Order Number 9521324

Synthesis and characterization of epidermal growth factor-like peptides derived from human blood coagulation factors IX and X

Yang, Yan, Ph.D.

City University of New York, 1995

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

SYNTHESIS AND CHARACTERIZATION OF EPIDERMAL GROWTH FACTOR-LIKE PEPTIDES DERIVED FROM HUMAN BLOOD COAGULATION FACTORS IX AND X

by

YAN YANG

A dissertation submitted to the Graduate Faculty in Chemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1995

This manuscript has been read and accepted for the Graduate Faculty in Chemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

1/24/95
Date

William Sweeney
Chair of Examining Committee

1/25/95
Date

Richard Pagni
Executive Officer

James Tam

Ruth E. Stark

Michael Blumenstein

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

SYNTHESIS AND CHARACTERIZATION OF EPIDERMAL GROWTH FACTOR-LIKE DOMAINS DERIVED FROM HUMAN BLOOD COAGULATION FACTORS IX AND X

by

Yan Yang

Advisor: Professor William V. Sweeney

There are four parts in this dissertation. The first chapter mainly reviews the pioneer work by Dr. Cohen's group, who discovered, isolated, and characterized the epidermal growth factor (EGF) molecule. EGF is a 53 residue, cysteine-rich peptide with disulfide pairing Cys 1-3, 2-4 and 5-6. The structural analysis of EGF through CD and NMR has been done, indicating a non-helical structure and two antiparallel β-sheet. The chemical synthesis has been successfully performed by Merrifield's group using a solid-phase method.

The second chapter of the dissertation describes how a two-step approach has been developed to resolve a multiple disulfide scrambling problem in EGF-like peptides. This method reduces the number of possible disulfide isomers from 15 to 3 and has been applied to the C-terminal EGF-like domain in human blood coagulation factor IX (fIX_{EGF-C}) and the N-terminal EGF-like domain in factor X (fX_{EGF-N}).

Chapter 3 illustrates a characterization of a side reaction found in a synthesis of an EGF-like peptide with Fmoc chemistry. It is the aspartimide formation from -Asp-X-, where X is Asn or Gly. This side reaction was thought not to occur when the side-chain of Asp is blocked with t-butyl which would avoid base-catalyzed aspartimide formation. The formation of aspartimide between two amino acids leads to a formation a piperidine adduct having an additional mass of 67 u.

Chapter 4 presents a preliminary structural analysis on EGF-like peptides using circular dichroism (CD). The samples include the N-terminal EGF-like domain of human factor IX ($fIX_{EGF\text{-}N}$), $fX_{EGF\text{-}N}$, $fIX_{EGF\text{-}C}$, and a misfolded peptide with the $fIX_{EGF\text{-}C}$ sequence. Peptides in different oxidized states have been used in this study. The results indicate that at pH 4.2, all the spectra of reduced and oxidized peptides, except that of fully oxidized $fIX_{EGF\text{-}C}$, have common CD features, similar to that of oxidized EGF. The CD spectrum of $fIX_{EGF\text{-}C}$ is completely unlike that of the other EGF-like peptides, suggesting a significantly different structure.

In memory of my father

JING-GUO YANG, Professor of Cardiacsurgery, M.D.

who encouraged me to pursue graduate work,

but passed away during my first semester of Ph.D. study

CONTENTS

ACKNOWLEDGMENTS

I would like to express my sincere thanks to many people and friends for their support during my Ph.D. studies, especially to following individuals:

To Professor William V. Sweeney, my Ph.D. advisor and mentor, for taking me into his Lab, providing support and encouragement, patience in teaching writing and training in giving seminars, his constant guidance throughout education including social life, and for his willingness to give me plenty of freedom to study in diverse areas.

To Professor James P. Tam, my Ph.D. co-advisor, for giving me the opportunity to work in his Lab on an exciting project, taking me to attend the International Chinese Peptide Symposium (Beijing, 1994), letting me visit his current Lab at Vanderbilt University as a part of his group, and for his encouragement and advice in peptide chemistry, including numerous feed-back faxes.

To Professor R. Bruce Merrifield, for his role model as a scientist and for his generous permission to work in his Lab, which I shall always remember with gratitude. I am lucky to have worked for several years in his lab at The Rockefeller University.

To Mrs. Libby Merrifield, for her encouragement, the tea party offerings, and the nice time we shared in our office.

To Dr. Seymour Koenig, a relaxometry NMR expert and now an IBM Research Emeritus, and Mrs. Harriet Koenig, for their parental concern for my career and personal development.

To Professor Brian T. Chait, for his comments, suggestion and encouragement during manuscript revision. To many members in his lab: Dr. Klaus Schneider, for his collaboration in weighing our peptides with MALDI-MS, for many discussions, and in particular for his suggestion of using peptide mapping in our peptides; Ms. Susanna Thörnqvist and Dr. Urooj Mirza, for the mass determination with ES-MS; Dr. David Fenyo, for his help in getting and treating original mass data after both Susanna and Klaus had left; Dr. Wenzhu Zhang, for her help to get the information from the Profile program; and other individuals for helpful discussions.

To many members in Dr. Merrifield's Lab, Dr. Tam's Lab, Dr. Sweeney's Lab and

Dr. Cowburn's Lab: Dr. Cecille Unson; Ms. Dee Graefe; Ms. Beverly Walson; Dr. Linda Huang, my first supervisor in Boc chemistry; Dr. Chuan-Fa Liu, my principal supervisor on solid phase peptide synthesis (many late night help sessions too); Dr. Jun Shao, my instructor in Fmoc chemistry, both orally and E-mailly; Dr. to be, Yuming Gong, for her suggestion on the side reaction mechanism; Dr. Lian-Shan Zhang, Mr. Jing-wen Zhang, Ms. Cui-Rong Wu and many individuals for discussion and teaching; Ms. Connie Cheung and others who helped with amino acid analysis; Dr. Sean Cahill, for his assistance with the operation of the CD instrument;.and others who provided good social interactions.

To Mr. Steven Bobin and his fellows, Elizabeth and Meeghan, in the Synthesis Facility at Hunter, for the synthesis of some peptide-resins using the ABI synthesizer.

To Dr. Mike Blumenstein, for his encouragement during the dissertation work, for teaching me to use the 300 MHz NMR instrument and for running some samples on the 400 and 500 MHz NMR instruments, and for being on my committee.

To Professor Ruth Stark, for her encouragement during the dissertation work and for being on my committee.

To Professor Fasman, for his kind providing his Chou-Fasman Algorithm program for the prediction of peptide secondary structure and the helpful discussion.

To many professors and individuals at West China University of Medical Sciences (Chengdu, China), for the fundamental training they offered, especially to Professor YIN Gong-Kuan with whom I did my graduate work, Master in Medicine.

At last, to my family: my mother Yu-Ying He, Professor in Nutrition, who dedicated the last three years to take care of my son so that I can work full-time in the Ph.D. program; my brother and sister-in-law, Drs. Hua Yang and Zhen Wang; and my son, Xiao-Ming Jia, who spends most weekends without me, and took his first Citywide Test, for New York City, with a rank of above 97% in reading and 99% in Math.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

1. Amino Acids:

3 letter	1 letter	Full name
Ala	A	alanine
Asn	N	asparagine
Asp	D	aspartic acid
Arg	R	arginine
Cys	C	cysteine
Glu	E	glutamic acid
Gln	Q	glutamine
Gly	G	glycine
His	H	histidine
Ile	I	isoleucine
Leu	L	leucine
Lys	K	lysine
Met	M	methionin
Phe	F	phenylalanine
Pro	P	proline
Ser	S	serine
Thr	T	threonine
Trp	W	tryptophan
Tyr	Y	tyrosine
Val	V	valine

2. Other abbreviations

Acm	acetamidomethyl
Boc	t-butyloxycarbonyl
Bu	butyl
Bzl	benzyl
CD	ultraviolet circular dichrois
COSY	homonuclear-correlated spectroscopy
Cz	benzyloxycarbonyl
DCC	dicyclohexylcarbodiimide
DCM	methylene chloride
DIEA	diisopropylethyl amine
DMF	N,N-dimethylformamide

DMS	dimethyl sulfide
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
EDT	1.2-ethanedithiol
EGF	epidermal growth factor
Fmoc	9-fluorenylmethyloxycarbonyl
fIX_{EGF-C}	the C-terminal EGF-like domain in human coagulation blood factor IX
fIX_{EGF-N}	the N-terminal EGF-like domain in human coagulation blood factor IX
fX_{EGF-N}	the N-terminal EGF-like domain in human coagulation blood factor X
GSH	reduced glutathione
GSSG	oxidized glutathion;
HF	hydrofluoric acid
HOBt	1-hydroxybenzotriazole
HMP resin	4-(hydroxymethyl)-phenoxymethyl-Copoly styrene resin
Meb	4-methylbenzyl
MALDI-MS	matrix-assisted laser desorption mass spectrometry
NGF	nerve growth factor
Pam	phenylacetamidomethyl
PDI	protein disulfide-isomerase
Pmc	pentamethylchroman
TFA	trifluoroacetic acid
TFE	trifluoroethanol
Tos	4-toluenesulfonyl (tosyl)
Trt	trityl (triphenylmethyl)
UV	ultraviolet spectrometry
VCD	vibrational circular dichroism

Chapter 1

Background

1.1. Epidermal Growth Factor (EGF): its discovery history (Mark, 1986)

For the discovery of EGF, Dr. Stanley Cohen shared the 1986 Nobel prize for Physiology or Medicine with Dr. Rita Levi-Montalcini, who identified nerve growth factor (NGF). Cohen's discovery of EGF was a direct outgrowth of his research on NGF and largely accomplished after 1959. Cohen had noted a peculiar effect when the NGF containing extracts of salivary glands were injected into newborn mice. The eyelids of the animals opened sooner than they usually do and their teeth also grew in ahead of schedule. These effects did not occur with pure NGF, thereby indicating that some additional factor in the extracts must be the cause. By 1962, Cohen had isolated the causative factor from male mice, proved it to have growth-stimulatory effects, and named it EGF. The amino acid sequence of the protein was determined in the early 1970's. Cohen and his colleagues found the molecule contains 53 amino acids and the cysteinyl residues were linked as disulfides in the positions 6-20, 14-31, and 33-42 (Savage et al., 1972a, 1972b, 1973; Gregory, 1975; Figure 1.1).

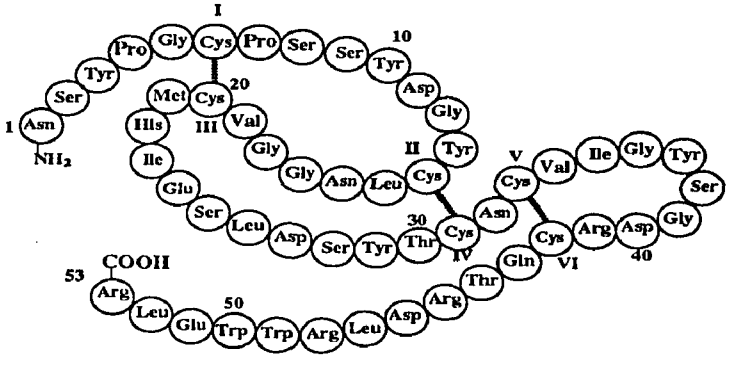


Figure 1.1. The primary structure of mouse epidermal growth factor.

Redrawn from Savage et al., J *Biol Chem* **248**, 7669-7672, 1973.

The discovery of EGF is important, not just for understanding normal growth control, but also because disturbances in cell growth and differentiation can lead to the development of cancer. A better understanding of how growth factors work may thus eventually lead to better cancer therapies. Meanwhile, with EGF becoming available in larger quantities as a result of chemical synthesis or gene cloning, the factor may find application in growing skin cells for treating burn victims and, Cohen notes, may have a potential application in ulcer treatment (Cohen and Taylor, 1974).

1.2. Physicochemical studies on EGF

Some physical and chemical properties for mouse EGF, and apparent thermodynamic parameters of unfolding of EGF, reported by Cohen and coworkers (Taylor et al., 1972; Holladay et al., 1976) are listed as follows.

Property	Value
Extinction coefficient ($E^{1\%}_{1cm}$ at 280 nm)	30.9
Isoelectric point	pH 4.60
Polarity*	40
Missing amino acids	Lys, Ala, Phe
Hexosamine content	None detected
Neutral sugar content	None detected
Antigenicity	Antigenic
Apparent free energy, ΔG°** (kcal/mol)	18.0
Apparent enthalpy, ΔH** (kcal/mol)	24.4
Apparent entropy, ΔS** (cal $mol^{-1}deg^{-1}$)	20.4

* As determined from the amino acid sequence using the method of Capaldi and Vanderkooi (1972). Thus, based on this parameter, EGF is a "typical" soluble protein and does not contain an unusually high content of hydrophobic side chains.
**Apparent thermodynamic parameter of unfolding at 40°C.

1.3. Structural analysis by CD and NMR

Circular dichroic (CD) studies, initially by the Cohen group, revealed the secondary structure of EGF (Taylor et al., 1972). The dominant feature of the spectrum is the strong negative dichroic extremum at 200 nm having an ellipticity of approximately -12,700 deg cm^2 per decimole. This dichroic band has been assigned to a π-π^* transition of the peptide bond (Holzwarth and Doty, 1965) and is a characteristic property of nonhelical polypeptides and proteins (Beychok, 1966). The n-π^* peptide bond transition which is characteristic of helical structures is absent in EGF spectra. With a computer-assisted curve-fitting program based on the polylysine model (Greenfield and Fasman, 1969), EGF was evaluated to be principally nonhelical with a random coil content of 74% and β form content of 25%.

Cohen's group and other groups further resolved through deconvolution the CD bands and assigned the major secondary structure in the CD spectrum of EGF (Taylor et al., 1972; Holladay et al., 1976; Kohda and Inagaki, 1992):

Wavelength (nm)	Structural feature
<220:	peptide groups
202	aperiodic structure
213	β structure
>220:	side-chain chromophores (Tyr, Trp and S-S bonds)
225-230, 253, 280	aromatics and disulfides
or: 240-290	aromatics and disulfides

Kohda and Inagaki (1992) studied pH dependence of conformation change by CD and NMR. Figure 1.2 shows the CD spectra at various values of pH. They compared the

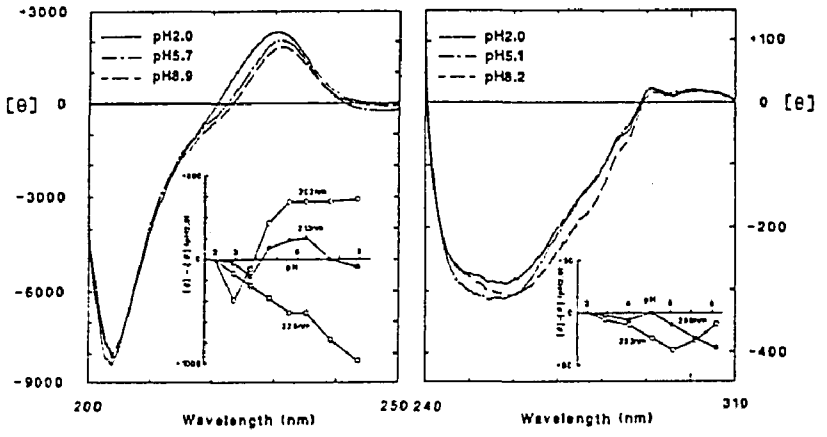


Figure 1.2. CD spectra (mean residue ellipticity, deg.cm^2/dmol) of mouse EGF at various pHs. Insets show the pH dependence of [θ] at 202 nm (aperiodic structure), at 213 nm (β structure), and at 225 , 253, and 280 nm (aromatics and disulfides). A difference between [θ] at a given pH and [θ] at pH 2.0 was plotted as a function of pH.
Reproduced from Kohda D. and Inagaki F., *Biochem* **31**, 11928-11939, 1992, Fig.1, with the permission of the publisher.

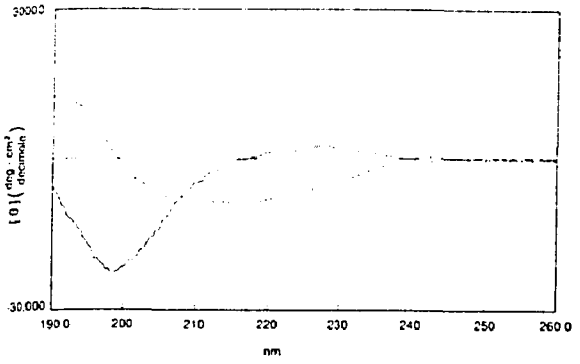


Figure 1.3. Far UV CD spectra of recombinant human EGF. The spectra of 0.44 mg/mL of reduced (---) and oxidized (—) EGF in 20 mM sodium phosphate pH 3.0 were recorded using a 0.02 cm cuvette, and the molar ellipticities were determined assuming a mean residue weight of 117. (Reproduced from Narhi et al., *Int J Peptide Protein Res* **39**, 182-187, 1992, Fig. 2, by permission of the publisher). Another almost identical CD figure for reduced and oxidized EGF molecules, at pH 7.5, by the same group (Prestrelski et al.) is shown in *J Biol Chem*, **267**, 319-322, 1992, Fig. 1B.

structure of mouse EGF at pH 2.0 and 6.8, where EGF is physiologically active. The conformations at two pHs were found to be nearly identical except for the C-terminal tail region detected by CD and the chemical shift comparison. The positions of the side chains of Leu^{47}, Arg^{48}, Trp^{49}, and Trp^{50} were changed probably by the effect of the deprotonation of Asp^{46}. Considering the fact that Leu^{47} is essential in EGF binding to the receptor, this conformational difference may be important in receptor recognition.

Interestingly, Cohen and coworkers (Holladay et al., 1976) suggested the assignment for disulfide bridges based on CD spectra. The 280 nm band may result from disulfide 6-20 (Cys 1-3) and the 253 nm CD band from disulfide bridges 14-31 and 33-42 (Cys 2-4 and Cys 5-6). According to them, the fact that the latter two disulfide bonds are separated only by a single residue (Asn^{32}) may introduce conformational constraints not present in less hindered disulfide bonds and may account for their existence as a single conformer.

Narhi and coworkers (Narhi et al., 1992; Prestrelski et al., 1992) studied reduced and oxidized human EGF by CD (Figure 1.3). Oxidized and reduced EGF display different far UV CD spectra. These spectra are pH independent between pH 3.0 and 7.5. It appears that the human EGF has a more regular secondary structure in the absence of the disulfide bonds than in their presence. The addition of the α-helix inducer TFE to the reduced molecule resulted in an enhancement of α-helix, while the oxidized molecule was unaffected by the presence of this reagent.

A number of groups have reported the NMR determined solution structure of EGF (Cooke et al., 1987; Makino et al., 1987; Kohda and Inagaki, 1988, 1992; Hommel et al., 1991; Baron et al., 1992). A rigid core structure is formed by the interplay of the three disulfide bridges and an antiparallel β-sheet consisting of Val^{19} to Glu^{24} and Asp^{27} to Asn^{32}. Furthermore, the hydrophobic amino acid residues of the long C-terminal segment fold back to interact locally with residues in the β-sheet. It is suggested that the C-terminal residues play an important role in the formation of the receptor-binding site.

1.4. Chemical synthesis of EGF

The successful solid-phase synthesis of EGF has been reported in 1986 (Heath and Merrifield). It was used as an example to emphasize the importance of new chemistry and the need to pay attention to details in Dr. Merrifield's Nobel Lecture (1986). The peptide was allowed to fold and oxidized under air by gentle stirring for 7 days. The yield of folded peptide was about 10%. The synthetic peptide could be readily isolated by HPLC at exactly the same retention time as natural EGF. In the sensitive and discriminating Leydig cell growth assay the synthetic and natural EGF had identical activity. It is worth mentioning that, when the native EGF was denatured, reduced, and then reoxidized and renatured, only about 10% of active peptide with the native structure could be recovered. The synthetic material also yields about 10% of the correct material (Merrifield, 1994).

1.5. EGF-like peptides: homology and classification

An EGF molecule is a cysteine-rich 53-residue peptide. The six cysteines have a disulfide bridging of Cys 1-3, 2-4, and 5-6. Between Cys 4 and 5 there is only one intervening residue. A wide range of peptides sharing EGF structure have been observed. Table 1.1 is a list, obtained from a database PROSITE (Bairroch, 1993), of proteins with an EGF-like domain cysteine pattern signature. The proteins currently known to contain one or more copies of an EGF-like pattern are believed to number over 260.

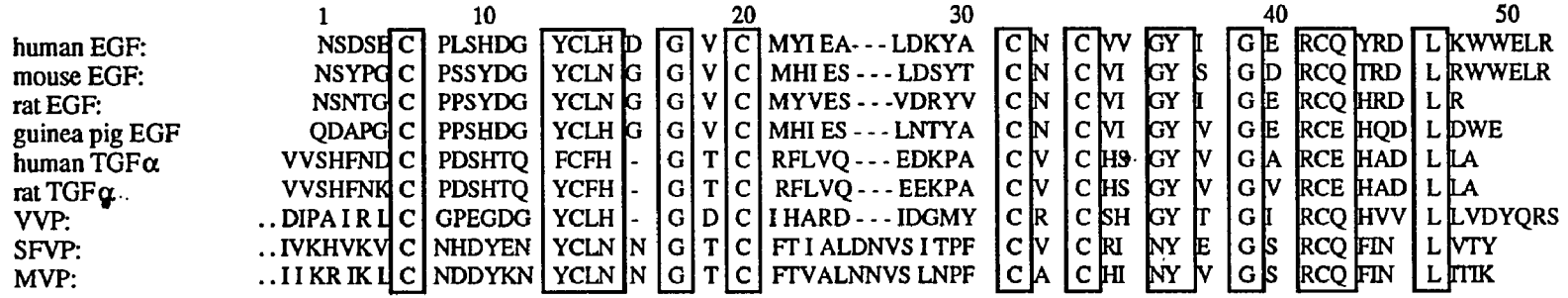
There are at least two classes of EGF-like units. The first group has high affinity to the EGF-receptor and are known to act as growth factors. The second group of molecules are from extracellular multi-domain proteins with diverse biological functions and exhibit no EGF-receptor binding activity. Among these functionally unrelated proteins are vitamin K-dependent clotting proteins (factors VII, IX, X and XII, protein C and protein S) and several non-vitamin K-dependent proteins. Classifications of EGF-like domains have been made by many investigators according to similarity of length between paired cysteines, identity of amino acid sequence, similarities in three regions (or to say inaccurately: loops),

and the biological activities of the proteins (Doolittle et al., 1984; Patthy, 1985; Appella et al., 1988; Rees et al., 1988; Campbell et al., 1989). One of these classifications is shown in Figure 1.4.

Table 1.1. EGF-like domain cysteine pattern signature. (From PROSITE: PDOC00022; EGF [documentation]).

1.	Transforming growth factor alpha (TGF-alpha).
2.	Amphiregulin, a growth factor.
3.	Schwannoma-derived growth factor (SDGF).
4.	Betacellulin, a growth factor.
5.	Growth factor related proteins from Vaccinia, Myxoma, and Shope fibroma viruses.
6.	Coagulation factors VII, IX, X (once) and XII (twice).
7.	Coagulation associated proteins C, S and Z (once).
8.	Urokinase and tissue plasminogen activator (TPA) (once).
9.	Complement components C6, C7, C8 alpha and beta chains, and C9 (once).
10.	Fibronectin (twice).
11.	Laminin subunits A (15 times), B1 (13 times) and B2 (12 times).
12.	Tenascin, an extracellular matrix protein (14 times).
13.	Nidogen (also called entactin), a basement membrane protein (once).
14.	Agrin, a basal lamina protein that causes the aggregation of acetylcholine receptors on cultured muscle fibers (4 times).
15.	Aggrecan (once) and versican (twice), two large proteoglycans.
16.	Selectins, cell adhesion proteins such as ELAM-1 (E-selectin), GMP-140 (P-selectin), or the lymph-node homing receptor (L-selectin) (once).
17.	Transforming growth factor beta-1 binding protein (TGF-B1-BP) (16 times).
18.	Drosophila neurogenic proteins: Notch (36 times), Delta (9 times), and Slit (7 times).
19.	Drosophila epithelial development protein Crumbs (30 times).
20.	Drosophila ectodermal development protein Serrate (14 times).
21.	Caenorhabditis elegans developmental proteins lin-12 (13 times) and glp-1 (10 times).
22.	Sea urchin protein uEGF-1 (at least 9 times).
23.	Human teratocarcinoma-derived growth factor 1 (TDGF-1) (CRIPTO protein) (once).
24.	Milk fat globule-EGF factor 8 (MFG-E8) (twice).
25.	Tyrosine-protein kinase receptors tek and tie (3 times).

A



VVP: vaccinia virus protein (C-terminal); SFVP: Shope fibroma virus protein (C-terminal); MVP: myxoma virus protein (C-terminal).

B

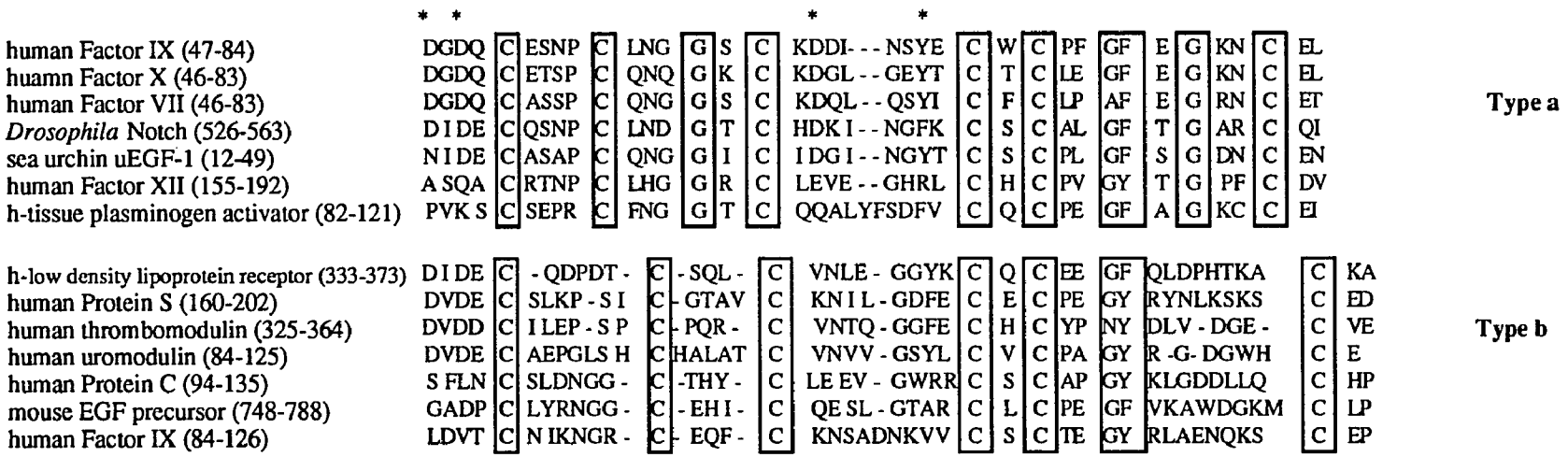


Figure 1.4 (A) Aligned sequence of the EGF-like units known to act as growth factors. The numbers correspond to the human EGF sequence and the boxed residues are the sixteen conserved and conservatively changed residues. (B) Aligned sequences of some representative EGF-like units from extracellular multi-domain proteins. The sequences are divided into **Type a** and **Type b** units following the classification of Doolittle et al. (1984); the sequences in (A) more closely resemble **Type a** units. Conserved residues are boxed; those being involved in β-hydroxylation and calcium binding are indicated by asterisks. (Redrawn from Campbell et al., 1989).

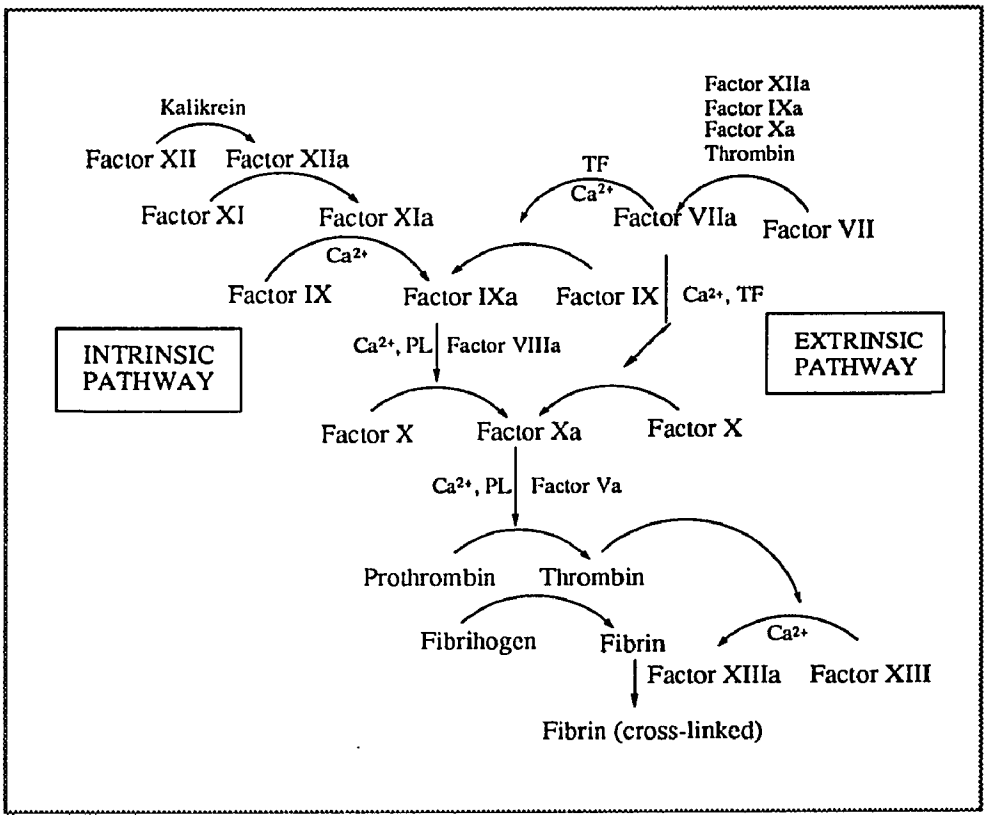


Figure 1.5. Diagram of the basic mechanism of blood coagulation. Redrawn from Kurachi K. Recombinant antihemophilic factors. In *Biotechnol. Ser.* 1991, **19** (Biotechnol. Blood), pp.177-195, Figure 9-1.

1.6. Human blood coagulation factors IX and X and the cascade system

Both human blood coagulation factors IX and X are involved in the intrinsic and extrinsic cascade pathways (Figure 1.5; Furie and Furie, 1988). A defect in the intrinsic clotting factor IX can cause haemophilia B (Christmas disease) which is an inherited, recessive, sex-linked, haemorrhagic condition (Rees et al., 1985). Appendix A shows amino acid sequence and domain structure for human factors IX and X. All the vitamin-K dependent blood coagulation factors have the same domain structures. For example, factor IX contains three distinct classes of domains, including a vitamin K-dependent domain having γ-carboxyglutamic acid (Gla) residues, two tandemly linked EGF-like domains, and a serine protease domain (Yoshitake et al., 1985). The Gla domain binds Ca^{2+} and allows the protein to interact with phospholipids on a membrane surface (Astermark et al., 1991a,

1991b). The function of the EGF-like domains is less clear. The N-terminal EGF-like domain binds Ca^{2+} (Huang et al., 1989, 1991; Handford et al., 1990, 1991), but little is known about the C-terminal EGF-like domain.

1.7. The C-terminal EGF-like domain in factor IX adopts an EGF-like disulfide pairing

It is widely believed that the C-terminal EGF-like domain from factor IX has a disulfide pairing analogous to EGF, although the disulfide structure in the peptide has not been directly determined yet. It was defined previously on the basis of homology with EGF (Young et al., 1978; 1984; Yoshitake et al, 1985). According to one hypothesis, the growth of the noncatalytic region in many proteins has been accomplished by a series of small tandem duplications (Doolittle, 1979) and subsequent evolution of the different domains of the noncatalytic region. Doolittle and coworkers (1984) published their computer-based characterization results for EGF-like domains in blood coagulation factors and the EGF precursor. Based on the suggestion (Patthy et al., 1984) that the presence of homologous units in those unrelated proteins may be due to exon-shuffling, Patthy (1985) constructed an evolution tree which indicates that the disulfide structure of two tandemly linked EGF-like domains in the blood coagulation factors adopts the EGF disulfide pattern. Højrup and Magnusson (1987) identified the Cys 1-3 pair in the C-terminal EGF-like domain from factor X, but the other disulfide pairs remain ambiguous, since either Cys 4 or 5 could be bound to Cys 2. They stated that there was no reason to believe that the second domain has a different disulfide bond pattern from that of the first EGF domain and mouse EGF.

Chapter 2

Selective formation of three disulfide bridges in epidermal growth factor-like peptides: a two-step approach and examples

Abstract

It is not uncommon to encounter problems in inducing synthetic peptides to fold disulfide bridges correctly. Attempts to synthesize the C-terminal epidermal growth factor (EGF)-like domain in human blood coagulation factors IX (fIX_{EGF-C}) and the N-terminal EGF-like domain in factor X (fX_{EGF-N}) have failed to yield peptides with proper disulfide pairings (Cys 1-3/2-4/5-6). Therefore a two-step approach to form three disulfide bridges selectively has been developed. The first model used was fIX_{EGF-C}. In the synthesis, four out of six cysteines are blocked with Trt (or Meb), and the remaining two cysteines are blocked with Acm. During the first step, this approach treats peptide as if it contains two pairs of disulfides, while the third disulfide bridge is formed in the second step. The choice of which pair of cysteines to block with Acm is critically important. To determine the optimal strategy, three peptides of fIX_{EGF-C} sequence have been synthesized, each with Acm blocking one of the three pairs of cysteines involved in disulfide bridges. Only the peptide having the Cys 2-4 pair blocked with Acm forms the desired disulfide isomer (Cys 1-3/5-6) in a high yield after the first-step folding, as identified by proteolytic digestion in conjugation with mass spectrometric peptide mapping. This result has been applied successfully to fX_{EGF-N}. Thus, the two-step approach limits the choice of disulfide formation and reduces the number of possible disulfide isomers from 15 to 3. Furthermore, by placing Acm groups appropriately, the desired EGF-like peptides can be obtained in high yield.

2.1. Introduction

Although the N-terminal domain in factor IX has been successfully synthesized

(Huang et al., 1989), to our knowledge no C-terminal EGF-like domain has been obtained by either biosynthesis or chemical synthesis. Thus, the synthesis of C-terminal EGF-like domain in human blood coagulation factor IX (84-128) (fIX_{EGF-C}) was performed to obtain enough material for biophysical characterization. Figure 2.1 shows the primary structure of the C-terminal EGF-like domain derived from human blood coagulation factor IX. For convenience, the amino acids are numbered 1-45, corresponding to 84-128 in the human blood coagulation factor IX sequence (Appendix A; Yoshitake et al., 1985).

The initial attempts to synthesize fIX_{EGF-C} failed to produce a large proportion of the peptide with the desired disulfide pairings. The amino acid composition and the molecular weight of the linear peptide were confirmed, but the peptide did not yield a single major peak by HPLC under a wide range of folding conditions (Figure 2.2). In the NMR spectrum, the apparent line widths were broad, consistent with the presence of a mixture. Thus, the C-terminal EGF-like synthetic peptide exhibited a significant multiple conformation problem that led to misformed disulfides, even though the N-terminal EGF-like synthetic peptide folded in a well-behaved manner to good yield (Huang et al., 1989). This was the starting point which motivated a two-step approach for selective formation of three disulfide bridges to be developed.

2.2. Materials and methods

2.2.1. Peptide synthesis

Synthetic peptides A, B and C (all three peptides having the same fIX_{EGF-C} sequence but each with a different pair of cysteines protected by Acm, see Table 2.1) were prepared using either Boc or Fmoc chemistry with stepwise solid-phase technology (Appendix B and C; Merrifield, 1963; Erickson and Merrifield, 1976; Barany and Merrifield, 1979; Stewart and Young, 1984).

Peptide A was synthesized manually using the Boc approach and standard coupling and deprotection techniques (Appendix C). N-Boc-Val-OCH2-Pam resin (0.76 mmol/g)

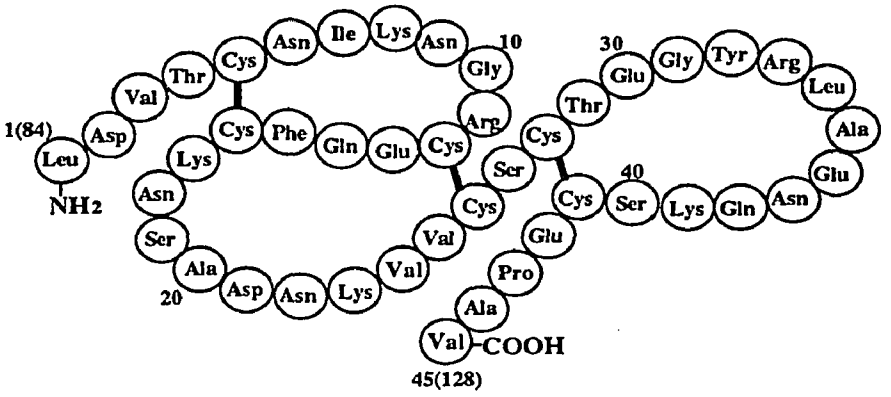


Figure 2.1. Primary structure of the C-terminal EGF-like domain in human blood coagulation factor IX (84-128). The numbering system used is based on the synthetic peptide, with the corresponding factor IX sequence position shown in parentheses.

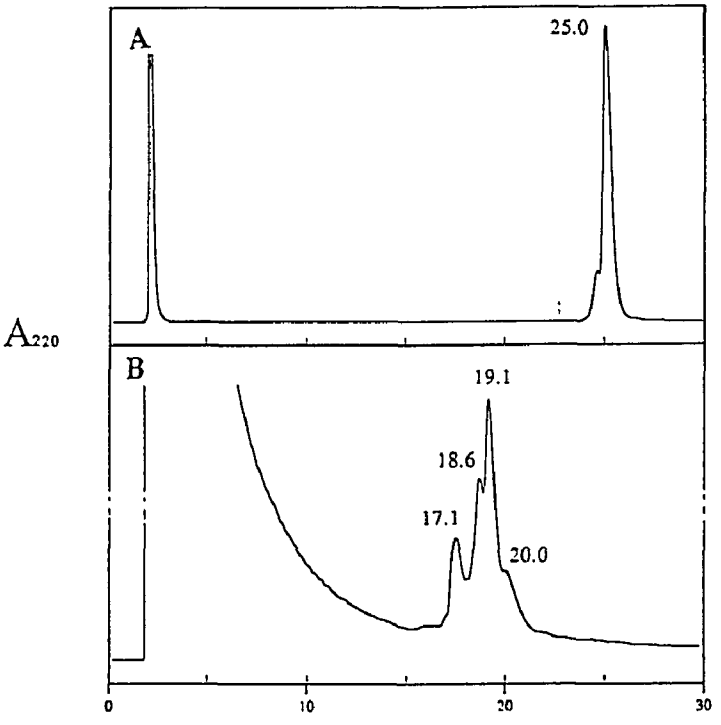


Figure 2.2. Reversed-phase HPLC profile showing the folding results of the C-terminal EGF-like domain in human blood coagulation factor IX resulting from using a single-step approach. Peptides were eluted with a 10-40% buffer B linear gradient over 30 min at 1.5 ml/min, monitored at 220 nm. Buffer A contained 5% acetonitrile in 0.045% TFA. Buffer B contained 60% acetonitrile in 0.037% TFA.

