664198; B) Ligand SFA22637552; C) As in B, but with the IGF1R monomer superimposed.

model was lower scoring than the site identified for the TACOS model, indicating that the AlphaFold site is less attributed to ligand binding.

Many high scoring poses featured polar H-bonds or electrostatic interactions with D562 (Figure 30A, B). Although this residue was not predicted as a hotspot, such poses typically featured hydrophobic contacts with hotspot residues L528 and L596

(Figure 30A, B). At the top of the binding site poses typically featured H-bonds with backbone C-O of R525, enhancing polar interactions (Figure 30A). Superimposing the IGF1R monomer revealed clashes with hotspot residues R391 and H392, which were predicted to insert into the ligand-occupied grooves. Additional clashes were often observed with D365 (Figure 30C). Consequently, the most promising compounds were selected for biological evaluation based on their predicted ability to inhibit the IR-IGF1R PPI.

2.5   Ligand Based Screening to Design HI2 Analogues.

As an alternative strategy to identify novel modulators of IR-IGF1R hybrid formation, a ligand-based screening approach focussed on the **HI2** scaffold was used to identify alternative chemical scaffolds. The small-molecule **HI2** has previously been shown to reduce hybrid formation by around 50% in HUVECs when tested at concentrations of 100 μM[113]. However, **HI2** suffers from poor aqueous solubility, making it unamenable to structural or *in vivo* studies. The scaffold hopping tools ROCs[168, 169] and EON[170] were utilised in the design of **HI2** analogues with the potential for improved ADME properties.

This workflow was used to identify small molecules with similar shape and electrostatic properties to **HI2**, which are therefore predicted to similarly inhibit hybrid formation. For the input library, the Chembridge diversity library was chosen as the large variety of chemical structures included in this library provided improved chance of identifying alternative scaffolds. Focus was placed upon altering the quinoline core of **HI2**, as this moiety had not been varied in previous analogues. Previous SAR studies of the HI series had established that the molecule could only tolerate limited modifications whilst retaining its biological activity. Therefore, the triazole and tolyl portions of **HI2** were retained to avoid large scale

structural changes. This approach identified several alternative scaffolds (Figure 31) which were targeted for synthesis and testing in Section 3.3
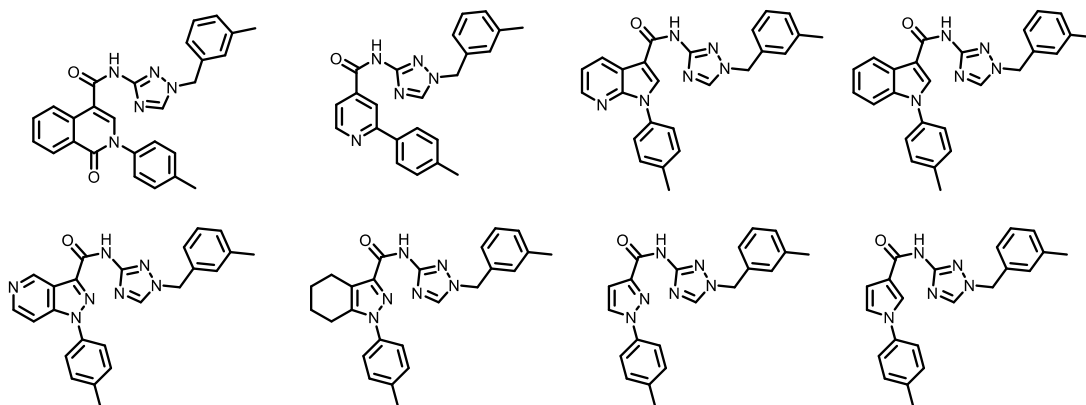
Figure 31 – Structures of synthetic targets derived from ligand-based screening of HI2

2.6 Conclusions

In this chapter, two homology models of the IR-IG1F1R hybrid ectodomain were generated utilising TACOS and AlphaFold Multimer. These methodologies have independently predicted the formation of an interface between the L2 and FnIII-1 domains in the hybrid receptor. This interface was evaluated by SiteMap and KFC2, identifying suitable drug-binding regions and hotspot residues for small molecule intervention. A virtual high throughput screen of 67,000 ligands was targeted to these hotspot epitopes, and molecules predicted were prioritised for biological evaluation according to their predicted ability to inhibit hybrid formation. Additionally, a ligand-based screening method has been utilised to design HI2 analogues which are similarly expected to inhibit hybrid formation. Biological evaluation of these molecules will be discussed in Chapter 3.

A limitation of this approach is that the large size of the IR-IGF1R hybrid receptor precluded modelling of the entire receptor. Whilst it is likely that interactions between individual receptor domains is somewhat independent from the entire receptor, a complete model of the receptor would be preferable.

It was also notable that the AlphaFold and TACOS models showed limited agreement on the identity of hotspot residues when evaluated by KFC2. To assess which model is the most accurate, further experimental validation would be required. The experimental validation of hotspot residues by point mutagenesis will be a focus of Chapter 4.

2.6.1 Progress Towards Project Goals

This chapter has made significant progress towards the overall project goals 1 and 2. Firstly, the generation of two updated homology models which incorporate novel structural information from published structure of the IGF1R satisfies the first aim of the project. Secondly, significant progress has been made towards aim 2, by using vHTS to identify putative small molecule modulators of hybrid formation. Both criteria provide insight into the molecular mechanisms that govern hybrid formation.

3 Synthesis and Biological Evaluation of Predicted Hybrid Modulators

3.1 Introduction

3.1.1 Evaluation of Protein-Protein Interactions using a Bioluminescence Resonance Energy Transfer Assay

The transient nature of many protein-protein interactions (PPIs) necessitates the use of specific assay techniques for their characterization and monitoring. Amongst the most popular of these techniques are those based on resonance energy transfer (RET). RET relies on the interaction of a compatible energy donor and energy acceptor, typically fluorescent tags fused to the proteins of interest[171]. For RET compatibility, the emission spectrum of the energy donor must overlap with the excitation spectrum of the energy acceptor. RET based techniques offer several advantages over alternative methodologies such as co-immunoprecipitation or proximity ligation assays, including high sensitivity, suitable throughput, and compatibility with both biochemical and cellular assay screening formats[172].

At a given temperature, the intensity of the RET interaction depends on three parameters[172, 173]. Firstly, the transfer signal is inversely proportional to the distance between the RET pair, up to a maximal distance of 10 nm. This parameter is particularly relevant to monitoring PPIs as RET assays typically aim to monitor the distance between two proteins. Second - the relative orientations of the RET partners, as RET is reliant on dipole-dipole interactions. This highlights the importance of any linker regions fusing the RET partners to their respective proteins of interest. Linkers should be flexible to allow the RET partners to adopt an appropriate orientation for resonance. Finally, RET is proportional to the donor/acceptor ratio in each experiment. PPIs occur in finite spaces, whether in the cellular environment or *in vitro*. As the concentration of two proteins rises, the probability of them coming into proximity also rises, even in the absence of a specific interaction. Although non-specific interactions tend to be weak, they rise linearly with increasing concentrations. Therefore, any RET experiment is subject to

the appropriate controls to ensure that the measured interaction is specific between the protein pair.

 Two of the most popular RET techniques are Forster Resonance Energy Transfer (FRET) and Bioluminescent Resonance Energy Transfer (BRET)[174]. FRET relies on external excitation to stimulate donor emission (Figure 32A), whilst in BRET the donor is a bioluminescent enzyme. These enzymes are typically luciferases, which catalyse the oxidation of a substrate molecule resulting in the emission of a photon to promote RET with a nearby acceptor molecule (Figure 32B). The main advantage of BRET is that it eliminates the need for external excitation, resulting in reduced donor photobleaching, cell autofluorescence and background excitation. This is of utility in cell-based formats, as BRET is sensitive at endogenous levels of protein expression, minimising artefacts associated with overexpression. BRET is adaptable to high-throughput drug screening assays, and has been successfully employed in the identification of PPI[175] and small molecule protein modulators[176]. Therefore, BRET appeared a promising technique for assessing the impact of our predicted small-molecule inhibitors on their ability to modulate IR-IGF1R hybrid formation.
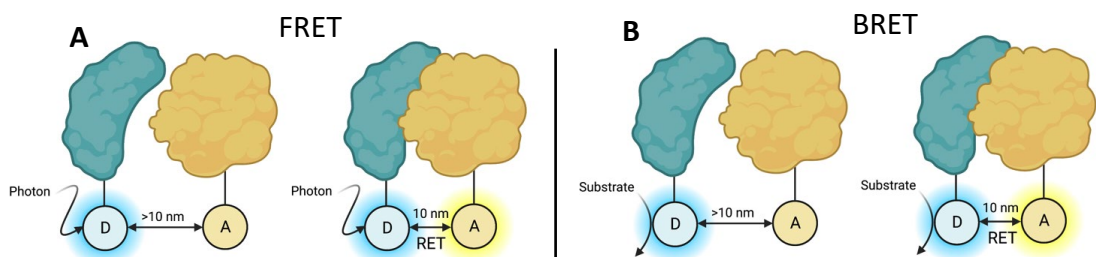
Figure 32 - The principle of FRET and BRET assays. Two representative proteins are shown in blue and yellow. The labels D and A indicate RET donor and acceptor tags respectively; A) Schematic diagram of FRET interaction between two tagged proteins; B) Schematic diagram of BRET interaction between two tagged proteins. Created with BioRender.com.

3.2   BRET Assay to Assess Modulation of the IR-IGF1R PPI

To evaluate the effect of potential bioactive small molecules on hybrid formation, a cell-based bioluminescence resonance energy transfer screening assay based on the system previously described by Blanqaurt *et. al*[177] was utilised. This assay measures

the spatial proximity of two proteins and can be used to evaluate the interaction between the IR and IGF1R in a hybrid receptor.

In this assay, HEK293 cells were co-transfected with cDNAs encoding the IR fused to *Renilla Luciferase* (*Rluc*) and IGF1R fused to second generation yellow fluorescent protein (YPET). These plasmids were a gift from the Issad lab (INSERM, Paris). The fluorescence of YPET is dependent on a resonance energy transfer interaction between the *Rluc* and YPET tags when in spatial proximity and the presence of the *Rluc* substrate coloenterazine-h. The BRET assay can differentiate between homo and heteromeric receptors, as only hybrid receptors containing both the Rluc and YPET tags are BRET competent (Figure 33). BRET signal is expressed as a ratio of the signal corresponding to the luminescence of *Rluc* at 485 nm and the luminescence of YPET at 535 nm. To account for the spectral overlap of Rluc and YPET luminescence, the background signal is determined by transfecting the *Rluc* tagged receptor alone. This is used to determine the background luminescence ratio
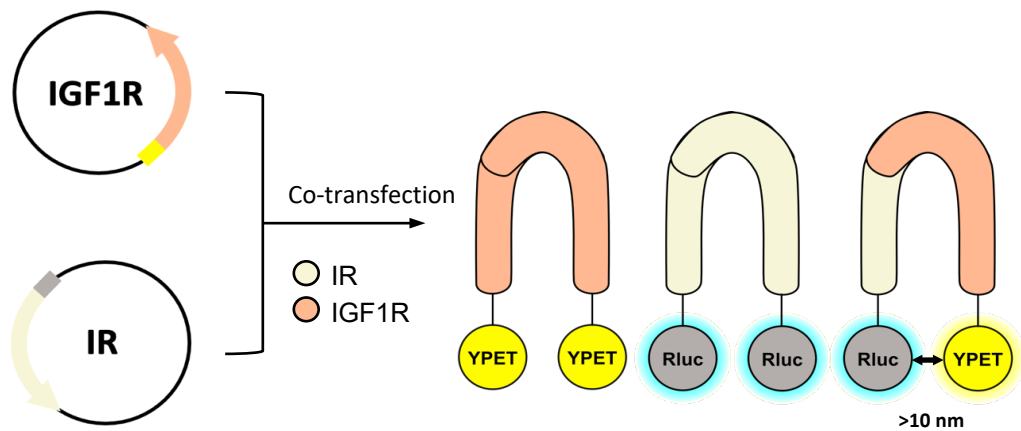
Figure 33 -Schematic of BRET assay to specifically detect IR-IGF1R hybrid formation in live cells. Plasmids encoding the IR and IGF1R tagged with Rluc and YPET respectively are transfected into HEK293 cells, allowing detection of hybrid receptors by a RET interaction between the Rluc and YPET tagged monomers.

between 485 nm and 535 nm and subtracting this from the BRET ratio calculated for both partners.
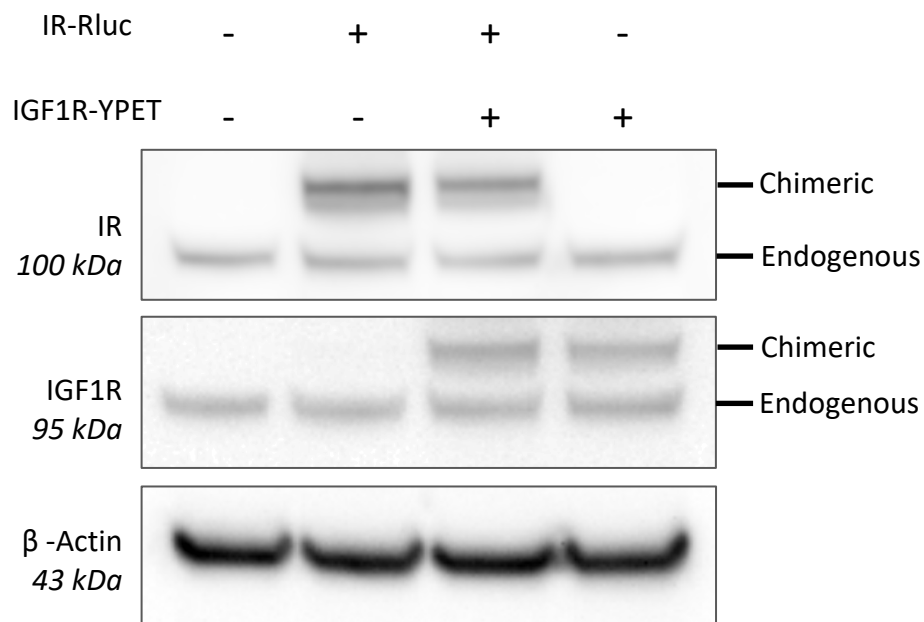
| IR-Rluc | - | + | + | - |
| IGF1R-YPET | - | - | + | + |

Figure 34 – Representative western blot of cell lysates from HEK293 cells transfected with IR-Rluc and IGF1R-YPET. Blots were probed with the IR specific antibody 48B and IGF1R specific antibody D23H3

Initial validation of the BRET assay was performed by transfecting the IR-Rluc and IGF1R-YPET constructs into HEK293 cells and assessing their expression by western blotting (Figure 34). Under reducing/denaturing conditions, the IR and IGF1R separate into their α and β subunits. Western blotting with the IR-β specific antibody (48B, Cell Signalling) revealed bands corresponding to the expected molecular weight of the IR-β subunit tagged with *Rluc* when IR-Rluc was transfected alone or in combination with IGF1R-YPET. Similarly, probing with the IGF1R antibody (D23H3, Cell Signalling) showed a band corresponding to the IGF1R-β subunit fused with YPET when IGF1R-YPET was transfected alone or in combination with IR-Rluc. This indicates robust expression of the constructs, which expressed at similar levels to the endogenous receptors in HEK293 cells when transfections were performed with 0.3 µg of DNA per well. Therefore, the chimeric receptors are expressed at a physiologically relevant level, avoiding artefacts associated with protein overexpression.

The specificity of the BRET interaction between the IR-Rluc and IGF1R-YPET constructs was confirmed by a donor-saturation assay, in which the ratio of the
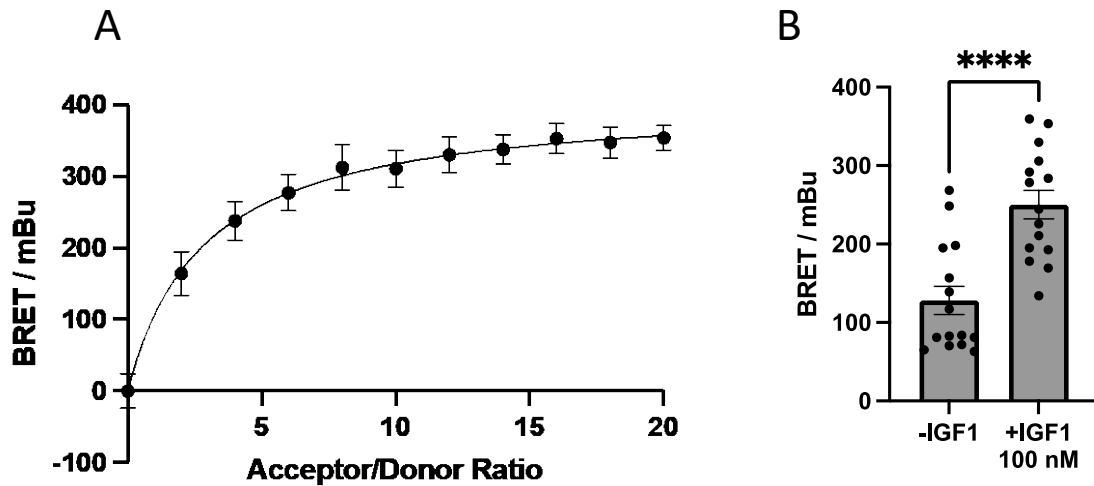
Figure 35 - Validation of BRET assay: A) Donor saturation assay curve for BRET constructs IR-Rluc and IGF1R-YPET (±SEM); (B) BRET measurements showing the effect of treating cells co-transfected with IR-Rluc and IGF1R-YPET with IGF1 (100 nM, 5 min treatment, n=8,2; p<0.0001 ±SEM)

BRET acceptor (IGF1R-YPET) was varied relative to the BRET donor (Figure 35A). The hyperbolic increase in BRET signal with increasing acceptor/donor ratio is typical of a specific BRET interaction, with the final plateau representing the saturation of all donor molecules. This contrasts with the expected donor-saturation relationship for non-specific interactions, where increasing the acceptor/donor ratio results in a linear increase in BRET signal[172]. From this experiment it was concluded that the BRET signal observed between IR-Rluc and IGF1R-YPET indeed arises from a specific BRET interaction between the two constructs and is not due to random collisions between BRET partners in the cell. Therefore, the BRET signal resulting from these constructs accurately represents the formation of hybrid receptors.

Additionally, in cells transfected with IR-Rluc and IGF1R-YPET, an enhanced BRET signal could be observed upon stimulation with 100 nM of IGF1 (Figure 35B). This BRET increase was comparable in magnitude to that previously reported in hybrid receptors similarly tagged with Rluc and YPET[177]. Cells were serum-starved for 24 hrs prior to IGF1 treatment, to minimise background IGF1R activation from IGF1 present in serum. The increase in BRET signal in response to IGF1 is consistent with dimerisation of the IR and IGF1R kinase domains; this response is conserved in the activation of each receptor by its primary ligand. Complexation of the kinase domains in the chimeric hybrid receptors reduces the distance of the Rluc and YPET tag, which in turn will increase resonance between the BRET pair. Therefore, this

experiment confirms the chimeric receptors are activated by IGF1, resulting in functional activation of the receptor kinase domains.

Finally, the subcellar location of the chimeric IR-IGF1R hybrid receptors was assessed to ensure they were correctly trafficked to the cell membrane. Using fluorescence microscopy, YPET signal was observed localised to the cell membrane in HEK293 cells co-transfected with IR-Rluc and IGF1R-YPET (Figure 36A). This was readily distinguishable from the signal observed in the same cells transfected with YPET, which is not trafficked to the cell membrane and primarily located in the cell cytoplasm. This result confirms that the attachment of BRET tags has not hindered the robust processing and trafficking of the chimeric hybrid receptors.
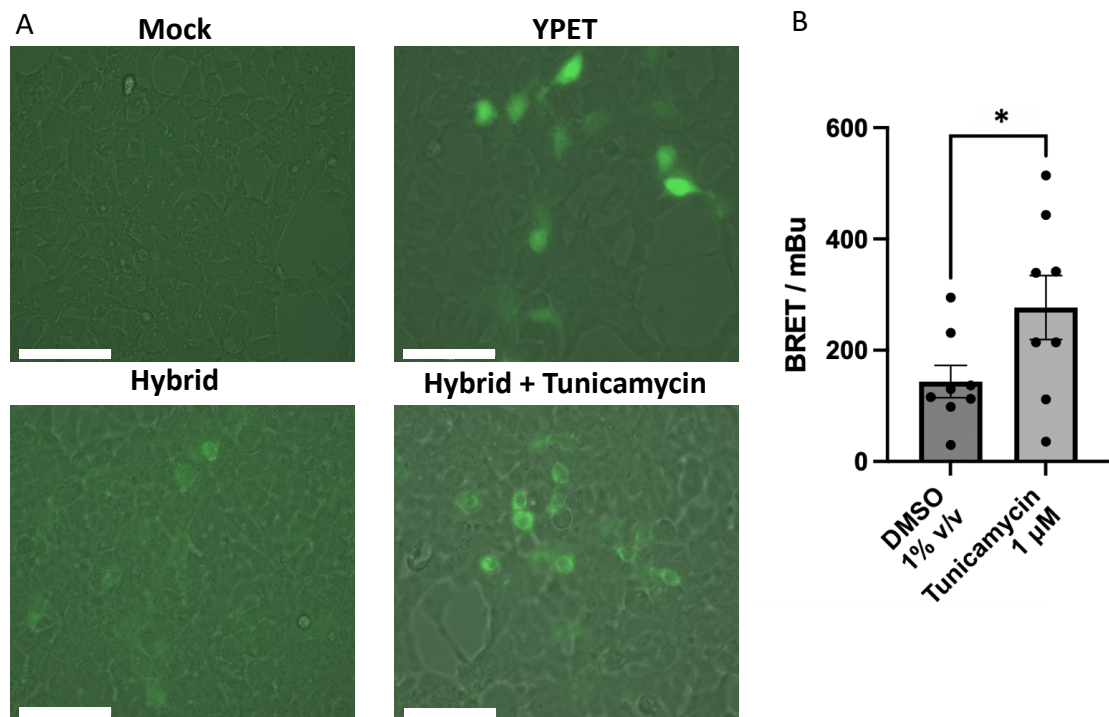
Figure 36 – Subcellular location of the IR-Rluc and IGF1R-YPET: A) Representative fluorescence microscopy images of HEK293 cells transfected as indicated, indicating the subcellar location of YPET, scale bar = 75 µm; B) BRET signal of HEK293 cells treated with Tunicamycin (n=8, p<0.05, ± SEM).

Additionally, the impact of ER stress on the trafficking of chimeric IR-IGF1R hybrid receptors to the cell membrane was assessed, as ER stress has been reported to inhibit the delivery of newly synthesised insulin receptors to the cell membrane[178]. ER stress was induced by treatment with the nucleoside tunicamycin, which is known to block N-linked glycosylation and result in subsequent ER stress. Treatment of HEK293 cells with 1 µM of Tunicamycin for 24 hrs induced a large

increase in BRET signal (Figure 36B). This is likely caused by accumulation of IR-Rluc and IGF1R-YPET in the ER, resulting in non-specific BRET signal from build-up of the chimeric receptors in the confinement of the endoplasmic reticulum. Accordingly, fluorescence microscopy of cells expressing chimeric receptors and treated with 1 µM tunicamycin showed build-up of yellow fluorescence in the endoplasmic reticulum (Figure 36A). These results indicate that agents inducing cellular ER stress result in a corresponding increase in BRET signal through accumulation of BRET constructs in the endoplasmic reticulum.

For screening purposes, the assay was performed with a 1:1 transfection ratio of IR-Rluc and IGF1R-YPET, following literature precedent[177] and allowing the detection of positive and negative modulation of the hybrid populations. Cells were treated with 100 µM of compound for 24 hours, chosen as an appropriate timeframe for compound treatments to affect hybrid levels given the half-life of the IR (*ca.* 7 hours)[179]. BRET signal from compound treated cells was compared to vehicle control and increases or decreases in BRET signal relative to the vehicle control were interpreted as a relative increase or decrease in hybrid receptor abundance due to compound treatment respectively.

3.3 Synthesis and Biological Evaluation of **HI2** Analogues
3.3.1 Synthesis of **HI2** Analogues

For initial screening, the **HI2** analogues identified by ligand-based screening (Section 2.5) were synthesised (Figure 37). Nitro-triazole **1** was alkylated with 3-methyl benzyl chloride **2** to give the nitro-triazole **3**, which was then subjected to
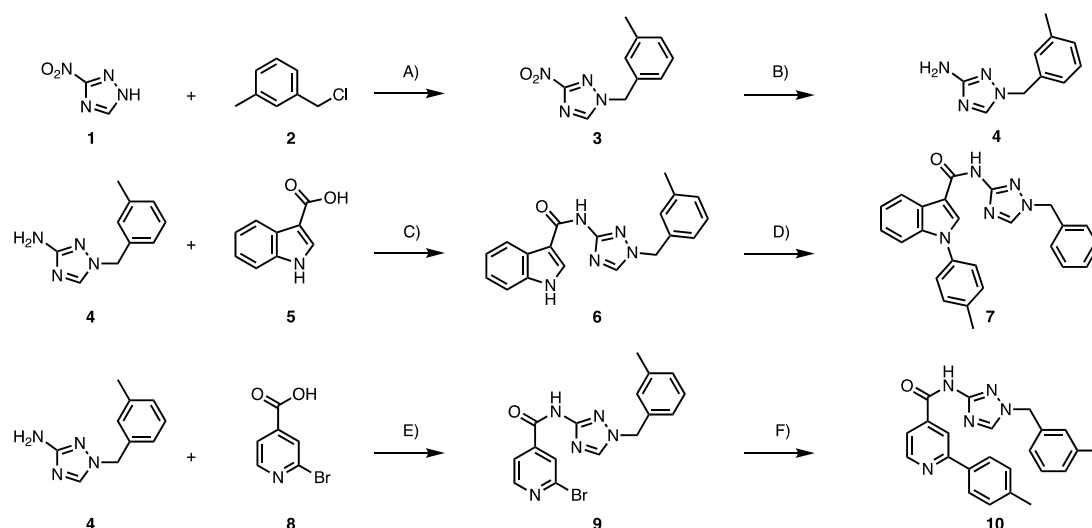
*Figure 37 - Synthesis of **HI2** analogues. a) NaOEt, EtOH, 80 °C, 18 h, 99% b) H$_2$, Pd/C, MeOH r.t., 8 h, 95% c) i) SOCl$_2$, 80 °C, 1 h ii) 4, NEt$_3$, DCM, r.t., 18 h, 74% d) 4-Bromo-toluene, CuI, N,N'-dimethylethylenediamine, K$_3$PO$_4$, DMF, 110 °C, 18 h, 17%, e) T$_3$P, NEt$_3$, EtOAc, 80 °C, 18 h, 59%, f) 4-Tolylboronic acid, PdCl$_2$(PPh$_3$)$_2$, CsCO$_3$, DMF, 100 °C, 18 h, 51%.*

nitro reduction to afford the amine **4**. This provided a key amine intermediate for successive coupling to core moieties derivatised with carboxylic acid in the appropriate positions.

Subsequently, **4** was coupled with indole-3-carboxylic **5** acid to generate the indole **6**. This reaction was sluggish when trialled with a range of coupling reagents, including T3P, HCTU and DCC/HOBT. TLC and LCMS analysis indicated that the carboxylic acid failed to be activated by these reagents. However, formation of the acyl chloride and subsequent addition of 4 proved to be successful, with the reaction proceeding in 74% yield under these conditions. A final Buchwald-type N-arylation with 4-bromotoluene afforded **7**. However, this reaction was hindered by the poor solubility of **6** in organic solvents even at elevated temperatures, resulting in poor yields of 17%.

For pyridyl analogue **10**, compounds **4** and **8** could be coupled using T3P, with final Suzuki reaction of **9** using 4-tolyl boronic acid affording **10**.

3.3.2 Biological evaluation of HI2 analogues

The **HI2** analogues were screened for inhibition of hybrid formation in the BRET assay (Figure 38A & B). Four additional **HI2** analogues were synthesised by Dr. Martin McPhillie and added to compounds **7** and **10** to give a total of six **HI2**
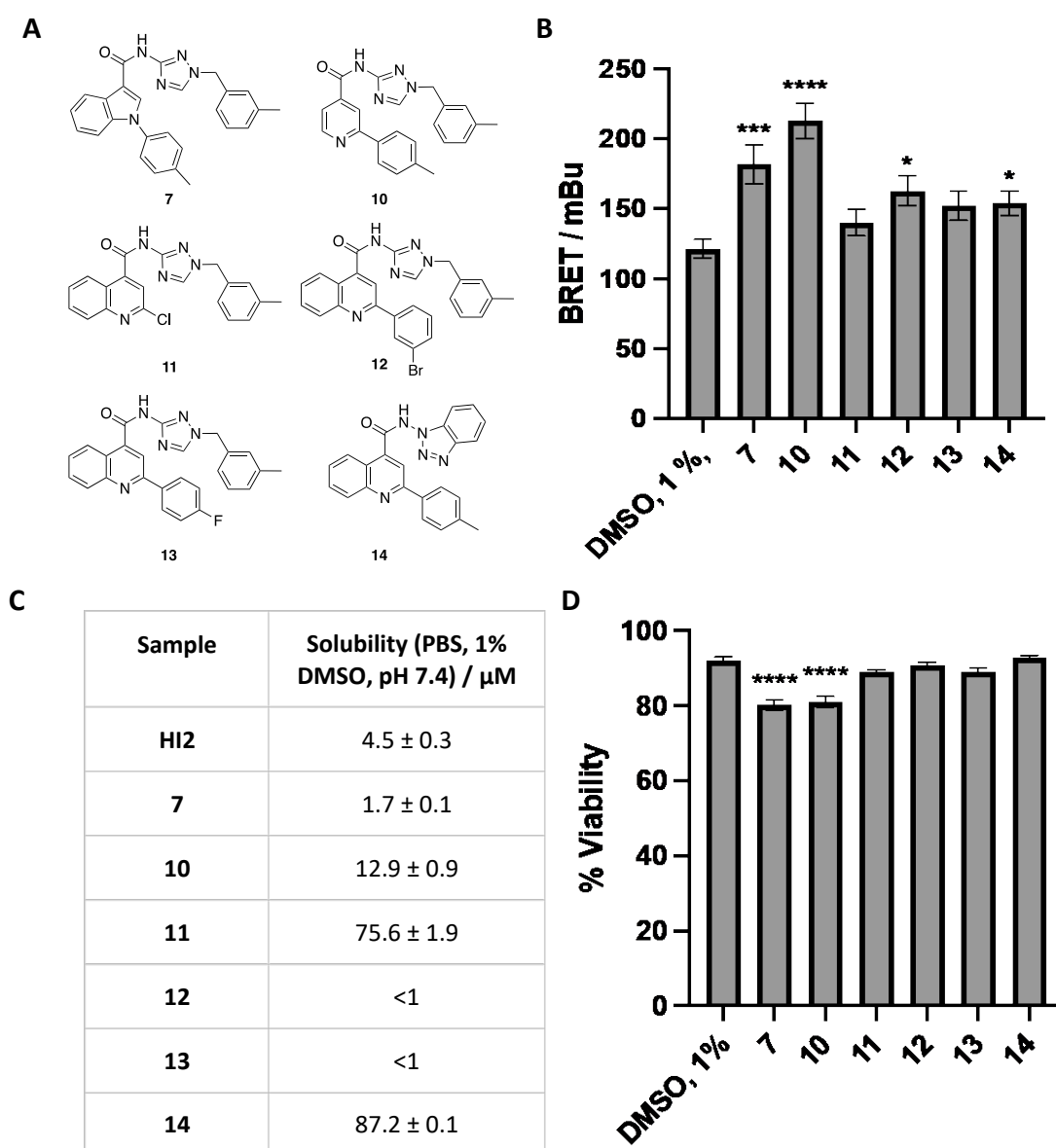
*Figure 38 - A) Structures of **HI2** analogues evaluated in the BRET assay. B) Average milliBRET (mBu) signal for **HI2** analogues tested in the hybrid BRET assay. (100 µM, 24 hr treatment, n= 8,3, ±SEM); C) Solubility values obtained for **HI2** analogues pH 7.4 phosphate buffered saline (n=3, ±SEM); D) Cell viability of **HI2** analogues tested in the LIVE-DEAD cell assay (n=8, ±SEM).*

analogues for testing. Compounds **7**, **10**, **12** and **14** showed a significant increase in BRET signal when compared to vehicle-control treated samples, with compound **10** showing 2-fold increase in BRET signal relative to the control sample. All other analogues showed no significant change in BRET signal. Therefore, BRET screening determined that none of the **HI2** analogues tested were inhibitors of hybrid formation.

The **HI2** analogues were also evaluated in a LIVE-DEAD mammalian cell cytotoxicity assay (Figure 38C). This two-colour fluorescence-based cell viability assay allows the

simultaneous determination of live and dead cells. Calcein AM and EthD-1 are utilised to measure esterase activity and plasma membrane integrity, which represent recognised parameters of cell viability. This assay determined **7** and **10** exhibit minor but statistically significant cytotoxicity in HEK293 cells at 100 μM, whilst the other **HI2** analogues have no significant cytotoxicity. It was noted that **7** & **10** similarly showed the largest increases in BRET signal of the **HI2** analogues tested in the BRET assay. Therefore, we were conscious that the increase in BRET signal from these two analogues may represent a cytotoxic artefact rather than representing an increase in hybrid formation post treatment, and it was decided not to pursue these compounds further.

3.3.3 Solubility of HI2 Analogues

The primary aim for ligand-based screening of **HI2** had been to generate analogues with improved solubility. To evaluate this, a HPLC solubility assay based on that reported by Belleni et. al[180] was utilised. This assay relies on the addition of pre-dissolved compound to buffered saline solution before incubation and equilibration at thermodynamic solubility. Solids were then pelleted before the concentration of the compound retained in the supernatant was determined by comparison of UV/Vis response to a calibration curve of the same compound. This assay was of utility as it allowed measurement of aqueous solubility using minimal quantities of screening compound pre-dissolved in DMSO, in comparison to alternative assays which required more than 1 mg of solid material.

In this assay, compounds **7**, **12** and **13** showed comparable or reduced solubility relative to **HI2** (Figure 38D). However, compound **14** showed approximately 45-fold increase in solubility over **HI2** through substitution of the benzyl-triazole tail with a benzotriazole unit. This suggests the low solubility of **HI2** is largely determined by the benzyl-triazole tail, rather than the quinoline core. In comparison, compound **10**, which substituted the quinoline core to a quinoline showed only modest 6-fold improvement in solubility, whilst substitution for an indole core as in compound **7** gave a reduction in aqueous solubility. Therefore, any design of further **HI2**

analogues should prioritise substitution of the benzyl-triazole portion of the molecule for maximal improvements in solubility.

3.4 Biological Evaluation of Molecules Identified using vHTS of the TACOS Homology Model

Additionally, compounds identified by vHTS of the TACOS homology model in section 2.3.7 were screened in the BRET assay described above to quantify hybrid formation post compound treatment. The LIVE-DEAD cytotoxicity assay was also utilised to evaluate compound cytotoxicity in HEK293 cells.

Of the 42 small molecules predicted by the TACOS model and screened in the BRET assay, none were found to significantly reduce BRET signal (Figure 39A). However, it
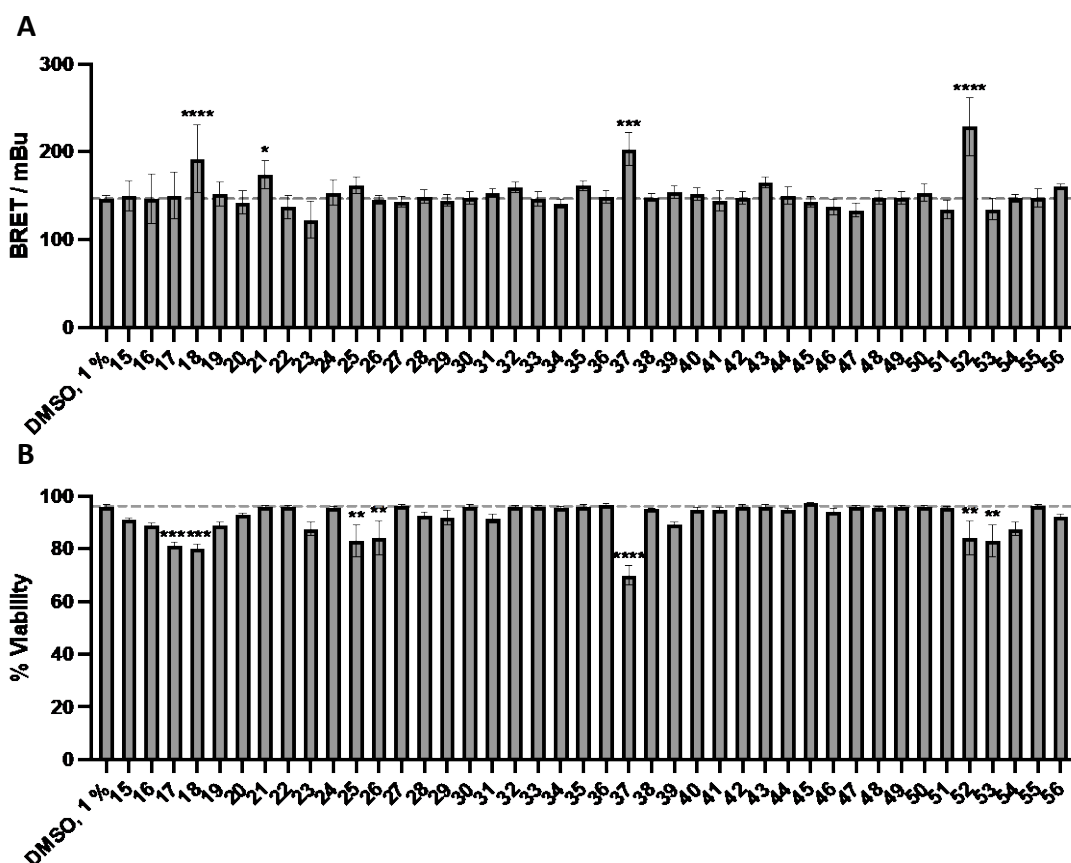
Figure 39 - Biological evaluation of small molecules prioritised from using vHTS of the TACOS homology model. Grey lines indicate mean value for DMSO control; (A) BRET measurements to quantify hybrid formation post-treatment with screening compounds (100 μM, 24 hr treatment, n=8,3; p<0.05, ±SEM); (B) LIVE-DEAD cell viability to determine cytotoxicity of screening compound (100 μM, 24 hr treatment, n=8,3; p<0.05, ±SEM).

was observed that several of the compounds increased BRET signal upon treatment (Figure 39A), despite these compounds being chosen on their predicted ability to

block hotspot residues at the IR: IGF1R interface. This appeared to be coupled with cytotoxicity in some cases; in compounds **18**, **37** and **52**, which show large increases in BRET signal but substantial decreases in cell viability at the same concentrations (Figure 39B). In these cases, we were wary of the increase in BRET representing a cytotoxic artefact rather than a *bona fide* interaction with the IR: IGF1R PPI. However, compounds such as **21** gave a smaller increase in BRET relative to **18**, **37** or **52**, but no corresponding cell toxicity (Figure 39). Therefore, it was decided further investigation into **21** was warranted, as this compound may be able to promote hybrid formation. From here on compound **21** will be referred to as **Z13** **(**Figure 40**)**.
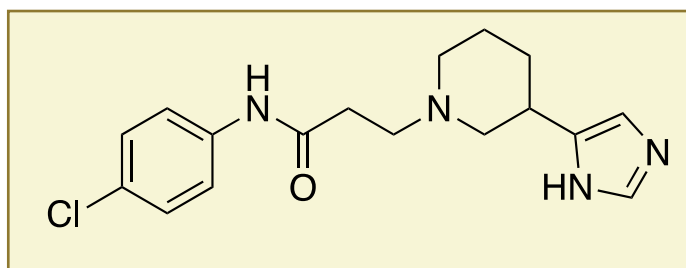
Figure 40 - Chemical structure of small molecule Z13 shown to promote hybrid formation at 100 µM in the BRET assay.

