Transforming Growth Factor β1: NMR Signal Assignments of the Recombinant Protein Expressed and Isotopically Enriched Using Chinese Hamster Ovary Cells†

Sharon J. Archer,‡ Ad Bax,§ Anita B. Roberts,∥ Michael B. Sporn,∥ Yasushi Ogawa,⊥ Karl A. Piez,# James A. Weatherbee,▲ Monica L.-S. Tsang,▲ Roger Lucas,▲ Bo-Ling Zheng,▲ Jeff Wenker,▲ and Dennis A. Torchia*,‡

Bone Research Branch, National Institute of Dental Research, Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, and Laboratory of Chemoprevention, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, Celtrix Pharmaceuticals, Inc., Santa Clara, California 95054, Department of Biochemistry and Molecular Biology, Jefferson Medical College, Philadelphia, Pennsylvania 19107, and R&D Systems Inc., Minneapolis, Minnesota 55413

Received September 21, 1992; Revised Manuscript Received November 16, 1992

ABSTRACT: The transforming growth factor βs are a homologous family of multifunctional cytokines that regulate cell growth and differentiation. As a prelude to studies of the solution structure and dynamics of TGF-β1, we report virtually complete assignment of ^1H and ^{15}N resonances for this 25-kDa homodimeric protein. Recombinant TGF-β1 was expressed in Chinese hamster ovary cells. The cells were grown either in a completely ^{15}N-enriched medium or in a medium containing selectively ^{13}C,^{15}N-labeled amino acids to obtain either uniformly or specifically labeled protein, respectively. Two- and three-dimensional heteronuclear edited magnetic resonance spectra of the uniformly ^{15}N-labeled protein and three samples selectively labeled with ^{13}C and ^{15}N yielded assignments for 96% of the backbone amide and Cα protons and 87% of the side chain protons. To our knowledge, this is the first report of the use of an animal cell expression system to obtain extensive isotopic enrichment in order to sequentially assign a protein. The methodology described herein for the isotopic enrichment and resonance assignments of TGF-β1 should be generally applicable to other eukaryotic proteins expressed by animal cells.

The transforming growth factor βs (TGF-βs)[1] are important regulators of numerous physiological processes including normal tissue growth and wound repair (Roberts & Sporn, 1990; Sporn & Roberts, 1990; Massagué, 1990). The diverse activities of these proteins include their ability to act as multifunctional regulators of the growth and differentiation of many types of cells. The TGF-βs are important in maintenance of normal epithelial structures, in the formation and repair of connective tissue and bone, and in the formation and function of the cells of the blood and the immune system. TGF-βs and their receptors are present in nearly all cells. Five homologous forms of TGF-β have been characterized, two of which have been isolated from natural sources in significant quantities: TGF-β1 and TGF-β2. TGF-β1 was originally isolated from human platelets but also has been purified from additional sources including bovine kidney, bovine bone, and porcine platelets. Recently, TGF-β1 has been cloned and expressed in Chinese hamster ovary (CHO) cells (Gentry et al., 1987), a development that has opened the way to produce large quantities of isotopically enriched protein.

TGF-β1 is synthesized as a 390-residue latent glycosylated precursor (Derynck et al., 1985; Gentry et al., 1987). Processing of the latent precrusor results in the loss of the N-terminal portion of the peptide and release of the mature TGF-β1 dimer (Gentry et al., 1988). Mature TGF-β1 is a 25K homodimer in which each chain has 112 amino acid residues, nine of which are cysteines. Reduction of the active dimer results in formation of inactive monomers. Activity is not recovered by simple reoxidation of the protein.

As a first step in using NMR spectroscopy to determine the solution structure of TGF-β1, we have sequentially assigned the ^1H and ^{15}N resonances of TGF-β1. Powerful new triple-resonance NMR techniques enable one to assign proteins having M_r in the range 15–30K in an efficient manner, provided that the protein is uniformly enriched with ^{13}C and ^{15}N (Ikura et al., 1990a, 1991; Pelton et al., 1991; Grzesiek et al., 1992). Proteins expressed in bacterial systems can be uniformly enriched with ^{15}N and ^{13}C by growing the bacterial cells on minimal media with ^{15}N-labeled ammonium chloride and ^{13}C-labeled glucose. However, it has been difficult to express and refold TGF-β1 in any bacterial system. For our purposes, TGF-β1 was properly expressed and folded in animal cells grown in medium containing all the naturally occurring amino acids. Because the necessary $^{13}C/^{15}N$-labeled amino acids were not available, we prepared a sample of TGF-β1 uniformly ^{15}N enriched, and three additional samples ^{13}C and ^{15}N enriched at selective sites for a number of specific types of amino acids. Herein we report the labeling methodology and the combination of heteronuclear edited NMR spectra used to obtain the sequential assignments for recombinant TGF-β1 obtained from a CHO cell expression system. Our approach

† This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to D.A.T.) and Public Health Service National Research Service Award GM13620 (to S.J.A.). Part of this work was done while K.A.P. was a Scholar-in-Residence at the Forgarty International Center, National Institutes of Health.

‡ National Institute of Dental Research.
§ National Institute of Diabetes and Digestive and Kidney Diseases.
∥ National Cancer Institute.
⊥ Celtrix Pharmaceuticals, Inc.
Jefferson Medical College.
▲ R&D Systems Inc.

[1] Abbreviations: CHO, Chinese hamster ovary cells; DQF-COSY, double quantum filtered correlated spectroscopy; HMQC, heteronuclear multiple quantum correlation; HNHB, proton to amide nitrogen to β-proton correlation; HOHAHA, homonuclear Hartmann–Hahn; HSQC, heteronuclear single quantum correlation; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; TFA, trifluoroacetic acid; TGF-β1, transforming growth factor β1; 2D, two-dimensional; 3D, three-dimensional.

0006-2960/93/0432-1152$04.00/0

is generally applicable to other proteins synthesized in animal cells. In a subsequent paper, we will report the secondary structure of TGF-β1 on the basis of the assignments reported herein.

MATERIALS AND METHODS

Cloning and Amplification Procedures. For preparation of labeled material, human TGF-β1 was cloned from COLO 320HSR cells using procedures essentially the same as those described previously for the cloning of human TGF-β2 (Caltabiano et al., 1989). Subsequently, plasmids containing the TGF-β1 cDNA and the gene for DHFR were used to cotransfect DHFR-deficient CHO cells. Amplified levels of TGF-β1 expression were obtained from growth of the transfected cells on medium containing increasing levels of methotrexate (Kingston et al., 1989).

Bioreactor Growth Conditions. For the production of recombinant TGF-β1, the CHO cell clone was grown in an S-51 ceramic matrix Opticore culture chamber using the Opticell culture system (Charles River Biotechnical Services, Inc., Wilmington, MA) according to the manufacturer's suggested protocols. During production phase, MEM-α, containing no nucleosides, 1% dialyzed heat-inactivated fetal calf serum, and 500 nM methotrexate and substituted with the appropriate ^{15}N- or ^{15}N-/^{13}C-labeled compounds was used. When switching between growth schemes using unlabeled or differentially labeled media, the bioreactor was perfused with 2–3 L of balanced salt solution to reduce cross-contamination between the schemes.

To obtain completely ^{15}N-labeled TGF-β1, the medium contained uniformly ^{15}N-labeled choline and uniformly ^{15}N-labeled amino acids (except Trp which was ^{15}N-labeled only on the backbone position) as described in the tables in the supplementary material. To obtain samples containing specific types of amino acids labeled with ^{15}N and ^{13}C, cells were grown in three different media containing the amounts of labeled and not labeled amino acids listed in the tables in the supplementary material. Labeled amino acids were obtained from Merck Isotopes (St. Louis, MO) and Cambridge isotopes (Cambridge, MA) as detailed in the supplementary material. Approximately 1 mg of labeled TGF-β1 was obtained per liter of labeled growth medium. The cost of the labeled media was ca. $1000/L. Although we seriously considered the prospect of preparing a uniformly ^{15}N/^{13}C-labeled protein, the necessary labeled amino acids were not available.

TGF-β1 Purification Procedure. All recombinant TGF-β1 samples were purified to greater than 97% homogeneity from acidified CHO cell conditioned medium using a two-step procedure consisting of cation-exchange chromatography on S Sepharose Fast Flow (Pharmacia) and HPLC on SynChropak C4 reverse-phase columns (SynChrom, Inc., Lafayette, IN). During purification, the biological activity of TGF-β1 was monitored by its ability to inhibit the IL4-dependent proliferation of HT-2 cells (Tsang et al., 1990).

Gel Electrophoresis Procedure. The purity of the labeled TGF-β1 was estimated using SDS–PAGE. Samples were electrophoresed on 15% polyacrylamide gels under nonreducing conditions using the Laemmli procedure (Laemmli, 1970). The gels were silver-stained as described by Morrisey (Morrisey, 1981). After purification was complete, the biological activity was again tested using the HT-2 assay (see above).

TGF-β1 from Bovine Bone. Nonlabeled natural TGF-β1 was isolated from bovine bone as described by Ogawa and Seyedin (1991). The protein samples were purified using gel filtration chromatography (Sephacryl S-200, Pharmacia), cation-exchange column chromatography (CM52, Whatman), and C18 reverse-phase HPLC (Vydac, Hesperia, CA) in acetonitrile/trifluoroacetic acid (TFA) solvent systems as described previously (Ogawa & Seyedin, 1991). The purity of the protein was determined using gel electrophoresis under nonreducing conditions (Laemmli, 1970) and staining of the gel with silver (Morrissey, 1981). The specific biological activity of the protein was determined by assaying for the inhibition of proliferation of mink lung epithelial cells (Mv1Lu CCL64; ATCC, Rockville, MD) in culture (Ogawa & Seyedin, 1991).

