Transforming Growth Factor β1: Three-Dimensional Structure in Solution and Comparison with the X-ray Structure of Transforming Growth Factor β2[†,‡]

Andrew P. Hinck,[§] Sharon J. Archer,[§,||] Su Wen Qian,[⊥] Anita B. Roberts,[⊥] Michael B. Sporn,[§,#] James A. Weatherbee,[○] Monica L.-S. Tsang,[○] Roger Lucas,[○] Bo-Ling Zhang,[○] Jeff Wenker,[○] and Dennis A. Torchia*[,§]

Molecular Structural Biology Unit, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland 20892-4326, Laboratory of Chemoprevention, National Cancer Institute, Bethesda, Maryland 20892, and R & D Systems, Inc., Minneapolis, Minnesota 55413

Received February 28, 1996; Revised Manuscript Received April 18, 1996[⊗]

ABSTRACT: The three-dimensional solution structure of human transforming growth factor β1 (TGF-β1) has been determined using multinuclear magnetic resonance spectroscopy and a hybrid distance geometry/simulated annealing algorithm. It represents one of the first examples of a mammalian protein structure that has been solved by isotopic labeling of the protein in a eukaryotic cell line and multinuclear NMR spectroscopy. The solution structure of the 25 kDa disulfide-linked TGF-β1 homodimer was calculated from over 3200 distance and dihedral angle restraints. The final ensemble of 33 accepted structures had no NOE or dihedral angle violations greater than 0.30 Å and 5.0°, respectively. The RMSD of backbone atoms for the ensemble of 33 structures relative to their mean structure was 1.1 Å when all residues were used in the alignment and 0.7 Å when loop regions were omitted. The solution structure of TGF-β1 follows two independently determined crystal structures of TGF-β2 (Daopin et al., 1992, 1993; Schlunegger & Grütter, 1992, 1993), providing the first opportunity to examine structural differences between the two isoforms at the molecular level. Although the structures are very similar, with an RMSD in backbone atom positions of 1.4 Å when loop regions are omitted in the alignment and 1.9 Å when all residues are considered, there are several notable differences in structure and flexibility which may be related to function. The clearest example of these is in the β-turn from residues 69−72: the turn type found in the solution structure of TGF-β1 falls into the category of type II, whereas that present in the X-ray crystal structure of TGF-β2 is more consistent with a type I turn conformation. This may be of functional significance as studies using TGF-β chimeras and deletion mutants indicate that this portion of the molecule may be important in receptor binding.

Cytokines which belong to the transforming growth factor β (TGF-β)[1] family of peptide growth factors have potent effects on the proliferation, differentiation, and adhesive properties in a wide variety of cell types. Although TGF-β was initially described as a factor which induced rat kidney fibroblasts to proliferate (Roberts et al., 1981), it is now known to be a multifunctional cytokine with both stimulatory and inhibitory effects on a wide range of cells (Boyd & Massagué, 1989; Suardet et al., 1992). TGF-β affects the adhesive properties of cells by regulating key components involved in cell adhesion, such as collagen, as well as other components of the extracellular matrix including integrins, fibronectins, thrombospondin, tenascin, and laminin [reviewed by Massagué (1990)]. TGF-βs mediate their effects by binding to cell surface receptors, two of which (types I and II) have been shown to possess serine/threonine kinase activity [reviewed by Derynck (1994)]. Two other cell surface TGF-β binding proteins of known sequence include betaglycan (López-Casillas et al., 1991), also known as TGF-β type III receptor, and endoglin (Moren et al., 1992). Betaglycan and endoglin each consist of a homologous transmembrane glycoprotein domain and have a short cytoplasmic tail. Recent results suggest that TGF-β signaling occurs by formation of an interdependent heteromeric complex; TGF-β type I receptor (TβRI) requires the TGF-β

[†] This research was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health.

[‡] PDB ID codes: 1KLC (minimized average), 1KLA and 1KLD (33 accepted structures), and R1KLCMR (X-PLOR restraint tables).

* Corresponding author. Phone: 301-496-5750. Fax: 301-402-1512. Email: torchia@yoda.nidr.nih.gov.

[§] National Institutes of Health.

[||] Present address: The DuPont-Merck Pharmaceutical Co., Wilmington, DE.

[⊥] National Cancer Institute.

[#] Present address: Department of Pharmacology, Dartmouth Medical School, Hanover, NH.

[○] R & D Systems, Inc.

[⊗] Abstract published in *Advance ACS Abstracts,* June 15, 1996.

[1] Abbreviations: ASA, solvent-accessible surface area; CN-SED, carbon−nitrogen spin-echo difference; CCO-SED, carbon−carbonyl spin-echo difference; CT-HSQC, constant-time HSQC; DG, distance geometry; LRCH, long-range carbon−proton correlation; LRCC, long-range carbon−carbon correlation; HNHA, amide proton−α-proton correlation; HNHB, amide proton−β-proton correlation; HAHB, α-proton−β-proton correlation; HSQC, heteronuclear single-quantum correlation; NMR, nuclear magnetic resonance spectroscopy; NOESY, NOE spectroscopy; RMSD, root mean square deviation; 2D, two dimensional; 3D, three dimensional; 4D, four dimensional; SA, simulated annealing; TGF-β, transforming growth factor β; {TGF-β1}, the ensemble of accepted TGF-β1 solution structures; ⟨TGF-β1⟩, mean structure of {TGF-β1}; ⟨TGF-β1⟩$_r$, the restrained minimized value of ⟨TGF-β1⟩; {TGF-β2}, the ensemble of accepted TGF-β2 model structures; ⟨TGF-β2⟩, mean structure of {TGF-β2}; ⟨TGF-β2⟩$_r$, the restrained minimized value of ⟨TGF-β2⟩; TβRI, TGF-β type I receptor; TβRII, TGF-β type II receptor.

S0006-2960(96)00494-1 CCC: $12.00

```
            1                   21                  41                  61                  81                 101       112
TGF-β1 Hum  ALDTNYCFSS TEKNCCVRQL YIDFRKDLGW KWIHEPKGYH ANFCLGPCPY IWSLDTQYSK VLALYNQHNP GASAAPCCVP QALEPLPIVY YVGRKPKVEQ LSNMIVRSCK CS
TGF-β1 Mur  ALDTNYCFSS TEKNCCVRQL YIDFRKDLGW KWIHEPKGYH ANFCLGPCPY IWSLDTQYSK VLALYNQHNP GASASPCCVP QALEPLPIVY YVGRKPKVEQ LSNMIVRSCK CS
TGF-β1 Bov  ALDTNYCFSS TEKNCCVRQL YIDFRKDLGW KWIHEPKGYH ANFCLGPCPY IWSLDTQYSK VLALYNQHNP GASAAPCCVP QALEPLPIVY YVGRKPKVEQ LSNMIVRSCK CS
TGF-β1 Por  ALDTNYCFSS TEKNCCVRQL YIDFRKDLGW KWIHEPKGYH ANFCLGPCPY IWSLDTQYSK VLALYNQHNP GASAAPCCVP QALEPLPIVY YVGRKPKVEQ LSNMIVRSCK CS
TGF-β1 Mky  ALDTNYCFSS TEKNCCVRQL YIDFRKDLGW KWIHEPKGYH ANFCLGPCPY IWSLDTQYSK VLALYNQHNP GASAAPCCVP QALEPLPIVY YVGRKPKVEQ LSNMIVRSCK CS
TGF-β1 Rat  ALDTNYCFSS TEKNCCVRQL YIDFRKDLGW KWIHEPKGYH ANFCLGPCPY IWSLDTQYSK VLALYNQHNP GASASPCCVP QALEPLPIVY YVGRKPKVEQ LSNMIVRSCK CS

TGF-β2 Hum  ALDAAYCFRN VQDNCCLRPL YIDFKRDLGW KWIHEPKGYN ANFCAGACPY LWSSDTQHSR VLSLYNTINP EASASPCCVS QDLEPLTILY YIGKTPKIEQ LSNMIVKSCK CS
TGF-β2 Mur  ALDAAYCFRN VQDNCCLRPL YIDFKRDLGW KWIHEPKGYN ANFCAGACPY LWSSDTQHTK VLSLYNTINP EASASPCCVS QDLEPLTILY YIGNTPKIEQ LSNMIVKSCK CS
TGF-β2 Bov  ALDAAYCFRN VQDNCCLRPL YIDFKRDLGW KWIHEPKGYN ANFCAGACPY LWSSDTQHSR VLSLYNTINP EASASPCCVS QDLEPLTILY YIGKTPKIEQ LSNMIVKSCK CS
TGF-β2 Chk  ALDAAYCFRN VQDNCCLRPL YIDFKRDLGW KWIHEPKGYH ANFCAGACPY LWSSDTQHSR VLSLYNTINP EASASPCCVS QDLEPLTILY YIGKTPKIEQ LSNMIVKSCK CS
TGF-β2 Xen  ALDAAYCFRN VQDNCCLRPL YIDFKKDLGW KWIHEPKGYN ANFCAGACPY LWSSDTQHSR VLSLYNTINP EASASPCCVS QDLDSLTILY YIGNKPKIEQ LSNMIVKSCK CS

TGF-β3 Hum  ALDTNYCFRN LEENCCVRPL YIDFRQDLGW KWVHEPKGYY ANFCSGPCPY LRSADTTHST VLGLYNTLNP EASASPCCVP QDLEPLTILY YVGRTPKVEQ LSNMVVKSCK CS
TGF-β3 Mur  ALDTNYCFRN LEENCCVRPL YIDFRQDLGW KWVHEPKGYY ANFCSGPCPY LRSADTTHST VLGLYNTLNP EASASPCCVP QDLEPLTILY YVGRTPKVEQ LSNMVVKSCK CS
TGF-β3 Por  ALDTNYCFRN LEENCCVRPL YIDFRQDLGW KWVHEPKGYY ANFCSGPCPY LRSADTTHSS VLGLYNTLNP EASASPCCVP QDLEPLTILY YVGRTAKVEQ LSNMVVKSCK CS
TGF-β3 Chk  ALDTNYCFRN LEENCCVRPL YIDFRQDLGW KWVHEPKGYF ANFCSGPCPY LRSADTTHST VLGLYNTLNP EASASPCCVP QDLEPLTILY YVGRTPKVEQ LSNMVVKSCK CS
```

FIGURE 1: Sequence alignment of the mature form of the three major TGF-β isoforms. Species correspond as follows: Hum, human; Mky, monkey; Rat, rat; Mur, murine; Bov, bovine; Por, porcine, Chk, chicken; Xen, *Xenopus laevis*. The set of nine conserved cysteine residues is indicated by bold type; species variation of the amino acid sequence within a single isoform is indicated by underlined text.

type II receptor (TβRII) to bind ligand, and both receptors are essential for efficient signal transduction (Wrana et al., 1992, 1994).

The mature form of TGF-β is homodimeric, with 112 residues and 9 conserved cysteines per monomer. Each of the three major isoforms of TGF-β identified in mammalian cells, TGF-β1, -β2, and -β3, is synthesized *in vitro* as 390 amino acid precursors. The precursor is cleaved, transported to the cytoplasm, and eventually released as the mature 25 kDa homodimer. TGF-β isoforms show strong sequence identity relative to one another (72−79%) and across species (>96.5%) (Figure 1). The TGF-β isoforms also share significant sequence homology (~30%) and seven conserved cysteines with other members of the TGF-β superfamily, which include the inhibins, bone morphogenic proteins (BMPs), Müllerian inhibiting substance (MIS), and the growth/differentiation factors (GDFs) (Burt & Law, 1994). One exception to this overall pattern is the activins, which share with the TGF-βs the same set of nine conserved cysteines.

In a number of *in vitro* cell proliferation assays, the activity and potency of TGF-β1, TGF-β2, and TGF-β3 are often indistinguishable (Danielpour & Sporn, 1990; Suardet et al., 1992). However, the existence of three highly similar TGF-β sequences, each of which is essentially invariant across species, suggests important specific roles for each of the isoforms. The main line of evidence that supports such a conclusion follows from readily apparent differences in activity among the different TGF-β isoforms. Examples of this for TGF-β1 and TGF-β2, the two isoforms that have been the most thoroughly examined, include growth inhibition of aortic endothelial (Jennings et al., 1988) and LS513 colorectal cancer cells (Suardet et al., 1992).

One aspect of the TGF-β's thought to play an important role in conferring specificity is the differential affinity of the mature forms of the protein for the TGF-β receptors (Massagué, 1990; Roberts & Sporn, 1990). Endoglin, for example, is known to bind TGF-β1 and TGF-β3 with high affinity, but not TGF-β2 (Cheifetz et al., 1992). Similarly, several lines of evidence now suggest that the affinity of the type II receptor for TGF-β1 and TGF-β3 is much greater than that for TGF-β2. Recombinant forms of TGF-β in which amino acid sequences among the different isoforms have been interchanged (Burmester et al., 1993; Qian et al., 1992), omitted (Qian et al., 1994), or substituted on a residue-by-residue basis (Amatayakul-Chantler et al., 1994; Qian et al., 1994) have been used to explore how differences in primary sequence affect the functional properties of the molecule. Results obtained from these studies indicate that residues 69−72 may contribute in part to receptor binding in some systems (Qian et al., 1992) and that differential sequestration of TGF-β2 by α2 macroglobulin (α_2M), an abundant serum protein that appears to recognize A45 and A47 of TGF-β2 (Burmester et al., 1993), may account for the differential potencies of the different TGF-β isoforms in some assay systems.

Clearly detailed characterizations of the three-dimensional structures of the different TGF-β isoforms are essential to understanding their activity at the molecular level. The three-dimensional structure of the TGF-β2 homodimer has been reported at 2.2 (Schlunegger & Grütter, 1992), 2.1 (Daopin et al., 1992), 1.95 (Schlunegger & Grütter, 1993), and 1.8 Å (Daopin et al., 1993) resolution and has revealed an unusual structure with four intrachain disulfide bonds and one interchain disulfide formed by oxidation of the thiol group of cysteine 77. A recent X-ray structure of OP-1 (Griffith et al., 1996), another TGF-β superfamily member with 36% amino acid identity with TGF-β2, has shown that, even for the most distantly related members of the TGF-β superfamily, the two proteins share the same overall fold. We previously reported the NMR backbone assignments and secondary structure of TGF-β1 in solution (Archer et al., 1993a,b), and although a detailed comparison of the three-dimensional structure was not possible, we indicated that the secondary structure of TGF-β1 in solution did not differ significantly from that of crystalline TGF-β2.

