Order Number 9315468

Structure and function of the first epidermal growth factor-like module in human factor IX

Huang, Linda Hao Tsai, Ph.D.

City University of New York, 1993

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106

Structure and Function of the First Epidermal Growth Factor-like Module in Human Factor IX

by

Linda Hao Tsai Huang

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

1993

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

January 15, 1993
Date

William Stevens
Chair of Examine Committee

January 25, 1993
Date

Horst Schulz
Executive Officer

Thomas C. Strekas

Dixie Goss

James P. Tam

Michael Blumenstein
Supervisory Committee

The City University of New York

Abstract

Structure and Function of the First Epidermal Growth Factor-like Module in Human Factor IX

By
Linda Hao Tsai Huang

Adviser: Professor William V. Sweeney

Epidermal growth factor (EGF) is a small mitogen which consists of 53 amino acids. During the last decade, a large number of EGF-like molecules have been found. There are at least two groups of EGF-like modules. Both classes of modules have three pairs of disulfide bonds in homologous arrangement in a pattern of cysteine residues 1-3, 2-4, and 5-6. The major difference between these two groups of modules is that the modules in the first group have high affinity for the EGF-receptor, and can lead to mitogenic responses in EGF-sensitive cells; the second group of modules are found in a wide range of proteins with diverse biological function. Among these proteins are vitamin K-dependent clotting factors VII, IX, and X, protein C and protein S, and the non-vitamin K-dependent proteins urokinase (uPA), tissue-type plasminogen activator (tPA), and factor XII. The functions of these modules are not fully understood.

To investigate the biological role of the EGF-like modules in the second group mentioned above, the first EGF-like module (45-87) of human factor IX (FIX-EGF1) has been chemically synthesized by the solid phase method. The synthesized peptide has been refolded, and characterized.

Efforts have been made to mimic putative EGF-like activities by synthesizing different chimeric EGF analogs with the first EGF-like module in factor IX (FIX-EGF1) as a non-active frame. However, no analog with significant EGF-like activity has been obtained.

Calcium plays an important role as a cofactor in the blood clotting system. The calcium binding site in the EGF-like region in factor IX has been reported to be of functional importance. The calcium binding with the synthesized FIX-EGF1 has been examined, leading to the identification a calcium binding site in this module. The function of the metal ion binding is more likely to orient, and to stabilize the factor IX molecule as a whole rather than to change the structure of the module itself.

The two dimensional ^1H nuclear magnetic resonance spectra (2D ^1H NMR) of FIX-EGF1 have been acquired. The sequence specific resonance assignment and the secondary structure of the module have been determined. It has been found that the peptide has two antiparallel β-sheets. The first sheet consists of residues 16-20 and 25-29, the second sheet consists of 32-34 and 40-42. A β-turn has been observed at one end of the each sheet. The tertiary structure has been determined by distance geometry calculation with the NOE constraints using DSPACE, followed by energy minimization refinement with the program AMBER. The solution structure indicates that there are two subdomains in the peptide consisting of 1-28 and 29-43. The structure is substantially similar to that of EGF, but differences in the degree of the NH_2-terminal β-sheet twist, and the C-terminal tail orientation are observed.

Acknowledgment

At the end of my study stage, I looked back the life-road which I have passed, I feel that many people deserved acknowledgment for their guidance, their support and their encouragement. However, there are a number of people who merit special appreciation.

To Prof. Sweeney, my supervisor and my mentor, for his patience, his kind support, encouragement, and generously and intensely sharing his knowledge in physical biochemistry with me.

To Prof. James Tam for inducing me into the peptide chemistry, for giving me the chance to study in the interesting area of the biochemistry in the excellent working facilities. Without his kind support, I would not have today's thesis done.

To Prof. Bruce R. Merrifield, for his kindness, his generous support and for his example to be a good scientist.

To Prof. Arthur Pardi, for his valuable guidance and excellent techniques in 2D NMR. To Jack Skalicky, for his cooperation during the solution structure calculation.

To Dr. Michael Blumenstein, for his guidance and valuable discussion during my study.

To Prof. D. Goss, for her many helps and valuable commands on the thesis.

To all members in Prof. Merrifield's laboratory of The Rockefeller University, past and present, for their many helps over the years.

To all members in Chemistry Department of Hunter college who guiding and supporting me during my study.

I also wish to express my deep gratitude to all my friends and my family for their love and their encouragement.

Especially, I would like to thank my husband Wolin for his understanding, his love and his support; to thank my parents not only for bringing us up, and also for their encouragement, their tutelage and their selfless support; to thank my mother in-law for her kindness help; to thank my daughters Alina and Grace, for their love and their admirable endurance.

TABLE OF CONTENTS

List of figures x

List of tables xiv

List of abbreviations xiv

Chapter One 1

1. Background 1

1.1. Blood coagulation 1

1.1.1. Initiation of coagulation 1

1.1.2. Regulation of coagulation 7

1.2. Human factor IX 8

1.2.1. Gla region 9

1.2.2. EGF-like region 10

1.2.3. Activation peptide region 10

1.2.4. Catalytic region 11

1.3. Structural feature of the first EGF-like module 11

1.4. Epidermal growth factor 15

1.5. Calcium in the clotting system 19

1.6. 2D-NMR in determination of protein solution structure 22

Chapter Two 30

2. Structure and function relationship study of the first EGF-like module of human factor IX 30

2.1. Experimental procedures 44

2.2. Results 52

2.3. Discussion 69

Chapter Three 71

3. Solution structure determination by ^{1}H NMR spectroscopy and refined by energy minimization with restraints 71

3.1. Sequence-specific assignment and secondary structure determination 71

3.1.1. Materials and Methods 71

3.1.2. Results 75

3.2. Solution structure determination and refinement 112

3.2.1. Materials and Methods 113

3.2.2. Results 115

Appendix

Distance constraints for solution structure determination of FIX-EGF1. 146

References 163

LIST OF FIGURES

Figure 1.1. Schematic representation of blood coagulation cascade 6

Figure 1.2. Schematic diagram of activated factor IX structural domains. 9

Figure 1.3. Schematic diagram of genetic defects on FIX-EGF1. 14

Figure 1.4. Schematic diagram of EGF precursor. 17

Figure 1.5. Sequence comparison by aligning cysteine residues of hEGF, hTGFα, human factor IX first EGF-like module and bovine factor X first EGF-like module. 19

Figure 1.6. Schematic illustration of the sequential assignment procedure 25

Figure 1.7. NMR parameters used to identify secondary structure in proteins 26

Figure 1.8. A schematic diagram of a peptide fragment with dihedral angles 28

Figure 2.1. Schematic diagram of the structures of FIX-EGF1 , its 15 analogs, and hEGF 35

Figure 2.2. Schematic illustration of synthetic chimeric analogues. 43

Figure 2.3. The dicyclohexylcarbodiimide mediated amino acid coupling reaction 47

Figure 2.4. C_{18} reverse-phase analytic HPLC profiles of FIX-EGF1 after HF and dialysis. 54

Figure 2.5. C_{18} reverse-phase HPLC profiles of FIX-EGF1 and analog 10. Comparing the refolding methods. 57

Figure 2.6. HPLC analysis on a C_{18} reverse-phase column of crude and purified FIX-EGF1 and its analogs. 59

Figure 2.7. HPLC profiles of all purified synthetic products 61

Figure 2.8. EGF-receptor binding inhibition assay results of FIX-EGF1 and its analog with A431 cell. 68

Figure 2.9. 1D ^1H NMR spectra of analog 14 and analog 15. 71

Figure 3.1. 1D ^1H NMR spectra of FIX-EGF1 at different pH values. 77

Figure 3.2. A phase-sensitive contour plot of the aliphatic proton region of the DQF-COSY spectrum (25 °C, pH 4.2). 79

Figure 3.3. A contour plot portion of the relayed-COSY spectrum in H_2O (25 °C, pH 4.2). 81

Figure 3.4. A contour plot portion of the TOCSY spectrum in H_2O (25 °C, pH 4.2). 82

Figure 3.5. A contour plot portion of the TOCSY spectrum (25 °C, pD 4.2) in D_2O. 83

Figure 3.6. A contour plot portion of the TQ-COSY spectrum (25 °C, pH 4.2). 85

Figure 3.7. The ring proton assignments by using both the NOESY and the TOCSY spectra. 88

Figure 3.8. Two sections of the H_2O NOESY contour plot (25 °C, pH 4.2) showing the connectivities between γ-NH and CβH for Asn^{37}. 90

Figure 3.9. The NOESY (25 °C, pD 4.2, in D_2O) contour plot showing the intraresidue NOE connections for Pro^{11} and Pro^{30} 93

Figure 3.10. The contour plot of the "fingerprint" region of the H_2O DQF-COSY (25 °C, pH 4.2), showing the amide/α-proton cross-peaks. 94

Figure 3.11. A contour plot section of the NOESY spectrum in H_2O (25 °C, pH 4.2), showing the sequential connections for Leu^{13} to Ser^{17} and Pro^{30} to Gly^{32} paths through $C_\alpha H(i)$-$NH_{(i+1)}$ cross-peaks. 96

Figure 3.12. A contour plot section of the NOESY spectrum in H_2O (25 °C, pH 4.2), showing the sequential connections through $NH_{(i)}$-$NH_{(i+1)}$ cross-peaks for Gly^4 to Asn^{10}. 97

Figure 3.13. Proline configuration in the peptide. 98

Figure 3.14. Summary of the sequential connectivities and antiparallel β-sheet segments found for FIX-EGF1. 100

Figure 3.15. A portion of the NOESY (25 °C, pD 4.2) contour plot showing the cross-strand $d_{\alpha\alpha}$ cross-peaks arising from the antiparallel β-sheet structure. 104

Figure 3.16. Schematic diagram of the antiparallel β-sheets in FIX-EGF1. 105

Figure 3.17. Calcium effect on the proton chemical shifts in the DQF-COSY spectrum at pH 4.7. 110

Figure 3.18. A portion of the contour plot of NOESY spectra at various mixing times 117

Figure 3.19. Plot of cross-peak volume versus mixing time for the $Asn^{37}\beta H$-βH cross-peak in the D_2O NOESY spectra 118

Figure 3.20. A portion of the PE-COSY contour plot. 123

Figure 3.21. The distribution of the distance constraints in the peptide 128

Figure 3.22. Superimposition of backbone atoms of 10 conformers fragments 131

Figure 3.23. The representative ribbon drawing for the structure of FIX-EGF1 with the indication of residues 12-14, and 35-39 134

Figure 3.24. The representative ribbon drawing for the structure of FIX-EGF1 with the indication of the fragment of 34-36 136

Figure 3.25. Ramachandran plot for residue 2-42 in 10 calculated AMBER structures 138

Figure 3.26. The representative ribbon drawing of the FIX-EGF1. The side chains of Gly^4-Gln^6, Asp^{20} and Asp^{21} are shown 140

Figure 3.27. The comparison of ^1H chemical shift of NH for FIX-EGF1, and mEGF , and for FIX-EGF1 and the EGF-like module of factor X 143

LIST OF TABLES

Table 2.1. Solid Phase Synthesis Protocol (for Boc-amino acid). 45

Table 2.2. Molecular weight determination of FIX-EGF1, 2 and analogous by Cf-252 fission ionization mass spectrometry. 63

Table 2.3. Disulfide bond location of synthetic products by enzymatic digestion method 65

Table 2.4. EGF-receptor binding competitive activity of synthetic FIX-EGF1 and analogous. 67

Table 3. 1. Proton chemical shifts for FIX-EGF1 obtained from the aliphatic region of the TQ-COSY (25 °C, pH 4.2) contour plot. 86

Table 3.2. ^{1}H NMR chemical shifts of FIX-EGF1 (pH 4.2). 101

Table 3.3. $^{3}J_{HN\alpha}$ coupling constants (≥ 8.0 Hz) determined from the DQF-COSY spectrum in H_2O. 107

Table 3.4. The criteria for classification of type I turn and the experiment data of the turn in the first β-sheet between amino acid residues 21 to 24. 108

Table 3.5. Chemical shifts for residues with protons which change resonance position by more than 0.02 ppm after addition of calcium (at pH 4.7). 111

Table 3.6. Some geminal pair cross-Peak volumes from NOESY spectra in water. 119

Table 3.7. Some geminal pair distance calculation data from NOESY spectra in water. 121

Table 3.8. The criteria of the stereospecific assignment and $\chi 1$ determination. 124

Table 3.9. Stereospecific $C^{\beta}H_2$ assignments of FIX-EGF1 125

Table 3.10. Numbers of experimental distance constraints used in determining the solution structure of FIX-EGF1. 129

Table 3.11. Summary of conformational energies for the final 10 FIX-EGF1 conformers 132

Table 3.12. The observed NOEs between the two subdomains 135

Table 3.13. Average RMSDs of the 10 FIX-EGF1 conformers generated by AMBER from one of the conformer 137

List of Abbreviations

1. common amino acid

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
Hya		β-hydroxyaspartic acid
Ile	I	isoleucine
Leu	L	leucine
Lys	K	lysine
Phe	F	phenylalanine
Pro	P	proline
Ser	S	serine
Thr	T	threonine
Trp	W	tryptophan
Tyr	Y	tyrosine
Val	V	valine

2. Other Abbreviations

Boc	tert-butyloxycarbonyl
BrZ	2-bromobenzyloxycarbonyl
Bzl	benzyl
CH_3CN	acetonitrile

ClZ-	2-chlorobenzyloxycarbonyl
COSY	homonuclear-correlated spectroscopy
DCC	dicyclohexylcarbodiimide
DCU	dicyclohexylurea
DIEA	diisopropylethyl amine
DMF	dimethylformamide
DMS	dimethyl sulfide
DNA	deoxyribonucleic acid
Dnp	2,4-dinitrophenyl
DQF-COSY	double quantum filtered COSY
DTT	dithiothreitol
EDT	ethanedithiol
For	formyl
hEGF	human epidermal growth factor
HOBt	1-hydroxybenzotriazole
HPLC	high performance liquid chromatography
mEGF	murine epidermal growth factor
4-MeBzl	4-methylbenzyl
NOESY	nuclear Overhauser enhancement
spectroscopy	
OBzl	benzyl ester
Pam	phenylacetamidomethyl
PE-COSY	primitive elegant correlation spectroscopy
TFA	trifluoroacetic acid
TGFα	transforming growth factor-α
TOCSY	total correlated spectroscopy
Tos	4-toluenesulfonyl (tosyl)

1. Background

The theme of this thesis is structure and function studies of the first epidermal growth factor-like module in human factor IX (FIX-EGF1). Chemically synthesized FIX-EGF1(45-87) and other analogs have been examined for their putative EGF-like activities. The calcium binding property of FIX-EGF1 has been analyzed, and its solution conformation has been determined by two dimensional nuclear magnetic resonance (2D-NMR).

For the purpose of giving a general background to this topic, the initiation and regulation of the blood coagulation, the structural features and the function of factor IX are briefly reviewed. A concise survey of recent studies on EGF and its family is given, and the significance of 2D-NMR in protein solution structure studies, and of calcium in the clotting system are also mentioned.

1.1. Blood coagulation

Blood coagulation is one of the host defense systems that specifically assists in keeping the mammalian circulatory system in a perfectly closed and constantly pressurized condition after blood vessel injury.

1.1.1. Initiation of coagulation

Normally, the endothelial cells that lie on the vascular wall function to inhibit the adherence and activation of platelets and to inhibit the clotting cascade. Thus, there is no clot formation in normal blood flow. When vascular damage happens, not only the procoagulant subendothelial structures of the vessels are exposed, but also the endothelial cell membrane can, under the influence of cytokins, change from the resting state to an active procoagulant form by expressing receptors for blood clotting proteins and by suppressing anticoagulant functions. Thus, clotting, as an immediate response, can start rapidly (1). There are two principal clotting mechanisms in response to the vascular damage. The first response is the platelet reaction. Platelets circulating in the blood will be activated from the resting state and adhere to the site of damage, then aggregate to form a platelet plug. The platelet plug reduces or temporarily stops the loss of blood. The activation of platelets also induces a series of physiological responses including the release of numerous proteins and small molecules which speed up and increase platelet plug formation and begin the tissue repair (2). The second response is blood coagulation. An intimate relationship between platelet reaction and blood coagulation is that aggregated platelets with damaged cell membranes provide the negatively charged phospholipid surface for the subsequent coagulation cascade to occur (3, 4), and for deposition and crosslinking of fibrin.

The coagulation mechanism has been studied since early in the century by Morawitz (5). In that "classic theory", it was envisioned that clotting resulted from a two-step reaction. In 1964, the coagulation cascade was proposed by Davie et al. and MacFarlane separately (6-7). This hypothesis set a valuable foundation for understanding the clotting mechanism even though a number of modifications have been made.

The process of coagulation consists of a series of sequential activation of certain plasma proenzymes to their active enzyme forms by either the intrinsic or extrinsic clotting pathway. Recent investigation suggests that the intrinsic pathway plays an important role in the growth and maintenance of fibrin formation in the clotting pathway, while the extrinsic pathway is crucial for the initiation of fibrin formation (4). Figure 1.1 shows the clotting cascade and fibrin formation by either the intrinsic or extrinsic pathway.

The trigger of the extrinsic pathway is the tissue factor which is located in tissue adventitia and contacts with clotting factors circulating in blood only after vascular injury (8-9). Tissue factor initiates the coagulation by being a cofactor. Tissue factor has high affinity to factor VIIa circulating in blood. It binds to factor VIIa in a one-to-one ratio in the presence of calcium ion. The factor VIIa-tissue factor complex activates factor X and factor IX when calcium ion is present. Factor Xa activates prothrombin to thrombin, and thrombin converts the soluble fibrinogen to fibrin monomer.

Polymerization of the fibrin monomers generates the insoluble fibrin clot.

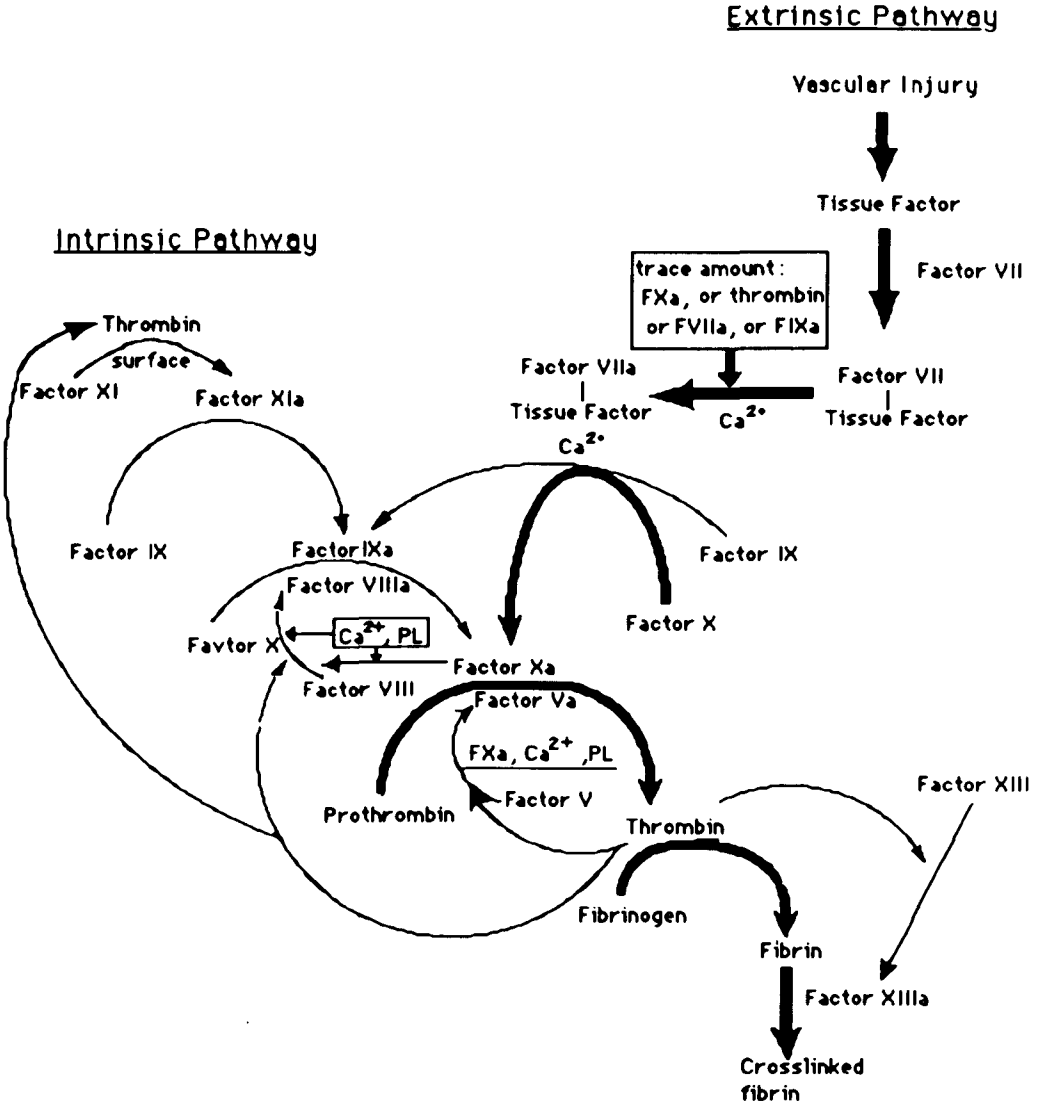
The activation of factor IX or factor X by tissue factor-factor VIIa complex depends on the concentration of tissue factor. When the tissue factor concentration is low, factor IX is a better substrate than factor X (10, 11). Factor IXa forms a complex with factor VIIIa (which is activated by factor Xa when there is calcium ion and phospholipid present, or by thrombin alone). The complex of factor IXa-factor VIIIa can convert factor X to factor Xa through hydrolysis of the same peptide bonds as the complex of VIIa-tissue factor does. A deficiency of factor VIII or factor IX will cause hemophilia **A**, or **B**, respectively.

After the discovery of lipoprotein-associated coagulation inhibitor (LACI) in blood, the importance of the intrinsic pathway was clarified. LACI inactivates the factor VIIa-tissue factor complex by binding to factor VIIa-tissue factor, and then binding to factor Xa (12-14). This inhibits the activation of factor X. However, interestingly, LACI does not block the activation of factor IX by the factor VIIa-tissue factor complex. Therefore, factor X can only be activated by factor IXa through the intrinsic pathway.

As long as the extrinsic pathway is diminished by LACI, the intrinsic pathway turns into a major route for the continuation of fibrin clot formation. It has been shown that the intrinsic pathway requires factor XIa which is activated by thrombin on a negatively charged surface. In addition, trace amounts of factor XIa can

activate factor XI in an autocatalytic manner (15). Another mechanism of activating factor XI is by factor XIIa, plasma prekallikrein, or high molecular weight kininogen. This conversion is observed only when blood plasma contacts with a synthetic surface such as glass or kaolin *in vitro*. Then, factor XIa converts factor IX to factor IXa. Factor IXa starts to activate the next step in the cascade (Figure 1.1.).

Figure 1.1. Schematic representation of blood coagulation cascade. (redrawn from reference 4). The heavy arrows link the extrinsic pathway; the thin arrows link the intrinsic pathway. FXa, FIXa, FVIIa, FVa, FXIIIa, and FXIa represent activated forms of factor X, factor IX, factor VII, factor V, factor XIII and factor XI. PL represents phospholipid.



1.1.2. Regulation of coagulation

Blood coagulation is under regulation of a number of protease inhibitors. Some patients with deficiencies of clotting blockers suffer different degrees of thrombotic episodes. Regulation of coagulation is quite complicated. Coagulation can be terminated by inactivating the proteases, for instance, by the antithrombin III-heparin system, or by inactivating the cofactors involved (16). The primary mechanism for degrading cofactors of the coagulation cascade is the thrombomodulin-protein C - protein S system. Thrombin is the only known physiologic activator of protein C. With phospholipids and calcium ion together, this anticoagulation system can selectively inactivate factor Va and factor VIIIa, inhibiting thrombin generation. This inhibition of thrombin formation leads to reduced platelet activation and clot generation (17). Antithrombin III can inhibit thrombin, factor IXa, factor Xa and factor XIa by forming one-to-one irreversible complexes with them. This inhibition is enhanced if heparin is present. Fortunately, this is a slow inhibition process. The coagulation system has enough time to generate fibrin before those proteases are blocked. Some other less important inhibitors are also noted in plasma, such as heparin cofactor II (18), α2-macroglobulin (19), activated protein C inhibitor (20), and α1-antitrypsin (21). In addition, coagulation may be slowed down by the dilution of activated clotting factors by circulating blood as well as the trapping of thrombin on the fibrin network. Moreover, the endothelial cells on the vascular wall may

have multiple functions in the regulation of the clotting system (22).

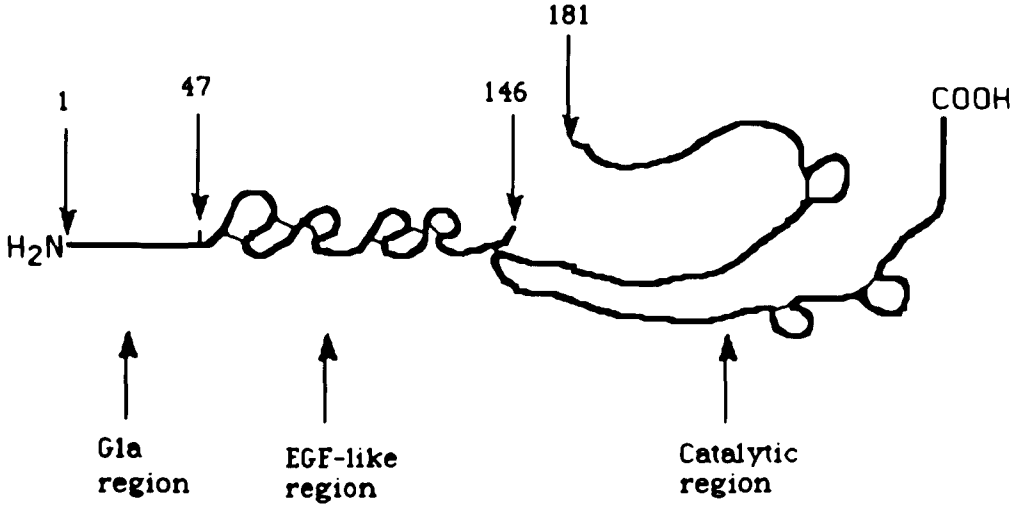
1.2. Human factor IX

Factor IX (Christmas factor) is one of the important proteins in blood coagulation. Individuals lacking this protein may bleed spontaneously into their skin, soft tissues, and joints, and the bleeding is often serious after minor injury. This genetic defect disorder is called Hemophilia **B**, or Christmas disease.

Human factor IX, a single chain plasma glycoprotein (M_r 56,000), is a serine protease zymogen. It contains about 17% carbohydrate (23). This protein is synthesized in the liver as a precursor molecule that undergoes a series of modifications prior to secretion (24). These post-translational processing reactions include glycosylation, cleavage of pre- and propeptides, vitamin K-dependent γ-carboxylation of the 12 glutamic acid residues at the amino-terminal site (25), and β-hydroxylation of aspartic acid-64 (26). Mature factor IX consists of 415 amino acid residues which are divided into several regions. Each region has its characteristic functions. The first fragment comprising 46 amino acids from the NH-terminal, a product of second and third exons, is called the Gla region. The region from residue 47 to 127 represents the two epidermal growth factor-like modules. Residue 128 to 195 are coded by exon six where the activation peptide (which is cleaved after activation) is located between 146 to 181. The last 220

residues are products of the last two exons, and contain the catalytic center of factor IXa (24). Figure 1.2 gives a schematic picture of the general structure of factor IX.

Figure 1.2. Schematic diagram of activated factor IX structural domains. The numbers indicate the amino acid positions. (Redrawn from reference 27.)



1.2.1. Gla region

Factor IX is a vitamin K-dependent protein. Its activity requires vitamin K-dependent carboxylase system associated with the microsomes of hepatocytes. This carboxylase converts 12 glutamic acids on the NH-terminal fragment of the protein to

gamma-carboxyl-glutamyl residues (Gla) during the post-translational modification. This modification is also found in other vitamin K-dependent clotting proteins, such as prothrombin, protein C, factor X, protein S, and protein Z (24).

When the Gla region binds calcium, it undergoes certain sequential conformational changes which are necessary for exposing membrane-binding sites, and for the clotting activity (28). The importance of the calcium-dependent conformation change in the Gla region was indicated by chemically modification study on Gla residues. Once the Gla residues have been chemically modified, the clotting activity of factor IX is inhibited (29).

1.2.2. EGF-like region

The region next to the Gla domain is composed of two tandem repeats of EGF-like modules. Little is known about the biological function of this fragment. However, it has been demonstrated that EGF-like modules in factor IX, X, and protein C bind one calcium ion. Genetic mutations in this region of factor IX seriously affect the factor IX clotting activity. More details about this region will be given in section 1. 3.

1.2.3. Activation peptide region

The activation peptide region is next to the EGF-like modules. Two activation sites, an Arg-Ala bond (residues 145 and 146) and an Arg-Val bond (residues 180 and 181), are located in this

region. The cleavage of these two internal peptide bonds results in the formation of an activation peptide consisting of 35 amino acids and factor IXa, a serine protease. There is a cysteine residue at position 132 which forms a disulfide bond with the cysteine residue at position 289. This disulfide linkage can account for the binding between light and heavy chains of factor IXa. During the activation of factor IX to factor IXa by factor XIa, calcium ion is the only required cofactor (30). Factor XI can be found in platelets but factor XIa does not bind to the platelet surface (24). Factor IX can be also activated by the factor VIIa-tissue factor complex, which cleaves the same peptide bonds as factor XIa in the presence of calcium ion (31). The activated peptide with two Asn-linked carbohydrate side chains has a molecular weight about 10,000. Factor IXa is 20% smaller than factor IX as demonstrated by their migration on polyacrylaminde gels (24).

1.2.4. Catalytic region

In the catalytic region of factor IXa, His^{221}, Asp^{269} and Ser^{365} are three essential components in the active site (24), homologous with chymotrypsin which has His^{57}, Asp^{102}, and Ser^{195} in the active site (30). The catalytic mechanism of FIXa, therefore, may be similar to chymotrypsin and trypsin.

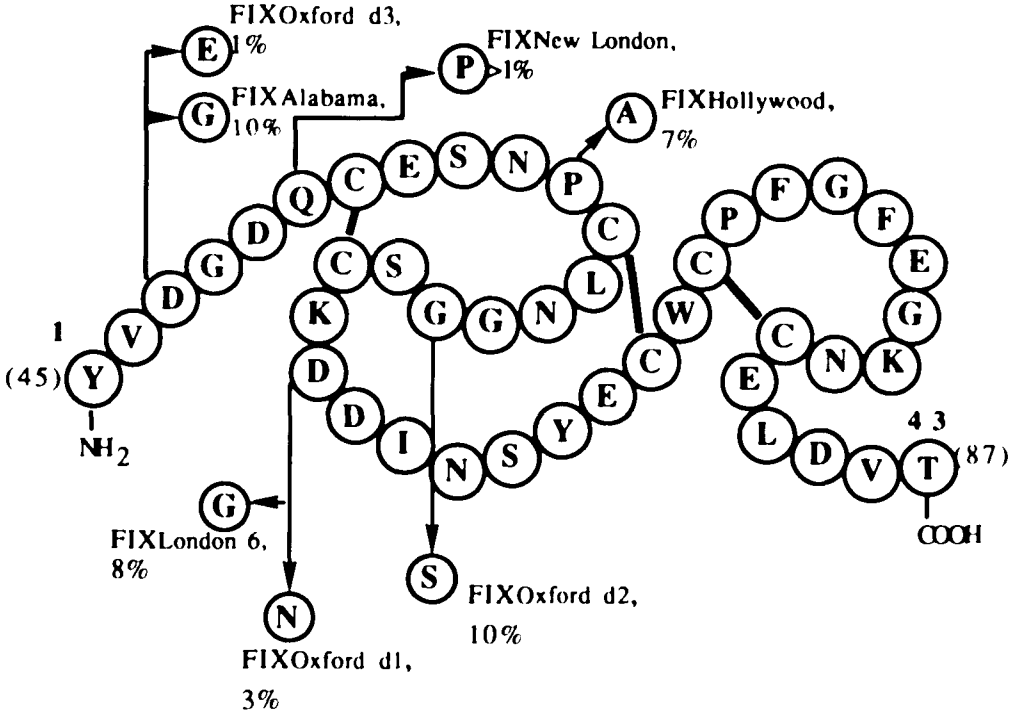
1.3. Structural feature of the first EGF-like module

The fourth and fifth exons in the human factor IX gene code for two EGF-like modules. This is a cysteine-rich region which contains 12 of the 22 cysteines found in the entire protein. In each module three disulfide bonds are arranged in homologous order to EGF. An unusual amino acid, β-hydroxyaspartic acid (Hya), at position 64 of factor IX, is also present in homologous positions in factor X and protein C (32). A liver microsome enzyme, Asp-β-hydroxylase, is responsible for β-hydroxylation. It has been shown that the synthetic first EGF-like module can be a substrate of the hydroxylase (33). In human factor IX the hydroxylation is about 30% (34). The significance of this modification is not clear. Post-translational glycosylation is also found in the first EGF-like module. The carbohydrate units are found in serine 53 and 61 in human factor IX, similar to what is found for factor VII, and protein Z. Factor X and protein C lack a serine residue in position 53, and do not have glycosylation. The function of the carbohydrate unit is unclear. The glycosylation enzyme has not yet been determined (27).

A number of genetic defects in the EGF-like region of factor IX cause hemophilia B. The gene mutation at position 47 that generates Glu^{47} (factor $IX_{Oxford\ d3}$) instead of Asp^{47} will reduce the clotting activity of factor IX to only 1% of normal human plasma clotting activity (27). Another mutation at position 47 giving rise to Gly^{47} (factor $IX_{Alabama}$) instead of Asp^{47} drops clotting activity to about 10% (35-36). The mutation in which Pro^{50} replaced Gln^{50}

is called factor $IX_{New\ London}$. In this case, the clotting activity is reduced to less than 1% (37). At position 55 the replacement of Pro^{55} by Ala^{55} generates a mutant named factor $IX_{Hollywood}$. This gene defect makes the factor IX clotting activity down to about 7% (38). The substitution on Gly^{60} by Ser^{60} reduces the factor IX clotting activity to 10% (39). This mutation is named factor $IX_{Oxford\ d2}$. There are two different mutations known at position 64, one of them is substitution Asp^{64} by Gly^{64} that is called factor $IX_{London\ 6}$ with 8% clotting activity (40), another is substitution by Asn^{64} named factor $IX_{Oxford\ d1}$ with 3% clotting activity (39). The fact that mutations on this module dramatically effect factor IXa's clotting activity shows the importance of this EGF-like module in factor IX molecule. The possible explanation for those mutation effects may be that the amino acid replacement causes the conformational change of the structure or disturbs the calcium binding site which is required for its protease catalytic activity. Figure 1.3 depicts a schematic diagram of mutations known to lead to hemophilia B in the first EGF-like module of human factor IX.

Figure 1.3. Schematic diagram of genetic defects on FIX-EGF1. Amino acid one letter symbols are used. The mutations at different positions are pointed out by arrows (redrawn from reference 27).



There is a calcium binding site in the first module of factor IX with a k_d of 0.4 mmol/L (41-42). This module has been chemically synthesized in this study, and has Asp^{64} instead of hydroxylated Asp^{64} (Hya). The module has also been expressed in yeast. And it binds to calcium (43).

It has been proposed that one important function of the EGF-like modules is to be a spacer between the Gla region and the catalytic region (27). This spacer can mediate the distance between the active site and the phospholipid surface. A similar role is possible for the four EGF-like modules of protein S. In the case of receptor proteins, such as thrombomodulin and the LDL-receptor, there are carbohydrate side chains adjacent to the membrane, followed by the EGF-like modules. This structural feature may be beneficial for proper separation of the ligand-binding site from the membrane surface (27). Recent research data suggest that factor IXa and its substrate factor X interact directly through their EGF-like modules and that the Gla module of factor IXa and the serine protease part interact with factor VIIIa (44).

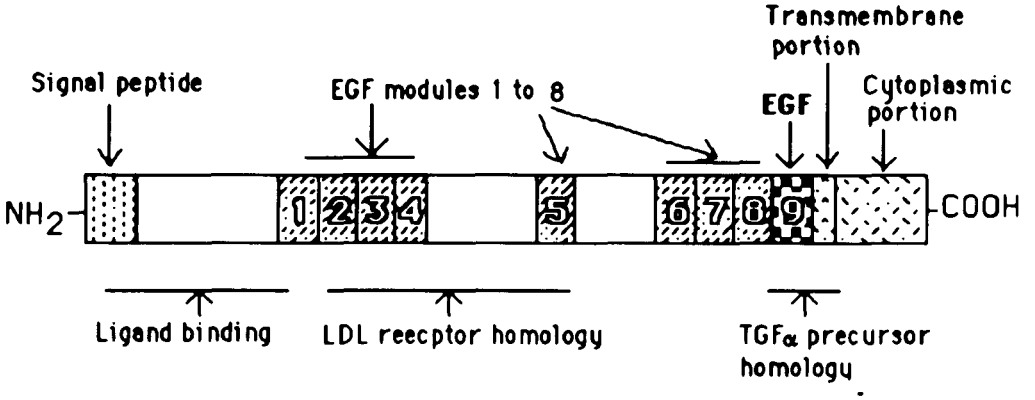
1.4. Epidermal growth factor

The EGF precursor is the archetype of proteins that contain EGF-like modules. This large membrane-spanning protein contains nine EGF modules in the extracellular NH_2-terminal region (45-48). The active EGF molecule is the one that locates next to the cell membrane and has arginine residues on either side after a trypsin-like endopeptidase cleavage. Figure 1.4 gives a schematic diagram of EGF-precursor. The biological function of the other eight NH_2-terminal EGF modules is unclear.

EGF is a small mitogenic protein containing 53 amino acids. Three disulfide bonds between cysteine residues 6-20, 14-31, and 33-42 forming three intramolecular loops are a characteristic structural feature of EGF-like modules commonly called the A, B, and C loops. In addition, another characteristic of EGF-like molecules is the capacity to compete with EGF to bind to the EGF-receptor and to mimic biological activities of EGF.

The proliferation of cells *in vivo* and in culture is largely regulated by growth factors, such as EGF and TGFa. Those growth factors, like polypeptide hormones, initiate their action by binding to specific cell-surface receptors. The activated receptor then mediates a cascade of rapid biochemical and physiological changes in the cell, which ultimately lead to DNA synthesis and physiological changes. EGF is one of the most extensively studied growth factors. It can cause a wide range of physiological and mitogenic responses after binding to cell-surface receptor.

Figure 1.4. Schematic diagram of EGF precursor. The portions of protein and EGF modules from 1 to 9 are indicated.



In vitro, EGF is a potent mitogen for a number of cultured cells; *in vivo*, it induces the proliferation and differentiation of various epidermal tissues. Physiological responses include accelerating wound healing, ulcer healing, reducing gastrointestinal tract acid secretion, and inhibiting hair growth. It also has complicated effects on many organs such as the kidney and liver as well as the nervous system, pituitary hormone secretion, and the reproductive system (45).

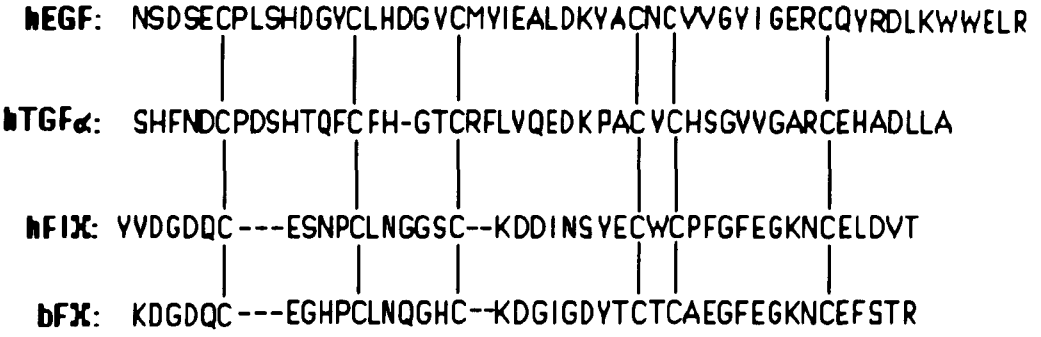
Since EGF and its family members show diverse biological functions, many efforts have been made to explore their reaction mechanisms and structural features. In recent years high

resolution multiple dimensional NMR studies have provided structural details about EGF and its family members. In general, there are two subdomains in the EGF molecule: residues 1 to 32, and residues 33 to 53. The NH_2-terminal subdomain is made of loops A and B, and contains one antiparallel β-sheet. The COOH-terminal subdomain is made of the C loop, which contains one shorter β-sheet. The structure of TGFα has also been obtained by 2D-NMR spectroscopy and was found to be similar to that of EGF. Synthetic peptide fragments containing COOH-terminal domain with intact disulfide bonds have been found to have little or much lower putative biological activity (49). This may indicate that to obtain the full biological activity of EGF, right conformation and correct amino acid sequence for interacting with the receptor is required.

Similar cysteine pairing structures are found in a diverse array of proteins. These include extracellular proteins, including vitamin K-dependent blood coagulation proteins (50) and several complement system proteins (51). This module is also seen in a number of cell-surface proteins, such as the low density lipoprotein (LDL) receptor (52), and the *Drosophila Notch* protein (53). These EGF-like modules are classified into different groups due to their different sequences and functions. One group of the modules includes EGF, TGFα, and several viral proteins. The modules in this group can bind to EGF receptor and have mitogenic activity. Another group of the EGF-like modules, including blood clotting

proteins and receptor proteins, may be involved in cellular interactions. Biological functions of EGF-like modules in the later group is still under investigation. Figure 1.5 illustrates sequence homology between EGF and some of the EGF-like modules.

Figure 1.5. Sequence comparison by aligning cysteine residues of hEGF, hTGFα, human factor IX first EGF-like module and bovine factor X first EGF-like module. (hEGF: human EGF, hTGFα: human TGFα, hFIX: the first EGF-like module of human factor IX , bFX: the first EGF-like module of bovine factor X .)



1.5. Calcium in the clotting system

The importance of calcium ion in biology and physiology has been noticed since the last century. In 1882, Ringer first

demonstrated that millimolar calcium is required for maintaining contraction of eel and frog hearts (54).

Calcium ion is one of the four significant alkali and alkaline earth metal ions (the other three ions are K^+, Na^+, and Mg^{2+}) in living systems. It exists plentifully in shells and corals as $CaCO_3$, and occurs as hydroxyapatite, $Ca_{10}(PO_4)_6(OH)_2$, in skeletons of insects and vertebrates (74). In biological fluids, Ca^{2+} is involved in many biological functions, such as muscle contraction, blood clotting, neurotransmitter release, microtubule formation, intercellular communication, hormonal responses, exocytosis, fertilization, cell fusion, adhesion and growth. Many of these Ca^{2+} related biological activities occur by interaction with proteins (55).