To prepare the protein for NMR spectroscopy, 8–10 mg of TGF-β1 in 30% acetonitrile/70% H_2O/0.1% TFA was lyophilized and then dissolved in 30% [D_3]acetonitrile/70% H_2O and lyophilized to remove any residual TFA and protonated acetonitrile. The protein was then dissolved in approximately 0.5 mL of 90% H_2O/10% D_2O and the pH was adjusted to 4.2 with 0.1 N NaOH. To prepare a sample in D_2O, the protein was dissolved in 0.5 mL of D_2O, incubated at 37 °C for 2 h, and lyophilized. The protein was lyophilized from 0.5 mL of D_2O at least two more times, dissolved in 99.996% D_2O, and transferred to an NMR sample tube which was then flame sealed. Unless noted otherwise, the protein concentration of the samples was approximately 1 mM (dimer). Samples were stored at 4 °C.

NMR Spectroscopy. NMR spectra of TGF-β1 were acquired on Bruker AMX 500 and 600 spectrometers and on a Bruker AM 500 spectrometer modified to reduce overhead time (Kay et al., 1990). 2D NMR spectra were processed using Bruker and NMRi (New Methods Research, Inc. Syracuse, NY) software available on ASPECT 1000 and SUN data stations, respectively. The NMR spectra were processed using a combination of in-house (Kay et al., 1989; S. Grzesiek, unpublished) and NMRi software and were analyzed and plotted using in-house software (Garrett et al., 1991).

All spectra were acquired at 45 °C unless noted otherwise. A 2D DQF-COSY (Rance et al., 1983) spectrum was acquired in D_2O. 2D HOHAHA (Braunschweiler & Ernst, 1983; Bax & Davis, 1985) spectra were acquired in D_2O and in H_2O with a WALTZ-17 mixing sequence (Bax & Davis, 1985) with mixing times of ca. 16 ms, 30 ms, and 50 ms. For HOHAHA spectra in H_2O, solvent suppression was accomplished using a jump return sequence (Bax et al., 1987). 2D NOESY spectra in D_2O and in H_2O were acquired at both 37 °C and 45 °C with a 100-ms mixing time. In the H_2O sample, solvent presaturation ($\gamma B_2/2\pi$ = 20 –25 Hz) was used.

2D ^1H–^{15}N HSQC (Bodenhausen & Ruben, 1980) spectra were acquired on uniformly ^{15}N-labeled TGF-β1 at 37 °C and at 45 °C and on specifically ^{13}C/^{15}N-labeled samples at 45 °C. The water signal was suppressed using 25 Hz of presaturation or ca. 15 Hz solvent presaturation and a 1-ms spin-lock pulse (Messerle et al., 1989). 2D ^1H–^{13}C HMQC (Bax et al., 1990) spectra were acquired on natural abundance ^{13}C TGF-β1 and selectively ^{13}C/^{15}N-labeled samples in D_2O. 2D ^1H–^{13}C HMQC–RELAY (Lerner & Bax, 1986) spectra were recorded on the specifically ^{13}C/^{15}N-labeled samples using a ca. 30-ms WALTZ-17 mixing sequence. 2D ^{13}C-edited NOESY (Bax & Weiss, 1987) and 2D ^1H–^{13}C HSQC–NOESY spectra were acquired on the specifically ^{13}C/^{15}N-labeled samples with either a 100-ms or a 200-ms mixing time. The 2D HSQC–NOESY experiment consists of a ^1H–^{13}C HSQC pulse sequence (which yields ^{13}C chemical shift evolution in t_1) followed by an NOE mixing period and proton detection in t_2. In all heteronuclear experiments, WALTZ-

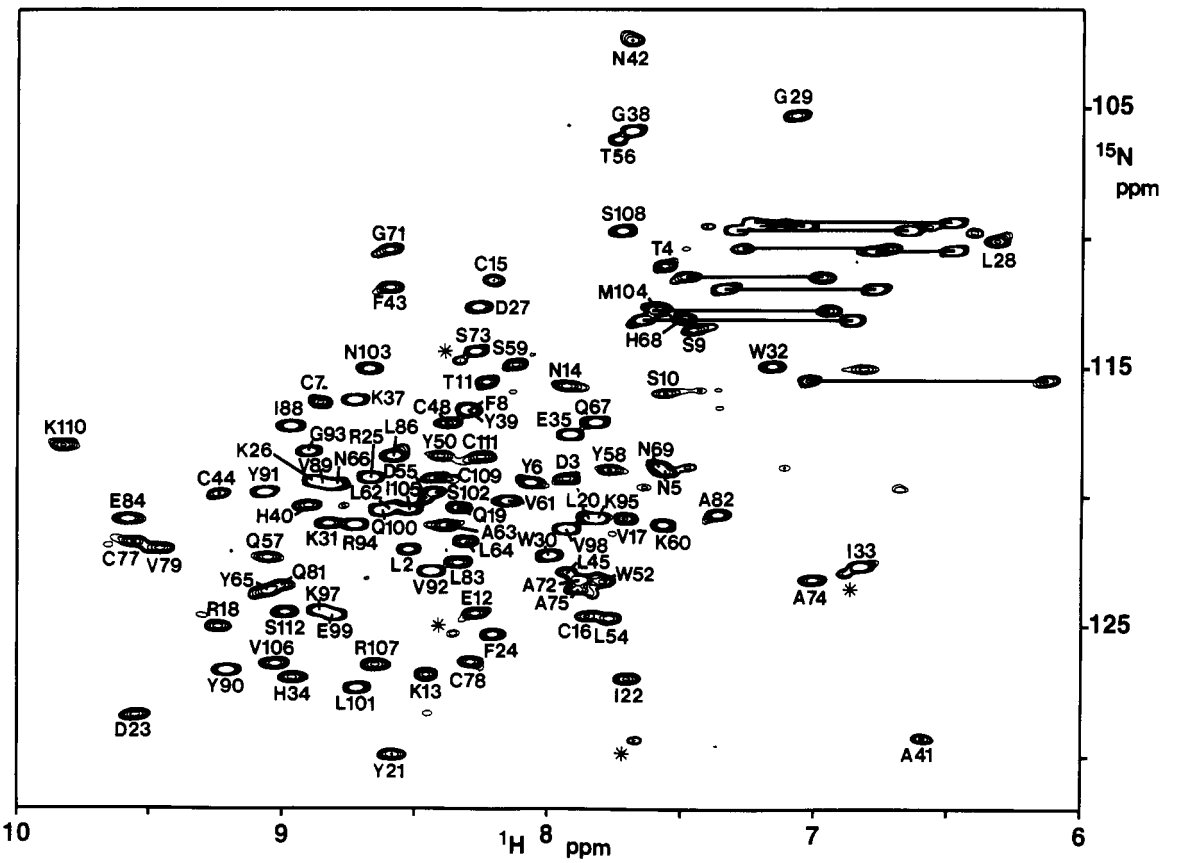

FIGURE 1: ^1H–^{15}N HSQC spectrum of uniformly ^{15}N-enriched TGF-β1. Backbone amide resonances are labeled according to sequential assignments. Gln and Asn side chain NH_2 groups are indicated by lines between the proton chemical shifts. Minor peaks are noted with an asterisk. The spectrum was acquired at 500 MHz with 512 complex points in t_2 and 300 complex points in t_1. The spectrum was processed with Lorentzian to Gaussian digital filtering and zero-filled once in each dimension.

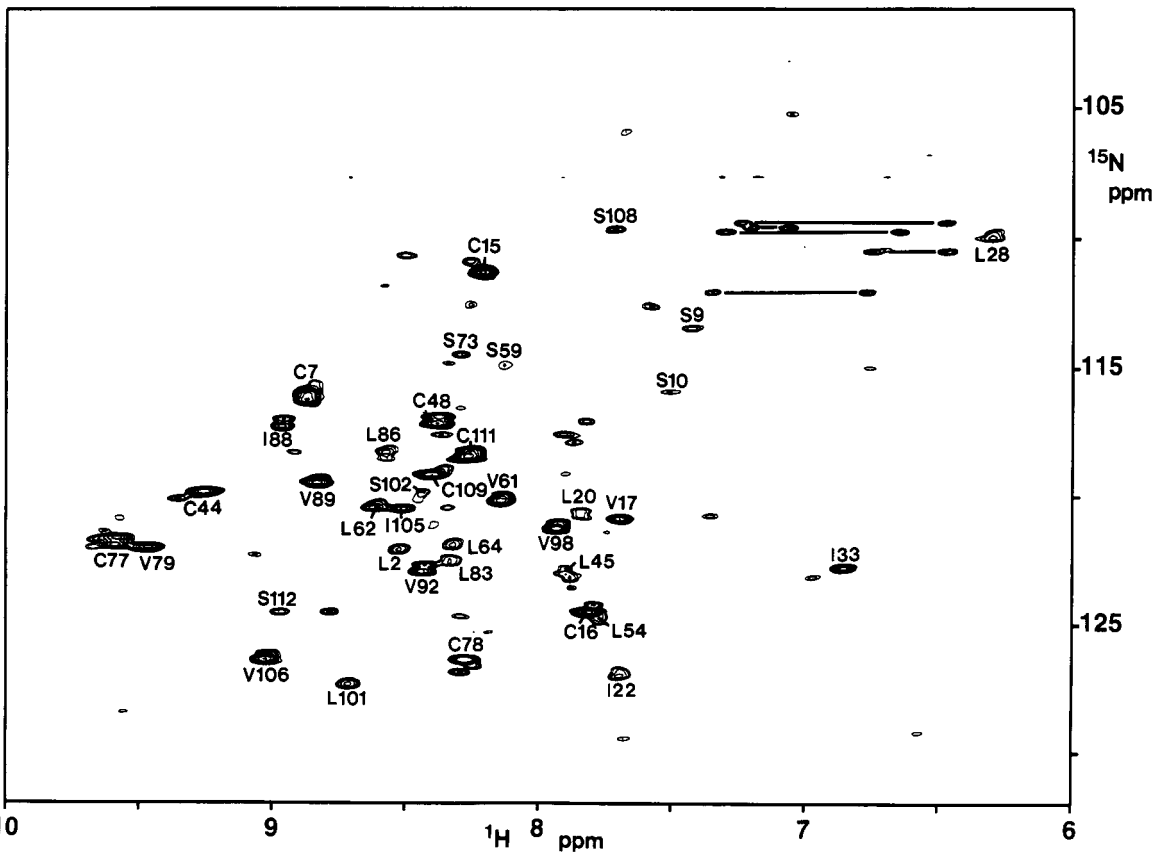

FIGURE 2: ^1H–^{15}N HSQC spectrum of selectively labeled TGF-β1 (scheme 3). The major peaks are labeled according to sequential assignments. Gln side chain NH_2 groups are indicated by lines between the proton chemical shifts. The spectrum was acquired at 500 MHz with 512 complex points in t_2 and 300 complex points in t_1. The spectrum was processed with Lorentzian to Gaussian digital filtering in both dimensions and zero-filled once in t_1.

