Structure and Properties of the DNA Binding Domain of Interferon Regulatory Factor-2 Studied by Heteronuclear Multidimensional NMR

Koichi Uegaki

Institute for Protein Research
Osaka University
1994

Contents

General Introduction

In the course of gene transcription, several protein factors appear and play critical roles for the regulation of the gene expression. Many investigations that focus on the structure and the interaction of these transcriptional factors with DNA have been reported to understand the mechanisim of transcription at molecular levels. Complete understanding of the protein function can be accomplished with the knowledge of its three-dimensional (3D) structure at atomic resolution. The recent progress of structural analyses of the DNA-binding domains of transcriptional factors, especially of eukaryotic ones, is relied on the finding that the whole molecule of a transcriptional factor can be functionally separable into two domains; the transcriptional activation or suppression domain and the DNA-binding domain (for review see Ptashne., 1988). These domains are interexchangeable with their counterdomains of other transcriptional factors, and functionally chimeric proteins are readily synthesized. The DNA-binding domains of eukaryotic transcriptional factors consist of relatively short stretches of amino acids (around 100 amino acids) with relativly high content of basic amino acids. This structural feature is an advantage for NMR analyses of transcriptional factors because of small size and solubility. At present, NMR is difficult to use for structure determination of larger proteins (more than 200 amino acid residues) because of signal degeneracy and large linewidth.

Historically, at the first stage, the three dimensional structure analyses of DNA-binding proteins were done on λ phage cro (Anderson *et al.*, 1981) and cAMP binding protein (CRP) (MaKay *et al.*, 1981) by X-ray crystallography. Then, the rapid progress in NMR and X ray techniques enabled us to analyse the structure of a variety of transcriptional regulatory proteins, their DNA-binding domains and their complexes with DNA. These analyses led to the elucidation of several DNA-binding motifs such as helix-turn-helix (HTH), zinc fingers, basic leucine zipper (bZip), basic helix-loop-helix (bHLH), β-ribbon *etc*. Based on these particular motifs, the structural principle of DNA recognition by transcrptinal factors has been discussed (for review see Harrison, 1991; Pabo and Sauer, 1992; Wolberger, 1993).

However, the common feature of DNA recognition has not yet been established, because still new binding motifs come out in the strcture analyses of transcriptional factors. Interferon regulatory factors (IRF) are such transcription factors in which DNA-binding motif does not have a resemblance with the known motif. Therefore, the author has undertaken to analyse the structure of the IRF's DNA binding domain in order to understand its DNA recognition mechanism at atomic levels.

Interferon regulatory factors have been originally identified by Fujita *et al.* (1988). Two DNA-binding factors are known which regulate the expression of the type I interferon and inducible genes. There are interferon regulatory factor-1 (IRF-1), a transcriptional activator, and IRF-2, its antagonistic repressor, (Fujita *et al.*, 1988; Miyamoto *et al.*, 1988; Harada *et al.*, 1989). These factors are structually similar to each other, particularly in the amino-terminal region which confers DNA binding character. Each factors binds to the specific target promoter sequences of the IFN-α and IFN-β genes as well as

other interferon-inducible genes. Recently, some IRF-related transcriptional factors were cloned (Driggers *et al.*, 1990; Veals *et al.*, 1992). A low degree of amino acids sequence homology showed that the interferon regulatory factors should be classified into a new family of DNA-binding motifs.

In many cases, the DNA binding protein exists as a homo- or hetero-dimer and binds to palindromic promoter elements (for review see Pabo and Sauer, 1992). However, IRFs exist as monomers and bind to direct repeats with non-palindromic sequence. The promoter elements of IRFs have unique consensus sequence repeats $[G(A)AAA(G \text{ or } C)(T \text{ or } C)GAAA(G \text{ or } C)(T \text{ or } C)]_n$ (Harada *et al.*, 1989; Tanaka *et al.*, unpublished data). This is an interesting feature eligible for the investigation of the three dimensional structure of IRFs DNA-binding domain and the interaction with DNA.

The author aimed at analysing the structure of the DNA-binding domain of IRF using NMR to clarify the recognition mechanisim of a specific DNA sequence. The process of the structure analysis by NMR is divided into three stages; the first stage is the preparation of a sample(s) specifically labelled woth a stable istope(s), the second stage is the signals of resonance assignment to specific nuclei and the third stage is the structure calculation employing distance information obtained from the NMR data.

First, the author constructed an over-expression vector of the recombinant DNA-binding domain of IRF-2 (This domain is reffered to IRF-2(113) as follows.). IRF-2 was exclusively used in the present investigation because recombinant IRF-1 was difficult to form native comformation from inclusion body in *E.coli*.

In the conventional NMR analysis of a small protein, the complete assignment of the backbone and side chain protons is performed by using well known homonuclear 2D spectra, COSY, NOESY, and TOCSY (Wüthrich, 1986). However, for the resonance assingment of a relatively large protein or a small protein but rich in helices like the DNA binding domain, the conventional assignment strategy is hardly applicable due to chemical shift overlap and large linewidth associated with high molecular weight.

Thus as the second stage, the author employed novel NMR techniques (double- and triple-resonances NMR) which introduced additional dimensions in heteronuclear chemical shifts to solve the problem of ambiguity in homonuclear 2D NMR. All of these double- and triple-resonance experiments rely on the magnetization transfer via heteronuclear one-bond J couplings. The sensitivity of these experiments depends strongly on the ratio of the size of the J coupling to the linewidth of the nuclei involved in the magnetization transfer process. For that purpose, the sample whose carbon and nitrogen nuclei are replaced by ^{15}N and/or ^{13}C must be prepared.

Third, the author performed the calculation of the three-dimensional structure of IRF-2(113) using nuclear Overhauser effect (NOE) and other NMR parameters. At the structure calculation, the distance geometry method with program DIANA developed by Güntert *et al.*, (1991) was not successful for IRF-2(113) because of the large number of residues and the complexity of the folding topology.

Therefore, the author employed another calculation method which is called a simulated annealing method (Nilges *et al.*, 1988) with program X-PLOR V3.1 (Brünger, 1993). By using this program, the global folding of IRF-2(113) could be obtained. Based on the calculated structure, the author tried to propose the binding mode of the domain with DNA.

This thesis consists of four chapters. Chapter 1 describes the identification and expression of the DNA-binding domain of IRF-2. Chapter 2 describes the techniques of the assignment of ^{1}H, ^{15}N and ^{13}C resonances to individual amino acid residues. This assignment is the most important for the structure analysis by NMR and is described in detail by means of the new techniques (heteronuclear double- and triple-resonance methods). Chapter 3 includes the result of the secondary structure determination of the DNA-binding domain derived from the NMR analyses descrived in Chapter 2. Finally, in Chapter 4 the three dimensional global folding of the DNA-binding domain of IRF-2 in solution calculated from the NMR data is given and the mode of interaction with DNA is discussed.

References

Anderson,W.F., Ohlendorf,D.H Takeda,Y. and Matthews,B.W. (1981) *Nature*, **290**, 718-723.

Driggers,P.H.,Ennist,D.L., Gleason,S.L.,Mak,W.H., Marrks,M.S., Levi,B.Z., Flanagan,J.R., Appella,E. and Ozato,K. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3743-3747.

Fujita,T., Sakakibara,J., Sudo,M., Miyamoto,M., Kimura,Y., and Taniguchi,T. (1988) *EMBO.* **7**, 3397 - 3405

Brünger, A.T. (1993) XPLOR Manual (New Haven, Connecticut: Yale University).

Güntert,P., Braun,W. and Wüthrich,K. (1991) *J.Mol.Biol.* **217**, 517-530.

Harada,H., Fujita,T., Miyamoto,M., Kimura,Y., Murayama,M., Furia,A., Miyata,T. and Taniguchi,T. (1989) *Cell*, **58**, 729 - 739

Harrison,S.C (1991) *Nature*, **353**, 715-719

Makay,D.B. and Steitz,T.A. (1981) *Nature*, **290**, 744-749.

Miyamoto,M., et al., (1988) *Cell*, **54**, 903 - 913

Nilges,M., Clore,G.M. and Gronenborn,A.M. (1988) *FEBS Lett.* **239**, 129-136

Pabo,C.O. and Saure,R.T. (1992) *Ann.Rev.Biochem.* **61**, 1053-1095.

Ptashne,M. (1988) *Nature*, **335**, 683-689

Veals,A.V., Schindler,C., Leonard,D., Fu,X-Y., Aebersold,R., Darnell,JR.J.E., and Levy,D.E. (1992) *Mol.Cell.Biol.* **10**, 3315-3324.

Wolberger,C. (1993) *Curr.opin.Struct.Biol.* **3**, 3-10.

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

Chapter-1

Characterization of the DNA Binding Domain of the Mouse IRF-2 Protein

Abstract

The DNA binding domain of the Iterferon Regulatory Factor-2 protein (IRF-2) has been produced and characterized. α-Chymotrypsin digestion of the purified IRF-2 protein bound to a synthetic binding site yields a peptide fragment of 14 K in molecular weight. N-Terminal analysis showed that the sequence of this peptide fragment is the same as that of intact IRF-2. An about 14 K peptide fragment, IRF-2(113), which corresponds to the N-terminal 113 amino acids of intact IRF-2, has been successfully expressed in a functional form in *E.coli*. The initiator methionine was removed during the expression and purified IRF-2(113) is composed of 112 amino acids. DNase I footprinting and gel retardation assay have shown that IRF-2(113) binds to a synthetic
DNA having the consensus binding site and to the upstream regulatory sequence of the IFN-β gene as well as intact IRF-2 does. These results showed that this peptide fragment, IRF-2(113), may be a good material for investigation of the DNA binding domain of IRF-2 and of the DNA-protein interaction.

Introduction

Two structurally related nuclear factors, termed interferon regulatory factors-1 and -2 (IRF-1 and IRF-2) have been discoverd and studied extensively in the context of the interferon regulation. Both IRF-1 and IRF-2 bind to the same regulatory elements within the IFN and IFN-inducible gene (Miyamoto *et al.*, 1988; Harada *et al.*, 1989). IRF-1 is a transcriptional activator (Fujita *et al.*, 1989), whereas IRF-2 functions as a repressor (Harada *et al.*, 1989). They also bind to synthetic oligonucleotides which contain unique consensus hexamer repeats. Their DNA binding domains are located within a 170 amino acid stretch in their N-terminal regions, but the precise locations are unknown (Fujita *et al.*, 1989; Harada *et al.*, 1989). Recently, a cDNA of the IFN consensus sequence binding protein (ICSBP) was cloned which also binds to the IFN-b gene regulatory element. The sequence homology indicates that this ICSBP belongs to the IRF gene family (Driggers *et al.*, 1990). The most conserved amino acids between ICSBP, and IRF-1 and -2 are located in the N-terminal 115 amino acid sequence. Driggers *et al.* pointed that this N-terminal sequence corresponds to a putative DNA binding domain. However, experimental evidence has not been provided. In this domain, there is no sequence homology to the known DNA binding motifs. Therefore, analyses of the structures of these DNA binding domains and the interaction with DNA are very interesting and important for understanding the mechanism underlying the IFN gene regulation at the molecular level. In this chapter, the author describes the characterization of the DNA binding domain of IRF-2 and the establishment of the over-expression system for the structure analyses. The DNA binding domain of IRF-2 has been produced and characterized. This manipulation is most important for the structure analyses by NMR. NMR measurement requires high sample cocentration (usually over 1mM (10-20mg/ml) protein concentration), relatively smaller molecular weight (under 20K) and sample stability (one 3D NMR measurement requires 2-4 days). Therefore, this step has a key role for the success of structure analyses by NMR spectroscopy. For these reasons, the author tried to prepare the sample of a stable and suitable size of DNA binding domain for NMR measurement.

Materials and Methods

Materials

Restriction enzymes, T4 polynucleotide kinase, deoxyribonuclease I and a DNA ligation kit were purchased from Takara Shuzo Co. (Japan). S-Sepharose, Superdex 75 and Mono-S were the products of Pharmacia (Sweeden). Oligonucleotides were synthesized with an automated DNA synthesizer 380B (Applied Biosystems, USA) according to the protocol of the manufacturer. All other reagents were of the highest grade available.

Purification of the IRF-2 protein and the truncated IRF-2 protein

A T7 expression system (Rosenberg *et al.*, 1987) was used to produce the IRF-2 protein (Harada *et al.*, 1989). Plasmid pET-IRF-2 was introduced into *E.coli.* BL21(DE3) cells (Studier and Moffatt, 1986). The transformed bacteria were grown in a flask (2L) and then induced with 1 mM IPTG when they had grown to 0.4 OD (530 nm). After 3 hr, they were harvested and frozen below -20 ℃ until they were needed. The frozen cell paste was suspended in 20 ml of lysis buffer (50 mM phosphate, pH 6.8, 300 mM KCl, 1 mM phenylmethylsulfonyl-fluoride and 7 mM 2-mercaptoethanol), followed by sonication on ice for 20 min. Insoluble materials were removed by centrifugation at 20,000 g for 30 min at 4 °C. Polyethyleneimine (MW 60,000-80,000) was added to the supernatant to a final concentration of 0.5%. The mixture was kept standing for 30 min on ice and then centrifuged at 20,000 g for 30 min at 4 °C. The supernatant was dialyzed against buffer A (50 mM phosphate, pH 6.8, 50 mM KCl, 1 mM EDTA and 7 mM 2-mercaptoethanol) plus 1 mM phenylmethylsulfonyl fluoride for 12 hr at 4 °C. The dialysate was loaded onto a S-Sepharose column pre-equilibrated with buffer A. The protein was eluted with a linear gradient of 50 mM-1 M KCl in buffer A. The IRF-2 protein was eluted in the 0.4-0.5 M KCl fraction and then dialyzed against buffer A. The dialysate was concentrated on a Mono-S column with a KCl gradient as before. Superdex 75 gel filtration was performed. The column was equilibrated with buffer A and the protein was eluted with buffer A (flow rate, 1 ml/min). The IRF-2 peak fraction was loaded onto the Mono-S column and then eluted with a linear KCl gradient. The purity was checked by 15% SDS-PAGE.

SDS-polyacrylamide gel electrophoresis

15% SDS-polyacrylamide gel electrophoresis was performed as described previously (Laemmli, 1970). Staining was performed with Coomassie Brilliant Blue.

Limited proteolysis of IRF-2 with a-chymotrypsin

IRF-2 was mixed with C1 oligomer DNA (protein/C1 oligomer = 1:1 mol/mol) in 50 ll of binding buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 7 mM 2-mercaptoethanol and 10 % glycerol) for 30 min at room temparature. The structure of the C1 oligomer, which has 4 repeats of a consensus hexamer sequence (dTCGAAGTGAAAGTGAAAGTGAAAGTGAGACTCTAGAG/dGATCCTCTAGGTCTCACTTT-CACTTTCACTTTCACT), was described by Fujita *et al.* (1987). The mixture was incubated with α-chymotrypsin at various enzyme/substrate ratios for 12 hr at 37 °C. The results of the proteolysis were analyzed by 15% SDS-polyacrylamide gel electrophoresis.

N-Terminal analysis of a 14 K fragment

A fragment of 14 K was collected from the 15% poly-acrylamide gel and its N-terminal sequence was determined by the Edman degradation method.

Amino acid analysis

Amino acid analysis was performed with a Hitachi 835-S amino acid analyzer on a protein hydrolysate after treatment with 6 N HCl containing 0.2% phenol at 110 °C for 24 hr. The optical absorption at 280 nm for a 1 mg/ml protein solution was calculated from the amino acid analysis data to be 0.418.

Construction of a plasmid for the truncated IRF-2 protein

Plasmid pET-IRF2(113) was generated by ligating a synthetic DNA (dCTACCGGATGCTGCCCTGA/dAGCTTCAGGGCAGCATCCGGT) into the AccI-HindIII backbone fragment of the pET-IRF2 vector.

DNase I footprinting and gel retardation assays

DNA binding experiments were carried out in essentially the same way as described previously (Yamamoto *et al.*, 1991). Plasmid puc-55C1A, which contains the AAGTGA hexamer repeats, $(AAGTGA)_4$, was constructed from pSV-55C1A (Fujita *et al.*, 1988) and puc119. The Eco47III and HindIII (fill in) fragment was ligated into the SmaI site of puc119. Plasmid puc-125IFN, which has a natural IFN-β promoter region, was constructed from p-125cat (Fujita *et al.*, 1988) and puc119. The

HindIII and SalI fragment was ligated into the HindIII and SalI site of puc119. Probe DNA labeling was performed as follows. Plasmids puc-55C1A and puc-125IFN were digested with EcoRI and SphI (for puc-55C1A) or SalI (for puc-125IFN), and then filled in with [α-32P] dATP and the Klenow enzyme.

The footprint assay was performed as follows. The binding mixture (20 ll) contains 50 mM phosphate, pH 6.8, 50 mM KCl, 1 mM EDTA, 7 mM 2-mercaptoethanol, 10 % glycerol, 1 lg poly dI-dC, 20 nmol ^{32}P-fill in labeled probe and the purified protein at different amount. Each mixture was incubated for 30 min at room temparature and then treated with DNaseI and analyzed by 6% denaturing gel electrophoresis.

Gel retardation assays were performed under the following conditions: 50 mM phosphate buffer, 50 mM KCl, 1 mM EDTA, 7 mM 2-mercaptoethanol, 10 % glycerol, 1 lg poly dI-dC, 20 fmol ^{32}P end-labeled C1 oligomer DNA, the indicated amounts of the purified protein and a competitor). The binding mixture (20 ll) was incubated at pH 6.8 for 30 min at room temperature and then the resultant mixture was loaded on a 8 % polyacrylamide gel.

CD measurement

CD spectra were recorded with a JASCO J-720 spectropolarimeter. The temperature of the sample cell was controlled with a circulating water thermostatic bath (Haake F3). Each spectrum was obtained by accumulating 16 times scanning.

Results

Purification and proteolysis of the IRF-2 protein

The purification procedure for IRF-2 reported by Harada *et al.* (1989) was modified as described in Materials and Methods. The IRF-2 protein was mixed with the C1 oligomer in which the AAGTGA sequence was repeated 4 times or with synthetic DNA containing an AP-1 binding site (dTTTCTGACTAATCCT/dAGGATTAGTCAGAAA) as control (Toda *et al.*, 1992). After incubation at 37 °C for 30 min, α-chymotrypsin was added at various S/E ratios, followed by further incubation at 37 °C for 12 hr. The digested mixture was analysed by 15% SDS-PAGE. Several sizes of α-chymotrypsin resistant fragments appeared when IRF-2 was incubated with the C1 oligomer and then digested with α-chymotrypsin. However, when IRF-2 was incubated with the AP-1 site DNA or without DNA, no detectable fragment appeared (Figure 1). The molecular weight of the smallest peptide fragment was about 14 K. We collected this peptide fragment from the gel and analyzed its N-terminal amino acid sequence. The results showed that its N-terminal sequence is Pro-Val-Glu-Arg-Met, which corresponds

to that of native IRF-2.

Expression and purification of the truncated IRF-2 protein

An expression vector for the truncated IRF-2 (IRF-2(113)) (Figure 2) was constructed. The synthetic oligomer, which has a stop codon at the end, was ligated into the AccI-HindIII site of the pET-IRF2 backbone. Overproduction of the truncated IRF-2 protein was obtained by adding IPTG (1 mM) to BL21(DE3) cells haboring the expression plasmid, pET-IRF2(113), at the mid-logarithmic phase (A530=0.6) at 37 °C. The cells were collected 3 hr later. The amount of induced IRF-2(113) was estimated to be 30% of the total cell protein by SDS-polyacrylamide gel electrophoresis (Figure 3).

The harvested cells were lysed by sonication. IRF-2(113) was detected in the soluble fraction. After column chromatographies for purification, the final purity of IRF-2(113) was more than 95% (Figure 3). By use of this procedure, we could obtain about 10 mg of IRF-2(113) from 1 liter of culture.

A small amount of IRF-2(113) was subjected to amino acid and N-terminal analyses. The N-terminal analysis showed that its N-terminal sequence was Pro-Val-Glu-Arg-Met and the first formylmethionine was processed by an endogenous enzyme in *E. coli*. This result is consistent with that of the amino acid composition analysis which gave the number of methionine as 5.02, although six methionine residues were expected for the peptide fragment from 1 to 113. Thus we can say that IRF-2(113) is composed of 112 amino acids.

DNA binding properties of the IRF-2(113) protein

The specific interaction of IRF-2(113) with DNA was comfirmed by gel retardation and DNase I footprinting analyses. In the gel retardaton assay involving C1 oligomer DNA, two distinct shifted bands with low and high mobilities were observed. Depending on the concentration of IRF-2(113), the number of shifted bands changed from one to three (Figure 4). To determine whether the binding of IRF-2(113) is specific to the DNA sequence or not, a competition experiment was performed, by adding an excess of unlabeled C1 oligomer or AP-1 site DNAs. As shown in Figure-5, the shifted bands appear to reflect a specific interaction, since they were competed out by the former but not the latter DNA. For the footprinting assay, we used a ^{32}P-labeled EcoRI-SphI fragment of puc-55C1A and the EcoRI-SalI fragment of puc-125IFN (Figures 6 and 7). The patterns of DNase I protection (Figure-6, lanes 2-6; Figure-7, lanes 3-6) were very similar to those previously reported for native IRF-1 (Fujita *et al.*, 1988) and IRF-2 (Harada *et al.*, 1989). Furthermore, the protected region expanded by increasing the protein concentration as reported previously with the native IRFs (Figure 7).

CD spectra

CD spectra of the IRF-2(113) protein in 5 mM phosphate buffer, pH 6.8, 20 mM KCl were measured in order to obtain imformation on the secondary structure. The spectral pattern at 30 °C showed the existence of some secondary structure (Figure 8). The fractions of each structure, α-helix and β-sheet were calculated to be 30% and 50%, respectivly, by the method of Chang *et al.* (1978).

Discussion

For the structure analyses using NMR, several hurdles exist. 1) Sample concentration. NMR measurement requires high sample cocentration. Usually it requires over 1mM (10-20mg/ml) protein concentration because NMR spectroscopy does not have high sensitivity like ESR, UV-CD, and fluorescence spectroscopy, *etc.* 2) Molecular weight. NMR sepectroscopy for the structure analyses can not deal with a protein with large molecular weight because of large linewidth and signal degeneracy. Therefore, it requires relatively small molecular weight (under 20K) 3) Sample stability. Usually, one multi-demensional NMR experiment time is needed about 2-4 days. In addition, pH condition needs under pH 6.0 to suppress the amide protons exchange to the solvent water. Therefore, sample is required for long time stability through the NMR measurement. For these reasons, sample preparation has a key role for success of the structure analyses by NMR spectroscopy. Then, the author tried the sample preparation to obtain a stable and suitable size of DNA binding domain for NMR measurement.

To determine the DNA binding domain of IRFs, the author selected IRF-2 protein not IRF-1. Because IRF-1 protein appeared in less expression level than IRF-2 and inclusion boody was formed when expressed in *E.coli*. Therefore, IRF-1 was not suitable for this work. IRF-2 protein was expressed as soluble form in *E.coli.*, but the purification procedure for IRF-2 reported by Harada *et al.* (1989) was not effective and time consuming for the author's purpose. So, the procedure was modified as described in Materials and Methods. The modified method was very simple and easy to get the sample of high purity and quantity (Figure 1; lane 2). Moreover, this procedure does not use denaturation process, and hence, purified IRF-2 protein seems to have higher DNA-binding activity than that obtained by the previous procedure.

To determine the DNA binding domain of IRF-2 protein, limited proteolysis experiment was performed. The IRF-2 protein was mixed with the C1 oligomer in which the AAGTGA sequence was repeated 4 times or with synthetic DNA containing an AP-1 binding site (dTTTCTGACTAATCCT/dAGGATTAGTCAGAAA) as control (Toda *et al.*, 1992). α-Chymotrypsin was added at various S/E ratios, followed by further incubation at 37 °C for 12 hr. The digested mixture was analysed by 15% SDS-PAGE. The result showed that limited proteolysis of IRF-2 bound to the C1 oligomer by α-chymotrypsin gave a small α-chymotrypsin resistant fragment (MW=14K).

This fragment was not further digested by α-chymotrypsin. In contrast, when IRF-2 was incubated with a non-specific DNA oligomer, like AP-1 site DNA or without DNA, IRF-2 was easily digested by α-chymotrypsin, and no detectable fragment was obtained. This result strongly suggests that the α-chymotrypsin resistant fragment which has affinity to a specific DNA sequence contains the DNA binding domain of IRF-2. N-Terminal analysis of this fragment showed that its amino acid sequence is Pro-Val-Glu-Arg-Met, which corresponds to the native IRF-2 N-terminal sequence lacking the Met at the first position. Previous report indicated that the DNA binding domains of IRF-1 and IRF-2 are located somewhere in a 170 amino acid stretch in their N-terminal regions (Fujita *et al.*, 1989; Harada *et al.*, 1989). Furthermore, the sequence similarity between IRF-1 and IRF-2 is very high in the N-terminal 120 amino acids. Our proteolysis results well agree with these findings.

Based on these lines of evidence we constructed an expression vector for the truncated IRF-2 (IRF-2(113)), the molecular weight of which was estimated 14K from DNA sequence. It is about the same size as that of the α-chymotrypsin resistant fragment, but the size of the truncated protein checked by SDS-PAGE was a little smaller than 14K marker (data not shown). Amino acid and N-terminal analysis of the IRF-2(113) obtained indicated that the N-terminal methionine was lacking probably because of endogenous *E. coli.* proteases. Thus, the IRF-2(113) protein consists of 112 amino acids.

Throughout purification, IRF-2(113) was stable as to the endogenous *E. coli.* protease, in contrast to native IRF-2. We also constructed an IRF-2(108) expression system in a similar way to that for the IRF-2(113) system, but IRF-2(108) was not expressed in *E. coli* (data not shown). It is assumed that the C-terminal portion of IRF-2(113) is important for the stability of the conformation of the DNA binding domain (detailed disscussion will be done in chapter 3).

Analysis of the interaction of IRF-2(113) with DNA was performed by gel retardation assay. It was shown that the IRF2(113) protein specifically bound to the C1 oligomer DNA, because the shifted bands were competed out by the cold C1 oligomer but not by the AP-1 site DNA. In this assay, two shifted bands were observed. Interestingly, with the increase in the protein concentration, an additional retarded band appeared (Figure 4). Ultracentrifugation showed that IRF-2(113) is a monomer in solution (data not shown). Thus, these results suggest that the shifted bands correspond to monomer, dimer and trimer proteins complexed with the C1 oligomer, respectively. The trimer band did not appear until the protein concentration exceeded 0.1 mM. In contrast, the monomer and dimer bands were easily detected at low protein concentrations. These results indicate that the C1 oligomer DNA contains two IRF-2(113) binding sites adjacent to each other. The C1 oligomer DNA has an extra 17 mer oligonucleotide sequence, other than the concensus binding sequence, and the third protein may be bound through a weak interaction with the extra sequence portion.

To further characterize the binding specificity of truncated protein IRF-2(113), DNase I footprinting was performed. It showed that IRF-2(113) exactly interacts with the same sequence in the

C1 oligomer DNA cloned in a plasmid, puc118, as that in the natural IFN-β upstream regulatory region (Fujita *et al.*, 1988). With the increase in the protein concentration, expansion of the protected region was observed. This is consistent with the results of the gel shift assay. It may be suggested that IRF-2 masks the region of the IFN regulatory sequence and then silences the IFN gene at a higher concentration in vivo.

Recently, UV-CD experiment showed that several DNA binding proteins caused comformational change when they bound to the target DNA sequences and bend them (for example, jun and fos, GCN4, pap1 and pho4 protein)(Patel *et al.*, 1990; Weiss *et al.*, 1990; Suzuki, *et al.*, unpublished. Shimizu, *et al.*, unpublished). Although, for the case of IRF-2(113) protein, UV-CD experiment showed that IRF-2(113) protein did not cause comformational change even when it bound to its target DNA (data no shown). In addition, the circular permutation assay showed that IRF-2(113) did not bend DNA (data not shown).

UV-CD measurements showed the existence of a large amount of secondary structure in the IRF-2(113) protein. The estimated contents of secondary structure were about 30% for α-helix and about 50% for β-sheet. IRF-2(113) contains 21 basic and 13 acidic amino acid residues among 112 total residues. Isoelectric point is about 10.0. It has a typical feature of the DNA binding domain. It contains 10 proline and 15 aromatic amino acid residues. However, the role of proline residues is hardly expected. The tertiary structure of the domain should be elucidated and it is expected that a new binding motif is expected to appear in structural analyses.

References

Chang,C.T., Wu,C.S.C. and Yang,J.T. (1978) *Anal. Biochem.*, **91**, 13-31

Driggers,P.H., Ennist,D.L., Gleason,S.L., Mak,W.H., Marrks,M.S., Levi,B.Z., Flanagan,J.R., Appella,E. and Ozato,K. (1990) *Proc. Natl. Acad. Sci.* USA, **87**, 3743-3747.

Fujita,T., Shibuya,H., Hotta,H., Yamanishi,K. and Taniguchi,T.(1987) *Cell*, **49**, 357-367.

Fujita,T., Sakakibara,J., Sudo,M., Miyamoto,M., Kimura,Y. and Taniguchi,T. (1988) *EMBO J.*, **7**, 3397-3405.

Fujita,T., Kimura,Y., Miyamoto,M., Barsoumian,E.L. and Taniguchi,T. (1989) *Nature*, **337**, 270-272.

Harada,H., Fujita,T., Miyamoto,M., Kimura,Y., Murayama,M., Furia,A., Miyata,T. and Taniguchi,T. (1989) *Cell*, **58**, 729-739.

Laemmli,U.K. (1970) *Nature*, **227**, 680-685.

Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.

Miyamoto,M., Fujita,T., Kimura,Y., Maruyama,M., Harada,H., Sudo,Y., Miyata,T. and Taniguchi,T. (1988) *Cell*, **54**, 903-913.

Patel,L., Abate,C. and Curran,T. (1990) *Nature*, **347**, 572-575

Rosenberg,A.H., Lade,B.N., Chui,D., Lin,S., Dunn,J.J. and Studier,F.W. (1987) *Gene*, **56**, 125-135.

Studier,F.W. and Moffatt,B.A. (1986) *J.Mol.Biol.*, **189**, 113-130

Toda,T., Shimanuki,M., Saka,Y., Yamano,H., Adachi,Y., Shirakawa,M., Kyogoku,Y. and Yanagida,H. (1992) *Mol.Cell.Biol.* (in press).

Weiss,M.A., Ellenderger,T ., Wobb,C.R., Lee,J.P, Harrison,S.C. and Struhl.K. (1990) *Nature*, **347**, 575-578.

Yamamoto,K., Mori,S., Okamoto,T., Shimotohno,K. and Kyogoku,Y.(1991) *Nucl.Acids.Res.*, **19**, 6107-6112.