A. Reduced peptide, with all six cysteines reduced, before oxidation.

B. Peptide after folding for 29 hr at pH 8.0 in the presence of reduced and oxidized glutathione (1mM, ratio 1:1).

Reproduced from Fig 2 in *Protein Science*, 3, 1267-1275, 1994, with the permission of the publisher.

Table 2.1. Location of the cysteine residues and their blocking groups* in three synthetic peptides homologous to the C-terminal EGF-like domain in human blood coagulation factor IX (fIX_{EGF-C}). The numbering system used is based on the synthetic peptide, with the corresponding factor IX sequence position shown in parentheses.

Residue number	1(84)	5	12	16	26	28	41	45(128)
Sequence	H-LDVT	C	NIKNGR	C	EQF	C	KNSADNKVV	C S C TEGYRLAENQKS C EPAV-OH
Cysteines		1	2	3	4	5	6	
Peptide A		Meb	Meb	Meb	Meb	Acm	Acm	
Peptide B		Trt	Acm	Trt	Acm	Trt	Trt	
Peptide C		Acm	Trt	Acm	Trt	Trt	Trt	

* In peptide A Meb protecting group was used in place of Trt, because Boc chemistry was employed in peptide synthesis, while peptides B and C were synthesized with Fmoc chemistry.

was used, where Pam is phenylacetamidomethyl. All amino acids were purchased from Applied Biosystems, Inc. (Foster City, California). Side-chain protecting groups were as follows: Asp and Glu by OBzl; Ser and Thr by Bzl; Tyr by 2-bromo-Cz; Lys by 2-chloro-Cz; and Arg by Tos. Both Meb and Acm were used for blocking Cys. To remove the peptide from the resin, a low/high HF method (Tam et al., 1983) was used. In the low step, 6.5 ml of DMS, 0.75 ml of p-cresol and 0.25 ml of p-thiocresol were premixed with 0.3-0.5 g of peptide resin in the HF reaction vessel. 50 mg cysteine·HCl was added to minimize the side reactions with thiol moiety of the peptide. After the HF reaction vessel was cooled down to -78°C in an acetone-dry ice bath for 30 min, HF was transferred to reach the final volume of 10 ml to give a reaction mixture of HF:DMS:p-cresol:p-thiocresol 25:65:7.5:2.5 (v/v) and a 2 hr reaction proceeded at 0°C. After the removal of HF and DMS, the high HF step was initiated by charging liquid HF to 15 ml to give HF:p-cresol:p-thiocresol 93.3:5:1.7 (v/v). After 1 hr reaction at 0°C, HF was removed. The resulting resin and crude peptide contained aromatic scavengers and organic byproducts. It was washed with cold ethyl ether/mercaptoethanol 98:2 (v/v) to remove organic contaminants.

The crude peptide was then extracted with 8 M urea/0.2M DTT/0.1 M Tris-HCl buffer, pH 8.4. The peptide solution was dialyzed (Spectrapor, molecular weight cut-off, 2000) against deaerated and N_2-purged 8 M, 5 M, 3 M, and 1.5 M urea, all in 0.1 M Tris-HCl (pH 8.4). The dialyzed solution was then diluted with the same 0.1 M Tris-HCl buffer to a final urea concentration 0.3 M for the first-step folding (Heath and Merrifield, 1986). The peptide was purified to homogeneity by preparative HPLC. Some purified peptide was then reduced and used to examine various folding conditions.

Linear peptides B and C were synthesized by the RCMI Synthesis and Sequencing Facility at Hunter College. An ABI 430 synthesizer with FastMoc chemistry was used. All amino acids were purchased from Applied Biosystems, Inc. (Foster City, California). Side-chain protecting groups were as follows: Asp and Glu by OtBu; Ser, Thr and Tyr by t-Bu; Asn and Gln by Trt; Lys by Boc; and Arg by Pmc. Both Trt and Acm were used for blocking Cys. A 4-(hydroxymethyl)phenoxymethyl-Copoly resin (HMP resin/Wang resin) with a substitution of 0.88 mmol/g was used. For peptide B, the initial weight of the HMP resin was 0.284 g, and the final resin containing the 45-residue peptide was 2.53 g (99% of theoretical yield). TFA cleavage simultaneously deblocked all the side chain protecting groups except Acm. The cleavage mixture used for peptides containing Arg and Trt protecting groups was: 0.75 g crystalline phenol, 0.25 ml 1,2-ethanedithiol, 0.5 ml thioanisole, 0.5 ml H_2O and 10 ml TFA (Applied Biosystems, 1990). The crude peptide, including the Acm blocking group, was precipitated in cold methyl t-butyl ether and collected by centrifugation. From 0.30 g resin B, 0.128 g of crude peptide B was obtained. Crude peptide B, in some folding experiments, was used without further purification. For peptide C, the quality of the resin was not as good as peptide B, so it demanded that crude peptide be purified before use for subsequent folding experiments.

Protein disulfide-isomerase (PDI) is purchased from Takara Biochemical Inc.

Semipreparative C18 reverse phase HPLC was used sometimes for the purpose of purification. All the big scale purification was conducted by preparative C18 reverse phase

HPLC in the RCMI Synthesis and Sequencing Facility at Hunter College.

2.2.2. Disulfide oxidation

Disulfide oxidation and purification were monitored by analytical C18 reversed-phase HPLC. Peptides were eluted with a 10-40% buffer B linear gradient over 30 min at 1.5 ml/min, monitored at 220 nm. Buffer A contained 5% acetonitrile in 0.045% TFA. Buffer B contained 60% acetonitrile in 0.037% TFA.

Formation of two disulfide bridges in the first-step oxidation

Peptides folding at 10^{-5}M were conducted in buffers (10-50 mM) over a range of pH values between 4 and 10 under air oxidation, and monitored by HPLC. However, typical folding was carried out at pH 7.5 in 10 mM phosphate buffer at room temperature with slow stirring. The time for completion varied from 28 to 48 hr, depending on the temperature and pH. At room temperature and pH 7.5, peptide folding was usually complete within 24-30 hr. Because there is always a small portion of misfolded isomers present, the major product should usually be purified before carrying on the second step.

Formation of the third disulfide bridge in the second step

Removal of Acm with iodine is a known method which can induce direct formation of disulfide bonds (Appendix D). Acetic acid was added to the peptide solution to give a 10% acid solution. The solution was bubbled with nitrogen for 10 min before adding iodine. Addition of 5 mM iodine in methanol was dropwise until a brown color remained. The reaction mixture was bubbled continuously with nitrogen in a darkened vessel. The reaction was complete in 1 hr, as monitored by HPLC. After the solution was cooled in an ice-bath, 10 mM sodium thiosulfate was added to reduce excess iodine. The solution containing the crude peptide mixtures was purified by preparative C18 reversed-phase HPLC (2.5 x 30 cm) in the same manner as described for the analytical HPLC. The product was eluted with

aqueous acetonitrile/0.1 % TFA at a flow rate of 20 ml/min. The overall synthetic yield was about 10%.

2.2.3. Enzymatic digestion

Identification of the disulfide bridge pairings utilized proteolytic digestion combined with matrix-assisted desorption mass spectrometry for the assignment of the disulfide bridge locations. Readers may refer to appendix E for specificity of various proteases.

Enzymes were purchased as follows: *Staphylococcus aureus* protease V8 (Sigma), trypsin (Sigma), and endoproteinase Lys-C (Boehringer Mannheim). In each case, the monitoring and separation of individual fragments of enzymatic digestion were performed by analytical C18 reversed-phase HPLC.

Protease V8 digestion: A 50 mM ammonium bicarbonate buffer (pH 8) was used. The ratio of enzyme to substrate (w/w) was 1:50. The mixture of peptide and protease V8 was incubated at 37°C for 5 hr. *Trypsin digestion*: A 50 mM ammonium bicarbonate buffer (pH 8) was used. The ratio of enzyme to substrate (w/w) was 1:25. The mixture of peptide and trypsin was incubated at 37°C for 4 hr. *Endoproteinase Lys-C digestion*: A 50 mM ammonium bicarbonate buffer (pH 8.5) was used. The ratio of enzyme to substrate (w/w) was 1:20. The mixture of peptide and endoproteinase Lys-C was incubated at 35°C for 4 hr.

2.2.4. Mass spectrometry (MS)

HPLC-purified peptides, enzymatic digestion mixtures of peptides and enzymatically generated peptide fragments isolated by reversed-phase HPLC, were subjected to analysis on a matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectrometer constructed at The Rockefeller University and described elsewhere (Beavis and Chait, 1989, 1990). The mass spectra were collected by adding individual spectra obtained from 200 laser shots. Samples were prepared for laser desorption mass analysis

as follows: The laser desorption matrix material 4-hydroxy-α-cyano-cinnamic acid was dissolved in 0.1% TFA/acetonitrile 2:1 (v/v) to a concentration of 50 mM. A 20 μM aqueous solution of the sample was then added to the matrix solution to give a final sample concentration of approximately 2 μM. A small aliquot (0.5 μl) of this mixture was applied to the metal probe tip and dried at room temperature with forced air. The sample was then inserted into the mass spectrometer and analyzed. Bovine insulin was used to calibrate the mass spectra. Usually two runs are necessary for one sample in order to calibrate the mass spectra. The first run is a sample spectrum with TOF scale. The second one is a calibration spectrum containing calibrant and sample with mass vs. charge scale (M/Z). Since TOF is proportional to M/Z, by selecting two peaks in the sample spectrum the value of M/Z for those two peaks can be introduced with a transfer from the calibration spectrum and the scale in spectrum can be changed from TOF to M/Z. The sample spectrum also serves as a probe for concentration, so that the calibrant can be added in an amount comparable to that of the sample. The masses for peptide fragments after protease digestions were calculated using a MASS program, written by Dr. Detlev Suckau. Readers may refer to Appendix F.

One limitation should be mentioned. Only molecular weights ≥ 1000u can be valuably determined by MALDI-MS, because matrix compounds show up in the range of ≤ 1000u.

2.2.5. NMR

Proton NMR spectra were recorded using a GE/Bruker QE 300 MHz FT spectrometer at 25°C. Peptide samples ($fX_{EGF\text{-}N}$) were prepared in D_2O at pH 4.2. Uncorrected pH meter readings were used to obtain pD values.

2.3. Results

2.3.1. Synthetic approach

In general there are two approaches for the formation of intramolecular disulfide bridges in synthetic peptides: the single-step approach (Barany and Merrifield, 1979b; Tam

et al., 1991) and the sequential approach (Van Rietschoten et al., 1977; Akaji et al., 1992). In the first approach, all disulfide bridges are formed in a single-step either by oxidation with an oxidant (such as DMSO or air) or disulfide exchanges with reduced/oxidized glutathione. Air oxidation is most commonly used. It usually requires a long duration and highly dilute peptide in basic or neutral pH for completion. A variation of the air oxidation method is the thiol-disulfide interchange reaction with a mixture of reduced and oxidized glutathione. It is usually effective at the basic range of pH. Because the air oxidation and the mixed disulfide interchange method are slow processes, they allow equilibration of different conformers to produce thermodynamic-controlled products. DMSO is a mild oxidizing agent. Because of its widely applicable pH range (pH 3-8 or even broader), DMSO oxidation was particularly suitable for basic and hydrophobic peptides. Success of these methods in the single-step approach depends on the ability of the peptide to fold into the native state. The second approach is to form the sulfur-sulfur bridges sequentially. Various schemes using this approach have been employed to guide the formation of two disulfide bridges (Barany and Merrifield, 1979; Akaji et al., 1992). Usually the first pair of cysteines is freed and the disulfide bond formed as a consequence of the acidic cleavage of all non-cysteine protecting groups. The second pair is then liberated by a method that will not interfere with the stability of the previously formed disulfide peptide bridge using an electrophilic agent such as I_2 under acidic conditions. The sequential approach is a method leading to an unambiguous building of disulfide bridges.

For a three-disulfide containing peptide it is difficult to use a fully sequential approach, because of the need for three orthogonal sets of thiol protecting groups. Among other problems, a low yield may result after three individual steps. We have therefore sought to provide a hybrid two-step approach that utilizes the selectivity of the sequential approach and the generality of the single-step approach. The fundamental idea of this two-step approach is to limit the number of possibilities for disulfide bond formation in the different folding stages. There are fifteen possible ways to form three disulfide bonds

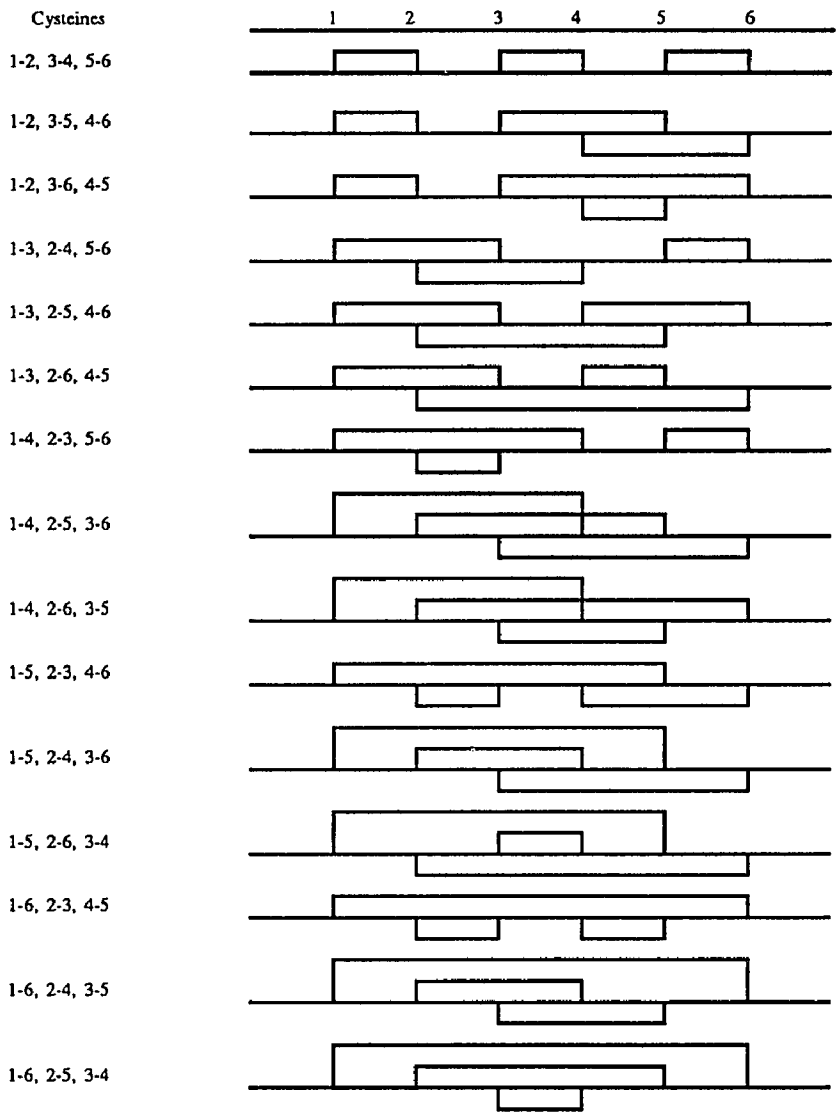


Figure 2.3.A. The fifteen possible disulfide bridging patterns for peptides containing six cysteines.

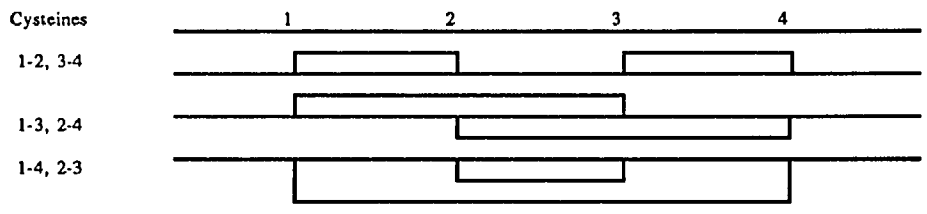


Figure 2.3.B. Three disulfide bridging patterns for peptides containing four cysteines.

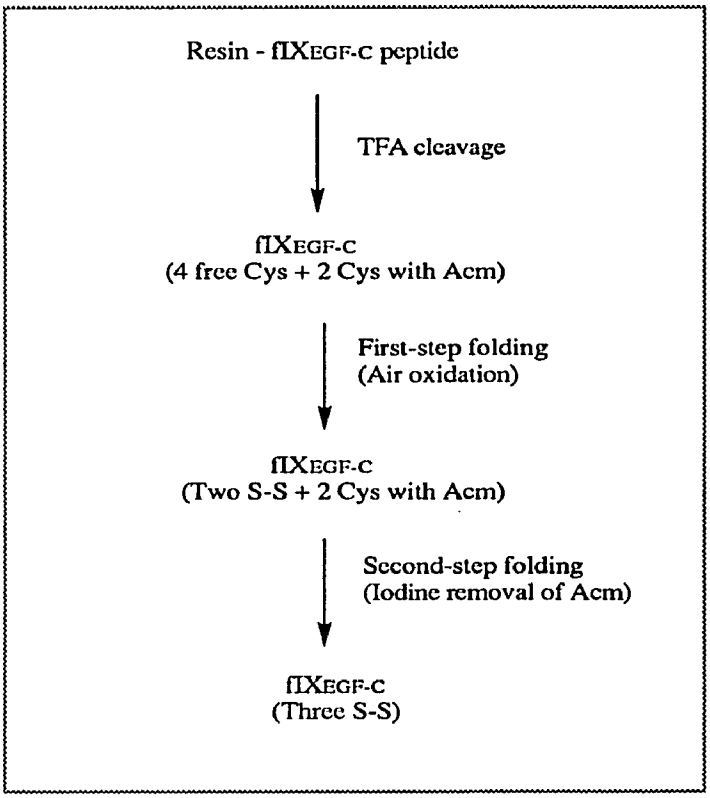


Figure 2.4. Scheme for two-step selective formation of three disulfide bridges.

(Figure 2.3.A); but only three ways to form two disulfide bonds (Figure 2.3.B). The two-step approach takes advantage of this, allowing only two sulfur-sulfur pairs to form in the first step, while the remaining disulfide bridge is formed in the second step. Thereby the number of possible disulfide isomers is reduced from fifteen to three.

The strategy is chemically feasible. In peptide chemistry, the sulfhydryl group of cysteines must be blocked during peptide synthesis, and a wide range of S-blocking groups is available. Here two different protecting groups for cysteines are used. One is the conventional protecting group, 4-methylbenzyl (Meb) in Boc chemistry or trityl (Trt) in Fmoc chemistry, which comes off from the cysteines after the peptide is cleaved from the resin. The other is acetamidomethyl (Acm) which remains on the cysteines after acidic cleavage and can be removed by treatment with iodine (Veber et al., 1972; Hiskey, 1981:

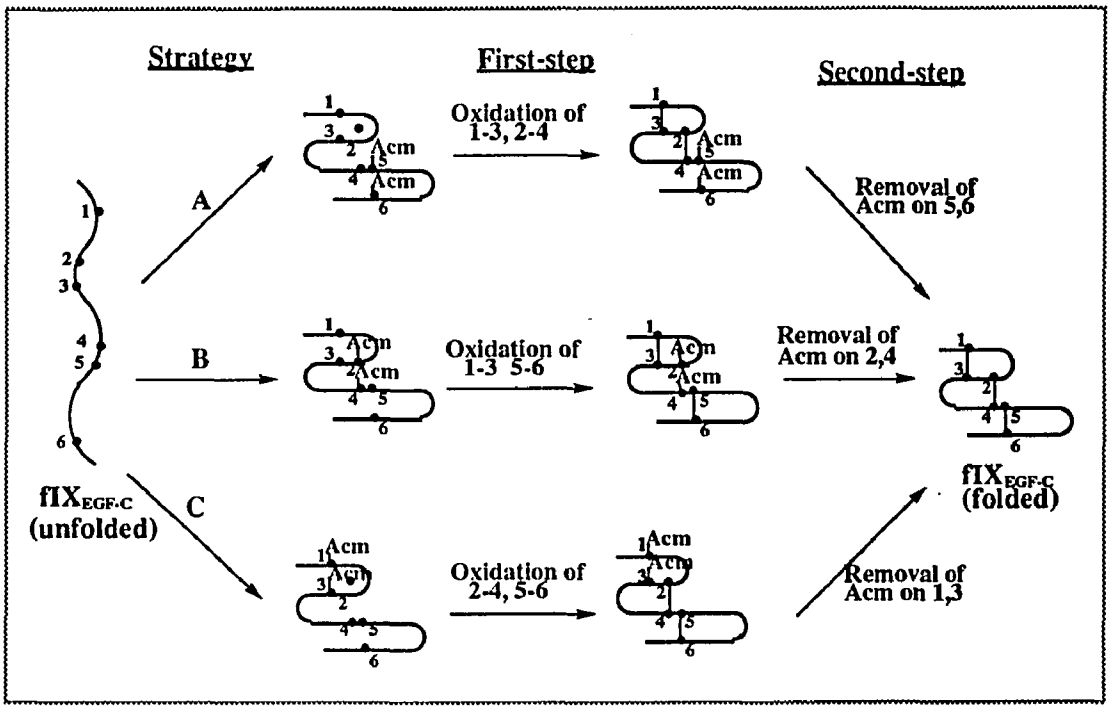


Figure 2.5. Strategy for synthesis of $fIX_{EGF\text{-}C}$.

Appendix D), in which simultaneous oxidation to the disulfide occurs. By using different protecting groups and removing them selectively, three S-S bridges can be formed correctly. In the first step, either TFA or HF was used to cleave the peptide from the resin and simultaneously deblock all the side-chain protecting groups except the Acm moiety. Thus, after cleavage, four cysteines were free and available to form two disulfide bridges in a maximum of three different disulfide isomers. In the second step, iodine was used to remove the Acm group from the last two cysteines to yield the third disulfide bridge. Because this step is conducted in acidic solution ($pH \leq 2$), disulfide scrambling is assumed not to occur. The two-step experimental procedure is shown in Figure 2.4.

Since $fIX_{EGF\text{-}C}$ has three disulfide bridges, there are three rational choices for where to place Acm protecting groups. Figure 2.5 and Table 2.1 show these choices and give an overview of the strategy for the two-step approach. To examine the effect of the position of

Table 2.2. Amino acid analysis for peptide A with 2 Acm on cysteines. The calculated number of residues was based on the average of all amino acid residues (except Cys) per mole of peptide. Sample was hydrolyzed in 6N HCl for 24 hr.

	Number of Residues	
Amino Acid	Theoretical	Found
Asp	7	6.8
Thr	2	2.1
Ser	3	1.9*
Glu	6	5.9
Pro	1	1.7
Gly	2	2.4
Ala	3	3.3
Cys	6	-**
Val	4	3.4
Leu	1	1.2
Phe	2	2.1
Tyr	1	1.1
Phe	1	1.2
Lys	4	3.9
Arg	2	1.9

* Ser is usually not stable and may found in a lesser amount than expected.
** Cys is the most unstable amino acid in hydrolysis.

the Acm groups on the disulfide pairing pattern, three peptides of fIX_{EGF-C} sequence have been synthesized, each with a different pair of cysteines protected by Acm. In peptide A Acm has been used to block the last pair of disulfide (Cys 5 and 6), in peptide B Acm blocks the middle pair of disulfide (Cys 2 and 4), and in peptide C Acm blocks the first pair of disulfide (Cys 1 and 3). Ideally all three peptides should yield the same final product with an EGF-like disulfide pairing of Cys 1-3, 2-4 and 5-6.

2.3.2. Synthesis of the C-terminal EGF-like domain in factor IX (fIX_{EGF-C})

2.3.2.1. Characterization of linear and folded peptides

The amino acid composition of the linear peptides was examined by amino acid analysis and was found to agree with the sequence (e.g. Table 2.2). The molecular weight

of the synthetic peptide with 2 Acm protecting groups was determined by matrix-assisted laser desorption mass spectrometry (MALDI-MS). The mass found for peptide A with 2 disulfides was 5108.1 ± 1.0 u (calculated mass 5107.8 u) (Figure 2.6). The masses found for peptides B and C, prior to disulfide formation, were 5110.9 ± 1.0 u (Figure 2.7) and 5111.5 ± 1.0 u, respectively (calculated mass 5111.8 u). The fully folded final products of peptide A and B were also examined. Molecular weights of 4963.0 ± 1.0 u and 4964.0 ± 1.0 u (Figure 2.8) were found, consistent with the calculated molecular weight of 4963.8 u for both peptides.

2.3.2.2. Determination of the two disulfide bridges after first-step oxidation

Since no disulfide scrambling was expected in the second step which was conducted under acidic condition (pH ≤ 2), special attention was given to identification of the disulfide bridge pairings after the first step folding. The standard approach (Lee and Shively, 1990) for obtaining disulfide structure in peptides is to digest the peptide with one or two selected endoproteinases, then to fractionate the resulting peptides by reversed-phase HPLC. There

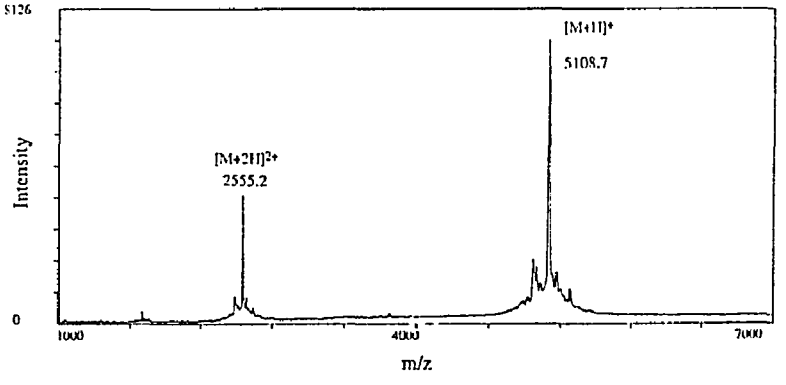


Figure 2.6. Matrix-assisted laser desorption mass spectrum for peptide A with two disulfide bridges formed after the first-step. Molecular weight was obtained from the average of the masses from the singlet-charged peak ($[M+H]^+$) and the doublet-charged peak ($[M+2H]^{2+}$), [5107.7+ (2555.2 x 2 - 2)] ÷ 2 = 5108.1. The calculated mass is 5107.8.
Reproduced from Fig.4 in *Protein Science*, **3**, 1267-1275, 1994, with the permission of the publisher.

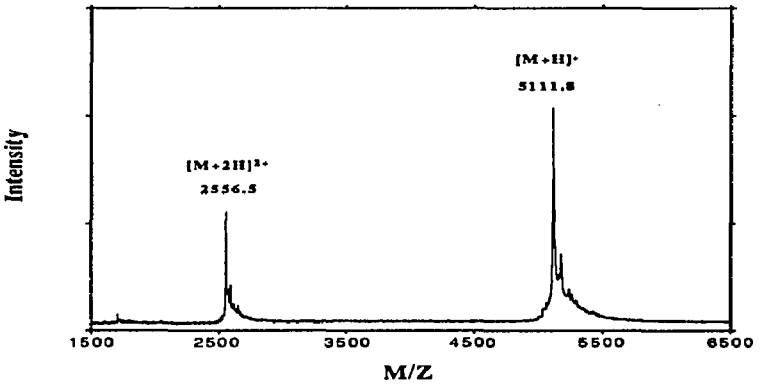


Figure 2.7. Matrix-assisted laser desorption mass spectrum for peptide B prior to disulfide formation.

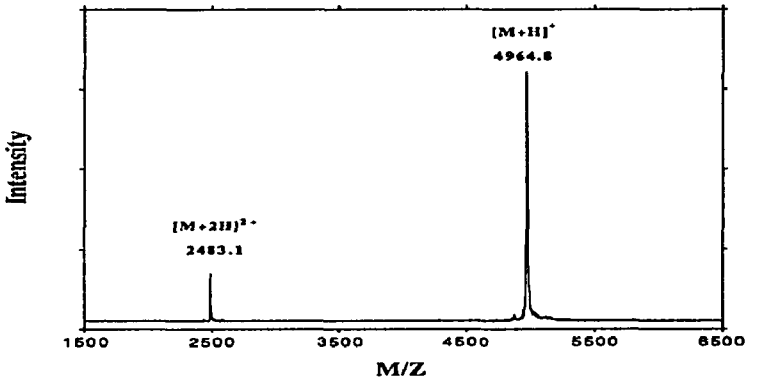


Figure 2.8. Matrix-assisted laser desorption mass spectrum for peptide B after all disulfide bonds have been formed.

are three options to analyze the resulting peptide fragments. First is to sequence the peptides with Edman chemistry. An alternate analysis is to determine the amino acid composition which reveals the amount of sample present and the relative molar ratios of each amino acid. Last choice is mass spectrometry. MS is a new technique that allows the rapid, accurate, and inexpensive analysis of peptides and proteins (Hillenkamp et al., 1991; Chait and Kent, 1992). The sample for MS may be a digestion mixture without further separation. Based on these advantages (see Table 2.3), proteolytic digestion in conjunction

with mass spectrometric peptide mapping has been used in this work to determine the disulfide structure.

A summary of the pattern of blocking groups used for peptides A, B, and C, together with the final folded products found, is given in Table 2.4.

Table 2.3. Comparison of three techniques for analysis of endoproteinase digested peptides.

Techniques	Sequencing	Amino acid analysis	Mass spectrometry
Sensitivity	20-100 pmole	1 nmole	1 pmole
Operating time	1 hour/residue	1.5 hr/sample	< 1 hr/sample
Expenses	expensive	inexpensive	inexpensive
Sample	separated fraction	separated fraction	separated fraction, or digestion mixture

Table 2.4. Summary of the disulfide isomers found after first-step folding of peptides A, B and C.

Peptide	Blocked cysteines	Major folding isomers		Results
		Number	Pattern	
A	5, 6	1	Cys 1-2/3-4	undesired pairing
B	2, 4	1	Cys 1-3/5-6	desired pairing
C	1, 3	3	Cys 2-6; Cys 2-5/4-6; Cys 2-4/ 5-6	mixture

For peptide A

In peptide A, Cys 5 and Cys 6 were blocked with Acm. The first-step folding and oxidation were performed under a wide range of conditions including pH ranging between 4 and 10, different solvents, with or without oxidized/reduced glutathione in aqueous solution, in 15% DMSO, or in 15% DMSO/60% TFE, in the absence or presence of urea, and in the absence or presence of protein disulfide-isomerase (PDI). Only one major product, accounting for 80-85% of the starting peptide, was observed by reversed-phase HPLC. The folded peptide eluted at 20.4 min as compared to the unfolded form which eluted at 23.0 min in HPLC with a 10-40% buffer B linear gradient over 30 min at 1.5 ml/min (data not shown).

Table 2.5. Identification by matrix-assisted laser desorption mass spectra (Figure 2.10) of four HPLC peaks (Figure 2.9) generated in the protease V8 enzymatic digestion of peptide A after first-step oxidation.

Peak Number (Retention Time)	Determined Mass	Calculated Mass [1]	Corresponding Fragments	Disulfide Pairing
A	1045.7	1046.2	[37-45]	_
(3.2 min)	(1845.8)	_	unidentified [2]	_
B	1945.5	1945.2	[14-30]	Cys 3-4
(4.4 min)				
C	1463.3	1462.7	[1-13] [3]	Cys 1-2
(6.8 min)	1735.7	1735.9	[31-45]	_
	(1606.8)	_	unidentified [2]	_
D	1463.5	1462.7	[1-13] [3]	Cys 1-2
(8.8 min)	1928.5 [4]	1945.2 [4]	[14-30]	Cys 3-4

1. Masses are calculated assuming oxidized cysteines.
2. No disulfide pairings matches this mass.
3. These fragments have the same mass but elute with different retention times. The reason for this is not known.
4. The difference between calculated and determined mass, -17, is possibly caused by the loss of NH_3 and the formation of pyr-Glu[14] at the N-terminal.

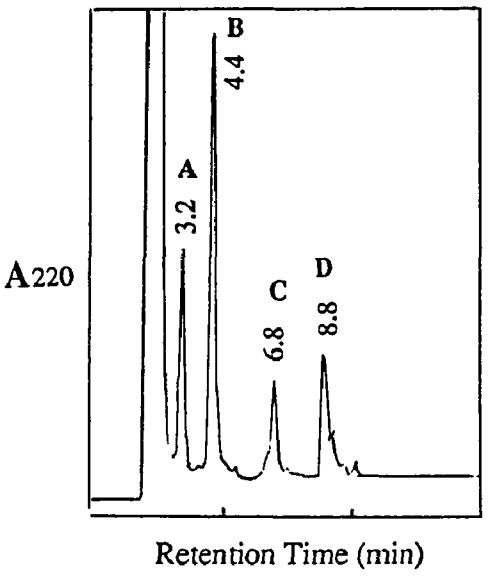


Figure 2.9. Reversed-phase HPLC profile showing the protease V8 digestion for peptide A after first-step oxidation. Peptide fragments were eluted with a 5-20% buffer B linear gradient over 15 min at 1.5 ml/min, monitored at 220 nm. See Materials and Methods for digestion conditions and buffer A and B compositions.

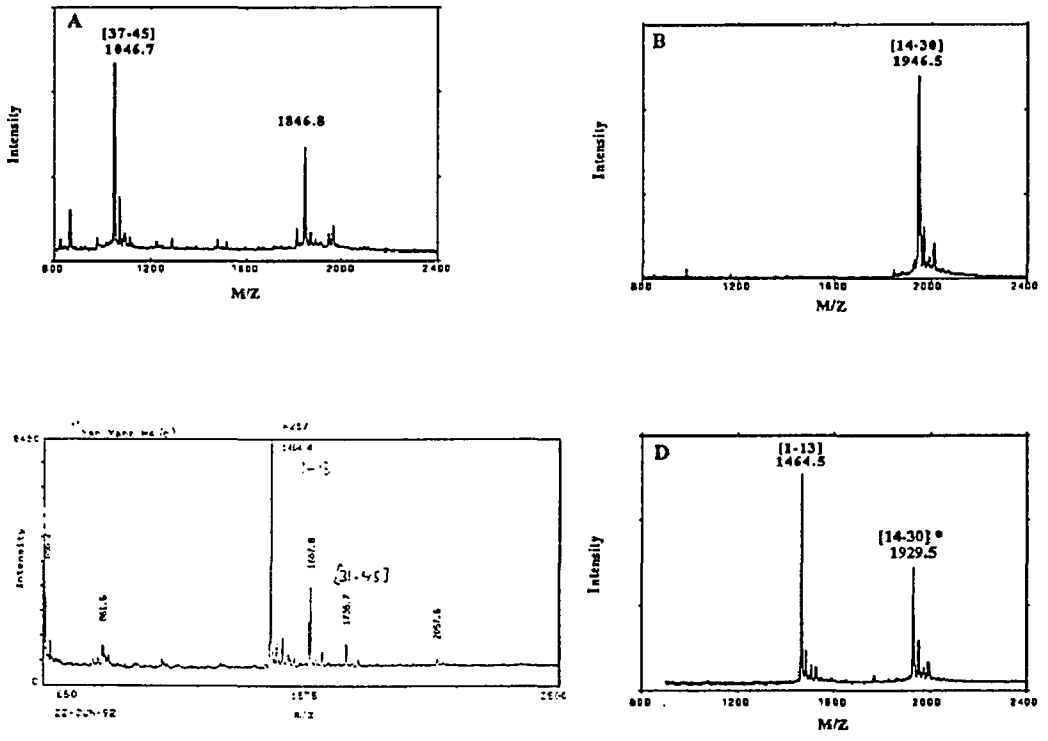


Figure 2.10. Portion of the matrix-assisted laser desorption mass spectrum of peptide A after digestion by protease V8. Fragments, ([1-13], [14-30], [37-45] and [31-45]), are found that indicate the presence of an undesired Cys1-2 and 3-4 disulfide pairing in peptide A after first-step folding with 15% DMSO.

A. Sample from HPLC peak A in Figure 2.9. Fragment [37-45] demonstrates that no disulfide Cys 5-6 is formed.

B. Sample from HPLC peak B in Figure 2.9.

C. Sample from HPLC peak C in Figure 2.9. Only original print-out is available.

D. Sample from HPLC peak D in Figure 2.9. * See Table 2.5 footnote 4.

The connectivity of the disulfide bridges formed by DMSO oxidation was determined by proteolytic digestion in conjunction with mass spectrometric peptide mapping. Enzymatic digestion of this product by *S. aureus* protease V8, which cleaved the peptide after Glu residues, yielded 4 peaks by HPLC (Figure 2.9). Each of the 4 peaks examined by matrix-assisted laser desorption mass spectrometry (Figure 2.10, Table 2.5). If peptide A had adopted an EGF-like pattern (Cys 1-3 and 2-4, while Cys 5 and 6 blocked with Acm), two or three peptide fragments should have been found, ([1-13]+[14-30] and [31-45] or [37-45], see Figure 2.11.A), but the experimental data didn't match this pattern. Instead four peptide fragments were observed, ([1-13], [14-30], [31-45] and [37-45]), corresponding to an undesired isomer having cysteine pairing 1-2 and 3-4 (Figure 2.11.B).

The folding product under air oxidation was also analyzed. The V8 digestion mixture was subjected to matrix-assisted laser desorption mass spectrometric analysis. Three peptide fragments were identified, ([1-13], [14-30] and [31-45]), as shown in Figure 2.12, consistent with a disulfide pairing of Cys 1-2/3-4 in DMSO folding condition. Therefore, both the studies on different folding conditions (DMSO or air oxidation) and the results from different digestion sampling (digestion mixture or separated HPLC peaks) revealed the same unwanted disulfide pattern for peptide A after first-step oxidation.