Table I: Percentages of Specifically ^{13}C- and ^{15}N-Enriched Amino Acids Added to the Growth Medium in Order to Selectively Label TGF-β1

scheme 1	% labeled	scheme 2	% labeled	scheme 3	% labeled
[1-^{13}C,^{15}N]Leu	100	[1-^{13}C]Ile	100	[1-^{13}C]Pro	100
[1-^{13}C]Val	80	[1-^{13}C]Tyr	80	[1-^{13}C]Tyr	80
[1-^{13}C]Pro	67	[1-^{13}C]Lys	67	[1-^{13}C]Lys	67
				[1-^{13}C]Ile	50
[^{15}N]Gly	66	[^{15}N]Trp	100	[^{15}N]Cys	100
[^{15}N]Tyr	33	[^{15}N]His	60	[^{15}N]Val	60
[^{15}N]Lys	20	[^{15}N]Ser	40	[^{15}N]Ile	30
		[^{15}N]Gly	25	[^{15}N]Ser	15
		[^{15}N]Phe	25		
		[^{15}N]Arg	12		
[2-^{13}C,^{15}N]Pro	33	[2-^{13}C]Ala	100	[2-^{13}C]Leu	100
[3-^{13}C]Cys	100	[γ-^{13}C]Pro	100	[3-^{13}C]Ala	100
[*methyl*-^{13}C]Met	100	[D_{10}]Leu	100	[3-^{13}C]Phe	100
[*amide*-^{15}N]Asn	100			[3-^{13}C]Ser	77
[*ring*-D_5]Phe	100			[*ring*-2-^{13}C]His	100
[*phenol*-3,5-$^{13}C_2$]Tyr	67				

Table II: Comparison of Percentage Enrichment of ^{15}N Amides in the Growth Media and in the Corresponding Labeled Samples of TGF-β1

scheme 1			scheme 2			scheme 3		
	% labeled			% labeled			% labeled	
residue	medium	TGF-β1	residue	medium	TGF-β1	residue	medium	TGF-β1
L	100	46	W	100	100	C	100	100
G	66	41	H	60	68	V	60	34
Y	33	23	S	40	31	I	30	19
K	20	16	G	25	31	S	15	6
I		29	F	25	31	L		9
S		20	R	12	15	M		8
V		16	L		8	A, D, E, Q		4–5
A, D, E, Q		8–13	Y		8	G		3
$N\gamma NH_2$	100	107	I		6	$Q\delta NH_2$		3
$Q\delta NH_2$		16	A, D, E, Q		5–7			
			K		4			
			V		3			
			$N\gamma NH_2$		21			
			$Q\delta NH_2$		5			

16 modulation (Shaka et al., 1983) was used to decouple ^{13}C or ^{15}N from the protons.

3D Spectra were acquired on uniformly ^{15}N-labeled TGF-β1. 3D 600-MHz HOHAHA–HSQC (Marion et al., 1989a) and 3D NOESY–HMQC (Marion et al., 1989b; Kay et al., 1989) experiments were acquired at both 37 °C and 45 °C. A DIPSI-2 mixing sequence (Shaka et al., 1988) was used in the 3D HOHAHA–HSQC with mixing times of 39 ms at 37 °C and 58 ms and 80 ms at 45 °C. The HOHAHA–HSQC experiments were acquired with spectral widths of 8.33, 21.9, and 15.43 ppm in F_1 (^1H), F_2 (^{15}N), and F_3 (^1H), respectively, and with 100 complex points in t_1, 31 complex points in t_2, 512 complex points in t_3, and 16 scans per t_3 point. The 3D NOESY–HMQC spectrum was acquired at 45 °C at 600 MHz with a 90-ms mixing time and at 37 °C at 500 MHz with a 100-ms mixing time. The 600-MHz NOESY–HMQC experiment was acquired with spectral widths of 21.9 ppm in F_2 (^{15}N) and 10.82 ppm in F_1 (^1H) and F_3 (^1H). The 500-MHz NOESY–HMQC experiment was acquired with spectral widths of 10.00, 22.9, and 16.12 ppm in F_1 (^1H), F_2 (^{15}N), and F_3 (^1H), respectively. Both experiments were acquired with 128 complex points in t_1, 32 complex points in t_2, 384 complex points in t_3 at 600 MHz and 1024 real points in t_3 at 500 MHz, and 16 scans per t_3 point. The 3D HNHB spectrum used to identify Hβ protons was acquired at 600 MHz as described previously (Archer et al., 1991). A 500-MHz 3D ^{15}N–^{15}N–^1H HMQC–NOESY–HMQC (Ikura et al., 1990b; Frenkiel et al., 1990) spectrum was acquired with a 100-ms mixing time at 37 °C. The spectrum was acquired with spectral widths of 29.6, 22.9, and 16.12 ppm in F_1 (^{15}N), F_2 (^{15}N), and F_3 (^1H), respectively, and with 64 complex points in t_1, 32 complex points in t_2, 1024 real points in t_3, and 16 scans per t_3 point. All 3D spectra were acquired with the ^1H carrier set on water and the ^{15}N carrier at 116.5 or 120 ppm.

RESULTS AND DISCUSSION

Isotopic Enrichment. A ^1H–^{15}N HSQC spectrum for uniformly ^{15}N-labeled TGF-β1 is shown in Figure 1. The excellent signal to noise ratio is comparable to that observed in HSQC spectra of uniformly enriched staphyloccocal nuclease (Baldisseri et al., 1991) and *Escherichia coli* protein III^{glc} (Pelton et al., 1991) indicating that the sample is highly and uniformly ^{15}N-enriched. A comparison of 500-MHz (a) ^{15}N edited with (b) ^{15}N filtered one-dimensional spin echo difference spectra showed that the level of ^{15}N enrichment was greater than 90%. A total of 35 mg of purified, fully active, ^{15}N-enriched TGF-β1 was obtained from three cell growths. The yields of purified TGF-β1 selectively enriched with ^{13}C and ^{15}N were 5 mg, 7 mg, and 10 mg, respectively, for schemes 1, 2, and 3. The lower than expected yields for schemes 1 and 2 could be due to contaminants in the labeled amino acids or slow adaptation of the cells to the labeled medium.

The extent of ^{15}N-enrichment of the selectively labeled samples was estimated by comparing the intensities of the cross-peaks in the ^1H–^{15}N HSQC spectra of the selectively labeled samples with the intensities in the spectrum of the uniformly labeled sample. The ^1H–^{15}N HSQC spectrum of TGF-β1, labeled according to scheme 3 in Table I, is shown in Figure 2. Comparison of this spectrum with the spectrum of uniformly ^{15}N-enriched protein (Figure 1) shows that the protein is selectively ^{15}N labeled. The levels of ^{15}N incor-

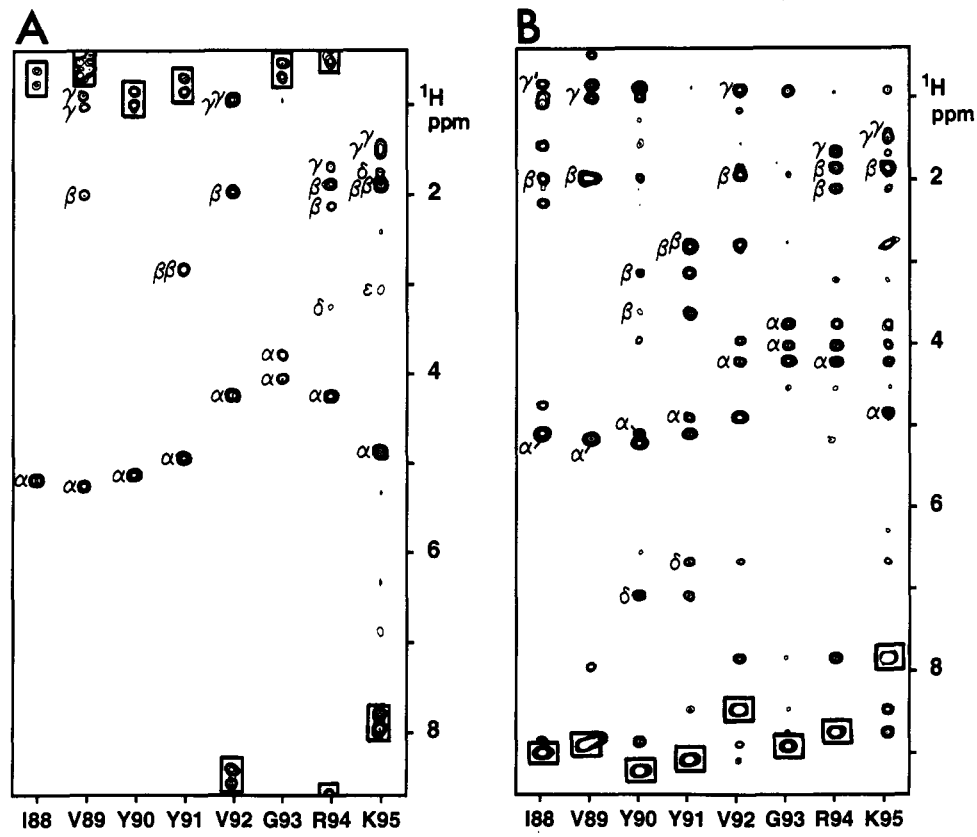

FIGURE 3: Sequential alignment of strips extracted from ^{15}N planes of 600-MHz 3D spectra. The ^{15}N chemical shifts of the diagonal peaks (delineated with boxes) are in Table III. (A) Strips from a 3D HOHAHA–HSQC spectrum acquired with a 58-ms mixing time. To increase resolution, a narrow sweep width was used in t_1 such that the diagonal peaks alias (fold) into the upfield region of the spectrum. The diagonal peaks are doublets because no decoupling of ^{15}N was used during t_1 evolution. (B) Strips from a 3D NOESY–HMQC spectrum acquired with a 90-ms mixing time. Both spectra were processed with 60°-shifted sinebell-squared digital filtering in t_1 and t_3 and 60°-shifted sinebell digital filtering in t_2.

poration in the protein are listed in Table II together with the levels of ^{15}N enrichment determined for samples labeled according to schemes 1 and 2. Comparison of Tables I and II shows that the levels of ^{15}N enrichment of most amino acids in the protein samples are approximately that expected given the enrichment in the media.