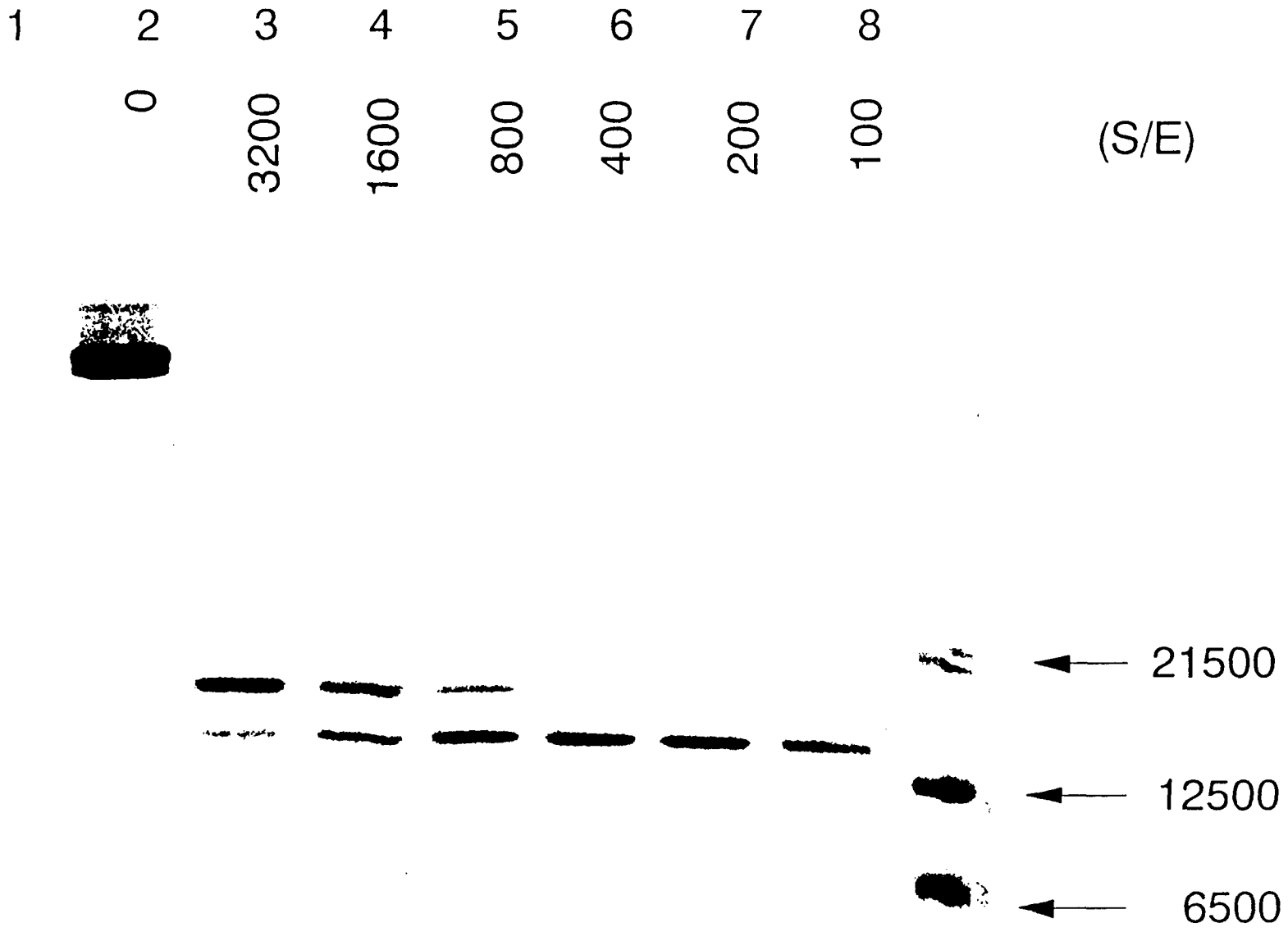
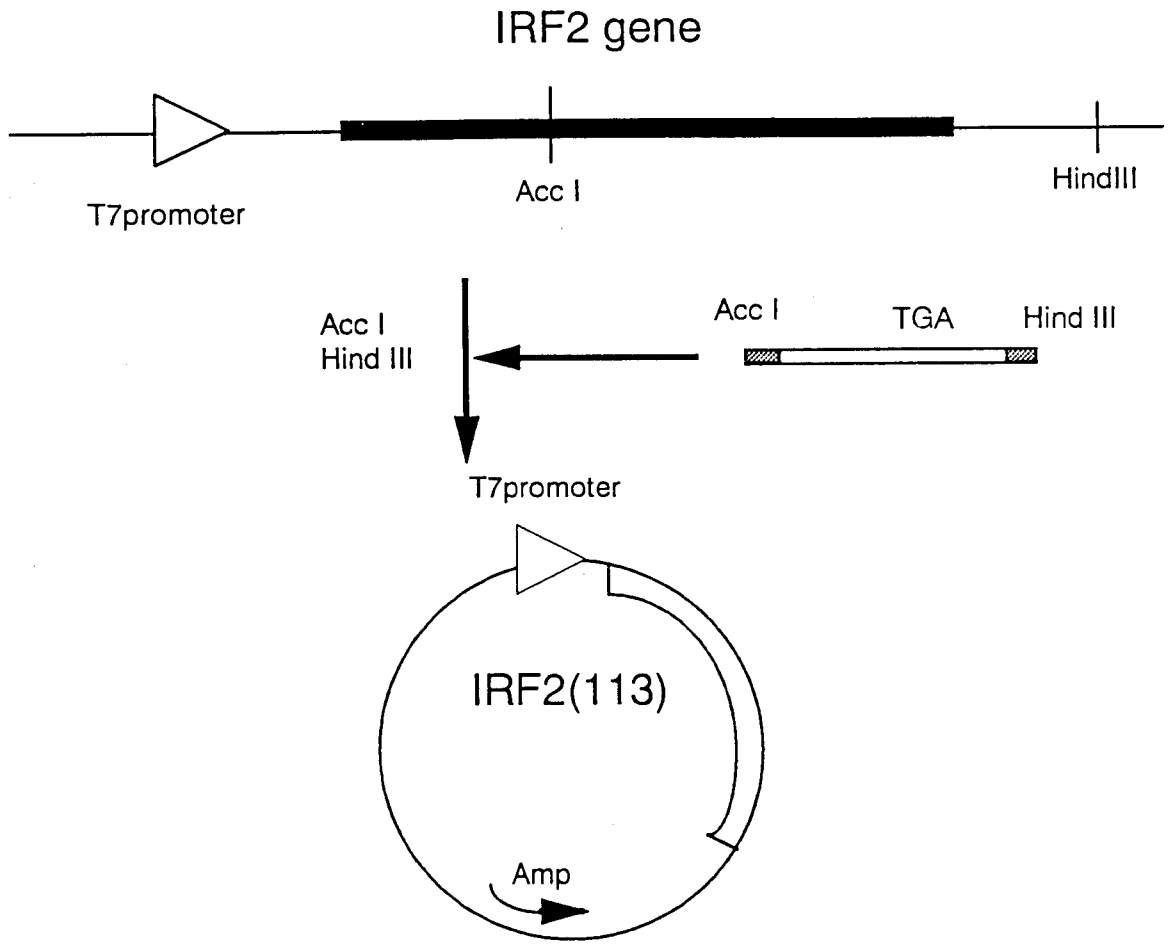


Figure 1

Figure 2



IRF2(113)

```
          10        20        30        40
MPVERMRMRPWLEEQINSNTIPGLKWLNKEKKIFQIPWMHAA
      50        60        70        80
RHGWDVEKDAPLFRNWAIHTGKHQPGIDKPDPKTWKANFRC
    90        100       110
AMNSLPDIEEVKDRSIKKGNNAFRVYRMLP
```

Figure 3

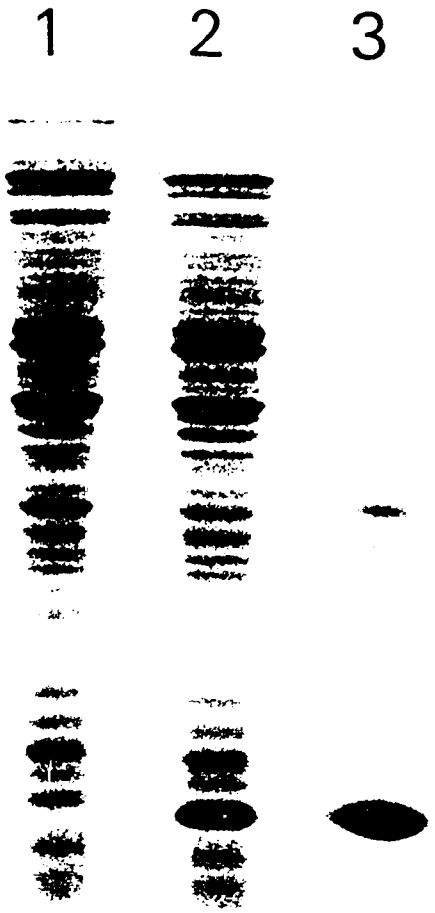
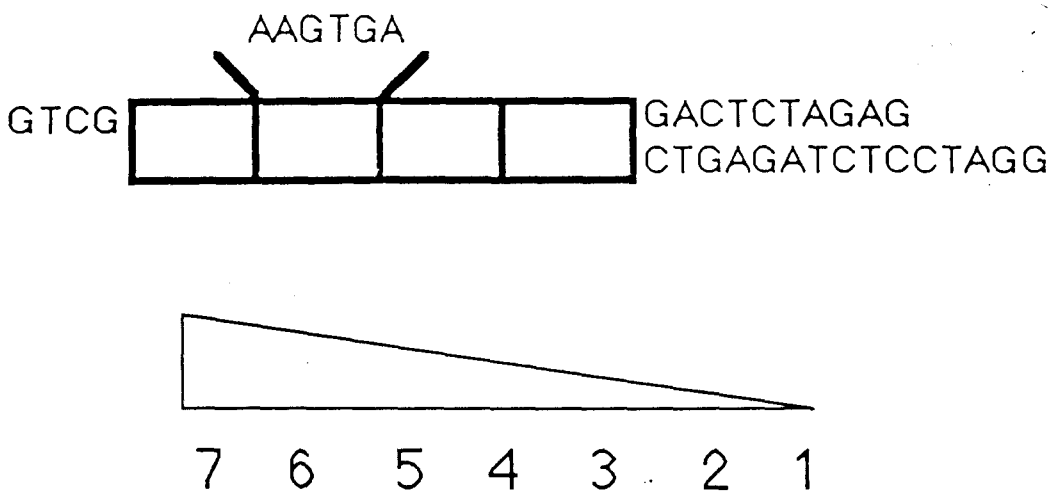


Figure 4



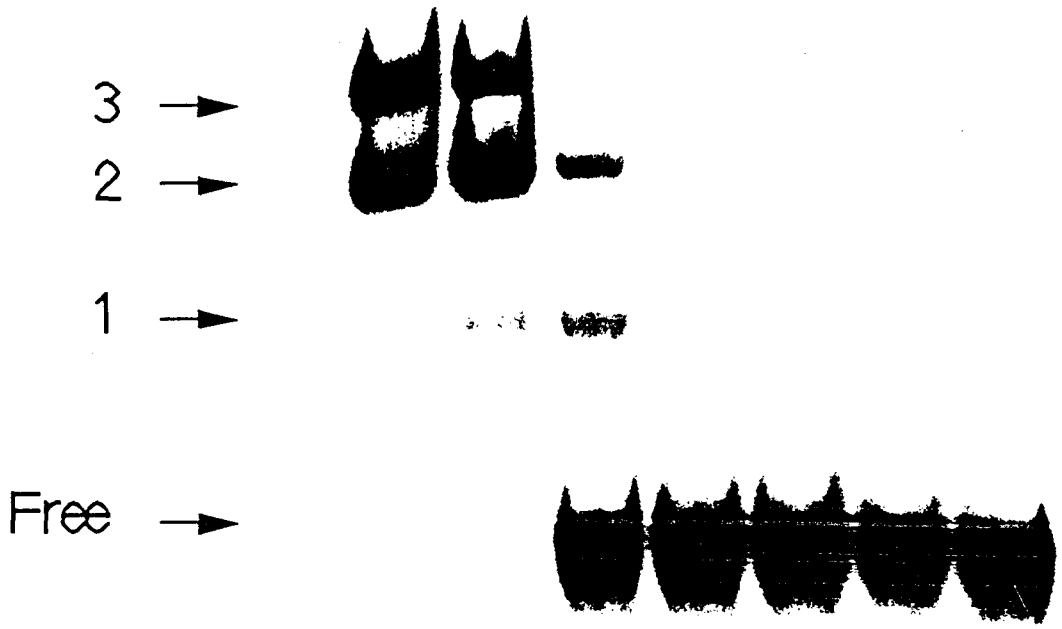


Figure 5

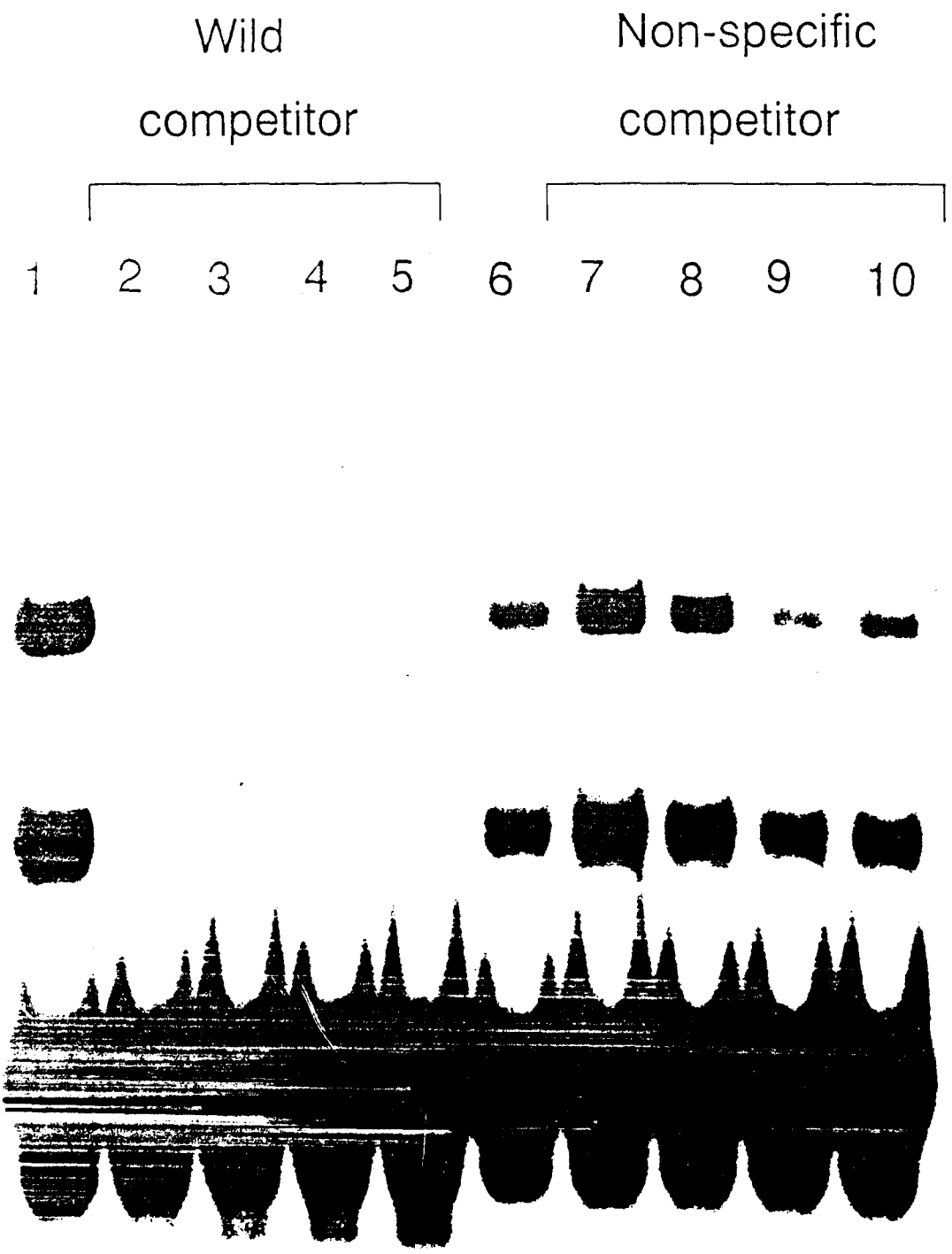


Figure 6

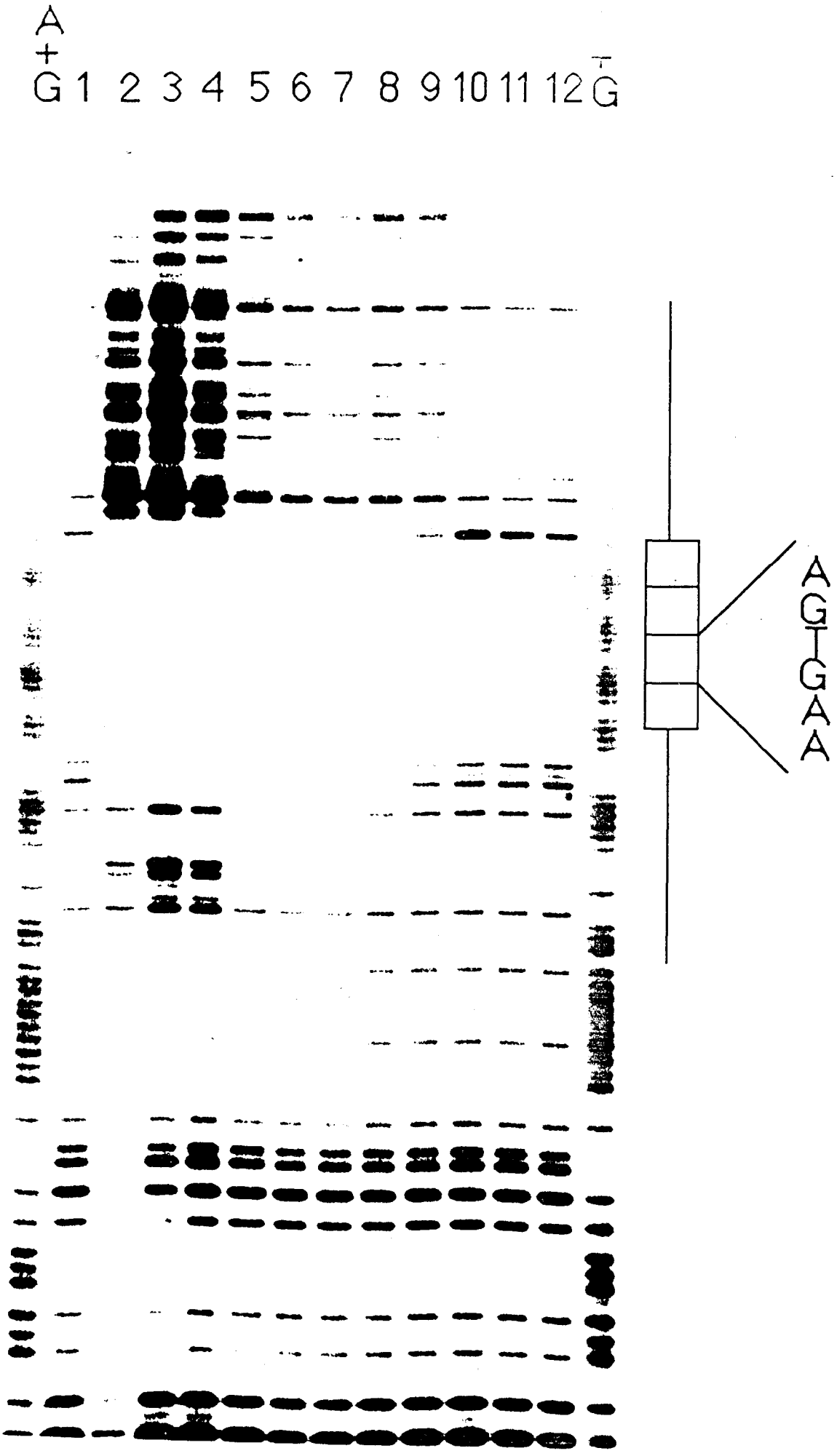


Figure 7

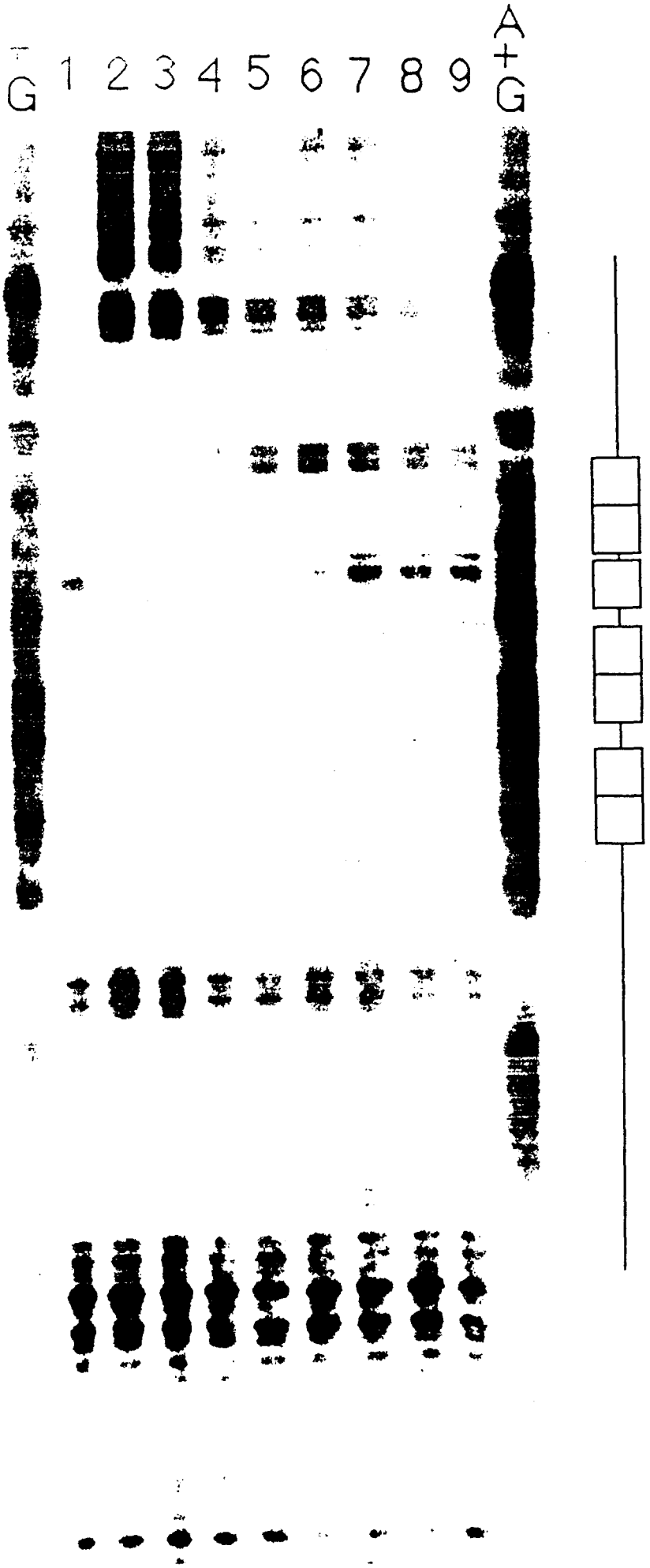


Figure 8

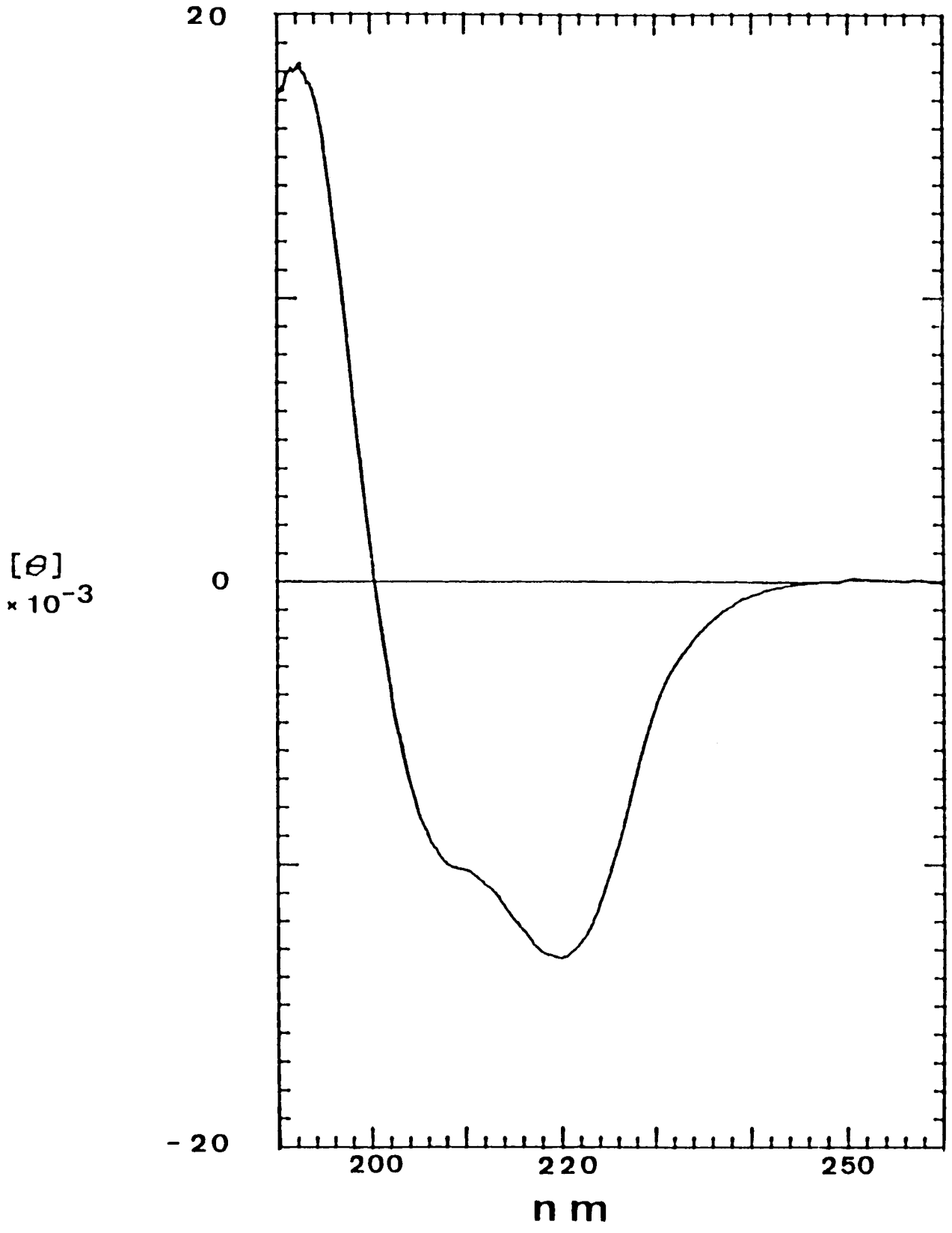


Figure legends

Figure 1.

Proteolysis of the IRF-2 protein. IRF-2 was incubated with the C1 oligomer (1:1=mol/mol)(lanes 3-8). α-Chymotrypsin was added at various ratios, followed by incubation at 37 oC for 12 hr. The digested mixture was analyzed by 15% SDS-PAGE. Lane 2; purified IRF-2 protein. It gives a satellite band due to autodegradation after prolonged incubation. Lane 1; without DNA. S/E ratios = 200. Lane 3; S/E =3200. Lane 4; S/E = 1600. Lane 5; S/E = 800. Lane 6; S/E = 400. Lane 7; S/E = 200. Lane 8; S/E = 100.

Figure 2.

Schematic representation of the construction of plasmid pET-IRF2(113). The lower panel shows the amino acid sequence of IRF-2(113). The first Met was processed during expression. + and - denote positively and negatively charged amino acids, respectively. Stars denote tryptophan residues.

Figure 3.

Purity of the truncated IRF-2 protein, as analysed by 15% SDS-PAGE. Lane 1; lysate of uninduced BL21(DE3) containing pET-IRF2(113). Lane 2; lysate of induced BL21(DE3) containing pET-IRF2(113). Lane 3; purified IRF-2(113).

Figure 4.

Gel shift assay of the C1 oligomer DNA with IRF-2(113) at different concentrations. 20 fmol of 32P-labeled C1 probe 37 mer was used. The concentration of the purified IRF-2 protein increases from right to left. Lane 1; no protein. Lane 2; 0.63 ng purified protein. Lane 3; 1.25 ng. Lane 4; 2.5 ng. Lane 5; 5 ng. Lane 6; 0.25 lg. Lane 7; 0.5 lg.

Figure 5.

Competition assaying of IRF-2(113) using a gel shift assay. 20 fmol of 32P-labeled C1 probe 37 mer and 3 ng of the purified IRF-2(113) were used. Cold C1 oligomer (lanes; 2-5) and cold AP-1 site DNA (16 mer) (lanes; 7-10) were used as competitors. Lanes 1 and 6; without a competitor. Lanes 2 and 7; 10-fold excess competitor. Lanes 3 and 8; 100-fold excess. Lanes 4 and 9; 200-fold excess. Lanes 5 and 10; 400-fold excess.

Figure 6.

Binding properties of IRF-2(113) as to plasmid DNA containing the synthetic consensus sequence, as analysed by DNase I footprinting. A fragment of puc-55C1A containing the C1 oligomer

was incubated with 0.06 pmol-0.2 nmol of purified IRF-2(113) as described under Materials and methods. After DNase I digestion, the digested mixture was analyzed on a 10% polyacrylamide sequencing gel. The boxes represent the hexamer repeats. The DNase I control and A/G ladder are shown in parallel lanes. A+G; Maxam-Gilbert sequencing reaction (Maxam *et al.*, 1980) used as size marker. Lane 1; no protein. Lane 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12; 0.2nmol, 0.08nmol, 0.04nmol, 8pmol, 4pmol, 2pmol, 1pmol, 0.5pmol, 0.25pmol, 0.125pmol, 0.06pmol purified protein, respectively.

Figure 7.

Binding properties of IRF-2(113) as to the native IFN-β gene sequence, as analysed by DNase I footprinting. A fragment of puc-125IFN containing the upstream region of the IFN-β gene sequence was incubated with 0.06 pmol-0.2 nmol of purified IRF-2(113) as described under Materials and methods. The boxes represent the putative consensus sequence. The DNase I control and A/G ladder are shown in parallel lanes. A+G; Maxam-Gilbert sequencing reaction (Maxam *et al.*, 1980) used as size marker. Lane 1; no protein. Lane 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12; 0.2nmol, 0.08nmol, 0.04nmol, 8pmol, 4pmol, 2pmol, 1pmol, 0.5pmol, 0.25pmol, 0.125pmol, 0.06pmol purified protein respectively.

Figure 8.

UV-CD spectrum of the IRF-2(113) protein. Purified IRF-2(113) was dissolved in a 5 mM sodium phosphate buffer solution with 20 mM KCl, pH 6.8, at 30°C.

Chapter-2

Assignment of ^{1}H, ^{15}N and ^{13}C Resonances of the IRF-2 DNA Binding Domain Using Double- and Triple-resonance Heteronuclear Magnetic Resonance Spectroscopy.

Abstract

The assignment of the ^{1}H, ^{15}N, ^{13}C resonances of the DNA binding domain of recombinant mouse interferon regulatory factor-2(IRF-2), 112 amino acids residues of the molecular weight of 13.5 K, has been achieved using a combination of two- and three dimensional double- and triple-resonance NMR experiments. A 3D constant-time (CT) version of the triple-resonance HNCA, HN(CO)CA which correlates the HN(i), $^{15}N(i)$ and $^{13}C_{\alpha}(i)$ and $^{13}C_{\alpha}(i-1)$ chemical shifts, HNCO which does the HN(i), $^{15}N(i)$ and $^{13}CO(i-1)$ chemical shifts, HN(CA)CO correlating the HN(i), $^{15}N(i)$ and $^{13}CO(i)$ shifts, and HCACO which takes the correlation of the $H_{\alpha}(i)$, $^{13}C_{\alpha}(i)$, and $^{13}CO(i)$ experiments were mainly used for obtaining the sequential connectivity of the backbone resonances. The side-chain spin systems were identified using 3D HCCH-TOCSY and 3D ^{1}H-^{15}N TOCSY-HSQC experiments. The deviations of $^{13}C_{\alpha}$ chemical shifts of IRF-2(113) from their random-coil chemical shifts correspond well to the positions of the helical structure and the β-strand or the extended structure determined by NMR short and medium-range NOEs.

Introduction

Interferon regulatory factor-2 (IRF-2) has been identified as a negative transcriptional regulator of type I interferon and interferon-inducible genes. As the author discussed in the previous chapter, the DNA binding domain of IRF-2 is expected to span from the N-terminal to the 113th amino acid residue predicted from the DNA sequence. However, the methionine residue at the N-terminal was removed in the expressed protein in *E. coli* and the total number of the amino acid is 112. The recombinant DNA binding domain (IRF-2(113)) binds to the consensus sequence in the same manner as the native IRF-2 does. So, the recombinant IRF-2(113) is a good model for analysis of DNA-binding domain of IRF-2 by NMR. This DNA binding domain does not show apparent sequence homology with the known DNA binding motifs. In order to understand its interaction with DNA, the author has started the structure analysis to determine its three dimensional structure in solution by NMR spectroscopy.

For that purpose, a sequential assignment of the backbone nuclear magnetic resonances is the first step. In the past years, this sequential assignment was done by the combined use of homonuclear ^{1}H-^{1}H NOESY and HOHAHA spectra which employ the spacial connection of protons and ^{1}H-^{1}H vicinal scalar couplings (3-10 Hz). However, this conventional approach is not applicable for large proteins like IRF-2(113) protein, because chemical shift overlap and degeneracy are so severe and larger linewidths associated with the increase of molecular weight disturb signal resolution. A few years ago stable isotope aided, 3D triple resonance techniques were proposed by the NIH group (Ikura *et al*,. 1990; Kay *et al*,. 1990). These techniques employ one-bond large heteronuclear couplings for magnetization transfer and thus, are less sensitive to linewidths. In the techniques, signal separation becomes much better by the introduction of heteronuclear chemical shift dimensions (^{15}N and ^{13}C chemical shift). In this chapter, nearly complete ^{1}H, ^{15}N, and ^{13}C resonances assingment of IRF-2(113) by using these techniques will be reported.

Materials and Methods

Sample preparation

Uniformly ^{15}N-labeled recombinant mouse IRF-2(113) was expressed in a BL21(DE3) in M9 medium with $^{15}NH_4Cl$ (0.5g/L) as sole nitrogen source and uniformly ^{15}N / ^{13}C-doubly labeled protein was expressed in M9 medium with $^{15}NH_4Cl$ (0.5g/L) and $[^{13}C_6]_D$-glucose (1g/L) as nitrogen and carbon sources. Specifically labeled IRF-2(113) (^{15}N-lysine, arginine and deuterium aromatic ring proton) was obtained by adding labeled amino acids (100mg/L) to the medium. The purification procedure was described in the previous chapter. The samples for NMR measurements comprised 1-3mM protein in 90% H_2O/10% D_2O, or in D_2O, containing 10 mM phosphate buffer and 50 mM KCl at pH 5.5 (direct meter reading).

NMR spectroscopy

Two- and three-dimensional (2D and 3D) spectra were acquired on a Brucker AMX-500 spectrometer or a Varian Unity-600 spectrometer equipped with a H-X inverse probe or a triple-resonance $^1H(^{15}N, ^{13}C)$ probe for the ^1H detection. Proton chemical shifts were referenced to water signal at 25°C (4.75ppm) and the ^{15}N and ^{13}C chemical shifts were indirectly referenced to $HCO^{15}NH$ and ^{13}C-aspartic acid. 2D or 3D data were processed with an X32 computer or transferred to a Silicon Graphics Iris 4D/35GT or a Indigo R4000 workstation for processing and analysis using a FELIX (Hare Research, Inc) or a NMRZ (New Methods Research) software package.

The 2D ^1H-^{15}N HSQC spectra (Bodenhausen and Ruben, 1980; Bax *et al.*, 1990; Norwood *et al.*, 1990) were recorded at 25°C with spectral widths of 2787 Hz and 12,500 Hz in the ^{15}N and ^1H dimensions, respectively. Solvent suppression was achieved by the spin lock purge pulse (Messerle *et al.*, 1989). ^{15}N decoupling during acquisition was done using GARP-1 phase moduration (Shaka *et al.*, 1985) and a pure absorption mode spectrum was obtained using the States method for t1 period (States *et al.*, 1982). The complex data matrix was 512(t1) × 1024(t2).

