A study on the interactions of synthetic IGF-II analogues with the type 1 IGF and insulin receptors

A thesis

submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy in Chemistry

at

The University of Adelaide

School of Chemistry and Physics

by

Jade Misty Cottam

THE UNIVERSITY
of ADELAIDE

May 2014

Table of contents

Abstract

Insulin-like growth factor II (IGF-II) is a unique regulatory peptide containing 67 residues and three disulfide bonds. It binds with high affinity to three receptors, the insulin receptor (IR), the type 1 insulin-like growth factor receptor (IGF-1R) and the type 2 insulin-like growth factor receptor (IGF-2R). Binding of IGF-II to these receptors signals mitogenic responses, such as cell proliferation, differentiation and migration. The interactions of IGF-II with the IR and IGF-1R have recently been identified as potential therapeutic targets for the treatment of cancer. Thus, an increased understanding of the interactions of IGF-II with the IGF-1R and the IR-A is required for the improved design and development of potential anticancer therapeutics.

A crystal structure of IGF-II bound to either the IGF-1R or the IR-A has not been reported. Thus, the precise location of IGF-II within the receptor binding pocket remains undefined. A fluorescence resonance energy transfer (FRET) approach was proposed to investigate the binding location and orientation of IGF-II within the IGF-1R. Two fluorescent IGF-II analogues, the F19Cou IGF-II and F28Cou IGF-II proteins, were synthesised for use in the desired FRET studies.

These FRET experiments first required the synthesis of an appropriate coumarin-based probe for incorporation into IGF-II. The synthesis of a range of fluorescent coumaryl amino acids is described in Chapter 2, and an analysis of the spectroscopic properties of these coumaryl amino acids is also detailed.

Site-specific incorporation of the coumarin-based probe into IGF-II was then undertaken. Three complementary methods were used for the preparation of the desired fluorescent IGF-II analogues. Chapter 3 describes the use of the nonsense suppression methodology for the expression of the novel F19Cou IGF-II protein. This was followed by an improved chemical synthesis of the F19Cou IGF-II protein using a linear solid phase peptide synthesis (SPPS) approach and is detailed in Chapter 4. A robust native chemical ligation approach was developed in Chapter 5, which allowed for the facile incorporation of the coumarin-based

probe at various locations within the IGF-II protein. Chapter 5 also details the synthesis of the native IGF-II, F19Cou IGF-II and F28Cou IGF-II proteins. The biological activity of the resultant IGF-II analogues was evaluated by competition binding assays. The fluorescent IGF-II analogues bind with low nanomolar affinity to the IR and IGF-1R, and as such were deemed suitable for use in the desired FRET-based experiments.

The FRET-based investigation into the binding interactions of the native IGF-II, F19Cou IGF-II and F28Cou IGF-II proteins to the IGF-1R is described in Chapter 6. FRET interactions were observed for both the F19Cou IGF-II and F28Cou IGF-II proteins. The results show the fluorophore binds in close proximity to Trp residues within the IGF-1R receptor and suggest the location of IGF-II binding within the IGF-1R is consistent with what is proposed in the literature. These experiments provide a basis for further investigations for determining the precise binding location and orientation of IGF-II within the IGF-1R.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree. I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Jade M Cottam Date

Publications

Work in this thesis has appeared in the following publication:

Cottam, J.; Scanlon, D.; Karas, J.; Calabrese, A.; Pukala, T.; Forbes, B.; Wallace, J.; Abell, A. *International Journal of Peptide Research and Therapeutics* **2013**, *19*, 61.

Acknowledgements

The Supervisors: I would like to thank *all* my supervisors, official and unofficial. Thank-you, Prof. Andrew Abell for you guidance and support and giving me a scholarship and sticking with me till the end. Thank-you Dr. Denis Scanlon, if thank-you is enough. Since your arrival at The University of Adelaide, my PhD project took a dramatic change of direction, one that I cannot thank-you enough for. You have taught me so many new skills and introduced me to people that I could have only ever imagined to have met. Your direction to my project has been so valuable and has put me in good stead for my future career. Dr. Tara Pukala, thank-you for the support role you have played in supervising my PhD. I appreciate all the conversations we had both productive and "non-productive" and all the MS help and guidance you have given me, it will not be forgotten. Thank-you Assoc. Prof Briony Forbes for your supervision, guidance, assistance, proof-reading and allowing me to work in your laboratory. I would like to thank Prof. John Wallace for initiating this project, which has allowed me to be exposed to many new techniques, people and have unforgettable experiences, which I will carry with forever.

Thank-you to my supervisors abroad: Dr Paul Harris and Prof. Margaret Brimble, thank-you for allowing me to visit your laboratory and use all your facilities. Thank-you Paul for teaching me and sharing all your skills and giving up you time to help me, I learnt so many things. Thank-you to the Brimble peptide lab for having me, helping me with all my questions, for your friendship and providing me with experiences I will never forget.

Friends and Family: I would like to thank my friends and family for the patience, support and help throughout my PhD and all my studies. Especially Seth, you have been a rock for me. You are always willing to listen, offer help where you can and put up with all my nonsense without any question thank-you so much.

The chemists: To Drs Jo, Anton, Courtney and Claire thank-you all for your friendship, I don't know what I would have done without some to vent to, laugh with and drink with. We have created so many memories together and made life-long friendships that I

hope will never be broken. Dr. Anton Calabrese thank-you so much for all your MS help I would never have finished without it. Special mention to Dr. Joanna Duncan and Dr. Scott Walker, thank-you for your patience and help with proof-reading. Thank-you to John Karas, for the synthesis of some IGF-II (31-67) resins and all the additional helpful advice he provided, without it I might not have finished.

The Abell Laboratory, members past and present: Thank-you all for your friendship and help throughout the last few years. We have so many laughs and put all with a lot of nonsense but none the less it has been a good few years.

The Biochemists: The Forbes/Wallace Lab - Thank-you to all the members past and present for all you help and allowing me to annoy you about stupid little things. Thank-you for putting up with a chemist, and teaching me biochemical techniques that will be invaluable for my career. Thank-you to Shee Chee, Carlie and Peter for all your help with my binding assays. Special thanks to Clair Alvino for her friendship and invaluable help and assistance through the course of my PhD research, I would not have finished without you.

Finally, I would like to acknowledge that work carried out in Chapter 3 was done in collaboration with Ms Clair Alvino, Assoc. Prof. Briony Forbes and Prof. John Wallace, in the Forbes/Wallace laboratory in the School of Molecular and Biomedical Sciences, at The University of Adelaide, South Australia; and the research carried out in Chapter 5 was done in collaboration with Prof. Margaret Brimble and Dr Paul Harris, in the Brimble peptide laboratory in the School of Chemistry at The University of Auckland, New Zealand.

Abbreviations

$[\alpha]^{23}_{D}$	specific rotation at the sodium D line (589 nm) at 23 °C
2-Br-Z	*2*-bromobenzyloxycarbonyl
2-Cl-Z	*2*-chlorobenzyloxycarbonyl
2-HED	2-hydroxyethyl disulfide
2xYT	2x yeast extract and tryptone
4-MeBzl	4-methyl benzyl
aaRS	aminoacyl-tRNA synthetase
Acm	acetamidomethyl
AIM	auto-inducing medium
Ala (A)	alanine
Amp	ampicillin
approx.	approximately
Ar	aromatic
Arg (R)	arginine
Asn (N)	asparagine
Asp (D)	aspartic acid
Bn	benzyl
Boc	*tert*-butoxycarbonyl
Br	broad (spectroscopy)
Bz	benzyl
calcd.	calculated
Cbz	benzyloxycarbonyl
CDI	1,1-carbonyldiimidazole
cDNA	coding DNA
CH_2N_2	diazomethane
conc.	concentrated
Cou	coumarin fluorophore **2.2**
CR	cysteine-rich domain
CT	C-terminal domain
Cys (C)	cysteine
DCM	dichloromethane
DIC	*N*,*N*′-diisopropylcarbodiimide
DIPEA	*N*,*N*-diisopropylethylamine
Dmb	*N*-α-(2,4-dimethoxybenzyl)
DMF	*N*,*N*-dimethylformamide
DNA	deoxyribonucleic acid
Dnp	2,4-dinitrophenyl
DODT	3,6-dioxa-1,8-octanedithiol
DTT	dithiothreitol
E	efficiency of the energy transfer (in FRET)
EDTA	ethylenediaminetetraacetic acid
equiv.	equivalent
ESI-MS	electrospray ionisation mass spectrometry
Et_2O	diethyl ether
EtOAc	ethyl acetate
EuIGF-II	europium labelled IGF-II
Ex11	exon 11
Fmoc	9-fluorenylmethyloxycarbonyl
FnIII	fibronectin type III domain
FRET	fluorescence resonance energy transfer

Gln (Q)	glutamine
Glu (E)	glutamic acid
Gly (G)	glycine
GnHCl	guanidine hydrochloride
h	hour(s)
HATU	2-(7-aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HBTU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
HCTU	2-(6-chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate
HEPES	*N*-2-hydroxyethylpiperazine-*N*-2-ethanesulfonic acid
HF	hydrogen fluoride
His (H)	histidine
HOBt	*N*-hydroxybenzotriazole
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
Hz	hertz (in NMR)
IB	inclusion bodies
ID	insert domain
IGF	insulin-like growth factor
IGF-1R	type 1 insulin-like growth factor receptor
IGF-2R	type 2 insulin-like growth factor receptor
IGFBP	insulin-like growth factor binding Protein
IGF-I	insulin-like growth factor I
IGF-II	insulin-like growth factor II
Ile (I)	isoleucine
IPTG	isopropyl-β-D-thiogalactoside
IR	insulin receptor
IR-A	insulin receptor isoform A
IR-B	insulin receptor isoform B
JM	juxatamembrane
kan	kanamycin
L1	large domain 1
L2	large domain 2
LB	Luria Bertani
LCMS	liquid chromatography mass spectrometry
Leu (L)	leucine
lit.	literature value
Lys (K)	lysine
m/z	mass to charge ratio
Me	methyl
$MeONH_2$ •HCl	methoxyamine hydrochloride
Met (M)	methionine

MHz	megahertz (in NMR)
MIN	minimal medium
min	minute(s)
mp	melting point
MPAA	mercaptophenylacetic acid
mRNA	messenger RNA
NCL	native chemical ligation
NIM	non-inducing medium
NIR	near infrared
NMM	*N*-methylmorpholine
NMP	*N*-methylpyrrolidine
NMR	nuclear magnetic resonance
O-2-Ada	2-adamantyl
o-aaRS	orthogonal aminoacyl-tRNA synthetase
OcHx	cyclohexyl ester
OD_{600nm}	optical density at 600 nm
o-tRNA	orthogonal tRNA
PAL	5-[3,5-dimethoxy-4-(fmoc-aminomethyl)phenoxy]pentanoic acid
PAM	4-hydroxymethylphenylacetamidomethyl
Pbf	2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl
PDB	protein data bank
PEG	polyethylene glycol
pet. spirit	petroleum spirit
PG	unspecified protecting group
Phe (F)	phenylalanine
ppm	parts per million
Pro (P)	proline
r	distance between the donor and acceptor (in FRET)
RF1	release factor 1
RNA	ribonucleic acid
R_0	Förster distance
RP-HPLC	reverse phase high performance liquid chromatography
rt	room temperature
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
semi-prep	semi-preparative
Ser (S)	serine
SPE	solid phase extraction
SPPS	solid phase peptide synthesis
*t*Bu	*tert*-butyl
TCEP	tris(2-carboxyethyl)phosphine hydrochloride
TEMED	*N,N.N,N'N'*-tetramethylethylenediamine
tet	tetracycline
TFA	trifluoroacetic acid

THF	tetrahydrofuran
Thr (T)	threonine
TIPS	triisopropylsilane
TK	tyrosine-kinase domain
TLC	thin-layer chromatography
TM	transmembrane domain
TNBSA	2,4,6-trinitrobenzene sulfonic acid
Tos	tosyl
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
Trp (W)	tryptophan
Trt	trityl
Tyr (Y)	tyrosine
Uaa	unnatural amino acid
UV	ultraviolet
v/v	volume per unit volume
Val (V)	valine
w/v	mass per unit volume
Xaa	amino acid
Xan	xanthyl
αCT	C-terminal region of the α-subunit
λ_{Em}	emission maximum
λ_{Ex}	excitation maximum

Chapter 1

Introduction

1.1 The insulin growth factor (IGF) system

The insulin growth factor (IGF) system is responsible for the regulation of cellular processes such as cell growth, proliferation, differentiation and apoptosis.[1-5] It is also associated with the development and progression of cancer.[6-11] Thus, the IGF system has been identified as a potential target for the development of novel cancer therapeutics.[9-17] Figure 1 depicts the overall IGF system consisting of three structurally related ligands; insulin, the insulin-like growth factors I and II (IGF-I and IGF-II), the cognate receptors, and six high affinity binding proteins (IGFBPs).[1,18]

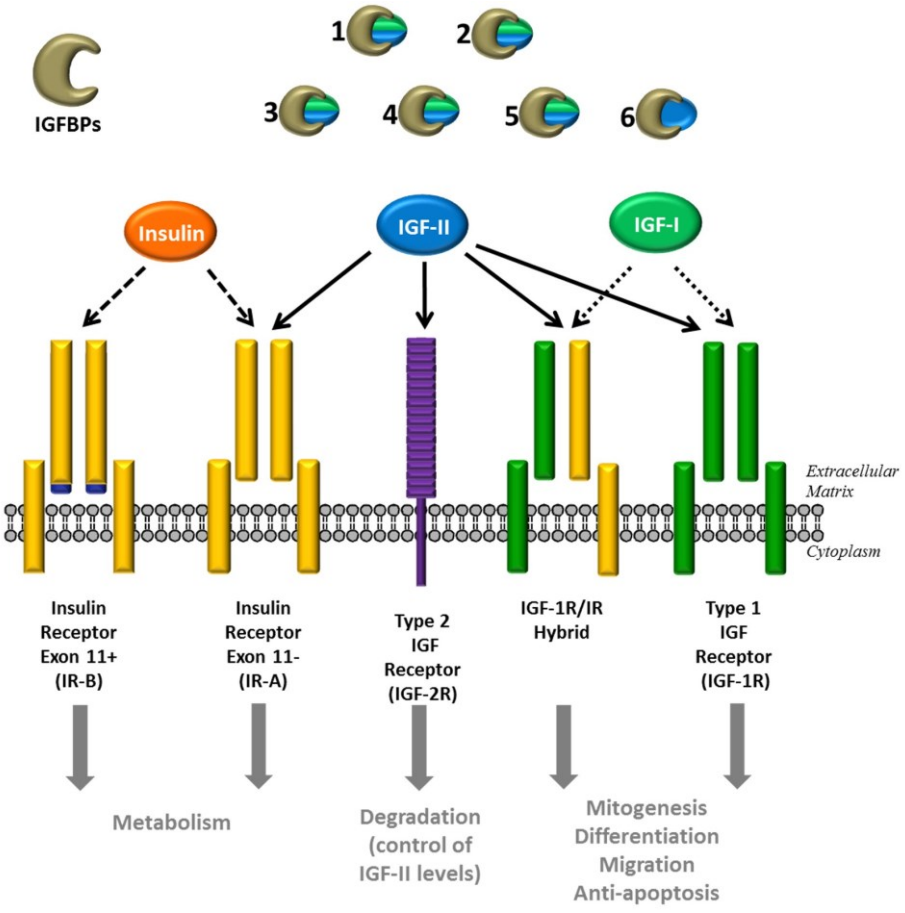


Figure 1: Schematic representation of the IGF system. The IGF system consists of three peptide ligands (insulin (*orange*), IGF-I (*green*) and IGF-II (*blue*)); cognate receptors, IR (two isoforms), IGF-1R, hybrid IR:IGF-R, IGF-2R) and six high affinity binding proteins (IGFBPs (1-6)). The receptors which insulin (*dashed arrows*), IGF-II (*bold arrows*) and IGF-I (*dotted arrows*) bind are represented by *black arrows*. IGFBP's are *numbered* and the proteins to which each bind to are represented by the full or semi-circles coloured in *green* (IGF-I) or *blue* (IGF-II). Figure adapted from Rajapaksha *et al.*[18] with permission from authors.

The biological activity of the IGF's are mediated through interactions with the cell surface membrane receptors, namely, the type 1 insulin-like growth factor receptor (IGF-1R) and the insulin receptor (IR). The insulin receptor (IR) exists in two isoforms that arise from alternative splicing of exon 11 (IR-A (exon 11 -) and IR-B (exon 11 +)), where the IR-A lacks the 12 residues normally encoded for by exon 11.[19-21] The ligands bind to each of these receptors with varying affinities and selectivity, as summarised in Table 1.[22,23] The IR-B, binds insulin with high affinity, the IR-A binds both insulin and IGF-II with strong affinity, while the IGF-1R binds both IGF-I and IGF-II strongly (refer to Table 1).[22,23] The availability of IGF-I and IGF-II to bind these receptors is further controlled through interactions with six high-affinity binding proteins (IGFBP (1-6)) and the type 2 insulin-like growth factor receptor (IGF-2R). The IGF-2R is a mannose 6-phosphate receptor[5,24,25] and is responsible for regulating the level of IGF-II in the cell.[26,27] Finally, the high structural homology between the IGF-1R and IR can result in the hybrid forms of the IGF-1R and IR, to which the IGF's can also bind (refer to Figure 1).[3,28]

Table 1: Summary of the selectivity and binding affinities of the insulin receptor isoforms (IR-A/IR-B) and the IGF-1R.

	IC_{50} (nM)		
	IR-B[a]	IR-A[a]	IGF-1R[b]
Insulin	1.4 ± 0.1	2.8 ± 0.3	> 100
IGF-I	366 ± 15	120.4 ± 34.1	1.4 ± 0.3
IGF-II	68 ± 11	18.2 ± 2.4	3.4 ± 0.2

[a] Data obtained from Denley *et al.*[22]
[b] Data obtained from Delaine *et al.*[23]

Deregulation of the IGF system is thought to be involved in the development and progression of cancer.[6-11] Deregulation is caused by the elevated production of IGF-I and IGF-II ligands, overexpression of the IGF-1R and IR-A, and possible mutation and/or down-regulation of the IGF-2R.[3,6-11] This results in increased binding of the IGFs to the IR-A and IGF-1R, and decreased clearance of IGF-II by the IGF-2R, culminating in increased cancer cell growth and migration.[3,8-10,14,29,30] Furthermore, the IGF system is known to promote the survival,

proliferation and migration of cancer cells, even in the presence of chemotherapeutics.[2,3,8,14,15,31,32]

Several strategies have been explored to inhibit IGF action as a possible treatment of cancer, and include the use of small molecule inhibitors and anti-IGF-1R monoclonal antibodies to prevent ligand binding.[10,12-17] A number of such therapeutics are currently in clinical trials (reviewed in ref. 12-14,17). More recently, the IGF-II:IGF-1R interaction has been identified as potential therapeutic target for the treatment of cancer. This is based on the fact that inhibition of IGF-II binding is known to decrease tumour growth and migration.[9,11,14] IGF-II expression is up-regulated in many cancers, and an overexpression of the IR-A has been observed in a number of different cancers including breast[33-37] and colon[31,35,38-40]. An increased understanding of the interactions of IGF-II with the IGF-1R and the IR-A is required for the improved design and development of potential cancer therapeutics.

1.1.1 Insulin-like growth factor II (IGF-II)

1.1.1.1 Structure of IGF-II

IGF-II is a single chain protein containing 67 residues and three disulfide bonds, and it has high structural and sequence similarity to both insulin and IGF-I (Figure 2).[41-43] It has four structural domains, B, C, A, and D (labelled in order from the N to the C terminus, Figure 2).[41] The A and B domains, corresponding to the A and B chains of insulin, are connected by domain C comprising residues 33-40 in IGF-II (the corresponding C domains of IGF-I and proinsulin contain 12 and 35 residues respectively). The D domain is comprised of the C terminal residues 62-67 in IGF-II and the analogous residues (62-70) in IGF-I.[1,41-43]

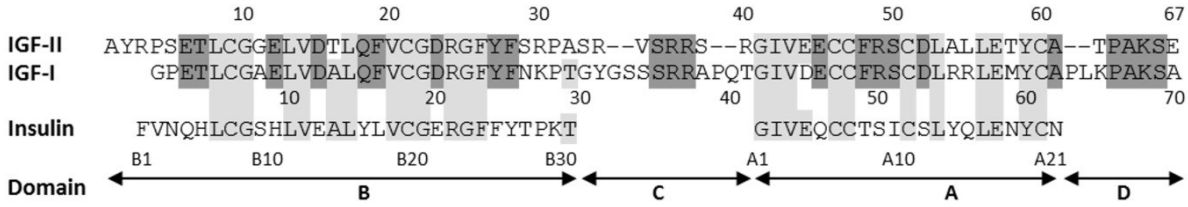


Figure 2: Sequence and domain alignment of insulin, IGF-I and IGF-II.[44] The residues that are conserved between IGF-II, IGF-I and insulin are shaded in *light gray* and conserved residues between IGF-II and IGF-I are shaded in *dark grey*. The domains are indicated with *double headed arrows* below the sequences. Residues numbers are provided above or below each sequence. Figure adapted from Alvino *et al.*[44] with permission from authors.

The primary amino acid sequence of IGF-II was first elucidated by Rinderknecht and Humbel in 1978.[41] The NMR structure of IGF-II was reported in 1994 by Terasawa *et al.*[42] with a higher resolution NMR structure published by Torres *et al.*[43] a year later. Several crystal structures of IGF-II in complex with binding partners, such as the IGF-2R[27] have also been described; however, crystal structures for IGF-II bound to either IGF-1R or IR-A remain elusive.

Depictions of the three-dimensional structures of insulin, IGF-I and IGF-II are shown in Figure 3.[43,45-47] Each ligand contains three α-helices, two in the A domain (residues 42-49 (*pink*) and 53-59 (*red*)), and one in the B domain (residues 12-21 (*blue*)).[i] The helices within the A domain are arranged in an antiparallel fashion, and perpendicular to the B domain α-helix. Together these helices form the basis for the hydrophobic core of the peptide.[48] In contrast, the C and D domains of IGF-II and IGF-I are relatively unstructured, compared to the A and B domains.[42,43]

[i] Residue numbers are for IGF-II

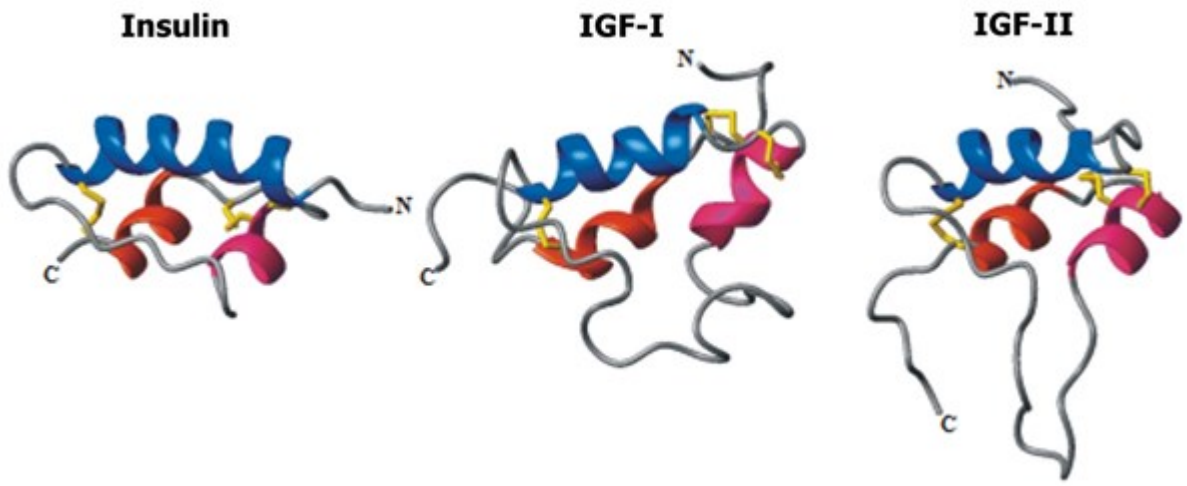


Figure 3: Comparison of the three-dimensional structures of insulin, IGF-I and IGF-II. The structure of insulin[45] (PDB: 1ZNI), IGF-I[47] (PDB: 1BQT) and IGF-II[43] (PDB: 1IGL) are aligned along the B domain helix shown in *blue*. Where the N- and C-termini are denoted as N and C respectively; the first and second A domain helices are shown in *pink* and *red* respectively and the three disulfide bonds are shown in *yellow*. Figure adapted from Denley *et al.*[1] with permission from authors.

1.1.1.2 Binding partners of IGF-II

As mentioned previously, IGF-II is a unique regulatory peptide that binds with high affinity to the IGF-1R, IGF-2R and IR-A.[22,23,35] Unlike IGF-II, insulin and IGF-I only bind with high affinity to the IGF-1R and IR respectively, and are unable to bind (or bind extremely poorly) to other receptors in the IGF system.[22,49] Activation of the IGF-1R and the IR-A by IGF-II results in mitogenic activities, such as cell proliferation, differentiation and migration (refer to Figure 1).[2,22,35,50]

1.1.2 IGF-1R and IR-A

1.1.2.1 Structures of the IGF-1R and IR-A

The IGF-1R and IR-A are transmembrane receptor tyrosine kinases that exhibit a high degree structural similarity.[51,52] The receptors consist of two disulfide linked monomers that associate in a β-α-α-β arrangement as depicted in Figure 4.[51,52] Each monomer contains an α-subunit linked to a β-subunit by a disulfide bond.[52] The α-subunit is approximately 130-135 kDa and

constitutes the extracellular component of the receptors. By comparison, the β-subunit is approximately 90-95 kDa and spans the intra- and extracellular regions.[51-53]

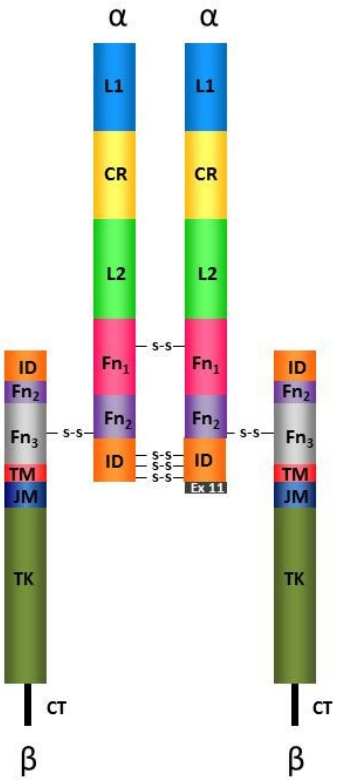


Figure 4: Schematic representation of the domain organisation of the IGF-1R and IR. The IGF-1R and IR share the same homodimer domain structure, and the domains are defined as follows: L1 and L2, large domains 1 and 2 (leucine-rich repeat domains); CR, Cys-rich domain; Fn1, Fn2, Fn3, fibronectin type III domains (FnIII (1-3)); ID, insert domain; Ex11, residues encoded by exon 11 for the IR; TM, transmembrane; JM, juxtamembrane; TK, tyrosine kinase; CT, C-terminal domains and disulfide bonds are shown.

Each monomer of the IGF-1R and IR is comprised of 11 domains as illustrated in Figure 4. The extracellular domain (or ectodomain) consists of two large domains (approx. 150-160 residues) that are rich in leucine repeats (L1 and L2). These are separated by a cysteine rich domain (CR, approx. 150 resides) and followed by three fibronectin type III domains (FnIII-1-3, approx. 100 resides each). The FnIII-2 domain is split over the C-terminal region of the α-subunit (FnIII-2α; αCT) and the N-terminal region of the β-subunit (FnIII-2β) and also comprises a large insert domain (ID, approx. 120 residues). The ID also contains the residues

that encode Exon11 (Ex11), which is present in the IR-B but not the IR-A. Furthermore, the intracellular domain of IGF-1R and IR contain a catalytic tyrosine kinase domain (TK) flanked by two regulatory regions, the juxtamembrane (JM) and the C-terminal or C-tail (CT; 108 residues). Of these domains the L1, CR, L2 and αCT have been shown to contain the main ligand binding determinants for insulin, IGF-I and IGF-II.

A conformational change is induced on the binding of insulin, IGF-I or IGF-II to the extracellular domain of the IGF-1R or IR-A. This results in autophosphorylation of tyrosine residues within the so-called activation loop of the tyrosine kinase domain. This then activates the tyrosine kinase activity of the receptor, which in turn triggers mitogenic responses such as cell proliferation, differentiation and migration.[54-59]

The crystal structures of the L1-CR-L2 domains for the IR-A and the IGF-1R were reported in 1998 and 2006 respectively.[60,61] These structures show that L1-CR-L2 domains for the IR-A and the IGF-1R exist as extended bilobal structures, where the L1-CR-L2 domains are arranged around a central cavity large enough to accommodate the binding ligand. However, L1-CR-L2 constructs are unable to bind the ligand due to the lack of key binding determinants normally present in the intact receptor, namely the C-terminal region of the α-subunit (αCT segment; residues 692-700; refer to Figure 4).[62,63]

A full ectodomain structure for the IR-A was subsequently reported.[64-66] Unlike the L1-CR-L2 construct discussed above, this construct binds native insulin with a K_d of 1 nM.[64] The structure as depicted in Figure 5 displays the two monomers in a folded over conformation with an inverted "V" arrangement.[64] Each arm of the receptor is comprised of the L1, CR and L2 domains from one monomer and the FnIII (1-3) domains, from the other monomer (refer to Figure 5).[64-67]

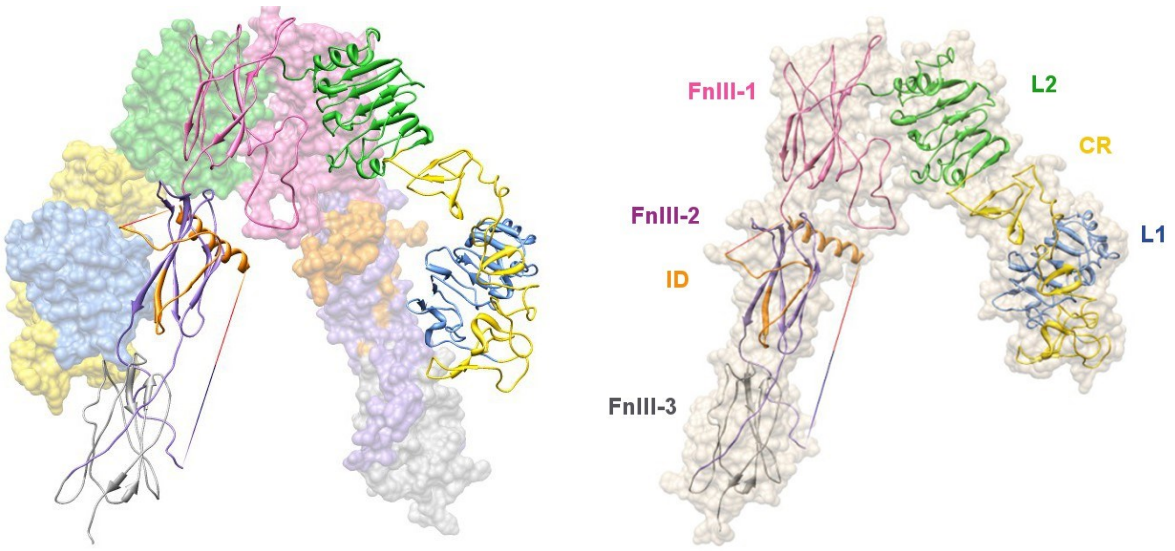


Figure 5: IR-A ectodomain structure.[66] The dimeric folded over conformation of the IR-A ectodomain (PDB: 3LOH), where each arm is formed from the crossover of the two monomers. Monomer 1 is shown in *ribbon structure* and monomer 2 is shown as a *space filling model.* A monomer of the ectodomain of the IR-A showing the domain organisation (*right*). Where domains are defined as follows: L1 and L2, large domains 1 and 2 (leucine-rich repeat domains); CR, Cys-rich domain; FnIII (1-3), fibronectin type III domains (1-3) and ID, insert domain.

Until recently, only crystal structures of the *holo*-ectodomain structure had been reported, where the receptors are not in complex with a ligand.[60,61,64-66] However, Menting *et al.*[68] recently reported a crystal structure of a homodimeric construct of the IR-A in complex with insulin. This construct (L1-CR-L2-(FnIII-1)-αCT(704-719)), consists of the L1-CR-L2 domains connected to the FnIII-1 domain and linked to the αCT segment, and contains all the predicted binding determinants for insulin. This L1-CR-L2-(FnIII-1)-αCT(704-719) construct is reported to bind insulin with a K_d of 17 nM.[68] These results confirmed the importance of the FnIII-1 domain and αCT(704-719) segment for ligand binding.

1.1.3 Binding of IGF-II to the IGF-1R and IR-A

1.1.3.1 Identifying the binding pocket of the IGF-1R and IR-A

To date crystal structures of IGF-II bound to either the IGF-1R or the IR-A have not been reported, and as such little is known about the details of IGF-II binding to these receptors. Thus, an increased understanding of the interactions of IGF-II with its two high affinity receptors is needed. Elucidation of these binding interactions will allow for the better design of anticancer therapeutics.

Two faces define the binding pockets of the IR-A and IGF-1R. The L1, CR and L2 domains of one monomer form one face and the FnIII (1-3) domains together with the αCT segment of the other monomer form the opposing face.[64-67] Two binding sites have been elucidated within the binding pocket and these are depicted in Figure 6. The primary IGF-II binding site, designated site 1, comprises the L1 domain and αCT segment (site 1; Figure 6). IGF-II binding site 2 is less well defined and it is proposed to be formed by the loops at the junction of the FnIII-1 and FnIII-2 domains (site 2; Figure 6).[64-67]

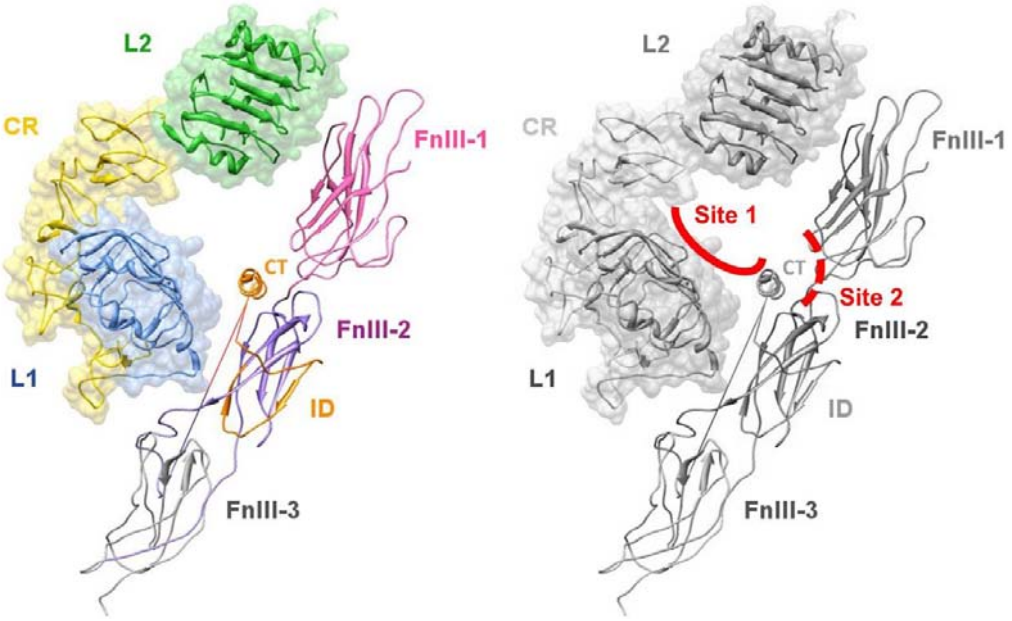


Figure 6: IR-A binding pocket (*left*) and location of the IR-A receptor binding sites (*right*).[66] The IR-A binding pocket (PDB: 3LOH) is formed from the L1, CR and L2 domains of one monomer and the αCT segment and FnIII (1-2) domains of the other monomer. The IR-A binding pocket (*right*), shows the putative location of the two receptor binding sites. Where site 1 is shown by a *solid red line* and site 2 is shown with a *red dashed line*. L1 and L2 refer to the large leucine rich domains 1 and 2; CR is the Cys-rich region; FnIII (1-3) are the type three fibronectin domains 1-3 and ID is the insert domain and contains the C-terminal segment (αCT).

Domains within the IR-A and IGF-1R, which contribute to the high affinity binding of IGF-II, have been identified using a combination of chimeric, hybrid and soluble IGF-1R/IR receptor variants[49,50,63] chemical crosslinking[66,69,70] and alanine scanning mutagenesis.[71] These studies demonstrate that IGF-II does not interact with the CR domain of the IGF-1R, and that a lack of this interaction is a binding determinant unique to IGF-II.[71]

The interactions associated with the binding of insulin and IGF-I to the IR-A and IGF-1R are better understood than those of IGF-II to either the IR-A or IGF-1R.[68,72-77] Since IGF-II exhibits high structural and sequence homology to IGF-I and insulin, it is expected to exhibit a similar mode of binding, and possess similar binding sites on the IGF-1R and IR-A.

1.1.3.2 IGF-1R:IGF-II Interaction

The functional IGF-II binding epitope on the IGF-1R was defined by Sorensen and co-workers using alanine scanning mutagenesis and is depicted in Figure 7.[71] These studies showed that the binding epitope of IGF-II lies primarily on the L1 and αCT for the IGF-1R (refer to Figure 7).[71] Residues in the L1 and αCT segments are thought to indirectly contribute to the high affinity binding of IGF-II by stabilising the interactions between the L1 and αCT segment.[59,66,68]

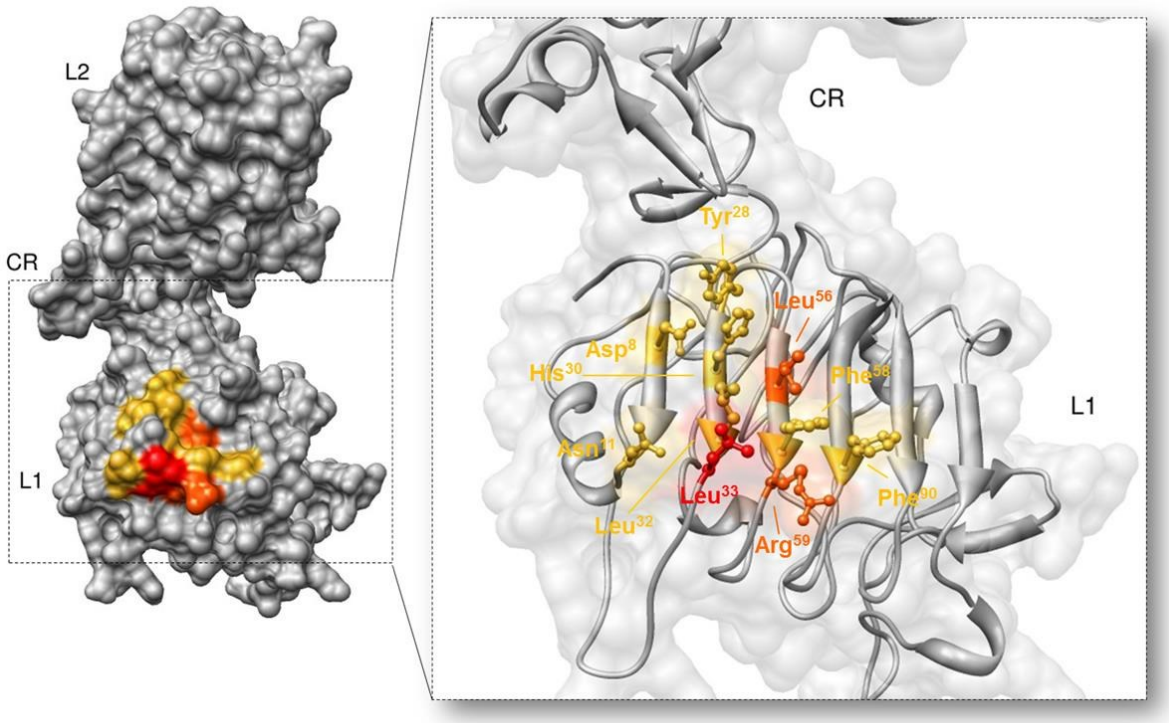


Figure 7: The functional IGF-II site 1 binding epitope on the L1 domain of the IGF-1R as determined by alanine scanning mutagenesis by Sorensen *et al.*[71] Residues identified as important for IGF-II binding to the IGF-1R are shown on the surface structure (*left*) and an expansion of these key residues in the L1 domain are shown as *ribbon* structure (*right*). The binding epitope is mapped on to the L1-CR-L2 crystal structure of the IGF-1R reported by Garret *et al.*[60] (PDB: 1IGR). Where residues in *yellow* demonstrated a 2-5 fold decrease in affinity; residues in *orange* demonstrated a 5-10 fold decrease in affinity and Leu^{33} is shown in *red* and demonstrated a greater than 10-fold decrease in affinity.

Alanine scanning mutagenesis revealed the binding epitope of IGF-II on the IGF-1R is predominantly localised to the second and third β sheets of the L1 domain (refer to Figure 7).[71] Mutation of Asp^{8}, Asn^{11}, Tyr^{28}, His^{30}, Leu^{32}, Phe^{38} and Phe^{90} to alanine, gave analogues of the IGF-1R which had between a 2-5-fold reduction in affinity for IGF-II (yellow; Figure 7). However, recent crystallographic data suggests these residues indirectly contribute to IGF-II binding by stabilising the interactions between the L1 domain and αCT segment.[59,66,68] Furthermore, mutation of Leu^{33}, Leu^{36} and Arg^{39} to alanine gave analogues with an observed 5-fold or greater decrease in affinity for IGF-II (orange and red; Figure 7).[71] From these studies it can be concluded that Leu^{33}, Leu^{36} and Arg^{39} are critical for the binding of IGF-II to IGF-1R.

Residues in the αCT segment of the IGF-1R have also been identified important for IGF-II binding.[59,66,68,71] Specifically, mutation of Phe^{701} to alanine produced an analogue that failed to bind IGF-II, indicating at least a 100-fold decrease in affinity.[71] This suggests Phe^{701} is important for IGF-II binding.[59,66,68] Alanine mutagenesis studies have also identified Phe^{692}, Glu^{693}, Asn^{694}, Phe^{695}, Leu^{696}, His^{697}, Asn^{698} and Ile^{700} as key residues for IGF-II binding.[71] Recent reports suggest that Phe^{692}, Asn^{694}, Phe^{695}, Leu^{696} and Ile^{700} stabilise the L1/αCT interaction, rather than directly engaging with IGF-II.[59,66,68,71] In contrast, the strong sequence and structural homology between the αCT segments of IR-A and IGF-1R indicates that the Glu^{693}, $His^{697}$and Asn^{698} residues are also likely to be involved in IGF-II binding.[59,66,68]

1.1.3.3 IR-A:IGF-II interaction

The IGF-II binding epitope on the IR-A has not been mapped. However, IGF-II binding is proposed to involve interactions at the two sites as depicted in Figure 6.[66] Site 1 is proposed to be formed by the L1 domain and the αCT segment, while site 2 is predicted to involve the loops at the junction of the FnIII-1 and FnIII-2 domains.[19,49,59,66-68,76,78] The high level of sequence and structural homology between both insulin and IGF-II suggests these ligands will bind similarly.[68] Specifically, the residues in the L1 domain of the IR-A, which are suggested important for ligand binding include Asp^{12}, Arg^{14}, Gln^{34}, Leu^{36}, Leu^{37}, Phe^{39}, Phe^{64}, Phe^{88}, Phe^{89}, Asn^{90}, Tyr^{91}, Glu^{120} and Lys^{121}. Of these residues, Gln^{34}, Leu^{36}, Leu^{37}, Phe^{64}, Phe^{88}, Phe^{89} and

Glu[120] are thought to stabilise the interactions between the L1 domain and αCT segment (blue; Figure 8),[59,66,68] whereas, Asp[12], Arg[14], Phe[39], Asn[90], Tyr[91] and Lys[121] are thought to directly contact the ligand (red; Figure 8).[59,66,68]

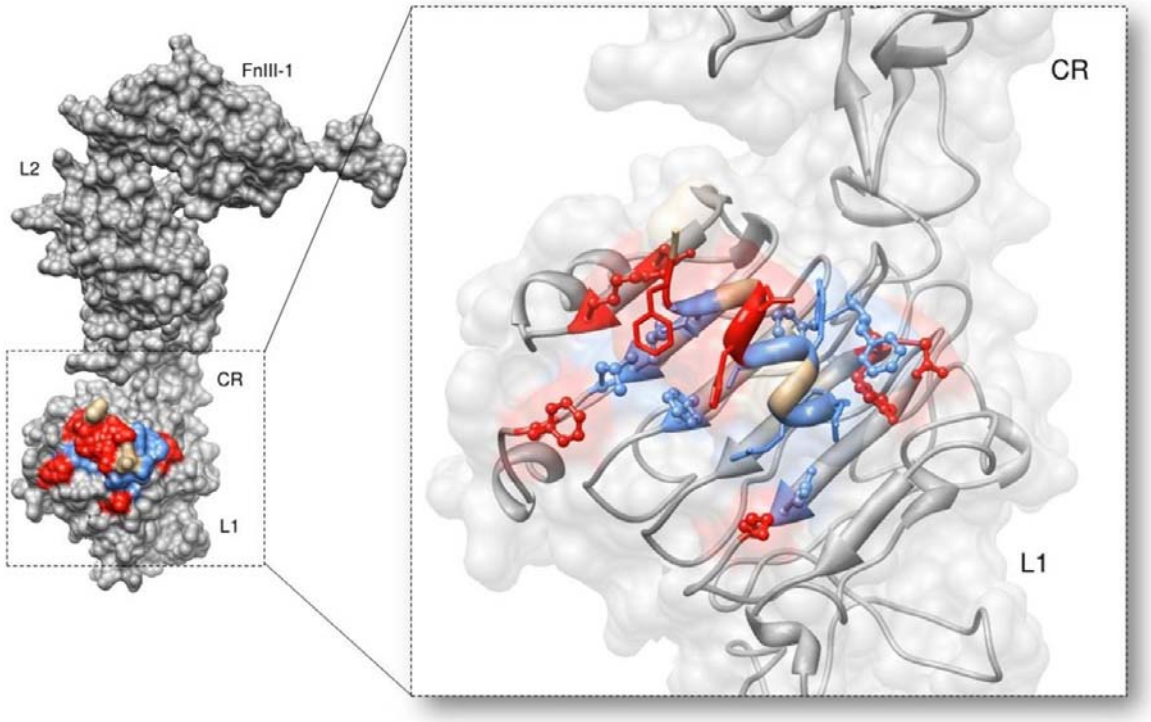


Figure 8: Schematic representation of site 1 binding interactions of IGF-II with the IR-A. Residues proposed important for IGF-II site 1 binding to the IR-A are shown on a *surface* structure (*left*) and an expansion of these key residues in the L1 domain are shown as *ribbon* structure (*right*). Binding interactions are based on those proposed by Lawrence and co-workers[59,66,68] for insulin binding to the IR-A. IGF-II (PDB: 1IGL)[43] is shown in the binding pocket of a homodimeric IR-A construct (IR593αCT), which consists of the L1-CR-L2-FnIII-1 domains linked to the αCT (709-719) fragment (PDB:3W14).[68] In this model IGF-II is superimposed and replacing insulin. The L1-CR-L2-FnIII-1 domains are shown in *grey*; the αCT (709-719) fragment is shown in *tan*; residues shown in *red* are suggested important for ligand binding; residues shown in *blue* are suggested important for the L1/αCT interaction; residues shown in *ball and stick model* are from the L1 domain and residues shown in *stick model* are from the αCT (704-719) segment.

Residues in the αCT segment of the IR-A, Thr[704], Phe[705], Glu[706], Tyr[708], Leu[709], His[710], Asn[711], Val[713] and Phe[714] were identified by alanine mutagenesis as critical for ligand binding.[59,66,68] Of these residues, His[710], Asn[711] and Phe[714] are suggested to directly engage the binding ligand

(red; Figure 8).[59,66,68] In contrast, Thr[704], Phe[705], Glu[706], Tyr[708], Leu[709] and Val[713] are predicted to be involved in the stabilisation of L1/αCT interaction (blue; Figure 8).[59,66,68]

1.1.4 Mapping of receptor binding sites on the IGF-II protein

Identification of the key residues involved in the high affinity binding of IGF-II to the IGF-1R and IR-A would lead to a better understanding of the binding interactions and thus, an opportunity for the improved design of targeted therapeutics.

The C and D domains (residues 33-40 and 62-67) (refer to Figure 2) of IGF-II have been identified using chimeric IGF-I/IGF-II ligands, as important determinants of the selectivity of binding of IGF-II to both the IGF-1R and IR-A.[22] Interestingly, these domains do not contribute to the main IGF-II binding sites. IGF-II's binding sites are the two key locations where IGF-II interacts with the IGF-1R and IR-A and are designated as site 1 and site 2 and are shown in Figure 9.[44,73,79,80]

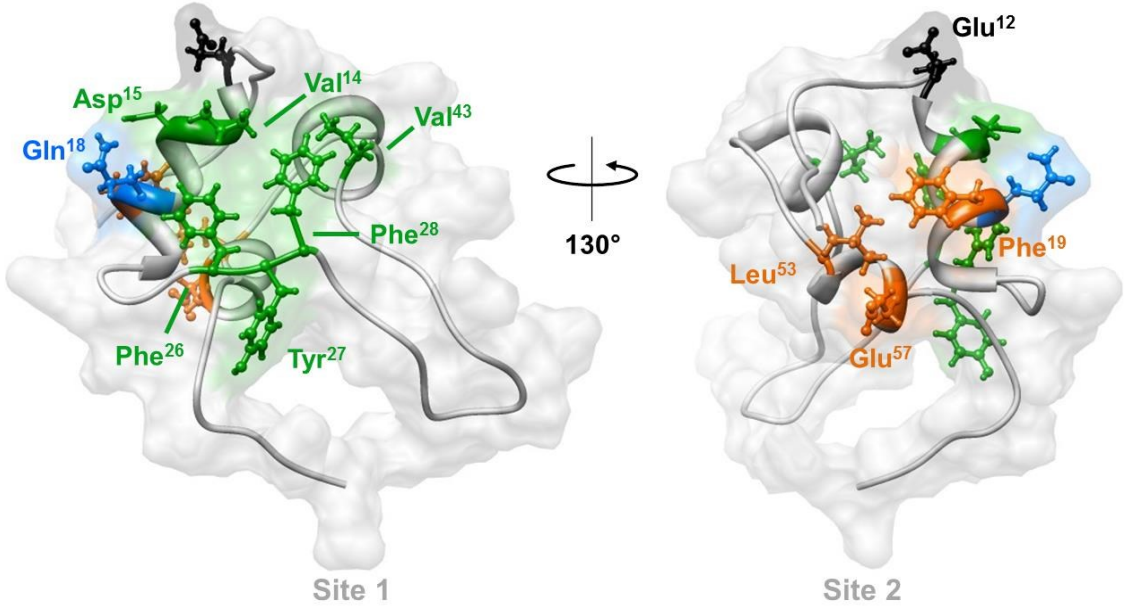


Figure 9: IGF-II binding sites. IGF-II (PDB: 1IGL)[43] is shown in *ribbon* structure and the key residues are shown as *ball and stick model*. Where residues defined as site 1 are shown in *green*; residues defined as site 2 are shown in *orange*; Gln[18] is specific for IGF-1R binding (site 1) and is shown in *blue*; Glu[12] is a site 2 residue which is critical for signalling and activation of the IGF-1R and is shown in *black*.

The key residues that form the IGF-II binding sites have been identified by site directed mutagenesis and through the use of minimised receptor constructs.[44] Site 1 residues are shown in green in Figure 9 and include Val^{14}, Asp^{15}, Phe^{26}, Tyr^{27}, Phe^{28} and Val^{43}. These residues form a hydrophobic cluster defined by the α-helix of the β-chain.[23,44,81-84] Site 1 IGF-II analogues (Val^{14}Thr, Phe^{28}Leu and Val^{43}Met) have been shown to exhibit a minimum of a 2.5-fold decrease in binding affinity for both the IGF-1R and the IR-A.[44] In particular, the binding affinity of a Val^{43}Met IGF-II analogue was less than 1% of the native IGF-II protein.[44] Similarly, a Tyr^{27}Leu IGF-II mutant displayed greater than a 20-fold decrease in binding affinity for the IR-A and more than a 5-fold decrease in affinity for the IGF-1R compared to the native IGF-II protein.[81-84] Mutation of Asp^{15} to alanine, produced an IGF-II analogue (Asp^{15}Ala IGF-II) which had a 10- and 16-fold decrease in binding affinity for the IGF-1R and IR-A respectively.[44] Gln^{18} is shown in Figure 9 in blue, and was identified as important for the selective binding of IGF-II to the IGF-1R.[23,44,81] Specifically, mutation of Gln^{18} to alanine produced an IGF-II analogue (Gln^{18}Ala IGF-II) which had a 2-fold decrease in binding affinity to the IGF-1R, but its binding to the IR-A was unaffected.

Site 2 residues were identified through the use of minimised receptor constructs and alanine mutagenesis.[44] Site 2 residues are highlighted in orange in Figure 9, and include Glu^{12}, Phe^{19}, Leu^{53} and Glu^{57}.[44] These residues are thought to interact with the FnIII domains of the IGF-1R and IR-A (refer to Section 1.1.3). In particular, Glu^{12} is crucial for receptor binding and activation.[85] Specifically, mutation of Glu^{12} to alanine gave an IGF-II analogue (Glu^{12}Ala IGF-II) that failed to activate the IR-A and IGF-1R.[85] In summary, the site 1 (Val^{14}, Asp^{15}, Phe^{26}, Tyr^{27}, Phe^{28} and Val^{43}) and site 2 (Glu^{12}, Phe^{19}, Leu^{53} and Glu^{57}) residues described above form the two primary binding surfaces where IGF-II engages the IGF-1R and IR-A.

The interactions of IGF-II with its two high affinity receptors, the IGF-1R and IR-A have been identified as potential therapeutic targets for the treatment of cancer (refer to Section 1.1.1). The key residues involved in the IGF-II binding sites have been elucidated (refer to Section 1.1.4). However the exact binding locations of IGF-II to the IGF-1R and IR-A still remain elusive (refer to Section 1.1.3). Identification of these interactions is important for the development of anti-cancer therapeutics. Thus, a FRET-based investigation of the binding

interactions between IGF-II and its two high affinity receptors was proposed. This study required the synthesis of the native IGF-II protein and a range of IGF-II analogues with fluorescent probes incorporated at specific sites.

1.2 Protein synthesis

As discussed above, access to IGF-II analogues with fluorescent probes incorporated, would allow a fluorescence resonance energy transfer (FRET) based investigation into the binding of IGF-II to the IGF-1R and IR-A. Several methods exist for the synthesis of such analogues. These techniques are summarised in Figure 10, and are suitable for the current study as they allow for the facile incorporation of fluorescent probes into proteins.[86-99] Both recombinant protein expression (refer to Chapter 3) and chemical synthesis have been used in this thesis to synthesise the native IGF-II protein and non-native IGF-II analogues with fluorescent probes incorporated. Two methods of chemical synthesis were employed and involve the use of SPPS (refer Chapters 4 and 5) and native chemical ligation (refer to Chapter 5).

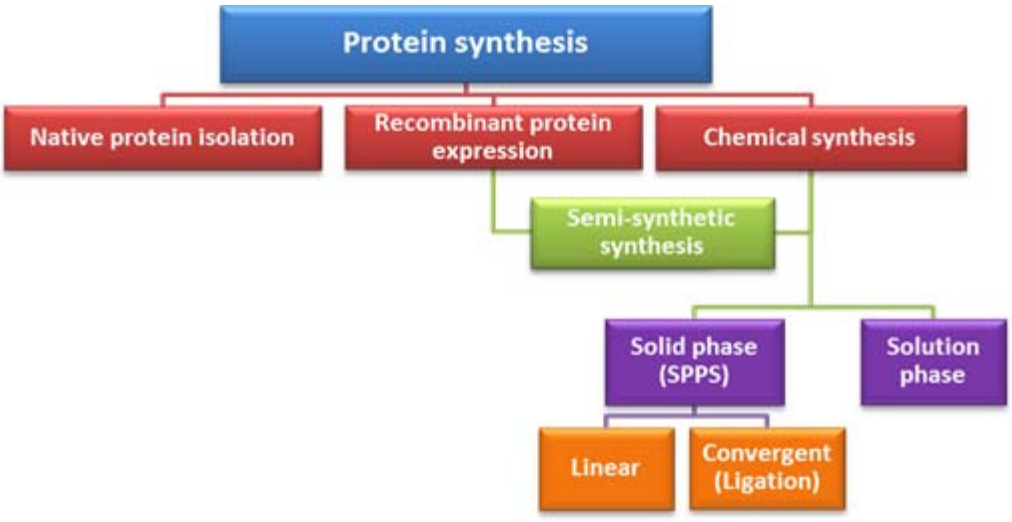


Figure 10: Methods of protein synthesis.[86-99] Three main methods of protein synthesis are summarised; protein isolation, recombinant protein expression and chemical synthesis. Synthetic approaches based on chemical synthesis are the most versatile and include semi-synthetic, solid-phase and solution phase methods.

1.3 Fluorescence resonance energy transfer (FRET)

Protein-protein interactions and the analysis of conformational changes, such as those associated with the binding of IGF-II to the IR-A and IGF-1R, can be investigated using fluorescence resonance energy transfer (FRET).[100-107] FRET is a non-radiative energy transfer from a donor fluorophore to an acceptor fluorophore. The energy transfer process is dependent on the distance between the donor and acceptor fluorophores (10-100 Å), the spectral overlap between the donor emission and acceptor absorption spectra (Figure 11), and the relative orientation of the fluorophores.[105,107-112]

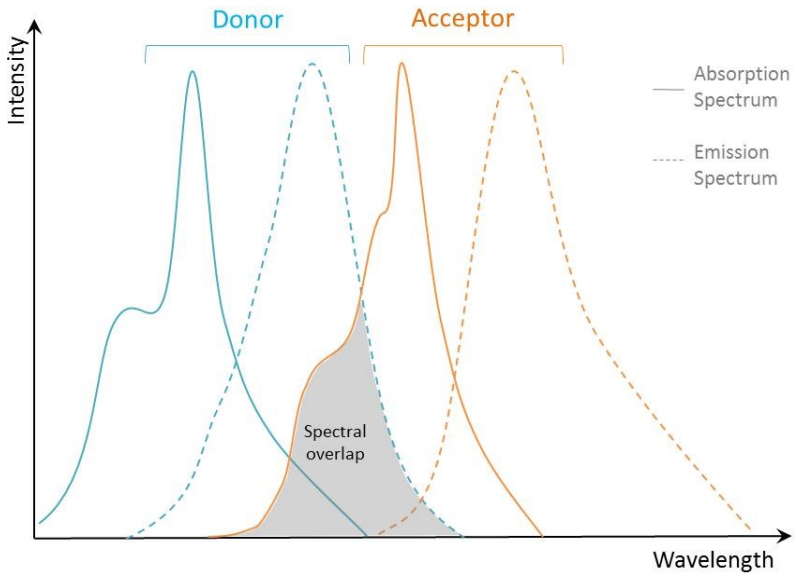


Figure 11: Spectral overlap between a donor emission and acceptor absorption spectra.[108,110] Where the donor spectra are shown in *blue*, the acceptor spectra are shown in *orange*, absorption spectra is designated by a *solid* line and emisson spectra is designated by a *dotted* line and the spectral overlap is shown in *grey*.

FRET can occur in both intra- and intermolecular systems as illustrated in Figure 12. In all cases the occurrence of the FRET interaction is distance dependent and only occurs when the donor and acceptor fluorophores are in close proximity (10-100 Å). A more detailed description of the process is discussed in Chapter 6.

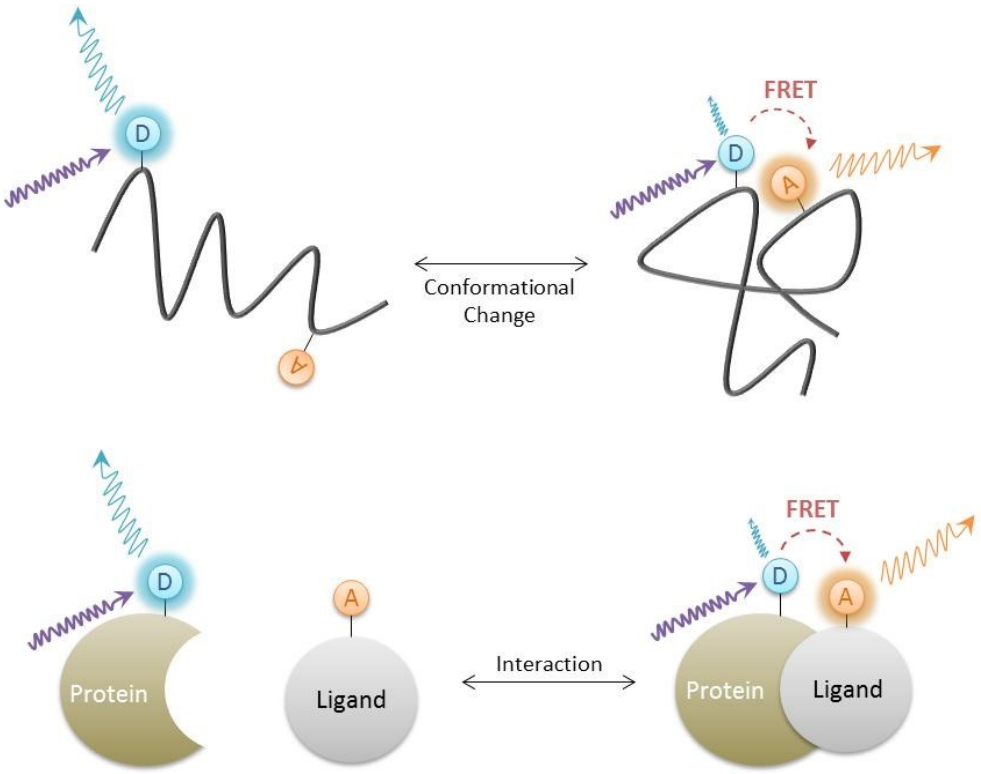


Figure 12: Examples of intramolecular and intermolecular FRET. Intramolecular FRET (*above*) and intermolecular FRET (*below*) interactions are shown. Where D is a donor fluorophore; A is an acceptor fluorophore and zig-zag arrows in *purple*, *blue* and *orange* indicate the excitation light, donor emission and acceptor emission respectively. Figure adapted from Miyawaki.[106]

As discussed in Section 1.1.3 and illustrated in Figure 13, IGF-II interacts with two sites on the IGF-1R and IR-A. To investigate the binding of IGF-II to these receptors using FRET, the selection of an appropriate FRET donor and acceptor pair, which could be positioned adjacent to or within these key binding sites was required (refer to Section 1.1.4).

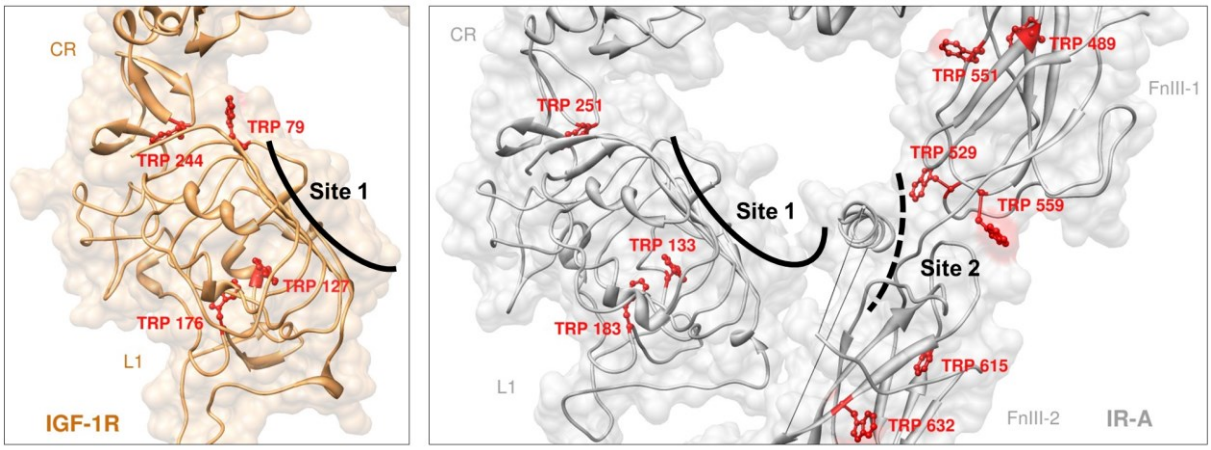


Figure 13: Location of Trp residues adjacent to the IGF-1R and IR-A binding sites. Trp residues adjacent to the IGF-1R (PDB:1IGR)[60] binding site 1 (*brown, left*), and the IR-A (PDB:3LOH)[66] binding pocket (*grey, right*) are shown. Where site 1 is designated by a *solid black line* and site 2 is designated with a *black dashed line* and the location of nearby tryptophan residues are *numbered* and shown in *red* as *ball and stick model*. L1 refers to the large leucine rich domain 1; CR is the Cys-rich region and FnIII 1 and 2 are the type three fibronectin domains 1 and 2.

Of all the proteinogenic amino acids, Trp has the most suitable spectroscopic properties for use in FRET. This is discussed in detail in Chapter 2, but briefly Trp has stronger fluorescence with higher quantum yields than the other proteinogenic amino acids. Fortunately, both the IGF-1R and IR-A contain an abundance of naturally occurring tryptophan residues which can behave as FRET donors. Furthermore, several of these Trp residues are predicted to be positioned either within the receptor binding sites, or closely adjacent. These residues are illustrated in Figure 13, and include Trp^{79}, Trp^{127}, Trp^{176} and Trp^{244} (site 1) for the IGF-1R, and Trp^{133}, Trp^{183} and Trp^{251} (site 1) and Trp^{489}, Trp^{551}, Trp^{529}, Trp^{559}, Trp^{615} and Trp^{632} (site 2) for the IR-A.

While the abundance of Trp residues within both the IGF-1R and IR-A increases the possibility of a FRET interaction, it also complicates the identification of the precise FRET partner. Thus, the work in this thesis endeavours to establish the feasibility of a FRET-based approach in providing information about the binding location of IGF-II within the IGF-1R and IR-A. These preliminary studies will first involve the detection of a FRET interaction between the donor and Trp residues within these receptors, and will provide the basis for more comprehensive FRET studies.

1.4 Scope of this thesis

The key residues that form the IGF-II binding sites have been elucidated. However, the absence of a crystal structure of IGF-II bound to either the IGF-1R or the IR-A means the precise location of IGF-II within the receptor binding pocket remains undefined. The work in this thesis describes the synthesis of the F19Cou IGF-II and F28Cou IGF-II analogues which have fluorescent coumarin-based probes incorporated. These analogues were used in a FRET study, in order to gain an increased understanding of the binding interactions of IGF-II with its two high affinity receptors, the IGF-1R and IR-A.

Chapter 2 describes the synthesis of a range of fluorescent coumarin-based probes, bearing traditional proteinogenic handles, and includes the analysis of the spectrochemical properties of these probes. The site-specific incorporation of a fluorescent coumarin-based probe into the IGF-II protein is then described. Three complementary methods were used for the preparation of IGF-II and its fluorescent analogues. First, the expression of the F19Cou IGF-II analogue using recombinant protein expression is discussed in Chapter 3. An improved chemical synthesis of the fluorescent F19Cou IGF-II protein using a linear solid phase peptide synthesis (SPPS) approach is detailed in Chapter 4. Chapter 5 describes the development of a robust native chemical ligation approach for incorporation of the coumarin-based probe at a various locations within the IGF-II protein, and the application of this approach in the synthesis of the native IGF-II protein and two fluorescent analogues, F19Cou IGF-II and F28Cou IGF-II proteins is also detailed in Chapter 5. Finally, Chapter 6 describes the preliminary FRET-based investigation into the binding interactions of the F19Cou IGF-II and F28Cou IGF-II proteins with a soluble form of the IGF-1R.

Chapter 2

Synthesis and characterisation of coumaryl amino acids

2.1 Introduction

As outlined in Chapter 1, the proposed FRET-based investigation into the binding interactions of IGF-II with the IGF-1R and IR-A required the synthesis of IGF-II analogues with an appropriate acceptor fluorophore incorporated.

2.1.1 Fluorophores

The choice of fluorophore was dictated in part by its spectrochemical properties, which for the current study a strong quantum yield (ϕ) and large Stokes shift were preferred. Other key properties include the absorption (λ_{Ex}) and emission (λ_{Em}) maxima, molar extinction coefficients (ε) and fluorescence lifetimes (τ). The molar extinction coefficient is a measure of absorptivity of a molecule at a given wavelength, while the fluorescence lifetime refers to the time that a molecule spends in the excited state before returning to its ground state. Quantum yield (ϕ) is a measure of the fluorescence efficiency of a molecule and is defined as the ratio of photons fluoresced to photons absorbed. Stokes shift is a measure of the overlap between the excitation and emission spectra and is defined as the difference (in nm) between the absorption and emission maxima (Figure 14).[110,113] These spectral properties in addition to the chemical structure and size of the fluorophore, determine its application.[114]

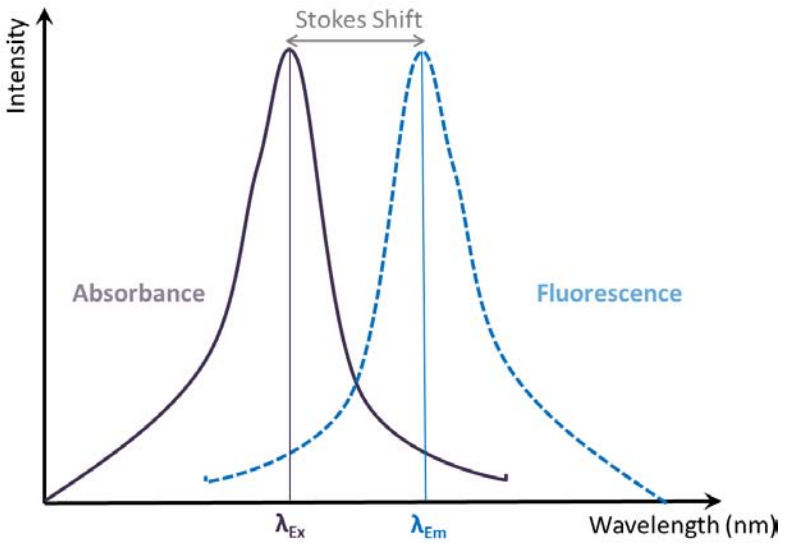


Figure 14: Stokes shift. Absorbance spectrum is indicated with a *solid purple* line and the peak maximum is identified by λ_{Ex}. Fluorescence spectrum is designated by a *dotted blue* line and the peak maximum is identified by λ_{Em}. Stokes shift is designated by a *grey double headed arrow* and is defined as the difference (in nm) between the λ_{Em} and λ_{Ex}. Figure adpated from Lavis *et al.*[113]

Fluorophores have found wide use in the investigation of biological systems as enzyme substrates, cellular stains, environmental indicators and molecular probes.[113] A fluorophore can be classified as intrinsic, extrinsic covalently bound, or extrinsic associating based on the nature of its interaction with a target protein.[115,116]

Endogenous fluorophores Phe, Tyr and Trp are examples of intrinsic probes. These probes can be used without manipulation of the protein, and as such do not disrupt the native protein structure or any of its interactions. However Phe and Tyr are not strongly fluorescent. Consequently, the fluorescence of these endogenous fluorophores is quenched or masked by the natural fluorescence of the protein or biological system.[110,117] In contrast, Trp exhibits stronger fluorescence and higher quantum yields. However the use of Trp as an endogenous fluorophore is limited by its low natural abundance in proteins.[117] The undesirable spectrochemical properties and low abundance of intrinsic probes has led to the development and use of extrinsic probes.[118]

Extrinsic covalently bound probes are covalently linked to a protein at a defined location, and consequently are often used to investigate site-specific interactions.[116] Conversely, extrinsic

associating probes are not covalently linked to the protein and associated through non-covalent interactions such as hydrophobic/hydrophobic or hydrophilic/hydrophilic interactions. Consequently the information gained from FRET studies using extrinsic associating probes is representative of the surrounding microenvironment rather than site-specific information.[115,116] In this work an extrinsic covalently bound probe was utilised, as site-specific information about the binding of IGF-II was desired.

Small molecular probes (< 1 kDa) are less likely to perturb the structure and function of proteins, thus are preferred for investigating protein-protein interactions. Such probes have a diverse and tuneable range of spectroscopic properties and are also easily incorporated or attached to proteins.[113,118-121]

2.1.2 Small organic fluorophores

Small molecular probes are generally organic fluorophores derived from a narrow range of "core" organic dye structures. Synthetic manipulation of these core structures allows the spectroscopic properties of the probe to be tuned to provide an array of probes for use in a variety of applications.[111,113,122]

Coumarins, quinolines, xanthenes, pyrenes, boron-dipyrromethenes (BODIPY) and cyanines are some of the most common classes of fluorescent organic dyes used in chemistry and biology (Figure 15). These dyes possess a wide range of spectrochemical properties spanning UV, visible and IR wavelengths.

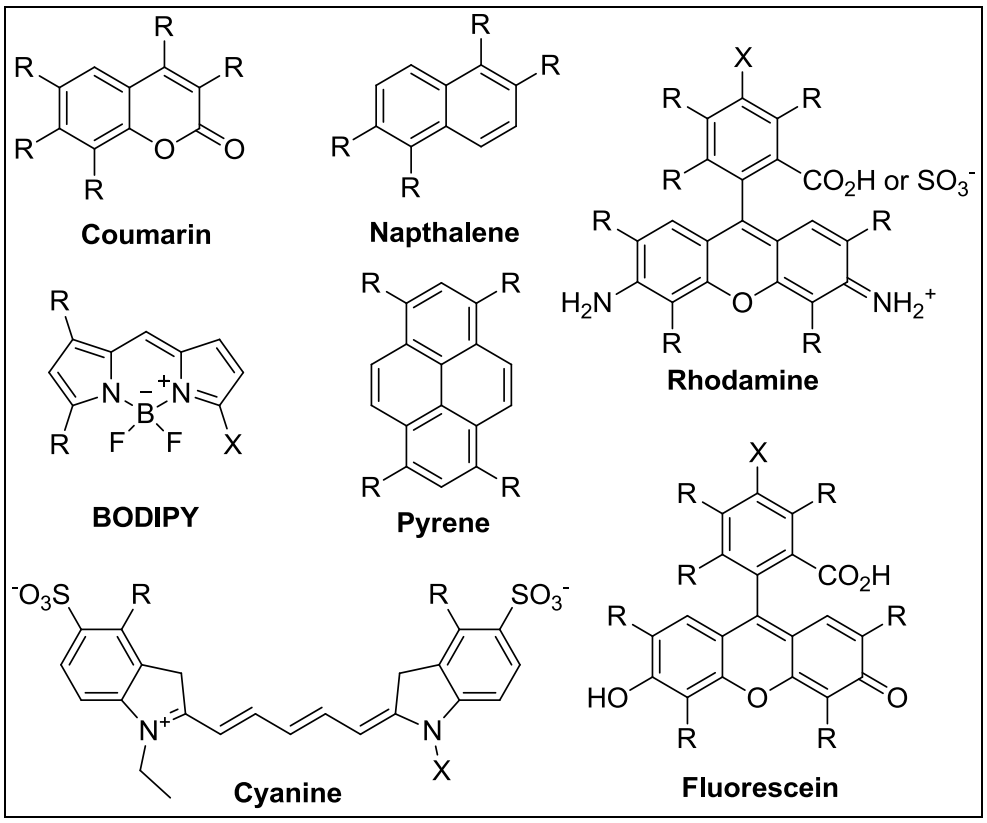


Figure 15: Core structures of common fluorescent dyes. Common sites for substitution are annotated by an R and the common site of conjugation is annotated by X (where known). Figure adapted from Sapsford *et al.*[111]

2.1.3 Coumarins

Coumarins are ideal for use as small molecular probes.[118,122-125] In general, coumarins exhibit high quantum yields, low photobleaching and large Stokes shifts, and have spectral properties which are sensitive to changes in pH and polarity.[110,122,123,126,127] Furthermore coumarins are relatively easily synthesised by the Pechmann[128,129], Perkin[130] and Knoevenagel[131] reactions. These methods allow incorporation of a variety of functional groups such as acids, amines, maleimides, alkynes or azides, which allow the coumarin moiety to be easily conjugated to a protein.[107,111,122] When incorporated into peptides and proteins, coumarins have been used to monitor site-specific conformational changes, protein-protein interactions and probe the molecular microenvironment.[132-137] For these reasons a coumarin-based fluorophore was chosen for incorporation into IGF-II, allowing the binding of IGF-II to the IR-A and IGF-1R to be investigated using FRET.