There was once a question about the proteolytic digestion results: since no cross-linking disulfide bridges were found, was this disulfide structure from an authentic enzyme digestion or because of the effect from some reducing agents possibly packed together with the enzyme V8? To clarify this question, peptide A containing all three disulfides was subjected to the same analysis, V8 digestion. If the reducing agents had existed in the enzyme package, the same digestion results should have been observed by HPLC. Figure 2.13 shows HPLC profile for the digestion mixture of the three-disulfides containing peptide A. It differed from Figure 2.9 for peptide A with two disulfide bridges. Therefore, the possibility of an improper interpretation for digestion data was excluded. Mass spectra are shown in Figure 2.14 (for HPLC peaks A, B and E only). The peptide fragments [1-

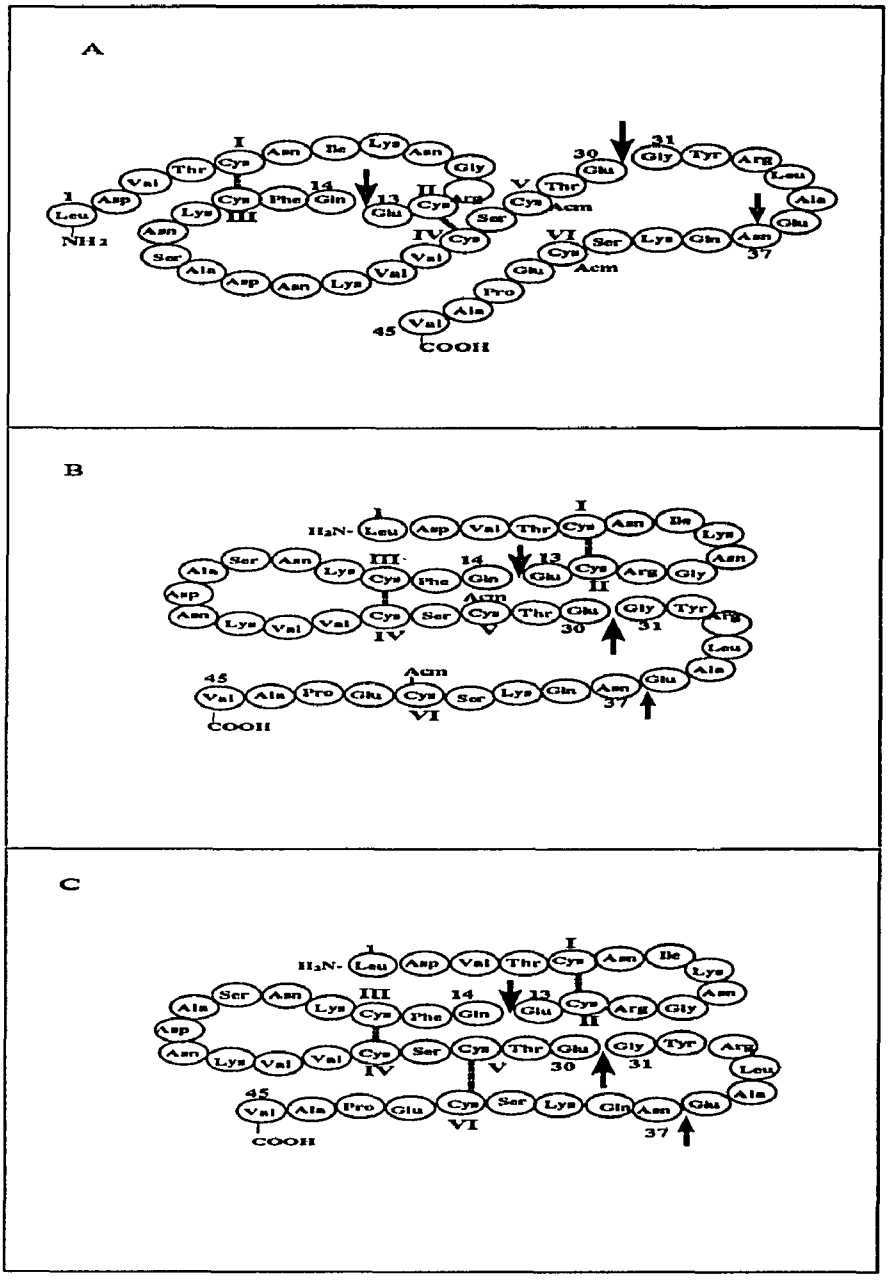


Figure 2.11. The assignment of disulfide bridges located in Peptide A with fIX_{EGF-C} sequence. Arrows indicate the position of *S. aureus* protease V8 cleavage.

A. Expected peptide with EGF pairing after first-step folding;

B. Experimentally found disulfide pattern in peptide A after first-step folding;

C. Experimentally found disulfide pattern in peptide A after second-step folding.

13], [14-30]+[37-45], and [14-45] found after second-step oxidation also indicate that the final folding product was an undesired peptide isomer with Cys 1-2, 3-4 and 5-6 (Figure

2.11.C). Thus, by blocking Cys 5 and Cys 6, the major product was a misfolded disulfide isomer (Figure 2.15).

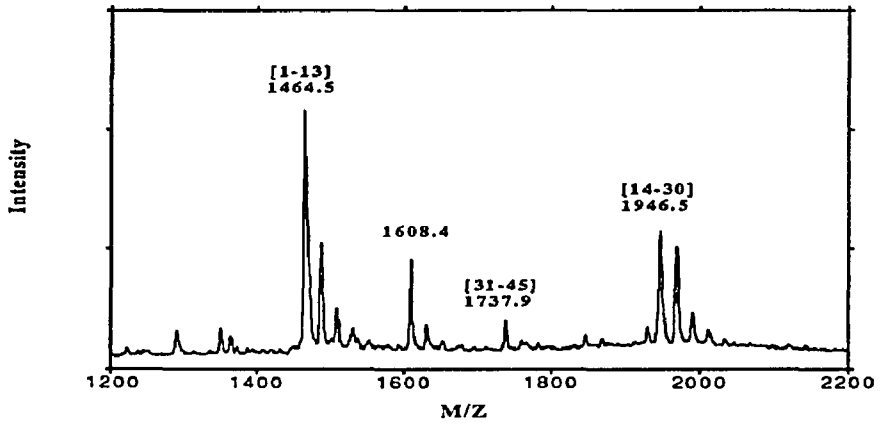


Figure 2.12. Portion of the matrix-assisted laser desorption mass spectrum of peptide A after digestion by protease V8. Fragments [1-13], [14-30] and [31-45] are found that indicate the presence of undesired Cys1-2 and 3-4 disulfide pairing in peptide A after first-step oxidation with air. The sample was an aliquot from the digestion mixture without further separation.

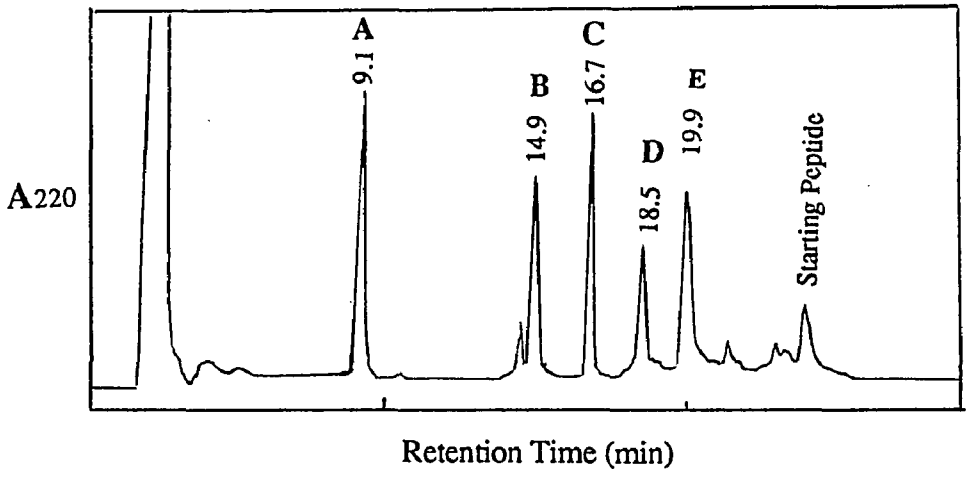


Figure 2.13. Reversed-phase HPLC profile showing the protease V8 digestion for peptide A after second-step oxidation. Peptide fragments were eluted with a 5-33% buffer B linear gradient over 28 min at 1.5 ml/min, monitored at 220 nm. See Materials and Methods for digestion conditions and buffer A and B compositions.

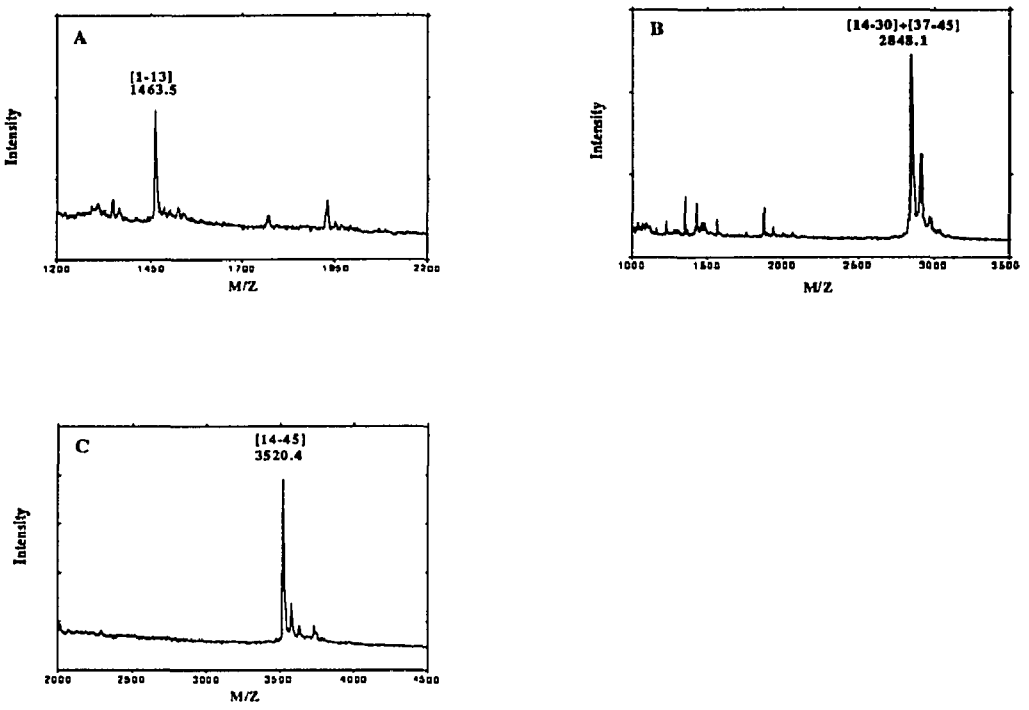


Figure 2.14. Portion of the matrix-assisted laser desorption mass spectrum of peptide A containing three S-S after digestion by protease V8. Fragments [1-13], [14-45] and [14-30]+[37-45] are found that indicate the presence of an undesired Cys1-2, 3-4 and 5-6 disulfide pairing in peptide A after second-step oxidation.
A. Sample from HPLC peak A in Figure 2.13. **B.** Sample from peak B in Figure 2.13. **C.** Sample from HPLC peak E in Figure 2.13.

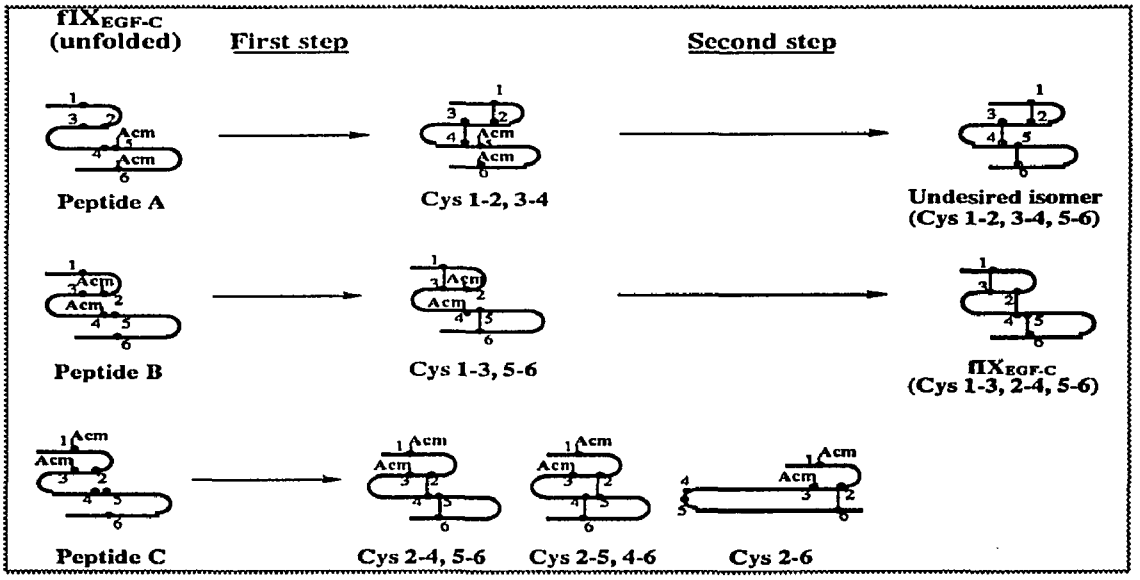


Figure 2.15. The results of two-step selective formation of disulfide bridges in peptides A, B and C. Reproduced from Yang et al. (1994 b), with the permission of the publisher.

For peptide B

Peptide B contained Cys 2 and Cys 4 blocked with Acm. To obtain an EGF-like folding pattern, a Cys 1-3/5-6 pairing would be expected after first-step folding. Under a wide range of folding conditions, similar to those used for peptide A, a single major peak was found by HPLC (Figure 2.16.D). It accounted for 80-85% of starting peptide and had a retention time of 18.1 min as compared to 23.5 min for the linear peptide. Three different enzyme digestions were performed to determine the disulfide pairing pattern (Figure 2.17).

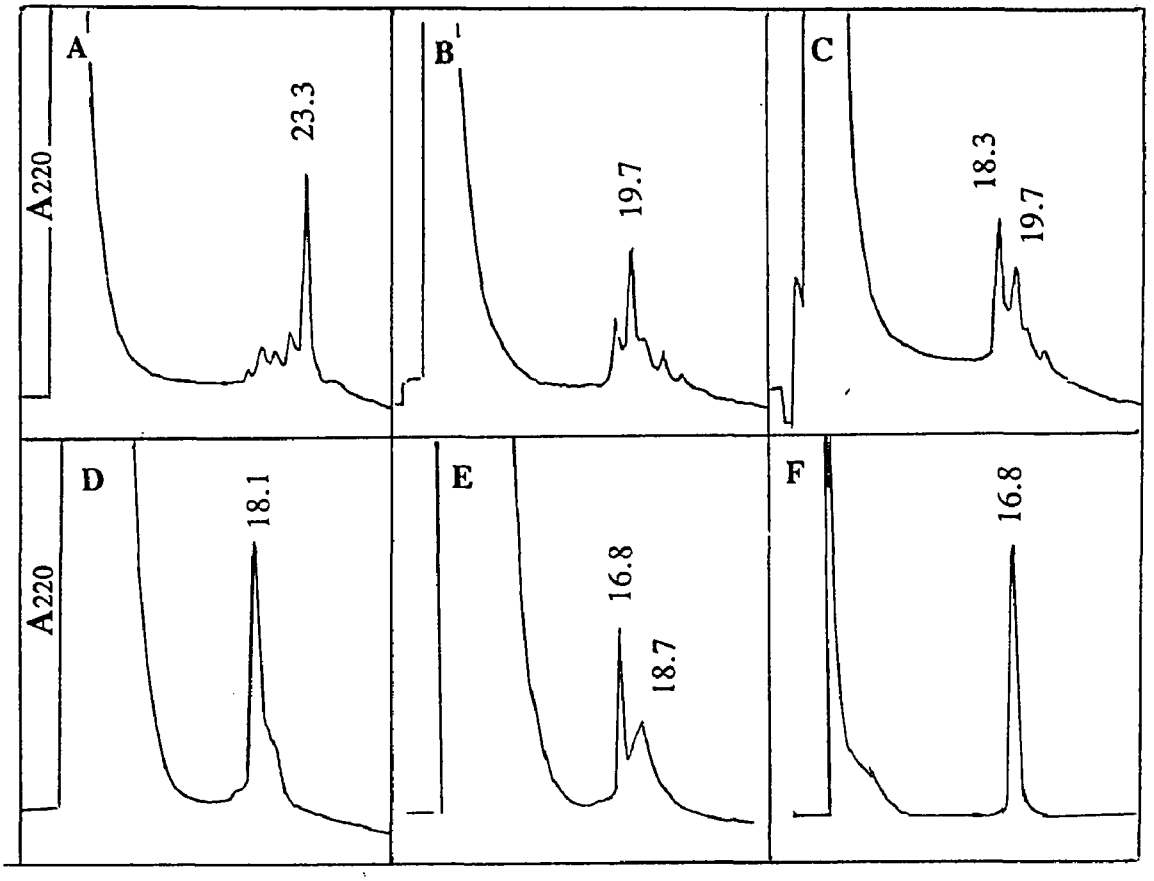


Figure 2.16. Time course plot of the two-step oxidation of three disulfide bridges in $fIX_{EGF\text{-}C}$ monitored by reversed-phase HPLC. Crude peptide was used as starting material, and folded under air in the first step. Peptides were eluted with a 10-35% buffer B gradient over 25 min at 1.5 ml/min, 220nm. See Materials and Methods for folding conditions and buffer A and B composition. Panels A to D represent the first step oxidation and panels E and F are for the second step oxidation.
A. Folding 3 hr; **B.** Folding 18 hr; **C.** Folding 26 hr; **D.** Folding 40 hr (completed); **E.** Removal of Acm with iodine; **F.** Final product after purification with preparative HPLC.

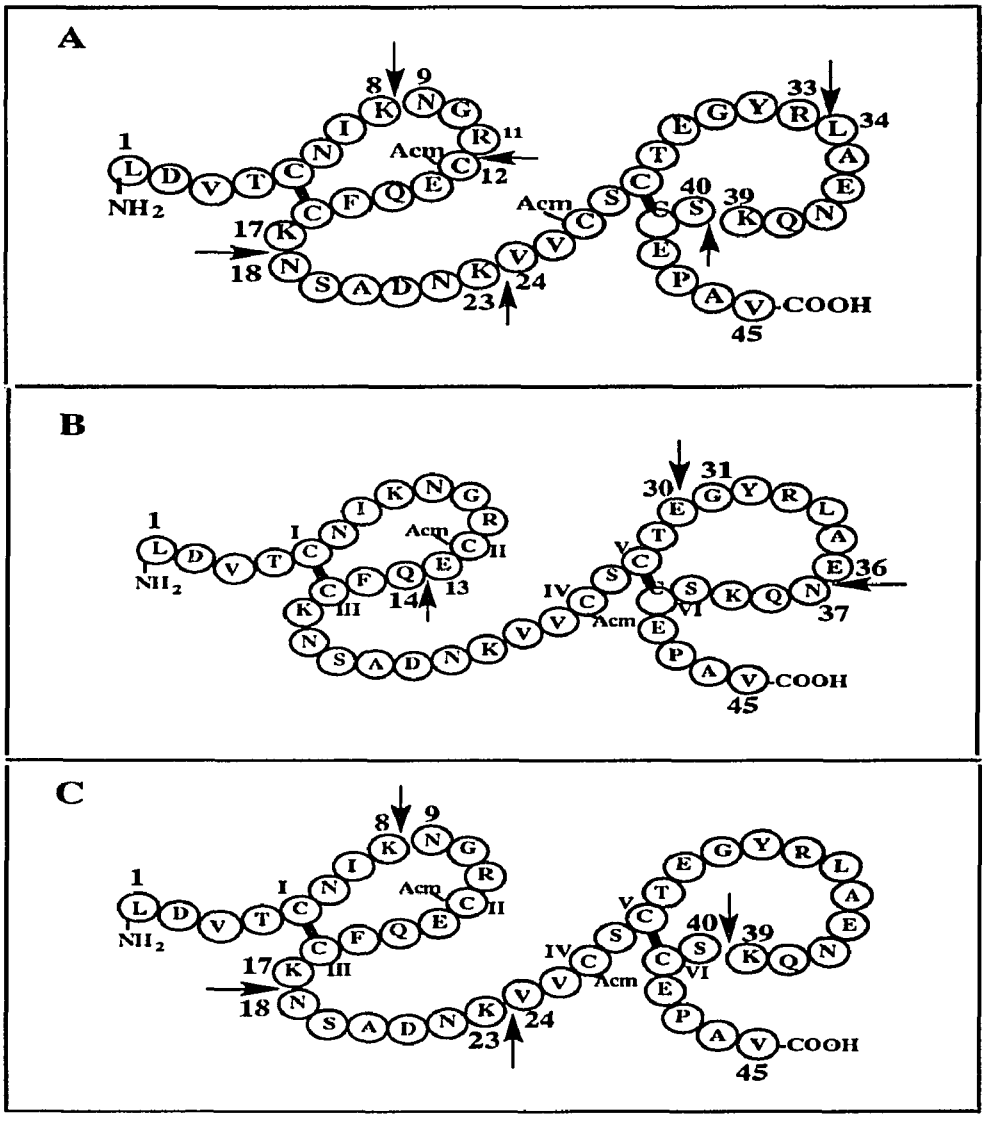


Figure 2.17. The assignment of disulfide bridges located in Peptide B. Arrows indicate the position of protease cleavage.

A. Tryptic digestion;
B. Protease V8 digestion;
C. Endoproteinase Lys-C digestion.

Tryptic digestion, which cleaves the peptide after Lys and Arg, gave one major peak and at least two minor peaks in reversed-phase HPLC (Figure 2.18). The results from mass spectral analysis, are shown in Figure 2.19 and Table 2.6. One of the minor peaks gave a fragment corresponding to either [1-8]+[9-17] or [1-11]+[12-17], indicating a Cys 1-3 bridge. Similarly, the major HPLC peak contained a [1-8]+[12-17] fragment, once again consistent with a Cys 1-3 disulfide bridge. Another of the minor peaks was found to be a misfolded fragment having a Cys 3-5 disulfide link (fragment [12-17]+[24-33]). The misfolded component accounted for less than 15% of the total digestion. To define the other disulfide, protease V8 was used. Figure 2.20 shows HPLC profile for digestion mixture. After V8 digestion, a [1-30]+[37-45] fragment was identified in the digestion mixture, confirming that the Cys 5-6 disulfide was also present (Figure 2.21, Table 2.6).

In order to corroborate these results, the peptide was digested by a third enzyme, endoproteinase Lys-C, which cleaves after lysine. Figure 2.22 shows HPLC profile for digestion mixture. Using this enzyme, both disulfide bridges Cys 1-3 and Cys 5-6 were identified in the digestion mixture (fragment [1-8]+[9-17], measured mass 2058.7+1.0 u, calculated mass 2058.4 u; and fragment [24-39]+[40-45], measured mass 2474.3+1.0 u, calculated mass 2473.8 u) (Figure 2.23). Thus, peptide B gave predominately the properly folded product (Cys 1-3/5-6) (Figure 2.15).

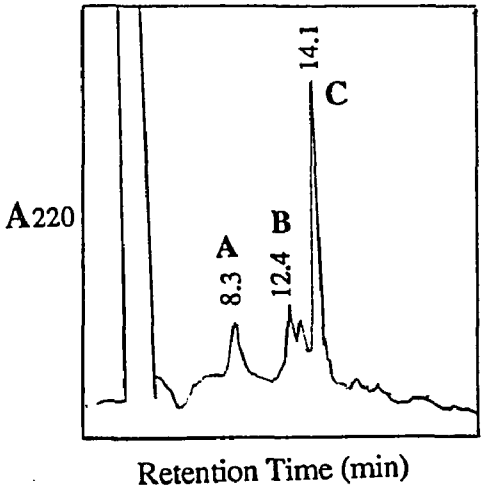


Figure 2.18. Reversed-phase HPLC profile showing the tryptic digestion for peptide B after first-step oxidation. Peptide fragments were eluted with a 10-35% buffer B linear gradient over 25 min at 1.5 ml/min, monitored at 220 nm. See Materials and Methods for digestion conditions and buffer A and B compositions.

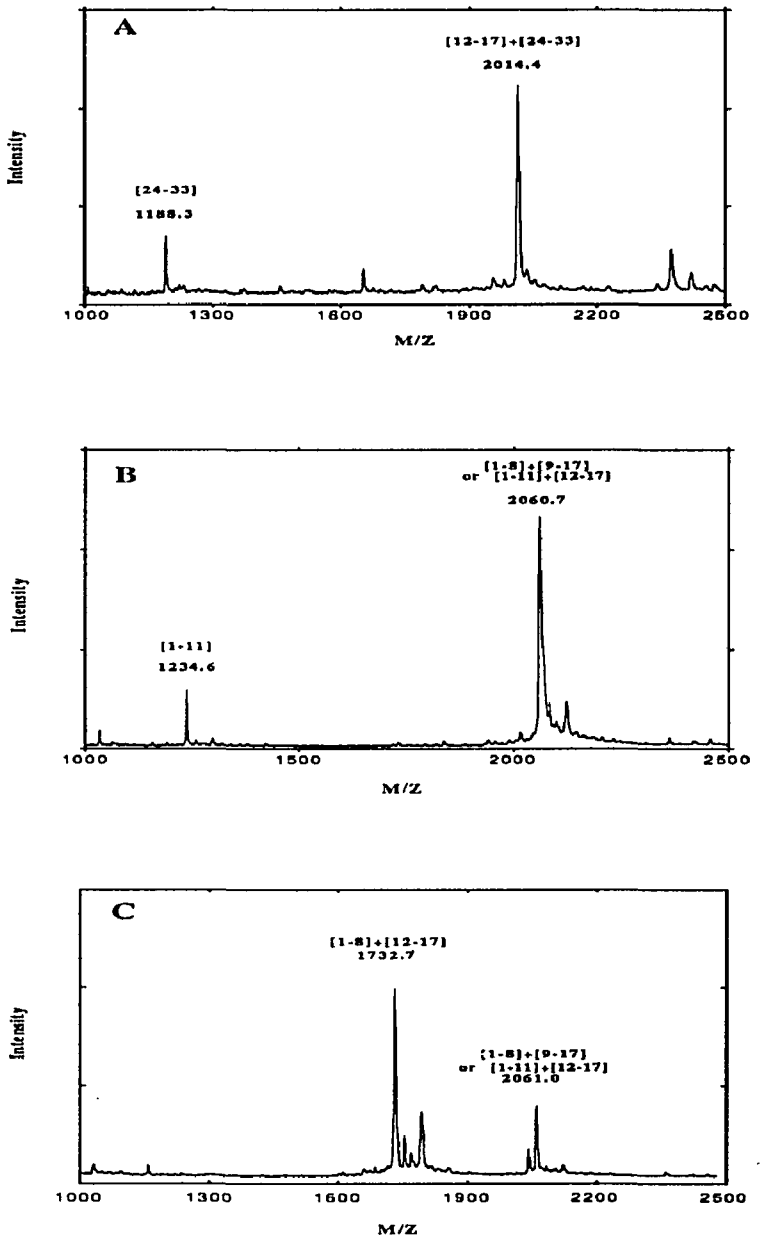


Figure 2.19. Portion of the matrix-assisted laser desorption mass spectrum of peptide B after digestion by trypsin. Fragments [1-8]+[9-17] (or [1-11]+[12-17]) and [1-8]+[12-17] are found that demonstrate the presence of the desired Cys 1-3 bridge (Cys^5-Cys^{16}) in peptide B. Misfolded Cys 3-5 disulfide link (fragment [12-17]+[24-33]) is also found in panel A which is accounted for less than 15% of the total digestion.

A. Sample from HPLC peak A in Figure 2.18.
B. Sample from HPLC peak B in Figure 2.18.
C. Sample from HPLC peak C in Figure 2.18.

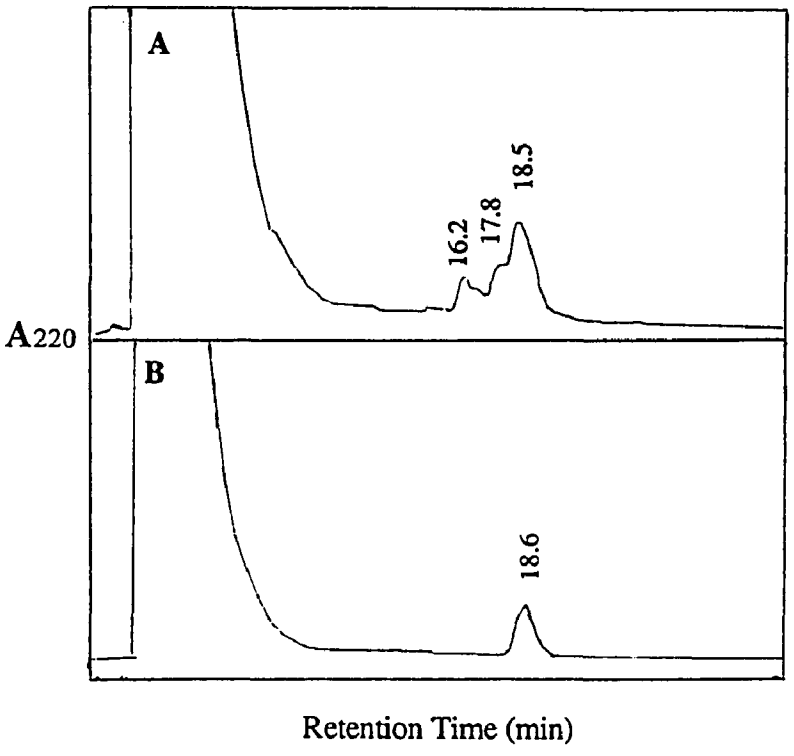


Figure 2.20. Reversed-phase HPLC profile showing the protease V8 digestion for peptide B after first-step oxidation. Peptide fragments were eluted with a 10-40% buffer B linear gradient over 30 min at 1.5 ml/min, monitored at 220 nm. See Materials and Methods for digestion conditions and buffer compositions. A: Enzyme digestion mixture; B: Control (peptide only). The weak intensity of the peak was due to the low amount of sample loaded. This peak serves as an indication of the retention time for starting peptide.

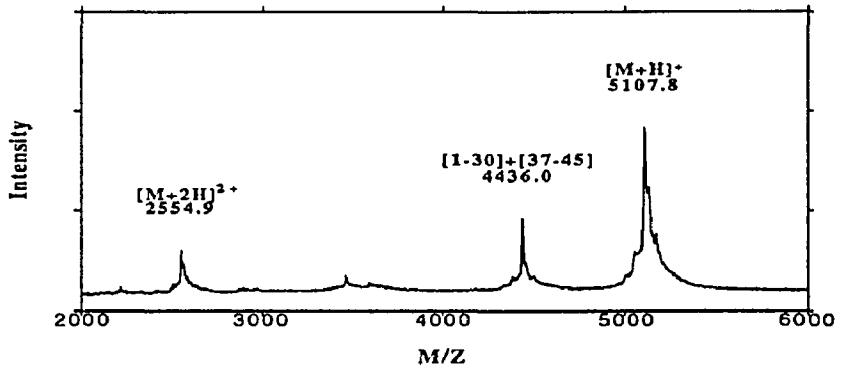


Figure 2.21. Portion of the matrix-assisted laser desorption mass spectrum of peptide B after digestion by protease V8. A [1-30]+[37-45] fragment is found, demonstrating the presence of the desired Cys 5-6 bridge (Cys^{28}-Cys^{41}) in peptide B. Presence of $[M+H]^+$ and $[M+2H]^{2+}$ peaks are due to incomplete digestion.

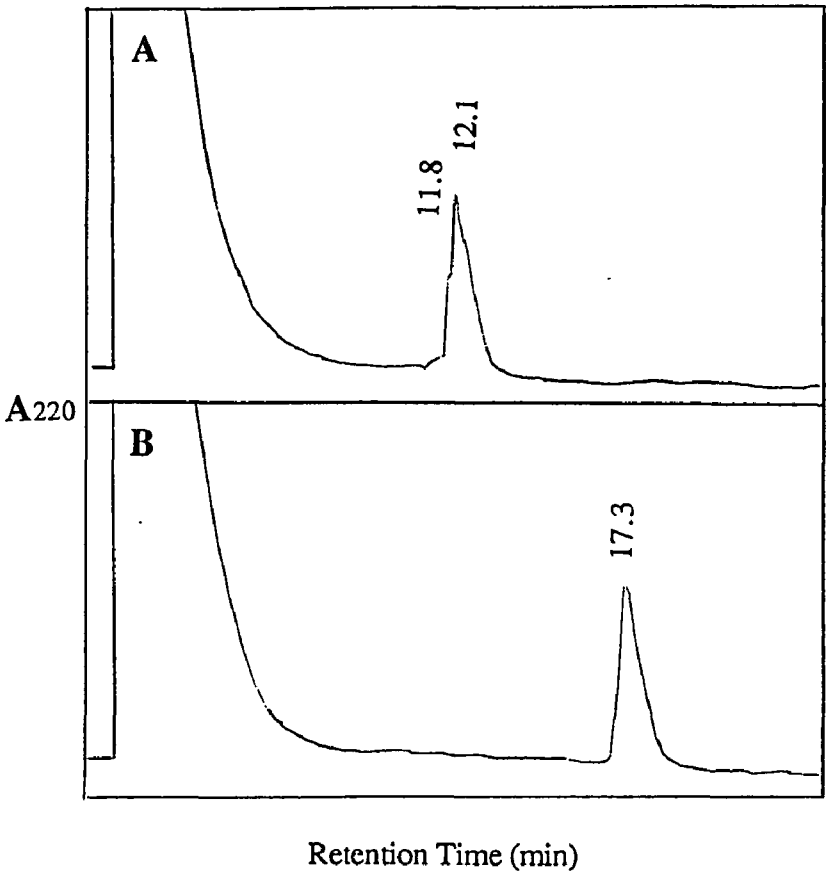


Figure 2.22. Reversed-phase HPLC profile showing the endoproteinase Lys-C digestion for peptide B after first-step oxidation. Peptide fragments were eluted with a 10-35% buffer B linear gradient over 25 min at 1.5 ml/min, at 220 nm. See Materials and Methods for digestion conditions and buffer A and B compositions.
A: Enzyme digestion mixture;
B: Control (peptide only).

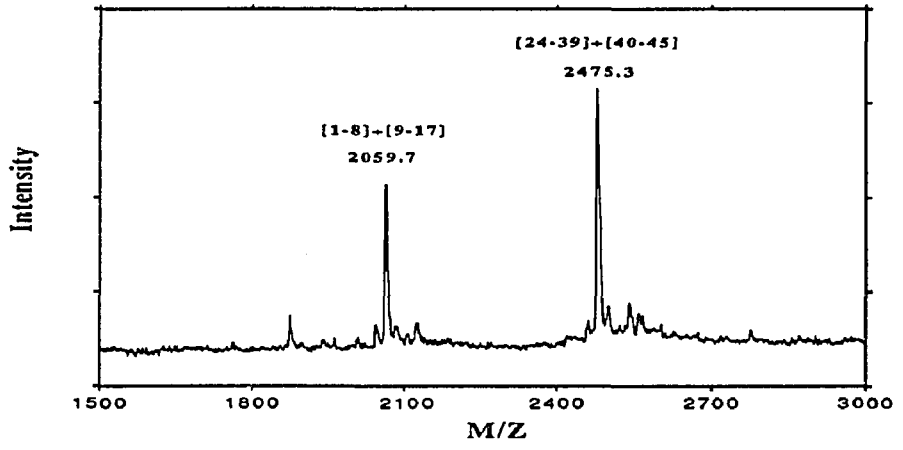


Figure 2.23. A portion of the matrix-assisted laser desorption mass spectrum of peptide B after digestion by endoproteinase Lys-C. Two fragments, [1-8]+[9-17] and [24-39]+[40-45], are found that demonstrate the presence of desired Cys^{5}-Cys^{16}/Cys^{28}-Cys^{41} disulfide pairing in peptide B.

Table 2.6. Identification by matrix-assisted laser desorption mass spectra (Figures 2.19 and 2.21) of HPLC peaks in Figures 2.18 and 2.20 generated in the tryptic digestion and V8 digestion of peptide B after first-step oxidation.

Enzyme	Peak Number (Retention Time)	Determined Mass	Calculated Mass [1]	Corresponding Fragments	Disulfide Pairing
Trypsin (Figure 2.18)	A (8.3 min)	2013.4	2013.3	[12-17]+[24-33]	Cys 3-5
	B (12.4 min)	2059.7	2058.4	[1-8]+[9-17] (or [1-11]+[12-17])	Cys 1-3
	C (14.1 min)	1731.8 2060.0	1731.1 2058.4	[1-8]+[12-17] [1-8]+[9-17] (or [1-11]+[12-17])	Cys 1-3
V8 (Figure 2.20)	digestion mixture	4435.0	4436.0	[1-30]+[37-45]	Cys 5-6

1. Masses are calculated assuming oxidized cysteines.

For peptide C

In peptide C (Cys 1 and Cys 3 blocked with Acm), a mixture of folding products was observed. There were at least 3 peaks in the HPLC profile (Figure 2.24). Tryptic digestion was performed for each of the three HPLC peaks separately. Peak 1, accounting for 28% of the total area under the HPLC peaks, was relatively sharp. A fragment [9-17]+[40-45] was found and identified as an isomer containing only one disulfide bridge (Cys 2-6) and two free cysteines (Cys 4 and Cys 5) (Figure 2.25, Table 2.7). Similar to peaks 2 and 3, peak 1 also contained isomers with two disulfide bridges. Peaks 2 and 3 (23% and 39%) were partially overlapped. These peaks resulted from two isomers with closely matched retention times, differing by about 1 min. Each isomer contained two disulfide bridges. Theoretically, isomers with Cys 2-4/5-6, Cys 2-5/4-6, and Cys 2-6/4-5 were possible, but no fragment consistent with Cys 2-6/4-5 was found (Figure 2.26, Figure 2.27, Table 2.7).

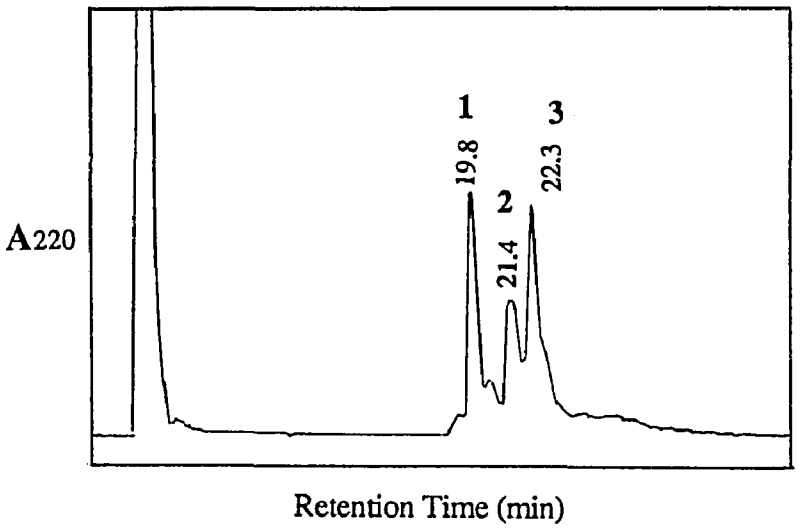


Figure 2.24. Reversed-phase HPLC profile showing peptide C after first-step oxidation. Peak 1, accounting for 28% of the total area. Peaks 2 and 3 are overlapped. Peptide was eluted with a 10-45% buffer B linear gradient over 35 min at 1.5 ml/min, monitored at 220 nm. See Materials and Methods for buffer A and B compositions.

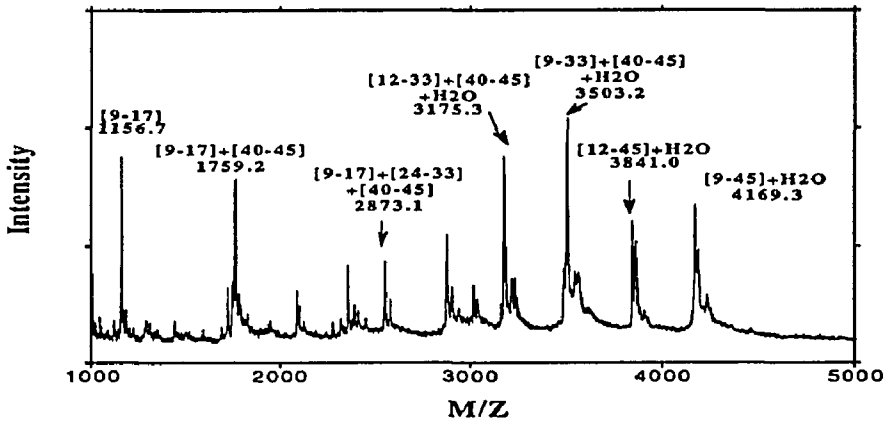


Figure 2.25. Portion of the matrix-assisted laser desorption mass spectrum of peptide C after digestion by trypsin. The sample is from peak 1 in the HPLC profile (Figure 2.24). The [9-17]+[40-45] fragment demonstrates the presence of one disulfide Cys 2-6 isomer (Cys^{12}-Cys^{41}) in peptide C. The presence of other digestion peaks are due to 2 isomers each containing two S-S or incomplete digestion.