In order to achieve a complete structural comparison of TGF-β1 and TGF-β2, we have now used additional selective isotope labeling coupled with heteronuclear NMR techniques to elucidate the 3D structure of TGF-β1. Because TGF-β1 must be expressed in a mammalian cell system, we designed a selective labeling scheme that afforded abundant structural information yet minimized the cost of purchasing isotopically labeled amino acids. Herein, we report the three-dimensional solution structure of TGF-β1, which provides the first opportunity to explore structure−function relationships among the different TGF-β isoforms in a detailed manner.

MATERIALS AND METHODS

Isotopic Enrichment, Purification, and NMR Sample Preparation. Isotopically enriched TGF-β1 was produced from a stably transfected CHO cell line by culturing the cells in an Opticore bioreactor and by including a mixture of isotopically labeled amino acids as part of the growth medium as first described by Archer et al. (1993a). The composition of the labeled amino acids used to prepare the medium is provided in the Supporting Information section. Purification and assessment of the purity and specific activity of the recombinant protein were carried out as previously described (Archer et al., 1993a). Approximately 10.0 mg of the recombinant, isotopically enriched protein was recovered from 15 L of growth medium, and its purity was found to be greater than 97%. The labeling pattern of the protein that resulted from the use of the amino acid mixture described here is designated as "scheme D".

The purified protein was initially obtained as a mixture dissolved in 30% acetonitrile/70% H_2O/0.1% TFA. Samples were prepared for NMR spectroscopy by lyophilizing a volume of the mixture containing 5.0 mg of the protein, by dissolving the protein in 30% [$^2H_3$]acetonitrile/70% H_2O, and by repeating the lyophilization step. The sample was then dissolved in 1.0 mL of 95% H_2O/5% ^2H_2O, and polymeric impurities present in the sample were quantitatively removed by alkaline precipitation of the protein by the addition of 0.05 M NaOH until a final pH of 9.5 was attained, followed by immediate centrifugation to recover the insoluble protein, and finally by dissolving the recovered pellet in 1.0 mL of acidified 95% H_2O/5% ^2H_2O (pH 4.2). The concentration of salts present in the sample was then lowered approximately 1000-fold by ultrafiltration (10K MWCO; Centricon, Amicon, Beverly, MA) using 95% H_2O/5% ^2H_2O, pH 4.2, as the solvent. The volume of the sample was then reduced to 150 μL, and the sample was transferred to a 5 mm NMR microcell (Shigemi, Allison Park, PA). The tube and transfer pipet were rinsed with a total volume of 80 μL of 95% H_2O/5% ^2H_2O, pH 4.2, to bring the final volume to 230 μL. Samples prepared for spectroscopy in ^2H_2O were handled in a similar fashion with the exceptions that the sample was lyophilized three times from 99.98% ^2H_2O, pH* 4.0 (CIL, Cambridge, MA), following the ultrafiltration treatment and the sample was finally dissolved in 99.997% ^2H_2O (CIL, Cambridge, MA). The TGF-β1 samples were prepared at acidic pH since the protein becomes insoluble at higher pH values. Moreover, this pH value miminizes any major structural differences that might result between the pH conditions under which X-ray diffraction data for TGF-β2 were collected (pH 4.2, Daopin et al., 1993; pH 4.5−5.5, Schlunegger & Grütter, 1992) and the present NMR data for TGF-β1.

NMR Spectroscopy. NMR spectra were acquired with Bruker AMX-500, DMX-500, and AMX-600 spectrometers equipped with carbon-optimized triple-resonance probes and pulsed Z-gradients at a sample temperature setting of 45 °C. During the course of data collection, there were no detectable changes in the 2D $^1H/^{15}N$ HSQC spectrum of scheme D TGF-β1, indicating that the protein is stable under these sample conditions. All experiments employed the States−TPPI method for quadrature detection in the indirectly detected dimensions (Marion et al., 1989). Chemical shifts of the ^1H, ^{13}C, and ^{15}N signals were referenced to $^1H^2HO$ (4.58 ppm at 45 °C), indirectly to TSP (Bax & Subramanian, 1986), and to liquid ammonia (Live et al., 1984), respectively. Parameters employed for the various NMR experiments are provided in Table 2 of the Supporting Information section. The majority of experiments conducted utilized scheme D TGF-β1, although data previously collected using unlabeled, [U-^{15}N]labeled, and selectively labeled scheme A, B, and C TGF-β1 samples (Archer et al., 1993a) were also used in the structure determination process.

Data processing and analysis were accomplished using the nmrPipe (Delaglio et al., 1995) and PIPP (Garrett et al., 1991) software packages, respectively. Quantitative estimates of coupling constants from HNHA (Vuister & Bax, 1993a), HAHB (Grzesiek et al., 1995), CCO-SED (Grzesiek et al., 1993a), CN-SED (Vuister et al., 1993), and LRCC (Bax et al., 1992) experiments were achieved by recording "diagonal" and "cross peak" peak intensities with the cubic-spline interpolation facility of the PIPP software package and by relating these quantities to coupling constants by the appropriate trigonometric relationships. Quantitative values of couplings estimated from the LRCH (Vuister & Bax, 1993b) experiment were derived using the equation

$$\frac{V_{2D}f}{V_{3D}} = \tan^2 [\pi J_{C-H}(2T - \zeta - \epsilon)] \qquad (1)$$

where the quantity $2T - \zeta - \epsilon$ was equal to 22.6 ms and V_{2D}/V_{3D} is the intensity ratio recorded in the 2D reference and 3D LRCH experiments, respectively. The scaling factor, f, which relates peak intensities in the 2D and 3D spectra acquired under otherwise identical conditions, is a statistical quantity calculated for 18 of the 20 leucine $C^{\delta 1}-H^{\delta 2}$ and $C^{\delta 2}-H^{\delta 1}$ correlations assuming a uniform value of $^3J_{C\delta-H\delta}$ equal to 5.2 Hz. Qualitative estimates of $^3J_{N-H\beta 1}$ and $^3J_{N-H\beta 2}$ were obtained from an HNHB data set by calculating cross peak/diagonal ratios as previously described (Archer et al., 1991).

^{15}N Relaxation Measurements. Backbone amide nitrogen T_1, T_2, and $^1H-^{15}N$ NOE relaxation parameters were measured at 45 °C using [U-^{15}N]TGF-β1 and the ^{1}H-detected pulse schemes previously described (Kay et al., 1992). The four T_1 and five T_2 data sets were acquired over a period of several months using a Bruker AM-500 or AMX-500 spectrometer, each operating at a proton resonance frequency of 500.13 MHz. The NOE data set was obtained using only the AMX-500. The T_1 and T_2 data sets consisted of eight interleaved time points with variable relaxation delays between 20−1500 and 8.2−98.7 ms, respectively, whereas the NOE experiment was carried out by recording signal intensities in two experiments, one with a series of proton presaturation pulses and one with the presaturation period replaced by a delay of equal length (2.9 s). The data were then analyzed by integrating volumes as a function of the variable relaxation delay in a pseudo-three-dimensional manner to obtain time-dependent intensities. These were then fitted, in the case of T_1 and T_2, to a decaying exponential, $I(t) = I_o \exp(-t/T_i)$ for each site i, using the conjugate gradient minimization technique (Press et al., 1988) with the uncertainties in the extracted parameters, I_o and T_i, obtained using a Monte Carlo approach (Kamath & Shriver, 1989). $^1H-^{15}N$ NOE values were calculated from the intensity ratio of the spectrum obtained with presaturation to that obtained without presaturation modified by a correction factor that

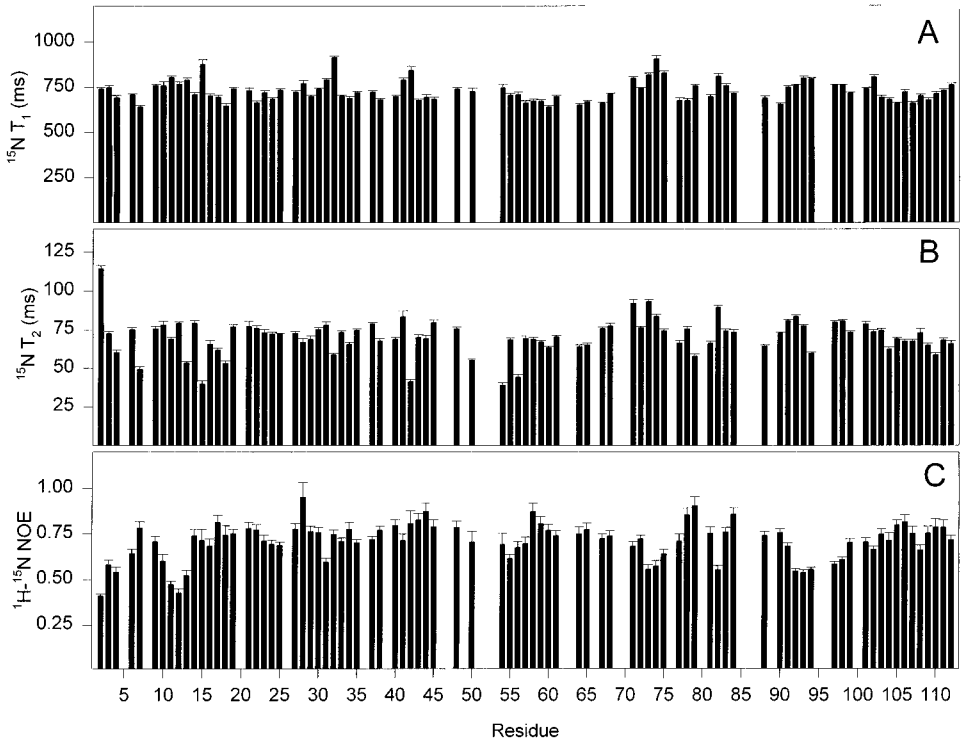

FIGURE 2: Graphical representation of amide nitrogen T_1 (A), T_2 (B), and $^{15}N-^1H$ NOE values (C) and their corresponding errors for TGF-β1 at a field strength of 11.7 T and 45 °C.

takes into account incomplete magnetization recovery during the recovery period (Grzesiek & Bax, 1993). The reported T_1 and T_2 relaxation rates correspond to their mean values determined over four and five independent data sets, respectively. The corresponding errors were obtained by calculating the RMSD relative to the average value. A graphical representation of the relaxation parameters and their errors is provided (Figure 2); a complete listing is provided as part of the Supporting Information (Table 3, Supporting Information).

Structure Calculations. NOE connectivities were classified as strong (1.8−2.8 Å), medium (1.8−3.3 Å), or weak (1.8−4.3 Å). The upper distance bound was corrected for restraints which included groups which could not be stereospecifically assigned or those involving methyl protons (Wüthrich et al., 1983). Hydrogen bond distance restraints were applied to backbone N^H and O atoms (2.5−3.7 Å) and H^N and O atoms (1.5−2.7 Å) for residues whose amide proton resonances displayed slow amide exchange rates and which displayed NOE patterns consistent with regular secondary structures of α-helices or β-sheets (Archer et al., 1993b). Intramonomer NOEs were initially distinguished from intermonomer NOEs by comparing the observed NOE connectivities with the corresponding distances in the 1.8 Å resolution crystal structure of TGF-β2 (Daopin et al., 1993). On the basis of the X-ray structure it was anticipated that several resonances near the crystallographic 2-fold symmetry axis, most notably those of C77 and V79, would display a combination of intermonomer as well as intramonomer interresidue and intraresidue NOE connectivities. Therefore, inter- as well as intramonomer distance restraints involving these residues were omitted throughout the structure determination process.

Torsion angle restraints for ϕ were derived from $^3J_{HN-H\alpha}$ coupling constants and intraresidue and sequential NOE data. ϕ was restrained to −60° for $^3J_{HN-H\alpha} < 6$ Hz and to −120° for $^3J_{HN-H\alpha} > 8$ Hz, with a range that varied between ±45° for $8.0 < {}^3J_{HN-H\alpha} < 9.0$ Hz and $5.0 < {}^3J_{HN-H\alpha} < 6.0$ Hz, ±35° for $9.0 < {}^3J_{HN-H\alpha} < 10.0$ Hz and $4.0 < {}^3J_{HN-H\alpha} < 5.0$ Hz, and ±25° for $^3J_{HN-H\alpha} > 10.0$ Hz or $^3J_{HN-H\alpha} < 4.0$ Hz. Restraints for χ_1 and χ_2 were derived from a variety of 2D and 3D experiments (described below) and were set to one of the three standard staggered rotamers with a range of ±30°. χ_1 angles for residues which contained a β-methylene group and which were uniformly $^{13}C/^{15}N$ enriched in the scheme D TGF-β1 sample were derived from qualitative estimates of $^3J_{N-H\beta}$ from an HNHB (Archer et al., 1991) data set and quantitive estimates of $^3J_{H\alpha-H\beta}$ derived from an HAHB (Grzesiek et al., 1995) data set. χ_1 restraints for residues that did not fall into this category were based upon qualitative estimates of $^3J_{N-H\beta}$ as before and additional information obtained from a 2D TOCSY, a 3D ^{15}N-separated HOHAHA, and intraresidue NOE patterns as described by Clore et al. (1991a). χ_1 restraints for valine, isoleucine, and threonine residues were based upon $^3J_{N-C\gamma}$ and $^3J_{Co-C\gamma}$ coupling constants obtained from 2D CN-SED (Vuister et al., 1993) and CCO-SED (Grzesiek et al., 1993a) CT-HSQC experiments and were confirmed by $^3J_{H\alpha-H\beta}$ (Grzesiek et al., 1995) and $^3J_{C\gamma-H\alpha}$ (Vuister & Bax, 1993b) coupling constants and intraresidue NOE patterns. LRCH (Vuister & Bax, 1993b) and LRCC (Bax et al., 1992) data sets provided estimates for $^3J_{C\alpha-C\delta}$ and $^3J_{C\delta-H\beta}$ for leucine residues and $^3J_{C\alpha-C\delta}$, $^3J_{C\delta-C\gamma2}$, and $^3J_{C\gamma2-H\gamma1}$ for isoleucine residues.