A fluorophore can simply be conjugated to, or incorporated into a protein by attaching it to either the N or C terminus of a protein. However, this can have significant drawbacks, as the termini of a protein are often involved in recognition or interactions with binding partners. Furthermore, the termini of a protein can be far removed from the area of interest, particularly the binding site. With this in mind, a coumarin-based fluorophore was incorporated into IGF-II as the side chain of a residue adjacent to the IGF-II binding sites (refer to Section 1.1.4, Chapter 1). Thus, allowing more site-specific information about the binding of IGF-II to its high affinity receptors (IGF-1R and IR-A) to be attained.

Incorporation of a coumarin fluorophore into IGF-II required the synthesis of a coumarin moiety with a suitable handle. Coumarin-based amino acids have been reported by Brun *et al.*[138] and Wang *et al.*[134] These amino acids are easily synthesised in three steps from commercially available starting materials and display desirable spectrochemical properties, including strong quantum yields (ϕ) and large Stokes shifts. Crucially, the basic amino acid structure allows the coumarin to be easily incorporated into a protein using recombinant protein expression or SPPS. With this in mind, a series of coumaryl amino acids were synthesised for incorporation into IGF-II and are shown in Figure 16.

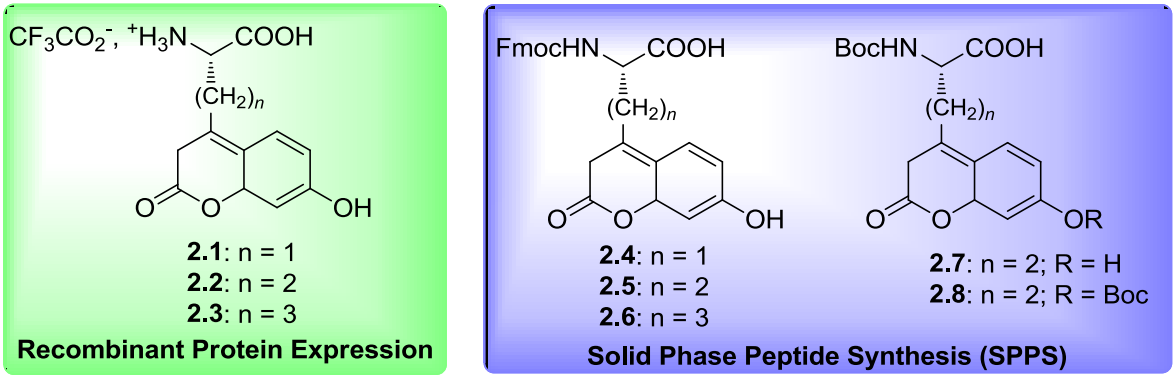
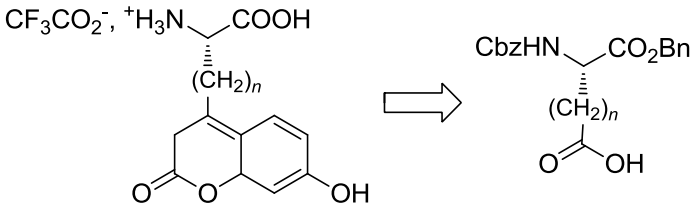


Figure 16: Fluorescent coumaryl amino acids synthesised for use in recombinant protein expression and solid phase peptide synthesis (SPPS). Where Boc = *tert*-butoxycarbonyl; Fmoc = 9-Fluorenylmethyloxycarbonyl and $CF_3CO_2^-$ = trifluoroacetate salt

2.2 Results and discussion

2.2.1 Synthesis of fluorescent coumaryl amino acids (2.1-2.3)

Fluorescent coumaryl amino acids **2.1** and **2.2** were synthesised as shown in Scheme 1, from aspartic acid (n=1) (**2.9**) and glutamic acid (n=2) (**2.10**) using methods described by Brun *et al.*[138] and Wang *et al.*[134] Furthermore adaptation of these procedures allowed the synthesis of the previously unreported coumaryl amino acid **2.3** from homoglutamic acid (n=3) (**2.11**).

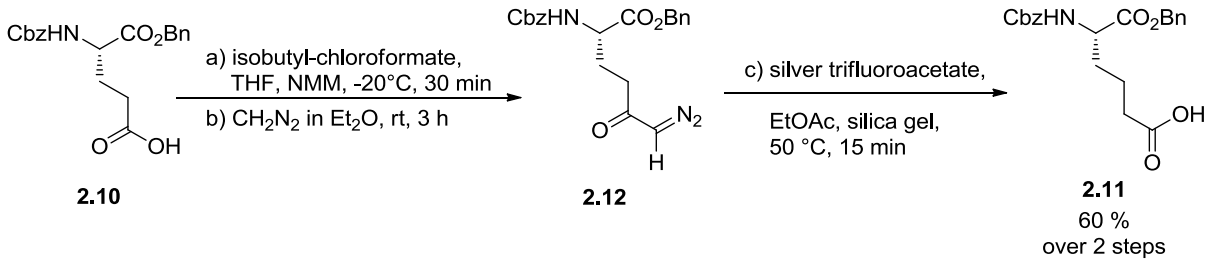


2.1: n = 1
2.2: n = 2
2.3: n = 3

2.9: n = 1
2.10: n = 2
2.11: n = 3

Scheme 1: Strategy for the synthesis of fluorescent coumaryl amino acids 2.1, 2.2 and 2.3. Coumaryl amino acids (**2.1-2.3**) were derived from aspartic (n=1) (**2.9**), glutamic (n=2) (**2.10**) and homoglutamic (n=3) (**2.11**) acids. Where Cbz = benzyloxycarbonyl, Bn = benzyl and $CF_3CO_2^-$ = trifluoroacetate salt.

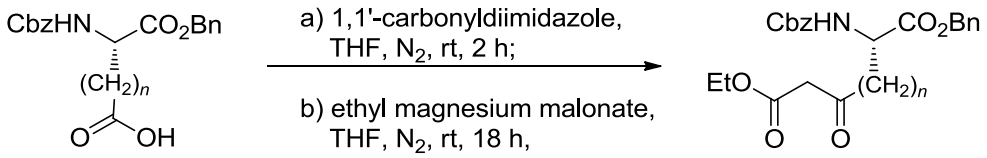
The key di-protected homoglutamic acid derivative, Cbz-HGlu-OBn (**2.11**) was readily prepared from the commercially available di-protected glutamic acid (Cbz-Glu-OBn) (**2.10**), via an *Arndt-Eistert* Homologation sequence.[139,140] As outlined in Scheme 2, **2.10** was reacted with isobutylchloroformate to give the corresponding mixed anhydride that was immediately reacted with an ethereal solution of diazomethane to give diazoketone **2.12**.[141] Subsequent Wolff rearrangement[142,143] on reaction with silver trifluoroacetate, in the presence of silica gel at 50 °C, under reduced pressure for 15 min, gave the desired homologated product (**2.11**). This material was obtained in a 60% yield over two steps, after purification by silica-based column chromatography.



2.10 **2.12** **2.11**
60 %
over 2 steps

Scheme 2: Synthesis of the di-protected homoglutamic acid derivative (2.11) via an *Arndt-Eistert* homologation reaction. Where NMM = *N*-methylmorpholine, Cbz = benzyloxycarbonyl and Bn = benzyl.

The desired β-ketoesters (**2.13-2.15**) were synthesised as depicted in Scheme 3, from the di-protected L-amino acids (**2.9-2.11**) *via* an acylation reaction.[144] The carboxylic acids were separately reacted with 1,1'-carbonyldiimidazole followed by treatment with an ethyl magnesium malonate salt to give the side chain β-ketoester derivatives (**2.13-2.15**) in high yields (84-95%) after purification by silica-based column chromatography.



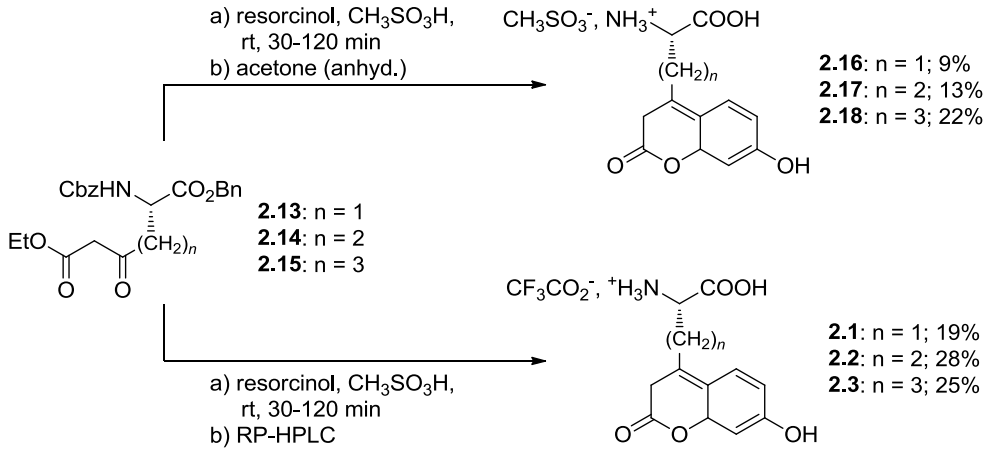
2.9: n = 1
2.10: n = 2
2.11: n = 3

2.13: n = 1; (93%)
2.14: n = 2; (95%)
2.15: n = 3; (84%)

Scheme 3: Synthesis of β-ketoesters 2.13, 2.14 and 2.15. Where Cbz = benzyloxycarbonyl and Bn = benzyl.

Coumaryl amino acids **2.16**, **2.17** and **2.18** were synthesised as shown in Scheme 4, following the methods outlined by Brun *et al.*[138] and Wang *et al.*[134] The β-ketoesters **2.13**, **2.14** and **2.15** were individually reacted with resorcinol in the presence of methanesulfonic acid for 30, 120 and 90 min respectively. The reactions were quenched by precipitation with cold ether, the precipitate was isolated by vacuum filtration and the crude residue was dissolved in water and lyophilised. These conditions gave concomitant formation of the fluorescent 7-

hydroxycoumarin moiety and deprotection of the Cbz and Bn protecting groups, to afford the previously reported coumaryl amino acids **2.16** and **2.17**, and an unreported coumaryl amino acid **2.18** in isolated yields of 9%, 13% and 22% after purification as described below.



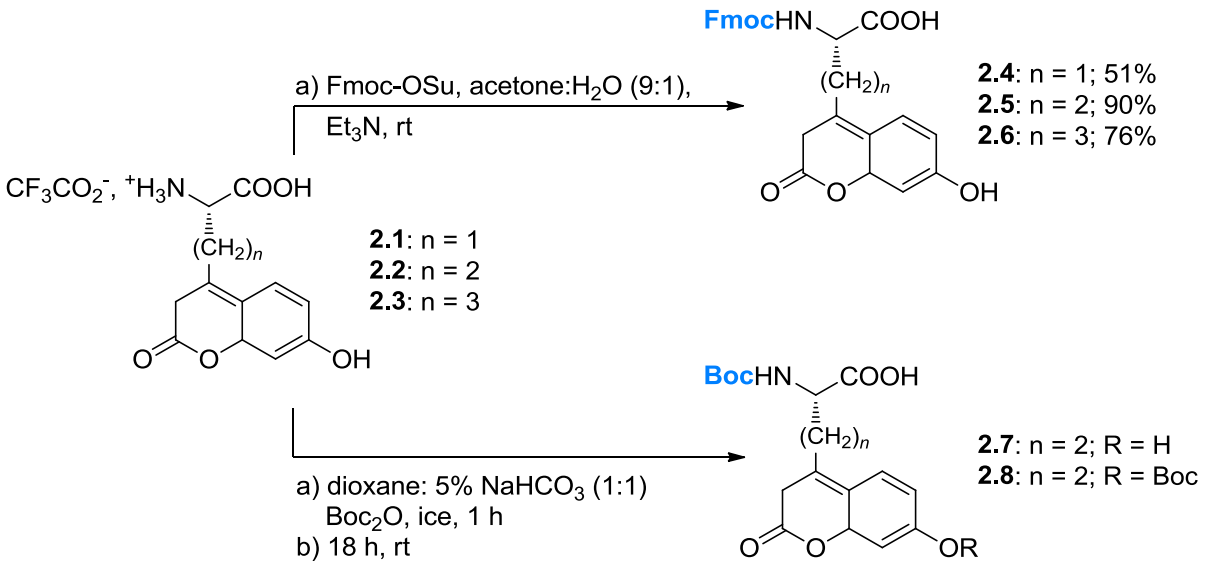
Scheme 4: Synthesis of the fluorescent coumaryl amino acids (2.1-2.3 and 2.16-2.18). Purification by semi-preparative RP-HPLC, was performed on a Phenomenex Luna C18 (10 µm, 50 x 10 mm, 100 Å) column with a linear gradient (1% D/min) (where the buffers were A (MilliQ water + 0.1% TFA) and D (acetonitrile + 0.08% TFA)). Where Cbz = benzyloxycarbonyl, Bn = benzyl, TFA = trifluoroacetic acid, $CH_3SO_3^-$ = methansulfonate salt and $CF_3CO_2^-$ = trifluoroacetate salt

The coumaryl amino acids **2.16-2.18** were initially purified by precipitation with anhydrous acetone.[ii] This afforded coumaryl amino acids **2.16, 2.17** and **2.18** as methanesulfonate salts in high purity. An alternative semi-preparative RP-HPLC based purification of crude **2.16, 2.17** and **2.18** gave coumaryl amino acids **2.1, 2.2** and **2.3** as TFA salts in improved yields of 19, 28 and 25% (refer to Scheme 4). Semi-preparative RP-HPLC was the preferred method of purification for coumaryl amino acids **2.1-2.3**, due to the ease of isolation and improved yields relative to the precipitation method.

[ii] Semi-preparative RP-HPLC facilities were not available at the time.

2.2.2 Synthesis of *N*α-protected coumaryl amino acids (2.4-2.8) for use in solid phase peptide synthesis (SPPS)

The *N*α-protected coumaryl amino acids **2.4**-**2.8** for use in SPPS were synthesised as shown in Scheme 5. Specifically, the Fmoc-protected coumaryl amino acid **2.4** was afforded from the treatment of a solution of **2.1** in aqueous acetone, with triethylamine and 9-fluorenylmethyloxycarbonyl succinimide (Fmoc-OSu). This gave the desired *N*α-Fmoc protected coumaryl amino acid **2.4** in a moderate yield (51%). Similarly, coumaryl amino acids **2.2** and **2.3** were subjected to *N*α-Fmoc protection, which afforded **2.5** and an unreported Fmoc-protected coumaryl amino acid **2.6** in improved yields of 90% and 76% respectively.



Scheme 5: Synthesis of *N*α-protected coumaryl amino acids (2.4-2.8) for use in SPPS. Where Boc = *tert*-butoxycarbonyl, Fmoc = 9-Fluorenylmethyloxycarbonyl and $CF_3CO_2^-$ = trifluoroacetate salt.

Treatment of **2.2** with aqueous sodium bicarbonate and di-*tert*-butyl dicarbonate (Boc_2O) gave a mixture of the mono-Boc protected (**2.7**) and the di-Boc protected (**2.8**) coumaryl amino acids in a ratio of 10:1 as determined by ^{1}H NMR. This material was acceptable for use in SPPS as the phenolic Boc group is concomitantly removed during the *N*α-Boc deprotection

using TFA (refer to Chapter 5). Coumaryl amino acids **2.1** and **2.3** were not used in Boc-SPPS, thus *N*-Boc protection was not undertaken.

2.2.3 Spectroscopic characterisation of coumaryl amino acids 2.1-2.3 and 2.16-2.18

As discussed in Section 2.1.3, coumaryl amino acids **2.1-2.3** were synthesised in anticipation of obtaining a fluorophore with a strong quantum yield (ϕ) and large Stokes shift, as these properties are important for the proposed FRET study. Thus, the spectroscopic properties of coumaryl amino acids **2.1-2.3** and **2.16-2.18** were analysed and compared with literature values.

The absorption (λ_{Ex}) and emission (λ_{Em}) maxima, molar extinction coefficients (ε), Stokes shift and quantum yields (ϕ) of synthetic coumaryl amino acids **2.1-2.3** were determined and the results are summarised in Table 2. A comparison of the absorbance and fluorescence spectra of these coumaryl amino acids (**2.1-2.3**) at a fixed concentration of 50 μM is also shown in Figure 17.

Table 2: Physical and spectrochemical properties of coumaryl amino acids 2.1, 2.2 and 2.3. All data was obtained at a concentration of 50 μM.

	n	$[\alpha]^{23}_{D}$ [a]	λ_{Ex} [b] (nm)	Abs [b] (a.u)	ε [b] (L mol^{-1} cm^{-1})	λ_{Em} [b] (nm)	F.I [b]	Stokes shift (nm)	ϕ [b],[c]
Tyr [d]	-	-	275		-	304	-	29	0.14
Trp [d]	-	-	295	-	-	353	-	58	0.13
2.1	*1*	17.25	328	0.053	10700	480	173	152	0.10
2.2	*2*	24.18	323	0.059	11700	485	279	162	0.12
2.3	*3*	13.80	323	0.072	14500	491	404	168	0.15

[a](c. 1.0, 1M HCl);[b] determined in 1M HCl; [c] Quinine sulfate used as a standard;
[d] Data taken from Lakowicz *et al.*[110] and determined at a pH of 7.2

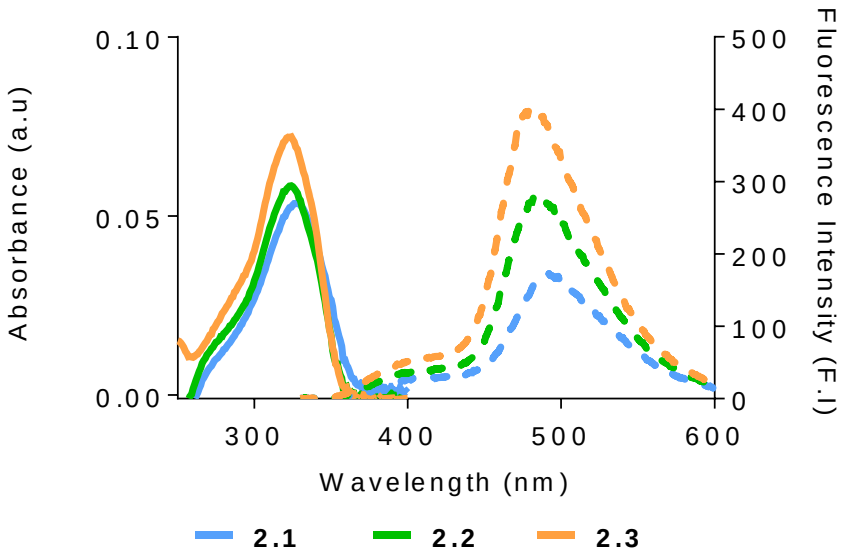


Figure 17: Comparison of the absorbance and fluorescence spectra of coumaryl amino acids 2.1, 2.2 and 2.3. Where the absorbance spectra are represented by *solid lines*; the fluorescence spectra are represented by *dashed lines* and spectra were collected at a concentration of 50 µM in 1 M HCl.

Each of the coumaryl amino acids **2.1**, **2.2** and **2.3** displayed large molar extinction coefficients and strong fluorescence emission maxima (λ_{Em}) and fluorescence intensities (F.I). The quantum yields for compounds **2.1**, **2.2** and **2.3** were determined to be 0.10, 0.12 and 0.15 respectively, which are comparable to the values reported for Tyr, Trp and other unsubstituted coumarin derivatives (refer to Table 2).[113,123,138] Finally, compounds **2.1**, **2.2** and **2.3** reported large Stoke shifts of 152, 162 and 168 nm respectively.

As expected the spectroscopic properties of coumaryl amino acids **2.1** and **2.2** are in agreement with those reported by Wang *et al.*[134] and Brun *et al.*[138] Likewise the spectrochemical properties of the unreported coumaryl amino acid **2.3** are consistent with **2.1** and **2.2**. Coumaryl amino acids **2.1**, **2.2** and **2.3** displayed enhanced properties compared to endogenous fluorophores Tyr and Trp. The large molar extinction coefficients, fluorescence emission maxima (λ_{Em}) and Stokes shifts make these coumaryl amino acids (**2.1**-**2.3**) suitable probes for use in FRET studies.

As discussed in Section 2.1.3, the spectroscopic properties of coumarin-based fluorophores such as those synthesised here (**2.1**-**2.3**), are highly sensitive to changes in pH and polarity of

the surrounding microenvironment.[118,126,127,134,136,145,146] To measure the sensitivity of coumaryl amino acids **2.1**-**2.3** to pH and polarity, the absorbance spectra of **2.16**, **2.17** and **2.3** were collected in a range of solvents and the results are summarised in Table 3. Under acidic conditions (pH of 2) the excitation maxima (λ_{Ex}) of **2.16**, **2.17** and **2.3** are redshifted to a higher wavelength while in basic conditions (pH of 12) the excitation maxima (λ_{Ex}) are blue shifted to a lower wavelength. These changes in the absorbance spectra of **2.17** are displayed in Figure 18. The solvatochromic data for coumaryl amino acids **2.16**, **2.17** and **2.3** is previously unreported, however it is consistent with other reported coumarin-based fluorophores.[147]

Table 3: Solvatochromic properties of coumaryl amino acids (2.16, 2.17, 2.3). The excitation maxima (λ_{Ex}) of each compound in various solvents are listed.

Compound	n	1M NaOH	MeOH	EtOH	1M HCl
2.16	1	301	327	327	329
2.17	2	300	325	326	324
2.3	3	304	n.d	325[a]	324

[a] Determined from as the methanesulfonate salt (CH_3SO_3); [n.d] Not determined.

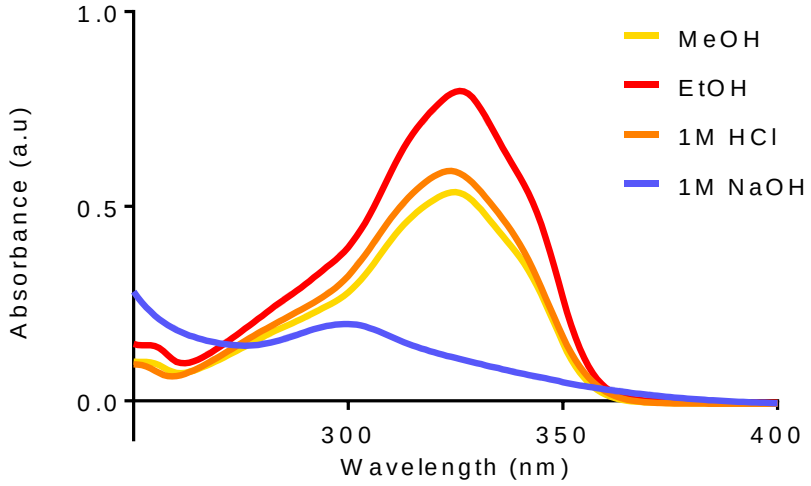


Figure 18: Solvatochromic properties of the glutamic acid derived coumaryl amino acid (2.17). Absorbance spectra were determined at a concentration of 50 µM.

The unreported homoglutamic acid (n=3) derived coumaryl amino acid (**2.3**) displays the strongest absorbance and fluorescence intensity, largest Stokes shift and highest quantum yield (refer to Table 2 and Figure 17). Despite these favourable spectroscopic properties, the synthesis of **2.3** was the longest and lowest yielding (13% over 4 steps). The shortest chain (n=1) coumaryl amino acid (**2.1**) was the least favoured for subsequent FRET study because it displays the least ideal spectroscopic properties (refer to Table 2 and Figure 17). Lastly, the glutamic acid (n=2) derived coumaryl amino acid (**2.2**) displays comparable fluorescent properties to **2.3** and was synthesised in a reasonable yield (27 % over 2 steps) (refer to Table 2 and Figure 17). Thus, **2.2** was used in the current study due to the ease with which it could be synthesised, and its suitable spectroscopic properties.

2.3 Conclusion

A range of fluorescent coumaryl amino acids **2.1**, **2.2** and **2.3** have been synthesised in moderate yields (19-28%). These amino acids displayed large molar extinction coefficients (ε), fluorescence emission maxima (λ_{Em}) and Stokes shifts, thus making **2.1**, **2.2** and **2.3** suitable probes for use in FRET experiments. The coumaryl amino acids **2.1**, **2.2** and **2.3** were transformed in good to excellent yields (51-90%) into the *N*α-Fmoc (**2.4**, **2.5** and **2.6**) and *N*α-Boc (**2.7** and **2.8**) derivatives for use in solid phase peptide synthesis (SPPS). Of these, β-ketoester derivative **2.14**, coumaryl amino acid **2.3** and the *N*α-Fmoc protected derivative **2.8** were previously unreported.

The glutamic acid derived fluorescent coumaryl amino acid (**2.2**) was synthesised in a good overall yield (27% over 2 steps) and displayed a high molar extinction coefficient (ε), large Stokes shift and strong fluorescence emission intensity. Thus, **2.2** was preferred for the proposed FRET-based investigation into the binding of IGF-II to the IR-A and IGF-1R. Transformation of coumaryl amino acid **2.2** into the *N*α-Fmoc (**2.5**) *N*α-Boc (**2.7-2.8**) protected derivatives, provided analogues which were suitable for use in SPPS (refer to Chapters 4 and 5). Studies on the incorporation of these glutamic acid derived fluorescent coumaryl amino acids into the IGF-II protein and analysis of the resulting fluorescent proteins is the focus of the reminder of this thesis.

Chapter 3

Genetic incorporation of a fluorescent amino acid into the IGF-II protein

3.1 Introduction

The incorporation of unnatural amino acids into peptides and proteins has led to an increased understanding of how proteins interact at a molecular level.[148-150] A variety of methods have been developed for the site-specific incorporation of unnatural amino acids into proteins (refer to Chapter 1).[88,151,152] As discussed in Chapter 2, an unnatural coumaryl amino acid **2.2** has been synthesised for use in recombinant protein expression. Herein the incorporation of this coumaryl amino acid (**2.2**) into the IGF-II protein using recombinant methods is discussed.

3.1.1 Recombinant protein expression

Protein which is expressed in a host organism from recombinant DNA is regarded as a recombinant protein. There are many established expression systems, each of which has its own distinct advantages and disadvantages. Expression systems are characterised based on the host used for expression and the method of delivery of the recombinant DNA into the host. Some common hosts include bacteria, yeast or eukaryotic cells and recombinant DNA can be delivered into these hosts in the form of viruses, plasmids or bacteriophage. For the expression of the fluorescent IGF-II analogues described in this chapter, *Escherichia coli* (*E. coli*) was used as the host and plasmids were used as vectors for DNA delivery.

3.1.2 Incorporation of unnatural amino acids into proteins

The incorporation of unnatural amino acids into proteins provides a range of invaluable tools to probe protein structure and function, both *in vitro* and *in vivo*.[153] A number of biosynthetic strategies have been developed for both the *in vitro* and *in vivo* incorporation of unnatural amino acids into proteins. Both these methodologies use nonsense suppression to precisely control the site of incorporation of the unnatural amino acid, but differ by the way in which the unnatural amino acid is attached to the transfer RNA (tRNA).[152,154,155]

3.1.2.1 *In vitro* incorporation of unnatural amino acids

In the *in vitro* method, acylation of the tRNA with the unnatural amino acid is performed external to the translation system. The amber suppressor tRNA, is chemically aminoacylated with the desired unnatural amino acid, and then added to a cell-free transcription/translation system.[95] This method has been successful at synthesising a number of proteins of varying sizes and structures, with the site-specific incorporation of greater than 80 novel amino acids into different proteins.[95,156-159] However the *in vitro* methodology has many disadvantages. The method is low yielding due to the low efficiency of incorporation of the unnatural amino acid and generation of the chemically aminoacylated tRNA is a complex process, which makes large scale expression using this methodology challenging.[155] Thus, an *in vivo* approach is often preferred for synthesising large quantities of non-native proteins.

3.1.2.2 *In vivo* incorporation of unnatural amino acids

An *in vivo* approach for the incorporation of unnatural amino acids has many advantages over the *in vitro* method. These include higher yields, improved fidelity and the ability to study protein structure and function both *in vivo* and *in vitro*.[155] Furthermore acylation of the tRNA, by the unnatural amino acid, is performed *in situ* during the expression, by an aminoacyl-tRNA synthetase (aaRS). Thus, an *in vivo* approach was used in the current study, to incorporate the unnatural coumaryl amino acid **2.2** into the IGF-II protein.

3.1.3 Nonsense suppression methodology

To investigate the binding of IGF-II to its two high affinity receptors using FRET, the site-specific incorporation of the unnatural coumaryl amino acid **2.2** into IGF-II was required. As discussed above, site-specific incorporation coumaryl amino acid **2.2** could be achieve through the use of nonsense suppression, and an *in vivo* approach was preferred as the tRNA is aminoacylated *in situ* during expression.[152,154,155]

The nonsense suppression methodology was pioneered by Schultz[152] and Chamberlain[154] in 1989, and allows for site-specific incorporation of unnatural amino acids into proteins in living cells. The methodology requires the integration of several key components so that the unnatural amino acid is compatible with existing translational machinery.[95] Specifically, incorporation requires a mutated version of the gene of interest, an orthogonal aminoacyl-tRNA synthetase/tRNA pair (*o*-aaRS/*o*-tRNA) that is selective for the unnatural amino acid and does not recognise endogenous amino acids, and an unnatural amino acid (Uaa) which is easily transported into the cell, non-toxic and stable to metabolic degradation (Figure 19).[153]

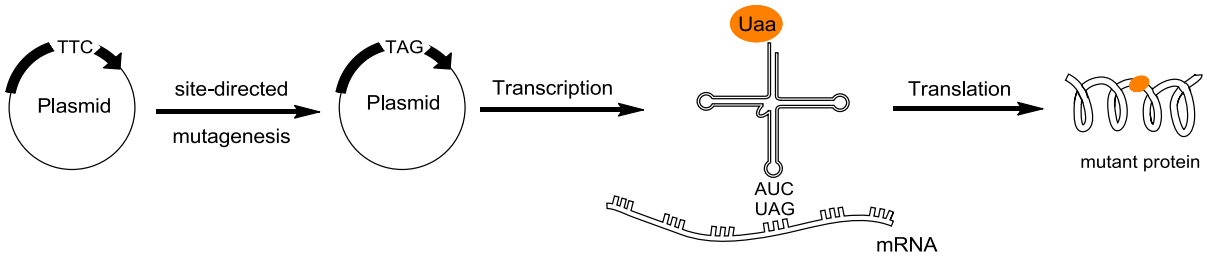


Figure 19: Schematic representation of the incorporation of unnatural amino acid by the nonsense suppression methodology. General approach to the site-specific incorporation of unnatural amino acids (Uaa) into proteins. Figure adapted from Wang and Schultz[160]

Mutation of the gene of interest can be easily accomplished by site directed mutagenesis, as long as the sequence for the gene is known. Once the site of incorporation of the unnatural amino acid is selected the codon corresponding to that site is mutated to a non-coding or "nonsense" codon.[160] This codon must be unique and not encoded for by any of the 20 endogenous amino acids. Typically the non-coding codon is a stop codon or a frame-shift codon (4 or 5-base codon).[158,161,162] In the case of a stop codon, the degenerate UAG codon is generally used as it the least common.[163] In the current study the UAG stop codon was used.

Successful incorporation of the unnatural amino acid is strongly dependent on the orthogonal aaRS/tRNA pair (*o*-aaRS/*o*-tRNA). The methodology requires the *o*-aaRS only aminoacylates the *o*-tRNA with the desired unnatural amino acid and not any cellular tRNA. Likewise it is desirable that the *o*-tRNA is not aminoacylated by endogenous aaRS. Optimally the unnatural

amino acid should not be a substrate for endogenous aaRS, and the *o*-aaRS/*o*-tRNA pair must also function efficiently in translation.[6,95,149,150,152-154,158,160]

The suppressor tRNA technology (*o*-aaRS/*o*-tRNA) is limited by its low efficiency to incorporate unnatural amino acids into proteins, on average the efficiency of incorporation is 20-30%.[164,165] The low efficiency is attributed to problems associated with aminoacylation of the *o*-tRNA by the *o*-aaRS, the misaminoacylation of the *o*-tRNA by the endogenous aaRS and competition between the release factor 1 (RF-1) and the *o*-tRNA for the unique codon.[92,149,164-170] Binding of the RF-1 to the unique codon before the *o*-tRNA results in termination of protein synthesis and the undesired truncated protein results (Figure 20).

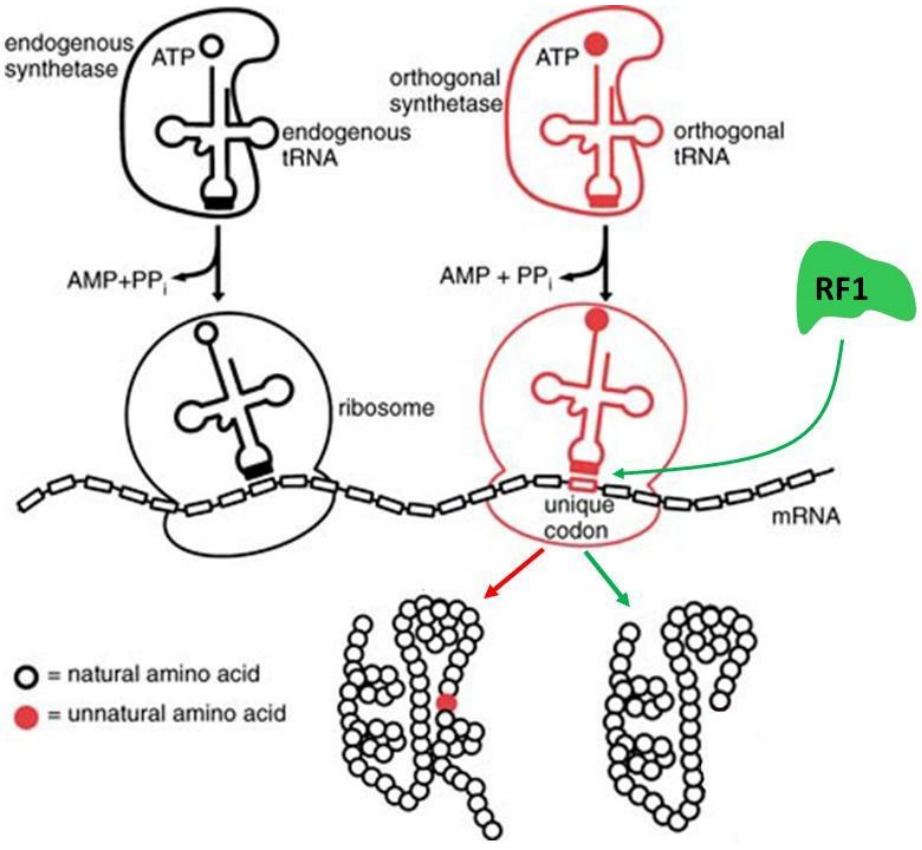


Figure 20: Site-specific, *in vivo* incorporation of unnatural amino acids in proteins. The competition for the unique codon on the mRNA between *o*-tRNA and the RF1 is shown. When the aminoacylated *o*-tRNA binds the mRNA first, full length protein with the unnatural amino acid incorporated is produced (*red*). However, if the RF1 binds to the mRNA first, protein synthesis is terminated and truncation of the protein results (*green*). Figure adapted from Wang and Schultz[160]

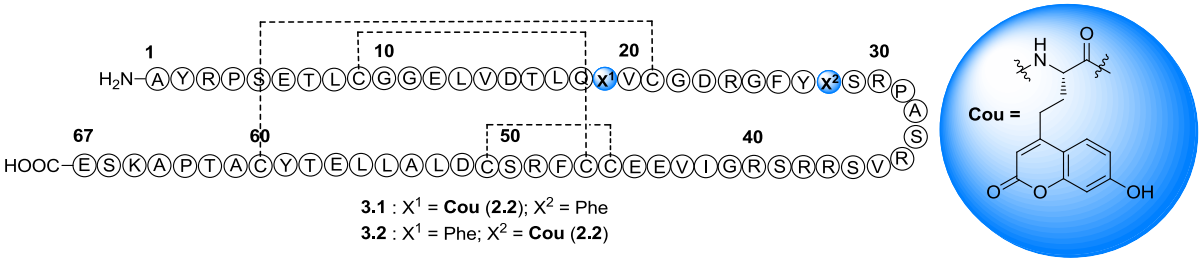
The nonsense suppression methodology allows for the *in vivo* site-specific incorporation of unnatural amino acids into proteins. This methodology has been employed for the incorporation of the unnatural coumaryl amino acid **2.2** into the IGF-II protein and is discussed herein.

3.2 Results and discussion

3.2.1 Choosing a site for incorporation of coumaryl amino acid 2.2

With a methodology selected to site-specifically incorporate the unnatural coumaryl amino acid **2.2**) into the IGF-II protein, the focus shifted to selecting an appropriate site for incorporation. Selecting a site nearby or within the IGF-II binding site was desirable, as it would allow more accurate information about the binding location of IGF-II within the IGF-1R and IR-A binding pockets, to be determined. Furthermore substitution of a structurally similar residue was also critical to prevent perturbation of the three-dimensional structure and biological activity of the resultant non-native IGF-II analogue.

A fluorescent coumarin-based probe was selected because of its ideal spectroscopic properties (refer to Chapter 2). The coumarin moiety is sterically and chemically similar to Phe and Tyr. Thus, it was thought substitution of Phe or Tyr, with the coumarin-based probe, would result in minimal disruption to the three-dimensional structure and biological activity of the IGF-II analogues. Two sites were selected for substitution of the coumarin-based probe, Phe^{19} (site 2) and Phe^{28} (site 1). Phe^{19} was chosen to examine IGF-II binding site 2 and Phe^{28} was selected for investigating IGF-II binding site 1 (refer to Section 1.1.4, Chapter 1). The sites of incorporation of **2.2** and the target non-native IGF-II analogues, F19Cou IGF-II (**3.1**) and F28Cou IGF-II (**3.2**) are summarised in Scheme 6.



Scheme 6: Target fluorescent IGF-II proteins (3.1 and 3.2). The sequence of IGF-II is shown with the residue numbers annotated above the sequence, disulfide bonds are shown as *dashed lines* and the coumarin-based fluorophore (Cou; **2.2**) is illustrated within the *blue* circle.

3.2.2 IGF-II expression vector

The expression and purification protocols for the synthesis of recombinant IGF-II analogues have been optimised by Forbes and collaborators.[23,44,46,171-173] For high level expression of the IGF-II analogues used in this work the pET32a (Novagen) expression vector was used.[174] The IGF-II constructs used in this thesis (F19X IGF-II, F28X IGF-II and L53X IGF-II) were generated "*in house*" by the Forbes/Wallace laboratory (by Ms Carlie Delaine), using the QuikChange™ site directed mutagenesis kit (Stratagene) as previously described.[23,44,171,173] Finally, protein expression was carried out in *E. coli* BL21 (DE3) (F– *ompT hsdSB* (*rB– mB–*) *gal dcm* (DE3)).

In the pET vector, the IGF-II genes are under the control of the T7*lac* promoter and T7 terminator, and the plasmid includes an ampicillin resistance marker (*ampR*).[174] Expression of the IGF-II gene was induced by the addition of isopropyl-β-D-thiogalactoside (IPTG) (Figure 21).

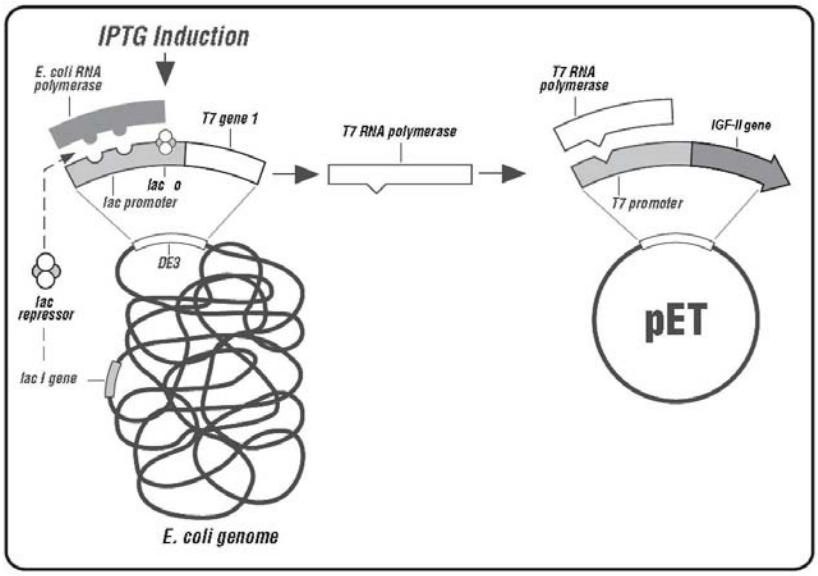


Figure 21: Regulation of the IGF-II gene by the T7*lac* promoter in *E. coli* BL21 (DE3). Figure adapted from Novagen pET System manual.[175]

Using the expression system described above, the IGF-II protein is produced as a fusion protein, where the N-terminus is fused to the first 11 amino acids of the porcine growth hormone ([Met^{1}]pGH-(1–11)-PAPM). The N-terminal fusion partner enhances the efficiency of expression, improves solubility during purification and increases recovery of correctly refolded IGF-II.[171,172] Following IPTG induction the expressed fusion protein is found in insoluble inclusion bodies.[171] During purification the fusion partner is removed by enzymatic cleavage using a highly selective mutant α-lytic protease called Prag A9, which cleaves C-terminal of the PAPM motif.

3.2.3 pEB-CouRS expression vector

The *in vivo* incorporation of coumaryl amino acid **2.2** at defined sites in the IGF-II protein required an *o*-aaRS/*o*-tRNA pair which is specific for **2.2**. The pEB-CouRS expression vector containing the genes for the *o*-aaRS (*Mj*CouRS) and *o*-tRNA (*Mj*tRNA^{Cou}) was a kind gift from Professor P. G. Schultz at the Scripps Research Institute, San Diego. The expression vector was evolved from the archeabacterium, *Methanocaldococcus jannaschii* (*M. jannaschii*)

by Schultz and co-workers, and uniquely incorporates **2.2** in response to a TAG codon.[87,134,153,160,176] The *Mj*CouRS and *Mj*tRNA^{Cou} genes were encoded on a pEB-CouRS plasmid, under the control of the *lpp* promoter and *rrn* terminator and the plasmid includes a tetracycline (*tetR*) resistance marker. Similar to the pET vector described above (refer to Section 3.2.2), the *Mj*CouRS and *Mj*tRNA^{Cou} genes were induced into expression by the addition of IPTG.

3.2.4 Optimisation of the IGF-II expression system using the pEB-CouRS expression vector

With the required plasmids in hand, an investigation into the optimal conditions for the expression of the F19X IGF-II and F28X IGF-II constructs was undertaken. In general, the pET vector for either the F19X IGF-II or F28X IGF-II construct was transformed into competent *E. coli* BL21 (DE3) cells containing the pEB-CouRS vector (refer to Section 3.2.3). The transformant was plated on Luria Bertani (LB) agar plates supplemented with ampicillin (100 μg/ml) and tetracycline (50 μg/ml), and incubated at 37 °C overnight. Overnight cultures of a single colony were grown in LB supplemented with ampicillin and tetracycline. An aliquot of the culture was then subcultured into pre-warmed growth medium, supplemented with the appropriate antibiotics and the unnatural coumaryl amino acid **2.2** (where required). The expression culture was monitored, using the optical density at 600 nm (OD_{600nm}) as a measure of culture growth.[m] The *E. coli* BL21 (DE3) cultures were induced into protein expression when the OD_{600nm} of the culture was approximately 0.6 (in log phase), by the addition of IPTG. The expression cultures were incubated overnight and the results of the expression were analysed by denaturing sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

[m]Larger OD_{600nm} indicates higher density cell cultures.

Initially the expression conditions developed by Forbes and co-workers for the native IGF-II and other mutant IGF-II constructs were investigated (Table 4).[22,23,171,173] Expressions were carried out in LB broth supplemented with **2.2** (2 mM), ampicillin (100 μg/ml) and tetracycline (50 μg/ml), incubated at 37 °C and the *E. coli* BL21 (DE3) cultures were induced into protein expression using IPTG (0.1 mM). These conditions were found to be less than optimal as the bacterial growth was extremely slow and in some cases expressions were abandoned due to the minimal culture growth (OD_{600nm} < 1). In the same experiment the conditions outlined by Wang *et al.*[134] for the incorporation of the same coumaryl amino acid (**2.2**) were also investigated. Expressions carried out by Wang and co-workers were performed in 2xYT (2x yeast extract and tryptone) medium, supplemented with **2.2** (1 mM), ampicillin (100 μg/ml) and tetracycline (50 μg/ml), incubated at 37 °C and cultures were induced into protein expression using IPTG (0.1 mM).[134] The results of the expressions are summarised in Table 4 and show the expressions performed in 2xYT medium had increased cell culture growth for all constructs compared to the expressions carried out in LB medium.

Table 4: Conditions investigated for the expression of the F19X IGF-II and F28X IGF-II constructs. Expressions were carried out in either LB or 2xYT medium and induced into protein expression using IPTG (0.1 mM). Where F19X refers to incorporation at position 19; F28X refers to incorporation at position 28; *Mj*TyrRS/*Mj*tRNA^{Tyr} refers to the incorporation of tyrosine; *Mj*CouRS/*Mj*tRNA^{Cou} refers to the incorporation of **2.2** (refer Section 3.2.3); 0 mM refers to the absence of **2.2** from the medium and OD_{600nm} refers to growth in log phase.[iv]

IGF-II construct	*o*-aaRS/tRNA pair	Growth Medium	[2.2] mM	Final OD_{600nm} after 24 h
F19X	*Mj*CouRS/*Mj*tRNA^{Cou}	LB	2	< 1
	*Mj*TyrRS/*Mj*tRNA^{Tyr}		0	< 1
F28X	*Mj*CouRS/*Mj*tRNA^{Cou}		2	< 1
	*Mj*TyrRS/*Mj*tRNA^{Tyr}		0	< 1
F19X	*Mj*CouRS/*Mj*tRNA^{Cou}	2xYT	1	1-2
	*Mj*TyrRS/*Mj*tRNA^{Tyr}		0	≈ 2-4
F28X	*Mj*CouRS/*Mj*tRNA^{Cou}		1	1-2
	*Mj*TyrRS/*Mj*tRNA^{Tyr}		0	≈ 2-4

[iv] Larger OD_{600nm} indicates higher density cell cultures.

Post-induction samples (+) of the F19X IGF-II and F28X IGF-II constructs displayed bands at 7.5 kDa in the SDS-PAGE analysis shown in Figure 22. These results indicate the presence of full length protein (1-67). However, the expression controls (no **2.2** in the medium) also exhibited the same distinctive bands at 7.5 kDa in the SDS-PAGE analysis (refer to Figure 22). These results suggested the *o*-tRNA (*Mj*tRNA^{Cou}) was not functioning efficiently and was likely being misaminoacylated with endogenous amino acids present in the media, such as tyrosine and not the desired unnatural amino acid (**2.2**).[92] It was proposed that misaminoacylation was a consequence of using rich media and could be overcome by changing the medium.[92]

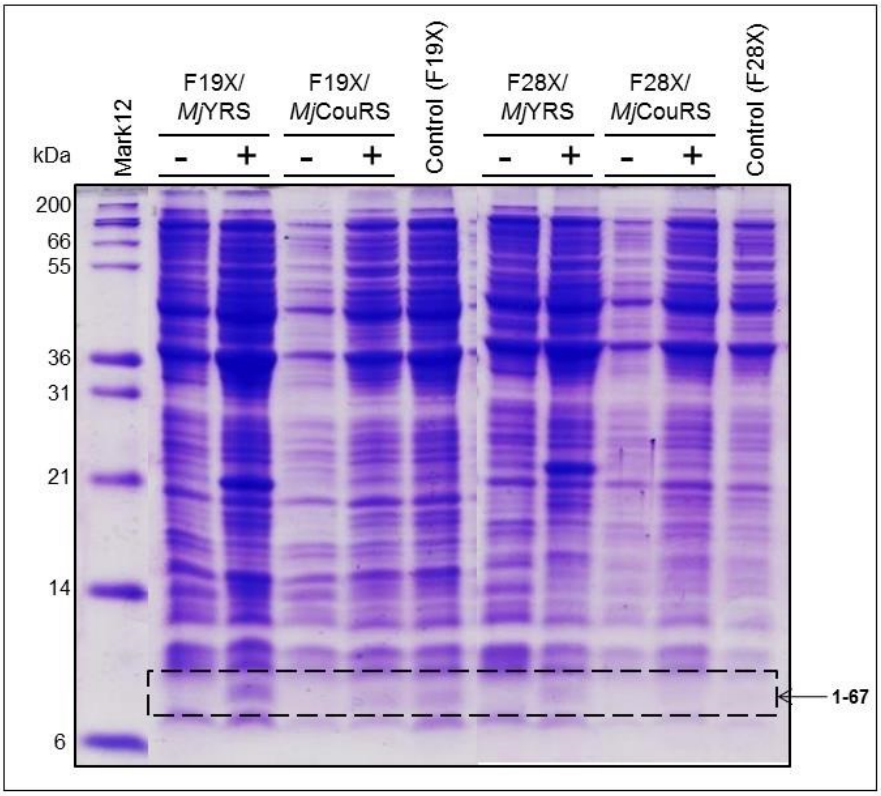


Figure 22: SDS-PAGE analysis of the expression of the F19X and F28X IGF-II constructs in 2xYT medium. BL21 (DE3) cultures were sampled prior and post-induction, analysed by denaturing SDS-PAGE and stained with Coomassie brilliant blue. Where F19X refers to incorporation at position 19; F28X refers to incorporation at position 28; *Mj*YRS refers to the incorporation of tyrosine; *Mj*CouRS refers to the incorporation of **2.2** (refer to Section 3.2.3); (-) = pre-induction sample; (+) = post-induction sample; control refers to the absence of **2.2** from the medium and 1-67 indicates a band corresponding to full length protein.

Several growth media were investigated for the incorporation of **2.2** into the IGF-II protein and are summarised in Table 5. For these experiments the L53X IGF-II construct was used. This construct contains the nonsense (TAG) codon at position 53 and was used as it has previously shown high levels of incorporation for other non-natural amino acids into the IGF-II protein.[177]

Table 5: Investigating the ideal growth medium for the incorporation of coumaryl amino acid 2.2 into the L53X IGF-II construct. Where L53X refers to the incorporation at position 53; *Mj*TyrRS/*Mj*tRNA^{Tyr} refers to the incorporation of tyrosine; *Mj*CouRS/*Mj*tRNA^{Cou} refers to the incorporation of **2.2** (refer to Section 3.2.3); NIM = non-inducing medium; AIM = auto-inducing medium*; MIN = minimal medium; 2xYT = 2 x yeast and tryptone; 100, 50, 25 and 10% refers to the volume of AIM medium in water; T = band observed indicating truncated protein (1-53); F = band observed indicating possible full length protein (1-67) and control refers to the absence of **2.2** from the medium.

Medium	IGF-II construct and *o*-aaRS/*o*-tRNA pair	
	L53X/ *Mj*TyrRS/*Mj*tRNA^{Tyr}	L53X/ *Mj*CouRS/*Mj*tRNA^{Cou}
NIM (control)	T	T
MIN	TF	F
2xYT	TF	TF
***AIM 100%**	T	T
***AIM 50% (v/v)**	T	T
***AIM 25% (v/v)**	T	T
***AIM 10% (v/v)**	T	T

*Does not require IPTG induction.

Four key media were investigated, and included auto-inducing medium (AIM), minimal medium (MIN), 2xYT medium and non-inducing medium (NIM, used as a control) (Table 5). MIN, AIM and NIM were investigated as they have been shown to be suitable for the incorporation of unnatural amino acids.[156,178] Small scale expressions were carried out for two colonies of each construct in medium, supplemented with **2.2** (1 mM), ampicillin (100 μg/ml) and tetracycline (50 μg/ml), and incubated at 37 °C. Where required, *E. coli* BL21 (DE3) cultures were induced into protein expression with IPTG (0.1 mM), when the OD_{600nm} was approximately 0.60 (in log phase).

Cultures grown in AIM did not require IPTG induction as the medium is formulated to induce protein expression automatically when the culture approaches cell growth saturation. The medium contains a supply of glucose, which is preferentially metabolised by the *E. coli* BL21 (DE3) during the growth phase (early log phase) and thus prevents the uptake of lactose. Once the glucose is depleted, usually mid to late log phase, the lactose in the medium is metabolised, which induces protein expression through the regular pathways (refer to Figure 21, Section 3.2.2).[178]

The results of the variable media expressions are summarised in Table 5 and Figure 23. The constructs grown in AIM produced bands at 6 kDa in the SDS-PAGE analysis (refer to Table 5). This result indicated the expression of truncated protein (1-53). In contrast, the constructs grown in MIN and 2xYT media displayed bands at 7.5 kDa in the SDS-PAGE analysis (refer to Figure 23). These results indicated MIN and 2xYT media supported the expression of the full length protein (1-67). However, previous expression data suggested 2xYT medium was not a suitable growth medium due to misaminoacylation of the *o*-tRNA (*Mj*tRNA^{Cou}) by endogenous amino acids. Consequently MIN was further investigated as a potential growth medium for the expression of the F19Cou IGF-II (**3.1**) and F28Cou IGF-II (**3.2**) proteins.

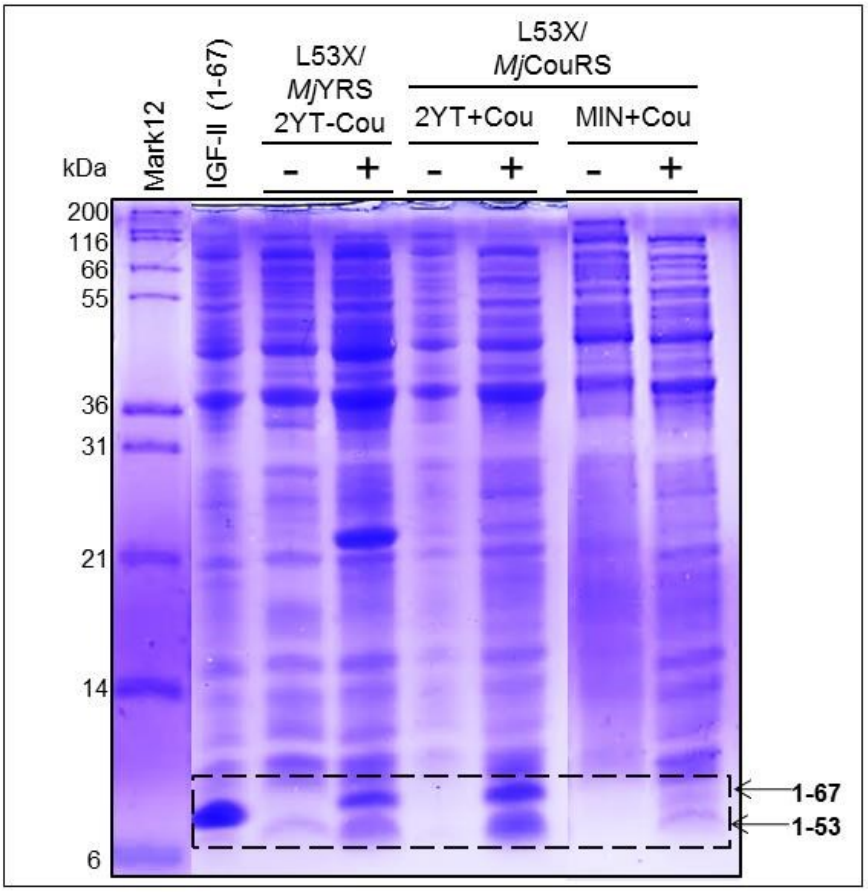


Figure 23: SDS-PAGE analysis of the expressions of the L53X IGF-II construct in 2xYT and MIN media. BL21 (DE3) cultures were sampled prior and post-induction, analysed by denaturing SDS-PAGE and stained with Coomassie brilliant blue. Where L53X refers to the incorporation at position 53; *Mj*YRS refers to the incorporation of tyrosine; *Mj*CouRS refers to the incorporation of **2.2** (1 mM) (refer to Section 3.2.3); (-Cou) refers to the absence of **2.2** from the medium; (+Cou) refers to the presence of **2.2** in the medium; (-) = pre-induction sample and (+) = post-induction sample.; 1-53 indicates a band corresponding to truncated protein; 1-67 indicates a band corresponding to full length protein and IGF-II (1-67) is a reference.

In addition to the misaminoacylation of the *o*-tRNA (*Mj*tRNA^{Cou}), it was also hypothesised the rate of protein expression could be contributing to the inefficient of incorporation of **2.2**. Thus, in an attempt to slow the rate of protein expression and improve the yield of full length protein (1-67), the expression temperature and IPTG concentration were investigated. Expressions were carried out using the L53X IGF-II construct, in MIN supplemented with **2.2** (1 mM), ampicillin (100 µg/ml), tetracycline (50 µg/ml) and induced into protein expression using IPTG.

It was predicted lowering the temperature would decrease the rate of cell growth and allow sufficient time for the expression of the *o*-aaRS and *o*-tRNA, and as such increase the expression of the full length protein (1-67). Two temperatures, 25 °C and 37 °C were investigated and the results are summarised in Table 6 and Figure 24. The expressions carried out at 25 °C produced higher density cell culture growth than expressions carried out at 37 °C (refer to Table 6). Furthermore, the SDS-PAGE analysis of the cultures grown at 25 °C, displays strong bands at 7.5 kDa, which indicates the expression of full length protein (1-67) (refer to Figure 24). In contrast, the cultures grown at 37 °C exhibited strong bands at 6 kDa, which indicates the expression of truncated protein (1-53) (refer to Figure 24). These results demonstrate the expression of the full length L53X IGF-II construct was optimal when the *E. coli* BL21 (DE3) cultures were grown in MIN and incubated at 25 °C.

Table 6: Effect of temperature and IPTG concentration on the expression of the L53X IGF-II construct. Where the OD_{600nm} refers to growth in log phase[v] and 0 mM refers to the absence of **2.2** from the medium.

Final [IPTG] mM	[2.2] mM	Temperature (°C)	Final OD_{600nm} after 24 h
0.1	0	25	3.2
	1		3.1
	0	37	2.1
	1		1.6
0.05	0	25	3.0
	1		3.4
	0	37	1.7
	1		2.1
0.01	0	25	3.5
	1		3.9
	0	37	2.5
	1		2.5

[v] Larger OD_{600nm} indicates higher density cell cultures.

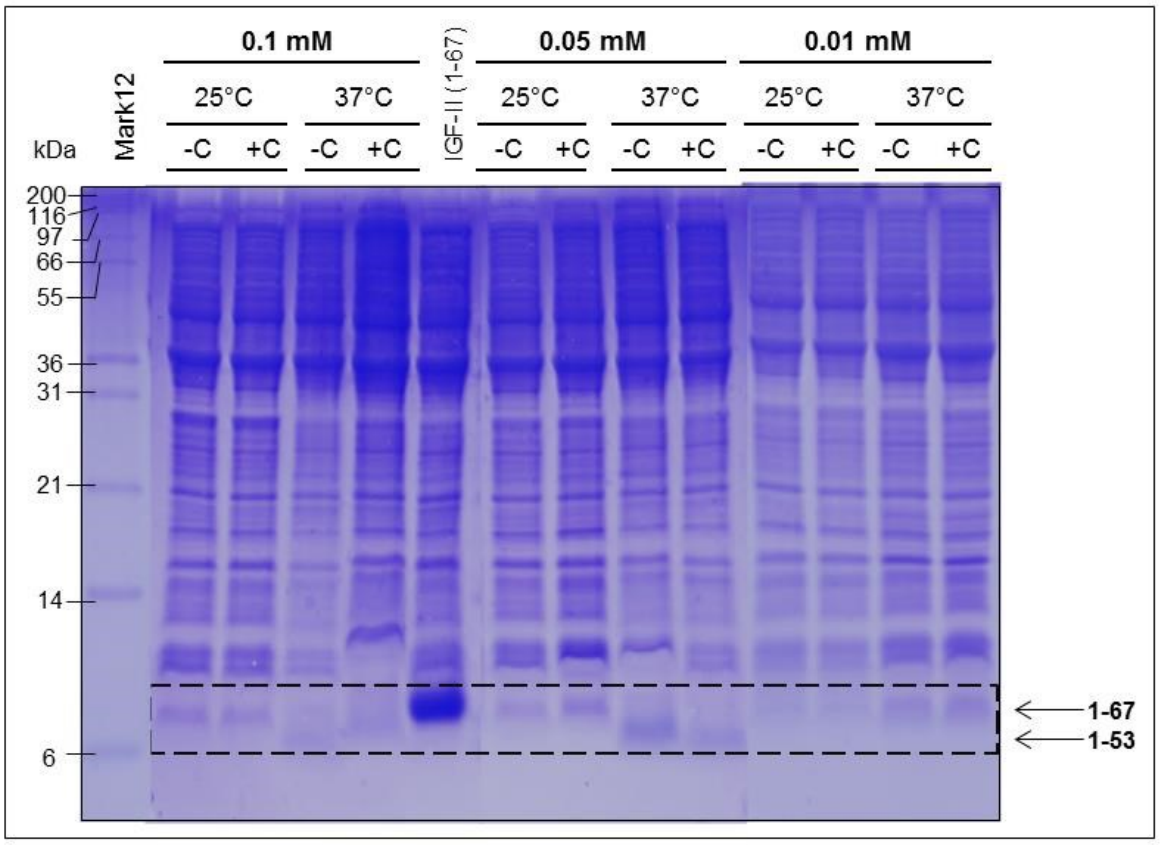


Figure 24: SDS-PAGE analysis of post-induction samples from the variable temperature and variable IPTG concentration expressions of the L53X IGF-II construct. BL21 (DE3) cultures were sampled post-induction, analysed by denaturing SDS-PAGE and stained with Coomassie brilliant blue. Where L53X refers to the incorporation at position 53; (-C) refers to the absence of **2.2** from the medium; (+C) refers to the presence of **2.2** (1 mM) in the medium; 1-53 indicates a band corresponding to truncated protein; 1-67 indicates a band corresponding to full length protein and IGF-II (1-67) is a reference.

The effect of IPTG concentration on protein expression was next investigated, with the results summarised in Table 6 and Figure 24. It was proposed lowering the concentration of IPTG used to induce the culture into protein expression would decrease the rate of protein expression, and thus produce more full length protein (1-67). A range of IPTG concentrations were examined (0.01-0.1 mM) (refer to Table 6). When a final concentration of 0.01 mM IPTG was used, faint bands at 7.5 kDa were observed in the SDS-PAGE analysis (refer to Figure 24). This result indicates expression of the IGF-II gene was extremely low and was likely to be comprised at such low concentrations of IPTG. In contrast, when a final concentration of 0.1 mM IPTG was used, only bands at 6 kDa, corresponding to truncated

protein (1-53) were observed in the SDS-PAGE analysis (refer to Figure 24). Finally, expressions carried out at 0.05 mM IPTG exhibited bands at 7.5 kDa in the SDS-PAGE analysis shown in Figure 24, indicating the expression of full length protein (1-67). Thus, full length protein expression was strongest when a final concentration of 0.05 mM of IPTG was used.

After optimising the conditions for the incorporation of **2.2** into the L53X IGF-II construct, the suitability of these conditions for the expression of the desired F19Cou IGF-II (**3.1**) and F28Cou IGF-II (**3.2**) proteins was investigated. Small scale expressions of the F19X IGF-II and F28X IGF-II constructs were carried out in MIN supplemented with **2.2** (1 mM), ampicillin (100 μg/ml), tetracycline (50 μg/ml), incubated at 25 °C and induced into protein expression using IPTG (0.05 mM).

Problems relating to low culture growth (OD_{600nm} < 1) were again encountered during the expression of the F28X IGF-II construct. Thus, expression of the F28Cou IGF-II protein (**3.2**) was abandoned. However expression of the F19X IGF-II construct gave encouraging results. SDS-PAGE analysis of the expression is shown in Figure 25 and displays a strong band at 7.5 kDa (for the culture grown at 25 °C), indicating the expression of full length protein (1-67). In light of the successful small scale expression of the F19X IGF-II construct a larger scale expression was attempted.

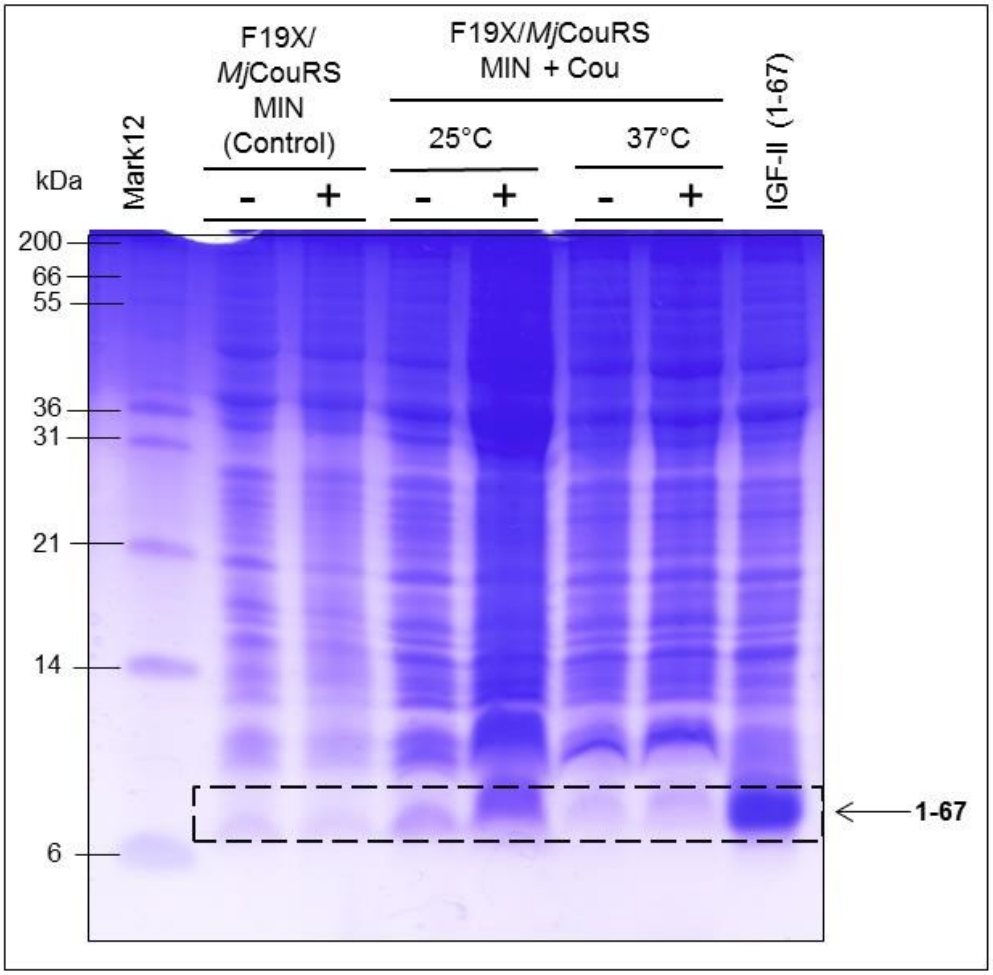


Figure 25: SDS-PAGE analysis of the variable temperature expressions of the F19X IGF-II construct. BL21 (DE3) cultures were sampled prior and post-induction analysed by denaturing SDS-PAGE gel and stained with Coomassie brilliant blue. Where F19X refers to the incorporation at position 19; *Mj*CouRS refers to the incorporation **2.2** (refer to Section 3.2.3); (-) = pre-induction sample; (+) = post-induction sample; control refers to absence of **2.2** from the medium; 1-67 indicates a band corresponding to full length protein and IGF-II (1-67) is a reference.

A medium scale expression (250 ml) of the F19X IGF-II construct was carried out in MIN supplemented with **2.2** (1 mM), ampicillin (100 µg/ml) and tetracycline (50 µg/ml), incubated at 25 °C and induced into protein expression using IPTG (0.05 mM). However these expressions condition led to the death of the *E. coli* BL21 (DE3) cultures overnight. It was concluded the death of the *E. coli* BL21 (DE3) cultures was due to the cultures becoming stressed from the lack of nutrients in the MIN. To overcome this problem the *E. coli* BL21 (DE3) were matured to grow in MIN.

Maturations of the *E. coli* BL21 (DE3) were carried out to evolve the bacteria to grow in MIN (refer to Section 7.3.2, Chapter 7 for more detail). Once the *E. coli* BL21 (DE3) were matured another medium scale expression of the F19X IGF-II construct was attempted. The expression was carried out using the same conditions as described above. However maturation proved unsuccessful, and death of the *E. coli* BL21 (DE3) was again observed. Thus, despite extensive investigation only conditions suitable for the small scale expressions of F19Cou IGF-II (**3.1**) and F28Cou IGF-II (**3.2**) proteins were discovered.

A colleague at the time was also experiencing similar difficulties with the inefficient incorporation of **2.2** and low level expression of full length protein using the pEB-CouRS vector. Thus, improvement of the pEB-CouRS expression vector was carried out in an attempt to overcome these problems.[179]

3.2.5 Improving protein expression: Generation of the pRFSDuet™ vector

It was hypothesised the difficulties with the inefficient incorporation of **2.2** were related to the pEB-CouRS vector and the low-level expression of the *o*-aaRS and *o*-tRNA (*Mj*CouRS and *Mj*tRNA^{Cou}).[179] Low expression of the *o*-aaRS and *o*-tRNA would lead to limiting concentrations of *o*-tRNA available to bind the unique UAG codon. As a consequence the RF1 would bind the mRNA and terminate protein synthesis, causing truncation of the protein (refer to Figure 20, Section 3.1.3).[179] Thus, it was proposed increasing the expression of the *o*-tRNA, would minimise protein truncation.

Improvement of the pEB-CouRS expression vector was carried out by Jitrapakdee.[179] The cDNA coding for the *Mj*CouRS and *Mj*tRNA^{Cou} genes was removed from the pEB-CouRS expression vector and ligated into a pRFSDuet™ vector using restriction enzyme cloning. The *Mj*CouRS and *Mj*tRNA^{Cou} genes were placed under the control of a T7*lac* promoter and T7 terminator, and the plasmid included a kanamycin resistance gene (*kan*R). The pRFSDuet™ vector was selected as it is compatible for co-expression with the pET vector (used for

expression of the IGF-II gene), and the encoded T7*lac* promoter drives high level expression.[180]

The pEB-CouRS and pRFSDuet™ vectors differ by the type of promoter which drives the expression of the *Mj*CouRS and *Mj*tRNA^{Cou} genes and the replication origin (*ori*). In the pEB-CouRS vector, expression of the *o*-aaRS and *o*-tRNA is driven by the *lpp* promoter, whereas the pRFSDuet™ vector uses a T7*lac* promoter. The T7*lac* promoter drives faster and stronger expression than the *lpp* promoter. The vector also contains a high copy number RSF replicon, which has the potential to generate greater than 100 copies of the *o*-aaRS and *o*-tRNA per cell. Thus, it was anticipated the pRFSDuet™ vector would generate more copies of the *o*-aaRS and *o*-tRNA compared to pEB-CouRS vector.

Several approaches have been reported for improving the efficiency of incorporation of unnatural amino acids into proteins. Many of these approaches focus on biasing the competition from RF1-mediated protein synthesis termination towards *o*-tRNA binding for full length protein synthesis.[167-169,181-183] Specifically, these methods focus on the suppression or elimination of the RF1 from the culture. However these methods are either expensive, only suitable for small-scale expression or produce low expression yields. In contrast, the pRFSDuet™ approach described above, focuses on outcompeting the RF1, opposed to elimination or suppression. This method is cheap, scalable and uses commercially available plasmids and cell lines and thus is a good alternative to using RF1 elimination. Application of the pRFSDuet™ vector for improving the efficiency of incorporation of **2.2** is now described.

3.2.6 Expression of fluorescent IGF-II analogues using the pRFSDuet™ vector.

With the pRFSDuet™ expression vector in hand, the focus shifted to investigating the ideal expression conditions for the incorporation of **2.2** into the IGF-II protein. The pET expression vector (refer to Section 3.2.2) containing the desired mutated IGF-II gene was transformed into competent *E. coli* BL21 (DE3) cells containing the pRFSDuet™ vector (refer to Section 3.2.5). The transformant was plated on LB agar plates, supplemented with **2.2** (1 mM), ampicillin (100 μg/ml) and kanamycin (30 μg/ml), and incubated at 37 °C overnight. The following day single colonies were selected and amplified for 8 h before subculturing for overnight expression in AIM.

SDS-PAGE analysis of the expression of the F19X IGF-II and F28X IGF-II constructs in AIM is shown in Figure 26 and displays weak bands at 7.5 kDa, which suggests the presence of full length protein (1-67). However further analysis of the F19X IGF-II and F28X IGF-II expression samples was required to conclusively determine the presence on any of full length protein (1-67). Visualisation of the expression samples under UV light (254nm)[vi] was used to confirm if the incorporation of **2.2** was successful and thus if full length protein (1-67) was being expressed. The SDS-PAGE after irradiation at 254 nm is displayed in Figure 27, and shows fluorescent bands at 7.5 kDa. These bands confirm the successful incorporation of **2.2** and as such expression of the full length F19Cou IGF-II (**3.1**) and F28Cou proteins (**3.2**).

[vi] The coumarin-based fluorophore (**2.2**) fluoresces when excited at 254 nm (refer to Chapter 2), thus any protein with **2.2** incorporated will display fluorescence.

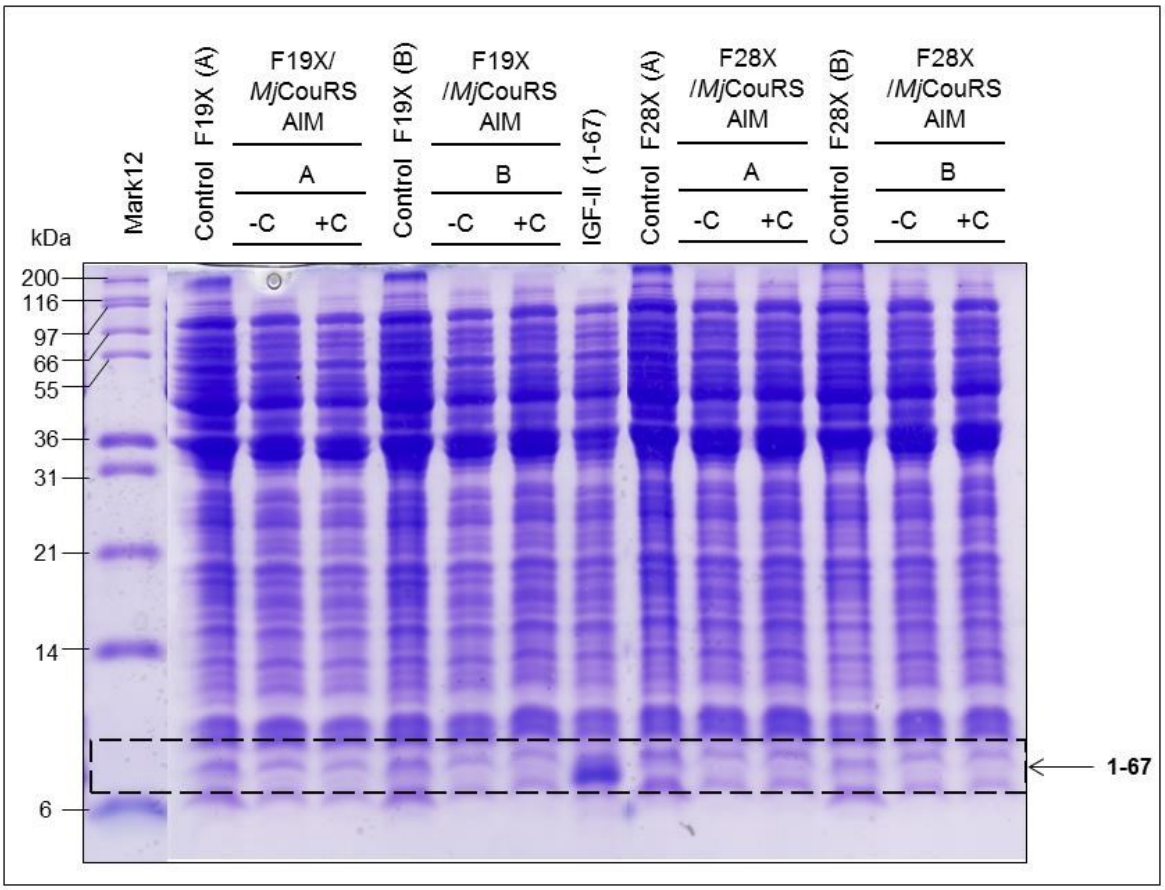


Figure 26: SDS-PAGE analysis of the expression of the F19X IGF-II and F28X IGF-II constructs in AIM using the pRFSDuet™ expression vector. BL21 (DE3) cultures were sampled post-induction and analysed by denaturing SDS-PAGE and stained with Coomassie brilliant blue. Where F19X refers to the incorporation at position 19; F28X refers to incorporation at position 28; *Mj*CouRS refers to the incorporation **2.2** (refer to Section 3.2.3); A and B refer to independent colonies selected from the same agar plate; (-C) refers to the absence of **2.2** from the medium; (+C) refers to the presence of **2.2** (1 mM) in the medium; control refers to an expression carried out in non-inducing medium (NIM) and in the absence of **2.2**; 1-67 indicates a band corresponding to full length protein and IGF-II (1-67) is reference.