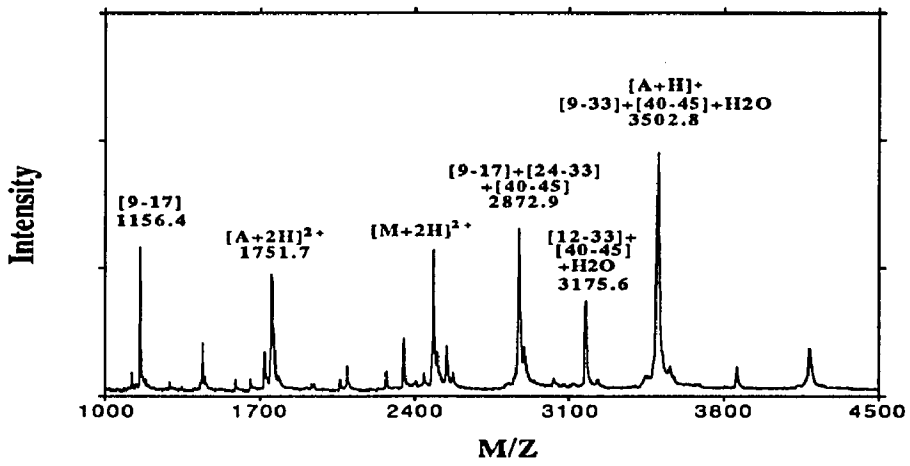


Figure 2.26. Portion of the matrix-assisted laser desorption mass spectrum of peptide C after digestion by trypsin. The sample is from peak 2 in the HPLC profile (see Figure 2.24). A peptide with mass 4964.8 is used for calibration, which is seen as the $[M+2H]^{2+}$ peak with mass 2483.0. Fragments [9-17]+[24-33]+[40-45], [12-33]+[40-45]+H_2O and [9-33]+[40-45]+H_2O are found that demonstrate the presence of 2 isomers each containing two S-S: Cys 2-4/5-6 (Cys^{12}-Cys^{26}/Cys^{28}-Cys^{41}) and Cys 2-5/4-6 (Cys^{12}-Cys^{28}/Cys^{26}-Cys^{41}).

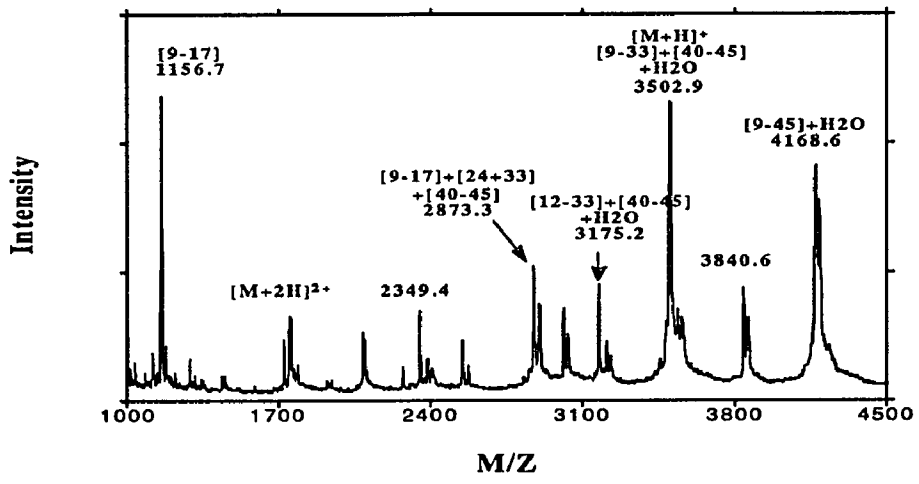


Figure 2.27. Portion of the matrix-assisted laser desorption mass spectrum of peptide C after digestion by trypsin. Sample is from peak 3 in HPLC profile (Figure 2.24). Similar to Figure 2.25 and 2.26, fragments [9-17]+[24-33]+[40-45], [12-33]+[40-45]+H_2O, [9-33]+[40-45]+H_2O and [9-45]+H_2O are found that demonstrate the presence of 2 isomers each containing two S-S. They are Cys 2-4/5-6 (Cys^{12}-Cys^{26}/Cys^{28}-Cys^{41}) and Cys 2-5/4-6 (Cys^{12}-Cys^{28}/Cys^{26}-Cys^{41}) in peptide C.

Table 2.7. Identification by matrix-assisted laser desorption mass spectra (Figures 2.25, 2.26 and 2.27) of three HPLC peaks in Figure 2.24 generated in the tryptic digestion of peptide C after first-step oxidation.

Peak Number (Retention Time)	Determined Mass [1]	Calculated Mass [2]	Corresponding Fragments	Disulfide Pairing Indicated
1	1155.7	1155.3	[9-17]	-
(19.8 min)	1758.2	1758.0 [3]	[9-17]+[40-45]	Cys 2-6
	2872.1	2872.3	[9-17]+[24-33]+[40-45]	Cys 2-4/5-6 (or Cys 2-5/4-6)
28% of total peptide	3174.3	3174.5 [4]	[12-33]+[40-45]+H_2O	Same as above
	3502.2	3501.9 [4]	[9-33]+[40-45]+H_2O	Same as above
2	1155.4	1155.3	[9-17]	-
(21.4 min)	2871.9	2872.3	[9-17]+[24-33]+[40-45]	Cys 2-4/5-6 (or Cys 2-5/4-6)
23% of total peptide	3174.6	3174.5 [3]	[12-33]+[40-45]+H_2O	Same as above
	3501.6	3501.9 [3]	[9-33]+[40-45]+H_2O	Same as above
3	1155.7	1155.3	[9-17]	-
(22.3 min)	2872.3	2872.3	[9-17]+[24-33]+[40-45]	Cys 2-4/5-6 (or Cys 2-5/4-6)
39% of total peptide	3174.2	3174.5 [3]	[12-33]+[40-45]+H_2O	Same as above
	3501.9	3501.9 [3]	[9-33]+[40-45]+H_2O	Same as above

1. The data indicating incomplete digestion are not listed in this table.
2. Masses are calculated assuming oxidized cysteines.
3. This mass is calculated assuming only two cysteines oxidized.
4. Mass 18, named H_2O for convenience, is assembled if the digestion cleavage occurs in a peptide loop. Although new C- and N-terminals are produced, the two fragments are still linked together by disulfide bridge.

Fragments [9-17]+[24-33]+[40-45], [12-13]+[40-45]+H_2O and [9-33]+[40-45]+H_2O were generated from 2 isomers each containing two S-S. However, the Cys 2-4/5-6 and Cys 2-5/4-6 could not be distinguished by tryptic digestion because trypsin did not cleave between Cys 4 and Cys 5 (Cys^{26}-Ser^{27}-Cys^{28}). Thus, the desired Cys 2-4/5-6 was present in either peak 2 or peak 3, although with the misfolded Cys 2-5/4-6 isomer contained under the other peak. Attempts to use partially purified fractions to proceed with the second-step

folding failed to yield a significant amount of peptide containing the desired disulfide pairing pattern.

2.3.2.3. Disulfide oxidation in second-step oxidation and folding

The second-step oxidation was performed by using iodine in acidic solution to remove the Acm. The oxidation was under kinetic control and was usually completed within an hour. The determination of the location of the all three disulfide bridges after formation of the third disulfide bond was not achieved. Attempts were made to locate these bonds using tryptic and V8 digestion, but the required cleavage between Cys 4 and Cys 5 (Cys^{26}-Ser^{27}-Cys^{28}) was not accomplished. In any case, however, thiol-disulfide exchange in this final step is prevented by the acidic condition (pH ≤ 2) employed. Figure 2.28 shows a chromatogram of the second-step folding process for peptide B. (Also see Figure 2.16 panels E and F which show a similar profile except that the starting peptide was crude). A single major peak accounting for 60% of the starting peptide is observed by reversed-phase HPLC in the second step oxidation.

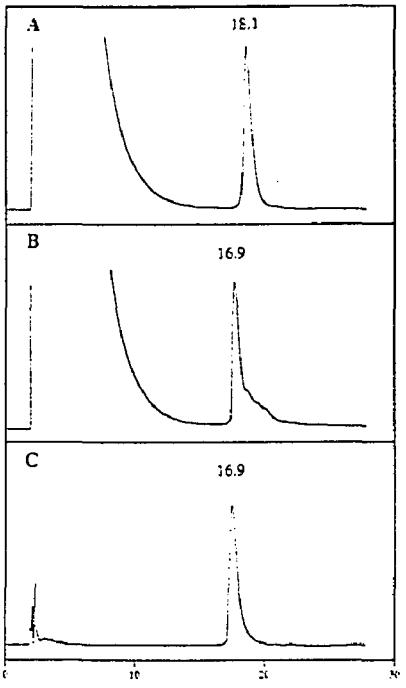


Figure 2.28. Reversed-phase HPLC profile showing the formation of the third disulfide bridge for peptide B. Peptides were eluted with a 10-40% buffer B linear gradient over 30 min at 1.5 ml/min, monitored at 220 nm. Buffer A contained 5% acetonitrile in 0.045% TFA. Buffer B contained 60% acetonitrile in 0.037% TFA.
A. Purified peptide B with two disulfide bridges formed after the first-step.
B. Peptide B after the third disulfide bridges was formed by removal of Acm and oxidation in 10% HAc solution with 5 mM iodine. Sodium thiosulfate (10 mM) was added to react with the excess iodine.
C. The final product with three disulfide bridges purified by preparative HPLC. (Reproduced from Fig. 7 in *Protein Science* 3, 1267-1275, 1994, with the permission of the publisher).

2.3.3. Synthesis of the N-terminal of EGF-like domain in factor X (fX_{EGF-N})

The results from the step-wise formation of disulfide bonds in the synthesis of fIX_{EGF-C} have been applied for synthesis of a number of other peptides (e.g. Spetzler et al, 1994). An example is the N-terminal EGF-like domain in human factor X (fX_{EGF-N}).

Before the current two-step method was employed for fX_{EGF-N}, an attempt to synthesize the peptide had been made through a conventional single-step approach. There were a couple of reasons supporting this synthetic design. First, both N-terminal EGF-like domains in factors IX and X have the same loop size and homologous sequences differing by only a few residues (Appendix A). Secondly, fIX_{EGF-N} had a well-behaved manner in folding (Huang et al., 1989). Consequently, peptide fX_{EGF-N} should have had similar folding ability to that of fIX_{EGF-N}. However, the peptide resulting from a conventional single-step method displayed multiple overlapping peaks in the HPLC (data not shown). It was difficult to determine which peak with the desired disulfide pairing and it was also not easy to separate them. Although many attempts were made we were unable to identify the disulfide pattern for this peptide. NMR studies also indicated the HPLC-pure fraction was NMR-impure (see section 2.3.3.5). Therefore, a two-step method was employed to obtain pure and properly disulfide-paired peptide.

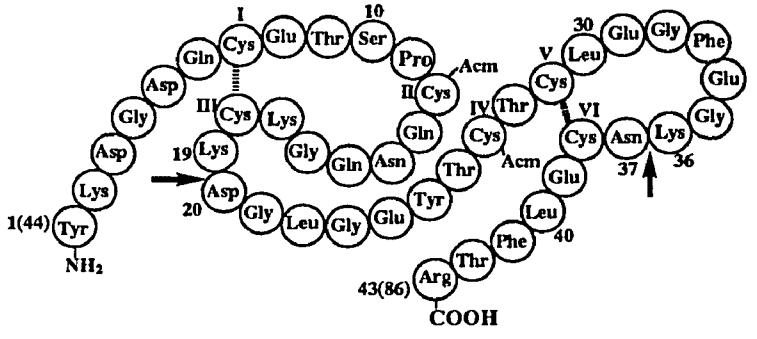


Figure 2.29. Primary structure of the N-terminal EGF-like domain in human blood coagulation factor X (44-86). The numbering system used is based on the synthetic peptide, with the corresponding factor X sequence shown in parenthesis. The synthetic peptide contains aspartic acid in place of the naturally occurring β-hydroxyaspartic acid in position 20. Following the two-step method, cysteines 2 and 4 are blocked with Acm. Arrows show where endoproteinase Lys-C cleaves the peptide.

The sequence of $fX_{EGF\text{-}N}$ (Leytus et al., 1986), in which cysteines 2 and 4 are blocked with Acm, is shown in Figure 2.29. The synthetic peptide contains aspartic acid in place of the naturally occurring β-hydroxyaspartic acid in position 20.

2.3.3.1. Peptide with desired disulfide pairing found predominantly after first-step oxidation

Various conditions were used in the first step oxidation, including pH ranging between 4 and 10, different solvents, with or without oxidized/reduced glutathione in aqueous solution or in 15% DMSO. The folding results did not change much within the range of given conditions. In the first step oxidation the peptide folded very fast compared with $fIX_{EGF\text{-}C}$. The time for $fX_{EGF\text{-}N}$ to complete first-step folding was less than 5 hr at room temperature instead of more than 24 hr for $fIX_{EGF\text{-}C}$. Figure 2.30.A presents HPLC profiles for first-step under air oxidation. Figure 2.30.B shows second-step folding process. After first-step oxidation the peptide exhibited three peaks by HPLC. There were two small peaks, named peak 1 and 2, and one dominant peak, named peak 3. The retention times were 15.4 min, 16.1 min and 17.0 min respectively, as compared to reduced peptide eluted at 23.4 min. The relative areas of the three peaks were 4%, 17%, and 78%, as determined by HPLC. It was not difficult to separate them from each other.

Disulfide pairing of the major peak after first-step folding of $fX_{EGF\text{-}N}$ was identified by endoproteinase Lys-C digestion, which cleaves after Lys, in conjunction with mass spectrometric peptide mapping. To determine if the dominant peak adopted the EGF pattern, the purified peak 3 was subjected to proteolytic analysis. The HPLC profile for the digestion mixture is shown in Figure 2.31 and peptide mapping results are in Figure 2.32 and Table 2.8. Two fragments, [20-43] and [20-36]+[37-43], were observed, which identified the Cys 5-6 disulfide bridge as existing in this peptide. This, in turn, indicates that the Cys 1-3 should also be present. In fact, the formation of the Cys 1-3 bridge was

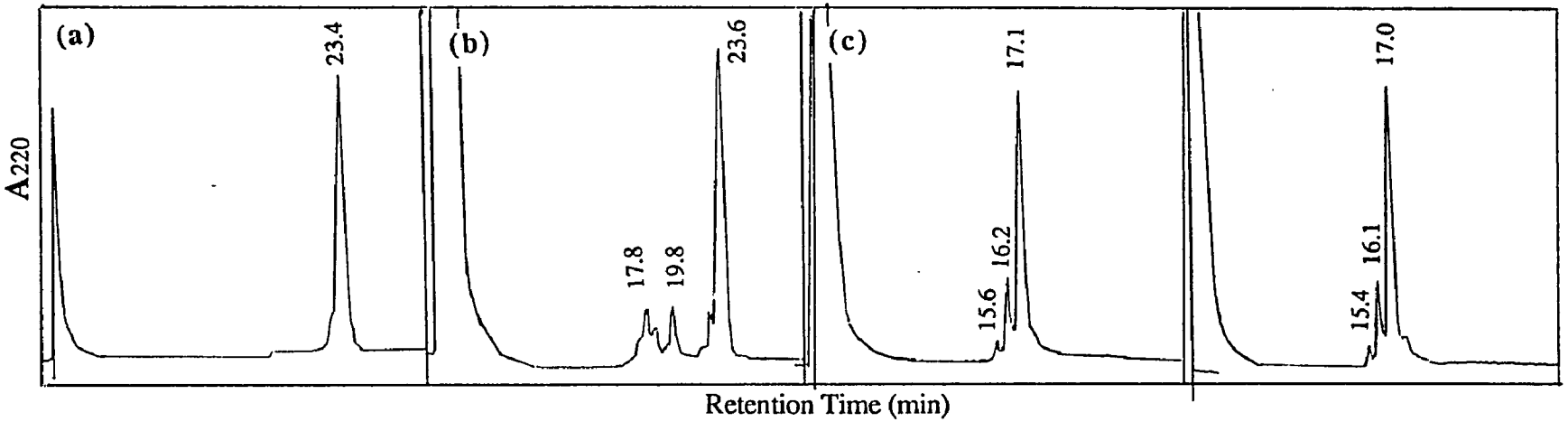


Figure 2.30.A. HPLC profile showing the folding process in the first-step oxidation for $fX_{EGF\text{-}N}$. Conditions: air oxidation, pH 8.2, 10 mM Tris-HCl buffer, and at room temperature. Peptide was eluted with a 20-50% buffer B linear gradient over 30 min at 1.5 mL/min, monitored at 220 nm. See Materials and Methods for buffer A and B compositions.
(a) Reduced and purified peptide before the folding; (b) Folding 5 min; (c) Folding 4 hr; (d) Folding 24 hr.

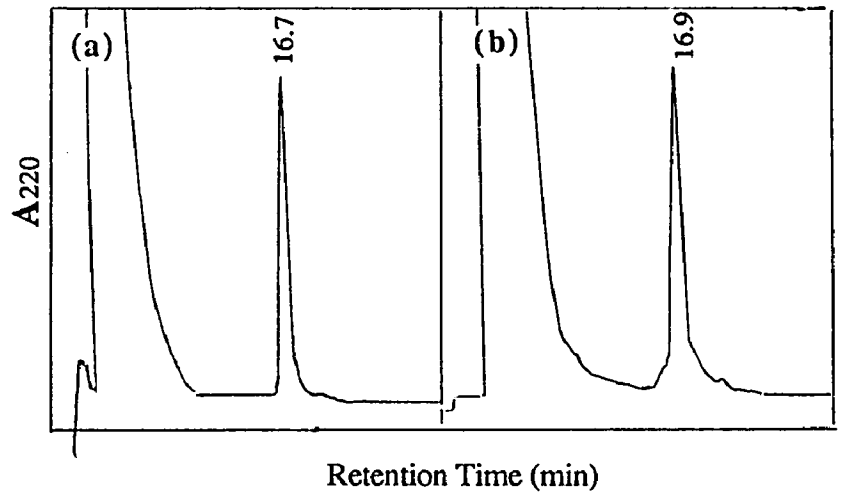


Figure 2.30.B. HPLC profile showing the formation of the last disulfide bridge in the second-step folding for $fX_{EGF\text{-}N}$. Peptide was eluted with a 20-45% buffer B linear gradient over 25 min at 1.5 mL/min, monitored at 220 nm. See Materials and Methods for buffer A and B compositions.
(a) The purified peptide containing two disulfide bridges after first-step oxidation;
(b) The resulting peptide after second-step oxidation.
Conditions: 10% HAc solution with iodine, 5 mM I_2, and then 10 mM $Na_2S_2O_3$ for react the excess I_2.

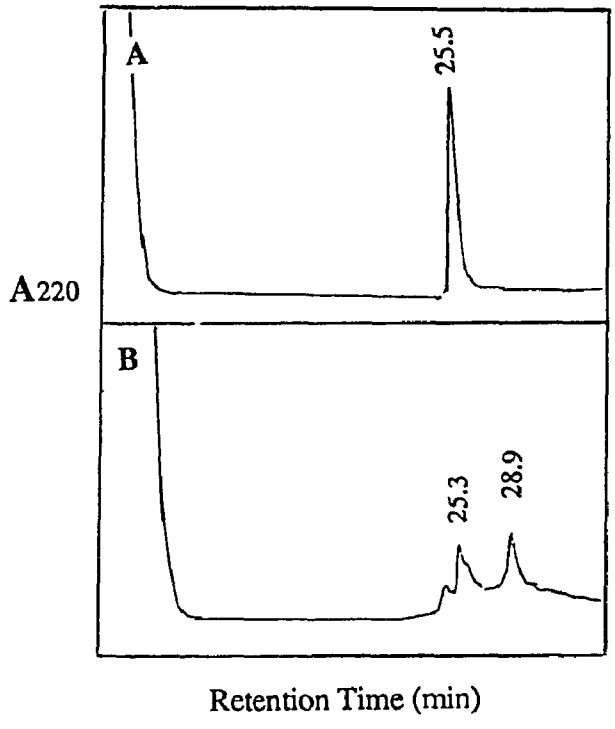


Figure 2.31. Reversed-phase HPLC profile showing the endoproteinase Lys-C digestion for $fX_{EGF\text{-}N}$ after first-step oxidation. Peptide fragments were eluted with a 10-45% buffer B linear gradient over 35 min at 1.5 ml/min, monitored at 220 nm. See Materials and Methods for digestion conditions and buffer A and B compositions. A: Control (peptide only); B: Enzyme digestion mixture.

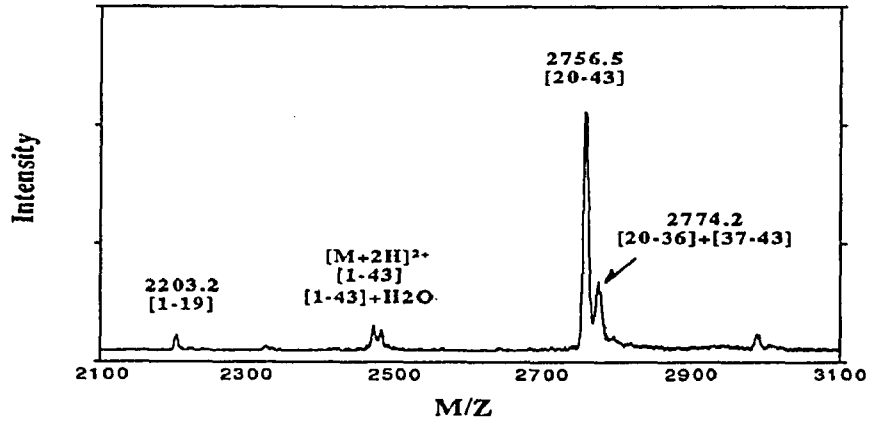


Figure 2.32. Portion of the matrix-assisted laser desorption mass spectrum of $fX_{EGF\text{-}N}$ containing two disulfide bonds after digestion by endoproteinase Lys-C. Two fragments, [20-43] and [20-36]+[37-43], are observed that demonstrate the presence of the desired Cys 5-6 bridge (Cys^{29}-Cys^{38}) in peptide $fX_{EGF\text{-}N}$. The fragment [1-19] shows the presence of another desired disulfide bridge (Cys^{7}-Cys^{18}).

Table 2.8. Disulfide pairing after first-step folding of fIX_{EGF-C} identified by endoproteinase Lys-C digestion in conjunction with mass spectrometric peptide mapping (Figure 2.32).

Determined Mass	Calculated Mass [1]	Fragments	Disulfide Pairing
2202.2	2201.3	[1-19]	Cys 1-3
2755.5	2754.9	[20-43]	Cys 5-6
2773.2	2772.9	[20-36]+[37-43]	Cys 5-6

1. Masses are calculated assuming oxidized cysteines.

verified by the presence of fragment [1-19], a small peak with mass 2202.2 ± 1.0 u. (The reason for different intensities from similar amounts of peptide digestion fragments is unknown). It is clear that the desired disulfide bridges (Cys 1-3, 2-4, and 5-6) have been formed by this method and peptide fX_{EGF-N} has been obtained in high yield. Thus, this work demonstrates a practical strategy to block Cys 2 and Cys 4 with Acm in order to obtain a peptide with an EGF-like pattern.

2.3.3.2. Second-step

The second step folding was carried out and monitored by HPLC as described before. The results are shown in Figure 2.30.B. There were no significant changes in retention time for HPLC peaks before and after the second step (see panels a and b). Nevertheless, the difference of the peptides after first-step and second-step was obvious in terms of removal the Acm blocking groups. The completion of second-step folding was illustrated by determining the molecular mass for the final product. The major peak in Figure 2.30.B (b) was collected and the mass was found to be 4793.4 u (calculated mass for the oxidized peptide without Acm groups is 4794.3 ± 1.0 u). Thus, the second step oxidation in the synthesis of the N-terminal EGF-like domain in factor X was successful.

2.3.3.3. Reversibility of reduced and oxidized peptides in first-step folding

There were at least two small peaks besides the major peak observed in the HPLC profile in the first-step oxidation of $fX_{EGF\text{-}N}$. Since the possibility that small folded peaks were from impure starting peptide could not be excluded, a refolding experiment was carried out to investigate the reason (Figure 2.33). One of the small peaks, peak 2 in HPLC Figure 2.33.A, was purified, reduced with DTT, and then refolded under the previous first-step oxidation conditions. Interestingly, starting from a reduced pure peak 2, the same folding products, two small peaks and one large peak, were observed. The composition of three peaks in both folding and refolding experiments were identical (peak 1 : peak 2 : peak 3, 5:17:78%). Therefore, we could rule out the possibility that the appearance of unwanted folding peaks in the HPLC profile was due to impure starting peptide. What we observed was not only the reversibility between reduced peptide and oxidized peptide, but also the equilibrium existed between the three folding isomers. Because the air oxidation and the mixed disulfide interchange method are slow processes, they allow equilibration of different conformers to produce thermodynamically-controlled products.

2.3.3.4. Folding conditions

Various oxidation conditions were tested as described above. The ratio of the three isomers varied slightly with folding conditions. Table 2.9 shows the studies on thiol-disulfide interchange method. The composition of folded peaks didn't depend on the ratio of oxidized and reduced glutathione (GSSG and GSH), but instead on the concentration of GSSG. Results indicated that the lower the concentration of GSSG used, the higher the percentage of major peak (peak 3) that could be obtained. Therefore, the thiol-disulfide interchange reagent does not produce a higher yield of peak 3. Both air oxidation and 15% DMSO could be used for folding. Air oxidation gave a slightly lower yield of peak 3 than DMSO oxidation, 5% less, but offered the advantage that no separation from DMSO was necessary. The folding was actually performed under air and pH 8 Tris-HCl buffer.

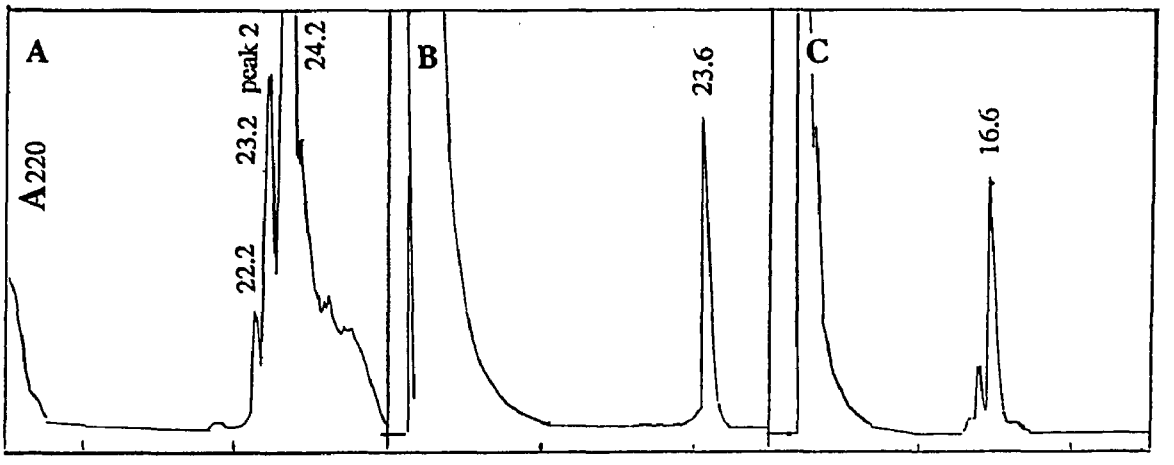


Figure 2.33. Reversed-phase HPLC profile illustrating the reversibility of peptide $fX_{EGF\text{-}N}$ folding in first-step oxidation. Panel A is a semi-preparative HPLC profile, C18 column. Peptide was eluted with a 15-45% buffer B linear gradient over 30 min at 2.5 ml/min, monitored at 220 nm. Panels B and C are profiles by analytical HPLC, C18 column. Peptide was eluted with a 20-45% buffer B linear gradient over 25 min at 1.5 ml/min, monitored at 220 nm. See Materials and Methods for buffer A and B compositions.

A. Separation of HPLC peaks. Peak 2 was collected for further refolding experiments shown in panels B and C. The retention time range in the profile was 5-30 min.

B. The peptide from the peak 2 was reduced with DTT.

C. Starting from the reduced peptide of peak 2 fraction, the same resulting peaks and the similar composition of three peaks is observed after the first-step folding.

Table 2.9. Composition of three disulfide isomers observed by HPLC. The first-step oxidation for $fX_{EGF\text{-}N}$ was performed using oxidized and reduced glutathione (GSSG and GSH), folding 44 hr, at room temperature and pH 8.

Concentration (mM)		Ratio	Percentage of isomers		
GSSG	GSH	GSSG / GSH	peak 1	peak 2	peak 3
5	1	5	9	29	62
10	1	10	13	34	53
10	0.5	20	10	34	56
1	5	0.2	8*	27*	65*
1	10	0.1	7	19	74
0.5	10	0.05	8	16	76
1	1	1	5	22	73
5	5	1	9	32	59

*This data was an exception. Otherwise, all results indicate that, the less concentration of GSSG was used, the higher percentage of peak 3 could be obtained.

2.3.3.5. 1D NMR spectra

To demonstrate the advantage of our two-step method over the traditional one-step approach, proton NMR studies were carried out (Figure 2.34). The NMR samples were prepared with the peptides which have the same $fX_{EGF\text{-}N}$ sequence but made by different approaches: conventional single-step method and the two-step method. The peptides were measured under pH 4.2 in D_2O and at room temperature. The proton NMR spectrum of the peptide from a single-step folding showed broad peakwidth and low resolution, indicating a mixture of disulfide isomers. In contrast, the spectrum from our two-step peptide shows much higher resolution.

2.3.3.6. Unsuccessful enzymatic digestions revealed tightly folded peptide made by single-step approach

A number of enzymatic cleavages for a $fX_{EGF\text{-}N}$ sample, HPLC-pure peptide with three disulfide bridges formed by the conventional single-step method, were performed (data not shown). However, no significant information for identifying the location of disulfide bridges was provided by proteolysis, since fewer cleavages were observed than what were expected in the peptide. Sequential enzyme cleavages were also carried out. It was based on the proposal that after a few cuts had been made on the surface of the peptide by the first enzyme, partial denaturation might expose more sites for the second enzyme cleavage. However, this also failed. One known difficulty is that, Thr is the residue between Cys 4 and 5, which cannot be cleaved by any specific protease. However, appearing to be resistant to specific enzymes and nonspecific cleavage reagents might be due to the peptide fold. This problem is sometimes encountered with proteins that are tightly coiled. All those unsuccessful proteolysis probed significantly the tightly folded structure of this peptide.

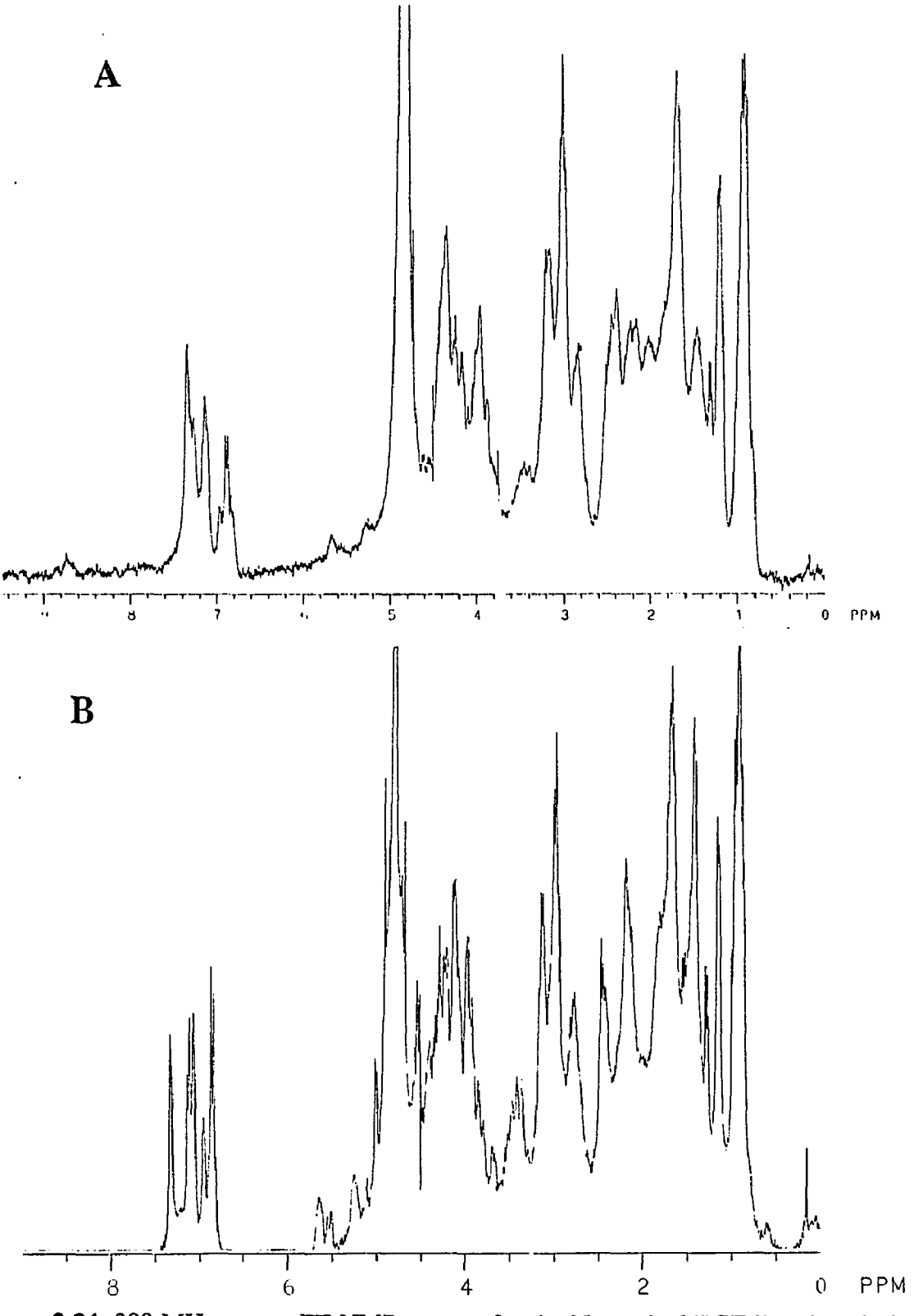


Figure 2.34. 300 MHz proton FT-NMR spectra for the N-terminal EGF-like domain in factor X at pH 4.2, 25°C in D_2O.
A. The peptide made with a conventional single-step approach;
B. The peptide made with the two-step approach.

2.4. Discussion

2.4.1. The necessity for developing a two-step approach

EGF-like domains are common mosaic blocks in proteins. The results from the synthesis $fIX_{EGF\text{-}C}$ and $fX_{EGF\text{-}N}$ show that not all such domains are capable of independently folding to tertiary structures with EGF-like disulfide pairings. There are a number of reasons to expect that a synthetic peptide may not adopt the native peptide folding pattern to give the desired disulfide isomer. Natural peptides can fold sequentially as they come off the ribosomes, perhaps with the aid of chaperone peptides (Ellis, 1994). Also, the folding of individual domains within a larger protein may be influenced by the presence of the rest of the protein. Both of these factors are absent in the folding of synthetic peptides. A two-step strategy has been developed to aid the formation desired disulfide bonds when it is not possible using the conventional chemical approach of simultaneously forming all three disulfides.

The number of disulfide bridges in single-chain proteins varies from 0 to more than 12 (Mao, 1989). A molecule containing N disulfide bridges can be cross-linked intramolecularly in $(2N)!/(2^N N!)$ different ways. An EGF-like peptide contains three disulfide bridges and therefore has 15 different possible disulfide cross linking patterns (Figure 2.3).

It is not uncommon to encounter problems in inducing synthetic peptides to fold correctly. Several groups have reported that misfolded isomers are found together with the desired disulfide-pairing peptide during synthesis of EGF-like domains when a single-step folding approach is used (Violand et al., 1991; Hunter et al., 1993). A coincident problem is the separation and identification of the isomers, always a difficult task. Also, determination of the location of the three disulfide bridges in a given isomer can be particularly challenging. Thus, significant advantages are achieved by the present two-step approach to disulfide bridge formation for synthetic peptides.

In addition to the peptide synthesis described in this dissertation, this two-step approach has been used to synthesize a malaria erythrocyte protein known as a K1 that has an EGF-like disulfide structure (Spetzler et al., 1994). This strategy has also been shown to be successful in the synthesis of a neurotoxin (scratcher peptide) that contains a 1-6, 2-4, and 3-5 disulfide structure (C. Rao and J.P. Tam, unpubl. data).

2.4.2. The choice of where to place Acm

The results of this work demonstrate that the choice as to which cysteines to block with Acm is important. Only one of the three possible choices produces a properly folded product in high yield (Table 2.4, Figure 2.15). To rationally design the synthesis, it is necessary to make an educated guess concerning the most probable disulfide pairing patterns to expect for a given placement of the Acm blocking groups.

There are three patterns for a peptide containing four Cys. A cartoon has been included in Figure 2.35, showing disulfide patterns 1-2/3-4, 1-3/2-4, and 1-4/2-3. In general, the 1-2, 3-4 pattern of two intrachain disulfide bridges appears to provide much less synthetic difficulty than the other patterns, implying a preference of synthetic peptides to adopt this pairing pattern. A survey (Benham and Jafri, 1993) of disulfide bonding patterns among native peptides with two disulfide bonds, found that the cysteine pairing pattern 1-2/3-4 was by far the most common. However, an examination of peptide synthesis papers reporting difficulty with disulfide formation (i.e., those using a sequential approach for disulfide bond formation as opposed to single-step formation) shows that peptide with crossed disulfide bonds as Cys 1-3/2-4, and Cys 1-4/2-3 pairing patterns to be reported much more frequently than peptides with Cys 1-2/3-4 pattern (Van Rietschoten et al., 1977; Nishiuchi and Sakakibara, 1982; Gray et al., 1984; Kumagaye et al., 1988; Tam et al., 1990; Zhang and Snyder, 1991; Akaji et al., 1992). This suggests that an effort to force the EGF-like peptide into the desired Cys 1-3/2-4 pattern by blocking the 5-6 pair with Acm will probably fail because of the observed tendency to produce 1-2/3-4 disulfide

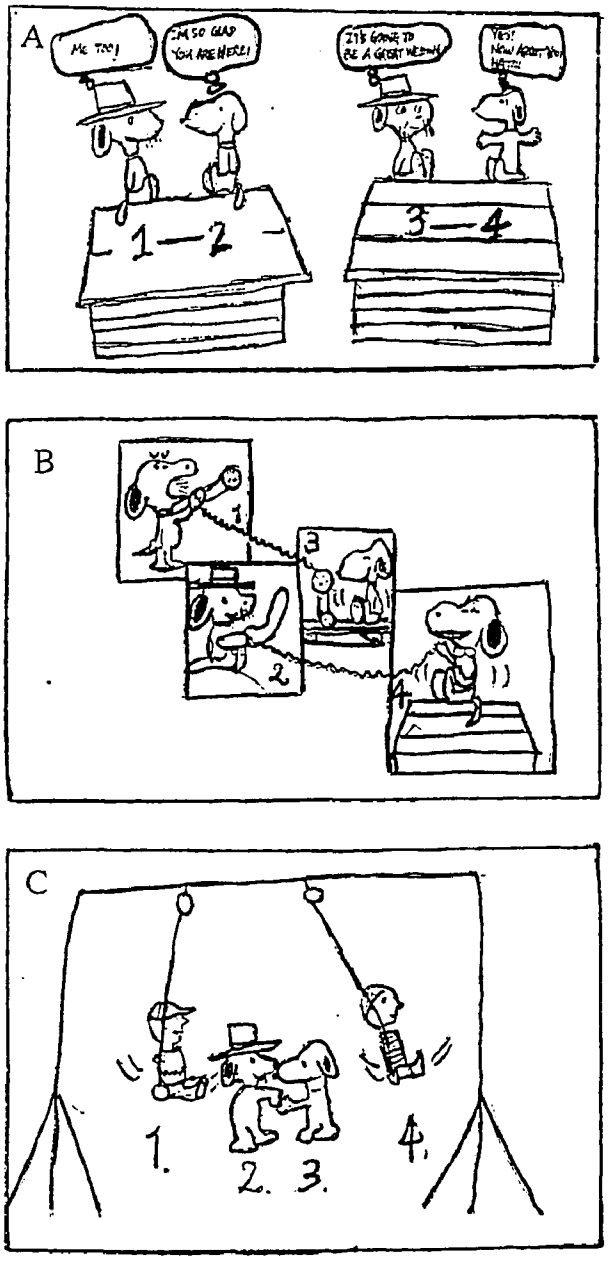


Figure 2.35. Cartoon showing the three disulfide patterns for peptides intramolecularly containing 4 cysteines. (Reproduced with the permission of the author).