The 3D ^1H-^{15}N TOCSY-HSQC and NOESY-HSQC spectra of uniformly ^{15}N-labeled IRF-2(113) were recorded at 25°C and pH 5.5. The pulse sequences used were minor modifications of those previously described by Fairbrother *et al.*,(1991). The isotropic mixing time using DIPSY-2 (Shaka *et al.*, 1988) in the 3D TOCSY-HSQC was 70ms and the mixing time in the NOESY-HSQC was 100ms. Water suppression was achieved by using a spin lock purge pulse (Messerle *et al.*, 1989). GARP-1 phase modulation was used for ^{15}N decoupling during acquisition (Shaka *et al.*, 1985). Quadrature detection in t1 and t2 in these spectra was achieved using TPPI-States method (Marion *et al.*, 1989). The acquired data matrix of each 3D spectrum has 128(t1) × 32(t2) × 1024(t3) complex

data points, and the spectra widthes were 6002.4 Hz (t1), 750.12 Hz (t2) and 12,500 Hz (t3), respectively. The data matrix were zero-filled to 256(t1) ×64(t2) ×1024(t3), and the Fourier-transformed data were plotted as 64 series of F2 slices.

For the 3D CT-HNCA, CT-HN(CO)CA, CT-HNCO and CT-HN(CA)CO triple-resonance experiments (Grzesiek *et al.*, 1992), the spectral width in the indirectly detected ^{15}N, $^{13}C_{\alpha}$ and ^{13}CO dimensions were 750.1 Hz, 3269.9 Hz and 9443.4 Hz, respectively. These experiments were performed in essentially the same way as described by Grzsiek *et al.* (1992). ^{15}N decoupling during acquisition was achieved using GARP-1 phase modulation (Shaka *et al.*, 1985).

For the CT-HCACO triple-resonance (Powers *et al.*, 1991) experiments were performed in essentially the same way as described by Powers *et al.* (1991). GARP-1 phase modulation was used for ^{13}C decoupling during acquisition (Shaka *et al.*, 1985). The spectra width in the indirectly detected ^{13}C and ^{13}CO dimensions were 2,765.4 Hz and 1,257.8 Hz, respectively. ^{1}H acquisition dimension was 12,500 Hz with the carrier frequency at the water signal in these experiments. Quadrature detection for all of these experiments in t1 and t2 dimensions were achieved using TPPI-States method (Marion *et al.*, 1989).

The t1 and t2 sampling delays for these triple-resonance experiments were adjusted to one-half of the respective increment, resulted in a 90° (zero-order) and -180° (first-order) phase value (Marion *et al.*, 1988). The size of the acquired data matrix were 128(t1) × 16(t2) × 1024(t3) for CT-HCACO, and 64(t1) × 16(t2) × 1024(t3) complex data points for the other experiment. One zero-filling was used in t1 (^{13}C or ^{13}CO) and t2 (^{15}N) dimensions and linear prediction by mean of the mirror image technique (Zhu *et al.*, 1990) was used to further extention of the data in t2 (^{15}N or ^{13}C) dimension. Thus the final data matrix consisted of (128 or 256) × 64 ×1024 points and were Fourier-transformed into 64 series of F2 slices.

For the HCCH-TOCSY experiments were performed as described by Bax *et al.* (1990b) except that the carbonyl inversion pulse during the t_2 period was applied on resonance using the off-resonance DANTE pulse and the ^{13}C isotropic mixing was achieved using DIPSY-2 sequence (Shaka *et al.*, 1988). ^{13}C decoupling during acquisition was carried out with GARP-1 phase modulation. The ^{1}H carrier was placed at 2.55 ppm and the ^{13}C carrier at 41.33 ppm. The isotropic mixing was achieved by using DIPSI-2 sequence and mixing time was 21.8ms. Quadrature detection for this experiment in t1 and t2 dimension was achieved using TPPI-States method and States method, respectively. Folding occurs in the F2 dimension of the HCCH experiment, so that the folded peaks had opposite phase to unfolded peaks (Bax *et al.*, 1991). The t1 and t2 sampling delays for this experiment was also adjusted to one-half of the respective increment, resulted in a 90° (zero-order) and -180° (first-order) phase values. The acquired data matrix had 128(t1) × 32(t2) × 1024(t3) complex data. The spectra widths for t1, t2 and t3 periods were 3,750.9 Hz, 4,464.3 Hz and 12,500 Hz, respectively. One zero-filling was used in t1 (^{1}H) and t2 (^{13}C) dimensions. Thus the final data

matrix consisted of 256 × 64 × 1024 points and were Fourier-transformed in to 64 series of F2 slices.

Results and Discussion

Overview of sequential assignment strategy

The outline of the strategy for the assignment was as follows (Figure-1).

1) IRF-2(113) was labeled uniformly or in an amino acid specific manner with ^{15}N. The specific labeling enabled the amino acid specific assignment of the ^{1}H-^{15}N cross-peaks. The uniformly ^{15}N labeling with specific [1-^{13}C] proline labeling facilitated the assignment of the residues following the proline residues (Kainosho and Tsuji, 1982). Then the author could assign several cross-peaks in the ^{1}H-^{15}N HSQC spectrum of the uniformly ^{15}N labeled protein.

2) The 3D ^{1}H-^{15}N TOCSY-HSQC and NOESY-HSQC spectra were employed to obtain the sequential connectivities of the main chain signals. However, it was still difficult to obtain the sequential connectivities for many parts of the protein because of signal degeneracy and relatively large linewidths due to larger molecular weight of the protein.

3) Finally, in order to get the complete assignment for the main chain nuclei, a series of the 3D triple-resonance experiments was performed.

Figure 1 represents a flow chart of the sequential assignment strategy. Figure 2 shows the summary of the correlations and connectivities observed in the hetero nuclear NMR experiments employed for the assignment of the spectra of IRF-2(113).

Specific and double-labeling experiments

The ^{1}H-^{15}N HSQC spectrum of uniformly ^{15}N-labeled IRF-2(113) protein was used as the master spectrum for the assignment of the backbone resonances (Figure 3).

For the classification of the amide proton resonances observed in the 2D HSQC spectrum to the amino acid types, the author performed the amino acid specific labeling of the protein.

The IRF-2(113) protein was labeled with ^{15}N-lysine and ^{15}N-arginine. Figures 4a and b show the ^{1}H-^{15}N HSQC spectra of the IRF-2(113) protein labeled with ^{15}N-lysine and ^{15}N-arginine, respectively. In these spectra, one-bond correlations between the amide proton and amide nitrogen of lysine and arginine residues were observed. By comparison of these spectra with ^{1}H-^{15}N HSQC spectrum of the uniformly ^{15}N labeled IRF-2(113) protein, the two amino acid types were identified.

Figure 3 shows a two-dimensional ^{1}H-^{15}N HSQC spectrum of uniformly ^{15}N - labeled IRF-2(113) protein in which 101 of 103 expected backbone ^{1}H-^{15}N correlations were identified.Two residues (Val 3 and Asp 73) do not show corresponding cross peaks in the two-dimensional and three-dimensional

double and triple resonance experiments. This can be explained by very fast amide proton exchange or by intermediate chemical exchange between different conformational states resulting in line broadening of the amide proton and/or ^{15}N resonance (chemical exchange will be discussed in the next chapter; 'Interaction with DNA' section). In Figure 3, and Figure 4a, some ^{1}H-^{15}N correlations detected (for example, Thr 62 in Figure 3, Lys 71, Lys 78 and Lys 50 in Figure 4a) have lower intensity than the others, indicating that flexibility difference occurs within the protein. Furthermore, in Figure 3, and Figure 4a, Lys 31, Lys 75, and Asp 90 generate two peaks (some peaks have lower intensity than the others). These facts directly indicate that the IRF-2(113) has different conformations at specific part in solution. As discussed in the next chapter, Lys 31, and Asp 90 locate at the edge of the β-sheet and Lys 75 locates at the edge of the α-helix (see Chapter-3).

The IRF-2(113) protein contains 10 proline residues. This fact means difficulty in the sequential assignment in a conventional procedure, because the sequential assignment is interrupted at the proline residue due to the lack of its amide proton. Therefore, the author performed the double labeling experiment (in this case, the protein was labeled with [1-^{13}C]proline and uniformly with ^{15}N). Figure 5 shows the ^{1}H-^{15}N HSQC spectrum of the doubly labeled protein. The $^{1}H/^{15}N$ cross peaks of the amino acid residues which are located next to the proline residue split into doublet along the ^{15}N axis due to the ^{13}C-^{15}N one bond J-coupling (11-15 Hz). Seven of 9 residues which are next to proline residues could be identified except Val3 and Asp73. Moreover, the amide proton of Lys 75 could be identified by comparison of this experiment with ^{15}N-lysine labeled spectrum (Figure 4a and 5). This spectrum provided very useful information to us, because the author could extend the sequential assignment starting from these identified chemical shifts. In some cases, site specific amino acid assignment could be done.

Analyses of the triple-resonance spectra (CT-, HNCA, HN(CO)CA, HNCO, HN(CA)CO, HCACO)

Starting from the definite assignments obtained from the above double labeling procedure, the sequential assignments by means of 3D ^{1}H-^{15}N TOCSY-HSQC and NOESY-HSQC spectra were performed. However, aliphatic ^{1}H resonances of many residues could not be assigned because extensive chemical shift overlap occurred in the finger print region and moreover coherence transfer was not efficient in 3D ^{1}H-^{15}N TOCSY-HSQC experiment. Therefore, the author combined the triple-resonance experiment with the result of specific labeling method for complete sequential assignment. Many kinds of triple-resonance pulse trains have been reported in the past few years. It is important for us to choose which triple-resonance pulses are efficient for the use. Since the ^{1}H-^{15}N HSQC spectrum was employed as the master spectrum for total assignment (Figure 3), following several triple-resonance experiments were chosen; CT-HNCA, CT-HN(CO)CA, CT-HNCO, and CT-HN(CA)CO. These triple-resonance experiments were all edited as to ^{1}H and ^{15}N chemical shifts

and measured in H_2O, and thus these spectra could be compared with the master spectrum of ^1H-^{15}N HSQC (strategy by using the triple resonance experiments is shown in Figure 1).

The advantage of a use of the constant-time (CT) acquisition scheme for indirectly detected frequencies is to eliminate the decay of magnetization during the acquisition period and to permit the use of a mirror image linear prediction (Zhu *et a*l., 1990). Therefore, the point resolution could be enhanced in this method.

Figure 6 shows the sequential connectivity using the CT-HNCA and CT-HN(CO)CA spectra. The sequential connectivity from Asp 96 to Glu 92 through C_α(i, i-1) is shown. By combining with HNCA and HN(CO)CA spectra,, almost all expected HN(i)-N(i)-C_α(i,i-1) sequential correlations were extracted for the signals observed in the ^1H-^{15}N HSQC spectrum. The spin systems of five glycine residues were identified by their distinctive $^{13}C_\alpha$ chemical shift (44 ppm) in the HNCA spectrum. The H_α resonances were found in the three dimensional HCCH-TOCSY experiment. As shown in Figure 6, the combination of these two spectra is very useful to establish the sequential connectivity through $^{13}C_\alpha$. CT-HNCA spectra give correlation for both the inter-residue connectivity, HN(i)-^{15}N(i)-C_α(i-1), via the one-bond $^1J_{NC\alpha}$ coupling and the intra-residue connectivity HN(i)-^{15}N(i)-C_α(i) via the two-bond $^2J_{HC\alpha}$ coupling, while CT-HN(CO)CA spectra give correlation for the intra-residue HN(i)-^{15}N(i)-C_α(i-1) resonances via the one-bond $^1J_{NCO}$ and $^1J_{C\alpha CO}$. In the CT-HNCA experiment, the signal intensity of the inter-reidue cross-peak, HN(i)-^{15}N(i)-C_α(i-1), was weaker than that of the intra-residue cross-peak, HN(i)-^{15}N(i)-C_α(i-1) because $^1J_{NCO}$ and $^1J_{C\alpha CO}$ coupling constants are larger than $^2J_{HC}$ coupling constant, so the author could easily distinguish the most part of the cross peaks of intra-residues from those of inter-residue ones. However, some signals degenerate and some cross peaks for inter-residue correlation are missing in the HNCA spectrum. In such a case it was difficult to make reliable assignment only using the HNCA spectrum. The CT-HN(CO)CA was very useful for such a case. Alternative pathway for the HN(i)-^{15}N(i)-C_α(i-1) correlation was provided by this experiment through the large $^1J_{NCO}$ and $^1J_{C\alpha CO}$ coupling constants ($^1J_{NCO}$, $^1J_{C\alpha CO} > {}^2J_{NC\alpha}$). Therefore, the signal intensity of the HN(i)-^{15}N(i)-C_α(i-1) correlation is higher than that of the CT-HNCA spectrum. Moreover, as only one correlation was observed in the CT-HN(CO)CA experiment, any ambiguity in distinguishing intra and inter-residue correlations in the CT-HNCA experiment could not be solved by this experiment. Some correlation was still difficult to be interpretated by signal degeneracy. The author employed the CT-HNCO and CT-HN(CA)CO experiments to establish the sequential connectivity through the carbonyl carbon resonance. The HNCO experiment correlates the inter-residue nuclei, HN(i)-N(i)-CO(i-1) via the one-bond $^1J_{NCO}$ coupling, and the HN(CA)CO experiment correlates the intra-residue nuclei, HN(i)-N(i)-CO(i), via the one-bond $^1J_{NC\alpha}$ and $^1J_{C\alpha CO}$ coupling and the inter-residue nuclei, HN(i)-N(i)-CO(i-1), via the two-bond $^2J_{C\alpha CO}$ coupling. Figure 7 shows an example of the sequential connectivity using the CT-HNCO and CT-HN(CA)CO experiments. The sequential connectivity from Gly 102 to Ser 98 through CO(i, i-1) connectivity is shown. Since

the HN(CA)CO experiment correlates the inter-residue, HN(i)-^{15}N(i)-CO(i-1), and the intra-residue HN(i)-^{15}N(i)-CO(i) connectivities in the same way as the HNCA experiment does. Therefore, the sequential connectivity was obtained in the same manner as the combination of the HNCA and HN(CO)CA experiments did.

In the CT-HNCO experiment, sensitivity was very high due to large one-bond $^{1}J_{NCO}$ coupling constants. The sensitivity of the CT-HN(CA)CO experiment was lower than that of CT-HNCO, due to the two step magnetization transfer and the relaxation of $^{13}C_{\alpha}$ nuclei. From these reasons, some resonance peaks disappeared in the CT-HN(CA)CO experiment. In such a case, the sequential connectivity through the carbonyl carbon was interrupted. To avoid this situation, the author employed the CT-HCACO experiment which correlates H_{α}(i)-C_{α}(i)-CO(i) via one-bond $^{1}J_{C\alpha CO}$ coupling, to compensate the lack of information on the intra-residue carbonyl carbon resonances. In addition, the combination of the HCACO, HNCO, and TOCSY-HSQC experiments could establish the sequential connectivity through the CO and Hα resonances. Figure 8 shows the sequential assignment from Glu 93 to Lys 95 using this combination.

These triple resonance experiments provided a large amount of imformation on the inter- and intra-residue chemical shifts. For example, for each nucleus of one residue in the backbone so many informations like HN(i, i±1), N(i), H_{α}(i, i-1), C_{α}(i, i-1), CO(i, i-1), are obtained. Thus the interpretation of each possible sequential connectivity by manual judgement was not efficient. For this reason, a program which sorts a chemical shift table and searches the sequential connectivities from 3D triple-resonance peak table was developed. The combination of this program and the result from the previously descrived specific labeling experiment provided the nearly complete sequential connectivities of the IRF-2(113) protein. Table 1 shows the result of the resonance assignment for the IRF-2(113) protein.

Side chain assignment by the HCCH-TOCSY experiment

Assignments of the side chain resonances were obtained from the HCCH-TOCSY experiment. In this experiment, ^{1}H magnetization that has frequency labeled in t1 evolution period is transferred from ^{1}H to its directly bound ^{13}C nucleus via $^{1}J_{CH}$ coupling, and after ^{13}C evolution t2 period, magnetization was transferred through the ^{13}C spin system via $^{1}J_{CC}$ coupling by an isotropic mixing pulse. The ^{13}C magnetization is transferred back to the directly attached ^{1}H and detected in t3 period. The resulting spectrum has the feature of a 2D ^{1}H-^{1}H TOCSY spectrum ($F_1(^{1}H)$-$F_3(^{1}H)$) which was edited by the ^{13}C chemical shift in F_2 dimension. Figure 9 shows the correlations of isoleucine residues in the $F_1(^{1}H_{\gamma methyl})$-$F_2(^{13}C_{\gamma methyl})$-$F_3(^{1}H)$ and $F_1(^{1}H_{\delta metyl})$ - $F_2(^{13}C_{\delta metyl})$-$F_3(^{1}H)$ regions. In this spectrum, the $^{1}H_{\alpha}$, $^{1}H_{\beta}$, $^{1}H_{\gamma,\delta\text{-metyl}}$ and $^{1}H_{\gamma}$ resonances can be distinguished from the other ^{1}H resonances of the same isoleucine residue.By using this HCCH-TOCSY experiment, approximately 80% of the

side chain resonances could be assigned.

Assignment of the aromatic ring protons

Other spin systems which include the aromatic ring protons of Trp, Phe and His are assigned by the combination of conventional homonuclear 2D spectra and the deuterium labeling method. IRF-2(113) has 6 Trp, 4 Phe and 4 His residues and so the author could not assign them only by 2 D ^{1}H-^{1}H HOHAHA and COSY spectra. The IRF-2(113) protein was labeled with deuterated Trp and Phe and 2D homonuclear ^{1}H-^{1}H HOHAHA spectra were measured (Figures 10a and b). The result of the ring-proton resonances are shown in Table1. Figure 10a shows the aromatic ring proton region of homonuclear ^{1}H-^{1}H HOHAHA spectrum of the IRF-2(113) protein which was labeled with deuterated Trp residues. In this spectrum, the spin systems due to Phe and His residues were only observed in this region (2H ring proton region is eliminated from this figure). In this figure, Phe 34 and Phe 106 ring proton spin systems are indicated as examples. As shown in Figure 10a, Phe 34 ring proton spin system shows a typical correlation pattern which often could be seen in a text book (for example, Wüthrich, (1986)). Typical order of arrangement of Phe ring proton resonances are (2H, 6H), 4H, and (3H, 5H) proton from upfield to downfield resonance and cross-peaks are observed between these resonances (see Phe 34 cross-peak pattern). On the other hand, Phe106 ring proton spin system is very unusual. The 4H ring proton resonance degenerates on the (2,6H) or (3,5H) proton resonances. This phenomenon is explained by the ring current effect from another aromatic ring. Indeed, as shown in the next chapter, Phe 106 is located near Trp 38 in the terminal of the β-sheet.. Figure10b shows the same region of homonuclear ^{1}H-^{1}H HOHAHA spectrum of IRF-2(113) which was labeled with deuterated Phe residues. In this spectrum in contrast to Figure10a, the spin systems due to the Trp and His residues are only observed (His 2H ring proton region is eliminated from this figure). In this figure, Trp 26 and Trp 38 ring proton spin systems are indicated as examples. As same as the Phe 34 spin system, the Trp26 ring proton spin system shows typical correlation pattern of Trp ring proton resonances. They are generally arranged in the ordered of 5H, 6H, 7H, and 4H from upfield and cross-peaks among these resonances (see Trp 26 cross-peak pattern). However, for the Trp 38 ring proton resonances, this arrangement could not be observed. Apparently, the cross-peak pattern looks like a normal pattern, but the 7H ring proton resonance shifts to very upfiled and the 5H ring proton shifts to lowerfied (this feature was confirmed by the ^{1}H-^{13}C HSQC spectrum in the aromatic carbon and proton region (data not shown)). As discussed above, Trp 38 and Phe 106 are located at near positions in the terminal region of the β-sheet, so their chemical shifts may be affected by the ring current effect. By the combination of these spectra and homonuclear ^{1}H-^{1}H NOESY spectra (data not shown), the most of ring proton resonances could be assigned.

Analysis of ^{13}C chemical shift deviations IRF-2(113) from random-coil chemical shifts

Recently, empirical correlations between $^{13}C_a$ chemical shifts and secondary structures have been reported for some proteins (Spera and Bax, 1991; Wishart *et al*., 1991; Fairbrother *et al.*, 1992). Downfield shifts for $^{13}C_a$ (about 3 ppm in average) are correlated with a helical conformation and upfieled shifts for $^{13}C_a$ (about 2 ppm in average) are correlated with a β-strand or extended conformation. Figure 11 shows the chemical shift differences between the $^{13}C_a$ chemical shifts of IRF-2(113) and those of random-coil chemical shifts plotted versus residue numbers.

$^{13}C_a$ chemical shifts of IRF-2(113) leads us to the estimation of the positions of the secondary structures of IRF-2(113). In the regions from Arg 7 to Ile 21, from Leu 54 to Ile 60, from Lys 75 to Leu 88 were assumed to be in helical structures and the regions from Gly 23 to Leu 27, from Lys 29 to Lys 50, from Ile 90 to Met 111 were thought to have β-strands or extended conformations. Actually, the prediction based on the ^{13}C chemical shifts deviation is quite consistent with that determined by short- and medium range NOEs discussed in the next chapter. The information of the chemical shifts provides us another insight into the secondary structure. Unusual chemical shifts reflect specific environments where residues are located. ^{1}H chemical shifts of Ile 16 shows remarkable upfield shift (Table 1). This phenomenon can be explained by the ring current effect. Therefore, we can assume that Ile16 residue may be located closely to the aromatic side chain and form a hydrophobic core.

Concluding Remark

For the structure analysis by the NMR experiment, a sequential assignment of the backbone nuclear magnetic resonances is the first step for the determination of the structure. In this chapter, the author has described the use of a series of 2D and 3D double- and triple- resonance NMR experiments for the assignments of ^{1}H, ^{15}N, ^{13}C, and ^{13}CO resonances of the IRF-2(113) protein. This information is absolutely essential for the determination of the secondary and tertiary structure of the IRF-2(113) protein. The side-chain assignments were provided by the HCCH-TOCSY experiment for the removal of ambiguity. The author could assign approximately 80 % of the side chain resonances by this procedure.

References

Bax,A., Ikura,M., Kay,L.E., Torchia,D.A. and Tschudin,R (1990) *J.Magn.Reson.* **36**, 304-318.

Bax,A., Clore,G.M. and Gronenborn,A.M. (1990b) *J.Magn.Reson.* **88**, 425-431.

Bax,A., Ikura,M., Kay,L.E. and Guang,Z. (1991) *J.Magn.Reson.* **91**, 174-178.

Bodenhausen,G., and Ruben,D.J. (1980) *Chem.phys.Lett.* **69**, 185-189.

Fairbrother,W.J., Cavanagh,J., Dyson,H.J., Palmer,A.G., Sutrina,S.L., Reizer,J., Saier,M.H. and Wright,P.E. (1991) *Biochemistry* **30**, 6896-6907.

Grzesiek,S. and Bax,A. (1192) *J.Magn.Reson.* **96**, 432-440.

Harada,H., Fujita,T., Miyamoto,M., Kimura,Y.,Murayama,M., Furia,A., Miyata,T. and Taniguchi,T. (1989) *Cell*, **58**, 729-739.

Ikura.M., Kay.L.E., and Bax,A. (1990) *Biochemistry* **29**, 4659-4667.

Kay.L.E., Ikura,M., Tschudin.R and Bax,A. (1990) *J.Magn.Reson.* **89**, 496-514.

Kainosho,M. and Tsuji,T. (1982) *Biochemistry*, **21**, 6273-6299.

Marion,D. and Bax,A. (1988) *J.Magn.Reson.* **79**, 352-356

Marion,D., Ikura,M. and Bax,A. (1989) *J.Magn.Reson.* **84**, 425-430.

Messerle,B.A., Wider,G., Otting,G., Weber,C. and Wüthrich,K. (1989) *J.Magn.Reson.* **85**, 608-613.

Norwood,T.J., Boyd,J., Heritage,J.E., Sogge,N. and Campbell,I.D. (1990) *J.Magn.Reson.* **87**, 488-501

Powers,R., Gronengorn,A.M., Clore,G.M. and Bax,A. (1991) *J.Magn.Reson.* **94**, 209-213.

Richarz,R., and Wüthrich,K. (1978) *Biopolymers* **17**, 2133-2141.

Shaka,A.J., Barder,P.B. and Freeman,R. (1985) *J.Magn.Reson.* **64**, 547-552.

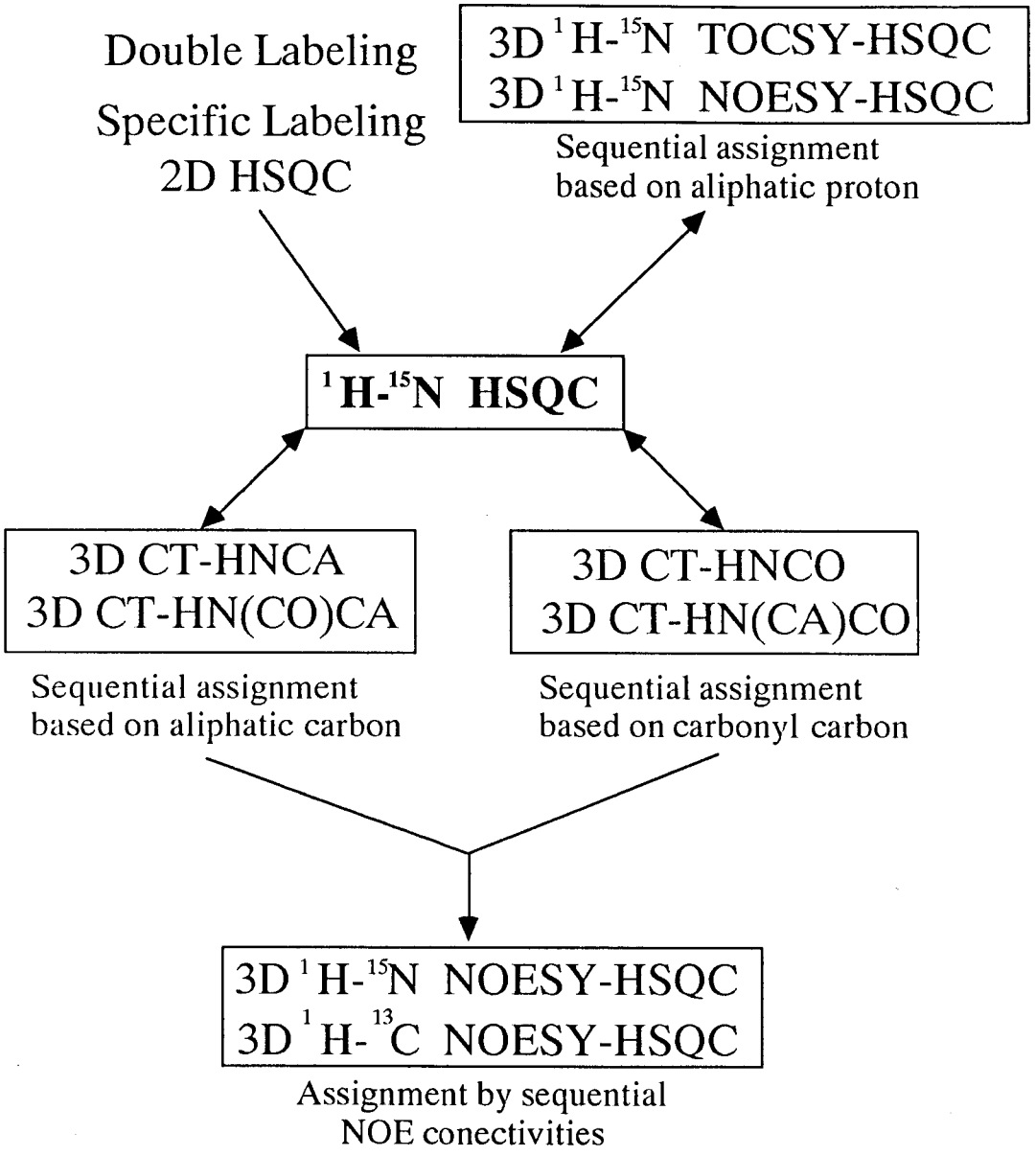
Shaka,A.J., Lee,C.J. and Pines,A. (1988) *J.Magn.Reson.* **77**, 274-293.

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

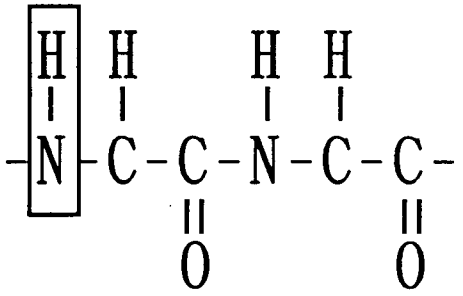
Zhu.G. and Bax,A. (1990) *J.Magn.Reson.* **90**, 405-410.

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

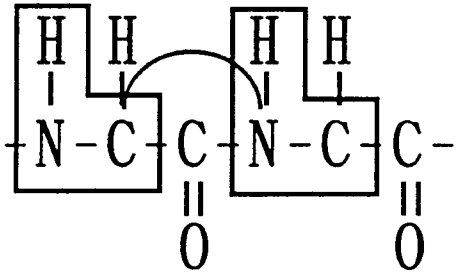
Figure 1


Double Labeling
Specific Labeling
2D HSQC
3D ^{1}H-^{15}N TOCSY-HSQC
3D ^{1}H-^{15}N NOESY-HSQC
Sequential assignment
based on aliphatic proton
^{1}H-^{15}N HSQC
3D CT-HNCA
3D CT-HN(CO)CA
3D CT-HNCO
3D CT-HN(CA)CO
Sequential assignment
based on aliphatic carbon
Sequential assignment
based on carbonyl carbon
3D ^{1}H-^{15}N NOESY-HSQC
3D ^{1}H-^{13}C NOESY-HSQC
Assignment by sequential
NOE conectivities

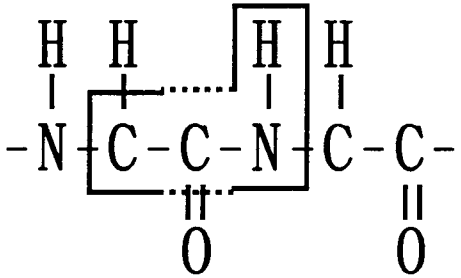
a) HSQC



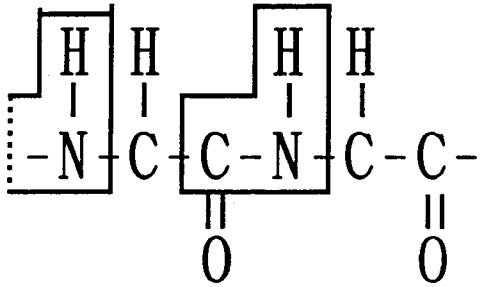
b) HNCA



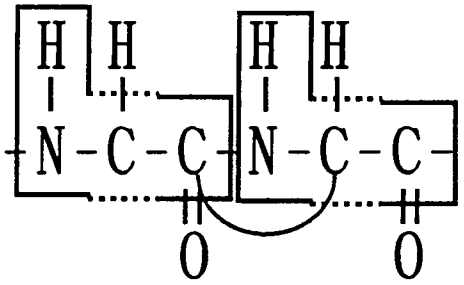
c) HN(CO)CA



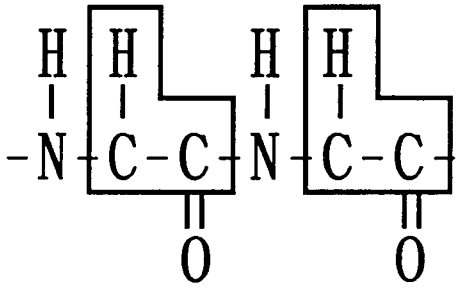
d) HNCO



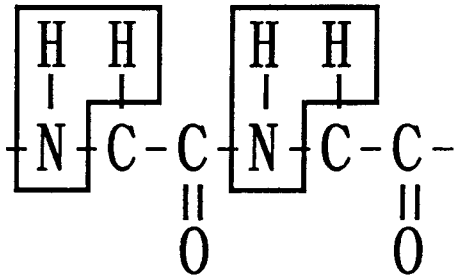
e) HN(CA)CO



f) HCACO



g) TOCSY-HSQC



h) HCCH-TOCSY

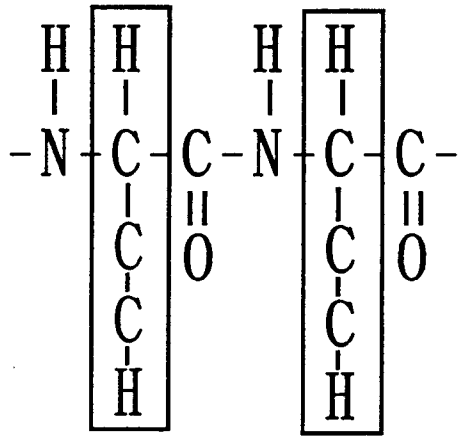


Figure 2

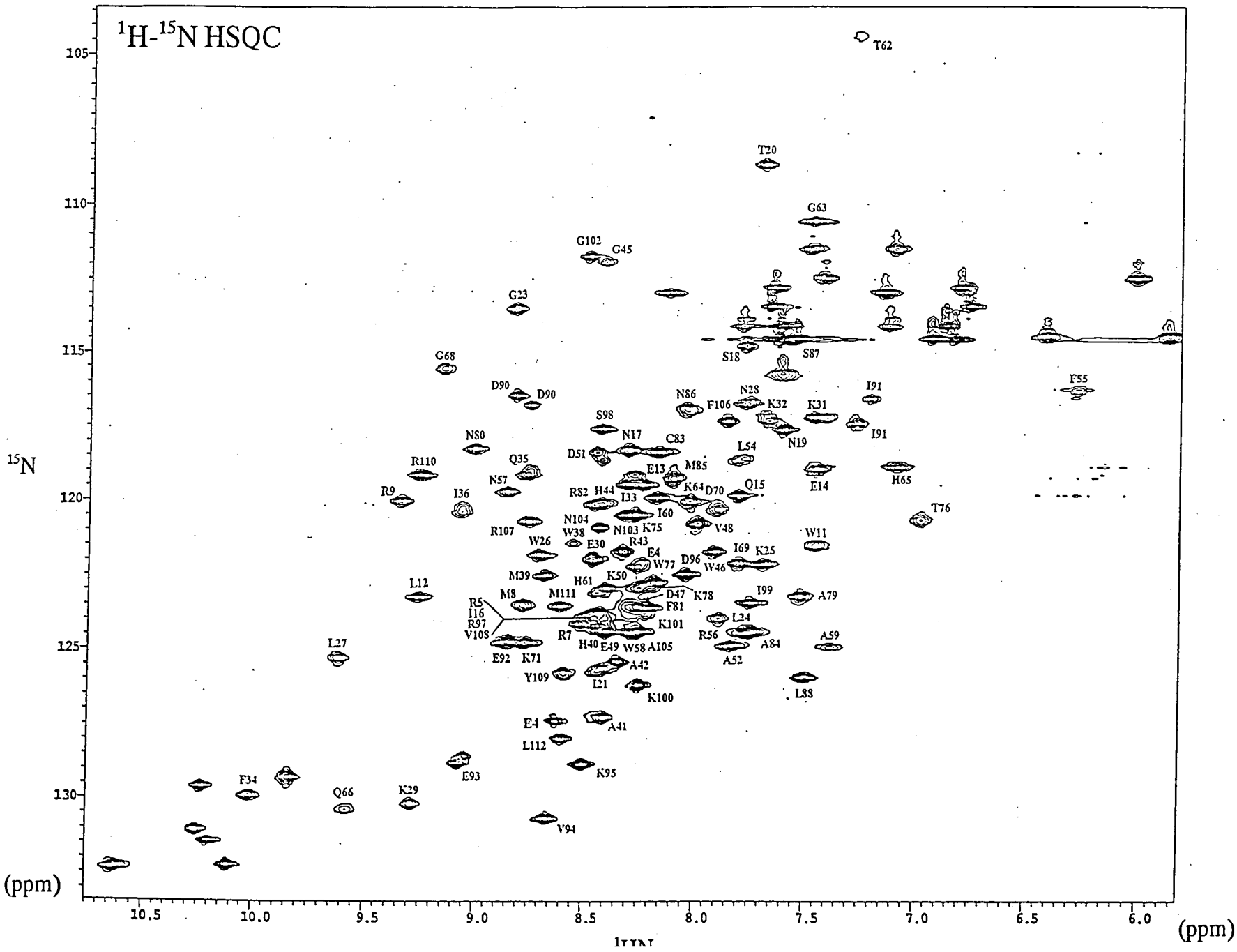

^{1}H-^{15}N HSQC
^{15}N
(ppm)
(ppm)
105
110
115
120
125
130
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
T62
T20
G63
G102
G45
G23
S18
S87
G68
D90
D90
N86
F106
N28
K32
K31
I91
I91
F55
S98
N17
C83
N19
L54
N80
D51
Q35
M85
E13
R110
N57
E14
H65
R9
I36
R82
H44
I33
K64
D70
Q15
T76
I60
N104
R107
W38
N103
K75
V48
W11
W26
E30
R43
E4
I69
K25
W77
D96
W46
M39
H61
K50
L12
M8
M111
D47
K78
I99
A79
R5
I16
R97
V108
F81
L24
K101
R7
R56
A59
L27
H40
E49
W58
A105
A84
E92
K71
A42
A52
Y109
L21
K100
L88
E4
A41
L112
E93
K95
F34
Q66
K29
V94

