Secondary Structure and Backbone Dynamics of Human Granulocyte Colony-Stimulating Factor in Solution†

Jörn M. Werner,‡ Alexander L. Breeze,§ Bo Kara,‖ Gina Rosenbrock,‖ Jonathan Boyd,‡ Nick Soffe,‡ and Iain D. Campbell*,‡

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K., and Protein Structure Laboratory and Biotechnology Department, Zeneca Pharmaceuticals, Mereside, Alderley Park, Cheshire SK10 4TG, U.K.

Received January 18, 1994; Revised Manuscript Received March 23, 1994⊛

ABSTRACT: The secondary structure and backbone dynamics of the cytokine, human granulocyte colony-stimulating factor (hG-CSF) have been determined by heteronuclear nuclear magnetic resonance (NMR) techniques. Virtually complete NH, $C^{\alpha}H$, $C^{\beta}H$ ^{15}N, $^{13}C\alpha$, and $^{13}C\beta$ assignment of the 175-residue recombinant protein, methionyl-[Cys-17-Ser]-hG-CSF, was achieved by use of three-dimensional (3D) heteronuclear ^{1}H–^{15}N and triple-resonance ^{1}H–^{15}N–^{13}C experiments. Spectra recorded at 750 MHz aided the assignment of severely overlapped regions. The structures of G-CSF from several species have recently been determined by X-ray diffraction [Hill, C. P., Osslund, T. D., & Eisenberg, D. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90,* 5167–5171; Lovejoy, B., Cascio, D., & Eisenberg, D. (1993) *J. Mol. Biol. 234,* 640–653]. Like several cytokines, hG-CSF has a four-helix topology (A–D) with overhand loop connections, but with an additional helical segment (A′) identified in the connection between helix A and helix B. The solution-state determination of the secondary structure is based on short- and medium-range NOEs, backbone *J*-couplings, and NH exchange data and is corroborated by $^{13}C\alpha$ secondary shifts. The helices are defined as follows: A, 10–38; A′, 44–53; B, 71–91; C, 102–123; D, 143–172. The dynamics of the amide backbone resonances, investigated using ^{1}H–^{15}N heteronuclear NMR, indicate a rigid protein core with some increased mobility in the AB loop and more pronounced mobility in the CD loop. Slow conformational exchange due to proline cis–trans isomerization is suggested by satellite resonances observed in two-dimensional heteronuclear spectra for some resonances from residues near prolines.

The colony-stimulating factors (CSFs[1]) and interleukins (ILs) form a diverse family of protein growth factors, many of whose members are involved in regulation of the proliferation, differentiation, and maturation of the various hemopoietic stem cell lines of the bone marrow (Clark & Kamen, 1987; Nicola, 1989). One of the most discriminating of this hemopoietic growth factor family is granulocyte colony-stimulating factor (G-CSF). Human G-CSF (hG-CSF) is a 174-residue protein of molecular weight 18 700, which is a highly specific growth and maturation factor for the granulocytic neutrophil lineage. The native protein carries a glycosyl moiety on Thr-133 whose enzymatic removal does not affect biological activity (Oh-eda et al., 1990).

With the recent publications of the three-dimensional structures of a number of hemopoietic growth factors and cytokines [for example, interferon β (IFN-β) (Senda et al., 1990, 1992), IL-2 (Bazan, 1992; McKay, 1992; Mott et al., 1992), IL-4 (Smith et al., 1992; Powers et al., 1992; Wlodawer et al., 1992), IL-5 (Milburn et al., 1993), granulocyte macrophage colony-stimulating factor (GM-CSF) (Diederichs et al., 1991; Walter et al., 1992), macrophage colony-stimulating factor (M-CSF) (Pandit et al., 1992), and growth hormone (GH) (Abdel-Meguid et al., 1987; de Vos et al., 1992)], a common fold is emerging for many of the members of this class, despite limited homology at the sequence level (Nicola, 1989; Parry et al., 1988). The core of this cytokine fold is an up–up–down–down, antiparallel, left-handed, four-α-helical bundle with a double-overhand loop topology. Secondary structure predictions (Bazan, 1990, 1991) suggested that G-CSF shares the double-overhand, four-helix bundle topology. This topology has been confirmed by the X-ray structures of hG-CSF (Hill et al., 1993), determined at 2.2-Å resolution, and those of canine and bovine G-CSFs, determined at 2.2 and 1.7 Å, respectively (Lovejoy et al., 1993). In spite of the relatively high resolution, no electron density was observed for some portions of the double-overhand loop regions, leaving both the precise local conformations of these sections of the polypeptide chain and their relationship to the remainder of the structure in doubt.

