INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

DETERMINATION OF THE SOLUTION STRUCTURE OF THE N-TERMINAL EGF-LIKE DOMAIN OF HUMAN BLOOD CLOTTING FACTOR IX AND INVESTIGATION OF ITS CALCIUM BINDING SITE USING 2D-NMR TECHNIQUES

BY

YUMIN GONG

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

1998

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

7/2/98 — Date

William D. Gray — Chair of Examining Committee

9/3/98 — Date

Gerald Kaufman — Executive Officer

Supervisory Committee

The City University of New York

Abstract

Determination of the solution structure of the N-terminal EGF-like domain of human blood clotting factor IX and investigation of its calcium binding site using 2D-NMR techniques

by

Yumin Gong

Advisor: Professor William V. Sweeney

2D-NMR techniques have been used to determine the solution structure of a 43-amino acid synthetic peptide, that is homologous to the portion (45-87) of the N-terminal EGF-like domain of human blood coagulation factor IX (FIX-EGF_N). The solution structures of FIX-EGF_N, FX-EGF_N, mEGF and hTGFα are compared. The results show that their tertiary structures are strikingly similar, with the exception of the termini. A previously reported structure of factor IX (46-84) does not conclude the second beta-sheet in our structure due to a shorter sequence.

Among FIX-EGF_N, FX-EGF_N, mEGF and hTGFα, FIX-EGF_N and FX-EGF_N have the highest sequential similarity, while FIX-EGF_N and mEGF have the highest structural similarity. The six cysteines and a few consensus residues should be important for the structural integrity of these EGF-like structures. A few residues conserved only in EGF and TGFα may be functionally distinct. It is likely that the different roles of these EGF-like polypeptides in biological function is due to the differences of the electronic environment of the surface.

The chemical shift titration was used to identify possible calcium ligands. The results show that Asp3, Gln6 and Gly4 are likely to be involved in calcium binding. The possibility of the sidechain of Asp20 for calcium ligand is ambiguous due to the complicated titration curve. The pH titration data is insufficient to determine if the backbone carbonyls of Asp5, and Asp20 are calcium ligands. A potential calcium binding site between the sidechains of Asp3, Gln6, Asp20, and backbone of Gly4 was inferred from a visual inspection of the solution structure of FIX-EGF_N. The backbone carbonyls of Asp5 and Asp20 are pointed away from the potential calcium binding site, making them less likely to be calcium ligands. The potential calcium binding site of FIX-EGF_N from this study is similar to that suggested by a previous crystal structure study. This potential calcium binding site is similar to that proposed for FX-EGF_N. The involvement of Asp3, Gln6, and Asp20 in calcium binding is consistent with the finding that mutations occur at these specific residue positions in hemophilia B patients.

In memory of my beloved father
Gong Xi-Wei
who was always proud of me
particularly when I entered the doctorate program,
and who passed away during the second semester
of my Ph.D. study.

Acknowledgments

I would like to express my sincere thanks to many people behind my success in obtaining my Ph.D degree, of whom I can only mention a few. I especially thank the following individuals:

Professor William Sweeney, my supervisor and mentor, for taking me into his lab, supervising and encouraging my Ph.D. research project, guiding me to be an independent scientist, and his generous friendship.

Professor Gary Quigley, for his kind discussions and suggestions for my structural computation work.

Professor Klaus Grohmann, for his generous encouragement and support while finishing my Ph.D. degree.

Dr. Michael Blumenstein, for his generous effort in instructing me to run Varian Unity-plus 500 MHz spectrometer, and for helpful discussions about NMR problems encountered in my thesis research work.

Dr. Linda Huang and Dr. Yan Yang, who graduated from Professor W. Sweeney's lab in 1993 and 1995 respectively, for synthesis of human factor IX (45-87).

Professor Peter Lipke, for his kindness in sharing his knowledge of protein homologous alignment and database searches.

Mr. Waldemar Cieniewicz, for his solid support with the computer workstation.

All members in the Chemistry Department of Hunter College who have guided and supported me during my Ph.D study.

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

Especially, I would like to thank my beloved father Gong Xi-Wei and my beloved mother Qu Qi who brought me up with deep love and always encouraged and supported me in my advanced education and career achievements; to my husband Ren Guang-Yi, for his deep love, solid support and encouragement; to my parent-in-laws, Ren Qing-Li and Wang Ai-Qing, for taking care of my newborn daughter while I undertook my study; to my brothers Gong Bo-Min and Gong Zheng-Min for their love and encouragement. I feel that I owe my daughter, Ren Yue, dearly. I wish I could have spent more time with her. I am also deeply grateful for her. She brightened up the successful pursuit of my Ph.D.

Table of Contents

List of Figures

List of Tables

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	histine
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

COSY	homonuclear-correlated spectroscopy
DGII	The distance geometry module of InsightII Software

DQF-COSY	double quantum filtered COSY
EGF	epidermal growth factor
FIX-EGF_N	the N-terminal EGF-like domain in human coagulation blood factor IX
FX-EGF_N	the N-terminal EGF-like domain in human coagulation blood factor X
Gla	γ-Carboxylglutamic acid
hEGF	human epidermal growth factor
hTGFα	human transforming growth factor-α
mEGF	mouse epidermal growth factor
NMR	proton nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
NOSEY	nuclear Overhauser enhancement spectroscopy
PE-COSY	primitive elegant correlation spectroscopy
RMSD	root mean square deviation
TGFα	transforming growth factor-α
TOCSY	total correlated spectroscopy

Chapter I. Background.

1.1 Structure-function relationships in EGF-like polypeptide.

The epidermal growth factor (EGF) is a small protein and an important hormone which stimulates cell proliferation. Both human and mouse EGF contain 53 amino acids with six cysteine residues. The native structure of EGF is stabilized by disulfide bridges, and reductive cleavage causes loss of function (Carpenter et al., 1979; Komoriya et al., 1984). A polypeptide with about 40 amino acid residues in length and has a significant sequence homology to EGF, including six cysteine residues that form three disulfide bonds, is known as an EGF-like module or domain. In larger proteins there can be single or multiple copies of EGF-like modules. Some small proteins are homologous to EGF, such as transforming growth factor-α (TGFα) with 50 amino acids. The EGF-like domain appear to have evolved from a common ancestor by gene duplication, point mutations, and so-called exon shuffling (Patty, L. 1991).

The cysteine residues of EGF are linked by a disulfide bonding motif of 1-3, 2-4, and 5-6 (Savage et al., 1973), a pattern characteristic of EGF-like modules (Figure 1.1.1). The pairing of disulfide bonds in EGF-like modules was routinely inferred from sequence homology (Stenflo, 1991). Although more than 300 EGF-like domains have been identified by sequence similarity, there are relatively few for which the disulfide pattern is unambiguous. So far, the actual EGF-like disulfide bonding pattern, which is Cys1-

Cys3, Cys2-Cys4, and Cys5-Cys6, has been determined only for TGFα (Winkler, 1986), factor X (Höjrup and Magnusson, 1987) and the fourth EGF-like domain of thrombomodulin (Meininger et al., 1995). Under appropriate conditions, some EGF-like domains fold spontaneously to attain a native conformation with the inferred correct disulfide bond pairing. These include the N-terminal EGF model in factor IX, EGF modules expressed in yeast (Val46-Leu84) and chemically synthesized (Tyr45-Thr87) (Handford et al., 1990), the synthesized fourth EGF-like domain of thrombomodulin (Meininger, et al., 1995), and the expressed EGF-like domain of Heregulin-α (HRG-α)(Jacobsen et al., 1996). Recently, a novel disulfide bond pattern (1-2, 3-4, 5-6) has been unambiguously determined in the fifth EGF-like domain of thrombomodulin (White et al., 1996).

EGF-like domains have been found in a large number of proteins with diverse functions. Rat and human TGFα compete with the EGFs in receptor binding (Marquardt and Todaro, 1982). In addition to the family of growth factors, there are some viral proteins contain sequences similar to the EGF/TGFαs. These EGF-like modules have been reported to display growth factor activity (Stroobant et al., 1985, Lin et al., 1988). The single EGF-like domain in urokinase mediates binding to the urokinase receptor (Appella et al., 1987). One or two of the EGF domains in the thrombin cofactor, thrombomodulin, binds thrombin with high affinity (Kurosawa, et al., 1988; Suzuki et al., 1989). In *Drosophila*, the protein products of neurogenic loci called notch (Whaton, et al., 1985), delta (Vässin, et al., 1987) and serrate

(Fleming, et al., 1990) are transmembrane proteins that contain up to 36 EGF-like domains. EGF domains 11 and 12 in the notch protein mediate the calcium-dependent cell-surface interaction with the delta protein, thus determining the embryonal fate of the cell (Fehon, 1990; Rebay, et al., 1991). Fibrillin is a large extracellular glycoprotein with 43 EGF-like domains. It is a constituent of the 10-12 nm diameter microfibrillar network. Mutations in the gene encoding this protein cause the Marfan syndrome (Lee et al., 1991; Maslen et al., 1991). Some multi domain proteins of the extracellular matrix (ECM) such as laminin, tenascin and thrombospondin have growth promoting activity (Engel, J., 1989). Vitamin K-dependent blood clotting proteins including factor VII, IX, and X, protein C and S are involved in the initiation and regulation of coagulation (Davie et al., 1991). Low density lipoprotein (LDL) receptor and urokinase (uPA) are involved in receptor-ligand interactions (Appella et al., 1988). The *C. elegans* spe-9 gene encodes a sperm transmembrane protein that contains ten EGF-like repeats and is required for fertilization (Singson et al., 1998).

Because of its widespread distribution in proteins, there has been considerable interest in the biological role of the EGF-like domain, which is presumed to be involved in mediated protein-protein interactions (Campbell and Bork, 1993). The EGF-like domains in vitamin K-dependent proteins do not contain the conserved amino acids in EGF and TGFα that appear to be involved at the receptor/growth factor interface (Campbell, 1990). It has been shown that the synthetic N-terminal EGF-like domain of factor IX does not bind to the EGF receptor (Huang,

et al., 1989).

Sequence homology identifies three subclass of EGF-like domains (Rees et al., 1988). One of the subclasses includes domains from the vitamin K-dependent coagulation factors as well as thrombomodulin, *Drosophila* notch, delta gene products (Handford et al., 1991) and fibrillin-1 (Pereira et al., 1993; Corson et al., 1993). This group of peptides is characterized by a series of conserved residues, which includes Asp, Asp/Asn, Gln/Glu, Asp*/Asn*, Tyr/Phe, where * denotes a β-hydroxylated residue (Rees et al., 1988; Handford et al., 1991) (Figure 1.1.2). Calcium binding to this type of EGF-like domains was demonstrated in factor IX and X, protein C, and protein S (Öhlin et al., 1988; Huang et al., 1989; Persson et al., 1989; Handford et al., 1990; Dahlbäck et al., 1990). Subsequently, calcium was shown to bind EGF-like domains from the connective tissue protein fibrillin-1 (Glanville et al., 1994; Handford et al., 1995). The function of these calcium binding EGF-like domains has been proposed as calcium mediated protein-protein interactions (Rao et al., 1995). The conserved residues are proposed to be responsible for calcium binding (Handford et al., 1990).

The functional importance of calcium binding was explored by the identification of mutations in consensus residues. Mutations altering the calcium binding consensus residues in the N-terminal factor IX EGF-like domain cause hemophilia B (Winship and Dragon, 1991; Mayhew et al., 1992). Four mutations of the conserved residues have been found in patients

with hemophilia B, i.e., factor IX Alabama, factor IX Oxford d1, factor IX Oxford d3, and factor IX London (Figure 1.1.3). The biological activity of the mutant factor IX molecules varies between less than 1% and 10% (Stenflo, 1990). More recently, a number of different mutations have been characterized in the fibrillin genes of individuals with Marfan syndrome (Dietz et al., 1993; Hewett et al., 1993; Kainulaine et al., 1994). Many of these occurred in the proposed calcium-binding EGF-like domains (Corson et al., 1993; Maslen et al., 1993). Moreover, the transmembrane proteins notch and delta have been shown to interact in *vitro* at the cell surface via their extracellular domains in a Ca^{2+}-dependent manner (Fehon et al., 1990). Each of these proteins contain multiple extracellular EGF repeats.

Figure 1.1.1 The primary structure of mEGF (Redrawn from Savage et al., 1973).

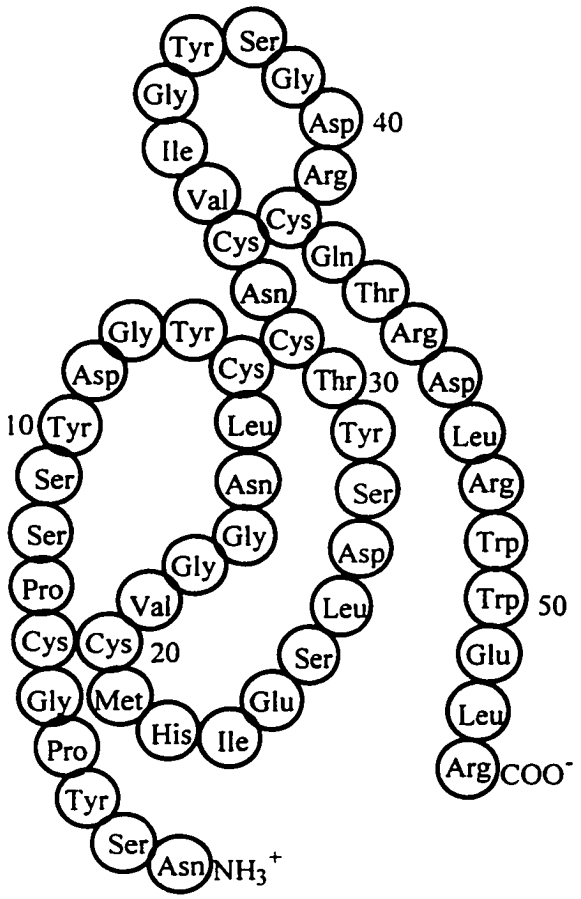
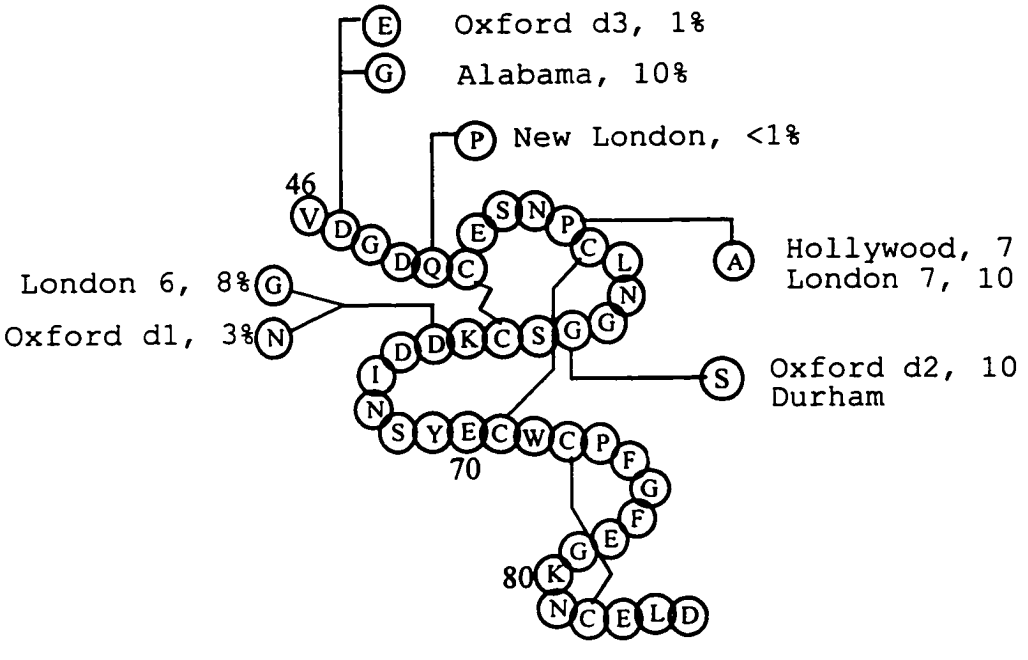


Figure 1.1.2. The sequence of EGF-like domains with homology to the first domain of human factor IX. The Asp, Asp/An, Asp*/An* (the asterisk donates a β-hydroxylated residue, Tyr/Phe consensus together with six conserved cysteine residues are marked with an asterisk (*). HFIX (human factor IX, HFX (human factor X), HFVII (human factor VII), HPS (human protein S), HTM (human thrombomodulin), NOTCH (*Drosophilia notch*). (Redrawn after Handford et al., 1990).

```
                  * ***      *     * *    * * *            *
HFIX    (47-84)  DGDQCE..SNPCLNGGSCKDDINSYECWCPFGFEG....KNCEL
HFX     (46-84)  DGDQCE..TSPCQNQGKCKDGLGEYTCTCLEGFEG....KNCELF
HFVII   (46-84)  DGDQCA..SSPCQNGGSCKDQLQSYICFCLPAFEG....RNCETH
HPS   (160-201)  DVDECSLKPSICG.TAVCKNILGDFECECPEGYRYNLKSKSCE
HTM   (325-365)  DVDDCILEPSPCP..QRCVNTQGGFECHCYPNYDL..VDGECVEP
NOTCH(449-487)  DIDECDQ.GSPCEHNGICVNTPGSYRCNCSQGFTGP....RCET
```

Figure 1.1.3. Schematic diagram of mutations known to cause hemophilia B in the N-terminal EGF-like domain of human factor IX (Redrawn after Stenflo, 1990).



1.2 Factor IX and blood clotting

Factor IX is a vitamin K-dependent blood clotting protein. There are two clotting pathways, intrinsic and extrinsic pathways (Figure 1.2.1). Blood clots are formed by a series of proteolytic activations in a cascade, the activated form of one clotting factor catalyzes the activation of the next. This allows a very small initial signal to be amplified so that it can result in a macroscopic blood clot.

Calcium mediates the binding of factor IX to the surface of a phospholipid membrane, activating it into its enzymatic form, factor IXa. This can occur either through the proteolytic action of factor XIa (intrinsic pathway), or through the proteolytic action of the factor VIIa/tissue factor complex (extrinsic pathway) (Davie et al., 1991). Factor IX is thus the crossover point between the two pathways. Activated factor IXa in the presence of factor VII and a membrane surface activates factor X to further propagate the clotting cascade. Calcium ions are required for all of these reactions. The mechanisms involve numerous interactions between the blood clotting factors, their cofactor, and phospholipid membranes. In the case of factors IX and X, the Gla which contains γ-carboxyglutamate residues and EGF modules are known to be involved in such interactions (Davie et al., 1991).

Factor IX is also known as Christmas because a hemophilia patient named Stephen Christmas was found to lack factor IX. Defects in factor IX are the second most common form of hemophilia. This form of hemophilia is called Christmas disease, or hemophilia

B.

Figure 1.2.2. shows a schematic structure of factor IX. There are four regions in this protein. The first region at the N-terminal is Gla domain, which contains 12 glutamine residues which are carboxylated to γ-caboxylglutamic acid by a carboxylase in the presence of vitamin K. The Gla domain binds to a phospholipid membrane surface in a calcium mediated activation by factor XIa. At the C-terminal is a serine protease region. A serine protease, factor IXa, is released from this region upon the activation of factor IX. Between these two regions are EGF-like domains. It has been found that this domain binds calcium (Huang et al., 1989; Handford et al., 1990).

Figure 1.2.1. Blood clotting by the intrinsic and extrinsic pathways. "a" denotes activated forms (Redrawn after Davie et al., 1991).

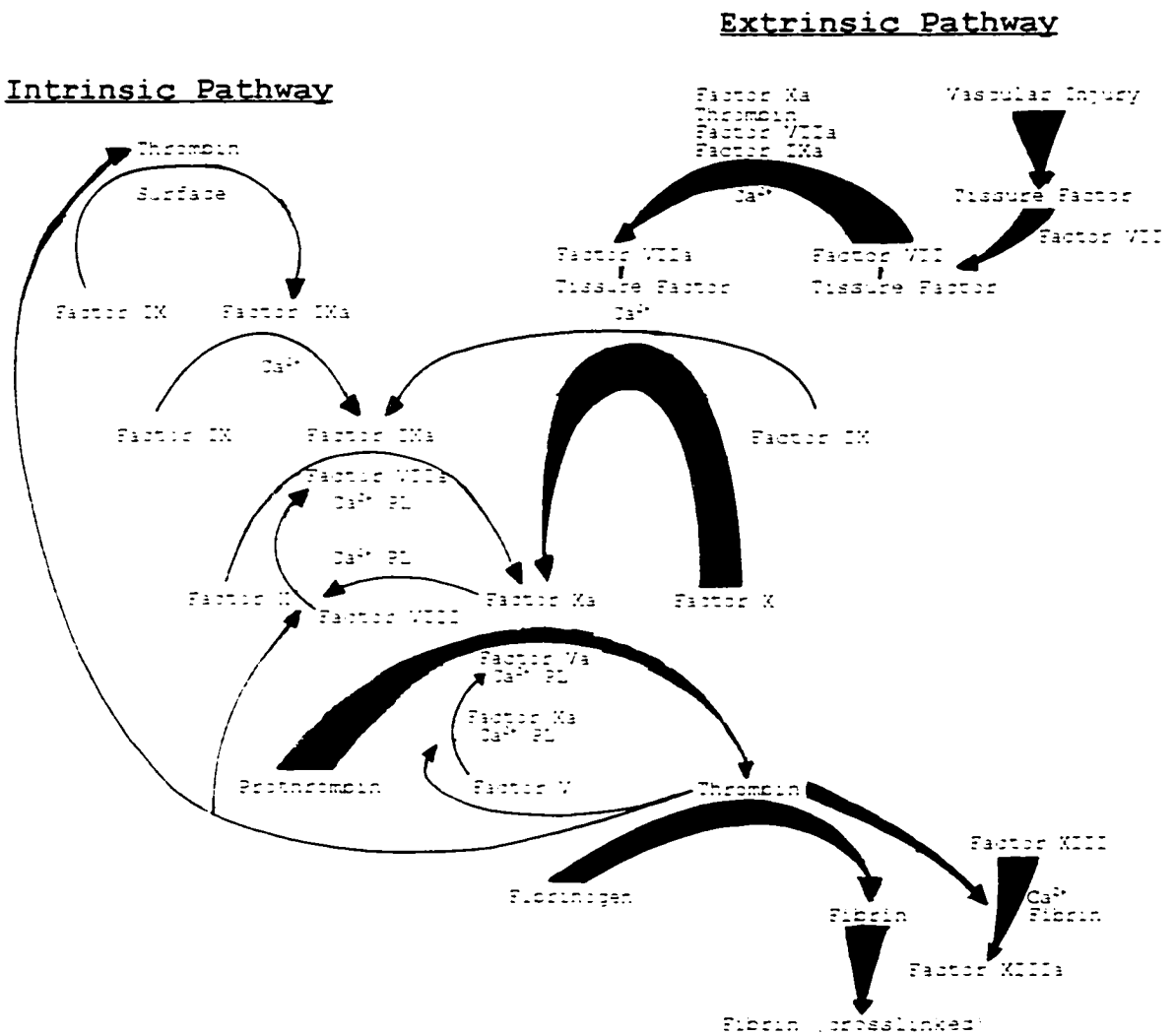
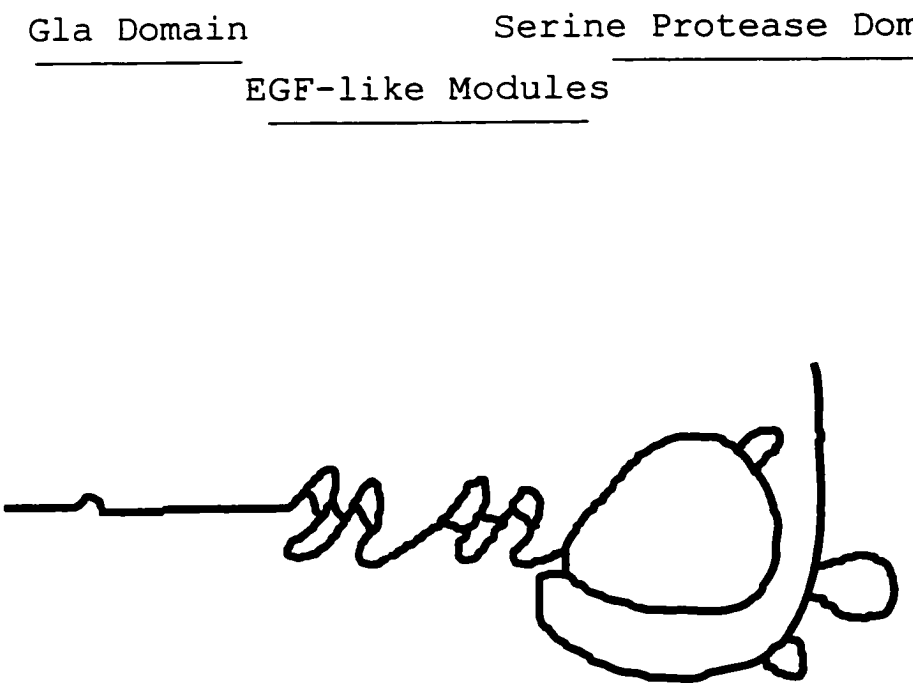


Figure 1.2.2. Schematic structure of factor IX.



1.3. Solution structure studies of EGF and EGF-like modules by NMR techniques

Although the primary structures of EGF modules are similar to EGF, it is important to determine the similarities of the tertiary structures to that of EGF, the biological functions of the EGF-like domain and the relationships of structure to function. An important approach to understanding structure-function relationships of the EGF module is to use high-resolution NMR. Solution structures derived from NMR measurements are available for human EGF (hEGF) (Carver et al., 1986; Cooke et al., 1987), murine EGF (mEGF) (Montelione et al., 1986, 1987, 1992; Konda and Inagaki, 1988, 1992a; Kohda et al., 1988), micelle-bound mEGF (Konda and Inagaki, 1992b), rat EGF (Mayo et al., 1989), human TGFα (hTGFα) (Brown et al., 1989; Campbell et al., 1989; Konda et al., 1989; Montelione et al., 1989; Tappin et al., 1989; Kline et al., 1990; Harvey et al., 1991; Moy et al., 1993), the N-terminal EGF-like domain of bovine coagulation factor X (Selander et al., 1990; Ullner et al., 1992), the N-terminal EGF-like module of human coagulation factor IX (46-84)(Baron et al 1992), and the EGF-like module of human complement protease C1r (Bersch et al., 1998).

1.4. Polypeptide backbone conformations and Ramachandran diagram

Polypeptide's backbone conformation can be specified by the torsion angles or dihedral angles about the Cα-N bond (ϕ) and the Cα-C bond (ψ) of each of its amino acid residues. These angles, ϕ and ψ,

are both defined as 180° when the polypeptide chain is in its planar, fully extended (all trans) conformation (Figure 1.4.1).

For the allowed conformations are limited by restrictions on the allowed contact distances between different atoms. An extremely useful device for studying protein conformation is the Ramachandran plot (Figure 1.4.2), a conformation diagram showing the range of allowed conformations, which is based on the ideas of the nonbonded contacts between atoms (Ramachandran et al., 1963). A Ramachandran diagram in Figure 1.4.2 shows the sterically allowed ϕ and ψ angles for poly-L-alanine. The diagram was calculated under the van der Waals distances in Table 1.4.1. There are three regions of freely allowed values of ϕ and ψ for an alanyl residue in a polypeptide, which are circled with the solid lines. The dotted lines enclose "outer limit" ϕ and ψ values based on the shortest known van der Waals radii in related structures (Table 1.4.1). The conformation angles, ϕ and ψ, of several secondary structures are indicated in Table 1.4.2.

Glycine, the only residue without a C_β atom, is much less sterically hindered than other amino acid residues. This is clearly apparent in comparing the Ramachandran diagram for glycine in a polypeptide chain (Figure 1.4.3). In fact, glycine often occupies positions where a polypeptide backbone makes a sharp turn which, with any other residue, would be subject to steric interference.

Figure 1.4.1. A portion of a polypeptide chain indicating the torsion degrees of freedom of each peptide unit. The only reasonably free movements are rotations about the C_α-N bond (ϕ) and the C_α-C bond (ψ). The torsion angles are both 180^0 in the conformation shown and increase in a clockwise manner when viewed from C_α.

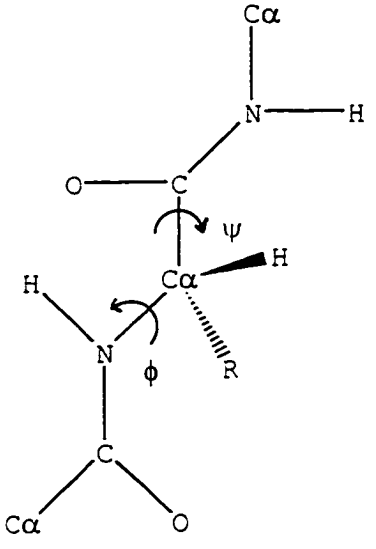


Figure 1.4.2. A Ramachandran diagram of the sterically allowed ϕ and ψ angles for poly-L-alanine. Regions of "normally allowed" ϕ and ψ angles are shaded in dark, whereas the lighter shaded regions correspond to conformations having "outer limit" van der Waals distances. The symbol of parallel arrow represents for parallel β-sheet; antiparallel arrow for antiparallel β-sheet; II for left-handed polyglycine II and poly-L-proline II helices; C for collagen; αL for left-handed α helix; and αR for right-handed α helix. (After Voet, D. and Voet, J. G., Biochemistry, Second Edit., John Wiley & Sons, Inc., New York).

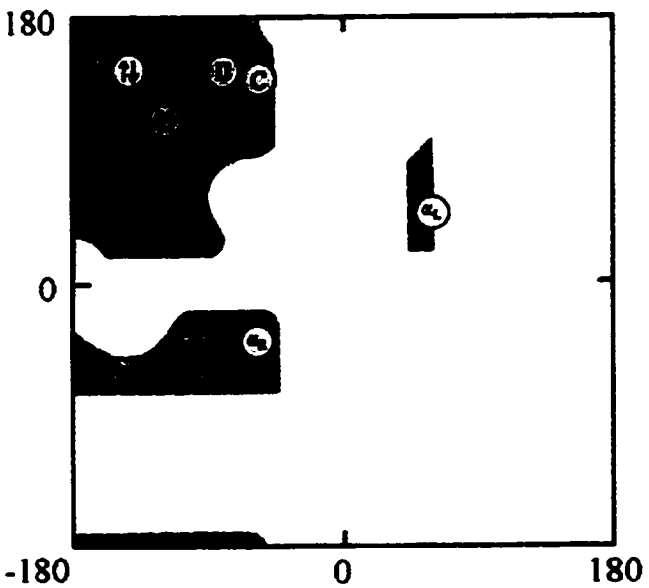


Figure 1.4.3. A Ramachandran diagram of glycine residues in polypeptide chain. "Normally allowed" regions are shaded in dark, whereas light-shaded regions correspond to "outer limit" atomic distances. Glycine residues have far greater conformational freedom than do other (bulker) amino acid residues as the comparison of this figure with Figure 1.4.2 indicates. (Redrawn after Voet, D. and Voet, J. G., Biochemistry, Second Edit., John Wiley & Sons, Inc., New York)

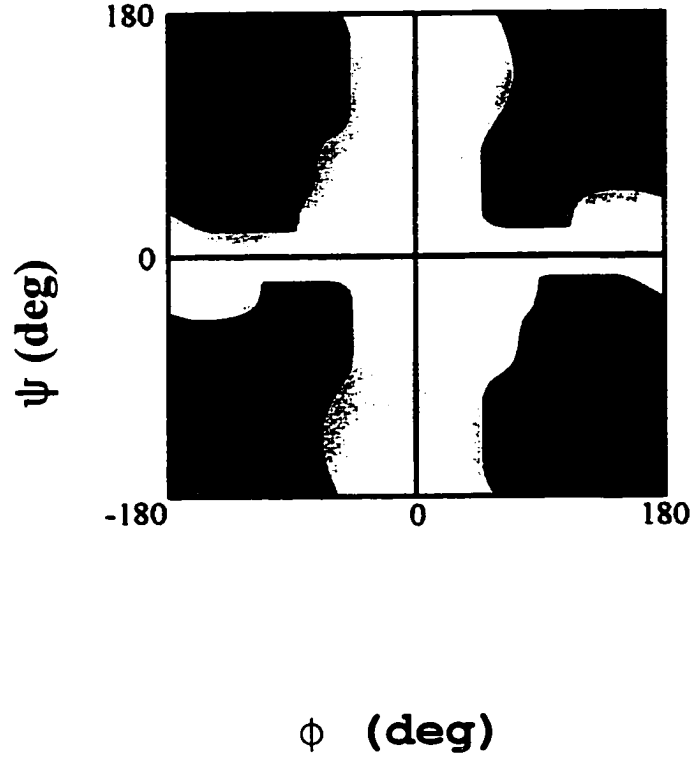


Table 1.4.1. van der Waals Distances for interatomic contacts (After Voet, D. and Voet, J., 1995).

Contact Type	Normally Allowed (Å)	Outer Limit (Å)
H...H	2.0	1.9
H...O	2.4	2.2
H...N	2.4	2.2
H...C	2.4	2.2
O...O	2.7	2.6
O...N	2.7	2.6
O...C	2.8	2.7
N...N	2.7	2.6
N...C	2.9	2.8
C...C	3.0	2.9
C...CH_2	3.2	3.0
CH_2...CH_2	3.2	3.0

Table 1.4.2. ϕ and ψ angles of several secondary structures (After Voet, D. and Voet, J., 1995).

Secondary Structure	ϕ (deg)	ψ (deg)
Right-handed α helix	-57	-47
Parallel β pleated sheet	-119	+113
Antiparallel β pleated sheet	-139	+135
Right-handed 3_{10} helix	-49	-26
Right-handed π helix	-57	-70
2.2_7 ribbon	-78	+59
Left-handed polyglycine II and poly-L-proline II helices	-79	+150
Collagen	-51	+153
Left-handed α helix	+57	+47

1.5. Turns in protein structures

A turn is defined as a site where the polypeptide chain reverses its overall direction. The term β and γ turn have more restricted definitions and describe turns of four or three residues, respectively. These turns may or may not be stabilized by an intra-turn hydrogen bond (Rose, et al., 1985). The possible ϕ, ψ angles for H-bonded β turns were determined by Nemethy and Printz (1972). Table 1.5.1 lists these turn classifications for both β and γ turns (Rose, et al., 1985).

Table 1.5.1 H-bonded β and γ turns: dihedral angles

	i + 1		*i* + 2	
Turn	ϕ	ψ	ϕ	ψ
β turns				
Type I	-60	-30	-90	0
Type I′	60	30	90	0
Type II	-60	120	80	0
Type II′	60	-120	-80	0
Type III	-60	-30	-60	-30
Type III′	60	30	60	30
Type VIa	-60	120	-90	0
Type VIb	-120	120	-60	0
γ turns				
Turn	70~85	-60~-70		
Inverse turn	-170~-85	60~70		

1.6. 2D-NMR solution structure determination.

During the last decade a method for the determination of the complete three-dimensional structure of proteins was developed, which uses NMR for data collection and distance geometry, or other mathematical techniques, for the structural interpretation of the NMR data (Wütherich et al., 1982; Wütherich, K., 1986).

1.6.1. Sequence-specific resonance assignments.

Sequence-specific resonance assignments are essential for the study of the solution structure of proteins by NMR. The structure of an amino acid residue is shown in Figure 1.6.1 and Table 1.6.1.

Figure 1.6.1. Amino acid residue. R is the side chain. The circle identifies the labile, but often NMR observable amide proton. (Redrawn from Wütherich, K., 1986).

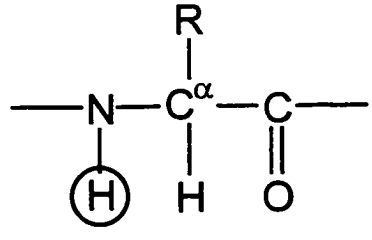
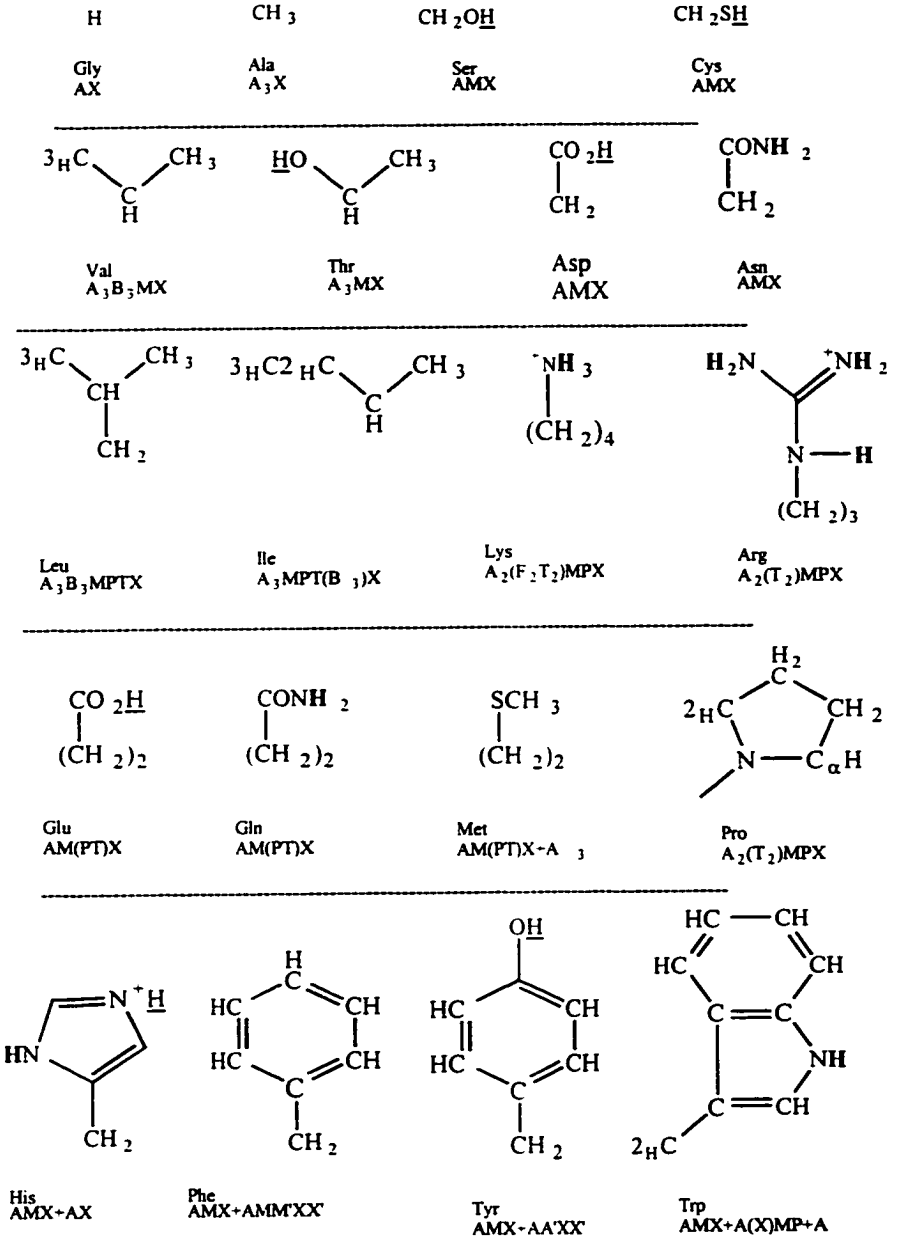


Table 1.6.1. Side chains R (see Figure 1.6.1) and three-letter symbol for the 20 common amino acids, and spin systems of the nonlabile hydrogen atoms in the molecular fragments H-Cα-R. (Redrawn after Wüthrich, K., 1986.)



Labile protons that can not be observed by NMR in aqueous solution are shown by underline. Labile protons that can be observed by NMR under certain conditions are bolded.

Usually, a protein NMR spectrum is run in D_2O or H_2O for different purposes. In D_2O solution the labile protons can be replaced by deuterium. In this case, only the carbon bound hydrogen resonances are obtained in the ^1H NMR spectrum, which simplifies the spectrum. For each amino acid residue, these nonlabile protons constitute one or more than one spin systems (Wütherich, K., 1986). A spin system is a group of spins that are connected by scalar (through-bond) spin-spin couplings J.

COSY and NOESY are the most commonly employed spectral techniques in protein 2D NMR. In a 2D spectrum, the diagonal peaks with $\omega_1 = \omega_2$ represent the 1-D spectrum. In addition, there are a large number of cross peaks with ω_1 not equal to ω_2. Each cross peak establishes a correlation between two diagonal peaks. The cross peaks of COSY arise from through-bond connectivity, while the cross peaks of NOESY from proton-proton interaction through space. The distance required for this through-space interaction must be less than 5 Å. A number of various forms of COSY experiments are designed for different purposes. Commonly used COSY experiments include relayed-COSY, total correlated spectroscopy (TOCSY), and double quantum filtered COSY (DQF-COSY). The DQF-COSY experiment gives connectivities between directly coupled proton separated by two or three bonds. In a relayed-COSY spectrum, the through-bond connectivities between four bonds are observed. The TOCSY experiment can connect almost all side chain protons together in the same system.

To identify the ^1H spin system of the individual

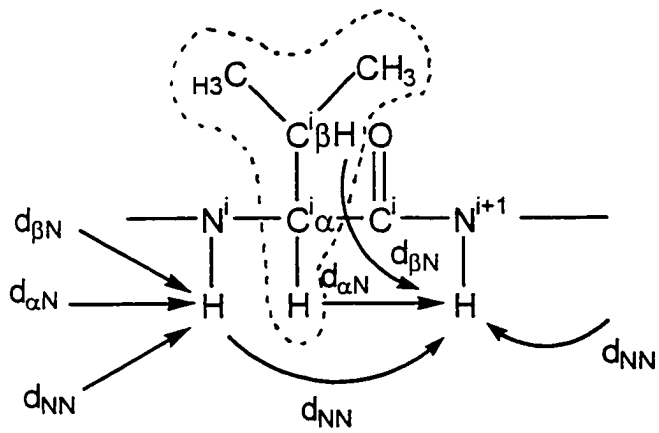
amino acid residues in a protein, one will normally start with homonuclear-correlated spectroscopy (COSY) in D_2O solution after replacement of all labile protons with deuterium. The spin systems consist of the non-labile protons in the same amino acid residues (Table 1.6.1). In the COSY spectral analysis, one identifies those groups of J-coupled resonances all of which arise from the same amino acid residue.

For the further studies, the amino acid spin systems of the individual residues are completed through identification of the connectivity with the labile protons in H_2O solution of the native protein. The backbone NH-αH region of the H_2O spectra is of special interest, since it presents a fingerprint of the amino acid sequence (Wüthrich, K., 1986). Once the spin systems are identified, further differentiation into the amino acids that contain the same spin system is important. NOESY spectra are used to resolve this problem. For example, there are eight amino acids that contain an $C^{\alpha}H-C^{\beta}H_2$ fragment (Asp, Asn, Ser, Cys, His, Phe, Tyr and Trp). Among them, the four aromatic side chain can reliably be identified from observation of the short $C^{\beta}H_2$-ring proton distances by NOESY, provided that both the AMX systems and the aromatic spin systems were independently identified. NOESY spectra in D_2O simplify the observation of cross-peaks to aromatic protons for the identification of the spin systems of the aromatic residues.

Having identified the spin systems as much as possible, the next major work is sequential

assignments. The information needed for obtaining the sequential connectivities between neighboring spin systems is obtained from analysis of the regions containing the cross peaks of $d_{\alpha N}$, d_{NN}, and $d_{\beta N}$ in NOESY spectra recorded in H_2O (Figure 1.6.2). The last step for identification and resonance assignment, which are obtained by the information on amino acid type, along with nuclear Overhauser effects (NOEs) between the known amino acid sequence.

Figure 1.6.2. Polypeptide segment with indication of spin systems of nonlabile protons in the individual residual (inside dotted lines), the αH-NH COSY connectivities (broken lines), and the sequential NOE connectivities (arrows). (Redraw from Wüthrich, K., 1986).



1.6.2. Stereo-specific Resonance Assignments.

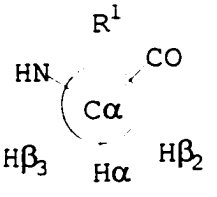
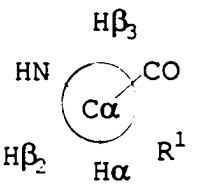
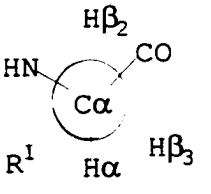
In a study of solution structure of proteins by NMR, the lack of stereo-specific assignments of diastereotopic protons in CH_2 groups and of methyl groups in leucine and valine residues can create the large uncertainties in the distance constraints. In many cases, stereo-specific assignments are possible, by analyzing the data from COSY spectrum, along with the observation of certain NOESY cross peaks (Basus, V. J., 1989).

i. Stereo-specific Assignment of β Protons.

Figure 1.6.3 shows the most energetically favorable conformations, g^2g^3, g^2t^3, and t^2g^3, around the C^α-C^β bond in a staggered conformation and the criteria of the stereo-specific assignment and χ^1 angle determination. *g* and *t* refer to gauche or trans orientation of the β-protons with respect to the α-proton, where β^2 and β^3 are defined according to the IUB-IUPAC convention (IUB-IUPAC Commission of Biochemical Nomenclature (1970) J. Mol. Biol. 52, 1). Next, the coupling constants between α and β-protons, and the intra-residue NOEs between α and β-protons and between amide and β-protons, respectively, are also listed in Figure 1.5.5. In the case of g^2g^3, one expects two small H^α-H^β coupling constants, two strong H^α-H^β NOEs, a weak NH-$H^{\beta2}$ NOE and a strong NH-$H^{\beta3}$ NOE. For g^2t^3, one expect a small H^α-$H^{\beta2}$ coupling constant, a large H^α-$H^{\beta3}$ coupling constant, a strong H^α-$H^{\beta2}$ NOE, a weak H^α-$H^{\beta3}$ NOE, and two strong NH-H^β NOEs. For t^2g^3, one expect a large H^α-$H^{\beta2}$ coupling constant, a small

H^{α}-$H^{\beta 3}$ coupling constant, a weak H^{α}-$H^{\beta 2}$ NOE, a strong H^{α}-$H^{\beta 3}$ NOE, a strong NH-$H^{\beta 2}$ NOE and a weak NH-$H^{\beta 3}$ NOE. With these criteria, the stereo-specific assignments of β protons can be obtained.

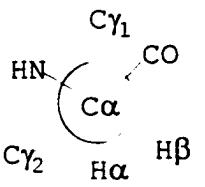
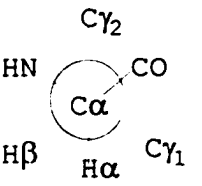
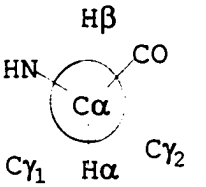
Figure 1.6.3. Characterization of the torsion angle χ^1 by NMR data (Redraw from Basus, V. J. 1989). Gauche and trans orientations are indicated as *g* and *t*, respectively. The β^2 and β^3 positions are defined according to the IUB-IUPAC convention. A Newman projection is shown for each of the conformations and the χ^1 angle is indicated. The coupling constants are obtained with relation: $^3J_{\alpha\beta}$ = (9.5 $\cos^2$ θ-1.6cos θ + 1.8)Hz (de Marco et al., 1978).

Conformation	g^2g^3	g^2t^3	t^2g^3
χ^1	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 (α,β_2)	S	S	W
NOE (α,β_3)	S	W	S
NOE (NH,$\alpha\beta_2$)	W	S-M	S
NOE (NH,$\alpha\beta_3$)	S-M	S	W

ii. Stereo-specific Assignments of Valine Methyl Protons

It has been demonstrated that stereo-specific assignments of the valine methyl protons can be obtained in some cases (Zuiderweg et al., 1985). Figure 1.6.4. shows three energetically favorable rotameric states about the C^{α}-C^{β} bond of a valine residue, g^-, t, and g^+. Both g^- and t rotamers would be expected to exhibit a small $^3J_{\alpha\beta}$ coupling constant (less than 5 Hz) and so it usually not possible to distinguish these two cases from each other. In contrast g^+ would have a large $^3J_{\alpha\beta}$ (about 12 Hz) and is quite distinct from the other two rotamers. The only way to distinguish between g^- and t conformations would be through the NOEs from the NH to the β proton, which should be large for the t conformation and small for g^- (Basus, V. 1989). Both of the H^N-$C^{\gamma}H_3$ NOEs in a g^- conformation would be relatively strong, while in the t and g^+ conformations one would be strong and the other would be weak. Of these, the g^+ conformation is the most predominant, as is shown in a statistical analysis of valine side-chain conformations of proteins with known X-ray crystal structures (Janin et al., 1978).

Figure 1.6.4. Characterization of the three possible rotamers around the C^{α}-C^{β} bond for valines. A Newman projection is shown for each conformation. The γ^1 and γ^2 methyl protons are defined in accordance with IUB-IUPAC convention. (Redrawn from Basus, V. 1989).

Conformation	*g*⁻	*t*	*g*⁺
χ^1	60⁰	180⁰	-60⁰
			
$^3J\alpha\beta$ (Hz)	2.6-5.1	2.6-5.1	11.8-14.0
NOE (α,γ^1)	W	S	S
NOE (α,γ^2)	S	W	S
NOE (NH,γ^1)	S	W	S
NOE (NH,γ^2)	S-M	S	W

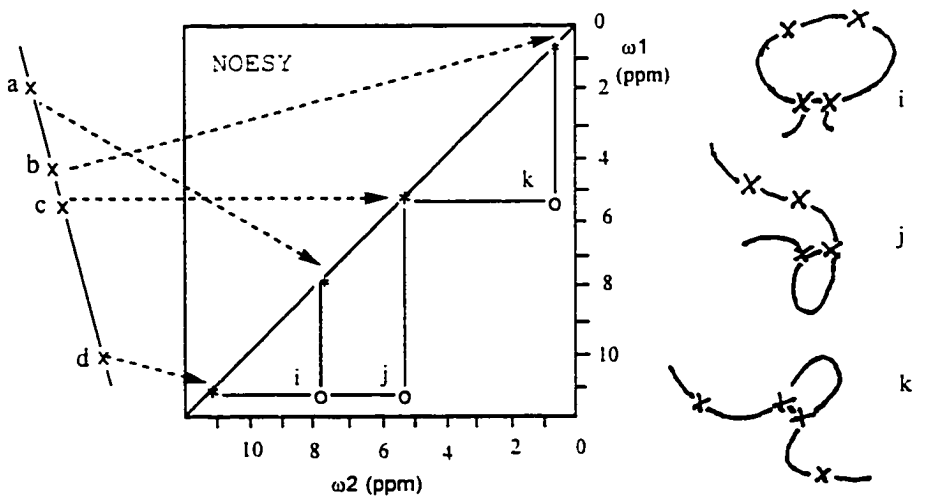
1.6.3. Protein Structure Determination by Distance Geometry.

A set of constraints could be obtained after completing as much as proton resonance assignments, including distance constraints from NOESY cross peaks; torsion angle constraints from coupling constants; and H-bond constraints. These constraints will be used as input data to calculate 3D protein structure using distance geometry (William et al., 1985; Havel, T. F. and Wüthrich, K. 1984; Braun et al., 1981; 1983). Restrained molecular dynamics calculations either can be used as an alternative to distance geometry or as a supplementary technique for refinement of the distance geometry conformations (Scheek et al., 1989).

Figure 1.6.5 illustrates the description of the NMR method for protein structure determination in solution. The sequence-specific resonance assignments have been obtained at this stage. That is for all the protons in the polypeptide chain, the corresponding diagonal peaks have been identified, as is indicated in Figure 1.6.5 by the dashed arrows linking the protons a to d with their diagonal positions in the NOESY spectrum. Each NOESY cross peak indicates that two specific protons in the polypeptide chain are separated by a distance of less than 5 Å in the three-dimensional protein structure. On the right of Figure 1.6.5, it is schematically indicated by the formation of the three circular structures, arise from the three NOESY cross peaks i-k. With as much as possible NOE cross peaks from NOESY spectrum, the tertiary structure of the polypeptide could be computed, which satisfies all the experimental

constraints (Wüthrich, K., 1989).

Figure 1.6.5. Illustration of the description of the NMR method for protein structure determination in solution. In the center, a schematic NOESY spectrum of a protein is shown, with two frequency axes ω1 and ω2. Three cross peaks are marked I-k and linked by horizontal and vertical lines with the diagonal positions of the protons connected by the corresponding NOEs. On the left, an extended polypeptide chain is represented by a straight line, and four protons in this chain are identified by circles and the letters a-d. The broken arrows connect three protons with their resonance positions on the diagonal of the NOESY spectrum. On the right, there is a schematic representation of three circular structures formed by the polypeptide chain, which are manifested by the cross peaks i-k (after Wüthrich, K., 1989).



1.6.4. DGII Program in Insight II.

The distance geometry program used (DGII, Insight II) is based on the EMBED algorithm, which consists of the following three steps:

1. Bound smoothing. The procedure starts with the incomplete and imprecise set of distance bounds that are experimentally available and computes a complete and more precise set of bounds. This step predicts the range of values that the interatomic distance can assume in any conformation of the molecule which also satisfies the given distance constraints. The DGII package includes two programs for bounds smoothing, one of which does triangle bound smoothing, which is fast, and the other of which does tetra angle bound smoothing, which is much slower but produce a better approximation.

2. Embedding. In this step, a random guess is made at the values of the distances from within the bounds obtained by bound smoothing, and atomic coordinates are obtained such that the distances calculated from these coordinates are a "best-fit" to this guess. Embedding consists of three steps: metrization, which generates a matrix of random trial distances that satisfies the inequality limits; embedding, which converts this matrix into a set of random atomic coordinates and majorization improves a weighted least squares fit between the trial distances and the coordinates with 4D.

3. Optimization. This step minimizes the deviations of the coordinates from the distance bounds as well as the chirality constraints by any of

the many available methods. The purpose of this procedure is to reduce the violations of the constraints by the embedded coordinates to an acceptable level. Optimization involves minimizing a function (called an error function) which measures the total violation of the constraints by the coordinates.

Simulated annealing as a Global optimization technique is employed by DGII to avoid the problems of many local minima. DGII imposes an upper bound T_{max} on the temperature, which is gradually reduced to zero according to the schedule:

$$T_{max} = T^{0}_{max} \ (3x^{2}_{i} - 2x^{3}_{i})$$

Here, T_{max} is the initial upper bound on the temperature (typically 200K), and x_i is the extent of the annealing on the I-th step given by $(S_{max} - I)/(S_{max})$, where S_{max} is the total number of steps taken by the annealing.

The best initial energy to use can only be determined by a process of trial and error. First, if the temperature does not rise far enough (i.e., the temperature, column 3 of output, never rises to the Max_Temp degrees), the initial energy must be increased; if it tries very hard to rise past the upper bound on the temperature (as evidenced by a cooling factor, column 5 of output, significantly less than 1), the initial energy must be decreased.

The coordinates obtained from the above annealing procedure are generally close to a minimum of the error function (hopefully near the global one). The

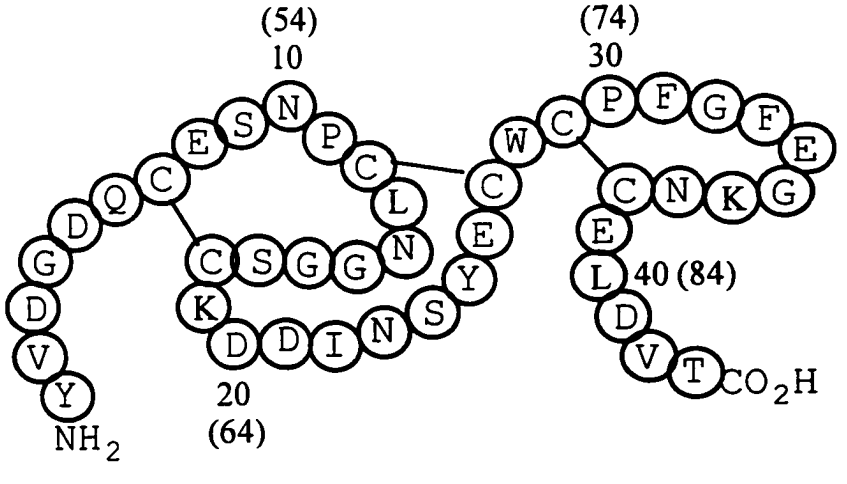
error function is then further minimized to via a conjugate gradient procedure.

Chapter II. Solution structure of the N-terminal EGF-like domain of blood clotting factor IX (FIX-EGF_N) by 2D ^1H NMR techniques and distance geometry approach.

2.1. The primary structure of FIX-EGF_N.

The primary structure of this EGF-like peptide is shown in Figure 2.1.1. There are forty three amino acids with six cysteins, which form three disulfide bonds in the pattern of 1-3, 2-4, and 5-6 in this peptide.

