THE OXIDATIVE FOLDING OF INSULIN-LIKE

GROWTH FACTOR-I ANALOGUES

by

Steven John Milner, B.Sc.(Hons)

A thesis submitted to the

University of Adelaide, South Australia

for the degree of

Doctor of Philosophy

March, 1996

Department of Biochemistry

University of Adelaide

South Australia

Awarded 1996

TABLE OF CONTENTS

SUMMARY

Many unfolded proteins can spontaneously refold to their biologically active form *in vitro* in a thermodynamically controlled process that is governed by the amino acid sequence. However, the production of many recombinant proteins has shown that not all have the capacity to refold quantitatively to their biologically active form *in vitro*. Insulin-like growth factor (IGF)-I, a small globular protein containing three disulphide bonds, is such an example. The *in vitro* refolding of IGF-I showed that 45% of the total protein is correctly folded with the remainder comprising a stable three-disulphide alternative isomer (24%) and numerous stable intermediates.

In this thesis, the factors that contribute to the high stability of the three-disulphide alternative isomer was investigated by examining the refolding properties of highly purified IGF-I mutants. The refolding studies of these mutants showed that the outcome was facilitated by a charge reversal at Glu^3 and by steric factors involving a 13-amino acid N-terminal extension. The binding of the alternative three-disulphide isomer to IGF-binding proteins and to its receptor highlighted the fact that switching two disulphide bonds affected the position of many important residues required for binding. Some preliminary NMR data and molecular modelling carried out on the alternative isomer of IGF-I showed that many of the conformational changes due to switching the disulphide bonds could explain the altered binding interactions. An analogue of IGF-I containing the 13-amino acid N-terminal extension and a charge reversal at Glu^3 (Long-[Arg^3]IGF-I), which shows greatly facilitated refolding properties, was used to investigate the oxidative folding pathway. This study showed that the major flow of the pathway appeared to be through native-like intermediates. However, other non-native intermediates were also involved.

In vitro, the oxidative folding pathway of IGF-I diverges as a result of elements that stabilize the alternative isomer. *In vivo*, it is possible that oxidative folding is guided to the native molecule by cellular machinery that prevents the contact of these non-native-like regions that lead to unproductive folding.

STATEMENT OF ORIGINALITY

This thesis contains no material which has been accepted for the award of any other degree or diploma in any other University. 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 is made. The author consents to the thesis being made available for photocopying and loan.

Steven John Milner

ACKNOWLEDGMENTS

In addition, I would like to thank members of staff and the students of the laboratories of C.S.I.R.O., the Biochemistry Department, and Gro*Pep* for their help, sharing of resources, and camaraderie throughout the course of my study. I would also like to thank the various members of the lunchtime running team for helping me keep fit.

Last, but not least, I especially thank my wife Fran and my daughter Siân for their love, support and encouragement.

PUBLICATIONS ARISING FROM THESIS RESEARCH

Published manuscripts:

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.

Milner, S.J., Thomas, S.M., Ballard, F.J. and Francis, G.L. (1996). Optimization of the hydroxylamine cleavage of an expressed fusion protein to produce recombinant human insulin-like growth factor (IGF)-I. Biotechnology and Bioengineering **50**, (in press).

Milner, S.J., Wallace, J.C., Ballard, F.B. and Francis, G.L. (1996). Probing the disulphide folding pathway of an analogue of insulin-like growth factor (IGF)-I. Biochemistry (in preparation).

Published Abstracts:

Milner, S.J., Francis, G.L., Wallace, J.C. and Ballard, F.J. (1993). Disulphide folding and biological activity of insulin-like growth factor (IGF)-I is influenced by changes in the amino acid sequence. Abstract. *Proceedings: 18th Annual Lorne Conference on Protein Structure and Function,* B14. (Lorne, Australia).

Milner, S.J., Francis, G.L., Wallace, J.C. and Ballard, F.J. (1993). Disulphide folding and biological activity of insulin-like growth factor (IGF)-I is influenced by changes in the amino acid sequence. Abstract. *Proceedings: 37th Annual Conference of the Australian Society for Biochemistry and Molecular Biology,* **25**, 149. (Adelaide, Australia).

* This poster was awarded one of six 'Best Student Poster' prizes.

Milner, S.J. and Francis, G.L. (1994). Changes in the primary structure of IGF-I affect the *in vitro* folding pathway. Abstract. *Third International Symposium on Insulin-like Growth Factors: Growth Regulation.* **4** (Supplement 1), 102. (Sydney, Australia).

Milner, S.J. and Francis, G.L. (1994). Changes in the primary structure of insulin-like growth factor (IGF)-I influence the *in vitro* folding pathway. Abstract. *Proceedings: Eighth Symposium of the Protein Society: Protein Science.* **3** (Supplement 1), 63. (San Diego, U.S.A.).

Milner, S.J., Francis, G.L., Wallace, J.C., Magee, B.A. and Ballard, F.J. (1995). B-domain analogues of insulin-like growth factor (IGF)-I with altered disulphide bridges exhibit modified biological properties. Abstract. *Proceedings: 20th Annual Lorne Conference on Protein Structure and Function,* 201. (Lorne, Australia).

Milner, S.J., Francis, G.L., Wallace, J.C., Magee, B.A. and Ballard, F.J. (1995). The B-domain of insulin-like growth factor (IGF)-I influences the *in vitro* oxidative folding pathway. Abstract. *Proceedings: 20th Annual Lorne Conference on Protein Structure and Function,* 215. (Lorne, Australia).

Milner, S.J., Francis, G.L., Wallace, J.C., Magee, B.A. and Ballard, F.J. (1995). Protein engineering insulin-like growth factor (IGF)-I to control the oxidative folding pathway. Abstract. *Proceedings: Ninth Symposium of the Protein Society: Protein Science.* **4** (Supplement 2),106. (Boston, U.S.A.).

Milner, S.J., Wallace, J.C., Ballard, F.J. and Francis, G.L. (1996). The *in vitro* oxidative folding pathway of Long-[Arg^3]-Insulin-Like Growth Factor-I. Abstract. *Proceedings: 21st Annual Lorne Conference on Protein Structure and Function,* B8. (Lorne, Australia).

ABBREVIATIONS

In addition to the standard abbreviations accepted for use in the *Journal of Biological Chemistry*, the following will be assumed:

CD	Circular Dichroism
DMEM	Dulbecco's Modified Eagle's Medium
DTT	Dithiothreitol
FBS	Foetal Bovine Serum
GuHCl	Guanidinium hydrochloride
2-HED	2-Hydroxyethyl Disulphide
HFBA	Heptafluorobutyric acid
HPLC	High Performance Liquid Chromatography
IGFBP(s)	Insulin-Like Growth Factor Binding Protein(s)
IGF-I	Insulin-Like Growth Factor-I
IGFs	Insulin-Like Growth Factors
[Gly^3]IGF-I	Glu^3 to Gly mutation of IGF-I
[Arg^3]IGF-I	Glu^3 to Arg mutation of IGF-I
[Lys^9]IGF-I	Glu^9 to Lys mutation of IGF-I
Long-IGF-I	[Met^1]pGH(1-11)-Val-Asn-IGF-I
Long-[Gly^3]IGF-I	[Met^1]pGH(1-11)-Val-Asn-[Gly^3]IGF-I
Long-[Arg^3]IGF-I	[Met^1]pGH(1-11)-Val-Asn-[Arg^3]IGF-I
Long-[Lys^9]IGF-I	[Met^1]pGH(1-11)-Val-Asn-[Lys^9]IGF-I
NOESY	Nuclear Overhauser Enhancement Spectroscopy
NMR	Nuclear Magnetic Resonance
pGH	Porcine Growth Hormone

TCA	Trichloroacetic Acid
TFA	Trifluoroacetic Acid
TOCSY	Total Correlation Spectroscopy
UV	Ultra-Violet

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1: INTRODUCTION

To be biologically active and interact with other molecules, a protein must adopt a specific three-dimensional tertiary structure. As many unfolded proteins can spontaneously refold *in vitro* to their biologically active form, the three-dimensional structures of these proteins must be determined by their amino acid sequences and the process is thermodynamically controlled. *In vivo* protein folding is expected to driven by the same principles. While this view of protein folding is maintained today, recent research indicates that it is incomplete. For example, the advent of gene fusion technology for the production of recombinant proteins has highlighted the fact that not all proteins can be refolded quantitatively to the native fold. Furthermore, the need to correctly form disulphide bonds complicates the already complex folding task. It has become clear that *in vitro* and *in vivo* refolding properties are quite different, although the fundamental forces that drive folding are the same. The difference lies in the role of the interactions between the nascent polypeptide chain and the cellular machinery that serves to overcome unproductive folding events which occur. Nevertheless, *in vitro* studies are essential to dissect and understand the physical and chemical interactions that are involved in all stages of folding. Investigations into the very early folding events and the transition states that link them as well as the factors involved in the late stages of folding that lead to the native fold are crucial if we are to understand protein folding in a global sense. Our understanding of how the linear sequence of amino acids folds to form the functional protein is advancing but there are many unknown facets. This thesis focuses on the oxidative folding of insulin-like growth factor(IGF)-I, a small single domain protein which contains three disulphide bonds. The folding of this molecule *in vitro* is unusual in that it forms two stable disulphide isomers, yet *in vivo* only one has been

observed. This property therefore highlights the potential of this molecule to provide a wider understanding of folding from the studies of the factors that influence the formation of native disulphide bonds.

In this chapter I will review events involved in the folding of small single domain proteins to the native state both *in vitro* and *in vivo*. This review is divided into three parts; the stable conformational states of proteins, folding structures and pathways, and the role of covalent interactions (the disulphide bond) in the folding process. The review will encompass the literature available in 1994; the most recent literature relating to IGF-I folding and to my PhD research will be included in the appropriate chapters.

1.2: PROTEIN FOLDING *IN VITRO* and *IN VIVO*

Protein folding is the process by which a protein in an unfolded state reaches the biologically active native folded state. How this occurs is not understood. However, there is a growing body of evidence from the study of early and late intermediates and their transition states in the folding 'pathway' that provides some important clues. In this section I will discuss elements and models of protein folding pathways.

1.2.1: Stable Conformational States of Proteins

1.2.1.1: The Unfolded State

The ideal unfolded protein is a random coil, and is operationally defined as that state that the polypeptide chain enters after a 'major, cooperative breakdown' of the native state in the absence of any changes in covalent structure (Tanford, 1968).

In the unfolded state the rotation angle about each bond of the backbone and the side-chains are therefore independent of distant parts of the molecule. *In vitro* studies of protein folding usually begin with the protein in a chemical environment that is designed to disrupt non-covalent interactions that stabilize the native state. Chaotrophic agents such as guanidinium hydrochloride (GuHCl) and urea are effective in disrupting these non-covalent interactions (Pace, 1986; Hagihara *et al.*, 1993) by diminishing the stabilizing hydrophobic effect (Section 1.2.1.3) and interacting with the unfolded protein (Kamoun, 1988). The resultant unfolded proteins have the average hydrodynamic properties typical of a random coil. However, these strong denaturants may not completely unfold the protein as local hydrophobic clusters can exist even in 7 M urea (Neri *et al.*, 1992). Other methods of unfolding proteins using extremes of pH (Goto *et al.*, 1990), pressure (Pyrse *et al.*, 1992) or temperature (Robertson and Baldwin, 1991; Plaza del Pino *et al.*, 1992), or organic co-solvents (Dufour *et al.*, 1994) have been used but these too may not completely unfold the protein (Dobson, 1992). Whichever method is used, the unfolded state is likely to be slightly different (Shortle, 1993), representing alternative structures with weak local structure and fluctuations expected of a nearly random coil (Fersht, 1994). Furthermore, the residual structure of unfolded proteins is important as it may serve as a nucleus for organizing the folding process (Neri *et al.*, 1992).

An unfolded protein with intact disulphide bonds is not as unfolded as the same protein with reduced disulphide bonds. In these situations the disulphide bond decreases the conformational flexibility and confers stability to the molecule. The contribution of stability by disulphide bonds will be discussed in more detail in section 1.2.3. In view of the role of disulphide bonds, the definition of the unfolded

state indicated above needs to be modified to specify the conditions required to maintain cysteine residues in the reduced form.

1.2.1.2: The Molten Globule State

Intermediate to the unfolded state and the native state is a third stable state which has become to be known as the 'molten globule'. The molten globule is a state of protein molecules that is neither fully unfolded nor fully folded and exists usually under mild denaturing conditions. It is characterized by being almost as compact as the native state and has similar secondary structure, but differs from the native state by virtue of fluctuations and loose packing of amino acid side-chains in the protein core (Kuwajima, 1989). The molten globule is of particular interest as it is preferentially bound by GroEL (Martin et al., 1991) and other molecular chaperones that play an ancillary role in the folding process in vivo. Furthermore, the molten globule state is possibly involved in protein transport and in membrane insertion.

A molten globule, as defined above must be compact, have secondary structure and little persistent tertiary structure, parameters that can be established using techniques such as fast size-exclusion liquid chromatography (Uversky, 1993; Uversky et al., 1992), far-UV circular dichroism (Uversky et al., 1992; Mann and Matthews, 1993; Baum et al., 1989), NMR, fluorescence spectroscopy and 8-anilino-1-naphthalene sulphonate (ANS) binding (Hughson et al., 1990; Elöve et al., 1992; Mann and Matthews, 1993). From these studies and from measurements of ^2H - ^1H exchange the topology of a molten globule shows that helical regions are the least mobile and probably contribute to the stabilization of native-like local structures.

1.2.1.3: The Native State

The native state, or the folded or biologically active structure, is that state of a protein attained under defined physiological conditions. The forces that contribute to stabilization of the native state are hydrophobic interactions, hydrogen bonds, van der Waals forces and electrostatic interactions, and those that oppose stabilization are conformational entropy and charge repulsions (Dill, 1990). However, it is the fine balance of all these forces that dictates the final structure or the fold of the protein. Moreover, the balance of these forces is dependent on the environment, a consideration that is important in protein folding both *in vitro* and *in vivo*.

Native conformations of proteins are complex, comprising of structural domains that consist of secondary structure elements packed together and stabilized by tertiary interactions. While it is easy to think of proteins as rigid structures, thermodynamic considerations and functional requirements demonstrate that the native fold of a protein is flexible (Karplus and McCammon, 1983; Bennett and Huber, 1984), Furthermore, evidence from structures determined by X-ray crystallography or by NMR demonstrate well defined (rigid) regions and regions of flexibility. In general, the native structure can be considered as existing in a range of conformations that readily interchange and where the average structure is similar to that average structure determined in crystals or in solution.

As indicated above, the native state of a protein is a result of the balance between opposing forces. While there have been significantly different conceptual views of the relative roles of the forces that contribute to conformational stability (Rose and Wolfenden, 1993), a concensus opinion appears to have been reached and this will be discussed below. The problem has been difficult as the conformational stability of a protein, which is small, is the sum of much larger

individual forces. The major forces of folding at physiological temperature are determined mainly by the entropy gain due to release of ordered water from nonpolar and polar residues, and the establishment of favourable van der Waals interactions (Dill, 1990; Jaenicke, 1991; Murphy and Gill, 1991; Yu *et al.*, 1994). The effect of the release of ordered water is three-fold. In the unfolded protein intramolecular hydrogen bonds occur with water; during folding and with the exclusion of water, hydrogen bonds become intramolecular. In this situation hydrogen bonds contribute to the entropic effect as well as conferring structural specificity (Seckler and Jaenicke, 1992; Shirley *et al.*, 1992; Creighton, 1991). The third effect is due to the tendency of non-polar atoms to remain in a non-polar environment that favour van der Waals forces rather than in an aqueous environment where van der Waals interactions disrupt the water structure (Privalov and Gill, 1989), the latter effect is synonymous with 'hydrophobic interactions'. From a thermodynamic point of view, the enthalpies and entropies of unfolding are temperature sensitive because the heat capacity of the unfolded state is significantly greater than the folded state, this arises largely from the temperature-dependent ordering of water molecules around the non-polar portions of the protein which are more solvent accessible in the unfolded state (Privalov and Makhatadze, 1992). A further consideration of the stability of folded proteins is the fundamental observation that the interiors of proteins are tightly packed (Richards, 1977). As a result, the free energy of stabilization is approximately proportional to the burial of water-accessible surface area (Eisenberg and McLachlan, 1986; Livingstone *et al.*, 1991).

Other forces, although minor, play a role in protein folding. Many lines of evidence (Dill, 1990; Seckler and Jaenicke, 1992) indicate that electrostatic interactions do not play a dominant role in protein folding. However, they are

relevant in view of mutagenic analysis of protein stability (Matthews, 1993) and research covered later in this thesis (Chapter 4). Long-range electrostatic interactions are negligible (Dao-pin *et al.*, 1991) and surface ion pairs (salt bridges) involved in stabilization are generally weak (Dill, 1990) being in the order of 0.16 - 1 kcal / mol (Akke and Forsén, 1990). Buried ion pairs are relatively rare but can provide a greater contribution to stability; 1 - 3 kcal / mol (Fersht, 1972). A striking example of the contribution of an ion-pair is a 3 - 5 kcal / mol increase of stability for T4 lysozyme (Anderson *et al.*, 1990). Other electrostatic interactions such as helix dipoles and helix capping contribute to stability (Matthews, 1993), but these effects are not dominant forces in folding.

1.2.2: Folding Structures and Pathways

Levinthal in 1969 posed the question 'How long does it take for a protein to fold to its native structure?' and concluded that a random search of all possible conformations is not effective and would take an impossibly long time (Levinthal, 1968; Levinthal, 1969). Yet, biological folding takes place in seconds or less; this is the Levinthal Paradox. A clue to the resolution of the paradox is by imposing a bias such as an energy cost for locally incorrect bond configurations (Zwanzig *et al.*, 1992). However, this view is an over-simplification of the enormously complex and subtle interactions between the forces that initiate and guide protein folding *in vitro* and *in vivo* as well as those that guide the folding of the emerging nascent polypeptide chain. Furthermore, this view does not consider the necessary role of long-range interactions that stabilize the tertiary structure. An answer is possibly provided by a lattice Monte Carlo model system (Sali *et al.*, 1994); the elements of folding that lead to the resolution of the paradox are the reduced number of

conformations that need to be searched in the semi-compact globule and the existence of many transition states. In the following sections, protein folding pathways are explored by examining the transition from the unfolded state to the native state, via early and late intermediates, as well as the properties of the transition states linking them.

1.2.2.1: General Properties of Protein Folding: The Two-State Equilibrium

At equilibrium, the unfolding transitions of most small proteins are two-state, with only the fully folded (N) and fully unfolded (U) states populated so that N $\leftrightarrow$ U. Intermediates are generally not present as they are energetically less stable than either N or U (Privalov, 1979). However, for some proteins, a third state may exist under some folding conditions (the molten globule state; Section 1.2.1.2, Jennings and Wright, 1993; Uversky, 1993; Uversky and Ptitsyn, 1994). The conformational stability of a small globular protein is the difference in free energy between N and U conformations under physiological conditions $\Delta G(H_2O)$; for small proteins this value is in the order of 5 - 15 kcal / mol (Pace, 1990). The small difference in stability means that the native conformation is stabilized by the equivalent of a few hydrogen bonds, ion pairs or hydrophobic patches (Seckler and Jaenicke, 1992). The native state is therefore sensitive to small changes in conditions, such that the addition of denaturants, heat, change of pH or pressure will affect the equilibrium distribution of N and U. The kinetics of the relationship also means that the folded protein (N) must spontaneously unfold completely under all conditions and then spontaneously refold. From an evolutionary point of view there must be a reason for the small difference in conformational stability and it may be related to protein turnover, function or folding of the nascent protein.

The interactions that stabilize proteins are individually weak but when they act in concert in the absence of competing water, the co-operative effect is greater. Changes in temperature, pH or denaturant concentration usually do not alter the structure of the folded state until a critical threshold is reached. Subsequent small changes then lead to the complete unfolding of the protein and in the transition region the protein is either fully folded or fully unfolded, Two possible reasons for the co-operativity in folding transitions are that there are unfavourable interactions present in partly folded states and that the interactions that stabilize the folded state are co-operative (Creighton, 1990). Thus, breaking one or more interactions of the folded state, destabilizes others, leading to complete unfolding of the protein.

1.2.2.2: Early Events in Protein Folding: Secondary Structure

The early events in refolding of small globular proteins *in vitro* after they have been exposed to denaturing conditions can be monitored by stopped flow methods that allow very rapid mixing and sampling (Labhardt, 1984; Kuwajima *et al.*, 1987; Elöve *et al.*, 1992). When interfaced with circular dichroism (CD) spectrophotometry, folding events can be monitored on the milli-second time-scale. When α-helical dichroism is monitored (at 222 nm), α-helical content has been observed to develop for many proteins within the dead time of the experiment (usually the first 2 - 4 ms), much faster than the formation of the native state (Elöve *et al.*, 1992; Chaffotte *et al.*, 1992a). Elaborate stopped-flow CD experiments have been devised that simultaneously monitor different wavelengths to resolve different kinetic phases and to reconstruct the spectra of the early folding intermediates (Evans and Radford, 1994; Chaffotte *et al.*, 1992a). In these experiments it may be possible to deconvolute early forming intermediates in terms of their secondary structure intermediates.

Another way to explore folding events that occur in the milli-second time-scale is to use pulsed deuterium / hydrogen exchange (Baldwin, 1993). This complex technique relies on the naturally occurring pH-dependent exchange of hydrogen between water and the peptide group amide hydrogens distributed throughout the protein molecule. In these experiments the protein is first unfolded in a ^2H_2O buffer with denaturant. Folding is initiated by rapid dilution to a low denaturant concentration but the pH is set to minimise ^2H - ^1H exchange (pH 4 - 7 and 10°C). After the initial folding time (t_f), between 3 ms and several seconds, ^2H - ^1H exchange is promoted by raising the pH (pH 8 - 10 and 10°C). In this phase the amides that are freely exposed to the solvent will have exchange rates equivalent to small extended peptides, whereas the amides protected by intermediate structure will exchange more slowly. After the pulse time, the pH is finally lowered (pH 4 - 5) to quench exchange and folding to the native state continues. The ^2H - ^1H exchange pattern locked into the protein can then be analysed by determining the ^1H occupancy at each probe amide, determined by two-dimensional NMR (Englander, 1993). This technique is very powerful. In kinetic experiments where the folding time is varied and the pulse pH and pulse time is held constant, plots of ^1H occupancy *versus* the folding time (t_f) for different probe amides display different decay constants, indicating the presence of structured intermediates within milliseconds after the initiation of folding (Roder *et al.*, 1988). In other experiments, where the pulse intensity is varied (pH or time) and the folding time (t_f) is constant, it may be possible to differentiate between parallel and serial folding pathways (Woodward, 1994).

The comparison of data from stopped flow CD spectroscopy and pulsed ^2H - ^1H exchange can provide important information about early folding

intermediates. In the case of hen lysozyme, a pulse labelling experiment showed there to be little protection of ^{2}H - ^{1}H exchange in the dead time of around 3 ms, even though the stopped flow CD experiment suggested that formation of native-like secondary structure was complete within this time (Radford *et al.*, 1992). Similar observations were made for human lysozyme (Hooke *et al.*, 1994) and cytochrome *c* (Elöve *et al.*, 1992). In these situations, it appears the structure formed early in folding is too labile to afford protection from the intense labelling pulse. Pulsed ^{2}H - ^{1}H exchange studies of peptide fragments of regions that would otherwise be protected from exchange early in folding show that the structure in these peptides is particularly labile and does not protect the amides concerned from exchange (Wu *et al.*, 1993; Waltho *et al.*, 1993). However, the structure can be stabilized by tertiary interactions between complementary fragments providing protection (Wu *et al.*, 1993; Waltho *et al.*, 1993; Shin *et al.*, 1993a; Shin *et al.*, 1993b). Thus, stopped flow CD spectroscopy and pulsed ^{2}H - ^{1}H exchange methods demonstrate that early intermediates have an abundant, but rapidly fluctuating, secondary structure.

1.2.2.3: Early Events in Protein Folding: Side Chain Clustering

Another technique used to probe the early events of folding is fluorescence. As fluorescence is sensitive to solvation, it can be used to perhaps answer a major unresolved question about the early folding events, namely at which points in time do hydrophobic side-chain interactions play a role in folding; are they formed concomitantly with secondary structure, or before, or after? Hydrophobic tryptophan residues have been widely used as an intrinsic probe as both the wavelength and intensity of fluorescence are sensitive to the environment (Itzhaki *et al.*, 1994; Elöve *et al.*, 1992; Bastiras and Wallace, 1992). If there is only one tryptophan residue in the protein then fluorescence can be monitored in terms of changes of the

environment (Staniforth *et al.*, 1993; Kiefhaber *et al.*, 1992b). Alternatively, protein engineering can be used to place a unique tryptophan residue at a single site (Khorasanizadeh *et al.*, 1993) or at different sites in the same protein (Staniforth *et al.*, 1993). For cytochrome *c* (Elöve *et al.*, 1992), hen lysozyme (Itzhaki *et al.*, 1994) and ubiquitin (Khorasanizadeh *et al.*, 1993), there is a substantial quenching of fluorescence in the dead time of the experiment, followed by large changes in fluorescence associated with late events of folding. As mentioned above, for each of these proteins, little protection from ^{2}H - ^{1}H exchange by secondary structure was detected in the dead time of the pulse-labelling experiments (Roder *et al.*, 1988; Radford *et al.*, 1992; Briggs and Roder, 1992), suggesting that some association of side-chains occurs at an earlier stage than the formation of structures that protect amides from exchange. If tryptophan residues behave as hydrophobic residues in general, it seems that hydrophobic association or collapse occurs in the first few milliseconds of refolding, as does the formation of an abundant rapidly fluctuating secondary structure, well ahead of the formation of interactions that protect hydrogen exchange.

The hydrophobic dye, 8-anilino-1-naphthalene sulphonic acid (ANS) can be used to measure hydrophobic side-chain association (Mann and Matthews, 1993; Elöve *et al.*, 1992). ANS fluorescence is largely quenched in water but is enhanced on binding to hydrophobic surfaces. For many proteins (Matthews, 1993), ANS fluorescence enhancement is slow, reaching a maximum when partly folded intermediates are populated either transiently or at equilibrium, such as in the molten globule state; finally ANS is excluded in the transition to the native state. For α-lactalbumin and ß-lactamase (Ptitsyn *et al.*, 1990) there is little fluorescence enhancement of ANS in the first few milliseconds of refolding suggesting that there is

little hydrophobic collapse of early intermediates; in these cases fluorescent enhancement develops after the development of secondary structure. However, in the case of hen lysozyme (Itzhaki *et al.*, 1994) and other proteins (Matthews, 1993), fluorescence enhancement is observed in the first few milliseconds of folding. In this situation ANS has been sequestered in a hydrophobic environment suggesting, in accord with the tryptophan fluorescence results, that these early intermediates have substantial secondary structure as well as tertiary structure (Evans and Radford, 1994; Matthews, 1993).

1.2.2.4: Late Events in Folding: Consolidation of Side Chain Interactions

The next stage of folding involves consolidation of side-chain interactions in a manner such that mutual protection of secondary structures from ^2H - ^1H exchange occur in concert. Examples of such secondary association and side-chain consolidation are found in the α-helices of apomyoglobin (Jennings and Wright, 1993), the α-helices of cytochrome *c* (Roder *et al.*, 1988) or between the ß-sheet and α-helix of ubiquitin (Briggs and Roder, 1992), however, for the last mentioned protein the interactions are much faster. ^2H - ^1H exchange studies of the folding of hen and human lysozyme from their GuHCl denatured states has shown that the two proteins, whose sequence differs by 51 amino acids and one insertion for the human lysozyme, have very different folding sequences. In hen lysozyme, the four α-helices which form one domain in the native structure are simultaneously protected from ^2H - ^1H exchange, whereas the other domain, which is mainly ß-sheet, is slower to assemble (Radford *et al.*, 1992). In contrast, human lysozyme which folds much faster than hen lysozyme, forms a stable sub-domain involving the two N-terminal α-helices and the C-terminal 3_{10} helix in the first few milliseconds of folding, in advance of the second two helices of the α-domain and the ß-domain (Hooke *et al.*, 1994).

Although circular dichroism analysis of both proteins demonstrates the presence of extensive secondary structure in the dead time of the experiment, this is accompanied by tertiary interactions, probably involving tryptophan residues, only in human lysozyme (Hooke *et al.*, 1994). In hen lysozyme, many molecules fold via an intermediate where the α-domain amides are protected and the ß-domain amides are not (Radford *et al.*, 1992), yet there are populations where the amides in both domains acquire protection rapidly and simultaneously (Miranker *et al.*, 1993). Thus, protein folding is not necessarily a simple sequential association of sub-domains, but can be a mixture of parallel folding pathways. Another feature of the analysis of late events of folding is the occurrence of far-UV circular dichroism or fluorescence spectra that indicate transient side-chain interactions which do not persist in the native state (Chaffotte *et al.*, 1992a; Itzhaki *et al.*, 1994). These non-native interactions have an unknown role in productive folding.

1.2.2.5: Transition to the Native State

The late stages of folding exhibit a great deal of variability, depending on the protein and the folding conditions. At one end of the scale, ubiquitin (Briggs and Roder, 1992) attains its native structure within 100 milliseconds and at the opposite end of the scale ribonuclease T_1 (Kiefhaber *et al.*, 1992a), and other large proteins, fold in a time-scale of hours. The final stages of folding may involve the fixing of contacts and alignment within subdomains and domains as in hen lysozyme. In this example, the binding of a competitive inhibitor increases with the increasing order of the binding site, at a rate similar to the acquisition of complete amide protection and of a native-like near-UV circular dichroism signal (Radford *et al.*, 1992; Chaffotte *et al.*, 1992a; Itzhaki *et al.*, 1994). Furthermore, organization of tertiary interactions

such as the fixing of aromatic rings and the establishment of ionic interactions takes place at this late stage.

Among the late events of folding, disulphide bond formation and isomerization and proline *cis-trans* isomerization have not yet been discussed. The complexity of the oxidative process involving cysteine residues and their contribution to folding and stability warrant a complete description and are dealt with in a major section of this chapter (Section 1.2.3). However, the covalent aspects of proline *cis-trans* isomerization are discussed in the next section.

1.2.2.6: Proline Isomerization

In a two-state folding reaction, which allows for fully unfolded and fully folded species only, a single kinetic phase of folding is expected. However, for some proteins there are multiple kinetic phases comprised of fast phases, which occur in the millisecond to one second time-scale (which are discussed above), and slow phases which take place in the ten second to one thousand second time-scale. This type of kinetic heterogeneity was first demonstrated for ribonuclease A and the slow phase was attributed to unfolded molecules which differ in their proline imide bond isomerization states (Brandts *et al.*, 1975; Nakano *et al.*, 1993). Furthermore, since the kinetics of proline isomerization are slow, molecules that contain non-native prolyl isomers will remain highly populated and serve as transient kinetic blocks to complete folding (Kiefhaber *et al.*, 1992b).

While most amino acids are linked together by a *trans* amide bond, as in the nascent polypeptide chain, proline amino acids may be linked to the preceding amino acid in the folded protein in either *cis* or *trans* orientations in what is termed an 'imide' bond (Schmid, 1993). Proline *cis* and *trans* equilibria are largely temperature independent so they have a similar enthalpy and are expected to exist

in roughly equal proportions, a situation that occurs in an unfolded protein. However, in native proteins *cis* isomers are less common than *trans* isomers so an unfolded protein therefore contains non-native isomeric states. Thus, in a folding reaction, molecules with proline residues that have the native conformation will fold in the fast kinetic phase. Those molecules with one or more prolines that exist with non-native conformations must wait for isomerization, and will refold in the slow kinetic phase (Brandts *et al.*, 1975).

1.2.2.7: Kinetic Aspects of Folding

In the above discussion, it is clear that proteins do not fold randomly and the native structure arises through a coalescence of different interactions. However, the folded state that results, at least early in folding, may not be the most stable conformation possible but could be the most accessible kinetically. While kinetic and thermodynamic aspects of folding may be compatible, leading rapidly to the native state, other kinetically inaccessible but more thermodynamically stable conformations may exist.

The thermodynamic hypothesis of folding (Anfinsen, 1973) is that native conformations of proteins are at global free energy minima relative to all other states having identical bonded chemistry, such that the folding / unfolding reactions of many small proteins are reversible. If a protein folds under thermodynamic control, then the outcome, the native fold, will be independent of the starting configuration. However, if there are energy barriers in the many-dimensional free energy landscape, and if they are sufficiently high, then molecules may become trapped in local energy minima. Furthermore, some molecules with other conformations may not need to progress over such energy barriers and rapidly access the global energy minimum. In this scheme the folding outcome is dependent on the starting

configuration and the reaction is under kinetic control (Baker and Agard, 1994). While it is difficult to demonstrate that a process is under thermodynamic control, demonstrating kinetic control is much easier as the final state will be dependent on the initial conditions. Examples of kinetic control of folding are the pro region dependent folding of α-lytic protease and subtilisin, and the folding of serpins. Other examples of kinetic control may include the folding of influenza hemagglutinin and luciferase (Baker and Agard, 1994; Baldwin *et al.*, 1993).

The two most extensively studied cases of pro region dependent folding are the serine proteases α-lytic protease and subtilisin. In both cases the pro sequence, which is 166 amino acids long for the α-lytic protease (Silen *et al.*, 1988) and 77 amino acids for subtilisin (Ikemura *et al.*, 1987), is required for proper folding of the mature protease *in vivo*. The pro regions facilitate folding either *cis* as in the natural proenzyme (Ikemura *et al.*, 1987; Silen *et al.*, 1989), or *trans* as part of a separate polypeptide chain (Silen and Agard, 1989). Similar folding properties have also been observed *in vitro* (Baker *et al.*, 1992b; Zhu *et al.*, 1989). These studies have shown that the pro regions do not simply confer stability to the folded state as the mature protease is very stable after the removal of the pro region. Denatured α-lytic protease folds to an intermediate state that has substantial secondary structure but little organized tertiary structure (similar to a molten globule state) and is stable for weeks at physiological pH in the absence of denaturant without any conversion to the native state. However, the addition of the pro region permits the intermediate to rapidly convert to the native state (Baker *et al.*, 1992a). Here, the pro region appears to function by directly reducing the relatively large (>27 kcal / mol) free energy barrier that blocks access to the native state, thereby increasing the rate of folding at least 10^7-fold (Baker *et al.*, 1992a). Furthermore, as both the intermediate and the

native state are stable under identical conditions with no detectable interconversion (Baker *et al.*, 1992a), the folding reaction must be under kinetic and not thermodynamic control as the intermediate is trapped in a local free energy minimum. Similar observations have also been made for subtilisin (Eder *et al.*, 1993; Strausberg *et al.*, 1993) suggesting a similar role for the pro region in folding. Proteases such as the serine protease carboxypeptidase Y (Winther and Sorensen, 1991; Sorensen *et al.*, 1993), the serine protease of *Serratia marcescens* (Ohnishi *et al.*, 1994), the cysteine protease cathepsin L (Tao *et al.*, 1994) and the aspartic proteinase-I of *Rhizopus niveus* (Fukuda *et al.*, 1994) are other examples where the pro region is important in guiding the folding of the mature enzyme. The pro regions of proteases have been called 'intramolecular' or 'co-translational' chaperones in the way that they assist folding. However, these terms may be misleading as pro sequences are highly specific and act as catalysts of folding because they increase the rate of on-pathway folding. In contrast, true chaperones (Section 1.2.3.5) block aggregation or other off-pathway reactions with little substrate specificity. The pro regions of other non-protease proteins such as transforming growth factor-ß1, activin A, insulin and bovine pancreatic trypsin inhibitor facilitate folding to the native state but perhaps by different mechanisms. These proteins are discussed in a later section (1.2.3.6).

A second example of kinetic control of folding is that of the serpin plasminogen activator inhibitor 1 (PAI-1). After synthesis *in vivo* or refolding from denaturant *in vitro* PAI-1 folds to the active state but after a few hours it folds to an inactive latent form (Franke *et al.*, 1990). As the latent form of PAI-1 can be unfolded by denaturants (Heckman and Loskutoff, 1985) or by temperature (Katagiri *et al.*, 1988) and refolded back to the active inhibitory form, the active form is therefore at

the lowest energy state accessible during folding. The fact that PAI-1 slowly converts to the latent form indicates that it is not at the lowest energy state but at a metastable state and that the energy barrier to the latent form is relatively low. If the energy barrier were 1 - 2 kcal / mol higher then there would be no indication that the active form was metastable (Baker and Agard, 1994). While protein folding provides one route to produce a latent form, other serpins, which do not appear to exist in a metastable form but are in a local energy minimum, are proteolytically transformed to the latent form of PAI-1 (Stein and Chothia, 1991).

1.2.2.8: Models of Folding

How does a protein fold? Clearly, a protein does not arrive at the native structure as a result of a random search of all possible conformations, but it does arrive at the native state through different intermediates and transitions that are driven by local or long-range interactions with some degree of cooperativity. While every protein is different and appears to exhibit different folding properties and seemingly different folding pathways, it is also possible that proteins with common folds may have common folding patterns.

Many different theoretical models of protein folding have been proposed such as the jigsaw puzzle model (Harrison and Durbin, 1985), the nucleation / rapid growth model (Wetlaufer, 1973), the diffusion-collision-adhesion model (Karplus and Weaver, 1979) and related models. Two other models of folding, the framework model and the collapse model are introduced here in a little more detail. In the framework model (Baldwin, 1989; Kim and Baldwin, 1990), the unfolded molecule is imagined to form the secondary structure elements which then pack together to form an intermediate compact state which in turn coalesces to establish long range interactions and other tertiary interactions present in the native state. A feature of

this model is the importance of local interactions that drive the formation of the secondary structure early in folding. In the collapse model (Dill, 1985), folding is driven by non-local interactions which drive concurrent secondary structure formation very early in folding. The rapid hydrophobic collapse continues to a state similar to the molten globule state, and in this compact form the final tertiary interactions that stabilize the native state are more likely to be established (Gregoret and Cohen, 1991).

A general scheme for protein folding, proposed by Creighton (1994), probably involves elements of all the above folding models. In this scheme, the unfolded molecules rapidly equilibrate under folding conditions with a few partly folded, marginally stable intermediates, which are also in rapid equilibrium. All the molecules pass through a common slow step, which involves going through a transition state that is a distorted form of the native-like conformation. In this scheme, any intermediates that occur after the rate limiting step are probably very unstable relative to the native form.

1.2.3: The Role of Covalent Interactions: The Disulphide Bond

The preceding sections have concentrated on the role of non-covalent interactions that drive folding, and hence have largely ignored the chemical events of disulphide bond formation and isomerization that lead to the stabilization of the final native fold of a protein.

1.2.3.1: The Complexity of Interactions that Lead to Correct Disulphide Bonds

In the discussions presented above the subtle interplay of forces that drive the formation of the native fold of a protein are complex. The chemical oxidation of cysteine residues that lock the native fold together further complicates our view of

folding. For proteins that spontaneously refold *in vitro* and quantitatively form the correct disulphide bonds, such as ribonuclease (Anfinsen, 1973), the process appears to be thermodynamically controlled. In this example, pairs of cysteine residues are placed together at covalent bonding distances that permit the disulphide bonds to form. The precision by which the native disulphide bonds are established demonstrates that a random process is not involved because a simple calculation reveals that for ribonuclease, which contains eight cysteine residues, there are 105 possible permutations for linking pairs of cysteine residues so that one or more disulphide bonds are formed. The oxidative folding process of ribonuclease and another small globular protein, bovine pancreatic trypsin inhibitor, are well documented because they form their native bonds spontaneously. However, the literature is starting to reveal other proteins that do not quantitatively form their disulphide bonds *in vitro*, and it appears that for these proteins additional factors are required for complete folding. Today, with the advent of protein engineering and the requirement for large amounts of correctly folded protein, the investigation of oxidative folding is of both fundamental and practical importance (Fischer *et al.*, 1993).

As indicated above, much is known about oxidative folding from the study of two small globular proteins, but many questions remain. These questions can be summarized in two broader questions: (i) What are the interactions that lead to the establishment of the correct disulphide bond? and (ii) For proteins that do not quantitatively yield the native disulphide structure *in vitro*, what are the factors that facilitate their correct folding *in vivo*? In the following sections, disulphide chemistry will be explored *in vitro* and *in vivo* and the above questions will be expanded and discussed.

1.2.3.2: Thiol / Disulphide Exchange: In Vitro

Disulphide bond formation *in vitro* from two thiols is a two electron oxidation reaction that requires an oxidant. Molecular oxygen, catalyzed by trace metal ions, will oxidize thiols but the reaction is complex. However, oxidizing equivalents can be transferred from low molecular weight disulphides to protein thiols in a simple disulphide exchange process (Gilbert *et al.*, 1990). Thiol / disulphide exchange occurs via direct attack of a nucleophilic thiolate anion on one of the sulphurs of the disulphide bond. The rate constant for the reaction increases as the basicity of the attacking thiolate nucleophile increases (pK_a increases) and as the basicity of the leaving thiolate decreases (pK_a decreases), therefore the reaction rates will be pH dependent. The pK_a of a typical cysteine sulphydryl group is 8.6, so at pH 8.6 one-half of the thiols will be ionized and therefore reactive.

Protein thiols in a redox buffer (a low molecular weight disulphide (RSSR) and its corresponding thiol (RSH)) first react with the low molecular disulphide forming a mixed disulphide intermediate which in turn undergoes intramolecular displacement to complete disulphide bond formation.

$$
\underset{\text{Reduced Protein}}{\text{Protein} \begin{array}{l} - SH \\ - SH \end{array} + RSSR} \;\;\overset{K_{mix}}{\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}}}\;\; \underset{\text{Mixed Disulphide}}{\text{Protein} \begin{array}{l} - S\text{-}SR \\ - SH \end{array} + RSH} \;\;\overset{K_{int}}{\underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}}}\;\; \underset{\text{Protein Disulphide}}{\text{Protein} \begin{array}{l} - S \\ \;\; | \\ - S \end{array} + 2\,RSH}
$$
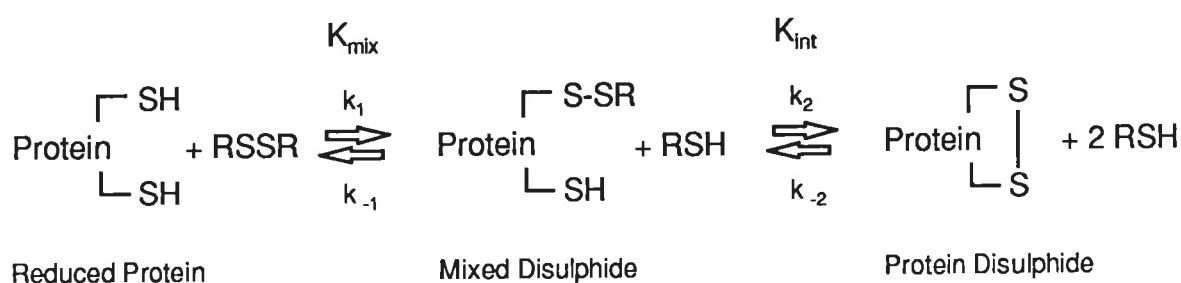
In this reaction, the thiol component of the redox buffer provides the means to reduce protein disulphides, allowing for disulphide rearrangement as the protein reaches its native conformation or to reverse the formation of the mixed disulphide intermediate (Gilbert, 1994). A common redox buffer used is that of oxidized and reduced glutathione (GSSG and GSH respectively). It is used primarily because the

components are known to exist at significant concentrations in biological systems (Saxena and Wetlaufer, 1970) and because it exhibits a suitable redox potential. Another redox buffer that has equilibrium properties similar to the glutathione system is oxidized ß-mercaptoethanol and its corresponding monothiol form (Lees and Whitesides, 1993; Szajewski and Whitesides, 1980). The formation of the mixed disulphide intermediate with glutathione has an equilibrium constant (K_{mix}) near unity. However, the equilibrium constant for the formation of the protein disulphide by intramolecular displacement (K_{int}) varies depending on the energetics of structural changes that accompany the covalent bond formation (Creighton, 1986; Gilbert *et al.*, 1990). For the above equation, under conditions where mixed disulphides do not accumulate significantly at equilibrium, the overall equilibrium constant (K_{ox}) is given by:

$$K_{ox} = \frac{[\text{Protein Disulphide}][\text{RSH}]^2}{[\text{Reduced Protein}][\text{RSSR}]} = K_{mix} K_{int} = \frac{k_1 k_2}{k_{-1} k_{-2}}$$

Under certain conditions, using the glutathione system, when the [GSH]/[GSSG] ratio is less than K_{mix}, and when [GSH] is greater than K_{int}, glutathione can form stable mixed disulphides with the protein thiols. This property greatly complicates studies of regeneration pathways. In contrast to the glutathione redox buffer system, another oxidizing system which employs a cyclic disulphide (oxidized dithiothreitol: DTT^{ox}) does not form stable mixed disulphides because the DTT cyclization reaction is very fast (Cleland, 1964; Lees and Whitesides, 1993; Rothwarf and Scheraga, 1992). This property simplifies oxidative folding studies as the observed rate of protein disulphide formation will reflect the thermodynamic stability of the protein (Gilbert, 1994).[*]

* Refer to Addendum

The equilibrium conversion of a dithiol to a specific disulphide depends on the composition of the redox buffer and the K_{ox} for disulphide formation. For the oxidative folding of a protein, the most stable disulphide bond will form first[*] (the one with the highest K_{ox}) if the composition of the redox buffer ($[GSH]^2/[GSSG]$ for glutathione) is well below the K_{ox}. Thus, at equilibrium, protein structures will contain more and more disulphides in a more oxidizing redox buffer (lower $[GSH]^2/[GSSG]$ values). However, among protein structures that have the same number of disulphide bonds, the most stable disulphide structure[*] will have the highest equilibrium concentration independent of the redox buffer composition. This is because disulphide rearrangement, which is driven by the thermodynamic properties of the structure, does not involve the net use or production of oxidizing equivalents.

1.2.3.3: Thiol / Disulphide Exchange: In Vivo

In aerobic biological systems, the two electron oxidation of two thiols is ultimately linked to the reduction of oxygen. Biologically, the reverse reaction of reducing a disulphide, occurs at the expense of reducing equivalents, generally provided by nicotinamide dinucleotides. As some proteins require oxidized thiols and others require reduced thiols for activity, changes in the thiol / disulphide oxidation state can be employed as a regulatory mechanism (Gilbert et al., 1990). Furthermore, oxidation states vary according to compartmentation in the cell.

A cellular redox buffer is one that is present in both the thiol and disulphide oxidation states that functions to maintain the oxidation state of cellular proteins. The tripeptide glutathione (γ-glutamyl-cysteinyl-glycine) is the most abundant thiol / disulphide redox buffer present in mammalian, plant and some bacterial cells. Other examples of cellular thiol / disulphide redox buffers are thioredoxins and glutaredoxins which are small (8 - 12 kDa) proteins. For glutathione, cytoplasmic

levels are in the order of 1 - 7 µmoles / gram of tissue for reduced glutathione (GSH) and 10-25 nmoles / gram tissue for oxidized glutathione (GSSG). However, the actual levels will vary from tissue to tissue (Harisch and Mahmoud, 1980; Ishikawa and Sies, 1984). Without going into too much detail about the redox status in other compartments in the cell and the maintenance of cellular redox buffers, it is pertinent here to briefly consider the endoplasmic reticulum and the plasma redox states. In the endoplasmic reticulum, where proteins destined for the external medium or cell surface are cotranslationally inserted for modification and secretion, the redox state is much more oxidizing than the cytosol as GSH and GSSG are present at concentrations of approximately 2 mM and 1 mM respectively (Hwang *et al.*, 1992; Freedman *et al.*, 1994), providing a $[GSH]^2/[GSSG]$ value of 4 mM compared to 2 - 5 M for the cytosol (Gilbert *et al.*, 1990; Freedman *et al.*, 1994). Other components in the endoplasmic reticulum involved in oxidative folding are protein disulphide isomerases, an aspect of folding that is discussed in greater detail later in this chapter (Section 1.2.3.5). In rat plasma, the concentration of reduced glutathione and oxidized glutathione are very low, being in the order of 17 ± 4 µM for GSH and 1.3 ± 0.4 µM for GSSG ($[GSH]^2/[GSSG] = 0.2$ mM), a composition that is considerably more oxidizing than the intracellular environment (Lash and Jones, 1985). While cystine and cysteine are present in plasma at about twice the level of glutathione, and form mixed disulphides with glutathione, the oxidation potential in plasma is maintained at low levels (Lash and Jones, 1985).

1.2.3.4: The Effect of Structure

The reactivity of a protein thiol at a particular pH can vary by many orders of magnitude depending on many factors, for example, pK_a, charge and steric factors and entropy. The thiolate anion is the reactive nucleophile in thiol / disulphide

exchange. Thus, the observed rate constant for thiol-disulphide exchange at neutral pH depends on the fraction of the total thiol present as the anion. Any electron-withdrawing substituent or nearby positive charge, that lowers the pK_a of a thiol, will increase the concentration of the thiolate anion at neutral pH (Gilbert *et al.*, 1990). A nearby negative charge should have an opposite effect. The different rate constants observed for reactions between different disulphides and monothiols to proteins or to small molecules (Gilbert *et al.*, 1990; Szajewski and Whitesides, 1980) indicate that local charge (other than that which affects pK_a), steric factors and binding interactions also affect the reactivity of the protein thiol. In small peptides, the equilibrium constants for forming intramolecular disulphides also depends on the number of intervening amino acid residues (m) (Zhang and Snyder, 1989). For these short peptides with less than eight intervening amino acids, disulphides with an even m are favoured over peptides with an odd m. Although the formation of a disulphide bond between adjacent cysteine residues requires a *cis* peptide bond, the equilibrium constant (K_{ox}) for this reaction is greater than that for the formation of disulphide bonds when $m = 1$ or 3 (Zhang and Snyder, 1989). However, as the number of intervening amino acids increases, the difference between odd and even values of m diminishes (Zhang and Snyder, 1991), as does K_{ox} (Gilbert *et al.*, 1990). The decreasing K_{ox} with increasing sequence separation arises from the loss in conformational entropy of the peptide backbone that accompanies disulphide bond formation (Flory, 1956).

In treating the denatured state of proteins as a random coil, Flory (1956) predicted that disulphide bonds increase the free energy of the denatured state by decreasing its configurational entropy. This sort of analysis predicts that the most substantial stabilization occurs for disulphides that link distant parts of the molecule,

an analysis that correlates well with that obtained experimentally for the removal of disulphide bonds from ribonuclease T_1 (Pace *et al.*, 1988). The thermodynamics of this increased stability are complex and involve enthalpic as well as entropic effects. While Doig and Williams (1991) suggest that the dominant effect of disulphide bonds on stability is enthalpic, recent studies conclude that the dominant effect is entropic (Betz, 1993), but add that given the subtleties of protein structure this view is incomplete.

The thermodynamics of disulphide bonds in proteins have been studied using two approaches. One approach involves removing natural disulphide bond(s) by mutagenesis or by blocking cysteines and the other is by engineering novel disulphides into a protein. The mutagenic approach for removing disulphide bonds has been carried out for many proteins including hen lysozyme (Tachibana *et al.*, 1994), human interleukin 6 (Rock *et al.*, 1994), human lysozyme (Kuroki *et al.*, 1992), bovine pancreatic trypsin inhibitor (BPTI) (Goldenberg, 1988; Staley and Kim, 1992), ribonuclease T_1 (Mayr *et al.*, 1994) and ribonuclease A (Talluri *et al.*, 1994). The alternative method of removing disulphide bonds by blocking cysteine residues has been carried out for many proteins such as ribonuclease T_1 (Pace *et al.*, 1988), α-lactalbumin (Ewbank and Creighton, 1993b; Igekuchi *et al.*, 1992) and human interleukin 6 (Rock *et al.*, 1994). A problem with the latter approach is that the reactive thiol is irreversibly chemically modified, often with a large polar group (e.g. carboxymethyl or carboxamidomethyl) which almost certainly affects the packing and stability of the protein (Schwartz *et al.*, 1987). Nevertheless, for ribonuclease T_1 where one of the disulphide bonds was reduced and chemically modified, the conformational stability (free energy) was 3.4 kcal / mol lower than that of the native protein (Pace *et al.*, 1988). Similar studies for BPTI (Schwartz *et al.*, 1987) and α-

lactalbumin (in the presence of Ca^{2+}; Igekuchi *et al.*, 1992) show that the reduction and modification of a single disulphide bond reduces the free energy by 8 kcal / mol and 3 kcal / mol respectively, compared to the native protein. The mutagenic approach to removing disulphides is also problematic as it is difficult to replace cysteine residues with ones that have a similar polarity and packing volume; typical mutations used are cysteine to serine or to alanine. Using this method, the effect of removing a disulphide bond in human lysozyme (Kuroki *et al.*, 1992) results in a protein that is approximately 2 kcal / mol less stable than the native protein, but for ribonuclease T_1 (Mayr *et al.*, 1994) the energy cost is 10 kcal / mol. Although free energy values were not tabulated for other proteins (hen lysozyme (Tachibana *et al.*, 1994), human interleukin 6 (Rock *et al.*, 1994), bovine pancreatic trypsin inhibitor (Staley and Kim, 1992; Darby *et al.*, 1991) and ribonuclease A (Talluri *et al.*, 1994)), it is apparent that the species lacking the disulphide bond are less stable than the parent molecules. Another thermodynamic consideration is the effect of burial of cysteine residues. Many disulphide bonds are buried within proteins (Thornton, 1981; Srinivasan *et al.*, 1990) and at room temperature the burial of a disulphide is favoured over the burial of two cysteines by 0.5 kcal / mol. This difference can account for as much as one-fifth of the increase in stability supplied by a disulphide bond to a protein (Saunders *et al.*, 1993).

The second approach to investigating the thermodynamics of disulphide bonds in proteins is to measure the increase in stability of a protein by incorporating a novel disulphide bond. This approach has met with mixed success. T4 lysozyme, which contains no disulphide bonds, has been engineered by adding a number of novel disulphide bonds. While the disulphide-containing variants are 0.4 - 1.6 kcal / mol more stable than their reduced counterparts, they are actually

0.6 - 1.1 kcal / mol less stable than the wild-type T4 lysozyme (Betz, 1993). A successful example of engineering stabilizing disulphide bonds has been carried out for barnase which, like T4 lysozyme, contains no disulphide bonds in the native protein. In this example, one disulphide bond was engineered into a region of the molecule that was known to pack early in the folding pathway, and another was engineered to link two secondary structures that packed together after the rate limiting step. Both engineered proteins were more stable than the wild-type by 4.2 kcal / mol and 1.2 kcal / mol respectively (Clarke and Fersht, 1993).

Another way of looking at the thermodynamics of the disulphide bond is from the reactivity of protein thiols (see above). Two proteins that vary in their oxidation potential (K_{α}), the active dithiol of disulphide isomerase (7×10^{-5} M) and the 5 - 55 disulphide bond of BPTI (1.1×10^{7} M) (Gilbert, 1994), reflect the effects of favourable and unfavourable interactions that accompany disulphide bond formation. Clearly, disulphide bond formation and the formation of native structures are linked thermodynamically.

1.2.3.5: Catalysis of Oxidative Folding In Vivo

In the early folding experiments which led to the conclusion that folding was thermodynamically controlled, there still remained the disconcerting fact that the time taken for a protein to regain its native structure *in vitro* (in the order of hours) was much longer than was relevant for *in vivo* protein synthesis rates (White, 1967; Epstein *et al.*, 1963). This discrepancy led to the discovery of an enzyme system in the endoplasmic reticulum that when used *in vitro* catalyses the formation of native disulphide bonds at a rate similar to the protein synthesis rate (Goldberger *et al.*, 1963; Venetianer and Straub, 1963). This enzyme later became known as protein disulphide isomerase (PDI). Other catalysts of oxidative folding involve cellular redox

buffers such as reduced and oxidized glutathione (Saxena and Wetlaufer, 1970). However, they are only pertinent for cellular compartments that have the appropriate levels of the particular redox buffer (Section 1.2.3.3). Another pertinent factor relating to the catalysis of oxidative folding involves peptidyl prolyl *cis-trans* isomerase (PPI) which can have a synergistic effect with PDI (Schönbrunner and Schmid, 1992). Lastly, other accessory proteins such as molecular chaperones which serve to suppress unproductive (non-native) folding events may also be involved in oxidative folding, although they do not actually catalyse the thiol / disulphide event. In this section I will discuss the catalysis of oxidative folding *in vivo* by PDI and PPI and consider the possible role of molecular chaperones.

Protein disulphide isomerase (PDI) isolated from bovine liver (Lambert and Freedman, 1983) is a 57 kDa protein located in the lumen of the rough endoplasmic reticulum of cells that actively synthesize disulphide-bonded cell-surface and secretory proteins (Noiva and Lennarz, 1992). Apart from its disulphide isomerase activity, PDI is involved in the N-glycosylation of proteins, proline hydroxylation, and is also a part of the microsomal triglyceride transfer protein complex (Noiva and Lennarz, 1992; Bulleid, 1993). The majority of PDI is found as a homodimer with no interchain disulphide bonds. PDI is also found in other mammalian cells and *Saccharomyces cerevisiae* (Farquhar *et al.*, 1991; Scherens *et al.*, 1991). A prokaryotic equivalent, located in the periplasm of *Escherichia coli* (DsbA) is also required for the catalysis of disulphide bonds (Bardwell *et al.*, 1991; Ostermeier and Georgiou, 1994). The cloning and sequencing of PDI revealed two internal sequence repeats with similarity to thioredoxin (Edman *et al.*, 1985). This repeat sequence -Trp-Cys-Gly-His-Cys-Lys- (-WCGHCK-) contains the redox active cysteine residues, as chemical modification of these cysteines completely abolishes its

activity (Hawkins and Freedman, 1991). Furthermore, site directed mutagenesis of a single cysteine in either repeat sequence reduces its activity by 50 % demonstrating that either repeat sequence can function independently (Vuori *et al.*, 1992). The same active-site domain is also found in other proteins such as the ß-subunits of follicle stimulating hormone and leuteinizing hormone (Boniface and Reichert, 1990), and similar domains have been found in calcium-binding protein-2 (Van *et al.*, 1993); the endoplasmic reticulum protein, ERp72 (Mazzarella *et al.*, 1990); *EUG1* of *Saccharomyces cerevisiae*, (Tachibana and Stevens, 1992); and other endoplasmic proteins (Gething and Sambrook, 1992). However, each of these proteins are much less efficient catalysts than either PDI or thioredoxin.

Although PDI is a far better oxidant than thioredoxin with oxidation potentials (K_{ox}) of 70 µM and 2 M respectively, the differences do not reside solely in the active-site sequences (-WCGHCK- and -WCGPCK- respectively) as mutations around the active-site thiols of PDI have only a marginal effect on activity (Lu *et al.*, 1992). However, if the proline of thioredoxin active-site is mutated to a histidine, thioredoxin becomes more PDI-like by a factor of 10 (Krause *et al.*, 1991; Lundström and Holmgren, 1993). If the same proline is mutated to a serine (Lin and Kim, 1991) thermodynamic studies reveal that the net effect is to increase the stability of the reduced form compared to parent molecule, and the stability of the oxidized forms remain the same. Furthermore, amino acids around the active site have no strong effect on the redox properties in small model peptides (Siedler *et al.*, 1993), so the active-site structure in the parent molecule must be maintained by conformational restraints imposed by the whole structure.

In the relatively oxidizing environment of the lumen of the endoplasmic reticulum provided by glutathione (4 mM; Section 1.2.3.3), the active-site of PDI is

predominantly reduced and active. *In vitro*, PDI catalyses the formation and rearrangement of disulphide bonds of ribonuclease A and bovine pancreatic trypsin inhibitor (BPTI), not by altering the mechanism of folding but by accelerating the process (Lyles and Gilbert, 1991b; Creighton and Freedman, 1980; Huth *et al.*, 1993). In this oxidative folding process, a redox buffer is required to maintain PDI in a reduced state (Creighton and Freedman, 1980; Lyles and Gilbert, 1991a). This is because at different concentrations of glutathione, where the $[GSH]^2/[GSSG]$ value is greater than 150 µM, PDI accelerates the formation of disulphide bonds by a constant factor compared to the uncatalyzed reaction. However, under more oxidizing conditions (low $[GSH]^2/[GSSG]$ values) the PDI active-site becomes oxidized and loses its catalytic effectiveness (Lyles and Gilbert, 1991a). Furthermore, PDI has the remarkable property of being able to rescue kinetically trapped folding intermediates (Weissman and Kim, 1993). Although PDI has a small effect on the oxidative folding rate of BPTI disulphide intermediates that readily progress to the native fold, PDI dramatically increases the rate of folding of kinetically trapped intermediates in which the structure impedes further progress. In this situation, PDI appears to access buried thiols in a protein that has already gained substantial structure. Thus, an important function of PDI may be to provide an *in vivo* mechanism to rescue 'off-pathway' folding events.