Three-dimensional structures of the mature 25 kDa TGF-β1 homodimer were calculated in two stages using a hybrid distance geometry/dynamical simulated annealing approach (Nilges et al., 1988) as implemented by the X-PLOR 3.1 package installed on a DEC α 3000-400 (Digital Equipment Corp., Maynard, MA) computer. The distance and dihedral angle restraints used in the calculation are summarized in Table 1. In the first stage of the calculation, a random extended structure was used as input, and the restraints

Table 1: Distance and Dihedral Restraint Statistics for the NMR Solution Structure of TGF-β1 and the TGF-β2 Model Structure

	experimental restraints for TGF-β1	simulated restraints for TGF-β2
total[a]	3250	3222
total NOE restraints	2810	2800
intramonomer	2602	2596
intraresidue ($i - j = 0$)	786	686
sequential ($\|i - j\| = 1$)	678	572
short range ($1 < \|i - j\| \leq 5$)	330	404
long range ($\|i - j\| > 5$)	808	934
intermonomer	208	204
long range	208	204
total hydrogen bond restraints	112	112
intramonomer	112	112
intermonomer	0	0
total dihedral restraints	328	310
ϕ	176	174
ψ	4	4
χ_1	118	108
χ_2	30	24
accepted structures	33	15
NOE violations >0.3 Å	0	0
dihedral violations >5°	0	0

	RMSDs from Idealized Covalent Geometry and Experimental Restraints			
	$\langle$TGF-β1$\rangle_r$[b]	{TGF-β1}[b]	$\langle$TGF-β2$\rangle_r$[b]	{TGF-β2}[b]
bonds (Å)	0.003	0.003 ± 0.00001	0.002	0.002 ± 0.00003
angles (deg)	0.74	0.73 ± 0.0095	0.61	0.61 ± 0.0024
impropers (deg)	0.44	0.43 ± 0.0089	0.35	0.36 ± 0.0046
distance restraints (Å)	0.022	0.023 ± 0.001	0.008	0.009 ± 0.0004
angle restraints (Å)	0.395	0.428 ± 0.043	0.075	0.092 ± 0.014
F_{NOE} (kcal/mol)[c]	67.7	75.2 ± 6.6	9.7	11.3 ± 0.9
F_{DIH} (kcal/mol)[d]	3.1	3.7 ± 0.7	0.11	0.16 ± 0.05

[a] Total for both monomers. [b] {TGF-β1} is the ensemble of 33 accepted structures; $\langle$TGF-β1$\rangle_r$ is obtained by restrained minimization of the average of {TGF-β1}; {TGF-β2} is the ensemble of 15 accepted structures; $\langle$TGF-β2$\rangle_r$ is obtained by restrained minimization of the average of {TGF-β2}. [c] Calculated using a square well potential and center averaging with a scale factor of 50, a square well constant of 1.0, and a ceiling of 1000 kcal/mol (Brünger, 1992). [d] Calculated using a square well potential and a scale factor of 50 (Brünger, 1992).

consisted of intramonomer distance and angle restraints. Disulfide S−S bonds were included as standard distance restraints (2.01−2.03 Å). Restraints involving the backbone heavy and side chain C^β atoms were used in distance geometry to generate the initial set of structures (*dg_sub_embed* protocol). The set of structures were then subjected to all-atom distance geometry with covalent disulfide bonds, followed by 4.9 ps of restrained dynamical simulated annealing (*dgsa* protocol). Finally, the resulting set of structures was refined by conducting 10 ps of restrained dynamical simulated annealing and energy minimization (*refine* protocol). This resulted in a low-resolution monomer structure that resembled that of TGF-β2 (approximately 6.0 Å RMSD for backbone atoms).

In the second stage of the calculation, intermonomer distance restraints along with a disulfide bond linking cysteine 77 from the two monomer subunits were introduced. Structures were then generated in a manner similar to that outlined above, with the exceptions that a noncrystallographic symmetry function (Brünger, 1992) and distance symmetry restraints (Nilges, 1993) were applied. The former restrained the RMSD between the two monomers to a minimum, whereas the latter required symmetry-related distances to fall within a specified tolerance (0.01 Å). The structures were then improved in an interative fashion as errors or ambiguities present in the restraints were removed or new restraints were added. Due to the wide variation in diagonal peak intensities, a large proportion of the NOEs were initially categorized as weak. Those NOEs which proved to be consistent with the remaining NOE, hydrogen bond, and dihedral angle restraints as the structure refinement progressed and which displayed upper restraint bounds which were 2.0 Å or larger than the actual distance specified by the restraint were subsequently reexamined and reclassified, as appropriate. The final set of structures began with 40 initial structures, 33 of which after refinement had distance and dihedral violations of less than 0.30 Å and 5.0°, respectively. Restraint tables used in the final round of calculations have been submitted to the Protein Data Bank (code R1KLCMR).

Estimation of Precision and Accuracy. Estimates for the precision and the accuracy of the X-PLOR calculations have been obtained by measuring our ability to reproduce the 1.8 Å resolution X-ray structure of TGF-β2 (Daopin et al., 1993) using a set of distance and dihedral angle restraints which were derived using the coordinates of the crystal structure. The large extent of sequence and structural similarity between TGF-β1 and TGF-β2 meant it was possible to maintain a similar distribution of distance and dihedral restraints for the simulated TGF-β2 structure relative to the TGF-β1 structure determined from NMR data. This was accomplished by using a tabular listing of TGF-β1 distance restraints as the basis for selecting a corresponding set of interproton distances from TGF-β2 crystal structure. Interatomic distances corresponding to each of the restraints were either translated into distance restraints in the X-PLOR format using the distance ranges and pseudoatom corrections previously described or were discarded when either the distance exceeded the upper distance bound possible for the atom types involved or the corresponding atom in TGF-β2 did not exist. The former category comprised 78.5% and 64.7% of the final intramonomer and intermonomer distance

restraints, respectively. The remaining 21.5% of intramonomer restraints and 35.3% of intermonomer restraints were randomly chosen from a pool of all possible intra- and intermonomer TGF-β2 distance restraints, excluding those already assigned. Dihedral angle restraints were applied in a similar fashion, with the exception that the actual dihedral angle was changed to either −60, −90, or −120 in the case of ϕ or one of the three staggered rotamers in the case of χ_1 and χ_2. Table 1 contains a summary of the simulated restraint data set. X-PLOR calculations for the simulated TGF-β2 model were carried out in an otherwise identical fashion to that used for the experimental TGF-β1 structures.

RESULTS

Signal Assignments. Nonstereospecific proton and nitrogen backbone and side chain signal assignments for the TGF-β1 homodimer have been previously reported and were based on various NMR experiments performed using unlabeled, [U-^{15}N] labeled, and three selectively ^{15}N- and ^{13}C-labeled TGF-β1 samples, designated as schemes A, B, and C (Archer et al., 1993a). In order to obtain a high-precision NMR structure, we have prepared a fourth selectively ^{15}N/^{13}C-labeled sample, designated scheme D, by overexpression of the mature form of the protein in a stably transfected CHO cell line using a medium containing a mixture of labeled amino acids (Table 1, Supporting Information). In this mixture, all hydrophobic amino acids except tryptophan were enriched uniformly with ^{13}C and ^{15}N, whereas the remaining amino acids, excluding arginine, were labeled with ^{15}N at the backbone amide position. The yield of purified, labeled protein obtained from 15 L of growth medium was 10.0 mg and was sufficient to prepare two NMR samples contained in 5 mm susceptibility-matched microcells. Although the final cost of the TGF-β1 samples was relatively high, approximately $2000 per milligram of labeled protein, it is likely that the expense of preparing such samples in the future can be significantly reduced (a) since the composition of the growth medium has not been optimized to maximize expression and (b) because only of a small fraction (10−20%) of the expressed TGF-β is released as the mature form of the protein.

Two-dimensional ^{1}H−^{15}N HSQC spectra of scheme D TGF-β1, acquired with ^{13}C decoupling in the indirectly detected dimension, demonstrated the presence of high-level ^{15}N enrichment throughout the protein backbone for all residues, except arginine. A similar spectrum, obtained with a low-power ^{13}C pulse that effectively inverts only the C$^{\alpha}$ resonances during the ^{15}N evolution period, resulted in skewed doublets for those residues sequentially following those sites targeted for ^{13}C enrichment. This splitting is caused by the one-bond and two-bond scalar couplings between C^{O}−N^{H} and C^{O}−H^{N}, respectively, and confirms the presence of high-level residue-specific incorporation of ^{13}C. There were, however, exceptions to this general finding involving glycine, serine, and alanine residues. For glycine and serine, an asymmetry in the H^{N}−N^{H} doublet was apparent, indicative of incomplete ^{13}C enrichment. Intensity ratio measurements suggest that the level of ^{13}C enrichment for these residues lies in the range of 60−70% and was consistent with our later observations that ^{1}H/^{13}C correlations involving these residues were lower in intensity when compared to other residues targeted for ^{13}C enrichment. For alanine, the level of ^{13}C enrichment at all sites was undetectable and contradicts our previous experience with successful incorporation of this amino acid into the protein. The cause of this discrepancy remains unknown.

A quantitative measure for the level of ^{13}C enrichment in the recombinant protein was obtained by recording a 2D NOESY spectrum of scheme D TGF-β1 in $^{2}H_2O$ with ^{13}C decoupling in the F_1 dimension. In this spectrum, diagonal as well as off-diagonal resonances for ^{13}C-bound protons are split in the F_2 dimension by the large one-bond ^{1}H−^{13}C scalar coupling, thereby revealing the presence of any component of the resonance arising from ^{12}C-bound protons. Intensity measurements of the ^{13}C doublet versus the central ^{12}C singlet provided our estimates for the level of ^{13}C enrichment of the various amino acids incorporated into scheme D TGF-β1 (Table 1, Supporting Information). We note that the values obtained by this method likely underestimate the level of enrichment due to the strong ^{1}H−^{13}C dipolar interaction, which is expected to lead to a significant broadening of resonances arising from ^{13}C-bound, versus ^{12}C-bound, protons.

Signal assignments for a majority of backbone and side chain carbon-13 resonances were obtained by extending the existing set of backbone ^{1}H and ^{15}N assignments using the 2D CT-HSQC and CT-HSQC-relay and 3D CBCA(CO)NH, C(CO)NH, and HCCH-TOCSY experiments. The approach we used was to extend the assignments from the H^{N} position to C$^{\alpha}$ and out to the side chain carbons using the CBCA-(CO)NH and C(CO)NH data sets. Confirmation of the carbon assignments was accomplished by comparing the newly obtained ^{13}C chemical shifts with those for H$^{\alpha}$ and side chain protons previously made using the HCCH-TOCSY data in combination with the ^{1}H/^{13}C CT-HSQC correlation map. The CT-HSQC data set was obtained with CT evolution times of 26.0 and 52.0 ms to maximize sensitivity of the H$^{\alpha}$/C$^{\alpha}$ and H$^{\beta}$/C$^{\beta}$ resonances while maintaining sufficient separation of resonances in the methyl region of the 2D ^{1}H/^{13}C correlation map. This was particularly important for a cluster of seven methyl groups near 0.95/21.5 ppm arising from four separate valine residues (inset, Figure 3a). The ring carbon assignments of Phe and Tyr residues were obtained by extending the existing set of ring proton chemical shifts using 2D CT-HSQC and CT-HSQC-relay experiments. Assignment of the eight Tyr C$^{\delta}$ and C$^{\epsilon}$ resonances proceeded with relative ease, whereas assignment of C$^{\delta}$, C$^{\epsilon}$, and C$^{\zeta}$ resonances of the Phe residues was considerably more difficult owing to their low intensities, which is presumably a consequence of strong coupling effects among the ring ^{13}C nuclei. A summary of the assignments is provided by three portions of the 2D ^{1}H/^{13}C CT-HSQC correlation map (Figure 3). A tabular listing of the complete set of assignments used in the work is provided in the Supporting Information (Table 4, Supporting Information).

NOESY Data, Hydrogen Bonds, and Distance Restraints. Interproton distance restraints were obtained by a series of NOESY experiments carried out using unlabeled, [U-^{15}N] labeled, and scheme D TGF-β1 at pH 4.2 and were classified as weak (1.8−4.3 Å), medium (1.8−3.3 Å), or strong (1.8−2.8 Å). The assignment of NOE cross peaks was facilitated in part by isotope filtering experiments which limited the subset of observable resonances in a given experiment. As an example, the 4D HCHC NOESY spectrum obtained from the scheme D sample provides NOE connectivities that are restricted to pairs of residues isotopically enriched with ^{13}C

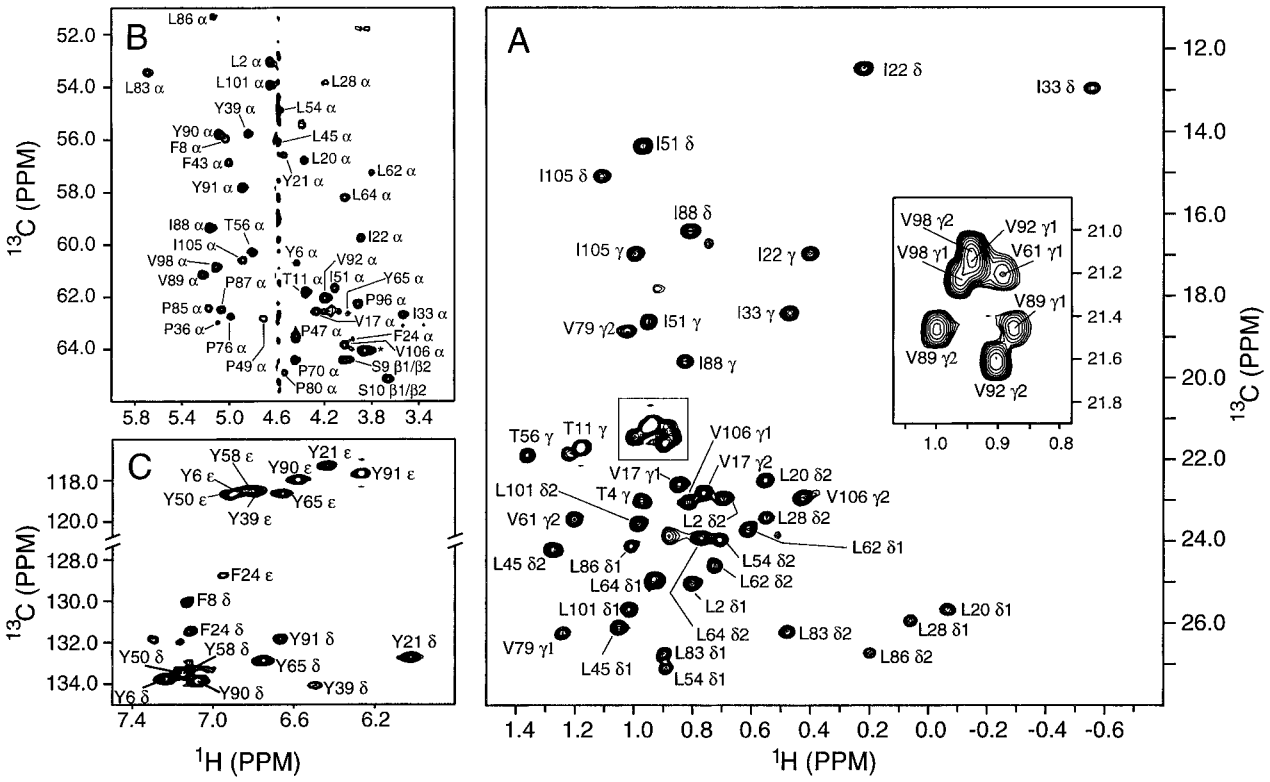

FIGURE 3: Three portions of the ^{1}H/^{13}C CT-HSQC map for scheme D TGF-β1 at 11.7 T and 45 °C. Resonance assignments are indicated by the one-letter amino acid code and residue number. (A) The methyl region and an expansion of several overlapping methyl groups acquired with a ^{13}C CT evolution period of 26 and 52 ms, respectively. (B) The α-proton region acquired with a ^{13}C CT evolution period of 26 ms. (C) The aromatic region acquired with a CT evolution period of 18.2 ms.