We have been engaged in a study of the solution structure of a recombinant hG-CSF by multinuclear NMR spectroscopy. Here, we report resonance assignments for the backbone atoms (NH, ^{15}N, $C^{\alpha}H$, and $^{13}C\alpha$) and for the majority of side-chain $C^{\beta}H$ and $^{13}C\beta$ resonances, including the two long loop regions. We describe the secondary structure of hG-CSF, together with a qualitative analysis of the backbone dynamics of the protein in solution as inferred from measurements of ^{1}H–^{15}N heteronuclear Overhauser enhancements (NOEs). The lim-

† This is a contribution from the Oxford Centre for Molecular Sciences, which is supported by SERC and MRC. This project is also partly funded by the LINK Protein Engineering programme. J.M.W. was supported by the German National Government Foundation and later by a fellowship from the German Academic Exchange Service from funds of the German Federal Ministry for Research and Technology.

* Address correspondence to this author.
‡ University of Oxford.
§ Protein Structure Laboratory, Zeneca Pharmaceuticals.
‖ Biotechnology Department, Zeneca Pharmaceuticals.

⊛ Abstract published in *Advance ACS Abstracts,* May 1, 1994.

[1] Abbreviations: hG-CSF, human granulocyte colony-stimulating factor; NMR, nuclear magnetic resonance; IFN-β, interferon β; IL-2, interleukin 2; IL-4, interleukin 4; IL-5, interleukin 5; GM-CSF, granulocyte macrophage colony-stimulating factor; M-CSF, macrophage colony-stimulating factor; GH, growth hormone; HPLC, high-performance liquid chromatography; NOE, nuclear Overhauser effect; HMQC, heteronuclear multiple-quantum correlation; HSQC, heteronuclear single-quantum correlation; NOESY, nuclear Overhauser enhancement spectroscopy; HOHAHA, homonuclear Hartmann–Hahn spectroscopy; 3D, three-dimensional; RF, radio frequency.

0006-2960/94/0433-7184$04.50/0

ited chemical shift dispersion, somewhat broad line widths, and very high content of leucine residues (33), together with the comparatively large size of hG-CSF, have hampered spectral assignment by ^1H–^{15}N 3D NMR methods alone and have necessitated a combination of approaches, including amino acid-specific labeling and ^1H–^{15}N–^{13}C triple-resonance 3D techniques. A secondary structure description of hG-CSF derived from NMR data was also recently reported by Zink et al. (1992).

EXPERIMENTAL PROCEDURES

Expression and Purification of Recombinant Human G-CSF. A recombinant hG-CSF containing a Cys-17 to Ser mutation and an additional N-terminal Met residue was expressed in insoluble form, accumulating as inclusion bodies in the cytoplasm of the prototrophic *Escherichia coli* host strain MG 1655 (CGSC 6300, *E. coli* Genetic Stock Center, Yale University, New Haven, CT; Guyer et al., 1981). The protein was expressed constitutively from the tryptophan promoter. Correctly folded protein was obtained following dissolution of the inclusion bodies in buffer containing *N*-lauroyl sarcosinate and purification by gel filtration chromatography and/or reversed-phase high-performance liquid chromatography (HPLC) (Souza et al., 1986). Uniformly ^{15}N- and ^{15}N,^{13}C-labeled material, accumulating at ca. 30% of the microbial protein, was produced from cultures grown in minimal medium containing $(^{15}NH_4)_2SO_4$ and/or [U-^{13}C]glucose as the sole nitrogen and carbon sources. Material (accumulating at ca. 20% of the microbial protein) specifically labeled with [^{15}N]Leu and [^{13}C]Ala [labeled at the carbonyl (C′) position] was produced from cultures grown in M9 minimal medium supplemented with labeled L-leucine (300 mg/L) and L-alanine (300 mg/L). Incorporation of the labeled amino acids was maximized by supplementing the medium with an excess of all other amino acids and by reduction of the incubation time from 16 h (used for the uniformly ^{15}N-labeled sample) to 4.5 h. Sample purity was assessed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis, reversed-phase HPLC, and, in some cases, capillary zone electrophoresis. The recombinant protein was biologically active in a colony stimulating assay.

Protein samples were prepared at 1.6–1.8 mM concentration (except the specifically labeled sample: 0.3 mM) in 0.5–0.6 mL of 50 mM deuterioacetic acid and 90% H_2O/10% D_2O or 100% D_2O, with a final pH (uncorrected) of 3.9. Because of the limited solubility of the lyophilized protein, buffer exchange from H_2O to D_2O was effected by repeated dilution and reconcentration in C10-Centricon ultrafiltration devices (Amicon) at 4 °C.