A second protein implicated in *in vivo* oxidative folding, which is also found in the endoplasmic reticulum, is the cyclophillin-type peptidyl prolyl *cis-trans* isomerase (PPI). PPI accelerates the folding of proteins by facilitating the rate-limiting isomerization of non-native proline conformations (Section 1.2.2.6). When PPI is added to denatured and reduced ribonuclease T_1, which contains two disulphide bonds and four *cis* prolyl peptide bonds in the native protein, the effect is to

accelerate oxidative folding (Schönbrunner and Schmid, 1992). This effect is probably indirect as the molecules that contain the native-like *cis* prolyl isomers will form the correct disulphides at the fastest rate, and those molecules with one or more prolines that exist with non-native *trans* conformations must wait for isomerization, and hence will refold in a slower kinetic phase unless PPI accelerates the isomerization (Brandts *et al.*, 1975). In the absence of PPI, the addition of PDI increases the rate of oxidative folding of ribonuclease T_1, and when increasing amounts of PPI are added, the oxidative folding rate is enhanced (Schönbrunner and Schmid, 1992). It is possible that unfolded or partially folded molecules with the correct *cis* prolyl isomers are better substrates for catalysis by PDI; alternatively, chaperone-like interactions are involved (see later in this section). Whichever mechanism is involved, the synergistic effect of the two enzymes will be most pronounced when proline isomerization and disulphide bond formation occur at similar rates (Schönbrunner and Schmid, 1992).

Finally, other accessory proteins such as molecular chaperones which serve to suppress unproductive (non-native) folding events may also be implicated in oxidative folding, although they do not actually catalyse the thiol / disulphide event. It is not the purpose here to review the explosion of research activity related to molecular chaperones in general but to focus on a small group of chaperones contained in the endoplasmic reticulum and how they may relate to oxidative folding. Molecular chaperones are defined as a family of unrelated classes of protein that mediate the correct assembly of other polypeptides, but are not themselves components of the final functional structure nor do they necessarily possess steric information specifying assembly (Ellis, 1987). A subset of molecular chaperones called 'heat shock proteins' (hsp), which are constitutively and abundantly expressed

under normal growth conditions and over-expressed under conditions of stress, include hsp70 (Georgopoulos and Welch, 1993). A hsp70 homologue present in the endoplasmic reticulum of mammalian cells is the glucose-regulated protein (Grp78: Pouysségur *et al.*, 1977) which is also known as the immunoglobulin heavy chain binding protein (BiP: Haas and Wabl, 1983). In the endoplasmic reticulum of *Saccharomyces cerevisiae*, the grp78 / BiP chaperone equivalent is KAR2 (Rose *et al.*, 1989). Also present in the endoplasmic reticulum is grp94, a glycosylated member of the hsp90 family. However, its role in protein folding in the endoplasmic reticulum is less clearly defined than the hsp70 homologues (Freedman *et al.*, 1994). During translocation into the endoplasmic reticulum grp78 / BiP associates with a number of proteins during their folding and assembly (Bole *et al.*, 1986). For example, *in vitro* grp78 / BiP associates with the non-glycosylated form of yeast invertase but not with the processed protein (Kassenbrock *et al.*, 1988). Furthermore, misfolded forms of influenza virus haemagglutinin which have correct glycosylation patterns and are retained in the endoplasmic reticulum, induce the synthesis of grp78 / BiP (and grp94) (Kozutsumi *et al.*, 1988). Using a stable mixed disulphide between glutathione and the proteins ribonuclease T_1 or α-lactalbumin to mimic partially disulphide folded intermediates, grp78 / BiP was shown to bind to the mixed disulphides but not to the fully folded proteins (Ruoppolo and Freedman, 1994). Thus, grp78 / BiP acts as a classical chaperone in being able to distinguish between unfolded and folded polypeptides, and does not interact with the native form.

PDI may also have a chaperone-like activity. In a series of *in vitro* experiments, PDI was added at various levels to denatured and reduced lysozyme (Puig and Gilbert, 1994). Under certain conditions lysozyme folds correctly, but

significant off-pathway aggregation occurs reducing the efficiency of folding. The effect of PDI was to either increase or decrease the yield of the native protein depending on the concentrations of lysozyme and PDI or the order in which they were combined (Puig and Gilbert, 1994). However, in the presence of a large molar excess (5 - 10-fold) of PDI when denatured lysozyme is diluted to initiate refolding, PDI prevents aggregate formation and promotes correct folding. Another more convincing observation of chaperone-like behaviour is that PDI associates with misfolded mutant human lysozyme molecules *in vivo* but not with the wild-type protein (Otsu *et al.*, 1994). PPI may also act as a chaperone as it can prevent the very early aggregation of carbonic anhydrase by binding to an aggregation-prone intermediate well ahead of the slow isomerase activity (Freskgård *et al.*, 1992).

In summary, the concerted action of molecular chaperones, PPI and PDI activity and the oxidizing environment serves to drive the oxidative folding of nascent proteins inserted into the endoplasmic reticulum. The premise that there is sufficient information in the amino acid sequence of a protein to determine the fold and that thermodynamics control the folding process are still applicable in the cellular environment. However, it may be that the cellular environment serves to overcome non-productive folding pathways and kinetic traps that can occur for proteins that have been observed *in vitro*, thereby facilitating correct oxidative folding. Relatively little is known about the cellular interactions that facilitate folding and they are currently the focus of intense research.

1.2.3.6: The Effect of the Pro Sequence

In addition to the pro regions of proteases that serve to overcome kinetic traps during folding (Section 1.2.2.7), the pro regions of other proteins are also involved in folding, but perhaps in a different manner. An early example of the requirement for a

pro sequence in folding is found in insulin (Steiner and Clark, 1968). The mature form of insulin is a disulphide-linked heterodimer which refolds inefficiently *in vitro*. The pro sequence however, which links the two chains, enables the protein to fold more efficiently (as a monomer) before it is excised in a later processing step. Other examples of pro region assisted folding are found in the biosynthesis of transforming growth factor-ß1 and activin A, two structurally related disulphide-linked homodimers synthesized as large precursors (Gray and Mason, 1990). In these examples, the mature protein was only secreted as its biologically active homodimer when the pro region was expressed either *cis* as *in vivo* or *trans* as a co-expressed protein. These pro-proteins appear to have analogous properties to the pro-enzymes discussed above (Section 1.2.2.7). The pro sequence of BPTI also appears to catalyse the rate of folding *in vitro* compared to the mature form and acts via a tethered thiol group in the N-terminal pro region (Weissman and Kim, 1992). However, another study using microsomes for the expression of pro-BPTI shows that the pro sequence does not substantially alter the folding rate, nor does a mutation of the N-terminal pro region cysteine to a serine (Creighton *et al.*, 1993). In this example, the fate of pro-BPTI is similar to that of BPTI in that a significant portion of the folding products end in a kinetically trapped form. Moreover, PDI overcomes the kinetic trapped form of pro-BPTI, as it does for BPTI, and catalyses the formation of the native fold (Creighton *et al.*, 1993; Weissman and Kim, 1993). The latter example provides a cautionary tale; the synthesis of a protein as a pre-pro-protein does not necessarily imply that the pro sequence of a protein affects the folding of the mature protein, it may have other roles in the biosynthesis of the protein. For example, the N-terminal pro sequence of barnase which is an unstructured 13-amino acid extension of the mature protein does not affect the folding of the protein nor does it affect binding at the active site.

Although the chaperone GroEL retards the folding of pro-barnase by a factor of 2 compared to the wild-type enzyme, the function of the pro sequence *in vivo* is unknown (Gray *et al.*, 1993).

1.2.3.7: Oxidative Folding Pathways

In this section, small proteins are organized into structural groups and are briefly considered in terms of their oxidative folding properties. I shall restrict this discussion to a number of proteins with known *in vitro* oxidative folding pathways that fall into the class of small proteins in the structural classification of proteins (SCOP) database (Murzin *et al.*, 1995). These include plant inhibitors of proteinases (which contain a beta-hairpin with two adjacent disulphides), the thrombin inhibitor-like fold (nearly all-beta), the epidermal growth factor-like module fold (nearly all-beta), the BPTI-like fold (alpha plus beta: where the α-helices and ß-strands are largely segregated), and the insulin-like fold (nearly all-alpha).

In the class of small proteins, the plant inhibitors of proteinases are a subset (superfamily) of the fold which contains a beta-hairpin with two adjacent disulphides. This group includes the 28 amino acid, three disulphide *Ecballium elaterium* trypsin inhibitor II (EETI II), the structurally similar potato carboxypeptidase inhibitor (PCI) and trypsin inhibitors from the bitter gourd and squash. Reduced and denatured PCI, which contains 39 amino acids and three disulphides, refolds spontaneously *in vitro* to its native structure (Chang *et al.*, 1994). The oxidative folding pathway reveals that PCI proceeds through an initial stage of non-specific disulphide formation (packing), followed by disulphide rearrangement of partially packed intermediates to the native structure (consolidation). This process involves a sequential flow of reduced PCI through heterogeneous one-, two- and three-disulphide species. Furthermore, at least 23 of the 75 possible one-, two- and three-disulphide isomers

were observed. The final heterogeneous three-disulphide group undergoes rearrangement to the native structure in a rate-limiting step under all redox conditions investigated, including folding conditions in the presence of protein disulphide isomerase (Chang *et al.*, 1994). The slow consolidation process was attributed to the exposure of hydrophobic residues in the native protein or the difficulty of forming the unusual 'knotted' disulphide topology. Nevertheless, quantitative yields of the native species were obtained. In contrast, the oxidative folding of reduced and denatured EETI II revealed a simple pathway which involved the transient accumulation of a single native-like two-disulphide intermediate (Le-Nguyen *et al.*, 1993). In this example, the last disulphide bond to form is the one that establishes the 'knotted' topology and occurs in a rate limiting step. Apart from the number of intermediates involved, the main difference between the two trypsin inhibitors appears to be the relative hydrophobicity of the native form compared to the reduced form.

The thrombin inhibitor fold is a unique protein fold which includes hirudin, a 65 amino acid, three-disulphide thrombin-specific inhibitor isolated from the leech *Hirudo medicinalis*. It is the most potent thrombin inhibitor known through its dual binding to the catalytic site and to the fibrinogen recognition site of the enzyme. The N-terminal domain (49 amino acids) which binds to the catalytic site is stabilized by the three disulphide bonds which are essential for inhibitory activity. In a number of studies (Chatrenet and Chang, 1992; Chatrenet and Chang, 1993; Chang, 1993; Chang, 1994), at least 60 of the 75 possible one-, two- and three-disulphide isomers were observed. Furthermore, the initial stages of folding appear to follow non-specific disulphide bond formation (packing) to a number of misfolded species which then rearrange (consolidate) to the native form. This packing and

consolidation is similar to that observed for PCI (Chang *et al.*, 1994). However, unlike PCI, the latter stage of oxidative folding is responsive to the redox potential of the refolding buffer (Chang, 1994).

The epidermal growth factor module fold is present in a number of proteins which include E-selectin, factor-IX, factor-X, transforming growth factor-α, plasminogen activator, epidermal growth factor (EGF) and others. However, the oxidative folding properties have only been studied for human EGF. Human EGF is a 53 amino acid protein that contains three disulphide bonds. In a study to elucidate the effect of N-terminal truncated derivatives of EGF, five variants of EGF were synthesized, each containing one less N-terminal amino acid (Shin *et al.*, 1994). The resultant peptides quantitatively yielded only the native disulphide bonded form when oxidized in air. Another more detailed study of the oxidative folding of EGF has recently been published (Chang *et al.*, 1995b) that confirms that reduced EGF quantitatively yields the native isomer. However, the analysis of well-populated intermediates indicates that regardless of whether the intermediates contain native or non-native disulphide bonds, they do not necessarily represent the productive species and specify the folding pathway, as has been proposed for the oxidative folding of BPTI. This topic is the source of some controversy in the literature and is not covered in this section. However, it is pertinent to work presented later in this thesis and will be discussed further in the relevant chapters.

Another disulphide-rich fold; the BPTI-fold includes a number of snake dendrotoxins, Alzheimer's amyloid B-protein precursor, a trypsin inhibitor from sea anemone, tick anticoagulant protein and bovine pancreatic trypsin inhibitor. By far, the best studied of these proteins is BPTI which contains 58 amino acids and three disulphide bonds. In this section, only the most general features of the folding

pathway will be introduced. The *in vitro* oxidative folding mechanism of BPTI, starting with the fully reduced molecule quickly establishes one native disulphide bond, followed by a second native disulphide bond. However, the order in which the disulphides are connected influences the folding outcome and can lead to a kinetically trapped two-disulphide species. Furthermore, the productive native-like two-disulphide intermediate rearranges to non-native-like two-disulphide species in a fast off-pathway reaction. Nevertheless, the kinetics of the pathway indicate that a native-like two-disulphide species rearranges to another native-like two-disulphide species in a rate limiting step which then quickly forms the last disulphide bond forming the biologically active molecule (Gilbert, 1994).

The last group of proteins covered in this section are those that contain the insulin-like fold. This group (family) includes bombyxin-II, relaxin, insulin and insulin-like growth factors. The insulin-like fold is characterized by three α-helices stabilized by three conserved disulphide bonds. Apart from the insulin-like growth factors (IGFs), the mature proteins contain separate A and B chains. For bombyxin-II, relaxin and insulin the proteins are folded *in vivo* as their pro-proteins which link the two chains. After correct folding *in vivo*, the pro sequence is excised by cellular proteases and then exported (Csorba, 1991). *In vitro* studies of the oxidative folding of pro-insulin show that the recovery of the correctly folded molecule is much more efficient than the oxidative folding of the separate chains (Steiner and Clark, 1968). The analogous pro sequence of the IGFs is not removed *in vivo* and the mature peptide contains N- and C-terminal extensions to the equivalent A and B chains of the other members of the family. The *in vitro* studies of the oxidative folding of IGF-I show a surprising result in that two major three-disulphide isoforms are present at equilibrium, irrespective of the refolding

conditions used (Hejnæs *et al.*, 1992; Iwai *et al.*, 1989; Forsberg *et al.*, 1990; Samuelsson *et al.*, 1991). The oxidative folding pathway of IGF-I has been partly characterized (Hober *et al.*, 1992; Miller *et al.*, 1993) and these studies demonstrate that the two isoforms have a similar thermodynamic stability and *in vitro*, at least, exhibit a bifurcating oxidative folding pathway. As only the native form of IGF-I has been found *in vivo*, this suggests that some factor that stabilizes the mis-folded form during *in vitro* refolding is overcome *in vivo* by the cellular machinery. Research into the oxidative folding of IGF-I is at its infancy and little is known about its folding either *in vitro* or *in vivo*. The available literature relating to the oxidative folding of IGF-I is dealt with in greater depth in the appropriate chapters of this thesis and in the General Discussion.

The above inspection of the oxidative folding of small single domain proteins paints a picture of molecules that undergo a search of many possible disulphide arrangements which then rearrange to the native fold in a thermodynamically controlled process. This process is analogous to packing and consolidation that occurs very early in the folding of non-disulphide containing proteins.

1.3: THE AIMS OF MY RESEARCH PROJECT

The investigation of folding covered in previous sections has utilized the novel properties of individual proteins to gain a picture of folding as a whole. It has become apparent that the fundamental forces that govern folding are applicable in general, but each protein folding case needs to be considered on its own and in the context of folding *in vivo*.

IGF-I is an unusual protein in the sense that it falls in between the group of proteins that can quantitatively recover their native fold *in vitro* and those that clearly

are unable to form native disulphide bonds without the help of the cellular machinery that overcome non-productive events. This property therefore highlights the potential of this molecule to provide a wider understanding of folding from the studies of the factors that influence the formation of the native disulphide bonds. In this thesis, I will investigate the effect of mutations and an N-terminal extension on the oxidative folding pathway of IGF-I, look at the structure of the stable mis-folded molecule in terms of its biological interactions, examine the kinetics of the late stages of oxidative folding and finally attempt to dissect the folding pathway of a mutant of IGF-I. In the course of this work I will discuss some folding problems pertaining to oxidative folding in general and IGF-I in particular.

CHAPTER 2

MATERIALS AND METHODS

2.1: MATERIALS

2.1.1: General Chemicals and Reagents

General chemicals and reagents were purchased from the following suppliers: 2-hydroxyethyl disulphide and Tris: Aldrich Chemicals, Milwaukee, WIS, U.S.A.; trichloroacetic acid: Ajax Chemicals, Auburn, NSW; EDTA: BDH Laboratory Supplies, Poole, U.K.; dithiothreitol and Hepes: Boehringer, Mannheim, Germany; oxidized and reduced glutathione: ICN Biochemicals, Cleveland, OH, U.S.A.; urea: Merck, Darmstadt, Germany; porcine pepsin (EC 3.4.23.1; 3500 U/mg protein) and protamine sulphate sodium azide and activated charcoal: Sigma Chemical Company, Castle Hill, NSW, Australia. Pharmacia Phastsystem™ isoelectric focusing (IEF) gels pH 3-9 and IEF standards with a pI range of 5.20 to 10.25 were purchased from Pharmacia Biotech, Uppsala, Sweden. All reagents were of analytical reagent grade or better. MilliQ™ water (Millipore, Bedford, MA, U.S.A.) was used throughout.

2.1.2: Radiochemicals

Carrier-free $Na^{125}I$, L-[4,5-^{3}H]Leucine (specific activity 71 Ci / mmole) and Iodo[2-^{3}H]acetic acid (specific activity 172 mCi / mmole) were obtained from Amersham Australia Pty. Ltd., Sydney, N.S.W., Australia. Recombinant IGF-I was iodinated to specific activities of 30-40 Ci/g with Chloramine-T (Van Obberghen-Schilling and Pouysségur, 1983). *The iodinated IGF-I was prepared by S. Knowles, Cooperative Research Centre for Tissue Growth and Repair, Adelaide, SA, Australia.*

2.1.3: Proteins and Peptides

2.1.3.1: Recombinant Proteins

Recombinant IGF-I, [Arg³]IGF-I, and [Met¹]pGH(1-11)-Val-Asn-[Arg³]IGF-I (Long-[Arg³]IGF-I) were supplied by GroPep Pty. Ltd., the commercial arm of the Cooperative Research Centre for Tissue Growth and Repair, Adelaide, S.A., Australia. These proteins were either provided as crude inclusion bodies, crude refolded fusion protein or as purified receptor grade material. Refolded [Met¹]pGH(1-11)-Val-Asn-IGF-I (Long-IGF-I), [Gly³]IGF-I and [Met¹]pGH(1-11)-Val-Asn-[Gly³]IGF-I (Long-[Gly³]IGF-I) (Francis *et al.*, 1992) were kindly provided by G. L. Francis, Cooperative Research Centre for Tissue Growth and Repair, Adelaide, S.A., Australia. Inclusion bodies containing [Met¹]pGH(1-11)-Val-Asn-[Lys⁹]IGF-I (Long-[Lys⁹]IGF-I) were kindly provided by B. A. Magee, also from the Cooperative Research Centre for Tissue Growth and Repair, Adelaide, S.A., Australia.

2.1.3.2: Chemically Synthesized Peptides

A chemically synthesized peptide containing the sequence Met-Phe-Pro-Ala-Met-Pro-Leu-Ser-Ser-Leu-Phe-Val-Asn-Gly-Pro-Arg-NH$_2$ was purchased from Chiron Mimotopes Peptide Systems, Clayton, VIC, Australia. The manufacturer showed the peptide to be greater than 77 % pure as analysed by mass spectrometry.

2.1.3.3: Native Proteins

Two vials of bovine IGF-I purified from colostrum (Francis *et al.*, 1988), each containing 20 µg of protein, was generously provided by G. L. Francis, Cooperative Research Centre for Tissue Growth and Repair, Adelaide, S.A., Australia.

2.1.4: Chromatography Materials

2.1.4.1: High Performance Liquid Chromatography

Pre-packed reverse-phase high performance liquid chromatography (HPLC) columns were obtained from Applied Biosystems, Foster City, CA, U.S.A. (C_4 Aquapore™: 7μm particle size and 300 Å pore size); J. T. Baker Co., Phillipsburg, NJ, U.S.A. (Bakerbond Wide-Pore Polyethyleneimine: 5 μm or 15 μm particle size, 300 Å pore size) and Millipore-Waters, Milford, MA, U.S.A. (C_4 DeltaPak™: 15 μm particle size and 300 Å pore size, and C_{18} NovaPak®: 4 μm particle size).

Acetonitrile (HPLC grade) was either purchased from Ajax Chemicals Pty. Ltd., Auburn, NSW, Australia or from BDH Laboratory Supplies, Poole, U.K.; propan-1-ol (HPLC grade) was purchased from EM Science, Gibbstown, NJ, U.S.A.; trifluoroacetic acid was purchased from BDH Laboratory Supplies, Poole, U.K. and heptafluorobutyric acid was obtained from Beckman Instruments, Palo Alto, CA, U.S.A. All HPLC buffers were filtered through a 0.22 μm GV filter (Millipore, Bedford, MA, U.S.A.) prior to use.

2.1.4.2: Low Pressure Chromatography

The chromatography columns were obtained from Pharmacia Biotech, Uppsala, Sweden. Sephadex G-25 (M) and SP-Sepharose Fast Flow were also purchased from Pharmacia Biotech, Uppsala, Sweden. Silica C_{18} LC (50 μm particle size, 100 Å pore size) was supplied by Amicon, Danvers, MA, U.S.A. The Sephadex G-25 (M) and SP-Sepharose Fast Flow chromatographic media were packed in the columns according to the manufacturers' instructions. Before and after each use, the Sephadex and the Sepharose columns were sanitized by pumping 1 M NaOH

through the column (0.5 column volumes) followed by MilliQ™ water (3 or 4 column volumes).

2.1.5: Cell Culture Materials

Rat L6 myoblast cells (ATTC CRL 1458) were purchased from American Type Tissue Culture, Rockville, MD, U.S.A. The cells were cultured at 37°C, 5 % CO_2 in Dulbecco's Modified Eagle's Medium (Gibco, Glen Waverly, NSW, Australia) containing 10 % (v/v) foetal bovine serum (FBS: Cytosystems Pty. Ltd., Castle Hill, N.S.W., Australia) and 100 μg / ml streptomycin and 60 μg / ml penicillin (both from Glaxo, Boronia, VIC, Australia). Hanks' salts were purchased from Gibco, Glen Waverly, NSW, Australia and the plasticware used for cell culture was purchased from Nunc, Roskilde, Denmark. *Routine maintenance and subculturing of cells was carried out by staff in the Tissue Culture Facility at the Cooperative Research Centre for Tissue Growth and Repair, Adelaide, S.A., Australia.*

2.2: METHODS

2.2.1: Preparation of Recombinant IGF-I Analogue Clones

2.2.1.1: Plasmid Construction and Bacterial Strain

The expression vectors of IGF analogues used were kindly provided by colleagues at the Cooperative Research Centre for Tissue Growth and Repair, Adelaide, S.A., Australia. The downstream processing of the IGF fusion proteins [Met^1]pGH(1-11)-Val-Asn-IGF-I (Long-IGF-I), [Met^1]pGH(1-11)-Val-Asn-[Gly^3]IGF-I (Long-[Gly^3]IGF-I), [Met^1]pGH(1-11)-Val-Asn-[Arg^3]IGF-I (Long-[Arg^3]IGF-I) and

[Met^1]pGH(1-11)-Val-Asn-[Lys^9]IGF-I (Long-[Lys^9]IGF-I) was facilitated using an efficient

Escherichia coli expression system developed by this group (Francis *et al.*, 1992).

The vector used in this expression system is based on the parent plasmid pKT52

(Vize and Wells, 1987) which contains an origin of replication (*ori*), the *trc* promoter,

an optimized ribosome binding site, the modified porcine growth hormone

(pGH) / human-IGF-I cDNA sequences (Vize and Wells, 1987; King *et al.*, 1992;

Francis *et al.*, 1992) and a transcription termination sequence (*rrnT1T2*) (Figure 2.1).

Codons were optimized for expression in *E. coli* (Sproat and Gait, 1985). When

propagated in a strain of *E. coli* which overproduces the *lac* repressor (JM101 (lacI^q);

Messing, 1979), addition of isopropyl-ß-D-thiogalactoside (IPTG) induces high-level

expression of the IGF-I fusion protein as inclusion bodies.

2.2.1.2: Fermentation and Isolation of Inclusion Bodies

Inclusion bodies containing the required IGF fusion proteins were prepared by

colleagues at the Cooperative Research Centre for Tissue Growth and Repair,

Adelaide, S.A., Australia. The bacterial strain containing the required plasmid was

selected as a single colony and used to inoculate a 20 ml starter culture containing

ampicillin. After the absorbance at 600 nm reached 50, this, in turn, was used to

inoculate a 15 litre fermenter (Chemap, Volketswil, Switzerland). The bacteria were

grown in a medium without ampicillin under automated control at 37°C, 55% pO_2 ,

pH 7.0 with glucose feed. Expression was induced on reaching an absorbance value

of 60 with 0.32 mM IPTG for 5 hours. Inclusion body formation was monitored by

phase-contrast microscopy. Cells were harvested by centrifugation and

homogenized at 62 MPa and the inclusion bodies were collected from the

homogenate by centrifugation. The inclusion bodies were washed by suspension in

a solution containing 20 mM Tris, 5 mM EDTA and 0.02% (w/v) lysozyme at pH 8.0

Figure 2.1 The Expression Vector Containing Long-IGF-I

Schematic representation of the construct engineered for the expression of [Met^1]pGH(1-11)-Val-Asn-IGF-I (Long-IGF-I). The vector used in this expression system is based on the parent plasmid pKT52 which contains an origin of replication (*ori*), the *trc* promoter, an optimised ribosome binding site, the modified porcine growth hormone (pGH)-IGF-I cDNA sequences and a transcription termination sequence (*rrnT1T2*). Restriction sites used to facilitate further site-directed mutagenesis (EcoR1, Hpa I, Hind III) are shown.

[Met^1]pGH(1-11)-Val-Asn [human IGF-I]

AGGAGGTAATATATG

Hpa I   Hind III

EcoR1

trc

rrnT1T2

pKT52

ori

amp
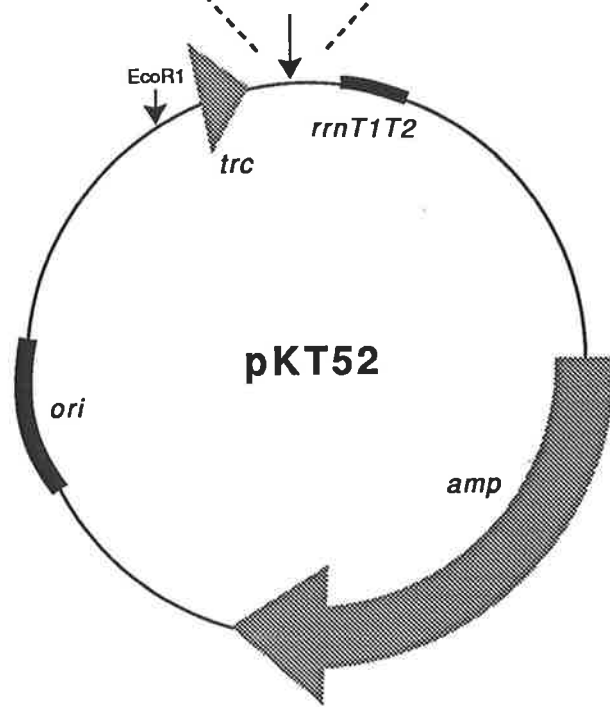
for 3h at 20°C. Finally, the inclusion bodies were collected by centrifugation and the wet paste stored at -80°C (King *et al.*, 1992).

2.2.2: Purification Protocols

2.2.2.1: Chromatography Instruments

Preparative reverse-phase high performance liquid chromatography was either carried out using a Waters Delta Prep 3000 chromatography system which employs a single pump and a low pressure mixer (Millipore-Waters, Milford, MA, U.S.A.), or using a GBC Scientific HPLC (GBC Scientific, Dandenong, VIC, Australia) two-pump system and a high pressure mixer. The same GBC Scientific HPLC system was used for analytical microbore reverse-phase HPLC but using an automated sample injector for multiple analyses. A Pharmacia-LKB HPLC system (Pharmacia Biotech, Uppsala, Sweden) similar to the GBC HPLC system was used for the evaluation of purity of IGF preparations. All chromatography was carried out a room temperature and the elution of protein was monitored using variable wavelength detectors either at 214 nm or 280 nm.

Low pressure liquid chromatography was carried out using a BioPilot™ preparative chromatography system (Pharmacia Biotech, Uppsala, Sweden) which employs dual syringe pumps. The elution of proteins was monitored using a fixed wavelength detector at 280 nm.

2.2.2.2: Purification of Fusion Proteins from Inclusion Bodies

Washed inclusion bodies, in 20 g batches, were dissolved by stirring at 10 % (w/v) in a 8 M urea buffer containing 40 mM glycine, 0.1 M Tris, 1 mM EDTA and 40 mM dithiothreitol, pH 9.0. After 15 minutes solubilization the mixture was

centrifuged at 30,000 g for a further 15 minutes and then filtered through a 1 µm filter (Multigrade GMF 150: Whatman, Maidstone, U.K.). The solubilized inclusion bodies were then desalted at 40 ml / min. on a column packed with Sephadex G-25(M) (10 cm diameter x 30 cm long), previously equilibrated with the above solubilization buffer but containing 1.6 mM dithiothreitol. After dilution and refolding (Sections 2.2.4.4 and 2.2.4.5), the crude refolded mixture was loaded on a SP-Sepharose Fast Flow (FFS) cation-exchange column to remove endotoxins. The FFS column (5cm diameter x 11.5 cm long: 0.23 l volume) was previously primed with 1 M acetic acid and equilibrated with a buffer containing 50 mM ammonium acetate, pH 4.8 and 8 M urea. The protein bound to the column was washed free of UV absorbing material using approximately 10 column volumes of the above equilibration buffer. Protein was then eluted with a gradient of 0 to 1 M NaCl over 30 minutes at 20 ml / min. using a buffer containing 50 mM ammonium acetate, pH 4.8, 8 M urea and 1 M NaCl. To exchange into a volatile buffer, the protein peak obtained from the ion-exchange step was adjusted to pH 2.5 with HCl, acetonitrile was added to 10 % (v/v), and the mixture loaded on to a Silica C_{18} Matrex column (5cm diameter x 8.5 cm long). The bound protein was eluted at 20 ml / min. employing a 10 to 80 % (v/v) acetonitrile gradient over 10 minutes in the presence of 0.1 % TFA. The crude refolded IGF fusion protein was then freeze-dried in convenient batches for subsequent purification. In each chromatographic step, the elution of protein was monitored by absorbance at 280 nm and the purification process was monitored by analysis using microbore C_4 reverse-phase HPLC (Section 2.2.2.5).

2.2.2.3: Purification of Refolded Fusion Proteins

Refolded IGF fusion proteins (Section 2.2.4.5) contain molecules with the correct disulphide bond arrangement, misfolded molecules and variants with altered

molecular mass. Most of these forms can be separated by one or more rounds of preparative reverse-phase HPLC. Typically a 200-300 mg batch of crude refolded IGF fusion protein was separated using a Waters Delta Prep 3000 chromatography system employing a C_4 DeltaPak™ column (47 mm diameter x 300 mm long) and a 31.2 % to 55.2 % (v/v) acetonitrile gradient over 240 minutes at 25 ml / min in the presence of 0.1 % TFA. Fractions containing the eluted proteins were either analysed using a microbore C_4 reverse-phase HPLC (Section 2.2.2.5) or using a C_{18} NovaPak® column (10 mm diameter x 100 mm long) employing a 27.5 % to 38 % (v/v) propan-1-ol gradient in 105 minutes in the presence of 0.13 % HFBA (Francis et al., 1988). The major peak fractions of IGF fusion protein or variants were pooled, diluted 1:3 with a buffer containing 10 % (v/v) propan-1-ol and 0.13 % HFBA in MilliQ™ water and reloaded on to the C_4 DeltaPak™ column (47 mm diameter x 300 mm long) that had been re-equilibrated with the diluent solution. A 27.5 % to 38 % (v/v) propan-1-ol gradient over 105 minutes, in the presence of 0.13 % HFBA, was used to elute the bound fusion protein. Fractions were analysed as for the previous step and the selected fractions were freeze-dried in convenient batches.

When the amount of crude refolded fusion protein was in the range of 20 to 50 mg, the same procedure was used except that the GBC scientific HPLC system was used with a C_4 DeltaPak™ column (25 mm diameter x 100 mm long) and flow rates of 5 ml / min.

2.2.2.4: Purification of Cleaved Fusion Proteins

The hydroxylamine chemical cleavage of IGF fusion proteins (Section 2.2.5) contains the desired product and many variants generated by the cleavage buffer. Many of the variants can be separated using the purification process described in the previous Section (2.2.2.3) but using modified HPLC gradients. The crude

cleavage mixture was first separated using a 24 % to 36 % (v/v) acetonitrile gradient in the presence of 0.1 % TFA and the selected peaks were then separated using a 27.5 % to 33.8 % (v/v) propan-1-ol gradient over 63 minutes in the presence of 0.13 % HFBA.

The reverse-phase separation steps do not separate some hydroxamic acid variants of IGF-I (Canova-Davis *et al.*, 1992), thus a weak anion-exchange HPLC separation was used to separate the IGF-I variants with altered charge. The IGF-I product purified by reverse-phase chromatography was re-chromatographed using a HPLC polyethyleneimine (PEI) column (4.6 mm diameter x 300 mm long) to separate the various molecular forms present. For this purpose, 5 milligrams of heterogenous IGF-I was dissolved at 1 mg / ml in water and then diluted with an equal volume of 20 mM Tris-acetate pH 9.0 prior to loading on the column. The IGF forms were separated at a flow rate of 1 ml / min using buffers containing 20 mM Tris-acetate pH 9.0 and either 0 and 200 mM sodium acetate (Buffers A and B respectively). After 20 minutes using isocratic conditions (Buffer A), a 0 to 15 % Buffer B gradient over 30 minutes was initiated. At 50 minutes the sodium acetate concentration was raised to 200 mM (100 % Buffer B). 1 ml fractions were collected and acidified with 50 μl of glacial acetic acid. The collected IGF peaks were pooled and desalted using a preparative reverse-phase HPLC step, employing a C_4 DeltaPak™ column (25 mm diameter x 100 mm long) and a 24 % to 36 % (v/v) acetonitrile gradient in the presence of 0.1 % TFA.

2.2.2.5: Analysis of the Purification Process

Samples taken at various points in the purification process were acidified to pH 2.1 with TFA and spun at 10,000 g for 3 minutes prior to analysis on a microbore C_4 Aquapore™ reverse-phase HPLC column (Applied Biosystems: 2.1 mm diameter

x 100 mm long). The samples were analysed using a 20 % to 45 % (v/v) acetonitrile gradient over 25 minutes in the presence of 0.1 % TFA at a flow rate of 0.5 ml / min. The eluted protein was detected by its absorbance at 214 nm. Alternatively, a 20 % to 50 % (v/v) acetonitrile gradient over 30 minutes in the presence of 0.1 % TFA at a flow rate of 0.5 ml / min was used. When required, the area under the absorbance profile was converted to protein concentration using calculated extinction coefficients (Buck *et al.*, 1989). See Table 2.1 for a summary of the extinction coefficients for the different IGF analogues.

This reverse-phase analytical HPLC gradient method was modified for the quantitation and the assessment of purity of IGF forms (Section 2.2.3.1), the analysis of refolding reactions (Section 2.2.4.5) and for the separation of enzymatic digestion mixtures (Section 2.2.3.4).

2.2.3: Characterization of IGFs and Their Folding Isomers

2.2.3.1: Reverse-Phase HPLC

Each purified IGF analogue was analysed as described in Section 2.2.2.5 except that a gradient of 25 % to 40 % (v/v) acetonitrile over 150 minutes in the presence of 0.1 % TFA was used.

2.2.3.2: N-Terminal Sequence Analysis

N-terminal sequence analysis was determined for each purified IGF by Edman degradation using a gas-phase sequencer (model 470A; Applied Biosystems, Foster City, CA, U.S.A.) by employing the methods of (Hunkapillar *et al.*, 1983). For proteins containing ^{3}H-labelled S-carboxymethylcysteine residues, the elution of the labelled amino acid was confirmed by collecting fractions from the

Insulin-Like Growth Factor	Extinction Coefficient (M^{-1} cm^{-1}, 214nm)
IGF-I	32.820
[Gly^3]IGF-I	33.131
[Arg^3]IGF-I	32.703
[Lys^9]IGF-I	32.822
Long-IGF-I	33.605
Long-[Gly^3]IGF-I	33.874
Long-[Arg^3]IGF-I	33.505
Long-[Lys^9]IGF-I	33.608

Table 2.1: Extinction Coefficients

Extinction coefficients of different IGF analogues used to estimate protein concentrations from microbore C_4 reverse-phase HPLC absorbance profiles. The values were derived from (Buck *et al.*, 1989).

sequencer and counting using a liquid scintillation counter (Packard Tri-Carb® 1500, Downers Grove, IL, U.S.A.). *N-terminal sequence analyses were performed by D. Turner, Department of Biochemistry, University of Adelaide, SA, Australia.*

2.2.3.3: Electrospray Mass Spectrometry

Electrospray mass spectrometry was carried out on protein samples by using a VG Biotech Quattro mass spectrometer (VG Biotech Ltd., Altrincham, Cheshire, U.K.). *Mass spectrometry was carried out by L. Hick, Department of Chemistry, University of Wollongong, Wollongong, NSW, Australia.*

2.2.3.4: Structural Analysis by Pepsin Digestion

The purified IGFs were cleaved by pepsin at an enzyme substrate ratio of 1:20 (w/w) in 10 mM HCl (Forsberg *et al.*, 1990). After 3 hours at 22°C the reaction was stopped by freezing with liquid nitrogen and drying immediately under vacuum. Electrospray mass spectrometry was then carried out on the pepsin digestion mixture as described in Section 2.2.3.3. Alternatively, the reaction was stopped by injection on a microbore C_4 Aquapore™ reverse-phase HPLC column (Applied Biosystems: 2.1 mm diameter x 100 mm long) equilibrated with 0.1 % TFA. The bound pepsin digestion fragments were then separated using a 0 to 60 % acetonitrile gradient over 60 minutes in the presence of 0.1 % TFA and a flow rate of 0.5 ml / min. The eluted pepsin cleavage fragments were collected for N-terminal sequence analysis (Section 2.2.3.2).

2.2.3.5: Structural Analysis by NMR

The native and alternative isomers of IGF-I were analysed by one-dimensional and two-dimensional ^{1}H-NMR. Both folding isomers were prepared in 10 % CD_3COOD / 90 % H_2O at pH ~ 3 and all spectra were acquired at 40°C on a

Varian Unity 400 NMR Spectrometer. The conditions were set up to mimic those of Sato *et al.* (1992). *The NMR spectroscopy and the skilled interpretation of data was generously carried out by Dr. John Carver, Department of Chemistry, University of Wollongong, Wollongong, NSW, Australia.*

2.2.4: Oxidative Folding of IGF-I and IGF-I Analogues

2.2.4.1: Redox Buffer

Except for one experiment using reduced glutathione (GSH) and oxidized glutathione (GSSG), the redox couple used for the oxidative folding of IGF-I and IGF-I analogues was dithiothreitol (DTT) and 2-hydroxyethyl disulphide (2-HED). Unless stated otherwise, DTT was present at 0.4 mM and 2-HED at 1 mM. The other refolding buffer constituents were 2 M urea, 10 mM glycine, 1 mM EDTA and 0.1 M Tris pH 8.7, unless stated otherwise.

2.2.4.2: Reversible Trapping of Disulphide Isoforms

Because the thiolate anion is the reactive nucleophile in thiol-disulphide exchange in aqueous solution (Gilbert *et al.*, 1990), it is possible to quench thiol-disulphide exchange very rapidly by lowering the pH (Weissman and Kim, 1991). However, acid trapping does not completely stop disulphide exchange. For example, a disulphide exchange reaction that occurs with a half-life of one second at pH 8.7 will exchange with a half life of 10^6 seconds (approximately 280 hours) at pH 2.7. Thus acid quenched disulphide folding intermediates have to be analysed rapidly at low pH, or alternatively separated in a volatile buffer, lyophilized and stored at - 20°C. In all experiments that may involve fast folding intermediates, careful control

experiments were set up to establish that further disulphide exchange did not take place during the HPLC analyses.

2.2.4.3: Non-Reversible Trapping of Disulphide Intermediates

Lyophilized acid-trapped intermediates containing free thiolate anions may be irreversibly alkylated with iodoacetate to facilitate identification of their disulphide architecture (Weissman and Kim, 1991). In these reactions, 400 µl of a solution containing 500 mM sodium iodoacetate, 1 mM EDTA, 200 mM KCl and 1 M Tris at pH 8.7 was added to 200 µg of lyophilized protein and thoroughly mixed for 30 seconds. The mixture was immediately desalted by injection onto a microbore C_4 Aquapore™ reverse-phase HPLC column previously equilibrated with 20 % (v/v) acetonitrile in the presence of 0.1 % TFA. The bound protein was eluted using a swift 20 % to 50 % acetonitrile gradient over 3 minutes in the presence of 0.1 % (v/v) TFA. The fractions containing the single peak were then re-chromatographed using a similar 20 % to 50 % acetonitrile gradient but over 30 minutes. The major peak was selected and lyophilized for subsequent analysis by electrospray mass spectrometry (Section 2.2.3.3) or for digestion with pepsin (Section 2.2.3.4).

To facilitate the identification of pepsin digestion fragments that contain the alkylated cysteine residues, lyophilized acid-trapped intermediates were alkylated in the same manner but using a buffer containing 100 µCi iodo[2-^{3}H]acetic acid per 50 µl. After reverse-phase HPLC purification, pepsin digestion and separation of the pepsin digestion fragments (Section 2.2.3.4), radioactive peptide fragments were sent for N-terminal sequence analysis (Section 2.2.3.2).

2.2.4.4: Reducing Purified IGFs

Each purified IGF-I analogue was reduced for 60 minutes at 22°C, at a protein concentration of 4 mg / ml, in a 8 M urea buffer containing 40 mM glycine, 0.1 M Tris, 1 mM EDTA and 16 mM dithiothreitol (DTT) at pH 8.7. When lyophilized reduced IGFs were required, the solution of reduced protein was acidified to pH 2.5 with HCl and loaded on to a C_4 DeltaPak™ column (25 mm diameter x 100 mm long) and eluted using a linear 0.1 % (v/v) acetonitrile per minute gradient at 5 ml / min. Otherwise, the solution of reduced protein was diluted to 0.1 mg protein / ml, 2 M urea, 10 mM glycine, 0.1 M Tris, 1 mM EDTA and 0.4 mM DTT, pH 8.7 at 22°C prior to refolding.

2.2.4.5: Oxidative Folding from the Fully Reduced Form

All refolding reactions were carried out at 0.1 mg protein / ml to prevent aggregation (Kotlarski *et al.*, 1995). To initiate the oxidative folding reaction, 2-hydroxyethyl disulphide (2-HED) was added to the fully reduced protein at a final concentration of 1 mM. The progress of the reaction was monitored and the folding intermediates stabilized by removing samples at various times and acidifying them to pH 2.1 with TFA. Analysis was carried out by microbore C_4 Aquapore™ reverse-phase HPLC employing a gradient of 20 % to 45 % (v/v) acetonitrile over 25 minutes in the presence of 0.1 % TFA at a flow rate of 0.5 ml / min. Alternatively, a gradient of 20 % to 45 % (v/v) acetonitrile over 75 minutes in the presence of 0.1 % TFA was used. The reverse-phase HPLC elution profiles were used to calculate the percentage abundance of folding products at each time.

For preparation of larger quantities of folding isomers, the folding reaction was stopped after 180 minutes by acidification to pH 2.5 with concentrated HCl. The refolding isomers were then purified by one or two preparative reverse-phase HPLC steps,

employing a C$_4$ DeltaPak™ column (25 mm or 47 mm diameter x 100 mm long) and the chromatography conditions described in Section 2.2.2.3.

2.2.4.6: Oxidative Folding of Reversibly Trapped Disulphide Isoforms

To facilitate detailed analysis of oxidative folding, isolated disulphide isomers or folding pathway intermediates trapped by acid-quenching were allowed to rearrange or continue folding by raising the pH. This was done by dissolving lyophilized disulphide isoforms in a buffer containing 2 M urea, 10 mM glycine, 0.1 M Tris, 1 mM EDTA, 0.4 mM dithiothreitol (DTT) and 1 mM 2-hydroxyethyl disulphide (2-HED) at pH 8.7. In control reactions, the same buffer was used but without DTT or 2-HED present. The progress of the reactions were monitored and analysed as described in the previous Section (2.2.4.5).

2.2.4.7: Analysis of the Disulphide Folding Pathway of Long-[Arg^3]IGF-I

Ten milligrams of purified Long-[Arg^3]IGF-I was reduced and refolded as described in Sections 2.2.4.4 and 2.2.4.5. However, after one minute of refolding the reaction was stopped by acidification and the resultant mixture purified by reverse-phase HPLC using a C$_4$ DeltaPak™ column (25 mm diameter x 100 mm long) and a 25 % to 50 % acetonitrile gradient over 500 minutes in the presence of 0.1 % TFA. One minute fractions were collected. Fractions containing the eluted protein peaks were analysed using the analytical microbore C$_4$ HPLC system described in Section 2.2.2.5. The selected fractions were pooled and lyophilized. Each of the trapped intermediates were then alkylated, separated and analysed as described in Section 2.2.4.3.

2.2.5: Cleavage of Fusion Proteins

2.2.5.1: Hydroxylamine Cleavage

The IGF fusion proteins have been designed to be cleaved with hydroxylamine (Bornstein and Balian, 1977) between the asparagine and glycine amino acids that link the fusion protein partner to the IGF molecule. The initial chemical cleavage reactions were carried out according to the method of (Bornstein and Balian, 1977) which was originally designed to aid sequencing of unknown proteins. However, this method generates undesirable side reactions. To help overcome these problems the cleavage of IGF fusion proteins with hydroxylamine was optimized.

2.2.5.2: Optimization of the Cleavage Reaction

For all of the hydroxylamine cleavage reactions the fusion protein concentration was 0.25 mg / ml and the cleavage buffer contained EDTA (1 mM) and Tris (0.1 M). To develop the cleavage conditions the following parameters were varied; pH (8.3, 8.65 and 9.0), urea concentration (0, 0.5 M, 1.0 M and 2.0 M), hydroxylamine concentration (0, 0.25 M, 0.5 M, 1 M and 2 M) and temperature (24, 37 and 45°C). For each experiment the cleavage buffer was sparged with nitrogen at the reaction temperature before adding to the freeze-dried fusion protein. The cleavage vessel was then closed under a nitrogen atmosphere. The progress of the reaction was monitored and the cleavage products stabilized by removing samples of the cleavage reaction at various times and acidifying them to pH 2.1 with 5 % (v/v) TFA. The cleavage reaction was finally stopped by lowering the pH to 5 with glacial acetic acid and cooling to 2°C. The cleavage products were then purified by one or two preparative reverse-phase HPLC steps, employing a C_4 DeltaPak™ column

(47 mm diameter x 300 mm, or 25 mm diameter x 100 mm) and a linear 0.1 % (v/v) acetonitrile per minute gradient in the presence of 0.1 % TFA (Section 2.2.2.4). The samples taken during the cleavage reaction were analysed by microbore C_4 reverse-phase HPLC (Section 2.2.2.5). The elution of protein on reverse-phase HPLC was measured by absorbance at 214 nm to calculate the percentage abundance of cleavage products.

2.2.5.3: Isoelectric Focussing

IGF-I samples produced by the altered hydroxylamine cleavage methods were analysed by isoelectric focusing (IEF) using a Pharmacia PhastSystem (Pharmacia Biotech, Uppsala, Sweden) and PhastGel IEF 3-9 media using IEF standards with a pI range of 5.20 to 10.25 The IGF-I samples were either dissolved in water or in 10 mM HCl, loaded in the middle of the gel and separated according to the Manufacturer's instructions (Separation Technique File Number 100). The gels were fixed in 20 % (w/v) trichloroacetic acid (TCA) for 30 minutes and washed twice with 20 % (w/v) TCA / 10 % (v/v) ethanol / 5 % (v/v) glacial acetic acid for 15 minutes. The fixed protein was then silver stained according to the methods of Merril et al. (1984).

2.2.6: In Vitro Biological Assays

2.2.6.1: Quantitation of IGF Proteins

For the in vitro biological assays, the amount of each IGF tested was determined by calculating the area under their absorbance profile when eluted from a microbore C_4 reverse-phase HPLC column (Section 2.2.2.5). The area under the absorbance profile was converted to protein concentration using calculated

extinction coefficients (Buck *et al.*, 1989). See Table 2.1 for a summary of the extinction coefficients for the different IGF analogues. The HPLC instrumentation was standardized by analysis of a reference Long-[Arg^3]IGF-I preparation.

2.2.6.2: Effect of IGF Proteins on Protein Synthesis

Stimulation of protein synthesis in rat L6 myoblasts by IGF proteins was measured over 18 hours using the method described by Francis *et al.* (1986). Monolayers of rat L6 myoblasts in 24-place multiwell plates were used when the cells approached confluence. Before incubating the cells with the IGF proteins, the DMEM culture medium containing foetal bovine serum (FBS) was replaced with DMEM without FBS and incubated for 2 h at 37°C. The medium in each well was then replaced with 900 µl of DMEM containing 1 µCi / ml L-[4,5-^{3}H]Leucine and 100 µl of diluted IGF protein. The IGF proteins had previously been dissolved in 10 mM HCl and diluted to the desired range with a 10 mM potassium phosphate buffer, pH 7.4 containing 0.09 % (w/v) NaCl and 0.1 % (w/v) bovine serum albumin (BSA). After incubation at 37°C for 18 hours, the monolayers were washed twice with Hanks' salts, twice with 5 % (w/v) trichloroacetic acid over a 10 minute period, and once with water before solubilizing in 0.5 M NaOH containing 0.1 % (v/v) Triton X-100. The solubilized cells were sampled in duplicate and counted using a liquid scintillation counter (Packard Tri-Carb® 1500, Downers Grove, IL, U.S.A.). The stimulation of protein synthesis was measured as the increased incorporation of L-[4,5-^{3}H]Leucine into total cell protein during the 18 hour incubation period compared with a buffer control. The biological potencies were calculated as the amount of added protein required to give 50 % of the maximum stimulation achieved by IGF-I.

2.2.6.3: Binding of IGF Proteins to Cell Receptors

The relative binding of various IGF proteins to confluent monolayers of rat L6 myoblasts was measured by their competition for the binding of 0.2 ng ^{125}I-labelled recombinant human IGF-I (Ross *et al.*, 1989). ^{125}I-labelled IGF-I was added in the presence of increasing concentrations of unlabelled IGF proteins in a total volume of 0.5 ml Hepes buffer (0.1 M Hepes, 0.12 M NaCl, 5 mM KCl, 1.2 mM $MgSO_4.7H_2O$ and 8 mM glucose at pH 7.6) containing 0.5 % (w/v) BSA to cell monolayers in 24-place multiwell plates. IGF proteins had previously been dissolved in 10 mM HCl and diluted to the desired range with the above Hepes buffer containing 0.5 % (w/v) BSA. Following incubation for 18 hours at 4°C the cell monolayers were washed with Hanks' salts at 0°C to remove unbound ligand. The cell monolayers were dissolved in 0.5 M NaOH containing 0.1 % (v/v) Triton X-100. The solubilized cells were then counted using a gamma counter (LKB-Wallac 1272 Clinigamma, Turku, Finland). Binding was expressed as the percentage of ^{125}I-labelled IGF-I bound in the absence of added unlabelled IGF protein. ED_{50} values were calculated as the concentration of added protein giving 50 % displacement of labelled ligand.

2.2.6.4: Binding of IGF Proteins to IGFBPs

The relative binding of IGF proteins to IGF binding proteins (IGFBPs) present in L6 myoblast conditioned medium was measured as described by (Szabo *et al.*, 1988). L6 myoblast IGFBPs were prepared from serum-free Dulbecco's Minimal Essential Medium conditioned by confluent monolayers of cells for 24 hours. The medium was filtered through a 0.22 μm filter and stored at -20°C until used. The binding of various IGF proteins to IGFBPs was measured by their competition for the binding of 0.2 ng ^{125}I-labelled IGF-I for 18 hours at 4°C. IGF proteins had previously been dissolved in 10 mM HCl and diluted to the desired range with a 50 mM sodium phosphate buffer,

pH 6.5, containing 0.1 % (w/v) BSA and 0.02 % (w/v) sodium azide. After incubation, free ^{125}I-labelled IGF-I was separated from IGFBP-bound label by incubating the reaction mixture with charcoal (5 % (w/v) activated charcoal, 0.2 % (w/v) protamine sulphate in the above phosphate buffer, pH 6.5) at 4°C for 30 minutes and then centrifuging the mixture. The supernatant was removed and counted using a gamma counter (LKB-Wallac 1272 Clinigamma, Turku, Finland). Binding was expressed as the percentage of ^{125}I-labelled IGF-I bound in the absence of added unlabelled IGF protein. ED_{20} values were defined as the concentration of added IGF protein giving 20 % displacement of labelled ligand.

2.2.6.5: Stability of Long-[Arg^3]IGF-I in the Receptor Binding Assay.

Five micrograms of the major alternative folding isomer of Long-[Arg^3]IGF-I was added in 0.1 M Hepes buffer pH 7.6, in the presence of 0.5 % (w/v) BSA to confluent monolayers of rat L6 myoblasts under the assay conditions described for binding of IGF proteins to cell receptors (Section 2.2.6.3). After incubation for 18 hours at 4°C, the medium was removed and centrifuged at 10000 g for 3 minutes; 400 µl was removed, acidified with 100 µl of 5 % (v/v) TFA and centrifuged again. Portions of this sample were either mixed with an equivalent amount of receptor binding activity of the correctly folded isomer of Long-[Arg^3]IGF-I in 10 mM HCl, or diluted with 10 mM HCl. The treated samples were analysed by microbore C$_4$ Aquapore™ reverse-phase HPLC employing a linear gradient of 28 % to 40 % (v/v) acetonitrile over 60 minutes in the presence of 0.1 % TFA. Fractions were collected for the analysis of receptor binding activity. The elution times of the main alternative and native folding isomers of Long-[Arg^3]IGF-I from the HPLC column were confirmed by mixing control incubation samples that contained no IGFs with the appropriate folding isomers of Long-[Arg^3]IGF-I.

2.2.6.6: Analysis of Data from Biological Assays

Curves were drawn for the biological assays using a computer-fitted logistic dose-response function employing TableCurve version 3.1 software (Jandel Scientific, Corte Madera, CA, U.S.A.). This logistic dose-response function was used to calculate biological potency data in each assay.

CHAPTER 3

THE PURIFICATION AND CHARACTERIZATION OF IGF-I AND IGF-I ANALOGUES

3.1: INTRODUCTION

The application of gene fusion technology to the production of heterologous proteins in *E. coli* has led to the development of many techniques to recover the protein of interest (Enfors, 1992; Fischer *et al.*, 1993; Chaudhuri, 1994). For some expression systems that direct the nascent protein to the periplasm for folding and secretion, the correctly folded and processed molecule can be recovered from the media. While there are distinct advantages provided by cellular processing, recoveries can be low as a result of low expression rates, degradation by cellular proteases and problems recovering the protein from the media. Alternative expression systems that result in the accumulation of the expressed protein as inclusion bodies may provide high yields, but the protein is both insoluble and generally inactive (Marston, 1986; Kane and Hartley, 1988). Thus, the inclusion bodies need to be extracted from the cell and processed *in vitro* to recover the biologically active form. Inclusion bodies are viewed as dense aggregates of misfolded protein with little ordered structure (De Bernardez-Clarke and Georgiou, 1991). However, some inclusion bodies such as those containing endoglucanase D may have some tertiary structure and biological activity (Tokatlidis *et al.*, 1991; Chaffotte *et al.*, 1992b). The mechanism that directs the formation of inclusion bodies is largely unknown but is almost certain to involve a combination of the lack of interactions with cellular proteins such as chaperones as well as the highly reducing environment of the cytoplasm (Chaudhuri, 1994; Thatcher and Hitchcock, 1994). Although there are significant problems associated with the inclusion body route of production of heterologous proteins, the yields are potentially very high and at least *in vivo* the protein is protected from cellular proteases. The expression system used for the production of IGFs by this research group requires the use of an

N-terminally linked fusion partner derived from porcine growth hormone to facilitate high level expression in *Escherichia coli* (King *et al.*, 1992; Francis *et al.*, 1992). This expression system results in the formation of inclusion bodies in the cytoplasm which require downstream processing to recover the biologically active protein.

The insulin-like growth factors (IGFs) used for refolding studies in subsequent chapters were supplied as inclusion bodies containing the IGF fusion protein, as crude refolded fusion protein, or as purified protein. To use these proteins for refolding experiments, they needed to be purified to a similar stage and fully characterized using techniques such as HPLC, mass spectrometry and N-terminal sequence analysis. In the course of these analyses it became apparent that the crude refolded fusion protein contained variants and that when these fusion proteins were cleaved using hydroxylamine, other modifications were evident. As the nature of the IGF variants were unknown, it was therefore important to further characterize the proteins and establish that the variants did not affect the oxidative folding process, which is the focus of this thesis. This chapter follows the purification of a batch of IGF-I from inclusion bodies through well-established protocols of recovery, refolding, removal of the fusion protein partner and 'polishing' of the final product. Particular attention is made to the variants of IGF-I present at various stages of the purification. The chemical cleavage of the fusion protein which proved to be a major source of heterogeneity in the final product was reviewed and optimized to reduce the chemical modification. Furthermore, purification protocols were developed to separate the IGF-I variants so that they could be compared in refolding experiments and in biological assays. Although this chapter refers mainly to the processing of IGF-I from Long-IGF-I inclusion bodies, the techniques used and the observations

therein are similar for the processing of other fusion proteins such as Long-[Gly^3]IGF-I, Long-[Arg^3]IGF-I and Long-[Lys^9]IGF-I.

3.2: METHODS

The methods used are described in detail in Chapter 2. However, as appropriate some details are included in the text of the Results section.

3.3: RESULTS

3.3.1: Recovery of Insulin-Like Growth Factors from Inclusion Bodies

3.3.1.1: Recovery of Fusion Protein from Inclusion Bodies

The plasmid containing the expression vector p[Met^1]pGH(1-11)-Val-Asn-IGF-I was transformed into *Escherichia coli*, fermented and induced as described in section 2.2.1 and (Francis *et al.*, 1992). The expressed fusion protein accumulated in the cells as dense aggregates called inclusion bodies as a result of the high level expression. The inclusion bodies were harvested by disruption of the cells and by differential centrifugation. In subsequent steps, the inclusion bodies were solubilized in a denaturing and reducing buffer containing 8 M urea buffer pH 9.0 and 40 mM dithiothreitol (DTT), centrifuged to remove insoluble cellular debris and then chromatographed by size-exclusion chromatography to remove low molecular mass material (section 2.2.2.2). The reduced and denatured fusion protein eluted as a single peak (Figure 3.1a). However, the microbore C$_4$ reverse-phase HPLC analysis (Figure 3.1b) showed the reduced fusion protein to have two major components (R) and (X).