3.4.1 Further Evaluation of Small Molecule Z13

The predicted binding mode of **Z13** showed it bound into the groove of the IGF1R L2 and FnIII-1 domains, into the region occupied by IR K487 in the TACOS homology model (Figure 41). As this residue is part of an interdomain linker region, it

appeared plausible that K487 could shift to accommodate the binding of **Z13**, and this potentially results in a favourable interaction to stabilise the interface.

Whilst the increase in BRET afforded by **Z13** could represent a genuine interaction with the IR: IGF1R hybrid receptor, promoting heterodimer formation and a subsequent increase in BRET signal, there were other several alternatives plausible mechanisms for its observed increase in BRET signal: **Z13** could be interacting with the hybrid receptor in a way which alters the receptors conformation and bring the
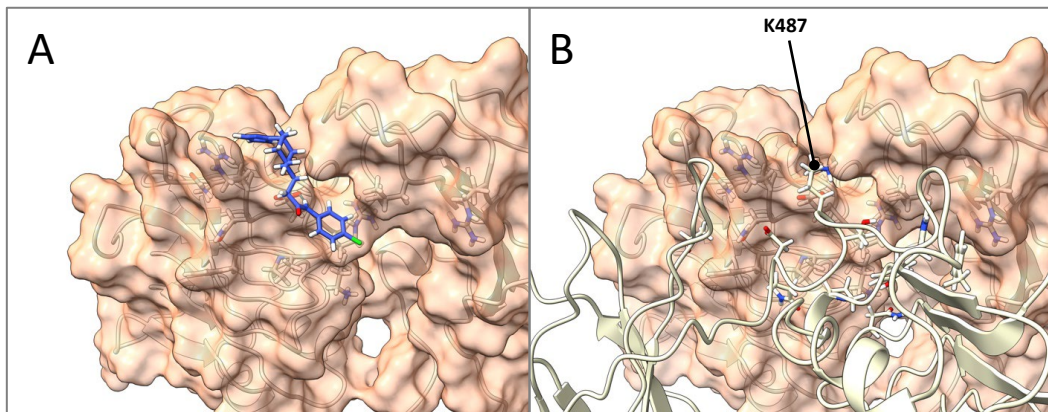
*Figure 41 - Predicted binding pose of small-molecule **Z13**; A) Docked pose of **Z13** (blue) with the IGF1R monomer from the TACOS homology model (orange); B) TACOS homology showing the IR (yellow ribbons), IGF1R (orange surface) and the orientation of K487.*

Rluc and YPET tags closer together; **Z13** may be acting indirectly on hybrid formation through altering gene expression or through any of the multitude of

compound interference affects possible in cell based screening assays such as compound fluorescence, aggregation or redox reactivity.

Therefore, a dose response curve was generated using the BRET assay, and this confirmed that the interaction was dose dependent with and EC_{50} of 196 ± 23 µM (Figure 42A). A dose response cytotoxicity experiment was also generated; this determined that **Z13** was cytotoxic at high concentrations (Figure 42B). However, as no cytotoxic affects were consistently observed at concentrations below 250 µM, it
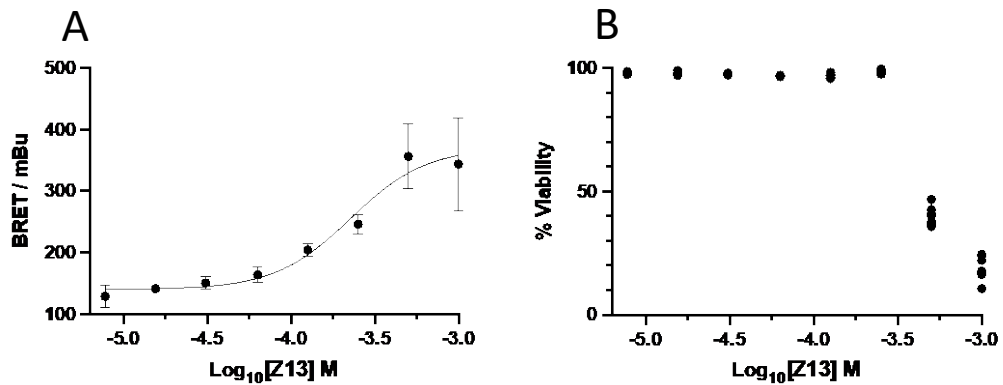
*Figure 42 - BRET dose-response measurements for HEK293 cells treated with Z13; A) BRET dose-response curve for HE293 cells treated with **Z13** (n=8, ±SEM); B) LIVE-DEAD cell viability dose-response for HEK293 cells treated with **Z13** (n=8). Individual values are shown as magnitude of ±SEM is too small to plot.*

was decided that any cytotoxic effects occurred at concentrations sufficiently above those at which the compound exerted its hybrid promoting function.

Next, western blotting was utilised in an orthogonal assay to further corroborate **Z13** as a promotor of hybrid formation. This was performed in HUVECs with compound treatments again at 100 µM for 24 hrs. This cell line was chosen as HUVECs represent a more relevant endothelial model and express high levels of IR and IGF1R. An immunoprecipitation procedure[181] was utilised to differentiate between heterodimer and homodimer formation (Figure 43A). With this methodology, it was possible to separate receptors containing the IGF1R subunit with the selective D23H3 IGF1R antibody bound to protein A-coupled sepharose
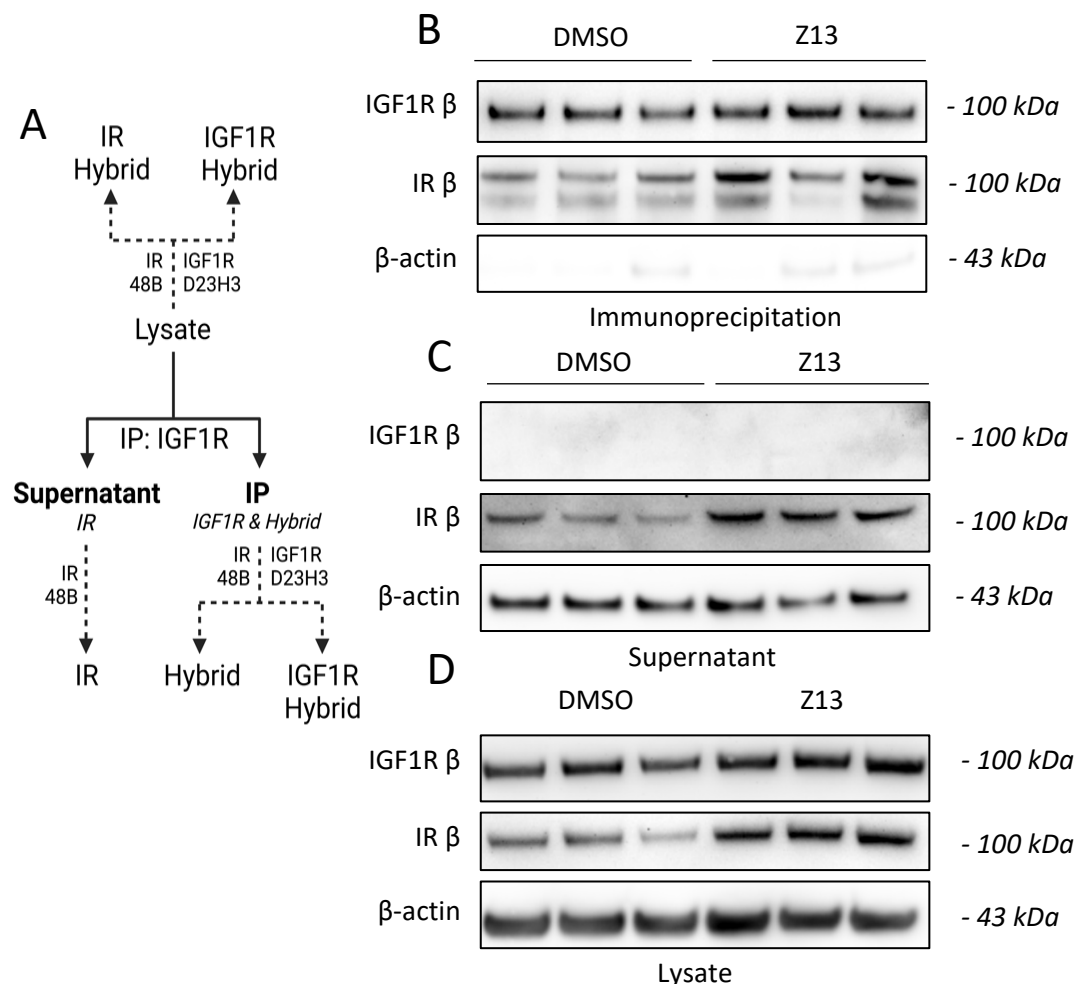
Figure 43 - Immunoprecipitation experiments on HUVECs treated with Z13: A) Schematic representation of fractions analysed in IP experiment. Full lines indicate fractionation pre-electrophoresis whilst dashed lines indicate immunostaining post electrophoresis; B) Representative western blot of receptors immunoprecipitated by IGF1R selective D23H3 antibody; C) Representative western blot of whole cell lysates; D) Representative western blot of cell lysates.
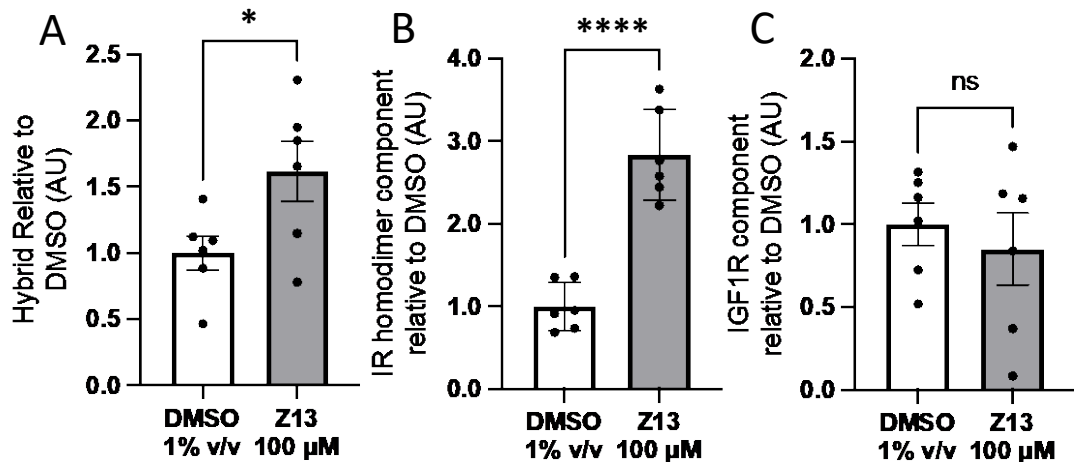
Figure 44 - Densitometric analysis of western blots from immunoprecipitation experiments on HUVECs treated with Z13. All values are normalised against the corresponding β-actin loading control, with mean DMSO values normalised to 1: A) Quantification of hybrid levels in IP fraction (n=3,2; p<0.05; ± SEM); B) Quantification of IR homodimer levels in supernatant fraction (n=3,2; p<0.0001; ± SEM) C) Quantification of IGF1R component in IP fraction (n=3,2, p≥0.05; ± SEM).

beads. Both the IGF1R immunoprecipitate and the unbound supernatant were collected and separated by SDS-PAGE under reducing condition. Subsequently, immunoblotting with the IR specific 48B antibody allowed hybrid receptors to be quantified in the immunoprecipitate, and IR quantified from the supernatant (Figure 43B). Densitometric analysis of these blots determined that **Z13** induced a significant increase in IR (Figure 44A) and hybrid (Figure 44B) populations when compared to the vehicle control but no significant change in IGF1R protein levels.

Further BRET analysis of **Z13** was performed to further corroborate its selectivity profile. This followed a similar procedure to the BRET hybrid assay, but instead utilising IR and IGF1R Rluc and YPET tagged proteins such that only homodimeric receptors were BRET competent. This confirmed that treatment with **Z13** at 100 µM
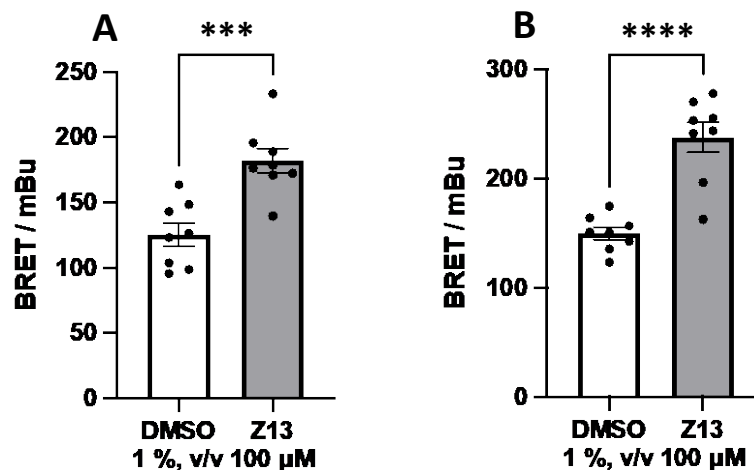
Figure 45 – BRET assay to evaluate the selectivity of Z13 treatment on IR and IGF1R homodimer levels; A) BRET measurements of IR homodimer formation in HEK293 cells transfected with IR-Rluc and IR-YPET (n=8; p<0.001. ±SEM); B) BRET measurements of IGF1R homodimer concentration in HEK293 cells transfected with IGF1R-Rluc and IGF1R -YPET (n=8; p<0.0001; SEM).

was promoting the formation of the IR (Figure 45A), whilst also establishing that **Z13** similarly promoted IGF1R formation (Figure 45B).

3.4.2   Synthesis and Biological Evaluation of Z13 Analogues

To establish the structure-activity relationship of **Z13**, several structural analogues of **Z13** were synthesised. The **Z13** analogues **57**, **58**, **59** and **60** were synthesised by Hannah McCarrick for evaluation in the BRET assay. The imidazole functionality of **Z13** was altered for both a phenyl and methyl substituent to determine if the imidazole functionality was necessary for potency. Similarly, substitution of the chloro-group from the 3 to 4 position was trialled to determine if this had an effect on potency. This generated four analogues (**57**-**60**) for biological evaluation (Figure 46A).

Additionally, **Z13** was used as a query file for ligand-based screen to identify alternative scaffolds which retained its biological activity, using the scaffold hopping tools ROCS[168, 169] and EON[170] (OpenEye Scientific). ROCS allows rapid identification of alternative scaffolds through shape comparison, aligning and scoring a database of small molecules against a query structure. EON compares electrostatic potential
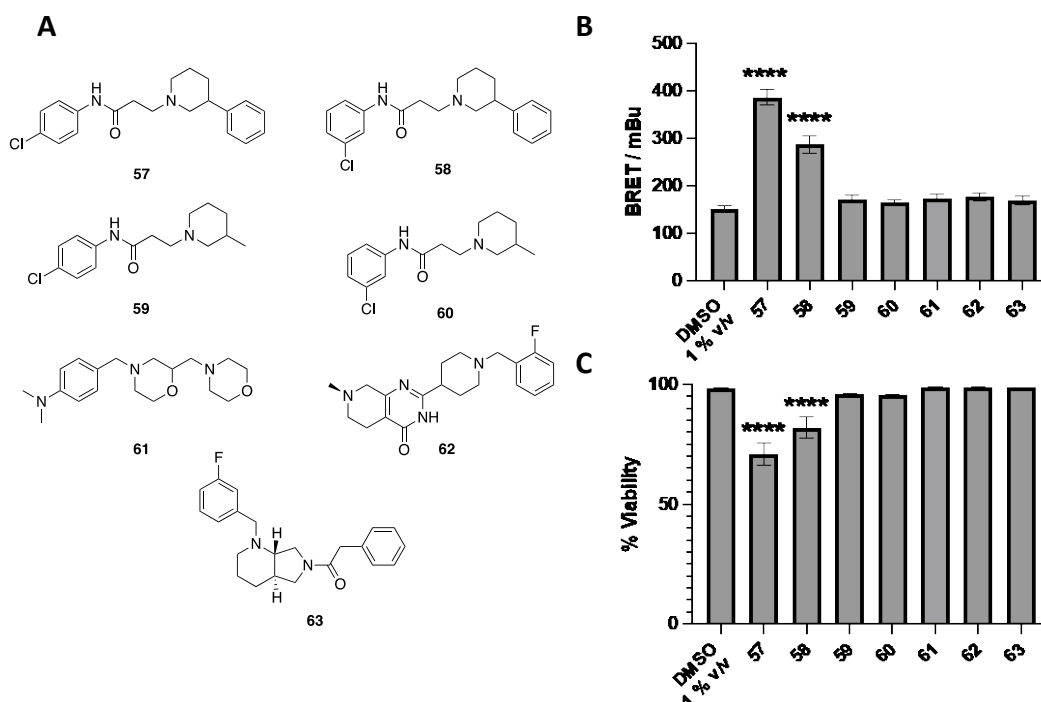
Figure 46 – Assessment of Z13 analogues A) Structures of small molecule Z13 analogues; B) BRET measurements to quantify hybrid formation post compound treatment (100 μM treatment, n = 3,8, p<0.001; ± SEM); C) LIVE-DEAD cytotoxicity measurements of Z13 analogues (100μM treatment, n=3,8, p<0.001; ± SEM).

maps of pre-aligned molecules to a query structure, producing a Tanimoto similarity score for comparison. Therefore, EON can be used to analyse ROCS aligned output structures in a workflow to identify scaffolds with similar steric and electrostatic shape to an input query. Shape similarity screening tool ROCs was used to computationally align and score molecules from the MCCB in-house library. The top 1,000 scoring aligned output structures screened with the electrostatic comparison tool EON. An additional three of the top scoring scaffolds were chosen (**61-63)** from this method for subsequent biological evaluation (Figure 46A).

Of the **Z13** analogues evaluated by BRET, only the phenyl-analogues **57** and **58** were determined to increase hybrid formation at 100 μM (Figure 46B). Both gave large increases in of *ca.* 150 mBu and 250 mBu, respectively; larger than *ca.* 50 mBu increase in response to **Z13** at the same concentration. However, these analogues showed cytotoxicity in the LIVE-DEAD assay when tested at concentrations of 100 μM (Figure 46C), with a 27% and 15% reduction in cell viability at this concentration

for compounds **57** and **58,** respectfully. This is consistent with the observation that **Z13** shows cytotoxicity, albeit at concentrations above 250 µM.

3.4.3 RT qPCR to Evaluate the Effect of **Z13** on INSR and IGF1R mRNA Levels

To further evaluate the mechanisms by which **Z13** was promoting hybrid formation, reverse transcription quantitative polymerase chain reaction (RT qPCR) was used to determine if **Z13** was influencing IR and IGF1R levels by affecting gene expression. This experiment was performed on HUVECs, again treated with **Z13** at 100 µM for 24 hrs. This determined that **Z13** treatment resulted in elevated mRNA levels of the *INSR* (Figure 47A) and *IGF1R* (Figure 47B) genes. Therefore, it is likely that **Z13** affects hybrid formation through promoting IR and IGF1R transcription, which results in increased formation of the heteromeric and homomeric receptors. This result confirmed that the BRET assay is sensitive to increases of IR: IGF1R hybrid levels. However, it also emphasised the necessity to validate hits with secondary
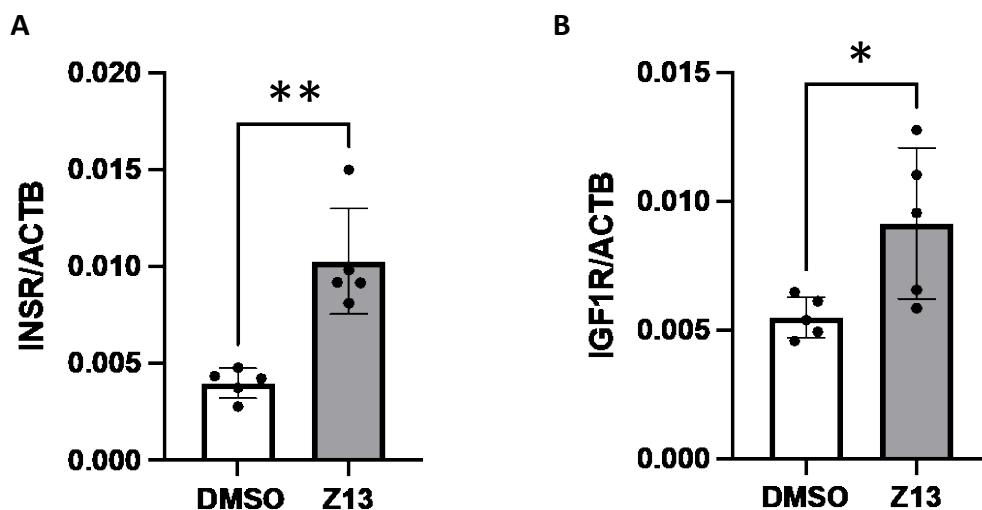
Figure 47 – RT qPCR quantification of genes encoding IR and IGF1R in HUVECs treated with Z13. Values are shown as cycles to threshold (cT) and normalised to the corresponding β-actin value for each sample A) RT qPCR quantification of gene expression encoding the IR in HUVECs treated with Z13 (100 µM, 24 hr treatment, n=5; p<0.01; ±SEM); RT qPCR quantification of gene expression encoding the IGF1R in HUVECs treated with Z13 (100 µM, 24 hr treatment, n=5, p<0.05 ± SEM).

orthogonal assays and evaluate their off-target activity to ensure a genuine interaction at the target of interest.

3.5 Biological Evaluation of Molecules Identified by vHTS of the AlphaFold Homology Model

An additional 42 small molecules prioritised from the AlphaFold screening campaign (Section 2.4.6) were assessed in the BRET and LIVE-DEAD screening assays. None of the molecules produced a significant decrease in BRET ratio at 100 µM concentration (Figure 48A). Two of the molecules, **71** and **83** showed a significant increase in BRET ratio when treating cells at 100 µM. However, these molecules also resulted in a significant decrease in cell viability in the LIVE-DEAD assay,
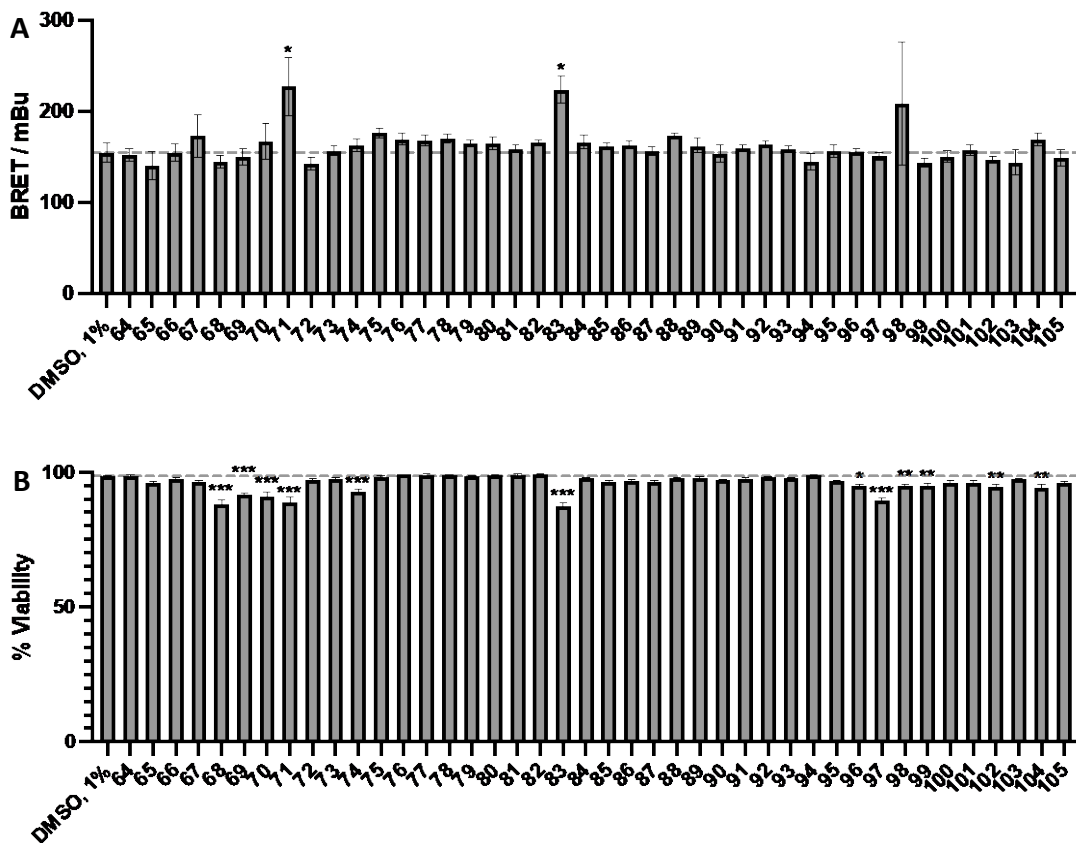
Figure 48 - Biological evaluation of small molecules prioritised from vHTS of the AlphaFold homology model. Grey lines indicate mean value for DMSO control; (A) BRET measurements to quantify hybrid formation post-treatment with screening compounds (100 µM, 24 hr treatment, n=8,3; p<0.05; ±SEM); (B) LIVE-DEAD cell viability to determine cytotoxicity of screening compound (100 µM, 24 hr treatment, n=8,3; p<0.05 ± SEM).

indicating increases in BRET ratio may occur due to cytotoxic artefacts (Figure 48B). Therefore, it was decided not to pursue any of these compounds further.

3.6   Conclusions

In this chapter, a BRET assay was validated and optimised to screen small-molecules for their activity to modulate IR-IGF1R hybrid formation. Small molecules that had been identified in Chapter 2 were screened in this assay. Although none of the tested molecules were determined to inhibit IR-IGF1R hybrid formation, several appeared to promote IR-IGF1R hybrid formation, as indicated by an increase in BRET signal. Amongst these, the compound **Z13** uniquely enhanced hybrid formation with no associated cytotoxicity. The ability of **Z13** to increase hybrid formation was confirmed using an orthogonal immunoprecipitation experiment. **Z13** was shown to similarly increased IR and IGF1R expression using both the BRET and immunoprecipitation-based assays when tested at concentrations of 100 µM. qRT PCR determined that **Z13** upregulated expression of the mRNA encoding both the IR and IGF1R suggesting its mechanism of action involves promoting expression of these genes. Analogues of **Z13** were synthesised, but cytotoxic effects were observed with all analogues that showed a corresponding increase in BRET signal. These findings indicated that **Z13** was unlikely to be acting directly on the hybrid receptor at the L1: FnIII-1 interface, according to the design hypothesis, and it was decided not to pursue this series any further.

The lack of activity of compounds identified by screening against the TACOS and AlphaFold models could be due to several factors: the homology models may not accurately model the L2-FnIII-1 interface, hotspot prediction may not have been accurate, or the chemical libraries utilised in screening have not provided sufficient diversity. To address the first and second of these possibilities, it was decided to experimentally validate some of the predicted hotspot residues in Chapter 4 using site-directed mutagenesis.

Additionally, analogues of the small molecule inhibitor **HI2** were designed and synthesised with the aim of improving their solubility, whilst retaining their ability to inhibit hybrid formation. However, none of the analogues showed any activity in the BRET assay at 100 µM, which is consistent with the relatively flat SAR of the HI

series. Several of the analogues did show improved solubility, particularly those which substituted the benzyl-triazole tail. Therefore, any further development of the HI series should prioritise modification in this region to improve solubility.

3.6.1   Progress Towards Project Goals

In this chapter, progress has been made towards objectives 3 and 4 outlined in the project goals. Several rounds of synthesis have been used to identify novel modulators of IR-IGF1R formation, specifically with the synthesis of analogues of **Z13**. Additionally, both synthesised molecules and those identified by vHTS have been evaluated for their ability to modulate hybrid formation *in vitro.* This was performed using a combination of BRET and immunoprecipitation-based procedure, aligning with the aims of objective number 3 in the project goals.

4 Characterisation of IR-IGF1R Hotspot Epitopes and Protein Production

4.1 Evaluation of Predicted Hotspot Residues Using Mutagenesis

4.1.1 Generation of Hybrid Receptor Mutants

To experimentally validate the hotspot residues identified using the KFC2 server[152], site directed mutagenesis was employed to generate a series of chimeric IR-IGF1R receptors containing point mutations at selected hotspot residues. Alanine mutations were chosen as they represent the truncation of the amino acid side chain to the β-carbon, whilst retaining the backbone dihedral angle preferences of most amino acids. These mutant receptors could then be characterised biochemically to determine the importance of each mutation to the IR-IGF1R PPI.

Mutants were generated by a PCR-based approach involving the amplification of the IR-Rluc and IGF1R-YPET plasmids (Figure 49). To introduce the desired mutation, back-to-back primers were designed; one of the primers contained the desired mutation but was flanked by sequences complimentary to the relevant
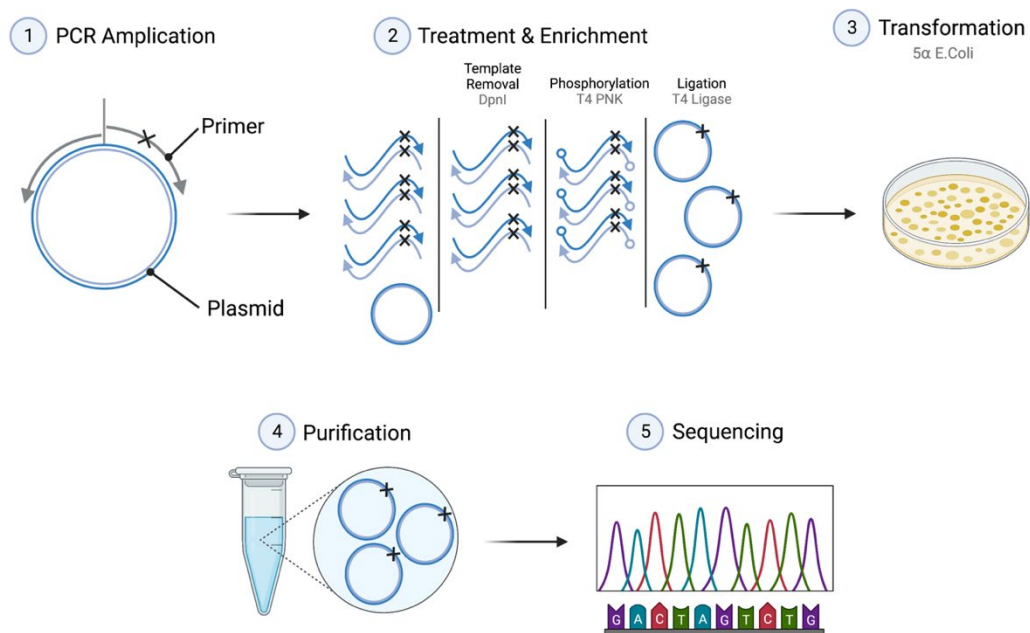
Figure 49 – Mutagenesis workflow for generating IGF1R-YPET mutants to validate KFC2 hotspot prediction. Point mutations are indicated by a cross. Created with BioRender.com.

surrounding plasmid sequences. During PCR amplification, the desired mutation was incorporated via the primer containing the specific mutation.
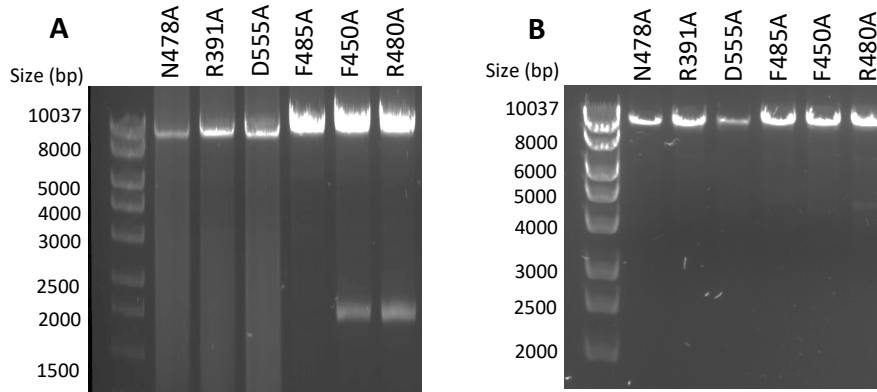
Figure 50 – Representative agarose gel electrophoresis of mutated IGF1R-YPET PCR products A) PCR product when PCR was performed utilising NEB Q5 polymerase B) PCR product when PCR was performed using SuperFi II Platinum mastermix.

Initial attempts at PCR were performed with Q5 polymerase and proved to have limited success (Figure 50A). This was attributed to the large size of the IR-Rluc and IGF1R-YPET plasmids, each approximately 10 kb. The subsequent long-range PCR proved challenging, with several non-specific products generated. PCR optimisation was attempted through alteration of the annealing temperature; however, this did little to prevent non-specific PCR bands. Success was instead achieved through utilising Platinum SuperFi II polymerase, which amplified IR-Rluc and IGF1R-YPET vectors more specifically at a universal annealing temperature of 60 °C (Figure 50B). The high-fidelity nature of this polymerase was crucial to reducing replication errors and ensuring only the desired point mutation was incorporated.

An aliquot of the purified PCR product was subsequently treated with DpnI, T4 Polynucleotide kinase (PNK) and T4 ligase to remove the template and circularise the PCR product. DpnI ensured efficient template removal, as this enzyme cleaves only dam methylated DNA at Gm6A^TC sites. Therefore, DpnI removes the template plasmid DNA derived from the dam+ 5α *E. coli* strain without degrading the PCR product, which is not dam methylated. This step reduces wild-type background in the subsequent transformation step. T4 PNK was utilised for

5'phosphorylation of the linear PCR product before ligation by T4 ligase to generate a circular plasmid. This was then transformed into 5α *E. coli*, from which DNA was purified. The IGF1R gene was sequenced to ensure the desired mutations had been incorporated and no additional mutations had been erroneously generated during PCR (Figure 51).
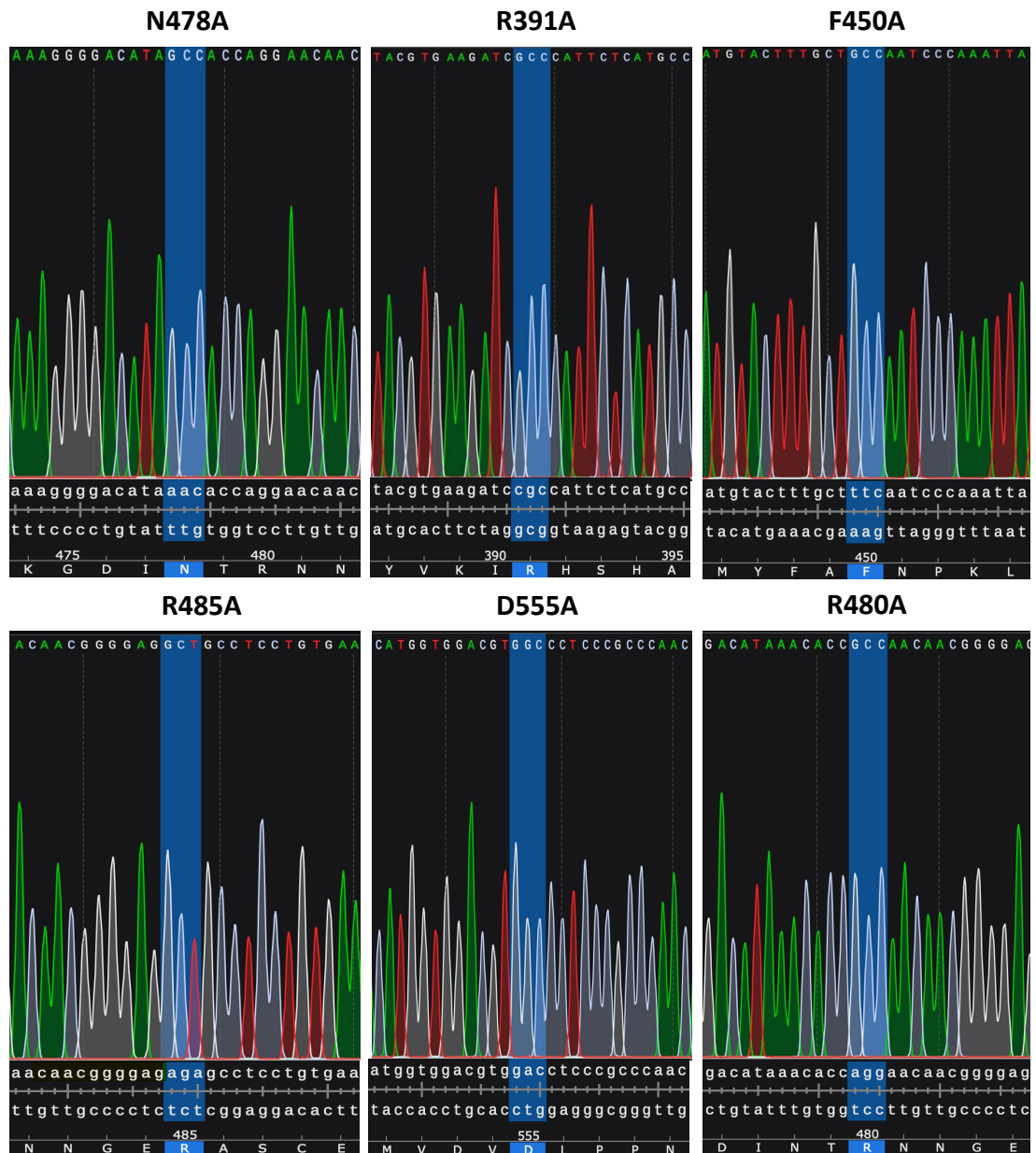
Figure 51 – Sequencing of mutant vectors to verify point mutagenesis. Mutant sequencing traces are shown above, and relevant region of the wild-type sequences are shown below.