NMR Spectroscopy. All ^1H–^{15}N 2D and 3D NMR experiments except the 2D ^1H–^{15}N HMQC-J, the 2D ^1H-detected ^1H–^{15}N heteronuclear NOE experiments, and the NH exchange experiments were performed at 500 MHz (^1H) on a spectrometer comprising an Oxford Instruments magnet, home-built radio frequency (RF) electronics, and Bruker probes interfaced to a GE-Nicolet 1280 computer. The 2D ^1H-detected ^1H–^{15}N heteronuclear NOE and NH exchange experiments and the two triple-resonance 3D experiments, CBCA(CO)NH and HBHA(CBCACO)NH, were performed at 600 MHz (^1H) on an unmodified three-channel Varian Unity 600 spectrometer equipped with waveform generators and a triple-resonance probe. A 2D ^1H–^{15}N HSQC and an ^1H–^{15}N HMQC-J experiment were also acquired on the 750-MHz spectrometer of the Oxford Centre for Molecular Sciences (OCMS). This instrument comprises an Oxford Instruments 17.6-T magnet, probes and RF electronics designed and built in-house, and a GE Omega acquisition console and PSG pulse programmer. The sample temperature for NMR experiments was 30 or 32 °C.

2D ^1H–^{15}N HSQC (Bodenhausen & Ruben, 1980; Norwood et al., 1990; Bax et al., 1990) spectra were recorded at 500 and 750 MHz on the uniformly ^{15}N-labeled sample with acquisition times of 219.5 (t_1, ^{15}N) and 169.8 ms (t_2, NH) [500 MHz] and 219.5 (t_1, ^{15}N) and 112.6 ms (t_2, ^1H) [750 MHz]. A similar experiment was recorded at 500 MHz for the sample selectively labeled with [^{15}N]Leu and [$^{13}C'$]Ala. The H_2O resonance was suppressed in these experiments by application of a spin–lock purge pulse (Messerle et al., 1989). 3D ^1H–^{15}N NOESY–HMQC (Kay et al., 1989a) and ^1H–^{15}N HOHAHA–HMQC (Driscoll et al., 1990) spectra were recorded on the uniformly ^{15}N-labeled sample with mixing times of 120 and 30 ms (WALTZ-16), respectively, and acquisition times of 21.1 (t_1, ^1H), 21.3 (t_2, ^{15}N), and 84.8 ms (t_3, NH). Water suppression in these experiments was achieved by presaturation with a weak RF field and, for the 3D NOESY–HMQC, a jump-and-return pulse at the end of the NOE mixing period (Plateau & Guéron, 1982).

2D ^1H-detected ^1H–^{15}N heteronuclear NOE spectra (Kay et al., 1989b) were acquired at 600 MHz with acquisition times of 70 (t_1, ^{15}N) and 105 ms (t_2, NH). Water suppression was achieved by means of spin–lock purge pulses in both the control (no NOE) and experimental (with NOE) spectra. ^1H saturation was effected in the experiment with the NOE by application of a train of 120- or 135-deg flip-angle pulses (RF field strength: 29 kHz) at 20-ms intervals for 3 s.

A 2D ^1H–^{15}N HMQC-J experiment (Kay & Bax, 1990) was acquired at 750 MHz (^1H). Water suppression was by presaturation, and acquisition times were 176.6 (t_1, ^{15}N) and 112.6 ms (t_2, ^1H). A similar experiment was also acquired at 600 MHz. Amide proton exchange rates were assessed from ^1H–^{15}N HMQC spectra acquired at 1 h and 2.5 days following exchange of the protein from H_2O into D_2O.

Two triple-resonance 3D experiments, CBCA(CO)NH (Grzesiek & Bax, 1992) and HBHA(CBCACO)NH (Grzesiek & Bax, 1993), were acquired on an ^{15}N,^{13}C-double-labeled sample. The first experiment correlates the $^{13}C\alpha$ and $^{13}C\beta$ resonances of residue i with the amide ^{15}N and NH of residue $i+1$; the second yields correlations from $C^\alpha H/C^\beta H(i)$ to $^{15}N/NH(i+1)$. Spectra were recorded with t_1 acquisition times of 6.1 ($^{13}C\alpha/\beta$; CBCA(CO)NH) and 14.3 ms ($C^\alpha H/C^\beta H$; HBHA(CBCACO)NH); for both spectra, the t_2 (^{15}N) and t_3 (NH) acquisition times were 20.7 and 57.6 ms, respectively. The ^{13}C carrier was placed in the middle of the $C\alpha/C\beta$ region (45.52 ppm); $^{13}C\alpha$ and $^{13}C'$ pulses were generated as off-resonance amplitude- and phase-modulated pulses using the waveform generators. The RF field strengths of the $^{13}C\alpha/\beta$ and $^{13}C\alpha$ pulses were adjusted to give minimal excitation in the $^{13}C'$ region and *vice versa*. ^1H decoupling during the sequences was achieved with a 5.5-kHz WALTZ-16 field; the ^1H carrier was placed at the H_2O frequency throughout. Suppression of the water resonance was achieved by a combination of spin–lock purge pulses and low-level presaturation ($\sim$10 Hz RF field) during the 0.95-s relaxation delay.