A. pattern 1-2, 3-4
B. pattern 1-3, 2-4
C. pattern 1-4, 2-3

pairing. Indeed, for this case we obtained a high yield of the undesired Cys 1-2/3-4/5-6 isomer (Figures 2.11, 2.12, Table 2.4).

The remaining two choices for positioning the Acm blocking groups are the Cys 1-3 pair (peptide C) or Cys 2-4 pair (peptide B). Both peptides were made, but only the peptide B yielded the desired disulfide isomer. An explanation of the difference in behavior between the peptide B and C is found by consideration of the number of residues separating the cysteines. Because there is only one residue between the fourth and fifth cysteines (Cys^{26} and Cys^{28}) in peptide C, the likelihood of forming an undesired 2-5 cysteine bond might be expected to be almost as large as that for forming the desired 2-4 cysteine linkage. Thus, a peptide with Acm blocking Cys 1-3 (peptide C) would be expected to yield a mixture of peptides, with both 2-5 and 2-4 disulfide bonds present in the isomer mixture. This is precisely what we observed experimentally (Tables 2.4 and 2.7).

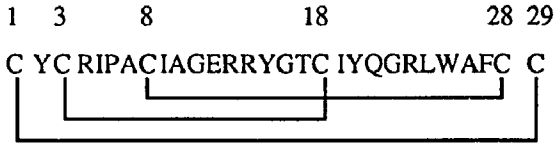
An additional consideration is that pairs of cysteines with only one or three intervening residues are very unlikely to form disulfide bonds (Zhang and Snyder, 1989). This is evidenced by the observation that, after the first folding step, peptide C (Acm on Cys 1 and 3) yielded a peptide, among others, with a 2-6 disulfide bond and two free cysteines (Cys 4 and Cys 5). These two cysteines are separated by only one residue, and the small ring that would be formed by a disulfide bond was so unfavorable as to prevent formation.

The considerations outlined above indicate that a rational choice for positioning Acm blocking groups can often be made for peptides containing three disulfides with a single crossing of disulfide bonds. It is better to place the Acm blocking groups on one of the pairs of cysteines involved in the crossing of disulfide bonds, preferably on the center loop (if this is a choice). Supplementary considerations regarding the number of intervening residues between cysteines must also be taken into account.

2.4.3. Generality of this method

It may be sensible to generally apply the two-step approach for synthesis of peptides with three disulfide bonds, even for peptides that are not yet known to exhibit difficulties in folding. The approach is feasible and it provides protection against potential unwanted disulfide isomers. If, at a later date, a peptide without Acm blocking groups is desired, it can be easily obtained by treatment of the blocked peptide with iodine.

There is no one method that is universal for all the possible situations encountered. For example, the two-step approach may not be successful for a defensin peptide with disulfide paring Cys^{1}-Cys^{29}/Cys^{3}-Cys^{18}/Cys^{8}-Cys^{28} (see sequence below). In this peptide Cys^{28} and Cys^{29} are so close that one of them should be blocked in order to minimize disulfide scrambling. After either the Cys^{1}-Cys^{29} pair or the Cys^{8}-Cys^{28} pair is blocked by Acm, the remaining two pairs of disulfide bonds appear cross-linking pattern of either Cys 1-3/2-4 or Cys 1-4/2-3 which is not the most common disulfide pattern in the native peptides (Benham and Jafri, 1993). Therefore, the application of the two-step method to this peptide is not a good choice. The sequence of this 29-residue peptide, human defensin 2, is as follows.



Recently we are pleased to read a preliminary report about a three-step method being developed (Durieux and Nyfeler, 1995). The overall yield of the method might be much lower than what achieved in the present two-step work, but it provides a new dimension for selective formation of multiple disulfide bridges in peptides.

Chapter 3

Characterization of a side reaction in 9-fluorenylmethoxycarbonyl chemistry: aspartimide formation followed by piperidine adduction

Abstract

Although 9-fluorenylmethoxycarbonyl (Fmoc) chemistry is one of the most widely used approaches for peptide synthesis, only a few side reactions have been characterized. This work demonstrates aspartimide formation from Asp-X, where X is Asn and Gly. This side reaction was thought not to occur when the aspartic acid is blocked with t-butyl (t-Bu), a commercially available and sterically hindered protecting group that would avoid base-catalyzed aspartimide formation. Contrary to the conventionally held view, however, during the synthesis of an EGF-like domain in a human blood coagulation factor, this side reaction has been found to occur in a significant extent. It indicates that in Fmoc chemistry, even when aspartic acid is protected with t-Bu, this side reaction is still a serious problem in the synthesis of long peptides having Asp-X sequences. It can form not only aspartimide between two amino acids but also a piperidine adduct having an additional mass of 67. An extensive study with pentapeptides has found that -Asp(Ot-Bu)-Asn(Trt)- and -Asp(Ot-Bu)-Gly- are problematic sequences in Fmoc chemistry.

3.1. Introduction

In solid phase peptide synthesis, it is important that the repetitive steps proceed rapidly, in high yield, and with minimal side reaction to prevent the accumulation of by-products (Merrifield, 1963). Fmoc chemistry and t-Boc chemistry are two major methodologies in solid phase peptide synthesis. The Fmoc approach offers an easy TFA cleavage technique, provides a wide choice of linkers and resins, and has become very popular in the last decade (Carpino and Han, 1972; Fields and Noble, 1990). However, Fmoc as an alpha-amino protecting group requires basic conditions for its removal, and so may induce base-catalyzed side reactions. Thus side reactions found in Fmoc chemistry may differ significantly from those found in Boc chemistry, which requires repetitive acidic

deprotecting conditions. Many side reactions in Fmoc chemistry will need to be investigated, while the potential side reactions in the acidic condition of Boc chemistry have been extensively studied. From a failed synthesis with Fmoc approach a side-reaction has been found and extensive studies have been followed. There is only one previously brief mention about this side reaction in Fmoc chemistry (Nicolás et al., 1989). This work demonstrates that aspartimide formation, well known in Boc chemistry (Barany and Merrifield, 1979a; Tam et al., 1988) but thought not to occur under the conditions of Fmoc, in fact occurs with both approaches.

3.2. Materials and methods

3.2.1. Peptide synthesis

An initial attempt to synthesize an epidermal growth factor (EGF)-like domain in human blood coagulation factor X (residues 83-130) was made at the RCMI Peptide Synthesis Facility of Hunter College using solid-phase methodology on an ABI 430A synthesizer. A standard single coupling with HOBT/HBTU and deprotection with 20% piperidine were used. All amino acids and resins were purchased from Biochem, Inc. (Torrance, California). The side-chain protecting groups for Fmoc amino acids were as follows: Asp and Glu by OtBu; Ser, Thr and Tyr by t-Bu; Asn, Gln and His by Trt; Lys by Boc; Arg by Pmc; and Cys by either Trt or Acm. A 4-(hydroxymethyl)-phenoxymethyl-Copoly styrene resin (HMP resin/Wang resin) with a substitution of 0.88 mmol/g was used. The cleavage mixture used for peptides containing both Arg and Trt protecting groups was: 0.75 g crystalline phenol, 0.25 ml 1,2-ethanedithiol (EDT), 0.5 ml thioanisole, 0.5 ml H_2O and 10 ml TFA (Applied Biosystems, 1990). The crude peptide, including the Acm blocking group, was precipitated in cold methyl t-butyl ether and collected by centrifugation.

Other related peptides derived from factor X (residues 106-130, residues 98-130) were later synthesized manually by the stepwise solid-phase method with Fmoc-Tyr(t-Bu)-

HMP resin (Wang resin) at 0.42 mmol/g substitution. Stepwise coupling of Fmoc amino acids using DCC/HOBT was performed first, followed by TBTU as a second coupling when necessary. Deblocking the NH_2-terminal Fmoc protecting group with 20% piperidine in DMF (v/v) was typically carried out for 20 min. Since Pro was the second residue from the C-terminal, 50% piperidine in DMF (v/v) was applied for 5 min to the dipeptide resin to minimize the formation of diketopiperazine.

Manual peptide synthesis was also used to make model pentapeptides. Fmoc-Ala-HMP resin (Wang resin) with a substitution of 0.62 mmol/g was used for the synthesis of model peptides. For model peptide studies, the initial weight of the HMP resin was 0.162 g corresponding to 0.1 mmole, which was enough for three microscale TFA cleavages and the piperidine treatment test.

Analytical C18 reverse-phase HPLC was performed. (1) For EGF-like peptides: Buffer A contained 5% acetonitrile in 0.045% TFA. Buffer B contained 60% acetonitrile in 0.037% TFA. Peptides were eluted with a 10-40% buffer B linear gradient over 30 min at 1.5 ml/min, monitored at 220 nm. (2) For model peptides: Buffer A contained H_2O in 0.045% TFA. Buffer B contained 60% acetonitrile in 0.037% TFA. The model peptides were eluted with 100% buffer A for 3 minutes and 0-22% buffer B linear gradient over 22 min at 1.5 ml/min, monitored at 220 nm. Semipreparative C18 reverse phase HPLC was used for purification.

3.2.2. Micro-scale TFA Cleavage

A sample of approximately 10-15 mg of resin was treated with a small portion of cleavage mixture (1 to 1.5 ml) in a sealed 20 ml scintillation vial. The cleavage mixture used for peptides not containing Arg and Trt protecting groups was: 0.5 ml H_2O and 9.5 ml TFA; for peptides containing Trt protecting groups was: 0.25 ml EDT, 0.25 ml H_2O and 9.5 ml TFA; for peptides containing both Arg and Trt protecting groups was: 0.75 g crystalline phenol, 0.25 ml EDT, 0.5 ml thioanisole, 0.5 ml H_2O and 10 ml TFA (Applied

Biosystems, 1990). The cleavage reaction was performed at room temperature with stirring for 1.5 hr. Crude peptides were filtered in a Buchner funnel with a fritted disc to remove the solid support, precipitated in about 6-8 ml of cold ethyl ether in a centrifuge tube (10 ml size), and then collected by centrifugation. The crude peptide was washed repeatedly by centrifugation with cold ethyl ether (at least three times) and then dissolved in water for lyophilization. A number of samples can be processed simultaneously in this manner.

3.2.3. Mass spectrometry (MS)

Mass spectra of individual peptide samples and model peptides were analyzed in MS group of The Rockefeller University with an electrospray mass spectrometer constructed at The Rockefeller University and described elsewhere (Chowdhury, et al., 1990). Briefly, the charged droplets produced by electrospray at atmospheric pressure are focused and transported through a 20 cm long stainless steel capillary tube (0.5 mm i.d.) into a region maintained at 1-10 torr. Desolvation of the ions to be analyzed is achieved in part by controlled heating of the capillary and in part by collisional activation brought about by an electrostatic field in the intermediate pressure region (1-10 torr) between the capillary exit and a coaxial skimmer. All the solutions used in the present study were electrosprayed at rates of 0.5-1.0 μL/min using a stainless steel syringe needle. The peptide samples were dissolved in a mixture of water, methanol and acetic acid (20:19:1) to a concentration of 10 μM and sprayed at a voltage of 3.5-4.5 KV.

Peptide ladder samples were analyzed on a matrix-assisted laser desorption time-of-flight mass spectrometer constructed at The Rockefeller University and described elsewhere (Beave and Chait, 1989, 1990). The individual peptides (9 residues to 18 residues in one vial and 19 to 32 residues in another vial) were mixed in approximately equal amounts and dissolved in water. The ladder mixtures were added to the matrix material (4-hydroxy-α-cyano-cinnamic acid (4HCCA) in formic acid/water/isopropanol 1:3:2) to a final concentration of 1-5 μM for each peptide component. The complete peptide ladder, which

ranged from the 9 mer to the 32 mer (except 16 mer and 28 mer) was measured from 4HCCA in water/acetonitrile 2:1 and collected from 200 laser shots. The final concentration of each peptide component was in the range of 0.2-1 μM. A small aliquot (0.5 μl) of this mixture was applied to the metal probe tip and dried at room temperature with forced air. The sample was then inserted into the mass spectrometer and analyzed. Bovine insulin and substance P were used as internal calibrants.

3.2.4. NMR

Proton NMR spectra were recorded using a GE/Bruker QE 300 MHz FT spectrometer at 25°C. Two peptide samples were prepared from 17 mer corresponding to the desired peptide and the unknown peptide which has a 67 u larger mass. They were both at pH 2.8 and in D_2O. Acetylpiperidine sample was also in D_2O with unadjusted pH (pH 4.5). Uncorrected pH meter readings were used to obtain pD values.

3.3. Results

3.3.1. Initial Synthetic Attempts

The initial synthesis of the EGF-like peptide carried out on a synthesizer using an Fmoc approach failed. Following is the primary structure of this peptide (Appendix A: Leytus et al., 1986), the C-terminal EGF-like domain from human blood coagulation factor X (fX_{EGF-C}):

1(83) 48(130)

LFTRKLCSLDNGDCDQFCHEEQNSVVCSCARGYTLADNGKACIPTGPYA

The numbering system used is based on the synthetic peptide, with the corresponding factor X sequence position shown in parentheses.

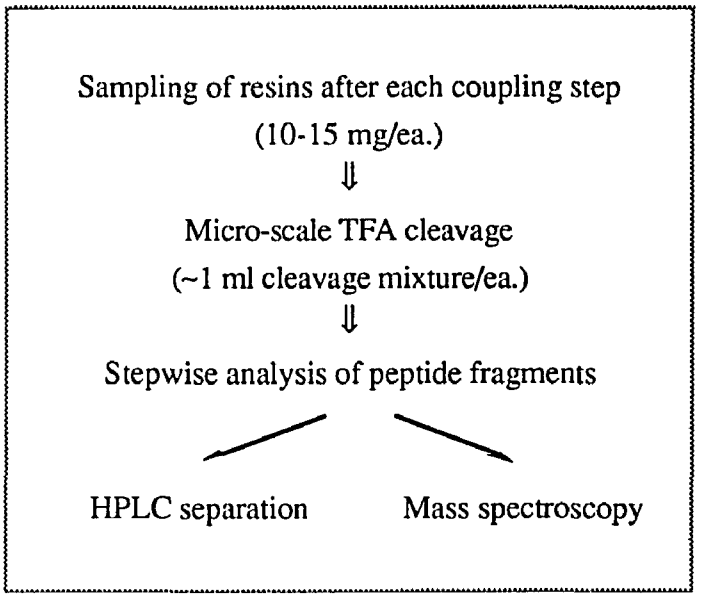
As identified by both HPLC and MS, the resulting peptide with 48 residues was found to be a mixture containing no detectable amount of the desired peptide. Very similar results were obtained when the same experiment was repeated using the synthesizer. Manual synthesis was then attempted. Since the diketopiperazine formation at the onset of

the synthesis could have been responsible for the failure of the synthesis, an attempt was made to minimize formation of diketopiperazine. The piperidine treatment time was shortened to about 7 minutes at the dipeptide resin stage using 50% piperidine in DMF. The substitution of the resulting resin containing the first three residues (GPY) was examined by amino acid analysis, and compared with that of the purchased resin containing Tyr (data not shown). No significant variation in the substitution of the tripeptide-resin was found, suggesting the formation of diketopiperazine was not a major cause for the failed synthesis. As the synthesis progressed, ninhydrin tests on the deblocked peptide resins indicated that there was a problem in the synthesis. Examination of the deprotected crude peptides produced after coupling of the 17th residue failed to yield a single dominant peak in the HPLC profile, indicating the presence of a mixture of products. In addition, an electrospray MS analysis of the crude peptide product did not show the presence of the desired peptide.

3.3.2. Stepwise detection side reaction

To determine where the problematic sequence was, a stepwise detection method was then employed while performing manual synthesis. It is an efficient and direct method to monitor the progress of side reactions in Fmoc chemistry. The analysis involves stepwise micro-scale TFA cleavage in conjunction with HPLC and MS for identification of products (Scheme 3.1).

The resins were sampled after each coupling step. In Fmoc chemistry micro-scale TFA cleavage can be conveniently carried out on a small amount of sample. The peptide fragments generated by TFA cleavages were then examined by HPLC and MS. By comparing each peptide fragment, the side reaction was detected and eventually defined. Peptide ladder mass spectrometric analysis was used for a mixture of the collected peptide fragments to provide further corroborative evidence for the aspartimide formation.



Scheme 3.1. Strategy for detecting side reaction during synthesis.

3.3.3. Peptide synthesis monitored with HPLC and MS

A portion of the desired EGF-like peptide (Figure 3.1) was then synthesized manually. After coupling of the first 7 amino acid residues, a sample of peptide resin was removed. Thereafter a sample of peptide resin was removed subsequent to each coupling step (except after the 16th and 28th cycles). The peptide resin samples were cleaved to obtain crude peptides for analysis by both HPLC and MS.

A plot collecting the HPLC results from the peptide samples containing 10 mer to 24 mer is shown in Figure 3.2. These HPLC profiles indicate that the synthesis proceeded well up to the 11th residue. However, after the 13th residue from the C-terminal a new peak appeared in the HPLC profile (shown shaded in Figure 3.2). This unknown peak eluted approximately 5 minutes later than the desired peptide peak. The new peak grew with each additional coupling, so that after the 23rd residue had been coupled it exhibited almost the same area as the peak from the desired peptide. Figure 3.3 shows the ratio of the area of the unknown peak and the expected peptide peak calculated from HPLC compared

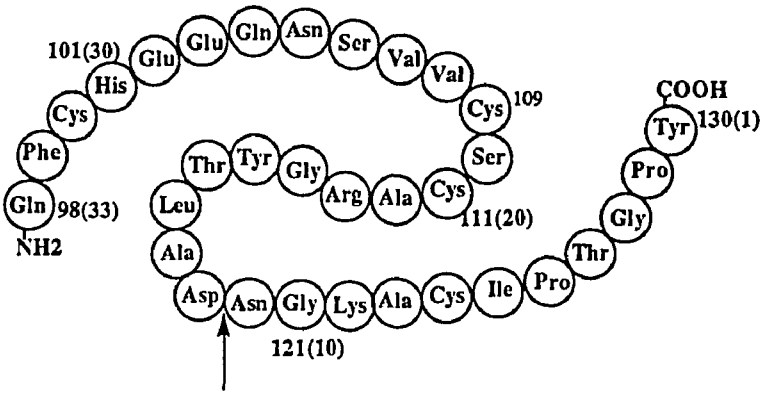


Figure 3.1. Primary structure of a manually synthesized epidermal growth factor-like peptide derived from human blood coagulation factor X (Leytus et al., 1986). The numbering system used is based on the sequence position in factor X, with numbers relative to the C-terminus of the synthetic peptide shown in the parentheses. The arrow indicates the location of aspartimide formation and subsequent ring opening by piperidine to form an adduct. All Cys residues are protected with Trt except Cys^{109} blocked with Acm.

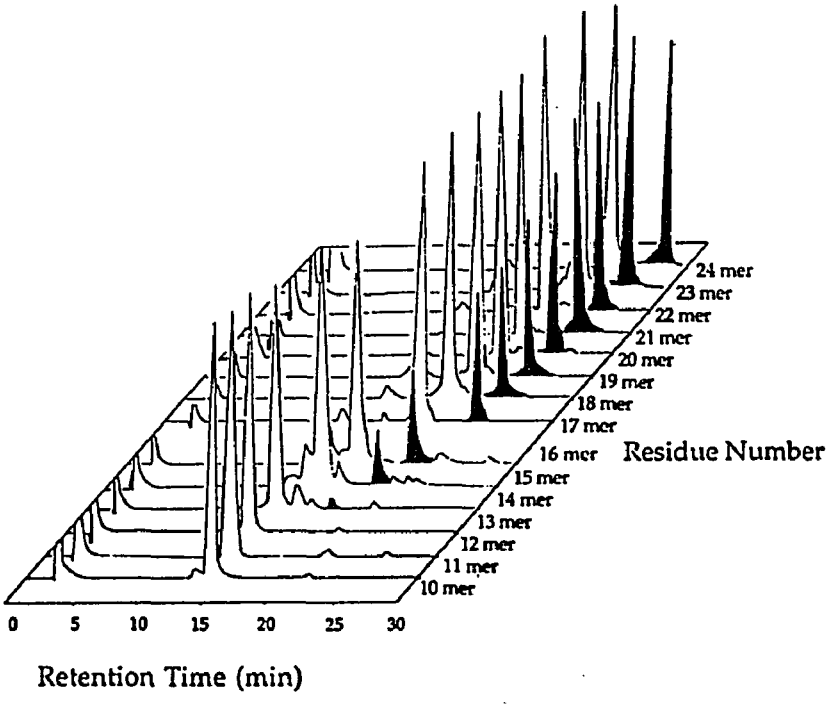


Figure 3.2. Reverse phase HPLC profiles showing the stepwise analysis of peptide samples containing the C-terminal 10 mer to 24 mer (except the 16 mer). The shaded peaks have masses 67 u larger than that of the corresponding normal peptide peaks which are unshaded.

with the ratio of peak heights obtained from MS analysis. Mass spectrometric analysis on individual peptide samples showed that the unknown peptide had a mass 67u larger than that of the expected peptide (see Table 3.1). The first Δ = + 67 u peak observed in the mass spectrum was that for 13 mer peptide (Figure 3.4). Another small peak (Δ = - 18u) was also present in the same spectrum.

3.3.4. Defining the side reaction

There is no such an amino acid, a coupling reagent, or a solvent having molecular weight 67 u. However, 67 u plus a molecular weight of H_2O yields 85 u corresponding to piperidine. Piperidine is frequently used for deprotection of Fmoc group. Before each coupling cycle resin is treated with 20% piperidine in DMF for 20 min. This difference in mass, +67 u, suggests that the unknown peak was possibly caused by piperidine adduction.

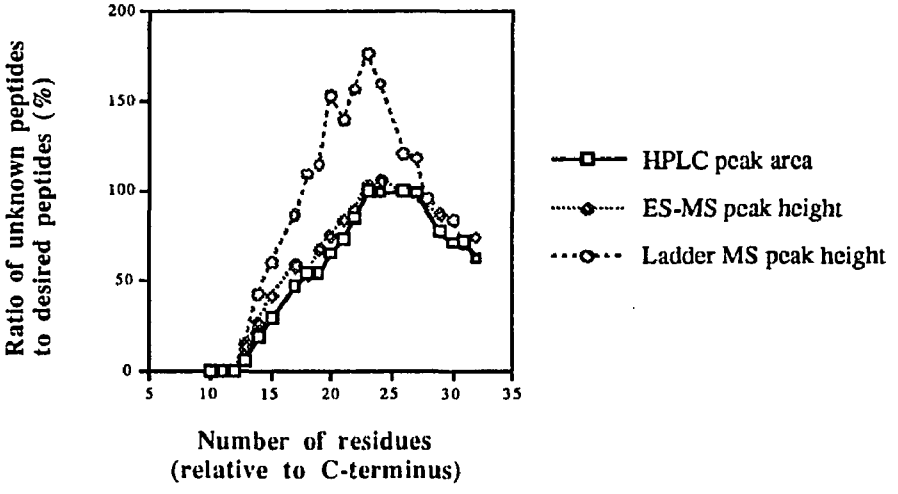


Figure 3.3. Plots of the ratio of unknown peak to the normal peptide peak in three different ways: HPLC profiles; electrospray mass spectra (ES-MS); peptide ladder by matrix-assisted laser desorption/ionization mass spectrometry (Ladder MS).

Table 3.1. Electrospray mass spectrometry results of individual peptides.

Peptide samples	Mass of expected peptides		Mass found in unknown species	ΔMass between expected and unknown peptides
	Theory	Found		
11 mer	1120.3	1120.3	-	-
12 mer	1235.4	1235.3	-	-
13 mer	1306.5	1306.3	1373.3	67.0
14 mer	1419.6	1419.3	1486.5	67.2
15 mer	1520.7	1520.5	1587.2	66.7
17 mer	1741.0	1740.5	1808.3	67.8
18 mer	1897.1	1896.7	1963.7	66.7
19 mer	1968.2	1968.4	2035.9	67.9
20 mer	2071.4	2069.3	2136.0	66.7
21 mer	2158.4	2157.8	2224.5	66.7
22 mer	2332.5	2331.4	2398.5	67.1
23 mer	2431.7	2430.7	2497.8	67.1
24 mer	2530.8	2529.3	2597.6	68.3
26 mer	2732.0	2731.5	2799.8	68.3
27 mer	2860.1	2857.1	2925.0	67.9
29 mer	3118.3	3116.6	3183.8	67.2

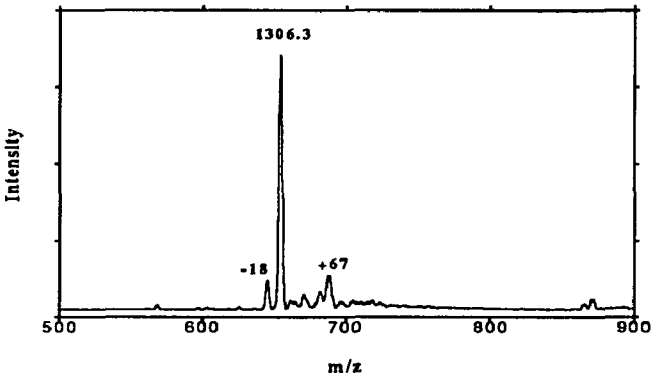


Figure 3.4. Electrospray mass spectrum for crude 13 mer peptide. The major peak represents the desired peptide. The peak with Δ= -18 indicates the loss of water, consistent with aspartimide formation. The peak of Δ= +67 in this 13 mer peptide is the first appearance of piperidine adduct.

Although the 13th residue is Ala where the piperidine adduct initially appeared, the first indication of the problem is originally from the 12 mer peptide. For the crude peptide containing 12 residues, a small peak in the electrospray mass spectrum (Figure 3.5) shows a mass 18 u lower than the desired peptide. The sequence, in which 12th and 11th amino acids are Asp and Asn, is a susceptible portion in Boc chemistry. It is known to form aspartimide between them. Therefore, the mass difference (+67 u) is due to the loss of water, formation of aspartimide (-18 u), and subsequent ring opening by piperidine adduct (+85 u). The problem sequence in this EGF-like peptide is -Ala-Asp-Asn-, the 13th, 12th and 11th residues from the C-terminus.

3.3.5. MS analysis of the synthetic peptide ladder

Matrix-assisted laser desorption mass spectrometric analysis of synthetic peptide ladders, first introduced for the purpose of sequencing in solid phase synthesis (Walker et al., 1993; Chait et al., 1993), was used here for tracking side-reactions. Sample was prepared by pooling the individual crude peptides. A ladder spectrum gives a single readout data set that can be interpreted in a straightforward manner. Following spectrum (Figure 3.6) is not perfect, because not enough amount of 14 mer is added. However, from this

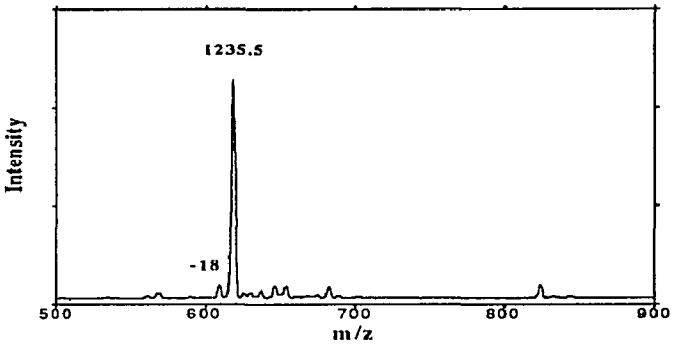


Figure 3.5. Electrospray mass spectrum for crude 12 mer peptide. The major peak represents the desired peptide. A small peak of Δ= -18 is the first indication of aspartimide formation.

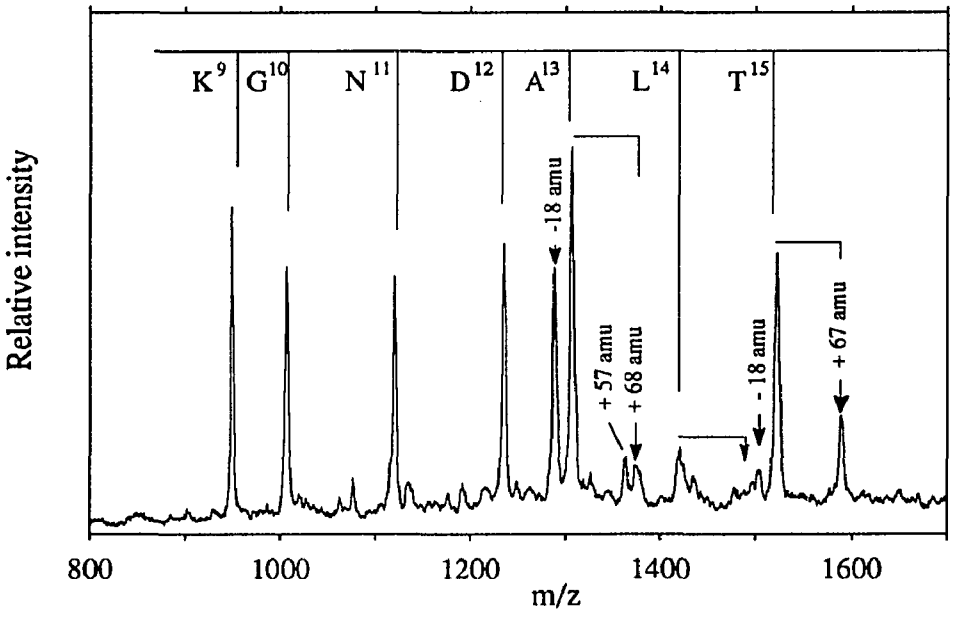


Figure 3.6. Partial view of the matrix-assisted laser desorption mass spectrum of the synthetic peptide ladder from 9 to 32 residues. The formation of aspartimide (loss of water, -18 u) and the piperidine adduct (+67 u) are observed after the synthesis of the 13th residue. The weak intensity of the peak corresponding to the 14 mer is due to the low amount of 14 mer added.

peptide ladder mass spectrum, it is clear that the Δ=+ 67 u peak first appeared at the 13th residue from the C-terminal bearing the sequence -Ala[13]-Asp[12]-Asn[11]-. The intensive peak, Δ= -18 u, is also observed in 13 mer, although it appeared earlier in 12 mer in electrospray mass spectrum (Figure 3.5). Table 3.2 shows the results obtained from this ladder spectra for the normal peptide and the unknown species.

3.3.6. NMR studies

Proton NMR studies (Figure 3.7) were consistent with the presence of a piperidine adduct in the peptide. Two NMR samples were prepared from 17 mer peptides corresponding to the desired peptide and the unknown peptide with a 67 u larger mass. The proton NMR spectrum of the unknown peptide showed increased peak intensity at 1.6 and

Table 3.2. A tabular presentation of the peptide ladder mass spectrometry. The data was obtained from two mixtures. The first mixture contained the individual peptides from the 9 mer to the 18 mer and the second mixture from the 19 mer to the 32 mer.

Peptide samples	Mass of expected peptide		Mass found in unknown species	ΔMass between expected and unknown peptides
	Theory	Found		
9 mer	949.1	948.9	-	-
10 mer	1006.2	1006.1	-	-
11 mer	1120.3	1120.0	-	-
12 mer	1235.4	1234.9	-	-
13 mer	1306.5	1306.3	1374.3	68.0
14 mer	1419.6	1419.5	1486.8	67.3
15 mer	1520.7	1520.4	1587.5	67.1
17 mer	1741.0	1741.4	1808.0	66.6
18 mer	1897.1	1897.3	1964.9	66.7
19 mer	1968.2	1968.2	2035.9	67.7
20 mer	2071.4	2071.3	2138.9	67.6
21 mer	2158.4	2158.6	2226.0	67.4
22 mer	2332.5	2332.6	2399.4	66.8
23 mer	2431.7	2431.2	2499.4	68.2
24 mer	2530.8	2530.4	2598.3	67.9
26 mer	2732.0	2729.9	2799.4	69.5
27 mer	2860.1	2859.1	2927.5	68.4
29 mer	3118.3	3117.6	3185.2	67.6
30 mer	3255.5	3254.3	3323.8	69.5
31 mer	3358.6	3358.1	3425.1	67.0
32 mer	3505.8	3505.2	3572.0	66.8

3.6 ppm relative to the spectrum of the desired peptide. Panel C in Figure 3.7 is an NMR spectrum of acetylpiperidine with a structure similar to that of the piperidine adduct. The middle peak in the spectrum is assigned to a methyl group, which is not present in the peptide after piperidine addition, and therefore is not visible in panel B. The other two peaks in the spectrum of acetylpiperidine display similar chemical shift (1.6 and 3.5 ppm) to the two peaks identified through comparison of the spectra from the known and unknown peptides. Thus, the NMR results were consistent with the formation of a piperidine adduct.

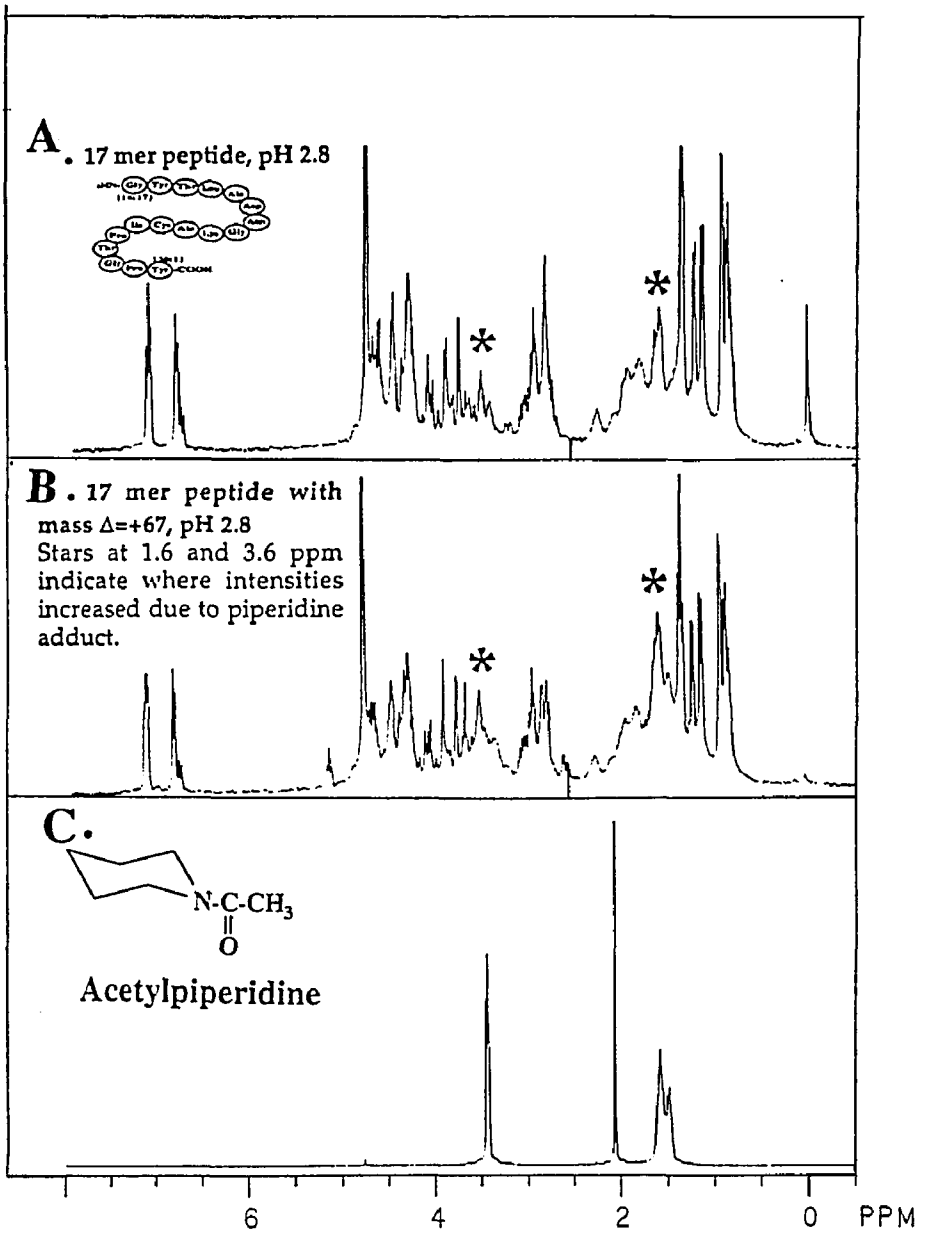


Figure 3.7. 300 MHz proton FT-NMR spectra at 25°C in D_2O.
A. 17 mer peptide at pH 2.8; B. 17 mer peptide containing piperidine at pH 2.8; C. Acetylpiperidine, pH 4.5.

3.3.7. Two possible pathways for aspartimide formation in Fmoc chemistry

Two possible pathways for piperidine addition are proposed in Figure 3.8. The side-chain of aspartic acid is protected with t-butyl. In pathway *a*, piperidine acts as both a base and a nucleophilic reagent. Under basic conditions, aspartimide is formed between Asp and Asn. Then the piperidine adduct is created from the nucleophilic attack at the imide ring.

Since t-butyl is a good leaving group, an alternative pathway can not be excluded. The formation of the piperidine adduct could be derived by a direct substitution reaction, as shown in pathway *b* of Figure 3.8.

There are two possible products, α and β aspartyl peptides, which rearrange upon imide opening. The α aspartyl peptide retains its normal peptide bond, while the β peptide is an isomer in which the peptide backbone is reconstructed through the Asp side-chain β-carboxyl group. A determination of whether an α or β peptide is present as major product was not completed during this work. It is noticed that an enzyme reported as an analytical probe to identify unknown aspartyl peptides (George-Nascimento et al., 1990), which might be able to answer this question.