4.1.2 Functional Evaluation of Hybrid Receptor Mutants

Once the mutated plasmid sequences had been verified, functional analysis of the resulting mutants was carried out. Mutations in receptor tyrosine kinases can

interfere with the robust expression and proper subcellular localisation of the protein. Therefore, it was important to ensure the mutant receptors were targeted to the cell-membrane and functional before further evaluation. The ability of the mutant receptors to be transported in the cell membrane was evaluated by fluorescence microscopy, which detected the YPET labelled IGF1R monomer (Figure 52). This experiment confirmed that yellow fluorescence was predominantly

located at the cell membrane for wild-type IGF1R-YPET co-transfected into HEK293 cells with IR-Rluc. Similarly, each of the mutant receptor construct displayed fluorescence localised to the cell membrane when co-transfected with IR-Rluc. This contrasted to YPET transfected alone, in which high levels of fluorescence could be
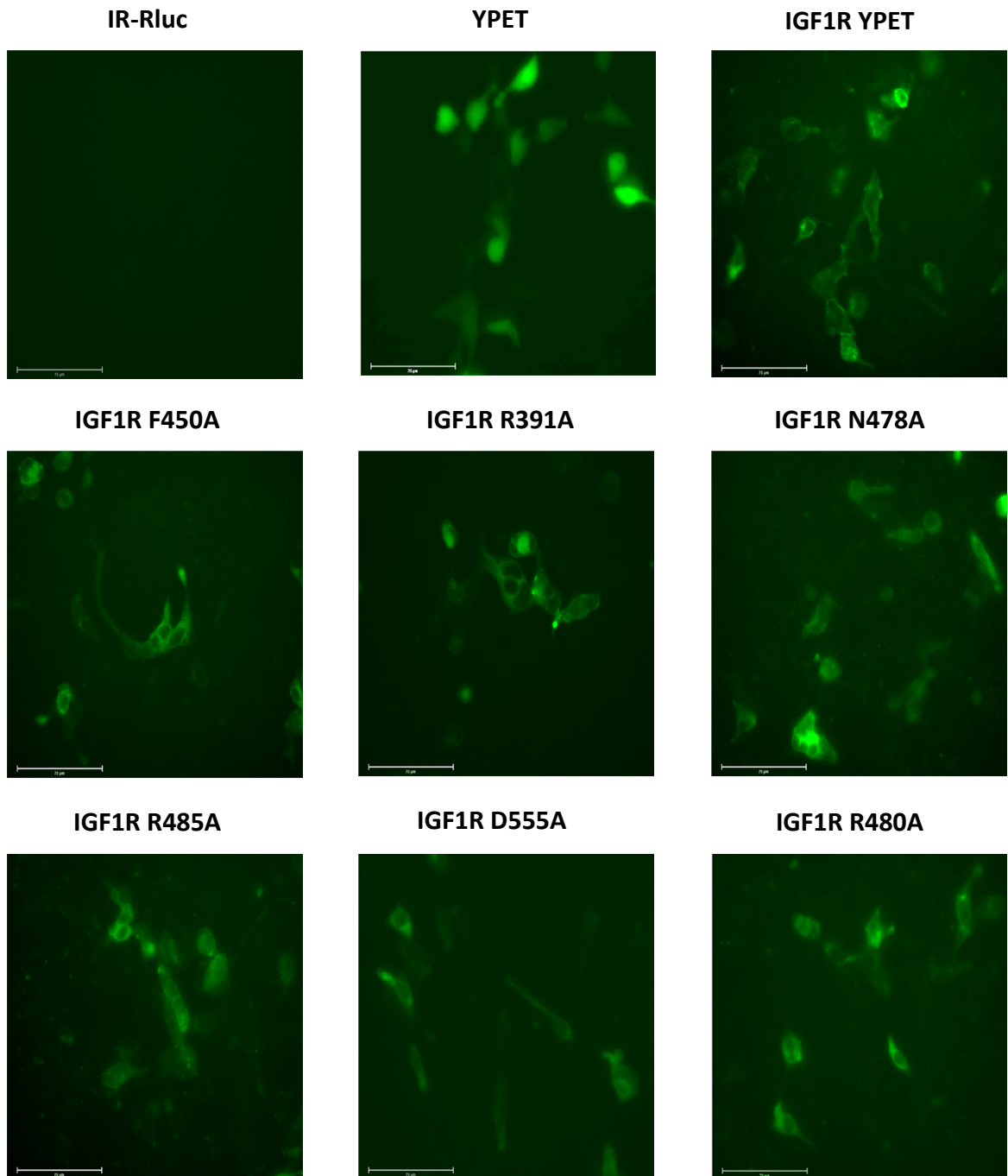
Figure 52 - Representative fluorescence microscopy images of cells transfected with combinations of IR-Rluc and the indicated IGF1R mutant constructs, showing the subcellular location of the YPET fusion protein. Scale bar = 75 μm.

detected in the cell cytoplasm. These experiments imply that the introduced

mutations do not affect the targeting of the chimeric hybrid receptors to the cell membrane.

Construct	Mock		WT		R480A		D555A		F450A	
IGF1 (100nM)	+	-	+	-	+	-	+	-	+	-

120 kDa		pIGF1Rβ (M)
95 kDa		pIGF1Rβ (WT)
120 kDa		IGF1R-β (M)
95 kDa		IGF1R-β (WT)

Construct	Mock		WT		N478A		R391A		R485A	
IGF1 (100 nM)	+	-	+	-	+	-	+	-	+	-

120 kDa		pIGF1Rβ (M)
95 kDa		pIGF1Rβ (WT)
120 kDa		IGF1R-β (M)
95 kDa		IGF1R-β (WT)
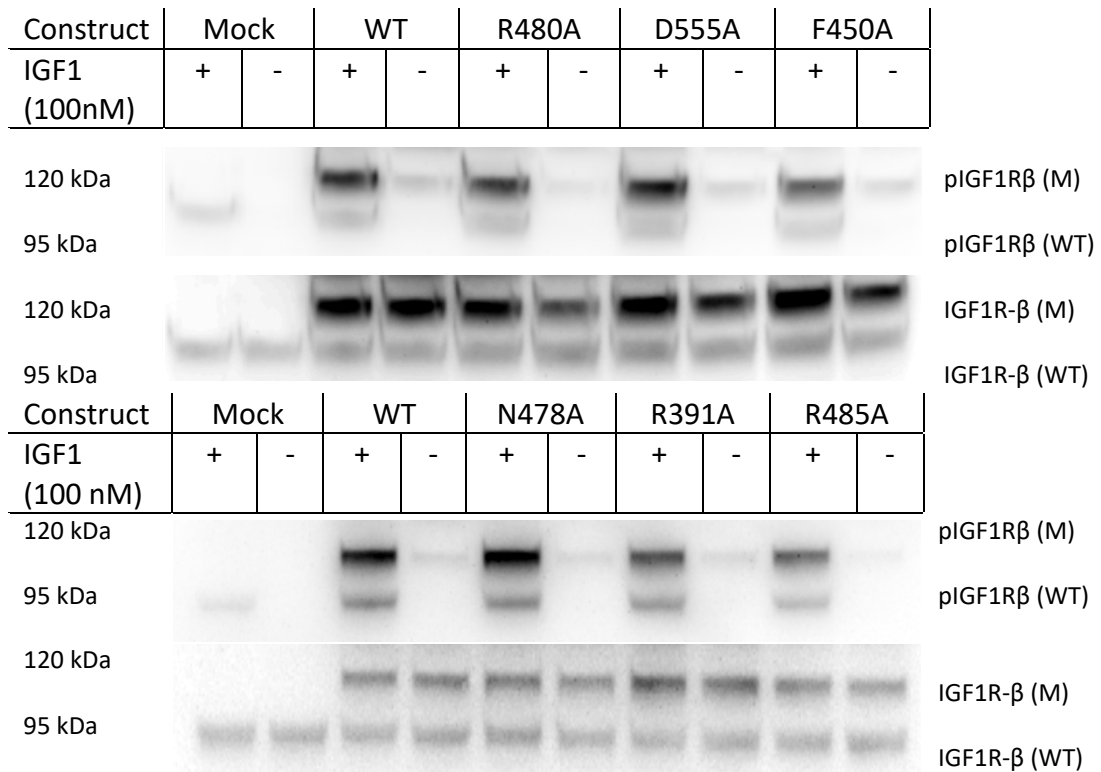
Figure 53 – Representative western blots of HEK293 cells transfected with combinations of IGF1R-YPET mutants and treated with 100 nM IGF1, showing the levels of IGF1R pY1135 and total IGF1R when probed by the pIGF1R specific DA7A8 antibody (top) and the IGF1R specific D23H3 antibody (bottom). The labels (M) and (WT) refer to the mutant and wild-type bands respectively.

Additionally, western blotting was utilised to ensure the mutant IGF1R receptors signalled robustly in response to IGF1 stimulation (Figure 53). HEK293 cells were co-transfected with IR-Rluc and the mutant IGF1R-YPET constructs before 24 hr serum starvation and treatment with 100 nM IGF1. Subsequent extraction and immunoblotting with the pIGF1R antibody DA7A8 confirmed that the receptors were robustly phosphorylated to a comparable extent as the wild-type IGF1R in response to IGF1. Whilst this assay will also detect an IGF1R homomeric

component, the presence of the mutant IGF1R constructs did not alter this signalling when compared to the signalling of the wild-type receptor.

With confirmation that the mutated receptors did not show altered signalling or subcellular localisation, the effect of the mutations on the receptor binding interface could be evaluated. This was achieved using a BRET based donor
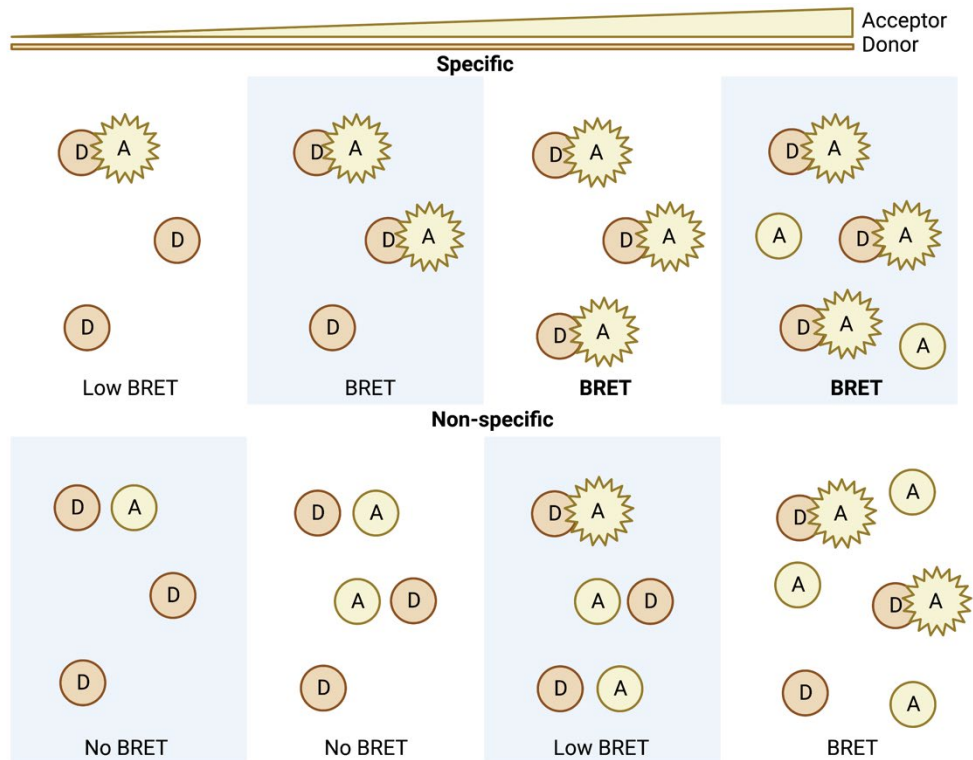
Figure 54 - Schematic showing the principle of the donor saturation assay. Acceptor molecules are denoted by an A and donor molecules depicted with a D. The DSA effect is described for specific and non-specific PPIs. Created with BioRender.com.

saturation assay[173] (Figure 54). This assay was performed by co-transfecting a fixed amount of BRET donor (Rluc tagged) with increasing amounts of BRET acceptor (YPET tagged). BRET signal is then measured at increasing ratios of BRET acceptor to donor. BRET signal is a hyperbolic function of acceptor to donor ratio for a specific protein-protein interaction, but a linear function of donor-acceptor ratio for a non-specific interaction. The hyperbolic relationship can be rationalised by considering that low levels of BRET acceptor will result in low BRET signal, and subsequent increase in BRET acceptor concentration will cause a rapid increase in BRET signal until most BRET donor molecules are saturated. Subsequently increasing the BRET acceptor concentration only results in small increases in BRET signal due to non-specific interactions. Alternatively, a nonspecific interaction results in a linear BRET

signal derived from random collisions of the BRET pair, which increase in frequency as a function of BRET acceptor concentration.

4.1.3 Radioligand Binding Assay to Quantify IR-Rluc and IGF1R-YPET
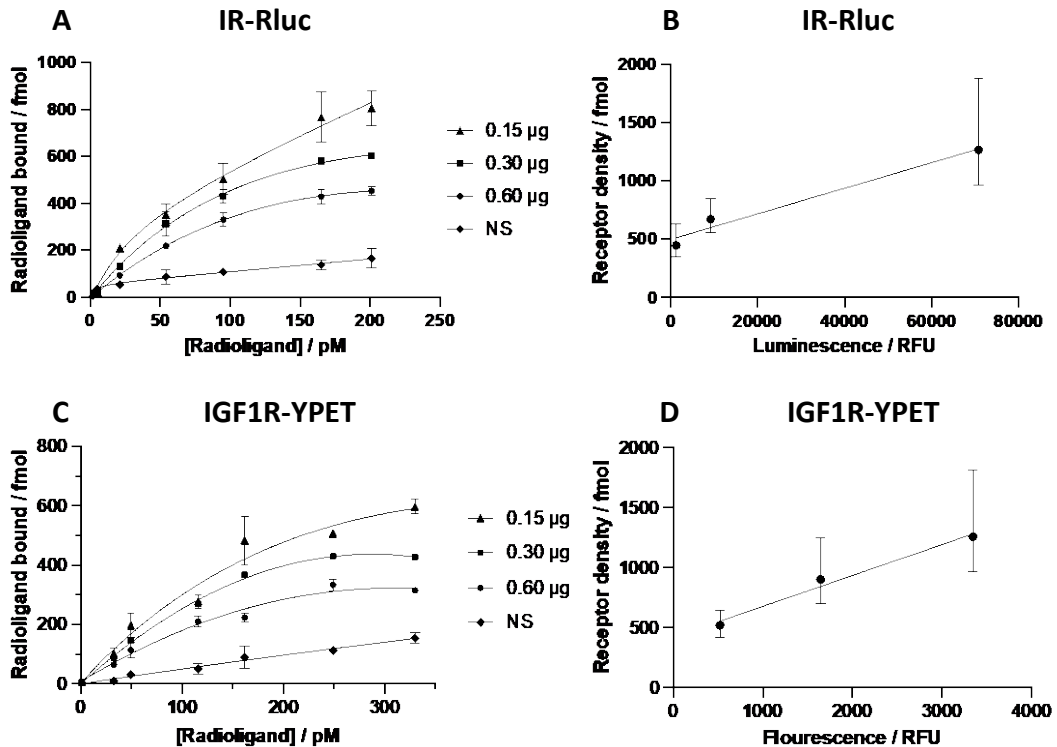
Figure 55 – Radioligand binding assays to determine how IR-Rluc and IGF1R-YPET receptor quantities are correlated with their relative luminescence and fluorescence respectively. Saturation isotherms (A,C) depict non-specific radioligand binding to untransfected cells (NS), or specific radioligand binding to cells transfected with increasing amounts of IR-Rluc or IGF1R-YPET cDNAs over increasing radioligand concentrations. A) Saturation radioligand binding curves for the IR-Rluc constructs (n=2); B) Calibration of IR-Rluc luminescence in the presence of 5 µM coelenterazine to total receptor numbers; C) Saturation radioligand bind curves for the IGF1R-YPET constructs (n=2); D) Calibration of IGF1R fluorescence when excited at 513 nm to total receptor numbers.

To obtain receptor/donor ratios that equate to real protein quantities in the DSA assay, luciferase and YPET fluorescence must be correlated to receptor numbers to allow conversion. Receptor numbers for the IR-Rluc and IGF1R-YPET constructs were determined by a radioligand saturation binding assay. Cells were transfected with increasing quantities of cDNA encoding IR-Rluc or IGF1R-YPET. Subsequently, the luminescence and fluorescence values corresponding to IR-Rluc in the presence of 5 µM coelenterazine or YPET excited at 513 nm were determined for each transfection. Cells expressing the chimeric receptors were then probed with increasing concentrations of ^{125}I-Insulin or ^{125}I-IGFI before bound probes were recorded. Non-specific binding was also evaluated by incubating cells with ^{125}I-Insulin or ^{125}I-IGFI in the presence of unlabelled insulin or IGFI at 500 nM to block

specific binding. A one-site non-linear regression model was fit to each dataset, allowing the determination of the dissociation constant (K_d) and number of binding sites present in the assay (B_{max}) (Figure 55A, C). From these assays, the K_d for the IR-Rluc and IGF1R-YPET constructs were determined as 140 ± 25.2 pM and 122.8 ± 24.2 pM, which are in good accordance with the apparent K_d values of 190 pM and 120 pM reported[182] for the IR and IGF1R respectively. It is noted in the case of IR-Rluc, the K_d value may be altered by the presence of the fused Rluc tag.

The B_{max} values for each transfection ratio were found to vary linearly with luminescence and fluorescence values for the IR-Rluc and IGF1R-YPET constructs respectively (Figure 55B, D). This in turn allows the conversion between luminescence, fluorescence values and actual protein quantities for the IR-Rluc and IGF1R-YPET constructs respectively. It was reasonable to assume that point mutations in the receptor ectodomains would not affect luminescence and fluorescence originating from the C-terminal Rluc and YPET tags. Therefore, this analysis could similarly be utilised to convert between luminescence, fluorescence, and actual protein quantities for the chimeric mutant receptors.

4.1.4   Donor Saturation Assays for Hybrid Receptor Mutants

The DSA was undertaken by co-transfection of WT-IR-Rluc and mutant IGF1R-YPET, incorporating mutant IGF1R-YPET as part of a hybrid receptor. Initial optimisation of the assay established a range of acceptor: donor ratios between 0-20 was suitable. A non-linear regression model was fit to the data and the used to calculate BRET$_{50}$
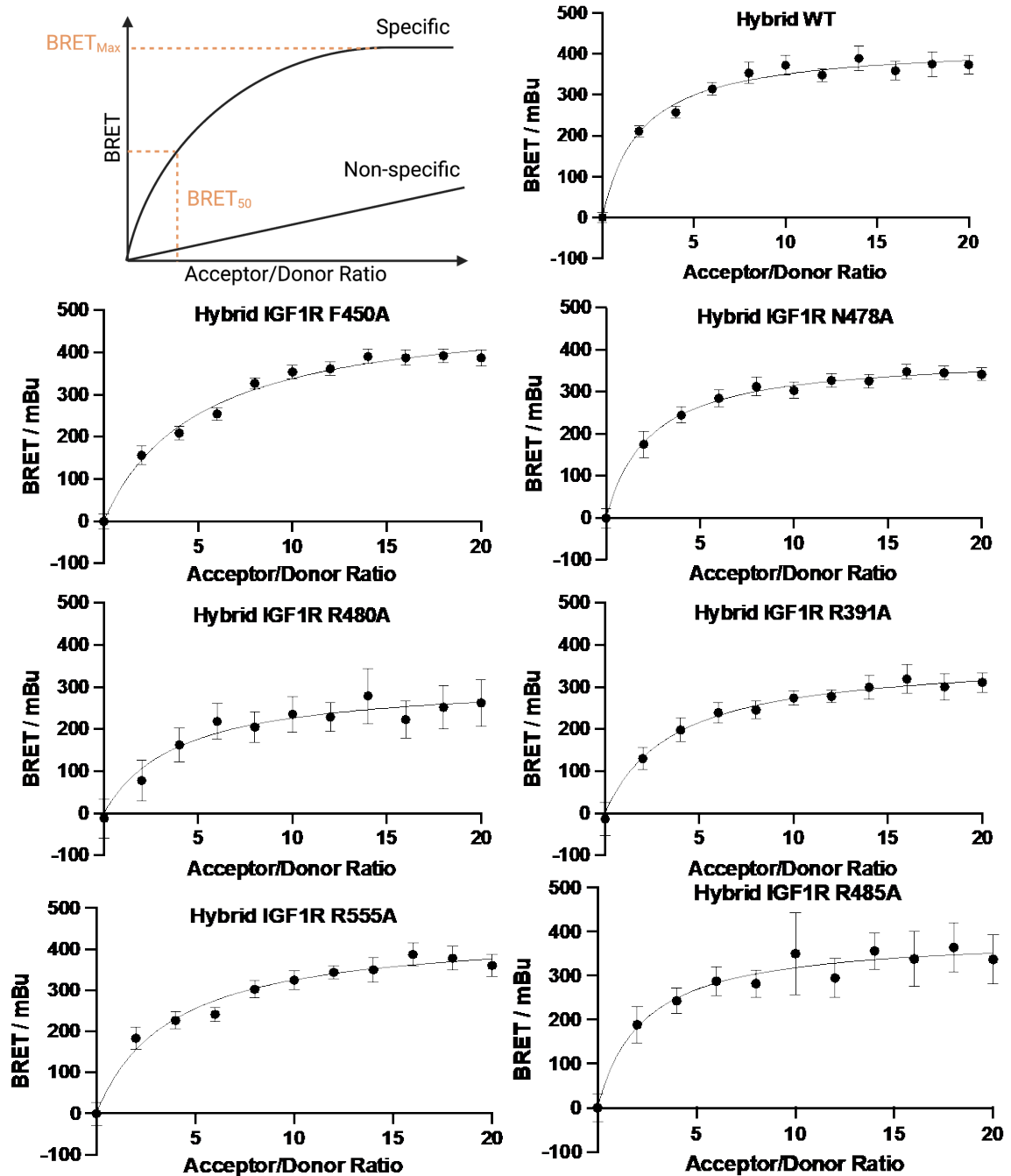
Figure 56 – Representative donor saturation assay curves performed for IR-IGF1R-hybrid mutants (n=8; ± SEM). The top left graph is representative and indicates the derivation of the $BRET_{50}$ and $BRET_{Max}$ parameters, as well as expected relationships for specific and non-specific interactions.

and $BRET_{max}$ values. The $BRET_{max}$ value represents the maximum BRET signal at the asymptote, with changes in $BRET_{max}$ interpreted as changes in the conformation or orientation of the BRET pair. The $BRET_{50}$ value refers to the donor acceptor ratio at

which 50% of the $BRET_{max}$ is achieved and is a measure of affinity between the BRET pair[172, 183].

IGF1R construct	$BRET_{50}$	$BRET_{max}$
WT	**2.08 ± 0.40**	**423.1 ± 16.9**
F450A	5.02 ± 0.69	507.6 ± 22.8
R391A	3.60 ± 0.92	371.3 ± 26.6
N478A	2.33 ± 0.44	388.4 ± 15.8
R485A	2.28 ± 0.98	391.0 ± 38.0
D555A	3.74 ± 0.80	446.5 ± 27.0
R480A	3.95 ± 1.75	316.1 ± 43.2

Figure 57 - $BRET_{max}$ and $BRET_{50}$ parameters determined for each of the IR-IGF1R hybrid mutants in the donor saturation assay. Error ranges represent the 95% confidence interval determined for the non-linear regression curve calculated for from three separate experiments. (n=8,3).

The donor saturation assays determined that mutation of IGF1R residues F450A, R391A and D555A did reduce the affinity of the IR IGF1R complex, as these mutants showed increased $BRET_{50}$ values in the assay relative to the value for the hybrid WT receptor (Figure 56, Figure 57). This result implies that each of these residues plays a significant role in stabilising the IR- IGF1R L2 FnIII-1 interface.

IGF1R residue F450 was predicted as a hotspot residue in both the AlphaFold and TACOS models (Figure 58A). In both models, the phenylalanine side chain inserts into a hydrophobic pocket on the IR monomer, stabilising the interface through hydrophobic interactions. Mutation of phenylalanine to alanine represents removal of the phenyl ring, which likely results in reduced hydrophobic contact and diminished interaction strength between the hybrid monomers. Noticeably, the
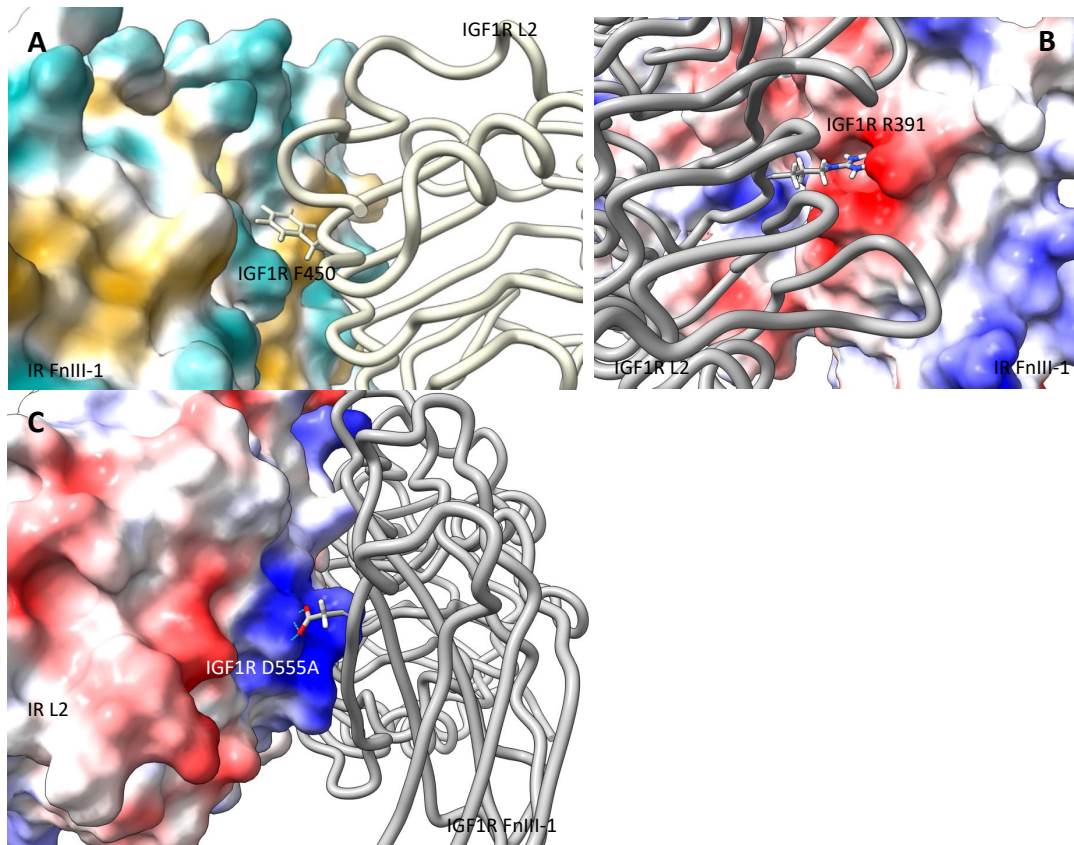
Figure 58 – Predicted interactions of the IGF1R residues for which mutation to alanine cause an increase in BRET$_{50}$. All images are taken from the AlphaFold homology model; A) Interaction of IGF1R F450 with the IR. The IGF1R is shown in yellow and IR with surface coloured by hydrophobicity; B) The interaction of IGF1R R391 with the IR. The IGF1R is shown in grey and IR with an electrostatic surface; C) The interaction of D555 with the IR. The IGF1R is shown in grey and IR with and electrostatic surface.

mutant receptor also shows a substantial increase in BRET$_{max}$ relative to the wild-type receptor, indicating that the mutation results in a change in the receptor conformation. As the BRET$_{50}$ measurement is proportional to BRET$_{max}$, the change in BRET$_{50}$ cannot definitively be attributed to a change in IR and IGF1R affinity and may be due to the inherent relationship between the two parameters. Conversely, a change in receptor affinity could be affected by a large-scale conformational change of the receptor facilitated by the mutation F450A, which may explain the two-fold change in BRET$_{50}$ from the point mutation of a single residue. It is notable that no change in subcellular location or signalling in response to IGF1 treatment was observed with this receptor. Therefore, any conformational change occurring

due to the F450A mutation does not appear to substantially alter the receptors processing or signalling properties.

The residue R391 was predicted as a hotspot residue by the AlphaFold model and extends into a positively charged pocket on the IR monomer to form electrostatic interactions with surrounding residues (Figure 58B). Mutation to alanine would result in the removal of electrostatic interactions and steric bulk and therefore reduce the affinity with the IR monomer. Similarly, residue D555 was predicted as a hotspot residue on by the AlphaFold model (Figure 58C). This residue extends to form electrostatic interactions with a cluster of residues positively charged residues located on the IR monomer. Mutation to alanine would result in the removal of these electrostatic interactions. Both mutations show no significant change in $BRET_{max}$ relative to the wild-type receptor, meaning that the change in $BRET_{50}$ can be interpreted as a change in the receptor affinity upon mutation of the receptor.

The residues R480, R485 and N478 were similarly predicted as hotspot residues. However, there was no significant deviation in $BRET_{50}$ observed with these mutants relative to the wild-type. It is possible that these were hotspots, but the DSA was not sensitive enough to detect changes in affinity caused by these mutations. Alternatively, these residues may not be hotspots and were modelled erroneously due to the inherent shortcomings in protein computational prediction. It is notable that residues F450, R485 and D555 were all predicted from KFC2b on the AlphaFold model. The scoring function for the KFC2b hotspot prediction software is designed for lower sensitivity but higher specificity to reduce false positives. It appears that the combination of the AlphaFold model of the hybrid receptor and KFC2b provided the greatest accuracy in predicting hotspot residues, albeit for this relatively small sample size of six mutations.

The ability of residues IGF1R R391 and D555A to alter the apparent affinity of the hybrid receptor is significant, as it confirms that hybrid receptor formation can be modulated pre-dimerisation by changes to residues at the L2: FnIII-1 interface. Once the receptors are dimerised, the formation of disulfide bonds affectively locks their dimeric arrangement[58]. Whilst mutations to residues at the L2:FnIII-1 interface are unlikely to be the major cause of IR-IGF1R upregulation in response to

metabolic disease[109], the fact that receptor dimerisation can be modulated by this confirms epitopes at the L2:FnIII-1 interface are a suitable target for therapeutic intervention.

4.2   Protein Production

4.2.1   Hybrid Purification Strategy

As mutagenesis experiments revealed disparities between the homology models and the IR-IGF1R interface, an experimental structure of the IR-IGF1R hybrids was sought to provide a protein model for identifying hotspot residues. Whilst the structure of the hybrid receptor ectodomain bound to IGF1 has been elucidated by cryo-EM[112], no structure of the full length or apo-hybrid receptor has yet been determined. Contemporary advances in cryo-electron microscopy mean that producing high-resolution experimental models of membrane proteins has recently become more feasible, as crystallisation of these challenging targets is not required.

To attempt cryo-electron microscopy studies on the IR-IGF1R hybrid receptor, it was necessary to produce IR-IGF1R hybrid protein. Production of IR-IGF1R hybrid protein presents several challenges: membrane proteins are typically lower abundance than cytosolic proteins and require the use of detergent to extract and solubilise them in aqueous solution for purification and structure determination. Additionally, hybrid receptors represent a distinct challenge as their expression leads to the inevitable production of IR and IGF1R homodimers. This both lowers the yield of hybrid receptors and adds an additional complexity of separating hybrid receptors from homodimeric protein.

In previous studies[112], the IR-IGF1R hybrid ectodomain was produced for the cryo-EM structure of IGF1 bound to the IR-IGF1R ectodomain. This study utilised a stable cell line co-expressing IR-B (residues 1-928) fused to a 33-residue leucine zipper sequence and IGF1R (residues 1-905) again fused to a 33-residue leucine zipper followed by a three-serine space and a c-myc affinity tag[112]. The inclusion of leucine zipper fragments aimed to stabilise the soluble hybrid ectodomain, a strategy which has previously been employed to generate structures of the soluble IR and IGF1R ectodomains[112, 184]. However, the inclusion of the leucine-zipper fragment resulted

in the production of receptor tetramers due to non-specific interactions between leucine zipper regions. Indeed, the size-exclusion chromatograms included with this study indicate that the majority of receptors formed tetramers, with their subsequent removal significantly reducing the overall protein yield. Whether the inclusion of the leucine zipper fragments augments or diminishes IR-IGF1R heterodimer formation is unknown. The use of a soluble ectodomain construct allowed continuous harvesting of protein secreted directly into the cell media. The development of a stable cell line co-expressing the IR and IGF1R constructs likely had the advantage of reduced batch to batch variability compared to transient co-expression of the same constructs.

A salient feature of the protein production methodology utilised in this study was the use of multiple rounds of purification: the first specific to IGF1R component and the second specific to the IR component of the hybrid receptors. This methodology sequentially removed IR and IGF1R homodimers, an inevitable by-product of the co-expression necessary to produce hybrid receptors. Firstly, the hybrid receptors and IGF1R were retained by a c-myc antibody-linked affinity resin before the hybrids were captured on an mAb 18-44 antibody-linked resin directed against a linear epitope in the IR β-chain. Therefore, the co-expression of IR and IGF1R followed by a two-step purification to remove homomeric receptors appears a feasible strategy to purify the hybrid receptors.

This two-step methodology provided precedent for purifying the hybrid receptors using two rounds of purification specific to each receptor monomer. However, it was decided to make several alterations to the above strategy. The usage of antibody -linked resins for purification appears to offer no significant advantages over standard affinity-tag based purification. Generating several IR and IGF1R constructs linked to a variety of epitope tags meant standard capture and release affinity purification could be used in a similar two-step purification to remove homomeric receptors.

It was also reasoned that transient co-transfection of such constructs would provide considerable flexibility over the generation of stable cell lines overexpressing the IR and IGF1R. Transient transfection would allow for more rapid screening of IR and

IGF1R constructs to optimise expression. Whilst transient transfections necessitate larger amounts of cDNA, the methodologies to produce this DNA are well optimised and efficient.

Finally, it was decided full-length IR and IGF1R constructs were favourable over production of soluble ectodomain constructs or ectodomain constructs fused to a leucine zipper fragment. As noted above, the flexibility of IR and IGF1R soluble ectodomain constructs poses a significant challenge in their characterisation by cryo-EM. However, leucine zipper stabilisation results in a loss of protein due to the leucine zipper forming higher order oligomers. Furthermore, the absence of an unliganded hybrid structure in this report[112] indicates that even a leucine zipper construct requires additional ligand stabilisation to generate a high-resolution structure. Alternatively, a full-length construct can be stabilised by membrane mimicking structures such as liposomes or detergent micelles, which may allow the characterisation of the apo-state hybrid receptor.

In conclusion, it was reasoned that the co-expression of full-length IR and IGF1R featuring an orthogonal affinity purification tag would provide the highest flexibility and cost effectiveness.