The major effect due to association and dissociation of calcium binding to protein is often a conformation change. Usually, there are seven or eight, rarely six, coordination sites for calcium (55). Calcium binding sites in proteins are composed of negatively charged and neutral oxygen donors; nitrogen donors are unlikely and none have been found in proteins (55). The major donor groups from the protein to the calcium ion are carboxylate groups, carbonyl oxygens of the amide backbone, and hydroxyl groups of serine and threonine side chains. These donor groups bind to calcium through oxygen atom in different ways. Some are close to each other sequentially, some are distributed in distant in amino acid sequence. The later depends largely on a predetermined fold which gathers those donors close in space to generate a rather

immobile cavity for calcium. The amino acid side chains involved in calcium binding are those of Asp, Glu, Asn, Thr, and Ser in most cases (56).

Calcium ion plays an important and unique role in clotting. The clotting cascade (figure 1.1.) shows that calcium ion is a cofactor in many activation steps, including activation of FX to FXa by FIXa. There are at least two calcium-binding regions in calcium related clotting proteins. One is the Gla domain, in which calcium ion binding is thought to bridge carboxylate groups of Gla residues of the membrane binding structure and phosphate groups of a membrane surface (57). The other is the tight calcium binding region. The location of a high affinity calcium binding site in protein C was published recently by Rezaie et al. (58). It was reported that the high affinity Ca^{2+} - binding site important for the activation of protein C is located outside of the first EGF-like module, and the Ca^{2+} binding site in the first EGF-like module may not be a high affinity site in intact protein C (58). Since factor IX and protein C are the same family members, they both have similar domain structures. A calcium binding site in the first EGF-like module was reported for both proteins. It is interesting to know the property of the calcium binding site in this region, and, of course, to locate the high affinity Ca^{2+} binding site out of the EGF-like module. The calcium binding at FIX-EGF1 region was studied in this project, and it will be discussed in section 3.3.

1.6. 2D-NMR in determination of protein solution structure

During the last decade, nuclear magnetic resonance (NMR) has been introduced as another method for protein structure determination in addition to the well-established X-ray crystallography techniques. NMR enables one to study the structural details in solution and other noncrystalline states.

The fundamentals of 2D-NMR are introduced in many excellent texts (59-64). Basically, most 2D-NMR experiments involve four time periods: a preparation period which is a relaxation delay to ensure that the spin system is in equilibrium; an evolution period (t_1) during which spins process at their own characteristic frequencies; a mixing period or mixing pulse to transfer coherence between spins; and a detection period (t_2) during which the nuclear magnetization is sampled. A number of experiments are recorded with successive small increments in the evolution time t_1 to generate a time-domain matrix $S(t_1, t_2)$. Then, two-dimensional Fourier transformation (FT) generates the frequency-domain matrix $S(\omega_1, \omega_2)$ which is a two-dimensional spectrum (65). There are two type of experiments which serve as major approaches for structure determination. One is homonuclear-correlated spectroscopy (COSY). The cross-peaks of COSY arise from through-bond connectivity (65). COSY is used to identify the characteristic networks of spin-spin coupled protons associated with different amino acids in a protein. A number of various forms of COSY experiments are designed for different

purposes. Commonly used COSY experiments include relayed-COSY, total correlated spectroscopy (TOCSY or HOHAHA), and double quantum filtered COSY (DQF-COSY).

The other major experimental approach is nuclear Overhauser enhancement spectroscopy (NOESY), which provides information about proton-proton interaction through space (66-67). The cross-peaks of NOESY arise from through-space correlations. For short mixing times in NOESY experiments (less than 50 ms for small proteins of fewer than 100 amino acid residues), only relatively short inter-proton distances (less than 3-3.5Å) will give rise to observable NOEs. If the mixing time is set longer, more NOEs can be observed (distance between protons up to 5Å) (64). However, at longer mixing times, spin diffusion may yield misleading cross-peaks.

Sequence-specific assignment: This is the first step for structure determination of proteins. Assignment usually starts with identification of resonances belonging to the same amino acid residue. This information can be acquired from COSY experiments which provide through-bond J-coupling connectivities between protons belonging to the same spin system. The assigned spin system is then classified to a specific amino acid type or a group of amino acid types. In addition to through-bond connectivity information, through-space NOEs can also provide information for identification of certain amino acid types. For instance, one can look for cross-peaks of C^{β} protons and aromatic protons for aromatic

amino acids in NOESY. After the amino acid type has been identified, NOEs are used for connecting one amino acid residue to another, often aided by a known protein sequence. A schematic illustration of NOE connection between amino acid residues are given in Figure 1.6. In this figure, distance separation which leads to the NOE connection between the alpha proton of a proceeding amino acid and the amide proton of the next amino acid is presented as: $d_{\alpha N}=d(C^{\alpha}H_i, NH_{i+1})$. The distance between amide protons next to each other is presented as: $d_{NN}=d(NH_i, NH_{i+1})$. The distance between the beta proton of proceeding amino acid and the amide proton of the amino acid following it is presented as: $d_{\beta N}=d(C^{\beta}H_i, NH_{i+1})$. Those notations are also used elsewhere in this text.

Secondary structure determination. Secondary structure of a protein can be assigned from long range NOEs, such as non-sequential $d_{\alpha\alpha}$ (NOE cross peak between C^{α}Hi and C^{α}Hx; x represents an arbitrary position, but not a sequential one) in D_2O solution, and $d_{\beta N}$, $d_{\alpha N}$, and d_{NN} NOEs in H_2O solution. The parameters used for secondary structure assignment are listed in Figure 1.7.

Figure 1.6. Schematic illustration of the sequential assignment procedure is listed. The amino acid 1, 2 and 3 are identified first by through-bond 2D-COSY experiments. Adjacent amino acid residues are then identified by looking for NOE connections between them. Those NOE connections include $d_{\alpha N}=d(C^{\alpha}H_i, NH_{i+1})$, $d_{NN}=d(NH_i, NH_{i+1})$, and $d_{\beta N}=d(C^{\beta}H_i, NH_{i+1})$ (68).

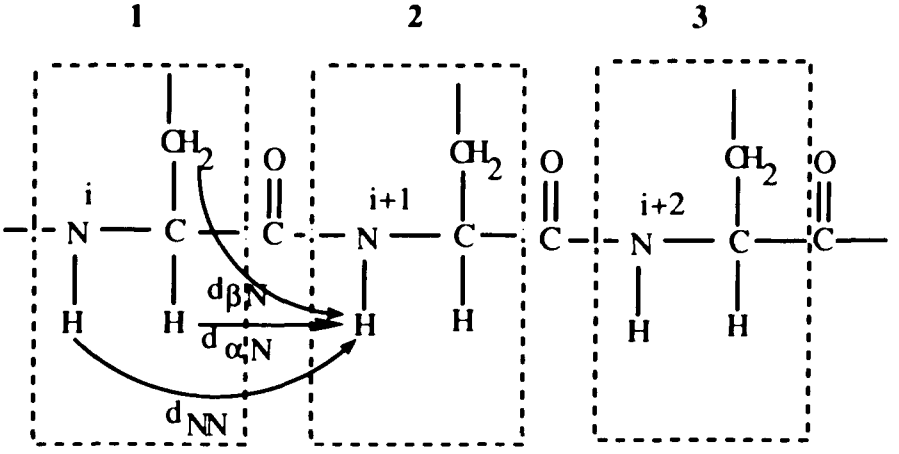


Figure 1.7. NMR parameters used to identify secondary structure in proteins (69).

β-sheet : $d_{\alpha N} = 2.2Å$ $^3J_{HN\alpha} > 8$ Hz

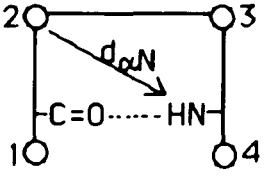
α-helix : $d_{NN} = 2.8Å$ $^3J_{HN\alpha} < 5.5$ Hz

$d_{\alpha N}$ (i, i+3)= 3.4 Å

$d_{\alpha\beta}$ (i, i+3) = 2.5 - 4.4 Å

β-turn : d_{NN} (3,4) = 2.4 Å

$d_{\alpha N}$ (2,4)= 3.3 Å



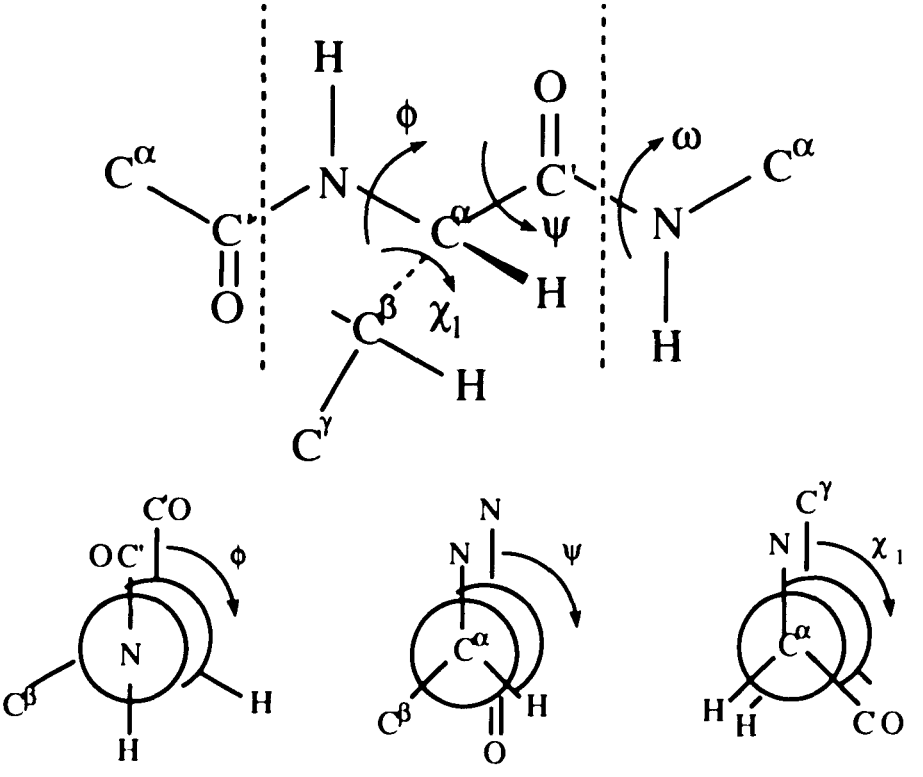
Another valuable complement to NOE distance data for peptide backbone conformation study is the use of vicinal spin-spin coupling constants, which are torsion angle dependent. The backbone conformation of a peptide is defined by three torsion angles for each constituent amino acid residue. The conformation about a peptide bond itself is described by the dihedral angle ω,

while those about the Cα-N and Cα-CO bonds are described by ϕ and ψ, respectively. Figure 1.8 shows a peptide segment with these angles (70). The Karplus equation gives a relationship between coupling constant and angle θ: $^3J_{HN\alpha} = A + B \cos \theta + C \cos^2\theta$ (70-72), with $\theta = |\phi - 60°|$ and $^3J_{HN\alpha}$ given in hertz. In this equation, A, B and C are constants used for electronegativity correction. Based on correlating with $^3J_{HN\alpha}$ and the ϕ angles in the crystal structure of BPTI (bovine pancreatic trypsin inhibitor), the equation for $^3J_{HN\alpha}$ was given as: $^3J_{HN\alpha} = 6.4 \cos^2\theta - 1.4 \cos \theta + 1.9$ (68). Individual small (< 6.0 Hz) or large (> 8.0 Hz) $^3J_{HN\alpha}$ couplings are combined with NOE data to identify helical or β-sheet structures, respectively (68).

The formation of hydrogen bonds between backbone amide protons and backbone carbonyl oxygens is one of the features in peptide secondary structures. The observation of the amide exchange rate during an NMR experiment is a common method to determine the location of the hydrogen bonds. In a β-sheet structure, a hydrogen bonding network between neighboring strands is expected.

Calculation of three dimensional structure by distance geometry. The final step of a structural determination is to find

Figure 1.8. A schematic diagram of a peptide fragment with dihedral angles shown. A peptide fragment (L-amino acid residue) in the extended conformation ($\phi = \psi = \omega = 180°$) and the initial projections $\phi = \psi = \omega = 0°$ for torsion angle determination (Redrawn from reference 69).



three-dimensional structure from NMR-derived constraints. There are two general approaches to generate structures from NMR-derived constraints. One way is distance geometry. It uses a matrix method, named embedding, to convert distances between points

into coordinates (73). Another way is to use a model structure or a random structure, and applying constraints directly, without the embedding step (73). In the FIX-EGF1 structure determination, the distance geometry method was used. Distance geometry can produce one or more molecular structures that meet a set of constraints from NMR experiments. The molecular structure generated from distance geometry methods is called a "starting structure". A starting structure is produced by fitting with the trial distances, which are chosen as randomly distributed between upper and lower boundary conditions (73). Generally, a starting conformation does not meet all the constraints. It may not be adjusted to the best value for energy, angles, or fit to NOE intensities. After generating a set of starting conformations with almost equally good agreement to distance constraints, the conformers with sufficiently high energy can be discarded directly, and a small number of conformers with lower energy will be used as input for more sophisticated refinement methods (73).

Refinement. This is an important step for obtaining a solution structure. A structure generated using the distance geometry method has some undesirable features, such as unfavorable potential energy due to the large number of small distortions in geometry and close nonbonded contacts. These are acceptable errors in the distance geometry objective function, but they contribute a large repulsion to the atom pair potential functions (73). An objective (or error) function is used to express

constraints, or to deal with violations of constraints. There are two common ways to do this. The first method assigns a target value for the parameter of interest. The objective function measures deviations from the optimum value. Alternatively, one can specify upper and lower bounds on a particular parameter. No penalty is assessed when a distance or an angle falls within its bounds, but a term is added to the objective function when the boundary conditions are violated. Most versions of distance geometry use this second method for handling constraint violations. In the literature, "objective function " and "error function" are often used interchangeably. There are a number of computational programs, such as AMBER, CHARMM, or ECCEP, that can dramatically reduce the unfavorable energy without much change in the fit to the distance constraints. These programs allow minimization of the energy while restraining the atomic positions close to their initial values (73).

Chapter Two

2. Structure and function relationship study of the first EGF-like module of human factor IX

The first step of this research project was to synthesize the first EGF-like module of human factor IX (45-87) (FIX-EGF1). Figure 2.1. depicts the amino acid sequence of this module with disulfide linkages. After finding that the synthetic FIX-EGF1 is devoid of any *in vitro* EGF-like activity, a number of chimeric EGF analogs using FIX-EGF1 as a framework were synthesized. For the purpose of this study, EGF was divided into five subdomains: the NH terminal, from the first amino acid to before the first Cys residue; the A loop, (amino acid residues between the first and the third Cys); the B loop, (amino acid residues between the second and the fourth Cys residues); the C loop, (amino acid residues between the fifth and sixth Cys residues); and the C-terminal (the amino acid residues after the sixth Cys residue). The chimeric analogs were designed to analyze the EGF subdomain contribution to EGF biological activity. Knowledge of the three-dimensional structures of human and mouse EGF, as well as of human TGFa, together with amino acid sequence comparisons led to a classification of two different kinds of important residues in EGF (74). Residues that seem to play their major role in stabilizing the structural integrity of the protein are Cys^{6}, Cys^{14}, Cys^{20}, Cys^{31}, Cys^{33}, Cys^{42}, Gly^{18}, Gly^{36}, Gly^{39} and Tyr^{37} in EGF. There are other residues that may be responsible for specific receptor recognition. All these residues:

Tyr^{13}, Leu^{15}, His^{16}, Arg^{41}, and Gln^{43} except Leu^{47} are on one side of the molecule (74). Therefore, they are likely to be directly involved in a specific interaction between EGF and its receptor. Based on sequence comparisons, FIX-EGF1 shares about 30% sequence homology with EGF and closely related proteins or peptides. However, those homologous amino acids mostly are structural residues, such as cysteines and glycines. Some conserved residues such as Tyr^{13}, Tyr^{37}, Arg^{41}, Asp^{46} and Leu^{47}, are found in EGF and its closely related molecules with EGF-like activity, but they are not in FIX-EGF1. The distribution of these conserved amino acids in EGF is in the A-loop, C-loop and carboxyl terminal. B-loop in EGF seems to play a structural role. In this project, four groups, consisting of a total of 15 analog peptides, were synthesized (Figure 2.1.). In group 1 three analogs were designed to examine the role of the A-loop in EGF-like activity. The size of the A-loop in FIX-EGF1 is three residues shorter than EGF. Therefore, in analog 1 seven residues from mEGF (mEGF 7-13) were coupled instead of FIX-EGF1 residue 8-11 to enlarge the loop size, and to insert a conserved Tyr^{13} into it. In analog 2 a fragment of hEGF 7-16 was coupled to replace a larger portion of the A-loop. For analog 3, hEGF 1-16 was used to replace most of the A-loop and all the NH-terminal of FIX-EGF1.

Group 2 was designed to examine the role of the A and B-loops. Analog 4 was synthesized with hEGF A-loop and B-loop sequence replacement, while analog 5 had the additional insertion

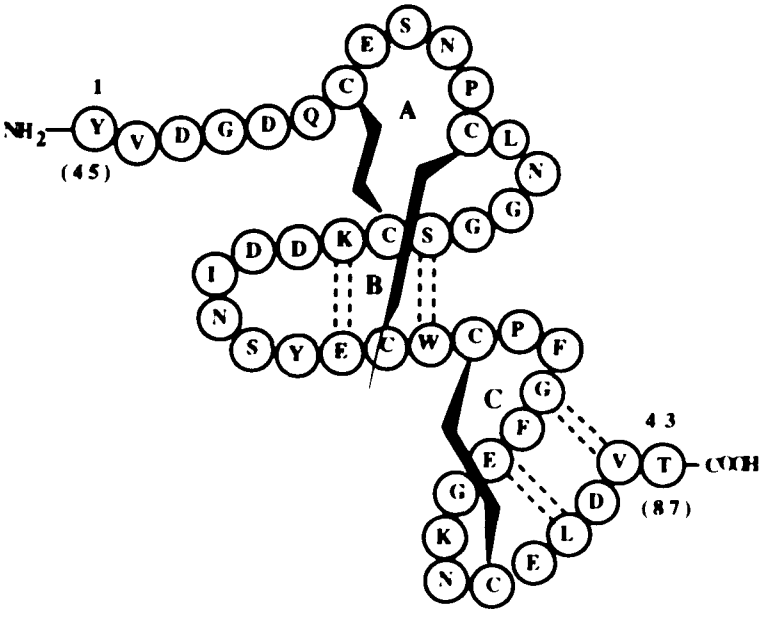
of hEGF Val^{34} and Tyr^{37} into the corresponding position in FIX-EGF1. The replacement Pro^{30} by Val^{34} was to avoid an unwanted bend at this position, the replacement of Phe^{33} by Tyr^{37} was to examine if the polarity of the side chain in the residue plays any role for the structural integrity.

The group 3 contained 7 analogs of varying structure. In analog 6, Arg^{41} of hEGF was the only residue used to replace Asn^{37} in FIX-EGF1 is considered as an important residue that plays both roles of receptor recognition and structural stabilization in EGF (75). Analog 7 had whole hEGF C-loop sequence replacing the FIX-EGF1 C-loop. Analog 8 is similar to analog 7, but with the addition of the YGDY sequence from mEGF inserted to enlarge the A-loop and also kept a conserved residue Tyr^{13}. Pro^{11} of FIX-EGF1 was deleted to avoid a possible bend at this position. In analog 9 two residues (Arg-Phe) from TGFa was inserted to enlarge the B-loop in addition to the same changes in analog 8. This sequence was chosen from TGFa because it contains a positively charged Arg residue at that position, which made the entire molecule less negatively charged. In analog 10 a Pro was placed next to the first Cys, as is found in hEGF, in addition to the same changes as in analog 9. In analog 11 whole C-loop and carboxyl tail of FIX-EGF1 had been replaced by hEGF sequences, in addition, its A-loop had been enlarged as in analog 9, but the B-loop was unchanged. In analog 12 only the B-loop of FIX-EGF1 remain, the rest of the sequence is that of hEGF.

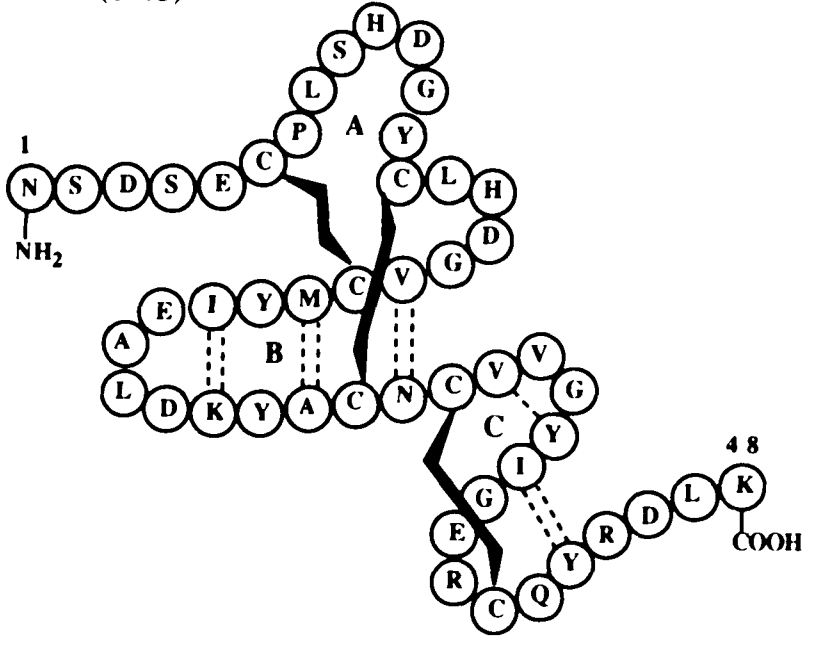
Group 4 was designed with the consideration of the results of structure-function studies of TGFa (Ala- and D-amino acid scan) by Professor J. P. Tam, and his coworkers at The Rockefeller University (76). Those amino acids considered as potential contributors to EGF-like activity were used to replace the corresponding amino acids in FIX-EGF1. Ala residue replacement at certain positions in TGFα showed increasing EGF-like activity slightly. Thus, corresponding Ala- replacements in FIX-EGF1 were performed. The positions where Ala replacement for the native amino acid residues were made are shown by arrows in Figure 2.1. The major difference between analog 13 and 14 is that the A loop in analog 14 was enlarged by adding in Pro^{8}, His^{11}, Ser^{12} and Tyr^{14} (deleted the native Pro^{14}). The difference between the analog 14 and 15 is only the Tyr/Phe replacement at position 38, and Ala/Glu replacement at position 41. The amino acid sequences of all 15 analogs are depicted in Figure 2.1. Figure 2.2. illustrates the schematic diagram of the subdomain replacement on the synthetic analogues. The lines indicate the FIX-EGF1 sequence, the boxes represent the replacement by EGF sequence, and the thin lines indicate the gaps. The synthetic products were characterized and their EGF-like activities were examined.

Figure 2.1. Schematic diagram of the structures of FIX-EGF1 (45-87), its 15 analogs, and hEGF (1-48) The shaded amino acid residues in each analog present the alteration in sequence from native FIX-EGF1. In Group 4, the arrows indicate the position where the native amino acid has been replaced by alanine (74). The dashed lines in analogs indicate the hydrogen bonding found in native sequences of EGF and FIX-EGF1.

FIX-EGF1 1-43 (45-87).

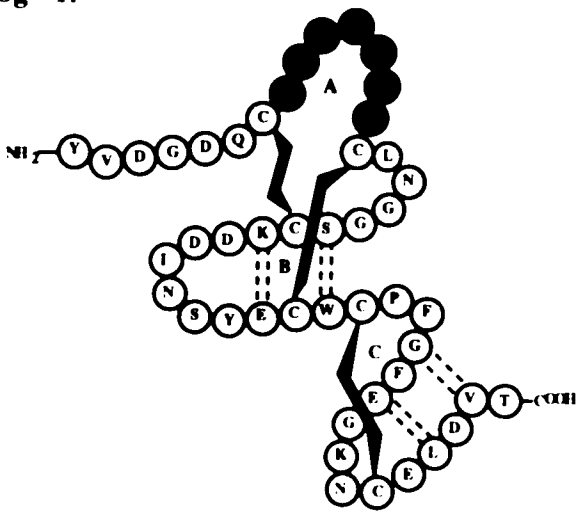


hEGF (1-48)

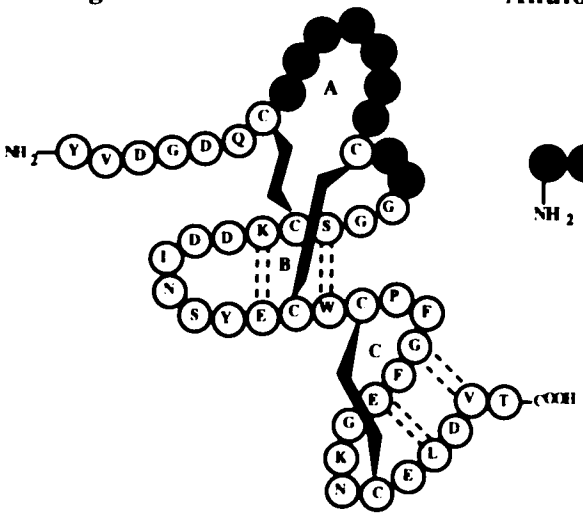


A. Group 1 analogs

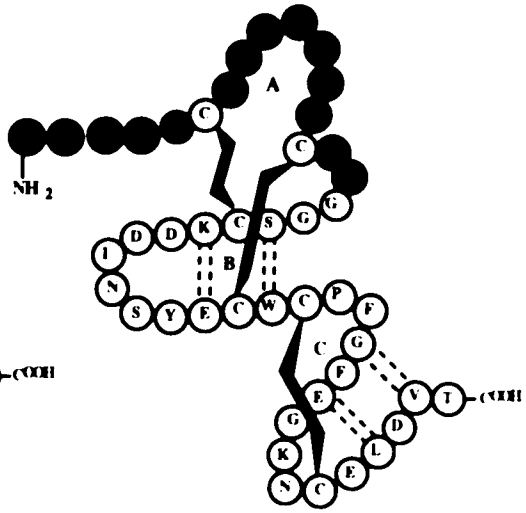
Analog 1.



Analog 2.

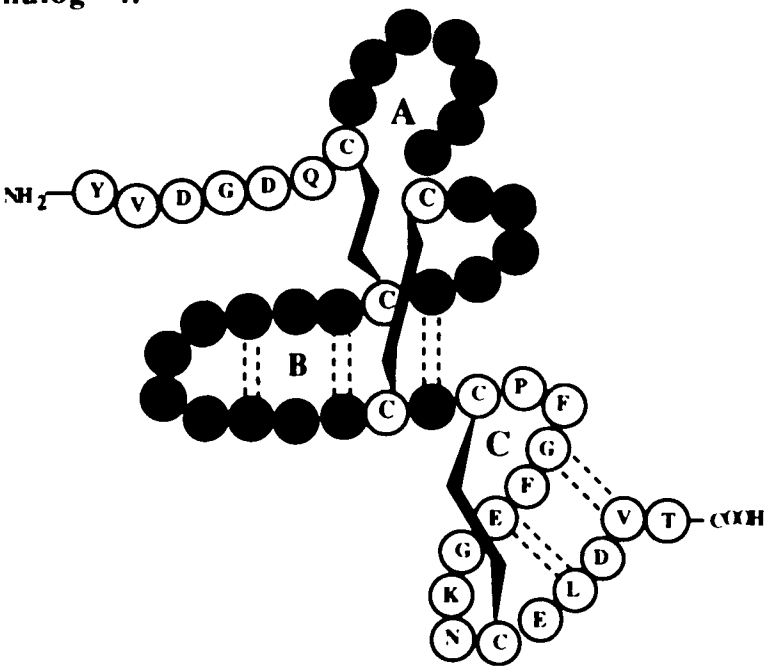


Analog 3.

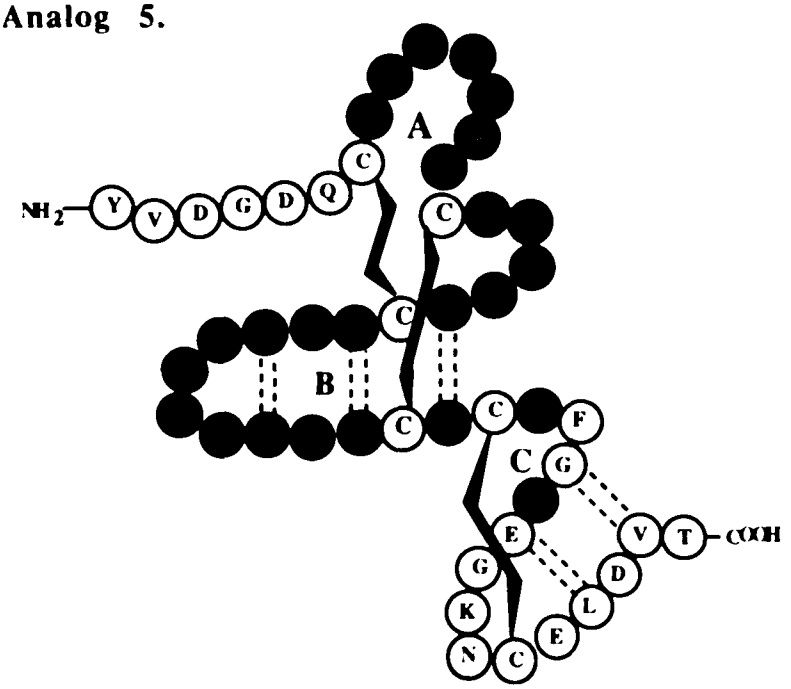


B. Group 2 analogs

Analog 4.

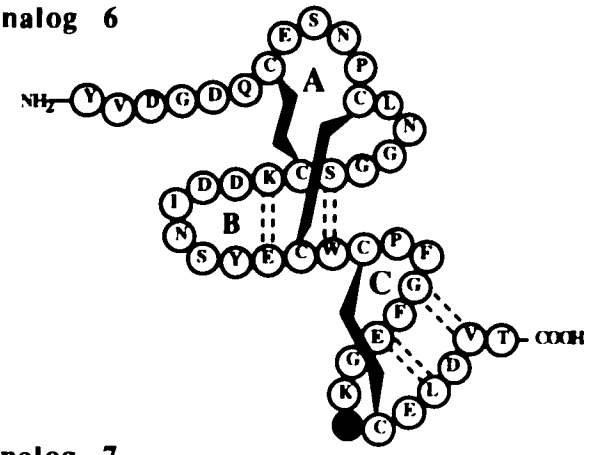


Analog 5.

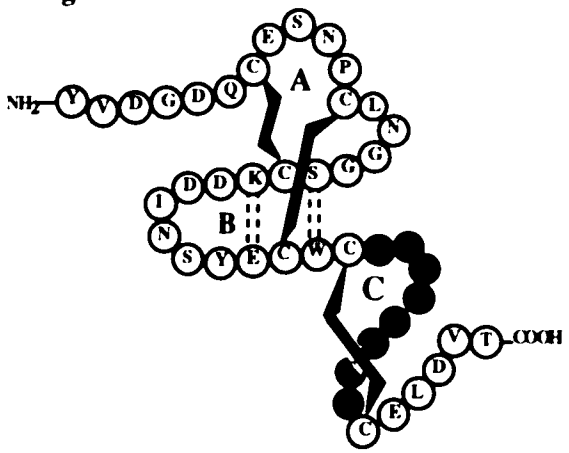


Group 3 Analog

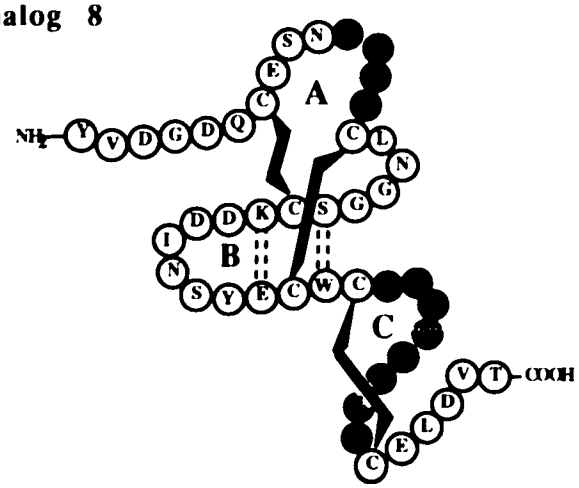
Analog 6



Analog 7

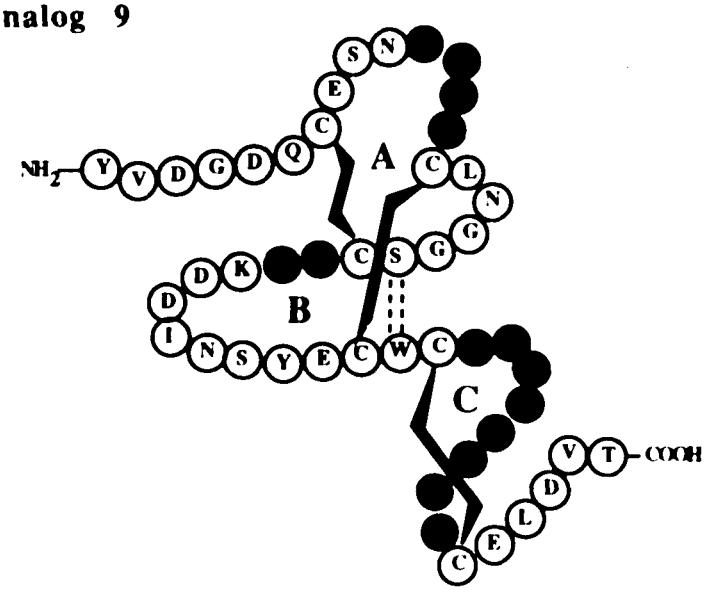


Analog 8

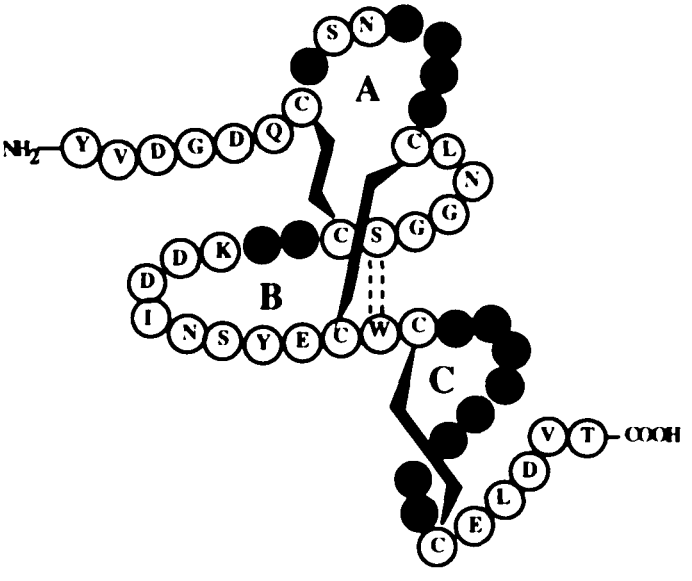


Group 3 analogs

Analog 9

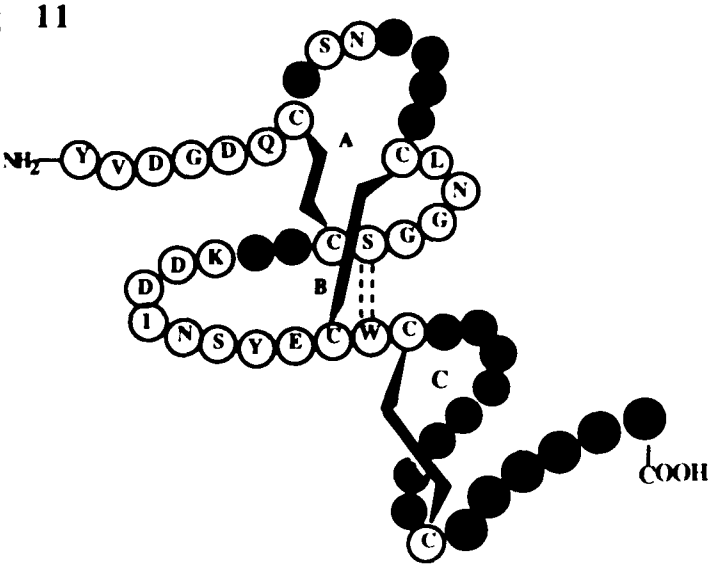


Analog 10

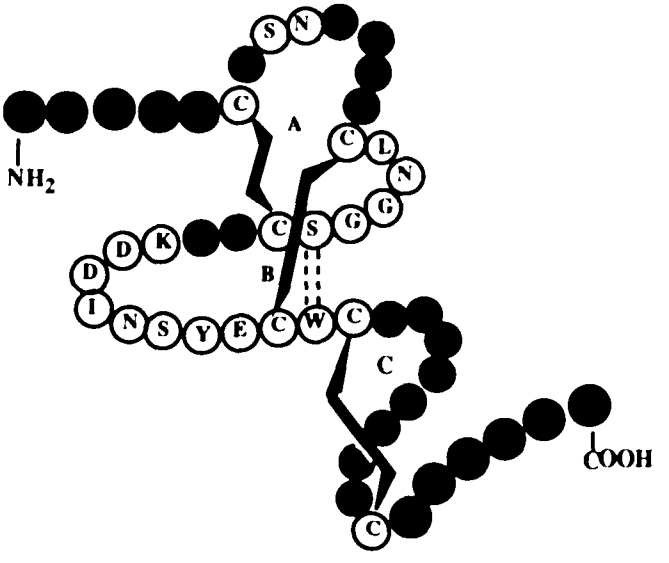


Group 3 analogs

Analog 11

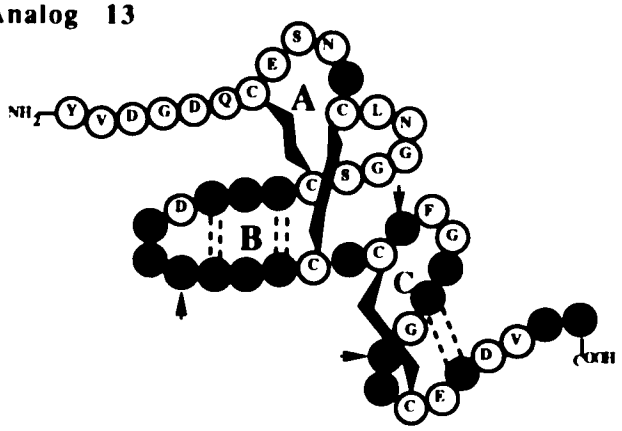


Analog 12

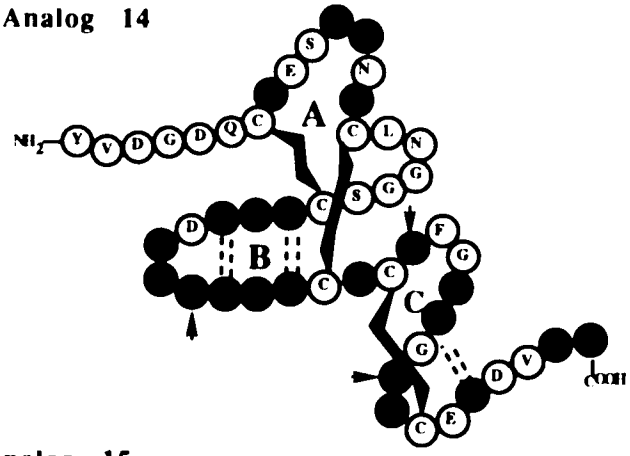


Group 4 analogs

Analog 13



Analog 14



Analog 15

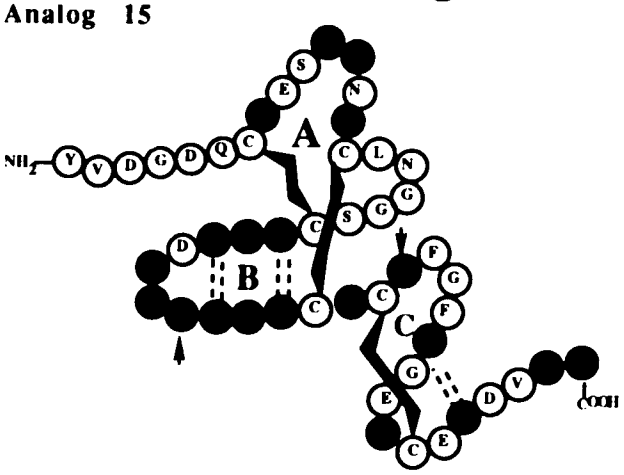
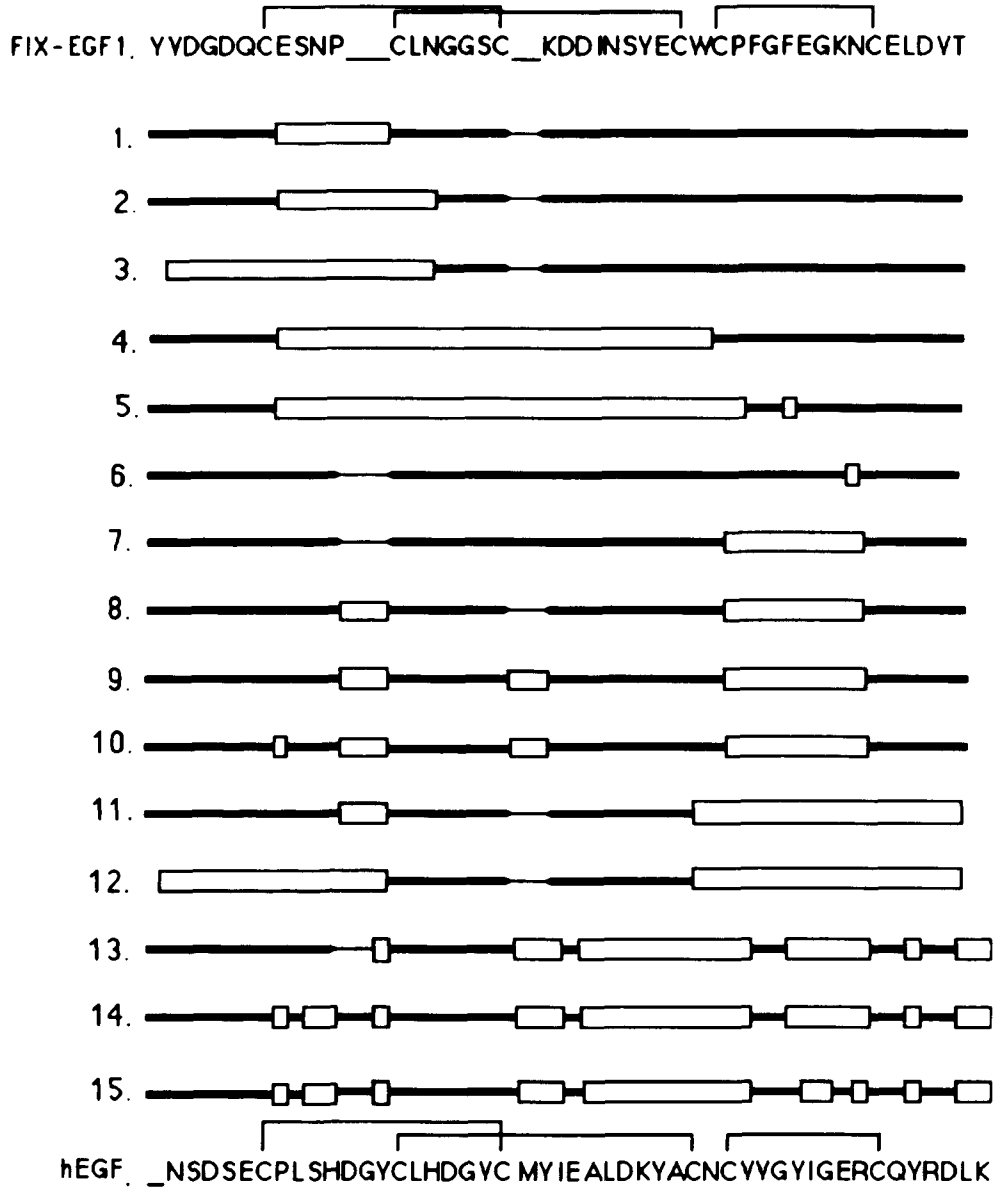


Figure 2.2. Schematic illustration of synthetic chimeric analogues. FIX-EGF1 sequence is on the top, while hEGF sequence is on the bottom. In the analog sequence, the lines represent FIX-EGF1 sequence, the boxes represent hEGF sequence and the thin lines represent the gaps.