Figure 2.1.1. The primary structure of FIX-EGF_N (45-87). For convenience, the residues in this EGF-like domain of FIX (45-87) are numbered sequentially from 1 to 43 in this thesis. The numbers in the parenthesis indicate factor IX sequences.



2.2. Materials and NMR Experiments

Peptide sample: The N-terminal EGF-like domain of factor IX (45-87) (Figure 2.1.1) was synthesized chemically in our group using a solid-phase method (Huang et al., 1989).

2D NMR spectra: Peptide samples in H_2O were prepared in 90% H_2O/10% D_2O solution at pH 4.2 with 2 mM (Huang, L. 1991). All NOESY and PECOSY spectra were acquired in Professor Pardi's laboratory on a Varian VXR-500S spectrometer operating at 499.84 MHZ at 25^{0}C at University of Colorado. All NOESY spectra were recorded with a spectral width of 7000 Hz and 2K x 0.3K complex time domain data points. The NOESY spectra were acquired with mixing times of 1, 40, 80, 120, and 200 ms. The detailed experimental conditions have been described elsewhere (Huang et al., 1991). DQF-COSY experiments were performed using a Varian Unity-plus 500 MHZ spectrometer operating at 25 ^{0}C. DQF-COSY spectrum was collected with a spectral width of 6000 Hz and 2K X 0.5K complex time domain data points.

Time domain convolution and linear prediction of the first point were used for NOESY and DQF-COSY spectral processing on the t2 time-domain data. Time domain convolution using a Gaussian window with window width 41 to eliminate huge solvent lines and the effect of the "tails" of these lines on less intense signals. Linear prediction of the first point(s) was used to estimate the values of incorrectly acquired point(s) from subsequent data points. Values of one through four data points were predicted. Both time-domain data sets of each

spectrum were multiplied by a 75 or 90 degree phase-shifted sinebell-squared window function for NOESY. A zero degree phase-shifted sinebell window function was used for DQF-COSY and PECOSY. Zero-filling to 2K x 2K and 4K x 1K frequency domain points were performed respectively for NOESY spectra and DQF-COSY spectrum. A 5th or 6th order polynomial baseline correction using manually selected baseline points was applied to NOESY and DQF-COSY in the t2 dimension. All NMR data is processed and analyzed using Felix (Biosym Technologies, San Diego, CA) installed on an IBM RS6000 platform.

2.3. Stereospecific Resonance Assignments.

2.3.1. Stereo-specific assignments of β-methylene protons.

The stereo-specific assignments of β-methylene protons were determined by consideration of the HN-H^{β}, H^{α}-H^{β} NOE intensities and the $^{3}J(H^{\alpha}-H^{\beta})$ coupling constants (Wagner et al., 1987). Figure 1.6.4 in chapter I shows the criteria of the stereo-specific assignment and χ^{1} determination. The χ^{1} angles for the 3 staggered side-chain rotamers and the corresponding conformations in a Newman projection are given in Figure 1.6.4. Next, the coupling constant criteria between α and β-protons, and the intra-residue NOEs between α and β-protons and between amide and β-protons, respectively, are also listed in Figure 1.6.4. The HN-H^{β}, H^{α}-H^{β} NOEs were estimated from NOESY cross-peak intensities in the 90% H_2O/10% D_2O or 100% D_2O NOESY spectrum respectively (with τm= 40ms)

(Figure 2.3.1 and Figure 2.3.2). $^{3}J(H^{a}-H^{\beta})$ coupling constants were measured from PE-COSY or DQF-COSY measured in 100% D_2O (Figure 2.3.3 and Figure 2.3.4).

Figure 2.3.1. A portion of the NOESY contour plot in 90% H_2O/10% D_2O

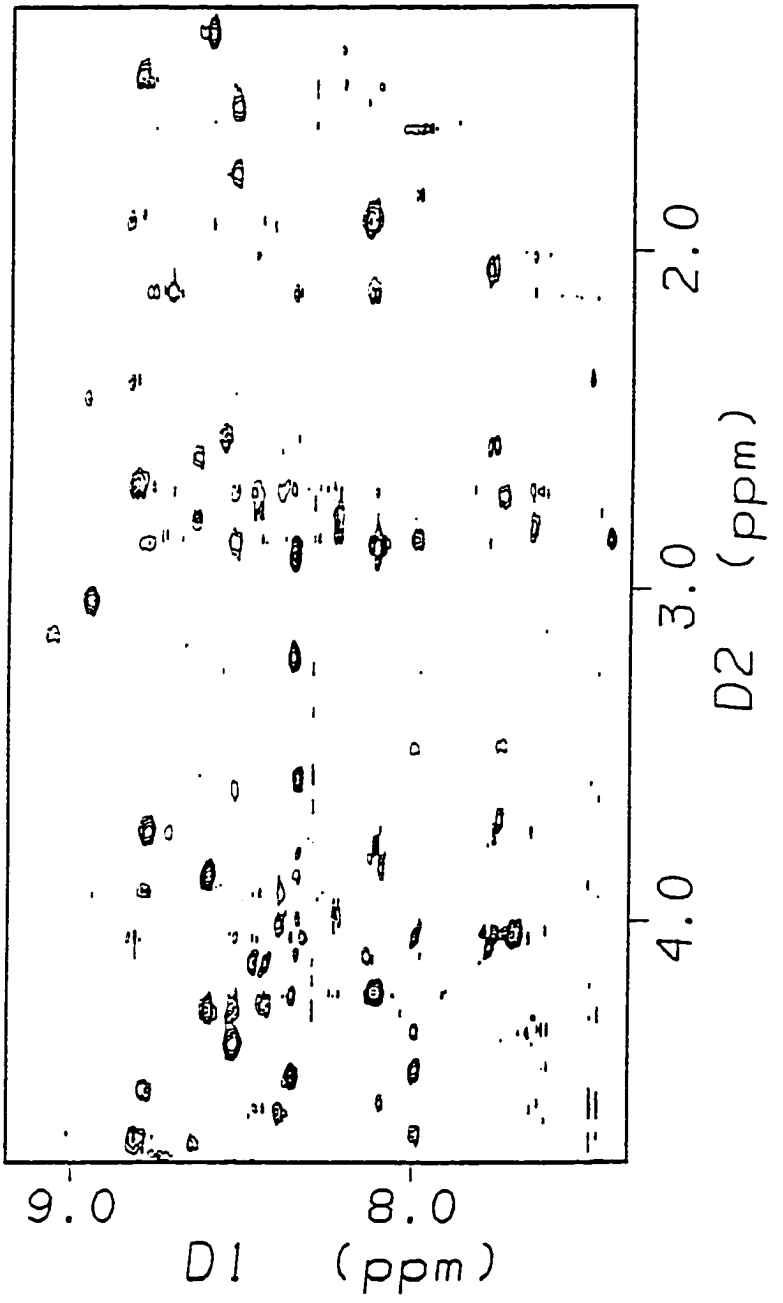


Figure 2.3.2. A portion of the NOESY contour plot in 100% D_2O

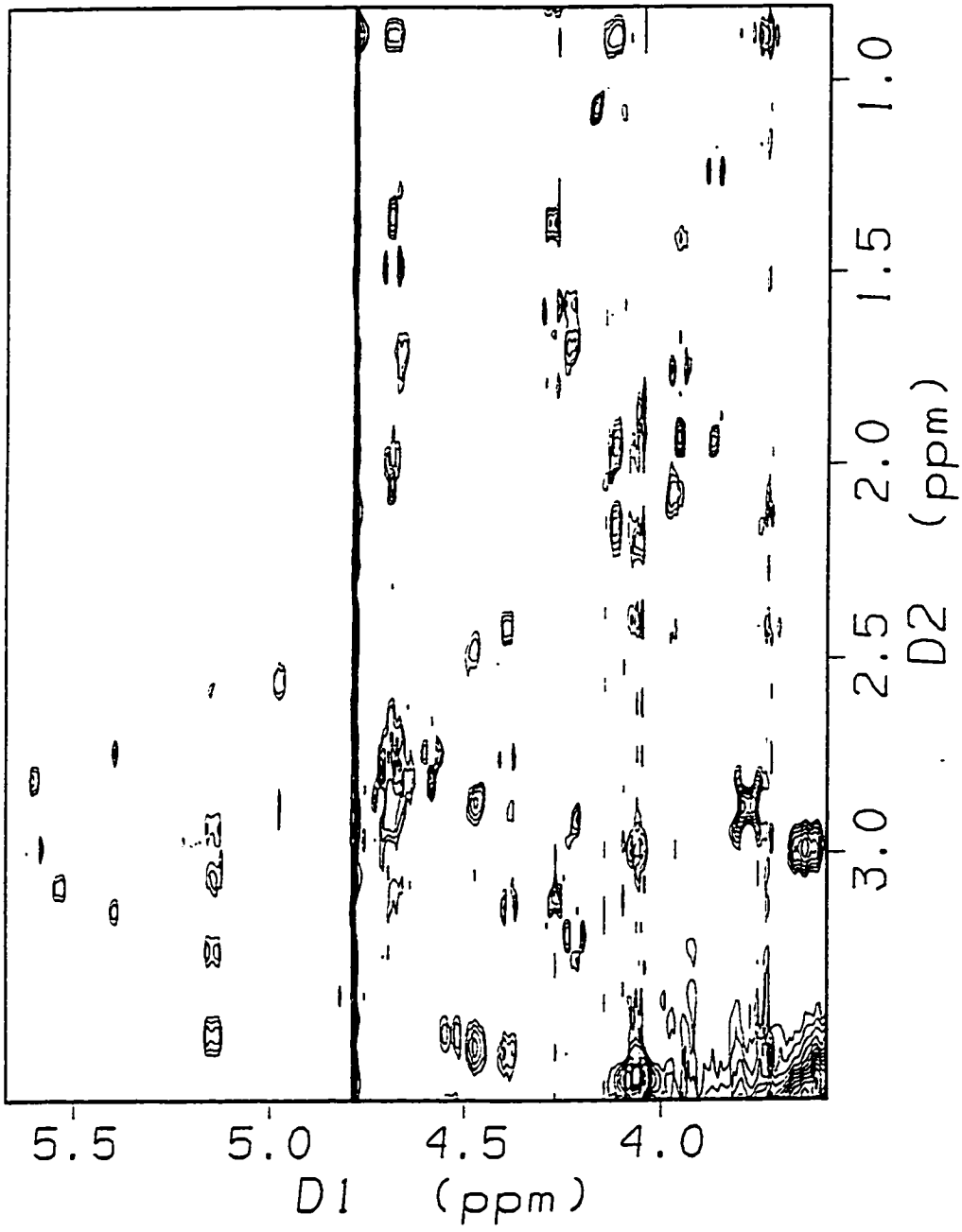


Figure 2.3.3. A portion of the PE-COSY contour plot in 100% D_2O

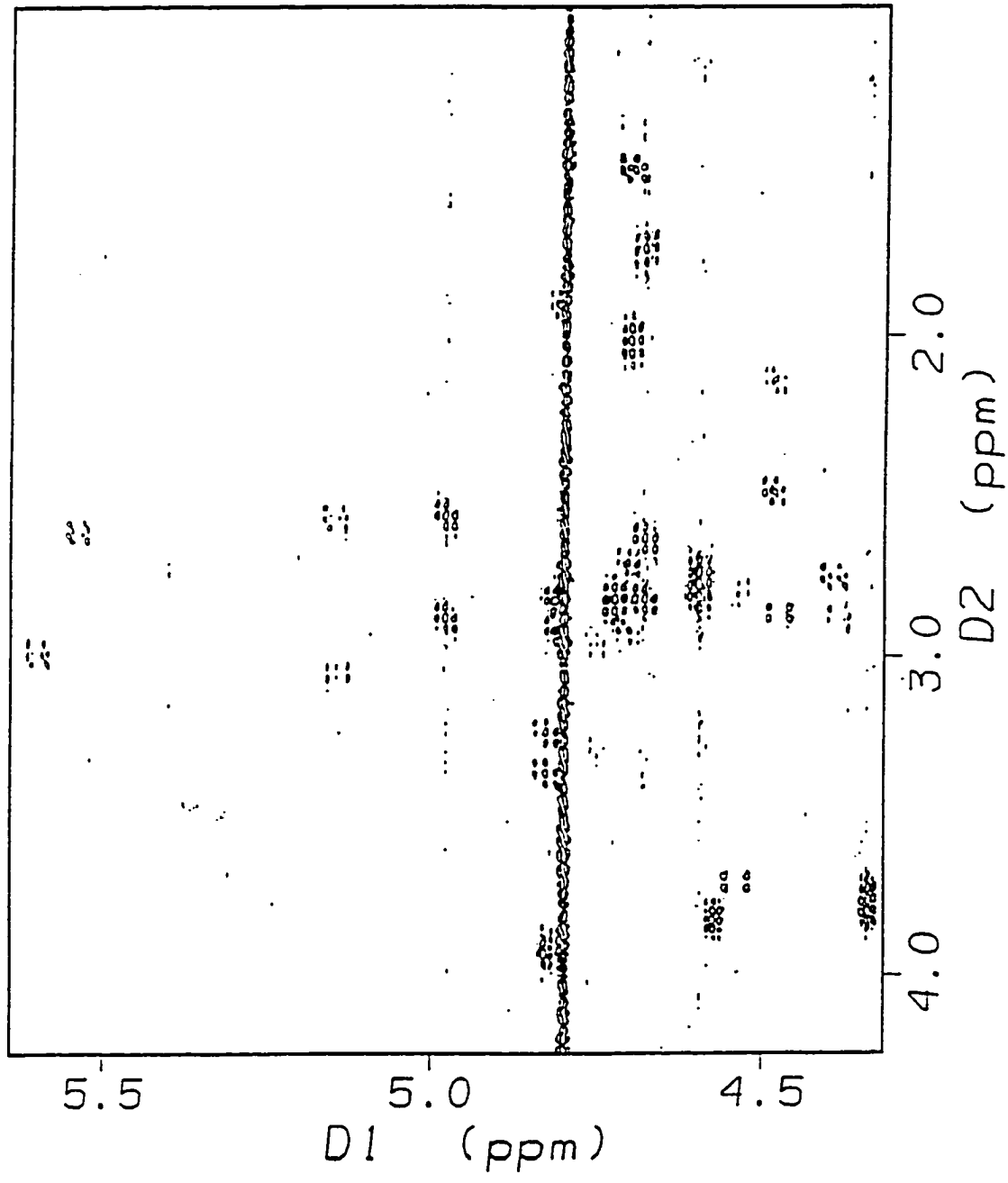
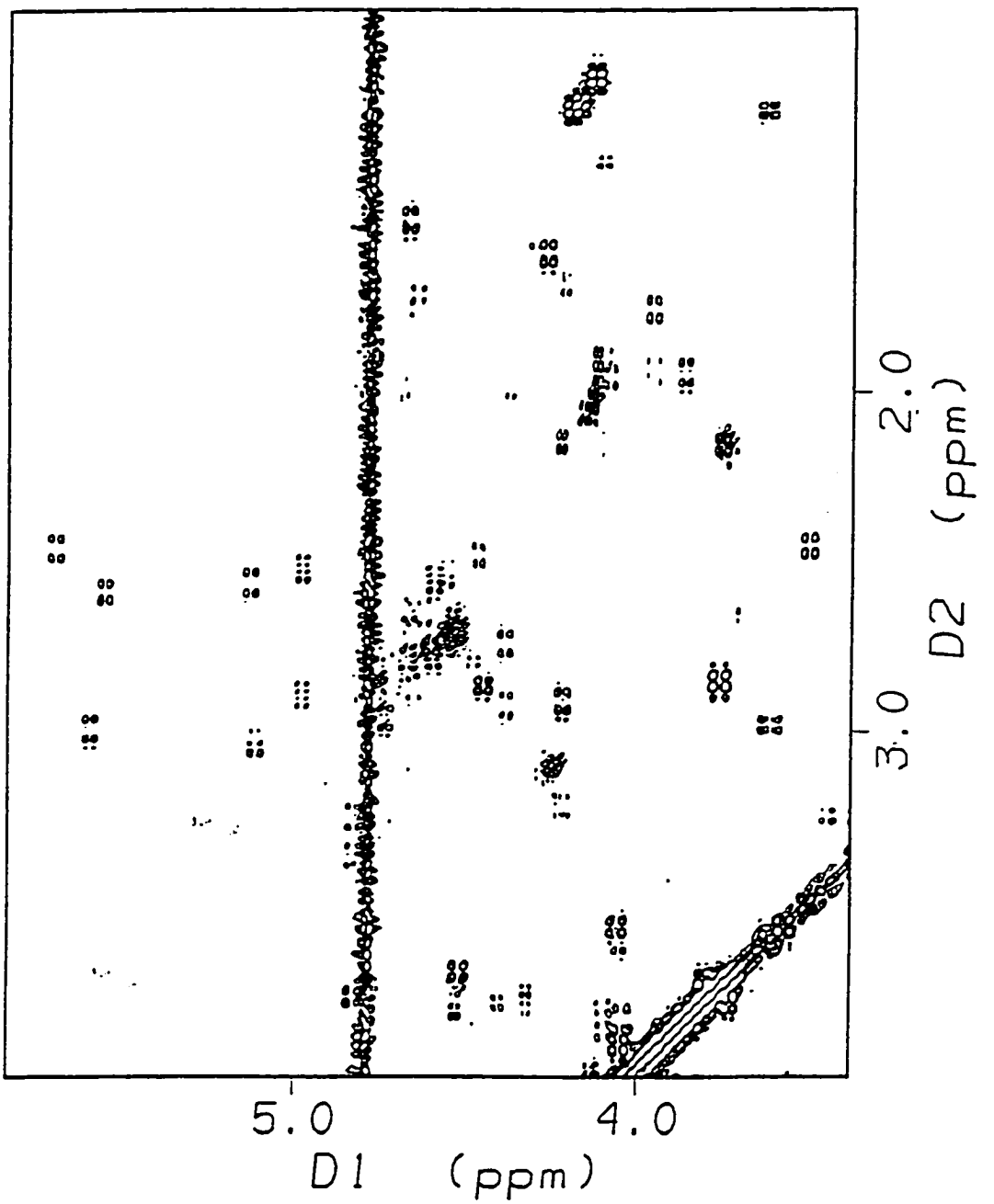


Figure 2.3.4. A portion of the DQF-COSY contour plot in 100% D_2O



2.3.2. Stereospecific assignment of valine methyl groups.

The stereo-specific assignment of the valine methyl groups were deduced from the $^{3}J(H^{\alpha}-H^{\beta})$ coupling constant and the relative magnitude of the $H^{N}-C^{\gamma}H_{3}$ NOEs (Figure 1.6.5). For the criteria used for stereo-specific assignment for valine methyl groups, refer to section 1.6.2. ii, and Figure 1.6.5. The relative intensities of $H^{\alpha}-C^{\gamma}H_{3}$ NOEs can be useful supplementary information (Montelione et al., 1992).

2.3.3. Stereospecific assignment of α-methylene protons of Gly residues.

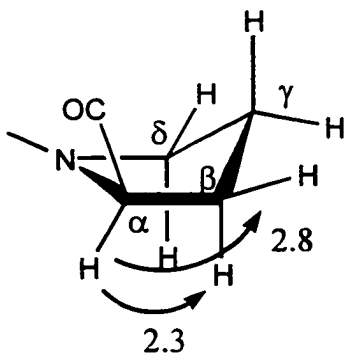
The α-methylene protons of Gly residues are floated between pro-R and pro-S configuration (Weber et al., 1988) using the Insight II DGII module. In DGII, restraints to HA^{1} and HA^{2} for glycine are "floating" between the two atoms, because DGII does not impose chirality restraint unless explicitly set as R and S. For example, if you designate an HA^{1} restraint to be 1 Å and HA^{2} restraint to be 2 Å, at the end of the run DGII will call the atom closer to 2 Å HA^{2} and call the one closer to 1 Å HA^{1}. There is no energy barrier to doing this for DGII. If the observed chirality of a particular HA^{1} and HA^{2} is consistent at the 95% confidence level of the calculated structures, the stereo-specific assignment would be made. HA^{1}, HA^{2} nomenclature was used arbitrarily for the down-field shifted protons and up-field shifted protons respectively. For a given germinal pair of protons, the larger of the two upper-bound NOE distance constraints was used for

both protons.

2.3.4. Stereo-specific assignment of proline β-methylene protons.

The prochirality of proline β-methylene protons were determined by the relative $C^{\alpha}H$-$C^{\beta}H$ NOE intensities from the 40 ms mixing time NOESY spectrum, based on the fact that in proline the $C^{\alpha}H$ proton is closer to the $C^{\beta}H_S$ proton than to the $C^{\alpha}H$-$C^{\beta}H_R$ proton. The distances between $C^{\alpha}H$-$C^{\beta}H_R$ protons and between $C^{a}H$-$C^{\beta}H_S$ protons with an energy minimized proline are shown 2.3 Å. and 2.8 Å respectively (Figure 2.3.5).

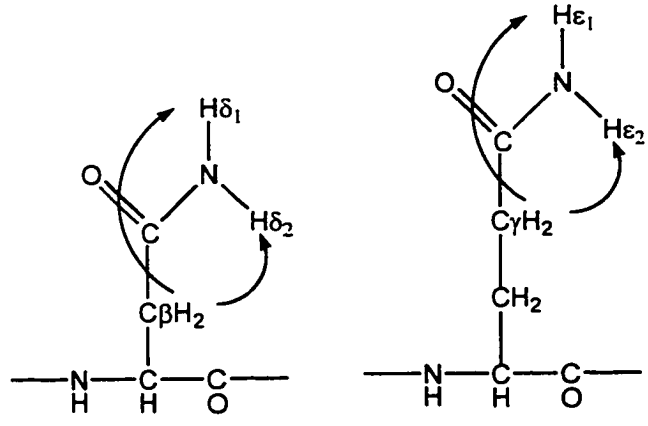
Figure 2.3.5. The energetic favorable conformation of proline (the distances of $C^{\alpha}H$-$C^{\beta}H_R$ and $C^{\alpha}H$-$C^{\beta}H_S$ of proline measured in Å from Insight II)



2.3.5. Individual assignments of sidechain amide protons.

Side-chain amide protons of asparagine and glutamine that are cis to the amide carbonyl oxygen atom are named respectively $H^{\delta 1}$ and $H^{\varepsilon 1}$, and trans protons are named $H^{\delta 2}$ and $H^{\varepsilon 2}$ consistent with IUPAC convention. Individual assignments for these side-chain amide protons were obtained by comparing the relative NOE intensities of the pairs of H^{β}-H^{δ} and H^{γ}-H^{ε}, respectively, in asparagine and glutamine side chains (Montelione et al., 1992). The inter-nuclear distances between the β protons in asparagine (or the γ protons in glutamine) and $H^{\delta 2}$ (or $H^{\varepsilon 2}$) are shorter than the corresponding distances to $H^{\delta 1}$ (or $H^{\varepsilon 1}$) (Figure 2.3.6). To minimize effects of spin diffusion, a short NOESY mixing time of τ_m=40 ms was used.

Figure 2.3.6. Criteria for individual assignment of side-chain amide protons.



2.4. Evaluation of NOE cross peaks.

Since the number of contours is increased as the intensity of the peak increased, the upper-bound distance constraints in both 90% H_2O/10% D_2O and D_2O solutions were obtained by counting the number of contours from NOESY with mixing time of 200 ms. Appropriate corrections were made for chemical shift degenerate methylene, methyl, and aromatic Hδ and Hε protons (Table 2.4.1). Non-stereo-specifically assigned methylene protons were treated as pseudo atoms and the weaker NOE cross peaks were set for the upper-bound distance constraints (Table 2.4.1) (Wüthrich, et al., 1983). All lower bound constraints were set to 1.8 astron.

Table 2.4.1. Pseudo-atom corrections.

Group	Correction (Å)
CH_2 degenerate	0.5
CH_2 non-degenerate	use the weaker upper bound
CH_3	1
$2CH_3$	2
aromatic Hδ and Hε	1
NH_2	0.5

2.5. Reiterative evaluation of NOESY data and monitoring the assignments of NOE cross peaks based on the calculated DG structures.

The possible assignment of some cross peaks could be ambiguous due to chemical shift degeneracies. With the unambiguously assigned NOEs, including medium and long range NOE's, initial peptide conformations were calculated using a distance geometry approach. These conformations were then analyzed by checking distance violations. For those with a high percentage of violated constraints, the cross-peak assignments in the NOESY spectrum were further checked and correction would be made if the assignment was wrong. The structures were used to interpret the NOESY cross peaks which has more than one possible assignments. If the interpretations showed one assignment is possible than the other(s) in the structures, this NOE constraints would be added to the constraint file for further calculation.

2.6. Structure computation.

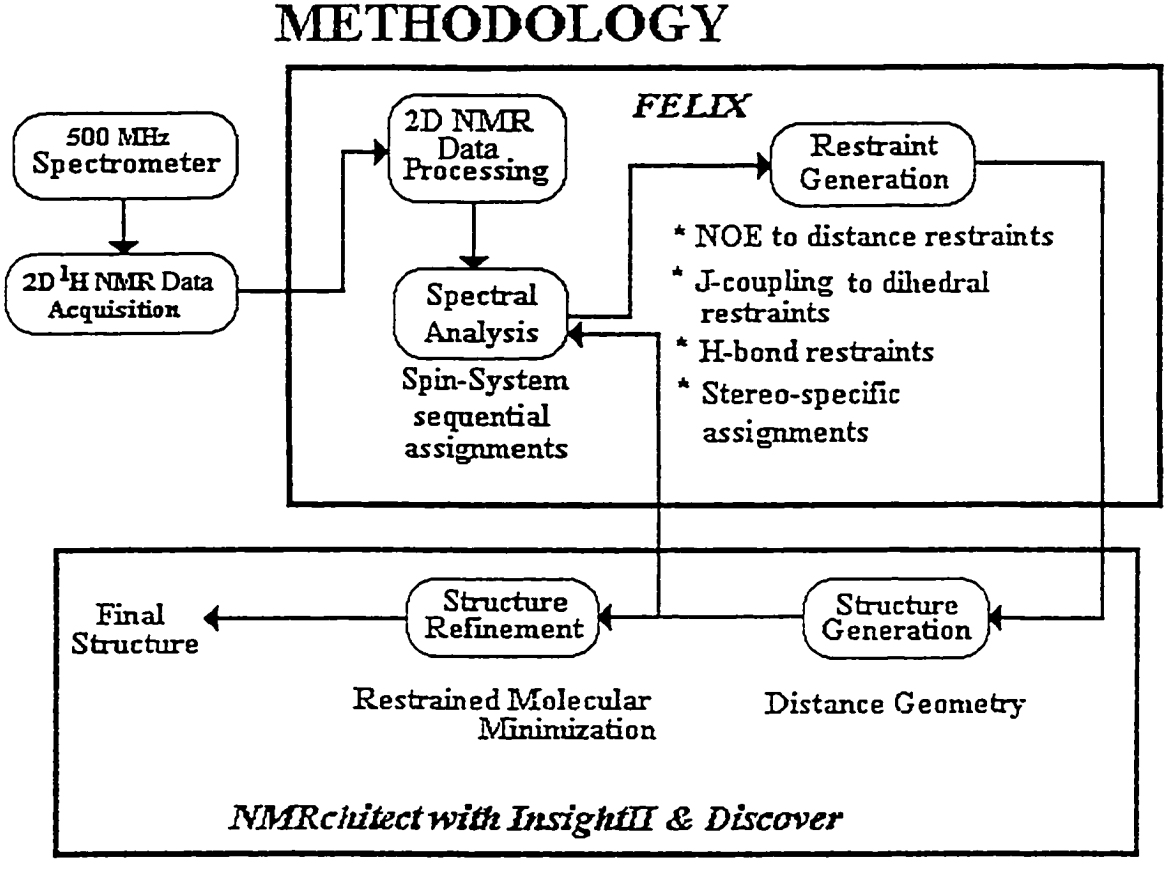
All structure calculations were carried out using the package of programs included in Insight II (BIOSYM technology, San Diego). The calculations were performed using an IBM RISC-6000.

The tertiary structure was generated by means of distance geometry methods (Havel 1991) via DGII in the NMR refine module of Insight II, which includes bound smoothing, embedding and optimization. Triangle bound smoothing, metrization, prospective embedding, majorization in 4 dimensions were used. The structures are then optimized by simulated annealing

and conjugate gradient energy minimization. In this procedure, the maximum temperature was set to 200K, the initial energy was set to 1,000 kcal and the step size was set up 10,000. In the final stage of calculations, the maximum temperature was set to 800K. The total time for each structure was 13 ns with 1.65×10^{-13} step size and 80,000 steps. For the nest step the initial energy term was set to 5,000 kcal. Conjugate gradient minimization in 3 dimensions using the CVFF parameters was performed after simulated annealing with 1,000 iterations, which would terminate if RMS gradient norm falls below 0.001.

Structures produced by distance geometry methods are required to be further energy minimization with constraints, in order to reduce repulsive nonbonded contacts. CVFF (consistent-valence forcefield) is a generalized valence forcefield (Dauber-Osguthorpe, et al., 1988) provided with the DISCOVER program (Biosym/MSI, Inc., San Diego). The final DGII structures were then minimized using 100 steps of steepest decent minimization, followed by 1000 steps of conjugate gradient minimization using CVFF with 12 astron non-bond cutoff. The non-bond and covalent bond force scaling constants were set to 1.0 and the force constants for the experimental NOE distance, the hydrogen bond and dihedral restraints were set to 100 kcal/mol/Å^2, 30 kcal/mol/Å^2, and 100 kcal/mol/rad^2 respectively. Figure 2.6.1 shows a summary of the methodology used for structure determination by this work.

Figure 2.6.1. Methodology of structure determination.



Chapter III. Results

3.1. Stereospecific Assignments.

3.1.1. Stereo-specific Assignment of C^{β} methylene protons

Of the 43 residues in the N-terminal EGF-like domain of factor IX, 34 have C^{β} methylene protons, including 2 proline residues. Among the non-proline residues, 12 have nearly degenerate (δ less than 0.1 ppm) $C^{\beta}H_2$ chemical shifts: Tyr1, Asp3, Asp5, Glu8, Ser9, Asn10, Ser17, Cys18, Lys19, Asn23, Ser24, and Glu26. Of the remaining 20 non-proline residues, stereo-specific assignments of 8 $C^{\beta}H_2$ methylene proton pairs were not obtained because of insufficient information from either COSY or NOESY spectra. Individual $C^{\beta}H_2$ methylene protons were assigned if the combined $^3J_{\alpha\beta}$ β vicinal coupling constants, H^{α}-H^{β} NOE data, and H^{N}-H^{β} NOE data were consistent with one of the three staggered conformations of χ^1 = +60°, 180°, or -60° (see Figure 1.6.4 in Chapter 1). Using these data, reliable stereospecific assignments could be made for residue Gln-6, Cys-7, Cys-12, Leu-13, Asp-21, Tyr-25, Cys-27, Cys-29, Phe-31, Phe-33, Asn-37, and Leu-40 (Table 3.1.1).

Table 3.1.1. Summary of stereo-specific $C\beta H_2$ assignments of FIX-EGF_N.

residue	chemical shifts (ppm) HN	H^α	H^β	$^3JH^\alpha$-H^β(Hz)	HN-H^β NOE[a]	H^α-H^β NOE[a]	stereo-specific assignments of χ^1(deg)	H^β
Gln-6	8.81	4.06	2.19	5	<10		-60 ± 30	β3
			1.93	12	66			β2
Cys-7	8.49	4.37	3.13	4	<10		-60 ± 30	β3
			2.89	11	168			β2
Cys-12	7.73	4.38	2.75	12	138		-60 ± 30	β2
			2.42	3	16			β3
Leu-13	8.51	4.28	1.59	13	184		-60 ± 30	β2
			1.38	3	64			β3
Asn-14[b]	8.58	3.87	1.92	4	26			
			1.23	12	<10			
Asp-21	8.33	4.97	2.89	5	62		60 ± 30	β3
			2.55	5	20			β2
Tyr-25	8.10	5.39	3.14	5	<10		60 ± 30	β2
			2.74	4	80			β3
Cys-27	9.22	5.59	2.98	12	78		-60 ± 30	β2
			2.81	4	29			β3
Trp-28[b]	9.54	4.83	3.37	8	<10			
			3.25	8	<10			
Cys-29	8.54	5.13	2.93		26	107	-60 ± 30	β3
			2.57	12	130	60		β2
Phe-31	8.35	4.22	3.22	5	201	113	180 ± 30	β2
			2.92	11	146	63		β3
Phe-33	7.77	5.52	3.07		<10	82	-60 ± 30	β3
			2.60	12	91	23		β2
Lys-36[b]	9.33	3.96	1.94	7	15			
			1.76	9	<10			
Asn-37	8.93	5.69	3.52	5	<10		-60 ± 30	β2
			2.60	12	66			β3
Glu-39[b]	10.38	4.11	2.15		41	158		
			1.95	10	74	177		
Leu-40	8.78	4.69	1.49	10	156		-60 ± 30	β2
			1.35	3	48			β3
Asp-41[b]	8.63	4.66	2.81	7	12			
			2.62	7	10			

[a]Peak intensity in a spectrum with τm = 40 ms
[b]Nonstandard or multiple rotamer states

3.1.2. Stereospecific Assignment of Valine Isopropyl Methyl Groups

There are two valine residues in FIX-EGF_N, Val2 and Val42. The chemical shifts of valine methyl groups of Val2 are degenerate (δ less than 0.1 ppm). The $^3J(H^\alpha-H^\beta)$ coupling constant of Val42 is 11 Hz. $H^N-C^\gamma H_3$ and $H^\alpha-C^\gamma H_3$ NOE crosspeaks were measured in a NOESY spectrum with a mixing time of 40ms in D_2O and H_2O solutions respectively. The intensities of $H^N-C^\gamma H_3$ crosspeaks of 7.99/0.68 and 7.99/0.53 (in ppm) are 15 and 143 in arbitrary units. In addition, the intensities of the $H^\alpha-C^\gamma H_3$ NOE crosspeaks of 4.05/0.68 and 4.05/0.53 (in D_2O solutions) are 75 and 43 in arbitrary units. The downfield and upfield methyl resonances of Val42 were assigned to $C^{\gamma 2}H_3$ and $C^{\gamma 1}H_3$ respectively. χ^1 was identified as 180 ± 30° (Table 3.1.2). (For the criteria, refer to section 1.6.2. ii, and Figure 1.6.5).

Table 3.1.2. Summary of stereo-specific $C^\gamma H_3$ assignments of valine 42 of FIX-EGF_N

	chemical shifts (ppm)								stereo-specific assignments of
residue	HN	H^α	H^γ	$^3JH^\alpha-H^\beta$(Hz)	HN-$C^\gamma H_3$	NOE[a]	$H^\alpha-C^\gamma H_3$	NOE[a]	χ^1(deg)	$H^\gamma H_3$
Val-2	8.41	4.14	0.86							
			0.90							
Val-42	7.99	4.05	0.68	11	15		75		180±30	$C^{\gamma 2}H_3$
			0.53		143		43			$C^{\gamma 1}H_3$

[a]Peak intensity in a τm = 40 ms
[b]Nonstandard or multiple rotamer states

3.1.3. Stereo-specific assignment of C^{α} methylene protons of glycine.

There are five glycine residues (residues 4, 15, 16, 32, and 35) in FIX-EGF_N. The α-methylene protons of glycine 4 are degenerate. Five structures were calculated by floating the α-methylene protons of glycines 15, 16, 32, and 35 (see section 2.3.3). The floating results for glycine were assayed as following: the downfield shifted protons were taken as proS chiralities and the upfield shift protons as proR chiralities arbitrarily. Then all the NOE distance constraints were compared to the distances observed with 95% confidence level (Weber, et al., 1988). If the distance pairs were consistent with the constraints for all of the structures, the stereo-specific assignment could be made and the initial random assignment was assumed to be correct. If the pairs were opposite with regard to the constraints for all of the structures, the stereo-specific assignment could be made and would be made the opposite to the initial assignment.

Using this method stereo-specific assignment could only be made for Gly16. Glycine 15, 32 and 35 did not show a single possible chirality. The floating results of Gly16 are listed in Table 3.1.3.

Table 3.1.3. The result of floating Gly16 α-methylene protons. The distances were measured according to the NOE distance constraints.

NOE[a]	Cal[b]	atom1	atom2	Mol#	dist[c]
C29Hα/G16Hα1	3	C29Hα	G16H αS	1	3.22
				2	3.20
				3	3.38
				4	3.06
				5	3.51
C29Hα/G16Hα2	3.5	C29Hα	G16H αR	1	2.68
				2	1.94
				3	2.49
				4	2.37
				5	2.50
G16Hα1/C12Hβ2	[d]	G16HαS	C12H βR	1	3.54
				2	2.48
				3	3.16
				4	2.54
				5	2.19
G16Hα2/C12Hβ2	3.5	G16HαR	C12H βR	1	5.18
				2	4.25
				3	4.71
				4	4.30
				5	3.93
G16Hα1/P30Hδ1	4	G16HαS	P30H δX[e]	1	4.53
				2	4.51
				3	4.52
				4	4.46
				5	4.58
G16Hα2/P30Hδ1	4.5	G16HαR	P30H δX	1	3.66
				2	3.32
				3	3.90
				4	3.74
				5	3.95
G16HN/G16H α1	2.9	G16HN	G16H αS	1	2.39
				2	2.42
				3	2.66
				4	2.41
				5	2.59
G16HN/G16H α2	3.1	G16HN	G16H αR	1	2.98
				2	3.05
				3	3.06
				4	3.04
				5	3.04
S17HN/G16H α1	3	S17HN	G16H αS	1	1.97
				2	2.00
				3	3.30
				4	1.99
				5	3.32
S17HN/G16H α2	3	S17HN	G16H αR	1	3.12
				2	3.07
				3	2.00
				4	3.07
				5	2.03

[a]NOE Cross peak
[b]Converted from the intensity of the cross peak
[c]Measured from the floating structure
[d]Threated as a pseudoatom

3.1.4. Stereospecific Assignment of β-methylene protons of proline

The absolute configurations of the β-methylene protons of Pro-11 and Pro-30 were determined by the relative NOE intensities of H^{α}-H^{β} from NOESY with mixing time of 40 ms. The relative NOE cross peak intensities of 4.24 ppm/1.69 ppm (Pro-11 $C^{\alpha}HC^{\beta}H_1$) and 4.24 ppm/1.58 ppm (Pro-11 $C^{\alpha}HC^{\beta}H_2$) are 108 and 58 respectively in arbitrary units. The relative NOE cross peak intensities of 4.47 ppm/2.49 ppm (Pro-30 $C^{\alpha}HC^{\beta}H_1$) and 4.47 ppm/2.15 ppm (Pro-30 $C^{\alpha}HC^{\beta}H_2$) are 131 and 54 respectively in arbitrary units. In both prolines, the downfield β protons were assigned to $C^{\beta}H_S$ and the upfield β protons to $C^{\beta}H_R$ (Table 3.1.4. Summary of stereo-specific $C^{\beta}H_2$ assignments of proline residues of FIX-EGF_N).

Table 3.1.4. Summary of stereo-specific $C^{\beta}H_2$ assignments of proline residues of FIX-EGF_N.

	chemical shifts (ppm)			stereo-specific assignments of
residue	H^{α}	H^{β}	H^{α}-H^{β} NOE[a]	H^{β}
Pro-11	4.24	1.69	108	$C\beta H_S$
		1.58	58	$C\beta H_R$
Pro-30	4.47	2.49	131	$C\beta H_S$
	2.15		54	$C\beta H_R$

[a]Peak intensity in a τm = 40 ms except where indicated

3.1.5. Individual Assignments of Side-Chain Amide Protons

There are four asparagines and one glutamine in this peptide. With an exception of Asn-14, the individual side-chain amide protons of these residues have distinct chemical shifts because of the double bond characteristics of amide-bond (Table 3.1.5). The individual assignments of side-chain amide protons of Gln6, Asn10, Asn23, and Asn37 were made based on the relative NOE intensities of H^{β}-H^{δ} or H^{γ}-H^{ε} (Table 3.1.5). (For the criteria, refer to section 2.3.5 and Figure 2.3.6). In all of these four side chains, the upfield and downfield amide protons are assigned to $H^{\delta 1}$ (or H^{q}) and $H^{\delta 2}$ (or $H^{\varepsilon 2}$), respectively. Those of Asn14 could not be made because the NOE intensities were too similar to determine the configuration (Table 3.1.5).

Table 3.1.5. Individual Side-Chain NH Assignments for FIX-EGF_N.

residue	chemical shifts (ppm) H^β	H^γ	H^δ	H^ε	H^β-H^δ NOE[a]	H^γ-H^ε NOE[a]	Stereospecific assignments of NH
Gln-6		2.40		7.49		309	ε2
		2.24				<50	ε2
		2.40		6.69		<50	
		2.24				<50	
Asn-10	2.85[b]		7.44		654		δ2
			6.65		<50		δ1
Asn-14[c]	1.92		7.22		132		
	1.23				100		
	1.92				114		
	1.23				61		
Asn-23	2.91		7.66		718		δ2
	2.83				155		δ1
	2.91		6.93		<50		
	2.83				<50		
Asn-37	3.52		7.36		185		δ2
	2.44				424		δ1
	3.52		6.91		<50		
	2.44				<50		

[a]Peak intensity in a τm = 40 ms spectrum
[b]Chemical shift degenerate
[c]NOE intensities were too similar to determine the configuration

3.2. Description of the Structural Constraints.

There are four kinds of constraints which were used in the input for the structure determination: (I) chiral constraints based on the configuration of chiral carbon Cα of each amino acid and Cβ of isoleucine and threonine; (ii) hydrogen bond constraints; (iii) distance constraints derived from NOE cross peaks; (iv) dihedral constraints including backbone ϕ dihedral angle constraints derived from $^3J(H^N-H^\alpha)$ coupling constants, side-chain χ^1 dihedral angle constraints based on the assignments of the three classical side-chain rotamers by NMR data (see Figure 1.6.4), and backbone ω dihedral angle.

Nearly all the cross peaks of H_2O and D_2O NOESY spectra were assigned. A total of 440 NOE upper constraints derived from NOESY spectra, which include 209 intra- and 231 inter-residue NOE constraints. Among the inter-residue NOE restraints, there are 124 of sequential NOE, 29 medium range NOE constraints including 11 of i, i+2, 7 of i, i+3, and 11 of i, i+4; and 78 long range NOE constraints (i, i+>4) (Table 3.2.1). Table 3.2.1 shows the distribution of the distance constraints in FIX-EGF_N.

Figure 3.2.1. The distribution of the distance constraints in FIX-EGF$_N$.

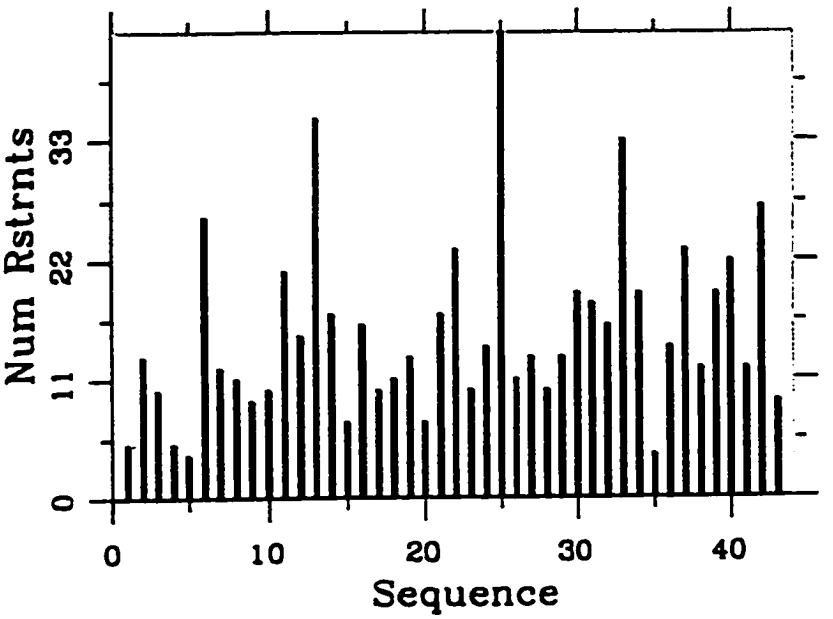


Table 3.2.1. Summary of upper bound NOE constraints

Resid.	Total	Intra	Inter	i+1	i+2	i+3	i+4	Med.	Long
Tyr1	5	2	3	3	0	0	0	0	0
Val2	12	4	8	8	0	0	0	0	0
Asp3	10	3	7	7	0	0	0	0	0
Gly4	4	1	3	3	0	0	0	0	0
Asp5	4	1	3	3	0	0	0	0	0
Gln6	26	9	17	4	0	0	0	0	13
Cys7	12	5	7	4	0	0	2	2	1
Glu8	11	5	6	5	1	0	0	1	0
Ser9	9	3	6	5	1	0	0	1	0
Asn10	10	4	6	5	1	0	0	1	0
Pro11	27	8	19	7	1	0	2	3	9
Cys12	17	5	12	7	0	0	2	2	3
Leu13	37	12	25	12	2	1	0	3	10
Asn14	19	4	15	12	1	0	0	1	2
Gly15	8	1	7	5	2	0	0	2	0
Gly16	15	2	13	5	1	1	2	4	4
Ser17	10	1	9	5	0	0	0	0	4
Cys18	11	2	9	5	0	0	0	0	4
Lys19	13	8	5	4	0	0	0	0	1
Asp20	7	3	4	2	0	0	0	0	2
Asp21	17	5	12	6	0	1	5	6	0
Ile22	23	10	13	10	3	0	0	3	0
Asn23	10	3	7	7	0	0	0	0	0
Ser24	14	2	12	7	0	0	0	4	1
Tyr25	50	11	39	10	0	0	5	5	24
Glu26	11	4	7	7	0	0	0	0	0
Cys27	13	5	8	3	0	0	0	0	5
Trp28	10	4	6	2	0	0	0	0	4
Cys29	13	5	8	5	0	0	1	1	2
Pro30	19	5	14	7	0	5	0	5	2
Phe31	18	11	7	6	1	0	0	1	0
Gly32	12	1	11	6	0	0	0	0	5
Phe33	32	10	22	7	1	5	1	7	8
Glu34	19	2	17	6	0	0	1	1	10
Gly35	4	1	3	3	0	0	0	0	0
Lys36	15	9	6	4	0	0	0	0	2
Asn37	25	8	17	5	2	0	0	2	10
Cys38	12	3	9	6	0	0	1	1	2
Glu39	19	4	15	8	2	0	0	2	5
Leu40	22	8	14	8	0	0	0	0	6
Asp41	12	3	9	5	0	0	0	0	4
Val42	26	9	17	5	0	0	0	0	12
Thr43	8	3	5	4	0	0	0	0	1

$^3J(H^N-H^\alpha)$ coupling constants were measured from DQF-COSY spectrum at pH 4.2 and 25 °C with resolution of 2.93 Hz/point (Figure 3.2.2). Thirty five $^3J_{N\alpha}$ coupling constants were collected from COSY spectrum in H_2O solution including three glycine residues (Table 3.2.2). Seventeen $^3J(H^N-H^\alpha)$ coupling constants > 8.0 Hz were measured, including those of residues Val-2, Ser-9, Cys-12, Cys-18, Lys-19, Asp-21, Tyr-25, Glu-26, Cys-27, Cys-29, Phe-33, Asn-37, Cys-38, Glu-39, Leu-40, Val-42 and Thr-43. The backbone torsion angels ϕ $(C_{i-1}-N-C\alpha-C)_i$ of these residues were constrained to the range $-160° < \phi < -80°$ (Pardi et al., 1984).

Dihedral angle constraints from the side-chain dihedral angle χ^1 $(N_i-C\alpha_i-C\beta i-C\gamma_i)$ were derived from the stereo-specific assignments of methylene protons, by adding ± 30° at the corresponding angles of 60°, 180° and -60° for the conformations of g^2g^3, g^2t^3 and t^2g^3 respectively (see Table 3.1.1).

Figure 3.2.2. A portion of the DQF-COSY contour plot in 90% H_2O/10% D_2O

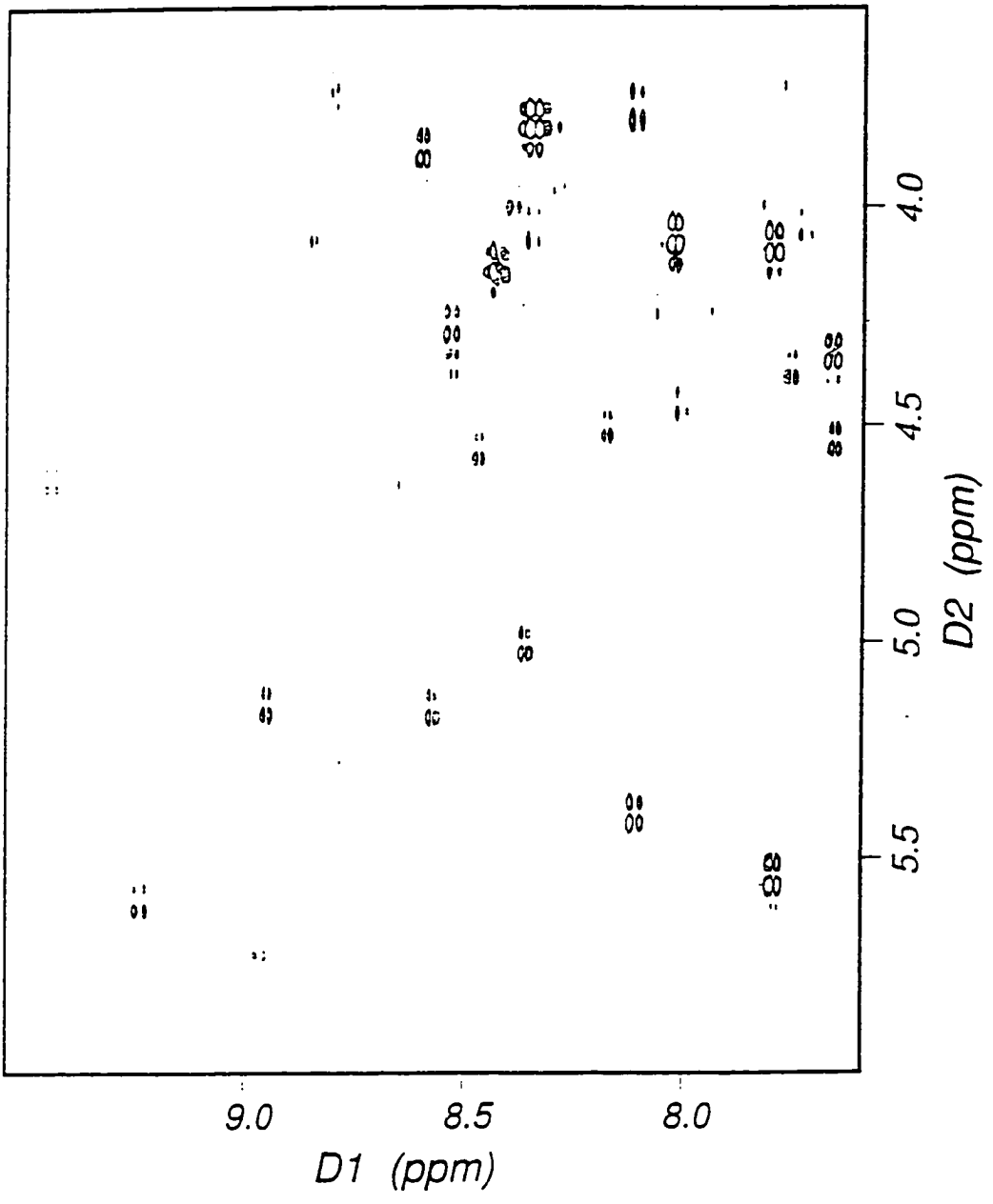


Table 3.2.2. The $^3J(H^N-H^\alpha)$ coupling constants measured from DQF-COSY spectrum at pH 4.2 and 25 °C

residue	$^3JH^N-H^\alpha$ (Hz)	Residue	$^3JH^N-H^\alpha$ (Hz)
Val-2	8.2	Ser-24	6.9
Asp-3	5.6	Tyr-25	8.8
Asp-5	7.6	Glu-26	10.0
Gln-6	6.4	Cys-27	9.8
Cys-7	7.2	Trp-28[a]	
Glu-8	6.3	Cys-29	8.2
Ser-9	8.8	Phe-31	7.8
Asn-10	7.4	Phe-33	11.8
Cys-12	8.2	Glu-34[a]	
Leu-13	7.9	Lys-36	5.7
Asn-14	7.4	Asn-37	9.1
Ser-17[a]		Cys-38	8.5
Cys-18	8.0	Glu-39	9.8
Lys-19	10.6	Asp-40[b]	9.7
Asp-20[b]	9.7	Asp-41	7.3
Asp-21	9.4	Val-42	8.7
Ile-22	5.9	Thr-43	8.8
Asn-23[a]			

[a]Chemical shift falls within H_2O peak

[b]The assignment can be exchanged or it could be both. Because of this uncertainty, it is not used as a ϕ dihedral angle constraint for both residues.

The β carbon of isoleucine is a chiral center. Among the three energetically favorable conformations around Cα-Cβ bond (*g*-, *t*, and *g*+) (Figure 3.2.3), the *g*+ conformation could be easily identified by large value for $^3J_{\alpha\beta}$ coupling constant (10-12 Hz) corresponding to *trans* orientation for α and β protons (Zuiderweg, et al., 1985). Both of the *g*- and *t* conformations will have a relative small value (2-5 Hz) for $^3J_{\alpha\beta}$ coupling constant corresponding to *cis* orientations. These two rotamers could be distinguished by the relative NOE intensities of HN-CγH_2 and HN-CγH_3 (Figure 3.2.3). There is one isoleucine in this peptide (Ile22). The $^3J_{\alpha\beta}$ coupling constant of Ile22 was measured from DQF-COSY as approximately 10 Hz. It indicated that the *g*+ conformation was favorable. The side-chain dihedral angle was constrained for Ile22 -60 ± 30°.

The total number of dihedral constraints was 31, including ϕ (17), χ^1 (14) dihedral constraints.

To detect slowly exchanging amide protons, the peptide sample in H_2O was lyophilized and dissolved in D_2O, and a COSY spectrum was immediately acquired (Huang, et al., 1991). Slow exchanging amide protons were detected for residues 13, 16, 19, 20, 26, 28, 33, 38 and 39. Slow amide proton exchange was taken as an indication for the existence of an intra-molecular hydrogen bond. To further identify Hydrogen bonds, the established criteria was used to characterize the Hydrogen bonds in regular secondary structures (Wüthrich et al., 1984; Wüthrich, k.

1986). In antiparrallel β-sheets, for identification of a hydrogen bond between the amide proton of residue *1* on one strand and the carbonyl oxygen of residue *i* on the other strand, the following NOEs would be observed: NH(*1*)-NH(*i*), NH(*1*)-H^{α}(*i*+1), and H^{α}(*1*-1)-H^{α}(*i*+1). Two antiparrallel β-sheets had been predicted (Huang et al., 1991) from typical NOE patterns (Wüthrich, k. 1986) and shown in the calculated tertiary structures. Hydrogen bonds Glu26 NH...OC Lys19, Lys19 NH...OC Glu26, SER17 HN...OC Trp28, Trp28 HN...OC Ser17, and Leu40 HN...OC Glu34 were predicted as interstrand hydrogen bonds, with the following NOE interactions observed: HN_{K19}-NH_{E26}, NH_{K19}-αH_{C27}, αH_{K19}-αH_{C27}, αH_{Y25}-αH_{D20}, HN_{S17}-NH_{W28}, αH_{G16}-αH_{C29}, NH_{W28}-αH_{C28}, αH_{C27}-αH_{C18}, HN_{L40}-NH_{E34}, αH_{D41}-αH_{G33}. For Hydrogenbonds SER17 HN...OC TRP28 and Glu26 NH...OC Lys19, the NH(1)-H^{α}(i+1) NOEs were not observed. The amide protons of Ser17 and Leu40 were not detected as a slowly exchanging proton in our COSY experiment, however, they were observed by Baron's group, which data was detected in a HOHAHA spectrum (Baron, et al., 1992). In HOHAHA (TOCSY), the coupled signals are in phase while in cosy they are anti phase. If the signal is broad (as in the case of exchangeable protons), the two anti-phase components are invisible due to negligible intensity. The two in-phase components in HOHAHA adds up and make it visible. For this reason, the amide protons of Ser17 and Leu40 were treated as slowly exchange protons.

For a tight turns, it would be stabilized by a backbone hydrogen bond between the CO of residue n and the NH of residue n + 3 (Venkatachalam, C. M. 1968. Hydrogen bonds associated with turns were identified based on the calculated tertiary

structures and the observed typical NOE interactions (Wüthrich, k. 1986). Two Hydrogen bonds involved the tight turns were found. They are Gly16 HN...OC Leu13 and Phe33 HN...OC Pro30 (Table 3.2.3 and Figure 3.2.4).

Figure 3.2.3. Three energetically favorable conformations around Cα-Cβ bond of an isoleucine.