4.2.2   Design of IR and IGF1R Constructs for Hybrid Purification.

To generate the desired constructs, the full-length IR-B and IGF1R were cloned into four separate expression vectors containing a series of purification tags. The backbone vectors were based on the pOPINGS, pHL-AviTag3, pHLFcHis and pOPINE-3C HALO7 vectors, which featured C-terminal 6xHis tag, followed by a StrepII, AviTag3, Fc tag and HALO7 tag, respectively. The pOPINGS and pOPINE-3C-HALO7 vectors were gifts from Prof. Ray Owens[185], whilst the pHL-AviTag3 and pHL-AviTag3 and pHL-FcHis vectors were gives from Prof. Edith Yvonne Jones[186]. These represent a vector suite optimised for expression of mammalian membrane proteins. Each features a C-terminal affinity-tag to prevent interference with the N-terminal signal sequence, which is necessary for the receptor processing and membrane trafficking. The pOPINGS, pHL-AviTag3 and pHLFcHis vectors all utilise a signal sequence based on that of RTPTμ, which has been shown to give higher expression over native signal sequences in many instances[186]. Alternatively,

pOPINE-3C-HALO7 utilises the native IR and IGF1R signal sequence. Each vector contains a hybrid CMV enhancer/chicken β actin promotor to drive overexpression in mammalian cells.
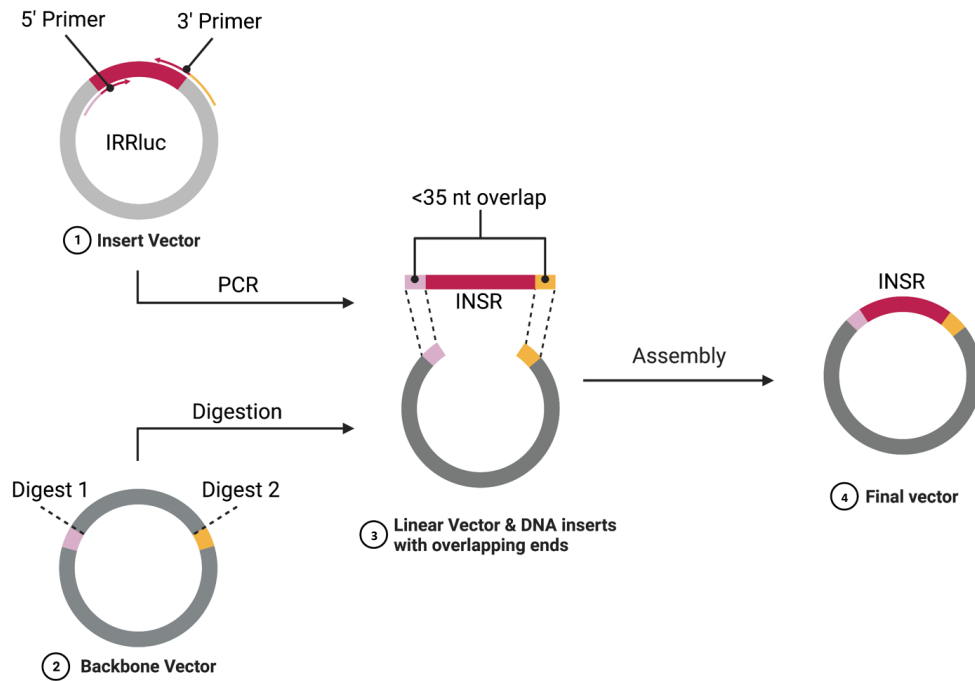
Figure 59 - Schematic overview of the HiFi assembly procedure utilised to generate affinity tagged IR and IGF1R constructs. Created with BioRender.com.

NEB HiFi assembly was utilised as this methodology allows scarless cloning without requiring the presence of compatible restriction sites on backbone and insert (Figure 59). Briefly, backbone vectors were first linearised by restriction digestion and purified by agarose gel electrophoresis and silica spin column.
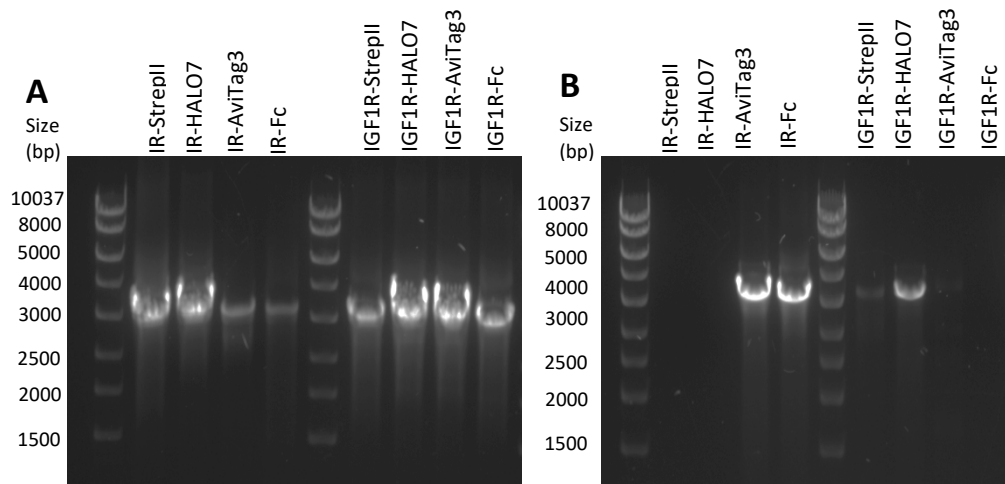
Figure 60 - Agarose gel electrophoresis of PCR products generated by amplifying IR-Rluc and IGF1R-YPET A) PCR performed with an annealing temperature of 72 °C B) PCR performed with an annealing step of 65°C

In parallel, the IR-B and IGF1R insert sequences were generated by PCR amplification of the IR-Rluc an IGF1R-YPET vectors. Optimisation of the PCR reaction determined that an annealing temperature of 72 °C resulted in robust amplification of all inserts (Figure 60A, B). The primers used in PCR contained a 5' overlap sequence, homologous to the 5' end of the relevant backbone vector, which was incorporated into each insert to allow subsequent assembly. It was determined that these overlap sequence needed to be >35 nucleotide pairs for the subsequent ligation reaction to be successful (Figure 61).

To avoid scarring in the subsequent constructs, the overlap sequence was designed so that touchdown occurred upstream of the nucleotide overhand generated by
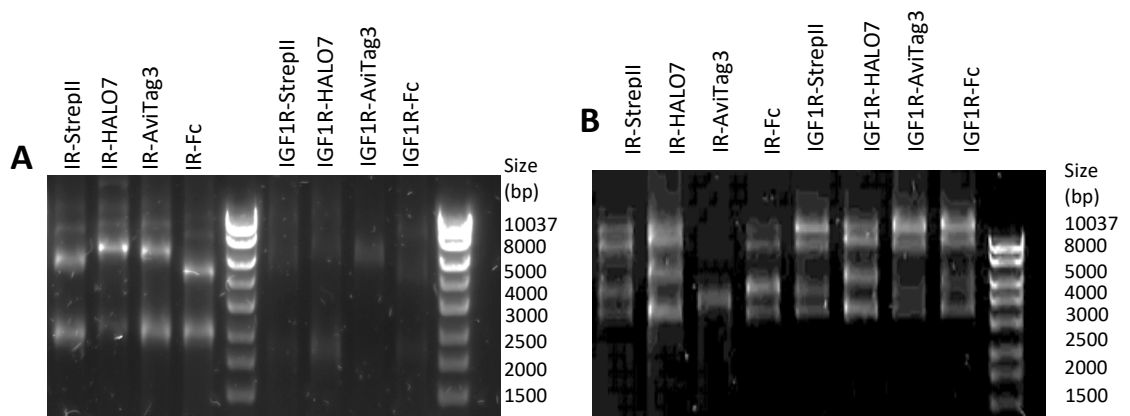
Figure 61- Agarose gel electrophoresis of HiFi assembly reaction with A) 15 nucleotide overlap between nucleotide between insert and backbone vector and B) >35 nucleotide overlap between insert and backbone sequence.

restriction digestion of the backbone vectors. The maximum distance from the touchdown to the restriction site was six nucleotides from the 3' end.

The linearised vector and insert were combined at a molar ratio of 1:2 for subsequent assembly, before direct transformation of DH5α cells with the assembly mixture containing a 5' exonuclease, a polymerase and DNA ligase. In the assembly, the 5' exonuclease generates a 3' overhang on the DNA fragments, allowing them to anneal. The polymerase subsequently fills in the resulting 5' gaps and the ligase seals the nick in the assembled DNA. Due to the presence of a LacZα expression cassette the pOPIN vectors could be pre-screened for insertion by blue-white screening. Chosen colonies were subjected to further colony PCR utilising primers
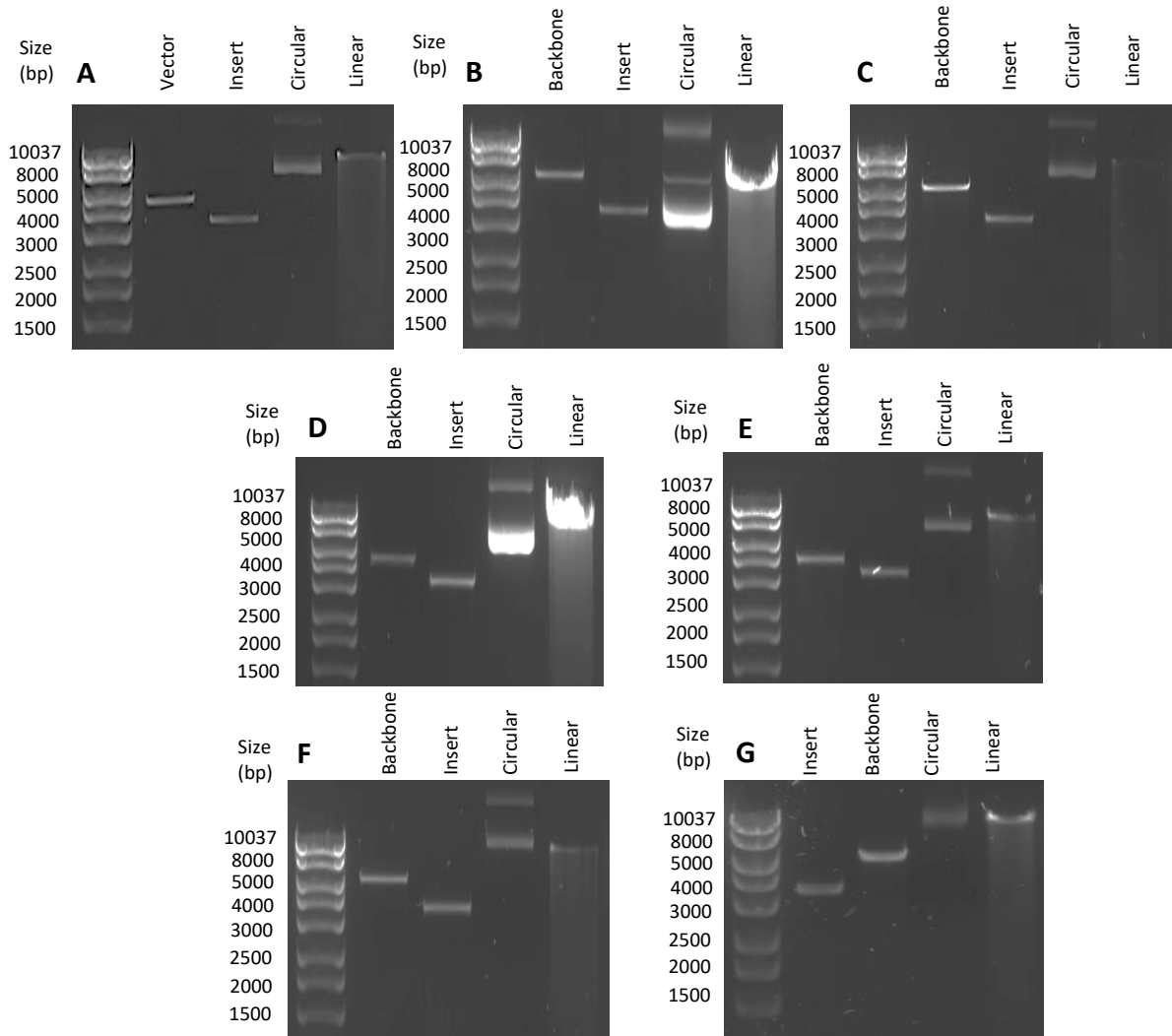
Figure 62 - Agarose gel electrophoresis showing components of the HIFI assembly reaction and products. Circular and linear lanes contain the product of the HiFi reaction and the NheI digestion product of the HiFi reaction respectively: A) IR-StrepII; B) IR-HALO7; C) IR-Fc D) IGF1R-StrepII; E) IGF1R-HALO7 F) IGF1R-AviTag3; G) IGF1R-Fc

based on the β-globin poly(A) signal common across all backbone vectors and an IR or IGF1R sequence. Colonies for which PCR indicated the presence of the correct insertion were cultured overnight before subsequent DNA purification. Purified DNA was analysed by agarose gel electrophoresis to determine the purity and size. In each case, a band was observed at the correct size for the relevant DNA constructs (Figure 62). Cloning was further verified by sequencing across the backbone and insert junction in all cases.

4.2.3   Small Scale Expression Trials

To deduce optimal conditions for large-scale protein expression experiment, small-scale expression trials were performed to evaluate construct expression and

integrity. The transfection reagent polyethylenimine (PEI) was utilised as this represented a cost-effective transfection reagent that would allow scale up for large scale expression trials. Small-scale expression was performed in adherent HEK293 cells instead of the suspension adapted 293 freestyle cells utilised for large scale culture due to cost considerations. Regardless, HEK293 cells would still provide a useful indication of the construct viability. The lysates from HEK293 cells transfected with each of the chimeric receptors were analysed by western blotting with IR-β and IGF1R-β specific antibodies under reducing conditions. The StrepII tagged constructs encode the IR and IGF1R followed by the StrepII sequence WSHPQFEK and a 6 x His. These tags have a total molecular weight of 1880 Da. Bands corresponding to the expected weight of the β subunit with an additional StrepII 6 x His sequence were observed in Figure 63A, E. Similarly, the HALO7 constructs feature a HRV 3C cleavage site, HALO7 sequence followed by 6 x His with a total molecular weight of 35 kDa. Bands corresponding to the IR and IGF1R β-subunit

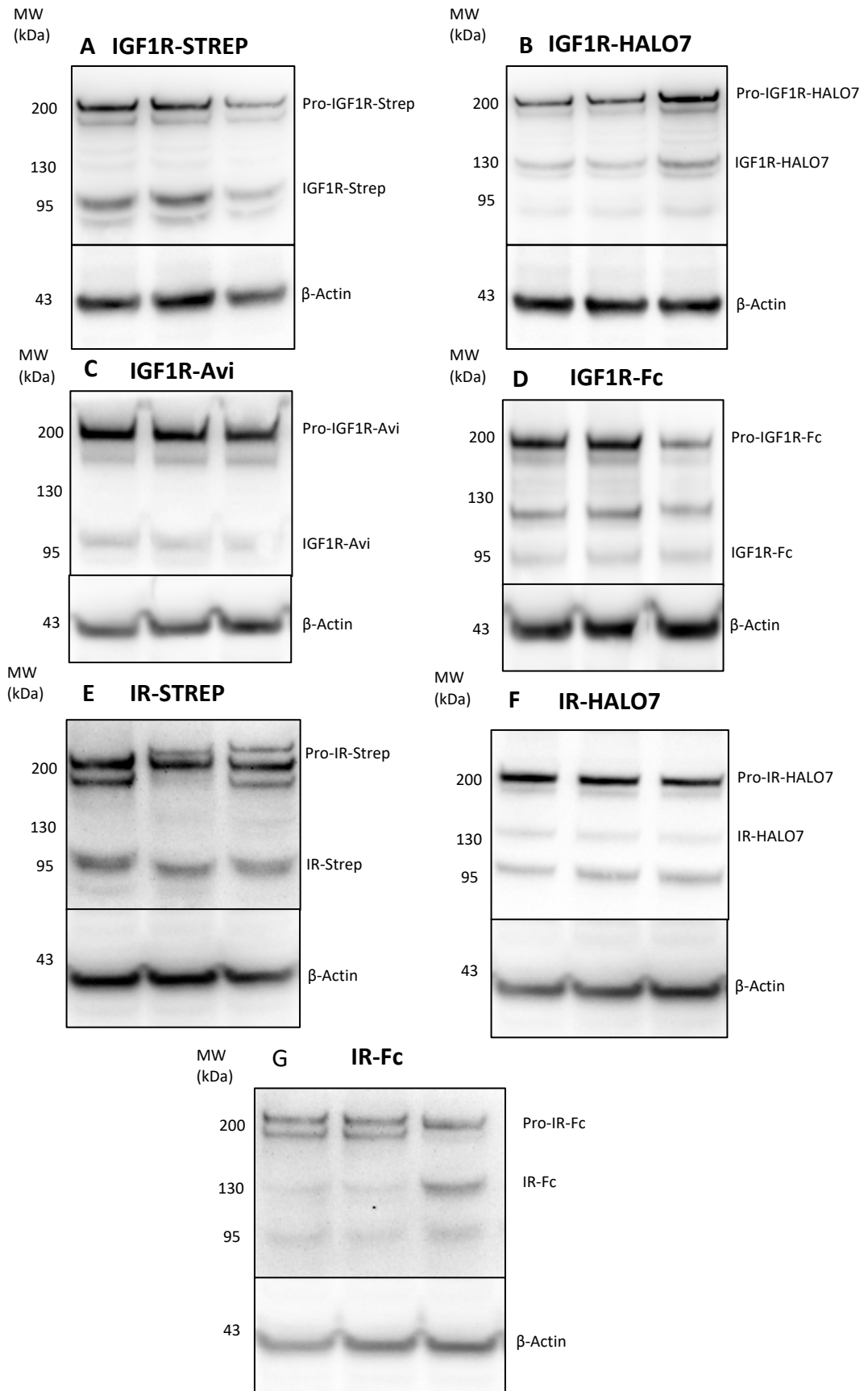*Figure 63 – Small-scale expression trials on full-length IR and IGF1R constructs. Transfections were performed in triplicate. Blots A-D (top) were probed with the IGF1R-β specific antibody D23H3 and blots E-G probed with IR-β specific antibody 48B. Blots A-G (bottom) were probed with β-actin specific antibody C4: A) IGF1R-Strep; B) IGF1R HALO7; C) IGF1R-Avi; D) IGF1R-Fc; E) IR-Step; F) IR-HALO7; G) IR-Fc*

followed by this sequence were observed in Figure 63B, F. The IGF1R-AviTag3 construct consists of the IGF1R followed by the AviTag3 sequence, GLNDIFEAQKIEWHE, a linker sequence (GRTK) and 6 x His, with an expected molecular weight of 3 kDa. A band corresponding to the IGF1R-β subunit with this additional sequence was observed in Figure 63C. Finally, the Fc-tagged constructs encodes a C-terminal HRV 3C protease sequence, an Fc encoding region and a 6 x His tag, with an expected molecular weight of 27.5 kDa. Bands corresponding to the weight of the IR and IGF1R β-subunit with this sequence attached could be observed in Figure 63D, G. In each of the blots, a band corresponding to the affinity tagged IR and IGF1R pro-receptor was observed relatively intensely, in several cases with greater intensity than the corresponding band for the mature receptor (Figure 63C, F). The presence of these bands indicates the accumulation of pro-receptor in the transfected cells, potentially due to the cells being unable to robustly process the overexpressed chimeric receptors. Such build up may prove problematic in receptor production, as it will lead to lower yields of the mature receptor, and potential difficulties in separating the mature receptor from the pro-receptor during affinity purification. It was decided to continue with scale-up to assess the severity of pro-receptor accumulation on a large-scale.

4.2.4   Large Scale Expression Trials

Large-scale trials were performed, transfecting only the IR or IGF1R construct individually, to evaluate the expression each construct in isolation. This would then allow the combination of promising constructs for co-expression and IR-IGF1R hybrid production. Large-scale expression trials were initially performed with the IR.Fc and IGF1R.Fc, as in small-scale trials these constructs showed relatively low levels of pro-receptor accumulation. Additionally, a control construct IR-TAP was utilised, as this was known to express well under the purification conditions. IR-TAP encodes the full-length IR with a c-terminal tandem affinity purification tag and was a gift from the Prof. Ünal Coskun. Briefly, HEK293 Freestyle cells were transfected
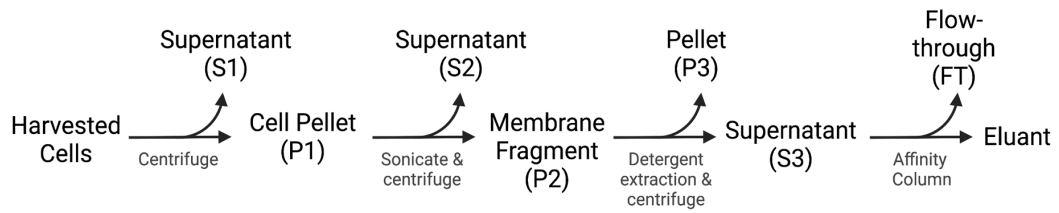
Figure 64 – Schematic overview of large-scale purification procedure. Symbols in brackets indicate the points at which aliquots were taken for assessment by western blotting (Figure 65)

with the relevant expression construct complexed with PEI and harvested after 72 hrs. The cells were pelleted, and a sample taken to assess the total protein expression, before disruption of the cell membranes by sonication. Cell debris was pelleted by gentle centrifugation before the supernatant containing the cell membranes was collected and subjected to ultracentrifugation. The pelleted cell membranes were resuspended to homogeneity and before solubilisation with the addition of 2% CHAPS. CHAPS was chosen as previous surfactant screens on the full-length purified IR showed this solvent produced minimal receptor autophosphorylation in the absence of insulin whilst permitting robust phosphorylation upon insulin stimulation[187]. Insoluble material was removed by a further round of centrifugation and the supernatant incubated with the relevant affinity purification beads to allow protein binding. The beads were then washed with buffer containing 5% glycerol and 1% CHAPS. These additives increased the hydrophobicity of the wash buffer and aid in removal of any contaminants non-specifically bound to the beads. Two further washes were carried out; the first contained ATP to remove chaperone proteins bound to the protein of interest, with the addition of ATP causing them to obtain a closed conformation and no longer bind to the protein of interest. The second wash contained EDTA to remove calmodulin binding proteins which can be eluted with the removal of Ca^{2+}. A final wash of 20% glycerol was used to remove residual hydrophobic contaminants. Next, GST-tagged 3C protease was applied to the column in 20% glycerol; this allowed on column cleavage of the protein of interest from the affinity tag whilst equilibrating the beads in a hydrophobic buffer to efficiently solubilise the released protein. The protein was then eluted from the column over several fractions, and the fractions assessed for purity and protein content by SDS-page with subsequent

Coomassie staining. Fractions containing the desired protein were pooled and incubated with glutathione Sepharose beads and subsequent elution to remove the GST-3C protease. Additionally, aliquots were taken at each stage of purification and analysed by western blotting to determine protein yield and quality at each stage.

Western blotting of the IR.Fc purification under reducing conditions showed the expected molecular weight bands for the IR pro-receptor and beta-subunit when probed with a 6 x His specific antibody. However, the band corresponding to the pro-receptor was much stronger than that for the beta subunit (Figure 65A), indicating that most of the receptor had not undergone furin cleavage and maturation inside the cell. Subsequently, only a limited amount of IR.Fc could be observed post elution from the affinity column (Figure 66B). The purification of IGF1R.Fc showed similar accumulation of the IGF1R pro-receptor by western
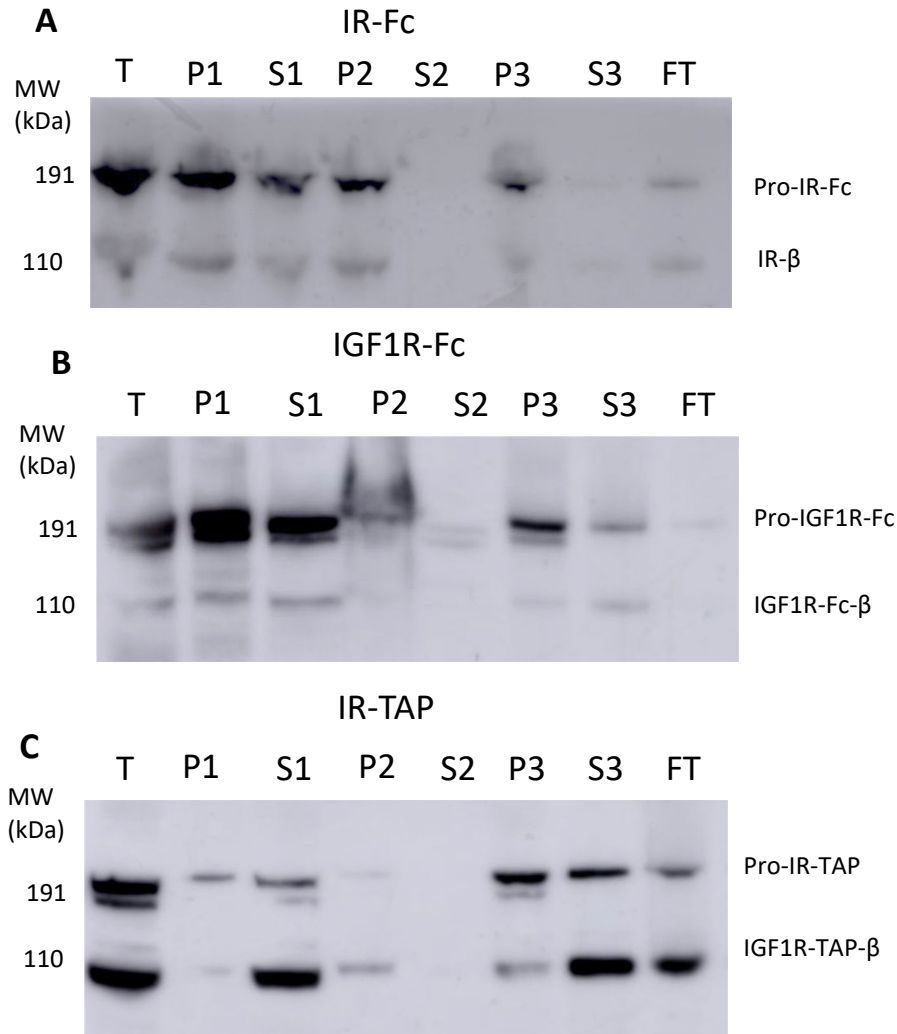
Figure 65 – Western blot analysis of IR-Fc and IGF1R-Fc purification. Labelling for fractions is T (Total), P1,2,3 (pellet 1,2,3) and S1,2,3 (supernatant 1,2,3) and FT (column flow-through), with the stage each aliquot taken indicated in Figure 64; A) Western blot for IR-Fc with 6 x His specific antibody; B) Western blot for IGF1R-Fc with 6 x His specific antibody; C) Western blot of IR-TAP with 6 x His specific antibody.

blotting (Figure 65B), and no bands could be observed eluting from the affinity column (Figure 66B). This is in contrast to the control construct IR-TAP which showed similar levels of pro-receptor and cleaved receptor in the total cell lysates (Figure 65C) and a much greater quantity of IR eluting from the affinity column (Figure 66A).

The cause of the differences in subcellular processing for the two IR constructs is not immediately clear as both constructs use identical nucleotide sequences to encode for the IR. Typically, the accumulation of receptor precursor is attributed to the nature or size of the affinity tag interfering with protein folding and transport, or excessive expression of protein precursor overloading cellular machinery[188]. The
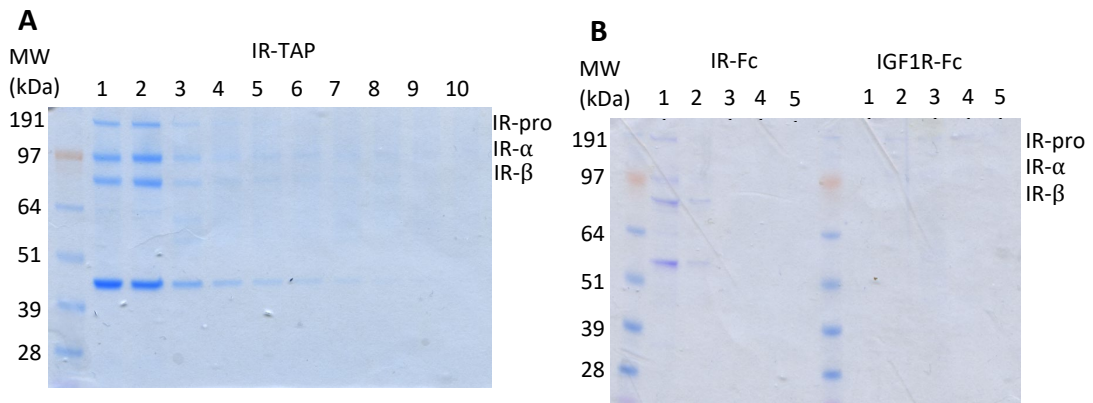
Figure 66 – Coomassie stained SDS gel of the elution fraction from IR and IGF1R affinity column purification. The top numbers indicate the elution fractions. The relative positions of the pro-IR (IR-pro), IRα subunit and IRβ subunits are shown A) Stained gel from the IR-TAP purification; B) Stained gel from the IR-Fc (left five columns) and IGF1R-Fc (right five columns) purifications.

differences between the expression vectors which are most likely to have resulted in different subcellular processing are the promotor region and the affinity tag. The IR-Fc and IR-TAP vectors use chicken β-actin and CMV promotors respectively, which are both suitable for mammalian expression. However, the ideal promotor region must be determined empirically for each receptor construct, and in this case, it may be that the CMV promotor is more suited to the IR construct. Alternatively, it may be the Fc tag is not compatible with the folding and processing of the IR and this results in the accumulation of receptor precursor. Due to time constraints, none of the alternative receptor constructs were further evaluated on large scale. In combination with the small-scale expression trials, these results indicate further optimisation of the receptor constructs is necessary to reduce pro-receptor accumulation. As the IR-TAP construct expresses well, it would be prudent to use the features of this vector as the basis of further construct design.

4.3   Conclusions

In this chapter, six IGF1R receptor mutants have been generated and their ability to disrupt hybrid formation has been assessed. All the generated mutants were expressed robustly, and the chimeric receptors IGF1R F450A, R391A and D555A all showed reduced affinity as part of a hybrid receptor. This shows for the first time that hybrid receptor formation can be manipulated at the protein level and highlights the importance of the L2: FnIII-1 interface in controlling hybrid receptor dimerisation. All the receptor mutants that showed reduced $BRET_{50}$ relative to the wild-type receptor were predicted by the KFC2b method on the AlphaFold homology model, implying that this model can be used to predict hotspot features at the L2: FnIII-1 interface with good accuracy. Additionally, optimisation of protein purification for the IR-IGF1R hybrid receptor was undertaken. IR and IGF1R constructs featuring several affinity tags were generated, for use in sequential rounds of affinity purification to purify hybrid protein for structural studies. However, in small and large-scale expression trials the chimeric receptors showed pro-receptor accumulation when transiently expressed, which was not observed with the control construct IR-TAP at large-scale. IR-TAP differs to the cloned constructs in its promotor region and the nature of its affinity tag, implying that IR and IGF1R cellular processing is extremely sensitive to one or both features. These features will require further optimisation to generate receptor constructs suitable for the proposed hybrid purification.

4.3.1   Progress Towards Project Goals

In this chapter, progress has been made towards aims 1 and 4. The IR-IGF1R homology models have been evaluated experimentally by the production of point mutants. This indicates the AlphaFold homology model generated in Chapter 2 can be successfully utilised in combination with KFC2b to predict hotspot residues which reduce the affinity of the IR-IGF1R L2: FnIII-1 interface. Additionally, progress has been made towards aim 4, by evaluating protein production methodologies for the generation of IR-IGF1R hybrid protein. The IR and IGF1R constructs tested showed pro-receptor accumulation, which was likely due to incompatibility of the attached affinity tags or vector promotor with robust expression of the receptor.

Further optimisation of construct design will be necessary to identify suitable IR and IGF1R constructs for protein production.

5 Conclusions and Future Work

5.1 Aims

This thesis has aimed to evaluate the molecular mechanisms affecting IR-IGF1R hybrid formation. This was performed by developing improved structural models of the IR-IGF1R hybrid receptor and evaluating these through a combination of structure-based drug design and mutagenesis studies. Additionally, methodologies for producing IR-IGF1R hybrid protein were investigated, with the aim of using such protein for structural studies to gain a high-resolution structure of the IR-IGF1R hybrids. These were undertaken to evaluate the possible role of the IR-IGF1R hybrid receptors in the development of insulin resistance and type 2 diabetes.

Hypothesis: Understanding the structure of IR-IGF1R hybrids and how to modulate their formation will further our understanding of their mechanism of action.

5.1.1 Aims

1. To improve the current hybrid homology model based on recent structural studies.

2. To use virtual high-throughput screening to identify novel small molecule modulators of hybrid formation.

3. To evaluate these small molecules using established screening cascades and structural studies.

4. To produce protein for structural studies on the IR-IGF1R complex.

5.1.1.1 *Improve the Current Homology Model Based Upon Recent Structural Studies.*

The discovery of **HI2**[113] utilised a homology model of the hybrid receptor based on the published structure[60] of the IR ectodomain dimer utilising the ITASSER server[114-116]. Subsequent publication of the IGF1R ectodomain dimer structure[61] and significant improvements in computational protein prediction[118] meant there was scope to improve this model. In this study, two separate models of the FnIII1-L2

interface were generated using TACOS and AlphaFold, incorporating updated structural information from the IGF1R ectodomain structure[61]. Computational assessment of these models showed good confidence in the accuracy of the predicted structures. The models were further evaluated experimentally by evaluating the affinity of several mutant hybrid receptor constructs, which contained point mutations chosen by KFC2 on the TACOS and AlphaFold model. Three of the six-point mutants evaluated showed reduced affinity of the IR and IGF1R relative to the wild-type receptor in donor saturation assay experiments. All the point mutants which showed reduced affinity were predicted from the AlphaFold model by KFC2b, indicating that this methodology is most promising for further studies.

This study could be expanded in several ways. Firstly, generating a full-length structure of the IR-IGF1R hybrid with AlphaFold multimer[167] would be beneficial. Whilst this study focussed on the FnIII1-L2 interface of the hybrid dimer, it is likely that further interfaces occur based on those seen in the IGF1R and IR dimer structures. Evaluation of these interfaces by hotspot prediction would provide further information on the function of the hybrid receptor. Additionally, mutagenesis studies could be expanded to include all the hotspot residues which have been computationally identified in this study, with priority given to those implicated by KFC2b on the AlphaFold model. This would provide a thorough evaluation of the generated models, whilst providing a more comprehensive insight into hotspot epitopes for further drug-discovery efforts.

5.1.1.2 *Virtual High-Throughput Screening to Identify Novel Small Molecule Modulators of Hybrid Formation*

The identification of small molecules **HI1** and **HI2** had set precedent for virtual screening directed against a homology model for the identification of modulators of hybrid formation[113]. The unfavourable physicochemical and pharmacokinetic properties of **HI1** and **HI2** meant that the identification of further hybrid modulators would be beneficial, both as tool compounds and leads for potential structure-based drugs design directed against the hybrid receptor. In this study, a virtual high-throughput screen was directed against the TACOS and AlphaFold

homology models. Whilst the FnIII1-L2 interface was relatively flat in both models, SiteMap[161] calculations indicated the presence of druggable sites at the receptor interface, which overlapped with hotspot epitopes computationally predicted by KFC2. This implies the potential to modulate hybrid formation at the L2:FnIII1 interface. Molecular docking of libraries containing 67,000 small molecules was performed with Glide[123-125, 163]. Docked poses were clustered and visually examined for their potential to block the formation of the FnIII-1: L2 interface. This prioritised sets of compounds for further biological evaluation. Additionally, the scaffold hopping software ROCs and EON were utilised **HI2** to identify alternative scaffolds that could be synthesised and tested.

Utilising larger screening collections in this workflow would result in larger coverage of chemical space, which in turn should result in an increase in the number and quality of potential hits found. However, it may be more prudent to focus on the validation of the protein models and predicted hotspots before any further screening is undertaken; the accuracy of virtual screening is reliant on each of these factors, and this will allow identification of the most effective sites to direct screening against. Additionally, studies identifying the mode of action for **HI2** would be beneficial, as if this molecule binds directly to either the IR or IGF1R and the binding epitope is characterised, then it would be preferable to target this region. However, the low solubility of **HI2** and the lack of related analogues which retain biological activity make such studies challenging.

5.1.1.3 *Evaluation of Small Molecules using Established Screening Cascades and Structural Studies*

A challenge in studying IR: IGF1R biology is differentiating between heterodimeric and homomeric proteins. In this study, BRET assays and immunoprecipitation procedures were utilised to specifically detect changes in hybrid formation. Screening of molecules identified by vHTS in the BRET assay identified the small molecule **Z13**, which reproducibly resulted in increases in hybrid formation, despite initial screening focussing on the identification of inhibitors. Further evaluation of **Z13** confirmed that hybrid formation was promoted in orthogonal immunoprecipitation-based assays at 100 µM. However, **Z13** was determined to

similarly promote IR and IGF1R homodimer formation to a similar extent. qRT PCR determined that **Z13** promoted the expression of mRNAs encoding the IR and IGF1R. Therefore, it is likely that **Z13** is promoting hybrid formation through increasing IR and IGF1R expression, rather than directly binding to the receptor. As this did not follow our original design hypothesis and would make eliciting selectivity extremely challenging, it was decided not to proceed with this series. However, the screening cascade utilised in this thesis was found to be effective at evaluating the effects of small molecules on hybrid formation and will be of use in further study.

5.1.1.4   *Production of IR: IGF1R Hybrid Protein for Structural Studies*

To further validate homology models, production of IR: IGF1R hybrid protein was attempted for use in cryo-EM studies. A specific challenge of IR: IGF1R hybrid purification is the separation of IR and IGF1R homodimers. To combat this, a protocol utilising two rounds of affinity purification on alternatively tagged IR and IGF1R receptor monomers was designed. Several IR and IGF1R constructs containing four separate affinity tags were cloned for this purification. However, small-scale expression trials of the tagged constructs showed the accumulation of receptor precursor inside the cell. When scaled up, this resulted in very low yields protein post purification. In contrast, purification of the control construct IR-TAP in parallel generated protein in good yield. This discrepancy led us to conclude that the accumulation of receptor precursor in the cloned constructs was likely due to incompatibility of the protein tag or vector promotor with robust receptor processing inside the host cells.

Whilst the current two-step purification strategy appears feasible, further optimisation of the IR and IGF1R fusion constructs is necessary to identify a suitable combination of compatible affinity tags which allow robust expression. As purification of the IR-TAP construct was efficient, it would only be necessary to further optimise the IGF1R construct. A potential next step could include subcloning the tagged IGF1R constructs generated in thesis into mammalian expression vectors containing a CMV promotor, which has proven to be an efficient promotor for the expression of IGF1R-TAP. Additionally, exploring alternative affinity tags for fusion

with the IGF1R may be necessary. Inspiration may be taken from the literature: for example, the expression and purification of the full-length IGF1R using a CMV promotor and C-terminal 3 x FLAG epitope tag for cryo-EM studies has been reported[75]. However, as this study involved generating a stable cell-line, transient expression of this construct would need to be evaluated. In summary, the current work provides a valuable foundation for studies aimed at optimising the expression of IR-IGF1R protein, with several avenues for future exploration.