(Figure 1, Supporting Information). 3D spectra which illustrate this point are the 3D ^{15}N-separated NOESY acquired with and without ^{12}C selection in the F_1 dimension (Figure 4) and the 3D ^{13}C-separated NOESY with and without ^{13}C selection in the F_1 dimension. Additional spectra which complement those outlined above include a 3D ^{15}N-separated NOESY with ^{13}C evolution in F_1 and a 2D NOESY with ^{12}C selection in both F_1 and F_2. Analysis of the NOESY data yielded a total of 1301 distance restraints per monomer. A total of 56 hydrogen bonds per monomer involving main chain amide protons were identified on the basis of those amide resonances which displayed significant protection against hydrogen exchange and were located in regular elements of secondary structure as determined by sequential and long-range NOE connectivities (Archer et al., 1993b).

Coupling Constants, Dihedral Angles, and Stereospecific Assignments. Estimates for the backbone dihedral angle, ϕ, for 88 residues in each monomer were determined by recording cross peak to diagonal ratios in a 3D HNHA spectrum (Vuister & Bax, 1993a) and by relating these to $^{3}J_{HN-H\alpha}$. ψ values for Pro[70] and Gly[93] were estimated from intraresidue and short-range NOE patterns and were also included as dihedral angle restraints. The side chain dihedral angle, χ_1, and stereospecific assignments of H$^{\beta}$ protons for those residues containing a β-methylene were determined from a semiquantitative interpretation of diagonal cross peak ratios in an HNHB data set along with quantitative H$^{\alpha}$−H$^{\beta}$ couplings determined from an HAHB data set (Figure 2, Supporting Information). For those residues which lacked ^{13}C enrichment, a 2D TOCSY and a 3D ^{15}N-separated TOCSY were used in place of the HAHB data set. A total of 45 such χ_1 restraints were obtained. A summary of these results is provided as part of the supplementary material (Table 5, Supporting Information).

For Val, Thr, and Ile residues, HNHB and HAHB data were complemented by estimates of $^{3}J_{N-C\gamma}$ and $^{3}J_{Co-C\gamma}$ coupling constants obtained from 2D CN-SED (Vuister et al., 1993) and CCO-SED (Grzesiek et al., 1993a) CT-HSQC experiments to provide 14 additional χ_1 restraints as well as stereospecific C$^{\gamma}$ assignments for 6 Val residues (Figure 3, Table 6, Supporting Information). Side chain χ_2 restraints as well as stereospecific C$^{\delta}$ carbon assignments for Leu residues were obtained using estimates of $^{3}J_{C\alpha-C\delta}$ and $^{3}J_{C\delta-H\beta}$ (Figure 4, Supporting Information). χ_2 restraints and stereospecific assignments of the H$^{\gamma 1}$ resonances for Ile residues were obtained using estimates of $^{3}J_{C\alpha-C\delta}$, $^{3}J_{C\delta-C\gamma 2}$, and $^{3}J_{C\gamma 2-H\gamma 1}$. A total of 15 χ_2 restraints per monomer as well as stereospecific assignments of the C$^{\delta}$ resonances for 8 Leu residues and the H$^{\gamma 1}$ resonances of the 5 Ile residues in TGF-β1 were obtained from this analysis. A summary of the couplings and the assigned rotomer configurations is provided as part of the supplementary material (Tables 7 and 8, Supporting Information).

Tertiary Structure of TGF-β1. The two monomer chains which comprise the mature form of the TGF-β1 homodimer each contain 112 residues and display indistinguishable ^{1}H, ^{13}C, and ^{15}N chemical shifts. This indicates that in solution the protein is symmetric on the chemical shift time scale (microsecond to millisecond). One consequence of this is that the subset of NOEs arising as a result of intermonomer contacts could not be distinguished from those arising from intramonomer NOE contacts. Typically, this problem is circumvented by preparing an isotopic heterodimer and by carrying out experiments which select for NOEs which arise

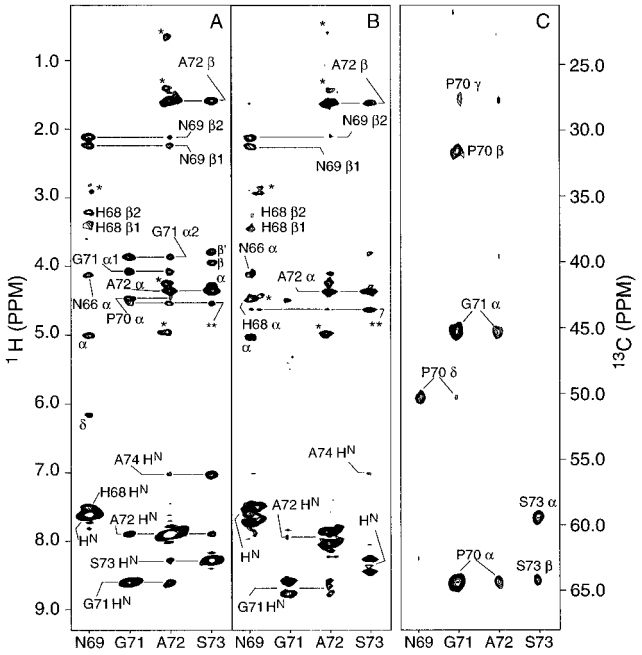

FIGURE 4: A series of strip plots for residues N69-P70-G71-A72-S73 of TGF-β1 taken from various 3D ^{15}N-separated NOE data sets which illustrate the conformation adopted by residues 69−73. (A) A 3D ^{15}N-separated NOE data set acquired using [U-^{15}N]TGF-β1 at 14.1 T and 45 °C. NOE cross peaks are designated by the one-letter amino acid code and residue number to which they have been assigned. Asterisks designate NOEs which arise from nearby overlapping diagonal peaks, whereas double asterisks indicate NOEs that arise as a result of exchange with water. (B) A 3D ^{15}N-separated ^{13}C-filtered NOE data set acquired using scheme D TGF-β1 at 11.7 T and 45 °C. NOE cross peaks are labeled as in panel A. (C) A 3D ^{15}N-separated ^{13}C-separated NOE data set acquired using scheme D TGF-β1 at 11.7 T and 45 °C. NOE cross peaks are labeled as in panel A.

between isotopically bound and non isotopically bound protons. This approach, however, requires the ability to reversibly disassociate the monomers, which to our knowledge cannot be done with significant recovery of TGF-β1 due to the presence of a disulfide bond linking Cys^{77} in the two monomers.

In order to proceed, we considered applying a computational method that does not necessitate monomer- or dimer-specific assignments (Nilges, 1993) or comparison of our NOESY data to that of the crystal structure of TGF-β2 in order to distinguish intra- versus intermonomer NOE contacts. The former has the disadvantage of requiring additional computations, whereas the latter approach could potentially bias the resulting model toward that of crystalline TGF-β2. Nevertheless, we initially chose the latter approach in our studies of TGF-β1 in solution since (a) previous NMR results indicated that the secondary structure and disulfide-bonding pattern of TGF-β1 were nearly equivalent to that of TGF-β2 (Archer et al., 1993b) and (b) the proportion of intermolecular restraints was only a small fraction of the total number of restraints (Table 1). Subsequent calculations, in which the set of intermonomer distance restraints were treated in an ambiguous fashion (Nilges, 1993), have shown that the family of accepted structures is indistinguishable, on a statistical basis, from the family of accepted structures calculated using the approach outlined above.

Three-dimensional structures of the TGF-β1 homodimer were calculated using a distance geometry/simulated annealing approach and proceeded in two stages. In the first stage, a monomer structure was generated by eliminating intermonomer distance restraints from the initial full set of distance restraints obtained from a manual examination of NOESY data, primarily 4D HCHC and 3D ^{15}N-separated NOE data sets. After several rounds of refinement, the monomer structure adopted a fold which was essentially the same as that found for the monomer of crystalline TGF-β2. In the second stage of the calculation, the three-dimensional structure of the TGF-β1 homdimer was generated by starting the calculation from two extended, overlapping, 112-residue monomer chains joined by a disulfide bond linking C77, by reintroducing the intermonomer distance restraints, and by applying a noncrystallographic symmetry function (Brünger, 1992) and distance symmetry restraints (Nilges, 1993) which minimized the RMSD between the two monomers and imposed symmetry. The structures were then further improved by semiautomated analysis of NOE data, using the program PIPP, by screening possible NOE assignments on the basis of calculated interproton distances within the current set of structures. This approach, along with stereospecific H$^\beta$, Ile H$^{\gamma 1}$, Val C$^\gamma$, and Leu C$^\delta$ assignments and the incorporation of χ_1 and χ_2 dihedral angle restraints for a number of residues (Tables 5−8, Supporting Information), resolved a number of ambiguities and approximately doubled the number of restraints used in the calculation.

The final family of structures was calculated using the NMR distance and dihedral angle restraints summarized in Table 1. Thirty-three of the 40 extended input structures had acceptably small NOE (<0.30 Å) and dihedral angle (<5.0°) violations following the final round of SA refinement. Structure statistics are provided in Table 2. An overlay of the backbone atom tracing of accepted structures for the TGF-β1 monomer and dimer (Figure 5) demonstrates that the structures superimpose well, over a large portion of the sequence. Exceptions to this finding are a solvent-exposed loop bridging residues 8−14, an extended loop segment from residue 29 to 39, a reverse turn including residues 69−72, and an extended, twisted antiparallel β-sheet with an intervening reverse turn from residue 87 to 102. Relative to the mean structure (⟨TGF-β1⟩), the RMSD of backbone atoms for the family of structures {TGF-β1} is 0.71 Å when these disordered regions are excluded and 1.07 Å for all residues.

Structure Simulation of TGF-β2. Estimates for overall and residue-specific RMSDs for both the precision and accuracy of the calculation were obtained by calculating dihedral and distance restraints based on the 1.8 Å X-ray structure of TGF-β2. These restraints, which were chosen to be analogous to those used in the determination of the solution structure of TGF-β1, were nearly equal in number and were employed in a manner identical to those used in the structural calculations outlined above. A summary of the distance and dihedral angle restraints used in the simulation is provided in Table 1. All 15 of the extended TGF-β2 structures used as input for the simulation had acceptably small NOE (<0.30 Å) and dihedral angle (<5.0°) violations following the final round of SA refinement. Structure statistics are provided in Table 2. A plot of backbone RMSDs as a function of sequence (Figure 6A) shows that the simulated TGF-β2 structures superimpose well over a large portion of the sequence. Relative to the mean structure (⟨TGF-β2⟩), the RMSD of backbone atoms for the family of accepted

Table 2: Coordinate RMSDs[a] of Experimental TGF-β1 and Simulated TGF-β2 Structures

	experimental TGF-β1 structures			
	all residues		loops omitted[d]	
structures[b]	bb atoms[c]	heavy atoms	bb atoms	heavy atoms
{TGF-β1} vs ⟨TGF-β1⟩	1.07	1.35	0.71	1.00
⟨TGF-β1⟩ vs ⟨TGF-β1⟩$_r$	0.23	0.45	0.19	0.41
⟨TGF-β1⟩$_r$ vs X-ray TGF-β2	1.90		1.39	
	simulated TGF-β2 structures			
	all residues		loops omitted[d]	
structures[b]	bb atoms[c]	heavy atoms	bb atoms	heavy atoms
{TGF-β2} vs ⟨TGF-β2⟩	0.91	1.17	0.60	0.83
⟨TGF-β2⟩ vs ⟨TGF-β2⟩$_r$	0.13	0.34	0.10	0.29
⟨TGF-β2⟩$_r$ vs X-ray TGF-β2	0.81	1.20	0.71	1.01
⟨TGF-β2⟩$_r$ vs ⟨TGF-β1⟩$_r$	1.52		1.17	

[a] In units of angstroms. [b] {TGF-βn} represents the ensemble of accepted structures; ⟨TGF-βn⟩ is the mean structure calculated from the family of accepted structures best-fitted to one another by omitting loop regions (see footnote *d*); ⟨TGF-βn⟩$_r$ is the structure obtained by restrained minimization of ⟨TGF-βn⟩; X-ray TGF-β2 refers to the crystal structure of TGF-β2 (Daopin et al., 1993). [c] Backbone atoms including N^H, C^α, and C^O. [d] Residues omitted in the superposition include 8−14, 29−39, 68−75, and 87−102.

structures ({TGF-β2}) is 0.60 Å when disordered regions (residues 8−14, 29−39, 68−75, and 87−102) are excluded and 0.91 Å for all residues (Table 2). Relative to the X-ray structure of TGF-β2 (from which the restraints were calculated), the backbone RMSD for the minimized average structure (⟨TGF-β2⟩$_r$) is 0.71 Å when disordered regions are excluded and 0.81 Å when all residues are considered (Table 2).