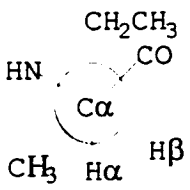
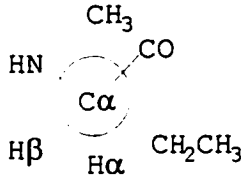
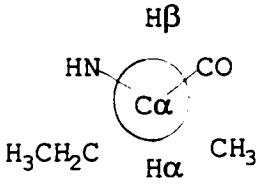
Conformation	*g*-	*t*	*g*+
χ^1	60°	180°	-60°
			
$^3J\alpha\beta$ (Hz)	2.6-5.1	2.6-5.1	10-12
Observed			10
NOE (HN,CH_3)	S	S	W
Observed			W
NOE (HN, $C\gamma H_2$)	S	W	S-M
Observed			S

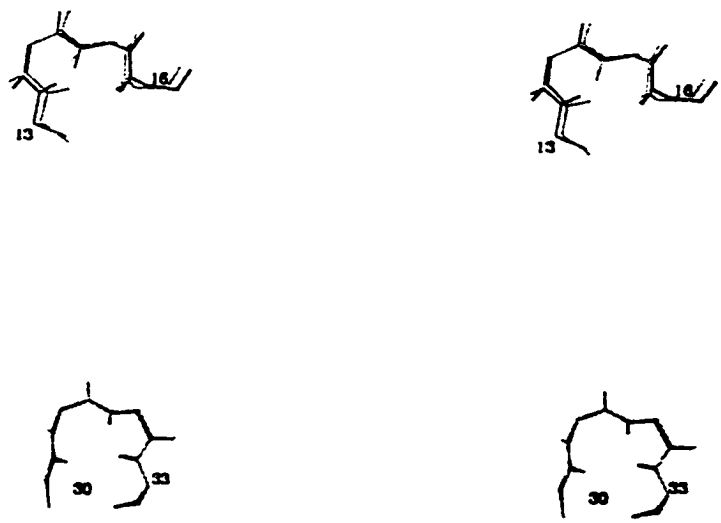
Table 3.2.3. Summery of identification of tight turns

Distance	Turn I	Turn II	L13-G16	P30-F33
$d\alpha_N(2,3)$	3.4	2.2	M	M
$d\alpha_N(3,4)$	3.2	3.2	M	M
$d\alpha_N(2,4)$	3.6	3.3	a	M
$d\alpha_N(1,4)$	3.1-4.2	3.8-4.7	a	a
$d_{NN}(2,3)$	2.6	4.5	S	b
$d_{NN}(3,4)$	2.4	2.4	S	S
$d_{NN}(2,4)$	3.8	4.3	VW	M

a. not observed

b. too close to diagonal peak to be calibrated.

Figure 3.2.4. Stereo views of the two identified tight turns.



Hydrogen bonds Glu39 HN...OC Glu34 and Val42 HN...OC Gly32, were identified from the calculated structures. The first one occurred in a five amino acids turn. The last happened in the C-terminal antiparallel β-sheet. The geometrical criteria for the hydrogen bonding employed required the hydrogen-acceptor distance to be less than 3.0 Å and the angle formed by the donor-hydrogen-acceptor to be greater than 120° with at least 50% (Üllner, et al., 1992). The hydrogen-acceptor distances and the angles of the donor-hydrogen-acceptor of both Hydrogen bonds are anticipated by this criteria (Table 3.2.4). Both Glu39 HN...OC Glu34 and Val42 HN...OC Gly32, were used as constraints although the amide proton of Val42 is not detected in the slow exchange experiment.

Each Hydrogenbond was treated as a pair of distance constraints with 2.7-3.3 Å for N_i-O_j and 1.7-2.3 Å for HN_i-O_j. (Kohda, D., and Inagaki, F. 1992a). There are 18 generic distance constraints arising from 9 hydrogen bonds.

Table 3.2.4. Hydrogen bond identifications for Glu39 NH...OC Glu34 and Val42 NH...OC Gly32.

Donor	Acceptor	Conformer	$d_{H,O}$(Å)	Angle(N-H,O)
Glu39	Glu34	1	2.57	136
		2	2.60	142
		3	2.74	138
		4	2.58	136
		5	2.63	142
		6	2.55	137
		7	2.73	139
		8	2.61	141
		9	2.59	142
		10	2.74	140
		11	2.52	140
		12	2.56	138
		13	2.40	135
		14	2.54	138
Val42	Gly32	1	2.30	154
		2	2.29	153
		3	2.29	153
		4	2.29	153
		5	2.29	153
		6	2.29	154
		7	2.29	153
		8	2.29	153
		9	2.29	152
		10	2.29	153
		11	2.29	153
		12	2.29	153
		13	2.30	154
		14	2.30	154

Table 3.2.5 gives the summary of the experimental input constraints used in determining the solution structure of FIX-EGF_N.

Table 3.2.5 The summary of the experimental input constraints.

Constraint	Total
upper-bound	440
Hydrogen bond	18
torsion angle (ϕ)	17
side-chain dihedral angle (χ^1)	14

3.3. The solution structure determination

The structures were generated using distance geometry method employing DGII package in the Insight II NMR-Refine module.

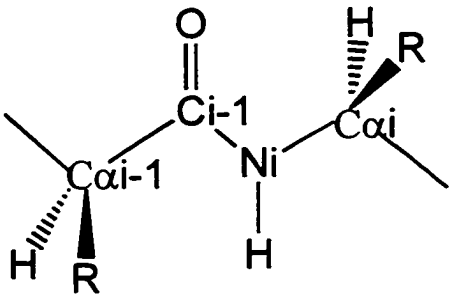
An initial structure without constraints was generated using Biopolymer module of Insight II with the consensus disulfide bridge pattern of EGF. To ensure that the starting geometry is regular, the initial structure was minimized with Discover without constraints before starting DGII. In the resulting structure several peptide bonds were found to

seriously deviate from planar. These bonds were constrained to trans and the initial structure then further minimized.

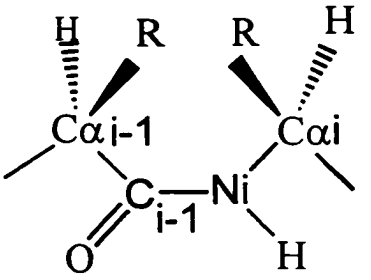
After each constrained structure calculation, NOE distance violations > 0.3 Å were listed and the assignments corresponding to the violated constraints were rechecked. This process was repeated in an interactive manner until there were no more detectable assignment errors.

Since the ω dihedral angle (C^{α}_{i-1}-C_{i-1}-N_i-C^{α}_{i}) is very close to 180°, producing a trans, planar peptide bond (Figure 3.3.1) except proline, with ω = 0°, which can occur perhaps 25% of the time (Richardson, J. S. 1981). In EGF-FIX_N, both of the peptide bonds involving the two proline residues, Pro11 and Pro30, were identified the trans conformations in our early study (Huang, L. H. 1995) by showing the strong NOEs between α proton of Asn10 and δ protons α of Pro11 and between α proton of Cys20 and δ protons of Pro30 (Wüthrich, K. 1986). All of the ω dihedral angle were constrained in the planar trans conformation (ω = 180 ± 5^0) with the force constants of 100 kcal/mol/rad^2 including proline 11 and 30.

Figure 3.3.1 Trans and cis conformations of peptide bonds. (After Voet, D. and Voet, J. G., Biochemistry, 2nd edition, John Wiley & Sons, Inc., New York, Chichester Brisbane, Toronto, Singapore)



Trans-Peptide Group



Cis-Peptide Group

14 DGII structures were generated in the final calculation. These DGII structures were then minimized using a CVFF force field. For the conditions for both DGII and minimization calculations, see section 2.6.

3.4. The quality of the tertiary structure of FIX-EGF_N

There were no NOE violations > 0.5 Å out of 440 NOE-derived upper-bound constraints. The number of violations > 0.3 Å varied between 1 to 4 among the 14 structures. There were 23 to 32 violations > 0.1 Å among the 14 individual structures. No lower bound distance constraint violations and no hydrogen bond constraint violations were larger than 0.1 Å. There is no dihedral violation > 5 degree out of 31 ϕ and χ^1 dihedral constraints. Figures 3.4.1 and 3.4.2 show summary of the distance violations and dihedral angle violations from conformer 1. The violations shown in these figures are not individual violation, but the sum of all violations found for each residual.

Figure 3.4.1 Summary of the distance violations.

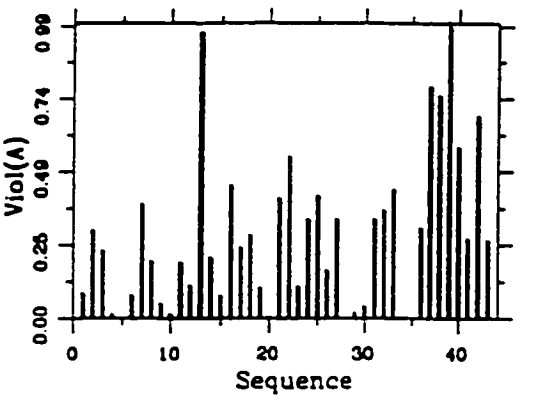
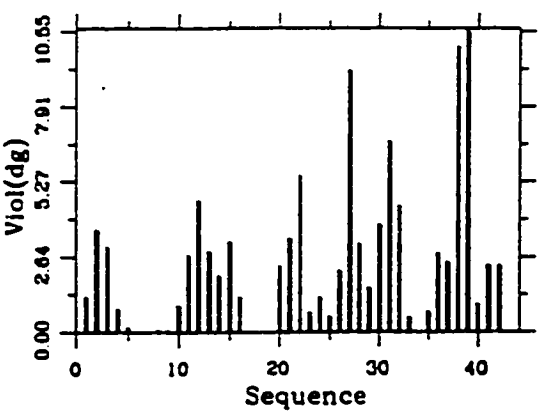


Figure 3.4.2 Summary of dihedral angle violations.



There are only intra and sequential NOE restraints for residues 1-5 (see Table 3.2.1. in section 3.2), and these residues appear to be significantly more disordered than the rest of the structure. Therefore backbone and all atom superimpositions used only residues 6 to 43, and the RMSDs (root mean square deviations) were calculated based on residues 6 to 43 as well. The average RMSD against the mean conformation from the 14 minimized structures for the backbone atoms (N, Cα, C, O) is 0.66 ± 0.10 Å and 1.22 ± 0.097 Å for all atoms.

RMSD of each residue of every conformer against the average structure obtained from NMR-refine model of Insight II. Figure 3.4.3. and Appendix 1 and 2 show the RMSD of each residue of individual conformation on both backbones and all atoms of the average conformation of all the 14 structures. Best fit super-positions of residues 6-43 of the final structures on the backbone of the mean coordinates are shown in Figure 3.4.4. It shows that the positions of the N-terminal residues 1-5 are clearly defined poorly. For the rest of the residues, the positions of residues 6-11, 20-24, and 35-38 are relatively less well defined compared with other residues. The average RMSD superimposed on the backbone atoms of residue 12-19, 25-34, 39-43 of the mean conformer is 0.44 ± 0.12 Å. The position of the C-terminal residuals is better defined than that of the N-terminal. For the polypeptide segment P30-T43, the RMSD of the backbone is significantly lower than the corresponding values for the entire segment 6-43, while the RMSD of Gln6-Pro30 is close to that of 6-43 (Table 3.4.1 and Figure 3.4.5).

Figure 3.4.3. RMSD of each residue for 14 conformers superimposed with the average structure. Part A shows the results based on the backbones. Part B shows the results based on the all atoms.

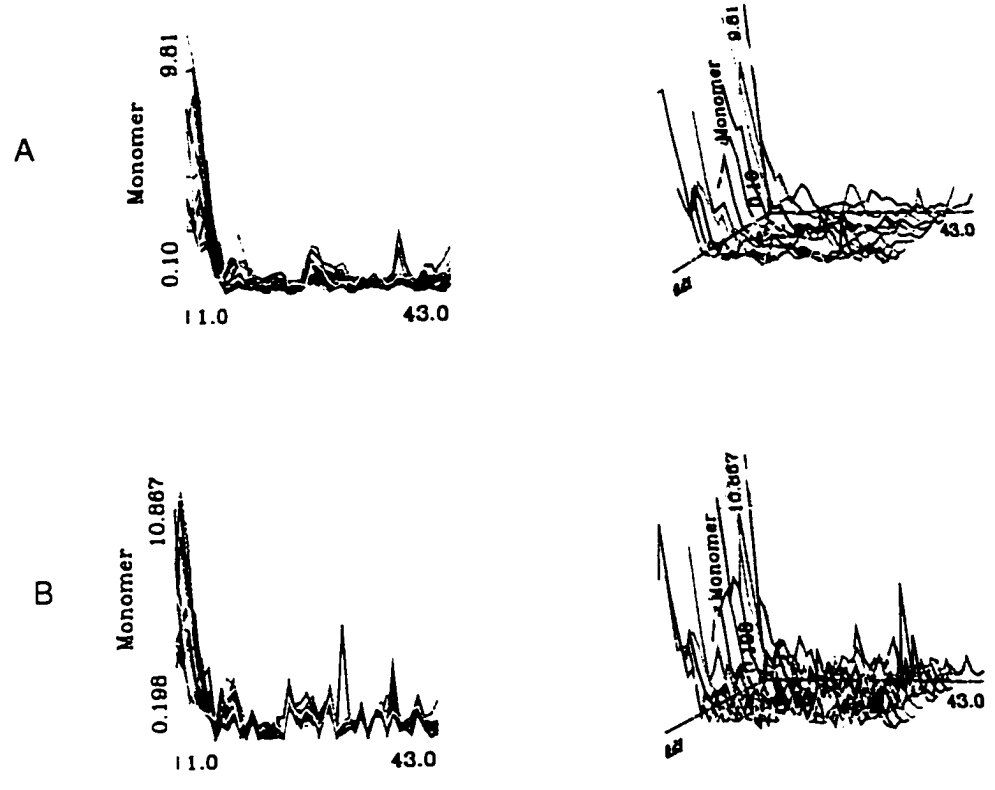


Figure 3.4.4. A stereo view of FIX-EGF_N backbones of 14 conformers. Each conformer was superimposed on the average structure backbones of residue 6-43.

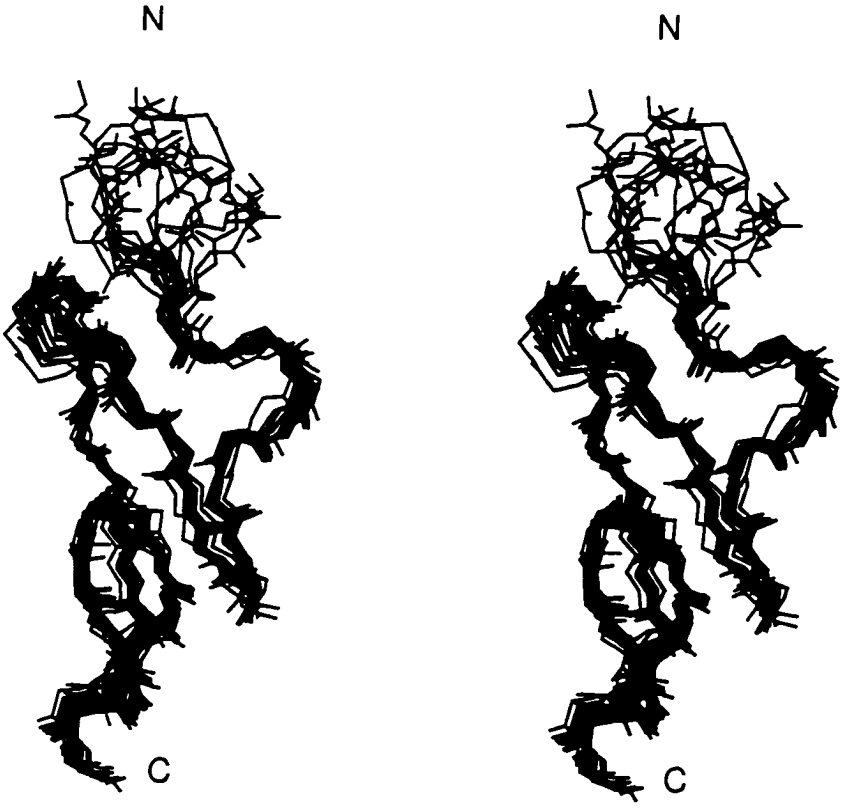


Figure 3.4.5. Stereo diagrams showing super-impositions for all heavy atoms of the polypeptide segments of 6-30 and 29-43.

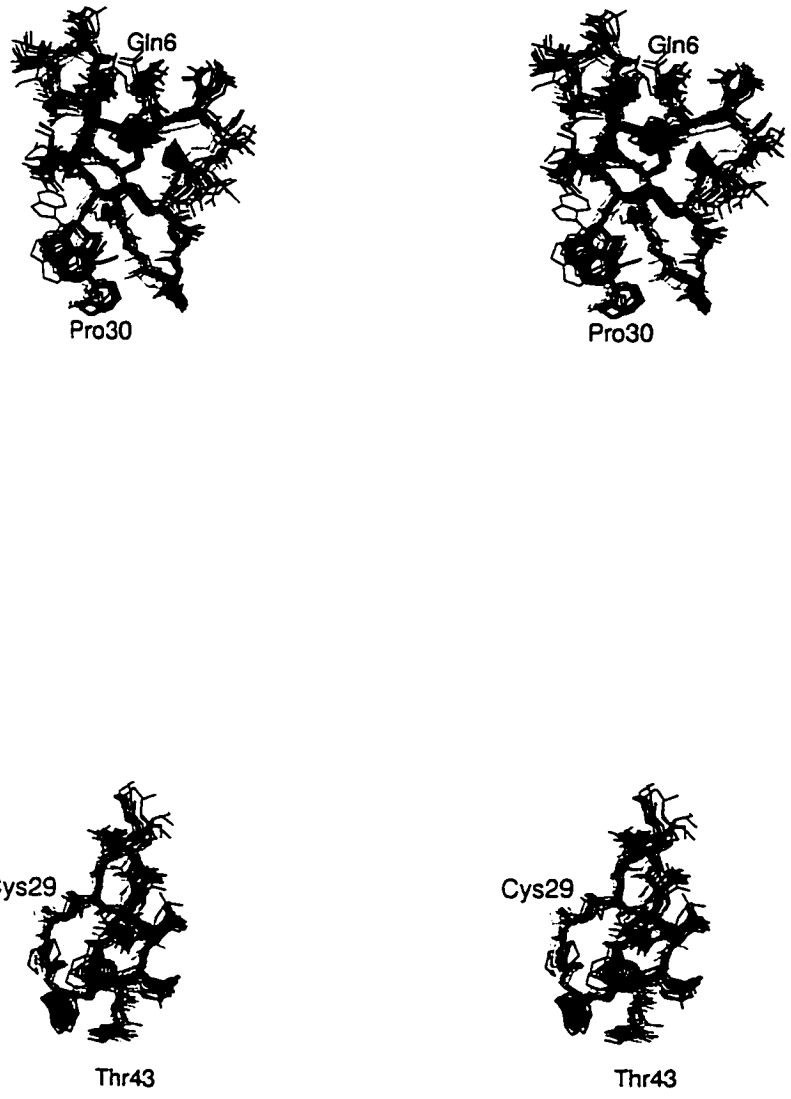


Table 3.4.1 Comparison RMSD values of backbone for polypeptide segments of FIX-EGF_N.

Residues	RMSD (N, Cα, C, O)
6-43	0.66 ± 0.10
12-19, 25-34, 39-43	0.44 ± 0.12
6-13	0.53 ± 0.080
6-30	0.61 ± 0.091
30-43	0.35 ± 0.086
16-20, 25-29	0.33 ± 0.098
32-34, 40-42	0.23 ± 0.1

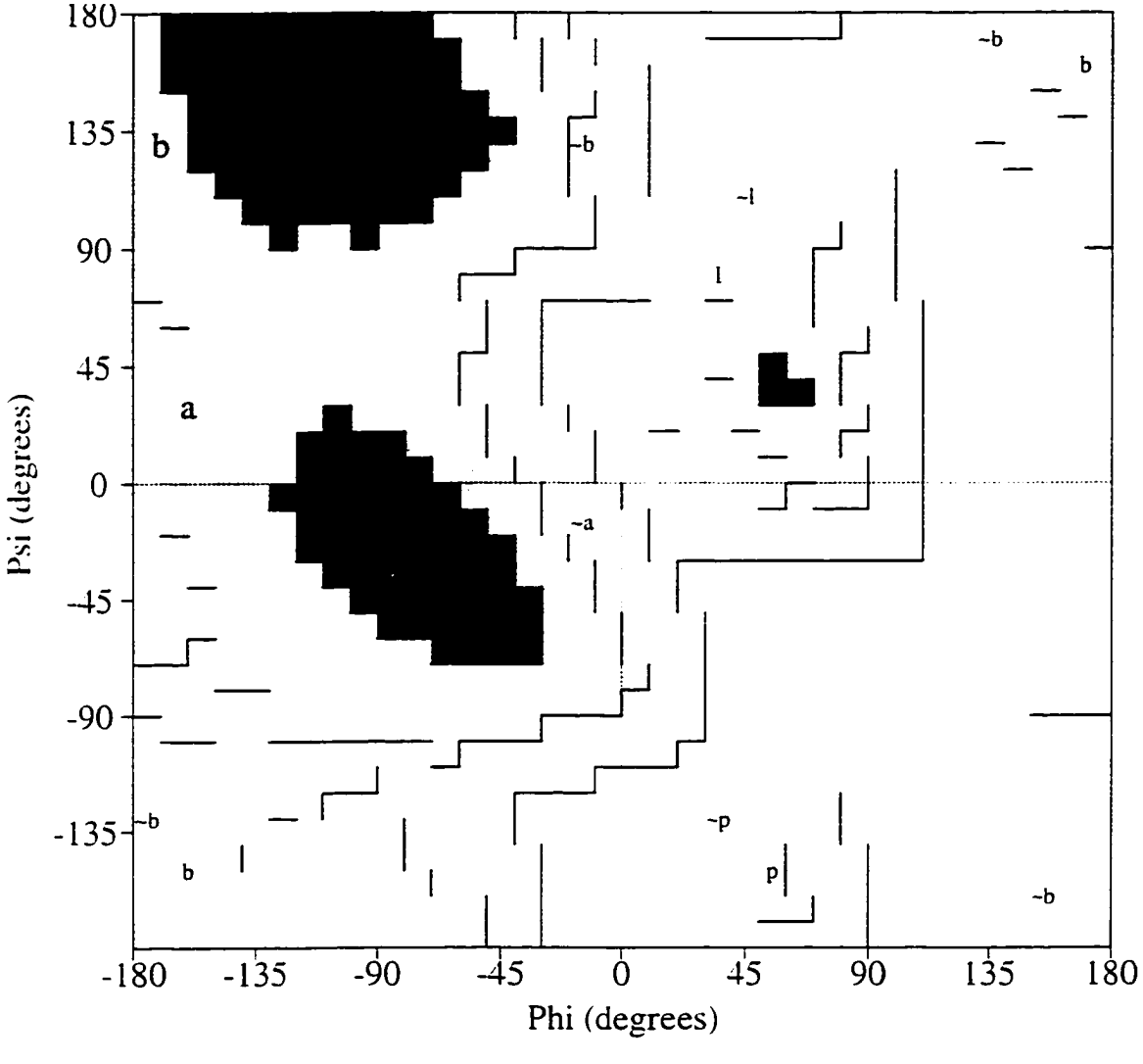
It has been discussed in section 1.4 that ϕ-ψ plots provide a useful tool for evaluating the quality of a protein structure determination, since amino acid residues in proteins generally adopt backbone conformations corresponding to low-energy minima of amino acid residues in small peptides (Roterman et al., 1989). ϕ and ψ angles of the fourteen conformations of FIX-EGF_N are plotted (Figure 3.4.6) and analyzed using the program PROCHECK (Laskowski et al., 1993). The energy dependence on ϕ and ψ angles of polypeptides in this program is based on an analysis of 118 polypeptides structures at resolution $\leq$ 2.0 Å and downloaded at "http://biotech.pdb.bnl.gov:8400/". The background of the energetic favorable distributions of ϕ and ψ angles of non-glycine and non-proline residues is shown in Figure 3.4.6. Figure 3.4.7 shows the results of the analysis for FIX-EGF_N by PROCHECK program. Among the non-glycine and non-proline residues of, 41% residues lie within the most favored regions of

a Ramachandran phi, psi plot, 52% lie within the additionally allowed regions, 6% lie in the generally allowed regions, and 1% is found in disallowed region (Laskowski et al., 1993).

The ϕ and ψ angles of each residue from the fourteen structures are measured from Insight II. There are 2 to 3 Gly residues in each conformer and 2 to 4 non-Gly residues with positive phi angles before and after energy minimization. A summary of the frequency of positive ϕ values in these conformers is listed in Table 3.4.2. The ϕ-ψ plot for residues 6-42 of all conformers is shown in Figure 3.4.7. In all conformers, Asn10, Gly32, and Gly35 have positive ϕ values. Furthermore, in 12 of the refined structures, Asn14 and Gly15 adopt positive ϕ values. For the remaining residues listed in Table 3.4.2, there remains some uncertainty regarding the backbone ϕ conformation. This is largely due to the scarcity of local conformational constraints (Montelione et al, 1992). Table 3.4.3 lists the and dihedral angles of the average structure of FIX-EGF_N.

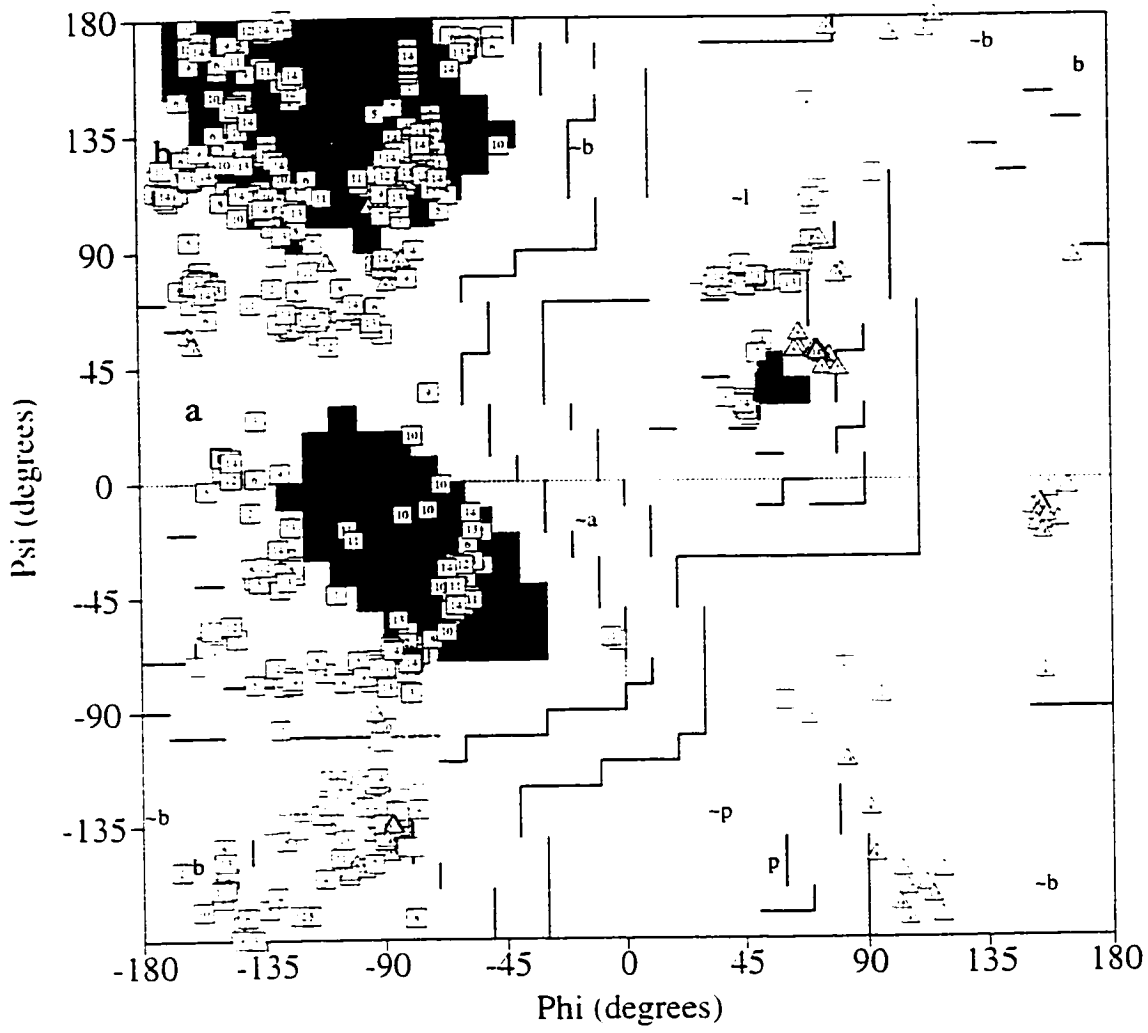
Peptide bonds are planar because of the double bond character due to its π molecular orbital system. Trans peptide bonds are preferred, with the dihedral angle $\omega = 180°$, but there is evidence that ω does vary slightly from 180° in real structures. Deviations from 180° as large as 10° have been observed (Ramachandran, G. N. and Sasisekharan, V, 1968). Cis-peptide bonds for any residues except proline are very rare. In FIX-EGF_N, most of residues show standard peptide bonds (deviate by less than 10 degrees for 180°), with the exception of residue Glu39, which exhibited ω dihedral angles in all conformers of about 165.

Figure 3.4.6. The energetic favorable background of ϕ–ψ map from PROCHECK program.



Residues in most favoured regions [A.B.L]
Residues in additional allowed regions [a.b.l.p]
Residues in generously allowed regions [~a.~b.~l.~p]
Residues in disallowed regions

Figure 3.4.7. Ramachandran plots of the 14 energy refined conformers. (Based on an analysis of 118 structures of resolution at 2.0 Å or better. Glycine residues are shown with triangles symbols.)



Plot statistics

Residues in most favoured regions [A.B.L]	194	40.8%
Residues in additional allowed regions [a.b.l.p]	246	51.7%
Residues in generously allowed regions [~a.~b.~l.~p]	30	6.3%
Residues in disallowed regions	6	1.3%
Number of non-glycine and non-proline residues	476	100.0%
Number of end-residues (excl. Gly and Pro)	28	
Number of glycine residues (shown as triangles)	70	
Number of proline residues	28	
Total number of residues	602	

Table 3.4.2. Summary of positive ϕ values among the 14 conformers of FIX-EGF_N.

	# of conformations for which > 0	
	before	after
residue	energy refinement	energy refinement
Cys-7	2	2
Asn-10	14	14
Asn-14	12	12
Gly-15	14	12
Gly-32	14	14
Gly-35	14	14
Lys-36	2	2
Asp-41	1	1

Table 3.4.3. ϕ and ψ dihedral angles of the average conformer.

Residue	ϕ	ψ
Tyr1		135.8
Val2	-135.3	105.0
Asp3	148.4	173.4
Gly4	-150.3	69.7
Asp5	-90.2	120.5
Gln6	-104.5	36.8
Cys7	-133.0	68.5
Glu8	-101.2	-67.1
Ser9	-112.7	57.7
Asn10	44.2	78.8
Pro11	-58.9	-25.8
Cys12	-84.4	168.5
Leu13	-134.2	-178.0
Asn14	39.7	20.6
Gly15	84.6	50.0
Gly16	-91.0	-145.7
Ser17	-164.3	115.2
Cys18	-81.8	128.5
Lys19	-126.7	117.5
Asp20	-70.6	119.4
Asp21	-90.0	-78.3
Ile22	-110.0	-81.3
Asn23	-124.3	-30.4
Ser24	-91.4	-135.7
Tye25	-158.0	168.0
Glu26	-131.6	104.8
Cys27	-84.9	117.1
Trp28	-72.8	133.1
Cys29	-126.1	155.9
Pro30	-57.9	168.4
Phe31	-88.9	84.7
Gly32	155.8	-14.2
Phe33	-143.2	147.0
Glu34	-145.3	-165.0
Gly35	-104.0	174.6
Lys36	-52.1	-50.8
Asn37	-94.5	-121.8
Cys38	-151.2	76.3
Glu39	-143.5	2.7
Leu40	-108.7	-123.9
Asp41	-158.6	124.1
Val42	-80.0	-70.6
Thr43	-125.3	

Chapter IX. Discussion

4.1. Description of the solution structure of FIX-EGF_N.

In order to compare with the solution structures of mEGF, hTGFα, FX-EGF_N, and FIX-EGF_N (Baron et al., 1992), the pdb files of mEGF (Montelione et al., 1992, : 3egf), hTGFα (Kline et al., 1990, Bruckhaven Data Bank file name: 4tgf), FX-EGF_N (Ullner et al., 1992, Bruckhaven Data Bank file name:1apo), and FIX-EGF_N (Baron et al., 1992, Bruckhaven Data Bank file name:1ixa) are obtained from the protein database of the Brookhaven National Laboratory (www.pdb.bnl.gov). The average conformer of each species was calculated from Insight II software for the purpose of comparison. The sources for the structure information in the following discussion will be referred from these pdb files.

One of the most distinct features of β-sheet as it occurs in known protein structures is its twist (Chothia, 1973). This twist always has the same handedness, although it has unfortunately been described by two conflicting conventions in the literature (Richardson, 1981). If defined in terms of the angle at which neighboring β strands cross each other, then the twist is left-handed (e.g., Quioch et al., 1977; Shaw and Muirhead, 1977); if defined in terms of the twist of the hydrogen bonding directions or of the peptide planes as viewed along a strand, then the twist is right-handed (eg., Schulz et al., 1974; Chothia et al., 1977). The right-handed definition will be used here.

The major secondary element of FIX-EGF_N in the N-terminal is an anti-parallel β-sheet. There are five residues in each β-strand, constituted by Gly16-Asp20 and Tyr25-Cys29. The β sheet has a right-handed twist, as expected. The first β-strand contains Cys^3 and the second strand contains Cys^4 and Cys^5. The values for the dihedral angles ϕ and ψ found in this β-sheet lie within the largest allowed region of the Ramachandran plot (the top left region in Figure 1.4.2), see Table 3.4.3.

A β turn is formed at polypeptide segment Asp21-Ser24. No hydrogen bond was found in this turn. The expected slow exchanging amide proton of residue 4 (Ser24) in the tight turn was not detected in this study. Turns are characterized based on ϕ and ψ angles of residue 2 and 3 in each turn (Table 1.5.1). In the structures, the average ϕ, ψ values of residue 2 and 3 in the polypeptide segment Asp21-Ser24 are ϕ_{22} =-110°, ψ_{22} =-81°; ϕ_{23} =-124°, ψ_{23} =-30°. Thus the turn between these two β strands is not standard (Ullner et al., 1993). In EGF and TGFα, the homologous position are type I turns (Moy et al., 1993, Montelione et al., 1992).

There is no evidence of N-terminal residues forming a third strand attached to this β sheet as has been observed in marine EGF (Montelione et al., 1987, 1992), human EGF (Cooke et al., 1987) and human TGF-α (Moy et al., 1993). On the contrary, residue 1-5 are poorly defined because there are only sequential constraints among these residues. An

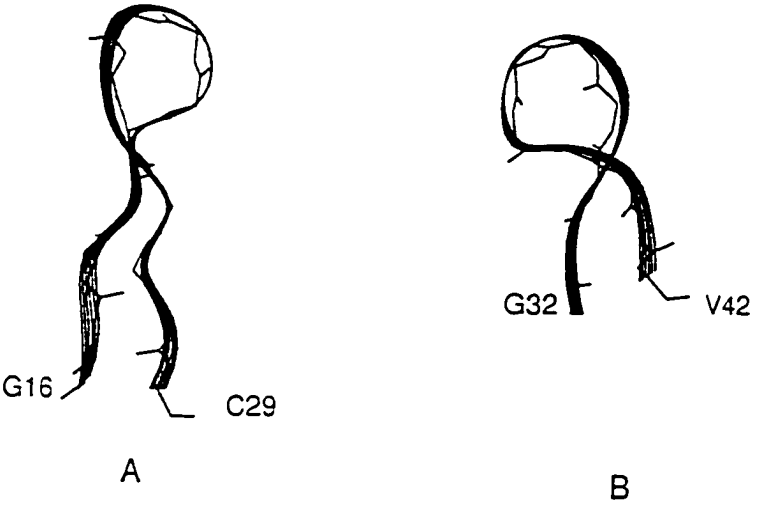
irregular turn is formed between Cys7-Pro11. The direction of the turn is away from the β-sheet (Figure 4.1.1). The segment Gln6-Cys12 is structured like a Ω turn. This Ω turn is almost perpendicularly to the surface of the β-sheet. Both arms of the turn are somewhat aligned with the direction of the β-sheet. The average ϕ and ψ dihedral angles of Asn10 are 44 and 79, which falls in the αL conformation. This conformation is observed in residue *i*+2 of type III turn (Rose et al., 1985).

Figure 4.1.1. A view of the backbone conformation of the polypeptide in FIX-EGF_N (made by RalMol 2.6 software). The N-terminal is on the top position. The Ω like turn is shown sitting at the top of the major β-sheet. The extensions of this turn are aligned with the directions of the β-sheet.

Residue segment Leu13-Gly16 turns at Asn14 and Gly15 (Figure 4.1.6). The ϕ and ψ angles of them are 40 and 21; 85 and 50, respectively, forming a type III′ β turn, in which both of the *i*+2 and *i*+3 residues has the left-handed α-helical conformation, which typical ϕ and ψ values are 57 and 47 (Table 1.4.2). Since the ideal ϕ and ψ values of I′ and III′ are so close, as shown in Table 1.5.1, that they could be distinguished only in the most highly refined protein structures and it was suggested by Richardson (1981) to eliminate type III as a distinct category. Therefore, residues Leu13-Gly16 would be categorized as a type I' turn. This turn was found to contain a hydrogen bond (slowly exchanging amide proton of glycine 16). In Baron's structure, the amide proton of Gly 60 (the homologous residue to Gly16 in our structure) was found to contain hydrogen bonds to the carbonyls of Cys 56 and Leu 57 (in 10 and 5 structures, respectively).

In the C-terminal of FIX-EGF_N, there is a minor β-sheet, with three amino acids in each strand. They are Gly32-Glu34 and Glu40-Val42 respectively. This β sheet is a right-handed twist like the β major sheet in the N-terminal. It is more twisted than the major β-sheet (Figure 4.1.2).

Figure 4.1.2. The two sheets in FIX-EGF_N showing right-handed twist: (a) the N-terminal anti-parallel β-sheet; (b) the C-terminal anti-parallel β-sheet.



There is a five residue turn between the two strands of the C-terminal β-sheet, with residue Gly 35-Glu 39. The turn is bent at Cys 38, where the backbone is pulled off by the disulfide bridge with Cys 29. Baron's structures show that Cys 38 has a positive ϕ angle (Baron et al., 1992), significantly different from the ϕ -151^0 found in our structure (Table 3.4.3). Glu39 is found as a β-bulge as in our study. A β-bulge, defined as a region between two consecutive β-type hydrogen bonds that includes two residues on one strand opposite a single residue on the other strand (Richardson, 1981), is found between residue Glu39, Leu40 and Glu34. Both the amide protons of Glu39 and Leu40 are hydrogen-bonded with the carbonyl group of Glu34. (Figure 4.1.3). This β-bulge is also described in Barron's paper and over half of the calculated structures show the same hydrogen bonds. In FX-EGF_N, a type I β-turn is distinguished with the homologous segment, residues 78-81 (Ullner, et al, 1992). In FIX-EGF_N, although the amide H of the homologous residue Cys38 is a slow-exchange NH, no hydrogen bond acceptor is found with this hydrogen bond candidate.

The hydrogen bond of Val42 HN...OC Gly32 was observed in the calculated solution structure. The conformations of the minor β sheet are the same with or without the constraint of this hydrogen bond (Figure 4.1.4). The backbone atoms of residue 32-43, and 6-43 between the average conformers with and without this hydrogen bond are superimposed one to another. The RMSDs are 0.18 Å and 0.52 Å, respectively. However, the amide proton of Val42 is not slowly exchanged. It is likely that the calculated structure is in the optimally low energy state, while the conformation in the solution is somehow dynamic.

The two β sheets are bent at residue Phe31 and Gly32, forming a β turn with residue Pro30 and Phe33 together. Although the ψ angle of *i*+1 and the ϕ angle of *i*+2 residues are not typical (refer to Table 1.5.1) this turn is identified as a type II turn (Table 4.1.1). Since Phe33 has a slowly exchange amide proton, this turn contains a hydrogen bond. This type II turn is highly conserved with mEGF (Kohda and Inagaki, 1992; Montelione et al., 1992), hTGFα (Kline et al., 1990, Moy et al., 1993), and FX-EGF_N (Ullner, et al., 1992). Glycine is conserved in position 3 of this turn in these EGF-like peptides (Figure 4.1.5), which is distinctive for type II turn. In type II turns, the carbonyl oxygen of the *i*+1 residue is very close to the β-carbon of the side chain of the *i*+3 residue, but the barrier is relieved if residue 3 is glycine (Richardson, 1981). It has been surveyed that 61% of the type II turns among 459 tight turns had a glycine in position 3 (Chou and Fasman, 1977). Phe/Tyr is identified as a consensus residue in this turn.

Figure 4.1.3. β-Bulge conformation of Glu 39 in FIX-EGF_N.

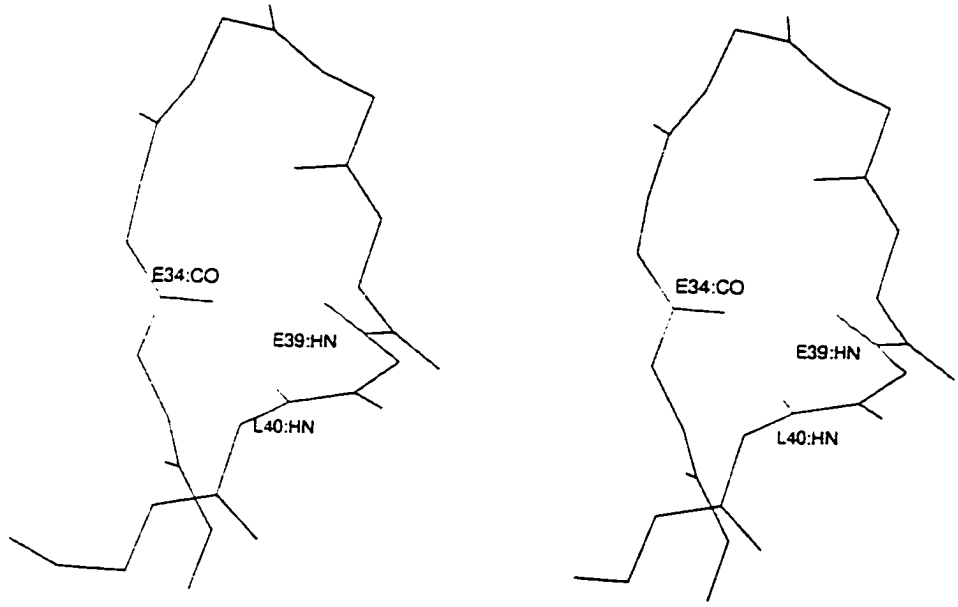


Figure 4.1.4. Superimpose the average conformers of the backbone atoms of residue 32-43 with and without the hydrogen bond of Val42:HN...OC:Gly32.

In summary, there is a major antiparallel β-sheet in N-terminal and a minor one in the C-terminal. A third β-strand is not found for both sheet as has been found in mEGF and hTGFα. Table 4.1.1 gives dihedral angles and type for the turns in FIX-EGF_N. Figures 4.1.6 and 4.1.7 show the secondary structure and the tertiary structure of FIX-EGF_N, respectively.

Figure 4.1.5. Sequence alignment of the conserved turn. The conserved glycine is bolded. The sequence number of the first residue in the turn is shown in the bracket.

FIX-EGF_N: (30)PF**G**F
FI-EGF_N: (73)AE**G**F
mEGF: (34)VI**G**Y
hTGFα: (35)HS**G**Y

Table 4.1.1. Identification of β turns in FIX-EGF_N

	i + 1		*i* + 2		
Turn	ϕ	ψ	ϕ	ψ	type
L13-G16	40	21	85	50	I′
D21-S24	-110	-81	-124	-30	nonstandard
P30-F33	-89	85	156	-14	II
G35-C38	-52	-51	-95	-122	nonstandard

Figure 4.1.6. A schematic diagram of the secondary structure of FIX-EGF_N.

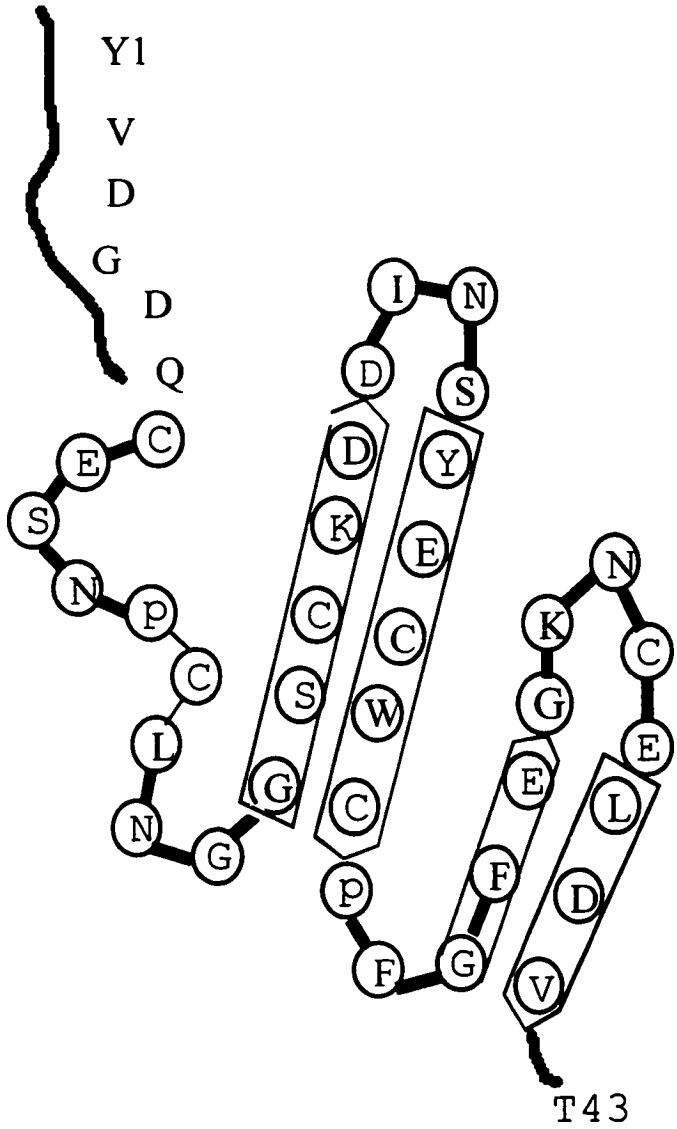
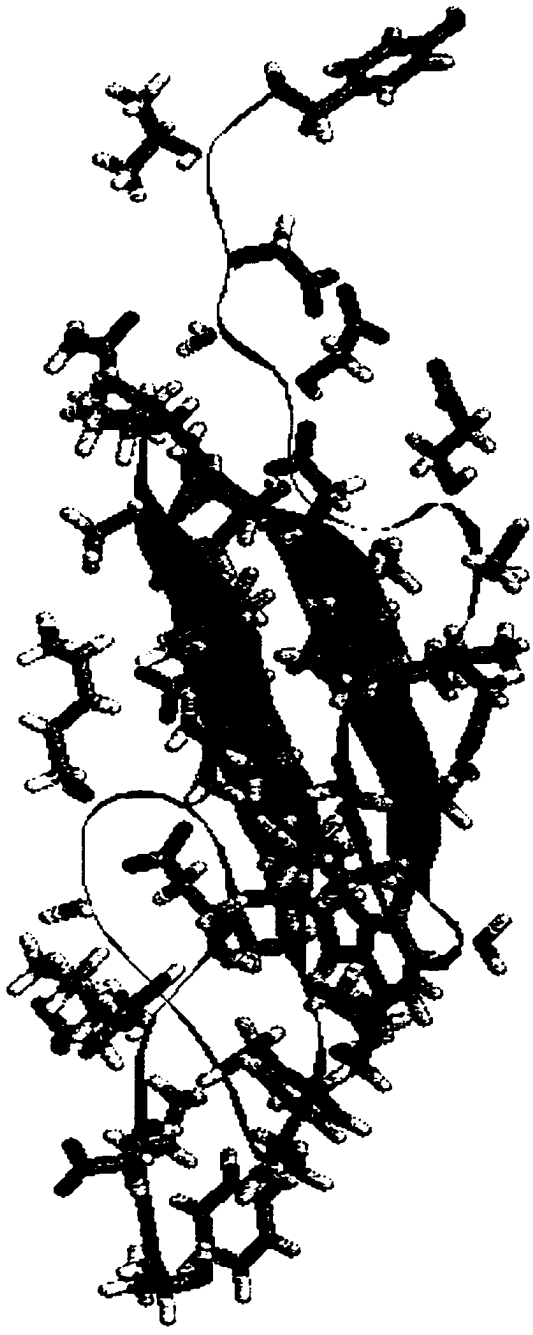


Figure 4.1.7. Tertiary structure of FIX-EGF_N. The major β-sheets are shown by ribbon bands. (This figure is made by RasMol 2.6).



4.2. Long range NOEs in the solution structure of FIX-EGF$_N$.

The backbone $C^{\alpha}H$ of Cys7 has long NOE interactions with the side chain protons of Pro11. Both Cys7 and Pro11 have one pair of long-rang NOE interactions with Cys18. Tyr25 shows long range interactions with several residues in addition to beta strand interactions, including Gln6, Pro11 and Lys36. For the Gln6 and Tyr25 residue pair, the inter residue NOE's include the backbone α-protons and the side-chain protons interactions. The Pro11 and Tyr25 residue pair has only side-chain interactions. There is a hydrophobic interaction between this pair. Figure 4.2.1 shows that the side-chain ring of Pro-11 is under the aromatic ring of Tyr-25. The γ-protons of Pro-11 are pointing the center of the aromatic ring of Tyr-25. This explains that the chemical shifts of the two γ-protons are up-field shifted for 0.71 and 1.39 ppm respectively, compared with random coil conformation (wüthrich, 1986). This hydrophobic interaction also observed in FX-EGF$_N$. The side-chain ring of Pro-54 is similarly located below the aromatic ring of Tyr-68 (Figure 4.2.1). γ-Protons of Pro-54 are up-field shifted 0.60 and 1.25 ppm respectively. This is interesting because this tyrosine residue had been shown to be essential for clotting activity (Hughes, et al., 1993). In addition to that, binding of calcium causes significant shifts in the 2,6 ring protons of Tyr25, and the homologous residues in FX-EGF$_N$ and in one of the EGF-like domains of human fibrillin-1 (Handford et al., 1995). Furthermore, it was reported that substitution of Pro11 with Gln causes severe Hemophila B (Lozier, et

al., 1990) and with Ala causes Hollywood and London 7 with only 7% and 10% of clotting activity respectively (Stenflo, 1991). It had been suggested by our experimental results (see Section 5.4.2) that Gln6 is one of the calcium binding ligands (Gong et al., 1993). The mutant of Qln6→Pro was found in New London disease with <1% clotting activity (see Figure 1.1.3).

There are a couple of long range NOE interactions between residues Cys12, Leu13, Asn14 and Asn37, Cys38, Glu39. Figure 4.2.2 shows that the backbones of C12-N14 goes parallel with these of N37-E39. This is similar to Baron's study (Barron et al., 1992). The amide proton of Leu13, which is slowly exchanging, is close to the carbonyl group of Asn37 at the distance of about 5.9 Å. A hydrogen bond has been found between these two residues in FIX-EGF_N by Baron and the consensus residues Leu56 and Asn80 in (Ullner et al., 1992).

There are a number of NOEs between the side-chain of Val42 with the backbone protons of Gly32 and the side-chains of Glu34, the side chain of Thr43 with the backbone α protons of Gly32, and the side-chain of F33 with the side-chain of D41. The side chains of residues E34, L40, and V42 are in the opposite side of the minor β sheet with F33, and D41 (Figure 4.2.3).

Figure 4.2.1. A stereo view of the hydrophobic interaction between the conserved proline and tyrosine in both FIX-EGF_N and FX-EGF_N.

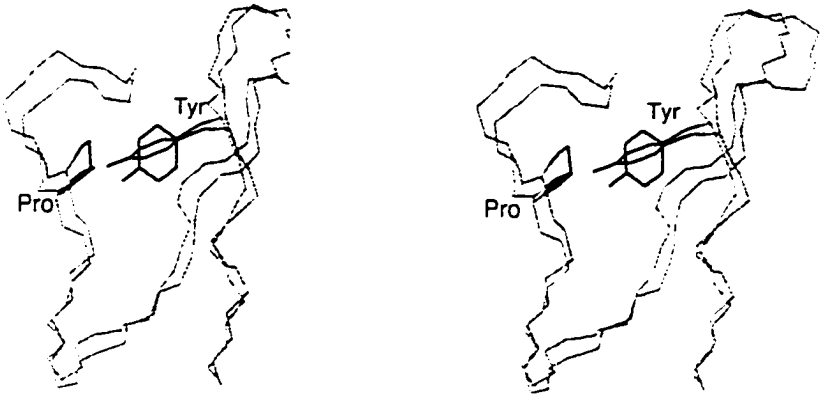


Figure 4.2.2. Long range NOE interactions between the N and C-terminals.

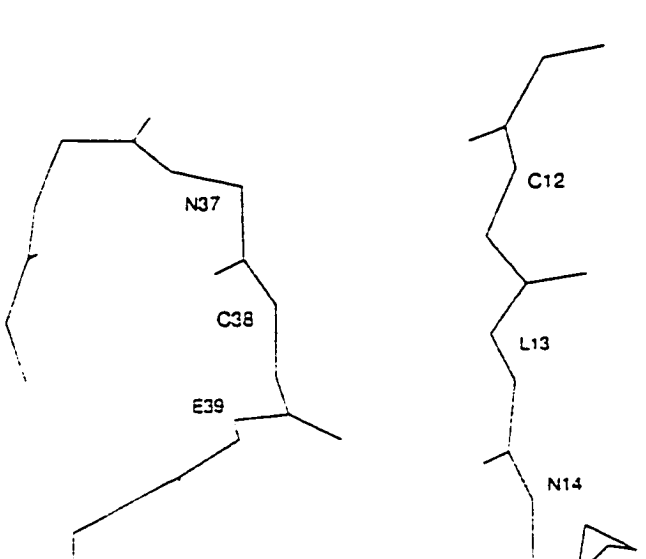
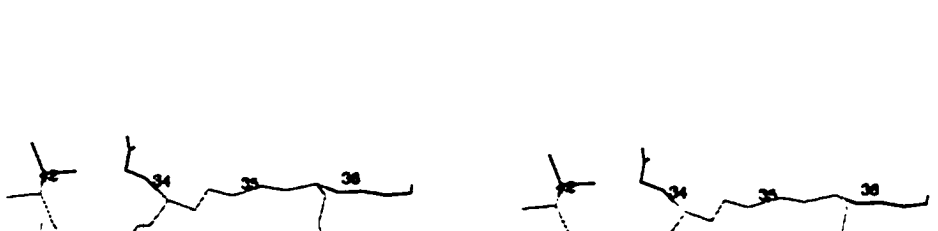


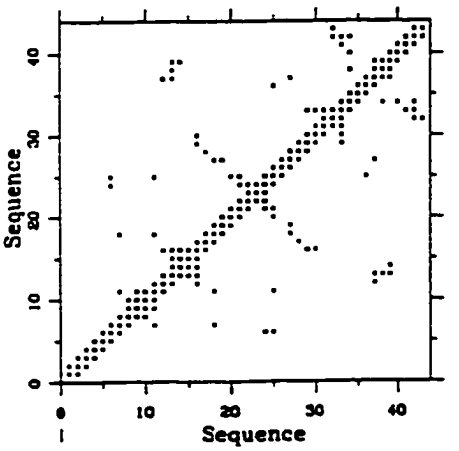
Figure 4.2.3 The side-chain relative directions of the minor b-sheet in FIX-EGF_N



4.3. Secondary structure and NOE distribution

Secondary structure can be identified from NOE distribution (Wüthrich, 1986). Figure 4.1.3 shows the distance constraints measured from FIX-EGF$_N$. The squares immediately adjacent to the diagnol are sequential NOE's. The regular anti-parallel β-sheets formed by the strands 16-20 and 25-29; 32-34 and 40-42 respectively are manifested by two lines oriented perpendicular to the diagonal. The turns including the loop 7-11 can be recognized from the shape of arrow along with the diagonal. It is formed typically by the serial NOE constrains *i*+1, *i*+2, *i*+3 (between Asp21 to Ser24) or plus *i*+4 (eg. between residue Cys12 to Gly16 and residue Cy29 to Phe33). There is no such NOE pattern between turn Gly35-Glu39, which is untypical. In addition, NOE distribution diagram of FIX-EGF$_N$ also shows the long range NOE's between residue 12-14 and 37-39.