Figure 3

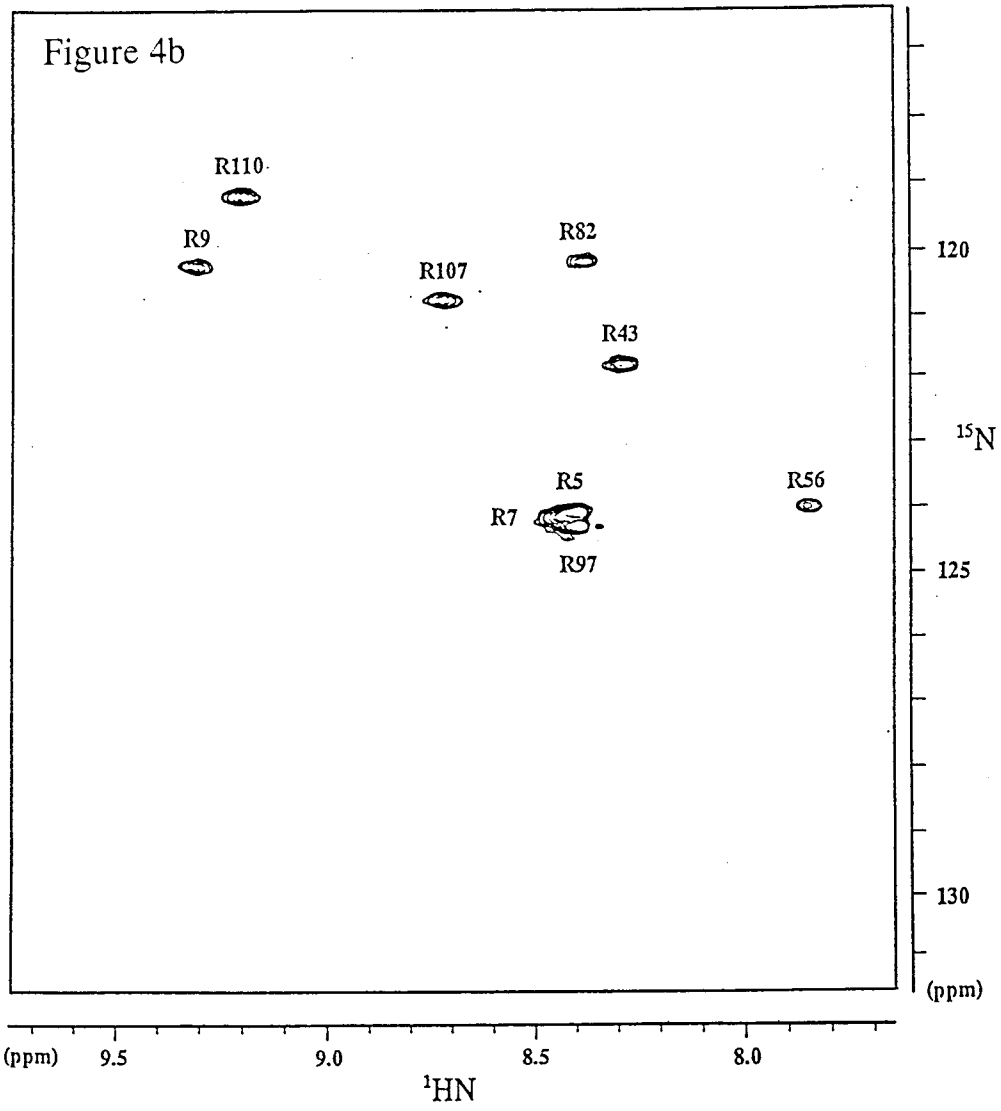


Figure 4b
R110
R9
R82
R107
R43
R5
R7
R97
R56
120
125
130
(ppm)
15N
(ppm)
9.5
9.0
8.5
8.0
1HN

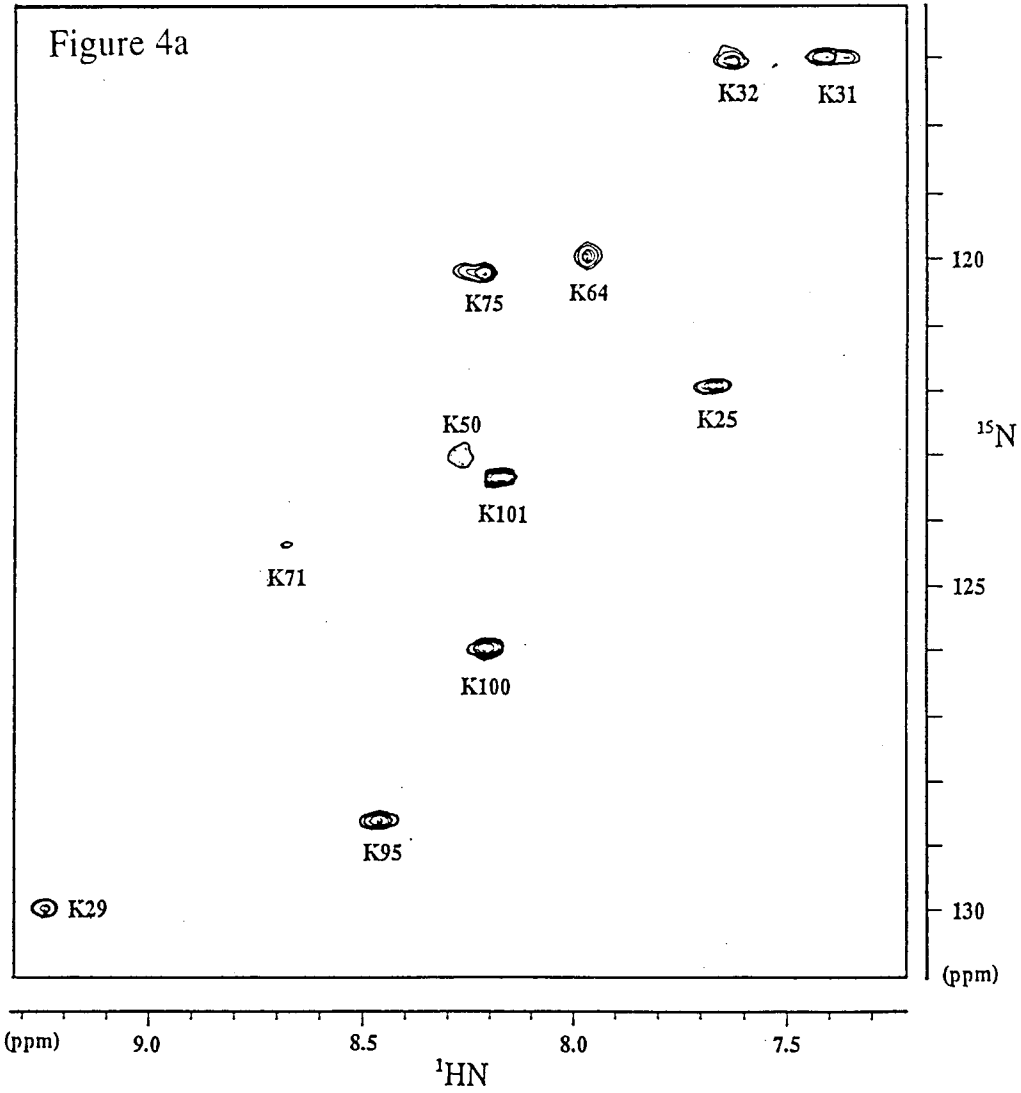


Figure 4a
K32
K31
K75
K64
K25
K50
K101
K71
K100
K95
K29
120
125
130
(ppm)
15N
(ppm)
9.0
8.5
8.0
7.5
1HN

Figure 4

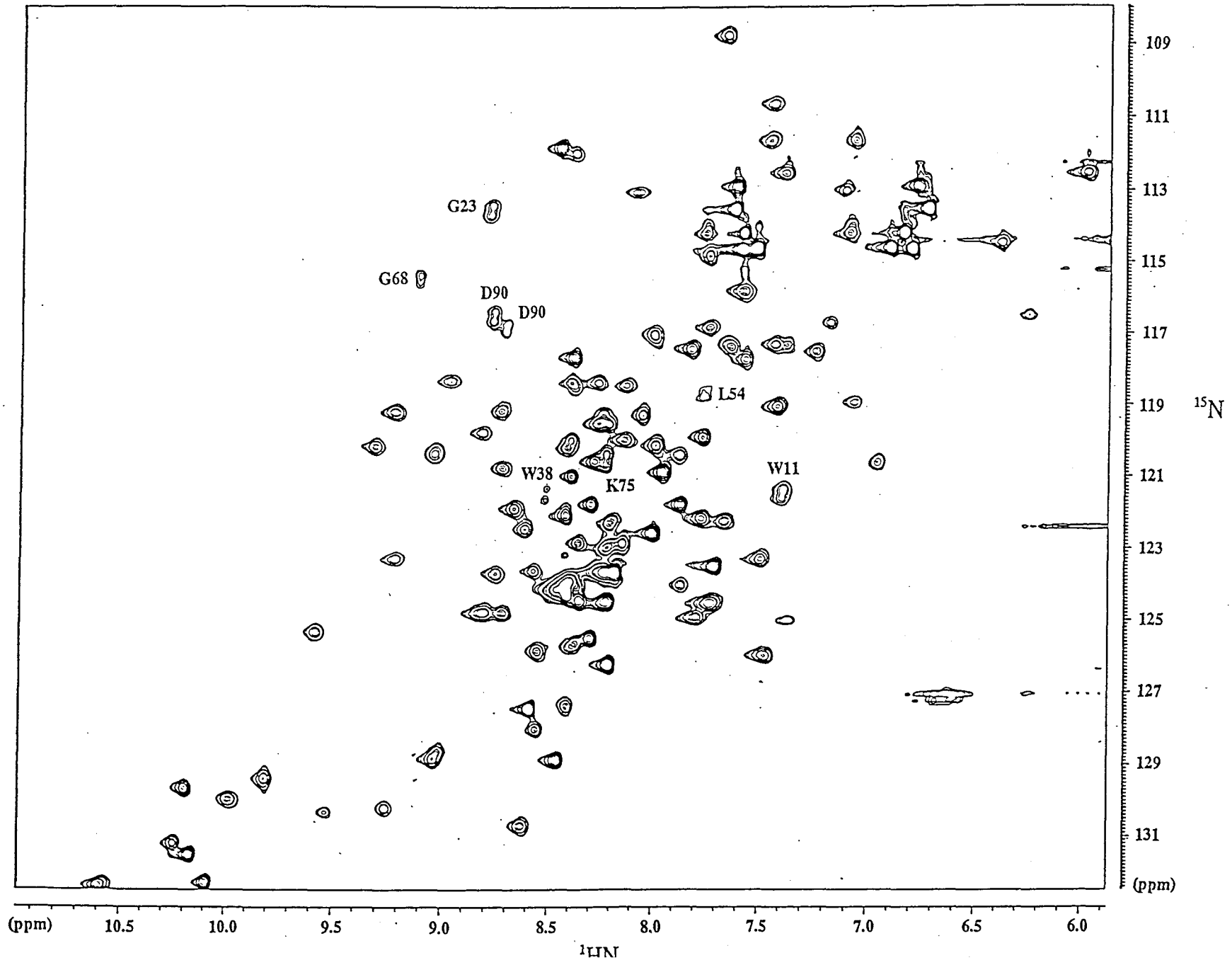

G23
G68
D90
D90
L54
W38
K75
W11
109
111
113
115
117
119
121
123
125
127
129
131
(ppm)
^{15}N
(ppm)
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
^{1}HN

Figure 5

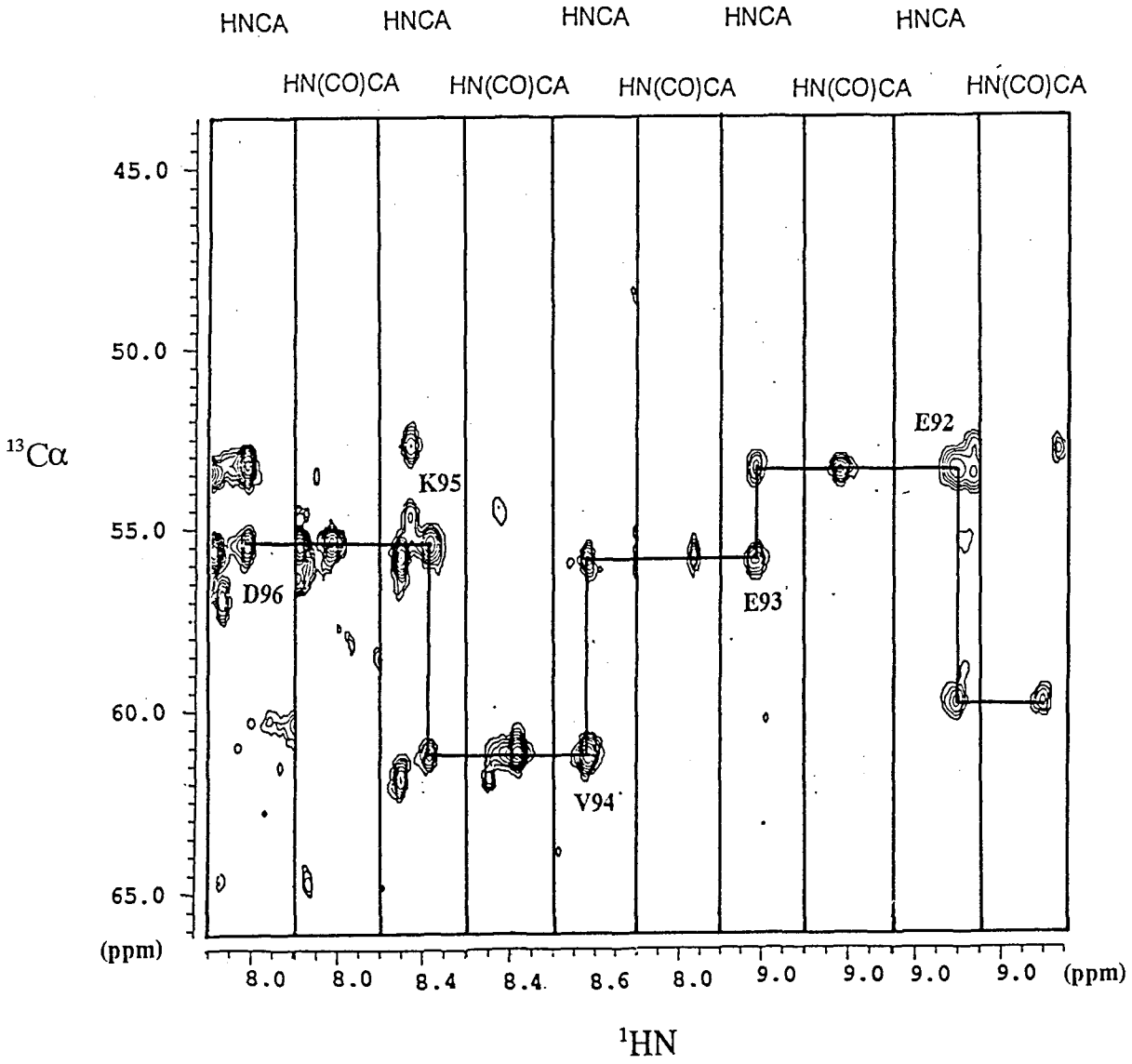

HNCA
HNCA
HNCA
HNCA
HNCA
HN(CO)CA
HN(CO)CA
HN(CO)CA
HN(CO)CA
HN(CO)CA
45.0
50.0
55.0
60.0
65.0
(ppm)
$^{13}C\alpha$
K95
D96
V94
E93
E92
8.0
8.0
8.4
8.4
8.6
8.0
9.0
9.0
9.0
9.0
(ppm)
^{1}HN

Figure 6

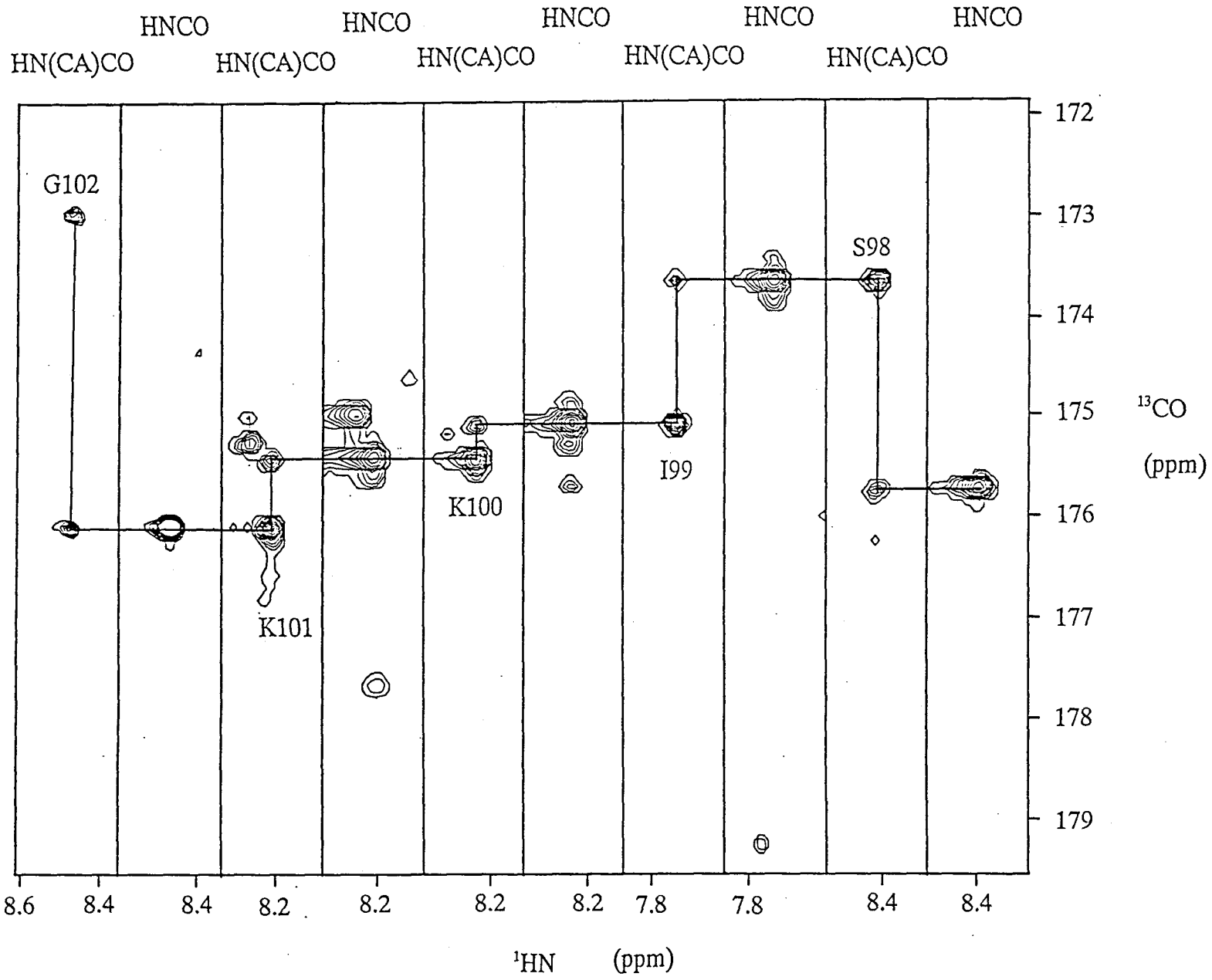

HN(CA)CO
HNCO
HN(CA)CO
HNCO
HN(CA)CO
HNCO
HN(CA)CO
HNCO
HN(CA)CO
G102
K101
K100
I99
S98
172
173
174
175
176
177
178
179
13CO (ppm)
8.6
8.4
8.4
8.2
8.2
8.2
8.2
7.8
7.8
8.4
8.4
1HN (ppm)

Figure 7

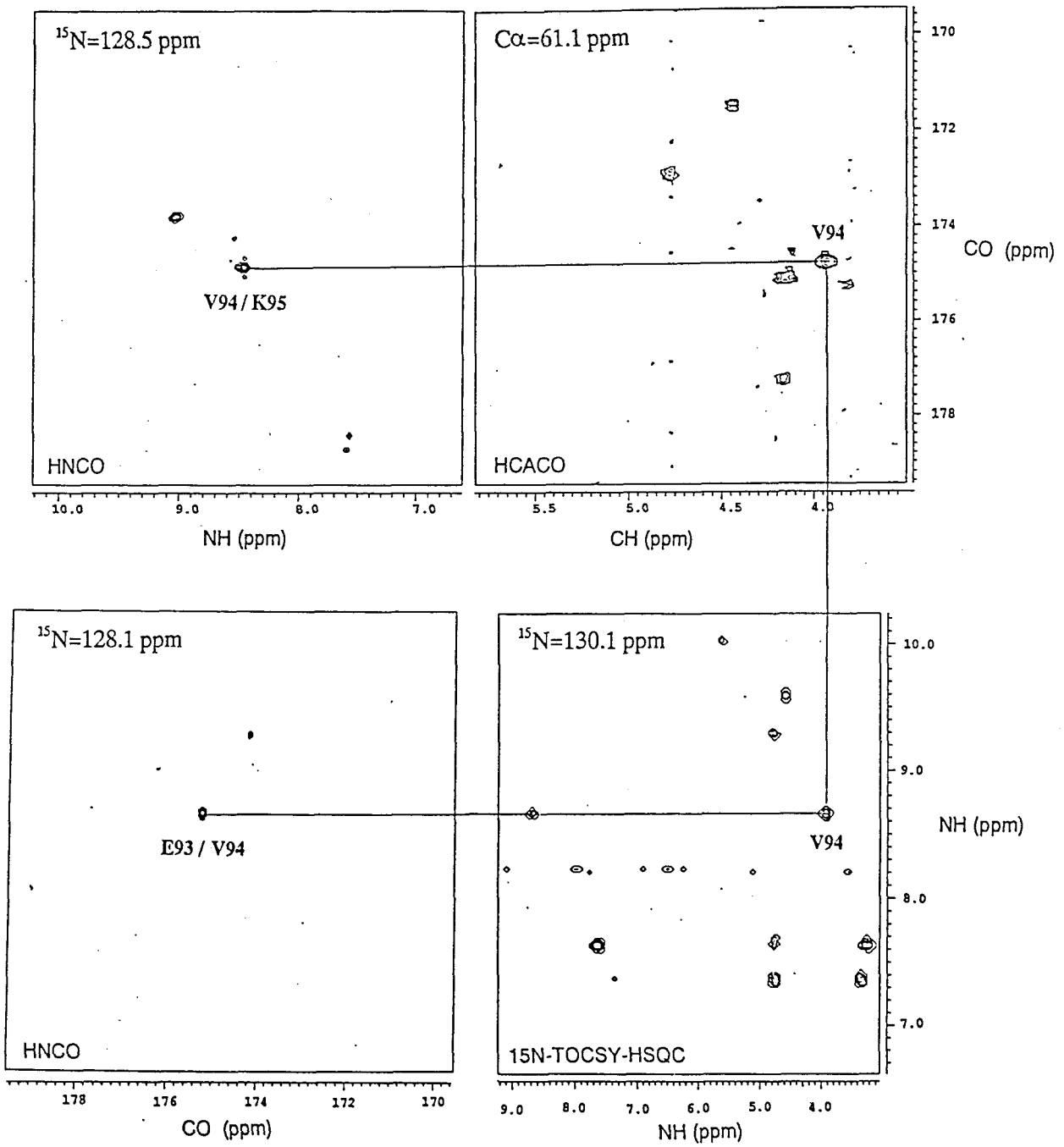

^{15}N=128.5 ppm
V94 / K95
HNCO
10.0
9.0
8.0
7.0
NH (ppm)
Cα=61.1 ppm
V94
HCACO
5.5
5.0
4.5
4.0
CH (ppm)
170
172
174
176
178
CO (ppm)
^{15}N=128.1 ppm
E93 / V94
HNCO
178
176
174
172
170
CO (ppm)
^{15}N=130.1 ppm
V94
15N-TOCSY-HSQC
9.0
8.0
7.0
6.0
5.0
4.0
NH (ppm)
10.0
9.0
8.0
7.0
NH (ppm)

Figure 8

Figure 9

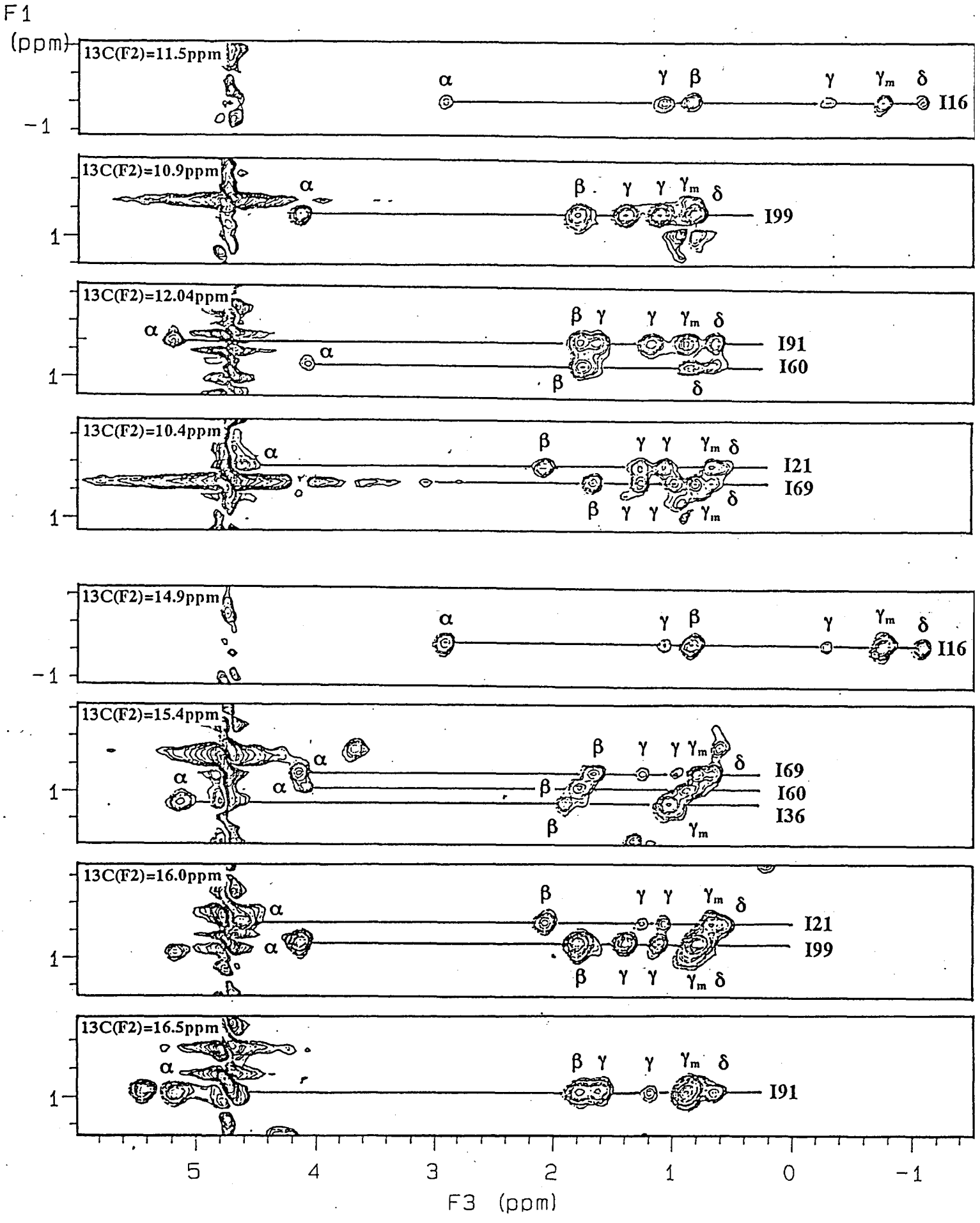

F1
(ppm)
13C(F2)=11.5ppm
α
γ β
γ γm δ
I16
−1
13C(F2)=10.9ppm
α
β γ γ γm δ
I99
1
13C(F2)=12.04ppm
α
β γ
γ γm δ
I91
α
I60
β
δ
1
13C(F2)=10.4ppm
α
β
γ γ γm δ
I21
I69
δ
β γ γ γm
1
13C(F2)=14.9ppm
α
γ β
γ γm δ
I16
−1
13C(F2)=15.4ppm
α
α
α
β γ γ γm δ
I69
β
I60
β
I36
γm
1
13C(F2)=16.0ppm
α
α
β
γ γ γm δ
I21
I99
β γ γ γm δ
1
13C(F2)=16.5ppm
α
β γ
γ γm δ
I91
1
5
4
3
2
1
0
−1
F3 (ppm)