Figure 3.1: Recovery of IGF-I from Inclusion Bodies

Denatured and reduced inclusion bodies containing the fusion protein [Met^1]pGH(1-11)-Val-Asn-IGF-I were solubilized and centrifuged prior to separation by size-exclusion chromatography (a). The eluted protein (hatched) was pooled and analysed by microbore C_4 reverse-phase HPLC using a 20 % to 50 % acetonitrile gradient over 30 minutes in the presence of 0.1 % TFA (b). After the refolding step, the hydroxylamine cleavage step and the final purification steps, samples were analysed by microbore C_4 reverse-phase HPLC (c, d and e respectively). The abbreviations used are R (reduced), X (variant X), A (alternative isomer), N (native isomer) and XN (native fold of variant X).
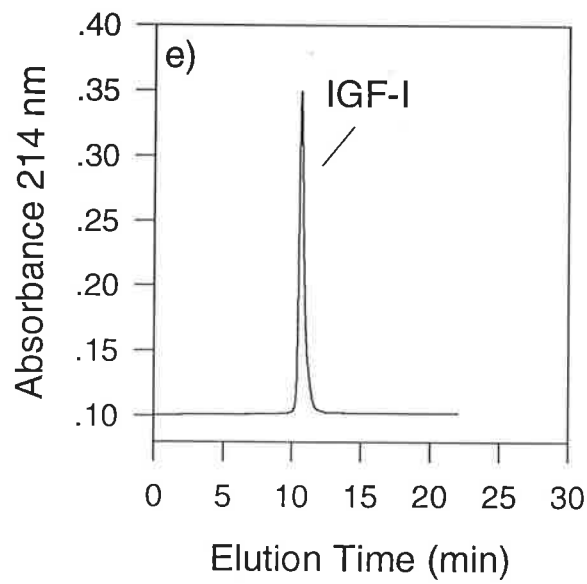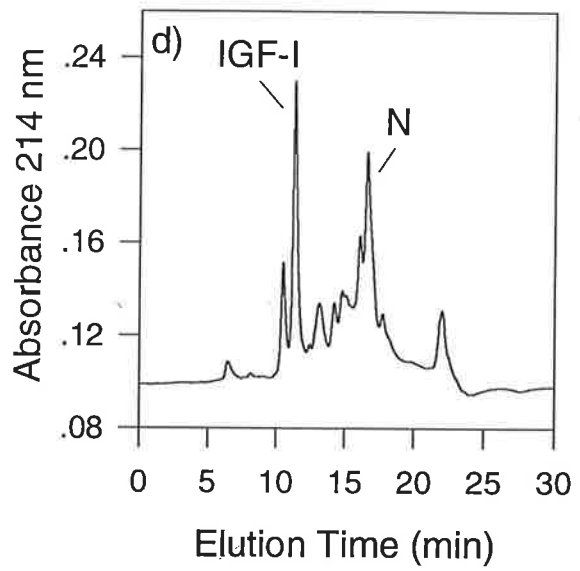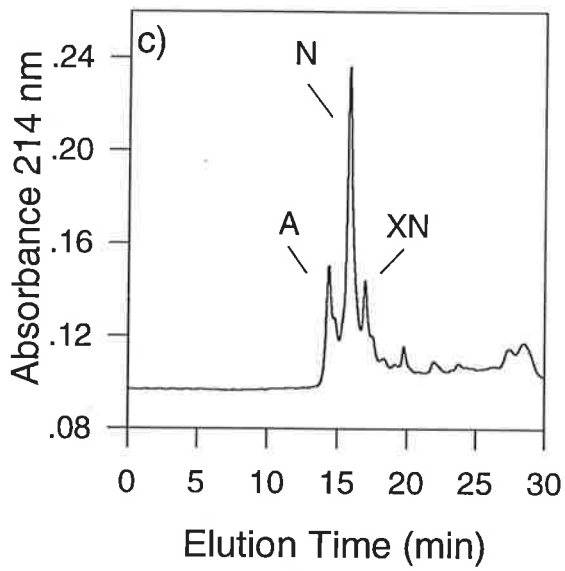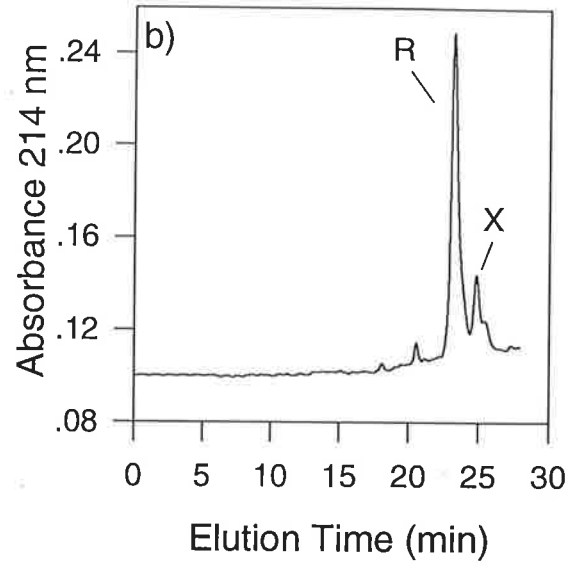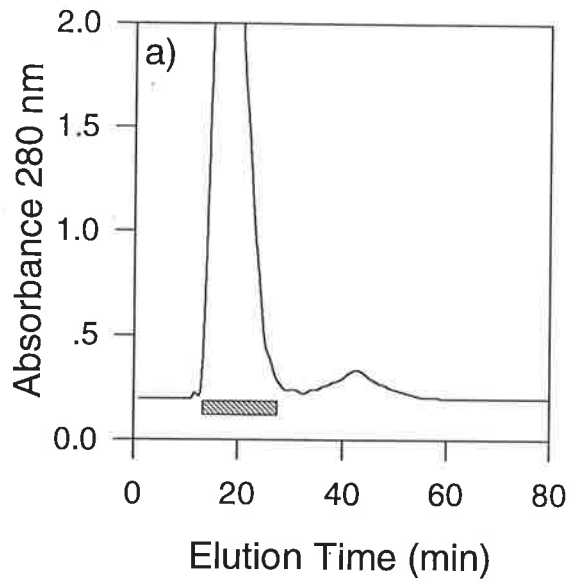
3.3.1.2: Refolding the IGF Fusion Protein

The reduced and denatured fusion protein mixture was diluted to 2 M urea and 0.4 mM DTT, at a protein concentration of 0.1 mg / ml prior to the addition of the oxidizing agent (2-hydroxyethyl disulphide) used to initiate refolding (Section 2.2.2.2). After 100 minutes the refolding reaction was stopped by acidification. A sample was analysed by microbore C_4 reverse-phase HPLC which showed the reduced fusion proteins R and X to have refolded to components that elute earlier (less hydrophobic) than the reduced forms (Figure 3.1c). Three major components were identified; A, N and XN. This refolding pattern was similar for the other IGF fusion proteins investigated (Long-[Gly^3]IGF-I, Long-[Arg^3]IGF-I and Long-[Lys^9]IGF-I). A simple analysis of these folded forms revealed that both A and N could be fully reduced to a form that elutes at the same position as R and that XN could be fully reduced to a form that elutes at the same position as X (results not shown). These variants were later shown to be an alternative misfolded molecule (A), the native folded molecule (N) and a correctly refolded variant (XN) (Sections 3.3.3.1 and 5.3.1). The refolded protein mixture was then loaded on to a SP-Sepharose Fast Flow column to remove any co-chromatographed endotoxins (Section 2.2.2.2). After elution, the protein was re-chromatographed on a large scale C_{18} reverse-phase column to exchange the protein into a volatile buffer for freeze drying (Section 2.2.2.2). The area under the absorbance profile for some of the microbore reverse-phase HPLC analyses were converted to protein concentration using calculated extinction coefficients (Buck *et al.*, 1989). From 20.5 g (wet weight) of inclusion bodies the yield of refolded freeze-dried peptide was 1.39 g.

3.3.1.3: Large scale Hydroxylamine Cleavage of Long-IGF-I

The refolded Long-IGF-I produced above (Section 3.3.1.2) was cleaved by hydroxylamine in a large scale reaction developed to minimize chemical heterogeneity (Section 3.3.2). In this reaction the Long-IGF-I variant A cleaves to produce misfolded IGF-I and the variants N and XN cleave to produce native IGF-I (See Section 3.3.3.1). The cleavage conditions were 1 mg protein / ml in 1M hydroxylamine, 0.1M Tris pH 8.65, 2M urea for 24 hours at 37°C. The yield of native IGF-I was calculated to be approximately 31 % of the theoretical maximum (from Figure 3.1d). Analysis by mass spectrometry showed the IGF-I product to contain approximately 70% unmodified IGF-I (7648.7 ± 0.5 m.u.), approximately 20% modified IGF-I with a mass 16 units higher (7665.1 ± 1.2 m.u.) and approximately 10% of a variant with a mass of 7630.5 ± 0.63. Analysis by isoelectric focusing showed that the main species present was IGF-I with the correct pI of 8.7 and that the variant with a mass 16 units higher had a pI of 8.5 (Section 3.3.3.2). The overall yield of unmodified IGF-I was 23%. The IGF-I product was purified by preparative C_4 reverse-phase HPLC and then by polyethyleneimine (PEI) anion-exchange chromatography to remove the altered molecular forms (Section 2.2.2.4). The final product was analysed by microbore C_4 reverse-phase HPLC (Figure 3.1e); IEF gel analysis showed the correct pI and mass spectrometry showed that it was greater than 95 % pure.

3.3.2: Optimization of the Hydroxylamine Cleavage of Long-IGF-I

To remove the 13-amino acid N-terminal fusion protein partner from the required IGF molecule, hydroxylamine is used to specifically cleave the fusion protein between the asparagine and glycine amino acids. The methods previously

used in our laboratory (King *et al.*, 1992; Upton *et al.*, 1992) yield an IGF product that appears homogenous by microbore C_4 reverse-phase HPLC. However, a retrospective analysis of the IGF products by mass spectrometry showed that side-chain modification of amino acids is common. For example, hydroxylamine cleavage of an IGF-I fusion protein in 2M urea, 2M hydroxylamine, at pH 9.0 and 45°C yields an IGF containing a small amount of unmodified IGF-I and up to 80 % modified peptide. As the potential impact of the modifications on refolding was unknown, I set out to establish a hydroxylamine cleavage method that minimises side-chain modifications and to purify the various modified peptides for comparative studies. To simplify the investigation of the hydroxylamine cleavage reaction, the correctly folded Long-IGF-I species (N) was purified from the other refolded Long-IGF-I variants by preparative reverse-phase HPLC (Section 2.2.2.3) for use as a substrate.

3.3.2.1: Optimization of the Cleavage Reaction

In the initial cleavage trial, the hydroxylamine concentration was fixed at 1M while urea concentration (0, 0.5M and 1M), pH (8.3 and 8.65) and temperature (24°C and 37°C) were manipulated; other conditions are as described in the Methods Section (2.2.5.2). For comparison, conditions similar to that used previously (2M urea and pH 9.0) were also carried out. These trials showed that addition of urea did not improve the cleavage of Long-IGF-I, since the IGF-I yield, determined by microbore C_4 reverse-phase HPLC, was similar in the absence or presence of 0.5M or 1M urea (Figure 3.2a, 3.2b and 3.2c). Furthermore, these reactions showed that the cleavage was reduced at lower pH and at lower temperature. The reaction carried out at 2M urea and pH 9.0 showed a slightly greater yield of IGF-I after 24 hours (Figure 3.2d), a result consistent with higher pH. Electrospray mass spectrometry of IGF-I product peaks after 48 hours cleavage at pH 8.3 and 37°C, at
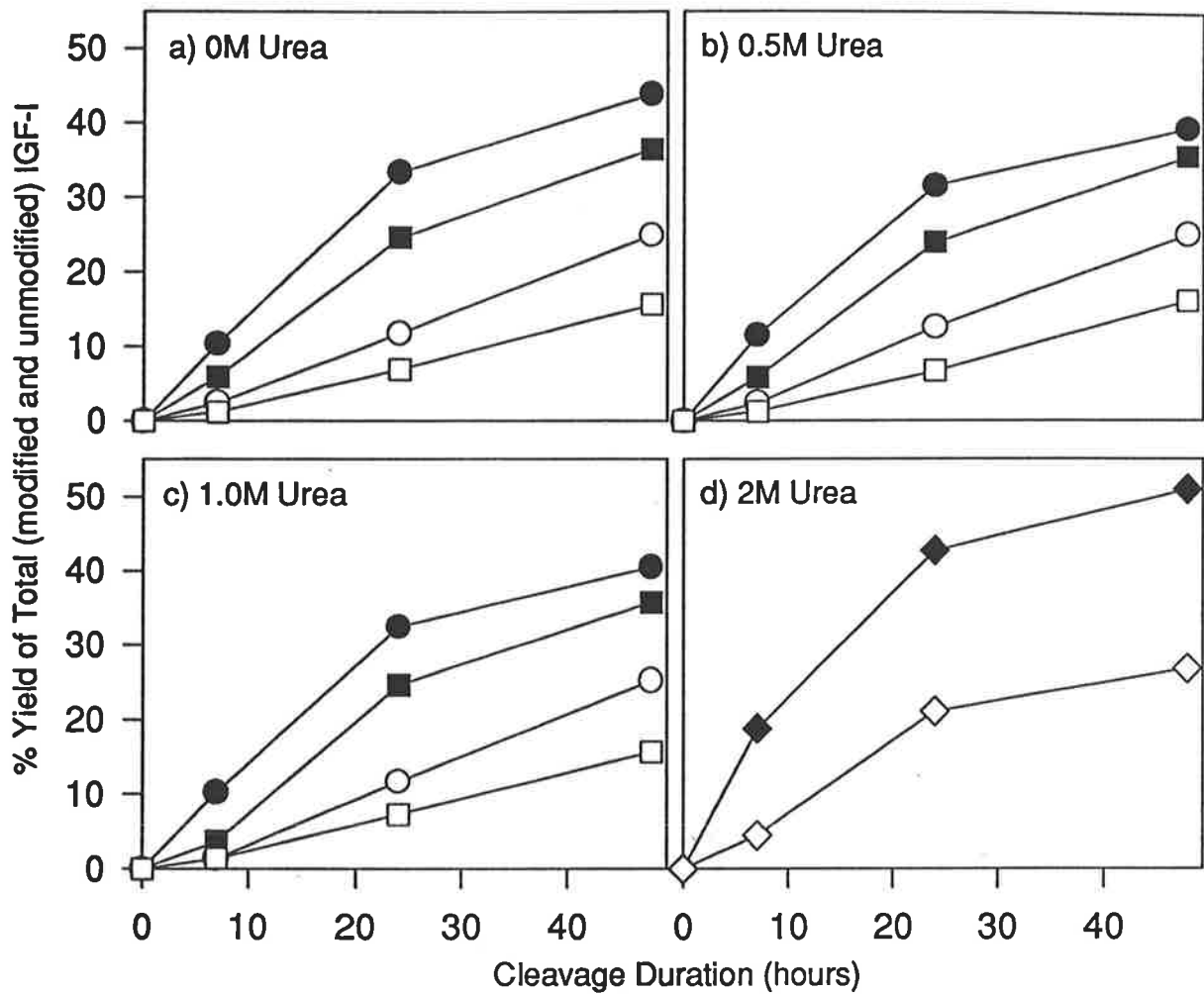
Figure 3.2: Effect of Urea, pH, and Temperature on the Hydroxylamine Cleavage of Long-IGF-I

The reactions were carried out at pH 8.3 (■,□), 8.65 (●,○) and 9.0 (◊,♦). The temperature was 24°C (open symbols) and 37°C (filled symbols). The yield of IGF-I was calculated from the microbore C_4 reverse-phase HPLC elution profiles taken at various time points during the cleavage reactions.

pH 8.65 and 24°C and at pH 8.65 and 37°C showed that in each case two species were present in an approximate ratio of 2:1: the major species was unmodified IGF-I (7649.2 ± 1.2 m.u.) and the minor species was an IGF-I variant containing an extra 16 mass units (7664.9 ± 0.8 m.u.). The presence of an extra 16 mass units is consistent with modification of a single asparagine or glutamine amino acid to its hydroxamic acid form in the hydroxylamine cleavage buffer (Canova-Davis *et al.*, 1992).

The next step in the development of the cleavage reaction was to test the effect of hydroxylamine concentration on IGF-I yield and modification. In this experiment Long-IGF-I was incubated in the absence of urea but with different concentrations of hydroxylamine. The temperature was elevated to 45°C to increase the rate of cleavage, a lower pH of 8.65 was chosen to reduce potential deamidation reactions (Wright, 1991) while still providing a reasonable cleavage rate (Figure 3.2a). Analysis of the different reaction mixtures showed that the yield of IGF-I product in the first 8 hours of incubation was proportional to the hydroxylamine concentration (Figure 3.3). For the 2M hydroxylamine reaction mixture the yield increased for 20 hours, but thereafter declined; presumably due to degradation in the cleavage buffer. Similarly, the yield of IGF-I in the 1M hydroxylamine cleavage reaction peaked at 28 hours and appeared to decline at 48 hours, suggesting degradation of the product. Clearly, an incubation of no more than 24 hours at the higher hydroxylamine concentrations is desirable. The hydroxamic acid modification of IGF-I was monitored for some of the cleavage reactions using isoelectric focusing (IEF; Canova-Davis *et al.*, 1992). This analysis showed that 0.5M hydroxylamine did not appear to cause any modification of the IGF-I product even after 48 hours (Figure 3.4a, Lanes 2 and 3). However, the IGF-I purified from the 1M hydroxylamine
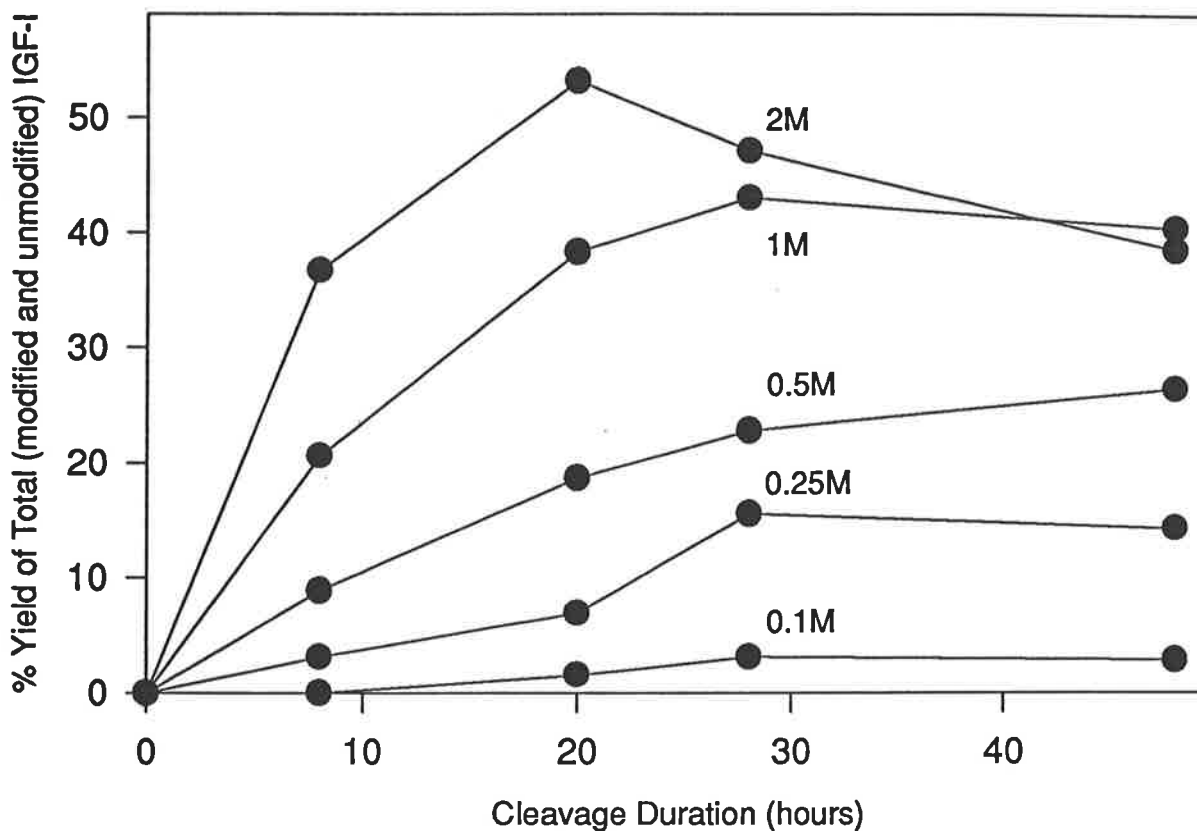
Figure 3.3: Effect of Hydroxylamine Concentration on the Cleavage of Long-IGF-I at 45°C

The reactions contained the indicated concentrations of hydroxylamine (0.1 M, 0.25M, 0.5 M, 1.0 M and 2.0 M), Tris (pH 8.65) was present at one-tenth the concentration of hydroxylamine. EDTA was also present (1 mM). The yield of IGF-I was calculated from the microbore C_4 reverse-phase HPLC elution profiles taken at various time points during the cleavage reactions and expressed as the percentage of total protein.

Figure 3.4: Isoelectric focusing of IGF forms obtained after cleavage of Long-IGF-I

(a) IGF-I products obtained using different concentrations of hydroxylamine at pH 8.65 and 45°C. Lanes: *(1)* Isoelectric focusing (IEF) standards; *(2)* 0.5 M hydroxylamine for 24 hours; *(3)* 0.5 M hydroxylamine for 48 hours; *(4)* 1.0 M hydroxylamine for 24 hours; *(5)* 2.0 M hydroxylamine for 24 hours; *(6)* IEF standards. (b) Purified IGF products obtained after cleavage of Long-IGF-I at pH 8.65 and 37°C using 1 M hydroxylamine for 24 hours. Lanes: *(1)* IEF standards; *(2)* IGF-I product peak; *(3)* alternative product peak P; *(4)* Long-IGF-I peak; *(5)* Long-IGF-I before incubation; *(6)* IGF-I obtained following cleavage at pH 9.0 and 45°C in 2 M urea and 2 M hydroxylamine; *(7)* IEF standards. (c) IGF-I purified by large scale polyethyleneimine (PEI) anion-exchange chromatography. Lanes: *(1)* IEF standards; *(2)* IGF-I obtained following large-scale cleavage at pH 8.65 and 37°C in 1 M hydroxylamine; *(3)* unmodified IGF-I obtained following PEI chromatography. (d) IGF-I produced by cleavage at pH 9.0 and 45°C in 2 M urea and 2 M hydroxylamine was separated using PEI chromatography (Figure 3.6) and analysed by isoelectric focusing. Lanes: *(1)* Peak A; *(2)* Peak B; *(3)* Peak C; *(4)* Peak D; *(5)* IEF standards.
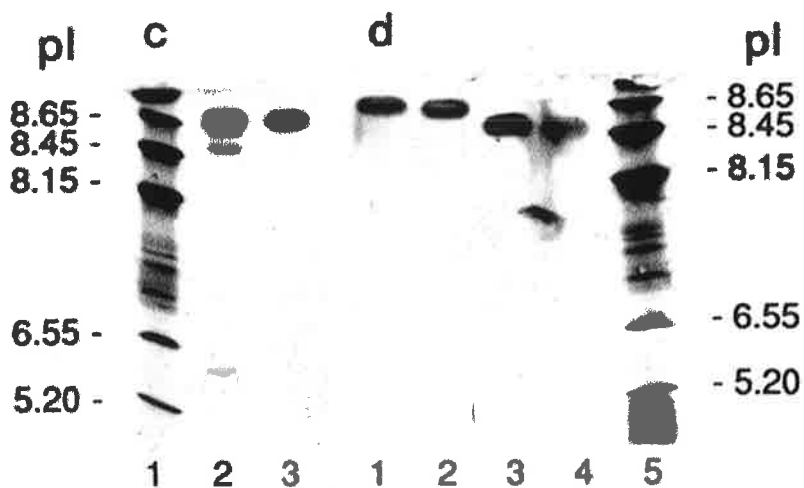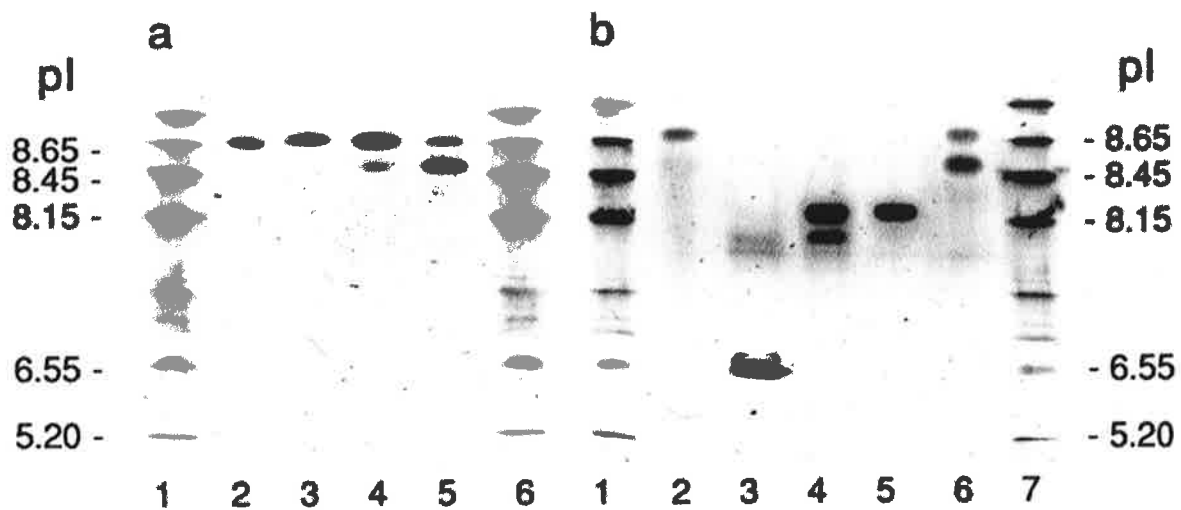
cleavage reaction after 24 hours showed approximately 30 % heterogeneity with a band focusing 0.2 pH units lower than that of unmodified IGF-I (Figure 3.4a, Lane 4). A similar analysis of the 2M hydroxylamine cleavage reaction showed that most of the IGF-I was modified (Figure 3.4a, Lane 5). The yields of unmodified IGF-I after 24 hours incubation at 45°C and pH 8.65 were calculated from the IEF gel for the different hydroxylamine conditions; the respective yields of unmodified IGF-I in 0.5M, 1M and 2M hydroxylamine were 21 %, 29 % and 20 % of the total protein. For comparison, the yields of unmodified IGF-I after 24 hours incubation at 24°C and 37°C using 1M hydroxylamine at pH 8.65 were 8 % and 22 % of the total protein respectively (Figure 3.2a). Clearly, the simplest conditions for producing unmodified IGF-I are 0.5M hydroxylamine, pH 8.65 and 45°C, although hydroxylamine concentrations intermediate between 0.5M and 1M (Figure 3.4a) or higher temperatures may improve the yield.

3.3.3 Investigation of IGF-I Variants

3.3.3.1: Investigation of IGF Variants Present After Refolding

The protein present in fully reduced inclusion bodies of Long-IGF-I (R and X) refolds to yield the native molecule (N), the alternative folding isomer (A) and a third variant XN (Figure 3.1c), a refolding pattern common to each of the IGF fusion proteins. The properties of the folding isomers A and N are investigated in detail in Chapters 4 and 5. To investigate the third type of IGF variant derived from the reduced species (X), Long-[Arg³]IGF-I was refolded and its variant XN was isolated. Long-[Arg³]IGF-I was chosen because the crude refolded peptide that contains the three variants A, N and XN proved to be almost as potent as N alone in biological

assays (results not shown). For this purpose, 20 mg of refolded Long-[Arg^3]IGF-I was separated on a semi-preparative C$_{18}$ reverse-phase HPLC column employing a linear 0.1 % (v/v) propan-1-ol per minute gradient in the presence of 0.13 % (v/v) HFBA. Peaks A, N and XN were collected and dried for analysis. Electrospray mass spectrometry of these variants showed only the expected mass for A and N (9111.5 ± 0.3 and 9111.3 ± 0.6 respectively, theoretical mass 9111.6 m.u.) whereas XN had a mass 28 units higher (9139.6 ± 0.6 m.u.). Mass increases of 16 mass units were not observed indicating that the potential oxidation of methionine residues had not taken place. The increase of mass of 28 units is consistent with a formylation at the N-terminus or elsewhere (Hirel *et al.*, 1989; Doonan *et al.*, 1978). The N-terminal sequence analysis of XN returned a sequence that accounted for only 15 % of the protein on a molar basis, indicating that most of the protein was N-terminally blocked. To demonstrate that the N-terminal methionine was formylated, the putative formyl group was removed by gentle hydrolysis with 0.5 M HCl in methanol at room temperature for 48 hours (Sheehan and Yang, 1958). After drying, N-terminal sequence analysis returned a sequence that accounted for 86 % of the protein on a molar basis confirming that the blocking formyl group had been removed. Furthermore, in a radio-immunoassay, an antibody raised against Long-[Arg^3]IGF-I bound equally well to N and XN, but poorly to A, indicating that N and XN have similar structures and that the epitope is disorganized in the alternative misfolded form *(S. Knowles, CRC for Tissue Growth and Repair, personal communication).* Thus, the crude refolded fusion proteins contain an alternative misfolded form (A), the native molecule (N) and a correctly folded N-terminally modified fusion protein (XN). Presumably the misfolded form of X is present in the refold mixture, but it was not resolved in this study.

3.3.3.2: Investigation of Variants Generated During Hydroxylamine Cleavage

To investigate the IGF-I variants generated during the hydroxylamine cleavage of Long-IGF-I, a set of conditions developed above (Section 3.3.2.1: 1M hydroxylamine, 0.1M Tris pH 8.65 and 37°C) was selected and compared to the conditions previously used for the cleavage of Long-IGF-I, adapted from the method of Bornstein and Balian (1977) (2M urea, 2M hydroxylamine, 0.1M Tris pH 9.0 and 45°C). After 24 hours incubation the reaction mixtures were separated by reverse-phase HPLC and the collected IGF fractions were analysed by isoelectric focusing and mass spectrometry. The total yield of IGF-I in the cleavage carried out at the lower pH and temperature was 33% of the total protein (Figure 3.5), a result consistent with the trial cleavage reaction (Figure 3.2a). The isoelectric focusing gel showed that the major product was IGF-I with the correct pI of 8.7 plus a faint band which focused at pI 8.5 suggesting only low levels of heterogeneity (Figure 3.4b, Lane 2). Analysis by mass spectrometry showed that the IGF-I peak contained approximately 70% unmodified IGF-I (7649.5 ± 0.5 m.u.) and approximately 30% of a variant with a mass 16 units higher (7664.8 ± 0.3 m.u.). While the yield of unmodified IGF-I (23 %) is only slightly higher than the hydroxylamine cleavage reaction investigated above (Section 3.3.2.1: 0.5M hydroxylamine / pH 8.65 / 45°C for 24 hours), the cleavage rate is higher, improving the overall yield of IGF-I. The cleavage reaction carried out at the higher urea, hydroxylamine, pH and temperature conditions increased the yield of IGF to 52% of the total protein (results not shown). Isoelectric focusing showed heterogeneity (Figure 3.4b, Lane 6) and mass spectrometry revealed that although 25% of this IGF-I product was unmodified, major components with masses of 7665.8 ± 0.7 (35%), 7682.2 ± 0.6 (30%) and 7699.2 ± 1.2 (10%) were present.
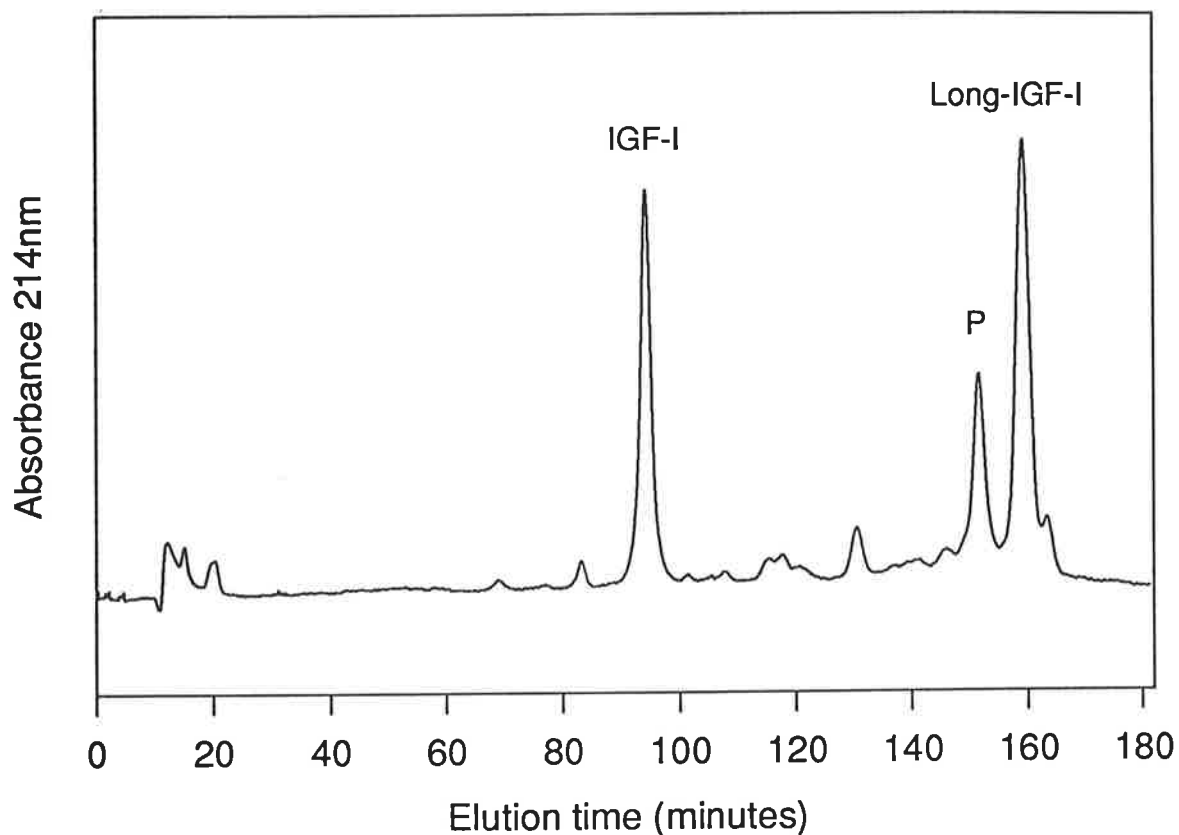
Figure 3.5: Reverse-Phase HPLC Purification of Cleaved Long-IGF-I

Reverse-phase HPLC purification of the cleavage reaction of Long-IGF-I after 24 hours at 37°C in the presence of 1 M hydroxylamine, 0.1 M Tris pH 8.65 and 1 mM EDTA. The mixture was separated on a C_4 column, 25mm by 10cm using a linear 20-38 % (v/v) acetonitrile gradient in the presence of 0.1 % TFA. The substrate (Long-IGF-I), the major product (IGF-I) and the alternative product (P) are indicated.

The remaining Long-IGF-I substrate present after 24 hours incubation in the modified cleavage reaction (Figure 3.5) was also analysed by isoelectric focusing and mass spectrometry. The IEF gel (Figure 3.4b, lane 4) showed that Long-IGF-I was modified in the reaction mixture as a more acidic band, not present in the starting material (Figure 3.4b, lane 5) was observed. Furthermore, mass spectrometry showed that the main species was Long-IGF-I (9085.6 ± 0.2 m.u.) and the other species present (40% of total) had a mass 16 units higher (9101.2 ± 0.1). The alternative product peak P (Figure 3.5) had components with masses of 9086.6 ± 0.2 and 9102.2 ± 0.4, values one unit higher than that of the remaining Long-IGF-I. Peak P also exhibits an isoelectric point (pI) approximately 1.7 pH units less than that of Long-IGF-I (Figure 3.4b, lane 3).

3.3.3.3: Analysis of the IGF-I variants

To investigate properties of the altered molecular IGF-I forms, heterogenous IGF-I produced using the relatively harsh cleavage conditions of 2M urea, 2M hydroxylamine, 0.1M Tris pH 9.0 and 45°C was characterized. The IGF-I mixture was separated using a PEI anion-exchange column (Section 2.2.2.4, Figure 3.6) and the eluted peaks were analysed by mass spectrometry and isoelectric focusing (Table 3.1 and Figure 3.4d). The heterogenous IGF-I loaded on to the PEI column contained three major forms: unmodified IGF-I and variants containing an extra 16 and 32 mass units. The first peak eluted from the PEI column (A) was unmodified IGF-I and the subsequent peaks B and C, which have an extra 16 m.u., contain single hydroxamic acid modifications (Table 3.1). The different elution times of peaks B and C suggest that they have different hydroxamic acid modifications. These results are consistent with IEF gel analysis in which peak B has a pI slightly less than that of IGF-I and peak C has a pI of 8.5 (Figure 3.4d). Mass spectrometry of
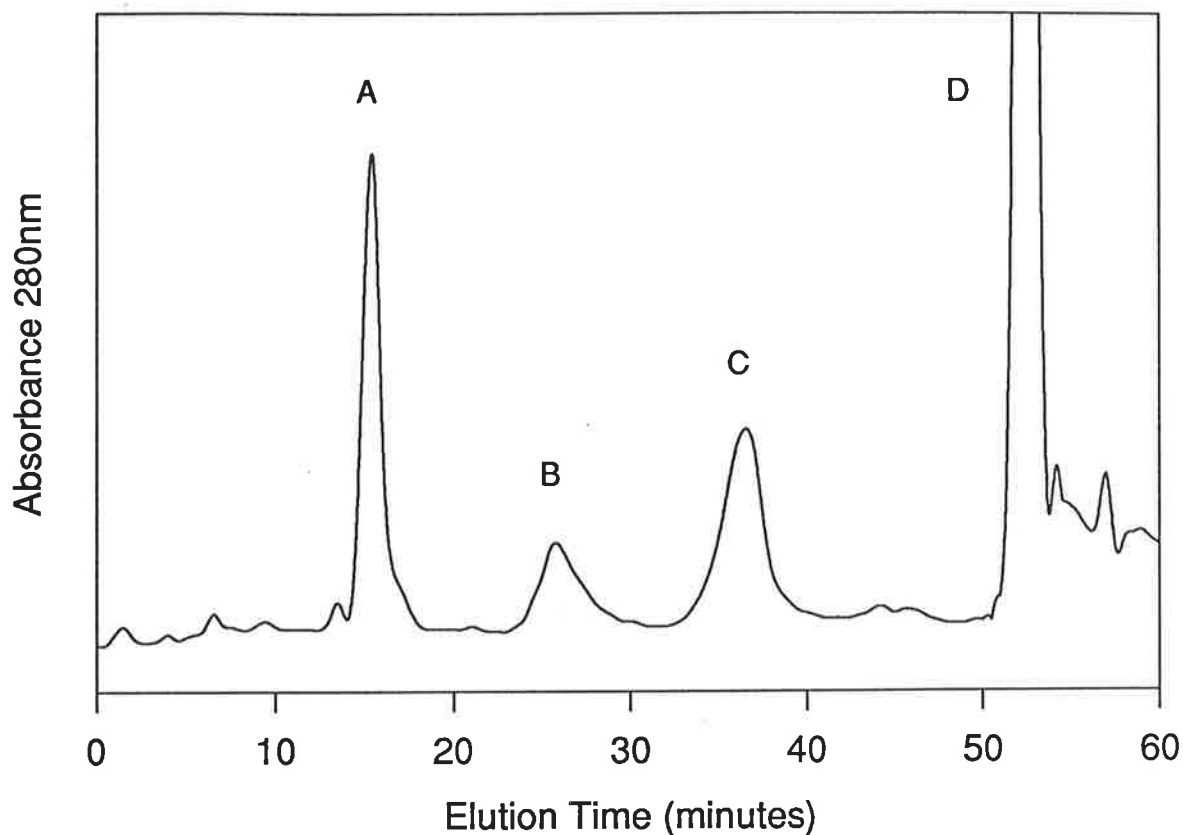
Figure 3.6: PEI Chromatography of Heterogeneous IGF-I

Polyethyleneimine (PEI) anion-exchange chromatography of IGF-I variants obtained from the cleavage of Long-IGF-I using 2 M urea, 2 M hydroxylamine at pH 9.0 and 45°C as described in Section 2.2.2.4. The eluted peaks (A to D) were collected and characterized by mass spectrometry (Table 3.1), isoelectric focusing (Figure 3.4d) and by digestion with pepsin (Figure 3.7).

Sample	Analysis by Mass Spectrometry	Approximate Composition (%)
IGF-I before ion-exchange	7648.2 ± 0.6	19
	7664.9 ± 0.4	43
	7680.3 ± 0.3	26
	7696.6	<4
	7733.9	<4
	7749.1	<4
IGF-I peak A	7631.5	<4
	7650.1 ± 0.2	88
	7666.8	<4
	7705.5	<4
IGF-I Peak B	7647.5	<4
	7664.2 ± 1.2	88
	7680.0	<4
	7717.3	<4
IGF-I Peak C	7647.5	<4
	7665.1 ± 0.2	92
	7718.9	<4
IGF-I Peak D	7649.5	<4
	7666.2 ± 0.8	12
	7681.5 ± 0.6	28
	7734.2 ± 0.8	47
	7750.7 ± 0.8	9

Table 3.1: Analysis of IGF-I Before and After Anion-exchange Chromatography.

IGF-I obtained from the cleavage of Long-IGF-I using 2 M urea, 2 M hydroxylamine at pH 9.0 and 45°C was separated on a polyethyleneimine column as described in section 3.2.3.3. The IGF-I loaded onto the column and the eluted peaks (A to D in Figure 3.6) were analysed by electrospray mass spectrometry.

peak D demonstrates the presence of multiple hydroxamic acid modifications (+32 m.u.) and other unidentified IGF-I variants (Table 3.1).

To determine which amino acids were modified to their hydroxamic acid forms, the IGF-I peaks A, B and C eluted from the PEI column (Figure 3.6) were digested by pepsin and analysed by mass spectrometry (Section 2.2.3.4). Both IGF-I peaks A and B yielded fragments of IGF-I with identical masses (Figure 3.7 and Table 3.2). However, the fragment containing residues 11-15: Val-Asp-Ala-Leu-Gln (Figure 3.7) was not resolved by mass spectrometry. To account for the difference of 16 m.u. in IGF-I peak B, Gln^{15} must have been modified to the hydroxamic acid form. A similar analysis of the pepsin digestion products of IGF-I peak C showed that the increase of mass of 16 m.u. was associated with the ADE fragment (Figure 3.7 and Table 3.2). As the sub-fragment D1 showed only the expected mass, the extra 16 m.u. present in IGF-I peak C must be a result of modification of Asn^{26}.

3.3.3.4: Properties of Asparagine-26 Modified IGF-I

The major IGF-I products produced when Long-IGF-I is cleaved using the altered cleavage reaction are unmodified IGF-I and IGF-I modified at Asn^{26} (Section 3.3.1.3 and Figure 3.4c, lane 2). To assess the effect of these molecular changes on the critical refolding step in downstream processing of recombinant proteins, unmodified IGF-I and IGF-I modified at Asn^{26} were fully reduced and refolded as described in Sections 2.2.4.4 and 2.2.4.5. In this experiment (Figure 3.8) the ratio of the alternative disulphide mismatched form to the native form at equilibrium (0.505 for IGF-I peak A and 0.498 for IGF-I peak C) was similar. However, the yield of native and mismatched isomers for IGF-I modified at Asn^{26} was slightly lower than that for unmodified IGF-I.
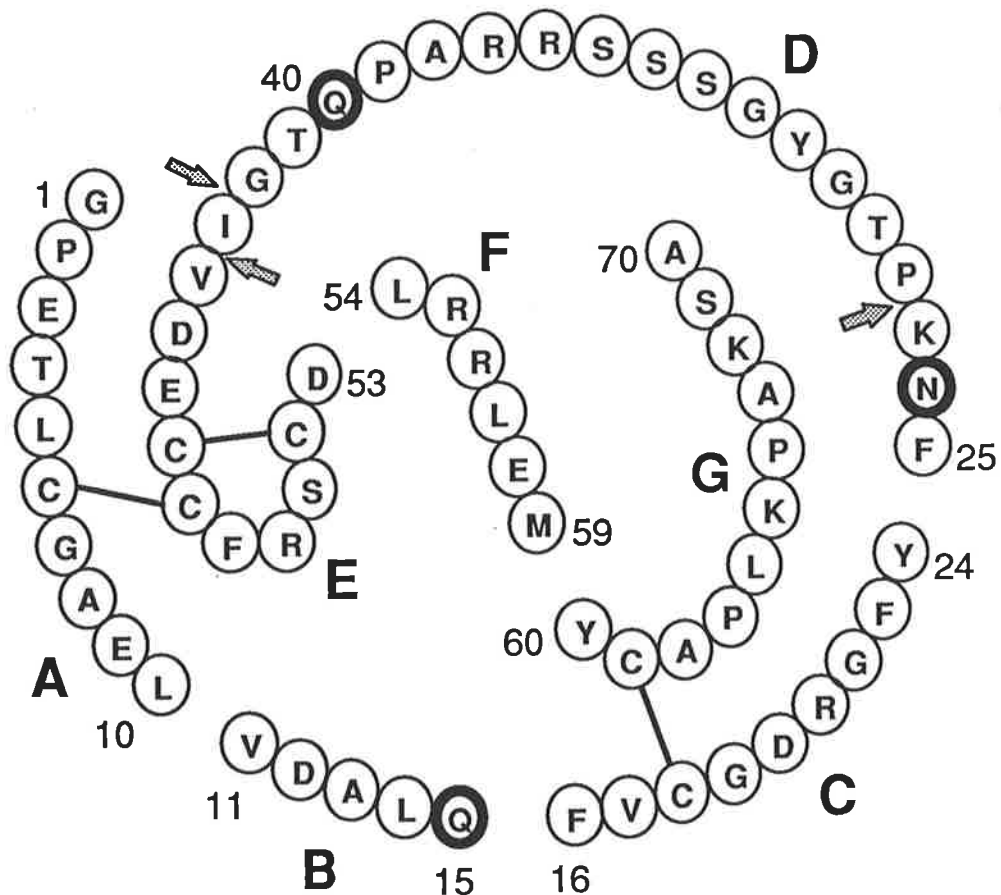
Figure 3.7: Digestion of IGF-I and variants with pepsin

The major fragments (Fragments A to G) produced following the digestion of IGF-I by pepsin are shown using the standard single letter amino acid code. Arrows are used to indicate other minor pepsin cleavage sites within the A-D-E fragment and bold circles show the potentially hydroxylamine-sensitive asparagine (N) and glutamine (Q) amino acids.

Pepsin Fragment	IGF-I residues	Expected Mass	Observed Mass		
			IGF-I peak A	IGF-I peak B	IGF-I peak C
D1	28-43	1634.8	1633.7 ± 0.1	1634.5 ± 0.3	1634.2 ± 0.2
D	25-42	1911.1	1911.0 ± 0.4	1911.6 ± 0.1	N.D.
C-G	16-24, 60-70	2209.6	2209.9 ± 0.4	2210.0 ± 0.1	2210.0 ± 0.1
A-E-F	1-10, 43-59	3073.6	3074.5 ± 0.3	3074.4 ± 0.1	3074.4 ± 0.1
A-D-E	1-10, 25-53	4167.6	4168.3 ± 0.9	4170.0 ± 0.8	4185.0 ± 0.3*
A-D-E-F	1-10, 25-59	4966.7	4967.2 ± 0.5	4968.5 ± 0.2	4984.2 ± 0.1*

*   hydroxamic acid modified fragments

N.D.   not detected

Table 3.2: Analysis of IGF-I Variants by Mass Spectrometry

IGFs separated by polyethyleneimine anion-exchange chromatography (peaks A to C in Figure 3.6) were digested by pepsin and analysed by mass spectrometry as described in Section 3.2.3.3.
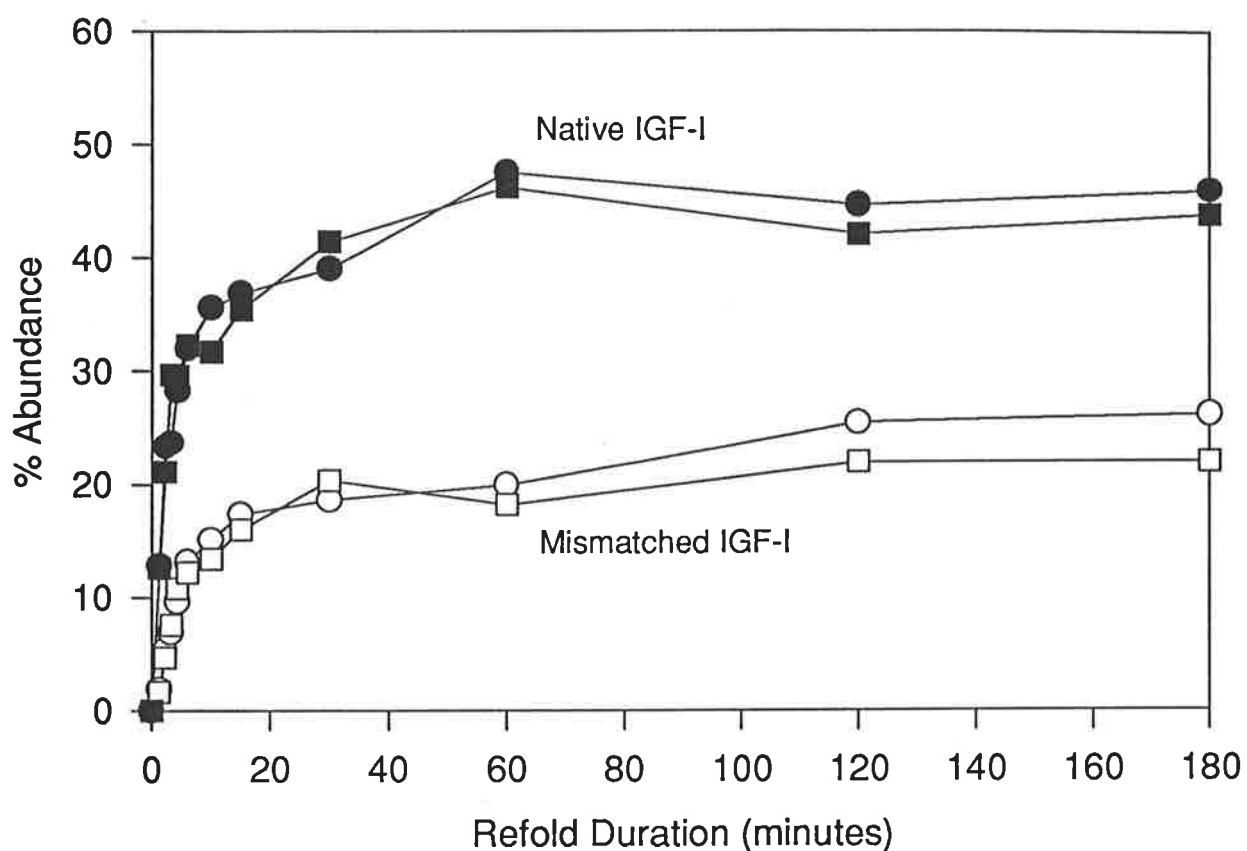
Figure 3.8: Analysis of refolding of IGF-I variants

Unmodified IGF-I (●,O) and IGF-I containing a hydroxamic acid modification at Asn[26] (■,□) were fully reduced and refolded. The yield of native IGF-I and the alternative disulphide mismatched form was calculated from the reverse-phase HPLC elution profiles taken at various time points during the refolding reactions.

To determine the influence of the modification at Asn[26] in biological applications, it was assayed for biological activity and compared to unmodified IGF-I. These assays showed that both IGF-I and IGF-I modified at Asn[26] had similar potencies: the ED_{50} value obtained for binding to the type 1 IGF receptor was 2.5 ng/ml and 3.1 ng/ml respectively (Figure 3.9a) and for the stimulation of protein synthesis the amount of peptide required to give half maximal stimulation was 3.5 ng/ml and 3.2 ng/ml respectively (Figure 3.9b).

3.4: DISCUSSION

The downstream processing of inclusion bodies has three areas of difficulty: (i) sensitivity to bacterial proteases, (ii) refolding to the native conformation and (iii) removal of the co-expressed fusion protein partner. Problems associated with the first two areas are minor for fusion proteins containing human IGF-I and human IGF-I analogues as they appear to be resistant to bacterial proteases and refold relatively easily (King *et al.*, 1992; Francis *et al.*, 1992). However, similar fusion proteins containing chicken IGF-I (Upton *et al.*, 1992) or human IGF-II (Francis *et al.*, 1993) are both sensitive to proteases and refold less well than their human IGF-I counterparts. The third area identified above is problematic as the very reaction that results in the cleavage of the fusion protein can also degrade the desired product. For the IGF-I fusion proteins used in this study, hydroxylamine is used to cleave the sensitive asparaginyl-glycyl peptide bond linker to yield IGF-I with a native glycine N-terminus. While the chemical cleavage reaction is specific, other reactions may occur that modify sensitive amino acid side-chains to their hydroxamic acid forms (Canova-Davis *et al.*, 1992).

Figure 3.9: Biological properties of IGF-I variants

Unmodified IGF-I (●) and IGF-I containing a hydroxamic acid modification at Asn^{26} (■) were compared in biological assays: (a) Binding to the type-1 IGF receptor was measured as the percentage of ^{125}I-labelled IGF-I bound in the absence of added unlabelled IGF peptide; (b) Stimulation of protein synthesis was measured as the increased incorporation of ^{3}H-labelled leucine into total cell protein compared to a buffer control. Each IGF was tested in triplicate at each peptide concentration. The standard errors of the estimate of the mean value (S.E.M.) are indicated by descending bars where they are larger than the symbols.
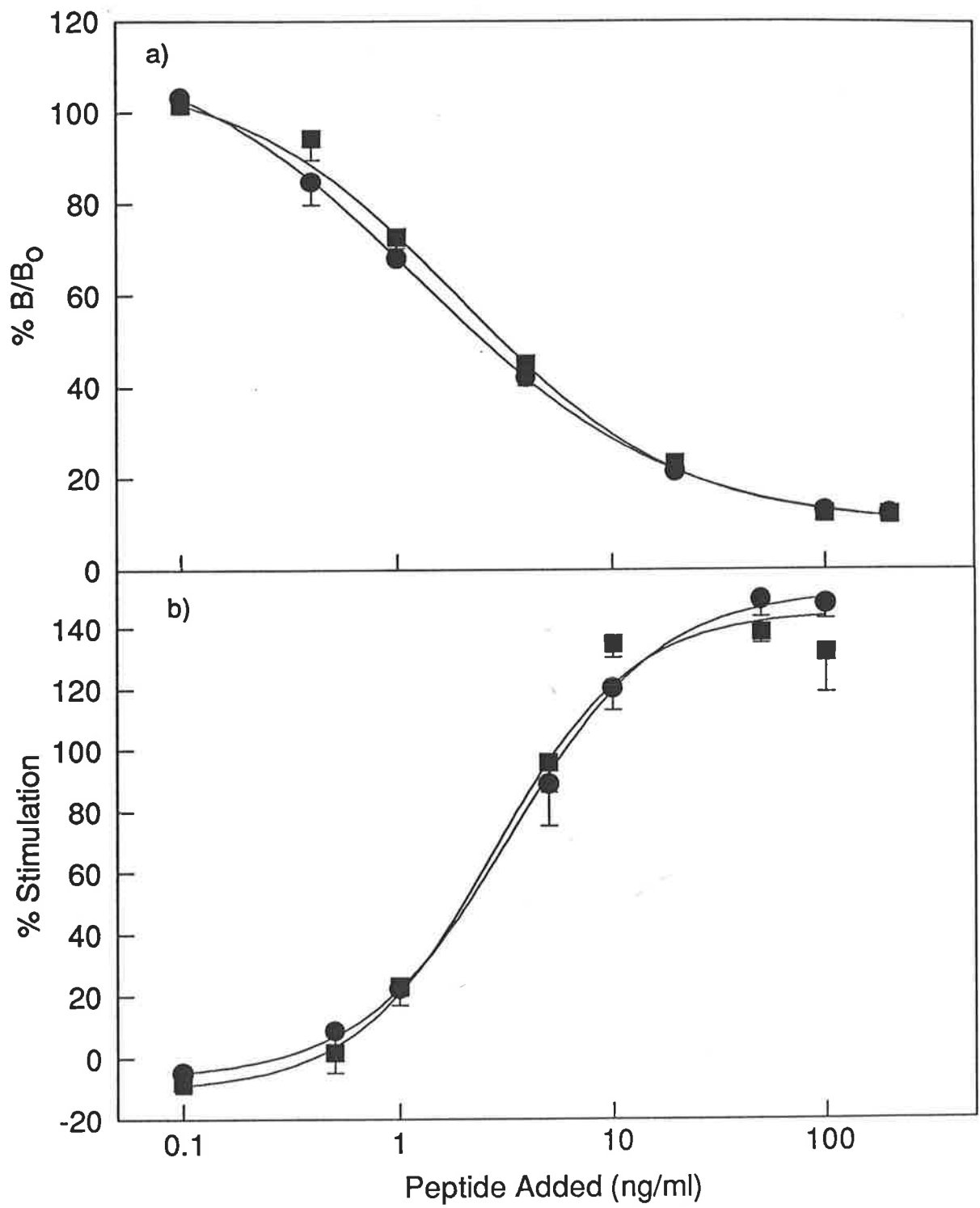
The observed hydroxylamine cleavage reaction of the Long-IGF fusion protein is consistent with known reaction mechanisms. The hydroxylamine cleavage of Long-IGF-I after 24 hours at 37°C in the presence of 1 M hydroxylamine, 0.1 M Tris pH 8.65 and 1 mM EDTA (Figures 3.4 and 3.5) yielded native IGF-I and an alternative product P that has a mass one unit greater than Long-IGF-I. A competing reaction to the cleavage of the Asn-Gly bond is the formation a ß-Asp-Gly (iso-aspartic acid) product (Bornstein and Balian, 1977) that has a mass one unit greater than the parent sequence, a more acidic character and is more hydrophilic (Geiger and Clarke, 1987; Stephenson and Clarke, 1989). The alternative product P exhibits the same properties and is likely to be Long-IGF-I converted to a ß-Asp form ([Met^1]-pGH(1-11)-Val-(ß-Asp)-IGF-I).

The formation of a stable cyclic imide between Asn and the following Gly precedes nucleophilic attack by hydroxylamine, rearrangement and release of the amino-terminal Gly in the cleavage reaction (Bornstein and Balian, 1977). While Asn can form cyclic imides with other flanking carboxy-terminal amino acids (Geiger and Clarke, 1987; Stephenson and Clarke, 1989; Tyler-Cross and Schirch, 1991), cleavage between Asn and other neighbouring amino acids by hydroxylamine has not been observed. However, it is possible that cyclic imide formation provides a route to formation of a hydroxamic acid and other variants. Analysis of heterogenous IGF-I showed that the hydroxamic acid modified residues could be at one or more different sites. The pepsin digestion of IGF-I variants with an increase of 16 mass units showed that Gln^{15} or Asn^{26} were modified to hydroxamates (Table 3.2). Furthermore, peak D (Table 3.1) contained variants with incremental increases of 16 mass units, suggesting that in some cases all susceptible residues were present as hydroxamates. The observation that Asn^{26} in IGF-I is particularly susceptible to

modification compared to Gln[15] and Gln[40] (Figure 3.4c and 3.4d) may be due to the ease of formation of the cyclic imide for Asn residues compared to Gln residues (Blodgett *et al.*, 1985). An alternative explanation for the occurrence of hydroxamic acids in IGF-I, in particular those of Gln residues produced under the more harsh cleavage conditions, is the direct hydroxylaminolysis of amide to hydroxamate groups (Ramachandran and Narita, 1958). However, this reaction is probably slow as generally more severe conditions are required (3 M hydroxylamine, 60°C, pH 6.5 - 7.5, 24 hours) (Ramachandran and Narita, 1958).

The application of the optimized hydroxylamine chemical cleavage reaction employing lowered pH, temperature and hydroxylamine concentration results in mild conditions where only unmodified IGF-I is produced. However, under these conditions the cleavage of the labile asparaginyl-glycyl bond was reduced. To improve the overall yield, the cleavage reaction was developed to produce IGF-I that contains a high proportion of unmodified peptide with an acceptable cleavage rate. The former conditions recommended for cleavage of Long-IGF-I produced an IGF-I product containing a complex mixture of one or more hydroxamic acid variants whereas the new conditions produce only one hydroxamic acid derivative at Asn[26], present at a significantly reduced level. While the improved yield of unmodified IGF-I is modest (from 13 to 23%) the greatest advantage is that the heterogeneity of IGF-I isolated after reverse-phase HPLC is reduced from 75% to 30%. This factor is important if a subsequent ion-exchange chromatography step is used to separate the IGF-I variants as the total amount of protein that can be loaded on to the high-performance column is limited.

The characterization of the refolding properties of the IGF-I variant containing a hydroxamate at Asn[26] demonstrates that this variant is essentially similar to the unmodified form by virtue of the relative stability of the native and the alternative

forms at equilibrium. Furthermore, the hydroxamic acid modification of Asn[26] in IGF-I does not affect the folding process as the kinetics of the oxidative folding reaction is the same as that of unmodified IGF-I (Figure 3.8). Because there was no difference in the folding reactions of the two IGF-I forms, IGF-I which contained 30 % of the hydroxamic acid form was used in the refolding reactions investigated in the subsequent chapters. Other IGF-I analogues containing hydroxamic acid modifications were also used in these experiments. [Gly[3]]IGF-I produced by the hydroxylamine cleavage of [Met[1]]pGH(1-46)-Val-Asn-[Gly[3]]IGF-I (King *et al.*, 1992) and [Arg[3]]IGF-I produced by the hydroxylamine cleavage of [Met[1]]pGH(1-11)-Val-Asn-[Arg[3]]IGF-I using the same method were shown by mass spectrometry to contain up to 40 % of a single variant with a mass 16 units higher. [Lys[9]]IGF-I which was produced using the optimized hydroxylamine cleavage reaction developed above showed the expected mass with approximately 10 % of the hydroxamic acid modified form. While the hydroxamic acid variants of these three IGF-I analogues were not identified directly, they are likely to be modifications at Asn[26] since this residue is particularly sensitive to modification in the IGF-I molecule. Like IGF-I, the heterogeneity present in each IGF-I analogue preparation is not expected to influence the refolding properties. Indeed, the influence of the amino acid substitutions on the refolding properties is far greater than could be attributed to any hydroxamic acid modifications (Chapter 4).

