Improved production of biopharmaceuticals by site-specific cleavage of fusion proteins expressed in *Escherichia coli*.

A thesis submitted for the degree of Doctor of Philosophy at the University of Adelaide, Department of Molecular Bioscience

Adam Charlton,

B.Biotech. (Hons)

Flinders University, South Australia

May 2007

TABLE OF CONTENTS

CHAPTER 3: SCREENING THE α-LYTIC PROTEASE MUTANTS FOR DESIRABLE SPECIFICITY

CHAPTER 4: CLEAVAGE OF MET-PGH(1-11)VN-PAPM-IGF-I

ABSTRACT

The recombinant expression of heterologous proteins in microorganisms, such *Escherichia coli*, is often improved by producing the protein of interest translationally linked to another, often unrelated, protein giving rise to a "fusion protein" construct. For many applications it is desirable or imperative to separate the extraneous material from the protein of interest. An increasingly popular approach to this task is the use of site-specific endoproteases to excise the protein product. A number of commercially available site-specific proteases exist, but many are not capable of generating an authentic N-terminus for the product, display unsatisfactory specificity leading to adventitious cleavage of the product, or they are unsuitable for an industrial process.

Mutants of the serine protease α-Lytic protease have been shown to satisfy many of the criteria for an industrially suitable protease and have been applied to the cleavage of some important fusion proteins used in the production of members of the Insulin-like Growth Factor (IGF) family. Lacking from these examples, however, is any viable proteolytic solution for the liberation of human IGF-I from fusion proteins. This has been primarily attributed to the Proline bearing N-terminal tripeptide sequence of this protein, which is known to be refractory to the activity of many site-specific proteases.

It has been suggested that, in the generation of two combinatorial mutant libraries of α-Lytic protease, the preference for amino acids C-terminal to the cleavage site may have been altered. It is the purpose of this work to first determine if such an alteration has been made in any of the mutants so as to

allow cleavage immediately before the N-terminus of human IGF-I, and then to task the lead mutant(s) to the cleavage of the full-length fusion protein. All members of the two mutant libraries were cultured and their activity confirmed and quantified against a generic β-casein substrate in a high-throughput assay. A second high-throughput technique was then employed to query the mutant proteases for their ability to catalyse proteolysis at the required sequence in a peptide model. Finding that many mutants appeared successful at this task, the findings were verified on a longer peptide model of the cleavage site.

Initially the yields achieved by cleavage of the full-length IGF-I fusion protein by a lead candidate mutant α-Lytic protease were not sufficient to satisfy the requirements of an industrial process, despite alteration of the reaction conditions. However, the insight gained from these reactions could be applied to the redesign of the protein structure around the intended site of cleavage, significantly improving site-specific proteolysis. The IGF-I generated by this cleavage has been shown to be bioequivalent to commercial reference standard to cultured mammalian cells and the yield of this process is approximately 5-fold improved over the existing cleavage system.

DECLARATION

This thesis contains no material that has been accepted for the award of any other degree or diploma 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. To the best of my knowledge and belief it contains no material that has previously been published by any other person, except where due reference has been made in the text. I give consent to a 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.

Adam Charlton

ACKNOWLEDGEMENTS

I would like to express my appreciation to my supervisors Prof. John Wallace and Mr. Geoff Francis for the opportunity to undertake this work, and for their advice and assistance throughout the project.

This project was funded by a Strategic Partnerships with Industry – Research and Training scheme (SPIRT) grant. I would therefore like to acknowledge the financial support of the Australian Research Council and Novozymes GroPep Ltd.

I would like to thank the past and current staff of Novozymes GroPep Ltd. for their support and technical assistance, with specific thanks to Graham Hobba and Phil Elliott. I would also like to thank the past and current members of the Wallace laboratory for being available with materials and helpful advice whenever the need arose. Finally, I would like my wife Leah Charlton for her tireless support and understanding.

ABBREVIATIONS

ACN	acetonitrile
CV	column volume
DMSO	dimethyl sulphoxide
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EIA	enzyme immunoassay
FITC	fluorescein isothiocyanate
FP	Fluorescence polarisation
FRET	Fluorescence resonance energy transfer
2-HED	2-hydroxyethyl disulphide
HEPES	N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)
IGF	Insulin-like growth factor
MES	2-(*N*-morpholino)ethanesulfonic acid
Met-pGH(1-11)VN	Methionine followed by the first 11 amino acids of porcine growth hormone terminated by an valine-asparagine dipeptide
PBS	Phosphate Buffered Saline
PBS-T	Phosphate Buffered Saline with Tween
SDS	sodium dodecyl sulphate
SEM	Standard error of the mean
TFA	trifluoroacetic acid

Tris	tris-(hydroxymethyl)methyl-amine
Tween-20	polyoxyethylene (20) sorbitan monolaurate

The fusion protein reported as Met-pGH(1-11)VN-Arg^3-IGF-I is registered under the trademark LONG®R^{3}IGF-I by Novozymes GroPep, Ltd.

1.1 FUSION PROTEINS

From the earliest examples of recombinant protein expression in microorganisms it has been apparent that there are a number of factors involved in achieving consistent yields. Proteins expressed in heterologous host systems may be subject to intracellular proteolysis (Murby *et al.,* 1996) or sequestration as insoluble inclusion bodies (Schein, 1989) which pose potentially difficult renaturation challenges. Even should production of intact, active protein be achieved, there remains the difficulty of its purification away from the extremely complex background environment of all other host-derived proteins.

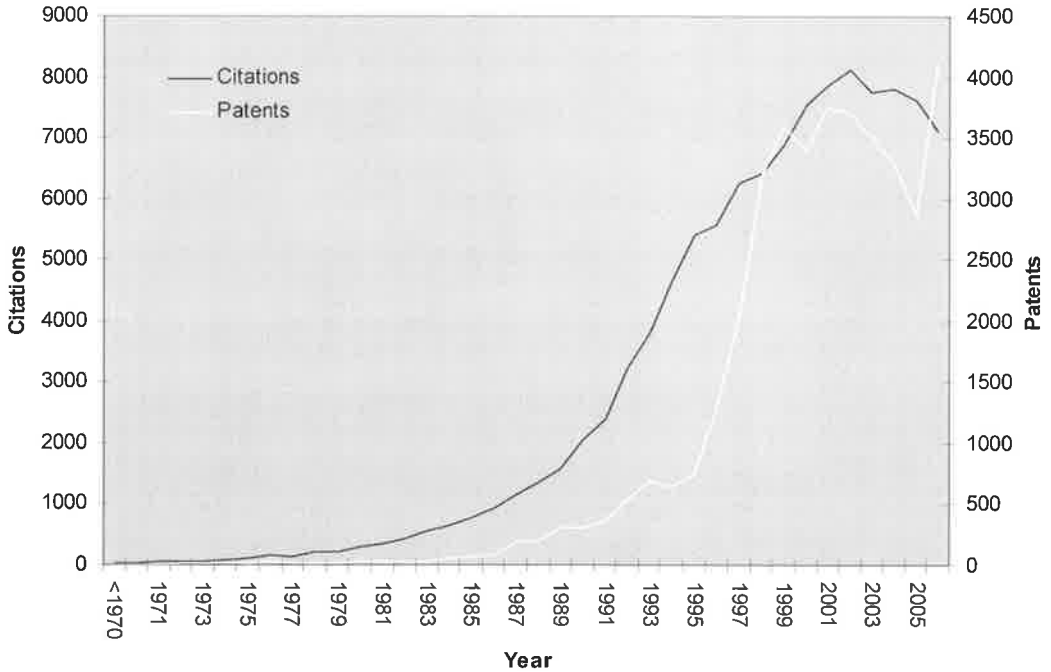
One approach employed to remedy these issues is the fusion protein strategy, where the intended protein product is translationally linked to an unrelated protein. A fusion protein in its simplest form is any recombinant protein expressed with extra amino acids attached to it, with the latter commonly referred to as a fusion partner. Such chimaeric proteins are generated by the recombinant linkage, or fusion, of the nucleotide coding sequence of the target protein in frame with that of the fusion partner. When this fused coding sequence is translated, the resulting product consists of the protein of interest and the fusion partner.

A fusion partner can be as small as a single residue, such as an unremoved N-terminal methionine from initiation of bacterial expression (Marston, 1986), or a short peptide sequence, as is the case with many affinity tags (Terpe, 2003). The fusion partner can be attached to the N- or C-terminus of the desired protein product. As discussed later, N-terminal

fusions do offer significant advantages should the removal of the fusion partner be required.

The fusion protein strategy has become ubiquitous in modern recombinant protein expression and invaluable to biotechnology. A wide variety of modular expression vectors are available to facilitate their generation, with the coding sequence of popular fusion partners up- or down-stream of the multiple cloning site (see Table 1.1). The widespread adoption of this approach can be appreciated by the number of citations indexed by Medline. Figure 1.1 shows the total citations returned for the search parameter "fusion protein" over the past three decades. A similar explosion in commercial application of fusion protein technology has been seen, with the growth in number of patents pertaining to fusion proteins approved by the USPTO showing a similar pattern to that of citations in the scientific literature over the past decade in particular.

Figure 1.1 - Citations and patents returned from Medline and USPTO databases, respectively, for the search term "fusion protein" as at 28/02/2007



Many popular fusion partners employed for the routine cloning and expression of recombinant proteins, particularly at small-batch scale or in the research environment are multifunctional, simultaneously providing solutions to some or all of the aforementioned challenges to recombinant expression. The fusion protein strategy has been applied to improve expression, to confer protection from proteolysis, to promote soluble expression or provide assistance in refolding, and is the subject of many reviews (Marston, 1986; Uhlen and Moks, 1990; Uhlen *et al.*, 1992; LaVallie and McCoy, 1995; Makrides, 1996; Nilsson *et al.,* 1997; Hearn and Acosta, 2001; Terpe, 2003; Lichty *et al.,* 2005; Waugh, 2005; Arnau *et al.,* 2006). However, in many fusion protein systems these attributes are consequential, with the primary goal of the fusion partners being the provision of a "handle" for facile affinity purification of the recombinant protein. As can be seen in Table 1.1, a non-exhaustive survey of the commercial suppliers of recombinant expression systems that incorporate a fusion partner suggests that almost all have an affinity purification function.

Table 1.1 – Selected examples of commercial prokaryotic expression vectors for fusion protein production.

Commercial supplier	Vector series	Fusion partner	Fusion partner type	Cleavage site provided
Invitrogen	pET100	Xpress™, hexahistidine	Affinity	Enterokinase
	pET102	His-Patch Thioredoxin, V5, Hexahistidine	Solubilisation & Affinity	Enterokinase
	pET151	V5, Hexahistidine	Affinity	TEV proteinase
	pET SUMO	SUMO, Hexahistidine	Solubilisation & Affinity	SUMO protease
	pThioHis	His-Patch Thioredoxin	Solubilisation & Affinity	Enterokinase
	pBAD	V5, Hexahistidine	Affinity	Enterokinase
	pTrcHis	Xpress™, hexahistidine	Affinity	Enterokinase
Merck Biosciences (Novagen)	pET30 & 46	Hexahistidine, S-tag	Affinity	Enterokinase
	pET32	Thioredoxin, Hexahistidine, S-tag	Solubilisation & Affinity	Factor Xa or

Commercial supplier	Vector series	Fusion partner	Fusion partner type	Cleavage site provided
				Enterokinase
	pET39 & 40	DsbA, Hexahistidine	Solubilisation & Affinity	Thrombin
	pET41	Glutathione-S-Transferase (GST)	Solubilisation & Affinity	Enterokinase
	pET44	Nus-Tag, Hexahistidine	Solubilisation & Affinity	Enterokinase
	pET47	Hexahistidine, S-tag	Affinity	3Cpro
	pET48	Thioredoxin, Hexahistidine, S-tag	Solubilisation & Affinity	3Cpro
	pET49	GST, Hexahistidine, S-tag	Solubilisation & Affinity	3Cpro
	pET50	Nus-tag, Hexahistidine, S-tag	Solubilisation & Affinity	3Cpro
	pET52	Strep-Tag II	Affinity	3Cpro
GE Healthcare (Amersham	pGEX-4	GST	Affinity	Thrombin
	pGEX-5	GST	Affinity	Factor Xa

Commercial supplier	Vector series	Fusion partner	Fusion partner type	Cleavage site provided
Biosciences)	pGEX-6	GST	Affinity	3Cpro
New England Biolabs	pIMPACT	Intein	Affinity	Self cleaving
	pMAL-2E	Maltose Binding Protein (MBP)	Solubilisation & Affinity	Enterokinase
	pMAL-2G	Maltose Binding Protein (MBP)	Solubilisation & Affinity	Genenase I
	pMAL-2X	Maltose Binding Protein (MBP)	Solubilisation & Affinity	Factor Xa
Stratagene	pCAL	Calmodulin Binding Peptide, FLAG-tag	Affinity	Thrombin and/or Enterokinase
Clontech	pHAT	HAT™ (Histidine Affinity Tag)	Affinity	Enterokinase
Qiagen	pQE	Hexahistidine	Affinity	Factor Xa
Promega	pFN2	GST	Solubilisation & Affinity	TEV proteinase
	pFN6	$(HQ)_3$	Affinity	Nil

1.2 FUSION PARTNER ROLES

The theme underlying the exploitation of a protein as a fusion partner is that it possesses an attribute that is desired to be conferred on the protein product of interest. However, many popular fusion partners employed today confer multiple functionalities, such as improved solubility and the provision of an affinity tag rather than any single attribute. For example, maltose binding protein attachment both improves solubility and offers an affinity purification solution (di Guan *et al.,* 1988). In many cases the manner in which a fusion partner fulfills its intended function involves a number of factors. As detailed below for example, a fusion partner can improve the yield of a recombinant protein indirectly by enabling expression of the protein in a natively folded, soluble form. The correctly folded protein, in turn, may have decreased susceptibility to intracellular proteolysis.

Fusion partners fall into two broad categories, either short peptide sequences or whole proteins, each with their own advantages and disadvantages. As a rule of thumb, whole proteins are required to confer improved product solubility, whereas short peptides are often employed as affinity tags (Waugh, 2005).

The extremely large and diverse range of fusion partners makes evaluation of each individual example impossible; instead discussion here will focus on examples illustrating the attributes commonly associated with a successful fusion partner strategy.

1.2.1 The fusion protein strategy to prevent intracellular proteolysis

Many fusion proteins are able to directly protect the protein product from digestion by intracellular proteases. In some cases, even protection of the N- and C-termini by a short peptide fusion partner has proven sufficient to decrease proteolysis, presumably by decreasing the action of exoproteases on the protein product (Bowie, 1989; Rajan et al., 1998).

It has been suggested that the degradation of heterologous proteins in E. coli may be a defence mechanism in response to the recognition of the non-native protein (Marston, 1986). The use of a fusion partner derived from the host organism can overcome this problem, as such host proteins are often well expressed in a native conformation in their natural host, allowing more intact fusion protein to be obtained (Shine *et al.*, 1980). As a result, many of the fusion partners commonly used, particularly in *E. coli*, were originally derived from that organism.

In keeping with the notion of using a fusion partner to confer a desirable characteristic to the protein of interest, highly protease resistant proteins have been used to prevent the intracellular proteolysis of the product. One such example is ubiquitin, itself quite protease resistant and, contrary to its role in eukaryotes of targeting proteins for proteasomal digestion (Baek, 2003); has been successfully employed as a fusion partner in prokaryotes on a number of occasions to produce recombinant protein (Butt *et al.*, 1989; Power *et al.*, 1990; Koken *et al.*, 1993; Baker, 1996; Catanzariti *et al.*, 2004; Moon *et al.*, 2007). In a classic example, fusion of ubiquitin to the N-terminus of two unrelated proteins, yeast metallothionein and the α-subunit of the GTP-binding stimulatory protein of adenylate cyclase, increased the yield of either

protein from undetectable levels to 20% of total cellular protein (Butt *et al.,* 1989). Likewise, the small ubiquitin-related modifier (SUMO) (Malakhov *et al.,* 2004; Butt *et al.,* 2005) has also been used as a protease resistant fusion partner.

If intracellular proteolysis of soluble heterologous proteins cannot be avoided, one of the simplest mechanisms to prevent excessive degradation is to deny the proteases access to the recombinant protein. The accumulation of recombinant proteins into insoluble inclusion bodies (Kane and Hartley, 1988; Schein, 1989) within the bacterial cell is one way to physically partition the protein product from host proteases (Sassenfeld, 1990; Uhlen and Moks, 1990). Similarly, secretion of the product into the periplasmic space can offer protection from proteolysis. In a classic example, the half-life of recombinant proinsulin was found to be 10-fold longer in the periplasm than in the cytoplasm (Talmadge and Gilbert, 1982).

Finally, partially or incorrectly folded proteins are frequently degraded by host proteases (Gottesman, 1989; Baneyx and Mujacic, 2004) and therefore the use of fusion partners may mitigate this problem by allowing correct folding to occur which, in turn, can prevent such intracellular proteolysis by minimising the formation of misfolded intermediates, as discussed in a later section.

1.2.2 The fusion protein strategy for the expression of insoluble protein

It has been argued that insoluble expression of recombinant protein in *E. coli* is the rule rather than the exception (Marson, 1986), that is unless special lengths are taken to avoid it the most likely location of a novel

expression construct will be largely in the insoluble fraction. Estimates have put the proportion of insoluble recombinant proteins at roughly 33% (Chayen, 2004), although others have put this number as low as 13% (Chambers *et al.,* 2004).

Insoluble expression is characterised by the aggregation of the expressed protein into intracellular "inclusion bodies" (Kane and Hartley, 1988; Schein, 1989). A number of mechanisms for this phenomenon have been postulated, including:

(1) That the higher numbers of protein molecules present during recombinant protein over-expression increases the chances of intermolecular collisions that can initiate aggregate formation via hydrophobic interactions (Nilsson and Anderson, 1991).

(2) Overwhelming of the natural protein folding apparatus of the host-cell (Nilsson and Anderson, 1991; Rudolph and Lilie, 1996).

(3) Sequestration of partially folded or misfolded protein, as such aggregates often consist of misfolded proteins or partially folded intermediates, but rarely fully unfolded or natively folded protein (Mitraki and King, 1989; Kapust and Waugh, 1999).

Hypotheses (1) and (2) in particular are supported by the observation that some native *E. coli* proteins are sequestered in inclusion bodies when over-expressed (Cheng, 1983). These mechanisms would tend to suggest that insoluble formation of recombinant protein inclusion bodies is a consequence of the nature of the target protein itself or the act of over-expressing a protein, but may or may not be promoted by an associated fusion partner. However, there are a few fusion partners that are thought to

promote the formation of inclusion bodies such as TrpE protein (Spindler *et al.,* 1984; Yansura, 1990), the cII protein of λ phage (Nagai and Thorgersen, 1984) and truncations of porcine growth hormone (Francis *et al.,* 1992; King *et al.,* 1992). It could be crudely thought that a fusion partner that increases the total hydrophobicity of the fusion protein complex will tend to promote inclusion body formation and vice versa, in keeping with scenario (1), above.

Insoluble expression of recombinant proteins has some merits for recombinant protein expression, including not only the previously mentioned protection from proteolysis, but also a strong tendency towards higher total expression yields (Marston, 1986). Not being reliant on the often rate-limiting folding process (Esposito and Chatterjee, 2006), recombinant proteins can be expressed recombinantly to over 26% of total cellular protein (Marston, 1986). Insoluble expression of recombinant proteins also offers a convenient early partial purification step, as the concentration of recombinant protein within the inclusion bodies can be as high as 50% (Murby *et al.,* 1996).

The most significant, and sometimes insurmountable disadvantage is the requirement for a refolding process should the native structure of the protein be desired. Such refolding protocols usually involve the application of reducing agents with high concentrations of chaotropes (guanidine-HCl, urea) or detergents, followed by dilution or buffer exchange to more physiological conditions (Rudolph and Lilie, 1996). The refolding of recombinant proteins is not always necessary though; there are circumstances when the production of folded protein may be undesirable, such as for eliciting an immune response (Uhlen and Moks, 1990), or for expression of proteins toxic to the host, as has been shown for the production of bacteriocidal peptides (Lee *et al.,* 2000).

In summary, where an effective *in vitro* refolding protocol exists for a given fusion protein, or is simply not relevant, the significant benefits provided by an insoluble expression strategy make it an attractive solution for recombinant protein expression.

1.2.3 The fusion protein strategy for the soluble expression of protein

The *in vitro* refolding of proteins is a complex task, with no guarantee of a successful outcome in economic terms. It is therefore of little surprise that a significant number of popular fusion partners are able to promote *in vivo* folding of the protein into the native conformation while ensuring solubility.

Early theories suggested a passive role for solubilising fusion partners, wherein almost any protein that was highly soluble in the host organism could transfer this property to the protein product (LaVallie and McCoy, 1995; Weickert *et al.,* 1996). Whilst this field is still not clearly understood the current understanding is that the influence of the fusion partner is not passive (Waugh, 2005). Moreover, recent evidence suggests that the fusion partner takes an active role in the folding of the fusion protein. Comparative studies of highly soluble fusion partners have shown the phenomenon of recombinant protein solubility to be far more complicated than originally thought. For example, Kapust and Waugh, (1999) investigated 6 aggregation prone proteins with thioredoxin (Trx), Glutathione-S-Transferase (GST) and Maltose-Binding Protein (MBP) fusion partners and found MBP to be superior at promoting solubility in all 6 cases, followed by thioredoxin and GST, which were similar in their effectiveness. The improvements in solubility of the MBP fusions over the other two ranged from roughly 3-fold to over 80-fold. By

contrast, Hammarstrom *et al.,* (2002), studied 27 human gene products for solubility, each with seven solubilising fusion partners. In this case it was observed that thioredoxin gave the highest probability of achieving a soluble product, with thioredoxin > MBP ≈ Gb1 domain of Protein G > the ZZ derivative of the IgG binding domain of protein A (ZZ) > N utilisation Substance A (NusA) > GST > hexahistidine at the promotion of soluble fusion protein recovery. The difficulty in drawing a general conclusion is further compounded by the findings of Shih *et al.,* (2002), who investigated 32 proteins from diverse natural hosts paired with 8 solubilising fusion partners. In this study the relative effectiveness of creating soluble fusion products was NusA ≈ MBP > GST > thioredoxin > hexahistidine > intein ≈ calmodulin-binding protein (CBP) > cellulose-associated protein (CAP) at creating soluble fusion products. Such variability in the solubilising ability of the fusion partners for these diverse protein products suggests that their action is not simply the result of being highly soluble themselves. Indeed, if that were the case, the relative solubilising ability of the fusion partners would not vary so dramatically.

The only clear trend to emerge from the above studies is that MBP is frequently one of the best solubilising fusion partners (Esposito and Chatterjee, 2006) and has thus been the subject of considerable attention as concerning its mode of action. It is hypothesised that the solubilising action of the MBP fusion partner may be due to binding of hydrophobic surfaces of the incompletely folded proteins in the deep hydrophobic cleft of MBP that is the natural ligand binding site (Kapust and Waugh, 1999). This mechanism would thus function in a manner akin to that of the endogenous *E. coli* GroEL

chaperone (Fenton *et al.,* 1994; Buckle *et al.,* 1997) and could therefore conceivably prevent aggregation due to mechanisms (1) and (3) listed in section 1.1.2, instead allowing the nascent polypeptide sufficient time in solution to adopt a folded conformation. It has been argued that the length and composition of the linker peptide between the MBP and protein product moieties may have an impact on the ability of MBP to actively assist in the folding of the fusion protein (Nallamsetty and Waugh, 2006). The other leading solubilising fusion partners assist in the generation of the disulphide bond structure in the protein, with the disulphide reductase thioredoxin (Holmgren, 1985; LaVallie *et al.,* 1993) and the disulphide isomerase NusA (Davis *et al.,* 1999) often able to assist in the solubilisation of a wide range of proteins in the comparative studies discussed above.

It is interesting to note that the assistance to protein folding offered by solubilising fusion partners can extend beyond the intracellular environment. The *in vitro* refolding of insoluble pepsinogen and procathespin-D were improved by the fusion to either MBP or thioredoxin (Sachdev and Chirgwin, 1998). Similarly, Interferon γ receptor α-chain could only be refolded when fused to thioredoxin (Williams et al., 1995).

Secretion, often into the periplasmic space of E. coli, is frequently associated with soluble expression strategies. The main reasons for this are twofold: (i) the periplasmic space offers a more favourable redox environment for disulphide bond formation in the protein than the reducing cytoplasm (LaVallie and McCoy, 1995; Baneyx and Mujacic, 2004), with evidence that disulphide bond formation occurs upon secretion (Pollitt and Zalkin, 1983); (ii) the periplasm also offers a simplified purification option, given that the protein

content in that region only represents 4% of total cellular proteins (Nossal and Heppel, 1966; Pugsley and Schwartz, 1985).

Whilst not recovered as a fusion protein *per se*, the use of a leader signal peptide sequence to achieve secretion is conceptually identical to the traditional fusion partner, involving the genetic linkage of the sequence encoding the leader N-terminal to that of the protein product. One such sequence is *pelB* (Lei *et al.,* 1987) and is a common feature on many commercial vectors for soluble expression. A useful side-effect of the secretion leader is that, when correctly auto-excised, the protein product is left with a native N-terminus.

1.2.4 *The fusion protein strategy for the provision of an affinity tag*

As mentioned earlier, many of the most commonly used fusion partners provide a mechanism for affinity purification of the fusion protein, and are chosen for this reason. This specific type of fusion partner is commonly referred to as an "affinity tag" and it has been claimed that affinity tags are the most common application of any fusion protein technology (Nilsson *et al.,* 1997).

Three classes of affinity tag have make up this diverse group (Lichty *et al.,* 2005). The first of these classes consists of those peptide or protein tags that bind a small molecule ligand and includes such members as the hexahistidine (Hochuli *et al.,* 1987; Hochuli *et al.,* 1988) and GST (Smith and Johnson, 1988) tags, which bind chelated metal ions and glutathione moieties, respectively. The second class are peptide or protein affinity tags that bind to a protein acceptor and would include such examples as Protein A

(and its binding domain variants) which binds to the Fc portion of immunoglobulin G (Bjork *et al.,* 1972; Sjoquist *et al.,* 1972; Uhlen *et al.,* 1983) or the calmodulin-binding peptide which specifically recognises calmodulin (Stofko-Hahn *et al.,* 1992). The final classification could be considered a specialised subset of the second, representing epitope affinity tags such as the FLAG peptide which binds reversibly to the M1 monoclonal antibody (Hopp *et al.,* 1988).

The extremely large and diverse group of affinity tag fusion partners are highly application specific, with each tag representing a unique mechanism of action. Furthermore, many affinity tagged fusion proteins have their own unique properties. As such, a comprehensive review of affinity tag fusion protein technology is well beyond the scope of this discussion, but as a group they provide an excellent example of fusion partners conferring a desirable trait to the fusion protein construct.

1.2.5 Conclusion

An extensive range of fusion protein systems are available in the modern biotechnology toolkit, and there exists virtually unlimited scope for the development of novel strategies by conferring a desirable characteristic to a target protein by such linkage. There is no single universally applicable fusion protein solution for all recombinant protein products. In the following section this concept is examined with an important family of recombinant proteins.

1.3 THE FUSION PROTEIN STRATEGY IN THE PRODUCTION OF INSULIN-LIKE GROWTH FACTORS.

Insulin-like Growth Factor(s) I and II (IGF-I and –II) are 70 and 67 amino acid polypeptide growth factors, respectively. Their names reflect the fact that they not only mimic some of the effects of insulin, but also exhibit some mitogenic actions such as the stimulation of DNA and protein synthesis (Humbel, 1990; Schmid, 1995; Jones and Clemmons, 1995).

Early research into the IGF system was significantly hampered by the lack of material, with Rinderknecht and Humbel (1976) reporting only milligrams of IGF-I purified from tons of human serum. Although this was somewhat improved by isolation from bovine colostrum of 70mg of IGF-I per 1000L (Francis *et al.*, 1986), there remained an imperative for facile and high yielding recombinant production. However, from the earliest attempts at their recombinant expression (Itakura *et al.,* 1977), it was clear that the expression of IGF proteins was plagued by many of the problems that have beset recombinant protein production detailed above. Recombinant IGF proteins have frequently suffered low yields due to either poor transcription and/or translation (Buell *et al.,* 1985), intracellular proteolysis (Buell *et al.,* 1985; Saito *et al.,* 1987) or both. More recently these problems have been largely overcome by various strategies; for example, a 14-fold increase in IGF-I yields was achieved when expressed in a host strain lacking the degP, ptrA and ompT cellular proteases (Joly *et al.,* 1998). Additionally, a further two-fold increase in yield was obtained by overexpressing the DsbA protein. Interestingly, despite the common use of this protein as a solubilising fusion partner, this secondary increase in yield was attributed to more rapid

aggregation of fusion protein into inclusion bodies, which further protected the protein from residual proteolysis (Joly *et al.,* 1998).

IGF proteins are frequently expressed in an insoluble form as inclusion bodies, which may actually have benefits to product recovery, beyond the partial purification and proteolysis prevention that inclusion bodies offer. Whilst low expression or proteolytic degradation of the product are of concern for total process efficiency, it is this issue of folding that makes the IGF's an interesting model group of recombinant protein. Attempts at the oxidative refolding of solubilised and denatured IGF protein under favourable conditions revealed a non-Anfinsen folding pattern (Hober *et al.,* 1992; Milner *et al.,* 1995). That is, the amino acid sequence itself did not contain sufficient information for the *in vitro* folding of the protein into one unique minimal energy state (Anfinsen, 1973). In addition to the correctly folded protein, one or more non-native folding isomers are commonly encountered. In the case of human IGF-I, the dominant alternative folding isomer has divergent disulphide bonds between Cys^{6}-Cys^{47} and Cys^{48}-Cys^{52} compared to the disulphide pattern in the native protein of Cys^{6}-Cys^{48} and Cys^{47}-Cys^{52} (both isoforms share the native Cys^{18}-Cys^{61} bond) (Iwai *et al.,* 1989). This has a significant impact on the biological potency of the protein, with 50 to 100-fold reduction in the bioactivity of the alternate isomer in cultured cells (Milner *et al.,* 1995).

Owing to this divergent *in vitro* folding behaviour of recombinantly expressed IGF, some investigators have opted to attempt to avoid the necessity of such refolding events by producing soluble protein. Efforts into IGF production have thus been relatively evenly divided between soluble and

insoluble expression strategies. At face value, insoluble approaches seem to achieve significantly higher raw productivity, with reported yields of recombinant protein ranging 10mg per litre of microbial ferment (Saito *et al.,* 1987) up to 8.5g/L (Joly *et al.,* 1998), with most in the hundreds of mg/L range (Saito *et al.,* 1987; King *et al.,* 1992; Francis *et al.,* 1992; Francis *et al.,* 1993). This compares starkly with the productivity of soluble expression systems, with outputs ranging from tens of micrograms per litre (Bayne *et al.*, 1989; Hammarberg *et al.,* 1989) up to tens of milligrams per litre (Moks *et al.*, 1987; Elliott *et al.,* 1990; Hammarberg *et al.,* 1991; Kim and Lee, 1996; Vai *et al.,* 2000), with isolated examples reaching 1g/L (Forsberg *et al.,* 1990). However, the most common yields are in the high hundreds of micrograms to unit milligram per litre range (Moks *et al.*, 1987; Bayne *et al.*, 1988; Cascieri *et al.*, 1988; Cascieri *et al.*, 1989; Forsberg *et al.,* 1990). This suggests that not only is insoluble expression of IGF proteins on average 10 to 1000-fold more productive than soluble expression, but that the outcomes are more predictable, with less variability apparent in the productivity of the various insoluble systems.

The issue of productivity becomes more complicated when the considerable downstream processing that is required to produce high purity IGF's from inclusion bodies is considered. Insoluble proteins require a downstream refolding event, which tends to decrease the final yield of active protein, particularly with the aforementioned divergent folding behaviour of IGF's. Additionally, fusion protein strategies require removal of the fusion partner if fully authentic IGF is the goal (see section 1.5). The associated losses experienced in these steps can reduce the overall process yield to be

comparable to that of direct soluble expression strategies. However, it must be noted that the impact of misfolded IGF's on the final yields obtained from soluble production are rarely considered, despite documented accounts of their occurrence (Forsberg *et al.,* 1990; Stuebe *et al.*, 1991; Joly *et al.,* 1998; Vai *et al.,* 2000).

Given the above issues associated with recombinant expression of the IGF's, they have themselves been used as a basis to evaluate the effectiveness of different novel fusion partner constructs. For example, Zhang *et al.* (1998) used IGF-I, IGFBP-3, 3C proteinase and TGF-β2 to examine the utility of the periplasmic protein DsbA as a solubilising fusion partner. Likewise, Moks *et al.* (1987) used IGF-I as a model recombinant protein to evaluate the IgG binding domains of *Staphylococcus* Protein A as a solubilising fusion partner and affinity tag. This latter fusion partner has proven very successful for the soluble expression of IGF protein, with the synthetic IgG binding domain designated the Z domain implemented on a number of occasions for IGF production (Moks *et al.,* 1987; Forsberg *et al.,* 1989; Hammarberg *et al.*, 1991; Forsberg *et al.,* 1992; Samuelsson *et al.,* 1994; Hodgson *et al.,* 1995). Over the years, IGF's have also been successfully produced as fusion proteins with partners as diverse as thioredoxin (Wilkinson *et al.,* 2004), Protein A (Luthi *et al.*, 1992), DsbA analogues (Zhang *et al.*, 1998), β-galactosidase (Saito *et al.*, 1987; Jeong and Lee, 2003; Choi *et al.,* 2003) and interleukin 1 (Sakano *et al.*, 1991).

1.3.1 The Met-pGH fusion partner for IGF production

One of the most important developments in IGF production came from the work of King *et al.* (1992) and Francis *et al.* (1992), where truncations of porcine growth hormone (pGH) were employed as a fusion partner with IGF-I. The expression system for the production of pGH in *E. coli* had been previously optimised and shown to express this protein to high levels (Vize and Wells, 1987). The fusion of IGF-I to an N-terminal fragment of pGH significantly increased yields, with 0.8-1.2g of IGF-I fusion protein produced per litre of bacterial culture medium, whereas attempts at production of IGF-I without the fusion of the pGH leader sequence were unsuccessful (King *et al.*, 1992). This compares very favourably to other instances of insoluble expression of human IGF-I, with yields of 0.24g/L (Jeong and Lee, 2003) and 0.13g/L (Saito *et al.*, 1987) reported. It was suggested that the improvement in productivity provided by the pGH constructs could be due to improved translational efficiency, increased product stability or decreased cellular toxicity (King *et al.*, 1992). It is interesting to note that the Met-pGH truncation and IGF-I fusions were expressed better than the full-length pGH itself (King *et al.*, 1992).

The first instance of the pGH fusion partner included the first 46 of the 191 residues of the full-length hormone, having the additional features of an N-terminal Methionine for translation initiation (King *et al.*, 1992). The N-terminal fusion partner construct also consisted of a linking C-terminal Valine-Asparagine dipeptide motif which provided a convenient restriction site and provided a motif for cleavage by hydroxylamine (discussed later). The leader peptide construct was designated Met-pGH(1-46)VN (King *et al.*,

1992). In subsequent experiments the length of the pGH sequence was reduced to the first 11 amino acids (Francis *et al.* 1992), giving rise to the Met-pGH(1-11)VN-IGF-I fusion protein, or "Long®-IGF-I". It was found that the reduction in the length of the fusion partner did not significantly alter the productivity of the construct on a mass yield basis, but given the 30% reduction in the size of the fusion protein, this equates directly to an improvement in the yield of IGF-I product on a molar basis.

The value of the methionyl porcine growth hormone truncations as a fusion partner is not limited to direct improvements in the recovery of the recombinant protein. The refolding of the insoluble protein is also facilitated by the presence of this N-terminal extension, or "leader peptide". Milner *et al.* (1995) shows the protein lost to the misfolded form decreased from 35% to 25% when the leader peptide was present, with the 10% difference quantitatively translated into improved yield of the native folded protein. Although lacking any actual chaperone function, it is hypothesized that in the case of IGF-I the presence of the leader peptide possibly sterically inhibits formation of a salt bridge between Glu^{3} and Arg^{56} that can form following the creation of the Cys^{18}-Cys^{61} disulphide bond (Hober *et al.,* 1992; Milner *et al.,* 1995). If formed, this salt bridge would stabilise the incorrect structure and allows the formation of the non-native disulphide bonds.

Since the initial production of IGF-I, the Met-pGH leader peptide fusion system has proven useful in the production of a host of other recombinant proteins, including an extensive variety of other IGF proteins ranging from mammalian IGF's such as human IGF-II (Francis *et al.,* 1993) through avian IGF's such as chicken IGF-II (Upton *et al.,* 1995) to fish IGF's such as

Barramundi IGF-II (Degger *et al.,* 2001) and Hagfish IGF (Upton *et al.,* 1997). The utility of this fusion protein system is not limited to the production of IGF's and analogues, with the Met-pGH system also employed successfully for the production of epidermal growth factor (Xian *et al.,* 1996), betacellulin (Dunbar *et al.,* 2001) and transforming growth factor β3 (G. Francis, personal communication).

1.4 CLEAVAGE OF FUSION PROTEINS

Removal of the fusion partner may be necessary, depending on the intended use for the recombinant protein, particularly when the protein is destined for use as a biopharmaceutical, in structural studies of the native protein or in multimeric protein production.

Reliance on endogenous host mechanisms for the removal of non-native modifications, such as a bacterial translation initiating formyl-methionine or a secretion signal may not be 100% efficient, leading to undesirable product heterogeneities (Uhlen and Moks, 1990). Such issues can be handled by the use of an *in vitro* site-specific cleavage system if not avoidable during expression of the product.

Site-specific cleavage of a fusion protein involves targeting the cleavage to the junction between the protein of interest and the fusion partner by the insertion of a "linker" peptide motif that is highly favourable for the cleavage system chosen. This cleavage system must be specific enough for the C-terminal linkage between the amino acid motif at this junction and the N-terminal residue of the product. This will minimise losses of recombinant protein yield due to the incorrect product being formed and internal

degradation of the product. Furthermore it is more likely to display catalytic rates and hence process economics that are industrially viable.

The strategies employed to achieve site-specific cleavage of fusion proteins are clearly delineated into two approaches, being the use of either chemical agents or proteases.

1.4.1 Chemical cleavage of fusion proteins

Some of the earliest examples of the site-specific cleavage of proteins involved the use of chemical agents to selectively break the peptide backbone of the protein at defined locations. Although having fallen out of favour in recent times (for reasons discussed below) chemical agents can offer rapid and cost effective cleavage of proteins at highly specific locations.

Among the chemical mechanisms most frequently employed for protein cleavage are cyanogen bromide which cleaves after methionine residues (Goeddel *et al.,* 1979), hydroxylamine, acting between Asn-Gly dipeptide motifs (Bornstein & Balian, 1977), N-chlorosuccinimide which cleaves C-terminal to tryptophan residues (Shechter *et al.,* 1976) and acid hydrolysis which cleaves between Asp-Pro dipeptide motifs (Schultz, 1967).

The above examples illustrate one significant disadvantage of the chemical cleavage agents; the motifs recognised are very short amino acid sequences (only one or two residues), which in many cases effectively precludes the uniqueness of the site within the fusion protein. The agents are also often quite toxic, with the above examples rated from hazardous to very toxic (Sigma-Aldrich MSDS for product numbers 481432, 379921 and 109681), posing handling and environmental concerns, particularly at

manufacturing scale. Perhaps chief amongst the issues faced by the use of chemical cleavage systems, is that the extreme environments can result in modifications to other residues in the protein product; such as the well documented modification by hydroxylamine with the formation of hydroxymates of asparagine and glutamine residues (Canova-Davis *et al.,* 1992). The heterogeneities resulting from such modifications are of concern for any recombinant protein, with potential implications for protein function. Moreover, they are absolutely untenable in a biopharmaceutical product and will require extensive and often difficult purification regimes, given the overall physiochemical similarity between the authentic and modified forms of the protein.

1.4.2 Proteolytic cleavage of fusion proteins

In recent years attention has turned towards the use of site-specific proteases for the cleavage of fusion proteins. Proteases are enzymes that catalyse the hydrolysis of peptide bonds and may be categorised as either endo- or exo-proteases, with the former cleaving the protein internally and the latter removing amino acids stepwise from the N- or C-terminus of a protein. Although some examples of exopeptidase use have been documented for fusion protein processing (Hejnaes *et al.,* 1992; Pedersen *et al.,* 1999), discussion here will be limited to the more widespread application of endoproteases in this field.

Of the large number of proteases known (estimated to represent 2% of all gene products (Rawlings and Barrett, 1999)), very few can be considered suitable for the site-specific cleavage of fusion proteins. To be applicable to

this application, a protease must satisfy a number of key criteria. It is advantageous that the proteases lack the requirement for cofactors or quaternary structure, as the former would require replacement or regeneration for continued activity, and the latter could be subject to disruption, restricting the range of operating conditions and the useful in-process lifespan of the enzymes. However, the foremost criterion is that they must be highly specific for the cleavage of protein substrates at a discrete location. Moreover, this location should be sufficiently unique so as to avoid adventitious cleavage elsewhere within the protein product. As a consequence, the majority of proteases employed for the site-specific cleavage of proteins must recognise a particular sequence of amino acids, with such "recognition sequences" or motifs generally made up of three to six residues. Furthermore, it is highly desirable that the proteases cleave C-terminal to this recognition sequence, with no interactions of residues downstream of the cleavage site. This condition will be rarely met leading to variations in cleavage rate at the preferred site. The consequence of the idealised circumstances is that the protease will be able to excise N-terminal fusion partners leaving the protein product with a native N-terminal sequence, with all non-target sequences upstream of the cleavage site including motif itself removed.

An emerging property that is becoming important to the cleavage of fusion proteins in the biopharmaceutical industry is that the proteases be capable of being produced by recombinant means themselves. With the use of animal derived in-process elements becoming increasingly undesirable in the current regulatory environment, this issue will likely gain prominence in the

future. For totally new products this would probably mean rejection by the appropriate regulatory bodies.

It is not surprising that many examples of enzymes that fulfill the specificity criterion are isolated as proprotein activating or regulatory enzymes, where there has been an evolutionary imperative towards site-specificity. More recently, protein engineering technology has enabled some examples of proteases with engineered site-specificity. Table 1.4.2.1 lists a range of site-specific proteases, and the most common proteases used for fusion protein cleavage (as indicated by the frequent appearance of their cleavage motifs on fusion protein expression vectors, Table 1.1) are discussed below.

A major limitation to the adoption of proteases for fusion protein cleavage, particularly for industry, is the almost universally high cost of the enzymes. Commercially available site-specific proteases are either extracted from natural sources, such as bovine plasma, which therefore require extensive purification, or they are a recombinant product and often encumbered by the manufacturer's licensing and patent landscape.

The protease substrate-subsite nomenclature that will be used in this work is that of Schechter and Berger (1967), where the amino acids of the substrate N-terminal to the scissile bond are labeled P_X, and those C-terminal are labeled P_X', where X position of the residue in relation to the site of cleavage, increasing with distance from the site of cleavage. Thus P_1 is the amino acid immediately preceding the scissile bond, whereas P_2' would be the second residue C-terminal to that position. Correspondingly, the subsites

on the enzyme that interact with these amino acids are labeled S and S' where, for example, S_2 binds P_2 and S_1' binds P_1'.

Table 1.4.2.1 – Examples of site-specific proteases. Amino acids represented by one letter code.

Protease	Cleavage site	Reference	Commercial supplier	Notes
Enterokinase	D-D-D-D-K↓	Maroux *et al.*, 1971 Anderson *et al.*, 1977	I, R, M, N, S	FLAG tag contains cleavage site. All recombinant products are the catalytic subunit.
Factor Xa	I-(E/D)-G-R↓	Nagai & Thorgersen, 1984 Jenny *et al.*, 2003	S, M, Pi, R, Q, N, G, Pr	
Thrombin	X-X-(P/G)-R↓G X=hydrophobic	Chang, 1985	M, G, S, R	Pre-Immobilised to agarose beads from Sigma. Biotinylated from Merck
H64A subtilisin	(A/F)-A-H-Y↓	Carter & Wells, 1987 Carter *et al.*, 1989	N	Sold as "Genenase I"
TEV proteinase	E-N-L-Y-F-Q↓(G/S/A/M)	Carrington & Dougherty, 1988 Dougherty & Parks, 1989	I, U	Invitrogen AcTEV is a stabilized mutant with N-terminal his-tag.

Protease	Cleavage site	Reference	Commercial supplier	Notes
3C proteinase	L-E-V-L-F-Q↓G-P	Ivanoff *et al.*, 1986 Cordingley *et al.*, 1990	M, G	GE offers a GST-fusion as "PreScission"
Plasmin	X-(K/R)↓ X=Aromatic	Harris *et al.*, 2000 Backes *et al.*, 2000	S, U	
Kallikrein	P-F-R↓	MacDonald *et al.*, 1988	M, U	
Trypsin	K↓, R↓	Halfon *et al.*, 2004	S, N, Pr, Pi, U	
Chymotrypsin	F↓, W↓, Y↓	Graf *et al.*, 2004	S, U	

G = GE Healthcare (Amersham Biosciences)	Pr = Promega
I = Invitrogen	Q = Qiagen
M = Merck Biosciences	R = Roche Diagnostics
N = New England Biolabs	S = Sigma Aldrich
Pi = Pierce	U = U.S. Biological

Supplier legend

1.4.3 Site-specific proteases of the mammalian blood clotting cascade

The proteases that activate the mammalian blood clotting cascade have been a source of two of the most well studied and established proteases employed for the site-specific cleavage of fusion proteins, namely the serine proteases; Factor Xa and Thrombin (Jenny *et al.,* 2003). Unfortunately, neither Factor Xa nor Thrombin is currently commercially available as recombinant preparations, limiting their application in industrial biopharmaceutical manufacture. Other proteases of the blood-clotting system such as plasmin and urokinase have received far less attention. Despite their *in vivo* site-specificity, the limited attempts at their use in fusion protein cleavage have not met with success, with adventitious cleavage of the protein product observed. (Forsberg et al., 1992).

In vivo Factor Xa (EC 3.4.21.6), activated by Factor IXa (Hultin and Nemerson, 1978) or Factor VIIa (Nemerson, 1966), in turn activates the next enzyme in the cascade by cleavage of prothrombin to liberate active Thrombin. Factor Xa is highly specific for cleavage C-terminal to its natural tetrapeptide recognition sequence Ile-(Glu/Asp)-Gly-Arg (Butkowski *et al.,* 1977; Walz *et al.,* 1977), allowing for the generation of an authentic N-terminus for the protein product. This was first demonstrated by the release of authentic human β-globin from a λ phage cII protein fusion partner. (Nagai and Thorgersen, 1984; Nagai *et al.,* 1985).

Thrombin (EC 3.4.21.5), acting immediately downstream of Factor Xa in the cascade, *in vivo* cleaves fibrinogen to generate fibrin (Blomback *et al.,* 1967; Takagi and Doolittle, 1974). Different to most other specific proteases used for fusion protein cleavage, thrombin does not have a long defined

recognition sequence, with the only absolute requirement for Thrombin cleavage being that it occurs after a basic residue, much like Trypsin (Halfon *et al.,* 2004). Unlike Trypsin however, cleavage by Thrombin requires the interactions of amino acids at other subsites. By studying the rate of Thrombin cleavage against 30 protein substrates, Chang (1985) determined that the optimum Thrombin cleavage motif should contain a P_2 Pro or Gly and that the P_3 and P_4 positions should be occupied by hydrophobic residues. Any non-acidic amino acid is tolerated in the P_1' and P_2' positions (Chang, 1985), but Thrombin distinctly prefers cleavage prior to Gly, so much so that it is considered to cleave within a recognition sequence, with interaction of the residue on the C-terminal side of the scissile bond required. As such, a protein released from a fusion by this protease will have a residual N-terminal Glycine extension. Whilst there are examples of fusion protein cleavage by Thrombin prior to residues other than Gly, these are the exception (Chang, 1985; Forsberg *et al.,* 1992). These requirements can be distilled into a consensus sequence for Thrombin, which mimics its *in vivo* substrate, of Leu-Val-Pro-Arg↓Gly-Ser, with the cleavage site indicated by a downward arrow (Takagi and Doolittle, 1974).

Neither Factor Xa nor Thrombin would enjoy the popularity they currently have for the excision of fusion partners if examples of their successful application did not far outweigh incidences of non-specific or failed cleavages (Jenny *et al.,* 2003). However, such cases do exist in the published literature and thus require careful monitoring of unwanted cleavage reactions to ensure the generation of a truly intact authentic product.

As Thrombin and Factor Xa are almost exclusively isolated from natural sources, some early examples non-specific cleavage could be attributed to the presence of contaminating proteases (Suzuki and Sakuragawa, 1989; Forsberg *et al.,* 1992), but the limited nature of the non-target cleavages suggest aberrant site-specific cleavage in others (Ellinger *et al.,* 1989; Quinlan *et al.,* 1989; Lauritzen *et al.,* 1991). Additionally, with the increased prevalence of commercially available highly purified preparations, it has become clear that cleavage at non-canonical sites is indeed possible by both of these enzymes, as is the case with many claimed site-specific proteases.

It could be reasonably anticipated that Thrombin, with its more permissive recognition sequence requirements, would be more likely to adventitiously cleave the target protein than Factor Xa, but this does not appear to be the case, with reports of cleavage at non-canonical sites well documented for both Thrombin (Chang, 1985; Forsberg *et al.,* 1991; Forsberg *et al.,* 1992; Raftery *et al.,* 1999; Cool *et al.,* 2002; Castagne *et al.,* 2004; Kwon *et al.,* 2005) and Factor Xa (Ellinger *et al.,* 1989; Quinlan *et al.,* 1989; Lauritzen *et al.,* 1991; Zaitseva *et al.,* 1996; Muse and Bender, 1999; Holowachuk and Ruhoff, 1995; Nakashima *et al.,* 1999; Wang *et al.,* 1999; Teng *et al.,* 1999; Mitra *et al.,* 2004).

1.4.4 Mammalian gastric site-specific serine proteases

The classic examples of site-specific digestive proteases are the serine proteases Trypsin and Chymotrypsin, widely used to predicatively fragment proteins into short peptide sequences. Although site-specific, cleavage by these enzymes requires only the recognition of one residue from a class of

amino acids, namely basic (Lys, Arg) for Trypsin and aromatic (Phe, Tyr, Trp) for Chymotrypsin. Although limited examples of the use of these enzymes for fusion protein cleavage do exist (Varadarajan *et al.,* 1985; Dahlman *et al.,* 1989; Wang *et al.,* 1989; Jonasson *et al.,* 1996), their utility for this application is limited as the residues they recognise are common in virtually all proteins.

The most useful enzyme in this category is Enterokinase (EC 3.4.21.9), although it is a gastric serine protease its function is regulatory rather than digestive, with its *in vivo* function being the activation of Trypsin by cleavage of the trypsinogen zymogen to its active form (Maroux *et al.,* 1971). The cleavage site for this enzyme within its natural substrate is C-terminal to the pentapeptide sequence $(Asp)_4$-Lys (Maroux *et al.,* 1971). As Enterokinase cuts C-terminal to its recognition sequence, without requiring the interaction of residues on the other side of the scissile bond, it is capable of generating a native N-terminus for a protein product.

The high charge density of the recognition sequence will increase the likelihood of solvent exposure at the site, assisting protease accessibility and also serving to improve the overall solubility of the fusion protein (Hopp *et al.,* 1988; Prickett *et al.,* 1989). Additional polar residues comprising the epitope for the M1 monoclonal antibody have been added N-terminal to the Enterokinase cleavage site to generate the so called FLAG tag, which further improves fusion protein solubility whilst simultaneously providing an affinity purification mechanism and a means for its removal (Hopp *et al.,* 1988). The improvement in solubility conferred to fusion proteins by such inclusions is clearly demonstrated in the work of Chen *et al.,* (1998), wherein removal of a FLAG tag by Enterokinase digestion of a soluble fusion of the influenza

haemagglutinin subunit HA2 with the FLAG tag caused the HA2 protein to become insoluble.

The unique nature of the cleavage motif should all but preclude its occurrence within a protein product, however, Enterokinase largely recognises the charge density of its recognition sequence rather than the precise amino acid sequence. Cleavage by Enterokinase is possible down to sequences as short as Asp-Asp-Lys (Maroux *et al.,* 1971), with the rate of hydrolysis increasing by the addition of the 3rd and 4th repeat Asp residues (Light *et al.,* 1980). Enterokinase activity is retained with substitution of the canonical motif residues with their charge equivalents (Light and Janska, 1989). Therefore, similar apparent charge densities in the target protein may also be susceptible to Enterokinase cleavage. Further evidence of the mutability of the Enterokinase cleavage site has emerged from recent combinatorial peptide library cleavages, where an S-G-D-R motif was actually cleaved 17-fold faster than the canonical $(D)_4$-K sequence (Boulware and Daugerty, 2006).

Much like the blood-clotting proteases discussed above, early investigations with Enterokinase relied on purified protease extracted from natural sources, and thus non-specific cleavage could be suspected to be due to the likely presence of contaminating proteases (Dykes *et al.*, 1988; Forsberg *et al.,* 1992). Enterokinase is now available as a recombinant enzyme, in many cases as simply the catalytic subunit of the holoenzyme (LaVallie *et al.,* 1993; Choi *et al.,* 2001; Yuan and Hua, 2002) and bearing affinity tags for the removal or immobilisation of the enzyme (Choi *et al.,* 2001). However, many reports of cleavage at non-canonical sites still occur (Forsberg *et al.,* 1992; Nakashima *et al.,* 1999; Coler *et al.,* 2000; Ranatunga

et al., 2001; Tenno *et al.,* 2004; Liew *et al.,* 2005; Yin *et al.,* 2005; Bowman *et al.,* 2006).

Enterokinase, along with Factor Xa and Thrombin discussed in the previous section, represent arguably the three most commonly employed proteases for the site-specific cleavage of fusion proteins, and with few well established alternatives for the generation of native product N-termini, there is little expectation of this role to diminish in the foreseeable future.

1.4.5 *Viral site-specific cysteine proteases*

To isolate novel site-specific proteases, attention has turned to the enzymes of RNA viruses. Upon infection, the genomes of these viruses are translated as one large polyprotein (Allison *et al.,* 1986; Krausslich and Wimmer, 1988). The proteases act to specifically cleave the polyprotein into its individual structural and functional components. A major feature that distinguishes this group of proteases is that they employ a cysteine residue at the core of their catalytic mechanism, as opposed to the serine of the proteases discussed in the other sections (Matthews *et al.,* 1994; Malcolm, 1995). The overall fold of these viral enzymes is very similar to that of the serine proteases, in some cases the active site cysteine can be substituted with serine to achieve an active enzyme, albeit with significantly diminished activity (Lawson and Semler, 1991).

Many viral proteases are highly specific for very long recognition sequences, but the two that have made the greatest impact in fusion protein cleavage are the NIa proteinase of Tobacco Etch Virus (referred to as TEV proteinase or protease) (Parks *et al.,* 1994) and the 3C proteinases of the

picornaviridiae (3Cpro), of which human Rhinovirus (HRV) is a common representative (Walker *et al.,* 1994). The recognition sequence for these enzymes spans at least 7 and 8 residues respectively, with little divergence from the wild-type sequence of the natural polyprotein junctions possible. The minimum cleavage site for TEV protease is of the form E-X-X-Y-X-Q↓(G/S), with a consensus sequence of E-N-L-Y-F-Q↓(G/S) (Carrington and Dougherty, 1988; Dougherty and Parks, 1989). The site for 3Cpro follows a similar general theme, with a consensus sequence of L-E-V-L-F-Q↓G-P (Ivanoff *et al.*, 1986; Cordingley *et al.,* 1990). As can be seen from these sequences, the viral proteases cleave within their recognition sequences and will hence in almost all cases leave a non-natural mono- or dipeptide extension on the N-terminus of the target protein. TEV protease is somewhat flexible in its P_1' requirements, with peptide studies suggesting that it may tolerate Glycine, Serine, Alanine or Methionine at P_1' (Kapust *et al.,* 2002). Emerging structural (Phan, *et al.,* 2002) and empirical evidence of protein substrates (Shih *et al.,* 2002; Tolbert *et al.,* 2005) suggest that the P_1' specificity for this enzyme may be even wider than this. However, most examples of fusion protein cleavage have retained the wild-type substrate P_1' Glycine or Serine (Parks *et al.,* 1994; Routzahn and Waugh, 2002; White *et al.,* 2004; Cabrita *et al.,* 2006; Biedendieck *et al.,* 2007).

High purity recombinant preparations of TEV and HRV proteases are available for fusion protein cleavage. Many manufacturers' implementations of these enzymes also bear an affinity tag to facilitate later removal of the protease from the protein preparation, such as the well known fusion of HRV

3Cpro with GST to give the construct marketed as PreScission protease by Amersham Bioscience (Walker *et al.,* 1994).

Aside from high specificity, the viral proteinases offer activity at very low temperatures, a feature that may promote product stability. For TEV protease, incubation at 4°C results in only a 3-fold reduction in overall activity compared to its optimum temperature of 20°C (Nallamsetty *et al.,* 2004). Recent examples do offer one caveat to the use of either of these proteases, in that incorporation of the relatively hydrophobic cleavage motif into a previously soluble model fusion protein can render it insoluble (Tsunoda *et al.,* 2005; Kurz *et al.,* 2006).

The use of viral cysteine proteases has become very popular for applications that do not demand a native N-terminal sequence, such as structural investigations where the additional residues can be taken into account (Stols *et al.,* 2002; Kristelly *et al.,* 2003).

Owing to the relative novelty of this class of protease for fusion protein cleavage and that their use is understandably limited to applications tolerant of N-terminal substitution, there have been fewer examples of the use of these enzymes than the more traditional proteases of section 1.3.3 and 1.3.4, but to date there have been no reports of non-specific cleavage events. As the body of data grows, developments arising from the continued use of these enzymes will surely be interesting to review in the future.

1.4.6 Engineered site-specificity: H64A subtilisin

H64A Subtilisin (or Genenase I) is unique amongst the proteases discussed in this chapter, as it represents only commercial example of both

bacterial site-specific protease and an enzyme with engineered specificity. The parent enzyme for this rationally designed protease is the broadly specific subtilisin BPN' from the bacterium *Bacillus subtilis* (Carter and Wells, 1987). The mutation of the catalytic triad Histidine residue to Alanine, resulted in a non-functional enzyme. Structural modelling suggested that the functionality of the protease can be restored if the side-chain of the His residue is supplied by the substrate at the P_2 position, this mechanism is known as substrate-assisted catalysis (Carter and Wells, 1987; Carter *et al.,* 1989). It was subsequently determined that the substrate-assisted catalysis effect could also be triggered if the His residue was supplied at P_1' (Forsberg *et al.,* 1991; Matthews and Wells, 1993). Furthermore, an unexpected cleavage site detected by Forsberg, *et al.,* (1991) lacked either a P_2 or P_1' histidine. The unexpected site offered a histidine residue at either P_3 or P_3' and thus it was suspected that the histidine side-chain necessary for substrate-assisted catalysis was supplied by the three dimensional conformation of the substrate.

The modified protease was shown to have greatly improved specificity, with a 400-fold difference in activity observed against substrate peptides that were all processed with equal efficiency by the wild-type enzyme (Carter and Wells, 1987). This selectivity has come at the cost of a catalytically compromised enzyme, with at best an almost 800-fold reduction in activity against cognate substrates compared to the wild-type protease (Carter *et al.,* 1989). The compromised nature of the mutant protease translates to the requirement for high enzyme:substrate ratios and/or long incubation times to effect complete cleavage. Mass ratios as high as 1:10 enzyme:substrate have been reported (Upton *et al.,* 1997), with ratios around 1:50 frequently

required (Degger *et al.,* 2001; Wilkinson *et al.,* 2004; Zhou *et al.,* 2004) and incubations times extending up to 48 hours (Degger *et al.,* 2001; Wilkinson *et al.,* 2004).

Cleaving C-terminal to its ideal recognition sequence, H64A Subtilisin is capable of producing the correct N-terminus for the product. As this sequence is not based around a charged amino acid, as is the case with many of the other serine proteases, H64A Subtilisin offers a considerably different cleavage site and mechanism. It appears that mutations have had little impact on the substrate binding of the other subsites (Carter and Wells, 1987; Carter *et al.,* 1989), which conveniently retain the wild-type specificity. This therefore provides recognition sequence requirements other than the provision of the histidine residue, with a consensus recognition sequence of (Ala/Phe)-Ala-His-Tyr. Although reports of cleavage at unexpected sites does occasionally occur, these can usually be attributed to a proximal histidine residue (Forsberg *et al.,* 1991; Upton *et al.,* 1995; Upton *et al.,* 1997; Degger *et al.,* 2001)

Far fewer examples of the use of this interesting protease exist than for any of the others, presumably due to its relative novelty and the related issues of its high cost (even relative to other site-specific proteases) and single supplier availability.

1.4.7 Strategies to correct suboptimal cleavage

As alluded to in the preceding sections, there are many examples of non-specific cleavage of fusion proteins by several of the otherwise site-specific proteases. Conversely, but not unrelated, is the phenomenon of

the failure of the protease to cleave a fusion protein, despite the presence of its idealised recognition motif (Belmouden *et al.,* 1993; Ko *et al.,* 1993; Sharma and Rose, 1995; Upton *et al.,* 1995; Polyak *et al.,* 1997; Lewis *et al.,* 1999). As such, many strategies have been applied to promote proteolytic cleavage at the intended site.

The first recourse in many situations of suboptimum cleavage is to alter the cleavage reaction conditions, although there are few clear examples of the ultimate success of this approach. Reduction in the reaction pH, ionic strength and non-ionic detergent concentration, minimised, but did not prevent, non-specific cleavage of the ribose transport protein RsbC by Factor Xa (Zaitseva *et al.,* 1996). Likewise, reducing the reaction pH and temperature improved by 50% the recovery of correctly cleaved human parathyroid hormone by Thrombin (Forsberg *et al.,* 1991). By comparison, cleavage with Factor Xa of a Maltose Binding Protein fusion to a domain of the cystic fibrosis transmembrane conductance regulator was not possible despite the presence of the requisite sensitive linker motif until the fusion protein was preincubated in 0.08% SDS (Ko *et al.,* 1993).

One of the simplest, and often most successful solutions to adventitious cleavage of the target protein by one protease has been its subsequent substitution with another site-specific protease, with an accompanying change in the cleavage motif. Substitution of Factor Xa with Thrombin has allowed the recovery of intact glial fibrillary acidic protein (Quinlan *et al.,* 1989), rat Granulocyte Macrophage-Colony Stimulating Factor (Holowachuk and Ruhoff, 1995) and a domain and linker region of the human multidrug resistance gene product (Wang *et al.,* 1999). Enterokinase has

been used in place of Factor Xa to successfully produce phenylalanine hydroxylase (Martinez *et al.,* 1995) and conversely, Factor Xa was used to produce pituitary adenylate cyclase activating polypeptide where Enterokinase had been shown to cleave it non-specifically (Tenno *et al.,* 2004). TEV proteinase replaced Thrombin as the default cleavage system for fusions to bacterial haemoglobin after the latter protease was observed to degrade a number of target proteins (Kwon *et al.,* 2005). Finally, Forsberg *et al.,* (1992) queried six site-specific proteases for the production of an analogue of IGF-I; selecting Thrombin for further study, with H64A subtilisin also showing promise.

It is a truism to state that for site-specific proteases the structure of the substrate will have an impact on cleavage. The above examples stand as evidence that the activity of the various proteases are influenced either positively or negatively by different structural elements in the protein substrate and that non-specific cleavage events are not merely the product of highly protease sensitive regions on the surface of a given protein. The poor performance of one protease against a particular substrate is no indication of general protease susceptibility.

An extension to the concept of the structural influence to cleavage by site-specific cleavage proteases is the use of polypeptide spacers to promote protease activity at the intended site. Improvements in protease activity have been achieved by the inclusion of flexible Glycine-rich linker sequences upstream (Hakes and Dixon, 1992; Holowachuk and Ruhoff, 1995) or downstream (Guan and Dixon, 1991; Rodriguez and Carrasco 1995) of a Thrombin cleavage site. Such sequences have also proven beneficial for the

activity of H64A subtilisin (Polyak *et al.*, 1997). Liew *et al.*, (2007) correctly targeted the activity of Enterokinase to the desired location by inserting the upstream sequence of a favoured internal cleavage site C-terminal to the intended cleavage motif.

1.4.8 Conclusion

The site-specific cleavage of fusion proteins, especially to liberate an authentic target protein, remains a difficult and highly application specific process. Much like the use of fusion partners in general, there is no one universal cleavage system that will work in all situations, but unlike the choice of fusion partners, it is unlikely that there will exist multiple valid solutions to a particularly difficult fusion protein cleavage. With each application presenting its own unique set of requirements and challenges, the issue of fusion protein cleavage demands careful attention, particularly as it applies in an industrial or biopharmaceutical setting.

1.5 α-LYTIC PROTEASE

An example of an emerging engineered site-specific protease can be found in α-Lytic protease. α-Lytic protease (E.C. 3.4.21.12) is a 20kDa serine protease first isolated as an extracellular enzyme from the soil bacteria *Lysobacter enzymogenes* (formerly *Myxobacter* 495) (Whitaker, 1970). Displaying the characteristic chymotrypsin-family double β-barrel fold (Brayer *et al.*, 1979; Fujinaga *et al.*, 1985; Fuhrmann *et al.*, 2006), and an accessible hydrophobic cleft substrate binding pocket (Mace and Agard, 1995), α-Lytic

protease has long been a model system for the relationship between structure and function in this class of enzymes. Many high-resolution crystal (Brayer *et al.,* 1979; Fujinaga *et al.,* 1985; Bone *et al.,* 1987; Bone *et al.,* 1991a; Rader and Agard, 1997; Fuhrmann *et al.,* 2004; Fuhrmann *et al.,* 2006) and solution (Bachovchin, 1986; Bachovchin *et al.,* 1988; Davis *et al.,* 1997; Ivanov *et al.,* 2002) structures are available for both the free and inhibitory substrate bound enzyme.

The wild-type enzyme has high activity and well defined substrate preference: favouring small, uncharged residues such as Alanine at P_1 with activity falling almost linearly with increased side-chain size (Kaplan *et al.*, 1970; Bone *et al.*, 1989). This suggests a role for the protease in its native host, of degradation of proteins as a food source and elimination of competing microorganisms.

At face value, α-Lytic protease may not seem a particularly useful candidate for fusion protein processing, with a preference for cleavage after frequently occurring residues. Aside from this apparent drawback, it does have the virtues of being highly stable in both wild-type and mutant forms (Graham *et al.*, 1993; Jaswal *et al.,* 2002) and is well expressed in *E. coli* at high levels (Haggett *et al.,* 1994c; Unutmaz *et al.*, 1997) with simple downstream purification (Haggett *et al.,* 1994a; Haggett *et al.,* 1994b). Additionally, P_1 specificity is not the only consideration, as the binding of peptide substrates to α-Lytic protease has been shown to involve at least as many as six subsites (Bauer *et al.*, 1981); and a large amount of background information concerning the structural determinants behind α-Lytic protease's

S_1 specificity is available. This wealth of structural and subsite data has greatly facilitated both rational and combinatorial mutagenesis of this enzyme.

1.5.1 Rational α-Lytic protease mutants

The true utility of α-Lytic protease began to be realised in studies by Bone *et al.* (1989) that attempted to relate structure to substrate specificity. In this work two methionine residues in the S_1 active site were mutated to alanines, with surprising results. The M190A and M213A mutants display drastically altered substrate specificity profiles. The side-chains of these methionine residues effectively face each other across the active site, each occurring on one of the two polypeptide segments that line the S_1 active site pocket (Fujinaga *et al.,* 1985). Both mutations were expected to increase the size of the active site by the same volume, but with different geometry.

The M213A mutant shows almost a total loss of the ability to discriminate between P_1 resides, with only a 6-fold difference in the activity against the best (Met) and worst (Leu) P_1 residues, as opposed to the over 55,000-fold variation observed for the wild-type enzyme's discrimination between its most favoured (Ala) and least favoured (Phe) residues. The catalytic ability of this mutant is severely compromised though, with a reduction in activity of 35-fold against equivalent Ala substrates (Bone *et al.,* 1989).

By comparison, the M190A mutant appears the more interesting of the two, with its substrate specificity almost the opposite to that of the wild-type enzyme. Instead of preference for small residues only, with activity falling with P_1 residue size, this mutant prefers to cleave after very large resides. An

800,000-fold increase in activity toward P_1 Phenylalanine is observed at the cost of only a 2-fold loss of activity against P_1 Alanine (Bone *et al.,* 1989). Considerable discrimination between resides has therefore been lost for this increased activity, with a 115-fold difference between the best (Met) and the worst (Val) substrates tested as opposed to the aforementioned 55,000-fold of the wild-type.

The surprise comes in that the altered specificity is not solely attributable to steric considerations. X-ray crystallography shows the S_1 active site of wild-type α-Lytic protease to be shallow, with the side chain of Met190 occluding what would otherwise be a deep binding pocket (Fujinaga *et al.,* 1985). If the altered specificity profile of M190A was due to the reduction of steric hindrance then the increase in the size of the P_1 residue accommodated would only be to a residue as large as methionine, as was expected. Instead, the altered specificity has been shown to be a result of increased active site plasticity, with the M190A S_1 active site more flexible than that of the wild-type. Structural analysis of the M190A mutant with bound peptide boronic acid inhibitors shows that the side-chains of other S_1 residues have more conformational freedom than observed for the wild-type enzyme (Bone *et al.*, 1991b). Mathematical and structural modelling suggests that the walls of the S_1 active site are capable of asymmetrical motion in the M190A mutant, whereas those of the wild-type enzyme are strictly symmetrical (Miller & Agard, 1999; Ota and Agard, 2001). This would have the consequence of a fixed active site volume in the wild-type, with the M190A mutant instead able to adopt varying volumes, according to the P_1 residue being accommodated.