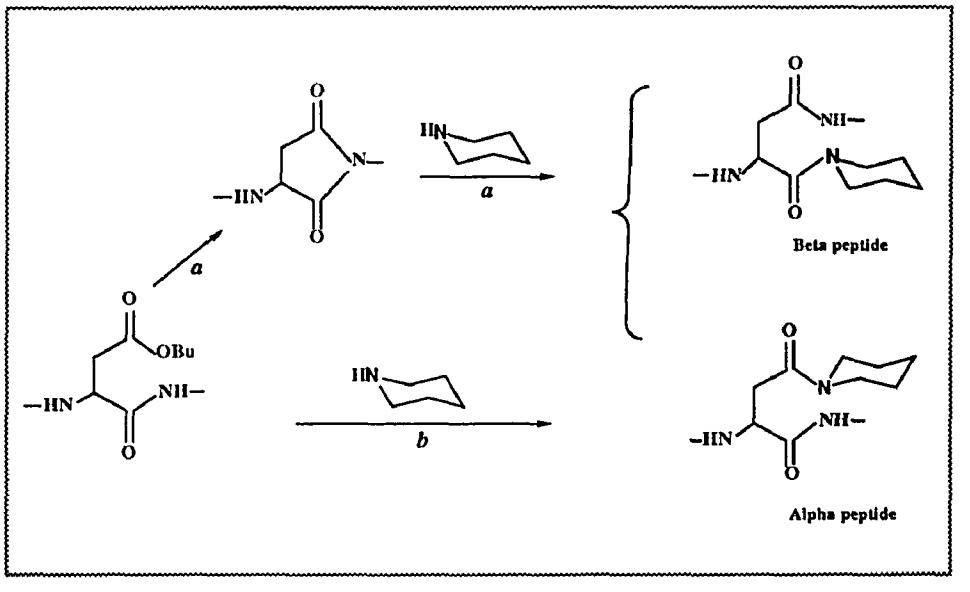


Figure 3.8. Two possible pathways for piperidine addition to the aspartimide formed between -Asp(OtBu)-Asn(Trt)- or -Asp(OtBu)-Gly-.

3.3.8. Model peptide studies

To investigate which amino acids in the sequence -Asp(Ot-Bu)-X- are susceptible to the aspartimide side reaction, seven model pentapeptides were synthesized: Lys(Boc)-Ala-Asp(Ot-Bu)-X-Ala, where X was Ala, Asn(Trt), Gly, Gln (Trt), His (Trt), Ser(t-Bu), Thr (t-Bu). The -Ala-Asp(Ot-Bu)-X- portion of pentapeptides was designed to mimic the sequence -Ala-Asp-Asn-, the 13th, 12th, and 11th amino acids from the C-terminal in the EGF-like peptide mentioned previously. The C-terminal residue in the model peptides was Ala, a simple amino acid to attach to the resin. Lys was added in the sequence because of its easy protonation which facilitates the use of positive electrospray mass spectrometer. Most of the residues examined (those which are represented by X) have been reported in Boc chemistry to be susceptible to aspartimide formation.

The formation of aspartimide and its piperidine adduct in these model peptides was monitored using HPLC and MS analysis of the crude peptides. Aspartimide formation was also found to increase with piperidine concentration (see section 3.3.10: Figure 3.15). To simulate a lengthy peptide synthesis but under milder deprotection conditions, the pentapeptide resins were treated for 4 hours with 10% piperidine in DMF (v/v) rather than 20% piperidine.

Table 3.3. Piperidine adduction in Fmoc chemistry studied with model pentapeptides: Lys(Boc)-Ala-Asp(Ot-Bu)-X-Ala

X	Presence of mass Δ=+67 found after synthesis	Presence of mass Δ=+67 found after treating target resin with 10% piperidine for 4 hours
Ala	-	-
Asn(Trt)	-	+
Gly	+	+
Gln(Trt)	-	-
His(Trt)	-	-
Ser(t-Bu)	-	-
Thr(t-Bu)	-	-

The amount of aspartimide and piperidine adduct formed after four hours incubation with 10% piperidine, as determined by HPLC, was between 15 to 25 % for -Asp(Ot-Bu)-Asn(Trt)- and -Asp(Ot-Bu)-Gly-. This extent is less pronounced in the model peptide studies than in the synthesis of the EGF-like peptide (fXEGF-C) containing -Asp-Asn-. A smaller amount of aspartimide formation, without piperidine adduct formation, was observed for the -Asp(Ot-Bu)-Ala- and Asp(Ot-Bu)-Gln(Trt)- sequences. Table 3.3 shows the results for detection of piperidine adduct in model peptide studies in Fmoc chemistry.

Since serious side reaction has been found in two model peptides {Lys(Boc)-Ala-Asp(Ot-Bu)-Asn(Trt)-Ala and Lys(Boc)-Ala-Asp(Ot-Bu)-Gly-Ala}, a detailed study was conducted. Figure 3.9 and Table 3.4 show the HPLC and electrospray MS analysis results for Lys(Boc)-Ala-Asp(Ot-Bu)-Asn(Trt)-Ala. Figure 3.10 and Table 3.5 show the HPLC and electrospray MS analysis results for Lys(Boc)-Ala-Asp(Ot-Bu)-Gly-Ala. The data demonstrates significant aspartimide formation in these two peptides.

Table 3.4. HPLC (Figure 3.9) and MS analysis of model peptide: Lys(Boc)-Ala-Asp(Ot-Bu)-Asn(Trt)-Ala

Peaks separated by HPLC	Retention time (min)	HPLC area (%)	Mass difference Δ*	Identification
Major	2.9 and 3.2	82.8	0	Desired peptide
(a)	5.6	2.9	-18	Aspartimide**
(b)	14.2	9.1	-18	Aspartimide**
(c)	18.8	5.1	+67	Piperidine adduct

* Compared with mass of desired peptide, 517.5 u.
** These side reaction products have same mass but elute with different retention times. The reason for this is not known.

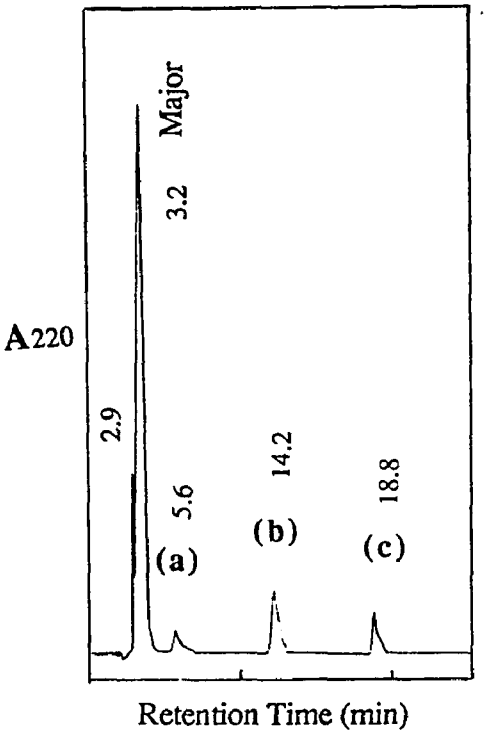


Figure 3.9. Reversed-phase HPLC profile for model peptide Lys(Boc)-Ala-Asp(Ot-Bu)-Asn(Trt)-Ala. Resin has been treated with 10% piperidine in DMF for 4 hrs prior to TFA cleavage. Peptides were first eluted with 100% of buffer A for 3 min, then eluted with a 0-22% buffer B linear gradient over 22 min at 1.5 ml/min, monitored at 220 nm. See Materials and Methods for buffer A and B compositions.

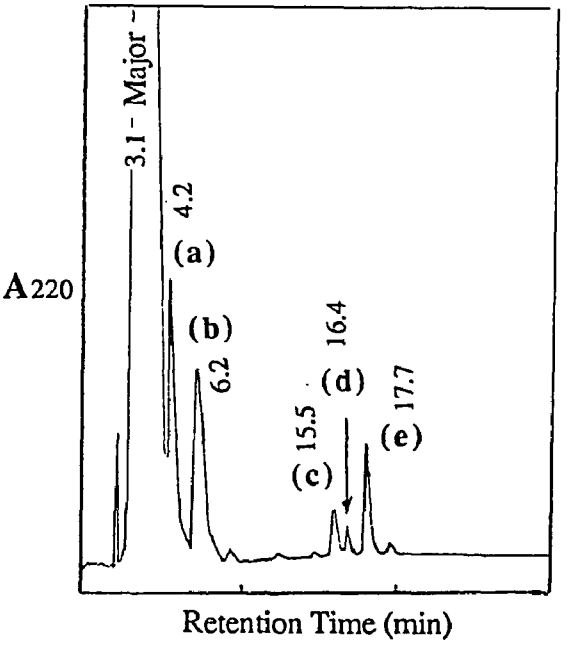


Figure 3.10. Reversed-phase HPLC profile for model peptide Lys(Boc)-Ala-Asp(Ot-Bu)-Gly-Ala. Resin has been treated with 10% piperidine in DMF for 4 hrs prior to TFA cleavage. Peptides were first eluted with 100% of buffer A for 3 min, then eluted with a 0-22% buffer B linear gradient over 22 min at 1.5 ml/min, monitored at 220 nm. See Materials and Methods for buffer compositions.

Table 3.5. HPLC (Figure 3.10) and MS analysis of model peptide: Lys(Boc)-Ala-Asp(Ot-Bu)-Gly-Ala

Peaks separated by HPLC	Retention time (min)	HPLC area %	Mass difference Δ*	Identification
Major	3.1	72.9	0	Desired peptide
(a)	4.2	11.2	-18;** +71*** (small peak)	Aspartimide; An extra Ala coupled;
(b)	6.2	9.9	-18**	Aspartimide
(c)	15.5	1.5	142 (=+67+57-18)	Piperidine adduct plus Gly minus a H_2O
(d)	16.4	0.8	+67****	Piperidine adduct
(e)	17.7	3.7	+67;**** +124 (small peak) (=+67+57)***	Piperidine adduct; Piperidine adduct plus Gly

* Compared with mass of the desired peptide, 460.5.
** These side reaction products have same mass but elute with different retention times. The reason for this is not known.
*** Detectable amount only, i.e. in fraction (a) about one fourth the MS peak height of the other species, aspartimide; and in fraction (e) about one third the MS peak height of the other species, piperidine adduct.
****These side reaction products have same mass but elute with different retention times. The reason for this is not known.

3.3.9. Comparison of aspartimide formation in Boc and Fmoc chemistry

The formation of aspartimide is a side reaction often observed in the synthesis of aspartic acid-containing peptides either in solution or in solid phase with t-Boc chemistry (Barany and Merrifield, 1979; Tam et al., 1988). It is well known to be sequence dependent and occurs under either strongly acidic or basic conditions. Aspartimide formation is known to occur readily with Asp-Gly or Asp-Ser in strong acids such as HF or CH_3SO_3H. Similarly these sequences are also susceptible to base-catalyzed aspartimide formation during solution synthesis using benzyl protecting groups (Bodanszky et al.,

1978a, 1978b). A comparison of aspartimide formation in base-driven Fmoc and in acid-driven Boc chemistry is shown in Table 3.6. In Boc chemistry hydroxyamino acids (Ser and Thr) and His have serious aspartimide formation problems, especially for Ser. However in Fmoc chemistry, when Ser is protected by t-butyl, there is no significant side reaction. In contrast to this, Gln is not susceptible in Boc chemistry, but may be in Fmoc chemistry. A small amount of aspartimide was detected when X=Gln, yet no piperidine adduct was found. A small amount of aspartimide formation was also found for -Asp-Ala-containing peptide in both Fmoc and Boc chemistry, but no piperidine adduct was detected. Both -Asp-Asn- and -Asp-Gly- are susceptible to aspartimide formation in either Boc or Fmoc approach with piperidine adduct formed. Therefore, the nature of the amino acid following aspartic acid in the sequence influences not only the rate and amount of aspartimide formation, but also the formation of piperidine adducts.

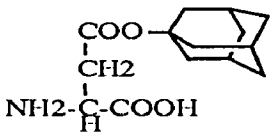
Table 3.6. Occurrences of Aspartimide Formation in Fmoc- and Boc- Chemistry.

Susceptible residue	Aspartimide formation[1]		
	Fmoc chemistry[2] -Asp(OtBu)-	Boc chemistry[3] -Asp(OBzl)-	
Ala	+[4]	+	(Yang and Merrifield, 1976; Marshall and Merrifield, 1965)
Asn	+++[5]	++	(Tam et al., 1988)
Gly	+++[5]	+++	(Blake, 1979; Yang and Merrifield, 1976; Wang et al., 1974; Nicolás et al., 1989; Ondetti et al., 1968)
Gln	+[4]	-	-
His	-	++	(Baba et al., 1973)
Ser	-	+++	(Yang and Merrifield, 1976; Suzuki and Endo, 1978)
Thr	-	++	(Suzuki and Endo, 1978)

1. Arbitrary scale. 2. This work. 3. Data from references in parentheses.
4. No piperidine adduct was observed. 5. Piperidine adduct was observed.

3.3.10. Influence of protection groups

A side-chain protection group for Asp, β-1-adamantyl aspartate [H-Asp(O-1-Ada)-OH], has been reported to be unaffected by treatment of 55% piperidine and can suppress aspartimide formation under acidic and basic conditions (Okada et al., 1987, 1988).



Asp-(O-1-Ada)

Okada and coworkers hoped that the adamantyl group would be rigid and bulky enough to prevent aspartimide formation. However, the peptide they studied contains -Asp-Ser- sequence which has no significant side reaction in Fmoc chemistry (see Table 3.6). When this protecting group is used in following model peptides: Lys(Boc)-Ala-Asp(O-1-Ada)-Asn(Trt)-Ala and Lys(Boc)-Ala-Asp(O-1-Ada)-Gly-Ala, no evidence of reduced aspartimide formation was found.

TFA cleavages were carried out for model peptide-resins containing the Ada protecting group before and after the treatment with 20% piperidine for 4 hours. Figures 3.11 and 3.12 represent the HPLC and MS figures for peptide Lys(Boc)-Ala-Asp(O-1-Ada)-Asn(Trt)-Ala. Figures 3.13 and 3.14 are for peptide Lys(Boc)-Ala-Asp(O-1-Ada)-Gly-Ala. HPLC results are summarized in Tables 3.7 and 3.8, respectively for two model peptides. The identification for HPLC peaks in these tables are made according to previous model peptides showed in Table 3.4 and Table 3.5. Both piperidine adduction and aspartimide formation were detected in trace amount for either peptide even before the long hour piperidine treatment. Of the piperidine treatment the total formation of aspartimide and piperidine adduct was 48% and 33% in the two peptides, KADNA and KADGA. Therefore, O-1-Ada is not a desired side-chain protecting group for Asp to provide protection against aspartimide formation in Fmoc chemistry.

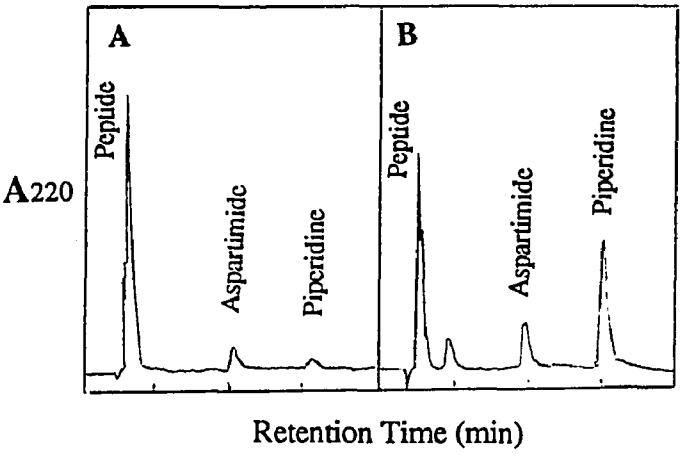


Figure 3.11. Reversed-phase HPLC profile for model peptide Lys(Boc)-Ala-Asp(O-1-Ada)-Asn(Trt)-Ala before and after tretment with piperidine. See Materials and Methods for gradient and buffer compositions.
A. Peptide without post-synthesis piperidine tretment.
B. Peptide-resin treated with 20% piperidine for 4 hrs.

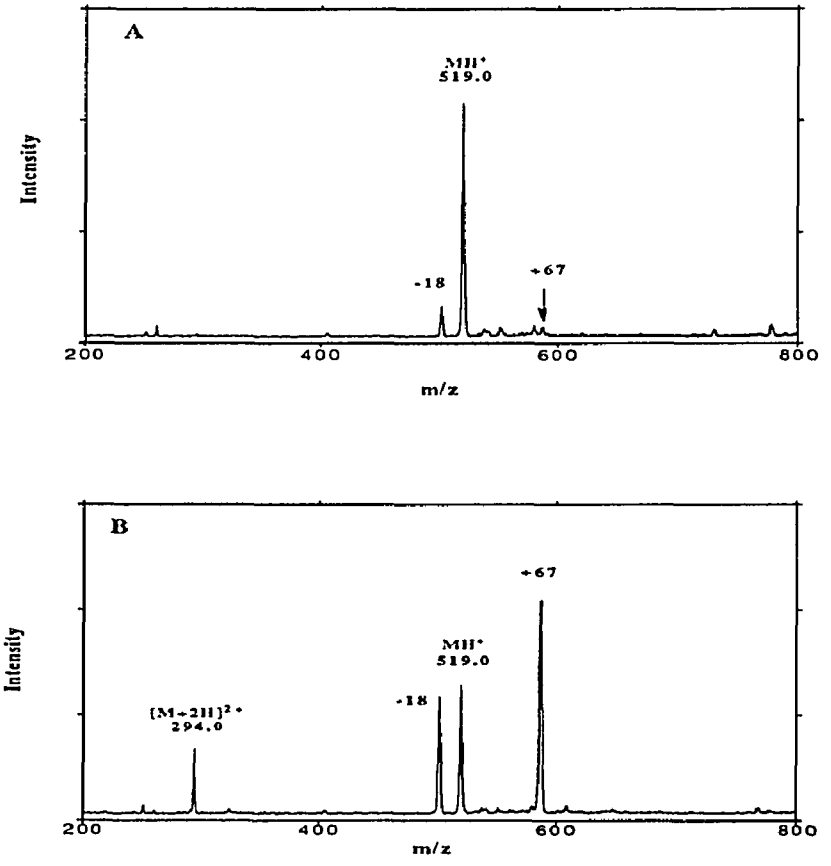


Figure 3.12. Portion of the electrospray mass spectra of peptide Lys(Boc)-Ala--Asp(O-1-Ada)-Asn(Trt)-Ala. The theoretical mass of the peptide is 517.5 u.
A. Peptide without post-synthesis piperidine treatment.
B. Peptide resin treated with 20% piperidine for 4 hrs.

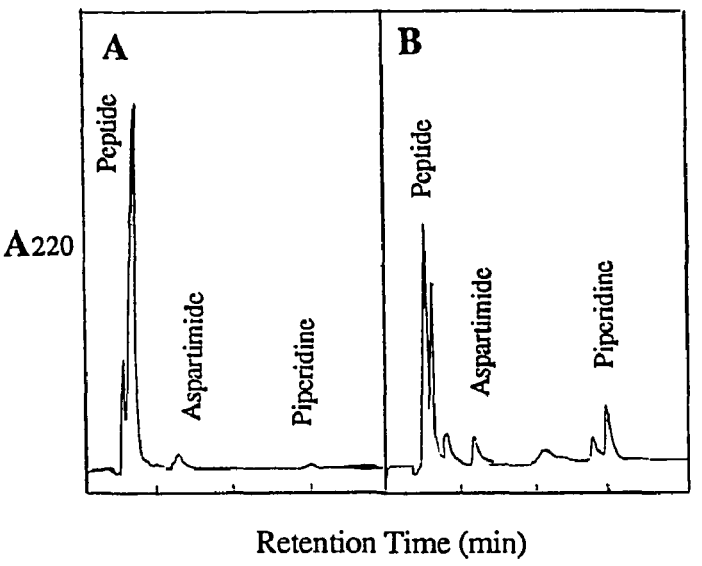


Figure 3.13. Reversed-phase HPLC profiles of peptide Lys(Boc)-Ala-Asp(O-1-Ada)-Gly-Ala before and after treatment with piperidine. Peptides were first eluted with 100% of buffer A for 3 min, then eluted with a 0-22% buffer B linear gradient over 22 min at 1.5 ml/min, monitored at 220 nm. See Materials and Methods for buffer compositions.
A. Peptide without post-synthesis piperidine treatment.
B. Peptide-resin treated with 20% piperidine for 4 hrs.

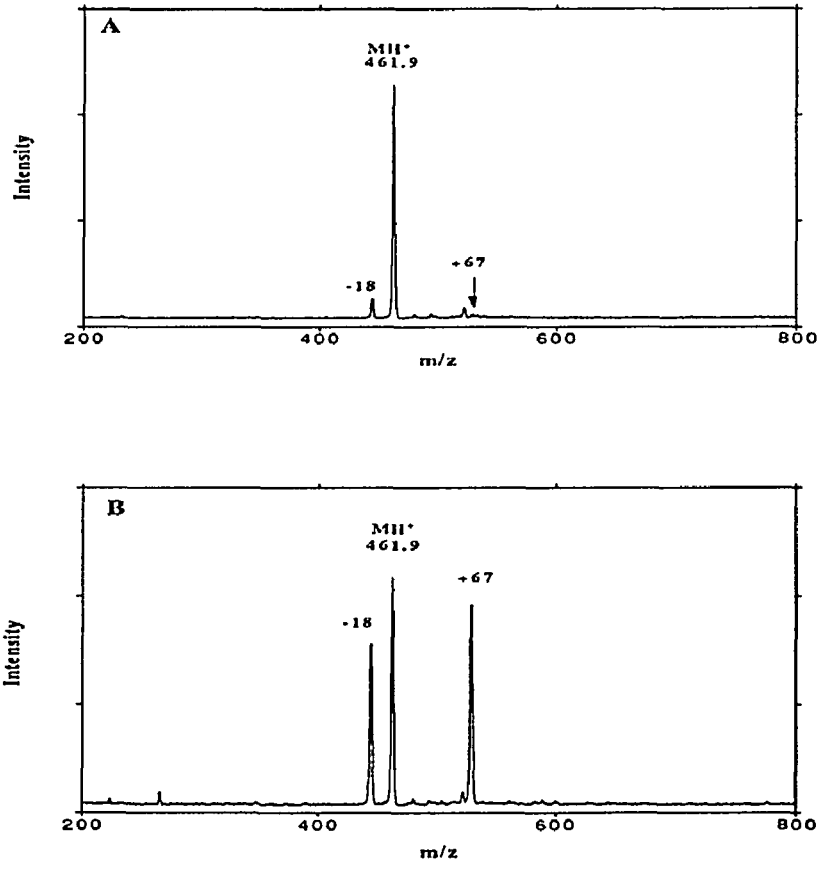


Figure 3.14. Portion of the electrospray mass spectra of peptide Lys(Boc)-Ala--Asp(O-1-Ada)-Gly-Ala. The theoretical mass of the peptide is 460.5 u.
A. Peptide without post-synthesis piperidine treatment.
B. Peptide resin treated with 20% piperidine for 4 hrs.

Table 3.7. HPLC (Figure 3.11) and MS analysis of model peptide: Lys(Boc)-Ala-Asp(O-1-Ada)-Asn(Trt)-Ala

Peaks separated by HPLC	Retention time (min)	Effect of 20% piperidine treatment on HPLC area (%) Before	After	Mass difference Δ*	Identification
Major	2.9/3.2	90.1	51.6	0	Desired peptide
(a) and (b)	5.5/12.1	7.6	16.8	-18**	Aspartimide
(c)	18.6	2.3	31.6	+67	Piperidine adduct

* Compared with mass of the desired peptide, 517.5 u.
** These side reaction products have the same mass but elute with different retention times. The reason for this is not known.

Table 3.8. HPLC (Figure 3.13) and MS analysis of model peptide: Lys(Boc)-Ala-Asp(O-1-Ada)-Gly-Ala

Peaks separated by HPLC	Retention time (min)	Effect of 20% piperidine treatment on HPLC area (%) Before	After	Mass Difference Δ*	Identification
Major	2.9/3.4	94.6	66.4	0	Desired peptide
(a) and (b)	4.7/7.3	4.5	17.8	-18**	Aspartimide
(c) and (d)	17.0/18.1	0.8	15.8	+67**	Piperidine adduct

* Compared with mass of the desired peptide, 460.5 u.
** These side reaction products have the same mass but elute with different retention times. The reason for this is not known.

An intriguing possibility is that the side-chain amide in Asn could form a hydrogen bond with its backbone NH group, and the thus formed six-member ring would prevent the aspartimide formation. It may be easier to form the ring when the side-chain in Asn without protection.

To test this idea, an experiment was carried out using Asn with and without the side chain blocking group, Trt. The model peptide was Lys(Boc)-Ala-Asp(Ot-Bu)-Asn-Ala. The resin-attached peptides were treated with piperidine for 4 hours. Two concentrations of piperidine, 10% and 20%, were used. Figure 3.15 and Table 3.9 show the HPLC results

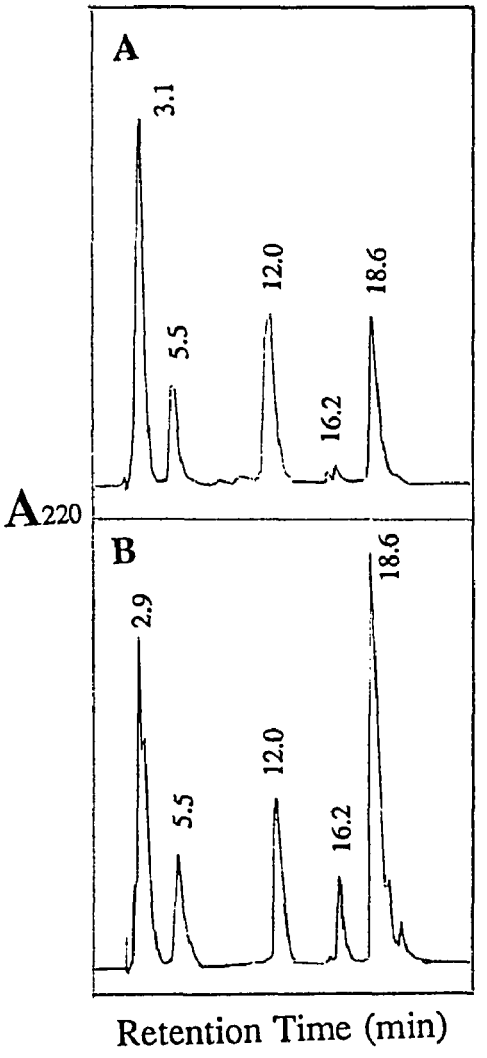


Figure 3.15. Reversed-phase HPLC profiles of peptide Lys(Boc)-Ala-Asp(Ot-Bu)-Asn--Ala after treatments with piperidine. Peptides were first eluted with 100% of buffer A for 3 min, then eluted with a 0-22% buffer B linear gradient over 22 min at 1.5 ml/min, monitored at 220 nm. See Materials and Methods for buffers composition.
A. Peptide-resin treated with 10% piperidine for 4 hrs.
B. Peptide-resin treated with 20% piperidine for 4 hrs.

Table 3.9. HPLC (Figure 3.15) and MS analysis of model peptide: Lys(Boc)-Ala-Asp(Ot-Bu)-Asn-Ala

Peaks separated by HPLC	Retention time (min)	Effect of piperidine treatment on HPLC area (%) Before	After 10%	After 20%	Mass Difference Δ*	Identification
Major	2.9/3.1	84.4	42.9	32.6	0	Desired peptide
(a) and (b)	5.5/12.0	15.4	37.6	26.9	-18**	Aspartimide
(c)	16.2/18.6	0.2	19.5	40.5	+67**	Piperidine adduct

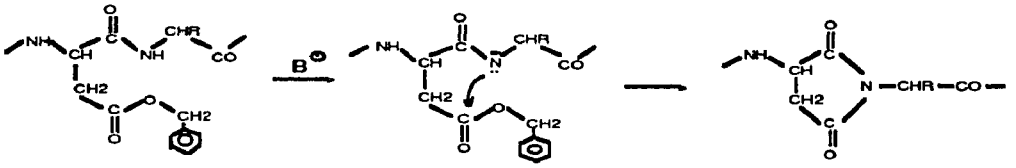
* Compared with mass of the desired peptide, 517.5 u.
** These side reaction products have same mass but elute with different retention times. The reason for this is not known.

for the peptide without the Trt protecting group. In contrast with the results of using Asn with Trt shown in Figure 3.9, the peptide with bare Asn (Figure 3.15.A) exhibited a greater extent of side reaction. Under 10% piperidine treatment, the total amount of aspartimide and piperidine adduct is 17% with Asn(Trt), and 57% with bare Asn. Thus, the proposed improvement of omitting protection of side chain of Asn is of no synthetic value. Possibly the hydrogen bond can not form between amides in side chain and backbone because the nitrogen lone-pair electrons are delocalized by the carbonyl group and both amides are very poor nucleophiles. Therefore, protection for the Asn side chain is necessary and reduces aspartimide formation in the peptide containing -Asp-Asn- sequence.

The influence of the piperidine concentration used in synthesis was also studied. As shown in two HPLC profiles of Figure 3.15, the total amount of aspartimide and piperidine adduct increases from 57% to 67% when concentration of piperidine for 4 hour treatment of the same resin increases from 10% to 20%. Among these the amount of piperidine adduct increases from 20% to 41%. It is strong evidence for the importance of the piperidine concentration which significantly enhances the extent of this side reaction.

3.4. Discussion

Bodanszky and coworkers (1978a, 1978b, 1978c) proposed a mechanism for base-catalyzed aspartimide formation in β-benzylaspartyl peptides. The nucleophilic attack to ring closure is intramolecular and leads to a thermodynamically favored structure. This offers an explanation for the aspartimide formation in the presence of a β-carboxyl protecting group removable with nucleophiles:



(Redrawn from Bodanszky et al., 1978a, 1978b, 1978c. The side-chain of Asp is protected with benzyl).

In current study, aspartimide formation has been found for two problematic sequences in -Asp-X-, when X=Asn or Gly. In general, the imide formation, when X=Gly, can be explained as the result of the absence of side chain steric hindrance. Base in this case may abstract the proton in backbone amide group. However, a bulky amino acid will also lead to the same side reaction, e.g. when X=Asn(Trt). Therefore, a corresponding explanation for -Asp-Asn- may be given based on the side chain structure of Asn. The side-chain amide group may remove the proton from the backbone amide group through resonance effect, and then the amide group in the backbone acts as a nucleophile to attack intramolecularly the side-chain of Asp. The proposed mechanism is shown in Figure 3.16..

Imide formation is enhanced by electron withdrawing side-chain protecting groups of Asp, and inhibited by electron donating or sterically hindered protecting groups (Tam et al., 1988). Of currently commercially available side-chain protecting groups in Fmoc chemistry, t-Bu is still a better choice, as it satisfies the both requirements of being sterically hindered and having electron rich characteristic. Nevertheless, for synthesis of

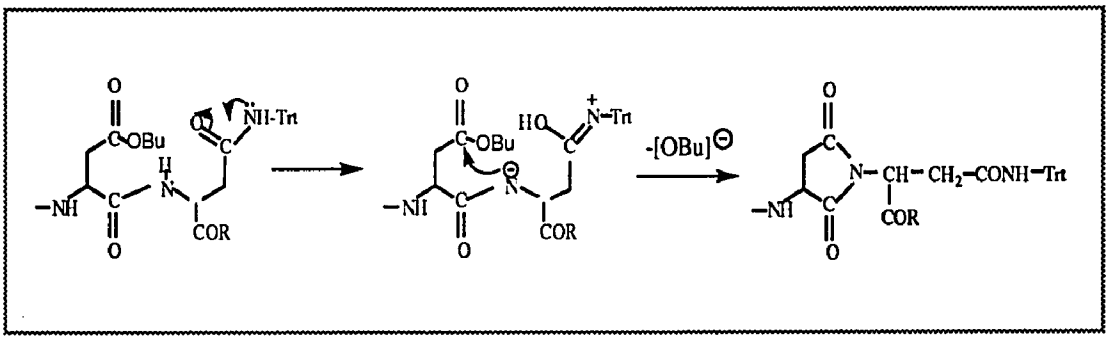


Figure 3.16. Proposed mechanism for aspartimide formation between -Asp-Asn-.

long peptides, it should no longer be considered a good candidate for protection in all cases, because the small amount of products from the side reaction will accumulate and significantly interfere for peptides containing -Asp(t-Bu)-X- sequences, where X=Asn or Gly. An argument against the common belief in the importance of the bulkiness of the protecting group could be made from our experimental results. The results using Asp(O-1-Ada) also support this point, where adamantyl is a side-chain protecting group which only provides bulkiness and no electric effect thus less resistant to imide formation. Therefore, the consideration regarding to electric effect of aspartimide related residues may be more important.

There are a number of approaches which may lead to minimize this side reaction: (i) finding better protecting groups for Asp and related amino acids; (ii) changing the solvent system (Adler et al., 1963); (iii) using additives (Martinez and Bodanszky, 1978); (iv) replacing piperidine with a milder base to deblock the Fmoc group, or washing resin with certain acids to get rid of trace piperidine (Bodanszky et al., 1978a; Carter and Lyon, 1990; Das and Lindstrom, 1991). The consideration on the peptide secondary structure which may influence the imide formation is also important (George-Nascimento et al., 1990).

Chapter 4

Preliminary structural analysis: some peculiar behaviors of the C-terminal EGF-like domain of factor IX

Abstract

C-terminal epidermal growth factor (EGF)-like domain in human blood coagulation factor IX (fIX_{EGF-C}) has been synthesized and folded with a desired EGF disulfide pattern (Cys 1-3, 2-4 and 5-6) using a two-step strategy. Preliminary structural analysis has been carried out for this target peptide and other available EGF-like peptides. Chou-Fasman secondary structure predictions based on the amino acid sequences show that for cysteine-rich domains, secondary structure is not well predicted, based on comparison with NMR determined structures. To further investigate the secondary structures of EGF-like domains, three EGF-like peptides and one related peptide have been examined using circular dichroism (CD). The samples include the N-terminal EGF-like domain of human factor IX (fIX_{EGF-N}), the N-terminal EGF-like domain of human factor X (fX_{EGF-N}), the C-terminal EGF-like domain of human factor IX (fIX_{EGF-C}) (target peptide), and a misfolded peptide with the fIX_{EGF-C} sequence but with Cys 1-2, 3-4, and 5-6 pairing. Peptides in different oxidized states including reduced peptides (no disulfides), fully oxidized peptides (three disulfides), and two-disulfide containing peptides (Cys 1-3 and 5-6 for EGF-like, or Cys 1-2 and 3-4 for misfolded), have been used in this study. This work presents a systematic attempt to study secondary structures in the initial, final, and one intermediate folding stages. The results indicate that at pH 4.2, all the spectra of reduced and oxidized peptides, except that of the oxidized fIX_{EGF-C}, have common CD features, which are similar to that of oxidized EGF molecule. The strong negative dichroic band observed at 200-202 nm is characteristic of aperiodic structure, and has been reported previously for EGF. The CD spectrum of fIX_{EGF-C} is completely unlike that of the other EGF-like peptides, suggesting a significantly different structure. In this work the 225-231

nm region of the CD spectra has been tentatively assigned to peptide conformations due to disulfide bridges in EGF-like peptides. Those three disulfide-containing EGF-like peptides demonstrate a weak positive band in that region. An attempt to determine the structure through 2D NMR experiment (pH 5.0) failed because only about 50% of the expected cross peaks were observed in the fingerprint region of NOESY spectrum. This may be due to rapid amide proton exchange with solvent water, suggesting an open structure. A comparison of primary structures, secondary structure predictions and CD spectra for all three EGF-like peptides with that of EGF is presented.

4.1. Introduction

An array of physicochemical techniques can be used for structural analysis, such as CD, vibrational CD (VCD), fluorescence, UV, IR, Raman, NMR, or x-ray crystallography. Among them CD is a sensitive, convenient, and powerful tool for studying protein conformation in solution. In particular CD is very useful for α-helix evaluation and less accurate for β-sheet and β-turns. Nevertheless, the most significant secondary structure in EGF family is β-sheet rather than α-helix. Another technique available is VCD which may play a critical role in the study of β-turns (Xue and Deem, 1995). However, to date the application of VCD has largely focusing on short model peptides. In an initial structural study, the Chou-Fasman secondary structure prediction method is applied to EGF family peptides.

4.2. Experimental

4.2.1. Secondary structure prediction

A Chou-Fasman-Algorithm program provided by Dr. Gerald D. Fasman at Brandeis University, constructed from a database of 64 proteins, was used. Following are the standard conformational parameters for α-helix, β-sheet, β-turn and random coil used in

the program. The assignments of helical structure (α) are given under each residue: H (strong helix former), h (helix former), I (weak helix former), i (helix indifferent), b (helix breaker), and B (strong helix breaker). The respective assignments for β-sheet under each residue are similar to that of α-helix.

Amino acid	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
$P\alpha$	139	95	106	144	111	64	100	99	121	130	132	78	55	112	100	72	78	97	103	73
$P\beta$	79	107	66	51	123	87	87	157	73	117	101	66	62	100	94	94	133	164	124	131
P_t	66	119	146	74	60	156	95	47	101	59	60	156	152	98	95	143	96	50	96	114
α	H	i	h	H	h	B	I	i	h	H	H	i	B	h	h	b	i	i	I	B
β	i	h	b	b	h	i	i	H	b	h	I	b	B	I	i	i	h	H	h	h

Where $P\alpha$, $P\beta$ and P_t are potentials of α-helix-forming, β-sheet-forming and β-turn-forming. α and β are α-helix and β-sheet structures.

4.2.2. Circular dichroism spectra

CD spectra were recorded on an Aviv 62DS CD spectropolarimeter at The Rockefeller University. The measurements were made at room temperature (18-20°C) using a circular quartz cell with a path length of 1-mm. The estimated percent of α helix, β sheet, β turn, and random coil structure in the peptides was calculated with Prosec, a program based on fifteen proteins (Yang et al., 1986). The spectra were measured from 250 to 190 nm at 1nm intervals, average time 4 seconds per interval, and ten scans were averaged. EGF-like peptides were dissolved in distilled water at a concentration of 0.1 ± 0.05 mg per ml (2.0 x 10^{-5} M ± 1.0 x 10^{-5} M). pH was adjusted to 4.2 ± 0.2 by addition of diluted HCl or NaOH. The CD data were recorded by the instrument directly in terms of degrees ellipticity, θ. Molar ellipticity [θ], in deg cm^2 $dmol^{-1}$, was calculated using the measured data multiplied by a normalization factor.

$$\text{Normalization factor} = \frac{1}{10 \times 0.1 \times \text{number of residues} \times \text{concentration in M}}$$

where 10 is a conversion factor and 0.1 is the path length of the cell in cm. Concentrations were calculated through both UV measurements and amino acid analysis. The molar extinction coefficients, ε_{280}, for UV data treatments are taken from the literature (Gill and Hippel, 1989): Trp, $\varepsilon_{280} = 5690\ M^{-1}\ cm^{-1}$; Tyr, $\varepsilon_{280} = 1280\ M^{-1}\ cm^{-1}$; and for cystine $\varepsilon_{280} = 120\ M^{-1}\ cm^{-1}$. The calculated molar extinction coefficients and number of residues for various test peptides are given bellow.