A notable exception to this statement occurs for Gly and Ser residues which are closely related metabolically and scramble ^{15}N between themselves. For example, in scheme 1, the levels of ^{15}N enrichment for Gly and Ser were 66% and 0% in the medium, whereas in TGF-β1 the corresponding enrichment levels were 41% and 20%. Comparison of Tables I and II shows that extensive scrambling of ^{15}N between Gly and Ser also occurred in schemes 2 and 3. The scrambling of the amide nitrogens of Gly and Ser is not surprising and is seen in *E. coli* expression systems.

In addition to Gly/Ser scrambling, ^{15}N scrambling of essential residues, Leu, Ile, and Val, took place in both schemes 1 and 3 to a surprising extent. For example, the scheme 1 medium contained [^{15}N]Leu but no [^{15}N]Ile or Val. However, the protein obtained in scheme 1 had Leu, Ile, and Val levels of enrichment in the ratio 3:2:1. We note that in scheme 2 these residues were not labeled in the medium but were labeled in the protein in approximately the same 3:2:1 ratio but at a 6-fold lower level of enrichment. This low level "feed-through" of label presumably occurred because the same cells were used to make all three protein samples, and a small amount of labeled material was carried over from scheme 1 into the scheme 2 medium. No carryover of label was seen from the scheme 2 to the scheme 3 preparation. We note that the total ^{15}N labeling of Leu, Val, and Ile residues in TGF-β1 samples derived from schemes 1 and 3 reflects the total amount of ^{15}N in the respective media. The same remark applies to the total level of Ser plus Gly ^{15}N labeling observed in each of the three specifically enriched protein samples.

All three protein samples exhibited low levels of ^{15}N enrichment (less than 10%) of the backbone amides of Ala, Gln, Glu, and Asp residues. In addition, in scheme 3, which contained ^{15}N-labeled Cys, the Met amide enrichment was ca. 8%.

Limited scrambling of labels between the side chain amides of Asn and Gln was also observed. In the medium for scheme 1 only the side chain of Asn was labeled with ^{15}N, but the side chain amides of both Asn and Gln in TGF-β1 were labeled in the ratio of Asn to Gln of 7:1. Due primarily to the feed-through of label discussed above, low levels of Asn and Gln side chain labeling were found for the protein obtained from scheme 2, with a ratio of Asn to Gln enrichment of 4:1. Scheme 3 exhibited a very low level (less than 5%) of ^{15}N-labeled Gln side chain enrichment (presumably due to scrambling from labeled backbone amides) and insignificant levels (ca. 1%) of labeled Asn.

In contrast with the scrambling observed for the ^{15}N labels, no evidence for scrambling of any of the various ^{13}C labels was found in the ^{1}H–^{13}C double resonance spectra of the three specifically ^{13}C enriched TGF-β1 samples. This result is in agreement with observations made on proteins labeled in *E. coli* which show that ^{13}C labels scramble to a much lower degree than ^{15}N labels. We did of course observe the feed-through of ^{13}C labels from scheme 1 to scheme 2 to the same extent found for the ^{15}N labels.

Signal Assignments, General Considerations. Approximately 100 major backbone amide signals are observed in the

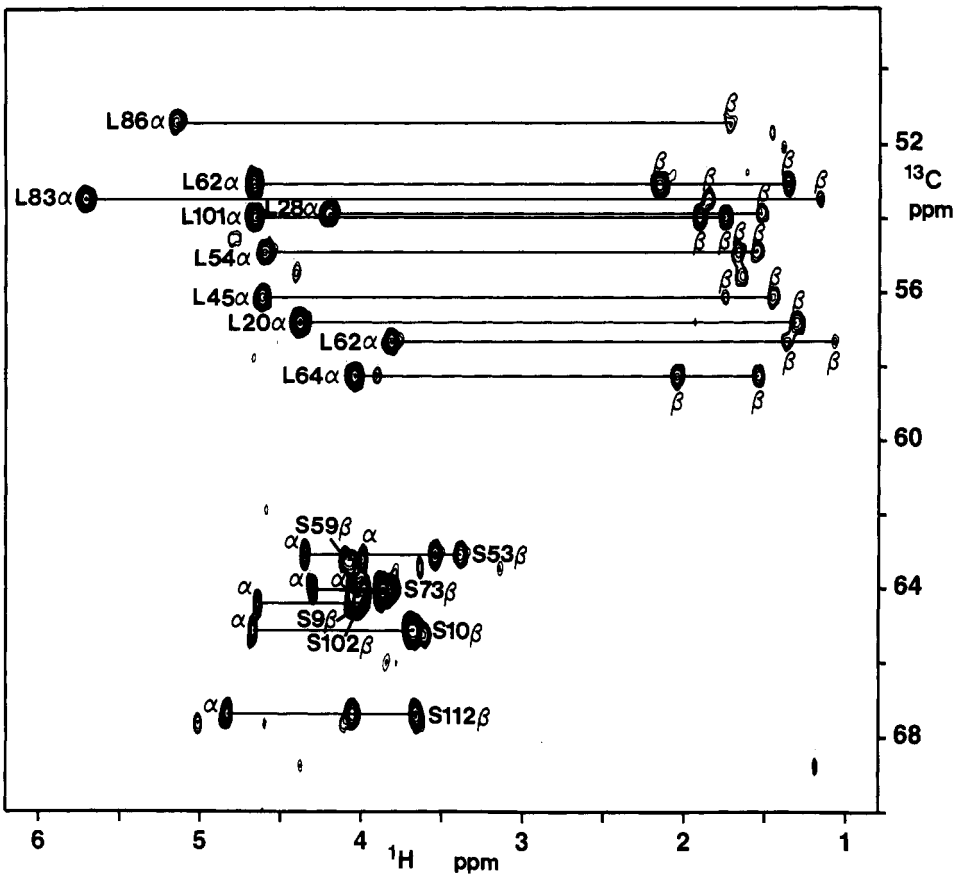

FIGURE 4: Region from ^1H–^{13}C HMQC–RELAY of selectively labeled TGF-β1 (scheme 3), 1 mM (dimer) in D_2O. Leu Cα carbons and Ser Cβ carbons are ^{13}C-enriched. The spectrum was acquired at 500 MHz with a 28-ms mixing time and with 512 complex points in t_2 and 128 complex points in t_1. The spectrum was processed with Lorentzian to Gaussian digital filtering in both dimensions and zero-filled once in t_2 and twice in t_1.

^1H–^{15}N HSQC spectrum of uniformly ^{15}N labeled TGF-β1, showing that the TGF-β1 dimer forms a 2-fold symmetric structure in solution. The possibility of a second structural form is suggested by the observation of a large number of minor signals, having intensities ca. 20% those of the major peaks. The ^{15}N linewidths, 6–8 Hz, are in the range expected for a 25-kDa protein at 45 °C, and ^1H and ^{15}N chemical shifts are well dispersed. These observations indicate that the protein has a well-defined structure in solution that could be elucidated by means of NMR techniques once sequential signal assignments were made.

In principle, sequential signal assignments of proteins can be obtained from homonuclear 2D COSY, HOHAHA, and NOESY experiments (Wüthrich, 1986, 1989; Clore & Gronenborn, 1987, 1989). However, complete signal assignments for proteins having M_r larger than 12K is extremely difficult using 2D homonuclear spectroscopy due to (a) resonance overlap and (b) small T_2 values that cause inefficient magnetization and/or coherence transfer. Resonance overlap is greatly reduced by recording 3D ^{15}N-edited spectra on uniformly ^{15}N-enriched proteins (Marion et al., 1989b; Driscoll et al., 1990). However, TGF-β1 presents several difficulties for obtaining sequential assignments solely on the basis of ^{15}N-edited spectra. One difficulty arises because the backbone amide proton rotating frame T_1s ($T_1\rho$s) are approximately 20 ms, as is expected for a 25-kDa protein; hence, transverse magnetization decays significantly during mixing periods in COSY/HOHAHA pulse sequences. This is a particular problem in the 3D ^{15}N-edited HOHAHA experiment where long mixing times are necessary to transfer magnetization from the amide proton to side chain protons. The problem of small HN–CαH and/or small CαH–CβH couplings is partially overcome using the 3D HNHB experiment, which relies on the $J_{N-H\beta}$ coupling, to correlate Hβ protons with intraresidue amide protons (Archer et al., 1991). The nine prolines in each TGF-β1 monomer pose a further problem for ^{15}N-directed assignments, because there is a break in the amide-directed sequential connectivities wherever a proline occurs in the sequence. A third difficulty in assigning TGF-β1 arises because it is a structurally symmetric homodimer; hence, one has the problem of distinguishing intra- from intermonomer NOE's. These assignment difficulties have been overcome by augmenting the information obtained from 3D spectra of the uniformly ^{15}N-enriched sample with information obtained from heteronuclear edited 2D spectra of the three specifically labeled samples. The sequential assignments were obtained using the general strategy of Wüthrich (1986) which consists of two basic steps. First, the NMR signals are assigned to specific types of amino acids, and second, these signals are linked to signals in adjacent residues in the amino acid sequence to obtain the sequential assignments.