Backbone Amide Nitrogen Relaxation Parameters. Backbone ^{15}N T_1, T_2, and ^{1}H−^{15}N NOE relaxation parameters have been recorded for 85 of the 103 non-proline residues in TGF-β1. A graphical representation of the relaxation data and their errors is presented in Figure 2. The data indicate that the relaxation behavior is uniform throughout most of the protein backbone, with average T_1 and T_2 values of 730 ± 6 and 70.6 ± 1.3 ms, respectively. There are several portions of the sequence that have been identified as loop regions, however, where the relaxation parameters appear to systematically deviate from their respective average values. These regions include residues 2−4, 10−13, 50−56, 71−75, and 91−98. If these residues are excluded in calculating the average T_1 and T_2, we obtain average T_1 and T_2 values of 722 ± 7 and 72.1 ± 0.8 ms, respectively, which, assuming isotropic tumbling, corresponds to an overall correlation time of 11.6 ns.

The N-terminal residues, 2−4, exhibit elevated T_2 and diminished NOE values, indicating the presence of significant internal motion and flexibility in solution. Residues 10−13, which are located in the solvent-exposed loop spanning the C7/C16 disulfide, exhibit diminished NOE and larger than average T_1 and T_2 values, indicating the presence of rapid internal motions ($\tau_i \leq 1$ ns). Residues 50−56, an extended loop segment, display diminished T_2 and average NOE and T_1 values. This behavior indicates the likely presence of an exchange broadening effect. Possible sources include *cis*−*trans* isomerization about the C48−P49 peptide bond or C48−C111 disulfide bond isomerization. The last two regions that we consider are residues 71−75 and 91−98, each of which display diminished NOE and elevated T_1 and T_2 values. This indicates the presence of rapid internal motion ($\tau_i \leq 1$ ns) consistent with the observation that each comprises a solvent-exposed turn segment in the TGF-β1 solution structure.

DISCUSSION

The 3D solution structure of TGF-β1 was calculated from a total of 3250 distance and dihedral angle restraints, summarized in Table 1, using a hybrid distance geometry/simulated annealing (DG/SA) algorithm. As indicated previously, a major goal of the present work is to discern local structural differences that might explain differences in function between TGF-β1 and TGF-β2. The presence of such differences would provide clues regarding the identity of receptor binding sites, whereas their absence would indicate that the differences in activity are mediated largely by substituted residues located on the protein surface. The latter hypothesis has been suggested on the basis of analyses of the X-ray crystal structures of TGF-β2, which indicate that the majority of amino acid differences between the different TGF-β isoforms are localized to surface regions and loops; conserved residues include those which comprise the hydrophobic dimer interface, residues which lend unique conformational properties to the protein backbone such as G46 and *cis* P36, and the conserved set of nine cysteine residues.

In order to compare the structures of the two isoforms, we considered (a) structural differences that arise between the liquid and solid phases and (b) any systematic differences introduced by the completely independent approaches used to generate the X-ray and NMR structures. We approached the former by examining fluctuations in the position of the protein backbone, as revealed by an analysis of backbone nitrogen T_1, T_2, and NOE relaxation parameters in solution, and comparing these results with crystallographic *B*-factors. The latter was approached by assessing the ability of the DG/SA protocol to calculate the structure of an appropriate model protein using a set of distance and angle restraints derived from the coordinates of the model itself. The model we chose to simulate was the 1.8 Å resolution X-ray structure of TGF-β2 (Daopin et al., 1993). The distance and dihedral restraint sets used for the simulation were nearly equal in number and distributed similarly to those used in the calculation of TGF-β1 structures (Table 1).

Description of the Protein Fold. A ribbon diagram representing the C^α backbone trace of the restrained minimized average NMR TGF-β1 structure, designated here as ⟨TGF-β1⟩$_r$, is shown in Figure 6. The chain begins with

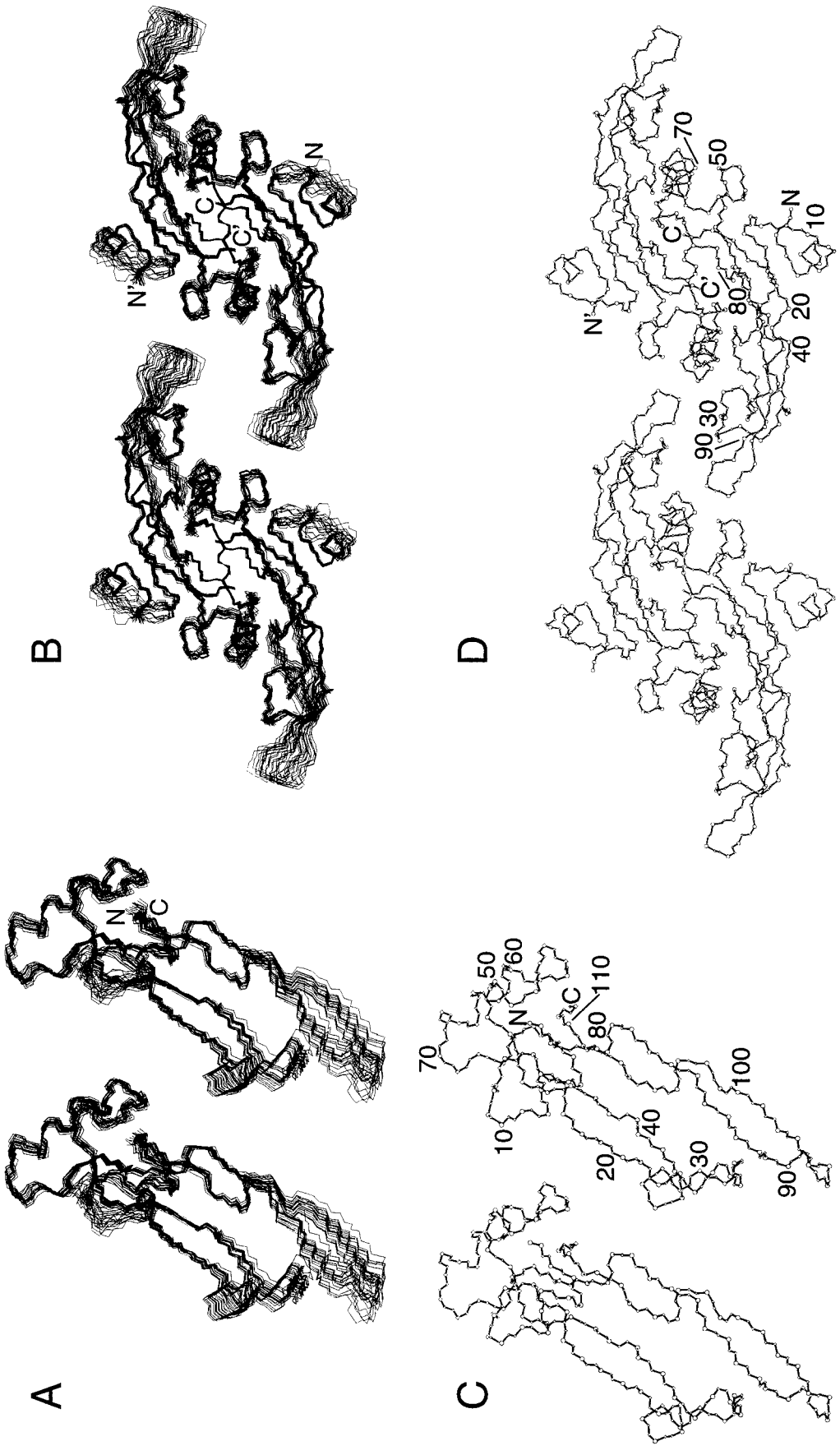

FIGURE 5: Several representations of the solution structure of TGF-β1. (Panels A, B) Stereo representation of the backbone heavy atom coordinates for the family of accepted structures overlaid versus their mean. The monomer is shown in panel A, whereas the TGF-β1 homodimer is shown in panel B. (Panels C, D) Stereo representation of the backbone heavy atom coordinates for the restrained minimized average NMR structure ($\langle$TGF-β1$\rangle_r$). The monomer is shown in panel C, whereas the TGF-β1 homodimer is shown in panel D.

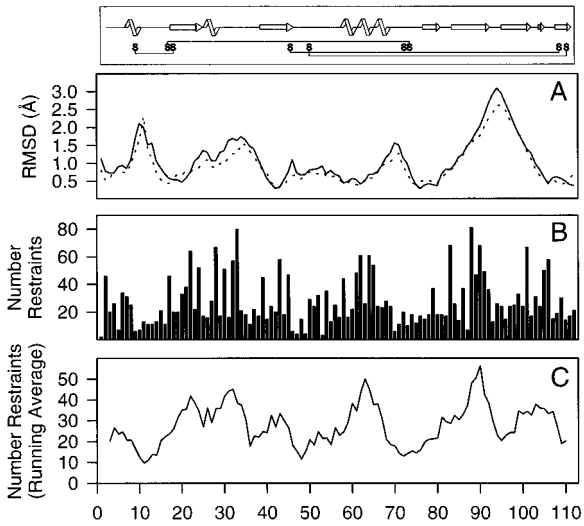

FIGURE 6: Statistics for the family of accepted structures generated by X-PLOR and the restraints used in the calculation. (A) The RMSD in backbone atom positions for the family of accepted structures versus the mean structure when loop regions (residues 8−14, 29−39, 68−75, and 87−102) were omitted in the superposition. The solid line reflects residue-based RMSDs for the experimentally determined TGF-β1 homodimer, whereas the dashed line corresponds to the simulated TGF-β2 model structure. (B) The total number of distance and dihedral restraints on a per residue basis that were used to calculate the NMR structure of the TGF-β1 homodimer. (C) A running average of the number of restraints per residue shown in panel B; it was calculated using $\sum_{i-2}^{i+2} n_i/5$.

A1, which displays very few NOE contacts and is flexible. This region is followed by residues 2−4, which form a relatively well ordered extended structure and whose side chains display several long-range hydrophobic contacts. The increase in order adopted by the backbone as it proceeds from residue 1 to 4 is accompanied by a decrease in flexibility. Residues 5−8 form a one-turn α-helix (α1) which is followed by a disordered, solvent-exposed loop lying between the disulfide bond formed by C7 and C16. Residues in this range display relatively few NOE contacts and appear to be flexible. Following this disordered region is a β-strand (β1) from residue 16 to 23, which pairs with residues 38−45 to form a somewhat irregular antiparallel β-sheet. This section of β-strand is followed by a second one-turn α-helix (α2), spanning residues 24−28 and an extended loop segment encompassing residues 29−37. Numerous hydrophobic contacts are made by the side chain atoms of W32-I33-H34-K35 with side chain atoms of those residues residing within the antiparallel β-sheet formed by residues 82−92 and 95−102. This extended loop region is followed by a β-strand from residues 38 to 45 (β2) and an extended loop region spanning residues 46−56. A well-defined three-turn α-helix spanning residues 58−68 (α3) follows, several of whose side chains participate in the hydrophobic dimer interface. This helix is terminated by residues 69−72, which participate in a type II β-turn (Sibanda & Thornton, 1991) and an extended loop structure from residue 73 to 76. The C-terminal portion of the molecule from residue 77 to 112 is largely extended and forms an irregular antiparallel β-structure punctuated by a crossing of the strands near residues 80−82, a severe twist near residues 102−104, and a type II′ β-turn from residue 92 to 95. Residues 77−80 (β3) pair with residues 108−112 (β7), and residues 83−92 (β4) pair with residues 95−102 (β5) and 104−106 (β6).

The topology of the monomer structure has been described by analogy to a curled left hand (Daopin et al., 1993). The two antiparallel β-sheets comprise the fingers, the smaller of the two sheets containing β1 and β2 comprise fingers 1 and 2, respectively, and the larger, extended antiparallel β-sheet containing β3, β4, β5, β6, and β7 give rise to fingers 3 and 4. Finger 3 is comprised of β3 and β4, whereas β5, β6, and β7 belong to finger 4. The thumb is comprised of the N terminus, α1, the flexible loop from residue 8 to 14, and the C7/C16 disulfide. The palm and heel of the hand are formed by α3 and the well-defined portion of the structure restrained by the C15/C78, C44/C109, and C48/C111 disulfides, respectively. The dimer is formed by a head-to-tail orientation of the two monomers with a inter-monomer covalent link formed by oxidation of the side chain thiols of C77/C277.

Precision of the Structure. A summary of the statistics for the family of accepted structures is presented in Table 2. The RMSD of backbone atoms for the 33 accepted structures relative to the mean structure (⟨TGF-β1⟩) is 0.71 Å when the disordered regions are excluded and 1.07 Å for all residues. A plot of the backbone RMSD as a function of residue number, Figure 6A, reveals that the structures superimpose well over a large portion of the sequence, with the exception of residues 8−14, 29−37, 68−72, and 85−103 where the RMSD exceeds 1.0 Å. The variation in backbone RMSD as a function of residue number reveals that disorder is generally inversely correlated to the number of restraints, as shown by plots presented in Figure 6B,C. Results obtained for the set of 15 accepted simulated TGF-β2 models relative to the mean model structure (⟨TGF-β2⟩) closely parallel the experimental results for TGF-β1 (Figure 6A). The observed overall RMSD for the model calculation is lower than the experimentally determined RMSD, the overall RMSD being 0.60 Å when the disordered regions are excluded and 0.91 Å for all residues (Table 2). This result was expected, as each of the restraints in the model calculation has been categorized exactly, whereas some of the restraints used in the calculation of the TGF-β1 NMR structures have been conservatively categorized.

Disorder, as revealed by elevated RMSD values, is typically ascribed to an insufficient number of restraints to precisely define the structure. A lack of restraints is often associated with extended loop regions because of low proton spin density or flexibility, as revealed by ^{15}N relaxation parameters, which indicates the presence of conformational freedom in solution. Disorder and flexibility are often correlated because the conformation of loop regions is often poorly defined with respect to the rest of the protein as well as being locally flexible on a time scale which affects the ^{15}N relaxation parameters. On the other hand, disorder and flexibility can also be independent of one another, as portions of a structure can be poorly defined, yet rigid, because the density of protons is low.