Figure 10a

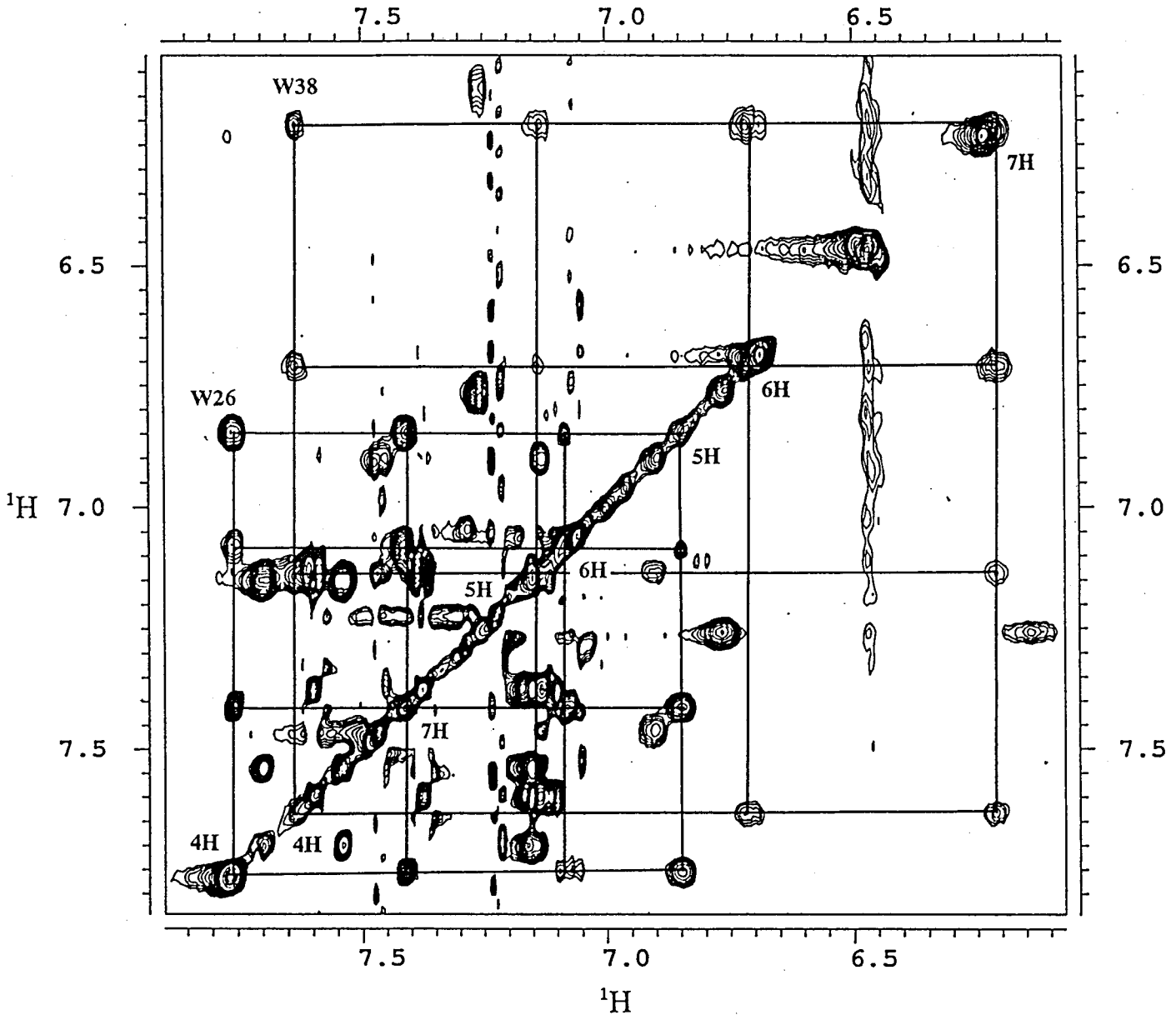
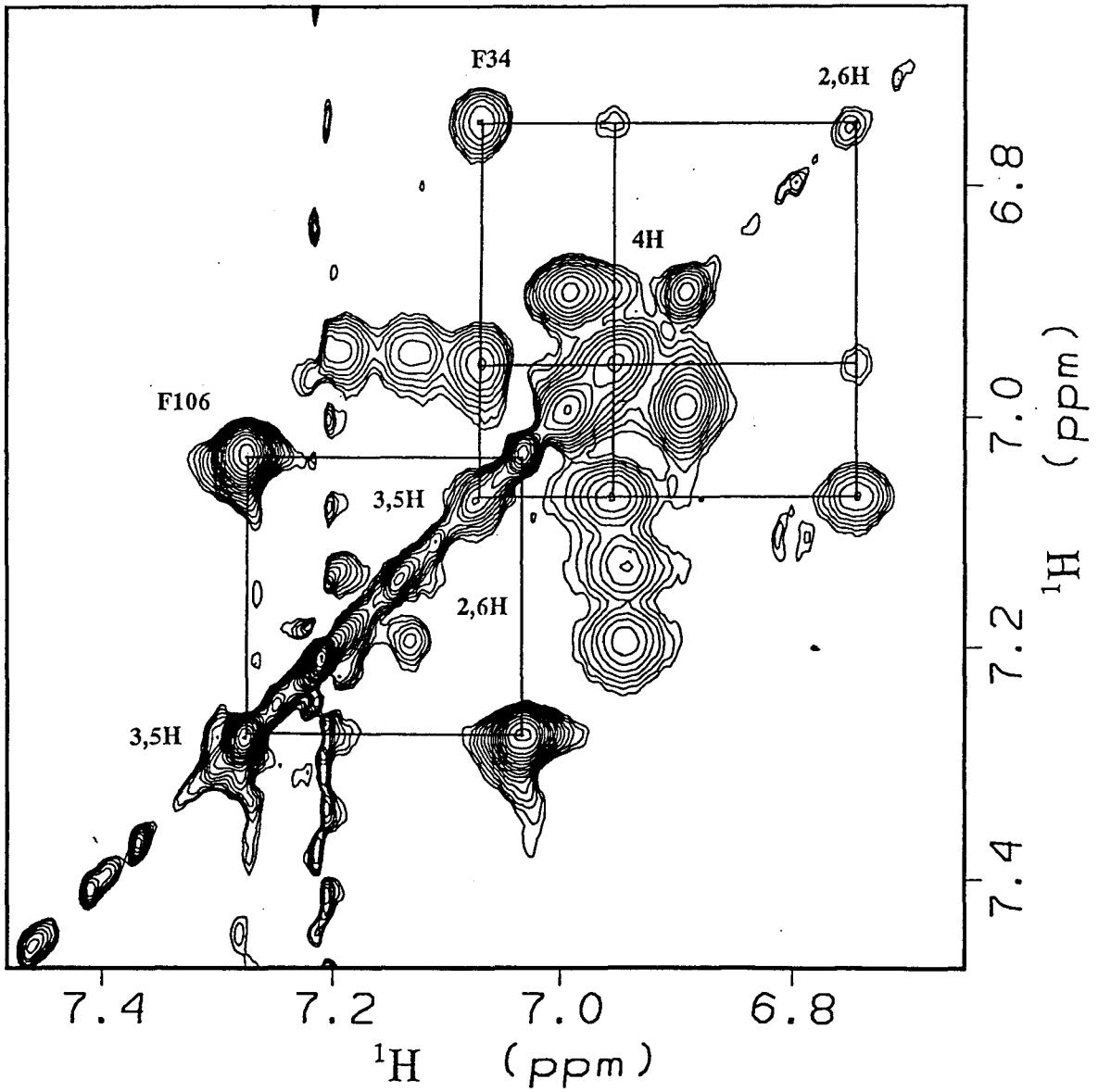

7.5
7.0
6.5
W38
7H
6.5
6.5
W26
6H
5H
^{1}H 7.0
7.0
6H
5H
7H
7.5
7.5
4H
4H
7.5
7.0
6.5
^{1}H

Figure 10b



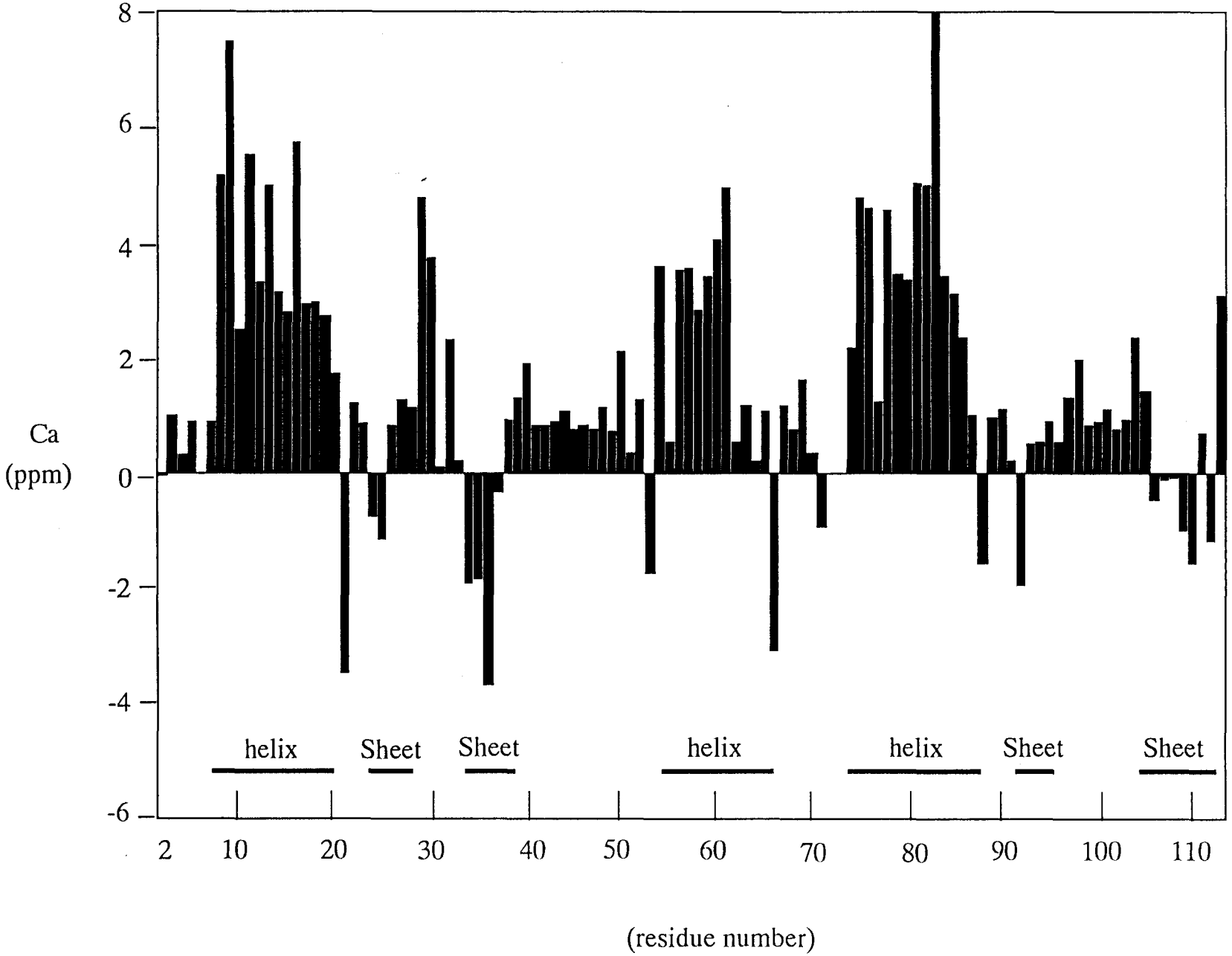

8
6
4
2
0
-2
-4
-6
Ca
(ppm)
helix
Sheet
Sheet
helix
helix
Sheet
Sheet
2
10
20
30
40
50
60
70
80
90
100
110
(residue number)

Figure 11

Figure legends

Figure 1.

Overview of assignment strategy.

Figure 2.

Summary of correlations and connectivities observed in heteronuclear NMR experiments used for the assignment of the spectra of IRF-2(113)

Figure 3.

500-MHz ^{1}H-^{15}N HSQC spectrum of uniformly ^{15}N-labeled IRF-2(113) at 25 °C and pH 5.5, with spin lock pulse for solvent suppression. The assignments are indicated by amino acids numbers.

Figure 4

Sections of 500-MHz ^{1}H-^{15}N HSQC spectra of the IRF-2(113) labeled with (a) [^{15}N] lysine and (b) [^{15}N] arginine, at 25 °C and pH 5.5, with spin lock pulse for solvent suppression. Assignments are indicated by amino acids numbers. The box corresponds to the cross-peak of lower intensity.

Figure 5

500-MHz ^{1}H-^{15}N HSQC spectrum of the IRF-2(113)doubly labeled with ^{15}N and [1-^{13}C] proline, at at 25 °C and pH 5.5, with spin lock pulse for solvent suppression. Assignments are indicated by amino acids numbers.

Figure 6.

Composite spectrum consisting of strips taken from ^{15}N planes of 500-MHz 3D HNCA and HN(CO)CA spectra of the ^{15}N-^{13}C labeled IRF-2(113). The sequential assignment from Asp96 to Glu92 is indicated.

Figure 7

Composite spectrum consisting of strips taken from ^{15}N planes of 500-MHz 3D HN(CA)CO and HNCO spectra of ^{15}N - ^{13}C labeled IRF-2(113). The sequential assignment from Gly102 to Ser98 is indicated.

Figure 8.

Composite spectrum consisting of strips taken from ^{15}N and $^{13}C_{\alpha}$ planes of 500-MHz 3D

HNCO, HCACO and TOCSY-HSQC spectra of ^{15}N and ^{15}N-^{13}C labeled IRF-2(113). The sequential assignment from Lys95 to Glu93 is indicated.

Figure 9.

Selected $^{1}H(F1)$-$^{1}H(F3)$ planes at different ^{13}C chemical shifts of the 600-MHz HCCH-TOCSY spectrum of ^{15}N-^{13}C labeled IRF-2(113). Connectivities originate from the δ–methyl- and γ–methyl protons of Ile residues.

Figure 10.

Aromatic ring proton regions of 500-MHz homonuclear HOHAHA spectra of the IRF-2(113) labeled with a) ^{2}H-Phe labeled IRF-2(113) and b) ^{2}H-Trp labeled IRF-2(113). Trp26 and Trp38 spin system and Phe34 and Phe106 spin system are indicated, respectively.

Figure 11.

Plot of the differences between $^{13}C_{\alpha}$ chemical shifts of each amino acid residue in IRF-2(113) and randam coil shifts (Richarz and Wüthrich, 1978).

Chapter-3

Determination of Secondary Structure and Folding Topology of IRF-2(113) by Using Heteronuclear Magnetic Resonance Spectroscopy

Abstract

The secondary structure of the DNA-binding domain of the mouse IRF-2 protein has been elucidated by three-dimensional (3D) ^{15}N- and ^{13}C-edited nuclear Overhauser (NOE) spectroscopy on the basis of the ^{1}H, ^{15}N, ^{13}C assignments described in the previous chapter. From the NOE connectivities between the HN, $C_{\alpha}H$, and $C_{\beta}H$ protons, as well as $^{3}J_{HN\alpha}$ coupling constants, amide exchange rates and $^{13}C_{\alpha}$ chemical shift deviation, IRF-2(113) has been shown to have a unique secondary structure composition different from those of known DNA-binding proteins. This protein has three helices (residue7- 20, 54-64, and 75-88), and an antiparallel β-sheet consisting of four strands (residue 24-27, 32-37, 91-94, and 106-112). Based on the long-range NOEs between helices and the β-sheet obtained from ^{13}C-edited NOESY and the homonuclear NOESY experiment, the topological arrangement of the secondary structure elements of IRF-2(113) has been shown.

Introduction

Interferon Regulatory Factor-2 (IRF-2), a transcriptional repressor, has been identified as a negative transcriptional regulator of type I interferon and interferon-inducible genes. Its DNA binding domain locates in the N-terminal region which consists of 112 amino acid residues, and binds to the consensus sequence AAGTGA repeats (see Chapter 1). This DNA binding domain does not show apparent sequence homology with the known DNA binding motifs (helix-turn-helix (HTH), basic-leucine zipper (bZip), Zn-finger, *etc*). As was discussed in Chapter 1, the DNA binding domain is expected to span from the N-terminal to the 113th amino acid residue predicted from the DNA sequence. The recombinant DNA binding domain (IRF-2(113)) binds to the consensus sequence in the same manner as the native IRF-2 does. So, the recombinant IRF-2(113) is a good model for the analysis of the DNA-binding property of IRF-2 by NMR. In the preceding Chapter 2, the ^{1}H,^{15}N, ^{13}C, and ^{13}CO assignments of IRF-2(113) which comprise the essential base for structure analyses were obtained. In this chapter, the author will present the determination of the secondary structures of the IRF-2(113) protein derived from the analyses of 3D heteronuclear-edited NOE spectra, ^{13}C chemical shifts, the amide proton exchange rates, and $^{3}J_{HN\alpha}$ coupling constants. It will be shown that IRF-2 has a unique topology of the folding of the secondary structure.

Materials and Methods

Sample Preparation

Uniformly, ^{15}N and $^{15}N/^{13}C$ - labeled IRF-2(113) were expressed in *E.coli.* and purified in the same way as described in the previous chapter. Samples for NMR measurements contain 0.7 mM- 3 mM of the IRF-2(113) protein, in either 90% H_2O/10% D_2O or 99.996% D_2O at pH 5.5.

NMR Spectroscopy

All NMR spectra were recorded at 25 °C on a Brucker AMX 500 spectrometer equipped with a H-X probe. Quadrature detection in the indirectly detected dimension was obtained in all 3D NMR experiments using TPPI-States method (Marion *et al.*, 1989) except the t1 dimension in ^{13}C-edited NOESY experiment. For the t1 dimension in ^{13}C-edited NOESY experiment, States method (States *et al.*, 1982) was used for quadrature detection. The t1 and t2 sampling delays for all NMR experiments were adjusted to one - half of the respective increment, resulted in a 90 ° (zero-order) and -180 ° (first-order) phase value. The data were processed with an X32 computer, or were transferred to a Silicon Graphics Iris 4D/35GT workstation for processing and analysis using a FELIX software package (Hare Research, Inc.) or a NMRZ software package (New Methods Research, Ind., Syracuse, NY).

3D ^1H-^{15}N NOESY-HSQC experiment was performed in essentially the same way as described in the previous chapter. 3D ^1H-^{13}C NOESY-HSQC experiment was performed as described by Shirakawa *et al.* (1993) with mixing time 100 ms. The ^1H carriers (t1 and t3) were placed at 2.55 ppm and at 4.75 ppm on the water resonance, respectively. ^{13}C carrier (t2) was set at 41.33 ppm. The spectra widths were 3750,9 Hz (t1), 4464.3 Hz (t2), and 12500 Hz (t3), respectively. The acquired data matrix had 128 (t1) × 32 (t2) × 1024 (t3) complex points. One zero-filling was used in t1 (^1H) and t2 (^{13}C) dimension. The final data matrix consisted of 256 × 64 × 1024 points and Fouier - transformed into 64 series of F2 slice.

3D HMQC-NOESY-HMQC experiment was performed in essentially the same way as described by Ikura *et al.*, (1990) with mixing time 100 ms. The proton carrier was placed on the water resonance at 4.75 ppm, and ^{15}N carrier (t1 and t2) was placed at 119.6 ppm. The spectra widths were 1165.7 Hz (t1), 750.13 Hz (t2), and 12500 Hz (t3), respectively. The acquired data matrix had 64 (t1) × 32 (t2) × 1024 (t3) complex points. One zero-filling was used in t1 (^{15}N) and t2 (^{15}N) dimension.

2D HMQC-J experiment (Kay *et al.*, 1990) of uniformly ^{15}N labeled IRF-2(113) was performed to determine dihedral angle $^3J_{HN}$ coupling constants. Quadrature detection in t1 dimension was obtained by the TPPI-States method (Marion *et al.*, 1989). The proton carrier in the recording dimension was

placed on the water resonance at 4.75 ppm, and the ^{15}N carrier (t1) was placed at 119.6 ppm. The spectra widths were 1165.7 Hz (t1), and 12500 Hz (t2), respectively. The acquired data matrix consisted of 512 (t1) × 1024 (t2) complex data points. One zero-filling was used in t1 (^{15}N) and t2 (^{1}H) dimensions.

Slowly exchangeable amide protons were identified by measuring a series of ^{1}H-^{15}N HSQC experiments (Bodenhausen and Ruben, 1980; Bax *et al.*, 1990; Norwood *et al.*, 1989) over 24 hrs starting from 10 min just after dissolving lyophilized protein in D_2O. The residual water signal was suppressed by preirradiation during the relaxation delay. The acquired data matrix had 64 (t1) × 1024 (t2) complex data points. Two zero-filling was used in t1 (^{15}N). The time taken per one experiment was 23 min.

Results and Discussion

Secondary structure determination

The location of the secondary structure elements can be estimated from the NOE connectivity involving the backbone protons which shows secondary structure specific patterns, together with the amide proton exchange rates and $^{3}J_{HN}$ coupling constants (Wüthrich, 1986; Clore and Gronenborn, 1987).

Especially, sequential NOE connectivities provide a basis for the identification of the type of secondary structure elements. Stretches of strong sequential HN(i)-HN(i+1) and weak $C_\alpha H$(i)-HN(i+1) NOEs in combination with $C_\alpha H$(i)-HN (i+1, 2, 3, 4), HN(i)-HN(i+2) and $C_\alpha H$(i)-$C_\beta H$(i+3) NOEs are characteristic of an α-helix. A stretch of strong sequential $C_\alpha H$(i)-HN(i) NOEs in combination with interstrand NOEs involving HN and $C_\alpha H$ protons is characteristic of a β-sheet.

As mentioned in Chapter 2, the dispersion of the ^{1}H chemical shifts of IRF-2(113) was very limited, especially the signals of the aliphatic protons severely overlapped. Therefore, it is not feasible to use only 2D spectra. Thus, the sequential NOE connectivity was obtained from 3D ^{1}H-^{15}N (^{13}C) NOESY-HSQC and ^{1}H-^{15}N HMQC-NOESY-HMQC spectra (Ikura *et al.*, 1990). Figures 1 shows a typical ^{1}H(F1)-^{1}HN(F3) plane at ^{15}N=118.3 ppm of the 3D NOESY-HSQC spectrum of IRF-2(113) with mixing time 100 ms. To introduce the ^{15}N dimension, the signal degeneracy was solved and medium and long range NOEs could be identified. As shown in Figure 1, the sequential and medium range NOEs between the aliphatic protons and amide protons of Asn 17, Asp 51 and Cys 83 were identified. The medium range NOEs between the aliphatic protons of Asn 80, Gln 15 and Glu 14, and the amide protons of Asn and Cys 83 were essential to determine the α-helix portion (Figure 4). On the other hand, Asp 51 did not have these NOE. This indicates the possibility that Asp 51 is not in α-helix region (Figure 4).

For the degenerate HN(i)-HN(i+1, 2) NOEs, we employed the H-^{15}N HMQC-NOESY-HMQC experiment. This experiment is particularly useful to obtain the NOE connectivity between the amide protons with degenerate or nearly degenerate chemical shifts. Figure 2 shows the slice of the ^{1}H-^{15}N HMQC-NOESY-HMQC spectrum at ^{15}N=114.43, 125.71, 112.08, 125.24, and 121.48 ppm. In this figure, sequential connectivities of HN(i)-HN(i+1) NOEs from Ser 87 to Leu 88 (left panel) and from Ala 41 to Arg 43 (right panel) are shown. These connectivities could not be observed in the ^{1}H-^{15}N NOESY-HSQC spectrum for the amide protons (see Chapter 2, Figure 3). These types of NOE connectivities, which indicate the existence of α-helices, were observed for other amino acid sequences extending from Arg 7 to Ile 21 and from Lys 75 to Leu 88 (Figure 4). Although, in the region from Phe55 to Lys64, the pattern of the NOE connectivities for C_{α}H(i)-HN(i+3) and C_{α}H(i)-C_{β}H(i+3) is not clear compared to those of the other two regions (from Arg7 to Ile21 and from Lys75 to Leu88). The author defined these three helix parts as α1 (Arg 7 - Ile 21), helix-like-α2 (Phe 55 - Lys 64), and α3 (Lys 75 - Leu 88).

As mentioned above, another important NOE pattern is what characterizes a "β-sheet". It involves a stretch of strong sequential C_{a}H(i)-HN(i+1) NOEs in combination with interstrand NOEs for HN and C_{α}H protons. Strong sequential C_{a}H(i)-HN(i+1) NOEs and weak C_{α} H(i)-HN(i) NOEs with interstrand NOEs were observed for the residues from Leu 24 to Leu 27, from Lys 32 to Pro 37, from Asp 90 to Lys 95, and from Phe 106 to Leu 112. They were referred to β1, β2, β3, and β4, respectively. Especially, the pattern of the interstrand C_{α}H(i)-C_{α}H(j), C_{α}H(i)-HN(j), and HN(i)-HN(j) NOE connectivities shows a typical feature for an antiparallel β-sheet (Wüthrich, 1986). Figure 3 shows the backbone NOE connectivities which indicate the exsistence of an antiparallel β-sheet in IRF-2(113).

Summary of the sequential and medium range NOEs involving the HN, C_{α}H, and C_{β}H protons is given in Figure 4.

Analysis of $^{3}J_{HNa}$ coupling constants obtained from HMQC-J spectra

The three bond $^{3}J_{HNa}$ coupling constants are related to the ϕ backbone torsion angles by the Karplus type relationship. Especially, the secondary structure is related to specific coupling constant values (Pardi *et al.*, 1984). Therefore, the coupling constant values provide a useful additional information for the assessment of the secondary structure elements. For ideal secondary structure elements, the backbone torsion angle ϕ of α-helix is -57° which corresponds to $^{3}J_{HNa}$=3.9 Hz, and the ϕ of antiparallel β-sheet is -139° which corresponds to $^{3}J_{HNa}$=8.9 Hz. In many cases of small proteins, these coupling constants are generally estimated from the antiphase components of HN-C_{α}H cross-peaks in 2D ^{1}H-^{1}H COSY spectrum. However, in the case of IRF-2(113) protein, this method could not be applied for the reasons of resolution, signal degeneracy, and ^{1}H linewidth. As discussed above, IRF-2(113)

is a relatively large protein for NMR, and the total number of HN-C_αH cross-peaks are more than one hundred. These HN-C_αH cross-peaks are concentrated in the specific region called a finger-print region. The antiphase components of HN-C_αH cross-peaks cancele each other in the crowded region. Second, ^{1}H linewidth of IRF-2(113) is relatively broad compared with those of a smaller protein (under 60-70 amino acids residues). This also causes cancelation between antiphase components each other. Therefore, the measured coupling constants no longer reflect true coupling constants. For these reasons, we employed the HMQC-J experiment (Kay *et al.*, 1990) to measure the coupling constants from the spliting in the ^{15}N (F1) dimension. Figure 5 shows a part of the HMQC-J spectrum. For example, HN-C_αH cross-peaks of Ile 69, Ile 99, Leu 24, and Asp 96 show a large splitting in the ^{15}N (F1) dimension (11.4Hz, 8.5Hz, 7.4Hz, and 6.7Hz, respectively), the HN-C_αH cross-peaks of Leu 88, Ala 52, Arg 56, Ala 79, and Trp 11 showed a small splitting in ^{15}N (F1) dimension (< 6Hz). The result of HMQC-J experiment is summarized in Figure 4. These results are in good agreement with α-helices positions derived from NOE connectivities as discussed above. In contrast to the portions of α-helices, the $^3J_{HNa}$ coupling constant values in the β-sheet region, are relatively small compared with ideal $^3J_{HNa}$ coupling constants of an antiparallel β-sheet. The $^3J_{HNa}$ coupling constant values are dispersed in a relatively wide range for many proteins. This is due to the extended structure which differs from a typical β-strand. Thus, the observed $^3J_{HNa}$ coupling constant values are not used for structure assessment of the β-sheet.

Determination of the slowly exchanging amide protons

In order to determine the exchange rates of the amide protons of individual residues, we measured ^{1}H-^{15}N HSQC spectra of the ^{15}N-uniformly labeled IRF-2(113) protein in D_2O solution at pH 5.5 and at 25°C. The labeled IRF-2(113) was lyophilized from H_2O and dissolved in D_2O. As immediately as possible, ^{1}H-^{15}N HSQC spectra were recorded at 10 min, 60 min, and 157 min after dissolution. The time taken for one experiment was 23 min. At 10 min after dissolution, 81 of the total 112 backbone amide proton signals disappeared by the exchange with D_2O, but residual 31 signals could be observed. In the spectrum taken at 60 min after dissolution, only two more signals vanished, and sixteen amide proton peaks were left in the spectrum measured at 157 min after dissolution. These residual sixteen peaks were still recognized even at 5 hr after dissolution. The results are summarized in Figure 4. This experiment clearly indicated that the overall exchange rates of the amide protons of IRF-2(113) are rather rapid, because most of the amide protons signals (approximately 70%) could not be observed only in 10 min after dissolution. However, some amide protons have relatively slow exchange rates. As will be discussed below, these slow exchange rates are, in many cases, explained by the involvement to hydrogen bonds. (For a α-helix, a hydrogen bond can be formed between the amide proton of HN(i) and the carbonyl oxgen of CO(i+4) and for β-sheet, a hydrogen bond can be

formed between an amide proton of HN(i) and the carbonyl oxgen of CO(j) in a different strand.) Therefore, IRF-2(113) has tightly packed secondary structures or hydrophobic cores.

The rapid exchange rates indicate that the most part of the IRF-2(113) protein is flexible and exposed to solvent. As mentioned above, some amide protons with medium or slowly exchange rates were identified. The residues which have the amide protons with medium exchange rates correspond to those from Arg 7 to Asn 17 in α1 helix, and the residues which have the amide protons with slow exchange rates, correspond to those Lys 25, Leu 27, Lys32 - Ile36, Glu 92, Val 94, and Val 108 - Leu 112 in the β-sheet. Therefore, the present results support the existence of α1 helix and an antiparallel β-sheet consisting of β1, β2, β3, and β4 strands. However, some parts of the secondary structures, helix-like α2 and α3, are not consistent with the data of exchange rates. In these parts, the cross-peaks immediately vanished within 10 min after dissolution. Therefore, in contrast to α1 helix which may be tightly packed structure, helix-like α2 and α3 helix are more flexible and easily accessible to water.

Summary of the NMR parameters which define the secondary structures

In the previous sections, we demonstrated the parameters which define the secondary structures obtained from NMR measurements. Systematic NOE connectivities were observed and indicated the existence of the secondary structure elements (Figure 4). Other parameters such as ^{13}C chemical shift deviations, $^{3}J_{HN\alpha}$ coupling constants, amide proton exchange rates, participate in the identification of these secondary structures. As Spera and Bax (1991) speculated, the ^{13}C chemical shifts are closely related to the secondary structure. As shown in Chapter 2, the ^{13}C chemical shifts indicated the existence of β-strands and α-helices in IRF-2(113). Figure 11 in Chapter 2 indicates a perfect correlation with the results from NOEs and gives a more reliable definition of the secondary structure elements.

Character of the secondary structures

a) Arrangement of the secondary structure elements

Figure 6 shows the location of the secondary structure elements in IRF-2(113). From the N terminal, the arrangement of the secondary structures are α-helix (α1), turn, β-strand (β1), turn, β-strand (β2), long extended loop or strand, helix-like structure (α2), turn, α-helix (α3), β-strand (β3), loop, and β-strand (β4). These four β-strands constitute a single β-sheet.

b) Helices

Figure 7 shows the helical wheel models of the three helices (α1, α2, and α3) of IRF-2(113). Shaded areas indicate the basic amino acid region and dotted areas show the hydrophobic amino acids region. For α1-helix, the hydrophobic region locates at the one side formed by Met 8, Leu 12, Ile 16, and Asn 19 residues. In α2-helix, the region is formed by Phe 55, Trp 58, and Thr 62 residues. For α3-helix, the hydrophobic region is formed by Trp 77, Phe 81, Ala 84, and Leu 88 residues. As will be discussed in later chapter, α1 and α3-helices are arranged in antiparallel direction and these hydrophobic regions come to the same side to each other. Therefore, we could easy to deduce that these hydrophobic amino acid regions form a hydrophobic core. The hydrophilic basic amino acid core is only observed in α3-helix and locates at the opposite side of the hydrophobic region. According to the concept of the recognition helix, this α3-helix seems to be the helix which may interact with DNA. DNA-protein interaction will be discussed later in detail. Other different characters are noticed for these three helices. The amide protons of each helix have different exchanging rates. The amide protons of α1-helix revealed the slowest exchange rates (the amide proton signal still remained 10 min after dissolution in D_2O), while those of α-2, and α3-helices showed much faster exchage rates (the amide proton signal had already vanished by 10 min after dissolution in D_2O) (Figure 4). α1-helix is most tightly packed, while α2 and α3-helices are more loose. Also, as discussed later, some of these amide protons disappeared in the DNA-titration experiments. It is difficult, however, to correlate the felixibility with DNA-recognition at present.

c) β-sheet

Figure 3 shows short and long range NOE networks among β1, β2, β3 and β4-strands. In Figure 8, a schematic diagram of the amino acid arrangement in the β-sheet. Characterstic features of this β-sheet are

1) IRF-2(113) reveals a unique arrangement of β-strands. The arrangement of β-strands is β1-β2-β4-β3 and these β-strands form an antiparallel β-sheet. After all, β4-strand turns back and goes into the space between β2- and β3-strands. This kind of β-strand arrangement is the first example found in the DNA-binding domains. There are some DNA-binding proteins that contain β-sheets (see a review by Harrison (1991)), but none of them contains such type of β-strand arrangement like this. In many cases, β-strand arrangements are β1-β2-β3..... from the N-terminal in whether antiparallel- or parallel-sheet, or four antiparallel β strands are frequently arranged in a pattern so called a Greek key motif like β4-β1-β2-β3 or β1-β4-β3-β2 (Branden and Tooze,1991).

2) The cluster with strong aromaticity exsists in the β-sheet. The side-chains of the amino acids in a β-sheet come alternately to the opposite sides of the β-sheet. The aromatic residues concentrate at

the center of the one side of the β-sheet and form a strong hydrophobic core. As mentioned in Chapter 1, the author tried to construct the expression vecter of the truncated IRF-2(108) protein (numbering is acoording to IRF-2(113)), but failed to do so in *E. coli*. We explained this reason that the deleted five residues in the C-terminal portion of IRF-2(113) were essential for the domain formation. This explanation has been proved true. The deleted five residues correspond to those of β4-strand including Tyr 109 which locates at the center of the hydrophobic core. The core including three aromatic residues, Trp 26, Phe 34 and Tyr 109, is critical to the stabilization of the domain structure.

d) Turns: role of proline residues

One of the unique features of the IRF-2(113) protein is that it contains many proline residues(total ten prolines). Proline residues often disrupt regular secondary structures and are observed at the edges of the secondary structures to form turn structures. In fact, nine of the ten proline residues in IRF-2(113) are distributed in the unordered or loop regions. The contribution of Pro10 shows a unique feature. This Pro 10 residue is included in α1 helix and does not disturb the helical structure (detailed discussion will be done in the next chapter). For the other proline residues, if they are involved in turn structure, it could be identified from NOE between d(i,i+2) and the backbone torsion angles ϕ and ψ. However, at the present level of analyses, it is difficult to mention about it. In addition, proline residues cause cis-trans isomelization frequently. This isomelization occurs in the X-Pro sequence in many cases. This isomelization gives multi-comformations to the protein. In the NMR measurement, the proline cis-trans isomelization could be detected by NOESY spectrum. In the trans comfomation, NOEs are observed between XCα–PCδ protons, while in the cis comformation NOEs appear between XCα-PCα protons. In the IRF-2(113) protein, NOEs between XCα–PCδ protons were observed throughout the sequence. Therefore, all proline residues in IRF-2(113) take all trans comformation. Cis-trans isomelization does not occur at any proline residue in IRF-2(113).

Folding topology of IRF-2(113)

Figure 9 shows a schematic diagram of the arrangement of the secondary structure element of IRF-2(113) protein. For the determination of the folding topology of the three helices and four β-strands, the author observed long range NOEs among the secondary structure elements. For the interaction between helix α1 and the β-sheet, NOEs were observed for the pairs of residues, Ile16-Trp 26, Ile 16-Lys 25, and Ile 16-Phe 34. For helix α3 and the β-sheet, NOEs were seen between Phe 81-Tyr 109 and for the interaction between helix-like α2 and helix α1 or α3, NOEs appeared for the pair of Leu 54-Phe 81, and His 61-Trp 11. Thus the three helices are located on the one side of the β-sheet which forms the hydrophobic core. Detailed discussion on the global folding will be done in the next chapter.