States-TPPI (Marion et al., 1989) was applied in the indirectly detected dimensions of all experiments to give quadrature detection. Data were processed using FELIX (Hare Research/Biosym) on Sun Sparc or Silicon Graphics Indigo workstations. Linear prediction (using the algorithm supplied in FELIX 2.1) was employed to extend the data

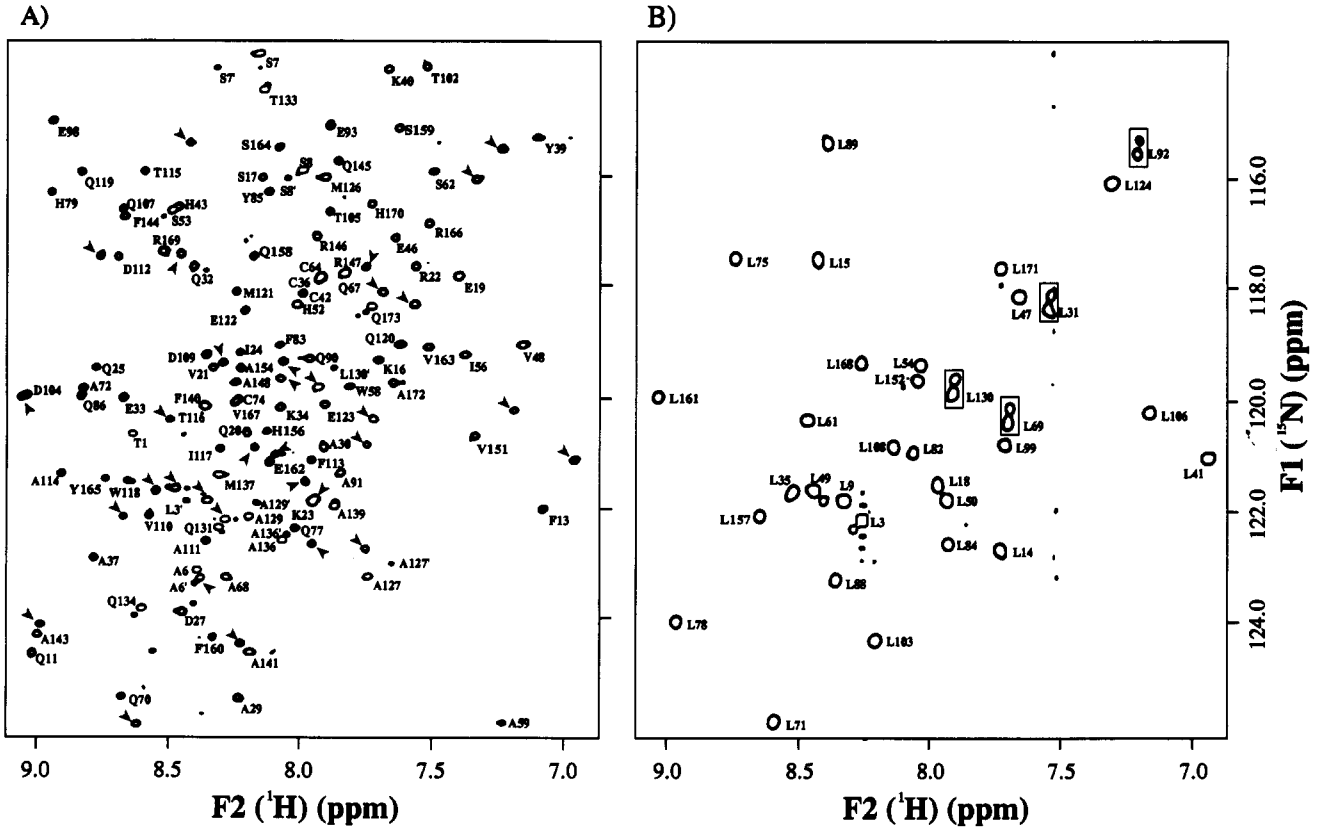

FIGURE 1: Central section of the 2D ^1H–^{15}N HSQC of hG-CSF. (A) Spectrum of uniformly ^{15}N-labeled hG-CSF recorded at 750-MHz ^1H frequency and 32 °C. All backbone amide resonances are shown except Gly-28, Glu-45, Ser-63, Ser-80, and the Ser-Gly region upfield of 113.5 ppm (^{15}N). Labels are as close as possible to the appropriate resonance. Satellite resonances carry an additional prime (′) in the label. Leucines are marked by an arrow (except for L3′ and L130′). (B) Spectrum recorded at 500 MHz and 30 °C. Same region as in A of the ^1H–^{15}N HSQC spectrum of hG-CSF specifically labeled with [^{15}N]Leu and [$^{13}C'$]Ala. The 33 leucine NH correlations are labeled according to sequence position, with the four correlations showing splitting due to a preceding $^{13}C'$ of Ala boxed.

table from 32 to 64 complex points in the ^{15}N (t_2) dimension of the 3D ^1H–^{15}N NOESY–HMQC and HOHAHA–HMQC experiments. The constant time evolution domains of the CBCA(CO)NH and HBHA(CBCACO)NH experiments were extended 2-fold, by "mirror image" linear prediction (Zhu & Bax, 1990) executed within FELIX 2.1. Data extension in the semiconstant time ^1H evolution period of the HBHA(CBCACO)NH experiment was achieved using linear prediction without the mirror image constraint. Data in this dimension were deliberately 'folded' in acquisition by use of a small (4 kHz) sweep width; the apparent carrier frequency was shifted, and the frequency domain data unfolded during processing by the application of a frequency-dependent phase correction to the t_1 interferogram.