5.1.2 Future Research on the IR: IGF1R Hybrid Receptor

This thesis has utilised a structure-based approach to evaluate mechanisms affecting IR-IGF1R hybrid formation. There has been a specific focus on the L2: FnIII-1 interface of the hybrid receptor, as structures of the IR and IGF1R homodimers suggested this region may be important in dimerisation of this receptor family. The prediction and evaluation of hotspot residues at this interface has determined that this region is important in the formation of the hybrid dimer and can be modulated at the protein level through point mutagenesis. Whether this region can be chemically modulated is not established: computational techniques indicated the presence of druggable sites in this region, however virtual-screening and subsequent biological testing of *ca.* 70,000 small molecules did not detect any ligands that could bind to and modulate this epitope. Synthesis and testing of several **HI2** analogues did not identify any related molecules with similar activity. Therefore, this molecule appears to be a poor starting point for further elaboration.

Future work in this area should focus on further characterisation of the IR-IGF1R hybrid receptor, particularly by structural techniques to fully characterise the receptor apo-state and pro-receptor. Whilst the hybrid receptor represents a formidable structural challenge, the recent characterisation of IGF1 bound to the hybrid receptor indicates such tasks are possible with current experimental techniques. Additionally, further hotspot validation would be beneficial to structurally map the most promising areas to target small molecule modulators. These experiments would provide the structural basis for rational design of hybrid

receptor modulators, increasing the prospect of identifying a small molecule that specifically interacts with this target.

6 Methodology

6.1 Cell Culture Protocols

6.1.1 Adherent cells

6.1.1.1 Thawing Procedure

Aliquoted cells were stored under vapour phase liquid nitrogen. Cells were thawed by warming the cryovial in a water bath (37 °C) and seeding into prewarmed cell culture medium (10 mL) in a 75 cm^2 cell culture flask. A full medium change was undertaken once the cells had adhered, no longer than 24 hrs post thawing.

6.1.1.2 Passaging Procedure

Cells were passaged at 90% confluency. Cells were washed in prewarmed Dulbecco's Phosphate Buffered Saline (DBPS, Sigma Aldrich, 10 mL, 37 °C) before the addition of prewarmed trypsin (Gibco, 0.05%, 2 mL, 37 °C) and subsequently incubated for approximately 2 mins until detached from the flask. Prewarmed culture medium (37 °C) was added to achieve the appropriate seeding density for subsequent subculture.

6.1.1.3 Cell Maintenance

Cells were maintained in a 75 cm^2 cell culture flask in a humidified incubator (37 °C, 5% CO$_2$). Cell incubations detailed in this thesis refer to these conditions unless otherwise specified.

6.1.1.4 Transfections – Lipofectamine 2000

HEK293 cells were seeded into a 6 well plate at 2.5 x 10^6 cells per well and incubated (24 hrs) to achieve approximately 70% confluency by transfection. For transfections, Lipofectamine 2000 (Invitrogen) and cDNAs were used at a ratio of 5:1 v/w. To prepare the transfection mixture, Lipofectamine 2000 and the relevant cDNA were each diluted into OptiMEM (50 μL per well to be transfected) (Gibco). The diluted Lipofectamine solution was then mixed with the diluted cDNA and incubated (5 mins, rt). The resulting transfection mixture was then added dropwise to each well (100 μL) and the plate rocked to ensure even distribution. The

transfected cells were then incubated for at least 12 hrs before harvesting or replating for further assays.

6.1.2 Adherent Cell Lines

6.1.2.1 Human Embryonic Kidney 293 Cells

HEK293 cells were cultured in DMEM (4.5 g L^{-1} glucose, GlutaMAX supplement, 10% Fetal Bovine serum, Penicillin 50 units mL^{-1}, Streptomycin 0.5 mg mL^{-1}, 37 °C) (Gibco) until passage 30.

6.1.2.2 Human Umbilical Vein Endothelial Cells

HUVECs from pooled donors were maintained in Endothelial Cell Growth Medium 2 (ECGM2) (PromoCell) supplemented with Growth Medium 2 Supplement Mix (PromoCell) containing the following growth factors: fetal calf serum (0.02 mL mL^{-1}), Epidermal Growth Factor (recombinant human, 5 ng mL^{-1}) Basic Fibroblast Growth Factor (recombinant human, 10 ng mL^{-1}), Insulin-like Growth Factor (Long R3 IGF, 20 ng mL^{-1}), Vascular Endothelial Growth Factor 165 (recombinant human, 0.5 ng μL^{-1}), Ascorbic Acid (1 μg mL^{-1}), Heparin (22.5 μg mL^{-1}), Hydrocortisone (0.2 μg mL^{-1}). HUVECs were cultured until passage five.

6.1.3 Suspension Cell Lines

6.1.3.1 FreeStyle 293F

FreeStyle 293F cells were stored in vapor-phase liquid nitrogen until use. Aliquoted cells were thawed by warming a cryovial (37 °C) and seeding into 125 mL polycarbonate Erlenmeyer shaker flask containing 30 mL of pre-warmed FreeStyle 293F Expression Medium (Gibco). Cells were incubated in an orbital shaker incubator (37 °C, 100 rpm, 76% humidity, 8.0% CO_2). Cells were subcultured once they had reached 1 x 10^6 viable cells mL^{-1}

6.1.3.2 Cell Maintenance

Freestyle 293F cells were maintained in FreeStyle™ 293 Expression Medium (Gibco) (Penicillin 50 units mL^{-1}, Streptomycin 0.5 mg mL^{-1}) and incubated in an orbital shaker incubator (37 °C, 100 rpm, 76% humidity, 8.0% CO_2). Cells were subcultured at *ca.* 2 x 10^6 cells mL^{-1} and seeded at 0.1 x 10^6 cells mL^{-1}.

6.2 BRET Screening Assay

HEK293 cells were seeded into 6-well plates at a density of 2.5 x 10^5 cells per well and incubated. Cells were transfected according to protocol 6.1.1.4. For experiments detecting hybrid receptors, cells were co-transfected with IR-Rluc cDNA and either IGF1R-YPET cDNA or empty pUC19 cloning vector cDNA to bring the total DNA to 0.6 μg per well. For experiments detecting homodimeric receptors, the combinations of IR-Rluc/ IR-YPET or IGF1R-Rluc/IGF1R-YPET were utilised, with the Rluc tagged construct co-transfected with empty pUC19. Transfected cells were incubated for 24 hrs. After 24 hours the transfected cells would typically be imaged using an Incucyte® Zoom (Sartorius) live cell imager to check the expression of YPET labelled protein and ensure transfection efficiency. Subsequently the cells were washed with DPBS (1 mL) (Sigma Aldrich) and the transfected cells pooled to reduce the effects of variation in well to well transfection efficiency. The cells were seeded into 96 well plate at a density of 1.5 x 10^4 cells per well and incubated. After 24 hrs, cells were treated with relevant compounds in DMEM media (100 μM, 200 μL), diluted from DMSO stock (10 mM). After a further 24 hrs incubation, treated media was removed, wells were washed with DPBS (200 μL) and cells retained in DPBS (50 μL) for measurements. Plasmids encoding the constructs IR-Rluc, IGF1R-Rluc, IR-YPET and IGF1R-YPET were gifts from Dr. Tariq Issad (INSERM institute, Paris).

6.2.1.1 BRET Measurements

Measurements were performed on a Envision® 2105 Multimode Plate Reader (Perkin Elmer). A solution of coelenterazine-h in ethanol (250 μM, 1 μL) was added to each well to give a final concentration of 5 μM in each well. Light-emission acquisition at 485 nm and 535 nm was started immediately. BRET signal was expressed in milliBRET unit (mBU). The BRET unit has been defined previously as the ratio 535 nm/485 nm obtained when the two partners are present, corrected by the ratio 535 nm/485 nm obtained under the same experimental conditions, when only the partner fused to *Renilla luciferase* is present in the assay. This can be represented as:

$$BRET = (E_{535} \div E_{485}) - Cf$$

Where E_{535} corresponds to the luminescence at 535 nm, E_{485} corresponds to the luminescence at 485 nm, and Cf corresponds to the ratio E_{535}/E_{485} for the Rluc tagged construct transfected alone.

6.2.2   IGF1 Treatments

Cell culture and transfections were performed as in section 6.2. After 24 hrs post transfection, cells were changed into serum free DMEM (Gibco), containing no FBS supplement, and incubated (24 hrs). IGF1 (Antibodies.com) dissolved in deionised water was added to each well at a final concentration of 100 nM and incubated (37 °C, 15 mins). Subsequently, the BRET ratio was measured as described in the section 6.2.1.1.

6.3   LIVE-DEAD Cytotoxicity Assay

Assays was performed using the LIVE-DEAD Viability/Cytotoxicity Kit (Invitrogen) with minor modifications to the manufacturers protocol. HEK293 cells were seeded into transparent Poly-L-Lysine coated 96-well plates at a density of 60,000 cells per well. After 24 hrs, cells were treated with relevant compounds in DMEM media (100 μM, 200 μL) (Gibco), diluted from DMSO stock (10 mM). After a further 24 hours, treated media was removed and wells were washed with DPBS (200 μL) (Sigma Aldrich). In a separate flask, Ethidium Homodimer-1 stock solution (EthD-1, component B, 2 mM, 20 μL) was added to DPBS (10 mL). To this, Calcein AM stock solution (Component A, 4mM, 5 μL) was added to give a solution of EthD-1 and Calcein AM in PBS at approximately 4 μM and 2 μM respectively. 100 μL of this solution was added to each well and cells were incubated for 30 minutes. Wells were imaged using an Incucyte® ZOOM (Sartorius) live cell analysis system, with fluorescence measurements at 530nm and 645 nm recorded. Images were analysed using Incucyte® Zoom software and cell viability calculated with the following formula:

$$Cell\ Viability = \frac{\%\ Red\ Area}{\%\ Green\ Area}\ x\ 100$$

6.4   Western Blotting

6.5   Buffer Preparations

6.5.1.1   Tris-buffered Saline Tween 20 (TBST)

20 mM Tris, 150 mM NaCl, 0.1% Tween 20, pH 7.4 with HCl/NaOH

6.5.1.2   Stripping Buffer

37.2 mM Glycine, 34.7 mM SDS, pH 2.2 with HCl

6.5.1.3   IP Binding Buffer

100 mM HEPES, pH 7.8,100 mM NaCl, 10 mM MgSO4, 0.02% Tween-20

6.5.1.4   Phosphate Buffered Saline Tween 20 (PBS-T)

PBS-T (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM $KHPO_4$, 0.02% Tween-20, 200 mM Na_3VO_4, 1 µL/mL Protease inhibitor cocktail)

6.5.2   Sample Preparation

6.5.2.1   Transfected Cells

HEK293 cells were transfected according to 6.1.1.4 with 0.3 µg of the relevant cDNA per well. After 24 hrs, cells were washed with ice-cold DPBS (1 mL) (Sigma Aldrich) and each well lysed by addition of lysis buffer (80 µL) (Invitrogen) on ice. The lysates were pre-cleared by centrifugation (14,000 g for 10 min at 4°C).

6.5.2.2   Compound-treated Cells

HUVECs were seeded into 6 well plates 1.5 x 10^5 cells per plate. After 48 hours cells were treated with relevant compounds dissolved in ECGM2 media (1 mL, 100 µM) (PromoCell), diluted from DMSO stock (10 mM). After 24h the cells were lysed in cell lysis buffer (80 µL) (Invitrogen) over ice. Cell lysates of 2 wells were pooled into a total volume of 80 µL. The lysates were pre-cleared by centrifugation at 14,000 g for 10 min at 4°C.

6.5.2.3   Protein Quantification

The protein concentration of cell lysates was determined using the Pierce BCA Protein Assay Kit (Thermo Scientific). Diluted albumin standards were prepared according to the manufacturers protocol. For each sample, 8 µL of protein lysate

was diluted into 52 µL of water. 25 µL of diluted sample or albumin standard was plated into a clear 96 well plate in duplicate. The BCA working reagent was prepared by mixing BCA reagent A with BCA reagent B at a 50:1 ratio, and 200 µL of working reagent added to each sample or BCA standard. The plate was sealed and incubated for 30 mins at 37 °C. Subsequently, the absorbance of each well at 562 nm was determined. A standard concentration curve was generated from the albumin standards and cell lysate protein concentrations calculated by comparison to this concentration curve.

6.5.3   Gel Electrophoresis Procedures

6.5.3.1   Immunoblotting Procedure 1

50 µg of protein were mixed with NuPAGE™ LDS sample loading buffer (4X) (Invitrogen), NuPAGE™ sample reducing buffer (10X) (Invitrogen) and water to give the appropriate sample volume, before heating to 95 °C for 5 mins to ensure protein denaturation. Samples were loaded onto NuPAGE™ 4 to 12%, Bis-Tris gels (Invitrogen) alongside molecular weight markers samples and separated by SDS-PAGE electrophoresis for 90 mins at 110 V. Samples were then transfered to nitrocellulose membranes (Bio-Rad) using the TransBlot® Turbo Transfer System (Bio-Rad). The membrane was dried and incubated in 2% BSA TBST for 1hr to block non-specific binding. Membranes were treated overnight with primary antibodies at the relevant concentrations described in section 7. Subsequently, the membranes were washed 3 x 10 mins in TBST before treatment with horseradish-peroxidase conjugated secondary antibodies of the relevant species (1:5000) in 5% BSA TBST for 1 hr. The membranes were then washed in 3 x 10 mins in TBST before visualisation.

6.5.3.2   Protein Visualisation

The immunoreactive proteins were visualised using the Western Chemiluminescent HRP substrate reagents (Immobilon) before imaging on the G:Box (Syngene) system. Blots were then stripped by application of stripping buffer for 15 mins before 3 x 10 mins of washes with TBST. Membranes were stored dry at 4°C. Images were analysed using the Image J software[189], with each band normalised to the signal of β-actin from the same sample for analysis.

6.5.4   Immunoprecipitation

Immunoprecipitation of receptors containing an IGF1R component was performed by the addition of Protein G-Agarose beads (30 µL) (Roche), IGF1R D23H3 antibody (3 µL) (Cell Signalling) and cell lysate (80 µg total protein) to 3 ml of binding buffer. Samples were incubated (16 h, 4°C) with gentle mixing (22 rpm). The samples were briefly centrifuged (500 x g, 1 min, 4 °C) and supernatant removed and collected. The precipitates were washed six times in PBS-T with brief centrifuging between washes (500 x g, 1 min, 4 °C).

For gel electrophoresis, immunoprecipitants were resuspended in NuPAGE™ LDS sample loading buffer (10 µL) (Invitrogen), NuPAGE sample reducing buffer (4 µL) (Invitrogen) and water (26 µL) before heating (98 °C, 5 min).

Supernatant samples (28 µL) were similarly combined with NuPAGE™ LDS sample loading buffer (10 µL) (Invitrogen), NuPAGE sample reducing buffer (4 µL) before heating (98 °C, 5 min) (Invitrogen). Samples were kept on ice or stored at -20 °C until electrophoresis.

6.6   Reverse-Transcription Quantitative Polymerase Chain Reaction

HUVECs were maintained as in section 6.1.2.2 and treated with the relevant compound in ECGM2(100 µM, 24 hr) (PromoCell). Subsequently, cells were washed in DPBS (1 mL) (Sigma Aldrich), lysed by addition of 0.05% trypsin (400 µL) (Gibco) and neutralized using ECGM2 (1 mL). Two wells of the 6-well plate, sharing the same compound or vehicle-treatment were combined and centrifuged (1000xg, 8 min, 4°C). The cells were resuspended in DPBS (1 mL) (ThermoFisher) and centrifuged once more under the same conditions. RNA extraction was performed using the Monarch Total RNA Miniprep Kit (NEB), according to manufacturer's protocol. The resultant RNA samples were placed on ice for duration of the experiment. RNA quantitation was performed using the Nanodrop DS-11 FX+ Spectrophotometer/Fluorometer (DeNOVIX) system. Reverse transcription of RNA to cDNA was performed by combining the RNA samples (500 ng/ mL) with the LunaScript™ RT SuperMix (5X) (NEB) in 20 µL aliquots. cDNA samples were diluted in nuclease-free water (NEB) before use in qPCR reactions. The qPCR samples were prepared by combining the cDNA samples with iTaq™ Universal SYBR® Green Supermix (Bio-Rad) and analysed by the LightCycler® 96 system (Roche) using the

following primers: IGF1R - PrimePCR™ SYBR® Green Assay: IGF1R, Human (Bio-Rad); INSR - PrimePCR™ SYBR® Green Assay: INSR, Human (Bio-Rad) and ACTB - PrimePCR™ SYBR® Green Assay: ACTB, Human (Bio-Rad). The cycles to threshold (cT) was measured for each well.

6.7 HPLC Solubility Assay

Solubility samples (100 µM solution in PBS, 1% DMSO) were prepared by mixing compound stock solution in DMSO (10 µL, 10 mM) with DPBS buffer (990 µL, pH 7.4) (Sigma Aldrich) for 5 seconds on vortex mixer. Samples were agitated for 2 hrs on a shaker (500 rpm, 20 °C) and then separated by centrifugation (14000 rpm, 15 min). The supernatant (200 µL) was transferred to a vial (2 mL) containing DMSO (50 µL) and vortexed for 5 seconds to avoid precipitation from the saturated solution.

External standards were prepared by mixing the same compound DMSO stock solution used in solubility samples (10 µL, 10 mM) with further DMSO (990 µL) to produce a 100 µM solution of the relevant compound in DMSO.

The concentration of the solubilized compound in the solubility sample determined by high-performance liquid chromatography (HPLC) with UV detection. A calibration curve was prepared by injecting 0.5, 2.5, and 5 µL of compound external standard. The compound solubility value is obtained by injecting 6 and 20 µL of compound solubility sample. The compounds propranolol (Stratech Scientfic) and ketoconazole (Flourochem) were used as solubility standards.

The HPLC method is as follows: 5-minute gradient elution from 90:10 to 10:90 water: methanol (both modified with 0.1% formic acid) at a flow rate of 1.5 mL/min at 30°C.

6.8 Fluorescence Microscopy

HEK293 cells were maintained and transfected as in 6.2. Cells were transfected with IR-Rluc and empty pUC19 vector, or IR-Rluc and IGF1R-YPET mutants and incubated (24 hrs). The media was then replaced with OptiMEM (1 mL) (Gibco), and cells imaged on an EVOS cell imaging system (ThermoFisher) utilising the GFP filter.

6.9   Mutagenesis Procedure

Point mutations were integrated by including the desired nucleotide change into the forward primer, with at least 10 complimentary nucleotides present on the 3' end of the mutation. Reverse primers were designed so the 5' end of the forward and reverse primer anneal back-to-back. Primers were designed with NEBasechanger (https://nebasechanger.neb.com/) in the first instance. PCR was performed with Platinum™ SuperFi II PCR Master Mix (Invitrogen). PCR reactions were set up as follows:

Reagent	Volume / µL	Final Concentration
2X Platinum SuperFi II PCR mastermix	25	1X
Nuclease Free Water	19	-
Forward Primer (10 µM)	2.5	0.5 µM
Reverse Primer (10 µM)	2.5	0.5 µM
Template DNA (1 ngµL^{-1})	1	1 ng total DNA

Thermocycling conditions are detailed in below:

Cycles	Step	Temp/ °C	Time / s
1	Initial denaturation	98	30
25	Denaturation	98	5
	Annealing	60	10
	Extension	72	300

| 1 | Final extension | 72 | 300 |
| 1 | Hold | 4 | ∞ |

The PCR reactions were evaluated by agarose gel electrophoresis (1% TAE) to ensure robust amplification. PCR reactions which resulted in a single band of the correct size were purified using QIAquick PCR purification kit (Qiagen). The purified PCR reaction (1 μL, 5-10 ng) was then added to a mixture of nuclease free water (5 μL) (NEB), 10X T4 Ligase buffer (1 μL) (NEB), T4 PNK (1 μL), T4 DNA ligase (1 μL) (NEB) and DpnI (1 μL) (NEB). The reaction was incubated at room temperature for 1 hr. 5α competent E-coli (50 μL) (NEB) stored at -80 °C were thawed on ice. Immediately upon thawing, reaction mixture (1 μL) was added to the 5α competent E-coli (50 μL). The sample was flicked 2-3 to mix and incubated on ice (30 mins). The sample was heat shocked (42 °C, 45 s) and then cooled on ice for a further 2 minutes. An aliquot of SOC media (950 mL) (NEB) was added to each sample and then incubated (37 °C, 1 hr, 220 rpm) on a benchtop incubator shaker. The final mixture (100 μL) was spread onto selective agar plates (100 μg mL^{-1} AmpC) and incubated (37 °C, 16 hr). After incubation, single colonies were picked from the agar plates and used to inoculate overnight cultures in selective LB broth (10 mL, 100 μg mL^{-1} AmpC) which were incubated (37°C, 16 hr, 230 rpm). An aliquot of the overnight culture (500 μL) was mixed with 50% glycerol in water (500 μL) and frozen at -80 °C to make a glycerol stock. The overnight cultures were centrifuged (4500 rpm, 4 °C, 10 min) and supernatant removed. The pelleted bacteria were then miniprepped according to section 6.11.2.4.

6.10 Donor Saturation Assay Procedure

HEK293 cells were seeded into a 12 well plate with 2 x 10^5 cells per well. The cells were incubated for 24 hrs. The cells were transfected with plasmid using 110 ng of donor, increasing ratios of acceptor (x 0-20), and empty pUC19 maintain total DNA at 2130 ng per well (12 well). Transfections were performed at approximately 60% confluency. Plasmid DNA was diluted in OptiMEM (200 μL) (Gibco) and PEI (11.5 μL of a 1 mg mL^{-1}, 5:1 w/w PEI: DNA) (Polysciences) added to the DNA solution. The solution was immediately vortexed (5s) and incubated (RT, 15 mins). The

transfection mixture (200 µL) was added dropwise to each well and the plate rocked to ensure even distribution over the well. The cells were incubated with the transfection reagents for 4 hrs, after which the media was exchanged for fresh media and the cells incubated for a further 44 hrs. Cells were then seeded at 5×10^5 cells per well into white 96-well plate and incubated for 24 hrs. Simultaneously, the cells were seeded into a black 96 well plate. Cells were washed with DPBS (200 uL per well) (Sigma Aldrich) and resuspended in DPBS (50 µL). BRET measurements were recorded as in section 6.2.1.1. Fluorescence measurements were recorded at 535 nm after excitation at 513 nm on cells in black plates. Luminescense and flourescense measurements were then converted to protein quantities using the relationships established in section 4.1.3.

6.10.1 Radioligand Binding Assay Procedure

HEK293 cells were seeded into 6-well plates at 2.5×10^6 cells per well. After 24 hrs, the cells were transfected with IR-Rluc or IGF1R-YPET according to protocol 6.1.1.4. Several transfections were performed with 0.15-0.6 µg cDNA per well. 24 hours later, cells were plated into a 96-well plate with 50,000 cells per well and incubated for a further 24 hours. Subsequently, the cells were washed with DPBS (100 µL per well) (Sigma Aldrich) and then maintained in DPBS (50 uL). For IR-Rluc transfected cells, coelenterazine-h was added to a final concentration of 5 µM and the luminescence recorded at 485 nm. For IGF1R-YPET transfected cells, the fluorescence at 535 nm was recorded for each well after excitation at 513 nm.

For radioligand binding, ^{125}I-insulin or ^{125}I-IGF1 (PerkinElmer) were diluted to the working concentration with unlabelled insulin or IGF1 respectively. This was then further diluted in DPBS to give six separate ligand concentrations spanning 0.5-10 times the Kd value for the relevant receptor-ligand complex. An aliquot of each radioligand was then removed and unlabelled insulin or IGF1 added at a final concentration of 500 nM (*ca.* 100 x Kd) to allow the measurement of non-specific radioligand binding. A further aliquot of radioligand was retained to determine the specific activity of each radioligand concentration. The radioligand was then added to the cells in duplicate wells for each concentration value and the cells incubated (3 hrs, 37 °C). Post-incubation, the cells were washed with ice-cold DPBS (100 µL)

and lysed by addition of cell lysis buffer (50 μL) (Invitrogen). Cell lysates were removed into counting vials, and each well washed further lysis buffer (50 μL). Vials were counted on an AMG Automated Gamma Counter (Hidex). The K_d and B_{max} values were determined using GraphPad Prism using non-linear regression model with single site fitting.

6.11 Cloning Procedures

6.11.1 Cloning Protocol for the Generation of Tagged IR and IGF1R Constructs

6.11.1.1 Primer Design

NEBuilder Assembly Tool (Ver. 2.7.1, https://nebuilder.neb.com/#!/) was used to inform primer design. Primers were designed with two components: the 3' end contained a gene specific sequence to amplify the full-length IR or IGF1R sequence from the IR-Rluc or IGF1R-YPET vectors respectively; the 5' end contained an overlap region, which added a *ca.* 40 base pair overlap with each of the POPINGS, POPIN-3C-HALO7, pHL-avitag3 and pHL-FcHis vectors upon PCR amplification of the gene. The primer sequences used to amplify each insert are detailed below:

Gene	Backbone Vector	Direction	Sequence (5' → 3')
IR	POPINGS	For	cgtgagccttctctccgtgctgctgatgggttgcgtagctCACCTGTACCCCGGAGAG
IR	POPINGS	Rev	gatggtgatggtgatgttttcgaactgcgggtggctccaGGCTAGAGCTAGCGCGGAAG
IR	POPIN-3C-HALO7	For	ctgtaaattacattttatttacaatcaaaggagatataccATGGGCACCGGGGGCCGG
IR	POPIN-3C-HALO7	Rev	taccgatttcggatcccgggccctgaaacagaacttccagGGCTAGAGCTAGCGCGGAAGGATTGGAC
IR	pHL-avitag3	For	cgtgagccttctctccgtgctgctgatgggttgcgtagctCACCTGTACCCCGGAGAG
IR	pHL-avitag3	Rev	gccattcaatcttctgagcttcaaagatatcattcagaccGGCTAGAGCTAGCGCGGAAG
IR	pHL-FcHis	For	cgtgagccttctctccgtgctgctgatgggttgcgtagctCACCTGTACCCCGGAGAG
IR	pHL-FcHis	Rev	ttttgtcacaagatttggggccctggaacagcacctccagGGCTAGAGCTAGCGCGGAAG

IGF1R	POPINGS	For	cgtgagccttctctccgtgctgctgatgggttgcgtagctGAAATCTGCGGGCCAGGC
IGF1R	POPINGS	Rev	gatggtgatggtgatgttttttcgaactgcgggtggctccaGTAGCGACCGGTGGTGGAA
IGF1R	POPIN-3C-HALO7	For	ctgtaaattacattttatttacaatcaaaggagatataccGTAGCGACCGGTGGTGGAA
IGF1R	POPIN-3C-HALO7	Rev	taccgatttcggatcccgggccctgaaacagaacttccagGTAGCGACCGGTGGTGGAA
IGF1R	pHL-avitag3	For	cgtgagccttctctccgtgctgctgatgggttgcgtagctGAAATCTGCGGGCCAGGC
IGF1R	pHL-avitag3	Rev	gccattcaatcttctgagcttcaaagatatcattcagaccGTAGCGACCGGTGGTGGAA
IGF1R	pHL-FcHis	For	cgtgagccttctctccgtgctgctgatgggttgcgtagctGAAATCTGCGGGCCAGGC
IGF1R	pHL-FcHis	Rev	ttttgtcacaagatttggggccctggaacagcacctccagGTAGCGACCGGTGGTGGAA

6.11.1.2 PCR Generation of Gene-inserts

Vector specific inserts were generated by PCR amplification of the IR-Rluc or IGF1R-YPET plasmid utilising the relevant primer pairs, with the components of these PCR reactions as follows:

Component	Volume / µL
Q5 Reaction Buffer (5X)	5
dNTPs (10 mM)	0.5
Q5 DNA polymerase	0.25
Nuclease Free Water	10.75
Q5 High GC Enhancer (5X)	5
Forward Primer (10 µM)	1.25
Reverse Primer (10 µM)	1.25
Template DNA (1 ng uL^{-1})	1
Total Volume	**25**

Thermocycling was performed as follows:

Number of cycles	Step	Temperature /°C	Time /s
1 cycle	Initial Denaturation	98	30
30 cycles	Denaturation	98	5
	Annealing	72	20
	Elongation	72	80
1 cycle	Final Extension	72	120 s
1 cycle	Hold	4	∞

The primer annealing temperatures were calculated to be higher than 72 °C by NEB T_m Calculator (Ver 1.15.0, https://tmcalculator.neb.com/#!/main) due to the length of the primers. The manufacturer recommends annealing temperatures should not exceed 72 °C for Q5 DNA polymerase (NEB) and therefore the annealing temperature were set at 72 °C for this protocol.

Once PCR reactions were completed, the reactions were treated with FastDigest DpnI (1 µL) (Invitrogen) and incubated (37 °C, 1 hr) to remove residual template vector.

A portion (10 µL) of PCR reactions were separated and visualised by agarose gel electrophoresis (1% agarose, TAE). If this analysis showed the PCR product as a single clean band of the correct size, then the reaction was purified using the QIAquick PCR purification protocol. If agarose gel analysis showed smearing or multiple bands in the PCR sample, then the band of the expected size was extracted using the PureLink Gel Extraction protocol (6.11.1.4) and then further purified with the QIAquick PCR Purification protocol (6.11.1.5) to remove residual guanidinium salts from the gel extraction which inhibited the subsequent HIFI assembly ligation.

6.11.1.3 Linearisation of Backbone Vectors by Restriction Digest

Each of the pOPINGS, POPIN-3C-HALO7, pHL-avitag3 and pHL-FcHis vectors were linearised by double restriction digest. The restriction enzymes utilised to digest each plasmid are detailed below:

Plasmid	Enzyme 1	Enzyme 2
pHL-FcHis	AgeI	KpnI
pHl-Avi	AgeI	KpnI
pOPIN-3C-HALO7	NcoI	PmeI
pOPINGS	AgeI	NheI

Reactions were prepared on ice, containing uncut vector DNA (500 ng μL^{-1}, 32 μL), FastDigest Green Buffer (4 μL) (Invitrogen) and the two relevant restriction enzymes (2 μL each) as indicated in a total reaction volume of 40 μL. The reactions were incubated (37 °C, 1 hr) before heat deactivation (80 °C, 15 mins). Reaction components were separated by agarose gel electrophoresis (1% agarose in TAE), loaded over multiple wells if necessary. The subsequent gel was imaged under a blue-light LED transilluminator (BT Labsystems), and relevant bands corresponding to the doubly digested vector excised with a scalpel.

The excised bands were purified according to the PureLink™ Quick Gel Extraction Kit protocol (6.11.1.4).

To remove residual guanidinium salts from the gel extraction procedure, which were found to inhibit the subsequent HiFi Assembly ligation procedure, resultant DNA was further purified with the QIAquick PCR Purification Kit (Qiagen) according to the manufactures protocol.

6.11.1.4 Purelink Quick Gel Extraction Kit Protocol

This procedure was performed with the Purelink Quick Gel Extraction Kit (Invitrogen). Gel slices were weighed and then dissolved in 3 volumes of gel solubilisation buffer (buffer L3) whilst being heated in a water bath (50°C, inverted every 3 mins to mix). Once the gel appeared dissolved, the gel slice was incubated (50 °C) for a further 5 mins to ensure complete dissolution of the gel slice. An additional 1 gel volume of IPA was added and the dissolved gel slice and the

solution mixed. The resulting solution was applied to a Quick Gel Extraction Spin Column and centrifuged (12,000 x g, 1 min). The flow through was discarded, and the spin column washed with Wash Buffer (buffer W1). The wash buffer was removed by centrifuging (12,000 x g, 1 min) and flow-through discarded. The spin column was centrifuged (12, 000 g, 2 min) to remove residual wash buffer, and the DNA eluted by adding Elution Buffer (buffer E5) to the centre of the column, incubating (5 mins, rt) and centrifuging (12,000 g, 1 min) to capture the flow-through in a clean Eppendorf tube.

6.11.1.5 QIAquick PCR Purification Kit Protocol

This procedure was performed with the QIAquick PCR Purification Kit (Qiagen). Five volumes of buffer PB were added to one volume of DNA with additional sodium acetate (3 M, 10 µL, pH 5.0), ensuring the resultant solution was yellow to indicate pH < 7.5. The sample was loaded onto a QIAquick column, centrifuged (17, 900 x g, 1 min) and flowthrough discarded. The column was washed with Buffer PE (750 µL), centrifuged (12,000 x g, 1 min) and flowthrough discarded. The column was centrifuged (17, 900 x g, 1 min) an additional time to remove residual wash buffer. DNA was eluted by adding Buffer EB (50 µL) directly to the spin column, incubating (2 min, rt) and centrifuging (17,900 x g, 1 min) to capture the flow-through in a clean Eppendorf tube. Resultant DNA concentration and purity was evaluated using a NanoDrop (DeNovix).

6.11.2 HiFi Assembly Reaction

The NEB HIFI assembly reaction was performed to ligate the PCR generated gene-inserts into restriction digested backbone vectors. The amount of DNA added was adapted to the varying concentrations of prepared insert and vectors but always remained between 0.075-1.5 pmol total DNA per reaction. Reactions were comprised of NEBuilder HiFi DNA Assembly Master Mix (2.5 µL) (NEB), insert DNA, vector DNA and nuclease-free water (NEB) up to a total volume of 5 µL. The reaction was incubated (50 °C, 1 hr) and placed on ice for subsequent transformation. 2 µL of HIFI reaction would be used to transform 50 µL of 5-alpha competent *E. Coli* (NEB).

6.11.2.1 Transformation Protocol

5-alpha Competent *E. Coli* (50 μL) (NEB) were thawed on ice. Once thawed, the appropriate volume of reaction mixture was added immediately to the cells. The sample was mixed by flicking the tube 4–5 times and placed on ice for 30 minutes. The mixture was heat shocked (42°C, 45 s). Tubes were immediately transferred back onto ice for 2 minutes. Subsequently room temperature SOC medium (950 μL) (NEB) was added samples incubated (37°C, 60 mins, 220 rpm) on a shaker incubator. Agar selection plates containing ampicillin (100 μgmL^{-1}) (Sigma Aldrich), X-Gal (20 μgmL^{-1}) (Thermo Scientific) and IPTG (1 mM) (Invitrogen) were warmed to 37°C for 30 minutes. 100 μL and 900 μL of each sample was spread onto two different selection plates to increase the chance of obtaining single colonies. The plates were incubated overnight (37 °C, *ca.* 18 hr) and checked for the formation of white colonies the following morning.

6.11.2.2 Colony PCR

To screen whether bacterial colonies contained the desired insert, colony PCR was performed on individual bacterial colonies. PCR reactions containing the forward primer (0.5 μL, 10 mM in nuclease free water), reverse primer (0.5 μL, 10 mM in nuclease free water), OneTaq® Quick-Load® 2X MasterMix with Standard Buffer (12.5 μL) (NEB) and nuclease-free water (11.5 μL) (NEB) were prepared on ice. The following primers were utilised for colony PCR:

Primer	Direction	Sequence (5' → 3')
IR	Forward	caagatgaggcaaccttcctggagattgt
IGF1R	Forward	gcagtataaccccaagatgaggccttcctt
Backbone	Reverse	gatttgccctcccatatgtccttccgagtg

Subsequently, a single colony grown overnight on a selective agar plate was picked with a sterile pipette tip, touched to a selective replica plate, and the submersed in the appropriate PCR reaction. Once all samples had been prepared, the replica plate was incubated (37 °C, *ca.* 8 hr) until colonies could be visualised. Simultaneously, the PCR reactions were thermocycled utilising the following conditions:

Number of cycles	Step	Temperature /°C	Time /s
1 cycle	Initial Denaturation	94	300
30 cycles	Denaturation	94	15
	Annealing	Variable	15
	Elongation	68	120 s
1 cycle	Final Extension	68	300 s
1 cycle	Hold	4	∞

PCR reactions were then separated and visualised by agarose gel electrophoresis (1% agarose, TAE). Samples giving bands of the correct size were then selected from the replica plate for the inoculation of liquid cultures.

6.11.2.3 Overnight Cultures

Bacterial colonies for which colony PCR had indicated they contained the correct insert were picked with a sterile pipette tip and placed into a falcon tube (50 mL) containing selective LB Broth (10 mL, 100 µg mL^{-1} ampicillin) (Fisher Bioreagents). The tops were fastened loosely, and the samples incubated overnight (*ca.* 18 hr, 37 °C, 220 rpm)

6.11.2.4 QIAprep Miniprep Protocol

This protocol was performed using the QIAprep Miniprep kit (Qiagen). The pipette tip was removed from overnight cultures and the cultures centrifuged (4000 x g, 10 mins). The supernatant was removed, and the pellet dried by inverting the falcon tube onto a paper towel for 5 mins. The dried pellet was resuspended in buffer P1 (250 µL), briefly vortexed to ensure no clumps of bacteria were remaining and transferred to a microcentrifuge tube. Buffer P2 was added (250 µL, containing lyse blue reagent) and the sample immediately inverted several times until a homogenously blue coloured suspension was achieved. Immediately, buffer N3 (350 µL) was added, and the sample inverted until a cloudy colourless suspension had formed. The sample was centrifuged (10 min, 17,900 x g). The supernatant (800 µL) was applied to a QIAprep 2.0 spin column, and the column centrifuged (1 min, 17,900 x g) with the flow-through discarded. The column was washed twice, firstly

with buffer PB (500 µL) and secondly with buffer PE (750 µL), by applying the appropriate buffer to the top of the column, centrifuging (1 min, 17,900 x g) and discarding the supernatant. Residual wash buffer was removed by centrifuging (1 min, 17,900 x g) before buffer EB was applied to the top of the column, incubated for 5 mins, and then centrifuged (1 min, 17,900 x g) into a clean Eppendorf tube. The resultant plasmid DNA was utilised directly for downstream applications.