Comparison of the secondary structure contents determined by NMR and UV-CD measurements

As shown in Chapter-1, UV-CD measurement showed the coexistence of α-helix and β-sheet (Chapter 1, Figure 8). The estimated contents of the secondary structure were approximately 30% for α-helix and 50% for β-sheet. However, NMR experiments indicated that the secondary structure contents were 35% for α-helix and 20% for β-sheet. For the α-helix contents measured by UV-CD and NMR data are in good agreement with each other. For the β-sheet portion, however, their discrepancy is remarkable.

Our method for analyzing CD spectrum of IRF-2(113) was that by Chang *et al.* (1978). In this method the CD spectrum of a protein was analysed by a linear combination of the reference CD spectra of poly(L-lysine) in α-helical, β-sheet, and random coil forms (see review Johnson, 1988). Several reasons can be thought for deviation of the estimated contents by this method from the real content. The typical standard [θ] values for α-helix are $[\theta]_{222nm}$=-4000, $[\theta]_{209nm}$=-4000, and $[\theta]_{190nm}$=+8000, for β-sheet, $[\theta]_{217nm}$=-2000, $[\theta]_{196nm}$=+5000, and for random coil, $[\theta]_{200nm}$=-4000, respectively (Hamaguchi and Takesada, 1971). Some unconsidered factors that disturb the typical [θ] value exists are as follows.

1) The contribution to the CD intensity at 220 nm by the side chains of aromatic residues may give an ambiguity to the estimation of the secondary structure. IRF-2(113) contains 6 Trp, 4 Phe, 4 His, and 1 Tyr residues which occupy 13% of the total residues.The aromatic rings give the positive circular dichroism in the 220 nm region (Hamaguci and Takesada, 1971). Therefore, these positive value causes the decrease of negative contribution around the 220 nm range. As a consequence, the content of the β-sheet was overestimated.

2) Contribution of the turn structure should be considered. It is not clear what kind of CD spectra should be used for various types of β-turns. Generally, turn structure are included in a disorder or random-coil structures. Therefore, when β-turns become predominant, the accurate estimation of the secondary structure becomes difficult. IRF-2(113) contains 10 Pro residues and is rich of β-turn type structures.

3) Unordered structures in IRF-2(113) may give an ambiguity to the estimation. As shown in Figures 1 and 6, there are many unknown structures (whether extended or loop structure) in IRF-2(113). Approximately 50% of the total residues are not classified into α-helix and β-sheet. It is hard to classify the portion to of unordred structure which may show a different type of CD spectrum pattern.

References

Bax,A., Ikura,M., Kay,L.E., Torchia,D.A. and Tschudin,R (1990) *J.Magn.Res.* **36**, 304-318.

Bodenhausen,G., and Ruben,D.J. (1980) *Chem.phys.Lett.* **69**, 185-189.

Branden,C. and Tooze,J. (1991) *Introduction to Protein Structure,* Garland Publishing, Inc.

Chang,C.T., Wu,C.S.C. and Yang,J.T. (1978) *Anal. Biochem*., **91**, 13-31

Clore,G.M. and Gronenborn,A.M. (1987) *Protein Eng.* **1**, 275-288

Hamaguchi,K. and Takesada,H. (1971) 蛋白質の旋光性、学会出版センター

Harrison,S.C (1991) *Nature*, **353**, 715-719

Ikura,M., Bax,A., Clore,G.M. and Gronenborn,A.M. (1990) *J.Am.Chem.Soc.* **112**, 9020- 9022.

Johnson,W.C. (1988) *Ann.Rev.Biophys.Biophys.Chem.* **17**, 145-166

Kanei-Ishii,C., Sarai,A., Sawazaki,T., Nakagoshi,H., He,D.-N., Ogata,K., Nishimura,Y. and Ishii,S. (1990) *J.Biol.Chem.* **265**, 19990-19995.

Kay.L.E. and Bax,A. (1990) *J.Magn.Res.* **84**, 72-84.

Marion,D., Ikura,M. and Bax,A. (1989) *J.Magn.Res.* **84**, 425-430.

Matso,H., Shirakawa,M., Okubo,T., Yamazaki,T. and Kyogoku,Y. (1991) *J.Biomol.NMR,* **1**, 191-204.

Norwood,T.J., Boyd,J., Heritage,J.E., Sogge,N. and Campbell,I.D. (1990) *J.Magn.Res.* **87**, 488-501.

Oneil,K.T., Shuman,J.D., Ampe,C. and Degrado,W.F. (1991) *Biochemistry* **30**, 9030-9034.

Pabo,C.O. and Saure,R.T. (1992) *Ann.Rev.Biochem.* **61**, 1053-1095.

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

Patel,L., Abate,C. and Curran,T. (1990) *Nature* **347**, 572-575.

Saudek,V., Pasley,H.S., Gibson,T., Gausepohl,H., Frank,R. and Pastore,A. (1991) *Biochemistry* **30**, 1310-1317.

Shirakawa,M., Fairbrother,W.J., Serikawa,Y., Okubo,T., Kyogoku,Y. and Wright, P.E. (1993) *Biochemistry,* **32**, 2144-2153.

Spera,S. and Bax,A. (1991) *J.Am.Chem.Soc*. **113**. 5490-5492.

States,D.J., Haberkorn,R.A. and Rubren,D.J. (1982) *J.Magn.Res.* **48**, 286-292.

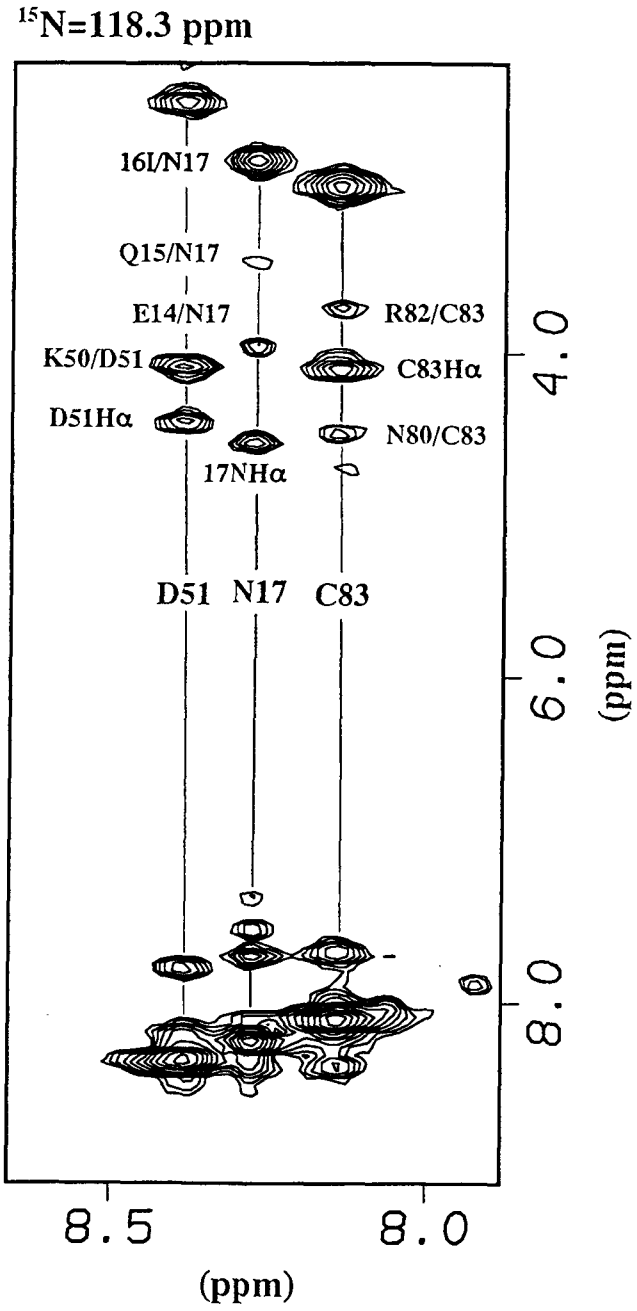
Veals,S.A., Schindler,C., Leonard,D., Fu,X-Y., Aebersold,R., Darnell,J.E.jr. and Levy,D.E. (1992) *Mol.Cell.Biol.* **12**, 3315-3324.

Weiss,M.A., Ellenberger,T., Wobbe,C.R., Lee,J.P., Harrison,S.C and Struhl,K. (1990) *Nature* **347**, 575-578.

Wolberger,C. (1993) *Curr.opin.Struct.Biol*. **3**, 3-10.

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

Figure 1


^{15}N=118.3 ppm
16I/N17
Q15/N17
E14/N17
K50/D51
D51Hα
17NHα
R82/C83
C83Hα
N80/C83
D51
N17
C83
4.0
6.0
8.0
(ppm)
8.5
8.0
(ppm)

^{1}H-^{15}N HMQC-NOESY-HMQC

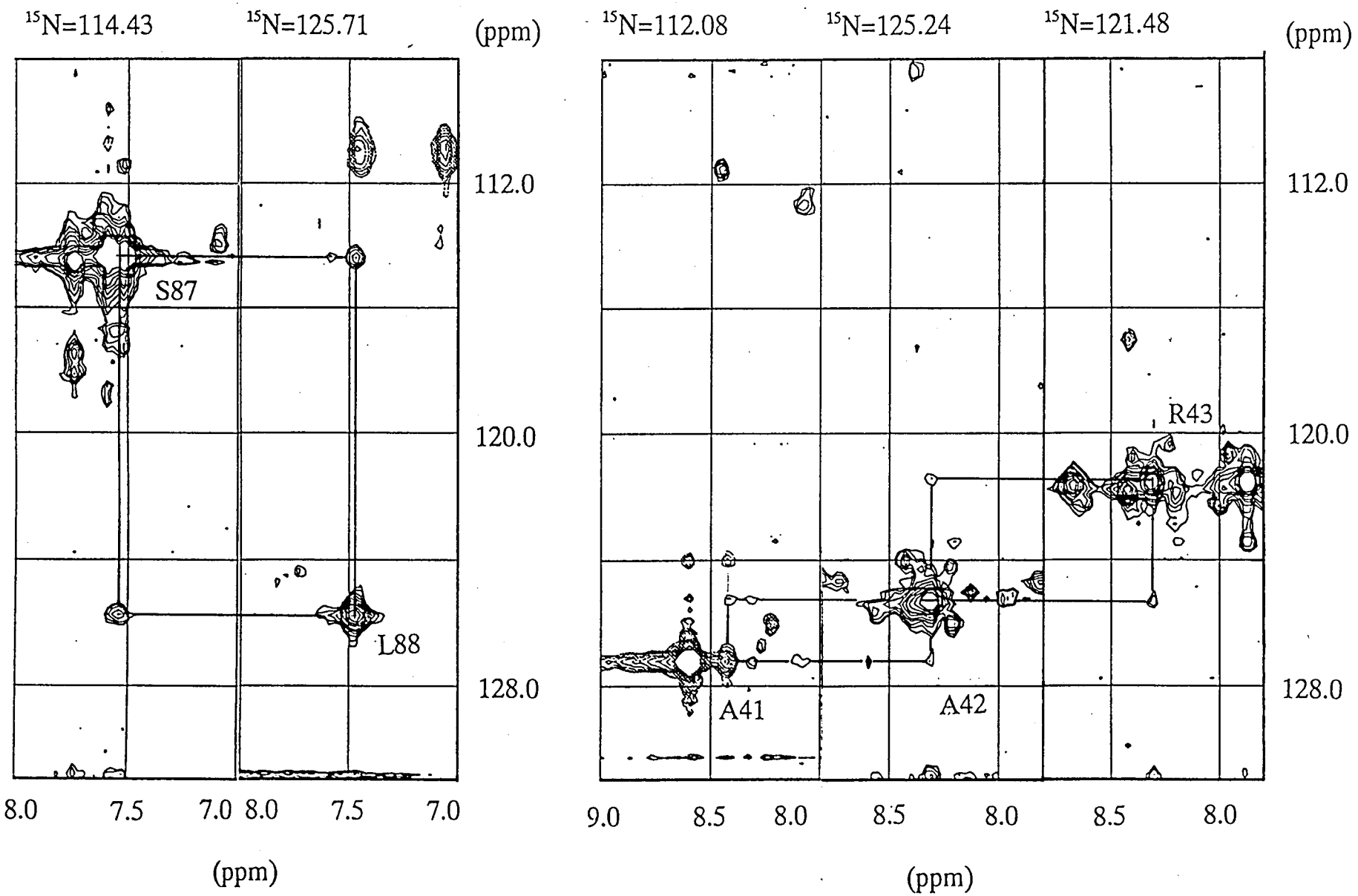
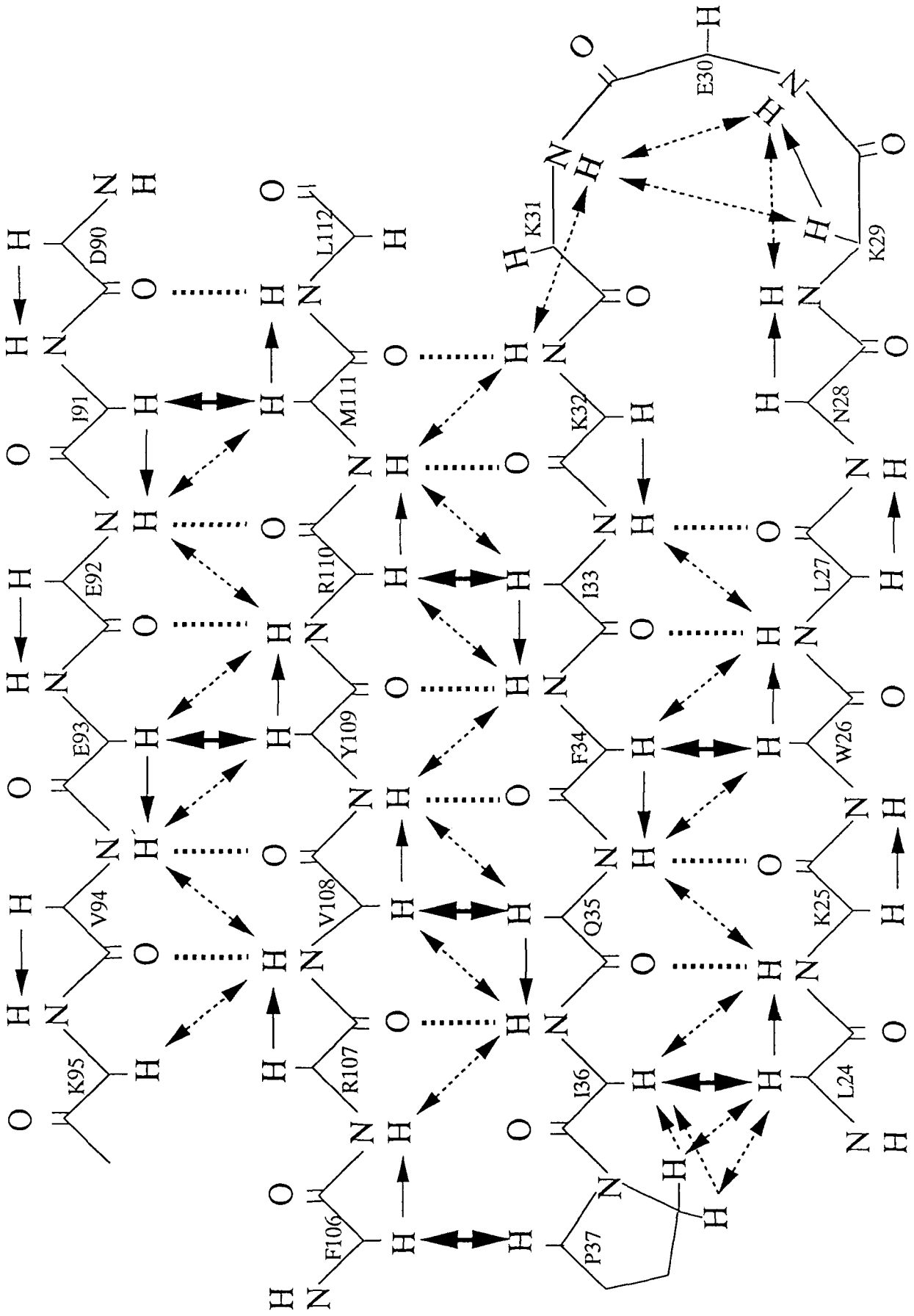


Figure 2

Figure 3


K95
V94
E93
E92
D90
I91
R107
V108
Y109
R110
M111
L112
F106
P37
I36
Q35
F34
I33
K32
K31
E30
K29
N28
L27
W26
K25
L24

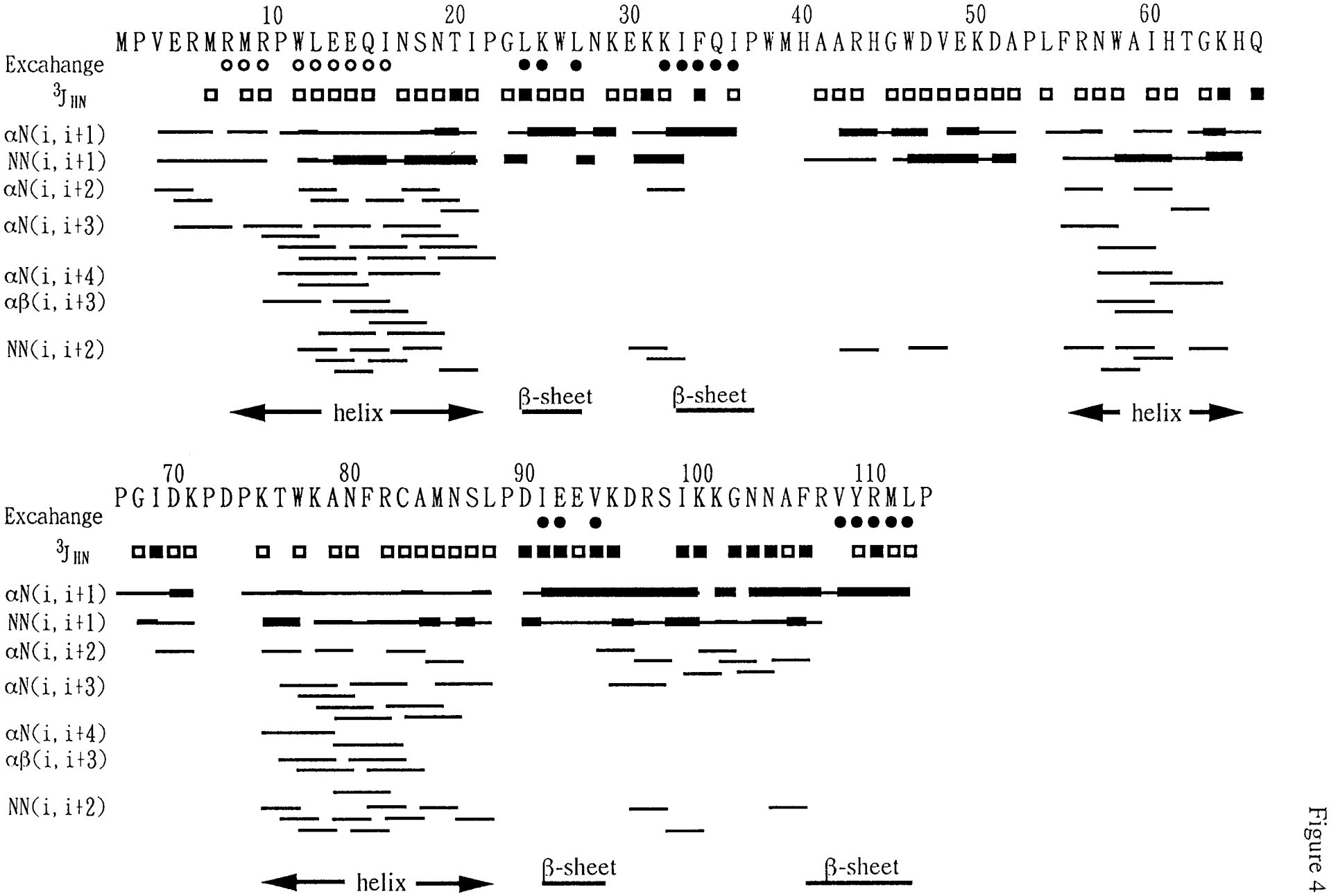
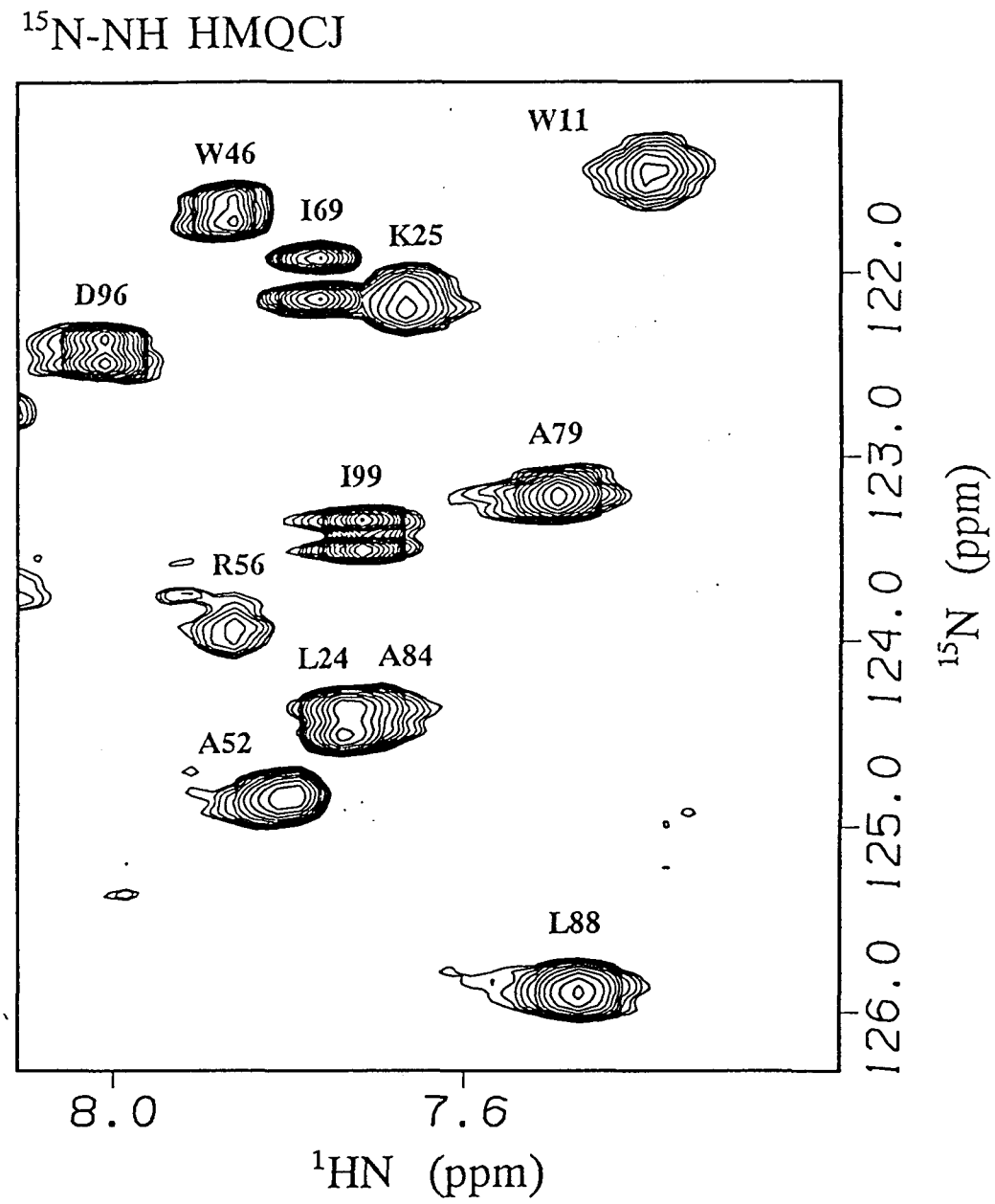
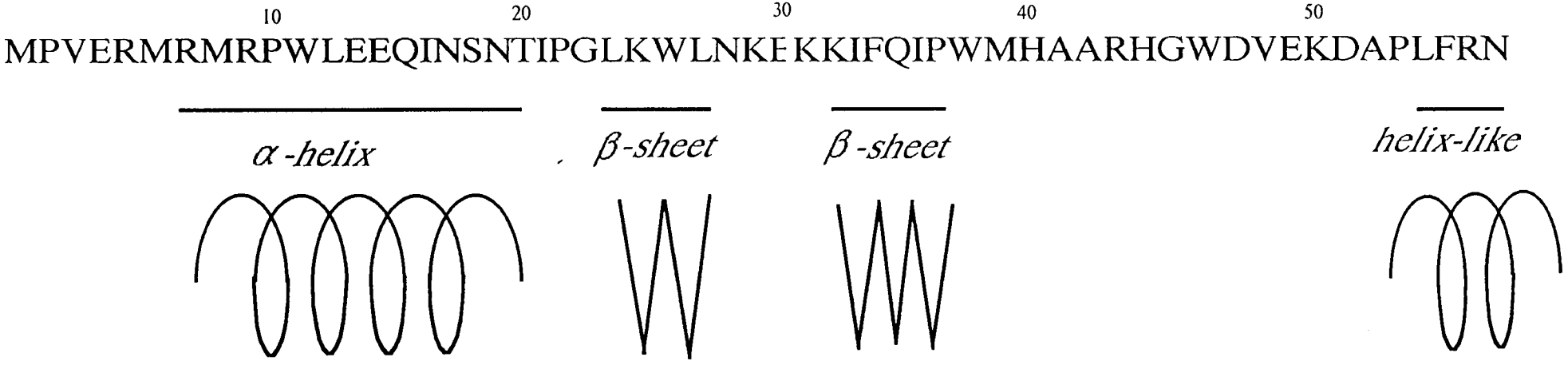

10
20
30
40
50
60
MPVERMRMRPWLEEQINSNTIPGLKWLNKEKKIFQIPWMHAARHGWDVEKDAPLFRNWAIHTGKHQ
Excahange
$^3J_{HN}$
αN(i, i+1)
NN(i, i+1)
αN(i, i+2)
αN(i, i+3)
αN(i, i+4)
αβ(i, i+3)
NN(i, i+2)
helix
β-sheet
β-sheet
helix
70
80
90
100
110
PGIDKPDPKTWKANFRCAMNSLPDIEEVKDRSIKKGNNAFRVYRMLP
Excahange
$^3J_{HN}$
αN(i, i+1)
NN(i, i+1)
αN(i, i+2)
αN(i, i+3)
αN(i, i+4)
αβ(i, i+3)
NN(i, i+2)
helix
β-sheet
β-sheet

Figure 4

Figure 5


15N-NH HMQCJ
W11
W46
I69
K25
D96
A79
I99
R56
L24
A84
A52
L88
122.0
123.0
124.0
125.0
126.0
15N (ppm)
8.0
7.6
1HN (ppm)


10
20
30
40
50
MPVERMRMRPWLEEQINSNTIPGLKWLNKEKKIFQIPWMHAARHGWDVEKDAPLFRN
α-helix
β-sheet
β-sheet
helix-like

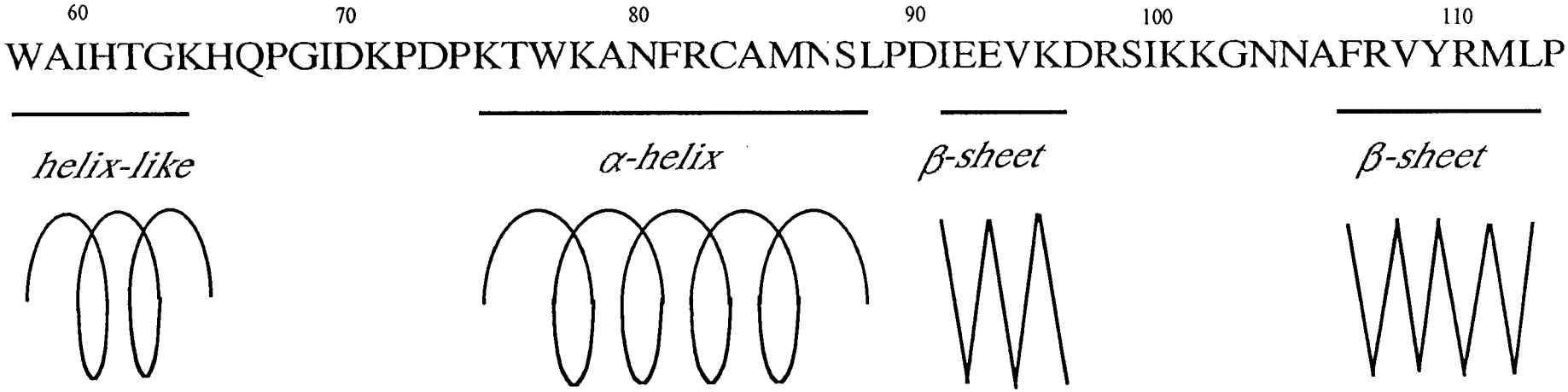

60
70
80
90
100
110
WAIHTGKHQPGIDKPDPKTWKANFRCAMNSLPDIEEVKDRSIKKGNNAFRVYRMLP
helix-like
α-helix
β-sheet
β-sheet

Figure 6

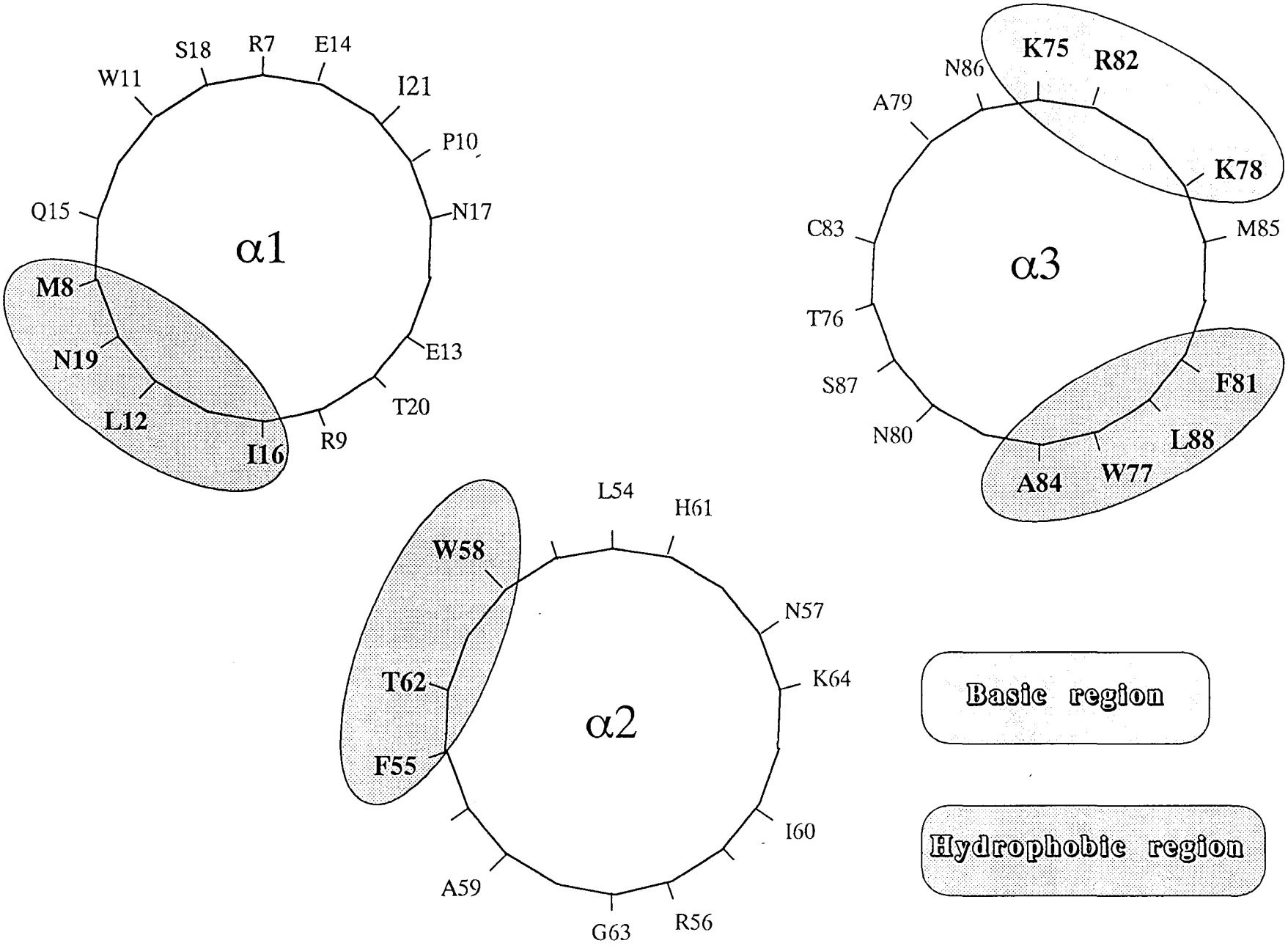

α1
S18
R7
E14
W11
I21
P10
Q15
N17
M8
E13
N19
T20
L12
R9
I16
α3
K75
R82
N86
A79
K78
C83
M85
T76
F81
S87
L88
N80
A84
W77
α2
L54
H61
W58
N57
T62
K64
F55
I60
A59
G63
R56
Basic region
Hydrophobic region