The 2D ^1H–^{15}N HMQC-J spectra were processed twice with 2- and 5-Hz exponential line narrowing in t_1 (^{15}N) to obtain J-coupling information (Kay & Bax, 1990). Resonances arising from residues other than Gly that showed either splitting or at least observable line-broadening due to spin–spin coupling were fitted with a nonlinear least-squares fitting procedure. Traces were taken through each autopeak and were fitted to theoretical line shapes using a modified version of the program NHfit (Redfield & Dobson, 1990; P. C. Driscoll, personal communication). Resonances in the two spectra of the ^1H-detected ^1H–^{15}N heteronuclear NOE experiment were picked automatically using the program peakpick (C. Redfield), and the ratios of their peak heights were evaluated.

RESULTS AND DISCUSSION

Backbone Assignment of hG-CSF. The assignment of hG-CSF was carried out using a combination of two different approaches. Our initial strategy was based on analysis of 3D ^1H–^{15}N NOESY–HMQC and HOHAHA–HMQC experiments together with the use of selectively labeled amino acids, i.e., [^{15}N]Leu and [$^{13}C'$]Ala; this approach was supplemented by information derived from the triple-resonance CBCA(CO)-NH and HBHA(CBCACO)NH experiments (Grzesiek & Bax, 1992, 1993).

In the high-resolution ^1H–^{15}N HSQC spectrum, 161 candidate backbone NH resonances of high intensity were tentatively identified, together with a number of low-intensity peaks (Figure 1A; supplementary material); the number of backbone amides in our recombinant hG-CSF (excluding the labile NH_2 of the N-terminal Met residue) is 161. Most of the spurious peaks of lower intensity could later be assigned as satellites of the major peaks and interpreted as multiple conformations (see below); some of the 161 peaks originally selected were also reclassified as satellites. Resonances that were severely overlapped at 500 MHz could be identified later in the course of the assignment, especially in the spectra acquired at 750 MHz (Figure 1A), which show better chemical shift dispersion in the ^1H and ^{15}N dimensions than comparable spectra recorded at 500 and 600 MHz (data not shown). Comparison of the ^1H–^{15}N HSQC of the uniformly labeled sample with the ^1H–^{15}N HSQC of the sample specifically labeled with [^{15}N]Leu and [$^{13}C'$]Ala allowed the identification of all 33 Leu NHs. The four cases where Ala-Leu occur as