Residues 8−14 of TGF-β1, a solvent-exposed loop spanning the disulfide bond formed by C7/C16, provide one example of correlated disorder and flexibility. These residues exhibit short-range and intraresidue NOE connectivities (Table 2, Supporting Information), display elevated RMSD values (Figure 6A), and appear to undergo internal motion on the nanosecond time scale, as evidenced by diminished ^{1}H−^{15}N NOE values for residues 10−13 (Figure 2). Residues 69−72 of TGF-β1, a solvent-exposed type II β-turn

that bridges the C-terminal portion of α3 with the antiparallel β-structure which comprises fingers 3 and 4, provide an example of uncorrelated disorder and flexiblity. The precision of this portion of the structure is limited by the fact that only short-range and intraresidue NOEs are observed. Nevertheless, NOE patterns (Figure 4) reveal a well-defined type II β-turn. Moreover, although the ^{15}N relaxation data indicate that residues 71−75 undergo small amplitude internal motion (10−100ps), this motion affects only the C-terminal portion of the type II β-turn and does not appear to affect the conformation that the turn adopts.

The remaining two segments with elevated RMSD values, residues 29−39 and 87−102, include portions of the two antiparallel β-sheets and intervening loop segments that comprise the tips of the four fingers. This motif is responsible for the unusually extended shape that characterizes both TGF-β isoforms. Disorder in this portion of the structure appears to be largely due to a hinge bending of the tips of the fingers inward toward the palm and base of the hand. Two lines of evidence support such a conclusion. (1) The local conformation in this portion of the structure appears to be well defined: comparable RMSDs are found when the set of accepted TGF-β1 structures is aligned on the basis of the backbone atom positions of residues which comprise the tips of the fingers (29−39 and 87−102; backbone RMSD 0.18 Å) versus those of residues located in a relatively well ordered portion of the structure, such as α3 (57−68; backbone RMSD 0.14 Å). Hydrophobic interactions, mediated by side chain atoms of W32-I33-H34-K35 in fingers 1 and 2 with the side chains of I88-V89-Y90, L101, and M104 in fingers 3 and 4, form a hydrophobic pocket which presumably stabilizes and precisely defines relative orientation of the fingers. (2) Similar hinge bending of the tips of the fingers was also observed in the simulated TGF-β2 structures. This strongly suggests that the disorder observed in this part of the structure is largely a consequence of the limited precision and number of restraints which define the orientation of the tips of the fingers relative to the base and palm of the hand. Further evidence to support this idea follows from additional simulations which showed that disorder in this part of the sequence is greatly diminished when the number of long-range distance restraints was increased.

Flexibility in solution, as determined by ^{15}N relaxation data, suggests that residues 91−98 undergo rapid local internal motion on the 10−100 ps time scale. This motion, however, does not appear to extend to other regions of the TGF-β1 fingers as average relaxation parameters are observed for residues 29−39, 87−90, and 98−102. It is interesting to note that residues 91−98 comprise the most extended portion of the TGF-β dimer, including the type II′ β-turn connecting the two antiparallel β-strands and two to three residues of flanking sequence on either side. Residues in this portion of the sequence are highly exposed to solvent and display mostly sequential and short-range NOE interactions with other residues in this segment. These observations indicate that, among the residues which comprise the four fingers, only residues 91−98 experience significant flexibility in solution. Finally, we note that elevated thermal *B*-factors and weak electron density are reported for residues 91−98 in the 1.8 Å X-ray structure of TGF-β2 (Daopin et al., 1992, 1993; Schlunegger & Grütter, 1992, 1993), indicating that this portion of the structure is flexible in the crystalline state as well.

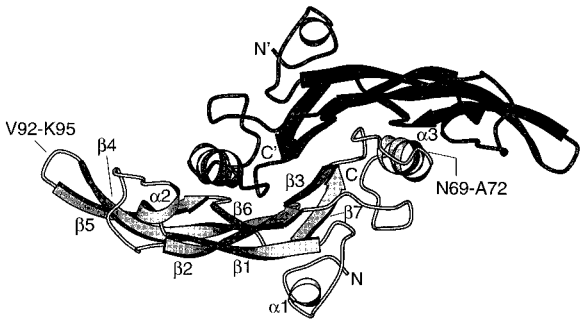

FIGURE 7: MOLSCRIPT (Kraulis, 1991) ribbon diagram representing the C^{α} backbone coordinates of the solution structure of the TGF-β1 homodimer. Regular elements of secondary structure are identified, along with the N and C termini of the two monomers.

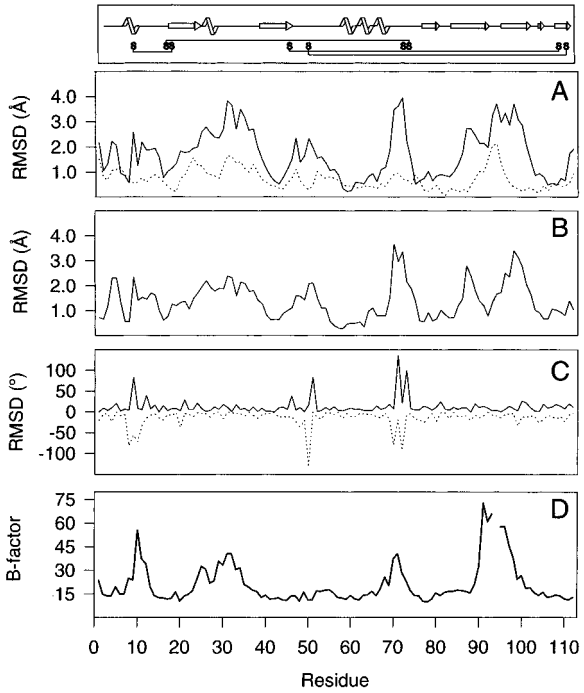

FIGURE 8: Comparison of the experimentally determined TGF-β1 solution structure and simulated TGF-β2 model structure versus the X-ray crystal structure of TGF-β2 (Daopin et al., 1993). (A) The RMSD in backbone atom positions for the restrained minimized average structures versus the 1.8 Å X-ray crystal structure of TGF-β2 when loop regions (residues 8−14, 29−39, 68−75, and 87−102) are omitted in the superposition. The solid line reflects residue-based RMSDs for the experimentally determined TGF-β1 homodimer, whereas the dashed line corresponds to the simulated TGF-β2 model structure. (B) The numerical difference between the backbone RMSDs shown in panel A. (C) RMSD differences in ϕ and ψ for the restrained minimized average TGF-β1 solution structure versus the 1.8 Å X-ray crystal structure of TGF-β2. RMSD differences in ϕ are indicated by a solid line and are shown as positive deviations, whereas RMSD differences in ψ are indicated by a dashed line and are shown as negative deviations. (D) Thermal *B*-factors reported for the 1.8 Å crystal structure of TGF-β2.

Conserved Structural Features of TGF-β. Overall, the backbone conformations of TGF-β1 and TGF-β2 are very similar, as shown by the small RMSD differences in ϕ/ψ throughout most of the protein backbone (Figure 8C and Figure 9). An analysis of the restrained minimized NMR structure by the PROCHECK program (Laskowski et al., 1993) indicated that 57% of the non-glycine residues occupy the most favorable regions of the Ramachandran map and

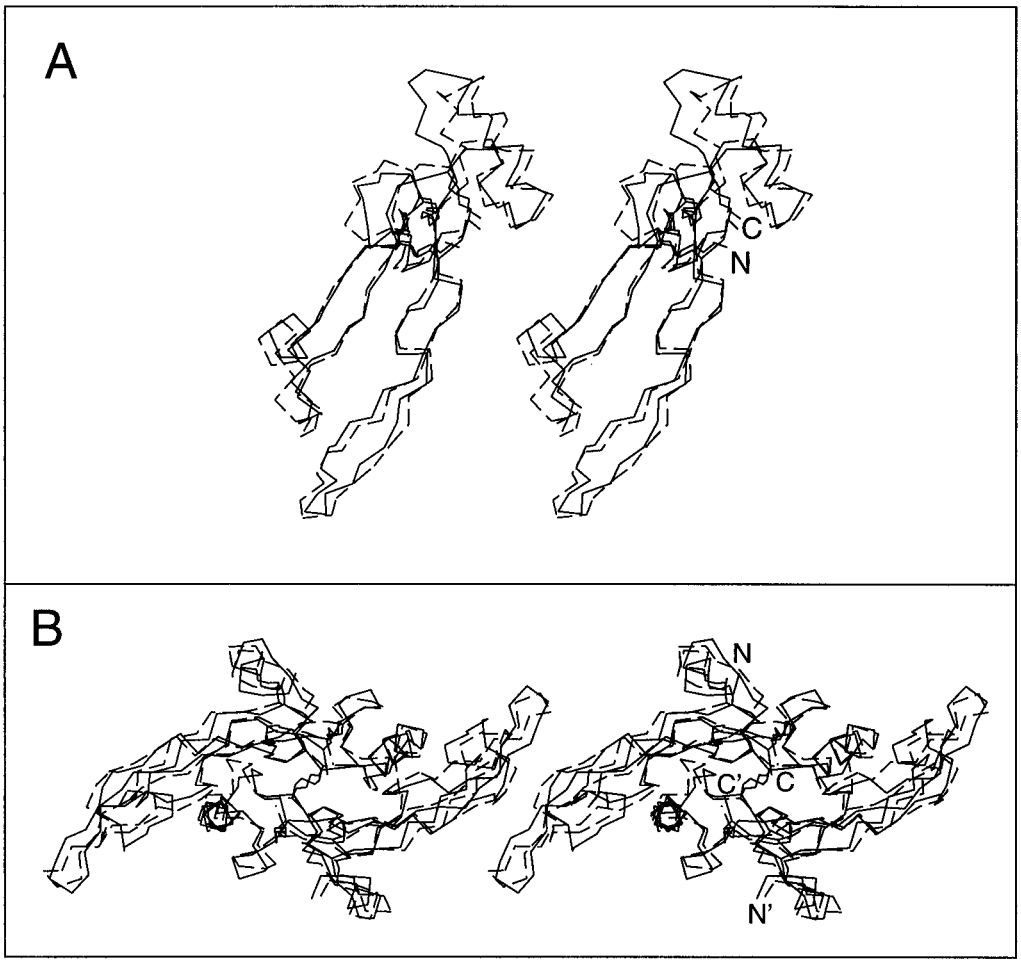

FIGURE 9: Stereo overlay of the backbone heavy atom coordinates of the restrained minimized average solution structure of TGF-β1 and the 1.8 Å X-ray crystal structure of TGF-β2. (A) An overlay of backbone heavy atoms (N, C^{α}, C) for the TGF-β monomer (TGF-β1, solid; TGF-β2, dashed) using the backbone coordinates of residues 1−68 and 76−112 in the alignment. (B) An overlay of C^{α} coordinates for the TGF-β homodimer (TGF-β1, solid; TGF-β2, dashed) using the backbone coordinates of residues 1−8, 15−28, 40−68, 76−86, and 103−112 from each monomer in the alignment.

that there are no residues found in the disallowed regions. Unusual conserved features of the protein backbone conformation include left-handed helical values for N103 and S53 and a *cis* peptide bond linking E35-P36. Moreover, the solution structure of TGF-β1 generally conforms to previous expectations regarding the structural significance and location of the 31 amino acid differences (Daopin et al., 1992, 1993; Schlunegger & Grütter, 1992, 1993). The dimer interface, for example, is similar in size (TGF-β1, 1377 $Å^2$; TGF-β2, 1316 $Å^2$) and is composed of a set of hydrophobic residues that are nearly identical: residues L20, I22, F24, L28, W30, Y39, A41, F43, M104, and V106 from the first monomer and Y58, V61, L62, L64, Y65, H68, A72, and A74 from the symmetry-related monomer. Among these, only Y58 and H68 differ, TGF-β2 having a histidine at position 58 and an isoleucine at position 68. Neither amino acid substitution appears to be structurally significant, as each maintains similar side chain χ_1 torsion angles and both display a similar array of hydrophobic contacts in the two structures. Moreover, the backbone nitrogen and side chain −OH group of Y58 in TGF-β1 is within hydrogen-bonding distance of the same array of acceptors that were reported for H58 in the TGF-β2 crystal structure (Daopin et al., 1993; Schlunegger & Grütter, 1993). Other examples of structural conservation between the two isoforms include the two remaining hydrophobic clusters, the larger of which is formed by residues which comprise fingers 1/2 and 3/4 and the smaller of which is comprised of several residues near the N-terminus. The former contains residues F24, L28, W30, W32, I33, and Y39 and residues L86, I88, Y90, L101, and M104 from fingers 1 and 2 and 3 and 4, respectively, and as indicated previously, appears to be important for stabilizing the relative orientation of the fingers. The latter may be important in anchoring the N-terminal portion of the molecule to the rest of the protein and is composed of L2, Y6, and W52 and the disulfide moiety of the C7/C16 disulfide bond. Neither hydrophobic pocket contains any residues that differ between TGF-β1 and TGF-β2.

Another feature that the TGF-β1 NMR structure and X-ray crystal structures of TGF-β2 has in common is a similar Lee and Richards (1971) type of solvent-accessible surface area (ASA, Figure 10). One interesting conserved feature is a hydrophobic surface patch comprised of residues W30, W32, Y90, V92, and L101 from the first monomer and Y50 and I51 from the symmetry-related monomer. Each of these residues is conserved among TGF-β1 and TGF-β2, except V92, which is an isoleucine residue in TGF-β2. The cysteine residues of TGF-β1 and TGF-β2 provide another example of a conserved structural feature; all are predominantly in the interior of the protein and are among the most well-defined portions of the protein backbone in both isoforms (Figure 6A). As in the X-ray structure of TGF-β2, C77 is

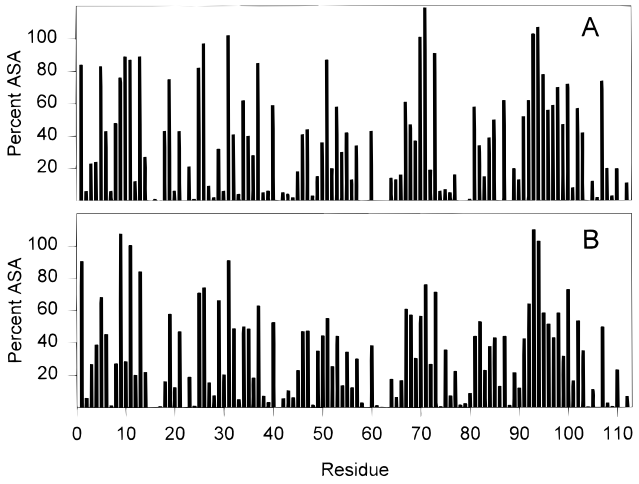

FIGURE 10: Comparison of the Lee and Richards type of solvent-accessible surface area for the minimized average solution structure of TGF-β1 and the 1.8 Å X-ray crystal structure of TGF-β2. Solvent-accessible surface areas were calculated using a water probe radius of 1.4 Å and are reported as a fraction of the solvent-accessible surface areas previously reported for Gly-X-Gly tripeptides (Chothia, 1975). Panel A includes data for the solution structure of TGF-β1, whereas panel B describes the surface of the crystal structure of TGF-β2 (Daopin et al., 1993).

exceptional in that its side chain atoms are considerably more exposed to solvent compared to those of the other cysteines. Finally, we note that while a similar pattern of ASA is observed between TGF-β1 and TGF-β2 (Figure 10), the distribution of charged residues on the surface of the two proteins differs (Figure 11).