2.1. Experimental procedures

Solid-phase peptide synthesis. All peptides were synthesized manually by the stepwise solid phase approach (77). The synthesis started with N^a-tert-Boc-amino acid-OCH_2-Pam resin (0.7 - 0.8 mmol/g of resin, Applied Biosystem Inc.). The amino acid on the Pam resin was varied, depending on the C-terminal sequence of each peptide. The Boc group was chosen for the protection of the N^a-amino terminus. The protecting groups for side chains of amino acids were as follows: Asp(OBzl), Cys(4-MeBzl), Glu(OBzl), His(Dnp), Lys(2-ClZ), Ser(Bzl), Thr(Bzl), and Tyr(2-BrZ). The conventional deprotection of the Boc group by TFA and neutralization before the coupling step was modified with the following: 1. 0.05% ethanedithiol (EDT) and 0.05% of dimethylsulfide (DMS) were added into the TFA prewashing solution (50% TFA in CH_2Cl_2); 2. the resin was washed with DMF one time before using the DIEA neutralization solution. Double coupling was carried out for all syntheses. Boc-amino acid and coupling reagents used for each coupling were in three fold excess to the substitution sites on resin. The first coupling was mediated by DCC only. The second coupling was by symmetrical anhydride in CH_2Cl_2 : DMF (1:1, v/v) for most amino acids except Boc-Asn, Boc-Gln and Boc-Arg(Tos), for which HOBt coupling in DMF was used. Another exception was for Boc-Gly which was mediated by the DCC method only. Each amino acid coupling was found to be satisfactory by the quantitative ninhydrin test (78) before going on to the next step. Resin hydrolysis was

performed at the middle stage of each syntheses to monitor the progress of the coupling reactions. The synthetic protocols are listed in Table 2.1. Figure 2.3 shows amino acid activation and coupling reactions.

Table 2.1. Solid Phase Synthesis Protocol (for Boc-amino acid).

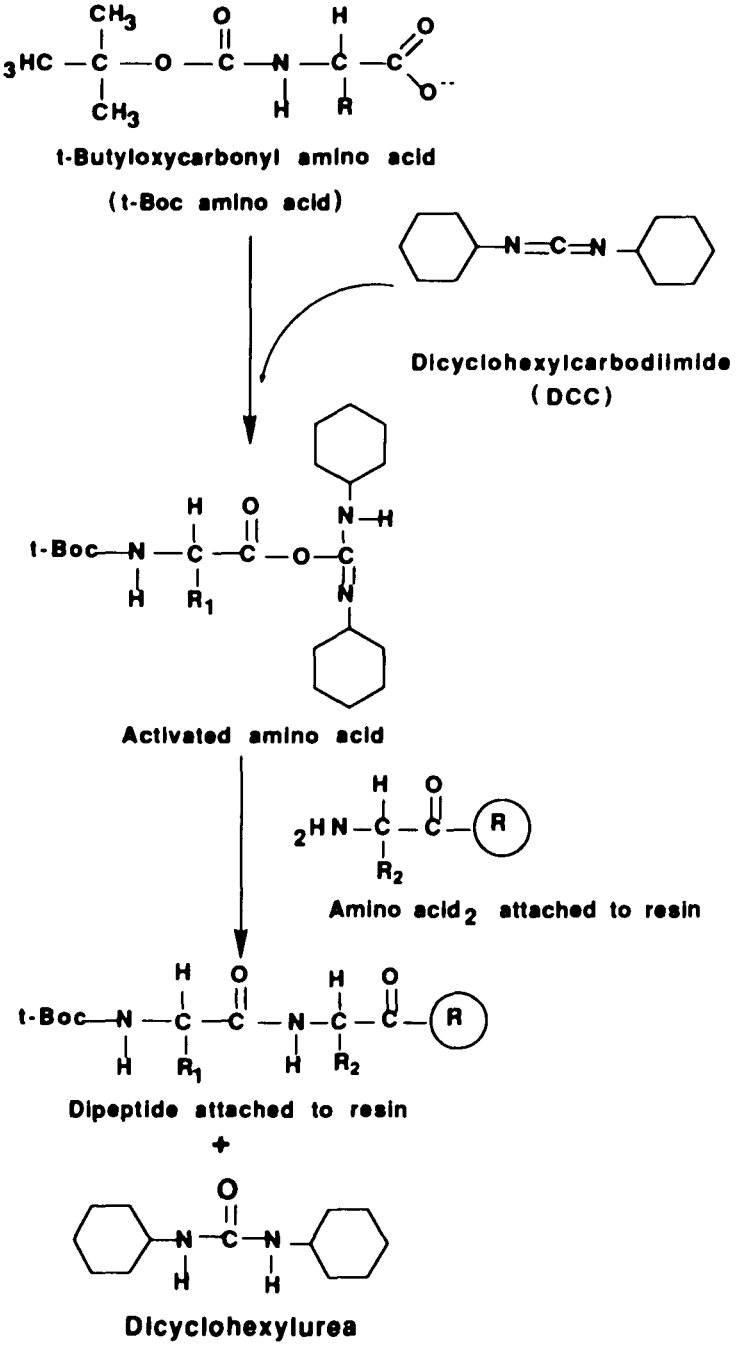
1. DCC coupling	
DeBoc:	
50% TFA/CH_2Cl_2 (+ 0.025% DMS, 0.025% EDT)	2 x 2 min
50% TFA/CH_2Cl_2	1 x 20 min
CH_2Cl_2	3 x 1 min
DMF	1 x 1 min
5% DIEA/DMF	3 x 2 min
CH_2Cl_2	3 x 1 min
Coupling:	
amino acid in CH_2Cl_2	0.5 min
DCC/CH_2Cl_2	60 min
Adding in 1/3 volume of DMF to the reaction mixture after 45 min coupling reaction.	
Washing:	
CH_2Cl_2	2 x 1 min
DMF	1 x 2 min
5 % DIEA/DMF	1 x 2 min
CH_2Cl_2	3 x 1 min
Ninhydrin test (78)	
Recoupling:	
Amino acid in CH_2Cl_2	1 min
DCC/CH_2Cl_2	60 min*

*: If the ninhydrin test indicated over 99.5% coupling, the second coupling time was shortened to 30 - 45 min.

2. HOBt coupling:

DeBoc:	
50% TFA/CH_2Cl_2 (+ 0.025% DMS, 0.025% EDT)	2 x 2 min
50% TFA/CH_2Cl_2	1 x 20 min
CH_2Cl_2	3 x 1 min
DMF	1 x 1 min
5% DIEA/DMF	3 x 2 min
CH_2Cl_2	3 x 1 min
Coupling:	
DMF	2 x 2 min
Amino acid in DMF	1 min (keep)
HOBt/DMF	1 min (keep)
DCC/CH_2Cl_2	2 h
Washing:	
DMF	2 x 2 min
5 % DIEA/DMF	1 x 2 min
CH_2Cl_2	3 x 1 min
Ninhydrin test	
Recoupling:	
DMF	2 x 2 min
Amino acid in DMF	1 min (keep)
HOBt/DMF	1 min (keep)
DCC/CH_2Cl_2	2 h
Washing:	
DMF	2 x 2 min
CH_2Cl_2	3 x 1 min

Figure 2.3. The dicyclohexylcarbodiimide (DCC) mediated amino acid coupling reaction are shown below.



Prior to the HF cleavage, the peptide-OCH_2Pam resin (300-500 mg) containing His(Dnp) residues was treated 5 to 6 times with about 15 to 20 mL of 1 M thiophenol in DMF for 8 h/time. It is an efficient way to remove Nim-Dnp protecting group from His. 50% TFA in CH_2Cl_2 (v/v) was used to remove the Boc group. After drying, the peptide-resin was ready for the HF cleavage.

HF cleavage: The low-high HF cleavage procedure (79) was carried out to cleave the peptide from the resin. In the low HF step, DMS, p-cresol, and p-thiocresol were transferred into the HF reaction vessel to premix with the peptide resin in the presence of 20 mg cysteine. The addition of cysteine was to minimize the side reactions of cysteinyl thiols in the peptide during the workup. 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 a final volume of 10 mL to give a reaction mixture in the following ratio: HF : DMS : p-cresol : p-thiocresol, 25 : 65 : 8 : 2 (v/v). The reaction was allowed to proceed for 2 h at 0 °C. After the removal of HF and DMS, the high HF reaction was initiated by transferring 14 mL liquid HF to yield the following reaction mixture ratio: HF : p-cresol : p-thiocresol, 94 : 5 : 1 (v/v). The reaction went for 1 h at 0 °C. HF was then evaporated. The presence of equimolar dimethyl sulfide and HF (3:1 by volume) changes the cleavage mechanism from the usual S_N1, which produces carbonium and nitronium ions, to S_N2. Dimethyl sulfide traps the initial protonated intermediate before separation of a discrete carbonium ion can occur (79). This new

procedure prevents alkylation of tyrosine by O-benzyl or O-dichlorobenzyl groups, formation of succinimide peptide from Asp-Gly sequences, and other possible side reactions caused by cleaved side-chain protecting groups. Para-cresol is recommended as a scavenger in this reaction. The addition of p-thiocresol in the reaction mixture is to cleave formyl groups from tryptophan residues.

The resulting crude peptide with resin containing aromatic scavengers and organic byproducts was extracted twice with cold ethyl ether/mercaptoethanol (98:2, v/v). The addition of mercaptoethanol was to prevent oxidation due to any peroxide contaminants. The crude peptide was then extracted into 100 mL of 8 M urea with 0.2 M dithiothreitol (DTT) in a 0.1 M Tris-HCl buffer at pH 8.2 - 8.5.

Refolding and Purification. A sequential dialysis was used as the first step in purification to remove the residual cresol in the crude peptide solution. The dialysis condition was as follows: Spectra por 6 dialysis tubing, molecular weight cutoff, 1000, against 2 L each of 0.1 M Tris-HCl (pH 8.2-8.5) buffer with decreasing urea concentration from 8 M, 5 M, 3 M to 1 M, at room temperature for 24 h. The buffers were kept oxygen free by vacuum degassing, and then N_2-purging to prevent premature disulfide formation. During the dialysis, no precipitation was observed. After dialysis, the peptide solution was diluted to 500 mL (about 0.1 mg peptide/mL) with 0.1 mM Tris-HCl (pH 8.2-8.5). Mixed disulfide refolding and air

oxidation methods were compared. The mixed disulfide method was initiated by incubation with 1 mM oxidized glutathione and 1 mM reduced glutathione in the diluted peptide solution (pH 8.2-8.5). The air oxidation method was started by simply allowing air to get into the diluted peptide solution. The solution was stirred slowly at room temperature. The refolding process was monitored by analytic HPLC at the following conditions: Shimadzu HPLC system on a VYDAC C_{18} reverse-phase (Separation Group) column, 4.6 mm x 25 cm; monitoring at UV_{225nm} and elusion with a 30 min. linear gradient of 15% to 85% of solution B at a flow rate of 1.5 mL/min. (solution A: 5% CH_3CN/0.045% TFA, v/v. solution B: 60% CH_3CN/0.030% TFA in aqueous solution, v/v).

The refolded peptide in 500 mL solution was purified by a preparative HPLC (Water system, VYDAC C_{18} reverse-phase column: 2.5 cm x 30 cm), at UV_{225nm} and flow rate of 15 mL/min. with a linear gradient of CH_3CN/0.04% TFA from 10.5% to 38% CH_3CN. A second purification was applied if the product was not satisfactory.

Characterization. The synthetic products were analyzed for their purity and integrity by the following methods: **1**. Analytic HPLC: The collected fraction from preparative HPLC was analyzed by analytic HPLC with similar conditions to those mentioned above. **2**. Amino acid analysis: Peptide samples were hydrolyzed in 5.7 N HCl at 110 °C for 24 h. The hydrolyzing tube was kept loose during the first couple of minutes after the HCl boiled, to allow the air out.

Analysis was carried out with Beckman System 6300. **3**. Mass spectrometry: The molecular mass of each peptide was identified by Cf-252 fission ionization mass spectrometry at The Rockefeller University Mass Spectrometry service center (80). **4**. Enzymatic digestion to determine disulfide linkages: Thermolysin was used for the enzymatic digestion to establish the disulfide linkages of the peptide products. Commercially purchased thermolysin (Kuhl und Trocken Lagern, Germany) was recrystallized 4 times in 0.02 M Ca-acetate buffer by the following procedures. Thermolysin crystals (about 30 mg) were suspended in 10 mL of cold 0.02 M Ca-acetate buffer, pH 7.0. The pH of the suspension was adjusted to 11.5 in an ice bath by dropwise addition of cold 0.2 N NaOH. The solution was centrifuged (3000 g, 2 min) to remove any precipitate in the suspension at this pH. The pH of the solution was then adjusted to 8.5 - 9.0 with cold 0.2 N CH_3COOH, added dropwise. Another centrifugation was carried out to clarify the solution. Afterward, the pH of the clear solution was adjusted back to 7.0 - 7.2 with cold 0.2 N CH_3COOH dropwise. Crystals appeared after standing in cold room (about 4 °C) overnight (81). The enzymatic digestion was performed in 0.1 mL of 0.1 M pyridine-acetic acid buffer at pH 6.5 and 45 °C for 18 hr. The ratio of the peptide to enzyme was 6 to 1 (mol/mol). The digested sample was filtered (Waters, pore size 0.22 μm) and separated on analytic HPLC (VYDAC C_{18} reverse-phase column, 4.6 mm x 25 cm; monitoring at UV_{225nm}). The flow rate was set at 0.5 mL/min. to get better separation. Each peak was collected, and hydrolyzed with 5.7 N HCl at 110 °C for 24 hr. The

amino acid content on each digested fragment was determined using an amino acid analyzer (Beckman System 6300).

Biological assays. The ability to compete with ^{125}I-EGF binding to EGF receptor of the synthetic products was tested on a subconfluent monolayer of formalin-fixed A-431 cells after 1 h incubation at 22 °C. A mitogenic assay was performed on normal kidney cells (clone 49F) by determining the amount of ^{3}H-thymidine incorporation (49).

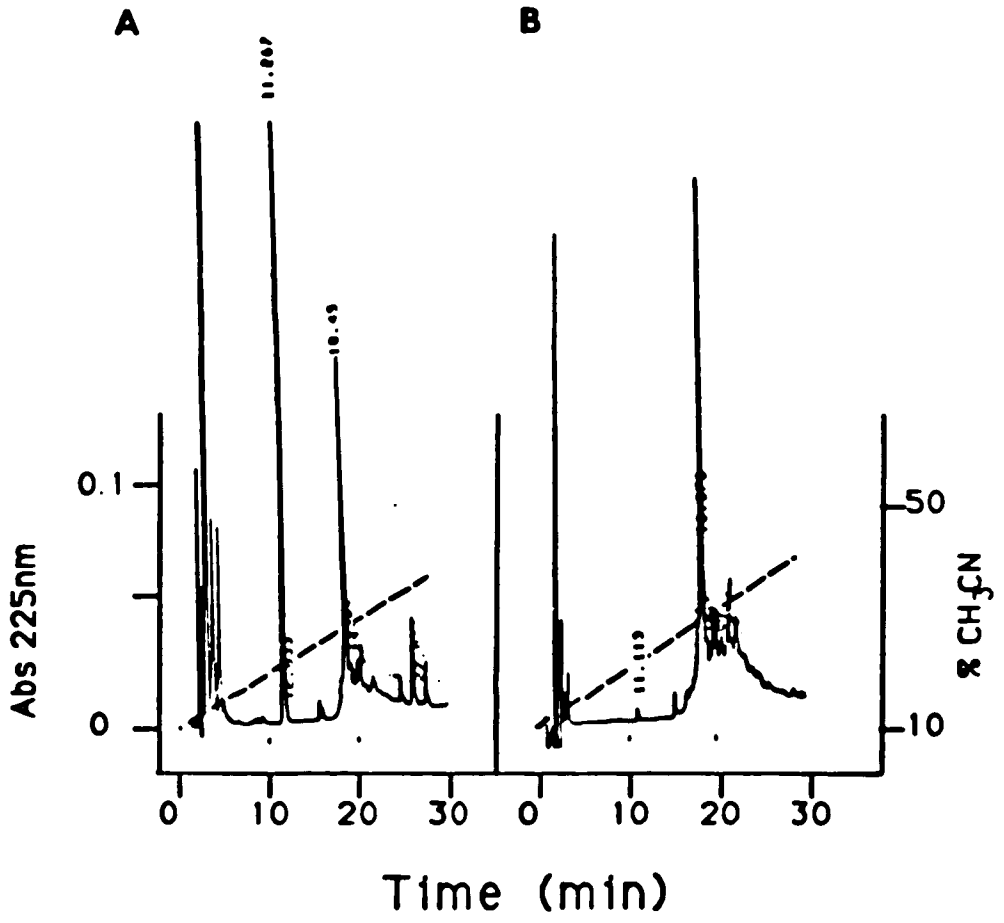
2.2. Results

Solid phase peptide synthesis and HF cleavage. During the syntheses, the major attention was to preserve the integrity of the thiol ether linkage of the six cysteinyl residues. Serious side reactions, such as modification of the cysteinyl side chain by S-tert-butylation during the TFA deprotection of the N^{a}-Boc group (82), as well as oxidation throughout the synthesis, could lead to predictable failure. To avoid the S-tert-butylation side reaction on cysteine residues, 0.05% EDT and 0.05% DMS were added as scavengers in the first 4 min of the prewashing step. Because the half-life of the cleavage of the Boc-group in TFA is shorter than 2 min, most Boc-groups will be removed after prewashing by TFA. The presence of the thiol scavengers in TFA for the 20-min deprotection cycle may not be desirable since they will accelerate the loss of benzyl side chain protecting groups on cysteine residues

(83). This compromise adjusted the use of thiol as a scavenger and suppressed the danger of losing benzyl side chain protecting groups. The synthesis yield was about 80%.

The low-high HF cleavage method gave a satisfactory result for all peptide-resin products. The low HF removed the benzyl-type protecting groups under a S_N2 condition which minimized the risk of alkylation reactions. The high HF was applied to remove those acid-resistant protecting groups, such as the MeBzl group on cysteinyl side chains, and to cleave peptide from its solid support (79). The cleavage yields were obtained from the resin hydrolysis after HF in the range of 90-94%. HPLC was followed after extraction of peptide from HF treatment. Figure 2.4 shows the HPLC profile of native FIX-EGF1 after HF cleavage extracted in 8 M urea solution, in which the major peptide peak and scavenger peak are seen; and the HPLC profile of native FIX-EGF1 after sequential dialysis, in which the scavenger peak at 11 min is almost vanished from sight.

Figure 2.4. C_{18} reverse-phase analytic HPLC profiles of FIX-EGF1. (A) After HF cleavage. (B) After the sequential dialysis with decreasing the concentrations of urea. It shows that most p-cresol (peak at elution time 11 min) from HF cleavage was removed after dialysis. The peptide in the dialysis solution was still in the reduced state, since the HPLC profile shows that the elution time of the peptide peak is kept the same as the one after HF with DTT at 18 min.



Refolding and purification. Refolding was conducted in a stepwise gradient of linearly lowered urea concentration under dialytic condition while DTT and other small impurities were removed. At 1 M urea concentration, the peptide solution was diluted into 500 mL with 1 mM reduced and oxidized glutathione (about 10 to 15 fold excess). The disulfide formation was monitored by HPLC. The reduced peak after HF was used as a reference to compare with the newly formed peak. After the refolding ended, the HPLC peak corresponding to the refolded peptide eluted 2 to 5 min earlier than the peak of reduced form. The time required for refolding varied from 1.5 h to 3-5 days. Figure 2.5. depicts the HPLC elution profile of native FIX-EGF1 and analog 10. In this figure the refolding results by mixed disulfide method or air oxidation are compared. It shows that the mixed disulfide refolding strategy gives better yields. The peptides were allowed to reach a thermodynamically controlled equilibrium. Various intermediates could be interconverted to the corrected products through constant opening and reforming disulfide bonds when reduced and oxidized glutathione are present. This is not as likely to occur in the air oxidation conditions. However, the refolding duration and the yield for those analogs were very different. Analogs 1, 6, 10, and 13-15 refolded within 2 to 5 h, and gave relatively high yield. Analogs 2, 3, 8, and 9 refolded in about 40 to 50 h, and gave lower yields. Analogs 4, 5, 11 and 12 ended the refolding in about 15 to

20 h, but they did not give a predominant refolded product. The duration for analog 7 lasted for about 5 days, and the yield was low.

Figure 2.5. C_{18} reverse-phase HPLC profiles of FIX-EGF1 and analog 10. The refolding methods of mixed disulfide-exchanging and air oxidation are compared. **A. 1.**: FIX-EGF1 refolded with reduced and oxidized glutathione. **A. 2.**: FIX-EGF1 refolded by air oxidation. **B. 1.**: analog 10 refolded with mixed disulfide method. **B. 2.**: the air oxidation product of analog 10.

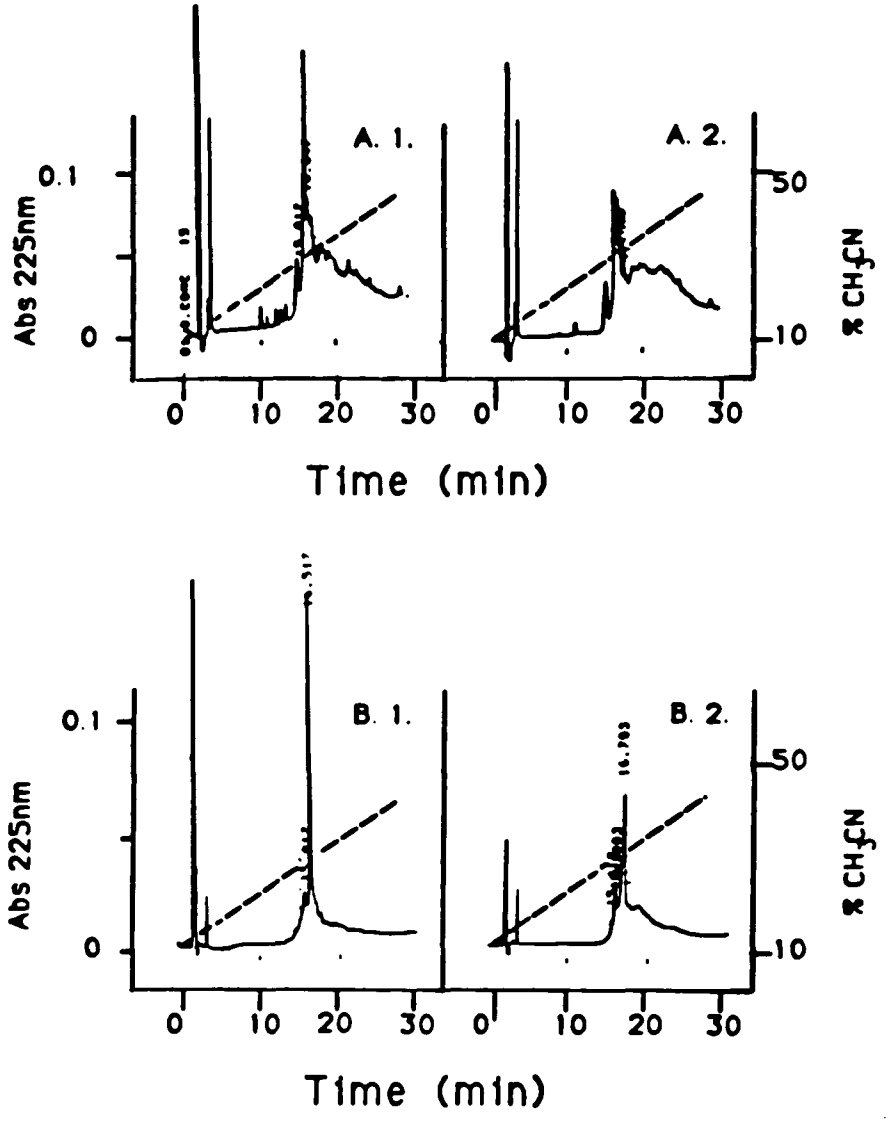


Figure 2.6. illustrates some refolded peptides before purification compared with purified peptides on HPLC. With the overall improvements in the synthesis, all synthetic products showed a distinct major peak by HPLC in the crude peptide solution. FIX-EGF1 and six analogs (1, 6, 10, 13, 14, and 15) required only a single preparative HPLC step to obtain a homogeneous product. However, the other nine analogs (2-5, 7-9, and 11-12) required two sequential HPLC (C_{18} reverse-phase) purifications to obtain homogeneous product. The overall yields of refolding and purification of synthetic peptides were in the range of 15 - 35 %. HPLC profiles of all purified synthetic products are shown in Figure 2.7.

Cf-252 fision ionization spectrometry yields precise information about the integrity of synthetic products. Peaks corresponding to $(M+H)^+$, $(M+2H)^{2+}$, $(M+3H)^{3+}$, or $(M+4H)^{4+}$, and $(M+5H)^{5+}$ were found for all peptides. No peak corresponding to $(2M+3H)^{3+}$ was found in any of the synthetic products. This excluded the possibility of dimmer formation in the refolded peptides. Table 2.2. gives mass spectrometry data for these synthetic peptides.

Figure 2.6. HPLC analysis on a C18 reverse-phase column of crude and purified FIX-EGF1 and its analogs. A. 1., B. 1., C. 1., and D. 1. are crude products of FIX-EGF1, analog 6, analog 5, and analog 15 after disulfide refolding, respectively. A. 2., B. 2., C. 2., and D. 2 are purified products of FIX-EGF1, analog 6, analog 5, and analog 15, respectively.

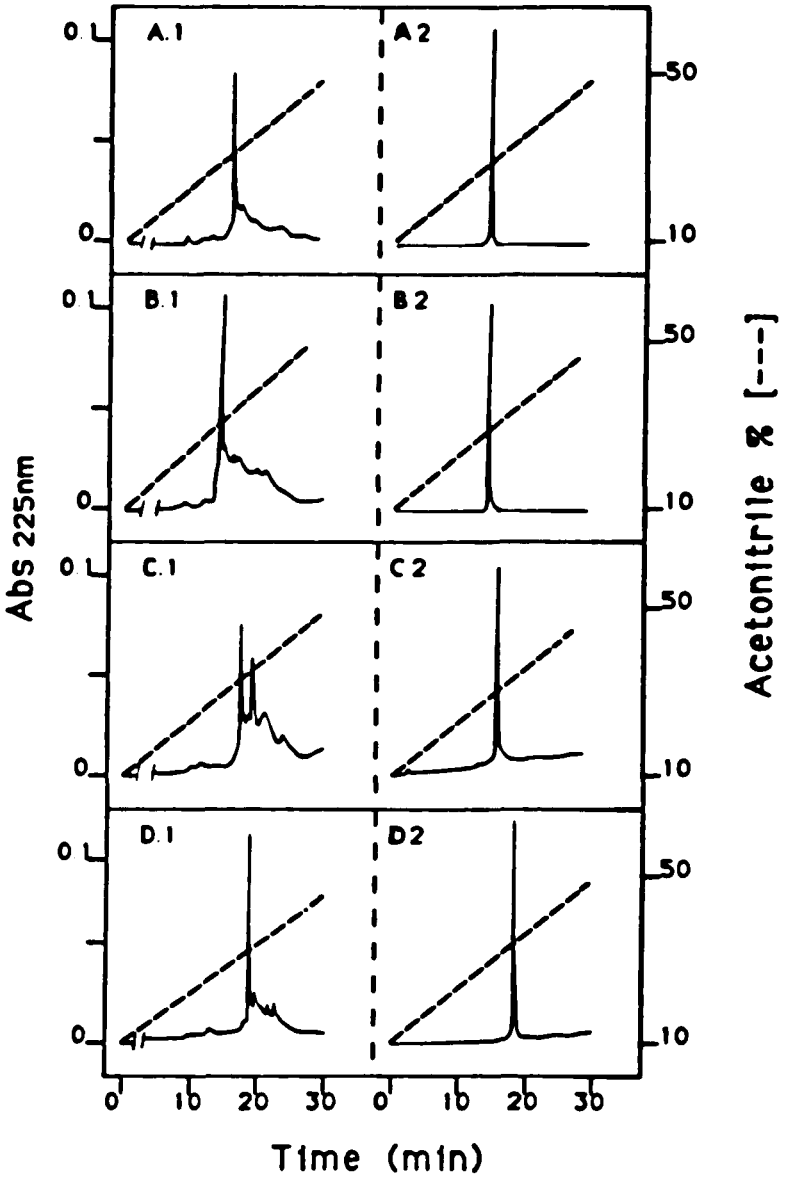
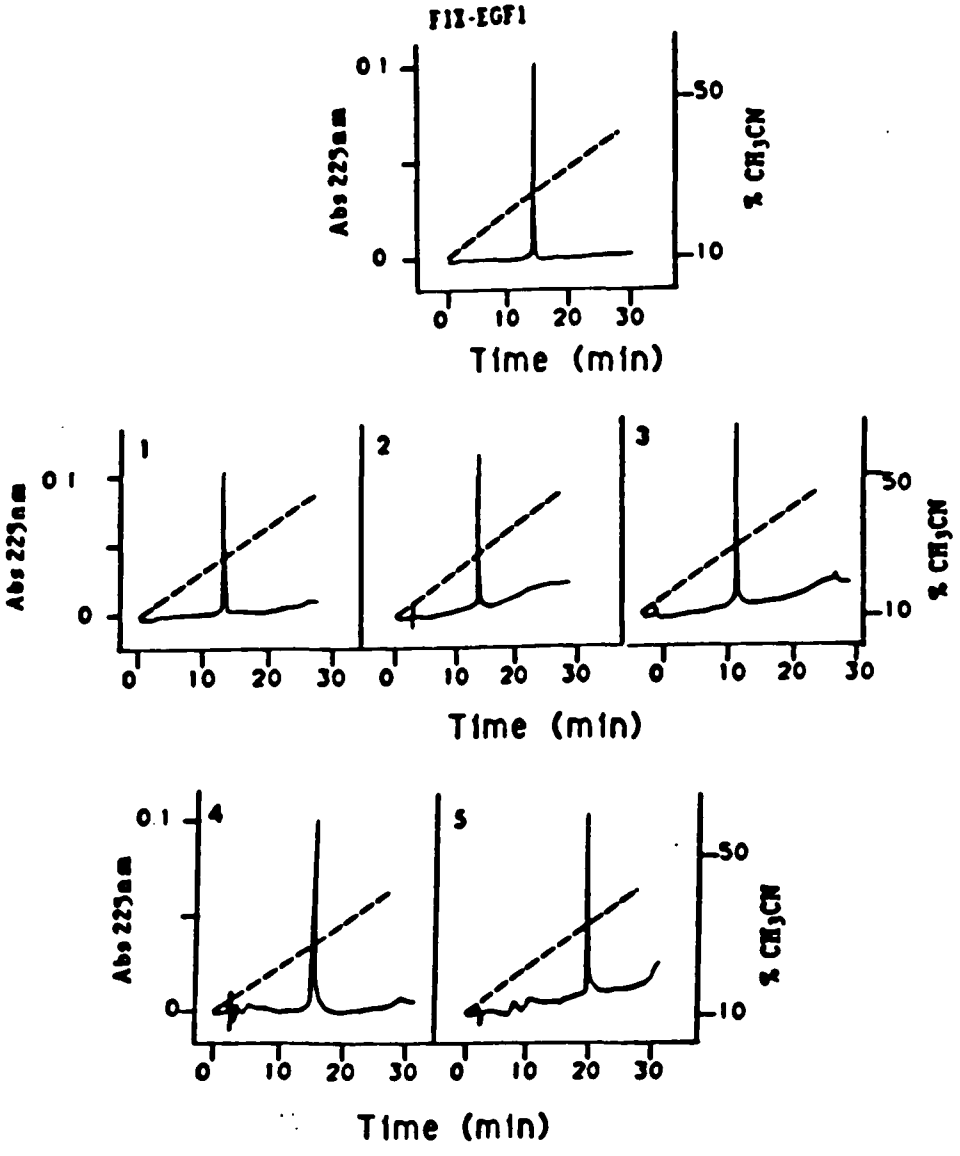

A.1
A2
B.1
B2
C.1
C2
D.1
D2
0.1
0
50
10
Abs 225nm
Acetonitrile % [---]
0
10
20
30
Time (min)

Figure 2.7. HPLC profiles of all purified synthetic products. The analogs are labeled by numbers. The HPLC condition are stated in the experimental section above.



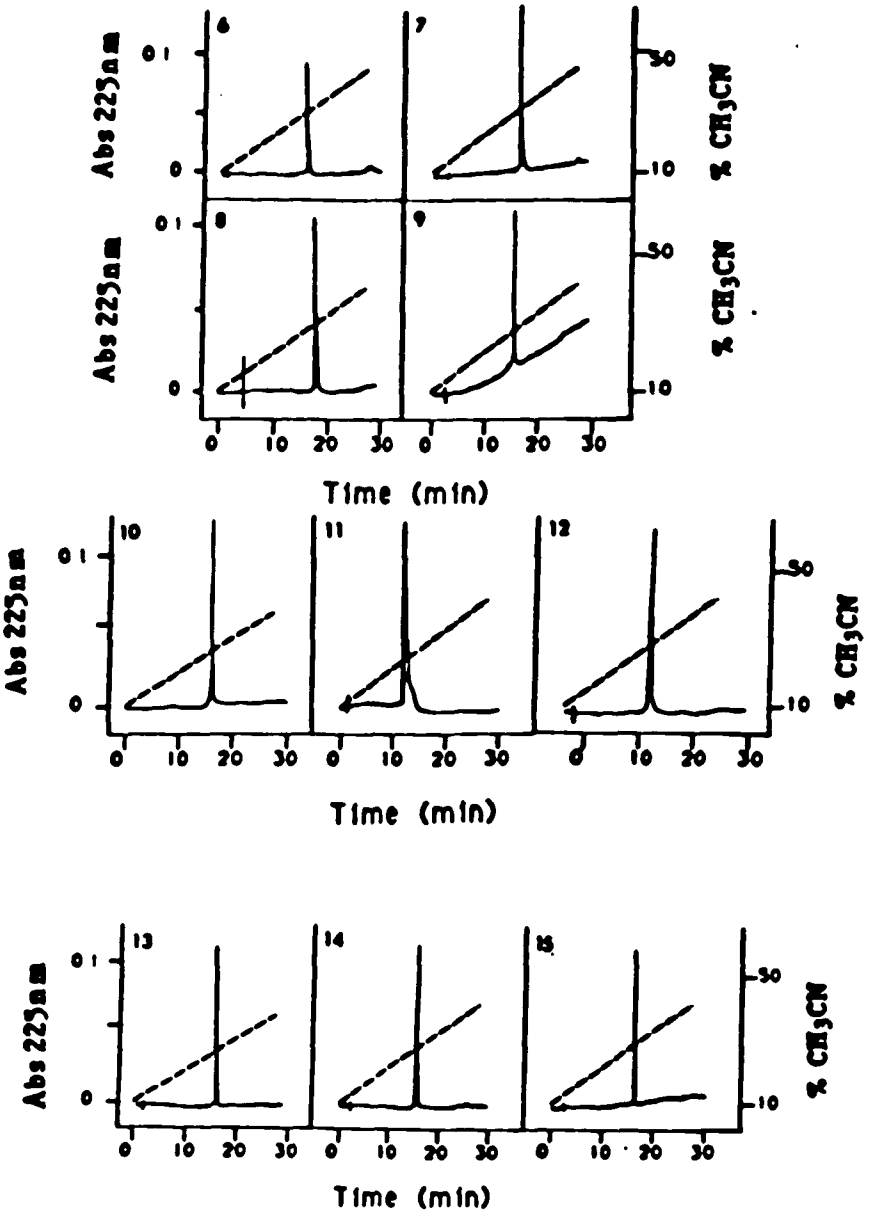

Abs 225nm
% CH3CN
Time (min)
Abs 225nm
% CH3CN
Time (min)
Abs 225nm
% CH3CN
Time (min)

Table 2.2. Molecular weight determination of FIX-EGF1, 2 and analogous by Cf-252 fission ionization mass spectrometry.

Peptide	Ion species	Mea. M.W.	Calc. M.W.	Δ(M-C)
FIX-EGF1	$(M+H)^+$	4750.6	4749.9	+0.7
	$(M+2H)^{2+}$	4749.8	4749.9	-0.1
FIX-EGF2	$(M+H)^+$	4963.0	4963.2	-0.2
	$(M+2H)^{2+}$	4963.2	4963.2	0
Analog 1	$(M+H)^+$	5093.7	5092.7	+1.1
	$(M+2H)^{2+}$	5093.4	5092.7	+0.2
Analog 2	$(M+H)^+$	5115.8	5115.6	+0.2
	$(M+2H)^{2+}$	5115.8	5115.6	+0.2
Analog 3	$(M+H)^+$	4971.0	4970.5	-0.5
	$(M+2H)^{2+}$	4969.8	4970.5	-0.7
	$(M+3H)^{3+}$	4968.9	4970.5	-1.6
Analog 4	$(M+H)^+$	4789.9	4792.2	-2.3
	$(M+2H)^{2+}$	4791.6	4792.2	-0.6
Analog 5	$(M+H)^+$	4747.7	4747.0	+0.7
	$(M+2H)^{2+}$	4747.2	4747.0	+0.2
Analog 6	$(M+H)^+$	5148.0	5148.6	-0.6
	$(M+2H)^{2+}$	5148.4	5148.6	-0.2
Analog 7	$(M+H)^+$	5450.3	5451.9	-1.6
	$(M+2H)^{2+}$	5450.0	5451.9	-1.9
	$(M+3H)^{3+}$	5452.2	5451.9	+0.3
Analog 8	$(M+H)^+$	5419.6	5419.9	-0.3
	$(M+2H)^{2+}$	5419.2	5419.9	-0.7
	$(M+3H)^{3+}$	5420.1	5419.9	+0.2
Analog 9	$(M+H)^+$	5322.8	5322.1	+0.7
	$(M+2H)^{2+}$	5321.2	5322.1	-0.9
Analog 10	$(M+4H)^{4+}$	5117.8	5117.9	-0.9
	$(M+5H)^{5+}$	5118.3	5117.9	-0.4
Analog 11	$(M+H)^+$	5345.6	5346.9	-1.3
	$(M+2H)^{2+}$	5344.6	5346.9	-2.3
Analog 12	$(M+H)^+$	5363.9	5365.0	-1.1
	$(M+2H)^{2+}$	5364.4	5365.0	-0.6
	$(M+3H)^{3+}$	5360.1	5365.0	-4.9
Analog 13	$(M+3H)^{3+}$	5068.5	5068.7	-0.2
	$(M+4H)^{4+}$	5067.6	5068.7	-1.1
	$(M+Na+3H)^{4+}$	5068.4	5068.7	-0.3
Analog 14	$(M+4H)^{4+}$	5388.8	5389.3	+0.5
	$(M+5H)^{5+}$	5389.3	5389.3	0
Analog 15	$(M+4H)^{4+}$	5430.8	5432.0	-1.2
	$(M+5H)^{5+}$	5431.3	5432.0	-0.7

The results from enzymatic digestion provided evidence of correct disulfide pairing for some peptides. Some were not available due to a failed digestion. Thermolysin hydrolyzes peptide bonds between hydrophobic amino acids. The reason in failure cases could be due to the peptides with higher hydrophobicity. After digestion the fragments were too small to be analyzed. It was seen in the elution profile of digested sample, a big peak came out in first 1-3 min. Some sample data from successful digestion were listed in Table 2.3.

Table 2.3. Disulfide bond location of synthetic products by enzymatic digestion method(1).

1. Results of FIX-EGF1:

Fragment 11		Corresponding section
amino acid ratio		
Cys	2.00	
(2)Asx	1.75	Asp^{3}-Gly-Asp-Gln-Cys^{7}-Glu^{8}
Ser	0.66	|
(2)Glx	1.29	Cys^{18}-Ser-Gly^{16}
Gly	1.66	

Fragment 7		Corresponding section
Cys	2.00	
(2)Asx	1.44	Cys^{29}-Pro^{30}-Phe^{31}-Gly^{32}
(2)Glx	2.19	|
Pro	0.55	Glu^{39}-Cys^{38}-Asn-Lys-Gly-Glu^{34}
Gly	1.67	
Lys	1.23	
Phe	0.94	

2. Result of analog 2:

Fragment 19		Corresponding section
amino acid ratio		
Cys	2.00	
(2)Asx	1.60	Cys^{32}-Pro-Phe-Gly^{35}
(2)Glx	2.27	|
Pro	1.00	Cys^{41}-Asn-Lys-Gly-Glu^{37}
Gly	1.70	|
Phe	0.82	Glu^{42}
Lys	1.00	

Fragment 31		Corresponding section
Cys	2.00	
(2)Asx	2.62	Asp-Gln-Cys^{7}-Pro-Leu
Ser	0.79	|
(2)Glx	2.11	Cys^{21}-Ser-Gly-Gly^{18}
Pro	0.95	|
Gly	1.92	Lys^{22}-Asp-Asp^{24}
Leu	1.00	
Lys	1.21	

3. Results of analog 3:

Fragment 9 Amino acid ratio		Corresponding section
Cys	2.00	
(2)Asx	3.00	Asp^{3}-Ser-Glu-Cys^{6}-Pro-Leu^{8}
Ser	2.06	|
(2)Glx	1.85	Cys^{20}-Ser-Gly-Gly^{17}
Pro	1.13	|
Gly	1.81	Lys-Asp-Asp^{23}
Leu	1.02	
Lys	0.82	
Cys	2.00	
(2)Asx	1.46	Cys^{31}-Pro^{32}
(2)Glx	1.85	|
Pro	0.98	Cys^{40}-Asn-Lys-Gly-Glu^{36}
Gly	1.24	|
Lys	1.00	Glu^{41}

4. Results of analog 6.

Fragment 9 Amino acid ratio		Corresponding section
Cys	2.00	
(2)Asx	4.00	Gln-Cys^{7}-Glu-Ser^{9}-Asn-Pro^{11}
Ser	1.42	|
(2)Glx	2.20	Cys^{18}-Ser-Gly-Gly-Asn-Leu^{13}
Pro	0.78	|
Gly	2.46	Lys^{19}-Asp-Asp^{21}
Leu	0.80	
Lys	1.03	
Cys	2.00	
(2)Glx	1.14	Cys^{29}-Pro-Phe-Gly^{32}
Pro	0.82	|
Gly	0.96	Cys^{38}-Arg-Lys^{36}
Phe	0.61	|
Lys	1.00	Glu^{39}
Arg	1.05	

(1). The number of amino acid used here are according to the amino acid position in the synthetic products.