Identification of Amino Acid Spin Systems. In order to link the amide proton signal of each amino acid residue in TGF-β1 to its corresponding aliphatic proton signals, 3D HOHAHA–HSQC spectra were recorded at 600 MHz with mixing times of 39 ms at 37 °C and 58 and 80 ms at 45 °C. Spectra were obtained at two temperatures in order to locate signals close to the water signal. A total of 97 (of a possible 103) HN–Hα correlations were identified in these spectra. Because of the short T_2 values of the TGF-β1 backbone and β protons, only about 100 HN–Hβ connectivities (out of a possible 171) were observed in the HOHAHA–HSQC spectra. Approximately 70 of the HN–Hβ correlations observed in the HOHAHA spectra were also identified in the 3D HNHB

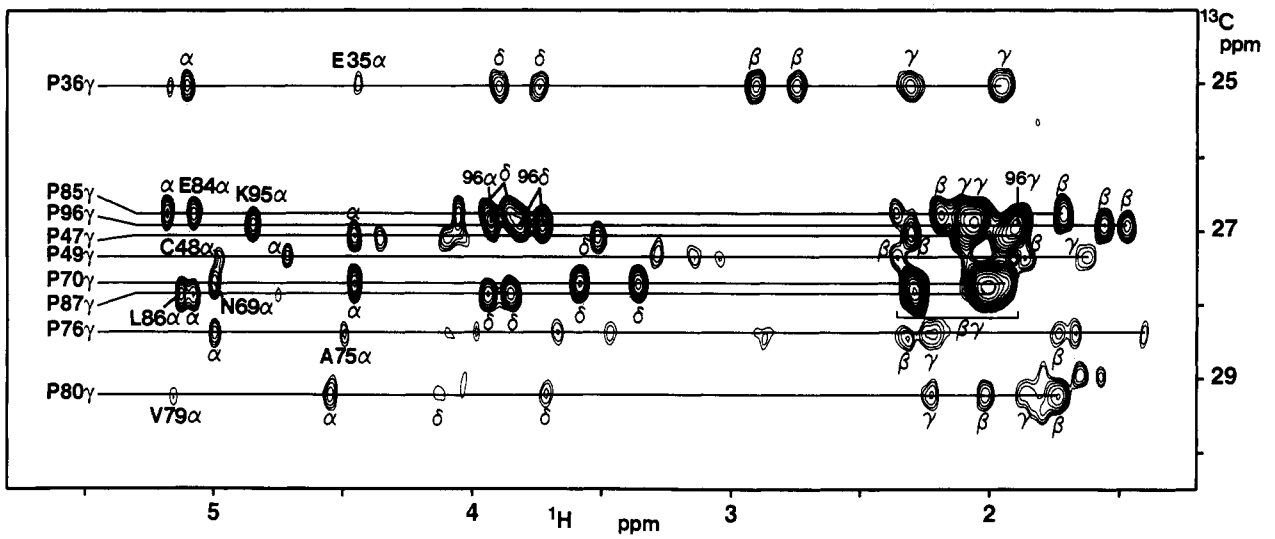

FIGURE 5: Proline $^{13}C\gamma$ region from ^1H–^{13}C HSQC–NOESY of selectively labeled TGF-β1 (scheme 2), 0.68 mM (dimer) in D_2O. Intraresidue NOE cross-peaks are indicated with Greek symbols. The protons in preceding residues are also indicated. The spectrum was acquired at 500 MHz with a 200-ms mixing time, 512 complex points in t_2, and 184 complex points in t_1. The spectrum was processed with Lorentzian to Gaussian digital filtering and zero-filled once in both dimensions.

FIGURE 6: Summary of sequential connectivities for TGF-β1. Sequential connectivities $H\alpha_i$–HN_{i+1}, $H\beta_i$–HN_{i+1}, and HN_i–HN_{i+1} are designated αN, βN, and NN, respectively. The αN and βN connectivities were observed in the 3D NOESY–HMQC. The NN connectivities were observed in the 3D NOESY–HMQC and/or the 3D ^{15}N–^{15}N–^1H HMQC–NOESY–HMQC. The J_{CN} connectivities are from splitting patterns in 2D ^1H–^{15}N HSQC spectra of selectively ^{13}C- and ^{15}N-labeled samples of TGF-β1. The footnote *a* indicates that for proline residues, the α_{i-1}–δ_i NOE correlations from a 2D NOESY spectrum were used. Narrow lines indicate that the proton chemical shifts of H_i and H_{i-1} were degenerate.

spectrum, in which 20 additional HN–Hβ correlations, not observed in the 3D HOHAHA spectra, were also found. Approximately 30 additional Hα–Hβ connectivities were identified in uncrowded regions of 2D ^1H–^1H HOHAHA spectra.

Although the difficulty of observing the HN–aliphatic correlations significantly increases as the number of intervening bonds increases, we were able to observe most of the possible HN–$C\gamma H_3$ correlations for Val and Ile residues in the 3D HOHAHA–HSQC spectrum (Figure 3A), presumably because (a) the methyl protons in the spin I = 1/2 manifolds have large T_2 values and (b) the Hβ–Hγ coupling is sizable, ca. 7 Hz. In addition, we were able to trace out virtually the entire spin systems for most of the Lys and Arg residues because their side chains are often flexible and have large proton T_2 values. We now discuss in more detail the spin system assignments that were obtained by combining the information derived from the spectra of the uniformly and specifically labeled samples.

Identification of Amino Acid Spin Systems. Ala, Ser, and Gly. The HN–Hα–Hβ connectivities of six of the seven Ala residues were observed in both the 3D HNHB and the 3D HOHAHA–HSQC spectra. The single intense correlation to the upfield Ala methyl protons indicates the presence of the Ala spin system. The Ala spin system assignments derived from the 3D spectra were confirmed by 2D ^1H–^{13}C HMQC–HOHAHA spectra of samples that were labeled with either [2-^{13}C]Ala or [3-^{13}C]Ala. These spectra also provided spin system assignments for the aliphatic signals of the N-terminal Ala.

The HN–Hα–Hβ spin systems of seven of the eight Ser residues in TGF-β1 were identified in a straightforward fashion from (a) the 3D HOHAHA–HSQC spectra of uniformly ^{15}N enriched TGF-β1, (b) the 2D ^1H–^{15}N HSQC spectrum of a sample containing [^{15}N]Ser, and (c) a 2D HSQC spectrum and a HMQC–RELAY spectrum (Figure 4) of the sample labeled with [3-^{13}C]Ser. The Hα–Hβ spin system of the remaining serine (Ser53) was identified in 2D HOHAHA spectra of nonlabeled protein and 2D HSQC and HSQC–RELAY of protein labeled with [3-^{13}C]Ser. The HN–Hα connectivity has not been identified for Ser53. After assigning the Ser spin systems, four of the five Gly spin systems were identified using the 3D HOHAHA–HSQC data and a 2D HSQC spectrum of a sample labeled with [^{15}N]Gly. The spin system of Gly46 has not been identified.

Ile, Leu, Thr, and Val. The complete set of connectivities, HN–Hα–Hβ–Hγ, was observed for one of the three Thr spin systems in the 3D HOHAHA–HSQC spectra. This assignment was confirmed, and the Hβ signals of the other two Thr residues were identified on the basis of the observation of the three expected Thr $C\beta H\beta$ correlations in their unique locations in the natural abundance 2D ^1H–^{13}C HMQC spectrum of TGF-β1. The spin system identifications were extended by observation of Hα–Hβ, Hβ–Hγ, and $C\gamma H\gamma$ correlations in

2D HOHAHA and HMQC spectra. Linking the Hα–Hβ–Hγ and HN–Hα–Hβ connectivities completed the assignments.

The connectivities linking six of the seven Val HN protons to their backbone and side chain aliphatic protons were seen in the 3D HOHAHA–HSQC spectra (for example, see V89 and V92 in Figure 3A). The corresponding connectivities linking the Val CαH to the Val side chain protons were also observed in the natural abundance 2D HOHAHA spectrum. The Hβ–Hγ correlations for the seventh Val were readily identified by comparing the 2D HOHAHA spectrum of nonlabeled protein with the spectrum of the protein labeled with deuterated Leu. The Hβ–Hγ correlation was linked to the HN–Hα correlation via the chemical shift of the Hβ proton as well as correlation between the amide and the side chain protons in the 3D NOESY–HMQC spectrum. The Val spin system assignments were confirmed by the observation of the seven expected ^{15}N–HN connectivities in the HSQC spectrum of TGF-β1 labeled with [^{15}N]Val.

Only four of the expected five Ile ^{15}N–HN connectivities were found in the ^{1}H–^{15}N HSQC spectrum of TGF-β1 labeled with [^{15}N]Ile. In the 3D HOHAHA–HSQC spectra, three of the four Ile amide protons showed connectivities to the Cγ′H$_3$ and CβH protons while four amides showed connectivities to the Hα protons. The fourth HN–Hβ connectivity was present in the HNHB spectrum. These data together with the four Ile Hα–Hβ–Hγ′ connectivities identified in the natural abundance 2D HOHAHA spectrum assign four of the five Ile spin systems. The spin system of Ile 51 has not been identified.

All ten Leu ^{15}N-HN correlations were identified in the ^{1}H–^{15}N HSQC spectrum of TGF-β1 labeled with [^{15}N]Leu. All Leu amide protons were linked to their α protons, five were linked to both of their Hβ protons, and three were linked to one Hβ proton by correlations seen in the 3D HOHAHA–HSQC spectra. All Leu Hα protons and 18 Hβ protons were identified from 2D HSQC and 2D HMQC–RELAY spectra of [2-^{13}C]Leu-labeled TGF-β1 (Figure 4). All Leu methyl and γ protons were identified by comparing a HOHAHA spectrum of natural abundance TGF-β1 with the identical spectrum recorded for a sample in which the Leu methyl and methylene groups were deuterated. Four Leu side chains were completely assigned from correlations in the 2D HOHAHA spectrum. A ^{13}C-edited NOESY spectrum of [2-^{13}C]Leu-labeled TGF-β1 linked the α protons of the remaining six leucine residues to their γ and methyl protons.