CHAPTER 4

THE EFFECT OF CHANGES TO THE AMINO ACID SEQUENCE ON THE OXIDATIVE FOLDING OF IGF-I

4.1: INTRODUCTION

The *in vitro* oxidative folding pathway of insulin-like growth factor-I (IGF-I) has been partially characterized (Hober *et al.*, 1992). A feature of this pathway is that the fully reduced IGF-I molecule refolds to produce *two* major three-disulphide folding isomers at equilibrium (Hober *et al.*, 1992; Miller *et al.*, 1993). One isomer has the native disulphide arrangement of Cys^{18} - Cys^{61}, Cys^{47} - Cys^{52}, Cys^{6} - Cys^{48} (Raschdorf *et al.*, 1988; Axelsson *et al.*, 1992) and the other has an alternative disulphide arrangement of Cys^{18} - Cys^{61}, Cys^{48} - Cys^{52}, Cys^{6} - Cys^{47}. The non-native isomeric form has not been observed *in vivo*. The two products have a similar thermodynamic stability with the native conformation being slightly more stable than the alternative isomer. Calculations of the free energy difference ($\Delta\Delta G$) between the two isomers vary slightly but are in the region of 0.25 kcal / mol (Miller *et al.*, 1993) or 0.5 kcal / mol (Hober *et al.*, 1992). A striking observation is that the relative proportion of the two isomers is similar irrespective of the conditions used to refold the protein (Saito *et al.*, 1987; Iwai *et al.*, 1989; Hober *et al.*, 1992; Miller *et al.*, 1993). Furthermore, IGF-I variants such as [Met^1]IGF-I (Hodgkinson *et al.*, 1989) and Ala-Glu-IGF-I (Hejnæs *et al.*, 1992) exhibit similar refolding properties. It is notable that the *Saccharomyces cerevisiae* secretion of recombinant human IGF-I also yields the two isomeric forms (Raschdorf *et al.*, 1988; Elliot *et al.*, 1990; Axelsson *et al.*, 1992). *In vitro*, the purified isomers will readily undergo disulphide rearrangement at physiological pH in the presence of a redox buffer to yield the same ratio of products as when refolded from the fully reduced species (Miller *et al.*, 1993). This observation demonstrates an oxidative folding pathway governed by thermodynamic control.

In the development of an efficient expression system for analogues of IGF-I with novel biological properties we observed that a 13-amino acid N-terminal fusion protein partner facilitated downstream refolding (Francis *et al.*, 1992). Thus, when a fusion

protein analogue of IGF-I with a change in the B-domain ([Met^1]pGH(1-11)-Val-Asn-[Arg^3]IGF-I: Long-[Arg^3]IGF-I) was refolded from the fully reduced form, most of the protein was correctly folded (Francis *et al.*, 1992). This folding outcome was significantly different from that of IGF-I (Saito *et al.*, 1987; Iwai *et al.*, 1989). However, it was unclear whether the 13-amino acid N-terminal fusion protein partner or the Arg substitution at Glu^3, or both, were responsible for the favourable folding outcome. To investigate the role of the 13-amino acid N-terminal extension and the charge substitution at Glu^3 on the refolding of Long-[Arg^3]IGF-I, the folding properties of Long-IGF-I and [Arg^3]IGF-I were compared with those of Long-[Arg^3]IGF-I and the native sequence IGF-I. To further examine the contribution of charge at position 3 of IGF-I, the folding properties of the neutral amino acid substitution, [Gly^3]IGF-I and its 'Long' counterpart Long-[Gly^3]IGF-I were studied. The folding properties of another analogue of IGF-I, containing a charge substitution of Lys for Glu^9 ([Lys^9]IGF-I) and its 'Long' counterpart Long-[Lys^9]IGF-I were also investigated as the charge reversal in helix 1 (IGF-I residues 8-18: Cooke *et al.*, 1991; Sato *et al.*, 1993; Sato *et al.*, 1992) is potentially disruptive and may have a detrimental effect on refolding. To examine the possibility that the N-terminal extension alone may facilitate refolding of IGFs *in trans*, a chemically synthesized 'Long' peptide was added to the refolding reaction of IGF-I and compared to that of Long-IGF-I. As buffer constituents may affect the refolding outcome, the effect of the urea concentration and the effect of redox reagents were also investigated for IGF-I or Long-[Arg^3]IGF-I.

4.2: METHODS

The materials and methods are described in detail in Chapter 2. However, for clarity the methods employed in this chapter are summarized here.

4.2.1: The Oxidative Folding of IGFs

4.2.1.1: The Oxidative Folding of IGFs from the Fully Reduced Form

The crude refolded fusion proteins Long-IGF-I, Long-[Gly^3]IGF-I, Long-[Arg^3]IGF-I and Long-[Lys^9]IGF-I containing a mixture of the native molecule, a misfolded form and a N-terminally modified form (Chapter 3) were purified by preparative reverse-phase HPLC (Section 2.2.2.3) to isolate the most abundant variant present (the native isomer). Each purified fusion protein was then fully reduced (Section 2.2.4.4) and then refolded at a protein concentration of 0.1 mg / ml at 22°C in a buffer containing 2 M urea, 10 mM glycine, 0.1 M Tris pH 8.7, 1 mM EDTA, 0.4 mM dithiothreitol and 1 mM 2-hydroxyethyl disulphide (Section 2.2.4.5). The progress of the reaction was monitored and the folding intermediates stabilized by removing samples at various times and acidifying them to pH 2.1 with TFA. Analysis was carried out by microbore C$_4$ reverse-phase HPLC employing a gradient of 20-45 % (v/v) acetonitrile over 25 minutes in the presence of 0.1 % (v/v) TFA at a flow rate of 0.5 ml / min. The reverse-phase HPLC elution profiles were used to calculate the percentage abundance of folding products at each time. In a similar manner, purified IGF-I and its analogues [Gly^3]IGF-I, [Arg^3]IGF-I and [Lys^9]IGF-I were fully reduced and refolded. Control incubations of each of the IGFs were carried out using the same refolding buffer but without the oxidizing agent 2-hydroxyethyl disulphide present. After 120 minutes the control reactions were sampled for reverse-phase HPLC analysis as described above.

4.2.1.2: The Oxidative Folding of IGF-I in the Presence of a Synthetic 'Long' Peptide

IGF-I was fully reduced and refolded as described in section 4.2.1.1. However, the reaction was carried out in the absence and presence of increasing

concentrations of the synthetic peptide containing the sequence MFPAMPLSSLFVNGPR-NH$_2$ (Section 2.1.2.2). The molar ratios of IGF-I to the synthetic peptide in the refolding reactions were 1 to 0.1, 1, and 10. The reactions were analysed by microbore C$_4$ reverse-phase HPLC after 180 minutes (Section 2.2.2.5).

4.2.2: The Effect of Buffer Constituents on Refolding

4.2.2.1: The Effect of Urea on the Refolding Reactions of IGF-I and Long-[Arg^3]IGF-I

To determine the effect of the urea on the refolding reaction when refolding from the fully reduced form, IGF-I and Long-[Arg^3]IGF-I were refolded as described in Sections 2.2.4.4 and 2.2.4.5 but in the presence of increasing concentrations of urea. After 180 minutes of refolding, the refold mixtures were sampled for analysis by microbore C$_4$ reverse-phase HPLC (Section 2.2.2.5). In a separate experiment, IGF-I and Long-[Arg^3]IGF-I were refolded from the fully reduced form in 8 M urea for 180 minutes. At 180 minutes the refold mixture was divided, one portion was diluted 1:4 with the refold buffer without urea (10 mM glycine, 0.1 M Tris pH 8.7, 1 mM EDTA, 0.4 mM dithiothreitol and 1 mM 2-hydroxyethyl disulphide) and the other portion was diluted 1:4 with the same buffer but containing 8 M urea. After a further 180 minutes the mixtures were sampled for analysis by microbore C$_4$ reverse-phase HPLC.

4.2.2.2: The Effect of Redox Buffer on the Refolding Reaction of Long-[Arg^3]IGF-I

To determine the effect of the redox buffer on the refolding of Long-[Arg^3]IGF-I, the protein was refolded using either dithiothreitol (DTT) and 2-hydroxyethyl disulphide (2-HED) or reduced (GSH) and oxidized (GSSG) glutathione. In each refolding reaction, freeze-dried reduced Long-[Arg^3]IGF-I (Section 2.2.4.4) was reacted in a 2 M urea, 0.1 M tris pH 8.7, 10 mM glycine, 1 mM

EDTA buffer at 22°C containing different ratios of the redox constituents. After 180 minutes refolding, the mixtures were acidified with TFA and analysed using microbore C_4 reverse-phase HPLC employing a gradient of 20-45 % (v/v) acetonitrile over 75 minutes in the presence of 0.1 % (v/v) TFA. The folding products were identified by their elution times.

4.3: RESULTS

4.3.1: Refolding Reactions of IGF-I and IGF-I Analogues

To investigate the role of the 13-amino acid N-terminal extension and the charge substitution at Glu^3 on the refolding of Long-[Arg^3]IGF-I, the folding properties of Long-IGF-I and [Arg^3]IGF-I were compared with those of Long-[Arg^3]IGF-I and the native sequence IGF-I (Figure 4.1). To further examine the contribution of charge at position 3 of IGF-I, the folding properties of the neutral amino acid substitution, [Gly^3]IGF-I and its 'Long' counterpart Long-[Gly^3]IGF-I were studied. Another analogue of IGF-I was constructed containing both the N-terminal extension and a charge reversal, but at a position in the B-domain other than that of Glu^3. A charge substitution of Lys for Glu^9, near the start of helix I (IGF-I residues 8-18), was chosen. Lys was used instead of Arg because of the more similar packing volume to Glu (171 $Å^3$ versus 155 $Å^3$; Chothia, 1975) compared with Arg (202 $Å^3$). Since such a charge reversal in the helix is potentially disruptive and may have a detrimental effect on refolding, the possibility that the N-terminal extension could also modulate the refolding of this analogue was also investigated. In the subsequent refolding experiments the most abundant product of folding, which also exhibits the greatest biological potency (results not shown), is referred to as the native isomer and the most abundant non-native folding product is referred to as the alternative isomer.
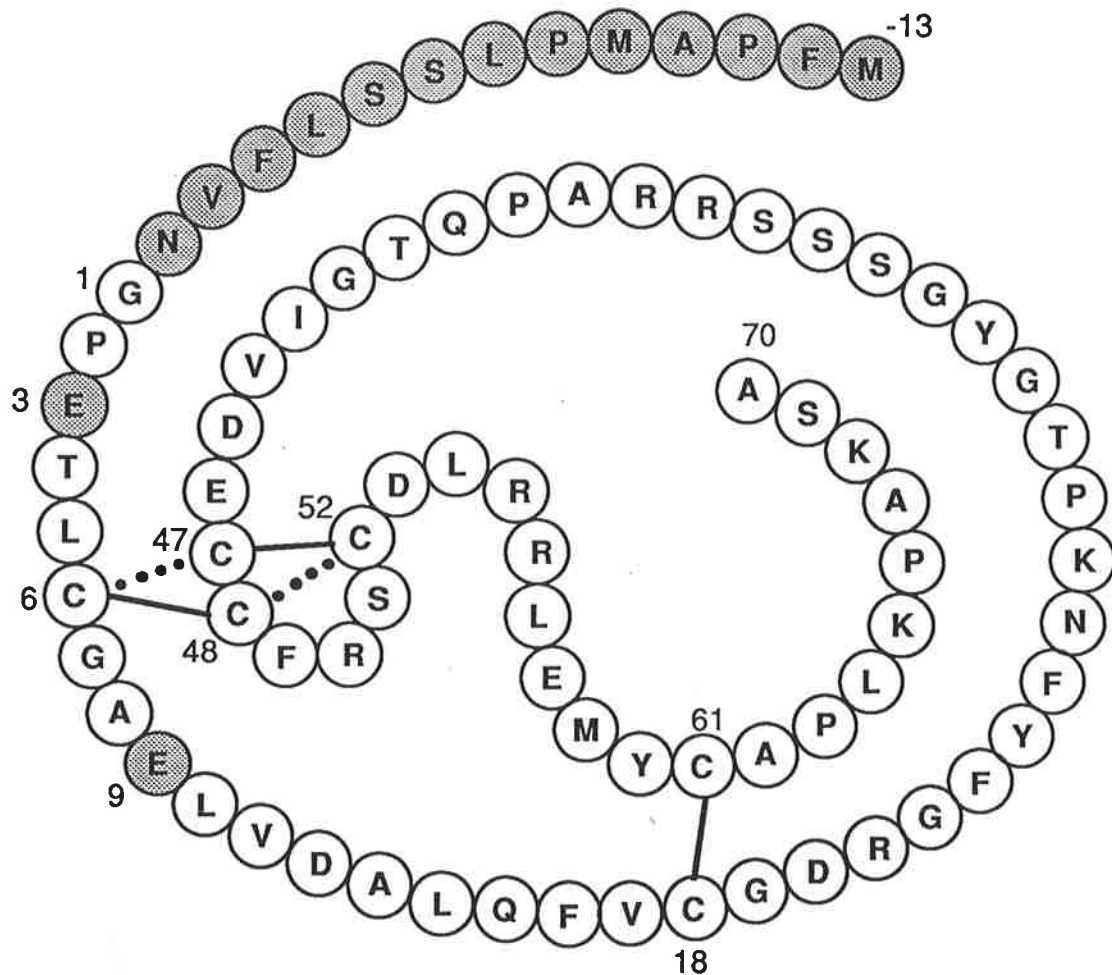
Figure 4.1: Analogues of IGF-I

The analogues of IGF-I are substitutions of Glu (E) at position 3 by Gly (G) or Arg (R) and at position 9 by Lys (K) in the B-domain, either separated from or still attached to the 13-amino acid N-terminal extension. The primary structure of IGF-I is shown using standard single letter codes with the native disulphide assignment (solid lines) and with a mismatched disulphide arrangement involving Cys^6, Cys^{47}, Cys^{48} and Cys^{52} that corresponds to the alternative folding isomer (dotted lines).

Within 1.25 minutes of initiating refolding, reduced IGF-I formed at least 15 different folding intermediates (Figure 4.2). These intermediates rearranged with time becoming less abundant as the two major forms present at equilibrium became more abundant. After 10 minutes of folding, the native and alternative isomers were present in similar proportions to those observed at 120 minutes. At equilibrium, the native and alternative isomers comprised 69 % of the total folding products (Table 4.1). The remaining folding products (31 % of the total) are presumably stable folding intermediates. The analysis of the oxidative refolding reaction of [Gly³]IGF-I (Figure 4.3 and Table 4.1) showed that at equilibrium the abundance of the native isomer for [Gly³]IGF-I was greater than that of IGF-I, whereas the abundance of the alternative isomer was approximately the same. In this refolding reaction, the folding intermediates present at equilibrium were less well-populated. Like [Gly³]IGF-I, folding intermediates of [Arg³]IGF-I at equilibrium were not well-populated (Figure 4.4 and Table 4.1). However, the native isomer was more stable than the alternative isomer with the percentage abundance of the two isomers being 75 % and 11 % of the total folding products respectively. In contrast to the refolding properties of [Gly³]IGF-I and [Arg³]IGF-I, the refolding of [Lys⁹]IGF-I showed an increased stability of the folding intermediates at the expense of the native and alternative isomers (Figure 4.5 and Table 4.1). Thus, the IGF-I analogues with mutations in the B-domain clearly affect folding; the charge reversal at Glu⁹ decreased the relative stability of the native isomer, whereas the charge reversal at Glu³ increased the stability of the native isomer.

Although the oxidative refolding reactions of the IGF fusion proteins Long-IGF-I, Long-[Gly³]IGF-I, Long-[Arg³]IGF-I and Long-[Lys⁹]IGF-I (Figures 4.6, 4.7, 4.8 and 4.9 respectively) showed elements in common with the folding of IGF-I, there were a few important differences. Like the normal length IGFs, the fully-reduced fusion protein quickly formed a number of folding intermediates. However, there appeared to be fewer

% Abundance of folding isomers at equilibrium			
IGF-I Analogue	Native	Alternative	Others
IGF-I	45,45	24,24	31,31
[Gly^3]IGF-I	64,67	22,22	14,11
[Arg^3]IGF-I	75,74	10,12	15,14
[Lys^9]IGF-I	35,38	13,14	52,48
Long-IGF-I	62,65	23,27	15,8
Long-[Gly^3]IGF-I	58,57	23,26	19,17
Long-[Arg^3]IGF-I	80,79	8,10	12,11
Long-[Lys^9]IGF-I	73,72	15,17	12,11

Table 4.1: Folding Patterns at Equilibrium

The percentage abundance of the folding isomers at equilibrium for each of the IGF-I analogues was calculated from the reverse-phase HPLC profiles for two separate folding experiments.

Figures 4.2 to 4.9: Analysis of the refolding reactions of IGF-I and IGF-I Analogues

The oxidative refolding reactions of reduced IGF-I, [Gly^3]IGF-I, [Arg^3]IGF-I, [Lys^9]IGF-I, Long-IGF-I, Long-[Gly^3]IGF-I, Long-[Arg^3]IGF-I and Long-[Lys^9]IGF-I were monitored by acid-quenching samples with time and analysis on microbore C_4 reverse-phase HPLC. Elution of protein was monitored as absorbance at 214nm, performed on a gradient of 20 - 45 % (v/v) acetonitrile over 25 minutes in the presence of 0.1 % (v/v) TFA. The elution profiles were analysed to calculate the percentage abundance of folding isomers and intermediates with time (insets). The reduced protein (R), the native folding isomer (N) and the alternative folding isomer (A) are indicated.
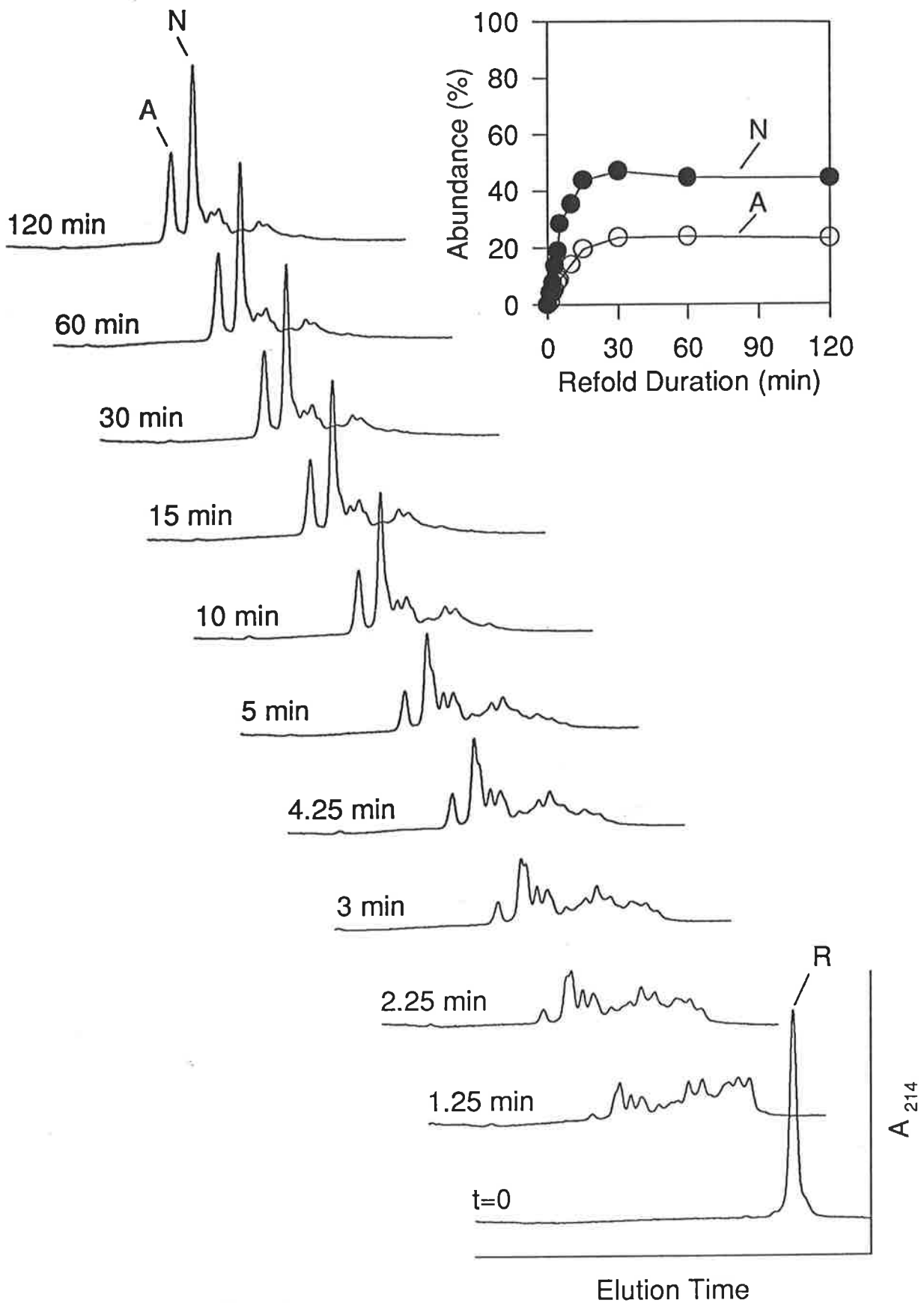
Figure 4.2: IGF-I
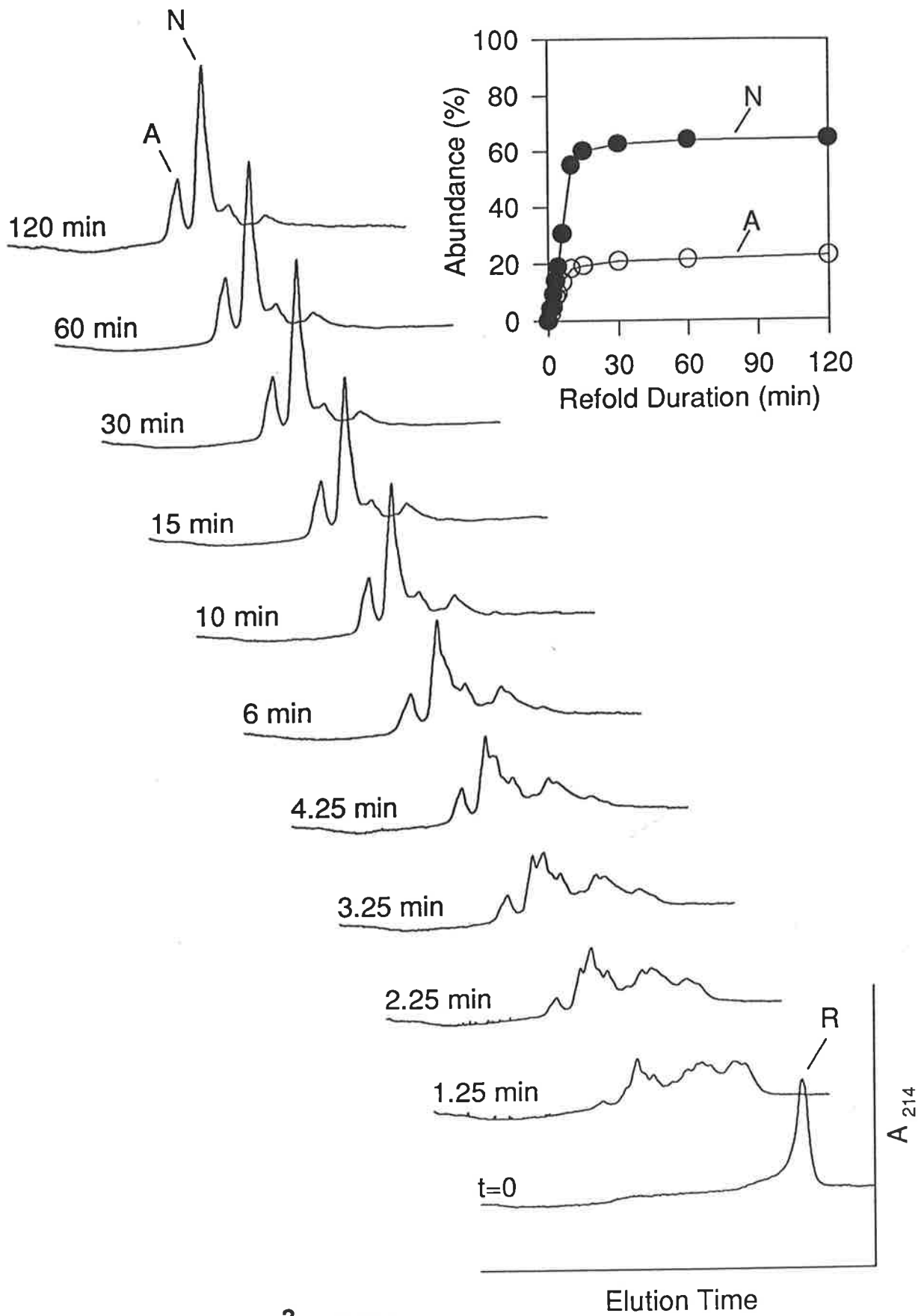
Figure 4.3: [Gly³]-IGF-I
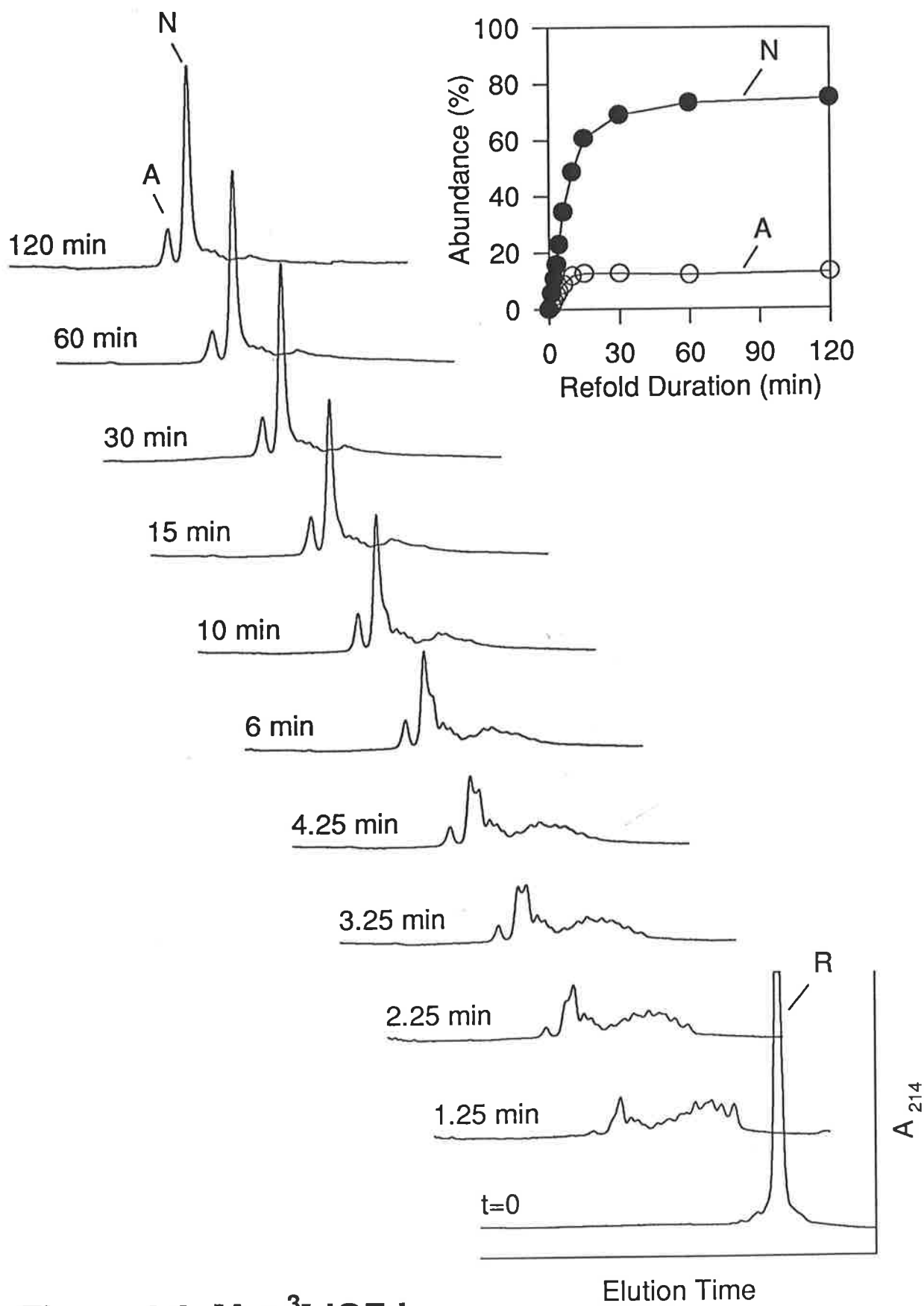
Figure 4.4: [Arg^3]-IGF-I

Figure 4.5: [Lys^9]-IGF-I

N

A

120 min

60 min

30 min

15 min

10 min

5 min

4.25 min

3 min

I

2.25 min

1.25 min

R

A_{214}

t=0

Elution Time

Abundance (%)

100

80

60

40

20

0

N

A

0   30   60   90   120

Refold Duration (min)
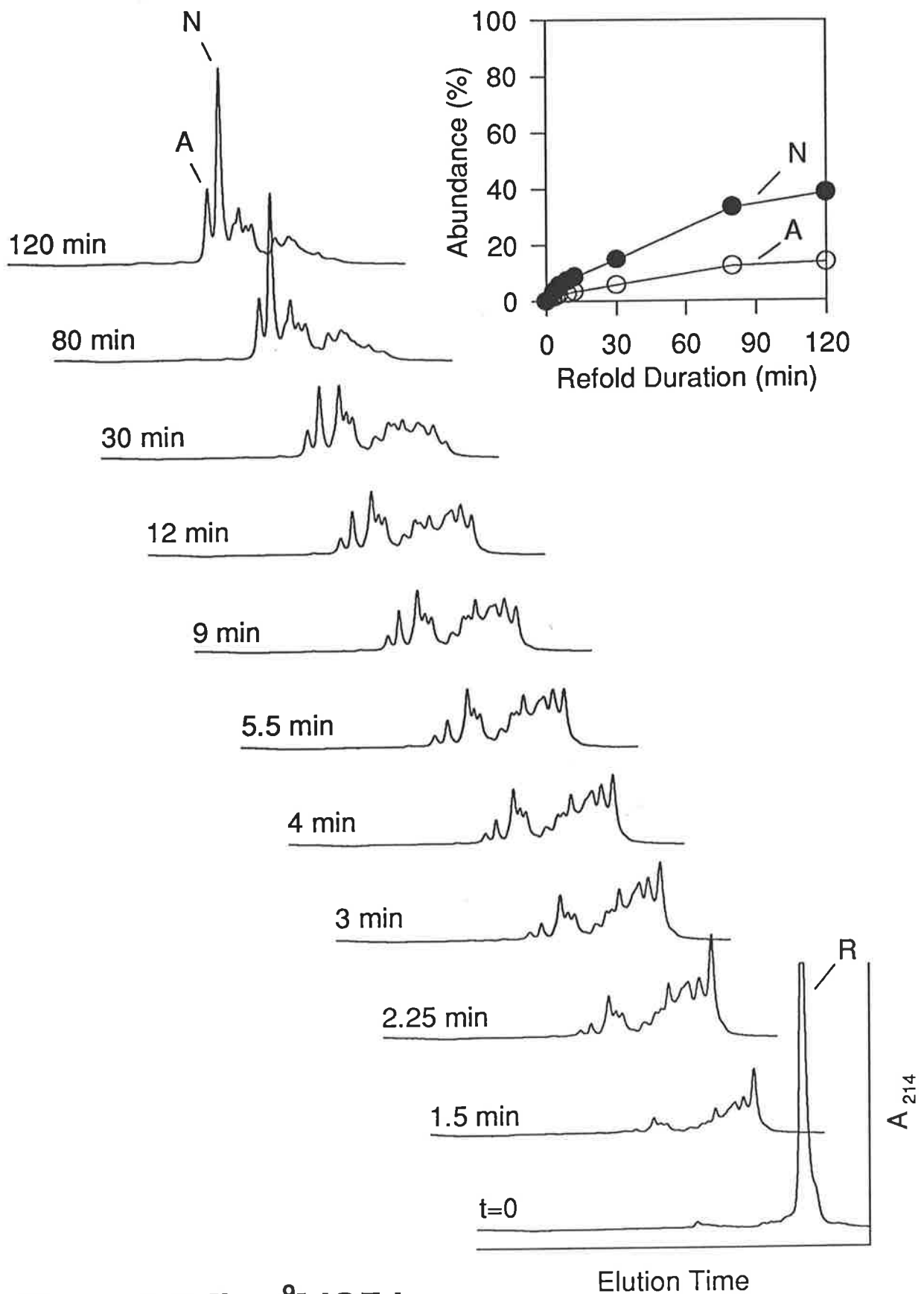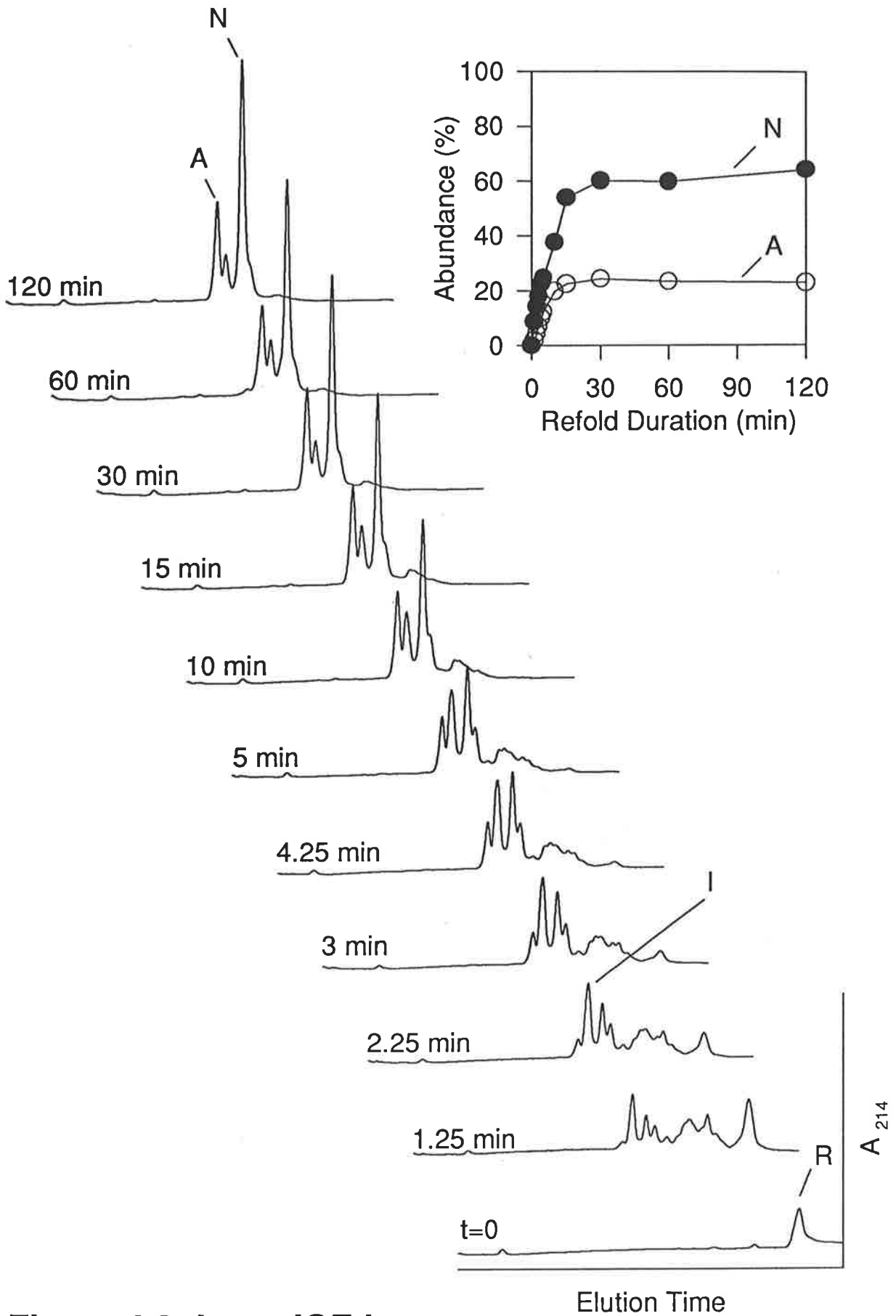
Figure 4.6: Long-IGF-I
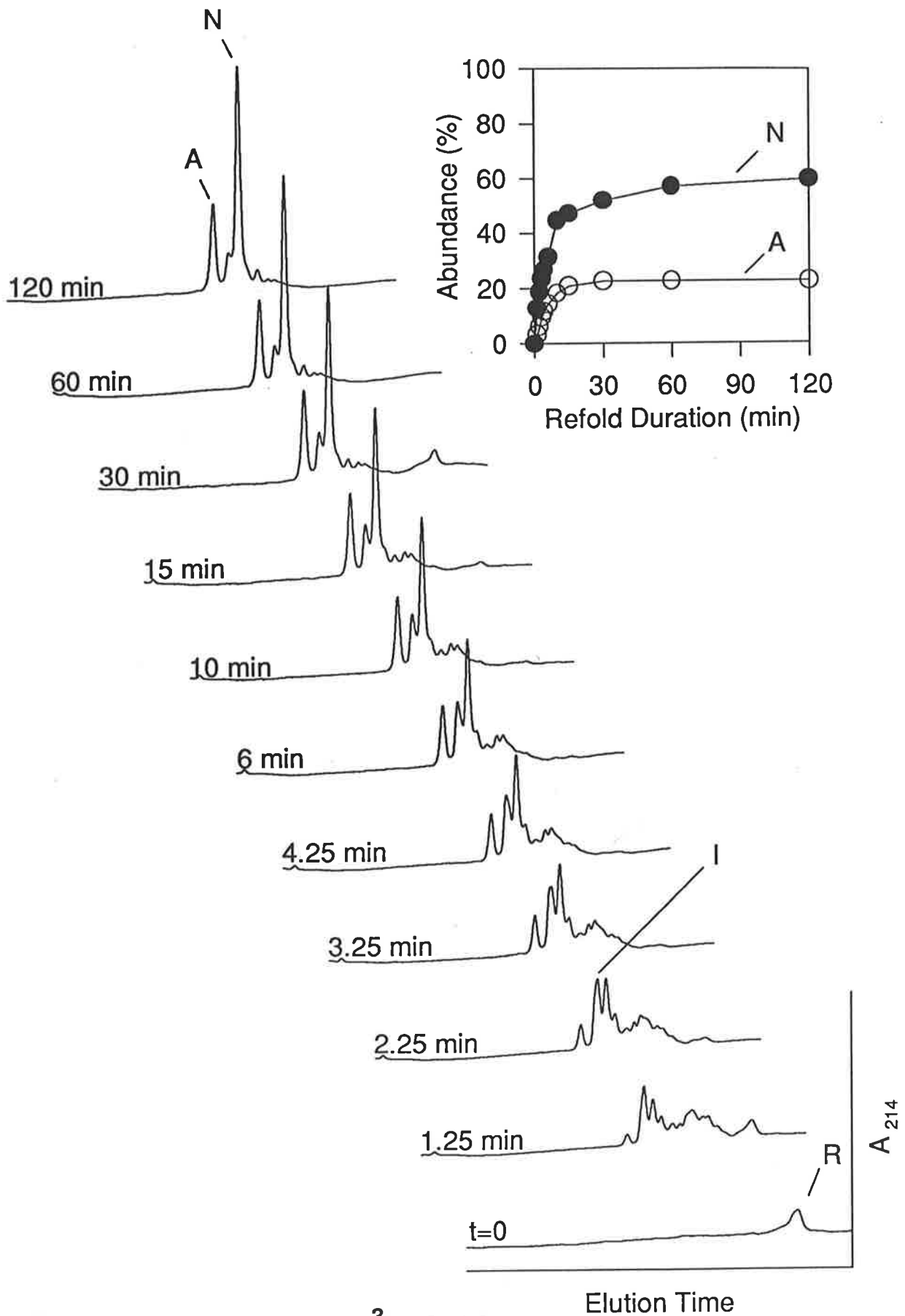
Figure 4.7: Long-[Gly^3]-IGF-I
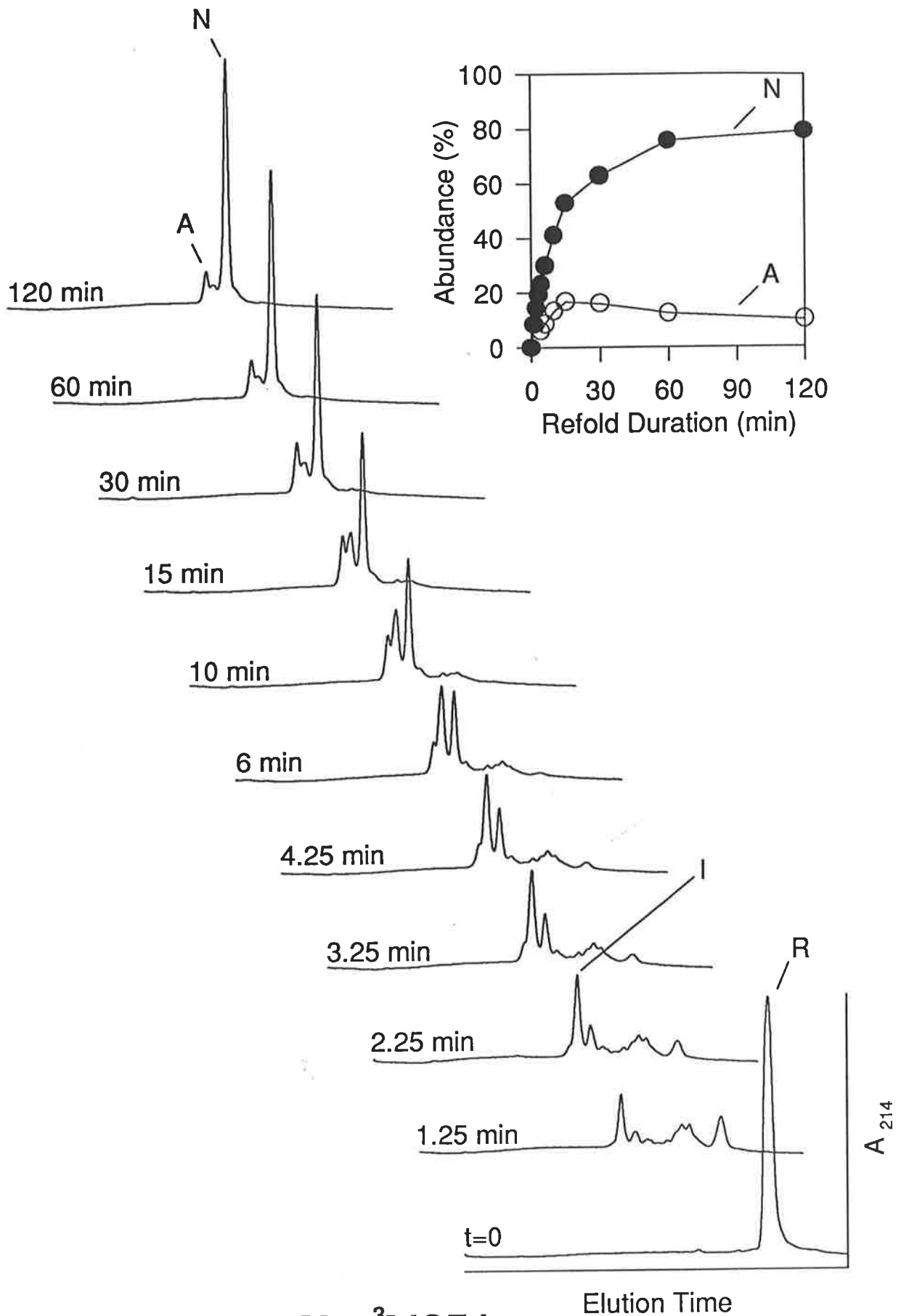
Figure 4.8: Long-[Arg³]-IGF-I
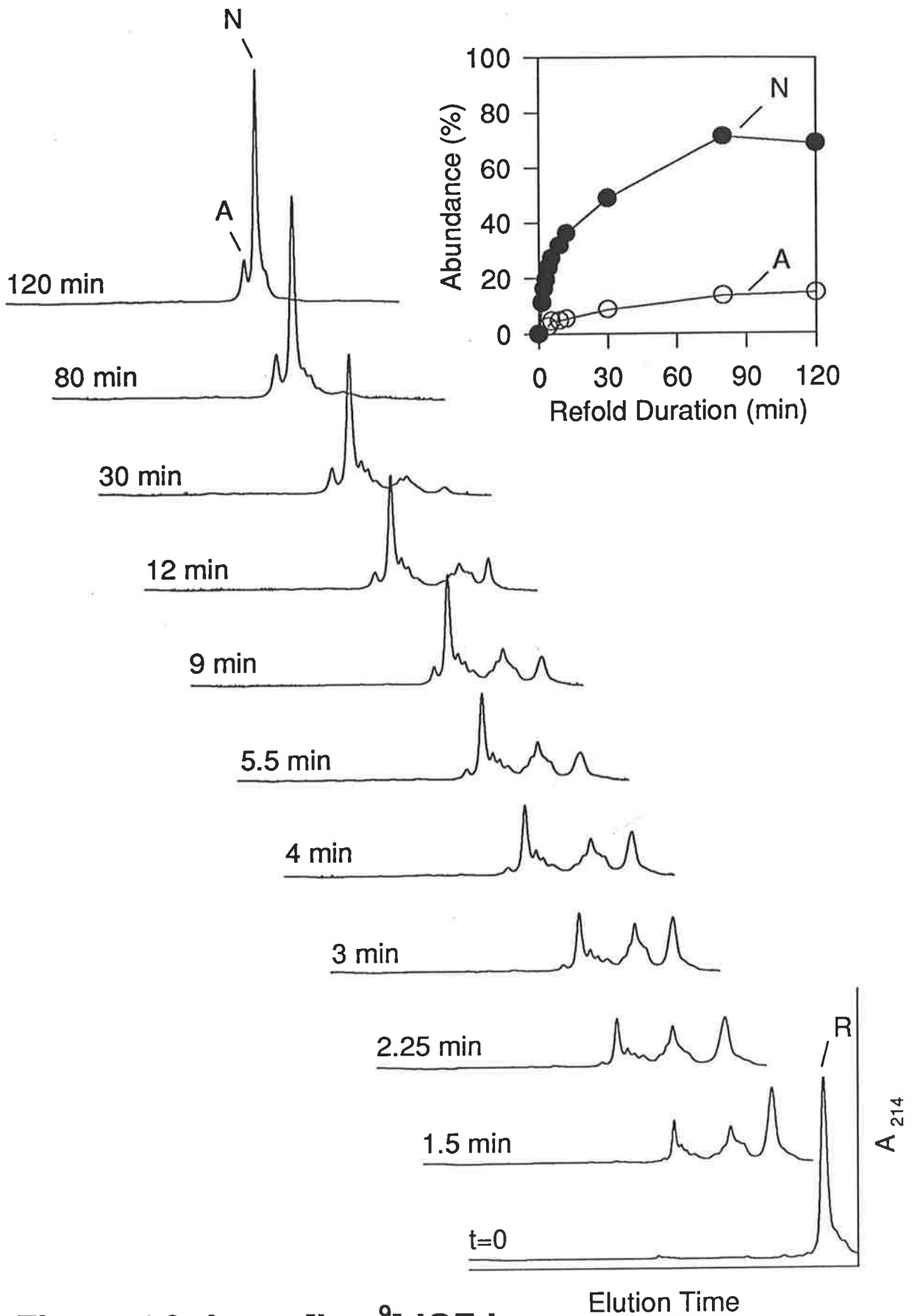
Figure 4.9: Long-[Lys^9]-IGF-I

well-populated intermediates. Furthermore, folding intermediates were present at equilibrium at only very low levels or not at all. A striking difference in the pathways of Long-IGF-I, Long-[Gly^3]IGF-I and Long-[Arg^3]IGF-I compared to those of IGF-I, [Gly^3]IGF-I and [Arg^3]IGF-I was the appearance of a well-populated intermediate that eluted between the alternative and the native isomers. For example, the folding reaction of Long-[Arg^3]IGF-I showed that the early products of folding were dominated by an intermediate that constituted up to 45 % of the total products at 4.25 minutes (Figure 4.8). The rapid appearance of this intermediate was followed by a decline and the emergence of the native and alternative folding isomers of Long-[Arg^3]IGF-I. An equivalent intermediate was not well-populated in the refolding of Long-[Lys^9]IGF-I. The equilibrium distribution of folding products for the Long-IGFs (Table 4.1) shows that the 13-amino acid N-terminal extension increased the abundance of the native isomer for Long-IGF-I, Long-[Arg^3]IGF-I and Long-[Lys^9]IGF-I but not for Long-[Gly^3]IGF-I.

A regression analysis of the *in vitro* oxidative folding reactions of IGF-I and Long-[Arg^3]IGF-I using a simplified kinetic scheme (below) gives very similar values for the rate constants k_1 to k_4. However, the value of k_5 for Long-[Arg^3]IGF-I is approximately one-quarter of that for IGF-I (results not shown).

An expansion of this scheme for Long-[Arg^3]IGF-I where the major intermediate (I) is separated from all other intermediates reveals that the formation of the intermediate (I) from this group of intermediates is very fast with a back reaction that is approximately one-half the rate. In this scheme, the rate limiting step in the formation of the native isomer from the reduced protein is between the intermediate (I) and the native isomer

(results not shown). *This analysis was kindly provided by Ben Hunt, Department of Chemical Engineering, University of Adelaide, Adelaide SA 5005.*

In each refolding reaction (Figures 4.2 to 4.9), the amount of purified protein that was fully reduced and refolded was recovered at equilibrium, demonstrating quantitative refolding. However, the microbore C_4 reverse-phase HPLC analyses of the samples taken at t=0, showed extremely variable analytical recoveries of the reduced protein. For the samples taken at 1.25 minutes, an estimate of the total protein recovered in the HPLC analysis demonstrated at least 80 % recovery of protein which increased to 100 % for the samples taken one or two minutes later (results not shown). These results indicate that the fully reduced protein may be readily lost during the sampling procedure, possibly due to aggregation, but there is no loss of protein in the refolding reaction. When the fully reduced IGFs were incubated in the refolding buffer but without the addition of 2-hydroxyethyl disulphide used to start the refolding reaction, only the fully reduced species was observed after 120 minutes (results not shown), indicating that the DTT concentration was sufficient to maintain reducing conditions and that oxidation by atmospheric oxygen did not contribute to the refolding reaction. The recovery of fully reduced protein in these analyses was also variable (results not shown).

4.3.2: The Effect of the 'Long' Peptide on the Refolding Reactions of IGF-I

IGF-I was refolded in the presence of increasing concentrations of the synthetic peptide MFPAMPLSSLFVNGPR-NH$_2$ (16mer) in an experiment designed to investigate the possibility that the 'Long' N-terminal part of the IGF fusion protein sequence could affect refolding in an analogous manner to the addition of the pro peptide in the folding of serine proteases (Section 1.2.2.7). The synthetic peptide was designed to copy the hydrophobic N-terminal sequence MFPAMPLSSLFVN, but

with three extra amino acids to be more similar to the peptide in *cis* and a C-terminal amide to aid solubility. In this experiment, the 16mer was added to the fully reduced IGF-I molecule prior to the oxidizing reagent used to initiate refolding. After refolding the mixtures were analysed by analytical microbore C_4 reverse-phase HPLC (Figure 4.10). Fully reduced IGF-I and the purified 16mer peptide were also analysed by microbore C_4 reverse-phase HPLC; the single peaks eluted respectively 6.5 minutes and 0.7 minutes after the native IGF-I molecule (results not shown). At low concentrations the 16mer was well resolved from the native IGF-I molecule, but at higher concentrations the added protein completely swamped the analysis (Figure 4.10d). For the HPLC analyses which were well resolved (Figure 4.10a, 4.10b and 4.10c) only the expected yields were recorded. Although the shape of the HPLC profiles changed slightly, the addition of the 16mer caused no difference in the yield of either the native or the alternative IGF-I isomers.

4.3.3: The Effect of Urea on the Refolding Reactions of IGF-I and Long-[Arg^3]IGF-I

To determine the effect of urea on the refolding reaction, IGF-I and Long-[Arg^3]IGF-I were refolded from the reduced form in the presence of increasing concentrations of urea. The yield of the native and alternative isomers of IGF-I or Long-[Arg^3]IGF-I at equilibrium is dependent on the urea concentration (Figure 4.11). Furthermore, the yield of folding intermediates present at equilibrium increases as the urea concentration increases. For example, the refolding reactions of IGF-I and Long-[Arg^3]IGF-I carried out in 8 M urea have analytical microbore C_4 reverse-phase HPLC profiles identical to those shown in Figures 4.12b and 4.12d respectively. For IGF-I, the yield of both the alternative and native isomers decrease by the same

Figure 4.10: Analysis of the refolding reaction of IGF-I in the presence of increasing concentrations of the peptide MFPAMPLSSLFVNGPR-NH₂.

IGF-I was refolded from the fully reduced form for 180 minutes and analysed by microbore C_4 reverse-phase HPLC employing a 20 % to 45 % acetonitrile gradient over 25 minutes in the presence of 0.1 % TFA. IGF-I was refolded: (a) in the absence of the peptide MFPAMPLSSLFVNGPR-NH₂ (16mer); (b) in the presence of the 16mer at the molar ratio of 1:0.1 (IGF-I:16mer); (c) at the molar ratio of 1:1; and (d) at the molar ratio of 1:10. The native (N) and the alternative (A) isomers and the peptide MFPAMPLSSLFVNGPR-NH₂ (16mer) are indicated.
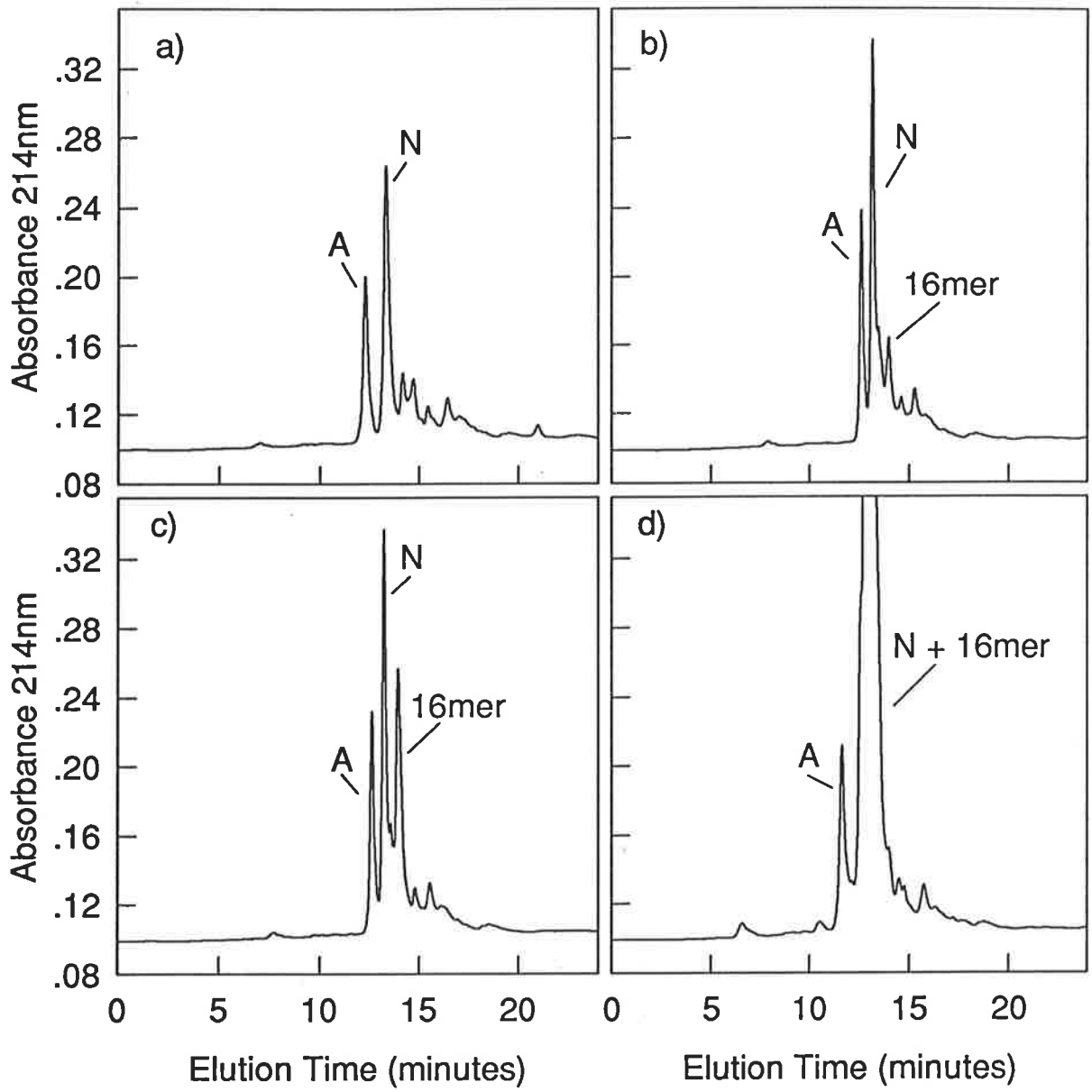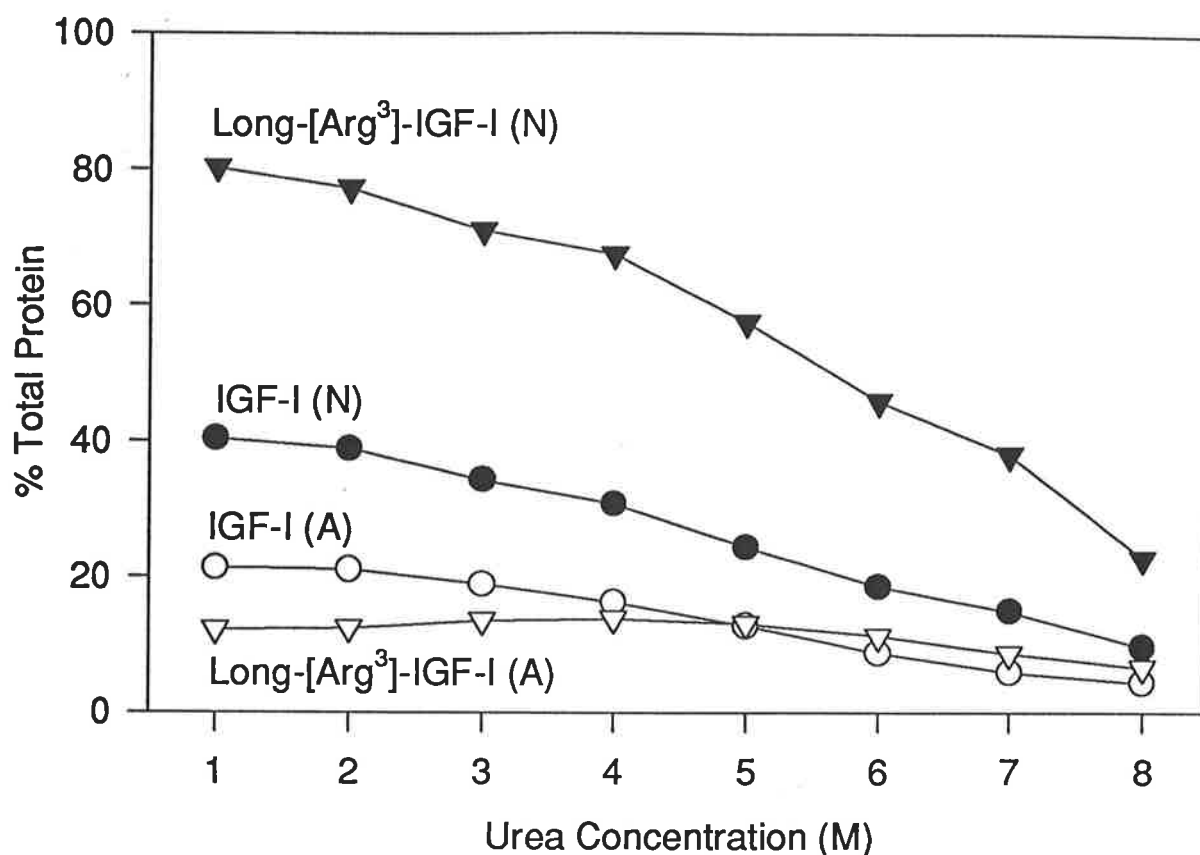
Figure 4.11: The effect of urea on the refolding reactions of IGF-I and Long-[Arg³]IGF-I

IGF-I (●,○) or Long-[Arg³]IGF-I (▼,▽) were refolded from the fully reduced peptide in the presence of increasing concentrations of urea. At equilibrium, the mixtures were analysed by microbore C_4 reverse-phase HPLC to determine the amount of native (N) or alternative (A) folding isomers present. The percentage of total protein present as either N or A, determined by calculating the area under the HPLC absorbance profile, was plotted against the urea concentration of the refolding buffer. The data plotted is the mean of two separate experiments.