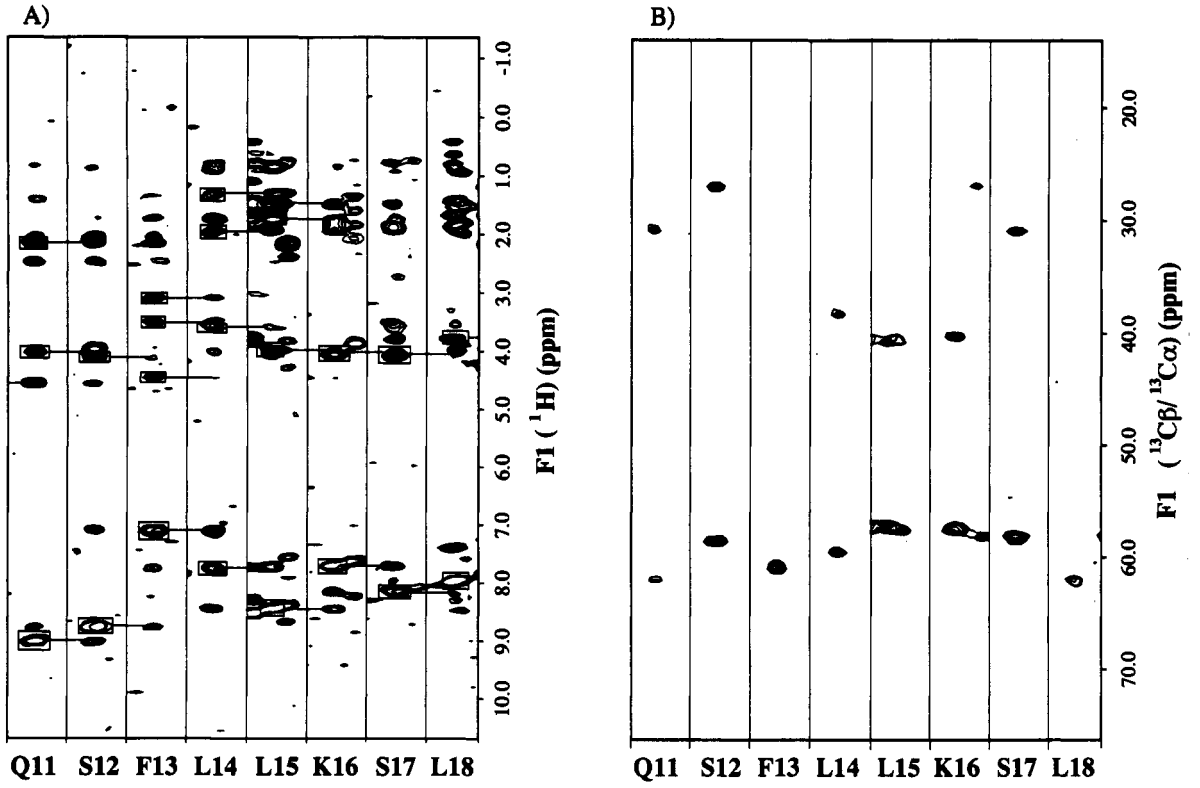

FIGURE 2: Sections from the ^{1}H–^{15}N and ^{1}H–^{15}N–^{13}C 3D experiments showing resonances from the beginning of helix A, as indicated. (A) Strips from the 3D ^{1}H–^{15}N NOESY–HMQC spectrum recorded at 500 MHz and 30 °C. Strips of 18 points in F_3 were taken at the ^{15}N chemical shift of each NH resonance. No window function was used to suppress the sides of the strips. NH, C$^\alpha$H, and C$^\beta$H resonances are boxed where they give rise to NH(i)–NH(i + 1), C$^\alpha$H(i)–NH(i + 1), or C$^\beta$H(i)–NH(i + 1) connectivities. Connectivities are indicated by thin lines. (B) Strips from the 3D CBCA(CO)NH spectrum recorded at 600 MHz and 30 °C. Strips were extracted as in A. ^{13}Cα and ^{13}Cβ resonances of residue i appear in strip i + 1.

pairs in the sequence were also identified by splittings in the ^{15}N dimension due to the one-bond J-coupling to the ^{13}C-labeled carbonyl carbon of alanine (Figure 1B).

Strips were extracted from the 3D ^{1}H–^{15}N NOESY–HMQC and HOHAHA–HMQC spectra at the NH and ^{15}N coordinates of the correlations identified from the ^{1}H–^{15}N HSQC. These were matched on the basis of their NH(i)–NH(i + 1), C$^\alpha$H(i)–NH(i + 1), and, if possible, C$^\beta$H(i)–NH(i + 1) NOEs. Generally, only patches of 3–6 contiguous strips could be found because of a combination of substantial overlap in both the NH and C$^\alpha$H regions of the 3D spectra and the presence of 13 prolines. A representative patch of strips from the 3D ^{1}H–^{15}N NOESY–HMQC spectrum is shown in Figure 2A. Throughout much of the sequence, NH(i)–NH(i + 2) NOEs helped to confirm the assignment. Because of broad line widths and the consequently poor scalar transfer efficiency, side chains could only be identified from the 3D HOHAHA–HMQC spectrum for Gly, Ala, and, in some cases, Ser with any confidence. This residue-type-specific information was combined with that for the leucines (and for the four leucines preceded by alanine), which served as anchor points in the sequence, enabling about 60–70% of the residues to be assigned in this manner.