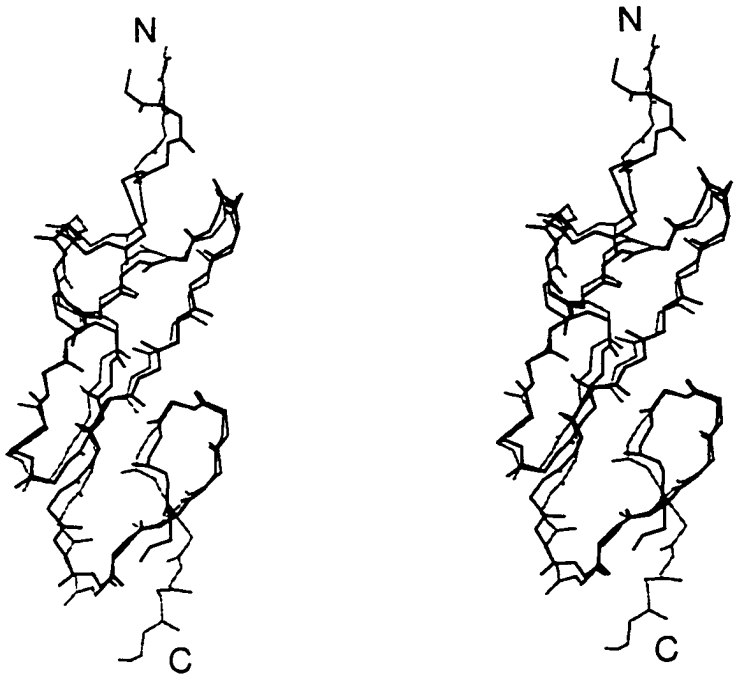
Figure 4.3.1. Summary of NOE distance constraints residue.



4.4. Comparison with the solution structure of FIX-EGF_N by Baron.

Baron et al., had published the solution structure of FIX-EGF_N (46-84), corresponding residue 2-40 in this study. Our result shows similar but more refined structures. The coordinates of Baron's structure were downloaded from Brookhaven National Institute' protein database. The results of both studies were compared by superimposing the backbones of Gln6-Glu40 (50-84) for the whole molecules, Gly16-Cys29 (60-73) for the major β-sheet, and Pro30-Glu40 (74-84) for the minor β-sheet on each other (Figure 4.4.1). The RMSD values are 1.0, 0.67, and 0.93, respectively. Because the sequence is shorter, Baron's structure did not conclude the second β-strand of the short β-sheet in our structure.

4.4.1. Stereo views of the solution structures of FIX-EGF_N from both studies. The black one with a shorter length is from Baron's study.



4.5. Comparison of the solution structures of mEGF, hTGFα, FX-EGF_N, and FIX-EGF_N.

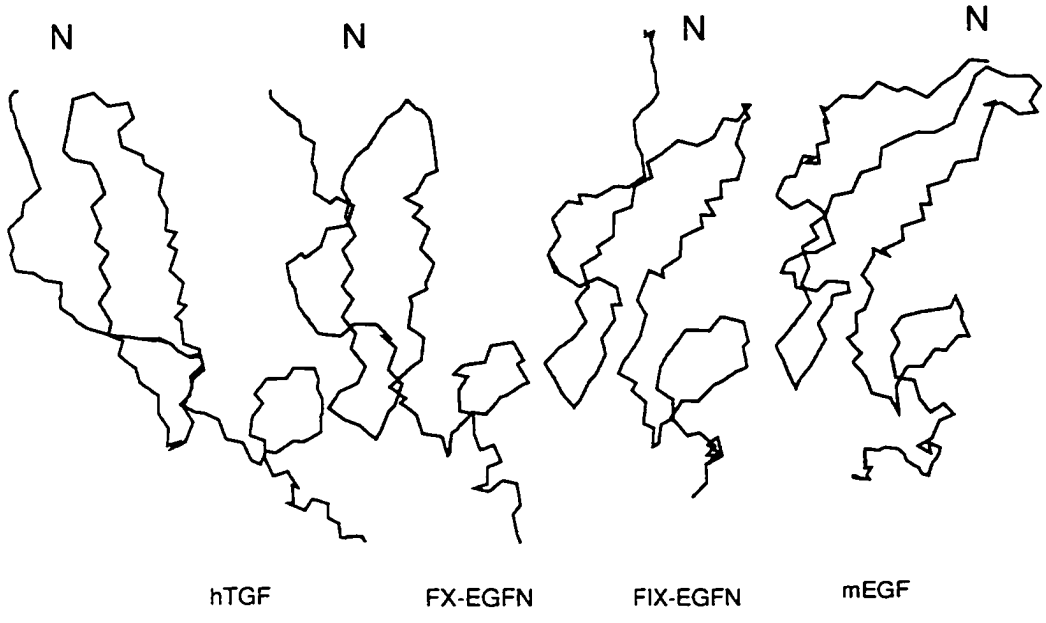
The secondary structures in these EGF-like peptides are similar to each other (see Table 4.5.2). For mEGF (Montelione et al, 1987, 1992) and hTGFα (Moy et al., 1993), the backbone structures have a three-stranded anti-parallel β-sheet in the N-terminal sequences, while in FX-EGF_N (Ullner et al., 1992) and FIX-EGF_N there are only two β-strands. In the C-terminal, a "double hairpin" structure are described in both mEGF (Montelione et al, 1987, 1992) and hTGFα (Moy et al., 1993) with two turns and a minor two-stranded antiparallel β-sheet. While in FX-EGF_N (Ullner et al., 1992) and FIX-EGF_N with the consensus two turns and a minor two-stranded antiparallel β-sheet. In general, the N-terminal of FIX-EGF_N and FX-EGF_N are flexible, while the C-terminals of EGF and TGFα are flexible.

Structures of the major β-sheet region in EGF-like peptides were compared by superimposing residues 18-22 and 29-35 of mEGF, 19-23 and 30-36 of hTGFα, and 59-63 and 68-75 of FX-EGF_N on the backbones of residue 16-20 and 25-31 of FIX-EGF_N. The RMSD are 0.93, 1.19, and 1.11 for mEGF, hTGFα, and FX-EGF_N respectively. Figure 4.5.1 shows that the sheets are right-hand twist.

Figure 4.5.1. A stereo view of the N-terminal β-sheets in the EGF-like polypeptides showing that the two β-strands are cross over each other. The ones with longer length are mEGF and hTGFα respectively.

In the homologous peptide segment between Cys^1 and Cys^2 in FX-EGF_N, residue Cys50-His53 is identified as a type I β turn (Ullner et al., 1992). There is an irregular turn (Cys7-Pro11) in this homologous segment of FIX-EGF_N (see Table 1.4.2 in Section 1.4). Both hTGFα and mEGF observed an irregular turn in this homologous segment from the pdb coordinates, which are reported as an irregular α-helix (Moy et al., 1993, Montelione, et al., 1992). They are similarly positioned above the β-sheet perpendicularly. The lengths of the turns are longer in mEGF and hTGFα than neither FIX-EGF_N or FX-EGF_N. Cys^1 and Cys^2 are located at the beginning and the end of the homologous polypeptide turn in all the EGFlike peptides. In addition, the first few N-terminal residues in FIX-EGF_N and FX-EGF_N are randomly set above the turns of the β-sheets, while those in mEGF and hTGFα are aligned next to the first β-strand due to the third β-strands of Tyr3-Pro4 in mEGF and Phe5-Asn6 in hTGFα, respectively (see Figure 4.5.2).

Figure 4.5.2. The backbone solution structures of the EGF-like poly-peptides.



These EGF-like structures contain a conservative turn (Figure 4.5.3 and 4.5.4). This conserved turn is identified as a type I′ turn in both FIX-EGF_N and FX-EGF_N (Ullner et al., 1992). Konda and Inagaki (1992) reported a type I′ turn, L15-N16-G17-G18, in mEGF. The last residues of this conserved turn are glycine, which is the beginning of the first β strand in the N-terminal in all cases. Both Figure 4.5.3 and 4.5.4 indicate that there is one residue deletion in hTGFα between the conserved cysteine and glycine. For this reason the turn starts at Cys^2 instead of one residue later. The conserved cystine and glycine residues are aligned. Both mEGF and FIX-EGF_N turn at residues asparagine and glycine, FX-EGF_N at asparagine and glutamine, and TGF at phenylalanine and histidine.

EGF-like peptides contain six cysteine residues in a 1-3, 2-4, and 5-6 bonding pattern. The structures of the EGF-like peptides are shown superimposed only on the backbones of the six cysteines of FIX-EGF_N (Figure 4.5.5). The RMSDs from the deviation of FIX-EGF_N are 0.92, 1.51, and 0.85 for mEGF, hTGFα, and FX-EGF_N, respectively. Figure 4.5.5 gives a stereo-view of residue 6-46 of mEGF, 8-47 of hTGFα, 50-85 of FX-EGF_N, and 7-42 of FIX-EGF_N. It shows that the backbones of these residues are approximately aligned each other. However, the minor β-sheets of mEGF and FIX-EGF_N, and the ones of hTGFα and FX-EGF_N are better aligned as pairs (Figure 4.5.5). Similarly, the alignments of the three disulfide bonds from the EGF-like peptides are compared through superimposing 18-22 and 29-46 of mEGF, 19-23 and 30-47 of hTGFα, and 59-63 and 68-85 of FX-EGF_N on the backbones of 16-20 and 25-42 of FIX-EGF_N (Figure 4.5.6). The RMSD deviated from FIX-EGF_N are 1.26, 2.21, and 1.58, respectively. It shows that three disulfide bonds are similarly positioned in all 4 peptides (Figure 4.5.6). However, the conformations of the second and third disulfide bonds of mEGF and FIX-EGF_N are more similar each other.

Figure 4.5.3. Stereo structures of the first conserved turn in these EGF-like poly-peptides.

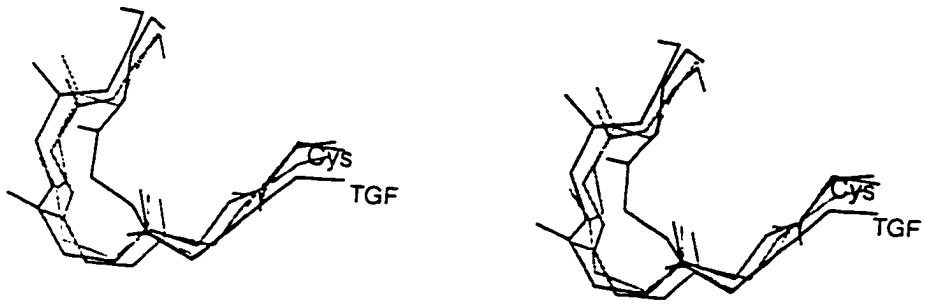


Figure 4.5.4. Sequence alignment of the conserved turn at potion of C12-G16 in FIX-EGF_N. Underline indicates the turn sequences. Notice that glycine is a conserved residue in addition to cysteine in this turn.

FIX-EGF_N:	(12)C<u>LNGG</u>
FX-EGF_N:	(55)C<u>LNQG</u>
mEGF:	(14)C<u>LNGG</u>
hTGFα:	(16)<u>CF.HG</u>

Figure 4.5.5. A Stereo view of the EGF-like peptide segments superimposed on the backbones of the six cysteine residues. It shows that the structures of these homologous segments are very similar.

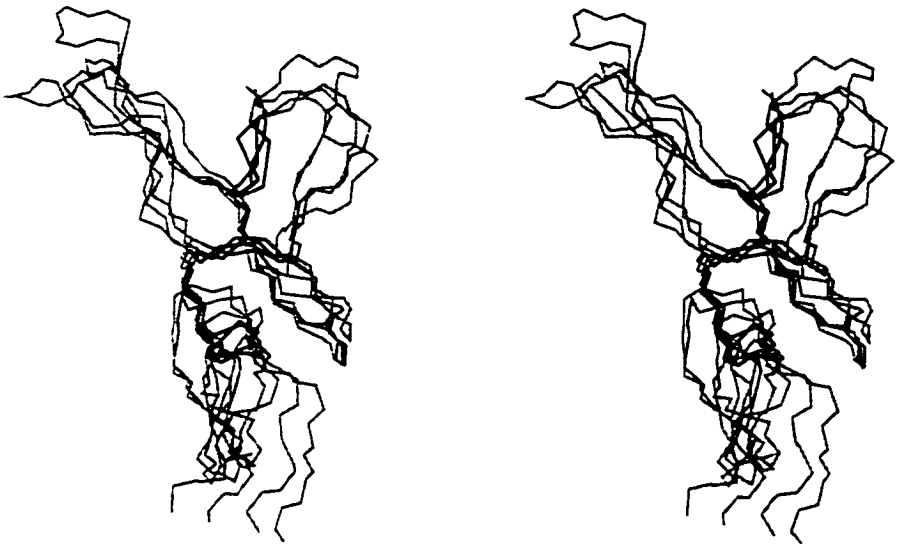
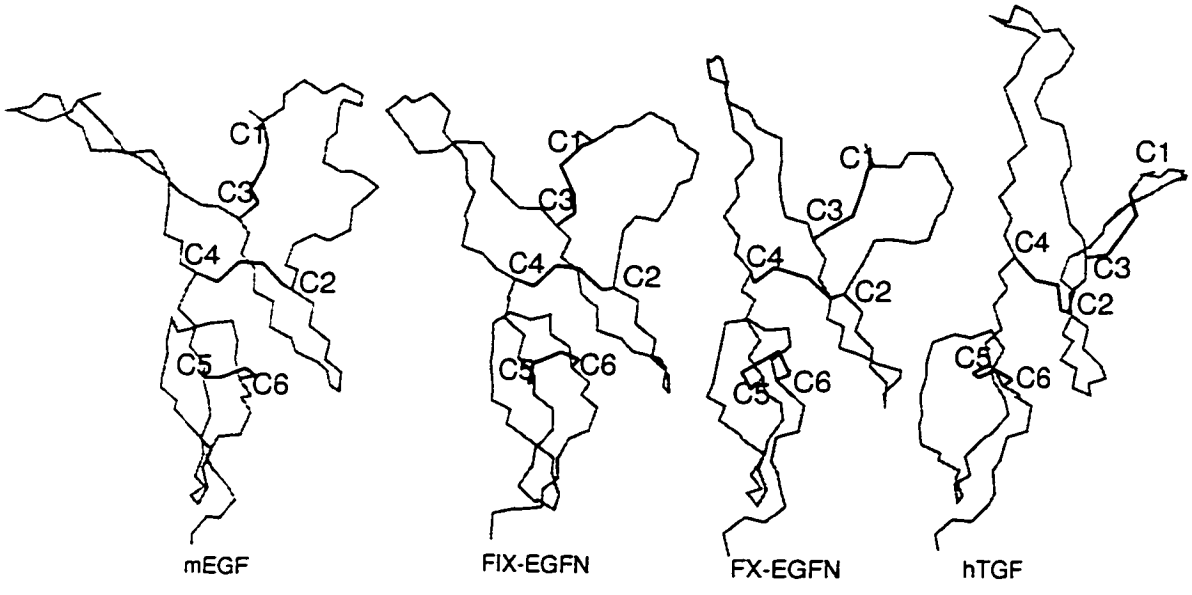


Figure 4.5.6. Alignment of three disulfide bonds of the EGF-like peptides result from superimposing on the backbones of residue 16-20 and 25-42 in FIX-EGF_N.



The C-terminal segment has a "double hairpin" architecture as described in mEGF (Montelione et al., 1992) and hTGFα (Moy et al., 1993). The average structures of mEGF, hTGFα, and FX-EGF_N were superimposed on the backbone chains of residues C29-V42 in FIX-EGF_N. The RMSD values are 1.34, 1.80, and 1.28, respectively. Figure 4.5.7 shows a similar "double hairpin" structure in each molecules. However, there is no NOE evidence showing that the homologous residue Cys 29 and Pro 30 in FIX-EGF_N form the third β-strand of the minor β-sheet as in mEGF (Montelione et al., 1992) and hTGFα (Moy et al., 1993). The arrow on the bottom of Figure 4.5.7 indicates the conserved turn (the position of Pro30-Phe33 in FIX-EGF_N), which has been discussed in section 4.1. Another conserved turn is observed in the double hairpin structure, as indicated by an arrow on the top of Figure 4.5.7 (position of Gly35-Cys38 in FIX-EGF_N). No typical turn was identified in FIX-EGF_N. Type I turn (G78-C81) was reported in FX-EGF_N (Ullner et al., 1992). This segment together with the following residue in both mEGF and hTGFα has been described as a left-hand helix due to one or more residues with positive value of ϕ (Montelione et al., 1992, Moy et al., 1993). In FIX-EGF_N, positive values were found in Gly35 and Lys36 of 14 and 2 conformers, respectively (Table 3.4.2). The sequence alignment of this conserved turn are shown in Figure 4.5.9. Interestingly glycine is detected as a conserved residue in this turn as the others (Figures 4.1.5 and 4.5.4). mEGF, hTGFα, and FX-EGF_N show an analogous β-bulge at residue Q39, E45, and E82, respectively, which are right after residue Cys^6 in

all the cases (Figure 4.5.7). The amide protons of the homologous residues are significantly downfield shifted compared with the random coil form (Table 4.5.1, Figures 4.7.1 and 4.7.2). Cys^6 is pulled back due to forming the disulfide bond with Cys^5. This disulfide bond might be responsible for stabilizing the unusual segment. All C-terminal sheets exhibit a similar right-handed twist (Figure 4.5.8).

The relative orientations of the EGF-like peptides are also compared by superimposing on the backbones of the second β-sheet (P30-V42) of FIX-EGF_N. Figures 4.5.10 and 4.5.11 show that these structures are similar. However, there are really two pairs of very similar orientation: mEGF and FIX-EGF_N; hTGFα and FX-EGF_N.

The directions of side-chains in these EGF-like solution structures are compared. The most of the side-chains are shown in the same directions. On the top of Figure 4.5.12. shows a stereo view of the polypeptide segment between Cys^2 and Cys^5 in EGF, TGFα, FX-EGF_N and FIX-EGF_N. In the bottom shows the C-terminal, starting with Cys^5.

Figure 4.5.7. A stereo view of comparison with the backbone conformations of the C-terminals among the EGF-like peptides

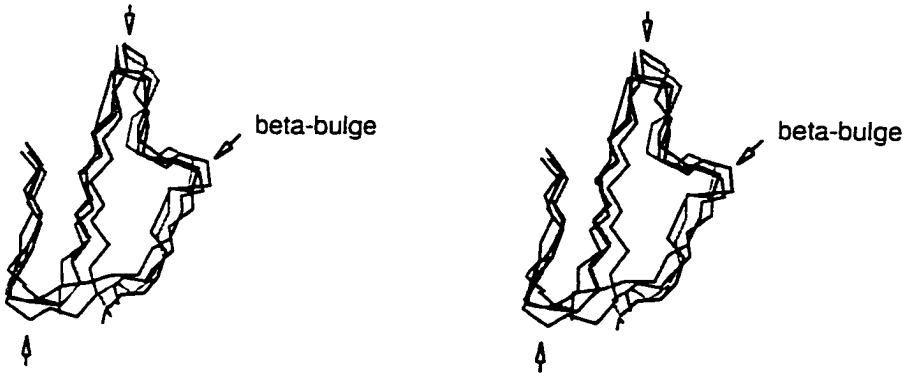


Figure 4.5.8. Structures of the C-terminal β-sheets in the EGF-like poly-peptides, showing that the minor β-strands are twisted

Figure 4.5.9. Sequence alignment of the conserved turn at potion of G35-C38 in FIX-EGF_N.

FIX-EGF_N: (35)GKNC
FX-EGF_N: (78)GKNC
mEGF: (39)GDRC
hTGFα: (40)GARC

Table 4.5.1. The chemical shifts of the homologous residues in EGF-like peptides and the random coil conformation.

Name	Residue	ppm(peptide)	ppm(random coil)[a]
mEGF	Gln43	9.99[b]	8.41
hTGFα	Glu45	9.55[c]	8.37
		9.96[d]	
FIX-EGF_N	Glu39	10.38	
FX-EGF_N	Glu82	10.63[e]	

[a]Wüthrich, 1986.
[b]Montelione et al., 1988.
[c]Moy et al., 1993.
[d]Kline et al., 1990
[e]Selander et al., 1990.

Figure 4.5.10. A stereo view of the relative orientations of the two β-sheets in the EGF-like poly-peptides, the backbones of which are superimposed on residues 29-42 in the C-terminal of FIX-EGF_N. From left to right are hTGFα, FX-EGF_N, FIX-EGF_N, and mEGF, respectively.

Figure 4.5.11. Structures of the EGF-like polypeptides were X-transferred and shown individually.

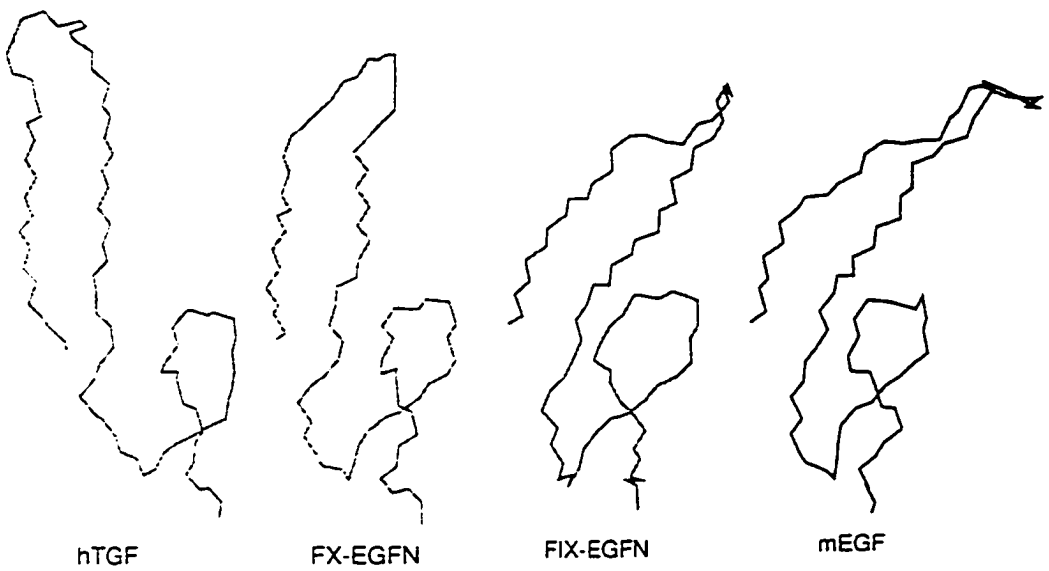


Figure 4.5.12. Stereo views of the superimposed backbone conformations of the EGF-like polypeptides with sidechains. On the top shows the segment between Cys^2 and Cys^5. The segment starting with Cys^5 shows on the bottom. The flexible segment in the C-terminals (labeled as "C") are not considered.

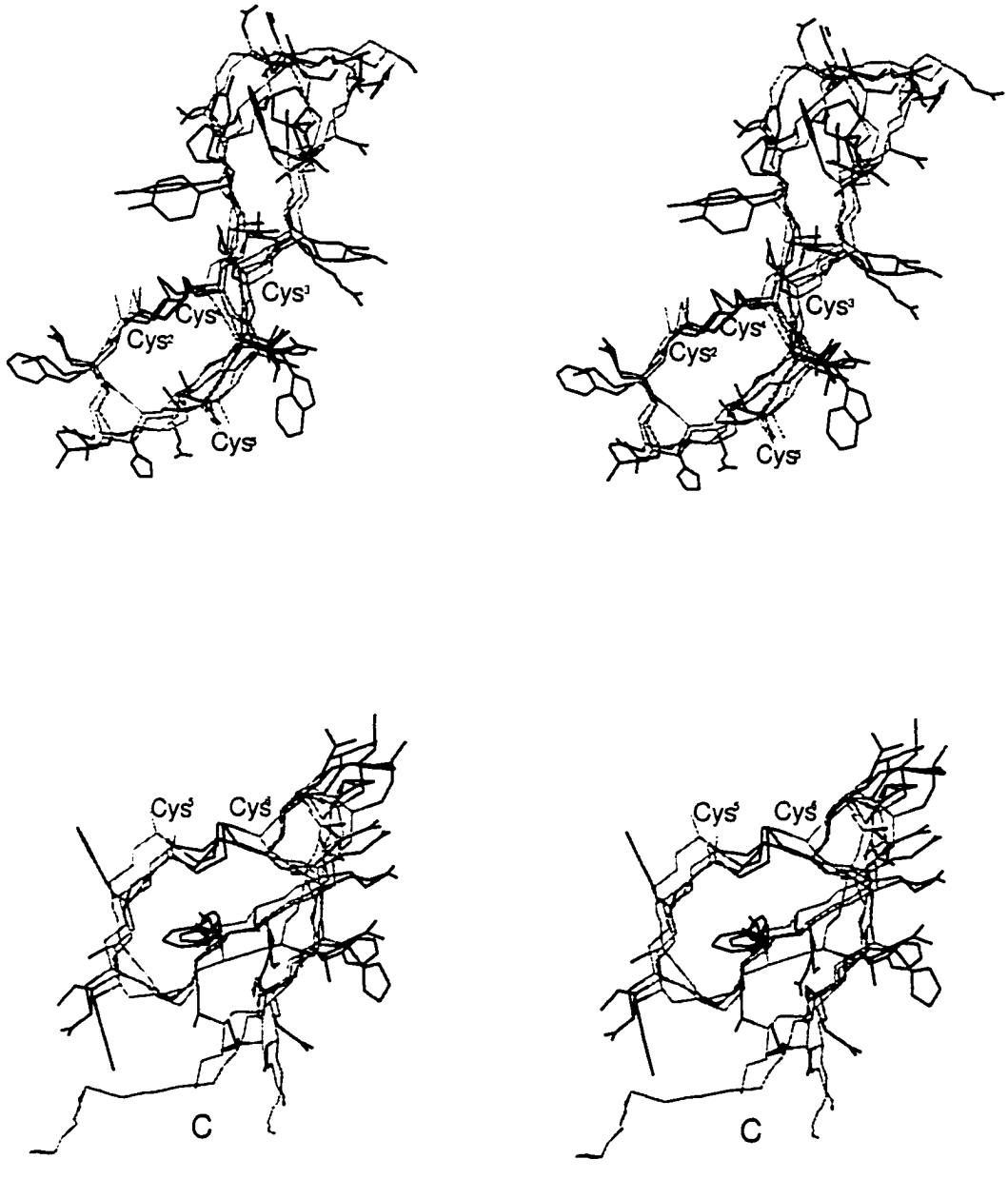


Table 4.5.2. Summary of homologous secondary elements occur in EGF-like peptides.

Peptide	Segment	Identification
FIX-EGF$_N$		
	C7-P11:	An irregular turn
	L13-G16:	A type I′ turn
	G16-D20, Y25-C29:	Antiparallel β sheets
	D21-S24:	A non-standard β turn
	P30-F33:	A type II′ like turn
	G32-E34, L40-V42:	Antiparallel β sheets
	G35-C38:	A non-standard β turn
	E39:	A β bulge
mEGF[a]		
	C6-C14:	An irregular helical conformation
	P7-L15:	An irregular right-handed helical structure
	L15-G18:	Type I′ turn[b]
	Y3-P4, G18-I23, S28-C33:	A three-strnaded antiparallel β sheet
	E24-D27:	Type I turn
	V34-Y37:	Type II turn
	Gly39-Gln43:	A Left-handed helix[c]
	C33-V34,Y37-S38,Thr44-D46:	A tree-stranded antiparallel β sheet
hTGFα[d]		
	F5-N6 :	A third antiparallel β strand
	D7-C16:	An irregular helical conformation
	G19-L24, K29-C34:	Antiparallel β sheet
	V25-D28:	Type I
	H35-Y38:	Type II
	Gly40-Glu44:	A Left-handed helix
	H34-H35, Y38-V39, H45-Ala46:	A third-stranded antiparallel β sheet
FX-EGF$_N$[e]		
	G59-β63, Y68-C72:	Antiparallel β sheet (right-handed twist)
	C50-H53:	Type I β turn
	L56-G59	Type I′ β turn
	G64-D67	Nonstandard
	F76-E77, F83-S84	Antiparallel β sheet
	A73-F76	Type II β turn
	G78-C81	Type I β turn

[a]Montelione et al., 1992. (pH 3.1).

[b]Kohda, D. and Inagaki, F., (1992) *Biochem. 31,* 11928-11939. (pH 2.0)

[c]12 out the 16 conformers.

[d]Moy et al., 1993

[e]Ullner et al., 1992

4.6. Sequence alignment of EGF-like peptides

Sequence alignments of EGF, TGFα, FX-EGF_N and FIX-EGF_N are done pairewise by BESTFIT program (Table 4.6.1 and Figure 4.6.1). It shows that FIX-EGF_N and FX-EGF_N have the highest identity and similarity. Surprisingly, the solution structures of the EGF-like domains from the two blood clotting factors are not the most similar, as shown in Figure 4.5.2. In fact, FIX-EGF_N is more like mEGF, and FX-EGF_N is more like hTGFα.

Table 4.6.1. The percentage of similarity and identity of sequence alignments by BESTFIT program.

Peptides	Identity* (%)	Similarity* (%)
FIX-EGF_N/FX-EGF_N	65	78
FIX-EGF_N/mEGF	36	56
FIX-EGF_N/hTGFα	23	49
FX-EGF_N/mEGF	37	60
FX-EGF_N/hTGFα	25	48
mEGF/hTGFα	33	52

*Identity percentages are calculated over the total number of all the aligned residues. Similarity ones are done by comparing the total number of the aligned residues which are only not identity.

Figure 4.6.1. Sequence alignment results by BESTFIT program. ("|" stands for identity, ":" for very similar, and "." for similar.)

FIX_EGF$_N$/FX-EGF$_N$:

```
              3 DGDDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDV 42
                ||||||||::||||.| |||:|..|.| |: ||||||||:::
              2 DGDDQCEGHPCLNQGHCKDGIGDYTCTCAEGFEGKNCEFST 41
```

FIX_EGF$_N$/mEGF:

```
              1 YVDGDQCESNPCLNGGSCK..DDINSYECWCPFGFEGKNCE 39
                |.:.. : .. ||||| |.  :.::||.| |.:|:.|..|:
              3 YPGCPSSYDGYCLNGGVCMHIESLDSYTCNCVIGYSGDRCQ 43
```

FX_EGF$_N$/mEGF:

```
              3 GDQCEGHPCLNQGHCK..DGIGDYTCTCAEGFEGKNCE 38
                .. ::|. |||.| |.  ::::.|||.|. |:.|..|:
              7 PSSYDGY.CLNGGVCMHIESLDSYTCNCVIGYSGDRCQ 43
```

mEGF/hTGFα:

```
            1 NSYPGCPSSYDGYCLNGGVCMHIESLDSYTCNCVIGYSGDRCQTRDL 47
              . : :||.|...:|::|.. : ::. |. .| |  || |.||:  ||
            3 SHFNDCPDSHTQFCFHGTCRFLVQE.DKPACVCHSGYVGARCEHADL 48
```

FIX_EGF$_N$/hTGFα:

```
             1 YVDGDQCESNPCLNGGS...CKDDINSYECWCPFGFEGKNCELD 41
               : |..::... |::|.:   ..:|  .: | :  |: | .|| .
             5 FNDCPDSHTQFCFHGTCRFLVQEDKPACVCHS..GYVGARCEHA 46
```

FX_EGF$_N$/hTGFα:

```
             1 KDGDQCEGHPCLNQGHCKDGIGDY..TCTCAEGFEGKNCEFST 41
               .|..:...: |: :| |:  :.:   .|.| .|: | .|| ..
             6 NDCPDSHTQFCF.HGTCRFLVQEDKPACVCHSGYVGARCEHAD 47
```

Sequence alignments were also obtained based on the structure superimposion. The backbones of mEGF, hTGFα, and FX-EGF_N were superimposed on the backbones of G16-D20, Y25-V42 of FIX-EGF_N. Small differences of relative orientation of the β-sheets are ignored in this sequence alignment. The result is shown in Figure 4.6.2. The aligned sequences are bolded. Turns are indicated by underline. some turns are observed from the structures, which are not classified or described by authors (see Table 4.5.2). β-Strands are shown by italic. Sequences in parenthesis in Figure 4.6.2 were not considered for alignment. The observed bulge residue is indicated by larger size of the letter symbol, which is E, E, Q, and E in FIX-EGF_N, FX-EGF_N, mEGF, and hTGFα, respectively.

Figure 4.6.2. Sequence alignments of the EGF-like peptides by structure comparisons in the order of FIX-EGF_N, FX-EGF_N, mEGF, and hTGFα.

(YVDGDQ)**C..E(.)SNPCLNGGSCKDD.IN.SYECWCPFGFEGKNCELDV**(T)
(KDGDQ)**C..E(.)GHPCLNQGHCKDG.IG.DYTCTCAEGFEGKNCEFST**(R)
(N*SYPG*)**CPSS(Y)DGYCLNGGVCMHI**ESLD**SYTCNCVIGYSGDRCQTRD**(LRWWELRO)
(VVSH*FND*)**CPDS(H)TQFCF.HGTCRFL**VQED**KPACVCHSGYVGARCEHAD**(LLA)

Similarities and identities are compared in pair between these EGF-like poly-peptides. Only the aligned amino acid sequences are accounted. Similar amino acid pairs are defined as the ones scoring non-negatively in an amino acid scoring substitution matrix (Figure 4.6.3.) developed by Henikoff et al., (1992).

Figure 4.6.3. A. Matrix of scoring similarities of amino acid pairs. Negative values indicate non-similarity, "0" means similar, and positive values indicate very similar. For example , the similarity of Leu and Glu is scored for -3, which indicates they are very unsimilar. The score is located between "L" and "E" in both dimensions.

	A	B	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	X	Y	Z
A	4	-2	0	-2	-1	-2	0	-2	-1	-1	-1	-1	-2	-1	-1	-1	1	0	0	-3	-1	-2	-1
B	-2	6	-3	6	2	-3	-1	-1	-3	-1	-4	-3	1	-1	0	-2	0	-1	-3	-4	-1	-3	2
C	0	-3	9	-3	-4	-2	-3	-3	-1	-3	-1	-1	-3	-3	-3	-3	-1	-1	-1	-2	-1	-2	-4
D	-2	6	-3	6	2	-3	-1	-1	-3	-1	-4	-3	1	-1	0	-2	0	-1	-3	-4	-1	-3	2
E	-1	2	-4	2	5	-3	-2	0	-3	1	-3	-2	0	-1	2	0	0	-1	-2	-3	-1	-2	5
F	-2	-3	-2	-3	-3	6	-3	-1	0	-3	0	0	-3	-4	-3	-3	-2	-2	-1	1	-1	3	-3
G	0	-1	-3	-1	-2	-3	6	-2	-4	-2	-4	-3	0	-2	-2	-2	0	-2	-3	-2	-1	-3	-2
H	-2	-1	-3	-1	0	-1	-2	8	-3	-1	-3	-2	1	-2	0	0	-1	-2	-3	-2	-1	2	0	
I	-1	-3	-1	-3	-3	0	-4	-3	4	-3	2	1	-3	-3	-3	-3	-2	-1	3	-3	-1	-1	-3
K	-1	-1	-3	-1	1	-3	-2	-1	-3	5	-2	-1	0	-1	1	2	0	-1	-2	-3	-1	-2	1
L	-1	-4	-1	-4	-3	0	-4	-3	2	-2	4	2	-3	-3	-2	-2	-2	-1	1	-2	-1	-1	-3
M	-1	-3	-1	-3	-2	0	-3	-2	1	-1	2	5	-2	-2	0	-1	-1	-1	1	-1	-1	-1	-2
N	-2	1	-3	1	0	-3	0	1	-3	0	-3	-2	6	-2	0	0	1	0	-3	-4	-1	-2	0
P	-1	-1	-3	-1	-1	-4	-2	-2	-3	-1	-3	-2	-2	7	-1	-2	-1	-1	-2	-4	-1	-3	-1
Q	-1	0	-3	0	2	-3	-2	0	-3	1	-2	0	0	-1	5	1	0	-1	-2	-2	-1	-1	2
R	-1	-2	-3	-2	0	-3	-2	0	-3	2	-2	-1	0	-2	1	5	-1	-1	-3	-3	-1	-2	0
S	1	0	-1	0	0	-2	0	-1	-2	0	-2	-1	1	-1	0	-1	4	1	-2	-3	-1	-2	0
T	0	-1	-1	-1	-1	-2	-2	-2	-1	-1	-1	-1	0	-1	-1	-1	1	5	0	-2	-1	-2	-1
V	0	-3	-1	-3	-2	-1	-3	-3	3	-2	1	1	-3	-2	-2	-3	-2	0	4	-3	-1	-1	-2
W	-3	-4	-2	-4	-3	1	-2	-2	-3	-3	-2	-1	-4	-4	-2	-3	-3	-2	-3	11	-1	2	-3
X	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
Y	-2	-3	-2	-3	-2	3	-3	2	-1	-2	-1	-1	-2	-3	-1	-2	-2	-2	-1	2	-1	7	-2
Z	-1	2	-4	2	5	-3	-2	0	-3	1	-3	-2	0	-1	2	0	0	-1	-2	-3	-1	-2	5

Figure 4.6.4. Identities and similarities of the EGF_like peptides in pair compared with the sequence alignments from the structures.

FIX_EGF$_N$/FX-EGF$_N$:

```
                (YVDGDQ)CESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDV(T)
                        ||.:|||| | ||| |..| | |  ||||||||...
                 (KDGDQ)CEGHPCLNQGHCKDGIGDYTCTCAEGFEGKNCEFST(R)
```

FIX_EGF$_N$/mEGF:

```
(YVDGDQ)C..E(.)SNPCLNGGSCKDD(.IN.)SYECWCPFGFEGKNCELDV(T)
        |  .   .. ||||| |           || | | .|:.|..|:
 (NSYPG)CPSS(Y)DGYCLNGGVCMHI(ESLD)SYTCNCVIGYSGDRCQTRD(LRWWELRO)
```

FX_EGF$_N$/mEGF:

```
  (KDGDQ)C..E(.)GHPCLNQGHCKDG(.IG.)DYTCTCAEGFEGKNCEFST(R)
         |  .      ||| | |          .|||.|. |:.| .|:
  (NSYPG)CPSS(Y)DGYCLNGGVCMHI(ESLD)SYTCNCVIGYSGDRCQTRD(LRWWELRO)
```

mEGF/hTGFα:

```
   (NSYPG)CPSSYDGYCL(N)GGVCMHIESLDSYTCNCVIGYSGDRCQTRD(LRWWELRO)
          ||.|:  :|.    |.|  :   |. .| |  || | ||:  |
 (VVSHFND)CPDSHTQFCF(.)HGTCRFLVQEDKPACVCHSGYVGARCEHAD(LLA)
```

FIX_EGF$_N$/hTGFα:

```
   (YVDGDQ)C..E(.)SNPCL(N)GGSCKDD(.IN.)SYECWCPFGFEGKNCELDV(T)
           |  .   :: |.    |:|:         .  | |  |: | .||
  (VVSHFND)CPDS(H)TQFCF(.)HGTCRFL(VQED)KPACVCHSGYVGARCEHAD(LLA)
```

FX_EGF$_N$/hTGFα:

```
     (KDGDQ)C..E(.)GHPCL(N)QGHCKDG(.IG.)DYTCTCAEGFEGKNCEFST(R)
            |  .    . |.   .| |:          .|.| .|: | .||  :
   (VVSHFND)CPDS(H)TQFCF(.)HGTCRFL(VQED)KPACVCHSGYVGARCEHAD(LLA)
```

This structure alignment gave a relatively different result compared with BESTFIT program (see Figure 4.6.1). However this result again shows that FIX-EGF_N and FX-EGF_N have the highest identity and similarity (Table 4.6.2). Interestingly, mEGF and hTGFα do not show a relative higher identity and similarity.

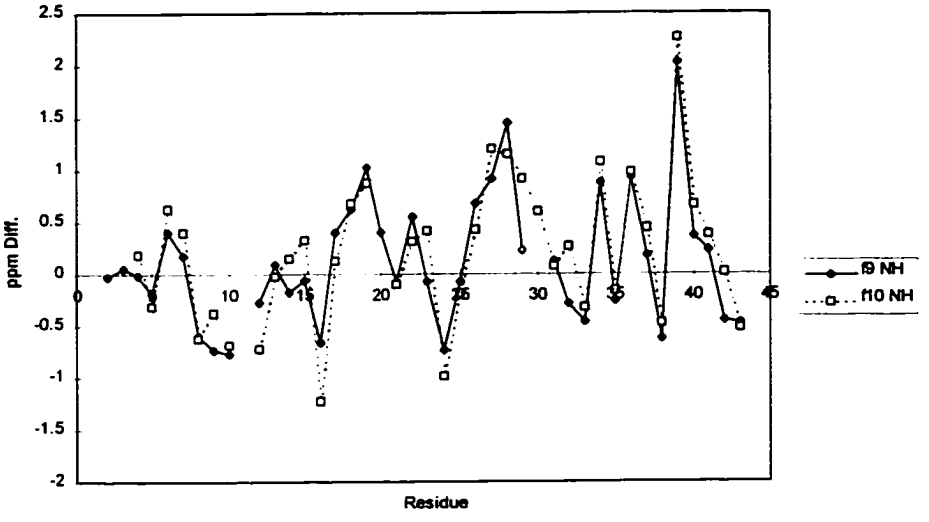
Table 4.6.2. The percentage of similarity and identity of sequence alignments from structure alignments.

Peptides	Identity (%)	Similarity (%)
FIX-EGF_N/FX-EGF_N	61	19
FIX-EGF_N/mEGF	41	27
FIX-EGF_N/hTGFα	36	25
FX-EGF_N/mEGF	38	24
FX-EGF_N/hTGFα	36	30
mEGF/hTGFα	37	23

4.7. Chemical shift differences of the backbone protons compared with random coil conformations.

The chemical shifts of the backbone hydrogens (amide and alpha protons) are affected by primary, secondary, and tertiary structures (Wüthrich, 1986). In order to compare the contributions of conformation influences on the chemical shifts between these EGF-like polypeptides, the chemical shifts of backbone hydrogens (HN and $C_\alpha H$) are subtracted by the corresponding random-coil chemical shifts (Wüthrich, 1986). These conformation-dependent shifts are plotted with the sequence alignments based on the results of structure superimpositions (Figure 4.7.1 to 4.7.6). The results indicate that the conformation influences of them are distinctly similar in the major b-sheets and the C-terminals. In the N-terminals, the conformation-dependent chemical shift patterns of mEGF and hTGFα are similar to each other as are those for FIX-EGF_N and FX-EGF_N. This is consistent with the fact that both FIX-EGF_N and FX-EGF_N do not have the third β-strand in the N-terminal found in mEGF and hTGFα.

Figure 4.7.1. Conformation-dependent chemical shifts of FIX-EGF_N and FX-EGF_N.



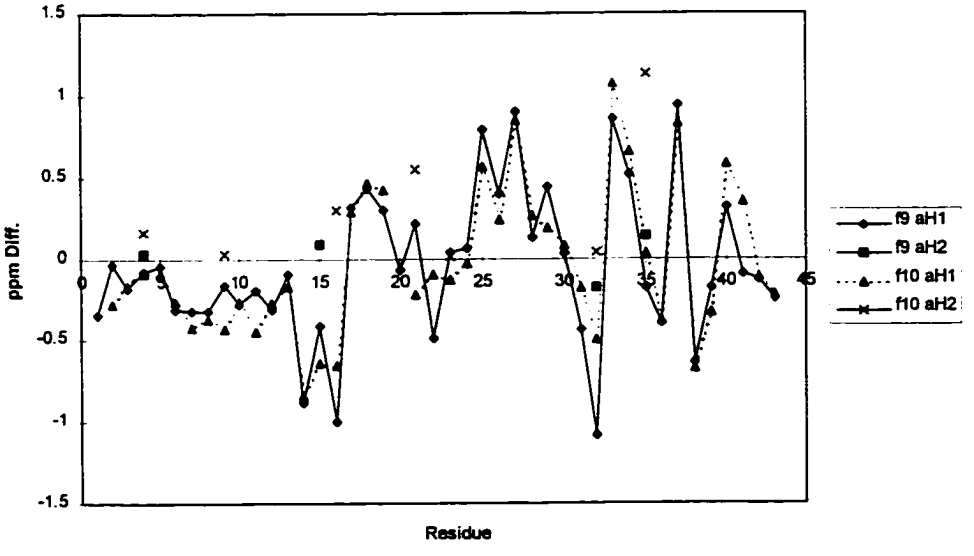
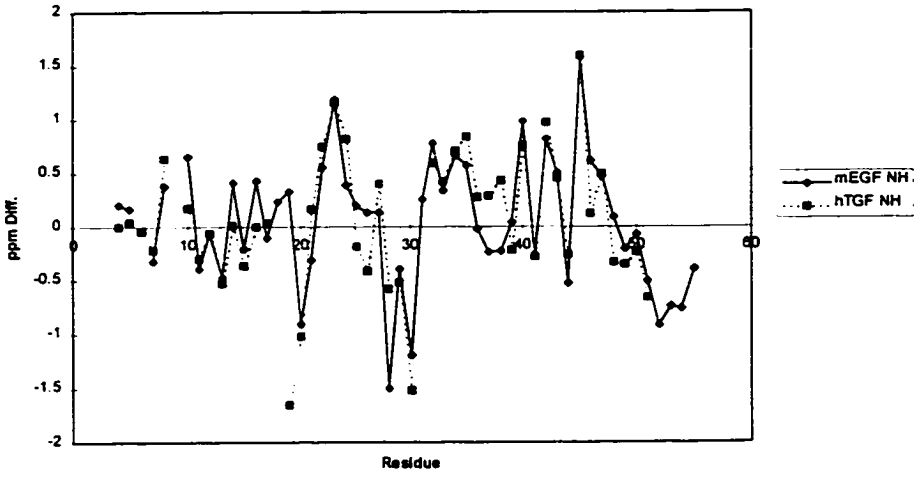


Figure 4.7.2. Conformation dependent chemical shifts of mEGF and hTGFα.



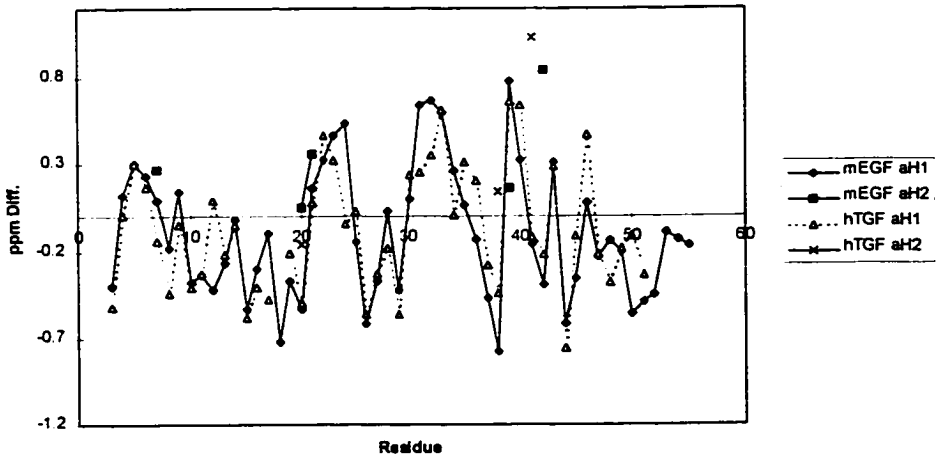
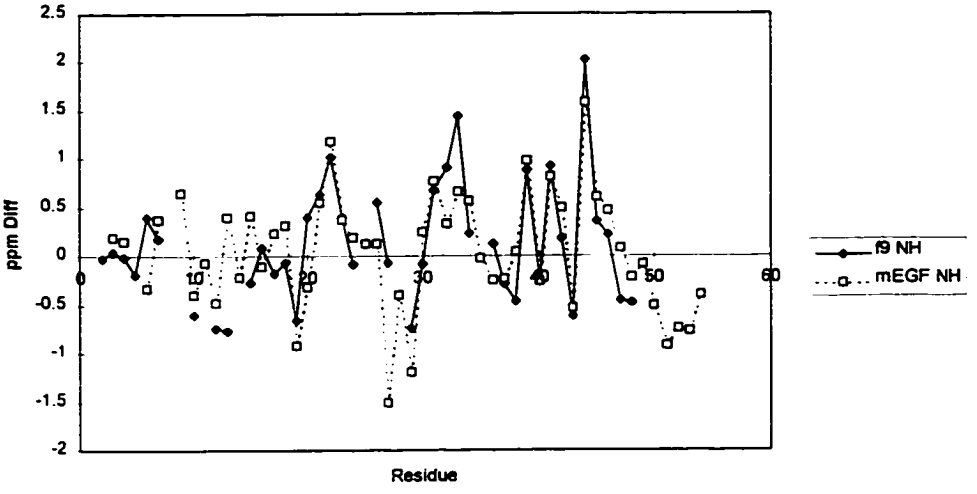


Figure 4.7.3. Conformation dependent chemical shifts of FIX-EGF_N and mEGF.



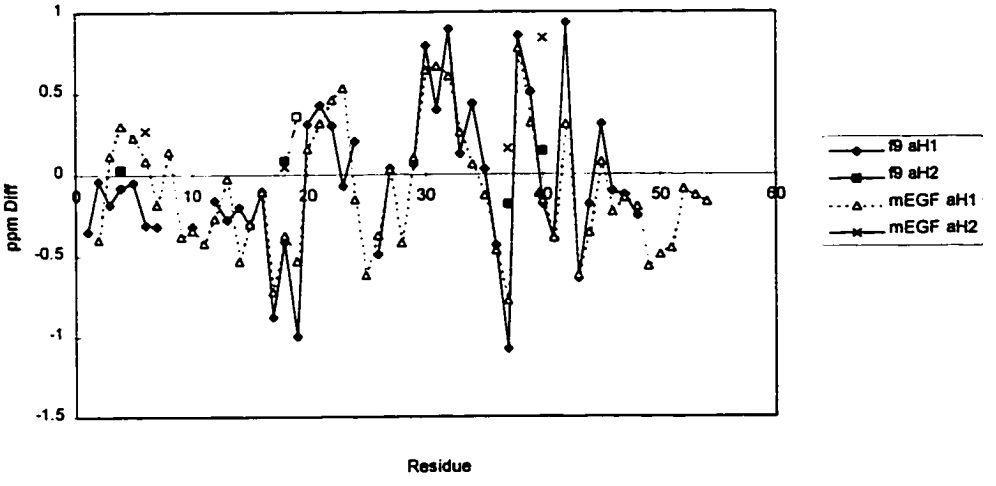
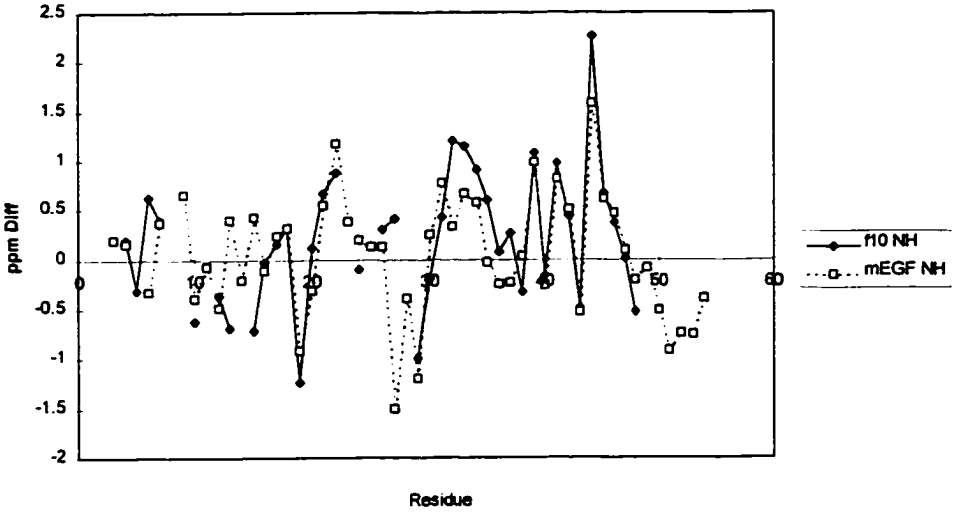


Figure 4.7.4. Conformation dependent chemical shifts of FX-EGF_N and mEGF.



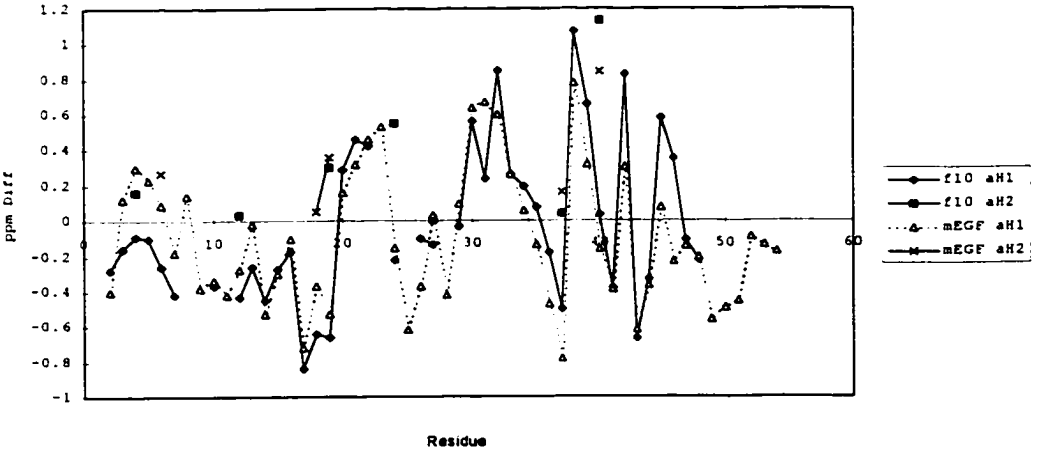


Figure 4.7.5. Conformation dependent chemical shifts of FIX-EGF_N and hTGFα.