Peptide	Molar Extinction Coefficients ($M^{-1}\ cm^{-1}$)			
	Number of residues	No disulfide	Two disulfides	Three disulfides
$fIX_{EGF\text{-}N}$	43	8250	-	8970
$fIX_{EGF\text{-}C}$	45	1280	1760	2000
$fX_{EGF\text{-}N}$	43	2560	3040	3280

4.2.3. General Methods

Amino acid hydrolysis of purified free peptides were performed with 6N HCl/propionic acid in evacuated, sealed tubes at 110°C for 24 hr. Then samples were diluted to a concentration of about 0.1 nmol/µl, and 50 µl of the diluted solution was injected in a Beckman amino acid analyzer (System 6300).

4.3. Results and Discussion

4.3.1. Prediction of the secondary structure for EGF-like peptides based on Chou-Fasman model

Table 4.1 lists primary structures of human EGF and EGF-like domains in human factors IX and X. The amino acids and the loop size are varied, but the pattern of cysteine pairing is conserved.

Chou-Fasman secondary structure prediction based on statistical surveys of proteins of known structure has achieved a great deal of success because of its relative simplicity

Table 4.1. Comparison of the primary structure of EGF-like domains in human factors IX and X with human EGF.

Cysteines and loops / Peptides		1	pre A-Loop	2	Loop A	3	Loop B	4		5	Loop C	6	
EGF	NSDSE	C	PLSHDGY	C	LHDGV	C	MYI EALDKYA	C	N	C	VVGYIGER	C	QYRDLKWWELR
fIX_{EGF-N}	YVDGDQ	C	ESNP	C	LNGGS	C	KDDI NSYE	C	W	C	PFGFEGKN	C	ELDVT
fX_{EGF-N}	YKDGDQ	C	ETSP	C	QNQGK	C	KDGLGEYT	C	T	C	LEGFEGKN	C	ELFTR
fIX_{EGF-C}	LDVT	C	NIKNGR	C	EQF	C	KNSADNKVV	C	S	C	TEGYRLAENQKS	C	EPAV

* Disulfide bridges in EGF peptides are Cys 1-3, 2-4, and 5-6.

Note: Pro, a α helix and β sheet breaker and a good turn former, is in bold.

Gly, a α helix breaker and a good turn former, is underlined.

and reasonably high degree of accuracy (Prevelige and Fasman, 1989). To our knowledge no prediction for EGF-like peptides has been reported, although studies have been carried out for EGF (Holladay et al., 1976; Narhi et al., 1992). An attempt to predict EGF-like peptide secondary structure by using the Chou-Fasman algorithm has been made. Table 4.2 shows the main part of the original output file from the Chou-Fasman algorithm. The potentials for α-helix and β-sheet predicted for EGF and three EGF-like peptides are shown in Table 4.3. The probability of forming β-turns for these peptides are shown in Figure 4.1. To judge how valuable this program is for cysteine-rich EGF peptides, the predicted structures are compared with reported structures for EGF, $fIX_{EGF\text{-}N}$ and $fX_{EGF\text{-}N}$ as determined by NMR (Makino K. et al., 1987; Cook et al., 1987; Kohda and Inagaki, 1992; Huang et al., 1991; Baron et al., 1992; Selander et al., 1990; Ullner et al., 1992).

For EGF, NMR results found no α-helix, but did find antiparallel β-sheet consisting of 19-24 (VCMYIE) and 27-32 (DKYACN) (Makino et al., 1987; Kohda and Inagaki, 1992). The prediction indicates 16-23 (HDGVCMYI) and 31-37 (CNCVVGY) having high $\langle P\beta \rangle$ values and therefore possible β-sheet. The solution structure of the N-terminal EGF-like domain of factor IX is also different from the predicted secondary structure. Two antiparallel β-sheets are found in the NMR determined structure (Huang et al., 1991). The first one consists of 16-20 (GSCKD) and 25-29 (YECWC) and the second shorter one involves 32-34 (GFE) and 40-42 (LDV). The prediction misses an entire part (25-29) in the first sheet and the entire piece (32-34) in the second sheet. The third comparison is for $fX_{EGF\text{-}N}$. The NMR data (Ullner et al., 1992) is reported for *bovine* species, rather than *human* species, and the secondary structure is found to be almost identical to that of $fIX_{EGF\text{-}N}$. The prediction results are also similar to that for $fIX_{EGF\text{-}N}$. One entire part in each of the antiparallel β-sheets is mis-predicted. Predictions, in general, indicate a number of α-helix in EGF peptides (Table 4.3), but α-helix has never been found in an EGF family protein. Therefore, for these peptides predictions are poorly correlated to actual NMR determined structures. The β-turn profiles (Figure 4.1) have different patterns for

Table 4.2. The Chou-Fasman-Algorithm prediction of secondary structure for human EGF and EGF-like domains in human blood coagulation factors IX and X. The program is provided by Dr. Fasman at Brandeis University. There is no data available for the last three residues in EGF due to the limitation of the prediction window.

	$fIX_{EGF\text{-}N}$					$fX_{EGF\text{-}N}$				
	<Pa>	<Pb>	<Pt>	<pt>		<Pa>	<Pb>	<Pt>	<pt>	
1 Y	85	112 *	116	1.07e-004	*	91	89	129	2.57e-004	*
2 V	93	95	124	1.05e-004	*	98	73	137	9.31e-005	*
3 D	97	79	136	2.19e-004	*	97	79	136	2.19e-004	*
4 G	94	90	129	5.31e-005		94	90	129	5.31e-005	
5 D	114 *	81	109	1.08e-004	*	114 *	81	109	1.08e-004	*
6 Q	105 *	88	108	3.20e-005		107 *	97	96	2.39e-005	
7 C	97	79	123	1.02e-004	*	97	96	108	6.16e-005	
8 E	87	68	131	1.01e-004	*	87	85	116	5.14e-005	
9 S	75	82	142	4.33e-005		75	99	127	5.20e-005	
10 N	89	88	121	3.97e-004	*	93	90	128	4.14e-004	*
11 P	89	88	121	1.77e-005		85	83	131	1.82e-005	
12 C	91	94	122	1.08e-004	*	99	93	117	2.73e-004	*
13 L	84	89	131	1.46e-004	*	91	88	127	3.45e-005	
14 N	69	93	152	2.76e-004	*	93	81	127	2.95e-004	*
15 G	73	93	143	1.39e-004	*	98	91	118	5.80e-005	
16 G	88	90	129	1.58e-004	*	100 *	85	119	1.30e-004	*
17 S	98	85	127	3.71e-005		110 *	79	116	1.70e-005	
18 C	107 *	78	128	2.48e-004	*	96	83	130	4.66e-004	*
19 K	108 *	90	110	6.06e-005		105 *	85	115	8.05e-005	*
20 D	97	88	123	1.91e-005		91	89	129	6.84e-005	
21 D	88	95	123	1.01e-004	*	100 *	85	111	3.10e-005	
22 I	80	112 *	115	5.58e-005		102 *	96	100	4.99e-005	
23 N	91	85	121	1.63e-004	*	89	100 *	110	5.51e-005	
24 S	96	95	112	7.69e-005	*	97	105 *	100	3.03e-005	
25 Y	103 *	103 *	100	9.61e-005	*	81	126 *	106	8.19e-005	*
26 E	109 *	97	102	2.42e-005		96	120 *	107	3.79e-005	
27 C	87	100 *	121	1.54e-005		99	116 *	98	1.32e-004	*
28 W	91	104 *	106	9.02e-006		111 *	102 *	87	1.05e-005	
29 C	81	94	121	4.43e-004	*	108 *	90	102	4.36e-005	
30 P	85	98	107	5.16e-005		112 *	94	87	4.52e-005	
31 F	107 *	96	87	2.09e-005		115 *	78	91	1.98e-005	
32 G	95	87	111	4.89e-005		95	87	111	4.89e-005	
33 F	110 *	83	97	6.39e-005		110 *	83	97	6.39e-005	
34 E	101 *	69	121	3.12e-005		101 *	69	121	3.12e-005	
35 G	89	83	133	2.87e-004	*	89	83	133	2.87e-004	*
36 K	109 *	74	112	3.42e-005		109 *	74	112	3.42e-005	
37 N	111 *	85	102	4.60e-005		111 *	85	102	4.60e-005	
38 C	118 *	85	99	2.61e-005		120 *	99	78	2.09e-005	
39 E	119 *	99	82	1.33e-005		115 *	106 *	72	7.19e-006	
40 L	102 *	120 *	87	1.48e-005		104 *	116 *	77	1.38e-005	
41 D	94	117 *	102	5.87e-005		102 *	105 *	88	5.99e-005	
42 V	87	117 *	105	7.13e-005		107 *	104 *	87	4.59e-005	
43 T	87	115 *	104	4.88e-005		111 *	97	93	3.71e-005	

fIX_{EGF-C}

		<Pa>	<Pb>	<Pt>	<pt>
1	L	102 *	120 *	87	1.48e-005
2	D	94	117 *	102	5.87e-005
3	V	87	117 *	105	7.13e-005
4	T	87	115 *	104	4.88e-005
5	C	98	100 *	105	1.53e-005
6	N	94	90	115	3.59e-005
7	I	90	95	115	1.44e-004 *
8	K	90	80	127	7.37e-005
9	N	84	88	131	1.73e-004 *
10	G	100 *	84	111	8.10e-005 *
11	R	112 *	88	96	2.80e-005
12	C	115 *	95	87	2.15e-005
13	E	115 *	95	87	4.57e-005
14	Q	109 *	100 *	94	3.37e-005
15	F	101 *	92	109	2.05e-005
16	C	91	85	129	3.47e-004 *
17	K	102 *	78	116	3.31e-005
18	N	98	76	127	6.34e-005
19	S	98	76	127	1.49e-004 *
20	A	111 *	71	117	1.20e-004 *
21	D	100 *	92	113	4.60e-005
22	N	98	116 *	89	2.75e-005
23	K	102 *	127 *	80	9.46e-006
24	V	90	132 *	90	3.69e-005
25	V	89	118 *	107	5.26e-005
26	C	85	110 *	119	1.91e-004 *
27	S	97	96	108	2.65e-005
28	C	95	94	111	1.88e-004 *
29	T	89	100 *	110	1.23e-004 *
30	E	95	90	109	4.61e-005
31	G	91	107 *	106	4.59e-005
32	Y	110 *	105 *	83	1.81e-005
33	R	128 *	85	73	3.92e-006
34	L	122 *	78	88	3.25e-005
35	A	118 *	74	98	6.74e-005
36	E	113 *	72	107	1.63e-005
37	N	95	83	124	1.20e-004 *
38	Q	100 *	93	115	1.36e-004 *
39	K	108 *	81	109	5.72e-005
40	S	91	78	122	3.33e-005
41	C	108 *	74	102	1.76e-005
42	E	108 *	89	85	3.13e-005
43	P	86	91	105	1.48e-005
44	A	100 *	107 *	82	6.36e-006
45	V	79	102 *	103	8.25e-005 *

Human EGF

		<Pa>	<Pb>	<Pt>	<pt>
1	N	82	80	147	4.25e-004 *
2	S	98	76	126	1.06e-004 *
3	D	104 *	79	120	2.01e-004 *
4	S	91	78	122	5.73e-005
5	E	106 *	84	101	7.06e-006
6	C	88	95	118	1.71e-004 *
7	P	92	89	112	1.72e-005
8	L	105 *	90	110	6.39e-005
9	S	88	82	135	1.53e-004 *
10	H	88	91	127	3.66e-004 *
11	D	84	97	133	1.82e-004 *
12	G	90	110 *	112	5.43e-005
13	Y	102 *	109 *	96	8.45e-006
14	C	110 *	93	104	2.81e-005
15	L	103 *	88	114	7.80e-005 *
16	H	94	100 *	111	1.55e-004 *
17	D	90	106 *	117	4.48e-005
18	G	97	114 *	96	3.15e-005
19	V	99	125 *	85	5.75e-006
20	C	99	124 *	85	7.80e-005 *
21	M	112 *	110 *	73	3.68e-006
22	Y	113 *	104 *	75	1.25e-005
23	I	128 *	101 *	61	6.32e-006
24	E	129 *	78	86	1.24e-005
25	A	124 *	83	93	2.55e-005
26	L	107 *	96	105	6.04e-005
27	D	109 *	87	106	1.12e-004 *
28	K	107 *	97	100	1.60e-005
29	Y	96	95	113	6.64e-005
30	A	101 *	89	115	7.77e-005 *
31	C	91	111 *	111	7.67e-005 *
32	N	91	125 *	93	1.27e-005
33	C	88	130 *	93	3.04e-005
34	V	82	136 *	92	7.07e-005
35	V	83	134 *	91	3.36e-005
36	G	75	115 *	118	1.31e-005
37	Y	95	106 *	97	3.39e-005
38	I	101 *	97	93	2.39e-005
39	G	100 *	84	111	7.76e-005 *
40	E	112 *	88	96	6.81e-005
41	R	95	108 *	106	1.72e-005
42	C	95	108 *	106	1.41e-004 *
43	Q	97	97	113	3.86e-005
44	Y	102 *	102 *	103	1.09e-004 *
45	R	114 *	87	100	2.63e-005
46	D	115 *	95	100	4.42e-005
47	L	114 *	109 *	88	7.50e-005
48	K	117 *	93	91	2.93e-006
49	W	120 *	104 *	81	5.40e-006
50	W	119 *	96	81	1.41e-005

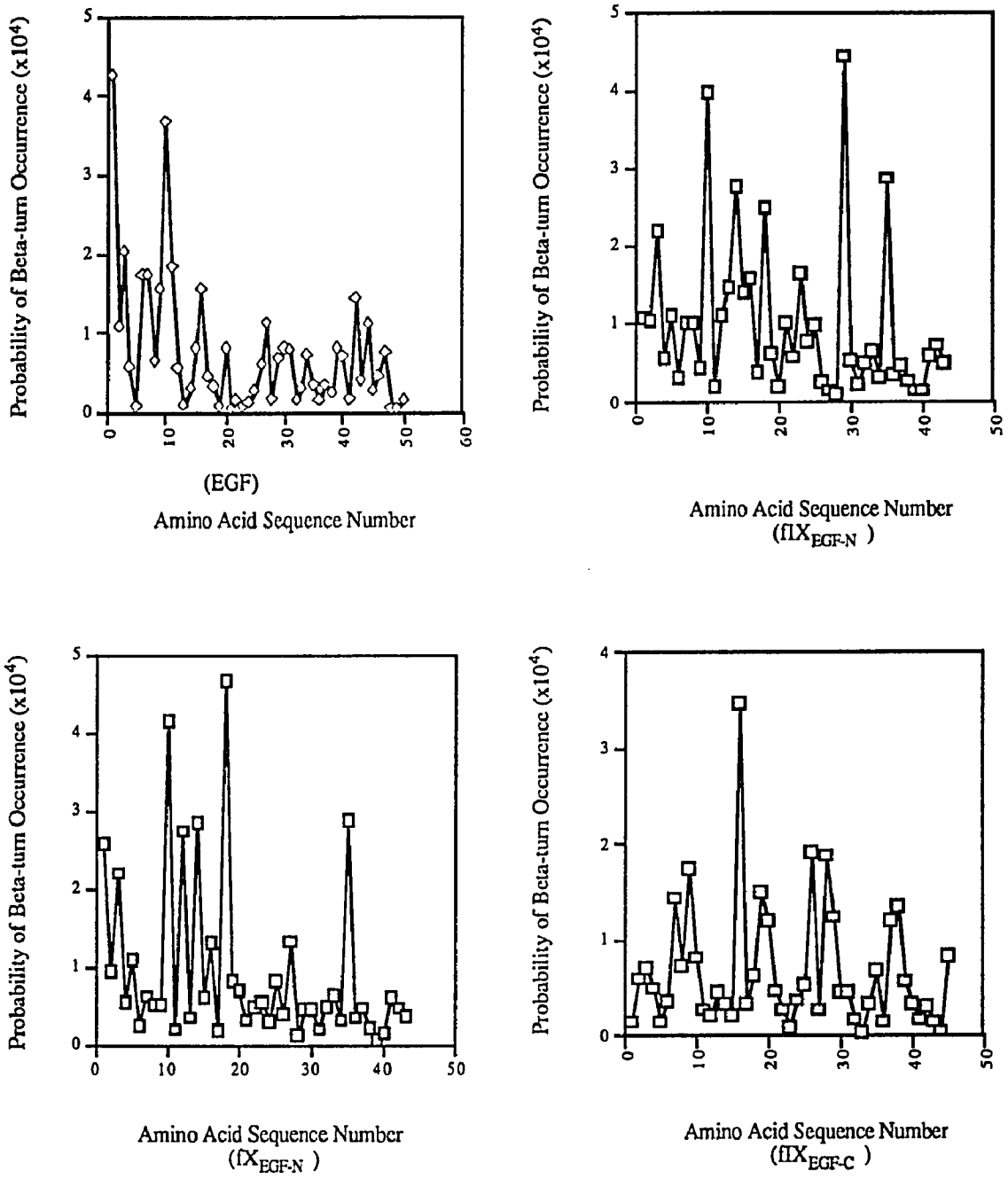


Figure 4.1. Probability of predicted β-turns in EGF and EGF-like peptides.
Upper left: EGF; upper right: fIX$_{EGF-N}$; bottom left: fX$_{EGF-N}$; bottom right: fIX$_{EGF-C}$.

Table 4.3. Conformational prediction for EGF and EGF-like peptides.

EGF			fIX_{EGF-N}			fX_{EGF-N}			fIX_{EGF-C}		
Predicted	$<P_\alpha>$	$<P_\beta>$	Predicted	$<P_\alpha>$	$<P_\beta>$	Predicted	$<P_\alpha>$	$<P_\beta>$	Predicted	$<P_\alpha>$	$<P_\beta>$
13-15, α	1.05	0.97	17-20, α	1.02	0.85	16-22, α	0.99	0.86	1-5, β	0.94	1.14
16-23, β	1.04	1.10	23-26, β	0.98	1.01	23-28, β	0.94	1.11	10-21, α	1.04	0.86
24-30, α	1.10	0.89	31-39, α	1.07	0.85	29-43, α	1.07	0.91	22-32, β	0.95	1.09
31-37, β	0.86	1.22	40-43, β	0.93	1.17	39-43, β	1.08	1.06	33-42, α	1.09	0.81
41-44, β	0.97	1.04							43-45, β	0.88	1.00
45-53, α	1.17	0.97									

$<P_\alpha>$ and $<P_\beta>$ are the average conformational potential for the predicted region to be in the helical and β-sheet conformation. They are calculated by the sum of the $<P_\alpha>$ or $<P_\beta>$ values of the individual residues divided by the total residues in the segment under consideration, respectively. The obtained values are then divided by a factor of 100, according to Chou-Fasman method.

each peptide, despite the fact that EGF and the two N-terminal EGF-like peptides are known to have similar secondary structure (Huang et al., 1991; Baron et al., 1992; Selander et al., 1990; Ullner et al., 1992).

The Chou-Fasman algorithm for prediction of protein secondary structure is one of the most widely used predictive schemes. A previous finding by Narhi and coworkers (1992) is instructive: the secondary structure predictions correlate most closely with the structure of the reduced (no disulfides) EGF molecule. Our results suggest that secondary structure prediction for cysteine-rich peptides of unknown structure, such as fIX_{EGF-C}, will probably fail. This is consistent with a statistical examination (Sheridan et al., 1985), in which prediction of the secondary structure for normal domains (cysteine content below 4.5%) is found meaningful, but for cysteine-rich domains may not be as reliable.

4.3.2. CD data analysis of secondary structure for EGF-like peptides

The CD spectra of two peptides, fX_{EGF-N} and fIX_{EGF-C}, were recorded in their different oxidation states (containing no disulfide, two disulfide, or three disulfide bonds).

These peptides were available since they had been synthesized with a two-step approach. Besides these two peptides, $fIX_{EGF\text{-}N}$ was also investigated in its fully reduced form and fully oxidized form. A misfolded $fIX_{EGF\text{-}C}$ (peptide A, described in chapter 2 of this dissertation), synthesized through the two-step method but with a disulfide pairing of Cys 1-2, 3-4, and 5-6, was studied as well. For convenience these peptides are given laboratory labels as follows:

Peptide	$fIX_{EGF\text{-}N}$	$fX_{EGF\text{-}N}$	$fIX_{EGF\text{-}C}$	Misfolded $fIX_{EGF\text{-}C}$
Label	9N	10N	9C	A
No disulfide	0ss9N	0ss10N	0ss9C	0ssA
Two disulfides	-	2ss10N	2ss9C (Cys 1-3/5-6)	2ssA (Cys 1-2/3-4)
Three disulfides	3ss9N	3ss10N	3ss9C (Cys 1-3/2-4/5-6)	3ssA (Cys 1-2/3-4/5-6)

where 9N and 10N are assigned to the N-terminal EGF-like peptides in factors IX and X, respectively, and 9C is assigned to the C-terminal EGF-like peptide in factor IX. A is for the misfolded peptide A described in chapter 2 of this dissertation, which has the sequence of peptide 9C but different disulfide pairing. Unfortunately no EGF peptide was available to make parallel measurements under identical conditions.

Figures 4.2, 4.3, 4.4 and 4.5 show the CD spectra for $fIX_{EGF\text{-}N}$ (peptide 9N), $fX_{EGF\text{-}N}$ (peptide 10N), misfolded $fIX_{EGF\text{-}C}$ (peptide A), and $fIX_{EGF\text{-}C}$ (peptide 9C). The observed difference in CD spectra between 0ss10N and 2ss10N is not as much as that among other peptides (Figure 4.3). The reason may be either that the 0ss10N peptide was partially oxidized, containing a component of disulfide-containing peptide, or that the conformations of 0ss10N and 2ss10N are similar. The latter possibility is consistent with the observed rapid folding behavior of reduced peptide 10N in the 1st-step oxidation after synthesis.

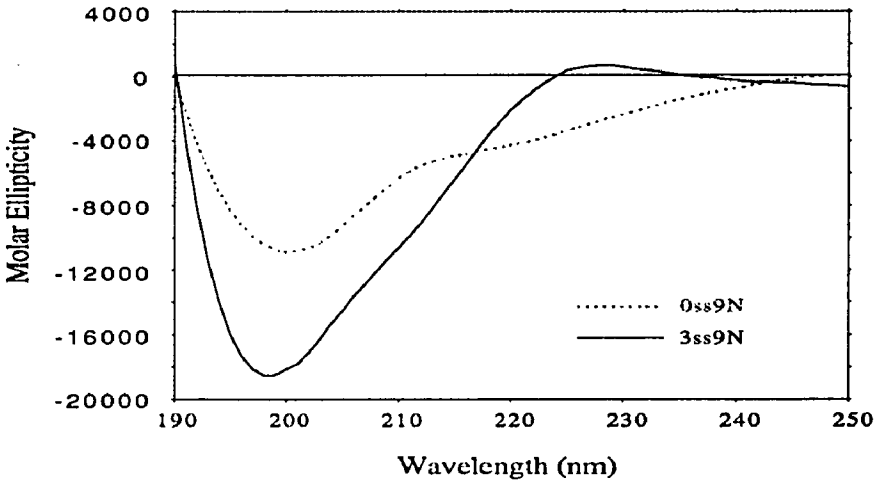


Figure 4.2. CD spectra of $fIX_{EGF\text{-}N}$ at pH 4.2 ± 0.2. 3ss9N is oxidized peptide. 0ss9N is reduced peptide.

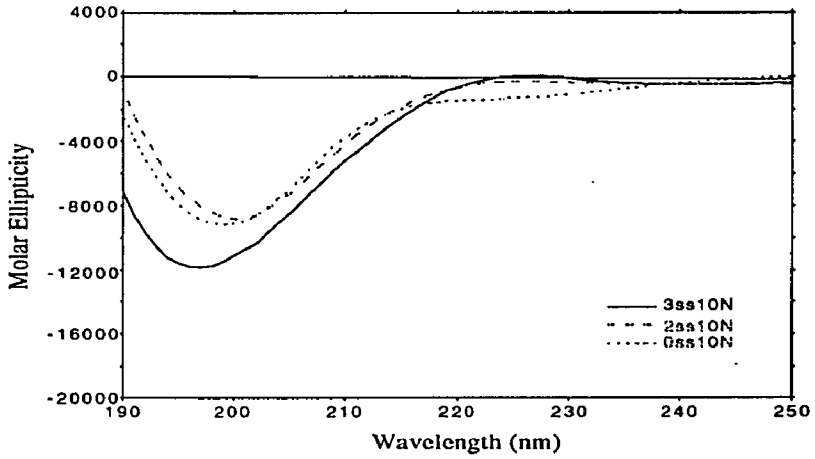


Figure 4.3. CD spectra of $fX_{EGF\text{-}N}$ at pH 4.2 ± 0.2. 0ss10N is reduced peptide. 2ss10N contains Cys 1-3 and 5-6, Cys 2 and 4 are protected by Acm. 3ss10N is oxidized peptide.

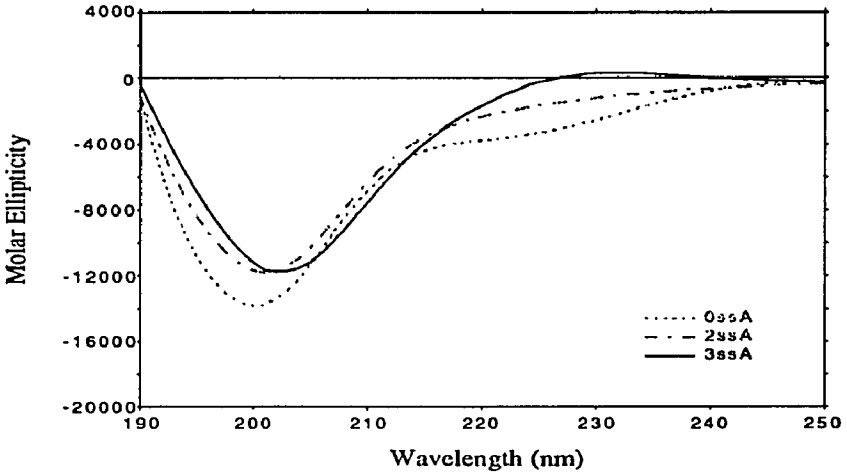


Figure 4.4. CD spectra of misfolded $fIX_{EGF\text{-}C}$ at pH 4.2 ± 0.2. 0ssA is reduced peptide. 2ssA contains Cys 1-2 and 3-4, Cys 5 and 6 are protected by Acm. 3ssA is oxidized peptide with Cys 1-2, 3-4, and 5-6.

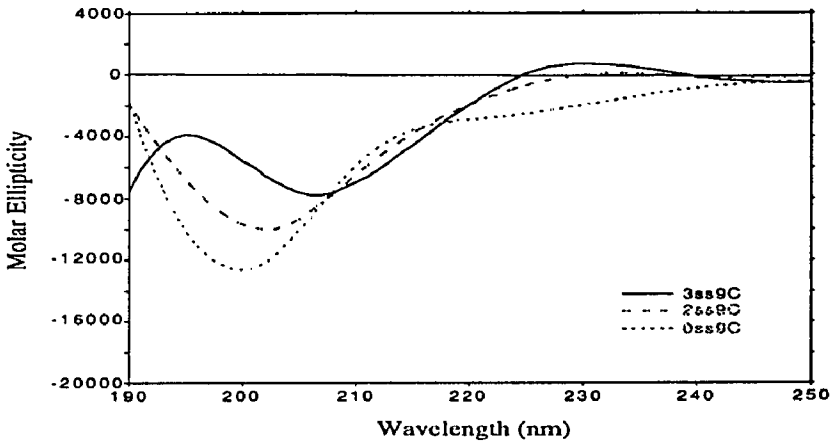


Figure 4.5. CD spectra of $fIX_{EGF\text{-}C}$ at pH 4.2 ± 0.2. 0ss9C is reduced peptide. 2ss9C contains Cys 1-3 and 5-6, Cys 2 and 4 are protected by Acm. 3ss9C is oxidized peptide with Cys 1-3, 2-4, and 5-6.

All of the peptides, except 3ss9C, exhibit CD spectra having a strong negative dichroic band with a minimum in the range 200-202 nm. This dichroic band is a characteristic property of nonhelical polypeptides and proteins (Beychok, 1966), indicating aperiodic structure. This is consistent with the known structure of 2ss9N and 3ss10N, and strongly suggests non-helical structure for 3ss9C and 3ssA. The spectra also have a distinguishing feature in the region 218-232 nm, the band magnitude varying with number of disulfide bridges. Only fully oxidized peptides with three disulfide bonds show a weak positive band in this region. The smaller the number of disulfides, the more negative the CD spectra appear in the region. Therefore, the spectra in the region of 218-232 nm may depend on the disulfide conformation of EGF-like peptides. The observations from other groups of CD spectra for EGF related to this region were as follows. Cohen and coworkers (Holladay et al., 1976) assigned the region of the spectrum above 225 nm as arising mainly from side-chain chromophores (e.g. tyrosines, tryptophans, and disulfides in EGF). They also reported that the 232 nm extremum seen in EGF spectra is particularly sensitive to guanidinium chloride, a strong denaturant. Prestrelski's group (1992) compared the difference in the CD spectra between EGF and TGFα and stated that the positive band at 220 nm in the EGF spectrum may be ascribed to the transition of aromatic groups in this region: EGF has three Trp and TGFα has none.

CD spectra from current studies are also classified into three catalogues according to no-disulfide, two-disulfide, and three-disulfide peptides. They are shown in Figures 4.6, 4.7, and 4.8. Unlike the spectra of EGF (Figure 1.3, in which reduced molecule is very different from the oxidized EGF), all EGF-like domains but 3ss9C peptide, no matter whether they are reduced or oxidized, have similar CD spectra to that of oxidized EGF. Since the structure of EGF is nearly identical in a broad pH range, from pH 2 to 8 (Kohda and Inagaki, 1992), and since the oxidized and reduced EGF molecules are known to exhibit dissimilar CD spectra at either acidic or neutral pH (Narhi et al., 1992; Prestrelski et al., 1992), the different observations between EGF and EGF-like peptides in their

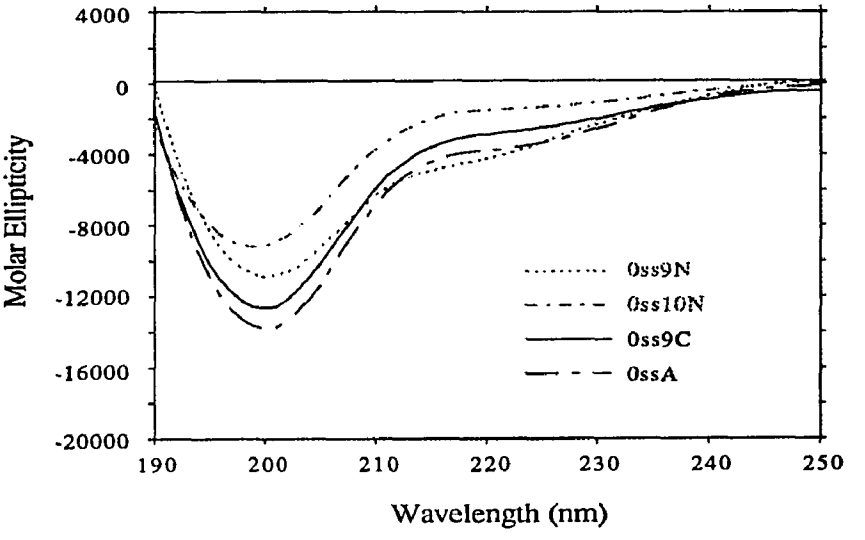


Figure 4.6. CD spectra of no disulfide-containing EGF-like peptides at pH 4.2 ± 0.2. 0ss9N is reduced $fIX_{EGF\text{-}N}$ peptide. 0ss10N is reduced $fX_{EGF\text{-}N}$ peptide. 0ssA is reduced $fIX_{EGF\text{-}C}$ peptide with Acm blocking Cys 5 and 6. 0ss9C is reduced $fIX_{EGF\text{-}C}$ peptide with Acm blocking Cys 2 and 4.

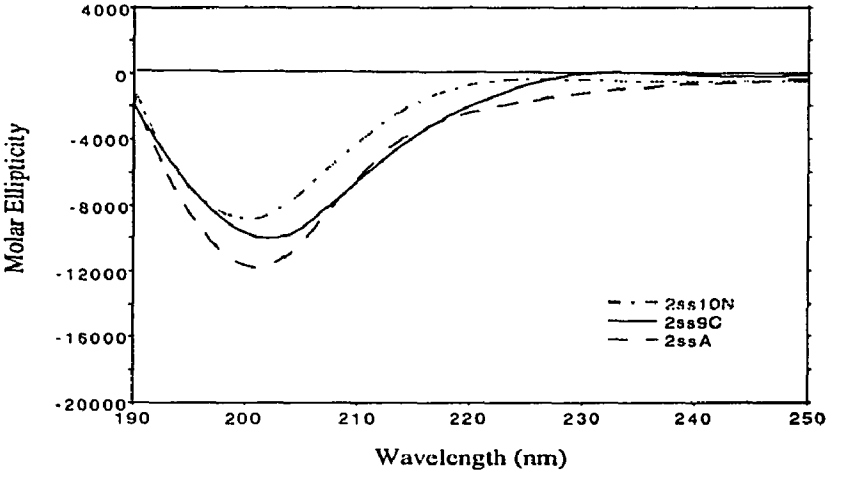


Figure 4.7. CD spectra of two disulfide-containing EGF-like peptides at pH 4.2 ± 0.2. 2ss10N is $fX_{EGF\text{-}N}$ peptide with Cys 1-3 and 5-6. 2ssA is $fIX_{EGF\text{-}C}$ peptide with Cys 1-2 and 3-4, and Acm blocking Cys 5 and 6. 2ss9C is $fIX_{EGF\text{-}C}$ peptide with Cys 1-3 and 5-6, and Acm blocking Cys 2 and 4. No 2ss9N peptide is available for this study due to synthetic approach.

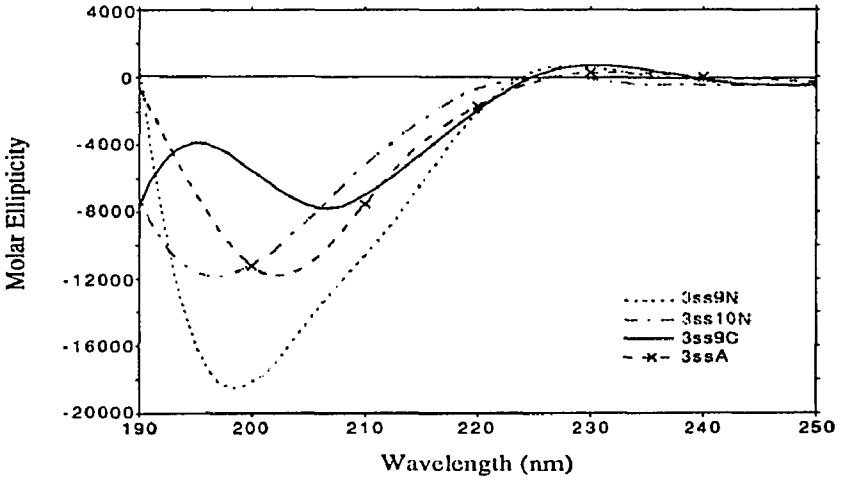


Figure 4.8. CD spectra of three disulfide-containing EGF-like peptides at pH 4.2 ± 0.2. 3ss9N is $fIX_{EGF\text{-}N}$ peptide. 3ss10N is $fX_{EGF\text{-}N}$ peptide. They both adopt an EGF disulfide pattern. 3ssA is misfolded $fIX_{EGF\text{-}C}$ peptide with Cys 1-2, 3-4, and 5-6 pairing. 3ss9C is $fIX_{EGF\text{-}C}$ peptide with Cys 1-3, 2-4, and 5-6 structure.

reduced state, compared to their oxidized state, do not arise as a result of differences in pH conditions. All of the CD spectra of reduced peptides (except peptide 10N) in the Figure 4.6 show a common strong negative band at 200 nm having an ellipticity of approximately -12,500 ± 1500 deg cm^2 per decimole. Peptide 0ss10N is a somewhat irregular as discussed before, showing a dichroic extremum at 199 nm with an ellipticity of -9,000 deg cm^2 per decimole.

In Figure 4.6, showing the spectra of fully reduced peptides, if consideration of peptide 0ss9N and 0ss10N is excluded, the spectra of 0ssA and 0ss9C are nearly identical, and spectrum of 0ss9N is similar to them. The only difference between peptides 0ss9C and 0ssA is the position of Acm protecting groups in cysteines. This may reflect a small difference in CD spectra in Figure 4.6. Unlike the other peptides the CD curve of peptide 0ss9N exhibits a small negative shoulder in the region of 211-221 nm. A negative band at 213 nm has been identified as a β-sheet indication (Taylor et al., 1972; Hollarday et al.,

1976), suggesting that peptide 0ss9N may have a greater content of β-sheet than the other non disulfide-containing EGF-like peptides in Figure 4.6.

Figure 4.7 shows the CD spectra of two-disulfide containing peptides. Peptides 2ssA and 2ss9C have same sequence and same number of disulfide bonds, but they have different disulfide connections:

Peptide 2ssA Peptide 2ss9C

Therefore, their secondary structures are not expected to be similar. The spectra have one short segment in common in 208-218 nm, a region indicating β-sheet structure in EGF. Both observations from CD spectra and calculations using a Prosec program indicate that peptide 2ss9C has a little more β-sheet than peptide 2ssA. Accordingly, the CD spectrum of peptide 2ssA shows more aperiodic structure at 200-202 nm than peptide 2ss9C, and the minimum of the peptide 2ss9C spectrum has a bigger shift (203 nm) than that of peptide 2ssA (201 nm), as compared with that of the reduced EGF-like peptides commonly at 200 nm. In the region about 230 nm, possibly attributable to disulfides, the peptide 2ss9C spectrum displays a larger magnitude than peptide 2ssA, although they have same number of disulfides. Peptide 2ss10N has a different sequence from peptides 2ssA/2ss9C, and it also displays a different CD spectrum. The origins of the difference are not understood.