1.5.2 Combinatorial α-Lytic protease mutant libraries

Building on the findings of the M190A mutant, Graham *et al.* (1993) and Graham *et al.* (1994) developed of two libraries of α-Lytic protease mutants. These libraries both involve combinatorial mutation to four residues, two on each of the two polypeptide segments (residues 190 – 194 and 213 – 226, Fujinaga *et al.*, 1985) that line the S_1 active site of α-Lytic protease.

The first library (designated PragA(X), where X is the number of the mutant (Graham *et al.*, 1993) was based on the M190A mutant, with this mutation held constant. The residues Gly191, Arg192, Met213 and Val218 were subjected to combinatorial mutagenesis. As a result of the procedure used for the generation of the library, many of the active enzymes produced were a product of mutagenesis to either one of the polypeptide segments only.

The second library (labeled as RadX(X), with X again representing the mutant number (Graham *et al.*, 1994), used the observation that all the active enzymes (except one) maintained the wild-type glycine at position 191; this residue was held constant with Met190 the subject of combinatorial mutagenesis in its stead. Additionally, the mutagenesis technique was altered so that all active enzymes generated had to be a result of both polypeptide segments being modified.

The active mutant enzymes of both libraries can be divided into three groups according to their preferred substrate P_1 residue, determined by overlay experiments consisting of short peptide sequences with chromogenic or fluorogenic leaving groups. One group contains those enzymes that have a similar substrate specificity profile to the M190A mutant. The second group

(which is largely a phenomenon of the second library) again displays a similar profile to M190A, but with an increase in activity toward Alanine at the expense of Phenylalanine. The third group is perhaps the most interesting, with increased activity toward His and Met at P_1 to the exclusion of other residues. With this third group considered, α-Lytic protease has come full-circle: from an enzyme specific for common residues, to a relatively non-specific enzyme with the M190A mutation, back to a very specific protease, favouring rarer amino acids and thus it has become a potentially attractive candidate for fusion protein processing.

1.5.3 The other subsites of α-Lytic protease

As mentioned above, substrate recognition by α-Lytic protease is actually thought to take place over at least six subsites, S_4 – S_2', with contacts extending well beyond even this range. Hydrolysis of peptides of increasing length shows that there is an exponential increase in enzyme activity in going from one residue to four residues N-terminal to the scissile bond, with negligible activity increase observed with the addition of a P_5 residue (Bauer *et al.*, 1981), although structural models suggest the enzyme makes substrate contacts up to this position (Bone, *et al.,* 1991a). Additionally, Schellenberger *et al.* (1994) showed that the rate of acyl-transfer reactions was improved from undetectable for one P_1' residue to greater than 15-fold with four P' residues. No improvement was observed with the addition of a P_5' residue. This data suggests that substrate binding, but not necessarily for substrate recognition or discrimination, can require up to 9 subsites. Binding to the subsites outside of S_4-S_2' improves catalysis by α-Lytic protease, but these subsites are very

permissive, accepting equally well all of the limited number of amino acids that have been presented in these positions in substrates.

Given the general degradative role of α-Lytic protease in its host organism, there is no single wild-type evolutionary selected long recognition motif to draw on, as for the regulatory proteases of the previous section. It has in fact been suggested that it is advantageous for a digestive enzyme to have broad specificity at the non-S_1 substites, a trend established in other digestive proteases such as trypsin, chymotrypsin and elastase (Davis and Agard, 1998).

Although limited data exists on the S_4-S_2 specificity of α-Lytic protease, the long employed empirical (Ala/Pro)-Ala-Pro sequence (Bauer *et al.,* 1981; Bone *et al.,* 1989; Epstein and Abeles, 1992; Graham *et al.,* 1993) offers many advantages. Whilst Proline is not the ideal P_2 or P_4 residue in α-Lytic protease substrates, it is well tolerated (Epstein and Abeles, 1992; Lien *et al.,* 2003), but it is not accepted at either P_3, P_1 or P_1' (Bauer *et al.,* 1981; Schellenberger *et al.,* 1994; Lien *et al.,* 2003). The use of Proline at P_2 and P_4 in substrates thus constrains the protease into cleavage in one potential binding mode, providing confidence in the site of cleavage on peptides or proteins (Bauer *et al.*, 1981).

The specificities of the S' subsites of α-Lytic protease have received far less attention, but are of the most importance to industrial proteolysis. The comprehensive study by Schellenberger *et al.* (1994) utilised acyl-transfer reactions to map the individual S' subsite preferences of α-Lytic protease. The S_2' subsite was shown to be the most selective of the three probed, with Leu strongly favoured over all others. S_1' was more tolerant, catalyzing more

than half the substrates evaluated with similar activity. The S_3' subsite was shown to be the most permissive, with only two residues unfavourable at this position. The only generalisations that can be made about the S_1' and S_2' subsites are that they have a strict requirement for larger, non-acidic, sterically normal residues, with Gly, Ala, Asp, Glu and Pro among the worst substrates for both. The only significant requirement for S_3' is for residues to be larger than Ala. In direct contrast to these findings, cleavage of the peptide Ac-PAPA↓A-NH_2 has been shown to be twofold higher than Ac-PAPA↓F-NH_2 or Ac-PAPA↓G-NH_2 (at almost equivalent activity) (Bauer *et al.*, 1981). The comparison of these two sets of findings serves to illustrate that proteolysis (particularly by α-Lytic protease) involves the interactions of binding at multiple subsites.

1.5.4 Conclusion

Providing high level recombinant expression, simple purification and high stability, α-Lytic protease offers many features desirable in a protease for an industrial process. With engineered S_1 specificity for uncommon amino acids and the likelihood of substrate recognition extending over six subsites, the full potential of this enzyme in this role is emerging.

1.6 CLEAVAGE OF IGF FUSIONS

Just as the expression of recombinant IGF's can be seen as a showcase microcosm of fusion protein technologies, the cleavage of these fusions has also seen the application of a wide variety of site-specific

cleavage systems. With the specific case of IGF-I aside (discussed in section 1.5.1), IGF fusion proteins have been cleaved successfully on many occasions by both chemical agents and site-specific proteases.

Lacking an internal Met residue, cleavage of fusion proteins of IGF-II from various species by cyanogen bromide has been favoured for many production efforts (Furman *et al.,* 1987; Hammarberg *et al.,* 1989; Misoka *et al.,* 1989; Rhee *et al.,* 1990; Hammarberg *et al.,* 1991; Nilsson *et al.,* 1991; Sakano *et al.,* 1991; Luthi *et al.,* 1992), with yields ranging from 0.5% to 65%, although it seems that the apparent variability is due to the reporting of either purified yields or raw cleavage productivity, rather than inherent variation in the cleavage reaction. The rise in prominence of site-specific proteases has seen a shift towards their adoption for the cleavage of these fusions. The use of H64A subtilisin has resulted in 50-100% cleavage productivity for human IGF-II, with a 17% purified product yield obtained (Francis *et al.,* 1993), an outcome that was mirrored for salmon IGF-II with the same protease delivering a 20% yield of purified protein product (Wilkinson *et al.,* 2004). In the case of human IGF-II this has been further improved by the use of a combinatorial mutant of α-Lytic protease, PragA9 (*i.e.:* mutant number 9 of the first library of Graham *et al.,* 1993) to near 100% correct product liberation using less enzyme and a shorter incubation than the previous example (Lien *et al.,* 2001).

An IGF-I analogue lacking the first three residues, des(1-3)-IGF-I has seen a similar paradigm shift, with an early example of its liberation from a ZZ domain fusion protein by cyanogen bromide chemical cleavage reporting a ~30% cleavage yield, but only 0.8% purified yield (Forsberg *et al.,* 1989). This

was improved by the use of site-specific proteases, with Thrombin and H64A subtilisin both improving the raw cleavage yields to 60% and 100% respectively (Forsberg *et al.,* 1992).

1.6.1 Cleavage of IGF-I fusions

The variety of successful site-specific cleavage solutions enjoyed by the other IGF's has not been echoed in IGF-I, with only a select number of strategies proving useful. Chief amongst these is cleavage by chemical agents. Cleavage of IGF-I fusion proteins has been attempted by a variety of chemical methods including cyanogen bromide, despite an internal Met (Saito *et al.,* 1987) and acid hydrolysis (Nilsson *et al.,* 1985), but by far the prevailing method is the use of hydroxylamine (Moks *et al.,* 1987; Forsberg *et al.,* 1990; Canova-Davis *et al.,* 1992; King *et al.,* 1992; Hodgson *et al.,* 1996; Kim and Lee, 1996; Antorini *et al.,* 1997). Taking advantage of the N-terminal glycine residue, the application of this agent requires only the insertion of an asparagine immediately N-terminal to the IGF-I sequence to create the cleavage motif. The yields of such cleavage reactions vary from 5 – 80%, with again the main reason for the seeming variability being that the outcomes reported vary between raw cleavage reaction output and the yield of purified IGF-I recovered.

The use of hydroxylamine for IGF-I fusion protein cleavage is not without significant disadvantages, despite its apparent efficiency and specificity. The harsh chemical environment required for hydroxylamine cleavage to occur (*e.g.*: 3M hydroxylamine, pH > 9, 45°C, 4 hours (Bornstein and Balian, 1977; King *et al.,* 1992)), has been known to result in both

oxidation of the single methionine of IGF-I and the formation of hydroxymates of internal amino acids. Three hydroxamation sensitive residues in IGF-I have been identified; Gln^{15}, Asn^{26} and Gln^{40}, with their abundance in a standard fusion protein cleavage reaction potentially as high as 15% of total IGF-I (Canova-Davis *et al.,* 1992). Additionally, in the monitoring of hydroxylamine cleavage of a Z domain-IGF-I fusion, Antorini *et al.,* (1997) showed that Met^{59} sulphoxide and Gln^{15} hydroxamate modified forms may each be as high as 11 and 13% (respectively) of the total IGF-I liberated. Whilst such modified forms may be tolerated for many applications, especially given the biological equivalency of hydroxamated IGF-I and authentic IGF-I in cultured mammalian cells (Canova-Davis *et al.,* 1992), they are not acceptable in a high purity product and untenable in a potential biopharmaceutical.

Whilst not strictly an *in vitro* fusion protein cleavage approach, a popular and effective IGF-I production strategy has been its fusion to the mating factor α1 pre-pro sequence of *Saccharomyces cerevisiae* (Bayne *et al.,* 1988; Cascieri *et al.,* 1988; Bayne *et al.,* 1989; Elliott *et al.,* 1990; Chaudhuri *et al.,* 1992; Vai *et al.,* 2000). This approach allows recovery of authentic, soluble IGF-I from the yeast culture supernatant, as the leader fusion is cleaved off during export by the endogenous Kex2 protease (Elliott *et al.,* 1990). Yields from this system have been variable, ranging from 65 µg to 56 mg per litre of ferment. Aside from this variability, this system is mostly limited to soluble expression and yeast host, a disadvantage of which being that expression of IGF-I in yeast does pose the problem of potential glycosylation (Gellerfors *et al.,* 1989), as natural human IGF-I is unglycosylated (Rinderknecht and Humbel, 1978). Although this final issue

can be avoided by the use of yeast strains with deletions or mutations to glycosylation pathways (Finck *et al.,* 1996).

The various successful proteolytic cleavages of other IGF fusion proteins do serve to demonstrate the overall compatibility of proteins of this fold to site-specific proteolysis. Conversely, some examples also highlight the possible vulnerability of such proteins to non-specific proteolysis, as in the degradation of des(1-3)-IGF-I and salmon IGF-II by Enterokinase (Forsberg *et al.,* 1992; Wilkinson *et al.,* 2004).

Only two examples of proteolytic production of IGF-I have been reported to date, and only one of these by endoproteolysis. This is of little surprise, as the second amino acid of full-length human IGF-I is Proline, a residue which at this position is refractile to the action of many site-specific proteases. For example, at least two of the proteases discussed earlier that are capable of generating a native N-terminus for most protein products display this tendency. Factor Xa retains only 5% of its maximum activity against peptides displaying Proline at P_2' (Ludeman *et al.,* 2003) and catalysis by H64A subtilisin is prohibited against substrates containing Proline at this location (Carter, 1990).

In the first example of proteolytic IGF-I production, the dipeptidyl aminopeptidase Cathepsin C was used to convert a two residue N-terminally extended form of IGF-I (AE-IGF-I) to authentic IGF-I (Hejnaes *et al.,* 1992). Although no data on the recoveries from technique were presented, even assuming quantitative efficiency the use of this approach would most likely be limited to relatively short, even numbered extensions that lacked any of the terminators of this enzyme's activity such as Lys, Arg or Pro (McDonald *et al.,*

1969). Hence in this case the N-terminal sequence of IGF-I itself provided the required terminator.

To date the only example of *in vitro* site-specific endoproteolysis of an IGF-I fusion protein has been the use of Thrombin to release IGF-I from a human growth hormone truncation fusion partner (Nishikawa *et al.,* 1987). Whilst an 80% crude cleavage yield was stated, this could only be translated into a 3% yield of purified IGF-I. These findings have not been replicated to date, but even if borne out, the lack of a commercial source of recombinant enzyme would make its application in a biopharmaceutical production process questionable. Interestingly, cleavage of an equivalent fusion was attempted with Factor Xa without success in the same investigation, with no proteolysis products observed.

The IGF-I fusion proteins exemplify many of the difficulties faced in site-specific cleavage of fusion proteins. Displaying a susceptibility to both chemical modification and non-specific proteolysis and an N-terminal sequence refractile to the activity of many proteases, efficient IGF-I liberation remains a challenge.

1.7 Cleavage of the Met-pGH(1-11)VN-IGF-I fusion protein with α-Lytic protease: This project.

With the dramatic success of one α-Lytic protease combinatorial mutant in the release of IGF-II and des(1-3)-IGF-I from the Met-pGH(1-11) fusion partner (Lien *et al.,* 2001; G. Francis, personal communication), it is of significant industrial interest to determine if a proteolytic approach can be

devised for the equivalent fusion of full-length IGF-I. Such improvements are not without parallel in other Met-pGH-IGF fusion protein cleavages. Cyanogen Bromide was for a long time favoured for the cleavage of IGF-II fusions, but this role has been supplanted by the advent of site-specific proteases that release this protein efficiently with no adventitious cleavage, firstly with H64A subtilisin then later with PragA9 α-Lytic protease. If a similar enzyme system could be devised for IGF-I fusion proteins, the issues surrounding the use of hydroxylamine could be avoided, which as stated previously, includes the potential for chemically modified variants of the product which necessitates extensive downstream purification to achieve a homogeneous preparation. Additionally, the hydroxylamine cleavage of this fusion is not very efficient, with only 20% of the fusion protein cleaved in a typical reaction (King *et al.,* 1992). The cumulative effect of the poor raw cleavage output and a rigorous purification regime translated to a 0.5% process yield of purified authentic human IGF-I (King *et al.,* 1992).

The advantages for industrial application present in the parent α-Lytic protease provides the impetus to identify enzymes within the mutant libraries able to cleave the expressed Met-pGH(1-11)-IGF-I efficiently, and yield authentic human IGF-I.

The theoretical impediment to cleavage of full-length IGF-I with an α-Lytic protease mutant is the difficult N-terminal tripeptide sequence, Gly-Pro-Glu. Schellenberger *et al.* (1994) shows these three residues are individually unfavourable for wild-type α-Lytic protease activity in the positions that they occupy in IGF-I. Gly at P_1' displays 12% of the activity of the best substrate and Glu at P_3' has ≈50% of maximal activity (as stated previously,

S_3' is quite permissive). Most troubling is the fact that the reaction with P_2' Pro was almost below the minimum detection level.

However, there is some tantalising evidence that the S' enzyme-substrate interactions may be altered in the α-Lytic protease mutant libraries. The evidence comes from the substrate overlay experiments used initially to determine the S_1 substrate specificity of the mutants where P_1 Ala substrates with two different leaving groups were tested. It was found that differences existed in the mutant proteases' ability to cleave before these leaving groups. This difference would not normally be surprising, as various P' entities could be expected to be cleaved at different rates. However, the discrepancy in the rates between the two P' moieties was not constant for all the mutants, as would be the case if the S' subsites were unaffected by the S_1 mutagenesis (Graham *et al.,* 1993).

Additionally, the S' mapping experiments of Schellenberger *et al.* (1994) dealt with each position in isolation, with Ala making up the remainder of the peptide. No research into potential cumulative effects of multiple, different P' residues in α-Lytic protease has been carried out to date, or known is the relative importance of the P' residues compared to P sequences. Hence it is unknown whether a highly favourable P motif can overcome the effect of unfavourable P' residues. In short, there remains much to be understood in regard to the S' binding interaction of this increasingly prominent enzyme.

2.1 INTRODUCTION

In Chapter 1 the need to examine α-Lytic protease for IGF-I processing was established, so too was the improbability of achieving correct cleavage at the required sequence with the wild-type enzyme. If IGF-I cleavage by α-Lytic protease is indeed possible, the required cleavage specificity will most likely come from a member of the two libraries of S_1 active site mutants. Whilst it is known that these mutants all possess varying substrate specificity and intrinsic activity, it is also suspected that they display a high degree of variability in either their inherent expression level or intermediate term stability (Graham *et al.,* 1993).

The intrinsic activity, or catalytic efficiency, of an enzyme can be expressed as its turn-over number. This is defined as the number substrate molecules one molecule of enzyme can convert to product per second. Thus in the case of proteases the turn-over number is therefore the number of peptide bonds it can hydrolyse in a second.

It can be thought that given an excess of protease activity cleavage will be observed at alternative cleavage sites that would otherwise not be significantly cleaved, as discussed below. In this specific example of α-Lytic protease, the wild-type and select mutant enzymes have their preferred substrate sites (as discussed in Section 1.5.2), but are also known to cleave after other residues, albeit at a lower rate. Where cleavage at an unfavoured site may be so slow that it is not observed under certain circumstances, given more protease activity (or more time) this cleavage may become observable (Haggett *et al.,* 1994c; Lien *et al.,* 2001). Additionally, once the high level of protease activity has catalysed hydrolysis at all favoured sites and substrate

therefore becomes limiting, the unfavoured site provide available alternative targets.

If the α-Lytic protease library mutants are to be assessed for their ability to specifically cleave a given protein target, it is therefore of utmost importance to be able to compare the activity of equivalent amounts of protease protein present in cultures of the mutants. This will allow for meaningful comparison of the activity of the individual mutants against that specific substrate target. As the mutants vary in their intrinsic total activity (or turn-over number) and their absolute substrate specificity, such normalisation should be as generic as possible to avoid introducing prematurely undue bias into the determination by applying any preconceptions about the substrate requirements of the mutants. Given the large number of mutants to be investigated (126), it is desirable to screen them at an early stage of purification, with as little sample pretreatment as possible. Unfortunately, an affinity tag for convenient single-step purification was not included in the expression construct for the library enzymes. It would be useful to apply an iterative screening process where initially a protease activity normalisation assay, which functions in crude bacterial culture supernatants, allows selection of desirable mutants prior to any demanding purification steps.

Determination of protease activity by quantitation of protease protein concentration is one means to ensure comparability for subsequent specificity investigations. Given appropriate antisera, immunological detection of protease protein, by either immuno-blot or immunoassay, can function in very crude samples with no necessity for sample pretreatment. Whichever approach is employed, immunological techniques usually involve binding of

the protein samples to a solid support, followed by washing away of unbound material. This has the benefit of rendering such approaches relatively impervious to interference or high backgrounds from contaminating crude sample constituents. The most significant problem in such protease concentration determination is the potential inability to discriminate active protease from inactive or degraded protease material. Such inactivation may be accompanied by a change in the antigenicity of the enzyme, but this will not necessarily occur in every instance, nor can such an effect be readily incorporated into an antibody-antigen interaction by design. This issue is particularly noteworthy when α-Lytic protease is considered, given that production of functional protease involves an obligatory pro-region mediated folding (Silen *et al.,* 1989) into a non-ground state thermodynamic configuration (Baker *et al.,* 1992; Cunningham *et al.,* 1999; Cabrita and Bottomley, 2004). It is quite conceivable that this folding pathway could lead to the presence of misfolded, inactive α-Lytic protease isoforms. This has indeed been demonstrated in attempts to express α-Lytic protease at temperatures over 30°C, which interrupts the autoexcision of the mature protease from its inhibitory pro-region (Silen *et al.* 1989; Fujishige *et al.* 1992). An assay of specific activity needs the ability to exclude such misfolded forms or otherwise degraded inactive protease from the reported value, given that they will not contribute protease activity.

Measurement of actual protease activity is much more straightforward in this analysis, wherein by definition only mature, active α-Lytic protease is measured. In crude samples the presence of contaminants could inhibit the measurement of enzyme activity, but the assumption made is that any such

inhibition will be uniform across all mutants. Protease activity assays fall into two broad classes: the measurement of hydrolysis of either discrete peptides or whole proteins, each coupled with either colourimetric or fluorescent detection. Given the largely unknown substrate preference of the α-Lytic protease library mutants, the use of discrete peptides of restricted sequence for an activity assay could potentially bias any determinations. The limited number of potential cleavage sites in any given discrete peptide may not present a favourable motif for the detection of all the protease mutants available for screening. Assuming only a tetrapeptide minimum cleavage motif, known to be the minimum required for meaningful detection of α-Lytic protease activity (Bauer *et al.* 1981), complete randomisation of 20 amino acids at all four sites would need a 1.6×10^5 member peptide library. Whilst this number could be reduced by applying some of the shared substrate preferences of the α-Lytic protease library mutants, such as their shared ability to cleave after a Succinyl-Ala-Pro-Xaa motif (Graham *et al.* 1993; Graham *et al.,* 1994). It has been shown that optimum substrate preference of α-Lytic protease mutants cannot be predicted rationally, despite molecular modelling based on the available substrates (Lien *et al.*, 2003). Additionally, constraining protease activity to a single cleavage motif will artificially select against broad specificity mutants. Given that it is precisely these mutants that an activity normalisation strategy is intended to highlight, any such negative selection would undermine the power of the assay. It is therefore apparent that a peptide based activity assay would be too constraining and require too many assumptions regarding the substrate specificity of the mutants be made to provide a realistic generalisation of protease activity.

The use of a whole protein for activity determination overcomes many of the concerns regarding the introduction of bias towards a narrow range of peptide linkages, and provides a practical solution to activity normalisation. Not only are a great variety of overlapping sequence motifs presented simultaneously as potential proteolytic sites, they are presented in a natural protein three dimensional structure. The use of a whole protein substrate has the added advantage of offering a mechanism to account for the action of broad specificity mutants. These mutants will most likely act at a number of sites within a protein substrate, increasing the cleavage response accordingly. Not inconsequentially, the reagents for whole protein assays are readily generated in-house, without specialised synthesis equipment or complex chemistries, offering a considerably lower cost per assay and greater renewability than peptide-based methods.

The detection in protease assays may be either colourimetric or fluorescent. In the majority of cases, the generation of a colourimetric signal requires the liberation of a free chromophore with an enhanced extinction coefficient, thus limiting their utility to peptide based assays where they are located C-terminal to the proteolytic cleavage site. One approach for the use of chromogenic detection of proteolysis that does not require the liberation of a free chromophore, and thus could be applied to the monitoring of cleavage of a whole protein has been reported involving the use of trinitrobenzenesulphonic acid (TNBSA) (Hatakeyama *et al.*, 1992). Hydrolysis of the peptide backbone of a protein will generate a new primary amine at the N-terminus of the newly liberated peptide fragment cleavage product. Reaction of TNBSA with such primary amines results in the formation of a

coloured product. However, for such an approach to be successful, the already present primary amines in the sample and substrate protein would need to be derivatised or blocked in some other fashion to prevent their own reaction with the TNBSA and the generation of a high degree of background coloured signal. This approach would thus require highly modified reagents, which may no longer provide a reliable model of the activity of the natural enzyme upon its substrates.

Fluorescent technologies hold more promise for monitoring of protease cleavage of whole proteins, since they do not require the liberation of a free fluorophore for a signal to be observed. Two fluorescence technologies have been developed in recent years that offer the ability to monitor proteolysis of whole proteins, these being Fluorescent Resonance Energy Transfer (FRET) and Fluorescence Polarisation (FP). FRET technology makes use of two fluorophores, where the emission wavelength spectrum of the first (the donor) overlaps the excitation wavelength spectrum of the second (the acceptor). Any excitation energy supplied to the donor fluorophore is passed in a non-radiant manner to the acceptor. In this way the fluorescent signal of the donor fluorophore is effectively "quenched". In the protease assay scenario, the donor and acceptor fluorophores are spatially proximal to one another in the intact protein, prior to protease treatment. Upon exposure to the excitation wavelength of the donor fluorophore no emission is detected due to intramolecular quenching of the excitation energy. Where the protease has cleaved the labelled protein, the donor and acceptor fluorophores are physically separated and an excited donor fluorophore emits fluorescence in order to return to its energetic ground state. This fluorescent signal will be

proportional to the degree of proteolysis of the labelled protein or peptide (Matayoshi *et al.,* 1990). Since the spatial proximity required for FRET quenching is in the order of 1-10nm (Stryer and Haugland, 1967), multiple donor/acceptor pairs will be required for effective coverage of a whole protein molecule. Many examples of whole protein FRET protease substrates are generated by incubation with a 500x molar excess of fluorophore to maximise label incorporation (Voss *et al.,* 1996). The addition of a large number of bulky fluorophores to the surface of a protein does suggest at the possibility for masking potential enzyme cleavage sites by alteration of the amino acid side-chains surrounding a potential cleavage site in the unmodified protein. Such high degrees of labelling could also make the generation of equivalently quenched protein difficult to reproduce consistently, leading to the potential for significant batch-to-batch variation.

Another leading application of recent advances in fluorescence technology is fluorescence polarisation. The principle underlying this technique is that a theoretically immobile fluorophore, when excited with plane polarised light, will emit correspondingly plane polarised light. In reality, molecules are in constant motion and so when a population of fluorophores is excited by plane polarised light some molecules will have moved within the fluorescence lifetime of the fluorophore (the time between excitation and photon emission). This results in some depolarisation of the emitted light, which is detected as an increased signal in the plane perpendicular to that of excitation. At constant temperature the extent of the molecular motion that the fluorescently labelled species undergoes, and thereby the degree of depolarisation that is observed, is proportional to the size of that molecule

(Bolger and Checovich, 1994). The utility of this technology in a protease activity assay is clear, as the protease acts on a labelled whole protein substrate resulting in a reduction in the size of the resultant labelled proteolytic fragments. This has been applied on many occasions to quantitatively measure protease activity (Spencer *et al.* 1973; Maeda, 1979; Bolger and Checovich, 1994; Schade *et al.* 1996; Grys *et al.* 2000;) in complex biological samples ranging from bacterial lysate to dental plaques. Since FP is expressed as the ratio of the two planes of emitted light, it is largely insensitive to the total fluorescence yield of the samples. It is therefore not subject to interference from crude biological samples such as autofluorescence of signal quenching. Substrates for FP protease assays are readily synthesised and require only one label per protein molecule. The lower levels of labelling of FP substrates are actually desirable since some common fluorophores, such as fluorescein, have overlapping absorption and emission spectra, meaning that these fluorophores are capable of self-quenching of the fluorescent signal and potentially interfering with FP.

Given the requirement for a whole protein based activity assay that can operate in bacterial culture supernatant and with substrate that can be readily generated in-house, a fluorescence polarisation assay seems a logical choice. It is therefore the objective of the work in this chapter to quantify the level of α-Lytic protease activity in the crude culture supernatants of the library mutants. As appropriate α-Lytic protease antiserum becomes available, the activity data can be compared to the level of specific protease protein present.

2.2 MATERIALS

2.2.1 General reagents & materials

The following reagents were obtained from Sigma-Aldrich, Castle Hill, NSW, Australia: fluorescein isothiocyanate (FITC), bovine β-casein, bovine serum albumin (RIA grade), dimethyl sulphoxide (DMSO), ethylenediaminetetraacetic acid (EDTA), o-phenylenediamine (o-PD) and hydrogen peroxide. BDH chemicals, Kilsyth, Victoria, Australia, supplied: D-glucose, boric acid, N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES), NaCl, KCl, Na_2CO_3, $NaHCO_3$, citric acid, Na_2HPO_4, KH_2PO_4, polyoxyethylene (20) sorbitan monolaurate (Tween-20). All reagents were analytical grade unless otherwise specified.

PD-10 desalting columns were supplied by Amersham Bioscience (Uppsala, Sweden), as was donkey anti-rabbit-IgG horseradish peroxidase conjugate.

SDS-PAGE materials; NuPAGE Novex 4-12% Bis-Tris 1mm minigels, NuPAGE 4X LDS Sample buffer, NuPAGE MES SDS 20X Running Buffer, Mark12™unstained molecular weight standard, were supplied by Invitrogen, CA, USA.

Succinyl-Ala-Ala-Pro-Met-p-nitroaniline (suc-AAPM-pNA) peptide was synthesised by Bachem AG, Bubendorf, Switzerland.

2.2.2 Purified α-Lytic protease standard.

Purified PragA9 mutant α-Lytic protease was supplied by GroPep Ltd. It was quantified as 0.6 mg/ml by absorbance at 280 nm (ε = 17335 $M^{-1}.cm^{-1}$)

and stored at 4°C in 25 mM Na Tricine buffer (as in section 3.2.5 and 3.3.2), pH 6 with 0.02% (w/v) sodium azide.

2.2.3 Bacterial strains

E. coli JM109 (*recA*1 *supE*44 *endA*1 *hsdR*17 *gyrA*96 *relA*1 *thi* Δ(*lac-proAB*) F' [*traD*36 *proAB*$^+$ *lacI*^{q} *lacZ* Δ M15] clones bearing plasmid pBS-*pelB*-PROMAT that encoded the mutant α-Lytic protease libraries were kindly provided by Dr L. Graham, CSIRO, North Ryde, Sydney, Australia.

2.2.4 Bacterial culture media

Media were prepared with Milli-Q ultrapure water and autoclaved prior to the addition of sterile-filtered antibiotics. Luria-Bertani (LB broth) liquid medium contains 1.0% (w/v) tryptone, 0.5% (w/v) yeast extract, 85 mM NaCl and 100 mg/L ampicillin. LA100 agar plates were prepared by addition of 0.5% (w/v) bacteriological agar to LB broth. Skim milk agar plates were prepared by addition of 2% (w/v) skim milk powder and 1% (w/v) bacteriological agar to LB broth. Terrific broth (TB broth) liquid medium contains 1.2% (w/v) tryptone, 2.4% (w/v) yeast extract. 0.8% (w/v) glucose, 55 mM K_2HPO_4, 16 mM KH_2PO_4, and 50 mg/L ampicillin. Tryptone, yeast extract, bacteriological agar and ampicillin were purchased from Sigma-Aldrich, Castle Hill, NSW, Australia.

colony and grown overnight at 25°C in an orbital shaking incubator (Ratek Instruments model OM11, Victoria, Australia) at 200 rpm. The seed cultures were harvested by centrifugation at 400 g for 5 mins at 4°C (Sigma Laborzentrifugen GmbH model 6K15, Germany). The pellets were resuspended gently in 5 ml sterile saline (0.9% (w/v) NaCl) at 4°C. The bacterial cells were centrifuged as above and the pellets again resuspended in 6.6 ml fresh sterile saline. 500 µl of the washed bacterial cells were used to inoculate 50 ml TB expression cultures.

The expression cultures were grown at 25°C for 10 to 12 days. α-Lytic protease production was monitored by the inclusion of a control flask containing an expression culture of the PragA9 mutant, with the activity in this flask was measured by colourimetric peptide assay (see Section 2.3.2). Expression cultures were harvested when protease levels in the control flask had not altered for two days. Expression cultures were harvested by centrifugation at 4500 g for 45 mins at 4°C, the supernatant decanted and centrifuged for a further 10 mins at 4500 g , 4°C. The protease enriched cell-free supernatants were adjusted to pH 6.0 with 20% (v/v) acetic acid and sodium azide added to 0.02% as a preservative. The culture supernatants were stored at 4°C until needed.

2.3.2 Colourimetric peptide assay for PragA9 activity

Crude observation of the presence of protease activity in the expression cultures was made by the monitoring of the protease activity present in a control culture of PragA9 α-Lytic protease using a synthetic suc-AAPM-pNA peptide substrate. Peptide substrate at 500 µM was

incubated with culture supernatant in 100 mM HEPES, pH 8. Liberation of pNA was monitored by A_{410nm} over 5mins, with the activity present in the culture expressed as the increase in A_{410} in that time interval.

2.3.3 Protease culture on skim milk agar.

A single colony of α-Lytic protease mutants was taken from the LA100 streak plates prepared in Section 2.3.1 and used to inoculate 5 ml of LB broth. Cultures were grown overnight at 25°C. The following day, the optical density at 600 nm of the culture was measured, and the equivalent of 10 µl of 1.0 OD_{600} unit applied to the centre of a skim milk agar plate. Plates were grown at 25°C for 7 days.

2.3.4 Fluorescence polarisation conjugate preparation

The substrate for the fluorescence polarisation reaction was prepared in a manner similar to that of Schade et al. (1996). Briefly, FITC was dissolved to 0.1% (w/v) in DMSO. β-casein was incubated with fluorescein label overnight in conjugation buffer, at a 7-fold molar excess of protein over label. This modest molar excess of FITC label is to minimise the chances of multiply labelled β-casein proteins. With overlapping excitation and emission spectra, fluorescein is able to self-quench between spatially proximal fluorophores (Voss *et al.,* 1996). Keeping the labeling ratio below 1 per protein minimises the likelihood of such FRET-like auto-quenching that could potentially interfere with the assay. Unbound label was removed by buffer exchange on a PD-10 column according to the manufacturer's instructions,

with pre-equilibration and elution in cleavage assay buffer. The resulting fluorescein thiocarbamoyl (FTC)-β-casein conjugate was stored at 4°C until needed.

2.3.5 *Fluorescence Polarisation assay*

The FTC-casein conjugate was diluted in cleavage buffer to obtain a total fluorescence yield equivalent to that of a 25 nM solution of free fluorescein. In quadruplicate, 10 µl of purified PragA9 standards or 1 µl of culture supernatants were added to the bottom of a black 96 well microtitre plate (BMG Labtechnologies GmbH, Offenburg, Germany). To this was added 100 µl of diluted FTC-casein conjugate. The plates were incubated at 37°C in a polarising flourimeter (BMG Labtechnologies Polarstar, Offenburg, Germany) with the fluorescence polarisation read every minute for 30 mins at $\lambda_{excitation}$= 485 nm and $\lambda_{emission}$= 520 nm.

Enzymes that returned protease levels outside the appropriate range for the assay were diluted 1:10 in cleavage buffer and retested. Conversely, mutants with undetectable protease activity were retested with 5 µl of culture supernatant.

2.3.6 *SDS-PAGE of FTC-casein conjugate.*

The PragA9 standard reactions were reproduced as detailed in Section 2.3.5. 20 µl of the reactions were run on 4-12% NuPAGE gel in MES buffer under reducing conditions according to the manufacturer's instructions. After electrophoresis protein bands were visualised by staining with Coomassie

Blue as follows. Gels were first fixed in 50% methanol (v/v), 10% acetic acid (v/v) for 20 minutes before staining with 0.5% (w/v) Coomassie Blue R-250 in 10% (v/v) acetic acid for 1 hour. The gels were destained by incubation in three changes of 10% (v/v) acetic acid for one hour each.

2.3.7 Enzyme Immunoassay

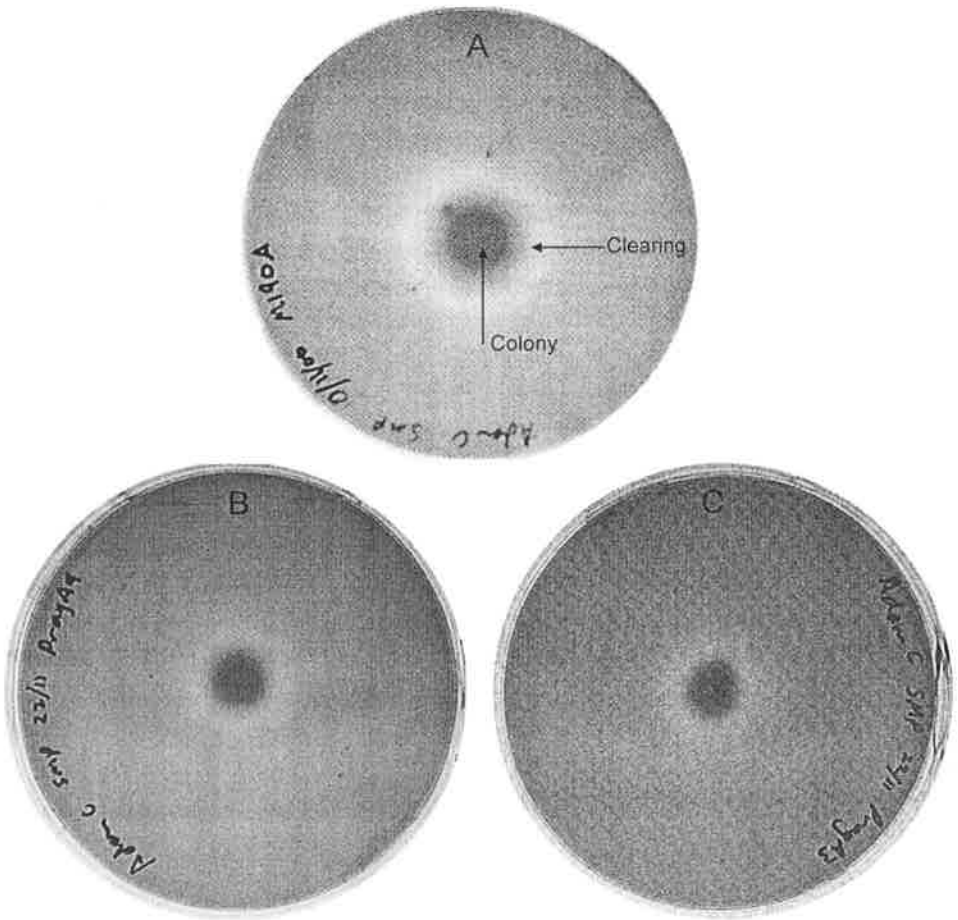
Rabbit polyclonal antiserum was raised against purified PragA9 α-Lytic protease by J. MacKrill, Dept. of Molecular Bioscience, University of Adelaide. Transparent Maxisorb 96-well microtitre plate (Nalge Nunc, Rochester, NY, USA) were coated by adding 10 µl of concentration standards or culture supernatant with 10 µl of 10x immunoassay coating buffer and 80 µl of water. Plates were incubated overnight at 4°C. Plates were washed with 3 changes of 300 µl PBS-T and blocked by addition of 200 µl 2% BSA in PBS-T with shaking for 1 hour at room temperature. Plates were washed as previously and 100 µl of antiserum was added to the blocked wells at a 1:4000 dilution in PBS-T containing 0.5% (w/v) BSA and shaken for 1 hour at room temperature. Plates were again washed thrice with fresh PBS-T and 100 µl added of the secondary antibody-horseradish peroxidase conjugate, diluted 1:4000 in PBS-T with 0.5% (w/v) BSA. Plates were shaken for 1 hour at room temperature, followed by a final wash. 100 µl of immunoassay substrate was added and plates incubated at 37°C for 5 to 10 mins to allow colour generation. Colour development was terminated by addition of 10 µl of 1 M H_2SO_4. Developed plates were read at 490 nm on a microtitre plate reader (Multiskan EX, MTX Lab Systems, Inc., Vienna, VA, USA).

2.4 RESULTS

2.4.1 Culture of mutants

A zone of clearing around the bacterial colony on skim milk agar plates demonstrated the secretion of active protease into the medium. Any active protease expressed by the clone degraded the casein that was clouding the agar, resulting in a zone of clearing. The cleared zones of various mutant protease expressing clones ranged from 0.5 to 1 cm farther than the radius of the colony. It was apparent that such qualitative data would lack the precision necessary for a protease activity assay. Figure 2.4.1.1 shows an example of some selected plates.

Figure 2.4.1.1 – Skim milk agar plate of selected mutants (A) M190A library parent, (B) PragA9, (C) PragA3. Protease secretion is observed by the zone of background clearing around the colony.



2.4.2 *Synthesis of conjugate*

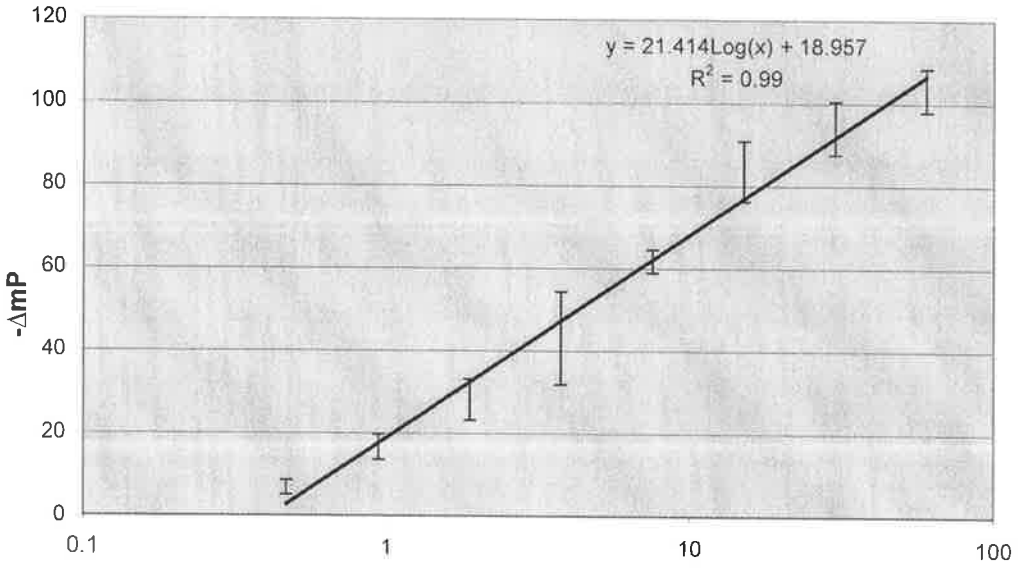
Preparation of conjugate was successful, with the degree of labelling found to be an average of 0.14 molecules of FTC per β-casein molecule, determined by the absorbance of the conjugate at OD_{490nm} using the extinction co-efficient of fluorescein (61,000 $mol^{-1}.cm^{-1}$) over the protein concentration of the conjugate preparation, calculated by BCA assay (data not shown). This confirms that the attempt at labelling the conjugate at less than one fluorescein label per protein molecule was successful, and at a little over 7 β-casein molecules per label (1/0.14 = 7.15) the seven-fold molar excess of protein used in the conjugate generation was labelled nearly quantitatively. This has minimised the likelihood of multiple, spatially proximal, fluorescein moieties on the surface of a significant population of β-casein molecules. The fluorescence polarisation (FP) of the FTC-casein conjugate was typically around 200 millipolarisation units (mP).

2.4.3 *FP standard curve*

The fluorescence polarisation of the conjugate was measured after proteolysis for 30 minutes with the standard amounts of PragA9. The fixed time was introduced as very little activity was observed in many examples. Physical limitations in the equipment prevented the measurement of instantaneous reaction rates, so a fixed time interval for the reaction was set. The 30 minute fixed time selected allowed more sensitive quantitation of low levels of protease activity, as little cleavage was observed in many of the unknown culture supernatants in short time intervals (see Section 2.4.5). The loss of polarisation that was observed upon proteolytic digestion of the

conjugate was proportional to the amount of protease added and reached a minimum polarisation of a little less than half of the initial level by the highest level of protease. A representative standard curve is displayed as Figure 2.4.3.1. Such curves were used for quantitation of the protease activity in the unknown culture supernatants of the other mutants (relative to a known amount of purified PragA9).

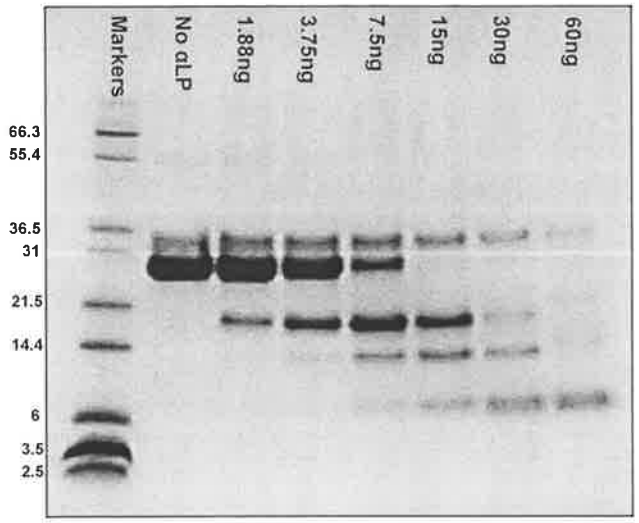
Figure 2.4.3.1 – Standard curve for calibration of FP assay with 95% confidence intervals.



2.4.4 *FP conjugate SDS-PAGE*

The proteolytic reactions of the conjugate that were used to generate the standard curve of Figure 2.4.3.1 were replicated for analysis by SDS-PAGE. Such a gel is shown as Figure 2.4.4.1, where it can be seen that the initial single dominant protein band is progressively degraded to smaller product species with the increasing protease concentrations. This confirms that the reduction in fluorescence polarisation of the conjugate is indeed due to the proteolytic fragmentation of the conjugate into smaller products.

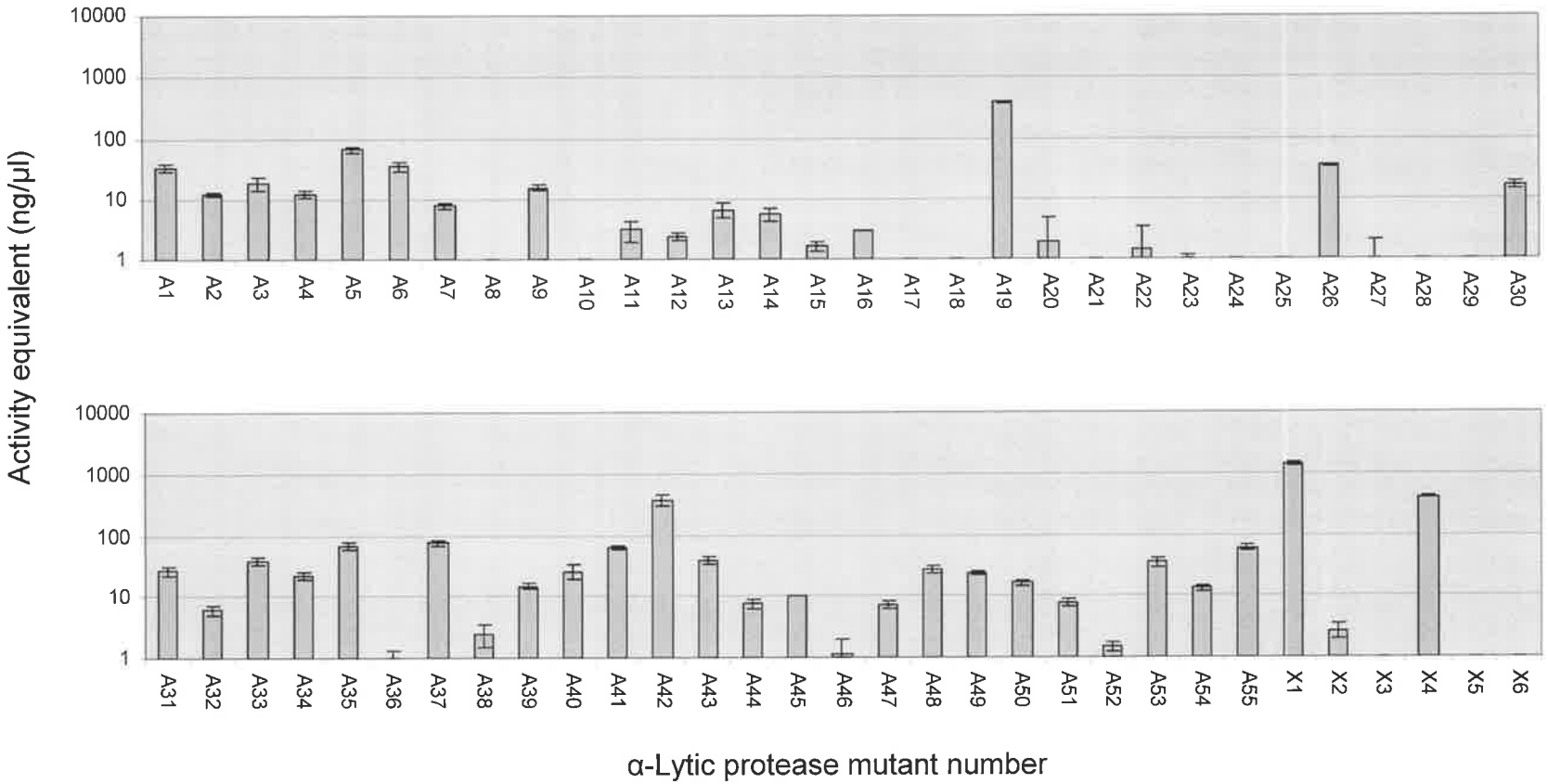
Figure 2.4.4.1 – SDS-PAGE of proteolysis products of the FTC-casein conjugate.

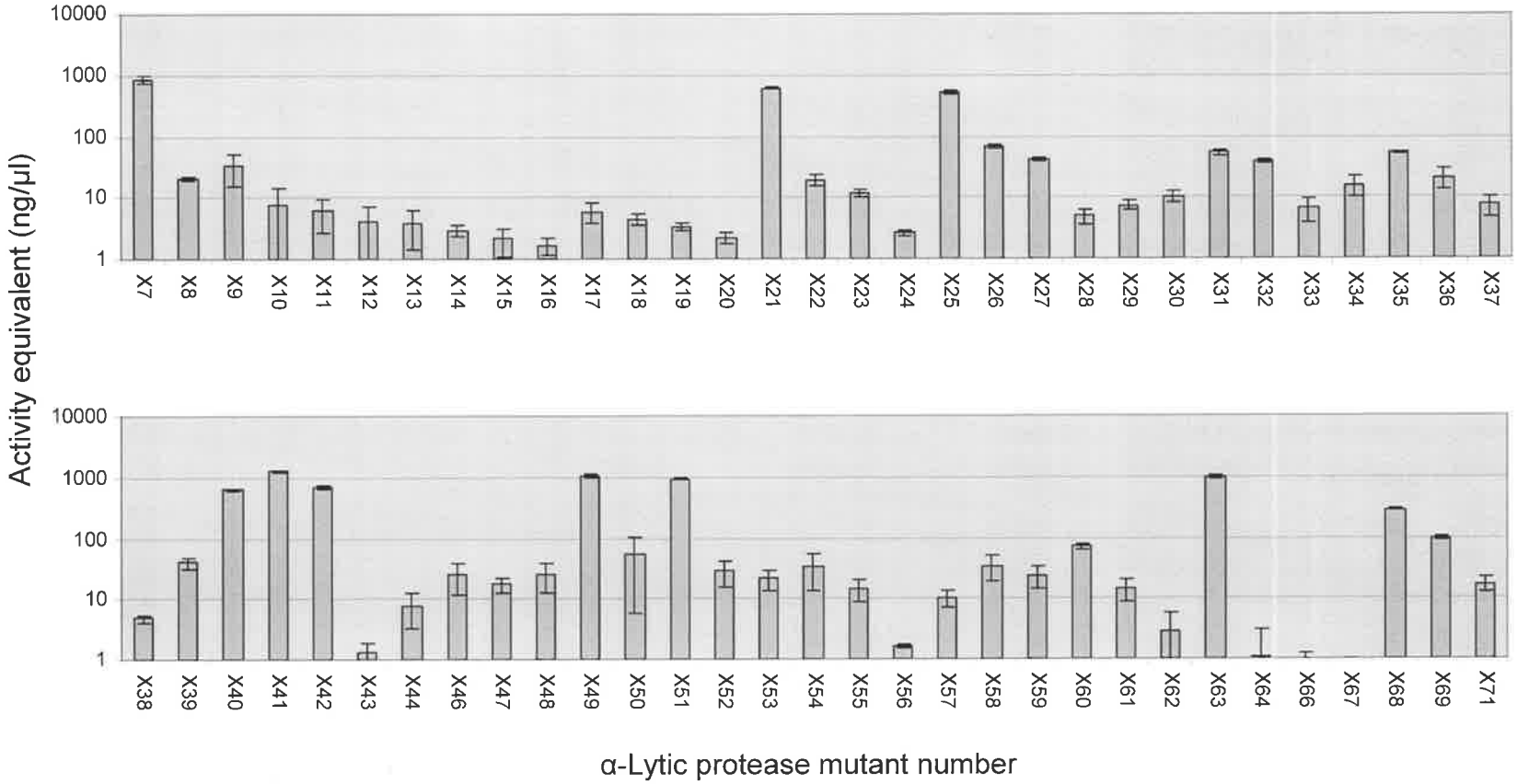


2.4.5 *Quantitation of protease activity in culture supernatants by FP.*

The two libraries of α-Lytic protease mutants were assessed for their total level of proteolytic activity using the FTC-β-casein fluorescence polarisation assay. As can be seen in Figure 2.4.5.1, the amount of protease activity was determined to be highly variable amongst the various mutants, ranging from undetectable to over 1000 ng/µl. This does not necessarily represent a 1000-fold difference in the level of protease present. Owing to the standardisation strategy employed (Section 2.4.3), the protease concentration present in the supernatants is expressed as the equivalent activity to PragA9, hence mutants with higher intrinsic activity or broader specificity than this standard mutant will report significantly higher "equivalent" activity at similar physical level of protease protein. This is, of course, the fundamental purpose of the generic protease activity assay, where the proteases can be compared on the basis of their ability to proteolyse a protein, rather than a physical amount of protease protein.

Figure 2.4.5.1 – Total protease activity measured in culture supernatants of the α-Lytic protease mutant libraries expressed in activity equivalent to purified PragA9 and ±1SEM.

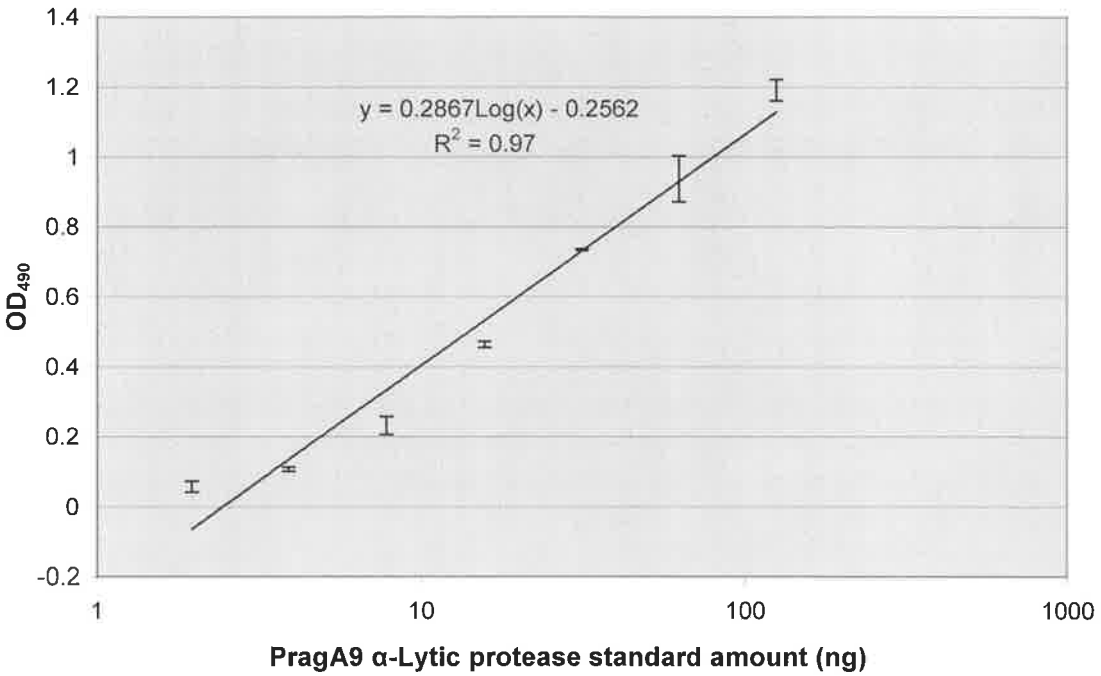



Activity equivalent (ng/µl)
10000
1000
100
10
1
X7 X8 X9 X10 X11 X12 X13 X14 X15 X16 X17 X18 X19 X20 X21 X22 X23 X24 X25 X26 X27 X28 X29 X30 X31 X32 X33 X34 X35 X36 X37
X38 X39 X40 X41 X42 X43 X44 X46 X47 X48 X49 X50 X51 X52 X53 X54 X55 X56 X57 X58 X59 X60 X61 X62 X63 X64 X66 X67 X68 X69 X71
α-Lytic protease mutant number

2.4.6 EIA standard curve

An enzyme immunoassay (EIA) was developed to measure the concentration of protease protein present in the libraries of the α-Lytic protease mutant's culture supernatants. As shown in Figure 2.4.6.1, the binding of the polyclonal antiserum was indeed proportional to the amount of α-Lytic protease immobilised on the plates.

Figure 2.4.6.1 – Standard curve for calibration of enzyme immunoassay with 95% confidence interval.

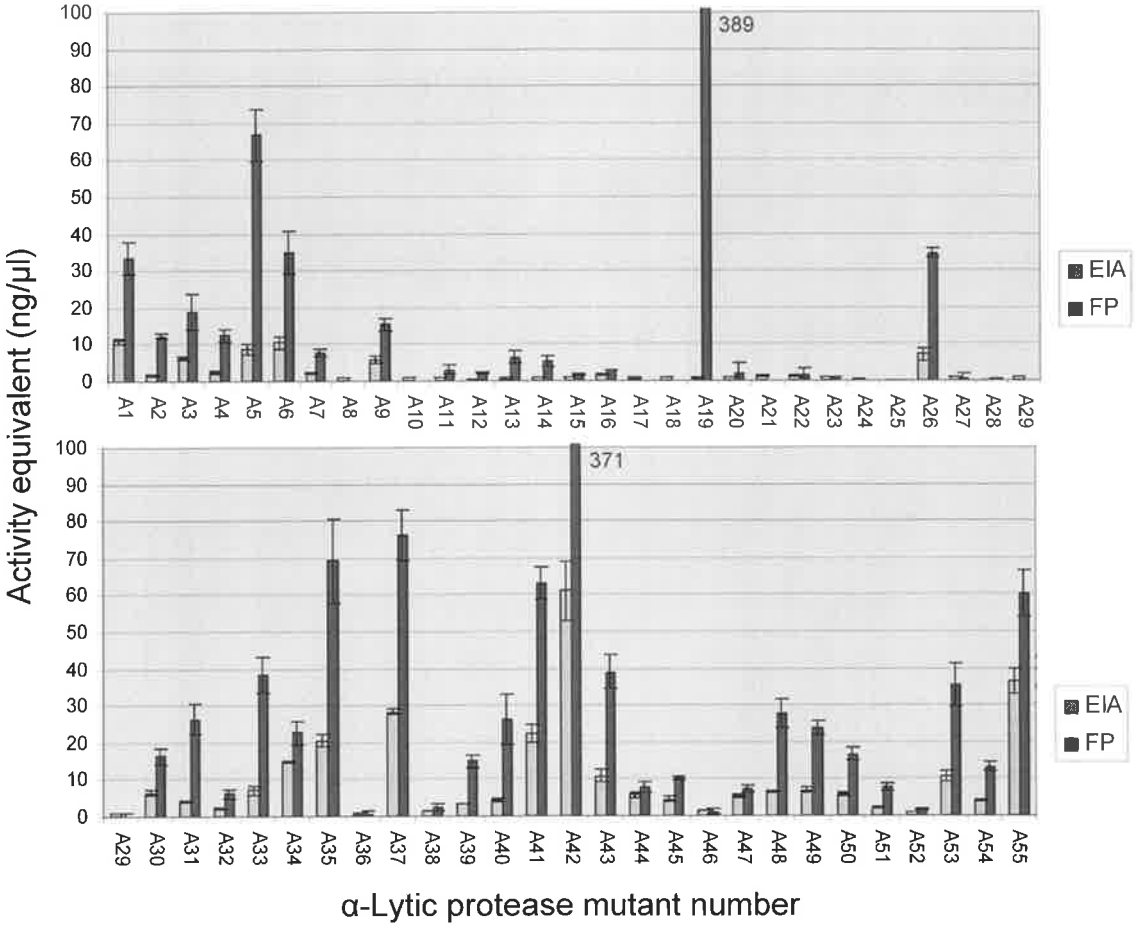


2.4.7 Quantitation of protease activity in culture supernatants by EIA.

The concentration of α-Lytic protease protein present in the culture supernatants was determined using an EIA and appears as Figure 2.4.7.1. Interestingly the protease protein concentration detected by the EIA was considerably more homogenous than that reported by the FP assay, where in

the case of the EIA protease concentrations for the most part ranged from low unit to low tens of nanograms per microlitre. As with the FP assay, due to the calibration against a single mutant, all concentrations are reported as equivalents to PragA9.

Figure 2.4.7.1 - α-Lytic protease protein concentrations determined by enzyme immunoassay ±1 SEM. For comparison, the concentrations determined by FP are also shown.



2.5 DISCUSSION

With the objective of examining all of the available α-Lytic protease mutants for a unique and desirable specificity, all members of the libraries have been initially screened for their general protease activity. To achieve

this result the mutants were cultured and their protease activity normalised to allow a direct comparison. Activity could not be observed in some isolates, allowing their exclusion from subsequent investigations.

Preliminary determination of a mutant's ability to secrete active protease was made by culturing on skim milk agar plates. This approach was in keeping with the assay originally used to evaluate the success of mutagenesis employed during the generation of the libraries (Graham *et al.,* 1993; Graham *et al.,* 1994). The clearing of the background cloudiness in the agar around the colony shows that the mutant is secreting active enzyme into the surrounding medium. This information suggests a candidate protein to use in the fluorescence polarisation assay, as all active mutants are known to be able to cleave casein. Representing approximately 33% of whole casein (Ekstrand and Larsson-Raznikiewicz, 1978), β-casein is an attractive candidate for a proteolytic assay for a two reasons. Firstly it contains no disulphide bonds (Ribadeau-Dumas *et al.*, 1972), and thus cannot offer a false negative result where a cleaved protein backbone could remain associated through an intact disulphide linkage. Secondly, as a subset of the whole casein mixture of isoforms, β-casein constrains the proteases to one single whole protein target, and thus whilst still presenting a large number of potential substrates, the number has been somewhat minimised, simplifying interpretation and reproducibility. At 209 amino acids in size (Kumosinski *et al.,* 1993), β-casein offers 206 potential four residue cleavage motifs, given that α-Lytic protease cleaves very poorly at sequences of less than four residues (Bauer *et al.,* 1981). So whilst far from presenting every possible

permutation of sequence, there remains significantly more randomisation than could be realistically achieved by discrete synthetic peptides.

The examination of the cleavage of the FTC-casein conjugate by SDS-PAGE reveals that the fluorescence polarisation data are indeed the result of a decrease in mass of the labelled protein. It is most interesting that the cleavage of the conjugate appears to be at discrete sites. This is in keeping with the existing knowledge of the PragA9 mutant, being known to cleave preferentially after a limited number of sequence motifs (Graham *et al.,* 1993; Lien *et al.*, 2003). Although not an actual timecourse, the protease concentration experiment does appear to show the substrate cleavage proceeded through a number of intermediates, where the initial intact 24 kDa protein is first cleaved to a ~19 kDa form, which is then itself cleaved to a ~13 kDa species, and finally to yield an ~8 kDa fragment. However, the 8 kDa form is itself not an absolute endpoint, as the total intensity of the bands in the 60ng reaction does not equal that of the no enzyme control, even when differential staining of smaller protein is considered. This suggests that the protease has continued to degrade the substrate down to small peptides not retained on the gel. The other possibility that explains the observed pattern is that the secondary cleavage sites that give rise to the smaller species are less favourable for the enzyme than that which is cut to produce the 19 kDa form, it therefore takes more enzyme to produce these products in the same time. In other words, cleavage at these secondary cleavage sites is rate-limiting. Examination of the amino acid sequence of β-casein reveals no clues to support either hypothesis, with amino acid sequences potentially labile to α-Lytic protease distributed relatively evenly throughout the protein. The

structure of the β-casein protein does potentially provide some insight, with the molecule displaying two "arm-like" open surface regions (Kumosinski *et al.,* 1993; Livney *et al.*, 2004) that could make attractive protease targets. The sequence of β-casein with these regions marked is shown as Figure 2.5.1. Cleavage at the Gln^{34}-Ser^{35} bond would give rise to a fragment of approximately 19.6 kDa. It is known that Gln can be an attractive P_1 residue for the PragA9 mutant of α-Lytic protease (Lien *et al.,* 2003) and that Ser is well tolerated at P_1' (Schellenberger *et al.*, 1994).

Figure 2.5.1 – Primary structure of bovine β-casein. The sequence elements that make up the open "arm-like" surface loop structures (Lys^{28} to Thr^{55} and Pro^{85} to Phe^{119}) are displayed in bold text (from Kumosinski *et al.,* 1993).

```
                    10                  20                  30
R E L E E L N V P G E I V E S L S S S E E S I T R I N K K I E K F Q S E E
    40                  50                  60                  70
Q Q Q T E D E L Q D K I H P F A Q T Q S L V Y P F P G P I P N S L P Q
                80                  90                  100
N I P P L T Q T P V V V P P F L Q P E V M G V S K V K E A M A P K H K
    110                 120                 130                 140
E M P F P K Y P V Q P F T E S Q S L T L T D V E N L H L P P L L Q S
                150                 160                 170
W M H Q P H Q P L P P T V M F P P Q S V L S L S Q S K V L P V P E K
    180                 190                 200
A V P Y P Q R D M P I Q A F L L Y Q Q P V L G P V R G P F P I I V
```

Throughout all reactions there also appears a highly protease resistant 35 kDa protein, the intensity of this band is only somewhat reduced by the highest concentrations of protease. It is unlikely that this band that actually represents the FTC-casein conjugate, as the most significant decreases in the intensity of this band only occur in association with very small changes in the fluorescence polarisation.

Enzyme immunoassay of culture supernatants of the first library of mutants shows average α-Lytic protease expression of 7 mg/L, with most falling in the range of 1-10 mg/L. Although these expression levels appear down on the published values using the same production regime (Haggett *et al.,* 1994a) of 14-77 mg/L, it should be noted that the values reported in those studies were determined by an activity based assay, calibrated against purified α-Lytic protease standard. This mode of α-Lytic protease quantification can report higher values than strict detection of protease protein, and indeed this trend was observed in the expression data of this chapter (see discussion below). If the activity data for the first library is considered, the agreement with the published results is much closer, with an average of 20 mg/L. Also, in the past only select mutants have been cultured in this manner and quantified, but it is known that α-Lytic protease mutants can display dramatic variability in expression. Mace *et al.* (1995) using an immunoassay for protease protein observed over a 1000-fold difference in the expression of Alanine scanning mutants of a surface loop of α-Lytic protease, despite all mutants being active.

One interesting trend in the data is that, where determined, the concentration of α-Lytic protease protein reported by enzyme immunoassay is invariably lower than the protease activity concentration demonstrated by the fluorescence polarisation assay. This is somewhat counter-intuitive, as the level of protease protein should represent the theoretical maximum attainable protease activity, and that could only be achieved in the unlikely scenario that the antiserum recognised only intact, mature, active protease. As both assays were calibrated against the same preparation of PragA9 α-Lytic protease

standard, the apparent discrepancy cannot be explained by incorrect quantitation of the standard. Any potential error introduced by this quantitation would be applied equally to both assays.

Background proteolysis of the FTC-casein substrate by contaminating, host-derived, proteases could be suspected of creating an artefactually high level of protease activity in the unknown culture supernatant samples. However, the large number of samples in which no protease activity was observed discredits this hypothesis, given that all mutants were cultured under the same conditions. Contaminating proteases, if present, do not appear to interact with the FTC-casein of the fluorescence polarisation assay.

One plausible explanation for the observed phenomenon comes from the idea that the mutants have variable total activities, independent of their specificity (Graham *et al.,* 1993; Graham *et al.,* 1994). Indeed, this was the rationale for conducting the activity normalisation assay from the outset. If a given mutant had higher activity, broader specificity, or both, than the PragA9 mutant used for calibration, it could present protease activity at a level higher than the protease protein concentration predicted by the reference mutant enzyme. Although it is not possible to determine if the higher activity is a function of broader specificity or a higher turn-over number, support for either possibility being involved in the phenomenon is provided by some of the data generated in the two screens. Indirect evidence for this is illustrated in the comparison of the mutants PragA19 and PragA42, which both showed similar levels of protease activity, yet varied in their protease protein concentration by over 50-fold (Figure 2.4.7.1). Whilst a higher level of inactive protease detected in the latter mutant could explain this variation in part, it is unlikely to

be solely responsible. Such a magnitude of the difference between activity and protease protein is not observed in any other group of mutants that display similar activity levels such as PragA1, PragA6, PragA26; PragA31 and PragA53; PragA5, PragA35 and PragA37; or the pair PragA41 and PragA55. Therefore at least some of the observed difference could be attributed to either: (1) A higher turn-over number, where the same unit concentration of protease is able to cleave more substrate in a given time; (2) Variation in the substrate specificity of the mutants, where FTC-casein better caters for the specificity requirements of some mutants than those of others, such that there are effectively more potential substrate sites per molecule.