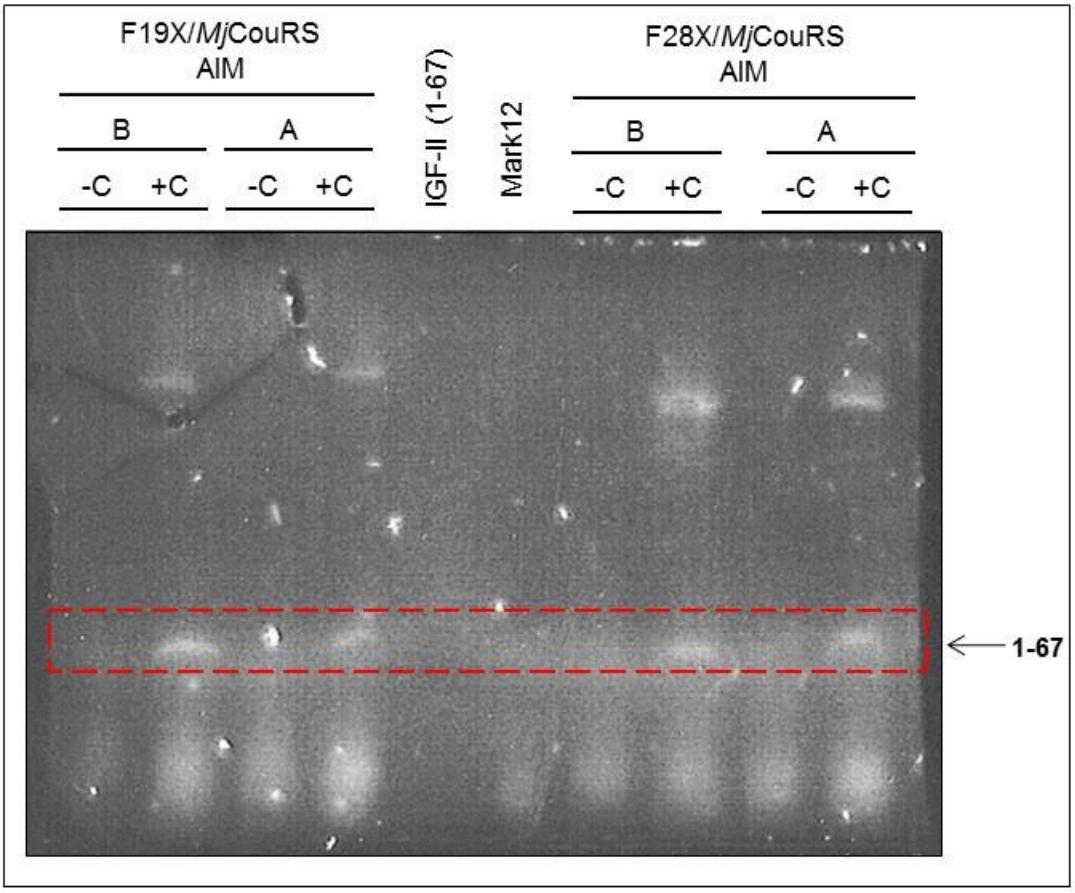


Figure 27: SDS-PAGE analysis of the expression of the F19X IGF-II and F28X IGF-II constructs in AIM using the pRFSDuet™ vector after irradiation at 254 nm. BL21 (DE3) cultures were sampled post-induction and analysed using denaturing SDS-PAGE and visualised under UV light (254 nm). Fluorescent bands at 7.5 kDa correspond to the expression of full length protein and are highlighted with a *dashed red box*. Where F19X refers to the incorporation at position 19; F28X refers to incorporation at position 28; *Mj*CouRS refers to the incorporation **2.2** (refer to Section 3.2.3); A and B refer to independent colonies selected from the same agar plate; (-C) refers to the absence of **2.2** from the medium; (+C) refers to the presence of **2.2** (1 mM) in the medium; 1-67 indicates a band corresponding to full length protein and IGF-II (1-67) is reference.

The presence of fluorescent bands at 7.5 kDa in the SDS-PAGE analysis shown in Figure 27, confirms the successful incorporation of **2.2** and thus, the expression of full length protein (1-67). In contrast, the same expression conditions with the pEB-CouRS vector produced truncated protein (refer to Table 5, Section 3.2.4). These results suggest the pRFSDuet™ vector improved the efficiency of incorporation of **2.2** into the F19X and F28X IGF-II constructs. Despite the expression of full length protein, the level of expression was too low for scale up and consequently further optimisation of the expression conditions in AIM was undertaken.

The concentration of **2.2** and the expression temperature were investigated in an attempt to improve the expression of the F19Cou IGF-II (**3.1**) and F28Cou IGF-II (**3.2**) proteins. Small scale expressions were conducted in AIM supplemented with ampicillin (100 μg/ml) and kanamycin (30 μg/ml) and controls for each construct were carried out in NIM. The results of the expressions are summarised in Table 7 and the related SDS-PAGE analysis is displayed in Figure 28.

Table 7: Optimisation of the expression of the F19X IGF-II and F28X IGF-II constructs in AIM. OD_{600nm} measurements after 24 h are reported for several colonies and constructs at varying temperatures and concentrations of **2.2**. Where the OD_{600nm} refers to growth in log phase[vii]; A and B refer to independent colonies selected from the same agar plate for each individual construct; F19X refers to the incorporation at position 19; F28X refers to the incorporation at position 28; L53X refers to the incorporation at position 53 and 0 mM refers to the absence of **2.2** from the medium.

	F19X IGF-II			**F28X IGF-II**		**L53X IGF-II**[viii]
	37 °C		**30 °C**	**37 °C**		**37 °C**
[2.2] mM	***A***	***B***	***B***	***A***	***B***	***A***
0	2.40	2.46	3.77	2.36	2.07	2.79
1	2.36	2.14	3.99	2.25	2.61	3.01
2	2.14	2.28	4.07	2.72	2.26	2.49
5	2.66	2.14	3.93	2.87	2.04	2.30

[vii] Larger the OD_{600nm} the higher cell culture density.

[viii] L53X IGF-II construct was a control as it has previously shown high levels of incorporation for other non-natural amino acids (refer to Section 3.2.4).

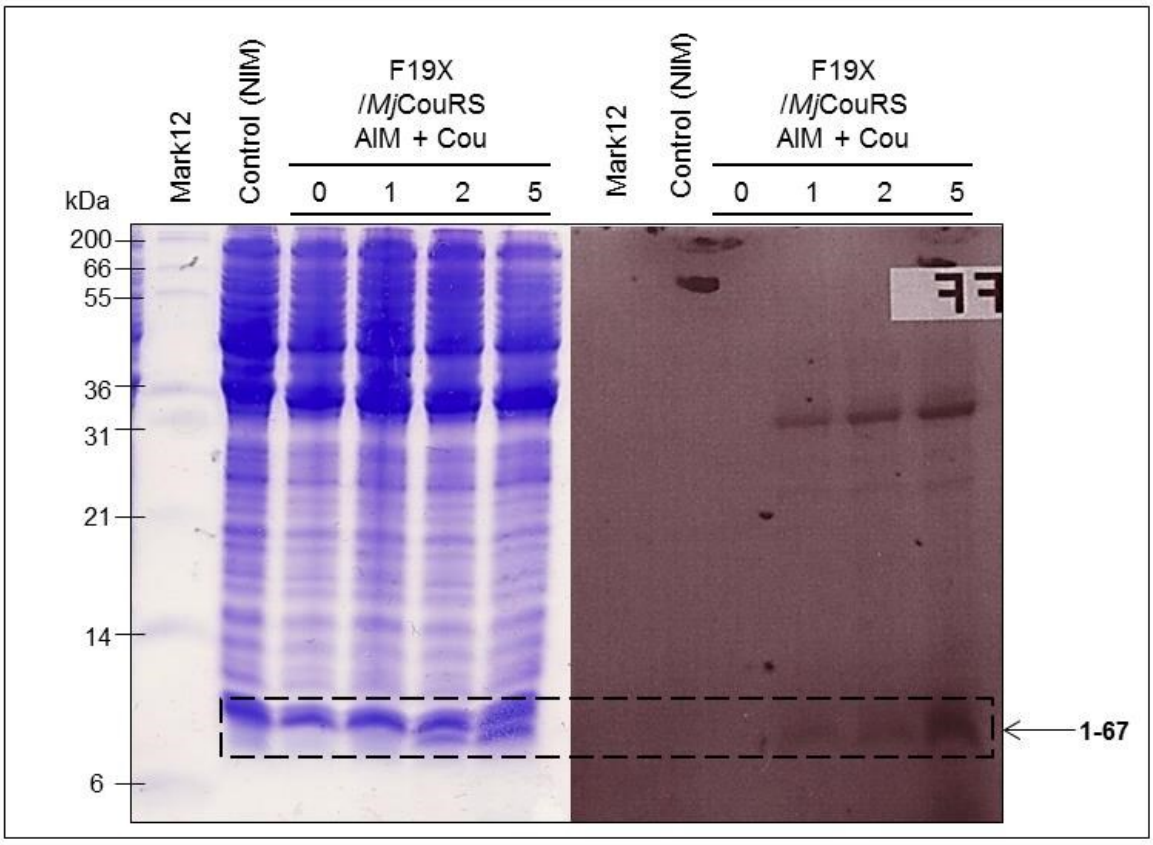


Figure 28: SDS-PAGE analysis of the expression of the F19X IGF-II construct in AIM supplemented with varying concentrations of 2.2 (mM) and incubated at 37 °C. BL21 (DE3) cultures were sampled post-induction, analysed by denaturing SDS-PAGE and visualised under UV light (254 nm) (*right*) before staining with Coomassie brilliant blue (*left*). Where F19X refers to the incorporation at position 19; *Mj*CouRS indicates incorporation of **2.2** (refer to Section 3.2.3); 0, 1, 2, and 5 refer to the concentration (in mM) of **2.2** in the medium; 0 refers to the absence of **2.2** from the medium; control refers to a culture grown in NIM and in the absence of **2.2** and 1-67 indicates a band corresponding to full length protein.

Expressions of the F19X IGF-II and F28X IGF-II constructs in AIM displayed strong cell culture growth (OD_{600nm} > 2), and produced bands at 7.5 kDa, indicating the expression of full length protein (refer to Table 7). In contrast, the analogous expressions carried out using the pEB-CouRS vector, produced low cell culture growth (OD_{600nm} < 1) and truncated protein (refer to Section 3.2.4). Full length protein expression was strongest when the expression was carried out in AIM supplemented with **2.2** (5 mM) and incubated at 37 °C (refer to Table 7 and Figure 28). These expression conditions gave rise to intense bands at 7.5 kDa which produced strong fluorescence when irradiated at 254 nm (refer to Figure 28). These results

confirmed the successful incorporation of **2.2** and thus expression of the full length F19Cou IGF-II (**3.1**) and F28Cou IGF-II (**3.2**) proteins.

These results demonstrate the pRFSDuet™ vector was successful at overcoming the limitations with protein truncation encountered with the pEB-CouRS vector. Therefore it can be concluded that increasing the expression of the *Mj*CouRS and *Mj*tRNA^{Cou} genes, allowed the *o*-tRNA (*Mj*tRNA^{Cou}) to outcompete the RF1 for the mRNA and permitted the synthesis of full length protein. This also demonstrates the suitability of the pRFSDuet™ vector as a generic expression vector for the expression of novel aaRS and tRNA pairs for the efficient incorporation of unnatural amino acids into proteins.

Due to the significant quantities of **2.2** required for large scale expression, only expression of the F19Cou IGF-II protein (**3.1**) was undertaken, as it exhibited stronger full length protein expression compared to the F28Cou IGF-II protein (**3.2**) (refer to Figure 28).

First a medium scale expression was conducted to determine the suitability of these conditions on a larger scale. The F19X IGF-II construct was expressed in AIM, supplemented with **2.2** (5 mM), ampicillin (100 µg/ml) and kanamycin (30 µg/ml), and incubated at 37 °C. SDS-PAGE analysis of the resultant cultures again revealed strong bands at 7.5 kDa, which indicated successful expression of full length protein (data not shown). Consequently a large scale expression of the F19Cou IGF-II protein (**3.1**) was undertaken.

3.2.7 Expression of the F19Cou IGF-II protein (3.1)

The F19X IGF-II construct was expressed in AIM, supplemented with **2.2** (4 mM)[ix], ampicillin (100 µg/ml) and kanamycin (30 µg/ml), and incubated at 37 °C for 18 h. The F19Cou IGF-II protein was expressed as a fusion protein with the first 11 amino acids of porcine growth

[ix] A final concentration of 4 mM was used due to limited quantities of **2.2**. Based on previous expression data (refer to Section 3.2.6), a decrease in concentration from 5 mM to 4 mM was not expected to comprise the expression.

hormone ([Met 1]-pGH-(1–11)-PAPM) (refer to Section 3.2.2). Inclusion bodies (IBs) were isolated as described by King *et al.*[171] and purified according to the methods described by Francis *et al.*[173] and Delaine *et al.*[23]

A summary of the purification and isolation of the recombinant F19Cou IGF-II protein (**3.1**) from IBs is summarised in Table 8. Firstly IBs as shown in Figure 29A, were purified using gel filtration and fractions containing the F19Cou IGF-II fusion protein were identified by analytical RP-HPLC and pooled prior to refolding (Figure 29B). The F19Cou IGF-II fusion protein was folded following the methods outlined by Delaine *et al.*[23] with the crude folded fusion protein shown in Figure 29C. The [Met1-pGH(1-11)-PAPM] fusion partner was removed from the folded F19Cou IGF-II protein by enzymatic cleavage using Prag A9.[172] The crude Prag A9 cleavage mixture is shown in Figure 29D and was purified using semi-preparative RP-HPLC.[184] The purified recombinant F19Cou protein (**3.1**) is shown in Figure 30A and contains a minor unknown impurity as evdenced by a small shoulder peak. Protein **3.1** was confirmed from the good agreement of the *m/z* ions ($[M+6H]^{+6}$, 1262.12 *m/z*) displayed in the HRMS shown in Figure 30B with the calculated *m/z* ions ($[M+6H]^{+6}$, 1262.24 *m/z*).

Table 8: Summary of the purification of the recombinant F19Cou IGF-II (3.1). Quantification of recombinant F19Cou IGF-II (**3.1**) was carried out at each step out by comparing analytical C4 RP-HPLC profiles with profiles of standard Long^{3}R^{3}IGF-I preparations.[185]

Figure 29 Reference	Stage of purification	F19Cou IGF-II (3.1) (µg)
A	IB's	875
B	Gel filtration	326
C	Refold	215
	Prag A9 cleavage	
	90 min	118
	135 min	153
D	165 min (stop)	230
	Final Purification	
Figure 30	RP-HPLC	44

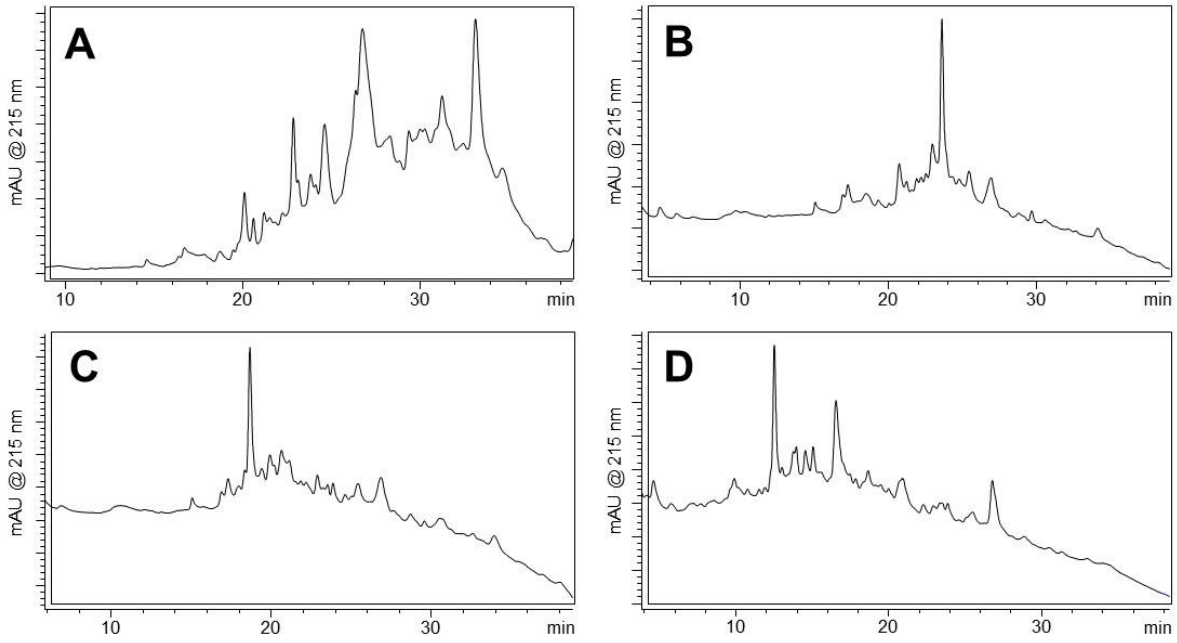


Figure 29: Isolation and purification of the recombinant F19Cou IGF-II protein (3.1). Analytical RP-HPLC traces at each stage of purification are shown. a) Representative sample of inclusion bodies; b) Isolation of F19Cou IGF-II fusion protein after gel filtration; c) Refolding of the F19Cou IGF-II fusion protein after 90 min; d) Enzymatic cleavage of the [Met1-pGH(1-11)-PAPM] fusion partner from the folded F19Cou IGF-II protein after 165 min.

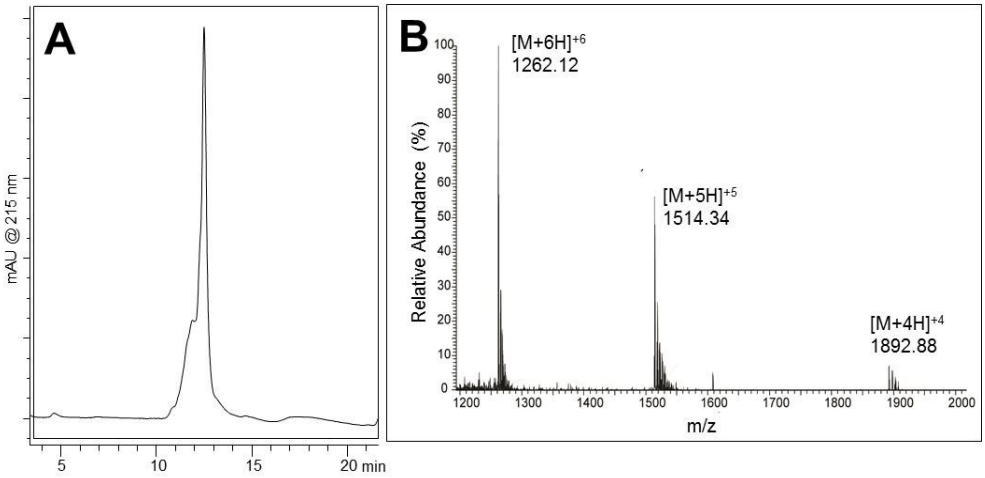


Figure 30: Characterisation of the recombinant F19Cou IGF-II protein (3.1). a) Analytical RP-HPLC trace of **3.1**; b) HRMS of **3.1**. Calcd. for $C_{325}H_{501}N_{93}O_{104}S_6$: 7567.4292 (average isotopes); observed: 1892.88 ($[M+4H]^{+4}$), 1514.34 ($[M+5H]^{+5}$) and 1262.12 ($[M+6H]^{+6}$).

3.2.8 Biological activity of the recombinant F19Cou IGF-II protein (3.1)

Competition binding assays were conducted to determine the affinity of the recombinant F19Cou IGF-II protein (**3.1**) for the IGF-1R and IR-A. Receptor binding experiments were carried out on immunocaptured IR-A and IGF-1R as described by Denley *et al.*[22], where increasing concentrations of the F19Cou IGF-II protein (**3.1**) were added to a constant concentration of europium labelled IGF-II (EuIGF-II).

The competition binding curves are illustrated in Figure 31 and the IC_{50} values are summarised in Table 9. The IC_{50} values for the binding of F19Cou IGF-II protein (**3.1**) to the IGF-1R and IR-A were determined to be 47.6 ± 0.3 nM and 96.9 ± 0.3 nM, respectively. Significantly, compared to the recombinant native IGF-II protein, the recombinant F19Cou IGF-II protein (**3.1**) exhibited only a 8-fold decrease in affinity for the IGF-1R (5.9 nM versus 47.6 nM) and a 10-fold decrease in affinity for the IR-A (8.9 nM versus 96.9 nM).

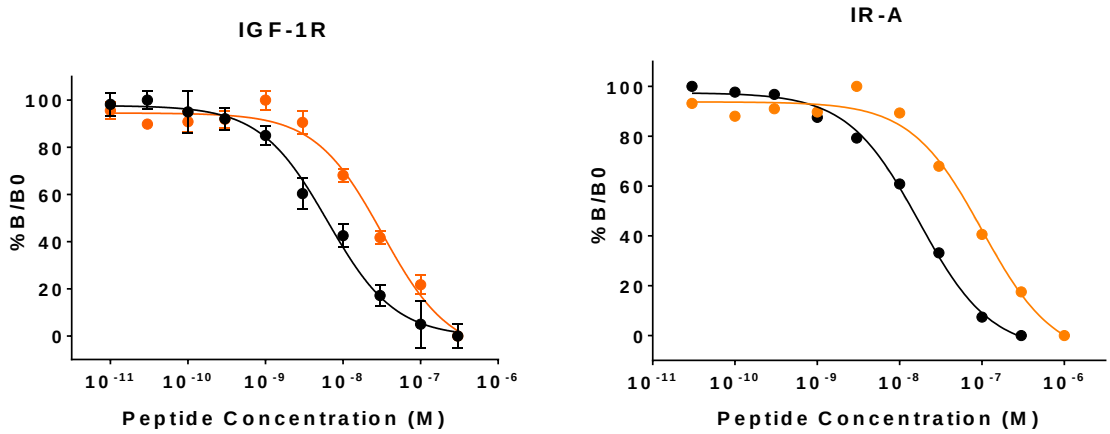


Figure 31: Competitive binding curves for the recombinant F19Cou IGF-II protein (3.1) binding to immunocaptured IGF-IR and IR-A. Immunocaptured IGF-1R or IR-A were incubated with Europium labelled IGF-II (EuIGF-II) in the presence of, or absence of, increasing concentrations of recombinant native IGF-II protein (•) (*black*) or **3.1** (•) (*orange*). Results are expressed as a percentage of binding in the absence of competing ligand (B_0). Graphs show data pooled from three separate experiments and each data point is measured in triplicate per experiment. Data is shown as the mean ± S.E. Error bars are shown when greater than the size of the symbols

Table 9: IC_{50} values derived from competitive binding assays of the recombinant native IGF-II protein and recombinant F19Cou IGF-II protein (3.1) to immunocaptured IGF-1R and IR-A. Where the affinity relative to IGF-II is the IC_{50} relative to that of IGF-II binding to the IGF-1R or IR-A (IC_{50} IGF-II/ IC_{50} IGF-II analogue) and is expressed as percentage of IGF-II binding. IGF-II ± S.E is derived from at least 3 separate experiments performed in triplicate.

	IC_{50} values from immunocaptured IGF-1R and IR-A (nM)		
	recombinant native IGF-II	F19Cou IGF-II (3.1)	Affinity relative to recombinant native IGF-II (%)
IGF-1R	5.9 ± 0.4	47.6 ± 0.3	12.5
IR-A	8.9 ± 0.4	96.9 ± 0.3	9.2

The decreases in receptor binding reported here for the F19Cou IGF-II protein (**3.1**) are consistent with the affinities reported for other recombinant F19X IGF-II analogues.[23,44] IGF-II analogues where Phe^{19} was mutated to Ala, Ser, Tyr or Leu, exhibited between a 1.4-2.8-fold lower affinity for the IGF-1R and 2.2-3.2-fold for the IR-A compared to the recombinant native IGF-II protein. Furthermore these decreases in affinity for the F19Cou IGF-II protein (**3.1**) binding to the IGF-1R and IR-A were expected, as Phe^{19} is involved in receptor binding (refer to Section 1.1.4, Chapter 1).[44] Specifically, these decreases in affinity are proposed to result from the inability of the receptor binding pockets (IGF-1R and IR-A) to accommodate a larger and bulkier group at position 19.

3.3 Conclusion

The pEB-CouRS expression vector was initially investigated for use in the expression of two novel fluorescent IGF-II proteins, the F19Cou IGF-II (**3.1**) and F28Cou IGF-II (**3.2**). Several conditions were optimised for the expressions carried out using the pEB-CouRS vector, and included the medium, expression temperature, IPTG concentration and concentration of **2.2**. Full length protein expression was strongest when expressions were carried out in MIN supplemented with **2.2** (1 mM), ampicillin (100 µg/ml) and tetracycline (50 µg/ml), incubated at 25 °C and induced into protein expression using IPTG (0.05 mM). Despite these encouraging results, problems with cell culture death, misincorporation of endogenous amino acids and protein truncation were continually encountered.

These problems resulted from the low level expression of the *o*-tRNA and *o*-aaRS (*Mj*tRNA^{Cou} and *Mj*CouRS) by the pEB-CouRS vector. Low expression reduces the concentration of the *o*-tRNA and *o*-aaRS in the cell, and thus increases the probability of the RF1 binding to the UAG codon before the *o*-tRNA. RF1 binding in turn causes protein truncation. It was concluded improving the expression of the *o*-tRNA and *o*-aaRS, would increase the concentration of *o*-tRNA and *o*-aaRS, reduce the competition between *o*-tRNA and the RF1 and thus prevent protein truncation by the RF1.

Expression of the *o*-tRNA and *o*-aaRS was improved by changing the expression vector from the pEB-CouRS to a superior pRFSDuet™ vector, which contains a high-copy number RSF replicon and a strong T7*lac* promoter. Conditions for the expression of the F19Cou IGF-II (**3.1**) and F28Cou IGF-II (**3.2**) proteins were optimised. Full length protein expression was strongest when expressions were carried out using AIM, supplemented with **2.2** (5 mM), ampicillin (100 μg/ml) and kanamycin (30 μg/ml), and incubated at 37 °C. Analogous expressions using the pEB-CouRS vector produced truncated protein. Thus, the successful expression of the F19Cou IGF-II (**3.1**) and F28Cou IGF-II (**3.2**) proteins was attributed to the use of pRFSDuet™ vector. Specifically, it was thought the increased expression of the *o*-tRNA and *o*-aaRS, allowed the *o*-tRNA (*Mj*tRNA^{Cou}) to outcompete the RF1 for the mRNA and thus enabling full length protein to be expressed.

After successful small scale expressions using the pRFSDuet™ vector, a large scale expression was carried out. The large quantities of **2.2** required for expression limited the large scale expression to one fluorescent IGF-II protein. The F19Cou IGF-II (**3.1**) protein was chosen as it exhibited stronger full length protein expression.

Expression of the F19Cou IGF-II protein (**3.1**) was carried out in AIM, supplemented with **2.2** (4 mM), ampicillin (100 μg/ml) and kanamycin (30 μg/ml), and incubated at 37 °C. The F19Cou IGF-II protein (**3.1**) was isolated from IB's and refolded into its biologically active form. The IC_{50} values for F19Cou IGF-II protein (**3.1**) binding to the IGF-1R and IR-A were determined to be 96.9 ± 0.3 nM and 47.6 ± 0.3 nM, respectively. These decreases in affinity for the IGF-1R and IR-A compared to the recombinant native IGF-II protein were consistent

with what has been reported for other F19X IGF-II analogues. Finally, the successful expression of the F19Cou IGF-II protein (**3.1**) using the pRFSDuet™ vector, also demonstrated the applicability of this vector as a generic method for outcompeting the RF1 and improving full length protein expression of non-native proteins.

In summary, the use of the pRFSDuet™ vector for the expression of the *Mj*CouRS and *Mj*tRNA^{Cou} genes, overcame the problems with misincorporation of endogenous amino acids and protein truncation. Furthermore, the pRFSDuet™ vector allowed for the successful expression of a novel fluorescent IGF-II protein, F19Cou IGF-II (**3.1**), which binds both the IGF-1R and IR-A with nanomolar affinity. However investigation of the high affinity binding of IGF-II to the IGF-1R and IR-A using FRET, required increased quantities of the F19Cou IGF-II protein (**3.1**). Thus, an alternative SPPS-based approach to the synthesis of the F19Cou IGF-II protein (**3.1**) and the unattained F28Cou IGF-II protein (**3.2**) was investigated and is now described.

Chapter 4

Linear synthesis of a novel fluorescent IGF-II analogue

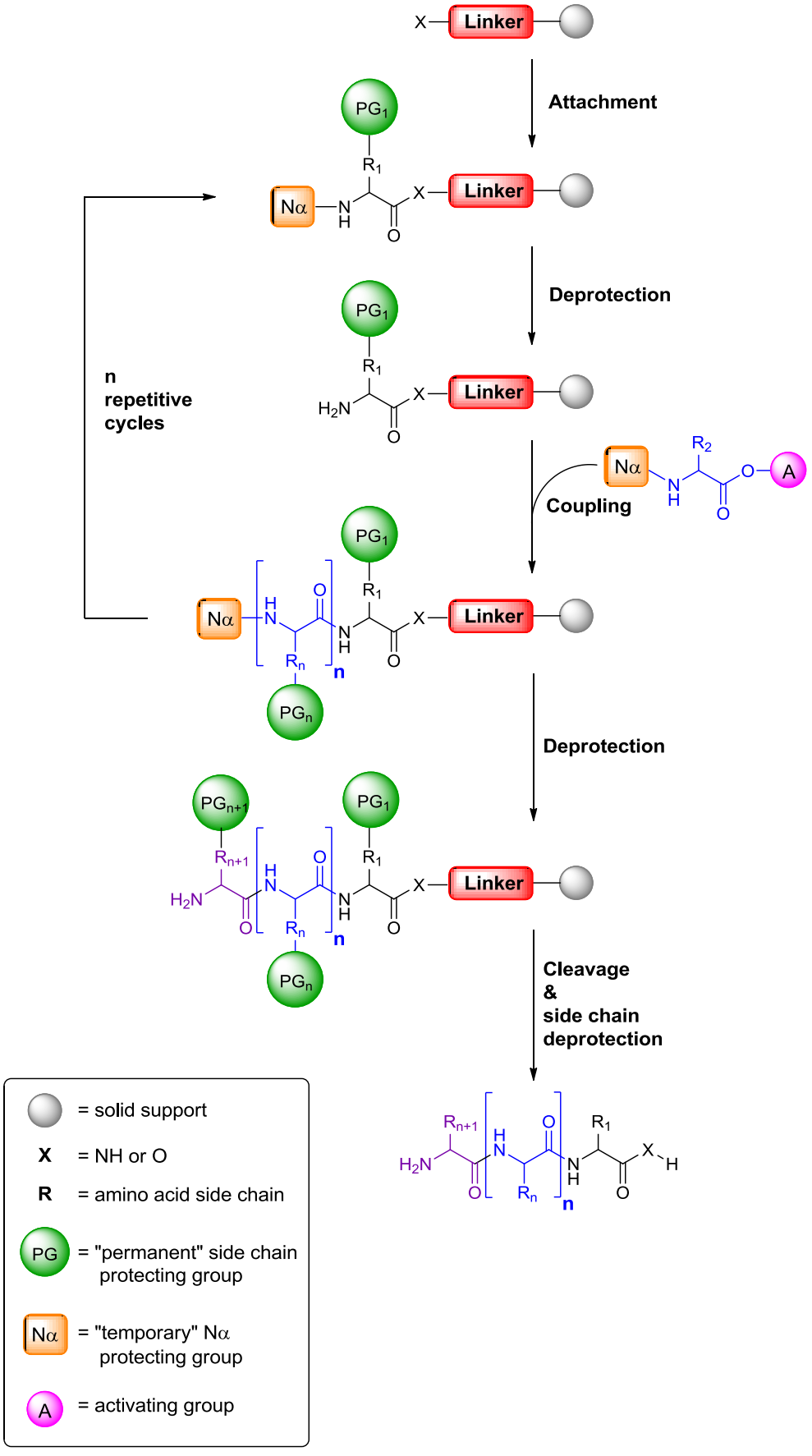
4.1 Introduction

As discussed in Chapter 3, recombinant protein expression yielded insufficient quantities of the F19Cou IGF-II protein (**3.1**) for the proposed FRET study. Furthermore expression of the F28Cou IGF-II protein (**3.2**) was impractical due to the large quantities (1-5 g) of coumaryl amino acid **2.2** required. Thus, the next logical approach was to investigate an alternative solid phase peptide synthesis (SPPS) methodology to overcome these limitations. SPPS has many advantages including increased efficiency, speed and yield, a simplified isolation process and allows facile introduction of the coumarin-based fluorophore at any stage during the synthesis. Furthermore it requires lower quantities (100-200 mg) of amino acid **2.2** than recombinant protein expression.

4.1.1 Background to solid phase peptide synthesis (SPPS)

4.1.1.1 Methodology

Solid phase peptide synthesis (SPPS), involves the sequential addition of amino acids to a solid support, through a series of iterative coupling and deprotection steps (Scheme 7).[186-188] The assembly is from the C to the N-terminus and the methodology uses an orthogonal protecting group strategy. The N-terminal amino functionality is protected with a 'temporary' protecting group (orange; Scheme 7), while the reactive amino acids side chain moieties are protected with more 'permanent' protecting groups (green; Scheme 7).[187,189] The N-terminal protecting group is removed in a deprotection step after each coupling step to allow for the next cycle. Once completed, the peptide is concomitantly deprotected and removed from the solid support on treatment with strong acid.



Scheme 7: Solid phase peptide synthesis methodology. Scheme adapted from Amblard *et al.*[190] and Cudic *et al.*[187]

Table 10 summarises the many advantages SPPS has over traditional solution phase chemistry. The use of a solid support in SPPS overcomes the solubility limitations associated with solution phase chemistry, while also minimising the number of by-products and simplifying product isolation. Solid phase synthesis can also be automated to greatly reduce the synthesis time, which is beneficial for the synthesis of long sequences (> 30 residues) like IGF-II (67 residues). Thus, a SPPS methodology was selected for the chemical synthesis of the native IGF-II protein (**4.1**) and the non-native IGF-II analogue, F19Cou IGF-II (**4.2**).

Table 10: Comparison of solution and solid phase methods for the chemical synthesis of peptides. The main differences between solution phase and solid phase methods are summarised. Table adapted from Kent.[188]

	Method	
	Solution Phase	**Solid Phase**
Strategy	Segment/convergent	Stepwise
Solubility problems	Frequent/serious	Occasional/minor
Side chain protection	Maximal	Maximal
Chemistry	Highly specialised	Simple/general
Automation	No	Yes
Purity	Excellent	Acceptable

4.1.1.2 Protecting group strategies: Boc/Bzl and Fmoc/*t*Bu

Two main protecting group strategies have been developed for SPPS, Boc/Bzl and Fmoc/*t*Bu.[189,191-193] Both these strategies have been employed in the work described in this thesis. Specifically, the Fmoc/*t*Bu approach was used in the work described in this chapter and the Boc/Bzl strategy was employed in the synthesis of the IGF-II analogues described in Chapter 5. A comparison of these strategies is presented in Table 11, and summarises the advantages and disadvantages of each method. In general, the Boc/Bzl approach is successful in the synthesis of long and difficult sequences. However this method requires HF to cleave the peptide from the resin. This is a major limitation of the Boc/Bzl approach, as HF requires specialised HF-resistant apparatus, which is expensive and uncommon to most laboratories.[187,194-196] In contrast, the Fmoc/*t*Bu strategy is often preferred for routine synthesis

as it does not require the use of specialised equipment. The base/acid orthogonality of the Fmoc/*t*Bu approach also provides the opportunity for on-resin side chain functionalisation.[187,194-196]

Table 11: Comparison of the Boc and Fmoc protecting schemes. The main differences between Boc-SPPS and Fmoc-SPPS are summarised. Table adapted from Kent.[187,188,194-196]

	Protecting group strategy	
	Boc/Bzl	**Fmoc/*t*Bu**
***N*α Deprotection**	TFA	Piperidine
Peptide Cleavage	HF	TFA
Specialised Equipment	HF-resistant polymer Kel-F	Standard glassware
Recommended for	Long or Difficult sequences prone to aggregation Base sensitive peptides	Acid sensitive peptides Labelled or side chain modified peptides

The Boc/Bzl strategy employs graduated acid lability for both the N-terminal and side chain protecting groups.[189,191] The N-terminal Boc deprotection is achieved by treatment with TFA, whereas contaminant cleavage from the resin and deprotection of the benzyl-based side chains is achieved with HF. In contrast, the Fmoc/*t*Bu strategy uses orthogonal protecting groups. It uses the base-labile N-terminal protecting group, Fmoc and acid-labile side chain protecting groups, such as *t*Bu and Boc.[189,192,193] A comparison of the acid-labile side chain protecting groups used in both strategies is shown in Table 12.[187,188,197]

Table 12: Common side chain protecting groups used in SPPS. The acid-labile side chain protecting groups used in Boc-SPPS and Fmoc-SPPS are summarised. Table adapted from Cudic *et al.*[187], Kent[188] and Isidro-Llobet *et al.*[197]

	Amino acid side chain protecting group	
	Boc-SPPS	**Fmoc-SPPS**
Arg	Tos	Pbf
Asn	Xan	Trt
Asp	OBzl, O-2-Ada, OcHx	O*t*Bu,
Cys	MeBzl	Acm, *t*Bu, Trt
His	Tos, Dnp	Trt
Gln	Xan	Trt
Glu	OBzl	O*t*Bu
Lys	2-Cl-Z	Boc
Met	Met(O)	Met(O)
Ser/Thr	Bzl	*t*Bu
Tyr	2-Br-Z	*t*Bu
Trp	CHO	Boc

A chemical synthesis of native human IGF-II protein has been reported by Li and co-workers.[198,199] The synthesis used an automated Boc/Bzl methodology in combination with acid stabile protecting groups for Cys and Lys. The protection of the Cys and Lys residues produced a crude peptide with limited solubility and necessitated a complicated, multi-step, post-cleavage deprotection. Furthermore, isolation of the desired synthetic IGF-II protein from the crude peptide mixture appeared inefficient as a lengthy, multi-step purification procedure was reported. In addition to the tedious purification, the synthetic IGF-II protein was produced in low yield (2%), however it was of high purity and indistinguishable from the human IGF-II protein.[198,199] The related chemical synthesis of four mutant IGF-II analogues (Leu^{27} IGF-II; Ser^{27} IGF-II; Gln^{6},Ala^{7},Tyr^{18},Leu^{19} IGF-II and Gln^{6},Ala^{7},Tyr^{18},Leu^{19}, Leu^{27} IGF-II) has been reported by Oh *et al.*[200] In this synthesis an automated Boc/Bzl approach was again employed and all residues were protected with standard protecting groups (refer to Table 12). The resulting non-native IGF-II proteins were isolated by a similar multi-step purification procedure to that described above by Li and co-workers[198,199], and were reported without yields and limited characterisation data. In both instances the methods described for the synthesis and isolation of the IGF-II analogues were lengthy, inefficient and low yielding. Thus, the development of a shorter, more efficient and robust method for the synthesis of the

native IGF-II protein and fluorescent IGF-II analogues, based on a linear Fmoc-SPPS approach was proposed.

This chapter describes the linear synthesis of the native IGF-II protein (**4.1**) and a fluorescent IGF-II analogue, F19Cou IGF-II (**4.2**). The target proteins are shown in Figure 32, and were synthesised using a linear Fmoc-SPPS protocol. The F19Cou IGF-II analogue (**4.2**) contains the coumaryl amino acid **2.2** at position 19 (refer to Figure 32), to facilitate the FRET-based investigation of the binding of IGF-II to IGF-1R and IR-A, as discussed in Chapters 1 and 3.

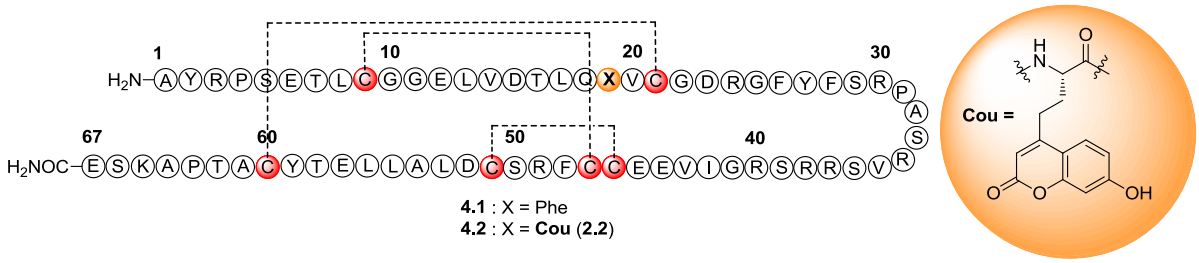


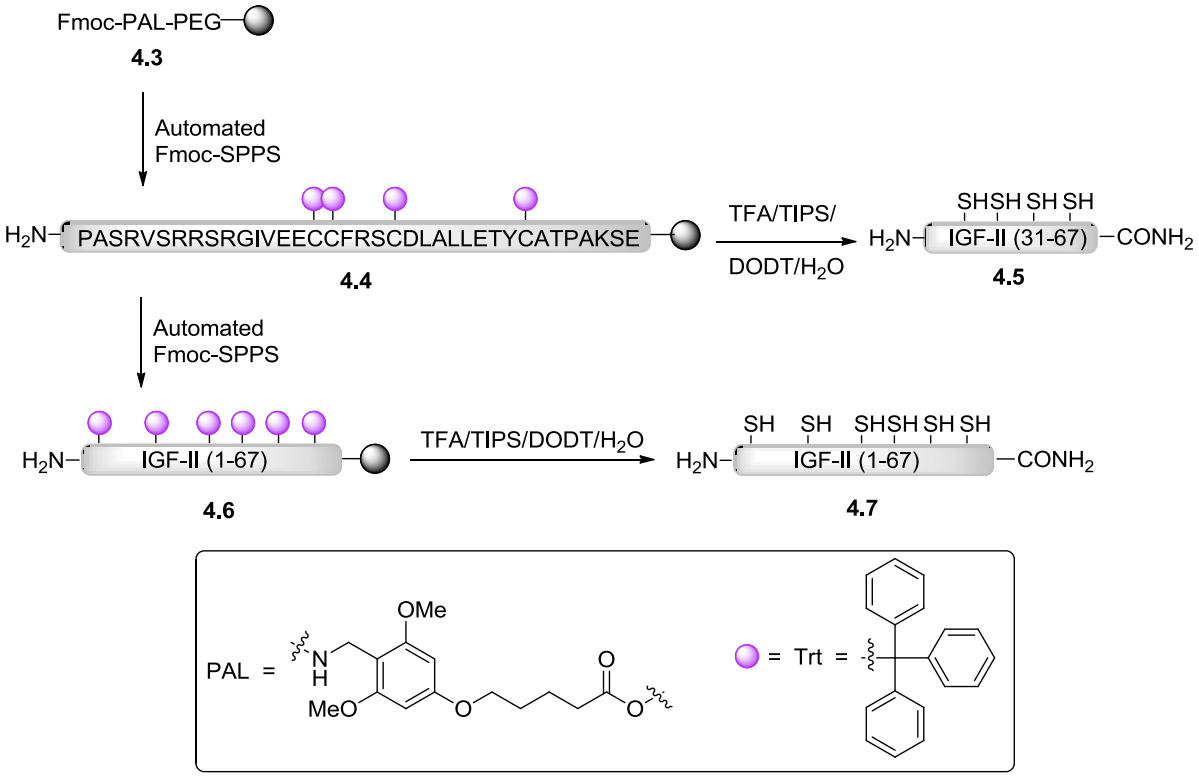
Figure 32: Target IGF-II proteins (4.1 and 4.2). Sequence of IGF-II is shown with the residue numbers annotated in *bold* above the sequence; Cys residues are shown in *red*; disulfide bonds are shown as *dashed lines* and the location of the coumarin-based fluorophore (Cou; **2.2**) is denoted by an *X*, highlighted in *orange* and illustrated within the *orange* circle (*right*).

4.2 Results and discussion

The synthesis of the native IGF-II (**4.1**) and the F19Cou IGF-II (**4.2**) proteins were attempted using a linear Fmoc-SPPS protocol. The native IGF-II sequence is depicted in Figure 32, with the Cys residues highlighted in red. Also displayed, is the position of coumaryl amino acid **2.2** in the F19Cou IGF-II protein (**4.2**) (orange; Figure 32).

4.2.1 Six trityl protecting group strategy (6 Trt)

Synthesis of native IGF-II protein (**4.1**) (refer to Figure 32) was initially attempted using a global trityl (Trt) protection approach as depicted in Scheme 8. In this approach all six cysteines of the native IGF-II peptide (Cys^{9}, Cys^{21}, Cys^{46}, Cys^{47}, Cys^{50} and Cys^{61}) (refer to Figure 32) were Trt protected. The Trt group is displayed in Scheme 8 and is a very acid-labile protecting group. Thus, it was expected to be deprotected during cleavage of the peptide from the resin, to give a peptide with six free sulfhydryl's (refer to Scheme 8).



Scheme 8: Synthesis of the native IGF-II (1-67) peptide (4.7), using the six trityl (Trt) protecting group strategy. Where PEG = Polyethylene glycol; TIPS = triisopropylsilane and DODT = 3,6-dioxa-1,8-octanedithiol.

The IGF-II (1-67; 6 Trt) resin-bound peptide (**4.6**) was assembled as depicted in Scheme 8, from commercially available Fmoc-Pal-PEG-PS resin (**4.3**) using automated peptide synthesis. First, resin **4.3** was elongated using automated peptide synthesis to give the IGF-II (31-67; 4 Trt) resin-bound peptide (**4.4**). A sample of resin **4.4** was cleaved with a cocktail of

TFA/TIPS/DODT/water and the resultant crude peptide mixture was analysed by ESI-MS. The ESI-MS is depicted in Figure 33A, and displays *m/z* ions ($[M+5H]^{+5}$; 822.00 *m/z*), that are consistent with the expected *m/z* ions ($[M+5H]^{+5}$; 821.94 *m/z*) for the IGF-II (31-67) peptide (**4.5**). This result indicates that the IGF-II (31-67) peptide (**4.5**) was present in the crude mixture. Encouraged by this result, the IGF II (31-67; 4 Trt) resin-bound peptide (**4.4**) was elongated to the full-length IGF-II (1-67; 6 trt) resin-bound peptide (**4.6**) using automated peptide synthesis. Resin **4.6** was cleaved using a TFA/TIPS/DODT/water mixture and the crude peptide was analysed by RP-HPLC and ESI-MS. The RP-HPLC analysis is shown in Figure 33C, and suggests a complex mixture of products, as indicated by the presence multiple overlapping peaks and the absence of a discernible major product. The ESI-MS is shown in Figure 33B, and displays *m/z* ions ($[M+5H]^{+5}$; 1496.72 *m/z*) that are in good agreement with the calculated *m/z* ions ($[M+5H]^{+5}$; 1495.89 *m/z*) for the IGF-II (1-67) peptide (**4.7**). These results suggest the desired IGF-II (1-67) peptide (**4.7**) was present as a minor component of a complex mixture.

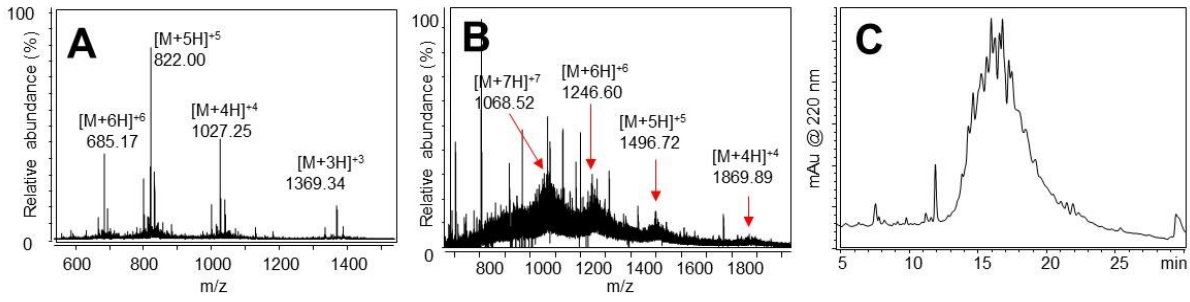


Figure 33: Analysis of the native IGF-II peptide (4.7) synthesised using the six trt protecting group strategy. a) ESI-MS of crude peptide **4.5**. Calcd. for $C_{171}H_{284}N_{54}O_{55}S_4$: 4104.6867 (average isotopes); observed: *m/z* 1369.34 ($[M+3H]^{+3}$), 1027.25 ($[M+4H]^{+4}$), 822.00 ($[M+5H]^{+5}$) and 685.17 ($[M+6H]^{+6}$). b) ESI-MS of the crude peptide **4.7**. Calcd. for $C_{321}H_{500}N_{94}O_{100}S_6$: 7468.3875 (average isotopes); observed: 1869.89 ($[M+4H]^{+4}$), 1496.72 ($[M+5H]^{+5}$), 1246.60 ($[M+6H]^{+6}$) and 1068.52 ($[M+7H]^{+7}$). c) RP-HPLC of crude peptide **4.7**.

Based on the presence of the expected *m/z* ions within the ESI-MS shown in Figure 33B, purification of **4.7** from the mixture was attempted using semi-preparative RP-HPLC.

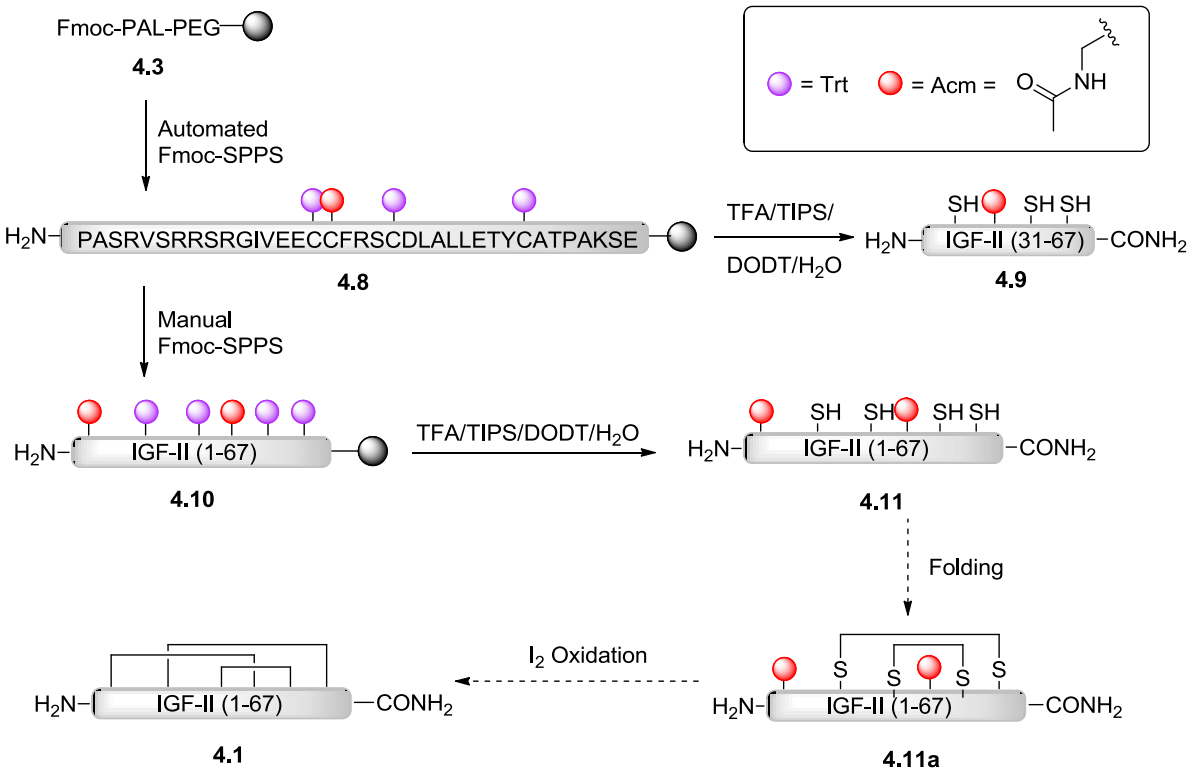
Multiple fractions were collected and analysed by ESI-MS (data not shown), however the calculated *m/z* ions for desired IGF-II (1-67) peptide (**4.7**) were absent in all fractions. The complexity of the RP-HPLC trace suggests that a mixture of both truncated and higher mass peptide by-products had formed. The truncated by-products likely resulted from incomplete Fmoc deprotection and/or incomplete couplings in the assembly of the N-terminus of the peptide. The higher mass products are thought to be the result of the free cysteine thiols forming complex oligomers and disulfide adducts during the TFA cleavage, ether precipitation or lyophilisation. The large number of by-products in the crude peptide mixture led us to investigate an alternative protecting group strategy in which the cysteine thiols were protected with a moiety which would be stable to acidic cleavage conditions (> 90% TFA). The acetamidomethyl (Acm) cysteine protecting group was selected as it is more stable than Trt to acid cleavage,[201] and can be selectively deprotected using iodine, thallium trifluoracetate or silver trifluoromethanesulfonate salts.[202-205] It was anticipated protection of the cysteine sulfhydryl's during the cleavage of the peptide from the resin, would potentially minimise disulfide adducts and assist in the isolation of the native IGF-II peptide (**4.7**).

4.2.2 Double Acm protecting group strategy (2 Acm, 4 Trt)

The large number of by-products observed in the crude mixture from the synthesis of the native IGF-II peptide (**4.7**) (refer Section 4.2.1) was potentially caused by the formation of disulfide adducts during cleavage of the peptide from the resin. Consequently, it was envisaged maintaining protection of one disulfide pair during the cleavage would minimise by-product formation.

The IGF-II sequence contains six cysteines and three disulfide bonds, Cys^{9}/Cys^{47}, Cys^{21}/Cys^{60} and Cys^{46}/Cys^{51} (refer to Figure 32). The Cys^{9}/Cys^{47} disulfide pair was selected for protection with Acm, as it is believed to be the final disulfide bond formed during the folding of IGF-II.[206] It was envisaged iodine oxidation of a partially folded IGF-II (1-67; 2 Acm; S-S) intermediate (**4.11a**), could be performed as the final step to concomitantly remove the Acm groups and form the last disulfide bond (Cys^{9}/Cys^{47}), to afford the desired IGF-II protein (**4.1**) (Scheme

9). Synthesis of the IGF-II (1-67; 2 Acm) peptide (**4.11**) with selective Acm protection of the Cys^{9}/Cys^{47} disulfide pair is now described, as depicted in Scheme 9.



Scheme 9: Synthesis of the IGF-II (1-67; 2 Acm) peptide (4.11), using the 2 Acm, 4 Trt protecting group strategy. Where PEG = Polyethylene glycol; TIPS = triisopropylsilane and DODT = 3,6-dioxa-1,8-octanedithiol.

The IGF-II (1-67; 2 Acm) peptide (**4.11**) was assembled from commercially available Fmoc-PAL-PEG-PS resin (**4.3**) using a combination of both automated and manual peptide synthesis as depicted in Scheme 9. First, the IGF-II (31-67; Acm, 3 Trt) resin (**4.8**) was synthesised from resin **4.3** by automated peptide synthesis. Resin **4.8** was then elongated to the full length IGF-II (1-67; 2 Acm, 4 Trt) resin-bound peptide (**4.10**) using a manual peptide synthesis protocol. The combination of automated and manual synthesis allowed the assembly of the N-terminal region of the peptide to be carefully monitored using a 2,4,6-

trinitrobenzene sulfonic acid (TNBSA) resin test.[207] This demonstrated that each coupling reaction and Fmoc deprotection had proceeded to completion.[x]

The full length IGF-II (1-67; 2 Acm) peptide (**4.11**) was released from resin **4.10** on treatment with TFA/TIPS/DODT/water and the resultant mixture was analysed by RP-HPLC and ESI-MS. Figure 34A reveals a complex mixture of products as evidenced by multiple overlapping and unresolved peaks. A small sample of the mixture was separated by semi-preparative RP HPLC and the resultant fractions were analysed by ESI-MS. ESI-MS analysis of a fraction collected between 12.5-13.5 min is shown in Figure 34B and displays *m/z* ions ($[M+6H]^{+6}$; 1270.00 *m/z*) which are consistent with the predicted *m/z* ions ($[M+6H]^{+6}$; 1270.43 *m/z*) for the IGF-II (1-67; 2 Acm) peptide (**4.11**). This result suggests peptide **4.11** was present as a minor component of the mixture, eluting between 12-14 min in the RP-HPLC trace shown in Figure 34A. Encouraged by this, the remaining peptide mixture was subjected to semi-preparative RP-HPLC. Unfortunately, this gave rise to various unidentified by-products and only small quantities (< 1 mg) of peptide **4.11** were isolated.[xi] The large number of by-products observed in the crude mixture displayed in Figure 34, are again thought to result from the formation of disulfide adducts during cleavage of the peptide from the resin. Consequently, an alternative synthesis in which all six cysteines (Cys^{9}/Cys^{47}, Cys^{21}/Cys^{60} and Cys^{46}/Cys^{51}) (refer to Figure 32) were protected with Acm was undertaken. It was anticipated that protection of the cysteine sulfhydryl's during cleavage would further simplify the crude reaction mixture by preventing the formation of oligiomeric disulfide adducts, thus allowing isolation of the IGF-II peptide.

[x] The TNSBA test is used for the quantification of free primary amines. In the presence of a free primary amine the resin bead appears red/orange in colour.

[xi] The small quantities (< 1 mg) of peptide **4.11** were consumed in subsequent trial Acm deprotection and oxidative folding steps, but no desired native IGF-II protein (4.1) was recovered from these steps.

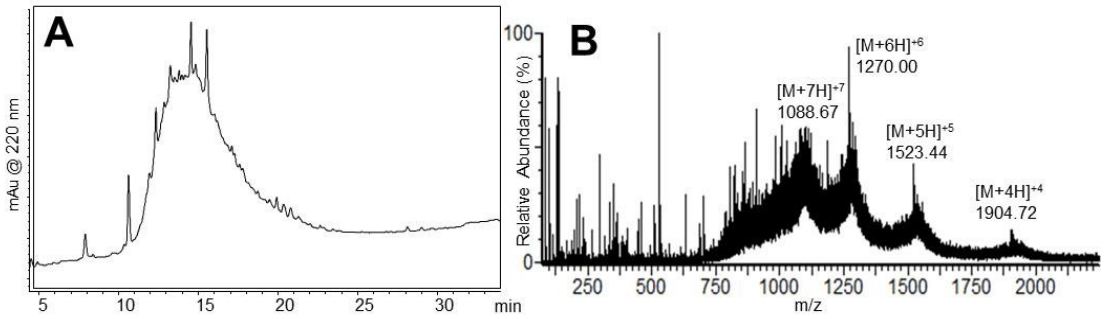
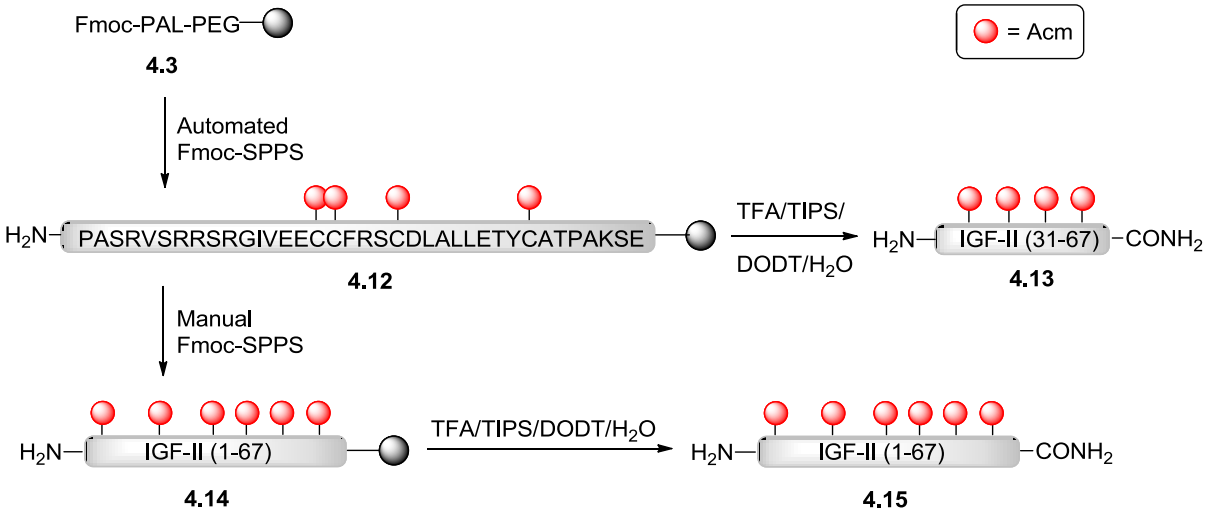


Figure 34: Analysis of the crude IGF-II (1-67; 2 Acm) peptide (4.11) synthesised using the 2 Acm, 4 Trt protecting group strategy. a) RP-HPLC trace of crude peptide **4.11** and b) ESI-MS of a fraction collected between 12.5-13.5 min for peptide **4.11**. Calcd. for $C_{327}H_{516}N_{96}O_{102}S_6$: 7617.5912 (average isotopes); observed: *m/z* 1904.72 ($[M+4H]^{+4}$), 1523.44 ($[M+5H]^{+5}$), 1270.00 ($[M+6H]^{+6}$) and 1088.67 ($[M+7H]^{+7}$).

4.2.3 Six Acm protecting group strategy (6 Acm)

Given the difficulties encountered in the purification of the full length IGF-II (1-67; 2 Acm) peptide (**4.11**), a synthesis of the IGF-II (1-67; 6 Acm) peptide (**4.15**) was undertaken and is outlined in Scheme 10. In this synthesis all six cysteines (Cys^{9}/Cys^{47}, Cys^{21}/Cys^{60} and Cys^{46}/Cys^{51}) (refer to Figure 32) were protected with Acm.



Scheme 10: Synthesis of the IGF-II (1-67; 6 Acm) peptide (4.15), using the 6 Acm protecting group strategy. Where PEG = Polyethylene glycol; TIPS = triisopropylsilane and DODT = 3,6-dioxa-1,8-octanedithiol.

Synthesis of the IGF-II (1-67; 6 Acm) peptide (**4.15**) was conducted as previously described for the IGF-II peptides **4.7** and **4.11** (refer Sections 4.2.1and 4.2.2). First the IGF-II (31-67; 4 Acm) resin-bound peptide (**4.12**) was assembled from resin **4.3** using automated peptide synthesis (refer to Scheme 10). A sample of resin **4.12** was treated with TFA/TIPS/DODT/water and the resulting mixture was analysed by RP-HPLC and ESI-MS. The RP-HPLC trace displayed in Figure 35A shows a major product at 7.5 min plus a few minor by-products. Analysis of the major component by ESI-MS is shown in Figure 35B, and displays *m/z* ions ($[M+4H]^{+4}$; 1098.39 *m/z*) which are consistent with the expected *m/z* ions ($[M+4H]^{+4}$; 1098.26 *m/z*) for the IGF-II (31-67; 4 Acm) peptide (**4.13**).

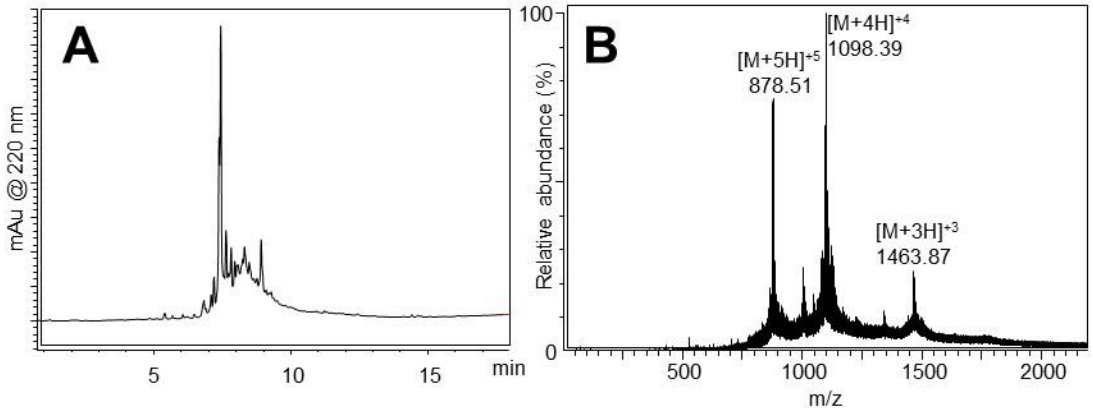


Figure 35: Analysis of the crude IGF-II (31-67; 4 Acm) peptide (4.13). a) RP-HPLC trace of crude peptide **4.13** and b) ESI-MS analysis of the crude peptide **4.13**. Calcd. for $C_{183}H_{304}N_{58}O_{59}S_4$: 4388.9990 (average isotopes); observed: *m/z* 1463.87 ($[M+3H]^{+3}$), 1098.39 ($[M+4H]^{+4}$) and 878.51 ($[M+5H]^{+5}$).

Encouraged by this result, resin **4.12** was elongated to the full length IGF-II (1-67; 6 Acm) resin-bound peptide (**4.14**) using manual peptide synthesis. Manual synthesis again allowed each coupling and deprotection step to be monitored using the TNBSA resin test[207]. Resin **4.14** was cleaved using a TFA/TIPS/DODT/water mixture and the resultant peptide mixture **4.15** was analysed by RP-HPLC and ESI-MS, as shown in Figure 36. The analytical RP-HPLC is shown in Figure 36A and shows a complex mixture as demonstrated by multiple overlapping and unresolved peaks. Surprisingly, ESI-MS analysis (refer to Figure 36B) of a

fraction collected between 13-14 min during a small scale semi-preparative RP-HPLC, displayed *m/z* ions ($[M+7H]^{+7}$; 1129.73 *m/z*) which are consistent with the predicted *m/z* ions ($[M+7H]^{+7}$; 1129.70 *m/z*) for the IGF-II (1-67; 6 Acm) peptide (**4.15**). Encouraged by these results, the crude peptide mixture **4.15** was purified by semi-preparative RP-HPLC. However isolation of peptide **4.15** from the complex mixture again proved challenging, and again only small quantities (< 1 mg) of the full length IGF-II (1-67; 6 Acm) peptide (**4.15**) were isolated.[xii]

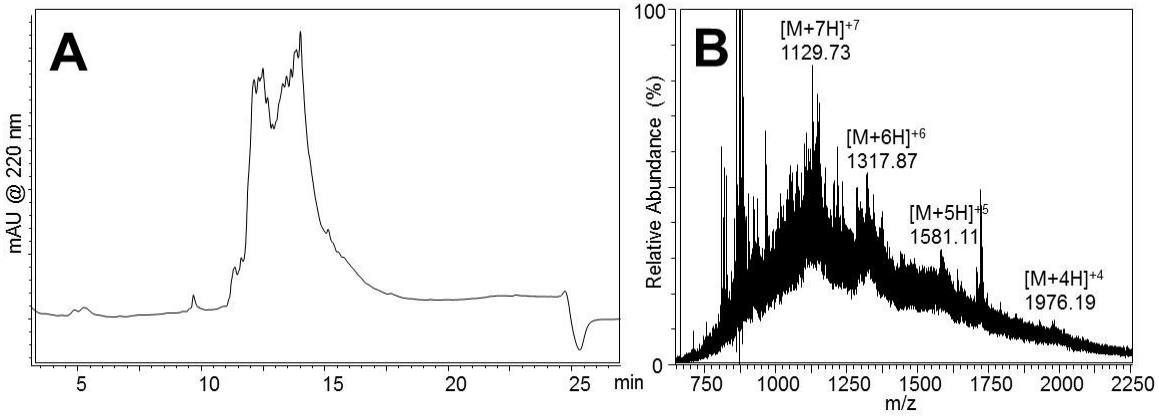


Figure 36: Analysis of the crude IGF-II (1-67; 6 Acm) peptide (4.15). a) RP-HPLC trace of crude peptide **4.15** and b) ESI-MS of a fraction collected between 13-14 min for peptide **4.15**. Calcd. for $C_{339}H_{536}N_{100}O_{106}S_6$: 7900.9035 (average isotopes); observed: *m/z* 1976.19 ($[M+4H]^{+4}$), 1581.11 ($[M+5H]^{+5}$), 1371.87 ($[M+6H]^{+6}$) and 1129.73 ($[M+7H]^{+7}$).

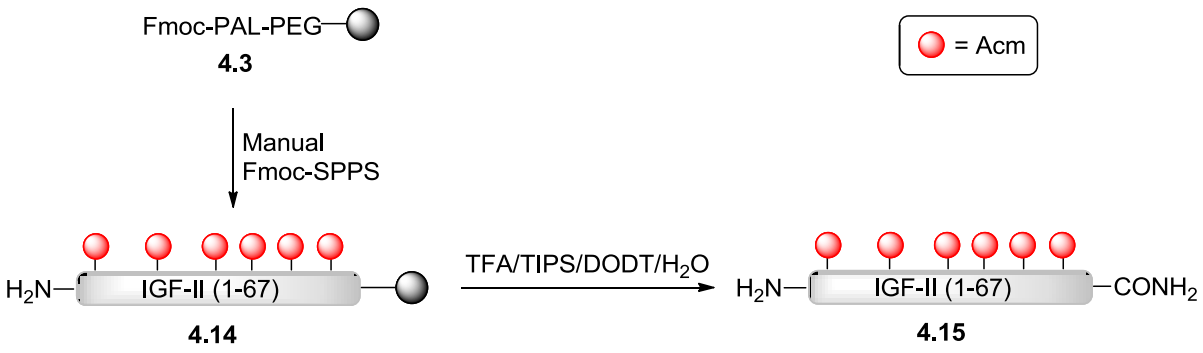
4.2.4 Analysing the assembly of the N-terminal region of the IGF-II peptide

ESI-MS analysis of the crude peptide mixtures confirmed the full length IGF-II peptides **4.7**, **4.11** and **4.15** could be synthesised using a linear Fmoc-SPPS approach (refer to Sections 4.2.1, 4.2.2 and 4.2.3). However, isolation of the purified peptides from their mixtures was hindered by the overwhelming presence of truncated and higher mass peptide by-products.

[xii] The small quantities (< 1 mg) of peptide **4.15** were consumed in subsequent trial Acm deprotection and oxidative folding steps, but no desired native IGF-II protein (**4.1**) was recovered from these steps.

Since RP-HPLC and ESI-MS analysis of the IGF-II (31-67) intermediates (**4.5**, **4.9** and **4.13**) displayed no significant impurities, it was postulated the majority of the truncations and by-products were forming during the assembly of N-terminal region (from residue 31). Consequently, a more systematic analysis of the assembly process was performed in an attempt to identify the problematic regions and residues in the synthesis of the IGF-II peptide.

Samples of resin (5-10 mg) were collected at regular intervals, every 5-10 residues during the manual synthesis of the IGF-II (1-67; 6 Acm) peptide (**4.15**) from resin **4.3** (Scheme 11). These resin samples were cleaved on treatment with TFA/TIPS/DODT/water, and the resultant peptide mixtures were analysed by RP-HPLC and ESI-MS. The analysis was performed to identify at which stage in the synthesis the truncations and by-products became the dominant species.



Scheme 11: Manual synthesis of the IGF-II (1-67) (6 Acm) peptide (4.15). Where PEG = Polyethylene glycol; TIPS = triisopropylsilane and DODT = 3,6-dioxa-1,8-octanedithiol.

The stages in the synthesis where resin samples were collected are depicted in Figure 37. RP-HPLC and ESI-MS analysis of the resin samples from the C-terminal region (residues 30-67) displayed the desired IGF-II peptide intermediates as the major components of the mixture (data not shown). These results suggest the assembly of the IGF-II (1-67; 6 Acm) peptide (**4.15**) was unproblematic up to this point, which was consistent with the synthesis of the other IGF-II (31-67) peptides (**4.5**, **4.9** and **4.13**) (refer to Sections 4.2.1, 4.2.2 and 4.2.3).

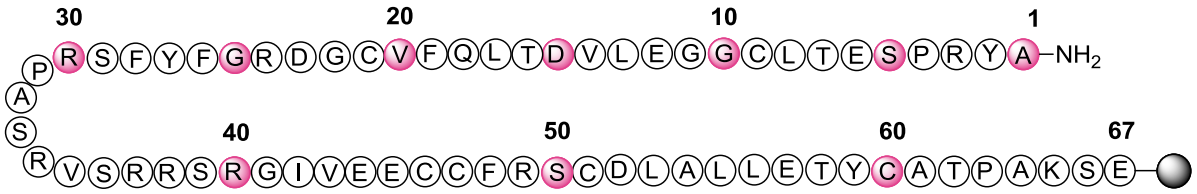


Figure 37: Resin samples collected during the manual synthesis of the IGF-II (1-67; 6 Acm) peptide (4.15). Sequence of IGF-II is shown with the residue numbers annotated above the sequence and residues where small scale cleavages occurred are highlighted in *pink*.

Analysis of the cleavage mixtures from the IGF-II (25-67; 4 Acm), IGF-II (15-67; 5 Acm) and full length IGF-II (1-67; 6 Acm) resin samples are displayed in Figure 38, Figure 39 and Figure 40 respectively. RP-HPLC analysis of the crude IGF-II (25-67; 4 Acm) peptide is displayed in Figure 38A, and shows a distinct peak at 12.2 min. ESI-MS analysis of this peak is shown in Figure 38B and displays *m/z* ions ($[M+5H]^{+5}$; 1130.16 *m/z*) which are in good agreement with the expected *m/z* ions ($[M+5H]^{+5}$; 1130.88 *m/z*) for the desired IGF-II (25-67; 4 Acm) peptide. This result suggests the IGF-II (25-67; 4 Acm) peptide was present as a major component of the crude peptide mixture.

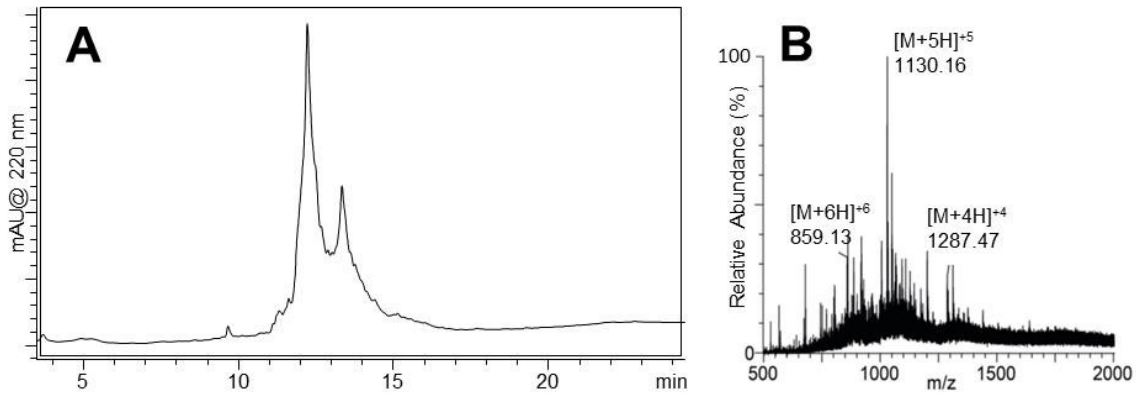


Figure 38: Analysis of the crude IGF-II (25-67; 4 Acm) peptide. a) RP-HPLC trace of the crude IGF-II (25-67; 4 Acm) peptide and b) ESI-MS of the crude IGF-II (25-67; 4 Acm) peptide. Calcd. for $C_{239}H_{381}N_{75}O_{74}S_5$: 5649.3834 (average isotopes); observed: *m/z* 1287.47 ($[M+4H]^{+4}$), 1130.16 ($[M+5H]^{+5}$) and 859.13 ($[M+6H]^{+6}$).

A similar result was also observed in the RP-HPLC and ESI-MS analysis of the crude IGF-II (15-67; 5 Acm) peptide as shown in Figure 39. The RP-HPLC is illustrated in Figure 39A shows a distinct peak at 12.4 min. ESI-MS analysis of this peak is shown in Figure 39B and displays *m/z* ions ($[M+5H]^{+5}$; 1271.64 *m/z*) which are consistent with the calculated *m/z* ions ($[M+5H]^{+5}$; 1271.63 *m/z*) for the IGF-II (15-67; 5 Acm) peptide. This result suggests the desired IGF-II (15-67; 5 Acm) peptide was a major component of the crude mixture.