(2). Asx includes Asp and Asn; Glx includes Gln and Glu.

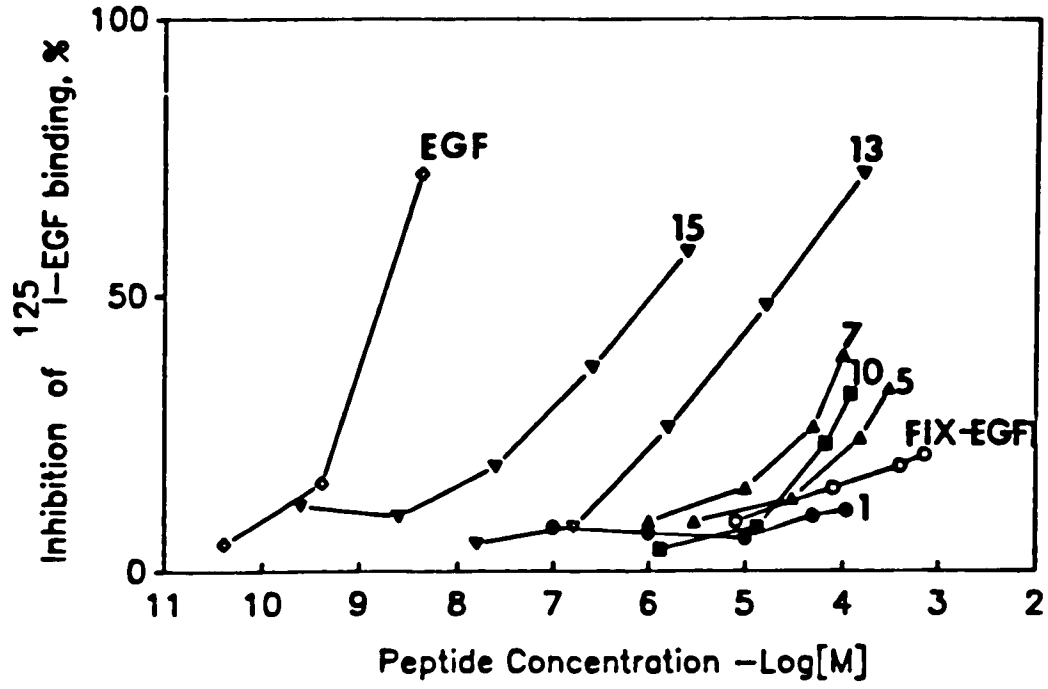
Biological assays. The putative EGF-like activities were tested. EGF-receptor binding competitive activity of the synthetic peptides were compared and results were listed in table 4. Some analogs could bind to EGF receptor at high concentration. No synthetic analogous showed more than slight EGF-like activity. The results of inhibition of ^{125}I-EGF binding (%) from each group of analogs are depicted in Figure 2.8.

Table 2.4. EGF-receptor binding competitive activity of synthetic FIX-EGF1 and analogous.

Peptide	IC_{50}	Relative to hEGF(%)
FIX-EGF1	N.A	N.A
Analog 1	N.A	N.A
Analog 2	N.A	N.A
Analog 3	N.A	N.A
Analog 4	N.A	N.A
Analog 5	N.A	N.A
Analog 6	N.A	N.A
Analog 7	$> 1x10^{-3}$	$1.5x10^{-5}$
Analog 8	$> 1x10^{-3}$	$1.5x10^{-5}$
Analog 9	$2x10^{-4}$	$8.3x10^{-4}$
Analog 10	$3x10^{-4}$	$2x10^{-4}$
Analog 11	$1.5x10^{-3}$	$1x10^{-4}$
Analog 12	$5x10^{-4}$	$3.2x10^{-4}$
Analog 13	$2.8x10^{-5}$	$5.5x10^{-3}$
Analog 14	$2.7x10^{-6}$	$5.6x10^{-2}$
Analog 15	$1x10^{-6}$	$1.7x10^{-1}$
hEGF	$1.6x10^{-9}$	100

N.A: Activity not detectable at peptide concentration higher than $1x10^{-3}$ M.

Figure 2.8. EGF-receptor binding inhibition assay results of FIX-EGF1 and its analog with A431 cell. Open circle represents FIX-EGF1, filled circle represents analog 1 (group 1), open triangle represents analog 5 (group 2), filled triangle represents analog 7 (group 3), filled sequare represents analog 10 (group 3), open inverse triangle represents analog 13 (group 4), filled inverse triangle represents analog 15 (group 4), and open diamond represents EGF.



2.3. Discussion

Structure and EGF-like activity correlation. The solution structure of hEGF and mEGF have been determined recently by NMR (84-90). The structure of EGF is highly rigid and compact since the presence of three disulfide bonds, and two b-sheet structures. The A-loop and C-loop of EGF are close together on one side of the molecule, and the B-loop and part of NH-terminal tail face on another side of the molecule. The conserved amino acids in EGF and its closely related molecules are located on the A-loop and C-loop side, which is considered as a receptor contact side.

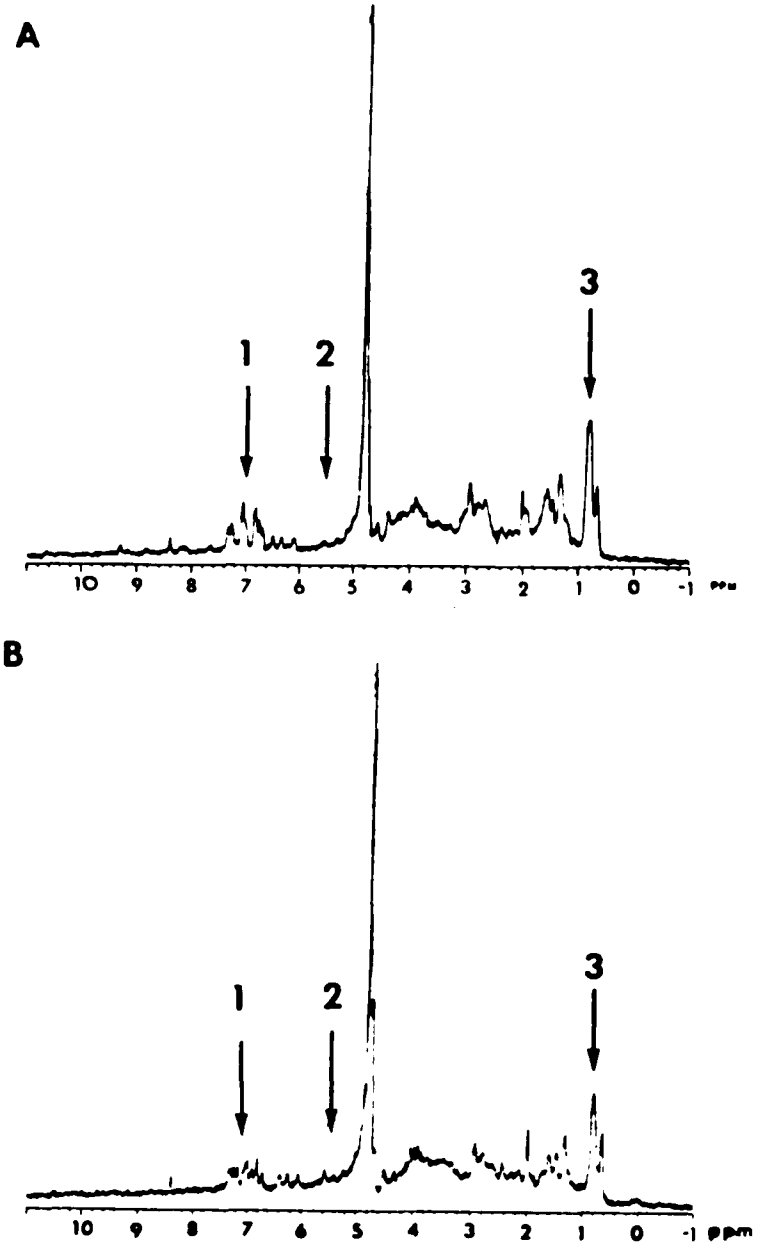
The group 1 chimeric analogs with EGF A loop replacement did not show any EGF-like activity. The group 2 analogs with EGF A and B loop replacement did not show any EGF-like activity. In the group 3 analogs 7 to 12 showed weak EGF-like activity at high peptide concentration. In analog 11 and 12 both A and C loops were replaced with EGF sequences, but no EGF-like activity was observed.

One possible explanation for low EGF-like activity could be that the tertiary structure may be very sensitive to sequence replacement. The native conformation, which is required for receptor binding, could be disrupted in all analogs. Another possible reason could be that the sequence changes make it difficult for the chimeric analogs to refold properly, perhaps yielding products with improper disulfide pairing. With the exception of analogs 2, 3, and

6, there is no direct evidence on the location of disulfide bonds. However, the relatively high yields observed during refolding of analogs 10, 13, 14, 15, suggests that they are likely to be properly folded. In addition, two group 4 analogs (14, and 15) , which did show some EGF-like activity, have been examined using 1D proton NMR. The 1D NMR spectra showed indications of a structured peptide: 1. chemical shift dispersion in aromatic proton region; 2. downfield-shifted alpha proton peaks; 3. resolved methyl proton region. Figure 2.9. shows the 1D NMR spectra for these two peptides.

These results suggest that EGF-like activity is a sensitive function of amino acid sequence. A more detailed explanation might be provided by NMR conformational analysis of these synthetic analogs.

Figure 2.9. 1D ^{1}H NMR spectra of (**A**) analog 14 and (**B**) analog 15. The arrow pointed region: **1**. chemical dispersion for aromatic protons; **2**. downfield-shifted alpha protons; and **3**. resolved methyl protons.



Chapter Three

3. Solution structure determination by NMR spectroscopy and refined by energy minimization with restraints

The goal of this NMR study is to determine the solution structure of FIX-EGF1, to locate the calcium binding region in this module, and to examine structural differences between EGF and FIX-EGF1. The solution structure of FIX-EGF1, and the comparison of the solution structures of FIX-EGF1 and EGF will be discussed in Section 3.3.

3.1. Sequence-specific assignment and secondary structure determination

In this section, the complete sequence specific resonance assignment of the ^1H NMR spectrum (at pH 4.2) of the Ca^{2+}-free FIX-EGF1, and its secondary structure will be reported. The effect of Ca^{2+} binding on the peptide conformation will be discussed.

3.1.1. Materials and Methods

Peptide Sample. FIX-EGF1 (45-87) was synthesized chemically as described in the second chapter of this thesis.

1D and 2D-NMR Experimental Conditions: About 6 mg purified peptide was dissolved in 0.6 mL of buffer containing 10 mM deuterated sodium succinate at pD 4.2. The pD values are the

uncorrected pH meter readings. For all NMR experiments in D_2O, the sample was lyophilized and redissoved in 99.8% D_2O (Stohler Isotope) a few times. The final NMR sample was prepared with 99.996% D_2O (Stohler Isotope). For the NMR experiments in H_2O, the peptide solution was lyophilized, and redissolved in H_2O with 10% D_2O as an internal lock. For determination of the slowly exchanging amide protons, a peptide solution in H_2O was lyophilized and redissolved in D_2O, and a COSY spectrum was immediately acquired. All experiments were thermostatted to 25 $\pm$ 0.2°C. A Varian VXR-500S spectrometer operating at 499.84 MHz for ^1H was used for all NMR experiments. All 2D spectra were recorded in the phase-sensitive absorption mode using the hypercomplex method (91). The carrier frequency was set on the water resonance with a sweep width of 7000 Hz in both dimensions, except for the triple-quantum experiment which had a sweep width of 10500 Hz in the t_1 dimension. The 2D data were collected as follows: 2048 complex points in t_2, 300-350 complex FIDs in t_1, 32-64 transients for each FID with a recycle time of 1.3-1.9 s. In all the spectra, low-power continuous-wave irradiation was applied to the water frequency in order to suppress the residual water signal. The 2D NMR experiments, DQF-COSY, (92-93), TOCSY (94-95), NOESY (96) and relayed-COSY (97-98) were performed in 90% H_2O/10% D_2O; triple quantum COSY (TQ-COSY) (99-100), DQF-COSY, TOCSY, NOESY and relayed-COSY were carried out in D_2O according to standard procedures (68). The NOESY spectra were acquired with mixing times of 1, 40, 80, 120, and 200 ms, and the

TOCSY spectra were performed with a spin lock time of 70 ms. The 2D NMR data were processed at a Vax Station 3100 or SUN4 computer using the FTNMR or FELIX programs (Hare Research, Inc.).

Resonance Assignment Strategy: Resonance assignment was performed by using the standard sequential assignment procedure (68): elucidation of type-specific assignment was accomplished by establishing through-bond connectivities among protons with the help of phase-sensitive DQF-COSY, TOCSY, relayed-COSY and TQ-COSY.

The first step in the assignment procedure is to identify resonances belonging to the same amino acid residue. Because no coupling between protons can be observed through a peptide bond, amino acid residues can be identified as individual spin systems. The information can be obtained from a variety of experiments that probe through-bond J-coupling connectivities between protons in the same spin system. The pattern of these connectivities is then used to classify the spin system to a specific amino acid type, or to a class of amino acid types. Along with J-coupling information, through space NOESY information allows identification of certain amino acid types. In the case of aromatic amino acid assignments, NOESY data were used to link the backbone and C^{β} protons in an individual spin system to their aromatic protons. When a sure identification of amino acid type, or a class of amino acid types, has been made, NOESY cross-peaks are used to connect one amino acid to next amino acid residue in an amino acid sequence. The majority

of the short distances between neighboring residues are distances of an amide proton with its neighboring amide proton, and C^{α} proton or C^{β} proton with its neighboring amide proton. The notations used often to indicate these distances are the followings: $d_{NN} = d(NH_i, NH_{i+1})$, $d_{\alpha N} = d(C^{\alpha}H_i, NH_{i+1})$, $d_{\beta N} = d(C^{\beta}H_i, NH_{i+1})$.

The information on amino acid type, along with nuclear Overhauser effects (NOEs) between neighboring residues, is then combined with the known amino acid sequence to give sequential resonance assignment.

Secondary structure determination. The extended secondary structure of an antiparallel β-sheet in the peptide is determined based on information from: (1) strong sequential H^{α}_i-H^{N}_{i+1} NOEs, and cross-strand NOEs between alpha-alpha, amide-amide, and alpha-amide protons; (2) large vicinal ^1H-^1H spin-spin coupling constants ($^3J_{HN\alpha}$); and (3) a pattern of slowly exchanging protons (68).

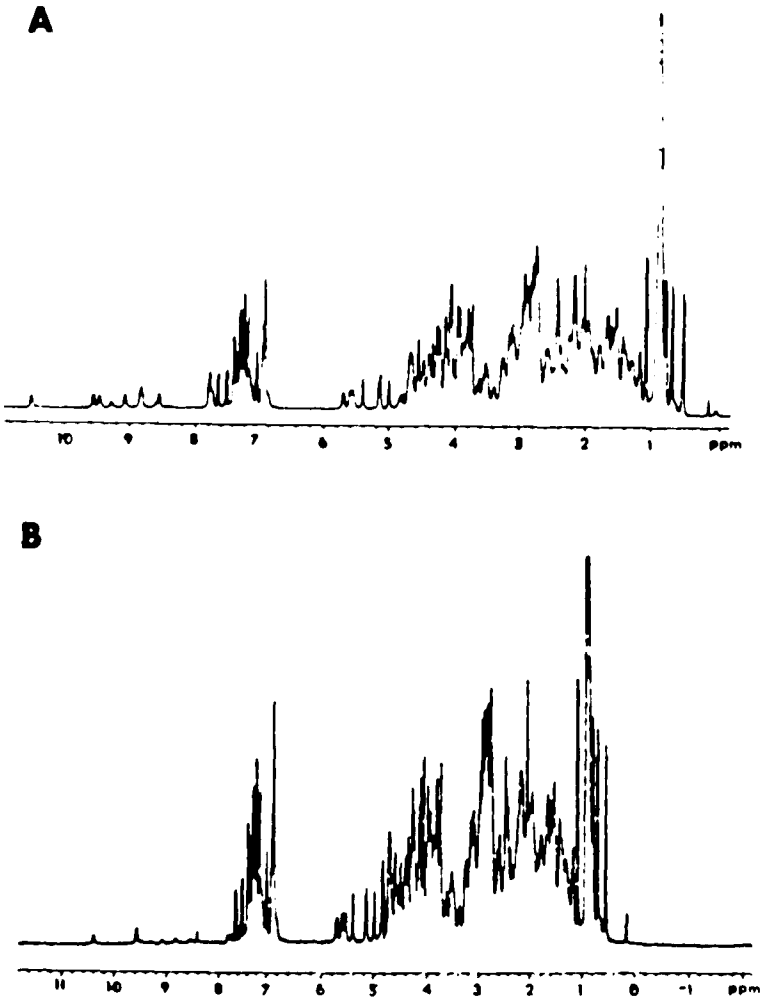
As a result of an extended structure, intraresidue NOEs are very weak or undetectable. The amide protons in the sheet are exchanging slowly with solvent due to the formation of the hydrogen bonds with the interstrand carbonyl oxygens. Vicinal ^1H-^1H spin-spin coupling constants have been applied to determine dihedral angles for studying peptide conformation. The most useful one is the coupling between $C^{\alpha}H$ and NH protons ($^3J_{HN\alpha}$) as determination of ϕ. A secondary structure of the peptide backbone

is determined from the short distances of d_{NN} and $d_{\alpha N}$ combined with $^3J_{HN\alpha}$ coupling constants (68). $^3J_{HN\alpha}$ coupling constants were estimated from ω2 cross-sections of H^N-H^α cross-peaks with DQF-COSY spectrum. Most H^N-H^α cross-peaks of FIX-EGF1 have line width of ca. 7.0 Hz at 25 °C, and the smallest apparent values of were ca. 4.5 Hz. Thus, only cross-peaks with apparent couplings ≥ 8.0 Hz were used in the structure determination.

3.1.2. Results

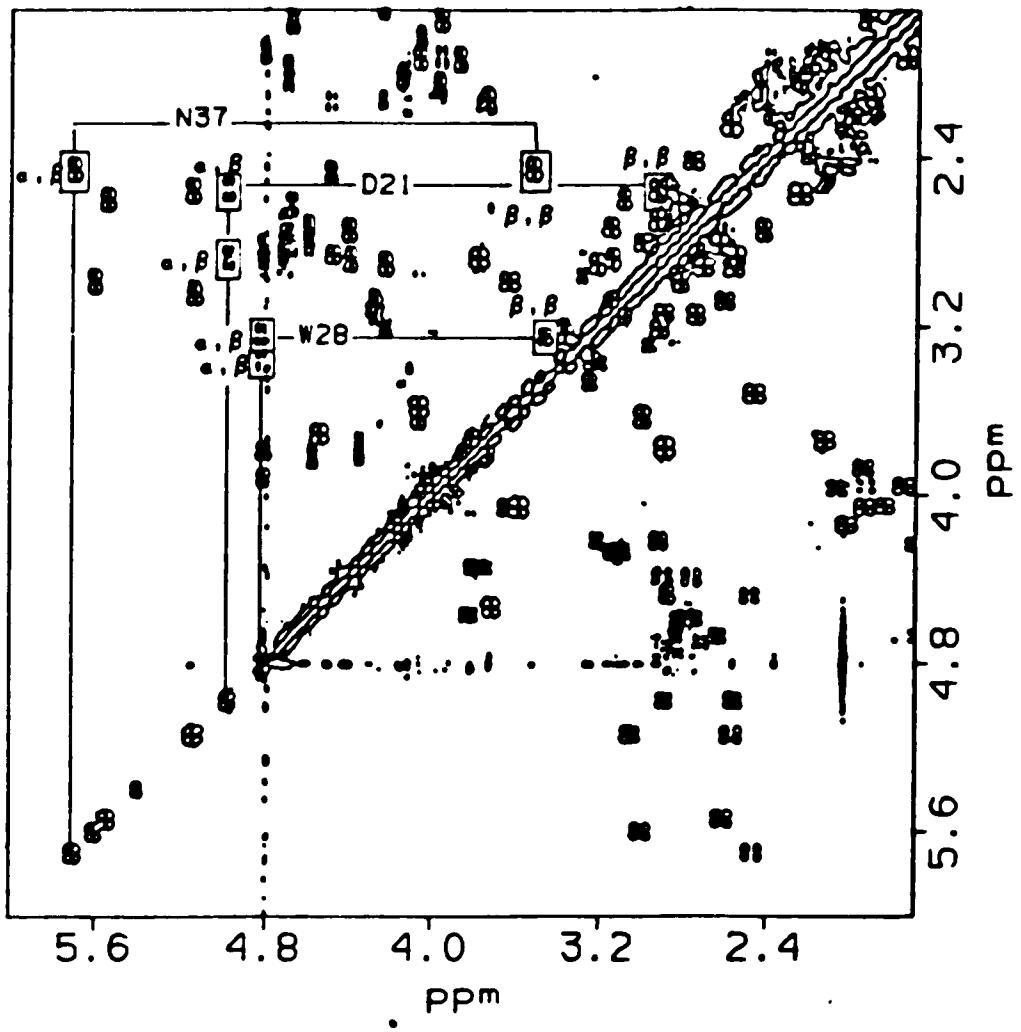
1D NMR. The 1D NMR spectrum of FIX-EGF1 is shown in Figure 3.1. The amide proton region is from 7.6 to 10.4 ppm. The aromatic protons are in the region of 6.8 - 7.5 ppm. The aliphatic proton region, including alpha protons, methylene protons and methyl protons, is approximately between 0.5 - 5.70 ppm.

Figure 3.1. 1D ^1H NMR spectra of FIX-EGF1 at different pH values. The spectra were collected on a Varian VXR-500S spectrometer. (A) shows 1D ^1H NMR spectrum at pH 4.7 in D_2O (25 °C) in which the slowly exchanging amide protons are observed in the region 7.6 - 10.4 ppm. (B) shows 1D ^1H NMR spectrum at pH 4.2 in D_2O (25 °C). This spectrum was acquired after the full DQF-COSY experiment, in which most slowly exchanging amide protons have vanished.



Spin system assignments and identification of amino acid type. There are several amino acid types with unique spin systems in FIX-EGF1 (45-87 in human factor IX): five Glu/Gln residues, five Gly residues, two Pro residues, one Ile residue, two Leu residues, one Thr residue, two Val residues, and two Lys residues. It also contains AMX spin systems (68) consisting of six Cys residues, three Ser residues, two Phe residues, two Tyr residues, five Asp residues, four Asn residues and one Trp residue. The identification of amino acid type was based on the data from experiments of 2D DQF-COSY, TOCSY in H_2O and D_2O, TQ-COSY in D_2O. The DQF-COSY experiment gives connectivities between directly coupled protons separated by two or three bonds. For example, $C^{\alpha}H_i/C^{\beta}H_i$, or $C^{\beta}H_i/C^{\beta'}H_i$ will show a cross-peak in a DQF-COSY spectrum, but it will not show cross-peak for $NH_i/C^{\beta}H_i$. Figure 3.2. shows a phase sensitive contour plot of the aliphatic proton region of the DQF-COSY spectrum. The proton-proton J-coupling connectivities of Asp^{21}, Trp^{28}, and Asn^{37} are indicated in this figure.

Figure 3.2. A phase-sensitive contour plot of the aliphatic proton region of the DQF-COSY spectrum (25 °C, pH 4.2). The α - β proton, and β–β proton J-coupling connectivities of Asp^{21}, Trp^{28}, and Asn^{37} are indicated.



For amino acids with a larger spin system than an AMX system, techniques for relaying magnetization between protons that are not directly coupled are extremely useful. The relayed-COSY experiment performs a single relay of magnetization beyond the directly coupled protons. Figure 3.3. shows a portion of a contour plot of a relayed-COSY spectrum in H_2O. Besides the coupling between NH_i-$C^{\alpha}H_i$, the connection between NH_i-$C^{\beta}H_i$ was also obtained. It shows a single relay couple beyond the direct proton-proton couple for the residues of Val^2, Asp^3, Ser^9, Asn^{14}, Ile^{22}, Phe^{33} and Val^{42}.

The TOCSY experiment can connect almost all side chain protons together in the same spin system. Figure 3.4. shows a contour plot of a TOCSY spectrum in H_2O. Figure 3.5. shows a contour plot portion of the TOCSY spectrum in D_2O. In this figure, the side chain connections with $C^{a}H$ for Leu^{40}, Lys^{19}, Glu^{26}, Pro^{30}, Leu^{13}, Pro^{11}, Val^2, Glu^{39}, Val^{42}, Gln^6, Glu^8, Lys^{36}, Asn^{14}, and Ile^{22} are indicated.

Difficulty arises when the resonances overlap in the spin system. A powerful method to help solve this problem is to use multiple quantum spectroscopy. Some overlapping resonances were resolved by acquisition of a 2D triple-quantum COSY experiment. This experiment selects for three-quantum coherence during the evolution time that is then transferred to single-quantum coherence on the individual spins. In a TQ-COSY Figure

Figure 3.3. A contour plot portion of the relayed-COSY spectrum in H_2O (25 °C, pH 4.2). The coupling between NH_i-$C^{\alpha}H_i$, and NH_i-$C^{\beta}H_i$ are observed in the region of the spectrum shown. A single relay couple beyond the direct proton-proton couple for the residues of Val^2, Asp^3, Ser^9, Asn^{14}, Ile^{22}, Phe^{33} and Val^{42} are depicted.

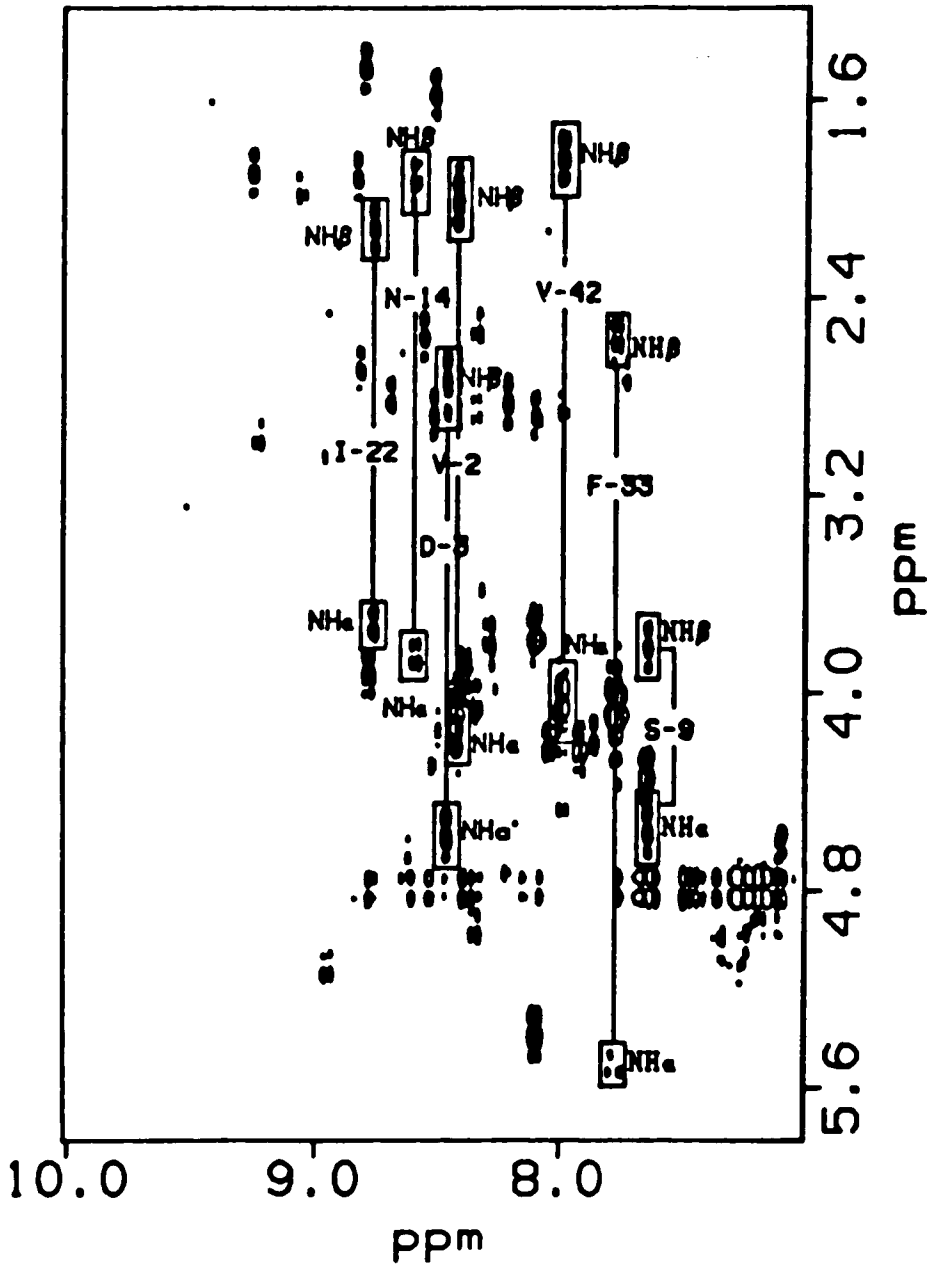


Figure 3.4. A contour plot portion of the TOCSY spectrum in H_2O (25 °C, pH 4.2). The spin system connectivities for the following are shown: NH, $C^{\alpha}H$, $C^{\beta}H$ and two $C^{\gamma}Hs$ of Val^{42}; NH and two $C^{\beta}Hs$ of Trp^{28}; NH, $C^{\alpha}H$, and two $C^{\beta}Hs$ of Cys^{27}; NH, $C^{\alpha}H$, and $C^{\beta}H$ of Cys^{18}; NH, two $C^{\beta}Hs$, and two $C^{\gamma}Hs$ of Lys^{19}; NH and two $C^{\beta}Hs$ of Leu^{40}; NH, $C^{\alpha}H$, $C^{\beta}H$ and $C^{\gamma}H$ of Val^{2}; NH and two $C^{\alpha}Hs$ of Gly^{32}; and NH, $C^{\alpha}H$, and two $C^{\beta}Hs$ of Phe^{33}.

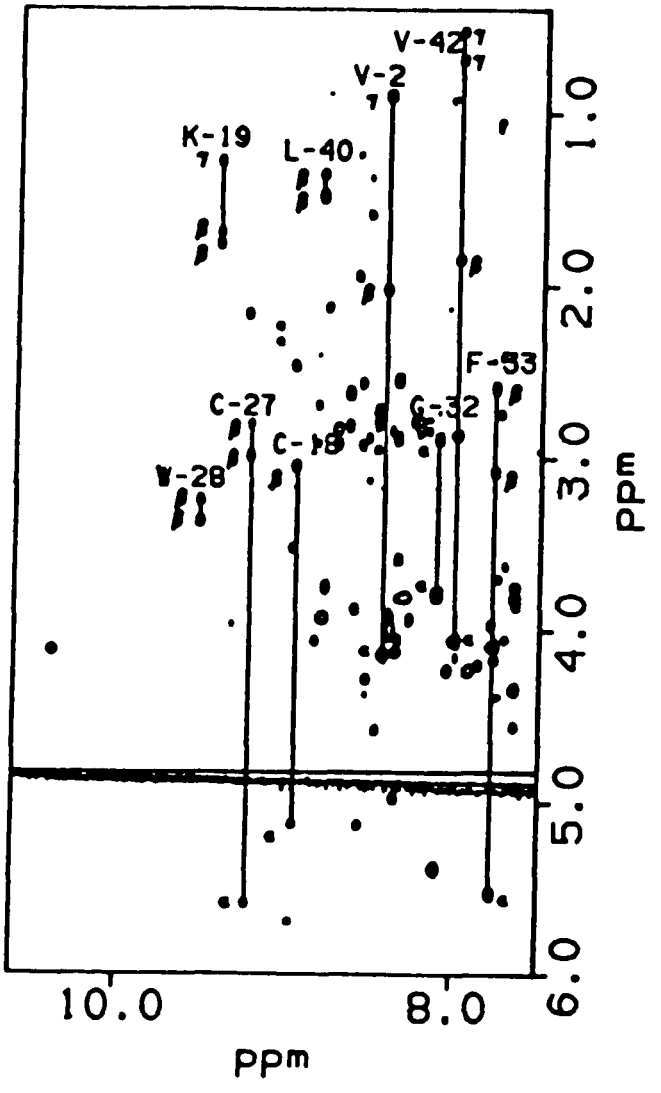
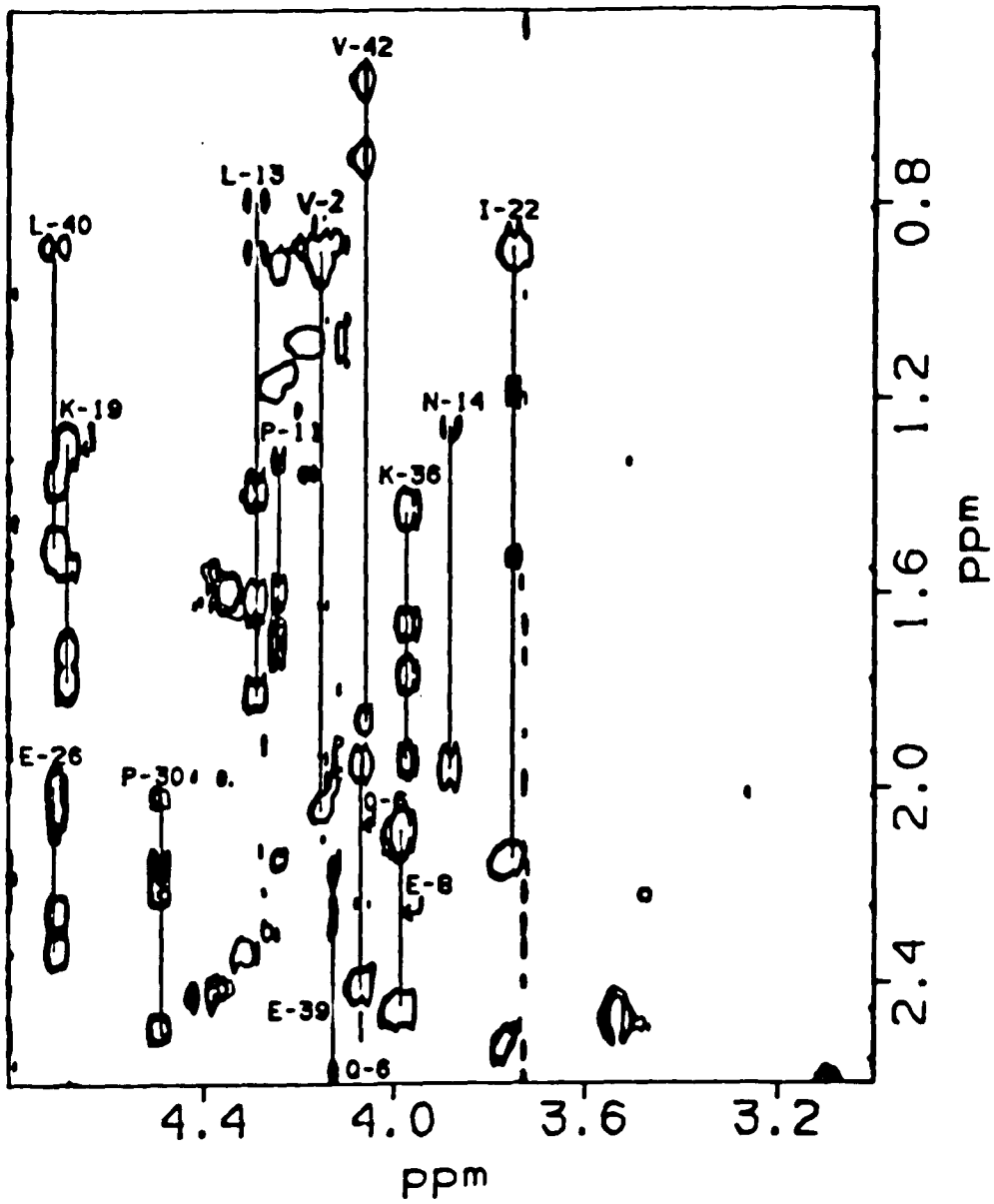


Figure 3.5. A contour plot portion of the TOCSY spectrum (25 °C, pD 4.2) in D_2O. The side chain connections with $C^{\alpha}H$ for a number of amino acids are indicated.



spectrum, the H_2O frequency is set at 0 Hz at $\omega 1$ and $\omega 2$ both axes, and for a set of α and β protons in the same amino acid, cross-peaks occur at the three-quantum frequency, $d\alpha + d\beta 1 + d\beta 2$. Triple-quantum frequencies are plotted along $\omega 1$, and single-quantum frequencies are plotted along $\omega 2$. Since the triple-quantum frequency at $\omega 1$ is the sum of three single-quantum frequencies: $d1 + d2 + d3 = \omega 1$, if the single-quantum frequency of only two of three protons are shown, the third one can be calculated from the measured triple-quantum frequency: $d1 = \omega 1 - (d2 + d3)$. For example, two $C^{\beta}Hs$ of Ser^{17}, and two $C^{\beta}Hs$ of Ser^{24} were very close in chemical shift, as identified from a TQ-COSY spectrum. Figure 3.6. shows a portion of a contour plot of the TQ-COSY spectrum. In figure 3.6. the connectivities of the following spin systems are shown: $C^{\alpha}H$ and $C^{\beta}Hs$ for Phe^{33}, Cys^{27}, Trp^{28}, Asn^{37}, Ser^{9}, Ser^{24} and Ser^{17}. The data gotten from TQ-COSY can also be used to confirm assignments made from DQF-COSY spectrum and TOCSY spectrum. This is useful information to map the $C^{\alpha}H$ to its two $C^{\beta}Hs$, or even a $C^{\beta}H$ to its two neighboring $C^{\gamma}Hs$. The spin systems observed using TQ-COSY spectrum are listed on Table 3.1.

Figure 3.6. A contour plot portion of the TQ-COSY spectrum (25 °C, pH 4.2). The connectivities between the $C^{\alpha}H$ and $C^{\beta}Hs$ in a number of amino acid spin systems are shown.

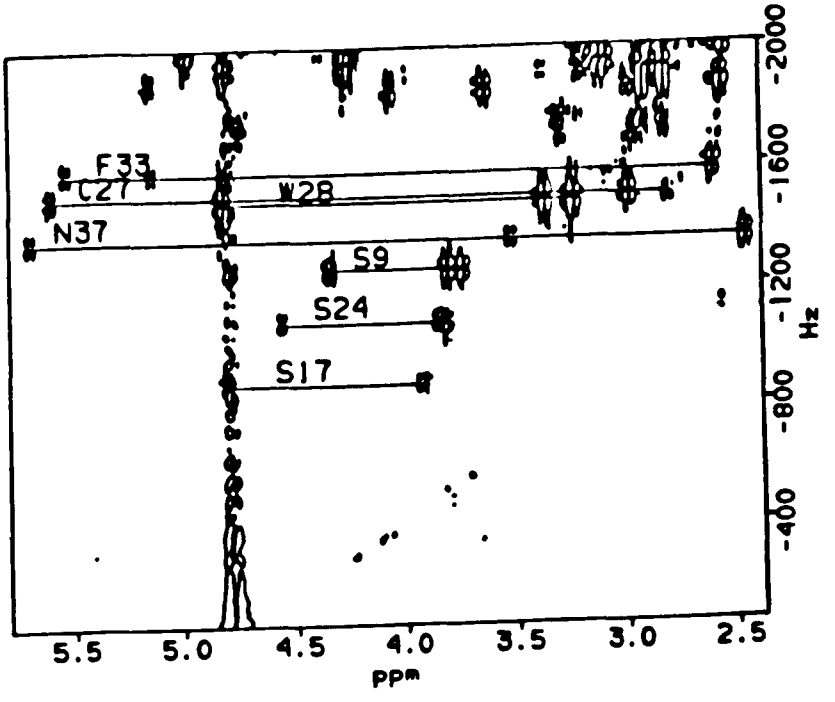


Table 3. 1. Proton chemical shifts for FIX-EGF1 obtained from the aliphatic region of the TQ-COSY (25 °C, pH 4.2) contour plot.

Amino acid	CαH	CβH	Cβ'H	CγH and others
3 Asp	4.58	2.82	2.74	
5 Asp	4.71	2.81	2.86	
6 Gln	4.06	1.93	2.19	
7 Cys	4.37	3.13	2.92	
8 Glu	3.97	2.07	2.04	
9 Ser	4.34	3.74	3.81	
10 Asn	4.47	2.90	2.87	
12 Cys	4.38	2.42	2.75	
13 Leu	4.28	1.59	1.38	
14 Asn	3.87	1.92	1.23	
17 Ser	4.81	3.93	3.89	
18 Cys	5.12	3.05	3.02	
19 Lys	4.66	1.77	1.70	
'20 Asp	4.68	2.92	2.70	
21 Asp	4.97	2.89	2.55	
22 Ile		2.12	1.51	1.16
23 Asn	4.79	2.91	2.83	
24 Ser	4.56	3.85	3.81	
25 Tyr	5.39	3.14	2.74	
26 Glu	4.69	1.99	2.04	
27 Cys	5.59	2.98	2.81	
28 Trp	4.83	3.37	3.25	
29 Cys	5.13	2 93	2.57	
30 Pro	4.47	2.49	2.15	
31 Phe	4.22	3.22	2.92	
33 Phe	5.52	3.07	2.60	
34 Glu	4.80	1.90	2.12	
36 Lys	3.96	1.94	1.76	
		1.94	1.76	1.42
				1.42 1.66 2.73
37 Asn	5.69	3.52	2.44	
38 Cys	4.05	3.60	2.99	
39 Glu	4.11	1.95	2.15	
40 Leu	4.69	1.49	1.35	
41 Asp	4.66	2.81	2.62	

The aromatic residues in the peptide were assigned by looking for the connectivities between the aromatic protons to aromatic protons from the TOCSY experiment, and the connectivities between the aromatic protons to their CβHs from a NOESY experiment in D_2O. Figure 3.7. **A**. shows the portion of a NOESY contour plot with the ring proton-CβH cross-peaks; Figure 3.7. **B** shows the TOCSY contour plot in the ring proton region. These two spectra were well correlated through ring proton frequencies. All five aromatic amino acid spin systems: Tyr^{1}, Tyr^{25}, Trp^{28}, Phe^{31} and Phe^{33} were identified. Tyr^{1} was assigned from these five aromatic amino acids by the absence of an NH resonance. From its characteristic TOCSY ring system pattern, Trp^{28} was identified. These unambiguous assignments provide reliable starting points for sequential assignment.

There are four Asn and one Gln in the peptide. They were distinguished from other amino acids by the observation of the strong side chain amide/amide proton resonances and the connection of amide proton to their aliphatic protons in the H_2O NOESY spectrum. Figure 3.8. shows Asn^{37} NOE connections between CβH and side chain NHs.