With the recently developed CBCA(CO)NH and HBHA-(CBCACO)NH experiments, an alternative approach to assignment is available for uniformly ^{15}N,^{13}C-labeled proteins in the 20-kDa range (Grzesiek & Bax, 1992, 1993). Using the relatively large one-bond J-couplings, the ^{13}Cβ and ^{13}Cα (CBCA(CO)NH) or C$^\beta$H and C$^\alpha$H (HBHA(CBCACO)-NH) resonances of one residue are correlated with the ^{15}N and NH resonances of the following residue via the intervening carbonyl nucleus. The spectra were well resolved and showed good signal-to-noise (e.g., Figure 2B), in contrast to some of the earlier out-and-back-type triple-resonance sequences (Kay et al., 1990) which gave poor sensitivity for hG-CSF (data not shown). In the CBCA(CO)NH experiments, virtually all of the expected resonances were present except the ^{13}Cβ's of the serines in helices, which consistently were not observed. Although the HBHA(CBCACO)NH experiment is less sensitive, the majority of the C$^\beta$H and C$^\alpha$H resonances could be identified from this spectrum (supplementary material). In some cases (e.g., Val-167 and Cys-74), it was possible to disentangle strips that were highly overlapped (especially in the 3D ^{15}N-based experiments recorded at 500 MHz) by using the CBCA(CO)NH and HBHA(CBCACO)NH experiments.

The combined ^{13}Cα and ^{13}Cβ chemical shifts do not uniquely define individual residue types, except for Ala, Ser, Thr, and Gly. By adding the C$^\beta$H and C$^\alpha$H shift information, Leu and Ile can also be identified confidently. In these cases, the combination of the chemical shifts of ^{13}Cβ and ^{13}Cα with shifts of the C$^\beta$H and C$^\alpha$H from the HBHA(CBCACO)NH experiment confirms the identification of the side chains. Matching the C$^\beta$H and C$^\alpha$H shifts from the HBHA-(CBCACO)NH with the 3D ^{1}H–^{15}N HOHAHA–HMQC and NOESY–HMQC experiments immediately yielded pairs of residues. Consequently, 12 C$^\alpha$H and 10 C$^\beta$H shifts could be identified for the 13 prolines. Although the combination of ^{13}Cβ and ^{13}Cα shifts for individual residues (with the exceptions mentioned above) is, in general, not sufficient to determine the amino acid type, for stretches of 3–4 residues the joint probabilities associated with the combined ^{13}Cβ and ^{13}Cα shifts should permit unique placement of the strips in the sequence (Grzesiek & Bax, 1993). For analysis of the ^{13}C shift data, we used the programs type_prob and seq_prob (Grzesiek & Bax, 1993). These employ a database of protein ^{13}Cα and ^{13}Cβ chemical shifts, with which the experimental

shifts are compared to calculate the probability that a stretch of (for example) 3–6 contiguous strips from the CBCA(CO)-NH experiment corresponds to a particular stretch of the sequence. In the case of hG-CSF, this approach worked well for stretches containing one or more residues with unique chemical shift combinations, but it was less reliable for some other segments. A number of factors could be responsible for this: (i) systematic shifts of the observed ^{13}C chemical shifts relative to the database values due to secondary shift contributions; (ii) the over-representation of certain amino acids in the sequence (e.g., Leu in this case) relative to the database proteins and the under-representation of others (e.g., Asn, of which there are none in hG-CSF); and (iii) the predominant occurrence of a particular amino acid in one type of secondary structure (as in the case of hG-CSF, where almost all of the leucines are in helices).

The assignment of low-intensity peaks present in the ^{1}H–^{15}N HSQC as satellite resonances was greatly helped by the 3D CBCA(CO)NH and HBHA(CBCACO)NH experiments, especially in cases where the corresponding spin system could not be identified from the 3D HOHAHA–HMQC spectrum (Figure 1A; supplementary material).