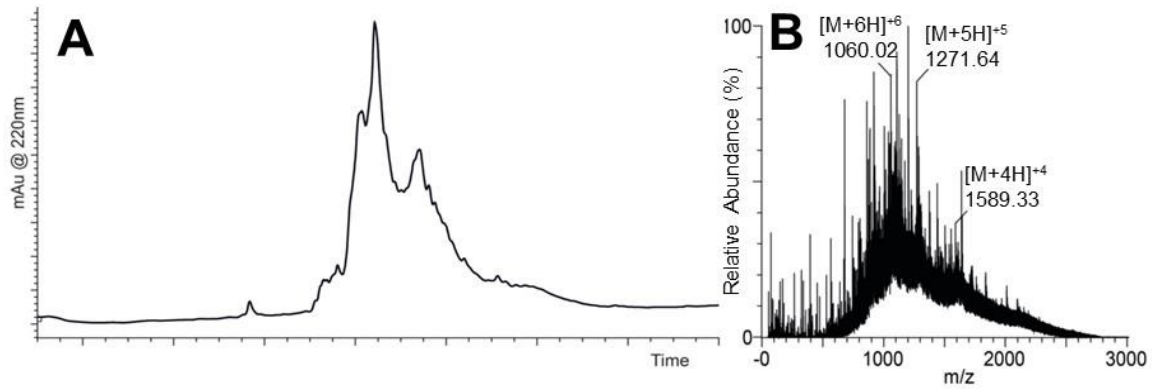


Figure 39: Analysis of the crude IGF-II (15-67; 5 Acm) peptide. a) RP-HPLC trace of the crude IGF-II (15-67; 5 Acm) peptide and b) ESI-MS of the crude IGF-II (15-67; 5 Acm) peptide. Calcd. for $C_{272}H_{430}N_{82}O_{82}S_5$: 6353.1681 (average isotopes); observed: *m/z* 1589.33 ($[M+4H]^{+4}$), 1271.64 ($[M+5H]^{+5}$) and 1060.02 ($[M+6H]^{+6}$).

RP-HPLC and ESI-MS analysis of the resin samples collected up to residue 15 (IGF-II (15-67; 5 Acm)) had been encouraging, as the desired IGF-II peptides were detected as the major components of the crude mixtures. However, analysis of the samples collected during the N-terminal elongation (residues 1-15) revealed the desired IGF-II peptides were only present as a minor component. RP-HPLC analysis the crude IGF-II (1-67; 6 Acm) peptide (**4.15**) is shown in Figure 40A, and shows a complex mixture. ESI-MS analysis of the crude peptide **4.15** is depicted in Figure 40B, and displays *m/z* ions ($[M+7H]^{+7}$; 1129.61 *m/z*) which are consistent with the expected *m/z* ions ($[M+7H]^{+7}$; 1129.70 *m/z*) for the IGF-II (1-67; 6 Acm) peptide (**4.15**). These results suggest the desired full length crude IGF-II (1-67; 6 Acm) peptide (**4.15**) was a minor component of the crude peptide mixture.

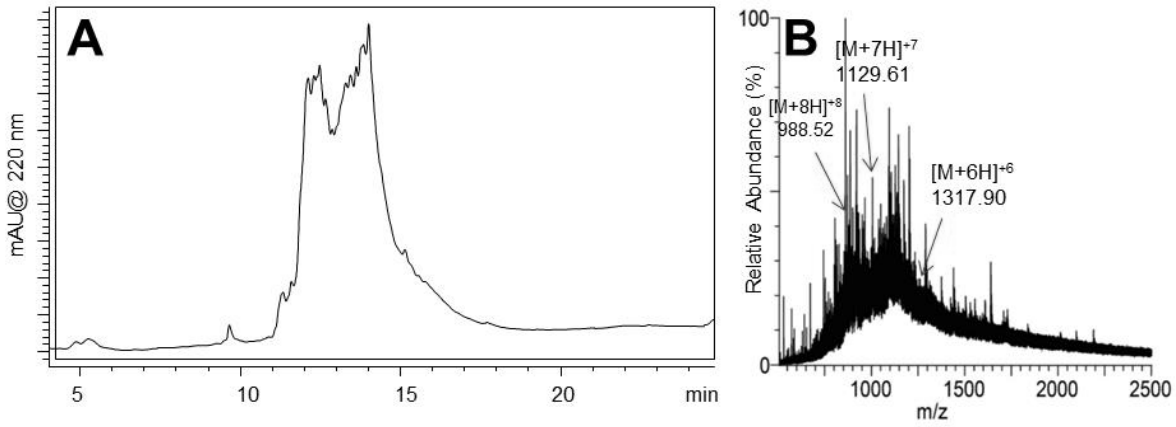


Figure 40: Analysis of the crude IGF-II (1-67; 6 Acm) peptide (4.15). a) RP-HPLC trace of crude peptide **4.15** and b) ESI-MS of crude peptide **4.15**. Calcd. for $C_{339}H_{536}N_{100}O_{106}S_6$: 7900.9035 (average isotopes); observed: *m/z* 1371.90 ($[M+6H]^{+6}$), 1129.61 ($[M+7H]^{+7}$) and 988.52 ($[M+8H]^{+8}$).

RP-HPLC and ESI-MS analysis of the resin samples collected during the assembly of the N-terminal region of the peptide, particularly the last 15 residues (from IGF-II (15-67)) showed a significant increase in the number of impurities and by-products. This increase was postulated to be caused by the peptide adopting secondary structure on-resin. The formation of secondary structure leads to aggregation of the growing peptides chains, which in turn causes incomplete Fmoc deprotection and couplings steps, resulting in truncated by-products.

4.2.5 Improving the assembly of the N-terminal region of the IGF-II peptide

As discussed above, the difficult assembly of the N-terminal region of the IGF-II peptide was determined to result from the formation on-resin secondary structure and aggregation. Several strategies have been developed to prevent or disrupt the formation of such aggregation. Improving the solvation of the growing peptide chain or the introduction of *N*α-alkylated amino acids or pseudoproline units are the most common strategies used to overcome aggregation.[195,196,208,209] Solvents such as NMP, DMSO, hexafluroisopropanol (HFIP) or the addition of chaotropic salts, have been shown to increase the solvation of the growing peptide chain and prevent aggregation. In contrast, *N*α-alkylated amino acids such as *N*-α-

(2,4-dimethoxybenzyl)-glycine ((Dmb)Gly) or pseudoprolines, which are derived from Cys, Ser and Thr, disrupt on-resin secondary structure by introducing structural changes into the peptide backbone. Conveniently, these bulky substituents and protecting groups are removed during cleavage of the peptide from the resin using TFA, to give the native amino acid as depicted in Figure 41. All three methods were employed in an attempt to overcome the formation of on-resin secondary structure and improve the assembly of the N-terminal region of the native IGF-II peptide (**4.7**).

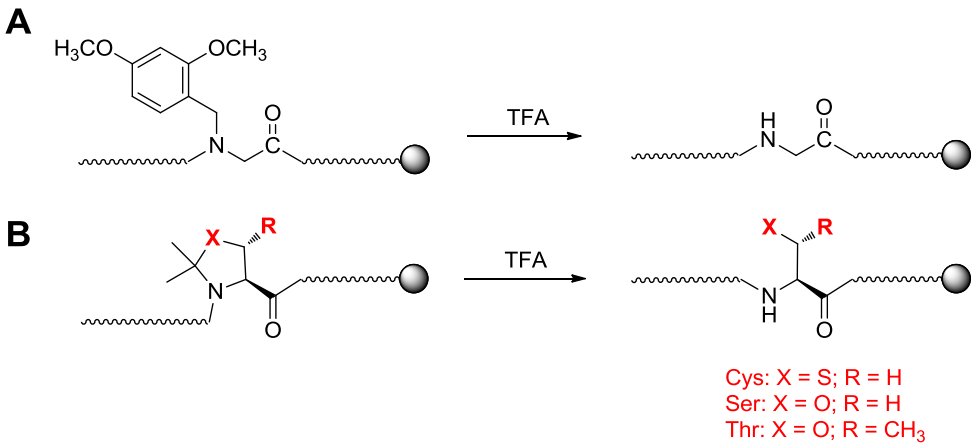
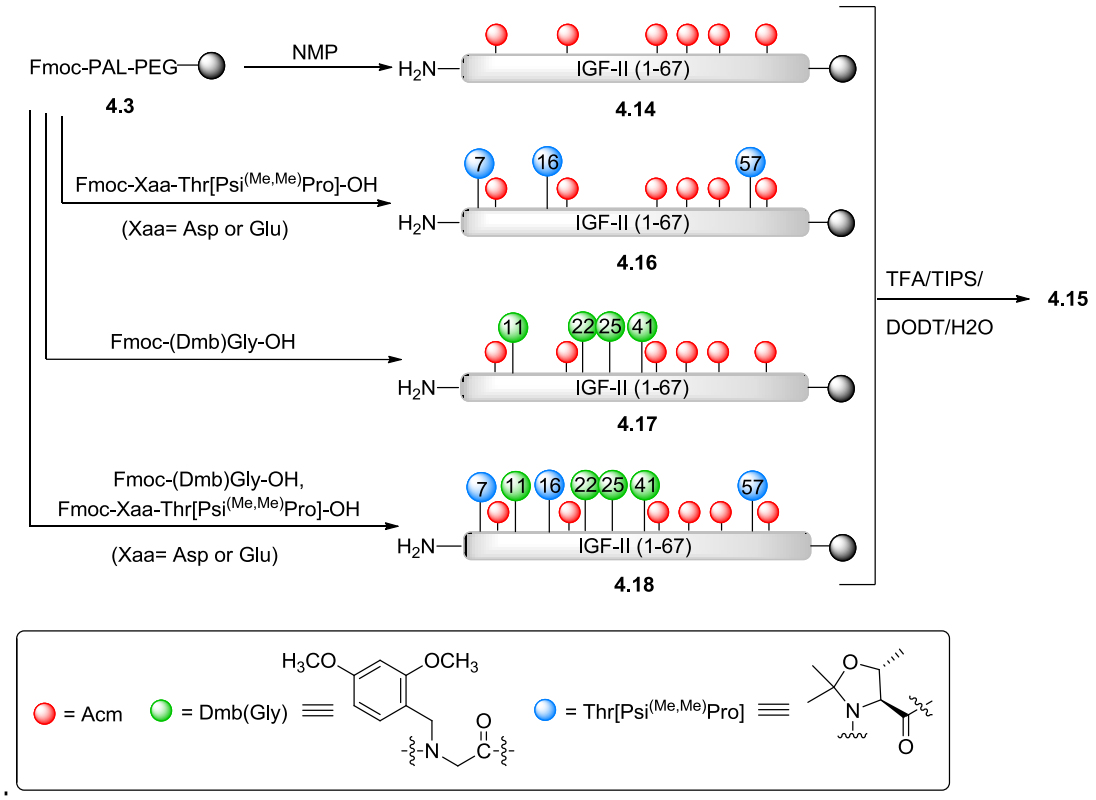


Figure 41: Strategies used to disrupt on-resin aggregation. Units shown disrupt aggregation by inducing a structural change in the peptide backbone. a) *Nα*-alkylated amino acids and b) pseudoprolines, Xaa[$\Psi^{(Me,Me)}$Pro]-OH; where Xaa refers to Cys, Ser or Thr.

The IGF-II resin-bound peptides (**4.14, 4.16-4.18**) were assembled from commercially available Fmoc-Pal-PEG-PS resin (**4.3**) as depicted in Scheme 12. Each of the resin-bound IGF-II peptides (**4.14, 4.16-4.18**) were synthesised with all six cysteines Acm protected. It was envisaged that protection of the Cys residues would prevent the formation of disulfide adducts, and other undesired by-products which could result from having free sulfhydryl's.



Scheme 12: Synthetic attempts to improve the N-terminal assembly of the IGF-II peptide (4.15), using manual Fmoc-SPPS. Where PEG = Polyethylene glycol; TIPS = triisopropylsilane and DODT = 3,6-dioxa-1,8-octanedithiol.

The resin-bound IGF-II peptide (1-67; 6 Acm) (**4.14**) was assembled from resin **4.3** using manual peptide synthesis (refer to Scheme 12). During the synthesis *N*-methylpyrrolidone (NMP) was used in place of DMF to increase the solvation of the growing resin-bound peptide. NMP is a better solvating agent than DMF and as such has been shown to improve coupling and deprotection reactions of difficult and aggregation prone sequences.[209] Resin **4.14** was cleaved using a TFA/TIPS/DODT/water mixture, and the resultant peptide mixture (**4.15**) was analysed by RP-HPLC and ESI-MS. The RP-HPLC analysis is shown in Figure 42, and displays a complex mixture from which the desired IGF-II (1-67; 6 Acm) peptide (**4.15**) was not detected. This was demonstrated by lack of any distinct peaks in the RP-HPLC trace shown in Figure 42 and the absence of the predicted *m/z* ions (1581 ($[M+5H]^{+5}$), 1317 ($[M+6H]^{+6}$) and 1129 ($[M+7H]^{+7}$)) in the ESI-MS (data not shown). These results suggest the use of NMP in the assembly of resin-bound peptide **4.14**, did not improve the N-terminal assembly of the IGF-II (1-67; 6 Acm) peptide (**4.15**).

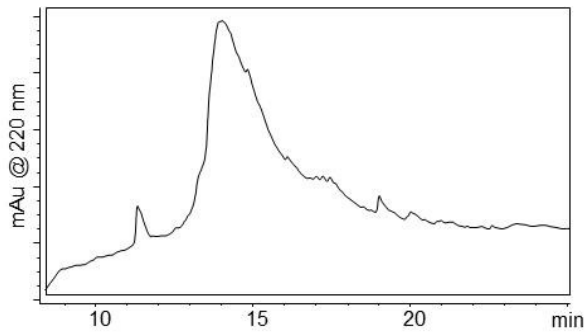


Figure 42: RP-HPLC analysis of the crude peptide mixture from the cleavage of resin 4.14.

As discussed above, the incorporation of pseudoproline units (Fmoc-Xaa-Yaa[$\Psi^{(Me,Me)}$Pro]-OH) can significantly improve the efficiency and yield of synthesis for 'difficult' and aggregation prone sequences.[208-212] To investigate the effect of pseudoprolines on the assembly of the N-terminal region of IGF-II, Thr^{7}, Thr^{16} and Thr^{58} were replaced with threonine derived pseudoproline dipeptides, Fmoc-Asp(O*t*Bu)-Thr[$\Psi^{(Me,Me)}$Pro]-OH and Fmoc-Glu(O*t*Bu)-Thr[$\Psi^{(Me,Me)}$Pro]-OH.

The resin-bound IGF-II (1-67; 6Acm; ΨT7, ΨT16, ΨT58) peptide (**4.16**) was elongated from resin **4.3** by manual peptide synthesis as depicted in Scheme 12. Resin **4.16** was cleaved on treatment with TFA/TIPS/DODT/water and the resultant peptide mixture **4.15**, was analysed by RP-HPLC and ESI-MS (refer to Figure 43). The ESI-MS analysis is shown in Figure 43B, and displays *m/z* ions ($[M+7H]^{+7}$; 1129.73 *m/z*) which are consistent with the calculated *m/z* ions ($[M+7H]^{+7}$; 1129.70 *m/z*) for the IGF-II (1-67; 6 Acm) peptide (**4.15**). However, the RP-HPLC analysis shown in Figure 43A, displays a complex mixture of products as indicated by a series of overlapping and unresolved peaks. These results suggest that the incorporation of pseudoproline dipeptides did not significantly improve the assembly of the N-terminal region of the IGF-II (1-67; 6 Acm) peptide (**4.15**).

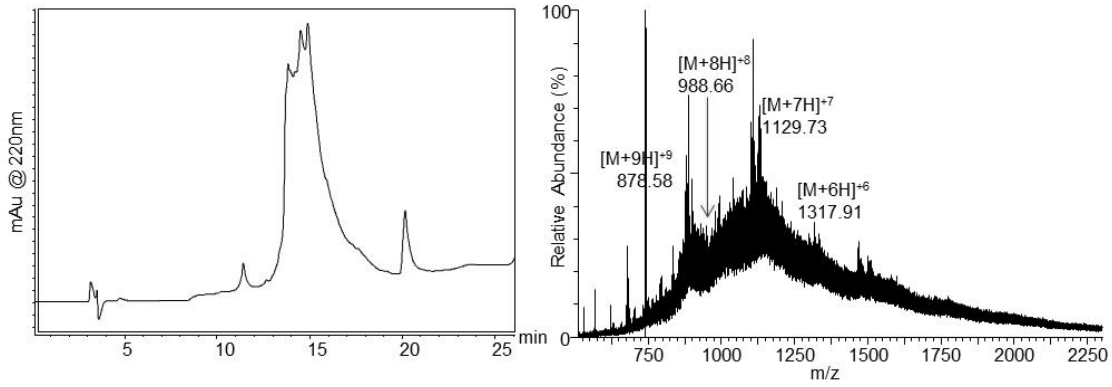


Figure 43: Analysis of the crude peptide mixture from the cleavage of IGF-II (1-67; 6Acm; ΨT7, ΨT16, ΨT58) resin (4.16). a) RP-HPLC trace of crude peptide **4.15** and b) ESI-MS of crude peptide **4.15**. Calcd. for $C_{339}H_{536}N_{100}O_{106}S_6$: 7900.9035 (average isotopes); observed: *m/z* 1317.91 ($[M+6H]^{+6}$), 1129.73 ($[M+7H]^{+7}$), 988.66 ($[M+8H]^{+8}$) and 878.58 ($[M+9H]^{+9}$).

The incorporation of the *N*-α-Fmoc-*N*-α-(2,4-dimethoxybenzyl)-glycine (Fmoc-(Dmb)Gly-OH) derivatives into the IGF-II (1-67; 6 Acm) peptide (**4.15**) was next investigated as a method for disrupting on-resin aggregation. The resin-bound IGF-II (1-67; 6Acm; DmbG11, DmbG22, DmbG25, DmbG41) peptide (**4.17**) was assembled from resin **4.3**, using a combination of automated and manual synthesis, as outlined in Scheme 12. Dmb glycine derivatives were incorporated into the IGF-II peptide (**4.17**) in place of Gly^{11}, Gly^{22}, Gly^{25} and Gly^{41}. Resin **4.17** was cleaved using a TFA cleavage cocktail and the resultant peptide mixture was analysed by RP-HPLC and ESI-MS and the results are displayed in Figure 44.

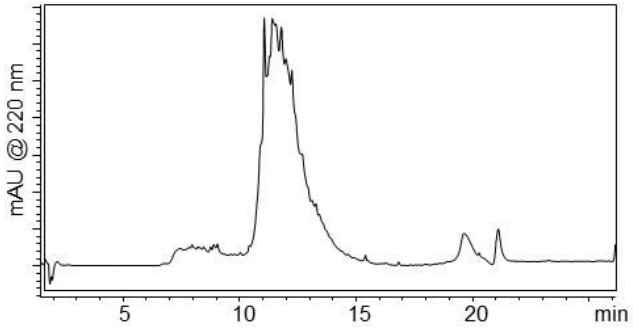


Figure 44: RP-HPLC analysis of the crude peptide from the cleavage of IGF-II (1-67; 6Acm; DmbG11, DmbG22, DmbG25, DmbG41) resin 4.17.

RP-HPLC analysis of crude peptide **4.15** is displayed in Figure 44 and shows a complex mixture of products as indicated by the presence of several overlapping and unresolved peaks. In addition, the desired *m/z* ions (878 ($[M+9H]^{+9}$), 988 ($[M+8H]^{+8}$), 1129 ($[M+7H]^{+7}$), 1317 ($[M+6H]^{+6}$)) for peptide **4.15,** were not detected in the ESI-MS (data not shown). Subsequent LCMS analysis[xiii] of the crude mixture is shown in Figure 45 and gave contrasting results. Firstly a distinct and resolved peak at approximately 15 min is present in the LCMS displayed in Figure 45A, which was not apparent in the initial RP-HPLC trace as shown in Figure 44. Analysis of the MS data collected between 14.9-15.1 min is shown in Figure 45B and displays *m/z* ions ($[M+7H]^{+7}$; 1129.4 *m/z*) which are in good agreement with the expected *m/z* ions ($[M+7H]^{+7}$; 1129.70 *m/z*) for the IGF-II (1-67; 6 Acm) peptide (**4.15**). These results suggest the IGF-II (1-67; 6 Acm) peptide (**4.15**) was the major component of the crude mixture shown in Figure 45A. Despite this encouraging result, no follow up experiments were conducted as the LCMS data was obtained much later than the initial RP-HPLC and ESI-MS analysis (refer to Figure 44). Consequently the work discussed for the remainder of this chapter was conducted after acquiring the results from the initial RP-HPLC and ESI-MS analysis (refer to Figure 44), and before obtaining the encouraging LCMS data (refer to Figure 45). Nonetheless, the LCMS results cannot be disregarded, and they suggest that the inclusion of Dmb glycine derivatives during the assembly resin-bound peptide **4.17** may have been successful at improving the assembly of the N-terminal region of the IGF-II (1-67; 6 Acm) peptide (**4.15**).

[xiii] LCMS analysis was unavailable until after the completion of the work in Chapter 4.

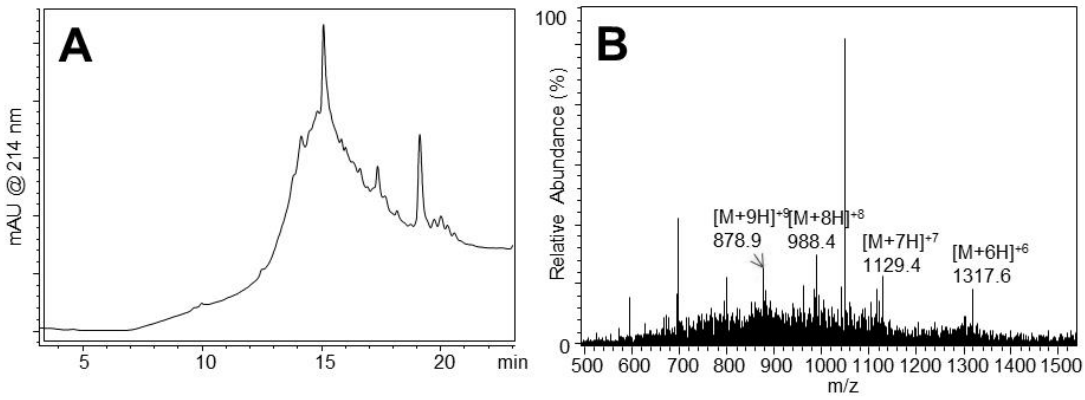


Figure 45: LCMS Analysis of the crude peptide mixture from the cleavage of IGF-II (1-67; 6Acm; DmbG11, DmbG22, DmbG25, DmbG41) resin 4.17. a) RP-HPLC trace of crude peptide **4.15** and b) ESI-MS of collected between 14.9-15.1 min for crude peptide **4.15**. Calcd. for $C_{339}H_{536}N_{100}O_{106}S_6$: 7900.9035 (average isotopes); observed: *m/z* 1317.6 ($[M+6H]^{+6}$), 1129.4 ($[M+7H]^{+7}$), 988.4 ($[M+8H]^{+8}$) and 878.9 ($[M+9H]^{+9}$).

In a final attempt to improve assembly of the N-terminal region of the IGF-II (1-67; 6 Acm) peptide (**4.15**), pseudoproline dipeptides and Dmb glycine derivatives were used in combination (refer to Scheme 12). The IGF-II (1-67; 6Acm; ΨT7, ΨT16, ΨT58, DmbG11, DmbG22, DmbG25, DmbG41) resin-bound peptide (**4.18**) was synthesised from resin **4.3** using manual peptide synthesis, where Thr^{7}, Thr^{16} and Thr^{58} were replaced with threonine derived pseudoproline dipeptides (Fmoc-Asp(O*t*Bu)-Thr[$\Psi^{(Me,Me)}$Pro]-OH and Fmoc-Glu(O*t*Bu)-Thr[$\Psi^{(Me,Me)}$Pro]-OH) and Gly^{11}, Gly^{22}, Gly^{25} and Gly^{41} were substituted with Dmb glycine derivatives. Resin **4.18** was cleaved on treatment with TFA/TIPS/DODT/water and the resultant peptide mixture was analysed by RP-HPLC and ESI-MS. RP-HPLC analysis is displayed in Figure 46 and reveals three distinctive overlapping peaks. Despite the promising RP-HPLC analysis, ESI-MS analysis (data not shown) revealed a complex mixture of peptide products, from which the desired *m/z* ions (878 ($[M+9H]^{+9}$), 988 ($[M+8H]^{+8}$), 1129 ($[M+7H]^{+7}$), 1317 ($[M+6H]^{+6}$)) were not observed.

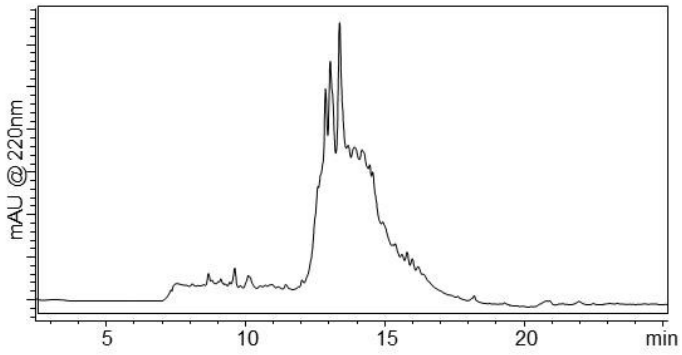


Figure 46: RP-HPLC analysis of the crude mixture peptide from the cleavage of the IGF-II (1-67; 6Acm; ΨT7, ΨT16, ΨT58, DmbG11, DmbG22, DmbG25, DmbG41) resin (4.18).

LCMS analysis[xiv] of the mixture is displayed in Figure 47. Specifically, the LCMS trace is shown in Figure 47A and shows three distinct peaks, which is comparable to the initial RP-HPLC analysis (refer to Figure 46). MS data collected between 14.9-15.1 min is shown in Figure 47B and shows *m/z* ions ($[M+6H]^{+6}$; 1317.6 *m/z*) which are consistent with the calculated *m/z* ions ($[M+6H]^{+6}$; 1317.82 *m/z*) for the desired peptide **4.15**. These results indicate the IGF-II (1-67; 6 Acm) peptide (**4.15**) was present as a major component of the crude mixture, and thus suggests the inclusion of pseudoproline dipeptides and Dmb glycine derivatives may have been effective at improving the N-terminal assembly of the IGF-II (1-67; 6 Acm) peptide (**4.15**). However, as the LCMS data was acquired after the initial RP-HPLC and ESI-MS analysis (refer to Figure 46) no follow up experiments were conducted.

[xiv] LCMS analysis was unavailable until after the completion of the work in Chapter 4.

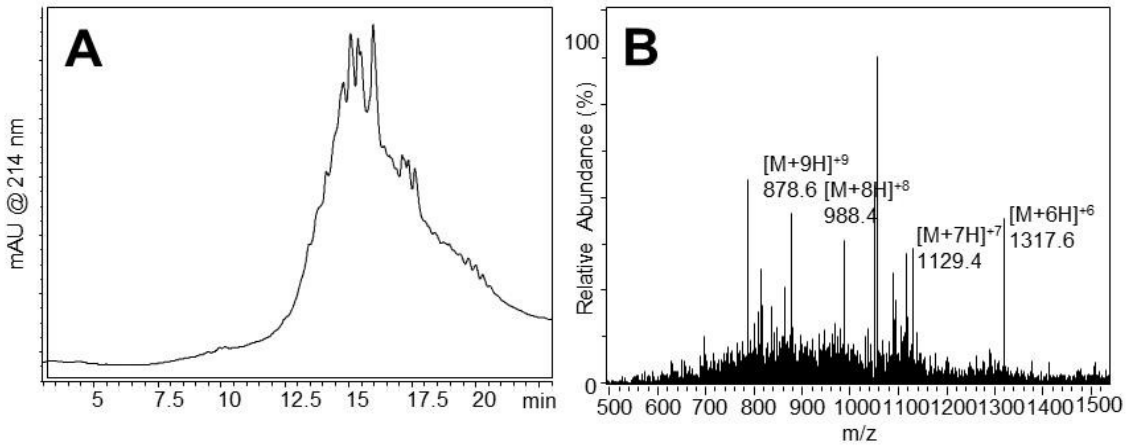
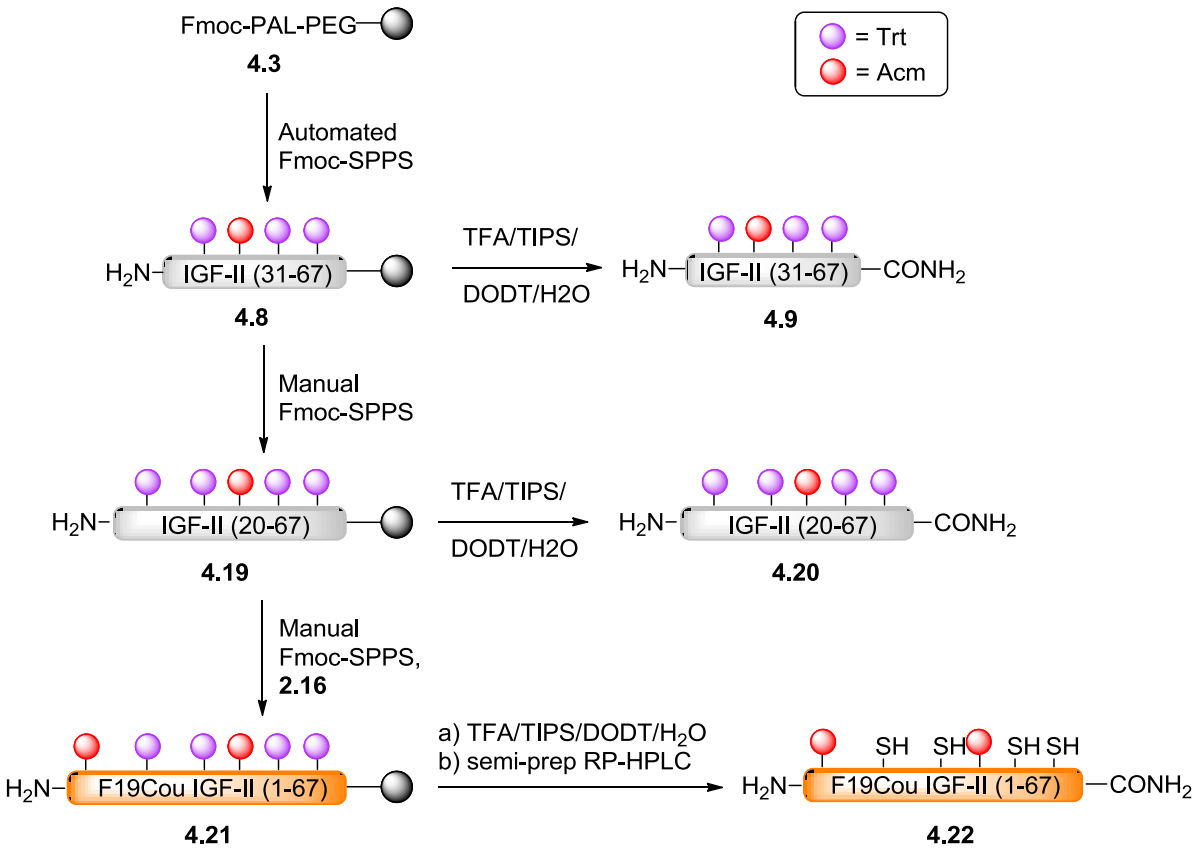


Figure 47: LCMS analysis of the crude peptide mixture from the cleavage of the IGF-II (1-67; 6Acm; ΨT7, ΨT16, ΨT58, DmbG11, DmbG22, DmbG25, DmbG41) resin 4.18. a) RP-HPLC trace of crude peptide **4.15** and b) ESI-MS of collected between 14.9-15.1 min for crude peptide **4.15**. Calcd. for $C_{339}H_{536}N_{100}O_{106}S_6$: 7900.9035 (average isotopes); observed: *m/z* 1317.6 ($[M+6H]^{+6}$), 1129.4 ($[M+7H]^{+7}$), 988.4 ($[M+8H]^{+8}$) and 878.6 ($[M+9H]^{+9}$).

In summary, small scale resin cleavages conducted during the manual assembly of the native IGF-II peptide (**4.7**), identified the N-terminal region of the peptide as the main source of difficulty in the synthesis. This difficult assembly of the N-terminal region is thought to have been caused by the formation of on-resin secondary structure and aggregation. Solvation, incorporation of pseudoproline dipeptides, and Dmb glycine derivatives were investigated as strategies to improve the assembly of the N-terminal region, and thus accomplish the synthesis of the IGF-II (1-67; 6 Acm) peptide (**4.15**). LCMS data from the strategies employing Dmb glycine derivatives (**4.17** and **4.18**) suggest the on-resin aggregation was disrupted and assembly of the N-terminal region of the IGF-II (1-67; 6 Acm) peptide was significantly improved. Further studies into the incorporation of Dmb glycine derivatives into the IGF-II peptide are required to confirm the suitability of these units as a robust strategy for improving the synthesis of the native IGF-II peptide. Interestingly, the difficulties in N-terminal assembly were only encountered in the synthesis of the native IGF-II peptide. In the parallel synthesis of the fluorescent IGF-II analogue, F19Cou IGF-II (**4.2**), which differs from the native IGF-II peptide by the substitution of coumaryl amino acid **2.2** at position 19 (refer to Figure 32), on-resin aggregation was not observed. The successful synthesis of the novel F19Cou IGF-II analogue (**4.2**) is now described.

4.2.6 Synthesis of the F19Cou IGF-II protein (4.2)

The synthesis of the fluorescent IGF-II analogue, F19Cou IGF-II (**4.2**), with the substitution of coumaryl amino acid **2.2** at residue 19, is outlined in Scheme 13. This approach employed the selective Acm protection of the Cys^{9}/Cys^{47} disulfide pair, as described in Section 4.2.2. The F19Cou IGF-II (1-67; 2 Acm) peptide (**4.22**) was assembled from resin **4.3** using a combination of automated and manual peptide synthesis (refer to Scheme 13). First, IGF-II (31-67; Acm, 3 Trt) resin (**4.8**) was assembled from resin **4.3** by automated peptide synthesis. Resin **4.8** was then elongated by manual peptide synthesis to residue 20, to afford the IGF-II (20-67; Acm, 3 Trt) resin-bound peptide (**4.19**). The *N*α-Fmoc protected coumaryl amino acid **2.5** (synthesis described in Section 2.2.2, Chapter 2) was incorporated at residue 19 in place of Phe. Elongation using manual peptide synthesis gave the full length F19Cou IGF-II (1-67; 2 Acm, 4 Trt) resin-bound peptide (**4.21**) (refer to Scheme 13).



Scheme 13: Synthesis of the novel fluorescent F19Cou IGF-II (1-67; 2Acm) peptide (4.22).

The F19Cou IGF-II (1-67; 2 Acm) peptide (**4.22**) was released from resin **4.21** on treatment with TFA/TIPS/DODT/water and the resultant mixture analysed by RP-HPLC and ESI-MS. RP-HPLC analysis of the crude peptide **4.22** is shown in Figure 48A, and shows a series of overlapping peaks with a major peak at 14.6 min. ESI-MS analysis (data not shown) of the peak at 14.6 min revealed *m/z* ions ($[M+6H]^{+6}$; 1286.00 *m/z*) that are consistent with the calculated *m/z* ions ($[M+6H]^{+6}$; 1286.80 *m/z*) for the F19Cou IGF-II (1-67; 2 Acm) peptide (**4.22**). These results suggest the desired **4.22** peptide was a major component of the mixture. Peptide **4.22** was subsequently isolated by semi-preparative RP-HPLC to afford the purified F19Cou IGF-II (1-67; 2 Acm) peptide in a reasonable yield (14%).

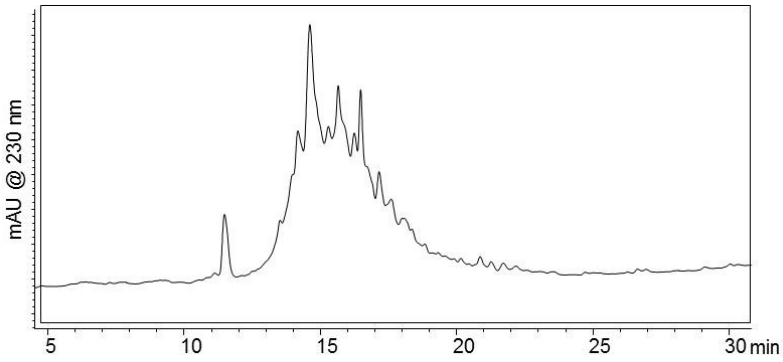
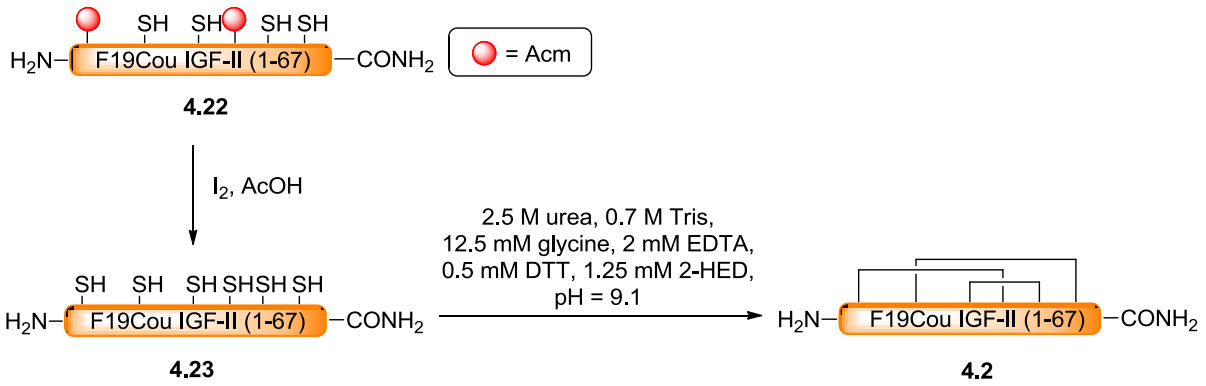


Figure 48: RP-HPLC analysis of the crude fluorescent F19Cou IGF-II (1-67; 2 Acm) peptide (4.22).

The Acm groups of **4.22** were deprotected using iodine oxidation[213] as shown in Scheme 14. Specifically, peptide **4.22** was dissolved in 80% aqueous acetic acid and the solution was treated with 20 mM iodine in 80% aqueous acetic acid. The reaction was quenched by the slow addition of Dowex 1X8 200, and the crude F19Cou IGF-II peptide (**4.23**) was isolated by semi-preparative RP-HPLC. This gave the desired F19Cou IGF-II (1-67) peptide (**4.23**) in excellent yield (90%).



Scheme 14: Synthesis of the novel fluorescent F19Cou IGF-II protein (4.2).

The F19Cou IGF-II (1-67) protein (**4.2**) was isolated following oxidative folding of the F19Cou IGF-II (1-67) peptide (**4.23**) as depicted in Scheme 14. First, the F19Cou IGF-II (1-67) peptide (**4.23**) was reduced with 20 mM dithiothreitol (DTT). Following this, peptide **4.23** was refolded according to the procedure described by Delaine *et al.*[23] In this, the reduced solution of **4.23** was diluted in a refolding buffer to give a solution with a pH of 9.1. Specifically, the refold was performed at a final concentration of 2.5 M urea, 0.7 M Tris, 12.5 mM glycine, 2 mM EDTA, 0.5 mM DTT and with a maximum protein concentration of 0.1 mg/ml. To this, 2-hydroxyethyl disulfide (2-HED) was added at a final concentration of 1.25 mM, and the refold was stirred slowly at rt for 180 min. The resultant F19Cou IGF-II protein (**4.2**) was isolated by semi-preparative RP-HPLC in an acceptable yield (5%).

The F19Cou IGF-II protein (**4.2**) was synthesised in high purity as demonstrated by the presence of a single peak in the RP-HPLC trace illustrated in Figure 49A, and confirmed by the strong agreement of the *m/z* ions ($[M+7H]^{+7}$, 1081.94 *m/z*) observed in the HRMS shown in Figure 49B with the expected *m/z* ions ($[M+7H]^{+7}$; 1081.92 *m/z*). Gel electrophoresis was carried out to demonstrate the fluorescent properties of the synthetic F19Cou IGF-II protein (**4.2**) in comparison to the recombinant native IGF-II protein. The resultant SDS-PAGE gel shown in Figure 49C, was irradiated at 254 nm (left) and shows a fluorescent band at 7.5 kDa, which corresponds to the synthetic F19Cou IGF-II protein (**4.2**).

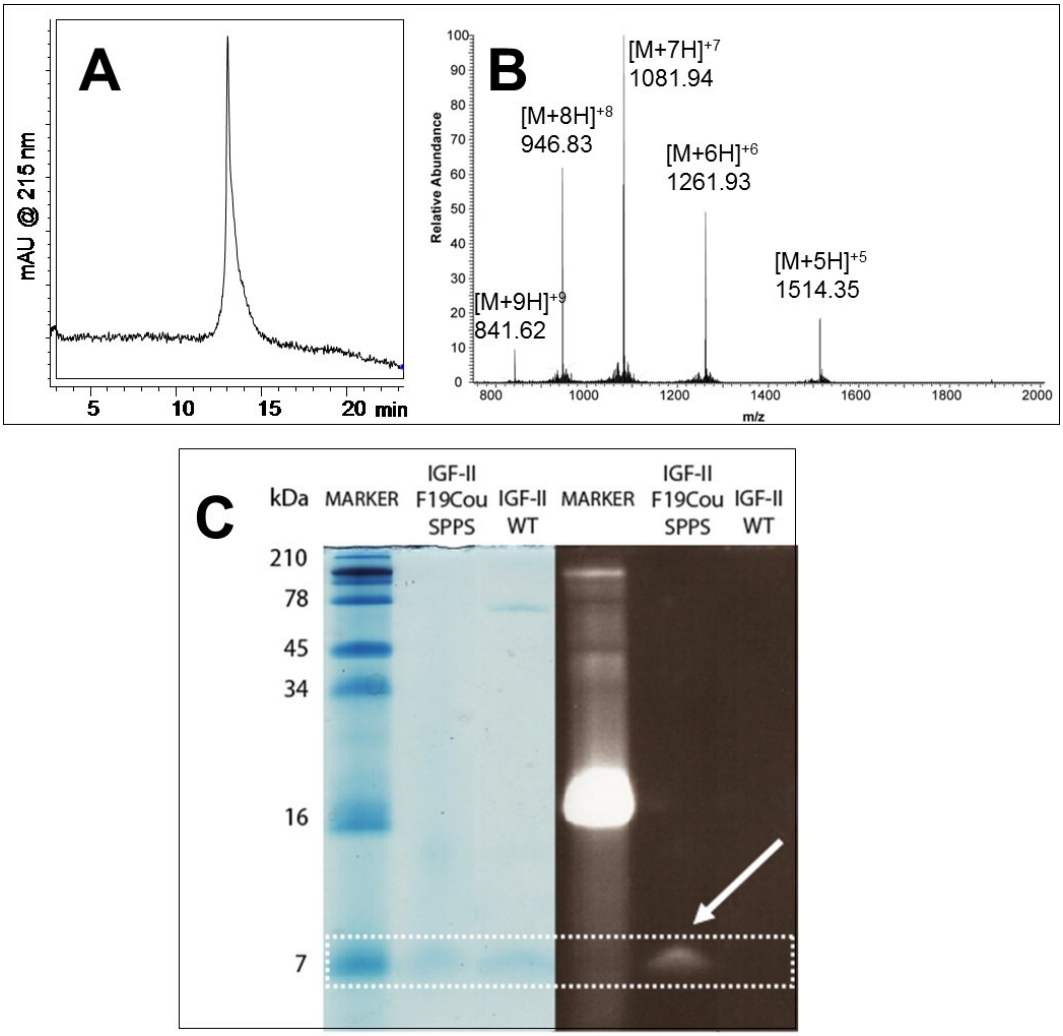


Figure 49: Characterisation of the F19Cou IGF-II protein (4.2). a) RP-HPLC trace of **4.2**; b) HRMS of **4.2**. Calcd. for $C_{325}H_{502}N_{94}O_{103}S_6$: 7566.4445 (average isotopes); observed: *m/z* 1514.35 ($[M+5H]^{+5}$), 1261.93 ($[M+6H]^{+6}$), 1081.94 ($[M+7H]^{+7}$), 946.83 ($[M+8H]^{+8}$) and 841.62 ($[M+9H]^{+9}$); c) SDS-PAGE gel stained with coomassie blue (*left*); unstained gel irradiated at 254nm (*right*); lanes (*from left to right*) correspond to marker, synthetic F19Cou IGF-II protein (**4.2**) and recombinant native IGF-II protein. The bands at 7.5 kDa correspond to the full length IGF-II proteins and are highlighted by a *dashed box* and the fluorescence of the synthetic F19Cou IGF-II protein (**4.2**) is highlighted with an *arrow*.

4.2.7 Biological activity of the synthetic F19Cou IGF-II protein (4.2)

Competition binding assays were conducted to determine the affinity of the synthetic F19Cou IGF-II protein (**4.2**) for IGF-1R and IR-A. Receptor binding experiments were carried out on immunocaptured IGF-1R and IR-A as described by Denley *et al.*[22] Briefly, increasing concentrations of **4.2** were added to a constant concentration of europium labelled IGF-II (EuIGF-II). The competition binding curves are illustrated in Figure 50 and the IC_{50} values are summarised in Table 13. The IC_{50} values for the binding of the F19Cou IGF-II protein (**4.2**) to the IGF-1R and IR-A were 45.5 ± 0.3 nM and 72.5 ± 0.3 nM, respectively. Importantly, compared to the recombinant native IGF-II protein, protein **4.2** had only an 18-fold decrease in affinity for the IGF-1R (2.5 nM versus 45.5 nM), and a 9.8-fold decrease in affinity for the IR-A (7.4 nM versus 72.5 nM). These binding affinities were also consistent with the binding affinities determined for the recombinant F19Cou IGF-II protein (**3.1**) binding to the IGF-1R (45.5 nM versus 47.6 nM) and IR-A (72.5 nM versus 96.9 nM) (refer to Section 3.2.8). Finally, the minor discrepancies between the binding affinities of **3.1** and **4.2** were attributed to the difference in the C-terminus of the proteins and the increased purity of **4.2** compared to **3.1**.

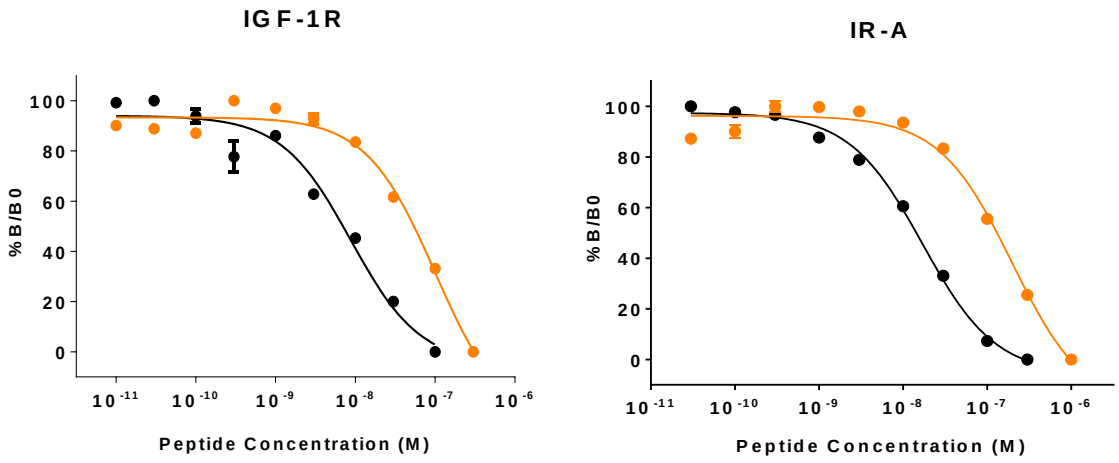


Figure 50: Competitive binding of F19Cou IGF-II (4.2) to immunocaptured IGF-1R and IR-A. Immunocaptured IGF-1R or IR-A were incubated with Europium labelled IGF-II (EuIGF-II) in the presence of, or absence of, increasing concentrations of recombinant native IGF-II (*black*) or **4.2** (*orange*).

Table 13: IC_{50} values derived from competitive binding assays of recombinant native IGF-II and synthetic F19Cou IGF-II (4.2) to immunocaptured IGF-1R and IR-A. Where the affinity relative to the recombinant native IGF-II is the IC_{50} relative to that of IGF-II binding to the IGF-1R or IR-A (IC_{50} native IGF-II/IC_{50} IGF-II analogue) and is expressed as percentage of IGF-II binding. IGF-II ± S.E is derived from at least 3 separate experiments performed in triplicate.

	IC_{50} values from immunocaptured IGF-1R and IR-A (nM)		
	recombinant native IGF-II	F19Cou IGF-II (4.2)	Affinity relative to recombinant native IGF-II (%)
IGF-1R	2.5 ± 0.3	45.5 ± 0.3	5.6
IR-A	7.4 ± 0.2	72.5 ± 0.3	10.2

It was anticipated the F19Cou IGF-II protein (**4.2**) would exhibit a lower binding affinity than the recombinant native IGF-II protein, as Phe^{19} has been shown to be involved in receptor binding (refer to Section 1.1.4, Chapter 1).[23,44] It is thought that the increase in side chain bulk, caused by the inclusion of the coumarin moiety, perturbs the site 2 interaction of IGF-II with both the IGF-1R and IR-A. However, it has also been demonstrated that changing Phe^{19} to a smaller and less hydrophobic residue, such as Leu or Ala in other IGF-II analogues also perturbs receptor binding.[23,44] Thus, these results are consistent with what has been reported in the literature and reinforce the importance of Phe^{19} for IGF-II binding to the IGF-1R and IR-A.

4.3 Conclusion

The native IGF-II peptide **4.7** and Acm protected peptides **4.11** and **4.15** were synthesised using a linear Fmoc-SPPS protocol. However isolation of these IGF-II peptides (**4.7**, **4.11** and **4.15**) was unsuccessful due to a large number of truncated and higher mass peptide by-products. A systematic investigation of the peptide synthesis found these by-products arose during the assembly of the N-terminal region of the IGF-II peptide. In particular, the assembly of the last 15 residues (1-15) was found to be very difficult.

Difficulties in the assembly of the N-terminal region of the native IGF-II peptides were attributed to the formation of on-resin secondary structure and aggregation. Increasing resin solvation and the incorporation of pseudoprolines and Dmb glycine derivatives were

investigated as strategies to overcome this on-resin aggregation. Data suggests that assembly of the IGF-II (1-67; 6 Acm) peptide (**4.15**) was significantly improved by the incorporation of Dmb glycine. However, further studies are required to determine the effectiveness of these Dmb glycine derivatives in improving the synthesis of the native IGF-II peptide.

A fluorescent IGF-II analogue, F19Cou IGF-II (**4.2**), was successfully synthesised using a linear Fmoc-SPPS protocol. The substitution of Phe^{19} with coumaryl glycine was also found to improve the N-terminal assembly of the F19Cou IGF-II (1-67; 2 Acm) peptide (**4.22**), which in turn allowed the full length F19Cou IGF-II protein (**4.2**) to be isolated by semi-preparative RP-HPLC. Thus, it is postulated that substitution of coumaryl amino acid **2.2** at residue 19 disrupts the formation of structures that were normally inhibitory to the assembly process in native IGF-II peptides (**4.7**, **4.11** and **4.15**).

The F19Cou IGF-II protein (**4.2**) bound its two high-affinity receptors the IGF-1R and IR-A with IC_{50} values of 45.5 ± 0.3 nM and 72.5 ± 0.3 nM respectively. The drop in affinity for the IGF-1R and IR-A compared to the recombinant native IGF-II protein was consistent with what has been observed in the literature for other F19X IGF-II analogues. These results further reinforce the importance of Phe^{19} for the binding of IGF-II to the IGF-1R and IR-A.

Initially, the aim of these studies was to develop a robust synthetic strategy that would allow the assembly of several IGF-II analogues, as part of an IGF-II receptor binding study. Due to the difficulties described, a convergent approach was proposed for the synthesis of the native IGF-II protein and other fluorescent IGF-II analogues. This is discussed further in Chapter 5.

Chapter 5

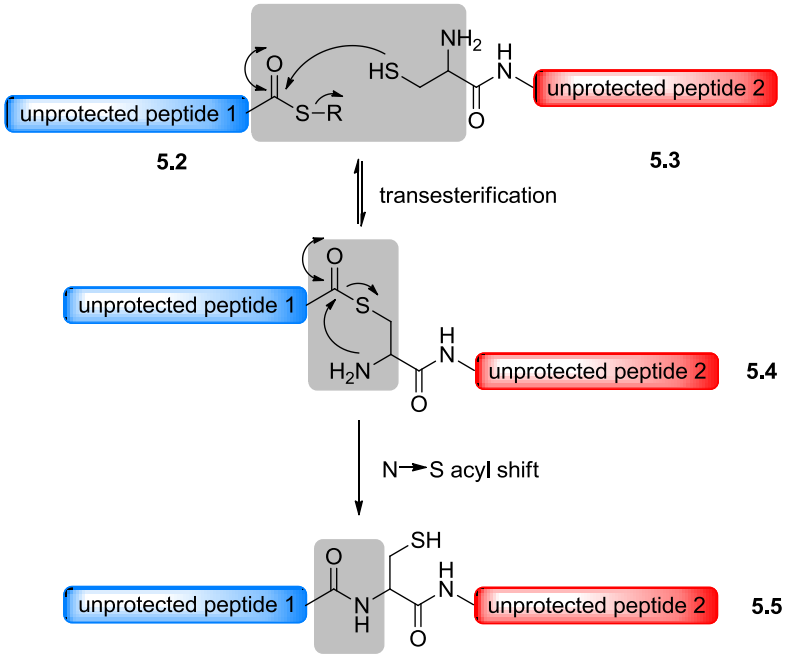
Synthesis of IGF-II analogues by native chemical ligation

5.1 Introduction

As discussed in Chapter 4, a linear Fmoc-SPPS strategy yielded the fluorescent F19Cou IGF-II protein (**4.2**). However difficulties with on-resin aggregation prevented isolation of the native IGF-II protein (**4.1**). It was envisaged synthesising smaller peptide fragments and assembling them in a convergent approach would overcome these problems with on-resin secondary structure and aggregation (refer to Section 4.2.4, Chapter 4). Native chemical ligation has been shown to be successful for synthesising the related IGF-I ligand.[214] Thus, it was proposed, a native chemical ligation methodology would be suitable for the synthesis of the native IGF-II (**4.1**), F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins.

5.1.1 Native chemical ligation

Native chemical ligation was first conceptualised by Dawson and co-workers,[86] and since has become the most widely used method for the total and semi-synthesis of proteins.[88,151,215,216] This methodology has been employed in the current study to synthesise the IGF-II analogues **4.1**, **4.2** and **5.1**. The native chemical ligation reaction is displayed in Scheme 15 and involves the coupling of unprotected peptide fragments in aqueous solution. The reaction is chemo- and regioselective and occurs between a peptide fragment bearing a C-terminal thioester (**5.2**) and a peptide fragment containing an N-terminal cysteine (**5.3**), and yields a product (**5.5**) with a native amide bond at the ligation site.



Scheme 15: Mechanism of the native chemical ligation reaction.[86] The ligation reaction involves the coupling of unprotected peptide segments in solution.

Native chemical ligation has been shown to overcome the solubility and size limitations of traditional solution phase chemistry, linear SPPS and molecular biology.[151,217] As such it was thought that native chemical ligation could be used to synthesise the IGF-II analogues **4.1**, **4.2** and **5.1** in high yield and purity. The ligation reaction involves two steps (refer to Scheme 15). The first is a reversible, chemoselective transthioesterification to form a thioester-linked intermediate (**5.4**). This is followed by an irreversible, rapid intramolecular N-to-S acyl shift to form a peptide (**5.5**) with a native amide bond at the ligation site.[86] The kinetics of the ligation reaction are affected by a number of factors, including pH, the nature of the C-terminal amino acid bearing the thioester, the nature of the thioester leaving group and the presence or otherwise of thiol additives and catalysts.[86,218]

Of each of the variables mentioned above, pH is the most critical. Ligation reactions are carried out in buffered aqueous solutions at neutral pH (pH of 7), as at higher pH the thioester moiety is hydrolysed. Similarly, strongly acidic conditions reduce the nucleophilic character of the cysteine thiol, which in turn decreases the rate of ligation reaction.[86]

When selecting ligation sites the influence of the C-terminal amino acid (Xaa) bearing the thioester was carefully considered. A comparison of different C-terminal thioesters is illustrated in Figure 51. Ligations where the C-terminal amino acid (Xaa) is Gly, His or Cys have been shown to proceed rapidly and to completion in less than 4 h. In contrast, ligations of C-terminal Pro and β-branched amino acids, such as Ile, Leu, Thr or Val proceed much more slowly and are often not complete after two days (refer to Figure 51).[219]

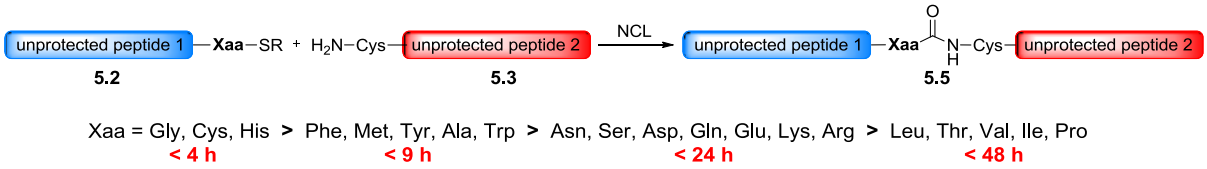


Figure 51: The influence of the amino acid forming the C-terminal thioester (5.2) on the time taken for completion of the ligation reaction. The C-terminal residue is denoted by Xaa and the times given in *red* correspond to the maximum time taken for the ligation reaction to proceed to completion. Figure adapted from Hackenberger *et al.*[93]

Generally, alkyl thioesters are less reactive than aryl thioesters. As such peptide thioesters are synthesised and purified with alkyl groups, which can be converted *in situ* during the ligation reaction to a more reactive aryl thioester.[93,218,220] This *in situ* conversion of an alkyl thioester (**5.6**) to the aryl thioester (**5.8**) is referred to as a thiol-thioester exchange and is illustrated in Figure 52.

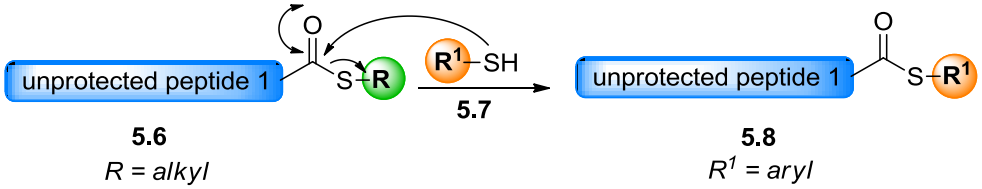


Figure 52: Thiol-thioester exchange reaction.[220] Conversion of an alkyl thioester (**5.6**) to a more reactive aryl thioester (**5.8**) occurs by treatment with an aryl thiol additive (**5.7**). Where R is an alkyl group and is shown in *green* and R^1 is an aryl group and is shown in *orange*.

Thiol-thioester exchange is accomplished by treating the ligation mixture with an excess of an exogenous aryl thiol (**5.6**) (refer to Figure 52).[220] After thiol-thioester exchange, the newly formed, reactive aryl thioester (**5.8**), rapidly undergoes transthioesterification to form a thioester linked intermediate (**5.4**; Scheme 15), followed by a rapid N-to-S acyl shift to give the desired ligation product (**5.5**; Scheme 15). When exogenous thiol additives are used, the thiol-thioester exchange is proposed to be the rate-limiting step of the ligation reaction.[218,220] Thus, an ideal exogenous thiol additive should undergo complete and rapid thiol-thioester exchange to form the more reactive thioester and have a leaving group which is easily displaced in the formation of the thioester-linked intermediate.[218] The exogenous thiol used in this work was mercaptophenylacetic acid (MPAA) and it is widely utilised in ligation chemistry.

This chapter describes the synthesis of the native IGF-II (**4.1**), F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins. The target proteins are shown in Figure 53 and were synthesised using native chemical ligation. The F19Cou IGF-II (**4.1**) and F28Cou IGF-II (**5.1**) proteins contain the coumaryl amino acid **2.2** at positions 19 and 28 respectively. As discussed in Chapter 1, access to these IGF-II analogues would facilitate a FRET-based investigation into the binding of IGF-II to its two high-affinity receptors.

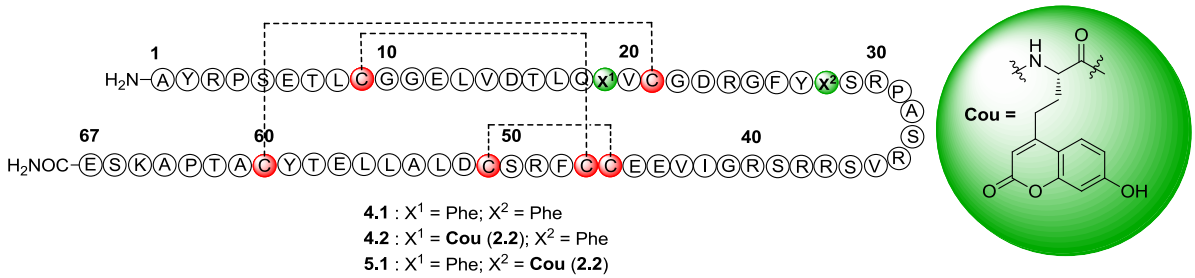
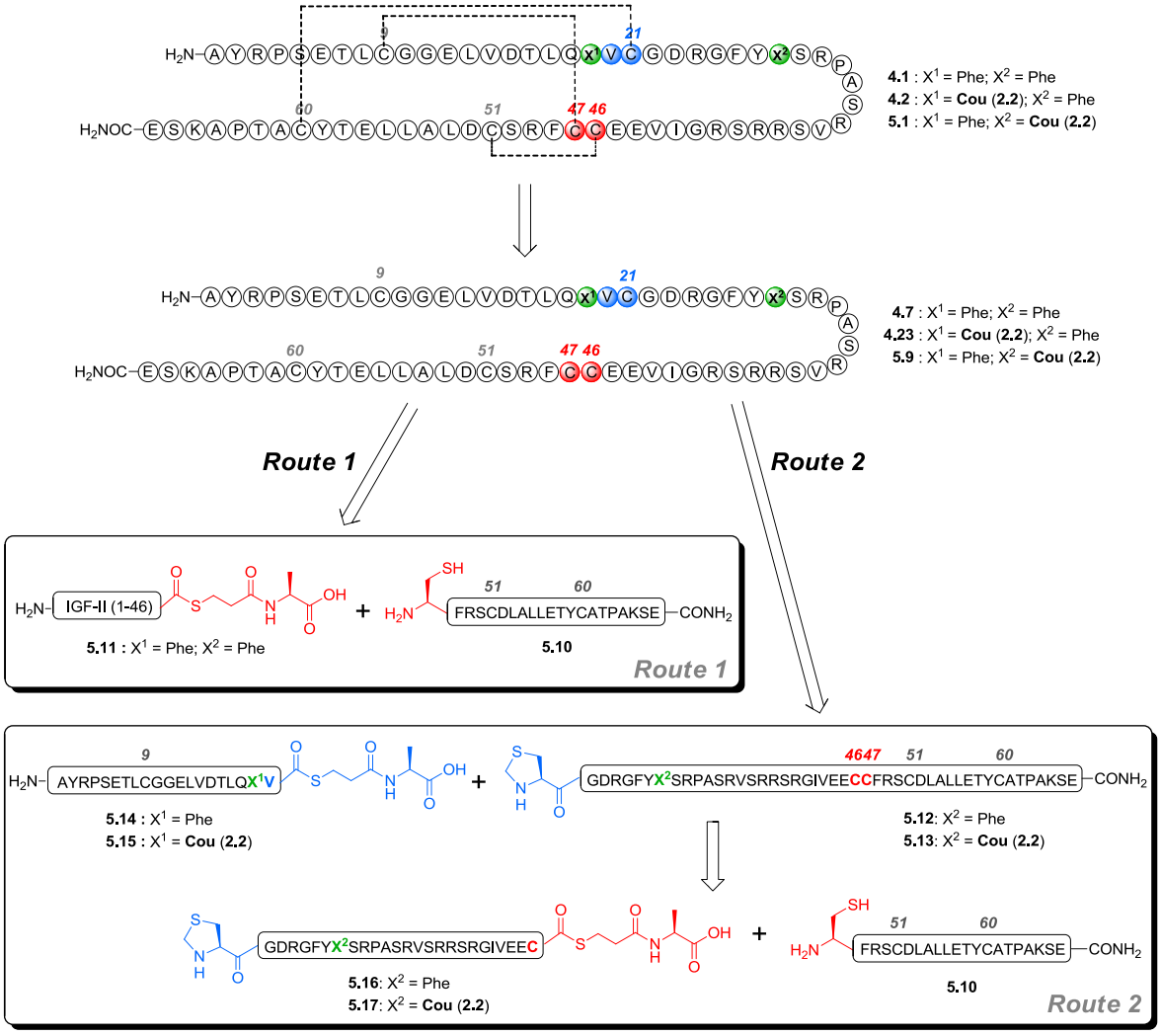


Figure 53: Target IGF-II proteins (4.1; 4.2 and 5.1). Sequence of IGF-II is shown with the residue numbers annotated above the sequence, Cys residues are highlighted in *red*; disulfide bonds shown are shown as *dashed lines* and the coumarin-based fluorophore (Cou; **2.2**) is illustrated within the *green* circle.

5.2 Results and discussion

5.2.1 Ligation strategy

As discussed in Section 5.1.1, native chemical ligation requires a Cys residue at the ligation site. The native IGF-II protein (**4.1**) contains six Cys residues (red; Figure 53) and as such is a good target for synthesis by native chemical ligation. The synthetic strategies adopted for the synthesis of the native IGF-II protein (**4.1**) and two non-native IGF-II analogues (**4.2** and **5.1**) are outlined in Scheme 16.



Scheme 16: Strategy for the synthesis of the native IGF-II protein (4.1) and fluorescent IGF-II analogues (4.2 and 5.1). Where ligation sites are shown in *red* and *blue*; location of fluorescent coumaryl amino acid **2.2** in the non-native analogues are shown in *green*; Cys residues are annotated in *bold italic numbering* above the sequence and disulfide bonds are shown as *dashed lines*.

Initially the two fragment approach, outlined in Route 1, Scheme 16, was proposed for the synthesis of the native IGF-II protein (**4.1**). In this approach the native IGF-II peptide (**4.7**) could be furnished from a single ligation between a 46 residue N-terminal fragment (**5.11**) and a 21 residue C-terminal fragment (**5.10**). This approach is discussed in Section 5.2.2, and was desirable as it afforded **4.7** from a single step. However synthesis of a 46 residue fragment is somewhat impractical due to the synthetic inefficiencies and difficulties with peptide purification which arise from long sequences. As such, an alternative, more modular approach (Route 2; Scheme 16) was subsequently proposed for the synthesis of the native IGF-II protein (**4.1**) and fluorescent IGF-II analogues **4.2** and **5.1**.

In the modular approach (Route 2; Scheme 16), the full length IGF-II peptides (**4.7**, **4.23** and **5.9**) were assembled through sequential cysteine and valine ligations from a common C-terminal fragment (**5.10**). The benefits of increased synthetic efficiency and improved product purity which arise from the synthesis of smaller peptide fragments were expected to outweigh any potential difficulties associated with increasing the complexity of the assembly process.

Ligation sites were selected based on the location of the cysteine residues in the IGF-II peptide, the predicted reactivity of the C-terminal residue bearing the thioester and the size of the resultant peptide fragment. Disconnection at Cys^{46}-Cys^{47} was selected for the C-terminal ligation site because studies of related systems has shown such ligations to be quantitatively complete in less than 4 h.[219] Two N-terminal ligation sites, Leu^{8}-Cys^{9} or Val^{20}-Cys^{21} were available. It was predicted that both of these sites would result in C-terminal thioesters with similar slow reactivity.[219] The Val^{20}-Cys^{21} junction was preferred for the N-terminal ligation site as it yielded peptide fragments (**5.14**, **5.15**, **5.16** and **5.17**) of similar size (approx. 20 residues), which was anticipated to result in increased synthetic efficiency and improved fragment purity.

Using this modular approach (Route 2; Scheme 16) it was envisaged the native IGF-II peptide (**4.7**) could be synthesised from the ligation of N-terminal thioester **5.14** and C-terminal fragment **5.12**. Fragment **5.12** could be assembled from ligation of thioester **5.16** and C-terminal fragment **5.10**. In a similar fashion, the F19Cou IGF-II peptide (**4.23**) could be

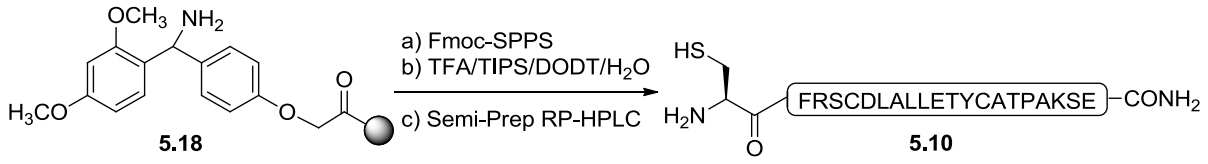
synthesised from ligation of the N-terminal thioester **5.15** and C-terminal fragment **5.12**. Fragment **5.12** could be furnished as described above. Finally, ligation of N-terminal thioester **5.14** with C-terminal fragment **5.13** could be used to synthesise the F28Cou IGF-II peptide (**5.9**). Fragment **5.13** could be from assembled from the ligation of thioester **5.17** and C-terminal fragment **5.10**. Application of these approaches in the synthesis of the native IGF-II (**4.1**), F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins is now described.

5.2.2 Two fragment ligation approach

A two fragment approach was initially investigated to determine the suitability of a convergent strategy for the chemical synthesis of the native IGF-II protein (**4.1**) and its fluorescent analogues (**4.2** and **5.1**). As discussed in Section 5.2.1 the Cys^{46}-Cys^{47} ligation site was selected. It was envisaged the desired native IGF-II peptide (**4.7**) could be furnished from a single ligation between thioester **5.11** and C-terminal fragment **5.10**. The synthesis of the required C-terminal fragment, IGF-II (47-67) (**5.10**) and the N-terminal IGF-II (Thz-46) thioester **5.11** is now described.

5.2.2.1 Synthesis of the C-terminal fragment: IGF-II (47-67) (5.10)

The C-terminal fragment (**5.10**) was synthesised as depicted in Scheme 17 from commercially available Tentagel® S RAM resin (**5.18**) using automated Fmoc-SPPS. The crude peptide was cleaved from the resin on treatment with TFA/TIPS/DODT/water and the resulting crude IGF-II (47-67) peptide (**5.10**) was analysed by LCMS as shown in Figure 54A.



Scheme 17: Synthesis of the C-terminal fragment, IGF-II (47-67) (5.10). Where TIPS = triisopropylsilane and DODT = 3,6-dioxa-1,8-octanedithiol.

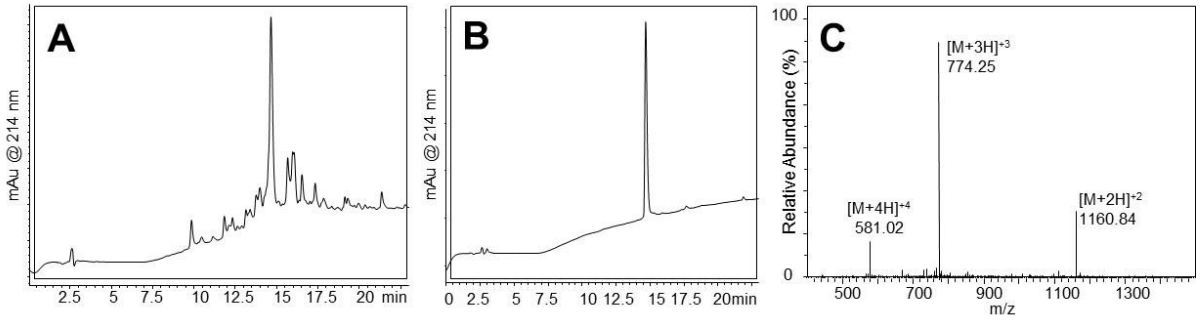
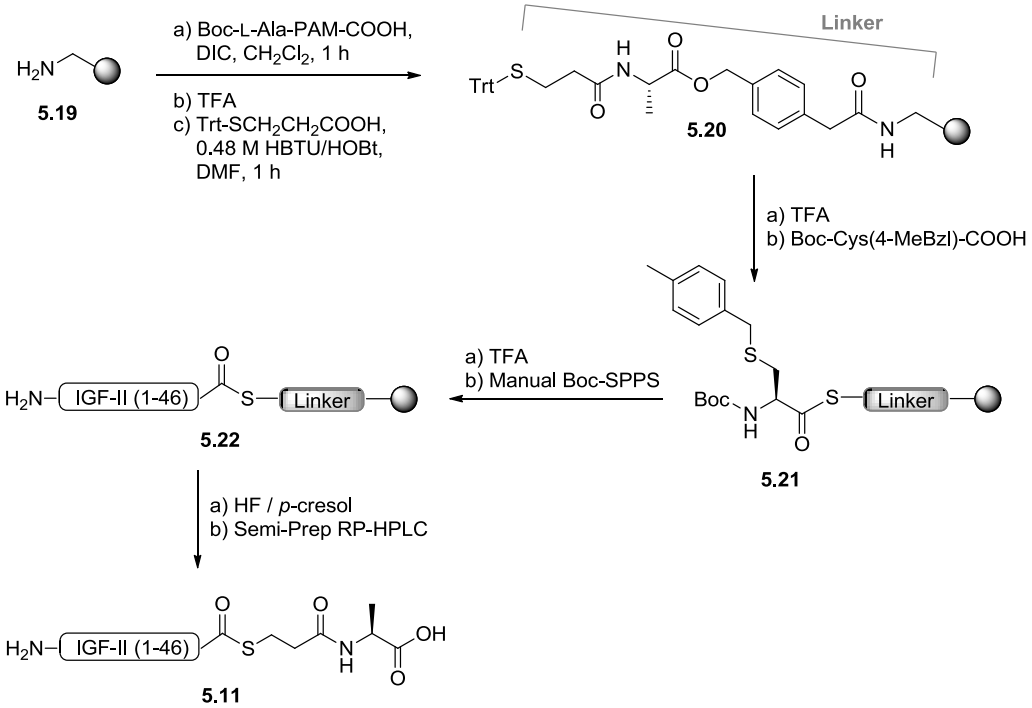


Figure 54: Analysis of the C-terminal fragment, IGF-II (47-67) (5.10). a) LCMS of crude **5.10**; b) LCMS of purified **5.10**; c) ESI-MS of purified **5.10**. Calcd. for $C_{99}H_{158}N_{26}O_{32}S_3$: 2320.6720 (average isotopes); observed: 1160.84 ($[M+2H]^{+2}$), 774.25 ($[M+3H]^{+3}$) and 581.02 ($[M+4H]^{+4}$).

LCMS analysis of crude peptide **5.10** is shown in Figure 54A and displays a major peak at 14.6 min. Analysis of the MS data for this peak revealed *m/z* ions ($[M+3H]^{+3}$; 774.41 *m/z*) that agreed with the calculated *m/z* ions ($[M+3H]^{+3}$; 774.57 *m/z*) for the IGF-II (47-67) fragment (**5.10**) (data not shown). This result suggested fragment **5.10**, was the major component of the mixture. Based on this, the crude peptide was purified by semi-preparative RP-HPLC to give the C-terminal fragment (**5.10**) in a respectable yield (13%). Fragment **5.10** was synthesised in high purity, as confirmed by the presence of a single peak in the LCMS shown in Figure 54B. This peak gave rise to *m/z* ions ($[M+3H]^{+3}$; 774.25 *m/z*) in the ESI-MS shown in Figure 54C, which are consistent with the expected *m/z* ions ($[M+3H]^{+3}$; 774.57 *m/z*). With the successful synthesis of the C-terminal fragment, the synthesis of the 46 residue N-terminal thioester **5.11** was undertaken.