The two lysine residues were assigned by following the connectivities along the side chain protons. The Lys^{19} spin system was very well shown in H_2O TOCSY spectrum except for a missing

$C^{\alpha}H$ cross-peak, that was assigned from the H_2O DQF-COSY spectrum. The Lys^{36} spin system did not show complete

Figure 3.7. The ring proton assignments by using both the NOESY and the TOCSY spectra. **A**. A portion of the NOESY contour plot (25 °C, pH 4.2) with the ring proton-$C^{\beta}H$ cross-peaks marked; **B** The TOCSY contour plot (25 °C, pD 4.2) in the ring proton region. The connectivities between ring protons and $C^{\beta}Hs$ for all five aromatic amino acid residues in the peptide are depicted.

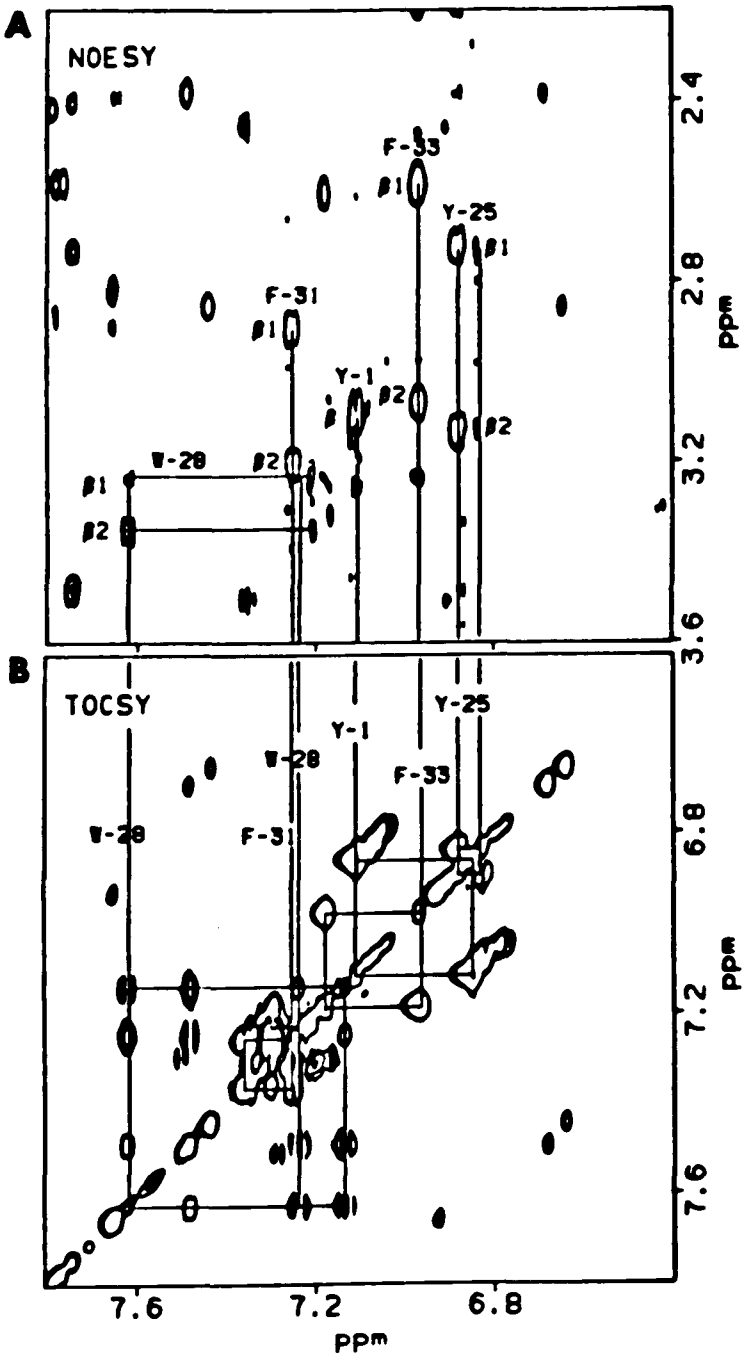
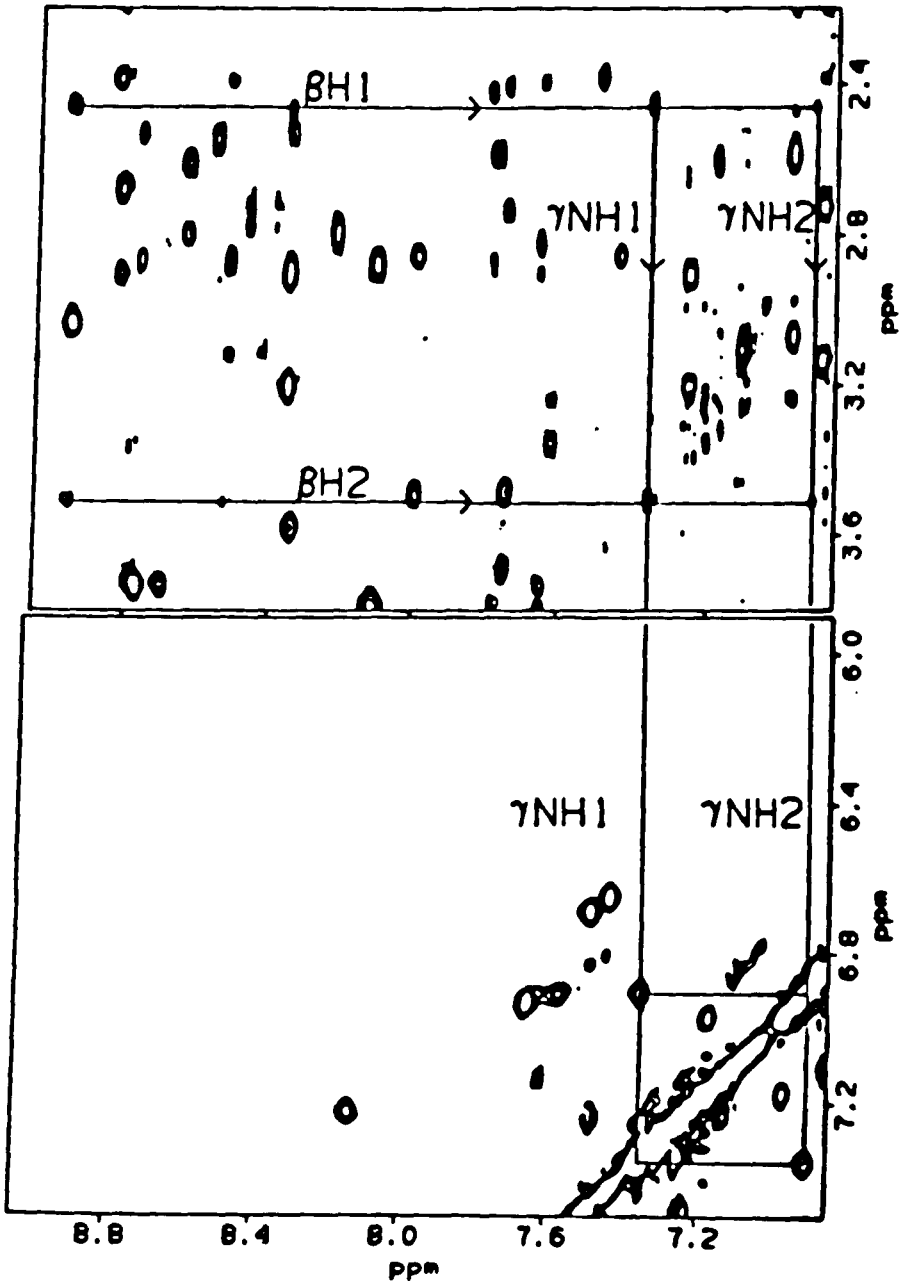

A
NOESY
F-33
β1
Y-25
β1
F-31
β1
Y-1
β2
β
β2
W-28
β2
β1
β2
2.4
2.8
3.2
3.6
ppm
B
TOCSY
Y-25
Y-1
W-28
F-33
W-28
F-31
6.8
7.2
7.6
ppm
7.6
7.2
6.8
ppm

Figure 3.8. Two sections of the H_2O NOESY contour plot (25 °C, pH 4.2) showing the connectivities between γ-NH and $C^{\beta}H$ for Asn^{37}.



connections of the side chain protons to its backbone amide proton, but the connection of $C^{\varepsilon}H$, $C^{\delta}H$, $C^{\gamma}H$ and $C^{\beta}H$ with $N^{\varepsilon}H$ were observed in H_2O TOCSY spectrum. The TOCSY spectrum in D_2O provided the correlation to the $C^{\alpha}H$ for the Lys^{36} residue. The amine protons of Lys^{19} and Lys^{36} were seen in H_2O TOCSY spectrum. The observation of these protons suggests that they may be involved in forming hydrogen bonds, or alternately that the side chain was hidden inside of the molecule.

The spin systems of two Val, two Leu and one Ile were identified through methyl proton connections in the DQF-COSY and the TOCSY spectra. Val and Leu residues each have two methyl groups coupled to one proton ($C^{\beta}H$ in the Val spin system, $C^{\gamma}H$ in the Leu spin system) while Ile residue has one methyl proton coupled to two $C^{\gamma}H$ protons and another methyl group coupled to a $C^{\beta}H$ proton. Leu spin systems can be distinguished from Val spin systems by the $C^{\gamma}H$ cross-peaks in the TOCSY spectrum recorded in D_2O. The sequence of two Val and two Leu residues were assigned by sequential NOE connectivities later.

Five Gly residues were assigned differently. Gly^{15}, Gly^{16} and Gly^{32} showed a characteristic pattern in the fingerprint region of the H_2O DQF-COSY spectrum. In addition, Gly^{15}, Gly^{16}, Gly^{32} and Gly^{4} showed characteristic butterfly-shaped strong cross-peaks in the TOCSY spectrum. Gly^{35} was assigned through sequential assignment.

The two Pro residues were sorted out by analysis of TOCSY and DQF-COSY spectra in D_2O. They showed the following cross-peaks: $C^{\alpha}H/C^{\beta}H$, $C^{\alpha}H/C^{\gamma}H$, and $C^{\alpha}H/C^{\delta}H$ in the NOESY spectrum with a feature of lacking amide proton connections in the spin system. Figure 3.9 shows intraresidue NOESY connections for Pro^{11} and Pro^{30} in **A** and **B**, respectively.

Sequential Resonance Assignment. The sequential resonance assignment focused on the backbone amide protons. It is very important to locate the amide protons of all spin systems. Figure 3.10 shows the NH-$C^{\alpha}H$ "fingerprint" region of FIX-EGF1 in the H_2O DQF-COSY. Of 40 expected cross-peaks, 34 were found in this region. Six missing cross-peaks were: Trp^{28}, Asn^{23}, Glu^{34}, Ser^{17}, Asp^{5} and Gly^{35}. These cross-peaks were not observed as a result of overlap with other intraresidue cross-peaks or as a result of radio-frequency irradiation used for suppression of the water resonance.

Figure 3.9. The NOESY (25 °C, pD 4.2, in D_2O) contour plot showing the intraresidue NOE connections for Pro^{11} and Pro^{30} in **A** and **B**, respectively.

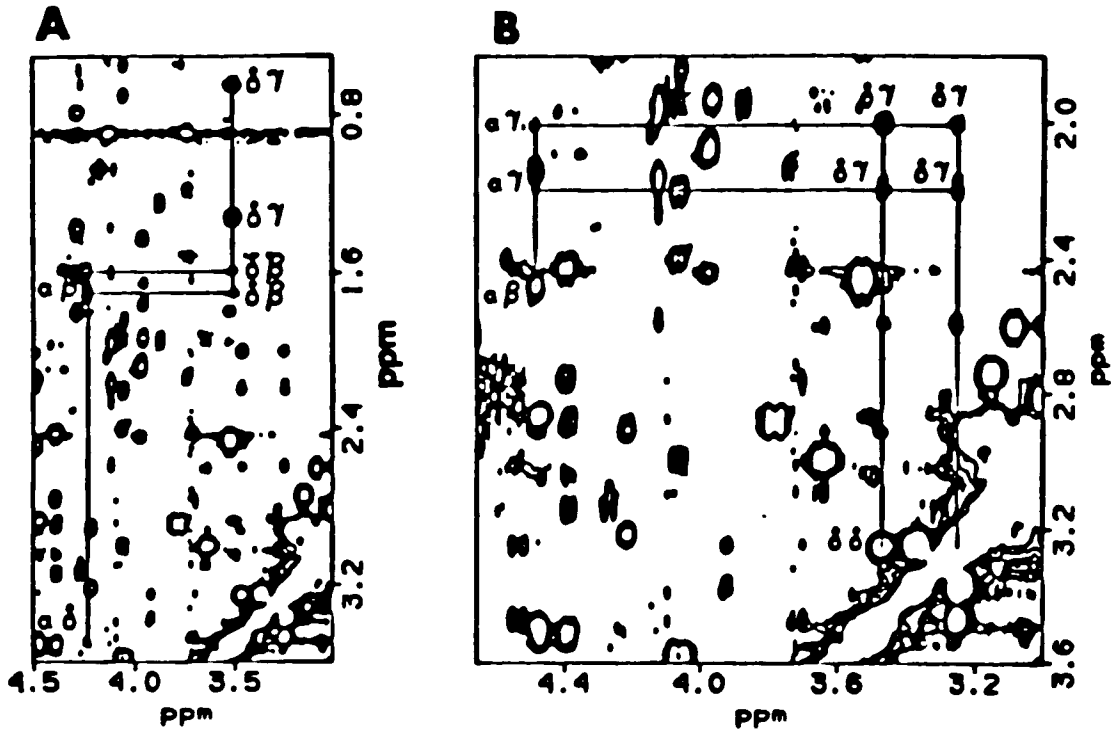
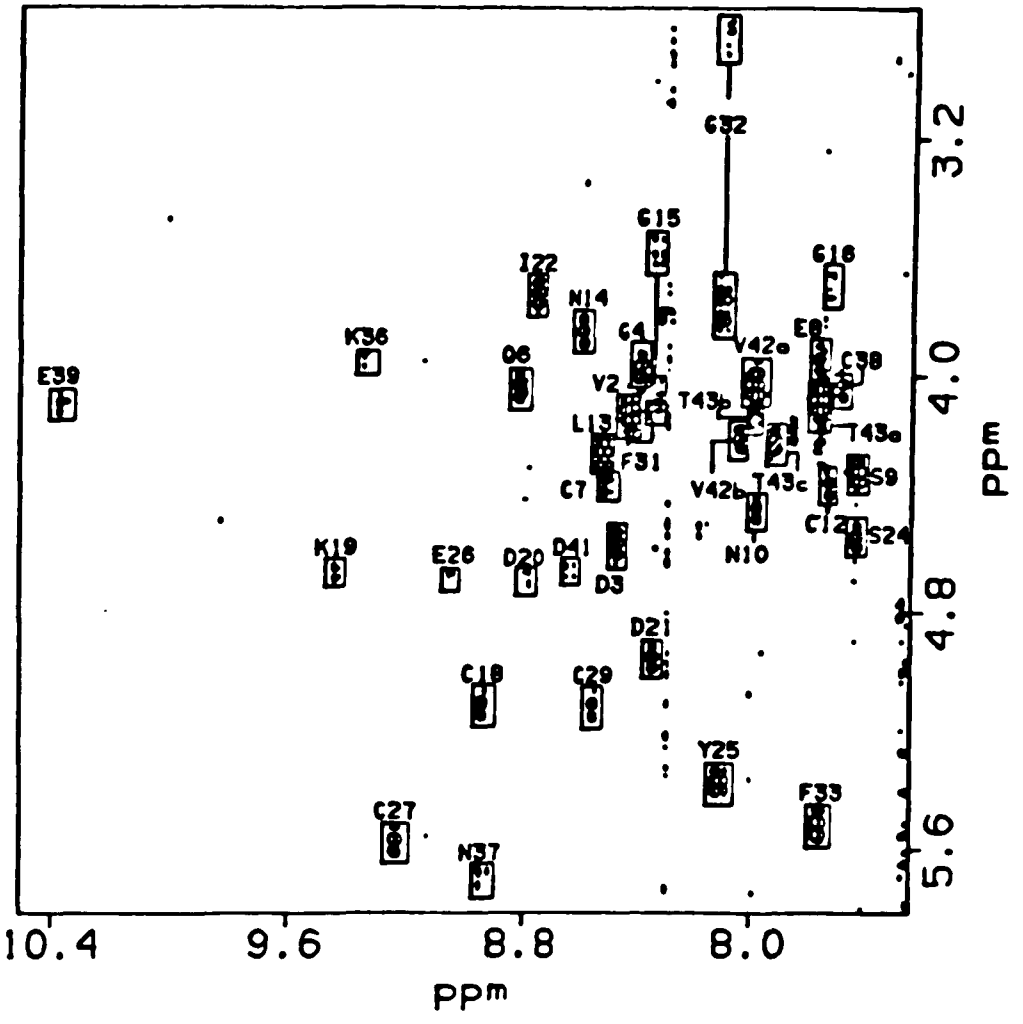


Figure 3.10. The contour plot of the "fingerprint" region of the H_2O DQF-COSY (25 °C, pH 4.2), showing the amide/α-proton cross-peaks.



The sequential assignment of the peptide can start with $NH_{(i)}$-$NH_{(i+1)}$ and $C^{\alpha}H_{(i)}$-$NH_{(i+1)}$ connections in the H_2O NOESY spectrum to "walk" down the peptide backbone. Figure 3.11 shows the sequential connection paths of Leu^{13} to Ser^{17} and Pro^{30} to Gly^{32} through $C^{\alpha}H_{(i)}$-$NH_{(i+1)}$ cross-peaks in the H_2O NOESY spectrum. Figure 3.12 shows the sequential connections through $NH_{(i)}$-$NH_{(i+1)}$ cross-peaks for Gly^{4} to Asn^{10} in the H_2O NOESY spectrum.

At Pro^{11} and Pro^{30}, the loss of $NH_{(i)}$-$NH_{(i+1)}$ connections is inevitable, but the connectivities between $C^{\delta}H$ of proline and $C^{\alpha}H$ of the previous residue were detected for both prolines. This indicated that the prolines were predominantly in the *trans* configuration (68). No indication of a cross-peak of Pro $C^{\alpha}H$ with the previous amino acid $C^{\alpha}H$ for either proline was observed, providing negative evidence that both proline were not in a *cis* configuration. Figure 3.13. **A** shows *trans* and *cis* proline configuration; figure 3.13. **B** shows sequential connections of both prolines.

Figure 3.11. A contour plot section of the NOESY spectrum in H_2O (25 °C, pH 4.2), showing the sequential connections for Leu^{13} to Ser^{17} and Pro^{30} to Gly^{32} paths through $C^{\alpha}H_{(i)}$-$NH_{(i+1)}$ cross-peaks.

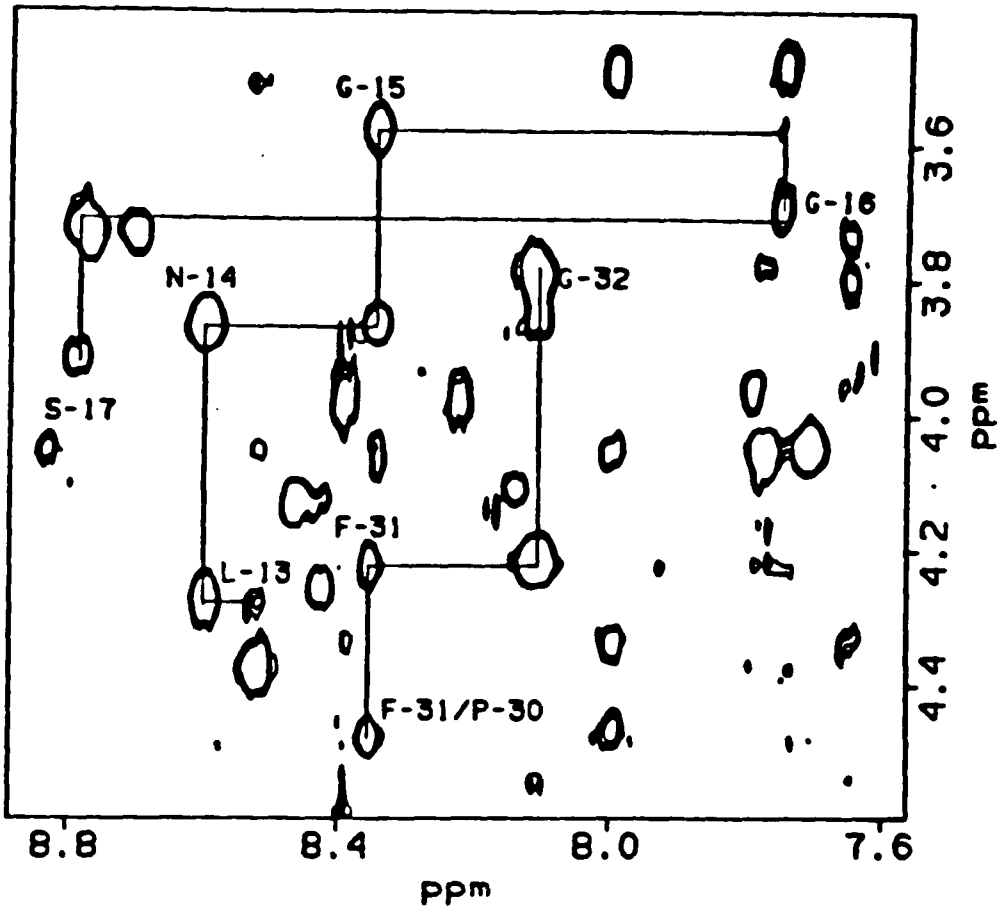


Figure 3.12. A contour plot section of the NOESY spectrum in H_2O (25 °C, pH 4.2), showing the sequential connections through $NH_{(i)}$-$NH_{(i+1)}$ cross-peaks for Gly^4 to Asn^{10}.

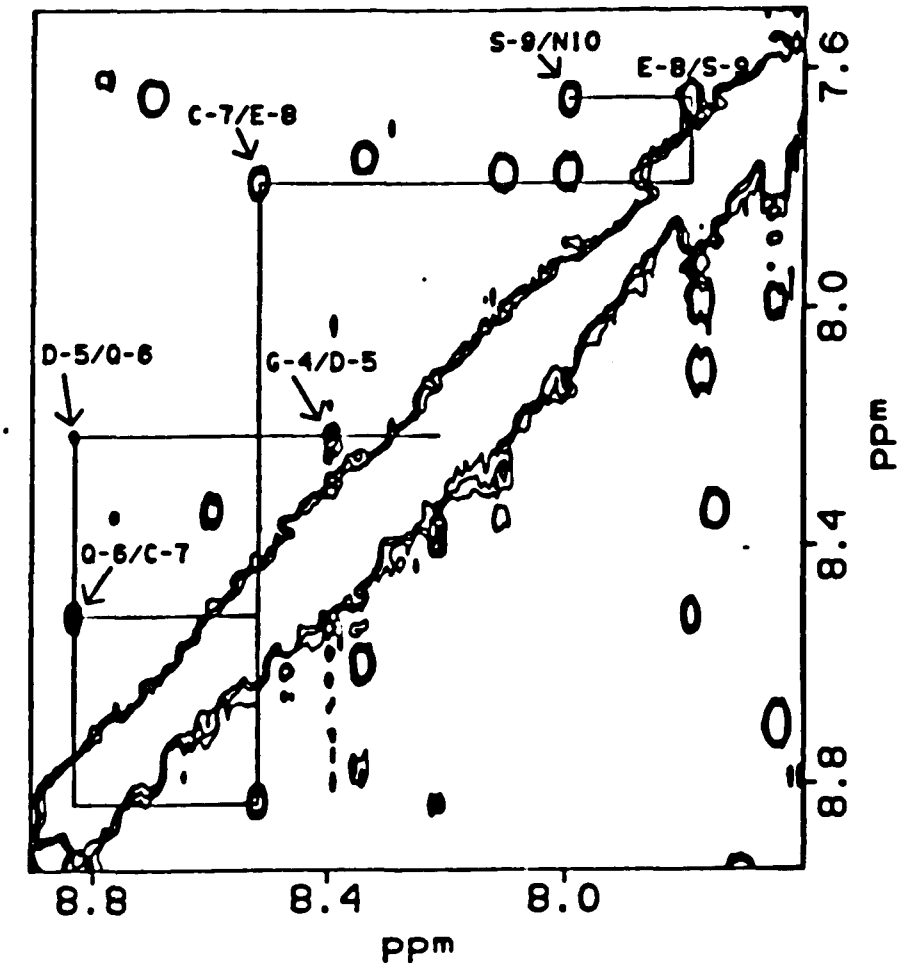
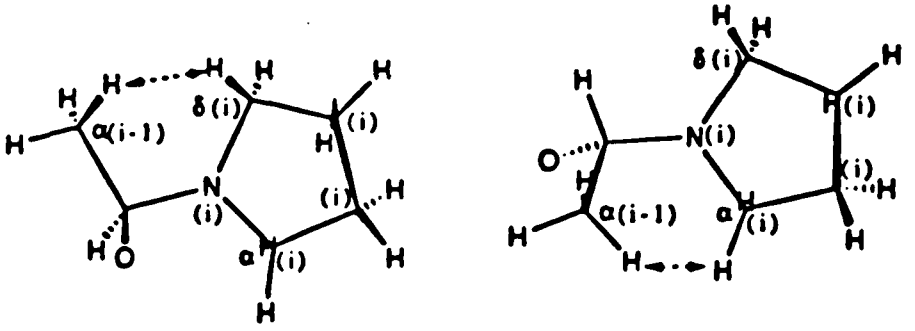


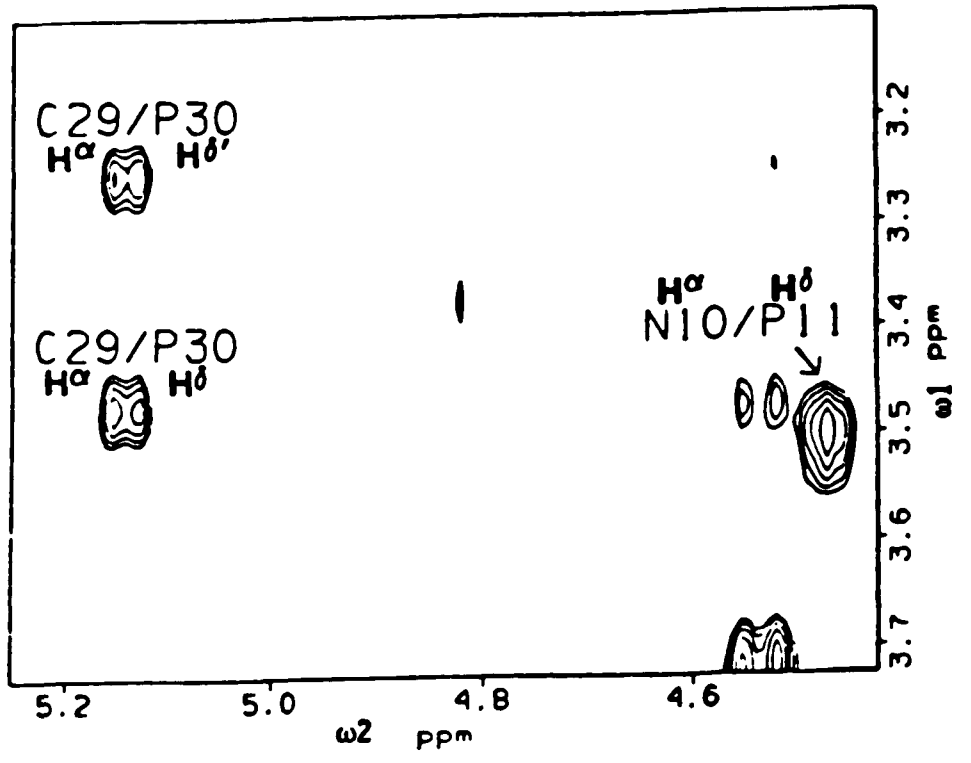
Figure 3.13. Proline configuration in the peptide. **A** shows *trans* and *cis* proline configuration (68); **B** shows sequential connections of both prolines. The NOE connectivities between $C^{\delta}H$ of Pro^{11} and $C^{\alpha}H$ of Asn^{10}, and between $C^{\delta}H$ of Pro^{30} and $C^{\alpha}H$ of Cys^{29} shown indicate the *trans* proline configurations for both residues.

A


α(i-1)
δ(i)
(i)
N
(i)
α
(i)
O
α(i-1)
N(i)
δ(i)
(i)
α
(i)

Trans Cis

B


C29/P30
Hα Hδ'
C29/P30
Hα Hδ
Hα Hδ
N10/P11
5.2
5.0
4.8
4.6
ω2 ppm
3.2
3.3
3.4
3.5
3.6
3.7
ω1 ppm

The NOE sequential connections of the peptide are summarized in Figure 3.14. The proton chemical shifts are listed in Table 3.2.

Figure 3.14. Summary of the sequential connectivities and antiparallel β-sheet segments found for FIX-EGF1. Disulfide linkages are indicated by single lines, and slowly exchanging amide resonances are depicted by an asterisk. Sequential NOE connectivities are indicated in three rows below the amino acid sequence. The row labeled dαN indicates the observed connectivities between an a proton and an amide proton by a thick line (in the case of Pro, the observed connectivities are dad, instead). Residues with observable amide/α proton coupling constants greater than 8 Hz are indicated by filled circles. The antiparallel β-sheet regions are shown by heavy checkered lines with thinner lines indicating the associated turns.

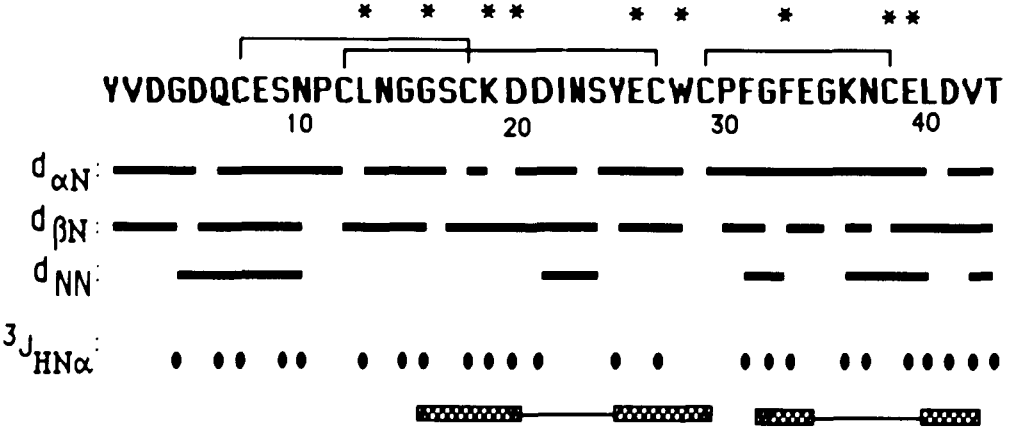


Table 3.2. ^{1}H NMR chemical shifts of FIX-EGF1 (pH 4.2).

Amino acid	chemical shifts (ppm)			
Residues	NH	αH	βH	others
Y-1		4.25	3.10, 3.13	2,6H: 7.10; 3,5H: 6.85
V-2	8.41	4.14	2.01	γCH_3: 0.86, 0.90
D-3	8.46	4.58	2.74, 2.82	
G-4	8.38	3.89, 4.00		
D-5	8.22	4.71	2.82, 2.86	
Q-6	8.81	4.06	1.93, 2.19	γCH_2: 2.24, 2.40; δNH_2: 6.69, 7.49
C-7	8.49	4.37	2.89, 3.13	
E-8	7.77	3.97	2.04, 2.07	γCH_2: 2.42,2.42
S-9	7.64	4.34	3.74, 3.81	
N-10	7.98	4.47	2.87, 2.90	γNH_2: 6.65, 7.43
P-11		4.24	1.58, 1.69	γCH_2: 0.64, 1.31; δCH_2: 3.50
C-12	7.73	4.38	2.42, 2.75	
L-13	8.51	4.28	1.38, 1.59	γCH_2: 1.80; δCH_3: 0.79, 0.89
N-14	8.58	3.87	1.23, 1.92	γNH_2: 7.17, 7.22
G-15	8.33	3.56, 4.06		
G-16	7.73	3.69, 4.53		
S-17	8.78	4.81	3.89, 3.93	
C-18	8.94	5.12	3.05, 3.02	
K-19	9.43	4.66	1.70, 1.77	γCH_2: 1.28; δCH_2: 1.53; γCH_2: 2.73; εNH_3: 7.37
D-20	8.81	4.68	2.70, 2.92	
D-21	8.33	4.97	2.55, 2.89	
I-22	8.74	3.74	2.12	γCH_2: 1.16, 1.51; γCH_3: 0.88
N-23	8.68	4.79	2.83, 2.91	γNH_2: 6.93, 7.66
S-24	7.64	4.56	3.79, 3.85	
Y-25	8.10	5.39	2.74, 3.14	2,6H: 6.88; 3,5H: 6.84
E-26	9.05	4.69	1.99, 2.04	γCH_2: 2.24, 2.33
C-27	9.22	5.59	2.81, 1.98	
W-28	9.54	4.83	3.25, 3.37	2H: 7.22; 4H: 7.62; 5H: 7.12; 6H: 7.22; 7H: 7.48; NH: 10.07
C-29	8.54	5.13	2.57, 2.93	
P-30		4.47	2.15, 2.49	γCH_2: 2.02, 2.21; δCH_2: 3.26, 3.48
F-31	8.35	4.22	2.92, 3.22	2,6H: 7.25; 3,5H: 7.35; 4H: 7.29
G-32	8.10	3.79, 2.89		
F-33	7.77	5.52	2.60, 3.07	2,6H: 6.97; 3,5H: 7.17
E-34	9.25	4.80	1.90, 2.12	γCH_2: 2.08, 2.17
G-35	8.13	4.11, 3.79		
K-36	9.33	3.96	1.94, 1.76	γCH_2: 1.42; δCH_2: 1.65; εCH_2: 2.74, 2.80; εNH_3: 7.34
N-37	8.93	5.69	3.52, 2.44	γNH_2: 6.91, 7.36
C-38	7.69	4.05	3.60, 2.99	
E-39	10.38	4.11	1.95, 2.15	γCH_2: 2.25, 2.56
L-40	8.78	4.69	1.35, 1.49	γCH_2: 1.18; γCH_3: 0.88
D-41	8.63	4.66	2.62, 2.81	
V-42	7.99	4.05	1.85	γCH_3: 0.68, 0.53
T-43	7.76	4.10		γCH_3: 1.07

Secondary Structure determination. The secondary structure determination for the peptide was based on nonsequential and sequential NOESY connectivities. No evidence for the presence of α-helical structure was detected as judged by the absence of $C^{\alpha}H\text{-}NH_{i+3}$ NOESY peaks (68). Extended β-structure is characterized by the presence of interstrand $d_{\alpha\alpha}$ and $d_{\alpha N}$ connectivities. The following short interstrand proton-proton NOE cross-peaks are expected in an antiparallel β-sheet: $d_{\alpha\alpha}$ =2.3 Å, $d_{\alpha N}$ = 3.2 Å, and d_{NN} = 3.3 Å (68). Residues that were involved in the β-sheet structure were pointed out by the presence of non-sequential $d_{\alpha\alpha}$ NOESY connectivities. Figure 3. 15 shows a portion of the NOESY contour plot where the cross strand cross-peaks for $d_{\alpha\alpha}$ of Gly^{16}/Cys^{29}, Phe^{33}/Asp^{41}, Tyr^{25}/Asp^{20} and Cys^{18}/Cys^{27} are illustrated. Other evidence for the β-sheet structure come from the presence of slowly exchanging amide protons, and $^{3}J_{HN\alpha}$ coupling constants ≥ 8 Hz for the residues in the β-sheet. Most of interstrand NOE's characteristic for an antiparallel β-sheet were observed, and three out of four hydrogens involved in hydrogen bonding exchange slowly with the solvent in the first sheet. The rate of solvent exchanging was not able to be determined for the amide proton of Ser^{17} in the first sheet, and the amide proton of Glu^{34} in the second sheet because the alpha protons of both amino acids are too close to the water frequency. Another amide proton involved in hydrogen bonding is the one of Gly^{32}. The cross-peak of $NH\text{-}C^{\alpha}H$ for Gly^{32} and Gly^{35} are almost overlapped, therefore, the measurement of the solvent exchange rate for the amide proton of

Gly^{32} is not unambiguous. The sequential NH_i-$C^{\alpha}H_{i+1}$ connectivities of the residues in the sheet are much stronger than the intra NH_i-$C^{\alpha}H$ connectivities. Two sections of proposed antiparallel β-sheet structure are shown in Figure 3.16. One sheet is between residues 16-20 and 25-29 (Figure 3.16 **A**). Another short sheet between residues 32-34 and 40-42 is shown in Figure 3.16 **B**. Heavy arrows indicate the stronger NOESY cross-peaks. Light arrows mean weaker NOESY cross-peaks. Dotted lines indicate proposed hydrogen bonds. All expected short interstrand NOE peaks for typical antiparallel β-sheets were observed, except the $C^{\alpha}H(Cys^{29})$/$NH(Ser^{17})$ cross-peak in the first sheet, and the $NH(Gly^{32})$/$NH(Val^{42})$ cross-peak in the second sheet. Since the β-sheet is an extended conformation, it should have very short $d_{\alpha N}(i, i+1) \approx 2.2$ Å, and an intra-residue $C^{\alpha}H(i)$-$NH(i)$ spin-spin coupling constant ($^3J_{HN\alpha}$) near 9 Hz (68).

Figure 3.15. A portion of the NOESY (25 °C, pD 4.2) contour plot showing the cross-strand $d_{\alpha\alpha}$ cross-peaks arising from the antiparallel β-sheet structure.

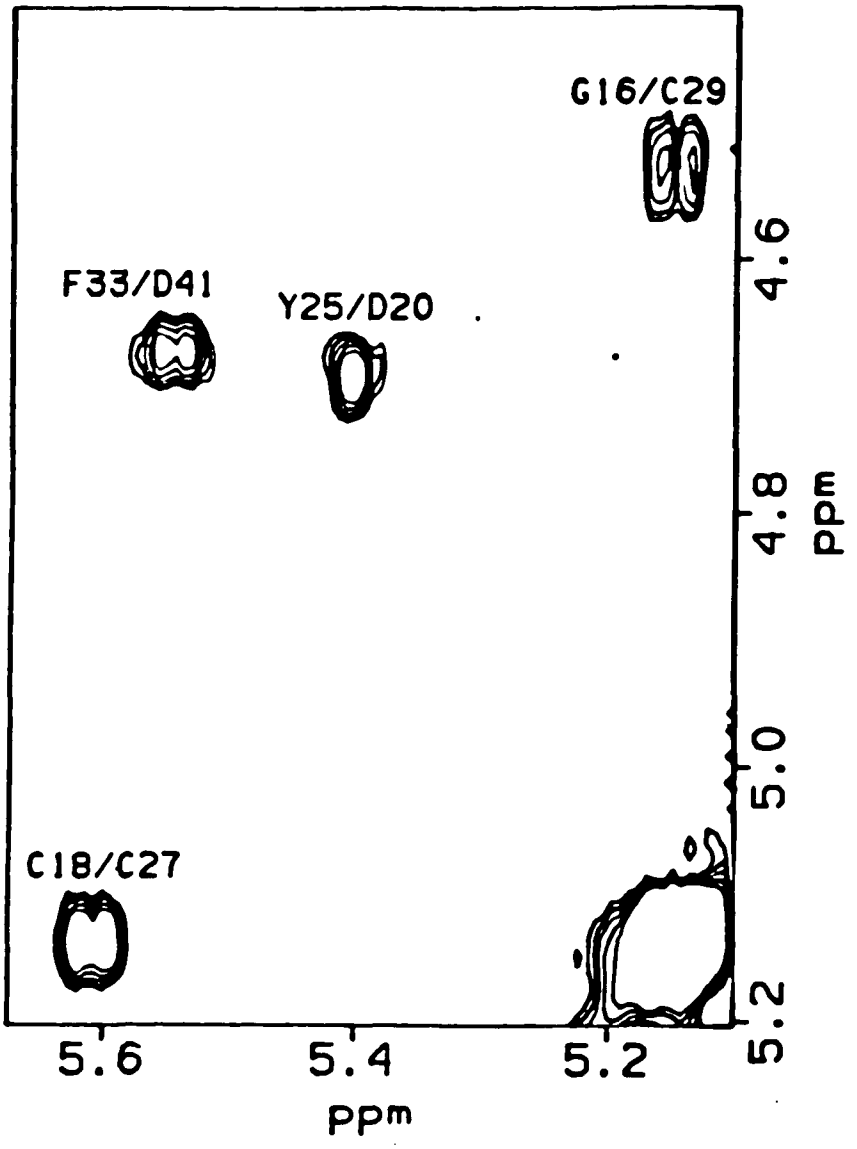
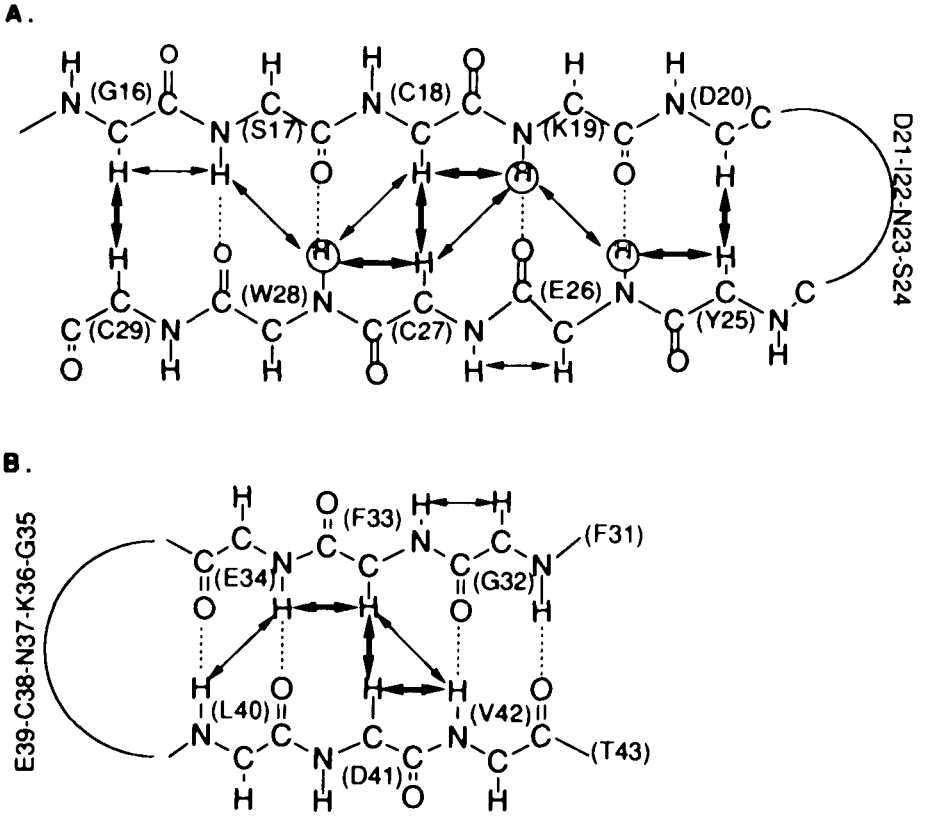


Figure 3.16. Schematic diagram of the antiparallel β-sheets in FIX-EGF1. NOEs are indicated by thicker arrows and thinner arrows for stronger and weaker NOEs, respectively. Putative hydrogen bonds are displayed by dashed lines. Slowly exchanging protons are encircled.