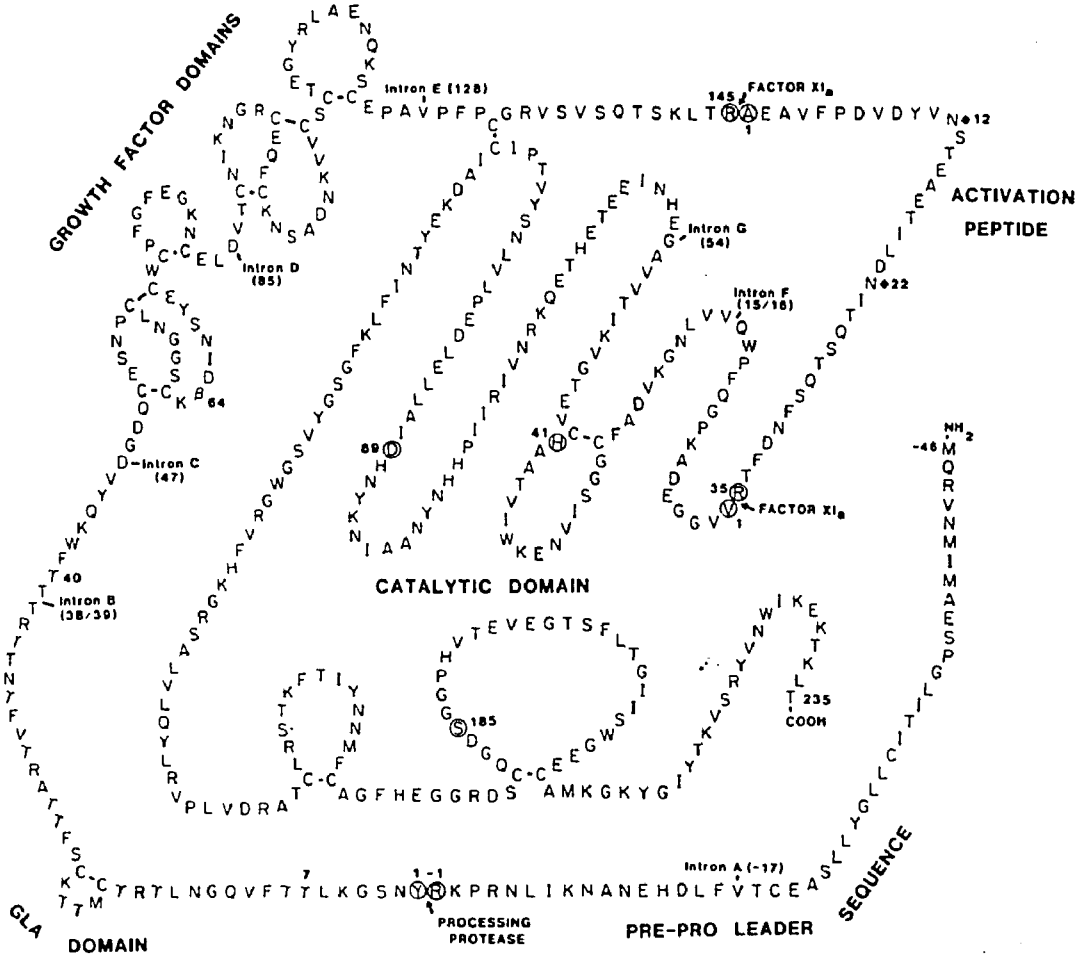
In Figure 4.8, the CD spectra for three disulfide-containing peptides are quite similar in the region above 225 nm, but very different in the part below 225 nm. All of the peptides have same number of disulfide bonds, and a Cys 5-6 bond in common. In the two peptides with the same sequence as that of the C-terminal EGF-like domain, peptide 3ssA has a misfolded bead disulfide configuration (Cys 1-2, 3-4, and 5-6), while peptide 3ss9C has an EGF-like global (Cys 1-3, 2-4, and 5-6). The CD spectra of 3ss9N, 3ss10N and 3ssA show strong negative minima at about 200 nm, although the negative minima of N-terminal

EGF-like peptides (3ss9N and 3ss10N) shift to shorter wavelength relative to the spectra of fully reduced peptides: from 200 nm to 196-199 nm. However, the CD spectrum of peptide 3ss9C is fundamentally different from the others, showing a positive extremum at 195 nm and a minimum at 207 nm. The feature of aperiodic structure (reflected in the CD spectra at about 200 nm) exists in the spectra of the two N-terminal EGF-like peptides and even in misfolded 3ssA spectra, but is not present in the 3ss9C spectrum. Possible explanations for the 3ss9C spectrum include: (a) new or different β-turns; (b) disulfide twist; (c) other unknown structure. A recent report (Perczel et al., 1991) demonstrates that β-turns can yield a CD spectrum very similar to that of peptide 3ss9C (fIX_{EGF-C}).

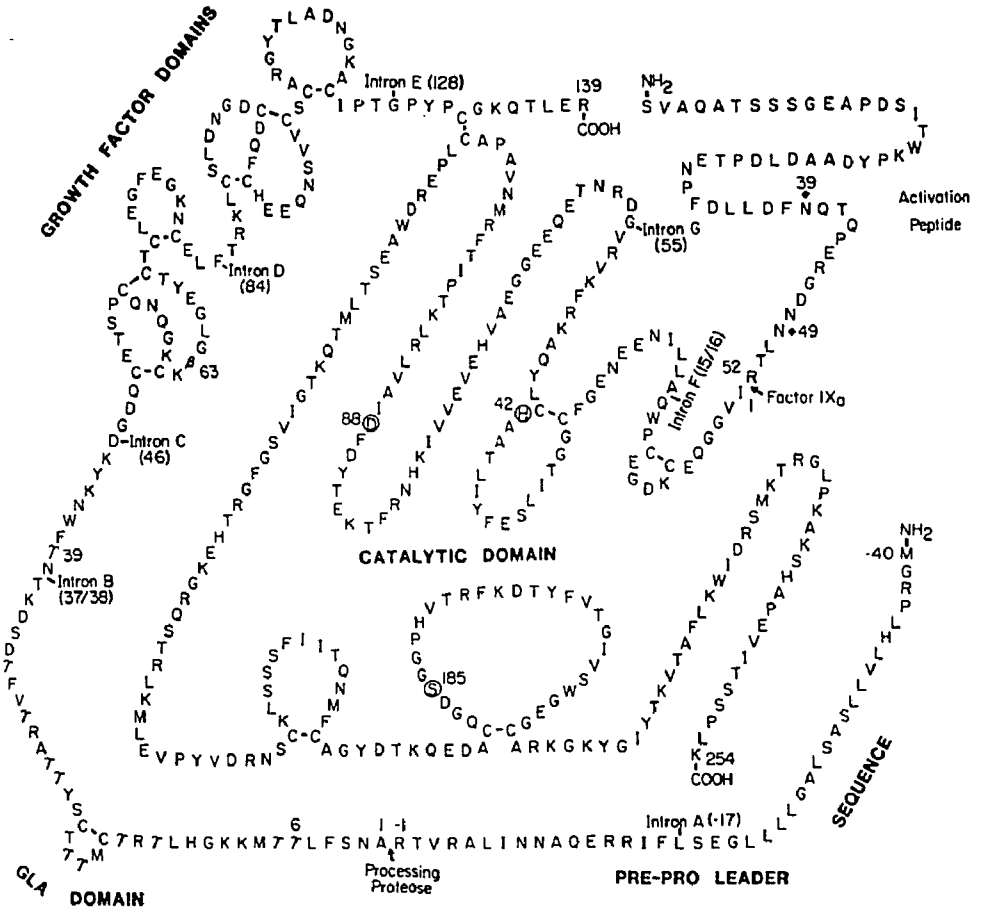
4.3.3. 2D NMR investigation of fIX_{EGF-C}

An attempt was made to determine the structure of fIX_{EGF-C} using 2D NMR techniques on 500-MHz and 600-MHz (proton frequency) spectrometers. However, the resulting spectra were not useful for a structure determination since too few amide protons were observed (data not shown). In the fingerprint region of the proton-proton NOESY spectrum (250-ms mixing time) in ^1H_2O at pH 5, only about half of the number of expected amide/alpha cross peaks could be observed, although the peaks in the NMR spectra were sharp and well resolved.

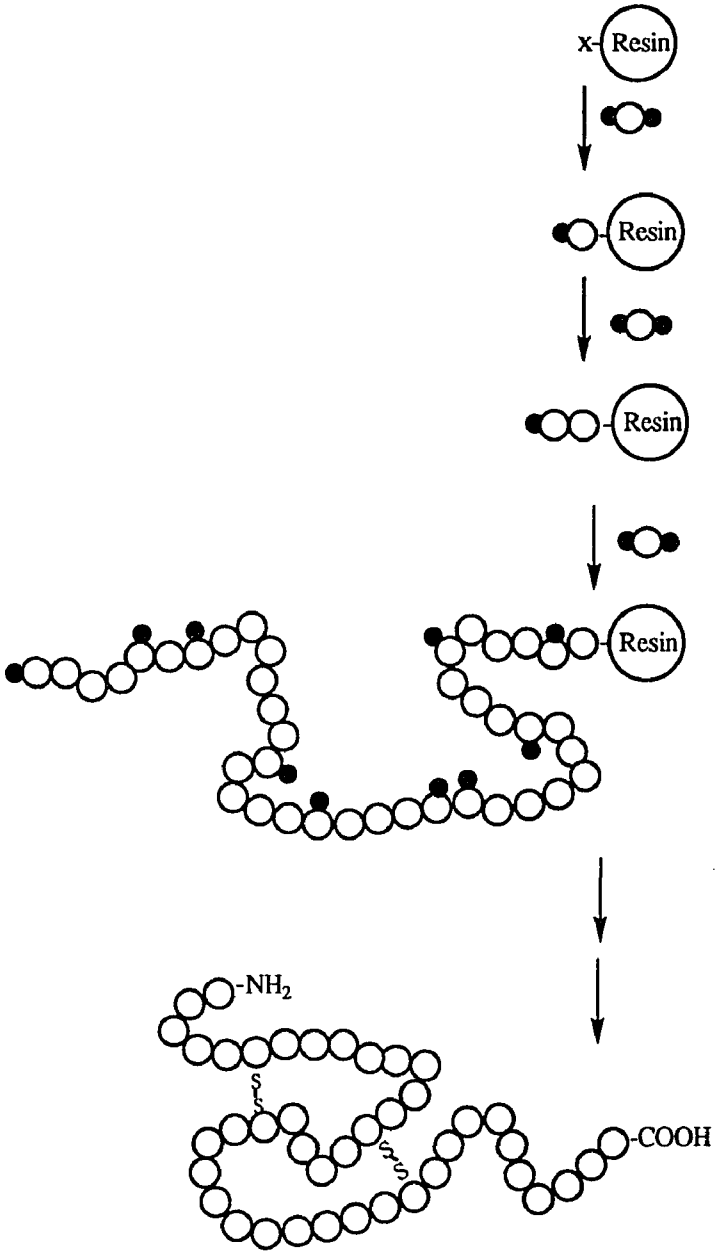
The reason for this peculiar phenomenon is not clear yet. Much more efforts need to be put in to understand the structure of the synthetic peptide fIX_{EGF-C}.



Appendix A.1. Amino acid sequence of human factor IX. β^{64} is β-hydroxyaspartic acid. γ refers to γ-carboxyglutamic acid. Authors stated: The proposed disulfide bonds in factor IX have been placed by analogy to those in bovine prothrombin and epidermal gowth factor. (Reproduced from Yoshitake et al., *Biochem* **24**, 3736-3750, 1985, Fig.7, with the permission of publisher).

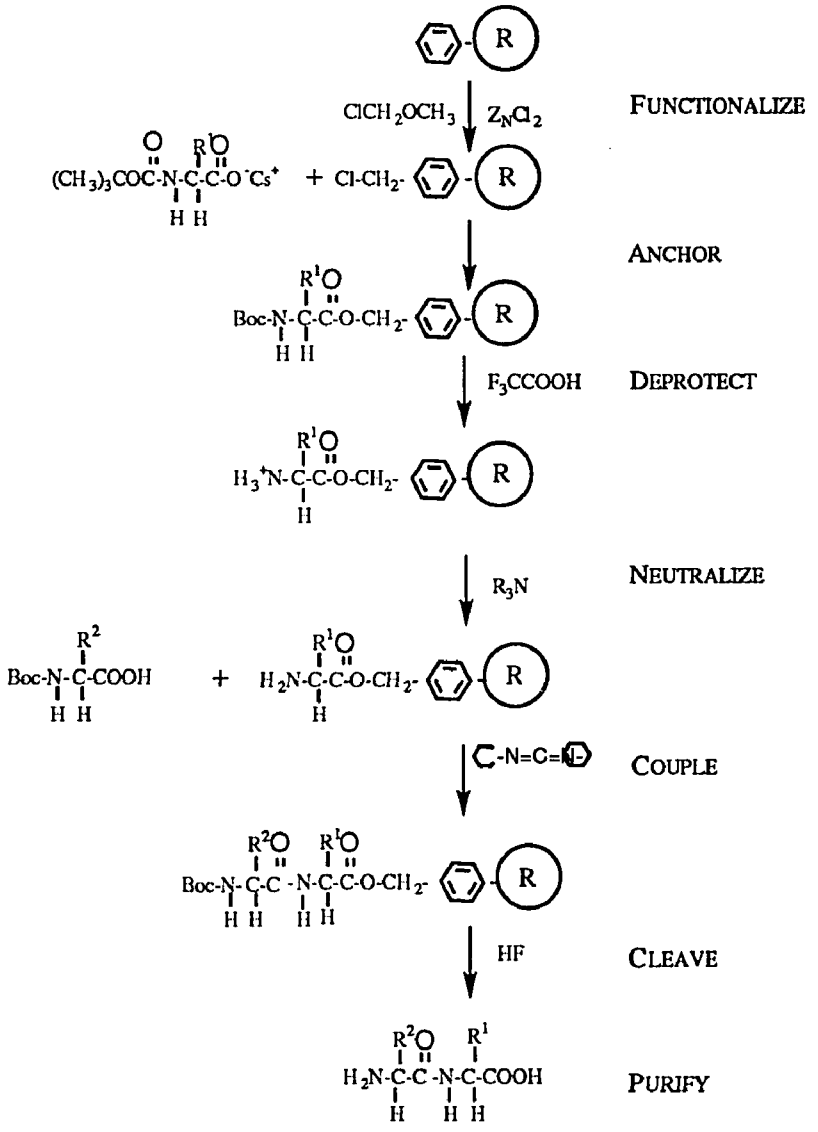


Appendix A.2. Amino acid sequence of human factor X. The Arg-Lys-Arg tripeptide that connects the light and heavy chains during biosynthesis is not shown. β^{63} is β-hydroxyaspartic acid. γ refers to γ-carboxyglutamic acid. Authors stated: The proposed disulfide bonds in factor X have been placed by analogy to those in bovine prothrombin and epidermal gowth factor. (Reproduced from Leytus et al., *Biochem* **25**, 5098-5102, 1986, Fig.2, with the permission of publisher).



Appendix B. 1. The general scheme for solid phase synthesis.

(Redrawn from: Bruce Merrifield, *Science*, **232**, 341-347, 1986)



Appendix B.2. A scheme for solid phase peptide synthesis.

(Redrawn from: Bruce Merrifield, *Science*, **232**, 341-347, 1986)

Appendix C. Peptide synthesis protocols

Table C.1. Boc chemistry

Step	Reagents	Repeat x Time(min)
Deprotection	50%TFA/DCM with 0.05%DMS	2 x 2
	50%TFA/DCM	1 x 20
Neutralization	DCM	3 x 1
	DMF	1 x 1
	5 % DIEA in DMF	3 x 2
	DMF	1 x 1
	isopropanol	2 x 1
	DCM	3 x 1
Coupling (3 methods)	(1) DCC coupling	
	3 equiv. Boc A.A. in DCM	1 x 1
	3 equiv. DCC in DCM	1 x 60
	(2) DCC/HOBt coupling (not for Asn in the His(Tos) containing peptide)	
	3 equiv. Boc A.A./DCM + 3 equiv. HOBt /DMF	1 x 1 1 x 15
	3 equiv. DCC in DCM	1 x 15
	equal vol DMF	
	(3) Symmetrical coupling (except for Gly, Gln, Asn and Arg)	
	4 equiv. Boc A.A. in DCM + 2 equiv. DCC in DCM, filter	1 x 45
Washing	DCM	2 x 1
	DMF	1 x 1
	DCM	2 x 1
Ninhydrin test*	A small amount of dried resin + 3 drops of reagent A + 1 drop of reagent B	100°C x 10min

Appendix C. Peptide synthesis protocols

Table C.2. Fmoc chemistry

Step	Reagent	Repeat x Time (min)
Deprotection	20 % piperidine/DMF	1 x 2
	20 % piperidine/DMF	1 x 20
Washing	DMF	3 x 1
	DCM	1 x 1
	isopropanol	2 x 1
	DMF	3 x 1
Coupling	3 equiv. Fmoc A.A. in DMF	1 x 1
	3 equiv. HOBt in DMF	1 x 1
	DCC in DCM	1 x 45
Washing	DMF	2 x 1
	isopropanol	1 x 1
	DMF	1 x 1
	DCM	1 x 1
Ninhydrin test*	A small amount of dried resin + 3 drops of reagent A + 1 drop of reagent B	100°C x 10min

HOBt and DCC are usually prepared in a 1M (1mM/ml) solution and stored avoid light.

*Ninhydrin test reagent preparation (Sarin et al., 1981)

Reagent A: mix solutions 1 and 2. It should be colorless and stored in the dark under nitrogen.

Solution 1 - 40 g phenol in 10 ml of absolute ethanol ---> add 4 g Amberlite mixed-bed resin (MB-3) ---> stir 45 min ---> filter.

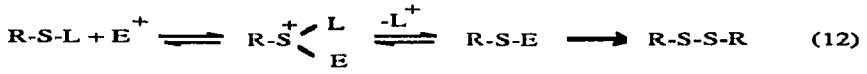
Solution 2 - KCN solution: 65 mg KCN ---> add 100 ml H2O.
Above KCN solution 2 ml ---> add pyridine to 100 ml ---> add 4g MB-3 resin ---> stir 45 min ---> filter.

Reagent B: 2.5 g ninhydrin in 50 ml absolute ethanol. It is yellowish and should be stored in the dark under nitrogen.

Appendix D. Iodine removal of Acm, a thiol protecting group for cysteine, induces direct conversion to disulfides

[Hiskey, R.G. (1981) in *The Peptides. Analysis, synthesis, Bology.* (E. Gross and J. Meienhofer eds.) Vol 3. Protection of funtional groups in peptide synthesis, pp 145.]

The conversion of a thioether of cysteine to an intermediate capable of direct conversion to cystine eliminates isolation of the thiol. More importantly, however, this strategy permits the stepwise formation of two or more disulfide bonds without the possibility of thiol-disulfide interchange. The essence of this approach is to utilize a sulfur protective group that will provide a highly stablized cation (L^+) upon interaction of the sulfur atom with an electrophile [eq. (12)]. The intermediate sulfenyl compound may then function as the electrophile and react with another molecule of substract or another added thioether or thiol to provide the disulfide.

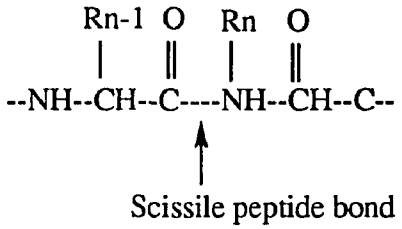


L^+: a highly stabilized cation such as Acm; E: electrophile such as I_2

Table 1. "Half-Times" (t_H) for Iodine Oxidations of R-S-Trt- and R-S-Acm

Solvent	t_{HT} (R-S-Trt)	t_{HA} (R-S-Acm)	t_{HA}/t_{HT}
MeOH	4 sec	1 min	15
MeOH-HOH (8:2)	< 1 sec	5 sec	10
AcOH	75 sec	45 min	35
AcOH-HOH (8:2)	2 sec	50 sec	25
Nitromethane	1 sec	5 sec	5
1,2-Dichloroethane	1.5 min	200 min	130
CCl_4	2.5 min	300 min	120
Dioxane	1 min	100 min	100
$CHCl_3$	1 sec	90 min	5,400
$CHCl_2$	1 sec	60 min	3,600
TFE-CH_2Cl_2 (3:1)	7 sec	>4 hr	>2,500
HFIP-CH_2Cl_2 (3:1)	1 sec	>> 3 hr	>10,000
MeOH-$CHCl_3$ (1:1)	1 sec	15 min	900

Appendix E. Specificity of Various Endopepteases



Enzyme	Specificity
S. aureus protease V8	R_{n-1} = Glu
Trypsin	R_{n-1} = positive charged residues: Arg, Lys; $R_n \neq$ Pro
Endoprotease Lys-C	R_{n-1} = Lys
Elastase	R_{n-1} = small neutral residues: Ala, Gly, Leu, Ser, Val; $R_n \neq$ Pro
Thermolysin	R_n=Ile, Leu, Met, Phe, Trp, Tyr, Val; $R_{n-1} \neq$ Pro

Appendix F. Calculated mass for peptide (fIX_{EGF-C}) digestion fragments from a MASS program, written by Detlev Suckau

Table F.1. Molecular weight* for peptide A fragments generated from protease V8 digestion after first-step oxidation

Sequence	$[M+H]^+$
1-13	1463.7
1-30	3390.8
1-45	5108.7
14-30	1946.2
14-36	2635.9
14-42	3396.7
14-45	3664.1
31-36	708.8
31-42	1469.6
31-45	1736.9
37-42	779.8
37-45	1047.2

* Masses are calculated assuming oxidized cysteines, and Cys^{28} and Cys^{41} with Acm.

Table F.2. Molecular weight* for peptide A fragments generated from protease V8 digestion after second-folding oxidation

equence	$[M+H]^+$
1-13	1463.7
1-30	3319.8
1-45	4964.6
14-30	1875.1
14-36	2564.9
14-42	3252.6
14-45	3520.0
31-36	708.8
31-42	1398.5
31-45	1665.9
37-42	708.8
37-45	976.1

* Masses are calculated assuming oxidized cysteines.

Table F.3. Molecular weight* for peptide B fragments generated from tryptic digestion after first-step oxidation

Sequence	$[M+H]^+$	Sequence	$[M+H]^+$
1-8	906.1	12-23	1458.6
1-11	1233.4	12-33	2625.9
1-17	2041.4	12-39	3309.7
1-23	2671.0	12-45	3896.3
1-33	3840.3	18-23	648.7
1-39	4524.1	18-33	1818.0
1-45	5108.7	18-39	2501.7
9-11	346.4	18-45	3086.4
9-17	1156.3	24-33	1188.3
9-23	1785.9	24-39	1872.1
9-33	2953.2	24-45	2456.8
9-39	3637.0	34-39	702.8
9-45	4223.7	34-45	1289.4
12-17	829.0	40-45	605.7

* Masses are calculated assuming oxidized cysteines, and Cys^{12} and Cys^{26} with Acm.

Table F.4. Molecular weight* for peptide B fragments generated from protease V8 digestion after first-step oxidation

Sequence	$[M+H]^+$
1-13	1536.7
1-30	3465.9
1-45	5108.7
14-30	1948.2
14-36	3637.9
14-42	3327.7
14-45	3595.0
31-36	708.8
31-42	1398.5
31-45	1665.9
37-42	708.8
37-45	976.1

* Masses are calculated assuming oxidized cysteines, and Cys^{12} and Cys^{26} with Acm.

Table F.5. Molecular weight* for peptide B fragments generated from protease Lys-C digestion after first-step oxidation

Sequence	$[M+H]^+$
1-8	906.1
1-17	2041.4
1-23	2671.0
1-39	4524.1
1-45	5108.7
9-17	1156.3
9-23	1785.9
9-39	3637.0
9-45	4223.7
18-23	648.7
18-39	2501.7
18-45	3086.4
24-39	1872.1
24-45	2456.8
40-45	605.7

* Masses are calculated assuming oxidized cysteines, and Cys^{12} and Cys^{26} with Acm.

Table F.6. Molecular weight* for peptide C fragments generated from tryptic digestion after first-step oxidation

Sequence	$[M+H]^+$	Sequence	$[M+H]^+$
1-8	977.1	12-23	1458.6
1-11	1304.5	12-33	2554.8
1-17	2114.4	12-39	3238.6
1-23	2744.1	12-45	3823.3
1-33	3840.2	18-23	648.7
1-39	4524.1	18-33	1744.9
1-45	5108.7	18-39	2428.7
9-11	346.4	18-45	3013.3
9-17	1156.3	24-33	1115.3
9-23	1785.9	24-39	1799.0
9-33	2884.2	24-45	2383.7
9-39	3565.9	34-39	702.8
9-45	4150.6	34-45	1289.4
12-17	829.0	40-45	605.7

* Masses are calculated assuming oxidized cysteines, and Cys^{5} and Cys^{16} with Acm.

REFERENCES

Adler A. J., Fasman G.D. and Blout E.R. (1963). *J Am Chem Soc* **85**, 90-97.

Akaji K., Fujino K., Tatsumi T., Kiso Y. (1992). *Tetrahedron Lett* **33**, 1073-1076.

Appella E., Weber I. and Blasi F. (1988). *FEBS Lett* **231**, 1-4.

Applied Biosystems (**1990**). *Introduction to cleavage techniques.* Strategies in peptide synthesis. Applied Biosystems, Inc.: Foster City, CA. pp.6-14.

Astermark J., Björk I., Öhlin AK. and Stenflo J. (1991a). *J Biol Chem* **266**, 2430-2437.

Astermark J. and Stenflo J. (1991b). *J Biol Chem* **266**, 2438-2443.

Baba T., Sugiyama H. and Seto S. (1973). *Chem Pharm Bull Japan* **21**, 207-209.

Bairoch A. (1993). *Nucleic Acids Res* **21**, 3097-3103.

Barany, G. and Merrifield, R.B. (1979a). Solid-phase peptide synthesis. In *The Peptides,* Vol **2**; (Gross E. and Meienhofer J. Eds.), Academic Press: New York. pp.193-196.

Barany, G. and Merrifield, R.B. (1979b). Solid-phase peptide synthesis. In *The Peptides,* Vol **2**; (Gross E. and Meienhofer J. Eds.), Academic Press: New York. pp.233-243.

Baron M., Norman D.G., Harvey T.S., Handford P.A., Mayhew M., Tse A.G.D., Brownlee G.G. and Campbell I.D. (1992). *Protein Science* **1**, 81-90.

Beavis R.C. and Chait B.T. (1989). *Rapid Commun Mass Spectrom* **3**, 233-237.

Beavis R.C. and Chait B.T. (1990). *Anal Chem* **62**, 1836-1840.

Benham C.J. and Jafri M.S. (1993). *Protein Science* **2**, 41-54.

Beychok S. (1966). *Science* **154**, 1288-1299.

Blake J. (1979). *Int J Peptide Protein Res* **13**, 418-425.

Bodanszky M., Tolle J.C., Deshmane S.S. and Bodanszky A. (1978a). *Int J Peptide Protein Res* **12**, 57-68.

Bodanszky M. and Kwei J.Z. (1978b). *Int J Peptide Protein Res* **12**, 69-74.

Bodanszky M and Martinez J. (1978c). *J Org Chem* **43**, 3071-3073.

Campbell I.D., Cook R.M., Baron M., Harvey T.S. and Tappin M.J. (1989). *Progress in Growth Res***1**, 13-22.

Capaldi R.A. and Vanderkooi G. (1972). *Biochem* **11**, 4120-4131.

Carpino L.A. and Han G.Y. (1972). *J Org Chem* **37**, 3404-3409.

Carter J.M. and Lyon J.A. (**1990**). In *Peptides: Chemistry, Structure and Biology*, Proceedings of the Eleventh American Peptide Symposium (Rivier J.E. and Marshall G.R. Eds.), ESCOM: Leiden. pp.721-723.

Chait B.T. and Kent S.B.H. (1992). *Science* **257**, 1885-1894.

Chait B.T., Wang R., Beavis R.C. and Kent S.B.H. (1993). *Science* **262**, 89-92.

Chowdhury S.K., Katta V. and Chait B.T. (1990). *Rapid Commun Mass Spectrom* **4**, 81-87.

Chou P.Y. and Fasman G.D. (1974). *Biochem* **13**, 222-245.

Cohen S. and Taylor J.M. (**1974**). In *Thirtieth Symposium of the Society of Developmental Biology*. (Hay E.D., King T.J. and Papaconstantinou J. Eds.), Academic Press: New York. pp.25-42.

Cooke R.M., Wilkinson A.J., Baron M., Pastore A., Tappin M.J., Compbell I.D., Gregory H. and Sheard B. (1987). *Nature* **327**, 339-341.

Das M.K. and Lindstrom J. (1991). *Biochemistry* **30**, 2470-2477.

Dolittle R.F. (1979). Protein evolution. In *The proteins*, Vol **4**. (Neurath H. and Hill R.L., Eds.), Academic Press: New York. pp.1-118.

Doolittle R.F., Feng D.F. and Johnson M.S. (1984). *Nature* 307, 558-560.

Durieux J-P and Nyfeler R. (**1995**). Selective bridging. An unambiguous synthesis of peptides having three disulfide bridges. *Peptides 1994*. Proceedings of the Twenty-Third European Peptide Symposium. ESCOM: Leiden. In press.

Ellis R.J. (1994). *Curr Opin Struct Biol* **4**, 117-122.

Erickson B.W. and Merrifield R.B. (1976). Solid-phase peptide synthesis. In *The proteins*, Vol **II**. (Neurath H. and Hill R.L., Eds.), Academic Press: New York. pp.255-527.

Furie B. and Furie B.C. (1988). *Cell* **53**, 505-518.

George-Nascimento C., Lowenson J., Borissenko M., Calderon M., Medina-Selby A., Kuo J., Clarke S. and Randolph A. (1990). *Biochem* **29**, 9584-9591.

Gill S.C. and Hippel P.H. (1989). *Anal Biochem* **182**, 319-326.

Gray W.R., Luque F.A., Galyean R., Atherton E., Sheppard R.C., Stone B.L., Reyes A., Alford J., McIntosh M., Olivera B.M., Cruz L.J. and Rivier J. (1984). *Biochem* **23**, 2796-2802.

Gregory H. (1975). *Nature* **257**, 325-327.

Greenfield N. and Fasman G.D. (1969). *Biochem* **8**, 4108-4116.

Handford P.A., Baron M., Mayhew M., Willis A., Beesley T., Brownlee G.G. and Campbell I.D. (1990). *EMBO J.* **9**, 475-480.

Handford P.A., Mayhew M., Baron M., Winship P.R., Campbell I.D. and Brownlee G.G. (1991). *Nature* **351**, 164-167.

Heath W.F. and Merrifield R.B. (1986). *Proc Natl Acad Sci, USA* . **83**, 6367-6371.

Hillenkamp F., Karas M., Beavis R.C. and Chait B.T. (1991). *Anal Chem* **63**, 1193A-1203A.

Hiskey R.G. (1981). Sulfhydryl group protection in peptide synthesis. *The peptides: Analysis, Synthesis, Biology*, Vol 3. Protection of functional groups in peptide synthesis. (Gross E., Meienhofer J. Eds.), Academic Press: New York, 137-167.

HØjrup P. and Magnusson S. (1987). *Biochem J.* **245**, 887-892.

Holladay L.A., Savage C.R.Jr., Cohen S. and Puett D. (1976). *Biochem* **15**, 2624-2633.

Holzwarth G. and Doty P. (1965). *J Am Chem Soc* **87**, 218-228.

Hommel U., Dudgeon T.J., Fallon A., Edwards R.M. and Campbell I.D. (1991). *Biochem* **30**, 8891-8898.

Huang L.H., Ke X.H., Sweeney W. and Tam J.P. (1989). *Biochem Biophys Res Commun* **160**, 133-139.

Huang L.H., Cheng H., Pardi A., Tam J.P. and Sweeney W.V. (1991). *Biochem* **30**, 7402-7409.

Hunter M.J., Plesniak L.A. and Komives E.A. (1993). *Protein Sci* **2** (Suppl 1), 128.

Kahn P.C. (1972). Ph.D. Dissertation, Columbia University.

Kohda D. and Inagaki F, (1988). *J Biochem* **103**, 554-571.

Kohda D. and Inagaki F. (1992). *Biochem* **31**, 11928-11939.

Kumagaye S-I., Kuroda H., Nakajima K., Watanabe T.X., Kimura T., Masaki T. and Sakakibara S. (1988). *Int J Pept Protein Res* **32**, 519-526.

Lee T.D. and Shively J.E. (1990). Enzymatic and Chemical digestion of proteins of mass spectrometry in Mass Spectrometry, in *Methods in Enzymology,* Vol **193**, Mass Spectrometry (McCloskey J.A., Ed.), Academic Press: San Diego. pp.361-374; 384.

Leytus S.P., Foster D.C., Kurachi K. and Davie E.W. (1986). *Biochem* **25**, 5098-5102.

Makino K., Morimoto M., Nishi M., Sakamoto S., Tamura A., Inooka H. and Akasaka K. (1987). *Proc Natl Acad Sci, USA*. **84**, 7841-7845.

Mao B. (1989). *J Am Chem Soc* **111**, 6132-6136.

Mark J.L. (1986). *Science*, **234**, 543-544.

Marshall G.R. and Merrifield R.B. (1965). *Biochem.* **4**, 2394-2401.

Martinet J. and Bodanszky M. (1978). *Int J Peptide Protein Res*, **12**, 277-283.

Merrifield R.B. (1963). *J Am Chem Soc* **85**, 2149-2154.

Merrifield B. (1986). *Science* **232**, 341-347.

Merrifield R.B. (**1993**). *Life During a Golden Age of Peptide Chemistry.* (Seeman J.I., Ed.), American Chemical Society: Washington, DC. pp.182.

Narhi L.O., Arakawa T., Mcginley M.D., Rohde M.F. and Westcott K.R. (1992). *Int J Peptide Protein Res* **39**, 182-187.

Nicolás E., Pedroso E. and Giralt E. (1989). *Tetrahedron Lett* **30**, 497-500.

Nishiuchi Y. and Sakakibara S. (1982). *FEBS Lett* **148**, 260-262.

Okada Y., Iguchi S. and Kawasaki K. (1987). *J Chem Soc, Chem Commun*, 1532-1534.

Okada Y. and Iguchi S. (1988). *J Chem Soc, Perkin Trans I*, 2129-2136.

Ondetti M.A., Deer A., Sheehan J.T., Pluscec J. and Kocy O. (1968). *Biochem* **7**, 4069-4075.

Patthy L. Trexler M., Váli Z., Bányai L. and Váradi A. (1984). *FEBS Lett* **171**, 131-136.

Patthy L. (1985). *Cell* **41**, 657-663.

Perczel A., Hollósi M., Foxman B.M. and Fasman G.D. (1991). *J Am Chem Soc* **113**, 9772-9784.

Prestrelski S.J., Arakawa T., Wu C-S C., O'Neal K.D., Westcott K.R. and Narhi L.O. (1992). *J Biol Chem* **267**, 319-322.

Prevelige P. and Fasman G. D. (**1989**). Chou-Fasman prediction of the secondary structure of proteins. In *Prediction of protein structure and the principles of protein conformation* (Fasman G. D, Ed.), Plenum Press: New York and London. pp.391-416.

Rees D.J.G., Rizza C.R. and Brownlee G.G. (1985). *Nature* **316**, 643-645.

Rees D.J.G., Jones I.M., Handford P.A., Walter S.J., Esnouf M.P., Smith K.J. and Brownlee G.G. (1988). *EMBO J.* **7**, 2053-2061.

Sarin V.K., Kent S.B.H., Tam J.P. and Merrifield R.B. (1981). *Anal Biochem* **117**, 147-157.

Savage C.R.Jr. and Cohen S. (1972a). *J Biol Chem* **247**, 7609-7611.

Savage C.R. Jr., Inagami T. and Cohen S. (1972b). *J Biol Chem* **247**, 7612-7621.

Savage C.R.Jr., Hash J.H. and Cohen S. (1973). *J Biol Chem* **248**, 7669-7672.

Selander M., Persson E., Stenflo J. and Drakenberg T. (1990). *Biochem* **29**, 8111-8118.

Sheridan R.P., Dixon J.S., Venkataraghavan R., Kuntz I.D. and Scott K.P. (1985). *Biopolymer* **24**, 1995-2023.

Spetzler J.C., Rao C. and Tam J.P. (1994). *Int J Pept Protein Res* **43**, 351-358.

Stewart J.M. and Young J.D. (**1984**). *Solid phase peptide synthesis*. Second Edition. Pierce Chemical Company: Rockford.

Suzuki K. and Endo N. (1978).*Chem Pharm Bull Japan* **26**, 2269-2274.

Tam J.P., Heath W.F. and Merrifield R.B. (1983). *J Am Chem Soc* **105**, 6442-6455.

Tam J.P., Rieman M.W. and Merrifield R.B. (1988). *Peptide Res* **1**, 6-18.

Tam J.P., Liu W., Zhang J.W., Galantino M. and De Castiglione R. (**1990**). In *Peptides 1990*. (Giralt E. and Andreu D., Eds.), ESCOM: Leiden. pp.160-163.

Tam J.P., Wu C.R., Liu W. and Zhang J.W. (1991). *J Am Chem Soc* **113**, 6657-6662.

Taylor J.M., Mitchell W.M. and Cohen S. (1972). *J Biol Chem* **247**, 5928-5934

Ullner M., Selander M., Persson E., Stenflo J., Drakenberg T. and Teleman O. (1992). *Biochem* **31**, 5974-5983.

Van Rietschoten J., Pedroso Muller E. and Granier C. (**1977**). In *Peptides: Chemistry, Structure and Biology*. Proceedings of the Fifth American Peptide Symposium. (Goodman M. and Meienhofer J., Eds.), John Wiley and Sons: New York. pp.522-524.

Veber D.F., Milkowski J.D., Varga S.L., Denkewalter R.G. and Hirschmann R. (1972). *J Am Chem Soc* **94**, 5456-5461.

Violand B.N., Tou J.S., Vineyard B.D., Siegel N.R., Smith C.E., Pyla P.D., Zobel J.F., Toren P.C. and Kolodziej E.W. (1991). *Int J Pept Protein Res* **37**, 463-467.

Walker S.M., Wang R., Milton S., Chait B. and Kent S. (**1993**). In *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics*, San Francisco, CA, May 30-June 4, 1993. pp.380a-380b.

Wang S.S., Yang C.C., Kulesha I.D., Sonenberg M. and Merrifield R.B. (1974). *Int J Peptide Protein Res* **6**, 103-109.

Warne N.W. and Laskowski M. Jr. (1990). *Biochem Biophys Res Commun* **172**, 1364-1370.

Yang C.C. and Merrifield R.B. (1976). *J Org Chem* **41**, 1032-1041.

Yang J.T., Wu C-S.C. and Martinez H.M. (1986). Calculation of protein conformation from circular dichroism. In *Methods in Enzymology*. Vol.**130**. (Hirs C.H.W. and Timasheff S.N., Eds.), Academic Press: New York. pp.208-269.

Yang Y., Sweeney W.V. and Tam J.P. (**1994**a). *Peptides: Chemistry, Structure and Biology*. (Hodges R.S. and Smith J.A. Eds.), ESCOM: Leiden. pp. 83-84.

Yang Y., Sweeney W.V., Schneider K., Chait B.T. and Tam J.P. (1994b). *Protein Science* **3**, 1267-1275.

Yang Y., Sweeney W.V., Schneider K., Chait B.T. and Tam J.P. (**1995**a). Selective formation of three disulfide bridges in EGF-like peptides. In *Peptides: Biology and Chemistry*. Proceedings of the 1994 Chinese peptide symposium. (Lu G-S. and Tam J. P., Eds.), ESCOM: Leiden. In press .

Yang Y., Sweeney W.V., Thörnqvist S., Schneider K., Chait B.T. and Tam J.P. (**1995**b). Characterization of a side reaction using stepwise detection in peptide synthesis with Fmoc chemistry. In *Techniques in Protein Chemistry* (Crabb J. Ed.), Academic press: San Diego. In press.

Yang Y., Sweeney W.V., Schneider K., Thörnqvist S., Chait B.T. and Tam J.P. (1995c). Aspartimide formation in base-driven 9-Fluorenylmethoxycarbonyl chemistry. *Tetrahedron Lett*, In press.

Yoshitake S., Schach B.G., Foster D.C., Davie E.W. and Kurachi K. (1985). *Biochem* **24**, 3736-3750.

Zhang R. and Snyder G.H. (1989). *J Biol Chem* **264**, 18472-18479.

Zhang R. and Snyder G.H. (1991). *Biochem* **30**, 11343-11348.