Figure 4.12: The effect of urea on the refolding reactions of IGF-I and Long-[Arg^3]IGF-I

Samples of refolding mixtures were analysed by microbore C_4 reverse-phase HPLC using a 20 % to 45 % acetonitrile gradient over 25 minutes in the presence of 0.1 % TFA. (a) IGF-I after refolding at 8 M urea for 180 minutes followed by refolding at 2 M urea for a further 180 minutes; (b) IGF-I refolded at 8 M urea for 180 minutes followed by dilution and refolding at 8 M for a further 180 minutes; (c) Long-[Arg^3]IGF-I after refolding at 8 M urea for 180 minutes followed by refolding at 2 M urea for a further 180 minutes; (d) Long-[Arg^3]IGF-I refolded at 8 M urea for 180 minutes followed by dilution and refolding at 8 M for a further 180 minutes.
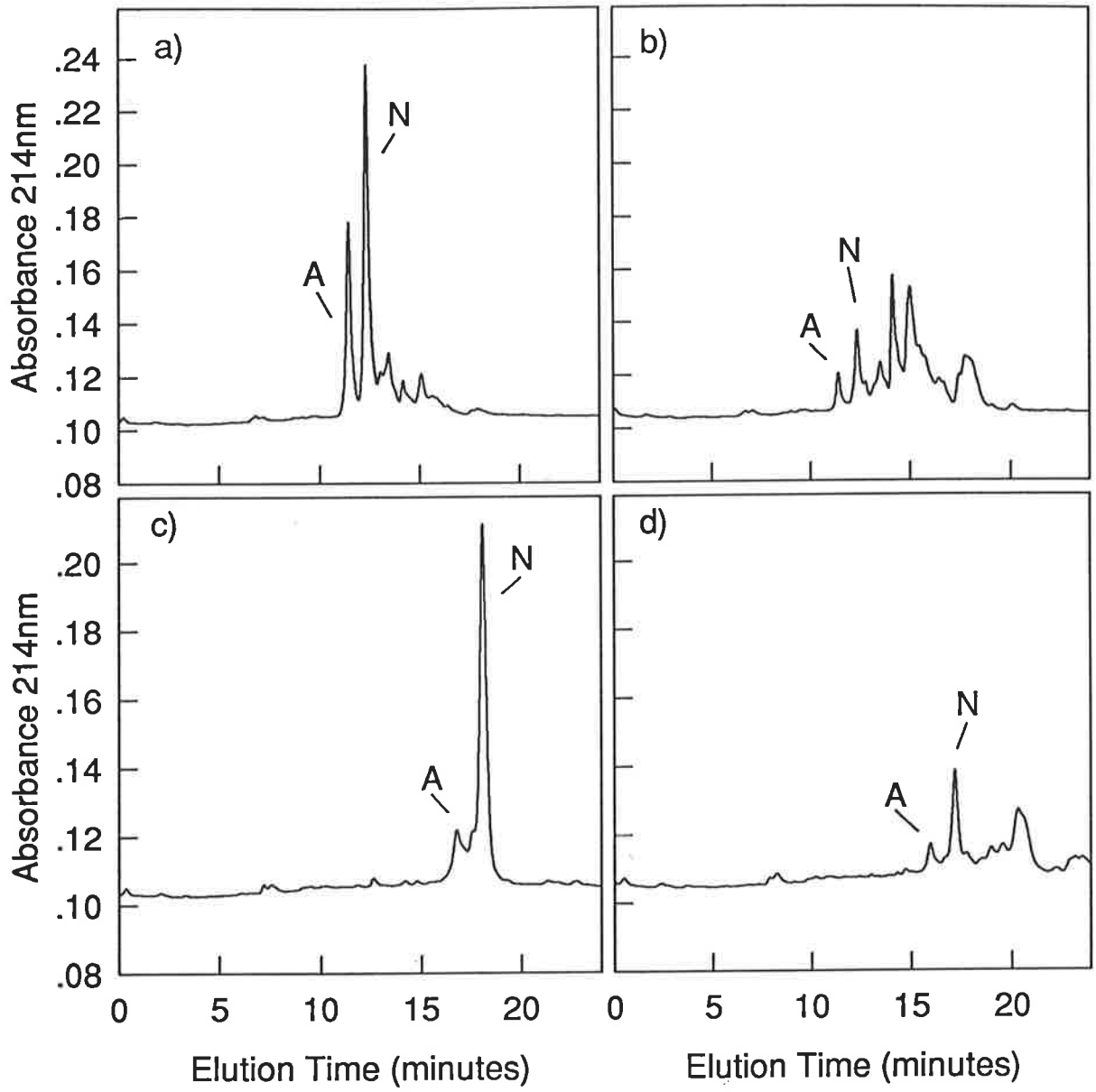
amount as the urea concentration increases, with the ratio between the two isomers remaining approximately constant (data not shown). However, for Long-[Arg^3]IGF-I, the yield of the native isomer decreases as the urea concentration increases, and the yield of the alternative isomer appears to increase slightly and then decrease slightly.

When either IGF-I or Long-[Arg^3]IGF-I was reduced and refolded in a 8 M urea refolding buffer and then diluted to 2 M urea refolding conditions, the equilibrium distribution of folding products is identical to that when refolded directly from the fully reduced form (Figures 4.12a and 4.12c). When the 8 M urea refolding mixtures were diluted by the same amount but with a 8 M urea refolding buffer, the equilibrium distribution was identical to that prior to dilution (Figures 4.12b and 4.12d). It is noteworthy that the major intermediate (I) present in the refolding of Long-[Arg^3]IGF-I from the fully reduced form at 2 M urea (Figure 4.8) is absent in the 8 M urea refolding reaction (Figure 4.12d). An observation that was made in three separate experiments.

4.3.4: The Effect of Redox Buffer on the Refolding Reaction of Long-[Arg^3]IGF-I

In a series of simple refolding reactions, purified, reduced and dried Long-[Arg^3]IGF-I was reacted in refolding buffers containing 2 M urea, 10 mM glycine, 0.1 M tris pH 8.7, 1 mM EDTA and different ratios of either DTT and 2-HED or GSH and GSSG, at 22°C for 180 minutes. The equilibrium distribution obtained for the refolding of Long-[Arg^3]IGF-I using DTT (0.4 mM) and 2-HED (1 mM) in the time course experiment (Figure 4.8 and Table 4.1) was reproduced in this experiment (Table 4.2). Furthermore, the same HPLC absorbance profile was obtained when the DTT and 2-HED concentrations were 0.4 mM and 0.8 mM respectively (Table 4.2).

Redox Buffer Constituents				% Abundance				
[DTT] mM	[2-HED] mM	[GSH] mM	[GSSG] mM	Native	Alt.	Int.	Others	Red.
0	0	0	0	36	12	42	8	2
0.4	0			0	0	0	3	97
0.4	0.1			0	0	0	4	96
0.4	0.2			0	0	5	15	80
0.4	0.4			71	9	20	0	0
0.4	0.8			81	10	9	0	0
0.4	1.0			79	11	10	0	0
0.4	10			0	0	0	0	0
0	1.0			0	0	0	0	0
		10	0	11	2	22	23	42
		10	1	48	18	25	5	4
		1	1	81	15	4	0	0
		1	10	72	26	2	0	0
		0	1	60	40	0	0	0

Table 4.2: The Effect of Redox Buffer Constituents on the Refolding of Long-[Arg^3]IGF-I

Long-[Arg^3]IGF-I was reacted in refolding buffers containing 2 M urea, 10 mM glycine, 0.1 M tris pH 8.7, 1 mM EDTA and different ratios of either DTT and 2-HED or GSH and GSSG, at 22°C for 180 minutes. The percentage abundance of the native and the alternative (alt.) isomers, the major intermediate (int.) and other intermediates, and the reduced (red.) peptide were calculated from the area under the absorbance peaks of the microbore C$_4$ reverse-phase HPLC analyses.

When the concentrations of DTT and 2-HED were both at 0.4 mM, refolding was nearly complete, but at lower concentrations of 2-HED the majority of protein was as the fully reduced form. When 2-HED was in large excess, no protein was recovered. The glutathione redox buffer also yielded the native and alternative isomers of Long-[Arg^3]IGF-I as judged by analytical microbore C$_4$ reverse-phase HPLC. When GSH and GSSG were both present at 1 mM the absorbance profile was almost identical to that obtained at equilibrium in the time course experiment (Figure 4.8 and Table 4.1). However, when the conditions were more oxidizing (1 mM GSH and 10 mM GSSG), the yield of the alternative isomer appeared to increase.

In the control refolding experiment, in the absence of any redox reagents, the HPLC analysis showed that oxidation by atmospheric oxygen had taken place. The reduced peptide had slowly refolded giving an absorbance profile similar to that obtained after 4.25 minutes refolding in the presence of 0.4 mM DTT and 1 mM 2-HED (Figure 4.8).

4.4: DISCUSSION

The oxidative folding of human IGF-I *in vitro* yields two major products of similar thermodynamic stability, the native isomer and an alternative isomer with a native-like Cys^{18} - Cys^{61} disulphide bond and two mismatched disulphide bonds (Cys^{48} - Cys^{52} and Cys^6 - Cys^{47}) (Miller *et al.*, 1993). Thus it is apparent that the amino acid sequence of IGF-I does not contain sufficient information to determine *uniquely* the native globular structure in accord with the classical Anfinsen concept (Anfinsen, 1973). The proportion of native and misfolded isomers present in the final equilibrium of IGF-I reported in this chapter is similar to that achieved when IGF-I is refolded by other techniques (Saito *et al.*, 1987; Iwai *et al.*, 1989; Hober *et al.*, 1992; Miller *et al.*, 1993). Furthermore, an

estimate of the free energy difference ($\Delta\Delta$G) between the two isomers is 0.37 kcal / mol, a value intermediate to the published values of 0.25 and 0.5 kcal / mol (Hober *et al.*, 1992; Miller *et al.*, 1993). In this study, the presence of the 13-amino acid N-terminal extension or the amino acid substitutions affects the refolding outcome (Table 4.1). The IGF analogue that refolds to the highest yield of the native isomer is Long-[Arg³]IGF-I (80 % of the total protein). For this analogue, an estimate of the free energy difference ($\Delta\Delta$G) between the native and alternative isomers present at equilibrium is 1.28 kcal / mol, a value much greater than that for IGF-I.

The folding of IGF-I is controlled by thermodynamic principles as the isolated native and alternative folding isomers will rearrange in the presence of redox reagents to yield the same ratio of folding products as when refolded from the fully reduced form (Miller *et al.*, 1993). In the absence of redox reagents the isolated isomers are stable indefinitely at physiological pH (Miller *et al.*, 1993; Joly and Swartz, 1994) thus representing alternative thermodynamic ground states that are effectively trapped by their disulphide bonds. The work presented in this chapter confirms that the folding of IGF-I is controlled by thermodynamic principles as the stable refolding products present at equilibrium in a refolding buffer containing 8 M urea, upon dilution to a lower urea concentration, will refold to yield the same ratio of folding products as when refolded from the fully reduced form (Figure 4.12). Furthermore, the refolding of Long-[Arg³]IGF-I is also thermodynamically controlled as it exhibits similar rearrangement properties to IGF-I. These rearrangement reactions of Long-[Arg³]IGF-I are investigated in more detail in Chapter 6. The effect of urea on the refolding of IGF-I or Long-[Arg³]IGF-I (Figures 4.11 and 4.12) demonstrates that urea stabilizes the folding intermediates that elute after the native isomer on C_4 reverse-phase HPLC, but not the major intermediate (I). Nevertheless, both the native and the alternative isomers are present in the refolding

reactions that occur at the highest urea concentrations. An important observation in the oxidative folding of IGF-I is that the synthetic 'Long' extension does not affect the relative abundance of the native and alternative isomers present at equilibrium. This suggests that any energy barriers in the free energy landscape (Section 1.2.2.7) of folding are unaltered in the presence of the 'Long' extension *in trans* and that all folding options are accessible thermodynamically providing that a redox buffer is present. This aspect of folding is explored in more detail in Chapter 7.

The choice of the redox buffer system used in the refolding of Long-[Arg^3]IGF-I had no effect on the folding outcome. In this analysis, the same equilibrium distribution of folding products was achieved when refolding in the presence of dithiothreitol (DTT) and 2-hydroxyethyl disulphide (2-HED) or reduced and oxidized glutathione (GSH and GSSG). For the glutathione redox buffer the highest yield of the native isomer was obtained when the K_{ox} was 1mM and for the DTT / 2-HED redox buffer the highest yield of the native isomer was obtained when the components were present at 0.4 mM and 1 mM respectively. The redox potential of this buffer system was not calculated as the equilibrium between the disulphide exchange reactions are complex. For example, prior to the addition of 2-HED (also known as oxidized ß-mercaptoethanol), the protein thiols are maintained in the reduced state by DTT. When the oxidizing reagent is added most of the DTT becomes oxidized to the cyclic disulphide (DTT^{ox}) and some of the 2-HED becomes reduced to two molecules of ß-mercaptoethanol (ßME). A complex equilibrium between DTT, DTT^{ox}, 2-HED and ßME is established and protein disulphide exchange reactions develop. Whatever the redox potential of this buffer system it is sufficiently oxidizing to permit disulphide exchange with the protein thiols. However, higher oxidation potentials may have a detrimental effect on folding as the refolding reaction with the highest level of 2-HED resulted in the loss of peptide.

Although the analysis reported in this chapter reveals only an overview of the folding process, it is clear that modifications in the B-domain (residues 3 - 29 of IGF-I) affect the folding outcome, a situation seen in the equilibrium distribution of folding products for the IGF-I analogues (Table 4.1). Moreover, the relative thermodynamic stability of the folding isomers is indicated by the composition of the equilibrium folding mixture (Hober *et al.*, 1992). For the charge substitutions at Glu^3 in IGF-I the native folding isomer is more stable than the alternative isomer and the other folding intermediates compared to IGF-I. A different situation is seen for the charge reversal at Glu^9 in helix 1 of the IGF-I molecule where the abundance of both the native and alternative isomers is decreased. Presumably this amino-acid substitution results in destabilisation of the helix structure by affecting the helix dipole (Matthews, 1993) or by affecting charge interactions with other regions of the IGF-I molecule. While the 13-amino acid extension does not appear to have any major effects on the relative stabilities of the alternative isomers, it appears to confer stability to the native isomers of Long-IGF-I, Long-[Arg^3]IGF-I and Long-[Lys^9]IGF-I resulting in an increased predominance of the native folding form at equilibrium. This effect is most striking for Long-[Lys^9]IGF-I where the N-terminal extension overcomes the disruptive effect of the charge reversal near the beginning of helix 1, nearly doubling the yield of the native isomer (Table 4.1).

The different proportions of native and alternative isomers in the Glu^3 substituted IGF-I analogues may perhaps be explained by structural observations based on the solution structure of native IGF-I (Cooke *et al.*, 1991) in which the amino acids Glu^3 and Arg^{56} are spatially close to one another (approximately 4.5Å) but too far away to form a salt bridge (Blundell *et al.*, 1978). However, it is conceivable that these two amino acids form a salt bridge in the alternative folded form of IGF-I, stabilising the folding isomer. If a stabilizing salt bridge exists in the alternative isomer of IGF-I, then one would expect that a charge substitution of Glu^3 by a neutral Gly would remove the salt bridge, thus reducing

its stability and hence its abundance. Moreover, substitution by Arg would cause charge repulsion, further reducing the stability. Similarly the presence of the hydrophobic 13-amino acid extension may prevent formation of the proposed salt bridge as a result of steric hindrance.

Thus the artificial 13-amino acid amino-terminal extension facilitates correct folding of IGF-I and IGF-I analogues, probably by imparting a steric constraint in the refolding reaction. This constraint may prevent the contact of regions, such as Glu^3 and Arg^{56} in IGF-I that would otherwise draw the folding pathway into stable non-native disulphide arrangements at the time the second disulphide bond is formed. As indicated above, the amino-terminal extension also exerts an influence over other disruptive forces as in the [Lys^9]IGF-I analogue. This study confirms that the amino acid sequence of IGF-I contains insufficient information to determine uniquely the native globular structure. However, it is possible that when part of the B-domain is prevented from contact with the rest of the molecule at a crucial point in folding, native disulphide bonds are formed in high yields. This concept is reinforced by the observation that the C-terminal E-peptide extension of pro-IGF-I does not affect the folding outcome in a similar fashion (Hober *et al.*, 1994). While undefined accessory proteins may be required to assist the folding of IGFs *in vivo*, it is also possible that the N-terminal signal sequence of pre-pro-IGF-I guides the folding of the emerging nascent polypeptide chain in the endoplasmic reticulum. Hence, the hydrophobic signal pre-sequence may function in an analogous way to the hydrophobic artificial N-terminal extension investigated in this chapter.

CHAPTER 5

STRUCTURAL ANALYSIS OF DISULPHIDE ISOFORMS OF IGF-I AND IGF-I ANALOGUES

5.1: INTRODUCTION

The *in vitro* oxidative folding reactions of IGF-I and the IGF-I analogues presented in Chapters 3 and 4 have relied upon the microbore C_4 reverse-phase HPLC elution profiles to identify the native and alternative isomers present at equilibrium. In these experiments the most abundant product of folding was assumed to be the native isomer and the second most abundant isomer, which was always less hydrophobic, was assumed to be the alternative isomer with the disulphide arrangement of Cys^{18} - Cys^{61}, Cys^{48} - Cys^{52}, Cys^{6} - Cys^{47}. When the refolding products of each of the IGF analogues were screened for biological activity, the greatest biological activity was always associated with the most abundant isomer. Other evidence was provided by the observation that hydroxylamine could be used to cleave a biologically active IGF fusion protein to yield a product that has a biological activity identical to the most abundant refolding product of the native length molecule. Furthermore, the biological activity of the recombinant IGF-I molecule was identical to that published for IGF-I isolated from serum or colostrum. Thus the assumption that the most abundant product of folding has the native disulphide arrangement of Cys^{18} - Cys^{61}, Cys^{47} - Cys^{52}, Cys^{6} - Cys^{48} was well founded. However, the assumptions that were used to identify the alternative folding form (hydrophobicity and abundance) may be incorrect as the effect of the amino acid changes and misfolding is unpredictable. The alternative isomer of IGF-I is known to have a reduced potency in a number of biological assays (Hodgkinson *et al.*, 1989; Forsberg *et al.*, 1990; Axelsson *et al.*, 1992; Miller *et al.*, 1993) and this property can be used to probe the structure of the putative alternative isomers of the IGF analogues. Thus, if the alternative isomers of each of the IGF-I analogues exhibit a

parallel reduction in potency compared to IGF-I then the structures may be considered equivalent.

In this chapter, I will investigate the disulphide conformations of the native and the alternative isomers of IGF-I analogues obtained in Chapters 3 and 4. I will then determine their potency in a number of biological assays. Finally, I will build a model of the structure of misfolded IGF-I and compare it to the biological data and that of preliminary two-dimensional total correlation spectroscopy (TOCSY) nuclear magnetic resonance data.

5.2: METHODS

The materials and methods are described in detail in Chapter 2. However, for clarity the methods employed in this chapter are summarized here.

5.2.1: Purification of the Folding Isomers

The purified proteins IGF-I, [Gly^3]IGF-I, [Arg^3]IGF-I, [Lys^9]IGF-I, Long-IGF-I, Long-[Gly^3]IGF-I, Long-[Arg^3]IGF-I and Long-[Lys^9]IGF-I were fully reduced and refolded as described in Section 4.2.1.1. After 180 minutes refolding the reaction was stopped by acidification and the folding products were isolated by one or two rounds of preparative reverse-phase HPLC using techniques described in Sections 2.2.2.3 and 2.2.2.4. To assess their purity, each native or alternative isomer was analysed by microbore C$_4$ reverse-phase HPLC using a 25 % to 40 % (v/v) acetonitrile gradient over 150 minutes in the presence of 0.1 % TFA. The purified isomers were analysed for quantity as described in Section 2.2.6.1, distributed as 6 µg or 12 µg lots and then lyophilized and stored at -20°C. To ensure that the

quantitation was accurate, an aliquot of each lyophilized isomer was recovered and re-analysed in duplicate on the same day using a single dedicated HPLC system.

5.2.2: Determination of the Disulphide Bond Linkages

Each purified folding isomer was digested with pepsin and analysed by electrospray mass spectrometry (Sections 2.2.3.3 and 2.2.3.4). To identify the peptide fragments, the observed mass was compared with the known amino acid sequence and with the published pepsin digestion patterns of IGF-I (Forsberg *et al.*, 1990). As the native and alternative isomers have different structures (Hober *et al.*, 1992; Miller *et al.*, 1993), different pepsin cleavage sites are exposed (Forsberg *et al.*, 1990), thereby facilitating the identification of the disulphide architecture. IGF-I purified from bovine colostrum (Section 2.1.3.3) was also digested and its disulphide architecture confirmed.

5.2.3: *In Vitro* Biological Assays

The native and alternative folding isomers of IGF-I and each of the IGF-I analogues (16 different peptides) were analysed in triplicate at seven concentrations for binding to IGF binding proteins (IGFBPs) secreted by rat L6 myoblasts, for binding to the rat type 1 IGF receptor and for the stimulation of protein synthesis in rat L6 myoblasts (Sections 2.2.6.2 to 2.2.6.4). To assess stability of Long-[Arg^3]IGF-I in the receptor binding assay, the alternative isomer of Long-[Arg^3]IGF-I was incubated in the presence of rat L6 myoblasts under the same assay conditions as for binding to the type 1 IGF receptor. After incubation, the peptide was removed and

assayed for the presence of the native isomer by employing a reverse-phase HPLC separation and detection using a receptor binding assay (Section 2.2.6.5).

5.2.4: Preliminary NMR Studies of Native and Alternative Isomers of IGF-I

Preliminary NMR studies on the native and alternative isomers of IGF-I were carried out by Dr. John Carver (Department of Chemistry, University of Wollongong, Wollongong, NSW, Australia) using two-dimensional TOCSY. For this purpose, 5 mg of each of the IGF-I isomers were prepared as described in section 5.2.1.1 and analysed by NMR as described in section 2.2.3.5.

5.2.5: Construction of a Model of Alternative IGF-I

A model of alternative IGF-I was created from the average NMR structure of native IGF-I in several steps (Brookhaven code 2GF1: Cooke *et al.*, 1991). The NMR structure of IGF-I was imported into Biosym software employing *InsightII, Builder* and *Discover* programs (Version 95.0: Biosym Technologies, San Diego, CA 92121), and missing hydrogen atoms were added and valences were corrected. The Cys^6 - Cys^{48} and Cys^{47} - Cys^{52} disulphide bonds were removed and rebuilt with Cys^6 - Cys^{47} and Cys^{48} - Cys^{52} disulphide bonds. The structure was then minimized using 500 iterations of steepest descent and 500 iterations of conjugate gradient minimization. Finally, the resultant structure was analysed for quality using the program *PROCHECK* (Laskowski et al., 1993).

5.3: RESULTS

5.3.1: Determination of Disulphide Linkages

The two most abundant products of folding present at equilibrium in the oxidative folding reactions of IGF-I and each of the IGF-I analogues were purified to homogeneity by reverse-phase HPLC. The most abundant isomer is the putative native isomer (N) with the greatest biological activity and the second isomer which is more hydrophilic is the putative alternative isomer (A). Each purified isomer was assessed by analytical microbore C_4 reverse-phase HPLC, as demonstrated in Figure 5.1 for IGF-I and Long-[Arg^3]IGF-I. The identity of each peptide was confirmed by N-terminal sequence analysis and mass spectrometry (results not shown).

To confirm that the putative native and alternative isomers contained the expected disulphide bond conformations, the proteins were digested by pepsin and the fragments produced were analysed by mass spectrometry. Pepsin was used to digest the IGF isomers for two reasons: (i) its effectiveness at low pH allows for digestion in conditions where potential disulphide exchange is minimized, and (ii) pepsin cleavage sites that are accessible in one folded form may be less accessible in the other folded form and exhibit different rates of cleavage. Thus the two disulphide isomers are expected to exhibit different cleavage patterns as a result of altered topology. In this experiment, the digestion of IGF-I with pepsin differed slightly from that published by Forsberg et al. (1990) in that cleavage between Lys^{27} and Pro^{28} was observed and that cleavage between Asp^{45} and Glu^{46} was absent. Nevertheless, the different cleavage patterns generated from the cleavage of the native and the alternative isomers of IGF-I was reproducible. In the present work, the
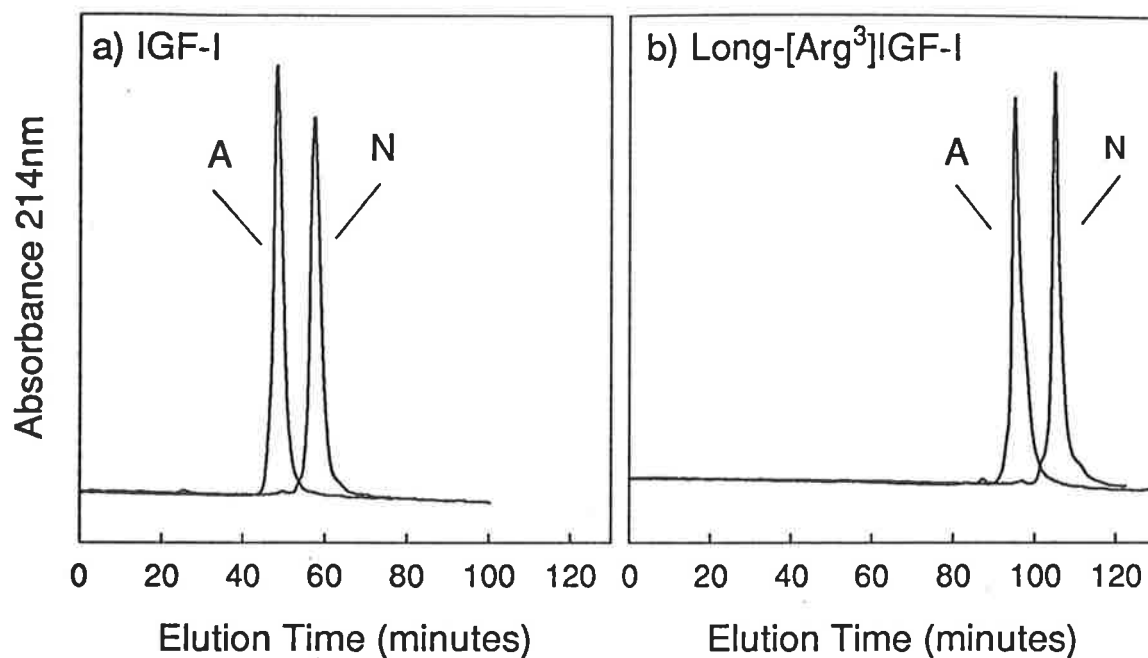
Figure 5.1: Purity of IGF-I and Long-[Arg³]IGF-I Folding Isomers

The purified alternative (A) and native (N) folding isomers of IGF-I and Long-[Arg³]IGF-I were assessed by microbore C_4 reverse-phase HPLC. Elution of protein was monitored as absorbance at 214nm, performed on a gradient of 25-40 % (v/v) acetonitrile over 150 minutes in the presence of 0.1 % (v/v) TFA. The chromatograms of the separate injections are overlapped to indicate the relative retention times of the alternative and native isomers for each analogue.

putative alternative isomer was digested at a slower rate than the putative native isomer (results not shown). The peptide fragments produced for each IGF isomer are summarized in Table 5.1. The mass spectrometry data is shown for Long-[Arg³]IGF-I (Figure 5.2) and a diagram of the deduced fragments is shown in Figure 5.3. The nomenclature used for the labelling of the peptide fragments is the same as that used by Forsberg *et al.* (1990). Each IGF isomer shows the presence of the CG fragment containing residues 16→24 and 60→70 linked by a disulphide bond between Cys^{18} and Cys^{61} (Table 5.1 and Figure 5.3). This demonstrates that the two major folding isomers present at equilibrium are linked by a common native-like disulphide bond. For IGF-I and each of the IGF-I analogues, the putative native isomer has a pepsin cleavage pattern different from that of the putative alternative isomer. The major difference in the cleavage patterns is that an AEF fragment is generated for the putative native isomer and that an ADE fragment is generated for the putative alternative isomer (Table 5.1 and Figures 5.2 and 5.3). The appearance of the different fragments can be accounted for by the relative rates of cleavage of the pepsin sensitive Ile^{43} - Val^{44} and Asp^{53} - Leu^{54} peptide bonds in the parent ADEF fragment (Figure 5.3). Thus for the putative native isomer fold, the Asp^{53} - Leu^{54} bond appears less accessible to the protease and cleavage occurs at the Ile^{43} - Val^{44} bond. The reverse appears to be true for the putative alternative isomer fold where the Ile^{43} - Val^{44} bond appears to be less accessible and cleavage occurs at the Asp^{53} - Leu^{54} bond. Except for trace levels in [Lys⁹]IGF-I, the AEF fragment was not detected in the putative alternative isomers (Table 5.1). Some cleavage occurred at the Asp^{53} - Leu^{54} bond in the putative native isomers of IGF-I and Long-IGF-I, resulting in the presence of low levels of ADE and AE fragments after digestion. However, ADE fragments were not observed after digestion of the other putative

Table 5.1: Peptide Fragments Detected after Cleavage with Pepsin

For each IGF, the native (N) and alternative (A) folding isomers present at equilibrium in the oxidative folding reaction were isolated, digested with pepsin and analysed by mass spectrometry. The resultant mass spectrometry profiles were used to identify the fragments generated by cleavage by pepsin. The identified fragments contain the following residues: Long, -13→ -4; A, -3→10 or 1→10; B, 11→15; C, 16→24; D, 25→42; D1, 28→43; E, 43→53; F, 54→59; G, 60→70.

	Long	D1	CG	AEF	ADE	ADEF	Intact	Other
Bovine IGF-I	n/a	+	+	+	-	-	-	AE+
IGF-I (N)	n/a	+	+	+	+[1]	-	-	D+
IGF-I (A)	n/a	+	+	-	+	+	+	D-
[Gly^3]IGF-I (N)	n/a	+	+	+	-	-	-	
[Gly^3]IGF-I (A)	n/a	+	+	-	+	-	+	
[Arg^3]IGF-I (N)	n/a	+	+	+	-	+[2]	-	B+
[Arg^3]IGF-I (A)	n/a	+	+	-	+	-	-	B+
[Lys^9]IGF-I (N)	n/a	+	+	+	-	+	+	
[Lys^9]IGF-I (A)	n/a	+	+	+[1]	+	+	+	
Long-IGF-I (N)	+	+	+	+	+[1]	+	-	D+
Long-IGF-I (A)	+	+	+	-	+	+	-	D-
Long-[Gly^3]IGF-I (N)	+	+	+	+	-	-	-	
Long-[Gly^3]IGF-I (A)	+	+	+	-	+	+	-	
Long-[Arg^3]IGF-I (N)	+	+	+	+	-	-	-	
Long-[Arg^3]IGF-I (A)	+	+	+	-	+	-	-	
Long-[Lys^9]IGF-I (N)	+	+	+	+	-	+	-	
Long-[Lys^9]IGF-I (A)	+	+	+	-	+	+	-	

n/a   Not applicable
1   Low levels detected
2   Contains modified peptide with +16 mass units

Figure 5.2: Mass Spectrometry Analysis of the Pepsin Digestion of Folding Isomers of Long-[Arg^3]IGF-I.

The two most abundant folding isomers of Long-[Arg^3]IGF-I were isolated, digested with pepsin and analysed by electrospray mass spectrometry. The mass of peptides generated by pepsin are shown for the most abundant folding isomer (a) and the alternative isomer (b). The identified fragments contain the following residues: Long, -13→ -4; A, -3→10 or 1→10; B, 11→15; C, 16→24; D, 25→42; D1, 28→43; E, 43→53; F, 54→59; G, 60→70. The asterix denotes fragments that result from alternative cleavage sites.
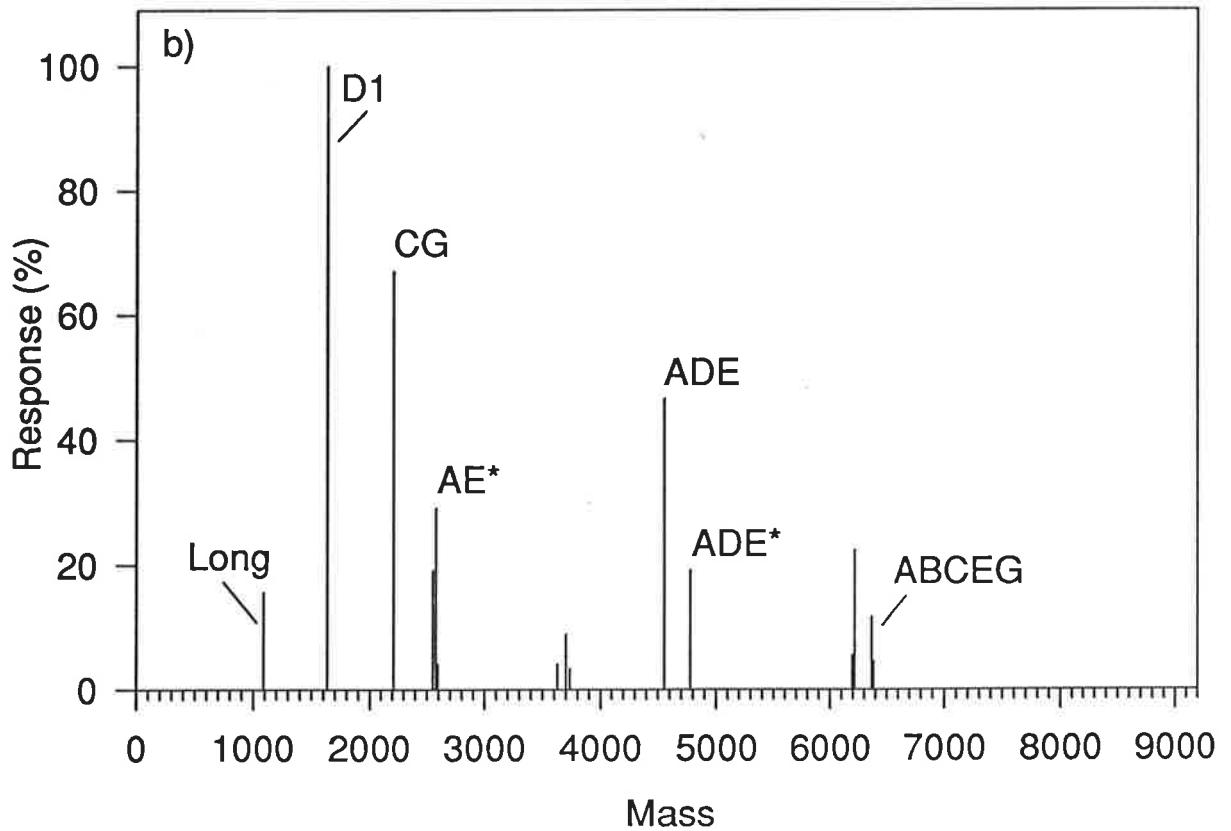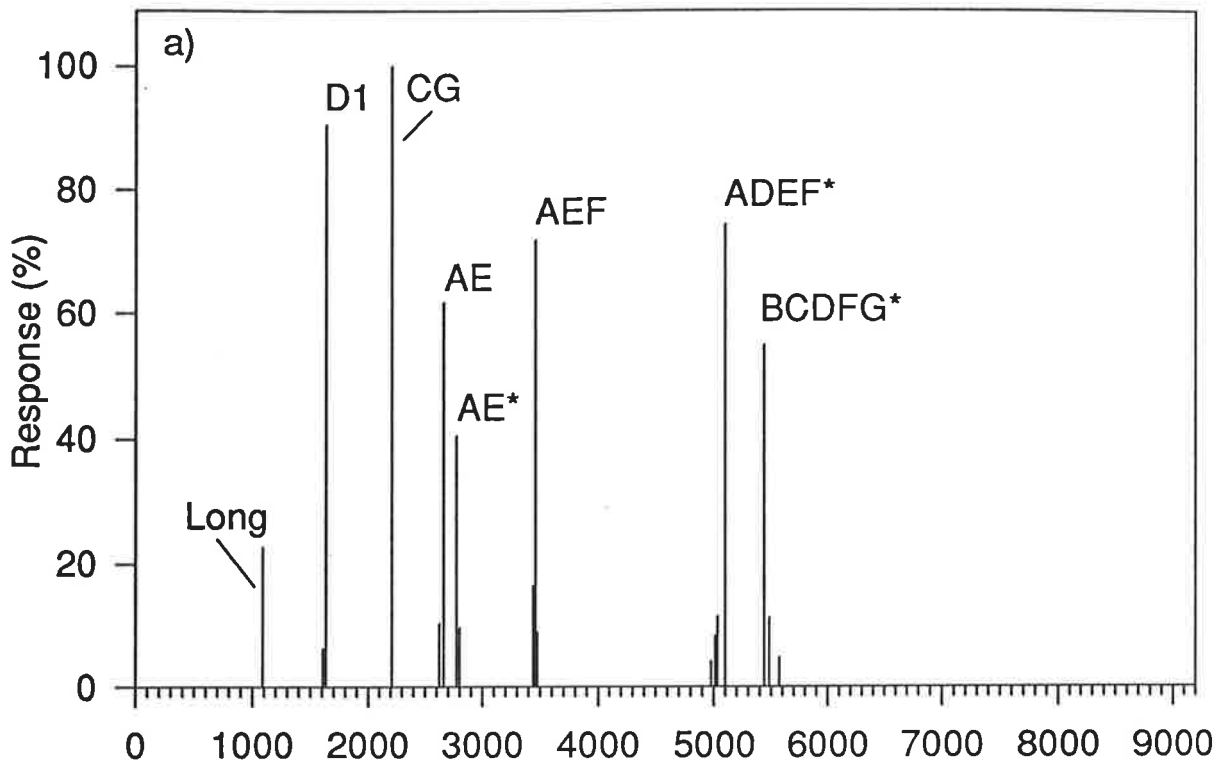
Figure 5.3: The Deduced Pepsin Digestion Patterns of Native and Alternative Isomers of Long-[Arg^3]IGF-I.

The major fragments produced following the digestion of the native and alternative folding isomers Long-[Arg^3]IGF-I by pepsin are shown using the standard single letter amino acid code. The cleavage of the native isomer yields the A-E-F fragment and the alternative isomer yields the A-D-E fragment. Arrows represent further cleavage sites in the A-D-E fragment which in turn yields an A-E fragment and a D-fragment containing residues 28-43.

Native

L S S L P M A P F M -13

G N V F

1 G

P 43 I

R V

T D

L E

6 C 47 C C 52

G C F R S

A D R

E L R L

A(1-10) L E

E(44-53) M

F(54-59)

G T Q P A R R S S S G Y G T P D(28-42)

70 A

S K N F

K

A

P

G(60-70) K

61 L

Y C A P Y

F V C G D R G F

18 C(16-24)

V D A L Q

B(11-15)

Alternative

L S S L P M A P F M -13

G N V F

1 G

P I

R V

T D

L E

6 C 47 C C 52 D

G C F R S

A D

E L R L

A(1-10) R L

E(43-53) E M

F(54-59)

I G T Q P A R R S S S G Y G T P D(25-42)

70 A K

S N F

K

A

P

G(60-70) K

61 L

Y C A P Y

F V C G D R G F

18 C(16-24)

V D A L Q

B(11-15)
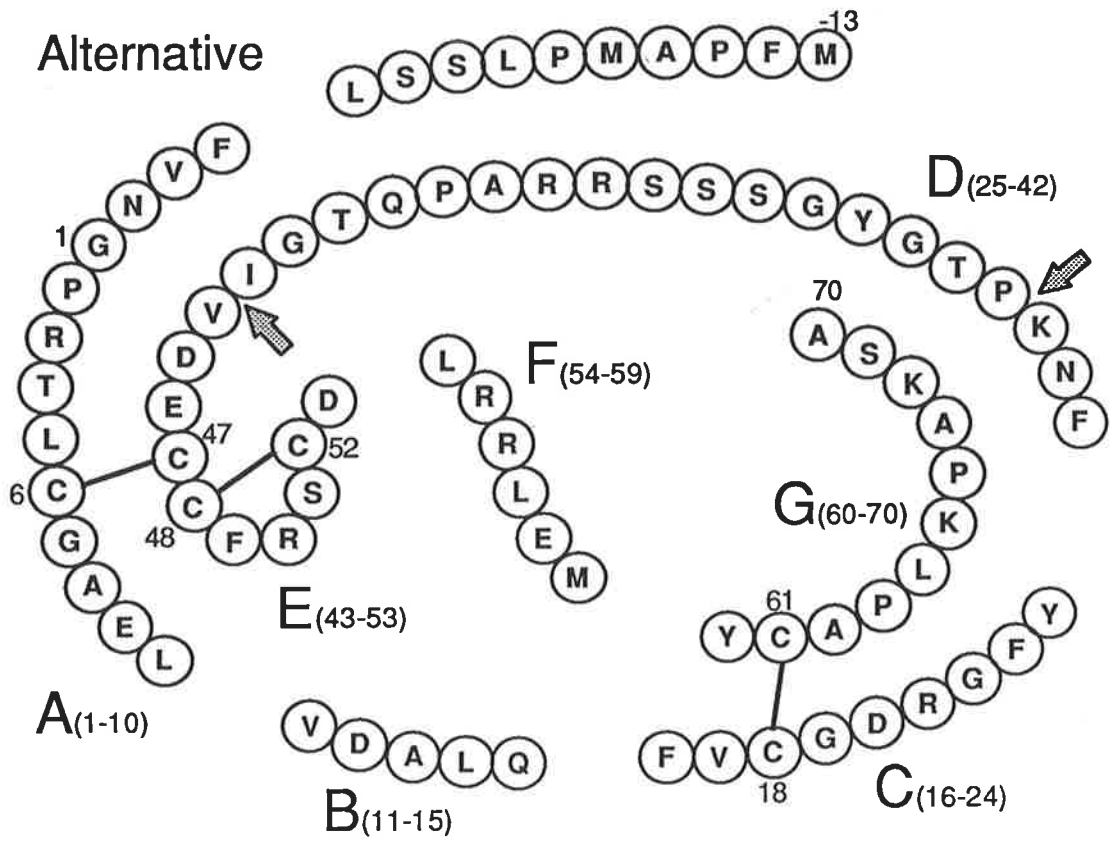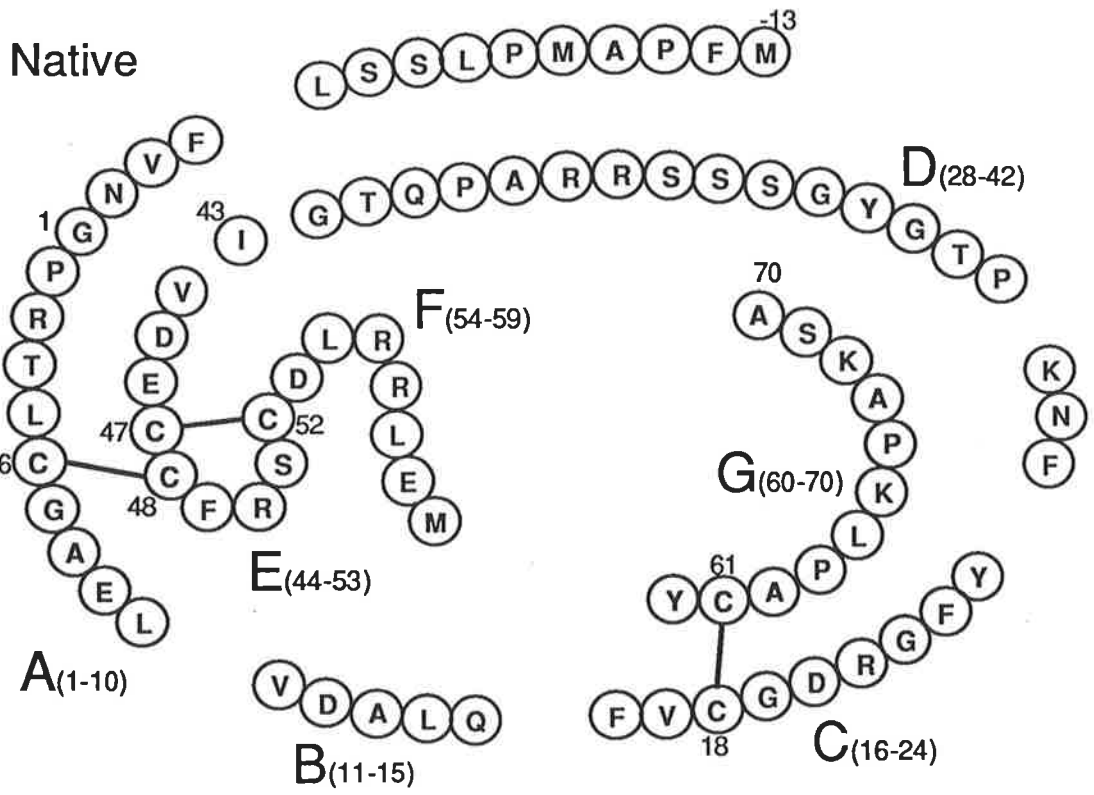
native IGFs. In summary, the putative native and alternative isomers of each IGF showed the expected pepsin cleavage patterns of IGF molecules with $Cys^{18} - Cys^{61}$, $Cys^{47} - Cys^{52}$, $Cys^6 - Cys^{48}$ and $Cys^{18} - Cys^{61}$, $Cys^{48} - Cys^{52}$, $Cys^6 - Cys^{47}$ bonds respectively.

5.3.2: Interactions of IGF Isomers with Rat L6 Myoblasts

To further examine the effect of changes in topology as a result of disulphide bond isomerization, the biological function of the alternative isomers was compared with the native isomers in rat L6 myoblasts. The binding of folding isomers of IGF-I and analogues to IGFBPs secreted by rat L6 myoblasts was measured in a competition binding assay (Figure 5.4 and Table 5.2). For the native folding isomer of the normal length IGFs, the order of potency was: IGF-I > $[Lys^9]$IGF-I > $[Gly^3]$IGF-I > $[Arg^3]$IGF-I. The alternative folding isomers of these analogues showed a similar order of potency with a reduction in binding of at least 200-fold (Table 5.2). The presence of the 13-amino acid N-terminal extension had the effect of reducing binding to IGFBPs for the native folding isomers of the fusion protein analogues. The alternative folding isomer of Long-$[Arg^3]$IGF-I had an apparent affinity for L6 conditioned medium IGFBPs similar to the native folding form. However, the alternative folding isomers of the other fusion protein analogues exhibited lower binding potencies compared with their native isomer forms (Table 5.2).

Binding of the folding isomers of the IGF-I analogues to the type 1 IGF receptor on L6 myoblasts was measured in a competition binding assay. Each of the alternative folding isomers competed with ^{125}I-IGF-I for binding to the type 1 IGF receptor less well than the corresponding native folding isomer (Figure 5.5 and Table 5.2). For the native folding isomer of the normal length IGFs, the order of potency was:

Figure 5.4: Binding of IGF Folding Isomers to IGFBPs

Competition for binding of ^{125}I-labelled IGF-I to IGFBPs in media conditioned by rat L6 myoblasts by native (filled symbols) and alternative (open symbols) folding isomers of (a) IGF-I ($\bullet$, $\circ$), [Gly^3]IGF-I ($\blacktriangle$, $\triangle$), [Arg^3]IGF-I ($\blacktriangledown$, $\triangledown$) and [Lys^9]IGF-I ($\blacksquare$, $\square$) and (b) Long-IGF-I ($\bullet$, $\circ$), Long-[Gly^3]IGF-I ($\blacktriangle$, $\triangle$), Long-[Arg^3]IGF-I ($\blacktriangledown$, $\triangledown$) and Long-[Lys^9]IGF-I ($\blacksquare$, $\square$). Each folding isomer was tested in triplicate at each peptide concentration. The standard errors of the estimate of the mean value (S.E.M.) are indicated by descending bars where they are larger than the symbols. B/B_o is the percentage of ^{125}I-labelled-IGF-I bound in the absence of added unlabelled peptide.
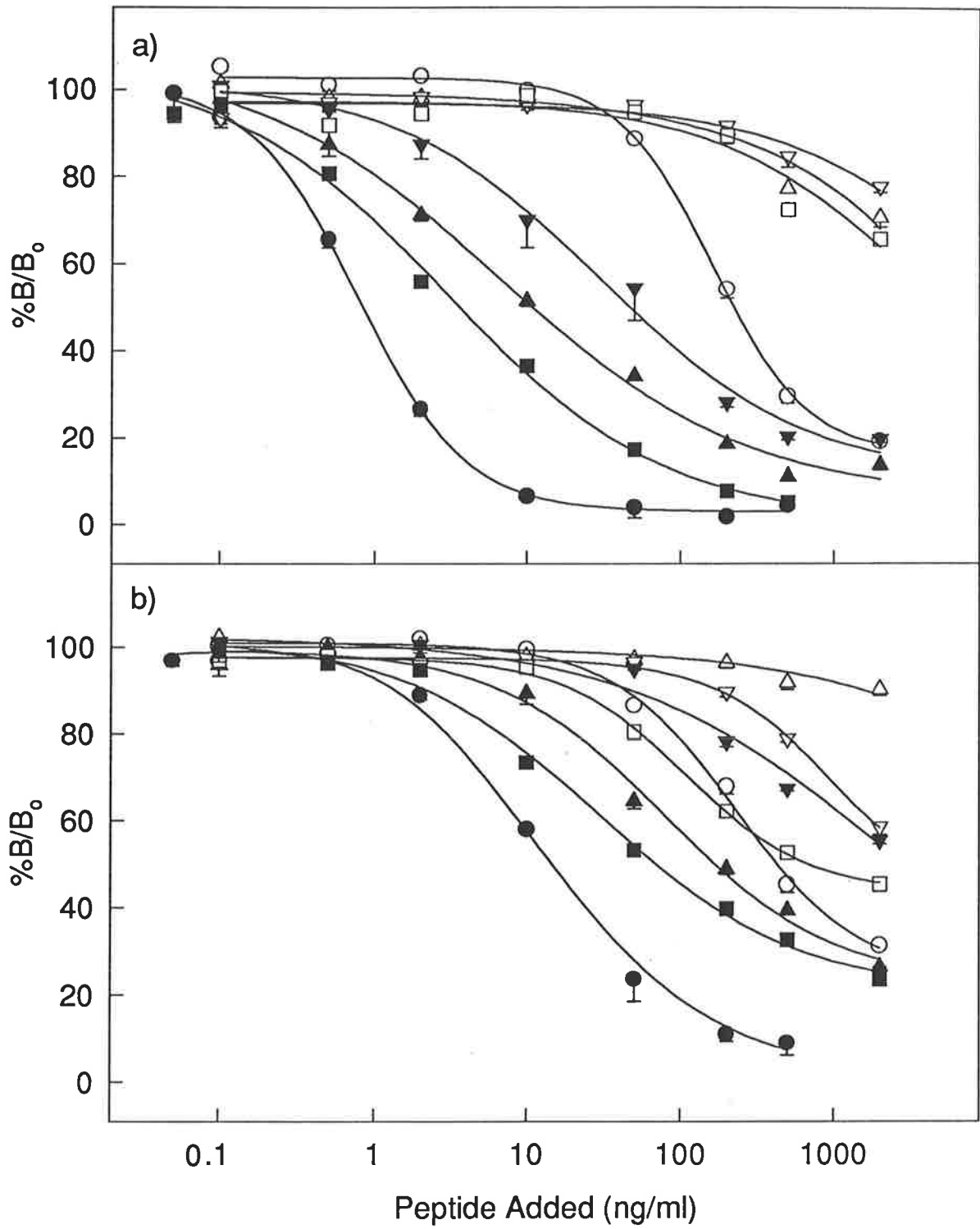
	Binding to IGFBPs (ED_{20})		Binding to the Type 1 IGF receptor (ED_{50})		Stimulation of Protein Synthesis (ED_{50})	
	Nat.	Alt.	Nat.	Alt.	Nat.	Alt.
IGF-I	0.3	79	6.8	1200	14	550
[Gly^3]IGF-I	1.0	430	5.3	610	2.5	100
[Arg^3]IGF-I	5.1	1050	5.4	610	1.9	80
[Lys^9]IGF-I	0.5	340	39	>2000	24	1600
Long-IGF-I	4.2	93	9.8	300	14	250
Long-[Gly^3]IGF-I	21	>2000	10	890	3.5	160
Long-[Arg^3]IGF-I	190	460	17	65	4.2	13
Long-[Lys^9]IGF-I	6.8	56	31	350	15	110

Table 5.2: Biological Properties of IGF Isomers in Rat L6 Myoblasts

The binding of IGF isomers to rat L6 myoblast IGFBPs in conditioned medium and to the type 1 IGF receptor were expressed respectively as ED_{20} and ED_{50} values (ng / ml). The stimulation of protein synthesis was expressed as ED_{50} (ng / ml).

Figure 5.5: Binding of IGF Folding Isomers to the Type 1 IGF Receptor

Competition for binding of ^{125}I-labelled IGF-I by native (filled symbols) and alternative (open symbols) folding isomers of (a) IGF-I (●, O), [Gly^3]IGF-I (▲, Δ), [Arg^3]IGF-I (▼, ∇) and [Lys^9]IGF-I (■, □) and (b) Long-IGF-I (●, O), Long-[Gly^3]IGF-I (▲, Δ), Long-[Arg^3]IGF-I (▼, ∇) and Long-[Lys^9]IGF-I (■, □). Each folding isomer was tested in triplicate at each peptide concentration. The standard errors of the estimate of the mean value (S.E.M.) are indicated by descending bars where they are larger than the symbols. B/B_o is the percentage of ^{125}I-labelled-IGF-I bound in the absence of added unlabelled peptide.
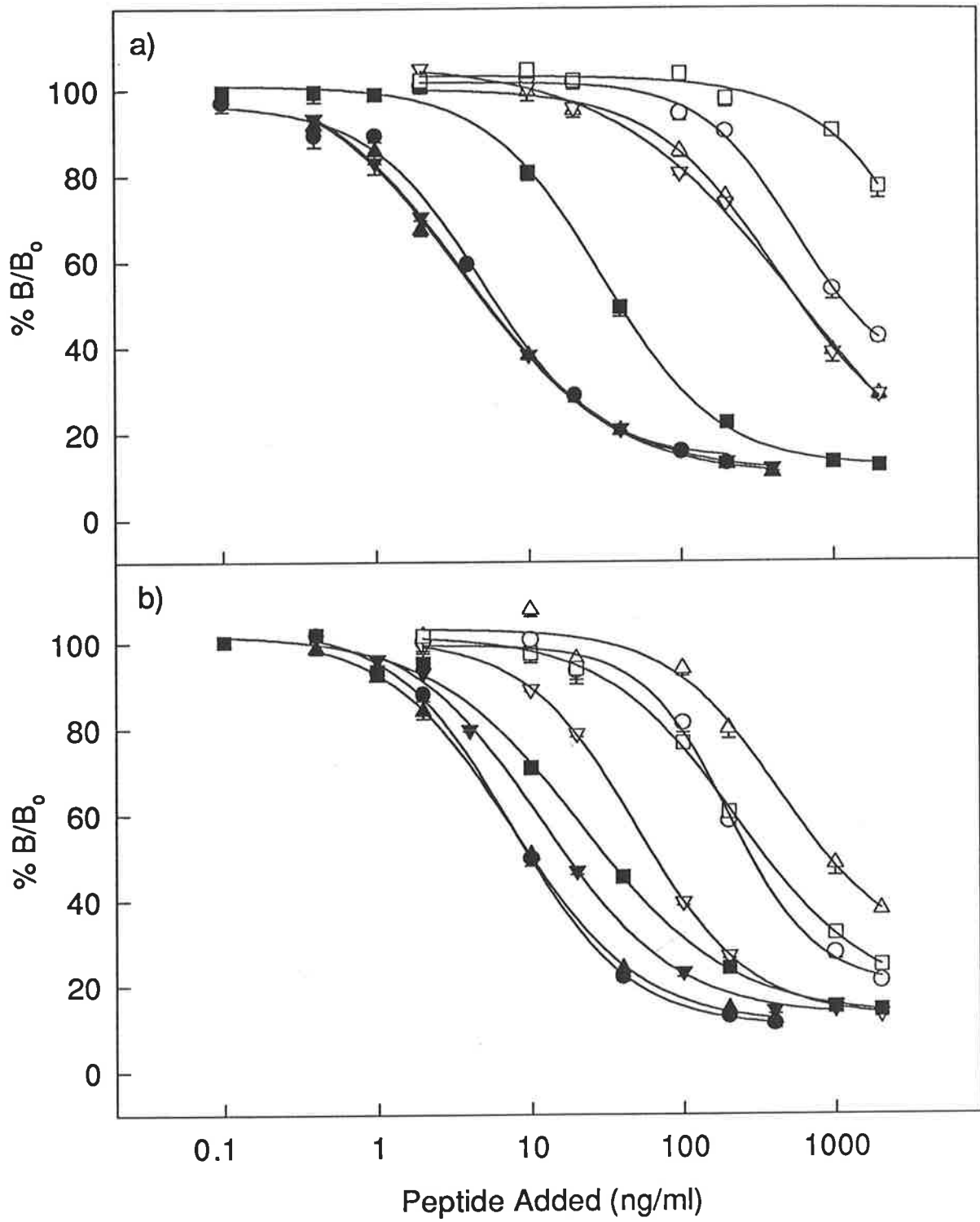
[Gly^3]IGF-I ≥ [Arg^3]IGF-I > IGF-I > [Lys^9]IGF-I. This order of potency was similar for the alternative folding isomers but with a reduction in binding of at least 100-fold (Table 5.2). Except for Long-[Lys^9]IGF-I, the presence of the N-terminal extension reduced the binding of the native folding isomers to the type 1 IGF receptor. The alternative folding isomer of Long-[Arg^3]IGF-I had one quarter the potency as the native folding isomer, whereas the alternative folding isomers of Long-IGF-I, Long-[Gly^3]IGF-I and Long-[Lys^9]IGF-I had only 3, 1 and 9 percent of the binding potency compared to the respective native folding isomers.

The effect of altered IGFBP and type 1 IGF receptor binding on the growth factor activity of the various native and alternative folding isomers was measured in a protein synthesis assay (Figure 5.6). In L6 myoblasts all folding isomers stimulated protein synthesis substantially, and with the exception of the alternative folding isomers of IGF-I and [Lys^9]IGF-I achieved responses nearly equivalent to those observed with 10 % (v/v) foetal bovine serum. Higher concentrations of all alternative folding isomers were required to stimulate protein synthesis compared with the respective native folding isomer (Figure 5.6). For the native folding isomer of the normal length IGFs, the order of potency was [Arg^3]IGF-I > [Gly^3]IGF-I > IGF-I > [Lys^9]IGF-I. The order of potency for the alternative isomers was the same but had potencies only 1-2 % of the native isomers (Figure 5.6a). The order of potency of the native isomers of the fusion protein analogues was: Long-[Gly^3]IGF-I > Long-[Arg^3]IGF-I > Long-IGF-I > Long-[Lys^9]IGF-I. The alternative folding forms of these Long isomers exhibited a wider range of potency differences than observed with the normal length IGFs. In particular, the alternative folding isomer of Long-[Arg^3]IGF-I was only moderately less potent than the native isomer and was as potent as the native folding isomers of Long-IGF-I and Long-[Lys^9]IGF-I (Figure 5.6b).