In the β-sheets of FIX-EGF1, sequential $d_{\alpha N}$s were consistent with an extended sheet structure, and $^3J_{HN\alpha} \geq 8$ Hz was measured for most residues. Some coupling constants were not detected due to water resonance suppression. Residues with coupling constants ≥ 8.0 Hz are listed in Table 3.3. Each β-sheet has an associated β-turn. A tight turn was observed between amino acid residues 21-24 in the first β-sheet. This turn does not contain a glycine residue. This indicates it is unlikely to be a type I', type II, or type II' turn (71). The NOESY connectivities and $^3J_{HN\alpha}$ coupling constants in the turn are not consistent with a type I turn either. The criteria for classification as a type I turn and the observed experiment data for the turn in the first β-sheet are listed in Table 3.4. Another β-turn associated with the second β-sheet consists of five amino acid residues which precluded classification as any standard class of β-turn, as these characteristically contain four amino acid residues.

Table 3.3. $^3J_{HN\alpha}$ coupling constants (≥ 8.0 Hz) determined from the DQF-COSY spectrum in H_2O.

Amino acid	$^3J_{HN\alpha}$ (Hz)
Gly^4	9.7
Gln^6	8.5
Cys^7	8.0
Ser^9	9.4
Asn^{10}	8.4
Cys^{12}	9.4
Leu^{13}	9.1
Gly^{15}	9.9
Gly^{16}	9.3
Cys^{18}	8.1
Lys^{19}	10.2
Asp^{20}	8.2
Asp^{21}	9.8
Tyr^{25}	9.0
Cys^{27}	9.5
Cys^{29}	8.4
Phe^{31}	9.5
Gly^{32}	9.8
Phe^{33}	10.0
Lys^{36}	8.7
Asn^{37}	9.2
Glu^{39}	11.4
Leu^{40}	9.2
Asp^{41}	9.6
Val^{42}	8.5
Thr^{43}	8.2

Table 3.4. The criteria for classification of type I turn (68) and the experiment data of the turn in the first β-sheet between amino acid residues 21 to 24.

	type I turn			FIX-EGF1				
amino acid position	1	2	3	4	21	22	23	24
$^3J_{HN\alpha}$ (Hz)		4	9			7.2	*	
$d_{\alpha N}(i, i+3)$		w**	w	w		Not	observed	
$d_{NN}(i, i+2)$			w	w		Not	observed	
$d_{\alpha N}(i, i+2)$			w	w		Not	observed	
d_{NN}			m**	s**	w	w		
$d_{\alpha N}$		w	w	w	m	s		

*: C^aH [23] is too close to the H_2O frequency to get the coupling constant.

**: "w" means weak NOESY connection; "m" means medium strong NOESY connection; "s" means strong NOESY connection.

Calcium Binding Study. The calcium binding for this peptide has been reported (41-42). The peptide binds calcium ion significantly only at pH near 5 or higher. DQF-COSY spectra at pD values of 4.2, 4.7 and 5.4 were compared. The NMR spectrum at pD 4.7 showed very similar chemical shift overall; The spectrum at pD 5.4 showed a significant resonance shifts, especially for acidic amino acid: Asp and Glu residues. To confidently use the assignment information at pD 4.2 as a calcium-free template, while providing a proper condition for calcium binding, pD 4.7 was chosen for the calcium binding study.

A DQF-COSY spectrum of FIX-EGF1 was acquired in D_2O at pD 4.7. Another DQF-COSY experiment was acquired after adding 20 mM $CaCl_2$ into the peptide solution at pD 4.7. Resonance assignment of the spectrum with Ca^{2+} was confirmed by using information from TOCSY and NOESY spectra in H_2O.

By analysis of the D_2O DQF-COSY and H_2O TOCSY spectra, it was shown that chemical shift of most resonances remained nearly unchanged after adding calcium. This indicated that the calcium binding did not cause a major conformation change on the peptide. Only 11 out of 43 residues have aliphatic proton resonances which are shifted by more than 0.03 ppm after adding calcium. The summary of the calcium binding effect on resonance changes is shown in Figure 3.17, and Table 3.5.

Figure 3.17. Calcium effect on the proton chemical shifts in the DQF-COSY spectrum at pH 4.7. The bar graph shows the chemical shift changes for each amino acid after adding calcium (20 mM $CaCl_2$, pD 4.7, 25 °C). The filled bars show the magnitude of the maximum shift observed among the α and β proton resonances, while the open bars indicated the sum of the absolute values of the chemical shift changes for these resonances. It was not possible to determine the chemical shift change for either Pro^{11} or Pro^{30},and so the values for positions 11 and 30 are unknown.

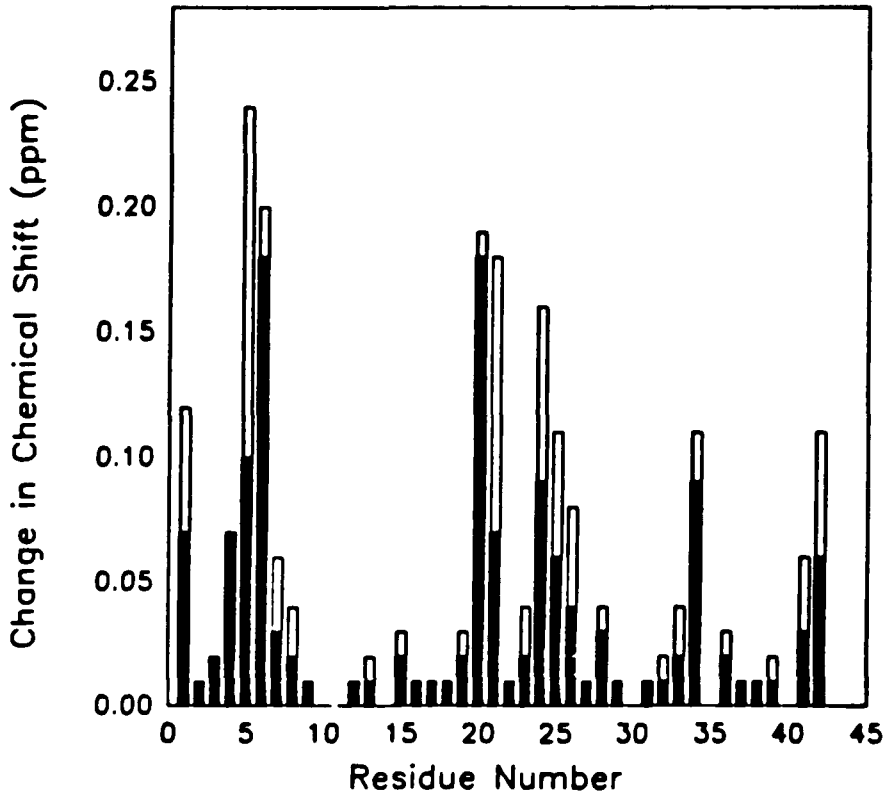


Table 3.5. Chemical shifts for residues with protons which change resonance position by more than 0.02 ppm after addition of calcium (at pH 4.7).

Amino acid residues		Chemical Shifts (ppm) αH	βH	γH	others
Y1	-Ca	4.26	3.08, 3.16		
	+Ca	4.30	3.09, 3.23		
G4	-Ca	3.94, 4.04			
	+Ca	3.94, 4.11			
D5	-Ca	4.71	2.78,.2.82		
	+Ca	4.67	2.68, 2.72		
Q6	-Ca	4.06	1.94, 2.19	2.22, 2.41	
	+Ca	4.06	1.93, 2.01	-	
K19	-Ca	4.67	1.76	1.27	1.53, 2.73
	+Ca	4.67	1.70, 1.78	1.28	1.55, 2.78
D20	-Ca	4.68	2.68, 2.90		
	+Ca	4.67	2.68, 2.72		
D21	-Ca	4.98	2.55, 2.87		
	+Ca	5.03	2.48, 2.93		
S24	-Ca	4.55	3.79, 3.81		
	+Ca	4.52	3.83, 3.90		
Y25	-Ca	5.40	2.74, 3.15		
	+Ca	5.34	2.72, 3.12		
E34	-Ca	4.80	1.90, 2.12	-	
	+Ca	4.78	1.92, 1.99	1.84, 2.09	
V42	-Ca	4.08	1.87	0.53, 0.69	
	+Ca	4.13	1.93	0.51, 0.69	

3.2. Solution structure determination and refinement

To establish the solution structure for a peptide, 2D NOESY spectra are required. The nuclear Overhauser effect (NOE) gives rise to a change in the integrated NMR absorption intensity of a nuclear spin when the NMR absorption of another spin is saturated (68). This change is due to nonequilibrium magnetization of a nucleus being transferred through dipole-dipole relaxation through space to neighboring nuclei. The 2D NOESY technique has been extensively applied to macromolecular structure since it gives one information about the relative dispositions of protons in space.

The structural constraints can be obtained by: 1. using NOESY to estimate the proton-proton distances through space; 2. determining the dihedral angels from the coupling constants; 3. locating possible hydrogen bonds from the rate of amide proton exchange in D_2O. Those constraints can be used to determine solution structure of a molecule.

In this section, results of proton-proton distance calculations from NOEs in D_2O and H_2O at different mixing times, and assignments of those NOE cross-peaks, will be presented. In addition, the stereospecific assignment of the side chain protons, and location of hydrogen bonds will be reported.

3.2.1. Materials and Methods

Peptide sample and NMR experimental conditions. The peptide sample and NMR experimental conditions were the same as described in section 3.1. except for the following: The 2D NMR spectra were collected in the following form: 2048 complex points were acquired in t_2, 330 and 350 complex FIDs in t_1 for H_2O NOESY experiment and D_2O NOESY experiment, respectively. The mixing time for both H_2O and D_2O experiments were 1, 40, 80, 120, and 200 ms. 2D-NMR data were transferred to IBM RS6000 computer, where spectrum processing was carried out using Felix 2.0 (Hare Research, Inc.).

Proton-proton distance calculation: Since the NOE buildup rate depends on r^{-6} (101), where r is the distance between a pair of protons, for pair **a** and pair **b** protons at distance r_a and r_b in a rigid structure, one would have:

$$s_a/s_b = (r_b/r_a)^6$$

(s_a and s_b represent the NOE intensities of peak **a** and **b**. r_a represents the distance of separation between the pair of protons giving rise to peak **a**. Similarly, r_b is the distance between protons giving rise to peak **b**.) This equation offers a path for determination of the distance ratios between different pairs of spins. It is useful for an empirical calibration of the correlation between NOE intensities and ^1H-^1H through-space distances (68, 102). NOE intensities were determined by integration of each peak with Felix program. The initial buildup rate of NOE's is time dependent, and

related to the inverse sixth power of the distance between the observed and the presaturated protons (102). Proton-proton distances were quantified by the following steps: **1**. integrating the peak intensity at different NOESY mixing times; **2**. plotting the integrated intensities versus mixing times to find out the initial buildup rate - its slope at time = 0; **3**. calculating the distance by using the GlyαH-αH' proton distance as a standard (known GlyαH-αH' = 1.75Å (68)), then applying the formula:

$$\text{xÅ} \;/\; \text{GlyÅ} = (\text{Gly slope} \;/\; \text{x slope})^{1/6}$$

to get the distances. The two sets of NOESY spectra in H_2O and D_2O serve for determination of peptide backbone conformation and side chain position in space, respectively.

NOESY cross-peak assignment: NOESY cross-peak assignments were based on previous sequence specific assignment information.

Hydrogen bond location: Hydrogen bonds were located based on the information from slowly exchanging protons.

Stereospecific assignment of β-methylene protons: The diastereotopic β-methylene protons of amino acid side chains can be distinguished by combined measurements of $^3J(H^{\alpha}\text{-}H^{\beta})$ coupling constants and $NH_i\text{-}C^{\beta}H_i$ nuclear Overhauser effects(103-105). $NH_i\text{-}C^{\beta}H_i$ nuclear Overhauser effects were estimated from cross-peak intensities in a NOESY spectrum recorded with 80 ms mixing time.

$^3J(H\alpha\text{-}H\beta2)$ and $^3J(H\alpha\text{-}H\beta3)$ vicinal coupling constants were measured from the $H\beta2/H\alpha$, and $H\beta3/H\alpha$ cross-peaks of PE-COSY (primitive elegant correlation spectroscopy) spectrum, respectively (106).

Reexamination of NOESY data. NOESY data were reexamined after generating some high energy starting structures by distance geometry calculation. Only cross-peaks with unique assignments were included, and a particularly critical examination was made for distance constraints assignments repeatedly causing high energy violations.

Structure generation. The solution structures were generated in the following way: first, all constraints were converted into coordinates - embedding to form an input file; then, DSPACE program was applied to generate a number of structures within the constraint limits; during the DSPACE calculation, structures were annealed (heated up), very slowly cooled down, then, energy minimization was applied. The output file listed the penalty for each structure. The structures with penalty numbers smaller than 2 were considered as candidates for further energy refinement by using the program AMBER.

3.2.2. Results

Peak volume determination. NOESY data processing needs extreme caution since the integration for those peaks are very sensitive to the baseline of a spectrum. This is particularly

important because the results of integration directly affect the calculation of the initial buildup rate; and the initial buildup rate was used to estimate the proton-proton distances. During the data processing, several different phase corrections were tried. Rephasing portions of H_2O NOESY matrices was also used to reduce noise. The contour plot of a portion of NOESY spectra at five different mixing time is shown in Figure 3.18. The similar intensities of geminal pair proton cross-peaks were observed in NOESY spectra at mixing time of 200 ms and 120 ms which could be due to the saturation of magnetization transfer at 200 ms mixing time. Figure 3.18 shows a portion of contour plot of D_2O NOESY spectrum at five different mixing times. An example of the cross-peak volume plotted versus mixing time for $Asn^{37}\beta H$-βH is shown in Figure 3.19.

Figure 3.18. A portion of the contour plot of NOESY spectra at various mixing times:1 ms, 40 ms, 80 ms, 120 ms and 200 ms shown in A, B, C, D and E, respectively. The peaks pointed by arrows are geminal pair proton cross-peaks.

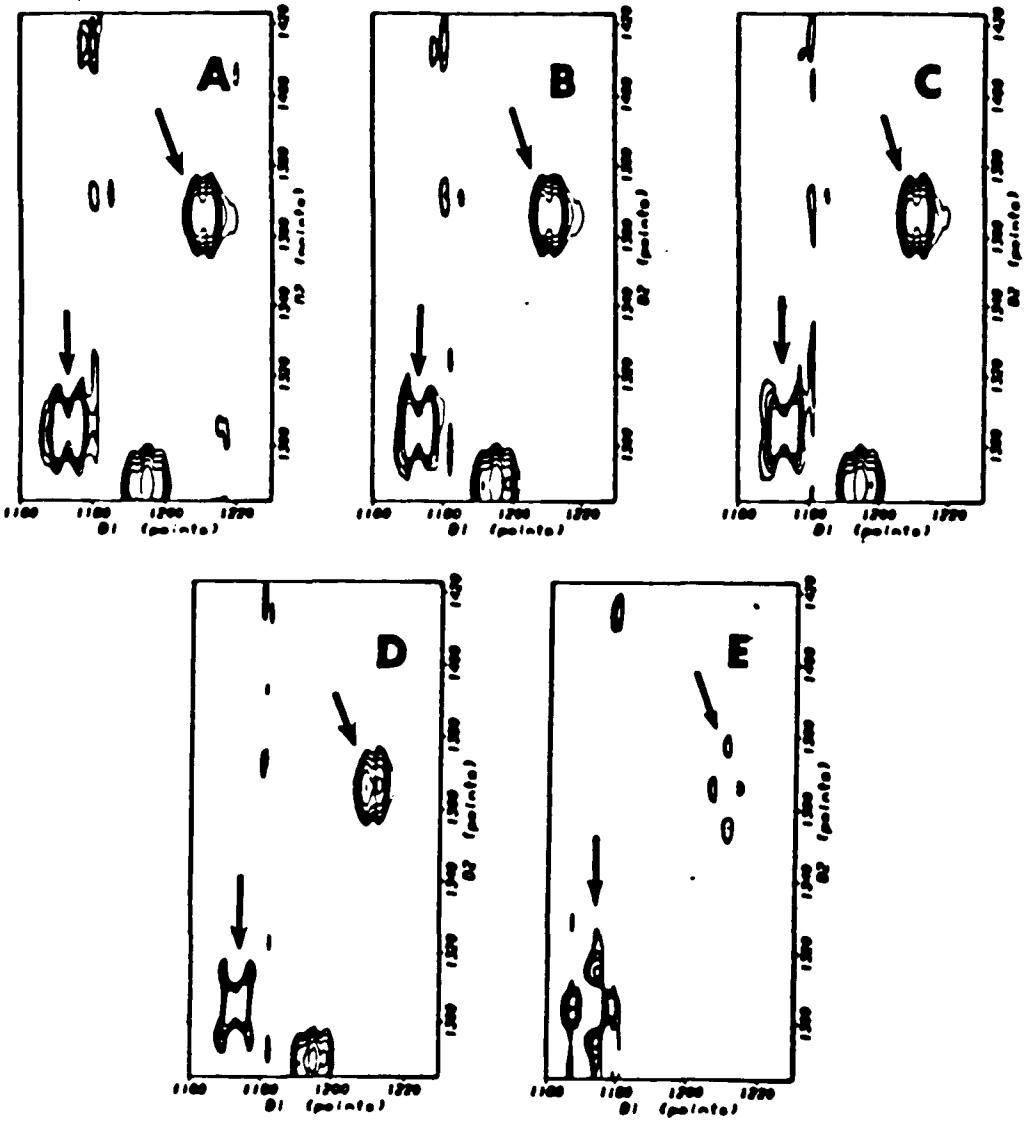
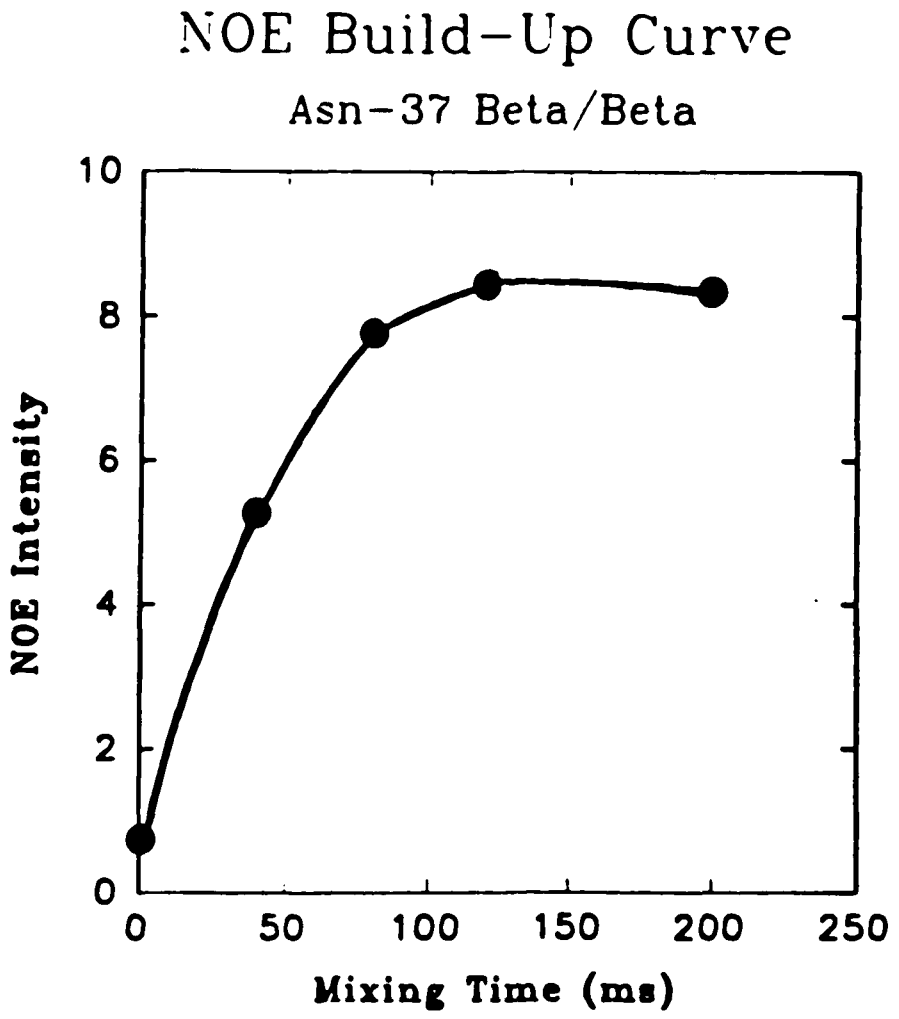


Figure 3.19. Plot of cross-peak volume versus mixing time for the $Asn^{37}\beta H$-βH cross-peak in the D_2O NOESY spectra.



Peak volume measurement. Each peak volume has been measured at 5 different mixing times. The data sheets for the integration results were given with cross-peak location. The example of some geminal pair proton cross-peaks in water NOESY are shown in Table 3.6.

Table 3.6. Some geminal pair cross-Peak volumes from NOESY spectra in water.

Cross-Peak	Integration Volume* at different mixing time				
(point)	200 ms	120 ms	80 ms	40 ms	1 ms
1133/1201	.738E8	.580E8	.398E8	.200E8	-.681E8
1174/1305	.695E8	.604E8	.474E8	.233E8	-.569E8
1197/1292	.485E8	.525E8	.466E8	.267E8	-.264E8
1202/1133	.950E8	.924E8	.663E8	.556E8	.248E8
1212/1368	.542E8	.488E8	.453E8	.280E8	-.210E8
1291/1197	.598E8	.662E8	.594E8	.440E8	.200E8
1304/1174	.989E8	.930E8	.716E8	.584E8	.346E8
1367/1212	.650E8	.571E8	.544E8	.361E8	.403E7

*: Peak volumes are in arbitrary units.

Estimation of the initial NOE buildup rate. To estimate the initial NOE buildup rate, the integration volumes were plotted versus mixing times; then the plots were fitted into an exponential decay formula:

$$a + b * (1 - \exp.\ (- c * x\ ))$$

Here **a** is the y intercept; **b** is the full range of volume; **c** a variable indicating degree of curvature. The limiting slope at a mixing time of 0.0 is given by multiplying **b** by **c**. Glycine residues in the peptide were considered as a best standard for distance calibration, since it has a pair of alpha protons with known distance: 1.75Å (68). Using the limiting slope at mixing time = 0.0, the unknown distances were calculated. Table 3.7 shows the results of some germinal pair proton initial rate slopes. The weaker NOEs usually did not integrate well because of the background noise. Thus, weaker NOEs were assigned in the distance range from 4 to 5 Å (the maximum proton separation that an NOE can be observed for is 5 Å (68)). All constraints derived from NOESY in H_2O and in D_2O are attached with cross-peak assignments in Appendix 1.

Table 3.7. Some geminal pair distance calculation data from NOESY spectra in water.

Amino acid	Peak position (p.t)	Slope*	Å
$Gly^{15}{}_{\alpha H/\alpha H}$	1133/1201	3.13E6	
	1202/1133	9.44E5	1.77
$Gly^{32}{}_{\alpha H/\alpha H}$	1174/1305	3.15E6	
	1304/1174	7.63E5	1.71
$Cys^{38}{}_{\beta H/\beta H}$	1197/1292	2.50E6	
	1291/1197	1.12E6	1.78
$Asn^{37}{}_{\beta H/\beta H}$	1212/1368	2.09E6	
	1367/1212	1.22E6	1.76

*: The slopes were calculated (see text) from the plot of NOE peak volume versus NOESY mixing times.

Stereospecific assignments of $C^{\beta}H_2$ methylene protons: Of the 43 amino acid residues in FIX-EGF1, 34 have C^{β} methylene protons. Of these 34 side chains, 10 have almost degenerate $C^{\beta}H_2$ chemical shifts ($\Delta\delta$ < 0.1 ppm), which include Tyr^1, Asp^5, Glu^8, Ser^9, Asn^{10}, Ser^{17}, Cys^{18}, Asn^{23}, Ser^{24}, and Glu^{26}. Their degenerated methylene proton resonances were confirmed by a TQ-COSY experiment (see Table 3.1.). Of remaining 22 non-proline residues, 9 well-resolved H-H cross-peaks were obtained. About 90% of the buried side chains in globular proteins adopt one of the three staggered conformations with $\chi 1$ = +60°, 180°, or -60° (107).

Surface side chains of FIX-EGF1 may have multiple conformations that interconvert quickly on the NMR time scale. It is important to avoid applying an averaged coupling constant derived from multiple side-chain conformations to this kind of analysis. Thus, when using $^3J(H\alpha\text{-}H\beta)$ vicinal coupling constants as conformational constraints, both $^3J(H\alpha\text{-}H\beta)$ coupling constants ($^3J(H\alpha\text{-}H\beta 2)$ and $^3J(H\alpha\text{-}H\beta 3)$) have to be consistent with one of the staggered rotamer states for c1 (105). After getting both coupling constants, the intraresidue NOEs were used to assign $C\beta H_2$ methylene protons and to identify a staggered rotamer conformation. $H\alpha$ and $H\beta$ atoms of a standard staggered $\chi 1$ rotamer of an amino acid has either a *gauche* or *trans* conformation, with $^3J(H\alpha\text{-}H\beta)$ vicinal coupling constants of 3 ± 2 Hz and 12 ± 2 Hz, respectively (70). Three side chain methylene protons within this range were identified and stereospecifically assigned. The criteria for stereospecificity assignment and $\chi 1$ determination are listed in Table 3.8. The data are summarized in Table 3.9. A portion of the PE-COSY contour plot is illustrated in Figure 3.20.

Figure 3.20. A portion of the PE-COSY contour plot. The measured coupling constants are indicated. (A) H^{α}-$H^{\beta 2}$ PE-COSY cross-peak of Tyr^{25} manifesting $^{3}J(H^{\alpha}\text{-}H^{\beta 2})$. (B) H^{α}-$H^{\beta 3}$ PE-COSY cross-peak of Tyr^{25} manifesting $^{3}J(H^{\alpha}\text{-}H^{\beta 3})$.

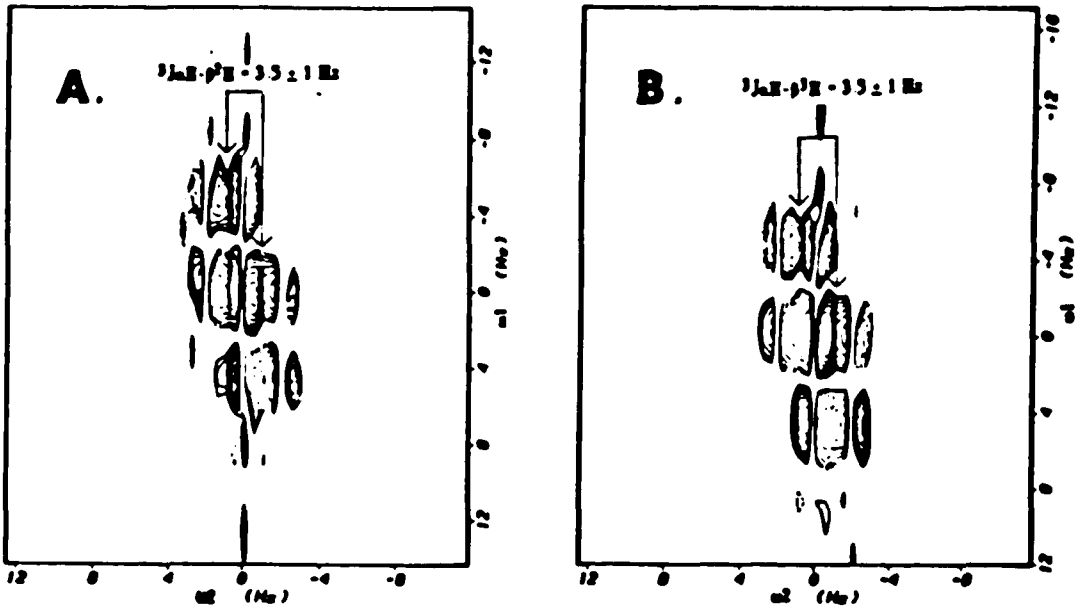


Table 3.8. The criteria of the stereospecific assignment and $\chi 1$ determination. (redrawn from reference 94.)

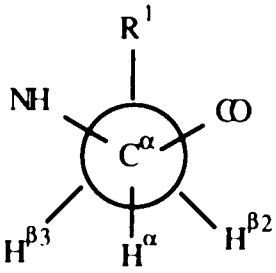
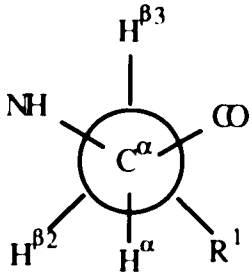
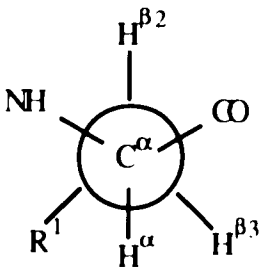
conformation	g^2g^3	g^2t^3	t^2g^3
c1	60°	180°	-60°
			
$^3J\alpha\beta^2$(Hz)	2.6-5.1	2.6-5.1	11.8-14.0
$^3J\alpha\beta^3$(Hz)	2.6-5.1	11.8-14.0	2.6-5.1
NOE (NH, β^2)	weak	strong-medium	strong
NOE (NH, β^3)	strong-medium	strong	weak

Table 3.9. Stereospecific $C^{\beta}H_2$ assignments of FIX-EGF1

Amino acid	Chemical shift (ppm) H^N	H^{α}	H^{β}	$^3J_{\alpha\beta}$ (Hz)	NOE (H^N-H^{β})	Stereo. assignment $\chi 1$	of H^{β}
Cys-7	8.49	4.37	2.89	3.97	500	(A)	
			3.13	10.94	800		
Cys-12	7.73	4.38	2.42	11.30	550	(A)	
			2.75	5.13	1000		
Leu-13	8.51	4.28	1.38	10.25	500	(A)	
			1.59	3.45	1000		
Asn-14	8.58	3.87	1.23	11.96	250	-60°±30°	β2
			1.92	5.06	50		β3
Tyr-25	8.10	5.39	2.74	3.91	400	60°±30°	β3
			3.14	4.38	100		β2
Cys-27	9.22	5.59	2.81	10.80	400	(A)	
			2.98	3.11	900		
Cys-29	8.54	5.13	2.57	5.02	1000	(A)	
			2.93	12.66	500		
Cys-38	7.69	4.05	2.99	11.92	50	180°±30°	β2
			3.60	4.79	25		β3
Leu-40	8.78	4.69	1.35	10.75	300	(A)	
			1.49	5.36	950		

(A) The data are not consistent with any standard rotamer states.

Some residue side chains with well resolved H^{α} and H^{β} cross-peaks can not be stereospecifically assigned due to either coupling constants or intraresidue NOEs that are inconsistent with any single staggered rotamer state (104). This may arise from a nonstaggered conformation or from rapid averaging of two staggered rotamers (105).

Experimental Conformational Constraints: There are five experimental conformational constraints used in the input for the structural determination. They are as follows: (1) NOE-derived upper and lower bound ^{1}H-^{1}H distance constraints; (2) constraints on ranges of backbone dihedral angles ϕ from vicinal $^{3}J(H^{N}\text{-}H^{\alpha})$ coupling constants; (3) constraints on ranges of side-chain dihedral angles $\chi 1$ determined by PE-COSY and NOESY data; (4) constraints from interstrand H-bonds in the β-sheet; and (5) constraints derived from disulfide bonds.

The NOE-derived upper and lower bound distance constraints were obtained as described in Material and Methods section. Backbone vicinal coupling constants $^{3}J(H^{N}\text{-}H^{\alpha})$ were estimated from $\omega 2$ cross-section of H^{N}-H^{α} cross-peaks of a DQF-COSY contour plot as described in Section 3.1. The residues with the corresponding dihedral angle ϕ restricted to the range ($-160\,^{\circ} < \phi > -80^{\circ}$) (68) are listed in Table 3.3. The constraints on the ranges of the side-chain dihedral angles $\chi 1$ are listed in Table 3.10. The $\chi 1$ angle for Val, Ile and Thr corresponding to the $^{3}J_{\alpha\beta} \geq 8.0$ Hz were attributed a range

of ± 30 ° about the *trans* position of αH relative to βH (68). It was not possible to determine the coupling constant for residues Val^{2}, Ile^{22}, and Thr^{43} in the PE-COSY spectrum, so the α/β coupling constants were measured in the DQF-COSY spectrum after zero filling to 16384 points. The results are: Val^{2}, 9.7 Hz; Ile^{22}, 10.4 Hz; and Thr^{43} , 11.5 Hz. Thus, the αHs of these three amino acid residues are *trans* relative to their βHs.

Interstrand hydrogen bonds in the β-sheet structures were located by both the rate of the amide proton exchange and the cross-strand NOEs. The slowly exchanging amide protons for Lys^{19}, Glu^{26} and Trp^{28} were observed, and a number of cross-strand NOEs within the two b-sheet regions were detected. Thus the following hydrogen-bonded pairs were identified on this basis: $Trp^{28}H^{N}$.....$Ser^{17}O'$; $Lys^{19}H^{N}$.....$Glu^{26}O'$; $Glu^{26}H^{N}$.....$Lys^{19}O'$; $Val^{42}H^{N}$.....$Gly^{32}O'$; $Leu^{40}H^{N}$.....$Glu^{34}O'$. The hydrogen-bonded pairs for $Ser^{17}H^{N}$....$Trp^{28}O'$; $Glu^{34}H^{N}$.....$Leu^{40}O'$; and $Gly^{32}H^{N}$.....$Val^{42}O'$ were located based on the interstrand NOEs, since it was not possible to obtain the solvent exchanging rate due to water frequency interference or overlapping with other cross-peak. The disulfide bonds constraints are fixed by S-S bond distance.

Totally, there were 535 conformational constraints used for the calculations of the final structures to represent the solution conformation of FIX-EGF1. Those constraints include 484 NOE-derived distance constraints, 29 dihedral angle constraints, and 22

upper and lower bound distance constraints associated with eight hydrogen bonds and the three disulfide bonds (Table 3.10). The distribution of the distance constraints in the peptide is plotted in Figure 3.21.

Figure 3.21. The distribution of the distance constraints in the peptide. The empty bars indicate the observed the number of short range ($i - j \leq 4$) NOEs, while the solid bars indicate the observed the number of long range ($i - j \geq 5$) NOEs for each amino acid residues.

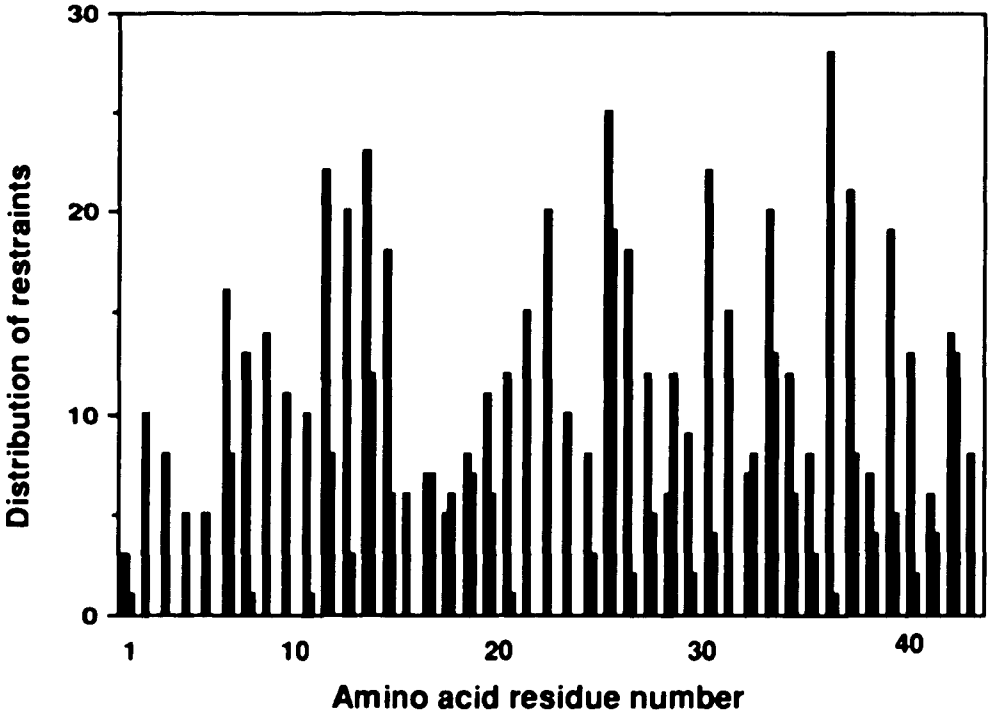


Table 3.10. Numbers of experimental distance constraints used in determining the solution structure of FIX-EGF1.

NOE-derived upper bound			dihedral angle			
constraints			constraints		H-bond	S-S bond
intraresidue	sequential	longer range	backbone	side chain	constraints	constraints
214	151	119	26	3	16	6

Determination of the solution structure: The DSPACE and AMBER calculations were carried out by Jack Skalicky at The University of Colorado. The interpretation of the data was performed in this laboratory. From the distance geometry program DSPACE, **75** structures were generated. Among those structures, **34** structures with DSPACE penalty number < 2.0 were chosen for further energy refinement calculation by the program AMBER. AMBER uses experimental constraints on the upper and lower values of interatomic distances, and of torsion angles to generate molecular structures, assuming standard bond lengths, bond angles, and van der Waals radii. Since distance constraints do not define the chirality of the structure, mirror images (local or global) of correct structure can occur. Among these **34** generated AMBER structures, **10** were chosen for further analysis based on low energy.

The total energy of each AMBER conformer before the refinement is around 10^3-10^7 kcal/mol. After the refinement the

total energy was lowered to approximately -10^2 kcal/mol. The total energy and distance constraint energies of 10 conformers generated from the program AMBER are listed in Table 3.11.

Backbone atoms of β-sheets and β-turns of 10 conformers generated by the program AMBER are superimposed in Figure 3.22.

Figure 3.22. Superimposition of backbone atoms of 10 conformers fragments for the major β-sheet (16-29) in (A), for the minor β-sheet (32-42) in (B), for the β-turn (21-24) in (C) and the β-turn (35-39) in (D). Only the C^{α} atoms are shown.

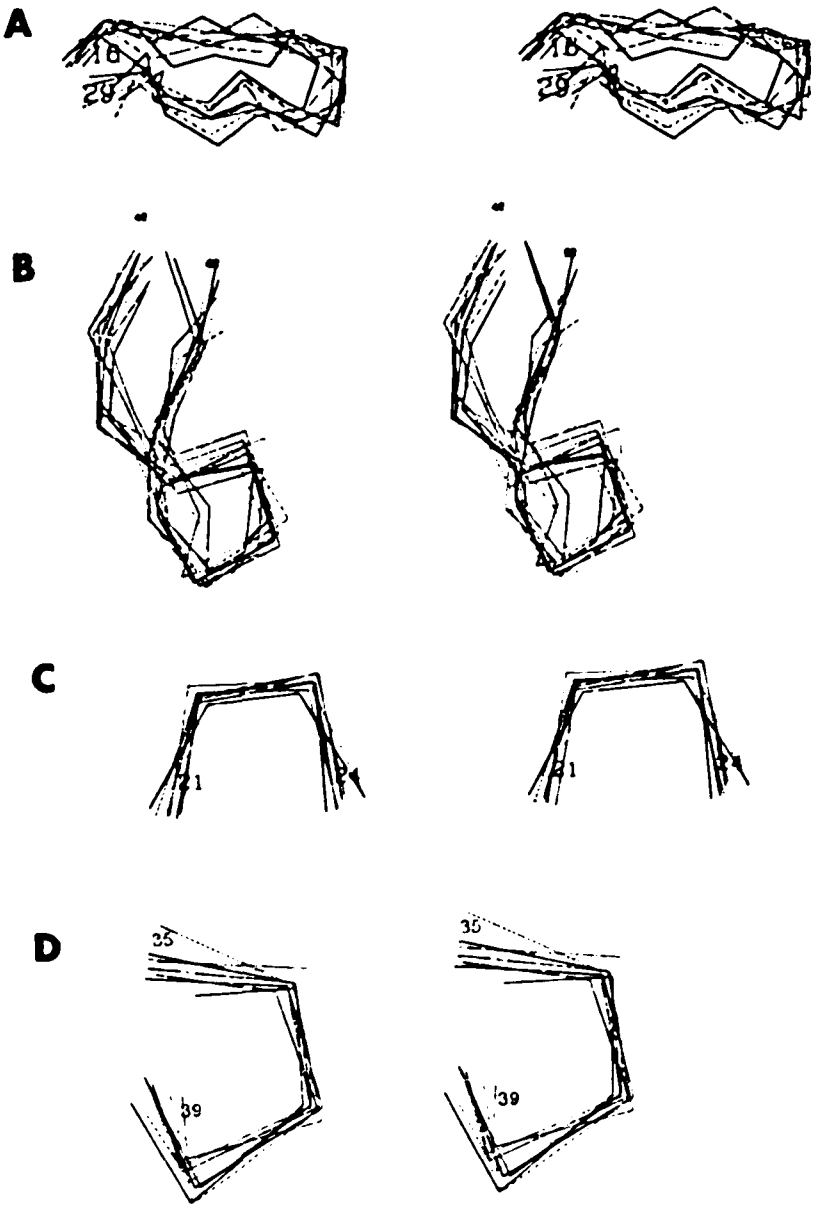


Table 3.11. Summary of conformational energies for the final 10 FIX-EGF1 conformers used to represent the solution structure conformation.

AMBER conformer number	AMBER constraint energy (kcal/mol)	AMBER total energy (kcal/mol)
1	18	-375
2	76	-150
3	45	-230
4	27	-295
5	16	-369
6	47	-235
7	66	-116
8	80	-155
9	35	-277
10	72	-100

Possible hydrogen bond locations for the observed slowly exchanging amide protons were examined among these 10 conformers with the program InsightII (Biosym). A low consistency for the proton acceptors among the 10 conformers was found, except those H-bonds in the β-sheets applied as constraints for the structure calculations. Therefore, little information can be drawn for the structural roles of the slowly exchanging amide protons outside of β-sheets region at this moment.

3.3. Discussion

Solution structure of FIX-EGF1: The main features of the secondary structure of FIX-EGF1 are a major antiparallel b-sheet connecting residues 16-20 and 25-29, and a second sheet connecting residues 32-34 and 40-42. Neither turn, a four residue-turn with the first sheet, and a five residue-turn with the second sheet, were classified as standard type I or type II turn after consideration of the amino acids involved, the observed NOESY patterns, and the magnitude of $^{3}J_{HN\alpha}$ coupling constants (see Table 3.4 for details) (68).