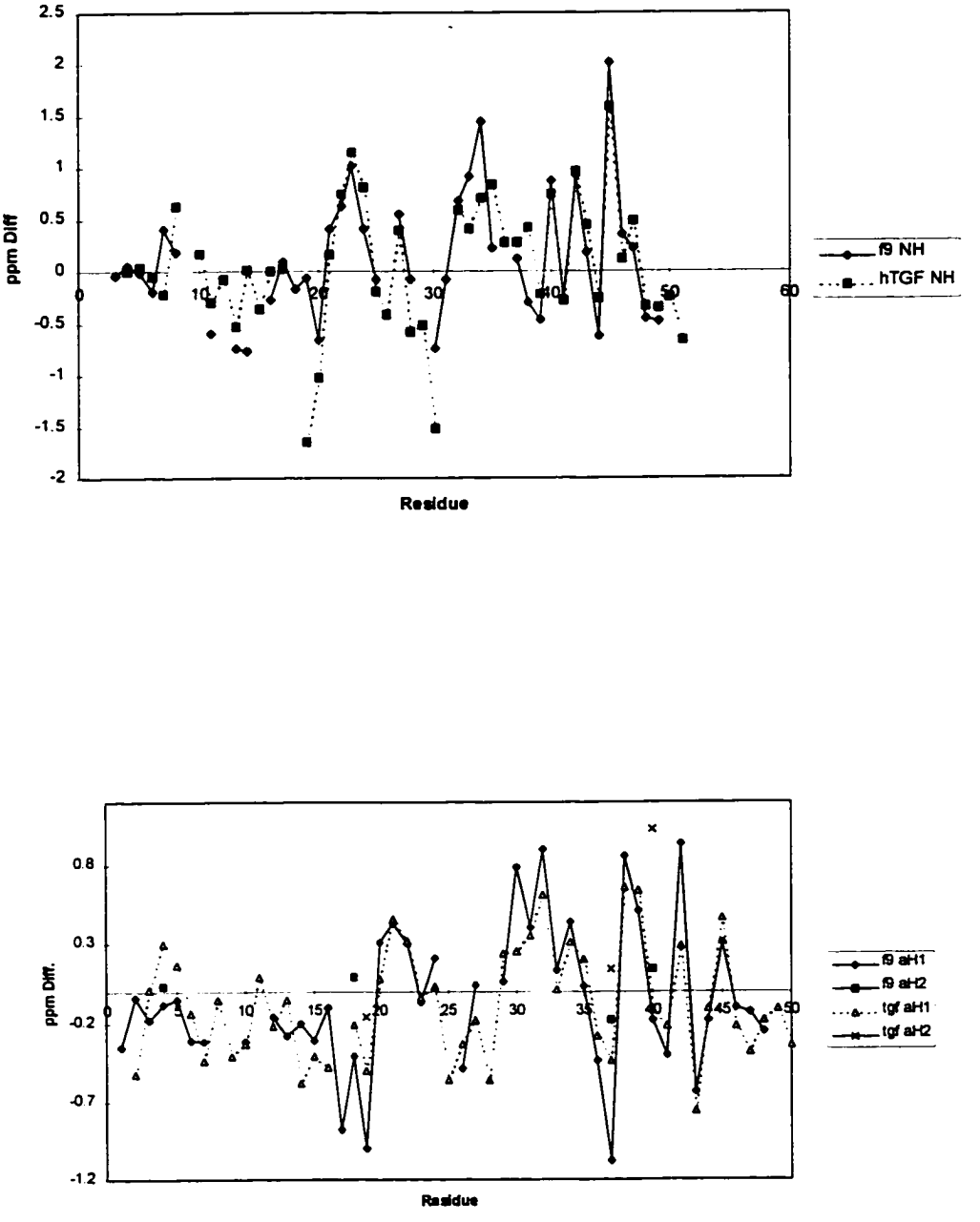
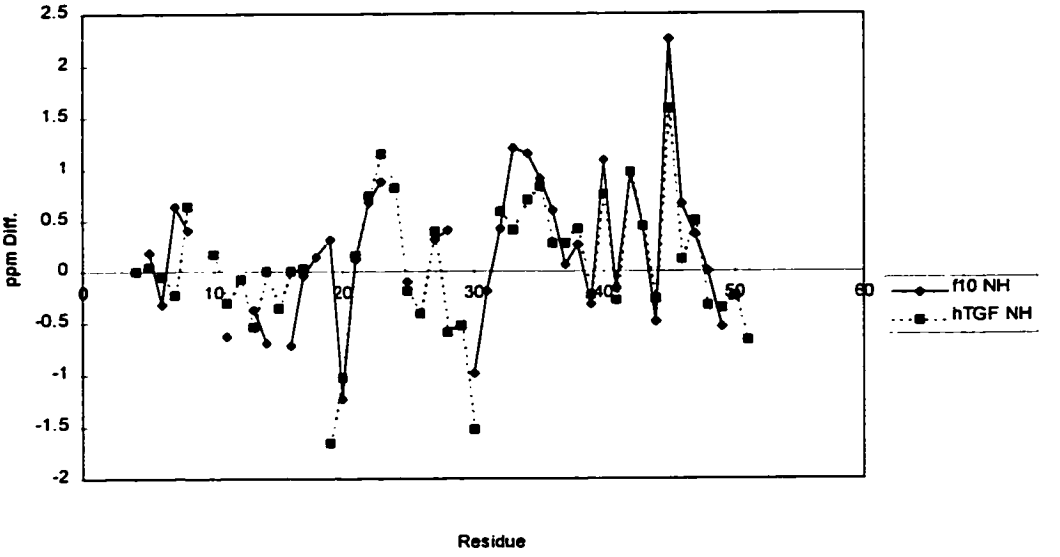
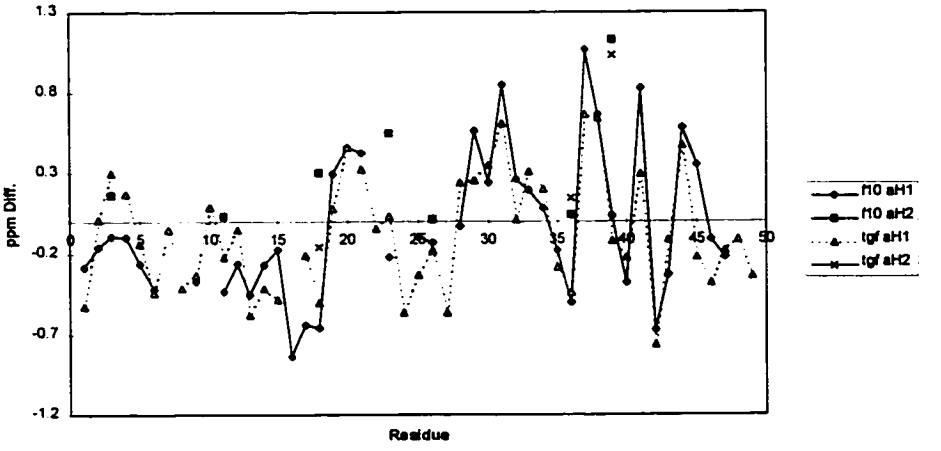


Figure 4.7.6. Conformation dependent chemical shifts of FX-EGF_N and hTGFα.





4.8. Function and structure relationships of the EGF-like polypeptides.

Based on the sequence alignment of the four EGF-like polypeptides (see Figure 4.6.2 in Section 4.6), there are five conserved residues (Gly, Gly, Phe/Tyr, Gly, and Glu/Gln as the positions of 16, 32, 33, 35, and 39 respectively in FIX-EGF_N, in addition to the six cysteines. These residues should be important for the structural integrity of the EGF-like structure, due to the fact of all of them are involved in the conserved structure elements. It has been confirmed that the conserved aromatic residue in hEGF is not obligatory for activity (Engler et al, 1991 Engler, D. A., Hauser, M. R., Cook, J. S., and Niyogi, S. K. (1991) Mol. Cell Biol. 11, 2425-2431).

EGF and TGFα do not contain the conserved residues that may be involved in calcium binding in the EGF-like domains in vitamin K-dependent proteins (refer to Section 1.1). The EGF-like domains in vitamin K-dependent proteins do not contain the conserved amino acids in EGF and TGFα that may be involved at the receptor/growth factor interface (Campbell, 1990).

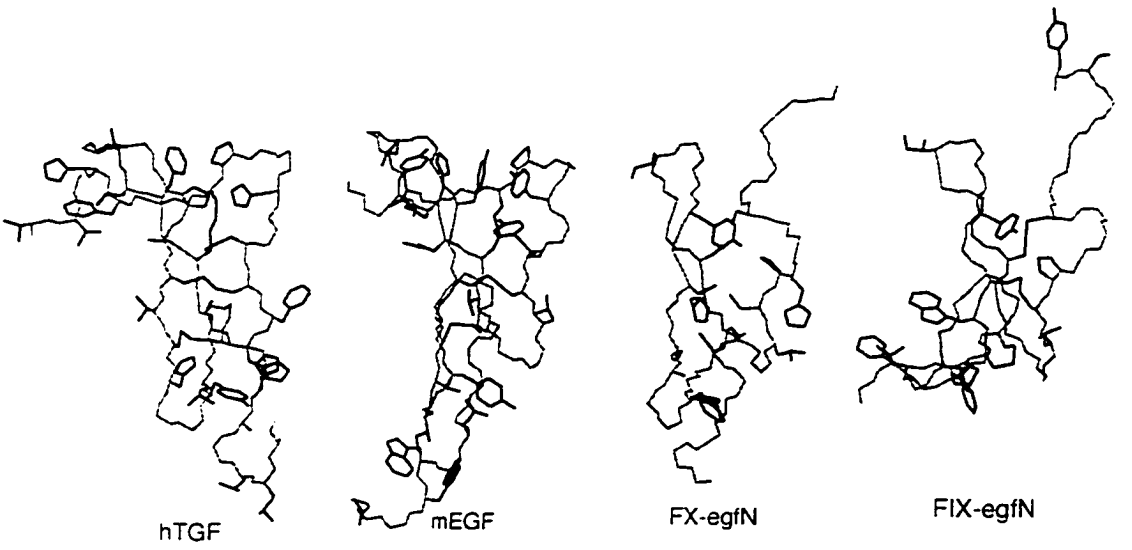
The percentage of hydrophobic residues (Ala, Gly, His, Ile, Leu, Phe, Pro, Trp, and Tyr) in various EGF-like peptides were calculated. The results show that mEGF has 45.3% of hydrophobic residues, hTGFα is 54.0%, FIX-EGF_N is 39.5%, and FX-EGF_N is 40.5%. Both FIX-EGF_N and FX-EGF_N are relatively less hydrophobic than mEGF and hTGFα. Figure 4.8.1 shows that the distribution of hydrophobic residues is quite

different in these EGF-like peptides. There are a lot of hydrophobic residues distributed on the top part of the structures (the N-terminals and the major β-sheet turns) for mEGF and hTGFα, while only few for FIX-EGF_N and FX-EGF_N. Actually the consensus sequence for calcium binding (refer to Section 1.1) is located in this area.

Similarly the EGF-like domains in vitamin K-dependent proteins do not contain the conserved amino acids in EGF and TGFα that may be involved at the receptor/growth factor interface (Campbell, 1990). Even though the mechanism of receptor binding of EGF has not been fully established at the molecular structure level, several amino acid residues have been found to be important for the receptor binding, including Tyr13 (Tadaki and Niyogi (1993)), Tyr22, (His for mEGF) (Campion et al., 1990), Ile23, (Koide et al., 1992), Tyr29 (Engler et al., 1988), Arg41 (Engler, et al., 1992), Arg41 (Engler et al., 1990), Leu47 (Groenen et al., 1994). It has been indicated that the hydrophobic residues on the exposed surface of the β-sheet have an important role in the formation of the active EGF-receptor complex (Campion et al., 1990).

It is likely that the biological activities of the EGF-like domains in vitamin K dependent proteins including FIX-EGF_N and FX-EGF_N are not the same with the growth factor family due to sequence differences, by which could make the surfaces of these polypeptides have a different electronic environments, even though they are structurally similar to each other.

Figure 4.8.1. Hydrophobic residue distributions of the EGF-like peptides.



4.9. Conclusion

For the EGF-like peptides mEGF, hTGFα, FIX-EGF_N and FX-EGF_N, the solution structures are strikingly similar. In both mEGF and hTGFα, there is a three-stranded β-sheet in the N-terminal, while in FIX-EGF_N and FX-EGF_N there is a two strand antiparallel β-sheet in the N-terminal region. In addition, there is a turn or an irregular loop between Cys^1 and Cys^2 in all these EGF-like structures, which sits over the β-sheet. There is a β-turn bended at the beginning of the β-sheet. In the C-terminal region, there is a small "double hairpin" structure in mEGF (Montelione et al., 1992) and hTGFα (Kline et al., 1990) which includes a three-stranded β-sheet, one β-turn, and one five residue turn. In FIX-EGF_N and FX-EGF_N, the homologous segment has similar conformations. However, the third β-strand is not defined in FIX-EGF_N due to the absence of NOE interactions. Glu39 is classified as a β-bulge in FIX-EGF_N. From the result of superimposing the homologous structures, a conserved β-bulge conformation is also observed in mEGF, hTGFα, and FX-EGF_N. The major β-strands of both mEGF and hTGFα are one residue longer than those of FIX-EGF_N and FX-EGF_N. These conserved disulfide bonds in the EGF-like structures are well aligned. In the solution structures of FIX-EGF_N and FX-EGF_N (Ullner et al., 1992), the N-terminals are flexible, similar to the C-terminal of mEGF (Montelione et al., 1992). In contrast, the N- and C-terminals in the solution structure of hTGFα are both not well defined (Kline et al., 1990). The solution structures of FIX-EGF_N

from both this study and Baron's group (Baron et al., 1992) are very similar. However, Baron's structure does not include the second β-strand of the minor β-sheet in the C-terminal due to a shorter sequence.

The sequence alignment as determined by the program BESTFIT shows similar results in the C-terminal of each pair of the EGF-like peptides, but different ones in the N-terminal alignments.

The sequence alignments show that FX-EGF_N and FIX-EGF_N have a higher similarity and identity. Interestingly, they do not show the most similar tertiary structures. In fact, those of mEGF and FIX-EGF_N are close to each other and those of hTGFα and FX-EGF_N are close to each other in the relative orientations of the two β-sheets. It is likely that the different roles of these EGF-like polypeptides in biological function is due to the differences of the electronic environment of the surface.

Chapter V . Calcium binding site of FIX-EGF_N by 2D ^1H NMR pH titration.

5.1. Materials and methods

The peptide sample was synthesized by Y. Yang as described previously (Huang et al., 1989). The peptide was purified by reverse phase HPLC (C18 column) using 0.1% of tri-fluoroacetic acid (TFA) in H_2O (buffer A) and 70% of CH_3CN in H_2O (buffer B) gradient mobile phase. The retention time of the peptide was about 16 minutus. Each fraction was ran by HPLC under the same condition. The pure sample fractions were combined and lyophilized. For calcium free titration experiments, 1.5 mg pure peptide was dissolved in 300 μL of 40 mM NaCl, 92% H_2O/8% D_2O, solution to yield a final concentration of approximately 1 mM. The sample was placed in a micro cell NMR tube for NMR spectroscopy. For titration experiments in the present of calcium, approximately 1.3 mg of pure peptide was dissolved in 300 μL of 20 mM $CaCl_2$, 92% H_2O/8% D_2O, solution and placed in a micro cell NMR tube. The concentration was approximately 0.9 mM. The pH meter was calibrated before each measurement using pH 7.40 and pH 4.00 buffers. The pH values of both +/- calcium solutions were adjusted by the addition of small amount of diluted HCl or NaOH. Measurements of the pH units of the sample were not corrected for deuterium isotope effects and were taken before and after the NMR experiment, with the latter measurement considered more accurate. The pH values of calcium free solution were measured to be 2.4, 3.2, 3.78, 4.2, 4.7, 4.9, 5.0, 5.2, 5.5, 5.7, 6.0, and 6.2; and calcium plus

solution to be 2.3, 2.7, 3.1, 3.7, 4.1, 4.6, 5.1, 5.5, 5.7, 6.1, and 6.2. 1D and 2D NOESY and TOCSY spectra were acquired for each pH value of the solutions. DQF-COSY spectra were acquired for pH 4.2 of calcium free solution and pH 4.6 and pH 5.1 of calcium plus solution.

All 1D and 2D ^{1}H NMR titration experiments were performed on a Varian Unity-plus 500 MHz spectrometer operating at 499.882 MHz at 25 ∓ 0.1 ^{0}C. All spectra were recorded in the phase-sensitive absorption mode using the hyper-complex method (States, et al., 1982). Presaturation was used to suppress the water signal. The carrier frequency was set on the water resonance with a spectral width of 6000 Hz in both dimensions. 2D NOESY and TOCSY data were collected as follows: 2048 complex points in *t2*, and 64 to 100 FID were acquired in t_1 for TOCSY and NOESY; 512 FID for DQF-COSY, and 32 to 400 transients for each FID. TOCSY, NOESY and DQF-COSY 2D NMR experiments were performed according to standard procedures (Wüthrich, 1986). The NOESY spectra were acquired using a mixing of time 200 ms, and TOCSY spectra were acquired with a spin lock time of 65 ms. The 2D NMR data were processed on a SUN workstation using Varian software and on a RISC 6000 workstation using FELIX 2.3 (Biosym/MSI, San Diego).

5.2. Location of calcium binding sites.

The chemical shift of a given nucleus is due to its electronic shielding. The dominant effects on the chemical shift arise from the electron negativity and partial charge of the atoms which are covalently bound, including the inductive effect and/or the

conjugate effect, since these have an immediate influence on the polarization of the electrons surrounding the nucleus observed in the NMR experiment. Other important influences on the chemical shift can be attributed to the relative position of the nucleus to charges on the ionizable groups, π-shell aromatic electrons, and to the collective charge potential of all electrons of the protein in the given conformational state (Forman-Kay, et al., 1992). When a peptide binds calcium, the titration curve might change due to the direct influence of the electrostatic effect, as a result of the conformation change of the peptide arising from binding calcium, or both. For the calcium binding ligand, substantial chemical shift changes could be observed for the alpha proton of calcium binding backbone carbonyl group and the i+1 amide proton and the side chain protons of a calcium binding carboxyl ligand (Akke, et al., 1991; Selander-Sunnerhagen, et al., 1992). Thus pH titrations in the presence and absence of calcium were performed and monitored using 2D-NMR techniques for the study of the calcium binding sites of FIX-EGF_N.

5.3. The chemical shift measurements of the resonances of the N-terminal EGF-FIX_N in the absence and presence of calcium solutions at different pH.

The sequence-specific resonance assignment at pH 4.2 for the N-terminal EGF-FIX_N has been completed (Huang et al., 1991). The sequence specific chemical shifts of protons at other pH values without calcium were measured by starting at a pH value close to 4.2, monitoring TOCSY, NOESY, and/or COSY crosspeaks based on the sequence-specific assignment at pH 4.2. Then

The sequence-specific assignment of spectra at increasing or decreasing pH were performed by tracking the cross-peaks assigned at the previous pH. The sequential NOE interactions were identified where it was necessary to aid the sequence-specific assignment.

There is no significant changes for the relative positions of the total correlation cross-peaks of each spin-system in TOCSY and the NOE cross-peaks in NOESY with calcium at pH 4.1 compared these without calcium at pH 4.2 (Figures 5.3.1 to 5.3.4). In fact most of them are identical. At the lower value of pH, the chemical shifts are even closer (Figure 5.3.5, and 5.3.6). This suggests that at low pH calcium may not be bound to the peptide. Based on this fact, the sequence-specific assignment with calcium are achieved according to the assignment made at about the same pH as the apo form at pH 4.2. The sequential NOEs were identified to confirm assignments as needed.

Figure 5.3.1. A portion of the TOCSY contour plot in 90% H_2O/10% D_2O in the absence of calcium at pH 4.2.

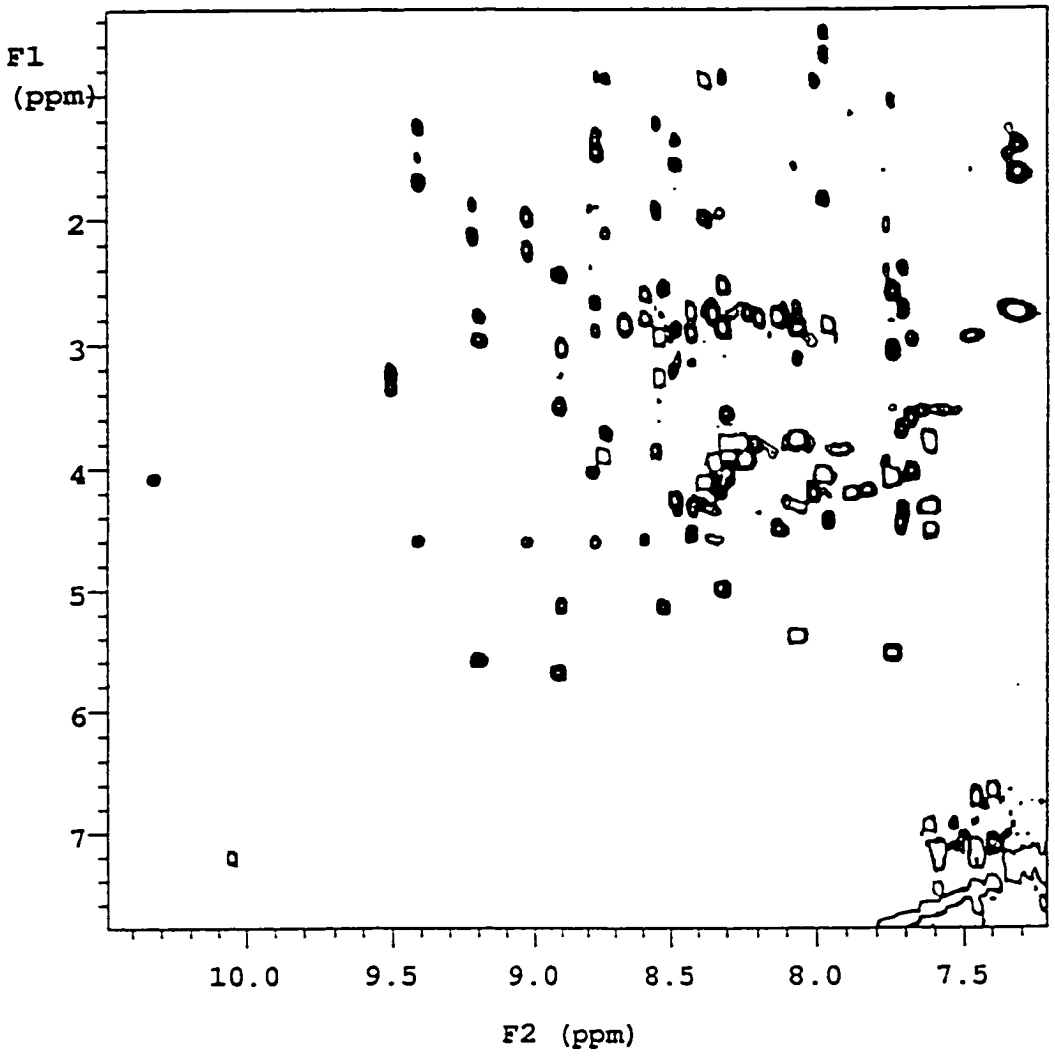


Figure 5.3.2. A portion of the TOCSY contour plot in 90% H_2O/10% D_2O in the presence of calcium at pH 4.1.

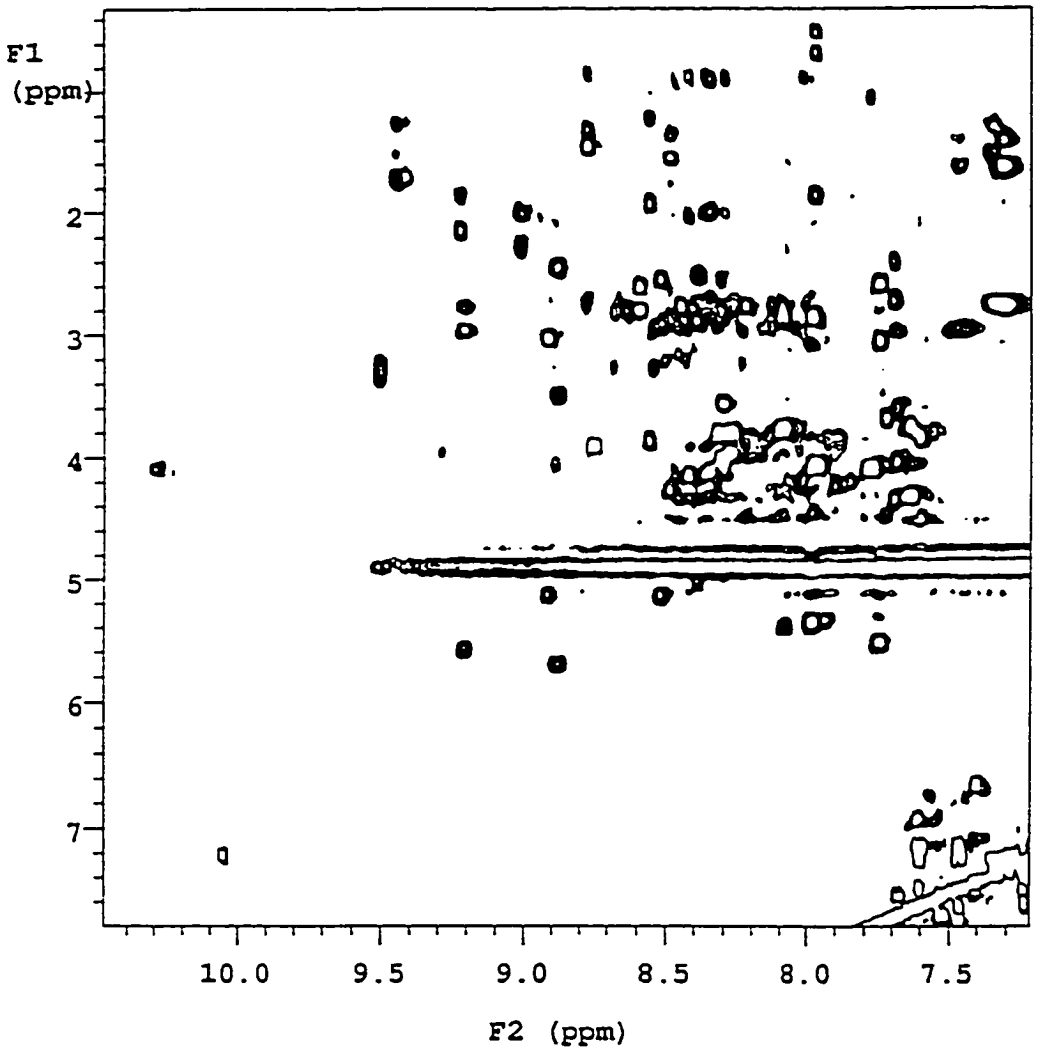


Figure 5.3.3. A portion of the NOESY contour plot in 90% H_2O/10% D_2O in the absence of calcium at pH 4.2.

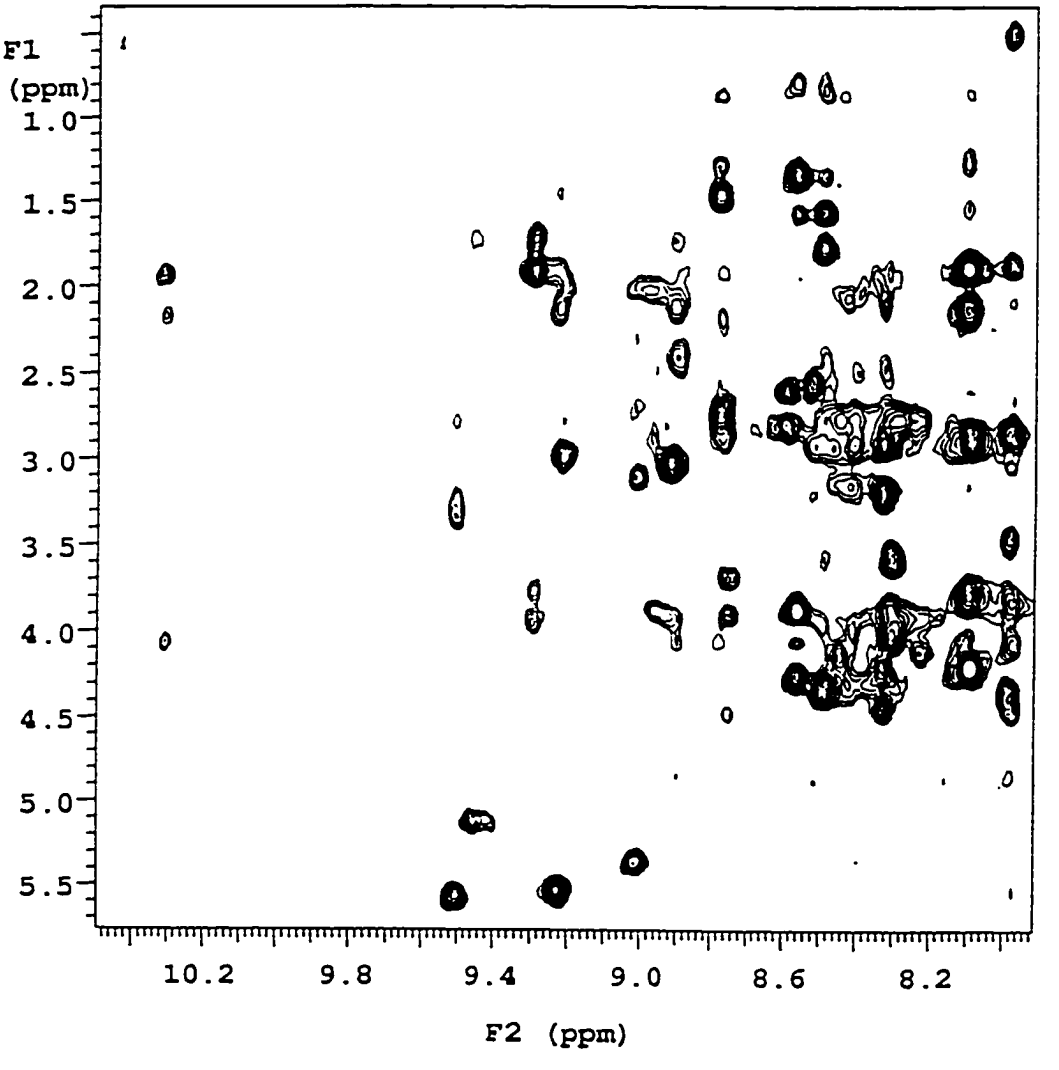


Figure 5.3.4. A portion of the NOESY contour plot in 90% H_2O/10% D_2O in the presence of calcium at pH 4.1.

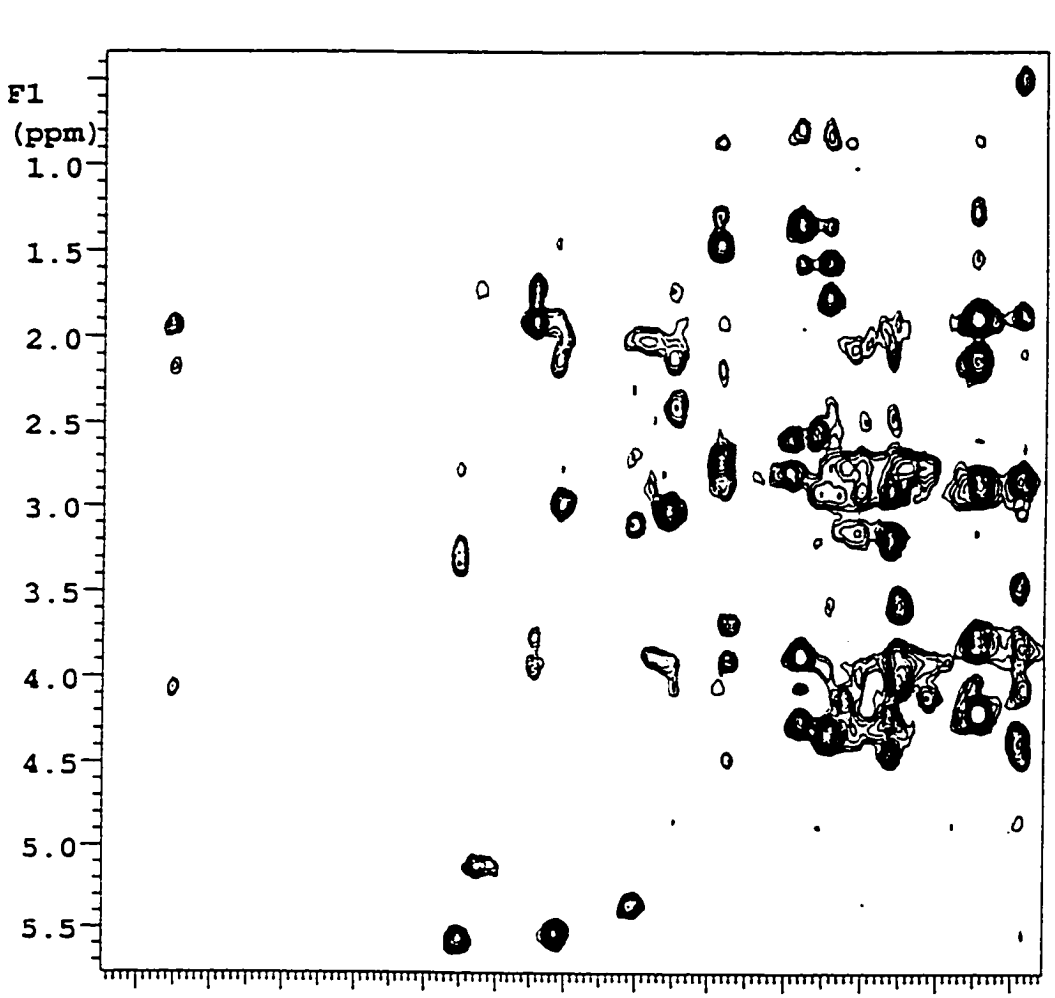


Figure 5.3.5. A portion of the NOESY contour plot in 90% H_2O/10% D_2O in the absence of calcium at pH 3.2.

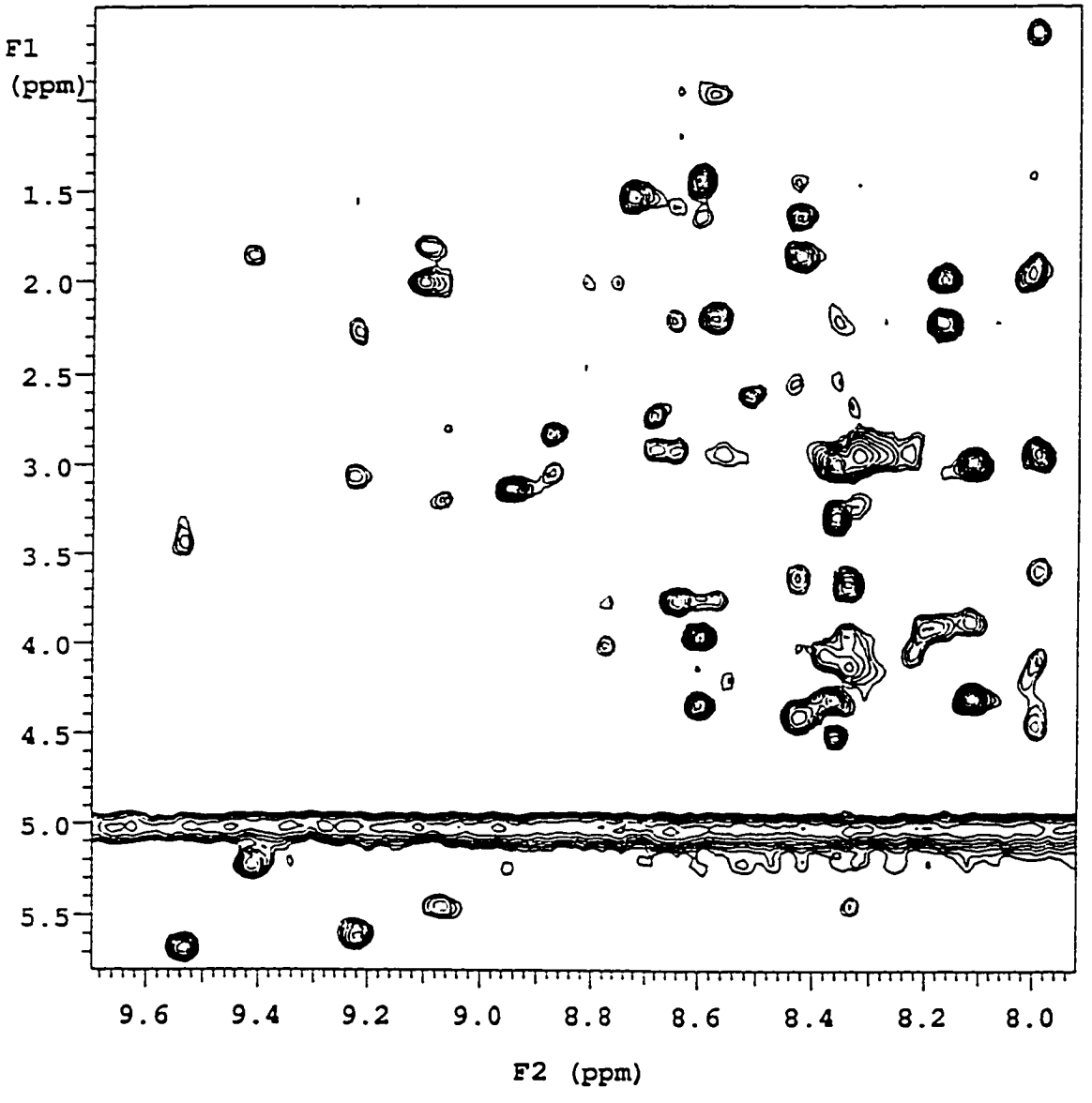
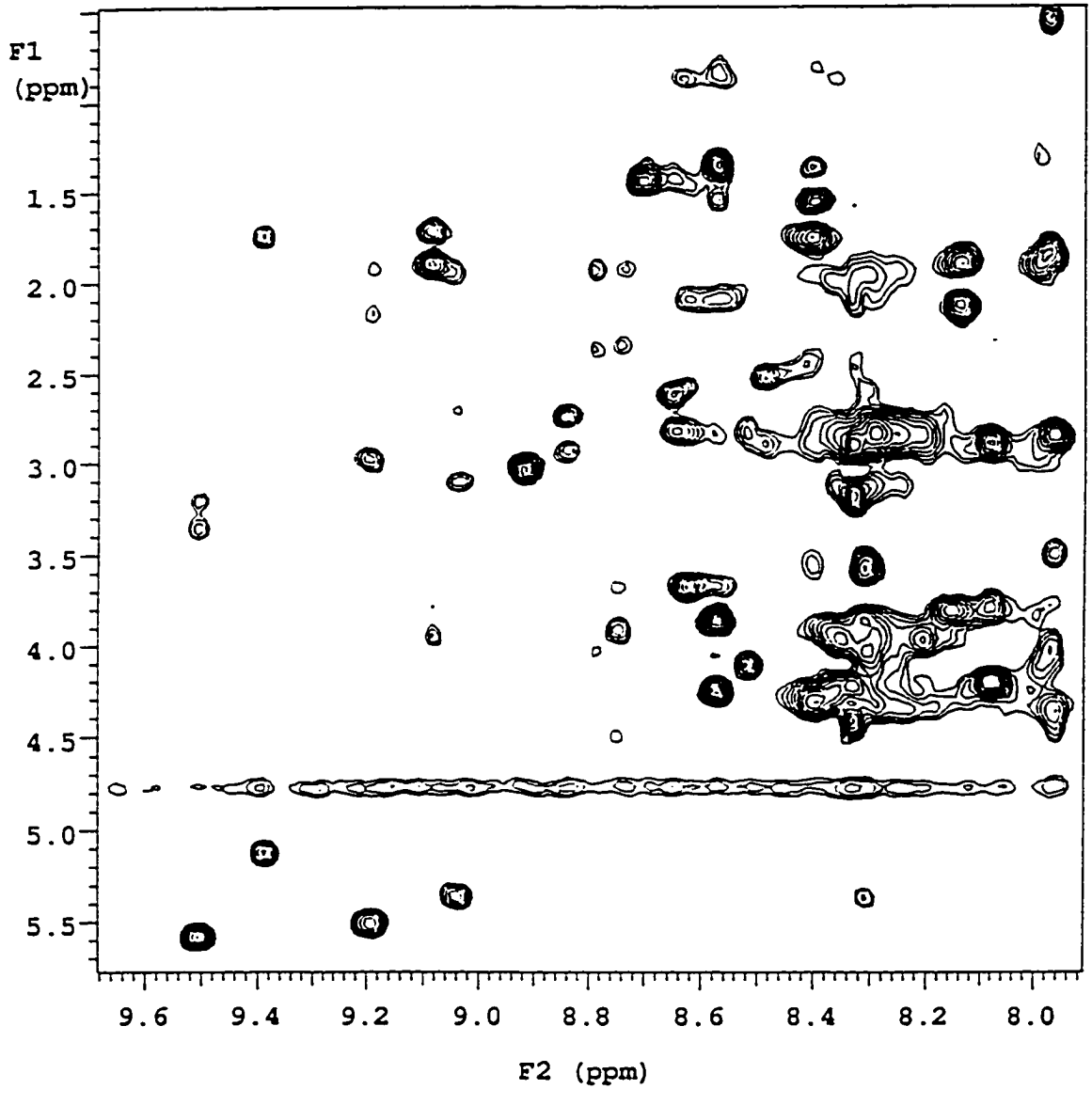


Figure 5.3.6. A portion of the NOESY contour plot in 90% H_2O/10% D_2O in the presence of calcium at pH 3.1.



5.4. Results and discussion.

5.4.1. Titration curves of the apo form.

Significant titration shifts between 0.2 ppm and up to 1.6 ppm for the backbone amide groups over the entire pH range were observed for these residues: Asp3, Cys7, Leu13, Ile22, Asn23, Tyr25, Lys36, Asn37, Glu39, and Asp41. Considerable titration changes for the aliphatic resonances of the backbone between 0.07 ppm and 0.36 ppm) were observed in these residues: Asp5, Glu8, Ile22, Ser24, Gly32, Phe33, Glu39, Asp41, Val42 and Thr43. Much smaller changes were observed for the chemical shifts of protons from the aliphatic sidechains (between 0.09 ppm and 0.3 ppm) of Asp3, Asp5, Cys7, Glu8, Asp20, Asp21, Glu26, Glu34, Glu39, and Asp41. The 10th H of Trp28 are shifted about 0.2 ppm. (Table 5.4.1 and Figures of the titration curves in Appendix 4.).

Conformation isomer was suggested due to the observation of two chemical shifts for the amide proton and aliphatic resonances of Asn14. This amide proton resonance is pH-dependent (Appendix 4.). This residue is part of a turn, Leu13-Gly16 (see Table 4.1.1.). This indicates the turn may have more than one stable conformations.

The chemical shift positions of the side chain amide HNs are relatively pH insensitive. There is no significant shift change in this pH range. The largest shift changes are the side chain NH_2 of Gln6, which is approximately 0.1 ppm (Appendix 4.).

Table 5.4.1. Largest observed chemical shift changes for resonances of FIX-EGF_N over the pH range 2.4-6.2.

proton	residue	Δ ppm
amide HN (≥0.2 ppm)	Asp3	0.25
	Cys7	0.29
	Leu13	0.21
	Ile22	0.46
	Asn23	0.28
	Tyr25	0.34
	Lys36	0.49
	Asn37	1.21
	Glu39	1.58
	Asp41	0.25
aliphatic $C^{\alpha}H$ (≥ 0.07 ppm)	Asp3	0.10
	Glu8	0.07
	Ile22	0.13
	Ser22	0.14
	Gly32	0.09
	Phe33	0.07
	Glu39	0.09
	Asp41	0.11
	Vla42	0.21
	Thr43	0.36
aliphatic side chain CH (≥ 0.09 ppm)	Asp3	0.21
	Asp5	0.15
	Cys7	0.18
	Glu8	0.20
	Asp20	0.23
	Asp21	0.16
	Glu26	0.09
	Glu34	0.11
	Glu39	0.31
	Asp41	0.22
Aromatic H	Trp28(10H)	0.24

The fact that the amide N-H bond is much easier to polarize than the C-H bond may account for the observation that the amide proton chemical shift are more sensitive to changes in the electronic environment due to pH changes. A qualitative evaluation of the amide data reveals that most of the large titration shifts occur in the turns (Leu13, Ile22, Asn23, Lys36, Asn37, and Glu39; refer to section 4.1). It is interesting that although the amide protons of Asn37 and Glu39 have significant pH-dependent chemical shifts, the chemical shifts of the amide proton of Cys38 is pH independent. Since pH-dependent chemical shifts could result from the local conformation change, it is likely that the pH independent behavior of Cys38 is due to its disulfide-bond with Cys29 and it is relative rigid.

All of the acidic residues (aspartic acid, and glutamic acid) show significant titration shifts and the pKas of these residues fall within the range of 3.4 to 5.1.

The pKa of the sidechain amine in lysine in polypeptides is about 10.5 (Cantor and Schimmel, 1980). The ε-NH_2 chemical shifts of Lys 19 and Lys 36 are pH insensitive, indicating they are protonated over the pH range of this study as expected.

5.4.2. Titration curves in the present of calcium and location of the calcium binding sites.

Most of the titration curves of the calcium form are nearly identical to the those of the apo form (Appendix 4). This indicates that binding calcium did not change the overall conformation significantly.

There are a few of residues which contain titration curves that are distinguished from these of the apo form at the higher values of pH. They are the amide HN and $C^{\beta}H$ of Asp3; $C^{\alpha}H$ of Gly4; the amide H of Asp5; the amide protons of both backbone and side chain of Gln6; the amide HN of Glu8, the amide HN and $C^{\beta}H$ of Asp20; the amide HN and $C^{\beta}H$ of Asp21; the $C^{\beta}H$ of Asn23; and the amide H and aromatic 2, 6H of Tyr25 (Appendix 4).

Since the chemical shift positions of the alpha proton of Gly4 and the amide proton of Asp3 exhibit a calcium-dependent downfield shift at the higher pH range, the carbonyl backbone of Gly4 is apparently one of the calcium binding ligands (Appendix 4.). The fact that the chemical shifts of sidechain $C^{\beta}H$ of Asp3 and the side chain amide proton of Gln6 exhibit a calcium-dependent downfield shifts suggests that the sidechain carboxylate group of Asp3 and the sidechain amide groups of Gln6 are bound to calcium (Appendix 4). However the chemical shifts of $C^{\gamma}H$ of Gln6 did not show significant differences between the apo and calcium forms (Appendix 4). Since the influences on chemical shift changes are complex, as discussed in section 5.2, the sidechain amide group of Gln6 is still a possible calcium binding ligand.

The chemical shifts of the amide protons of Gln6 and Asp21 are considerably downfield shifted in the present of calcium. Therefore it is possible that backbone carbonyls of the immediately preceding residues, Asp5 and Asp20 respectively, are involved in calcium binding. However, the possibility could not be clarified from the alpha proton resonances of these two residues, because their titration data were

incomplete (Appendix 4).

The chemical shift position of one sidechain proton of Asp20 in the calcium form is downfield shifted, while the titration curve of another one is upfield shifted relative to the chemical shifts exhibited by the apo form (Appendix 4). It is possible that these chemical shift changes are due to calcium binding, local conformational changes, or both.

The titration curve of one β-proton of Asp21 in the calcium form is similar to that of the apo form, while another one is upfield shifted (Appendix 4). The chemical shift of the degenerate backbone amide protons of Glu8 is upfield shifted 0.4 ppm in the presence of calcium. The titration curves of one β-proton of Asp21, the degenerate β-protons of Asn23, the amide proton and the 2, 6 aromatic protons of Tyr25 are also upfield shifted in the presence of calcium (Appendix 4). Since the electron density of these protons would be decreased as the result of calcium binding, it is expected that calcium binding would result in a downfield shift. These suggest that the sidechains of Asp21 and Asn23, the backbone carbonyl of Cys7, and the hydroxyl group of Tyr25 may not be involved in calcium binding.

The possible calcium binding site was inferred from the solution structure of FIX-EGF_N in conjunction with the pH titration data. Figure 5.4.1 shows Asp3, Gly4, Gln6, and Asp20 form a potential calcium binding site between the N-terminal and the beginning of the major β-sheet turn. The sidechain carboxyl groups of Asp3, Gln6, and Asp20 and the

backbone carbonyl of Gly4 are oriented toward the center of this calcium binding site. They are possibly involved in calcium binding. The backbone carbonyl of Asp5 is oriented away from the proposed calcium binding site, and therefore it is less likely to be involved in calcium binding. For the same reason, the backbone carbonyl of Asp20 is less likely to be a ligand. (Figure 5.4.1).

Figure 5.4.1. The possible candidates for calcium ligands in the solution structure of FIX-EGF_N. Beginning with the N-terminal on the top, the residues shown here are Asp3 (sidechain), Gly4 (backbone carbonyl), Asp5 (backbone carbonyl), Gln6 (sidechain), and Asp20 (backbone carbonyl and sidechain), respectively.

5.4.3. Discussion.

Potential calcium-ligands of FIX-EGF_N have been investigated in this study. The conclusion is similar to that reported in the recent crystallographic analysis of the Ca^{2+} binding EGF-like domain from human coagulation factor IX (Rao et al., 1995) (Table 5.4.2.). The results of Rao's study show that the geometry of the Ca^{2+} co-ordination is pentagonal bipyramidal, forming with seven ligands. Six of the ligands come from one molecule of the peptide, four are supplied by the side chain carboxylates of Asp3 and Asp20 and the side chain amide group of Gln6, and two are donated by the backbone carbonyl oxygen atoms of Gly4 and Asp21. The seventh ligand is the sidechain of Asn23, which belongs to another molecule.

The calcium-bound solution structure of FX-EGF_N has been studied by Selander-Sunnerhagen et al (Selander-Sunnerhagen et al., 1992). The calcium biding site was identified from the structure, in which there is a cavity of appropriate size for a calcium ion is arranged by four oxygens. They report that the backbone carbonyls of Gly47 and Gly64, the side chains of Gln49 and erythro-β-hydroxyasparic acid (Hya)63 are well defined-ligands, the sidechain of Asp46 is also possibly a calcium binding ligand (Table 5.4.2.).

Table 5.4.2. Summery of Calcium binding site studies of FIX-EGF_N and FX-EGF_N.

Calcium binding ligands					
FIX-EGF_N				FX-EGF_N	
NMR-pH Titration[a]		Crystal structure[b]		Solution structure[c]	
Residue	Group	Residue	Group	Residue	Group
Asp3	sidechain	Asp3	sidechain	Asp46	sidechain?
Gly4	backbone	Gly4	backbone	Gly47	backbone
Gln6	sidechain	Gln6	sidechain	Gln49	sidechain
Asp20	sidechain	Asp20	sidechain	Hya63	sidechain
		Asp21	backbone	Gly64	backbone

[a]Gong, in this thesis study.
[b]Rao, et al., 1995.
[c]Selander-Sunnerhagen, et al., 1992.

In the N-terminal EGF-like domain of factor IX, there are three aspartic acid residues, which are among the consensus sequence Asp/Asn, Asp/Asn, Gln/Glu, Asp/Asn, Tyr/Phe in the EGF-like polypeptides, corresponding to the positions of 47, 49, 50, 64, and 69 in human factor IX (Cook et al., 1987; Rees et al., 1988, Handford et al., 1991). Studies demonstrated the importance of the three aspartic acid residues, 47, 49, and 64 in human factor IX (Handford, et al., 1991; Mayhew, et al., 1992). Mutant Asp47/Asn in EGF-FIX_N reduced calcium binding dramatically. Mutant Asp64/Asn lost calcium affinity. However for mutant Asp49/Asn, there is only a relative small change in the K_d value for calcium binding. The glutamine at position 50 in factor IX is strictly conserved in a number of related coagulatant

proteins, including human factor IX, bovine factor IX, canine factor IX, human factor VII, and human factor X, as well as human protein C. It may be inferred that the residue is of critical importance. The mutation of Gln50/Glu has been studied. It shows that the affinity of EGF-FIX_N domain for calcium biding is doubled (Handford, et al., 1991), which indicates this residue is directly involved in calcium biding. Our NMR titration and the recent crystal structure study (Rao, et al., 1995) are consistent with these mutant study results. Our study shows that the side chain of Asp3 (47), Gln6 (50), and possibly of Asp20 (64) are the ligands among the five consensus residues, while the crystal structure study confirms that the side chains of Asp3 (47), Gln6 (50) and Asp20 (64) are ligands. Mutations altering calcium binding consensus residues in the FIX-EGF_N cause Hemophilia B (Davis, et al., 1987; Winship, et al., 1991; McCord, et al., 1990; Lozier, et al., 1990). Hemophilia B results from the lack of factor IX activity and accounts for 10% to 15% of congenital coagulant factor deficiencies.

Table 5.4.3 summarized mutations in the first EGF-like domain of factor IX studied either in patient or as isolated EGF-like domains. The reduced biological activities of these mutations can result from the loss of calcium ligands, which affect the affinity of calcium binding. When a calcium binding Asp residue is mutated to Asn, the negative charge is lost, and the Ca^{2+} binding affinity should decrease. Except the mutants of Asp3, Gln6, and 20, presuming cause the loss of a calcium ligand, the decreased biological activity of the mutant Pro11/Ala can result from the effects of an altered tertiary

structure of the domain. It has been observed in this study that Pro11 has a strong hydrophobic interaction with Tyr25 (see Figure 4.2.1 in Section 4.2), which is among the consensus residues. The mutagenic studies showed that mutant Tyr25/Val caused a reduction in the yield of correctly disulphide bonded material for clotting activity of recombinant factor IX (Hughes et al., 1993). Both of the mutations confirms that residue Pro11 and Tyr25 plays an important role in the tertiary structure. For the consensus residue Asp5, the side chain of this residue forms a crucial H-bond with the backbone amide nitrogen of Cys7, as detected in the crystal structure of FIX-EGF_N (Rao et al., 1995). This residue may play an indirect role on calcium biding. The influence of biological activity of mutant Asp5/Glu might be due to the loss of stabilizing H-bond (Rao et al., 1995), leading to a lowered calcium binding affinity due to local structural changes.

The most likely explanation for the recurrence of the multiple Ca^{2+} binding EGF-like domains within proteins is that they are involved in Ca^{2+} mediated protein-protein interactions (Rao et al., 1995). It is interesting that the recent crystal structure study of FIX-EGF_N (46-84) shows that the seventh ligand, Asn58, bounded with Ca^{2+}, is from a second molecule (Rao et al., 1995). It provides considerable evidence for the Ca^{2+} mediated protein-protein interaction mechanism. In both factor IX and X, a γ-carboxyglutamyl (Gla) domain is N-terminal to the EGF-like domain, linked to it by a short hydrophobic region (Stenflo, 1991). Calcium binding to the isolated EGF-like domains from factors IX and X is 20- to 40-fold reduced compared with the high

affinity binding exhibited by the native protein (Handford et al., 1991; Persson et al., 1989). It was suggested that the carboxylate groups of Gla-40 is calcium bound to the adjacent EGF-like domain based on modeling (Rao et al., 1995). A recent ^{1}H-NMR structural determination of the Gla-EGF domain from bovine factor X, performed in the presence of Ca^{2+}, shows that the Gla domain folds over the EGF domain in region of the Ca^{2+} binding site (Selander-Sunnerhagen, 1994). These are consistent with the calcium intermediate protein-protein interaction mechanism.

In the summery, the titration study indicates that the most sensitive protons versus pH are within the turns. The sidechains of Asp3 and Gln6 and the backbone carbonyl group of Gly4 are clarified as calcium ligands from the titration data. The possibility of the sidechain of Asp20 for calcium ligand is ambiguous due to the complicated titration curve. The backbone carbonyls of Asp5, and Asp20 for calcium ligands are inconclusive due to some insufficient pH titration data. The potential calcium ligands are inferred from consideration of the solution structure of FIX-EGF_N in conjunction with the pH titration data. A visual inspection is supported for the sidechains of Asp20 to be calcium ligands, but for the backbone carbonyls of Asp5 and Asp20. The results of calcium binding site of FIX-EGF_N from this study and the crystal structure study (Rao et al., 1995) are similar. The potential calcium biding site of FIX-EGF_N from both of these studies is similar to that of FX-EGF_N (Selander-Sunnerhagen et al., 1992). The involvment of Asp3, Gln6, and Asp20 in calcium binding is consistent with the finding

that mutation occur at these specific sequenses in hemophilia B patients.

Table 5.4.3. Summery of mutations in the first EGF-like domain of factor IX studied either in patient or in isolated domains.

Mutation	Ca^{2+} affinity (domain)	Source	Biological actitivity(%)	Ref.
Asp3/Glu	Defective	Hemophilia B	1	1,6
Asp3/Asn	-	Domain	N/A	2
Asp3/Gly	Defective	Haemophilia B	10	2-4
Asp5/Asn	+	Domain	N/A	2
Asp5/Glu	Defective	Recombinant	8	5,6
Gln6/Glu	+	Domain	N/A	6
Gln6/Pro	N/A	Haemophilia B	1	7
Pro11/Ala	N/A	Haemophilia B	7-10	8
Gly16/Ser	N/A	Haemophilia B	10	1
Asp20/Asn	-	Haemophilia B	3	1,6
Asp20/Gly	N/A	Haemophilia B	8	8
Asp20/Lys	N/A	Haemophilia B	2	5

1. Winship and Dragon, 1991.
2. Mayhew, et al., 1992.
3. McCord, et al., 1990.
4. Davis et al., 1987.
5. Rees et al., 1988.
6. Handford et al., 1991.
7. Lozier et al., 1990.
8. Green et al., 1989.

Appendix 1. RMSD of each residue of individual conformation on all atoms of the average conformation of all the 14 structures.