6.12   Protein production

6.12.1.1 Preparation of cDNAs for Transfection

Plasmids were transformed into D5Hα following the standard transformation protocol. Colonies were picked from the subsequent selective agar plate and used to start small (6 mL, 100 µg mL^{-1} AmpC, LB broth) which were incubated (37°C, 180 rpm, 8 hr). An aliquot (1 mL) of each culture was used to inoculate a LB broth (700 mL, 100 µg mL^{-1} AmpC) which was incubated overnight with shaking (37°C, 180 rpm).

The following day the cultures were purified with the NucleoBond Xtra Maxi Plus EF kit (Machery-Nagel) according to manufacturer's instructions.

Briefly, the overnight cultures (700 mL) were split into two aliquots and centrifuged (6,000 x g, 10 min, 4°C) and supernatant discarded. The pellet was resuspended in buffer RES-EF (12 mL), ensuring the pellet was completely resuspended and no clumps were remaining. Buffer LYS-EF (12 mL) was added to the resulting suspension, the suspension mixed gently by inverting five times and subsequently incubated (rt, 5 min). Meanwhile, buffer EQU-EF was added around the rim of Nucleobond Xtra Columns with inserted column filter and the column allowed to empty by gravity flow. To the bacterial suspension was added buffer NEU-EF (12 mL), the resulting suspension mixed by immediately inverting 5 times and incubated (on ice, 5 mins). Subsequently, the crude lysates were clarified by inverting to mix and applying to the equilibrated Nucleobond Xtra Column Filter, which was allowed to empty by gravity flow. The column and filter were washed with buffer FIL-EF (10 mL), allowed to empty by gravity flow, and the column filter removed. The column was washed with buffer Endo-EF (90 mL) and buffer Wash-EF (45 mL), each time allowing the column to empty by gravity flow. The DNA was

eluted from the column with buffer ELU-EF (15 mL). The DNA was precipitated with IPA (10.5 mL), mixed and allowed to sit (2 mins). The resulting suspension was loaded onto a NucleoBond Finalizer Filter with a syringe, by pushing the plunger slowly to pass the suspension through the filter with minimal force. The DNA was washed by passing endotoxin free 70% EtOH slowly through the filter. The DNA was dried by passing air forcefully through the filter at least six times, until any residual ethanol had been removed. The DNA was eluted by passing endotoxin free water (1 mL) dropwise through the filter, with the first eluate reloaded into the syringe and passed through the filter a second tie to increase yield. Air was passed through the filter to remove as much eluate as possible. DNA was analysed for purity and concentration on a nanodrop, with yields typically 1-2 mg of DNA per overnight culture.

6.12.1.2 Transfection Procedure

Transfections were typically performed in two batches (2 x 500 mL). The following protocol describes the quantities for a single 500 mL batch. Cells were centrifuged (600 x g, 10 min, rt) with a deceleration setting of 3 to avoid damaging the pellet on deceleration. Cells were resuspended into prewarmed FreeStyle™ 293 Expression Medium (500 mL per shaker bottle, 2×10^6 cells mL^{-1}, no Pen/strep) (Gibco). DNA (500 µg) was diluted into NaCl (150 mM, 25 mL) and mixed. 2.5 mL of PEI (1 mg mL^{-1} stock, linear, 25 kDa) (Polysciences) was diluted into NaCl (150 mM, 25 mL) and mixed. Subsequently, the diluted PEI was added to the diluted DNA, mixed, and incubated (rt, 10 mins). The subsequent suspension (50 mL) was added to the cell suspension, and then the cells incubated (31 °C, 90 rpm, 70% humidity, 8% CO_2) for 72 hrs before harvesting.

6.12.1.3 Purification Procedure

Typically, this procedure was performed with two batches (500 mL) of transfected cells. All steps were performed on ice or at 4 °C. Cells were harvested after 72 hr, by centrifuging (2,000 x g, 5 min, 4 °C, medium deceleration) (S). The cells were then either snap frozen at -80 °C or lysed immediately. The pellet was resuspended in resuspension buffer (50 mM HEPES, 150 mM NaCl, pH 7.25) with protease inhibitors (Roche). The cells were disrupted by sonication (17% amplitude, 2 x 90 s

in 1 s pulses, 60 s break, on ice). The lysate was aliquoted into two, into falcon tubes that had been prewashed with resuspension buffer and centrifuged (500 x g, 10 mins, 4 °C, medium deceleration). The supernatant was centrifuged (45, 000 rpm, 1 h, 4 °C) and supernatant discarded. The pellet was disrupted in extraction buffer (20 mL, 50 mM HEPES, 150 mM NaCl, 2% CHAPS). The suspension was stirred (1.5 h, 4°C), and then centrifuged (45,000 rpm, 1h, 4°C) (S). Sepharose bead (2 mL) (Roche) were equilibrated with extraction buffer (15 mL), centrifuged (300 g, 5 min, 4°C), before adding the disrupted pellet and incubating overnight (4°C).

The beads were loaded onto a prewashed polypropylene column. The beads were washed with 10 column volumes of wash buffer (50 mM HEPES, 150 mM NaCl, 1% CHAPS, 5% glycerol)

6.13 Computational Modelling and Screening

6.13.1 Generation of HI2 Analogues by ROCs and EON Analysis

3D ligand conformations were generated using the software OMEGA[190] (OMEGA 3.4. 0.4. OpenEye, Cadence Molecular Sciences, Santa Fe, NM). Ligand based screening was performed using the software ROCs[169] and EON.

Compounds were ranked by TanimotoCombo which quantifies both ShapeTanimoto (molecular shape overlay) and ColorTanimoto (chemical functionality overlaps). The top 1000 ranked compounds identified by ROCs were used as inputs for EON analysis. ET_Comb score, which is a combination of ShapeTanimoto and PB Electrostatic Tanimoto score were utilised to further rank compounds based on their electrostatic potential similarity with **HI2.** The top 100 ranked compounds were visually assessed, with a particular focus on scaffolds which altered quinoline core of **HI2**, as previous structure activity relationship testing had not altered this moiety.

6.14 Homology Modelling

6.14.1 TACOS

Sequences for residues 332-619 of the human IR-B and 331- 608 of the human IGF1R downloaded from the UniProt Databse[137, 145, 191]. These were submitted to the TACOS online structure prediction server (https://zhanglab.ccmb.med.umich.edu/TACOS/) with default settings.

6.14.2 AlphaFold

Sequences for residues 332-619 of the human IR-B and 331- 608 of the human IGF1R downloaded from the UniProt Databse[145, 146, 191]. These were submitted to the ColabFold[192] using AlphaFold2 multimer[167] using 20 recycles with tolerance = 0.05. pLDDT values and the PAE plot were downloaded from the results file UCSF ChimeraX[193] used to visualise the resulting model.

6.14.3 Homology Model Quality Assessment

Models were submitted to the SWISS-MODEL (https://swissmodel.expasy.org/) structure assessment server[138, 150, 194]. QMEANDisCO values were downloaded and visualised in UCSF ChimeraX[193].

6.14.4 Hotspot Prediction

Hotspot prediction using the resulting IR-IGF1R hybrid models was performed using KFC2 server (https://mitchell-web.ornl.gov/KFC_Server/index.php)[152]. Visualisation of the homology model and hotspot analysis was performed using UCSF ChimeraX[193].

6.15 Virtual High-throughput Screening

The graphical user interface Maestro[126] (Schrödinger Release 2019-3: Maestro, Schrödinger, LLC, New York, NY, 2023) was used to visualise the hybrid homology model. The hybrid homology model was prepared using the protein preparation wizard. 3D ligand conformations were prepared using Ligprep[160]. Ligand docking was performed using Glide[124, 125]. SiteMap[161] was used to identify sites suitable for ligand binding within proximity to hotspot residues, and 10x10x10 Å grids generated directly from the chosen site points. Screening was performed on ARC4, part of the High-Performance Computing Facility at the University of Leeds, UK.

6.16 Statistics

Data were analysed using GraphPad Prism software (version 9)[195]. P values > 0.05 were deemed statistically significant. Results are expressed as mean ±SEM unless otherwise stated.

Comparisons between two groups were performed using an unpaired Student's t-test. Comparisons between the mean values of multiple groups were performed using one-way analysis of variance (ANOVA), followed by Tukey's multiple comparisons test. Non-linear regression curves were calculated using using a single-site fitting model, accounting for non-specific binding if appropriate.

6.16.1 Synthetic Chemistry

All solvents and reagents were obtained from commercial suppliers and used without further purification. Thin layer chromatography was performed on aluminium backed silica gel and visualised using an ultraviolet lamp. Normal-phase flash-column chromatography was performed using silica gel 60 (40-63 μm particles). Automated normal-phase flash-column chromatography was performed on a Biotage® Isolera™ One machine using Biotage® Sfär columns. Automated reverse-phase flash-column chromatography was performed using C18 silica columns. ^{1}H and ^{13}C NMR data were collected on a Bruker Avance III 500. All shifts were recorded against an internal standard of tetramethyl silane. Solvents used for NMR (chloroform-d, methanol-d4 and DMSO-d6) were obtained from Sigma-Aldrich. 1H NMR data is reported in the following format: ppm (splitting pattern, coupling constant (Hz), number of protons, proton assignment). Signal assignments were deduced with the aid of TopSpin. LCMS data were recorded on a Donex Ultimate 3000 LC system with a MeCN/H_2O +0.1% formic acid gradient. HRMS data were recorded using a Bruker MaXis impact spectrometer using electron spray ionisation. Infrared spectra were recorded on a Perkin-Elmer one FTIR spectrometer.

6.16.1.1 1-[(3-Methylphenyl)methyl]-3-nitro-1H-1,2,4-triazole (3)
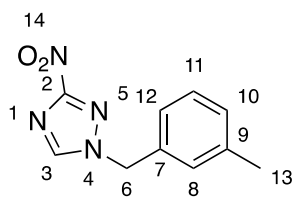
Anhydrous ethanol (19 ml) was added to sodium metal (456 mg, 19.8 mmol, 1.1 eq) under inert atmosphere. The resulting mixture was stirred at room temperature until all solids had dissolved. 3-nitro-1H-1,2,4-triazole (2.0 g, 17.5 mmol. 1.0 eq) was added against a counter flow of N_2, and the reaction allowed to stir until the solids had dissolved to form a yellow solution. 1-(chloromethyl)-3-methylbenzene (2.9 mL, 21.9 mmol, 1.3 eq.) was added, and the reaction heated to 80 °C for 18 h. The reaction was cooled to room temperature, quenched with water (2 mL) and concentrated. The residue was partitioned between water (20 mL), and EtOAc (3 x

20 mL). Organics were combined and washed with brine (20 mL), dried (Na_2SO_4), filtered and concentrated to yield a yellow oil. This was purified by column chromatography (20-50% EtOAc in Pet. Ether) and relevant fractions concentrated to yield <u>the title compound</u> as a colourless oil. (3.25 g, 14.9 mmol, 99%). **¹H NMR** (500 MHz, $CDCl_3$) δ = 8.06 (s, 1H, H_3), 7.31 (t, J = 7.92 Hz, 1H, H_{11}), 7.23 (d, J = 7.92, 1H, H_{12}), 7.15 (s, 1H, H_8), 7.07 (d, J = 7.92, H_{10}, 1H), 5.38 (s, 2H, H_6), 2.36 (s, 3H, H_{13}); **¹³C NMR** (125 MHz, $CDCl_3$) δ = 162.8 (C_2), 144.5 (C_3), 139.5 (C_9), 132.1 (C_7), 130.4 (C_{10}), 129.4 (C_8), 129.3 (C_{12}), 125.7 (C_{11}), 55.6 (C_6), 21.3 (C_{13}); **IR** v_{max} / cm⁻¹ (solid) 3137 (C-H, aryl), 2984 (C-H, alkyl), 2940 (C-H, alkyl), 1634 (C-C, aryl), 1584 (N-O), 1529 (C-C, aryl); **HRMS** (ESI-TOF) *m/z* [M+Na]⁺ calcd for $C_{10}H_{10}N_4O_2Na$ 241.0701, found 241.0686; **LCMS** RT = 0.6 min, 217.6 [M+H]⁺.

6.16.1.2 1-[(3-Methylphenyl)methyl]-1H-1,2,4-triazol-3-amine (*4*)
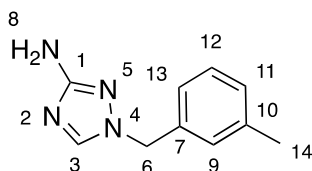
1-[(3-methylphenyl)methyl]-3-nitro-1H-1,2,4-triazole (3.01 g, 13.8 mmol, 1.0 eq.) was dissolved in MeOH (80 mL), and added to Pd/C (0.3g, 0.1 eq. w/w) under a N_2 atmosphere. The reaction was placed under a H_2 atmosphere, and stirred at room temperature for 18 h. The reaction was removed from the H_2 atmosphere, filtered through celite, and filtrate concentrated yield <u>the title compound</u> as a colourless solid (2.45 g, 13.0 mmol, 95%). Solids were used for subsequent reactions without further purification. **¹H NMR** (500 MHz, $CDCl_3$) δ = 7.67 (s,1H, H_3), 7.25 (t, J = 7.96 Hz, 1H, H_{12}), 7.14 (d, J = 7.96, 1H, H_{13}), 7.06 (s, 1H, H_9), 7.05 (d, 1H, H_{11}) 5.07 (s, 2H, H_6), 3.98 (bs, 2H, H_8), 2.33 (s, 3H, H_{14}); **¹³C NMR** (125 MHz, $CDCl_3$) δ = 163.7 (C_1), 142.3 (C_3), 138.8 (C_{10}), 134.7 (C_7), 129.3 (C_{11}), 128.9 (C_9), 128.7 (C_{13}), 125.0 (C_{12}), 53.3 (C_6), 21.3 (C_{14}); **IR** v_{max} / cm⁻¹ (solid) 3318, 3188, 3125, 3017, 2919, 1640, 1548;

HRMS (ESI-TOF) *m/z* [M+H]$^+$ calcd for $C_{10}H_{10}N_4O_2$ 188.1803, found 188.1805; **LCMS** RT = 0.5 min, *m/z* = 188.99 [M+H]$^+$.

6.16.1.3 *2-Bromo-N-{1-[(3-methylphenyl)methyl]-1H-1,2,4-triazol-3-yl}pyridine-4-carboxamide (9)*
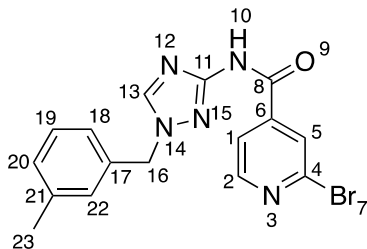
1-[(3-Methylphenyl)methyl]-1H-1,2,4-triazol-3-amine (168 mg, 0.9 mmol, 0.9 eq.) and 2-bromopyridine-4-carboxylic acid (200 mg, 1.0 mmol, 1.0 eq.) were added to EtOAc (5 mL), followed by T$_3$P (50% in EtOAc, 1.2 mL, 2.0 mmol, 2.0 eq.) and NEt$_3$ (0.38 mL, 3.0 mmol, 3.0 eq.). The reaction was refluxed under N$_2$ atmosphere for 18 h. The reaction was quenched with water (5 mL), and partitioned between water (30 mL) and DCM (3 x 10 mL). The organics were combined, washed with sat. aqueous NH$_4$Cl (20 mL), dried (Na$_2$SO$_4$), filtered and concentrated to yield a brown crystalline solid. This was further purified by column chromatography (4% MeOH in DCM). Relevant fractions were concentrated to yield <u>the title compound</u> as a colourless solid (219 mg, 0.58 mmol, 59%). **^{1}H NMR** (500 MHz, CDCl$_3$) δ = 10.09 (s, 1H, H$_{10}$), 8.45 (d, J = 4.97 Hz, 1H, H$_2$), 7.92 (s, 1H, H$_5$), 7.72 (d, J = 4.97 Hz, 1H, H$_1$), 7.63 (s, 1H, H$_{13}$), 7.27 (t, J = 7.36 Hz, 1H, H$_{19}$), 7.18 (d, J = 7.36, 1H, H$_{20}$), 7.11 (s, 1H, H$_{22}$), 7.09 (d, J= 7.36 Hz, 1H, H$_{18}$), 5.28 (s, 2H, H$_{18}$), 2.34 (s, 3H, H$_{18}$); **^{13}C NMR** (125 MHz, CDCl$_3$) δ = 161.9, (C$_8$) 155.9 (C$_{11}$), 151.1 (C$_2$) , 144.0 (C$_{13}$), 142.6 (C$_4$), 141.9 (C$_{21}$), 139.2 (C$_{17}$), 133.2 (C$_6$), 129.9 (C$_{20}$), 129.4 (C$_{22}$), 129.2 (C$_{18}$), 125.9 (C$_{19}$), 125.7 (C$_5$), 121.2 (C$_1$), 54.5 (C$_{16}$), 21.4 (C$_{23}$); **IR** v$_{max}$ / cm^{-1} (solid) 3241, 3118, 3093, 1683,

1597, 1571, 1528; **HRMS** (ESI-TOF) *m/z* [M+H]$^+$ calcd for $C_{16}H_{14}{}^{79}BrN_5O$ 372.0454, found 372.0450; **LCMS** RT = 0.5 min, *m/z* = 372.26 [M+H]$^+$

6.16.1.4 *2-(4-Methylphenyl)-N-{1-[(3-methylphenyl)methyl]-1H-1,2,4-triazol-3-yl}pyridine-4-carboxamide (10)*
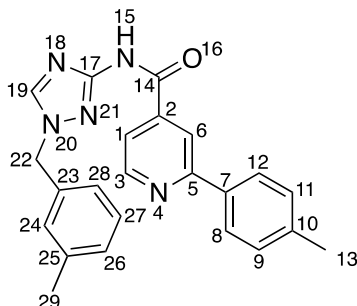
2-bromo-N-{1-[(3-methylphenyl)methyl]-1H-1,2,4-triazol-3-yl}pyridine-4-carboxamide (50 mg, 0.13 mmol, 1.0 eq.) and 4-methylbenzeneboronic acid (22 mg, 0.16 mmol, 1.2 eq.) were dissolved in dioxane (2.3 mL). The solution was purged with N$_2$, PdCl$_2$(PPh$_3$)$_2$ (10 mg, 0.01 mmol, 0.1 eq.) and CsCO$_3$ added, and the reaction purged with further N$_2$. The reaction stirred at 100 °C for 18 h. The reaction was cooled to room temperature, filtered through celite, and filtrate concentrated. The residue was dissolved in DCM (20 mL), and washed with water (20 mL), sat. aqueous NH$_4$Cl (20 mL), dried (Na$_2$SO$_4$) and concentrated to yield a brown solid. This was further purified by column chromatography (5% MeOH in DCM) and relevant fractions concentrated to yield the title compound as a colourless solid (24 mg, 62.6 μmol, 51%) **^{1}H NMR** (500 MHz, CDCl$_3$) δ = 9.50 (s, 1H, H$_{15}$), 8.83 (d, J = 5.09 Hz, 1H, H$_3$), 8.18 (s, 1H, H$_6$), 7.95 (d, J = 7.50, 1H, H$_8$/H$_{12}$), 7.73 (s, 1H, H$_{19}$), 7.64 (d, J = 5.09 Hz, 1H, H$_1$) 7.33 (d, 2H, H$_9$/H$_{11}$) 7.28 (t, H = 7.95 Hz, 1H, H$_{27}$) 7.19 (d, J = 7.95 Hz, 1H, H$_{26}$), 7.10 (s, 1H, H$_{24}$), 7.09 (d, J = 7.95 Hz, 1H, H$_{28}$), 5.26 (s, 2H, H$_{22}$), 2.45 (s, 3H, H$_{13}$), 2.37 (s, 3H, H$_{29}$) **^{13}C NMR** (125 MHz, CDCl$_3$) δ 163.3 (C$_{14}$), 158.6 (C$_5$), 156.0 (C$_{17}$), 150.5 (C$_3$), 142.1 (C$_{19}$), 141.8 (C$_{10}$), 139.9 (C$_{25}$), 139.0 (C$_7$), 135.5 (C$_2$), 133.5 (C$_{26}$) 129.7 (C$_{24}$), 129.7 (C$_8$/C$_{12}$), 129.2 (C$_9$/C$_{11}$), 129.1 (C$_{28}$), 126.9 (C$_{27}$), 125.5 (C$_6$), 119.1 (C$_1$), 117.8 (C$_{23}$), 54.3 (C$_{22}$), 21.3 (C$_{29}$), 21.3 (C$_{13}$); **IR** v$_{max}$ / cm^{-1} (solid) 3232.3, 3029.4, 2917.9, 1678.8, 1572.4, 15497, 1527.9; **HRMS**

(ESI-TOF) *m/z* [M+H]⁺ calcd for $C_{23}H_{22}N_5O$ 384.1817, found 384.1818; **LCMS** RT = 0.7 min, *m/z* = 384.35 [M+H]⁺
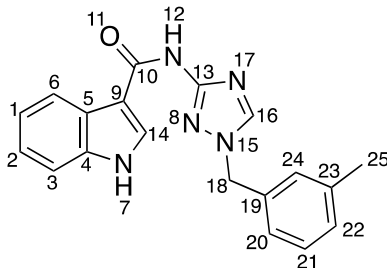
3-Indole-carboxylic acid (100 mg, 0.621 mmol, 1.0 eq.) was cooled to 0 °C, and $SOCl_2$ (0.5 mL) added dropwise. The reaction was stirred and allowed to warm to room temperature, and then refluxed for 1 hr, after which no solids were present. The reaction was concentrated to yield a pink waxy solid, to which was added dropwise a solution of 1-[(3-methylphenyl)methyl]-1H-1,2,4-triazol-3-amine (128 mg, 0.681 mmol, 1.1 eq.) and NEt_3 (0.24 mL, 1.86 mmol, 3.0 eq.) in DCM (5 mL). The reaction was allowed to stir at room temperature for 18 h, after which a white precipitate had formed. The solids were filtered, triturated in diethyl ether, and dried under vacuum to yield <u>the title compound</u> as a colourless solid (152 mg, 0.459 mmol, 74%). **¹H NMR** (500 MHz, DMSO) δ = 11.70 (s, 1H, H_7) 10.19 (s, 1H, H_{12}), 8.56 (s, 1H, H_{16}), 8.30 (d, J = 2.95 Hz, 1H, H_6), 8.16, (d, J = 7.66 Hz, 1H, H_3), 7.45 (d, J = 7.55, 1H, H_{21}), 7.27 (t, J = 7.55, 1H, H_{20}), 7.18 (dt, J_1 = 7.96 Hz, J_2 = 1.46 Hz, 1H, H_2) 7.15 (s, 1H, H_{14}), 7.14 (dt, J_1 = 7.96 Hz, J_2 = 1.46 Hz, 1H, H_1) 7.14 (s, 1H, H_{24}), 7.13 (d, J = 7.55, 1H, H_{20}) 5.32 (s, 2H, H_{18}) 2.31 (s, 3H, H_{25}) **¹³C NMR** (125 MHz, DMSO-D_6) δ = 163.0 (C_{10}), 157.2 (C_{13}), 144.1 (C_{23}), 138.3 (C_{23}), 136.7 (C_{19}), 136.6 (C_4), 129.6 (C_{22}), 129.0 (C_9), 128.9 (C_{24}), 127.0 (C_{20}), 125.5 (C_{21}),122.6 (C_5), 121.5 (C_{14}), 121.2 (C_2), 116.0 (C_1) 112.4 (C_6), 110.1 (C_3), 52.7 (C_{18}), 21.4 (C_{25}); **IR** v_{max} / cm⁻¹ (solid) 3223.8,

1665.2, 1555.8, 1519.7; **HRMS** (ESI-TOF) *m/z* [M+H]$^+$ calcd for $C_{19}H_{17}N_5O$ 332.1506, found 332.1508; **LCMS** RT = 0.6 min, m/z = 332.34 [M+H]$^+$

6.16.1.6 1-(4-Methylphenyl)-N-{1-[(3-methylphenyl)methyl]-1H-1,2,4-triazol-3-yl}-1H-indole-3-carboxamide (7)
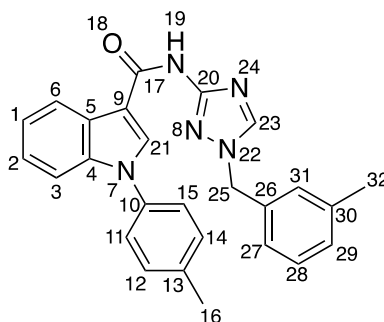
N-{1-[(3-methylphenyl)methyl]-1H-1,2,4-triazol-3-yl}-1H-indole-3-carboxamide (75 mg, 0.23 mmol, 1.0 eq.), 4-bromo-toluene (37 mg, 0.22 mmol, 1.2 eq.), copper iodide (8 mg, 0.04 mmol, 0.2 eq.) and K_3PO_4 (75 mg, 0.36 mmol, 2.1 eq.) were placed under N_2 atmosphere. Anhydrous DMF (2.5 mL) as added and the reaction degassed for 10 minutes. To this was added a solution of N,N'-dimethylethylenediamine (8 µL, 0.07 mmol, 0.4 eq.) in anhydrous DMF (0.5 mL), upon which the reaction instantaneously went from colourless to black. The reaction was heated to 110 °C for 18 h. After this, the reaction was judged to have not gone to completion, so was cooled to room temperature, and further 4-bromo-toluene (40 mg, 0.23 mmol, 1.3 eq.), copper iodide (20 mg, 0.11 mmol, 0.6 eq.) K_3PO_4 (35 mg, 0.17 mmol, 0.9 eq.) and N,N'-dimethylethylenediamine (20 µL, 0.18 mmol, 1.0 eq.) was added under N_2. The reaction was heated to 110 °C for 18 h. The reaction was cooled to room temperature, filtered through silica plug (eluted with 10% MeOH in DCM), and filtrate concentrated to give a pale green solid. This was purified by column chromatography (3% MeOH in DCM), and relevant fractions concentrated. The residue was dissolved in DCM (20 mL), washed with 10% Na_2SO_4 (2 x 10 mL), dried (Na_2SO_4) and concentrated to yield <u>the title compound</u> as a pale yellow solid (16 mg, 38.0 µmol, 17%). **^{1}H NMR** (500 MHz, MeOH-D$_4$) 8.31 (bs, 1H, H$_{19}$) 8.21 (s, 1H, H$_{23}$), 8.13 (s, 1H, H$_{21}$) 7.36 (d, J = 8.18 Hz, 1H, H$_6$), 7.33 (d, J = 8.18 Hz, 2H, H$_{11}$/H$_{15}$), 7.28 (d, J= 8.18 Hz, 2H, H$_{12}$/H$_{14}$) 7.17 (t, J = 8.18, 1H), 7.15 (d, J= 7.14 Hz, 1H, H$_3$) 7.14 (d, J = 7.69 Hz, 1H, H$_{27}$), 7.12 (t, J = 7.69 Hz, 1H, H$_{28}$), 7.08 (s, 1H, H$_{31}$), 7.03 (t, J = 8.18, 1H, H$_1$), 7.02 (d, J = 7.06, 1H, H$_{20}$) 5.22 (s, 2H, H$_{25}$), 2.33 (s,

3H, H$_{32}$), 2.21 (s, 3H, H$_{16}$); **^{13}C NMR** (125 MHz, MeOH-D$_4$) δ = 164.0 (C$_{17}$), 158.2 (C$_{20}$) 144.1 (C$_{23}$) 138.4 (C$_{10}$), 138.3 (C$_{30}$) 137.8 (C$_{26}$), 136.7 (C$_4$), 136.0 (C$_{13}$), 135.3 (C$_{11}$/C$_{15}$), 130.0 (C$_{12}$/C$_{14}$), 128.6 (C$_{29}$), 128.5(C$_{28}$), 128.4 (C$_9$), 127.4 (C$_5$) 124.9 (C$_{27}$), 124.3 (C$_{21}$), 123.3 (C$_2$), 121.9 (C$_1$), 121.5 (C$_6$), 110.6 (C$_3$), 53.1 (C$_{25}$) 20.0 (C$_{32}$), 19.7 (C$_{16}$); **IR** ν$_{max}$ / cm^{-1} (solid) 2990 (C-H, aryl) 2943 (C-H, alkyl), 1680 (C=O, amide), 1614 (C-C, aryl), 1547 (C-C, aryl); **HRMS** (ESI-TOF) *m/z* [M+H]$^+$ calcd for C$_{26}$H$_{24}$N$_5$O 422.1979, found 422.1975. **LCMS** RT = 0.8 min, *m/z* = 422.40 [M+H]$^+$

7   Materials

7.1   Generic Lab Reagents

Reagent	Cat #	Manufacturer
Tris HCl	T/P631/53	Fisher Chemical™
Tris Base	BP152-1	Fisher Bioreagents™
Ethanol	24194	HoneyWell™
Methanol	M/4056/17	Fisher Chemical™
NaCl	S7653	Sigma Aldrich®
$MgSO_4$	M/1000/60	Fisher Chemical™
Tween 20	P1379	Sigma Aldrich®
SDS	05030	Sigma Aldrich®
Skim Milk Powder	70166	Millipore®
Agarose	BIO-41025	Bioline
EDTA	10526383	Fisher Chemical™
Glycine	BP381-1	Fisher Bioreagents™
Sodium Acetate	S2889	Sigma Aldrich®

7.2 Cell Culture Reagents

Reagent	Cat #	Manufacturer
DMEM, high glucose, GlutaMAX™ Supplement	61965026	Gibco™
Fetal Bovine Serum	S-001A-BR	Life Science Production
Antibiotic, Antimycotic Solution (100X), Stabilised	A5955	Sigma Aldrich®
Dulbecco's Phosphate Buffered Saline	D8537	Sigma Aldrich®
Trypsin (0.25%), phenol red	15050065	Gibco™
DMSO	D5879	HoneyWell™
Endothelial Cell Growth Medium 2	C22011	PromoCell®
Growth Medium 2 SupplementMix	C39216	PromoCell®
FreeStyle™ 293 Expression Medium	12338018	Gibco™
Opti-MEM™ I Reduced Serum Medium	31985062	Gibco™
Polyethylenimine, Linear, (MW 25,000), Transfection Grade	23966	Polysciences™
IGF1 Human	A63373	Antibodies.com®
Insulin Human	A63248	Antibodies.com®

7.3   BRET assay Reagents

Reagent	Cat #	Manufacturer
Lipofectamine 2000	11668019	Invitrogen™
Coelenterazine-h	S2011	Promega™

7.4   LIVE-DEAD assay

Reagent	Cat #	Manufacturer
Poly-L-Lysine	P4707	Sigma Aldrich®
LIVE/DEAD™ Viability/Cytotoxicity Kit, for mammalian cells	L3224	Invitrogen™

7.5   Western blotting

Reagent	Cat #	Manufacturer
Cell Lysis Buffer	FNN0011	Invitrogen™
Pierce™ BCA Protein Assay Kit	23225	Thermo Scientific™
Protein G Agarose	11243233001	Roche
Protease Cocktail Inhibitor	P8340	Sigma Aldrich®
LDS Sample Buffer	11549166	Invitrogen™
NuPAGE™ sample reducing buffer	11569166	Invitrogen™
NuPAGE™ 4-12%, Bis-Tris, 1.0-1.5 mm, Mini Protein Gels	NP0322BOX	Invitrogen™
Running Buffer	NP0002	Invitrogen™
Transfer Buffer	1704271	Bio-Rad®
Immobilon Western Chemiluminescent HRP Substrate	WBKLS0500	Millipore®
0.2 µM Nitrocellulose	1704159	Bio-Rad®

7.6 QPCR

Reagent	Cat #	Manufacturer
Monarch® Total RNA Miniprep Kit	T2010S	New England BioLabs®
LunaScript® RT SuperMix Kit	E3010	New England BioLabs®
iTaq Universal SYBR Green Supermix	1725120	Bio-Rad®
IGF1R - PrimePCR™ SYBR® Green Assay: IGF1R, Human	10025636	Bio-Rad®
INSR - PrimePCR™ SYBR® Green Assay: INSR, Human	10025636	Bio-Rad®
ACTB - PrimePCR™ SYBR® Green Assay: ACTB, Human	10025636	Bio-Rad®

7.7 Cloning

Reagent	Cat #	Manufacturer
Q5® High Fidelity DNA Polymerase	M0491S	New England BioLabs®
Nuclease-free Water	B1500S	New England BioLabs®
Deoxynucleotide (dNTP) Solution Mix	N0447S	New England BioLabs®
FastDigest DpnI	FD1703	Thermo Scientific™
FastDigest BshTI (AgeI)	FD1464	Thermo Scientific™
FastDigest KpnI	FD0524	Thermo Scientific™
FastDigest NcoI	FD0574	Thermo Scientific™
FastDigest MssI (PmeI)	FD1344	Thermo Scientific™
FastDigest NheI	FD0974	Thermo Scientific™
PureLink™ Gel Extraction Kit	K210012	Thermo Scientific™
QIAquick PCR Purification Kit	28104	Qiagen
FastDigest Green Buffer	B72	Thermo Scientific™
NEBuilder HiFi DNA Assembly Cloning Kit	E5520S	New England BioLabs®
NEB 5-alpha Competent *E.Coli* (High Efficiency)	C2987H	New England BioLabs®
SOC Outgrowth Medium	B9020	New England BioLabs®

LB Agar	BP1425-500	Fisher Bioreagents™
LB Broth	BP1426-500	Fisher Bioreagents™
Ampicillin	A9518-5G	Sigma Aldrich®
X-Gal	R0404	Thermo Scientific
IPTG	15529019	Invitrogen
OneTaq® Quick-Load® 2X MasterMix	M0486S	New England BioLabs®
QIAprep Spin Minprep Kit	27104	Qiagen
QIAquick PCR Purification Kit	29104	Qiagen

7.8   Small Molecules

Reagent	Cat #	Manufacturer
Tunicamycin	BIT1006	Apollo Scientific
Picropodophyllin	S7668-SEL	Stratech
Rutecarpine	S2349-SEL	Stratech
Demethylasterriquinone B1	CAY21026	Cambridge Bioscience
Propranolol	B1346-APE	Stratech
Ketoconazole	M02048	Flourochem

7.9   Radiochemicals

Reagent	Cat #	Manufacturer
^{125}I-Insulin, human recombinant	NEX4200010UC	Apollo Scientific
^{125}I-IGFI, human recombinant	NEX241005UC	Stratech