Arg, Asn, Gln, and Lys. The backbone amide HN–Hα–Hβ–Hγ–Hδ and side chain amide HN–Hδ–Hγ connectivities for three of the four Arg spin systems were observed in the 3D HOHAHA–HSQC spectra (Figure 3A). In addition, all four Arg ^{15}N–HN correlations were observed in the ^{1}H–^{15}N HSQC spectrum of Arg, ^{15}N enriched at the backbone amide. Connectivities from the backbone amide all the way to the Hε proton were observed in the 3D HOHAHA–HSQC spectrum acquired with an 80-ms mixing time for five of the eight Lys (for an example, see K95 in Figure 3A). Additional side chain proton assignments were determined from Hε–Hδ–Hγ–Hβ correlations in the natural abundance 2D HOHAHA. All eight Lys ^{15}N–HN correlations were identified in the ^{1}H–^{15}N HSQC spectrum of TGF-β1 labeled with [^{15}N]Lys.

Backbone amide–Hβ correlations were observed in the 3D HOHAHA–HSQC spectra for nine of the eleven Gln/Asn residues. Additional Hβ signals were identified in the ^{1}H–^{1}H HOHAHA spectrum. These signals were linked to the side chain amide protons of all five Gln and five of the six Asn residues by NOE correlations in the 3D NOESY–HMQC spectrum. The side chain amide resonance assignments were confirmed by the ^{15}N–HN correlations seen in the HSQC spectrum of TGF-β1 labeled with ^{15}N on the side chain amides of Asn and Gln.

Table III: Predicted Relative Intensities of Amide Splitting Patterns Due to Scalar Coupling of the Backbone ^{15}N Amide with the ^{13}C-Enriched Carbonyl of the Preceding Residue[a]

splitting pattern intensities	scheme 1	scheme 2	scheme 3
1:1	Y21, G29, S101	H34, W52	C48, C77, I88
1.5:1	Y90, G93	H40, S59, Y91	C7, I22, I51, S59, V92
2:1	K37, G70, L86, I88, K97	W32, G38	V61, V98, C111
3:1			V89, V106

[a] Levels of ^{13}C enrichment: for scheme 1 L, 100%; V, 80%; P, 67%; for scheme 2 I, 100%; Y, 80%; K, 67%; for scheme 3 P, 100%, Y, 80%; K, 67%; I, 50%.

His, Phe, Trp, and Tyr. Initial assignments of aromatic ring protons were derived from 2D HOHAHA and 2D DQF-COSY spectra (Wüthrich, 1986). The ring protons were linked to Hβ–Hα–HN spin systems from Hδ–Hβ and Hδ–Hα correlations in the 2D NOESY of nonlabeled protein and from correlations in the 2D NOESY of TGF-β1 labeled with [*ring*-D$_5$]Phe. In addition, correlations between the amide and Hδ protons were observed in the 3D NOESY–HMQC spectrum (Figure 3B). The assignments were confirmed and extended using information from the natural abundance ^{1}H–^{13}C HMQC and from 2D heteronuclear spectra of selectively enriched samples.

In the natural abundance ^{1}H–^{13}C HMQC spectrum, Trp Hζ2 protons were identified by their characteristic upfield carbon chemical shifts, and connectivities to Trp Hε3, Hζ3, and Hη2 protons were identified in 2D HOHAHA spectra. Assignments of the backbone amides of all three Trp residues were confirmed by ^{15}N–HN correlations in the HSQC spectrum of TGF-β1 labeled with [^{15}N]Trp.

The Hε protons for seven of the eight Tyr residues were observed in the natural abundance HMQC spectrum, and the Tyr Hδ protons were identified by the strong Hε–Hδ correlations in the 2D HOHAHA and 2D DQF-COSY spectra. These assignments were confirmed by HSQC and HSQC–RELAY spectra of TGF-β1 labeled with [*phenol*-3,5-^{13}C$_2$]Tyr. ^{15}N–HN correlations were identified for all eight Tyr in the HSQC spectrum of TGF-β1 labeled with [^{15}N]Tyr.

The ^{15}N–HN correlations for all three Phe residues were observed in the HSQC of TGF-β1 labeled with [^{15}N]Phe. The Phe HN–Hδ NOE correlations were present in the 3D NOESY–HMQC, but because all three Phe residues have similar Hδ chemical shifts (7.17 ± 0.02 ppm), the Phe HN–Hβ connectivities were used to distinguish the spin systems. The Phe Hβ protons were readily identified in the HMQC spectrum of TGF-β1 labeled with [3-^{13}C]Phe. HMQC–RELAY, ^{13}C-edited NOESY, and HSQC–NOESY spectra of TGF-β1 labeled with [3-^{13}C]Phe were used to correlate the Hβ protons with Hα and Hδ protons in order to complete the Phe assignments.

All three His Hε1–Hδ2 correlations were observed in 2D HOHAHA spectra. For two of the three His, the Hε1 protons were readily identified in the natural abundance HMQC by their characteristic downfield carbon and proton chemical shifts. Assignments of the ring protons were confirmed by HMQC and HMQC–RELAY spectra of TGF-β1 labeled with [*ring*-2-^{13}C]His. The His amide assignments were confirmed

Table IV: Chemical Shifts (in Parts per Million) for TGF-β1 at 45 °C

residue	^{15}N	HN	Hα	Hβ	other H, N
A1			3.99	1.46	
L2	122.1	8.54	4.64	2.11, 1.42	Hγ 1.56; Hδ 0.83, 0.72
D3	119.2	7.92	5.17	3.11, 2.60	
T4	111.0	7.55	3.10	3.74	Hγ 1.01
N5	119.0	7.55	4.47	2.93, 2.83	HNγ 7.67, 6.87; Nγ 113.2
Y6	119.4	8.07	4.50	3.15, 3.44	Hδ 7.26; Hϵ 6.90
C7	116.3	8.88	4.28	3.50, 2.73	
F8	116.7	8.31	5.02	3.43, 3.40	Hδ 7.15; Hϵ 6.99
S9	113.5	7.44	4.65	4.07, 4.02	
S10	116.0	7.53	4.66	3.75, 3.72	
T11	115.7	8.26	4.39	4.29	Hγ 1.22
E12	124.7	8.30	4.42	2.57, 2.10	
K13	127.1	8.47	4.22	1.76, 1.43	Hγ 2.46; Hϵ 3.01
N14	115.7	7.98	4.43	1.69	
C15	111.3	8.22	4.15	3.27, 2.31	
C16	124.5	7.83	4.67	3.17, 3.12	
V17	120.9	7.71	4.28	1.60	Hγ 0.84, 0.79
R18	124.9	9.25	4.60		Hγ 1.48;[a] Hδ 3.00;[a] NHδ 6.07;[a] Nδ 84.4[a]
Q19	120.4	8.35	4.44	1.80	Hγ 1.99; HNδ 6.76, 6.48; Nδ 110.5
L20	120.8	7.85	4.42	1.33	Hγ 0.88; Hδ 0.59, –0.02
Y21	129.9	8.59	4.57	2.75, 2.39	Hδ 6.06; Hϵ 6.46
I22	127.0	7.71	3.92	1.08	Hγ 0.84, 0.48; Hδ 0.23; Hγm 0.42
D23	128.4	9.57	4.57	2.85, 2.50	
F24	125.3	8.20	3.99	3.30, 2.94	Hδ 7.18; Hϵ 6.93; Hζ 6.62
R25	119.3	8.67	4.05	2.00, 1.94	Hγ 1.81, 1.68; Hδ 3.34; HNδ 7.52; Nδ 85.1
K26	119.4	8.88	4.00	1.76, 1.65	Hγ 1.54, 1.39; Hϵ 3.09
D27	112.6	8.27	4.60	2.97, 2.48	
L28	110.0	6.32	4.25	1.54, 0.49	Hγ 1.17; Hδ 0.58, 0.09
G29	105.3	7.07	4.27, 4.01		
W30	122.2	7.99	5.19	3.07, 2.64	Hϵ_3 6.98; Hζ_3 6.74; Hη_2 7.17; Hζ_2 7.51
K31	121.1	8.87	4.45	1.87, 2.13	Hγ 1.55, 1.46; Hδ 1.82; Hϵ 3.14
W32	114.9	7.16	4.64	3.73	Hϵ_3 7.15; Hζ_3 6.08; Hη_2 6.56; Hζ_2 6.33
I33	122.8	6.86	3.59	1.01	Hγ –0.58; Hδ –0.52; Hγm 0.50
H34	127.0	8.97	4.72	3.31, 3.13	Hϵ_1 8.45; Hδ_2 7.19
E35	117.6	7.92	4.47	1.83, 2.11	Hγ 2.20, 2.28
P36			5.11	2.90, 2.76	Hγ 2.32, 1.96; Hδ 3.91, 3.76
K37	116.3	8.75	4.44	2.14, 1.96	Hγ 1.71, 1.66; Hδ 1.80; Hϵ 3.18
G38	105.9	7.69	4.02, 3.31		
Y39	116.6	8.31	4.88	3.00, 2.98	Hδ 6.51
H40	120.4	8.92	4.82	3.40, 3.38	Hϵ_1 8.64; Hδ2 7.51
A41	129.2	6.58	4.15	1.18	
N42	102.3	7.70	4.99	3.80, 3.08	HNγ 7.28, 6.73; Nγ 110.4
F43	111.8	8.60	5.02	3.19, 3.18	Hδ 7.17; Hϵ 7.33
C44	119.9	9.26	4.99	3.33, 2.64	
L45	123.0	7.91	4.63	1.76, 1.48	Hγ 1.74; Hδ 1.29, 1.07
G46					
P47			4.48	2.32, 1.97	Hδ 3.53
C48	117.1	8.39	5.00	2.93, 2.87	
P49			4.72	2.35, 2.09	Hγ 1.65; Hδ 3.30,[a] 3.16[a]
Y50	118.5	8.44	4.18	2.96, 2.88	Hδ 7.18; Hϵ 6.93
I51					
W52	123.4	7.81	4.92	3.44, 3.06	Hδ_1 7.41;[a] Hϵ_3 7.56; Hζ_3 7.27; Hη_2 7.44; Hζ_2 7.42
S53			4.36	3.55, 3.40	
L54	124.8	7.79	4.59	1.70, 1.57	Hγ 1.64; Hδ 0.93, 0.75
D55	120.1	8.47	4.58	2.95, 2.87	
T56	106.0	7.74	4.79	4.53	Hγ 1.39
Q57	122.3	9.08	4.24	2.41, 2.37	Hγ 2.86, 2.61; NHδ 7.22, 7.07; Nδ 109.6
Y58	119.0	7.75	4.20	3.60, 2.79	Hδ 7.14, Hϵ 6.83
S59	114.9	8.14	3.99	4.09, 4.06	
K60	121.1	7.57	4.13	2.28, 2.25	Hγ 2.16
V61	120.2	8.15	3.49	2.40	Hγ 1.21, 0.88
L62	120.4	8.60	3.83	1.36, 1.08	Hγ 1.31; Hδ 0.75, 0.64
A63	121.1	8.40	4.30	1.66	
L64	121.9	8.33	4.06	2.05, 1.56	Hγ 1.78; Hδ 0.97, 0.82
Y65	123.7	9.07	3.99	3.35, 2.73	Hδ 6.78; Hϵ 6.68
N66	119.5	8.81	4.16	3.29, 2.80	HNγ 7.50, 6.97; Nγ 111.5
Q67	117.1	7.83	4.05	2.26, 2.13	Hγ 2.48; NHδ 7.25, 6.49; Nδ 109.4
H68	113.1	7.49	4.50	3.47, 3.23	Hϵ_1 8.30; Hδ_2 7.36
N69	118.9	7.57	5.03	2.26, 2.14	NHγ 7.04, 6.12; Nγ 115.6
P70			4.47	2.31, 1.95	Hδ 3.60, 3.37
G71	110.5	8.63	4.09, 3.89		
A72	123.2	7.89	4.38	1.61	
S73	114.5	8.30	4.30	3.97, 3.81	
A74	123.2	6.99	4.27	0.68	
A75	123.6	7.89	4.54	1.42	
P76			5.00	2.32, 1.73	Hγ 2.23; Hδ 3.69,[a] 3.49[a]
C77	121.8	9.58	4.87	3.33, 3.25	