5.2.2.2 Synthesis of the N-terminal thioester: IGF-II (1-46) (5.11)

The N-terminal IGF-II (1-46) thioester (**5.11**) was synthesised as illustrated in Scheme 18. Thioester **5.11** was assembled from resin **5.19**[xv] using a manual *in situ* neutralisation Boc-SPPS protocol.[221-223] First the Boc-L-Ala-PAM linker was coupled to resin **5.19** and the Boc group removed on treatment with neat TFA. Trityl protected mercaptopropanoic acid was then coupled to the resultant free amine to give the resin-bound thioester linker **5.20**. Resin **5.20** was deprotected of its Trt group using neat TFA, and Boc-Cys(4-MeBzl)-COOH was coupled to the resultant free thiol, to afford the resin-bound C-terminal cysteine alkyl thioester **5.21**. Resin **5.21** was elongated using a manual *in situ* neutralisation Boc-SPPS protocol to give the resin-bound N-terminal thioester **5.22**. Resin **5.22** was cleaved on treatment with HF and *p*-cresol. The resultant crude N-terminal thioester **5.11** was analysed by LCMS, and the result is displayed in Figure 55.



Scheme 18: Synthesis of the N-terminal IGF-II (1-46) thioester (5.11), using *in situ* neutralisation Boc-SPPS.

[xv] Aminomethyl resin (**5.19**) was prepared 'in house' by Harris and co-workers.

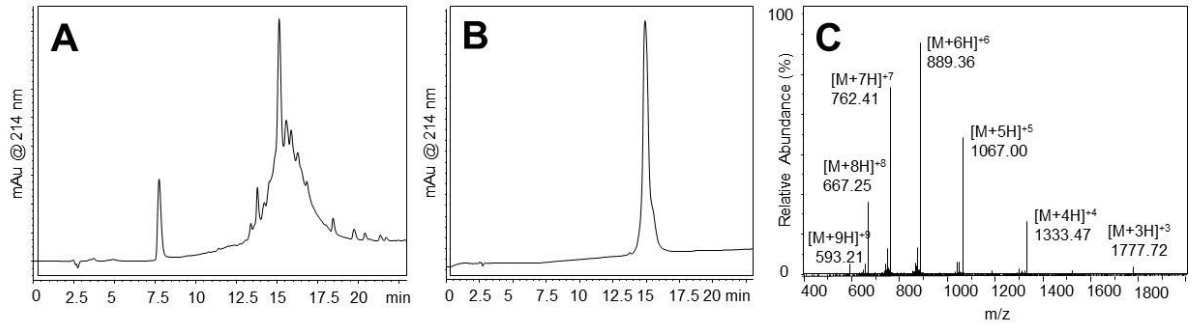
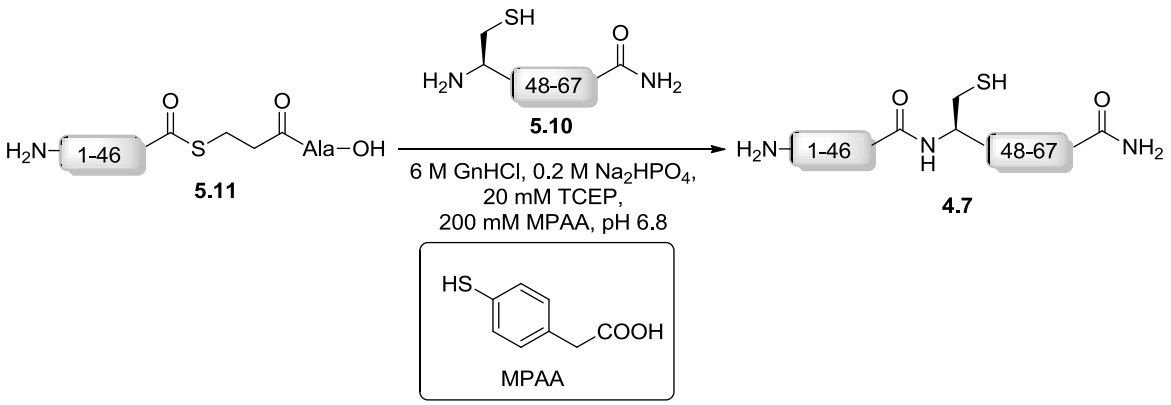


Figure 55: Analysis of the N-terminal IGF-II (1-46) thioester 5.11. a) LCMS analysis of crude **5.11**; b) LCMS analysis of purified **5.11**; c) ESI-MS of purified **5.11**. Calcd. for $C_{228}H_{359}N_{69}O_{71}S_4$: 5330.9859 (average isotopes); observed: 1777.72 ($[M+3H]^{+3}$), 1333.47 ($[M+4H]^{+4}$), 1067.00 ($[M+5H]^{+5}$), 889.36 ($[M+6H]^{+6}$), 762.41 ($[M+7H]^{+7}$), 667.25 ($[M+8H]^{+8}$) and 593.21 ($[M+9H]^{+9}$).

LCMS analysis of crude thioester **5.11** is shown in Figure 55A and displays a peak at 15.2 min. This peak gave rise to *m/z* ions ($[M+6H]^{+6}$; 889.40 *m/z*) which were consistent with the expected *m/z* ions ($[M+6H]^{+6}$; 889.50 *m/z*) for N-terminal thioester **5.11** (data not shown). This result indicated thioester **5.11** was the major component of the mixture. The N-terminal IGF-II (1-46) thioester (**5.11**) was isolated by semi-preparative RP-HPLC in a good yield (5%). The high purity of thioester **5.11** is demonstrated in the LCMS trace shown in Figure 55B. MS analysis of **5.11** is shown in Figure 55C and displays *m/z* ions ($[M+6H]^{+6}$; 889.36 *m/z*) that are consistent with the calculated *m/z* ions ($[M+6H]^{+6}$; 889.50 *m/z*) for the N-terminal IGF-II (1-46) thioester (**5.11**).

5.2.2.3 Synthesis of the native IGF-II protein (4.1) using the two fragment ligation approach

Having accomplished the synthesis of the C-terminal fragment **5.10** and the N-terminal thioester **5.11**, these fragments were assembled, using native chemical ligation into the native IGF-II peptide (**4.7**) as depicted in Scheme 19.



Scheme 19: Synthesis of the native IGF-II peptide (4.7), using a two fragment ligation approach.

Ligation of the N- and C- terminal fragments, **5.11** and **5.10** is outlined in Scheme 19, and was carried out using standard native chemical ligation conditions, 6 M GnHCl, 200 mM Na_2HPO_4, 20 mM TCEP, 200 mM MPAA at a pH of 6.8. MPAA is a water soluble and non-malodourous thiol catalyst.[218] It has been shown to have superior catalytic properties in ligation reactions compared to the widely used thiophenol and as such has been used in this study.[218] The progress of the ligation was monitored by LCMS as shown in Figure 56 and was found to be quantitatively complete within 2 h. Figure 56A shows the N- and C- terminal fragments, **5.11** and **5.10** after 3 min and Figure 56B shows the formation of the desired native IGF-II peptide (**4.7**) after 2 h. The formation of peptide **4.7** was confirmed by the agreement of the *m/z* ions ($[M+7H]^{+7}$; 1068.63 *m/z*) observed in the MS displayed in Figure 56C, with the expected *m/z* ions ($M+7H]^{+7}$; 1068.78 *m/z*) for the native IGF-II peptide (**4.7**). Finally, the trace amount of the C-terminal fragment **5.10** remaining in the LCMS shown in Figure 56B, was due to fragment **5.10** being used in excess (1:1.36 ratio) to ensure the ligation reached completion.

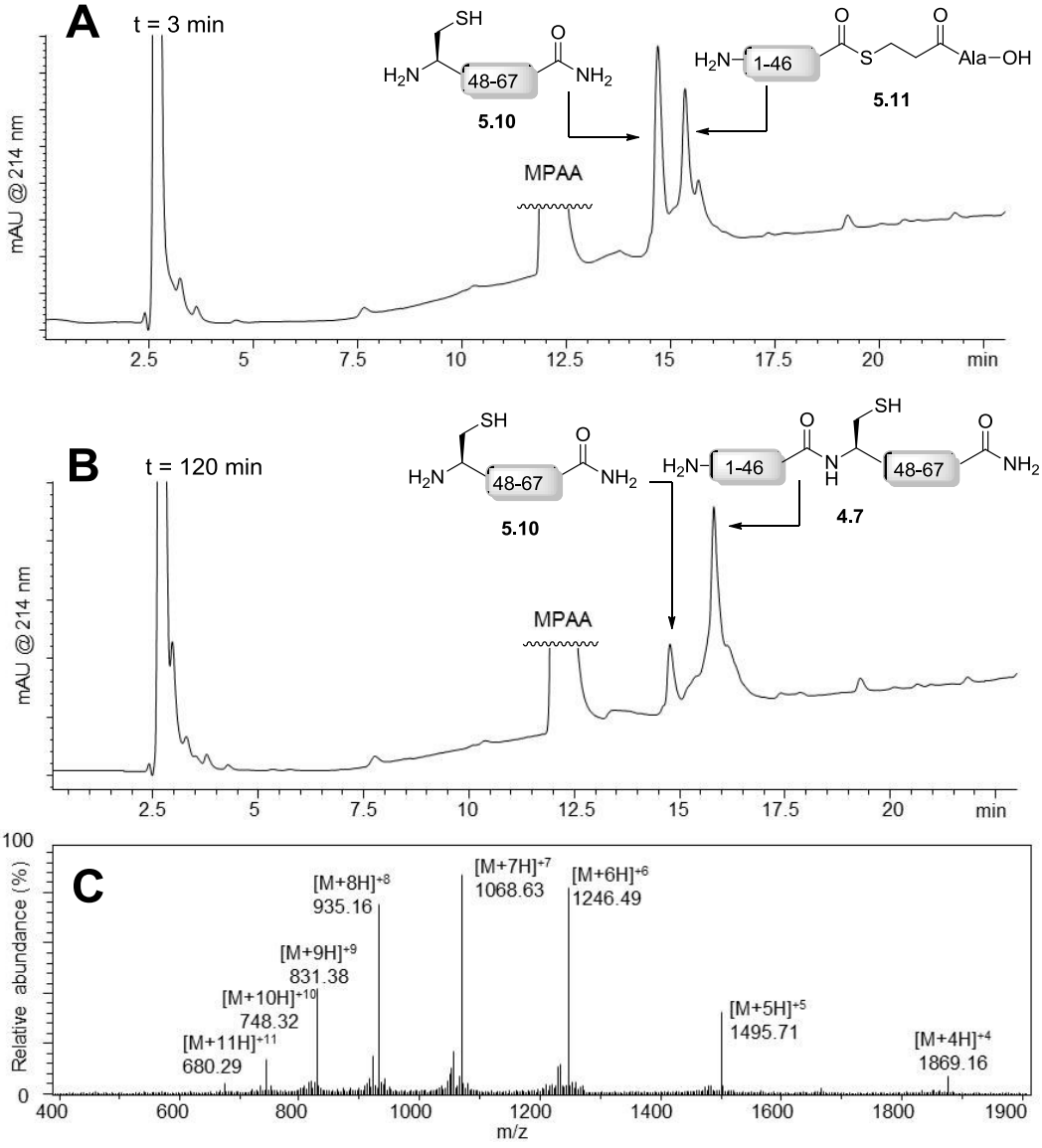


Figure 56: Two fragment ligation of native IGF-II peptide (4.7) at a concentration of 1 mM. a) LCMS analysis of the crude ligation mixture after 3 min; b) LCMS analysis of the crude ligation mixture after 120 min; c) ESI-MS of peptide **4.7**. Calcd. for $C_{321}H_{506}N_{94}O_{100}S_6$: 7474.4351 (average isotopes); observed: 1869.16 ($[M+4H]^{+4}$), 1495.71 ($[M+5H]^{+5}$), 1246.49 ($[M+6H]^{+6}$), 1068.63 ($[M+7H]^{+7}$), 935.16 ($[M+8H]^{+8}$), 831.38 ($[M+9H]^{+9}$), 748.32 ($[M+10H]^{+10}$) and 680.29 ($[M+11H]^{+11}$).

Encouraged by the successful ligation of fragments **5.11** and **5.10**, a brief optimisation of the ligation conditions was undertaken where fragment concentrations of 1, 2 and 3 mM were investigated with all other reaction variables kept constant. It was found that the ligation could be improved by increasing the concentration of the peptide fragments. By increasing the concentration of the peptide fragments from 1 mM to 3 mM, the reaction time decreased from 2-3 h to 1 h.

The LCMS analysis of the ligation performed at a concentration of 3 mM is displayed in Figure 57. Unlike the ligation performed at a concentration of 1 mM shown in Figure 56 a significant amount of product (**4.7**) was observed in the LCMS after only 2 min (refer to Figure 57A). More importantly, the ligation was determined to be complete after 1 h by the disappearence of **5.11** from the LCMS shown in Figure 57B. The MS data obtained for both the 1 mM and 3 mM ligations were essentially identical and the observed *m/z* ions ($[M+7H]^{+7}$; 935.15 *m/z*) displayed in Figure 57C are again consistent with the expected *m/z* ions ($[M+7H]^{+7}$; 935.31 *m/z*) of the native IGF-II peptide (**4.7**). Peptide **4.7** was isolated from the ligation mixture using solid phase extraction (SPE) and purified by semi-preparative RP-HPLC to give the desired linear native IGF-II peptide (**4.7**) in a moderate yield (4%) and in high purity, as shown in the RP-HPLC displayed in Figure 57D.

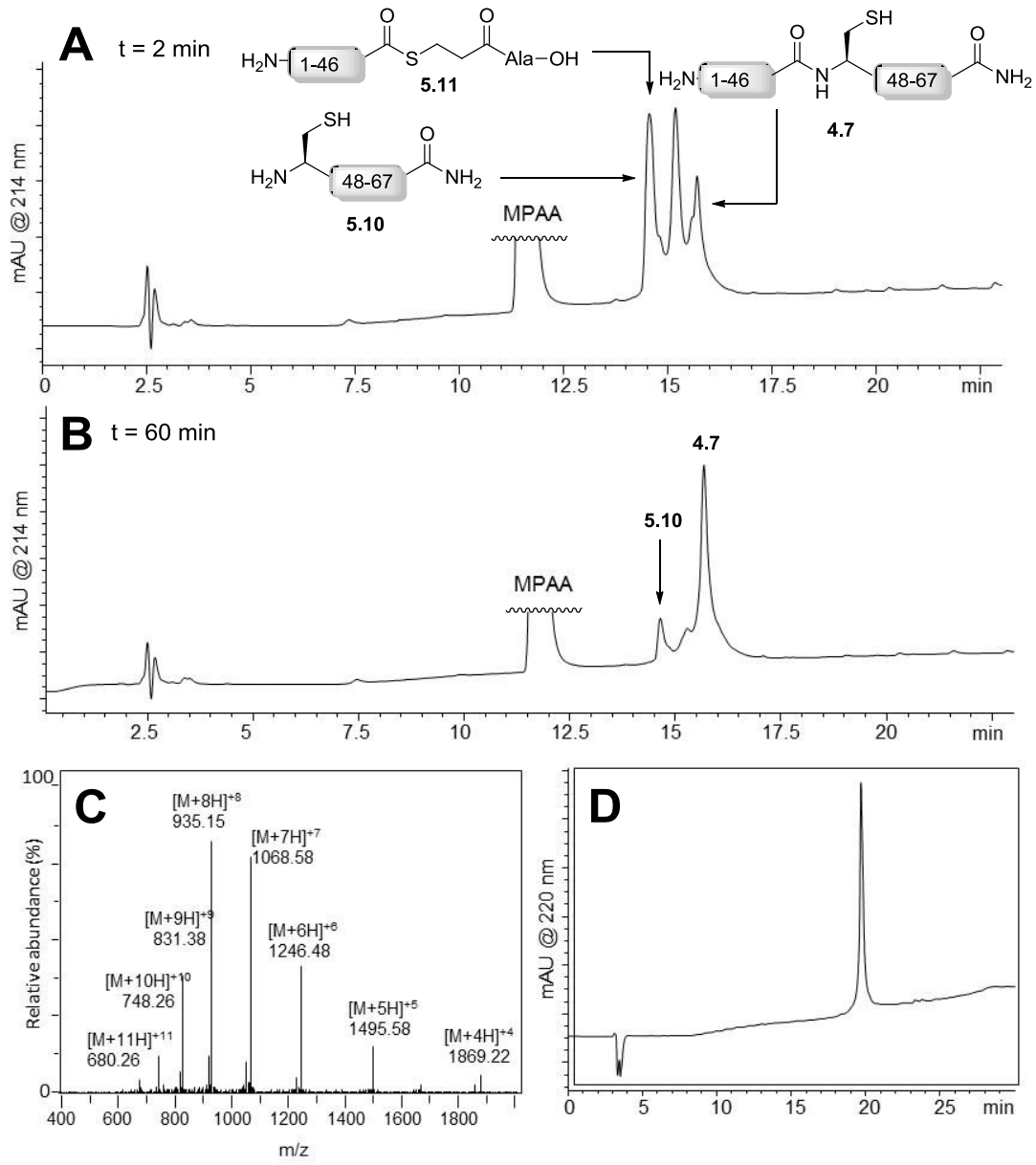
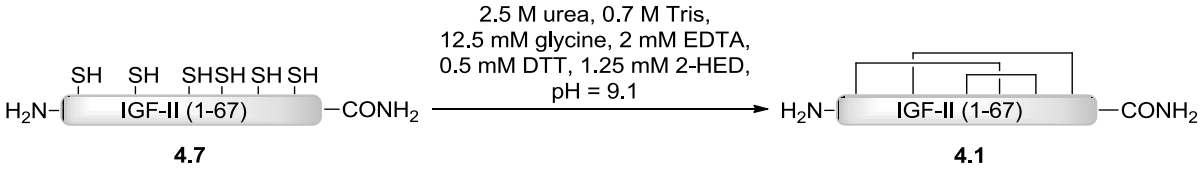


Figure 57: Two fragment ligation of native IGF-II peptide (4.7) at a concentration of 3 mM. a) LCMS analysis of the crude ligation mixture after 2 min; b) LCMS analysis of the crude ligation mixture after 60 min; c) ESI-MS of peptide **4.7**. Calcd. for $C_{321}H_{506}N_{94}O_{100}S_6$: 7474.4351 (average isotopes); observed: 1869.22 ($[M+4H]^{+4}$), 1495.58 ($[M+5H]^{+5}$), 1246.48 ($[M+6H]^{+6}$), 1068.58 ($[M+7H]^{+7}$), 935.15 ($[M+8H]^{+8}$), 831.38 ($[M+9H]^{+9}$), 748.26 ($[M+10H]^{+10}$) and 680.26 ($[M+11H]^{+11}$); d) RP-HPLC of the purified IGF II (1-67) peptide (**4.7**).

The resultant linear IGF-II peptide (**4.7**) was oxidised into its biologically active form, as outlined in Scheme 20, following the protocol reported by Delaine *et al.*[23] The native IGF-II protein (**4.1**) was isolated in a reasonable yield (5%) using semi-preparative RP-HPLC. This gave the desired native IGF-II protein (**4.1**) in high purity as depicted in the RP-HPLC shown

in Figure 58A. Successful synthesis of **4.1** was confirmed by the agreement of the *m/z* ions ($[M+6H]^{+6}$; 1245.79 *m/z*) displayed in the HRMS shown in Figure 58B, with the expected *m/z* ions ($[M+6H]^{+6}$; 1245.73 *m/z*) for the native IGF-II protein (**4.1**).



Scheme 20: Synthesis of the synthetic native IGF-II protein (4.1) from the two fragment ligation approach.

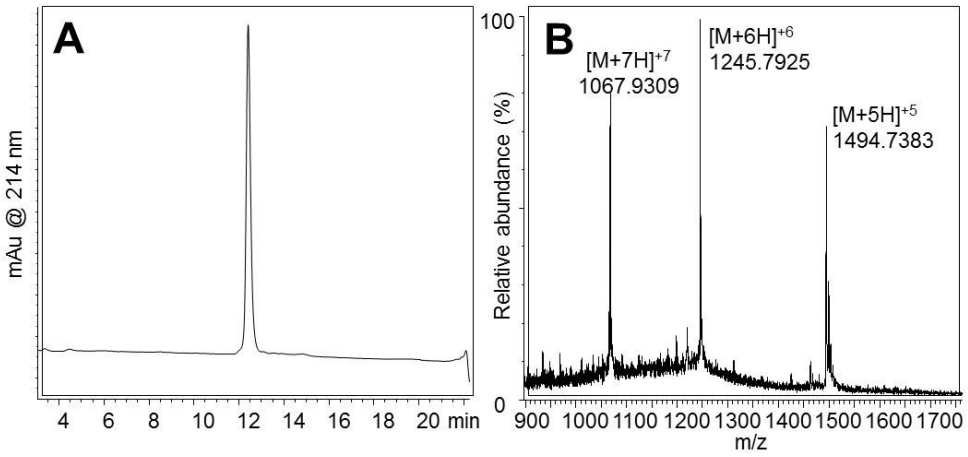


Figure 58: Characterisation of the synthetic native IGF-II protein (4.1) synthesised using the two fragment approach. a) RP-HPLC trace of **4.1**; b) HRMS of **4.1**. Calcd. for $C_{321}H_{500}N_{94}O_{100}S_6$: 7468.3875 (average isotopes); observed: 1494.7383 ($[M+5H]^{+5}$), 1245.7925 ($[M+6H]^{+6}$) and 1067.9309 ($[M+7H]^{+7}$).

The two fragment approach described above afforded the native IGF-II protein (**4.1**) in reasonable yields and high purity. However synthesis of the F19Cou IGF-II (**4.2**) and F28Cou IGF-II proteins (**5.1**) using this two fragment approach was not desirable, as it would require the synthesis of two additional 46-residue fragments. As discussed in Section 5.2.1, the synthesis of these large fragments is impractical, due to the synthetic inefficiencies and

difficulties with peptide purification which arise from long sequences. Thus, the modular approach was preferred for the synthesis of F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins due to the increased efficiency associated with synthesising smaller peptide fragments. The synthesis of the native IGF-II (**4.1**), F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins using the modular approach outlined in Section 5.2.1, is now described.

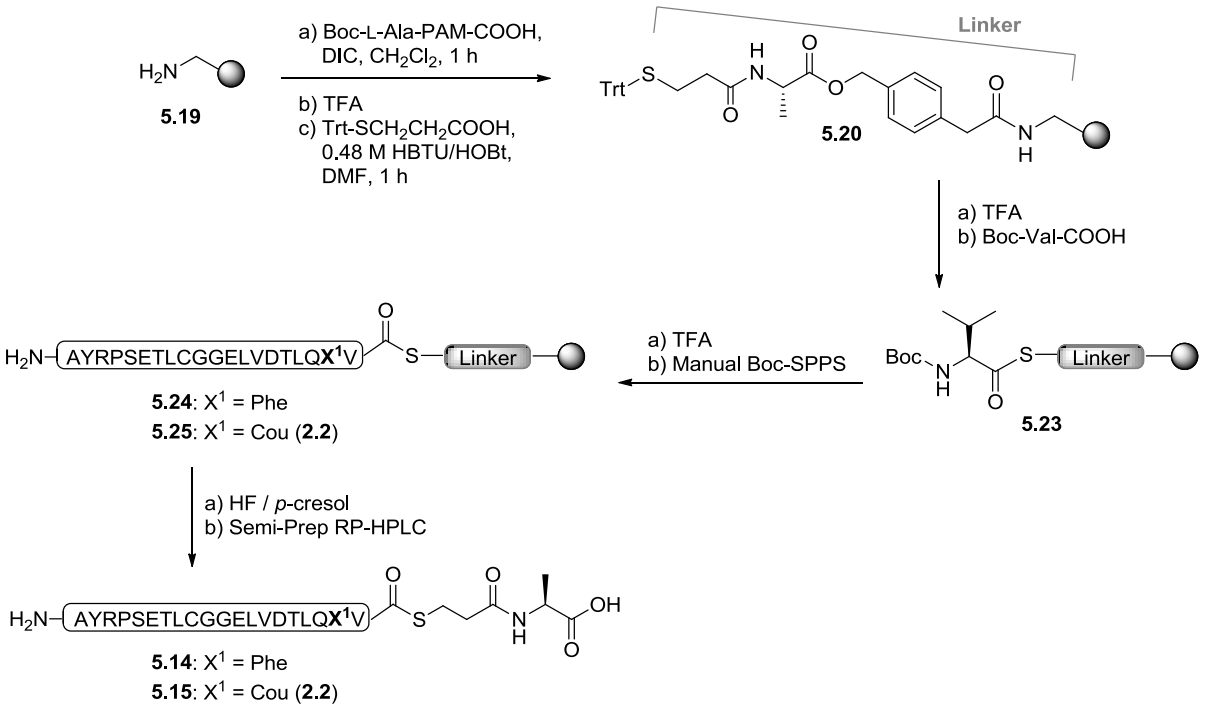
5.2.3 Three fragment ligation approach

Encouraged by the successful synthesis of the native IGF-II protein (**4.1**) using the two fragment approach, our attention was turned to developing a more efficient, modular approach. As discussed in Section 5.2.1, the Cys^{46}-Cys^{47} and Val^{20}-Cys^{21} ligation sites were selected, which gave rise to the fragments **5.14**, **5.15**, **5.16**, and **5.17** which were all of a similar size (approx. 20 residues). It was envisioned that the desired IGF-II analogues (**4.1**, **4.2** and **5.1**) could be assembled in one-pot, from three fragments using iterative cysteine and valine ligations. With the common C-terminal fragment (**5.10**) in hand (synthesised described in Section 5.2.2.1), the synthesis of thioesters **5.14**, **5.15**, **5.16**, and **5.17** was undertaken.

5.2.3.1 Synthesis of the N-terminal thioesters 5.14 and 5.15

The N-terminal thioesters **5.14** and **5.15** were synthesised as illustrated in Scheme 21. Thioesters **5.14** and **5.15** were assembled from resin **5.19**[xvi] using a manual *in situ* neutralisation Boc-SPPS protocol.[221-223] Elongation of resin **5.19** to the resin-bound thioester linker **5.20** was accomplished as detailed in Section 5.2.2.2. Coupling of Boc-Val-COOH to resin **5.20** yielded the resin-bound C-terminal valine alkyl thioester **5.23**. Elaboration of resin **5.23** by manual Boc-SPPS gave the N-terminal resin-bound thioesters, IGF-II (1-20) (**5.24**) and F19Cou IGF-II (1-20) (**5.25**).

[xvi] Aminomethyl resin (**5.19**) was prepared 'in house' by Harris and co-workers.



Scheme 21: Synthesis of the N-terminal thioesters 5.14 and 5.15, using *in situ* neutralisation Boc-SPPS.

Thioesters **5.14** and **5.15** were released from their respective resins (**5.24** and **5.25**) on treatment with HF and *p*-cresol. The resultant crude thioesters **5.14** and **5.15** were analysed by LCMS and the results are displayed in Figure 59A and Figure 60A respectively.

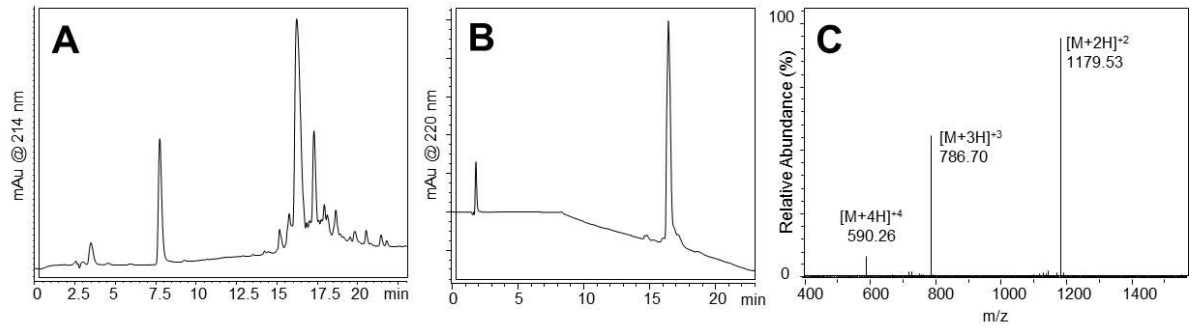


Figure 59: Analysis of the N-terminal IGF-II (1-20) peptide thioester 5.14. a) LCMS analysis of crude **5.14**; b) LCMS analysis of purified **5.14**; c) ESI-MS of purified **5.14**. Calcd. for $C_{103}H_{161}N_{25}O_{34}S_2$: 2357.6648 (average isotopes); observed: 1179.53 ($[M+2H]^{+2}$), 786.70 ($[M+3H]^{+3}$) and 590.26 ($[M+4H]^{+4}$).

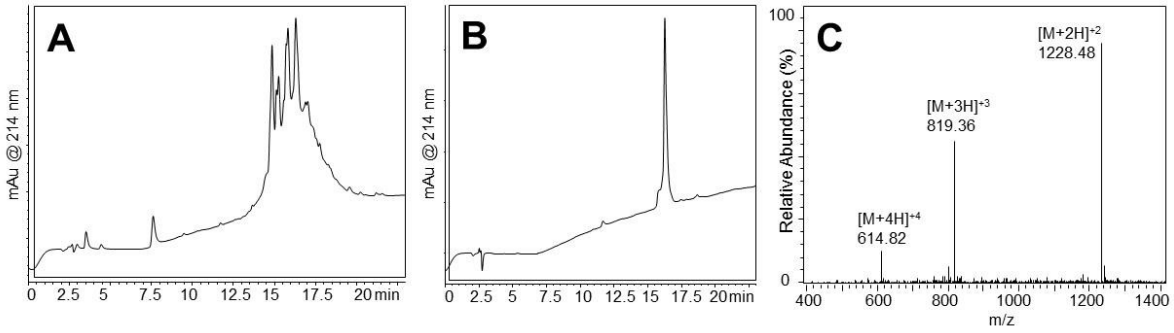


Figure 60: Analysis of the N-terminal F19Cou IGF-II (1-20) peptide thioester 5.15. a) LCMS analysis of crude **5.15**; b) LCMS analysis of purified **5.15**; c) ESI-MS of purified **5.15**. Calcd. for $C_{107}H_{163}N_{25}O_{37}S_2$: 2455.7218 (average isotopes); observed: 1228.48 ($[M+2H]^{+2}$), 819.36 ($[M+3H]^{+3}$) and 614.82 ($[M+4H]^{+4}$).

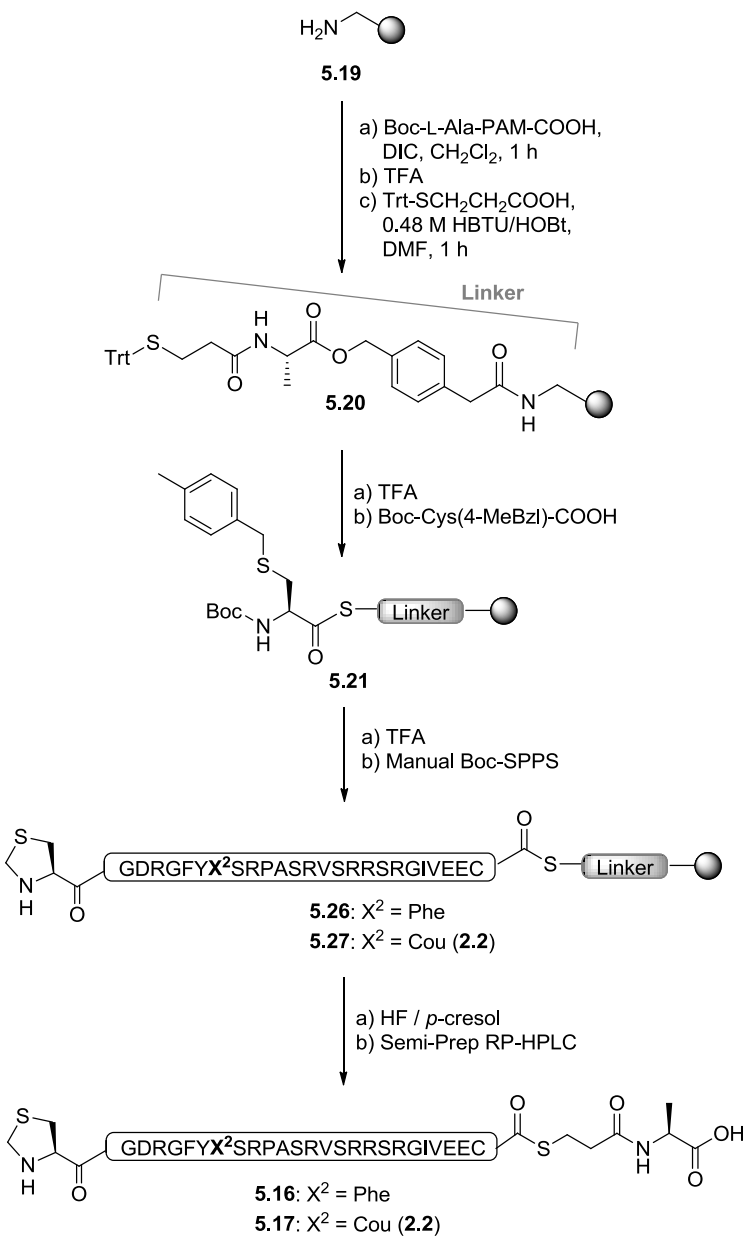
LCMS analysis of crude thioester **5.14** is shown in Figure 59A, and gave rise to a peak a 16.2 min. MS analysis of this peak revealed *m/z* ions ($[M+2H]^{+2}$; 1179.51 *m/z*) that agreed with the expected *m/z* ions ($[M+2H]^{+2}$; 1179.84 *m/z*) for desired N-terminal thioester **5.14** (data not shown). This result confirmed thioester **5.14** was present as a major component of the mixture. Crude thioester **5.14** was purified by semi-preparative RP-HPLC to afford the N-terminal IGF-II (1-20) thioester (**5.14**) in an good yield (17%). Purified thioester **5.14** is displayed in Figure 59B and gave rise to *m/z* ions ($[M+2H]^{+2}$; 1179.53 *m/z*) in the MS displayed in Figure 59C, which agreed with the expected *m/z* ions ($[M+2H]^{+2}$; 1179.84 *m/z*) for the IGF-II (1-20) thioester (**5.14**).

Similarly, LCMS analysis of crude thioester **5.15** is shown in Figure 60A and gave rise to three distinct peaks. MS analysis of these peaks revealed, the peak at 16.3 min gave rise to *m/z* ions ($[M+2H]^{+2}$; 1228.69 *m/z*) that were consistent with the expected *m/z* ions ($[M+2H]^{+2}$; 1228.87 *m/z*) for the desired N-terminal thioester **5.15** (data no shown). These results confirmed thioester **5.15** was a main component of the mixture. The N-terminal F19Cou IGF-II (1-20) thioester (**5.15**) was isolated by semi-preparative RP-HPLC in an acceptable yield (2%) and the purified thioester is shown in Figure 60B. Finally, MS analysis of thioester **5.15** is shown in Figure 60C and displays *m/z* ions ($[M+2H]^{+2}$; 1228.48 *m/z*) which are consistent with the expected *m/z* ions ($[M+2H]^{+2}$; 1228.87 *m/z*) for the F19Cou IGF-II (1-20) thioester (**5.15**). These results confirm the successful synthesis of the N-terminal thioesters **5.14** and **5.15**.

As discussed in Section 4.2.4 (refer to Chapter 4), difficulties with the N-terminal assembly of the native IGF-II peptides (**4.7**, **4.11** and **4.15**) had been encountered.[224] It is important to report that problems with on-resin aggregation were not encountered during the synthesis of the N-terminal thioesters **5.11** (refer to Section 5.2.2.2), **5.14** and **5.15**. The successful synthesis of these fragments was attributed to using a manual *in situ* neutralisation Boc-SPPS protocol.[223] Specifically, it was thought the use of trifluoroacetic acid (TFA) in the deprotection of the *N*α-Boc amino functionality was key, as this has previously been reported to disrupt on-resin aggregation.[223,225]

5.2.3.2 Synthesis of peptide thioesters 5.16 and 5.17

Peptide thioesters **5.16** and **5.17** were synthesised as detailed in Scheme 22, from aminomethyl resin[xvii] (**5.19**) using an *in situ* neutralisation Boc-SPPS protocol.[221-223]



Scheme 22: Synthesis of peptide thioesters 5.16 and 5.17 using *in situ* neutralisation Boc-SPPS.

[xvii] Aminomethyl resin (**5.19**) was prepared 'in house' by Harris and co-workers.

First, the resin-bound thioester linker **5.20** was assembled from resin **5.19** in a same manner as described in Section 5.2.2.2. Resin **5.20** was elongated using a manual *in situ* neutralisation Boc-SPPS protocol to give the resin-bound thioesters **5.26** and **5.27** (refer to Scheme 22). Resins **5.26** and **5.27** were cleaved on treatment with HF and *p*-cresol. The resultant crude peptide thioesters **5.16** and **5.17** were analysed by LCMS, and the results are displayed in Figure 61A and Figure 62A respectively.

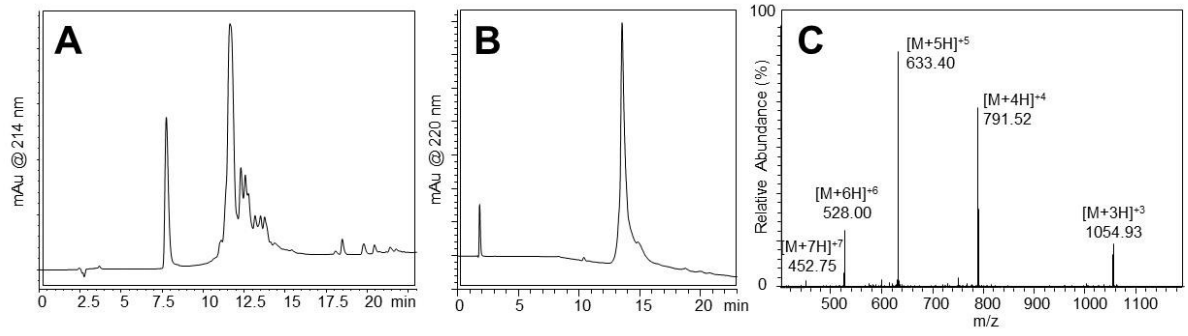


Figure 61: Analysis of the IGF-II (Thz-46) thioester (5.16). a) LCMS analysis of crude **5.16**; b) LCMS analysis of purified **5.16**; c) ESI-MS of purified **5.16.** Calcd. for $C_{136}H_{209}N_{45}O_{40}S_3$: 3162.5546 (average isotopes); observed: 1054.93 ($[M+3H]^{+3}$), 791.52 ($[M+4H]^{+4}$), 633.40 ($[M+5H]^{+5}$), 528.00 ($[M+6H]^{+6}$) and 452.75 ($[M+7H]^{+7}$).

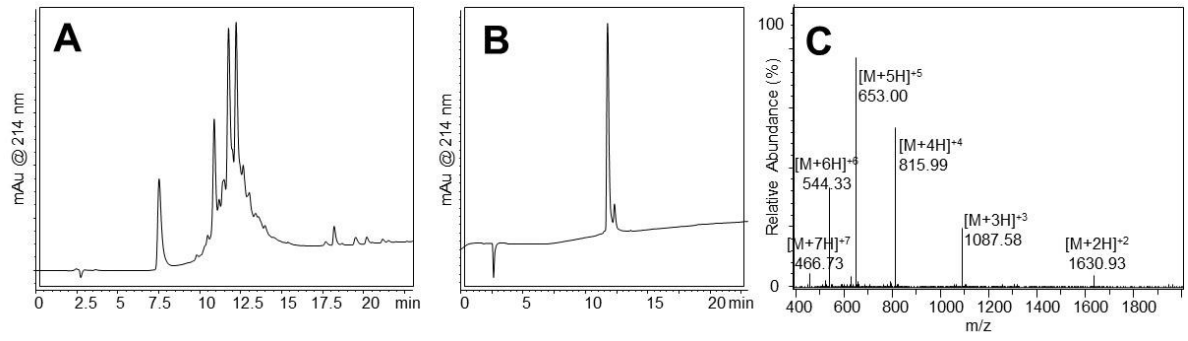


Figure 62: Analysis of the F28Cou IGF-II (Thz-46) thioester (5.17). a) LCMS analysis of crude **5.17**; b) LCMS analysis of purified **5.17**; c) ESI-MS of purified **5.17**. Calcd. $C_{136}H_{211}N_{45}O_{43}S_3$ for: 3260.6117 (average isotopes); observed: 1630.93 ($[M+2H]^{+2}$), 1087.58 ($[M+3H]^{+3}$), 815.99 ($[M+4H]^{+4}$), 653.00 ($[M+5H]^{+5}$), 544.33 ($[M+6H]^{+6}$) and 466.73 ($[M+7H]^{+7}$).

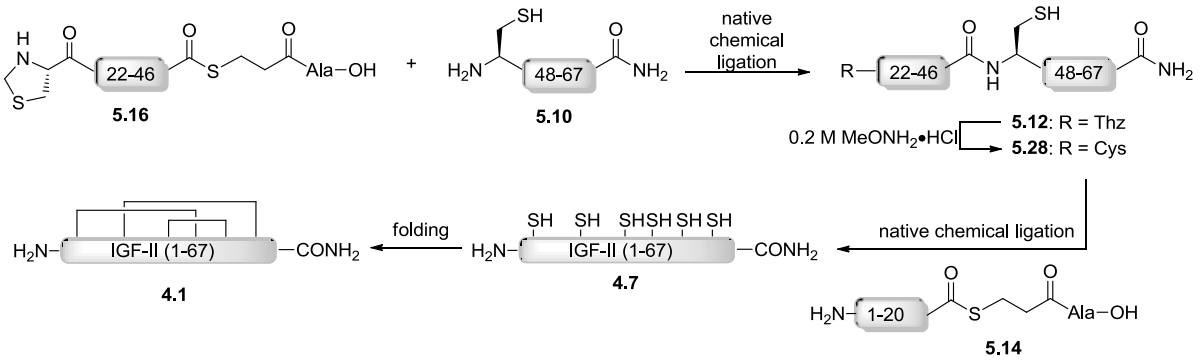
Analysis of crude thioester **5.16** furnished a peak at 11.6 min in the LCMS shown in Figure 61A. This peak gave rise to *m/z* ions ($[M+5H]^{+5}$; 633.40 *m/z*) which were consistent with the calculated *m/z* ions for the IGF-II (Thz-46) thioester **5.16** ($[M+5H]^{+5}$; 633.51 *m/z*) (data not shown). Likewise, a peak at 11.7 min in the LCMS shown in Figure 62A for crude thioester **5.17,** gave rise to *m/z* ions ($M+5H]^{+5}$; 653.00 *m/z*) which agreed with the expected *m/z* ions ($M+5H]^{+5}$; 653.12 *m/z*) for the F28Cou IGF-II (Thz-46) thioester **5.17** (data not shown). These results suggested that the desired thioesters **5.16** and **5.17** were present as the major component of their respective mixtures.

Encouraged by this, crude thioesters **5.16** and **5.17** were purified by semi-preparative RP-HPLC, to afford the desired thioesters, **5.16** and **5.17** in 11% and 3% yields respectively. Thioesters **5.16** and **5.17** were synthesised in high purity as indicated by the LCMS traces shown in Figure 61B and Figure 62B respectively. MS analysis of thioester **5.16** is shown in Figure 61C and displays *m/z* ions ($[M+5H]^{+5}$; 633.40 *m/z*) which agree with the calculated *m/z* ions ($[M+5H]^{+5}$; 633.51 *m/z*) for the IGF-II (Thz-46) thioester (**5.16**). Likewise MS analysis of thioester **5.17** is shown in Figure 62C and displays *m/z* ions ($[M+5H]^{+5}$; 653.00 *m/z*) which are consistent with the expected *m/z* ions ($[M+5H]^{+5}$; 653.12 *m/z*) for the F28Cou IGF-II (Thz-46) thioester (**5.17**). These results confirm the successful synthesis of peptide thioesters **5.16** and **5.17**. In these fragments Cys^{21} was replaced by a protected the N-terminal cysteine derivative, thiazolidone (Thz), to prevent head to tail cyclisation of the thioester fragments (**5.16** and **5.17**) during the ligation reaction.

With the desired thioesters in hand, these fragments were assembled into the native IGF-II (**4.1**), the F19Cou IGF-II (**4.2**) and the F28Cou IGF-II (**5.1**) proteins using native chemical ligation.

5.2.3.3 Synthesis of the native IGF-II protein (4.1) using the three fragment ligation approach

The assembly of the native IGF-II protein (**4.1**) from peptide fragments **5.14**, **5.16** and **5.10** is summarised in Scheme 23.



Scheme 23: Three fragment, one-pot synthesis of the native IGF-II protein (4.1). Native chemical ligation conditions: 6 M GnHCl, 200 mM Na_2HPO_4, 20 mM TCEP, 200 mM MPAA, pH of 6.9. Folding conditions: 2.5 M urea, 0.7 M Tris, 12.5 mM glycine, 2 mM EDTA, 0.5 mM DTT, 1.25 mM 2-HED at a pH of 9.1 and protein concentration of <0.1 mg/mL.

The assembly of the native IGF-II peptide (**4.7**) began with ligation of the C-terminal fragment **5.10** and the IGF-II (Thz-46)-thioester (**5.16**) at peptide concentration of 3 mM, using standard native chemical ligation conditions (6M GnHCl, 200 mM Na_2HPO_4, 20 mM TCEP, 200 mM MPAA, pH = 6.9). After only 3 min the IGF-II (Thz-67) fragment (**5.12**) could be detected in the LCMS shown in Figure 63A. The ligation was determined to be complete after 60 min, as evidenced by the absence of thioester **5.16** in the LCMS shown in Figure 63B. C-terminal fragment **5.10** was used in excess (1:1.12 ratio) to ensure complete conversion of thioester **5.16** to fragment **5.12**, thus trace amounts of fragment **5.10** were remaining at the completion of the ligation (refer to Figure 63B).

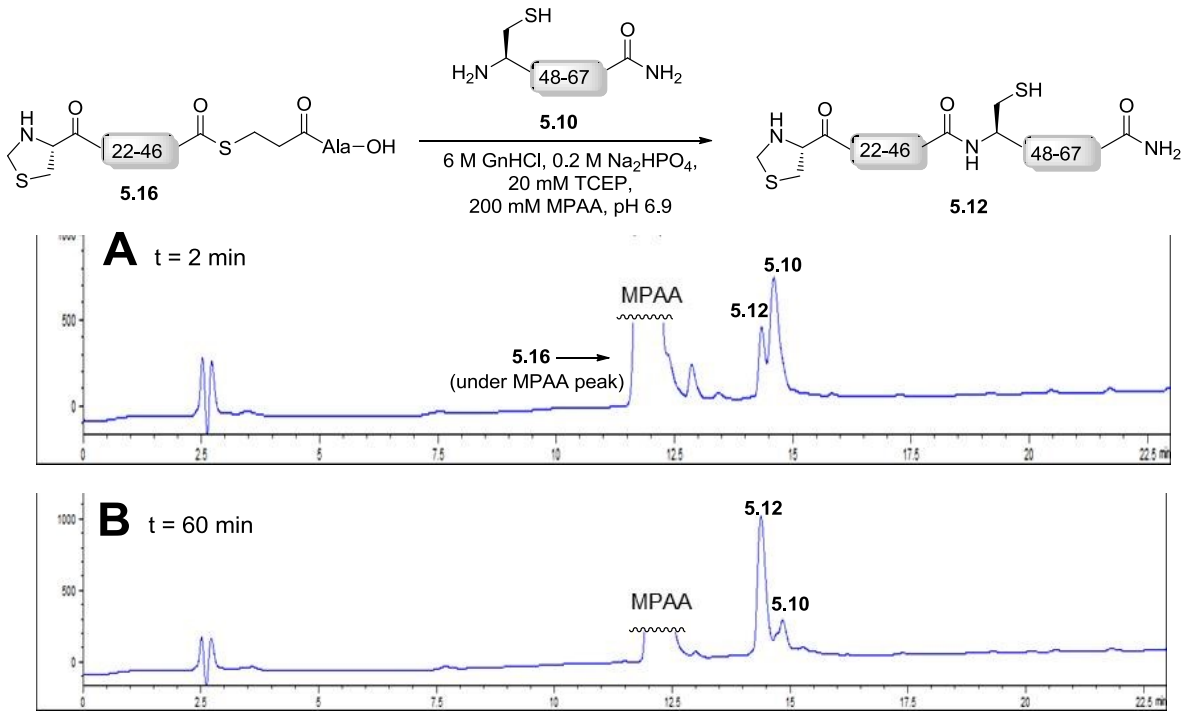


Figure 63: Analysis of the cysteine-based ligation between the IGF-II (Thz-47) thioester (5.16) and C-terminal fragment (5.10). a) LCMS analysis of the crude ligation mixture after 2 min; b) LCMS analysis of the crude ligation mixture after 60 min.

The N-terminal thiazolidone of fragment **5.12** was converted to an N-terminal cysteine (**5.28**) using methoxyamine hydrochloride (0.2 M). After 8 h fragment **5.12** was quantitatively converted to fragment **5.28**, which was confirmed by the absence on **5.12** in the LCMS shown in Figure 64A. ESI-MS analysis of deprotected fragment **5.28** is shown in Figure 64B and displays *m/z* ions ($[M+6H]^{+6}$; 833.20 *m/z*), that correlate to a loss of 12 Da, which is consistent with thiazolidone deprotection.

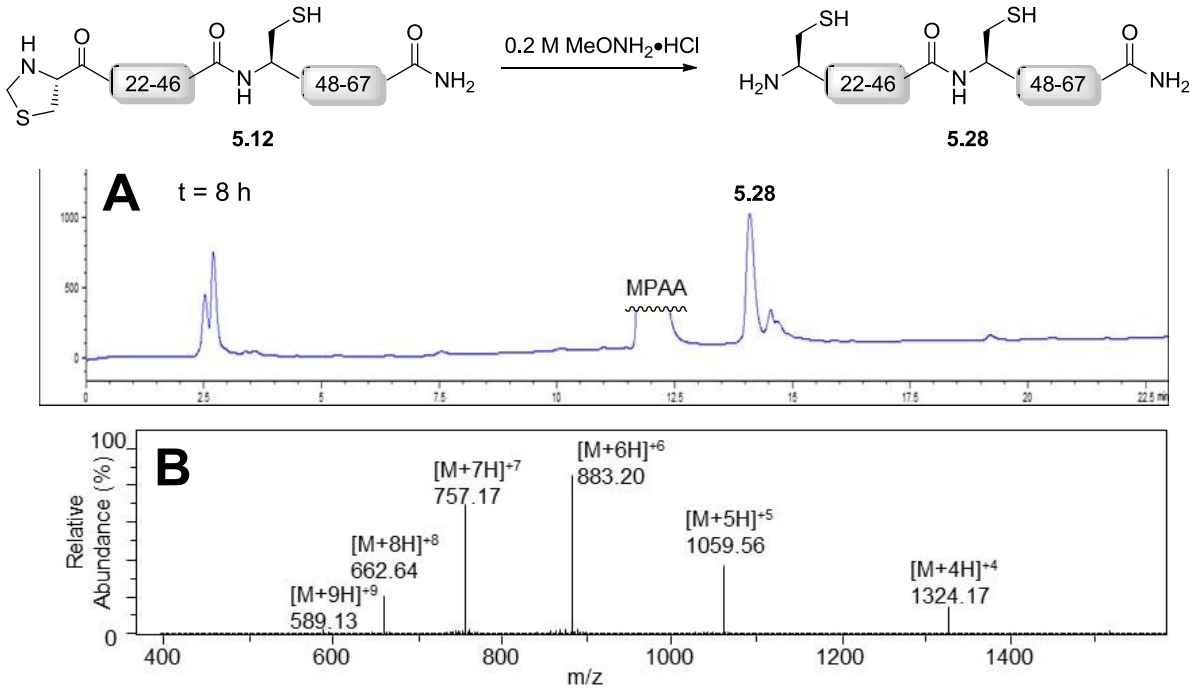


Figure 64: Analysis of the thiazolidone deprotection of fragment 5.12. a) LCMS of the crude thiazolidone deprotection after 8 h; b) ESI-MS of fragment **5.28**. Calcd. for $C_{224}H_{356}N_{70}O_{69}S_5$: 5293.9931 (average isotopes); observed: 1324.17 ($[M+4H]^{+4}$), 1059.56 ($[M+5H]^{+5}$), 883.20 ($[M+6H]^{+6}$), 757.17 ($[M+7H]^{+7}$), 662.64 ($[M+8H]^{+8}$) and 589.13 ($[M+9H]^{+9}$).

A second, valine-based ligation between the resultant IGF-II (21-67) fragment (**5.28**) and the IGF-II (1-20) thioester (**5.14**) was carried out as outlined in Figure 65. The ligation was determined to be complete after 36 h, by the absence of fragment **5.28** in the LCMS depicted in Figure 65A. The major peak at 14.2 min in the LCMS (refer to Figure 65A), gave rise to *m/z* ions ($[M+7H]^{+7}$; 1068.58 *m/z*) in the MS shown in Figure 65B. which agree with the expected *m/z* ions ($[M+7H]^{+7}$; 1068.78 *m/z*) for peptide **4.7**. These results confirm the desired native IGF-II peptide (**4.7**) was the major product of the one-pot, three fragment ligation.

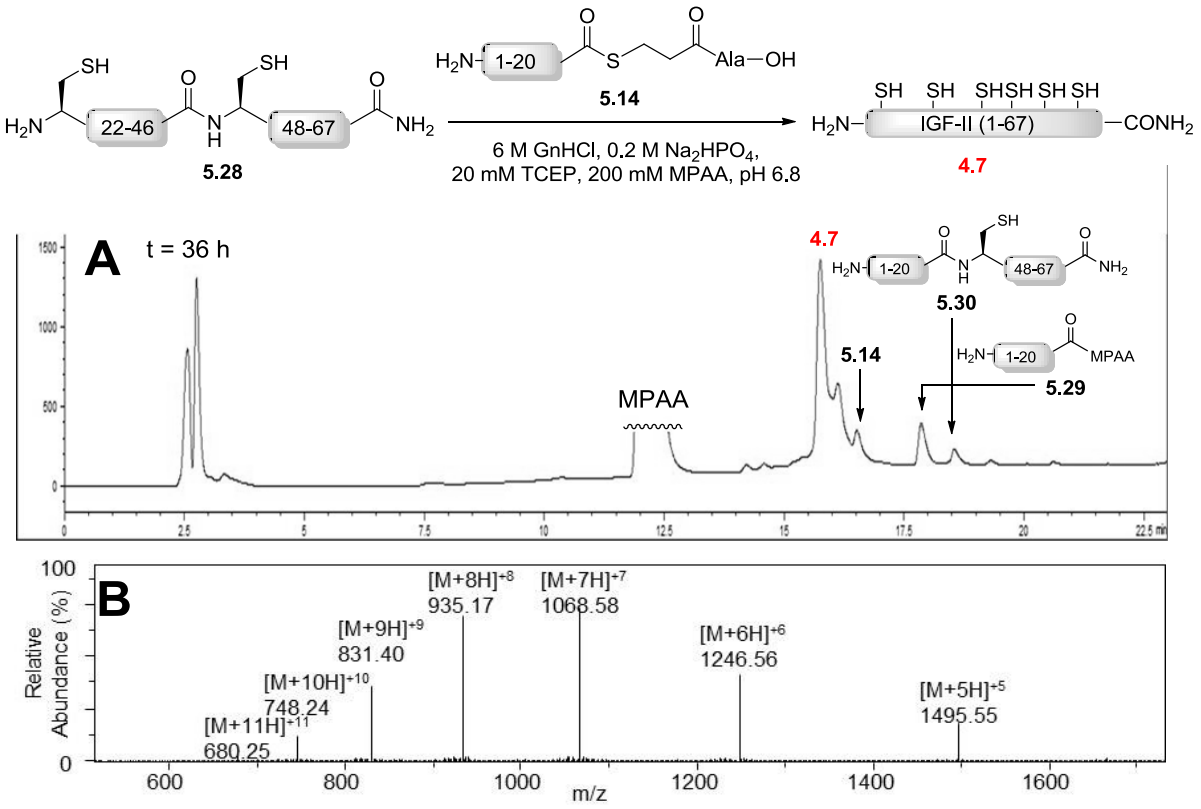


Figure 65: Analysis of the valine-based ligation between the IGF-II (21-67) fragment (5.28) and N-terminal IGF-II (1-20) thioester (5.14). a) LCMS analysis of the crude ligation mixture after 36 h; b) ESI-MS of peptide **4.7**. Calcd. for $C_{321}H_{506}N_{94}O_{100}S_6$: 7474.4351 (average isotopes); observed: 1495.55 ($[M+5H]^{+5}$), 1246.56 ($[M+6H]^{+6}$), 1068.58 ($[M+7H]^{+7}$), 935.17 ($[M+8H]^{+8}$), 831.40 ($[M+9H]^{+9}$), 748.24 ($[M+10H]^{+10}$) and 680.25 ($[M+11H]^{+11}$)

A number of intermediates were observed in LCMS analysis of the valine-based ligation of fragments **5.14** and **5.28** (refer to Figure 65). These included the MPAA exchanged thioester **5.29**, and mixed ligation product **5.30**, which is formed from the ligation of thioester **5.14** and residual fragment **5.10**. Despite this mixed ligation by-product **5.30** and the slow reactivity of the valine thioester, the ligation proceeded efficiently and to completion within 36 h.

The native IGF-II peptide (**4.7**) was isolated from the crude ligation mixture by SPE and the resultant peptide mixture was purified by semi-preparative RP-HPLC to afford peptide **4.7** in a reasonable yield (4%). Peptide **4.7** was subjected to oxidative folding following the protocol outlined by Delaine and co-workers to afford the native IGF-II protein (**4.1**).[23] The synthetic IGF-II protein (**4.1**) was isolated in moderate yield (6%) and high purity as indicated by the

RP-HPLC shown in Figure 66A. Successful synthesis of the native IGF-II protein (**4.1**) was confirmed by the agreement of the *m/z* ions ($[M+5H]^{+5}$; 1494.71 *m/z*) displayed in HRMS shown in Figure 66B, with the expected ions *m/z* ions ($[M+5H]^{+5}$; 1494.68 *m/z*) for the native IGF-II protein (**4.1**).

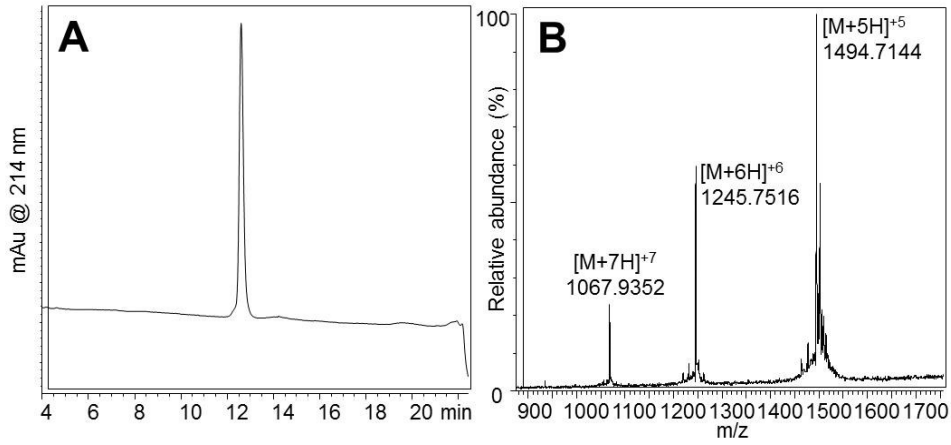
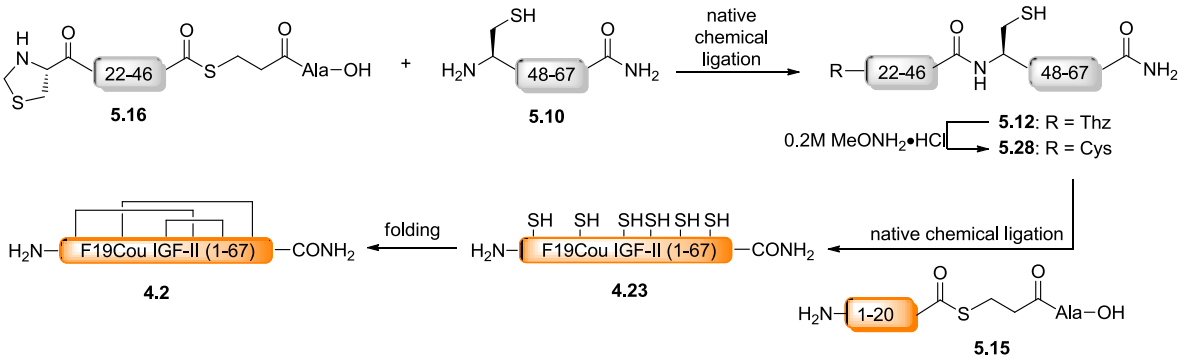


Figure 66: Characterisation of the synthetic native IGF-II protein (4.1) synthesised using the three fragment approach. a) RP-HPLC trace of protein **4.1**; b) HRMS of protein **4.1**. Calcd. for $C_{321}H_{500}N_{94}O_{100}S_6$: 7468.3875 (average isotopes); observed: 1494.7144 ($[M+5H]^{+5}$), 1245.7516 ($[M+6H]^{+6}$) and 1067.9352 ($[M+7H]^{+7}$).

Successful synthesis of the native IGF-II protein (**4.1**) demonstrated the suitability of a three fragment, one-pot native chemical ligation approach for the synthesis of IGF-II analogues. The methodology was employed for the synthesis of the F19Cou IGF-II protein (**4.2**) as it was anticipated a native chemical ligation approach would be more efficient and higher yielding than the previous methods detailed in Chapters 3 and 4. Similarly, this approach was expected to afford the F28Cou IGF-II protein (**5.1**) which was unattainable using the recombinant protein expression (refer to Chapter 3) and linear Fmoc-SPPS (refer to Chapter 4) approaches.

5.2.3.4 Synthesis of the F19Cou IGF-II protein (4.2)

The F19Cou IGF-II protein (**4.2**) was synthesised in the same manner descried for the native IGF-II protein (**4.1**). Specifically peptide **4.23** was assembled as outlined in Scheme 24, from three fragments (**5.15**, **5.16** and **5.10**) using iterative cysteine and valine ligations.



Scheme 24: Synthesis of the F19Cou protein (4.2), using the three fragment approach. Native chemical ligation conditions: 6 M GnHCl, 200 mM Na_2HPO_4, 20 mM TCEP, 200 mM MPAA, pH of 6.9. Folding conditions: 2.5 M urea, 0.7 M Tris, 12.5 mM glycine, 2 mM EDTA, 0.5 mM DTT, 1.25 mM 2-HED at a pH of 9.1 and protein concentration of <0.1 mg/mL.

First the IGF-II (Thz-46) thioester (**5.16**) and C-terminal fragment **5.10** were ligated using standard native chemical ligation conditions (6 M GnHCl, 200 mM Na_2HPO_4, 20 mM TCEP, 200 mM MPAA, pH = 6.9) as depicted in Figure 67. LCMS analysis of the cysteine-based ligation is shown in Figure 67, and shows the presence of the desired IGF-II (Thz-67) fragment (**5.12**) after only 4 min. The ligation was complete after 120 min, as determined by the absence of thioester **5.16** from the LCMS.

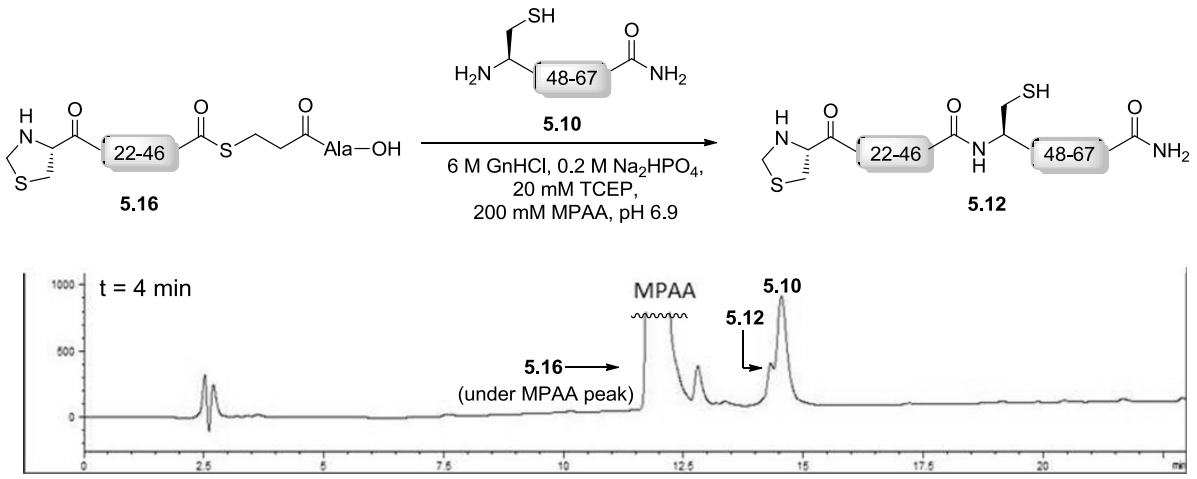


Figure 67: LCMS analysis of the cysteine-based ligation between the IGF-II (Thz-47) thioester (5.16) and C-terminal fragment (5.10) after 4 min.

The thiazolidone of the resultant ligation product **5.12** was converted to a free N-terminal cysteine (**5.28**) using methoxyamine hydrochloride (0.2 M) as illustrated in Figure 68. Fragment **5.28** was observed in the LCMS shown in Figure 68A after 3 min. The deprotection was determined quantitatively complete after 16 h, by the absence of fragment **5.12** from the LCMS shown in Figure 68B. ESI-MS analysis of resultant product **5.28**, is shown in Figure 68C and displays *m/z* ions ($[M+6H]^{+6}$; 883.20 *m/z*) are in agreement with the calculated *m/z* ions ($[M+6H]^{+6}$; 883.33 *m/z*) for the IGF-II (21-67) fragment (**5.28**).

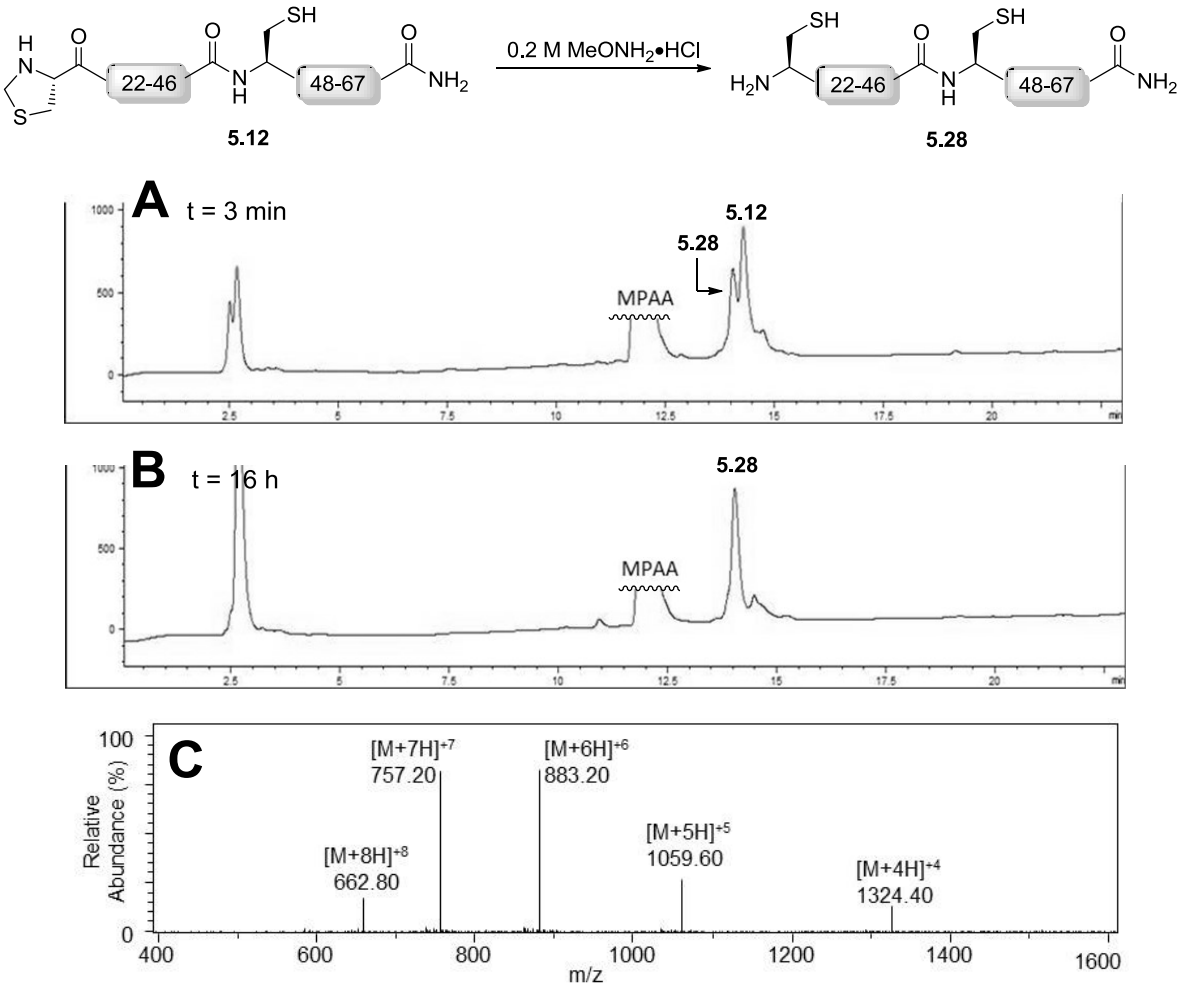


Figure 68: Analysis of the thiazolidone deprotection from fragment 5.12. a) LCMS analysis of the crude thiazolidone deprotection after 3 min; b) LCMS analysis of crude thiazolidone deprotection after 16 h; c) ESI-MS of fragment **5.28**. Calcd. for $C_{224}H_{356}N_{70}O_{69}S_5$: 5293.9931 (average isotopes); observed: 1324.40 ($[M+4H]^{+4}$), 1059.60 ($[M+5H]^{+5}$), 883.20 ($[M+6H]^{+6}$), 757.20 ($[M+7H]^{+7}$) and 662.80 ($[M+8H]^{+8}$).

A final ligation between the N-terminal F19Cou IGF-II (1-20) thioester (**5.15**) and the IGF-II (21-67) fragment (**5.28**) was used to furnish the desired F19Cou IGF-II peptide (**4.23**) and is depicted in Figure 69. Initial analysis of the valine-based ligation is displayed in Figure 69A and shows unreacted thioester **5.15** and fragment **5.28**. The ligation was determined to be complete after 27 h, by the absence of fragment **5.28** in the LCMS shown in Figure 69B. The F19Cou IGF-II peptide (**4.23**) was confirmed as the major product of the valine-based ligation, by the agreement of the *m/z* ions ($[M+7H]^{+7}$; 1082.56 *m/z*) displayed in Figure 69C with the expected *m/z* ions ($[M+7H]^{+7}$; 1082.78 *m/z*) for peptide **4.23**. In addition to peptide

4.23, a minor amount of the unreacted N-terminal thioester **5.15**, MPAA exchanged thioester **5.31** and mixed ligation by-product **5.32**, were also observed in the LCMS shown in Figure 69B.

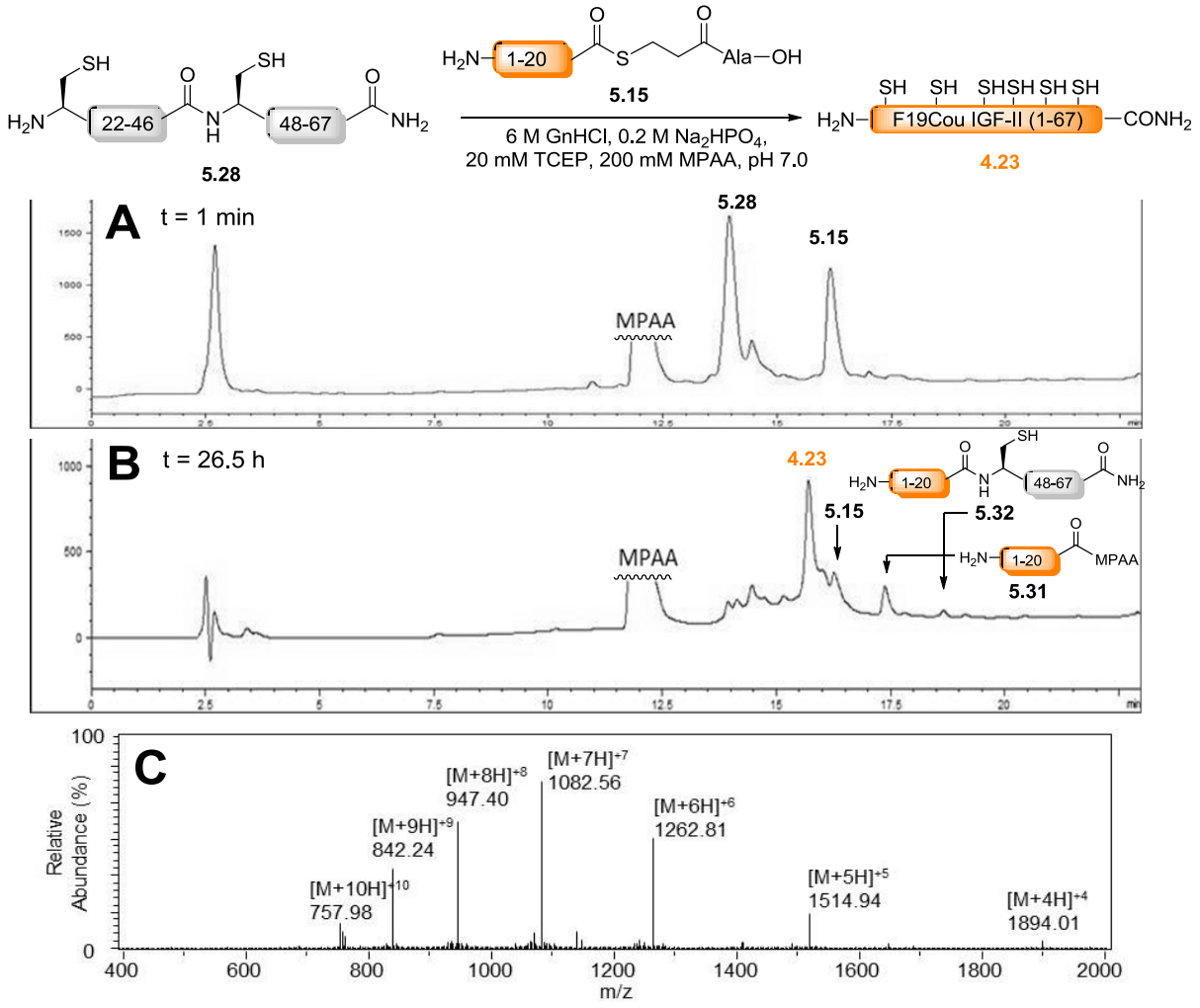


Figure 69: Analysis of the valine-based ligation between the IGF-II (21-67) fragment (5.27) and N-terminal F19Cou IGF-II (1-20) thioester (5.15). a) LCMS analysis of the crude ligation mixture after 1 min; b) LCMS analysis of the crude ligation mixture after 26.5 h; c) ESI-MS of the F19Cou IGF-II peptide (**4.23**). Calcd. for $C_{325}H_{508}N_{94}O_{103}S_6$: 7572.4921 (average isotopes); observed: 1894.01 ($[M+4H]^{+4}$), 1514.94 ($[M+5H]^{+5}$), 1262.81 ($[M+6H]^{+6}$), 1082.56 ($[M+7H]^{+7}$), 947.40 ($[M+8H]^{+8}$), 842.24 ($[M+9H]^{+9}$) and 757.98 ($[M+10H]^{+10}$).

The F19Cou IGF-II peptide (**4.23**) was isolated from the crude ligation mixture by SPE and the resultant peptide mixture was purified by semi-preparative RP-HPLC to give peptide **4.23** in a good yield (9%). Peptide **4.23** was folded following the methods described by Delaine *et al.*[23] to give the desired F19Cou IGF-II protein (**4.2**) in a respectable yield (11%). The F19Cou IGF-II protein (**4.2**) was isolated in high purity as shown in the RP-HPLC displayed in Figure 70A. Successful synthesis of protein **4.2** was confirmed by the agreement of the *m/z* ions ($[M+7H]^{+7}$; 1081.82 *m/z*) displayed in the HRMS shown in Figure 70B with the calculated *m/z* ions ($[M+7H]^{+7}$; 1081.92 *m/z*) for the F19Cou IGF-II protein (**4.2**).