Species	1	2	3	4	5	6	7
Residue	RMSD	RMSD	RMSD	RMSD	RMSD	RMSD	RMSD
:TYRN_1	5.14	5.57	7.03	3.86	6.47	3.28	4.60
:VAL_2	5.14	3.66	3.75	9.39	2.14	3.79	3.12
:ASP_3	2.14	3.14	2.94	6.85	5.91	3.46	3.65
:GLY_4	2.86	2.75	1.42	2.52	5.11	3.10	3.32
:ASP_5	2.92	2.07	1.42	2.23	2.33	1.47	3.55
:GLN_6	1.55	1.76	1.60	2.03	1.31	1.14	1.15
:CYS_7	0.82	1.16	0.84	1.49	1.24	1.70	0.93
:GLU_8	1.28	1.60	1.86	1.11	1.67	2.62	1.23
:SER_9	1.72	1.39	0.90	1.44	2.16	1.71	1.51
:ASN_10	2.41	1.44	1.95	1.46	3.50	1.62	1.64
:PRO_11	0.73	0.74	0.86	1.00	1.15	1.13	0.88
:CYS_12	0.80	0.28	0.63	0.40	1.79	1.16	0.89
:LEU_13	1.36	1.28	1.08	1.38	1.30	1.00	1.41
:ASN_14	0.58	0.32	0.75	0.70	0.87	0.38	0.62
:GLY_15	0.65	0.46	0.68	0.54	1.54	1.24	0.80
:GLY_16	0.68	0.35	0.71	0.33	2.27	1.21	0.74
:SER_17	0.61	1.12	0.76	0.59	0.87	1.29	0.70
:CYS_18	0.73	0.90	0.89	1.00	0.88	1.13	0.96
:LYS_19	2.71	3.10	2.20	2.25	1.54	1.93	1.72
:ASP_20	1.51	1.51	1.08	1.98	0.91	1.49	1.49
:ASP_21	2.07	1.00	1.04	1.37	0.98	2.07	1.05
:ILE_22	2.32	1.06	1.49	1.83	1.08	2.78	1.22
:ASN_23	1.29	1.26	2.00	3.25	1.16	1.73	1.00
:SER_24	1.05	1.09	1.57	1.85	0.81	1.26	1.15
:TYR_25	1.29	1.39	1.14	1.84	1.01	1.65	1.52
:GLU_26	2.14	1.89	1.11	2.02	1.02	2.59	1.86
:CYS_27	0.59	0.59	0.65	0.48	0.80	1.35	0.65
:TRP_28	1.03	0.58	0.76	0.92	1.37	1.63	4.38
:CYS_29	0.65	0.44	1.26	0.35	1.02	1.81	0.82
:PRO_30	0.55	0.79	1.76	0.71	0.93	1.53	0.87
:PHE_31	1.24	1.45	2.28	1.29	1.45	1.97	1.71
:GLY_32	0.57	0.86	1.01	0.83	0.79	1.24	1.05
:PHE_33	1.26	1.33	1.66	1.42	1.33	1.88	1.41
:GLU_34	1.49	1.28	0.89	1.02	1.40	1.44	1.66
:GLY_35	0.44	1.70	0.61	0.91	0.63	1.28	0.47
:LYS_36	1.32	1.32	1.60	1.87	2.88	1.69	1.27
:ASN_37	0.32	0.78	0.62	0.87	1.04	0.74	0.38
:CYS_38	0.44	0.46	1.03	0.56	0.73	1.03	0.46
:GLU_39	0.60	0.71	0.63	0.90	0.82	0.55	0.85
:LEU_40	1.11	1.51	1.04	1.52	1.81	0.94	1.23
:ASP_41	0.81	1.10	0.48	1.36	1.16	1.19	1.11
:VAL_42	0.57	0.79	0.91	0.67	0.87	1.22	1.13
:THRC_43	0.64	1.29	1.44	0.95	1.43	1.31	1.32

Species	8	9	10	11	12	13	14
Residue	RMSD	RMSD	RMSD	RMSD	RMSD	RMSD	RMSD
:TYRN_1	2.49	8.58	5.29	2.67	6.72	5.17	9.52
:VAL_2	5.75	5.19	4.87	3.42	4.37	1.70	7.40
:ASP_3	5.43	3.70	6.00	4.21	3.70	4.68	4.71
:GLY_4	2.93	2.21	2.81	2.83	2.32	2.58	4.89
:ASP_5	2.44	4.57	1.58	1.45	1.86	2.66	2.95
:GLN_6	1.97	1.92	1.02	2.08	1.89	1.21	2.98
:CYS_7	1.54	1.90	0.87	1.33	0.87	1.07	2.67
:GLU_8	2.12	2.02	1.75	2.52	1.87	1.59	2.10
:SER_9	1.63	2.17	1.45	2.22	1.63	2.11	1.49
:ASN_10	1.79	2.18	2.59	2.38	1.93	2.33	2.43
:PRO_11	0.80	1.84	1.68	0.65	0.69	0.51	1.34
:CYS_12	0.35	0.94	0.50	0.65	0.51	0.43	0.61
:LEU_13	1.26	1.72	1.11	1.54	1.06	1.00	1.02
:ASN_14	0.73	1.08	0.56	1.00	0.37	0.27	0.56
:GLY_15	1.10	0.95	0.50	1.44	0.55	0.27	1.07
:GLY_16	0.35	0.43	0.50	0.62	0.74	0.26	0.87
:SER_17	0.58	0.41	0.44	0.75	0.88	0.57	1.28
:CYS_18	0.94	1.10	0.61	0.67	0.88	0.59	1.53
:LYS_19	1.77	1.50	2.55	2.10	1.78	1.45	1.68
:ASP_20	0.82	2.42	0.80	1.40	0.72	0.91	2.52
:ASP_21	1.25	2.11	1.04	1.12	0.78	1.41	1.88
:ILE_22	2.06	2.74	1.69	1.34	1.02	1.96	2.89
:ASN_23	1.74	2.54	1.44	1.93	1.11	1.59	3.63
:SER_24	1.85	1.28	1.74	1.09	0.37	0.71	2.38
:TYR_25	1.62	1.36	1.16	1.24	1.31	1.02	1.98
:GLU_26	2.38	1.32	1.66	1.14	1.82	1.22	1.88
:CYS_27	0.75	0.40	0.29	0.46	0.30	0.39	0.62
:TRP_28	0.63	0.85	0.76	0.92	0.95	0.84	1.00
:CYS_29	0.49	0.44	0.37	0.53	0.46	0.39	0.64
:PRO_30	0.36	0.71	0.45	0.24	0.44	0.64	1.38
:PHE_31	1.47	1.53	1.30	1.18	1.29	1.40	2.00
:GLY_32	0.63	0.79	0.55	0.41	0.44	0.68	1.07
:PHE_33	1.20	1.31	1.28	1.20	1.14	1.31	1.30
:GLU_34	1.13	1.25	1.09	1.43	1.41	0.87	1.51
:GLY_35	0.35	1.29	0.79	1.07	0.47	0.39	0.92
:LYS_36	1.59	2.51	1.66	3.36	2.40	1.47	1.95
:ASN_37	1.02	1.51	0.82	0.53	0.47	0.56	0.84
:CYS_38	0.85	1.11	0.42	0.60	0.32	0.32	0.42
:GLU_39	1.03	1.30	0.72	0.68	0.57	0.48	0.48
:LEU_40	1.12	1.05	1.23	1.69	1.38	1.09	1.78
:ASP_41	1.21	0.80	1.54	0.96	0.70	0.81	0.87
:VAL_42	1.04	0.50	0.90	0.52	0.50	0.56	1.20
:THRC_43	1.43	0.89	1.13	0.61	0.53	0.75	1.56

Appendix 2. RMSD of each residue of individual conformation on the backbone of the average conformation of all the 14 structures.

Species	1	2	3	4	5	6	7
Residue	RMSD	RMSD	RMSD	RMSD	RMSD	RMSD	RMSD
:TYRN_1	2.85	4.07	5.08	4.51	4.31	2.55	2.60
:VAL_2	3.45	2.88	3.00	6.85	1.70	2.79	1.89
:ASP_3	2.51	2.34	2.05	5.38	4.60	2.51	2.50
:GLY_4	2.64	2.38	0.91	2.15	4.73	2.79	2.76
:ASP_5	2.25	1.18	0.74	1.17	2.11	1.20	1.08
:GLN_6	1.40	1.17	0.76	1.94	1.06	0.51	0.72
:CYS_7	0.57	1.12	0.91	1.27	1.10	1.41	0.59
:GLU_8	0.75	0.90	1.44	0.85	1.05	1.86	0.77
:SER_9	1.39	0.73	0.69	1.08	1.74	1.52	0.74
:ASN_10	1.30	0.22	1.26	0.90	2.13	1.28	0.43
:PRO_11	0.71	0.46	0.75	0.75	1.30	1.16	0.74
:CYS_12	0.75	0.30	0.61	0.41	1.43	0.99	0.84
:LEU_13	0.65	0.18	0.56	0.37	0.97	0.49	0.70
:ASN_14	0.66	0.25	0.64	0.53	0.96	0.31	0.71
:GLY_15	0.63	0.43	0.65	0.50	1.48	1.12	0.77
:GLY_16	0.65	0.36	0.69	0.31	1.97	1.18	0.75
:SER_17	0.48	0.55	0.65	0.39	0.87	1.18	0.69
:CYS_18	0.38	0.73	0.73	0.64	0.58	0.90	0.69
:LYS_19	0.81	1.07	0.72	1.05	0.62	0.83	0.42
:ASP_20	1.31	1.12	0.90	1.50	0.69	1.13	0.47
:ASP_21	1.93	0.87	0.90	1.49	0.92	1.86	0.77
:ILE_22	1.80	0.86	1.14	1.82	0.78	1.78	0.49
:ASN_23	0.94	0.91	1.65	2.34	0.63	0.84	0.65
:SER_24	0.75	0.99	1.08	1.92	0.56	0.89	0.92
:TYR_25	1.10	0.95	0.64	1.38	0.46	1.26	0.90
:GLU_26	1.25	0.89	0.55	0.76	0.75	1.44	1.02
:CYS_27	0.49	0.63	0.62	0.49	0.78	1.17	0.66
:TRP_28	0.41	0.59	0.41	0.34	0.79	1.12	0.89
:CYS_29	0.55	0.46	0.66	0.27	0.92	1.34	0.83
:PRO_30	0.55	0.71	1.30	0.61	0.72	1.12	0.94
:PHE_31	0.53	0.88	1.53	0.72	0.72	1.17	1.12
:GLY_32	0.56	0.84	1.01	0.80	0.74	1.24	1.05
:PHE_33	0.54	0.42	0.53	0.59	0.37	1.15	0.86
:GLU_34	0.37	0.60	0.32	0.62	0.74	0.80	0.59
:GLY_35	0.40	1.30	0.56	0.87	0.63	0.86	0.43
:LYS_36	0.68	0.92	0.85	1.04	0.87	0.55	0.37
:ASN_37	0.31	0.76	0.58	0.77	0.82	0.49	0.31
:CYS_38	0.27	0.45	0.39	0.58	0.55	0.28	0.37
:GLU_39	0.37	0.63	0.46	0.72	0.48	0.41	0.57
:LEU_40	0.50	0.86	0.33	0.86	0.84	0.63	0.74
:ASP_41	0.53	0.85	0.22	0.97	0.92	1.03	0.91
:VAL_42	0.57	0.86	0.73	0.79	0.97	1.21	1.12
:THRC_43	0.57	1.13	0.99	0.96	1.32	1.25	1.25

Species	8	9	10	11	12	13	14
Residue	RMSD	RMSD	RMSD	RMSD	RMSD	RMSD	RMSD
:TYRN_1	2.57	7.07	3.33	2.27	3.94	3.28	7.10
:VAL_2	4.41	4.80	3.62	2.26	2.84	1.28	5.69
:ASP_3	4.05	2.56	4.62	3.09	2.78	3.26	4.00
:GLY_4	2.57	1.85	2.73	2.59	2.03	2.29	4.53
:ASP_5	1.89	3.31	1.40	1.20	1.29	1.13	2.98
:GLN_6	1.22	2.34	0.78	0.48	0.88	0.64	2.89
:CYS_7	1.36	1.96	0.57	1.13	0.86	0.87	2.53
:GLU_8	1.37	2.10	0.97	2.01	1.04	1.38	2.15
:SER_9	1.44	2.04	1.01	1.85	0.87	1.74	1.51
:ASN_10	1.01	1.80	1.47	1.05	0.93	1.47	1.49
:PRO_11	0.60	1.74	0.99	0.52	0.42	0.59	0.85
:CYS_12	0.30	1.01	0.57	0.64	0.44	0.39	0.59
:LEU_13	0.31	1.00	0.58	0.88	0.35	0.21	0.44
:ASN_14	0.29	1.02	0.54	0.79	0.39	0.18	0.62
:GLY_15	0.85	0.89	0.49	1.22	0.53	0.22	1.01
:GLY_16	0.30	0.38	0.47	0.59	0.71	0.21	0.89
:SER_17	0.54	0.36	0.41	0.50	0.71	0.30	1.06
:CYS_18	0.61	0.82	0.35	0.26	0.73	0.24	1.27
:LYS_19	0.62	1.07	0.28	0.49	0.55	0.31	1.48
:ASP_20	0.55	1.92	0.44	0.65	0.56	0.71	1.97
:ASP_21	1.01	2.22	0.76	0.97	0.63	1.22	1.93
:ILE_22	1.26	2.36	0.76	1.15	0.65	1.01	2.43
:ASN_23	1.25	2.20	0.63	1.29	0.45	0.38	2.81
:SER_24	1.44	1.40	0.44	1.00	0.24	0.53	2.27
:TYR_25	1.14	0.86	0.40	0.94	0.21	0.61	1.61
:GLU_26	1.07	0.57	0.31	0.61	0.20	0.66	1.05
:CYS_27	0.79	0.24	0.21	0.30	0.27	0.33	0.58
:TRP_28	0.69	0.18	0.25	0.53	0.38	0.27	0.51
:CYS_29	0.46	0.37	0.29	0.50	0.43	0.32	0.64
:PRO_30	0.39	0.66	0.34	0.26	0.48	0.56	1.04
:PHE_31	0.59	0.89	0.47	0.32	0.57	0.71	1.35
:GLY_32	0.60	0.76	0.53	0.40	0.44	0.67	1.06
:PHE_33	0.40	0.54	0.48	0.53	0.35	0.54	0.74
:GLU_34	0.52	0.65	0.53	0.91	0.47	0.26	0.69
:GLY_35	0.24	1.23	0.73	1.00	0.33	0.30	0.87
:LYS_36	0.57	1.37	0.58	0.69	0.55	0.34	0.89
:ASN_37	0.70	1.11	0.44	0.53	0.38	0.29	0.58
:CYS_38	0.66	0.90	0.39	0.45	0.22	0.21	0.29
:GLU_39	0.61	1.11	0.53	0.45	0.20	0.31	0.35
:LEU_40	0.75	0.78	0.70	1.20	0.20	0.41	0.62
:ASP_41	0.99	0.63	1.03	0.77	0.36	0.44	0.78
:VAL_42	1.07	0.49	0.97	0.59	0.45	0.54	1.12
:THRC_43	1.32	0.83	1.05	0.64	0.45	0.69	1.37

Appendix 3. Constraints for solution structure determination of FIX-EGF_N.

```
#remote_prochiral_center
1:VAL_42:HG2*   1:VAL_42:HG1*   1:VAL_42:CG2   1:VAL_42:CG1   1:VAL_42:CB
#chiral
1:TYRN_1:CA       S
1:VAL_2:CA        S
1:ASP_3:CA        S
1:ASP_5:CA        S
1:GLN_6:CA        S
1:CYS_7:CA        R
1:GLU_8:CA        S
1:SER_9:CA        S
1:ASN_10:CA       S
1:PRO_11:CA       S
1:CYS_12:CA       R
1:LEU_13:CA       S
1:ASN_14:CA       S
1:SER_17:CA       S
1:CYS_18:CA       R
1:LYS_19:CA       S
1:ASP_20:CA       S
1:ASP_21:CA       S
1:ILE_22:CA       S
1:ILE_22:CB       S
1:ASN_23:CA       S
1:SER_24:CA       S
1:TYR_25:CA       S
1:GLU_26:CA       S
1:CYS_27:CA       R
1:TRP_28:CA       S
1:CYS_29:CA       R
1:PRO_30:CA       S
1:PHE_31:CA       S
1:PHE_33:CA       S
1:GLU_34:CA       S
1:LYS_36:CA       S
1:ASN_37:CA       S
1:CYS_38:CA       R
1:GLU_39:CA       S
1:LEU_40:CA       S
1:ASP_41:CA       S
1:VAL_42:CA       S
1:THRC_43:CA      S
1:THRC_43:CB      R
```

#distance

1:SER_17:O	1:TRP_28:HN	1.700	2.300	100.00	100.00	1000.000
1:SER_17:O	1:TRP_28:N	2.700	3.300	100.00	100.00	1000.000
1:TRP_28:O	1:SER_17:HN	1.700	2.300	100.00	100.00	1000.000
1:TRP_28:O	1:SER_17:N	2.700	3.300	100.00	100.00	1000.000
1:LYS_19:O	1:GLU_26:HN	1.700	2.300	100.00	100.00	1000.000
1:LYS_19:O	1:GLU_26:N	2.700	3.300	100.00	100.00	1000.000
1:GLU_26:O	1:LYS_19:HN	1.700	2.300	100.00	100.00	1000.000
1:GLU_26:O	1:LYS_19:N	2.700	3.200	100.00	100.00	1000.000
1:PRO_30:O	1:PHE_33:HN	1.700	2.300	100.00	100.00	1000.000
1:PRO_30:O	1:PHE_33:N	2.700	3.300	100.00	100.00	1000.000
1:GLY_32:O	1:VAL_42:HN	1.700	2.300	100.00	100.00	1000.000
1:GLY_32:O	1:VAL_42:N	2.700	3.300	100.00	100.00	1000.000
1:GLU_34:O	1:GLU_39:HN	1.700	2.300	100.00	100.00	1000.000
1:GLU_34:O	1:GLU_39:N	2.700	3.300	100.00	100.00	1000.000
1:GLU_34:O	1:LEU_40:HN	1.700	2.300	100.00	100.00	1000.000
1:GLU_34:O	1:LEU_40:N	2.700	3.300	100.00	100.00	1000.000
1:LEU_13:O	1:GLY_16:HN	1.700	2.300	100.00	100.00	1000.000
1:LEU_13:O	1:GLY_16:N	2.700	3.300	100.00	100.00	1000.000

#NOE_distance

1:ASN_10:HA	1:ASN_10:HB*	1.800	3.000	3.000	30.00	30.00	1000.000
1:ASN_10:HA	1:PRO_11:HG*	1.800	5.000	4.500	30.00	30.00	1000.000
1:ASN_10:HA	1:PRO_11:HD*	1.800	3.500	3.000	30.00	30.00	1000.000
1:ASN_10:HN	1:ASN_10:HA	1.800	2.900	2.900	30.00	30.00	1000.000
1:ASN_10:HN	1:ASN_10:HB*	1.800	3.700	3.200	30.00	30.00	1000.000
1:ASN_10:HN	1:GLU_8:HA	1.800	4.650	4.650	30.00	30.00	1000.000
1:ASN_10:HN	1:PRO_11:HD*	1.800	4.200	3.200	30.00	30.00	1000.000
1:ASN_10:HN	1:SER_9:HA	1.800	3.300	3.300	30.00	30.00	1000.000
1:ASN_10:HN	1:SER_9:HN	1.800	3.800	3.800	30.00	30.00	1000.000
1:ASN_10:HD21	1:ASN_10:HB*	1.800	3.300	2.800	30.00	30.00	1000.000
1:ASN_14:HA	1:ASN_14:HB*	1.800	3.000	3.000	30.00	30.00	1000.000
1:ASN_14:HD2*	1:LEU_13:HBS	1.800	5.300	4.700	30.00	30.00	1000.000
1:ASN_14:HD2*	1:LEU_13:HBR	1.800	5.300	4.700	30.00	30.00	1000.000
1:ASN_14:HD2*	1:ASN_14:HB*	1.800	5.000	4.000	30.00	30.00	1000.000
1:ASN_14:HN	1:LEU_13:HD*	1.800	6.300	4.300	30.00	30.00	1000.000
1:ASN_14:HN	1:LEU_13:HG	1.800	4.500	4.500	30.00	30.00	1000.000
1:ASN_14:HN	1:ASN_14:HA	1.800	2.700	2.700	30.00	30.00	1000.000
1:ASN_14:HN	1:ASN_14:HB*	1.800	4.800	4.300	30.00	30.00	1000.000
1:ASN_14:HN	1:GLY_15:HN	1.800	3.100	3.100	30.00	30.00	1000.000
1:ASN_14:HN	1:GLY_16:HN	1.800	5.000	5.000	30.00	30.00	1000.000
1:ASN_14:HN	1:LEU_1:HA	1.800	3.100	3.100	30.00	30.00	1000.000
1:ASN_14:HN	1:LEU_13:HBR	1.800	3.500	3.500	30.00	30.00	1000.000
1:ASN_14:HN	1:LEU_13:HBS	1.800	3.200	3.200	30.00	30.00	1000.000
1:ASN_14:HN	1:GLU_39:HA	1.800	5.000	5.000	30.00	30.00	1000.000
1:ASN_14:HD2*	1:GLU_39:HA	1.800	3.800	3.300	30.00	30.00	1000.000
1:ASN_14:HD2*	1:LEU_13:HD*	1.800	7.000	4.500	30.00	30.00	1000.000
1:ASN_23:HA	1:ASN_23:HB*	1.800	3.000	3.000	30.00	30.00	1000.000
1:ASN_23:HA	1:ILE_22:HG2*	1.800	5.500	4.500	30.00	30.00	1000.000
1:ASN_23:HN	1:ILE_22:HG2*	1.800	4.000	3.000	30.00	30.00	1000.000
1:TYRN_1:HD*	1:TYRN_1:HA	1.800	5.300	4.300	30.00	30.00	1000.000

1:ASN_23:HN	1:ASN_23:HB*	1.800	5.000	5.000	30.00	30.00	1000.000
1:ASN_23:HD22	1:ASN_23:HB*	1.800	4.000	3.500	30.00	30.00	1000.000
1:ASN_23:HN	1:ILE_22:HA	1.800	3.600	3.600	30.00	30.00	1000.000
1:ASN_23:HN	1:ILE_22:HB	1.800	3.300	3.300	30.00	30.00	1000.000
1:ASN_23:HN	1:SER_24:HN	1.800	3.100	3.100	30.00	30.00	1000.000
1:ASN_37:HA	1:ASN_37:HBR	1.800	2.700	2.700	30.00	30.00	1000.000
1:ASN_37:HA	1:ASN_37:HBS	1.800	3.100	3.100	30.00	30.00	1000.000
1:ASN_37:HA	1:CYS_12:HA	1.800	4.000	4.000	30.00	30.00	1000.000
1:ASN_37:HA	1:CYS_27:HBR	1.800	4.000	4.000	30.00	30.00	1000.000
1:ASN_37:HA	1:CYS_27:HBS	1.800	4.500	4.500	30.00	30.00	1000.000
1:ASN_37:HBR	1:GLU_39:HB*	1.800	4.700	4.200	30.00	30.00	1000.000
1:ASN_37:HBR	1:LEU_13:HD*	1.800	6.200	4.200	30.00	30.00	1000.000
1:ASN_37:HBS	1:LEU_13:HD*	1.800	5.200	3.200	30.00	30.00	1000.000
1:ASN_37:HBR	1:LEU_13:HG	1.800	4.000	4.000	30.00	30.00	1000.000
1:ASN_37:N	1:LEU_13:HD*	1.800	6.300	4.300	30.00	30.00	1000.000
1:ASN_37:HN	1:ASN_37:HA	1.800	3.100	3.100	30.00	30.00	1000.000
1:ASN_37:HN	1:ASN_37:HBR	1.800	4.300	4.300	30.00	30.00	1000.000
1:ASN_37:HN	1:ASN_37:HBS	1.800	3.700	3.700	30.00	30.00	1000.000
1:ASN_37:HD21	1:ASN_37:HBR	1.800	3.200	3.200	30.00	30.00	1000.000
1:ASN_37:HD21	1:ASN_37:HBS	1.800	3.200	3.200	30.00	30.00	1000.000
1:ASN_37:HD22	1:ASN_37:HBS	1.800	4.500	4.500	30.00	30.00	1000.000
1:ASN_37:HN	1:CYS_38:HN	1.800	4.000	4.000	30.00	30.00	1000.000
1:ASN_37:HN	1:LYS_36:HA	1.800	3.000	3.000	30.00	30.00	1000.000
1:ASN_37:HN	1:LYS_36:HB*	1.800	5.000	4.000	30.00	30.00	1000.000
1:ASN_23:HD22	1:ILE_22:HD1*	1.800	5.800	4.800	30.00	30.00	1000.000
1:ASP_20:HA	1:ASP_20:HB*	1.800	3.000	2.600	30.00	30.00	1000.000
1:ASP_20:HA	1:TYR_25:HBR	1.800	3.200	3.200	30.00	30.00	1000.000
1:ASP_20:HN	1:ASP_20:HA	1.800	3.100	3.100	30.00	30.00	1000.000
1:ASP_20:HN	1:LYS_19:HA	1.800	3.400	3.400	30.00	30.00	1000.000
1:ASP_20:HN	1:ASP_20:HB*	1.800	3.500	3.000	30.00	30.00	1000.000
1:ASP_21:HA	1:ASP_21:HBS	1.800	2.700	2.700	30.00	30.00	1000.000
1:ASP_21:HA	1:ASP_21:HBR	1.800	2.700	2.700	30.00	30.00	1000.000
1:ASP_21:HA	1:ILE_22:HG1*	1.800	5.600	4.600	30.00	30.00	1000.000
1:ASP_21:HN	1:SER_24:HN	1.800	4.500	4.500	30.00	30.00	1000.000
1:ASP_21:HN	1:ASP_20:HA	1.800	3.600	3.600	30.00	30.00	1000.000
1:ASP_21:HN	1:TYR_25:HBR	1.800	4.300	4.300	30.00	30.00	1000.000
1:ASP_21:HN	1:ASP_21:HA	1.800	2.900	2.900	30.00	30.00	1000.000
1:ASP_21:HN	1:ASP_21:HBR	1.800	3.000	3.000	30.00	30.00	1000.000
1:ASP_21:HN	1:ASP_21:HBS	1.800	4.200	4.200	30.00	30.00	1000.000
1:ASP_21:HN	1:TYR_25:HA	1.800	4.000	4.000	30.00	30.00	1000.000
1:ASP_21:HN	1:TYR_25:HBS	1.800	4.000	4.000	30.00	30.00	1000.000
1:ASP_3:HA	1:ASP_3:HB*	1.800	3.000	3.000	30.00	30.00	1000.000
1:ASP_3:HB*	1:VAL_2:HG*	1.800	6.200	3.200	30.00	30.00	1000.000
1:ASP_3:HN	1:ASP_3:HA	1.800	3.100	3.100	30.00	30.00	1000.000
1:ASP_3:HN	1:ASP_3:HB*	1.800	4.500	4.000	30.00	30.00	1000.000
1:ASP_3:HN	1:VAL_2:HA	1.800	3.200	3.200	30.00	30.00	1000.000
1:ASP_3:HN	1:VAL_2:HB	1.800	4.500	4.500	30.00	30.00	1000.000
1:ASP_3:HN	1:VAL_2:HG*	1.800	6.500	4.500	30.00	30.00	1000.000
1:ASP_41:HB*	1:ASP_41:HA	1.800	3.100	3.100	30.00	30.00	1000.000
1:ASP_41:HN	1:LEU_40:HA	1.800	3.000	3.000	30.00	30.00	1000.000

1:ASP_41:HN	1:LEU_40:HD*	1.800	6.200	4.200	30.00	30.00 1000.000
1:ASP_41:HN	1:LEU_40:HBR	1.800	4.500	4.500	30.00	30.00 1000.000
1:ASP_41:HN	1:LEU_40:HBS	1.800	3.500	3.500	30.00	30.00 1000.000
1:ASP_41:HN	1:ASP_41:HA	1.800	2.900	2.900	30.00	30.00 1000.000
1:ASP_41:HN	1:ASP_41:HB*	1.800	3.800	3.300	30.00	30.00 1000.000
1:ASP_5:HN	1:ASP_5:HB*	1.800	4.000	3.000	30.00	30.00 1000.000
1:ASP_5:HN	1:GLY_4:HA*	1.800	4.500	3.500	30.00	30.00 1000.000
1:CYS_12:HA	1:CYS_12:HBS	1.800	3.100	3.100	30.00	30.00 1000.000
1:CYS_12:HA	1:CYS_12:HBR	1.800	2.700	2.700	30.00	30.00 1000.000
1:CYS_12:HA	1:ASN_37:HBR	1.800	3.500	3.500	30.00	30.00 1000.000
1:CYS_12:HA	1:ASN_37:HBS	1.800	4.500	4.500	30.00	30.00 1000.000
1:CYS_12:HN	1:CYS_12:HA	1.800	2.900	2.900	30.0	30.00 1000.000
1:CYS_12:HN	1:CYS_12:HBS	1.800	3.100	3.100	30.00	30.00 1000.000
1:CYS_12:HN	1:CYS_12:HBR	1.800	4.000	4.000	30.00	30.00 1000.000
1:CYS_12:HN	1:PRO_11:HA	1.800	3.600	3.600	30.00	30.00 1000.000
1:CYS_12:HN	1:PRO_11:HBR	1.800	4.750	4.750	30.00	30.00 1000.000
1:CYS_12:HN	1:PRO_11:HD*	1.800	4.300	3.300	30.00	30.00 1000.000
1:CYS_12:HN	1:PRO_11:HG*	1.800	5.500	4.500	30.00	30.00 1000.000
1:CYS_18:HA	1:CYS_27:HBS	1.800	4.200	4.200	30.00	30.00 1000.000
1:CYS_18:HA	1:LYS_19:HA	1.800	4.450	4.450	30.00	30.00 1000.000
1:CYS_18:HN	1:CYS_18:HA	1.800	3.100	3.100	30.00	30.00 1000.000
1:CYS_18:HN	1:CYS_18:HB*	1.800	3.000	2.500	30.00	30.00 1000.000
1:CYS_18:HN	1:SER_17:HA	1.800	3.600	3.600	30.00	30.00 1000.000
1:CYS_18:HN	1:SER_17:HB*	1.800	5.300	4.300	30.00	30.00 1000.000
1:CYS_27:HA	1:CYS_18:HA	1.800	3.100	3.100	30.00	30.00 1000.000
1:CYS_27:HA	1:CYS_27:HBR	1.800	3.100	3.100	30.00	30.00 1000.000
1:CYS_27:HA	1:CYS_27:HBS	1.800	2.700	2.700	30.00	30.00 1000.000
1:CYS_27:HN	1:CYS_27:HA	1.800	3.100	3.100	30.00	30.00 1000.000
1:CYS_27:HN	1:CYS_27:HBR	1.800	3.900	3.900	30.00	30.00 1000.000
1:CYS_27:HN	1:CYS_27:HBS	1.800	4.650	4.650	30.00	30.00 1000.000
1:CYS_27:HN	1:GLU_26:HA	1.800	3.600	3.600	30.00	30.00 1000.000
1:CYS_27:HN	1:GLU_26:HB*	1.800	5.800	4.800	30.00	30.00 1000.000
1:CYS_29:HA	1:CYS_29:HBR	1.800	2.700	2.700	30.00	30.00 1000.000
1:CYS_29:HA	1:CYS_29:HBS	1.800	3.100	3.100	30.00	30.00 1000.000
1:CYS_29:HA	1:PRO_30:HG*	1.800	5.500	4.500	30.00	30.00 1000.000
1:CYS_29:HA	1:GLY_16:HAR	1.800	3000	3.000	30.00	30.00 1000.000
1:CYS_29:HA	1:GLY_16:HAS	1.800	3.500	3.500	30.00	30.00 1000.000
1:CYS_29:HA	1:PRO_30:HD*	1.800	3.800	2.800	30.00	30.00 1000.000
1:CYS_29:HN	1:CYS_29:HA	1.800	3.100	3.100	30.00	30.00 1000.000
1:CYS_29:HN	1:CYS_29:HBR	1.800	4.100	4.100	30.00	30.00 1000.000
1:CYS_29:HN	1:TRP_28:HA	1.800	3.600	3.600	30.00	30.00 1000.000
1:CYS_29:HN	1:CYS_29:HBS	1.800	3.200	3.200	30.00	30.00 1000.000
1:CYS_38:HA	1:CYS_38:HB*	1.800	3.100	2.800	30.00	30.00 1000.000
1:CYS_38:HA	1:GLU_39:HG*	1.800	4.500	4.500	30.00	30.00 1000.000
1:CYS_38:HB*	1:GLU_39:HG*	1.800	5.000	3.500	30.00	30.00 1000.000
1:CYS_38:HA	1:LEU_13:HD*	1.800	5.700	3.700	30.00	30.00 1000.000
1:CYS_38:HN	1:CYS_38:HB*	1.800	3.800	3.300	30.00	30.00 1000.000
1:CYS_38:HN	1:ASN_37:HA	1.800	3.600	3.600	30.00	30.00 1000.000
1:CYS_38:HN	1:CYS_38:HA	1.800	2.800	2.800	30.00	30.00 1000.000
1:CYS_7:HA	1:CYS_7:HBR	1.800	2.700	2.700	30.00	30.00 1000.000

1:CYS_7:HA	1:CYS_7:HBS	1.800	3.100	3.100	30.00	30.00	1000.000
1:CYS_7:HA	1:CYS_18:HB*	1.800	5.000	4.000	30.00	30.00	1000.000
1:CYS_7:HA	1:PRO_11:HG*	1.800	5.500	4.500	30.00	30.00	1000.000
1:CYS_7:HA	1:PRO_11:HD*	1.800	3.900	2.900	30.00	30.00	1000.000
1:CYS_7:HN	1:CYS_7:HA	1.800	2.700	2.700	30.00	30.00	1000.000
1:CYS_7:HN	1:CYS_7:HBR	1.800	4.100	4.100	30.00	30.00	1000.000
1:CYS_7:HN	1:CYS_7:HBS	1.800	3.000	3.000	30.00	30.00	1000.000
1:CYS_7:HN	1:GLN_6:HA	1.800	3.600	3.600	30.00	30.00	1000.000
1:CYS_7:HN	1:GLN_6:HN	1.800	4.150	4.150	30.00	30.00	1000.000
1:CYS_7:HN	1:GLU_8:HN	1.800	2.900	2.900	30.00	30.00	1000.000
1:GLN_6:HA	1:GLN_6:HBS	1.800	2.700	2.700	30.00	30.00	1000.000
1:GLN_6:HA	1:GLN_6:HBR	1.800	3.100	3.100	30.00	30.00	1000.000
1:GLN_6:HA	1:GLN_6:HG*	1.800	3.650	3.150	30.00	30.00	1000.000
1:GLN_6:HG*	1:GLN_6:HBR	1.800	3.700	2.700	30.00	30.00	1000.000
1:GLN_6:HN	1:ASP_5:HA	1.800	3.000	3.000	30.00	30.00	1000.000
1:GLN_6:HN	1:ASP_5:HN	1.800	5.000	5.000	30.00	30.00	1000.000
1:GLN_6:HN	1:GLN_6:HA	1.800	3.100	3.100	30.00	30.00	1000.000
1:GLN_6:HN	1:GLN_6:HBS	1.800	4.300	4.300	30.00	30.00	1000.000
1:GLN_6:HN	1:GLN_6:HBR	1.800	3.100	3.100	30.00	30.00	1000.000
1:GLN_6:HN	1:GLN_6:HG*	1.800	4.500	4.000	30.00	30.00	1000.000
1:GLN_6:HE21	1:GLN_6:HG*	1.800	3.500	3.000	30.00	30.00	1000.000
1:GLU_26:HA	1:GLU_26:HB*	1.800	3.100	3.100	30.00	30.00	1000.000
1:GLU_26:HA	1:GLU_26:HG*	1.800	5.000	4.500	30.00	30.00	1000.000
1:GLU_26:HN	1:GLU_26:HB*	1.800	4.000	3.500	30.00	30.00	1000.000
1:GLU_26:HN	1:GLU_26:HG*	1.800	4.500	4.000	30.00	30.00	1000.000
1:GLU_26:HN	1:TYR_25:HA	1.800	2.700	2.700	30.00	30.00	1000.000
1:GLU_26:HN	1:TYR_25:HBS	1.800	3.000	3.000	30.00	30.00	1000.000
1:GLU_26:HN	1:TYR_25:HBR	1.800	4.000	4.000	30.00	30.00	1000.000
1:GLU_26:HN	1:TYR_25:HD*	1.800	7.500	5.500	30.00	30.00	1000.000
1:GLU_34:HG*	1:LEU_40:HBR	1.800	5.200	4.200	30.00	30.00	1000.000
1:GLU_34:HG*	1:LEU_40:HBS	1.800	5.200	4.200	30.00	30.00	1000.000
1:GLU_34:HG*	1:VAL_42:HGR*	2.200	5.500	4.000	30.00	30.00	1000.000
1:GLU_34:HG*	1:VAL_42:HGS*	2.700	5.000	3.500	30.00	30.00	1000.000
1:GLU_34:HB*	1:VAL_42:HGS*	2.700	5.500	4.000	30.00	30.00	1000.000
1:GLU_34:HN	1:LEU_40:HD*	1.800	6.500	4.500	30.00	30.00	1000.000
1:GLU_34:HN	1:PHE_33:HD*	1.800	7.000	5.500	30.00	30.00	1000.000
1:GLU_34:HN	1:VAL_42:HGS*	1.800	5.750	4.750	30.00	30.00	1000.000
1:GLU_34:HN	1:CYS_38:HA	1.800	5.500	5.500	30.00	30.00	1000.000
1:GLU_34:HN	1:GLU_34:HB*	1.800	4.500	4.000	30.00	30.00	1000.000
1:GLU_34:HN	1:GLU_34:HG*	1.800	4.500	4.000	30.00	30.00	1000.000
1:GLU_34:HN	1:LEU_40:HBS	1.800	4.200	4.200	30.00	30.00	1000.000
1:GLU_34:HN	1:LEU_40:HBR	1.800	5.000	5.000	30.00	30.00	1000.000
1:GLU_34:HN	1:LEU_40:HN	1.800	4.650	4.650	30.00	30.00	1000.000
1:GLU_34:HN	1:PHE_33:HA	1.800	2.800	2.800	30.00	30.00	1000.000
1:GLU_34:HN	1:PHE_33:HBS	1.800	4.200	4.200	30.00	30.00	1000.000
1:GLU_39:HA	1:GLU_39:HB*	1.800	3.100	2.800	30.00	30.00	1000.000
1:GLU_39:HA	1:LEU_13:HD*	1.800	5.300	3.300	30.00	30.00	1000.000
1:GLU_39:HA	1:LEU_13:HBR	1.800	4.000	4.000	30.00	30.00	1000.000
1:GLU_39:HA	1:LEU_13:HBS	1.800	4.500	4.500	30.00	30.00	1000.000
1:GLU_39:HG*	1:GLU_39:HB*	1.800	5.000	4.000	30.00	30.00	1000.000

1:GLU_39:HG*	1:LEU_40:HBR	1.800	5.200	4.200	30.00	30.00	1000.000
1:GLU_39:HG*	1:LEU_40:HD*	1.800	6.700	5.200	30.00	30.00	1000.000
1:GLU_39:HN	1:GLU_39:HA	1.800	3.100	3.100	30.00	30.00	1000.000
1:GLU_39:HN	1:CYS_38:HA	1.800	3.600	3.600	30.00	30.00	1000.000
1:GLU_39:HN	1:ASN_37:HN	1.800	5.000	5.000	30.00	30.00	1000.000
1:GLU_39:HN	1:CYS_38:HN	1.800	4.300	4.300	30.00	30.00	1000.000
1:GLU_39:HN	1:GLU_39:HB*	1.800	4.500	4.000	30.00	30.00	1000.000
1:GLU_8:HA	1:GLU_8:HB*	1.800	3.100	3.100	30.00	30.00	1000.000
1:GLU_8:HA	1:GLU_8:HG*	1.800	3.600	3.100	30.00	30.00	1000.000
1:GLU_8:HB*	1:GLU_8:HN	1.800	3.700	3.200	30.00	30.00	1000.000
1:GLU_8:HN	1:SER_9:HN	1.800	2.800	2.800	30.00	30.00	1000.000
1:GLU_8:HN	1:CYS_7:HA	1.800	3.600	3.600	30.00	30.00	1000.000
1:GLU_8:HN	1:GLU_8:HA	1.800	3.100	3.100	30.00	30.00	1000.000
1:GLU_8:HN	1:GLU_8:HG*	1.800	4.750	4.250	30.00	30.00	1000.000
1:GLY_15:HN	1:ASN_14:HA	1.800	3.200	3.200	30.00	30.00	1000.000
1:GLY_15:HN	1:ASN_14:HB*	1.800	5.000	4.500	30.00	30.00	1000.000
1:GLY_15:HN	1:LEU_13:HBR	1.800	5.000	5.000	30.00	30.00	1000.000
1:GLY_15:HN	1:LEU_13:HBS	1.800	5.000	5.000	30.00	30.00	1000.000
1:GLY_15:HN	1:GLY_15:HA*	1.800	3.200	2.700	30.00	30.00	1000.000
1:GLY_15:HN	1:GLY_16:HN	1.800	3.000	3.000	30.00	30.00	1000.000
1:GLY_16:HAS	1:CYS_12:HBR	1.800	3.500	3.500	30.00	30.00	1000.000
1:GLY_16:HAR	1:PRO_30:HD*	1.800	4.000	3.000	30.00	30.00	1000.000
1:GLY_16:HAS	1:PRO_30:HD*	1.800	4.500	3.500	30.00	30.00	1000.000
1:GLY_16:HN	1:GLY_16:HAR	1.800	2.900	2.900	30.00	30.00	1000.000
1:GLY_16:HN	1:GLY_16:HAS	1.800	3.100	3.100	30.00	30.00	1000.000
1:GLY_16:HN	1:GLY_15:HA*	1.800	4.500	3.500	30.00	30.00	1000.000
1:GLY_32:HA*	1:VAL_42:HGS*	1.800	6.500	5.000	30.00	30.00	1000.000
1:GLY_32:HA*	1:VAL_42:HGR*	1.800	7.000	5.500	30.00	30.00	1000.000
1:GLY_32:HA*	1:THRC_43:HG*	1.800	7.000	5.500	30.00	30.00	1000.000
1:GLY_32:HN	1:PHE_31:HBR	1.800	5.000	5.000	30.00	30.00	1000.000
1:GLY_32:HN	1:GLY_32:HA*	1.800	3.000	2.850	30.00	30.0	1000.000
1:GLY_32:HN	1:PHE_31:HA	1.800	2.700	2.700	30.00	30.00	1000.000
1:GLY_32:HN	1:PHE_33:HN	1.800	2.800	2.800	30.00	30.00	1000.000
1:GLY_32:HN	1:VAL_42:HN	1.800	3.000	3.000	30.00	30.00	1000.000
1:GLY_32:HN	1:VAL_42:HGS*	1.800	5.000	4.000	30.00	30.00	1000.000
1:GLY_35:HN	1:GLU_34:HB*	1.800	4.800	3.800	30.00	30.00	1000.000
1:GLY_35:HN	1:GLU_34:HG*	1.800	5.500	4.500	30.00	30.00	1000.000
1:GLY_35:HN	1:GLY_35:HA*	1.800	3.000	3.000	30.00	30.00	1000.000
1:GLY_4:HN	1:ASP_3:HA	1.800	3.600	3.600	30.00	30.00	1000.000
1:GLY_4:HN	1:ASP_3:HB*	1.800	4.200	3.200	30.00	30.00	1000.000
1:GLY_4:HN	1:GLY_4:HA*	1.800	3.100	3.100	30.00	30.00	1000.000
1:ILE_22:HA	1:ILE_22:HB	1.800	3.100	3.100	30.00	30.00	1000.000
1:ILE_22:HA	1:ILE_22:HG1*	1.800	4.500	3.500	30.00	30.00	1000.000
1:ILE_22:HA	1:ILE_22:HG2*	1.800	5.000	4.000	30.00	30.00	1000.000
1:ILE_22:HB	1:ILE_22:HG1*	1.800	4.000	3.000	30.00	30.00	1000.000
1:ILE_22:HB	1:ILE_22:HG2*	1.800	4.200	3.200	30.00	30.00	1000.000
1:ILE_22:HG1*	1:ILE_22:HG2*	1.800	4.300	2.800	30.00	30.00	1000.000
1:ILE_22:HN	1:ILE_22:HG2*	1.800	4.000	3.000	30.00	30.00	1000.000
1:ILE_22:HN	1:ASP_21:HA	1.800	3.600	3.600	30.00	30.00	1000.000
1:ILE_22:HN	1:ASP_21:HBS	1.800	3.500	3.500	30.00	30.00	1000.000

1:ILE_22:HN	1:ASP_21:HBR	1.800	3.500	3.500	30.00	30.00	1000.000
1:ILE_22:HN	1:ASP_21:HN	1.800	3.300	3.300	30.00	30.00	1000.000
1:ILE_22:HN	1:SER_24:HN	1.800	3.500	3.500	30.00	30.00	1000.000
1:ILE_22:HN	1:ILE_22:HA	1.800	3.000	2.840	30.00	30.00	1000.000
1:ILE_22:HN	1:ILE_22:HB	1.800	3.800	3.800	30.00	30.00	1000.000
1:ILE_22:HN	1:ILE_22:HG1*	1.800	6.000	5.000	30.00	30.00	1000.000
1:LEU_13:HA	1:LEU_13:HBR	1.800	3.100	3.100	30.00	30.00	1000.000
1:LEU_13:HA	1:LEU_13:HBS	1.800	2.700	2.800	30.00	30.00	1000.000
1:LEU_13:HA	1:LEU_13:HD*	1.800	6.000	4.500	30.00	30.00	1000.000
1:LEU_13:HA	1:LEU_13:HG	1.800	3.200	3.200	30.00	30.00	1000.000
1:LEU_13:HBS	1:LEU_13:HD*	1.800	4.600	3.100	30.00	30.00	1000.000
1:LEU_13:HBR	1:LEU_13:HD*	1.800	4.500	3.000	30.00	30.00	1000.000
1:LEU_13:HG	1:LEU_13:HD*	1.800	5.000	3.500	30.00	30.00	1000.000
1:LEU_13:HN	1:ASN_37:HBR	1.800	4.500	4.500	30.00	30.00	1000.000
1:LEU_13:HN	1:GLY_16:HN	1.800	3.300	3.300	30.00	30.00	1000.000
1:LEU_13:HN	1:CYS_12:HA	1.800	3.600	3.600	30.00	30.00	1000.000
1:LEU_13:HN	1:CYS_12:HBS	1.800	4.300	4.300	30.00	30.00	1000.000
1:LEU_13:HN	1:CYS_12:HBR	1.800	4.000	4.500	30.00	30.00	1000.000
1:LEU_13:HN	1:CYS_38:HB*	1.800	5.000	4.000	30.00	30.00	1000.000
1:LEU_13:HN	1:LEU_13:HA	1.800	3.100	3.100	30.00	30.00	1000.000
1:LEU_13:HN	1:LEU_13:HBR	1.800	2.800	2.800	30.00	30.00	1000.000
1:LEU_13:HN	1:LEU_13:HBS	1.800	4.000	4.000	30.00	30.00	1000.000
1:LEU_13:HN	1:LEU_13:HD*	1.800	5.500	4.000	30.00	30.00	1000.000
1:LEU_13:HN	1:LEU_13:HG	1.800	3.500	3.000	30.00	30.00	1000.000
1:LEU_13:HN	1:ASN_14:HB*	1.800	5.500	4.500	30.00	30.00	1000.000
1:LEU_40:HA	1:LEU_40:HBR	1.800	3.100	2.600	30.00	30.00	1000.000
1:LEU_40:HA	1:LEU_40:HBS	1.800	2.600	2.600	30.00	30.00	1000.000
1:LEU_40:HA	1:LEU_40:HD*	1.800	5.700	4.200	30.00	30.00	1000.000
1:LEU_40:HBS	1:LEU_40:HD*	1.800	4.200	2.700	30.00	30.00	1000.000
1:LEU_40:HN	1:LEU_40:HA	1.800	2.900	2.900	30.00	30.00	1000.000
1:LEU_40:HN	1:LEU_40:HD*	1.800	5.000	3.500	30.00	30.00	1000.000
1:LEU_40:HN	1:GLU_39:HA	1.800	3.000	3.000	30.00	30.00	1000.000
1:LEU_40:HN	1:GLU_39:HN	1.800	4.000	4.000	30.00	30.00	1000.000
1:LEU_40:HN	1:LEU_40:HBR	1.800	3.000	3.000	30.00	30.00	1000.000
1:LEU_40:HN	1:LEU_40:HBS	1.800	4.000	4.000	30.00	30.00	1000.000
1:LYS_19:HA	1:LYS_19:HB*	1.800	3.100	3.100	30.00	30.00	1000.000
1:LYS_19:HA	1:LYS_19:HG*	1.800	3.500	3.000	30.00	30.00	1000.000
1:LYS_19:HE*	1:LYS_19:HA	1.800	3.600	3.100	30.00	30.00	1000.000
1:LYS_19:HE*	1:LYS_19:HD*	1.800	4.100	3.100	30.00	30.00	1000.000
1:LYS_19:HE*	1:LYS_19:HG*	1.800	5.500	4.500	30.00	30.00	1000.000
1:LYS_19:HN	1:CYS_27:HA	1.800	4.850	4.850	30.00	30.00	1000.000
1:LYS_19:HN	1:LYS_19:HG*	1.800	3.600	3.100	30.00	30.00	1000.000
1:LYS_19:HN	1:CYS_18:HA	1.800	2.300	2.300	30.00	30.00	1000.000
1:LYS_19:HN	1:CYS_18:HB*	1.800	5.500	4.500	30.00	30.00	1000.000
1:LYS_19:HN	1:LYS_19:HA	1.800	3.100	3.100	30.00	30.00	1000.000
1:LYS_19:HN	1:LYS_19:HB*	1.800	4.500	4.500	30.00	30.00	1000.000
1:LYS_36:HA	1:LYS_36:HB*	1.800	3.100	3.100	30.00	30.00	1000.000
1:LYS_36:HA	1:LYS_36:HD*	1.800	4.700	4.200	30.00	30.00	1000.000
1:LYS_36:HA	1:LYS_36:HG*	1.800	3.600	3.100	30.00	30.00	1000.000

1:LYS_36:HB*	1:LYS_36:HG*	1.800	4.000	3.000	30.00	30.00	1000.000
1:LYS_36:HE*	1:LYS_36:HG*	1.800	5.500	4.500	30.00	30.00	1000.000
1:LYS_36:HN	1:ASN_37:HN	1.800	4.250	4.250	30.00	30.00	1000.000
1:LYS_36:HN	1GLY_35:HA*	1.800	3.800	3.300	30.00	30.00	1000.000
1:LYS_36:HN	1:LYS_36:HA	1.800	3.100	3.100	30.00	30.00	1000.000
1:LYS_36:HN	1:LYS_36:HB*	1.800	3.500	3.000	30.00	30.00	1000.000
1:LYS_36:HN	1:LYS_36:HD*	1.800	5.200	4.700	30.00	30.00	1000.000
1:LYS_36:HN	1:LYS_36:HG*	1.800	4.700	4.200	30.00	30.00	1000.000
1:PHE_31:HA	1:PHE_31:HBR	1.800	2.700	2.700	30.00	30.00	1000.000
1:PHE_31:HA	1:PHE_31:HBS	1.800	3.100	3.100	30.00	30.00	1000.000
1:PHE_31:HD*	1:PHE_31:HA	1.800	6.500	5.500	30.00	30.00	1000.000
1:PHE_31:HD*	1:PHE_31:HBR	1.800	3.800	2.800	30.00	30.00	1000.000
1:PHE_31:HD*	1:PHE_31:HBS	1.800	3.600	2.600	30.00	30.00	1000.000
1:PHE_31:HE*	1:PHE_31:HBR	1.800	6.000	5.000	30.00	30.00	1000.000
1:PHE_31:HE*	1:PHE_31:HBS	1.800	5.500	4.500	30.00	30.00	1000.000
1:PHE_31:HN	1:PRO_30:HBS	1.800	4.000	4.000	30.00	30.00	1000.000
1:PHE_31:HN	1:PRO_30:HBR	1.800	4.000	4.000	30.00	30.00	1000.000
1:PHE_31:HN	1:GLY_32:HN	1.800	3.500	3.500	30.00	30.00	1000.000
1:PHE_31:HN	1:PHE_31:HA	1.800	3.000	3.000	30.00	30.00	1000.000
1:PHE_31:HN	1:PHE_31:HBR	1.800	2.800	2.800	30.00	30.00	1000.000
1:PHE_31:HN	1:PHE_31:HBS	1.800	2.800	2.800	30.00	30.00	1000.000
1:PHE_31:HN	1:PHE_31:HD*	1.800	5.500	4.500	30.00	30.00	1000.000
1:PHE_31:HN	1:PRO_30:HA	1.800	3.000	3.000	30.00	30.00	1000.000
1:PHE_33:HA	1:VAL_42:HA	1.800	4.450	4.500	30.00	30.00	1000.000
1:PHE_33:HA	1:ASP_41:HA	1.800	3.100	3.100	30.00	30.00	1000.000
1:PHE_33:HA	1:PHE_33:HBS	1.800	2.700	2.700	30.00	30.00	1000.000
1:PHE_33:HA	1:PHE_33:HBR	1.800	3.100	3.100	30.00	30.00	1000.000
1:PH_33:HA	1:VAL_42:HGS*	1.800	5.200	4.200	30.00	30.00	1000.000
1:PHE_33:HA	1:GLU_34:HG*	1.800	5.800	4.800	30.00	30.00	1000.000
1:PHE_33:HD*	1:CYS_29:HBR	1.800	6.500	5.000	30.00	30.00	1000.000
1:PHE_33:HD*	1:GLY_32:HA*	1.800	7.000	5.000	30.00	30.00	1000.000
1:PHE_33:HD*	1:ASP_41:HA	1.800	5.000	3.500	30.00	30.00	1000.000
1:PHE_33:HD*	1:PHE_33:HA	1.800	4.000	3.000	30.00	30.00	1000.000
1:PHE_33:HD*	1:PHE_33:HBS	1.800	4.000	3.000	30.00	30.00	1000.000
1:PHE_33:HD*	1:PHE_33:HBR	1.800	4.000	3.000	30.00	30.00	1000.000
1:PHE_33:HD*	1:PRO_30:HBS	1.800	5.700	4.200	30.00	30.00	1000.000
1:PHE_33:HD*	1:PRO_30:HBR	1.800	4.700	3.200	30.00	30.00	1000.000
1:PHE_33:HD*	1:PRO_30:HD*	1.800	6.000	4.000	30.00	30.00	1000.000
1:PHE_33:HD*	1:PRO_30:HG*	1.800	6.200	4.200	30.00	30.00	1000.000
1:PHE_33:HE*	1:PHE_33:HBS	1.800	5.500	4.500	30.00	30.00	1000.000
1:PHE_33:HE*	1:PRO_30:HG*	1.800	6.300	4.300	30.00	30.00	1000.000
1:PHE_33:HE*	1:ASP_41:HA	1.800	6.000	4.500	30.00	30.00	1000.000
1:PHE_33:HE*	1:ASP_41:HB*	1.800	6.500	4.500	30.00	30.00	1000.000
1:PHE_33:HN	1:GLY_32:HA*	1.800	4.000	3.500	30.00	30.00	1000.000
1:PHE_33:HN	1:PHE_33:HA	1.800	3.000	3.000	30.00	30.00	1000.000
1:PHE_33:HN	1:PHE_33:HBS	1.800	4.300	4.300	30.00	30.00	1000.000
1:PHE_33:HN	1:PHE_33:HBR	1.800	3.000	3.000	30.00	30.00	1000.000
1:PHE_33:HN	1:PHE_33:HD*	1.800	4.500	3.500	30.00	30.00	1000.000
1:PHE_33:HN	1:PHE_31:HA	1.800	3.500	3.500	30.00	30.00	1000.000
1:PHE_33:HN	1:VAL_42:HB	1.800	4.500	4.500	30.00	30.00	1000.000