However, variation in the specificity or turn-over number cannot be the only factor involved. When the reference mutant, PragA9, was cultured as part of the 1st library with the other mutants, it too showed the same pattern of reporting a higher concentration of protease by the FP activity assay than by the protein immunoassay. This mutant, by definition, has identical intrinsic activity and substrate specificity as was used for the calibration of both assays. Nevertheless, the variability in the enzyme turn-over number of the mutants does have a role to play in the diversity of the results observed, as shall be discussed later.

A plausible explanation for all the observed data is that the purified PragA9 α-Lytic protease standard itself may have quite low protease activity per mass unit of protein and that the antiserum does indeed recognise inactive protease protein. As stated above (Section 2.2.2), the quantification of the purified PragA9 α-Lytic protease standard was by the determination of protein concentration in the purified protease preparation. If a significant

proportion of this material was inactive, the protease activity assay standard curve would be generated from a lower actual level of active protease than thought, whilst the protease protein assay curve would be unaffected. Indeed, the purification protocol used for the generation of the standard is known to co-purify protease auto-digestion fragments (Haggett *et al.*, 1994b). Having undergone less potentially activity damaging purification steps and much shorter storage times, it could be argued that the culture supernatants would retain higher protease activity per unit protein. The protease protein levels reported in these samples by immunoassay would more likely be accurate relative to those of the standard. These protease levels, however, would appear more efficacious in the activity assay than was the standard on a protein mass basis.

2.6 Conclusion

The challenge in translating a concentration of protease protein into an ability to cleave a protein target has been highlighted by this work. It is for this reason that the level of protease is normalised with respect to its activity alone, irrespective of the protease protein concentration. Having cultured the α-Lytic protease mutants and normalised the level of generic activity present, they can now be applied to a target more relevant to the goal of this project. Although not directly pertinent to the main direction of the work, determination of the level of protease protein has provided some insight into the degree of variability in the intrinsic activity possible in the α-Lytic protease mutants. The lack of a correlation between protease concentration and activity should serve as a caution for future quantitation of purified α-Lytic protease.

3.1 INTRODUCTION

With the ultimate goal of releasing authentic human IGF-I from a Met-pGH(1-11)VN fusion partner using an α-Lytic protease mutant, the activity of the mutants has been quantified in Section 2.4.5. The next step toward the stated goal is to determine the ability of the mutants to act specifically at a cleavage motif that would enable such an outcome. The available literature on the S' substrate specificity of α-Lytic protease (Schellenberger *et al.*, 1994) suggests that wild-type α-Lytic protease should not be suitable for Met-pGH(1-11)VN-IGF-I cleavage. However, a systematic study of the suitability to this application of a library of α-Lytic protease mutants has not been performed, nor is the relative contribution of S' binding to overall α-Lytic protease catalysis known. Furthermore, there exists the possibility that modifications in the S' subsites have been created by the combinatorial mutagenesis to generate the library mutants (Graham *et al.*, 1993); thus presenting an opportunity to explore this aspect.

In the previous chapter, all members of the α-Lytic protease mutant libraries were cultured and the activities of the protease enriched culture supernatants normalised against a generic protein substrate. For the reasons stated in Section 2.1, the methodology employed in that work presented a protein substrate to the proteases, in an attempt to provide the largest range of random potential cleavage motifs as possible. The potential for modified S' specificity described above more than warrants an investigation into the ability of the mutants to cleave before the N-terminus of IGF-I, *i.e.*: to examine the mutant libraries for enzymes that can accept Gly-Pro-Glu as the cognate P_1' to P_3' residues.

In the peptide substrates that bear a P' motif based on the N-terminus of IGF-I, it is desirable to display an upstream cleavage motif (P_4 to P_1) that has the potential to be cleaved by the as many of the mutant enzymes as possible. This will permit the P' cleavage ability of a greater number of the mutants to be evaluated in one comparative assay system. Xaa-Ala-Pro-Xaa has been used as a P_4 to P_1 cleavage sequence framework on many occasions for a variety of α-Lytic protease investigations, on both wild-type enzymes and mutants (Bauer *et al.*, 1981; Bone *et al.*, 1989; Bone *et al.*, 1991; Epstein and Abeles, 1992; Mace and Agard, 1995). Significant among these was the generation of the libraries of mutant proteases that can cleave after such a sequence (Graham *et al.*, 1993; Graham *et al.*, 1994), even though it may not necessarily contain the optimal motif. Pro is often included at P_4 in such peptides as, although not the ideal residue for this subsite, it is well accommodated (Epstein and Abeles, 1992). However, Pro is not tolerated at P_3 or P_1 and thus a P-A-P-X motif should constrain the potential binding modes for the substrate to one, ensuring cleavage at the intended target site as shown by Lien *et al.* (2003).

As demonstrated by Graham *et al.* (1993) and Graham *et al.* (1994) and discussed in Section 1.4.2, although the members of the α-Lytic protease mutant libraries display quite variable P_1 specificity, they can be grouped according to commonalities in the types of P_1 amino acids after which they are able to cleave. Thus, if the P_1 position is randomised sufficiently to include at least one amino acid from each of these recognition groupings, the likelihood that the mutants will be presented with a favourable P_1 will be maximised, including the enzymes of largely uncharacterised specificity.

As in Chapter 2, given the large number of enzymes to be evaluated, the methodology must remain sufficiently high-throughput for the initial screening, with positive hits then examined more stringently in subsequent investigations. With the detection of internal cleavage of a peptide substrate required, the options available for screening of the enzymes are limited to the fluorescent techniques for practical reasons as discussed in Section 2.1. Other approaches require either release of a reporter molecule (such as a chromophore), which precludes the presence of any P' residues, or are too insensitive, time-consuming, costly or prone to interference to be applied to crude samples (*e.g.*: HPLC, Mass Spectrometry). Both FRET and FP have been used extensively to monitor internal proteolysis of peptides, with over 200 Medline citations for development of such assays in the last decade alone. Although the number is three-fold in favour of FRET assays, this is probably due to FRET not requiring the specialised equipment that FP does. FRET-type protease assays are versatile, and have been used in more diverse applications than just solution-phase cleavage monitoring alone. These include incorporation in microarrays (Gosalia *et al.,* 2005; Han *et al.,* 2006; Thomas *et al.,* 2006), flow cytometry sorting of peptide-coated microspheres (Saunders *et al.,* 2006) and FRET peptides expressed on cell surfaces (Olsen *et al.,* 2000; Boulware & Daugherty, 2006). However, having demonstrated the utility of FP for the study of these mutants in this particular crude sample matrix, it is logical to maintain the use of FP for further characterisation of the action of the mutants. Moreover, FP makes an attractive candidate for an assay of this type since binding of an N-terminally biotinylated peptide to the 68 kDa avidin molecule could be expected to

effectively prevent all rotation within the fluorescence lifetime of many fluorophores commonly used for FP. Furthermore, the binding of a comparatively small peptide to the large avidin protein may well abolish any residual protease acting on the peptide by blocking access to the cleavage motif; especially where that cleavage motif is relatively short. The end-point nature of such an assay should simplify the post-assay data analysis.

In this chapter the libraries of α-Lytic protease mutants will be screened against model peptide substrates for the potential to liberate authentic IGF-I. The ability of the enzymes to cleave the required sequence will be evaluated in a model system by a high-throughput fluorescence polarisation peptide assay. Positive hits obtained in this screening assay, or a subset thereof, will then be studied in more detail using a more discriminating peptide model of the final cleavage motif.

3.2 MATERIALS

3.2.1 General reagents & materials

The following reagents were obtained from Sigma-Aldrich, Castle Hill, NSW, Australia: egg white avidin, dimethyl sulphoxide (DMSO) and N-[tris(hydroxymethyl)methyl]glycine (tricine). BDH chemicals, Kilsyth, Victoria, Australia, supplied: N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES), NaCl, NaOH, trifluoroacetic acid (TFA) and acetonitrile. All reagents were of Analytical reagent grade or higher. SP-sepharose Fast Flow resin was supplied by Amersham Bioscience

(Uppsala, Sweden). Benzonase Nuclease (purity II) was obtained from Merck Biosciences KGaA. Darmstadt, Germany

3.2.2 FP peptide sets

The peptides for the FP assay were produced as two sets of five peptides each, as described in Section 3.4.1 and Figure 3.4.1.1. The peptide sets were synthesised by Mimotopes Pty. Ltd., Clayton, Victoria, Australia and were of "as synthesised" or ≥70% purity. Peptide sets were made up to 0.5 mg/ml (2.5 mM) in 20% DMSO, 100 mM HEPES, pH 8.5 and stored at 4°C until needed.

3.2.3 Discrete peptides

Long discrete peptides were synthesised by Auspep Pty. Ltd., Parkville, Victoria, Australia and were of ≥90% purity. The sequences are shown in Table 3.2.3.1. Discrete peptides were made up to 1 mg/ml (0.64 to 0.83 mM) in 10 mM HCl and stored at 4°C until needed.

Table 3.2.3.1 – Sequences of long, discrete peptides for characterisation of shortlisted α-Lytic protease mutants.

TLG-motif	GPE-motif
H-WSGSPAPMTLGW-OH	H-WSGSPAPMGPETLGW-OH
H-WSGSPAPATLGW-OH	H-WSGSPAPAGPETLGW-OH
H-WSGSPAPFTLGW-OH	H-WSGSPAPFGPETLGW-OH
H-WSGSPAPLTLGW-OH	H-WSGSPAPLGPETLGW-OH
H-WSGSPAPHTLGW-OH	H-WSGSPAPHGPETLGW-OH

3.2.4 General solutions.

Cleavage buffer

100 mM HEPES

pH 8.5

3.2.5 α-Lytic protease purification buffers

Sodium Tricine buffer (Na Tricine) was prepared as a 10x stock, at a sodium ion concentration of 250 mM. NaOH (0.5 mol) was added to 1.6 L Milli-Q water and solid tricine added until the solution reached pH 8.7 at 20°C. The buffer was made to 2 L and tricine added to give pH 8.6. Binding buffer for α-Lytic protease purification was prepared by diluting the above Na Tricine stock to 25 mM with Milli-Q water and the pH confirmed at 8.6. The elution buffer was identical to the binding buffer but for the inclusion of 1 M NaCl.

3.3 METHODS

3.3.1 FP peptide screening assay

Using the α-Lytic protease mutant concentrations determined in Chapter 2, the GPE and TLG motif bearing FP peptide sets were reacted with the protease at an enzyme to substrate molar ratio of 1:500. The peptide sets were diluted from the stock concentration to 50 μM in cleavage buffer and each peptide set was incubated with each protease mutant in a 10 μl reaction volume for 16 hours at 37°C in a microfuge tube. Peptide sets without included protease were incubated as the negative control. To the reacted

mixtures, 90 µl of cleavage buffer was added as a diluent and the sample applied to a well of a black polystyrene 96 well microtitre plate (BMG Labtech GmbH, Offenburg, Germany). To each well was added 10 µl of 50 pg/µl avidin in cleavage buffer and polarisation readings averaged over 10 minutes in a Polarstar fluorimeter (BMG Labtech GmbH, Offenburg, Germany) with $\lambda_{excitation}$= 485 nm and $\lambda_{emission}$= 520 nm.

3.3.2 *Purification of shortlisted mutants*

The purification of the shortlisted mutants (see Section 3.4.3 and Discussion) was performed largely as described in Haggett *et al.* (1994b), except in this case SP-sepharose Fast Flow resin was used as the stationary support in place of the MacroPrep High-S used in that study. Ten ml of the protease-enriched culture supernatants were adjusted to pH 8.0 and $MgCl_2$ added to 1 mM. The supernatants were then treated with benzonase nuclease at a rate of 5 µl nuclease per 100 ml and incubated at room temperature for 2 hours. Benzonase treated supernatants were diluted 1:10 with cold Milli-Q H_2O and adjusted to pH 8.6 with 2 M NaOH.

The resin for each mutant was washed in 3 changes of Milli-Q H_2O to remove preservative. New resin for each mutant was used to pack 1 ml of settled bed in an Econo-Pac column housing (BioRad Laboratories, California, USA) and the column operated under gravity flow. All steps in the chromatography were conducted at 4°C and all buffers chilled to 4°C prior to use. The column was pre-equilibrated with 5 column volumes (CV) of 10x Na Tricine buffer. The column was then equilibrated with 5 CV of binding buffer and the diluted, benzonase treated sample loaded. Unbound material was

washed from the column by addition of a further 5 CV of binding buffer. Elution was effected by a step-gradient, with incremental steps of 5 CV of elution buffer at 10%, 20%, 50% and 100% (v/v) in binding buffer. One CV fractions of all steps were captured and assessed for protease activity using the casein-FP assay of Chapter 2. Protease containing fractions were stored at 4°C until required.

3.3.3 Cleavage of discrete peptides.

The longer, discrete peptides (as opposed to the 5-member mixture of the FP peptide sets) were reacted with the purified shortlisted α-Lytic protease mutants. Peptides were diluted from the stock in cleavage buffer to 50 μM and reacted with the mutants at a molar enzyme to substrate ratio of 1:500 for 16 hours at 37°C. Reactions were terminated by acidification with 1.5 reaction volumes of 1.5% TFA in Milli-Q H_2O.

The cleavage reactions were analysed by reversed phase HPLC (rp-HPLC) on an Agilent 1100 HPLC system (Agilent technologies, California, USA) with using a C4 column (Brownlee Reversed Phase, BU-300, Aquapore Butyl C-4, 2.1 mm I.D. x 100mm length, Perkin Elmer, MA, USA.). Separation was by linear gradient of 0 – 50% acetonitrile over 30 minutes at 0.5 ml/min, with absorbance at 215 nm. All mobile phase solvents contained 0.1% TFA.

3.3.4 Mass Spectrometry of peptide cleavages.

Cleavage reactions were desalted by rp-HPLC and dried by vacuum centrifugation. Mass spectrometry was performed by L. Hick, University of Wollongong, NSW, Australia.

3.4 RESULTS

3.4.1 FP screening peptide design.

The peptides used for screening the α-Lytic protease mutant libraries for the required specificity have an architecture similar to those of Levine *et al.* (1997). A schematic representation of the peptides appears as Figure 3.4.1.1. Recognition of the biotinylated N-terminus by avidin is anticipated to provide a large increase in fluorescence polarisation in the intact peptide. Where a protease has cleaved a peptide, the fluorophore will no longer be physically associated with the biotin moiety and this increase in polarisation will thus not be observed.

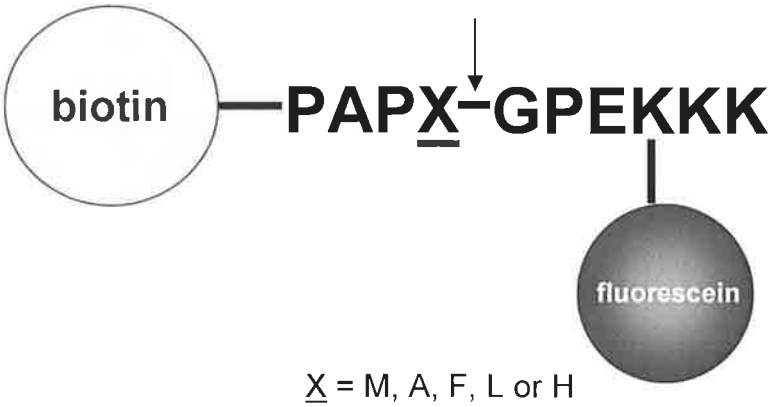
The FP screening peptides were designed to display the shortest possible cleavage motif to the protease mutants, minimising the opportunity for false positives from cleavage at sites other than the required one. As this assay is designed to evaluate the ability of the mutants of α-Lytic protease to release the N-terminus of IGF-I, one type of peptide bore a Gly-Pro-Glu sequence to mimic this site. Another, containing a Thr-Leu-Gly motif serves as a positive control. Thr-Leu-Cys are the next three residues of IGF-I, and as such are the first three residues of des(1-3)IGF-I, a protein that is able to be released quantitatively from its Met-pGH(1-11)VN fusion partner using

α-Lytic protease mutants (G. Francis, personal communication). Thr-Leu-Gly has been used as a model for this motif in the past (Lien *et al.*, 2001). Another important reason for the selection of this control is that it bears a Leu in the apparently highly selective P_2' position of the substrate. Leu at P_2' has been shown to have a nearly 3-fold higher second order rate constant than the next preferred residue (Met), and provides an overall discrimination factor of 173-fold between the most and least favourable residues at this site (Schellenberger *et al.*, 1994).

As discussed in Section 3.1, the upstream P residue motif for the peptides was the canonical Pro-Ala-Pro-Xaa, with Xaa representing one of five amino acids. These five amino acids correspond to the five most frequently recognised P_1 residues of the α-Lytic protease mutants (see Table 3.4.2.1), ensuring that each mutant would see at least one P_1 residue that it was able to act upon. The peptides of each P' sequence type are grouped into two mini-libraries or "sets", consisting of an equimolar mixture of five peptides, varying at the P_1 residue. This gives an assay system of two FP peptide sets, one bearing a GPE sequence and one with a TLG motif.

It was anticipated that solubility may be an issue for these small, largely non-polar peptides. To attempt to address this issue, a tri-lysyl sequence was included C-terminal to the P_1' to P_3' motif. The last of these was added during synthesis as a fluoresceinyl-lysyl to incorporate the required fluorophore.

Figure 3.4.1.1 – Schematic representation of the FP peptide screening sets. The intended scissile bond is marked with a downward arrow. Displayed is the GPE motif bearing peptide.



3.4.2 Peptide screening assay

Pilot cleavages with purified representative α-Lytic protease mutants (PragA9 and PragA55) were performed and the data shown in Figure 3.4.2.1. The polarisation observed in the full-length peptides is shown to increase by approximately 20-25% when avidin is added. Correspondingly, the polarisation of the peptides decreases by a similar margin after exposure to protease, and is largely unresponsive to the presence of avidin. In these early experiments there are apparent differences in the ability of mutants to act upon the two peptide sets, with both mutants reducing the FP of the TLG peptide set to the same level, but one enzyme (PragA55) able to reduce the FP of the GPE set nearly two-fold more that the other (PragA9).

The FP peptide screening assay was carried out on the culture supernatants of the α-Lytic protease mutant libraries, the outcome of which is displayed in Table 3.4.2.1. Cleavage of any of the peptides is indicated by the decrease in the polarisation of that peptide set relative to its corresponding

full-length peptide set. This investigation revealed that almost all the mutants were able to cleave the GPE peptide set to some extent. Interestingly, approximately one quarter of the enzymes registered a greater decrease in FP for the GPE-motif peptide set, suggesting that these mutants may actually favour cleavage before this motif. From this data a shortlist of mutants were selected for more detailed study (see Discussion). The enzymes can thus be grouped as either "GPE-favouring" or "TLG-favouring" according to the FP peptide set for which they showed the greatest decrease in polarisation. One trend seems to be that the GPE-favouring enzymes have a broader P_1 substrate specificity. Where nearly half of the TLG-favouring mutants have a single P_1 preference, a similar proportion of the GPE-favouring enzymes have 3 to 4 preferred P_1 residues (see Graham *et al.* (1993), Graham *et al.* (1994) and Table 3.4.2.2).

Figure 3.4.2.1 - Pilot raw FP of no-protease control and protease-treated peptide sets with and without added avidin

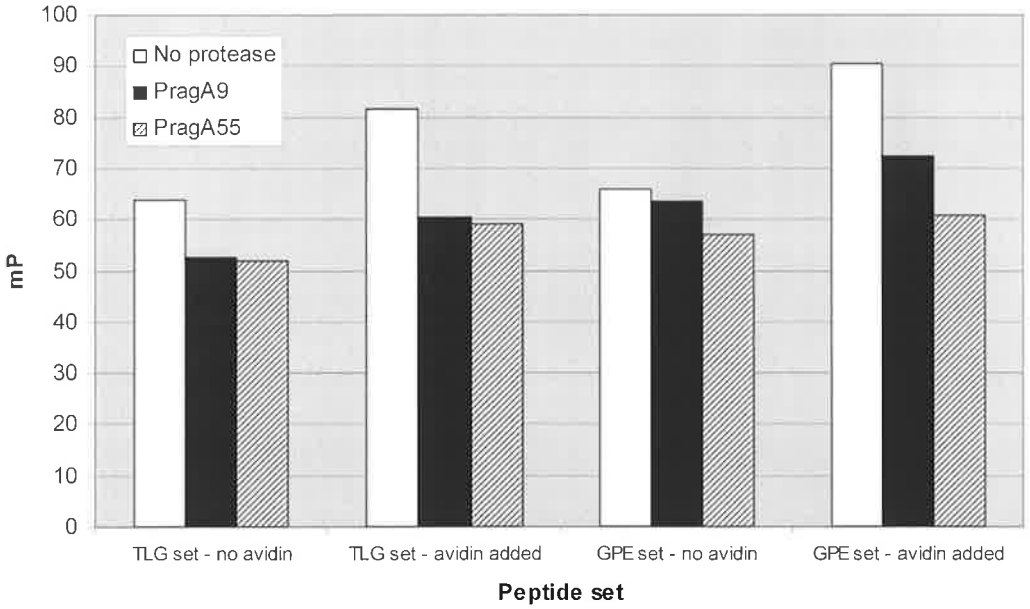


Table 3.4.2.1 – Results of the FP screening assay for the α-Lytic protease mutant libraries. Values are the mean of 2 replicates.

†Mutations and Favoured P_1, where known, are from Graham *et al.* (1993) and Graham *et al.* (1994) with blank cells signifying unknown values. Wild-type S_1 residues are shown in italics.

‡The index value is the GPE set score minus the TLG set score; hence a positive index value suggests a mutant that prefers to cleave the GPE set over the TLG set.

*Shortlisted mutants.

Enzyme	Mutation†					Favoured P_1†	TLG set ΔmP	GPE set ΔmP	Index value‡
	190	191	192	213	218				
PragA1	A	*G*	V	H	M	M	12.2	9.2	-3
PragA2	A	*G*	*R*	H	M	M/H	6.8	4.8	-2
PragA3	A	*G*				M/H	13.8	9.2	-4.6
PragA4	A	*G*				M/H	14.6	5.5	-9.1
PragA5	A	*G*					12.8	0	-12.8
PragA6	A	*G*				M/H	15.2	5.1	-10.1
PragA7	A	*G*	*R*	H	I	M/H	20.9	12.9	-8
PragA9*	A	*G*	*R*	H	L	M/H	18	8.2	-9.8
PragA13	A	*G*	*R*	H	S	H	16.2	6.9	-9.3
PragA19	A	*G*	H	H	T	M/H	19.2	13.6	-5.6
PragA26	A	*G*				M/A/F	16.8	19.7	2.9
PragA30	A	*G*				M/F	17.4	16.8	-0.6
PragA31	A	*G*	*R*	*M*	L	M/A/F	19.5	18.7	-0.8
PragA32	A	*G*				M	20.9	10.6	-10.3
PragA33*	A	*G*				M/F	16.4	22.2	5.8
PragA34	A	*G*				M/F	18	15.1	-2.9
PragA35	A	*G*	L	L	M	M/F	17.4	14.8	-2.6
PragA37	A	A	Q	T	M	M/F	20	14.5	-5.5
PragA39	A	*G*	*R*	F	M	F/A	20.7	21.4	0.7
PragA40*	A	*G*				M/F/A	14.9	15	0.1
PragA41	A	*G*	*R*	T	M	M/F/A/L	12.5	21.4	8.9
PragA42	A	*G*				M/F/A	13.5	15	1.5
PragA43*	A	*G*	*R*	*M*	I	M/F	11	19	8
PragA44	A	A	T	T	I	M/F/L	18.2	25	6.8
PragA45	A	A	W	T	L	M/F/A	18.6	15.1	-3.5
PragA47	A	*G*	*R*	S	L	M/F/A	10.6	15.8	5.2
PragA48	A	*G*				M/F/A	9.4	14.3	4.9
PragA49	A	*G*				M/F/A	11.9	15.5	3.6

Enzyme	Mutation†					Favoured P_1†	TLG set ΔmP	GPE set ΔmP	Index value‡
	190	191	192	213	218				
PragA50	A	*G*	*R*	*M*	M	M/F	10.5	9.5	-1
PragA51	A	*G*	*R*	L	L	M/F/A	12.3	17.4	5.1
PragA53	A	*G*				M/F/A	11.5	16.4	4.9
PragA54	A	*G*				M/F/L	12.6	11.5	-1.1
PragA55	A	*G*	*R*	T	L	M/F/A/L	10.7	16	5.3
RadX1	L	*G*	S	S	M	A	0.4	1.1	0.7
RadX2		*G*					0.7	0.6	-0.1
RadX4	V	*G*	K	S	M	A	12.4	4.1	-8.3
RadX7	V	*G*	A	T	L	A	15.1	3.7	-11.4
RadX8	V	*G*	L	H	*V*	N	18.3	12.2	-6.1
RadX9	S	*G*	*R*	H	L	A/M/H	19.1	12.9	-6.2
RadX10		*G*					0	1.8	1.8
RadX11	T	*G*	H	T	S	M/F	0	0.6	0.6
RadX12	S	*G*	T	S	Q	F	3.9	2.4	-1.5
RadX13	I	*G*	Y	A	T	M/V/A	3.6	0.7	-2.9
RadX14	L	*G*	G	S	T	A	6.1	0	-6.1
RadX17		*G*				F	21	2.4	-18.6
RadX18	I	*G*	Y	T	S	A/M	17.4	4.5	-12.9
RadX19	Q	*G*	L	S	*V*	M	24.6	11.4	-13.2
RadX21	V	*G*	*R*	T	L	A	27	10.8	-16.2
RadX22	A	*G*	*R*	T	M	M/F/A/L	16.2	18.9	2.7
RadX23	S	*G*	E	S	L	A	13	5.8	-7.2
RadX25	V	*G*	K	S	M	A/M	11	3	-8
RadX26	L	*G*	S	T	I	A	16	8.4	-7.6
RadX27	I	*G*	A	L	*V*	A	16.2	7.7	-8.5
RadX28	G	*G*	Q	S	*V*	M/F	12.9	9.3	-3.6
RadX29	V	*G*	N	T	G	M	15.4	12.1	-3.3
RadX30*	T	*G*	*R*	N	M	A	16.8	19.9	3.1
RadX31	T	*G*	K	A	L	A	18.2	15.2	-3
RadX32	A	*G*	V	S	M	M/F	16.5	12.7	-3.8
RadX33		*G*					0	0.7	0.7
RadX34	V	*G*	A	T	E	A	14.6	9.7	-4.9
RadX35	V	*G*	L	T	S	M/A	16	13.7	-2.3
RadX36	*M*	*G*	*R*	T	T	A	13.3	13.3	0
RadX37		*G*				A	2.8	8.8	6
RadX38		*G*					4.6	21.2	16.6
RadX39	T	*G*	W	S	L	A	13.1	6.6	-6.5
RadX40	V	*G*	N	S	I	A	21.8	4.8	-17
RadX41	L	*G*	G	S	L	A	5.2	0	-5.2
RadX42	L	*G*	Q	T	L	A	13.7	2.9	-10.8
RadX44		*G*					6.1	6.8	0.7
RadX46	L	*G*	E	S	I	A	17.3	12.8	-4.5
RadX47	V	*G*	H	T	S	M	19.6	7.3	-12.3
RadX48	I	*G*	M	T	S	M/A	12.8	6.2	-6.6
RadX49	I	*G*	E	T	I	A	14.7	9.6	-5.1

Enzyme	Mutation†					Favoured P_1†	TLG set ΔmP	GPE set ΔmP	Index value‡
	190	191	192	213	218				
RadX50		*G*				M	11.2	10.1	-1.1
RadX51	L	*G*	V	S	M	M/A/V	10.6	3	-7.6
RadX52	*M*	*G*	*R*	S	T	M/A	16.3	12.2	-4.1
RadX53	I	*G*	G	A	M	A	16.6	15.7	-0.9
RadX54		*G*					10.9	13.5	2.6
RadX55*	A	*G*	E	H	S	M/H	10.4	11.6	1.2
RadX57		*G*				A	12.7	12.9	0.2
RadX58	V	*G*	Q	T	T	M	8.2	6.6	-1.6
RadX59	A	*G*	A	*M*	M	M/F/A	7.9	3.6	-4.3
RadX60	V	*G*	V	T	S	M/A	21.6	16.2	-5.4
RadX61	L	*G*	*R*	S	T	A	22.5	14.9	-7.6
RadX62		*G*				M	14.8	11.2	-3.6
RadX63*	V	*G*	E	T	L	M/A	23	10.3	-12.7
RadX69		*G*				M/F/A	16.5	9	-7.5
RadX71		*G*				M/F	11.9	13.5	1.6
M190A	A	*G*	*R*	*M*	*V*	M/F	20.7	15.5	-5.2

Table 3.4.2.2 – P_1 residues recognised by the α-Lytic protease mutant libraries, sorted by their apparent P' preference from the screening data.

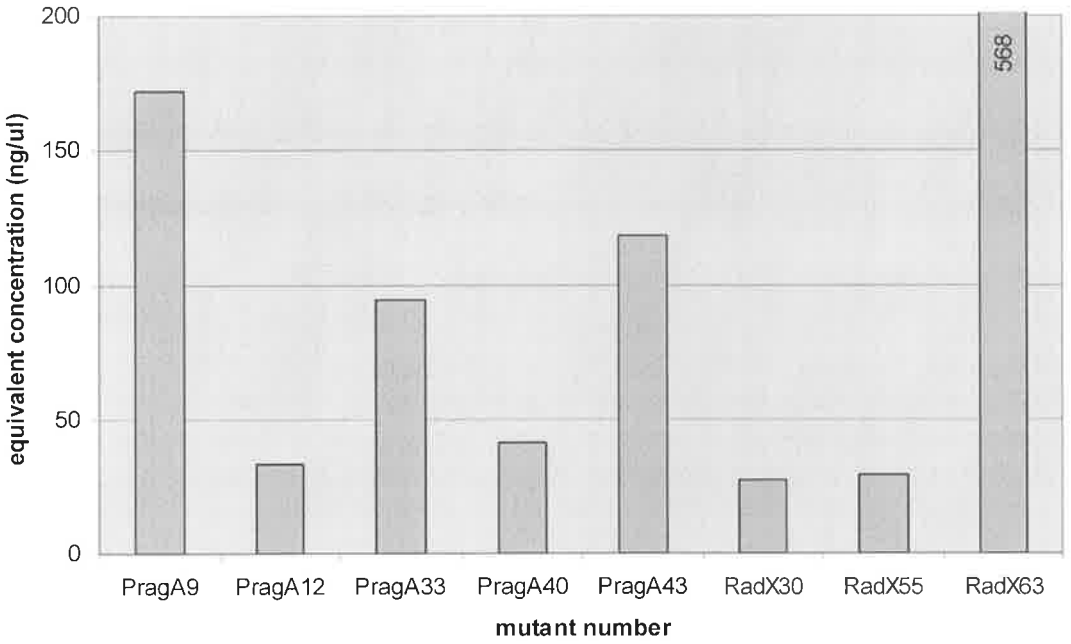
Number of preferred P_1 residues	GPE favouring	TLG favouring
1	15%	48%
2	22%	35%
3	33%	13%
4	11%	0
Unknown	19%	3%

3.4.3 α-Lytic protease shortlist selection and purification

A shortlist of α-Lytic protease mutants was selected based on the results of the fluorescence polarisation screening assay. The mutants were selected for either their P' substrate preference or their S_1 mutations (see Discussion). The shortlisted mutants were PragA9, PragA12 (see Discussion), PragA33, PragA40, PragA43, RadX30, RadX55 and RadX63.

During their purification the α-Lytic protease mutants were observed to elute in the first 2 fractions of 20% B (=200 mM NaCl). This is in good agreement with the findings of Haggett *et al.* (1994b) of elution at 250 mM NaCl. The α-Lytic protease activity in the peak fraction pools was determined using the β-casein FP assay of Chapter 2, with the activity thus reported in PragA9 equivalents. The protease activity in the purified samples is shown as Figure 3.4.3.1.

Figure 3.4.3.1 – Protease concentration detected in elution faction pool



3.4.4 Cleavage of long discrete peptides.

Cleavage of the peptides was monitored by HPLC. Some representative traces are shown as Figures 3.4.4.1 to 3.4.4.4, with the retention time of the dominant cleavage product peaks labelled. The retention times of these peaks in the reactions of each peptide with each shortlist

mutant protease are summarised in Table 3.4.4.1. The reoccurrence of certain peaks in reactions of peptides sharing particular features, such as the 18.16 (±0.15) minute peak in many TLG-motif peptide cleavages allows cautious assignment of identities to many peaks (see Discussion). These abstracted identities are indicated by the bracketed one letter code in Figures 3.4.4.1 to 3.4.4.4 and Table 3.4.4.1.

Figure 3.4.4.1 – HPLC trace of RadX30 protease cleavage of TLG-motif discrete peptide with P_1 Ala. The reaction trace is overlaid with full-length substrate incubated without added protease. A = N-terminal fragment of the cleaved peptide, T = C-terminal fragment of the cleaved peptide

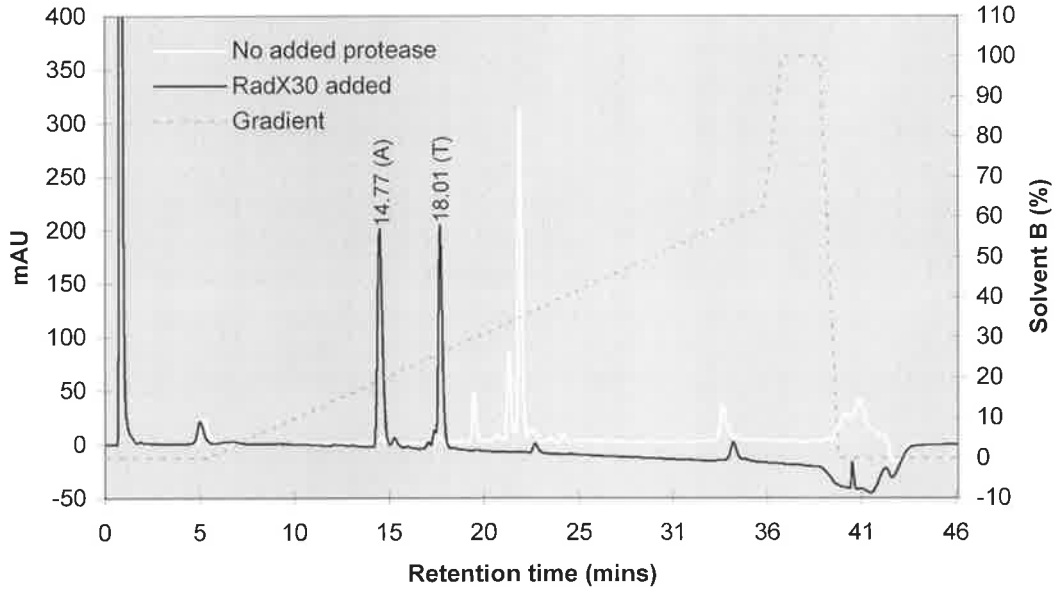


Figure 3.4.4.2 – HPLC trace of RadX55 protease cleavage of TLG-motif discrete peptide with P_1 Phe. The reaction trace is overlaid with full-length substrate incubated without added protease. F = N-terminal fragment of the cleaved peptide, T = C-terminal fragment of the cleaved peptide, D = Degradation of the peptide, U = Unidentified peak, I = Incomplete cleavage

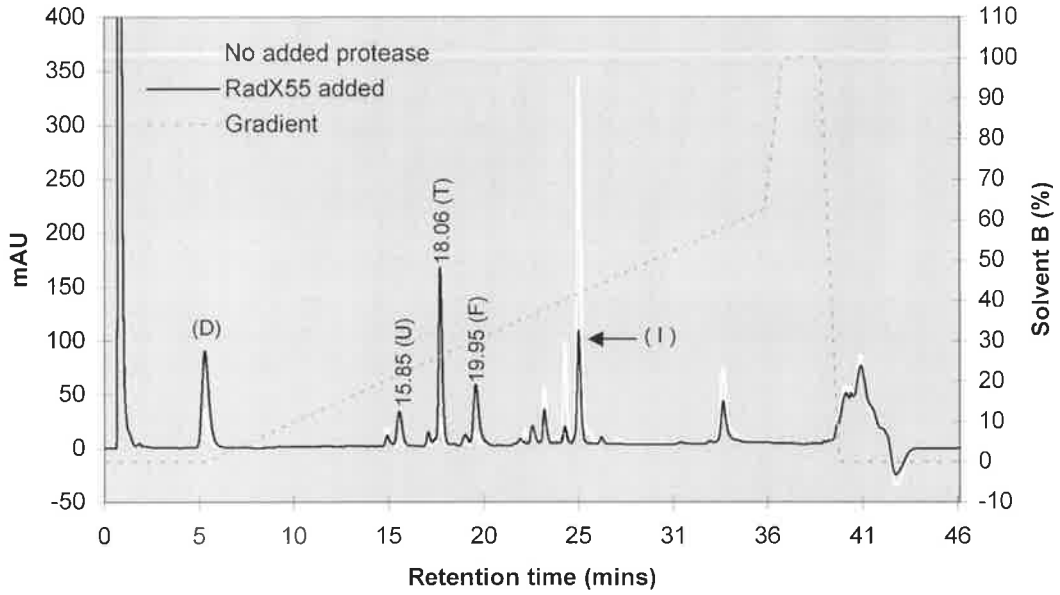


Figure 3.4.4.3 – HPLC trace of PragA9 protease cleavage of GPE-motif discrete peptide with P_1 Met. The reaction trace is overlaid with full-length substrate incubated without added protease. M = N-terminal fragment of the cleaved peptide, G = C-terminal fragment of the cleaved peptide.

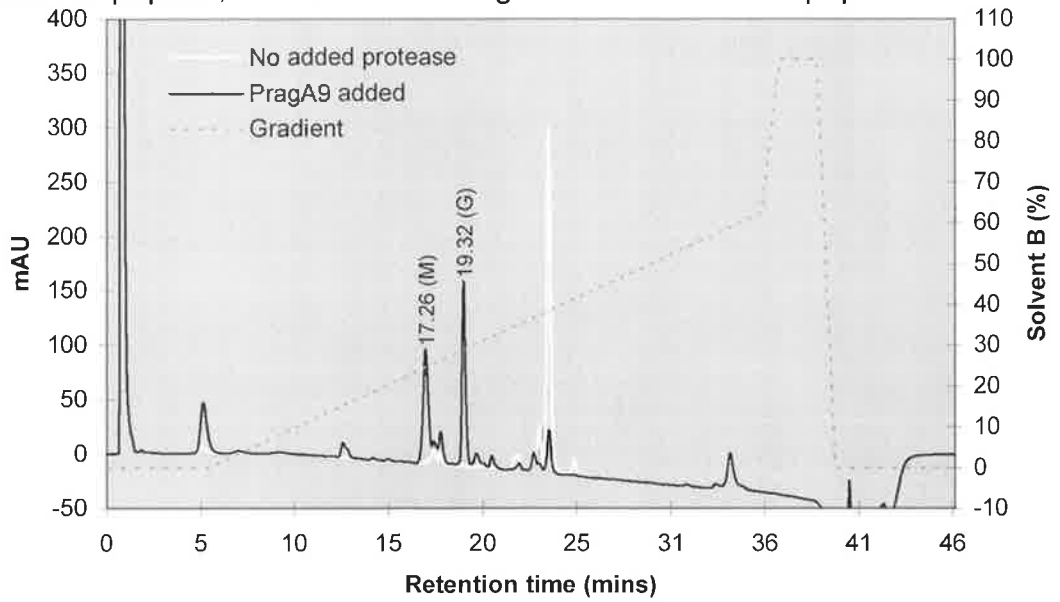


Figure 3.4.4.4 – HPLC trace of PragA33 protease cleavage of GPE-motif discrete peptide with P_1 Leu. The reaction trace is overlaid with full-length substrate incubated without added protease. L = N-terminal fragment of the cleaved peptide, D = Degradation of the peptide, U = Unidentified peak.

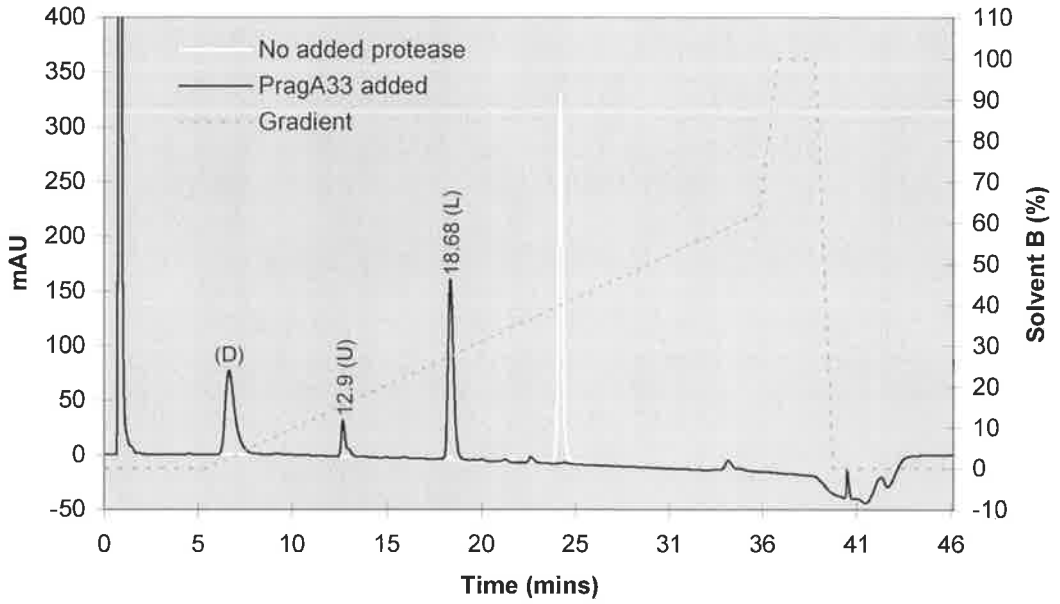


Table 3.4.4.1 - Retention times of the peaks detected by rp-HPLC in the cleavage reactions of the discrete peptides with the shortlisted α-Lytic protease mutants. Interpreted peak identities (see discussion) are indicated by the bracketed letters following the code outlined in the symbol legend.

Symbol legend:

M, A, F, L, H = Correct N-terminal fragment for each P_1 residue variant peptide.

T = Correct C-terminal fragment TLG-motif bearing peptides, NB: also possible in GPE-motif peptides (see discussion).

G = Correct C-terminal fragment GPE-motif bearing peptides.

U = Unidentified peak

I = Incomplete cleavage, a peak with the retention time of the full-length peptide is present in cleavage reaction.

D = Degradation, a large amount of unbound material on HPLC trace is detected.

tr. = Detection of a trace amount of this peak.

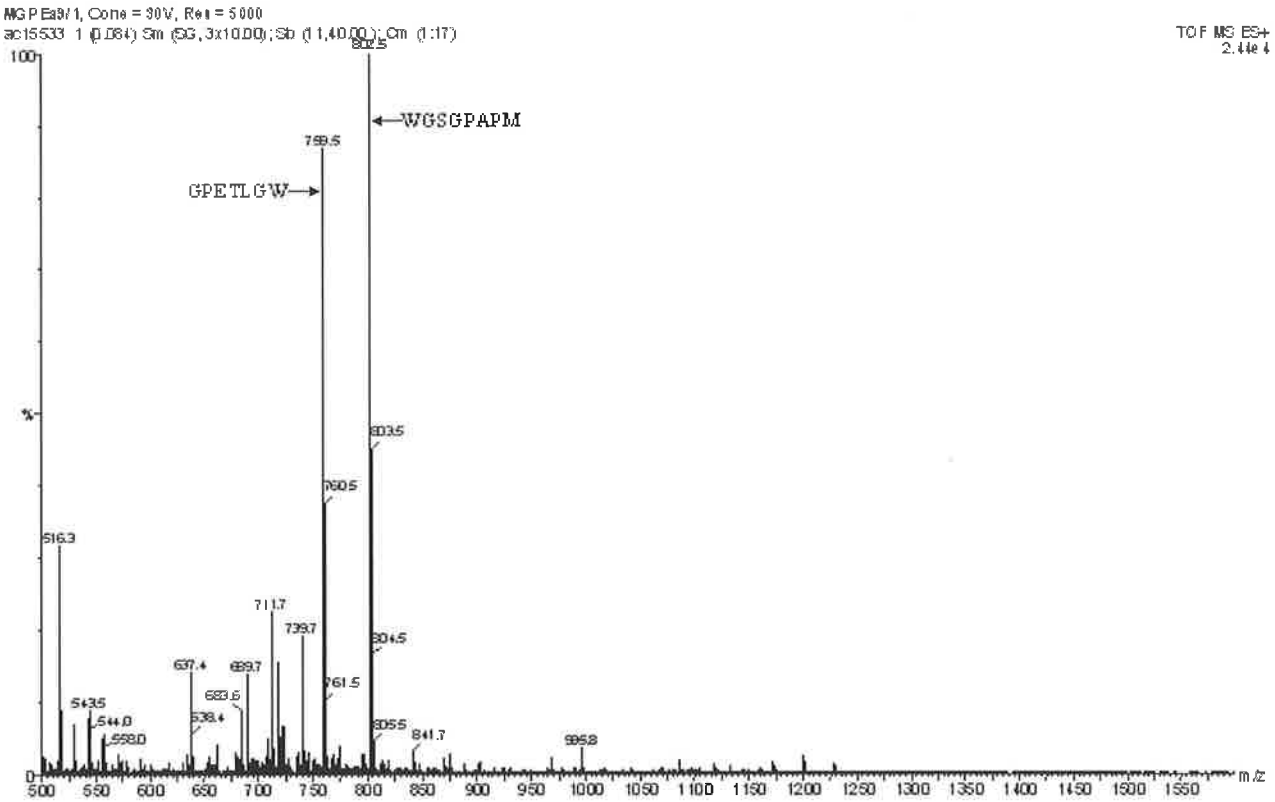
P_1 residue→	TLG motif peptides					GPE motif peptides				
	M	A	F	L	H	M	A	F	L	H
PragA9	17.26 (M) 18.08 (T)	14.82 (A) 18.12 (T)	18.16 (T) 20.02 (F)	18.16 (T) 18.59 (L)	14.15 (H) 18.15 (T)	17.26 (M) 19.32 (G)	tr.19.3 (G) (I)	18.19 (T) 20.0 (F) 22.6 (U) (D)	18.09 (T) 18.57 (L) 21.66 (U) (I)	14.20 (H) 19.41 (G) 20.1 (U)
PragA12	17.27 (M) 18.09 (T)	tr.14.8 (A) 18.07 (T) (D)	15.84 (U) 18.02 (T) (I)	18.04 (T) 22.72 (U) (I) (D)	14.15 (H) 18.14 (T)	17.75 (U) 18.13 (T) 19.38 (G) (I) (D)	15.87 (U) 18.01 (T) 20.46 (U) (I) (D)	17.96 (U) 19.77 (U) 23.64 (U) (I) (D)	18.04 (T) 18.57 (L) (I) (D)	14.18 (H) 15.85 (U) 19.4 (G)
PragA33	17.26 (M) 18.08 (T)	14.72 (A) 17.91 (T)	15.78 (U) 17.93 (T) 19.88 (U) (D)	17.95 (U) 18.1 (T) (D)	18.15 (T) 19.12 (U) (D)	13.0 (U) 17.27 (M) (D)	(D)	12.97 (U) 15.79 (U) 19.87 (F) (D)	12.9 (U) 18.68 (L) (D)	16.07 (U) 18.85 (U) (D)
PragA40	17.28 (M) 18.11 (T)	(I)	19.8 (U) (I)	18.16 (T) 20.12 (U) (I)	18.16 (T) 18.6 (U) (I)	(I)	(I) (D)	13.1 (U) 15.9 (U) 20.57 (U) (D)	12.9 (U) 18.6 (L) 21.7 (U) (D)	16.0 (U) 18.86 (U) 20.1 (U) (I) (D)
PragA43	12.8 (U) 17.3 (M) 18.17 (T)	14.88 (A) 18.16 (T) (D)	15.9 (U) 18.16 (T) 20.13 (F)	14.2 (U) 18.14 (T) (D)	14.4 (H) 17.1 (U) 18.18 (T)	13.08 (U) 17.3 (M) (D)	13.07 (U) 14.8 (A) 17.8 (U)	13.0 (U) 15.86 (U) (D)	12.9 (U) 18.6 (L) (D)	16.03 (U) (D)

P_1 residue→	TLG motif peptides					GPE motif peptides				
	M	A	F	L	H	M	A	F	L	H
			(D)		19.13 (U) (D)		(I) (D)			
RadX30	17.24 (M) 18.04 (T)	14.77 (A) 18.01 (T)	17.97 (T) 20.04 (F)	18.02 (T) 18.66 (L)	14.07 (H) 18.05 (T)	12.9 (U) 17.25 (M) (D)	12.98 (U) 14.8 (A) (D)	12.98 (U) 20.04 (F) (D)	18.6 (L) 21.67 (U) (D)	14.12 (H) tr.19.4 (G) (D)
RadX55	17.27 (M) 18.07 (T)	14.76 (A) 18.00 (T) (I)	15.85 (U) 18.06 (T) 19.95 (F) (I) (D)	(I) (D)	14.17 (H) 18.15 (T)	(I)	(I)	22.86 (U) (I) (D)	(I)	(I)
RadX63	17.26 (M) 18.07 (T)	14.78 (A) 18.06 (T)	17.74 (U) 19.7 (U) (I)	18.1 (T) 18.7 (L)	(I)	(I)	(I)	(I)	(I)	(I)

3.4.5 Mass Spectrometry of peptide cleavage.

Two dominant mass species were detected in the PragA9 cleavage reaction of the GPE-motif peptide with Met at P_1. These species had masses of 759.5 and 802.5 Da. These are in very close agreement (± < 1 Da) with the expected masses from correct proteolysis of the peptide of 758.8 and 801.9 Da, confirming cleavage at the correct site in this peptide by this mutant enzyme. The mass spectrum appears as Figure 3.4.5.1.

Figure 3.4.5.1 – Mass spectrometry of PragA9 cleavage of GPE-motif peptide with Met at P_1.



3.5 DISCUSSION

The least anticipated outcome of the screening assay was the very large number of mutants that appear to be able to cleave prior to the Gly-Pro-Glu motif. This result is contrary to the current understanding of the S' specificity of α-Lytic protease, which should suggest that Gly-Pro-Glu as P_1' to P_3' should be highly unfavourable (Schellenberger *et al.*, 1994). However, this latter work was performed with the wild-type protease, and was with acyl-transfer reactions rather than actual peptide hydrolysis. As such acyl-transfer reactions are essentially the reverse of peptide hydrolysis, with the acyl moiety transferred from the acyl-enzyme complex to a nucleophile peptide acceptor rather than released into solution by the attack from water (Schellenberger *et al.*, 1993). Since the alteration to the reaction sequence is limited to the order of events, and not the binding interaction of the substrates and enzyme, it has been suggested that the substrate specificity at the subsites should be identical to the proteolytic reaction (Fersht *et al.*, 1973). However, it is open as to how true a model of actual proteolysis these reactions represent.

Explanations for the observed pattern of results could be that either:

(1) The relative contribution of the P' residues is of insignificant importance compared to favourable P residues presence, as suggested by the work of Bauer *et al.* (1981). In that study the variation in the second order rate constant between the most and least favoured P_1' residues was only two-fold compared to over 16-fold for varying P_1 residues. The P_1 variation could have been much higher as well, as known unfavourable residues were not investigated. Or:

(2) The mutagenesis has indeed had an effect on the S' sites, as was suggested in their development (Graham *et al.*, 1993; Graham *et al.*, 1994), wherein otherwise equivalent peptides that varied only in their reporter leaving group (pNA chromophore or β-nap fluorophore) were hydrolysed at different rates, and that this rate discrepancy varied amongst the mutants.

Some confidence in the screening assay described in Section 3.4.2 can be gained from the mutant pairs RadX4/RadX25 and PragA41/RadX22. These enzyme pairs contained previously unnoticed duplicate mutations. In the case of the former pair the raw cleavage results agree quite strongly, with only unit differences in the cleavage of either of the peptide sets, and the abstracted index value reporting nearly identical properties. The latter pairing does display a larger variation, but this is primarily due to differences in the TLG peptide cleavage. Given that the GPE peptide was near the theoretical maximum of 20-25 mP units for both enzymes, small differences in the TLG peptide cleavage will have a large effect on the derived index value. The inherent technical difficulty of dispensing very small volumes of viscous culture supernatant could quite easily have given rise to the observed discrepancy. Despite this variability, both isolates still report as GPE favouring enzymes.

It is unfortunate that the screening assay was unable to provide greater discrimination between mutants favouring cleavage before the intended target and control motifs. This lack of discriminating power was due to the narrow operational window for the assay, of approximately 25% (around 25 mP units) difference between intact substrate peptide and the free fluorophore-bearing

peptide fragment. The difference in polarisation observed for the small free peptide versus the much larger avidin-peptide complex would have been expected to be much larger, with applications of the equivalent peptide-avidin system achieving a 150-250 mP difference between the free labelled peptide and the peptide-avidin complex (Levine *et al.*, 1997; Leissring *et al.*, 2003). It could be speculated that the lack of a large signal increase could be due the "propeller effect", wherein part of the fluorescently labelled entity freely rotates about a bond, resulting in a significantly lower polarisation signal than should be expected for its mass (Lynch *et al.*, 1997). Alternatively, it is plausible that the labelled peptide, whether full-length or cleaved fragment, could interact with buffer constituents or other proteins via the C-terminal lysines. If such electrostatic interaction occurred, it could effectively dampen the depolarisation of the free peptide by giving the peptide a larger apparent mass. However, this theory does not satisfy the lack of a larger polarisation increase for the full-length peptide in the presence of avidin. Whatever the underlying cause, it seems that the issues with this screening protocol are FP based technical difficulties. Attempts to circumvent the FP aspect of the assay, by binding the peptides to avidin coated microplates, washing then reading the residual fluorescence, were wholly unsuccessful. This could have been due to the in-house coating of the plates (using the coating protocol of Section 2.3.7) potentially having low efficiency or not presenting the avidin in an accessible orientation for optimal access to the biotin moiety. Recently similar assays have been successfully performed, wherein peptides with comparable structure to the FP peptide sets presented here, have been used for a non-FP based protease assay. In these examples, the intact peptides

were pulled down with streptavidin coated magnetic beads, and the remaining fluorescence signal in the supernatant measured and observed to be proportional to the degree of proteolysis of the peptide (Baechle *et al.*, 2005; Kozlov *et al.*, 2006).

The screening protocol achieved its goal, demonstrating that the GPE motif of the N-terminus of IGF-I does not appear to prevent cleavage by α-Lytic protease cleavage, at least in this peptide model. To attempt to make sense of the large number of hits, a shortlist of mutant enzymes with relevant properties was created. Mutants were selected for further investigation with insight into the number of equally preferred P_1 residues, and the mutations known to confer substrate specificity. Except for the unique case of PragA40 (see later), mutants were chosen where they had a maximum of two favoured P_1 residues, as these were most likely to be more specific enzymes, and likewise mutants with unknown P_1 preference were not considered. The three mutants that met these criteria and displayed the largest preference for cleavage before GPE were selected, except that PragA33 was included over RadX37 due to higher cleavage scores and higher overall protease activity. The final GPE-favouring mutant that was included, RadX55, although showing only slight preference for GPE, was selected due to it being the only enzyme that both favoured GPE cleavage and bore the M213H mutation. This mutation is known to confer a P_1 preference for Met and His (Graham *et al.*, 1993), and is one of the mutations present in the benchmark site-specific α-Lytic protease mutant, PragA9 (Lien *et al.*, 2001; Lien *et al.*, 2003.). The remaining shortlisted mutants were selected based on the mutations present in RadX55 compared to PragA9. Of the few other enzymes that bore the

R192E mutation of RadX55, RadX63 was selected, as it only differs from PragA9 by two mutations (aside from the R192E change under consideration). Moreover, it was amongst the most GPE-disfavouring, thus serving as a contrasting alternative for comparison. The other mutation of RadX55 that separates it from PragA9 is V218S, which in PragA9 is V218L. The mutant PragA12 contains all the mutations of PragA9 as well as having V218S, but was not found to have sufficient activity in the FTC-casein screen of Chapter 2 to warrant evaluation in the peptide screening assay. However, purification, and thereby concentration, of this mutant was performed, allowing its inclusion into the shortlist for investigation, despite the initially low activity and apparent expression. It was hoped that, in these selections, there may be an opportunity to gain insight into the role of the mutations at different positions, and the resulting structures that confer apparent GPE cleavage preference. The two other mutants that were included in the shortlist are PragA40 and PragA9. The PragA40 mutant represents a unique case, in that it appeared to show little discrimination between the P' sequences, cleaving both to the TLG and GPE-motif peptide sets to the same degree. PragA9 is included as the reference site-specific mutant against which the previous assays had been calibrated.

The action of the shortlisted α-Lytic protease mutants was studied using longer discrete peptide substrates, each individually bearing one of the five P_1 residue variants of the FP-peptide sets. The peptides utilised were a slight departure from the idealised IGF-I cleavage motif of the FP peptides; but peptides with this architecture have been shown in a previous study to be a good model for release of full-length IGF's from their fusion proteins (Lien *et*

al., 2001). The variable P_1 positions and P' motifs employed in the screening peptides allowed for rapid probing whilst avoiding the complications of analysing the complex cleavage pattern arising from the use of a full-length protein.

The activity of the α-Lytic protease mutants against the longer, discrete peptides allows interpretation of some aspects of the FP screening data. Table 3.4.4.1 displays the peaks detected in the cleavage reactions. The most striking feature of the data is the large number of mutants that rapidly degrade the peptides; where the term "degrade" is used to signify non-specific cleavage of the substrate at multiple sites. This phenomenon was particularly prevalent in those peptides that bear the GPE motif, as shown on HPLC by a significant level of unbound material eluting in the column load effluent (for examples see Figure 3.4.4.2 and 3.4.4.4). The putative GPE-favouring mutants from the FP-screening assay (PragA33, PragA43 and RadX30) were observed to degrade all the GPE-bearing peptides, regardless of the P_1 residue presented. However, these three mutants were also able to cleave TLG-motif peptides as in many cases this cleavage was observed to be specific (as opposed to non-specific degradation) yielding identifiable product peaks. In short, the GPE-favouring mutants identified in the FP-screen were able to cleave both GPE and TLG-motif bearing peptides, but with differing specificity. The converse appears to hold true, exemplified by the strongly TLG-favouring mutant, RadX63, which is unable to cleave the GPE-motif peptides at all, yet it did specifically cleave many of the TLG-motif peptides. PragA9, a slightly TLG-favouring mutant cleaved all TLG-motif peptides specifically and yet was unable to efficiently cleave two of the GPE-motif

peptides and degraded one. PragA40, the enzyme that showed little P' preference in the FP screen, produced mixed results. It produced product peaks in more P_1 variants of the TLG-motif peptides than the GPE-motif peptides, yet degraded those GPE-motif peptides on which it did act. Again, this is largely in keeping with its assignment from the FP assay. The only exception to this trend is RadX55, which specifically cleaved and/or degraded the TLG-motif peptides to a greater extent than it did the GPE-motif peptides. However, RadX55's deduced status as GPE-favouring was always marginal, at best, and its inclusion in the shortlist was simply because it was the only known enzyme that bore the M213H mutation that could be considered GPE-favouring.

It is therefore quite apparent that whilst the FP screening protocol did indeed return enzymes that could act on a peptide bearing a GPE-motif in positions P_1' to P_3', the selections it returned were not necessarily cleaving the peptide specifically. Assuming that the actions of the mutants on the discrete peptides are a reasonable model for the observations from the FP assay, it appears that the screening selected in favour of enzymes that non-specifically degrade their peptide substrates. This phenomenon could be partially attributed to the observation that the mutants that appeared to favour cleavage of the GPE peptide over the TLG peptide in the FP screen are generally those with broader P_1 substrate specificity. As shown in Table 3.4.2.2, the mutants that favour cleavage of the GPE peptide have a higher proportion of representatives showing broad P_1 specificity; with 44% having 3 to 4 equally favoured P_1 residues, compared to the 83% of TLG preferring enzymes that have only 1 to 2 favoured P1 residues. So even by restricting

shortlist selection to those mutants with a small number of favoured P_1 residues, the putative GPE-favouring mutants isolated by this FP assay system tend to be relatively non-specific enzymes in general. It is possible that the presence of the GPE sequence in the peptides forces the candidate mutants to search for secondary specificities for cleavage. This represents a quite different level of assay selection pressure compared with TLG, which accommodates the preferred P_1 site in strong competition to any secondary non-preferred sites.

The cleavage of the discrete peptides does provide insight into the utility of the enzymes for the required task of Met-pGH(1-11)VN-IGF-I cleavage. Careful examination of the retention times of the peaks generated provides a wealth of useful information. For example, the TLG-motif peptides, from which many of the mutants were able to specifically produce two cleavage products. Although identity of these cleavage products was not confirmed directly, they could be deduced to represent the predicted N and C-terminal fragments of the full-length peptide. Hence, we see that a peak at 18.03 to 18.17 mins occurs in all cases. As the C-terminal fragment would be identical regardless of which P_1 variant of the peptide was cleaved, it would always have the same retention time, within experimental error. It is reasonable to assume that this peak corresponds to the C-terminal fragment of the peptide; as the addition or deletion of amino acids would most likely produce a change in hydrophobicity and thus retention time in a 4 amino acid peptide. The other peak in the two product reactions should therefore correspond to the N-terminal fragment and be identical for that P_1 variant. This is indeed the case, with a peak of fixed retention time present in all two

product generating cleavage reactions of the TLG-motif peptide that share the same P_1 variant. The retention time range of this peak is:

- 17.24 to 17.3 minutes for P_1 Met.
- 14.77 to 14.88 minutes for P_1 Ala.
- 17.97 to 20.13 minutes for P_1 Phe.
- 18.59 to 18.7 minutes for P_1 Leu.
- 14.07 to 14.2 minutes for P_1 His.

Although much rarer, two product cleavage reactions for the GPE-motif peptides did occur, but not frequently enough for the above interpretation to be made. However, the N-terminal fragment arising from specific cleavage at the correct site of a GPE-motif peptide should be identical to its corresponding P_1 variant TLG-motif counterpart. This is observed, with some of the mutants giving rise to peaks of the correct retention time for a supposedly identical N-terminal fragment. By the above logic, with a correct N-terminal fragment, the other peak in a two product reaction must be the all important GPE C-terminal fragment or a derivative of it. Unfortunately, there are very few two product GPE-motif reactions that do not suffer from degradation. In the one case containing only two products, that of PragA9 acting with Met at P_1, there is indeed a second peak at 19.32 rninutes. The identity of these two peaks was determined to be correct by mass spectrometry (Figure 3.4.5.1), demonstrating that PragA9 is able to specifically cleave after PAPM prior to the N-terminal three residues of IGF-I.

3.6 CONCLUSION

The FP screening assay reported in this chapter has been shown to identify enzymes that are able to cleave a peptide bearing a GPE motif. This chapter demonstrates that whilst Gly-Pro-Glu remains a difficult motif for α-Lytic protease mutants to cleave specifically, it is indeed possible. Quite unexpectedly it was found that PragA9, a mutant included in the discrete peptide investigation simply as a control, turned out to be the one best able to correctly cleave at the correct site. Since PragA9 has been demonstrated to be able to correctly cleave at this site, the barriers to Met-pGH(1-11)VN-IGF-I cleavage by α-Lytic protease must be largely structural. Therefore, alteration of the structure of the substrate, within the constraints of the existing fusion protein sequence, will be investigated by manipulation of the cleavage conditions.

4.1 INTRODUCTION

In the last chapter it was established that, contrary to the limited literature on the S' specificity of α-Lytic protease, it is indeed possible using at least one mutant to cleave specifically before the Gly-Pro-Glu N-terminal sequence of IGF-I. This provides encouraging evidence that IGF-I fusion protein cleavage with an α-Lytic protease mutant may in fact be possible. With cleavage of a peptide model verified, it is now necessary to investigate the activity of PragA9 against the full-length fusion protein to determine both liberation of the desired product with an authentic N-terminus, and the absence of adventitious cleavage elsewhere within the protein.

There is anecdotal evidence that this selected mutant is unable to liberate IGF-I from the customary fusion protein (King *et al.*, 1992), even when provided the same P_4 to P_1 cleavage motif (Lien *et al.*, 2001) employed in both the FP peptide screening and long peptide reactions of Chapter 3. Since the amino acid sequence of the site itself is now known to not be an absolute barrier to successful cleavage, any inability to liberate the correct N-terminus of IGF-I from an equivalent cleavage motif must have some structural element. The variable cleavage of GPE-motif bearing peptides in Chapter 3 provides some preliminary support to this hypothesis.

There is significant evidence for the role that protein structure can play in the differential recognition of given amino acid sequence motifs. The literature is replete with examples of normally site-specific proteases failing to act at their purported ideal sequence, or more prevalently, cleaving the target protein at sites other than the expected recognition motif. Adventitious cleavage of the desired protein product has been frequently reported for three

most commonly employed site-specific endoproteases: Enterokinase (Light *et al.,* 1980; Jacquet *et al.,* 1999; Choi *et al.,* 2001; Liew *et al.,* 2005), Factor Xa (Nagai *et al,* 1985; Quinlan *et al.,* 1989; Belmouden *et al.,* 1993; Holowachuk and Ruhoff, 1995; Jacquet *et al.,* 1999) and Thrombin (Forsberg *et al.,* 1991; Thanabalu *et al.,* 1992; Kwon *et al,.* 2005; Park *et al.,* 2005). This limited subset of examples probably represents the "tip of the iceberg" with reporting of incorrect proteolytic processing likely to take second place to the reporting of successful resolution of such issues.

Where cleavage occurs at a noncanonical sequence, it can be concluded to be the result of altered cleavage specificity due to protein structure or sequence, or both. The structural effect could be thought of as a positive or negative influence on protease substrate recognition; by either increasing or decreasing the rate of cleavage at the expected site, or promoting cleavage at the secondary site(s). This effect was demonstrated elegantly by Liew *et al.* (2005), wherein enterokinase activity against an internal cleavage site in the target protein was found to be higher than against the intended site in the linker sequence. When the amino acid sequence of this preferred internal site was inserted in the linker region, in place of the traditional enterokinase cleavage motif, it too was found to be poorly processed when at this location.

Alternatively, in a fusion protein consisting of chicken IGF-II and a longer version of the Met-pGH studied in this chapter (46 residues in that example as compared to the 11 employed here), cleavage by H64A subtilisin has been reported at an unexpected Gln-Phe linkage in the fusion partner (Upton *et al.*, 1995). Cleavage at this non-canonical site has also been

observed in fusions of human IGF-I with this extended Met-pGH fusion partner (G. Francis, personal communication).

As structural analyses of IGF-I reveal an N-terminus that is free and solvent exposed in its native state (Cooke *et al.,* 1991; Laajoki *et al.,* 2000), this should render it an attractive target for proteolytic cleavage (Figure 4.1.1). However, this accessibility appears to be altered significantly by the introduction of the leader peptide in this fusion protein. The structure of Met-pGH(1-11)VN-Arg^{3}-IGF-I (Laajoki *et al.* 1997; Laajoki *et al.* 2000), a very closely related protein differing from the construct in question by only the Glu^{3} to Arg^{3} amino acid substitution and the absence of the Pro-Ala-Pro-Met cleavage motif, provides a plausible explanation for the inability of α-Lytic protease to release a native N-terminus for IGF-I. This structure, displayed as Figure 4.1.2, shows the leader peptide forming a tight loop at the site of the N-terminus of IGF-I proper, conceivably limiting the accessibility of α-Lytic protease to the site. This model of the solution structure of the fusion protein suggests a turn-like structural element between Asn^{-1} and Arg^{3}, although the Φ and Ψ bond angles do not correspond to any previously characterized β-turn (Laajoki *et al.*, 2000). Moreover, interactions between Phe^{-3} and Val^{-2} of the leader and Helix 1 (specifically Cys^{6} and Gly^{7}) of the IGF-I molecule were detected, suggesting that the leader may "fold-back" onto the hydrophobic core of the IGF-I structure, stabilising the constraint at the cleavage target site (Laajoki *et al.,* 2000).

Successful cleavage of this fusion protein with α-Lytic protease may therefore require a relaxation of the structure of the substrate. Of the forces that stabilise the tertiary structure of a protein, hydrophobic interaction is

probably the most significant in this case, given the hydrophobicity of the leader peptide and the "fold-back" phenomenon observed in Met-pGH(1-11)VN-Arg^3-IGF-I. Determination of environmental conditions that allow the leader sequence to become more solvent exposed may reduce the steric constraint around the cleavage site, although any conditions employed must be sufficiently mild to avoid denaturation of the protease.

In this chapter the lead mutant α-Lytic protease isolated by the peptide screening process of the previous chapters, PragA9, will be used to cleave the full-length fusion protein. The reaction conditions will be manipulated to optimise the production of authentic IGF-I.

Figure 4.1.1 – The solution structure of human IGF-I of Laajoki *et al.,* (2000). The N-terminus appears solvent exposed and relatively unstructured.