Figure 5.6: Stimulation of Protein Synthesis in Rat L6 Myoblasts by Folding Isomers of IGF-I Analogues

Stimulation of protein synthesis by native (filled symbols) and alternative (open symbols) folding isomers of (a) IGF-I (●, O), [Gly^3]IGF-I (▲, Δ), [Arg^3]IGF-I (▼, ∇) and [Lys^9]IGF-I (■, □) and (b) Long-IGF-I (●, O), Long-[Gly^3]IGF-I (▲, Δ), Long-[Arg^3]IGF-I (▼, ∇) and Long-[Lys^9]IGF-I (■, □). Each folding isomer was tested in triplicate at each peptide concentration. The standard errors of the estimate of the mean value (S.E.M.) are indicated by descending bars where they are larger than the symbols.
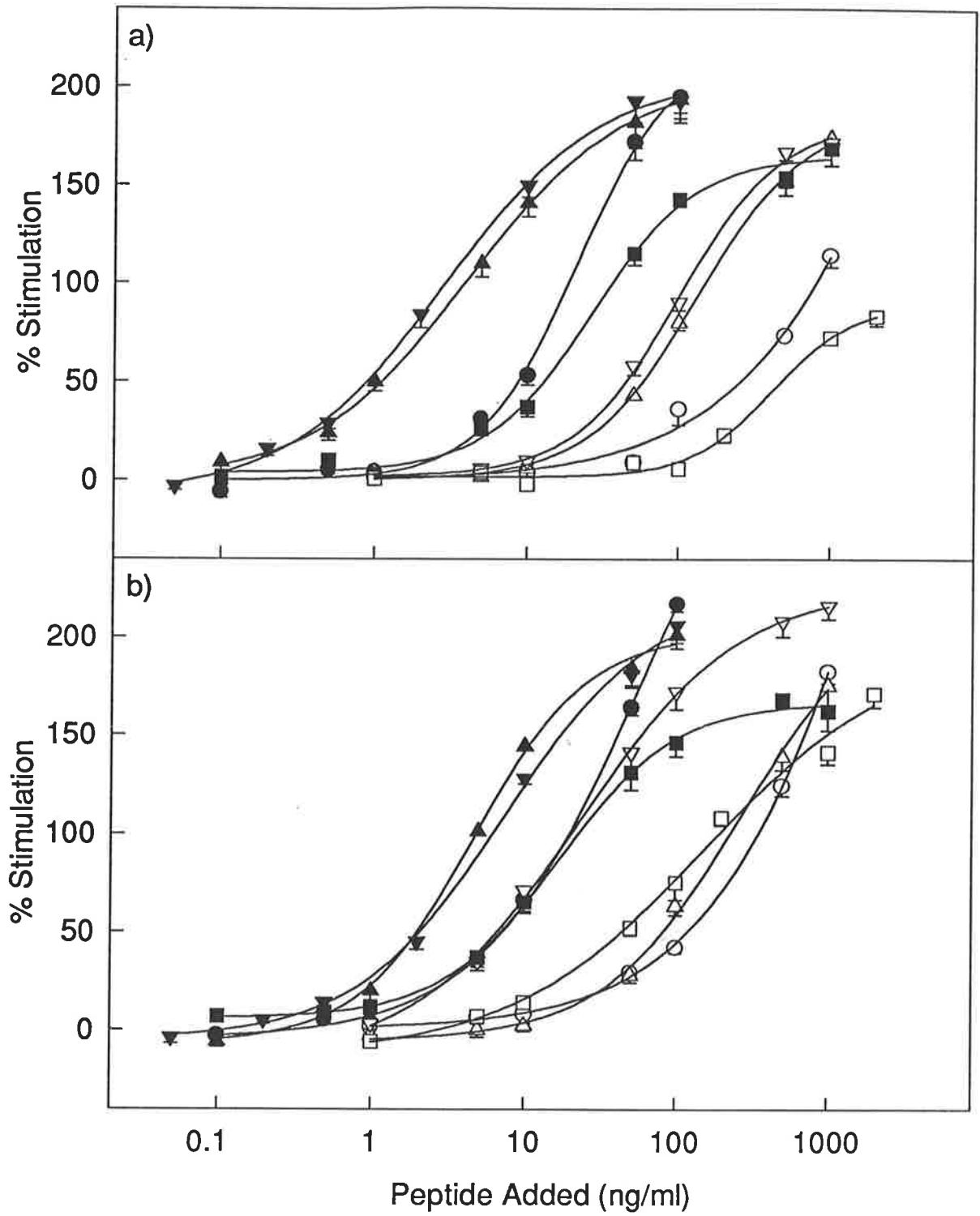
As the potency of alternative folding isomer Long-[Arg^3]IGF-I is only slightly less than the native isomer, it is possible that the alternative folding isomer rearranges to the native isomer during the assay. However, this is unlikely since no native Long-[Arg^3]IGF-I could be detected after incubation of the pure alternative folding isomer under the conditions of the receptor binding assay (Figure 5.7). This observation is also relevant for the IGFBP-binding assay where the lower pH would result in an even slower rate of disulphide rearrangement.

5.3.3: ^{1}H-NMR Spectroscopy of the Alternative Isomer of IGF-I

From the differences observed in the one-dimensional NMR spectra of the alternative and native isomers of IGF-I there are some structural changes caused by disulphide misfolding (Figure 5.8). In particular there appears to be more dispersion in the alternative isomer in the region downfield of the residual water peak at ~ 4.7 ppm which is indicative of α-CH protons in ß-sheet or turn-type regions. An expansion of the two-dimensional TOCSY spectrum for these peaks suggests that they may arise from residues such as Ala^8, Cys^{47}, Cys^{48}, Glu^{46} and Phe^{49} (Figure 5.9a). However, one cannot categorically assign these residues without further NMR spectroscopy and sequential assignment. Expansions of other regions of the TOCSY spectra indicate that there have been changes in the structure on misfolding, many of which may be quite small (results not shown). An expansion of the cysteine α to ß region for both isomers (Figures 5.9b and 5.9c) indicate that there have been significant changes which may be expected as a result of switching the disulphide bonds. The six Cys peaks of the native isomer (Figure 5.9b) identified using the published data of Sato *et al.* (1992) were present, but for the alternative isomer (Figure 5.9c) some Cys peaks were absent from this region. Although the Cys peaks
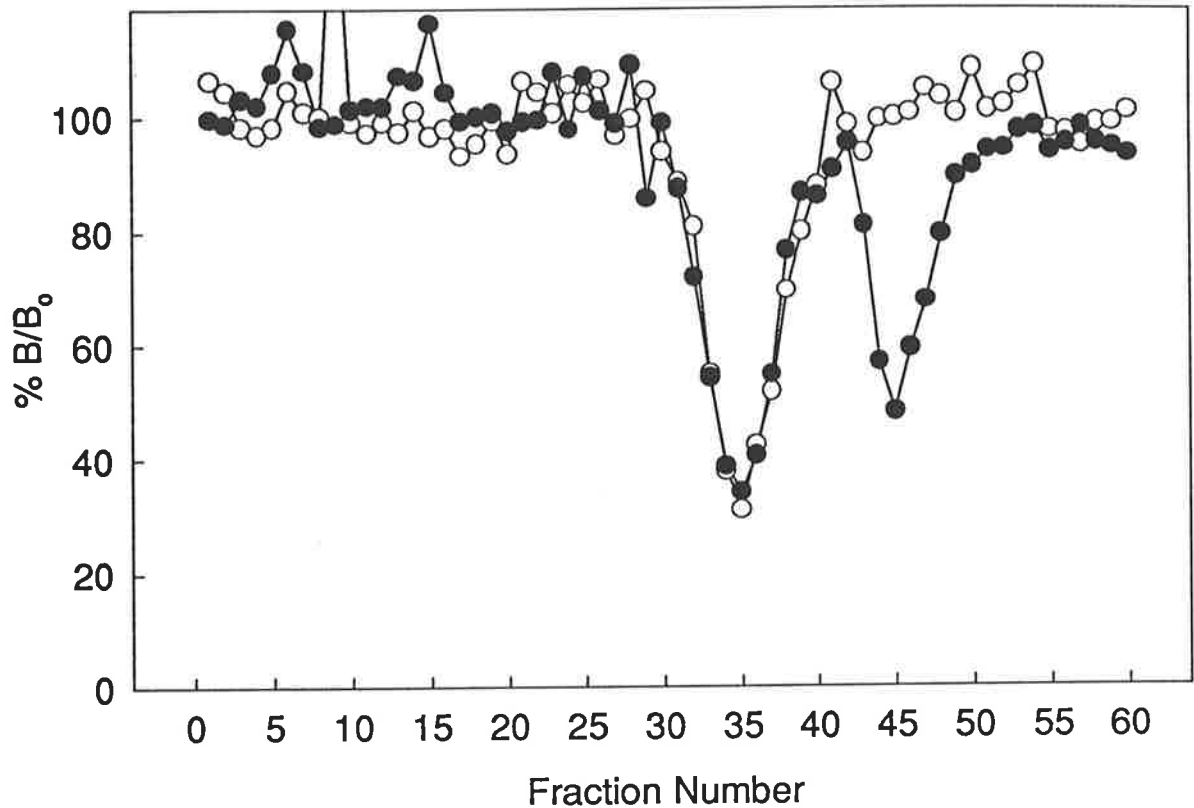
Figure 5.7: The Stability of the Alternative Isomer of Long-[Arg³]IGF-I in the Type 1 IGF Receptor Binding Assay

The alternative folding isomer of Long-[Arg³]IGF-I was incubated for 18 hours in the presence of rat L6 myoblasts under the conditions used for the receptor binding assay. The mixture was then separated by C_4 reverse-phase HPLC using a 28-40 % (v/v) acetonitrile gradient over 60 minutes in the presence of 0.1 % (v/v) TFA. Fractions were analysed for receptor binding activity (O). A second gradient separation was carried out with the native isomer included as an internal standard (●).

Figure 5.8: One-Dimensional ^{1}H-NMR Spectra of the Native and Alternative Isomers of IGF-I

One-dimensional NMR spectra of (a) the native isomer and (b) the alternative isomer of IGF-I were acquired at 40°C in 10 % CD_3COOD / 90 % H_2O at pH ~ 3. The region downfield of the residual water peak at ~ 4.7 ppm indicative of α-CH protons in ß-sheet or turn type regions is denoted by an asterix.

a)
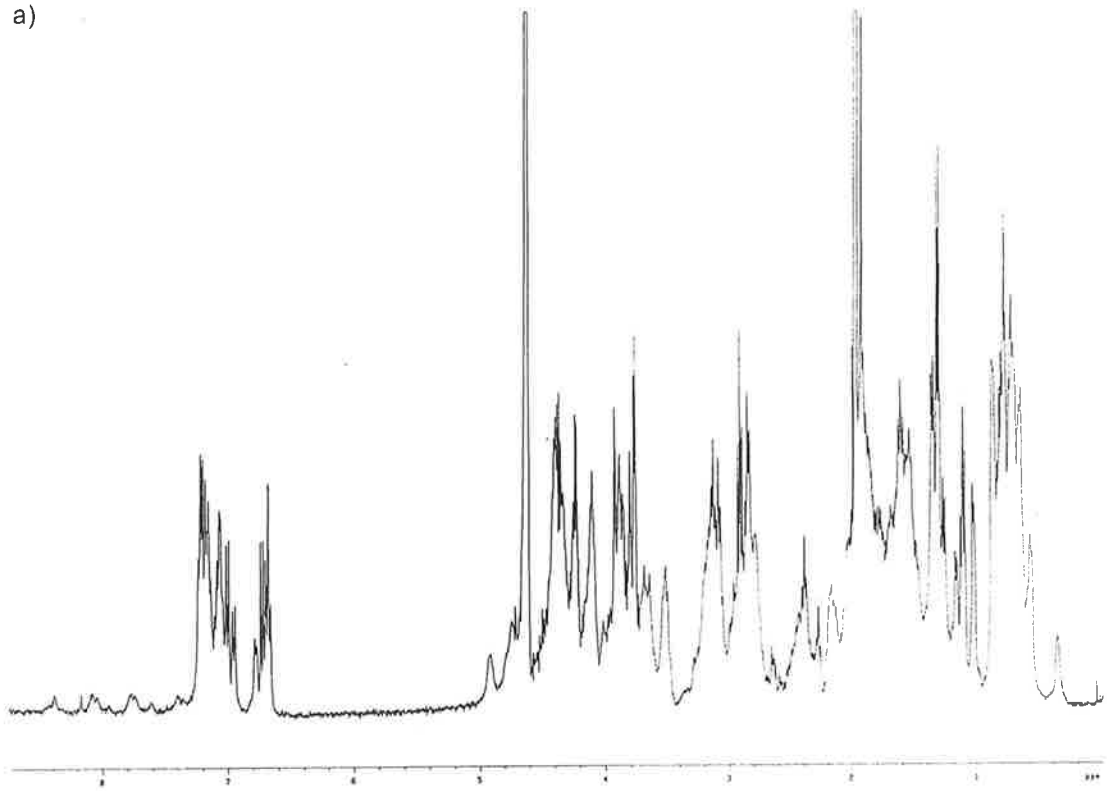
b)

*
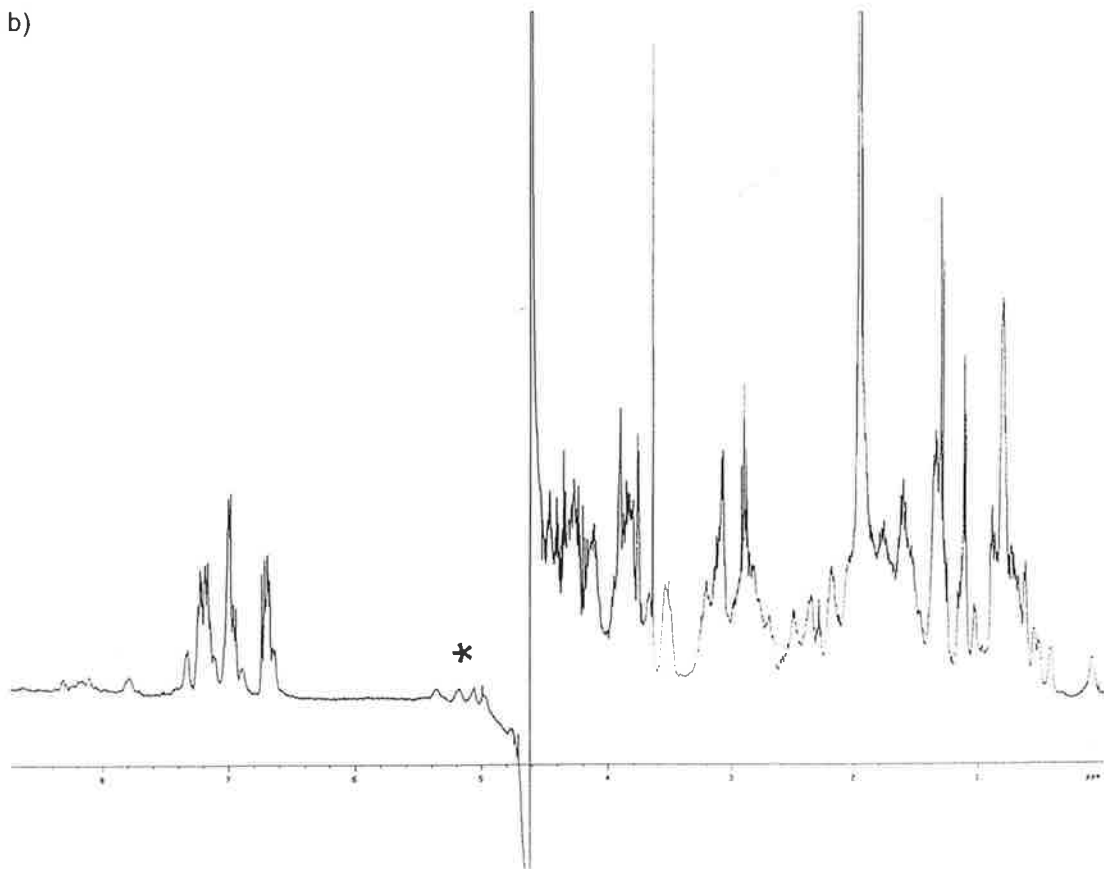
Figure 5.9: Two-Dimensional TOCSY ^{1}H-NMR Spectra of the Native and Alternative Isomers of IGF-I

Two-dimensional TOCSY NMR spectra of the native isomer and the alternative isomers of IGF-I were acquired at 40ºC in 10 % CD_3COOD / 90 % H_2O at pH ~ 3. The TOCSY spectrum downfield of the residual water peak (figure 5.8b) is shown for the alternative isomer (a) and the spectra of the cysteine α to ß regions are shown for both the native (b) and alternative (c) isomers. The residues indicated in panel (a) are tentative and need to be confirmed by further NMR spectroscopy and sequential assignment. The numbers in panel (b) refer to the six Cys peaks of the native isomer (Figure 5.9b) identified using the published data of Sato *et al.* (1992).

a)

Ala8

Glu46

Cys   Cys   Phe49

F1 (ppm)

1.5 2.0 2.5 3.0 3.5

F2 (ppm)

5.4 5.2 5.0

b)

61

52   47   18

6

48

F1 (ppm)

3.0

F2 (ppm)

5.0 4.8 4.6 4.4

c)

F1 (ppm)

3.0

F2 (ppm)

5.0 4.8 4.6 4.4
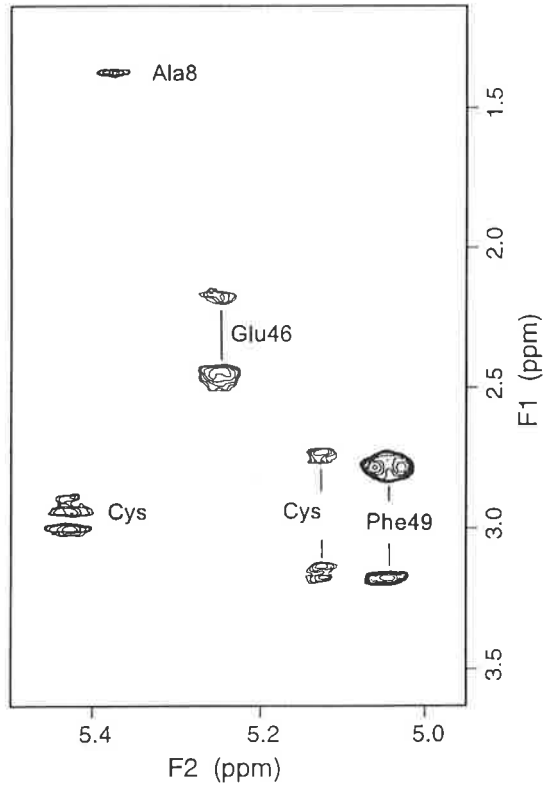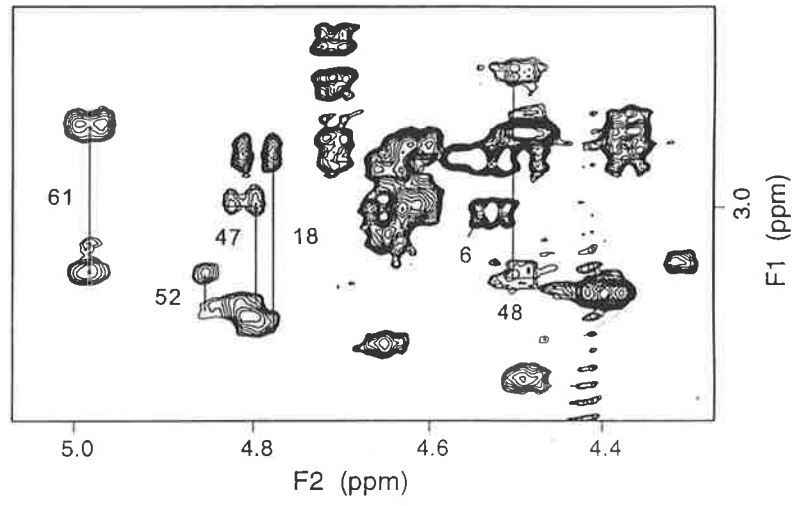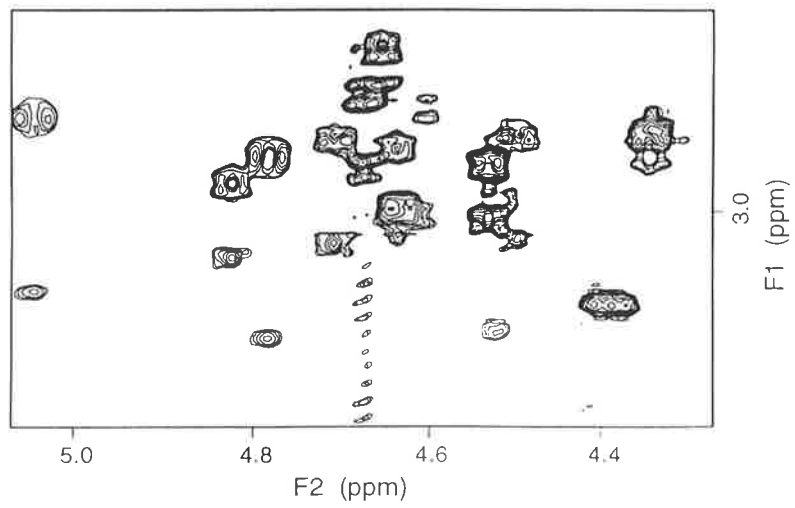
in the alternative isomer could not be identified with certainty, it has peaks that may correspond to Cys^{18}, Cys^{61} and Cys^{6} in the native molecule (Figure 5.9c). The presence of similar peaks for Cys^{18} and Cys^{61} in the alternative isomer may be expected as this molecule contains an identical Cys^{18} - Cys^{61} disulphide bond and that there are few structural changes in this region of the molecule (Miller *et al.*, 1993).

5.3.4: Construction of a Model of the Alternative Isomer of IGF-I

The model of alternative IGF-I was created from the average NMR structure of native IGF-I (Cooke *et al.*, 1991). The NMR structure of IGF-I was imported into Biosym software and missing hydrogen atoms were added and valences were corrected. The Cys^{6} - Cys^{48} and Cys^{47} - Cys^{52} disulphide bonds were removed and rebuilt with the alternative disulphide bond arrangement of Cys^{6} - Cys^{47} and Cys^{48} - Cys^{52}. Finally, the whole structure was energy minimized. The quality of the resultant structure was then analysed using the program *PROCHECK* (Laskowski et al., 1993). The Ramachandran plot of the model structure showed 32 residues (56.1%) to be in most favoured regions and 22 residues (38.6%) to be in additional allowed regions. This compares well to the original native structure which has 38 residues (66.7%) in the most favoured regions and 15 residues (26.3%) in additional allowed regions. Molecular dynamics experiments were also carried out on the model of alternative IGF-I, but in each experiment the quality of the resultant structures were much worse than the original alternative IGF-I model. For this thesis, the minimized structure presented above is used for discussion.

A close examination of the superimposed native IGF-I structure and the alternative model structure (Figures 5.10a and 5.10b) shows there to be only modest

Figure 5.10: A Model of Alternative IGF-I Superimposed Over That of Native IGF-I

A model of alternative IGF-I (red) based upon the NMR structure of native IGF-I (Cooke et al.,1991; green) was constructed by energy minimization as described in Section 5.3.3.2. Panel (a) shows a ribbon-diagram of alternative IGF-I superimposed over native IGF-I. The numbers in green refer to the six cysteine residues of native IGF-I, the helices 1, 2 and 3 are indicated and 'N' and 'C' denote the N-terminal and C-terminal ends of the molecule. Panel (b) shows a close-up of the region in which two disulphides are switched in the alternative molecule. For clarity, only the Cys^6, Cys^{48} and Cys^{52} residues are labelled.

a)

Helix 2

48   6

47

52

Helix 1

N

Helix 3   C

61

18

b)

Cys 6

Cys 48

Cys 6

Cys 48

Cys 52

Cys 52
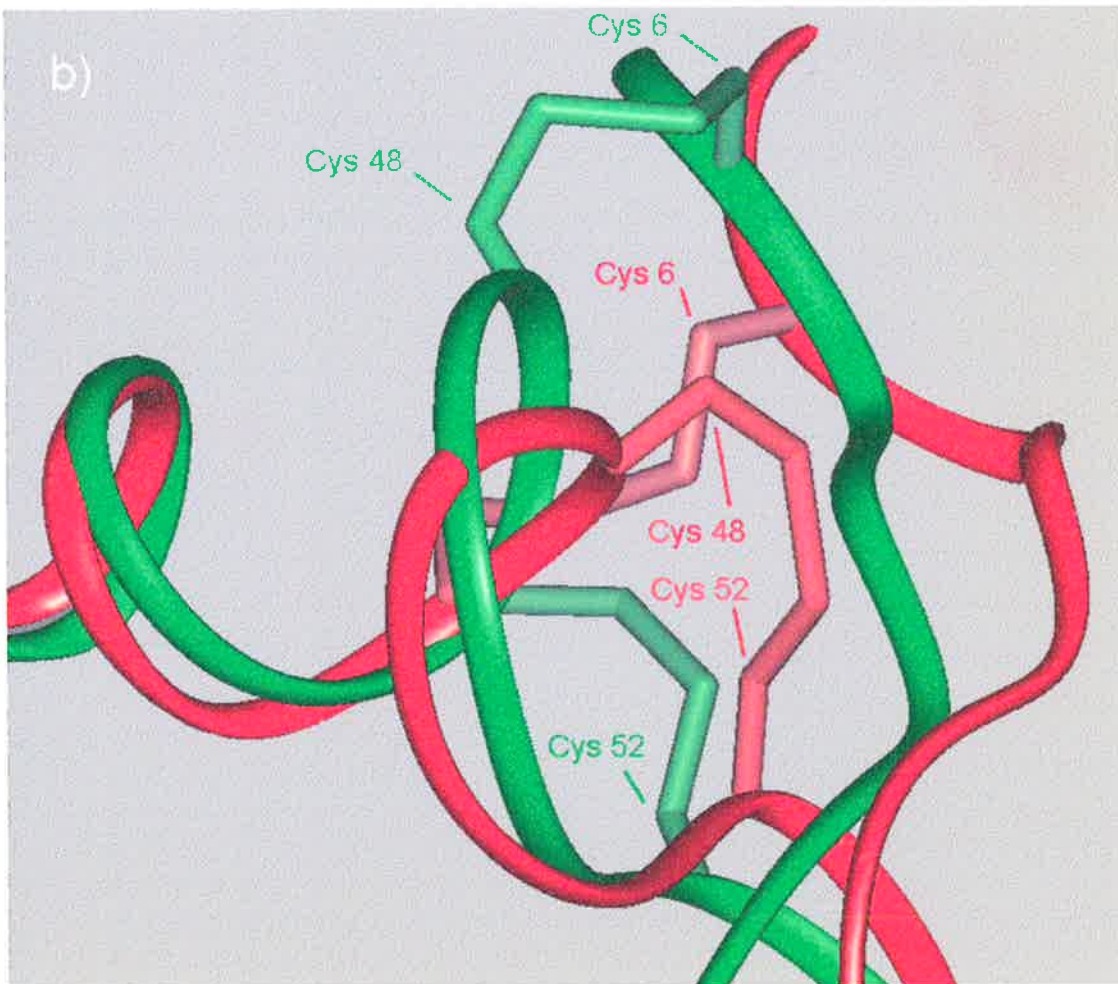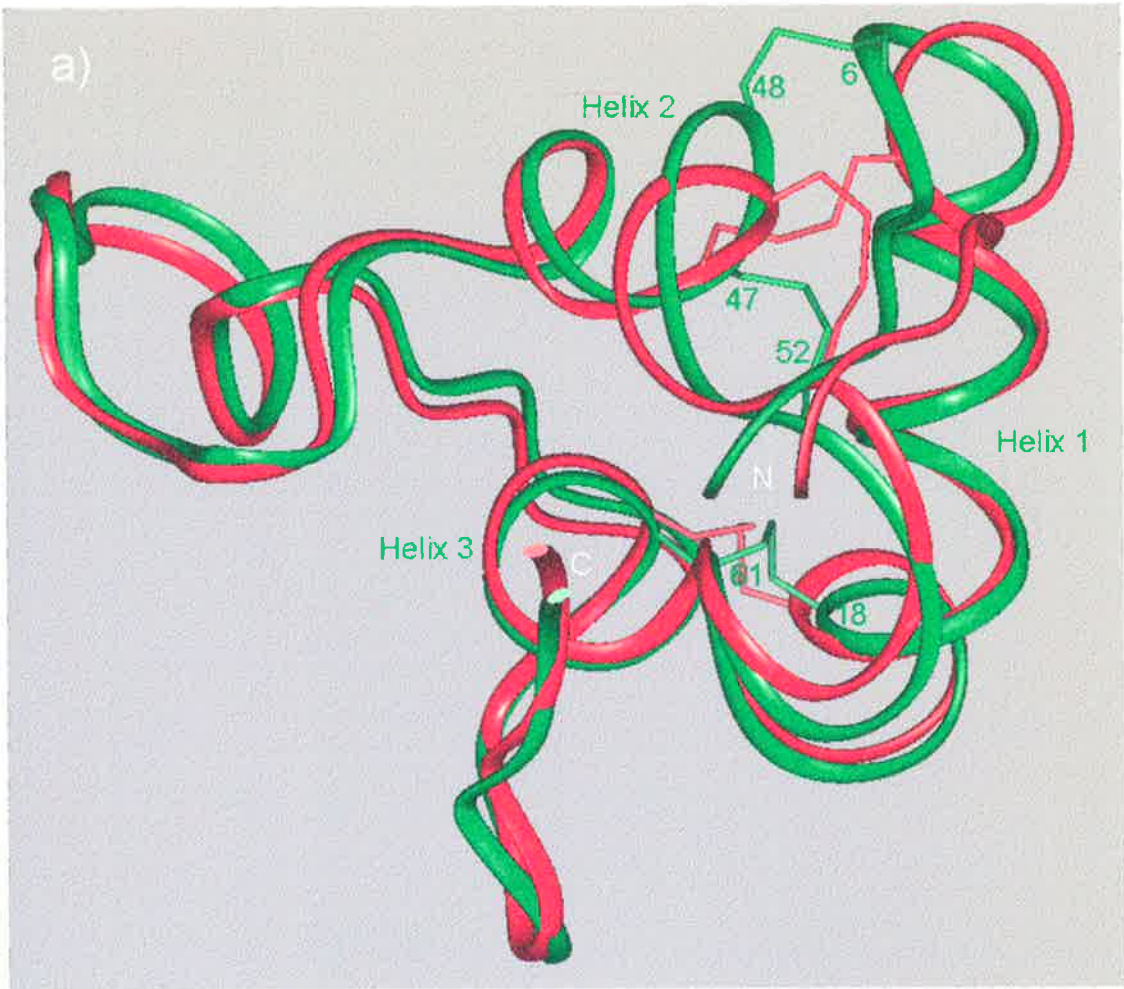
changes but the greatest changes occur at the N-terminus and at helix 2 Furthermore, the program *PROCHECK* (Laskowski et al., 1993) shows that residues Ile^{43} to Cys^{47} in alternative IGF-I (helix 2 in native IGF-I) do not possess helical structure. These two structures are compared further in the Discussion section.

5.4: DISCUSSION

The two major products of the oxidative folding of IGF-I are the native and alternative three-disulphide isomers containing Cys^{18} - Cys^{61}, Cys^{47} - Cys^{52}, Cys^{6} - Cys^{48} and Cys^{18} - Cys^{61}, Cys^{48} - Cys^{52}, Cys^{6} - Cys^{47} disulphide bonds respectively (Hober *et al.*, 1992; Miller *et al.*, 1993). The *in vitro* oxidative folding of IGF-I presented in this thesis (Chapter 4) showed two similar products to be formed at equilibrium. The digestion of these two isomers showed that the altered topology presented different pepsin sensitive peptide bonds, giving rise to altered cleavage patterns (Table 5.1). The cleavage patterns of the two isomers were almost identical to that obtained by Forsberg *et al.* (1990), where the native isomer is digested to form the AEF fragment and the alternative isomer is digested to form the ADE fragment. The different fragments generated are the result of the altered presentation of pepsin sensitive Ile^{43} - Val^{44} and Asp^{53} - Leu^{54} peptide bonds in the different folded forms. The digestion of each of the putative native and alternative isomers of [Gly^{3}]IGF-I, [Arg^{3}]IGF-I, [Lys^{9}]IGF-I, Long-IGF-I, Long-[Gly^{3}]IGF-I, Long-[Arg^{3}]IGF-I and Long-[Lys^{9}]IGF-I showed that the sensitivity of the Ile^{43} - Val^{44} and Asp^{53} - Leu^{54} peptide bonds was similar to the native and alternative isomers of IGF-I. Thus the native isomers and the alternative isomers of each of the IGF analogues represent two sets of proteins that differ by their structure.

The alternative and native isomers of IGF-I have been compared in biological assays for binding to binding proteins in sheep plasma (Hodgkinson *et al.*, 1989), for binding to the type 1 IGF receptor in rat L6 myoblasts (Hodgkinson *et al.*, 1989) and in human placental membranes (Axelsson *et al.*, 1992; Forsberg *et al.*, 1990; Miller *et al.*, 1993) and for the stimulation of protein synthesis using L6 myoblasts (Hodgkinson *et al.*, 1989). In each case the alternative isomer was demonstrated to have a lower potency compared to the native form. The altered biological property of the misfolded form can therefore be used as a probe to determine structure. Thus the native and the alternative isomers of each IGF were compared for binding to IGFBPs, to the type 1 IGF receptor and for the stimulation of protein synthesis in rat L6 myoblasts.

In correctly folded IGF-I, the important residues involved in binding to IGFBPs are Glu^3 (Bagley *et al.*, 1989; Bayne *et al.*, 1988), Phe^{49}-Arg^{50}-Ser^{51} (Cascieri *et al.*, 1989; Baxter *et al.*, 1992) and the α-helix between residues 8 and 18 of the B-domain (Baxter *et al.*, 1992). However, IGFBP-4 which is the major form of binding protein secreted by L6 cells (McCusker and Clemmons, 1994), requires IGF-I residues Phe^{49}-Arg^{50}-Ser^{51} for optimal binding (Clemmons *et al.*, 1992; Bach *et al.*, 1993). The observed reduced IGFBP binding properties of the alternative folding isomers is therefore consistent with a model of misfolded IGF-I (Hejnæs *et al.*, 1992) which shows modification of the IGF-I structure in the region of residues 47-52. The alternative folding isomers of IGF-I, [Gly^3]IGF-I, [Arg^3]IGF-I and [Lys^9]IGF-I show a parallel reduction in potency compared with the native folding form in binding to the rat IGFBPs of at least 200-fold. This shows that the same conformational changes occur in each of these alternative folding isomers, affecting the same residues on the IGFBP binding faces. Furthermore, the difference in the observed potency of both sets of isomers demonstrates that Glu^3 and Glu^9 are

involved in binding to IGFBP-4. The 'Long' fusion protein analogues do not exhibit this uniform reduction in potency between the alternative and native isomer forms. This is most striking for Long-[Arg^3]IGF-I where the alternative and native isomers have a similar low affinity for IGFBPs (Figure 5.4). These results suggest that the N-terminal extension affects the surface topology at least in the region of binding to IGFBP-4.

The important residues in the binding of native IGF-I to the type 1 IGF receptor are Phe^{23}, Tyr^{24} (Cascieri *et al.*, 1988), Arg^{21}, Val^{44} (Cooke *et al.*, 1991), Tyr^{31} and Tyr^{60} (Bayne *et al.*, 1990; Maly and Lüthi, 1988) and part of the C-domain (Bayne *et al.*, 1988). The native isomers of the normal length IGFs show that Glu^3 is not involved in binding to the type 1 IGF receptor, and that the charge reversal at Glu^9 has some effect. Furthermore, the presence of the N-terminal extension had only a marginal affect on binding for the native isomers. Disulphide isomerization of the normal length IGFs has a marked effect on the binding to the type 1 IGF receptor, reducing binding by at least 100-fold (Figure 5.5). For the alternative isomers containing the N-terminal extension, the reduction in potency was not as great as observed for the equivalent normal length proteins. Thus structural changes due to misfolding reduce binding to the type 1 IGF receptor, but this effect is ameliorated in the presence of the N-terminal extension.

The biological potency of the IGF isomers is dependent on their binding to IGFBPs and to the type 1 IGF receptor (Ross *et al.*, 1989). The properties of each IGF isomer in the protein synthesis assay can be explained by their relative binding to the IGFBPs and to the type 1 receptor. Native [Gly^3]IGF-I and [Arg^3]IGF-I are more potent than IGF-I which in turn is more potent than [Lys^9]IGF-I. The increased potency of the Glu^3 substituted IGF-I is due to the lower affinity to the IGFBPs and the lower potency of the Glu^9 substituted IGF-I is due to reduced binding to the receptor. This pattern is maintained for the alternative isomers of the normal length IGFs and for the native IGFs

containing the N-terminal extension (Figure 5.6). For the alternative isomers of the IGFs containing the N-terminal extension, the poor binding to the IGFBPs and to the type 1 IGF receptor results in a low activity in the protein synthesis assay. However, the alternative isomer of Long-[Arg^3]IGF-I is surprisingly potent in the protein synthesis assay as it binds poorly to the IGFBPs but reasonably well to the type 1 IGF receptor.

Using the biological properties of the IGF isomers as a probe of their structure, the alternative isomers of the normal length IGFs clearly have an identical disulphide conformation, a result consistent with the pepsin cleavage patterns presented above. The results for the alternative isomers of the IGF fusion proteins are not so clear even though the pepsin cleavage patterns suggest that their structures are equivalent. For these analogues the N-terminal extension appears to reduce the effect of misfolding but in an unpredictable manner, resulting in unusual biological properties. This is especially true for the alternative isomer of Long-[Arg^3]IGF-I.

Circular dichroism (CD) spectroscopy shows that the alternative isomer of IGF-I has a reduced helical content compared to the native isomer (Hober *et al.*, 1992; Miller *et al.*, 1993). The observation that the alternative disulphide linkage has a reduced α-helical content is not restricted to IGF-I. Molecules related to the insulin-like growth factors also have misfolded disulphide structures. For example, insulin, pro-insulin and bombyxin-II are susceptible to misfolding when folded *in vitro* (Morris *et al.*, 1990; Vértesy *et al.*, 1995 and Nagata *et al.*, 1992). For bombyxin-II, a disulphide bonded isomer containing equivalent disulphide bonds to the alternative isomer of IGF-I showed a helix content of 15 % compared to 28 % for the native isomer, results very similar to that for IGF-I (Nagata *et al.*, 1992).

In addition to the reduced α-helical content, alternative IGF-I has a greater reverse-turn intensity as judged by FTIR (Miller *et al.*, 1993). Two-dimensional NOESY

^{1}H-NMR studies (Miller *et al.*, 1993) show that the NOEs between Tyr[60], Leu[15] and Phe[23] is similar for the alternative and the native structures. Tyrosine fluorescence data and photochemical dynamic nuclear polarization (photo-CIDNP) indicate that the three tyrosines: Tyr[24], Tyr[31] and Tyr[60] have similar local structures in both isomeric forms (Miller *et al.*, 1993). The combined data indicates that there are few changes in structure around the Cys[18] - Cys[61] disulphide bond that is common to both isomeric forms. The effect of the amino acid substitutions at Glu[3] in the oxidative folding reactions suggest a possible stabilizing effect with Arg[56] in the alternative isomer. The pepsin cleavage study presented above indicates that the Asp[53] - Leu[54] peptide bond is more exposed and the Ile[43] - Val[44] peptide bond is less exposed in the alternative structure than in the native structure. The reduced binding of the alternative isomer to IGFBPs indicates that the structure is perturbed in the region of Glu[3] and Phe[49]-Arg[50]-Ser[51] and the reduced binding to the type 1 IGF receptor implicates an altered structure in the regions of Arg[21], Phe[23], Tyr[24], Tyr[31], Val[44] and Tyr[60]. The preliminary two dimensional TOCSY ^{1}H-NMR data indicates that the switching of the disulphide bonds has led to the introduction of non-helical structure, a result that is consistent with CD spectroscopy (Hober *et al.*, 1992; Miller *et al.*, 1993). Furthermore, residues Ala[8], Glu[46], Phe[49] and possibly Cys[47] and Cys[48] are implicated in ß-sheet or turn-type regions, suggesting that the structure of helix 2 (residues 42 to 48) is disrupted in the alternative isomer and that these changes affect the N-terminal end of helix 1 (residues 8 to 18). Thus the structure of the alternative isomer of IGF-I appears to be similar to the native isomer in the region of the common Cys[18] - Cys[61] disulphide bond and that the greatest changes occur near the binding sites of IGFBP-4 and the type 1 IGF receptor.

The model of the alternative isomer of IGF-I (Figure 5.10) was created from the average NMR structure of native IGF-I (Brookhaven code 2GF1: Cooke *et al.*, 1991) by switching two disulphide bonds and energy minimization. Thus Cys[6] moved

approximately 7 Å closer to Cys^{47} and Cys^{48} moved approximately 8 Å closer to Cys^{52}. The resultant model (Figure 5.10) shows relatively few changes to the structure in the region of the Cys^{18} - Cys^{61} disulphide bond. The greatest changes appear to occur at the opposite side of the molecule. The model of alternative IGF-I shows Glu^{3} to be closer to Arg^{56}, a result consistent with that predicted from the refolding studies of [Gly^{3}]IGF-I and [Arg^{3}]IGF-I (Chapter 4). In the model of alternative IGF-I the Cys^{6} - Cys^{47} disulphide bond alters the conformation of the N-terminal end of helix 1, and in conjunction with the Cys^{48} - Cys^{52} disulphide bond disrupts helix 2 (Ile^{43} to Cys^{47}). This model is therefore consistent with experimental data such as the preliminary NMR data which suggests that residues such as Ala^{8} in helix 1 and residues in helix 2 are in quite different environments (Section 5.3.3.). The significant changes of this part of the model are also consistent with the binding protein studies which suggest that Glu^{3} and Phe^{49}-Arg^{50}-Ser^{51} residues are perturbed (Section 5.3.2.). Furthermore, the loss of helical structure in helix 2, is consistent with the observation that the Asp^{53} - Leu^{54} peptide bond more exposed and that residues Ile^{43} and Val^{44} are less exposed in the pepsin digestion experiments (Section 5.3.1.). However, the model of alternative IGF-I does not readily explain the dramatic reduction in binding to the type 1 IGF receptor. Residues Arg^{21}, Phe^{23}, Tyr^{24}, Tyr^{31} and Tyr^{60} which are known to be important in binding to the receptor are only altered slightly in the alternative model compared to native IGF-I. But Ile^{43}, located in the altered helix 2, may affect crucial van der Waals interactions with Tyr^{60} (Bayne *et al.*, 1990; Sakano *et al.*, 1991), a residue that is important in binding to the type 1 IGF receptor. Furthermore, Glu^{9} has been shown to have an unexpected effect on binding to the type 1 IGF receptor (Figure 5.5) and changes at the N-terminal end of helix 1 in the alternative molecule may have a greater impact than was anticipated from studies of the native

protein reported in the literature (Cascieri *et al.*, 1988, Bayne *et al.*, 1988, Maly and Lüthi, 1988, Bayne *et al.*, 1990 and Cooke *et al.*, 1991). *

* Refer to Addendum

CHAPTER 6

THE DISULPHIDE FOLDING PATHWAY OF LONG-[ARG^3]IGF-I

6.1: INTRODUCTION

Studies on the oxidative folding of IGF-I have indicated the role of a native-like single disulphide species and a native-like two-disulphide species which leads to the formation of the native molecule in a sequential fashion (Hober *et al.*, 1992). The presence of the alternative isomer containing Cys^{18}-Cys^{61}, Cys^{6}-Cys^{47}, and Cys^{48}-Cys^{52} disulphide bonds at equilibrium indicates that other non-native-like disulphide intermediates must also exist in the folding pathway (Hober *et al.*, 1992), arising from a divergence of the pathway after the Cys^{18} - Cys^{61} disulphide bond has formed (Miller *et al.*, 1993). However, these descriptions of IGF folding are incomplete as they do not account for the very large number of intermediates that have been detected early in the folding of IGF-I (Chapter 4). In a similar manner to the oxidative folding of potato carboxypeptidase inhibitor (Chang *et al.*, 1994), hirudin (Chang, 1994) and epidermal growth factor (Chang *et al.*, 1995), it is possible that the *in vitro* oxidative folding of IGF-I may involve multiple parallel folding pathways that form a population of different one-, two- and three-disulphide species which then rearrange to form the native isomer in a thermodynamically controlled process. The observation that IGF-I analogues which contain mutations at Glu^3 in the B-domain affect the oxidative folding outcome is consistent with the latter view of oxidative folding as the mutations presumably stabilize the native fold (or alternatively destabilize the non-native-like intermediates and products), affecting the flow of intermediates to the native form in a thermodynamically controlled process. The 13-amino acid N-terminal extension also affects the folding outcome by imparting a steric constraint at a crucial point in the folding pathway, a striking example of which is provided by analysis of the folding of Long-[Lys^9]IGF-I (Chapter 4). Thus the relative thermodynamic stabilities of the isomers and intermediates play a crucial role in directing the folding pathway of the

IGFs. To investigate the *in vitro* oxidative folding pathway of IGFs, the analogue Long-[Arg^3]IGF-I was chosen for study rather than IGF-I because fewer intermediates occur during folding and unlike IGF-I, the intermediates that are present early in folding are not well-populated at equilibrium, thereby simplifying the study.

The approach that I used to establish elements of the *in vitro* oxidative folding pathway of Long-[Arg^3]IGF-I was first to identify the intermediates of folding that were well-populated after one minute of refolding and then to examine their appearance and disappearance with time when folded from the fully reduced form. Secondly, disulphide rearrangements reactions of the equilibrium folding products and of some of the folding intermediates were used to aid the construction of a folding scheme.

6.2: METHODS

The materials and methods are described in detail in Chapter 2. However, for clarity the methods employed in this chapter are summarized here.

6.2.1: Oxidative Folding Reactions

The oxidative folding reaction of Long-[Arg^3]IGF-I was investigated using the techniques described in Section 2.2.4. Briefly, purified Long-[Arg^3]IGF-I was fully reduced and refolded as described in Section 4.2.1.1. However, more early time points were sampled and a more discerning microbore C$_4$ reverse-phase HPLC analysis was used to analyse the folding products. To isolate folding intermediates a similar refolding reaction was stopped after one minute and the folding products were stabilized by acidification, purified by reverse-phase HPLC and lyophilized

(Section 2.2.4.7). The folding products were allowed to continue refolding by placing them in the refolding buffer in rearrangement reactions (Section 2.2.4.6). The isolated folding products were also alkylated to irreversibly trap any free thiols (Section 2.2.4.3) and then characterized by digestion and mass spectrometry (Sections 2.2.3.3. and 2.2.3.4). The rearrangement reactions of the alternative and native isomers of Long-[Arg^3]IGF-I were also examined and compared to those of IGF-I (Section 2.2.4.6).

6.3: RESULTS

6.3.1: Refolding Fully Reduced Long-[Arg^3]IGF-I

Fully reduced Long-[Arg^3]IGF-I was refolded using identical conditions to those described in Section 4.2.1.1. However, more samples were taken early in the folding reaction and a more discerning analytical microbore C$_4$ reverse-phase HPLC was used to separate the folding products (Figure 6.1). Although baseline separation was not achieved for each of the folding intermediates, it was possible to group the folding products into 'early' and 'late' intermediates. The fully reduced protein (R) quickly refolded and after 30 seconds a cluster of early folding intermediates was identified (peaks 8 to 14), peak 2 which is equivalent to the major intermediate (I) in Figure 4.8 was also present at this time but at a low concentration. The early intermediates were dominated by peaks 10 and 11. After one minute, the cluster of early intermediates increased slightly in concentration (32 % of the total protein) but the reduced species R decreased and peak 2 increased. At this time peaks 4 to 7 were present but not well-populated (Figure 6.2). At two minutes refolding the cluster of early intermediates started to decline and the concentration of peak 2 increased.

Figure 6.1: Analysis of the *in vitro* oxidative folding of Long-[Arg^3]IGF-I

Long-[Arg^3]IGF-I was fully reduced and refolded in the presence of 2 M urea, 0.1 M tris pH 8.7, 10 mM glycine, 1 mM EDTA, 0.4 mM dithiothreitol and 1 mM 2-hydroxyethyl disulphide. Samples of the folding mixture were stabilized at the indicated times by lowering the pH to 2.1 with trifluoroacetic acid (TFA) and were analysed by microbore C_4 reverse-phase HPLC using a 20 % to 45 % gradient over 50 minutes in the presence of 0.1 % TFA. The reduced protein (R), the folding intermediates 2, 4 and 11, and the native (N) and alternative (A) isomers are indicated.

t=180 min

N

A

t=10 min

t=9 min

t=8 min

t=7 min

t=6 min

t=5 min

t=4 min

t=3 min

t=2 min

R

t=1 min

2

4

11

t=30 sec

t=0
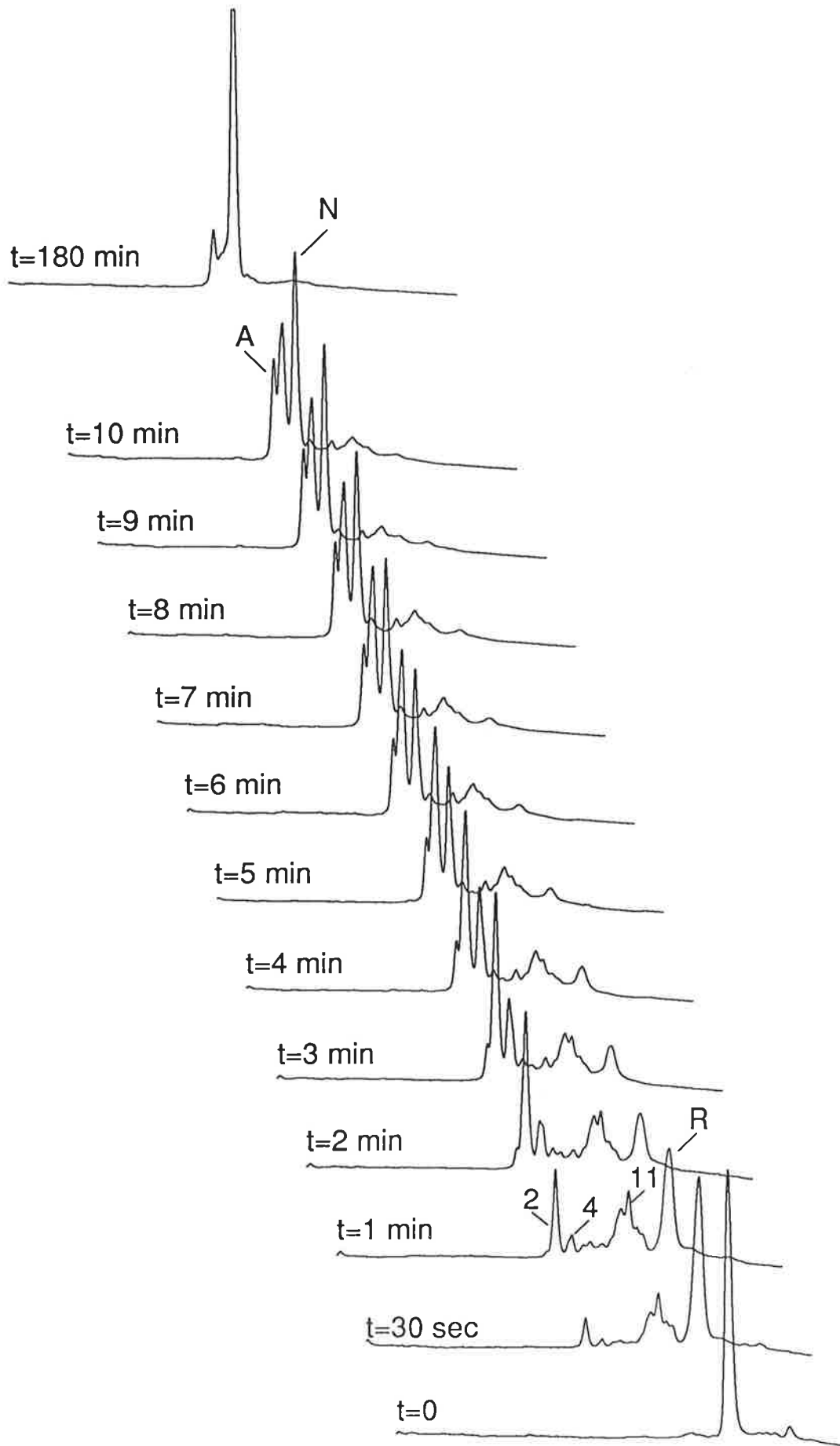
The subsequent time samples showed the cluster of early intermediates to continue to decline, peaks 4 to 7 reached a maximum at 3 minutes (11 % of the total protein) and peak 2 reached a maximum at 5 minutes (39 % of the total protein). Peak 2 could therefore be described as a well-populated late-intermediate. Thereafter the concentration of peak 2 declines steadily over 60 minutes (see figure 4.8). The concentration of the native isomer and the alternative isomers continued to rise reaching 43 % and 16 % of the total protein at 10 minutes. The concentration of the alternative isomer reached a maximum of 19 % at approximately 30 minutes (results not shown) but declined thereafter to 11 % of the total protein at equilibrium. The folding reaction reached equilibrium at 140 minutes. At equilibrium, peaks 5 and 6 were detected (approximately 2 % of the total protein) and low levels of peaks 7 to 14 were detected (less than 2 % of the total protein).

6.3.2: Identification of Folding Intermediates Present During the Refolding of Long-[Arg^3]IGF-I

6.3.2.1: Isolation of Folding Intermediates

Long-[Arg^3]IGF-I was fully reduced and refolded for one minute. The folding products were acidified to prevent disulphide exchange and were purified using preparative reverse-phase HPLC (Figure 6.3). After one or two further rounds of purification, the folding intermediates were assessed for quantity and lyophilized. The reverse-phase HPLC purification resolved most of the intermediates and the absorbance profile was almost identical to the analytical microbore C_4 reverse-phase HPLC profile shown in Figure 6.1. Except for peaks 12, 13 and 14 which proved difficult to purify as disulphide rearrangement appeared to continue slowly at low pH,
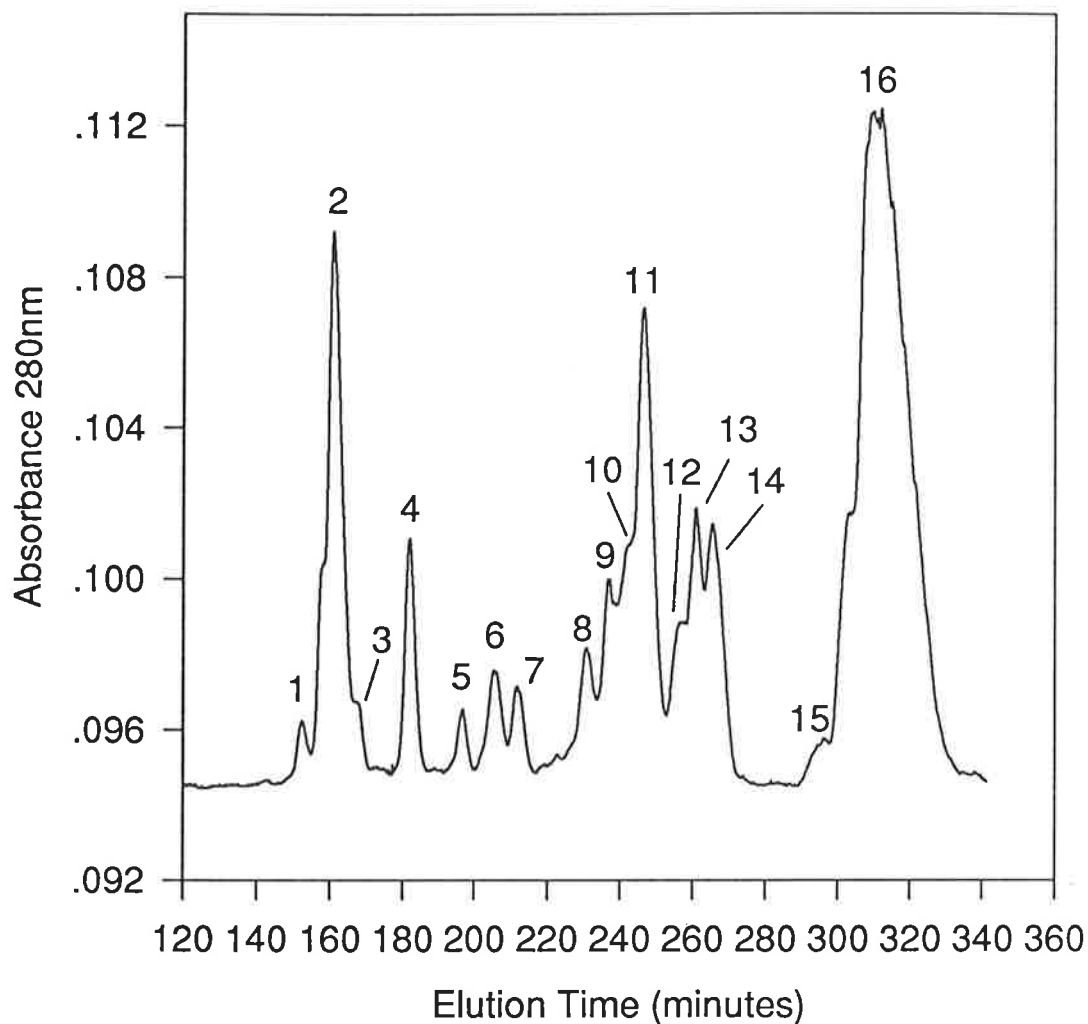
Figure 6.3: Purification of Long-[Arg^3]IGF-I Intermediates Present After Folding for One Minute

Fully reduced Long-[Arg^3]IGF-I was refolded for one minute in the presence of 2 M urea, 0.1 M tris pH 8.7, 10 mM glycine, 1 mM EDTA, 0.4 mM dithiothreitol and 1 mM 2-hydroxyethyl disulphide. The reaction was stopped by lowering the pH to ~ 2.1 with TFA and the mixture was purified using a C$_4$ preparative reverse-phase HPLC column and a linear 0.1 % per minute acetonitrile gradient in the presence of 0.1 % TFA. Peaks 1 to 16 are shown.

the folding intermediates were purified to homogeneity. Peaks 3 and 15 were recovered at only very low levels. Within a short period of time, a portion of each isolated peak was reacted with either iodoacetic acid or iodo[2-^{3}H]acetic acid at pH 8.7 to stabilize any free thiols.

6.3.2.2: The Fully Reduced Protein

A small quantity of lyophilized Peak 16 (Figure 6.3) was digested with pepsin for 3 hours at 22°C and the products were analysed by mass spectrometry (Table 6.1). The mass spectrometry analysis revealed peptide fragments that accounted for nearly all the residues present in Long-[Arg^3]IGF-I (residues 47 to 53 were not resolved by mass spectrometry). Furthermore, no disulphide linked fragments were detected. Peak 16 was also alkylated with iodo[2-^{3}H]acetic acid analysed by mass spectrometry. This analysis showed an increase of mass consistent with the presence of six S-carboxymethylcysteine residues (observed mass 9468.0 ± 1.0, theoretical mass 9466.1). However, approximately 30 % of the modified protein contained four S-carboxymethylcysteine residues (observed mass 9349.6 ± 0.8, theoretical mass 9349.9) indicating that a proportion of cysteine residues were not alkylated. This suggests that some disulphide bond formation had occured during the reaction with iodoacetate.

6.3.2.3: The Well-Populated Late Intermediate: Peak 2

Long-[Arg^3]IGF-I was reduced and refolded for 1.5 minutes to generate peak 2. The acidified mixture was purified by preparative reverse-phase HPLC (Figure 6.4a). Peak 2 was isolated (Figure 6.4b) and reacted with iodoacetic acid for 30 seconds at pH 8.7. The reaction was stopped by injection onto a microbore C$_4$ reverse-phase HPLC column previously equilibrated with 0.1 % TFA, and the bound

Amino Acid Residues	Expected Mass	Observed Mass
-13 to -4	1093.4	1092.6 ± 0.7
-3 to 9	1263.4	1262.8 ± 0.2
8 to 15	857.9	857.3 ± 0.4
16 to 24	1063.2	1063.0 ± 0.3
25 to 46	2367.6	2368.1 ± 0.7
54 to 59	817.0	816.7 ± 0.3
60 to 70	1147.4	1148.1 ± 0.2

Table 6.1: Pepsin Digestion Fragments Detected Following Cleavage of Fully Reduced Long-[Arg³]IGF-I.

A small quantity of fully reduced Long-[Arg³]IGF-I (Peak 16; Figure 6.3) was digested with pepsin for 3 hours at 22ºC and the products were analysed by electrospray mass spectrometry. Except for the fragment containing amino acids 47 to 53, all the other amino acids were observed.