1:PRO_11:HA	1:PRO_11:HBS	1.800	2.700	2.700	30.00	30.00 100.000
1:PRO_11:HA	1:PRO_11:HBR	1.800	3.100	3.000	30.00	30.00 1000.000
1:PRO_11:HA	1:PRO_11:HD*	1.800	5.200	4.200	30.00	30.00 1000.000
1:PRO_11:HA	1:PRO_11:HG*	1.800	4.500	4.500	30.00	30.00 1000.000
1:PRO_11:HBS	1:PRO_11:HG*	1.800	3.500	3.500	30.00	30.00 1000.000
1:PRO_11:HD*	1:ASN_10:HB*	1.800	5.500	4.000	30.00	30.00 1000.000
1:PRO_11:HD*	1:PRO_11:HBS	1.800	3.700	3.200	30.00	30.00 1000.000
1:PRO_11:HD*	1:PRO_11:HBR	1.800	3.500	3.000	30.00	30.00 1000.000
1:PRO_11:HD*	1:PRO_11:HG*	1.800	4.200	3.200	30.00	30.00 1000.000
1:PRO_11:HD*	1:CYS_18:HB*	1.800	5.500	4.000	30.00	30.00 1000.000
1:PRO_30:HA	1:PRO_30:HG*	1.800	5.000	4.500	30.00	30.00 1000.000
1:PRO_30:HA	1:PRO_30:HBS	1.800	2.700	2.700	30.00	30.00 1000.000
1:PRO_30:HA	1:PRO_30:HBR	1.800	3.100	2.900	30.00	30.00 1000.000
1:PRO_30:HBS	1:PRO_30:HG*	1.800	3.500	3.000	30.00	30.00 1000.000
1:PRO_30:HD*	1:CYS_29:HBR	1.800	4.500	4.000	30.00	30.00 1000.000
1:PRO_30:HD*	1:CYS_29:HBS	1.800	4.500	4.000	30.00	30.00 1000.000
1:PRO_30:HD*	1:PRO_30:HG*	1.800	4.000	4.000	30.00	30.00 1000.000
1:SER_17:HB*	1:TRP_28:HB*	1.800	5.000	4.000	30.00	30.00 1000.000
1:SER_17:HN	1:TRP_28:HB*	1.800	4.500	4.000	30.00	30.00 1000.000
1:SER_17:HN	1:GLY_16:HN	1.800	3.600	3.600	30.00	30.00 1000.000
1:SER_17:HN	1:GLY_16:HAR	1.800	3.000	3.000	30.00	30.00 1000.000
1:SER_17:HN	1:GLY_16:HAS	1.800	3.000	3.000	30.00	30.00 1000.000
1:SER_17:HN	1:SER_17:HB*	1.800	3.800	3.300	30.00	30.00 1000.000
1:SER_17:HN	1:TRP_28:HE3	1.800	4.000	4.000	30.00	30.00 1000.000
1:SER_24:HA	1:SER_24:HB*	1.800	3.100	3.000	30.00	30.00 1000.000
1:SER_24:HN	1:SER_24:HA	1.800	3.100	3.100	30.00	30.00 1000.000
1:SER_24:HN	1:ASN_23:HA	1.800	3.600	3.600	30.00	30.00 1000.000
1:SER_24:HA	1:GLN_6:HG*	1.800	5.000	4.000	30.00	30.00 1000.000
1:SER_24:HN	1:ILE_22:HG2*	1.800	4.000	3.000	30.00	30.00 1000.000
1:SER_24:HB*	1:ILE_22:HG2*	1.800	6.000	4.500	30.00	30.00 1000.000
1:SER_9:HA	1:SER_9:HB*	1.800	3.100	3.000	30.00	30.00 1000.000
1:SER_9:HN	1:GLU_8:HA	1.800	3.600	3.600	30.00	30.00 1000.000
1:SER_9:HN	1:GLU_8:HB*	1.800	5.000	4.000	30.00	30.00 1000.000
1:SER_9:HN	1:SER_9:HA	1.800	3.100	3.100	30.00	30.00 1000.000
1:SER_9:HN	1:SER_9:HB*	1.800	4.200	4.200	30.00	30.00 1000.000
1:SER_9:HN	1:PRO_11:HD*	1.800	5.000	4.000	30.00	30.00 1000.000
1:THRC_43:HA	1:THRC_43:HG*	1.800	4.100	3.100	30.00	30.00 1000.000
1:THRC_43:HN	1:THRC_43:HG*	1.800	5.500	4.500	30.00	30.00 1000.000
1:THRC_43:HN	1:THRC_43:HA	1.800	3.000	3.000	30.00	30.00 1000.000
1:THRC_43:HN	1:VAL_42:HGR*	1.800	3.900	2.900	30.00	30.00 1000.000
1:THRC_43:HN	1:VAL_42:HGS*	1.800	4.100	3.100	30.00	30.00 1000.000
1:TRP_28:HA	1:TRP_28:HB*	1.800	3.100	2.750	30.00	30.00 1000.000
1:TRP_28:HD1	1:TRP_28:HB*	1.800	3.500	3.000	30.00	30.00 1000.000
1:TRP_28:HE3	1:SER_17:HB*	1.800	5.500	4.500	30.00	30.00 1000.000
1:TRP_28:HE3	1:TRP_28:HB*	1.800	3.800	3.000	30.00	30.00 1000.000
1:TRP_28:HN	1:TRP_28:HB*	1.800	3.000	3.000	30.00	30.00 1000.000
1:TRP_28:HN	1:CYS_27:HA	1.800	2.800	2.800	30.00	30.00 1000.000
1:TYRN_1:HB*	1:VAL_2:HG*	1.800	6.500	4.000	30.00	30.00 1000.000
1:TYRN_1:HD*	1:TYRN_1:HB*	1.800	5.100	3.600	30.00	30.00 1000.000
1:TYR_25:HA	1:ASP_21:HBS	1.800	4.500	4.500	30.00	30.00 1000.000

1:TYR_25:HA	1:ASP_21:HBR	1.800	4.500	4.500	30.00	30.00	1000.000
1:TYR_25:HA	1:ASP_20:HA	1.800	3.000	3.000	30.00	30.00	1000.000
1:TYR_25:HA	1:GLN_6:HBR	1.800	4.500	4.500	30.00	30.00	1000.000
1:TYR_25:HA	1:GLN_6:HG*	1.800	7.000	6.000	30.00	30.00	1000.000
1:TYR_25:HA	1:TYR_25:HBR	1.800	2.700	2.700	30.00	30.00	1000.000
1:TYR_25:HA	1:TYR_25:HBS	1.800	2.700	2.700	30.00	30.00	1000.000
1:TYR_25:HBR	1:GLN_6:HBS	1.800	4.500	4.500	30.00	30.00	1000.000
1:TYR_25:HBR	1:GLN_6:HBR	1.800	4.500	4.500	30.00	30.00	1000.000
1:TYR_25:HBS	1:GLN_6:HBR	1.800	4.000	4.000	30.00	30.00	1000.000
1:TYR_25:HBS	1:GLN_6:HBS	1.800	4.000	4.000	30.00	30.00	1000.000
1:TYR_25:HD*	1:GLN_6:HA	1.800	5.800	4.300	30.00	30.00	1000.000
1:TYR_25:HD*	1:GLN_6:HBS	1.800	4.700	3.200	30.00	30.00	1000.000
1:TYR_25:HD*	1:GLN_6:HBR	1.800	5.500	4.000	30.00	30.00	1000.000
1:TYR_25:HD*	1:GLN_6:HG*	1.800	6.200	4.200	30.00	30.00	1000.000
1:TYR_25:HD*	1:PRO_11:HBS	1.800	7.000	5.500	30.00	30.00	1000.000
1:TYR_25:HD*	1:PRO_11:HBR	1.800	7.000	5.500	30.00	30.00	1000.000
1:TYR_25:HD*	1:PRO_11:HG*	1.800	6.300	4.300	30.00	30.00	1000.000
1:TYR_25:HD*	1:PRO_11:HD*	1.800	7.500	5.500	30.00	30.00	1000.000
1:TYR_25:HD*	1:GLU_26:HA	1.800	5.500	4.000	30.00	30.00	1000.000
1:TYR_25:HD*	1:SER_24:HA	1.800	5.800	4.300	30.00	30.00	1000.000
1:TYR_25:HD*	1:TYR_25:HA	1.800	5.000	4.000	30.00	30.00	1000.000
1:TYR_25:HD*	1:TYR_25:HBR	1.800	4.000	3.000	30.00	30.00	1000.000
1:TYR_25:HD*	1:TYR_5:HBS	1.800	4.000	3.000	30.00	30.00	1000.000
1:TYR_25:HD*	1:LYS_36:HD*	1.800	7.500	5.500	30.00	30.00	1000.000
1:TYR_25:HE*	1:PRO_11:HBS	1.800	6.500	5.000	30.00	30.00	1000.000
1:TYR_25:HE*	1:PRO_11:HBR	1.800	6.000	4.500	30.00	30.00	1000.000
1:TYR_25:HE*	1:PRO_11:HG*	1.800	6.500	4.500	30.00	30.00	1000.000
1:TYR_25:HE*	1:PRO_11:HD*	1.800	7.500	5.500	30.00	30.00	1000.000
1:TYR_25:HE*	1:GLN_6:HBS	1.800	6.000	4.500	30.00	30.00	1000.000
1:TYR_25:HE*	1:GLN_6:HG*	1.800	6.500	4.500	30.00	30.00	1000.000
1:TYR_25:HE*	1:LYS_36:HD*	1.800	6.500	4.500	30.00	30.00	1000.000
1:TYR_25:HE*	1:SER_24:HA	1.800	5.800	4.300	30.00	30.00	1000.000
1:TYR_25:HE*	1:TYR_25:HBR	1.800	5.300	4.300	30.00	30.00	1000.000
1:TYR_25:HE*	1:TYR_25:HBS	1.800	5.300	4.300	30.00	30.00	1000.000
1:TYR_25:HN	1:SER_24:HN	1.800	4.500	4.500	30.00	30.00	1000.000
1:TYR_25:HN	1:SER_24:HA	1.800	3.600	3.600	30.00	30.00	1000.000
1:TYR_25:HN	1:SER_24:HB*	1.800	4.200	3.200	30.00	30.00	1000.000
1:TYR_25:HN	1:TYR_25:HA	1.800	3.000	3.000	30.00	30.00	1000.000
1:TYR_25:HN	1:TYR_25:HBR	1.800	4.500	4.500	30.00	30.00	1000.000
1:TYR_25:HN	1:TYR_25:HBS	1.800	4.000	4.000	30.00	30.00	1000.000
1:TYR_25:HN	1:TYR_25:HD*	1.800	4.200	3.200	30.00	30.00	1000.000
1:VAL_2:HA	1:VAL_2:HG*	1.800	4.500	3.000	30.00	30.00	1000.000
1:VAL_2:HN	1:ASP_3:HB*	1.800	5.500	4.500	30.00	30.00	1000.000
1:VAL_2:HN	1:TYRN_1:HA	1.800	3.200	3.200	30.00	30.00	1000.000
1:VAL_2:HN	1:VAL_2:HG*	1.800	5.500	4.000	30.00	30.00	1000.000
1:VAL_2:HN	1:TYRN_1:HB*	1.800	5.500	4.500	30.00	30.00	1000.000
1:VAL_2:H	1:VAL_2:HA	1.800	3.100	3.100	30.00	30.00	1000.000
1:VAL_2:HN	1:VAL_2:HB	1.800	4.500	4.500	30.00	30.00	1000.000
1:VAL_42:HA	1:THRC_43:HG*	1.800	5.300	4.300	30.00	30.00	1000.000
1:VAL_42:HA	1:VAL_42:HGR*	1.800	3.800	2.800	30.00	30.00	1000.000

```
1:VAL_42:HA          1:VAL_42:HGS*       1.800  3.600  2.600 30.00 30.00 1000.000
1:VAL_42:HB          1:VAL_42:HA         1.800  3.100  3.100 30.00 30.00 1000.000
1:VAL_42:HB          1:VAL_42:HGR*       1.800  3.600  2.600 30.00 30.00 1000.000
1:VAL_42:HB          1:VAL_42:HGS*       1.800  3.400  2.400 30.00 30.00 1000.000
1:VAL_42:HN          1:PHE_33:HA         1.800  2.900  2.900 30.00 30.00 1000.000
1:VAL_42:HN          1:THRC_43:HN        1.800  2.600  2.600 30.00 30.00 1000.000
1:VAL_42:HN          1:VAL_42:HA         1.800  2.900  2.900 30.00 30.00 1000.000
1:VAL_42:HN          1:VAL_42:HB         1.800  3.800  3.800 30.00 30.00 1000.000
1:VAL_42:HN          1:VAL_42:HGR*       1.800  5.000  4.000 30.00 30.00 1000.000
1:VAL_42:HN          1:VAL_42:HGS*       1.800  4.000  3.000 30.00 30.00 1000.000
1:VAL_42:HN          1:ASP_41:HA         1.800  3.500  3.500 30.00 30.00 1000.000

#NMR_dihedral
1:TYRN_1:C          1:VAL_2:N          1:VAL_2:CA          1:VAL_2:C          -160.0
-80.0  100.00 100.00 1000.000
1:GLU_8:C           1:SER_9:N          1:SER_9:CA          1:SER_9:C          -160.0
-80.0  100.00 100.00 1000.000
1:PRO_11:C          1:CYS_12:N         1:CYS_12:CA         1:CYS_12:C         -160.0
-80.0  100.00 100.00 1000.000
1:SER_17:C          1:CYS_18:N         1:CYS_18:CA         1:CYS_18:C         -160.0
-80.0  100.00 100.00 1000.000
1:CYS_18:C          1:LYS_19:N         1:LYS_19:CA         1:LYS_19:C         -160.0
-80.0  100.00 100.00 1000.000
1:ASP_20:C          1:ASP_21:N         1:ASP_21:CA         1:ASP_21:C         -160.0
-80.0  100.00 100.00 1000.000
1:SER_24:C        1:TYR_25:N         1:TYR_25:CA         1:TYR_25:C         -160.0
-80.0  100.00 100.00 1000.000
1:TYR_25:C          1:GLU_26:N         1:GLU_26:CA         1:GLU_26:C         -160.0
-80.0  100.00 100.00 1000.000
1:GLU_26:C          1:CYS_27:N         1:CYS_27:CA         1:CYS_27:C         -160.0
-80.0  100.00 100.00 1000.000
1:TRP_28:C          1:CYS_29:N         1:CYS_29:CA         1:CYS_29:C         -160.0
-80.0  100.00 100.00 1000.000
1:GLY_32:C          1:PHE_33:N         1:PHE_33:CA         1:PHE_33:C         -160.0
-80.0  100.00 100.00 1000.000
1:LYS_36:C          1:ASN_37:N         1:ASN_37:CA         1:ASN_37:C         -160.0
-80.0  100.00 100.00 1000.000
1:ASN_37:C          1:CYS_38:N         1:CYS_38:CA         1:CYS_38:C         -160.0
-80.0  100.00 100.00 1000.000
1:CYS_38:C          1:GLU_39:N         1:GLU_39:CA         1:GLU_39:C         -160.0
-80.0  100.00 100.00 1000.000
1:GLU_39:C          1:LEU_40:N         1:LEU_40:CA         1:LEU_40:C         -160.0
-80.0  100.00 100.00 1000.000
1:ASP_41:C          1:VAL_42:N         1:VAL_42:CA         1:VAL_42:C         -160.0
-80.0  100.00 100.00 1000.000
1:VAL_42:C          1:THRC_43:N        1:THRC_43:CA        1:THRC_43:C        -160.0
-80.0  100.00 100.00 1000.000
1:GLN_6:N           1:GLN_6:CA         1:GLN_6:CB          1:GLN_6:CG          -90.0
-30.0  100.00 100.00 1000.000
1:CYS_7:N           1:CYS_7:CA         1:CYS_7:CB          1:CYS_7:SG          -90.0
-30.0  100.00 100.00 1000.000
1:CYS_12:N          1:CYS_12:CA        1:CYS_12:CB         1:CYS_12:SG         -90.0
-30.0  100.00 100.00 1000.000
1:LEU_13:N          1:LEU_13:CA        1:LEU_13:CB         1:LEU_13:CG         -90.0
-30.0  100.00 100.00 1000.000
```

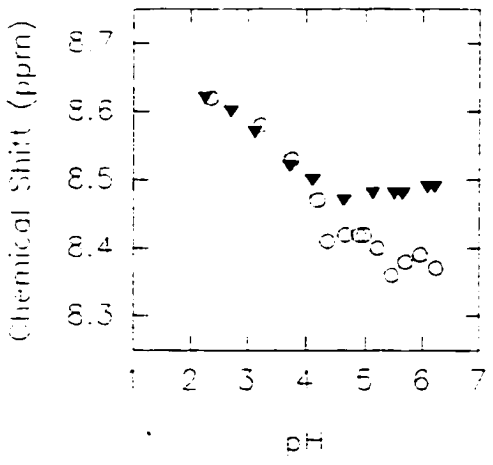
```
1:ASP_21:N          1:ASP_21:CA         1:ASP_21:CB         1:ASP_21:CG          30.0
 90.0  100.00 100.00 1000.000
1:TYR_25:N          1:TYR_25:CA         1:TYR_25:CB         1:TYR_25:CG          30.0
 90.0  100.00 100.00 1000.000
1:CYS_27:N          1:CYS_27:CA         1:CYS_27:CB         1:CYS_27:SG         -90.0
-30.0  100.00 100.00 1000.000
1:CYS_29:N          1:CYS_29:CA         1:CYS_29:CB         1:CYS_29:SG         -90.0
-30.0  100.00 100.00 1000.000
1:PHE_31:N          1:PHE_31:CA         1:PHE_31:CB         1:PHE_31:CG         150.0
-150.0  100.00 100.00 1000.000
1:PHE_33:N          1:PHE_33:CA         1:PHE_33:CB         1:PHE_33:CG         -90.0
-30.0  100.00 100.00 1000.000
1:ASN_37:N          1:ASN_37:CA         1:ASN_37:CB         1:ASN_37:CG         -90.0
-30.0  100.00 100.00 1000.000
1:LEU_40:N          1:LEU_40:CA         1:LEU_40:CB         1:LEU_40:CG         -90.0
-30.0  100.00 100.00 1000.000
1:VAL_42:N          1:VAL_42:CA         1:VAL_42:CB         1:VAL_42:CG1        150.0
-150.0  100.00 100.00 1000.000
1:ILE_22:N          1:ILE_22:CA         1:ILE_22:CB         1:ILE_22:CG1        -90.0
-30.0  100.00 100.00 1000.000
1:TYRN_1:CA         1:TYRN_1:C          1:VAL_2:N           1:VAL_2:CA          175.0
-175.0  100.00 100.00 1000.000
1:VAL_2:CA          1:VAL_2:C           1:ASP_3:N           1:ASP_3:CA          175.0
-175.0  100.00 100.00 1000.000
1:ASP_3:CA          1:ASP_3:C           1:GLY_4:N           1:GLY_4:CA          175.0
-175.0  100.00 100.00 1000.000
1:GLY_4:CA          1:GLY_4:C           1:ASP_5:N           1:ASP_5:CA          175.0
-175.0  100.00 100.00 1000.000
1:ASP_5:CA          1:ASP_5:C           1:GLN_6:N           1:GLN_6:CA          175.0
-175.0  100.00 100.00 1000.000
1:GLN_6:CA          1:GLN_6:C           1:CYS_7:N           1:CYS_7:CA          175.0
-175.0  100.00 100.00 1000.000
1:CYS_7:CA          1:CYS_7:C           1:GLU_8:N           1:GLU_8:CA          175.0
-175.0  100.00 100.00 1000.000
1:GLU_8:CA          1:GLU_8:C           1:SER_9:N           1:SER_9:CA          175.0
-175.0  100.00 100.00 1000.000
1:SER_9:CA          1:SER_9:C           1:ASN_10:N          1:ASN_10:CA         175.0
-175.0  100.00 100.00 1000.000
1:ASN_10:CA         1:ASN_10:C          1:PRO_11:N          1:PRO_11:CA         175.0
-175.0  100.00 100.00 1000.000
1:PRO_11:CA         1:PRO_11:C          1:CYS_12:N          1:CYS_12:CA         175.0
-175.0  100.00 100.00 1000.000
1:CYS_12:CA          1:CYS_12C          1:LEU_13:N          1:LEU_13:CA         175.0
-175.0  100.00 100.00 1000.000
1:LEU_13:CA         1:LEU_13:C          1:ASN_14:N          1:ASN_14:CA         175.0
-175.0  100.00 100.00 1000.000
1:ASN_14:CA         1:ASN_14:C          1:GLY_15:N          1:GLY_15:CA         175.0
-175.0  100.00 100.00 1000.000
1:GLY_15:CA         1:GLY_15:C          1:GLY_16:N          1:GLY_16:CA         175.0
-175.0  100.00 100.00 1000.000
1:GLY_16:CA         1:GLY_16:C          1:SER_17:N          1:SER_17:CA         175.0
-175.0  100.00 100.00 1000.000
1:SER_17:CA         1:SER_17:C          1:CYS_18:N          1:CYS_18:CA         175.0
-175.0  100.00 100.00 1000.000
1:CYS_18:CA         1:CYS_18:C          1:LYS_19:N          1:LYS_19:CA         175.0
-175.0  100.00 100.00 1000.000
1:LYS_19:CA         1:LYS_19:C          1:ASP_20:N          1:ASP_20:CA         175.0
```

```
-175.0  100.00 100.00 1000.000
1:ASP_20:CA        1:ASP_20:C        1:ASP_21:N        1:ASP_21:CA       175.0
-175.0  100.00 100.00 1000.000
1:ASP_21:CA        1:ASP_21:C        1:ILE_22:N        1:ILE_22:CA       175.0
-175.0  100.00 100.00 1000.000
1:ILE_22:CA        1:ILE_22:C        1:ASN_23:N        1:ASN_23:CA       175.0
-175.0  100.00 100.00 1000.000
1:ASN_23:CA        1:ASN_23:C        1:SER_24:N        1:SER_24:CA       175.0
-175.0  100.00 100.00 1000.000
1:SER_24:CA        1:SER_24:C        1:TYR_25:N        1:TYR_25:CA       175.0
-175.0  100.00 100.00 1000.000
1:TYR_25:CA        1:TYR_25:C        1:GLU_26:N        1:GLU_26:CA       175.0
-175.0  100.00 100.00 1000.000
1:GLU_26:CA        1:GLU_26:C        1:CYS_27:N        1:CYS_27:CA       175.0
-175.0  100.00 100.00 1000.000
1:CYS_27:CA        1:CYS_27:C        1:TRP_28:N        1:TRP_28:CA       175.0
-175.0  100.00 100.00 1000.000
1:TRP_28:CA        1:TRP_28:C        1:CYS_29:N        1:CYS_29:CA       175.0
-175.0  100.00 100.00 1000.000
1:CYS_29:CA        1:CYS_29:C        1:PRO_30:N        1:PRO_30:CA       175.0
-175.0  100.00 100.00 1000.000
1:PRO_30:CA         1:PR_30:C        1:PHE_31:N        1:PHE_31:CA       175.0
-175.0  100.00 100.00 1000.000
1:PHE_31:CA        1:PHE_31:C        1:GLY_32:N        1:GLY_32:CA       175.0
-175.0  100.00 100.00 1000.000
1:GLY_32:CA        1:GLY_32:C        1:PHE_33:N        1:PHE_33:CA       175.0
-175.0  100.00 100.00 1000.000
1:PHE_33:CA        1:PHE_33:C        1:GLU_34:N        1:GLU_34:CA       175.0
-175.0  100.00 100.00 1000.000
1:GLU_34:CA        1:GLU_34:C        1:GLY_35:N        1:GLY_35:CA       175.0
-175.0  100.00 100.00 1000.000
1:GLY_35:CA        1:GLY_35:C        1:LYS_36:N        1:LYS_36:CA       175.0
-175.0  100.00 100.00 1000.000
1:LYS_36:CA        1:LYS_36:C        1:ASN_37:N        1:ASN_37:CA       175.0
-175.0  100.00 100.00 1000.000
1:ASN_37:CA        1:ASN_37:C        1:CYS_38:N        1:CYS_38:CA       175.0
-175.0  100.00 100.00 1000.000
1:CYS_38:CA        1:CYS_38:C        1:GLU_39:N        1:GLU_39:CA       175.0
-175.0  100.00 100.00 1000.000
1:GLU_39:CA        1:GLU_39:C        1:LEU_40:N        1:LEU_40:CA       175.0
-175.0  100.00 100.00 1000.000
1:LEU_40:CA        1:LEU_40:C        1:ASP_41:N        1:ASP_41:CA       175.0
-175.0  100.00 100.00 1000.000
1:ASP_41:CA        1:ASP_41:C        1:VAL_42:N        1:VAL_42:CA       175.0
-175.0  100.00 100.00 1000.000
1:VAL_42:CA        1:VAL_42:C        1:THRC_43:N       1:THRC_43:CA      175.0
-175.0  100.00 100.00 1000.000
```

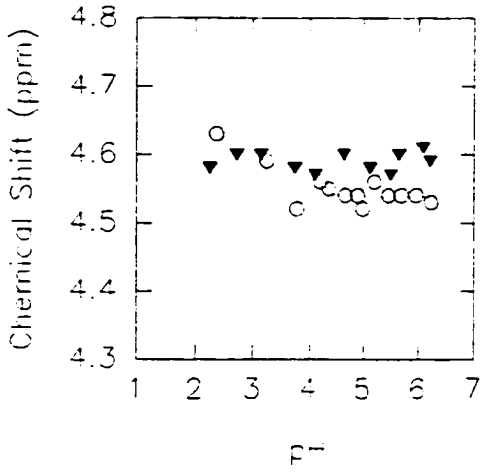
Appendix 4. Figures of titration curves of the pH-dependent residues in the absence and presence the calcium. The plain circles represent data collected in the absence of calcium, while the filled triangles represent data collected in the presence of calcium.

Appendix 4.1. pH Titration results of Asp3.

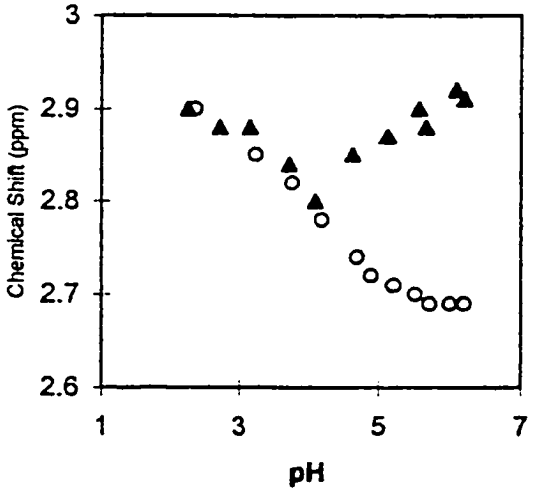
Asp3 NH



Asp3 αH

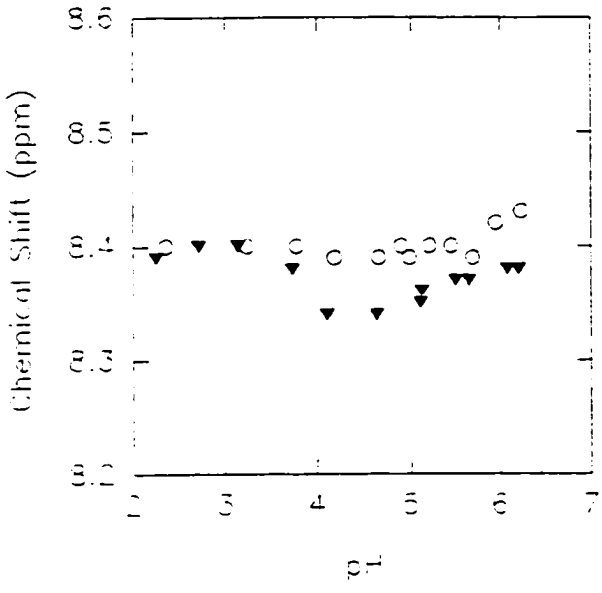


Asp3 βH

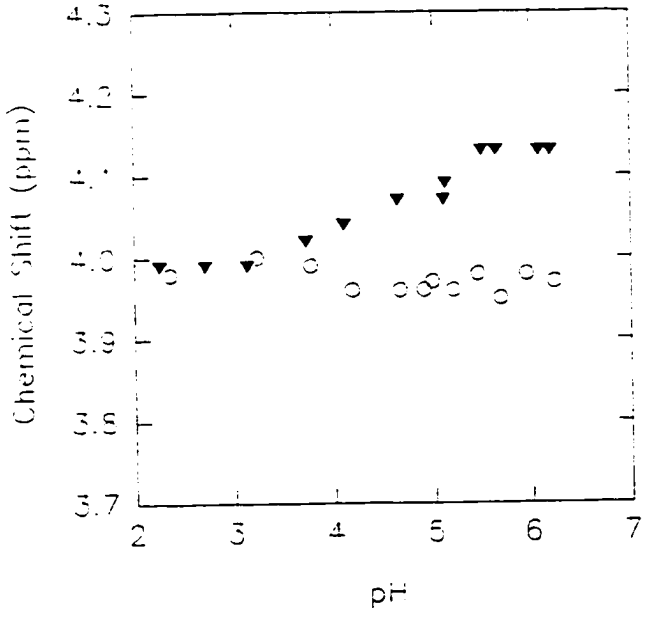


Appendix 4.2. pH Titration results of Gly4.

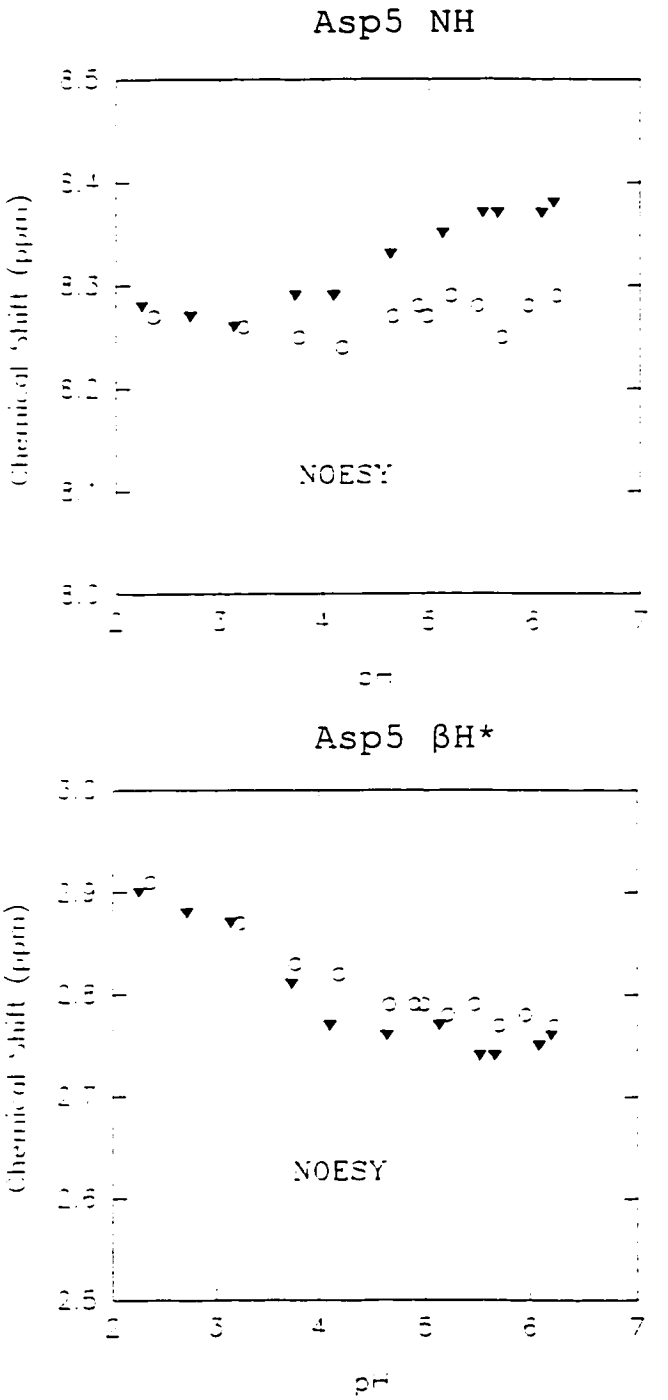
Gly4 NH



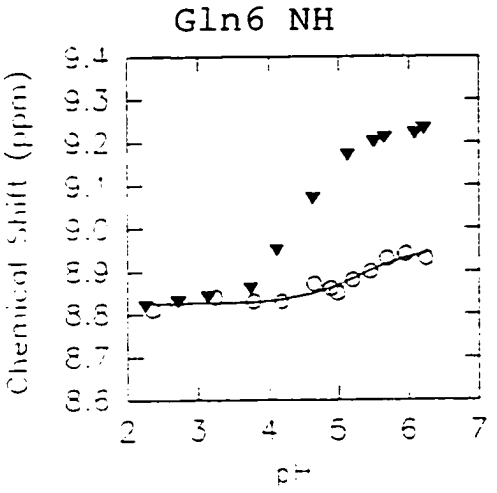
Gly4 αH

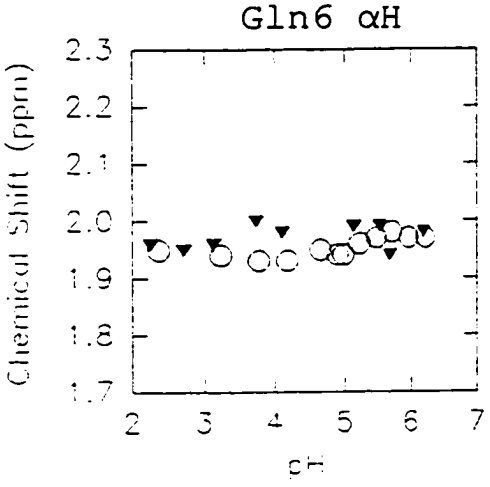


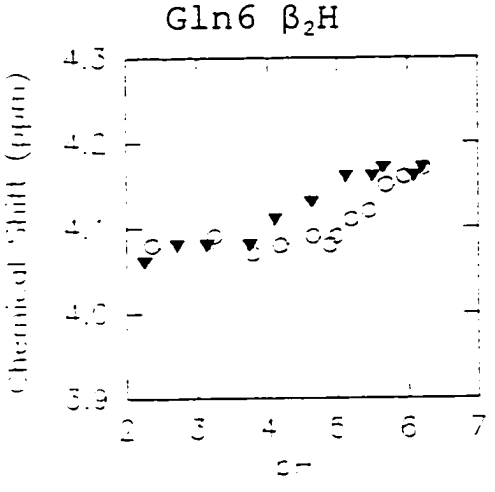
Appendix 4.3. pH Titration results of Asp5.

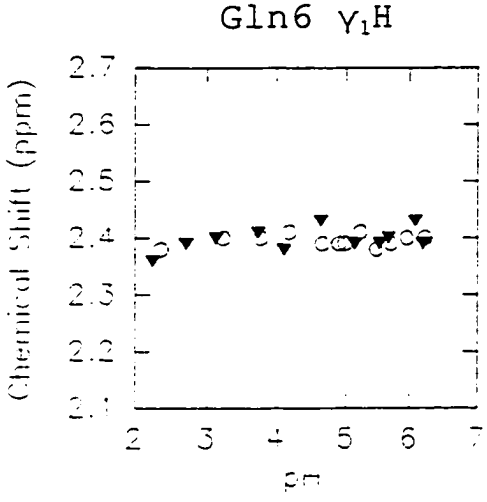


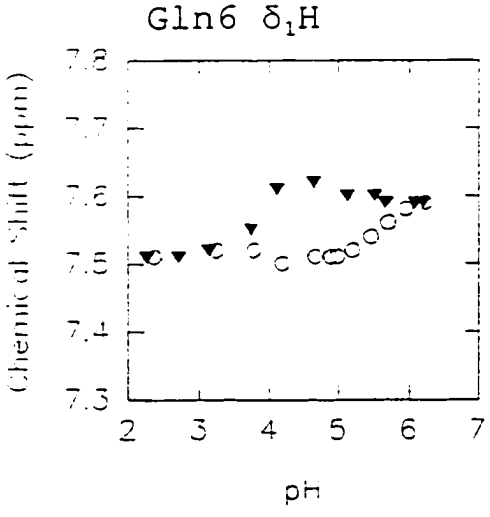
Appendix 4.4. pH Titration results of Gln6.

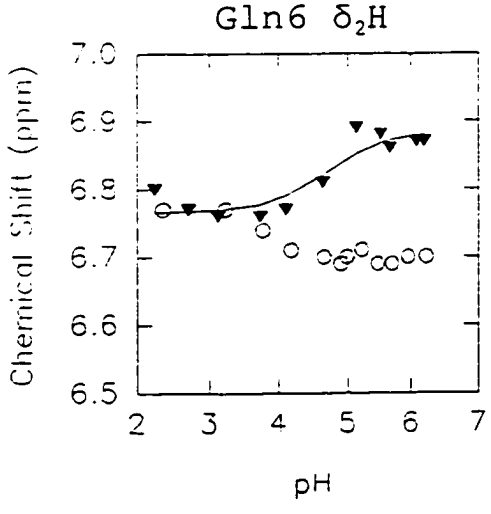






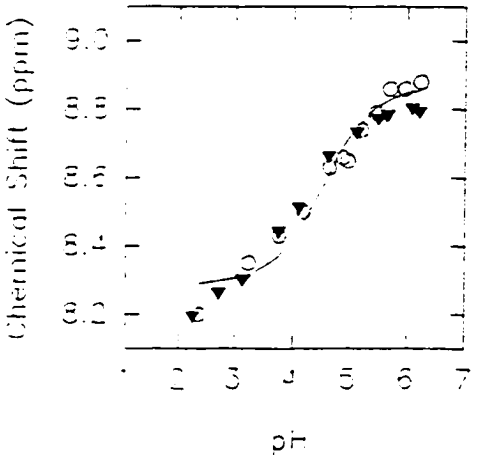




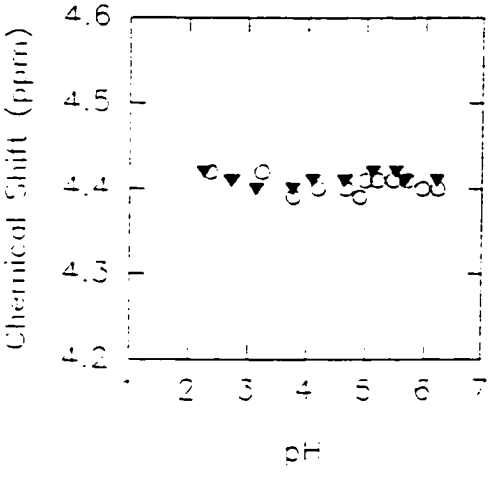


Appendix 4.5. pH Titration results of Cys7.

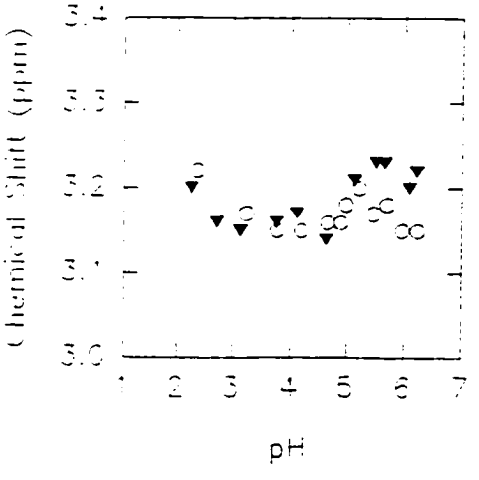
Cys7 NH



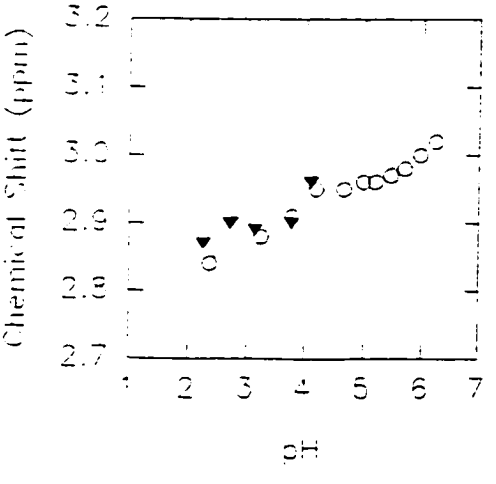
Cys7 αH



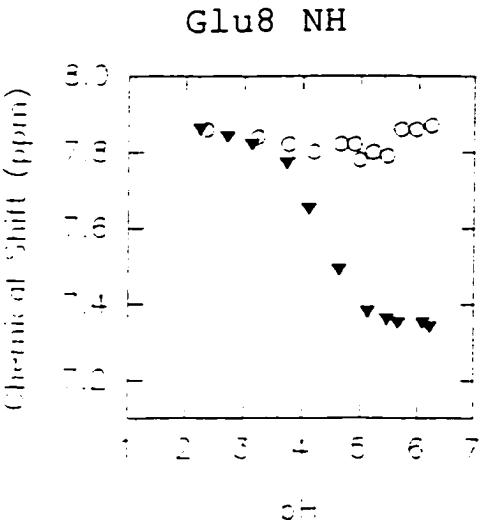
Cys7 β_1H

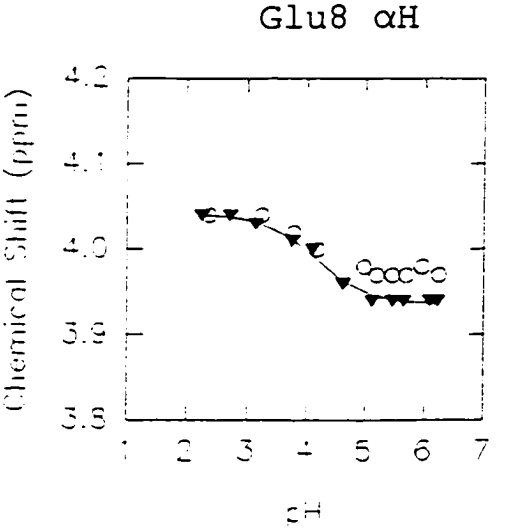


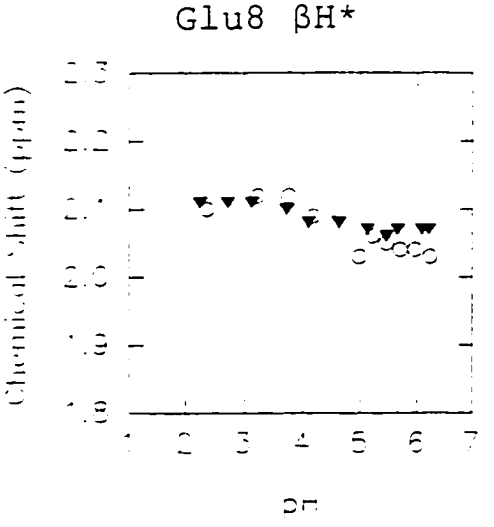
Cys7 β_2H

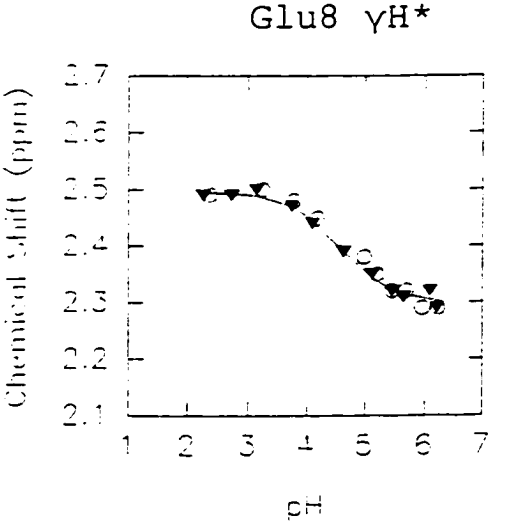


Appendix 4.6. pH Titration results of Glu8.



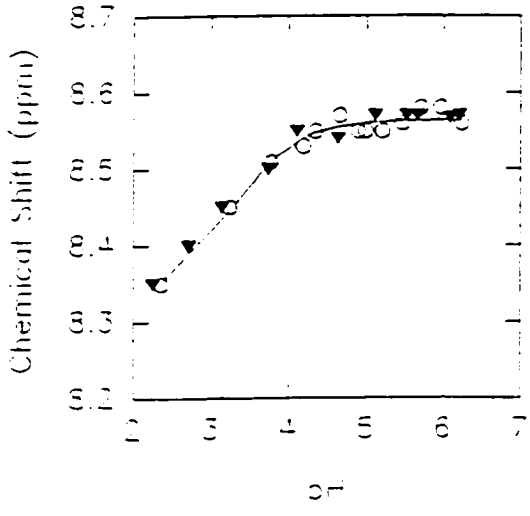




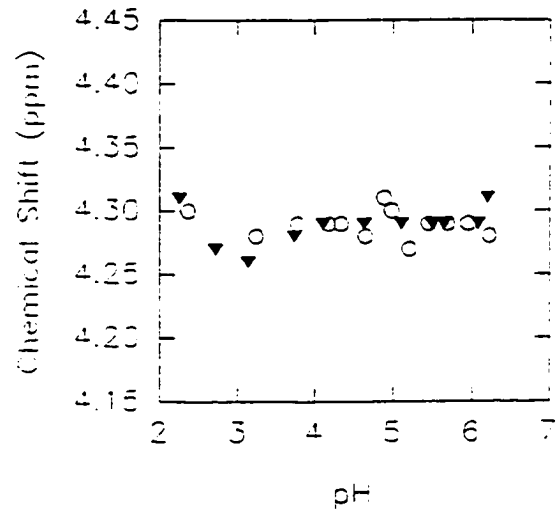


Appendix 4.7. pH Titration results of Leu13.

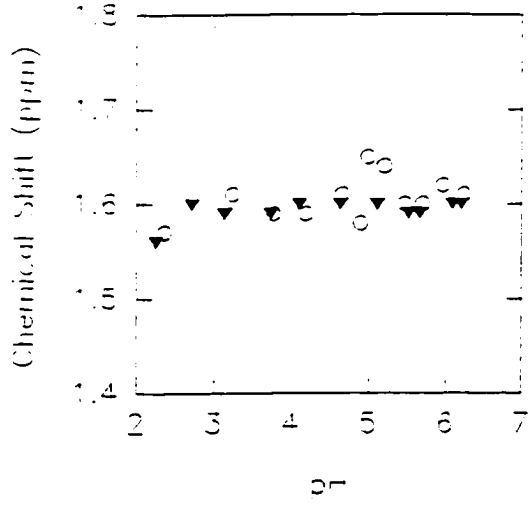
Leu13 NH



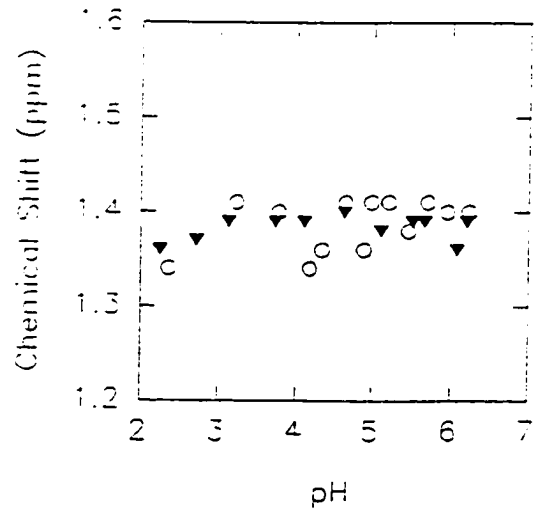
Leu13 αH



Leu13 β_1H

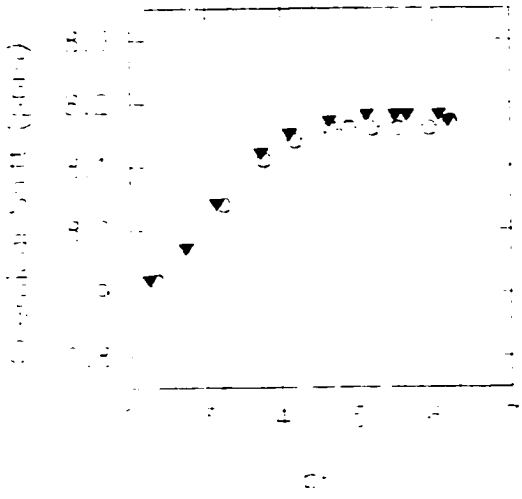


Leu13 β_2H

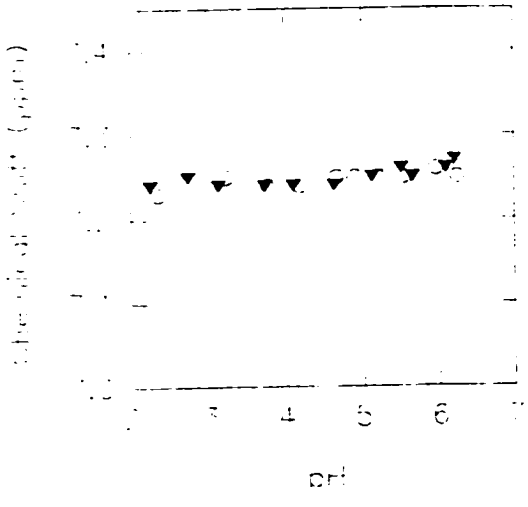


Appendix 4.8. pH Titration results of Asn14 conformation isomer.

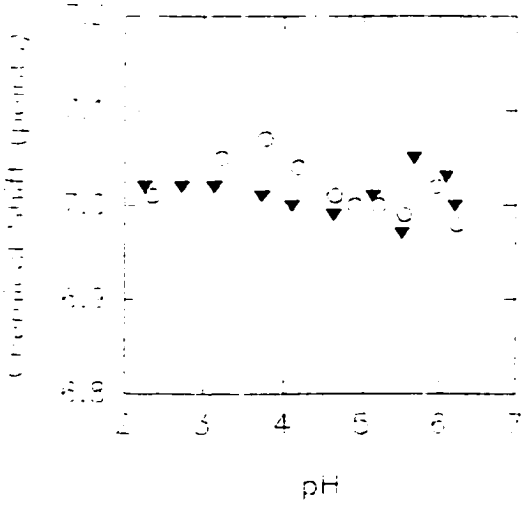
Asn14' NH



Asn14' δ_1NH

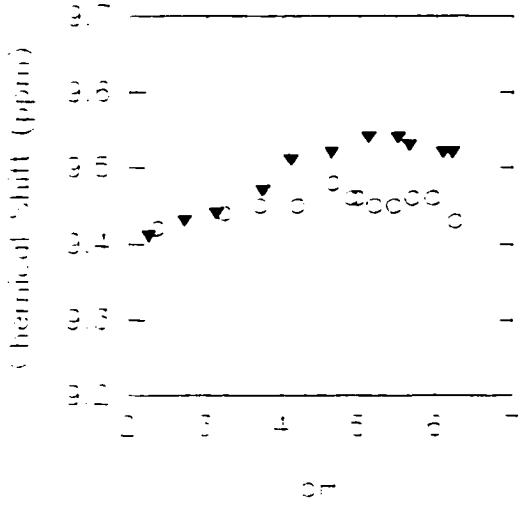


Asn14' δ_2NH

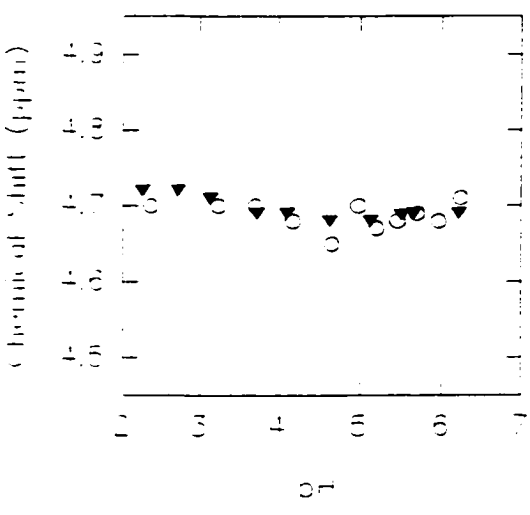


Appendix 4.9. pH Titration results of Lys19.

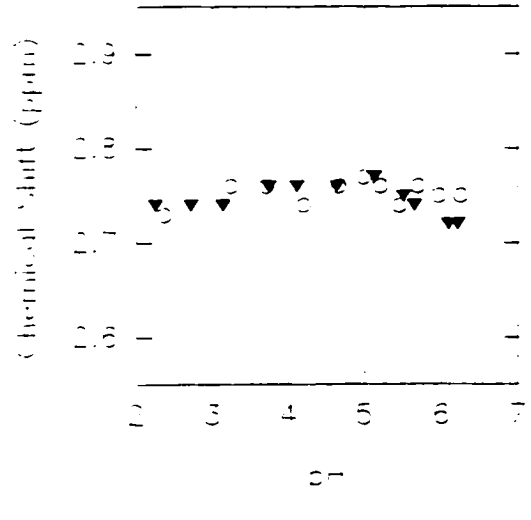
Lys19 NH



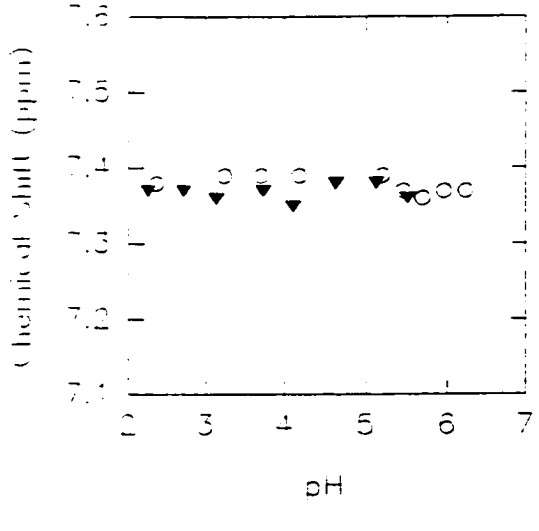
Lys19 αH



Lys19 εH*

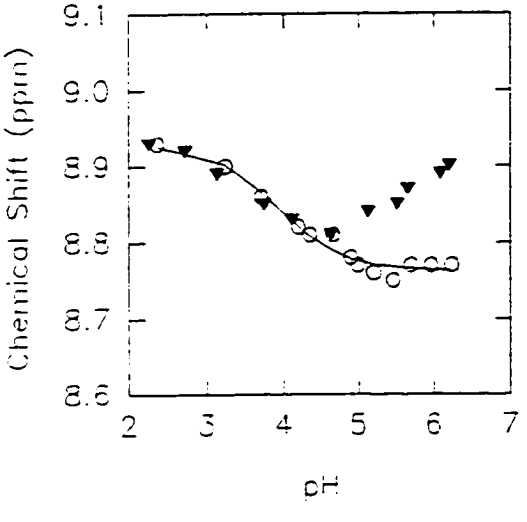


Lys19 NH_3

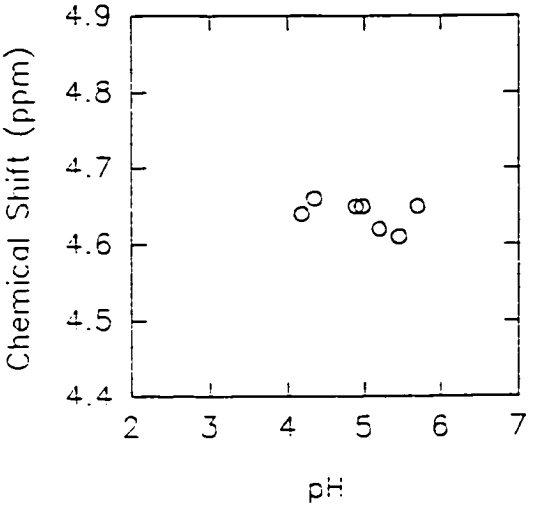


Appendix 4.10. pH Titration results of Asp20.

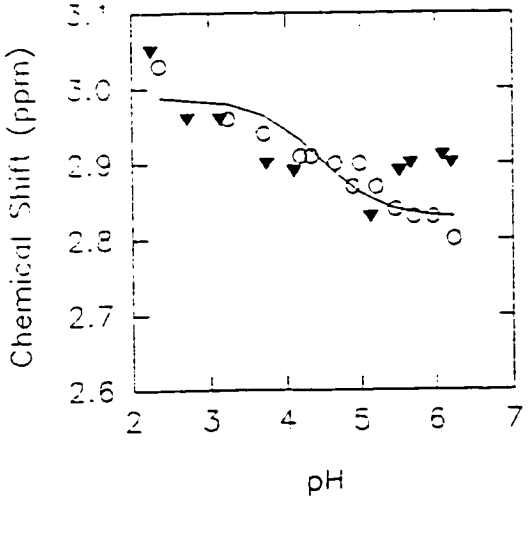
Asp20 NH



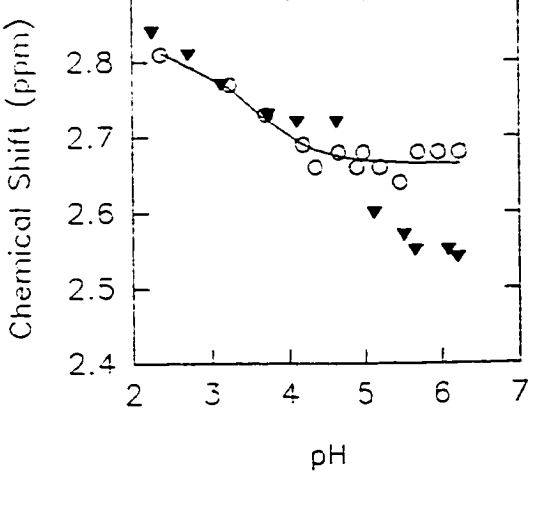
Asp20 αH



Asp20 β_1H

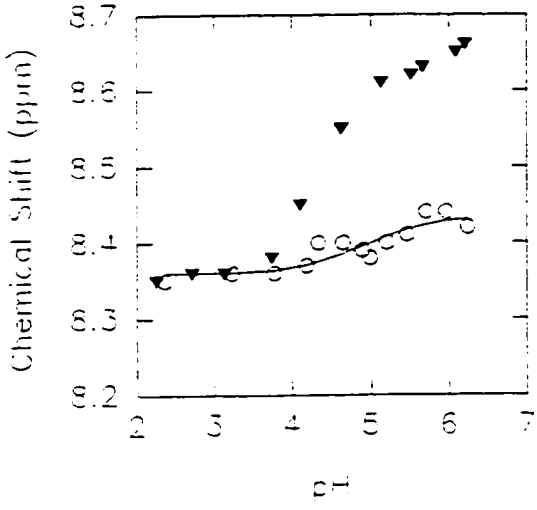


Asp20 β_2H

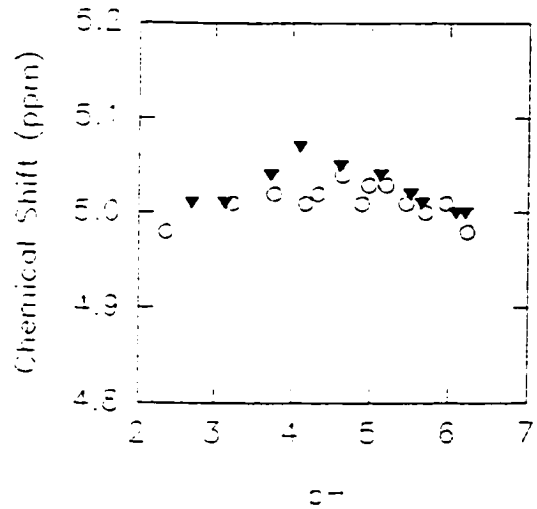


Appendix 4.11. pH Titration results of Asp21.