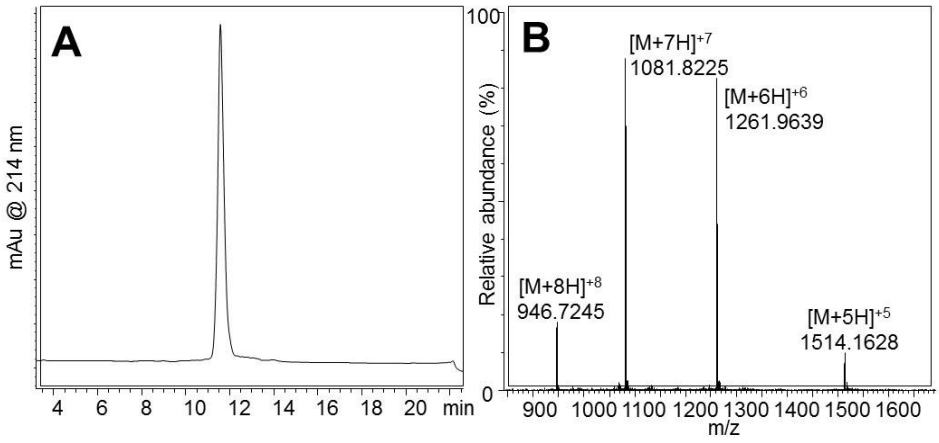
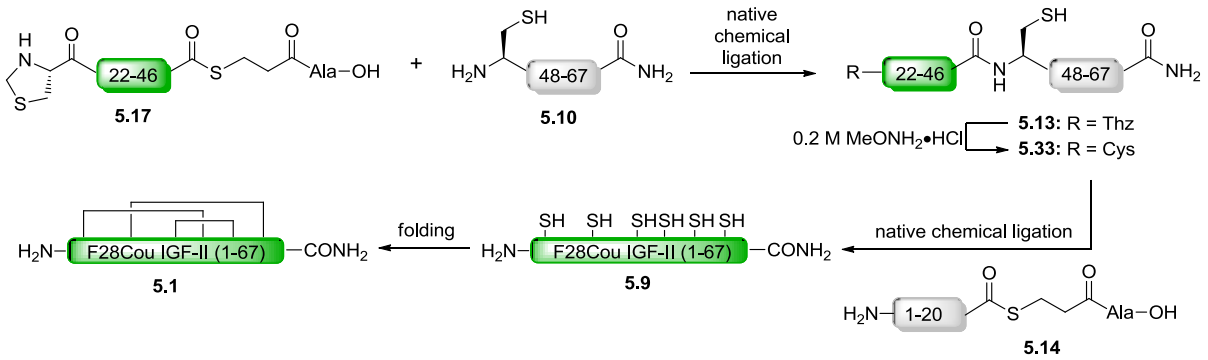


Figure 70: Characterisation of the synthetic F19Cou IGF-II protein (4.2). a) RP-HPLC trace of **4.2**; b) HRMS of **4.2**. Calcd. for $C_{325}H_{500}N_{94}O_{103}S_6$: 7566.4445 (average isotopes); observed: 1514.1628 ($[M+5H]^{+5}$), 1261.9639 ($[M+6H]^{+6}$), 1081.8225 ($[M+7H]^{+7}$) and 946.7245 ($[M+8H]^{+8}$).

5.2.3.5 Synthesis of the F28Cou IGF-II protein (5.1)

Assembly of the F28Cou IGF-II protein (**5.1**) using native chemical ligation was accomplished in a similar way to that described for both the native IGF-II protein (**4.1**) and the F19Cou IGF-II protein (**4.2**) and is summarised in Scheme 25.



Scheme 25: Synthesis of the F28Cou protein (5.1) using the three fragment approach. Native chemical ligation conditions: 6 M GnHCl, 200 mM Na_2HPO_4, 20 mM TCEP, 200 mM MPAA, pH of 6.9. Folding conditions: ; 2.5 M urea, 0.7 M Tris, 12.5 mM glycine, 2 mM EDTA, 0.5 mM DTT, 1.25 mM 2-HED at a pH of 9.1 and protein concentration of <0.1 mg/mL.

First, fragment **5.13** was assembled from a cysteine-based ligation of C-terminal fragment **5.10** with the F28Cou IGF-II (Thz-46) thioester (**5.17**), under standard native chemical ligation conditions (6M GnHCl, 200 mM Na_2HPO_4, 20 mM TCEP, 200 mM MPAA, pH of 6.9), as illustrated in Figure 71. The ligation was determined to be complete after 120 min by the absence of thioester **5.17** in the LCMS.

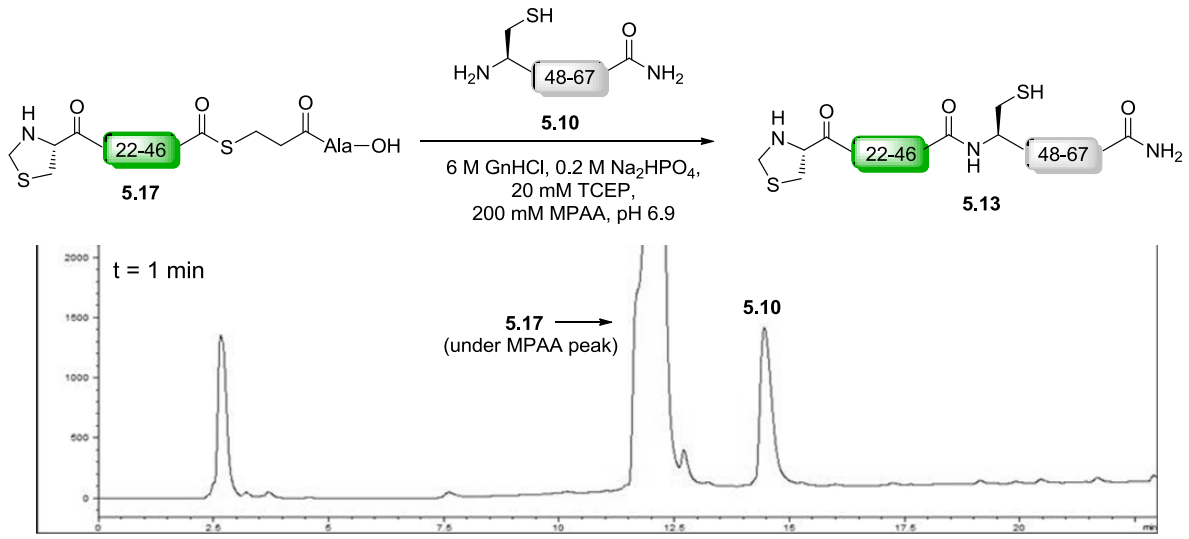


Figure 71: LCMS analysis of the cysteine-based ligation between the F28Cou IGF-II (Thz-47) thioester (5.17) and C-terminal thioester (5.10) after 1 min.

The N-terminal thiazolidone of the resultant fragment (**5.13**) was deprotected using methoxyamine hydrochloride (0.2 M) to furnish the F28Cou IGF-II (21-67) fragment (**5.33**). The deprotection was determined to be complete after 12 h by the absence of **5.13** in the LCMS shown in Figure 72. ESI-MS analysis of fragment **5.33** is shown in Figure 72 and displays *m/z* ions ($[M+5H]^{+5}$; 899.46 *m/z*) which are consistent with the calculated *m/z* ions ($[M+5H]^{+5}$; 899.68 *m/z*) for the F28Cou IGF-II (21-67) fragment (**5.33**).

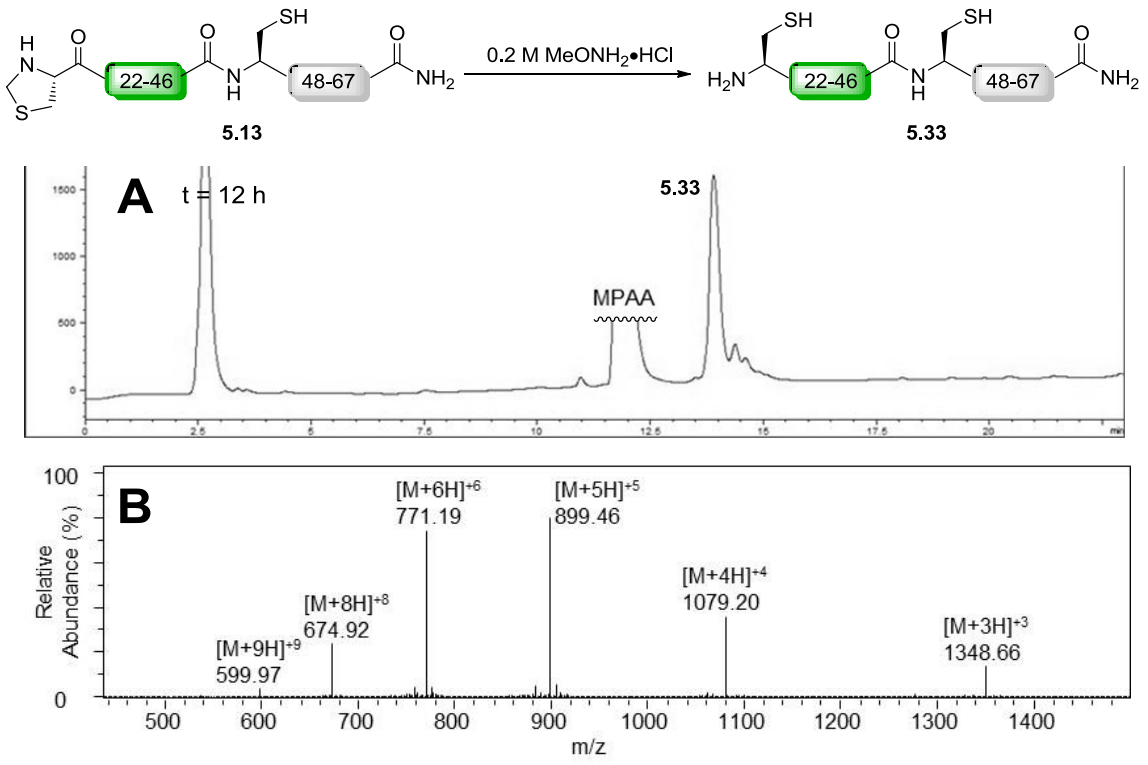


Figure 72: Analysis of the thiazolidone deprotection of the F28Cou IGF-II (Thz-67) fragment (5.13). a) LCMS analysis of the crude thiazolidone deprotection after 12 h; b) ESI-MS of fragment **5.33**. Calcd. for $C_{228}H_{358}N_{70}O_{72}S_5$: 5392.0502 (average isotopes); observed: 1348.66 ($[M+4H]^{+4}$), 1079.20 ($[M+5H]^{+5}$), 899.46 ($[M+6H]^{+6}$), 771.19 ($[M+7H]^{+7}$), 674.92 ($[M+8H]^{+8}$) and 599.97 ($[M+9H]^{+9}$).

A second, valine-based ligation between the resultant F28Cou IGF-II (21-67) fragment (**5.33**) and the IGF-II (1-20) thioester (**5.14**) was carried out as depicted in Figure 73. Initial LCMS analysis of the valine-based ligation is displayed in Figure 73A and shows unreacted thioester **5.14** and fragment **5.33** as expected. The full length F28Cou IGF-II peptide (**5.9**) was detected after 22 h by LCMS as shown in Figure 73B, however thioester **5.14** and MPAA exchanged thioester **5.29** also still remained. The ligation was determined complete after 36 h, by the disappearance of thioesters **5.14** and **5.29** from the LCMS shown in Figure 73C. Peptide **5.9** was confirmed as the major product of the valine-based ligation by the agreement of *m/z* ions ($[M+7H]^{+7}$; 1082.58 *m/z*) displayed in Figure 73D with the expected *m/z* ions ($[M+7H]^{+7}$; 1082.78 *m/z*) for the F28Cou IGF-II peptide (**5.9**). In addition to peptide **5.9** a minor amount of the mixed ligation product **5.30** and unreacted fragment **5.33** were also observed in the LCMS shown in Figure 73C.

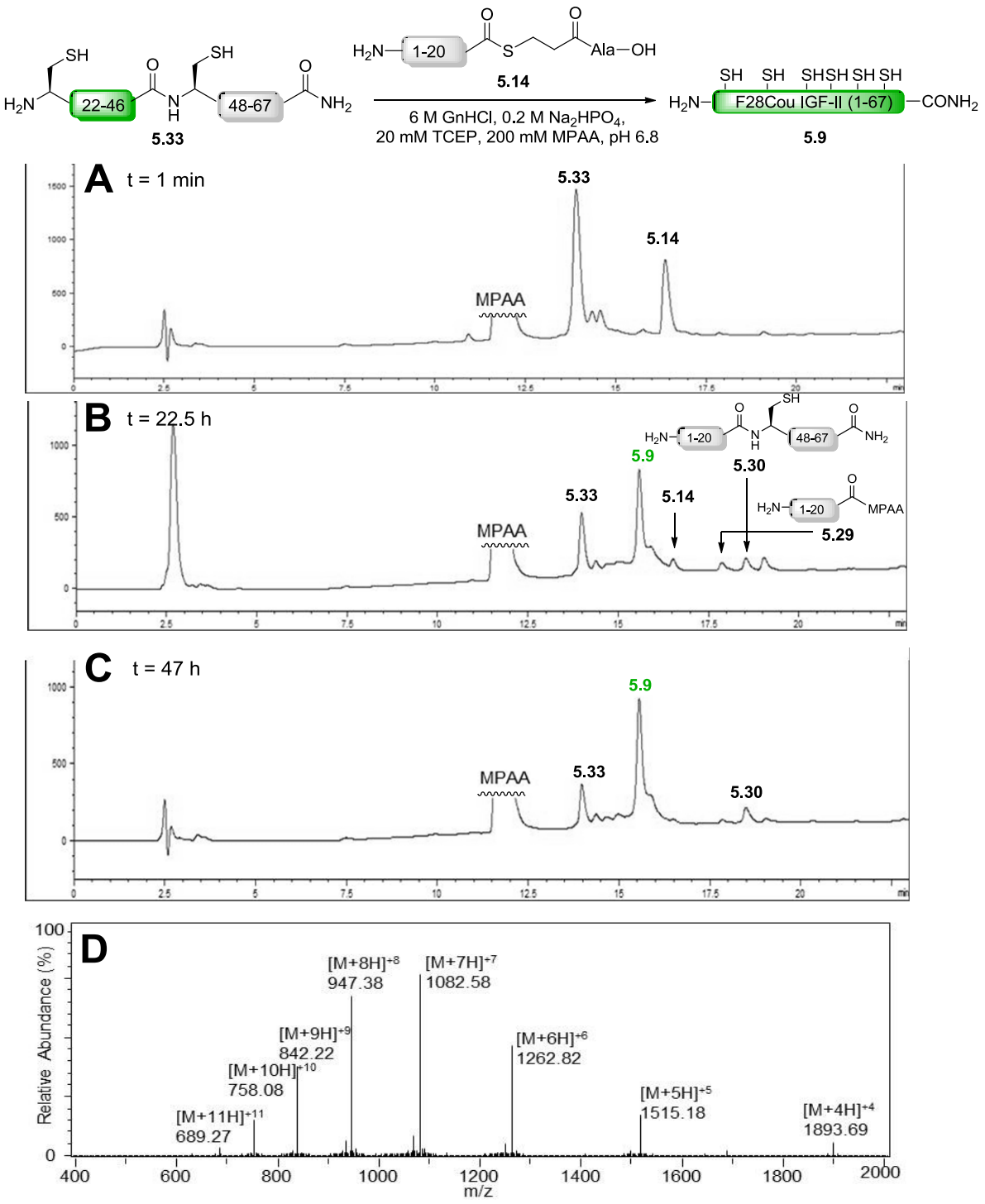


Figure 73: Analysis of the valine-based ligation between the F28Cou IGF-II (21-67) fragment (5.33) and N-terminal IGF-II (1-20) thioester (5.14). a) LCMS analysis of the crude ligation mixture after 1 min; b) LCMS analysis of the crude ligation mixture after 22.5 h; c) LCMS analysis of the crude ligation mixture after 47 h; d) ESI-MS of the F28Cou IGF-II peptide (**5.9**). Calcd. for $C_{325}H_{508}N_{94}O_{103}S_6$: 7572.4921 (average isotopes); observed: 1893.69 ($[M+4H]^{+4}$), 1515.18 ($[M+5H]^{+5}$), 1262.82 ($[M+6H]^{+6}$), 1082.58 ($[M+7H]^{+7}$), 947.38 ($[M+8H]^{+8}$), 842.22 ($[M+9H]^{+9}$), 758.08 ($[M+10H]^{+10}$) and 689.27 ($[M+11H]^{+11}$).

As discussed in Section 5.1.1, β-branched amino acids, such as Val have poor reactivity in the ligation reaction.[219] Typically, ligations performed at a C-terminal valine thioester, are slow and often do not go to completion even after 48 h.[219] Synthesis of the native IGF-II (**4.1**), F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins using the three fragment ligation approach required the use of C-terminal valine thioesters (**5.14** and **5.15**). Importantly, these valine-based ligations proceeded quantitatively to completion is less than 47 h.

The F28Cou IGF-II peptide (**5.9**) was isolated from the crude ligation mixture by SPE and the resultant peptide mixture was purified by semi-preparative RP-HPLC to give peptide **5.9** in a reasonable yield (3%). Peptide **5.9** was folded according to the protocol described by Delaine *et al.*[23] to give the desired F28Cou IGF-II protein (**5.1**). The F28Cou IGF-II protein (**5.1**) was isolated in a moderate yield (4%) and high purity as shown in the RP-HPLC displayed in Figure 74A. Successful synthesis of protein **5.1** was confirmed by the agreement of the *m/z* ions ($[M+6H]^{+6}$; 1261.88 *m/z*) displayed in HRMS shown in Figure 74B, with the expected ions *m/z* ($[M+6H]^{+6}$; 1261.96 *m/z*) for the F28Cou IGF-II protein (**5.1**).

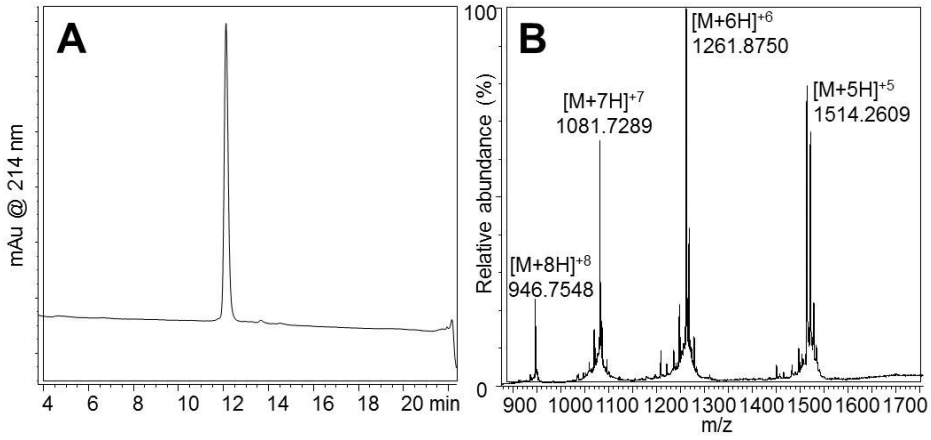


Figure 74: Characterisation of the synthetic F28Cou IGF-II protein (5.1). a) Analytical RP-HPLC trace of **5.1**; b) HRMS of **5.1**. Calcd. for $C_{325}H_{500}N_{94}O_{103}S_6$: 7566.4445 (average isotopes); observed: 1514.2609 ($[M+5H]^{+5}$), 1261.8750 ($[M+6H]^{+6}$), 1081.7289 ($[M+7H]^{+7}$) and 946.7548 ($[M+8H]^{+8}$).

5.2.4 Biological activity of the synthetic IGF-II proteins 4.1, 4.2 and 5.1

The affinities of the synthetic IGF-II analogues (**4.1**, **4.2** and **5.1**) for the IGF-1R were evaluated by competition binding assays. Receptor binding assays were carried out on immunocaptured IGF-1R, as described by Denley *et al.*[46] Briefly, increasing concentrations of the desired protein (**4.1**, **4.2** and **5.1**) were added to a constant concentration of Europium labelled IGF-II (EuIGF-II). Binding curves for the synthetic IGF-II proteins (**4.1**, **4.2** and **5.1**) binding to the IGF-1R are illustrated in Figure 75 and IC_{50} values are summarised in Table 14. The IC_{50} values for the binding of the synthetic native IGF-II (**4.1** (2 fragment and 3 fragment)), the F19Cou IGF-II (**4.2**) and the F28Cou IGF-II (**5.1**) proteins to the IGF-1R were determined to be 2.1 ± 1.6 nM, 2.0 ± 1.2 nM, 7.0 ± 1.3 nM and 6.5 ± 1.5 nM, respectively (refer to Table 14 and Figure 75).

Table 14: IC_{50} values derived from competitive binding assays of the synthetic native IGF-II (4.1), F19Cou IGF-II (4.2) and F28Cou IGF-II (5.1) proteins to immunocaptured IGF-1R. Where the affinity relative to IGF-II is the IC_{50} relative to that of IGF-II binding to the IGF-1R (IC_{50} IGF-II/ IC_{50} IGF-II analogue) and is expressed as percentage of IGF-II binding. IGF-II ± S.E is derived from at least 3 separate experiments performed in triplicate.

Protein	IC_{50} (nM)	Affinity relative to recombinant native IGF-II (%)
recombinant native IGF-II	2.1 ± 1.2	100
4.1 (2 Fragment)[a]	2.1 ± 1.6	100
4.1 (3 Fragment)	2.0 ± 1.2	105
4.2	7.0 ± 1.3	30
5.1	6.5 ± 1.5	32

[a] results derived from a single experiment performed in triplicate.

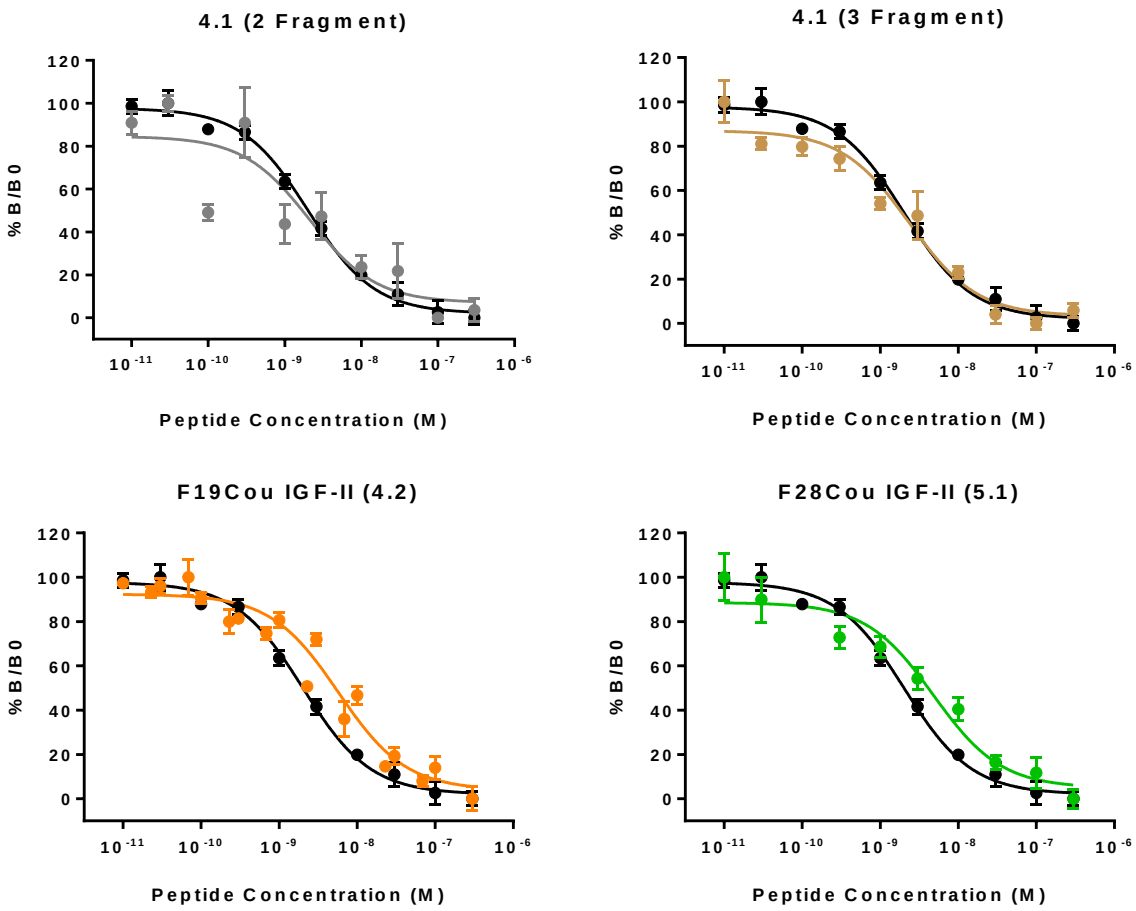


Figure 75: Competitive binding of the synthetic native IGF-II (4.1), F19Cou IGF-II (4.2) and F28Cou IGF-II (5.1) proteins to immunocaptured IGF-IR. Immunocaptured IGF-1R was incubated with Europium labelled IGF-II (EuIGF-II) in the presence or absence of increasing concentrations of recombinant native IGF-II (*black*), **4.1** (2 fragment: *grey*; 3 fragment: *brown*), **4.2** (*orange*) or **5.1** (*green*). Results are expressed as a percentage of binding in the absence of competing ligand (B_0). Graphs show data pooled from three separate experiments and each data point is measured in triplicate per experiment.[xviii] Data is shown as the mean ± S.E. Error bars are shown when greater than the size of the symbols.

Competitive binding assays showed the two synthetic native IGF-II analogues (**4.1**) had essentially the same IC_{50} values as the recombinant native IGF-II protein (2.0 nM versus 2.1 nM versus 2.1 nM) (refer to Table 14). In contrast, the F19Cou IGF-II protein (**4.2**) had only a 3.3-fold decrease in affinity for the IGF-1R (7.0 nM versus 2.1 nM), and the F28Cou protein

[xviii] Data from the binding of the native IGF-II protein (2 fragment) (**4.1**) was derived from a single experiment.

(**5.1**) had a 3.1-fold decrease in affinity for the IGF-1R compared to the recombinant native IGF-II protein (6.5 nM versus 2.1 nM).

The IC_{50} values reported here differ from those determined for the F19Cou IGF-II protein (**3.1** and **4.2**) synthesised by the recombinant protein expression (47.6 nM; refer to Section 3.2.8, Chapter 3) and linear Fmoc-SPPS (45.5 nM; refer to Section 4.2.7, Chapter 4) protocols. It is thought this increased binding affinity (7.0 nM) reflects the higher purity of the F19Cou IGF-II protein (**4.2**) obtained using the native chemical ligation. As well, batch-to-batch variations of the P6 IGF-1R and R$^-$IR-A cells, and 24-31 and 83-7 anti-antibodies (refer to Section 7.3.4.3, Chapter 7) could also be a minor contributing factor to the discrepancies between binding affinities (7.0 nM versus 47.6 nM versus 45.5 nM). The IC_{50} values determined for the F19Cou IGF-II protein (**4.2**) synthesised *via* native chemical ligation are consistent with binding affinities reported for other F19X IGF-II analogues.[23,44] These results further demonstrate the advantages of a native chemical ligation approach in the synthesis of the native IGF-II protein (**4.1**) and fluorescent IGF-II analogues (**4.2** and **5.1**).

The decreases in affinity of the F19Cou IGF-II protein (**4.2**) and F28Cou IGF-II protein (**5.1**) compared to native IGF-II protein are consistent with expectations, as both Phe^{19} and Phe^{28} are known to be involved in receptor binding (refer to Section 1.1.4, Chapter 1). Specifically, IGF-II analogues in which Phe^{19} was mutated to Ala, Ser, Tyr or Leu, exhibited between 1.4-2.8 fold lower affinities for the IGF-1R compared to the native IGF-II protein.[23,44] This decrease in affinity of the F19Cou IGF-II protein (**4.2**) is consistent with these F19X IGF-II analogues, and is attributed to the increase in side-chain bulk and polarity of the coumarin moiety. Likewise, an IGF-II analogue where Phe^{28} was mutated to Leu, this analogue exhibited greater than a 6-fold decrease in affinity for the IGF-1R compared to native IGF-II protein.[23,44] The affinities of the synthetic IGF-II proteins (**4.1**, **4.2** and **5.1**) are consistent with what has been reported for other recombinant IGF-II analogues, and reinforce the importance of Phe^{19} and Phe^{28} for high affinity binding of IGF-II to the IGF-1R.

5.3 Conclusions

The synthesis of two native IGF-II proteins (**4.1**) and the fluorescent IGF-II analogues, F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) has been accomplished using native chemical ligation. Protein **4.1** was synthesised using both a two and three fragment approach. The F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins were synthesised using a three fragment approach. These proteins (**4.1**, **4.2** and **5.1**) were isolated in reasonable yields (4-11%), which are comparable to those reported for the synthesis of the related IGF-I ligand.[214]

The three fragment approach utilised a Val^{20}-Cys^{21} ligation site, which required the use of a C-terminal valine thioester. Typically valine-based ligations are slow and often do not go to completion. Yet these valine-based ligations were quantitatively complete in less than 47 h.

The synthetic IGF-II proteins (**4.1**) produced by the two and three fragment approaches exhibited IC_{50} values of 2.1 ± 1.6 nM, 2.0 ± 1.2 nM respectively, which were the same as the IC_{50} values determined for the recombinant native IGF-II protein (2.1 ± 1.2 nM). The F19Cou IGF-II protein (**4.2**) and F28Cou IGF-II protein (**5.1**) bound the IGF-1R with IC_{50} values of 7.0 ± 1.3 nM and 6.5 ± 1.5 nM respectively, and had only a 3-fold decrease in affinity compared to the native IGF-II proteins. The decreases in affinities of the fluorescent IGF-II analogues, F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) were consistent with what had been reported for other F19X and F28X IGF-II analogues. These results reinforce the importance of Phe^{19} and Phe^{28} for the binding of IGF-II to the IGF-1R.

A convergent approach has yielded the native IGF-II protein (**4.1**) which was previously unattainable using a linear Fmoc-SPPS methodology (refer Chapter 4). The approach also provided an improved, higher yielding synthesis of the F19Cou IGF-II protein (**4.1**), while also providing a route to the novel F28Cou IGF-II analogue (**5.1**). The resultant synthetic proteins were of high purity and bound the IGF-1R with nanomolar affinity. Synthesis of these fluorescent F19Cou IGF-II (**4.1**) and F28Cou IGF-II (**5.1**) proteins allowed the proposed FRET-based investigation into the binding interactions of the IGF-II. Thus, the analysis of the

IGF-II analogues synthesised in this chapter (**4.1**, **4.2** and **5.1**), with a soluble form of IGF-1R using FRET is now described in Chapter 6.

Chapter 6

Investigating the binding of fluorescent IGF-II analogues to the IGF-1R using fluorescence resonance energy transfer

6.1 Introduction

As discussed in Chapter 1, Section 1.3 fluorescence resonance energy transfer (FRET) is a common technique used to investigate structural and conformational changes in proteins (intramolecular FRET) and protein-protein interactions (intermolecular FRET).[102,106,107] FRET is a fluorescence-based technique, and as such it is highly sensitive and only requires small quantities of the donor and acceptor fluorophore. The FRET-based investigation into the binding of the IGF-II proteins, **4.1**, **4.2** and **5.1** (synthesised in Chapter 5) to the IGF-1R is now described.

6.1.1 Principles of fluorescence resonance energy transfer (FRET)

FRET was first described by Förster in 1948.[226] The process of FRET is depicted in Figure 76 and the first step involves the irradiation of a donor at its excitation maximum (λ_{Ex}) (purple arrow; Figure 76). This causes excitation of the donor into a higher energy state. If an acceptor is within close proximity (10-100 Å) a non-radiative energy transfer from the donor to the acceptor will occur.[105,107-112] This in turn causes excitation of the acceptor molecule into a higher energy state, and subsequent relaxation results in an emission of light at the fluorescence emission maximum (λ_{Em}) of the acceptor (light blue arrow; Figure 76).[105,110,111] In the absence of the acceptor, or when the acceptor is not in close proximity (> 100 Å), only the fluorescence emission maximum (λ_{Em}) of the donor is observed (dark blue arrow; Figure 76).

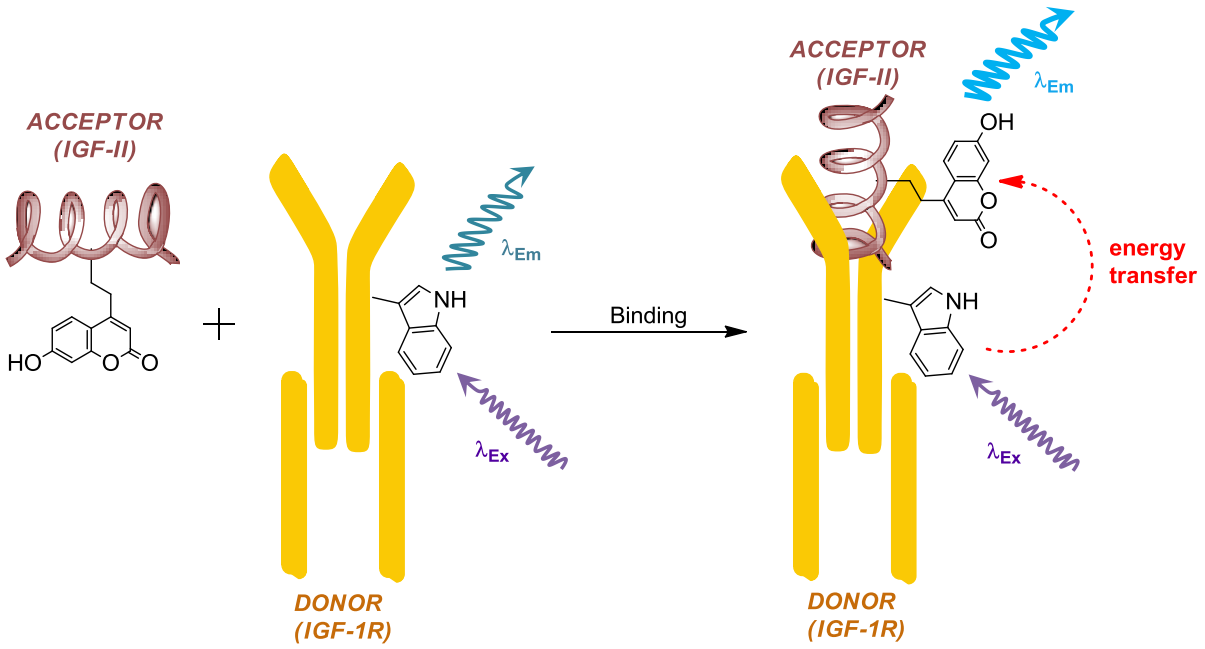


Figure 76: The process of FRET. The non-radiative energy transfer between the donor (Trp) and acceptor (Cou; **2.2**) occurs at distances between 10-100 Å.[105,107-112] Where; λ_{Ex} refers to the excitation maximum; λ_{Em} refers to the emission maximum; λ_{Ex} of the donor is depicted as a *purple zig-zag arrow*; λ_{Em} of the donor is depicted as a *dark blue zig-zag arrow* and λ_{Em} of the acceptor is depicted as a *light blue zig-zag arrow*.

The efficiency of the energy transfer and the distances over which a FRET interaction can occur, are determined by the spectral properties of the donor-acceptor pair (refer to Section 1.3, Chapter 1).[105,107-112,227,228] The efficiency of the energy transfer (E) can be expressed as a generalised 'quantum yield' for the FRET interaction (Eq. (1)), where:

$$E = \frac{\text{Energy transferred from donor to acceptor}}{\text{Energy absorbed by the donor}} \quad \textbf{(1)}$$

however E is more commonly defined as represented in Eq. (2):

$$E = \frac{1}{1 + \left({}^{r}/_{R_0}\right)^6} \quad \textbf{(2)}$$

In Eq (2), r refers to the distance between the donor and acceptor and R_0 is the Förster distance, which is the distance at which the FRET interaction is 50% efficient and varies for a given FRET pair.[105,107,108,110,228] For systems in which Trp is the donor, R_0 values between 12-40

Å have been reported.[110,111,123,228,229] In particular, Trp/Cou-based FRET pairs report R_0 values of between 20-30 Å, however the distance is dependent on the nature of the coumarin fluorophore.[110,111,228,230] Thus, for coumarin-based fluorophore **2.2**, the FRET efficiency was expected to be strongest between 10-30 Å, as depicted in Figure 77.

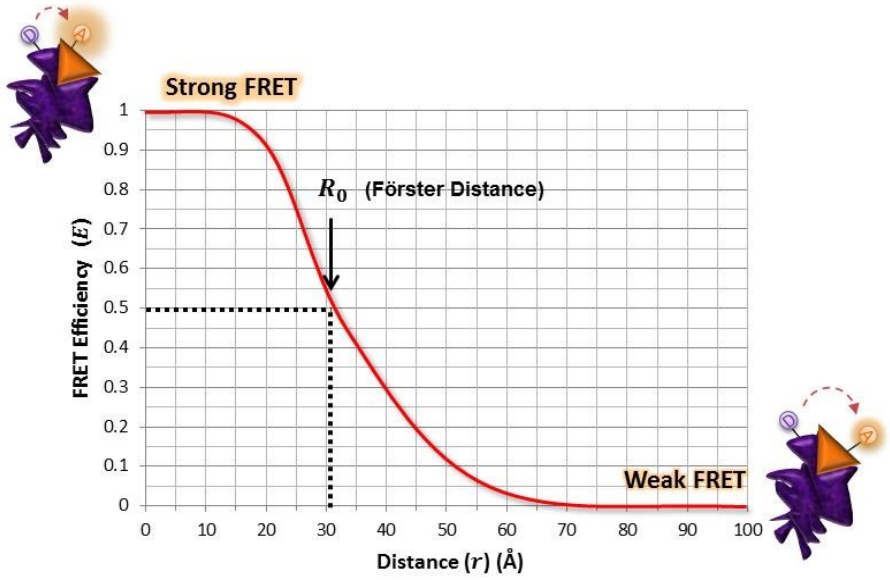


Figure 77: FRET efficiency as a function of distance. Where R_0 is the Förster distance, which is the distance at which the FRET interaction is 50% efficient. Figure adapted from Sahoo.[107]

As discussed above, the FRET efficiency is dependent on the distance between the donor and acceptor. Therefore FRET can be exploited as a 'spectroscopic ruler' for accurately determining the distance between a donor and acceptor.[105,107,108,110,228,231] Thus, by using FRET in the current study, it was anticipated the distances between Trp and Cou (**2.2**) could be determined. This would in turn reveal information about the binding distance and orientation of IGF-II to the IGF-1R. Several methods exist for detecting and analysing FRET interactions of this nature, and are discussed below.

6.1.2 Methods of FRET detection.

FRET interactions cause several changes to the fluorescent properties of the donor and acceptor.[232] The distance between the donor and acceptor can be determined by measuring these changes. Some of changes include the decrease in the donor emission intensity and fluorescence lifetime (τ), increases in the acceptor emission intensity and changes to the emission anisotropy of the donor and acceptor.[107,232-234]

The two main methods used to quantify FRET interactions are time-resolved fluorescence and steady-state fluorescence.[107,110,115,123,228,232-234] The method used is determined by the spectroscopic properties of the FRET pair, the fluorescent properties being measured, and the equipment available.[107,110,232-234] The main drawback of time-resolved fluorescence is that it requires high-speed detection systems and a pulsed laser or frequency-modulated light source.[110,232] Consequently, steady-state fluorescence methods are often preferred due to the ease with which the measurements can be taken.

6.1.2.1 Steady-state fluorescence methods

Steady-state fluorescence techniques monitor the changes in fluorescence emission intensity of the donor and acceptor. There are two common ways in which these changes are measured, acceptor photobleaching and sensitised emission.[107,110,123,229,232,234] Acceptor photobleaching involves measuring the changes in the fluorescence emission intensity of the donor before, and after acceptor photobleaching. Photobleaching involves the photochemical destruction of the acceptor. This is achieved by exposing the acceptor to light until it no longer displays any fluorescence. The main disadvantage of this method is the acceptor is destroyed, which prevents any further analysis of the FRET interaction. In contrast, sensitised emission measures the changes in the acceptor fluorescence emission intensity in the presence of, and absence of, the donor. Specifically, an increase in the acceptor fluorescence emission correlates to a FRET interaction (refer to Figure 76, Section 6.1.1). Both these techniques are easily performed on a fluorescence spectrophotometer and do not require any specialised

equipment.[107,110] Due to the equipment limitations associated with time-resolved methods, and the photochemical destruction of the acceptor using photobleaching, a steady-state-fluorescence, sensitised emission approach was preferred. The FRET interactions between the IGF-II proteins (**4.1**, **4.2** and **5.1**) and the IGF-1R were monitored using a sensitised emission approach and the results from these experiments are discussed herein.

6.2 Results and discussion

As previously discussed, Trp and coumarin-based fluorophore **2.2** were selected as the donor and acceptor for the proposed FRET study. Trp was selected due to its high natural abundance in the IGF-1R, thus no manipulation of the IGF-1R was required (refer to Section 1.3, Chapter 1). The coumarin-based fluorophore **2.2**, was a selected as a suitable acceptor for use with tryptophan as it exhibits the desired spectral overlap. Furthermore **2.2** was also preferred over other fluorophores as it is small, and thus it was anticipated its incorporation into IGF-II would not significantly disrupt its binding to the IGF-1R (refer to Section 3.2.1; Chapter 3). Coumaryl amino acid **2.2** also displays a large molar extinction coefficient, fluorescence emission maximum (λ_{Em}) and Stokes shift, which are spectroscopic properties that are well-suited for use in FRET (refer to Section 2.2.3; Chapter 2).

6.2.1 Biological activity of the synthetic IGF-II analogues (4.1, 4.2 and 5.1)

The IC_{50} values of the IGF-II analogues, **4.1**, **4.2** and **5.1** binding to the immunocaptured IGF-1R were previously determined, as discussed in Section 5.2.4 (refer to Chapter 5). The F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins displayed IC_{50} values of 7.0 nM and 6.5 nM respectively. The incorporation of the coumarin-based fluorophore **2.2** into the IGF-II proteins **4.2** and **5.1** had only a minor impact on the binding affinity. Both analogues reported only a 3-fold decrease in binding affinity compared to recombinant native IGF-II protein (IC_{50} = 2.1 nM). Although these affinities were reduced compared to the recombinant native IGF-II protein (IC_{50} = 2.1 nM) and the synthetic native IGF-II proteins (**4.1**) (IC_{50} = 2.10 nM (2

Fragment); 2.0 nM (3 Fragment)), these analogues still retained a significant degree of affinity. Thus, the F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins are suitable for investigating the binding of IGF-II to the IGF-1R using FRET.

The binding affinities summarised above were determined on immunocaptured IGF-1R. The immunocaptured receptor is a whole receptor construct, which means it contains both the ecto- and intracellular domains of the IGF-1R (refer to Section 1.1.2, Chapter 1). However for the proposed FRET study, using immunocaptured IGF-IR was not feasible, as the antibodies and blocking agents used to isolate the receptor could potentially interfere with the FRET experiment.[235] Furthermore use of a whole IGF-1R construct was also impractical due to the difficulties associated with the isolation and purification of transmembrane receptors. Thus an alternative source of the IGF-1R was used.

6.2.2 Soluble form of the IGF-1R (sIGF-1R) (6.1)

A high-affinity, soluble form of the IGF-1R has been reported by Surinya *et al.*[63] This soluble receptor, the sIGF-1R (**6.1**), is an ectodomain structure of the IGF-1R. Thus, it differs from the full length wild-type IGF-1R by the absence of the intracellular domains (TM, JM, TK and CT), as depicted in Figure 78. Importantly, the absence of these intracellular domains does not significantly affect the binding properties of the receptor. The sIGF-1R (**6.1**) contains the known site 1 and site 2 binding determinants required for high-affinity ligand binding and exhibits the same binding characteristics as the full length receptor, including negative cooperativity. The sIGF-1R (**6.1**) has only a 1.8-fold decrease in affinity for the recombinant native IGF-II protein (IC_{50} = 0.88 nM) compared to the full length wild-type receptor (IC_{50} = 0.49 nM).[63] Thus, the sIGF-1R (**6.1**) was used in the FRET experiments described. Although an investigation into the binding of these fluorescent IGF-II analogues (**4.2** and **5.1**) to the IR-A was intended, limitations associated with access to a similar soluble form of the IR-A receptor prevented the proposed FRET-based investigation.

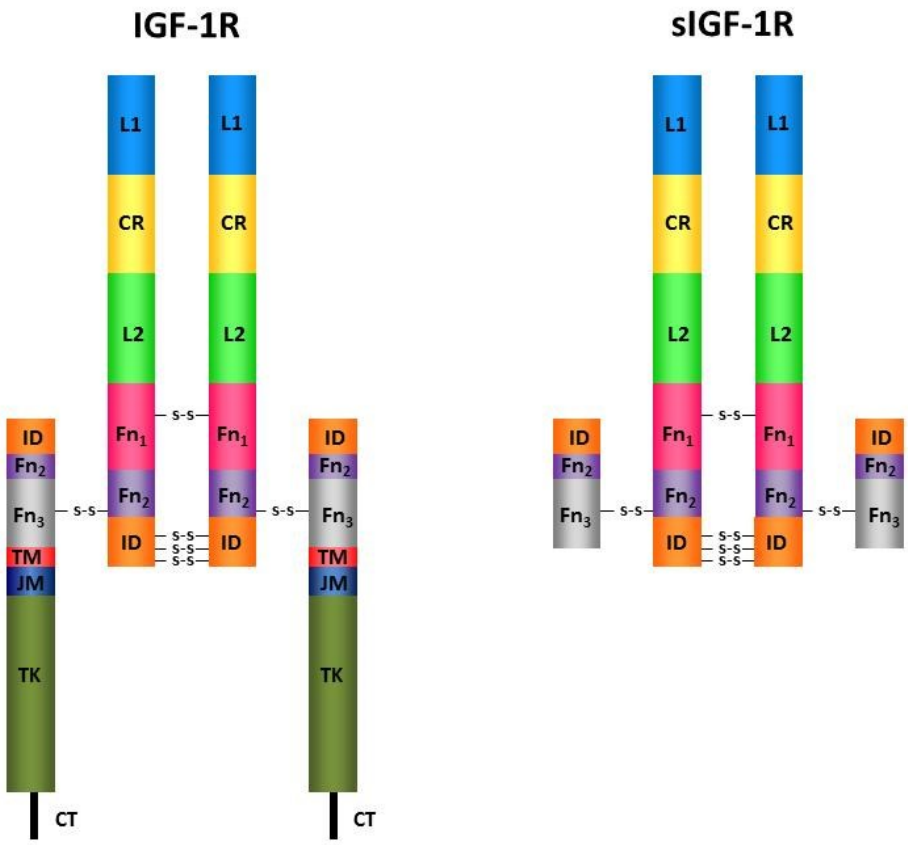


Figure 78: Comparison of the full length IGF-1R and a soluble form of the IGF-1R (sIGF-1R) (6.1). L1 and L2, large domains 1 and 2 (leucine-rich repeat domains); CR, Cys-rich domain; FnIII-1, FnIII-2, FnIII-3, fibronectin type III domains (1-3); ID, insert domain; TM, transmembrane; JM, juxtamembrane; TK, tyrosine kinase; CT, C-terminal domains and disulfide bonds are shown.

6.2.3 Fluorescence experiments

As described in Section 6.1.2.1, a sensitised emission approach was used to determine the location of the extracellular binding of IGF-II to the sIGF-1R (**6.1**). FRET experiments were performed on a Cary Eclipse Fluorescence Spectrophotometer at room temperature (25 °C), at a pH of 7, where the desired IGF-II analogue (**4.1**, **4.2** and **5.1**) was titrated against an aliquot of the sIGF-1R. The emission profiles of the IGF-II analogues (**4.1**, **4.2** and **5.1**) were monitored in the presence of, and absence of, the Trp donor (sIGF-1R) (**6.1**).

First the fluorescence emission maximum (λ_{Em}) of the sIGF-1R (**6.1**) was determined. The sIGF-1R (**6.1**) was excited at various wavelengths (250 – 300 nm) and the resulting fluorescence emission profiles are shown in Figure 79A. From this, the excitation maximum (λ_{Ex}) of the sIGF-1R (**6.1**) was determined to be 280 nm, which gave a fluorescence emission maximum (λ_{Em}) at 337 nm (refer to Figure 79A). These excitation (λ_{Ex}) and emission maxima (λ_{Em}) are consistent with the excitation (λ_{Ex} = 280 nm) and emission (λ_{Em} = 348 nm) maxima reported for Trp in aqueous solution at neutral pH (pH of 7).[110,236] In contrast, Tyr under the same conditions displays an excitation maximum (λ_{Ex}) of 280 nm and emission maximum (λ_{Em}) of 303 nm.[110,236] The small blue-shift in the emission maximum (λ_{Em}) of the sIGF-1R (**6.1**) compared to the reported value is thought to be the result of the sensitivity of Trp to its surrounding environment. For Trp residues located within proteins, as is the case for the sIGF-1R (**6.1**), the fluorescence emission maximum (λ_{Em}) for these Trp residues has been reported to shift from 320 nm to 355 nm depending on the surrounding microenvironment.[110,236]

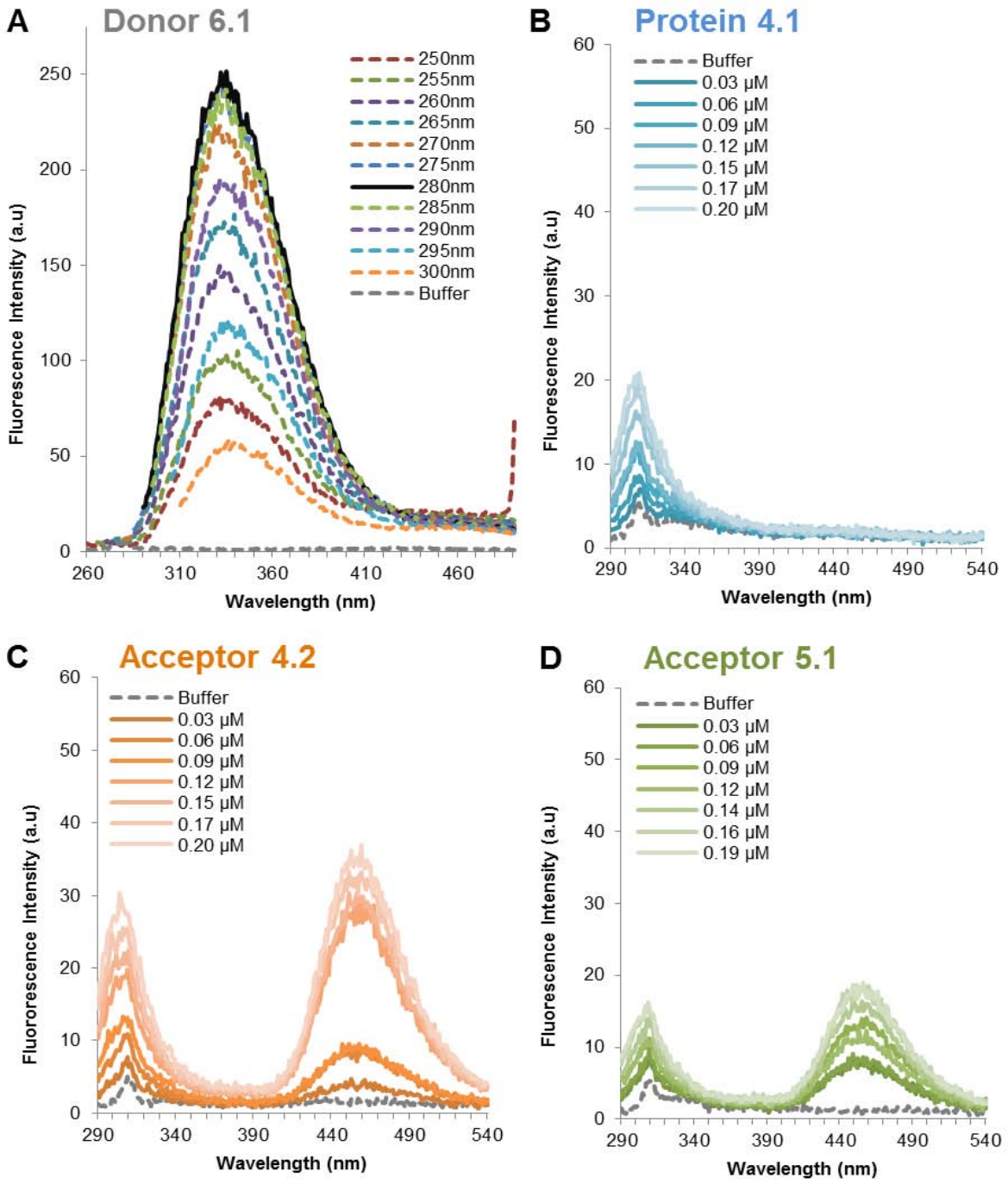


Figure 79: Fluorescence emission profiles for the sIGF-IR (6.1); synthetic native IGF-II protein (4.1); F19Cou IGF-II protein (4.2) and F28Cou IGF-II protein (5.1) after excitation at 280 nm. a) Emission profiles for the sIGF-1R (**6.1**) after excitation at varying wavelengths (250 - 300 nm). Where fluorescence emission spectra for irradiation at different wavelengths (250 - 300 nm) are shown as *coloured dotted lines* and the emission maxima after irradiation at 280 nm is shown as a *black solid line*; b) Emission profiles for the synthetic native IGF-II protein (**4.1**) at varying concentrations of **4.1** (0-0.20 µM) in buffer are shown as *blue lines*; c) Emission profiles for the F19Cou IGF-II protein (**4.2**) at varying concentrations of **4.2** (0-0.2 µM) in buffer are shown as *orange lines*; c) Emission profiles for the F28Cou IGF-II protein (**5.1**) at varying concentrations of **5.1** (0-0.19 µM) in buffer are shown as *green lines*. Where the buffer is shown as a *grey dotted line* in all spectra. All spectra were collected after excitation at 280 nm. Spectra are derived from a single experiment, where each spectrum is averaged from three consecutive scans and are corrected for background fluorescence.

As discussed in Chapter 2, Section 2.2.3 coumaryl amino acid **2.2** has a strong fluorescence emission maximum (λ_{Em}) at 485 nm. As such the F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins were expected to display similar fluorescence emission maxima (λ_{Em}). In contrast, the synthetic native IGF-II protein (**4.1**), which does not contain the coumarin-based fluorophore **2.2**, was not expected to display a fluorescence emission maximum (λ_{Em}).

The background fluorescence emission profiles for the three IGF-II analogues (**4.1**, **4.2** and **5.1**) in the absence of sIGF-1R (**6.1**) were determined. IGF-II analogues **4.1**, **4.2** and **5.1**, were titrated (0.03 μM aliquots) into buffer (pH of 7) until they reached a final concentration of 0.20 μM (**4.1** and **4.2**) or 0.19 μM (**5.1**). After the addition of each aliquot the samples were mixed, irradiated at 280 nm and the fluorescence emission spectra were recorded. The fluorescence emission profiles for IGF-II analogues **4.1**, **4.2** and **5.1** at varying concentrations are shown in Figure 79B, Figure 79C and Figure 79D respectively.

As expected the synthetic native IGF-II protein (**4.1**) did not display a fluorescence emission maximum (λ_{Em}) (refer to Figure 79B). All that was observed for protein **4.1** was background protein fluorescence (290 – 330 nm; Figure 79B).[110,116,122] In contrast, both the F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins displayed a fluorescence emission maximum (λ_{Em}) at 455 nm. This λ_{Em} is consistent with the fluorescence emission expected for coumarin-based fluorophore **2.2** (refer to Section 2.3.2, Chapter 2).[122,134,138]

FRET experiments were conducted by sequentially adding an IGF-II analogue (**4.1**, **4.2** and **5.1**) to an aliquot of the sIGF-1R (**6.1**) until the protein (**4.1**, **4.2** and **5.1**) and receptor (**6.1**) were present in an equimolar ratio. After the addition of each aliquot the sample was mixed and left to stand at rt for 1 h. After this time the sample was irradiated at 280 nm and the fluorescence emission spectrum was recorded. The changes in the fluorescence emission spectra with increasing amounts of the IGF-II analogues (**4.1**, **4.2** and **5.1**) are illustrated in Figure 80.

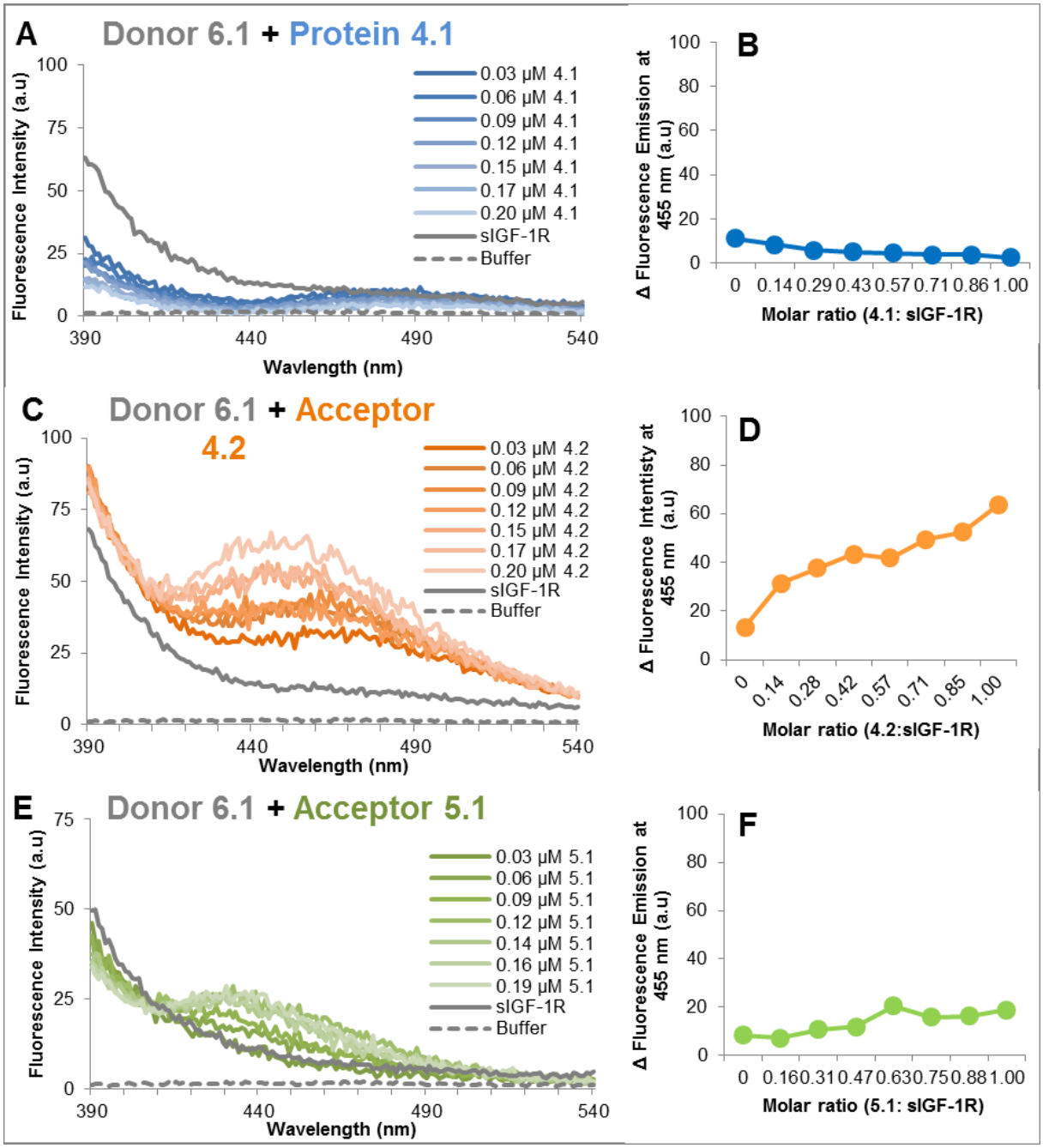


Figure 80: Changes in fluorescence emission spectra during the titration of IGF-II analogues (4.1, 4.2 and 5.1) against the sIGF-1R (6.1). a) The change in fluorescence emission spectra for the titration of the synthetic native IGF-II protein (**4.1**) against the sIGF-1R (**6.1**); b) The change in fluorescence intensity at 455 nm with increasing molar ratio of **4.1**; c) The change in fluorescence emission spectra for the titration of the F19Cou IGF-II protein (**4.2**) against the sIGF-1R (**6.1**); d) The change in fluorescence intensity at 455 nm with increasing molar ratio of **4.2**; e) The change in fluorescence emission spectra for the titration of the F28Cou IGF-II protein (**5.1**) against the sIGF-1R (**6.1**); f) The change in fluorescence intensity at 455 nm with increasing molar ratio of **5.1**. Where the sIGF-IR (**6.1**) in the absence of protein (acceptor) is shown as a *grey solid line* and buffer is shown as a *grey dotted line.* All spectra were collected after excitation at 280 nm. Spectra are derived from a single experiment, where each spectrum is averaged from three consecutive scans and are corrected for background fluorescence.

Figure 80A depicts the changes in the fluorescence emission when increasing amounts of the synthetic native IGF-II protein (**4.1**) was titrated against the sIGF-1R (**6.1**). As discussed above, the native IGF-II protein (**4.1**) does not contain coumaryl amino acid **2.2** and as such did not display any fluorescence emission at 455 nm (refer to Figure 79). This is further demonstrated in Figure 80B, where no change in fluorescence emission intensity at 455nm was observed with increasing molar ratio of the native IGF-II protein (**4.1**). As expected, these results indicate no FRET interaction occurred between synthetic native IGF-II protein (**4.1**) and the sIGF-1R (**6.1**).

The changes in the fluorescence emission spectra when the sIGF-1R (**6.1**) was titrated with increasing amounts of the F19Cou IGF-II protein (**4.2**) is depicted in Figure 80C. Protein **4.2** contains coumaryl amino acid **2.2** at position 19, which is adjacent to IGF-II binding site 2 (refer to Section 1.1.4, Chapter 1). A positive FRET interaction was expected if **2.2** came within close proximity (10-100 Å) to a Trp residue nearby to the IGF-1R binding site 2. Figure 80C shows the increase in the fluorescence emission at 455 nm after excitation at 280 nm. This change in fluorescence emission intensity at 455 nm with increasing molar ratio of the F19Cou IGF-II (**4.2**) is also depicted in Figure 80D. The increase in fluorescence intensity observed in both Figure 80C and Figure 80D is attributed to a positive FRET interaction between the site 2 tryptophan residues in sIGF-1R and the coumarin-based fluorophore (**2.2**) in the F19Cou IGF-II protein (**4.2**).

Figure 81 depicts a model for the binding of F19Cou IGF-II protein (**4.2**) to the IGF-1R and displays the Trp residues (Trp^{127}, Trp^{176} and Trp^{523}) which could be acting as potential donors.[43,60,68] Unfortunately, this model lacks the FnIII (2-3) domains, which are proposed to contain some of the site 2 binding determinants for the IGF-1R (refer to Section 1.1.3, Chapter 1). The F19Cou IGF-II protein (**4.1**) is a site 2 analogue, thus it is likely the Trp residues which are contributing most strongly to the FRET interactions observed in Figure 80C, will not be visible in the current model.

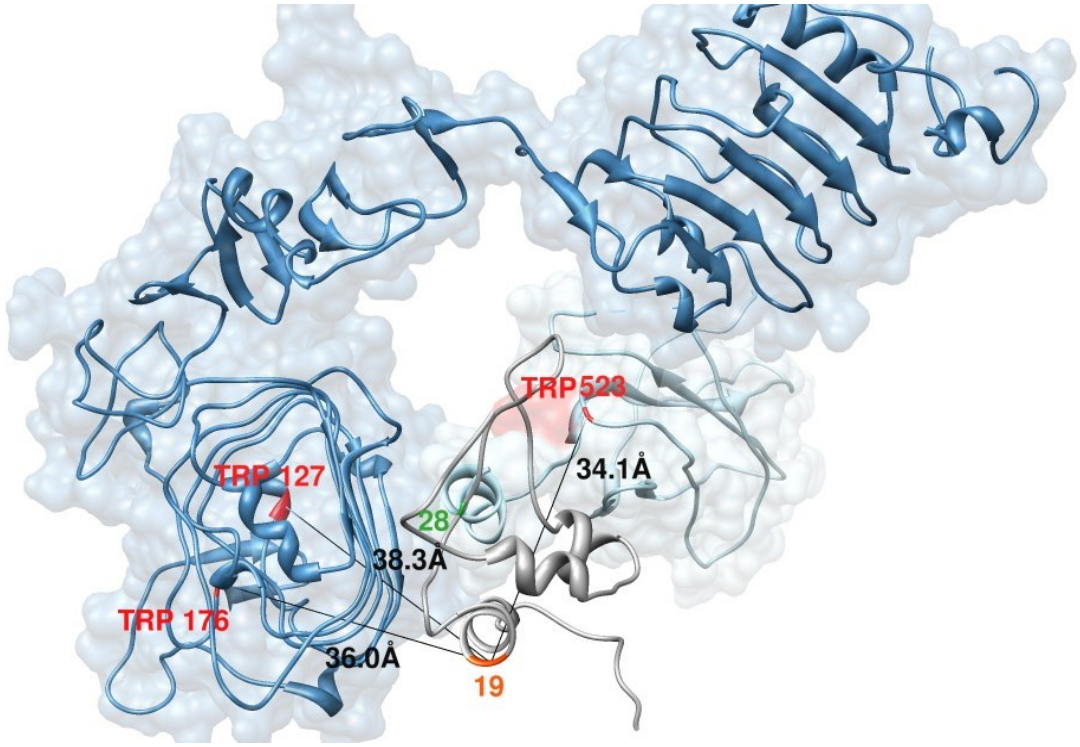


Figure 81: Model for the binding of the F19Cou IGF-II protein (4.2) to the IGF-1R. Model is based on the crystal structure reported by Menting *et al.*[68] for insulin binding to the IR-A (IR593.αCT construct) (PDB: 3W14). In this model the crystal structure for the L1-CR-L2 domains reported by Garrett *et al.*[60] is used for the IGF-1R (PDB: 1IGR) and the NMR solution structure reported by Torres *et al.*[43] is used for IGF-II (PDB: 1IGL). The IGF-1R is superimposed and replacing the IR-A (L1-CR-L2), IGF-II is superimposed and replacing insulin and the (FnIII-1)-α-CT (704-719) segment is unchanged. Where the IGF-1R is shown in *medium blue*; IGF-II is shown in *grey*; the (FnIII-1)-α-CT (704-719) segment is shown in *light blue*; coumarin-based fluorophore (**2.2**) is shown in *orange* (19) and *green* (28); Trp residues < 50 Å from residue 19 are shown in *red*; the distances (in Å) between Trp and **2.2** are shown for residues < 50 Å from residue 19 (IGF-II) and are designated by *black lines*.

Figure 80E depicts the changes in the fluorescence emission spectra when increasing molar equivalents of the F28Cou IGF-II protein (**5.1**) was titrated against the sIGF-1R (**6.1**). The F28Cou IGF-II protein (**5.1**) contains coumarin-based fluorophore **2.2** at position 28, which is adjacent to the IGF-II binding site 1 (refer to Section 1.1.4, Chapter 1). A positive FRET interaction was expected if **2.2** came within close proximity (10-100 Å) to a Trp residue adjacent to the IGF-1R binding site 1. After excitation at 280 nm, an increase in the fluorescent emission at 455 nm was observed and is shown in Figure 80E. The change in fluorescence emission intensity at 455 nm, with increasing molar ratio of F28Cou IGF-II (**5.1**), is also depicted in Figure 80F. This increase in fluorescence intensity at 455 nm, apparent in both Figure 80E and Figure 80F is attributed to a positive FRET interaction between the site 1

Trp residues in sIGF-1R (**6.1**) and coumarin-based fluorophore **2.2** in the F28Cou IGF-II protein (**5.1**).

A model for the binding of F28Cou IGF-II protein (**5.1**) to the IGF-1R (**6.1**) is displayed in Figure 82 and shows potential Trp donors.[43,60,68] As discussed in Section 1.1.3.2 (refer to Chapter 1) the site 1 analogue, the F28Cou IGF-II protein (**5.1**) is proposed to interact with the L1 and the αCT segment of the IGF-1R. Both these binding determinants are represented in the model shown in Figure 82. Based on the predicted R_0 values for the Trp/Cou pair (10-30 Å) (refer to Section 6.1.1), Trp residues (Trp^{76}, Trp^{127}, Trp^{176} and Trp^{523}) displayed in Figure 82 are proposed to be the residues involved in the FRET interaction with the F28Cou IGF-II protein (**5.1**).

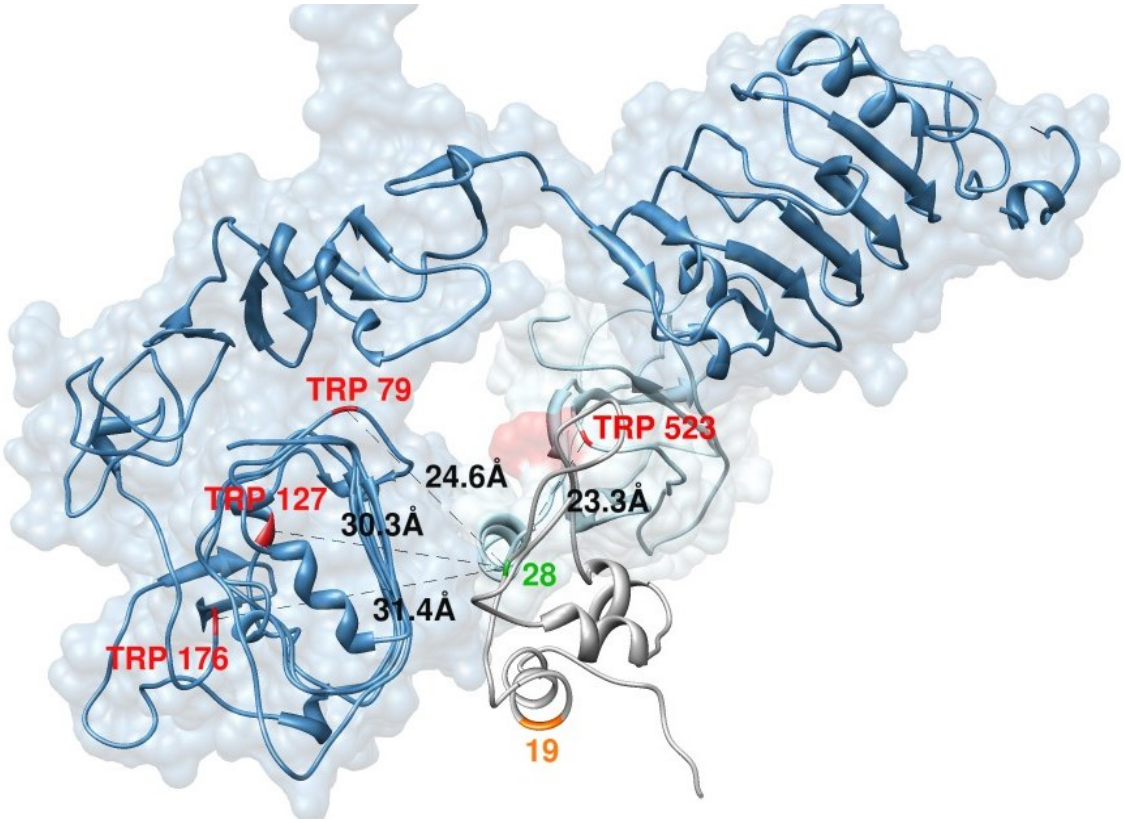


Figure 82: Model for the binding of the F28Cou IGF-II protein (5.1) to the IGF-1R. Model based on the crystal structure reported by Menting *et al.*[68] for insulin binding to the IR-A (IR593.αCT construct) (PDB: 3W14). In this model the crystal structure for the L1-CR-L2 domains reported by Garrett *et al.*[60] is used for the IGF-1R (PDB: 1IGR) and the NMR solution structure reported by Torres *et al.*[43] is used for IGF-II (PDB: 1IGL). The IGF-1R is superimposed and replacing the IR-A (L1-CR-L2), IGF-II is superimposed and replacing insulin and the (FnIII-1)-α-CT (704-719) segment is unchanged. Where the IGF-1R is shown in *medium blue*; IGF-II is shown in *grey*; the (FnIII-1)-α-CT (704-719) segment is shown in *light blue*; coumarin-based fluorophore (**2.2**) is shown in *orange* (19) and *green* (28); Trp residues < 50 Å from residue 28 are shown in *red*; the distances (in Å) between Trp and **2.2** are shown for residues < 50 Å from residue 28 (IGF-II) and are designated by *dashed black lines*.

Steady-state fluorescence measurements of the F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins displayed increases in fluorescence emission at 455 nm (refer to Figure 80). These increases in fluorescence are consistent with the presence of a FRET interaction between coumarin-based fluorophore **2.2** and the sIGF-1R (**6.1**). These results demonstrate that the fluorophore was positioned in close proximity to the IGF-II binding sites, thus establishing the F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins are suitable analogues for use in FRET-based investigations.

6.2.4 Future work

Determining the distance between the coumarin-based fluorophore **2.2** and its FRET partner would provide information about the binding location of IGF-II within the IGF-1R. As expected the high abundance of Trp residues in the IGF-1R receptor, prevented the identification of the exact FRET partner using the current experiments. Thus, further studies are required to identify which Trp residues in the receptor are contributing to the observed FRET interactions. Such experiments may require the synthesis or expression of mutant IGF-1R constructs, which is a significant undertaking.

One common approach is to use alanine mutagenesis. This would involve the mutation of all but a single Trp residue located within the IGF-1R putative binding sites, to alanine (L1, CR, L2, FnIII-2, FnIII-3 and αCT) (refer to Section 1.1.3.2, Chapter 1). This approach is beneficial as it gives site-specific control over the location of the donor Trp residue, which in turn would allow the desired distance measurement to be determined. However the mutation of such a large number of residues could result in perturbations to the three-dimensional structure and the binding properties of the receptor. Thus a better approach may involve the site-specific mutation of a single Trp residue, to a donor with contrasting spectroscopic properties. This approach would allow site-specific control over the donor location, which in turn would permit the distance between the donor and acceptor to be determined. Since this approach involves the mutation of only a single residue, it is less likely to cause large perturbations to the three-dimensional structure and binding of the receptor. Appropriate donors could include unnatural Trp analogues, such as 7-azatryptophan, 5-hydroxytrptophan or other halogenated or methylated tryptophan analogues.[164,237-241] These unnatural Trp analogues are sterically and chemically similar to Trp, and as such would minimise any perturbations to the three-dimensional structure of the receptor, while also providing a donor fluorophore with distinctive spectrochemical properties. Given the current advances of *in vivo*, site-specific incorporation of unnatural amino acids (refer to Chapter 3) this approach is likely to be rational and efficient. The Trp residues identified in Figure 81 and Figure 82 could be used a starting point for identifying receptor targets for site-specific mutations.

Many of the current methodologies for intermolecular protein distance determination are fluorescence based. Such methods are highly sensitive, thus requiring only small quantities of probe. However, complimentary techniques, such as X-ray crystallography and pulse electron paramagnetic resonance (EPR) spectroscopy could be explored in parallel to FRET for determining the binding location of IGF-II within the IGF-1R.

Obtaining an X-ray crystal structure of IGF-II bound to a receptor would provide an accurate depiction of the binding interactions. However acquiring a crystal structure is not straightforward and poses several challenges.[242-244] Large quantities of homogenously purified ligand and receptor are required. Predicting the crystallisation conditions, including the buffer, pH, and temperature, is immensely challenging and requires a lengthy screening process. Finally, even if crystallisation is achieved there is no guarantee a of high quality crystal structure. Given these difficulties, alternative structure determination techniques are preferred. Biophysical techniques such as FRET and EPR are suitable alternatives for studying these interactions.

Pulse EPR techniques, such as pulse electron-electron double resonance (known as DEER or PELDOR) and double quantum coherence (DQC) can provide accurate measurements of intermolecular distances of up to 80 Å.[245-249] These techniques would require the site-specific installation of spin labels into both IGF-II and the receptor. Site directed spin labelling of IGF-II could be easily achieved using the native chemical ligation approaches developed in this thesis. An appropriate spin label is 2,2,6,6-tetramethylpiperidine-1oxyl-4-amino-4-carboxylic acid (TOAC), as it is compatible with Fmoc-SPPS.[250] In contrast, spin labelling of the receptors could be achieved using the site directed mutagenesis approach discussed in Chapter 3, or through conjugation of the spin label post expression. While synthesis of the required labelled IGF-II and receptor is predicted to be somewhat challenging, EPR should still be a viable alternative to FRET and X-ray crystallography.

6.3 Conclusion

The functional IGF-II binding epitope on the IGF-1R, and the residues which define the IGF-II binding sites 1 and 2 have been identified. However the absence of a crystal structure of IGF-II bound to the IGF-1R or IR-A means the location and orientation of IGF-II within the receptor binding pocket remains elusive. In order to gain an increased understanding of IGF-II binding, experiments were conducted to establish FRET as a suitable method for determining the IGF-II binding location within the IGF-1R.