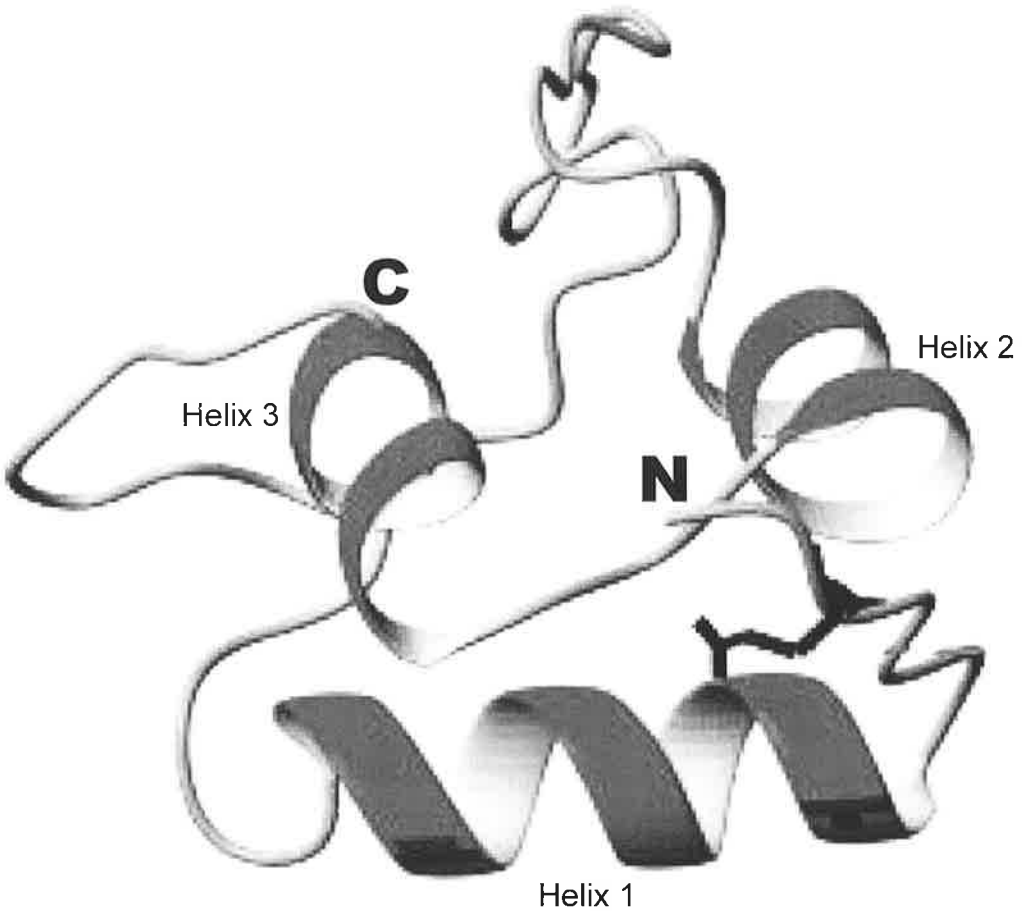
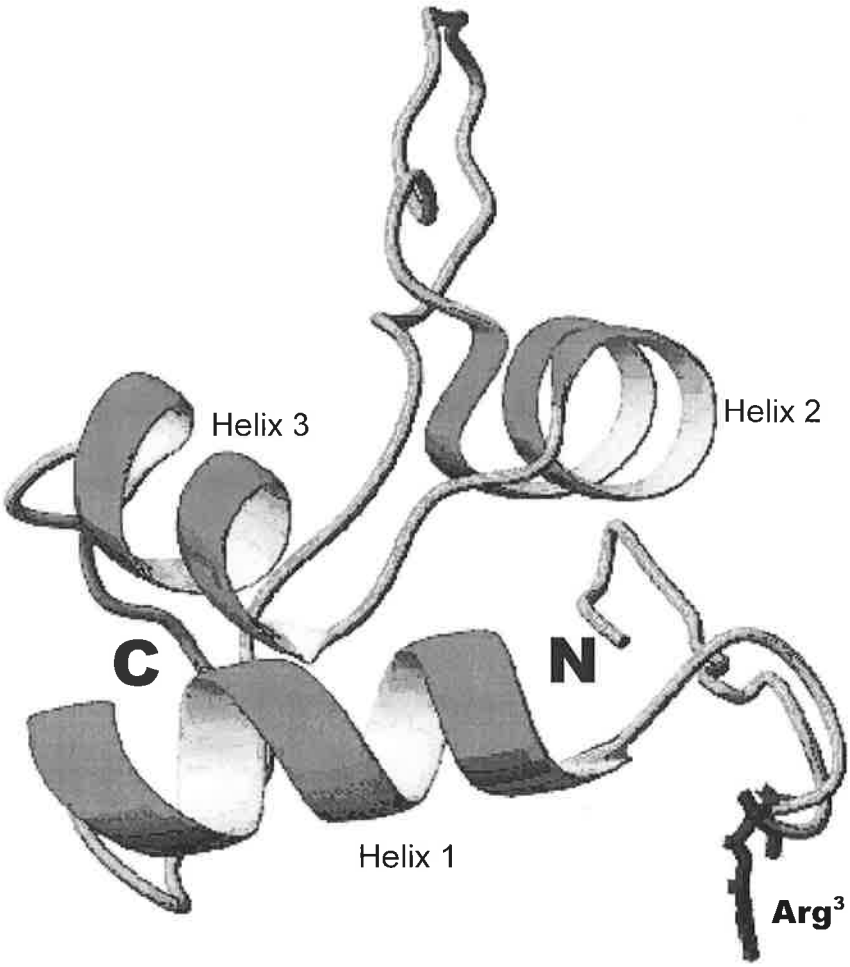


Figure 4.1.2 – The solution structure of Met-pGH(1-11)VN-Arg^3-IGF-I provided by Laajoki *et al.,* (2000). The fold-back constraint at the site that the cleavage motif would occupy is visible in the vicinity of the indicated Arg^3 residue.



4.2 MATERIALS

4.2.1 General reagents & materials

The following reagents were obtained from BDH chemicals, Kilsyth, Victoria, Australia: HEPES, glycine, arginine, tris-(hydroxymethyl)methyl-amine (Tris), polyoxyethylene (20) sorbitan monolaurate (Tween-20), sodium dodecyl sulphate (SDS), trifluoroacetic acid (TFA), acetonitrile and urea. All

were of analytical reagent grade or higher. Analytical grade iodoacetic acid and dithiothreitol (DTT) were supplied by Sigma-Aldrich, Castle Hill, NSW, Australia. PD-10 desalting columns were supplied by Amersham Bioscience (Uppsala, Sweden).

SDS-PAGE materials; NuPAGE Novex 4-12% Bis-Tris 1 mm minigels, NuPAGE 4X LDS Sample buffer, NuPAGE MES SDS 20X Running Buffer, Mark12™ unstained molecular weight standard, were supplied by Invitrogen, CA, USA.

4.2.2 General solutions

Cleavage buffer

4M Urea

50mM Tris

pH8.5

Denaturation buffer

8M Urea

25mM Tris

10mM DTT

pH8.5

IAA solution

64mg/ml Iodoacetic acid

1M Tris

pH8.5

4.2.3 Purified PragA9 α-Lytic protease.

Purified PragA9 mutant α-Lytic protease was supplied by Novozymes GroPep Ltd., Thebarton, S.A., Australia. It was quantified to 0.6 mg/ml by

absorbance at 280 nm and stored at 4°C in 25 mM Na Tricine buffer (as in Section 3.2.5 and 3.3.2), pH 6 with 0.02% w/v sodium azide.

4.2.4 Fusion protein substrates

Receptor grade Met-pGH(1-11)VN-IGF-I was obtained from Novozymes GroPep Ltd. Thebarton, S.A., Australia. Semi-pure Met-pGH(1-11)VN-PAPM-IGF-I was provided by I. Butler, Novozymes GroPep Ltd, Thebarton, S.A., Australia. Both substrates were made up from lyophilised protein to a stock concentration of 10 mg/ml in 10 mM HCl and stored at 4°C until needed.

4.3 METHODS

4.3.1 Conduct of cleavage reactions.

For all reactions 10 µg of substrate protein at 50 µM was incubated overnight (unless otherwise stated) with the buffer conditions, temperatures and enzyme to substrate ratio stated in Table 4.4.1.1. The initial reaction conditions replicated those of the discrete peptides of Chapter 3, and were altered in a stepwise, iterative fashion. Reactions were terminated by the addition of 130 µl of 1.5% TFA.

4.3.2 Analysis of cleavage reactions by HPLC

The cleavage reactions, with the exception of Met-pGH(1-11)VN-IGF-I and S-carboxymethylated substrate derived reactions, were analysed by

reversed phase HPLC (rp-HPLC) on an Agilent 1100 HPLC system (Agilent technologies, California, USA) using a C4 column (Brownlee Reverse-Phase, BU-300, Aquapore Butyl C-4, 2.1 mm I.D. x 100 mm length, Perkin Elmer, MA, USA.). Separation was by linear gradient of 20 – 50% acetonitrile over 30 minutes in 0.1% TFA at a flow rate of 0.5 ml/min, with monitoring of eluted protein by absorbance at 215 nm. The analyses were verified performed by subjecting the product of reaction condition #37 and authentic IGF-I to resolution on a slower gradient between 35 – 55% acetonitrile over 80 minutes at 0.5 ml/min.

Later in the work of this chapter the C4 Aquapore column was replaced and reactions of Met-pGH(1-11)VN-IGF-I were analysed on a C4 Vydac column (Vydac 214TP54, 4.6 mm ID x 250 mm length. Grace Vydac, California, USA) under the same conditions, but with a 1 ml/min flow rate. S-carboxymethylated cleavages and controls were also analysed on this column employing a 20 – 64% acetonitrile gradient over 50 minutes at 1 ml/min.

4.3.3 SDS-PAGE of cleavage reactions

Reaction condition #37 was repeated and the product compared with undigested substrate and authentic IGF-I on a 4-12% NuPAGE gel in MES buffer under reducing conditions according to the manufacturer's instructions. Protein bands were visualised as in Section 2.3.6.

4.3.4 Mass Spectrometry of peptide cleavages.

The whole cleavage reaction (no separation) of condition #37 (Table 4.4.1.1), but without included glycine, was desalted by rp-HPLC and the entire gradient elution dried by vacuum centrifugation. Mass spectrometry was performed by L. Hick, University of Wollongong, NSW, Australia.

4.3.5 S-Carboxymethylation reactions.

S-Carboxymethylation of Met-pGH(1-11)VN-PAPM-IGF-I and authentic IGF-I was carried out in a manner similar to that described by Crestfield *et al.* (1963). Briefly, 3 mg of protein was incubated in denaturation buffer at a concentration of 1 mg/ml for 1 hour at room temperature with gentle shaking. To this was added 600 µl of IAA reagent and the reactions incubated in a light-protected tube for 15 mins at room temperature with gentle shaking. The reaction was loaded immediately onto a PD-10 column to desalt the S-carboxymethylated proteins into cleavage buffer.

4.4 RESULTS

4.4.1 HPLC analysis of variations in cleavage conditions

The starting Met-pGH(1-11)VN-PAPM-IGF-I substrate used in the cleavage reactions of this chapter appears as Figure 4.4.1.1. Two dominant peaks are visible with retention times of 22.3 and 23.4 minutes. These peaks most likely correspond to two different folding isomers of Met-pGH(1-11)VN-

PAPM-IGF-I (see Discussion), as the sample runs as a single band on SDS-PAGE (Figure 4.4.2.1).

The earliest reactions of Met-pGH(1-11)VN-PAPM-IGF-I with α-Lytic protease varied the enzyme:substrate mass ratio (reaction conditions #1 to 6, Table 4.4.1.1), with 1:100 (condition #4) proving to be the optimum point between complete conversion of substrate to cleavage products without incurring significant overdegradation. This reaction was thus set as the default for the remainder of the study of cleavage conditions. The HPLC trace of this reaction appears as Figure 4.4.1.2. With a single broad peak at 18.15 minutes, this reaction contains no material with a retention time corresponding to authentic human IGF-I, shown to be 17.5 minutes in Figure 4.4.1.3.

The incorporation of urea into the cleavage reactions (conditions #7 to 34, Table 4.4.1.1) caused a shift in the pattern of peaks obtained from the reactions, specifically the observance of material eluting with a retention time very similar to that of authentic IGF-I. This material was observed as a leading shoulder on the initial broad peak of the default conditions when lower concentrations of urea were included (for example see Table 4.4.1.1, conditions #7, 11, 16, 20 and 31). The abundance of this material and its ability to be resolved into a distinct separate peak was increased with increasing urea concentrations (for example see Table 4.4.1.1 conditions #9, 22, 29, 26 and 33 and Figure 4.4.1.4). Alteration of the reaction pH (conditions #7 to 34, Table 4.4.1.1) did not appear to have an effect on the generation of this IGF-I-like material. The highest abundance of the IGF-I-like material with the greatest degree of resolution into a distinct peak was achieved by condition #26, at 4 M urea, pH 8.5 and incubation at 37°C. A

HPLC trace of this reaction appears as Figure 4.4.1.4, wherein a significant peak with a retention time of 17.34 minutes is visible.

Reduction of the incubation temperature of the reaction was also able to promote the formation of the IGF-I-like material. Condition #35 differed from condition #4 by only the decrease in temperature from 37°C to ambient (20 to 22°C), yet as shown by condition #35 in Table 4.4.1.1 and in Figure 4.4.1.5, a product profile displaying the leading shoulder was achieved. These two IGF-I-like peak promoting factors, ambient incubation and 4 M urea, could be combined to synergistically in condition #37 to improve the abundance and resolution of the IGF-I-like peak to its highest attainable yield in this set of experiments. The HPLC trace of reaction condition #37 is shown in Figure 4.4.1.6, wherein the resolution and abundance of the peak at 17.6 minutes is greater than in either condition #26 (4 M urea, 37°C) in Figure 4.4.1.4 or condition #35 (default condition but ambient incubation) in Figure 4.4.1.5. For comparison of the retention times of the IGF-I-like peak with that of authentic IGF-I, the traces of Figures 4.4.1.3 (authentic IGF-I) and 4.4.1.6 (reaction condition #37) are overlaid. It appears that the retention times of authentic IGF-I standard and the IGF-I-like peak are quite well matched given the complexity of the cleavage reaction sample. Furthermore, considering that variations in retention time of up to ± 0.15 minutes were encountered between comparable peak species in Chapter 3, Section 3.4.4, the differences observed here can be considered to not be significant.

Further confirmation of the identity of the peak as IGF-I comes from running the apparently successful cleavage reaction products and authentic IGF-I on a HPLC with a much slower gradient. Figure 4.4.1.8 shows a peak

with the same retention time persists in both samples under these separation conditions. Additionally, co-injection of authentic IGF-I standard with the cleavage reaction showed the putative and known IGF-I peaks co-eluting, although the very large peak that resulted tended to dominate the chromatogram, obscuring the remaining signal (data not shown).

Initial examination of Table 4.4.1.1 suggests that condition #36 (4°C, 48 hour incubation) was highly successful at the generation of the IGF-I-like peak. However, the apparently high productivity in this reaction was due to a dual peak species around the retention time of IGF-I. As only one half of the split peak at IGF-I-like retention time could correspond to authentic IGF-I, the actual yield of this reaction is likely to be closer to $\leq$ 20%. Furthermore, the presence of the complex peak pattern indicates that even if authentic IGF-I is generated by this reaction, it could be in association with a modified form of IGF-I. Such product-related contaminants would make subsequent purification events more difficult, and the extended incubation times are undesirable from a process perspective. Hence such conditions were not investigated further.

Table 4.4.1.1 – Key data of cleavage reactions of Met-pGH(1-11)VN-PAPM-IGF-I by PragA9 α-Lytic protease with varied incubation conditions. The enzyme:substrate mass ratio [Ratio (mass)] was first manipulated in reaction conditions #1 to 6, then the optimum held constant. In reaction conditions #7 to 34 urea was included at 1, 2, 3 and 4 M, and the pH at each of these urea concentrations tested from pH 6.5 to 9.5 in 0.5 unit increments. The effect of temperature on the reaction was examined by conditions #35 to 38. The inclusion of large ionic species, non-ionic and ionic detergents was tested by conditions #39 to 49. For clarity, only peaks not present in the starting substrate (with the exception of controls #1 to 3) are noted as "main peaks". The retention times (R.T.) of the main peaks detected are expressed in minutes, and the percentage of the total peak area of the chromatogram that each peak represents of shown in brackets following its retention time.

Condition number	**Ratio (mass)**	**Buffer, pH**	**Other conditions**	**Main peaks R.T. & % area**	**% IGF-I like peak**	**Observations**
Control 1	no protease	100mM HEPES, pH8.5	37°C	22.3 (32%), 23.42 (28%)	-	Two peaks visible. See Figure 4.4.1.1.

Condition number	Ratio (mass)	Buffer, pH	Other conditions	Main peaks R.T. & % area	% IGF-I like peak	Observations
Control 2	no protease	50mM Tris, 40mM Glycine, pH8.5	1M Urea, 37°C	22.3 (38%), 23.41 (29%)	-	No change from Control 1. Incubation at other pH values also gave the same outcome as this control.
Control 3	no protease	50mM Tris, 40mM Glycine, pH8.5	4M Urea, 37°C	22.29 (38%), 23.38 (27%)	-	No change from Control 1. Incubation at other pH values also gave the same outcome as this control.
1	1:1000	100mM HEPES, pH8.5	37°C	16.8 (23%), 18.2 (32%), 18.9 (16%)	-	Three peaks merged in large group, significant full-length substrate.
2	1:500	100mM HEPES, pH8.5	37°C	17.0 (17%), 18.25 (38%), 18.96 (17%)	-	Three peaks merged in large group, reduced separation, significant full-length substrate

Condition number	Ratio (mass)	Buffer, pH	Other conditions	Main peaks R.T. & % area	% IGF-I like peak	Observations
3	1:250	100mM HEPES, pH8.5	37°C	18.18 (43%), 18.95 (25%)	-	Two peaks almost merged
4	1:100	100mM HEPES, pH8.5	37°C	18.16 (82%)	-	Majority of material in one broad peak. See Figure 4.4.1.2.
5	1:50	100mM HEPES, pH8.5	37°C	18.2 (49%), 18.4 (36%)	-	Almost dual peak, total peak area somewhat reduced
6	1:25	100mM HEPES, pH8.5	37°C	18.3 (42%), 18.5 (43%)	-	Almost dual peak, total area considerably reduced.
7	1:100	50mM Tris, 40mM Glycine, pH6.5	1M Urea, 37°C	17.34 (22%), 18.49 (31%), 19.16 (27%)	22	IGF-I-like peak detectable as leading shoulder
8	1:100	50mM Tris, 40mM Glycine, pH6.5	2M Urea, 37°C	17.3 (20%), 18.45 (28%), 19.2 (17%)	20	Many minor peaks, IGF-I-like peak detectable as leading shoulder

Condition number	Ratio (mass)	Buffer, pH	Other conditions	Main peaks R.T. & % area	% IGF-I like peak	Observations
9	1:100	50mM Tris, 40mM Glycine, pH6.5	3M Urea, 37°C	17.26 (16%), 18.44 (20%), 19.17 (13%)	16	Some full-length, IGF-I-like peak almost separate peak
10	1:100	50mM Tris, 40mM Glycine, pH6.5	4M Urea, 37°C	17.34 (11%), 18.5 (15%), 19.23 (10%), 19.9 (8%)	11	Many minor peaks, much residual full-length substrate, IGF-I-like peak little more than shoulder
11	1:100	50mM Tris, 40mM Glycine, pH7	1M Urea, 37°C	17.33 (19%), 18.46 (34%), 19.13 (19%)	19	IGF-I-like peak detectable as leading shoulder
12	1:100	50mM Tris, 40mM Glycine, pH7	2M Urea, 37°C	17.24 (19%), 18.4 (32%), 19.0 (16%)	19	Trace residual full-length substrate, IGF-I-like peak detectable as leading shoulder
13	1:100	50mM Tris, 40mM Glycine, pH7	3M Urea, 37°C	17.37 (19%), 18.5 (30%), 19.15 (18%)	19	Some residual full-length substrate, IGF-I-like peak almost separate peak

Condition number	Ratio (mass)	Buffer, pH	Other conditions	Main peaks R.T. & % area	% IGF-I like peak	Observations
14	1:100	50mM Tris, 40mM Glycine, pH7	4M Urea, 37°C	17.26 (16%), 18.38 (22%), 19.07 (15%)	16	Significant full-length substrate, IGF-I-like peak detectable but not as distinct separate peak
15	1:100	50mM Tris, 40mM Glycine, pH7.5	1M Urea, 37°C	18.3 (39%), 19.0 (18%)	-	No IGF-I-like peak, leading shoulder present
16	1:100	50mM Tris, 40mM Glycine, pH7.5	2M Urea, 37°C	17.3 (20%), 18.5 (39%), 19.0 (16%)	20	IGF-I-like peak detectable as leading shoulder
17	1:100	50mM Tris, 40mM Glycine, pH7.5	3M Urea, 37°C	17.3 (22%), 18.46 (33%), 19.1 (16%)	22	IGF-I-like peak almost separate peak
18	1:100	50mM Tris, 40mM Glycine, pH7.5	4M Urea, 37°C	17.2 (18%), 18.4 (22%), 19.0 (15%)	18	Some residual full-length substrate, IGF-I-like peak detectable but not as distinct separate peak

Condition number	Ratio (mass)	Buffer, pH	Other conditions	Main peaks R.T. & % area	% IGF-I like peak	Observations
19	1:100	50mM Tris, 40mM Glycine, pH8	1M Urea, 37°C	18.16 (23%), 18.4 (19%), 19.15 (19%)	-	Majority of material in one broad peak
20	1:100	50mM Tris, 40mM Glycine, pH8	2M Urea, 37°C	17.4 (22%), 18.5 (35%), 19.26 (15%)	22	IGF-I-like peak detectable as leading shoulder
21	1:100	50mM Tris, 40mM Glycine, pH8	3M Urea, 37°C	17.5 (20%), 18.6 (31%), 19.2 (17%)	20	IGF-I-like peak almost separate peak
22	1:100	50mM Tris, 40mM Glycine, pH8	4M Urea, 37°C	17.5 (21%), 18.6 (25%), 19.2 (19%)	21	Some full-length substrate, IGF-I-like peak is clearly separate peak
23	1:100	50mM Tris, 40mM Glycine, pH8.5	1M Urea, 37°C	18.16 (51%), 19.16 (17%)	-	Majority of material in one broad peak
24	1:100	50mM Tris, 40mM Glycine, pH8.5	2M Urea, 37°C	17.57 (19%), 18.56 (36%), 19.26 (20%)	19	IGF-I-like peak detectable as leading shoulder

Condition number	Ratio (mass)	Buffer, pH	Other conditions	Main peaks R.T. & % area	% IGF-I like peak	Observations
25	1:100	50mM Tris, 40mM Glycine, pH8.5	3M Urea, 37°C	17.4 (20%), 18.53 (29%), 19.16 (19%)	20	IGF-I-like peak almost separate peak
26	1:100	50mM Tris, 40mM Glycine, pH8.5	4M Urea, 37°C	17.34 (20%), 18.53 (22%), 19.17 (19%)	20	Trace full-length substrate, IGF-I-like peak is clearly a separate peak. See Figure 4.4.1.4.
27	1:100	50mM Tris, 40mM Glycine, pH9	1M Urea, 37°C	18.12 (53%), 19.1 (19%)	-	Majority of material in one broad peak
28	1:100	50mM Tris, 40mM Glycine, pH9	2M Urea, 37°C	17.45 (17%), 18.46 (40%), 19.17 (18%)	17	Trace full-length substrate, IGF-I-like peak detectable as leading shoulder
29	1:100	50mM Tris, 40mM Glycine, pH9	3M Urea, 37°C	17.3 (20%), 18.4 (33%), 19.1 (15%)	20	Some full-length substrate, IGF-I-like peak almost separate peak
30	1:100	50mM Tris, 40mM	4M Urea,	17.34 (20%), 18.5	20	Significant full-length substrate, IGF-I-like

Condition number	Ratio (mass)	Buffer, pH	Other conditions	Main peaks R.T. & % area	% IGF-I like peak	Observations
		Glycine, pH9	37°C	(22%), 19.2 (18%)		peak almost separate peak
31	1:100	50mM Tris, 40mM Glycine, pH9.5	1M Urea, 37°C	17.2 (5%), 18.25 (55%), 19.25 (17%)	5	Small leading shoulder on main peak group
32	1:100	50mM Tris, 40mM Glycine, pH9.5	2M Urea, 37°C	17.5 (14%), 18.5 (40%), 19.2 (17%)	14	Trace full-length substrate, IGF-I-like peak detectable as leading shoulder
33	1:100	50mM Tris, 40mM Glycine, pH9.5	3M Urea, 37°C	17.48 (18%), 18.5 (32%), 19.2 (17%)	18	Some full-length substrate, IGF-I-like peak almost separate peak
34	1:100	50mM Tris, 40mM Glycine, pH9.5	4M Urea, 37°C	17.38 (17%), 18.5 (18%), 19.3 (12%)	17	Significant full-length substrate, IGF-I-like peak almost separate peak
35	1:100	100mM HEPES, pH8.5	Ambient temp.	18.3 (25%), 18.7 (42%)	-	Leading shoulder on main peak. See Figure 4.4.1.5

Condition number	Ratio (mass)	Buffer, pH	Other conditions	Main peaks R.T. & % area	% IGF-I like peak	Observations
36	1:100	100mM HEPES, pH8.5	4°C, 48hrs	17.6 (35%), 18.8 (31%), 19.03 (28%)	35	IGF-I-like peak almost as a double peak itself, leading shoulder from it visible
37	1:100	50mM Tris, 40mM Glycine, pH8.5	4M Urea, Ambient	17.6 (22%), 18.8 (24%), 19.3 (18%)	22	IGF-I-like peak the best resolved of all conditions examined. See Figure 4.4.1.6
38	1:100	50mM Tris, 40mM Glycine, pH8.5	4M Urea, 4°C, 48hrs	17.5 (4%), 18.8 (6%), 19.3 (11%)	4	Much residual full-length substrate, trace cleavage products
39	1:100	100mM HEPES, pH8.5	10mM Arginine, 37°C	17.3 (6%), 18.5 (90%)	6	Leading shoulder on main peak
40	1:100	100mM HEPES, pH8.5	100mM Arginine, 37°C	17.4 (6%), 18.5 (60%)	6	Small leading shoulder on main peak

Condition number	Ratio (mass)	Buffer, pH	Other conditions	Main peaks R.T. & % area	% IGF-I like peak	Observations
41	1:100	100mM HEPES, pH8.5	500mM Arginine, 37°C	17.47 (5%), 18.73 (52%), 19.5 (16%)	5	Small leading shoulder on main peak
42	1:100	100mM HEPES, pH8.5	10mM Arginine, Ambient	17.3 (8%), 18.77 (45%)	8	Small leading shoulder on main peak
43	1:100	100mM HEPES, pH8.5	100mM Arginine, Ambient	18.75 (45%), 19.5 (23%)	-	Small leading shoulder on main peak
44	1:100	100mM HEPES, pH8.5	500mM Arginine, Ambient	18.79 (43%), 19.5 (18%)	-	Small leading shoulder on main peak

Condition number	Ratio (mass)	Buffer, pH	Other conditions	Main peaks R.T. & % area	% IGF-I like peak	Observations
45	1:100	100mM HEPES, pH8.5	0.05% Tween-20, 37°C	18.07 (22%), 18.3 (23%), 19.03 (20%)	-	Small leading shoulder on main peak group
46	1:100	100mM HEPES, pH8.5	0.5% Tween-20, 37°C	18.07 (15%), 18.3 (17%), 19.03 (18%)	-	Small leading shoulder on main peak group
47	1:100	100mM HEPES, pH8.5	0.05% SDS, 37°C	17.5 (3%), 18.4 (20%), 19.5 (20%)	3	Much residual full-length substrate, very small leading shoulder
48	1:100	100mM HEPES, pH8.5	0.5% SDS, 37°C	-	Trace	Very little cleavage detected, very small leading shoulder
49	1:100	100mM HEPES, pH8.5	5% SDS, 37°C	-		No cleavage detected

Figure 4.4.1.1 – Met-pGH(1-11)VN-PAPM-IGF-I substrate incubated overnight at 37°C in 100 mM HEPES, pH 8.5.

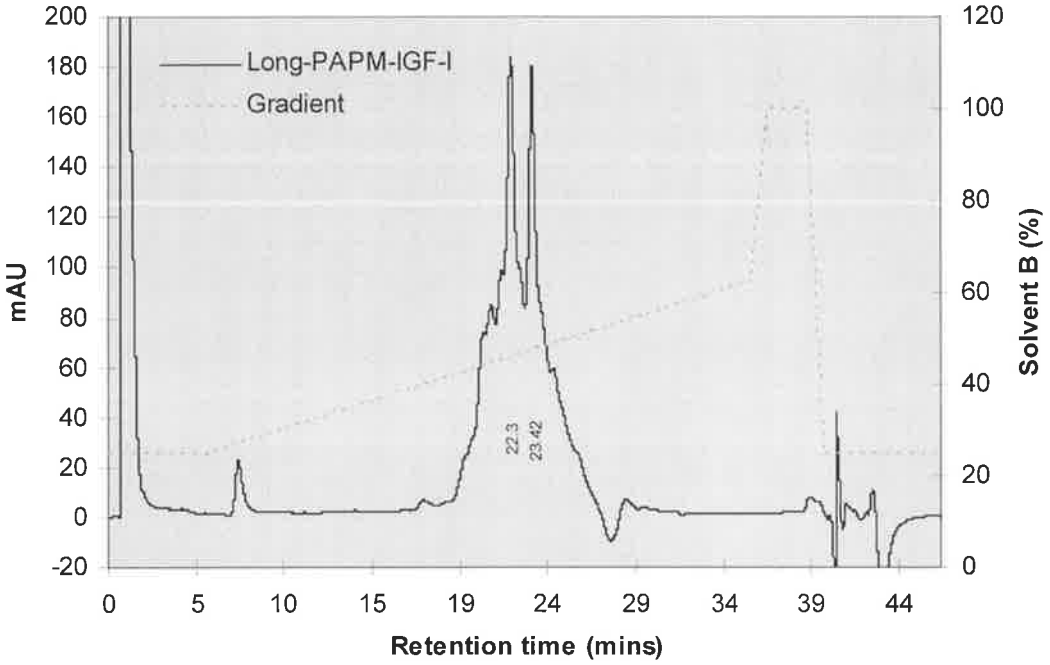


Figure 4.4.1.2 – Cleavage of Met-pGH(1-11)VN-PAPM-IGF-I by PragA9 under reaction condition #4, (100 mM HEPES, pH 8.5, 37°C).

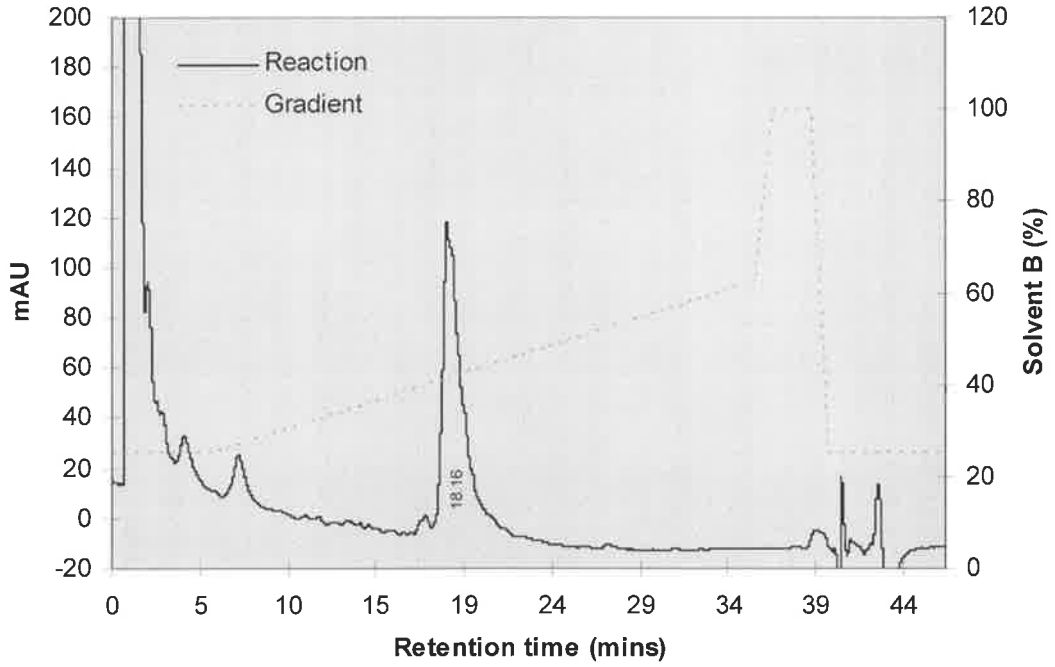


Figure 4.4.1.3 – HPLC trace of authentic IGF-I standard.

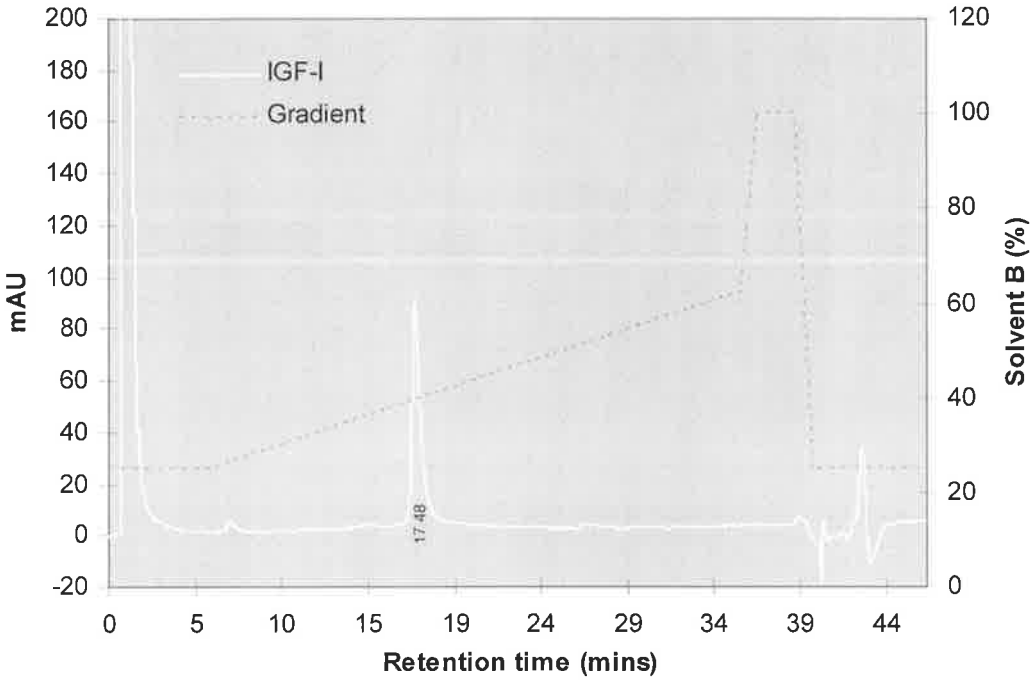


Figure 4.4.1.4 – Cleavage of Met-pGH(1-11)VN-PAPM-IGF-I by PragA9 under reaction condition #26, (4 M urea, pH 8.5, 37°C).

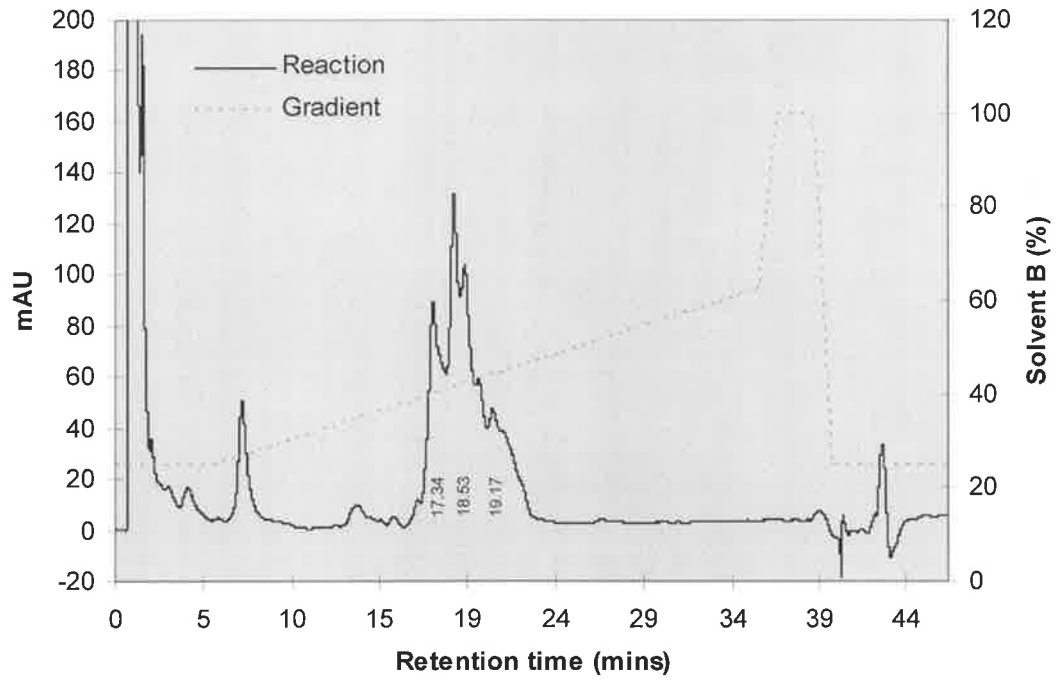


Figure 4.4.1.5 – Cleavage of Met-pGH(1-11)VN-PAPM-IGF-I by PragA9 under reaction condition #35, (100 mM HEPES, pH 8.5, ambient temperature). The peak varies from that of Figure 4.4.1.2 by the presence of a leading shoulder on the main peak, indicated by a downward arrow.

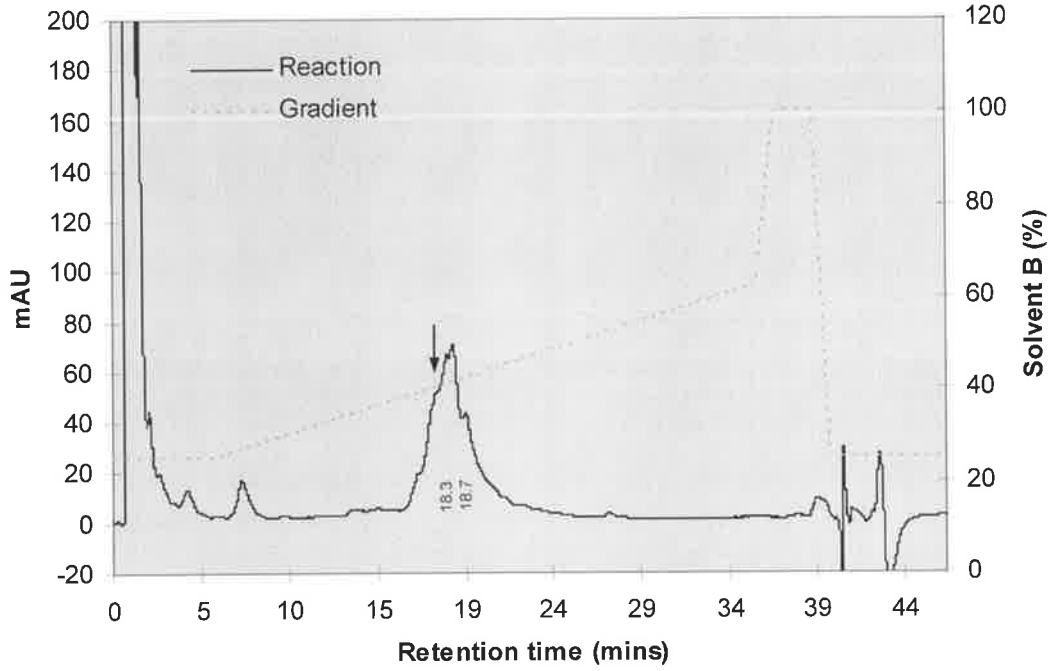


Figure 4.4.1.6 – Cleavage of Met-pGH(1-11)VN-PAPM-IGF-I by PragA9 under reaction condition #37 (4 M urea, pH 8.5, ambient temperature). The peak of IGF-I-like retention time is partially resolved into a discrete peak.

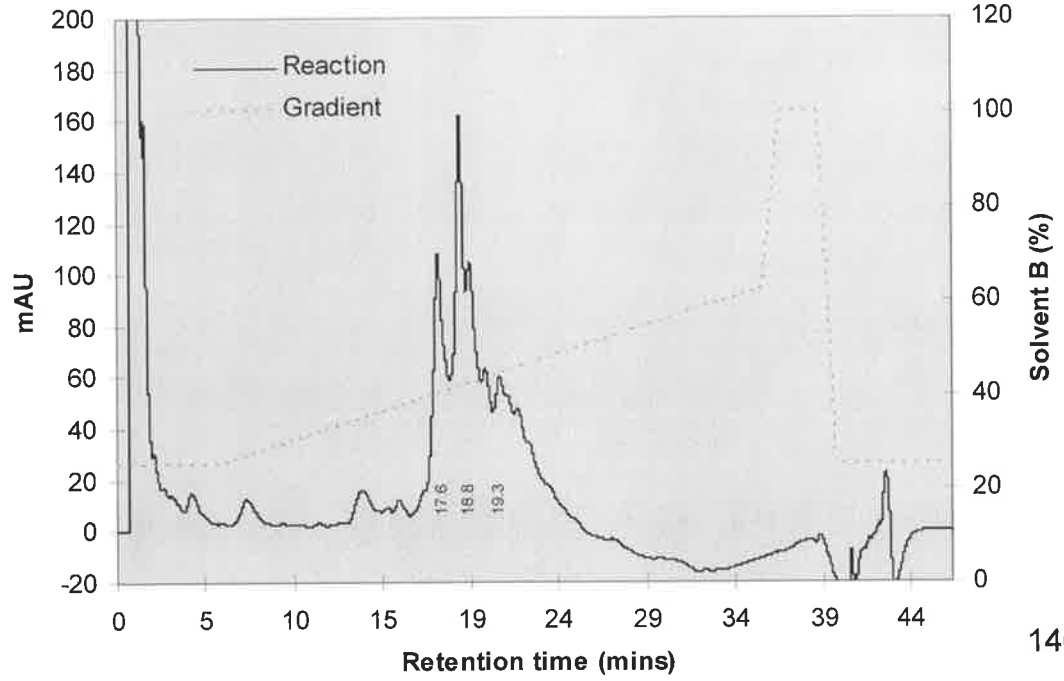


Figure 4.4.1.7 – Overlay of the chromatograms of Figures 4.4.1.3 and 4.4.1.6. The retention times of the leading peak and authentic IGF-I are well matched given the complexity of the cleavage reaction sample and system performance.

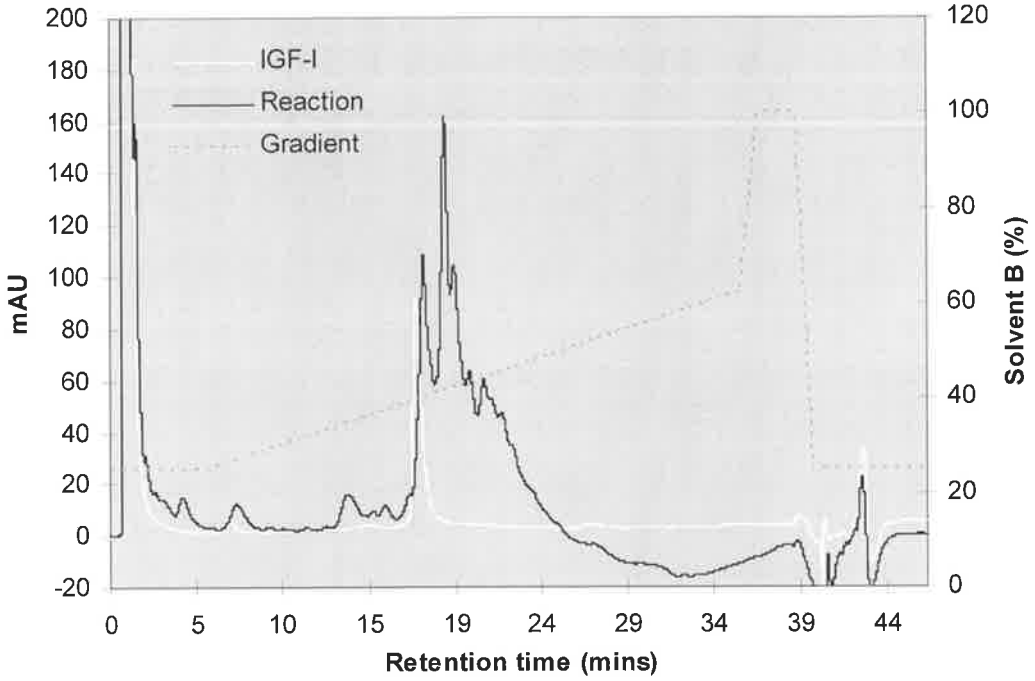
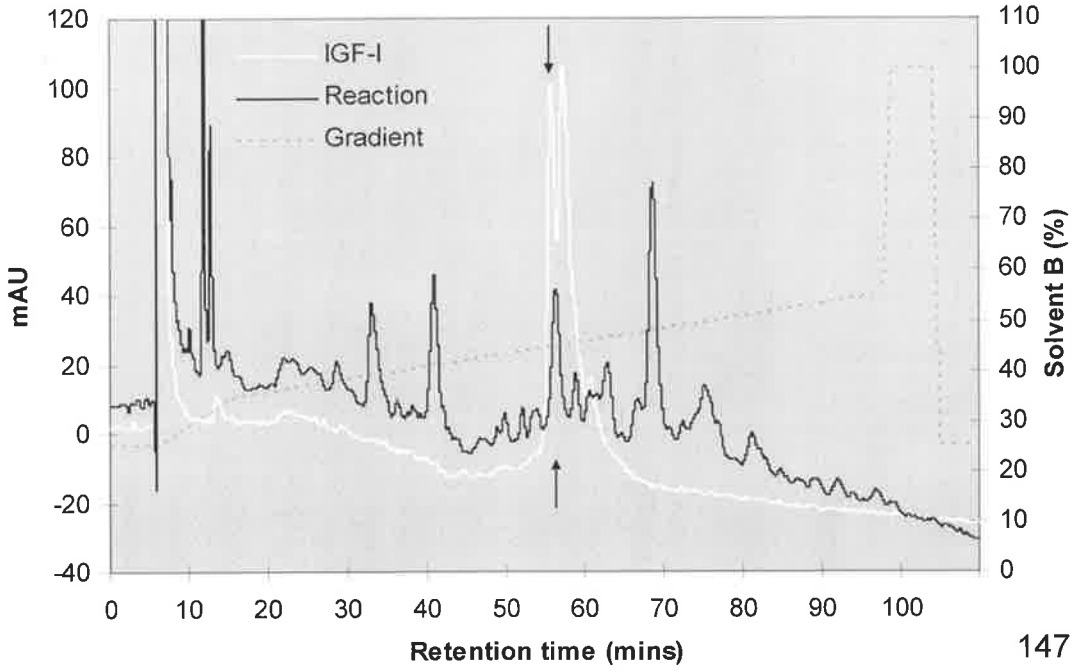


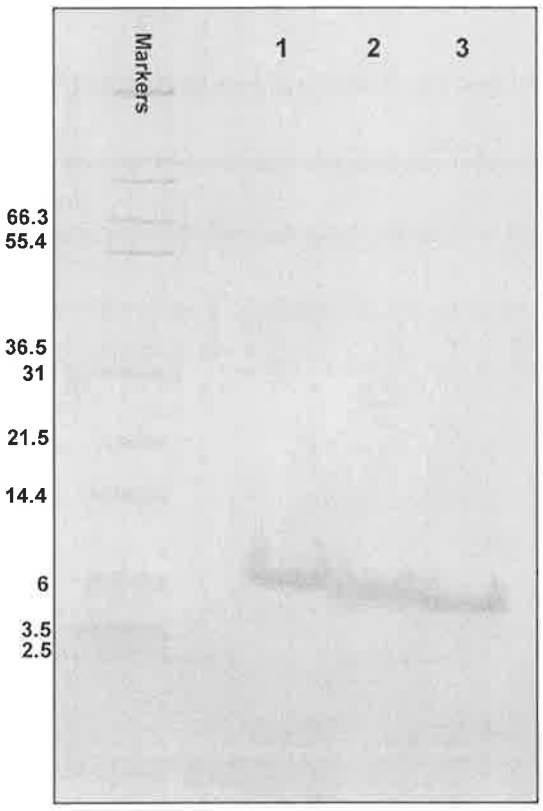
Figure 4.4.1.8 – HPLC of reaction condition #37 products and authentic IGF-I separated over a long gradient. A peak with comparable retention time appears in both traces. The split peak in the standard could correspond to a chemically modified form of IGF-I such as oxidation of Met[59].



4.4.2 SDS-PAGE of Met-pGH(1-11)VN-PAPM-IGF-I cleavage

Analysis of Met-pGH(1-11)VN-PAPM-IGF-I cleavage in 4 M urea, ambient temperature (condition #37 of Table 4.4.1.1) by SDS-PAGE was performed and is shown as Figure 4.4.2.1. The cleavage product appears to have a mass somewhere in between that of the full-length fusion protein and that of authentic IGF-I, but the somewhat diffuse band makes the absolute determination difficult.

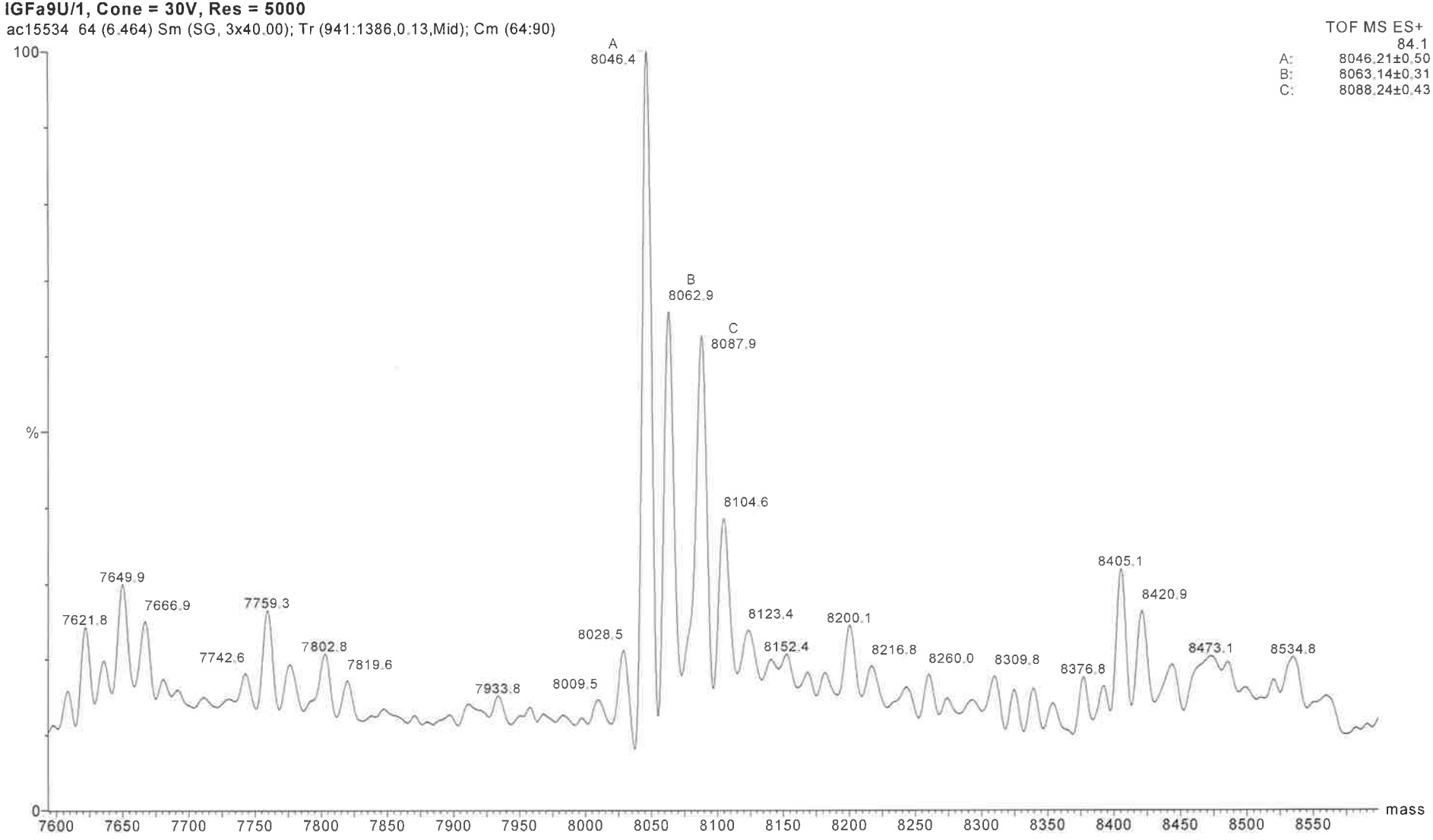
Figure 4.4.2.1 - SDS-PAGE of Met-pGH(1-11)VN-PAPM-IGF-I cleavage and controls. (1) Met-pGH(1-11)VN-PAPM-IGF-I substrate, (2) Cleavage reaction under condition #37, (3) Authentic IGF-I standard.



4.4.3 Mass spectrometry of Met-pGH(1-11)VN-PAPM-IGF-I cleavage

The protein species present in the cleavage of Met-pGH(1-11)VN-PAPM-IGF-I were determined by mass spectrometry and is shown in Figure 4.4.3.1. The major species present have masses 8046, 8063 and 8088 Da. Authentic IGF-I (7648 Da) may be present as represented by the 7650 Da species, but this abundance appears to be far lower than the 20% expected from the HPLC data. This could be due to non-quantitative recovery and resolution with mass spectrometric analysis.

Figure 4.4.3.1 – Mass spectrum of condition #37 (4 M urea, pH 8.5, ambient temperature) cleavage.



4.4.4 Cleavage of S-Carboxymethylated Met-pGH(1-11)VN-PAPM-IGF-I.

Met-pGH(1-11)VN-PAPM-IGF-I was reduced and the free cysteines then blocked by S-carboxymethylation. This was largely quantitative, with a single dominant peak at 39.1 minutes as shown in Figure 4.4.4.1, suggesting that the two folding isomers of the starting product are resolved into a single product as expected. The equivalent form of authentic IGF-I appears as Figure 4.4.4.2, with a single peak at 37 minutes visible. However, as shown in Figure 4.4.4.3, when the S-carboxymethylated Met-pGH(1-11)VN-PAPM-IGF-I is reacted with α-Lytic protease under the optimal cleavage conditions determined above (4 M urea, ambient temperature, pH 8.5), the products generated clearly do not co-elute with S-carboxymethylated authentic IGF-I.

Figure 4.4.4.1 – HPLC trace of S-carboxymethylated Met-pGH(1-11)VN-PAPM-IGF-I (source data not available for clearer plotting).

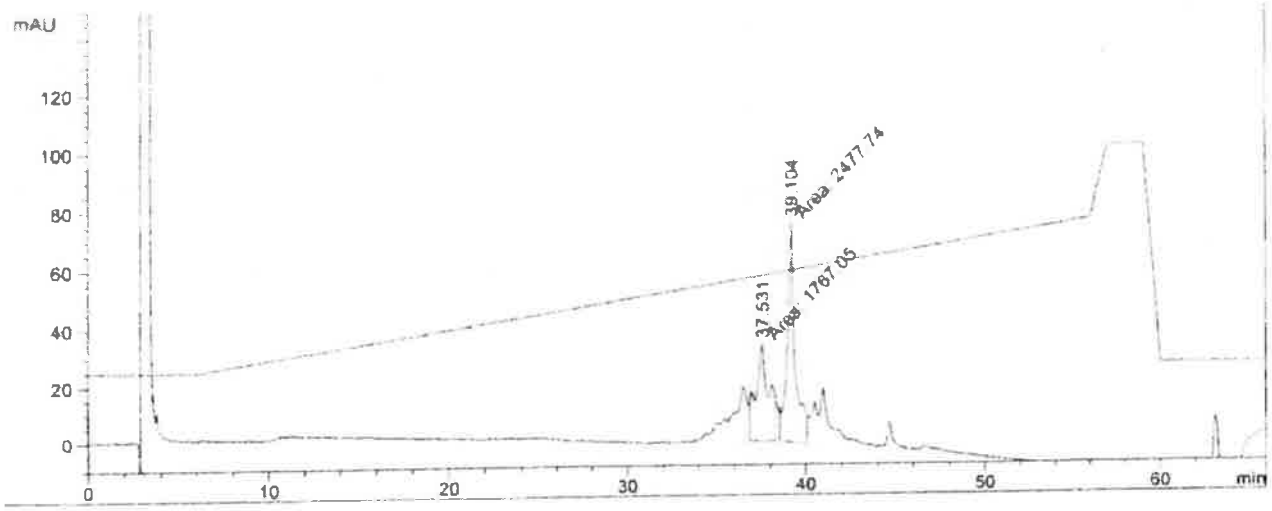


Figure 4.4.4.2 – HPLC trace of S-carboxymethylated authentic IGF-I (source data not available for clearer plotting).

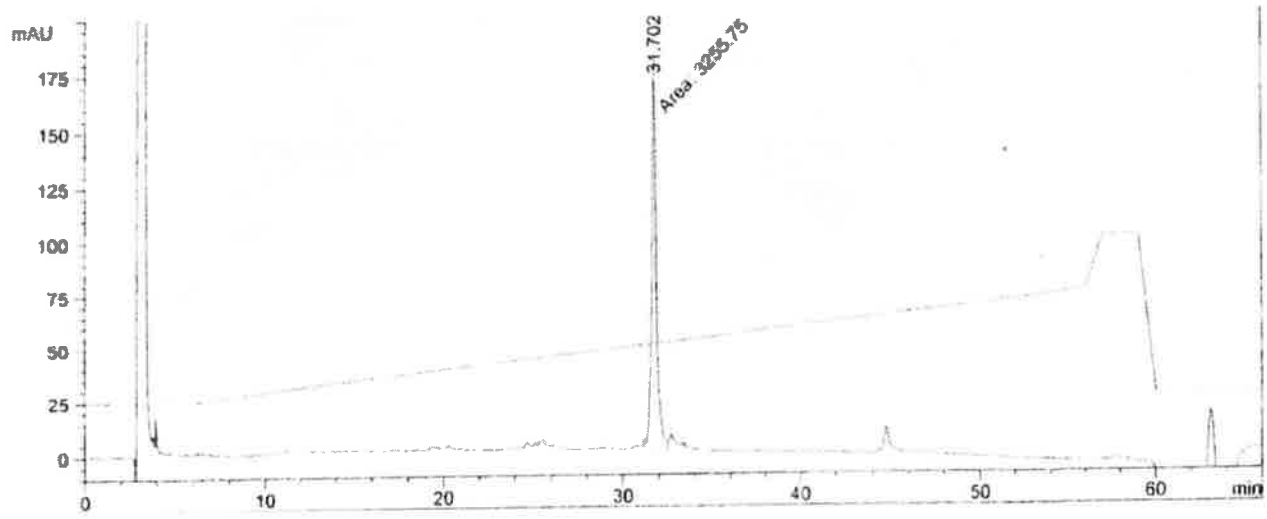
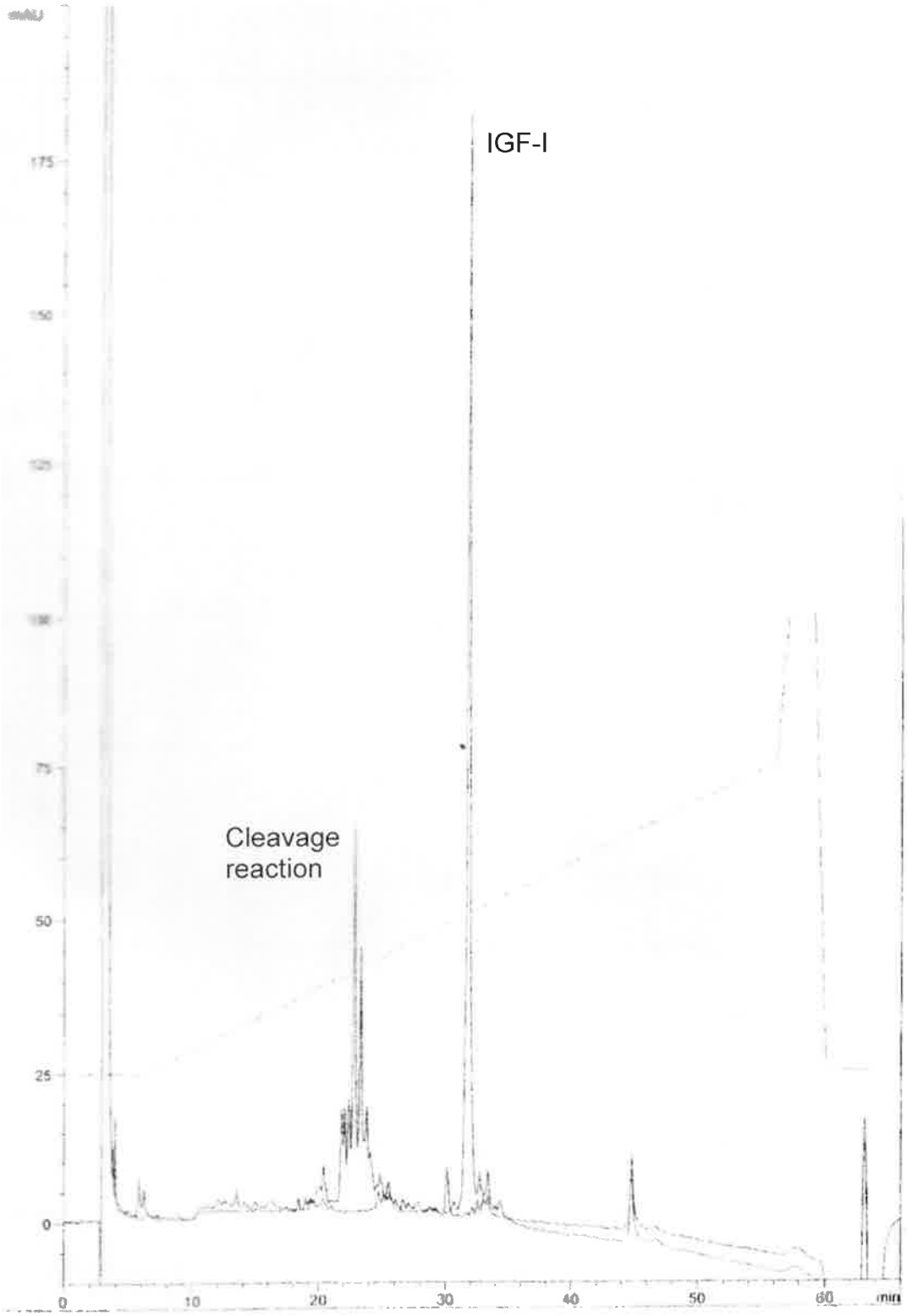


Figure 4.4.4.3 – HPLC trace of S-carboxymethylated Met-pGH(1-11)VN-PAPM-IGF-I cleavage by PragA9 under reaction condition #37 overlaid with that of S-carboxymethylated authentic IGF-I (source data not available for clearer plotting). It is clear that the retention times of the two sets of reactions products are dissimilar.



4.4.5 HPLC analysis of Met-pGH(1-11)VN -IGF-I cleavage reactions

Met-pGH(1-11)VN-IGF-I, bearing no specific cleavage site for PragA9 α-Lytic protease, was incubated with protease under reaction conditions #4 (buffer only, 37°C) and #37 (4 M urea, ambient temperature) of Table 4.4.1.1. The peak group of the full-length substrate, seen in Figure 4.4.5.1, shifts to a series of peaks upon exposure to protease under either reaction condition (Figure 4.4.5.2, condition #4 and 4.4.5.3, condition #37). Although both cleavage reactions generate a large number of product peaks, there does appear to be a trace presence of a peak with the retention time of authentic IGF-I in both reactions (21.5 minutes under the analytical techniques of these experiments, Figure 4.4.5.4).

Figure 4.4.5.1 – HPLC trace of Met-pGH(1-11)VN-IGF-I substrate.

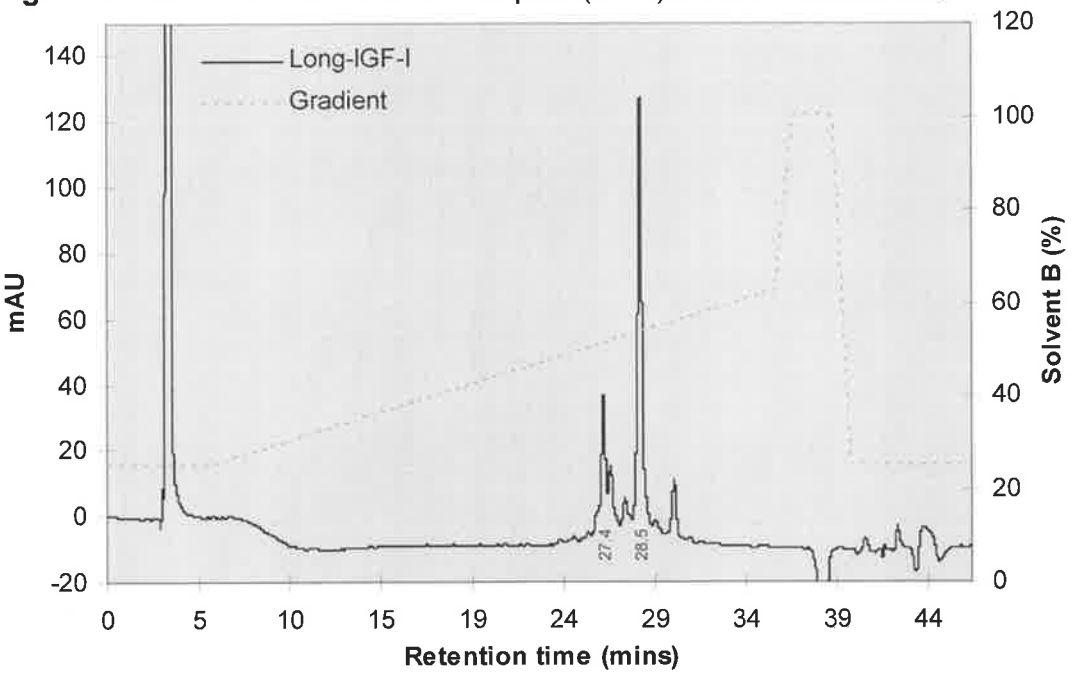


Figure 4.4.5.2 – Met-pGH(1-11)VN-IGF-I cleavage by PragA9 under reaction condition #4. The peak group of the substrate (Figure 4.4.5.1) is shifted forward. A much larger number of peaks are present compared to the substrate. The position of the IGF-I-like peak is indicated with a downward arrow.

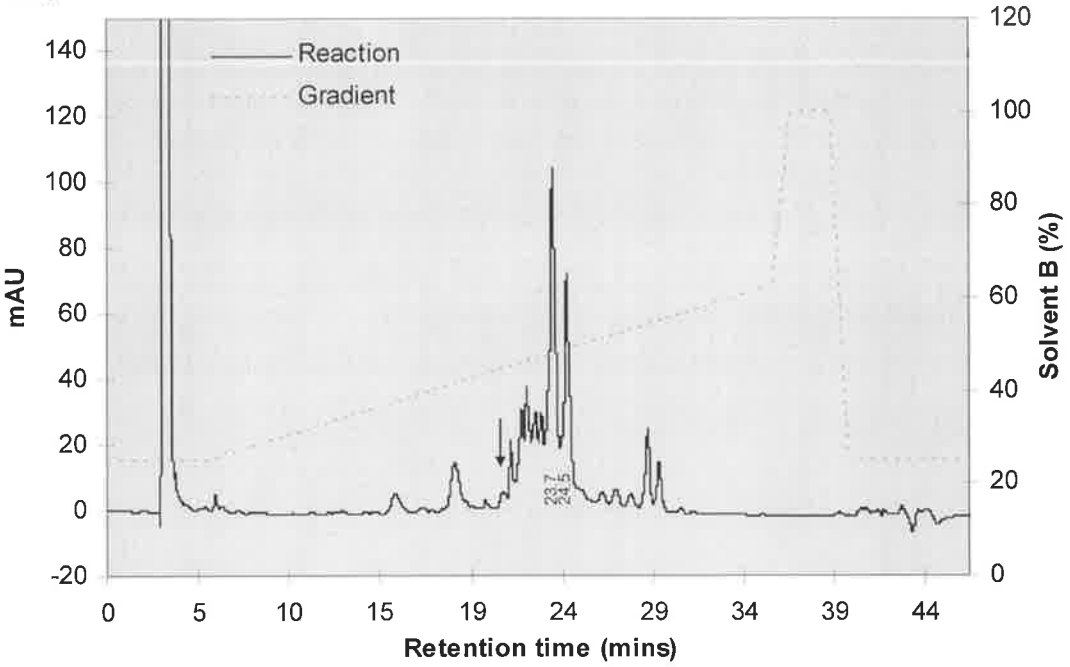


Figure 4.4.5.3 – Met-pGH(1-11)VN-IGF-I cleavage by PragA9 under reaction condition #37. Much residual substrate remains visible at 28.5 minutes. The trace presence of the IGF-I-like peak is indicated with a downward arrow.