Comparison of the Backbone Topology with the Crystal Structure of TGF-β2. The overall fold of the TGF-β1 homodimer closely resembles that of the X-ray structures of TGF-β2 (Daopin et al., 1992, 1993; Schlunegger & Grütter, 1993, 1993). A plot of the RMSD difference in backbone atom positions for the restrained minimized average NMR structure ($\langle$TGF-β1$\rangle_r$) and the 1.8 Å resolution X-ray structure of TGF-β2, presented in Figure 8, shows that there is generally good agreement between the two structures, the overall RMSD being 1.90 Å when all residues are considered and 1.39 Å when the disordered regions (8−14, 29−39, 68−75, and 87−102) are excluded in the alignment (Table 2). The close correspondence of the backbone topology is also demonstrated by the small pairwise RMSD differences in ϕ and ψ (Figure 8C). Interestingly, the precision of the TGF-β1 structure on a per residue basis (Figure 6A) generally parallels the pairwise RMSD difference between $\langle$TGF-β1$\rangle_r$ and the 1.8 Å X-ray structure of TGF-β2 (Figure 8A). This indicates that the structures, overall, are largely similar and that a component of the observed deviation between the two structures is the limited precision of the NMR structures.

In order to obtain a more quantitative basis for such a conclusion, we have estimated the accuracy of the calculated TGF-β1 structures by calculating families of TGF-β2 structures using the DG/SA protocol and a set of interproton distance restraints derived from the 1.8 Å resolution X-ray structure of TGF-β2. A plot of the RMSD difference in backbone atom positions for the restrained minimized average structure ($\langle$TGF-β2$\rangle_r$) and the 1.8 Å resolution X-ray structure of TGF-β2, Figure 8A, shows that the RMSD generally parallels that found for TGF-β1. The overall RMSD is 0.81 Å when all residues are considered and 0.71 Å when disordered regions (8−14, 29−39, 68−75, and 87−102) are excluded in the alignment (Table 2), lower than that found in the case of TGF-β1 (1.90 and 1.39 Å, respectively). Regions of the sequence which display deviations larger than ca. 0.5 Å include F8−Q12, R27−E35, P47−D55, H68−A72, P85−I88, and G93−Q100. A similar result is obtained when the $\langle$TGF-β1$\rangle_r$ and $\langle$TGF-β2$\rangle_r$ structures are compared on a pairwise basis (Figure 8B). Presumably, the larger RMSD observed when comparing the TGF-β1 NMR structure with the TGF-β2 X-ray structure is due to (a) small errors present in the set of TGF-β1 restraints, (b) the limited accuracy of the X-ray structure, or (c) structurally significant differences between the two structures. We now compare the TGF-β1 and TGF-β2 structures and identify the reasons for the differences in coordinates that are found.

(A) Residues F8−E12. The first region of the structure where significant RMSD and ϕ/ψ differences are observed includes residues 8−12. As previously noted, in solution these residues are solvent exposed and flexible on the nanosecond time scale. Disorder of the backbone in this region of the sequence was also reported for crystalline TGF-β2 on the basis of elevated crystallographic *B*-factors (Figure 8D). Thus, in this portion of the structure, deviations between the NMR structure of TGF-β1 and the X-ray structure of TGF-β2 result from local flexibility that limits the precision within which the structures can be defined.

(B) Residues R27−E35, P85−I88, and G93−Q100. Residues R27−E35 along with P85−I88 and G93−Q100 comprise the tips of the four fingers. As noted previously, relative to the X-ray structure of TGF-β2, the tips of the fingers undergo a hinge bending where the tips of the fingers move inward toward the palm of the hand. Evidence to support such a conclusion follows from the observation that comparable RMSDs in backbone atom positions are found when the monomers are aligned using residues which comprise either the four fingers (24−39 and 77−112, RMSD 1.38 Å) or the palm and heel of the hand (1−23 and 40−76, RMSD 1.47 Å). Here, we present two possible explanations that could account for the observed hinge bending. (1) The first follows from our observation that this behavior was observed for both the experimentally determined TGF-β1 and the simulated TGF-β2 structures (Figure 8A), thereby indicating that this phenomenon is to some extent arising from the limited accuracy of the NMR structures. (2) The second possibility is that lattice contacts reported in the 1.8 Å X-ray structure of TGF-β2 (Daopin et al., 1993) may serve to "freeze out" a single conformation of fingers 1−4 which differs from that populated in solution. Reported lattice contacts among residues in the four fingers include hydrogen bonds between A1 N−E35 $O^{\epsilon 2}$ and Y6 OH−T87 O, as well as van der Waals contacts between A1−E35, Y6−E35, Y6−L89, Y6−I98, N10−T87, Q12−T87, N69−P96, and E71−K94.

Finally, we note even if hinge bending is not an artifact, the local conformation of residues located both within and flanking the turn linking fingers 3 and 4 does not differ significantly between the NMR structure of TGF-β1 and the X-ray structure of TGF-β2: the RMSD in backbone atom positions when the X-ray structure of TGF-β2 and the NMR structure of TGF-β1 are aligned using residues 91−98 of a single monomer is 0.43 Å. As indicated previously, in TGF-

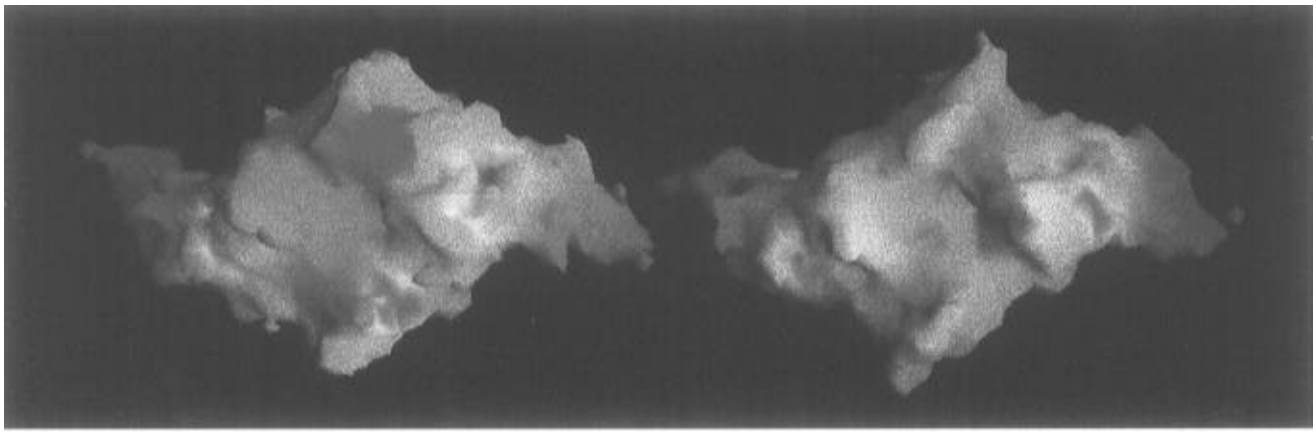

FIGURE 11: Electrostatic surface potential of the NMR solution structure of TGF-β1 (left) and the 1.8 Å X-ray structure of TGF-β2 (right) as calculated using GRASP (Nicholls et al., 1991). The two structures have been aligned using the backbone coordinates of residues 1−8, 15−28, 40−68, 76−86, and 103−112 and are displayed with the plane of the dimer interface perpendicular to the page. Charges in each of the structures have been calculated at pH 7.0 and are contoured at a constant electrostatic potential: negatively charged areas are identified in red, positively charged areas are identified in blue, and neutral areas are left white.

β1 residues 92−95 comprise a well-ordered, but flexible type II′ β-turn in solution. A similar turn conformation was reported in the 1.8 Å X-ray structure of TGF-β2 (Daopin et al., 1993). There is, however, significant coordinate error, ca. 0.5−1.0 Å, associated with the backbone conformation of residues 91−98 reported in this structure. Thus, while the overall conformation of fingers 3 and 4 is similar in TGF-β1 and TGF-β2, given the weak electron density found for residues 92−95 in the X-ray structure, it is concievable, for example, that the turn geometry differs between TGF-β1 and TGF-β2. This is potentially significant as a recent report (Bach et al., 1996) has shown that the specificity of a certain cyclic peptide for integrin receptors can be altered in a systematic fashion by residue changes within the peptide which promote a swap from a type II to type I β-turn.

(C) Residues 47−53. Substantial RMSD and ϕ/ψ differences are also reported for residues 47−53. These appear to be due to the previously noted observation that this amino acid segment undergoes motion on the millisecond time scale in solution, leading to resonance broadening and limiting our ability to assign several of the backbone resonances in this part of the sequence. In contrast, the 1.8 Å X-ray structure of TGF-β2 displays low thermal *B*-factors (Figure 8D) and well-defined electron density throughout this portion of the sequence. Crystal packing forces may in part be responsible for such ordering, as several residues in this portion of the sequence participate in intermolecular interactions in the crystalline lattice: the backbone oxygen of Y50 is hydrogen bonded to $N^{\delta 2}$ of N42, the side chain C^{β} of L51 is within van der Waals contact distance of Y21 C^{ϵ}, and the $C^{\zeta 3}$ of W52 contacts P36 C^{β} and C^{γ}, respectively.

(D) Residues 68−72. The last region of the sequence with coordinate and ϕ/ψ RMSDs that we consider is the turn segment spanning residues 69−72. In solution, portions of the 3D ^{15}N-separated, 3D ^{15}N-separated ^{12}C-filtered, and 3D ^{15}N/^{13}C-separated NOESY data sets obtained using [U-^{15}N] and scheme D TGF-β1, Figure 4, reveal weak $H^{N}_{i}-H^{\beta}_{i+4}$, strong $H^{\alpha}_{i+1}-H^{N}_{i+2}$, and medium $H^{N}_{i+3}-H^{N}_{i+4}$ NOE cross peaks for the four-residue amino acid segment, N69-P70-G71-A72. Such NOE patterns are consistent with a type II β-turn but not with the conformation reported in the 1.8 Å X-ray structure of TGF-β2 (Daopin et al., 1993), which is a type I β-turn. One plausible explanation that accounts for this difference in conformation is the substitution of glycine at position 71 in TGF-β1 by glutamate in TGF-β2. In TGF-β1, G71 adopts unusual ϕ/ψ values, $\phi = +120°$ and $\psi = +12°$, acceptable for a glycine residue but not allowed for a glutamate. In contrast, E71 in TGF-β2 adopts backbone dihedral angles of $\phi = -66°$ and $\psi = -23°$ that fall well within the allowed region of the Ramachandran surface. Thus, this structural difference appears to be largely mediated by the conformational freedom of the residue located at the third turn position: the glycyl residue permits the backbone to adopt a type II β-turn, whereas the more restrained glutamate causes the chain to adopt the type I β-turn configuration.

(E) N-Terminus. The N-terminal residues, 2−4, exhibit elevated T_2 and diminished NOE values, indicating the presence of significant internal motion and flexibility in solution. In the 1.8 Å X-ray structure of TGF-β2, residues 1−4 exhibit low *B*-factors and well-defined electron density. Crystal contacts may be responsible for these differences as a hydrogen bond was identified between the backbone N of A1 and the side chain $O^{\epsilon 2}$ of E35 in the neighboring molecule of the crystal lattice.

Potential Structure−Function Relationships. The TGF-β isoforms mediate their effects by binding to cell surface receptors. Two of these, TβRI and TβRII, possess a disulfide-bonded extracellular domain which recognizes the different TGF-β isoforms, a hydrophobic transmembrane domain, and a cytoplasmic domain which has been shown by amino acid homology to belong to the serine/threonine kinase class of signal transducing proteins. Several functional regions of TGF-β, thought to be important in receptor binding, have been mapped by generating TGF-β chimeras or deletion mutants and measuring their ability to inhibit growth of various cell lines that are known to display a differential response to the various TGF-β isoforms. Regions of the TGF-β sequence that have been mapped by this approach include the extended structure formed by residues L45−P47 (Burmester et al., 1993), the β-turn segment from H69 to S73 (Qian et al., 1994), and the C-terminal portion of the molecule including the type II′ β-turn of fingers 3 and 4 (Qian et al., 1995).

Previously, we indicated two possible mechanisms by which differences in amino acid sequence might translate to

differences in function among the three TGF-β isoforms: local structural differences in exposed loop regions or a preserved backbone topology with alterations in residues located on the protein surface. At the current level of precision, the conformation of a large portion of the TGF-β1 backbone conformation is essentially indistinguishable from that reported in the X-ray crystal structures of TGF-β2. One clear exception to this finding is made by residues 69−72, which adopt a well-defined type II β-turn in TGF-β1 and a type I β-turn in TGF-β2. This region may, therefore, provide one example of the former mechanism as there is evidence that such structural differences may be functionally relevant: deletion of residues 69−73 of TGF-β1 diminishes the potency of the molecule to inhibit the growth of LS513 colorectal cells several hundredfold to a level comparable to that of TGF-β2 (Qian et al., 1994). Since recently obtained receptor binding data indicate that this region is not important for high-affinity binding to the type II receptor (Qian et al., 1995), it has been proposed that this region of the molecule may, therefore, be important for recognition by another class of receptors, or another, as yet unidentified molecule, that binds TGF-β1 with high affinity.