Figure 7

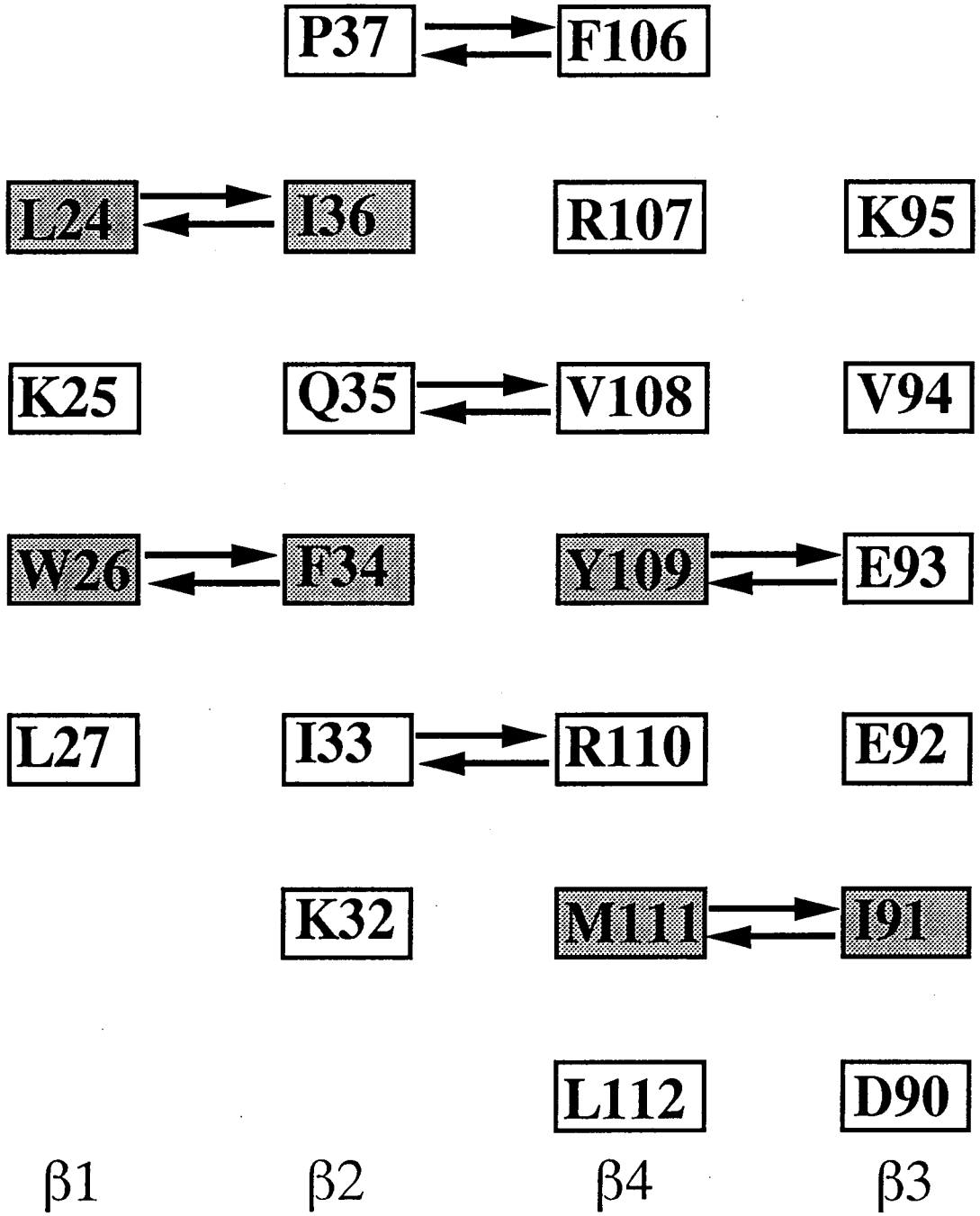

P37
F106
L24
I36
R107
K95
K25
Q35
V108
V94
W26
F34
Y109
E93
L27
I33
R110
E92
K32
M111
I91
L112
D90
β1
β2
β4
β3

Figure 8

Figure 9

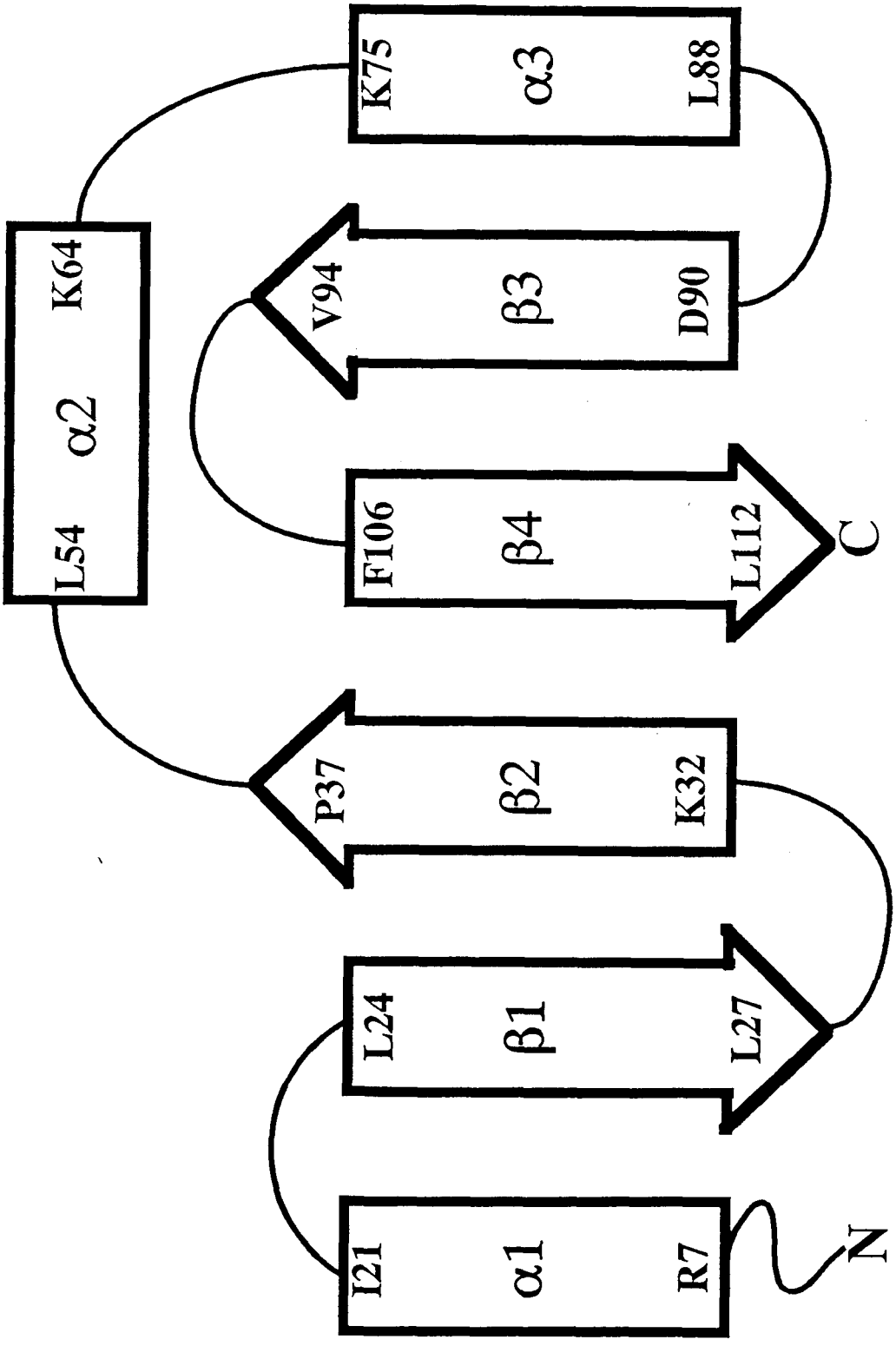

I21
α1
R7
N
L24
β1
L27
P37
β2
K32
L54
α2
K64
F106
β4
L112
C
V94
β3
D90
K75
α3
L88

Figure legends

Figure 1.

Typical $^{1}H(F1)$-$^{1}HN(F3)$ plane at ^{15}N=118.3 ppm of the 3D NOESY-HSQC spectrum of IRF-2(113) taken at 500 MHz with mixing time 100 ms.

Figure 2.

Composite spectrum consisting of strips taken from ^{15}N planes of the 3D ^{1}H-^{15}N HMQC-NOESY-HMQC spectrum of IRF-2(113) with 100 ms mixing time. Left spectrum shows the sequential connectivity between Ser 87 and Leu 88. Right spectrum shows the sequential connectivity through Ala41-Ala42-Arg43.

Figure 3.

Short- and long-range backbone NOE pattern indicating the presence of anti-parallel β-sheet involving the residues from Leu 24 to Leu 27, from Lys 32 to Pro 37, from Asp 90 to Val 94, and from Fhe 106 to Leu 112. Thick and dotted double - end arrows indicate interstrand NOE connectivities. Thin end arrows indicate the sequential NOE connectivities. Thick dotted lines show the probable position of hydrogen bonds estimated from the amide proton exchange rates.

Figure 4.

Summary of the sequential and medium range NOEs between the HN, CαH, CβH protons together with the amide proton exchange rates and $^{3}J_{HN\alpha}$ coupling constant data. The thickness of the lines reflects the intensity of NOEs. The residues whose amide protons still were present 10 min after dissolving in D_2O, but exchanged by 60 min are indicated by open circles. The residues whose amide protons were remain at 60 min after dissolving in D_2O are shown by closed circles. The closed and open squares indicate $^{3}J_{HN\alpha}$ coupling constants value more than 7 Hz and less than 6 Hz, respectively.

Figure 5.

A section of the 2D HMQC-J spectrum of IRF-2(113). The $^{3}J_{HN\alpha}$ coupling constants were read from splitting of the cross peaks in $^{15}N(F1)$ dimension.

Figure 6.

The secondary structure element of the IRF-2(113) protein. The positions of α-helices and β-strands are indicated.

Figure 7.

Helical wheel models of three helices of IRF-2(113). The shaded area indicates the basic region, and the dotted areas mean the hydrophobic regions. Helices are viewed from their C-terminals

Figure 8.

The arrangement of the amino acid residues in the β-sheet region of IRF-2(113). Each amino acid is given by one letter code. Shaded boxes indicate the amino acids which form the hydrohobic core.

Figure 9.

A shematic diagram of the arrangement of the secondary structure element of the IRF-2(113). α - Helices are shown as rectangles and β - strands as arrows.

Chapter-4

Global Folding of IRF-2(113) and Interaction with DNA

Abstract

The global folding structure of the IRF-2(113) has been calculated by the simulated annealing and energy minimization methods based on the NMR data. The input constraints for the structure calculation contained ^{1}H-^{1}H distance constraints from NOE intensities, dihedral angle constraints from coupling constants and hydrogen bond constraints. The result of the calculation shows that the IRF-2(113) is composed of two α - helices, one helix-like structure and four stranded antiparallel β-sheet. The α1, helix-like α2, and α3 helices include residues 7 to 21, 54 to 64, and 75 to 89, respectively, and the antiparallel β-sheet consists of β1, β2, β3 and β4 strands including residues 24 to 27, 32 to 37, 90 to 94 and 106 to 112, respectively. These calculated secondary structure elements are in good agreement with those predicted from NOE connectivities (see Chapter 3). The structure calculation showed that the IRF-2(113) protein was stabilized by a hydrophobic core including three aromatic residues at the center of the β - sheet and the hydrophobic residues in the three helices. The global folding of IRF-2(113) protein revealed no apparent structure similarity to the knwon DNA- binding domain. α2 and α3 helices arrange like those helix-turn-helix (HTH) motif, but the turn part is fairly longer and the relative orientation of the two helices is different from that of HTH. From the chemical shift perturbation experiments of the ^{15}N labeled protein upon complexation with the binding site DNA, it is inferred that α3 helix of IRF-2(113) directly interacts with DNA .

Introduction

The global folding of the DNA binding domain of IRF-2 was studied by means of nuclear magnetic resonance (NMR). In the preceding Chapter 2 and 3, the author presented the ^{1}H,^{15}N, ^{13}C, and ^{13}CO assignments of IRF-2(113) and determined the secondary structure of the IRF-2(113) protein. These processes confered the foundation of structure calculation. In this chapter, the author presents the result of structure calculation of IRF-2(113) protein derived from the analyses of 3D heteronuclear-edited NOE spectra, ^{13}C chemical shifts, the amide proton exchange rates and $^{3}J_{HN\alpha}$ coupling constant data. The input constraints for the structure calculation were the ^{1}H-^{1}H distance constraints from NOE intensities, the dihedral angle constraints and the hydrogen bond constraints. At the structure calculation, the distance geometry calculation with a program DIANA developed by Güntert *et al.* (1991) was not successful for IRF-2(113), because of a large number of residues and the complexity of the folding topology. Therefore, another calculation method called a simulated annealing method combined with energy minimization (Nilges *et al.*, 1988) was employed with program X-PLOR V3.1 (Brünger, 1993). By using this program, the global folding of IRF-2(113) could be obtained.

The interaction with DNA was investigated from the chemical shift perturbation experiments of the ^{15}N labeled protein upon complexation with the binding site DNA. It is inferred that the third helix of the DNA binding domain of IRF-2 interacts with DNA directly. Based on the calculated structure, the author tried to propose the binding mode of IRF-2 protein to the target DNA sequence.

Materials and Methods

Distance constraints

Distance constraints for the structure calculation were obtained from homonuclear 2D NOESY spectra in D_2O with mixing time of 100ms. These spectra were recorded with the sample which has deuterated Trp and Phe aromatic ring protons. 3D ^1H-^{15}N NOESY-HSQC and ^1H-^{13}C NOESY-HSQC spectra with mixing time 100 ms were also used to obtain distance constraints.

Upper bounds for NOE constraints were calibrated using aliphatic ^1H-^1H distance or sequential ^1H-^1HN distance in the β-sheet. Classification into three classes was performed according to Wüthrich (1986). For the sequential NOEs, NOE with strong intensity is set to 2.5Å, medium intensity one to 3.0Å, and weak intensity NOE to 4.0Å. For the medium and long range NOE, NOEs between backbone protons are uniformly set to 4.0Å. For the medium and long range NOE including side chain, all NOEs are uniformly set to 5.0 Å. Appropriate pseudoatom corrections were added for the prochiral methylene, methyl, and Leu dimethyl groups, the Tyr and Phe aromatic ring protons, according to Wüthrich (1986). Lower distance limits were defined as the sums of the van der Waals radii.

The positions of hydrogen bonds within the secondary structure elements that were identified from NOE connectivity and the amide proton exchange rates were assumed to be fixed in the range from 3.5Å to 2.5Å and were added to the distance constraints.

Torsion angle constraints

Torsion angle constraints for the structure calculation were obtained from the HMQC-J spectrum (Kay and Bax., 1990). This spectrum was already explained in Chapter 3. The backbone vicinal coupling constants ($^3J_{N\alpha}$) determined by the HMQC-J spectrum were converted to the torsion angle ϕ constraints. The $^3J_{N\alpha}$ values were classified into three groups, $^3J_{N\alpha} < 5.5$ Hz, 8.0 Hz $< {}^3J_{N\alpha} <$ 10 Hz and $^3J_{N\alpha} >$ 10 Hz, and were converted to the torsion angle ϕ ranging [-90°, -40°], [-160°, -80°], and [-140°, -100°], respectively.

Structure calculation

Three-dimensional structure of IRF-2(113) was computed from the experimental constraints with the program X-PLOR V3.1 (Brünger, 1993). Initial structures with approximate polypeptide fold are obtained using a procedure known as embeding and then these structures were used for dynamical simulated annealing calculations (Nigei, et al., 1988). These structures were refined using a simulated annealing protocal (Brünger, 1993). 5 structures were selected based on agreement with the experimental

distance data. They were the candidates for farther structural consideration. These structures calculations were carried out on Silicon Graphics Iris INDIGO workstation.

DNA synthesis and determination of DNA concentration

The synthetic 12 bp DNA sequence (dGAAAGTGAAAGT /CTTTCACTTTCAd) of the IRF-2 binding site were synthesized with an automated DNA synthesizer 380B (Applied Biosystems, USA) according to the protocol of the manufacturer.

The concentration of synthetic DNA was determined as follows. DNA solution of 0.8 - 20 OD/tube in water was lyophilized. 2 μl of Nuclease P1 (1mg/ml), 2μl of 1M anmonium acetate (pH 5.0) buffer, and 36 μl of water were added to this solution. After incubation (at 37 °C for overnight) for digestion, the sample solution was incubated at 90 °C for 3 min to inactivate the enzyme. 1 ml of water was added to this sample solution, and OD_{260} value was measured at 20 °C and calibrated to the ε of nucleotide monomers.

NMR titration of IRF-2(113) with DNA oligomers

For the experiment of the DNA - protein interaction, protein concentration was 0.7mM, and dissolved in 10mM phosphate buffer contain 200 mM KCl, 1mM EDTA, at pH 6.8. The high concentrate DNA solution (5 mM) was added to this sample to various DNA/protein ratios. Protein-DNA interaction was monitered by the change of the amide proton signals.

A series of ^{1}H-^{15}N HSQC spectra were measured at the various protein/DNA ratios. The water suppression was achieved by preiradiation during relaxation delay. The acquired data matrix consisted of 128 (t1) × 1024 (t2) complex data points. Two zero-filling was done in t1 (^{15}N).

Results and Discussion

Comparison with distance geometry and simulated annealing methods

Initially, the author attempted the structure calculation of IRF-2(113) by using a distance geometry method with a program DIANA (Güntert, *et al.*, 1991). However, a satisfactory result could not be accomplished with the program. A conceivable reason is as follows. DIANA program is based on the minimization of a variable target function. To reduce possibility being trapped in local minima, the target function is varied during the structure calculation. At the first stage, only local constraints with respect to the polypeptide sequence are considered, and in subsequent rounds of calculations, the constraints between atoms far apart with respect to the amino acid sequence are included in stepwise

fashion. Consequently, at the first stage of the structure calculation, local conformations are established, and the global folding of the peptide chain will be obtained only at the end of the calculation. Due to this character of the algorithm, the calculation of a turn-back structure like the β-strand arrangement β1-β2-β4-β3 in IRF-2(113) seems to be difficult. In fact, the portion of the interstrand β2-β4 and β4-β3 contacts always showed large violations in spite of accurate constraints were supplied. Therefore, the folded structure of IRF-2(113) was calculated using a dynamical simulated annealing method with the same interatomic distance and torsion angle constraints. The result of the structure calculation showed that IRF-2(113) contains three α-helices and four β-strands arranged in antiparallel β-sheet as predicted in the previous chapter. The global folding of IRF-2(113) is given in Figure 1a. A ribbon representation of the structure is shown in Figure 1b.

Quality of the NMR structure

The quality of calculated structure depends on the number of NOEs and torsion angle constraints used in the calculation. The present structure calculations employed a total of 601 experimental interproton distance constraints and 38 dihedral angle constraints. 260 of the total distance constraints were backbone sequential ($1 = |i-j|$) NOEs, 167 were medium rang ($1 < |i-j| \leq 4$) NOEs, and 174 were long range ($5 < |i-j|$) ones. In this calculation intraresidues NOEs were not included. In addition, 68 hydrogen-bonding constraints derived from slowly exchanging backbone amide resonances were added (Table 1A).

The precision of an NMR structure is usually expressed by the average root mean aquare deviation (RMSD) between individual structures and the mean structure. The RMSD for all the backbone heavy atoms of the calculated structures are 2.44 ± 0.22 Å at the present level. The individual α1, α3 helices and β-sheet have lower RMSD (1-1.6 Å), but some of the connecting loops and α2 helix are less defined by the present NMR data. Thus, the RMSD is 1.92 ± 0.17 Å for only the backbone atoms of the defined secondary structure elements including three α-helices and a β-sheet.

Structure of IRF-2(113)

IRF2-(113) consists of three α-helices and four β-strands arranged in a β-sheet packed to enclose an extensively hydrophobic core (Figure 2). Helix α1 contains residues from Met 7 to Ile 21, helix-like α2 is shortest and formed by residues from Leu 54 to Ile 64, and helix α3 contains residues from Lys 75 to Leu 88 (these helices were referred to α1, α2 and α3 in the previous chapter). In helices α1 and α3 have characteristic α- helical (CO_i-NH_{i+4}) hydrogen bonds are formed throughout. Helix α2 is somewhat disordered at the present level of the structure calculation. The three α helices give RMS deviations, 1.60 Å, 2.65 Å, and 1.5 Å, respectively (Table 1B). Helix α1 contains a proline

residue (Pro10) at the near center of the helix. As mentioned in previous chapter, in many cases, a proline residue causes helix to disrupt or bend by disturbing the local hydrogen bonding pattern. The proline residue in helix α1 does not cause such effect to its helical structure and the helix axis is nearly straight.

The β-sheet region contains residues from Leu 24 to Leu 27, from Lys 32 to Pro 37, from Asp 90 to Val 94, and from Phe 106 to Leu 112. RMSD is 1.38 Å for the sheet strucure (Table 1B). The β-sheet consists of four β-strands referred to β1, β2, β3 and β4 in the previous chapter. They are arranged like (β1-β2-β4-β3) forming an antiparallel β-sheet. This arrangement is consistent with that predicted from the NOE connectivities and the H-D exchange rate measurement as discussed in the previous chapter. The strands β1 and β2 are joined by a short hairpin structure that contains five residues from Asn 28 to Lys 31. The strands β3 and β4 are joined by a long poorly defined loop that has a feture of a helix like structure. The strand β4 turns back and inserts into the space between β2 and β3 strands.

The calculated global folding of IRF-2(113) at the present level, shows that the three helices form a bundle that packs onto the β-sheet and encloses a hydrophobic and highly aromatic core (Figure 2). The central core comprises residues Leu 12, Ile 16, and Gln 15 from helix α1; Leu 24, Trp 26, Ile 36, Phe 34, and Tyr 109 from β-sheet; Leu 54, Trp 58, His 61 from helix α2 and Phe 81 from α3. However, the loop regions, which consist of amino acids residues from Trp 38 to Pro 53 and from Arg 97 to Ala 105 are less difined at the present level. Thus, it is hard to mention about the folding feature to these regions. Generally, loop region is very difficult to determine the tertially structure by NMR because of less long range NOE informations by its flexibility. By this mean, at the present, the author could not be identified the longrange NOEs which play a keyrole to define the tertially structure, between these loop regions and the another portions. Therfore, this is the one of reasons that the structure determination of IRF-2(113) by NMR is very difficult from the present NMR information. To obtain the more fine structure, more detail analyses of NOESY spectra and stereospecific assignment in these regions must be done. Of cause, these analyses are necessary to the secondary structure region to obtain the more defined topology.

By now, the structures of many DNA binding proteins were solved by NMR and X-ray analyses. However, this arrangement of the β-strands observed in IRF-2(113) is the first example for the DNA binding proteins. Moreover, this feature of the β-strand arrangement is very rare in all structure known proteins (Branden and Tooze 1991).

Structure comparison with other IRF family proteins

There are some transcriptional factors (IRF-1, ISGF3γ and ICSBP) that show remarkable sequence homology to the DNA binding domain of IRF-2 (Miyamoto *et al.*, 1988; Harada *et al.*,

1989; Driggers *et al.*, 1990; Veals *et. al.*, 1992). IRF family member proteins (IRF-1,-2, ICSBP and ISGF3γ) participate in the regulation of IFN and IFN-inducible gene expression. Among these proteins, the putative DNA binding domains show 80% amino acids sequence similarities (Figure 3). As discussed above, the DNA binding domain of IRF-2(113) protein has extensive hydrophobic interaction in its core. The central core comprises residues Trp 11, Leu 12, Ile 16 and Gln 15 from helix α1; Leu 24, Trp 26, Phe 34, Ile 36, and Tyr 109 from β-sheet; Leu 54, Phe 55, Trp 58, His 61 from helix α2; Phe 81 from α3. These hydrophobic residues are conserved in IRF-1, ISGF3γ and ICSBP sequences(Figure 3). In addition to the conserved hydrophobic side chains which probably play a key role in packing and stabilization of the three dimensional structure, some residues on the protein surface are invariant in the sequences of IRF family members. Such invariant residues are found predominantly on the secondary structure elements defined by helix α1, α2, α3 and β-sheet. In addition, several hydrophilic residues are invariant in the available sequences of IRF family members (Figure 3). The localization of the invariant hydrophilic residues (Lys 75, Lys 78 and Arg 82) on the one face of helix α3 in the IRF-2(113) strongly suggests that they are essential for DNA binding activity (interactions with DNA will be discussed later). These sequence similarity strongly suggests that the IRF family transcriptional factors have identical three dimensional structures and may form a new class of DNA binding protein.

Veals *et al.* (1992) discussed that the IRFs (IRF-1, IRF-2, ICSBP and ISGF3γ) are related to the myb family of DNA binding proteins that have three α-helices arranged into helix-turn-helix (HTH) motif. Their discussion was based on the finding that the positions of three Trp residues among IRFs proteins and c-myb oncoprotein were identical with each other when IRF's and c-myb sequences were aligned. C-myb, which acts as a transcriptional regulator during hematopoietic cell differentiation, has a DNA binding domain which consists of three repeated related stretches of 51 to 52 amino acids repeated three times (reffered to R1, R2 and R3; Figure 4) and which is conserved among myb-related proteins (Gabrielsen *et al.*, 1991, Howe *et al.*, 1990, Kanei-Ishii, *et al.*, 1990). Each repeat contains three Trp residues spaced at intervals of 18 to 19 amino acids. This triple repeats of Trp residues are conserved in IRFs proteins (Figure 4). However, the similarity between IRFs and c-myb related proteins is limited to the repeats of Trp residues. None of other residues are not conserved among these proteins. In addition, structure analyses showed the striking difference between the proteins in two families. A recent NMR work on the DNA binding domain of c-myb protein showed that the side chain of the aromatic residues, Trp and His, form a hydrophobic core surrounded by three α-helices to stabilize the HTH motif (Ogata et al. 1992). This feature is quite different from the structual feature of IRF-2(113). As shown in Figure 4, the positions of α-helices of IRF-2(113) are quite different from these of the c-myb protein. Especially, the corresponding position of IRFs to the first helix of the c-myb protein is poorly defined as ordered structure. In this regard, the transcriptional factors in the IFN system belong to a different class of DNA binding protein from those of the c-myb family.

Interaction with DNA

Interaction between IRF2(113) and the synthetic 12 bp DNA dGAAAGTGAAAGA /CTTTCACTTTCTd to which IRF-2(113) specifically bind, was studied by ^{1}H-^{15}N HSQC experiment. To identify the DNA-binding region of IRF-2(113), the author employed the DNA titration experiment.The concentrated DNA solution was added to the protein solution at various DNA/protein ratios, and the ^{1}H-^{15}N HSQC spectra were taken for each solution. In comparison of the series of spectra with a reference spectrum of pure IRF-2(113), several peaks were found to change in intensity and linewidth. Figure 5 shows a typical spectrum of the DNA titration experiment. As shown here, main changes are peak broad-out phenomena insted of signal shifts. Some reasons are thought as follows.

1) Chemical exchange. The protein and DNA are in an equilibrium with the complexed state in solution like

$$[\text{protein}]+[\text{DNA}] \underset{k_2}{\overset{k_1}{\rightleftharpoons}} [\text{protein-DNA}]$$

The free protein gives a singnal at the frequency ν_f and the bound protein does at ν_b. Association rate denotes k_1 and dissociation one denotes k_2. At the equilibrium $k_1=k_2=1/2\tau$, where τ is defined as the mean life time at both free and bound states. When τ is extremely longer than $1/|\nu_f-\nu_b|$ (slow exchange), the protein gives new NMR peaks at different positions from those of a free protein. While, if τ is much shorter than $1/|\nu_A-\nu_B|$ (fast-exchange), NMR signals from each state can not be separately observed and the peaks of the protein shifts to the the center of ν_f and ν_b depending on the amount of each state. The most difficult situation for NMR measurement, however, is the middle of both cases (intermediate-exchange). At this state, resonance peaks give broad line in between ν_f and ν_b and in many cases, signals become hardly detectable. This phenomenon is called exchange broadening.

2) Catalytic effect. In the protein-DNA complex, the base nitrogen at polypeptide groups of DNA could be proton acceptors and take out amide protons from the protein backbone. Therefore, the exchange rates of the amide protons are accelated and their signals disappear from the original position by converging to the water signal.

3) Dispersion of the hydrogen bond chemical shifts. In DNA, many nuclei can form hydrogen bonds with amide protons. The chemical shifts of the hydrogen bonds at each site are different, and thus the observable signal becomes an overlap of each peaks. It causes the line broadening, but the situation becomes severe when the exchange broadening is accompanied.

In Figure 6, the residues which gave broad-out amide proton signals are indicated by circle. The residues of broad-out peaks are concentrated at α2, α3 helices and the N-terminal region near α1 helix. The peaks from three basic amino acid residues (Lys 75, Lys 78 and Arg 82) located on the

same side of α3 helix were broad-out easily. Such a cluster of the basic amino acid residues can not be observed in α1 and α2 helices (Figure 7 in Chapter 3). By analogy with the binding model of the helix-turn-helix motif, the basic residue rich α3-helix may be inserted into the major groove of DNA. The side chains of Arg and Lys are apt to form hydrogen bonds with guanine residues in double stranded DNA (Pabo and Sauer, 1992). Lys and Arg side chain also have possivilty to interact with the phosphate backbone by salt bridges.

Among the member of the IRF family, several residues that may contribute to the DNA binding judged from the chemical shift perturbation experiment are conserved. Despite the striking sequence similarity of α3 helix in IRF family, ISGF3γ does not bind to the IRF recognition sequence (Veals *et al.,* 1992). The apparent difference is the residue at position 76 in α3 helix (Figure 3). Thr 76 residues in IRF-1 and 2 are replaced by Val in ISGF3γ. The residue may contribute to the recognition specificity. Of course, this interpretation is very primitive, and more extensive experiments like site directed mutagenesis are required to solve the DNA recognition mechanism by the IRF-2 protein.

Helix-like α2 is also another candidate of the recognition helix. The amide proton signals of some residues in the helix were also broade-out on the addition of DNA. This helix may interact with DNA in either specific or non-specific way. The N terminal region which gives the broad-out signals may interact non-specifically with the phosphate backbone of DNA. In several DNA-protein complexes, the contacts between the protein and DNA take place at multiple sites, some of which contribute to the recognition of the base sequence and the other of which simply utilize electrostatic interaction to enhance the stability of the complex.

Does IRF-2(113) contains a new DNA binding Motif ?

By now the three dimensional structure of more than 20 DNA-binding proteins were solved by NMR and X-ray analyses. These results showed that most DNA binding proteins can be categolized into the families for specific DNA binding motifs. Examples of these DNA binding motifs are helix-turn-helix (HTH), several types of Zn-finger, basic-leucine zipper (bZip), basic helix-loop-helix (bHLH), β-ribbon, *etc* (for review see Harrison 1991; Pabo and Sauer 1992; Wolberger 1993). As mentioned above, Veals *et al.* (1992) speculated that IRFs could be classified into the Trp-cluster motif (a sub-family of HTH motif) family. However, the structure analyses revealed that the DNA binding domain of IRF-2(113) is not classified into any of the known family from the view point of topology. Then, is there any similarity in the way of recognition of the DNA sequence between HTH motif?. We will compare the recognition helix of IRF-2(113) with that of other motifs.

If we assume the α3-helix of IRF-2(113) as a recognition helix, the arrangement of α2 and α3-helices resembles with that of the helix-turn-helix motif in c-myb R3 (Ogata *et al.*, 1992). When we compared the locations and the lengths of the corresponding helices with c-myb repeats, there are

remarkable difference between them (Figure 4). Helix α3 in c-myb R3, the recognition helix, consists of Asp 70 and Ser 80 (numbering follows that of IRF-2). In IRF-2(113), the corresponding part (Asp 70 to Pro 74) is involved in a turn or a loop region because of the presence of a proline residue. The α2-helix of c-myb R3 is quite short (7 residues) and rigid, while that of IRF-2(113) is rather long (11 residues) and fairly flexible. Moreover helix-like α2 of IRF-2(113) may participate in the interaction with DNA. IRF-2(113) has a long loop or turn region (10 residues) between α2 and α3 helices where three proline residues locate. Lex A has a two-helix motif with a long turn (Lamerichs *et al.*, 1990), but it is a variant of HTH motif judged from the sequence homology with other regular HTH motif and its tertiary structure. Although the exact relative orientation of both α2 and α3 helices in IRF-2(113) has not yet been fixed by the calculation at the present stage, it is hard to say that IRF has a variant of HTH motif.

Three α-helices form a bundle on β-sheet in IRF-2(113). Such a overall feature of the structure a little resembles with that of papillomavirus-1 E2 DNA binding domain (Hegde et al., 1992), where two α-helices and one short β-strand sit on the antiparrallel four strands β-sheet. The topology of the arrangement of the β-strands differs from that of IRF-2(113). The first helix of E2 domain is the recognition one, but the domain forms a dimer structure through the interaction of the β-sheets and the helices of the two subunits fit into the major grooves of the DNA with a palindromic sequence. Thus E2 and IRF2 can not be classified into the same structural family in the DNA binding proteins.

At the present quality of the structure level, it is difficult to define the criterion of the motif of IRF family proteins in the canonical coordinates. We could only compare the overall secondary structure topology between IRF family proteins and the another DNA binding proteins. At the present, refined tertiary structure of IRF-2(113) has not yet been obtained. Thus detailed discussion on the structure should be done after the refinement. Based on the rerfined structure, we should determine the precise interaction site and then can propose a reliable binding model with DNA. For that purpose we should employ new NMR techniques (heteronuclear filter methods, relaxation and NOE measurements of the heteronuclei) together with gene manipulation and other physicochemical techniques. At this time the author hope that the present results described in this thesis would be helpful for the future work.

References

Branden,C. and Tooze,J. (1991) *Introduction to Protein Structure,* Garland Publishing,Inc.

Driggers,P.H., Ennist,D.L., Gleason,S.L., Mak,W.H., Marrks,M.S., Levi,B.Z., Flanagan,J.R., Appella,E. and Ozato,K. (1990) *Proc. Natl. Acad. Sci.* USA, **87**, 3743-3747.

Gabrielsen,O.S., Sentenac,A., and Fromageot,P. (1991) *Science*, **253**, 1140-1143.

Brünger, A.T. (1993) XPLOR Manual (New Haven, Connecticut: Yale University).

Güntert,P., Braun,W. and Wüthrich,K. (1991) *J.Mol.Biol.* **217**, 517-530.

Heged,R.S., Grossman,S.R., Laimins,L.A. and Sigler,P.B. (1992) *Nature*, **359**, 505-512.