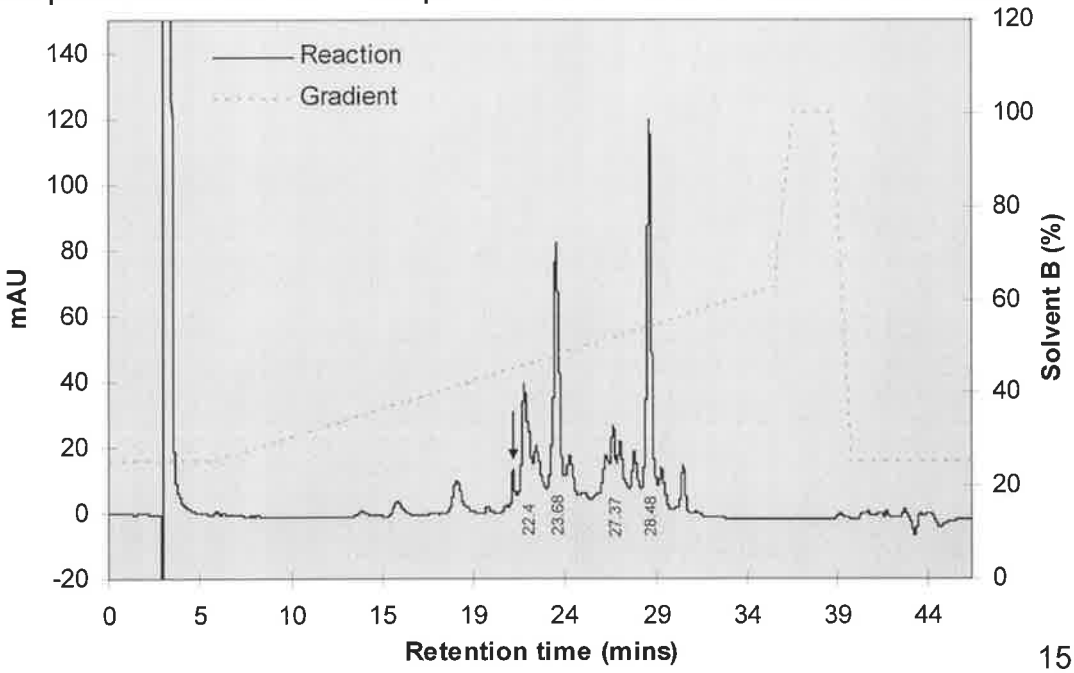
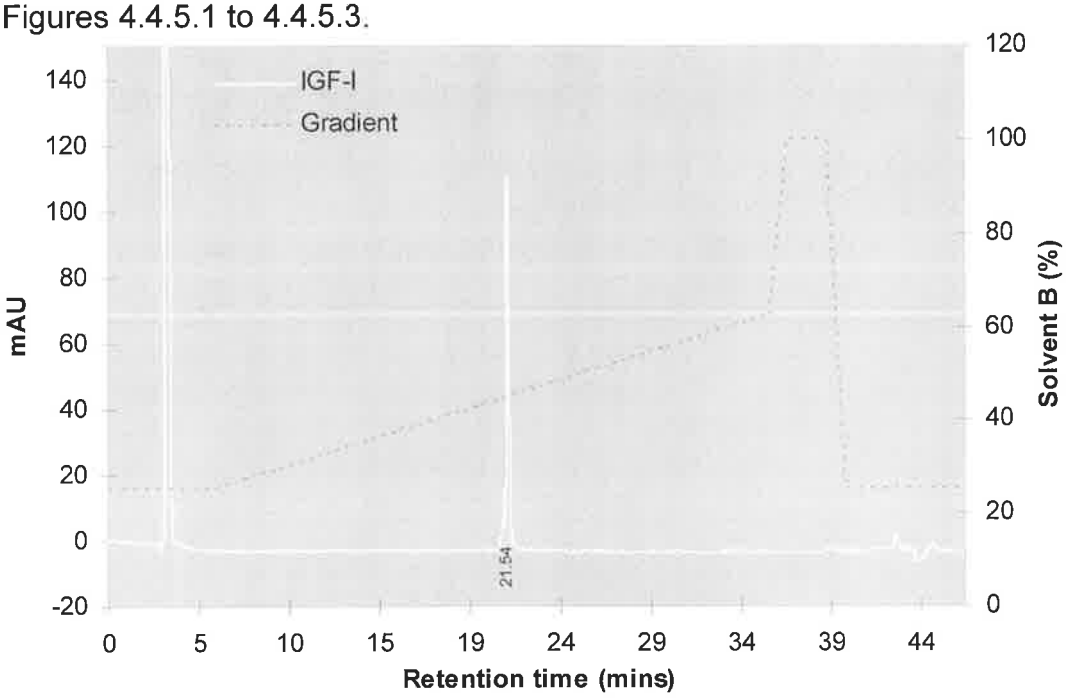


Figure 4.4.5.4 – HPLC trace of authentic IGF-I under the analytical conditions of Figures 4.4.5.1 to 4.4.5.3.



4.5 DISCUSSION

Authentic IGF-I could not be generated from this fusion protein with PragA9 α-Lytic protease by simply employing the same reaction conditions used in the peptide studies of Chapter 3 (approximately a 1:50 enzyme to substrate mass ratio). The two peaks of the starting full-length fusion protein (Figure 4.4.1.1), which most likely correspond to the two disulphide isomers frequently observed in IGF-I refolding events (King *et al.,* 1992; Hober *et al.,* 1992; Miller *et al.,* 1993; Samuelsson *et al.,* 1994; Milner *et al.,* 1995), are merged into one broad product peak. Conditions 4 and 5 of Table 4.4.1.1 and Figure 4.4.1.2 show that the products of such a reaction do not have the correct HPLC retention time to be authentic IGF-I. Preliminary mass spectrometry supported this finding (data not shown), with the only major species present having a mass of 8046 Da. Interestingly, the only possible

cleavage product from the given Met-pGH(1-11)VN-PAPM-IGF-I starting material with such a mass is PAPM-IGF-I at 8045.2 Da. This would represent specific PragA9 cleavage directly before the intended cleavage site, releasing IGF-I with a residual tetrapeptide N-terminal extension of the intended PAPM cleavage motif. Cleavage at this site is not without precedent, as Haggett *et al.,* (1994c) observed cleavage at the identical position in Met-pGH(1-11)VN-FAHY-IGF-II fusion protein. This construct lacked any inserted PragA9 cleavage motif making cleavage precisely at the N-terminus of IGF-II unlikely. Besides cleavage at the Asn residue preceding the FAHY motif, this reaction also produced a Y-IGF-II co-product at twice the molar yield. Interestingly, there were two other available Asn residues present but no cleavage at these was identified.

Decreasing the temperature of the cleavage reaction to ambient altered the appearance of the HPLC profile of the reaction (condition #35, table 4.4.1.1). The leading shoulder of the peak generated in these reactions approached the retention time of authentic IGF-I (Figures 4.4.1.3 and 4.4.1.5). Further reduction in the reaction temperature to 4°C (condition #36), with a commensurate increase in time, enabled resolution of this leading shoulder into a separate double peak, suggesting some heterogeneity. Similarly, the inclusion of urea in the cleavage reactions (conditions #7 to 34) altered the product profile, with a more distinct peak apparent that had a retention time almost identical to that of authentic IGF-I (Figure 4.4.1.4). These two factors also appeared to function synergistically to further improve the abundance and resolution of the putative IGF-I peak, with the maximal yield of this single peak achieved by the inclusion of 4 M urea in the reaction with incubation at

ambient temperature (condition #37) (Figure 4.4.1.6). SDS-PAGE of such apparently successful cleavage reactions (Figure 4.4.2.1) is largely inconclusive, given that the actual mass difference between even full-length fusion protein and authentic IGF-I is less than 2 kDa this is not the preferred analytical approach to apply to this question. Furthermore, even the most successful cleavage conditions are known to still contain the 8046 Da species, as indicated by the continued presence of the 18.5 minute HPLC peak. The resulting potential heterogeneity of somewhat similar mass species present as a broad unresolved band on SDS-PAGE that is impossible to interpret. Preliminary attempts to verify the identity of the putative IGF-I peak by running a more resolving HPLC gradient over a much longer time period than the one routinely employed for the optimal reaction condition scouting seemed to support the identity of the peak as authentic IGF-I (Figure 4.4.1.8).

The emergence of the ability of the protease to liberate authentic IGF-I by modification of the cleavage environment would seem to support the hypothesis that the structure of the fusion protein around the cleavage site was initially responsible for the inability to obtain correctly cleaved product. Since the structural constraints for cleavage appeared to be mitigated by factors that extrinsically affect the strength of hydrophobic interactions in a protein (temperature and chaotrope inclusion), it seems plausible that such interactions are indeed partially responsible for stabilisation of the structural elements that are refractory to protease cleavage. It is interesting that the inclusion of detergents, which will affect the hydrophobic bonding by direct interaction with the protein surface, did not appear to improve the cleavage reaction. Such detergent-substrate interactions could potentially interfere with

protease-substrate binding. More likely however, is that the inclusion of detergents resulted in a considerable loss of protease activity due to denaturation. This is supported by the presence of much residual full-length fusion protein substrate in many of the detergent-bearing reactions.

The manipulation of other factors that may influence protein structure did not enable production of the putative IGF-I peak. For example, inclusion of arginine (conditions #39 to 44), a basic ion-pairing agent, was without benefit to cleavage. As arginine is involved in both simple and complex salt-bridge formation (Musafia *et al.,* 1995), it was thought that inclusion of increasing concentrations of exogenous arginine may competitively interfere with the proposed salt-bridge network of the substrate, thereby perturbing the structure sufficiently to aid protease action.

Similarly ineffective at improving IGF-I generation was variation in the pH of the reactions. The pH 6.5 to 9.5 range investigated (conditions #7 to 34) would be expected to primarily affect histidine side-chains (side-chain pKa 6.0), shifting the equilibrium of the deprotonation of the imide moiety toward completion at increased pH. The S_1 active site of α-Lytic protease contains one non-catalytic histidine residue (His^{213}, Graham *et al.*, 1993). Under investigation was thus any potential effect of the ionisation state of this residue on the specificity of the enzyme. Conversely, as IGF-I lacks any internal histidine residues, the ionisation status of histidine alone was therefore not expected to have any effect on protein structure. However, it was hypothesised that localised charge density microenvironments, that could not be predicted rationally, may be promoted by such manipulations in pH. These could then perturb the protein structure sufficiently to enhance protease

cleavage at the required site. However, the variations in pH investigated did not significantly alter the range of products generated. Instead, there appeared to be a decrease in total protease activity, with residual full-length substrate present in many reactions where the pH fell on either side of the expected maxima of α-Lytic protease mutants (pH 8.0 to 8.5, Graham *et al.* 1993; Graham *et al.* 1994).

Mass spectrometry of the cleavage reaction conditions with the best resolved putative IGF-I peak (condition #37) showed that whilst it did contain detectable authentic IGF-I, it was much less than the approximately 20% apparent in the HPLC chromatogram (Figures 4.4.3.1 and 4.4.1.6, respectively). Given that the putative IGF-I peak comprises in excess of 20% of the total peak area in the HPLC profile of such reactions, it is clear that this peak cannot solely represent totally authentic IGF-I, but likely also includes other minor variants. Such minor variants could include the approximately +17 Da species (7667 Da, see Figure 4.4.3.1), which could co-elute with authentic IGF-I. One possible modification that fits the observed mass difference is oxidation of methionine to methionine sulphoxide (+16 Da). Alternatively, an additional point of hydrolysis of the protein would increase the mass by 18 Da if the resulting product remains held together through the existing disulphide bonds. Additional internal cleavage of the protein could be ruled out if the reducing SDS-PAGE result was clearer (Figure 4.4.2.1); as it would have shown smaller fragments had such cleavage events been occurred.

The most abundant species in the reaction the 8046 Da form and its variants, with major mass species detected at 8046.2 Da, 8063.1 Da and

8088.2 Da. The identities of these latter novel mass species are most likely to be modified forms of the PAPM-IGF-I product. This protein has a predicted mass of 8045.2 Da, providing very close agreement with the detected 8046.2 Da material. The difference in mass between this product and the other two are +16.7 Da and +42 Da for the 8063.1 Da and 8088.2 Da forms respectively. Possibilities for the identity of the 8063.1 Da (+16.7 Da) form include methionine sulphoxide generation or internal hydrolysis, as discussed above. The 8088.2 Da (+42 Da) form could be explained by carbamylation of the protein, which adds 42 mass units. The inclusion of 4 M urea was one of the most significant differences between the reaction condition employed to generate this material and that of the preliminary reaction which produced the 8045 Da form. Carbamylation is a well known phenomenon wherein isocyanic acid (a breakdown product of urea in solution) reacts with primary amines on a protein (Stark *et al.,* 1960). Glycine is frequently included in reactions containing urea to prevent such modifications (Jackowski *et al.,* 1976), but was omitted in this reaction in an attempt to simplify the reaction conditions.

It remained unclear whether the structure that was refractory to efficient cleavage was a phenomenon localised to the cleavage motif/IGF-I junction, or a more global effect derived from the tertiary structure of the whole protein. To address this issue, a sample of Met-pGH(1-11)VN-PAPM-IGF-I was denatured and reduced, the free cysteine thiols were then alkylated by S-carboxymethylation. This protein was then subjected to reaction condition #37 and, upon analysis by HPLC, it was observed to contain no detectable S-carboxymethylated authentic IGF-I (Figures 4.4.4.1 to 4.4.4.3). It is surprising that the cleavage products were grouped at a retention time eluting

much earlier than for the reference material (Figure 4.4.4.3). This contrasts with the comparison of non-carboxymethylated cleavage reactions, wherein the bulk of the cleavage products ran distal to the authentic IGF-I reference (see Figure 4.4.1.7). One explanation for this could be the introduction of one or more favourable cleavage sites with removal of the disulphide linkages and the associated "insulin-like fold". Thus it is possible that the correct product could form and then be degraded, or that new cleavage sites of greater preference may prevent the desired cleavage from occurring. An additional control, that is digestion of carboxymethylated IGF-I was not included, but even this could be problematical to interpret in the absence of the fusion partner sequence.

The inability of PragA9 to successfully cleave at the intended site in this linearised substrate is an encouraging finding. Had cleavage of this protein been successful, it would have indicated that the folded structure of the fusion protein itself was preventing cleavage. This, in turn, may then have necessitated cleavage prior to refolding, an undesirable undertaking as it would mean losing the benefits to refolding efficiency that the Met-pGH(1-11)VN leader provides (Francis *et al.,* 1992; Milner *et al.*,1995; Milner *et al.,* 1999).

The apparent preference of PragA9 to cleave immediately prior to the PAPM motif in the Met-pGH(1-11)VN-PAPM-IGF-I fusion protein construct in a variety of conditions is not solely a function of a favourable sequence immediately upstream. The insertion of the PAPM PragA9 cleavage motif was, logically, between the N-terminus of IGF-I proper and the C-terminal end of the leader sequence. The site of cleavage that gave rise to the 8046 Da

form in the reactions of Met-pGH(1-11)VN-PAPM-IGF-I was thus the C-terminal end of the Met-pGH(1-11)VN leader sequence as discussed earlier. Cleavage reactions of the parental Met-pGH(1-11)VN-IGF-I fusion protein were performed, with this protein lacking an introduced α-Lytic protease cleavage sequence. Thus this substrate had the seemingly favoured Leu-Phe-Val-Asn motif of the Met-pGH(1-11)VN leader immediately before IGF-I N-terminal Gly-Pro-Glu. Interestingly, these reactions also appear to show a small peak at IGF-I-like retention time (Figures 4.4.5.2 and 4.4.5.3). However, a very large number of other product peaks are also observed in such reactions with Met-pGH(1-11)VN-IGF-I, indicating that even if this peak is indeed a result of correct cleavage at the predicted site, it is far from favoured. For example, the cleavage of this substrate under reaction condition #4 produces eight distinct product peaks in the main peak grouping alone (Figure 4.4.5.2). Further characterisation of the cleavage products with Met-pGH(1-11)VN-IGF-I were not carried out. Thus it appears that LFVN is a favoured site when followed by PAPM, but in the absence of this sequence it is at best a minor secondary site.

4.6 CONCLUSION

The initially promising formation of an IGF-I-like peak observed on HPLC was improved by reduction of the reaction temperature and the incorporation of chaotropes. Whilst such reactions were shown to be indeed producing some authentic IGF-I, it was at a relatively low yield. The formation of IGF-I was also not quantitative with the abundance suggested by the magnitude of the IGF-I-like HPLC peak. From the work of this chapter it is

clear that alteration of the cleavage conditions alone is not sufficient to cause a predominance of correctly cleaved IGF-I to be formed by PragA9. The yield obtained from this system is far too low to provide the basis for a successful manufacturing process. Therefore manipulation of the local structure adopted by the leader peptide, or the recognition motif, by mutagenesis of the Met-pGH(1-11)VN-IGF-I fusion protein may be necessary if further gains in the yield of correctly cleaved fusion protein are to be made.

5.1 INTRODUCTION

The ability of the PragA9 mutant of α-Lytic protease to cleave specifically at a site with the N-terminal three residues of IGF-I at the corresponding P_1' to P_3' positions has been demonstrated in Sections 3.4.4 and 3.4.5 This result indicates that the amino acid sequence of the cleavage site itself is not an absolute barrier to successful site-specific proteolysis. However, as described in Section 4.4.3, this success could not be translated into liberation of amounts of authentic IGF-I from the full-length Met-pGH(1-11)VN-PAPM-IGF-I fusion protein that could be successfully scaled industrially. This would seem to indicate that a structural element that is adopted in the full-length protein was interfering with the ability of the protease to act at the required site. This structural inhibition of protease activity could not be alleviated to any significant degree by mild modifications to the cleavage reaction conditions. Empirical alteration of the structural factors preventing protease cleavage appears to be the only remaining solution to the problem at hand. Two main avenues present themselves for this purpose:

1. Reduction in the hydrophobicity of the leader peptide may reduce the tendency of the leader to adopt the "fold-back" structure observed in Met-pGH(1-11)VN-Arg^3-IGF-I and suspected in this Met-pGH(1-11)VN-PAPM-IGF-I construct.
2. Incorporation of a flexible spacer sequence upstream of the cleavage site may improve the effectiveness of any solution conditions aimed at reducing the cleavage site conformational constriction.

The hydrophobicity of the leader peptide, in combination with a flexible turn-like structure in Asn^{-1}-Gly^1-Pro^2-Glu^3 (see below for numbering system),

is possibly responsible for its ability to fold-back onto the hydrophobic core of the IGF-I protein proximal to helix 1. The contact of the leader with the remainder of the IGF-I protein is shown by the exclusion of solvent from the region between the leader sequence and IGF-I proper. A lack of solvent exchange in residues near the C-terminus of helix 2 and the N-terminus of helix 3 of IGF-I (Figure 4.1.1), that is present in authentic IGF-I supports this interaction (Laajoki *et al.,* 2000). However, there is some mobility detected in the solution structure of the leader sequence (Laajoki *et al.,* 2000). This situation suggests that the adopted conformation is the result of ongoing thermodynamic equilibrium rather than permanent burial of the residues during the refolding event. It is hypothesised that mutagenesis of some hydrophobic residues at the N-terminal end of the leader sequence will lead to greater solvent exposure of the leader. By increasing the polarity and flexibility of the leader it should generate a structure that will favour protease cleavage. The effectiveness of these modifications can then be determined for their ability to promote the liberation of authentic IGF-I from the novel fusion protein construct by α-Lytic protease. Furthermore, the influence of reaction conditions on the effectiveness of the mutational changes can also be evaluated.

Given the Proline-richness of the cleavage site, with 3 Prolines in a seven residue span, it is conceivable that the refractory structure created by this site will be quite rigid and bulky. This could have confounded the efforts to improve the cleavage at the site by the modification of the protease reaction conditions as described in Chapter 4, Section 4.4.1. The insertion of a flexible spacer sequence may help to resolve such issues by improving

access to the peptide at the constrained cleavage site. Such spacers have been used effectively when placed both C-terminal (Guan and Dixon, 1991; Rodriguez and Carrasco, 1995) and N-terminal (Hakes and Dixon, 1992; Holowachuk and Ruhoff, 1995) to the cleavage motif. For this application a spacer would obviously have to be N-terminal to the cleavage motif, so as to not leave residual amino acids at the new N-terminus. N-terminal spacers such as Ser-$(Gly)_5$ and $(Ser)_3$-Gly have been used to improve proteolytic processing of fusion proteins by Thrombin (Hakes and Dixon, 1992; Holowachuk and Ruhoff, 1995; respectively). The inclusion of a Ser-$(Gly)_4$-Ala flexible spacer N-terminal to a H64A subtilisin cleavage motif has proven effective in the cleavage of IGF-II from a Met-pGH(1-46)VN fusion partner (Polyak *et al.,* 1997). These Ser-containing, Gly-rich flexible linkers bear a striking resemblance to the upstream amino acid sequence of the long discrete peptides successfully cleaved in Chapter 3. This sequence would therefore make a logical insertion to satisfy the need for a flexible linker, as it is known from Chapter 3 to be successfully tolerated by PragA9.

The potential benefits of the inclusion of a spacer peptide may extend beyond improvements to the flexibility of the leader peptide. The provision of a region of unstructured sequence upstream of the cleavage motif to physically separate the cleavage site from the existing structural elements may well prove beneficial. This is even more pertinent given that there is some evidence for α-Lytic protease enzyme-substrate contact upstream of the P_4 residue, at least up to a P_5 residue. Inclusion of a P_5 residue, although having far less impact than the presence of any other residue in the range P_4 to P_2', has been shown to result in a 25% increase in the second order rate

constant for reactions of the wild-type enzyme against peptide substrates (Bauer *et al.,* 1981). Additionally, the insertion of an amino acid sequence upstream of the Pro-Ala-Pro-Met cleavage motif would effectively remove the interfering cleavage site that was identified in Chapter 4, Section 4.4.3.

By the above two disparate mechanisms the structure of the leader peptide may be altered to promote correct cleavage by PragA9 α-Lytic protease. The hydrophobicity of the leader can be changed by mutagenesis of two hydrophobic residues near the N-terminus to charged amino acids. The residues targeted for alteration are Phe^{-16} and Met^{-13}, using the numbering system of Laajoki *et al.* (1997). In this system the amino acids of Met-pGH(1-11)VN-IGF-I analogues are numbered with respect to the IGF-I portion of the fusion protein, *i.e.* the first three residues of IGF-I proper, Gly-Pro-Glu are Gly^{1}, Pro^{2} and Glu^{3} and the Pro-Ala-Pro-Met cleavage site is numbered Pro^{-4}, Ala^{-3}, Pro^{-2} and Met^{-1}. The N-terminus of the fusion protein is Met^{-17}. For the insertion of a flexible spacer, it is logical to include the upstream sequence from the successfully cleaved long discrete peptide sequence Trp-Gly-Ser-Gly. A schematic representation of the two intended constructs appears as Figures 5.1B and 5.1C.

Figure 5.1 – (A) Primary structure of Met-pGH(1-11)VN-PAPM-IGF-I. (B) Intended primary structure of the leader peptide charge mutant construct. (C) Intended primary structure of the insertional mutant. The deviations from the parental construct are shown in bold.

A
MFPAMPLSSLFVNPAPMGPETLCGAELVDALQFVCGDRGFYFNKPTGYGS
SSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSA

B
ME**E**PA**E**PLSSLFVNPAPMGPETLCGAELVDALQFVCGDRGFYFNKPTGYGS
SSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSA

C
MFPAMPLSSLFVN**WGSG**PAPMGPETLCGAELVDALQFVCGDRGFYFNKPT
GYGSSSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSA

5.2 MATERIALS

5.2.1 General reagents and materials

The following reagents were obtained from BDH chemicals, Kilsyth, Victoria, Australia: HEPES, glycine, Tris, polyoxyethylene (20) sorbitan monolaurate (Tween-20), trifluoroacetic acid (TFA), acetonitrile, urea and all salts and trace metals used in Sections 5.2.5 to 5.2.7. All reagents were of Analytical reagent grade or higher. Analytical grade ethylenediaminetetraacetic acid (EDTA), agarose, bovine serum albumin (RIA grade), dithiothreitol (DTT) and 2-hydroxyethyl disulphide (2-HED) were supplied by Sigma-Aldrich, Castle Hill, NSW, Australia.

5.2.2 *Molecular biology reagents*

Deoxynucleotide Solution Mix, restriction enzymes DnpI, HpaI and HindIII were supplied by New England Biolabs, Ipswich MA, USA. *PfuTurbo* DNA proofreading DNA polymerase, 10x cloned *Pfu* DNA polymerase reaction buffer were supplied by Stratagene, La Jolla, CA, USA. TrackIt™1kb Plus DNA Ladder markers were purchased from Invitrogen, CA, USA. QIAprep Spin Miniprep kit and QIAgen Plasmid Midi kit were purchased from QIAgen GmbH, Hilden, Germany and Wizard SV Gel and PCR Clean-Up kit was supplied by Promega, Madison, Wisconsin, USA.

5.2.3 *Oligonucleotides and Plasmids*

The plasmid CB-A-30 encoding pGH(1-11)VNPAPM-hIGF-I was kindly provided by GroPep, Ltd. Thebarton, SA, Australia. The following oligonucleotide primers were synthesised by Sigma Genosys, Castle Hill, NSW:

WGinsFwd	ATC CAG GTT AAC TGG GGT TCT GGT CCG GCG CCG ATG GGC CCG GAA
F2EM2Efwd	CAC ACA GGA GGT AAT ATA TGG AAC CAG CCG AAC CCT TGT CCA GCC
pGHseq_Fwd	GAG CGG ATA ACA ATT TC
pGHseq_Rev	CAG GCT GAA AAT CTT CTC TCA TCC GC

5.2.4 Bacterial strains

Escherichia coli strains JM109 and JM101 were provided by Novozymes GroPep, Ltd. Thebarton, SA, Australia. *Escherichia coli* XL-10 Gold Ultracompetent cells were purchased from Stratagene, La Jolla, CA, USA.

JM101	F'*traD36 proA*$^+$*B*$^+$ *lacI*^{q} *Δ*(lacZ)M15/*Δ*(*lac-proAB*) *glnV thi*
JM109	F'*traD36 proA*$^+$*B*$^+$ *lacI*^{q} *Δ*(lacZ)M15/*Δ*(*lac-proAB*) *glnV44 e14*$^-$ *gyrA96 recA1 relA1 endA1 thi hsdR17*
XL-10 Gold	Tet^r *D(mcrA)183 D(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F' *proAB lacI*^{q}*ZDM15 Tn10* (Tet^r) Amy Cam^r]

5.2.5 Bacterial culture media

Media were prepared with Milli-Q ultrapure water and autoclaved prior to the addition of sterile-filtered antibiotics. Luria-Bertani (LB broth) liquid medium contains 1.0% tryptone, 0.5% yeast extract, 85 mM NaCl and 100 mg/L ampicillin. LA100 agar plates were prepared by addition of 0.5% (w/v) bacteriological agar to LB broth. MinA minimal medium consists of 30 mM Na_2HPO_4, 9 mM KH_2PO_4, 50 mM NH_4Cl, 10 mM K_2SO_4, 0.5 mM $MgSO_4$, 150 µM Thiamine, 0.075% (w/v) D-glucose, 2 ml/L of trace metal stock solution (Table 5.2.5.1) and 50 mg/L ampicillin. FM11 defined medium contained 38 mM Na_2HPO_4, 38 mM $(NH_4)_2SO_4$, 90 mM KH_2PO_4, 1.0 mM Na_2CO_3, 1.5 mM $MgSO_4$, 1.5 mM tri-sodium citrate, 150 µM Thiamine, 1.5% (w/v) D-glucose, 0.1 ml/L of trace metal stock solution 1 (Table 5.2.5.2), 0.5 ml/L trace metal stock solution 2 (Table 5.2.5.2) and 50 mg/L ampicillin.

Tryptone, yeast extract, bacteriological agar, thiamine and ampicillin were purchased from Sigma-Aldrich, Castle Hill, NSW, Australia. Analytical reagent grade D-glucose was supplied by BDH chemicals, Kilsyth, Victoria, Australia.

Table 5.2.5.1 – MinA medium trace metal stock solution

1.6 mM $CuSO_4$	15 mM $ZnSO_4$
15 mM $MnSO_4$	170 mM tri-sodium citrate
36 mM $FeSO_4$	0.2 M HCl

Table 5.2.5.2 – FM11 medium trace metal stock solutions

FM11 Trace Metals 1	**FM11 Trace Metals 2**
1.5 mM $CuSO_4$	100 mM $FeSO_4$
1.6 mM $CoCl_2$	0.32 mM $ZnSO_4$
1.6 mM Na_2MoO_4	1.0 mM $MnSO_4$
1 mM H_3BO_3	0.1% (v/v) H_2SO_4
1 mM KI	
0.1% (v/) H_2SO_4	

5.2.6 Protein production and analysis

SDS-PAGE and western blot materials; NuPAGE Novex 4-12% Bis-Tris 1 mm minigels, NuPAGE 4x LDS Sample buffer, NuPAGE MES SDS 20x Running Buffer, Mark12™unstained molecular weight standard, SeeBlue™ prestained molecular weight standard and 20x NuPAGE Transfer buffer were supplied by Invitrogen, CA, USA. Hybond-C nitrocellulose,

donkey anti-rabbit horseradish peroxidase conjugate, SP-Big Bead resin and SOURCE 15rpc 80 ml prepacked column were purchased from Amersham Bioscience, Uppsala, Sweden.

Solutions for protein production

Dissolution buffer

8 M Urea

40 mM glycine

1 mM EDTA

100 mM Tris-HCl

40 mM DTT

pH 9.0

IB wash buffer

30 mM NaCl

10 mM KH_2PO_4

0.5 mM $ZnCl_2$

Dilution buffer

2 M Urea

40 mM glycine

1 mM EDTA

100 mM Tris-HCl

pH 9.0

SP-BB binding buffer

8 M Urea

50 mM sodium acetate

pH 4.8

SP-BB elution buffer

8 M Urea

50 mM sodium acetate

1 M NaCl

pH 4.8

5.2.7 *Protein synthesis assay*

All cell culture media were prepared with Milli-Q ultrapure water and filter sterilised prior to use. Dulbecco's modified Eagle's medium (DMEM) contains 1.35% (w/v) DMEM high glucose powder (Sigma-Aldrich, Castle Hill, NSW, Australia), 45 mM $NaHCO_3$, 20 mM HEPES and 1% (v/v) penicillin-streptomycin-fungizone. DMEM/FBS was made by addition of 10% foetal bovine serum to DMEM. Hank's balanced salt solution (HBSS) consists of 136 mM NaCl, 0.35 mM Na_2HPO_4, 0.35 mM KH_2PO_4, 5.5 mM KCl, 0.4 mM $MgSO_4$, 1.2 mM $CaCl_2$, 0.1% (w/v) D-glucose and 0.001% (w/v) phenol red. Resuspension buffer contained 0.5 M NaOH and 0.1% (v/v) Tween-20.

Ready-Solv HP™ liquid scintillant was purchased from Beckman Coulter, Gladesville, NSW, Australia. Nunclon 75 polystyrene tissue culture flasks and 24-well polystyrene cell culture dishes were obtained from Nalge Nunc, Rochester, NY, USA.

5.2.8 *General solutions*

Phosphate Buffered Saline with Tween (PBS-T)

136 mM NaCl

11.5 mM KH_2PO_4

80 mM Na_2HPO_4

25 mM KCl

0.05% (w/v) Tween-20

HEPES Cleavage buffer

100 mM HEPES

pH 8.5

Urea/Tris cleavage buffer

2 M Urea

200 mM Tris-HCl

pH 8.5

5.3 METHODS

5.3.1 Generation of the Met-pGH(1-11)VN-[F^{16}E/M^{13}E]-PAPM-IGF-I expression construct

A derivation of Quickchange PCR (Stratagene, La Jolla, CA, USA; Wang and Malcolm, 1999) was used to perform the site-directed mutagenesis of Phe^{-16} and Met^{-13} each to Glu. 50 ng CB-A-30 vector DNA, 90 pmol of primer F2EM2Efwd, 0.8 mM dNTPs, 2.5 U of *PfuTurbo* DNA polymerase and 5 µl of 10x cloned *Pfu* reaction buffer in a 50 µl reaction buffer, with the volume balance made up with sterile Milli-Q water. This was reacted in a heated lid thermocycler (Perkin Elmer, Waltham, MA, USA) for 1 minute at 95°C followed by 35 cycles of [95°C for 30 seconds, 52°C for 30 seconds, 68°C for 8 minutes] and completed with 68°C for 12 minutes. The reactions were analysed by agarose gel electrophoresis (see Section 5.3.3). Parental DNA was digested with DpnI, which recognises only methylated DNA, according to the manufacturer's directions.

Product-bearing DpnI-treated PCR reactions were used to transform XL-10 Gold Ultracompetent cells according to the manufacturer's directions. Putative positive transformants were labelled pGP-E2E5-XX/xl10, where XX was the incremental number of the clone in the order that they were sampled. Each such clone was used to inoculate 5 ml LB100 which was grown overnight at 37°C with shaking at 200 rpm (Ratek Instruments, Victoria, Australia). From these overnight cultures, plasmid DNA was prepared by miniprep and glycerol stocks were established by addition of sterile glycerol to 40% (v/v) and stored at -80°C.

An aliquot (5 µl) of purified pGP-E2E5-04/xl10 plasmid was used to transform 100 µl of $CaCl_2$ precipitation competent *E. coli* JM109 by heatshock, with positive transformants then designated pGP-E2E5-04/109 to reflect the new host strain. Glycerol stocks were established as above and stored at -80°C. To confirm the maintenance of the insertion, 4.5 µl of purified plasmid DNA was digested with HpaI/HindIII according to the manufacturer's instructions.

To maintain consistency with established IGF-I expressions (King *et al.,* 1992; Francis *et al.,* 1992), pGP-E2E5-04/109 plasmid DNA was purified and 10 µl used to transform 66 µl of *E. coli* JM101 cells. Positive transformants were then designated pGP-E2E5-04/101 to reflect this host strain. Glycerol stocks were established as above and stored at -80°C.

5.3.2 *Generation of the Met-pGH(1-11)VN-WGSGPAPM-IGF-I expression construct*

Insertion of the Trp-Gly-Ser-Gly sequence upstream of the PragA9 cleavage motif was performed by PCR of the PAPM-IGF-I coding sequence of CB-A-30. The forward primer WGSGinsFwd bore homology in its 3' half to the coding sequence of vector immediately 3' to the HpaI restriction site and had a 5' overhang bearing the *E. coli* codon optimised sequence encoding Trp-Gly-Ser-Gly upstream of this position, along with replacement of the HpaI site 5' to this. A schematic representation of this system appears as Figure 5.3.2.1. The reverse primer pGHseq_Rev was homologous to vector CB-A-30 18 bp downstream of the HindIII site that is immediately 3' to the stop codon of the IGF-I coding sequence.

The PCR reaction was conducted by mixing 70 ng CB-A-30 vector DNA, 30 pmol of each primer (WGSGinsFwd and pGPseq_Rev), 0.8 mM dNTPs, 2.5 U of *PfuTurbo* DNA polymerase and 5 µl of 10x cloned *Pfu* reaction buffer in a 50 µl reaction buffer, with the volume balance made up with sterile Milli-Q water. The most successful reaction was achieved by incubation in a heated lid thermocycler (Perkin Elmer, Waltham, MA, USA) for 2 minutes at 95°C followed by 35 cycles of [95°C for 30 seconds, 62°C for 30 seconds, 72°C for 2 minutes] and completed with 72°C for 10 minutes. The reactions were analysed by agarose gel electrophoresis (see Section 5.3.3).

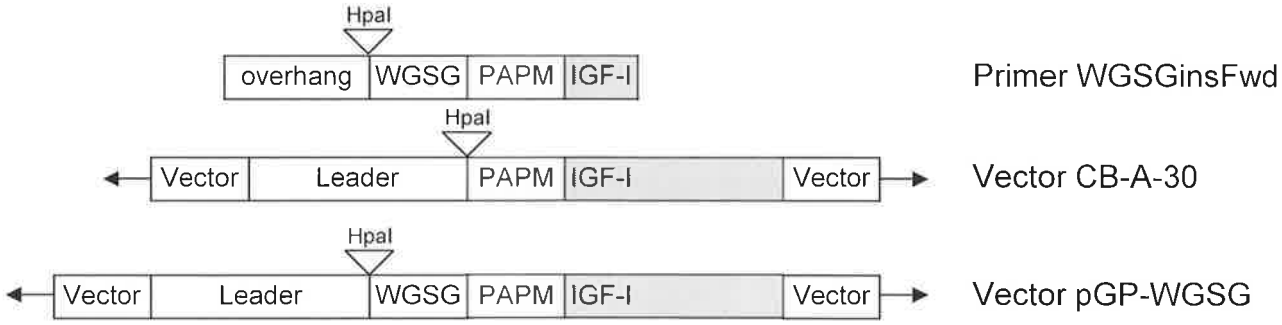
An aliquot (13.5 µl) of the PCR product and CB-A-30 parental vector were digested with HpaI and HindIII according to the manufacturer's instructions and separated on a large agarose gel. The bands corresponding to the insertion fragment and the linearised vector were excised and purified from the gel using a gel purification extraction kit.

An aliquot (30 ng) of the purified digested insert was ligated with 100 ng of the purified digested vector with T4 DNA ligase according to the manufacturer's directions. A 25 µl ligation reaction was used to transform 100 µl of $CaCl_2$ precipitation competent *E. coli* JM109 by heatshock.

Putative positive transformants were designated pGP-WGSG-XX/109 where XX was the incremental number of the clone in the order that they were sampled. Each such clone was used to inoculate 5 ml LB100 which was grown overnight at 37°C with shaking at 200 rpm (Ratek Instruments, Victoria, Australia). From these overnight cultures, plasmid DNA was prepared by miniprep and glycerol stocks were established by addition of sterile glycerol to 40% (v/v) and stored at -80°C. Plasmid DNA was redigested with HpaI/HindIII

as above to confirm insertion. To maintain consistency with established IGF-I expressions, 10 µl of pGP-WGSG-02 plasmid DNA was transformed into *E. coli* JM101 cells as described above. Positive transformants were then designated pGP-WGSG-02/101 to reflect the new host strain. Glycerol stocks were established as above and stored at -80°C.

Figure 5.3.2.1 – Schematic representation of WGSG insertion cloning strategy



5.3.3 *Agarose gel electrophoresis*

DNA separations were performed on 1% agarose gels. Gels were prepared and run in TE buffer (10 mM Tris, 1 mM EDTA, pH 8) at 100 V and visualised by ethidium bromide staining and ultraviolet illumination.

5.3.4 *Plasmid preparation and DNA sequencing*

Plasmid DNA of pGP-E2E5-04/109 and pGP-WGSG-02/109 was prepared from 50 ml of overnight culture in LB100 using a midiprep kit according to the manufacturer's instructions. Sequencing reactions and analysis from each plasmid using primers pGPseq_Fwd or pGPseq_Rev was

conducted by Flinders University DNA Sequencing Facility, Department of Haematology & Genetic Pathology (Flinders University, Bedford Park, SA, Australia).

5.3.5 Trial induction

Recombinant protein expression was trialled by culture of the construct bearing host bacteria (pGP-WGSG-02/101 and pGP-E2E5-04/101) in 25 ml MinA medium in 200 ml Erlenmeyer flasks. Duplicate flasks of each construct were inoculated with 100 µl of overnight culture in MinA medium and grown at 37°C with 200 rpm shaking (Ratek Instruments, Victoria, Australia). Log phase growth was typically achieved 6 hours post-inoculation, at which time one culture of each construct was induced by the addition of 250 µl of 0.1 M IPTG (final concentration 1 mM) and grown for a further 4 hours. 500 µl samples were withdrawn hourly and stored at -20°C for later analysis.

5.3.6 Fusion protein production

Production batch fermentations of fusion proteins took place in 200 ml FM11 medium in 2 L baffled Erlenmeyer flasks. Flasks were inoculated with a 3.5 ml overnight culture in FM11 medium, and grown at 37°C with 200 rpm shaking (Ratek Instruments, Victoria, Australia). Log phase growth was typically achieved 4 to 5 hours post-inoculation, at which time the cultures were induced by the addition of 200 µl of 0.1 M IPTG (final concentration 0.1 mM) and grown for a further 5 hours. The cultures were pooled and frozen at -20°C until needed.

5.3.7 *Harvesting and washing of Inclusion bodies*

Batches of bacterial ferment from each construct were thawed at room temperature prior to cell lysis by mechanical homogenisation. Homogenisation was performed by 10 passes through a high-pressure homogeniser at 800 bar (Rannie MINI-LAB Laboratory Homogenizer, type 8.3H, Albertslund, Denmark). Homogenate was chilled to 4°C between each pass. Inclusion bodies were separated from cellular debris by centrifugation at 18,000 g for 25 min at 4°C (Beckman Coulter, Gladesville, NSW, Australia). Pellets were frozen at -20°C until needed. The homogenate pellets containing the inclusion bodies were thawed at room temperature and pooled. The inclusion bodies were washed with four changes of 25 ml IB wash buffer.

5.3.8 *Inclusion body dissolution and oxidative refolding of novel Met-pGH(1-11)VN-PAPM-IGF-I constructs*

The dissolution and oxidative refolding of the Met-pGH(1-11)VN-PAPM-IGF-I constructs was achieved largely as described in Francis *et al.* (1992). Briefly, the insoluble protein in inclusion bodies was denatured and reduced in dissolution buffer at an inclusion body concentration of 10% (w/v). Dissolution was allowed to continue with stirring for 60 minutes at room temperature. Oxidative refolding of the recombinant protein was initiated by the dilution of the dissolution reaction 1:100 with dilution buffer, to bring the DTT concentration down to 0.4 mM and the total protein concentration to less than 0.15 mg/ml. 2-HED was added to a final concentration of 1 mM and the pH adjusted to 9.0. The refolding reaction proceeded for 90 minutes with stirring at room temperature before termination by acidification with 18 ml of

10 M HCl per litre of refold reaction. Both the dissolution and refolding reactions were monitored by rp-HPLC (see Section 5.3.14)

5.3.9 *Fusion protein purification*

Purification of the two folding isomers of Met-pGH(1-11)VN-WGSGPAPM-IGF-I and Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I from refolding reactions was effected by a two stage chromatographic process. In the first stage, the proteins were captured on SP-Big Beads. A column (25 ml) of packed bed resin (XK26 housing, Amersham Biosciences, Uppsala, Sweden) was equilibrated with 5 column volumes of binding buffer at 113 cm/hr on an AKTA Explorer 100 FPLC system (Amersham Biosciences, Uppsala, Sweden). The refold reaction was loaded and the bound protein eluted with 5 column volumes of elution buffer at 113 cm/hr.

In the second purification event, each isomer was separated from contaminating protein and isolated to near homogeneity. Separation was effected by binding to a prepacked 80 ml SOURCE 15RPC column (Amersham Biosciences, Uppsala, Sweden). The column was equilibrated with 5 column volumes of HPLC solvent A (aqueous 0.1% TFA) at 20 ml/min before loading the output of the prior SP-Big Bead column. Elution was effected by a linear gradient of 32% – 48% acetonitrile in aqueous 0.1% TFA over 10 column volumes. Fractions were analysed by rp-HPLC (see Section 5.3.14) and those corresponding to each isomer were pooled, lyophilised and stored at 4°C until needed

5.3.10 Fusion protein cleavage

Purified isomers of Met-pGH(1-11)VN-WGSGPAPM-IGF-I and Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I were resuspended at a concentration of 1 mg/ml in 10 mM HCl. Preliminary cleavage reactions were performed by incubating 15 µg of fusion protein at 50 µM overnight at room temperature (20 – 22°C) with PragA9 α-Lytic protease at an 1:100 enzyme:substrate mass ratio in either HEPES cleavage buffer or Urea/Tris cleavage buffer. Reactions were terminated by acidification with 1.5% TFA.

Met-pGH(1-11)VN-WGSGPAPM-IGF-I reaction conditions were varied as defined in Tables 5.4.10.1 and 5.4.10.3. Unless otherwise stated, all reactions were conducted overnight at room temperature (20 – 22°C). Large-scale cleavage reactions of Met-pGH(1-11)VN-WGSGPAPM-IGF-I were conducted in Urea/Tris cleavage buffer at a ratio of 1:30 overnight at room temperature. Reactions were terminated by acidification with 1.5% TFA.

5.3.11 Preparative cleavage product separation

The cleavage products generated from Met-pGH(1-11)VN-WGSGPAPM-IGF-I by PragA9 were separated on a C4 reverse phase radial pressure column (Waters Prep Pak 25 x 100 mm; Waters, Milford, MA, USA) on an AKTA Explorer 100 FPLC system (Amersham Biosciences, Uppsala, Sweden). Separation was by linear gradient from 28% to 52% acetonitrile in aqueous 0.1% TFA over 40 minutes at 10 ml/min. Fractions were analysed by rp-HPLC (see Section 5.3.14) and those corresponding to each product were pooled, lyophilised and stored at 4°C until needed.

5.3.12 L6 Myoblast protein synthesis assay

The biological potency of proteolytically generated IGF-I, along with cleavage intermediates, was compared to that of commercially available IGF-I for their ability to stimulate protein production in rat L6 myoblast cells. The conduct of the assay was largely as described in Francis *et al.* (1986). Specifically, L6 myoblast cell stocks were grown to confluent monolayers in 20 ml DMEM/FBS in T75 bottles at 37°C in an atmosphere containing 5% CO_2. Cell stocks were quantified and used to seed 24-well culture plates with 1.7 x 10^5 cells in 1 ml DMEM/FBS. Cells were grown for a further 48 hours until confluent. The medium was aspirated and replaced with 1 ml fresh serum-free DMEM and incubated for a further 2 hours. The serum-free medium was then replaced with 900 µl fresh serum-free DMEM containing ^{3}H-Leucine at 1 µCi/ml. A range of growth factor concentrations was added in 100µl sterile PBS, with all assays performed in triplicate. Growth factors and intermediate cleavage products were resuspended in 10 mM HCl prior to dilution in PBS. Sterile PBS (100 µl) containing no added growth factors served as a negative control and 100 µl of FBS was used as a positive control. Cell stimulation and label incorporation proceeded for 18 hours at 37°C in an atmosphere containing 5% CO_2. The medium was aspirated and the wells were washed twice with 1 ml of cold Hanks balanced salt solution (HBSS) then twice with 1 ml cold 5% trichloroacetic acid (TCA). Proteins were solubilised by the addition of 1ml resuspension buffer and incubated for 1 hour with shaking at room temperature. An aliquot (100 µl) from each well was added to duplicate polypropylene scintillation vials (POLY-Q, Beckman Coulter, Gladesville, NSW, Australia) and mixed with 1.9 ml liquid scintillation

cocktail. Total β-emissions from the vials was measured for 3 minutes each in a β counter (Beckman Coulter, Gladesville, NSW, Australia).

The biological potency of the growth factors was expressed as the percentage increase in ^{3}H-Leucine incorporation compared to the baseline negative control. Dose-response curves were plotted using Tablecurve 2D software (Version 5, SPSS software, Chicago, IL, USA) to calculate ED_{50} values for the peptides.

5.3.13 SDS-PAGE and Western blot analysis

Protein samples were analysed on 4-12% NuPAGE gels in MES buffer under reducing conditions according to the manufacturer's instructions. Protein bands were visualised as in Section 2.3.6.

Western blot transfers from the above gels onto Hybond-C nitrocellulose were conducted using an Invitrogen submerged transfer apparatus (Invitrogen, CA, USA) according to the manufacturer's instructions. Membranes were blocked overnight in 2% BSA in PBS-T at 4°C. Then, 5 µl of anti-hIGF-I (Novozymes GroPep Ltd., Thebarton, SA, Australia) was added and the membrane incubated for a further 2 hours at room temperature with shaking. The membrane was washed with 4 changes of fresh PBS-T for 10 minutes each. The membrane was then probed with a secondary antibody-HRP conjugate at a 1:2000 dilution of stock in PBS-T and incubated for 2 hours at room temperature with shaking. The membrane was again washed as above and developed with an Enhanced Chemi-Luminescence kit (Amersham Bioscience, Uppsala, Sweden) and used to expose X-ray film for (typically) 5–30 seconds.

5.3.14 Analysis of cleavage reactions by HPLC

Dissolution, refold and cleavage reactions were analysed by reversed phase HPLC (rp-HPLC) on an Agilent 1100 HPLC system (Agilent technologies, California, USA) using a C4 column (Vydac 214TP54, 4.6 mm ID x 250 mm length. Grace Vydac, California, USA). Separation was by linear gradient of 20 – 50% acetonitrile in aqueous 0.1% TFA over 30 minutes at 1.0 ml/min, with monitoring at 215 nm.

5.3.15 Mass Spectrometry of proteins.

Proteins to be analysed by mass spectrometry were desalted by rp-HPLC and dried by vacuum centrifugation. Mass spectrometry was performed by I. Milne, Hanson Institute Protein Core Facility, Division of Human Immunology, Institute of Medical and Veterinary Science, Adelaide, SA, Australia.

5.4 RESULTS

5.4.1 Generation of Met-pGH(1-11)VN-[$F^{16}E/M^{13}E$]-PAPM-IGF-I expression plasmid

Quickchange PCR produced a DpnI resistant plasmid, visible in Figure 5.4.1.1. Although the expected size of the plasmid is 3.1 kb (Francis *et al.*, 1992), the apparent smaller size detected for this product could be readily explained by supercoiling of the plasmid. Indeed, after a diagnostic restriction digest of the plasmid intended to excise the IGF-I coding sequence, shown as

Figure 5.4.1.2, the 243 bp IGF-I coding sequence is visible and the now linearised vector has assumed its correct position of approximately 2.9 kb. The undigested control (lane 2) of this reaction echoes the supercoiled plasmid of Figure 5.4.1.1. DNA sequencing of pGP-E2E5-04 plasmid confirmed successful mutagenesis.

Figure 5.4.1.1 – PCR generated mutant plasmid encoding Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I, post DpnI treatment.

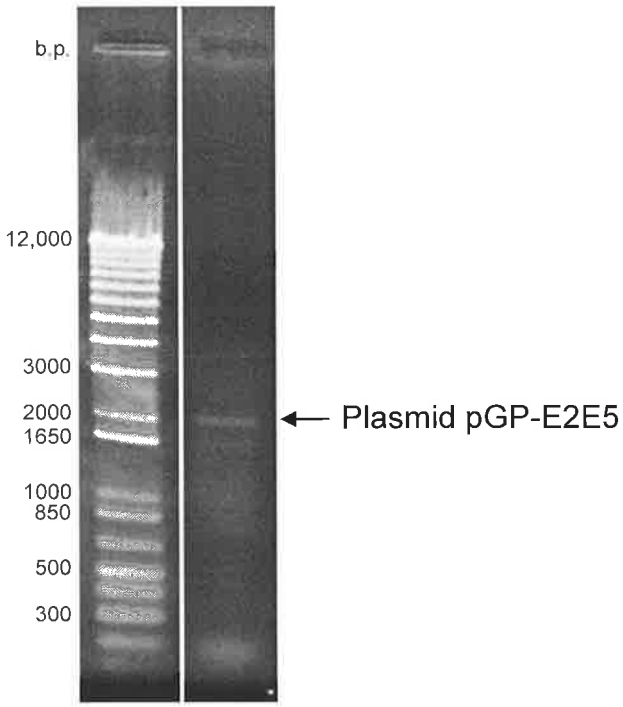
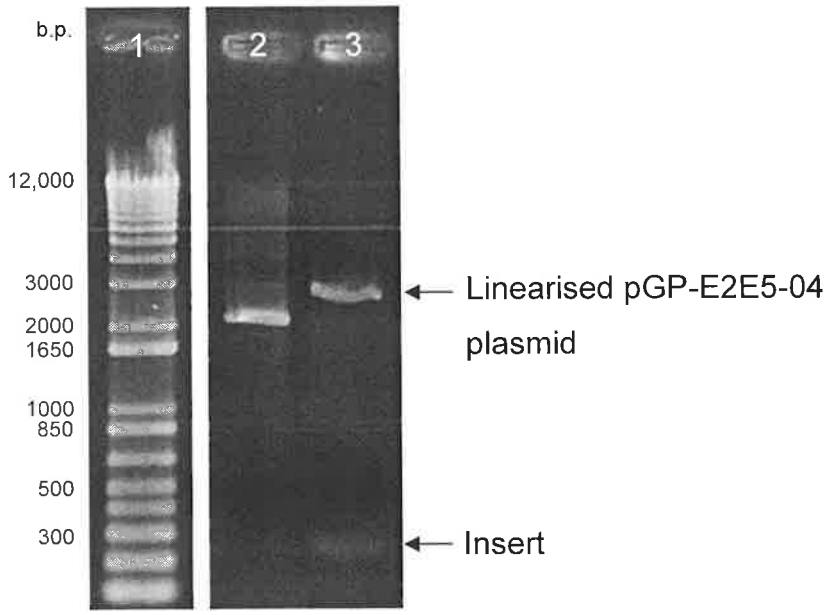


Figure 5.4.1.2 – Diagnostic HpaI/HindIII cleavage of pGP-E2E5-02 plasmid DNA. (1) DNA size markers (b.p.), (2) undigested plasmid, (3) digested pGP-E2E5-04 DNA.



5.4.2 *Generation of Met-pGH(1-11)VN-WGSGPAPM-IGF-I expression plasmid*

The DNA sequence encoding IGF-I with the insertion of Trp-Gly-Ser-Gly upstream of the Pro-Ala-Pro-Met PragA9 cleavage motif was generated by PCR. The 291 bp PCR product can be seen in Figure 5.4.2.1 lanes 3 and 5. More specific production of the fragment was possible by increasing the annealing temperature to 62°C from the 55°C shown in lane 2 of Figure 5.4.2.1. Following ligation and transformation into a suitable host strain, the resulting pGP-WGSG plasmid was redigested for confirmation of insertion of the coding sequence. The outcome of such a digest on plasmid pGP-WGSG-02 appears as Figure 5.4.2.2, where the supercoiling observed in the intact plasmid is relaxed upon digestion to reveal the 2.9 kb linearised vector backbone and the 243 bp insertion. DNA sequencing of

pGP-WGSG-02 plasmid confirmed the successful insertion of the desired sequence.

Figure 5.4.2.1 – PCR production of WGSGPAPM-IGF-I coding sequence insert. (1) DNA size markers (b.p.), (2) preliminary 55°C annealing temperature PCR reaction, (3) final 62°C annealing temperature PCR reaction, (4) no template DNA negative control of reaction 3. (5) Lane 3 sample with loading volume decreased 50% for improved clarity.

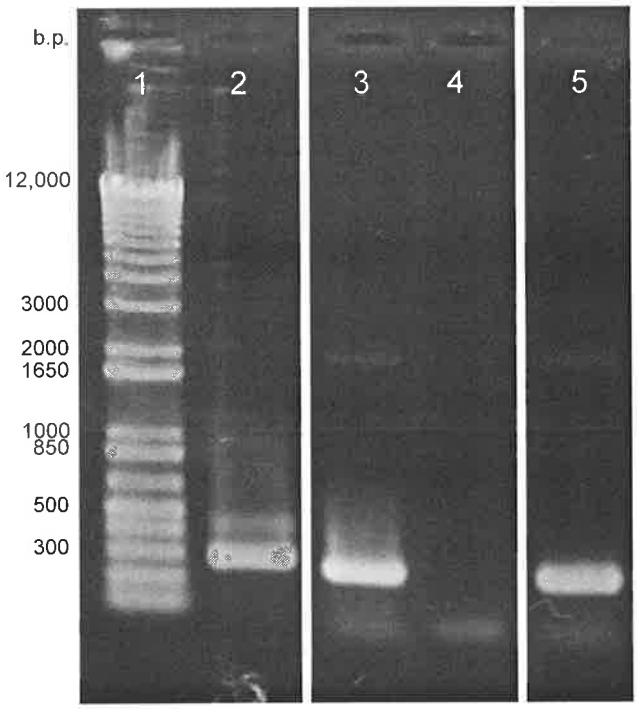
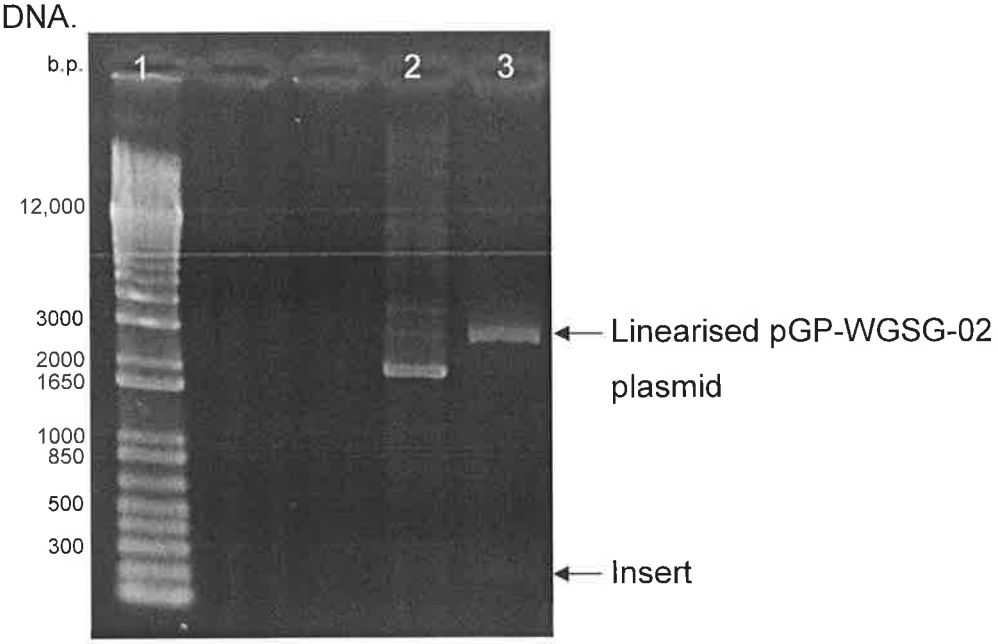


Figure 5.4.2.2 - pGP-WGSG-02 plasmid DNA and diagnostic HpaI/HindIII re-digestion. (1) DNA size markers (b.p.), (2) undigested DNA, (3) digested pGP-WGSG-02 DNA.



5.4.3 *Expression of Met-pGH(1-11)VN-WGSGPAPM-IGF-I and Met-pGH(1-11)VN-[F^{-16}E/M^{-13}E]-PAPM-IGF-I*

Trial induction of the two expression constructs for Met-pGH(1-11)VN-WGSGPAPM-IGF-I and Met-pGH(1-11)VN-[F^{-16}E/M^{-13}E]-PAPM-IGF-I, namely pGP-WGSG-02/101 and pGP-E2E5-04/101, respectively, was performed in MinA minimal medium, however no clear inducible band could be detected. Figure 5.4.3.1 shows the outcome of parallel cultures, differing only in the presence or absence of exogenous inducer (IPTG). Although the loading of all lanes was adjusted to equate the optical density of the cultures, there appears to be a greater level of total protein, particularly high molecular weight, in both cases of the induced cultures. However, no new bands appear in the induced cultures that are not already present in the non-induced. Western blot analysis of the trial expressions shows that

material that interacts with anti-IGF-I antibodies is present in both expression cultures, but these same timecourse analyses confirm the constitutive expression in this medium. Figures 5.4.3.2 and 5.4.3.3 show bands that respond to anti-IGF-I well before the time of induction. The majority of the positive detections in the western blots of Figures 5.4.3.2b and 5.4.3.3b appears to be much higher than the expected approximately 10 kDa of the Met-pGH(1-11)VN-IGF-I analogues. Material of this size is observed lower than the much more prominent higher molecular weight band. This material could possibly represent incomplete reduction of the loaded material or a novel artefact of the gel running or transfer process. As a higher molecular weight species (approximately 16 kDa) is also detected in the Met-pGH(1-11)VN-Arg^3-IGF-I control lane of Figure 5.4.3.4b, that is not observed in the equivalent SDS-PAGE gel, the latter explanation seems plausible.

The apparently constitutive expression of the Met-pGH(1-11)VN-IGF-I analogues in minimal defined medium appears to be a phenomenon brought about by a property of that medium particularly. Analysis of the expression of a very similar construct, differing only from Met-pGH(1-11)VN-WGSGPAPM-IGF-I in the upstream insertion of IQLEVLFQ in place of WGSGPAPM but with the same vector backbone and host *E. coli* strain, in a much richer defined medium (FM11) shows wholly inducible control. No material is detected at the expected size of a Met-pGH(1-11)VN-IGF-I analogue prior to the time of induction, nor are there any bands that interact with anti-IGF-I antibodies, but after the addition of exogenous inducer both are observed.

Figure 5.4.3.1 – SDS-PAGE of the trial expression of pGP-WGSG-02/101 and pGP-E2E5-04/101.

(1) Molecular weight markers, (2) pGP-E2E5-04/101 culture without IPTG, (3) pGP-E2E5-04/101 with IPTG, (4) pGP-WGSG-02/101 culture without IPTG, (5) pGP-WGSG-02/101 culture with IPTG

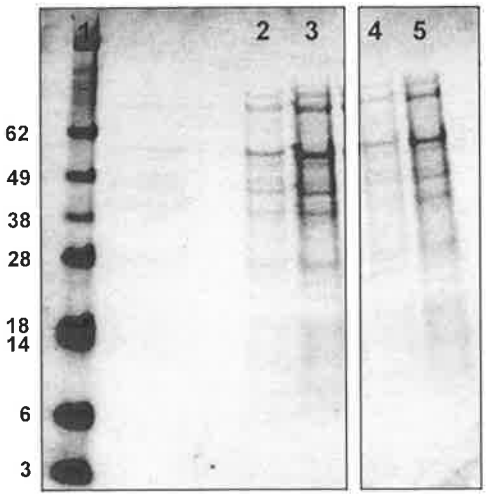


Figure 5.4.3.2 - SDS-PAGE (A) and Western blot (B) of the trial expression of clone pGP-E2E5-04/101 expressing Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I. IPTG inducer was added at hour 6. LR3 = Met-pGH(1-11)VN-Arg^{3}-IGF-I standard.

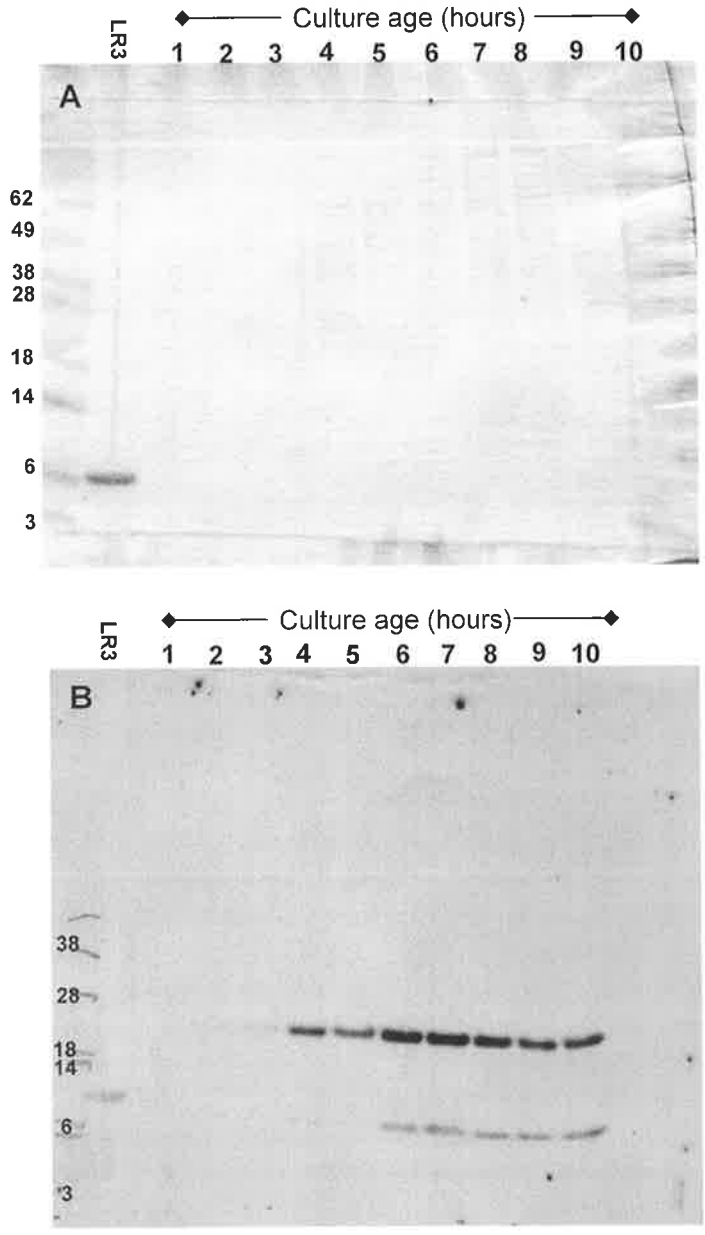


Figure 5.4.3.3 - SDS-PAGE (A) and western blot (B) of the trial expression of clone pGP-WGSG-02/101 expressing Met-pGH(1-11)VN-WGSGPAPM-IGF-I. IPTG inducer was added at hour 6. LR3 = Met-pGH(1-11)VN-Arg^3-IGF-I standard.

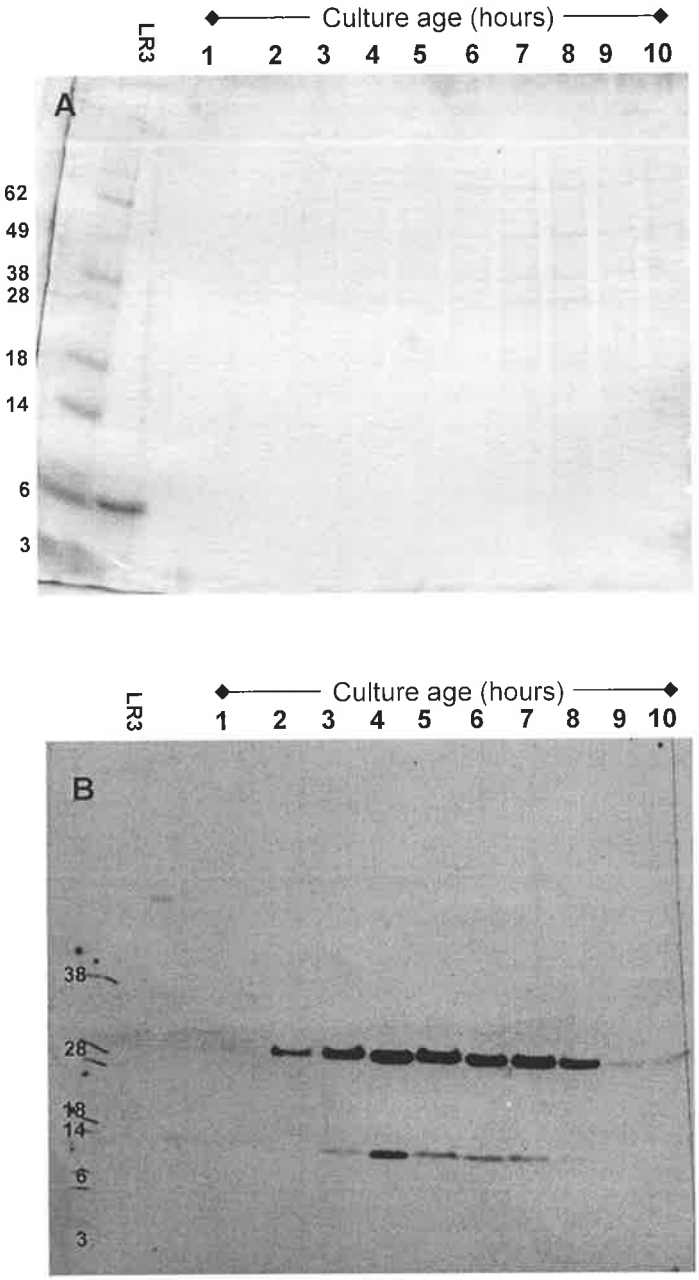
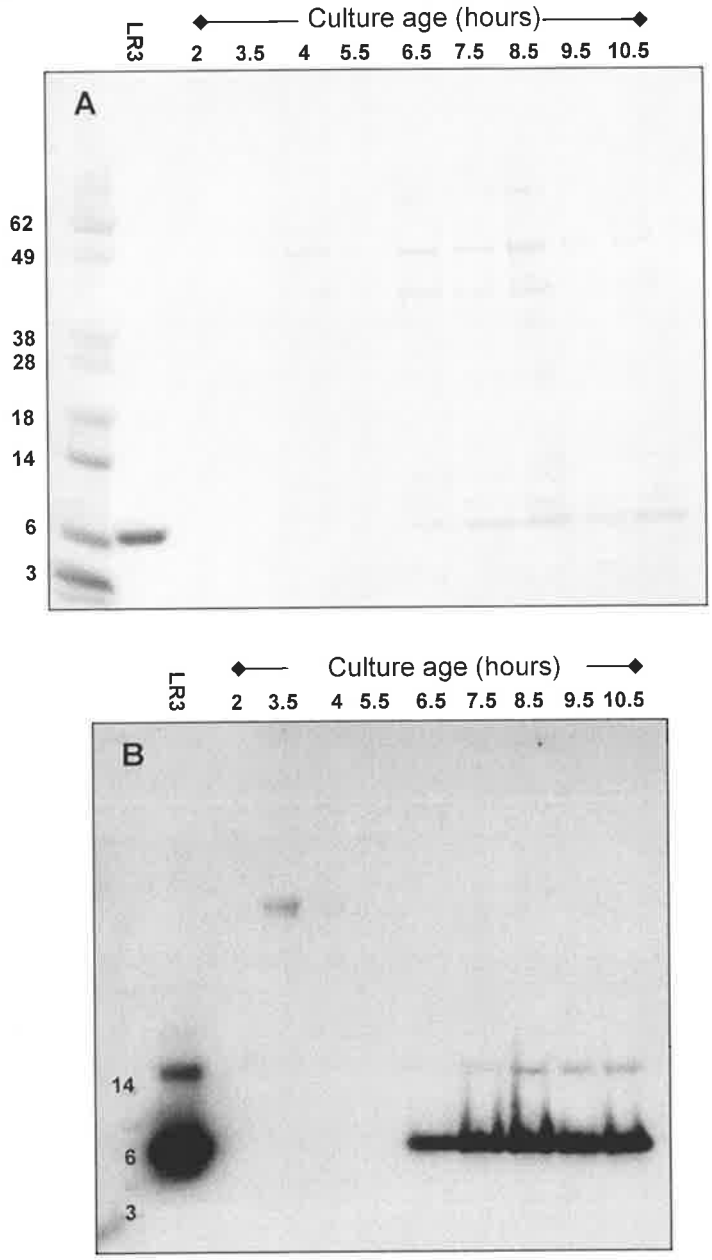


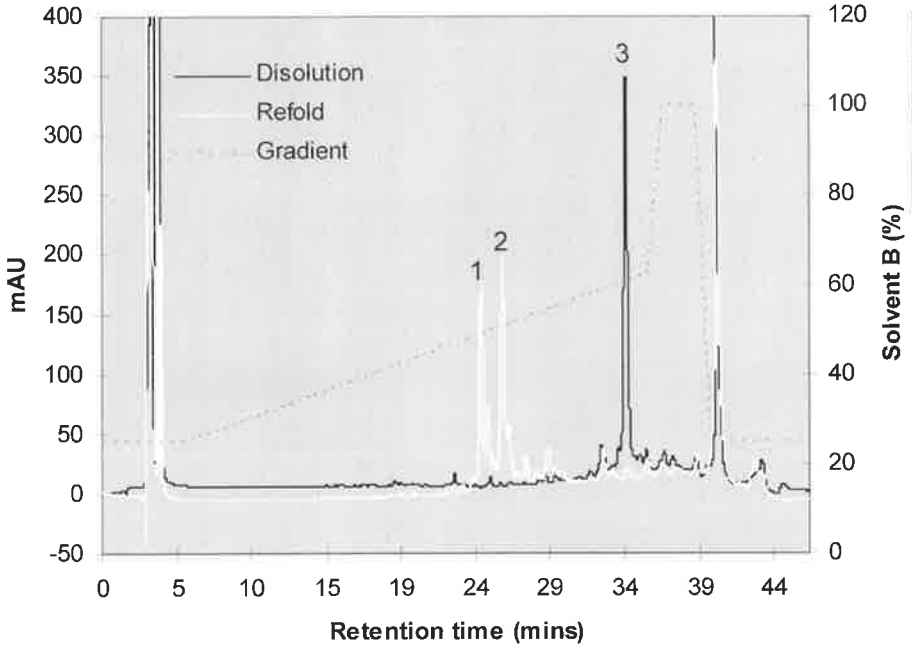
Figure 5.4.3.3 - SDS-PAGE (A) and western blot (B) of the trial expression of clone pGP-3C-01/101 expressing Met-pGH(1-11)VN-IQLEVLFQ -IGF-I in FM11 medium. IPTG inducer was added at hour 5.5. LR3 = Met-pGH(1-11)VN-Arg^3-IGF-I standard. Anti-IGF-I reactive species are visible from 6.5 hours onwards.