Figure 6.4: Purification and Alkylation of the Long-[Arg^3]IGF-I Folding Intermediate Peak 2

Long-[Arg^3]IGF-I was fully reduced and refolded for 1.5 minutes in the presence of 2 M urea, 0.1 M tris pH 8.7, 10 mM glycine, 1 mM EDTA, 0.4 mM dithiothreitol and 1 mM 2-hydroxyethyl disulphide. (a) The folding products were purified using a C$_4$ preparative reverse-phase HPLC column and a linear 0.1 % per minute acetonitrile gradient in the presence of 0.1 % TFA. (b) Microbore C$_4$ reverse-phase HPLC analysis was carried out on the purified intermediate. (c) The purified protein was was reacted with iodoacetate for 30 seconds and loaded on to a microbore C$_4$ reverse-phase HPLC column. The bound protein was eluted with a 20 % to 50 % acetonitrile gradient over 3 minutes in the presence of 0.1 % TFA. (d) The major peak was isolated and analysed by analytical microbore C$_4$ reverse-phase HPLC.
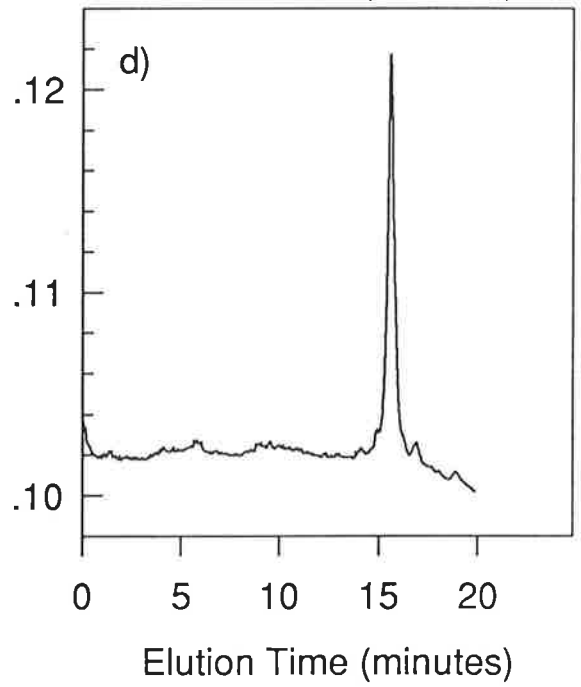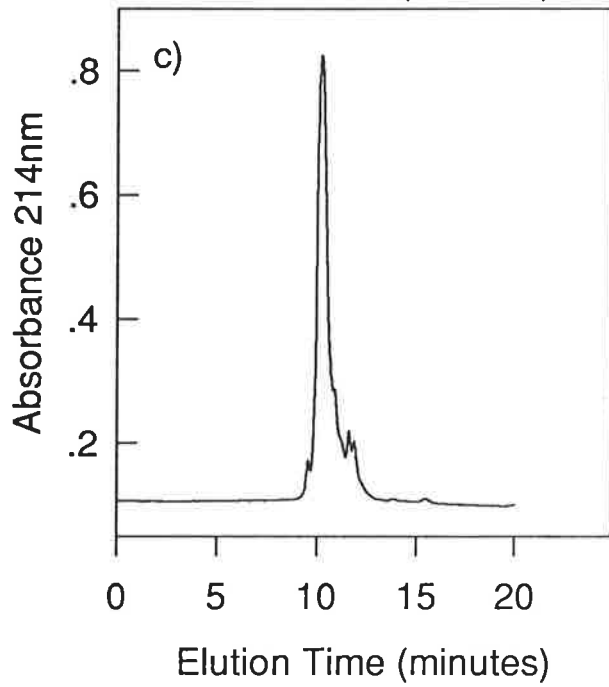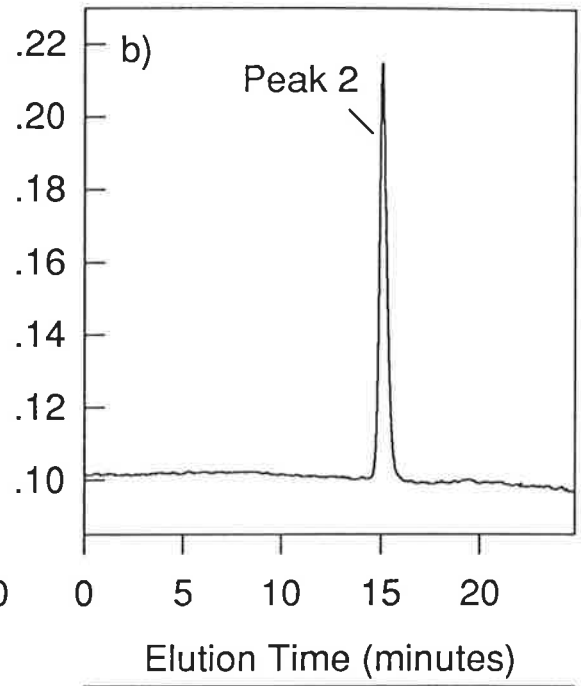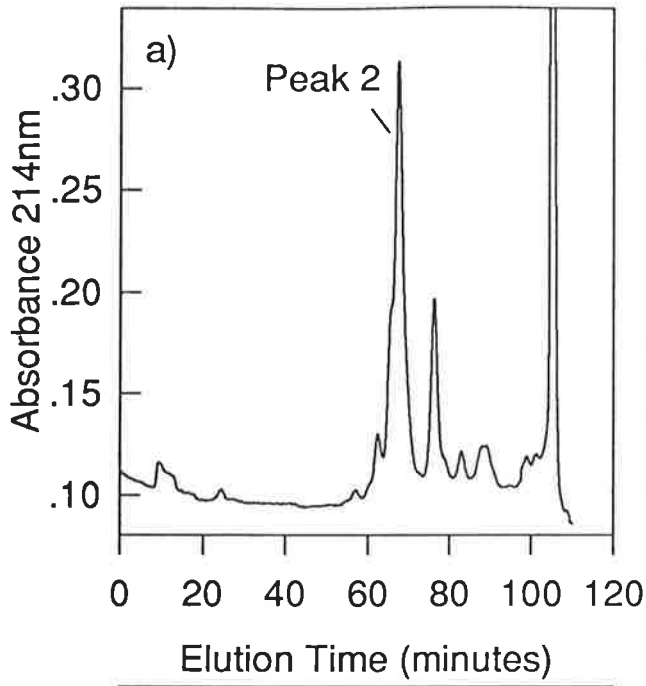
protein was eluted with an acetonitrile gradient (Figure 6.4c). The major product was isolated (Figure 6.4d) and mass spectrometry showed the major component to have a mass of 9230.5 ± 0.6. Native Long-[Arg^3]IGF-I has a mass of 9111.6, indicating that the modified protein has a mass increase consistent with the presence of two S-carboxymethylcysteine residues (theoretical mass 9229.8). The pepsin digest / mass spectrometry analysis of the modified protein showed the presence of a disulphide linked CG fragment (Figure 5.3) containing residues 16 to 24 and 60 to 70 (observed mass 2210.0 ± 0.1, theoretical mass 2209.6) and a disulphide linked AE fragment containing two S-carboxymethylcysteine residues within the sequence -3 to 10 and 47 to 53 (observed mass 2324.0 ± 0.4, theoretical mass 2324.7). The detection of these disulphide linked fragments confirms the presence of two disulphide bonds and two free thiols in peak 2. However, the AE fragment does not indicate which cysteine residue is linked to Cys^6. To identify the cysteine residues present as free thiols, peak 2 was modified with iodo[2-^{3}H]acetic acid at pH 8.7 to label the free thiols and then digested with pepsin. The fragments were purified by microbore C$_4$ reverse-phase HPLC (Figure 6.5) and the radioactive-labelled fragment was analysed by N-terminal sequence analysis. This analysis showed two disulphide linked fragments present in equal proportions, one with a sequence that corresponds to residues -3 to 10 linked to 46 to 53 and the other with a sequence that corresponds to residues -3 to 10 linked to 47 to 53. The N-terminal sequence analysis unambiguously showed the presence of S-carboxymethylcysteine at positions 47 and 52. Thus Cys^6 is linked to Cys^{48} and Cys^{18} is linked to Cys^{61} in Long-[Arg^3]IGF-I peak 2.
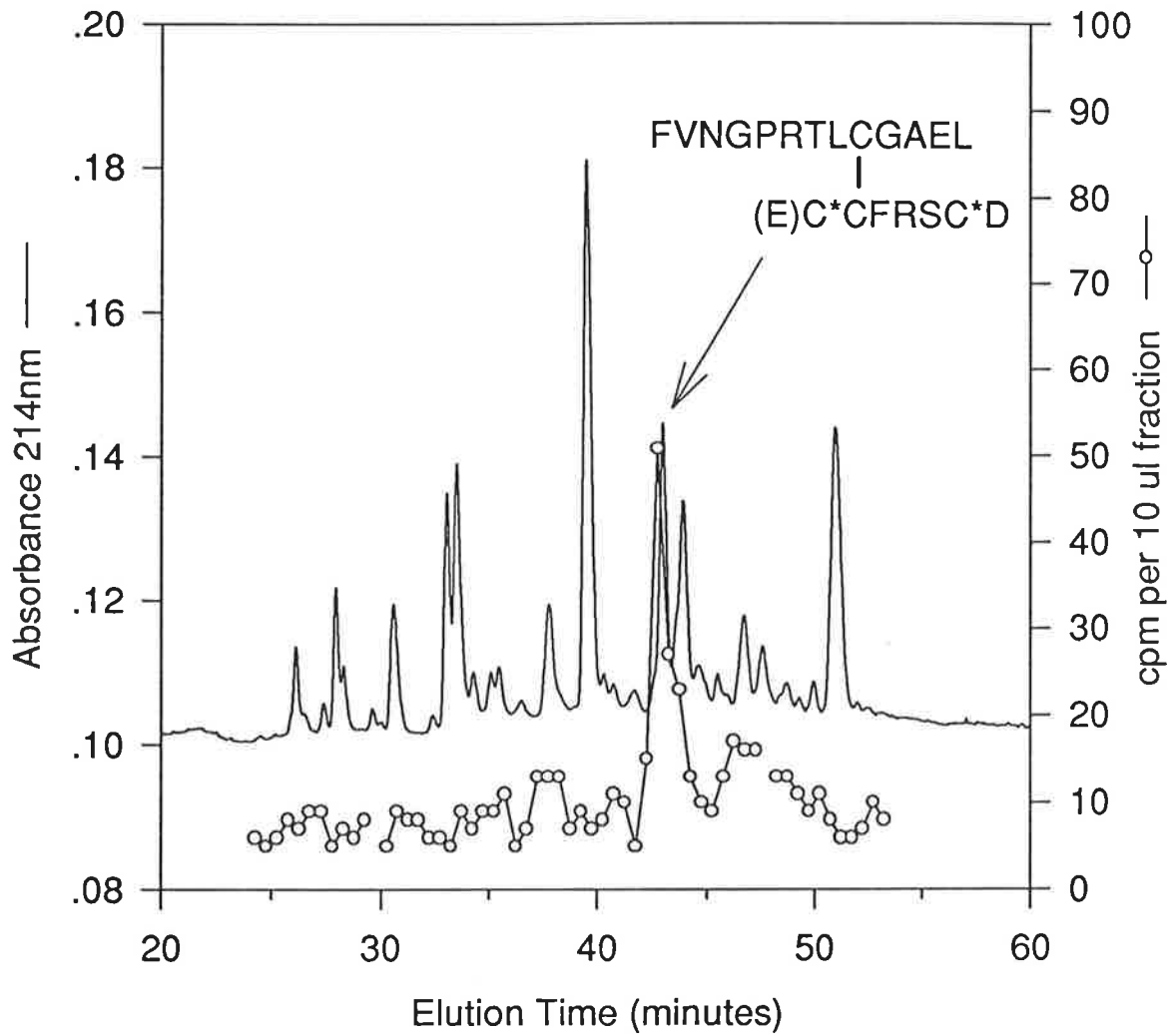
Figure 6.5: Pepsin Digestion of Alkylated Peak 2

The acid trapped folding intermediate of Long-[Arg³]IGF-I (peak 2) was alkylated with iodo[2-³H]acetate, desalted and digested with pepsin. After three hours incubation the digest mixture was loaded on to a microbore C_4 reverse-phase HPLC column and the bound protein was eluted with a 0 to 60 % acetonitrile gradient over 60 minutes in the presence of 0.1 % TFA. The eluted protein was collected and the fractions were analysed for the ³H-label. The single radio-active peptide fragment was dried and its sequence determined by N-terminal sequence analysis. The disulphide linked fragment contains two labelled S-carboxymethylcysteine residues (C*).

6.3.2.4: Analysis of Other Long-[Arg³]IGF-I Folding Intermediates

Acid trapped peaks 1, 4, 5, 6, 7, 8, 9, 11, 12, 13 and 14 isolated from the one minute refold of Long-[Arg³]IGF-I (Figure 6.3) were alkylated to irreversibly trap any free thiols. Control reactions were carried out with the native and alternative isomers of Long-[Arg³]IGF-I as well as peaks 2 and 16. The mass spectrometry results of the modified peptides show that the reaction with iodoacetate for 30 seconds at pH 8.7 allows some degree of rearrangement of disulphide bonds. The mass spectrometry of the native isomer of Long-[Arg³]IGF-I shows that 85 % of the protein does not react with iodoacetate (mass 9114.8 ± 0.8) and that approximately 15 % of the protein becomes unfolded reacting to form two S-carboxymethylcysteine residues (mass 9232.8 ± 0.6). However, the alternative isomer of Long-[Arg³]IGF-I does not react with iodoactate and peak 2 shows the presence of only the two expected S-carboxymethylcysteine residues (mass 9232.4 ± 0.2). Peak 16 shows some degree of folding as observed in Section 6.3.2.2. For each control reaction the presence of +16 and +32 modified proteins were noted. The folding intermediates present after one minute of folding showed the additions of 0, 2, 4 or 6 S-carboxymethylcysteine residues (Table 6.2) which correspond to the presence of 3, 2, 1 or 0 disulphide bonds respectively. Each isolated intermediate showed some degree of disulphide rearrangement during the alkylation reaction, although this was minor for peaks 1, 2, 5 and 7. The major product of peaks 1, 4, 5, 7, 8 and 12 in the alkylation reaction indicates that S-carboxymethylcysteine residues were not formed. The major alkylated product of peaks 2, 6, 9, 13 and 14 shows the presence of two S-carboxymethylcysteine residues. Peak 11 and peak 16 show the presence of four and six S-carboxymethylcysteine residues respectively.

	Mass of S-carboxymethylated Protein	Number of S-carboxymethylcysteine residues
Peak 1	9113.7 ± 0.5	0
Peak 2	9232.4 ± 0.2	2
Peak 4	9111.0 ± 0.6, 9230.7 ± 0.1	0*, 2
Peak 5	9113.3 ± 0.4	0
Peak 6	9112.5 ± 0.8, 9230.8 ± 0.9	0, 2*
Peak 7	9112.7 ± 0.5	0
Peak 8	9114.0 ± 1.4, 9231.2 ± 1.5	0*, 2
Peak 9	9230.8 ± 1.3, 9348.6 ± 0.7	2*, 4
Peak 11	9229.8 ± 0.9, 9348.1 ± 1.0	2, 4*
Peak 12	9113.0 ± 0.3, 9231.0 ± 0.8	0*, 2
Peak 13	9112.7 ± 0.4, 9232.4 ± 1.3	0, 2*
Peak 14	9230.9 ± 0.3, 9349.7 ± 1.6	2*, 4
Peak 16	9351.1 ± 0.8, 9468.8 ± 0.5	4, 6*

Table 6.2: Mass Spectrometry of S-carboxymethylated Long-[Arg^3]IGF-I Folding Intermediates

The folding intermediates of Long-[Arg^3]IGF-I present after oxidative folding for one minute were isolated, modified with iodoacetic acid and analysed by mass spectrometry. Minor products of the alkylation reaction (<10 % of the total products) are not shown. An asterix denotes the major alkylated product.

In a separate experiment peaks 2, 11 and 16 were reacted with iodo[2-^{3}H]acetate and the incorporation of ^{3}H-label was measured; in this analysis the proteins had measured specific activities of 37, 75 and 112 cpm per μg protein respectively. Peak 2 is known to contain two disulphide bonds and two free thiols (Section 6.3.2.3) and peak 16 contains six free thiols (Section 6.3.2.2). The level of incorporation of peak 11 is intermediate to that of peaks 2 and 16, therefore peak 11 contains one disulphide bond and four ^{3}H-S-carboxymethylcysteine residues. Pepsin digestion of peak 11 followed by mass spectrometry showed the presence of a disulphide linked fragment containing residues 16 to 24 and 60 to 70 (observed mass 2209.4 ± 0.7, theoretical 2209.6), a result that was noted in both alkylated and acid trapped peptides. No other disulphide-linked fragments were detected. Therefore, peak 11 contains a single disulphide between Cys^{18} and Cys^{61}.

The pepsin digestion / mass spectrometry analysis of peak 4 isolated after reacting with iodoacetate showed the presence of fragments consistent with that of a three-disulphide native molecule (results not shown). However, the pepsin digestion of acid trapped peak 4 shows the presence of a disulphide linked fragment containing residues 16 to 24 and 60 to 70 (observed mass 2208.4 ± 0.7, theoretical 2209.6) and a fragment consistent with residues 39 to 54 containing a disulphide bond and a free thiol (observed mass 1784.0 ± 0.2, theoretical 1784.1). No other disulphide linked fragments were detected. Thus peak 4 contains one disulphide bond between residues Cys^{18} and Cys^{61} and a second disulphide bond involving cysteine residues 47, 48 and 52. It is reasonable to speculate that the second disulphide bond is native-like between residues 47 and 52 as the molecule refolds very fast in the alkylation reaction to form the native molecule.

The pepsin digestion / mass spectrometry analysis of peak 1 showed a result consistent with the alternative isomer containing disulphide bonds between Cys^{18} - Cys^{61}, Cys^{48} - Cys^{52}, Cys^6 - Cys^{47} (results not shown). A similar analysis of the other acid trapped or alkylated intermediates isolated after one minute of refolding did not provide any conclusive evidence of disulphide linked fragments. However, a noteable feature of each of these digests was that a disulphide linked fragment linking Cys^{18} and Cys^{61} was not detected.

6.3.3: Rearrangement Reactions of Long-[Arg³]IGF-I

6.3.3.1: The Rearrangement Reactions of Alternative and Native Isomers of Long-[Arg³]IGF-I

When the purified native and alternative isomers of Long-[Arg³]IGF-I are added to a buffer containing 2 M urea, 10 mM glycine, 1 mM EDTA, 0.1 M tris pH 8.7 and the redox constituents dithiothreitol (0.4 mM) and 2-hydroxyethyl disulphide (1 mM), they rearrange to an equilibrium distribution identical to that when refolded from the fully reduced species. However, the rate at which they do so is slightly different for each isomer (Figures 6.6a and 6.6b). When the rearrangement reaction starts with the native isomer of Long-[Arg³]IGF-I, the equilibrium distribution is reached in approximately 100 minutes (Figure 6.6a). This time taken to reach equilibrium is similar to that observed when Long-[Arg³]IGF-I is refolded from the fully reduced species (Figure 4.8). When the rearrangement reaction starts with the alternative isomer of Long-[Arg³]IGF-I, the distribution approaches equilibrium after approximately 200 minutes but the final equilibrium distribution is reached after 12 hours (results not shown). For both rearrangement reactions, the only well-populated folding isomers present were the native and the alternative isomers. The

Figure 6.6: The Rearrangement Reactions of the Native and Alternative Isomers of Long-[Arg^3]IGF-I and IGF-I

The rearrangement reactions of (a) purified native Long-[Arg^3]IGF-I, (b) the alternative isomer of Long-[Arg^3]IGF-I, (c) native IGF-I and (d) the alternative isomer of IGF-I were carried out in a buffer containing 2 M urea, 0.1 M tris pH 8.7, 10 mM glycine, 1 mM EDTA, 0.4 mM dithiothreitol and 1 mM 2-hydroxyethyl disulphide. Samples of the rearrangement mixture were stabilized at various times by lowering the pH to 2.1 with trifluoroacetic acid (TFA) and were analysed by microbore C$_4$ reverse-phase HPLC using a 20 % to 45 % gradient over 25 minutes in the presence of 0.1 % TFA. The elution profiles, measured at 214 nm, were used to calculate the abundance of each isomer at each time point. The abundance of native (filled symbols) and alternative (open symbols) isomers of Long-[Arg^3]IGF-I (▼, ▽), and IGF-I (●, ○) are indicated.
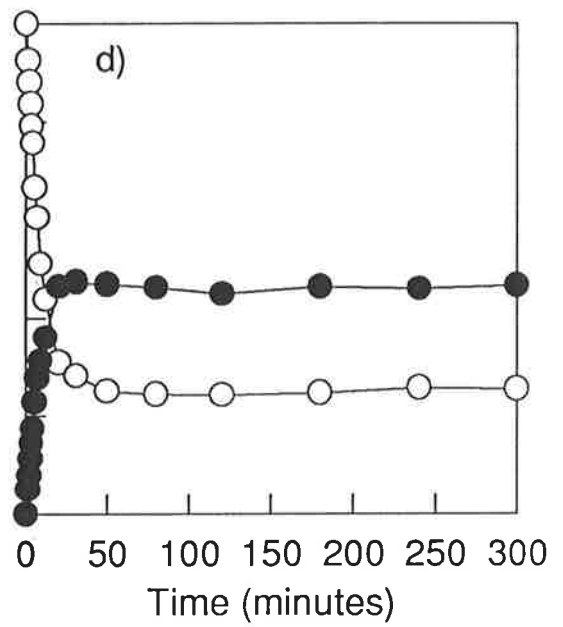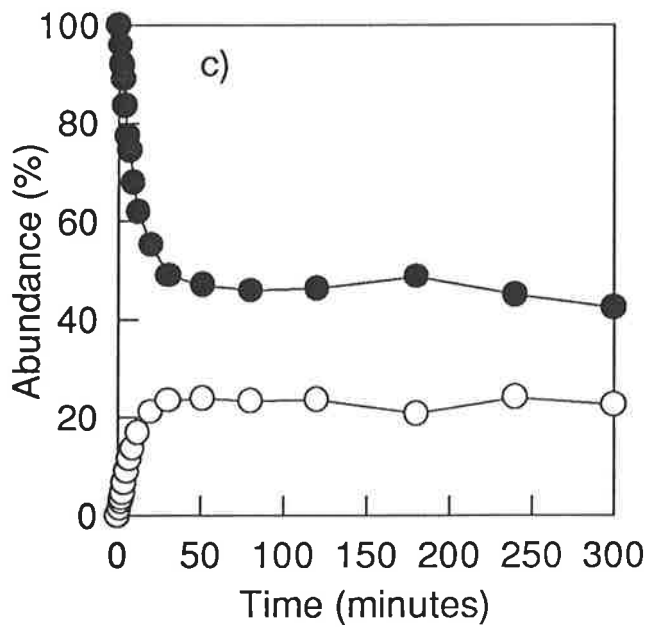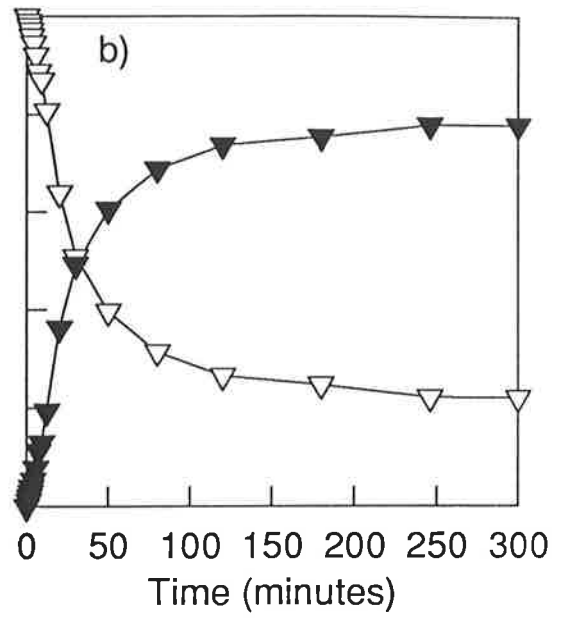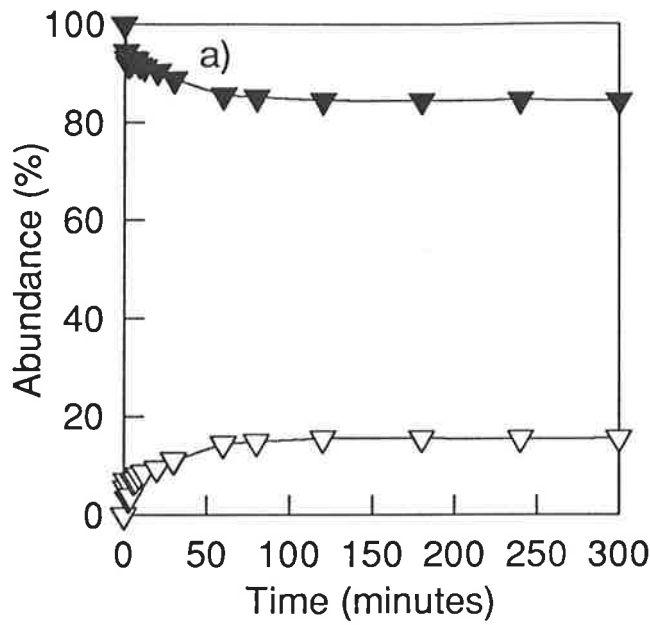
rearrangement reactions of the native and alternative isomers of IGF-I (Figures 6.6c and 6.6d) reach the same equilibrium distribution obtained when IGF-I is refolded from the fully reduced protein. Furthermore, their reaction times are similar, reaching completion within 30 minutes (see also Figure 4.2).

In control reactions using a buffer containing 2 M urea, 10 mM glycine, 1 mM EDTA and 0.1 M tris pH 8.7, but without the redox buffer constituents, the alternative and native isomers of Long-[Arg^3]IGF-I or IGF-I remain as their respective isomer forms. Even after 12 hours there are no detectable changes in their microbore C_4 reverse-phase HPLC absorbance profiles (results not shown).

6.3.3.2: The Rearrangement Reactions of Folding Intermediates of Long-[Arg^3]IGF-I

In a similar manner to the rearrangement reactions of the native and alternative isomers presented above, isolated folding intermediates (Section 6.3.2.1) were allowed to rearrange in a buffer containing 2 M urea, 10 mM glycine, 1 mM EDTA and 0.1 M tris pH 8.7, with and without the redox buffer constituents. The rearrangement reaction of peak 2 (Figure 6.7a) was similar to the refolding reaction of fully reduced Long-[Arg^3]IGF-I (Figure 4.8) and at equilibrium the distribution of folding products was identical to that of the refolding reaction. Furthermore, the time taken to reach the equilibrium distribution was approximately 50 minutes, much faster than for the refolding reaction. The rearrangement reactions of peaks 9 and 11 were similar to that of peak 2 (Figure 6.7b). However, the reaction of peak 4 was much faster with the protein rearranging to form large quantities of the native isomer within minutes (Figure 6.7b). A detailed analysis of the microbore C_4 reverse-phase HPLC absorbance profiles that were used to generate the data presented in Figure 6.7b indicates that most folding intermediates were involved in the rearrangement reactions of peaks 9 and 11 (results not shown). However, for the rearrangement

Figure 6.7: The Rearrangement Reactions of the Folding Intermediates of Long-[Arg^3]IGF-I

The purified acid-trapped folding intermediates, peaks 2, 4, 9 and 11 were reacted in a buffer containing 2 M urea, 0.1 M tris pH 8.7, 10 mM glycine, 1 mM EDTA, 0.4 mM dithiothreitol and 1 mM 2-hydroxyethyl disulphide. Samples of the rearrangement mixture were stabilized at various times by lowering the pH to 2.1 with trifluoroacetic acid (TFA) and were analysed by microbore C_4 reverse-phase HPLC using a 20 % to 45 % gradient over 25 minutes in the presence of 0.1 % TFA. The elution profiles, measured at 214 nm, were used to calculate the abundance of each isomer at each time point. Panel (a) shows the detailed rearrangement reaction of acid trapped peak 2 over 30 minutes. The abundance of the folding intermediate (♦), the native isomer (▼) and the alternative isomer (∇) are indicated. Panel (b) shows the abundance of the native isomer of Long-[Arg^3]IGF-I generated with time in the rearrangement reactions of Peak 2 (▼), Peak 4 (■), Peak 9 (▲) and Peak 11 (●).
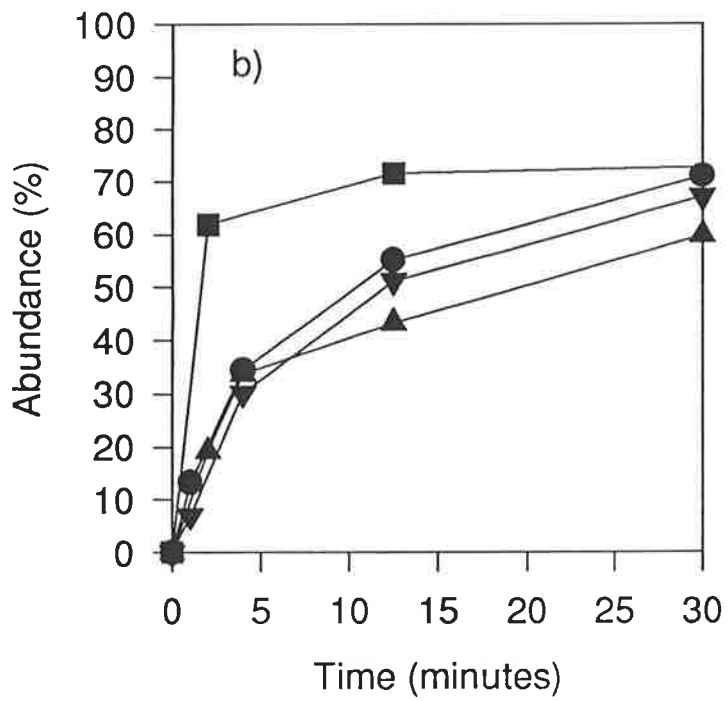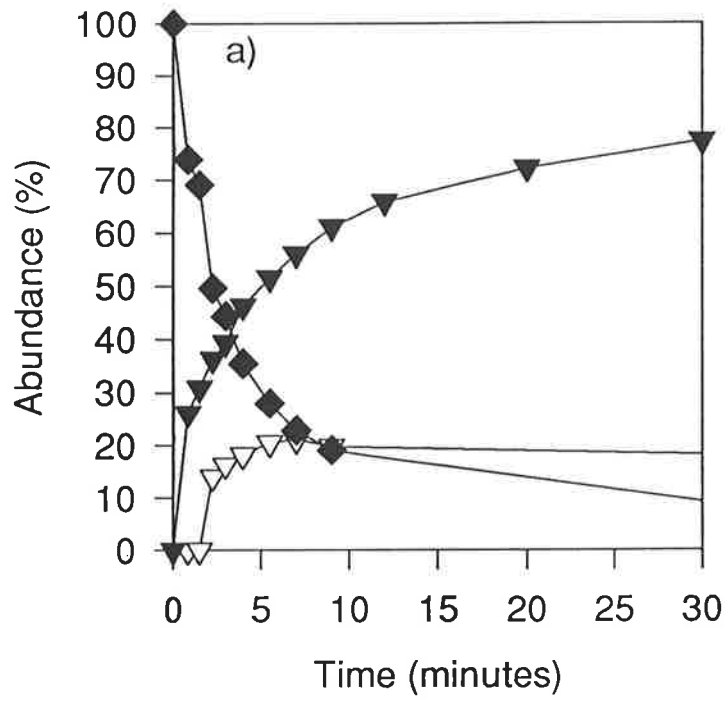
reaction of peaks 2 and 4, the various intermediates were not well-populated at any time point. After 180 minutes, each intermediate had rearranged to a similar equilibrium distribution as that obtained for the refolding of the reduced protein (results not shown). Unlike the control rearrangement reactions of the native and alternative isomers, the folding intermediates rearranged in the absence of redox reagents at a rate approximately 10 times slower than in the presence of the redox buffer dithiothreitol / 2-hydroxyethyl disulphide.

6.4: DISCUSSION

The *in vitro* oxidative folding pathway of Long-[Arg^3]IGF-I has many similarities and differences compared to the folding of IGF-I. Clearly the folding outcome of Long-[Arg^3]IGF-I is different from IGF-I in that the native three-disulphide isomer is well-populated at equilibrium and that folding intermediates are absent (Chapter 4). However, the disulphide folding intermediates that are involved in the folding 'pathway' are not necessarily different even though the reverse-phase HPLC analyses show different proportions of the transiently stable intermediates. An intermediate that is not well-populated or is apparently absent does not mean that it does not exist in the folding scheme. This is because a folding species may be highly reactive and progress to another state without accumulating. Conversely, a well-populated transient intermediate represents a folding species that is relatively slowly turned over in the folding pathway. Thus the kinetics of the individual steps in the folding pathway dictate the appearance of intermediates at any particular time. Furthermore, the thermodynamic stabilities of the folding intermediates and the products present at equilibrium will determine the observed kinetics of the pathway.

For a protein containing six cysteine residues there are theoretically many permutations for forming one or more disulphide bonds. For example, there are 15 possible permutations linking two out of the six cysteine residues, 45 permutations linking four cysteines and a further 15 permutations linking six cysteines. Thus, there is a total of 75 possible disulphide conformations containing one, two or three disulphide bonds. Of these 75 different possible disulphide conformations, only 7 have native-like disulphide architecture, one of which is the native three-disulphide isomer. The *in vitro* oxidative folding pathway of Long-[Arg^3]IGF-I, like IGF-I, clearly involves more than 7 different isomers (Figure 6.1) indicating that non-native-like isomers are present in the folding scheme. Furthermore, the presence of the alternative isomer at equilibrium implicates non-native-like disulphide intermediates. There appear to be three types of folding intermediates; the 'early' intermediates which are well-populated after refolding for one minute (peaks 8 to 14), a 'late' intermediate which is well-populated at 4 minutes and a group of intermediates that appear swiftly but disappear slowly (peaks 4 to 7: Figure 6.2). A crude analysis of the intermediates present after refolding for one minute shows the presence of the fully reduced protein and one-, two- and three-disulphide isomers (Figure 6.3 and Table 6.2). The native three-disulphide isomer was not detected this early in folding. The 'late' intermediate contained a two-disulphide isomer and the group of 'early' intermediates contained a mixture of three, two and one-disulphide isomers. The third group contained a mixture of three and two-disulphide isomers. A key feature of this type of analysis is that so many three-disulphide folding isomers appear to be present after folding for one minute. However, there remains the formal possibility that free thiols may be buried and thus not able to react with the alkyating reagent, thereby over-estimating the number of three-disulphide and two-disulphide folding

isomers. Another problem with this type of analysis is that the acid trapped intermediate may rearrange swiftly to another disulphide isoform during the course of the reaction with iodoacetate.

The analysis of the well-populated 'late' intermediate (peak 2) showed that it had a native-like two-disulphide conformation containing Cys^{18} - Cys^{61} and Cys^{6} - Cys^{48} disulphide bonds and two free thiols. The rearrangement reaction of this isomer was much faster than for the folding of the fully reduced protein and yielded identical folding products. The accumulation of this isomer in the refolding reaction from the fully reduced protein suggests that the formation of the disulphide link between Cys^{47} and Cys^{52} is a rate limiting step for the formation of the native three-disulphide IGF-I. Peak 4, which was was present in modest proportions in the first two minutes of folding contained a native-like Cys^{18} - Cys^{61} disulphide bond and second disulphide bond involving cyteines 47, 48 or 52.* The rearrangement rate of this isomer is much faster than for peak 2 (Figure 6.7) which suggests that the second disulphide bond is native-like, probably the Cys^{47} - Cys^{52} disulphide bond. Thus, the two possible native-like two disulphide folding intermediates that contain a common Cys^{18} - Cys^{61} disulphide bond have been identified. Furthermore, the speed at which the Cys^{6} - Cys^{48} disulphide bond is formed in peak 4 compared to that of the Cys^{47} - Cys^{52} disulphide bond in peak 2 suggests that the Cys^{47} - Cys^{52} disulphide bond is strained,* an observation similar to that for IGF-I (Hober *et al.*, 1992). It is possible that peak 4 is the 'productive' intermediate in the folding of Long-[Arg^{3}]IGF-I and that peak 2 represents a kinetically accessible but unproductive species; this possibility is explored further in the next chapter. The analysis of peak 11 showed the presence of a single disulphide between Cys^{18} and Cys^{61}, a native-like disulphide isomer which has been observed during the folding of IGF-I (Hober *et al.*, 1992).

* Refer to Addendum

The rearrangement reactions of the native and alternative isomers of Long-[Arg³]IGF-I demonstrate that in the presence of a redox buffer at pH 8.7, the purified folding isomers will rearrange to an equilibrium distribution identical to that when refolded from the fully reduced protein, a result similar to that obtained for IGF-I (Miller *et al.*, 1993). In the rearrangement reaction from the native isomer to the alternative isomer, or *vice versa*, at least three intermediates are involved (Figure 6.8). In both rearrangement reactions the only folding species present that are well-populated are the native and alternative isomers, the other intermediates do not accumulate significantly. The time taken to reach equilibrium is different for each isomer which provides some evidence that one step in the rearrangement reaction of the alternative isomer to the native isomer is slow. This data confirms that the *in vitro* oxidative folding of Long-[Arg³]IGF-I, like IGF-I, is thermodynamically controlled. A key point in the rearrangement reactions of Long-[Arg³]IGF-I (and IGF-I) is that a redox buffer (or at least a reducing agent) is essential to break the first disulphide bond so that disulphide rearrangement of other disulphide bonds continues driven in a thermodynamically controlled process. In an identical buffer, but without the redox constituents, the three-disulphide isomeric forms are stable for a long period of time with no detectable disulphide rearrangement. In contrast, the rearrangement reactions of Long-[Arg³]IGF-I folding intermediates isolated after one minute of folding do not require redox buffer constituents and fold relatively slowly at pH 8.7 to form the same folding products, thereby demonstrating that free thiols are present facilitating disulphide rearrangement.

The *in vitro* oxidative folding of Long-[Arg³]IGF-I involves a number of folding intermediates that contain native-like and non-native-like disulphide bonds that appear rapidly. Amongst these intermediates the native-like single-disulphide
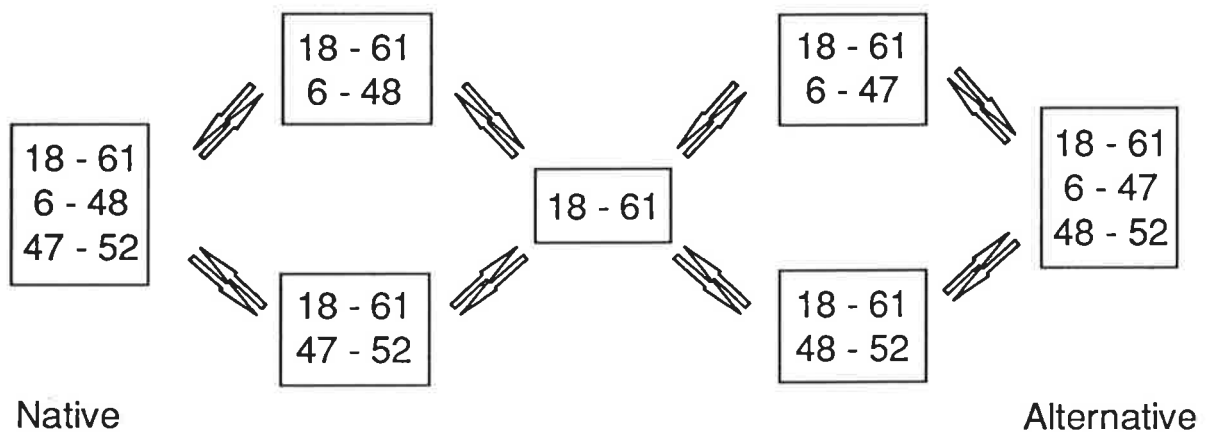
Figure 6.8: Schematic of the Rearrangement Reaction of Long-[Arg³]IGF-I *

In the presence of a redox buffer the native isomer of Long-[Arg³]IGF-I will rearrange to the alternative isomer and *vice versa* in a thermodynamically controlled process. This scheme involves a common 18 - 61 disulphide bond and shows the possible disulphide conformations that link cysteine residues 6, 47, 48 and 52 in the rearrangement reaction. The numbers refer to the amino acid positions in the sequence.

(Cys^{18} - Cys^{61}) intermediate is well populated. This intermediate declines with the emergence of the two-disulphide (Cys^{18} - Cys^{61}, Cys^{6} - Cys^{48}) intermediate. The next step in the folding pathway is not clear but it is possible that the Cys^{18} - Cys^{61}, Cys^{6} - Cys^{48} intermediate forms the native three-disulphide molecule directly in a rate-limiting step. Alternatively, this intermediate may unfold to the single-disulphide Cys^{18} - Cys^{61} intermediate and fold to the native isomer via the putative Cys^{18} - Cys^{61}, Cys^{47} - Cys^{52} intermediate (Figure 6.9). The folding pathway also includes 'off-pathway' reactions such as those that lead to the alternative three-disulphide isomer. Although 'off-pathway' products are formed, they are able to rearrange to the native form in a thermodynamically controlled process. It is also likely that other intermediates (not observed or identified in this study) may play a role in the folding of Long-[Arg^{3}]IGF-I, but until they are examined in detail, their role remains elusive. The interactions of a few intermediates in the folding reaction of Long-[Arg^{3}]IGF-I are summarized in Figure 6.9 and are discussed further in the next chapter.

In summary, the *in vitro* oxidative refolding pathway of Long-[Arg^{3}]IGF-I is complex, but the major 'flow' appears to be through a handful of native-like intermediates (Figure 6.9). Other parallel pathways must also exist as judged by the large number of intermediates present early in folding; however, their contribution is not clear. The *in vitro* oxidative refolding reaction of IGF-I probably involves the same intermediates as for Long-[Arg^{3}]IGF-I. But for IGF-I the main 'flow' of the pathway is in competition with interactions that stabilize 'off-pathway' reactions that lead to the alternative isomer and to stable intermediates at equilibrium.
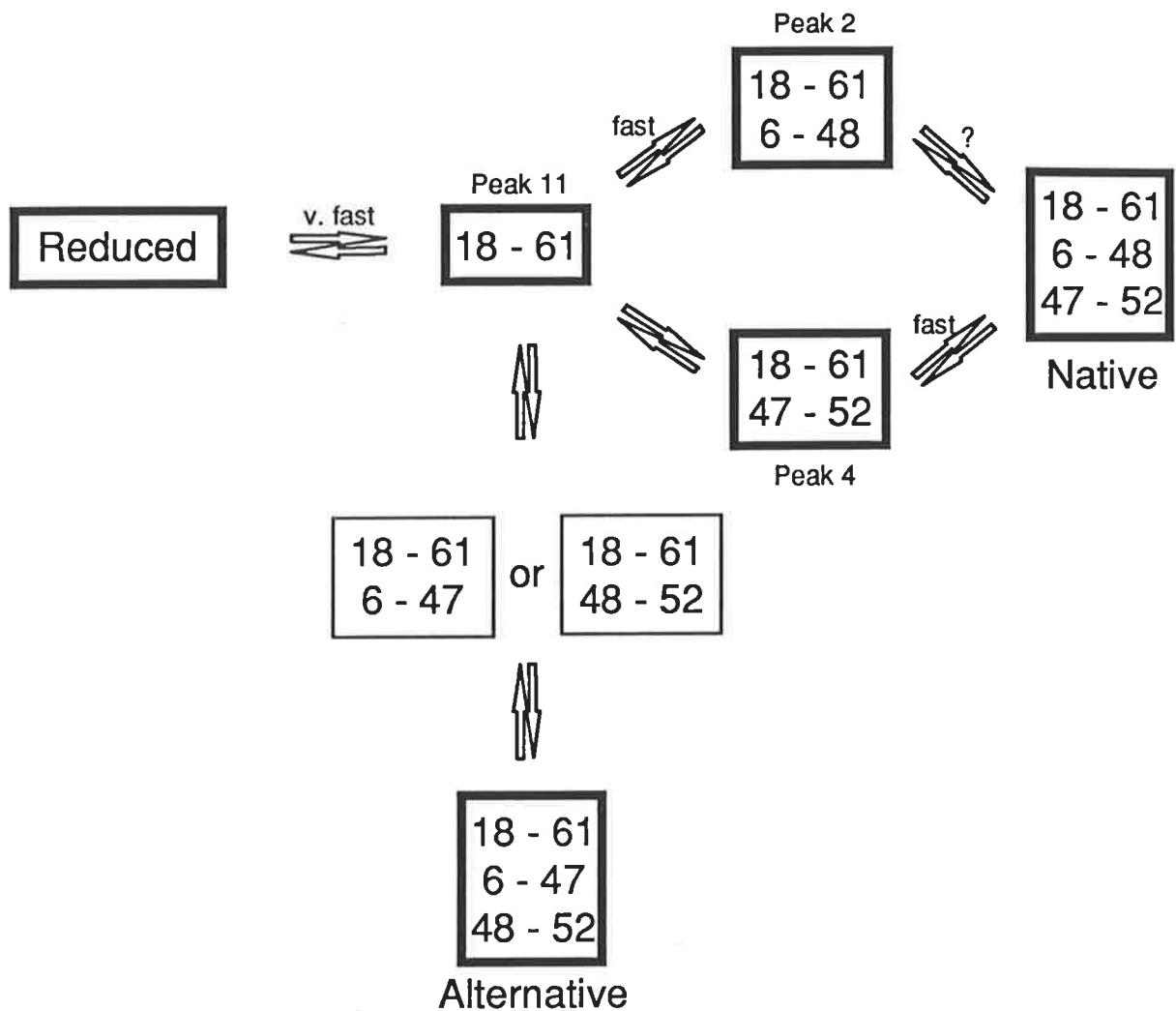
Figure 6.9: A Scheme of the Refolding Reactions of Long-[Arg³]IGF-I [*]

This scheme summarizes the possible interactions of a few intermediates in the oxidative folding reactions of Long-[Arg³]IGF-I. The numbers in the boxes refer to the disulphide linkages between cysteine residues in the amino acid sequence. The bold boxes indicate that the disulphide structure has been observed. The reactions are deduced from the oxidative folding reactions of the fully reduced protein and from the rearrangement reactions of intermediates and of products present at equilibrium.

* Refer to Addendum

CHAPTER 7

GENERAL DISCUSSION

At the time I commenced research for this PhD thesis (July 1992) little was known about the folding of IGF-I beyond the fact that the yield of the correctly folded recombinant molecule was poor regardless of the technique used to prepare the protein. The available evidence for IGF-I folding *in vivo* was and still is limited, with only the native IGF-I molecule being observed in plasma. Around this time the first NMR solution structures of the native IGF-I molecule were published (Cooke *et al.*, 1991; Sato *et al.*, 1992) which confirmed the structural similarity of the IGF molecule to insulin. The first study of the *in vitro* oxidative folding of IGF-I (Hober *et al.*, 1992) was also published which showed the presence of two native-like folding intermediates, an 'off-pathway' non-native intermediate and the non-native three-disulphide isomer which has a similar thermodynamic stability to the native isomer. The non-native three-disulphide isomer, known as misfolded IGF-I, scrambled IGF-I or iso-IGF-I, had been isolated by other workers and limited biological assays provided some indirect evidence about its structure (Hodgkinson *et al.*, 1989; Forsberg *et al.*, 1990; Axelsson *et al.*, 1992). But the data which showed reduced biological potencies compared to the native isomer was variable. The impetus for the research carried out in this thesis was the observation in our laboratory that a fusion protein analogue of IGF-I, prepared to determine binding interactions with IGF-binding proteins, refolded *in vitro* to a much higher yield than did IGF-I (Francis *et al.*, 1992). Furthermore, the misfolded form of this analogue had a similar biological activity to its native form. Thus the goals of my PhD research were two-fold. Firstly, I aimed to characterize the factors that led to the improved thermodynamic stability of the native structure compared to the misfolded structure by a systematic analysis of the *in vitro* oxidative folding of IGF-I mutants and by an analysis of the intermediates present during folding. Secondly, I wished to

investigate the structure of the misfolded IGF-I isomer to explain its unusual stability in the oxidative folding reaction.

While our laboratory had established considerable expertise in producing recombinant IGFs (King *et al.*, 1992; Francis *et al.*, 1992), some of the variants produced in this process remained unidentified. Furthermore, the hydroxylamine cleavage technique previously used to remove the N-terminal fusion protein partner caused hitherto undetected modifications of asparagine and glutamine amino acids. With the advent of the facile analysis of proteins using electrospray mass spectrometry, the hydroxylamine-modified proteins were identified and the problem was addressed. Therefore, I examined the procedures used to recover recombinant IGF-I and IGF-I mutants in order to characterize variants present during downstream processing. The *Escherichia coli* expression system resulted with the IGF-fusion protein accumulating in the cell as inclusion bodies. The inclusion bodies contained the expected fusion protein and a small quantity of a variant containing a formyl group at the N-terminal methionine, a result of altered processing in *E. coli*. These two proteins refolded to yield a mixture of native and misfolded protein which were easily separated by reverse-phase HPLC. The hydroxylamine cleavage reaction was modified by a systematic approach in order to reduce chemical modification while maintaining an adequate cleavage rate to release the native length IGF molecule (see Chapter 3). This study was, to my knowledge, the first time that the hydroxylamine cleavage reaction had been applied to produce a recombinant protein free from chemical heterogeneity. The experience gained from the investigation of the downstream processing techniques was used to prepare well characterized native isomers of IGF-I and a number of IGF-I analogues for *in vitro* oxidative folding studies.

An analysis of the *in vitro* oxidative folding of a variant of IGF-I containing a dipeptide at the N-terminus (Ala-Glu-IGF-I: Hejnæs *et al.*, 1992) or IGF-I containing mutations at Arg[55] and Arg[56] (Rosenfeld *et al.*, 1993) has confirmed that the amino acid sequence of IGF-I can 'code' for two different stable disulphide isomers (Hober *et al.*, 1992; Miller *et al.*, 1993). Yet a charge substitution at Glu[3] in the B-domain alters the thermodynamics of folding, stabilizing the native isomer resulting in an increased yield at equilibrium (Chapter 4). The observation that a single amino acid mutation in IGF-I can affect oxidative folding favourably is not unique; a single amino acid substitution of Leu for Tyr[35] in BPTI leads to an enhanced rate of formation of the native molecule and prevents the accumulation of species that act as kinetic traps (Zhang and Goldenberg, 1993). Another approach to refold IGF-I *in vitro* is the use of a N-terminal fusion protein partner containing a IgG-binding domain derived from staphylococcal protein A (Nilsson *et al.*, 1991; Samuelsson *et al.*, 1991) which improves the solubility of IGF-I. However, the native and misfolded isomers of IGF-I are produced in approximately the same ratio as IGF-I alone (Samuelsson *et al.*, 1994). Our use of a 13-amino acid N-terminal extension derived from porcine growth hormone in the fusion protein versions of the IGF-I mutants further facilitates correct folding. How this occurs is not clear but one may speculate that the N-terminal extension may prevent the contact of regions such as Glu[3] and Arg[56] in IGF-I (Chapter 4) that may otherwise draw the folding pathway into stable non-native disulphide arrangements at the time the second disulphide bond is formed. Furthermore, the N-terminal extension also exerts an influence over other disruptive forces as in the [Lys[9]]-IGF-I analogue (Chapter 4).

When IGF-I is expressed and secreted in *Saccharomyces cerevisiae* the same two disulphide isoforms are detected (Elliot *et al.*, 1990; Axelsson *et al.*, 1992)

as well as disulphide-linked aggregates (Chaudhuri *et al.*, 1991; Chaudhuri and Stephan, 1992). This suggests that the cellular export system lacks essential quantities of components that are required to guide formation of the native isomer and that the thermodynamic process that drives folding *in vitro* may occur in a similar manner in *S. cerevisiae*. However, the presence of misfolded IGF-I can cause an elevated level of secretion of the chaperone BiP (Chaudhuri *et al.*, 1991) which indicates the possibility that the cellular machinery is trying to compensate for the high level production of heterologous protein.

Other factors have been shown to influence the *in vitro* oxidative folding of IGF-I. For example, reduced *E. coli* protein disulphide isomerase (DsbA) can catalyze the conversion of misfolded IGF-I to the native isomer under physiological conditions (Joly and Swartz, 1994), thereby overcoming the kinetic barrier between the two isomer forms. Furthermore, the neuroendocrine protein 7B2 acts as a molecular chaperone in regenerating inactive yeast-derived IGF-I molecules (Chaudhuri *et al.*, 1995). Thus, misfolded IGF-I molecules appear to have the potential to be refolded to the functional form *in vivo*, but this property has not yet been confirmed directly. The apparent absence of misfolded IGF-I *in vivo* suggests that only the native form is secreted from the cell. A recent study by Hober and co-workers (Hober *et al.*, 1994) suggest that an insulin-like growth factor binding protein (IGFBP) is responsible for the maintenance of the native IGF-I form in serum in a thermodynamically controlled process. However, their conclusion, which is based on the property of IGFBP-1 to quantitatively yield native IGF-I *in vitro*, may be unfounded. Their results may be explained by the possibility that IGFBP-1, like IGFBP-3, may have a protein disulphide isomerase activity (Koedam and Van den Brande, 1994; Hober *et al.*, 1994). Alternatively, IGFBP-1 may act by providing a

binding face that facilitates IGF-I refolding in an analogous manner to the affinity maturation of human growth hormone (Lowman and Wells, 1993). Furthermore, the oxidation potential of serum is highly oxidizing (Lash and Jones, 1985) and IGF-I has been demonstrated to be very stable under physiological conditions (Joly and Swartz, 1994). Therefore, native IGF-I may still be considered as the preferred isomer that is secreted from the cell.

There is emerging evidence to suggest that in the endoplasmic reticulum (ER), the sequential action of chaperones is prominent in the kinetics and efficiency of proper protein folding. Rather than act in a quality control sense where the off-pathway folding products are degraded, the chaperones of the ER and the helper proteins appear to assist nascent protein folding, presumably by limiting counter-productive interactions or aggregation that can be essentially irreversible. An example of sequential binding of ER chaperones can be seen for the membrane-bound ER chaperone calnexin and the lumenal ER chaperone BiP on the productive folding of thyroglobulin (Kim and Arvan, 1995). An important observation of this study is that reduction of the ER which causes thyroglobulin aggregation did not involve significant protein degradation but instead redirected most of the protein into the normal folding pathway. Recent investigations into other accessory proteins that help productive folding reveal other aspects of folding in the ER. Protein disulphide isomerase (PDI) which is present at almost millimolar concentrations in the ER and which has high kinetic reactivity as an oxidant, can diminish the importance of kinetically trapped cysteines by either avoiding their formation or providing for rapid oxidation once the stabilizing structure spontaneously unfolds (Walker and Gilbert, 1995). Furthermore, PDI may bind covalently to the substrate, cause unfolding, and subsequent oxidation of the kinetically trapped species (Walker and Gilbert, 1995).

A study of the *E. coli* periplasmic proteins DsbA and prolyl isomerase A on the *in vitro* oxidative folding of a mutant of ribonuclease T_1 with a single disulphide bond and two *cis*-prolyl bonds (Frech and Schmid, 1995) showed that the rank order of the major processes (chain folding, prolyl isomerization and disulphide bond formation) is not fixed, but can differ, depending on the conditions for folding. In view of the emerging understanding of protein folding *in vivo*, it is possible that the IGF-I molecule may be correctly folded in the ER through the actions of chaperones and PDI which serve to overcome the kinetically trapped misfolded molecule and other non-productive folding molecules. It is also possible that prolyl isomerase plays an important role in IGF-I folding in the ER as five proline residues are present in the mature protein. However, most of these prolines are in flexible regions (Cooke *et al.*, 1991) and may not influence the oxidation of cysteine residues.

Little is known about the post-translational processing of IGFs. For example, is the nascent prohormone (Rotwein, 1986) translated in the cytoplasm and exported to the ER for folding and processing or is it inserted co-translationally? Furthermore, is pro-IGF-I (containing the E_a- or E_b-domain) processed in the ER to the mature length protein before or after folding? Regardless of the events that occur before the mature length protein is exported from the cell, it resides in the ER for a period of time under conditions that allow disulphide oxidation and isomerization, as discussed above. The observation that pro-IGF-I (IGF-I-E_a) shows *in vitro* disulphide exchange properties similar to IGF-I (Hober *et al.*, 1994), indicates that the C-terminal domain does not overcome the folding problem of IGF-I. This lends weight to the notion that correct folding of the molecule is provided by the cellular machinery rather than by structural factors, as has been described for some proteins (Section 1.2.3.6).

There has been an explosion of research activity in the area of *in vitro* oxidative folding. One approach is to compare the folding / unfolding properties of disulphide-intact and fully reduced proteins. An analysis of the role of disulphide bonds on the folding of lysozyme (Goldberg and Guillou, 1994) using continuous-flow circular dichroism shows that the molecule with intact disulphide bonds has significant secondary structure within 4 milliseconds. However, the reduced lysozyme molecule was still essentially unfolded at this early time. In contrast, a similar study of the folding of cysteine-modified and disulphide-intact human growth hormone (Youngman *et al.*, 1995) show remarkably similar folding properties. Thus, disulphide bonds are not absolutely required to assist the formation of secondary structure early in folding, an observation consistent with that of a ribonuclease T_1 mutant (Frech and Schmid, 1995). For the oxidative folding reaction of a protein from the fully reduced form, the molten globule state may be involved. The analysis of the disulphide folding properties of bovine α-lactalbumin (Ewbank and Creighton, 1993) and human α-lactalbumin (Creighton and Ewbank, 1994) reveals that the molten globule state which exists under certain conditions allows flexibility and the formation of non-native disulphide pairings. An analysis of the disulphide folding pathways of other proteins such as potato carboxypeptidase inhibitor (Chang *et al.*, 1994), hirudin (e.g. Chatrenet and Chang, 1992; Chang, 1994; Chang *et al.*, 1995a) and epidermal growth factor (EGF: Chang *et al.*, 1995b) all adopt a folding mechanism which involves a sequential flow of unfolded (reduced) molecules through three groups of intermediates, namely one-, two-, and three-disulphide (scrambled) isomers. The scrambled three-disulphide isoforms then undergo rearrangement to the native isomer in the presence of a redox buffer. In these studies, many of the possible disulphide isoforms were detected indicating that the folding process

involving disulphide bonds adopts a process of trial and error in a search for the thermodynamic minimum. A particular property of the folding of potato carboxypeptidase inhibitor, hirudin and EGF is the presence of non-native disulphide structures. For EGF folding only 25 % of the well-populated folding intermediates are native-like with the remainder being non-native-like. However, these well-populated intermediates do not necessarily represent the *productive* intermediates that specify the folding pathway (Chang *et al.*, 1995b). This observation is in contrast to that observed for the oxidative folding of BPTI where the well-populated folding intermediates are native-like (Weissman and Kim, 1991). However, a more recent study of the folding pathway of BPTI (Darby *et al.*, 1995) suggests that non-native structures can be productive in the folding pathway.

The analysis of the *in vitro* oxidative folding of Long-[Arg^3]IGF-I (and IGF-I) appears to have many similarities to the folding of EGF (Chang *et al.*, 1995b). A striking similarity is the appearance of many intermediates early in folding which consist of native- and non-native-like disulphide isomers. Some intermediates are well-populated, and after one minute of folding, up to five three-disulphide isomers may be present, although it is conceivable that this is an over-estimate due to the nature of the analysis (Section 6.4). At least one productive Long-[Arg^3]IGF-I folding intermediate has been identified. A first analysis suggests that it is peak 2 which is a native-like two-disulphide isomer containing Cys^{18} - Cys^{61} and Cys^6 - Cys^{48} disulphide bonds; it is also well-populated. This intermediate appears to undergo transformation to the native structure in a rate-limiting step. A stop / go analysis, in which the acid-trapped, purified and characterized intermediate is rapidly put back into refolding conditions, reveals that the native-like two-disulphide isomer is productive with the rapid appearance of equilibrium folding products (Figure 6.7a).

The kinetics of this reaction are much faster than when Long-[Arg³]IGF-I is refolded from the fully reduced species. However, one intermediate, designated as Peak 4, which contained a Cys^{18} - Cys^{61} disulphide bond and a second disulphide bond, presumably Cys^{48} - Cys^{52}, folded to the equilibrium distribution of products much faster than for peak 2 in the stop / go experiment (Figure 6.7b). This intermediate was not well-populated at any time during folding but nevertheless appears to be a productive species. The involvement of this intermediate therefore raises the possibility that the sequential flow of the Cys^{18} - Cys^{61} intermediate to the native isomer via the Cys^{18} - Cys^{61}, Cys^{6} - Cys^{48} intermediate may not adequately describe the folding pathway. Peak 4 may provide an alternative pathway to the native isomer via the two-disulphide Cys^{18} - Cys^{61}, Cys^{47} - Cys^{52} intermediate. The fast kinetics of peak 4 in the stop / go experiment raises the possibility that the Cys^{18} - Cys^{61}, Cys^{6} - Cys^{48} intermediate may not directly form native IGF-I but may unfold to the single-disulphide Cys^{18} - Cys^{61} intermediate before progressive folding via the Cys^{18} - Cys^{61}, Cys^{47} - Cys^{52} intermediate. This implies that peak 4 is the productive intermediate and that peak 2 represents an unproductive but kinetically accessible species. If the Cys^{47} - Cys^{52} disulphide bond is strained as has been suggested for IGF-I (Hober *et al.*, 1992) then the formation of the Cys^{47} - Cys^{52} disulphide bond may be easier if the Cys^{6} and Cys^{48} are present as free thiols. The observation that two other acid-trapped intermediates (peaks 9 and 11) have similar rearrangement kinetics to peak 2, lends support to the notion that peak 4 is the productive intermediate. Clearly much more detailed work needs to be carried out on the kinetics of stop / go experiments of intermediates present during the folding of Long-[Arg³]IGF-I, though this is a daunting task as large amounts of these intermediates are required for a complete analysis. As has been suggested by Creighton (1992),

Goldenberg (1992) and Chang *et al.* (Chang *et al.*, 1995b), one must be cautious when equating intermediates that are well-populated with intermediates that are productive even when the well-populated intermediates are native-like.

A study of genetic models of IGF-I folding intermediates in which a number of Cys to Ala mutations were used to create native-like models of one- and two-disulphide folding intermediates indicated that IGF-I intermediates may be viewed as molten globules (Narhi *et al.*, 1993). The Cys^{18} - Cys^{61} disulphide model results in a partially folded state that has low but significant biological activity. The introduction of the Cys^{6} - Cys^{48} or Cys^{47} - Cys^{52} disulphide bond provided additional but incomplete structural organization and biological activity. The thermodynamic stabilities of the models indicated that the Cys^{18} - Cys^{61}, Cys^{6} - Cys^{48} intermediate was more stable than the Cys^{18} - Cys^{61}, Cys^{47} - Cys^{52} model which had a stability similar to the Cys^{18} - Cys^{61} model. However, this type of analysis does not reveal which of the two two-disulphide intermediates may be preferred in the folding pathway.