7.10   Antibodies
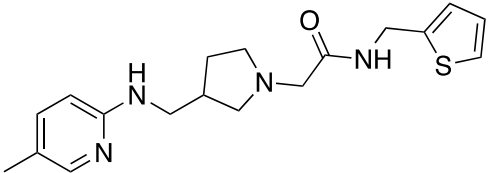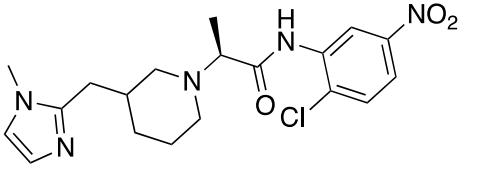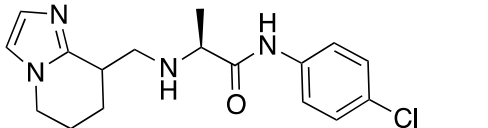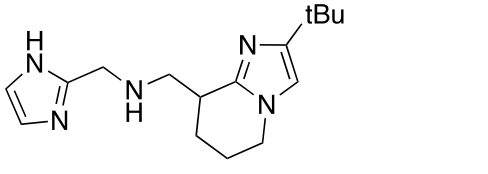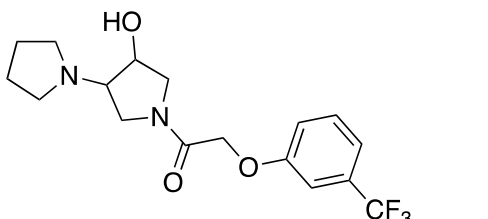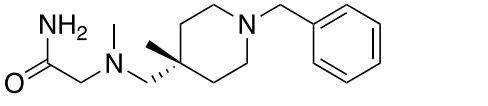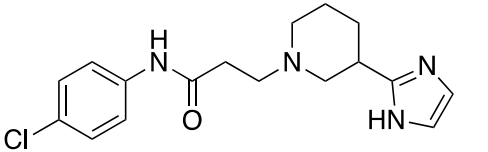
Antibody	Dilution (IP)	Dilution (WB)	Company	Source	Cat #
Insulin Receptor β (48B)	N/A	1:1000	Cell Signalling	Rabbit	3025
IGF1R β (D23H3)	1:200	1:1000	Cell Signalling	Rabbit	9750
β-Actin (C4)	N/A	1:5000	Santa Cruz Biotechnology	Mouse	Sc-376421
pIGF1R (DA7A8)	N/A	1:1000	Cell Signalling	Rabbit	3918
6 x His	N/A	1:1000	Cell Signalling	Rabbit	2365
ECL Anti-mouse IgG, Horseradish Peroxidase	N/A	1:4000	Cytiva	Sheep	NA931V
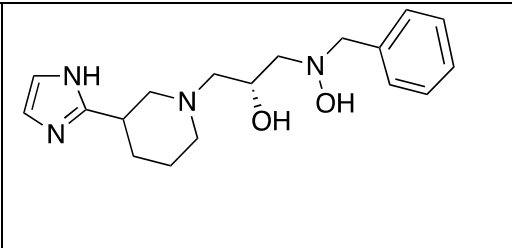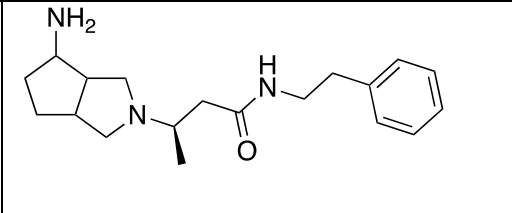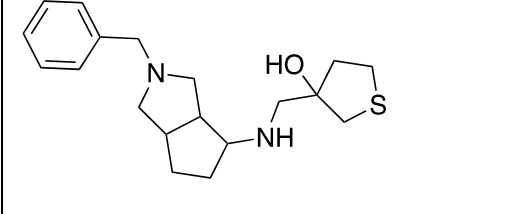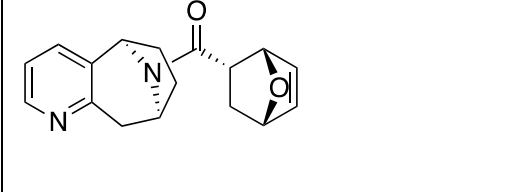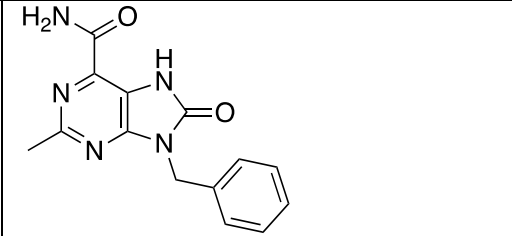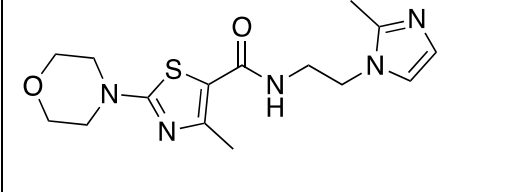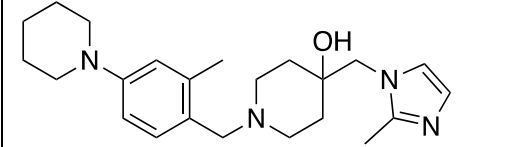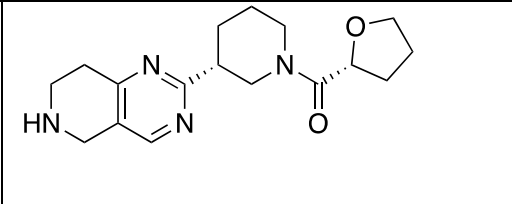
ECL Antib-Rabbit IgG, Horseradish Peroxidase	N/A	1:4000	Cytiva	Donkey	NA934V
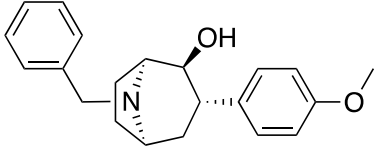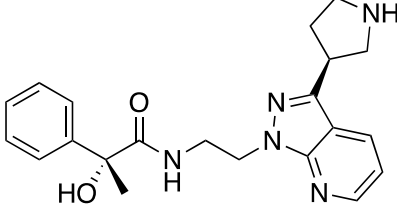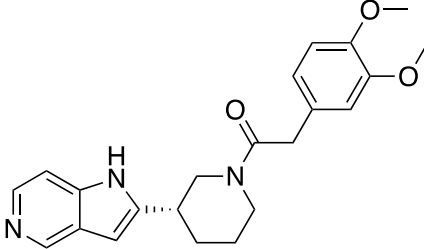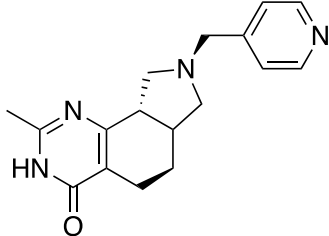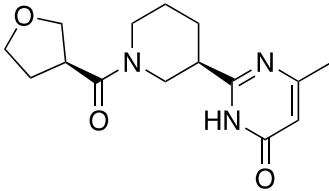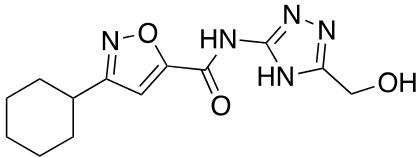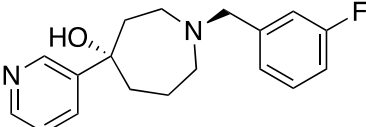
7.11   Plasmids
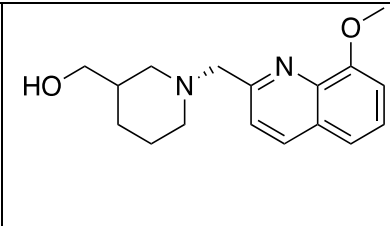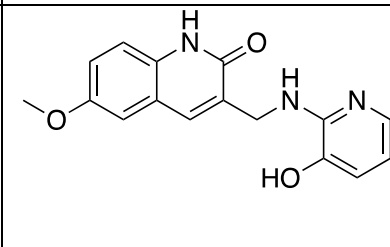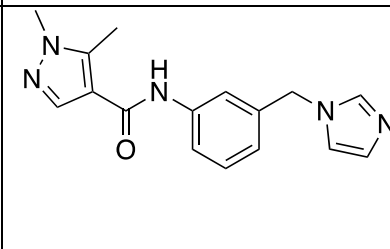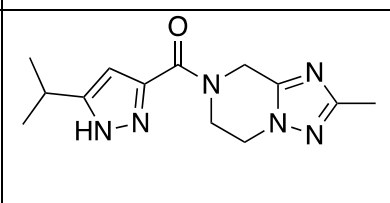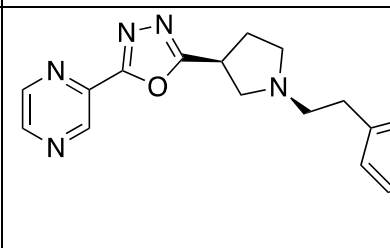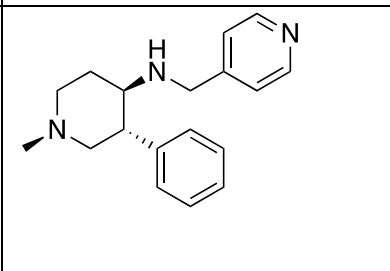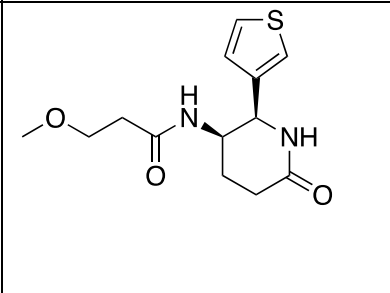
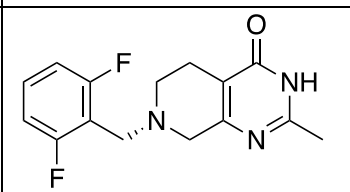| Vector	Addgene ID	Origin
IR-Rluc	N/A	Gift from Prof Tarik Issad
IR-YPET	N/A	Gift from Prof Tarik Issad
IGF1R-Rluc	N/A	Gift from Prof Tarik Issad
IGF1R-YPET	N/A	Gift from Prof Tarik Issad
pUC19	50005	New England BioLabs®
pOPINGS	41121	Gift from Prof. Ray Owens (Addgene)
pOPINE-3C-HALO7	41126	Gift from Prof. Ray Owens (Addgene)
pHL-FcHis	99846	Gift from Prof Edith Yvonne Jones (Addgene)
pHL-avitag3	99847	Gift from Prof Edith Yvonne Jones (Addgene)
IR-TAP	N/A	Gift from Prof. Ünal Coskun
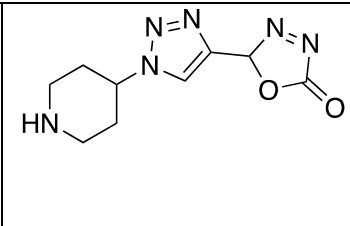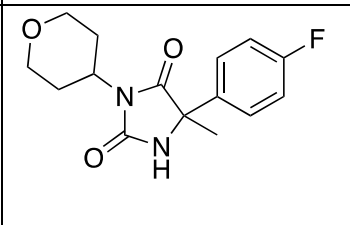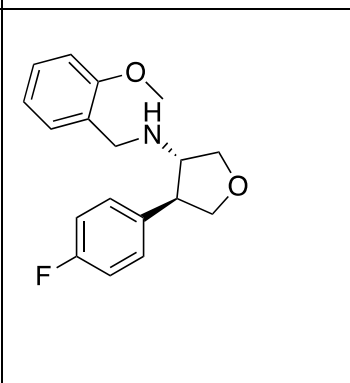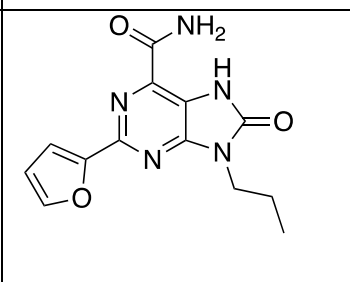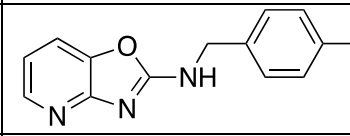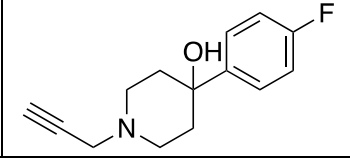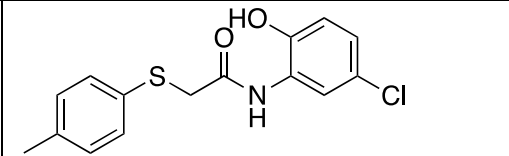
8 Appendices

Compound Number	Manufacturer ID	Vendor	Structure
15	Z990547890	Enamine	
16	Z1456216568	Enamine	
17	Z1268843608	Enamine	
18	Z1268664762	Enamine	
19	Z1130236778	Enamine	
20	Z1524881235	Enamine	
21	Z1317921254	Enamine	

22	Z1491395185	Enamine	
23	Z1762056851	Enamine	
24	Z1630799765	Enamine	
25	ART 17483471	Asinex	
26	ART 10298682	Asinex	
27	SYN 15638193	Asinex	
28	AEM 17368488	Asinex	
29	SFA 21715299	Asinex	

30	AEM 09584122	Asinex	
31	SFA 22636156	Asinex	
32	AOP 22322900	Asinex	
33	ART 13073366	Asinex	
34	ART 15348314	Asinex	
35	ART 18147998	Asinex	
36	LEG 13979153	Asinex	

37	LEG 13983547	Asinex	
38	LEG 07786328	Asinex	
39	ART 20296990	Asinex	
40	ART 13818717	Asinex	
41	LMG 20990341	Asinex	
42	ART 08966646	Asinex	
43	ART 14272385	Asinex	

44	ART 13776553	Asinex	
45	ART 20660267	Asinex	
46	SYN 19066813	Asinex	
47	ART 10298936	Asinex	
48	ART 12453331	Asinex	
49	ART 14132787	Asinex	
50	ART 09065762	Asinex	
51	7971285	ChemBridge	
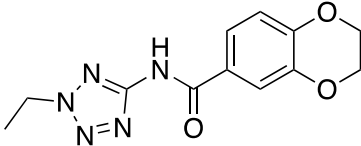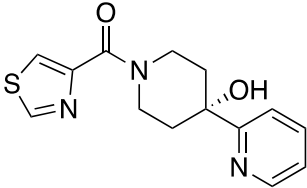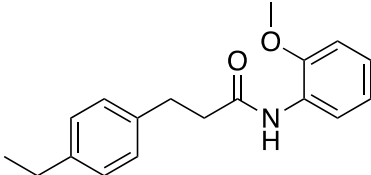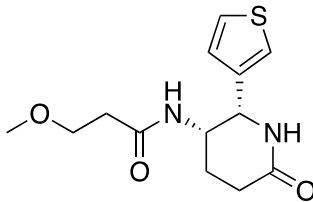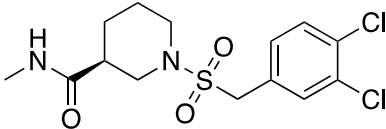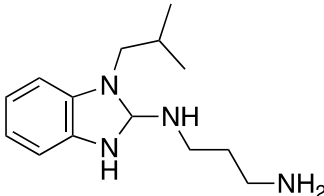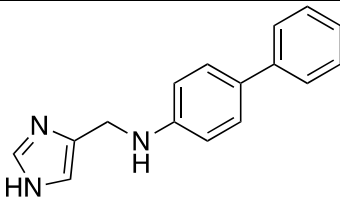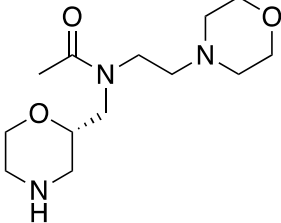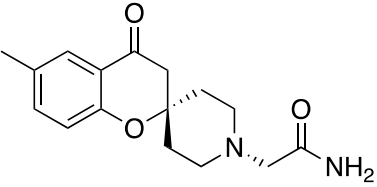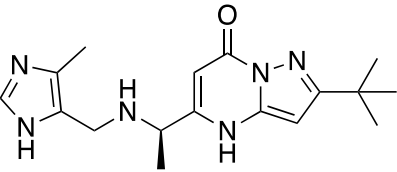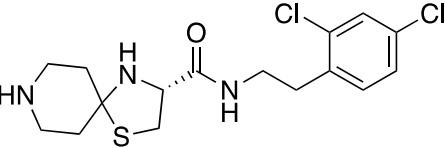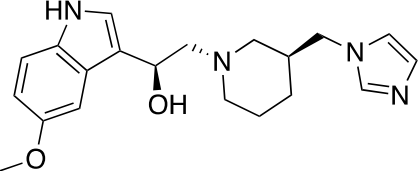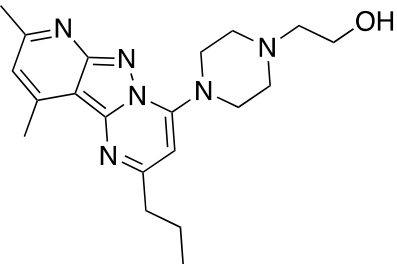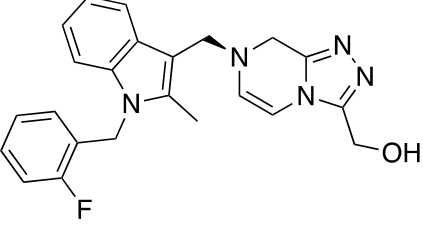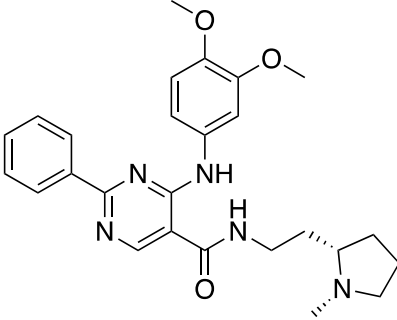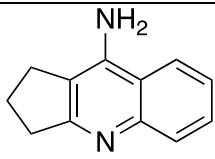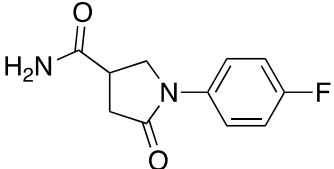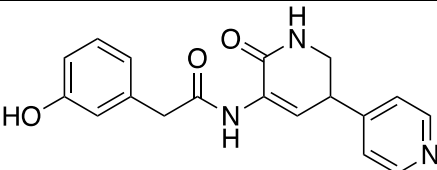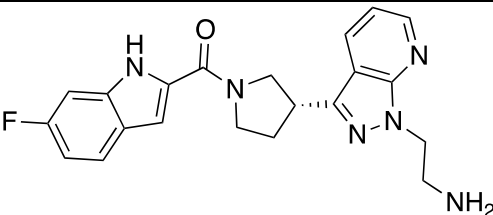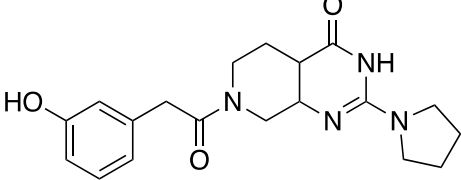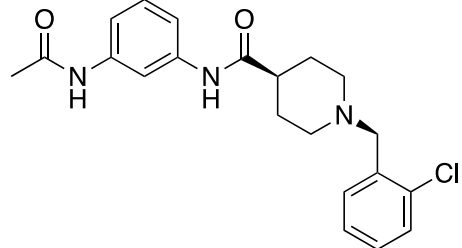
52	9111872	ChemBridge	
53	ART 11336028	Asinex	
54	9111779	ChemBridge	
55	ART 14272385	Asinex	
56	AAM-19379433	Asinex	
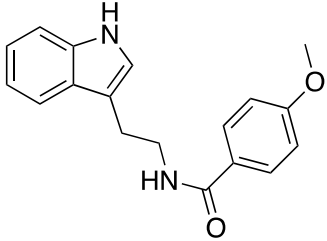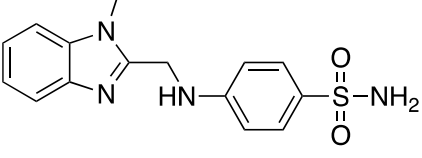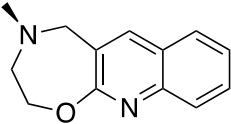
*Table 3 – Chemical structures of compounds **15-56** tested in the BRET assay in section 3.4 (TACOS model).*
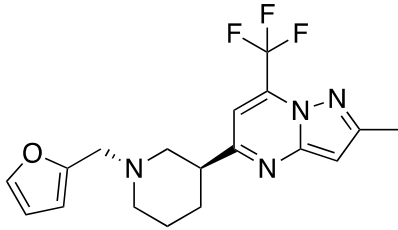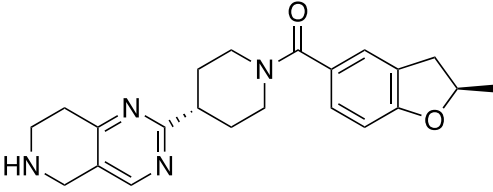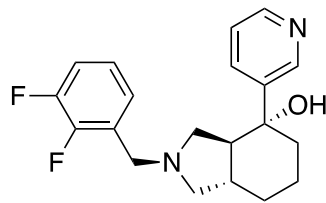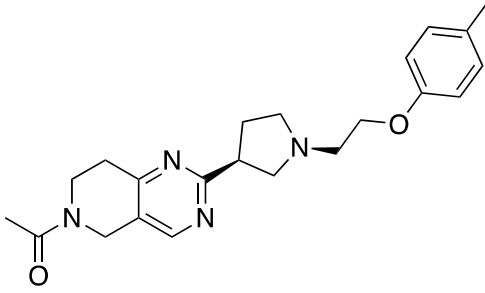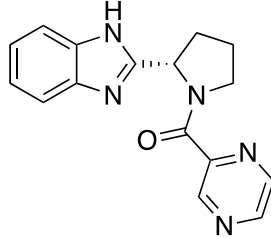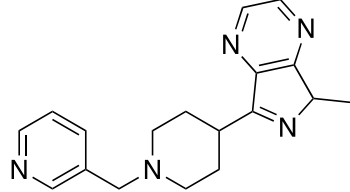
Compound Number	Manufacturer ID	Vendor	Structure
64	ALB H11275385	AMRI	
65	ALB H03206547	AMRI	
66	ALB H05713990	AMRI	
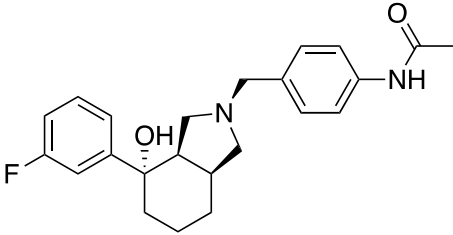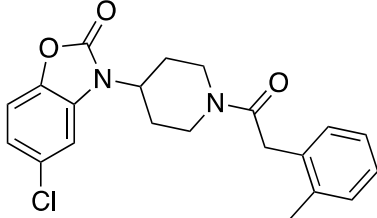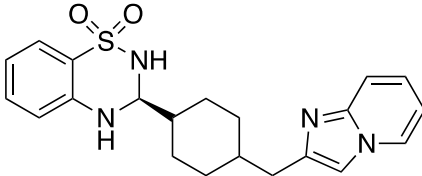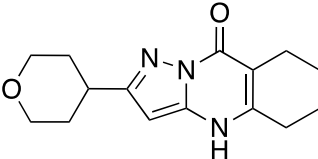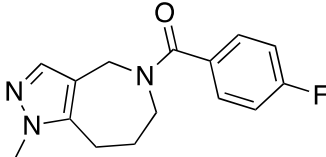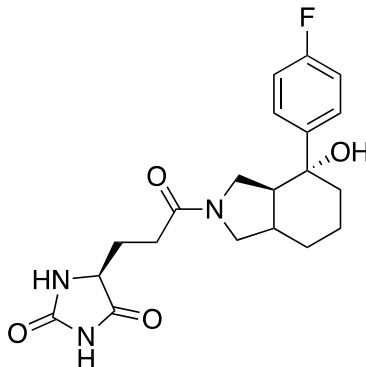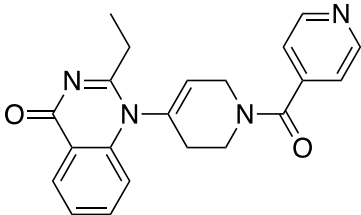
67	ART 10022182	Asinex	
68	AOP 19578164	Asinex	
69	ALB H02134954	AMRI	
70	LEG 19977799	Asinex	
71	9152831	ChemBridge	
72	SYN 17235387	Asinex	

73	ALB H11779526	AMRI	
74	5229962	ChemBridge	
75	9146559	ChemBridge	
76	8017-2926	ChemDiv	
77	SFA 22638034	Asinex	
78	AEM 11795328	Asinex	
79	9112043	ChemBridge	

80	5106764	ChemBridge	
81	5650922	ChemBridge	
82	ART 07439157	Asinex	
83	ART 12360474	Asinex	
84	AEM 11789246	Asinex	
85	AEM 12844846	Asinex	
86	ADM 11112569	Asinex	

87	AEM 14859046	Asinex	
88	SFA 21724159	Asinex	
89	AOP 14670416	Asinex	
90	SFA 21721286	Asinex	
91	ART 11364220	Asinex	
92	LEG 16470729	Asinex	

93	AEM 14663843	Asinex	
94	AEM 12674129	Asinex	
95	AEM 13697685	Asinex	
96	ART 14132762	Asinex	
97	ART 15417156	Asinex	
98	AEM 14658672	Asinex	
99	AEM 12941457	Asinex	
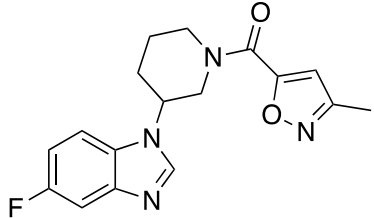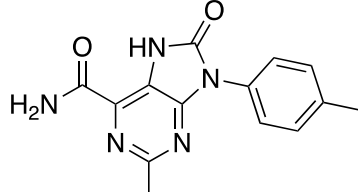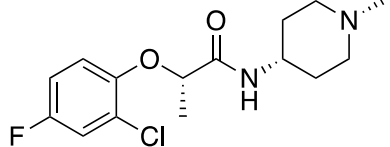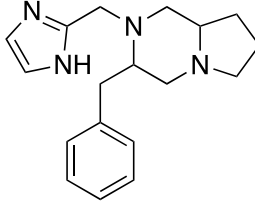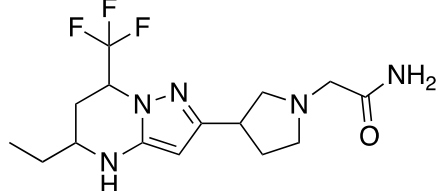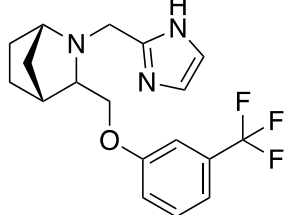
100	AOP 12841219	Asinex	
101	ART 10298689	Asinex	
102	9090359	ChemBridge	
103	SYN 15395045	Asinex	
104	ADM 19561463	Asinex	
105	AAM 19379433	Asinex	

*Table 4 – Chemical structures of compounds **64-105** tested in the BRET assay in section 3.5 (AlphaFold model).*

9   References

1.   Libby, P.; Buring, J. E.; Badimon, L.; Hansson, G. K.; Deanfield, J.; Bittencourt, M. S.; Tokgözoğlu, L.; Lewis, E. F., Atherosclerosis. *Nature Reviews Disease Primers* **2019,** *5* (1), e56.

2.   Benjamin, E. J.; Blaha, M. J.; Chiuve, S. E.; Cushman, M.; Das, S. R.; Deo, R.; de Ferranti, S. D.; Floyd, J.; Fornage, M.; Gillespie, C.; Isasi, C. R.; Jiménez, M. C.; Jordan, L. C.; Judd, S. E.; Lackland, D.; Lichtman, J. H.; Lisabeth, L.; Liu, S.; Longenecker, C. T.; Mackey, R. H.; Matsushita, K.; Mozaffarian, D.; Mussolino, M. E.; Nasir, K.; Neumar, R. W.; Palaniappan, L.; Pandey, D. K.; Thiagarajan, R. R.; Reeves, M. J.; Ritchey, M.; Rodriguez, C. J.; Roth, G. A.; Rosamond, W. D.; Sasson, C.; Towfighi, A.; Tsao, C. W.; Turner, M. B.; Virani, S. S.; Voeks, J. H.; Willey, J. Z.; Wilkins, J. T.; Wu, J. H.; Alger, H. M.; Wong, S. S.; Muntner, P., Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. *Circulation* **2017,** *135* (10), 146-603.
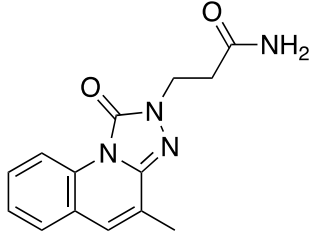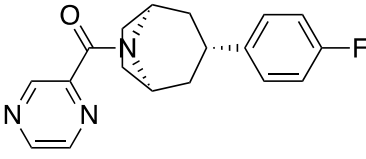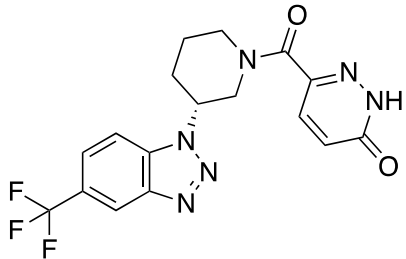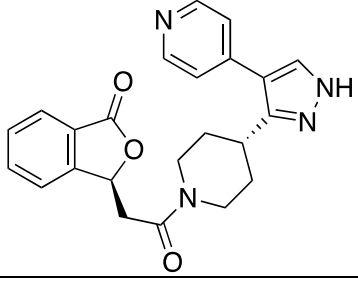
3.   Musunuru, K., Moving toward genome-editing therapies for cardiovascular diseases. *Journal of Clinical Investigation* **2022,** *132* (1), e148555.

4.   Gatsiou, A.; Stellos, K., RNA modifications in cardiovascular health and disease. *Nature Reviews Cardiology* **2023,** *20*, 325-346.

5.   Botts, S. R.; Fish, J. E.; Howe, K. L., Dysfunctional Vascular Endothelium as a Driver of Atherosclerosis: Emerging Insights Into Pathogenesis and Treatment. *Frontiers in Pharmacology* **2021,** *12*, 1-18.

6.   Mundi, S.; Massaro, M.; Scoditti, E.; Carluccio, M. A.; van Hinsbergh, V. W. M.; Iruela-Arispe, M. L.; De Caterina, R., Endothelial permeability, LDL deposition, and cardiovascular risk factors-a review. *Cardiovascular Research* **2018,** *114* (1), 35-52.

7.   Cahill, P. A.; Redmond, E. M., Vascular endothelium – Gatekeeper of vessel health. *Atherosclerosis* **2016,** *248*, 97-109.

8.   Davignon, J.; Ganz, P., Role of endothelial dysfunction in atherosclerosis. *Circulation* **2004,** *109* (23), 27-32.

9.   Sitia, S.; Tomasoni, L.; Atzeni, F.; Ambrosio, G.; Cordiano, C.; Catapano, A.; Tramontana, S.; Perticone, F.; Naccarato, P.; Camici, P.; Picano, E.; Cortigiani, L.; Bevilacqua, M.; Milazzo, L.; Cusi, D.; Barlassina, C.; Sarzi-Puttini, P.; Turiel, M., From endothelial dysfunction to atherosclerosis. *Autoimmunity Reviews* **2010,** *9* (12), 830-834.

10.   Förstermann, U.; Xia, N.; Li, H., Roles of Vascular Oxidative Stress and Nitric Oxide in the Pathogenesis of Atherosclerosis. *Circulation Research* **2017,** *120* (4), 713-735.

11.   Farah, C.; Michel, L. Y. M.; Balligand, J.-L., Nitric oxide signalling in cardiovascular health and disease. *Nature Reviews Cardiology* **2018,** *15* (5), 292-316.

12.   Noris, M.; Morigi, M.; Donadelli, R.; Aiello, S.; Foppolo, M.; Todeschini, M.; Orisio, S.; Remuzzi, G.; Remuzzi, A., Nitric Oxide Synthesis by Cultured Endothelial Cells Is Modulated by Flow Conditions. *Circulation Research* **1995,** *76* (4), 536-543.

13.	Quillard, T.; Franck, G.; Mawson, T.; Folco, E.; Libby, P., Mechanisms of erosion of atherosclerotic plaques. *Current Opinion in Lipidology* **2017,** *28* (5), 434-441.

14.	Libby, P., The changing landscape of atherosclerosis. *Nature* **2021,** *592* (7855), 524-533.

15.	Katsarou, A.; Gudbjörnsdottir, S.; Rawshani, A.; Dabelea, D.; Bonifacio, E.; Anderson, B. J.; Jacobsen, L. M.; Schatz, D. A.; Lernmark, Å., Type 1 diabetes mellitus. *Nature Reviews Disease Primers* **2017,** *3* (1), 17016.

16.	Muoio, D. M.; Newgard, C. B., Molecular and metabolic mechanisms of insulin resistance and β-cell failure in type 2 diabetes. *Nature Reviews Molecular Cell Biology* **2008,** *9* (3), 193-205.

17.	Poznyak, A.; Grechko, A. V.; Poggio, P.; Myasoedova, V. A.; Alfieri, V.; Orekhov, A. N., The Diabetes Mellitus–Atherosclerosis Connection: The Role of Lipid and Glucose Metabolism and Chronic Inflammation. *International Journal of Molecular Sciences* **2020,** *21* (5), e1835.

18.	Taskinen, M.-R.; Borén, J., New insights into the pathophysiology of dyslipidemia in type 2 diabetes. *Atherosclerosis* **2015,** *239* (2), 483-495.

19.	Berneis, K. K.; Krauss, R. M., Metabolic origins and clinical significance of LDL heterogeneity. *Journal of Lipid Research* **2002,** *43* (9), 1363-1379.

20.	Hagensen, M. K.; Mortensen, M. B.; Kjolby, M.; Palmfeldt, J.; Bentzon, J. F.; Gregersen, S., Increased retention of LDL from type 1 diabetic patients in atherosclerosis-prone areas of the murine arterial wall. *Atherosclerosis* **2019,** *286*, 156-162.

21.	Duvillard, L.; Pont, F.; Florentin, E.; Gambert, P.; Vergès, B., Significant improvement of apolipoprotein B-containing lipoprotein metabolism by insulin treatment in patients with non-insulin-dependent diabetes mellitus. *Diabetologia* **2000,** *43* (1), 27-35.

22.	Pizzino, G.; Irrera, N.; Cucinotta, M.; Pallio, G.; Mannino, F.; Arcoraci, V.; Squadrito, F.; Altavilla, D.; Bitto, A., Oxidative Stress: Harms and Benefits for Human Health. *Oxidative Medicine and Cellular Longevity* **2017,** *2017*, 1-13.

23.	Giacco, F.; Brownlee, M., Oxidative Stress and Diabetic Complications. *Circulation Research* **2010,** *107* (9), 1058-1070.

24.	Lander, H. M.; Tauras, J. M.; Ogiste, J. S.; Hori, O.; Moss, R. A.; Schmidt, A. M., Activation of the Receptor for Advanced Glycation End Products Triggers a p21 - dependent Mitogen-activated Protein Kinase Pathway Regulated by Oxidant Stress. *Journal of Biological Chemistry* **1997,** *272* (28), 17810-17814.

25.	Schmidt, A. M.; Hori, O.; Chen, J. X.; Li, J. F.; Crandall, J.; Zhang, J.; Cao, R.; Yan, S. D.; Brett, J.; Stern, D., Advanced glycation endproducts interacting with their endothelial receptor induce expression of vascular cell adhesion molecule-1 (VCAM-1) in cultured human endothelial cells and in mice. A potential mechanism for the accelerated vasculopathy of diabetes. *Journal of Clinical Investigation* **1995,** *96* (3), 1395-1403.

26.	Defronzo, R. A.; Ferrannini, E.; Groop, L.; Henry, R. R.; Herman, W. H.; Holst, J. J.; Hu, F. B.; Kahn, C. R.; Raz, I.; Shulman, G. I.; Simonson, D. C.; Testa, M. A.; Weiss, R., Type 2 diabetes mellitus. *Nature Reviews Disease Primers* **2015,** *1* (1), e15019.

27.   Rena, G.; Pearson, E. R.; Sakamoto, K., Molecular mechanism of action of metformin: old or new insights? *Diabetologia* **2013,** *56* (9), 1898-1906.

28.   Lamoia, T. E.; Butrico, G. M.; Kalpage, H. A.; Goedeke, L.; Hubbard, B. T.; Vatner, D. F.; Gaspar, R. C.; Zhang, X.-M.; Cline, G. W.; Nakahara, K.; Woo, S.; Shimada, A.; Hüttemann, M.; Shulman, G. I., Metformin, phenformin, and galegine inhibit complex IV activity and reduce glycerol-derived gluconeogenesis. *Proceedings of the National Academy of Sciences* **2022,** *119* (10), e212228119.

29.   Zhou, G.; Myers, R.; Li, Y.; Chen, Y.; Shen, X.; Fenyk-Melody, J.; Wu, M.; Ventre, J.; Doebber, T.; Fujii, N.; Musi, N.; Hirshman, M. F.; Goodyear, L. J.; Moller, D. E., Role of AMP-activated protein kinase in mechanism of metformin action. *Journal of Clinical Investigation* **2001,** *108* (8), 1167-1174.

30.   Seino, S., Cell signalling in insulin secretion: the molecular targets of ATP, cAMP and sulfonylurea. *Diabetologia* **2012,** *55* (8), 2096-2108.

31.   Shyangdan, D. S.; Royle, P.; Clar, C.; Sharma, P.; Waugh, N.; Snaith, A., Glucagon-like peptide analogues for type 2 diabetes mellitus. *Cochrane Database of Systematic Reviews* **2011,** *1* (10), CD0006423.

32.   Blicklé, J. F., Meglitinide analogues: a review of clinical data focused on recent trials. *Diabetes Metab* **2006,** *32* (2), 113-20.

33.   Miyazaki, Y.; He, H.; Mandarino, L. J.; DeFronzo, R. A., Rosiglitazone improves downstream insulin receptor signaling in type 2 diabetic patients. *Diabetes* **2003,** *52* (8), 1943-50.

34.   Gastaldelli, A.; Ferrannini, E.; Miyazaki, Y.; Matsuda, M.; Mari, A.; DeFronzo, R. A., Thiazolidinediones improve beta-cell function in type 2 diabetic patients. *The American Journal of Physiology - Endocrinology and Metabolism* **2007,** *292* (3), E871-83.

35.   Greenfield, J. R.; Chisholm, D. J., Experimental and clinical pharmacology: Thiazolidinediones - mechanisms of action. *Australian Prescriber* **2004,** *27* (3), 67-70.

36.   Nissen, S. E.; Wolski, K., Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New England Journal of Medicine* **2007,** *356* (24), 2457-71.

37.   Lewis, J. D.; Habel, L. A.; Quesenberry, C. P.; Strom, B. L.; Peng, T.; Hedderson, M. M.; Ehrlich, S. F.; Mamtani, R.; Bilker, W.; Vaughn, D. J.; Nessel, L.; Van Den Eeden, S. K.; Ferrara, A., Pioglitazone Use and Risk of Bladder Cancer and Other Common Cancers in Persons With Diabetes. *JAMA* **2015,** *314* (3), 265-77.

38.   Weber, A. E., Dipeptidyl Peptidase IV Inhibitors for the Treatment of Diabetes. *Journal of Medicinal Chemistry* **2004,** *47* (17), 4135-4141.

39.   Zheng, S. L.; Roddick, A. J.; Aghar-Jaffar, R.; Shun-Shin, M. J.; Francis, D.; Oliver, N.; Meeran, K., Association Between Use of Sodium-Glucose Cotransporter 2 Inhibitors, Glucagon-like Peptide 1 Agonists, and Dipeptidyl Peptidase 4 Inhibitors With All-Cause Mortality in Patients With Type 2 Diabetes. *JAMA* **2018,** *319* (15), 1580.

40.   Griffith, D. A.; Edmonds, D. J.; Fortin, J.-P.; Kalgutkar, A. S.; Kuzmiski, J. B.; Loria, P. M.; Saxena, A. R.; Bagley, S. W.; Buckeridge, C.; Curto, J. M.; Derksen, D. R.; Dias, J. M.; Griffor, M. C.; Han, S.; Jackson, V. M.; Landis, M. S.; Lettiere, D.; Limberakis, C.; Liu, Y.; Mathiowetz, A. M.; Patel, J. C.; Piotrowski, D. W.; Price, D. A.; Ruggeri, R. B.; Tess, D. A., A Small-Molecule Oral Agonist of the Human

Glucagon-like Peptide-1 Receptor. *Journal of Medicinal Chemistry* **2022,** *65* (12), 8208-8226.

41.   Fleming, I.; Busse, R., Signal transduction of eNOS activation. *Cardiovascular Research* **1999,** *43* (3), 532-541.

42.   Laine, H.; Yki-Jarvinen, H.; Kirvela, O.; Tolvanen, T.; Raitakari, M.; Solin, O.; Haaparanta, M.; Knuuti, J.; Nuutila, P., Insulin resistance of glucose uptake in skeletal muscle cannot be ameliorated by enhancing endothelium-dependent blood flow in obesity. *Journal of Clinical Investigation* **1998,** *101* (5), 1156-1162.