5.4.4 Downstream processing of Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I

Figure 5.4.4.1 shows the rp-HPLC profile of the dissolution and subsequent oxidative refolding of Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I. The single peak of the reduced fusion protein at 34 minutes is shifted to two peaks of retention times (R.T.) 24.6 and 26 minutes following the refolding reaction. Both peaks represent forms of Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I, indistinguishable from each other by SDS-PAGE (Figure 5.4.4.2). The peaks could be separated from each other to near homogeneity, as determined by rp-HPLC (Figure 5.4.4.3) and designated isomers 1 and 2 in order of their respective rp-HPLC retention time.

Figure 5.4.4.1 - Long-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I inclusion body processing



Peak	R.T.	Area
1	24.6	2850
2	26.0	3740
3	34.0	5980

Figure 5.4.4.2 – SDS-PAGE of the two peak forms of Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I. Two lanes of each peak are shown. Met-pGH(1-11)VN-Arg3-IGF-I is included as a standard (labelled LR3).

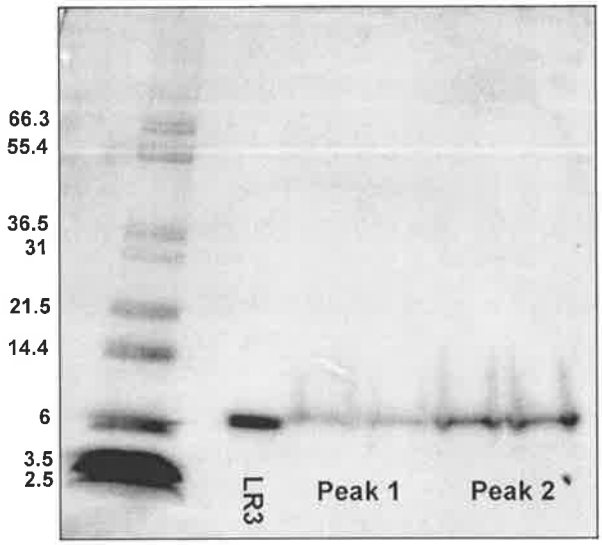
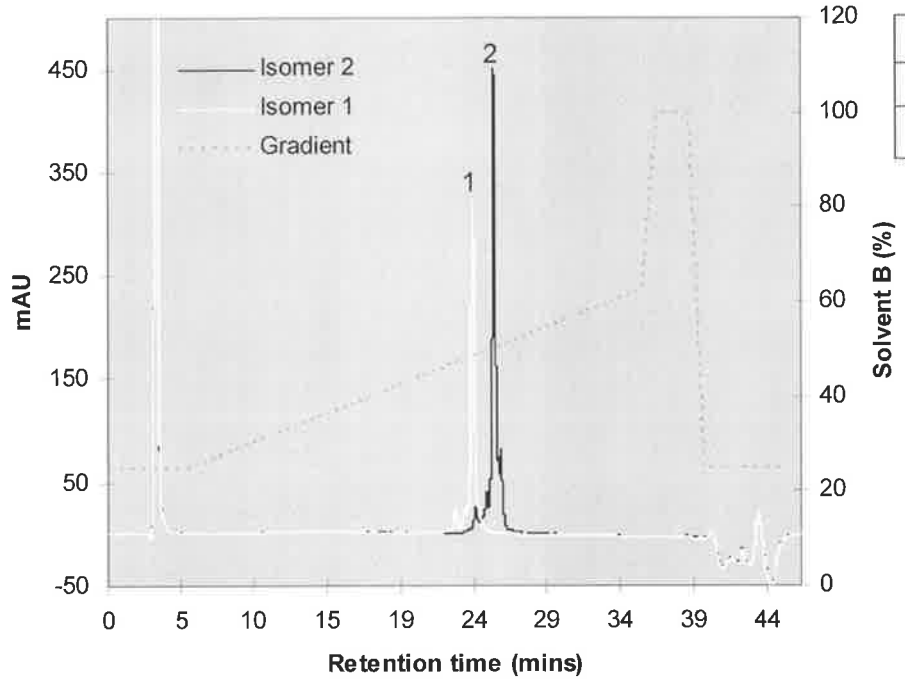


Figure 5.4.4.3 - Long-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I isomer separation

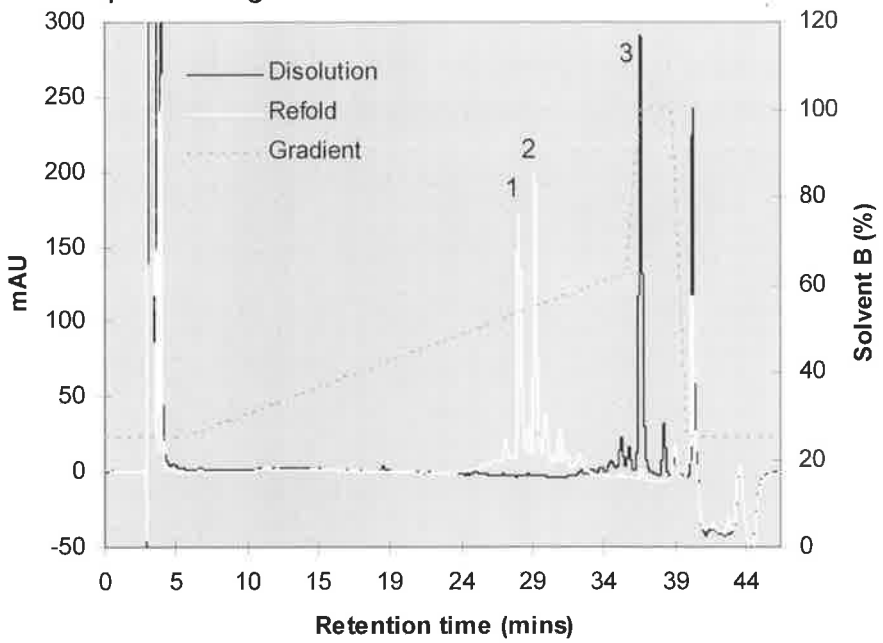


Peak	R.T.	Area
1	24.2	5890
2	25.6	8680

5.4.5 Downstream processing of Met-pGH(1-11)VN-WGSGPAPM-IGF-I

Figure 5.4.5.1 shows the rp-HPLC profile of the dissolution and subsequent oxidative refolding of Met-pGH(1-11)VN-WGSGPAPM-IGF-I. The single peak of the reduced fusion protein (3), eluting after the main gradient at 36.4 minutes, is shifted to two main peaks (1 and 2) at 28.2 and 29.3 minutes following the refolding reaction. Both peaks represent forms of Met-pGH(1-11)VN-WGSGPAPM-IGF-I, indistinguishable from each other by SDS-PAGE (Figure 5.4.5.2). The peaks could be separated from each other to near homogeneity, as determined by rp-HPLC (Figure 5.4.5.3) and designated isomers 1 and 2 in order of their respective rp-HPLC retention time.

Figure 5.4.5.1 – Met-pGH(1-11)VN-WGSGPAPM-IGF-I inclusion body processing



Peak	R.T.	Area
1	28.2	2780
2	29.3	2530
3	36.4	4920

Figure 5.4.5.2 – SDS-PAGE of the two peak forms of Met-pGH(1-11)VN-WGSGPAPM-IGF-I. Met-pGH(1-11)VN-Arg3-IGF-I is included as a standard (labelled LR3).

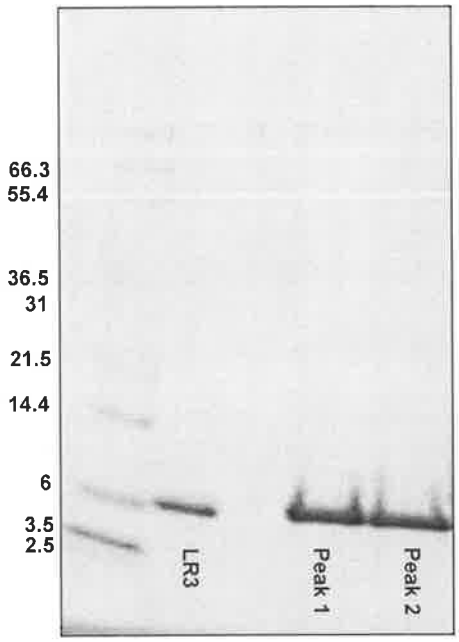
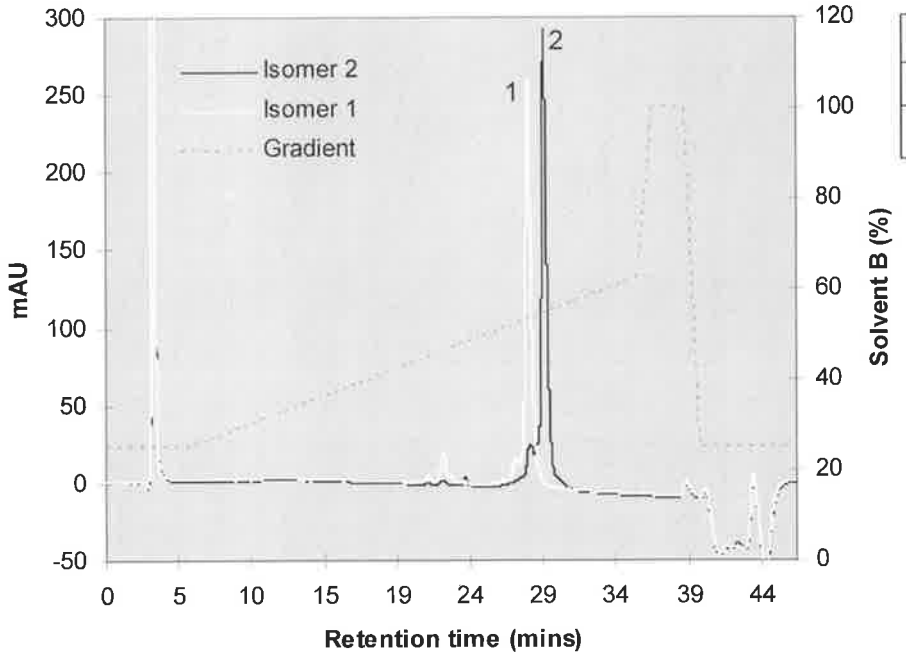


Figure 5.4.5.3 - Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer separation



Peak	R.T.	Area
1	28.1	5350
2	29.2	7200

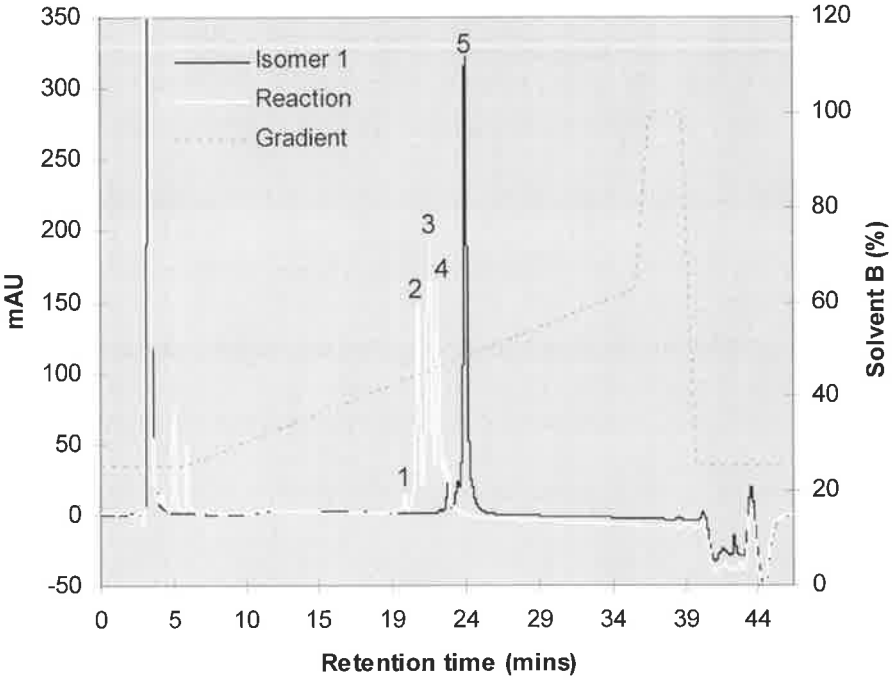
5.4.6 Cleavage of Met-pGH(1-11)VN-[F^{-16}E/M^{-13}E]-PAPM-IGF-I

Cleavage reactions of Met-pGH(1-11)VN-[F^{-16}E/M^{-13}E]-PAPM-IGF-I isomer 1 were performed in both HEPES and Urea/Tris cleavage buffers to evaluate the effectiveness of the charge substitution in the leader peptide. rp-HPLC analysis of these reactions are displayed as Figures 5.4.6.1 and 5.4.6.2. Three major peaks (2, 3 and 4) at 21.2, 21.9 and 22.4 minutes and one minor product peak (1) at 20.2 minutes are observed in the HEPES buffer reaction (Figure 5.4.6.1) and two major peaks (2 and 3) with retention times of 21.1 and 21.8 minutes and one minor peak at 20.2 minutes are detected in the Urea/Tris reaction (Figure 5.4.6.2). Material with the 22.4 minute retention time of the HEPES buffer reaction appears to be present in the Urea/Tris reaction, but at very low abundance. The retention times of the products of both reactions agree favourably with each other, albeit at differing abundances. Interestingly, the minor peak present in these reactions appears to have a close agreement in retention time to that of misfolded IGF-I at 20.2 minutes (see Figure 5.4.6.3).

The outcome of the reactions of isomer 1 are mirrored for the cleavage reactions of isomer 2, with rp-HPLC chromatograms of these reactions shown as Figures 5.4.6.4 and 5.4.6.5. However, isomer 2 tended to generate only two dominant product peaks (2 and 3) in both buffer systems, with retention times of 22.4/22.6 and 23.2/23.3 minutes. In cleavages of this latter isomer, an early minor peak (1) is again detected, with a retention time of 21.5 minutes. In this case however, the retention time is in very close agreement to that of authentic IGF-I. A chromatogram of commercial IGF-I reference

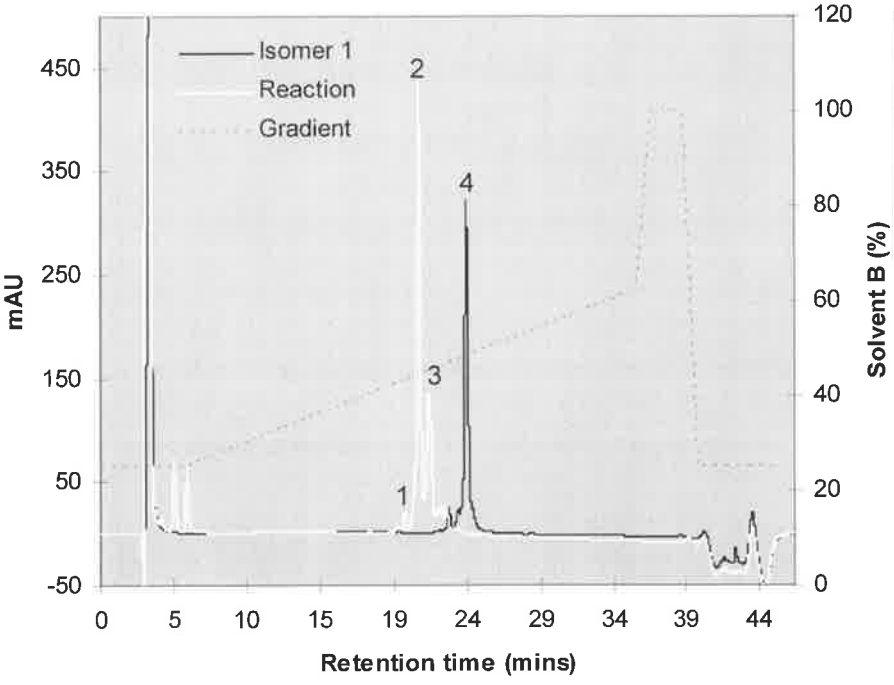
standard appears as Figure 5.4.6.6 and shows a retention time of 21.4 minutes.

Figure 5.4.6.1 - Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I isomer 1 cleavage in HEPES cleavage buffer.



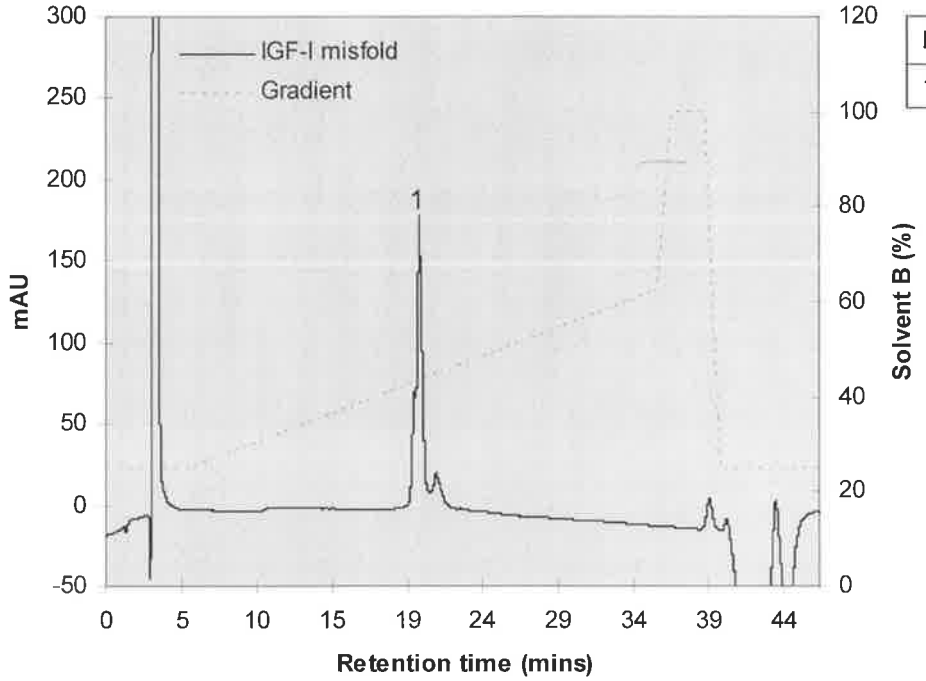
Peak	R.T.	Area
1	20.2	90
2	21.1	1020
3	21.9	1860
4	22.4	1070
5	24.2	5890

Figure 5.4.6.2 - Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I isomer 1 cleavage in Urea/Tris cleavage buffer.



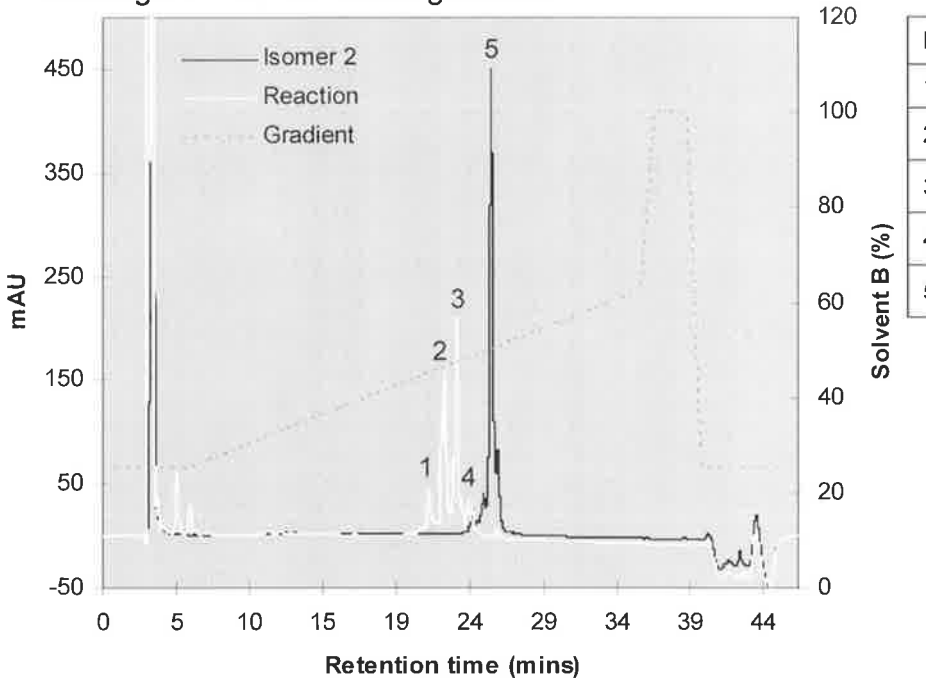
Peak	R.T.	Area
1	20.2	110
2	21.1	2760
3	21.8	1360
4	24.2	5890

Figure 5.4.6.3 – Misfolded IGF-I



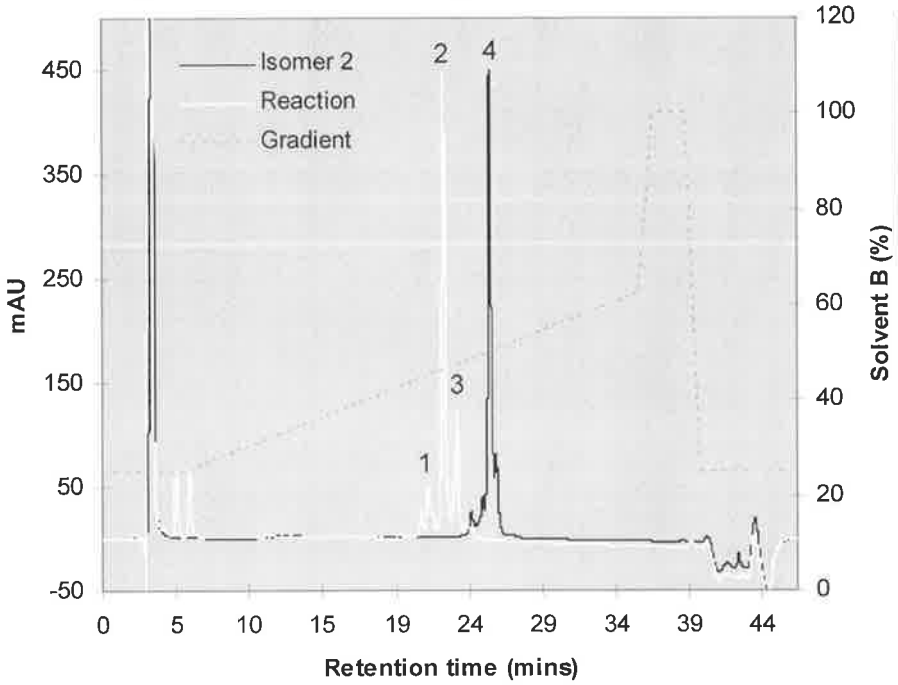
Peak	R.T.	Area
1	20.2	3760

Figure 5.4.6.4 - Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I isomer 2 cleavage in HEPES cleavage buffer.



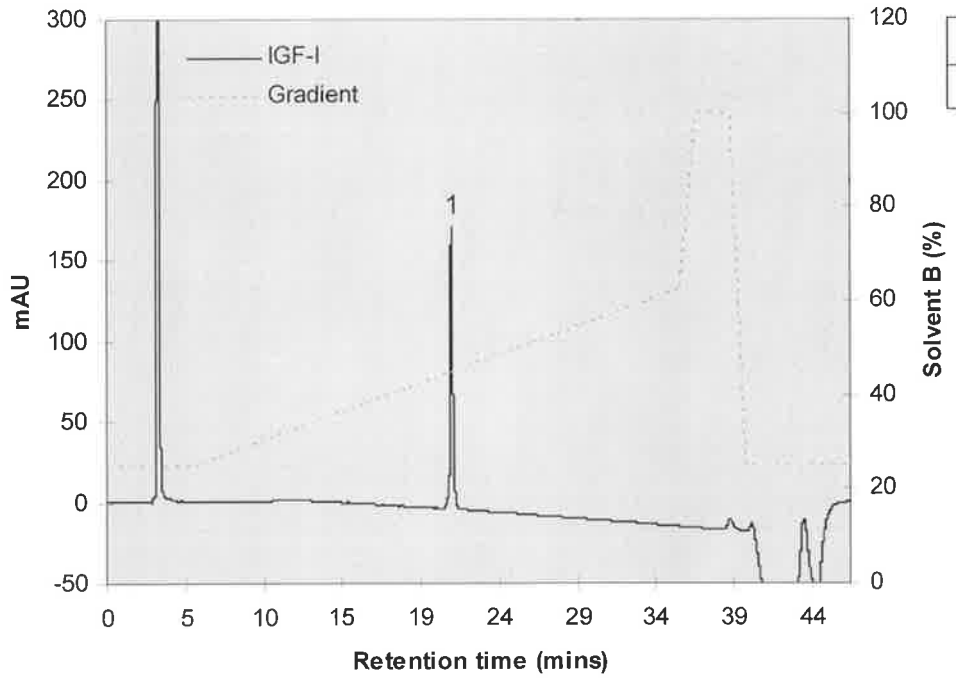
Peak	R.T.	Area
1	21.5	650
2	22.4	3080
3	23.2	3420
4	24.2	780
5	25.6	8680

Figure 5.4.6.5 - Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I isomer 2 cleavage in Urea/Tris cleavage buffer.



Peak	R.T.	Area
1	21.5	620
2	22.6	4250
3	23.3	2080
4	25.6	8680

Figure 5.4.6.6 – Commercial IGF-I reference standard



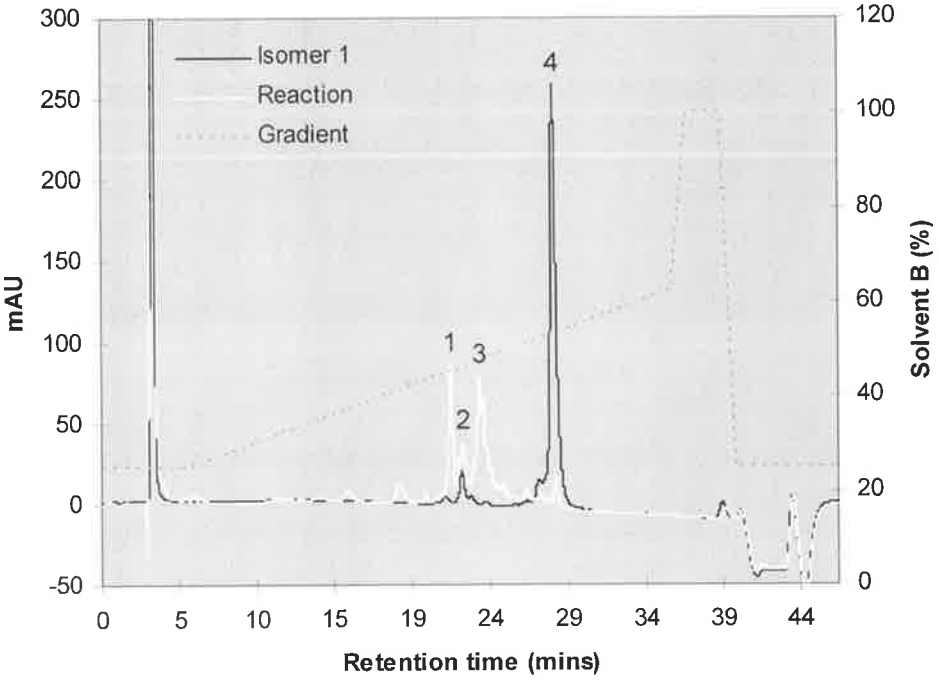
Peak	R.T.	Area
1	21.4	1910

5.4.7 Cleavage of Met-pGH(1-11)VN-WGSGPAPM-IGF-I

Three significant peaks (1, 2 and 3) of retention time 21.8, 22.5 and 23.6 minutes are visible in the chromatogram of the cleavage reaction of Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 1 (Figure 5.4.7.1). As in the cleavage reactions of the previous construct (Section 5.4.6), no peaks in the cleavage of Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 1 had a retention time comparable to that of authentic IGF-I.

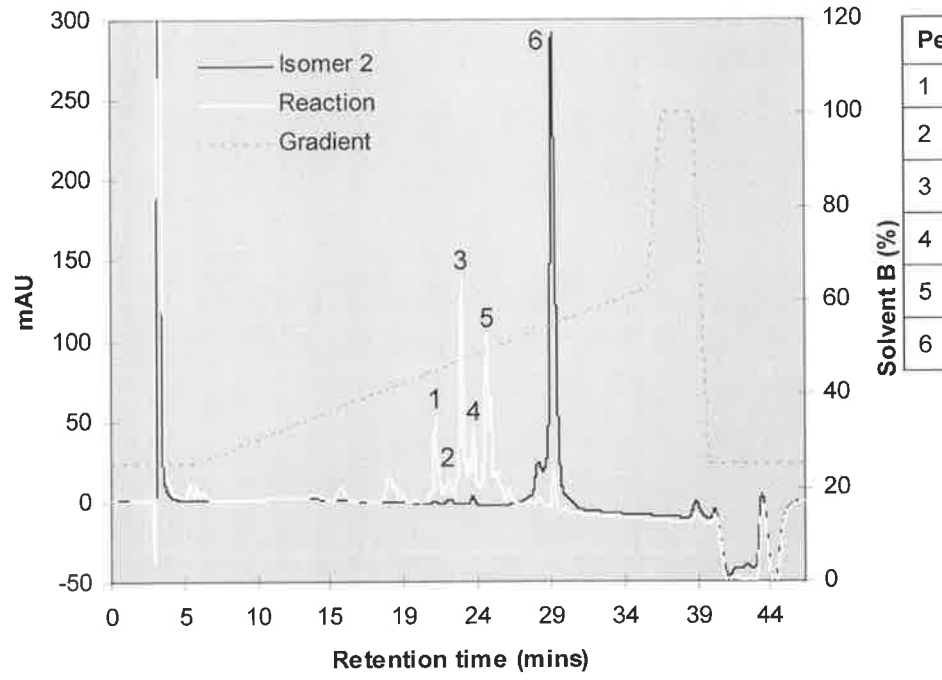
The outcome of the cleavage of Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 2 is considerably different again, with a cluster of peak species (1 – 5) detected (Figure 5.4.7.2). As with the cleavage of isomer 2 of Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I (Section 5.4.6), one of these products has a close agreement in retention time to authentic IGF-I reference standard. This reaction has the highest abundance of the putative authentic IGF-I peak of the two constructs as well as some residual full-length fusion protein substrate, suggesting that further product generation may be possible.

Figure 5.4.7.1 - Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 1 cleavage in Urea/Tris cleavage buffer.



Peak	R.T.	Area
1	21.8	1300
2	22.5	890
3	23.6	2290
4	28.1	5350

Figure 5.4.7.2 - Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 2 cleavage in Urea/Tris cleavage buffer.



Peak	R.T.	Area
1	21.6	820
2	22.4	430
3	23.3	1980
4	24.0	620
5	25.0	2700
6	29.2	9900

5.4.8 Separation of Met-pGH(1-11)VN-WGSGPAPM-IGF-I cleavage products

The first four peak species of Met-pGH(1-11)VN-WGSGPAPM-IGF-I cleavage observed in Figure 5.4.7.2 were separated from each other and labelled reaction products 1 to 4, in order of increasing retention time. rp-HPLC of these isolates appears as Figure 5.4.8.1 and SDS-PAGE as 5.4.8.2. An overlay of the chromatogram of the purified putative IGF-I peak (reaction product 1) of the cleavage reaction with that of commercial IGF-I reference standard is shown as Figure 5.4.8.3.

Figure 5.4.8.1 – Isolated Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 2 cleavage products

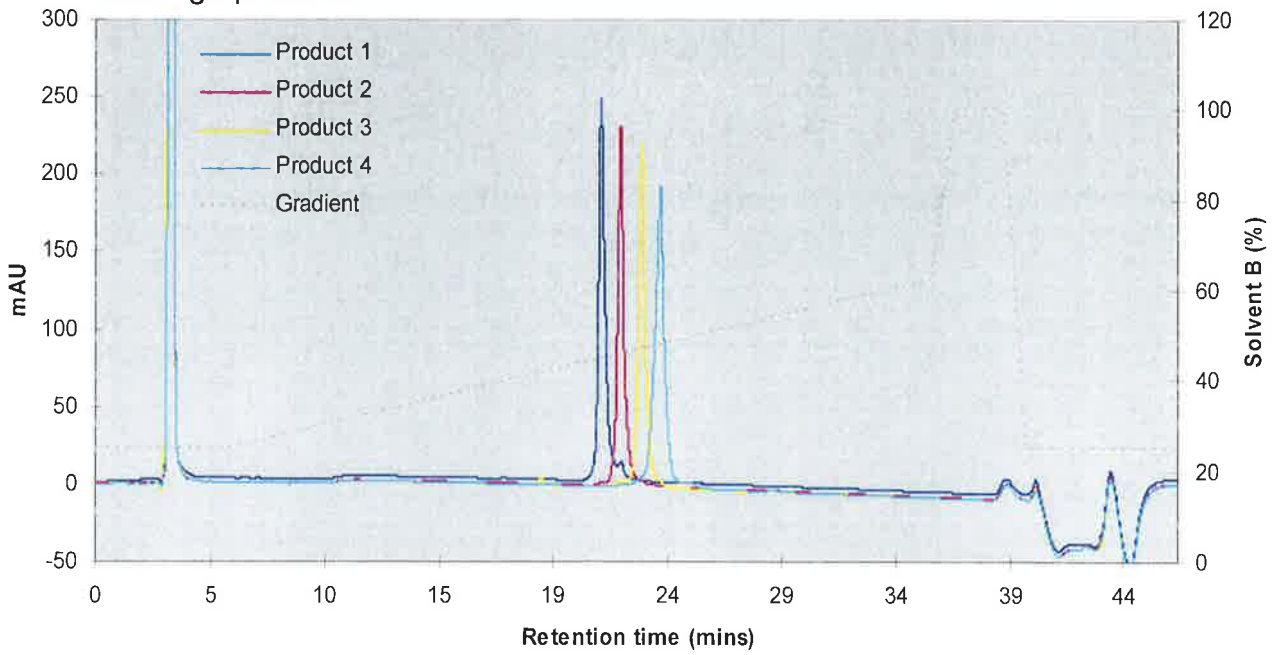
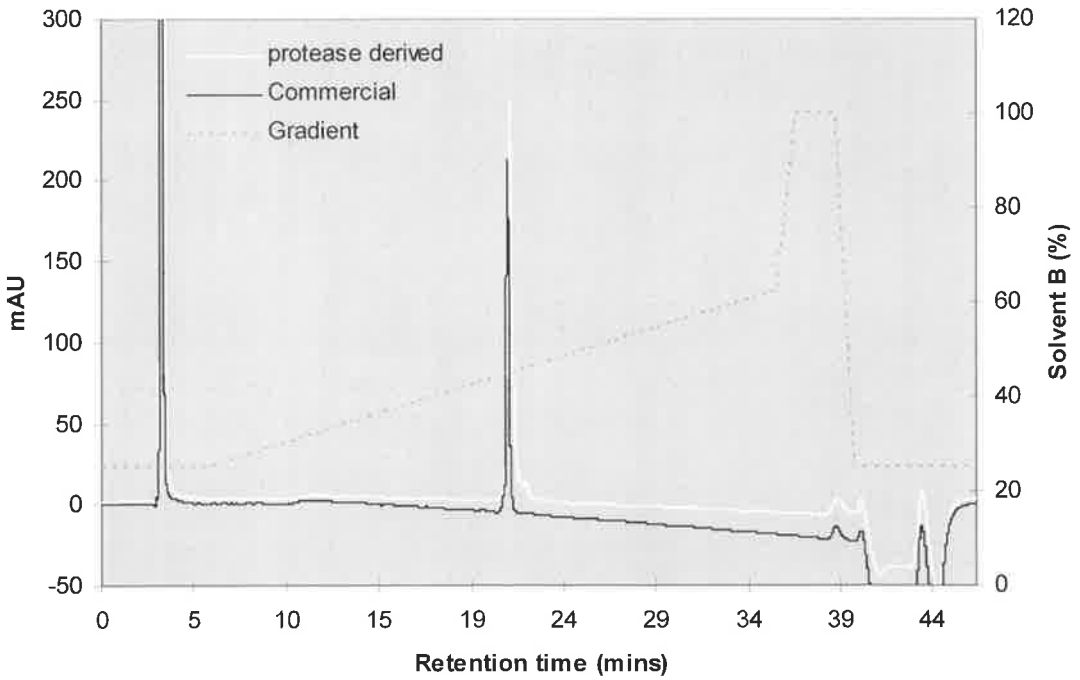


Figure 5.4.8.2 – SDS-PAGE of isolated Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 2 cleavage products. "Substrate" is the input isomer 2 used for the reaction.

Figure 5.4.8.3 – Overlaid chromatograms of commercial IGF-I reference standard and IGF-I-like peak of Met-pGH(1-11)VN-WGSGPAPM-IGF-I cleavage



5.4.9 Mass spectrometry of Met-pGH(1-11)VN-WGSGPAPM-IGF-I cleavage products

The isolated cleavage products displayed in Figure 5.4.8.1 were analysed by mass spectrometry. The spectrum of full-length Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 2 substrate appears as Figure 5.4.9.1 and is resolved to reveal a major mass species of 9868 Da and a minor species of 9589 Da. Figure 5.4.9.2 is the spectrum of the first of the four reaction products and is resolved to reveal a mass species of 7648 Da, which compares closely with the calculated mass of authentic IGF-I at 7649 Da.. Figure 5.4.9.3 displays the spectrum of reaction product 2 and is resolved to contain a mass species of 7362 Da, which compares well to the calculated mass of des(68-70) IGF-I at 7632 Da. The spectrum of reaction product 3 is shown as Figure 5.4.9.4 which is interpreted to contain a mass species of 8432 Da, which compares closely with the calculated mass of WGSGPAPM-IGF-I at 8433 Da. Finally, the mass spectrum of reaction product 4 is displayed as Figure 5.4.9.4, with a major mass species resolved at 8144Da, which compares closely with the calculated mass of des(68-70)-WGSGPAPM-IGF-I at 8146 Da.

Figure 5.4.9.1 - Mass spectrum of Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 2.

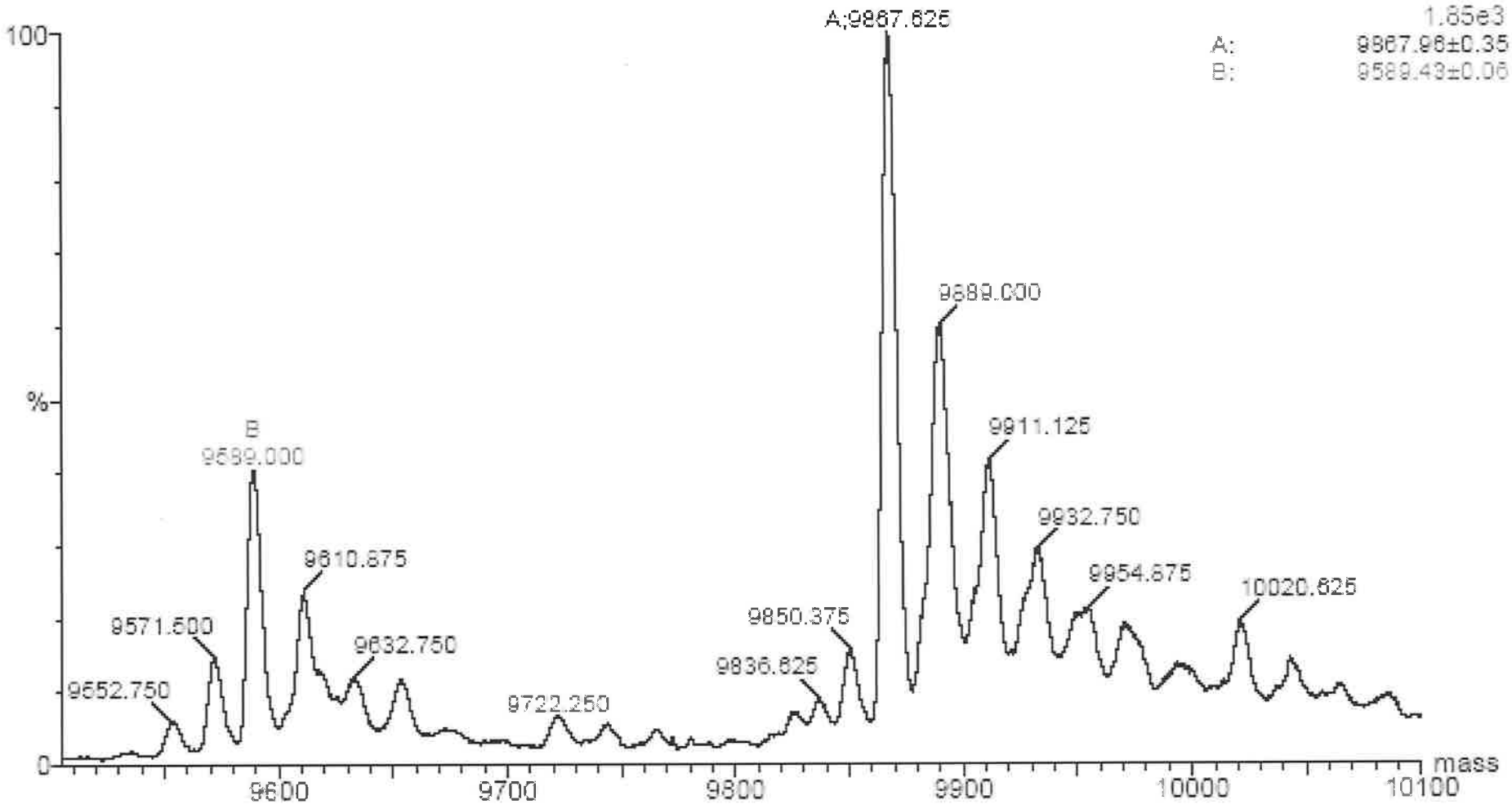


Figure 5.4.9.2 – Mass spectrum of reaction product 1

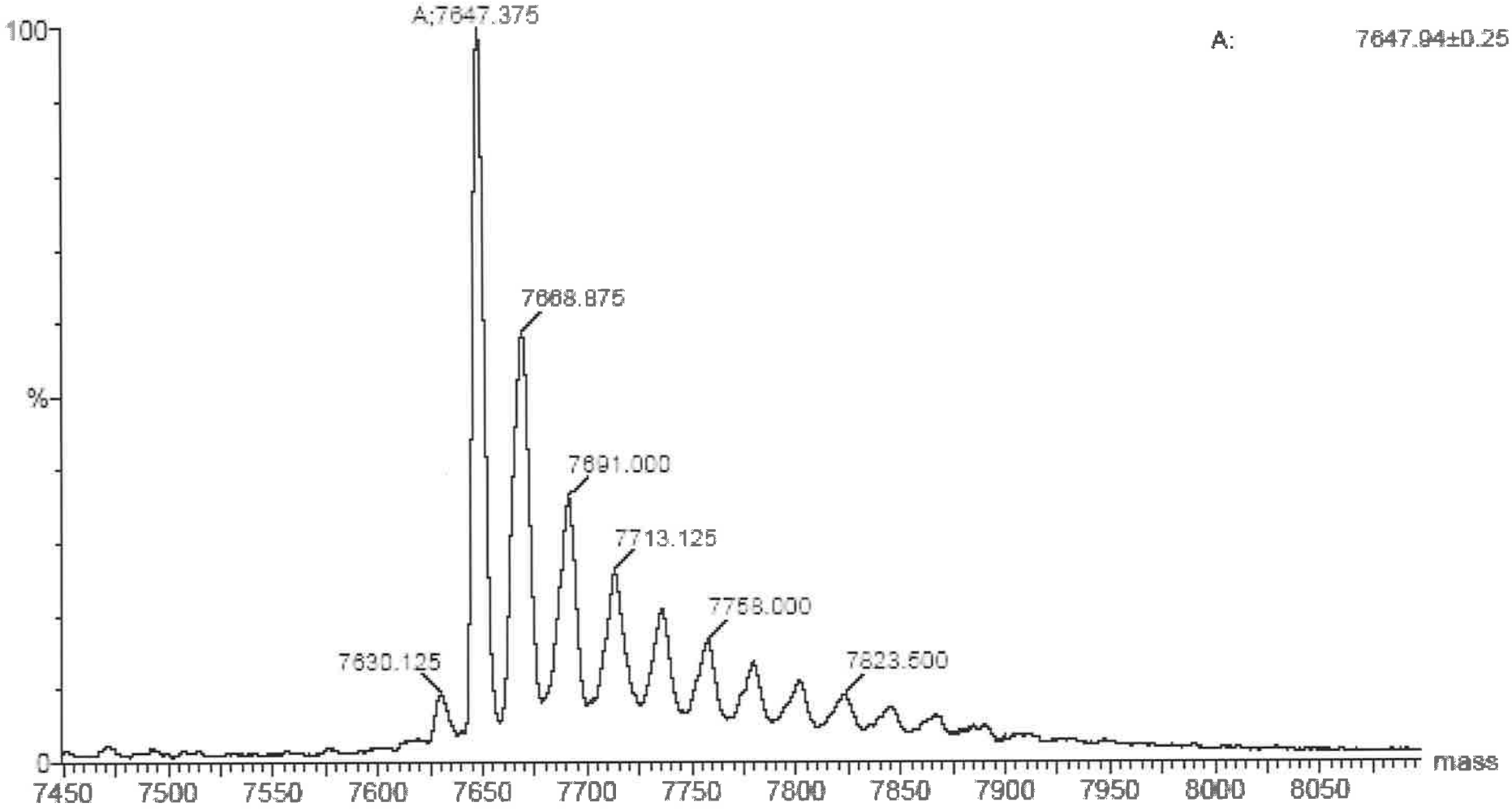


Figure 5.4.9.3 – Mass spectrum of reaction product 2

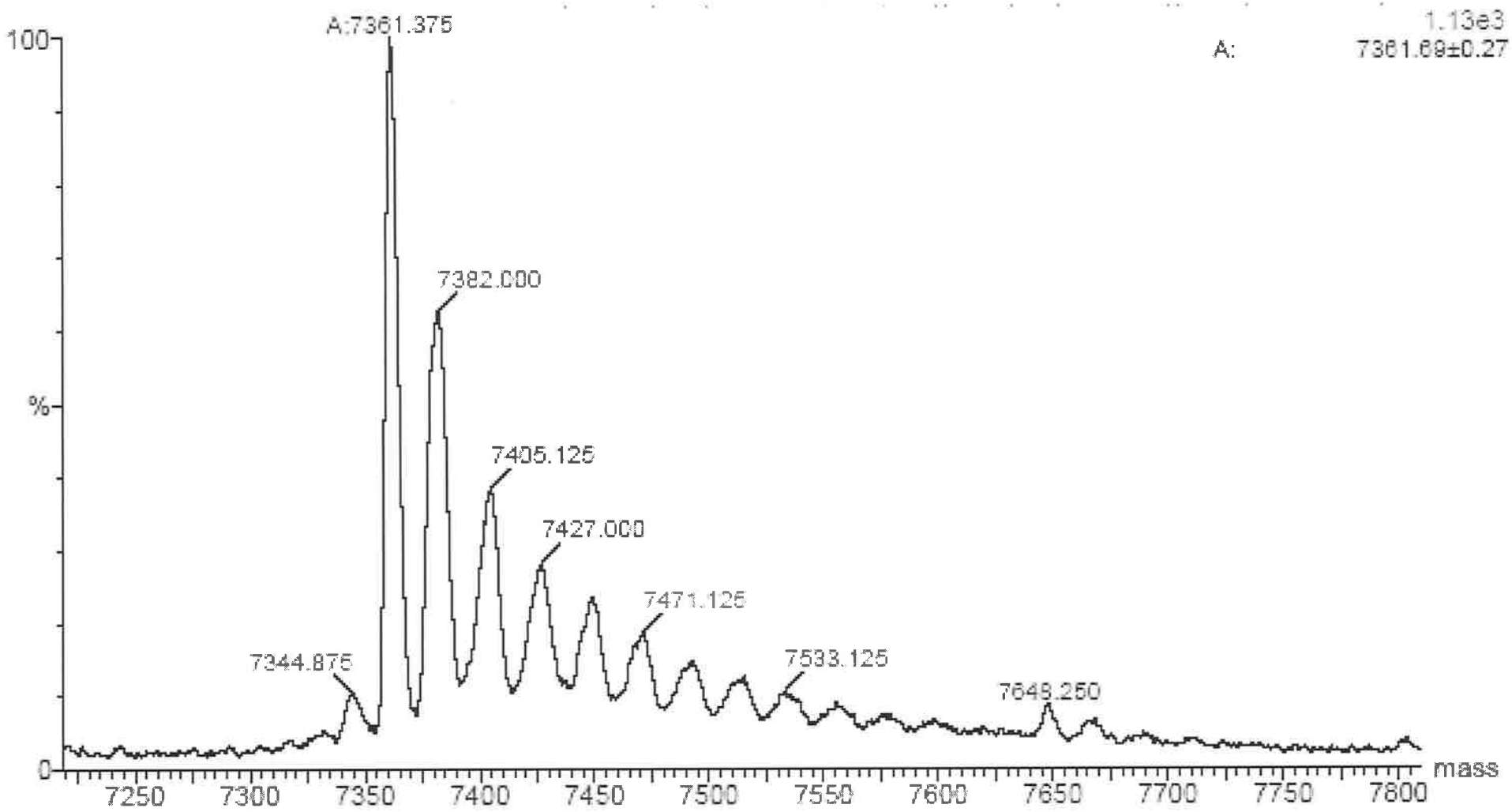


Figure 5.4.9.4 – Mass spectrum of reaction product 3

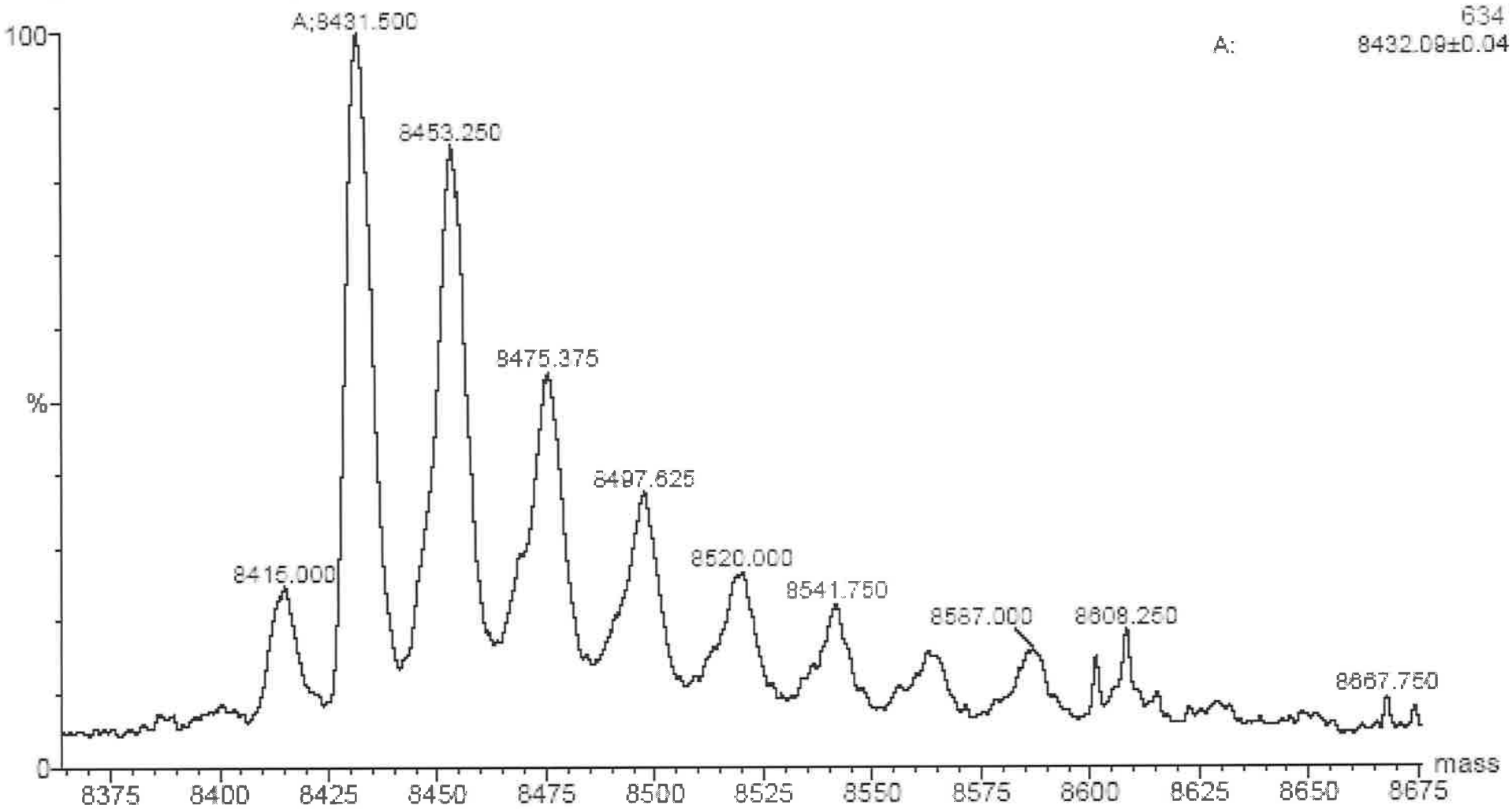
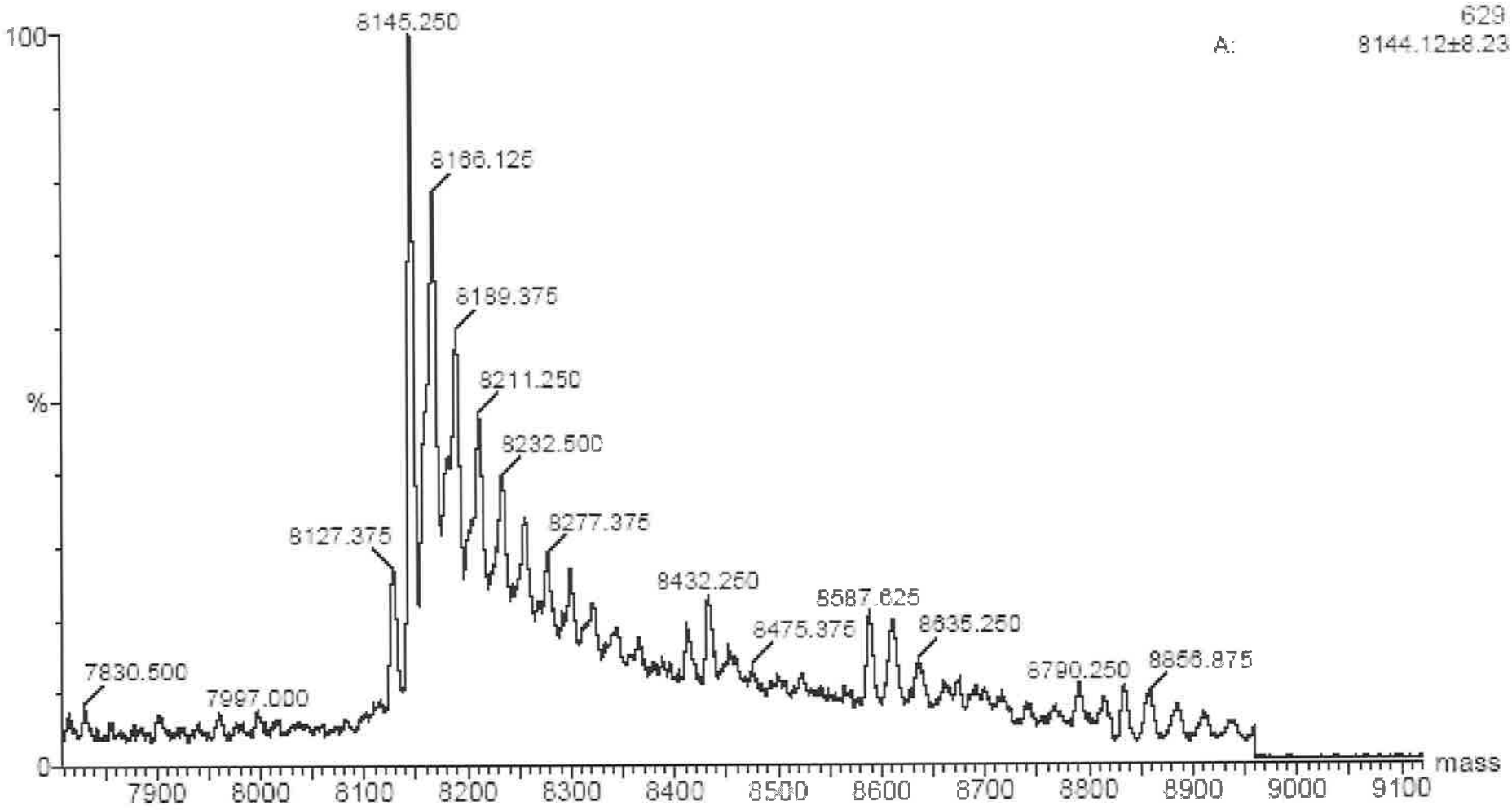


Figure 5.4.9.5 – Mass spectrum of reaction product 4



5.4.10 Optimisation of Met-pGH(1-11)VN-WGSGPAPM-IGF-I cleavage

In the data of this and subsequent sections the reaction products likely identities are based, in the absence of any sequencing data, on the mass spectrometry data (see Discussion). Thus "reaction product 1" compares closely with the molecular weight of authentic IGF-I, "reaction product 2" compares closely with the molecular weight of des(68-70)IGF-I, "reaction product 3" compares closely with the molecular weight of WGSGPAPM-IGF-I and "reaction product 4" compares closely with the molecular weight of des(68-70)-WGSGPAPM-IGF-I.

Various alterations of the cleavage conditions that gave rise to the products displayed in Figure 5.4.7.2 were employed in an attempt to increase the yield of authentic IGF-I. In Table 5.4.10.1 enzyme:substrate ratios, incubation times and temperatures were manipulated in an attempt to drive the reaction further towards completion. The abundance of authentic IGF-I could be increased to 20% of reaction products.

The order in which the various products are formed was investigated by incubating the Met-pGH(1-11)VN-WGSGPAPM-IGF-I substrate at a high enzyme:substrate ratio and observing their formation over time. This data is summarised in Table 5.4.10.2. It appears that the WGSGPAPM-IGF-I form is generated first, then authentic IGF-I. The des(68-70) forms of both WGSGPAPM-IGF-I and authentic IGF-I are generated following their native C-terminal counterparts, see discussion for more details. Whilst the abundance of all other species appear to be in flux, there is a steady accumulation of des(68-70)-IGF-I, suggesting that this is the terminal product of the reaction.

In the final set of the limited round of optimisations, the best set of conditions of Table 5.4.10.1 were held constant and other factors investigated such as the inclusion of organic solvents, the urea concentration and reaction pH in an attempt to change the relative ratio of the output products of the reaction, with these results summarised in Table 5.4.10.3. In none of the conditions employed could the des(68-70) truncation be avoided, nor could its abundance relative to authentic IGF-I be altered to favour recovery of more intact IGF-I. In fact, pH > 9 appeared to favour the formation of the des(68-70) truncation, but this is more likely to be a reflection of the reactions tending further toward completion, as there is a concurrent loss of the WGSGPAPM-IGF-I intermediate and its des(68-70) counterpart.

A chromatogram of the products generated by the optimum conditions determined by these experiments appears as Figure 5.4.10.1 and typical process yields of the optimum cleavage, from inclusion bodies through to highly purified authentic human IGF-I appears as Table 5.4.10.4. As an uninterrupted, continuous process was never performed on an entire single fermentation batch, with unit operations conducted on small samples of the bulk preparation, the theoretical process yield is calculated by compounding the yields from each step. The outcome for the process is thus determined to be 2.7%

Table 5.4.10.1 – Met-pGH(1-11)VN-WGSGPAPM-IGF-I cleavage optimisation round 1. Enzyme:substrate mass ratio, incubation temperature and time were varied to maximise the yield of authentic IGF-I. All reactions were performed in Urea/Tris cleavage buffer. The %Input peak area detected is relative to the initial substrate (*i.e.* 100% in the no protease control reaction) and the abundance of each cleavage product is its relative proportion of the total peaks detected in the cleavage reaction. nd = not detected, RT = room temperature (20 – 22°C).

Mass Ratio	Other conditions	%Input peak area detected	%Uncut substrate	% WGSGPAPM-IGF-I	% des(68-70) WGSGPAPM-IGF-I	% IGF-I	% des(68-70) IGF-I
1:1000	RT, 16 hrs	72	65	5	nd	nd	nd
1:500	RT, 16 hrs	71	52	8	1.7	1.1	nd
1:200	RT, 16 hrs	78	8.5	23	6.6	5.8	2.5
1:100	RT, 16 hrs	74	2	30	9.5	12	6
1:100	37°C, 16 hrs	68	nd	23	12	11	8
1:100	RT, 48 hrs	73	nd	31	11.5	15	9

Mass Ratio	Other conditions	%Input peak area detected	%Uncut substrate	% WGSGPAPM-IGF-I	% des(68-70) WGSGPAPM-IGF-I	% IGF-I	% des(68-70) IGF-I
1:50	RT, 16 hrs	72	nd	32	11	14.5	6
1:40	RT, 16 hrs	60	nd	31.5	13	18	13
1:35	RT, 16 hrs	64	nd	29	14	19	16.5
1:30	RT, 16 hrs	62	nd	28	14	20	19
1:25	RT, 16 hrs	61	nd	30	13	19	16
1:20	RT, 16 hrs	52	nd	13.5	14	19.5	30
1:10	RT, 16 hrs	43	nd	5.5	7.8	10.5	38

Table 5.4.10.2 – Met-pGH(1-11)VN-WGSGPAPM-IGF-I cleavage timecourse. Unless otherwise stated, conditions for these reactions were a 1:10 enzyme:substrate ratio. All reactions were performed in Urea/Tris cleavage buffer at room temperature (20 – 22°C). The %Input peak area detected is relative to the initial substrate and the abundance of each cleavage product is its relative proportion of the total peaks detected in the cleavage reaction. nd = not detected.

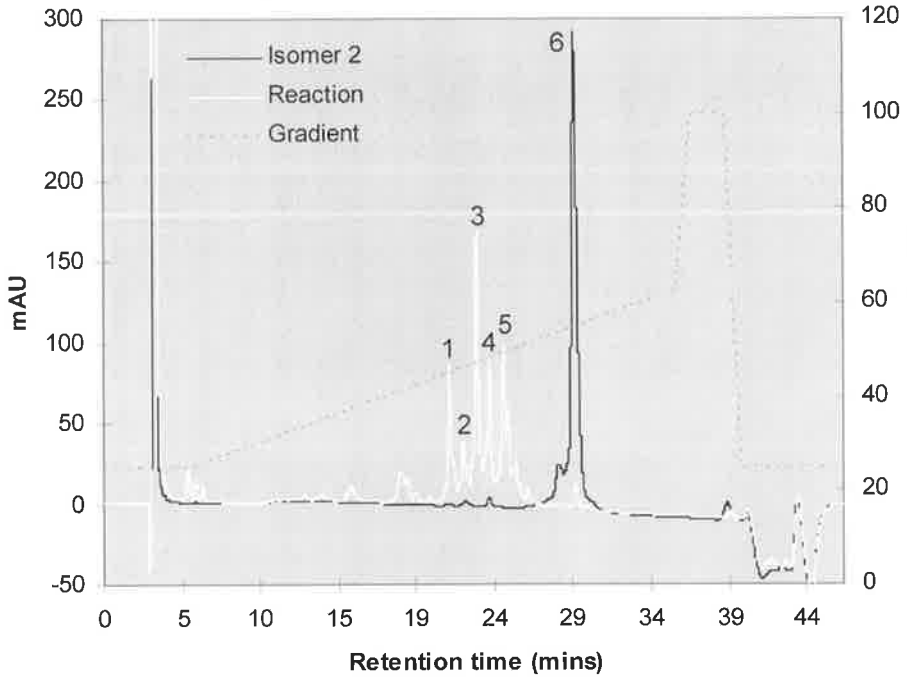
Time	%Input peak area detected	%Uncut substrate	% WGSGPAPM-IGF-I	% des(68-70) WGSGPAPM-IGF-I	%IGF-I	% des(68-70) IGF-I
5 min	85.5	78.5	4	nd	nd	nd
15 min	78.5	59	10	nd	3	nd
30 min	74	37	17	nd	6	nd
1 hr	73	14	24	7.6	8.5	4
2 hr	59	2.5	34	14	15	8
4 hr	60	nd	33	15.5	20	15
8 hr	48	nd	24	26	22	28
16 hr	43	nd	5.5	7.8	10.5	38
1:30 ratio, 4 hr	71	9	29	9.5	11	6

Table 5.4.10.3 – Met-pGH(1-11)VN-WGSGPAPM-IGF-I cleavage optimisation round 2. Reaction pH, urea concentration and inclusion of organic solvent were investigated to improve the yield of authentic IGF-I. All reactions were performed at an enzyme:substrate mass ratio of 1:30 at room temperature (20 – 22°C) for 16 hours. The %Input peak area detected is relative to the initial substrate and the abundance of each cleavage product is its relative proportion of the total peaks detected in the cleavage reaction. nd = not detected, ACN = acetonitrile.