A clear example of the second mechanism is provided by residues 45−47. These residues form an extended structure and were previously identified as being functionally important in mediating efficient growth inhibition of cultured fetal bovine heart endothelial (FBHE) cells by TGF-β1 but not by TGF-β2. It has been demonstrated that efficient sequestration of TGF-β2, but not TGF-β1, by α2 macroglobulin (α_2M) is responsible for such functional differences since TGF-β2 mutants in which A45 and A47 are simultaneously substituted by the corresponding amino acids of TGF-β1, L45 and P47, increase the potency of the molecule in FBHE growth inhibition assays to a level comparable to that of TGF-β1 (Burmester et al., 1993). In the solution structure of TGF-β1, the amino acid segment, L45-G46-P47, serves to connect the C-terminal portion of β2 with the N-terminal end of α3. Moreover, the backbone conformation is very similar to that reported in the X-ray structure of TGF-β2 for A45-G46-A47, as revealed by small RMSD differences in backbone atom coordinates and ϕ/ψ dihedral angles in this portion of the structure (Figure 8A,C and Figure 9). In addition to the similarity of the backbone conformation, we also note that G46 adopts ϕ/ψ angles that are typically disallowed for other residue types but compatible with glycyl residues. Thus, one mechanism by which α_2M may discriminate TGF-β1 and TGF-β2 is by differences in their surfaces in this portion of the structure. Lee and Richards type of ASA calculations of TGF-β1 and TGF-β2 reveal that although amino acids in this segment display a similar pattern of relative solvent exposure (Figure 10), the absolute magnitude of exposed hydrophobic surface area is larger in TGF-β1 as leucine and proline residues have considerably larger hydrophobic surface areas compared to alanine.

A final example illustrates a case in which both mechanisms may operate. This region includes the extended portions of fingers 3 and 4, including the type II′ β-turn formed by residues 92−95. Recent evidence indicates that residues 92, 94, and 98 are critical for efficient recognition of the TGF-β by the extracellular domain of the type II receptor: V92, R94, and V98 of TGF-β1 and TGF-β3 lead to high affinity binding by the extracellular domain of the type II receptor, whereas I92, K94, and I98 of TGF-β2 result in a TGF-β−TβRII receptor complex that is 2 orders of magnitude lower in stability (Qian et al., 1995). A comparison of the local conformation and flexibility of fingers 3 and 4 in TGF-β1 with the X-ray structure of TGF-β2 showed that there was reasonably good agreement between the two, even though the orientation of the fingers was not precisely defined. Thus, one possibility is that TβRII has the ability to discriminate the surface of TGF-β1 and TGF-β2, based largely on the these two conservative amino acid substitutions. Alternatively, on the basis of the very weak electron density for residues 92−95 reported in the X-ray structure, it is conceivable that residues 92−95 adopt different β-turn types between TGF-β1 and TGF-β2 and that this is the structural basis by which TβRII is able to distinguish TGF-β1 and TGF-β2.

Another solvent-exposed loop region that has not been fit into any of the above categories due to its inherent disorder in solution is the extended loop segment from residues 8 to 14 bridging the C7/C16 disulfide. It is interesting that this is the longest region of the sequence where TGF-β1 and TGF-β2 differ: S9-S10-T11-E12-K13 in TGF-β1 and R9-N10-V11-Q12-D13 in TGF-β2. A further interesting observation is that the C7/C16 is found only among the TGF-βs and the activins but not in other members of the TGF-β superfamily. Further work is required to ascertain if residues in this portion of the sequence either (a) are readily mutated because they contribute little to the structural integrity of the protein or (b) are tailored to meet the functional roles carried out by the various isoforms and families within the TGF-β superfamily.

Summary. The NMR structure of TGF-β1 presented herein is the first high-resolution three-dimensional solution structure that has been reported for a member of the TGF-β family of peptide growth factors. This structure provides information that complements that obtained from two independently determined crystal structures of TGF-β2 (Daopin et al., 1992, 1993; Schlunegger & Grütter, 1992, 1993) and the X-ray structure of OP-1 (Griffith et al., 1996).

A comparison of the TGF-β1 and TGF-β2 structures shows that the backbone conformation of TGF-β1 in solution closely resembles that of crystalline TGF-β2. There are, however, several notable differences in local structure and flexibility which may be related to differences in receptor binding and signal transduction. The clearest example of a structural difference is found for residues 69−72, which link the C terminus of the longest α-helix with the extended C-terminal portion of the protein. These residues form a type II β-turn in TGF-β1 and a type I β-turn in TGF-β2. This observation may be of functional significance because this portion of the molecule has been implicated (Qian et al., 1994) in receptor binding.

Residues 92−95 have recently been shown to be important in recognizing the extracellular domain of the TGF-β type II receptor (Qian et al., 1995). These residues form a surface-exposed turn in both TGF-β structures. However, in TGF-β1 the turn is a flexible but well-defined type II′ β-turn, while in TGF-β2 the turn is found among the most poorly defined portions of the electron density. Hence, it is possible that different turn conformations contribute to the 100-fold higher affinity of the TGF-β type II receptor for TGF-β1 over that of TGF-β2.

In contrast with the residues just discussed, residues 45−47 have essentially identical, well-ordered backbone conformations in TGF-β1 and TGF-β2. However, replacement of A45 and A47 in TGF-β2 with leucyl and prolyl residues (which are found at these positions in TGF-β1), respectively, yields a TGF-β2 chimera having TGF-β1 type binding affinity for α_2M. The different affinities of the TGF-β isoforms for α_2M are therefore ascribed to the fact that an exposed hydrophobic surface is presented by residues 45 and 47 in TGF-β1 but not in TGF-β2.

Two other regions where TGF-β1 and TGF-β2 may have local structural differences, as a consequence of chain flexibility in solution, include residues 8−12 and 48−53. At this time it is not possible to relate possible differences in structure and dynamics in these regions of the isoforms to function, since more information is required to identify which amino acids in the TGF-β isoforms are crucial for function. As this information is amassed, the structural information now available about the TGF-β1 and TGF-β2 isoforms should provide further insights as to the molecular basis for differences in function.

ACKNOWLEDGMENT

We gratefully acknowledge Frank Delaglio for developing and supporting the NMR processing package, nmrPipe/nmrDraw, and Dan Garrett for the data analysis package, PIPP. We also thank Ad Bax and Stephan Grzesiek, who acquired the 2D NOESY and 3D HAHB and HNHB data sets at 600 MHz, and John Kuszewski, who generously provided advice on using X-PLOR.

SUPPORTING INFORMATION AVAILABLE

A listing of the amino acid composition of the CHO cell growth medium, a summary of the NMR data used to determine the solution structure of TGF-β1, and experimentally determined relaxation parameters, distance restraints, and coupling constants for TGF-β1 (eight tables) plus NMR spectra which were used in the analysis of NOE data and for the determination of side-chain coupling constants (four figures) (27 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.

Amatayakul-Chantler, S., Qian, S. W., Gakenheimer, K., Bottinger, E. P., Roberts, A. B., & Sporn, M. B. (1994) *J. Biol. Chem. 269*, 27687−27691.

Archer, S. J., Bax, A., Roberts, A. B., Sporn, M. B., Ogawa, Y., Piez, K. A., Weatherbee, J. A., Tsang, L.-S. M., & Lucas, R. (1993a) *Biochemistry 32*, 1152−1163.

Archer, S. J., Bax, A., Roberts, A. B., Sporn, M. B., Ogawa, Y., Piez, K. A., Weatherbee, J. A., Tsang, M. L.-S., Lucas, R., Zheng, B.-L., Wenker, J., & Torchia, D. A. (1993b) *Biochemistry 32*, 1164−1171.

Archer, S. J., Vinson, V. K., Pollard, T. D., & Torchia, D. A. (1993c) *Biochemistry 32*, 6680−6687.

Bach, A. C., Espina, J. R., Jackson, S. A., Stouten, P. F. W., Duke, J. L., Mousa, S. A., & DeGrado, W. F. (1996) *J. Am. Chem. Soc. 118*, 293−294.

Bax, A., & Subramanian, S. (1986) *J. Magn. Reson. 67*, 565−569.

Bax, A., Clore, G. M., Driscoll, P. C., Gronenborn, A. M., & Ikura, M. (1990) *J. Magn. Reson. 87*, 620−627.

Bax, A., Max, D., & Zax, D. (1992) *J. Am. Chem. Soc. 114*, 6923−6925.

Boyd, F. T., & Massagué, J. (1989) *J. Biol. Chem. 264*, 2272−2278.

Brünger, A. T. (1992) *X-PLOR Manual Version 3.1: A System for X-ray Crystallography and NMR*, Yale University, New Haven, CT.

Burmester, J. K., Qian, S. W., Roberts, A. B., Huang, A., Amatayakul-Chantler, A., Suardet, L., Odartchenko, N., Madri, J. A., & Sporn, M. B. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 8628−8632.

Burt, D. W., & Law, A. S. (1994) *Prog. Growth Factor Res. 5*, 99−118.

Cheifetz, S., Bellon, T., Cales, C., Vera, S., Bernabeu, C., & Massagué, J. (1992) *J. Biol. Chem. 267*, 19027−19030.

Chothia, C. (1975) *J. Mol. Biol. 105*, 1−14.

Clore, G. M., Gronenborn, A., & Bax, A. (1991a) *J. Biomol. NMR 1*, 13−22.

Clore, G. M., Kay, L. E., Bax, A., & Gronenborn, A. M. (1991b) *Biochemistry 30*, 12−18.

Danielpour, D., & Sporn, M. B. (1990) *J. Biol. Chem. 265*, 6973−6977.

Daopin, S., Piez, K. A., Ogawa, Y., & Davies, D. R. (1992) *Science 257*, 369−374.

Daopin, S., Li, M., & Davies, D. R. (1993) *Proteins: Struct., Funct., Genet. 17*, 176−192.

Delaglio, F., Grzesiek, S., Vuister, G., Zhu, W., Pfeifer, J., & Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.

Derynck, R. (1994) *Trends Biochem. Sci. 19*, 548−553.

Garrett, D. S., Powers, R., Gronenborn, A. M., & Clore, G. M. (1991) *J. Magn. Reson. 95*, 214−220.

Griffith, D. L., Keck, P. C., Sampath, T. K., Rueger, D. C., & Carlson, W. D. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 878−883.

Grzesiek, S., & Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291−6293.

Grzesiek, S., & Bax, A. (1993) *J. Am. Chem. Soc. 115*, 12593−12594.

Grzesiek, S., Vuister, G. W., & Bax, A. (1993a) *J. Biomol. NMR 3*, 487−493.

Grzesiek, S., Anglister, J., & Bax, A. (1993b) *J. Magn. Reson. 101*, 114−119.

Grzesiek, S., Kuboniwa, H., Hinck, A. P., & Bax, A. (1995) *J. Am. Chem. Soc. 117*, 5312−5315.

Jennings, J. C., Mohan, S., Widstrom, R., & Baylink, D. (1988) *J. Cell Physiol. 137*, 167−172.

Kamath, U., & Shriver, G. W. (1989) *J. Biol. Chem. 264*, 5586.

Kay, L. E., Marion, D., & Bax, A. (1989) *J. Magn. Reson. 84*, 72−84.

Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A., & Torchia, D. A. (1992) *J. Magn. Reson. 97*, 359−375.

Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1−6.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.

Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379−400.

Live, D. H., Davis, D. G., Agosta, W. C., & Cowburn, D. (1984) *J. Am. Chem. Soc. 106*, 1934−1941.

López-Casillas, F., Cheifetz, S., Doody, J., Andres, J. L., Lane, W. S., & Massagué, J. (1991) *Cell 67*, 785−795.

Marion, D., Ikura, M., Tsuchdin, R., & Bax, A. (1989) *J. Magn. Reson. 85*, 393−399.

Massagué, J. (1990) *Annu. Rev. Cell Biol. 6*, 597−641.

Moren, A., Ichijo, H., & Miyazono, K. (1992) *Biochem. Biophys. Res. Commun. 189*, 356−362.

Nicholls, A., Sharp, K., & Honig, B. (1991) *Proteins: Struct., Funct., Genet. 11*, 281.

Nilges, M. (1993) *Proteins: Struct., Funct., Genet. 17*, 297−309.

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

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1988) *Numerical Recipes in C*, Cambridge University Press, Cambridge, U.K.

Qian, S. W., Burmester, J. K., Merwin, J. R., Madri, J. A., Sporn, M. B., & Roberts, A. B. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6290−6294.

Qian, S. W., Burmester, J. K., Sun, P. D., Huang, A., Ohlson, D. J., Suardet, L., Flanders, K. C., Davies, D., Roberts, A. B., & Sporn, M. B. (1994) *Biochemistry 33*, 12298−12304.

Qian, S. W., Burmester, J. K., Tsang, M. L.-S., Weatherbee, J. A., Hinck, A. P., Ohlsen, D. J., Sporn, M. B., & Roberts, A. (1995) *J. Biol. Chem.* (submitted for publication).

Roberts, A. B., & Sporn, M. B. (1990) in *Peptide Growth Factors and Their Receptors* (Roberts, M. S. A., Ed.) pp 421−472, Springer-Verlag, Heidelburg, Germany.

Roberts, A. B., Anzano, M. A., Lamb, L. C., Smith, J. M., & Sporn, M. B. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 5339−5343.

Schlunegger, M. P., & Grütter, M. G. (1992) *Nature (London) 358*, 430−434.

Schlunegger, M. P., & Grütter, M. G. (1993) *J. Mol. Biol. 231*, 445−458.

Sibanda, B. L., & Thornton, J. M. (1991) *Methods Enzymol. 202*, 59−82.

Suardet, L., Gaide, A.-C., Calmes, J.-M., Sordat, B., Givel, G. C., Eliason, J. F., & Odartchenko, N. (1992) *Cancer Res. 52*, 3705−3712.

Vuister, G., & Bax, A. (1992*) J. Magn. Reson. 98*, 428−435.

Vuister, G., & Bax, A. (1993a) *J. Am. Chem. Soc. 115*, 7772−7777.

Vuister, G., & Bax, A. (1993b) *J. Magn. Reson., Ser. B 102*, 228−231.

Vuister, G., Wang, A. C., & Bax, A. (1993) *J. Am. Chem. Soc. 115*, 5334−5335.

Wrana, J. L., Attisano, L., Carcomo, J., Zentella, A., Doody, J., Laiho, M., Wang, X., & Massagué, J. (1992) *Cell 71*, 1003−1014.

Wrana, J. L., Attisano, L., Wiesner, R., Ventura, F., & Massagué, J. (1994) *Nature 370*, 341−347.

Wüthrich, K., Billetter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949−961.

BI9604946