Like scrambled EGF (Chang *et al.*, 1995b), the stable three-disulphide alternative isomer of Long-[Arg³]IGF-I or IGF-I is trapped unless a redox buffer (or at least a reducing agent) is present to facilitate the return to two- and one-disulphide species before refolding to the native form. From a thermodynamic point of view, the scrambled EGF can be considered as a structure that has a more advanced packing and lower free energy than that of a two-disulphide folding intermediate. Their conversion to the native structure through disulphide re-organization, does not necessarily require substantial unfolding of the compactness already attained (Chang *et al.*, 1995b). For the Long-[Arg³]IGF-I three-disulphide isomers present in the first few minutes of refolding, a similar principle may apply. However, the

alternative isomer containing Cys^{18} - Cys^{61}, Cys^{48} - Cys^{52}, Cys^6 - Cys^{47} disulphide bonds may be an exception with an increased stability due to a significantly different structure. The NMR investigation of the alternative isomer of IGF-I (Chapter 5), while incomplete, indicates that rather than attempt to maintain helix 2, the constraints placed on the structure by switching two disulphide bonds results in an alternative structure that has a stable extended conformation. The analysis of the folding reactions of [Gly^3]IGF-I and [Arg^3]IGF-I suggest that Glu^3 in the native molecule is responsible for the stability of the alternative isomer (Chapter 4). Indeed, the model of the alternative isomer (Chapter 5) suggests that Glu^3 and Arg^{56} are close enough to one another for a stabilising salt bridge to occur, thereby accounting for the stability of the alternative isomer. Thus, one amino acid sequence can 'code' for two folds with a similar thermodynamic stability. This property is not unique to IGFs. For example, the plasminogen activator inhibitor 1 (PAI-1: Section 1.2.2.7) has a stable latent structure and an active metastable structure, though the metastable form is only accessible kinetically. Another example is provided by a glutathione-S-transferase fusion protein with the N-terminal domain of the lymphocyte cell adhesion molecule CD2 (Murray et al., 1995) in which much of the protein folds to a monomeric component and approximately 15 % is present as a metastable dimeric form. It has been suggested (Hober et al., 1992; Miller et al., 1993) that the difference in the structures between the native and alternative isomers of IGF-I may have some biological significance. However, to date the role of the alternative isomer in vivo has not been demonstrated.

In view of the literature available for the folding of IGFs and the research presented in this thesis, the oxidative folding pathway of IGF-I may be described as one in which the molecule first adopts a molten globule type structure that allows

flexibility and low stability as disulphide bonds form. As progressively more disulphide bonds form, the flow of many intermediates follow a thermodynamically controlled path to the native structure. *In vitro*, there is clearly a divergence in the folding pathway attested by the presence of a stable alternative structure. However, the factors that stabilize the route to the alternative isomer are sensitive to amino acid changes and steric hindrance that results in a lower thermodynamic stability of folding products in the alternative pathway. *In vivo*, the oxidative folding pathway is probably driven by the same forces that occur *in vitro*. However, in the endoplasmic reticulum the cellular folding machinery may influence the folding pathway by destabilizing the intermediates that lead to the alternative isomer, possibly by preventing the contact of non-native regions. How this may occur is unknown, but further research on the uniquely sensitive IGF folding pathway *in vivo* could yield much about the endoplasmic reticulum folding processes.

BIBLIOGRAPHY

Akke, M. and Forsén, S. (1990). Protein stability and electrostatic interactions between solvent exposed charged side chains. Proteins: Struct. Funct. Genet. *8*, 23-29.

Anderson, D.E., Becktel, W.J., and Dahlquist, F.W. (1990). pH-induced denaturation of proteins - a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4-lysozyme. Biochemistry *29*, 2403-2408.

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

Axelsson, K., Johansson, S., Eketorp, G., Zazzi, H., Hemmendorf, B., and Gellerfors, P. (1992). Disulphide arrangement of human insulin-like growth factor I derived from yeast and plasma. Eur. J. Biochem. *206*, 987-994.

Bach, L.A., Hsieh, S., Sakano, K., Fujiwara, H., Perdue, J.F., and Rechler, M.M. (1993). Binding of mutants of human insulin-like growth factor II to insulin-like growth factor binding proteins 1-6. J. Biol. Chem. *268*, 9246-9254.

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. and Agard, D.A. (1994). Kinetics versus thermodynamics in protein folding. Biochemistry *33*, 7505-7509.

Baker, D., Silen, J.L., and Agard, D.A. (1992a). Protease pro region required for folding is a potent inhibitor of the mature enzyme. Proteins: Struct. Funct. Genet. *12*, 339-344.

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

Baldwin, R.L. (1989). How does protein folding get started? Trends Biochem. Sci. *14*, 291-294.

Baldwin, R.L. (1993). Pulsed H/D-exchange studies of folding intermediates. Curr. Opin. Struct. Biol. *3*, 84-91.

Baldwin, T.O., Ziegler, M.M., Chaffotte, A.F., and Goldberg, M.E. (1993). Contribution of folding steps involving the individual subunits of bacterial luciferase to the assembly of the active heterodimeric enzyme. J. Biol. Chem. *268*, 10766-10772.

Bardwell, J.C.A., McGovern, K., and Beckwith, J. (1991). Identification of a protein required for disulphide bond formation *in vivo*. Cell *67*, 581-589.

Bastiras, S. and Wallace, J.C. (1992). Equilibrium denaturation of recombinant porcine growth hormone. Biochemistry *31*, 9304-9309.

Baum, J., Dobson, C.M., Evans, P.A., and Hanley, C. (1989). Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig α-lactalbumin. Biochemistry *28*, 7-13.

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 analogues 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., Chicchi, G.G., Miller, R.E., and Cascieri, M.A. (1990). The roles of tyrosines 24,31, and 60 in the high affinity binding of insulin-like

growth factor-I to the type 1 insulin-like growth factor receptor. J. Biol. Chem. *265*, 15648-15652.

Bayne, M.L., Applebaum, J., Underwood, D., Chicchi, G.G., Green, B.G., Hayes, N.S., and Cascieri, M.A. (1988). 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.

Bennett, W.S. and Huber, R. (1984). Structural and functional aspects of domain motions in proteins. Crit. Rev. Biochem. *15*, 291-394.

Betz, S.F. (1993). Disulphide bonds and the stability of globular proteins. Protein Science *2*, 1551-1558.

Blodgett, J.K., Loudon, G.M., and Collins, K.D. (1985). Specific cleavage of peptides containing an aspartic acid (ß-hydroxamic acid) residue. J. Am. Chem. Soc. *107*, 4305-4313.

Blundell, T.L., Bedarkar, S., Rinderknecht, E., and Humbel, R.E. (1978). Insulin-like growth factor: A model for tertiary structure accounting for immunoreactivity and receptor binding. Proc. Natl. Acad. Sci. USA *75*, 180-184.

Bole, D.G., Hendershot, L.M., and Kearney, J.F. (1986). Posttranslational association of immunoglobulin heavy chain binding protein with nascent heavy chains in nonsecreting and secreting hybridomas. J. Cell Biol. *102*, 1558-1566.

Boniface, J.J. and Reichert, L.E., Jr. (1990). Evidence for a novel thioredoxin-like catalytic property of gonadotrophic hormones. Science *247*, 61-64.

Bornstein, P. and Balian, G. (1977). Cleavage at Asn-Gly bonds with hydroxylamine. Methods Enzymol. *47*, 132-145.

Brandts, J.F., Halvorson, H.R., and Brennan, M. (1975). Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerism of proline residues. Biochemistry *14*, 4953-4963.

Briggs, M.S. and Roder, H. (1992). Early hydrogen-bonding events in the folding reaction of ubiquitin. Proc. Natl. Acad. Sci. USA *89*, 2017-2021.

Buck, M.A., Olah, T.A., Weitzmann, C.J., and Cooperman, B.S. (1989). Protein estimation by the product of integrated peak area and flow rate. Anal. Biochem. *182*, 295-299.

Bulleid, N.J. (1993). Protein disulfide-isomerase: role in biosynthesis of secretory proteins. Adv. Protein Chem. *44*, 125-150.

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.

Cascieri, M.A., Chicchi, G.G., Applebaum, J., Green, B.G., Hayes, N.S., and Bayne, M.L. (1989). Structural analogues 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.

Chaffotte, A.F., Guillou, Y., and Goldberg, M.E. (1992a). Kinetic resolution of peptide bond and side-chain far-UV circular dichroism during the folding of hen egg white lysozyme. Biochemistry *31*, 9694-9702.

Chaffotte, A.F., Guillou, Y., and Goldberg, M.E. (1992b). Inclusion bodies of the thermophilic endoglucanase D from *Clostridium thermocellum* are made from the native enzyme that resists 8M urea. Eur. J. Biochem. *205*, 369-373.

Chang, J.-Y. (1993). Identification of productive folding intermediates which account for the flow of protein folding pathway. J. Biol. Chem. *268*, 4043-4049.

Chang, J.-Y. (1994). Controlling the speed of hirudin folding. Biochem. J. *300*, 643-650.

Chang, J.-Y., Canals, F., Schindler, P., Querol, E., and Avilés, F.X. (1994). The disulfide folding pathway of potato carboxypeptidase inhibitor. J. Biol. Chem. *269*, 22087-22094.

Chang, J.-Y., Schindler, P., and Chatrenet, B. (1995a). The disulfide structures of scrambled hirudins. J. Biol. Chem. *270*, 11992-11997.

Chang, J.-Y., Schindler, P., Ramseier, U., and Lai, P.-H. (1995b). The disulfide folding pathway of human epidermal growth factor. J. Biol. Chem. *270*, 9207-9216.

Chatrenet, B. and Chang, J.-Y. (1992). The folding of hirudin adopts a mechanism of trial and error. J. Biol. Chem. *267*, 3038-3043.

Chatrenet, B. and Chang, J.-Y. (1993). The disulfide folding pathway of hirudin elucidated by stop/go folding experiments. J. Biol. Chem. *268*, 20988-20996.

Chaudhuri, B., Helliwell, S.B., and Priestle, J.P. (1991). A Lys^{27}-to-Glu^{27} mutation in the human insulin-like growth factor-1 prevents disulphide linked dimerization and allows secretion of BiP when expressed in yeast. FEBS Lett. *294*, 213-216.

Chaudhuri, B. and Stephan, C. (1992). A modified Kex2 enzyme retained in the endoplasmic reticulum prevents disulphide-linked dimerisation of recombinant insulin-like growth factor-I secreted from yeast. FEBS Letters *304*, 41-45.

Chaudhuri, B., Stephan, C., Huijbregts, R.P.H., and Martens, G.J.M. (1995). The neuroendocrine protein 7B2 acts as a molecular chaperone in the *in vitro*

folding of human insulin-like growth factor-1 secreted from yeast. Biochem.

Biophys. Res. Commun. *211*, 417-425.

Chaudhuri, J.B. (1994). Refolding recombinant proteins: Process strategies and

novel approaches. Ann. NY Acad. Sci. *721*, 374-385.

Chothia, C. (1975). Structural invariants in protein folding. Nature *254*, 304-308.

Clarke, J. and Fersht, A.R. (1993). Engineered disulfide bonds as probes of the

folding pathway of barnase: Increasing the stability of proteins against the

rate of denaturation. Biochemistry *32*, 4322-4329.

Cleland, W.W. (1964). Dithiothreitol, a new protective reagent for SH groups.

Biochemistry *3*, 480-482.

Clemmons, D.R., Dehoff, M.L., Busby, W.H., Bayne, M.L., and Cascieri, M.A. (1992).

Competition for binding to insulin-like growth factor (IGF) binding protein-2, 3,

4, and 5 by the IGFs and IGF analogs. Endocrinology *131*, 890-895.

Cooke, R.M., Harvey, T.S., and Campbell, I.D. (1991). Solution structure of insulin-

like growth factor I: a nuclear magnetic resonance and restrained molecular

dynamics study. Biochemistry *30*, 5484-5491.

Creighton, T.E. (1986). Disulphide bonds as probes of protein folding pathways.

Methods Enzymol. *131*, 83-106.

Creighton, T.E. (1990). Protein folding. Biochem. J. *270*, 1-16.

Creighton, T.E. (1991). Stability of folded conformations. Curr. Opin. Struct. Biol. *1*,

5-16.

Creighton, T.E. (1992). The disulphide folding pathway of BPTI. Science *256*, 111-

112.

Creighton, T.E. (1994). The protein folding problem. In *Mechanisms of protein

folding*. R.H. Pain, ed. (Oxford, New York, Tokyo: IRL Press), pp. 1-25.

Creighton, T.E., Bagley, C.J., Cooper, L., Darby, N.J., Freedman, R.B., Kemmink, J., and Sheikh, A. (1993). On the biosynthesis of bovine pancreatic trypsin inhibitor (BPTI). Structure, processing, folding and disulphide bond formation of the precursor *in vitro* and in microsomes. J. Mol. Biol. *232*, 1176-1196.

Creighton, T.E. and Ewbank, J.J. (1994). Disulfide-rearranged molten globule state of α-lactalbumin. Biochemistry *33*, 1534-1538.

Creighton, T.E. and Freedman, R.B. (1980). Catalysis by protein-disulphide isomerase of the unfolding and refolding of proteins with disulphide bonds. J. Mol. Biol. *142*, 43-62.

Csorba, T.R. (1991). Proinsulin: biosynthesis, conversion, assay methods and clinical studies. Clin. Biochem. *24*, 447-454.

Dao-pin, S., Söderlind, E., Baase, W.A., Wozniak, J.A., Sauer, U., and Matthews, B.W. (1991). Cumulative site-directed charge-change replacements in bacteriophage T_4 lysozyme suggest that long-range electrostatic interactions contribute little to protein stability. J. Mol. Biol. *221*, 873-887.

Darby, N.J., Morin, P.E., Talbo, G., and Creighton, T.E. (1995). Refolding of bovine pancreatic trypsin inhibitor *via* non-native disulphide intermediates. J. Mol. Biol. *249*, 463-477.

Darby, N.J., Van Mierlo, C.P.M., and Creighton, T.E. (1991). The 5-55 single-disulphide intermediate in folding of bovine pancreatic trypsin inhibitor. FEBS Lett. *279*, 61-64.

De Bernardez-Clarke, E. and Georgiou, G. (1991). Inclusion bodies and recovery of proteins from the aggregated state. In *Protein refolding*. G. Georgiou and E. De Bernardez-Clarke, eds. (Washington DC: American Chemical Society), pp. 1-20.

Dill, K. (1985). Theory for the folding and stability of globular proteins. Biochemistry *24*, 1501-1509.

Dill, K.A. (1990). Dominant forces in protein folding. Biochemistry *29*, 7133-7155.

Dobson, C.M. (1992). Unfolded proteins, compact states, and molten globules. Curr. Opin. Struct. Biol. *2*, 6-12.

Doig, A.J. and Williams, D.H. (1991). Is the hydrophobic effect stabilizing or destabilizing in proteins? The contribution of disulphide bonds to protein stability. J. Mol. Biol. *217*, 389-398.

Doonan, S., Garman, A.J., Hanson, J.M., Loudon, A.G., and Vernon, C.A. (1978). Identification by mass spectrometry of N'-formyl-lysine residues in a peptide from bee venom. J. Chem. Soc. Perkins Trans. I 1157-1160.

Dufour, E., Genot, C., and Haertlé, T. (1994). ß-Lactoglobulin binding properties during its folding changes studied by fluorescence spectroscopy. Biochimica et Biophysica Acta *1205*, 105-112.

Eder, J., Rheinnecker, M., and Fersht, A.R. (1993). Folding of subtilisin BPN': Characterization of a folding intermediate. Biochemistry *32*, 18-26.

Edman, J.C., Ellis, L., Blacher, R.W., Roth, R.A., and Rutter, W.J. (1985). Sequence of protein disulphide isomerase and implications of its relationship to thioredoxin. Nature *317*, 267-270.

Eisenberg, D. and McLachlan, A.D. (1986). Solvation energy in protein folding and binding. Nature *319*, 199-203.

Elliot, 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 Arakava, T. (1990). Yeast derived recombinant human insulin-like growth factor I:

production, purification and structural characterization. J. Protein Chem. *9*, 95-104.

Ellis, J. (1987). Proteins as molecular chaperones. Nature *328*, 378-379.

Elöve, G.A., Chaffotte, A.F., Roder, H., and Goldberg, M.E. (1992). Early steps in cytochrome *c* folding probed by time-resolved circular dichroism and fluorescence spectroscopy. Biochemistry *31*, 6786-6883.

Enfors, S.-O. (1992). Control of *in vivo* proteolysis in the production of recombinant proteins. TIBTECH *10*, 310-315.

Englander, S.W. (1993). In pursuit of protein folding. Science *262*, 848-849.

Epstein, C.J., Goldberger, R.F., and Anfinsen, C.B. (1963). The genetic control of tertiary protein structure: studies with model systems. Cold Spring Harbor Symp. Quant. Biol. *28*, 439-449.

Evans, P.A. and Radford, S.E. (1994). Probing the structure of folding intermediates. Curr. Opin. Struct. Biol. *4*, 100-106.

Ewbank, J.J. and Creighton, T.E. (1993a). Pathway of disulphide coupled unfolding and refolding of bovine α-lactalbumin. Biochemistry *32*, 3677-3693.

Ewbank, J.J. and Creighton, T.E. (1993b). Structural characterization of the disulphide folding intermediate of bovine α-lactalbumin. Biochemistry *32*, 3694-3707.

Farquhar, R., Honey, N., Murant, S.J., Bossier, P., Schultz, L., Montgomery, D., Ellis, R.W., Freedman, R.B., and Tuite, M.F. (1991). Protein disulphide isomerase is essential for viability in *Saccharomyces cerevisiae* . Gene *108*, 81-89.

Fersht, A.R. (1972). Conformational equilibria in α- and δ-chymotrypsin: The energetics and importance of the salt bridge. J. Mol. Biol. *64*, 497-509.

Fersht, A.R. (1994). Pathway and stability of protein folding. Biochem. Soc. Trans. *22*, 267-274.

Fischer, B., Sumner, I., and Goodenough, P. (1993). Isolation, renaturation, and formation of disulphide bonds of eukaryotic proteins expressed in *Escherichia coli* as inclusion bodies. Biotechnol. Bioeng. *41*, 3-13.

Flory, P.J. (1956). Theory of elastic mechanisms in fibrous proteins. J. Am. Chem. Soc. *78*, 5222-5235.

Forsberg, G., Palm, G., Ekebacke, A., Josephson, S., and Hartmanis, M. (1990). Separation and characterization of modified variants of 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 colostrum. 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. Endo. *8*, 213-223.

Francis, G.L., Upton, F.M., Ballard, F.J., McNeil, K.A., and Wallace, J.C. (1988). Insulin-like growth factors 1 and 2 in bovine colostrum. Biochem. J. *251*, 95-103.

Franke, A.E., Danley, D.E., Kaczmarek, F.S., Hawrylik, S.J., Gerard, R.D., Lee, S.D., and Geoghegan, K.F. (1990). Expression of human plasminogen activator inibitor type-1 (PAI-1) in *Escherichia coli* as a soluble protein comprised of active and latent forms. Isolation and crystallization of latent PAI-1. Biochim. Biophys. Acta *1037*, 16-23.

Frech, C. and Schmid, F.X. (1995). DsbA-mediated disulfide bond formation and catalyzed prolyl isomerization in oxidative protein folding. J. Biol. Chem. *270*, 5367-5374.

Freedman, R.B., Hirst, T.R., and Tuite, M.F. (1994). Protein disulphide isomerase: Building bridges in protein folding. Trends Biochem. Sci. *19*, 331-336.

Freskgård, P.-O., Bergenhem, N., Jonsson, B.-H., Svensson, M., and Carlsson, U. (1992). Isomerase and chaperone activity of prolyl isomerase in the folding of carbonic anhydrase. Science *258*, 466-468.

Fukuda, R., Horiuchi, H., Ohta, A., and Takagi, M. (1994). The prosequence of *Rhizopus niveus* aspartic proteinase-I supports correct folding and secretion of its mature part in *Saccharomyces cerevisiae*. J. Biol. Chem. *269*, 9556-9561.

Geiger, T. and Clarke, S. (1987). Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. J. Biol. Chem. *262*, 785-794.

Georgopoulos, C. and Welch, W.J. (1993). Role of the major heat shock proteins as molecular chaperones. Annu. Rev. Cell Biol. *9*, 601-634.

Gething, M.-J. and Sambrook, J. (1992). Protein folding in the cell. Nature *355*, 33-45.

Gilbert, H.F. (1994). The formation of native disulphide bonds. In *Mechanisms of Protein Folding.* R.H. Pain, ed. (Oxford, New York, Tokyo: IRL Press), pp. 104-136.

Gilbert, H.F., McLean, V., and McLean, M. (1990). Molecular and cellular aspects of thiol-disulphide exchange. In *Advances in Enzymology. 63.* (Wiley), pp. 69-172.

Goldberg, M.E. and Guillou, Y. (1994). Native disulfide bonds greatly accelerate secondary structure formation in the folding of lysozyme. Protein Sci. *3*, 883-887.

Goldberger, R.F., Epstein, C.J., and Anfinsen, C.B. (1963). Acceleration of the reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver. J. Biol. Chem. *238*, 628-635.

Goldenberg, D.P. (1988). Kinetic analysis of the folding and unfolding of a mutant form of bovine pancreatic trypsin inhibitor lacking the cysteine-14 and -38 thiols. Biochemistry *27*, 2481-2489.

Goldenberg, D. P. (1992). Native and non-native intermediates in the BPTI folding pathway. TIBS *17*, 257-261.

Goto, Y., Calciano, L.J., and Fink, A.L. (1990). Acid-induced folding of proteins. Proc. Natl. Acad. Sci. USA *87*, 573-577.

Gray, A.L. and Mason, A.J. (1990). Requirement for activin A and transforming growth factor-β1 pro-regions in homodimer assembly. Science *247*, 1328-1330.

Gray, T.E., Eder, J., Bycroft, M., Day, A.G., and Fersht, A.R. (1993). Refolding of barnase mutants and pro-barnase in the presence and absence of GroEL. EMBO J. *12*, 4145-4150.

Gregoret, L.M. and Cohen, F.E. (1991). Protein folding. Effect of packing density on chain conformation. J. Mol. Biol. *219*, 109-122.

Haas, I.G. and Wabl, M. (1983). Innunoglobulin heavy chain binding protein. Nature *306*, 387-389.

Hagihara, Y., Aimoto, S., Fink, A.L., and Goto, Y. (1993). Guanidine hydrochloride-induced folding of proteins. J. Mol. Biol. *231*, 180-184.

Harisch, G. and Mahmoud, M.F. (1980). The glutathione status in the liver and cardiac muscle of rats after starvation. Hoppe-Seyer's Z. Physiol. Chem *361*, 1859-1862.

Harrison, S.C. and Durbin, R. (1985). Is there a single pathway for the folding of a polypeptide chain? Proc. Natl. Acad. Sci. USA *82*, 4028-4030.

Hawkins, H.C. and Freedman, R.B. (1991). The reactivities and ionization properties of the active-site dithiol groups of mammalian protein disulphide-isomerase. Biochem. J. *275*, 335-339.

Heckman, C.M. and Loskutoff, D.J. (1985). Endothelial cells produce a latent inhibitor of plasminogen activators that can be activated by denaturants. J. Biol. Chem. *260*, 11581-11587.

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

Hirel, P.-H., Schmitter, J.-M., Dessen, P., Fayat, G., and Blanquet, S. (1989). Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed

by the side-chain length of the penultimate amino acid. Proc. Natl. Acad. Sci. USA *86*, 8247-8251.

Hober, S., Forsberg, G., Palm, G., Hartmanis, M., and Nilsson, B. (1992). Disulphide exchange folding of insulin-like growth factor I. Biochemistry *31*, 1749-1756.

Hober, S., Hansson, A., Uhlén, M., and Nilsson, B. (1994). Folding of insulin-like growth factor I is thermodynamically controlled by insulin-like growth factor binding protein. Biochemistry *33*, 6758-6761.

Hodgkinson, S.C., Napier, J.R., Davis, S.R., Patel, B., and Gluckman, P.D. (1989). Binding protein, radioreceptor and biological activities of recombinant methionyl insulin-like growth factor-I variants. Mol. Cell Endocrinol. *66*, 37-44.

Hooke, S.D., Radford, S.E., and Dobson, C.M. (1994). The refolding of human lysozyme: A comparison with the structurally homologous hen lysozyme. Biochemistry *33*, 5867-5876.

Hughson, F.M., Wright, P.E., and Baldwin, R.L. (1990). Structural characterization of a partly folded apomyoglobin intermediate. Science *249*, 1544-1548.

Hunkapillar, M.W., Hewick, R.M., Dreyer, W.J., and Hood, L.E. (1983). High-sensitivity sequencing with a gas-phase sequenator. Methods Enzymol. *91*, 399-413.

Huth, J.R., Perini, F., Lockridge, O., Bedows, E., and Ruddon, R.W. (1993). Protein folding and assembly *in vitro* parallel intracellular folding and assembly. Catalysis of folding and assembly of the human chorionic gonadotropin αβ dimer by protein disulfide isomerase. J. Biol. Chem. *268*, 16472-16482.

Hwang, C., Sinskey, A.J., and Lodish, H.F. (1992). Oxidized redox state of glutathione in the endoplasmic reticulum. Science *257*, 1496-1502.

Igekuchi, M., Sugai, S., Fujino, M., Sugawara, T., and Kuwajima, K. (1992). Contribution of the 6-120 disulphide bond of α-lactalbumin to the stabilities of its native and molten globule states. Biochemistry *31*, 12695-12700.

Ikemura, H., Takagi, H., and Inouye, M. (1987). Requirement of the pro-sequence for the production of active subtilisin E in *Escherichia coli*. J. Biol. Chem. *262*, 7859-7864.

Ishikawa, T. and Sies, H. (1984). Cardiac transport of glutathione disulfide and S-conjugate. J. Biol. Chem. *259*, 3838-3843.

Itzhaki, L.S., Evans, P.A., Dobson, C.M., and Radford, S.E. (1994). Tertiary interactions in the folding pathway of hen lysozyme: Kinetic studies using fluorescent probes. Biochemistry *33*, 5212-5220.

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

Jaenicke, R. (1991). Protein stability and molecular adaptation to extreme conditions. Eur. J. Biochem. *202*, 715-728.

Jennings, P.A. and Wright, P.E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. Science *262*, 892-896.

Joly, J.C. and Swartz, J.R. (1994). Protein folding activities of *Escherichia coli* protein disulfide isomerase. Biochemistry *33*, 4231-4236.

Kamoun, P.P. (1988). Denaturation of globular proteins by urea: breakdown of hydrogen or hydrophobic bonds? Trends Biochem. Sci. *13*, 424-425.

Kane, J.F. and Hartley, D.L. (1988). Formation of recombinant protein inclusion bodies in *Escherichia coli*. TIBTECH *6*, 95-101.

Karplus, M. and McCammon, J.A. (1983). Dynamics of proteins: elements and function. Annu. Rev. Biochem. *53*, 263-300.

Karplus, M. and Weaver, D.L. (1979). Diffusion-collision model for protein folding. Biopolymers *18*, 1421-1437.

Kassenbrock, C.K., Garcia, P.D., Walter, P., and Kelly, R.B. (1988). Heavy chain binding protein recognizes aberrant polypeptides translocated *in vitro*. Nature *333*, 90-93.

Katagiri, K., Okada, K., Hattori, H., and Yano, M. (1988). Bovine endothelial cell plasminogen activator inhibitor. Eur. J. Biochem. *176*, 81-87.

Khorasanizadeh, S., Peters, I.D., Butt, T.R., and Roder, H. (1993). Folding and stability of a tryptophan-containing mutant of ubiquitin. Biochemistry *32*, 7054-7063.

Kiefhaber, T., Grunert, H.-P., Hahn, U., and Schmid, F.X. (1992a). Folding of RNase T_1 is decelerated by a specific tertiary contact in a folding intermediate. Proteins: Struct. Funct. Genet. *12*, 171-179.

Kiefhaber, T., Schmid, F.X., Willaert, K., Engelborghs, Y., and Chaffotte, A. (1992b). Structure of a rapidly formed intermediate in ribonuclease T_1 folding. Protein Sci. *1*, 1162-1172.

Kim, P.S. and Arvan, P. (1995). Calnexin and BiP act as sequential molecular chaperones during thyroglobulin folding in the endoplasmic reticulum. J. Cell Biol. *128*, 29-38.

Kim, P.S. and Baldwin, R.L. (1990). Intermediates in the reactions of small proteins. Annu. Rev. Biochem. *59*, 631-660.

King, R., Wells, J.R.E., Kreig, 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. Endo. *8*, 29-41.

Koedam, J.A. and Van den Brande, J.L. (1994). Insulin-like growth factors (IGFs) and IGF binding protein-3 display disulfide isomerase activity. Biochem. Biophys. Res. Commun. *198*, 1225-1231.

Kotlarski, N., Yeates, R.S., Milner, S.J., Francis, G.L., O'Neill, B.K., and Middleberg, A.P.J. (1995). Studies into the scale-up process to produce a biosynthetic insulin-like growth factor. Trans IChemE *73*, 27-32.

Kozutsumi, Y., Segal, M., Normington, K., Gething, M.-J., and Sambrook, J. (1988). The presence of malfolded proteins in the endoplasmic reticulum signals the induction of glucose-regulated proteins. Nature *332*, 462-464.

Krause, G., Lundstrom, J., Barea, J.L., de la Cuesta, C.P., and Holmgren, A. (1991). Mimiking the active site of protein disulphide isomerase by substitution of proline 34 in *Escherichia coli* thioredoxin. J. Biol. Chem. *266*, 9494-9500.

Kuroki, R., Inaka, K., Taniyama, Y., Kidokoro, S., Matsushima, M., Kikuchi, M., and Yutani, K. (1992). Enthalpic destabilization of a mutant human lysozyme lacking a disulphide bridge between cysteine-77 and cysteine-95. Biochemistry *31*, 8323-8328.

Kuwajima, K. (1989). The Molten Globule State as a Clue for Understanding the folding and cooperativity of globular-protein stucture. Proteins *6*, 87-103.

Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S., and Nagamura, T. (1987). Rapid formation of secondary structure framework in protein folding studied by stopped flow circular dichroism. FEBS Lett. *221*, 115-118.

Labhardt, A.M. (1984). Kinetic circular dichroism shows that the S-peptide α-helix of ribonuclease S unfolds fast and refolds slowly. Proc. Natl. Acad. Sci. USA *81*, 7674-7678.

Lambert, N. and Freedman, R.B. (1983). Structural properties of homogenous protein disulphide-isomerase from bovine liver purified by a rapid high-yielding procedure. Biochem. J. *213*, 225-234.

Lash, L.H. and Jones, D.P. (1985). Distribution of oxidized and reduced forms of glutathione and cysteine in rat plasma. Arch. Biochem. Biophys. *240*, 583-592.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. *26*, 283-291.

Le-Nguyen, D., Heitz, A., Chiche, L., El Hajji, M., and Castro, B. (1993). Characterization and 2D NMR study of the stable [9-21, 15-27] 2 disulphide intermediate in the folding of the 3 disulphide trypsin inhibitor EETI II. Protein Sci. *2*, 165-174.

Lees, W.J. and Whitesides, G.M. (1993). Equilibrium constants for thiol-disulfide interchange reactions: A coherent, corrected set. J. Org. Chem. *58*, 642-647.

Levinthal, C. (1968). Are there pathways for protein folding? J. Chim. Phys. *65*, 44-45.

Levinthal, C. (1969). Proceedings of a Meeting held at Allerton House, Monticello, IL. In *Mossbauer Spectroscopy in Biological Systems*. P. Debrunner, J.C.M. Tsibris, and E. Münck, eds. (Urbana: University of Illinois Press), pp. 22-24.

Lin, T.-Y. and Kim, P.S. (1991). Evaluating the effects of a single amino acid substitution on both the native and denatured states of a protein. Proc. Natl. Acad. Sci. USA *88*, 10573-10577.

Livingstone, J.R., Spolar, R.S., and Record, M.T., Jr. (1991). Contribution to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area. Biochemistry *30*, 4237-4244.

Lowman, H.B. and Wells, J.A. (1993). Affinity maturation of human growth hormone by monovalent phage display. J. Mol. Biol. *234*, 564-578.

Lu, X., Gilbert, H.F., and Harper, J.W. (1992). Conserved residues flanking the thiol/disulphide centres of protein disulphide isomerase are not essential for catalysis of thiol/disulphide exchange. Biochemistry *31*, 4205-4210.

Lundström, J. and Holmgren, A. (1993). Determination of the reduction-oxidation potential of the thioredoxin-like domains of protein disulfide-isomerase from the equilibrium with glutathione and thioredoxin. Biochemistry *32*, 6649-6655.

Lyles, M.M. and Gilbert, H.F. (1991a). Catalysis of the oxidative folding of ribonuclease A by protein disulphide isomerase: dependence of the rate on the composition of the redox buffer. Biochemistry *30*, 613-619.

Lyles, M.M. and Gilbert, H.F. (1991b). Catalysis of the oxidative folding of ribonuclease A by protein disulphide isomerase: pre-steady-state kinetics and the utilization of the oxidizing equilivalents of the isomerase. Biochemistry *30*, 619-625.

Maly, P. and Lüthi, C. (1988). The binding sites of insulin-like growth factor I (IGF I) to type I IGF receptor and to a monoclonal antibody. J. Biol. Chem. *263*, 7068-7072.

Mann, C.J. and Matthews, C.R. (1993). Structure and stability of an early folding intermediate of *Escherichia coli trp* aporepressor measured by far-UV stopped-flow circular dichroism and 8-anilino-1-naphthalene sulfonate binding. Biochemistry *32*, 5282-5290.

Marston, F.A.O. (1986). The purification of eukaryotic polypeptides synthesized in *Escherichia coli.* Biochem. J. *240*, 1-12.

Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A.L., and Hartl, F.-U. (1991). Chaperonin-mediated protein folding at the surface of groEL through a 'molten globule'-like intermediate. Nature *352*, 36-42.

Matthews, B.W. (1993). Structural and genetic analysis of protein folding and stability. Curr. Opin. Struct. Biol. *3*, 589-593.

Matthews, C.R. (1993). Pathways of protein folding. Annu. Rev. Biochem. *62*, 653-683.

Mayr, L.M., Willbold, D., Landt, O., and Schmid, F.X. (1994). Role of the Cys 2-Cys 10 disulfide bond for the structure, stability, and folding kinetics of ribonuclease T_1. Protein Sci. *3*, 227-239.

Mazzarella, R.A., Srinivasan, M., Haugejorden, S.M., and Green, M. (1990). Erp72, an abundant luminal endoplasmic reticulum protein, contains three copies of the active site sequences of protein disulphide isomerase. J. Biol. Chem. *256*, 1094-1101.

McCusker, R.H. and Clemmons, D.R. (1994). Effects of cytokines in insulin-like growth factor-binding protein secretion by muscle cells *in vitro.* Endocrinology *134*, 2095-2102.

Merril, C.R., Goldman, D., and Van Keuren, M.L. (1984). Gel protein stains: silver stain. Methods Enzymol. *104*, 441-447.

Messing, J. (1979). A multipurpose cloning system based on the single-stranded dna bacteriophage M13. *2*, 43-48.

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.

Miranker, A., Robinson, C.V., Radford, S.E., Aplin, R.T., and Dobson, C.M. (1993). Detection of transient protein folding populations by mass spectrometry. Science *262*, 896-900.

Morris, H.R., Pucci, P., Panico, M., and Marino, G. (1990). Protein folding / refolding analysis by mass spectrometry. Biochem. J. *268*, 803-806.

Murphy, K.P. and Gill, S.J. (1991). Small model compounds and the thermodynamics of protein unfolding. J. Mol. Biol. *222*, 699-709.

Murray, A.J., Lewis, S.J., Barclay, A.N., and Brady, R.L. (1995). One sequence, two folds: A metastable structure of CD2. Proc. Natl. Acad. Sci. USA *92*, 7337-7341.

Murzin, A.G., Brenner, S.E., Hibbard, T., and Chothia, C. (1995). SCOP: A structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. *247*, 536-540.

Nagata, K., Maruyama, K., Nagasawa, H., Urushibata, I., Isogai, A., Ishizaki, H., and Suzuki, A. (1992). Bombyxin-II and its disulphide bond isomers: synthesis and activity. Peptides *13*, 653-662.

Nakano, T., Antonino, L.C., Fox, R.O., and Fink, A.L. (1993). Effect of proline mutations on the stability and kinetics of folding of staphylococcal nuclease. Biochemistry *32*, 2534-2541.

Narhi, L.O., Hua, Q.-X., Arakawa, T., Fox, G.M., Tsai, L., Rosenfeld, R., Holst, P., Miller, J.A., and Weiss, M.A. (1993). Role of native disulfide bonds in the structure and activity of insulin-like growth factor 1: Genetic models of protein-folding intermediates. Biochemistry *32*, 5214-5221.

Neri, D., Billeter, M., Wider, G., and Wüthrich, K. (1992). NMR determination of residual structure in a urea-denatured protein, the 434 repressor. Science *257*, 1559-1563.

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.

Noiva, R. and Lennarz, W.J. (1992). Protein disulphide isomerase: a multifunctional protein resident in the lumen of the endoplasmic reticulum. J. Biol. Chem. *267*, 3553-3556.

Ohnishi, Y., Nishiyama, M., Horinouchi, S., and Beppu, T. (1994). Involvement of the COOH-terminal pro-sequence of *Serratia marcescens* serine protease in the folding of the mature enzyme. J. Biol. Chem. *269*, 32800-32806.

Ostermeier, M. and Georgiou, G. (1994). The folding of bovine pancreatic trypsin inhibitor in the *Escherichia coli* periplasm. J. Biol. Chem. *269*, 21072-21077.

Otsu, M., Omura, F., Yoshimori, T., and Kikuchi, M. (1994). Protein disulfide isomerase associates with misfolded human lysozyme *in vivo*. J. Biol. Chem. *269*, 6874-6877.

Pace, C.N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. Methods Enzymol. *131*, 266-280.

Pace, C.N. (1990). Conformation stability of globular proteins. Trends Biochem. Sci. *15*, 14-17.

Pace, C.N., Grimsley, G.R., Thomson, J.A., and Barnett, B.J. (1988). Conformational stability and activity of ribonuclease T_1 with zero, one, and two intact disulphide bonds. J. Biol. Chem. *263*, 11820-11825.

Plaza del Pino, I.M., Pace, C.N., and Freire, E. (1992). Temperature and guanidine hydrochloride dependance of the structural stability of ribonuclease T_1. Biochemistry *31*, 11196-11202.

Pouysségur, J., Shiu, R.P.C., and Pastan, I. (1977). Induction of two transformation sensitive membrane polypeptides in normal fibroblasts by a block in glycoprotein synthesis or glucose deprivation. Cell *11*, 941-947.

Privalov, P.L. (1979). Stability of proteins. Small globular proteins. Adv. Prot. Chem. *33*, 167-245.

Privalov, P.L. and Gill, S.J. (1989). The hydrophobic effect: a reappraisal. Pure Appl. Chem. *61*, 1097-1104.

Privalov, P.L. and Makhatadze, G.I. (1992). Contribution of hydration and non-covalent interactions to the heat capacity effect on protein unfolding. J. Mol. Biol. *224*, 715-723.

Ptitsyn, O.B., Pain, R.H., Semisotnov, G.V., Zerovnik, E., and Razgulyaev, O.I. (1990). Evidence for a molten globule state as an intermediate in protein folding. FEBS Lett. *262*, 20-24.

Puig, A. and Gilbert, H.F. (1994). Protein disulfide isomerase exhibits chaperone and anti-chaperone activity in the oxidative refolding of lysozyme. J. Biol. Chem. *269*, 7764-7771.

Pyrse, K.M., Bruckman, T.G., Maxfield, B.W., and Elson, E.L. (1992). Kinetics and mechanism of the folding of cytochrome *c*. Biochemistry *31*, 5127-5136.

Radford, S.E., Dobson, C.M., and Evans, P.A. (1992). The folding of hen lysozyme involves partially structured intermediates and multiple pathways. Nature *358*, 302-307.

Ramachandran, L.K. and Narita, K. (1958). Reactions involving the amide and carboxyl groups of tobacco mosaic virus (TMV) protein. Biochimica et Biophysica Acta *30*, 616-624.

Raschdorf, F., Dahinden, R., Maerki, W., Richter, W.J., and Merryweather, J.P. (1988). Location of disulphide bonds in human insulin-like growth factors (IGFs) synthesized by recombinant DNA technology. Biomed. Mass Spectrom. *16*, 3-8.

Richards, F.M. (1977). Area, volumes, packing, and protein structure. Annu. Rev. Biophys. Bioeng. *6*, 151-176.

Robertson, A.D. and Baldwin, R.L. (1991). Hydrogen exchange in thermally denatured ribonuclease A. Biochemistry *30*, 9907-9914.

Rock, F.L., Li, X., Chong, P., Ida, N., and Klein, M. (1994). Roles of disulfide bonds in recombinant human interleukin 6 conformation. Biochemistry *33*, 5146-5154.

Roder, H., Elöve, G.A., and Englander, W. (1988). Structural characterization of folding intermediates in cytochrome *c* by H-exchange labelling and proton NMR. Nature *335*, 700-704.

Rose, G.D. and Wolfenden, R. (1993). Hydrogen bonding, hydrophobicity, packing, and protein folding. Annu. Rev. Biophys. Biomol. Struct. *22*, 381-415.

Rose, M.D., Misra, L.M., and Vogel, J.P. (1989). KAR2, a karogamy gene, is the yeast homologue of BiP/GRP78 gene. Cell *57*, 1211-1221.

Rosenfeld, R.D., Noone, N.M., Lauren, S.L., Rohde, M.F., Narhi, L.O., and Arakawa, T. (1993). Mutation of Arg55/56 to Leu55/Ala56 in insulin-like growth factor-I results in two forms different in disulfide structure and native conformation but similar under reverse-phase conditions. J. Protein Chem. *12*, 247-254.

Ross, M., Francis, G.L., Szabo, L., Wallace, J.C., and Ballard, F.J. (1989). Insulin-like growth factor (IGF)-binding proteins inhibit the biological activities of IGF-1 and IGF-2 but not des-(1-3)-IGF-1. Biochem. J. *258*, 267-272.

Rothwarf, D.M. and Scheraga, H.A. (1992). Equilibrium and kinetic constants for the thiol-disulphide interchange reaction between glutathione and dithiothreitol. Proc. Natl. Acad. Sci. USA *89*, 7944-7948.

Rotwein, P. (1986). Two insulin-like growth factor i messenger RNAs are expressed in human liver. Proc. Natl. Acad. Sci. USA *83*, 77-81.

Ruoppolo, M. and Freedman, R.B. (1994). Protein-S-S-glutathione mixed disulfides as models of unfolded proteins. Biochemistry *33*, 7654-7662.

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.

Sali, A., Shakhnovich, E., and Karplus, M. (1994). Kinetics of protein folding. A lattice model study of the requirements for folding to the native state. J. Mol. Biol. *235*, 1614-1636.

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.

Samuelsson, E., Wadensten, H., Hartmanis, M., Moks, T., and Uhlén, M. (1991). Facillitated *in vitro* refolding of human recombinant insulin-like growth factor I using a solubilizing fusion partner. BioTechnology *9*, 363-366.

Sato, A., Nishimura, S., Ohkubo, T., Kyogoku, Y., Koyama, S., Kobayashi, M., Yasuda, T., and Kobayashi, Y. (1992). ^{1}H-NMR assignment and secondary structure of human insulin-like growth factor-I (IGF-I) in solution. J. Biochem. *111*, 529-536.

Sato, A., Nishimura, S., Ohkubo, T., Kyogoku, Y., Koyama, S., Kobayashi, M., Yasuda, T., and Kobayashi, Y. (1993). Three-dimensional structure of human insulin-like growth factor-I (IGF-I) determined by ^{1}H-NMR and distance geometry. Int. J. Pept. Protein Res. *41*, 433-440.

Saunders, A.J., Young, G.B., and Pielak, G.J. (1993). Polarity of disulphide bonds. Protein Science *2*, 1183-1184.

Saxena, V.P. and Wetlaufer, D.B. (1970). Formation of three-dimensional structure in proteins. I. Rapid non-enzymic reactivation of reduced lysozyme. Biochemistry *9*, 5015-5022.

Scherens, B., Dubois, E., and Messenguy, F. (1991). Determination of the sequence of the yeast *YCL313* gene localized on chromosome III. Homology with the

protein disulphide isomerase (PDI) gene product of other organisms. Yeast *7*, 185-193.

Schmid, F.X. (1993). Prolyl isomerase: Enzymatic catalysis of slow protein-folding reactions. Annu. Rev. Biophys. Biomol. Struct. *22*, 123-143.

Schönbrunner, E.R. and Schmid, F.X. (1992). Peptidyl-prolyl *cis-trans* isomerase improves the efficiency of protein disulphide isomerase as a catalyst of protein folding. Proc. Natl. Acad. Sci. USA *89*, 4510-4513.

Schwartz, H., Hinz, H., Mehlich, A., Tschesche, H., and Wenzel, H.R. (1987). Stability studies on derivatives of the bovine pancreatic trypsin inhibitor. Biochemistry *26*, 3544-3551.

Seckler, R. and Jaenicke, R. (1992). Protein folding and protein refolding. FASEB J. *6*, 2545-2552.

Sheehan, J.C. and Yang, D.-D.H. (1958). The use of N-formylamino acids in peptide synthesis. J. Am. Chem. Soc. *80*, 1154-1158.

Shin, H.-C., Merutka, G., Waltho, J.P., Wright, P.E., and Dyson, H.J. (1993a). Peptide models of protein folding initiation sites. 2. The G-H turn region of myoglobin acts as a helix stop signal. Biochemistry *32*, 6348-6355.

Shin, H.-C., Merutka, G., Waltho, J.P., Tennant, L.L., Dyson, H.J., and Wright, P.E. (1993b). Peptide models of protein folding initiation sites. 3. The G-H helical hairpin of myoglobin. Biochemistry *32*, 6356-6364.

Shin, S.Y., Shimizu, M., and Munekata, E. (1994). Synthesis of N-terminal truncated peptides of human epidermal growth factor (h-EGF). Chemistry Letters 1997-2000.

Shirley, B.A., Stanssens, P., Hahn, U., and Pace, C.N. (1992). Contribution of

hydrogen bonding to the conformational stability of RNase T$_1$. Biochemistry

31, 725-732.

Shortle, D. (1993). Denatured states of proteins and their roles in folding and

stability. Curr. Opin. Struct. Biol. *3*, 66-74.

Siedler, F., Rudolph-Böhner, S., Doi, M., Musiol, H.-J., and Moroder, L. (1993).

Redox potentials of active-site *bis*(cysteinyl) fragments of thiol protein

oxidoreductases. Biochemistry *32*, 7488-7495.

Silen, J.L. and Agard, D.A. (1989). The α-lytic protease pro-region does not require

a physical linkage to activate the protease domain *in vivo*. Nature *341*, 462-

464.

Silen, J.L., Frank, D., Fujishige, A., Bone, R., and Agard, D.A. (1989). Analysis of

prepro-α-lytic protease expression in *Escherichia coli* reveals that the pro

region is required for activity. J. Bacteriol. *171*, 1320-1325.

Silen, J.L., McGrath, C.N., Smith, K.R., and Agard, D.A. (1988). Molecular analysis

of the gene encoding α-lytic protease: evidence for a preproenzyme. Gene

69, 237-244.

Sorensen, P., Winther, J.R., Kaarsholm, N.C., and Poulsen, F.M. (1993). The pro

region required for folding of carboxypeptidase Y is a partially folded

domain with little regular structural core. Biochemistry *32*, 12160-12166.

Sproat, B.S. and Gait, M.J. (1985). Chemical synthesis of a gene for somatomedin

C. Nucleic Acids Research *13*, 2959-2977.

Srinivasan, N., Sowdhamini, R., Ramakrishnan, C., and Balaram, P. (1990).

Conformations of disulphide bridges in proteins. Int. J. Pept. Protein Res. *36*,

147-155.

Staley, J.P. and Kim, P.S. (1992). Complete folding of bovine pancreatic trypsin inhibitor with only a single disulphide bond. Proc. Natl. Acad. Sci. USA *89*, 1519-1523.

Staniforth, R.A., Burston, S.G., Smith, C.J., Jackson, G.S., Badcoe, I.G., Atkinson, T., Holbrook, J.J., and Clarke, A.R. (1993). The energetics and cooperativity of protein folding: A simple experimental analysis based upon the solvation of internal residues. Biochemistry *32*, 3842-3851.

Stein, P. and Chothia, C. (1991). Serpin tertiary structure transformation. J. Mol. Biol. *221*, 615-621.

Steiner, D.F. and Clark, J.L. (1968). The spontaneous reoxidation of reduced beef and rat proinsulins. Proc. Natl. Acad. Sci. USA *60*, 622-629.

Stephenson, R.C. and Clarke, S. (1989). Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins. J. Biol. Chem. *264*, 6164-6170.

Strausberg, S., Alexander, P., Wang, L., Schwarz, F., and Bryan, P. (1993). Catalysis of a protein folding reaction: Thermodynamic and kinetic analysis of subtilisin BPN' interactions with its propeptide fragment. Biochemistry *32*, 8112-8119.

Szabo, L., Mottershead, D.G., Ballard, F.J., and Wallace, J.C. (1988). The bovine insulin-like growth factor (IGF) binding protein purified from conditioned medium requires the N-terminal tripeptide in IGF-1 for binding. Biochem. Biophys. Res. Commun. *151*, 207-214.

Szajewski, R.P. and Whitesides, G.M. (1980). Rate constants and equilibrium constants for thiol-disulphide interchange reactions involving oxidized glutathione. J. Am. Chem. Soc. *102*, 2011-2026.

Tachibana, C. and Stevens, T.H. (1992). The yeast *EUG1* gene encodes an endoplasmic reticulum protein that is functionally related to protein disulphide isomerase. Mol. Cell. Biol. *12*, 4601-4611.

Tachibana, H., Ohta, K., Sawano, H., Koumoto, Y., and Segawa, S. (1994). Relationship between the optimal temperature for oxidative refolding and the thermal stability of refolded state of hen lysozyme three-disulfide derivatives. Biochemistry *33*, 15008-15016.

Talluri, S., Rothwarf, D.M., and Scheraga, H.A. (1994). Structural characterization of a three-disulfide intermediate of ribonuclease A involved in both the folding and unfolding pathways. Biochemistry *33*, 10437-10449.

Tanford, C. (1968). Protein denaturation. Adv. Prot. Chem. *23*, 121-282.

Tao, K., Stearns, N.A., Dong, J., Wu, Q., and Sahagian, G.G. (1994). The proregion of cathepsin L is required for proper folding, stability, and ER exit. Arch. Biochem. Biophys. *311*, 19-27.

Thatcher, D.R. and Hitchcock, A. (1994). Protein folding in biotechnology. In *Mechanisms of Protein Folding*. R.H. Pain, ed. (Oxford: IRL Press), pp. 229-261.

Thornton, J.M. (1981). Disulphide bridges in globular proteins. J. Mol. Biol. *151*, 261-287.

Tokatlidis, K., Dhiurjati, P., Millet, J., Begun, P., and Aubert, J.P. (1991). High activity of inclusion bodies formed in *Escherichia coli* overproducing *Clostridium thermocellium* endoglucanase D. FEBS Lett. *282*, 205

Tyler-Cross, R. and Schirch, V. (1991). Effects of amino acid sequence, buffers, and ionic strength on the rate and mechanism of deamidation of asparagine residues in small peptides. J. Biol. Chem. *266*, 22549-22556.

Upton, F.Z., Francis, G.L., Ross, M., Wallace, J.C., and Ballard, F.J. (1992). Production and characterization of recombinant chicken insulin-like growth factor-I from *Escherichia coli*. J. Mol. Endo. *9*, 83-92.

Uversky, V.N. (1993). Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. Biochemistry *32*, 13288-13298.

Uversky, V.N. and Ptitsyn, O.B. (1994). "Partly folded" state, a new equilibrium state of protein molecules: Four-state guanidinium chloride-induced unfolding of b-lactamase at low temperature. Biochemistry *33*, 2782-2791.

Uversky, V.N., Semisotnov, G.V., Pain, R.H., and Ptitsyn, O.B. (1992). 'All-or-none' mechanism of the molten globule unfolding. FEBS Letters *314*, 89-92.

Van Obberghen-Schilling, E. and Pouysségur, J. (1983). Mitogen-potentiating action and binding characteristics of insulin and insulin-like growth factors in chinese hamster fibroblasts. Experimental Cell Research *147*, 369-378.

Van, P.N., Rupp, K., Lampen, A., and Söling, H.-D. (1993). CaBP2 is a rat homolog of ERp72 with proteindisulfide isomerase activity. Eur. J. Biochem. *213*, 789-795.

Venetianer, P. and Straub, F.B. (1963). The enzymic reactivation of reduced ribonuclease. Biochim. Biophys. Acta *67*, 166-168.

Vértesy, L., Tripier, D., Koller, K.-P., and Riess, G. (1995). Disulphide bridge formation of proinsulin fusion proteins during secretion in *Streptomyces*. Biol. Chem. Hoppe-Seyler *372*, 187-192.

Vize, P.D. and Wells, J.R.E. (1987). Spacer alterations which increase the expression of porcine growth hormone in *E. coli*. FEBS Letters *213*, 155-158.

Vuori, K., Myllylä, R., Pihlajaniemi, T., and Kivirikko, K.I. (1992). Expression and site-directed mutagenesis of human protein disulphide isomerase in *Escherichia coli*. This multifunctional polypeptide has two independently acting catalytic sites for the isomerase activity. J. Biol. Chem. *267*, 7211-7214.

Walker, K.W. and Gilbert, H.F. (1995). Oxidation of kinetically trapped thiols by protein disulfide isomerase. Biochemistry *34*, 13642-13650.

Waltho, J.P., Feher, V.A., Merutka, G., Dyson, H.J., and Wright, P.E. (1993). Peptide models of protein folding initiation sites. 1. Secondary structure formation by peptides corresponding to the G- and H-helices of myoglobin. Biochemistry *32*, 6337-6347.

Weissman, J.S. and Kim, P.S. (1991). Reexamination of the folding of BPTI: predominance of native intermediates. Science *253*, 1386-1393.

Weissman, J.S. and Kim, P.S. (1992). The pro region of BPTI facilitates folding. Cell *71*, 841-851.

Weissman, J.S. and Kim, P.S. (1993). Efficient catalysis of disulphide bond rearrangements by protein disulphide isomerase. Nature *365*, 185-188.

Wetlaufer, D.B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. Proc. Natl. Acad. Sci. USA *70*, 697-701.

White, F.H., Jr. (1967). Reduction and reoxidation at disulphide bonds. Methods Enzymol. *11*, 481-484.

Winther, J.R. and Sorensen, P. (1991). Propeptide of carboxypeptidase Y provides a chaperone-like function as well as inhibition of the enzyme activity. Proc. Natl. Acad. Sci. USA *88*, 9330-9334.

Woodward, C.K. (1994). Hydrogen exchange rates and protein folding. Curr. Opin. Struct. Biol. *4*, 112-116.

Wright, H.T. (1991). Nonenzymic deamidation of asparaginyl and glutaminyl residues in proteins. Critical Reviews in Biochemistry and Molecular Biology *26*, 1-52.

Wu, L.C., Laub, P.B., Elöve, G.A., Carey, J., and Roder, H. (1993). A noncovalent peptide complex as a model for an early folding intermediate of cytochrome *c*. Biochemistry *32*, 10271-10276.

Youngman, K.M., Spencer, D.B., Brems, D.N., and DeFelippis, M.R. (1995). Kinetic analysis of the folding of human growth hormone -Influence of disulfide bonds. J. Biol. Chem. *270*, 19816-19822.

Yu, Y., Makhatadze, G.I., Pace, C.N., and Privalov, P.L. (1994). Energetics of ribonuclease T_1 structure. Biochemistry *33*, 3312-3319.

Zhang, J.-X. and Goldenberg, D. P. (1993). Amino acid replacement that eliminates kinetic traps in the folding pathway of pancreatic trypsin inhibitor. Biochemistry *32*, 14075-14081.

Zhang, R. and Snyder, G.H. (1989). Dependence of formation of small disulphide loops in two-cysteine peptides on the number and types of intervening amino acids. J. Biol. Chem. *264*, 18472-18479.

Zhang, R. and Snyder, G.H. (1991). Factors governing selective formation of specific disulphides in synthetic variants of α-conotoxin. Biochemistry *30*, 11343-11348.

Zhu, X.L., Ohta, Y., Jordan, F., and Inouye, M. (1989). Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. Nature *339*, 483-484.

Zwanzig, R., Szabo, A., and Bagchi, B. (1992). Levinthal's paradox. Proc. Natl. Acad. Sci. USA *89*, 20-22.

ADDENDUM

- The statements which indicate a direct relationship between the rate of disulphide formation and thermodynamic stability on pages 24, 25 and 130 are incomplete as inverse relationships are known. For example, the Cys^5-Cys^{55} disulphide bond of bovine pancreatic trypsin inhibitor is slow to form, however, it appears to be the most stable of the three native disulphide bonds (van Mierlo et al., 1991, Creighton et al., 1996). This apparent anomaly highlights the need for caution when interpreting the IGF-I oxidative folding reactions especially when considering the possibility that certain disulphide bonds may be strained as has been discussed for the Cys^{47}-Cys^{52} disulphide bond on page 130.

- Intramolecular disulphide bond rearrangement, which was not considered in Chapter 6, may provide a possible mechanism for direct interconversion between intermediates in Figures 6.8 and 6.9. Also, the intermediate designated as peak 4 during the oxidative folding of Long-[Arg^3]IGF-I (page 130), contained a native Cys^{18}-Cys^{61} disulphide bond and second disulphide bond involving cysteines 47, 48 or 52. The assignment of the second disulphide bond of peak 4 was unknown; but could contain the native-like Cys^{47}-Cys^{52} disulphide bond or alternatively non-native-like Cys^{48}-Cys^{52} or Cys^{47}-Cys^{48} disulphide bonds. A point not discussed in this chapter was that intramolecular disulphide rearrangement could provide the route to the native Cys^{47}-Cys^{52} disulphide bond from either of the possible non-native configurations.

- A recent study that is relevant to the modelling work presented in Chapter 5 is that of misfolded insulin. This study shows the NMR solution structure of native insulin containing Cys^{A7}-Cys^{B7}, Cys^{A6}-Cys^{A11}, Cys^{A20}-Cys^{B19} disulphide bonds and that of an alternative isomer with Cys^{A6}-Cys^{B7}, Cys^{A7}-Cys^{A11}, Cys^{A20}-Cys^{B19} disulphide bonds (Hua et al., 1995). The alternative isomer of insulin, which is analogous to the alternative isomer of IGF-I in this thesis, has lost helix Val^{A3} to Thr^{A8} (equivalent to helix 2 in IGF-I), a feature similar to that proposed for the structure of the alternative isomer of IGF-I presented in Chapter 5. Unlike the model of alternative IGF-I which predicts small differences between the native and alternative isomers in the regions outside helix 2, the solution structures of native and alternative insulin exhibit a greater difference in the orientation of the equivalent helix 1 and helix 3. An explanation for this may be that the C-domain of IGF-I (residues 30 to 41), which may serve to constrain the structure of the IGF-I isomers, is absent in insulin. This is supported by the observation that the alternative isomer of IGF-I has a greatly reduced biological activity compared to that of the alternative isomer of insulin. Thus, in a type 1 IGF receptor binding assay during which there is no detectable disulphide rearrangement, the alternative isomer of IGF-I has less than 0.5 % of the activity of the native isomer (Milner et al., 1995). By comparison, the alternative isomer of insulin has 5-10% of the activity of the native molecule (Sieber et al., 1978), possibly as a result of flexibility of the molecule on binding to the insulin receptor (Hua et al., 1995).

Creighton, T.E., Darby, N.J., and Kemmink, J. (1996). The roles of partly folded intermediates in protein folding. *FASEB J. 10*:110-118.

Hua, Q.-X., Gozani, S.N., Chance, R.E., Hoffman, J.A., Frank, B.R., and Weiss, M.A. (1995). Structure of a protein in a kinetic trap. *Nature Struct.Biol. 2*:129-138.

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.

Sieber, P., Eisler, K., Riniker, B., Rittel, W., Märki, F., and de Gasparo, M. (1978). Synthesis and biological activity of two disulphide bond isomers of human insulin: [A7-A11,A6-B7-cystine]- and [A6-A7,A11-B7-cystine]insulin (human). *Hoppe-Seyer's Z.Physiol.Chem. 359*:113-123.

Van Mierlo, C.P.M., Darby, N.J., Neuhaus, D., and Creighton, T.E.(1991). Two-dimensional ^{1}H nuclear magnetic resonance study of the (5-55) single-disulphide folding intermediate of bovine pancreatic trypsin inhibitor. *J.Mol.Biol. 222*:373-390.

3/7/96