This work first required the synthesis of an appropriate fluorescent amino acid for incorporation into IGF-II. The coumarin-based probe **2.2** was selected as it could be synthesised in good yields and displayed a high molar extinction coefficient (ε), large Stokes shift and strong fluorescence emission intensity. Several methodologies were then explored for the incorporation of coumaryl amino acid **2.2** into the IGF-II protein to provide fluorescent IGF-II analogues. Recombinant protein expression and the linear SPPS methodologies provided access to small quantities of the F19Cou IGF-II protein (**4.2**). However, a convergent synthetic strategy using native chemical ligation was developed to provide the F19Cou IGF-II protein (**4.2**) in improved purity and with enhanced biological activity, compared to the F19Cou IGF-II protein (**4.2**) produced by the expression and linear SPPS approaches. The native chemical ligation strategy also provided access to the native IGF-II (**4.1**) and F28Cou IGF-II (**5.1**) proteins. In competition binding assays, these analogues displayed nanomolar binding affinities to both the IGF-1R and IR-A. These binding affinities were consistent with the affinities reported for other F19X and F28X IGF-II analogues and reinforce the importance of residues 19 and 28 for IGF-II binding.

Preliminary steady-state fluorescence measurements were conducted between the F19Cou IGF-II (**4.2**) and the F28Cou IGF-II (**5.1**) proteins to the sIGF-1R (**6.1**), in order to establish FRET as a viable method for determining the binding location and orientation of IGF-II within the IGF-1R. Both the F19Cou IGF-II (**4.2**) and the F28Cou IGF-II (**5.1**) proteins displayed an increase in fluorescence emission spectra at 455 nm in the presence of receptor **6.1**. This increase was attributed to the presence of a FRET interaction. These FRET

interactions confirm that the fluorophore is binding in close proximity to Trp residues within the receptor, and demonstrates that the fluorophore was incorporated at suitable sites in IGF-II for FRET experiments.

As expected the high abundance of Trp residues in the sIGF-1R prevented the elucidation of the specific FRET donors. Without the identification of the FRET partner, the distance between IGF-II and the receptor could not be determined. However based on a model of the F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins binding to the IGF-1R and the predicted R_0 for the Trp/Cou pair, several Trp residues could be identified as potential donors. These results provide the important basis for further investigations for determining the precise binding location of IGF-II within the IGF-1R. Such investigations are discussed in detail in Section 6.2.4 and may require the use of the mutant IGF-1R constructs, in which the precise location of the donor is known.

In conclusion, two novel IGF-II analogues, the F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins have been successfully synthesised. The site-specific incorporation of coumarin-based fluorophore **2.2** into these IGF-II proteins permitted the use these analogues in the desired FRET-based investigation into the binding of the IGF-II to the IGF-1R. In these FRET studies both the F19Cou IGF-II (**4.2**) and the F28Cou IGF-II (**5.1**) proteins displayed FRET interactions with the sIGF-1R (**6.1**). However as anticipated the position of the FRET partners could not be identified, which prevented the elucidation of the binding location of IGF-II within the IGF-1R. Despite this, these results demonstrate the suitability of a FRET-based approach and provide sound basis for further investigations into the binding interactions of IGF-II with the IGF-1R and IR-A.

Chapter 7

Experimental section

7.1 General experimental information

Nuclear magnetic resonance (NMR)

^{1}H and ^{13}C NMR spectra were determined on a Varian Gemini 2000 spectrophotometer (operating at 300 or 75 MHz) or Varian Inova 600 Spectrophotometer (operating at 600 or 150 MHz, with a delay (D1) of 1 s). All spectra were obtained at 23 °C and chemical shifts (δ) are reported in parts per million (ppm) and are referenced relative to a residual solvent peak ($CDCl_3$: δH (ppm) = 7.26 ppm, δC (ppm) = 77.0 ppm; D_2O: δH (ppm) = 4.79 ppm, δC: sr = -32.10 Hz; $DMSO_{-d6}$: δH (ppm) = 2.50 ppm, δC (ppm) = 39.5 ppm). Spin multiplicities are represented by the following signals: s (singlet), br s (broad singlet), d (doublet), dd (doublet of doublets), ddd (doublet of doublet of doublets), dt (doublet of triplets), t (triplet), td (triplet of doublets), q (quartet), qd (quartet of doublets), and m (multiplet).

Melting point

Melting points were determined on a Reichert Thermovar Kofler apparatus and were not corrected.

Specific rotation

Specific rotation values were determined on an Atago Automatic Polarimeter (AP-100 model), using 100 mm observation tube with 1 mL capacity.

Electrospray ionisation mass spectrometry (ESI-MS) analysis

Peptide masses were confirmed by MS analysis on a Thermo Finnegan LCQ mass spectrometer, using ESI in the positive mode or a Micromass Q-TOF 2 spectrometer using Nano-ESI in the positive mode. Accurate high resolution mass spectrometry (HRMS) data was collected on an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific) equipped with a Nano-ESI source in the positive mode.

Liquid chromatography mass spectrometry (LCMS)

LCMS analysis was carried out on an Agilent 1120 LC compact equipped with a variable wavelength detector and coupled to a Hewlett Packard 1100 MSD mass spectrometer, using

ESI in the positive mode. Gradient analysis was conducted on an Agilent Zorbax C3 (3.5 μm, 3 mm x 150 mm) column using a linear gradient of 5-65% D over 20 min with a flow rate of 0.3 mL/min, where the solvent system used was: A (MilliQ water + 0.1% formic acid) and D (acetonitrile + 0.1% formic acid). Elution of product was monitored at 220 nm and additional wavelengths (215 nm and 280 nm) were used as required.

Lyophilisation

The samples were lyophilised on a Christ Alpha 2-4 LSC freeze-dryer or a Christ Alpha 2-4 LSC freeze-dryer.

UV-Visible spectrophotometry

UV-Visible spectra were collected on a Cary 5000 UV-Vis-NIR Spectrophotometer using UV grade disposable plastic semi-micro cuvettes (range 190 nm - 900 nm) (Brand GMBH, Wertheim, Germany). Solvents used were of spectroscopic grade or higher. If spectroscopic grade was not available then solutions and solvents were filtered through a 0.22 μm filter before use.

Fluorescence

Fluorescence spectra were collected on a Cary Eclipse Fluorescence Spectrophotometer, using Spectrosil quartz micro cuvettes (Starna) (700 μL) with a 10 mm pathlength. The slit width was adjusted to of 5 nm. Solvents used were of spectroscopic grade or higher. If spectroscopic grade was not available then solutions and solvents were filtered through a 0.22 μm filter before use. Quinine sulfate was used as a standard for determining quantum yields.

7.2 Experimental described in Chapter 2

7.2.1 Experimental materials

Solvents and reagents were of reagent grade or higher and used as supplied unless otherwise stated. Solvents used for HPLC were of HPLC grade and used without further purification. THF was freshly distilled from sodium and benzophenone under an inert atmosphere of nitrogen. Anhydrous acetone was dried over $CaSO_4$ (25 g/L) for 48 h prior to use.[251] 5-(benzyloxy)-4-(((benzyloxy)carbonyl)amino)-5-oxobutanoic acid (Cbz-Asp-OBn), 5-(benzyloxy)-4-(((benzyloxy)carbonyl)amino)-5-oxopentanoic acid (Cbz-Glu-OBn) were purchased from ChemImpex (Illinois, U.S.A). Methanesulfonic acid, ethyl potassium malonate, 1,1-Carbonyldiimidazole (CDI), *N*-methylmorpholine (NMM), isobutyl chloroformate and silver benzoate were purchased from Sigma Aldrich (Castle Hill, NSW, Australia). Resorcinol was purchased from MP Biomedicals (Santa Ana, California, USA). Calcium sulfate, Diethyl ether, ethyl acetate, petroleum spirit (40-60 fraction) and magnesium chloride hexahydrate were purchased from ChemSupply (Gillman, SA, Australia). Isopropanol (AR grade) and acetonitrile (HPLC grade) were purchased from Scharlau Chemie (Sentmenat, Barcelona, Spain). 9-Fluorenylmethyl *N*-succinimidyl carbonate (Fmoc-OSu) was purchased from GL Biochem (Shanghai, China). Di-*tert*-butyl dicarbonate was purchased from Acros organics (Geel, Belgium). *N*-methyl-*N*-nitroso-p-toluenesulfonamide (Diazald) was purchased from WAKO chemicals (Richmond, VA, U.S.A). Diazomethane was prepared as described by Vogel[141] from *N*-methyl-*N*-nitroso-p-toluenesulfonamide (Diazald) as an ethereal solution immediately prior to use.

Oven dried glassware was used for reactions performed under an inert atmosphere (dry nitrogen). Analytical thin layer chromatography (TLC) analysis was conducted on Merck-aluminium plates coated with silica gel 60 F_{254}, with visualisation under ultraviolet light (254 nm), followed by staining with vanillin (6 g of vanillin in 95 mL of ethanol, then 2.5 mL conc. H_2SO_4), potassium permanganate (10 g of K_2CO_3, 1.5 g of $KMnO_4$, in 150 mL of water, then 1.5 mL of 10% aqueous NaOH) or ninhydrin (5 g of ninhydrin in 100 mL of ethanol). Flash column chromatography was carried out using Scharlau or Grace silica gel 60 (mesh size, 230-

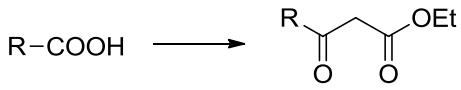
400 mesh). Where petroleum ether is stated, the fraction that is collected at 40-60 °C was used. Concentrated *in vacuo* refers to removal of solvent by rotary evaporation.

Analytical RP-HPLC was carried out on a HP 1100 series equipped with a diode array detector and an Agilent Eclipse C18 (5 µm, 4.6 x 150 mm) column. The product was eluted using a linear gradient of 0-70% D over 30 min and a flow rate of 1 mL/min, where the solvent system used was: A (MilliQ water + 0.1% TFA) and D (acetonitrile + 0.08% TFA). Elution was monitored at 220 nm, 280 nm and 320 nm, with the use of additional wavelengths in the range 190-400 nm as required.

Semi-preparative RP-HPLC was carried out on a HP 1100 series HPLC equipped with a diode array detector and a Luna C18 (C2) column (10 µm, 50 mm x 10 mm). The product was eluted using an appropriate linear gradient (1% D/min) with a flow rate of 5 mL/min, where the solvent system used was: A (MilliQ water + 0.01% TFA) and D (acetonitrile + 0.08% TFA). Gradient systems were adjusted according to the elution profiles and peak profiles obtained from the analytical RP-HPLC chromatograms. Fractions were analysed by ESI-MS and analytical RP-HPLC. Pure fractions were pooled, and lyophilised.

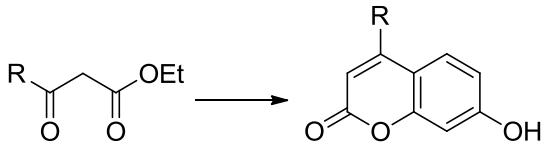
7.2.2 Methods

General Procedure I: Formation of β-keto esters via acylation reaction



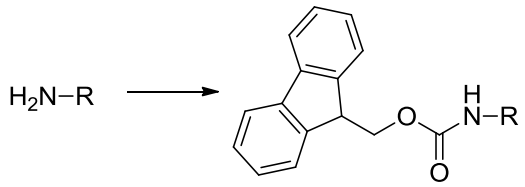
Under a nitrogen environment, the acid (1 equiv.) was dissolved in anhydrous THF (4 mL/mmol) and CDI (1.1 equiv.) was slowly added, the solution was then stirred at rt for 2 h. Ethyl magnesium malonate (0.55 equiv.) was added and reaction was stirred at rt for a further 16 h. The reaction mixture was extracted with diethyl ether (3 x 50 mL) and washed with saturated aqueous $NaHCO_3$ (3 x 50 mL), water (2 x 50 mL) and brine (3 x 50 mL), dried over Na_2SO_4 and concentrated *in vacuo*. The resulting oil was purified by silica column chromatography (1:1 PE:EtOAc).

General Procedure II: Pechmann condensation



The β-ketoester (1 equiv.) was slowly added to a stirring solution of resorcinol (5-10 equiv.) in methanesulfonic acid (25 equiv.) the solution was stirred vigorously at rt for 30-120 min. The reaction mixture was diluted with cold ether (20 volumes) and a precipitate was evolved. The mixture was then cooled in a dry ice/acetone bath for 30 min and the precipitate was filtered, dissolved in water, and the resulting solution was filtered and lyophilised. The resulting oil was then purified as specified for each compound.

General Procedure III: *N*α-Fmoc protection of coumaryl amino acids



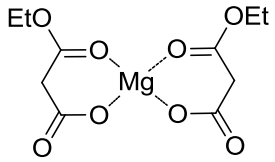
The amine (1 equiv.) was dissolved in 50% aqueous acetone (20 mL/mmol) and Fmoc-OSu (1.1 equiv.) was added, followed by triethylamine (2.0 equiv.), and the reaction mixture was stirred at rt for 16 h. The reaction mixture was diluted with water (5 volumes) and the mixture was extracted with diethyl ether (3 x 50 mL). The aqueous layer was separated and acidified to a pH of 3 using aqueous HCl (1 M).and the solution was extracted with ethyl acetate (3 x 50 mL). The organic phase was then separated and washed with aqueous HCl (1 M; 3 x 50 mL), brine (3 x 50 mL), dried over Na_2SO_4 and and concentrated *in vacuo*. The resulting residue was purified by recrystallisation (EtOAc: pet. spirit).

General Procedure IV: Precipitation of coumaryl amino acids

To the crude amino acid, anhydrous acetone (50 mL/100 mg) was added and the resulting mixture was sonicated for 30 min or until a precipitate evolved. The precipitate was filtered and washed with anhydrous acetone (50 mL/100 mg). The resulting solid was dissolved in water, filtered and lyophilised.

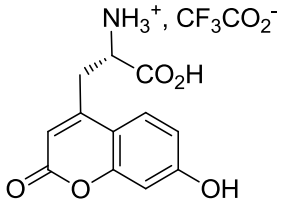
7.2.3 Experimental procedures

Ethyl magnesium malonate



An aqueous solution of magnesium chloride hexahydrate ($MgCl_2 \bullet 6H_2O$) (3.05 g, 10 ml, 1.5 M) was added to an aqueous solution of ethyl potassium malonate (5.10 g, 10 ml, 3 M). The resulting solution was stirred at rt for 30 min. The reaction mixture was diluted with isopropanol (10 volumes; 200 ml) and the mixture was stirred at rt for a further 30 min. The mixture was then filtered and the filtrate concentrated *in vacuo* to give ethyl magnesium malonate (4.21 g, 98%) as a white crystalline solid. The product dried under vacuum and used without further purification. ^{1}H NMR (300 MHz, D_2O, δ) ppm: 4.15 (app qd, *J* = 7.2, 0.9 Hz, CH_2CH_3, 4H), 3.26 (app d, *J* = 0.9 Hz, $CH_3CH_2OCOCH_2COO$, 4H), 1.22 (app td, *J* = 7.2, 1.0 Hz, CH_2CH_3, 6H).

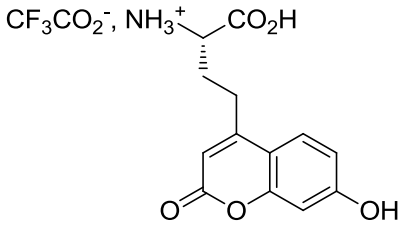
L-(7-hydroxycoumarin-4-yl) methylglycine trifluoroacetate salt (2.1)



The β-ketoester **2.13** (1.57 g, 3.7 mmol) was reacted with resorcinol (2.03 g, 18.5 mmol) and methanesulfonic acid (6 mL, 92.4 mmol) according to General Procedure II. The resulting residue was purified by semi-preparative RP-HPLC to give amino acid **2.1** (0.246 g, 19%) as an off-white solid. mp 168-174 °C. $[\alpha]^{23}_D$ = 17.25° (c. 1.0, 1 M HCl) (lit. [138] $[\alpha]^{20}_D$ = 8.4° (c. 0.485, 1 M HCl). ^{1}H NMR (300 MHz, DMSO-*d6*, δ) ppm: 10.67 (s, O*H*, 1H), 8.37 (br s, NH_3^+, 3H), 7.63 (d, *J* = 9.0 Hz, Ar*H*, 1H), 6.84 (dd, *J* = 8.7, 2.3 Hz, Ar*H*, 1H), 6.76 (d, *J* = 2.3 Hz, Ar*H*, 1H), 6.22 (s, CH_2CCHCO, 1H), 4.43 – 4.02 (m, -NHC*H*CH_2, 1H), 3.27 – 3.04 (app m, CH_2, 2H).

Spectral data matches literature:[138]

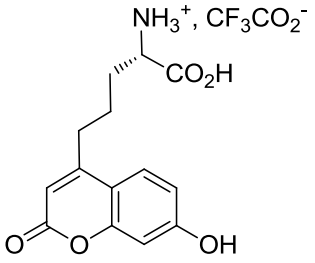
L-(7-hydroxycoumarin-4-yl) ethylglycine trifluoroacetate salt (2.2)



The β-ketoester **2.14** (2.4 g, 5.5 mmol) was reacted with resorcinol (3.0 g, 27.2 mmol) and methanesulfonic acid (9.0 mL, 0.136 mol) according to General Procedure II. The resulting residue was purified by semi-preparative RP-HPLC to give amino acid **2.2** (517 mg, 28%) as an off-white solid. mp > 220 °C. $[\alpha]^{23}_{D}$ = 24.18° (c. 1.0, 1 M HCl). ^{1}H NMR (600 MHz, DMSO-$_{d6}$, δ) ppm: 9.23 (br s, O*H*, 1H), 7.66 (d, *J* = 8.7 Hz, Ar*H*, 1H), 6.85 (dd, *J* = 8.6, 1.7 Hz, Ar*H*, 1H), 6.72 (d, *J* = 1.8 Hz, Ar*H*, 1H), 6.07 (s, CH_2CCHCO, 1H), 3.56 – 3.12 (m, $NHCHCH_2CH_2$, 1H), 3.01 – 2.67 (m, CH_2, 2H), 2.15 – 1.85 (m, CH_2, 2H). ^{13}C NMR (75 MHz, DMSO-$_{d6}$, δ) ppm: 169.70, 161.60, 160.35, 156.24, 155.19, 126.13, 113.01, 110.79, 109.19, 102.49, 102.46, 29.90, 27.40. MS (ESI+) *m/z* calculated for $C_{13}H_{13}NO_5$: 263.08; observed: 264.13 (M + H).

Spectral data matches literature:[134]

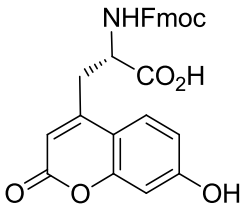
L-(7-hydroxycoumarin-4-yl) propylglycine trifluoroacetate salt (2.3)



The β-ketoester **2.16** (0.120 g, 0.26 mmol) was reacted with resorcinol (0.145 g, 1.32 mmol) and methanesulfonic acid (0.43 mL, 6.6 mmol) according to General Procedure II. The resulting residue was purified by semi-preparative RP-HPLC to give amino acid **2.3** (26.2 mg, 25%) as an off-white solid. mp > 220 °C. $[\alpha]^{23}_{D}$ = 13.80° (c. 1.0, 1 M HCl). ^{1}H NMR (600 MHz, DMSO-$_{d6}$, δ) ppm: 10.59 (br s, O*H*, 1H), 7.67 (d, *J* = 8.8 Hz, Ar*H*, 1H), 6.80 (dd, *J* = 8.7, 2.4 Hz, Ar*H*, 1H), 6.71 (d, *J* = 2.4 Hz, Ar*H*, 1H), 6.09 (s, CH_2CCHCO, 1H), 3.23 – 3.13 (m,

NHC*H*$CH_2CH_2CH_2$, 1H), 2.74 (qd, J = 14.7, 6.6 Hz, CH_2, 2H), 1.88 – 1.59 (m, CH_2CH_2, 4H). ^{13}C NMR (150 MHz, DMSO-d_6, δ) ppm: 161.07, 160.40, 156.88, 155.08, 126.52, 126.50, 112.87, 111.17, 109.24, 102.33, 53.82, 30.63, 30.56, 24.23. MS (ESI+) *m/z* calculated for $C_{14}H_{15}NO_5$: 277.27; observed: 277.89.

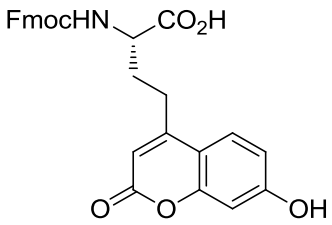
(2S)-2-(9H-fluoren-9-ylmethoxycarbonylamino)-4-(7-hydroxy-2-oxo-chromen-4-yl)propanoic acid (2.4)



Coumaryl amino acid **2.1** (0.298 g, 0.86 mmol) was reacted with Fmoc-OSu (0.305 g, 0.9 mmol) and triethylamine (0.36 mL, 2.6 mmol) according to General Procedure III. The resulting solid was recrystallised from EtOAc/pet. spirit to give the *N*α-Fmoc protected amino acid **2.4** (0.208 g, 51%) as an off-white solid. mp 138 - 144 °C. ^{1}H NMR (300 MHz, DMSO-d_6, δ) ppm: 7.98 – 7.79 (m, FmocAr*H*, 2H), 7.69 (d, *J* = 8.5 Hz, Ar*H*, 1H), 7.66 – 7.54 (m, FmocAr*H*, 2H), 7.50 – 7.07 (m, FmocAr*H*, 4H), 6.83 (dd, *J* = 8.8, 2.3 Hz, Ar*H*, 1H), 6.73 (d, *J* = 2.3 Hz, Ar*H*, 1H), 6.14 (s, CH_2CCHCO, 1H),), 4.79 – 4.64 (m, NHC*H*CH_2, 1H), 4.32 – 4.08 (m, FmocC*H*C*H*$_2$OCO, 3H), 3.07 – 2.84 (m, CH_2, 2H).

Spectral data matches literature:[138]

(2S)-2-(9H-fluoren-9-ylmethoxycarbonylamino)-4-(7-hydroxy-2-oxo-chromen-4-yl)butanoic acid (2.5)

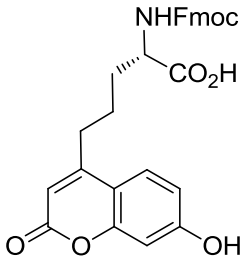


Coumaryl amino acid **2.1** (0.354 g, 0.94 mmol) was reacted with Fmoc-OSu (0.476 g, 1.4 mmol) and triethylamine (0.53 mL, 4.0 mmol) according to General Procedure III, to give *N*α-

Fmoc protected amino acid **2.4** as an off-white solid (0.409 g, 90%). mp 94 - 100 °C. ^{1}H NMR (300 MHz, DMSO-d_6, δ) ppm: 10.56 (s, O*H*, 1H), 7.90 (d, *J* = 7.5 Hz, FmocAr*H*, 2H), 7.84 (d, *J* = 8.3 Hz, FmocAr*H*, 1H), 7.74 (dd, *J* = 7.1, 0.8 Hz, FmocAr*H*, 1H), 7.67 (d, *J* = 8.8 Hz, Ar*H*, 1H), 7.42 (t, *J* = 7.5 Hz, FmocAr*H*, 2H), 7.38 – 7.28 (m, FmocAr*H*, 2H), 6.79 (dd, *J* = 8.7, 2.3 Hz, Ar*H*, 1H), 6.72 (d, *J* = 2.3 Hz, Ar*H*, 1H), 6.08 (s, CH_2CCHCO, 1H), 4.50 – 4.14 (m, FmocC*H*CH_2OCO, 3H), 4.07 (td, *J* = 9.2, 4.7 Hz, NHC*H*CH_2CH_2, 1H), 3.06 – 2.68 (m, CH_2, 2H), 2.25 – 1.72 (m, CH_2, 2H). ^{13}C NMR (75 MHz, DMSO-d_6, δ) ppm: 173.57, 161.24, 160.38, 156.21, 155.18, 143.88, 143.82, 140.77, 140.76, 127.68, 127.11, 126.35, 125.29, 125.27, 120.18, 113.00, 111.02, 109.68, 102.49, 79.21, 65.64, 53.56, 46.72, 30.01, 27.85.

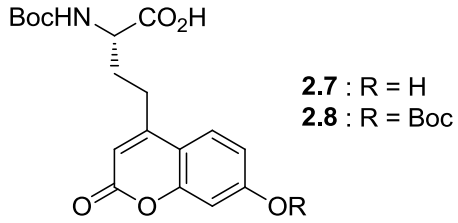
Spectral data matches literature:[134]

(2S)-2-(9H-fluoren-9-ylmethoxycarbonylamino)-4-(7-hydroxy-2-oxo-chromen-4-yl)pentanoic acid (2.6)



Coumaryl amino acid **2.1** (0.0687 g, 0.18 mmol) was reacted with Fmoc-OSu (0.651 g, 0.19 mmol) and triethylamine (0.08 mL, 0.055 mmol) according to General Procedure III, to give the *N*α-Fmoc protected amino acid **2.4** as an off-white solid (0.069 g, 76%).mp 107-114 °C. ^{1}H NMR (300 MHz, DMSO-d_6, δ) ppm: 10.55 (br s, O*H*, 1H), 7.88 (d, *J* = 7.3 Hz, FmocAr*H*, 2H), 7.70 (d, *J* = 7.5 Hz, FmocAr*H*, 2H), 7.65 (d, *J* = 7.5 Hz, Ar*H*, 1H), 7.40 (t, *J* = 7.4 Hz, FmocAr*H*, 2H), 7.34 – 7.19 (m, FmocAr*H*, 2H), 6.78 (dd, *J* = 8.7, 2.3 Hz, Ar*H*, 1H), 6.72 (d, *J* = 2.2 Hz, Ar*H*, 1H), 6.09 (s, CH_2CCHCO, 1H), 4.44 – 4.10 (m, FmocC*H*CH_2OCO, 3H), 4.10 – 3.81 (m, NHC*H*CH_2CH_2, 1H), 2.91 – 2.61 (m, CH_2, 2H), 1.89 – 1.37 (m, CH_2CH_2, 4H). ^{13}C NMR (75 MHz, DMSO-d_6, δ) ppm: 173.62, 161.07, 160.38, 156.77, 156.03, 155.13, 143.82, 140.67, 135.34, 128.76, 127.58, 127.00, 126.36, 125.18, 124.74, 121.17, 120.06, 112.89, 111.12, 109.35, 102.37, 65.50, 53.61, 46.63, 41.04, 30.46, 27.72.

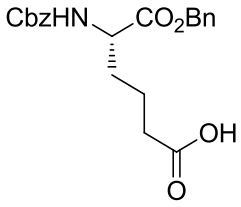
(S)-2-((tert-butoxycarbonyl)amino)-4-(7-hydroxy-2-oxo-2H-chromen-4-yl)butanoic acid (2.7) and (S)-2-((tert-butoxycarbonyl)amino)-4-(7-((tert-butoxycarbonyl)oxy)-2-oxo-2H-chromen-4-yl)butanoic acid (2.8)



Coumaryl amino acid **2.2** (87 mg, 0.23 mmol) was dissolved in 5% aqueous $NaHCO_3$/Dioxane (1:1 v/v) (20 mL). The solution was cooled in an ice bath and Boc anhydride (0.542 g, 2.4 mmol) was added. The reaction was stirred on ice for 1 h before stirring for a further 16 h at rt. The reaction mixture was acidified to a pH of 3 using 10% (w/v) aqueous citric acid (8 mL) and extracted with EtOAc (3 x 20 mL). The organic phase was separated and washed with water (3 x 50 mL), brine (3 x 50 mL), dried over Na_2SO_4 and concentrated *in vacuo*. The resultant oil was dissolved in a 30% aqueous acetonitrile and lyophilised to give a mixture of the *N*α-Boc protected coumaryl amino acid (**2.7**) (76 mg, 90%) and the di-Boc protected coumaryl amino acid (**2.8**) (8 mg, 8%) as an off-white solid in a ratio of 10:1. Compound **2.7**: ^{1}H NMR (600 MHz, $DMSO_{-d6}$, δ) ppm: 10.58 (br s, O*H*, 1H), 7.67 (d, J = 8.6 Hz, Ar*H*, 1H), 6.84 – 6.75 (m, Ar*H*, 1H), 6.75 – 6.68 (m, Ar*H*, 1H), 6.07 (s, Ar*H*, 1H), 4.06 – 3.92 (m, NHC*H*CH_2CH_2, 1H), 2.83 – 2.70 (m, CH_2, 1H), 2.07 – 1.78 (m, CH_2, 2H), 1.47 (s, (CH_3)$_3$, 9H). Compound **2.8**: ^{1}H NMR (600 MHz, $DMSO_{-d6}$, δ) ppm: 10.58 (br s, O*H*, 1H), 7.91 (d, J = 8.9 Hz, Ar*H*, 1H), 7.42 – 7.33 (m, Ar*H*, 1H), 7.00 - 6.84 (m, Ar*H*, 1H), 6.33 (s, Ar*H*, 1H), 3.92 – 3.80 (m, NHC*H*CH_2CH_2, 1H),), 2.93 – 2.81 (m, CH_2, 2H), 2.07 – 1.78 (m, CH_2, 2H), 1.51 (s, (CH_3)$_3$, 9H), 1.40 (s, (CH_3)$_3$, 9H). Mixture (**2.7** and **2.8**): ^{13}C NMR (75 MHz, $DMSO_{-d6}$, δ) ppm: 173.80, 161.19, 160.39, 158.49, 156.27, 155.70, 155.20, 126.41, 112.99, 111.07, 109.75, 106.23, 102.48, 78.23, 53.15, 29.91, 28.26, 27.96, 27.90, 27.26, 26.92.

Spectral data matches literature:[134,138]

(S)-6-(benzyloxy)-5-(((benzyloxy)carbonyl)amino)-6-oxohexanoic acid (2.11)



Protected acid (**2.2**) (1.01 g, 2.72 mmol) was dissolved in anhydrous THF (5 mL) and the solution was cooled to -20 °C. To this solution *N*-Methylmorpholine (0.32 mL, 2.8 mmol) and isobutyl chloroformate (0.37 mL, 2.8 mmol) were added and the reaction mixture was stirred on ice for a further 15 min. A solution of freshly prepared CH_2N_2[xix] in diethyl ether was added until a yellow colour persisted, and stirring continued for a further 3 h at rt. The excess diazomethane was quenched with acetic acid, and the reaction mixture was diluted with EtOAc (100 mL). The organic phase was separated and washed with saturated aqueous NH_4Cl (3 x 50 mL), saturated aqueous $NaHCO_3$ (3 x 50 mL) and brine (3 x 50 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The resulting residue was purified by silica column chromatography to give diazoketone **2.12** as a yellow oil.

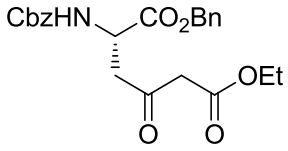
Diazoketone **2.12** was dissolved in EtOAc (5 mL/100 mg) and silver trifluoroacetate (27.6 mg, 0.125 mmol) and silica gel (1 g/100 mg) were added. The reaction mixture was agitated at 50 °C for 15 min with the exclusion of light.[xx] The mixture was filtered, washed with EtOAc, and the filtrate concentrated *in vacuo*. The resulting residue was recrystallised from EtOAc/pet. spirit to give the amino acid **2.11** (0.628 g, 60%) as a white solid. mp 78-84 °C. ^{1}H NMR (300 MHz, $CDCl_3$, δ) ppm: 7.48 – 7.26 (m, Ar*H*, 10H), 5.43 (d, *J* = 8.5 Hz, N*H*CHCH$_2$CH$_2$CH$_2$, 1H), 5.17 (s, COOC*H*$_2$Ph, 2H), 5.10 (s, PhC*H*$_2$OCONH$_2$, 2H), 4.55 – 4.33 (m, NHC*H*CH$_2$CH$_2$CH$_2$, 1H), 2.54 – 2.18 (m, C*H*$_2$, 2H), 2.02 – 1.49 (m, C*H*$_2$C*H*$_2$, 4H). ^{13}C NMR

[xix] WARNING: Highly explosive – avoid contact with ground or unsmooth glass, and keep cold. Prepared according to Vogel, A. I. *Vogel's Textbook of Practical Organic Chemistry*; 3rd ed.; Longman Scientific & Technical, 1956.

[xx] Agitation was achieved by rotation on a rotary evaporator and light was excluded by wrapping the flask in foil.

(75 MHz, $CDCl_3$, δ) ppm: 178.44, 172.06, 155.93, 136.05, 135.07, 128.57, 128.47, 128.24, 128.15, 128.05, 77.42, 76.99, 76.57, 67.26, 67.03, 53.59, 33.13, 31.73, 20.22.

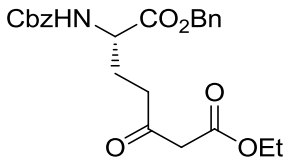
1-benzyl 6-ethyl 2-(((benzyloxy)carbonyl)amino)-4-oxohexanedioate (2.13)



Protected acid **2.9** (5 g, 0.014 mol) was reacted with CDI (2.5 g, 0.015 mol) and ethyl magnesium malonate (2.2 g, 0.008 mol) according to General Procedure I, to give β-ketoester **2.13** (5.54 g, 93%) as a white waxy solid. mp 74-77 °C. ^{1}H NMR (300 MHz, $CDCl_3$, δ) ppm: 7.39-7.28 (m, Ar*H*, 10H), 5.74 (d, *J* = 8.3 Hz, N*H*CHCH$_2$CH$_2$, 1H), 5.16 (s, COOC*H*$_2$Ph, 2H), 5.10 (s, PhC*H*$_2$OCONH, 2H), 4.75 – 4.50 (m, NHC*H*CH$_2$, 1H), 4.16 (q, *J* = 7.1 Hz, OC*H*$_2$CH$_3$, 2H), 3.40 (s, COC*H*$_2$CO, 2H), 3.23 (ddd, *J* = 53.7, 18.5, 4.4 Hz, NHC*H*$_2$CO,1H), 1.24 (t, *J* = 7.1 Hz, OCH$_2$C*H*$_3$, 3H). ^{13}C NMR (75 MHz, $CDCl_3$, δ) ppm: 200.69, 170.47, 166.37, 155.96, 155.93, 136.02, 135.09, 128.52, 128.45, 128.36, 128.16, 128.12, 127.98, 67.50, 67.02, 61.52, 49.85, 49.03, 44.52, 13.97.

Spectral data matches literature:[138]

1-benzyl 6-ethyl 2-(((benzyloxy)carbonyl)amino)-4-oxoheptanedioate (2.14)

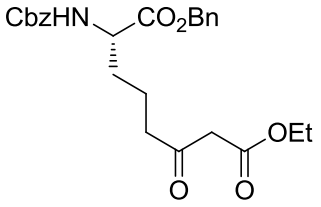


Protected acid **2.10** (10.98 g, 0.03 mol) was reacted with CDI (5.27 g, 0.033 mol) and ethyl magnesium malonate (4.7 g, 0.016 mmol) according to General Procedure I, to give β-ketoester **2.14** (12.42 g, 95%) as a white waxy solid. mp 52-59 °C. ^{1}H NMR (300 MHz, $CDCl_3$, δ) ppm: 7.42-7.28 (m, Ar*H*, 10H), 5.38 (d, *J* = 7.7 Hz, N*H*CHCH$_2$CH$_2$, 1H), 5.16 (s, -COOC*H*$_2$Ph, 2H), 5.10 (s, PhC*H*$_2$OCONH, 2H), 4.40 (dd, *J* = 13.6, 9.2 Hz, NHC*H*CH$_2$CH$_2$,1H), 4.17 (q, *J* = 7.1 Hz, OC*H*$_2$CH$_3$, 2H), 2.71 – 2.39 (m, NHCHCH$_2$C*H*$_2$CO, 2H), 2.38 – 1.80 (m, NHCHC*H*$_2$CH$_2$CO, 2H), 1.57 (s, COC*H*$_2$CO, 2H), 1.25 (t, *J* = 7.1 Hz, -

OCH_2CH_3, 3H). ^{13}C NMR (75 MHz, $CDCl_3$, δ) ppm: 201.65, 171.81, 167.04, 156.07, 136.21, 135.20, 128.72, 128.62, 128.61, 128.45, 128.29, 128.18, 67.44, 67.15, 61.50, 53.27, 49.24, 38.65, 26.29, 14.15.

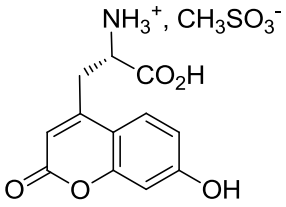
Spectral data matches literature:[134]

1-benzyl 6-ethyl 2-(((benzyloxy)carbonyl)amino)-4-oxooctanedioate (2.15)



Protected acid **2.11** (0.984 g, 2.6 mmol) was reacted with CDI (0.455 g, 2.8 mmol) and ethyl magnesium malonate (0.40 g, 1.4 mmol) according to General Procedure I, to give β-ketoester **2.15** (0.97 g, 84%) as a white waxy solid. mp 63-68 °C. ^1H NMR (300 MHz, $CDCl_3$, δ) ppm: 7.42-7.27 (m, Ar*H*, 10H), 5.40 (d, *J* = 8.2 Hz, N*H*CHCH$_2$CH$_2$CH$_2$, 2H), 5.17 (s, COOC*H*$_2$Ph, 2H), 5.10 (s, PhC*H*$_2$OCONH$_2$, 2H), 4.41 (dd, *J* = 13.3, 5.9 Hz, NHC*H*CH$_2$CH$_2$CH$_2$, 1H), 4.18 (q, *J* = 7.1 Hz, OC*H*$_2$CH$_3$, 2H), 3.38 (s, CH$_2$CC*H*CO, 2H), 2.75 – 2.37 (m, C*H*$_2$, 2H), 1.95 – 1.51 (m, C*H*$_2$C*H*$_2$, 4H), 1.26 (t, *J* = 7.1 Hz, OCH$_2$C*H*$_3$, 1H). ^{13}C NMR (75 MHz, $CDCl_3$, δ) ppm: 201.82, 171.86, 166.95, 155.75, 136.01, 135.02, 128.48, 128.37, 128.35, 128.20, 128.03, 127.93, 67.11, 66.87, 61.26, 53.45, 49.02, 41.81, 31.50, 18.74, 13.93.

L-(7-hydroxycoumarin-4-yl) methylglycine methanesulfonate salt (2.16)

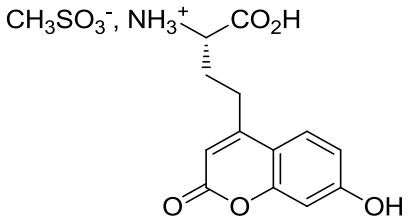


The β-ketoester **2.13** (0.57 g, 1.35 mmol) was reacted with resorcinol (1.48 g, 13.4 mmol) and methanesulfonic acid (2.2 mL, 33.8 mmol) according to General Procedure II. The resulting residue was purified according to General Procedure IV, to give amino acid **2.1** as an orange solid (43.2 mg, 9%). mp > 220 °C. ^1H NMR (300 MHz, $DMSO_{-d6}$, δ) ppm: 10.69 (br s, O*H*, 1H), 8.37 (br s, NH_3^+, 3H), 7.62 (d, *J* = 8.7 Hz, Ar*H*, 1H), 6.85 (dd, *J* = 8.8, 2.3 Hz, Ar*H*, 1H),

6.76 (d, *J* = 2.2 Hz, Ar*H*, 1H), 6.22 (s, CH_2CCHCO,1H), 4.37 – 4.01 (m, $NHCHCH_2$, 1H), 3.41 – 3.29 (app m, CH_2, 1H), 3.15 (dd, *J* = 15.1, 8.6 Hz, CH_2, 1H), 2.35 (s, $CH_3SO_3^-$, 3H). ^{13}C NMR (75 MHz, MHz, DMSO-*d6*, δ) ppm: 169.94, 161.53, 160.16, 155.38, 149.93, 126.20, 113.24, 112.85, 110.86, 102.73, 51.18, 31.98.

Spectral data matches literature:[138]

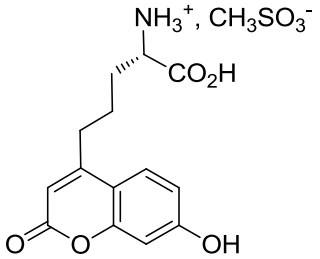
L-(7-hydroxycoumarin-4-yl) ethylglycine methanesulfonate salt (2.17)



The β-ketoester **2.14** (0.243 g, 0.55 mmol) was reacted with resorcinol (0.606 g, 5.5 mmol) and methanesulfonic acid (0.9 mL, 13.7 mmol) according to General Procedure II. The resulting residue was purified according to General Procedure IV, to give amino acid **2.2** as an orange solid (27.2 mg, 13%). mp 110 - 116 °C. ^{1}H NMR (300 MHz, DMSO-*d6*, δ) ppm: 10.61 (s, O*H*, 1H), 8.32 (br s, NH_3^+, 1H), 7.63 (d, *J* = 8.8 Hz, Ar*H*, 1H), 6.83 (dd, *J* = 8.7, 2.3 Hz, Ar*H*, 1H), 6.74 (d, *J* = 2.3 Hz, Ar*H*, 1H), 6.13 (s, CH_2CCHCO, 1H), 4.15 - 3.97 (m, $NHCHCH_2CH_2$, 1H), 3.03 – 2.60 (m, CH_2, 2H), 2.31 (s, $CH_3SO_3^-$, 3H), 2.23 – 1.89 (m, CH_2, 2H). ^{13}C NMR (75 MHz, DMSO-*d6*, δ) ppm: 170.72, 161.31, 160.33, 155.19, 155.10, 126.18, 113.09, 110.89, 109.74, 102.58, 51.64, 28.67, 26.74.

Spectral data matches literature:[134]

L-(7-hydroxycoumarin-4-yl) propylglycine methanesulfonate salt (2.18)



The β-ketoester **2.15** (0.387 g, 0.85 mmol) was reacted with resorcinol (0.468 g, 4.3 mmol) and methanesulfonic acid (2.04 mL, 21 mmol) as per General Procedure II. The resulting residue was purified according to General Procedure IV, to give amino acid **2.3** (68.7 mg, 22%) as an orange solid. ^{1}H NMR (300 MHz, DMSO-d_6, δ) ppm: 10.59 (br s, O*H*, 1H), 7.68 (d, *J* = 8.6 Hz, Ar*H*, 1H), 6.80 (dd, *J* = 8.7, 2.4 Hz, Ar*H*, 1H), 6.71 (d, *J* = 2.3 Hz, Ar*H*, 1H), 6.09 (s, CH_2CCHCO, 1H), 5.76 (s, O*H*, 1H), 3.23 – 3.13 (m, -NHC*H*$CH_2CH_2CH_2$, 1H), 2.79 – 2.61 (m, CH_2, 2H), 1.92 – 1.53 (m, CH_2,CH_2, 4H). MS (ESI+) *m/z* calculated for $C_{14}H_{15}NO_5$: 277.27; observed: 277.8.

7.3 Experimental described in Chapter 3

7.3.1 Experimental materials

All reagents were of analytical grade or higher. Ampicillin, acetic acid, 2-hydroxyethyl disulfide (2-HED), dithiothreitol (DTT), β-mercaptoethanol, urea, tricine, bromphenol blue, Coomassie Brilliant Blue R250, kanamycin, sodium dodecyl sulfate (SDS), tetracycline, trifluoroacetic acid (TFA) (HPLC grade), *N,N.N,N'N'*-tetramethylethylene-diamine (TEMED) and Tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl) and *N*-2-hydroxyethylpiperazine-*N*-2-ethanesulfonic acid (HEPES) were purchased from Sigma-Aldrich (Castle Hill, NSW, Australia). α-lytic protease Prag A9 was a kind gift from Gro*Pep* Pty Ltd. (Thebarton, South Australia, Australia). Long^{TM}Arg^{3}IGF-I was purchased from Novozymes *GroPep* Pty Ltd. (Adelaide, South Australia). HPLC grade acetonitrile and methanol were purchased from Scharlau Chemie (Sentmenat, Barcelona, Spain). Bis-acrylamide and acrylamide was purchased from BioRad Pty Ltd. (North Ryde, Australia). Yeast extract and tryptone were OXOID brand and purchased from Thermo Fisher Scientific (VIC, Australia.). Mark-12™ markers for Coomassie gels were purchased from NOVEX, (Carlsbad, CA, U.S.A). Durapore® 0.22 μm filter was purchased from Millipore Corp. (Belford, MA, U.S.A.).

Analytical RP-HPLC was performed on an Agilent 1100 series HPLC equipped with a variable wavelength detector and a Grace Vydac C4 column (5 μm, 2.1 x 100 mm). The product was eluted using a linear gradient of 25-55% acetonitrile over 30 min with a flow rate of 0.5 mL/min, where the solvent system used was: A (MilliQ water + 0.01% TFA) and B (80% acetonitrile + 0.08% TFA). Elution of product was monitored at 215 nm.

Semi-preparative RP-HPLC was carried out on a Waters Delta Prep 3000 chromatography system equipped with a multiple wavelength detector using a Grace Vydac C4 column (5 μm, 4.6 x 150 mm) The product was eluted using 25-55% acetonitrile over 80 min with a flow rate of 5 mL/min, where the solvent system used was: A (MilliQ water + 0.01% TFA) and B (80% acetonitrile + 0.08% TFA). Fractions were analysed by analytical RP-HPLC. Pure fractions were pooled, and lyophilised.

7.3.1.1 Bacterial strains and genotypes

E. coli BL21 (DE3)	F– *ompT hsd*SB(rB–, mB–) *gal dcm* (DE3)

7.3.1.2 Expression vectors and antibiotic resistance

pET32a-F19X-IGF-II	ampicillin (100 μg/ml)
pET32a-F28X-IGF-II	ampicillin (100 μg/ml)
pET32a-L53X-IGF-II	ampicillin (100 μg/ml)
pEB-CouRS	tetracycline (50 μg/ml)
pRFSDuet™-CouRS	kanamycin (30 μg/ml)

7.3.1.3 Standard solutions, buffers and growth media

The solutions listed below as standard media were prepared by the Central Services Unit, School of Molecular and Biomedical Science, The University of Adelaide. All solutions were prepared using water purified by the Milli-Q Ultra Pure Water System (Millipore Pty Ltd, North Ryde, Australia) and sterilised by autoclaving or sterilised by filtering through a 0.22 μm.

Standard media

Luria-Bertani (LB) Media	0.5% yeast extract, 1% (w/v) tryptone, 0.17 M NaCl, pH 7.0
Luria-Bertani (LB) Agar	LB media + 1.5% (w/v) agar
2 xYT Media	1% (w/v) yeast extract, 1.6 % (w/v) tryptone, 85 mM NaCl, pH 6.8.

Alternative media and buffers

Minimal media (MIN)	20% (v/v) 5 X M9 minimal salts solution, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% (v/v) glycerol and sterilised through 0.22 µm filter
M9 minimal salts solution (5 X concentrate)	6.4% (w/v) $Na_2HPO_4 \bullet 7H_2O$, 1.5% (w/v) KH_2PO_4, 2.5% (w/v) NaCl , 5% (w/v) NH_4Cl and sterilised through 0.22 µm filter
Auto-inducing media (ZYM-5052) (AIM) [178]	0.5% (w/v) yeast extract, 1% (w/v) tryptone, 25 mM Na_2HPO_4, 25 mM KH_2PO_4. 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.2 X trace metals, 0.5% (v/v) glycerol, 0.05% (v/v) glucose, 0.2% (w/v) lactose and sterilised through 0.22 µm filter
Non-inducing media (ZYM-505) (NIM)	0.5% (w/v) yeast extract, 1% (w/v) tryptone, 25 mM Na_2HPO_4, 25 mM KH_2PO_4. 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.2 X trace metals, 0.5% (v/v) glycerol, 0.05% (v/v) glucose and sterilised through 0.22 µm filter
1000 X trace metals[178]	50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM $MnCl_2$, 10 mM $ZnSO_4$, 2 mM $CoCl_2$, 2 mM $CuCl_2$, 2 mM $NiCl_2$, 2 mM Na_2MoO_4, 2 mM Na_2SeO_3 and 2 mM H_3BO_3 in 60 mM HCl

Transformation buffers

Transformation Buffer 1	30 mM K_2CO_3, 100 mM RbCl, 10 mM $CaCl_2 \bullet 2H_2O$, 50 mM $MnCl_2 \bullet 4H_2O$, 15% (v/v) Glycerol – pH to 5.8 and sterilised through 0.22 µm filter
Transformation Buffer 2	10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2 \bullet 2H_2O$, 15% Glycerol (v/v) – pH to 6.5 and sterilised through 0.22 µm filter.

Expression and purification buffers

Inclusion body wash buffer/ resuspension buffer	10 mM KH_2PO_4, 30 mM NaCl, pH 7.8 and sterilised through 0.45 µm filter.
Dissolution buffer	8 M urea, 0.1 M Tris, 40 mM glycine, pH 2 and sterilised through 1.0 µm filter.
Refold dilution buffer	1 M Tris-HCl, 2 mM EDTA, pH 9.1 and sterilised through 1.0 µm filter.
HPLC Buffer A	0.1% TFA in MilliQ water and filtered through a 0.22 µm filter.
HPLC buffer B	80% acetonitrile in MilliQ water, 0.08% TFA and filtered through a 0.22 µm filter.

Protein electrophoresis buffers

Tricine running buffer	0.1 M tricine, 0.01 M Tris-HCl (pH 8.3), 0.1% (v/v) SDS
2 x Protein loading buffer	125 mM Tris-HCl, 4% (v/v) SDS, 10% (v/v) glycerol, 0.1% Bromophenol Blue, pH 6.8
Gel Fixing Solution	50% (v/v) aqueous ethanol, 10% (v/v) aqueous acetic acid
Gel Destain Solution	10% (v/v) aqueous acetic acid
Coomassie Blue Stain	0.5% (w/v) Coomassie brilliant blue R-250, 10% (v/v) aqueous acetic acid, 50% (v/v) aqueous ethanol
Gel Drying Solution	5% (v/v) aqueous glycerol, 30% (v/v) aqueous ethanol

7.3.2 Methods

General Procedure V: Preparation of competent cells

A single colony of *E. coli* BL21 (DE3) competent cells containing a plasmid encoding the *Mj*CouRS and *Mj*tRNA^{Cou} genes (refer to Section 7.3.1.2) was grown overnight at 37 °C in LB (5 mL). The following day an aliquot of the overnight culture (330 µL) was subcultured into LB (10 mL) and incubated at 37 °C until the OD_{600nm} reached 0.6. The culture was then further subcultured (5 mL) into fresh pre-warmed LB (100 mL). Once this culture reached an OD_{600nm} of 0.6 the cells were pelleted at 4,000 rpm for 5 min at 4 °C, resuspended in transformation buffer 1 (refer to Section 7.3.1.3) (10 mL/25 mL culture) and incubated on ice for 5 min. The cells were pelleted at 4,000 rpm for 5 min at 4 °C, before resuspending the cells in transformation buffer 2 (refer to Section 7.3.1.3) (4 mL). The cells were incubated on ice for 15 min before aliquoting into 100 µL or 200 µL volumes and stored at –80 °C.

General Procedure VI: Transformation

Competent *E. coli* BL21 (DE3) cells (100 µL) prepared according to General Procedure V were placed on ice and an aliquot of pET expression plasmid (refer to Section 7.3.1.2) (0.5-1 µL) was added. The cells were incubated on ice for 10 min, heated to 42 °C for 2 min followed by incubation on ice for a further for 5 min. The freshly transformed cells were aseptically spread onto LB agar plates supplemented with the appropriate antibiotics (refer to Section 7.3.1.2).

General Procedure VII: Expression with IPTG induction

Expression plasmid (pET32a; refer to Section 7.3.1.2) (1 µL) containing the coding sequence for mutant human IGF-II gene with the desired codon mutated to the unique TAG codon was transformed into *E. coli* BL21 (DE3) (100 µL) containing the *Mj*CouRS and *Mj*tRNA^{Cou} genes according to General Procedure VI. The transformant was plated and incubated overnight at 37 °C. The following day a single colony was picked and used to inoculate medium (5 mL), containing appropriate antibiotics (refer to Section 7.3.1.2), and then incubated overnight at 37 °C. The following day an aliquot (to give starting OD_{600nm} of 0.18) of the overnight culture was subcultured into pre-warmed medium (5 mL) containing appropriate antibiotics (refer to

Section 7.3.1.2) and supplemented with coumaryl amino acid **2.2** where required. The cultures were incubated and shaken at 37 °C until the OD_{600nm} reached 0.6 at which time IPTG was added to induce protein expression. The induced cultures were incubated at 37 °C and shaken vigorously for 18 h. The expression was analysed by running pre- and post-induction samples of the cultures on a 15% tris-tricine SDS polyacrylamide gel according to General Procedure X.

General Procedure VIII: Expression in auto-induction medium (AIM)

Expression plasmid (pET32a; refer to Section 7.3.1.2) (1 μL) containing the coding sequence for mutant human IGF-II gene with the desired codon mutated to the unique TAG codon was transformed into *E. coli* BL21 (DE3) (100 μL) containing the *Mj*CouRS and *Mj*tRNA^{Cou} genes (refer to Section 7.3.1.2). The transformant was then plated onto an LB agar plate supplemented with appropriate antibiotics (refer to Section 7.3.1.2) and incubated overnight at 37 °C. The following day a single colony was picked and used to inoculate medium (5 mL), containing appropriate antibiotics (refer to Section 7.3.1.2), and then incubated overnight at 37°C. The following day an aliquot (to give starting OD_{600nm} of 0.18) of the overnight culture was subcultured into pre-warmed auto-inducing medium (ZYM-5052, AIM, refer to Section 7.3.1.3) (5 mL) containing appropriate antibiotics (refer to Section 7.3.1.2), and supplemented with coumaryl amino acid (**2.2**) where required. The induced cultures were incubated at 37 °C and shaken vigorously for 18 h. The expression was analysed by running pre- and post-induction samples of the cultures on a 15% tris-tricine SDS polyacrylamide gel according to General Procedure X.

General Procedure IX: Expression using matured E. coli BL21 (DE3) cells

Expression plasmid (pET32a; refer to Section 7.3.1.2) (1 μL) containing the coding sequence for mutant human IGF-II gene with the desired codon mutated to the unique TAG codon was transformed into *E. coli* BL21 (DE3) (100 μL) containing the *Mj*CouRS and *Mj*tRNA^{Cou} genes according to General Procedure VI. The transformant was plated and incubated overnight at 37 °C. The following day a single colony was picked and used to inoculate LB medium (5 mL), containing ampicillin (100 μg/ml) and tetracycline (50 μg/ml) and then incubated at 37 °C

overnight. The next day, an aliquot (170 µL) of the culture was subcultured into pre-warmed MIN (5 mL) (refer to Section 7.3.1.3) containing ampicillin (100 µg/ml) and tetracycline (50 µg/ml) and then incubated at 37 °C for 8 h. An aliquot (170 µL) of this culture was then subcultured into pre-warmed MIN (5 mL) containing ampicillin (100 µg/ml) and tetracycline (50 µg/ml) and then incubated at 37 °C for overnight. This culture was then further subcultured into pre-warmed MIN (5 mL) containing ampicillin (100 µg/ml) and tetracycline (50 µg/ml) for four more growth cycles (8 h, overnight, 8 h and overnight). After 6 growth cycles in MIN, an aliquot (to give starting OD_{600nm} of 0.18) of the overnight matured culture was subcultured into pre-warmed MIN (5 mL) containing ampicillin (100 µg/ml) and tetracycline (50 µg/ml) and supplemented with coumaryl amino acid (**2.2**) where required. The induced cultures were incubated at 37 °C and shaken vigorously for 18 h. The expression was analysed by running pre- and post-induction samples of the cultures on a 15% tris-tricine SDS polyacrylamide gel according to General Procedure X.

General Procedure X: SDS polyacrylamide gel electrophoresis

The OD_{600nm} of the pre- and post-induction cultures were measured and a sample of each ferment (1 mL) was collected. The samples were centrifuged at 13,200 rpm for 1 min and the supernatants removed. Each pellet was resuspended in lysis buffer (2% (v/v) SDS, 10% (v/v) β-mercaptoethanol) at a volume of 40 µL per unit OD_{600nm}. This standardised the concentration of bacteria that were loaded onto the gels and subsequent visualisation of the relative amount of protein expression in each flask could be made. To the lysed cells (5 µL), 2 X Loading buffer (5 µL) was added and the mixture was boiled at 100 °C for 5 min, centrifuged at 13,200 rpm for 1 min and then the entire volume was loaded onto the gel. All samples were analysed on a Mini-PROTEAN® 3 Electrophoresis Cell (BIORAD, NSW, Australia), using 15% tricine resolving gels with 12% tricine stacking gels as described by Schagger and von Jasgow [252]. Gels were run at a voltage of 40 V per gel in tricine running buffer (refer to Section 7.3.1.3) until the dye front had electrophoresed off the end of the gel. All gels were run against Mark12™ marker and once complete visualised under UV light (254 nm) where stated, before staining with Coomassie brillant blue. Once staining was complete the gel was placed in a destain

solution until the background stain had dissipated (all buffers used are detailed in Section 7.3.1.3).

General Procedure XI: Quantification of IGF-II analogues

Quantification of IGF-II analogues was performed by comparing analytical RP-HPLC C4 profiles with profiles of standard Long^{3}Arg^{3}IGF-I preparations, on an Agilent 1100 series RP-HPLC using the procedure described by Denley *et al.*[22]

7.3.3 Experimental procedures

7.3.3.1 Expression of the F19Cou IGF-II protein (3.1)

The pET32a-F19X-IGF-II expression plasmid (1 µL) was transformed into *E. coli* BL21 (DE3) containing the pRFSDuet™-CouRS expression vector (100 µL) according to General Procedure VI. The transformant was then plated onto an LB agar plate supplemented with ampicillin (100 µg/mL) and kanamycin (30 µg/mL) and incubated overnight at 37 °C. The following day single colonies were picked and used to inoculate non-inducing medium (5 mL) (ZYM-505; NIM, refer to Section 7.3.1.3) containing ampicillin (100 µg/ mL) and kanamycin (30 µg/mL). The culture was then incubated for 8 h at 37 °C. Aliquots (310 µL (1 in 10000 dilution)) of the culture were then subcultured into pre-warmed auto-inducing medium (8 x 310 mL) (ZYM-5052, AIM, refer to Section 7.3.1.3) containing ampicillin (100 µg/mL) and kanamycin (30 µg/mL) and supplemented with coumaryl amino acid **2.2** (4 mM). The cultures were then incubated at 37 °C and shaken vigorously for 18 h. At the end of the fermentation the bacterial cells were pelleted by centrifugation at 5,000 rpm for 15 min at 4 °C. The supernatant was discarded and the remaining pellet stored at -80 °C until further use. Successful expression was demonstrated by running pre- and post-induction samples of the cultures on a 15% tris-tricine SDS polyacrylamide gel according to General Procedure X.

7.3.3.2 Lysis of E. coli by French press

Pelleted bacterial cells were resuspended in inclusion body wash buffer (30 mL/500 mL culture) (refer to Section 7.3.1.3) and inclusion bodies (IBs) were released using two passes through a French press at 3,500 lb/in^2. The IBs and cell debris were spun at 4,500 rpm for 10 min at 4° C and the supernatant removed. The pellet was resuspended in inclusion body wash buffer (15 mL) and incubated at rt for 15 min before centrifugation at 4,500 rpm for 10 min at 4 °C. The supernatant was removed and the pellet once again resuspended in inclusion body wash buffer (15 mL) and finally spun at 6,000 rpm for 10 min at 4 °C. The supernatant was removed and the final pellet was stored at -80 °C until further use.

7.3.3.3 Gel filtration

Gel filtration was performed on a Superdex 75 (10 x 300 mm) column (GE Healthcare) using a GE ÄKTA™ (Amersham Pharmacia) FPLC system equipped with UV detection, with monitoring at 280 nm. Gel filtration column (Superdex 75) was sanitised with 0.5 column volumes of 0.5 M NaOH then washed with 0.5 column volumes of MilliQ water before equilibrating with 1 column volume of dissolution buffer containing 1.6 mM DTT (refer to Section 7.3.1.3) at a flow rate of 0.5 mL/min.

Inclusion bodies were dissolved in dissolution buffer containing 1.6 mM DTT for 30 min at rt at a final concentration of 10 mL/g (wet weight). Dissolved IBs were loaded on to the gel filtration column and eluted at a flow rate of 0.5 mL/min using dissolution buffer containing 1.6 mM DTT. Fractions were analysed by analytical RP-HPLC on an Agilent 1100 series. Fractions determined to contain the Long F19Cou IGF-II protein were pooled and subjected to refolding.

7.3.3.4 Folding of long F19Cou IGF-II protein

Folding of the Long F19Cou IGF-II protein was carried out as reported previously by Delaine *et al.*[23]. The gel filtration pool (refer to Section 7.3.3.3) was rapidly diluted in dissolution and refold buffers (refer to Section 7.3.1.3) to give a final concentration of 2.5 M urea, 0.7 M Tris, 12.5 mM glycine, 2 mM EDTA, 0.5 mM DTT at a pH of 9.1, with a maximum protein

concentration of 0.1 mg/mL. 2-hydroxyethyl disulfide (2-HED) was added at a final concentration of 1.25 mM. The refold mixture was stirred slowly at rt for 120 min and the reaction was monitored by analytical RP-HPLC on an Agilent 1100 series.

7.3.3.5 Liberation of the F19Cou IGF-II protein (3.1) by α-Lytic protease cleavage

The F19Cou IGF-II protein (**3.1**) was liberated from its porcine growth hormone fusion partner ([Met1-pGH(1-11)-PAPM]) by a mutant of the α-lytic protease called Prag A9.[172] This enzyme cleaves between the methionine of the linker PAPM motif and alanine which is the first residue of the IGF-II protein. The Prag A9 enzyme was added at a ratio of 1:330 (enzyme: total protein) to the refolding solution (refer to Section 7.3.3.4) and the cleavage reaction was incubated at 37 °C for 2 h. The cleavage reaction was monitored by analytical RP-HPLC and was stopped by acidification to pH 2.5 using conc. HCl.

The acidified cleavage reaction was filtered through a 0.22 μm filter and purified by semi-preparative HPLC. Fractions were analysed by analytical RP-HPLC and fractions containing pure F19Cou IGF-II protein (**3.1**) were pooled, lyophilised and quantified according to General Procedure XI.

7.3.4 Competition binding assays

7.3.4.1 Materials

Long^{1M}Arg^{3}IGF-I and human IGF-II were purchased from Novozymes *GroPep* Pty Ltd. (Adelaide, South Australia). Greiner Lumitrac 600 96-well plates were purchased from Omega scientific (Tarzana, USA). The DELFIA® europium-labeling kit and DELFIA® enhancement solution were purchased from PerkinElmer Life Sciences. EuIGF-II was produced as described by Denley *et al*[22] according to the manufacturer's instructions. Antibodies 83-7 and 24-31 were kind gifts from Prof. K. Siddle (Cambridge, UK). P6 IGF-1R cells (BALB/c3T3 cells overexpressing the human IGF-1R)[253] were a kind gift from Prof. R. Baserga (Philadelphia,

PA). IGF-1R-negative cells overexpressing the IR-A (R⁻IR-A) were generated as described by Denley *et al.*[22]

7.3.4.2 Competition binding assay buffers

Lysis Buffer	20 mM HEPES, 150 mM NaCl, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 1% (v/v) Triton X-100, 1 mM EGTA, and 1 mM phenylmethylsulfonyl fluoride, pH 7.5
TBST Buffer	20 mM Tris, 150 mM NaCl, and 0.1% (v/v) Tween 20

7.3.4.3 Competition binding assays

Receptor binding was measured essentially as described by Denley *et al.*[22]. Briefly, R⁻IR-A, or P6 IGF-1R cells were serum-starved for 4 h then lysed in lysis buffer (refer to Section 7.3.1.3) for 1 h at 4 °C. Lysates were centrifuged for 10 min at 3,500 rpm, then an aliquot (100 µL) was added per well to a white Greiner Lumitrac 600 96-well plate previously coated with anti-IR antibody 83-7 [254] or anti-IGF-1R antibody 24-31 [255] as appropriate. Approximately 50,000 fluorescent counts of Europium labelled IGF-II (EuIGF-II) were added to each well along with increasing concentrations of unlabelled competitor and incubated for 16 h at 4 °C. The next day wells were washed four times with TBST buffer (refer to Section 7.3.4.2), then twice with water, and then DELFIA enhancement solution (100 µL/well) was added and incubated at rt for 10 min. Time-resolved fluorescence was measured using 340 nm excitation and 612 nm emission filters with a Perkin Elmer VICTOR X5 Multilabel Plate Reader (Waltham, MA, U.S.A). IC_{50} values were calculated, using GraphPad Prism 6.01, by curve-fitting with a one-site competition model. The baseline used to calculate all IC_{50} values was set at the % bound/total value of the highest competing IGF-II concentration. Assays were performed in triplicate at least three times, unless otherwise stated.

7.4 Experimental described in Chapter 4

7.4.1 Experimental materials

Solvents and reagents were of reagent grade or higher and used as supplied unless otherwise stated. Solvents used for HPLC were of HPLC grade and used without further purification. 9-Fluorenylmethoxycarbonyl (*N*α-Fmoc) protected amino acids (Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asp(*t*Bu)-OH, Fmoc-Glu(*t*Bu)-OH, Fmoc-Gly-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Thr(*t*Bu)-OH, Fmoc-Tyr(*t*Bu)-OH, Fmoc-Val-OH), 2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU), *N*-Hydroxybenzotriazole (HOBt) were purchased from GL Biochem (Shanghai, China). *N*α-Fmoc protected amino acids (Fmoc-Arg(Pbf)-OH, Fmoc-Cys(trt)-OH, Fmoc-Cys(Acm)-OH Fmoc-Gln(trt)-OH, Fmoc-Ser(*t*Bu)-OH) were purchased from AusPep (Tullamarine, Victoria, Australia). *N*α-Fmoc protected pseudoproline dipeptides (Fmoc-Asp(O*t*Bu)-Thr($\Psi^{Me,Me}$Pro)-OH, Fmoc-Glu(O*t*Bu)-Thr($\Psi^{Me,Me}$Pro)-OH), and Fmoc-(Dmb)Gly-OH were purchased from NovaBioChem (Darmstadt, Germany). Dimethylformamide (DMF) (AR grade), *N*-methylpyrrolidine (NMP), piperidine (Reagent Grade) were purchased from Merck (Darmstadt, Germany). Diisopropylethylamine (DIPEA) and trifluoroacetic acid (TFA) were purchased from Alfa Aesar (Heysham Lancashire, UK). 3,6-dioxa-1,8-octanedithiol (DODT), triisopropylsilane (TIPS) were purchased from Sigma Aldrich (Castle Hill, NSW, Australia). Fmoc-PAL-PEG-PS resin (5-(4-*N*-Fmoc-aminomethyl-3,5-dimethoxyphenoxy)-pentanoic acid resin) (0.24 mmol/g) was purchased from Applied Biosystems (Foster City, CA, USA). acetonitrile (HPLC grade) was purchased from Scharlau (Sentmenat, Barcelona, Spain). Diethyl ether was purchased from ChemSupply (Gillman, SA, Australia).

Analytical RP-HPLC was conducted on a HP 1100 series HPLC equipped with a diode array detector and an Agilent Eclipse C18 (5μm, 4.6 x 150mm) column, or an Agilent 1200 Series HPLC equipped with a diode array detector and an Supelco Discovery Biowide Pore C5 column (5μm, 4.6 x 250mm). The product was eluted using a linear gradient of 0-70% D over 30 min and a flow rate of 1 mL/min, where the solvent system used was: A (MilliQ water +

0.1% TFA) and D (acetonitrile + 0.08% TFA). Elution of the peptide was monitored at 220nm, with the use of additional wavelengths (215 nm, 280 nm and 320 nm) as required.

Semi-preparative RP-HPLC was carried out on a HP 1100 series HPLC equipped with a diode array detector using a Supelco Discovery C18 column (5 μm, 250 mm x 10 mm). The product was eluted using an appropriate linear gradient with a flow rate of 5 mL/min, where the solvent system used was: A (MilliQ water + 0.01% TFA) and D (acetonitrile + 0.08% TFA). Gradient systems were adjusted according to the elution profiles and peak profiles obtained from the analytical RP-HPLC chromatograms. Fractions were analysed by ESI-MS and analytical RP-HPLC. Pure fractions were pooled, and lyophilised.

IGF-II analogues were quantified by comparing analytical RP-HPLC C4 profiles with profiles of standard Long^{3}Arg^{3}IGF-I preparations, on an Agilent 1100 series RP-HPLC using the procedure described by Denley *et al.*[22]

7.4.2 Methods

General Procedure XII: Automated peptide synthesis

Where stated, automated peptide synthesis was carried out on a Liberty Microwave Peptide Synthesiser (CEM Corporation) using the Fmoc/*t*Bu strategy [192,256]. Deprotection: The *N*α-Fmoc group was deprotected during each cycle by treatment of the resin with 20% v/v piperidine in DMF containing HOBt (0.1 M) for 30 s, followed by a second deprotection for 3 min. A maximum microwave power of 60 W and the maximum temperature was set to 50 °C for both deprotections. Amino acid couplings: The *N*α-Fmoc protected amino acid in DMF (5 equiv., 0.2 M), HATU in DMF (4.5 equiv., 0.45 M) and DIPEA in NMP (10 equiv., 2 M) were added to the resin and subjected to 25 W microwave irradiation for 6 min. Each coupling was performed twice with a maximum temperature set to 50 °C for both couplings. The exception was Fmoc-Arg(Pbf)-OH, which was coupled for 25 min at rt, followed by coupling for 5 min at 25 W microwave irradiation and the maximum temperature was set to 50 °C for both couplings.

General Procedure XIII: Manual peptide synthesis

Where stated, peptide elongation was performed manually on Fmoc-PAL-PEG-PS resin on a 0.05 mmol scale, using the Fmoc/*t*Bu strategy[192,256]. Deprotection: The *N*α-Fmoc group was removed by treatment of the resin with a solution of 20% v/v piperidine in DMF containing HOBt (0.1 M) for 3 min followed by a further treatment for 10 min. Amino acid coupling: The *N*α-Fmoc protected amino acids (10 equiv.) was pre-activated with a solution of HATU in DMF (10 equiv., 0.45 M) and DIPEA (neat) (20 equiv.) for 2 min at rt. The activated solution was added to the resin and the mixture was left for 20 min at rt. The solution was drained and the resin was washed with DMF (3 x 5 mL), DCM (3 x 5 mL) and DMF (3 x 5 mL). Unusual amino acid couplings: *N*α-Fmoc protected pseudoproline derivatives (Fmoc-Asp(O*t*Bu)-Thr($\Psi^{Me,Me}$Pro)-OH, Fmoc-Glu(O*t*Bu)-Thr($\Psi^{Me,Me}$Pro)-OH), Fmoc-(Dmb)Gly-OH and *N*α-Fmoc protected coumaryl amino acid (**2.5**) were coupled as detailed below. *N*α-Fmoc protected amino acid (1.5 equiv.) was pre-activated with a solution of HATU in DMF (1.4 equiv,) and DIPEA (neat) (5 equiv.) for 2 min at rt. The activated solution was added to the resin and the mixture was left for 16 h at rt. The solution was drained and the resin was washed with DMF (3 x 5 mL), DCM (3 x 5 mL) and DMF (3 x 5 mL).Coupling reactions were monitored *via* the TNBSA test[207] for primary amines and where applicable the acetaldehyde/chloranil test[257,258] for secondary amines was used.

General Procedure XIV: Cleavage of peptide from analytical resin samples

A cleavage solution (500 μL) comprised of TFA: TIPS: DODT: water (92.5: 2.5: 2.5: 2.5 v/v/v/v) was added to the resin-bound peptide (approx. 10 mg) and shaken at rt for 2-3 h. The resin beads were filtered and the filtrate was concentrated to approximately 100 μl under a stream of nitrogen. The crude peptide was precipitated by addition of cold diethyl ether (1 mL), and then isolated by centrifugation. The solid material was washed with cold diethyl ether (3 x 1 mL), dissolved in a 30% aqueous acetonitrile solution containing 0.1% TFA and lyophilised

General Procedure XV: Release of the final peptide from the resin

A cleavage solution comprised of TFA: TIPS: DODT: water (92.5: 2.5: 2.5: 2.5 v/v/v/v) was added to the resin-bound peptide (10 mL/250 mg) and shaken at rt for 3 h. The resin beads were filtered and the filtrate was concentrated to approximately 1 mL under a stream of nitrogen. The crude peptide was precipitated by addition of cold diethyl ether (40 mL), and isolated by centrifugation. The solid material was washed with cold diethyl ether (3 x 40 mL), dissolved in a 30% aqueous acetonitrile solution containing 0.1% TFA and lyophilised.

General Procedure XVI: Trinitrobenzenesulfonic acid (TNBSA) test[207]

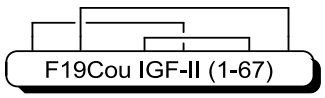
To a resin-bound peptide sample (10-20 beads) 25 µL of a 5% (v/v) solution of DIPEA in DMF was added followed by the addition of 25 µL of a 1% (w/v) solution of TNSBA in DMF. Where the test was positive, the *N*α-Fmoc protected amino acid was coupled again under the same conditions.

General Procedure XVII: Acetaldehyde/chloranil test [257,258]

To a resin-bound peptide sample (10-20 beads) 25 µL of a 2% (v/v) solution of acetaldehyde in DMF was added followed by the addition of 25 µL of a 2% (w/v) solution of chloranil in DMF. Where the test was positive, the *N*α-Fmoc protected amino acid was coupled again under the same conditions. Solutions were stored in the fridge and were kept for a maximum of 1 week.

7.4.3 Experimental procedures

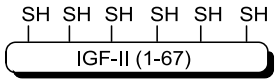
F19Cou IGF-II protein (4.2)



Folding of F19Cou IGF-II peptide (**4.7**) was carried out as reported previously by Delaine *et al.*[23]. The purified peptide (5.0 mg, 0.66 µmol) was dissolved in buffer (4 mg/mL) consisting of 8 M urea, 0.1 M Tris and 40 mM glycine containing 20 mM dithiothreitol (DTT). The

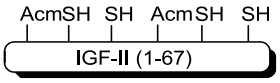
solution was gently stirred at rt for 1 h and monitored by analytical RP-HPLC. After 1 h, the solution was rapidly diluted with refold buffer (refer to Section 7.3.1.3) to give a solution with a final concentration of 2.5 M urea, 0.7 M Tris, 12.5 mM glycine, 2 mM EDTA, 0.5 mM DTT at a pH of 9.1 and a maximum protein concentration of 0.1 mg/mL. 2-hydroxyethyl disulfide (2-HED) was then added at a final concentration of 1.25 mM. The refold mixture was stirred slowly at rt and the reaction was monitored by analytical RP-HPLC and was determined complete when no more of the reduced F19Cou IGF-II peptide (**4.7**) remained. The reaction mixture was acidified to a pH of 3 using conc. HCl, filtered through a 0.45 μM syringe filter and purified by semi-preparative RP-HPLC, using a linear gradient of 0-40% acetonitrile over 40 min with a flow rate of 5 mL/min, to give the F19Cou IGF-II protein (**4.2**) (0.266 mg, 5%) as a white solid. HRMS (ESI+) *m/z* calculated for $C_{325}H_{502}N_{94}O_{103}S_6$: 7566.4445 (average isotopes); observed: *m/z* 1514.35 ($[M+5H]^{+5}$), 1261.93 ($[M+6H]^{+6}$), 1081.94 ($[M+7H]^{+7}$), 946.83 ($[M+8H]^{+8}$) and 841.62 ($[M+9H]^{+9}$).

IGF-II (1-67) peptide (4.7)



The IGF-II (31-67; 4 Trt) resin-bound peptide (**4.4**) was synthesised from Fmoc-PAL-PEG-PS resin (**4.3**) on a 0.05 mmol scale according to General Procedure XII. The resin-bound peptide (**4.4**) was then elongated according to the manual SPPS procedure detailed in General Procedure XIII, to give resin-bound peptide **4.6**. Peptide **4.6** was cleaved from the resin according to General Procedure XV. The crude peptide was purified by semi-preparative RP-HPLC using a linear gradient of 0-50% acetonitrile over 50 min, however IGF-II (1-67) peptide (**4.7**) was not afforded.