There are two subdomains in this structure. One consists of amino acid residues 1-28, and another consists of 29-43. Each domain has an antiparallel β-sheet associated with a non-standard β-turn. Table 3.12 lists all NOEs between the two subdomains. Figure 3.23 gives a representative drawing with indication of the positions for residues 12-14, and 35-39. The NH_2-terminal tail (residues 1-6) is not well defined.

Figure 3.23. The representative ribbon drawing for the structure of FIX-EGF1 with the indication of residues 12-14, and 35-39.

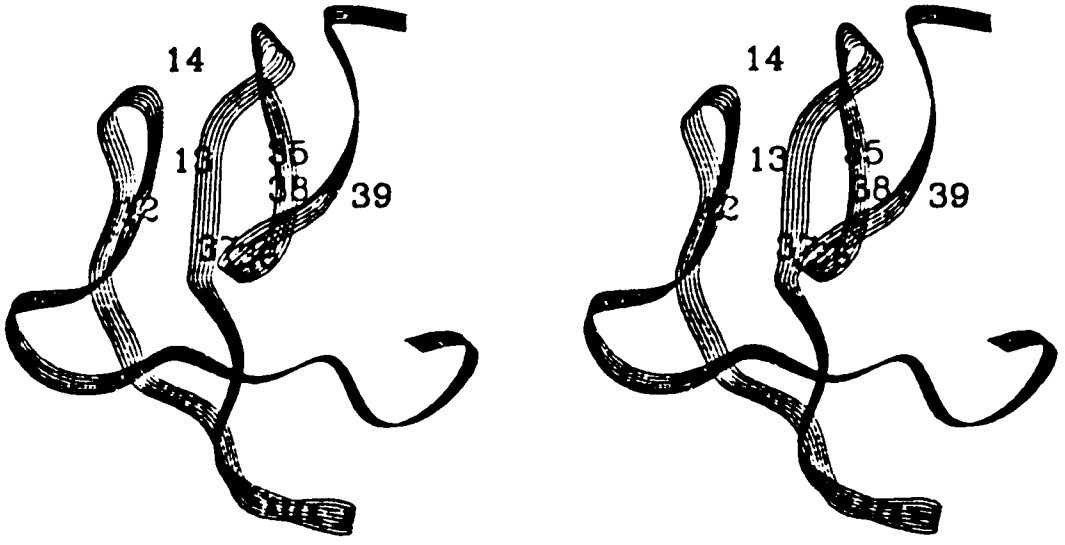


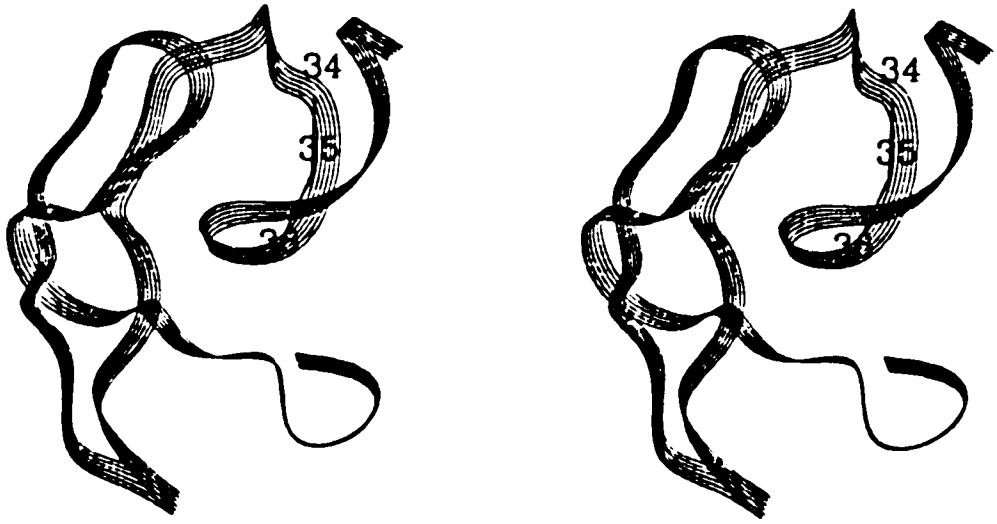
Table 3.12. The observed NOEs between the two subdomains.

12αH-37αH	1.8-4.2Å
12αH-37βrH	1.8-3.9Å
13NH-37βrH	1.8-6.5Å
14NH-35NH	1.8-5.5Å
14βH-35NH	1.8-3.5Å
35NH-13βsH	1.8-5.0Å
35NH-13βrH	1.8-4.3Å
37βHr-13γH	1.8-6.4Å
37αH-13mδHr	1.8-8.0Å*
37βHs-13γH	1.8-5.5Å
38βH-14βH	1.8-6.0Å
38βH-13βHr	1.8-6.5Å
39αH-13βHr	1.8-3.8Å
39αH-13βHs	1.8-5.0Å
39αH-13γH	1.8-4.7Å
39αH-13mδHr	1.8-7.6Å*
39αH-13mδHs	1.8-7.4Å*

* indicates the distances for methyl group protons which are multiplied by 1.3.

As showing in Figure 3.24 the C-terminal tail of the molecule turns under the fragment 34-36, while in the EGF-like module of factor X and in mEGF the C-terminal tail goes over this section (109-110).

Figure 3.24. The representative ribbon drawing for the structure of FIX-EGF1 with the indication of the fragment of 34-36.



Evaluation of the quality of the structure: The quality of the structure determination within the two β-sheets is better than for the overall molecules. Superpositions of the C^{α} coordinates for the β-sheets and β-turns of 10 AMBER structures are shown in Figure 3.22. The root-mean-square deviations (RMSDs) for the superposition of pairwise backbone atoms of four fragments are presented in Table 3.13.

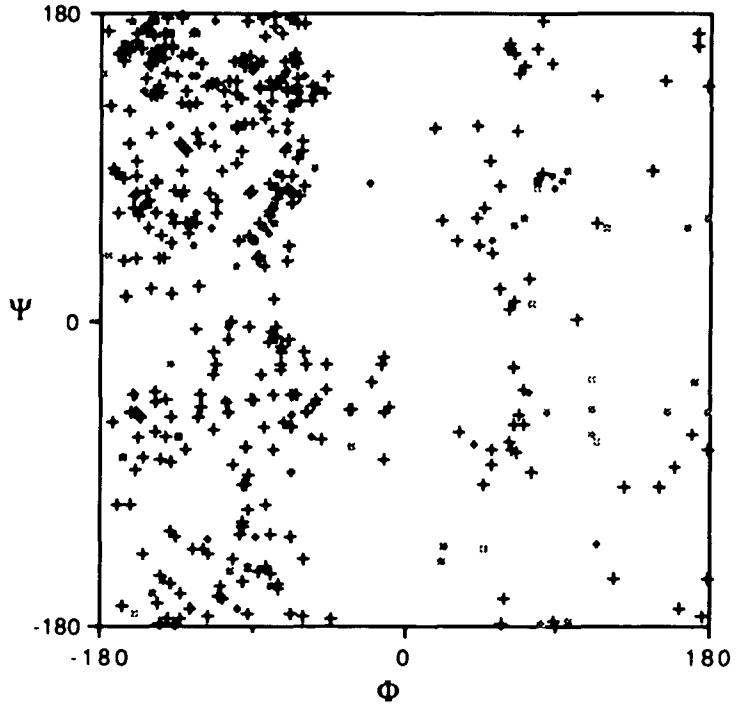
Table 3.13. Average RMSDs of the 10 FIX-EGF1 conformers generated by AMBER from one of the conformer

Atoms used for RMSD calculation	Average RMSD (Å)
N, C^{α}, C' of residues 16-29	1.71 ± 0.5 Å
N, C^{α}, C' of residues 32-42	1.42 ± 0.3 Å
N, C^{α}, C' of residues 21-24	0.8 ± 0.2 Å
N, C^{α}, C' of residues 35-39	0.73 ± 0.2 Å
N, C^{α}, C' of residues 1-43	3.68 ± 0.8 Å

Ramachandran plot (ϕ–φ) is a useful rule for evaluating the quality of a protein structure determination, since amino acid residues in proteins generally adopt backbone conformations corresponding to low-energy minima of amino acid residues in small peptides (109). Figure 3.25 displays the correlation between the ϕ and φ angles of the residues 2-42 of 10 AMBER structures. The scatters in this plot indicate the distribution of the

conformations of the peptide backbone in AMBER structures. Energetically favorable region expected for the extenuated conformation is found in the upper left quadrant. Glycyl residues (indicated by square signs) are more dispersed in the right half of the plot, since their greater conformational flexibility. Comparing the Ramachandran plot for the mEGF structure (109), the distribution of the conformation for FIX-EGF1 and mEGF are very similar. In the case of the EGF-like modules, this plot may not be a good indication for the quality of the structure.

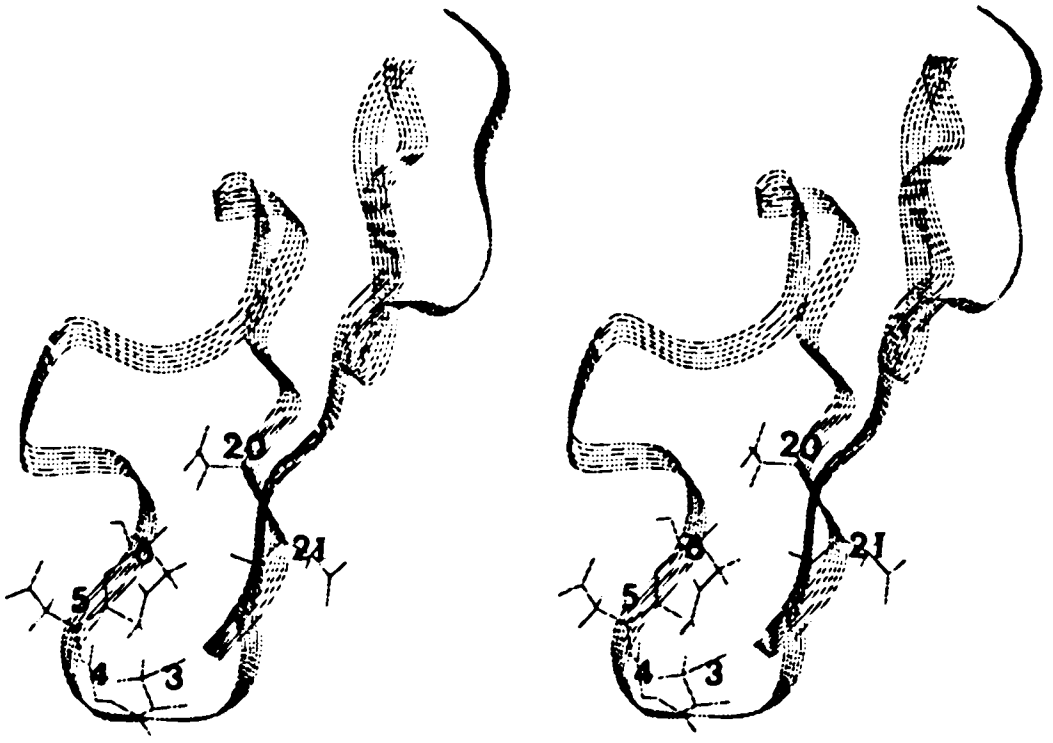
Figure 3.25. Ramachandran plot for residue 2-42 in 10 calculated AMBER structures. Glycyl residues are plotted by squares while all others are indicated by plus signs.



Calcium binding site: If chemical shifts of protons on calcium binding ligands will be affected by adding calcium ion, the proposed calcium binding site was suggested to the region of residue 20-21, 24-26, and 4-7. Asp^{20} (Hya^{64} in the protein, Hya: β-OH-Asp) has been previously considered as a calcium binding ligand in factor IX since Asn at position 20 instead of β-OH-Asp leads to severe hemophilia **B** with very low clotting activity. Genetic defects on position 46, 48, 49, 63 of factor IX (corresponding to position 3, 5, 6, 20 in FIX-EGF1) lead to severe hemophilia B with very low clotting activity (37). This indicates that these residues are important for the clotting activity. They may be involved in the calcium binding site, or may be very close to the binding site. Recently, a high resolution NMR structure of the calcium-bound form of the NH_2-terminal EGF-like module in factor X was reported (110). The calcium binding cavity in this module was found, which consists of four oxygens: backbone carbonyls of Gly^{47} and Gly^{64} as well as the side chain carbonyl of Gln^{49} and one of the carbonyl oxygens of Hya^{63}. In FIX-EGF1, this corresponds to residues Gly^{4}, Asp^{21}, Gln^{6} and Asp^{20}, respectively. Earlier work based on calcium-dependent chemical shifts suggested involvement of Ser^{24} or Tyr^{25} in the calcium binding (112). If the residues Ser^{24} and Tyr^{25} (corresponding to 67-68 in factor X) are not involved, the calcium induced chemical shift changes may be due to the conformation change on the Asp^{20}-Asp^{21} side of the β-sheet.

Figure 3.26 shows that the residues Asp^{3}-Gln^{6}, Asp^{20} and Asp^{21} are in close proximity to each other.

Figure 3.26. The representative ribbon drawing of the FIX-EGF1. The side chains of Asp^{3}-Gln^{6}, Asp^{20} and Asp^{21} are shown. The numbers indicate the residue positions.



If Asp^5 is not involved directly in the calcium binding site, the observed chemical shift change at this residue (substantial chemical shift change at a-proton) may be due to neighboring conformational changes of Gly^4 and Gln^6. It has been suggested that alteration of the positions of a single surface charge close to the calcium binding site can reduce calcium binding affinity (110). This may be the reason that mutation of Asp^5 to Glu causes severe hemophilia **B** with low clotting activity. Further study is necessary to identify its side chain location in space with the calcium ion present to confirm this explanation.

Mutation at the position of Asp^3 to Gly or Glu shows it is an important residue for clotting activity (27, 35). The NMR evidence of this study showed no observable chemical shift change after addition of calcium ion to FIX-EGF1 (112). However, the work of others (110) showed that the side chain of Asp^3 points toward the calcium binding site. This strongly suggests that Asp^3 is involved in the calcium binding in factor X. The different results for calcium binding by Asp^3 may be due to pH difference (pH 4.7 for factor IX-EGF1 versus pH 5.8 for the EGF-like module of factor X).

The function of β-hydroxyaspartic acid: The peptide FIX-EGF1 has aspartic acid at residue 20 (corresponding to residue 64 in human FIX protein) instead of a β-hydroxyaspartic acid. It has been demonstrated that β-hydroxyaspartic acid may not be required for calcium binding (34). This result has been confirmed

by the calcium dissociation constants of 0.4 mM (42), and 0.2-0.3 mM (43) from chemical synthetic and yeast-generated peptides with the first EGF-like module sequence without β-hydroxyaspartic acid, respectively. And it has been pointed out that the hydroxyl group points away from the putative calcium ion (110). The β-hydroxyaspartic acid in factor IX may contribute to the calcium binding affinity, but it is not necessary for calcium binding.

The effect of calcium binding: Calcium binding to proteins can either stabilize a protein's conformation or function as a joint to connect a protein to another molecule. In the case of FIX-EGF1, after addition of calcium to the peptide solution, the chemical shifts of most resonances are almost unchanged. This indicates that calcium ion does not induce any major conformational change in the peptide.

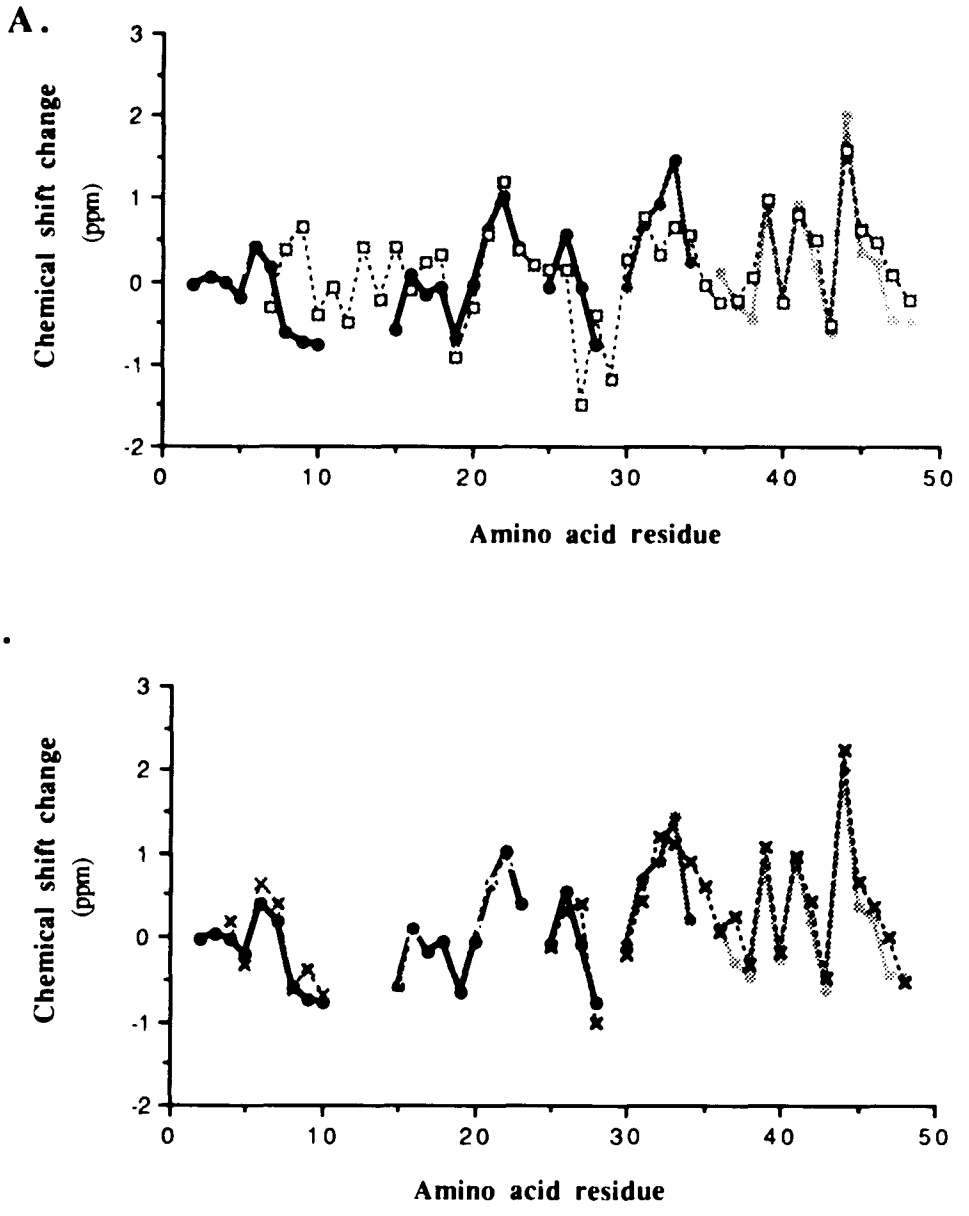
The calcium binding pocket on FIX-EGF1 is located on the major β-sheet and NH_2-terminal tail. The possible role of calcium is to make the NH_2-tail (Tyr^1-Gln^6) close to the β-sheet. It is known that the Gla-domain is connected to the membrane surface (50), possibly through the interaction of calcium and phospholipid. The NH_2-tail of FIX-EGF1 acts as a linker between Gla-domain and the first EGF-like module. The NH_2-terminal calcium-induced shift does not affect the global conformation of the peptide, but it may be important for the relative positioning of the NH_2-terminal neighboring region-Gla domain. From the structure of the calcium-bound EGF-like domain of factor X, it is clear that after binding to

calcium, the position of the NH_2-terminal tail is much better defined than the calcium-free form (110).

Is there a second binding site? Glu^{34} and Val^{42} are other two residues which showed noticeable sensitivity to the addition of calcium ion. Handford et al. reported a second aromatic resonance (corresponding to Phe^{33} in FIX-EGF1) in a shorter peptide (corresponding to residue 2 to 40 of FIX-EGF1) that showed sensitivity to calcium ion, but it was also sensitive to magnesium ion (43) with lower selectivity. Since the shorter peptide studied by Handford et al. has no Val^{42}, the consequence of calcium-dependent chemical shift of Phe^{33} is not clear. Based on structures gotten from distance geometry calculation, Glu^{34} and Val^{42} are too distant to participate in the calcium-binding site in the major β-sheet region. There is not enough evidence to make a conclusion on the presence of a second binding site involving Glu^{34} and Val^{42} in FIX-EGF1 at this moment.

Structural comparison between FIX-EGF1, FX-EGF and mEGF: The secondary shift is defined as the difference of the chemical shift of a proton resonance of an amino acid observed in the native protein minus its value found in a random coil peptide (109,113-115). Comparisons have been made for the amide protons of FIX-EGF1 vs. those of the first EGF-like module of FX, and similarly for FIX-EGF1 and murine EGF. The results are plotted in Figure 3.27. The results indicate that there are strong similarities in structure among these modules.

Figure 3.27. The comparison of ^{1}H chemical shift for amide protons of FIX-EGF1 (filled circle), vs. those of mEGF (box) (A), and of FIX-EGF1 (filled circle) vs. those of the EGF-like module of factor X (cross sign)(B).



The solution structures of these molecules are very similar, however, there are two significant differences. The C-terminal tail of the molecule turns under the residues 34-36, while in the EGF-like module of factor X and in mEGF, the C-terminal tail goes over this section. Secondly, the degree of the twist in the major β-sheet is different among these three molecules. Those conformational differences may be important for their functional specificity.

Appendix: Distance constraints for solution structure determination of FIX-EGF1.

The constraints from NOEs and slowly exchanging amide protons.

The experimental interproton distance was loosely classified into one of three groups: short (d<2.8Å), medium(2.8Å<=d<=3.8Å), or long(d>3.8Å). The upper and lower bounds for short distances were determined by adding ±0.5 Å to experimental data. The upper bounds for medium distances were generated by adding 1.0 Å and the lower bounds by adding -0.5 Å. The upper and lower bounds for a long distances were modified by adding ± 1.0 Å. Hd and He are used for Tyr and Phe aromatic ring proton 2,6H and 3,5H respectively. No other correction is added for this distance correction for methyl groups are: multiplying 1.2 for all methyl proton related cross peaks, execpt Leu40. x1.348 for Leu40 methyl related cross peaks. In addition, 1.0 Å is added to upper bounds for each methyl group involved in a distance constraint. In place of stereospecific assignments all pro chiral center are called "r" or "s" based upon their relative chemical shifts. The upfield resonance is "s" and the downfield resonance is "r".

Inter-residue NOE's

ha[1] hgr[26] 1.8 6.0 ; 1102/1390
(1);hd1[1] mgs[2] 1.8 10.0 ; 685/1595,
hn[2] hbs[3] 1.8 5.0 ; 492/1328, hb[25] 1328
hn[2] ha[1] 1.8 3.23 ; 492/1103
hn[3] mgs[2] 1.8 7.11 ; 488/1597
hn[3] ha[2] 1.8 3.20 ; 488/1123
hn[3] hb[2] 1.8 5.0 ; 488/1433
(1);hn[4] ha[9] 1.8 4.57 ; 499/1092
hn[4] hbs[3] 1.8 4.22 ; 499/1328, not fit very well
hn[4] ha[3] 1.8 4.05 ; 499/1056, not fit very well
hn[5] hbr[3] 1.8 5.0 ; 522/1319
hn[5] har[4] 1.8 4.07 ; 522/1142
hn[6] hn[5] 1.8 5.0 ; 434/522
hn[6] hn[7] 1.8 4.16 ; 434/480, not fit very well
hn[7] hbr[6] 1.8 4.5 ; 480/1405
hn[7] ha[6] 1.8 4.27 ; 480/1135

hn[7] hbr[6] 1.8 6.87 ; 480/1443, can't get from buildup,
hn[7] hn[8] 1.8 2.89 ; 480 588
ha[7] hbr[8] 1.8 6.0 ; 1083/1419
ha[7] hgs[6] 1.8 6.0 ; 1083/1401
(1);ha[7] hbs[9] 1.8 6.0 ; 1083/1179
ha[7] hgr[11] 1.8 6.0 ; 1083/1533
hn[8] hn[9] 1.8 3.20 ; 587/607, can't get from buildup
hn[8] ha[7] 1.8 5.0 ; 589/1087
ha[9] hbr[10] 1.8 6.0 ; 1091/1304. Changed from hbs[7] on 3/5/92.
ha[9] hbr[8] 1.8 6.0 ; 1091/1418 **.
(1);ha[9] hg[13] 1.8 8.0 ; 1091/1462. out 7/2/92
ha[9] hgr[8] 1.8 6.30 ; 1090/1369
(1);hbs[9] hbr[12] 1.8 6.0 ; 1179/1323, 1179 could be ha[22] needs to be checked
7/7/92
hn[9] hbs[8] 1.8 4.02 ; 607/1425
ha[9] hbs[11] 1.8 4.01 ; 1090/1495
ha[10] hdr[11] 1.8 2.66 ; 1072/1214
hn[10] hdr[11] 1.8 3.18 ; 556/1216
hn[10] ha[9] 1.8 3.28 ; 557/1147
hn[10] hn[9] 1.8 3.89 ; 557/606, can't get from buildup
(1);ha[11] hg[13] 1.8 4.01 ; 1106/1463
ha[11] mdr[13] 1.8 6.10 ; 1106/1596, x1.2 for correction
ha[11] mds[13] 1.8 6.14 ; 1106/1611, x1.2 for correction
hdr[11] hbr[10] 1.8 4.10 ; 1214/1305
hdr[11] hbr[18] 1.8 3.92 ; 1214/1279 **
(1);ha[11] hbr[39] 1.8 6.0 ; 1106/1412, 1412 could be hb[22]?
(1);ha[11] hb[22] 1.8 6.0 ; 1106/1416. Could be just the internal
ha[12] mdr[13] 1.8 6.8 ; 1083/1594, x1.2 for correction,
hbs[12] mds[13] 1.8 8.0 ; 1371/1612, 1371 could be hbr[8]
ha[12] hgr[11] 1.8 6.0 ; 1083/1534
ha[12] hgs[11] 1.8 6.0 ; 1083/1631
ha[12] hg[13] 1.8 6.0 ; 1083/1463
(1);ha[12] hbs[39] 1.8 6.0 ; 1083/1439, 1439 could be hbr[14]
ha[12] hg[13] 1.8 6.0 ; 1083/1461
ha[12] ha[37] 1.8 4.22 ; 1083/893

hbs[12] has[16] 1.8 3.20 ; 1371/1182
hbr[12] has[16] 1.8 4.0 ; 1323/1182
ha[12] hbr[37] 1.8 3.93 ; 1084/1212, 1084 could be ha[7],
ha[12] hbs[11] 1.8 4.20 ; 1084/1495, visual, change from
hn[12] hgr[11] 1.8 4.5 ; 594/1537
hn[12] hn[13] 1.8 5.0 ; 593/480. changed from hn[7] to hn[13].
hn[12] hn[10] 1.8 5.5 ; 594/556
hn[12] hbs[11] 1.8 5.5 ; 593/1497.change from hbr[13] to hbs[11]
hn[12] ha[11] 1.8 5.0 ; 593/1108
hn[12] hdr[11] 1.8 3.25 ; 593/1216, only see one hd for [11]
hn[13] hbs[12] 1.8 4.33 ; 480/1376, 480 could be hn[7],
hn[13] hbr[12] 1.8 4.50 ; 480/1326, not a good fit
hn[13] hbr[37] 1.8 6.5 ; 480/1214, 480 could be hn[7], 1216
hd2z[14] mdr[13] 1.8 8.2 ; 675/1600. changed from mgs[2]7/2/92.
hd2e[14] hbr[13] 1.8 5.5 ; 669/1527. Changed from hbs[40] to
hd2e[14] mdr[13] 1.8 7.6 ; 669/1600. changed from mgs[2]7/2/92.
hn[14] hn[35] 1.8 5.5 ; 469/536
hn[14] hn[16] 1.8 5.5 ; 468/592
hd2e[14] hbr[13] 1.8 7.5 ; 669/1495, 1495 could be hbr[11]
hn[14] mds[13] 1.8 5.64 ; 469/1613, x1.2 to correct distance.
hn[14] ha[13] 1.8 2.85 ; 469/1101
hn[14] hbs[13] 1.8 2.90 ; 469/1526
hn[14] hbr[13] 1.8 4.00 ; 469/1495, 1495 could be hbs[11]
hn[14] hn[15] 1.8 3.00 ; 469/506, can't get from buildup
hn[15] ha[14] 1.8 3.20 ; 506/1160
hn[15] hn[16] 1.8 3.00 ; 506/592
har[16] hdr[30] 1.80 3.79 ; 1062/1218
har[16] hds[30] 1.8 4.70 ; 1062/1250, 1250 could be
has[16] hdr[30] 1.8 4.35 ; 1181/1217
has[16] hds[30] 1.8 4.10 ; 1182/1251, 1251 could be
hn[16] has[15] 1.8 5.3 ; 593/1203
hn[16] hbr[28] 1.8 7.5 ; 592/1253. changed from hn[12]7/7/92.
hn[17] hn[16] 1.8 4.9 ; 440/592
hn[17] har[16] 1.8 3.00 ; 441/1065, not fit very well
hn[17] hn[28] 1.8 5.5 ; 441/330

hbs[17] hbs[28] 1.8 4.02 ; 1153/1251
hn[17] he3[28] 1.8 9.5 ; 441/611, not fit well ub was 11
(1);hn[17] mdr[13] 1.8 8.16 ; 439/1612, needs to be checked
hn[18] hbs[17] 1.8 4.30 ; 417/1152, can't get from buildup
hn[18] hd1[25] 1.88.9 ; 417/726. ub was 9.4, 7/9/92
hn[18] hn[20] 1.8 6.5 ; 417/437. out on 7/2/92 big violation.
hn[18] hbr[12] 1.8 6.0 ; 417/1326. changed from hbs[25]. 3/5/92
ha[18] hbr[7] 1.8 5.55 ; 974/1266
ha[18] hbr[27] 1.8 4.53 ; 974/1289
ha[18] hbs[27] 1.8 5.3 ; 974/1314, visual
hbr[18] ha[10] 1.8 7.03 ; 1280/1072, 1072 from ha[30] change to
hbr[18] ha[7] 1.8 5.95 ; 1280/1084, 1084 could be ha[12]
ha[18] ha[27] 1.8 3.6 ; 975/907
(1);ha[18] mdr[40] 1.8 8.0 ; 973/1597, 1597 could be mdr[13]
(1);ha[18] hbr[11] 1.8 6.0 ; 973/1477. change from hbs[19] 7/2.
ha[18] hbs[17] 1.8 6.0 ; 974/1152, 974 could be ha[29]
hn[18] hbr[18] 1.8 2.91 ; 417/1282
hn[19] ha[18] 1.8 3.1 ; 345/977
hn[19] hbr[18] 1.8 5.0 ; 345/1282
(1);hza[19] hd2z[10] 1.8 9.88 ; 646/755.not fit for any assigment
hn[19] hn[26] 1.8 6.0 ; 346/401
hn[19] hbr[25] 1.8 6.0 ; 345/1268. out due to violation,7/2/92.
(1);hn[19] hbs[25] 1.8 6.0 ; 345/1327, 1327 could also be hbs[3]
hn[19] ha[27] 1.8 6.0 ; 345/910
ha[20] ha[25] 1.8 3.5 ; 1039/936
hbr[20] mg2[22] 1.8 8.20 ; 1299/1596, x1.2 for correction,
hn[20] hgr[19] 1.8 4.8 ; 436/1539
hn[20] hbr[19] 1.85.4 ; 435/1470
hn[20] hbs[19] 1.8 5.8 ; 435/1479
hn[20] hbs[19] 1.8 4.60 ; 438/1477, can't get from buildup
hn[20] hbr[19] 1.8 4.60 ; 438/1469, can't get from buildup
hn[20] hbs[20] 1.8 3.00 ; 437/1334,
(1);ha[21] ha[19] 1.8 6.0 ; 998/1042.many choices on 1042, 7/2
(1);ha[21] hbr[3] 1.8 6.0 ; 999/1317 near 1317 is D5,D41,C27

(1);ha[21] her[19] 1.8 6.0 ; 999/1330, hbs[20] or hbs[3] or hes[36] are around 1327-1328
ha[21] hg1r[22] 1.8 6.0 ; 998/1503, 1503 could be hdr[19]
ha[21] hg1s[22] 1.8 6.0 ; 997/1551
ha[21] mg2[22] 1.8 9.0 ; 998/1598, x1.2 for correction
hn[21] ha[20] 1.8 5.0 ; 504/1042
hn[21] ha[25] 1.8 4.00 ; 504/939, 504 could be hn[31]
hn[22] hg1r[22] 1.8 5.9 ; 442/1597, x1.2 for correction
hn[22] ha[21] 1.8 5.25 ; 442/998
hn[22] hbr[21] 1.8 4.65 ; 442/1354
hn[22] hbs[21] 1.8 3.08 ; 441/1305
hn[22] hn[21] 1.8 3.67 ; 442/504
hn[23] ha[22] 1.8 3.84 ; 452/1180
hn[23] hn[24] 1.8 3.16 ; 452/608
hn[23] hb[22] 1.8 3.27 ; 452/1416
hn[23] hg1r[22] 1.8 5.6 ; 451/1598, x1.2 for correction
ha[23] mg2[22] 1.8 7.30 ; 1025/1597, visual, x1.2 for
ha[24] hgr[6] 1.8 6.00 ; 1057/1375, 1057 could be ha[3]
ha[24] hbr[6] 1.8 6.0 ; 1058/1404
hn[24] hgr[6] 1.8 4.75 ; 608/1375. Changed from hn[9] to hn[24]
ha[24] hd1[25] 1.8 8.5 ; 1057/719, ub was 9.2
hn[24] hn[25] 1.8 5.25 ; 608/542
(1);ha[24] ha[1] 1.8 6.5 ; 1058/1100, very weak, not sure.7/2/92
hn[25] ha[24] 1.8 4.05 ; 542/1061
he1[25] ha[24] 1.8 9.0 ; 723/1058,ub was 10.0
hd1[25] hgr[6] 1.8 8.22 ; 719/1376, up was 9.02
hd1[25] hbr[6] 1.8 7.25 ; 719/1403, ub was 9.05
he1[25] hbr[6] 1.8 9.12 ; 724/1404. Changed from hgr[30] to ,ub was 9.92
he1[25] hgr[6] 1.8 9.30 ; 722/1375,could be Y1 @ 722 was 9.6
he1[25] hbs[6] 1.8 9.10 ; 723/1442, ub was 9.9
hd1[25] hgs[11] 1.8 9.0 ; 718/1632,
hd1[25] hdr[11] 1.8 9.0 ; 719/1212,
hd1[25] hgr[11] 1.8 8.37 ; 719/1535, ub was 9.17
he1[25] hbr[11] 1.8 8.60 ; 723/1479,ub was 9.4
he1[25] hbs[11] 1.8 8.50 ; 723/1494,ub was 9.3

he1[25] hdr[11] 1.8 8.79 ; 725/1535,ub was 9.0
he1[25] hgs[11] 1.8 9.30 ; 722/1632,ub was 9.1
ha[25] hbr[21] 1.8 5.0 ; 936/1303
ha[25] hbs[21] 1.8 4.85 ; 937/1351
ha[25] hgs[26] 1.8 6.0 ; 937/1400, 1400 could be hgs[39]
he1[25] ha[26] 1.8 9.0 ; 723/1039,ub was 8.5
hd1[25] ha[26] 1.8 8.20 ; 718/1039,ub was 9.0
hbr[25] ha[26] 1.8 4.21 ; 1266/1041, 1041 could be ha[20]
he1[25] hbs[27] 1.8 8.87 ; 722/1314,
he1[25] hgr[36] 1.8 8.44 ; 725/1519, ub was 9.5
he1[25] ha[37] 1.8 9.14 ; 724/893,ub was 9.0
hn[26] hd1[25] 1.8 9.0 ; 401/720,
hgs[26] hbs[25] 1.8 3.96 ; 1401/1327, 1401 could be hgr[6]
(1);ha[26] hbr[18] 1.8 5.5 ; 1039/1278, 1039 could be ha[40], ha[20]
hn[26] hbr[25] 1.8 3.8 ; 401/1267
hn[26] hbs[25] 1.8 4.75 ; 401/1326
hn[26] ha[25] 1.8 3.4 ; 401/937
hn[27] ha[26] 1.8 4.70 ; 374/1042, cannot get from buildup
hn[27] hbr[26] 1.8 4.85 ; 375/1428, can't get from buildup
hn[27] hbs[26] 1.8 4.85 ; 375/1438, can't get from buildup
ha[27] hbs[28] 1.8 6.0 ; 906/1252
ha[27] hbr[18] 1.8 5.24 ; 907/1279,
hn[28] ha[27] 1.8 3.97 ; 330/910
he1[28] hbs[19] 1.8 6.0 ; 250/1477, ub was 12
he1[28] hbr[19] 1.8 6.3 ; 250/1470 ub was 12.3
hh2[28] hbs[19] 1.8 9.5 ; 671/1477 ub was 12.5
he3[28] hbr[17] 1.8 6.03 ; 611/1152
hd1[28] hbr[33] 1.8 11.55 ; 670/1274
ha[29] har[16] 1.8 3.13 ; 974/1063
ha[29] has[16] 1.8 3.5 ; 974/1184
ha[29] hdr[30] 1.80 2.75 ; 974/1219
ha[29] hds[30] 1.8 2.94 ; 974/1250
(1);ha[29] hbs[38] 1.8 6.0 ; 974/1201, 974 could be ha[18]
ha[29] hgs[30] 1.8 6.0 ; 974/1432
ha[29] hgr[30] 1.8 6.0 ; 974/1404

hdr[30] hbr[29] 1.8 4.20 ; 1219/1299, 1299 could be hbr[29].
hn[31] hbs[30] 1.8 4.06 ; 505/1414
hn[31] hbr[30] 1.8 4.05 ; 504/1363
hn[31] ha[30] 1.8 3.00 ; 504/1072
hn[31] hn[32] 1.8 4.00 ; 505/540, can not get from buildup
hd1[31] ha[30] 1.8 10.0 ; 665/1072
hd1[31] ha[30] 1.8 6.0 ; 666/1071, ub was 8.91
has[32] mgs[42] 1.8 7.36 ; 1304/1648, x1.2 for correction
has[32] hb[42] 1.8 4.0 ; 1304/1456
har[32] mgs[42] 1.8 9.0 ; 1173/1648, x1.2 for correction
har[32] hb[42] 1.8 6.0 ; 1173/1455
has[32] mgr[42] 1.8 9.0 ; 1305/1623, 1305 could be hbr[21]
har[32] mg2[43] 1.8 9.0 ; 1173/1569, x1.2 for correction
(1);def/bou hbr[33] hbs[39] 4.0 6.0 ; 1275/1441, 1441 could be hbr[14]
hn[32] hn[33] 1.8 2.80 ; 540/589
hn[32] ha[31] 1.8 2.67 ; 540/1109
hn[32] hbr[31] 1.8 4.85 ; 540/1257
hn[32] hb[42] 1.8 4.75 ; 539/1457, can't get from buildup
hn[32] mgs[42] 1.8 7.62 ; 539/1649, x1.2 for methyl group
hn[33] has[32] 1.8 3.90 ; 589/1306
hn[33] har[32] 1.8 4.80 ; 589/1172
ha[33] hbs[34] 1.8 6.0 ; 915/1446, 1446 could be hbr[14]
ha[33] hgr[34] 1.8 6.0 ; 915/1407, hbr[39] is 1409
(1);ha[33] hbr[27] 1.8 6.0 ; 916/1291
he1[33] hbr[28] 1.8 10.0 ; 675/1233,
he1[33] hbs[31] 1.8 9.0 ; 675/1300,ub was 10
he1[33] hbs[27] 1.8 8.95 ; 675/1313,
he1[33] hgr[30] 1.8 9.44 ; 675/1403, ub was 9.24,7/9
he1[33] hbr[34] 1.8 8.98 ; 675/1411, , 1411
he1[33] hbr[14] 1.8 8.95 ; 675/1440,change from hbs[39].
he1[33] hbs[14] 1.8 8.45 ; 675/1544,
he1[33] ha[41] 1.8 8.71 ; 676/1043,
hd1[33] ha[41] 1.8 8.07 ; 704/1042,
hd1[33] hds[30] 1.8 8.48 ; 704/1251, ,
hd1[33] hbr[27] 1.8 9.18 ; 704/1295,

hd1[33] hbr[14] 1.8 8.40 ; 704/1441, changed from hbs[39],7/2
hd1[33] hbr[30] 1.8 8.40 ; 704/1363,
hd1[33] hgr[30] 1.8 8.18 ; 704/1403,
hd1[33] hbr[34] 1.8 8.00 ; 704/1411,changed from hbs[30], 7/2
hd1[33] hgs[30] 1.8 8.40 ; 704/1433,
hd1[33] hbs[14] 1.8 8.74 ; 704/1544,
ha[33] mgs[42] 1.8 7.84 ; 916/1647, x1.2 for correction
(1);hbr[33] har[15] 1.8 3.96 ; 1274/1131, 1131 could be ha[38]
hbr[33] ha[41] 1.8 4.0 ; 1275/1043
hd1[33] hbs[37] 1.8 9.0 ; 706/1367,ub was 10.5
(1);hd1[33] hbs[26] 1.8 8.5 ; 706/1437,ub was 10.5 needs to be checked 7/7/92
hd1[33] hbs[34] 1.8 10.0 ; 707/1445,
hd1[33] hn[42] 1.8 8.5 ; 706/556,ub was 8.5
hn[34] hbs[33] 1.8 5.0 ; 371/1348
hn[34] hd1[33] 1.8 8.0 ; 372/707 ub was 9.0
hn[34] har[35] 1.8 6.0 ; 371/1173, 1173 could be har[32]
(1);hn[34] ha[36] 1.8 6.0 ; 371/1147
hn[34] hbs[40] 1.8 5.0 ; 371/1527
hn[34] mgr[42] 1.8 9.4 ;372/1626, x1.2 for correction **mgr
hn[34] mgs[42] 1.8 7.2 ; 372/1649, x1.2 for correction
hgr[34] mgs[42] 1.8 7.30 ; 1408/1647 **mgs for mg2
hn[34] mgs[42] 1.8 7.62 ; 372/1649, x1.2 for methyl group
hn[35] hbs[14] 1.8 3.58 ; 536/1546. out on 7/2. try 8/5.
hn[35] hbs[13] 1.8 5.0 ; 536/1527. out on 7/2. try 8/5
hn[35] hbr[13] 1.8 4.3 ; 536/1495. out 7/2. try 8/5.
hn[35] hbs[34] 1.8 3.00 ; 536/1447
hn[35] hbr[34] 1.8 3.85 ; 537/1411
hn[35] hgr[34] 1.8 3.85 ; 537/1419
hn[35] he1[33] 1.8 8.2 ; 536/677ub was 9.2
hn[35] hd1[33] 1.8 9.84 ; 536/706, not a good fit, real pk., try 8/5.
hn[36] has[35] 1.8 3.97 ; 358/1175
hn[36] hn[37] 1.8 4.25 ; 358/415
hbs[37] hg[13] 1.8 5.5 ; 1368/1464
(1);ha[37] hbs[27] 1.8 6.0 ; 892/1314, 1314 could be hbr[5] more than one choice, 7/2/92.