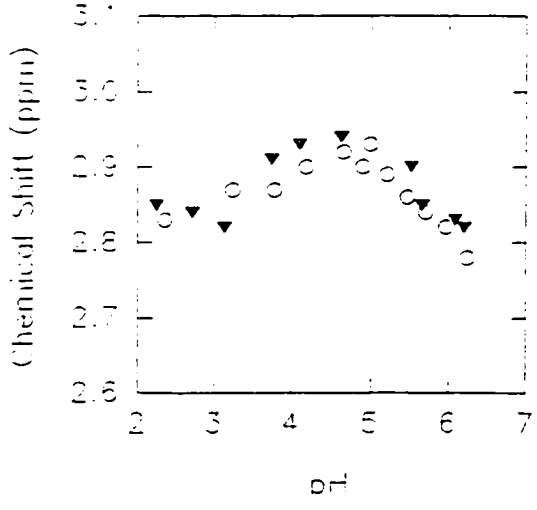
Asp21 NH



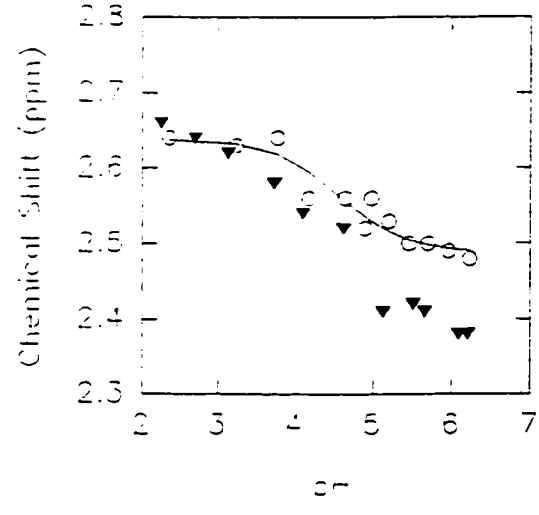
Asp21 αH



Asp21 β_1H

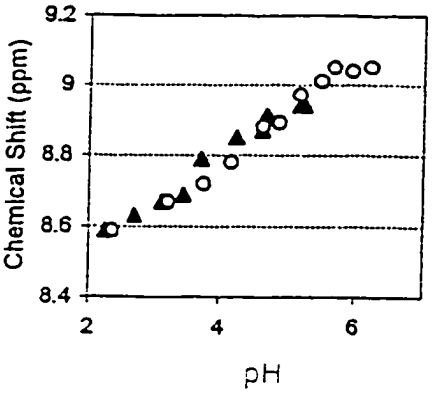


Asp21 β_2H

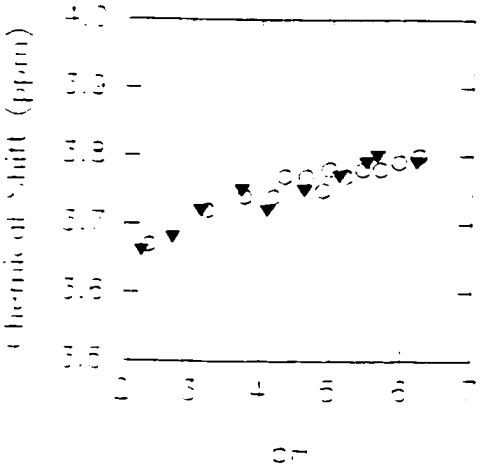


Appendix 4.12. pH Titration results of Ile22.

Ile22 NH

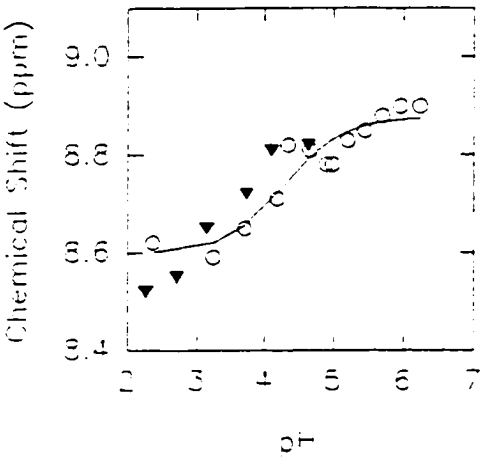


Ile22 αH

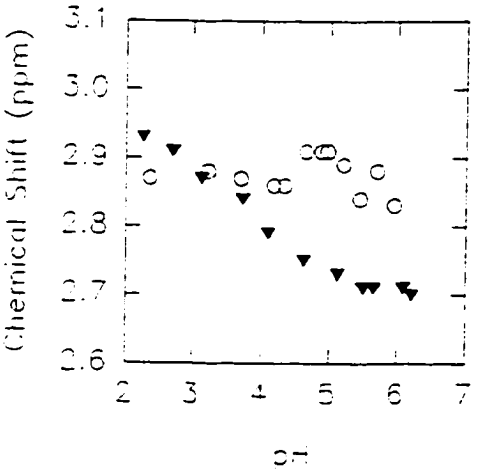


Appendix 4.13. pH Titration results of Asn23.

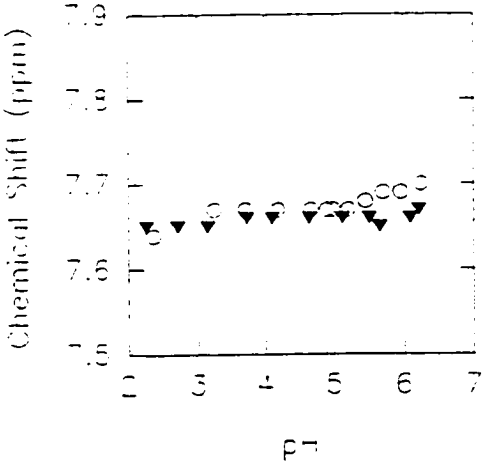
Asn23 NH



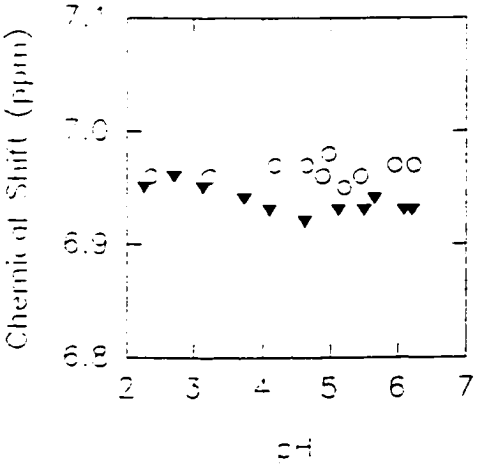
Asn23 βH*



Asn23 γ_1NH

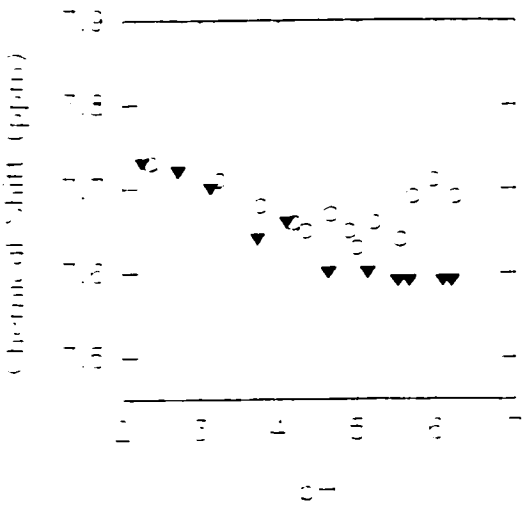


Asn23 γ_2NH

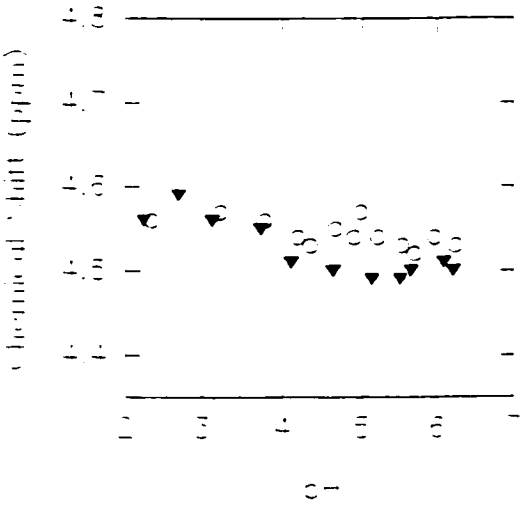


Appendix 4.14. pH Titration results of Ser24.

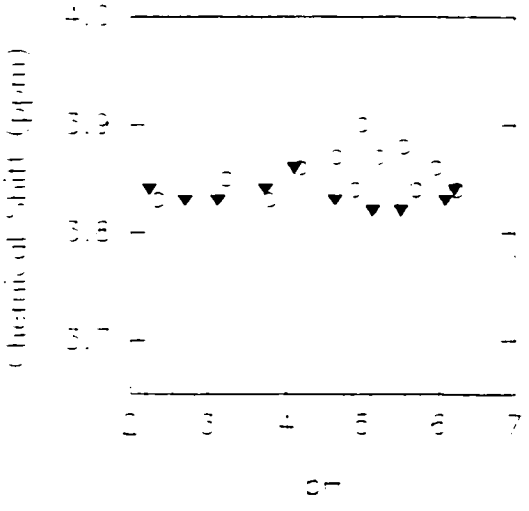
Ser24 NH



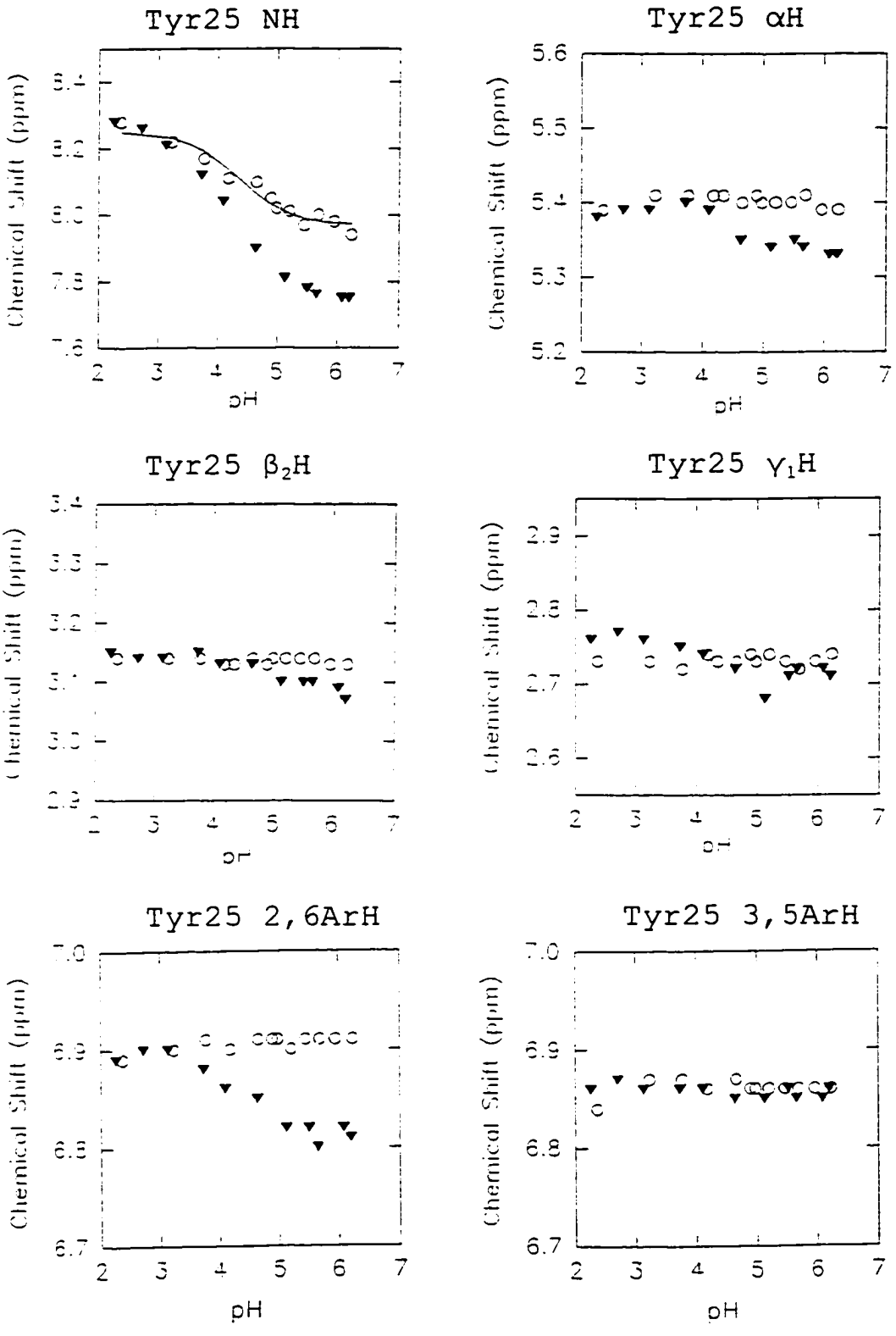
Ser24 αH



Ser24 βH*

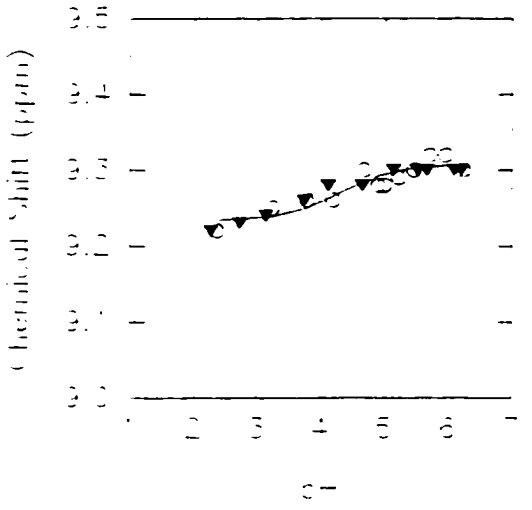


Appendix 4.15. pH Titration results of Tyr25.

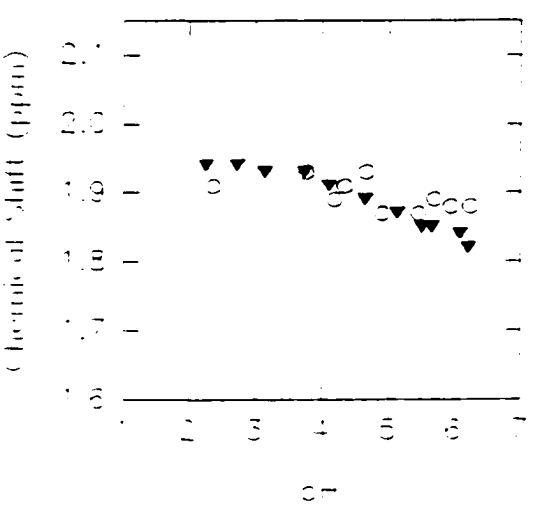


Appendix 4.16. pH Titration results of Glu34.

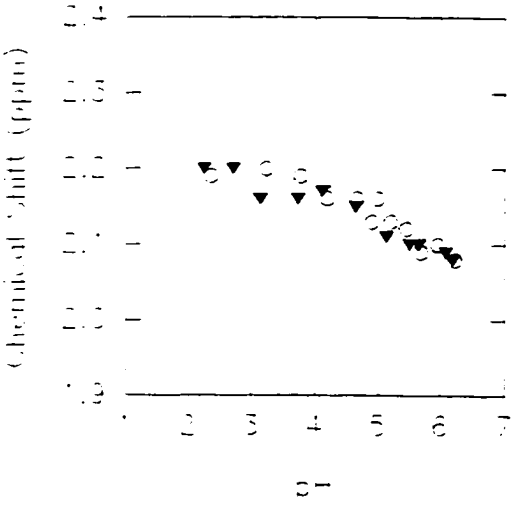
Glu34 NH



Glu34 βH*

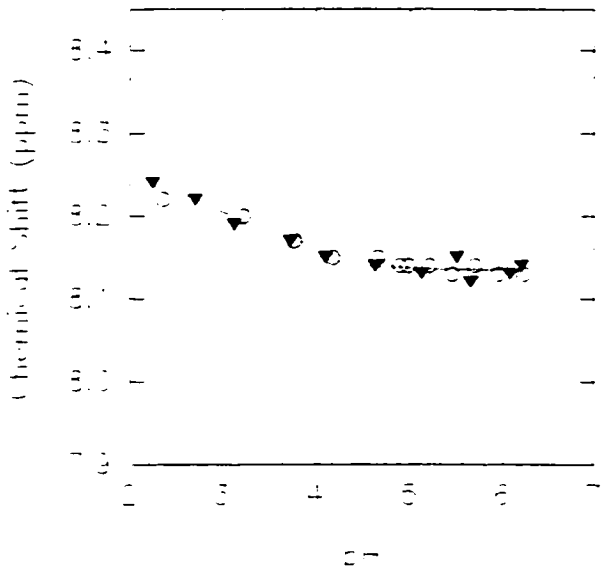


Glu34 γ_2H

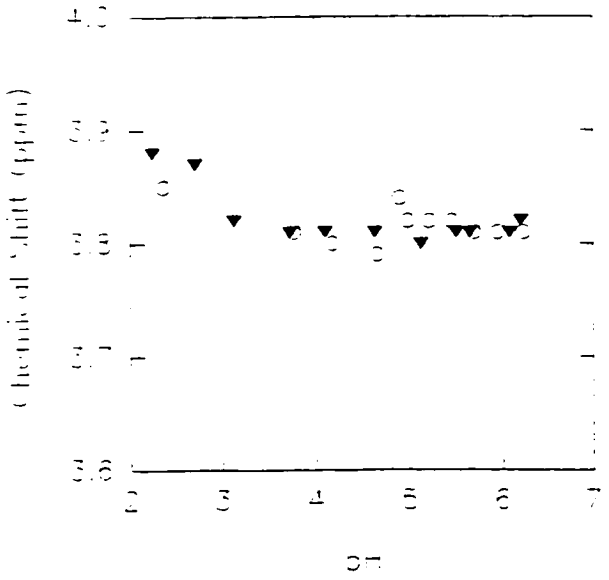


Appendix 4.17. pH Titration results of Gly35.

Gly35 NH

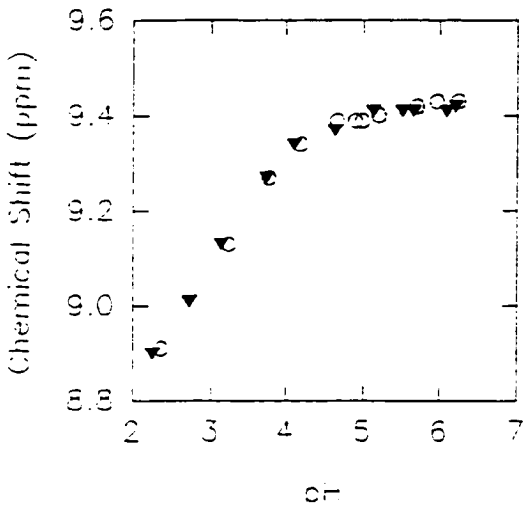


Gly35 α_2H

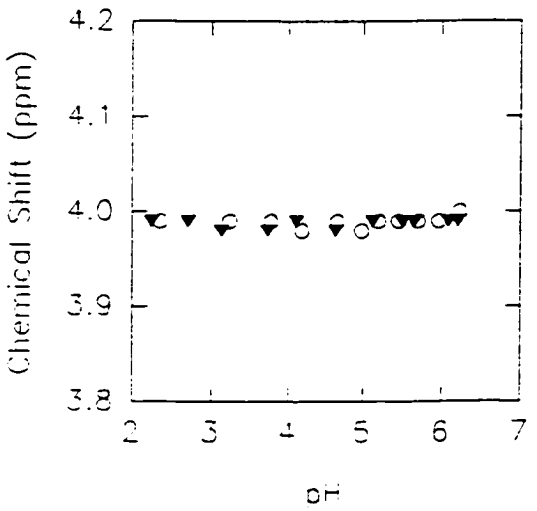


Appendix 4.18. pH Titration results of Lys36.

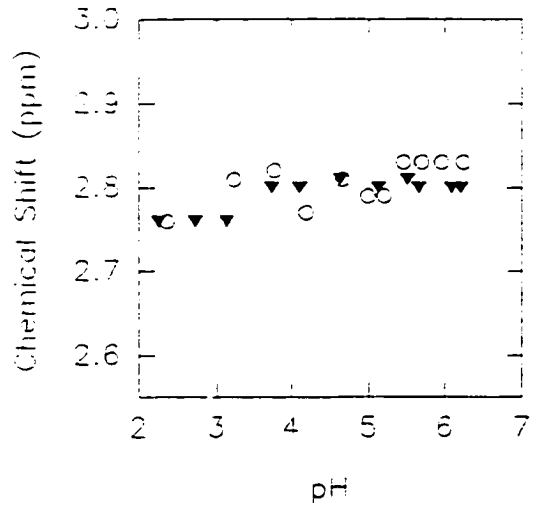
Lys36 NH



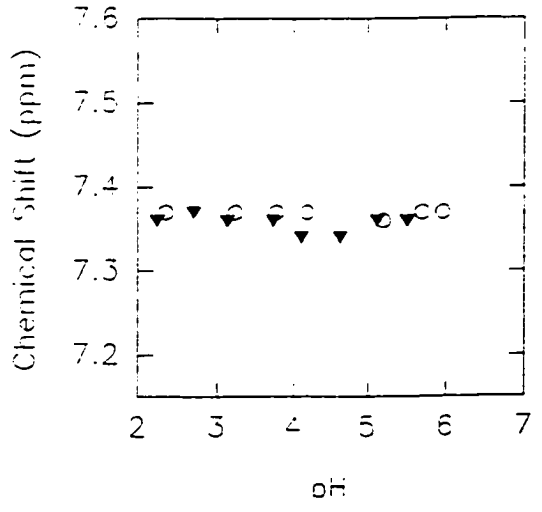
Lys36 αH



Lys36 εH*

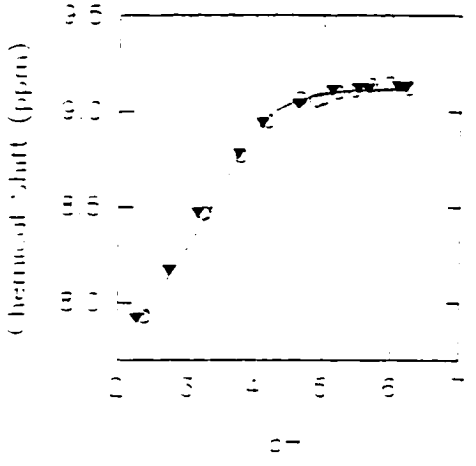


Lys36 NH_3

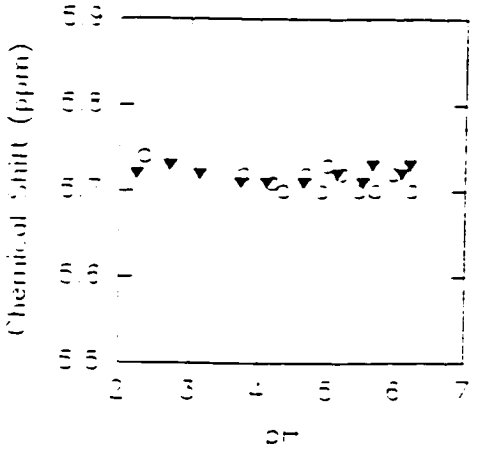


Appendix 4.19. pH Titration results of Asn37.

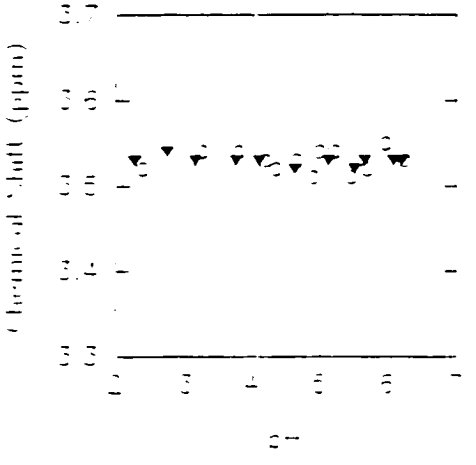
Asn37 NH



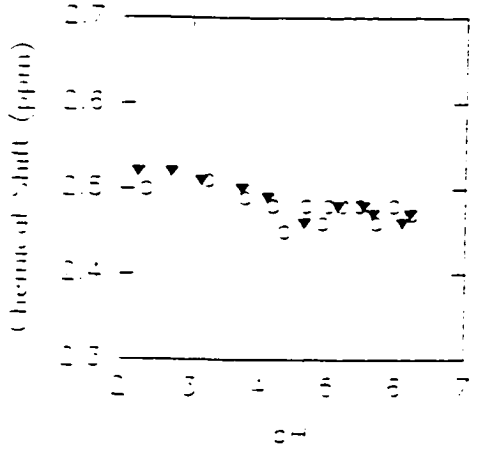
Asn37 αH



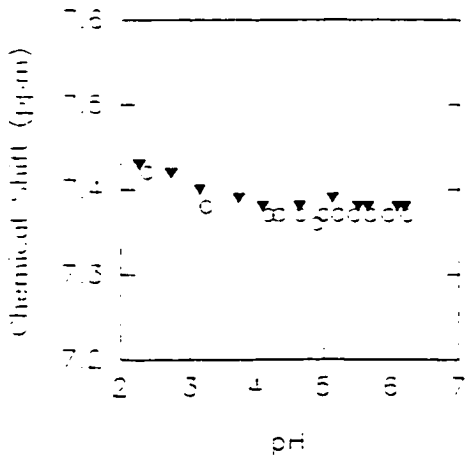
Asn37 β_1H



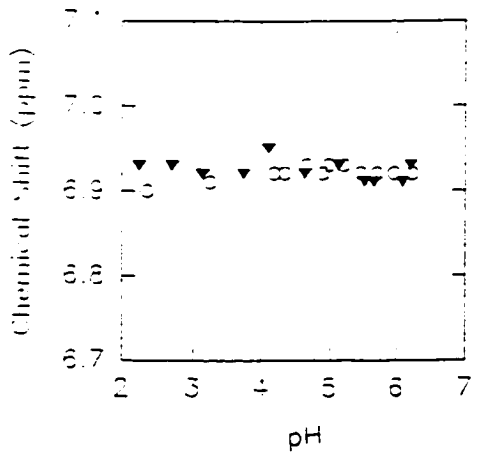
ASn37 β_2H



Asn37 γ_1NH

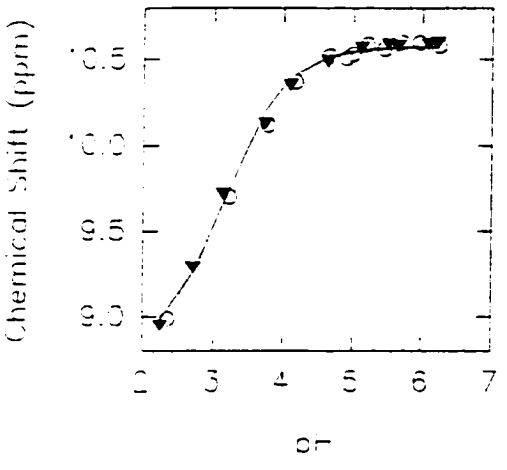


Asn37 γ_2NH

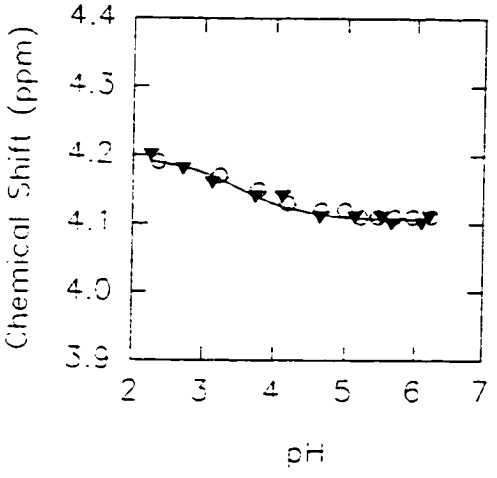


Appendix 4.20. pH Titration results of Glu39.

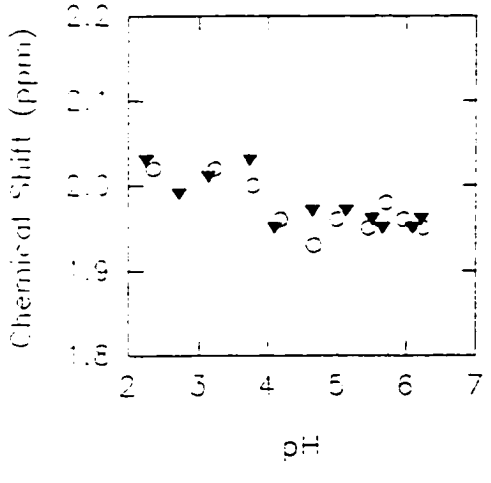
Glu39 NH



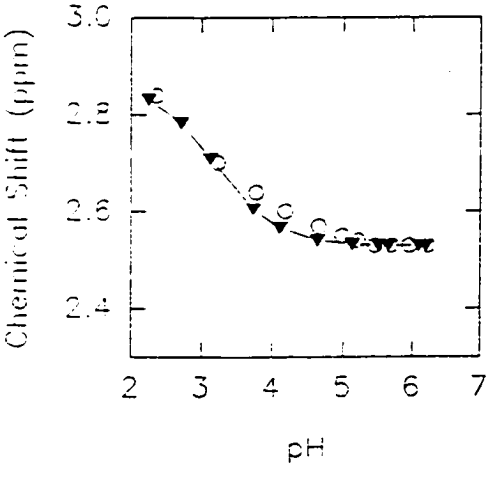
Glu39 αH



Glu39 β_2H

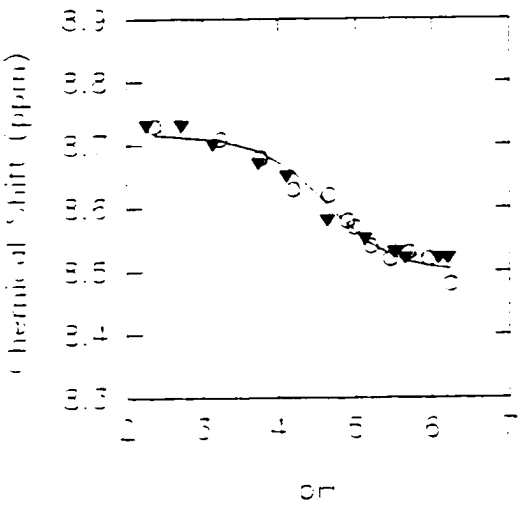


Glu39 γ_2H

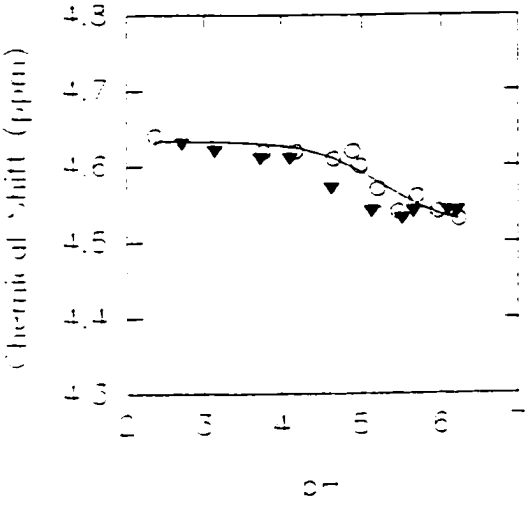


Appendix 4.21. pH Titration results of Asp41.

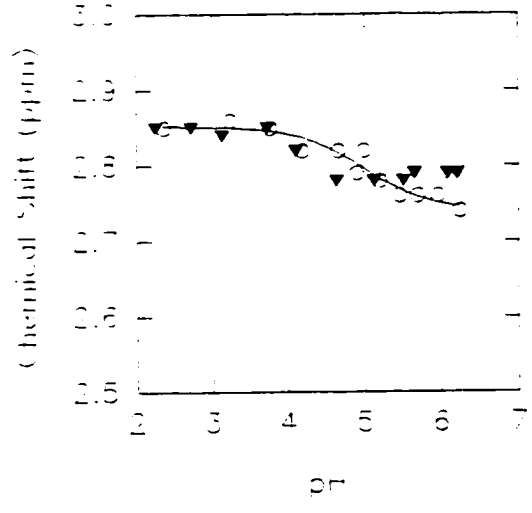
Asp41 NH



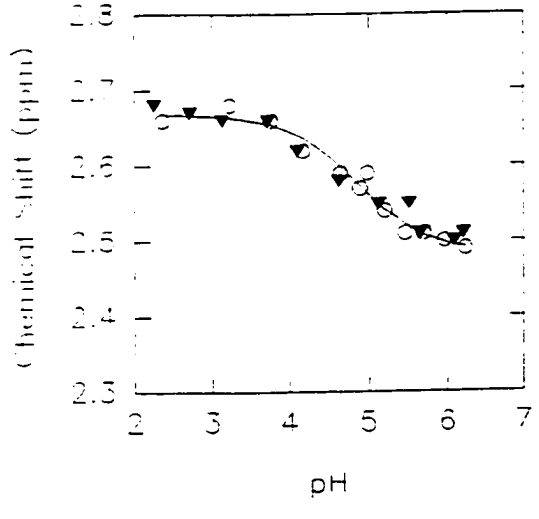
Asp41 αH



Asp41 β_1H

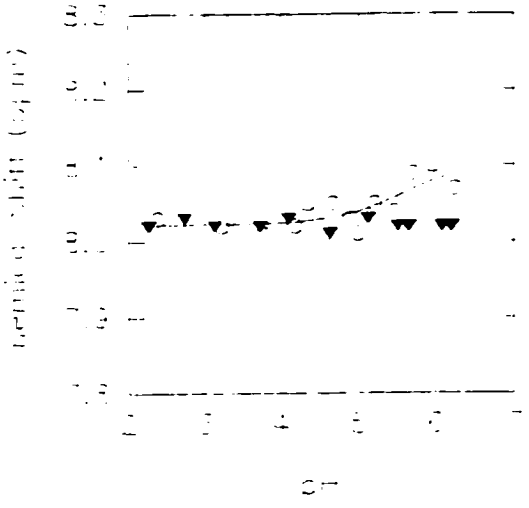


Asp41 β_2H

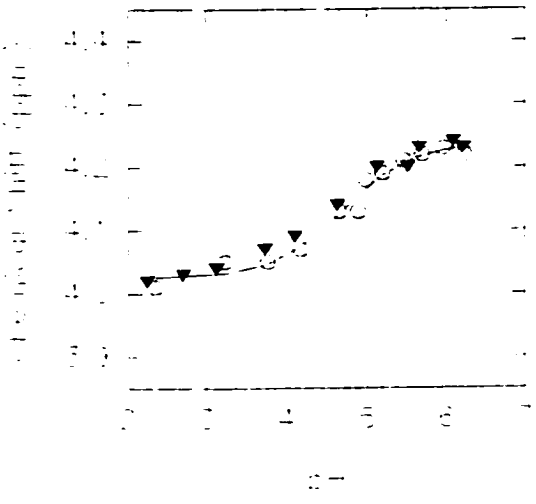


Appendix 4.22. pH Titration results of Val42.

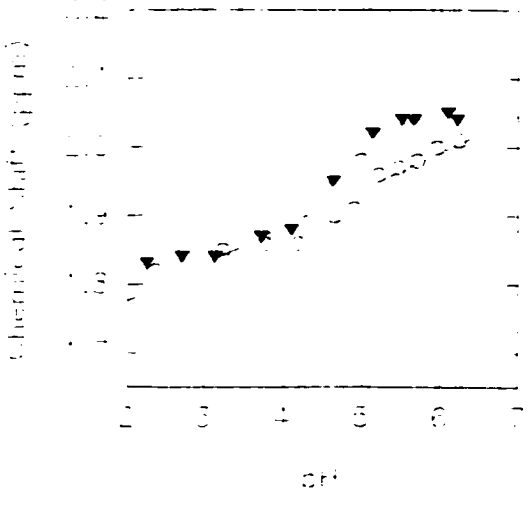
Val42 NH



Val42 αH



Val42 βH



References

Akke, M., Forsén, S. and Chazin, W. J. (1991) *J. Mol. Biol.* ***220***, 173-189.

Appella, E., Robinson, E. A., Ullrich, S. J., Stopelli, M. P., Corti, A., Cassani, G. and Blasi, F. (1987) *J. Biol. Chem.* ***262***, 4437-4440.

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

Astermark, J., Bjork, I., Ohlin, A. K., and Stenflo, J. (1991) *J. Biol. Chem.* ***266***, 2430.

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

Basus, V. J. (1989) *Methods in Enzymology* ***177***, 132-149.

Bersch, B., Hernandez, J. F., Marion, D., Arlaud, G. J. (1998) *Biochemistry* ***37***, 1204-1214.

Braun, W., Bösch, C., Brown, L. R., Gô, N., and Wüthrich, K. (1981) *Biochim. Biophys. Acta* ***667***, 377.

Braun, W., Wider, G., Lee, K. H., and Wüthrich, K. (1983) *J. Mol. Biol.* ***169***, 921.

Brown, S. C., Mueller, L., and Jeffs, P. W. (1989) *Biochemistry* ***28***, 593-599.

Bundi, A., and Wüthrich, K. (1979) *Biopolymers* ***18***, 299-311.

Campbell, I. D., and Bork, P. (1993) *Curr. Opin. Struct. Biol.* ***3***, 385-392.

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

Campion, S. R., Matsunami, R. K., Engler, D. A., and Niyogi, S. K. (1990) *Biochemistry* ***29***, 9988-9993.

Carver, J. A., Cook, R. M., Esposito, G., Campbell, I. D., Gregory, H., and Sheard, B. (1986) *FEBS Lett.* ***205***, 77-81.

Charles R. Cantor and Paul R. Schimmel, *Biophysical*

Chemistry, 1980, p50, W. H. Freeman and Company, San Francisco.

Chothia, C. (1973) *J. Mol. Biol.* ***75***, 295-302.

Chothia, C., Levitt, M., and Richardson, D. (1977) *Proc. Natl. Acad. Sci. U.S.A.* ***74***, 4130-4134.

Chou, P. Y., and Fasman, G. D. (1977), *J. Mol. Biol.* ***115***, 135-175.

Cook, R.M., Wilkinson, A.J., Baron, M., Pardy, A., Tappin, M. J., Campbell, I. D. and Sheard, B., (1987) *Nature* ***327***, 339-341.

Corson, G. M., Chalberg, S. C., Dietz, H. C., Charboneau, N. L., and Sakai, L. S. (1993) *Genomics* ***17***, 476-484.

Dahlbäck, B., Hildebrand, B., and Linse, S. (1990) *J. Biol. Chem.* ***265***, 18481-18489.

Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M., Hagler, A. T. (1988) *Proteins: structure, Function and Genetics* ***4***, 31-47.

Davie, E. W., Fujikawa, K., and Kisiel, W. (1991) *Biochemistry* ***30***, 10363-10370.

Davis, L. M., McGraw, R. A., Ware, J. L., Robert, H. and Stafford, D. W. (1987) *Blood* ***69***, 140-143.

de Marco, A., Llinás, M., and Wüthrich, K., (1978) *Biopolymers* ***17***, 617-636.

Dietz, H. C., McIntosh, I., Sakai, L. Y., Corson, G. M., Chalberg, S. C, Pyeritz, R. E., and Francomano, C. (1993) *Genomics* ***17***, 468-475.

Driscoll, P. C., Gronenborn, A. M., Beress, L., and Clore, G. M., (1989) *Biochemistry* ***28***, 2188-2198.

Engel, J. (1989) *FEBS Lett.* ***251***, 1-7.

Engler, D. A., Campion, S. R., Hauser, M. R., Cook, J. S., and Niyogi, S. K. (1992) *J. Biol. Chem.* ***267***, 2274-2281.

Engler, D. A., Matsunami, R. K., Campion, S. R., Stringer, C. D., Stevens, A., and Niyogi, S. K. (1988) *J. Biol. Chem.* ***263***, 12384-12390.

Engler, D. A., Montelione, G. T., and Niyogi, S. K.

(1990) *FEBS Lett.* ***271***, 47-50.

Fehon, R. G., Kooh, P. J., Rebay, I., Regan, C. L., Xu, T., Muskavitch, M. A. T., and Artavanis-Tsakonas, S. (1990) *Cell* ***61***, 523-534.

Fleming, R. G., Scottgale, T. N., Diederich, R. J., and Artavanis-Tsakonas, S. (1990) *Genes & Dev.* ***4***, 2188-2201. criterion.

Forman-Kay, J. D., Clore, G. M., and Gronenborn, A. M. (1992) *Biochemistry* ***31***, 3442-3452.

Giannelli, F., Green, P. M., Sommer, S. S., Lillicrap, D. P., Ludwig, M., Schwaab, R., Reitsma, P. H., Goossens, M., Yoshioka, A., and Brownlee, G. G. (1994). *Nucl. Acids Res.* ***22***, 3534-3546.

Glanville, R. W., Qian, R. Q., McClure, D. W., and Maslen, C. L. (1994) *J. Biol. Chem.* ***269***, 26630-26634.

Green, P.M., Bentley, D. R., Mibashan, R. S., Nilsson, I. M. and Gianeli, F. (1989) *EMBO J.* ***8***, 1067-1072.

Groenen, L. C., Nice, E. C., and Burgess, A. W. (1994) *Growth Factor* ***11***, 235-257.

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

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

Handford, P., Downing, A. K., Rao. Z., Hewett, D. R., Sykes, B. C., and Kielty, C. M. (1995) *J. Biol. Chem.* ***270***, 6751-6756.

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

Havel, T. F., (1991) *Prog. Biophys. Molec. Biol.* ***56***, 43-78.

Havel, T. F., and Wüthrich, K. (1984) *Bull. Math. Biol.* ***46***, 673.

Henikoff, S., and Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA.* ***89***, 10915-10919.

Hewett, D. R., Lynch, J. R., Smith, R., and Sykes, B. C. (1993) *Hum. Mol. Genet.* ***2***, 475-477.

Higashiyama S., and Klagsbrun, K. (1993) *Growth Factors, Pept., Recept.*, 93-100. Eidtor: Moody, T. W. Publisher: Plenum, New York, N.Y.

Higashiyama, S., and Taniguchi, N. (1994) *Gann Monogr. Cancer Res.* ***42***, 23-35.

Höjrup, P., and Magnusson, S., (1987) *Biochem. J.* ***245***, 887.

Huang, L. H., Thesis, 1993.

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

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

Hughes, P. E., Morgan, G., Rooney, E. K., Browlee, G. G., and Handford, P., (1993) *J. Mol. Biol.* ***268***, 17727-17733.

Jacobsen, N. E., Abadi, N., Sliwkowski, M. X., Reilly, D., Skelton, N. J., and Fairbrother, W. J. (1996) *Biochemistry* ***35***, 3402-3417.

Janin, J., Wodak, S., Levitt, M., and Maigrait, B. (1978) *J. Mol. Biol.* ***125***, 357.

Kainulainen, K., Karttunen, L., Puhakka, L., Sakai, L., and Peltonen, L. (1994) *Nat. Gene.* ***6***, 64-69.

Kline, T. P., Brown, S. C., Jeffs, P. W., Kopple, K. D., and Mueller, L. (1990) *Biochemistry* ***29***, 7805-7813.

Knott, V., Downing, A. K., Cardy, C. M., and Handford, P. (1996) *J. Mol. Biol.* ***255***, 22-27.

Koide, H., Muto, Y., Kasai, H., Kohri, K., Hoshi, K., Takahashi, S., Tsukumo, K., Sasaki, T., Oka, T., and Miyake, T. (1992) *Biochim. Biophys. Acta.* ***1120***, 257-261.

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

Konda, D., and Inagaki, F. (1992a) *Biochemistry* ***31***, 11928-11939.

Konda, D., and Inagaki, F. (1992b) *Biochemistry* ***31***, 677-685.

Kohda, D., Shimada, I., Miyake, T., Fuwa, T. and Inagaki, F., (1989) *Biochemistry* ***28***, 953-958.

Komoriya, A., Hortsch, M., Meyers, C., Smith, M., Kanety, H., and Schlessinger, J. (1984) *FEBS Lett.* ***231***, 1-4.

Konda, D., Gö, N., Hayashi, K., and Inagaki, F. (1988) *J. Biochem. (Tokyo)* ***103***, 741-743.

Konda, D., Shimada, I., Miyake, T., Fuwa, T., and Inagaki, F. (1989) *Biochemistry* ***28***, 953-958.

Kurosawa, S., Stearns, D. J., Jackson, K. W., and Esmon, C. T. (1988) *J. Biol. Chem.* ***263***, 5993-5996.

Lee, B., Maurice, G., Vitale, E., Hori, H., Mattei, M. G., Sarfari, M., Tsipouras, P., Ramirez, F., and Hollister, D. W. (1991) *Nature* ***352***, 330-334.

Lin, Y-Z., CaporSO, g., Chang, P-Y., Ke, X-H, and Tam, J. P. (1988) *Biochemistry* ***27***, 5640-5645.

Lozier, J. N., Monroe, D. M., Stanfield-Oakley, S., Lin, S. W., Smith, K. J., Roberts, H., R., and High, K. A. (1990) *Blood* ***74***, 1097-1104.

Marquardt, H. and Todaro, G. J. (1982) *J. Biol. Chem.* ***257***, 5220-5225.

Maslen, C. L., Corson, G. M., Maddox, B. K., Glanville, R. K., and Sakai, L. Y. (1991) *Nature* ***352***, 334-337.

Maslen, C. L., Qian, R-G., McClure, D. W., and Glanville, R. W. (1993) *9th Annual Conference of National Marfan Foundation* (Abstr. 4) Portland, OR.

Massagué J. (1990) *J. Biol. Chem.* ***265***, 21393.

Mayhew, M., Handford, P., A., Baron, M., Tse, A., Campbell, I. D., and Brownlee, G. G. (1992) *Protein Eng.* ***5***, 489-494.

Mayo, K. H., Cavalli, R. C., Peters, A. R., Boelens, R., and Kaptein, R. (1989) *Biochem. J.* ***257***, 197-205.

McCord, D. M., Monroe, D. M., Smith, K. J. and Robert, H. (1990) *J. Biol. Chem.* ***265***, 10250-10254.

Meininger, D. P., Hunter, M. J., and Komives, E. A. (1995) *Protein Sci.* ***4***, 1683-1695.

Montelione, G. T., Wüthrich, K., Nice, E. C., Burgess, A. W., and Scheraga, H. A. (1986) *Proc. Natl. Acad. Sci. USA.* ***83***, 8594-8598.

Montelione, G. T., Wüthrich, K., Nice, E. C., Burgess, A. W., and Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. U. S. A.* ***84***, 5226-5230.

Montelione, G. T., Wüthrich, K., and Scheraga, H. A. (1988) *Biochem.* ***27***, 2235-2243.

Montelione, G. T., Winkler, M. E., Burton, L. E., Rinderknecht, E., Sporn, M. B., and Wagner, G. (1989) *Proc. Natl. Acad. Sci. U. S. A.* ***86***, 1519-1523.

Montelione, G. T., Wüthrich, K., Burgess, A. W., Nice, E. C., Wagner, G., Gibson, K. D. and Scheraga, H. A., (1992) *Biochemistry* ***31***, 236-249.

Moy, F. J., Li, Y. C., Rauenbuehler, P., Winkler, M. E., Scheraga, H. A., and Montelione, G. T. (1993) *Biochemistry* ***32***, 7334-7353.

Nemethy, G., and Printx, M. P. (1972) *Macromolecules* ***5***, 755-758.

Öhlin, A. K., and Stenflo, J., (1987) *J. Biol. Chem.* ***262***, 13798-13807.

Öhlin, A. K., Linse, S., and Stenflo, J. (1988) *J. Biol. Chem.* ***263***, 7411-7417.

Pardi, A., Billeter, M., and Wüthrich, K. (1984) *J. Mol. Biol.* ***180***, 741-751.

Patty, L. (1991) *Curr. Opinion Struct. Biol.* ***1***, 351-361.

Persson, E., Selander, M., Linse, S., Drakenberg, T., Ohline, A. K., and Stenflo, J. (1989) *J. Biol. Chem.* ***264***, 16897-16904.

Persson, E., Bjork, I., and Stenflo, J. (1991) *J. Biol. Chem.* ***266***, 2444.

Qi, X. P., Stefano, D. L. D., and Wand, A. J. (1994) *Biochemistry* ***33***, 6408-6417.

Quiocho, F. A., Gilliland, G. L., and Phillips, G.

N., Jr. (1977) *J. Biol. Chem.* ***252***, 5142-5149.

Ramachadran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963) *J. Mol. Biol.* ***7***, 95-99.

Ramachandran, G. N. and Sasisekharan, V, (1968) *Adv. Protein Chem.* ***23***, 284-438.

Rao, Z., Handford, P. A., Mayhew, M., Knott, V., Brownlee, G. G. and Stuart, D. (1995) *Cell* ***82***, 131-141.

Rebay, I., Fleming, R. J., Fehon, R. G., Cherbas, L., Cherbas, P., and Artavanis-Tsakonas, S. (1991) *Cell* ***67***, 687-699.

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

Richardson, J. S. (1981) *Adv. Protein Chem.* ***34***, 167-339.

Rose, G. D., Gierasch, L. M., and Smith, J. A. (1985) *Avd. Protein Chem.* ***37***, 1-109

Roterman, I. K., Lambert, M. H., Gibson, K. D., and Scheraga, H. A. (1989) *J. Biomol. Struci. Dyn.* ***7***, 421-453.

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

Sceenk, R. M., van Gunsteren, W. F., and Kaptein, R. (1989) *Methods in Enzymology,* ***177***, 204.

Schultz, G. E., Elzinga, M., Marx, F., and Schirmer, R. H. (1974) *Nature (London)* ***250***, 120-123.

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

Selander-Sunnerhagen, M., Ullnert, M., Persson, E., Teleman, O., Stenflo, J. and Drakenberg, T. (1992) *J. Biol. Chem.* ***267***, 19642-19649.

Selander-Sunnerhagen, 1994. Ph.D thesis, University of Lund, Sweden.

Shaw, P. J., and Muirhead, H. (1977) *J. Mol. Biol.* ***109***, 475-485.

Singson, A., Mercer, K. B., and L'Hernault, S. W.

(1998) *Cell* ***93***, 71-79.

States, D. J., Harberkorn, R. A., and Reuben, D. J. (1982) *J. Magn. Reson.* ***48***, 286-292.

Stenflo, J., (1991) *Blood* ***78***, 1637-1651.

Stroobant, P., Rice, A. R., Gullick, W. J., Cheng, D. J., Kerr, I. M., and Waterfield, M. D. (1985) *Cell* ***42***, 383-393.

Suzuki, K., Hayashi, T., Nishioka, J., Kosaka, Y., Zushi, M., Honda, G., and Yamamoto, S. (1989) *J. Mol. Biol.* ***264***, 4872-4876.

Tadaki, D. K., and Niyogi, S. K. (1993) *J. Biol. Chem.* ***268***, 10113-10119.

Tappin, M. J., Cooke, R. M., Fitton, J. E., and Campbell, I. D., (1989) *Eur, J. Biochem.* ***179***, 629-637.

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

Vässin, H., Bremer, K. A., Knust, E., and Campos-Ortega, J. A. (1987) *EMBO J.* ***6***, 3431-3440.

Venkatachalam, C. M. (1968) *Biopolymers 6*, 1425-1436.

Voet, D., and Veot, J. (1995), Biochemistry, the 2nd Ed., John Wiley & Sons, Inc., New York.

Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Gõ, N. and Wüthrich, K. (1987) *J. Mol. Biol.* ***196***, 611-639.

Weber, P. L., Morrison R. and Hare, D. (1988) *J. Mol. Biol.* ***204***, 483-487.

Wharton, K. A., Johansen, K. M., Xu, T., and Artavanis-Tsakonas, S. (1985) *Cell* ***43***, 567-581.

White, C. E., Hunter, M. J., Meininger, D. P., Garrod, S., and Komives, E. A. (1996) *Proc. Natl. Acad. Sci. USA.* ***93***, 10177-10182.

Williamson, M. P., Havel, T. F., and Wüthrich, K. (1985) *J. Mol. Biol.* ***182***, 295

Winkler,(1986) *J. Biol. Chem.* ***261***, 13838-13843.

Winship, P. R., and Dragon, A. C. (1991) *Br. J. Haematol.* ***77***, 102-109.

Wu, Y., Bevilacqua, V. L. H. and Berg, J. M. (1995) *Curr. Biol.* ***2***, 91-97.

Wüthrich, K., (1989) *Acc. Chem. Res.* ***22***, 36-46.

Wüthrich, K., (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York.

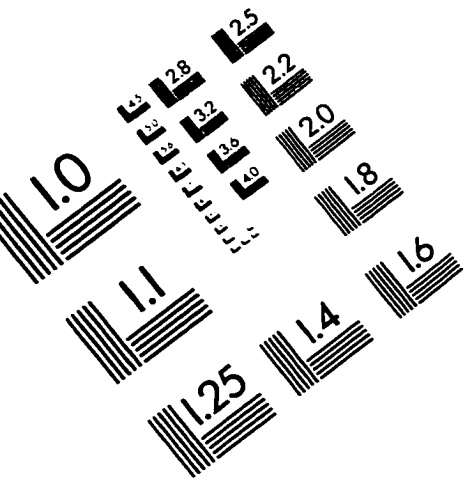
Wüthrich, K., Billeter, M., and Braun, W. (1983) *J. Mol. Biol.* ***169***, 949-961.

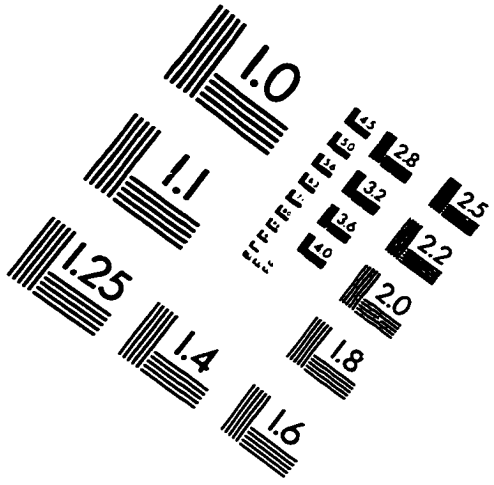
Wüthrich, K., Billeter, M., and Braun, W. (1984) *J. Mol. Biol.* ***180***, 715-740.

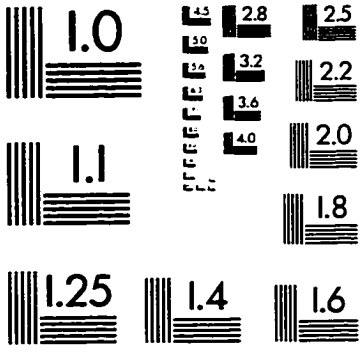
Wüthrich, K., Wilder, G., Wagner, G., and Brown, W. (1982) *J. Mol. Biol.* ***155***, 311.

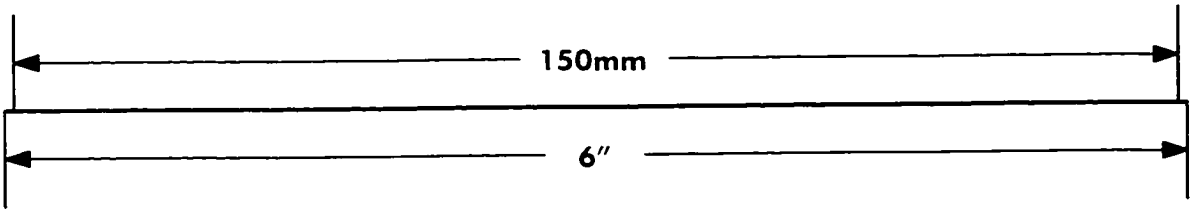
Zuiderweg, E. R. P., Boelens, R., and Kaptein, R. (1985) *Biopolymers* ***24***, 601-611.

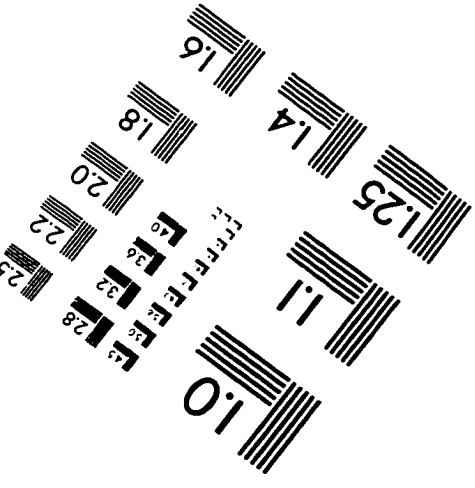
IMAGE EVALUATION
TEST TARGET (QA-3)











APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989