Harada,H., Fujita,T., Miyamoto,M., Kimura,Y., Murayama,M., Furia,A., Miyata,T. and Taniguchi,T. (1989) *Cell*, **58**, 729-739.

Harrison,S.C (1991) *Nature*, **353**, 715-719

Howe,K.M., Reakes,C.F.L. and Watson,R.J. (1990) *EMBO.J.* **9**, 161-169.

Kanei-Ishii,C., Sakai,A., Sawazaki,T., Nakagoshi,H., He,D.-N., Ogata,K., Nishimura,Y. and Ishii,S. (1990) *J.Biol.Chem.* **265**, 19990-19995.

Kay.L.E. and Bax,A. (1990) *J.Magn.Res.* **84**, 72-84.

Lamerichs,R.M., Padilla,A., Boelens,R., Kaptein,R., Ottleben,G., *et al.*, (1990) *Proc.Natl.Acad.Sci.* USA, **86**, 6863-6867,

Miyamoto,M., Fujita,T., Kimura,Y., Maruyama,M., Harada,H., Sudo,Y., Miyata,T. and Taniguchi,T. (1988) *Cell*, **54**, 903-913.

Nilges,M., Clore,G.M. and Gronenborn,A.M. (1988) *FEBS Lett.* **229**, 317-324

Nilges,M., Habazettl,J., Brünger,T., and Holak,T. (1991) *J.Mol.Biol.* **219**, 499-510.

Ogata,K., Hojo,H., Aimoto,S., Nakai,T., Nakamura,H., Sarai,A., Ishi,S. and Nishimura,Y. (1992) *Pro.Natl.Acad.Sci. USA*, **89**, 6428-6432.

Pabo,C.O. and Saure,R.T. (1992) *Ann.Rev.Biochem.* **61**, 1053-1095.

Veals,A.V., Schindler,C., Leonard,D., Fu,X-Y., Aebersold,R., Darnell,JR.J.E., and Levy,D.E. (1992) *Mol.Cell.Biol.* **10**, 3315-3324.

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

Figure 1a

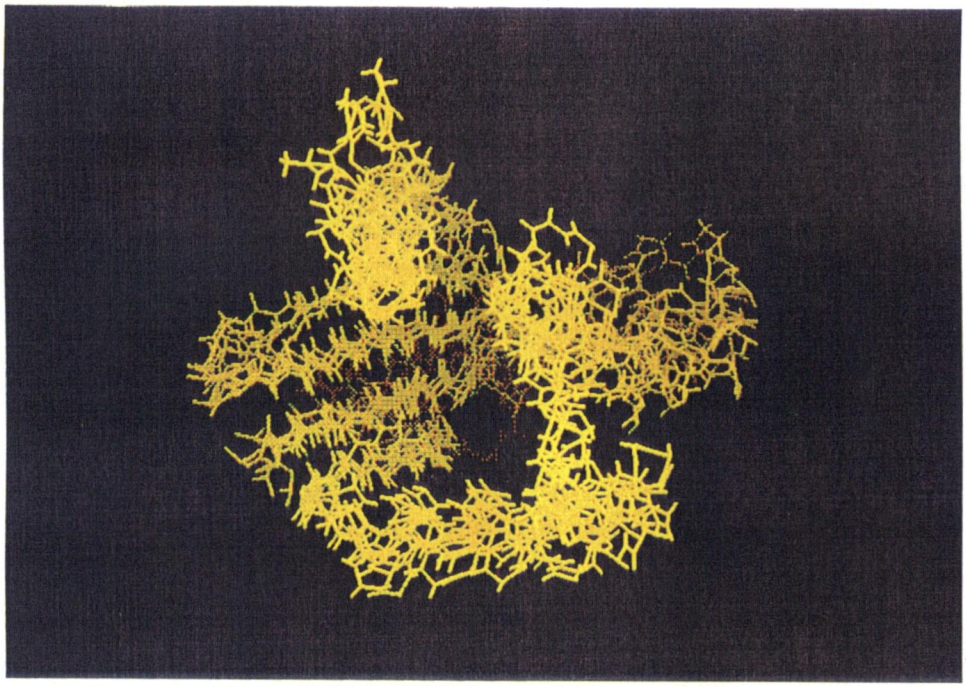
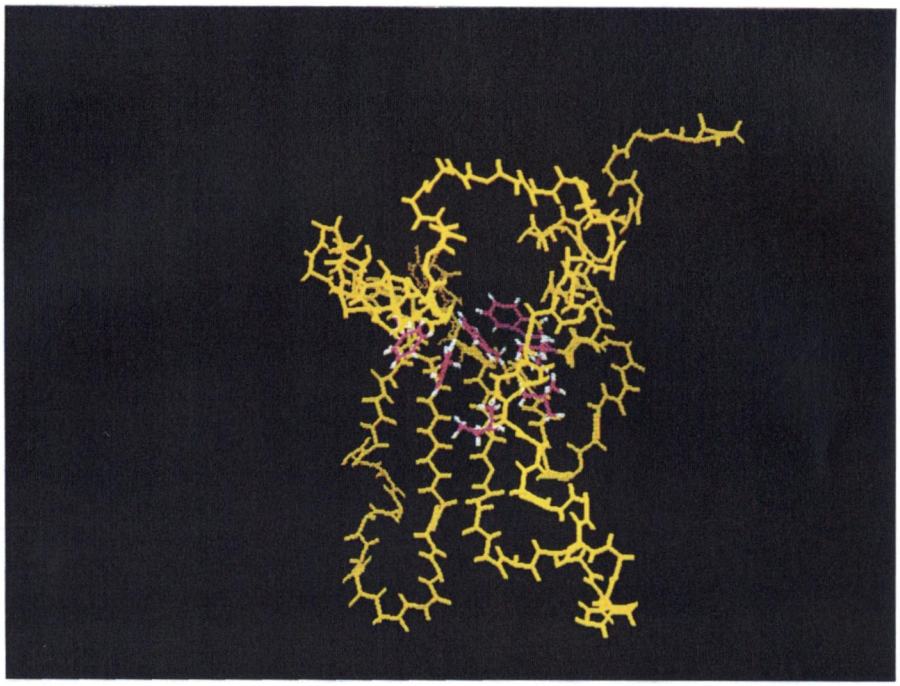


Figure 1b

Figure 2



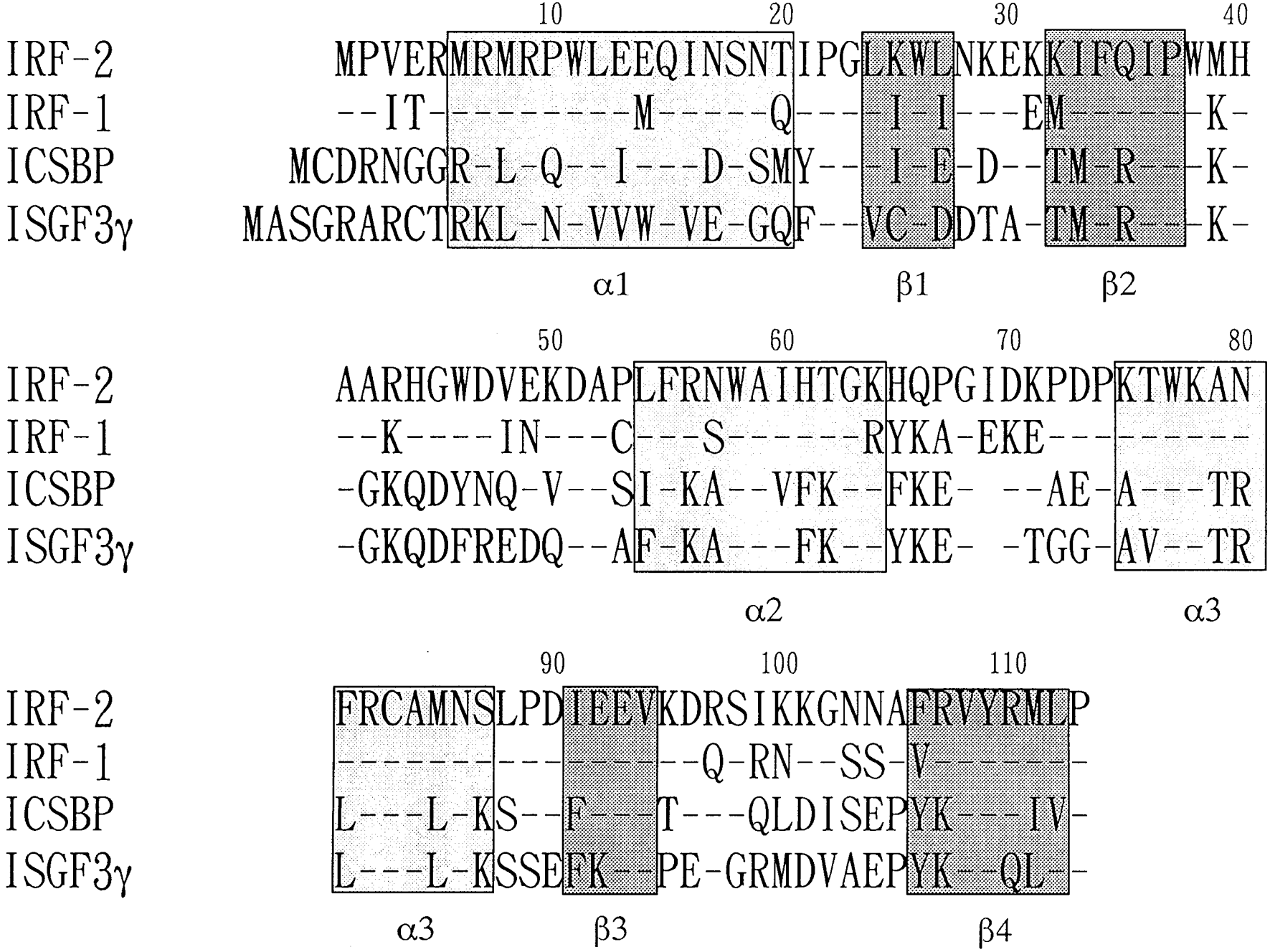
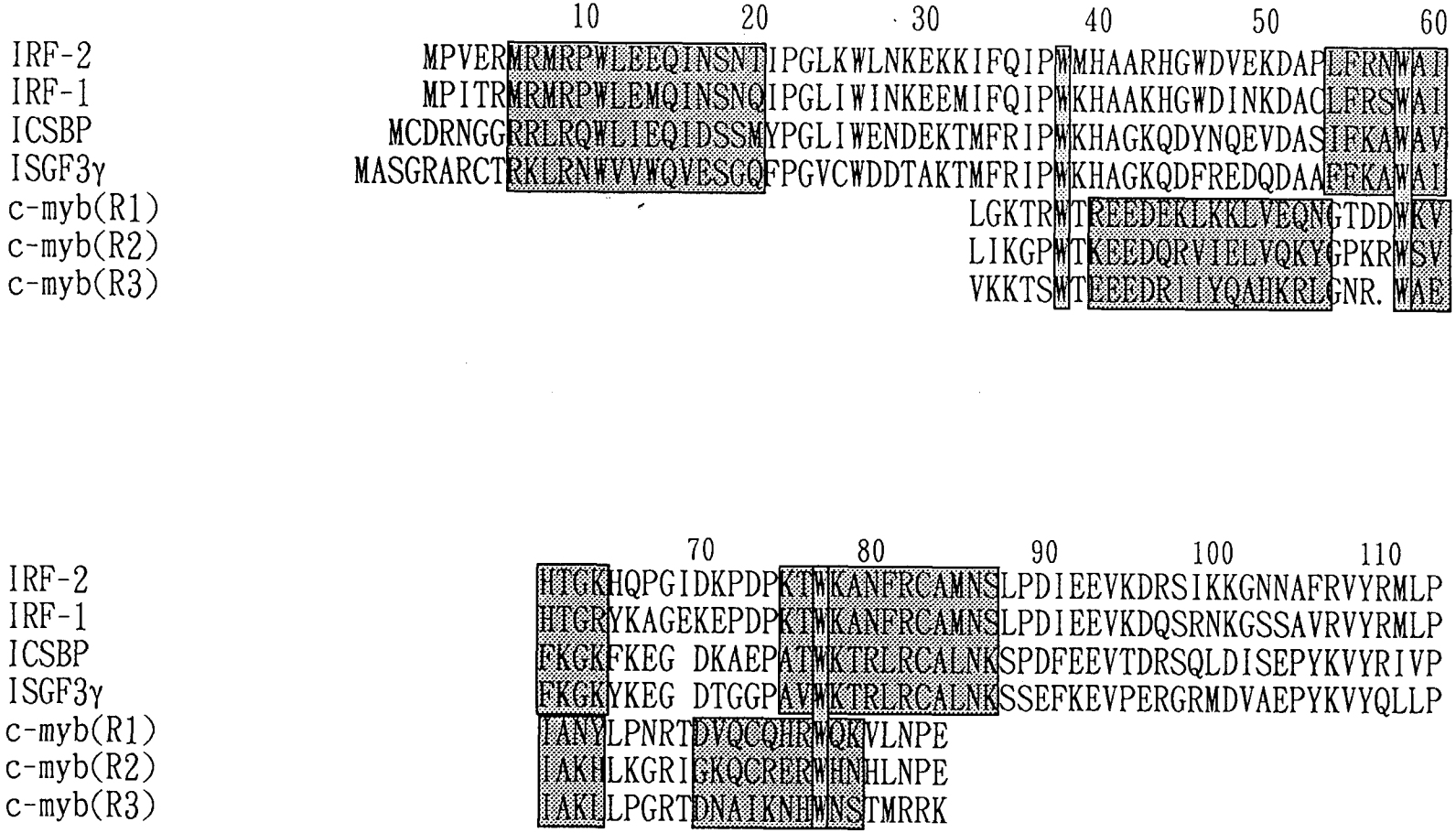

10
20
30
40
IRF-2 MPVERMRMRPWLEEQINSNTIPGLKWLNKEKKIFQIPWMH
IRF-1 --IT---------M-----Q----I-I---EM------K-
ICSBP MCDRNGGR-L-Q--I---D-SMY---I-E-D--TM-R---K-
ISGF3γ MASGRARCTRKL-N-VVW-VE-GQF--VC-DDTA-TM-R---K-
α1
β1
β2
50
60
70
80
IRF-2 AARHGWDVEKDAPLFRNWAIHTGKHQPGIDKPDPKTWKAN
IRF-1 --K----IN---C---S------RYKA-EKE----------
ICSBP -GKQDYNQ-V--SI-KA--VFK--FKE- --AE-A---TR
ISGF3γ -GKQDFREDQ--AF-KA---FK--YKE- -TGG-AV--TR
α2
α3
90
100
110
IRF-2 FRCAMNSLPDIEEVKDRSIKKGNNAFRVYRMLP
IRF-1 ---------------------Q-RN--SS-V-------
ICSBP L---L-KS--F---T---QLDISEPYK---IV-
ISGF3γ L---L-KSSEFK--PE-GRMDVAEPYK--QL--
α3
β3
β4

Figure 3



Veals *et. al.*, 1992 改変

Figure 4

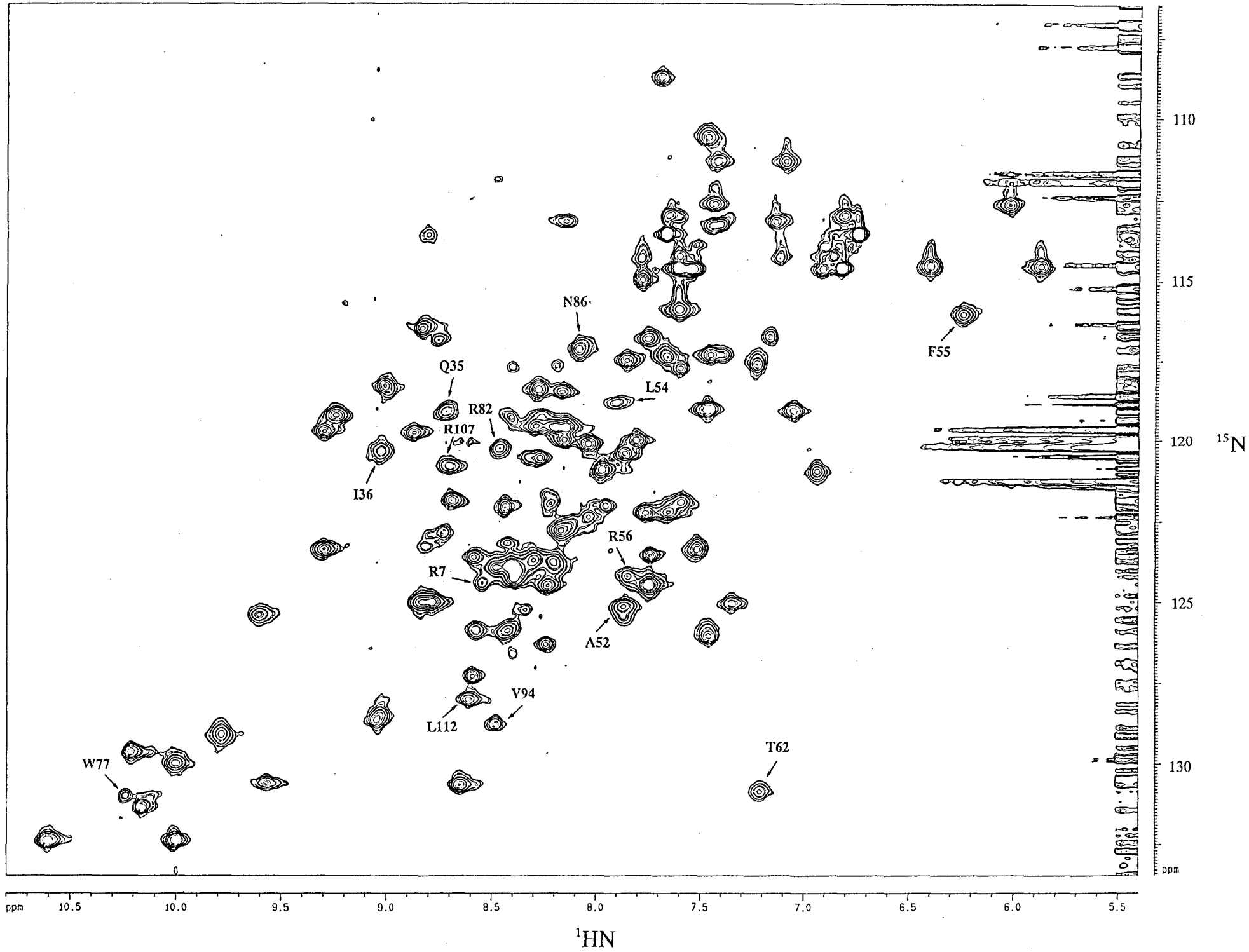

N86
Q35
L54
F55
R82
R107
I36
R56
R7
A52
V94
L112
T62
W77
110
115
120
125
130
ppm
^{15}N
ppm
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
^{1}HN

Figure 5a

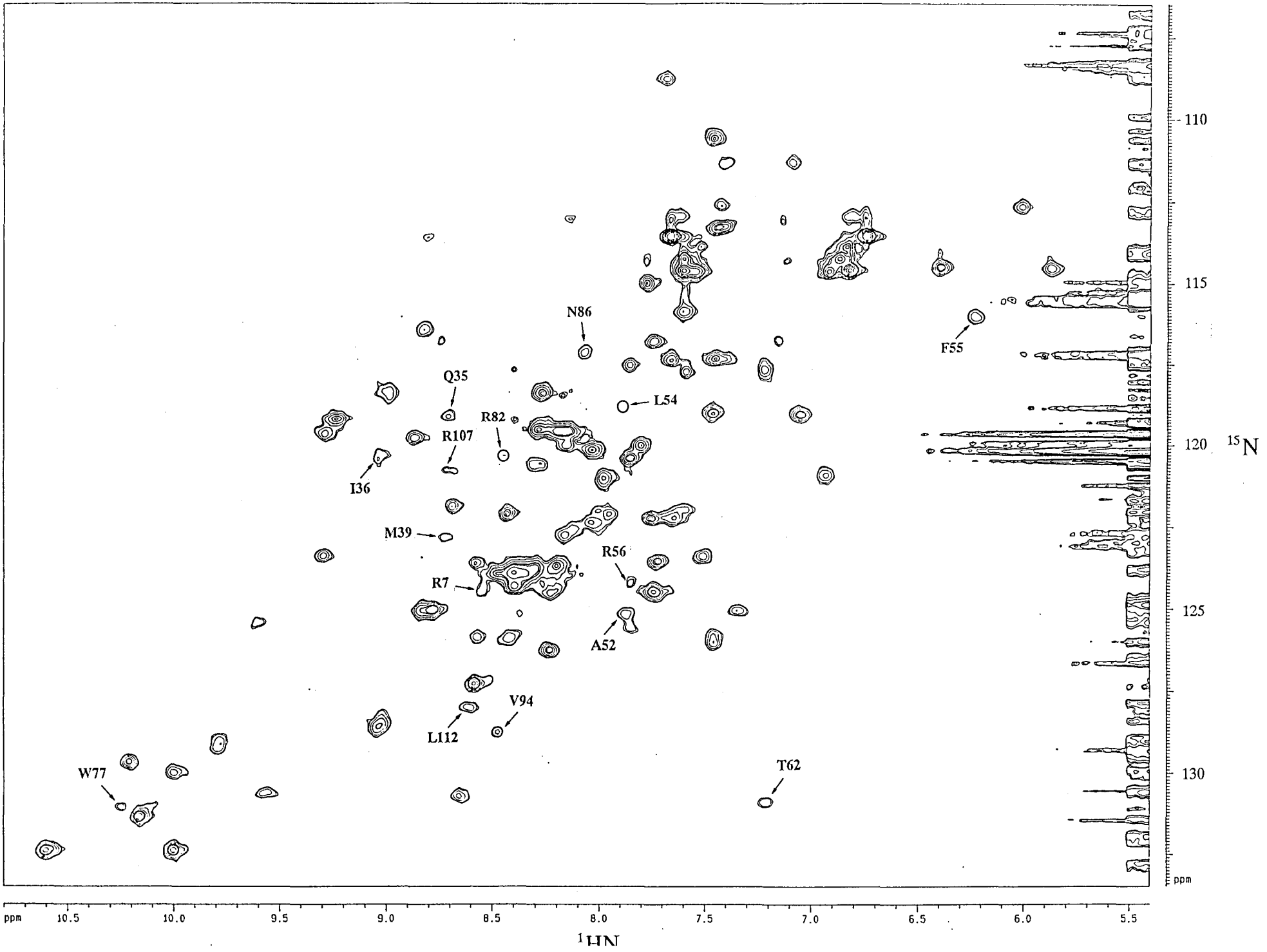

N86
Q35
L54
R82
R107
I36
M39
R56
R7
A52
V94
L112
W77
T62
F55
110
115
120
125
130
ppm
^{15}N
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
ppm
^{1}HN

Figure 5b

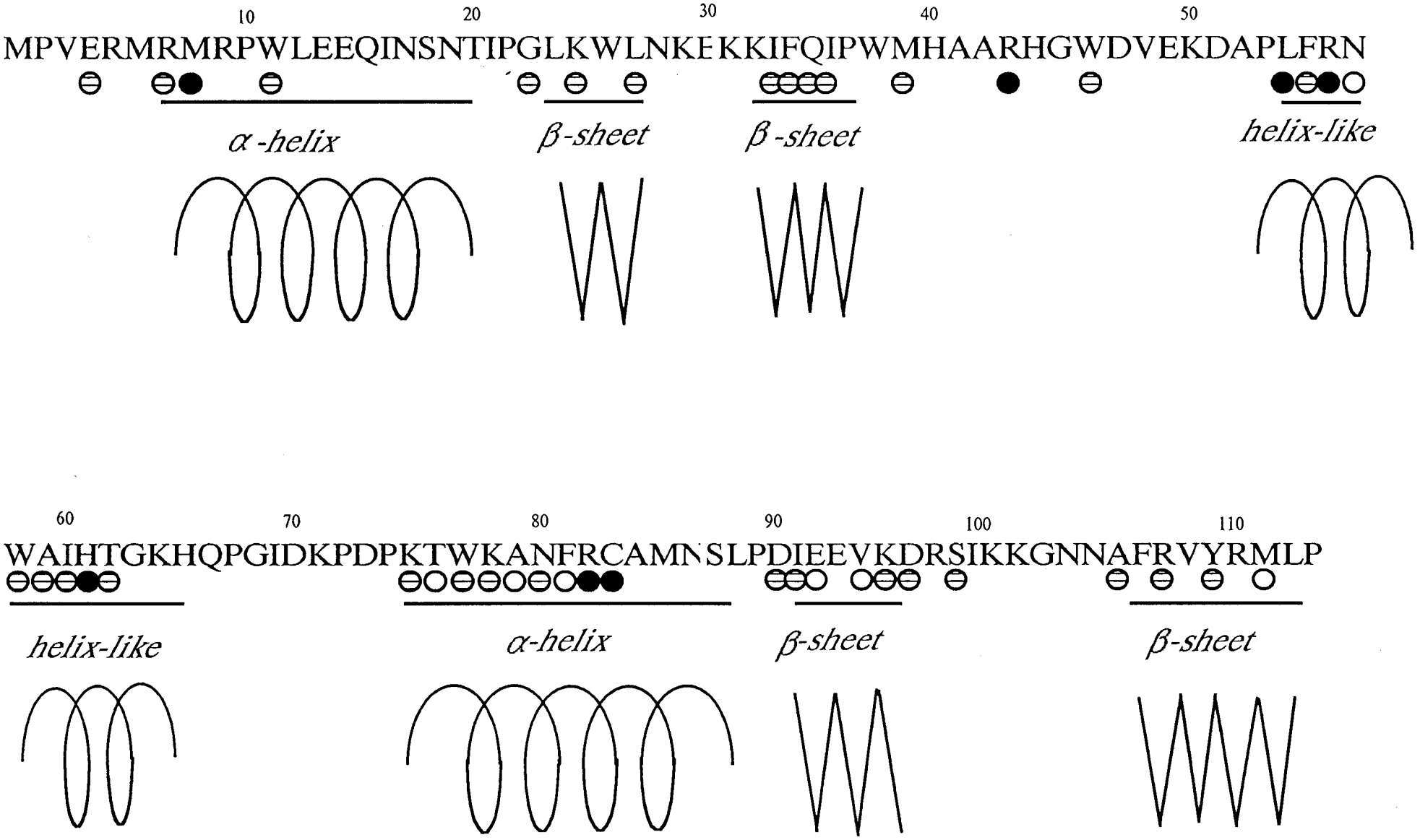

10
20
30
40
50
MPVERMRMRPWLEEQINSNTIPGLKWLNKEKKIFQIPWMHAARHGWDVEKDAPLFRN
α-helix
β-sheet
β-sheet
helix-like
60
70
80
90
100
110
WAIHTGKHQPGIDKPDPKTWKANFRCAMNSLPDIEEVKDRSIKKGNNAFRVYRMLP
helix-like
α-helix
β-sheet
β-sheet

Figure 6

Table 1.

A: NMR constraints

Distance Constraints	Number	Number of Dihedral Constraints	
Sequential ($\|i-j\|=1$)	260	ϕ constraints	38
Medium range ($\|i-j\|\leq 4$)	167	Total constraints	38
Long range	174		
Total constraints	601		

B: RMS Deviation from the Mean Structure (Å)

Residues	Backbone Heavy Atoms
Residues 2-113	2.44 ±0.22
Helix 1 (7-21)	1.60 ±0.36
Helix.2 (54-64)	2.65 ±0.46
Helix 3 (75-88)	1.50 ±0.80
β-sheet (24-27,32-37, 90-94,106-112)	1.38 ±0.19

Figure legends

Figure 1.

Preliminary result of structure calculation. a) Overlay of 5 refined simulated annealing structure of IRF-2(113). The backbone atoms are displayed. b) Ribbon representation of the average structure are shown. Yellow arrow represents the β-sheet and purple represents the α-helix. Upper and lower graphics are shown from another view points.

Figure 2.

The hydrophobic core of IRF-2(113) at the present level are indicated. Ile 16, Leu 24, Trp 26, Phe 34, Ile 36, Phe 81 and Tyr109 are shown.

Figure 3.

Sequence alignment diagram of IRF family members, IRF-2, IRF-1, ICSBP and ISGF3γ. Numbering of the residues is according to the IRF-2 sequence. Conserved residues in all proteins are indicated by one letter representation.

Figure 4.

Sequence alignment of several IRF family members and c-myb subdomains. Numbering of residues is according to the IRF-2 sequence. The consrved residues among these family members are indicated by shaded boxes.

Figure 5.

^{1}H-^{15}N HSQC spectra of DNA titration experiments. a) Control spectrum of IRF-2(113) with DNA free solution. b) The spectrum of DNA/protein solution. Ratios of DNA/protein is 0.2. Open circle indicates the vanished signals in the b spectrum.

Figure 6.

Correlation of the amide proton exchange rates and secondary structure. The positions of α-helices and β-strands are indicated. The residues whose amide protons vanished at DNA/prtotein = 0.1, 0.2, 0.4 are indicated as a close circle, open circle, and shaded circle, respectively.

Conclusion

DNA binding proteins recognize and interact with specific DNA sequence through special folding units called DNA binding domains. In many cases of the eukaryotic transcriptional factors, these domains can be interchanged among different proteins, indicating that these domains are independently folded units. Several families of DNA binding proteins have been identified on the basis of the primary sequence homology and the three dimensional structures. A familial relationship is a useful unifying theme in the study of DNA binding domains, because it relates to evolution, structure, DNA recognition mechanism, gene regulation and structure design. However, there are many DNA binding proteins that can not be classified into the known families. The author was interested in the structure and function of interferon regulatory factors which regulate the IFN-β and IFN inducible gene expression because they are not classified into the known family and bind to a unique direct repeat DNA sequence but not palindromic one. The author started the structure analyses of the DNA binding domain of IRF-2 to solve the familial relationship and DNA recognition mechanism by using NMR.

In this thesis, the author described the identification of the DNA binding domain of IRF-2 protein and three dimensional structure analysis of the domain by NMR. Then the interaction with DNA was examined.

To analyze the three dimensional structure of DNA binding domain, the author should start from the determination of the position of the DNA binding domain of IRF-2 and construction of the expression vector for its DNA binding domain. For that purpose, the author performed a limited proteolysis experiment and determined the position of the DNA binding domain. The overexpression system for the domain with functional form, was successfully obtained. An ordinal proton NMR spectrum of the DNA binding domain of IRF-2 showed that the conventional approach was difficult for this sample because of severe signal degeneracy and broad linewidth due to rapid relaxation. Therefore, the author performed newly developed NMR technique called triple-resonance multidimensional NMR to overcome these problems. By using the technique, nearly complete backbone assignment could be done. Based on the distance imformation obtained by NMR, the three dimensional structurer was calculated. The result showed that the DNA binding domain of IRF-2 has a new folding unit. This domain contains three α-helices α1, α2 and α3 and four β-strands arranged in a antiparallel β-sheet. This folding pattern was not classified into the known category. There are some transcriptional factors that show an apparent sequence homology with the DNA binding domain of IRF-2. The new folding pattern found by the present investigation suggest that these factors compose a new family of DNA binding proteins. All factors in this familly are related to the IFN and IFN inducible gene expression, so they may be originally evoluted differently from the other transcriptional factors.

Based on the analysis of the interaction of IRF-2(113) with DNA, a possible mode of DNA sequence recognition involving insertion of the 3 helix into the major groove of DNA was suggested.

More investigations will be demanded to understand the DNA recognition mechanism in detail and to establish the new category of DNA binding motif. To solve these problems, more refined structure and extensive analysis of interaction with DNA are required for getting more detailed aspects. The author wishes that the present results described in this thesis would be helpful for future analyses.

Acknowledgements

The present work has been performed under the direction of Professor Yoshimasa Kyogoku, Institute for Protein Research, Osaka University. The author would like to express his sincere gratitute to Professor Y.Kyogoku and to Dr. M. Shirakawa for their sincere guidence, discussions and intimate encouragements throughout the course of his study. He also wishes to his sincere thanks to Professor T.Taniguchi, Dr. N.Tanaka, and Dr.H.Harada, Institute for Molecular and Cellular Biology, Osaka University, and Dr. T.Fujita, Tokyo Metropolitan Institute of Medical Science, for generous gifts of the plasmids containing the IRF-1 and -2 gene and for precious discussions for IRFs. Thanks are also due to Dr. S. Tsunazawa and Ms. Y. Yagi, Institute for Protein Research, for the analyses of the N-terminal sequence and amino acid composition. Also wishes to Dr.F.Hayashi, Shionogi Corporation, for measurement the HCCH-TOCSY spectrum.

During the course of this study, the author has received a number of fruitful discussions, helpful comments and hearty encouragement from several persons. The author wishes to express his appreciation to Associate Professors H.Sugeta, and Y.Kobayashi. His thanks are also due to all the members of Division of Molecular Biophysics of Institute for Protein Research, especially, to Mr. H. Matsuo and Mr. C. Kojima for helpuful advices and supports to the NMR measurement and to Mr. N. Nemoto for helpuful advice and support to the structure caluculation. The author also thanks to Mr. M. Shimizu and Mr. T.Kodama for important technical and useful advices for biochemical techniques. Finally, the author thanks sincerely to his parents for their unfailing understanding and affectionate encouragements.

Koichi Uegaki

February, 1994

List of Publications

Characterization of the DNA Binding Domain of the Mouse IRF-2 protein.
Protein Engineering, **6**, 195-200, (1993)
K.Uegaki, M.Shirakawa, T.Fujita, T.Taniguchi and Y.Kyogoku

Archaerhodopsin-2, from Halobacterium sp.aus-2 Further Reveals Essential Amino Acid Residues for Light -Driven Proton Pumps.
Archives of Biochemistry and Biophysics, **286**, 107-110. (1991)
K.Uegaki, Y.Sugiyama, and Y.Mukohata

Australian Halobacteria and Their Retinal-Protein Ion Pumps.
Photochemistry and Photobiology, **54**, 1039-1045, (1991)
Y.Mukohata, K.Ihara, K.Uegaki, and Y.Sugiyama