Buffer, pH	Other conditions	%Input peak area detected	%Uncut substrate	% WGSGPAPM-IGF-I	% des(68-70) WGSGPAPM-IGF-I	% IGF-I	% des(68-70) IGF-I
200 mM Tris, pH 8.5	4 M urea	68	42	17	nd	3.5	nd
200 mM Tris, pH 10	2 M urea	55	nd	17	8.7	16	37
200 mM Tris, pH 9.5	2 M urea	58	nd	17	10	16	26
200 mM Tris, pH 9	2 M urea	62	nd	18	8.7	17	25
200 mM Tris, pH 8.5	2 M urea	61	nd	29	13	19	21
200 mM Tris, pH 8	2 M urea	57	nd	14	10	17	28

Buffer, pH	Other conditions	%Input peak area detected	%Uncut substrate	% WGSGPAPM-IGF-I	% des(68-70) WGSGPAPM-IGF-I	% IGF-I	% des(68-70) IGF-I
200 mM Tris, pH 7.5	2 M urea	53	3	32	7.1	16.7	6.5
200 mM sodium citrate, pH 7.5	2 M urea	61	1.5	26	7.6	14.5	7.5
200 mM sodium citrate, pH 7.0	2 M urea	65	4.5	24	7	13	6
200 mM sodium citrate, pH 6.5	2 M urea	59	17	21	4.4	11	4
200 mM sodium citrate, pH 6	2 M urea	61	44	13	1.9	6	trace
200 mM Tris, pH 8.5	2 M urea, 5% ACN	50	nd	20	12.5	19	18
200 mM Tris, pH 8.5	5% ACN	67	nd	11.5	5.5	4	67
200 mM Tris, pH 8.5	2 M urea, 15% ACN	74	13	22	6	7	5
200 mM Tris, pH 8.5	15% ACN	77	nd	23	17.5	9	19
200 mM Tris, pH 8.5	2 M urea, 25% ACN	74	69	8	1.2	2	trace
200 mM Tris, pH 8.5	25 % ACN	78	17	14	7.5	6	5.5
200 mM Tris, pH 8.5	37.5% ACN	77	72	6	2	2.5	2.5

Figure 5.4.10.1 – Optimum conditions for maximal generation of authentic IGF-I from Met-pGH(1-11)VN-WGSGPAPM-IGF-I.



Peak	R.T.	Area
1	21.6	1210
2	22.5	1090
3	23.3	1850
4	24.0	890
5	25.2	730
6	29.2	9900

Table 5.4.10.4 – Typical recoveries of Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 2 (steps 1 and 2) or authentic IGF-I (step 3). The theoretical yield of a continuous process is calculated by compounding the losses experienced at each step.

Step	Processing step	Step yield	Theoretical process yield
1	IB Dissolution to Isomer 2 in Refold	62%	62%
2	Refold to pure Isomer 2	67%	41.5%
3	Purified IGF-I from large scale cleavage	6.5%	2.7%

5.4.11 L6 myoblast protein synthesis

The biological potency of the two folding isomers of Met-pGH(1-11)VN-WGSGPAPM-IGF-I and the various cleavage product forms was examined by their ability to stimulate protein synthesis in rat L6 myoblasts. The dose-response curves for growth factors Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 1, isomer 2, WGSGPAPM-IGF-I, des(68-70)WGSGPAPMIGF-I, des(68-70)IGF-I, proteolytically generated authentic IGF-I and commercial IGF-I reference standard appear as Figures 5.4.11.1 through 5.4.11.7 respectively. The percent stimulation expressed is the increase in ^{3}H-Leucine uptake over the negative control, which lacked any exogenous growth factors or serum. Isomer 1 was clearly less potent than isomer 2 or any of the cleavage products. The proteolytically generated IGF-I was observed to be bioequivalent to the commercial reference standard IGF-I. The full-length isomer 2 fusion protein and all its derived cleavage products were of equivalent activity relative to each other at a slightly lower potency than that of authentic IGF-I. For comparison, the ED_{50} values and 95% confidence intervals of the data are provided in Table 5.4.11.1.

Figure 5.4.11.1 - Stimulation of protein synthesis in L6 myoblasts by Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 1

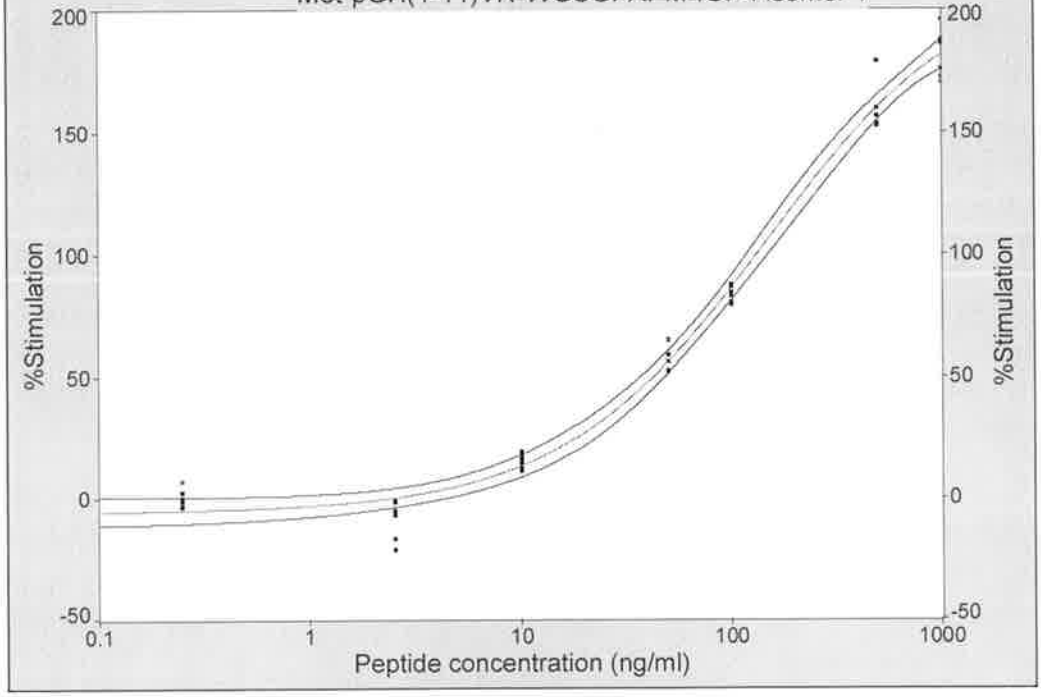


Figure 5.4.11.2 - Stimulation of protein synthesis in L6 myoblasts by Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 2

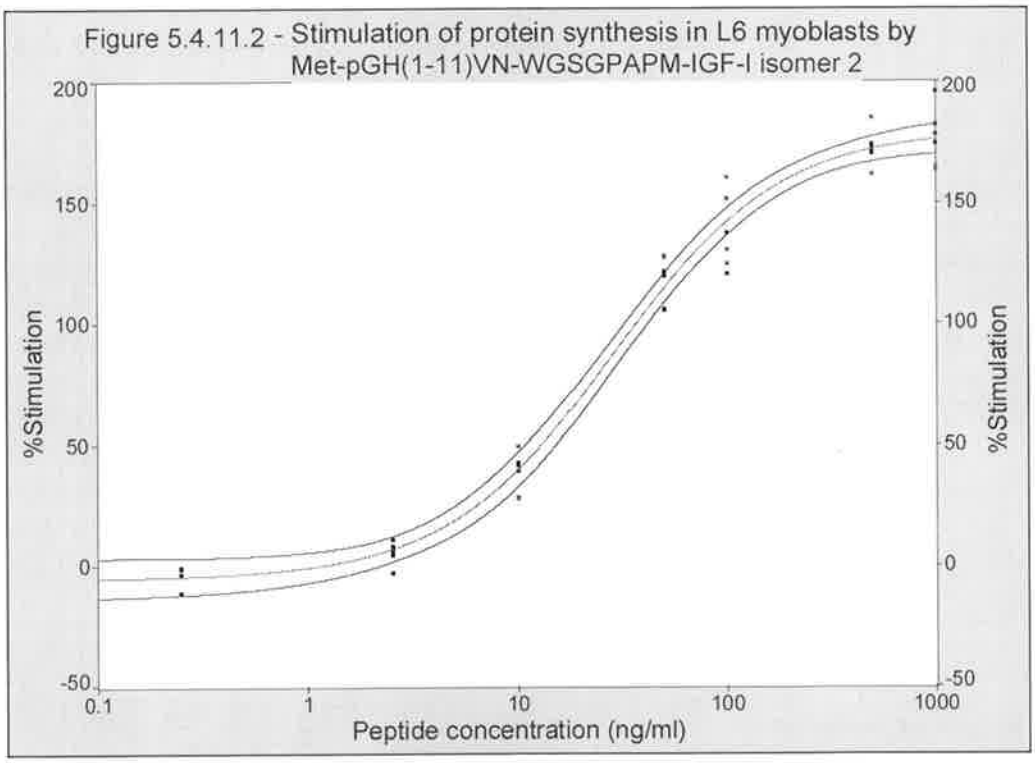


Figure 5.4.11.3 - Stimulation of protein synthesis in L6 myoblasts by WGSGIGF-I

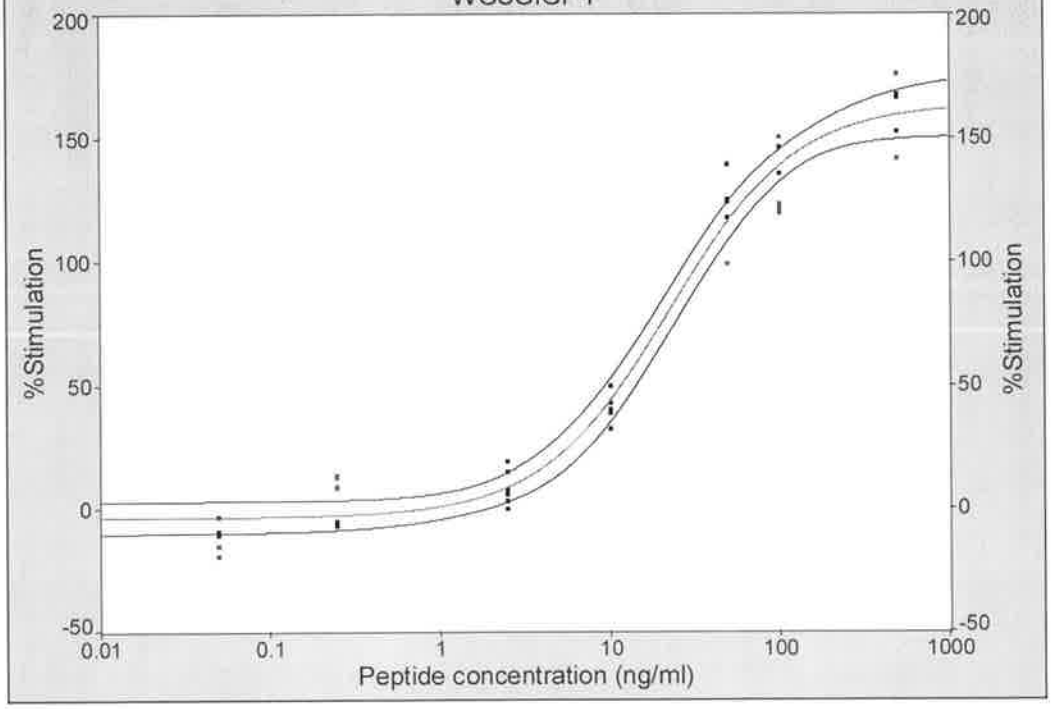


Figure 5.4.11.4 - Stimulation of protein synthesis in L6 myoblasts by des(68-70) WGSGIGF-I

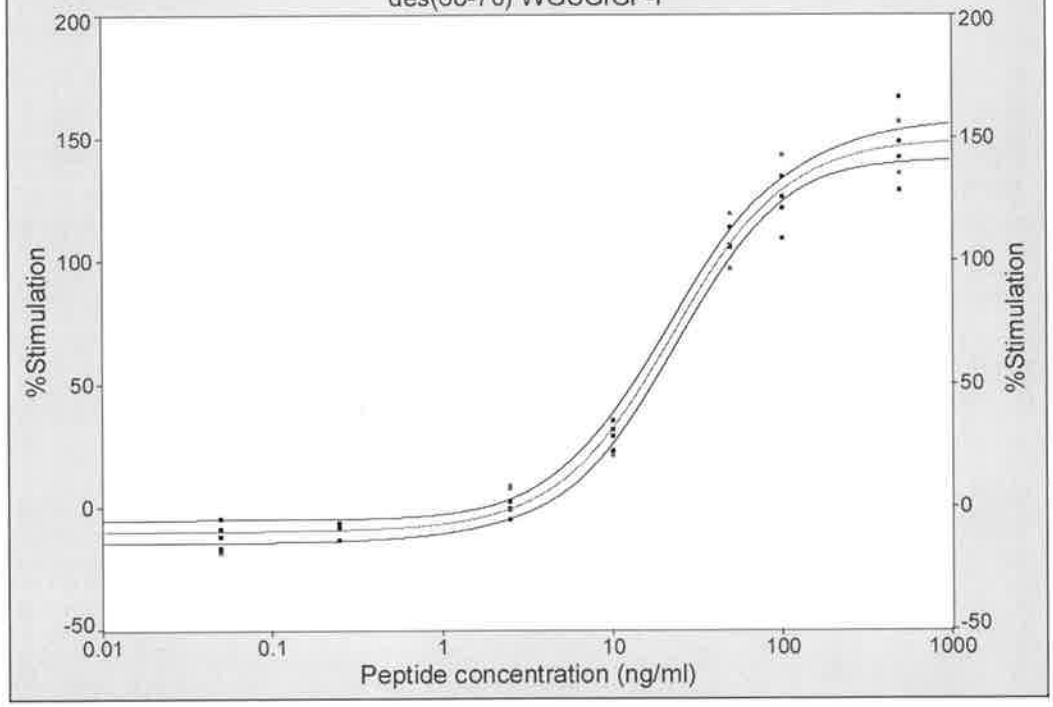


Figure 5.4.11.5 - Stimulation of protein synthesis in L6 myoblasts by des(68-70) IGF-I

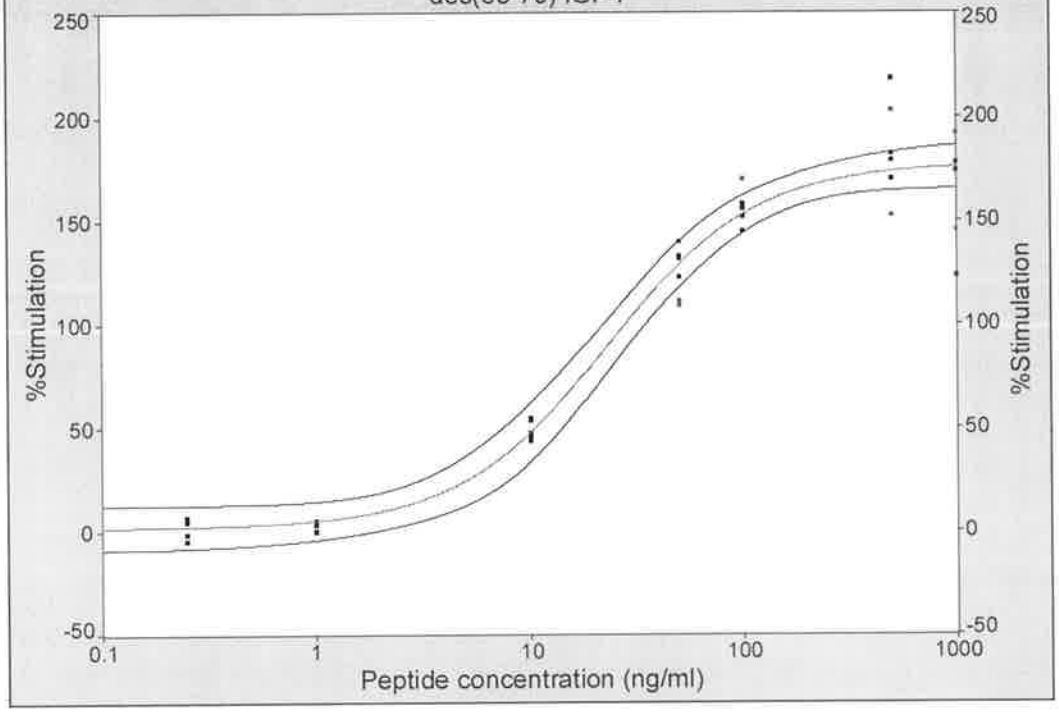
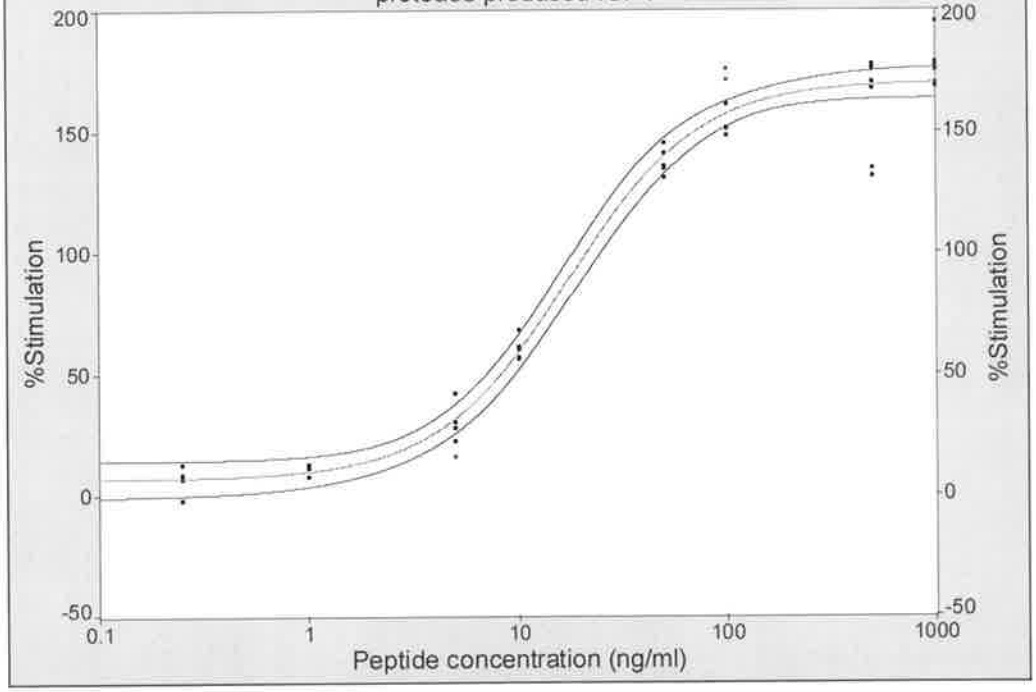


Figure 5.4.11.6 - Stimulation of protein synthesis in L6 myoblasts by protease produced IGF-I



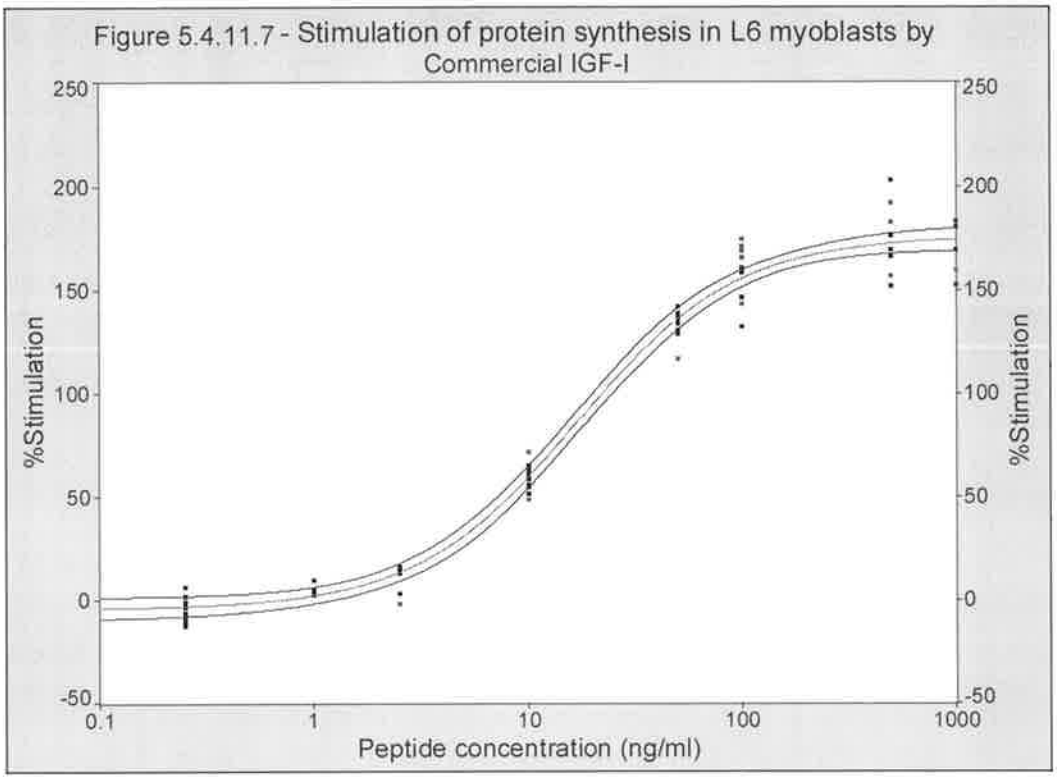


Table 5.4.11.1 – Concentrations of Met-pGH(1-11)VN-WGSGPAPM-IGF-I folding isomers, cleavage intermediates or products required for half maximal stimulation of protein synthesis in rat L6 myoblast cells.

Construct	**ED_{50} (ng/ml)**	**95% confidence interval (ng/ml)**
Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 1	146.8	93.8–199.8
Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 2	28.7	23.3–34.1
WGSGPAPM-IGF-I	22.7	16.5–28.8
des(68-70) WGSGPAPM-IGF-I	23.0	18.9–27.0
des(68-70)IGF-I	23.1	16.2–30.0
protease produced IGF-I	17.5	14.5–20.8
commercial IGF-I	16.9	14.5–19.2

5.5 DISCUSSION

The generation of novel variants of the Met-pGH(1-11)VN leader peptide has been investigated in this chapter, and these variants tested for their ability to yield authentic IGF-I upon digestion by the PragA9 mutant of α-Lytic protease.

5.5.1 Expression

Preliminary production of the constructs showed expression to be quite leaky, with material that interacts with anti-IGF-I antibody present prior to induction. This phenomenon was resolved empirically by protein expression in richer defined medium, as shown in a parallel induction experiment. It could be that the lower concentration of glucose in the minimal medium (MinA) of the induction trials triggered a starvation response in the cells, resulting in the upregulation of all carbon utilising metabolic pathways, including the *lac* system on which expression in this plasmid is based. In the richer FM11 medium, the cell would be receiving an abundance of glucose and as such the induction control was able to be imposed.

5.5.2 Folding

The refolding of the variants was successful using the established methods, with the characteristic two major folding isomers clearly visible in the refolding reactions of the leader peptide charge mutant (Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I; Figure 5.4.4.1) and the insertional mutant (Met-pGH(1-11)VN-WGSGPAPM-IGF-I; Figure 5.4.5.1). Given the novelty of these

two variants, the assignment of "native" and "alternate" isomer peaks could not be conclusively made until biological activity data was available (see later); but in all previous examples of human IGF-I refolding reactions that yield the canonical two folding isomers, the native isomer elutes later on rp-HPLC than the alternate isomer (Hejnaes *et al.,* 1992; Miller *et al.,* 1993; Samuelsson *et al.,* 1994; Milner *et al.,* 1995; Kim and Lee, 1996). However, no assumption regarding the identity of either isomer was made, as presumably only the correctly folded isomer will allow generation of authentic IGF-I. Cleavage at the correct site in the alternate isomer will simply result in the generation of alternative isomer IGF-I in the absence of any rearrangement of disulphide bonds.

The abundance of the two isomers was roughly 1:1 (±10%), in contrast to ratios of 2.5:1 native:alternate isomer that have been reported for refolding reactions of Met-pGH(1-11)VN-IGF-I (Milner *et al.,* 1995). Comparable yields to the latter were not attained in the unoptimised refolding reactions of the Met-pGH(1-11)VN-PAPM-IGF-I construct parental to those generated here. In the refolding of this construct the isomeric abundance is much closer to 1:1 (see Figure 4.4.1.1), and as this construct is the parent of the two novel constructs of this chapter, it serves as a more appropriate basis for comparison. The ratio between the folding isomers in the parental Met-pGH(1-11)VN-PAPM-IGF-I and the novel variants is unchanged by the above mutations.

Most interesting is the lack of an effect on folding by the alteration of net charge of the leader peptide in the Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I variant. This suggests that it is not the hydrophobicity of the leader

alone that provides its assistance to refolding. Milner *et al.* (1995) proposed that the presence of the leader sterically prevents the formation of a salt bridge between Glu^{3} and Arg^{56} early in the refolding pathway. Moreover, it was proposed that such a salt bridge, which is not observed in the mature folded protein, may form early in the folding of the denatured polypeptide, stabilising the structure sufficiently to enable the incorrect disulphide bond of the alternate isomer to form.

5.5.3 Selection rationale

The only assay available for preliminary screening of a successful cleavage reaction was HPLC, validated by comparison with the retention time of authentic IGF-I. However, as observed in Chapter 4, generation of peaks with identical retention times did not always translate into an identical sequence.

HPLC analysis of the preliminary cleavage reactions revealed that, for both constructs, only reactions of isomer 2 contained any material of IGF-I-like retention time. Comparison of the productivity of the two constructs showed that Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I isomer 2 generated 9% of the total peak area as putative authentic IGF-I (Figure 5.4.6.5), while the Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 2 produced 12% of its total products as putative authentic IGF-I (Figure 5.4.7.5). The presence of residual full-length fusion protein substrate in the initial screening reactions of Met-pGH(1-11)VN-WGSGPAPM-IGF-I suggests that this reaction has not reached completion under these conditions. This provides scope to further increase the yield of putative authentic IGF-I generated from this construct.

5.5.4 *Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I isomer 1 cleavage*

Published evidence and preliminary cleavage data suggest that isomer 2 represents the correctly folded protein. Some interesting insight into the structure of the cleavage site of the Met-pGH(1-11)VN-IGF-I constructs can be gained by consideration of the cleavage reactions of Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I isomer 1. Comparison of the retention time of a purified alternate folding isomer of authentic IGF-I (Figure 5.4.6.3) with the products of these cleavage reactions (Figures 5.4.6.1 and 5.4.6.2) supports the formation of alternate isomer in these reactions. Most striking is the low abundance of this peak compared to the supposed correctly folded IGF-I generated in the cleavage reactions of isomer 2. An explanation for this is gained by examination of the structure of the alternate isomer of IGF-I, where the N-terminus is shifted toward the interior of the molecule (Sato *et al.*, 2000; Figure 5.5.4.2). If this occurred in the Met-pGH(1-11)VN-IGF-I constructs, it is conceivable that the structural factors already reducing accessibility to the cleavage site would be compounded. It is also interesting that the cleavage reactions of the alternate folding isomer produce more products than those of the correct isomer. Again, the structure of the misfolded species provides a possible explanation, where the reduced helical content could allow for more potentially protease-labile sites in the greater unstructured or β-sheet conformations than are presented in the correctly folded molecule. The further decreased access to the preferred motif could thus promote cleavage at secondary sites.

Figure 5.5.4.1 – The solution structure of authentic human IGF-I of Sato *et al.* (2000). The N and C-termini are labelled accordingly and shown to be free and quite solvent exposed.

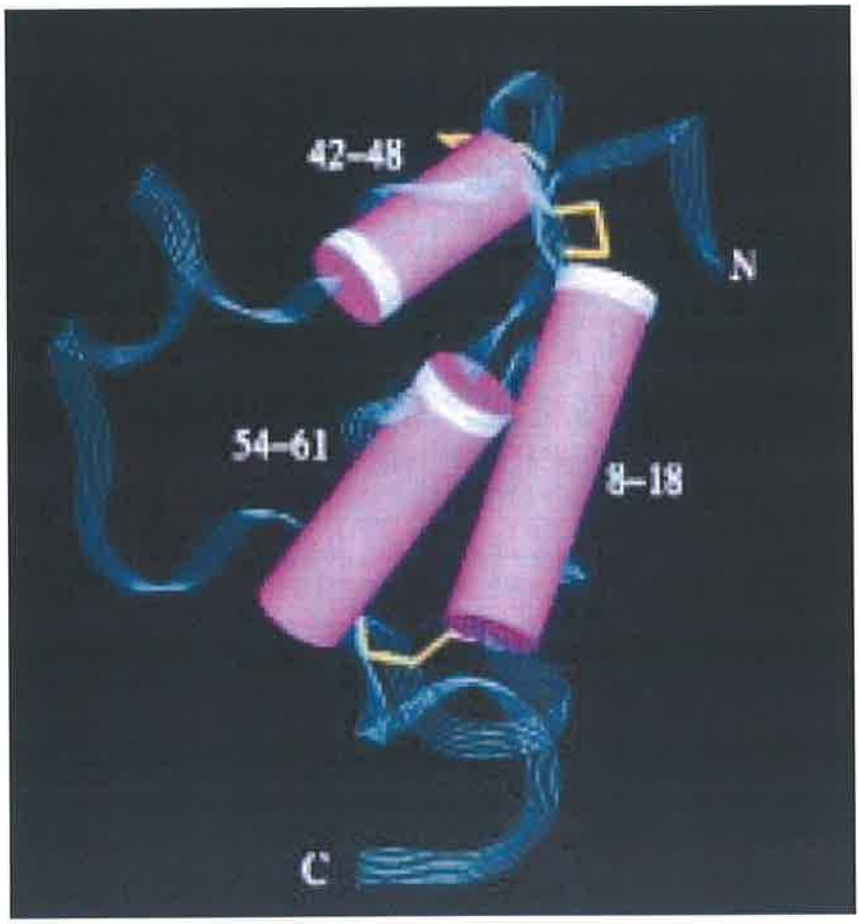
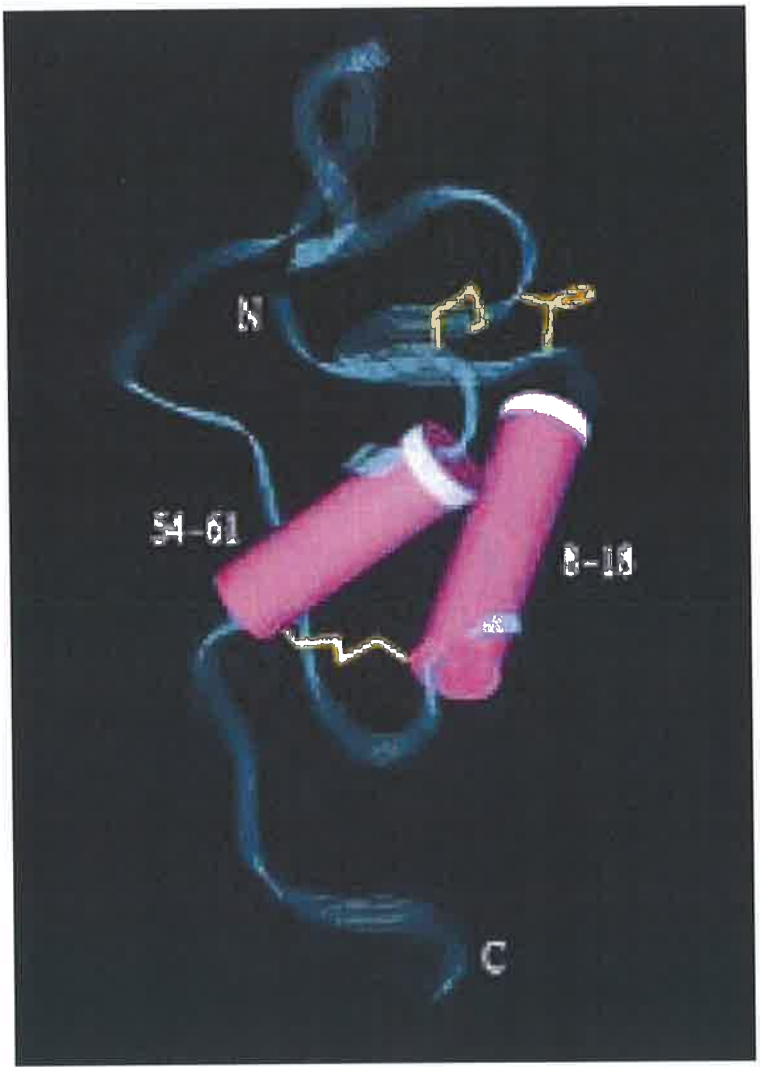


Figure 5.5.4.2 – The solution structure of misfolded human IGF-I of Sato *et al.* (2000). The N and C-termini are labelled accordingly, with the N-terminus observed to be somewhat buried within the structure.



5.5.5 *Met-pGH(1-11)VN-WGSGPAPM-IGF-I isomer 2 cleavage*

Four major peak species (1, 2, 3 and 4) are most frequently observed in cleavages of isomer 2 of Met-pGH(1-11)VN-WGSGPAPM-IGF-I. These were purified to single peaks (Figure 5.4.8.1) and analysed by mass spectrometry to determine their identities, which were, in order of increasing retention time:

(1) 7647.9 Da (Figure 5.4.9.2).

(2) 7361.7 Da (Figure 5.4.9.3).

(3) 8432.1 Da (Figure 5.4.9.4).

(4) 8144.1 Da (Figure 5.4.9.5).

Multiple masses, each separated by intervals of approximately 22 Da, are observed in the spectra which most likely correspond to sodium adducts of the main species. The mass of species (1) agrees quite favourably with authentic IGF-I (theoretical mass 7648.7 Da), (2) with authentic IGF-I lacking three residues from the C-terminus (theoretical mass 7632.4 Da), (3) with a cleavage product having only the Met-pGH(1-11)VN leader peptide excised, *i.e.* WGSGPAPM-IGF-I (theoretical mass 8432.6 Da) and (4) the latter product again but lacking three residues from its C-terminus (theoretical mass 8146.3 Da). With the exception of the final product, the detected masses were consistently within one Dalton of the theoretical, providing confidence in the identity assignments.

5.5.6 *Reaction pathway*

Timecourse analysis of the cleavage reaction of Met-pGH(1-11)VN-WGSGPAPM-IGF-I suggests a product generation pathway (Table 5.4.10.2). The first product is formed by cleavage at the end of the leader peptide (VN) as previously observed in Chapter 4, but in this case WGSGPAPM-IGF-I is liberated. This is followed by cleavage at the intended cleavage site to liberate authentic IGF-I. This two step reaction "pathway" is strongly suggested by the observation that authentic IGF-I is never detected in the absence of the WGSGPAPM-IGF-I form, whereas this intermediate is

generated comparatively rapidly, and is present at 5 minutes in the timecourse of Table 5.4.10.2.

Despite the disruption of the incorrect cleavage site that confounded the outcome in Chapter 4, cleavage after Val-Asn still occurs in this novel construct and appears to be a prerequisite for cleavage at the correct site. This would imply that the excision of the leader provides accessibility to the cleavage site, as suggested in Chapter 4, and that the insertion of a four residue spacer provides insufficient exposure of this steric constraint to promote cleavage at the intended site in the first instance. Where the construct of this chapter differs from that of Chapter 4 is that the resulting product of leader excision is able to be further cleaved to liberate authentic IGF-I. The PAPM-IGF-I generated in Chapter 4 was an N-terminal reaction endpoint product, unable to be further processed by α-Lytic protease. The residual four amino acid N-terminal extension of that product is apparently too short to be cleaved by this protease. By comparison, the residual eight amino acid N-terminal extension of the WGSGPAPM-IGF-I intermediate is not a terminal product, with the additional length of the remaining cleavage site now providing sufficient continuous sequence for α-Lytic protease binding and cleavage. This provides some support in a protein substrate of the peptide model findings of Bauer *et al.* (1981), which suggests substrate contacts up to at least a P_5 residue.

5.5.7 des(68-70) truncations

At any time during the above reaction "pathway" the C-terminal truncation mentioned earlier can occur, independent of the formation of any

precursor species. There is a suggestion that this product may appear in the initial full-length isomer 2 substrate, where a second major mass species is present at 9589 Da (Figure 5.4.9.1). This is only 7 Da removed from the expected mass of the full-length fusion protein possessing this C-terminal truncation. This product is not the result of contaminating host-derived protease activity as it does not increase over time in control reactions lacking added exogenous protease. Furthermore, the prevalence of such truncations is certainly exacerbated by incubation with α-Lytic protease (see Tables 5.4.10.1 to 5.4.10.3) suggesting that this truncation is α-Lytic protease derived.

Examination of the amino acid sequence around this secondary cleavage site does not reveal any specific barrier to α-Lytic protease cleavage. The upstream 4 residues, Leu-Lys-Pro-Ala, are neither exceptionally favourable, nor forbidden as the cognate P_4 to P_1 substrate positions (Lien *et al.,* 2003). Moreover, the three excised residues, Lys-Ser-Ala, could be considered more favourable as P_1' to P_3' amino acids than Gly-Pro-Glu with over 5- and 4.5-fold increased activity respectively, for Lys over Gly at P_1' and Ser over Pro at P_2' (Schellenberger *et al.,* 1994).

Cleavage at this C-terminal site must, however, be far less favourable than cleavages at either of the sites that give rise to the two reaction pathway products, otherwise correct IGF-I would never be observed, with all available authentic C-termini first being clipped. It does appear though, that given sufficient time authentic IGF-I will eventually undergo this C-terminal clip; as suggested by the 16 hr timepoint of Table 5.4.10.2, making the C-terminally

truncated IGF-I, or "des(68-70)IGF-I" the ultimate end-point product of the reaction.

A limited series of optimisation reactions were undertaken in an attempt to reduce the generation of des(68-70)IGF-I, and increase yields of authentic IGF-I (Tables 5.4.10.1 and 5.4.10.3). Although the tendency of PragA9 to perform the C-terminal clip could not be significantly altered, optimisation of the enzyme:substrate ratio did result in a significant improvement in the yield of authentic IGF-I. Thus at an enzyme:substrate ratio of 1:30, authentic IGF-I represented 20% of all proteolytic products.

5.5.8 Protein synthesis

Bioassay by stimulation of rat L6 myoblast protein synthesis confirmed the identities of isomer 2 as the correct fold and isomer 1 as the alternate fold, with the latter giving 5-fold lower activity (Table 5.4.11.1; Figures 5.4.11.1 and 5.4.11.2). There is reasonable agreement in the ED_{50} of the biological responses and the proposed structure of all cleavage products. The considerable overlap in the 95% confidence intervals of the ED_{50} data suggests no significant difference in their biological activity (Table 5.4.11.1). This is particularly encouraging for the protease derived IGF-I as compared to the commercial IGF-I derived by chemical cleavage. The ED_{50} of both of these fit well within the range of published activities for human IGF-I in the L6 myoblast assay of 8.4 ng/ml (Francis *et al.,* 1992), 13 ng/ml (Bagley *et al.,* 1989), 15.8 ng/ml (King *et al.,* 1992) and 18 ng/ml (Upton *et al.,* 1997), lending support to the findings.

The loss of the three C-terminal residues that gives rise to des(68-70)IGF-I had little effect on the biological potency of the protein. This is to be expected, as the involvement of the C-terminus of the IGF's is not implicated in any of the modulators of IGF-I bioactivity, such as binding protein recognition or receptor affinity (see below). The relatively minor contribution of the C-terminal three residues to biological activity in this assay system, is also observed in the intermediate cleavage products; WGSGPAPM-IGF-I and its des(68-70) counterpart, with ED_{50} values almost equivalent at 22.7 ng and 23.0 ng respectively.

The biological potency of the full-length fusion protein appears to be slightly less than that of authentic IGF-I, at around 1.7-fold lower ED_{50}. This is less than anticipated, as the biological activity of Met-pGH(1-11)VN-IGF-I is exactly equal to that of authentic IGF-I in Milner *et al.* (1995), and 1.4-fold greater in Francis *et al.* (1992), meaning that Met-pGH(1-11)VN-WGSGPAPM-IGF-I is 1.7 to 2.5-fold less active than the parental Met-pGH(1-11)VN-IGF-I. Given that WGSGPAPM-IGF-I is only 1.3-fold less active than authentic IGF-I, it appears then that any reduced potency of the full-length Met-pGH(1-11)VN-WGSGPAPM-IGF-I fusion protein is possibly due to the presence of this extra sequence.

The magnitude of this difference between isomers 1 and 2 of the full-length fusion protein falls within the range of biological activity difference between alternate and native isomers of IGF-I analogues reported by Milner *et al.* (1995). In that investigation the reduction in biological potency of alternate isomers ranged from approximately 3-fold for Met-pGH(1-11)VN-Arg^3-IGF-I isomer to many tens of folds for Met-pGH(1-11)VN-IGF-I isomer.

The differential potency of other IGF-I mutants folding isomers were reported falling between these two extremes. This non-uniformity in the activity difference between isoforms of other Met-pGH(1-11)VN-IGF-I mutants highlights the unpredictability of the biological response that can arise in novel IGF-I structures. The reduction in biological potency of misfolded isomers is a product of the affinity for both the inhibitory IGF-binding proteins and the growth stimulating type 1 receptor. For example, misfolded Met-pGH(1-11)VN-IGF-I has 22 and 30-fold lower affinity for binding proteins and receptor respectively than the correct isomer and this is the approximate magnitude of the reduction in potency of this protein. Similarly, for misfolded Met-pGH(1-11)VN-Arg^3-IGF-I the decreases in affinities were 2.4 and 3.8-fold compared to its cognate correct isomer, with a comparable low unit fold reduction in biological potency. The relative contribution of the altered affinity for IGF-I-binding proteins and the type 1 IGF receptor in the observed biological activity of the IGF-I mutants and their isomers reported here is unclear.

Interestingly, the presence of the Met-pGH(1-11)VN leader seems to reduce the "penalty" of misfolding on receptor and binding protein recognition affinity and biological potency. In comparison to the above data, misfolded IGF-I has 263 and 176-fold lower affinity for binding proteins and the type 1 IGF receptor respectively than the correct isomer (compared to 22 and 30-fold). This results in a concomitantly large decrease in biological potency, with the alternative isomer having only approximately 1% of the biological potency of the native form (Milner *et al.,* 1995).

5.5.9 *Met-pGH(1-11)VN-[F^{16}E/M^{13}E]-PAPM-IGF-I*

With knowledge of the mass identities corresponding to the four HPLC peaks detected in the cleavage reactions of Met-pGH(1-11)VN-WGSGPAPM-IGF-I (see Section 5.4.9 and Figure 5.4.8.1), further examination of the cleavage reaction of Met-pGH(1-11)VN-[F^{-16}E/M^{-13}E]-PAPM-IGF-I isomer 2 (Figure 5.4.6.5) becomes possible. It appears that the retention time (22.6 min) of the main peak in the cleavages of this construct corresponds very closely with that of des(68-70)IGF-I (typically 22.4 to 22.5 minutes, Figures 5.4.7.2 and 5.4.10.1). If this is borne out by subsequent analyses, it would appear that this construct was cleaved to the terminal product whereas Met-pGH(1-11)VN-WGSGPAPM-IGF-I cleavage was incomplete. This is supported by the greater abundance of the putative des(68-70)IGF-I product in the presence of chaotrope which could favour greater cleavage of a more open molecule (Figures 4.1.1, 5.5.4.1).

Although further characterisation of the cleavage reaction dynamics of the Met-pGH(1-11)VN-[F^{-16}E/M^{-13}E]-PAPM-IGF-I construct were not pursued, it appears from the preliminary screening reactions that production of IGF-I did not require the prior excision of the leader. Although inclusion of urea promoted formation of the putative des(68-70)IGF-I end-product, it was not absolutely required for the generation of either that species nor authentic IGF-I. Both of these outcomes seem to suggest that alteration of the net charge of the leader peptide caused the leader to relax its fold-back configuration from proximal to Gly^1, Pro^2 and Glu^3, and extend into solution. Moreover, this effect does not require the added presence of chaotrope for it to occur. As seen in Chapter 4, once the Met-pGH(1-11)VN leader has been

excised, no further α-Lytic protease action at the N-terminus is observed. This is in agreement with the suggestion that the residual four amino acid motif is too short for protease binding in a folded protein. Any authentic IGF-I N-termini that are generated from this fusion protein construct must therefore be produced in a single step, as the cleavage motif itself is unaltered from that of the parent Met-pGH(1-11)VN-PAPM-IGF-I of Chapter 4.

5.5.10 Process yields

The projected process yield to obtain highly purified IGF-I by this approach is 2.7%. This is 5.5-fold improved over the published value for hydroxylamine cleavage of Met-pGH and IGF-I fusions (King *et al.*, 1992). More recent improvements in the hydroxylamine cleavage of other IGF-I fusion systems have allowed process yields of 8% to be realised (Antorini *et al.*, 1997). However, this outcome was itself the result of many years of research into the production of clinical material.

The challenging steps determining final yield in the process developed in this chapter are the cleavage of the fusion and final purification of IGF-I. As IGF-I represents 20% of the cleavage products, further losses are experienced in the purification of the IGF-I from other cleavage products. However, the losses during purification at small scale as in this chapter, including losses due to adhesion to collection vessels and FPLC system hold-ups will not be significant at scale. For example, if the yield from the final purification could be improved by as little as 50% it would translate to a final theoretical process yield of over 4%.

5.6 CONCLUSION

The work of this chapter has demonstrated that it is the structure adopted by Met-pGH(1-11)VN-IGF-I that is refractory to cleavage by α-Lytic protease, and not the permanent obstacle of an unfavourable P_1' to P_3' sequence. Successful generation of significant levels of authentic human IGF-I are possible by manipulations that modify these structures upstream of the protease cleavage site. It has also been demonstrated that such changes to the composition of the Met-pGH(1-11)VN leader sequence are not detrimental to the recombinant expression or refolding of human IGF-I. Authentic human IGF-I has been generated by site-specific proteolysis of a Met-pGH(1-11)VN-IGF-I analogue in an industrially applicable manner for the first time. Moreover, this product has been shown to be bioequivalent to its chemical cleavage derived commercial counterpart.

6.1 INTRODUCTION

Advances in recombinant DNA technology have made the fusion of unrelated proteins a relatively straightforward undertaking. As discussed in Chapter 1, certain types of such fusions have been shown to be very beneficial to the production of heterologous proteins in *E. coli*. One very successful fusion protein system uses methionyl porcine growth hormone truncations to aid in the expression and refolding of Insulin-like Growth Factors. Once the non-target protein material has served its purpose it is often desirable to remove it, wherein lies what has been referred to as the Achilles' heel of the fusion protein strategy (Waugh, 2005). Many strategies exist for the site-specific cleavage of fusion proteins, but none is universally applicable to all circumstances.

The application of proteases to achieve a highly specific separation of the fusion protein partners under mild reaction conditions has been widely applied with considerable success. α-Lytic protease bears many of the attributes that would render it highly desirable for industrial-scale proteolysis, but the preference of this wild-type isoform for cleavage after very common amino acids precludes such use. With combinatorial S_1 active site mutagenesis, α-Lytic protease has undergone a shift toward more useful specificity, with some mutants having many fewer common residues as their primary recognition determinants (Graham *et al.,* 1993; Graham *et al.,* 1994). One mutant, PragA9, has been demonstrated to be highly effective at the cleavage of Met-pGH(1-11)VN fusion proteins, including human IGF-II (Lien *et al.,* 2001), des(1-3)-IGF-I and Transforming Growth Factor β3 (G. Francis, personal communications). Conspicuous by its absence from these examples

is IGF-I, given it is expressed by this fusion system (King *et al.,* 1992; Francis *et al.,* 1992). In fact, examples of proteolytic release of this protein by any protease in any fusion system remain extremely rare, and particularly any applicable to industrial or biopharmaceutical manufacture. With the status of IGF-I as the sole member of the IGF family to represent FDA approved biopharmaceuticals (Increlex™ from Tercica, Inc., Brisbane, CA, USA and Iplex™ from Insmed Inc., Richmond, VA, USA), improvements in the production of this growth factor could be of significant industrial merit.

Despite the utility of mutants of α-Lytic protease in other fusion protein cleavages, both the available literature and some preliminary work suggested that it was unable to process IGF-I fusions. This could by attributed primarily to the N-terminal tripeptide of IGF-I, Gly-Pro-Glu, where each residue is individually unfavourable at the corresponding P_1' to P_3' positions in α-Lytic protease substrates (Schellenberger *et al.,* 1994). The dearth of knowledge available regarding either the actual importance of S' binding relative to that of the better characterised S subsites, or the impact of substrate structure to α-Lytic protease catalysis presented a challenge to further improvements. Initially a screening system was developed to query all the combinatorial mutants of α-Lytic protease for the ability to cleave at an optimised motif linked to this refractory sequence. Then the strategic aim was to exploiting variability in S' specificity to achieve the desired outcome, as hinted at in Graham *et al.,* (1993).

The first step in the overall strategy was to verify the activity of the mutants that had been stored frozen for some years since last cultured. This was achieved by culture of the mutant vector bearing *E. coli* clones on skim

milk agar plates. A cleared halo surrounding the colony indicated cleavage of the whole casein substrate in these plates, demonstrating viable host secretion of the given protease. It was also clear from both these demonstrations and the background literature (Graham *et al.,* 1993; Graham *et al.,* 1994) that the mutants displayed either highly variable expression levels, differences in catalytic efficiency, or both. A laborious and qualitative assay such as skim milk agar clearance would not be suitable for normalising the enzymatic activity present in the protease bearing isolates required for subsequent steps. Instead, a quantitative fluorescence polarisation assay was developed and applied against all mutants. This assay also narrowed the available substrates, without undue bias, as only one of the casein isoforms (β-casein) was used, whereas this constitutes approx 33% of whole casein (Ekstrand and Larsson-Raznikiewicz, 1978).

The potential substrates provided to the proteases were further constrained in the subsequent assay of the screening campaign. End application-oriented peptide substrate mini-libraries were used to screen the mutants in another fluorescence polarisation assay. The results of this step were difficult to interpret (see Section 6.2), but it appeared that many of the mutants were cleaving efficiently at the motif of interest. A shortlist of mutants was selected based on both the assay data and knowledge of the mutations present. The final step in the screening strategy examined the action of these shortlisted mutants against a series of specifically designed peptide substrates. The outcome of this final analysis revealed a single mutant protease that was able to specifically cleave at the required peptide bond. This screening regime could conceivably be applied to any group of proteases

to investigate the cleavage of a given fusion protein, provided there was some knowledge of the upstream cleavage requirements of protease.

With a lead mutant protease selected from the libraries, cleavage of the full-length fusion protein was attempted. The preferred recognition sequence identified from the discrete peptide screening assay was incorporated into the product fusion protein. The ability to liberate some authentic IGF-I from this approach was well below the efficiency required for a viable industrial process. Moreover, the protease was surprisingly observed predominantly cleave immediately before the inserted recognition motif, liberating PAPM-IGF-I.

Alteration of the leader peptide's properties and manipulation of the site of cleavage were performed. Foremost, in both occasions this lead to the successful generation of authentic IGF-I, but also other side products. However, time constraints allowed the investigation of only one construct in detail, and in that example the cleavage at the correct site appeared to follow a two step pathway. This is not without precedent, with two different cleavage mechanisms (H64A subtilisin then hydroxylamine) used sequentially to liberate hagfish IGF from a Met-pGH fusion in the past (Upton *et al.,* 1997). The stepwise cleavage of the Met-pGH(1-11)VN-WGSGPAPM-IGF-I fusion parallels the earlier cleavage of the β-casein substrate in Chapter 2, where cleavage by PragA9 appeared to occur in a stepwise fashion.

The above culminated in the successful production of authentic human IGF-I, the identity of which was verified by equivalent mass spectrometry and biological activity compared to reference IGF-I. The yield of the non-optimised process, from bacterial expression to purified authentic IGF-I

was 2.7%. This represents a 5.5-fold improvement over the published hydroxylamine cleavage strategy (King *et al.,* 1992).

6.2 SUGGESTED IMPROVEMENTS TO THE STRATEGY

Despite the ultimate success in the isolation of a mutant α-Lytic protease that is able to liberate authentic IGF-I from this fusion protein, the existence of other, better suited enzymes in the library cannot be ruled out. Nor can this successful outcome be directly attributed to the screening system used. Although PragA9 was observed to have some activity against the GPE-motif peptide group, it was shortlisted as both an example of a TLG motif favouring mutant and as a known site-specific protease.

The "weak link" in the screening strategy is without question the fluorescence polarisation peptide assay. The large number of positive hits generated in this step (Chapter 3) was entirely unanticipated, given the available background literature on the requirements of the S' subsites of α-Lytic protease. Not only did almost all mutants queried appear to show some ability to cleave the Gly-Pro-Glu P' motif bearing peptide group, approximately one third appeared to favour this motif over the positive control. The low dynamic range in which the assay operated could be partly responsible for this lack of discriminating power, with only 25 – 30% difference between absolute negative and positive readings (Figure 3.4.2.1).

A lack of cleavage of the FP peptide groups may be responsible. The use of the shortest possible peptide motif in the assay was intended to prevent false positives due to cleavage at unexpected locations. However, as seen in Chapter 4 the tetrapeptide Pro-Ala-Pro-Met was insufficient for

cleavage by at least PragA9 against protein substrates, where PAPM-IGF-I was shown to be an endpoint product. This situation could even be compounded by the presence of the biotin moiety immediately preceding the cleavage motif, by potentially offering steric interference to an already unfavourable cleavage situation. If the effect observed for PragA9 against Met-pGH(1-11)VN-PAPM-IGF-I mentioned above were mirrored in the other mutants it could partly explain the poor selectivity of the FP assay peptide. Even the relatively specific PragA9 mutant has been observed to act at some non-target sequences. This is illustrated by the generation of unidentified products in some discrete peptide reactions in Chapter 3, cleavage at the LFVN motif in two separate constructs (Met-pGH(1-11)VN-PAPM-IGF-I and Met-pGH(1-11)VN-WGSGPAPM-IGF-I) and the generation of the C-terminally clipped IGF-I forms in Chapter 5.

At face value, one solution would seem to be the use of a peptide of increased length. It is known that α-Lytic protease makes substrate contacts extending out to at least P_5 (Bauer *et al.,* 1981). Moreover, the PragA9 mutant cannot remove a tetrapeptide motif in a fusion protein, at least when it precedes a Gly-Pro-Glu sequence. However, this is unlikely to have resolved the issue faced here, as the longer discrete peptides were cleaved non-specifically in most cases by the shortlisted mutants (Chapter 3). Hence the problem of false positives arising from the cleavage of longer peptide substrates would have remained.

The problem with the assay may reside in the lower fluorescence polarisation signal resulting from the binding of the peptide to avidin. Interference with either the binding interaction or the polarisation signal may

have occurred. The proximity of the biotin moiety to the fluorophore (separated by 7 residues) may have sterically impeded binding to the peptide. A previous example used a nonapeptide with an additional γ-aminobutyric acid spacer between the biotin and the fluorophore (Levine *et al.,* 1997). Therefore, increasing the length of the peptide may help to improve the assay system, but it may be at the cost of increased false-positive hits. Alternatively, an aliphatic carbon chain spacer arm could relieve steric inhibition of proteolysis, without adding extra cleavage sites.

A solution to this issue in the assay system would be to avoid any such post-cleavage complications by using a fluorescent resonance energy transfer system, where the cleavage event itself triggers the signal response. A FRET peptide, modelled after the discrete peptides, could be used in a screening assay, and then any hits could be confirmed by rp-HPLC. A FRET peptide also need not necessarily display a longer target sequence than the 7-mer used in the FP assay; endoproteases cleavage motifs in peptides as short as 6 residues have been used in FRET assays (Tanskul *et al.,* 2003; Gouvea *et al.,* 2007). Should steric inhibition due to the fluorophore or quencher be detected, a carbon chain spacer could be used as stated above.

6.3 OTHER OBSERVATIONS

6.3.1 *Cleavage at P_1' Proline*

Although not the main focus of this work, examination of some of the unexpected cleavage sites has provided insight into the substrate recognition

by PragA9 α-Lytic protease. This points to potential avenues of investigation to understand the substrate specificity of this useful protease.

The unique properties of PragA9 α-Lytic protease are borne out by its ability to cleave before a Pro residue. This is unusual amongst serine proteases, yet has been observed here with several protein substrates (Chapter 4). The trypsin or chymotrypsin fold endoproteases discussed throughout this work have all been shown to avoid substrates bearing a Pro residue at P_1'. Chymotrypsin and Trypsin themselves have undetectable activity against P_1' Pro in acyl-transfer reactions (Schellenberger *et al.*, 1994). Likewise, it has been shown to be unfavourable for Factor Xa, with peptides bearing this residue cleaved at a 50-fold lower rate than the favoured Phe residue at this position (Ludeman *et al.*, 2003). Although comprehensive studies are lacking, it is also strongly indicated that P_1' Pro is unfavourable for H64A subtilisin (Carter, 1990) and enterokinase (Stevens, 2000). However, this last enzyme has been observed to cleave N-terminally to a Pro in a fusion protein substrate (Hosfield and Lu, 1999), but even in this example it was the least favourable residue at P_1'.

In the case of α-Lytic protease, the cleavage before Pro is likely to be attributable to a highly favourable structure adopted by the substrate interacting outside of S_1', rather than a true example of favouring P_1' Pro (see below). This prospect remains intriguing and may merit further investigation.

6.3.2 *The influence of structure on cleavage*

Aside from the structural manipulations that were required to arrive at the final successful cleavage outcome, information can be gleaned from the

incorrect cleavages. These not surprisingly point towards the critical role of substrate structure in α-Lytic protease catalysis.

A further example of the strong impact that substrate structure has on α-Lytic protease activity is provided by the fortuitous cleavage at the Leu-Phe-Val-Asn sequence in two other separate fusion protein constructs, one in this work (Met-pGH(1-11)VN-WGSGPAPM-IGF-I) and previously in Met-pGH(1-11)VN-FAHY-IGF-II (Haggett *et al.,* 1994c). The other P_4 to P_2 substrate amino acid residues in this sequence offer no theoretical impediment to α-Lytic protease activity, Moreover, Asn has been reported as somewhat favourable in a direct assay of P_1 function for PragA9 (Graham *et al.,* 1993) and unfavourable in an indirect selection assay for P_1 residues (Lien *et al.,* 2003). This sequence could therefore be considered far from ideal, yet the combination of the Leu-Phe-Val-Asn sequence and local structure of the fusion proteins generate a highly favourable site for α-Lytic protease. The implication of this is clear, as cleavage at this Leu-Phe-Val-Asn motif occurs only in some situations. For example, neither the Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I fusion in this work, nor the Met-pGH(1-11)VN-PAPM-IGF-II of Lien *et al.,* (2001) display cleavage at the corresponding site. Thus, despite both bearing an identical sequence at the LFVN-PAPM junction to that in Met-pGH(1-11)VN-PAPM-IGF-I cleavage occurs at the required site. As is frequently observed with even the most specific proteases, the occupation of the enzyme binding site by a preferred motif (described by the K_m value) excludes binding and hence cleavage at other less preferred substrate sites. An example of this phenomenon, taken to the extreme, is provided in the action of Serpins. These proteins display

sequence and structural features that complement the binding site of their target protease so completely that they bind irreversibly and to the exclusion of all other substrates (Huntington, 2006).

6.4 FUTURE DIRECTIONS

The purpose of this work has been to establish a proof of concept, with the initial expectation of finding an α-Lytic protease mutant with altered structure at distal locations allowing cleavage of an IGF-I fusion protein. The ability of at least one mutant to liberate IGF-I with a native N-terminus has been confirmed. It has been argued this may be due to the structural accommodation provided by this mutant for substrate binding, and also reduced inhibition to catalysis by the unfavourable P' residues, Gly-Pro-Glu. More detailed study of this mutant and analysis of both the kinetics of binding and the cleavage reaction could provide greater insights into the design of other motifs.

6.4.1 *Short term goals*

A kinetic approach to the cleavage of the Met-pGH(1-11)VN-WGSGPAPM-IGF-I substrate could provide a simple and practical solution to some of the issues faced. Whilst it is clear that there is a difference in the protease activity at the various cleavage sites, it is unknown if this is due to differences in the K_m of the enzyme for these motifs or its ability to cleave the bonds, resulting in differences in K_{cat}. Manipulation of the substrate concentration (held constant in this work) could exploit any differences in K_m,

further reducing the relative reaction rate at the incorrect site and allowing the recovery of more full-length product. Separation of the individual intermediates and products, as undertaken in this work, would allow investigation of the subsequent reaction step in isolation (such as the conversion of WGSGPAPM-IGF-I to IGF-I).

A follow up of the preliminary findings of the Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I construct would be very useful. In hindsight this construct may have been far more successful than the Met-pGH(1-11)VN-WGSGPAPM-IGF-I that was investigated in detail. In the initial screening reaction of Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I, the dominant product species appeared to be the des(68-70) form of authentic IGF-I. Should the identity of this peak product be confirmed as des(68-70)IGF-I, investigation of the cleavage reaction the reaction conditions could lead to significant improvements in process yield in a very short time. For example, Met-pGH(1-11)VN-WGSGPAPM-IGF-I cleavage left approx. 10% full-length substrate and 25% of WGSGPAPM-IGF-I intermediate. In comparison, Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I left no starting material or intermediates, only a dominant putative des(68-70)IGF-I peak and a smaller authentic IGF-I peak.

The only challenging issue is the rapid progression of the latter construct to the des(68-70) form. It could be that the competitive substrate offered by the C-terminal motif in the presence of this leader renders the C-terminal clip more attractive. Consequently the reaction proceeds first through the des(68-70) form of Met-pGH(1-11)VN-[$F^{-16}E/M^{-13}E$]-PAPM-IGF-I then to liberation of des(68-70)IGF-I. Study of the reaction progression of this

construct would address such issues and variation in the position and number of the charge mutations (to residues other than F^{-16} or M^{-13}) may lead to the desired outcome. Moreover, the double benefit of combination of charge mutants with a flexible spacer should not be overlooked.

The positive result of the Met-pGH(1-11)VN-WGSGPAPM-IGF-I construct could possibly be further enhanced by increasing the length of the flexible spacer. Duplication of a similar spacer sequence by Polyak *et al.,* (1997) improved cleavage of Met-pGH(1-46)-FAHY-IGF-II by H64A Subtilisin nearly 20% over the single copy of the spacer.

6.4.2 *Longer term goals*

The other approach to the challenges addressed here could involve investigating proteases that not only tolerate, but require, the P' sequence of IGF-I for their function. Prime examples of such proteases are 3Cpro, which demands a P_1' Gly and a P_2' Pro (Ivanoff *et al.*, 1986; Cordingley *et al.,* 1990) and thus accommodates the N-terminal sequence of IGF-I. Alternatively, the IgA protease produced by *Neisseria gonorrhoeae* targets the IgA molecule at sites which invariably contain a Pro at P_2' (Pohlner *et al.,* 1992).

Substrate phage-display has been used previously to improve the recognition sequence for PragA9 α-Lytic protease (Lien *et al.,* 2003). This screening process could be repeated with phage bearing the Gly-Pro-Glu sequence downstream of a randomised linker. It would be extremely interesting to determine if this specific P' sequence altered the P motif optima.

The original structural models of α-Lytic protease (Fujinaga *et al.,* 1985; Bone *et al.,* 1987) led first to the generation of rational mutants with

interesting and unexpected properties (Bone *et al.,* 1989), and then to the combinatorial mutant libraries (Graham *et al.,* 1993; Graham *et al.,* 1994). A structural model of the PragA9 mutant will enable interpretation of this highly interesting enzyme, and possibly lay the foundation for further mutagenesis.

Another course of action directed toward the industrial application of α-Lytic protease would be to subject the S' subsites, particularly the most selective S_2', to combinatorial mutagenesis building on the PragA9 S_1 mutant as the library parent. Crystal structures of wild-type α-Lytic protease shows the substrate backbone at P_2' contacts the backbone of Leu^{16} (Bone *et al.,* 1991a). A later NMR model predicts that the side chain of P_1' would contact Ser^{15} and the side chain of P_2' sandwiches between those of Leu^{16} and Thr^{106} (Davis and Agard, 1998). These three residues (Ser^{15}, Leu^{16} and Thr^{106}) would therefore make promising targets for combinatorial mutagenesis. Recognition at these sites is critical to industrial proteolysis, and a library of active enzymes with mutations at this subsite would be an extremely useful resource. With a growing body of knowledge regarding the substrate requirements of this enzyme, even shorter development cycles for novel constructs con be envisaged.

In conclusion, while α-Lytic protease and in particular its PragA9 mutant, may not be a panacea for any given site-specific cleavage system, it does have a tight but malleable substrate specificity which can be manipulated.

REFERENCES

Allison, R., Johnston, R. E. and Dougherty, W. G. (1986). 'The Nucleotide Sequence of the Coding Region of Tobacco Etch Virus Genomic RNA: Evidence for the Synthesis of a Single Polyprotein.' *Virology*.; **154:** 9-20.

Anderson, L. E., Walsh, K. A. and Neurath, H. (1977). 'Bovine enterokinase. Purification, specificity, and some molecular properties.' *Biochemistry*; **16:** 3354-3360.

Anfinsen, C. B. (1973). 'Principles that govern the folding of protein chains.' *Science*; **181:** 223-230.

Antorini, M., Breme, U., Caccia, P., Grassi, C., Lebrun, S., Orsini, G., Taylor, G., Valsasina, B., Marengo, E., Todeschini, R., Andersson, C., Gellerfors, P. and Gustafsson, J. G. (1997). 'Hydroxylamine-induced cleavage of the asparaginyl-glycine motif in the production of recombinant proteins: the case of insulin-like growth factor I.' *Protein Expr. Purif.*; **11:** 135-147.

Arnau, J., Lauritzen, C., Petersen, G. E. and Pedersen, J. (2006). 'Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins.' *Protein Expr. Purif.*; **48:** 1-13.

Bachovchin, W. W. (1986). '15N NMR spectroscopy of hydrogen-bonding interactions in the active site of serine proteases: evidence for a moving histidine mechanism.' *Biochemistry*; **25:** 7751-7759.

Bachovchin, W. W., Wong, W. Y., Farr-Jones, S., Shenvi, A. B. and Kettner, C. A. (1988). 'Nitrogen-15 NMR spectroscopy of the catalytic-triad histidine of a serine protease in peptide boronic acid inhibitor complexes.' *Biochemistry*; **27:** 7689-7697.

Backes, B. J., Harris, J. L., Leonetti, F., Craik, C. S. and Ellman, J. A. (2000). 'Synthesis of positional-scanning libraries of fluorogenic peptide substrates to define the extended substrate specificity of plasmin and thrombin.' *Nat. Biotechnol.*; **18:** 187-193.

Baechle, D., Cansier, A., Fischer, R., Brandenburg, J., Burster, T., Driessen, C. and Kalbacher, H. (2005). 'Biotinylated fluorescent peptide substrates for the sensitive and specific determination of cathepsin D activity.' *J. Pept. Sci.*; **11:** 166-174.

Baek, K-H. (2003). 'Conjugation and deconjugation of ubiquitin regulating the destiny of proteins.' *Exp. Mol. Med.*; **35:** 1-7.

Bagley, C. J., May, B. L., Szabo, L., McNamara, P. J., Ross, M., Francis, G. L., Ballard, F. J. and Wallace, J. C. (1989). 'A key functional role for the insulin-like growth factor 1 *N*-terminal pentapeptide.' *Biochem. J.*; **259:** 665-671.

Baker, D., Sohl, J. L. and Agard, D. A. (1992). 'A protein-folding reaction under kinetic control.' *Nature*; **356:** 263-265.

Baker, R. T. (1996). 'Protein expression using ubiquitin fusion and cleavage.' *Curr. Opin. Biotechnol.*; **7:** 541-546.

Baneyx, F. and Mujacic, M. (2004). 'Recombinant protein folding and misfolding in *Escherichia coli*.' *Nat. Biotechnol.*; **22:** 1399-1408.

Bauer, C-A., Brayer, G. D., Sielecki, A. R. and James, M. N. G. (1981). 'Active site of α-Lytic Protease: Enzyme:Subsite interactions.' *Eur. J. Biochem.*; **120:** 289-294.

Baxter, R. C., Bayne, M. L. and Cascieri, M. A. (1992). 'Structural determinants for binary and ternary complex formation between insulin-like growth factor-I (IGF-I) and IGF binding protein-3.' *J. Biol. Chem.*; **267:** 60-65.

Bayne, M. L., Applebaum, J., Chicchi, G. G., Hayes, N. S., Green, B. G. and Cascieri, M. A. (1988). 'Structural analogs of human insulin-like growth factor I with reduced affinity for serum binding proteins and the type 2 insulin-like growth factor receptor.' *J. Biol. Chem.*; **263:** 6233-6239.

Bayne, M. L., Applebaum, J., Underwood, D., Chicchi, G. G., Green, B. G., Hayes, N. S. and Cascieri, M. A. (1989). 'The C region of human insulin-like growth factor (IGF) I is required for high affinity binding to the type 1 IGF receptor.' *J. Biol. Chem.*; **264:** 11004-11008.

Belmouden, A., Le, K. H. D., Lederer, F. and Garchon, H-J. (1993). 'Molecular cloning and nucleotide sequence of cDNA encoding rat kidney long-chain L-2-hydroxy acid oxidase. Expression of the catalytically active recombinant protein as a chimera.' *Eur. J. Biochem.*; **214:** 17-25.

Biedendieck, R., Yang, Y., Deckwer, W-D., Malten, M. and Jahn, D. (2007). 'Plasmid System for the Intracellular Production and Purification of Affinity-Tagged Proteins in *Bacillus megaterium*.' *Biotechnol. Bioeng.*; **96:** 525-537.

Bjork, I., Petersson, B. A. and Sjoquist, J. (1972). 'Some physiochemical properties of protein A from *Staphylococcus aureus*.' *Eur. J. Biochem.*; **29:** 579-584.

Blomback, B., Blomback, M., Hessel, B. and Iwanaga, S. (1967). 'Structure of N-terminal fragments of fibrinogen and specificity of thrombin.' *Nature*; **215:** 1445-1448.

Bolger, R. and Checovich, W. (1994). 'A new protease activity assay using fluorescence polarization.' *Biotechniques*; **17:** 585-589.

Bone, R., Sampson, N. S., Bartlett, P. A. and Agard, D. A. (1991a). 'Crystal Structures of a-Lytic Protease Complexes with Irreversibly Bound Phosphonate Esters.' *Biochemistry*; **30:** 2263-2272.

Bone, R., Fujishige, A., Kettner, C. A. and Agard, D. A. (1991b). 'Structural basis for broad specificity in alpha-lytic protease mutants.' *Biochemistry*; **30:** 10388-10398.