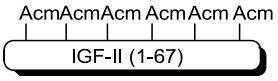
IGF-II (1-67) (2 Acm) peptide (4.11)



The IGF-II (31-67; Acm, 3 Trt) resin-bound peptide (**4.8**) was synthesised from Fmoc-PAL-PEG-PS resin (**4.3**) on a 0.05 mmol scale according to General Procedure XII. The resin-

bound peptide (**4.8**) was elongated according to the manual SPPS procedure detailed in General Procedure XIII, to give resin-bound peptide **4.10**. Peptide **4.10** was cleaved from the resin according to General Procedure XV. The crude peptide was purified by semi-preparative RP-HPLC using a linear gradient of 0-50% acetonitrile over 50 min to afford trace quantities of the IGF-II (1-67) (2 Acm) peptide (**4.11**) (< 1 mg) as a white solid. MS (ESI+) *m/z* calculated for $C_{327}H_{516}N_{96}O_{102}S_6$: 7617.5912 (average isotopes); observed: *m/z* 1904.72 ($[M+4H]^{+4}$), 1523.44 ($[M+5H]^{+5}$), 1270.00 ($[M+6H]^{+6}$) and 1088.67 ($[M+7H]^{+7}$).

IGF-II (1-67) (6 Acm) peptide (4.15)



Method A

The IGF-II (31-67; 4 Acm) resin-bound peptide (**4.3**) was synthesised from Fmoc-PAL-PEG-PS resin (**4.3**) on a 0.05 mmol scale according to General Procedure XII. The resin-bound peptide (**4.12**) was elongated according to the manual SPPS procedure detailed in General Procedure XIII, to give resin-bound peptide **4.14**. Peptide **4.14** was cleaved from the resin according to General Procedure XV. The crude peptide was purified by semi-preparative RP-HPLC using a linear gradient of 0-50% acetonitrile over 50 min to afford trace quantities of the IGF-II (1-67) (6 Acm) peptide (**4.15**) (< 1 mg) as a white solid. MS (ESI+) *m/z* calculated for $C_{339}H_{536}N_{100}O_{106}S_6$: 7900.9035 (average isotopes); observed: *m/z* 1976.19 ($[M+4H]^{+4}$), 1581.11 ($[M+5H]^{+5}$), 1371.87 ($[M+6H]^{+6}$) and 1129.73 ($[M+7H]^{+7}$)

Method B

The IGF-II (1-67; 6Acm; ΨT7, ΨT16, ΨT58) resin-bound peptide (**4.16**) was synthesised from Fmoc-PAL-PEG-PS resin (**4.3**) on a 0.025 mmol scale according to the manual SPPS procedure detailed in General Procedure XIII. Peptide **4.16** was cleaved from the resin according to General Procedure XV and the resultant crude peptide was analysed using analytical RP-HPLC and ESI-MS. The IGF-II (1-67) (6 Acm) peptide (**4.15**) was not detected.

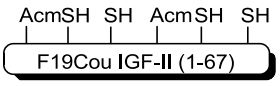
Method C

The IGF-II (1-67; 6Acm; DmbG11, DmbG22, DmbG25, DmbG41) resin-bound peptide (**4.17**) was synthesised from Fmoc-PAL-PEG-PS resin (**4.3**) on a 0.025 mmol scale according to the manual SPPS procedure detailed in General Procedure XIII. Peptide **4.17** was cleaved from the resin according to General Procedure XV and the resultant crude peptide was analysed using analytical RP-HPLC, ESI-MS and LCMS. LCMS (ESI+) *m/z* calculated for $C_{339}H_{536}N_{100}O_{106}S_6$: 7900.9035 (average isotopes); observed: *m/z* 1371.6 ($[M+6H]^{+6}$), 1129.4 ($[M+7H]^{+7}$), 988.4 ($[M+8H]^{+8}$) and 878.9 ($[M+9H]^{+9}$).

Method D

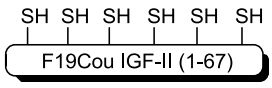
The IGF-II (1-67; 6Acm; ΨT7, ΨT16, ΨT58, DmbG11, DmbG22, DmbG25, DmbG41) resin-bound peptide (**4.18**) was synthesised from Fmoc-PAL-PEG-PS resin (**4.3**) on a 0.025 mmol scale according to the manual SPPS procedure detailed in General Procedure XIII. Peptide **4.18** was cleaved from the resin according to General Procedure XV and the resultant crude peptide was analysed using analytical RP-HPLC, ESI-MS and LCMS. LCMS (ESI+) *m/z* calculated for $C_{339}H_{536}N_{100}O_{106}S_6$: 7900.9035 (average isotopes); observed: *m/z* 1371.6 ($[M+6H]^{+6}$), 1129.4 ($[M+7H]^{+7}$), 988.4 ($[M+8H]^{+8}$) and 878.6 ($[M+9H]^{+9}$).

F19Cou IGF-II (1-67) (2 Acm) peptide (4.22)



The IGF-II (31-67; Acm, 3 Trt) resin-bound peptide (**4.8**) was synthesised from Fmoc-PAL-PEG-PS resin (**4.3**) on a 0.025 mmol scale according to General Procedure XII. The resin-bound peptide (**4.8**) was elongated according to the manual SPPS procedure detailed in General Procedure XIII, to give resin-bound peptide **4.21**. Peptide **4.21** was cleaved from the resin according to General Procedure XV. The crude peptide was purified by semi-preparative RP-HPLC using a linear gradient of 0-40% acetonitrile over 40 min to afford the F19Cou IGF-II (1-67) (2 Acm) peptide (**4.22**) (5.5 mg, 14%) as a white solid.

F19Cou IGF-II (1-67) peptide (4.23)



The purified F19Cou IGF-II (2 Acm) peptide (**4.22**) (5.5 mg, 0.73 μmol) was dissolved in 80% aqueous acetic acid solution (6.8 mL). A 20 mM solution of iodine in 80% aqueous acetic acid (2.65 mL) was added, followed by a 60 mM aqueous solution of HCl (530 μL) to give a final peptide concentration of approximately 0.5 mg/mL. Once the reaction was determined to be complete by analytical RP-HPLC and ESI-MS (typically 30-60 min), the reaction was quenched by slowly adding Dowex 1X8 200 until the deep red/orange colour of the iodine had dissipated. The solution was filtered and the resin washed with a small volume of 80% aqueous acetic acid. The solution was then diluted with water to reduce the acetic acid concentration to 20% and filtered through a 0.45 μM syringe filter before purification by semi-preparative HPLC using a linear gradient of 0-40% acetonitrile over 40 min to give the F19Cou IGF-II peptide (5.0 mg, 90%) as a white solid.

7.4.4 Competition binding assays

Competition binding assays for F19Cou IGF-II protein (**4.2**) were carried out as previously detailed in Section 7.3.4.

7.5 Experimental described in Chapter 5

7.5.1 Experimental materials

Solvents and reagents were of reagent grade or higher and used as supplied unless otherwise stated. Solvents used for HPLC were of HPLC grade and used without further purification. 9-Fluorenylmethoxycarbonyl (Fmoc) protected L-amino acids (Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asp(O*t*Bu)-OH, Fmoc-Cys(trt)-OH, Fmoc-Glu(O*t*Bu)-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Ser(*t*Bu)-OH), Fmoc-Thr(*t*Bu)-OH, Fmoc-Tyr(*t*Bu)-OH) were purchased from CEM Corporation or GL Biochem (Shanghai, China). 2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU), *N*-Hydroxybenzotriazole hydrate (HOBt•H_2O) were purchased from Advanced Chemtech (Louisville, KY). Diisopropylcarbodiimide (DIC) and 3-(tritylthio)propionic acid were purchased from GL Biochem (Shanghai, China).

*N*α-Boc-protected L-amino acids (Boc-Ala-OH, Boc-Asp(OcHex)-OH, Boc-Arg(Tos)-OH, Boc-Gly-OH, Boc-Ile-OH, Boc-Leu-OH, Boc-Phe-OH, Boc-Ser(Bzl)-OH, Boc-Thr(Bzl)-OH, Boc-Tyr(Br-Bzl)-OH, Boc-Val-OH) and Boc-L-Ala-OCH_2-Phi-CH_2-COOH (Boc-Ala-PAM-COOH) linker were purchased from Polypeptide Group (Strasbourg, France) Boc protected L-amino acids (Boc-Pro-OH and Boc-Gln-OH) were purchased from AusPep (Tullamarine, Victoria, Australia). Boc protected L-amino acids (Boc-Cys(4-MeBzl)-OH, Boc-Glu(OcHex)-OH, Boc-Thz-OH) and Fmoc-PAL-tentagel resin (0.21 mmol/g) were purchased from ChemImpex (Illinois, U.S.A). Aminomethyl polystyrene resin was made 'in house' according to the procedure described by Harris *et al.*[221,222] Trifluoroacetic acid (TFA) was purchased from Halocarbon (New Jersey). Diisopropylethylamine (DIPEA), *N*-methylpyrrolidine (NMP), piperidine, 3,6-dioxa-1,8-octanedithiol (DODT), triisopropylsilane (TIPS), tris(2-carboxyethyl)phosphine hydrochloride (TCEP), mercaptopropanoic acid (MPAA), methoxyamine hydrochloride ($MeONH_2$•HCl) were purchased from Sigma Aldrich. Dimethylformamide (DMF) (AR grade), acetonitrile (HPLC grade), guanidine hydrochloride

(GnHCl) were purchased from Scharlau (Sentmenat, Barcelona, Spain). Diethyl ether and disodium hydrogen phosphate (Na_2HPO_4) were purchased from Ajax chemicals. HF gas was purchased from Matheson Tri-Gas (Basking Ridge, NJ). Buffers were degassed by sparging with Ar for 30 min prior to use. pH measurements were carried out using an ISFET pocket pH meter with an accuracy of ± 0.01 pH units.

Analytical RP-HPLC was conducted on an Agilent 1200 Series HPLC equipped with a diode array detector and an Supelco Discovery Biowide Pore C5 column (5μm, 4.6 x 250mm). The product was eluted using a linear gradient of 0-70% D over 30 min and a flow rate of 1 mL/min, where the solvent system used was: A (MilliQ water + 0.1% TFA) and D (acetonitrile + 0.08% TFA). Elution of the peptide was monitored at 220nm, with the use of additional wavelengths (215 nm, 280 nm and 320 nm) as required.

Semi-preparative RP-HPLC was carried out on a Dionex Ultimate U3000 system equipped with a multiple wavelength detector and using a Phenomenex Gemini C18 column (5 μm, 250 mm x 10 mm) or on a HP 1100 series HPLC equipped with a diode array detector using a Supelco Discovery C18 column (5 μm, 250 mm x 10 mm). The product was eluted using a linear gradient of 1% D/min, with a flow rate of 5 mL/min, where the solvent system used was: A (MilliQ water + 0.01% TFA) and D (acetonitrile + 0.08% TFA). Typically, a 0-40% acetonitrile over 40 min gradient was used; however gradient systems were adjusted according to the elution profiles and peak profiles obtained from the analytical RP-HPLC chromatograms. Fractions were analysed by LCMS. Pure fractions were pooled, and lyophilised.

IGF-II analogues were quantified by comparing analytical RP-HPLC C4 profiles with profiles of standard Long$^{™}$Arg^{3}IGF-I preparations, on an Agilent 1100 series RP-HPLC using the procedure described by Denley *et al.*[22]

7.5.2 Methods

General Procedure XVIII: Automated peptide synthesis

Where stated, automated peptide synthesis was carried out on a Liberty Microwave Peptide Synthesiser (CEM Corporation) using the Fmoc/*t*Bu strategy [192,256]. Deprotection: The *N*α-Fmoc group was deprotected during each cycle by treatment of the resin with 20% v/v piperidine in DMF containing HOBt (0.1 M) for 30 s, followed by a second deprotection for 3 min. A maximum microwave power of 60 W and the maximum temperature was set to 50 °C for both deprotections. Amino acid couplings: The *N*α-Fmoc protected amino acid in DMF (5 equiv., 0.2 M), HATU in DMF (4.5 equiv., 0.45 M) and DIPEA in NMP (10 equiv., 2 M) were added to the resin and subjected to 25 W microwave irradiation for 6 min. Each coupling was performed twice with a maximum temperature set to 50 °C for both couplings. The exception was Fmoc-Arg(Pbf)-OH, which was coupled for 25 min at rt, followed by coupling for 5 min at 25 W microwave irradiation and the maximum temperature was set to 50 °C for both couplings.

General Procedure XIX: Manual peptide synthesis:-*in situ* neutralisation Boc SPPS [223]

Where stated, peptide synthesis was performed manually using the *in situ* neutralisation Boc SPPS methodology.[223] Deprotection: The *N*α-Boc group was removed by treatment with TFA (neat) (2 x 1 min) followed by flow washing of the resin with DMF for 30 s. Amino acid coupling: The *N*α-Boc protected amino acid (4 equiv.) was pre-activated with a solution of HBTU in DMF (3.8 equiv., 0.38 M) and DIPEA (neat) (20 equiv.) for 2 min at rt. The activated solution was added to the resin and the mixture was shaken for 10 min at rt. The solution was drained and the resin was flow washed with DMF for 30 s. Unusual amino acid couplings: The mixture of *N*α-Boc protected coumaryl amino acids (**2.16** and **2.17**) (1.2 equiv.) was pre-activated with a solution of HBTU in DMF (1.1 equiv., 0.4 M) and DIPEA (neat) (5 equiv.) for 2 min at rt. The activated solution was added to the resin and the mixture was shaken at rt for 18 h. Coupling reactions were monitored *via* the Kaiser test[259] for primary amines, and where applicable the acetaldehyde/chloranil test[257,258] for secondary amines was used.

General Procedure XX: Kaiser test[259]

To a resin-bound peptide sample (10-20 beads) 2-3 drops of each solution (A-C) were added (solution A: 5% (w/v) solution of ninhydrin in ethanol; solution B: phenol (80g) in ethanol (20 mL); and solution C: 0.001 M aqueous KCN (2 mL) in pyridine (98 ml)). The mixture was heated at 100 °C for 5 min. Where the test was positive, the *N*α-protected amino acid was coupled again under the same conditions.

General Procedure XXI: Acetaldehyde/chloranil test [257,258]

To a resin-bound peptide sample (10-20 beads) 25 µL of a 2% (v/v) solution of acetaldehyde in DMF was added followed by the addition of 25 µL of a 2% (w/v) solution of chloranil in DMF. Where the test was positive, the *N*α-protected amino acid was coupled again under the same conditions. Solutions were stored in the fridge and were kept for a maximum of 1 week.

General Procedure XXII: TFA cleavage

A cleavage solution comprised of TFA: TIPS: DODT: water (92.5: 2.5: 2.5: 2.5 v/v/v/v) was added to the resin-bound peptide (10 mL/250 mg resin) and the mixture was shaken at rt for 3 h. The resin beads were filtered and the filtrate was concentrated to approximately 2 mL under a stream of nitrogen. The crude peptide was precipitated by addition of cold diethyl ether (40 mL) and isolated by centrifugation. The solid material was washed with cold diethyl ether (3 x 40 mL), dissolved in a 30% aqueous acetonitrile solution containing 0.1 % TFA and lyophilised.

General Procedure XXIII: HF cleavage

Immediately prior to cleavage the resin the *N*α-Boc group was deprotected by treatment with TFA (neat) for 1 min followed by another treatment for another 1 min. The resin was flow washed with DMF (30 s), DCM (30 s), diethyl ether (30 s) and dried under vacuum for 30 min before being subjected to HF cleavage. The resin was cleaved using anhydrous HF and *p*-cresol (9:1 v/v) with stirring at 0 °C for 1 h. The HF was evaporated and the crude peptide was precipitated by addition of cold diethyl ether (40 mL) and isolated by centrifugation. The solid

material was washed with cold diethyl ether (3 x 40 mL), dissolved in 30% aqueous acetonitrile solution containing 0.1% TFA and lyophilised.

General Procedure XXIV: Solid phase extraction (SPE)

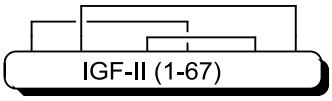
The crude peptide mixture was isolated from the ligation mixture using a solid phase extraction (SPE) cartridge (Alltech, C18LP, 600 mg). The cartridge was prepared by washing with MeOH (5 mL) and equilibrated with 5% aqueous acetonitrile containing 0.1% TFA (10 mL). The ligation mixture was diluted with water and acidified if necessary to pH 3 with TFA and loaded onto the prepared cartridge at a rate of approximately 2 mL/min. Salts and non-binding contaminants were eluted with 5% aqueous acetonitrile containing 0.1% TFA (10 mL) and the crude peptide was eluted by washing the cartridge with 50% aqueous acetonitrile containing 0.1% TFA (10 mL). Fractions containing the desired peptides were subsequently lyophilised.

General Procedure XXV: Oxidative folding

Oxidative folding of IGF-II peptides were carried out as described by Delaine *et al.*[23]. Purified peptide (4 mg/mL) was dissolved in a buffer consisting of 8 M urea, 0.1 M Tris and 40 mM glycine containing 20 mM dithiothreitol (DTT). The solution was incubated at rt for 1 h, with reduction of the peptide being monitored by analytical RP-HPLC. After 1 h The reduced solution was rapidly diluted with refolding buffer (refer to Section 7.3.1.3) to give a solution with a final concentration of 2.5 M urea, 0.7 M Tris, 12.5 mM glycine, 2 mM EDTA, 0.5 mM DTT, 1.25 mM 2-hydroxyethyldisulfide (2-HED) and a final protein concentration of 0.1 mg/mL This reaction mixture was slowly stirred at rt and monitored by analytical RP-HPLC. The reaction was determined complete when no more of the reduced peptide remained, typically 120 min. The reaction mixture was quenched by acidification to a pH of 3 using conc. HCl. The acidified solution was filtered through a 0.45 μM syringe filter and purified by semi-preparative RP-HPLC on an HP 1100 series HPLC.

7.5.3 Experimental procedures

Native IGF-II protein (4.1)



Method A: Two fragment

C-terminal fragment **5.10** (3.42 mg, 1.3 μmol, 3.5 mM), N-terminal thioester **5.11** (6.75 mg, 1.5 μmol, 3 mM), TCEP (0.02 M) and MPAA (0.2 M) were dissolved in degassed native chemical ligation buffer (6 M GnHCl and 0.2 M Na_2HPO_4, 423 μL). The pH of the solution was adjusted to 6.9 using aqueous NaOH (10 M and 2 M), the reaction mixture was sparged with Ar (10 s) and shaken at rt for 1 h. The products were isolated using SPE according to General Procedure XXIV and purified by semi-preparative RP-HPLC, using a linear gradient of 0-40% acetonitrile over 40 min to give the native IGF-II peptide (**4.7**) (0.4 mg, 4%) as a white solid. Peptide **4.7** (0.06 mg, 8 nmol) was subsequently folded according to General Procedure XXV to give the native IGF-II protein (**4.1**) (2.7 μg, 5%) as a white solid. HRMS (ESI+) *m/z* calculated for $C_{321}H_{500}N_{94}O_{100}S_6$: 7468.3875 (average isotopes); observed: 1494.74 ($[M+5H]^{+5}$), 1245.79 ($[M+6H]^{+6}$) and 1067.93 ($[M+7H]^{+7}$).

Method B: Three fragment

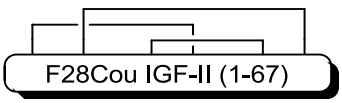
C-terminal fragment **5.10** (3.8 mg, 1.6 μmol, 3.35 mM), thioester **5.16** (4.6 mg, 1.4 μmol, 3 mM), TCEP (0.02 M) and MPAA (0.2 M) were dissolved in degassed native chemical ligation buffer (6 M GnHCl and 0.2 M Na_2HPO_4, 487 μL), pH of the solution was adjusted to 6.9 using aqueous NaOH (10 M and 2 M) and shaken at rt for 1 h. $MeONH_2$•HCl (8.1 mg, 97 μmol, 0.2 M) was then added, the pH of the solution was adjusted to 3.9 with aqueous HCl (5 M), the reaction mixture was shaken at rt for 6 h. The pH of the solution was adjusted to 6.8 with aqueous NaOH (10 M and 2 M) and the N-terminal thioester **5.14** (3.9 mg, 1.6 μmol, 3.37 mM) was added. The reaction mixture was shaken at rt for 47 h. The products were isolated using SPE according to General Procedure XXIV and purified by semi-preparative RP-HPLC to give the native IGF-II peptide (**4.7**) (0.44 mg, 4%) as a white solid. Peptide **4.7** (0.44 mg,

0.058 μmol) was subsequently folded according to General Procedure XXV to give the native IGF-II protein (**4.1**) (25 μg, 6%) as a white solid. HRMS (ESI+) *m/z* calculated for $C_{321}H_{500}N_{94}O_{100}S_6$: 7468.3875 (average isotopes); observed: 1494.7144 ($[M+5H]^{+5}$), 1245.7516 ($[M+6H]^{+6}$) and 1067.9352 ($[M+7H]^{+7}$).

F19Cou IGF-II protein (4.2)

C-terminal fragment **5.10** (3.15 mg, 1.36 μmol, 3.25 mM), thioester **5.16** (3.96 mg, 1.25 μmol, 3 mM), TCEP (0.02 M) and MPAA (0.2 M) were dissolved in degassed native chemical ligation buffer (6 M GnHCl and 0.2 M Na_2HPO_4, 418 μL), pH of the solution was adjusted to 6.9 using aqueous NaOH (10 M and 2 M) and shaken at rt for 1 h. $MeONH_2$•HCl (7 mg, 84 μmol, 0.2 M) was then added, the pH of the solution was adjusted to 4.0 with aqueous HCl (5 M), the reaction mixture was shaken at rt 16 h. The pH of the solution was adjusted to 7.0 using aqueous NaOH (10 M and 2 M) and the N-terminal thioester **5.15** (3.47 mg, 1.4 μmol, 3.38 mM) was added. The reaction mixture was shaken at rt for 27 h. The products were isolated using SPE according to General Procedure XXIV and purified by semi-preparative RP-HPLC to give the F19CouIGF-II peptide (**4.23**) (0.84 mg, 9%) as a white solid. Peptide **4.23** (0.84 mg, 0.11 μmol) was subsequently folded according to General Procedure XXV to give the F19Cou IGF-II protein (**4.2**) (89 μg, 11%) as a white solid. HRMS (ESI+) *m/z* calculated for $C_{323}H_{500}N_{94}O_{103}S_6$: 7566.4445 (average isotopes); observed: 1514.1628 ($[M+5H]^{+5}$), 1261.9639 ($[M+6H]^{+6}$), 1081.8225 ($[M+7H]^{+7}$) and 946.7245 ($[M+8H]^{+8}$).

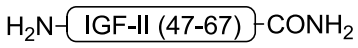
Synthetic F28Cou IGF-II protein (5.1)



C-terminal fragment **5.10** (7.09 mg, 3 μmol, 3.12 mM), thioester **5.17** (9.57 mg, 2.9 μmol, 3 mM), TCEP (0.02 M) and MPAA (0.2 M) were dissolved in degassed native chemical ligation buffer (6 M GnHCl and 0.2 M Na_2HPO_4, 976 μL), pH of the solution was adjusted to 6.9 using

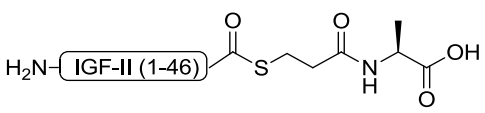
aqueous NaOH (10 M and 2 M) and shaken at rt for 1 h. $MeONH_2$•HCl (16.4 mg, 0.19 mmol, 0.2 M) was then added, the pH of the solution was adjusted to 4.0 with aqueous HCl (5 M) and the reaction mixture was shaken at rt 16 h. The pH of the solution was adjusted to 7.0 using aqueous NaOH (10 M and 2 M) and the N-terminal thioester **5.14** (7.54 mg, 3.2 µmol, 3.28 mM) was added. The reaction mixture was shaken at rt for 47 h. The products were isolated using SPE according to General Procedure XXIV and purified by semi-preparative RP-HPLC to give the F28CouIGF-II peptide (**5.9**) (0.7 mg, 3%) as a white solid. Peptide **5.9** (0.7 mg, 0.09 µmol) was subsequently folded according to General Procedure XXV to give the F28Cou IGF-II protein (**5.1**) (29 µg, 4%) as a white solid. HRMS (ESI+) *m/z* calculated for $C_{325}H_{500}N_{94}O_{103}S_6$: 7566.4445 (average isotopes); observed: 1514.2609 ($[M+5H]^{+5}$), 1261.8750 ($[M+6H]^{+6}$), 1081.7289 ($[M+7H]^{+7}$) and 946.7548 ($[M+8H]^{+8}$).

C-terminal IGF-II (47-67) fragment (5.10)



IGF-II (47-67) resin-bound peptide was synthesised on a 0.1 mmol scale from Tentgel® S RAM resin (**5.18**) according to General Procedure XVIII and the peptide was cleaved from the resin according to General Procedure XXII. The resulting crude peptide was dissolved in a 5% aqueous acetonitrile containing 0.1% TFA and purified by semi-preparative RP-HPLC to afford fragment **5.10** (30 mg, 13%) as a white solid. LCMS (ESI+) *m/z* calculated for $C_{99}H_{158}N_{26}O_{32}S_3$: 2320.6720 (average isotopes); observed: 1160.84 ($[M+2H]^{+2}$), 774.25 ($[M+3H]^{+3}$) and 581.02 ($[M+4H]^{+4}$).

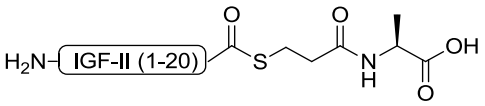
N-terminal IGF-II (1-46) thioester (5.11)



Cysteine alkyl thioester linked resin (**5.21**) was elongated according to General Procedure XIX to give the resin-bound IGF-II (1-46) thioester (**5.22**). Resin **5.22** was cleaved with HF according to General Procedure XXIII. The resulting crude peptide was dissolved in a 5% aqueous acetonitrile containing 0.1% TFA and purified by semi-preparative RP-HPLC to

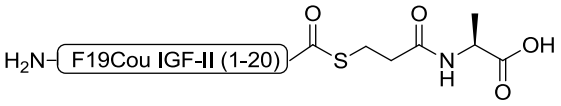
afford thioester **5.11** (12.5 mg, 7%) as a white solid. LCMS (ESI+) *m/z* calculated for $C_{228}H_{359}N_{69}O_{71}S_4$: 5330.9859 (average isotopes); observed: 1777.72 ($[M+3H]^{+3}$), 1333.47 ($[M+4H]^{+4}$), 1067.00 ($[M+5H]^{+5}$), 889.36 ($[M+6H]^{+6}$), 762.41 ($[M+7H]^{+7}$), 667.25 ($[M+8H]^{+8}$) and 593.21 ($[M+9H]^{+9}$).

IGF-II (1-20) thioester (5.14)



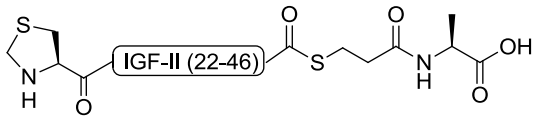
Valine alkyl thioester linked resin (**5.23**) was elongated according to General Procedure XIX to give the resin-bound IGF-II (1-20) thioester (**5.24**). Resin **5.24** was cleaved with HF according to General Procedure XXIII. The resulting crude peptide was dissolved in a 5% aqueous acetonitrile containing 0.1% TFA and purified by semi-preparative RP-HPLC to afford thioester **5.14** (41.3 mg, 17%) as a white solid. LCMS (ESI+) *m/z* calculated for $C_{103}H_{161}N_{25}O_{34}S_2$: 2357.6648 (average isotopes); observed: 1179.53 ($[M+2H]^{+2}$), 786.70 ($[M+3H]^{+3}$) and 590.26 ($[M+4H]^{+4}$).

F19Cou IGF-II (1-20) thioester (5.15)



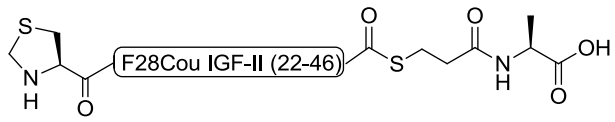
Valine alkyl thioester linked resin (**5.23**) was elongated according to General Procedure XIX to give the resin-bound F19Cou IGF-II (1-20) thioester (**5.25**). Resin **5.25** was cleaved with HF according to General Procedure XXIII. The resulting crude peptide was dissolved in a 5% aqueous acetonitrile containing 0.1% TFA and purified by semi-preparative RP-HPLC to afford thioester **5.15** (3.7 mg, 2%) as a white solid. LCMS (ESI+) *m/z* calculated for $C_{107}H_{163}N_{25}O_{37}S_2$: 2455.7218 (average isotopes); observed: 1228.48 ($[M+2H]^{+2}$), 819.36 ($[M+3H]^{+3}$) and 614.82 ($[M+4H]^{+4}$).

IGF-II (Thz-46) thioester (5.16)



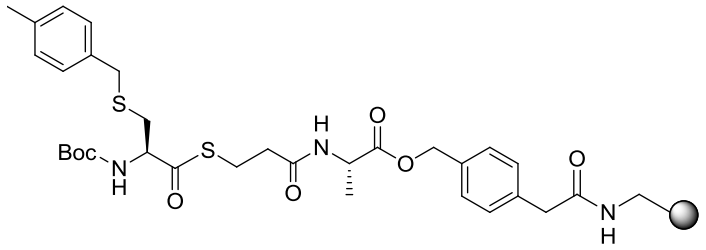
Cysteine alkyl thioester linked resin (**5.21**) was elongated according to General Procedure XIX to give the resin-bound IGF-II (Thz-46) thioester (**5.26**). Resin **5.26** was cleaved with HF according to General Procedure XXIII. The resulting crude peptide was dissolved in a 5% aqueous acetonitrile containing 0.1% TFA and purified by semi-preparative RP-HPLC to afford thioester **5.16** (12.3 mg, 11%) as a white solid. LCMS (ESI+) *m/z* calculated for $C_{136}H_{209}N_{45}O_{40}S_3$: 3155.5546 (average isotopes); observed: 1054.93 ($[M+3H]^{+3}$), 791.52 ($[M+4H]^{+4}$), 633.40 ($[M+5H]^{+5}$), 528.00 ($[M+6H]^{+6}$) and 452.75 ($[M+7H]^{+7}$).

F28Cou IGF-II (Thz-46) thioester (5.17)



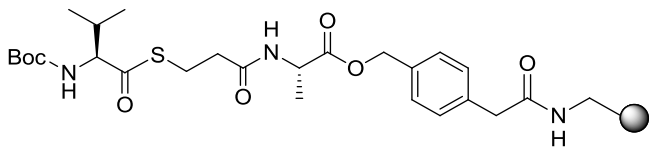
Cysteine alkyl thioester linked resin (**5.21**) was elongated according to General Procedure XIX to give the resin-bound F28Cou IGF-II (Thz-46) thioester (**5.27**). Resin **5.27** was cleaved with HF according to General Procedure XXIII. The resulting crude peptide was dissolved in a 5% aqueous acetonitrile containing 0.1% TFA and purified by semi-preparative RP-HPLC to afford thioester **5.17** (11.8 mg, 3%) as a white solid. LCMS (ESI+) *m/z* calculated for $C_{136}H_{211}N_{45}O_{43}S_3$ for: 3260.6117 (average isotopes); observed: 1630.93 ($[M+2H]^{+2}$), 1087.58 ($[M+3H]^{+3}$), 815.99 ($[M+4H]^{+4}$), 653.00 ($[M+5H]^{+5}$), 544.33 ($[M+6H]^{+6}$) and 466.73 ($[M+7H]^{+7}$).

Cysteine alkyl thioester linked resin (5.21)



Aminomethyl resin (**5.19**) (100 mg, 0.1 mmol)[221,222] was washed with DCM (3 x 5 mL), swollen in a DCM:DMF mixture (1:1) for 15 min and washed again with DCM (3 x 5 mL). A solution of Boc-L-Ala-PAM-COOH (67.8 mg, 0.2 mmol) and DIC (33 µL, 0.2 mmol) in DCM (3 mL) was added to the resin and the mixture was shaken for 1 h at rt. The solution was drained and the resin was flow washed with DCM (30 s) and DMF (30 s). The Boc group was removed by treatment of the resin with TFA (neat) (5 mL) (2 x 1 min). The resin was flow washed with DMF (30 s). A solution of S-Trityl-β-mercaptopropionic acid (191 mg, 0.55 mmol) in HBTU (0.4 M in DMF, 1.3 mL) was added to the resin followed by DIPEA (198 µL, 1.1 mmol). The resin was shaken for 20 min and then flow washed with DMF (30 s). The S-trityl group was removed by treatment of the resin **5.20** with a solution of TFA: TIPS: water (95:2.5:2.5, v/v/v) (5 mL) (2 x 1 min) and then flow washed with DMF (30 s). Thioesterification was achieved by adding a solution of Boc-Cys(4-MeBzl)-OH (180 mg, 0.55 mmol) in HBTU in DMF (0.4 M in DMF, 1.3 mL) and DIPEA (198 µL, 1.1 mmol) to the resin. The mixture was shaken for 1 h and then flow washed with DMF (30 s) to give the cysteine alkyl thioester linked resin (**5.21**).

Valine alkyl thioester linked resin (5.23)



Aminomethyl resin (**5.19**) (100 mg, 0.1 mmol)[221,222] was washed with DCM (3 x 5 mL), swollen in a DCM:DMF mixture (1:1) for 15 min and washed again with DCM (3 x 5 mL). A solution of Boc-L-Ala-PAM-COOH (71.6 mg, 0.21 mmol) and DIC (33 µL, 0.2 mmol) in DCM (3 mL) was added to the resin and the mixture was shaken for 1 h at rt. The solution

was drained and the resin was flow washed with DCM (30 s) and DMF (30 s). The Boc group was removed by treatment of the resin with TFA (neat) (5 mL) (2 x 1 min). The resin was flow washed with DMF (30 s). A solution of S-Trityl-β-mercaptopropionic acid (191 mg, 0.55 mmol) in HBTU (0.4 M in DMF, 1.3 mL) was added to the resin followed by DIPEA (198 µL, 1.1 mmol). The resin was shaken for 20 min and then flow washed with DMF (30 s). The S-trityl group was removed by treatment of resin **5.20** with a solution of TFA: TIPS: water (95:2.5:2.5, v/v/v) (5 mL) (2 x 1 min) and then flow washed with DMF (30 s). Thioesterification was achieved by adding a solution of Boc-Val-OH (119 mg, 0.55 mmol) in HBTU (0.4 M in DMF, 1.3 mL) and DIPEA (198 µL, 1.1 mmol) to the resin. The mixture was shaken for 1 h and then flow washed with DMF (30 s) to give the valine alkyl thioester linked resin (**5.23**).

7.5.3.1 Competition binding assays

Competition binding assays for the native IGF-II (**4.1**), F19Cou IGF-II (**4.2**) and F28Cou IGF-II (**5.1**) proteins were carried out as previously detailed in Section 7.3.4.

7.6 Experimental described in Chapter 6

7.6.1 Experimental materials

FRET experiments were performed on a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies, CA, USA), using Spectrosil quartz micro cuvettes (100 µL) (Starna) with a 10 mm pathlength at 25 °C, at a pH of 7. The excitation wavelength was set to 280 nm and emission spectra were collected between 290 to 600 nm with a scan speed of 600 nm/min. Excitation and emission slit widths were set to 5 nm. Each single spectrum was averaged from three consecutive scans. All solvents, buffers and reagents used were of spectroscopic grade or higher. If spectroscopic grade was not available then solutions and solvents were filtered through a 0.22 µm filter before use.

7.6.2 FRET experiments

The sIGF-1R (**6.1**) was prepared by Ms Carlie Delaine according to the procedure described by Surinya *et al.*[63] and used as a 0.2 µM solution in 0.1 M sodium phosphate buffer (pH 7.2). The IGF-II proteins (**4.1**, **4.2** and **5.1**) were dissolved in 10 mM HCl. FRET experiments were performed by titrating the IGF-II protein (**4.1** and **4.2**) (0.2 µM; 1.7 µL) against the sIGF-1R (**6.1**) (0.2 µM; 100 µL). Due to the limitations on the quantity of sIGF-1R (**6.1**) available, FRET experiments for the F28Cou IGF-II protein (**5.2**) were performed by titrating **5.2** (0.19 µM; 1.6 µL) against the sIGF-1R (**6.1**) (0.19 µM; 91 µL). After each addition the solution was manually mixed, incubated at rt for 1 h and then spectra were recorded. IGF-II protein (**4.1**, **4.2** and **5.1**) was added until the protein and sIGF-1R were present in an equimolar ratio. A positive FRET interaction was indicted by an increase in the fluorescence intensity at 455 nm.

Chapter 8

Bibliography

(1) Denley, A.; Cosgrove, L. J.; Booker, G. W.; Wallace, J. C.; Forbes, B. E. *Cytokine & Growth Factor Reviews* **2005**, *16*, 421.
(2) Harris, L. K.; Westwood, M. *Growth Factors* **2012**, *30*, 1.
(3) Belfiore, A.; Frasca, F.; Pandini, G.; Sciacca, L.; Vigneri, R. *Endocrine Reviews* **2009**, *30*, 586.
(4) Stewart, C. E. H.; Rotwein, P. *Physiological Reviews* **1996**, *76*, 1005.
(5) Kornfeld, S. *Annual Review of Biochemistry* **1992**, *61*, 307.
(6) Yu, H.; Rohan, T. *Journal of the National Cancer Institute* **2000**, *92*, 1472.
(7) Pollak, M. N.; Schernhammer, E. S.; Hankinson, S. E. *Nature Reviews Cancer* **2004**, *4*, 505.
(8) Samani, A. A.; Yakar, S.; LeRoith, D.; Brodt, P. *Endocrine Reviews* **2007**, *28*, 20.
(9) Frasca, F.; Pandini, G.; Sciacca, L.; Pezzino, V.; Squatrito, S.; Belfiore, A.; Vigneri, R. *Archives Of Physiology And Biochemistry* **2008**, *114*, 23.
(10) Pollak, M. *Nature Reviews Cancer* **2008**, *8*, 915.
(11) Gallagher, E. J.; LeRoith, D. *Endocrinology* **2011**, *152*, 2546.
(12) Arnaldez, F. I.; Heiman, L. J. *Hematology-Oncology Clinics of North America* **2012**, *26*, 527.
(13) Gao, J.; Chang, Y. S.; Jallal, B.; Viner, J. *Cancer Research* **2012**, *72*, 3.
(14) Pollak, M. *Clinical Cancer Research* **2012**, *18*, 40.
(15) Pollak, M. *Nature Reviews Cancer* **2012**, *12*, 159.
(16) Tognon, C. E.; Sorensen, P. H. *Expert Opinion on Therapeutic Targets* **2012**, *16*, 33.
(17) Buck, E.; Mulvihill, M. *Expert Opinion on Investigational Drugs* **2011**, *20*, 605.
(18) Rajapaksha, H.; Alvino, C.; McCarthy, P.; Forbes, B. E. *Growth Hormone & IGF Research* **2012**, *22*, 158.
(19) De Meyts, P. *Trends in Biochemical Sciences* **2008**, *33*, 376.
(20) Lawrence, M. C.; McKern, N. M.; Ward, C. W. *Current Opinion in Structural Biology* **2007**, *17*, 699.
(21) Seino, S.; Bell, G. I. *Biochemical and Biophysical Research Communications* **1989**, *159*, 312.
(22) Denley, A.; Bonython, E. R.; Booker, G. W.; Cosgrove, L. J.; Forbes, B. E.; Ward, C. W.; Wallace, J. C. *Molecular Endocrinology* **2004**, *18*, 2502.
(23) Delaine, C.; Alvino, C. L.; McNeil, K. A.; Mulhern, T. D.; Gauguin, L.; De Meyts, P.; Jones, E. Y.; Brown, J.; Wallace, J. C.; Forbes, B. E. *Journal of Biological Chemistry* **2007**, *282*, 18886.
(24) Morgan, D. O.; Edman, J. C.; Standring, D. N.; Fried, V. A.; Smith, M. C.; Roth, R. A.; Rutter, W. J. *Nature* **1987**, *329*, 301.
(25) MacDonald, R. G.; Pfeffer, S. R.; Coussens, L.; Tepper, M. A.; Brocklebank, C. M.; Mole, J. E.; Anderson, J. K.; Ellson, C.; Czech, M. P.; Ullrich, A. *Science* **1988**, *239*, 1134.
(26) Ghosh, P.; Dahms, N. M.; Kornfeld, S. *Nature Reviews Molecular Cell Biology* **2003**, *4*, 202.
(27) Brown, J.; Jones, E. Y.; Forbes, B. E. *Trends in Biochemical Sciences* **2009**, *34*, 612.

(28) Pandini, G.; Frasca, F.; Mineo, R.; Sciacca, L.; Vigneri, R.; Belfiore, A. *Journal of Biological Chemistry* **2002**, *277*, 39684.
(29) Pavelić, K.; Kolak, T.; Kapitanović, S.; Radošević, S.; Spaventi, Š.; Krušlin, B.; Pavelić, J. *The Journal of Pathology* **2003**, *201*, 430.
(30) Denley, A.; Wallace, J. C.; Cosgrove, L. J.; Forbes, B. E. *Hormone and Metabolic Research* **2003**, *35*, 778.
(31) Giovannucci, E. *Journal of Nutrition* **2001**, *131*, 3109S.
(32) Buck, E.; Gokhale, P. C.; Koujak, S.; Brown, E.; Eyzaguirre, A.; Tao, N.; Rosenfeld-Franklin, M.; Lerner, L.; Chiu, M. I.; Wild, R.; Epstein, D.; Pachter, J. A.; Miglarese, M. R. *Molecular Cancer Therapeutics* **2010**, *9*, 2652.
(33) Papa, V.; Pezzino, V.; Costantino, A.; Belfiore, A.; Giuffrida, D.; Frittitta, L.; Vannelli, G. B.; Brand, R.; Goldfine, I. D.; Vigneri, R. *Journal of Clinical Investigation* **1990**, *86*, 1503.
(34) Sachdev, D.; Yee, D. *Endocrine-Related Cancer* **2001**, *8*, 197.
(35) Frasca, F.; Pandini, C.; Scalia, P.; Sciacca, L.; Mineo, R.; Costantino, A.; Goldfine, I. D.; Belfiore, A.; Vigneri, R. *Molecular and Cellular Biology* **1999**, *19*, 3278.
(36) Belfiore, A.; Frasca, F. *Journal of Mammary Gland Biology and Neoplasia* **2008**, *13*, 381.
(37) Kalla Singh, S.; Brito, C.; Tan, Q. W.; De León, M.; De León, D. *Growth Factors* **2011**, *29*, 278.
(38) Lambert, S.; Collette, J.; Gillis, J.; Franchimont, P.; Desaive, C.; Golwinkler, R. *International Journal of Cancer* **1991**, *48*, 826.
(39) Li, S. R.; Ng, C. F. J.; Banerjea, A.; Ahmed, S.; Hands, R.; Powar, M.; Ogunkolade, W.; Dorudi, S.; Bustin, S. A. *Tumor Biology* **2004**, *25*, 62.
(40) Dziadziuszko, R.; Camidge, D. R.; Hirsch, F. R. *Journal of Thoracic Oncology* **2008**, *3*, 815.
(41) Rinderknecht, E.; Humbel, R. E. *FEBS Letters* **1978**, *89*, 283.
(42) Terasawa, H.; Kohda, D.; Hatanaka, H.; Nagata, K.; Higashihashi, N.; Fujiwara, H.; Sakano, K.; Inagaki, F. *Embo Journal* **1994**, *13*, 5590.
(43) Torres, A. M.; Forbes, B. E.; Aplin, S. E.; Wallace, J. C.; Francis, G. L.; Norton, R. S. *Journal of Molecular Biology* **1995**, *248*, 385.
(44) Alvino, C. L.; McNeil, K. A.; Ong, S. C.; Delaine, C.; Booker, G. W.; Wallace, J. C.; Whittaker, J.; Forbes, B. E. *Journal of Biological Chemistry* **2009**, *284*, 7656.
(45) Bentley, G.; Dodson, E.; Dodson, G. U. Y.; Hodgkin, D.; Mercola, D. A. N. *Nature* **1976**, *261*, 166.
(46) Denley, A. PhD, University of Adelaide, 2004.
(47) Sato, A.; Nishimura, S.; Ohkubo, T.; Kyogoku, Y.; Koyama, S.; Kobayashi, M.; Yasuda, T.; Kobayashi, Y. *International Journal of Peptide and Protein Research* **1993**, *41*, 433.
(48) Murray-Rust, J.; McLeod, A. N.; Blundell, T. L.; Wood, S. P. *Bioessays* **1992**, *14*, 325.
(49) Benyoucef, S.; Surinya, K. H.; Hadaschik, D.; Siddle, K. *Biochemical Journal* **2007**, *403*, 603.

(50) Pandini, G.; Medico, E.; Conte, E.; Sciacca, L.; Vigneri, R.; Belfiore, A. *Journal of Biological Chemistry* **2003**, *278*, 42178.

(51) Massagué, J.; Czech, M. P. *Journal of Biological Chemistry* **1982**, *257*, 5038.

(52) Czech, M. P. *Cell* **1982**, *31*, 8.

(53) Ullrich, A.; Gray, A.; Tam, A. W.; Yangfeng, T.; Tsubokawa, M.; Collins, C.; Henzel, W.; Lebon, T.; Kathuria, S.; Chen, E.; Jacobs, S.; Francke, U.; Ramachandran, J.; Fujitayamaguchi, Y. *Embo Journal* **1986**, *5*, 2503.

(54) Rubin, J. B.; Shia, M. A.; Pilch, P. F. *Nature* **1983**, *305*, 438.

(55) Ellis, L.; Clauser, E.; Morgan, D. O.; Edery, M.; Roth, R. A.; Rutter, W. J. *Cell* **1986**, *45*, 721.

(56) Lee, J.; O'Hare, T.; Pilch, P. F.; Shoelson, S. E. *Journal of Biological Chemistry* **1993**, *268*, 4092.

(57) White, M. F.; Shoelson, S. E.; Keutmann, H.; Kahn, C. R. *Journal of Biological Chemistry* **1988**, *263*, 2969.

(58) Yu, K. T.; Czech, M. P. *Journal of Biological Chemistry* **1984**, *259*, 5277.

(59) Whittaker, J.; Whittaker, L. J.; Roberts, C. T.; Phillips, N. B.; Ismail-Beigi, F.; Lawrence, M. C.; Weiss, M. A. *Proceedings of the National Academy of Sciences* **2012**, *109*, 11166.

(60) Garrett, T. P. J.; McKern, N. M.; Lou, M.; Frenkel, M. J.; Bentley, J. D.; Lovrecz, G. O.; Elleman, T. C.; Cosgrove, L. J.; Ward, C. W. *Nature* **1998**, *394*, 395.

(61) Lou, M.; Garrett, T. P. J.; McKern, N. M.; Hoyne, P. A.; Epa, V. C.; Bentley, J. D.; Lovrecz, G. O.; Cosgrove, L. J.; Frenkel, M. J.; Ward, C. W. *Proceedings of the National Academy of Sciences* **2006**, *103*, 12429.

(62) Kristensen, C.; Wiberg, F. C.; Andersen, A. S. *Journal of Biological Chemistry* **1999**, *274*, 37351.

(63) Surinya, K. H.; Forbes, B. E.; Occhiodoro, F.; Booker, G. W.; Francis, G. L.; Siddle, K.; Wallace, J. C.; Cosgrove, L. J. *Journal of Biological Chemistry* **2008**, *283*, 5355.

(64) McKern, N. M.; Lawrence, M. C.; Streltsov, V. A.; Lou, M.-Z.; Adams, T. E.; Lovrecz, G. O.; Elleman, T. C.; Richards, K. M.; Bentley, J. D.; Pilling, P. A.; Hoyne, P. A.; Cartledge, K. A.; Pham, T. M.; Lewis, J. L.; Sankovich, S. E.; Stoichevska, V.; Da Silva, E.; Robinson, C. P.; Frenkel, M. J.; Sparrow, L. G.; Fernley, R. T.; Epa, V. C.; Ward, C. W. *Nature* **2006**, *443*, 218.

(65) Whitten, A. E.; Smith, B. J.; Menting, J. G.; Margetts, M. B.; McKern, N. M.; Lovrecz, G. O.; Adams, T. E.; Richards, K.; Bentley, J. D.; Trewhella, J.; Ward, C. W.; Lawrence, M. C. *Journal of Molecular Biology* **2009**, *394*, 878.

(66) Smith, B. J.; Huang, K.; Kong, G.; Chan, S. J.; Nakagawa, S.; Menting, J. G.; Hu, S.-Q.; Whittaker, J.; Steiner, D. F.; Katsoyannis, P. G.; Ward, C. W.; Weiss, M. A.; Lawrence, M. C. *Proceedings of the National Academy of Sciences* **2010**, *107*, 6771.

(67) Ward, C.; Lawrence, M.; Streltsov, V.; Garrett, T.; McKern, N.; Lou, M. Z.; Lovrecz, G.; Adams, T. *Acta Physiologica* **2008**, *192*, 3.

(68) Menting, J. G.; Whittaker, J.; Margetts, M. B.; Whittaker, L. J.; Kong, G. K. W.; Smith, B. J.; Watson, C. J.; Zakova, L.; Kletvikova, E.; Jiracek, J.; Chan, S. J.; Steiner,

D. F.; Dodson, G. G.; Brzozowski, A. M.; Weiss, M. A.; Ward, C. W.; Lawrence, M. C. *Nature* **2013**, *493*, 241.

(69) Fabry, M. *Bioconjugate Chemistry* **1999**, *10*, 200.

(70) Kurose, T.; Pashmforoush, M.; Yoshimasa, Y.; Carroll, R.; Schwartz, G. P.; Burke, G. T.; Katsoyannis, P. G.; Steiner, D. F. *Journal of Biological Chemistry* **1994**, *269*, 29190.

(71) Sørensen, H.; Whittaker, L.; Hinrichsen, J.; Groth, A.; Whittaker, J. *FEBS Letters* **2004**, *565*, 19.

(72) De Meyts, P.; Whittaker, J. *Nature Reviews Drug Discovery* **2002**, *1*, 769.

(73) Whittaker, J.; Whittaker, L. *Journal of Biological Chemistry* **2005**, *280*, 20932.

(74) Keyhanfar, M.; Booker, G. W.; Whittaker, J.; Wallace, J. C.; Forbes, B. E. *Biochemical Journal* **2007**, *401*, 269.

(75) Whittaker, L.; Hao, C.; Fu, W.; Whittaker, J. *Biochemistry* **2008**, *47*, 12900.

(76) Ward, C. W.; Lawrence, M. C. *Bioessays* **2009**, *31*, 422.

(77) De Meyts, P. *Endocrinology* **2012**, *153*, 2054.

(78) Hao, C.; Whittaker, L.; Whittaker, J. *Biochemical and Biophysical Research Communications* **2006**, *347*, 334.

(79) Mynarcik, D. C.; Williams, P. F.; Schaffer, L.; Yu, G. Q.; Whittaker, J. *Journal of Biological Chemistry* **1997**, *272*, 18650.

(80) Gauguin, L.; Delaine, C.; Alvino, C. L.; McNeil, K. A.; Wallace, J. C.; Forbes, B. E.; De Meyts, P. *Journal of Biological Chemistry* **2008**, *283*, 20821.

(81) Sakano, K.; Enjoh, T.; Numata, F.; Fujiwara, H.; Marumoto, Y.; Higashihashi, N.; Sato, Y.; Perdue, J. F.; Fujita-Yamaguchi, Y. *Journal of Biological Chemistry* **1991**, *266*, 20626.

(82) Roth, B. V.; Bürgisser, D. M.; Lüthi, C.; Humbel, R. E. *Biochemical and Biophysical Research Communications* **1991**, *181*, 907.

(83) Bürgisser, D. M.; Roth, B. V.; Giger, R.; Lüthi, C.; Weigl, S.; Zarn, J.; Humbel, R. E. *Journal of Biological Chemistry* **1991**, *266*, 1029.

(84) Oh, Y.; Beukers, M. W.; Pham, H. M.; Smanik, P. A.; Smith, M. C.; Rosenfeld, R. G. *Biochemical Journal* **1991**, *278*, 249.

(85) Alvino, C. L.; Ong, S. C.; McNeil, K. A.; Delaine, C.; Booker, G. W.; Wallace, J. C.; Forbes, B. E. *Plos One* **2011**, *6*.

(86) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. *Science* **1994**, *266*, 776.

(87) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498.

(88) Muir, T. W. *Annual Review of Biochemistry* **2003**, *72*, 249.

(89) David, R.; Richter, M. P. O.; Beck-Sickinger, A. G. *European Journal of Biochemistry* **2004**, *271*, 663.

(90) Miranda, L. P.; Alewood, P. F. *Peptide Science* **2000**, *55*, 217.

(91) Hodgson, D. R. W.; Sanderson, J. M. *Chemical Society Reviews* **2004**, *33*, 422.

(92) Xie, J. M.; Schultz, P. G. *Nature Reviews Molecular Cell Biology* **2006**, *7*, 775.

(93) Hackenberger, C. P. R.; Schwarzer, D. *Angewandte Chemie International Edition* **2008**, *47*, 10030.

(94) Sletten, E. M.; Bertozzi, C. R. *Angewandte Chemie International Edition* **2009**, *48*, 6974.

(95) Young, T. S.; Schultz, P. G. *Journal of Biological Chemistry* **2010**, *285*, 11039.
(96) Wu, Y.-W.; Goody, R. S. *Journal of Peptide Science* **2010**, *16*, 514.
(97) Lim, R. K. V.; Lin, Q. *Chemical Communications* **2010**, *46*, 1589.
(98) Chalker, J. M.; Bernardes, G. J. L.; Davis, B. G. *Accounts of Chemical Research* **2011**, *44*, 730.
(99) Hao, Z.; Hong, S.; Chen, X.; Chen, P. R. *Accounts of Chemical Research* **2011**, *44*, 742.
(100) Selvin, P. R. *Nature Structural Biology* **2000**, *7*, 730.
(101) Kajihara, D.; Abe, R.; Iijima, I.; Komiyama, C.; Sisido, M.; Hohsaka, T. *Nature Methods* **2006**, *3*, 923.
(102) Giepmans, B. N. G.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y. *Science* **2006**, *312*, 217.
(103) Ciruela, F. *Current Opinion in Biotechnology* **2008**, *19*, 338.
(104) Miyake-Stoner, S. J.; Miller, A. M.; Hammill, J. T.; Peeler, J. C.; Hess, K. R.; Mehl, R. A.; Brewer, S. H. *Biochemistry* **2009**, *48*, 5953.
(105) Roda, A.; Guardigli, M.; Michelini, E.; Mirasoli, M. *Analytical and Bioanalytical Chemistry* **2009**, *393*, 109.
(106) Miyawaki, A. *Annual Review of Biochemistry* **2011**, *80*, 357.
(107) Sahoo, H. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews* **2011**, *12*, 20.
(108) Clegg, R. M. *Current Opinion in Biotechnology* **1995**, *6*, 103.
(109) Selvin, P. R.; Kenneth, S. In *Methods in Enzymology*; Academic Press: 1995; Vol. 246, p 300.
(110) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*; 3rd ed.; Springer, 2006.
(111) Sapsford, K. E.; Berti, L.; Medintz, I. L. *Angewandte Chemie International Edition* **2006**, *45*, 4562.
(112) Michalet, X.; Weiss, S.; Jager, M. *Chemical Reviews* **2006**, *106*, 1785.
(113) Lavis, L. D.; Raines, R. T. *ACS Chemical Biology* **2008**, *3*, 142.
(114) Giepmans, B. N. G.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y. *Science* **2006**, *312*, 217.
(115) Hawe, A.; Sutter, M.; Jiskoot, W. *Pharmaceutical Research* **2008**, *25*, 1487.
(116) Albani, J. R. *Principles and Applications of Fluorescence Spectroscopy*; John Wiley & Sons, 2008.
(117) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chemical Reviews* **2010**, *110*, 2579.
(118) Katritzky, A. R.; Narindoshvili, T. *Organic & Biomolecular Chemistry* **2009**, *7*, 627.
(119) Zhang, J.; Campbell, R. E.; Ting, A. Y.; Tsien, R. Y. *Nature Reviews Molecular Cell Biology* **2002**, *3*, 906.
(120) Hoffmann, C.; Gaietta, G.; Bunemann, M.; Adams, S. R.; Oberdorff-Maass, S.; Behr, B.; Vilardaga, J.-P.; Tsien, R. Y.; Ellisman, M. H.; Lohse, M. J. *Nature Methods* **2005**, *2*, 171.
(121) Wysocki, L. M.; Lavis, L. D. *Current Opinion in Chemical Biology* **2011**, *15*, 752.
(122) Haugland, R. P. *The Molecular Probes Handbook: A Guide to Fluorescent Probes and Labeling Technologies, 11th Edition*; Life Technologies, 2010.

(123) Valeur, B.; Berberan-Santos, M. N. *Molecular Fluorescence: Principles and Applications*; John Wiley & Sons, 2001.
(124) Katerinopoulos, H. E. In *Current Pharmaceutical Design*; Bentham Science Publishers Ltd.: 2004; Vol. 10, p 3835.
(125) Wagner, B. *Molecules* **2009**, *14*, 210.
(126) Christie, R. M.; Lui, C.-H. *Dyes and Pigments* **1999**, *42*, 85.
(127) Christie, R. M.; Lui, C.-H. *Dyes and Pigments* **2000**, *47*, 79.
(128) Sun, W. C.; Gee, K. R.; Haugland, R. P. *Bioorganic & Medicinal Chemistry Letters* **1998**, *8*, 3107.
(129) Pechmann, H. v. *Berichte der deutschen chemischen Gesellschaft* **1884**, *17*, 929.
(130) Johnson, J. R. In *Organic Reactions*; John Wiley & Sons, Inc.: 2004.
(131) Jones, G. In *Organic Reactions*; John Wiley & Sons, Inc.: 2004.
(132) Bennett, F. A.; Barlow, D. J.; Dodoo, A. N. O.; Hider, R. C.; Lansley, A. B.; Lawrence, M. J.; Marriott, C.; Bansal, S. S. *Tetrahedron Letters* **1997**, *38*, 7449.
(133) Cohen, B. E.; McAnaney, T. B.; Park, E. S.; Jan, Y. N.; Boxer, S. G.; Jan, L. Y. *Science* **2002**, *296*, 1700.
(134) Wang, J.; Xie, J.; Schultz, P. G. *Journal of the American Chemical Society* **2006**, *128*, 8738.
(135) Huang, F.; Lerner, E.; Sato, S.; Amir, D.; Haas, E.; Fersht, A. R. *Biochemistry* **2009**, *48*, 3468.
(136) Signore, G.; Nifosì, R.; Albertazzi, L.; Storti, B.; Bizzarri, R. *Journal of the American Chemical Society* **2010**, *132*, 1276.
(137) Goldberg, J. M.; Speight, L. C.; Fegley, M. W.; Petersson, E. J. *Journal of the American Chemical Society* **2012**, *134*, 6088.
(138) Brun, M.-P.; Bischoff, L.; Garbay, C. *Angewandte Chemie International Edition* **2004**, *43*, 3432.
(139) Koch, K.; Podlech, J. In *Synthetic Communications*; Taylor & Francis Ltd: 2005; Vol. 35, p 2789.
(140) Anderson, K. H. PhD, The University of Canterbury, 2007.
(141) Vogel, A. I. *Vogel's Textbook of Practical Organic Chemistry*; 3rd ed.; Longman Scientific & Technical, 1956.
(142) Kirmse, W. *European Journal of Organic Chemistry* **2002**, *2002*, 2193.
(143) Koch, K.; Podlech, J. *Synthetic Communications* **2005**, *35*, 2789
(144) Brooks, D. W.; Lu, L.; Masamune, S. *Angewandte Chemie International Edition* **1979**, *18*, 72.
(145) Li, C.; Henry, E.; Mani, N. K.; Tang, J.; Brochon, J.-C.; Deprez, E.; Xie, J. *European Journal of Organic Chemistry* **2010**, *2010*, 2395.
(146) Ammar, H.; Fery-Forgues, S.; El Gharbi, R. *Dyes and Pigments* **2003**, *57*, 259.
(147) Trenor, S. R.; Shultz, A. R.; Love, B. J.; Long, T. E. *Chemical Reviews* **2004**, *104*, 3059.
(148) Dougherty, D. A. *Current Opinion in Chemical Biology* **2000**, *4*, 645.
(149) de Graaf, A. J.; Kooijman, M.; Hennink, W. E.; Mastrobattista, E. *Bioconjugate Chemistry* **2009**, *20*, 1281.
(150) Ai, H.-w. *Analytical and Bioanalytical Chemistry* **2012**, *403*, 2089.

(151) Dawson, P. E.; Kent, S. B. H. *Annual Review of Biochemistry* **2000**, *69*, 923.
(152) Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182.
(153) Wang, L.; Xie, J.; Schultz, P. G. *Annual Review of Biophysics and Biomolecular Structure* **2006**, *35*, 225.
(154) Bain, J. D.; Diala, E. S.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R. *Journal of the American Chemical Society* **1989**, *111*, 8013.
(155) Hohsaka, T.; Sisido, M. *Current Opinion in Chemical Biology* **2002**, *6*, 809.
(156) Singh-Blom, A.; Hughes, R.; Ellington, A. In *Enzyme Engineering*; Samuelson, J. C., Ed.; Humana Press: 2013; Vol. 978, p 93.
(157) Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. *Current Opinion in Chemical Biology* **2010**, *14*, 774.
(158) Liu, C. C.; Schultz, P. G. *Annual Review of Biochemistry* **2010**, *79*, 413.
(159) Mendel, D.; Cornish, V. W.; Schultz, P. G. *Annual Review of Biophysics and Biomolecular Structure* **1995**, *24*, 435.
(160) Wang, L.; Schultz, P. G. *Angewandte Chemie International Edition* **2005**, *44*, 34.
(161) Hohsaka, T.; Ashizuka, Y.; Taira, H.; Murakami, H.; Sisido, M. *Biochemistry* **2001**, *40*, 11060.
(162) Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G. *Proceedings of the National Academy of Sciences* **2004**, *101*, 7566.
(163) Benzer, S.; Champe, S. P. *Proceedings of the National Academy of Sciences* **1962**, *48*, 1114.
(164) Cornish, V. W.; Benson, D. R.; Altenbach, C. A.; Hideg, K.; Hubbell, W. L.; Schultz, P. G. *Proceedings of the National Academy of Sciences* **1994**, *91*, 2910.
(165) Wang, K.; Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nature Biotechnology* **2007**, *25*, 770.
(166) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. *Journal of Molecular Biology* **2010**, *395*, 361.
(167) Johnson, D. B. F.; Xu, J.; Shen, Z.; Takimoto, J. K.; Schultz, M. D.; Schmitz, R. J.; Xiang, Z.; Ecker, J. R.; Briggs, S. P.; Wang, L. *Nature Chemical Biology* **2011**, *7*, 779.
(168) Loscha, K. V.; Herlt, A. J.; Qi, R.; Huber, T.; Ozawa, K.; Otting, G. *Angewandte Chemie International Edition* **2012**, *51*, 2243.
(169) Smolskaya, S.; Zhang, Z. J.; Alfonta, L. *Plos One* **2013**, *8*, 1.
(170) Johnson, D. B. F.; Wang, C.; Xu, J.; Schultz, M. D.; Schmitz, R. J.; Ecker, J. R.; Wang, L. *ACS Chemical Biology* **2012**.
(171) King, R.; Wells, J. R. E.; Krieg, P.; Snoswell, M.; Brazier, J.; Bagley, C. J.; Wallace, J. C.; Ballard, F. J.; Ross, M.; Francis, G. L. *Journal of Molecular Endocrinology* **1992**, *8*, 29.
(172) Lien, S.; Milner, S. J.; Graham, L. D.; Wallace, J. C.; Francis, G. L. *Biotechnology and Bioengineering* **2001**, *74*, 335.
(173) Francis, G. L.; Aplin, S. E.; Milner, S. J.; McNeil, K. A.; Ballard, F. J.; Wallace, J. C. *Biochemical Journal* **1993**, *293*, 713.
(174) Studier, F. W.; Moffatt, B. A. *Journal of Molecular Biology* **1986**, *189*, 113.

(175) Novagen *pET System manual*, 2003.
(176) Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. *Journal of the American Chemical Society* **2000**, *122*, 5010.
(177) Alvino, C. L.; Forbes, B. E. **2010**, personal communication
(178) Studier, F. W. *Protein Expression and Purification* **2005**, *41*, 207.
(179) Jitrapakdee, S.; Cottam, J. M.; Abell, A. D.; Alvino, C.; Evans, M.; Ong, S. C.; Delaine, C.; Forbes, B. E. **2013**, Optimisation of an expression system used to incorporate non-canonical amino acids into insulin-like growth factor-II (IGF-II), unpublished work
(180) Held, D.; Yaeger, K.; Novy, R. *inNovations* **2003**, *18*, 4.
(181) Short, G. F.; Golovine, S. Y.; Hecht, S. M. *Biochemistry* **1999**, *38*, 8808.
(182) Park, H.-S.; Hohn, M. J.; Umehara, T.; Guo, L.-T.; Osborne, E. M.; Benner, J.; Noren, C. J.; Rinehart, J.; Söll, D. *Science* **2011**, *333*, 1151.
(183) Isaacs, F. J.; Carr, P. A.; Wang, H. H.; Lajoie, M. J.; Sterling, B.; Kraal, L.; Tolonen, A. C.; Gianoulis, T. A.; Goodman, D. B.; Reppas, N. B.; Emig, C. J.; Bang, D.; Hwang, S. J.; Jewett, M. C.; Jacobson, J. M.; Church, G. M. *Science* **2011**, *333*, 348.
(184) Shooter, G. K.; Magee, B.; Soos, M. A.; Francis, G. L.; Siddle, K.; Wallace, J. C. *Journal of Molecular Endocrinology* **1996**, *17*, 237.
(185) Milner, S. J.; Francis, G. L.; Wallace, J. C.; Magee, B. A.; Ballard, F. J. *Biochemical Journal* **1995**, *308*, 865.
(186) Merrifield, R. B. *Journal of the American Chemical Society* **1963**, *85*, 2149.
(187) Cudic, M.; Fields, G. B. In *Molecular Biomethods Handbook*; Walker, J. M., Rapley, R., Eds.; Humana Press: 2008, p 515.
(188) Kent, S. B. H. *Annual Review of Biochemistry* **1988**, *57*, 957.
(189) Albericio, F. *Peptide Science* **2000**, *55*, 123.
(190) Amblard, M.; Fehrentz, J.-A.; Martinez, J.; Subra, G. *Molecular Biotechnology* **2006**, *33*, 239.
(191) Merrifield, B. *Science* **1986**, *232*, 341.
(192) Atherton, E.; Fox, H.; Harkiss, D.; Logan, C. J.; Sheppard, R. C.; Williams, B. J. *Journal of the Chemical Society, Chemical Communications* **1978**, 537.
(193) Carpino, L. A.; Han, G. Y. *Journal of the American Chemical Society* **1970**, *92*, 5748.
(194) Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *International Journal of Peptide and Protein Research* **1992**, *40*, 180.
(195) Lloyd-Williams, P.; Albericio, F.; Giralt, E. *Chemical Approaches to the Synthesis of Peptides and Proteins*; CRC Press, 1997.
(196) Chan, W. C.; White, P. D. *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*; Oxford University Press, 2000.
(197) Isidro-Llobet, A.; Alvarez, M.; Albericio, F. *Chemical Reviews* **2009**, *109*, 2455.
(198) Li, C. H.; Yamashiro, D.; Glenn Hammonds Jr, R.; Westphal, M. *Biochemical and Biophysical Research Communications* **1985**, *127*, 420.
(199) Yamashiro, D.; Li, C. H. *International Journal of Peptide and Protein Research* **1985**, *26*, 299.

(200) Oh, Y. M.; Muller, H. L.; Zhang, H. P.; Ling, N.; Rosenfeld, R. G. In *Current Directions in Insulin-Like Growth Factor Research*; LeRoith, D., Raizada, M. K., Eds.; Plenum Press: New York, 1993; Vol. 343, p 41.

(201) Veber, D.; Milkowski, J.; Varga, S.; Denkewalter, R.; Hirschmann, R. *Journal of the American Chemical Society* **1972**, *94*, 5456.

(202) Kamber, B. *Helvetica Chimica Acta* **1971**, *54*, 927.

(203) Kamber, B.; Hartmann, A.; Eisler, K.; Riniker, B.; Rink, H.; Sieber, P.; Rittel, W. *Helvetica Chimica Acta* **1980**, *63*, 899.

(204) Fujii, N.; Otaka, A.; Funakoshi, S.; Bessho, K.; Yajima, H. *Journal of the Chemical Society, Chemical Communications* **1987**, 163.

(205) Fujii, N.; Otaka, A.; Watanabe, T.; Okamachi, A.; Tamamura, H.; Yajima, H.; Inagaki, Y.; Nomizu, M.; Asano, K. *Journal of the Chemical Society, Chemical Communications* **1989**, 283.

(206) Hober, S.; Lundström Ljung, J.; Uhlén, M.; Nilsson, B. *FEBS Letters* **1999**, *443*, 271.

(207) Hancock, W. S.; Battersby, J. E. *Analytical Biochemistry* **1976**, *71*, 260.

(208) Wohr, T.; Wahl, F.; Netzi, A.; Rohwedder, B.; Sato, T.; Sun, X.; Mutter, M. *Journal of the American Chemical Society* **1996**, *118*, 9218.

(209) Coin, I.; Beyermann, M.; Bienert, M. *Nature Protocols* **2007**, *2*, 3247.

(210) Coïc, Y.-M.; Lan, C. L.; Neumann, J.-M.; Jamin, N.; Baleux, F. *Journal of Peptide Science* **2010**, *16*, 98.

(211) Coin, I.; Doelling, R.; Krause, E.; Bienert, M.; Beyermann, M. *Journal of Organic Chemistry* **2006**, *71*, 6171.

(212) Abedini, A.; Raleigh, D. P. *Organic Letters* **2005**, *7*, 693.

(213) Zhang, S.; Lin, F.; Hossain, M.; Shabanpoor, F.; Tregear, G.; Wade, J. *International Journal of Peptide Research and Therapeutics* **2008**, *14*, 301.

(214) Sohma, Y.; Pentelute, B. L.; Whittaker, J.; Hua, Q.-x.; Whittaker, L. J.; Weiss, M. A.; Kent, S. B. H. *Angewandte Chemie International Edition* **2008**, *47*, 1102.

(215) Nilsson, B. L.; Soellner, M. B.; Raines, R. T. *Annual Review of Biophysics and Biomolecular Structure* **2005**, *34*, 91.

(216) Kent, S.; Sohma, Y.; Liu, S.; Bang, D.; Pentelute, B.; Mandal, K. *Journal of Peptide Science* **2012**, *18*, 428.

(217) Kent, S. B. H. *Chemical Society Reviews* **2009**, *38*, 338.

(218) Johnson, E. C. B.; Kent, S. B. H. *Journal of the American Chemical Society* **2006**, *128*, 6640.

(219) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proceedings of the National Academy of Sciences* **1999**, *96*, 10068.

(220) Dawson, P. E.; Churchill, M. J.; Ghadiri, M. R.; Kent, S. B. H. *Journal of the American Chemical Society* **1997**, *119*, 4325.

(221) Harris, P. W. R.; Brimble, M. A. *Synthesis* **2009**, *2009*, 3460.

(222) Harris, P. W. R.; Yang, S. H.; Brimble, M. A. *Tetrahedron Letters* **2011**, *52*, 6024.

(223) Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. *International Journal of Peptide Research and Therapeutics* **2007**, *13*, 31.

(224) Cottam, J. M.; Scanlon, D.; Karas, J.; Calabrese, A.; Pukala, T.; Forbes, B.; Wallace, J.; Abell, A. *International Journal of Peptide Research and Therapeutics* **2013**, *19*, 61.
(225) Milton, R. C. d. L.; Milton, S. C. F.; Adams, P. A. *Journal of the American Chemical Society* **1990**, *112*, 6039.
(226) Förster, T. *Annalen der Physik* **1948**, *437*, 55.
(227) Andrews, D. L. *Canadian Journal of Chemistry* **2008**, *86*, 855.
(228) Wu, P. G.; Brand, L. *Analytical Biochemistry* **1994**, *218*, 1.
(229) Van Der Meer, B. W.; Coker, G.; Chen, S. Y. S. *Resonance Energy Transfer: Theory and Data*; John Wiley & Sons, 1994.
(230) Sinev, M.; Landsmann, P.; Sineva, E.; Ittah, V.; Haas, E. *Bioconjugate Chemistry* **2000**, *11*, 352.
(231) Stryer, L.; Haugland, R. P. *Proceedings of the National Academy of Sciences* **1967**, *58*, 719.
(232) Mayor, S.; Bilgrami, S. In *Evaluating Techniques in Biochemical Research*; Zuk, D., Ed.; Cell Press Cambridge, MA, 2007.
(233) Sekar, R. B.; Periasamy, A. *The Journal of Cell Biology* **2003**, *160*, 629.
(234) Lohse, M. J.; Nuber, S.; Hoffmann, C. *Pharmacological Reviews* **2012**, *64*, 299.
(235) Arkin, M. R.; Glicksman, M. A.; Fu, H.; Havel, J. J.; Du., Y. In *Assay Guidance Manual [Internet]*; Sittampalam, G. S., Gal-Edd, N., Arkin, M. R., Auld, D., Austin, C., Bejcek, B., Glicksman, M., Inglese, J., Lemmon, V., Li, Z., McGee, J., McManus, O., Minor, L., Napper, A., Riss, T., O. Joseph Trask, J., Weidner, J., Eds.; Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences; Available from: http://www.ncbi.nlm.nih.gov/books/NBK92000/: 2012.
(236) Albani, J. R. In *Principles and Applications of Fluorescence Spectroscopy*; Blackwell Publishing Ltd: 2008, p 88.
(237) Rich, R. L.; Chen, Y.; Neven, D.; Negrerie, M.; Gai, F.; Petrich, J. W. *The Journal of Physical Chemistry* **1993**, *97*, 1781.
(238) Steward, L. E.; Collins, C. S.; Gilmore, M. A.; Carlson, J. E.; Ross, J. B. A.; Chamberlin, A. R. *Journal of the American Chemical Society* **1997**, *119*, 6.
(239) Kwon, I.; Tirrell, D. A. *Journal of the American Chemical Society* **2007**, *129*, 10431.
(240) Rogers, J. M. G.; Lippert, L. G.; Gai, F. *Analytical Biochemistry* **2010**, *399*, 182.
(241) Petrović, D.; Leenhouts, K.; Roosmalen, M.; Broos, J. *Amino Acids* **2013**, *44*, 1329.
(242) Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M. *FEBS Journal* **2013**, *280*, 5705.
(243) Moraes, I.; Evans, G.; Sanchez-Weatherby, J.; Newstead, S.; Stewart, P. D. S. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2014**, *1838*, 78.
(244) Garman, E. F. *Science* **2014**, *343*, 1102.
(245) Rabenstein, M. D.; Shin, Y. K. *Proceedings of the National Academy of Sciences* **1995**, *92*, 8239.
(246) Tsvetkov, Y. D.; Grishin, Y. A. *Instruments and Experimental Techniques* **2009**, *52*, 615.
(247) Klug, C. S.; Feix, J. B. In *Methods in Cell Biology*; Dr. John, J. C., Dr. H. William Detrich, III, Eds.; Academic Press: 2008; Vol. Volume 84, p 617.
(248) Schiemann, O.; Prisner, T. F. *Quarterly Reviews of Biophysics* **2007**, *40*, 1.

(249) Eaton, S. S.; Eaton, G. R. In *Biomedical EPR, Part B: Methodology, Instrumentation, and Dynamics*; Eaton, S. R., Eaton, G. R., Berliner, L. J., Eds.; Springer US: 2005; Vol. 24/B, p 223.

(250) Schreier, S.; Bozelli, J., Jr.; Marín, N.; Vieira, R. F.; Nakaie, C. *Biophysical Reviews* **2012**, *4*, 45.

(251) Armarego, W. L. F.; Perrin, D. D. D. *Purification of Laboratory Chemicals*; 4th ed.; Butterworth Heinemann, 1997.

(252) Schägger, H.; von Jagow, G. *Analytical Biochemistry* **1987**, *166*, 368.

(253) Sell, C.; Dumenil, G.; Deveaud, C.; Miura, M.; Coppola, D.; DeAngelis, T.; Rubin, R.; Efstratiadis, A.; Baserga, R. *Molecular and Cellular Biology* **1994**, *14*, 3604.

(254) Soos, M. A.; O'Brien, R. M.; Brindle, N. P.; Stigter, J. M.; Okamoto, A. K.; Whittaker, J.; Siddle, K. *Proceedings of the National Academy of Sciences* **1989**, *86*, 5217.

(255) Soos, M. A.; Field, C. E.; Lammers, R.; Ullrich, A.; Zhang, B.; Roth, R. A.; Andersen, A. S.; Kjeldsen, T.; Siddle, K. *Journal of Biological Chemistry* **1992**, *267*, 12955.

(256) Sheppard, R. *Journal of Peptide Science* **2003**, *9*, 545.

(257) Christensen, T. *Acta Chemica Scandinavica Series B-Organic Chemistry and Biochemistry* **1979**, *33*, 763.

(258) Vojkovsky, T. *Peptide research* **1995**, *8*, 236.

(259) Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I. *Analytical Biochemistry* **1970**, *34*, 595.