43.   Steinberg, H. O.; Chaker, H.; Leaming, R.; Johnson, A.; Brechtel, G.; Baron, A. D., Obesity/insulin resistance is associated with endothelial dysfunction. Implications for the syndrome of insulin resistance. *Journal of Clinical Investigation* **1996,** *97* (11), 2601-2610.

44.   Tack, C. J. J.; Ong, M. K. E.; Lutterman, J. A.; Smits, P., Insulin-induced vasodilatation and endothelial function in obesity/insulin resistance. Effects of troglitazone. *Diabetologia* **1998,** *41* (5), 569-576.

45.   Melikian, N.; Chowienczyk, P.; MacCarthy, P. A.; Williams, I. L.; Wheatcroft, S. B.; Sherwood, R.; Gale, C.; Shah, A. M.; Kearney, M. T., Determinants of endothelial function in asymptomatic subjects with and without the metabolic syndrome. *Atherosclerosis* **2008,** *197* (1), 375-382.

46.   Hakuno, F.; Takahashi, S.-I., 40 YEARS OF IGF1: IGF1 receptor signaling pathways. *Journal of Molecular Endocrinology* **2018,** *61* (1), T69-T86.

47.   Denley, A.; Cosgrove, L. J.; Booker, G. W.; Wallace, J. C.; Forbes, B. E., Molecular interactions of the IGF system. *Cytokine & Growth Factor Reviews* **2005,** *16* (4-5), 421-439.

48.   Higashi, Y.; Gautam, S.; Delafontaine, P.; Sukhanov, S., IGF-1 and cardiovascular disease. *Growth Hormone & IGF Research* **2019,** *45*, 6-16.

49.   Cai, W.; Sakaguchi, M.; Kleinridders, A.; Gonzalez-Del Pino, G.; Dreyfuss, J. M.; O'Neill, B. T.; Ramirez, A. K.; Pan, H.; Winnay, J. N.; Boucher, J.; Eck, M. J.; Kahn, C. R., Domain-dependent effects of insulin and IGF-1 receptors on signalling and gene expression. *Nature Communications* **2017,** *8* (1), 14892.

50.   Belfiore, A.; Frasca, F.; Pandini, G.; Sciacca, L.; Vigneri, R., Insulin Receptor Isoforms and Insulin Receptor/Insulin-Like Growth Factor Receptor Hybrids in Physiology and Disease. *Endocrine Reviews* **2009,** *30* (6), 586-623.

51.   Higashi, Y.; Sukhanov, S.; Shai, S.-Y.; Danchuk, S.; Tang, R.; Snarski, P.; Li, Z.; Lobelle-Rich, P.; Wang, M.; Wang, D.; Yu, H.; Korthuis, R.; Delafontaine, P., Insulin-Like Growth Factor-1 Receptor Deficiency in Macrophages Accelerates Atherosclerosis and Induces an Unstable Plaque Phenotype in Apolipoprotein E– Deficient Mice. *Circulation* **2016,** *133* (23), 2263-2278.

52.   Nakao-Hayashi, J.; Ito, H.; Kanayasu, T.; Morita, I.; Murota, S.-i., Stimulatory effects of insulin and insulin-like growth factor I on migration and tube formation by vascular endothelial cells. *Atherosclerosis* **1992,** *92* (2), 141-149.

53.   Zeng, G.; Quon, M. J., Insulin-stimulated production of nitric oxide is inhibited by wortmannin. Direct measurement in vascular endothelial cells. *Journal of Clinical Investigation* **1996,** *98* (4), 894-898.

54.   Ullrich, A.; Bell, J. R.; Chen, E. Y.; Herrera, R.; Petruzzelli, L. M.; Dull, T. J.; Gray, A.; Coussens, L.; Liao, Y. C.; Tsubokawa, M.; Mason, A.; Seeburg, P. H.; Grunfeld, C.; Rosen, O. M.; Ramachandran, J., Human insulin receptor and its

relationship to the tyrosine kinase family of oncogenes. *Nature* **1985,** *313* (6005), 756-761.

55.   Ullrich, A.; Gray, A.; Tam, A. W.; Yang-Feng, T.; Tsubokawa, M.; Collins, C.; Henzel, W.; Le Bon, T.; Kathuria, S.; Chen, E.; et al., Insulin-like growth factor I receptor primary structure: comparison with insulin receptor suggests structural determinants that define functional specificity. *EMBO Journal* **1986,** *5* (10), 2503-12.

56.   Lane, M. D.; Ronnett, G.; Slieker, L. J.; Kohanski, R. A.; Olson, T. L., Post-translational processing and activation of insulin and EGF proreceptors. *Biochimie* **1985,** *67* (10-11), 1069-1080.

57.   Wu, J. J.; Guidotti, G., Construction and Characterization of a Monomeric Insulin Receptor. *Journal of Biological Chemistry* **2002,** *277* (31), 27809-27817.

58.   Wu, J. J.; Guidotti, G., Proreceptor Dimerization Is Required for Insulin Receptor Post-translational Processing. *Journal of Biological Chemistry* **2004,** *279* (24), 25765-25773.

59.   Turvey, S. J.; McPhillie, M. J.; Kearney, M. T.; Muench, S. P.; Simmons, K. J.; Fishwick, C. W. G., Recent developments in the structural characterisation of the IR and IGF1R: implications for the design of IR-IGF1R hybrid receptor modulators. *RSC Med Chem* **2022,** *13* (4), 360-374.

60.   Croll, T. I.; Smith, B. J.; Margetts, M. B.; Whittaker, J.; Weiss, M. A.; Ward, C. W.; Lawrence, M. C., Higher-Resolution Structure of the Human Insulin Receptor Ectodomain: Multi-Modal Inclusion of the Insert Domain. *Structure* **2016,** *24* (3), 469-76.

61.   Xu, Y.; Kong, G. K.; Menting, J. G.; Margetts, M. B.; Delaine, C. A.; Jenkin, L. M.; Kiselyov, V. V.; De Meyts, P.; Forbes, B. E.; Lawrence, M. C., How ligand binds to the type 1 insulin-like growth factor receptor. *Nature Communications* **2018,** *9* (1), 821.

62.   Williams, P. F.; Mynarcik, D. C.; Yu, G. Q.; Whittaker, J., Mapping of an NH2-terminal ligand binding site of the insulin receptor by alanine scanning mutagenesis. *The Journal of Biological Chemistry* **1995,** *270* (7), 3012-3016.

63.   Mynarcik, D. C.; Yu, G. Q.; Whittaker, J., Alanine-scanning mutagenesis of a C-terminal ligand binding domain of the insulin receptor alpha subunit. *The Journal of Biological Chemistry* **1996,** *271* (5), 2439-2442.

64.   Whittaker, J.; Whittaker, L., Characterization of the functional insulin binding epitopes of the full-length insulin receptor. *Journal of Biological Chemistry* **2005,** *280* (22), 20932-6.

65.   Hao, C.; Whittaker, L.; Whittaker, J., Characterization of a second ligand binding site of the insulin receptor. *Biochemical and Biophysical Research Communications* **2006,** *347* (1), 334-9.

66.   Keyhanfar, M.; Booker, G. W.; Whittaker, J.; Wallace, J. C.; Forbes, B. E., Precise mapping of an IGF-I-binding site on the IGF-1R. *Biochemical Journal* **2007,** *401* (1), 269-77.

67.   De Meyts, P., The structural basis of insulin and insulin-like growth factor-I receptor binding and negative co-operativity, and its relevance to mitogenic versus metabolic signalling. *Diabetologia* **1994,** *37* (S2), S135-S148.

68.   Weis, F.; Menting, J. G.; Margetts, M. B.; Chan, S. J.; Xu, Y.; Tennagels, N.; Wohlfart, P.; Langer, T.; Muller, C. W.; Dreyer, M. K.; Lawrence, M. C., The

signalling conformation of the insulin receptor ectodomain. *Nature Communications* **2018,** *9* (1), 4420.

69.   Gutmann, T.; Kim, K. H.; Grzybek, M.; Walz, T.; Coskun, U., Visualization of ligand-induced transmembrane signaling in the full-length human insulin receptor. *Journal of Cell Biology* **2018,** *217* (5), 1643-1649.

70.   Scapin, G.; Dandey, V. P.; Zhang, Z.; Prosise, W.; Hruza, A.; Kelly, T.; Mayhood, T.; Strickland, C.; Potter, C. S.; Carragher, B., Structure of the insulin receptor-insulin complex by single-particle cryo-EM analysis. *Nature* **2018,** *556* (7699), 122-125.

71.   Uchikawa, E.; Choi, E.; Shang, G. J.; Yu, H. T.; Bai, X. C., Activation mechanism of the insulin receptor revealed by cryo-EM structure of the fully liganded receptor-ligand complex. *Elife* **2019,** *8*, e48630.

72.   Gutmann, T.; Schafer, I. B.; Poojari, C.; Brankatschk, B.; Vattulainen, I.; Strauss, M.; Coskun, U., Cryo-EM structure of the complete and ligand-saturated insulin receptor ectodomain. *Journal of Cell Biology* **2020,** *219* (1).

73.   Horwitz, D. L.; Starr, J. I.; Mako, M. E.; Blackard, W. G.; Rubenstein, A. H., Proinsulin, insulin, and C-peptide concentrations in human portal and peripheral blood. *Journal of Clinical Investigation* **1975,** *55* (6), 1278-1283.

74.   Li, J.; Choi, E.; Yu, H.; Bai, X. C., Structural basis of the activation of type 1 insulin-like growth factor receptor. *Nature Communications* **2019,** *10* (1), 4567.

75.   Zhang, X.; Yu, D.; Sun, J.; Wu, Y.; Gong, J.; Li, X.; Liu, L.; Liu, S.; Liu, J.; Wu, Y.; Li, D.; Ma, Y.; Han, X.; Zhu, Y.; Wu, Z.; Wang, Y.; Ouyang, Q.; Wang, T., Visualization of Ligand-Bound Ectodomain Assembly in the Full-Length Human IGF-1 Receptor by Cryo-EM Single-Particle Analysis. *Structure* **2020,** *28* (5), 555-561 e4.

76.   Xu, Y.; Kirk, N. S.; Venugopal, H.; Margetts, M. B.; Croll, T. I.; Sandow, J. J.; Webb, A. I.; Delaine, C. A.; Forbes, B. E.; Lawrence, M. C., How IGF-II Binds to the Human Type 1 Insulin-like Growth Factor Receptor. *Structure* **2020,** *28* (7), 786-798.

77.   Li, Q.; Wong, Y. L.; Kang, C., Solution structure of the transmembrane domain of the insulin receptor in detergent micelles. *Biochimica et Biophysica Acta* **2014,** *1838* (5), 1313-21.

78.   Mohammadiarani, H.; Vashisth, H., All-Atom Structural Models of the Transmembrane Domains of Insulin and Type 1 Insulin-Like Growth Factor Receptors. *Front Endocrinol (Lausanne)* **2016,** *7*, 68-68.

79.   De Meyts, P., Insulin/receptor binding: The last piece of the puzzle? *BioEssays* **2015,** *37* (4), 389-397.

80.   Kuznetsov, A. S.; Zamaletdinov, M. F.; Bershatsky, Y. V.; Urban, A. S.; Bocharova, O. V.; Bennasroune, A.; Maurice, P.; Bocharov, E. V.; Efremov, R. G., Dimeric states of transmembrane domains of insulin and IGF-1R receptors: Structures and possible role in activation. *Biochimica et Biophysica Acta - Biomembranes* **2020,** *1862* (11), 183417.

81.   Kavran, J. M.; McCabe, J. M.; Byrne, P. O.; Connacher, M. K.; Wang, Z.; Ramek, A.; Sarabipour, S.; Shan, Y.; Shaw, D. E.; Hristova, K.; Cole, P. A.; Leahy, D. J., How IGF-1 activates its receptor. *Elife* **2014,** *3*.

82.   Hubbard, S. R.; Wei, L.; Hendrickson, W. A., Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* **1994,** *372* (6508), 746-754.

83.   Hubbard, S. R., Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. *EMBO journal* **1997,** *16* (18), 5572-5581.

84.   Favelyukis, S.; Till, J. H.; Hubbard, S. R.; Miller, W. T., Structure and autoregulation of the insulin-like growth factor 1 receptor kinase. *Nature Structural Biology* **2001,** *8* (12), 1058-1063.

85.   Munshi, S.; Kornienko, M.; Hall, D. L.; Reid, J. C.; Waxman, L.; Stirdivant, S. M.; Darke, P. L.; Kuo, L. C., Crystal structure of the Apo, unactivated insulin-like growth factor-1 receptor kinase. Implication for inhibitor specificity. *Journal of Biological Chemistry* **2002,** *277* (41), 38797-802.

86.   Munshi, S.; Hall, D. L.; Kornienko, M.; Darke, P. L.; Kuo, L. C., Structure of apo, unactivated insulin‐like growth factor‐1 receptor kinase at 1.5 Å resolution. *Acta Crystallographica Section D* **2003,** *59* (10), 1725-1730.

87.   Anastassiadis, T.; Duong-Ly, K. C.; Deacon, S. W.; Lafontant, A.; Ma, H.; Devarajan, K.; Dunbrack, R. L., Jr.; Wu, J.; Peterson, J. R., A highly selective dual insulin receptor (IR)/insulin-like growth factor 1 receptor (IGF-1R) inhibitor derived from an extracellular signal-regulated kinase (ERK) inhibitor. *Journal of Biological Chemistry* **2013,** *288* (39), 28068-77.

88.   Stauffer, F.; Cowan-Jacob, S. W.; Scheufler, C.; Furet, P., Identification of a 5-[3-phenyl-(2-cyclic-ether)-methylether]-4-aminopyrrolo[2,3-d]pyrimidine series of IGF-1R inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2016,** *26* (8), 2065-2067.

89.   Sanderson, M. P.; Apgar, J.; Garin-Chesa, P.; Hofmann, M. H.; Kessler, D.; Quant, J.; Savchenko, A.; Schaaf, O.; Treu, M.; Tye, H.; Zahn, S. K.; Zoephel, A.; Haaksma, E.; Adolf, G. R.; Kraut, N., BI 885578, a Novel IGF1R/INSR Tyrosine Kinase Inhibitor with Pharmacokinetic Properties That Dissociate Antitumor Efficacy and Perturbation of Glucose Homeostasis. *Molecular Cancer Therapy* **2015,** *14* (12), 2762-72.

90.   Chamberlain, S. D.; Redman, A. M.; Wilson, J. W.; Deanda, F.; Shotwell, J. B.; Gerding, R.; Lei, H.; Yang, B.; Stevens, K. L.; Hassell, A. M.; Shewchuk, L. M.; Leesnitzer, M. A.; Smith, J. L.; Sabbatini, P.; Atkins, C.; Groy, A.; Rowand, J. L.; Kumar, R.; Mook, R. A., Jr.; Moorthy, G.; Patnaik, S., Optimization of 4,6-bis-anilino-1H-pyrrolo[2,3-d]pyrimidine IGF-1R tyrosine kinase inhibitors towards JNK selectivity. *Bioorganic & Medicinal Chemistry Letters* **2009,** *19* (2), 360-4.

91.   Patnaik, S.; Stevens, K. L.; Gerding, R.; Deanda, F.; Shotwell, J. B.; Tang, J.; Hamajima, T.; Nakamura, H.; Leesnitzer, M. A.; Hassell, A. M.; Shewchuk, L. M.; Kumar, R.; Lei, H.; Chamberlain, S. D., Discovery of 3,5-disubstituted-1H-pyrrolo[2,3-b]pyridines as potent inhibitors of the insulin-like growth factor-1 receptor (IGF-1R) tyrosine kinase. *Bioorganic & Medicinal Chemistry Letters* **2009,** *19* (11), 3136-40.

92.   Parang, K.; Till, J. H.; Ablooglu, A. J.; Kohanski, R. A.; Hubbard, S. R.; Cole, P. A., Mechanism-based design of a protein kinase inhibitor. *Nature Structural Biology* **2001,** *8* (1), 37-41.

93.   Rosen, O. M.; Herrera, R.; Olowe, Y.; Petruzzelli, L. M.; Cobb, M. H., Phosphorylation activates the insulin receptor tyrosine protein kinase. *Proceedings of the National Academy of Sciences* **1983,** *80* (11), 3237-3240.

94. Ellis, L.; Clauser, E.; Morgan, D. O.; Edery, M.; Roth, R. A.; Rutter, W. J., Replacement of insulin receptor tyrosine residues 1162 and 1163 compromises insulin-stimulated kinase activity and uptake of 2-deoxyglucose. *Cell* **1986,** *45* (5), 721-732.

95. Huse, M.; Kuriyan, J., The Conformational Plasticity of Protein Kinases. *Cell* **2002,** *109* (3), 275-282.

96. Moxham, C. P.; Duronio, V.; Jacobs, S., Insulin-like growth factor I receptor beta-subunit heterogeneity. Evidence for hybrid tetramers composed of insulin-like growth factor I and insulin receptor heterodimers. *The Journal of Biological Chemistry* **1989,** *264* (22), 13238-13244.

97. Soos, M. A.; Siddle, K., Immunological relationships between receptors for insulin and insulin-like growth factor I. Evidence for structural heterogeneity of insulin-like growth factor I receptors involving hybrids with insulin receptors. *Biochemical Journal* **1989,** *263* (2), 553-563.

98. Slaaby, R., Hybrid Receptors Formed by Insulin Receptor (IR) and Insulin-like Growth Factor I Receptor (IGF-IR) Have Low Insulin and High IGF-1 Affinity Irrespective of the IR Splice Variant. *Journal of Biological Chemistry* **2006,** *281* (36), 25869-25874.

99. Pandini, G.; Frasca, F.; Mineo, R.; Sciacca, L.; Vigneri, R.; Belfiore, A., Insulin/Insulin-like Growth Factor I Hybrid Receptors Have Different Biological Characteristics Depending on the Insulin Receptor Isoform Involved. *Journal of Biological Chemistry* **2002,** *277* (42), 39684-39695.

100. Henderson, S. T.; Brierley, G. V.; Surinya, K. H.; Priebe, I. K.; Catcheside, D. E. A.; Wallace, J. C.; Forbes, B. E.; Cosgrove, L. J., Delineation of the IGF-II C domain elements involved in binding and activation of the IR-A, IR-B and IGF-IR. *Growth Hormone & IGF Research* **2015,** *25* (1), 20-27.

101. Bailyes, E. M.; Navé, B. T.; Soos, M. A.; Orr, S. R.; Hayward, A. C.; Siddle, K., Insulin receptor/IGF-I receptor hybrids are widely distributed in mammalian tissues: quantification of individual receptor species by selective immunoprecipitation and immunoblotting. *Biochemical Journal* **1997,** *327* (1), 209-215.

102. Treadway, J.; Morrison, B.; Goldfine, I.; Pessin, J., Assembly of insulin/insulin-like growth factor-I hybrid receptors in vitro. *The Journal of Biological Chemistry* **1990,** *264*, 21450-3.

103. Frattali, A. L.; Pessin, J. E., Relationship between alpha subunit ligand occupancy and beta subunit autophosphorylation in insulin/insulin-like growth factor-1 hybrid receptors. *The Journal of Biological Chemistry* **1993,** *268* (10), 7393-7400.

104. Abbas, A.; Imrie, H.; Viswambharan, H.; Sukumar, P.; Rajwani, A.; Cubbon, R. M.; Gage, M.; Smith, J.; Galloway, S.; Yuldeshava, N.; Kahn, M.; Xuan, S.; Grant, P. J.; Channon, K. M.; Beech, D. J.; Wheatcroft, S. B.; Kearney, M. T., The insulin-like growth factor-1 receptor is a negative regulator of nitric oxide bioavailability and insulin sensitivity in the endothelium. *Diabetes* **2011,** *60* (8), 2169-78.

105. Imrie, H.; Viswambharan, H.; Sukumar, P.; Abbas, A.; Cubbon, R. M.; Yuldasheva, N.; Gage, M.; Smith, J.; Galloway, S.; Skromna, A.; Rashid, S. T.; Futers, T. S.; Xuan, S.; Gatenby, V. K.; Grant, P. J.; Channon, K. M.; Beech, D. J.; Wheatcroft, S. B.; Kearney, M. T., Novel role of the IGF-1 receptor in endothelial

function and repair: studies in endothelium-targeted IGF-1 receptor transgenic mice. *Diabetes* **2012,** *61* (9), 2359.

106.   Valensise, H.; Liu, Y. Y.; Federici, M.; Lauro, D.; Dell'Anna, D.; Romanini, C.; Sesti, G., Increased expression of low-affinity insulin receptor isoform and insulin/insulin-like growth factor-I hybrid receptors in term placenta from insulin-resistant women with gestational hypertension. *Diabetologia* **1996,** *39* (8), 952-960.

107.   Federici, M.; Porzio, O.; Lauro, D.; Borboni, P.; Giovannone, B.; Zucaro, L.; Hribal, M. L.; Sesti, G., Increased Abundance of Insulin/Insulin-Like Growth Factor-I Hybrid Receptors in Skeletal Muscle of Obese Subjects Is Correlated withIn VivoInsulin Sensitivity1. *The Journal of Clinical Endocrinology & Metabolism* **1998,** *83* (8), 2911-2915.

108.   Federici, M.; Lauro, D.; D'Adamo, M.; Giovannone, B.; Porzio, O.; Mellozzi, M.; Tamburrano, G.; Sbraccia, P.; Sesti, G., Expression of Insulin/IGF-I Hybrid Receptors Is Increased in Skeletal Muscle of Patients With Chronic Primary Hyperinsulinemia. *Diabetes* **1998,** *47* (1), 87-92.

109.   Federici, M.; Giaccari, A.; Hribal, M. L.; Giovannone, B.; Lauro, D.; Morviducci, L.; Pastore, L.; Tamburrano, G.; Lauro, R.; Sesti, G., Evidence for glucose/hexosamine in vivo regulation of insulin/IGF-I hybrid receptor assembly. *Diabetes* **1999,** *48* (12), 2277-2285.

110.   Federici, M.; Zucaro, L.; Porzio, O.; Massoud, R.; Borboni, P.; Lauro, D.; Sesti, G., Increased expression of insulin/insulin-like growth factor-I hybrid receptors in skeletal muscle of noninsulin-dependent diabetes mellitus subjects. *Journal of Clinical Investigation* **1996,** *98* (12), 2887-2893.

111.   John; Callum; Geoffrey; Mai; Colin; Michael, Structural Congruency of Ligand Binding to the Insulin and Insulin/Type 1 Insulin-like Growth Factor Hybrid Receptors. *Structure* **2015,** *23* (7), 1271-1282.

112.   Xu, Y.; Margetts, M. B.; Venugopal, H.; Menting, J. G.; Kirk, N. S.; Croll, T. I.; Delaine, C.; Forbes, B. E.; Lawrence, M. C., How insulin-like growth factor I binds to a hybrid insulin receptor type 1 insulin-like growth factor receptor. *Structure* **2022,** *30* (8).

113.   Simmons, K. J.; Myers, C. G.; Viswambharan, H.; Haywood, N. J.; Bridge, K.; Turvey, S.; Armstrong, T.; Lunn, L.; Meakin, P. J.; Clavane, E.; Beech, D. J.; Cubbon, R. M.; Wheatcroft, S. B.; McPhillie, M. J.; Fishwick, C. W.; Kearney, M. T., A small molecule reveals role of insulin receptor-insulin like growth factor-1 receptor heterodimers [Pre-print]. *BioRχiv* **2021**.

114.   Zhang, Y. I-TASSER. https://zhanggroup.org/I-TASSER/ (accessed 04/03/2021).

115.   Roy, A.; Kucukural, A.; Zhang, Y., I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* **2010,** *5* (4), 725-38.

116.   Yang, J.; Zhang, Y., I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* **2015,** *43* (W1), W174-81.

117.   Murohara, T.; Witzenbichler, B.; Spyridopoulos, I.; Asahara, T.; Ding, B.; Sullivan, A.; Losordo, D. W.; Isner, J. M., Role of Endothelial Nitric Oxide Synthase in Endothelial Cell Migration. *Arteriosclerosis, Thrombosis, and Vascular Biology* **1999,** *19* (5), 1156-1161.

118.   Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moult, J., Critical assessment of methods of protein structure prediction (CASP)—Round XIV. *Proteins: Structure, Function, and Bioinformatics* **2021,** *89* (12), 1607-1617.

119.   Doman, T. N.; McGovern, S. L.; Witherbee, B. J.; Kasten, T. P.; Kurumbail, R.; Stallings, W. C.; Connolly, D. T.; Shoichet, B. K., Molecular Docking and High-Throughput Screening for Novel Inhibitors of Protein Tyrosine Phosphatase-1B. *Journal of Medicinal Chemistry* **2002,** *45* (11), 2213-2221.

120.   Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery* **2004,** *3* (11), 935-949.

121.   Bikker, J. A.; Narasimhan, L. S., Lead Discovery Using Virtual Screening. *Topics in Medicinal Chemistry* **2009,** *5*, 85-124.

122.   Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *Journal of Medicinal Chemistry* **2006,** *49* (21), 6177-6196.

123.   *Glide*, Release 2019-4; Schrödinger, LLC: New York, NY, 2019.

124.   Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **2004,** *47* (7), 1739-1749.

125.   Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry* **2004,** *47* (7), 1750-1759.

126.   *Maestro*, Release 2019-4; Schrödinger, LLC: New York, NY, 2019.

127.   Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D., Improved protein-ligand docking using GOLD. *Proteins* **2003,** *52* (4), 609-23.

128.   Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* **2009,** *30* (16), 2785-2791.

129.   Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, S. B.; Johnson, A. P., eHiTS: A new fast, exhaustive flexible ligand docking system. *Journal of Molecular Graphics and Modelling* **2007,** *26* (1), 198-212.

130.   Halgren, T., New Method for Fast and Accurate Binding-site Identification and Analysis. *Chemical Biology & Drug Design* **2007,** *69* (2), 146-148.

131.   Halgren, T. A., Identifying and Characterizing Binding Sites and Assessing Druggability. *Journal of Chemical Information and Modeling* **2009,** *49* (2), 377-389.

132.   Schmidt, T.; Bergner, A.; Schwede, T., Modelling three-dimensional protein structures for applications in drug design. *Drug Discovery Today* **2014,** *19* (7), 890-897.

133.   Huang, B.; Kong, L.; Wang, C.; Ju, F.; Zhang, Q.; Zhu, J.; Gong, T.; Zhang, H.; Yu, C.; Zheng, W.-M.; Bu, D., Protein Structure Prediction: Challenges, Advances, and the Shift of Research Paradigms. *Genomics, Proteomics & Bioinformatics* **2023.**

134.   Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Research* **2000,** *28* (1), 235-242.

135.   Protein Data Bank. https://www.rcsb.org/ (accessed 2019-2023).

136.   Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T., SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **2018,** *46* (W1), W296-W303.

137.   Bienert, S.; Waterhouse, A.; de Beer, Tjaart A. P.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T., The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Research* **2016,** *45* (D1), D313-D319.

138.   SWISS-MODEL. https://swissmodel.expasy.org/ (accessed 13/05/2021).

139.   Xu, J. RaptorX. http://raptorx6.uchicago.edu/ (accessed 04/03/2021).

140.   Wang, S.; Sun, S.; Li, Z.; Zhang, R.; Xu, J., Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. *PLOS Computational Biology* **2017,** *13* (1), e1005324.

141.   Haddad, Y.; Adam, V.; Heger, Z., Ten quick tips for homology modeling of high-resolution protein 3D structures. *PLOS Computational Biology* **2020,** *16* (4), e1007449.

142.   Ziu Liu, B. G., Srayanta Mukherjee, Chengxin Zhang, Aysam Guerler, Jeffrey Brender, Zhiping Weng, Dong-Jun Yu, Yang Zhang, Full-length protein quaternary structure prediction by multi-chain threading fragment reassembly. Submitted, 2022.

143.   Swendsen, R. H.; Wang, J. S., Replica Monte Carlo simulation of spin glasses. *Physical Review Letters* **1986,** *57* (21), 2607-2609.

144.   Wu, S.; Zhang, Y., MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins: Structure, Function, and Bioinformatics* **2008,** *72* (2), 547-556.

145.   Consortium, T. U., UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **2022,** *51* (D1), D523-D531.

146.   Consortium, U. UniProt. https://www.uniprot.org/ (accessed 2019-2023).

147.   Krissinel, E.; Henrick, K., Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **2007,** *372* (3), 774-97.

148.   EMBL-EBI Protein interfaces, surfaces and assemblies service PISA at the European Bioinformatics Institute (PDBePISA). http://www.ebi.ac.uk/pdbe/prot_int/pistart.html (accessed 13/04/2021).

149.   Day, E. S.; Cote, S. M.; Whitty, A., Binding Efficiency of Protein–Protein Complexes. *Biochemistry* **2012,** *51* (45), 9124-9136.

150.   Studer, G.; Rempfer, C.; Waterhouse, A. M.; Gumienny, R.; Haas, J.; Schwede, T., QMEANDisCo-distance constraints applied on model quality estimation. *Bioinformatics* **2020,** *36* (6), 1765-1771.

151.   Scott, D. E.; Bayly, A. R.; Abell, C.; Skidmore, J., Small molecules, big targets: drug discovery faces the protein–protein interaction challenge. *Nature Reviews Drug Discovery* **2016,** *15* (8), 533-550.

152.   Zhu, X.; Mitchell, J. C., KFC2: A knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. *Proteins* **2011,** *79* (9), 2671-2683.

153.   Mitchell, J. C. KFC Server. https://mitchell-web.ornl.gov/KFC_Server/index.php (accessed 04/03/2021).

154.   Darnell, S. J.; Legault, L.; Mitchell, J. C., KFC Server: interactive forecasting of protein interaction hot spots. *Nucleic Acids Research* **2008,** *36* (Web Server), W265-W269.

155.   Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R., Principles of Protein–Protein Interactions: What are the Preferred Ways For Proteins To Interact? *Chemical Reviews* **2008,** *108* (4), 1225-1244.

156.   Stenson, P. D.; Ball, E. V.; Mort, M.; Phillips, A. D.; Shiel, J. A.; Thomas, N. S.; Abeysinghe, S.; Krawczak, M.; Cooper, D. N., Human Gene Mutation Database (HGMD): 2003 update. *Human Mutations* **2003,** *21* (6), 577-81.

157.   Kadowaki, T.; Bevins, C. L.; Cama, A.; Ojamaa, K.; Marcus-Samuels, B.; Kadowaki, H.; Beitz, L.; McKeon, C.; Taylor, S. I., Two mutant alleles of the insulin receptor gene in a patient with extreme insulin resistance. *Science* **1988,** *240* (4853), 787-90.

158.   de Kerdanet, M.; Caron-Debarle, M.; Nivot, S.; Gaillot, T.; Lascols, O.; Fremont, B.; Bonaure, M.; Gie, S.; Massart, C.; Capeau, J., Ten-year improvement of insulin resistance and growth with recombinant human insulin-like growth factor 1 in a patient with insulin receptor mutations resulting in leprechaunism. *Diabetes Metabolism* **2015,** *41* (4), 331-337.

159.   Baell, J. B.; Holloway, G. A., New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *Journal of Medicinal Chemistry* **2010,** *53* (7), 2719-2740.

160.   *Ligprep*, Release 2019-4; Schrödinger, LLC: New York, NY, 2019.

161.   *SiteMap*, Release 2019-4; Schrödinger, LLC: New York, NY, 2019.

162.   *Protein Preparation Wizard*, Release 2019-4; Schrödinger, LLC: New York, NY, 2019.

163.   Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W., Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer Aided Molecular Design* **2013,** *27* (3), 221-34.

164.   Butina, D., Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets. *Journal of Chemical Information and Computer Sciences* **1999,** *39* (4), 747-750.

165.   Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021,** *596* (7873), 583-589.

166.   Callaway, E., DeepMind's AI makes gigantic leap in solving protein structures. *Nature* **2020,** *588*, 203-204.

167.   Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; Ronneberger, O.; Bodenstein, S.; Zielinski, M.; Bridgland, A.; Potapenko, A.; Cowie, A.; Tunyasuvunakool, K.; Jain,

R.; Clancy, E.; Kohli, P.; Jumper, J.; Hassabis, D., Protein complex prediction with AlphaFold-Multimer [Pre-print]. *BioRχiv* **2021**.

168.	*ROCS*, 3.5.0.1; OpenEye Scientific Software: Santa Fe, NM.

169.	Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A., Comparison of Shape-Matching and Docking as Virtual Screening Tools. *Journal of Medicinal Chemistry* **2007,** *50* (1), 74-82.

170.	*EON*, 2.3.7.1; OpenEye Scientific Software: Santa Fe, NM.

171.	Wu, P. G.; Brand, L., Resonance Energy Transfer: Methods and Applications. *Analytical Biochemistry* **1994,** *218* (1), 1-13.

172.	Bacart, J.; Corbel, C.; Jockers, R.; Bach, S.; Couturier, C., The BRET technology and its application to screening assays. *Biotechnol J* **2008,** *3* (3), 311-24.

173.	Pfleger, K. D. G.; Eidne, K. A., Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). *Nature Methods* **2006,** *3* (3), 165-174.

174.	Boute, N.; Jockers, R.; Issad, T., The use of resonance energy transfer in high-throughput screening: BRET versus FRET. *Trends Pharmacol Sci* **2002,** *23* (8), 351-4.

175.	Pavlinov, I.; Salkovski, M.; Aldrich, L. N., Beclin 1–ATG14L Protein–Protein Interaction Inhibitor Selectively Inhibits Autophagy through Disruption of VPS34 Complex I. *Journal of the American Chemical Society* **2020,** *142* (18), 8174-8182.

176.	Sakyiamah, M. M.; Nomura, W.; Kobayakawa, T.; Tamamura, H., Development of a NanoBRET-Based Sensitive Screening Method for CXCR4 Ligands. *Bioconjugate Chemistry* **2019,** *30* (5), 1442-1450.

177.	Blanquart, C.; Gonzalez-Yanes, C.; Issad, T., Monitoring the activation state of insulin/insulin-like growth factor-1 hybrid receptors using bioluminescence resonance energy transfer. *Mol Pharmacol* **2006,** *70* (5), 1802-11.

178.	Brown, M.; Dainty, S.; Strudwick, N.; Mihai, A. D.; Watson, J. N.; Dendooven, R.; Paton, A. W.; Paton, J. C.; Schröder, M., Endoplasmic reticulum stress causes insulin resistance by inhibiting delivery of newly synthesized insulin receptors to the cell surface. *Molecular Biology of the Cell* **2020,** *31* (23), 2597-2629.

179.	Reed, B. C.; Lane, M. D., Insulin receptor synthesis and turnover in differentiating 3T3-L1 preadipocytes. *Proceedings of the National Academy of Sciences* **1980,** *77* (1), 285-289.

180.	Bellenie, B. R.; Cheung, K.-M. J.; Varela, A.; Pierrat, O. A.; Collie, G. W.; Box, G. M.; Bright, M. D.; Gowan, S.; Hayes, A.; Rodrigues, M. J.; Shetty, K. N.; Carter, M.; Davis, O. A.; Henley, A. T.; Innocenti, P.; Johnson, L. D.; Liu, M.; De Klerk, S.; Le Bihan, Y.-V.; Lloyd, M. G.; McAndrew, P. C.; Shehu, E.; Talbot, R.; Woodward, H. L.; Burke, R.; Kirkin, V.; Van Montfort, R. L. M.; Raynaud, F. I.; Rossanese, O. W.; Hoelder, S., Achieving In Vivo Target Depletion through the Discovery and Optimization of Benzimidazolone BCL6 Degraders. *Journal of Medicinal Chemistry* **2020,** *63* (8), 4047-4068.

181.	Li, Q.; Fu, J.; Xia, Y.; Qi, W.; Ishikado, A.; Park, K.; Yokomizo, H.; Huang, Q.; Cai, W.; Rask-Madsen, C.; Kahn, C. R.; King, G. L., Homozygous receptors for insulin and not IGF-1 accelerate intimal hyperplasia in insulin resistance and diabetes. *Nature Communications* **2019,** *10* (1).

182. 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* **2009,** *5*, 243-243.

183. Borroto-Escuela, D. O.; Flajolet, M.; Agnati, L. F.; Greengard, P.; Fuxe, K., Bioluminescence Resonance Energy Transfer Methods to Study G Protein-Coupled Receptor–Receptor Tyrosine Kinase Heteroreceptor Complexes. Elsevier: 2013; pp 141-164.

184. Hoyne, P. A.; Cosgrove, L. J.; McKern, N. M.; Bentley, J. D.; Ivancic, N.; Elleman, T. C.; Ward, C. W., High affinity insulin binding by soluble insulin receptor extracellular domain fused to a leucine zipper. *FEBS Letters* **2000,** *479* (1-2), 15-18.

185. Berrow, N. S.; Alderton, D.; Sainsbury, S.; Nettleship, J.; Assenberg, R.; Rahman, N.; Stuart, D. I.; Owens, R. J., A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res* **2007,** *35* (6), e45.

186. Aricescu, A. R.; Lu, W.; Jones, E. Y., A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr D Biol Crystallogr* **2006,** *62* (Pt 10), 1243-50.

187. Delle Bovi, R. J.; Miller, W. T., Expression and purification of functional insulin and insulin-like growth factor 1 holoreceptors from mammalian cells. *Anal Biochem* **2017,** *536*, 69-77.

188. Nettleship, J. E.; Rada, H.; Owens, R. J., Overview of a High-Throughput Pipeline for Streamlining the Production of Recombinant Proteins. Springer New York: 2019; pp 33-49.

189. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W., NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **2012,** *9* (7), 671-675.

190. Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T., Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *Journal of Chemical Information and Modeling* **2010,** *50* (4), 572-584.

191. UniProt, C., UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **2021,** *49* (D1), D480-D489.

192. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M., ColabFold: making protein folding accessible to all. *Nature Methods* **2022,** *19* (6), 679-682.

193. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E., UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **2021,** *30* (1), 70-82.

194. SWISS-MODEL Structure Assessment https://swissmodel.expasy.org/ (accessed 13/05/2021).

195. *GraphPad Prism version 9.1.4 for Mac*, Boston, Massachusetts, USA.