Table IV (Continued)

residue	^{15}N	HN	Hα	Hβ	other H, N
C78	126.4	8.29	5.23	3.45, 2.96	
V79	122.0	9.47	5.16	2.48	Hγ 1.26, 1.02
P80			4.55	2.04, 1.75	Hγ 2.25, 1.86; Hδ 4.15, 3.73
Q81	123.5	9.02	4.48	2.25, 1.96	Hγ 1.96, 1.41; HNδ 7.36, 6.78; Nδ 112.1
A82	120.7	7.37	4.69	1.41	
L83	122.5	8.35	5.70	1.87, 1.21	Hγ 1.47; Hδ 0.93, 0.51
E84	120.9	9.59	5.12	2.41, 1.91	
P85			5.19	2.19, 1.72	Hγ 2.11, 2.05; Hδ 3.97, 3.86
L86	118.4	8.59	5.16	1.74	Hγ 1.52; Hδ 1.03, 0.23
P87			5.09	2.28, 1.99	Hδ 3.95, 3.86
I88	117.2	8.97	5.17	1.99	Hγm 0.83
V89	119.5	8.86	5.24	1.99	Hγ 1.02, 0.89
Y90	126.7	9.22	5.12	3.64, 3.17	Hδ 7.10; Hε 6.60
Y91	119.8	9.09	4.93	2.85, 2.77	Hδ 6.69; Hε 6.29
V92	122.9	8.44	4.23	1.96	Hγ 0.95, 0.90
G93	118.3	8.93	4.04, 3.77		
R94	121.1	8.75	4.24	2.12, 1.87	Hγ 1.68, 1.65; Hδ 3.24; HNδ 7.14; Nδ 84.9
K95	120.9	7.82	4.86	1.88, 1.86	Hγ 1.52; 1.44; Hδ 1.77; Hε 3.07
P96			3.95	1.56, 1.46	Hγ 1.91; Hδ 3.83, 3.74
K97	124.4	8.86	4.42	1.30, 1.00	Hγ 1.18, 1.15; Hδ 1.59; Hε 2.97
V98	121.2	7.94	5.15	2.00	Hγ 0.98, 0.97
E99	124.5	8.80	4.84	2.29, 2.15	Hγ 2.10
Q100	120.5	8.62	4.77	2.00, 1.87	Hγ 2.32, 2.13; HNδ 7.31, 6.66; Nδ 109.7
L101	127.4	8.73	4.69	1.94, 1.76	Hγ 1.79; Hδ 1.05, 1.01
S102	119.9	8.44	4.08	3.91	
N103	115.1	8.69	4.04	3.20, 3.02	HNγ 7.62, 6.95; Nγ 112.8
M104	112.8	7.58	5.17	2.03	Hδ 2.09
I105	120.5	8.52	4.92	1.70	Hδ 1.13;[a] Hγm 1.02
V106	126.4	9.04	4.06	1.83	Hγ 0.84, 0.47
R107	126.5	8.65	4.61	1.79, 1.72	Hγ 1.69; Hδ 3.26; HNδ 7.41; Nδ 84.4
S108	109.7	7.73	5.03	3.85, 3.78	
C109	119.2	8.42	5.36	3.01, 2.32	
K110	118.0	9.83	4.78		
C111	118.4	8.27	5.35	3.50, 2.71	
S112	124.5	8.98	4.81	4.05, 3.71	

[a] Tentative assignment.

by ^{15}N–HN correlations in 2D ^{1}H–^{15}N HSQC spectra of TGF-β1 labeled with [^{15}N]His.

Cys and Met. ^{15}N–HN correlations for all nine Cys residues were observed in the HSQC spectrum of TGF-β1 labeled with [^{15}N]Cys. Eight Cys HN–Hα correlations were observed in the 3D ^{15}N–^{1}H HOHAHA–HSQC spectra of uniformly ^{15}N-enriched TGF-β1, while only six (out of a possible 18) weak HN–Hβ correlations were observed. Conclusive identification of the Cys Hα–Hβ linkages was provided by the connectivities observed in the HMQC and HMQC–RELAY spectra of TGF-β1 labeled with [3-^{13}C]Cys.

The methyl carbon of the single Met in TGF-β1 was assigned in the HMQC spectrum of TGF-β1 labeled with [*methyl*-^{13}C]Met, and the Met Hα–CH_3 cross-peak observed in the ^{13}C edited HMQC–NOESY spectrum of this sample identified the Hα signal. The spin system assignments were extended by observations of Hα–Hβ and HN–Hα–Hβ correlations in 2D HOHAHA and 3D HOHAHA–HSQC spectra, respectively. This assignment was confirmed by sequential NOE's in the 3D NOESY–HMQC.

Pro. Conclusive assignments of Pro Hα and Hβ signals were derived from ^{13}C-Hα and ^{13}C-Hα–Hβ correlations observed in 2D HMQC and ^{13}C-edited HMQC–RELAY or HMQC–NOESY spectra of TGF-β1 labeled with [^{13}Cα]Pro. The proline Hγ and Hδ signals were assigned in a similar manner using the sample of TGF-β1 labeled with [γ^{13}C]Pro. The use of a long mixing time, 200 ms, in the ^{1}H–^{13}C HSQC–NOESY spectrum facilitated tracing connectivities from Hγ to Hβ to Hα (Figure 5).

Asp and Glu. The remaining HN–Hα–Hβ and HN–Hα–Hβ–Hγ correlations in the 3D HOHAHA–HSQC and NOESY–HMQC spectra that had not been assigned to other residue types were assigned, by elimination, to Asp and Glu, respectively. These HN signal assignments were in agreement with the positions of the weak ^{15}N–HN correlations observed in spectra of all three selectively labeled samples due to scrambling of ^{15}N into metabolically active amino acids such as Asp and Glu.

Sequential Assignments. Approximately 80% of the spin systems of TGF-β1 were uniquely identified on the basis of the observations just described. Of the remaining spin systems, all but six were assigned to two types of amino acid residues. This extensive set of spin system assignments together with the large number of dαN, dβN, and dNN sequential connectivities, observed in the 3D NOESY–HMQC and HMQC–NOESY–HMQC spectra (Figure 6), provided most of the information needed to make the sequential assignments. Additional crucial connectivities, linking proline Hδ signals to the Hα signals of the preceding residues, were observed in the long mixing time ^{13}C-edited NOESY spectrum of TGF-β1 labeled with [Cγ]Pro (Figure 5) for all proline residues except Pro36. A strong Glu35–Pro36 Hα–Hα connectivity was however observed in the ^{13}C-edited NOESY spectrum of TFG-β1 labeled with [Cα]Pro.

The ^{15}N and ^{1}H sequential assignments of TGF-β1 are listed in Figure 6. Briefly, the segments of the sequence were assigned in the following manner. The segments containing medium to long strings of dNN connectivities (e.g., Y58–N69) were identified first. Next, sequential segments in the C-terminal half of the molecule were identified by an extensive network of dαN and dβN connectivities (Figure 6) observed for residues in this portion of the protein. A similar network of connectivities provided assignments for residues 2–18 and 31–45 in the N-terminal half of the molecule. Segments including residues 19–24 and 46–54 were linked by the fewest connectivities and were assigned last.