ha[37] mdr[13] 1.8 8.0 ; 892/1595, x1.2 for correction
ha[37] hgs[36] 1.8 6.0 ; 892/1517
ha[37] hbr[36] 1.8 6.0 ; 892/1442, 1442 could be hbs[6] or hbr[14] or hbs[39]
hn[37] hgr[36] 1.8 5.0 ; 415/1522
hn[37] hdr[36] 1.8 5.5 ; 415/1486
hbr[37] hbs[39] 1.8 4.41 ; 1210/1440, 1440 could be
hbr[37] mdr[13] 1.8 6.44 ; 1210/1596, x1.2 for correction,
hbr[37] hg[13] 1.8 4.2 ; 1210/1463
ha[37] hbs[38] 1.8 4.91 ; 892/1288
hn[37] hn[38] 1.8 4.00 ; 415/599, can not get from buildup
hn[37] hbr[36] 1.8 4.15 ; 415/1471
hn[37] hn[36] 1.8 4.50 ; 415/359
hn[37] hbr[36] 1.8 4.40 ; 415/1444, bad fit. Changed from hbs[39]
hn[37] ha[36] 1.8 4.20 ; 415/1147, can't get from buildup
hn[38] ha[37] 1.8 5.3 ; 599/894
hbr[38] hbs[33] 1.8 6.0 ; 1193/1347, 1349 is hgr[39]
hbr[38] hbr[33] 1.8 5.5 ; 1193/1275
hbr[38] hbr[14] 1.8 6.0 ; 1195/1441,
hbr[38] hbr[13] 1.8 6.5 ; 1195/1492
(1);ha[38] hgs[36] 1.8 6.5 ; 1133/1519, very very weak
ha[39] hbr[13] 1.8 3.8 ; 1124/1494, changed from hbs[11] 1492
ha[39] hbs[13] 1.8 5.0 ; 1123/1524
ha[39] hg[13] 1.8 4.7 ; 1124/1462
hgr[39] hbr[40] 1.8 4.55 ; 1349/1505
hbr[39] hbr[40] 1.8 2.78 ; 1408/1507
hgr[39] mdr[40] 1.8 5.23 ; 1348/1596, x1.2 for correction
ha[39] mdr[13] 1.8 7.60 ; 1123/1596, 1596 could be mdr[40]
ha[39] mds[13] 1.8 7.44 ; 1123/1611, x1.2 for correction
hgs[39] hbr[40] 1.8 3.81 ; 1397/1506
hn[39] hn[37] 1.8 6.5 ; 206/405, very very weak
hn[39] hn[38] 1.8 6.0 ; 206/597
hn[39] hbr[37] 1.8 6.5 ; 206/1214 , very very weak
hn[39] ha[38] 1.8 5.5 ; 206/1134
hn[40] ha[39] 1.8 4.8 ; 438/1126
hbs[40] hbr[39] 1.8 4.1 ; 1527/1407

hn[40] hn[39] 1.8 5.0 ; 439/206, can not get from buildup
hn[40] hn[34] 1.8 4.64 ; 439/371, cannot get from buildup
hn[41] hbs[40] 1.8 4.02 ; 462/1526
ha[41] ha[33] 1.8 3.13 ; 1042/916
hb[42] has[32] 1.8 4.23 ; 1456/1302, 1302 could be hbr[21]
hn[42] ha[33] 1.8 4.14 ; 556/918, 556 could also be hn[10]
hn[42] ha[41] 1.8 3.85 ; 557/1045
hn[42] hn[43] 1.8 2.64 ; 557/589, not a good fit
hn[43] ha[42] 1.8 3.04 ; 589/1131, 589 maybe hn[33],
hn[43] mgs[42] 1.8 7.60 ; 589/1649, x1.2 for correction
hn[43] hb[42] 1.8 5.0 ; 587/1457, Changed from hn[33] on 7/2/92
hn[43] mgr[42] 1.8 6.9 ; 588/1627, x1.2 for correction **mgr

Intra-residue NOE's:

ha[1] hbr[1] 1.8 2.82 ; 1101/1267, 1101 could be ha[13]
ha[2] hb[2] 1.8 3.80 ; 1121/1429
ha[2] mgs[2] 1.8 8.56 ; 1122/1596, another pk overlap?
hn[2] hb[2] 1.8 5.0 ; 492/1433
hn[2] ha[2] 1.8 4.45 ; 492/1123, not fit
hn[2] mgs[2] 1.8 7.07 ; 492/1598, x1.2 for methyl group
hn[3] hbs[3] 1.8 3.90 ; 488/1328, not fit very well
hn[4] har[4] 1.8 3.06 ; 499/1142
hn[4] has[4] 1.8 4.24 ; 499/1151
ha[5] hbs[5] 1.8 3.20 ; 1037/1314, visual
hn[5] hbs[5] 1.8 3.0 ; 522/1314
hn[6] hbs[6] 1.8 3.09; 434/1446
hn[6] ha[6] 1.8 4.26 ; 434/1132
he2e[6] hgr[6] 1.8 6.27 ; 630/1377
hn[6] hbr[6] 1.8 4.75 ; 434/1406
he2z[6] hgr[6] 1.8 7.5 ; 747/1376
hn[6] hgr[6] 1.8 3.85 ; 434/1378, not fit very well
ha[6] hbs[6] 1.8 3.85 ; 1132/1442
ha[6] hgr[6] 1.8 3.16 ; 1132/1375
ha[6] hbr[6] 1.8 2.81 ; 1132/1404
ha[7] hbr[7] 1.8 3.12 ; 1084/1266
ha[7] hbs[7] 1.8 3.20 ; 1084/1303
hn[7] hbr[7] 1.8 4.08 ; 480/1271
hn[7] hbs[7] 1.8 3.0 ; 480/1305
hn[7] ha[7] 1.8 2.66 ; 480/1087
hn[8] hbr[8] 1.8 3.20 ; 587/1420
hn[8] hbs[8] 1.8 3.20 ; 587/1424
hn[8] hgr[8] 1.8 4.26 ; 587/1371
hn[8] ha[8] 1.8 3.95 ; 587/1147
ha[8] hgr[8] 1.8 4.00 ; 1145/1369, two hgr[8] are equiv.
ha[8] hbr[8] 1.8 3.200 ; 1146/1422
ha[8] hbs[8] 1.8 3.149 ; 1146/1426
ha[9] hbr[9] 1.8 2.94 ; 1091/1168

ha[9] hbs[9] 1.8 2.94 ; 1091/1179
hn[9] hbs[9] 1.8 4.20 ; 607/1178
hn[9] hbr[9] 1.8 4.20 ; 607/1169
hn[9] ha[9] 1.8 4.40 ; 608/1092
ha[10] hbr[10] 1.8 2.88 ; 1071/1307, almost equiv. hbr[10]
hd2e[10] hbr[10] 1.8 6.28 ; 638/1308, garma NH: 638
hn[10] hbr[10] 1.8 3.19 ; 556/1308
hn[10] ha[10] 1.8 3.24 ; 556/1074
hd2z[10] hbr[10] 1.8 7.89 ; 754/1308, can't get from buildup
ha[11] hbr[11] 1.8 2.88 ; 1106/1478
ha[11] hbs[11] 1.8 2.95 ; 1106/1494
hdr[11] hgr[11] 1.8 3.14 ; 1213/1534
hdr[11] hgs[11] 1.8 2.98 ; 1213/1632
hdr[11] hbr[11] 1.8 4.90 ; 1212/1477
hbr[11] hgs[11] 1.8 3.45 ; 1477/1630
hbs[11] hgs[11] 1.8 4.12 ; 1494/1630
hdr[11] hbs[11] 1.8 4.22 ; 1213/1495
hdr[11] hbr[11] 1.8 4.12 ; 1213/1477
ha[12] hbr[12] 1.8 3.50 ; 1084/1323
ha[12] hbs[12] 1.8 2.82 ; 1084/1371
hn[12] hbs[12] 1.8 4.00 ; 593/1375
hn[12] hbr[12] 1.8 3.12 ; 593/1326, not a good fit.
hn[12] ha[12] 1.8 4.47 ; 593/1086
hn[13] hbs[13] 1.8 3.91 ; 480/1528
hn[13] hbr[13] 1.8 2.75 ; 480/1495
hn[13] hg[13] 1.8 2.89 ; 480/1466
hn[13] mdr[13] 1.8 6.72 ; x 1.2 for methyl group correction.
hn[13] ha[13] 1.8 2.89 ; 480/1101
ha[13] hbr[13] 1.8 3.18 ; 1101/1492
ha[13] hbs[13] 1.8 2.85 ; 1101/1523
ha[13] mdr[13] 1.8 5.84 ; 1101/1596, x1.2 for correction
hg[13] mdr[13] 1.8 5.88 ; 1463/1596, x1.2 for correction
hbs[13] mdr[13] 1.8 5.00 ; 1526/1597, x1.2 for correction
ha[14] hbs[14] 1.8 3.19 ; 1159/1544
ha[14] hbr[14] 1.8 4.05 ; 1160/1442

hn[14] ha[14] 1.8 2.70 ; 469/1161
hn[14] hbs[14] 1.8 4.89 ; 469/1547
hd2e[14] hbr[14] 1.8 6.58 ; 669/1443
hd2e[14] hbs[14] 1.8 6.92 ; 669/1546
hd2z[14] hbs[14] 1.8 6.32 ; 677/1546
hn[15] har[15] 1.8 3.82 ; 506/1133
hn[15] has[15] 1.8 2.92 ; 506/1203
hn[16] has[16] 1.8 3.18 ; 592/1185
hn[17] hbr[17] 1.8 4.80 ; 441/1153, not fit very well
hn[18] ha[18] 1.8 5.0 ; 417/977
hn[19] hgr[19] 1.8 5.5 ; 346/1539
ha[19] hbs[19] 1.8 3.82 ; 1043/1476
ha[19] hdr[19] 1.8 3.81 ; 1043/1503
ha[19] hgr[19] 1.8 3.91 ; 1043/1538
ha[20] hbr[20] 1.8 3.10 ; 1039/1300, visual
ha[20] hbs[20] 1.8 3.80 ; 1039/1327, visual
hn[20] ha[20] 1.8 3.05 ; 437/1042
hn[20] hbr[20] 1.8 3.08 ; 436/1303, not fit very well
hn[21] hbr[21] 1.8 3.20 ; 504/1354,
ha[21] hbs[21] 1.8 3.00 ; 998/1352
ha[22] hb[22] 1.8 4.14 ; 1178/1413
ha[22] hg1r[22] 1.8 4.40 ; 1178/1504
ha[22] hg1s[22] 1.8 4.20 ; 1178/1554
ha[22] mg2[22] 1.8 4.99 ; 1179/1597
hb[22] mg2[22] 1.8 6.14 ; 1406/1596, x1.2 for correction
hb[22] hg1s[22] 1.8 4.30 ; 1414/1554
hb[22] hg1r[22] 1.8 2.90 ; 1414/1506
hn[22] hg1r[22] 1.8 3.00 ; 442/1506, not fit very well
hn[22] hb[22] 1.8 3.80 ; 442/1416
hn[22] ha[22] 1.8 2.84 ; 442/1180
hn[22] hg1s[22] 1.8 4.95 ; 442/1554, can't get from buildup
hd2e[23] hbs[23] 1.8 4.10 ; 606/1312
hd2e[23] hbr[23] 1.8 4.50 ; 606/1303
hn[23] hbr[23] 1.8 4.75 ; 452/1303
hn[23] hbs[23] 1.8 4.5 ; 452/1314

hd2z[23] hbs[23] 1.8 6.0 ; 713/1312
hn[24] ha[24] 1.8 5.0 ; 608/1060
hn[24] hbs[24] 1.8 5.3 ; 608/1168, visual
ha[24] hbr[24] 1.8 3.00 ; 1058/1168
hn[25] hbs[25] 1.8 4.68 ; 543/1326
hn[25] ha[25] 1.8 4.00 ; 542/939
hn[25] hd1[25] 1.8 8.03 ; 542/720, .
hd1[25] hbr[25] 1.8 7.17 ; 719/1265
hd1[25] hbs[25] 1.8 7.17 ; 719/1326
he1[25] hbr[25] 1.8 8.00 ; 723/1266
he1[25] hbs[25] 1.8 8.00 ; 723/1326
ha[25] hbr[25] 1.8 2.97 ; 937/1265
ha[25] hbs[25] 1.8 3.04 ; 937/1326
hn[26] hbr[26] 1.8 5.5 ; 400/1428
hn[26] hbs[26] 1.8 5.6 ; 400/1438
ha[26] hbr[26] 1.8 3.09 ; 1040/1426
ha[26] hbs[26] 1.8 3.02 ; 1040/1435
ha[26] hgr[26] 1.8 4.60 ; 1040/1387
ha[26] hgs[26] 1.8 4.02 ; 1040/1398
ha[27] hbs[27] 1.8 3.10 ; 906/1314
ha[27] hbr[27] 1.8 3.21 ; 907/1289
hn[27] hbs[27] 1.8 4.65 ; 375/1316, cannot get from buildup
hn[27] ha[27] 1.8 3.90 ; 375/910, can't get from buildup
hn[27] hbr[27] 1.8 3.90 ; 374/1291
hn[28] hbr[28] 1.8 5.5 ; 330/1235
hn[28] hbs[28] 1.8 5.7 ; 330/1252
ha[28] hbr[28] 1.8 3.10 ; 1021/1234
ha[28] hbs[28] 1.8 3.98 ; 1021/1250
ha[29] hbr[29] 1.8 2.91 ; 974/1299
ha[29] hbs[29] 1.8 3.24 ; 974/1351
hn[29] hbs[29] 1.8 3.18 ; 475/1353
hn[29] hbr[29] 1.8 4.10 ; 475/1301, can't get from buildup
ha[30] hbr[30] 1.8 3.16 ; 1070/1363
ha[30] hgr[30] 1.8 5.16 ; 1070/1403
ha[30] hbs[30] 1.8 3.14 ; 1070/1413

ha[30] hgs[30] 1.8 4.52 ; 1070/1432
hdr[30] hgs[30] 1.8 3.07 ; 1219/1433
hds[30] hgs[30] 1.8 4.05 ; 1251/1406
hds[30] hgr[30] 1.8 4.02 ; 1251/1433
hgr[30] hdr[30] 1.8 4.30 ; 1403/1218
ha[31] hbr[31] 1.8 3.02 ; 1109/1256
he1[31] hbr[31] 1.8 10.5 ; 650/1257, very very weak
he1[31] hbs[31] 1.8 10.5 ; 650/1301, very very weak
ha[31] hbs[31] 1.8 3.18 ; 1109/1301
hn[31] hbs[31] 1.8 2.94 ; 505/1301
hn[31] ha[31] 1.8 4.70 ; 505/1109
hn[31] he1[31] 1.8 8.21 ; 504/665, .
hn[32] har[32] 1.8 3.00 ; 540/1172
hn[32] has[32] 1.8 2.86 ; 540/1306
he1[33] hbs[33] 1.8 8.25 ; 675/1342,changed from hbs[41]
ha[33] hbr[33] 1.8 3.00 ; 916/1275
ha[33] hbs[33] 1.8 3.11 ; 916/1345
hn[33] hbs[33] 1.8 3.06 ; 589/1348
hn[33] ha[33] 1.8 4.14 ; 589/918, can't get from buildup.
hn[33] hbr[33] 1.8 4.10 ; 589/1277, can't get from buildup
hn[34] hbs[34] 1.8 4.05; 372/1447
hn[34] hbr[34] 1.8 3.94 ; 372/1411
hn[35] har[35] 1.8 3.81 ; 536/1126
hn[36] hbs[36] 1.8 4.29 ; 358/1472
hn[36] hbr[36] 1.8 2.95 ; 358/1444
hn[36] ha[36] 1.8 3.91 ; 358/1146
hn[36] hgr[36] 1.8 4.21 ; 358/1519
hn[36] hdr[36] 1.8 4.96 ; 358/1486, can't get from buildup
her[36] hdr[36] 1.8 3.95 ; 1317/1484
hes[36] hgs[36] 1.8 4.53 ; 1326/1519
her[36] hgs[36] 1.8 4.13 ; 1316/1519
hes[36] hgr[36] 1.8 3.97 ; 1326/1483
hbr[36] hes[36] 1.8 4.15 ; 1443/1326
hbr[36] hgs[36] 1.8 2.85 ; 1443/1517
her[36] hgs[36] 1.8 3.95 ; 1318/1520

hes[36] hdr[36] 1.8 4.30 ; 1325/1482, can't get from buildup
hes[36] hbr[36] 1.8 4.30 ; 1325/1442, can't get from buildup
ha[36] hbr[36] 1.8 3.51 ; 1146/1443
ha[36] hbs[36] 1.8 4.15 ; 1146/1469
ha[36] hdr[36] 1.8 4.23 ; 1146/1483
ha[36] hgr[36] 1.8 3.11 ; 1146/1519
ha[37] hbr[37] 1.8 3.18 ; 892/1210
ha[37] hbs[37] 1.8 3.29 ; 892/1366
hn[37] hbs[37] 1.8 3.92 ; 415/1367
hn[37] hbr[37] 1.8 4.28 ; 415/1212, can't get from buildup
hd2e[37] hbs[37] 1.8 4.18 ; 650/1367
hd2e[37] hbr[37] 1.8 4.20 ; 650/1212
hd2z[37] hbr[37] 1.8 5.5 ; 716/1214
hn[38] ha[38] 1.8 2.70 ; 599/1133
ha[38] hbs[38] 1.8 2.88 ; 1132/1288
ha[39] hgs[39] 1.8 3.95 ; 1124/1400
ha[39] hbr[39] 1.8 2.84 ; 1124/1409
ha[39] hbs[39] 1.8 3.25 ; 1124/1441
ha[39] hgr[39] 1.8 4.50 ; 1123/1348
hbs[39] hgr[39] 1.8 3.0 ; 1441/1349
hn[39] hbr[39] 1.8 5.5 ; 206/1411
hn[39] hbs[39] 1.8 5.3 ; 206/1441
(1);hn[39] hbs[29] 1.8 6.5 ; 206/1352 , very very weak
ha[40] hbr[40] 1.8 3.18 ; 1039/1507
ha[40] hbs[40] 1.8 3.15 ; 1039/1527
ha[40] mdr[40] 1.8 5.23 ; 1039/1597, x1.348 for correction
hn[40] hbr[40] 1.8 3.00 ; 439/1510, not fit very well
hn[40] hbs[40] 1.8 4.05 ; 439/1526, not fit very well
hn[41] ha[41] 1.8 4.08 ; 462/1043, fit not good
hn[41] hbr[41] 1.8 3.26 ; 462/1314
hn[41] hbs[41] 1.8 3.28 ; 462/1342
ha[41] hbs[41] 1.8 2.95 ; 1042/1341
hb[42] mgr[42] 1.8 6.64 ; 1456/1624, x1.2 for correction
hb[42] mgs[42] 1.8 6.88 ; 1456/1647, x1.2 for correction
ha[42] hb[42] 1.8 4.06 ; 1132/1455

ha[42] mgr[42] 1.8 7.12 ; 1132/1624, x1.2 for correction
ha[42] mgs[42] 1.8 7.50 ; 1132/1646, x1.2 for correction
hn[42] mgs[42] 1.8 7.56 ; 557/1648, x1.2 for distance
hn[42] ha[42] 1.8 3.23 ; 557/1131, can't get from buildup
hn[42] hb[42] 1.8 3.86 ; 557/1458
ha[43] hb[43] 1.8 2.80 ; 1115/1569
ha[43] mg2[43] 1.8 7.90 ; 1125/1570
hn[43] mg2[43] 1.8 7.0 ; 589/1571

hydrogen bond constraints are assigned between known slowly exchanging amide protons and a suitable carbonyl oxygen acceptor.

hn[17] o[28] 1.8 3.5
hn[28] o[17] 1.8 3.5
hn[19] o[26] 1.8 3.5
hn[26] o[19] 1.8 3.5
hn[32] o[42] 1.8 3.5
hn[42] o[32] 1.8 3.5
hn[34] o[40] 1.8 3.5
hn[40] o[34] 1.8 3.5

(1): indicates the NOE did not be applied for the structure calculation.

References

1. Nawroth, P. P., & Stern, D. M. (1986) In *Vascular Endothelium in Homeostasis and Thrombosis* Gimbrone, M. A. (ed.) pp. 14-39, Churchill Livingstone, Edinburgh, UK: Endothelial cells as active participants in procoagulant reactions.

2. Majerus, P. W. (1987) in *The Molecular Basis of Blood Diseases* (Stamatoyannopoulos, G., Nienhuis, A. W., Leder, P., & Majerus, P. W., Eds.) pp. 689-721, W. B. Saunders Co., Philadelphia, PA

3. Walsh, P. N. (1974) *Blood* , **43**, 597-605: Platelet coagulant activities and hemostasis: a hypothesis.

4. Davie, E. W., Fujikawa, K., & Kisiel, W. (1991) *Biochemistry*, **43**, 10363-10370: The coagulation cascade: initiation, maintenance, and regulation.

5. Morawitz, P. (1905) , *biologischen Chemie und experimentaellen Pharmakologie*, **4**, 307-422, Die Chemie der Blutgerinnerung, in Ergebnisse der Physiologie.

6. Davie, E. W., & Ratnoff, O. D. (1964) *Science* , **145**, 1310-1312: Waterfall sequence for intrinsic blood clotting.

7. MacFarlane, R. G. (1964) *Nature* , **202**, 498-499: An enzyme cascade in the blood clotting mechanism and its function as a biochemical amplifier.

8. Weiss, H. J., Turitto, V. T., Baumgartner, H. R., Nemerson, Y., & Hoffman, T. (1989) *Blood* , **73**, 968-975: Evidence for the presence of tissue factor activity on subendothelium.

9. Wilcox, J. N., Smith, K. M., Schwartz, S. M., & Gordon, D. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 2839-2843: Localization of tissue factor in the normal vessel wall and in the atherosclerotic plaque.

10. Zur, M., & Nemerson, Y. (1980) *J. Biol. Chem.* ,**255**, 5703-5707: Kinetics of factor IX activation via the extrinsic pathway.

11. Komiyama, Y., Pedersen, A. H., & Kisiel, W. (1990) *Biochemistry* , **29**, 9418-9425: Proteolytic activation of human factor IX and X by recombinant human factor VIIa: effects of calcium, phospholipids, and tissue factor.

12. Broze, G. J., Girard, T. J., & Novotny, W. F. (1990) *Biochemistry* , **29**, 7539-7546: Regulation of coagulation by a multivalent kunitz-type inhibitor.

13. Girard, T. J., Warren, L. A., Novotny, W. F., Likert, K. M., Brown, S. G., Miletich, J. P., & Broze, G. J. (1989) *Nature* , **338**, 518-520: Functional significance of the kunitz-type inhibitory domains of lipoprotein-associated coagulation inhibitor.

14. Rapaport, S. I. (1989) *Blood* , **73**, 359-365: Inhibition of factor VIIa/tissue factor-induced blood coagulation: with particular emphasis upon a factor Xa-dependent inhibitory mechanism.

15. Natio, K., & Fujikawa, K. (1991) *J. Biol. Chem.* **266**, 7353-7358: Activation of human blood coagulation factor XI independent of factor XII.

16. Travis, J., and Salvesen, G. S. (1983), *Ann. Rev. Biochem.* **52**, 655-709: Human plasma proteinase inhibitors.

17. Esmon, C. T. (1987) *Science*, **235**, 1348-1352: The regulation of natural anticoagulant pathways.

18. Tollefsen, D. M., Majerus, P. W., & Blank, M. K. (1982) *J. Biol Chem.*, **257**, 2162-2169: Heparin cofactor II.

19. Sottrup-Jensen, L. (1987) in *The Plasma Proteins* (Putnam, F. W., Ed.) Vol. **5**, pp. 191-291, Academic Press, New York.

20. Suzuki, K., Nishioka, J., & Hashimoto, S. (1983) *J. Biol. Chem.*, **258**, 163-168: Protein C inhibitor.

21. Heeb, M. J., & Griffin, J. H. (1988) *J. Biol. Chem.*, **263**, 11613-11616: Physiologic inhibition of human activated protein C by α1-antitrypsin.

22. Gimbrone, M. A. (1986): Vascular endothelium: nature's blood container. In *vascular Endothelium in Hemostasis and Thrombosis* (Gimbrone, M. A. Ed.) pp. 1-13, Churchill Livingstone, Edinburgh, UK

23. Hedner, U., Davie, E.W.: Factor IX, Colman, R. W., Hirsh, J., Marder, V. J., Salzman, E. W. (eds): *Hemostasis and Thrombosis*. Philadelphia, J. B. Lippincott, 1982, p 29.

24. Thompson, A. R.(1986) *Blood* , **67**, 565-572: Structure, function and molecular defects of factor IX.

25. DiScipio, R. G., and Davie, E. W. (1979) *Biochemistry*, **18**, 899-904: Characterization of protein S, a gamma-carboxyglutamic acid containing protein from bovine and human plasma.

26. McMullen, B. A., Fujikawa, K., and Kisiel, W. (1983) *Biochem Biophys Res Comm* **155**, 8-14: Occurrence of β-hydroxyaspartic acid in the vitamin K-dependent blood coagulation zymogen.

27. Stenflo, J. (1991) *Blood*, **78**, 1637-1651. Structure-function relationship of epidermal growth factor modules in vitamin K-dependent clotting factors.

28. Borowski, M., Furie, B. C., and Furie, B. (1986). *J. Biol. Chem.* **261**, 14969-14975: Prothrombin requires two sequential metal-dependent conformational transitions to bind phospholipid.

29. Straight, D. L., Sherrill G. B., Noyes, C. M., Trapp H. G., Wright, S. F., Roberts, H. R., Hiskey, R. G., Griffith, M. J. (1985) *J. Biol. Chem.* **260**, 2890: Structural and functional characteristics of activated human factor IX after chemical modification of γ-carboxyglutamic acid residues.

30. Davie, E. W., Fujikawa, K., Kurachi, K., and Kisiel, W. (1979) in Adv. *Enzymol.* **48**, 277-318: The role of serine proteases in the blood coagulation cascade.

31. Osterud, B., Rapaport, S. I. (1977) *Proc. Natl. Acad. Sci. USA* , **74**, 5260: Activation factor IX by the reaction product of tissue factor and factor VII: Additional pathway for initiating blood coagulation.

32. McMullen, B. A., Fujikawa, K., Kisiel, W. (1983) *Biochem. Biophys. Res Commun.* **115**, 8. The occurrence of bata-hydroxyaspartic acid in vitamin K-dependent blood coagulation zymogens.

33. Stenflo, J., Holme, E., Lindstedt, S., Chandramouli, N., Huang, L. H. T., Tam, J. P., and Merrifield, R. B. 1989 *Proc. Natl. Acad. Sci. U. S. A.* **86**, 444-447: Hydroxylation of aspartic acid in domains homologous to the epidermal growth factor precursor is catalyzed by a 2-oxoglutarate-dependent dioxygenase.

34. Morita, T. & Kisiel, W. (1985) *Biochem. Biophys. Res. Commun.* **130**, 841-847: Calcium binding to a human factor IXa derivative lacking γ-carboxyglutamic acid: Evidence for two high-affinity sites that do not involve β-hydroxyaspartic acid.

35. Davis, L. M., McGraw, R. A., Ware, J. L., Roberts, H. R., and Stafford, D. W. (1987) *Blood*, **69**, 140-143: Factor $IX_{Alabama}$: A point mutation in a clotting protein results in hemophilia B.

36. McCord, D. M., Monroe, D. M., Smith, K. J., and Roberts, H. R. (1990) *J. Biol. Chem.* **265**, 10250-10254: Characterization of the functional defect in factor IX Alabama. Evidence for a conformational change due to high affinity calcium binding in the first epidermal growth factor domain.

37. Lozier, J. N., Monroe, D. M., Stanfield-Oakley, S. A. (1990) *Blood*, **75**, 1097-1104: Factor IX new london: Substitution of proline for glutamine at position 50 causes severe hemophilia B.

38. Spized, S. G., Kuppuswamy, M. N., Saini, R., Kasper, C. K., Birktoft, J. J., Bajaj, S. P. (1990) *Blood*,.**76**, 1530-1537: Factor IX Hollywood: Substitution of Pro 55 by Ala in the first epidermal growth factor-like domain.

39. Winship, P. R., & Dragon, A. C. (1991) *British J. Haematology*, **77**, 102-109: Identification of haemophilia B patients with mutations in the two calcium binding domains of factor IX: importance of a β-OH Asp 64 to Asn change.

40. Green, P. M., Bentley, D. R., Mibashan, R. S., Nilson, I. M., and Giannelli, F. (1989) *EMBO J.* **8**, 1067-1072: Molecular pathology of haemophilia B.

41. Huang, L. H., Ke, X.-H., Sweeney, W., Tam, J. P. (1989) *Biochem. Biophys. Res. Commun.* **160**, 133-139: Calcium binding and putative activity of the epidermal growth factor domain of blood coagulation factor IX.

42. Huang, L. H., Sweeney, W., Tam, J. P. (1990) in *Peptides: Chemistry, Structure, and Biology* Rivier, J. E., & Marshall, G. R., Eds. Escom, Leiden, pp. 97-98: Biological activities of synthetic EGF-like domains in blood coagulation factor IX.

43. Handford, P. A., Baron, M. Mayhew, M., Wills, A., Beesley, T., Brownlee, G. G., Campbell, I. D. (1990) *EMBO J.* **9**, 475-480: The first EGF-like domain from human factor IX contains a high-affinity calcium binding site.

44. Astermark, J., Stenflo, J. (1991) *J. Biol. Chem.* **266**, 2438-2443: The epidermal growth factor-like domains of factor IX. Effect on blood clotting and endothelial cell binding of a fragment containing the epidermal growth factor-like domains linked to the γ-carboxyglutamic acid region.

45. Carpenter, G. & Wahl, M. L.(1990) *Handbook of Experimental Pharmacology*, Vol. **95**, 69-171: The epidermal growth factor family, in Sporn, M. B., Roberts, A. B. (eds), New York, NY, Springer.

46. Carpenter, G., Cohen, S. (1990) *J. Biol. Chem.* **265** 7709-7712: Epidermal Growth Factor.

47. Gary, A., Dull, T. J., Ullrich, A.(1983) *Nature* , **303**, 722-725: Nucleotide sequence of a epidermal growth factor cDNA predicts a 128,000-molecular weight protein precursor.

48. Bell, G. L., Fong, N. M., Stempien, M. M., Wormsted, M. A., Caput, D., Ku, L., Ureda, M. S., Rall, L. B., Sanchez-Pescador R.(1986) *Nucleic Acids Res.*, **14**, 8427-8446: Human epidermal growth factor precursor: cDNA sequence, expression in vitro and gene organization.

49. Heath, W. F., & Merrifield, R. B. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6367-6371: A synthetic approach to structure-function relationships in the murine epidermal growth factor molecule.

50. Furie, B., and Furie, B. C. (1988) *Cell*, **53**, 505-517: The molecular basis of blood coagulation.

51. Reid, K. B. M. and Day, A. J. (1989) *Immunol. Today*, **10**, 177-180: Structure function relationships of the complement system.

52. Scott, J. (1989) *Nature*, **338**, 118-119: Unravelling atherosclerosis.

53. Wharton, K. A., Johansen, K. M., Xu, T., & Artavanis-Tsakonas, S. (1985) *Cell*, **43**, 567-581: Nucleotide sequence from the neurogenic locus notch implies a gene product that shares homology with proteins containing EGF-like repeats.

54. Ringer, S., (1882) *J. Physiol.*, **4**, 29.

55. Seamon, K. B. and Kretsinger, R. H. (1983): Calcium-modulated proteins in *Calcium in biology* by Spiro, T. G. J. Wiley & Sons, Inc., pp. 45, New York.

56. Williams, R. J. P. (1986): *Calcium and the cell.*: The physics and chemistry of the calcium-binding proteins. Wiley. Chichester (Ciba Foundation Symposium 122.) pp. 149, New York.

57. Nelsestuen, G. L., (1984) in *Metal Ions in Biological Systems*, Vol. **17**, pp. 353-380, H. Sigel, Ed., Marcel Dekker, Inc., New York.

58. Rezaie, A. R., Esmon, N. L., and Esmon, C. T. (1992) *J. Biol. Chem.* **267**, 11701-11704: The high affinity calcium-binding site involved in protein C activation in outside the first epidermal growth factor homology domain.

59. Bax, A. (1982) *Two-dimensional Nuclear Magnetic Resonance*, Reidel, London.

60. Ernst, R. R., Bodenhausen, G. and Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon, Oxford University Press, London and New York.

61. Croasmun, W. R. and Carlson, R. M. K., eds (1987) *Two-dimensional NMR Spectroscopy*, VCH, New York.

62. Derome, A. E. (1987) *Modern NMR Techniques for Chemistry Research*, Pergamon.

63. Schraml, J. and Bellama, J. M. (1988) *Two-dimensional NMR Spectroscopy*, Wiley, New York.

64. Martin, G. E. and Zekter, A. S. (1988) *Two-dimensional NMR Methods for Establishing Molecular Connectivity*, VCH, New York.

65. Wright, P. E. (1990) in *Proteins: Form and Function*. Edited by Ralph A. Bradshaw and Mary Purton, Elsevier Trenda Journals, Cambridge. pp. 95-105: What can two-dimensional NMR tell us about proteins?

66. Noggle, J. M. and Schirmer, R. E. (1971) *The Nuclear Overhauser Effect*. Chemical Applications, Academic Press, New York.

67. Neuhaus, D. and Williamson, M. P. (1989) *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH, New York.

68. Wuethrich, K. (1986) in NMR of Proteins and Nucleic Acids, Wiley, New York.

69. Wuethrich, K., Billetre, M. and Braun, W. (1984) *J. Mol. Biol.* **180**, 715-740: Polypeptide secondary structure determinated by nuclear magnetic resonance observation of short proton-proton distances.

70. Bystrov, V. F. (1976) *Progr. Nucl. Magn. Reson. Spectros.*, **10**, 41-82: Spin-spin coupling and the conformational states of peptide systems..

71. A. Karplus, M. (1959) *J. Phys. Chem.* **30**, 11-15.

72. Creighton, T. (1984) Proteins: Structure and Molecular properties, pp. 236-237, W. H. Freeman and Company, New York.

73. Kuntz, I. D., Thomason, J. F., and Oshiro, C. M. (1989) in *Methods in Enzymology*, **177**, pp. 159-204: Distance geometry.

74. Campbell, I. D., Baron, M., Cooke, R. M., Budgeon, T. J., Fallon, A. F., Harvey, T. S., & Tappin, M. J. (1990) *Biochem. Pharmacol.* **40**, 35-40: Structure-function relationship in epidermal growth factor (EGF) and transforming growth factor-alpha (TGFa).

75. Hommel, U., Dudgeon, T. J., Fallon, A., Edwards, R. M., & Campbell, I. D. (1991) *Biochem.* **30**, 8891-8898: Structure-function relationship in human epidermal growth factor studied by site-directed mutagenesis and ^1H NMR.

76. Tam, J. P., Lin, Y.-Z., Wu, C. R., Shen, Z.-Y., Galantino, M., Liu, W., & Ke, X.-H., (1989), in *Peptides: Chemistry and biology*, Smith, J. A. & Rivier, J. E., Eds. pp.75-77: Systematic approach to study the structure-activity of transforming growth factor a. Escom, Leiden.

77. Merrifield, R. B., (1963) *J. Am. Chem. Soc.*, **85**, 2149-2154: Solid phase peptide synthesis. I. The synthesis of a tetrapeptide.

78. Sarin, V. K., Kent, S. B. H., Tam, J. P., and Merrifield, R. B. (1981) *Anal. Biochem.*, **117**, 147-157: Quantitative monitoring of solid-phase peptide synthesis by the ninhydrin reaction.

79. Tam, J. P., Heath, W. F., and Merrifield, R. B. (1983) *J. Am. Chem. Soc.* **105**, 6442-6455: S_N2 deprotection of synthetic peptides with a low contration of HF in dimethyl sulfide: Evidence and application in peptide synthesis.

80. Chait, B. T. Gisin, R. F. and Field, F. H. (1982) *J. Am. Chem. Soc.* **104**, 5157-5162: Fission fragment ionization mass spectrometry of alamethicin I.

81. Matsubara, H. *Method in Enzymology*, **19**, 642-651: Purification and assay of thermolysin.

82. Nakagawa, Y., Nishiuchi, Y., Emura, J. & Sakakibara, S. (1981) in *Peptide chemistry 1980* (Okawa, K., Ed.) pp.41-46, Protein Research Foundation, Osaka, Japan.

83. Tam, J. P. (1987) *Int. J. Peptide Protein Res.* **29**, 421-431: Synthesis of biologically active transforming growth factor alpha.

84. Montelione, G. T., Wuethrich, K., Nice, E. C., Burgess, A. W., & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 5226-5230: Solution structure of murine epidermal growth factor: Determination of the polypeptide backbone chain-fold by nuclear magnetic resonance and distance geometry.

85. Cooke, R. M., Wilkinson, A. J., Baron, M., Pastore, A., Tappin, M. J., Campbell, I. D., Gregory, H., & Sheard, B. (1987) *Nature* **327**, 339-341: The solution structure of human epidermal growth factor.

86. Kohda, D., Go, N., Hayashi, K., & Inagaki, F. (1988) *J. Biochem.* **103**, 741-743: Tertiary structure of mouse epidermal growth factor determined by two-dimensional ^{1}H NMR.

87. Kohda, D., & Inagaki, F. (1988) *J. Biochem.* **103**, 554-571: Complete sequence-specific ^{1}H nuclear magnetic resonance assignments for mouse epidermal growth factor.

88. Kohda, D., Shimada, I., Miyake, T., Fuwa, T., & Inagaki, F. (1989) *Biochemistry* **28**, 953-958: Polypeptide chain fold of human transformaing growth factor a analogous to those of mouse and human epidermal growth factors as studied by two-dimensional (^{1}H NMR).

89. Tappin, M. J., Cooke, R. M., Fitton, J. E., & Campbell, I. D. (1989) *Eur. J. Biochem.* **179**, 629-637: A high-resolution 1H-NMR study of human transforming growth factor a: structure and pH-dependent conformational interconversion.

90. Makino, K., Morimoto, M., Nishi, M., Sakamoto, S., Tamura, A., Inooka, H., Akasaka, K. (1987) *Proc. Natl. Acad. Sci. U.S. A.* **84**, 7841 : Proton nuclear magnetic resonance study on the solution conformation of human epidermal growth factor.

91 States, D. J., Harberkorn, R. A., & Reuben, D. J. (1982) *J. Magn. Reson.* **48**, 286-292: A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants.

92. Piantini, U., Sorensen, O. W., & Ernst, R. R., (1982) *J. Am. Chem. Soc.* **104**, 6800-6801: Multiple quantum filters for elucidating NMR coupling networks.

93. Shaka, A. J., & Freeman, R. (1983) *J. Magn. Reson.* **51**, 169-173: Simplification of NMR spectra by filtration through multiple-quantum coherence.

94. Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson.* **53**, 521-528: Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy.

95. Davis, D. G., & Bax, A. (1985) *J. Am. Chem. Soc.* **107**, 2820-2821: Assignment of complex 1H NMR spectra via two-dimensional homonuclear Hartmann-Hahn spectroscopy.

96. Macura, S., & Ernst, R. R. (1980) *Mol. Phys.* **41**, 95-117: Elucidation of cross relaxation in liquids by two-dimensional N.M.R. spectroscopy.

97. Eich, G., Bodenhausen, G., & Ernst, R. R. (1982) *J. Am. Chem. Soc.* **104**, 3731-3732: Exploring nuclear spin systems by relayed magnetization transfer.

98. Wagner, G. (1983) *J. Magn. Reson.* **55**, 151-156: Two-dimensional relayed coherence transfer spectroscopy of a protein.

99. Wokaun, A., & Ernst, R. R. (1977) *Chem. Phys. Lett.* **52**, 407-412: Selective detection of multiple quantum transitions in NMR by two-dimensional spectroscopy.

100. Rance, M., & Wright, P. E. (1986) *J. Magn. Reson.* **66**, 372-378.

101. Pardi, A., Billeter, M. & Wuethrich, K. (1984) *J. Mol. Biol.*, **180**, 741-751: Calibration of the angular dependence of the amide proton-C^{α} proton coupling constants, $^{3}J_{HN\alpha}$, in a globular protein.

102. Wider, G., Lee, K. H., & Wuethrich, K. (1982) *J. Mol. Biol.* **155**, 367-388: Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra - Glucagon bound to perdeuterated Dodecylphosphocholine micelles.

103. Kumar, A., Wagner, G., Ernst, R. R., and Wuethrich, K. (1981) *J. Am, Chem. Soc.* **103**, 3654-3658: Buildup rates of the nuclear Overhauser effect measured by two-dimensional proton magnetic resonance spectroscopy: Implication for studies of protein conformation.

104. Bax, A. (1989) *Annu. Rev. Biochem.* **58**, 223-256: Two-dimensional NMR and protein structure.

105. Basus, V. J. (1989) in *Methods in Enzymology*, **177**, pp. 132-149: Proton nuclear magnetic resonance assignments.

106. Mueller, Luciano (1987) *J. Magn. Reson.* **72**, 191-196: P. E. COSY, a simple alternative to E. COSY.

107. Ponder, J. W., & Richards, F. M. (1987) *J. Mol. Biol.* **193**, 775-791.

108. Rees, D. J. G., Jones, I. M., Handford, P. A., Walter, S. J., Esnouf, M. P., Smith, K. J., & Brownlee, G. G. (1988) *EMBO J.* **7**, 2053-2061: The role of β-hydroxyaspartate and adjacent carboxylate residues in the first EGF domain of human factor IX.

109. Montelione, G. T., Wuethrich, K., Burgess, A. W., Nice, E. C., Wagner, G., Gibson, K. D., and Scheraga, H. A. (1992) *Biochemistry*, **31**, 236-249: Solution structure of murine epidermal growth factor determined by NMR spectroscopy and refined by energy minimization with restraints.

110. Selander-Sunnerhagen, M., Ullner, M., Persson, E., Teleman, O., Stenflo, J., and Drakenberg, T. (1992) *J. Biol. Chem.* **267**, 19642-19649: How an epidermal growth factor (EGF)-like domain binds calcium.

111. Linse, S., Brodin, P., Johansson, C., Thulin, E., Grundstrom, T. & Forsen, S. (1991) *Nature*, **351**, 164-167: Key residues involved in calcium-binding motifs in EGF-like domains.

112. Huang, L. H., Cheng, H., Pardi, A., tam, J. P., & Sweeney, W. V. (1991) *Biochemistry*, **30**, 7402-7409: Sequence-specific 1H NMR assignments, secondary structure, and location of the calcium binding site in the first epidermal growth factor like domain of blood coagulation factor IX.

113. Campbell, I. D., Dobson, C. M., & Williams, R. J. P. (1975) *Proc. R. Soc. London A* **345**, 23-40.

114. Bundi, A., & Wuethrich, K. (1979) *Biopolymers* **18**, 285-298: ^{1}H-NMR parameters of the common amino acid residues measured in aqueous solution of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH.

115. Pardi, A., Wagner, G., & Wuethrich, K. (1983) *Eur. J. Biochem.* **137**, 445-454: Protein conformation and proton nuclear-magnetic-resonance chemical shifts.