Bone, R., Shenvi, A. B., Kettner, C. A. and Agard, D. A. (1987). 'Serine Protease Mechanism: Structure of an Inhibitory Complex of α-Lytic Protease and a Tightly Bound Peptide Boronic Acid.' *Biochemistry*; **26:** 7609-7614.

Bone, R., Silen, J. L. and Agard, D. A. (1989). 'Structural plasticity broadens the specificity of an engineered protease.' *Nature*; **339:** 191-195.

Bornstein, P. and Balian, G. (1977). 'Cleavage at Asn-Gly bonds with Hydroxylamine.' *Methods Enzymol.*; **47:** 132–145.

Boulware, K. T. and Daugherty, P. S. (2006). 'Protease specificity determination by using cellular libraries of peptide substrates (CLiPS).' *Proc. Natl. Acad. Sci. USA*; **103:** 7583-7588.

Bowie, J. U. and Sauer, R. T. (1989). 'Identification of C-terminal Extension That Protect Proteins from Intracellular Proteolysis.' *J. Biol. Chem.*; **264:** 7596-7602.

Bowman, B. R., Welschhans, R. L., Jayaram, H., Stow, N. D., Preston, V. G. and Quiocho, F. A. (2006). 'Structural characterization of the UL25 DNA-packaging protein from herpes simplex virus type 1.' *J. Virol.*; **80:** 2309-2317.

Brayer, G. D., Delbaere, L. T. and James, M. N. (1979). 'Molecular structure of the alpha-lytic protease from Myxobacter 495 at 2.8 Angstroms resolution.' *J. Mol. Biol.*; **131:** 743-775.

Buckle, A. M., Zahn, R. and Fersht, A. R. (1997). 'A structural model for GroEL-polypeptide recognition.' *Proc. Natl. Acad. Sci. USA*; **94:** 3571–3575.

Buell, G., Schulz, M. F., Selzer, G., Chollet, A., Movva, N. R., Semon, D., Escanez, S. and Kawashima, E. (1985). 'Optimizing the expression in E. coli of a

synthetic gene encoding somatomedin-C (IGF-I).' *Nucleic Acids Res.*; **13:** 1923-1938.

Butkowski, R. J., Elion, J., Downing, M. R. and Mann, K. G. (1977). 'Primary structure of human prethrombin 2 and alpha-thrombin.' *J. Biol. Chem.*; **252:** 4942-4957.

Butt, T. R., Edavettal, S. C., Hall, J. P. and Mattern, M. R. (2005). 'SUMO fusion technology for difficult-to-express proteins.' *Protein Expr. Purif.*; **43:** 1-9.

Butt, T. R., Jonnalagadda, S., Monia, B. P., Sternberg, E. J., Marsh, J. A., Stadel, J. M., Ecker, D. J. and Crooke, S. T. (1989). 'Ubiquitin Fusion Augments the Yield of Cloned Gene Products in *Escherichia coli*.' *Proc. Natl. Acad. Sci. USA.*; **86:** 2540-2544.

Cabrita, L. D and Bottomley, S. P. (2004). 'How do proteins avoid becoming too stable? Biophysical studies into metastable proteins.' *Eur. Biophys. J.*; **33:** 83-88.

Cabrita, L. D., Dai, W. and Bottomley, S. P. (2006). 'A family of E. coli expression vectors for laboratory scale and high throughput soluble protein production.' *BMC Biotechnol.*; **6:** 12.

Canova-Davis, E., Eng, M., Mukku, V., Reifsnyder, D. H., Olson, C. V. and Ling, V. T. (1992). 'Chemical heterogeneity as a result of hydroxylamine cleavage of a fusion protein of human insulin-like growth factor I.' *Biochem J.*; **285:** 207-213.

Carrington, J. C. and Dougherty, W. G. (1988). 'A viral cleavage site cassette: Identification of amino acid sequences required for tobacco etch virus polyprotein processing.' *Proc. Natl. Acad. Sci. USA.*; **85:** 3391-3395.

Carter, P. (1990). 'Site-Specific Proteolysis of Fusion Proteins.' In: *Protein Purification: From Molecular Mechanisms to Large Scale Processes.* (Ladish, M. R., Willson, R. C., Printon, C. C. and Builder, S. E. eds), p181-193, ACS, Washington, USA.

Carter, P., Nilsson, B. Burnier, J. P., Burdick, D. and Wells, J. A. (1989). 'Engineering subtilisin BPN' for site-specific proteolysis.' *Proteins*; **6:** 240-248.

Carter, P. and Wells, J. A. (1987). 'Engineering enzyme specificity by "substrate-assisted catalysis".' *Science*; **237:** 394-399.

Cascieri, M. A., Chicchi, G. G., Applebaum, J., Green, B. G., Hayes, N. S. and Bayne, M. L. (1989). 'Structural analogs of human insulin-like growth factor (IGF) I with altered affinity for type 2 IGF receptors.' *J. Biol. Chem.*; **264:** 2199-2202.

Cascieri, M. A., Chicchi, G. G., Applebaum, J., Hayes, N. S., Green, B. G. and Bayne, M. L. (1988). 'Mutants of human insulin-like growth factor I with reduced affinity for the type 1 insulin-like growth factor receptor.' *Biochemistry*; **27:** 3229-3233.

Castagne, N., Barbier, A., Bernard, J., Rezaei, H., Huet, J. C., Henry, C., Da Costa, B. and Eleouet, J. F. (2004). 'Biochemical characterization of the respiratory syncytial virus P-P and P-N protein complexes and localization of the P protein oligomerization domain.' *J. Gen. Virol.*; **85:** 1643-1653.

Catanzariti, A. M., Soboleva, T. A., Jans, D. A., Board, P. G. and Baker, R. T. (2004). 'An efficient system for high-level expression and easy purification of authentic recombinant proteins.' *Protein Sci.*; **13:** 1331-1339.

Chambers, S. P., Austen, D. A., Fulghum, J. R. and Kim, W. M. (2004). 'Highthroughput screening for soluble recombinant expressed kinases in *Escherichia coli* and insect cells.' *Protein Expr. Purif.*; **36:** 40-47.

Chang, J-Y. (1985). 'Thrombin specificity.' *Eur. J. Biochem.*; **151:** 217-224.

Chaudhuri, B., Latham, S. E. and Stephan, C. (1992). 'A mutant Kex2 enzyme with a C-terminal HDEL sequence releases correctly folded human insulin-like growth factor-1 from a precursor accumulated in the yeast endoplasmic reticulum.' *Eur. J. Biochem.*; **210:** 811-822.

Chayen, N. E. (2004). 'Turning protein crystallization from an art into a science.' *Curr. Opin. Struct. Biol.*; **14:** 577-583.

Chen, J., Skehel, J. J. and Wiley, D. C. (1998). 'A polar octapeptide fused to the N-terminal fusion peptide solubilizes the influenza virus HA2 subunit ectodomain.' *Biochemistry.*; **37:** 13643-13649.

Cheng, Y. S. (1983). 'Increased cell buoyant densities of protein overproducing *Escherichia coli* cells.' *Biochem. Biophys. Res. Commun.*; **111:** 104-111.

Choi, J. H., Lee, S. J. and Lee, S. Y. (2003). 'Enhanced Production of Insulin-Like Growth Factor I Fusion Protein in *Escherichia coli* by Coexpression of the Down-Regulated Genes Identified by Transcriptome Profiling.' *Appl. Environ. Microbiol.*; **69:** 4737-4742.

Choi, S. I., Song, H. W., Moon, J. W. and Seong, B. L. (2001). 'Recombinant Enterokinase Light Chain with Affinity Tag: Expression from *Sacchatomyces cerevisiae* and Its Utilities in Fusion Protein Technology.' *Biotechnol. Bioeng.*; **75:** 718-724.

Coler, R. N., Skeiky, Y. A., Ovendale, P. J., Vedvick, T. S., Gervassi, L., Guderian, J., Jen, S., Reed, S. G. and Campos-Neto, A. (2000). 'Cloning of a Mycobacterium tuberculosis gene encoding a purifed protein derivative protein that elicits strong tuberculosis-specific delayed-type hypersensitivity.' *J. Infect. Dis.*; **182:** 224-233.

Cooke, R. M., Harvey, T. S. and Campbell, I. D. (1991). 'Solution Structure of Human Insulin-Like Growth Factor 1 : A Nuclear Magnetic Resonance and Restrained Molecular Dynamics Study.' *Biochemistry*; **30:** 5484-5491.

Cool, R. H., Veenstra, M. K., van Klompenburg, W., Heyne, R. I., Muller, M., de Vries, E. G., van Veen, H. W. and Konings, W. N. (2002). 'S-decyl-glutathione nonspecifically stimulates the ATPase activity of the nucleotide-binding domains of the human multidrug resistance-associated protein, MRP1 (ABCC1).' *Eur. J. Biochem.*; **269:** 3470-3478.

Cordingley, M. G., Callahan, P. L., Sardana, V. V., Garsky, V. M. and Colonno, R. J. (1990). 'Substrate Requirements of Human Rhinovirus 3C Protease for Peptide Cleavage *in Vitro*.' *J. Biol. Chem.*; **265:** 9062-9065.

Crestfield, A. M., Moore, S. and Stein, W. H. (1963). 'The Preparation and Enzymatic Hydrolysis of Reduced and S-Carboxymethylated Proteins.' *J. Biol. Chem.*; **238:** 622-627.

Cunningham, E. L., Jaswal, S. S., Sohl, J. L. and Agard, D. A. (1999). 'Kinetic stability as a mechanism for protease longevity.' *Proc. Natl. Acad. Sci. USA*; **96:** 11008-11014.

Dahlman, K., Stromstedt, P. E., Rae, C., Jornvall, H., Flock, J. I., Carlstedt-Duke, J. and Gustafsson, J. A. (1989). 'High level expression in Escherichia coli of the DNA-binding domain of the glucocorticoid receptor in a functional form utilizing domain-specific cleavage of a fusion protein.' *J. Biol. Chem.*; **264:** 804-809.

Davis, G. D., Elisee, C., Newham, D. M. and Harrison, R. G. (1999). 'New fusion protein systems designed to give soluble expression in *Escherichia coli*.' *Biotechnol. Bioeng.*; **65:** 382–388.

Davis, J. H. and Agard, D. A. (1998). 'Relationship between Enzyme Specificity and the Backbone Dynamics of Free and Inhibited α-Lytic Protease.' *Biochemistry.*; **37:** 7696-7707.

Davis, J. H., Agard, D. A., Handel, T. M. and Basus, V. J. (1997). 'Alterations in chemical shifts and exchange broadening upon peptide boronic acid inhibitor binding to alpha-lytic protease.' *J. Biomol. NMR.*; **10:** 21-27.

Degger, B., Richardson, N., Collet, C. and Upton, Z. (2001). 'Production, in vitro characterisation, in vivo clearance, and tissue localisation of recombinant barramundi (*Lates calcarifer*) insulin-like growth factor II.' *Gen. Comp. Endocrinol.*; **123:** 38-50.

di Guan, C., Li, P., Riggs, P. D. and Inouye, H. (1988). 'Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose binding protein.' *Gene*; **67:** 21-30.

Dougherty, W. G. and Parks, T. D. (1989). 'Molecular Genetic and Biochemical Evidence for the Involvement of the Heptapeptide Cleavage Sequence in Determining the Reaction Profile at Two Tobacco Etch Virus Cleavage Sites in Cell-Free Assays.' *Virology*; **172:** 145-155.

Dunbar, A. J., Priebe, I. K., Sanderson, M. P. and Goddard, C. (2001). 'Purification and molecular characterization of recombinant rat betacellulin.' *J. Mol. Endocrinol.*; **27:** 239-247.

Dykes, C. W., Bookless, A. B., Coomber, B. A., Noble, S. A., Humber, D. C. and Hobden, A. N. (1988). 'Expression of atrial natriuretic factor as a cleavable fusion protein with chloramphenicol acetyltransferase in *Escherichia coli.*' *Eur. J. Biochem.*; **174:** 411-416.

Ekstrand, B. and Larsson-Raznikiewicz, M. (1978). 'The monomeric casein composition of different size bovine casein micelles.' *Biochim. Biophys. Acta.*; **536:** 1-9.

Ellinger, S., Glockshuber, R., Jahn, G. and Pluckthun, A. (1989). 'Cleavage of procaryotically expressed human immunodeficiency virus fusion proteins by factor Xa and application in western blot (immunoblot) assays.' *J. Clin. Microbiol.*; **27:** 971-976.

Elliott, S., Fagin, K. D., Narhi, L. O., Miller, J. A., Jones, M., Koski, R., Peters, M., Hsieh, P., Sachdev, R., Rosenfeld, R. D., Rohde, M. F. and Arakawa, T. (1990). 'Yeast-derived recombinant human insulin-like growth factor I: production, purification, and structural characterization.' *J. Protein Chem.*; **9:** 95-104.

Epstein, D. M. and Abeles, R. H. (1992). 'Role of serine 214 and tyrosine 171, components of the S2 subsite of alpha-lytic protease, in catalysis.' *Biochemistry*; **31:** 11216-11223.

Esposito, D. and Deb K Chatterjee, D. K. (2006). 'Enhancement of soluble protein expression through the use of fusion tags.' *Curr. Opin. Biotechnol.*; **17:** 353-358.

Fenton, W. A., Kashi, Y., Furtak, K. and Horwich, A. L. (1994). 'Residues in chaperonin GroEL required for polypeptide binding and release.' *Nature*; **371:** 614–619.

Fersht, A. R., Blow, D. M. and Fastrez, J. (1973). 'Leaving Group Specificity in the Chymotrypsin-Catalysed Hydrolysis of Peptides. A Stereochemical Interpretation.' *Biochemistry*; **12:** 2035-2041.

Finck, M., Bergmann, N., Jansson, B. and Ernst, J. F. (1996). 'Defective threonine-linked glycosylation of human insulin-like growth factor in mutants of the yeast *Saccharomyces cerevisiae*.' *Glycobiology*; **6:** 313-320.

Forsberg, G., Baastrup, B., Brobjer, M., Lake, M., Jornvall, H. and Hartmanis, M. (1989). 'Comparison of two chemical cleavage methods for preparation of a truncated form of recombinant human insulin-like growth factor I from a secreted fusion protein.' *Biofactors.*; **2:** 105-112.

Forsberg, G., Baastrup, B., Rondahl, H. Holmgren, E., Pohl, G., Hartmanis, M. and Lake, M. (1992). 'An Evaluation of Different Enzymatic Cleavage Methods for Recombinant Fusion Proteins, Applied on Des(1-3)Insulin-Like Growth Factor I.' *J. Prot. Chem.* **11:** 201-211.

Forsberg, G., Brobjer, M., Holmgren, E., Bergdahl, K., Persson, P., Gautvik, K. M. and Hartmanis, M. (1991). 'Thrombin and H64A subtilisin cleavage of fusion proteins for preparation of human recombinant parathyroid hormone.' *J. Protein Chem.*; **10:** 517–526.

Forsberg, G., Palm, G., Ekebacke, A., Josephson, S. and Hartmanis, M. (1990). 'Separation and characterizaton of modified recombinant human insulin-like growth factor I derived from a fusion protein secreted from *Escherichia coli*.' *Biochem. J.*; **271:** 357–363.

Francis, G. L., Aplin, S. E., Milner, S. J., McNeil, K. A., Ballard, F. J. and Wallace, J. C. (1993). 'Insulin-like growth factor (IGF)-II binding to IGF-binding proteins and IGF receptors is modified by deletion of the N-terminal hexapeptide or substitution of arginine for glutamate-6 in IGF-II.' *Biochem. J.*; **293:** 713-719.

Francis, G. L., Read, L. C., Ballard, F. J., Bagley, C. J., Upton, F. M., Gravestock, P. M. and Wallace, J. C. (1986). 'Purification and partial sequence analysis of insulin-like growth factor-1 from bovine colostrums.' *Biochem. J.*; **233:** 207-213.

Francis, G. L., Ross, M., Ballard, F. J., Milner, S. J., Senn, C., McNeil, K. A., Wallace, J. C., King, R. and Wells, J. R. E. (1992). 'Novel recombinant fusion protein analogues of insulin-like growth factor (IGF)-I indicate the relative importance of IGF-binding protein and receptor binding for enhanced biological potency.' *J. Mol. Endocrinol.*; **8:** 213–223.

Fuhrmann, C. A., Kelch, B. A., Ota, N. and Agard, D. A. (2004). 'The 0.83 Å Resolution Crystal Structure of α-Lytic Protease Reveals the Detailed Structure of the Active Site and Identifies a Source of Conformational Strain.' *J. Mol. Biol.*; **338:** 999-1013.

Fuhrmann, C. N., Daugherty, M. D. and Agard, D. A. (2006). 'Subangstrom Crystallography Reveals that Short Ionic Hydrogen Bonds, and Not a His-Asp Low-Barrier Hydrogen Bond, Stabilize the Transition State in Serine Protease Catalysis.' *J. Am. Chem. Soc.*; **128:** 9086-9102.

Fujinaga, M., Delbaere, L. T., Brayer, G. D. and James, M. N. (1985). 'Refined structure of alpha-lytic protease at 1.7 A resolution. Analysis of hydrogen bonding and solvent structure.' *J. Mol. Biol.*; **184:** 479–502.

Fujishige, A., Smith, K. R., Silen, J. L. and Agard, D. A. (1992). 'Correct Folding of α-Lytic Protease Is Required for Its Extracellular Secretion from *Escherichia coli*.' *J. Cell Biol.*; **118:** 33-42.

Furman, T. C., Epp, J., Hsiung, H. M., Hoskins, J., Long, G. C., Mendelsohn, L. G., Schoner, B., Smith, D. P. and Smith, M. C. (1987). 'Recombinant Human Insulin-Like Growth Factor II Expressed in *Escherichia coli*.' *Bio/Technol.*; **5:** 1047-1051.

Gellerfors, P., Axelsson, K., Helander, A., Johansson, S., Kenne, L., Lindqvist, S., Pavlu, B., Skottner, A. and Fryklund L. (1989). 'Isolation and characterization of a glycosylated form of human insulin-like growth factor I produced in *Saccharomyces cerevisiae*.' *J. Biol. Chem.*; **264:** 11444–11449.

Goeddel, D.V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. and Riggs, A. D. (1979). 'Expression in *Escherichia coli* of chemically synthesized genes for human insulin.' *Proc. Natl. Acad. Sci. USA*; **76:** 106–110.

Gosalia, D. N., Salisbury, C. M., Ellman, J. A. and Diamond, S. L. (2005). 'High throughput substrate specificity profiling of serine and cysteine proteases using solution-phase fluorogenic peptide microarrays.' *Mol. Cell Proteomics.*; **4:** 626-636.

Gottesman, S. (1989). 'Genetics of proteolysis in *Escherichia coli*.' *Annu. Rev. Genet.*; **23:** 163-198.

Gouvea, I. E., Izidoro, M. A., Judice, W. A., Cezari, M. H., Caliendo, G., Santagada, V., Dos Santos, C. N., Queiroz, M. H., Juliano, M. A., Young, P. R., Fairlie, D. P. and Juliano, L. (2007). 'Substrate specificity of recombinant dengue 2 virus NS2B-NS3 protease: Influence of natural and unnatural basic amino acids

on hydrolysis of synthetic fluorescent substrates.' *Arch. Biochem. Biophys.*; **457:** 187-196.

Graf, L., Szilagyi, L. and Venekei, I. (2004). 'Chymotrypsin.' In: *Handbook of Proteolytic Enzymes*, 2nd edition (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. eds), p.1495-1501, Elsevier, London, UK.

Graham, L. D., Haggett, K. D., Hayes, P. J., Schober, P. A., Jennings, P. A. and Whittaker, R. G. (1994). 'A new library of alpha-lytic protease S1 mutants generated by combinatorial random substitution.' *Biochem. Mol. Biol. Int.*; **32:** 831-839.

Graham, L. D., Haggett, K. D., Jennings, P. A., Le Brocque, D. S., Whittaker, R. G. and Schober, P. A. (1993). 'Random mutagenesis of the substrate-binding site of a serine protease can generate enzymes with increased activities and altered primary specificities.' *Biochemistry*; **32:** 6250-6258.

Grys, E. L., Schade, S. Z., Cohen, M. E., Geivelis, M., Robinson, P. J. and Simonson, L. G. (2000). 'Measurement of proteases in human subgingival dental plaque by fluorescence polarization.' *Arch. Oral. Biol.*; **45:** 1101-1106.

Guan, K. L. and Dixon, J. E. (1991). 'Eukaryotic proteins expressed in Escherichia coli: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase.' *Anal. Biochem.*; **192:** 262-267.

Haggett, K. D., Graham, L. D., Milner, S. J. and Whittaker, R. G. (1994a). 'Purification and characterization of S1 mutants of alpha-lytic protease having altered catalytic properties.' *Arch. Biochem. Biophys.*; **314:** 132-141.

Haggett, K. D., Graham, L. D. and Whittaker, R. G. (1994b). 'Rapid purification of α-lytic protease and its mutants from small cultures of recombinant *E. coli*.' *Biotechnol. Tech.*; **8:** 203-208.

Haggett, K. D., Graham, L. D., Wright, D. S. and Whittaker, R. G. (1994c). 'Overexpression of α-lytic protease and its mutants by recombinant *E. coli*.' *Biotechnol. Lett.*; **16:** 51-57.

Hakes, D. J. and Dixon, J. E. (1992). 'New Vectors for High Level Expression of Recombinant Proteins in Bacteria.' *Anal. Biochem.*; **202:** 293-298.

Halfon, S., Baird, T. T., Jr. and Craik, C. S. (2004) 'Trypsin.' In: *Handbook of Proteolytic Enzymes*, 2nd edition (Barrett, A. J., Rawlings, N. D. and Woessner, J. F. eds), p.1483-1488, Elsevier, London, UK.

Hammarberg, B., Nygren, P. A., Holmgren, E., Elmblad, A., Tally, M., Hellman, U., Moks, T. and Uhlen, M. (1989). 'Dual affinity fusion approach and its use to

express recombinant human insulin-like growth factor II.' *Proc. Natl. Acad. Sci. USA.*; **86:** 4367-4371.

Hammarberg, B., Tally, M., Samuelsson, E., Wadensten, H., Holmgren, E., Hartmanis, M., Hall, K., Uhlen, M. and Moks, T. (1991). 'Characterization of an extended form of recombinant human insulin-like growth factor II.' *J. Biol. Chem.*; **266:** 11058-11062.

Hammarstrom, M., Hellgren, N., van den Berg, S., Berglund, H. and Hard, T. (2002). 'Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*.' *Protein Sci.*; **11:** 313-321.

Han, A., Sonoda, T., Kang, J. H., Murata, M., Nliidome, T. and Katayam, Y. (2006). 'Development of a fluorescence peptide chip for the detection of caspase activity.' *Comb. Chem. High Throughput Screen.*; **9:** 21-25.

Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A. and Craik, C. S. (2000). 'Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries.' *Proc. Natl. Acad. Sci. USA.*; **97:** 7754-7759.

Hatakeyama, T., Kohzaki, H. and Yamasaki, N. (1992). 'A microassay for proteases using succinylcasein as a substrate.' *Anal. Biochem.*; **204:** 181-184.

Hearn, M.T.W. and Acosta, D. (2001). 'Applications of novel affinity cassette methods: use of peptide fusion handles for the purification of recombinant proteins.' *J. Mol. Recognit.*; **14:** 323-369.

Hejnaes, K. R., Bayne, S., Norskov, L., Sorensen, H. H., Thomsen, J., Schaffer, L., Wollmer, A. and Skriver, L. (1992). 'Development of an optimized refolding process for recombinant Ala-Glu-IGF-1.' *Protein Eng.*; **5:** 797-806.

Hober, S., Forsberg, G., Palm, G., Hartmanis, M. and Nilsson, B. (1992). 'Disulfide Exchange Folding of Insulin-like Growth Factor I.' *Biochemistry*; **31:** 1749-1756.

Hochuli, E., Banworth, W., Dobeli, H., Gentz, R., and Struber, A. (1988). 'Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent.' *Bio/Technol.* **6:** 1321-1325.

Hochuli, E., Dobeli, H. and Schacher, A. (1987). 'New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues.' *J. Chromatogr.*; **411:** 177–184.

Hodgson, D. R., May, F. E. and Westley, B. R. (1995). 'Mutations at positions 11 and 60 of insulin-like growth factor 1 reveal differences between its interactions

with the type I insulin-like-growth-factor receptor and the insulin receptor.' *Eur. J. Biochem.*; **233:** 299-309.

Hodgson, D. R., May, F. E. B. and Westley, B. R. (1996). 'Involvement of phenylalanine 23 in the binding of IGF-1 to the insulin and type I IGF receptor.' *Regul. Pept.*; **66:** 191-196.

Holmgren, A. (1985). 'Thioredoxin.' *Ann. Rev. Biochem.*; **54:** 237-271.

Holowachuk, E. W. and Ruhoff, M. S. (1995). 'Biologically active recombinant rat granulocyte macrophage colony-stimulating factor produced in *Escherichia coli*.' *Protein Expr. Purif.*; **6:** 588-596.

Hopp, T. P., Pricket, K. S., Price, V.L., Libby, R. T., March, C. J., Ceretti, D. P., Urdal, D.L. and Conlon, P. J. (1988). 'A short polypeptide marker sequence useful for recombinant protein identification and purification.' *Bio/Technol.* **6:** 1204–1210.

Hosfield, T. and Lu, Q. (1999). 'Influence of the Amino Acid Residue Downstream of $(Asp)_4Lys$ on Enterokinase Cleavage of a Fusion Protein.' *Anal. Biochem.*; **269:** 10-16.

Hultin, M. B. and Nemerson, Y. (1978). 'Activation of factor X by factors IXa and VIII; a specific assay for factor IXa in the presence of thrombin-activated factor VIII.' *Blood*; **52:** 928-940.

Humbel, R. E. (1990). 'Insulin-like growth factors I and II.' *Eur. J. Biochem.*; **190**: 445–462.

Huntington, J. A. (2006). 'Shape-shifting serpins – advantages of a mobile mechanism.' *Trends Biochem. Sci.*; **31:** 427-435.

Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F. and Boyer, H. W. (1977). 'Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin.' *Science*; **198:** 1056-1063.

Ivanoff, L. A., Towatari, T., Ray, J., Korant, B. D. and Petteway, S. R. Jr. (1986). 'Expression and site-specific mutagenesis of the poliovirus 3C protease in *Escherichia coli*.' *Proc. Natl. Acad. Sci. USA*; **83:** 5392–5396.

Ivanov, D., Bachovchin, W. W. and Redfield, A. G. (2002). 'Boron-11 pure quadrupole resonance investigation of peptide boronic acid inhibitors bound to alpha-lytic protease.' *Biochemistry*; **41:** 1587-1590.

Iwai, M., Kobayashi, M., Tamura, K., Ishii, Y., Yamada, H. and Niwa, M. (1989). 'Direct identification of disulfide bond linkages in human insulin-like growth factor I (IGF-I) by chemical synthesis.' *J. Biochem.*; **106:** 949-951.

Jacquet, A., Daminet, V., Haumont, M., Garcia, L., Chaudoir, S., Bollen, A. and Biemans, R. (1999). 'Expression of a recombinant Toxoplasma gondii ROP2 fragment as a fusion protein in bacteria circumvents insolubility and proteolytic degradation.' *Protein Expr. Purif.*; **17:** 392-400.

Jackowski, G., Suria, D. and Liew, C. C. (1976). 'Fractionation of nucleolar proteins by two-dimensional gel electrophoresis.' *Can. J. Biochem.* **54:** 9-14.

Jaswal, S. S., Sohl, J. L., Davis, J. H. and Agard, D. A. (2002). 'Energetic landscape of α-lytic protease optimizes longevity through kinetic stability.' *Nature*; **415:** 343-346.

Jenny, R. J., Mann, K. G. and Lundblad, R. L. (2003). 'A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa.' *Protein Expr. Purif.*; **31:** 1-11.

Jeong, K. J. and Lee, S. Y. (2003). 'Enhanced Production of Recombinant Proteins in *Escherichia coli* by Filamentation Suppression.' *Appl. Environ. Microbiol.*; **69:** 1295-1298.

Joly, J. C., Leung, W.S. and Swartz, J. R. (1998). 'Overexpression of *Escherichia coli* oxidoreductases increases recombinant insulin-like growth factor-I accumulation.' *Proc. Natl. Acad. Sci. USA*; **95:** 2773-2777.

Jonasson, P., Nilsson, J., Samuelsson, E., Moks, T., Stahl, S. and Uhlen, M. (1996). 'Single-step trypsin cleavage of a fusion protein to obtain human insulin and its C peptide.' *Eur. J. Biochem.*; **236:** 656-661.

Jones, J. I. and Clemmons, D. R. (1995). 'Insulin-like growth factors and their binding proteins: biological actions.' *Endocr Rev.*; **16:** 3-34.

Kane, J. F., and Hartley, D. L. (1988). 'Formation of recombinant inclusion bodies in *Escherichia coli.*' *Trends Biotechnol.*; **6:** 95-101.

Kaplan, H., Symonds, V. B., Dugas, H. and Whitaker, D. R. (1970). 'A comparison of the α-lytic protease of *Soragium* sp. and porcine elastase.' *Can. J. Biochem.*; **48:** 649-658.

Kapust, R. B., Tozer, J., Copeland, T. D. and Waugh, D. S. (2002). 'The P1' specificity of tobacco etch virus protease.' *Biochem. Biophys. Res. Comm.*; **294:** 949-955.

Kapust, R. B. and Waugh, D. S. (1999). '*Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused.' *Protein Sci.*; **8:** 1668-1674.

Kim, S-O. and Lee, Y. I. (1996). 'High-level expression and simple purification of recombinant human insulin-like growth factor I.' *J. Biotechnol.* **48:** 97-105.

King, R., Wells, J. R. E., Krieg, P., Snoswell, M., Brazier, J., Bagley, C. J., Wallace, J. C., Ballard, F. J., Ross, M. and Francis, G. L. (1992). 'Production and characterization of recombinant insulin-like growth factor-I (IGF-I) and potent analogues of IGF-I with Gly or Arg substituted for Glu^3, following their expression in *Escherichia coli* as fusion proteins.' *J. Mol. Endocrinol.*; **8:** 29–41.

Ko, Y. H., Thomas, P. J., Delannoy, M. R. and Pedersen, P. L. (1993), 'The cystic fibrosis transmembrane conductance regulator. Overexpression, purification, and characterization of wild type and DzF508 mutant forms of the first nucleotide binding fold in fusion with the maltose-binding protein.' *J. Biol. Chem.*; **268:** 24330-24338.

Koken, M. H., Odijk, H. H., van Duin, M., Fornerod, M. and Hoeijmakers, J. H. (1993). 'Augmentation of protein production by a combination of the T7 RNA polymerase system and ubiquitin fusion: overproduction of the human DNA repair protein, ERCC1, as a ubiquitin fusion protein in *Escherichia coli*.' *Biochem. Biophys. Res. Commun.*; **195:** 643-653.

Kozlov, I. A., Melnyk, P. C., Zhao, C., Hachmann, J. P., Shevchenko, V., Srinivasan, A., Barker, D. L. and Lebl, M. (2006). 'A method for rapid protease substrate evaluation and optimization.' *Comb. Chem. High Throughput Screen.*; **9:** 481-487.

Krausslich, H-G. and Wimmer, E. (1988). 'Viral Proteinases.' *Ann. Rev. Biochem.*; **57:** 701-754.

Kristelly, R., Earnest, B. T., Krishnamoorthy, L. and Tesmer, J. J. (2003). 'Preliminary structure analysis of the DH/PH domains of leukemia-associated RhoGEF.' *Acta Crystallogr. D. Biol. Crystallogr.*; **59:** 1859-1862.

Kumosinski, T. F., Brown, E. M. and Farrell, H. M. Jr. (1993). 'Three-Dimensional Molecular Modeling of Bovine Caseins: An Energy-Minimized β-Casein Structure.' *J. Dairy Sci.*; **76:** 931-945.

Kurz, M., Cowieson, N. P., Robin, G., Hume, D. A., Martin, J. L., Kobe, B. and Listwan, P. (2006). 'Incorporating a TEV cleavage site reduces the solubility of nine recombinant mouse proteins.' *Protein Expr. Purif.*; **50:** 68-73.

Kwon, S-Y, Choi, Y-J, Kang, T-H, Lee, K-H, Cha, S-S, Kim, G-H, Lee, H-S, Kim, K-T and Kim, K-J. (2005). 'Highly efficient protein expression and purification using bacterial hemoglobin fusion vector.' *Plasmid.*; **53:** 274-282.

L. Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F. R. and Donnelly, M. I. (2002). 'A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site.' *Protein Expr. Purif.*; **25:** 8-15.

Laajoki, L. G., Francis, G. L., Wallace, J. C., Carver, J. A. and Keniry, M. A. (2000). 'Solution Structure and Backbone Dynamics of Long-[Arg^{3}]insulin-like Growth Factor-I.' *J. Biol. Chem.*; **275:** 10009-10015.

Laajoki, L. G., Le Breton, E., Shooter, G. K., Wallace, J. C., Francis, G. L., Carver, J. A. and Keniry, M. A. (1997). 'Secondary structure determination of ^{15}N-labelled human Long-[Arg-3]-insulin-like growth factor 1 by multidimensional NMR spectroscopy.' *FEBS Lett.*; **420:** 97-102.

Lauritzen, C., Tuchsen, E., Hansen, P. E. and Skovgaard, O. (1991). 'BPTI and N-terminal extended analogues generated by factor Xa cleavage and cathepsin C trimming of a fusion protein expressed in Escherichia coli.' *Protein Expr. Purif.*; **2:** 372-378.

LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F. and McCoy, J. M. (1993). 'A Thioredoxin Gene Fusion Expression System That Circumvents Inclusion Body Formation in the *E. coli* Cytoplasm.' *Bio/Technol.*; **11**: 187-193.

LaVallie, E. R. and McCoy, J. M. (1995). 'Gene fusion expression systems in *Escherichia coli*.' *Curr. Opin. Biotechnol.*; **6:** 501-506.

Lawson, M.A. and Semler, B.L. (1991). 'Poliovirus thiol proteinase 3C can utilize a serine nucleophile within the putative catalytic triad.' *Proc. Natl. Acad. Sci. USA*; **88:** 9919-9923.

Lee, J. H., Kim, J. H., Hwang, S. W., Lee, W. J., Yoon, H. K., Lee, H. S. and Hong, S. S. (2000). 'High-level expression of antimicrobial peptide mediated by a fusion partner reinforcing formation of inclusion bodies.' *Biochem. Biophys. Res. Commun.*; **277:** 575-580.

Lei, S-P., Lin, H-C., Wang, S-S., Callaway, J. and Wilcox, G. (1987). 'Characterization of the *Erwinia carotovora pelB* Gene and Its Product Pectate Lyase.' *J. Bacteriol.*; **169:** 4379–4383.

Leissring, M. A., Lu, A., Condron, M. M., Teplow, D. B., Stein, R. L., Farris, W. and Selkoe, D. J. (2003). 'Kinetics of amyloid beta-protein degradation

determined by novel fluorescence- and fluorescence polarization-based assays.' *J. Biol. Chem.*; **278:** 37314-37320.

Levine, L. M., Michener, M. L., Toth, M. V. and Holwerda, B. C. (1997). 'Measurement of Specific Protease Activity Utilizing Fluorescence Polarization.' *Anal. Biochem.*; **247:** 83-88.

Lewis, V. O., Gehrmann, M., Weissbach, L., Hyman, J. E., Rielly, A., Jones, D. G., Llinas, M. and Schaller, J. (1999). 'Homologous plasminogen N-terminal and plasminogen-related gene A and B peptides. Characterization of cDNAs and recombinant fusion proteins.' *Eur. J. Biochem.* **259:** 618-625.

Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J. and Tan, S. (2005). 'Comparison of affinity tags for protein purification.' *Protein Expr. Purif.*; **41:** 98-105.

Lien, S., Francis, G. L., Graham, L. D. and Wallace, J. C. (2003). 'Isolating Substrates for an Engineered α-Lytic Protease by Phage Display.' *J. Protein Chem.*; **22:** 155-166.

Lien, S., Milner, S. J., Graham, L. D., Wallace, J. C. and Francis, G. L. (2001). 'Linkers for Improved Cleavage of Fusion Proteins with an Engineered α-lytic Protease.' *Biotechnol. Bioeng.*; **74:** 335-343.

Liew, O. W., Ching Chong, J. P., Yandle, T. G. and Brennan, S. O. (2005). 'Preparation of recombinant thioredoxin fused N-terminal proCNP: Analysis of enterokinase cleavage products reveals new enterokinase cleavage sites.' *Protein Expr. Purif.*; **41:** 332-340.

Liew, O. W., Chong, P. C. J., Lim, Z., Ang, C. X., Lau, Y. C. A., Yandle, T. G. and Brennan, S. O. (2007). 'An SRLLR motif downstream of the scissile bond enhances enterokinase cleavage efficiency.' *Biochimie*; **89:** 21-9.

Light, A. and Janska, H. (1989). 'Enterokinase (enteropeptidase): comparative aspects.' *Trends Biochem. Sci.*; **14:** 110-112.

Light, A., Savithri, H. S. and Liepnieks, J. J. (1980). 'Specificity of bovine enterokinase toward protein substrates.' *Anal. Biochem.*; **106:** 192-206.

Livney, Y. D., Schwan, A. L. and Dalgleish, D. G. (2004). 'A Study of β-Casein Tertiary Structure by Intramolecular Crosslinking and Mass Spectrometry.' *J. Dairy Sci.*; **87:** 3638-3647.

Ludeman, J. P., Pike, R. N., Bromfield, K. M., Duggan, P. J., Cianci, J., Le Bonniec, B., Whisstock, J. C. and Bottomley, S. P. (2003). 'Determination of the

P1', P2' and P3' subsite-specificity of factor Xa.' *Int. J. Biochem. Cell Biol.*; **35:** 221-225.

Luthi, C., Roth, B. V. and Humbel, R. E. (1992). 'Mutants of human insulin-like growth factor II (IGF II). Expression and characterization of truncated IGF II and of two naturally occurring variants.' *Eur. J. Biochem.*; **205:** 483-490.

Lynch, B. A., Loiacono, K. A., Tiong, C. L., Adams, S. E. and MacNeil, I. A. (1997). 'A Fluorescence Polarization Based Src-SH2 Binding Assay.' *Anal. Biochem.*; **247:** 77-82.

MacDonald, R. J., Margolius, H. S. and Erdos, E. G. (1988). 'Molecular biology of tissue kallikrein.' *Biochem. J.*; **253:** 313-321.

Mace, J. E. and Agard, D. A. (1995). 'Kinetic and structural characterization of mutations of glycine 216 in alpha-lytic protease: a new target for engineering substrate specificity.' *J. Mol. Biol.*; **254:** 720-736.

Mace, J. E., Wilk, B. J. and Agard, D. A. (1995). 'Functional linkage between the active site of alpha-lytic protease and distant regions of structure: scanning alanine mutagenesis of a surface loop affects activity and substrate specificity.' *J. Mol. Biol.*; **251:** 116-134.

Maeda, H. (1979). 'Assay of proteolytic enzymes by the fluorescence polarization technique.' *Anal. Biochem.*; **92:** 222-227.

Makrides, S. C. (1996). 'Strategies for Achieving High-Level Expression of Genes in *Escherichia coli*.' *Microbiol. Rev.*; **60:** 512-538.

Malakhov, M. P., Mattern, M. R., Malakhova, O. A., Drinker, M., Weeks, S. D. and Butt, T. R. (2004). 'SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins.' *J. Struct. Funct. Genomics.*; **5:** 75-86.

Malcolm, B. A. (1995). 'The picornaviral 3C proteinases: cysteine nucleophiles in serine proteinase folds.' *Protein Sci.*; **4:** 1439-1445.

Maroux, S., Baratti, J. and Desnuelle, P. (1971). 'Purification and Specificity of Porcine Enterokinase.' *J. Biol. Chem.*; **246:** 5031-5039.

Marston, F. A. (1986). 'The purification of eukaryotic polypeptides synthesized in *Escherichia coli*.' *Biochem. J.*; **240**: 1-12.

Martinez, A., Knappskog, P. M., Olafsdottir, S., Doskeland, A. P., Eiken, H. G., Svebak, R. M., Bozzini, M., Apold, J. and Flatmark, T. (1995). 'Expression of recombinant human phenylalanine hydroxylase as fusion protein in *Escherichia*

coli circumvents proteolytic degradation by host cell proteases. Isolation and characterization of the wild-type enzyme.' *Biochem. J.*; **306:** 589-597.

Matayoshi, E. D., Wang, G. T., Krafft, G. A. and Erickson, J. (1990). 'Novel fluorogenic substrates for assaying retroviral proteases by resonance energy transfer.' *Science*; **247:** 954-958.

Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L. and Worland, S. (1994). 'Structure of Human Rhinovirus 3C Protease Reveals a Trypsin-like Polypeptide Fold, RNA-Binding Site, and Means for Cleaving Precursor Polyprotein.' *Celll*; **77:** 761-771.

Matthews, D. J. and Wells, J. A. (1993). 'Substrate phage: selection of protease substrates by monovalent phage display.' *Science*; **260:** 1113–1117.

McCusker, R. H. and Clemmons, D. R. (1994) 'Effects of cytokines on insulin-like growth factor-binding protein secretion by muscle cells in vitro.' *Endocrinology*; **134:** 2095-2102.

McDonald, J. K., Zeitman, B. B., Reilly, T. J. and Ellis, S. (1969). 'New observations on the substrate specificity of cathepsin C (dipeptidyl aminopeptidase I). Including the degradation of beta-corticotropin and other peptide hormones.' *J. Biol. Chem.*; **244:** 2693-2709.

Miller, D. W. and Agard, D. A. (1999). 'Enzyme specificity under dynamic control: A normal mode analysis of α-Lytic Protease.' *J. Mol. Biol.;* **286:** 267-278.

Miller, J. A., Narhi, L. O., Hua, Q. X., Rosenfeld, R., Arakawa, T., Rohde, M., Prestrelski, S., Lauren, S., Stoney, K. S., Tsai, L. and Weiss, M. A. (1993). 'Oxidative refolding of insulin-like growth factor 1 yields two products of similar thermodynamic stability: a bifurcating protein-folding pathway.' *Biochemistry*; **32:** 5203-5213.

Milner, S. J., Carver, J. A., Ballard, F. J. and Francis, G. L. (1999). 'Probing the Disulfide Folding Pathway of Insulin-Like Growth Factor-I.' *Biotechnol. Bioeng.*; **62:** 693-703.

Milner, S. J., Francis, G. L., Wallace, J. C., Magee, B. A. and Ballard, F. J. (1995). 'Mutations in the B-domain of insulin-like growth factor-I influence the oxidative folding to yield products with modified biological properties.' *Biochem. J.*; **308:** 865-871.

Misoka, F., Miyake, T., Miyoshi, K., Sugiyama, M., Sakamoto, S. and Fuwa, T. (1989). 'Overproduction of human insulin-like growth factor in *Escherichia coli*.' *Biotechnol. Lett.*; **11:** 839-844.

Mitra, M., Lato, S. M., Ynalvez, R. A., Xiao, Y. and Moroney, J. V. (2004). 'Identification of a New Chloroplast Carbonic Anhydrase in *Chlamydomonas reinhardtii*.' *Plant Physiol.*; **135:** 173-182.

Mitraki, A., and King, J. (1989). 'Protein folding intermediates and inclusion body formation.' *Bio/Technol.*; **7:** 690-697.

Moks, T., Abrahmsen, L., Holmgren, E., Bilich, M., Olsson, A., Uhlen, M., Pohl, G., Sterky, C., Hultberg, H., Josephson, S., Holmgren, A., Jornvall, H. and Nilsson, B. (1987). 'Expression of human insulin-like growth factor I in bacteria: use of optimized gene fusion vectors to facilitate protein purification.' *Biochemistry*; **26:** 5239-5244.

Moon, W. J., Hwang, D. K., Park, E. J., Kim, Y. M. and Chae, Y. K. (2007). 'Recombinant expression, isotope labeling, refolding, and purification of an antimicrobial peptide, piscidin.' *Protein Expr. Purif.*; **51:** 141-146.

Murby, M., Uhlen, M. and Stahl, S. (1996). 'Upstream Strategies to Minimize Proteolytic Degradation upon Recombinant Production in *Escherichia coli*.' *Protein Expr. Purif.*; **7:** 129-136.

Musafia, B., Buchner, V. and Arad, D. (1995). 'Complex Salt Bridges in Proteins: Statistical Analysis of Structure and Function.' *J. Mol. Biol.*; **254:** 761-770.

Muse, W. B. and Bender, R. A. (1999). 'The amino-terminal 100 residues of the nitrogen assimilation control protein (NAC) encode all known properties of NAC from Klebsiella aerogenes and Escherichia coli.' *J. Bacteriol.*; **181:** 934-940.

Nagai, K. and Thorgersen, H. C. (1984). 'Generation of beta-globin by sequence-specific proteolysis of a hybrid protein produced in *Escherichia coli*.' *Nature.*; **309:** 810-812.

Nagai, K., Perutz, M. F. and Poyart, C. (1985). 'Oxygen binding properties of human mutant hemoglobins synthesized in *Escherichia coli*.' *Proc. Natl. Acad. Sci. USA*; **82:** 7252-7255.

Nakashima, A., Mori, K., Nagatsu, T. and Ota, A. (1999). 'Expression of human tyrosine hydroxylase type I in *Escherichia coli* as a protease-cleavable fusion protein.' *J. Neural Transm.*; **106:** 819-824.

Nallamsetty, S., Kapust, R. B., Tozser, J., Cherry, S., Tropea, J. E., Copeland, T. D. and Waugh, D. S. (2004). 'Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease *in vivo* and *in vitro*.' *Prot. Expr. Purif.*; **38:** 108-115.

Nallamsetty, S. and Waugh, D. S. (2006). 'Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners.' *Protein Expr. Purif.*; **45:** 175-182.

Nemerson, Y. (1966). 'The reaction between bovine brain tissue factor and factors VII and X.' *Biochemistry*; **5:** 601-608.

Nilsson, B. and Anderson, S. (1991). 'Proper and improper folding of proteins in the cellular environment.' *Annu. Rev. Microbiol.*; **45:** 607-635.

Nilsson, B., Forsberg, G. and Hartmanis, M. (1991). 'Expression and purification of recombinant insulin-like growth factors from *Escherichia coli*.' *Methods Enzymol.*; **198:** 3-16.

Nilsson, B., Holmgren, E., Josephson, S., Gatenbeck, S., Philipson, L. and Uhlen, M. (1985). 'Efficient secretion and purification of human insulin-like growth factor I with a gene fusion vector in *Staphylococci*.' *Nucl. Acids Res.*; **13:** 1151-1162.

Nilsson, J., Stahl, S., Lundeberg, J., Uhlen, M. and Nygren, P. A. (1997). 'Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins.' *Protein Expr. Purif.*; **11**: 1-16.

Nishikawa, S., Yanase, K., Tokunaga-Doi, T., Kodama, K., Gomi, H., Uesugi, S., Ohtsuka, E., Kato, Y., Suzuki, F. and Ikehara, M. (1987). 'Efficient cleavage by alpha-thrombin of a recombinant fused protein which contains insulin-like growth factor I.' *Protein Eng.*; **1:** 487-492.

Nossal, N. G. and Heppel, L. A. (1966). 'The release of enzymes by osmotic shock from *Escherichia coli* in exponential phase.' *J. Biol. Chem.*; **241:** 3055-3062.

Olsen, M. J., Stephens, D., Griffiths, D., Daugherty, P., Georgiou, G. and Iverson, B. L. (2000). 'Function-based isolation of novel enzymes from a large library.' *Nat. Biotechnol.*; **18:** 1071-1074.

Ota, N. and Agard, D. A. (2001). 'Enzyme specificity under dynamic control II: Principal component analysis of α-lytic protease using global and local solvent boundary conditions.' *Protein Sci.*; **10:** 1403-1414.

Park, S. J., Seo, M. D., Lee, S. K., Ikeda, M., Longnecker, R. and Lee. B. J. (2005). 'Expression and characterization of N-terminal domain of Epstein-Barr virus latent membrane protein 2A in *Escherichia coli*.' *Protein Expr. Purif.*; **41:** 9-17.

Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A. and Dougherty, W. G. (1994). 'Release of Proteins and Peptides from Fusion Proteins Using a Recombinant Plant Virus Proteinase.' *Anal. Biochem.*; **216:** 413-417.

Pedersen, J., Lauritzen, C., Madsen, M.T., and Dahl, S.W. (1999). ,Removal of N-terminal polyhistidine tags from recombinant proteins using engineered aminopeptidases.' *Prot. Expr. Purif.*; **15:** 389-400.

Phan, J., Zdanov, A., Evdokimov, A. G., Tropea, J. E., Peters, H. P. K., Kapust, R. B., Li, M., Wlodawer, A., and Waugh, D. S. (2002). 'Structural basis for the substrate specificity of tobacco etch virus protease.' *J. Biol. Chem.*; **277:** 50564-50572.

Pohlner, J., Klauser, T., Kuttler, E. and Halter, R. (1992). 'Sequence-specific cleavage of protein fusions using a recombinant *Neisseria* type 2 IgA protease.' *Biotechnology*; **10**: 799-804.

Pollitt, S. and Zalkin, H. (1983). 'Role of primary structure and disulfide bond formation in beta-lactamase secretion.' *J. Bacteriol.*; **153:** 27-32.

Polyak, S. W., Forsberg, G., Forbes, B. E., McNeil, K. A., Alpin, S. E. and Wallace, J. C. (1997). 'Introduction of spacer peptides N-terminal to a cleavage recognition motif in recombinant fusion proteins can improve site-specific cleavage.' *Protein Eng.*; **10:** 615-619.

Power, R. F., Conneely, O. M., McDonnell, D. P., Clark, J. H., Butt, T. R., Schrader, W. T. and O'Malley, B. W. (1990). 'High level expression of a truncated chicken progesterone receptor in *Escherichia coli.*' *J. Biol. Chem.*; **265:** 1419-1424.

Prickett, K. S., Amberg, D. C. and Hopp, T. P. (1989). 'A Calcium-Dependent Antibody for Identification and Purification of Recombinant Proteins.' *BioTechniques.*; **7:** 580-587.

Pugsley, A. P. and Schwartz, M. (1985). 'Export and secretion of proteins by bacteria.' *FEMS Microbiol Lett.*; **32:** 3-38.

Quinlan, R. A., Moir, R. D. and Stewart, M. (1989). 'Expression in Escherichia coli of fragments of glial fibrillary acidic protein: characterization, assembly properties and paracrystal formation.' *J. Cell Sci.*; **93:** 71-83.

Rader, S. D. and Agard, D. A. (1997). 'Conformational substates in enzyme mechanism: the 120 K structure of alpha-lytic protease at 1.5 A resolution.' *Protein Sci.*; **6:** 1375-1386.

Raftery, M. J., Collinson, L. and Geczy, C. L. (1999). 'Overexpression, oxidative refolding, and zinc binding of recombinant forms of murine S100 protein MRP14(S100A9).' *Protein Expr. Purif.*; **15:** 228–235.

Rajan, S. S., Lackland, H., Stein, S. and Denhardt, D. T. (1998). 'Presence of an N-Terminal Polyhistidine Tag Facilitates Stable Expression of an Otherwise Unstable N-Terminal Domain of Mouse Tissue Inhibitor of Metalloproteinase-1 in *Escherichia coli*.' *Protein Expr. Purif.*; **13:** 67-72.

Ranatunga, W., Jackson, D., Flowers, R. A. 2nd and Borgstahl, G. E. (2001). 'Human RAD52 protein has extreme thermal stability.' *Biochemistry*; **40:** 8557-8562.

Rawlings, N. D. and Barrett, A. J. (1999). 'MEROPS: the peptidase database.' *Nucl. Acids Res.*; **27:** 325-331.

Rhee, H. J., Lee, Y. I. and Yang, K. H. (1990). 'High-level expression of human insulin-like growth factor II in *Escherichia coli*.' *J. Biotechnol.*; **13:** 293-304.

Ribadeau-Dumas, B., Brignon, G., Grosclaude, F. and Mercier, J. C. (1972). 'Primary structure of bovine beta casein. Complete sequence.' *Eur. J. Biochem.*; **25:** 505-514.

Rinderknecht, E. and Humbel, R. E. (1976). 'Polypeptides with nonsuppressible insulin-like and cell-growth promoting activities in human serum: isolation, chemical characterization, and some biological properties of forms I and II.' *Proc. Natl. Acad. Sci. USA*; **73:** 2365-2369.

Rinderknecht, E. and Humbel, R. E. (1978). 'The Amino Acid Sequence of Human Insulin-like Growth Factor I and Its Structural Homology with Proinsulin.' *J. Biol. Chem.*; **253:** 2769-2776.

Rodriguez, P. L. and Carrasco, L. (1995). 'Improved factor Xa cleavage of fusion proteins containing maltose binding protein.' *Biotechniques*; **18:** 238, 241-243.

Routzahn, K. M. and Waugh, D. S. (2002). 'Differential effects of supplementary affinity tags on the solubility of MBP fusion proteins.' *J. Struct. Funct. Genomics.*; **2:** 83-92.

Rudolph, R. and Lilie, H. (1996). '*In vitro* folding of inclusion body proteins.' *FASEB J.*; **10:** 49-56.

Sachdev, D. and Chirgwin, J. M. (1998). 'Solubility of Proteins Isolated from Inclusion Bodies Is Enhanced by Fusion to Maltose-Binding Protein or Thioredoxin.' *Protein Expr. Purif.*; **12:** 122-132.

Saito, Y., Yamada, H., Niwa, M. and Ueda, I. (1987). 'Production and Isolation of Recombinant Somatomedin C.' *J. Biochem.*; **101:** 123-134.

Sakano, K., Enjoh, T., Numata, F., Fujiwara, H., Marumoto, Y., Higashihashi, N., Sato, Y., Perdue, J. F. and Fujita-Yamaguchi, Y. (1991). 'The design, expression, and characterization of human insulin-like growth factor II (IGF-II) mutants specific for either the IGF-II/cation-independent mannose 6-phosphate receptor or IGF-I receptor.' *J. Biol. Chem.*; **266:** 20626-20635.

Samuelsson, E., Moks, T., Nilsson, B. and Uhlen, M. (1994). 'Enhanced in vitro refolding of insulin-like growth factor I using a solubilizing fusion partner.' *Biochemistry.*; **33:** 4207-4211.

Sassenfeld, H. M. (1990). 'Engineering proteins for purification.' *Trends Biotechnol.*; **8:** 88-93.

Sato, A., Koyama, S., Yamada, H., Suzuki, S., Tamura, K., Kobayashi, M., Niwa, M., Yasuda, T., Kyogoku, Y. and Kobayashi, Y. (2000). 'Three-dimensional solution structure of a disulfide bond isomer of the human insulin-like growth factor-I.' *J. Pept. Res.*; **56:** 218-230.

Saunders, M. J., Kim, H., Woods, T. A., Nolan, J. P., Sklar, L. A., Edwards, B. S. and Graves, S. W. (2006). 'Microsphere-based protease assays and screening application for lethal factor and factor Xa.' *Cytometry A.*; **69:** 342-352.

Schade, S. Z., Jolley, M. E., Sarauer, B. J. and Simonson, L. G. (1996). 'BODIPY-α-Casein, a pH-Independent protein substrate for protease assays using flujorescence polarization.' *Anal. Biochem.*; **243:** 1-7.0

Schechter, I. and Berger, A. (1967). 'On the size of the active site in proteases. I. Papain.' *Biochem. Biophys. Res. Commun.*; **27:** 157-162.

Schein, C. (1989). 'Production of Soluble Recombinant Proteins in Bacteria.' *Bio/Technol.*; **7:** 1141-1149

Schellenberger, V., Turck, C. W., Hedstrom, L. and Rutter, W. J. (1993). 'Mapping the S' Subsites of Serine Proteases Using Acyl Transfer to Mixtures of Peptide Nucleophiles.' *Biochemistry*; **32:** 4349-4353.

Schellenberger, V., Turck, C. W. and Rutter, W. J. (1994). 'Role of the S' subsites in serine protease catalysis. Active-site mapping of rat chymotrypsin, rat trypsin, alpha-lytic protease, and cercarial protease from *Schistosoma mansoni*.' *Biochemistry*; **33:** 4251-4257.

Schmid, C. (1995). 'Insulin-like growth factors.' *Cell Biol. Int.*; **19**: 445–457.

Schultz, J. (1967). 'Cleavage at aspartic acid.' *Methods Enzymol.*; **11:** 255-263.

Sharma, S. and Rose, D. R. (1995). 'Cloning, overexpression, purification, and characterization of the carboxyl-terminal nucleotide binding domain of P-glycoprotein.' *J. Biol. Chem.*; **270:** 14085-14093.

Shechter, Y., Patchornik, A. and Burstein, Y. (1976). 'Selective chemical cleavage of tryptophanyl peptide bonds by oxidative chlorination with N-chlorosuccinimide.' *Biochemistry*; **15:** 5071-5075.

Shih, Y-P., Kung, W-M., Chen J-C., Yeh, C-H., Wang, A. H-J. and Wang T-F. (2002). 'High-throughput screening of soluble recombinant proteins.' *Protein Sci.*; **11:** 1714-1719.

Shine, J., Fettes, I., Lan, N. C., Roberts, J. L. and Baxter, J. D. (1980). 'Expression of cloned beta-endorphin gene sequences by *Escherichia coli*.' *Nature.*; **285**:456-63.

Silen, J. L., Frank, D., Fujishige, A., Bone, R. and Agard, D. A. (1989). 'Analysis of prepro-alpha-lytic protease expression in *Escherichia coli* reveals that the pro region is required for activity.' *J. Bacteriol.*; **171:** 1320-1325.

Sjoquist, J., Meloun, B. and Hjelm, H. (1972). 'Protein A isolated from *Staphylococcus aureus* after digestion with lysostaphin.' *Eur. J. Biochem.*; **29:** 572-578.

Smith, D. B. and Johnson, K. S. (1988). 'Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione-*S*-transferase.' *Gene*; **67:** 31–40.

Spencer, R. D., Toledo, F. B., Williams, B. T. and Yoss, N. L. (1973). 'Design, construction, and two applications for an automated flow-cell polarization fluorometer with digital read out: enzyme-inhibitor (antitrypsin) assay and antigen-antibody (insulin-insulin antiserum) assay.' *Clin. Chem.*; **19:** 838-844.

Spindler, K. R., Rosser, D. S. and Berk, A. J. (1984). 'Analysis of adenovirus transforming proteins from early regions 1A and 1B with antisera to inducible fusion antigens produced in *Escherichia coli*.' *J. Virol.*; **49:** 132-141.

Stark, G. R., Stein, W. H. and Moore, S. (1960). 'Reactions of the Cyanate Present in Aqueous Urea with Amino Acids and Proteins.' *J. Biol. Chem.*; **235:** 3177-3181.

Steube, K., Chaudhuri, B., Marki, W., Merryweather, J. P. and Heim, J. (1991). 'α-Factor-leader-directed secretion of recombinant human-insulin-like growth

factor I from *Saccharomyces cerevisiae*. Precursor formation and processing in the yeast secretory pathway.' *Eur. J. Biochem.*; **198:** 651-657.

Stevens, R. C. (2000). 'Design of high-throughput methods of protein production for structural biology.' *Structure Fold. Des.*; **8:** R177-185.

Stofko-Hahn, R. E., Carr, D. W. and Scott, J. D. (1992). 'A single step purifcation for recombinant proteins. Characterization of a microtubule associated protein (MAP 2) fragment which associates with the type II cAMP-dependent protein kinase.' *FEBS Lett.*; **302:** 274–278.

Stryer, L. and Haugland, R. P. (1967). 'Energy transfer: a spectroscopic ruler.' *Proc. Natl. Acad. Sci. USA*; **58:** 719-726.

Suzuki, S. and Sakuragawa, N. (1989). 'A study on the properties of commercial thrombin preparations.' *Thromb. Res.*; **53:** 271–277.

Takagi, T. and Doolittle, R. F. (1974). 'Amino acid sequence studies on factor XIII and the peptide released during its activation by thrombin.' *Biochemistry*; **13:** 750–756.

Talmadge, K. and Gilbert, W. (1982). 'Cellular Location Affects Protein Stability in *Escherichia coli*.' *Proc. Natl. Acad. Sci. USA.*; **79:** 1830-1833.

Tanskul, S., Oda, K., Oyama, H., Noparatnaraporn, N., Tsunemi, M. and Takada, K. (2003). 'Substrate specificity of alkaline serine proteinase isolated from photosynthetic bacterium, Rubrivivax gelatinosus KDDS1.' *Biochem. Biophys. Res. Comm.*; **309:** 547-551.

Teng, H., Pinon, J. D. and Weiss, S. R. (1999). 'Expression of Murine Coronavirus Recombinant Papain-LikeProteinase: Efficient Cleavage Is Dependent on the Lengths of both the Substrate and the Proteinase Polypeptides.' *J. Virol.*; **73:** 2658-2666.

Tenno, T., Goda, N., Tateishi, Y., Tochio, H., Mishima, M., Hayashi, H., Shirakawa, M. and Hiroaki, H. (2004). 'High-throughput construction method for expression vector of peptides for NMR study suited for isotopic labeling.' *Protein Eng. Des. Sel.*; **17:** 305-314.

Terpe, K. (2003). 'Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems.' *Appl. Microbiol. Biotechnol.*; **60:** 523-533.

Thanabalu, T., Hindley, J. and Berry, C. (1992). 'Proteolytic processing of the mosquitocidal toxin from Bacillus sphaericus SSII-1.' *J. Bacteriol.*; **174:** 5051-5056.

Thomas, D. A., Francis, P., Smith, C., Ratcliffe, S., Ede, N. J., Kay, C., Wayne, G., Martin, S. L., Moore, K., Amour, A. and Hooper, N. M. (2006). 'A broad-spectrum fluorescence-based peptide library for the rapid identification of protease substrates.' *Proteomics.*; **6:** 2112-2120.

Tolbert, T. J., Franke, D. and Wong, C-H. (2005). 'A new strategy for glycoprotein synthesis: ligation of synthetic glycopeptides with truncated proteins expressed in E. coli as TEV protease cleavable fusion protein.' *Bioorg. Med. Chem.*; **13:** 909-915.

Tsunoda, Y., Sakai, N., Kikuchi, K., Katoh, S., Akagi, K., Miura-Ohnuma, J., Tashiro, Y., Murata, K., Shibuya, N. and Katoh, E. (2005). 'Improving expression and solubility of rice proteins produced as fusion proteins in *Escherichia coli.*' *Protein Expr. Purif.*; **42:** 268-277.

Uhlen, M., Forsberg, G., Moks, T., Hartmanis, M. and Nilsson, B. (1992). 'Fusion proteins in biotechnology.' *Curr. Opin. Biotechnol.*; **3:** 363-369.

Uhlen, M. and Moks, T. (1990). 'Gene fusions for purpose of expression: an introduction.' *Methods Enzymol.*; **185:** 129-143.

Uhlen, M., Nilsson, B., Guss, B., Lindberg, M., Gatenbeck, S. and Philipson, L. (1983). 'Gene fusion vectors based on the gene for staphylococcal protein A.' *Gene*; **23:** 369-378.

Unutmaz, C., Cail, R. G. and Haggett, K. D. (1997). 'Increased α-lytic protease production from recombinant *Escherichia coli* grown with oxygen limitation.' *Biotechnol. Lett.*; **19:** 731-734.

Upton, Z., Francis, G. L., Chan, S. J., Steiner, D. F., Wallace, J. C. and Ballard, F. J. (1997). 'Evolution of insulin-like growth factor (IGF) function: production and characterization of recombinant hagfish IGF.' *Gen. Comp. Endocrinol.*; **105:** 79-90.

Upton, Z., Francis, G. L., Kita, K., Wallace, J. C. and Ballard, F. J. (1995). 'Production and characterization of recombinant chicken insulin-like growth factor-II from *Escherichia coli.*' *J. Mol. Endocrinol.*; **14:** 79-90.

Vai, M., Brambilla, L., Orlandi, I., Rota, N., Ranzi, B. M., Alberghina, L. and Porro, D. (2000). 'Improved Secretion of Native Human Insulin-Like Growth Factor 1 from *gas1* Mutant *Saccharomyces cerevisiae* Cells.' *Appl. Environ. Microbiol.*; **66:** 5477-5479.

Varadarajan, R., Szabo, A. and Boxer, S. G. (1985). 'Cloning, expression in Escherichia coli, and reconstitution of human myoglobin.' *Proc. Natl. Acad. Sci. USA*; **82:** 5681-5684.

Vize, P. D. and Wells, J. R. (1987). 'Spacer alterations which increase the expression of porcine growth hormone in *E. coli*.' *FEBS Lett.*; **213:** 155-158.

Voss, E. W. Jr, Workman, C. J. and Mummert, M. E. (1996). 'Detection of protease activity using a fluorescence-enhancement globular substrate.' *Biotechniques*; **20:** 286-291.

Walker, P. A., Leong, L. E., Ng, P. W., Tan, S. H., Waller, S., Murphy, D. and Porter, A. G. (1994). 'Efficient and rapid affinity purification of proteins using recombinant fusion proteases.' *Biotechnology*; **12:** 601-605.

Walz, D. A., Hewett-Emmett, D. and Seegers, W. H. (1977). 'Amino acid sequence of human prothrombin fragments 1 and 2.' *Proc. Natl. Acad. Sci. USA*; **74:** 1969-1972.

Wang, C., Castro, A. F., Wilkes, D. M. and Altenberg, G. A. (1999). 'Expression and purification of the first nucleotide-binding domain and linker region of human multidrug resistance gene product: comparison of fusions to glutathione S-transferase, thioredoxin and maltose-binding protein.' *Biochem. J.*; **338:** 77-81.

Wang, M., Scott, W. A., Rao, K. R., Udey, J., Conner, G. E. and Brew, K. (1989). 'Recombinant bovine alpha-lactalbumin obtained by limited proteolysis of a fusion protein expressed at high levels in *Escherichia coli*.' *J. Biol. Chem.*; **264:** 21116-21121.

Wang, W. W. and Malcolm, B. A. (1999). 'Two-Stage PCR Protocol Allowing Introduction of Multiple Mutations, Deletions and Insertions Using QuikChange™ Site-Directed Mutagenesis.' *Biotechniques*; **26:** 680-682.

Waugh, D. S. (2005). 'Making the most of affinity tags.' *Trends Biotechnol.*; **23:** 316-320.

Weickert, M. J., Doherty, D. H., Best, E. A. and Olins, P. O. (1996). 'Optimization of heterologous protein production in *Escherichia coli*.' *Curr. Opin. Biotechnol.*; **7:** 494-499.

Whitaker, D. R. (1970). 'The α-lytic protease of a Myxobacterium.' *Methods Enzymol.*; **19:** 599–613.

White, J. F., Trinh, L. B., Shiloach, J. and Grisshammer, R. (2004). 'Automated large-scale purification of a G protein-coupled receptor for neurotensin.' *FEBS Lett.*; **564:** 289-293.

Wilkinson, R. J., Elliott, P., Hohmann, A., Francis, G. and Carragher, J. (2004). 'Development and characterization of a competitive polyclonal antibody enzyme-immunoassay for salmon insulin-like growth factor-II.' *Comp. Biochem. Physiol. B.*; **139:** 193-201.

Williams, G , Ruegg, N., Birch, A., Weber, C., Hofstadter, K., Robinson, J. A., Aguet, M., Garotta, G., Schlatter, D. and Huber, W. (1995). 'Dissection of the extracellular human interferon receptor α-chain into two immunoglobulin-like domains. Production in an *Escherichia coli* thioredoxin gene fusion expression system and recognition by neutralizing antibodies.' *Biochemistry*; **34:** 1787-1797.

Xian, C. J., Upton, Z., Goddard, C., Shoubridge, C. A., McNeil, K. A., Wallace, J. C., Read, L. C. and Francis, G. L. (1996). 'Production of a human epidermal growth factor fusion protein and its degradation in rat gastrointestinal flushings.' *J. Mol. Endocrinol.*; **16:** 89-97.

Yang Zhang, Y., Olsen, D. R., Nguyen, K. B., Olson, P. S., Rhodes, E. T. and Mascarenhas, D. (1998). 'Expression of Eukaryotic Proteins in Soluble Form in *Escherichia coli*.' *Protein Expr. Purif.*; **12:** 159-165.

Yansura, D. G. (1990). 'Expression as trpE fusion.' *Methods Enzymol.*; **185:** 161-166.

Yin, X., Wei, D., Yi, L., Tao, X. and Ma, Y. (2005). 'Expression and purification of exendin-4, a GLP-1 receptor agonist, in *Escherichia coli*.' *Protein Expr. Purif.* **41:** 259-265.

Yuan, L. D., and Hua, Z. C. (2002). 'Expression, purification, and characterization of a biologically active bovine enterokinase catalytic subunit in *Escherichia coli*.' *Protein Expr. Purif.*; **25:** 300–304.

Zaitseva, J., Zhang, H., Binnie, B. A. and Hermodson, M. (1996). 'The proteins encoded by the rbs operon of Escherichia coli: II. Use of chimeric protein constructs to isolate and characterize RbsC.' *Protein Sci.*; **5:** 1100-1107.

Zhou, D., Yuen, P., Chu, D., Thon, V., McConnell, S., Brown, S., Tsang, A., Pena, M., Russell, A., Cheng, J. F., Nadzan, A. M., Barbosa, M. S., Dyck, J. R., Lopaschuk, G. D. and Yang, G. (2004). 'Expression, purification, and characterization of human malonyl-CoA decarboxylase.' *Protein Expr. Purif.*; **34:** 261-269.