We emphasize that, in addition to sequential NOE's, the $^{13}C'$–^{15}N splitting of the signals in the ^{1}H–^{15}N HSQC spectra of the selectively labeled samples were used to obtain sequential assignments of ^{15}N–HN correlations (Figure 6). In the case of the symmetric TGF-β1 homodimer, it is particularly important to confirm the NOE-based assignments using the sequential assignments derived from the J_{NC} couplings (Kainosho & Tsuji, 1982) because of the possibility of confusing an intra- with an intermonomer NOE. The use of multiple selective ^{13}C labeling, with different levels of ^{13}C enrichment, to obtain sequential resonance assignments of proteins has been described by Ikura et al. (1990c), who show that the ^{15}N–HN correlations will appear as doublets with relative peak intensities of 1:1, 1.5:1, 2:1, and 3:1 for 100%, 80%, 67%, and 50% ^{13}C incorporation, respectively (Ikura et al., 1990c). The calculated relative intensities for the selectively labeled TGF-β1 samples (Table III) were in agreement with the observed intensities in the ^{1}H–^{15}N HSQC spectra of the selectively labeled samples, with the following exceptions. In scheme 1, the ^{15}N–HN correlations of residues that were preceded by Pro were broad doublets having an intensity ratio of ca. 1.2:1, rather than 2:1 as predicted for 67% ^{13}C enrichment at the Pro carbonyl carbon. This observation is explained by the fact that, in scheme 1, the other 33% of proline was ^{13}C-labeled at the Cα carbon. Hence, in scheme 1, the amide ^{15}N signals of residues that are preceded by Pro residues are split by either the $^{1}J_{C'N}$ coupling to the Pro C′ or the $2J_{C\alpha N}$ ($\sim$7 Hz) coupling to the Pro Cα. The second exception to the predicted splitting intensities was observed for the Leu signals observed in the ^{1}H–^{15}N HSQC spectrum recorded for the scheme 3 sample. Although only [$^{13}C\alpha$]Leu was included in the scheme 3 growth medium, the protein contained [^{15}N]Leu due to scrambling from [^{15}N]Val and [^{15}N]Ile. In the HSQC spectrum, nearly all Leu ^{15}N–HN correlations were either broadened or exhibited small splittings because of intraresidue $^{1}J_{C\alpha N}$ ($\sim$11 Hz) coupling.

CONCLUSIONS

Using information from spectra of uniformly ^{15}N-enriched and a small number of selectively ^{13}C- and ^{15}N-labeled samples, 96% of the backbone protons and amide nitrogens of TGF-β1 have been assigned. In addition 87% of the side chain protons and side chain nitrogens for 13 of the 15 Asn, Gln, and Arg residues have been assigned. We have demonstrated the feasibility of using a CHO expression system to obtain either uniform or selective labeling of a eukaryotic protein. It has also been shown that one can simultaneously selectively label a large number of different amino acids with ^{2}H, ^{13}C, and ^{15}N and still retain selectivity. This has the desirable practical consequence that only a small number of labeled protein samples are required. Although some of the side chain labels (e.g., [3-^{13}C]Ala) provided redundant assignment information, the selectively labeled side chains provided additional NOE's that yielded information about TGF-β1 secondary and tertiary structure.

^{15}N scrambling such as was observed for [^{15}N]Ser and [^{15}N]-Gly complicated signal assignments but also provided additional information once the pattern of scrambling was recognized. Likewise, feed-through of ^{15}N and ^{13}C labels between schemes provided additional information. This feed-through between schemes could have been reduced by additional perfusion of the bioreactor with unlabeled medium. On the basis of our experience with TGF-β1, ^{13}C scrambling does not appear to be a problem in CHO cells. The protocol described herein for isotopic enrichment and sequential NMR assignments should be generally applicable to other eukaryotic proteins which, like TGF-β1, are difficult to express and fold in bacterial systems. This development should increase the number of proteins amenable to investigation by NMR spectroscopy. In the present case, the assignments obtained open the way to the study of the solution structure of TGF-β1 as described in the following paper.

ACKNOWLEDGMENT

We gratefully acknowledge Drs. D. Garrett, S. Grzesiek, M. Ikura, L. Kay, and D. Marion and F. Delaglio for providing computer software. We also thank R. Tschudin for expert technical support.

SUPPLEMENTARY MATERIAL AVAILABLE

Tables SI and SII, listing the amounts of labeled and unlabeled amino acids used in the CHO cell growth medium, and Figures S1 and S2, showing the ^{1}H–^{15}N HSQC spectra of Schemes I and II (10 pages). Ordering information is given on any current masthead page.

REFERENCES

Archer, S. J., Ikura, M., Torchia, D. A., & Bax, A. (1991) *J. Magn. Reson. 95*, 636–641.
Baldisseri, D. M., Torchia, D. A., Poole, L. B., & Gerlt, J. A. (1991) *Biochemistry 30*, 3628–3633.
Bax, A., & Davis, G. (1985) *J. Magn. Reson. 65*, 355–360.
Bax, A., & Weiss, M. A. (1987) *J. Magn. Reson. 71*, 571.
Bax, A., Sklenar, V., Clore, G. M., & Gronenborn, A. M. (1987) *J. Am. Chem. Soc. 109*, 6511–6513.
Bax, A., Ikura, M., Kay, L. E., Torchia, D. A., & Tschudin, R. (1990) *J. Magn. Reson. 86*, 304–318.
Bodenhausen, G., & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185–189.
Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.
Caltabiano, M. M., Tsang, M. L.-S., Weatherbee, J. A., Lucas, R., Sathe, G., Sutton, J., Johnson, D., & Bergsma, D. (1989) *Gene 85*, 479–488.
Clore, G. M., & Gronenborn, A. M. (1987) *Protein Eng. 1*, 275–288.
Clore, G. M., & Gronenborn, A. M. (1989) *Crit. Rev. Biochem. Mol. Biol. 24*, 479–564.
Derynck, R., Jarrett, J. A., Chen, E. Y., Eaton, D. H., Bell, J. R., Assoian, R. K., Roberts, A. B., Sporn, M. B., & Goeddel, D. V. (1985) *Nature 316*, 701–705.
Driscoll, P. C., Gronenborn, A. M., Wingfield, P. T., & Clore, G. M. (1990) *Biochemistry 29*, 3542–3556.
Frenkiel, T., Bauer, C., Carr, M. D., Birdsall, B., & Feeney, J. (1990), *J. Magn. Reson. 90*, 420–425.
Garrett, D. S., Powers, R., Gronenborn, A. M., & Clore, G. M. (1991) *J. Magn. Reson. 95*, 214–220.
Gentry, L. E., Webb, N. R., Lim, G. J., Brunner, A. M., Ranchalis, J. E., Twardzik, D. R., Lioubin, M. N., Marquardt, H., & Purchio, A. F. (1987) *Mol. Cell. Biol. 7*, 3418–3427.
Gentry, L. E., Lioubin, M. N., Purchio, A. F., & Marquardt, H. (1988) *Mol. Cell. Biol. 8*, 4162–4168.
Grzesiek, S., Döbeli, H., Gentz, R., Garotta, G., Labhardt, A. M., & Bax, A. (1992) *Biochemistry 31*, 8180–8190.
Ikura, M., Kay, L. E., & Bax, A. (1990a) *Biochemistry 29*, 4659–4667.
Ikura, M., Bax, A., Clore, G. M., & Gronenborn, A. M. (1990b) *J. Am. Chem. Soc. 112*, 9020–9022.
Ikura, M., Krinks, M., Torchia, D. A., & Bax, A. (1990c) *FEBS Lett. 266*, 155–158.
Ikura, M., Kay, L. E., Krinks, M., & Bax, A. (1991) *Biochemistry 30*, 5498–5504.
Kainosho, M., & Tsuji, T. (1982) *Biochemistry 21*, 6273–6279.
Kay, L. E., Marion, D., & Bax, A. (1989) *J. Magn. Reson. 84*, 72–84.

Kay, L. E., Ikura, M., Tschudin, R., & Bax, A. (1990) *J. Magn. Reson. 89*, 496–514.
Kingston, R. E., Kaufman, R. J., Bebbington, C. R., & Rolfe, M. R. (1989) *Current Protocols in Molecular Biology* (Ausubel, F. M., et al., Eds.) Vol. 2, pp 16.14.1–16.14.13, John Wiley & Sons, New York.
Laemmli, U. K. (1970) *Nature 227*, 680–685.
Lerner, L., & Bax, A. (1986) *J. Magn. Reson. 69*, 375.
Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., & Clore, G. M. (1989a) *Biochemistry 28*, 6150–6156.
Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., & Bax, A. (1989b) *J. Am. Chem. Soc. 111*, 1515–1517.
Massagué, J. (1990) *Annu. Rev. Cell. Biol. 6*, 597–641.
Messerle, B. A., Wider, G., Otting, G., Weber, C., & Wüthrich, K. (1989) *J. Magn. Reson. 85*, 608–613.
Morrissey, J. H. (1981) *Anal. Biochem. 117*, 307–310.
Ogawa, Y., & Seyedin, S. M. (1991) *Methods Enymol. 198*, 317–327.
Pelton, J. G., Torchia, D. A., Meadow, N. D., Wong, C.-Y., & Roseman, S. (1991) *Biochemistry 30*, 10043–10057.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Roberts, A. B., & Sporn, M. B. (1990) *Handbook of Experimental Pharmacology. Peptide Growth Factors and Their Receptors I* (Sporn, M. B., & Roberts, A. B., Eds.) pp 419–472, Springer-Verlag, New York.
Shaka, A. J., Keeler, J., & Freeman, R. (1983) *J. Magn. Reson. 53*, 313–340.
Shaka, A. J., Lee, C. J., & Pines, A. (1988) *J. Magn. Reson. 77*, 274–293.
Sporn, M. B., & Roberts, A. B. (1990) *Cell Regul. 1*, 875–882.
Tsang, M. L.-S., Weatherbee, J. A., Dietz, M., Kitamura, T., & Lucas, R. C. (1990) *Lymphokine Res. 9*, 670.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley, New York.
Wüthrich, K. (1989) *Acc. Chem. Res. 22*, 36–44.