Secondary Structure. The secondary structure of hG-CSF was determined using sequential and short-range NOEs, the three-bond J-coupling $^{3}J_{NH\alpha}$, and NH exchange data (Figure 3); corroborative evidence comes from the secondary shifts of the $^{13}C\alpha$ resonances (Figure 4A). We found a four-helix topology with two long loops between helices A and B and helices C and D. The data also suggest that there is a shorter helical segment between Pro-44 and Ser-53. We found a total of 384 sequential NOEs, which were classified into strong, medium, and weak, and 198 short- and medium-range NOEs, which have been interpreted only qualitatively. Quite frequently it was not possible to assign unambiguously a short- or medium-range NOE because of overlap. It was possible to perform line-shape fitting for resonances in the HMQC-J experiments, with J-couplings greater than approximately 5 Hz (depending on the line widths of the individual resonances). The spectrum acquired at 750 MHz was particularly useful in permitting the analysis of peaks that were overlapped at 600 MHz (for example, the NHs of His-43 and Leu-54), yielding a total of 41 three-bond J-couplings for non-Gly residues located in either extended regions or turns. The remaining resolved resonances were classified as having J-couplings of less than approximately 5 Hz and were largely located in helical regions. The sequential NH(i)–NH($i+1$), C$^{\alpha}$H(i)–NH($i+1$), and C$^{\beta}$H(i)–NH($i+1$) NOEs, together with the short- and medium-range NH(i)–NH($i+2$), C$^{\alpha}$H(i)–NH($i+2$), C$^{\alpha}$H(i)–NH($i+3$), and C$^{\alpha}$H(i)–NH($i+4$) NOEs, the $^{3}J_{NH\alpha}$ coupling constants, and the NH exchange data are summarized in Figure 3.

Following an unstructured N-terminal region of 10 residues with J-couplings of about 7 Hz and no medium-range NOEs, the data indicate that helix A commences at Pro-10 and runs to Thr-38. This is followed by a turn region containing four residues with J-couplings between 6.5 and 9.0 Hz, which is enclosed by the Cys-36–Cys-42 disulfide bridge. The following segment between His-43 and Ile-56 needs to be considered carefully. The C$^{\alpha}$H chemical shift of His-43 is shifted downfield by 0.9 ppm relative to its random coil value, and it gives rise to medium-range NOEs to the three subsequent residues Glu-45, Glu-46, and Leu-47. However, with an NH–C$^{\alpha}$H J-coupling of 8.9 Hz, His-43 is unlikely to be part of a helix. The C$^{\alpha}$H's of Pro-44 and Glu-45 both give rise to medium-range C$^{\alpha}$H(i)–NH($i+3$) NOEs, and the C$^{\alpha}$H of Glu-46 shows a short-range C$^{\alpha}$H(i)–NH($i+2$) NOE to Val-48, indicating the presence of helical structure. The NH autocorrelation peaks of Glu-45 and Glu-46 both show observable splittings in the ^{1}H–^{15}N HMQC-J experiment, but because of low signal-to-noise for the resonance of Glu-45, it was only possible to obtain a fitted J-coupling value (6.0 Hz) for Glu-46. Together with the relatively slow NH exchange rate of Val-48, the data suggest the presence of a loose helical turn between Pro-44 and Glu-46. This is interrupted by Leu-47, which has a J-coupling of 8.9 Hz. The short-range NOEs of the following residues suggest another two turns of helical structure. The predominance of C$^{\alpha}$H(i)–NH($i+2$) and the absence of C$^{\alpha}$H(i)–NH($i+3$) NOEs in this segment suggests a 3_{10} rather than an α-helical conformation. A J-coupling of 7.2 Hz for Leu-54 indicates that the helical structure extends only to Ser-53 (see below). Hence, we refer to the region between Pro-44 and Ser-53 as the helical segment A′.

The absence of short- and medium-range NOEs in conjunction with medium-to-strong C$^{\alpha}$H(i)–NH($i+1$) NOEs for the following 14 residues, as well as J-couplings of about 6.5 Hz, indicates a long, extended loop (designated here as the AB loop), which terminates with Gln-70. The second disulfide bridge between Cys-64 and Cys-74 encloses the end of the AB loop and the beginning of helix B. This helix is approximately one turn shorter than helix A and extends from Gln-71 to Ile-95 according to the NOE and NH exchange data. Helix B is followed by a three-residue stretch comprising Glu-98, Leu-99 (with J-couplings of 5.6 and 8.2 Hz), and Gly-100. Helix C, the second "short" helix, begins at Thr-102 and ends at Glu-123 and is followed by the 17-residue extended CD loop, with J-couplings of about 7 Hz. This leads into helix D, extending from Ala-143 to Ala-172. In general, the NH exchange data concur well with the secondary structure, as inferred from the NOEs and J-couplings, and demonstrate that helix C is particularly well protected, whereas the NHs of helix D seem to be, on average, more accessible to the solvent.

Further support for this definition of the secondary structure is given by the secondary shifts of the $^{13}C\alpha$ nuclei (Figure 4A). Positive (i.e., downfield) $^{13}C\alpha$ secondary chemical shifts with an average of 3.09 ± 1.00 ppm tend to be observed for α-helical regions within proteins, whereas negative $^{13}C\alpha$ secondary shifts with an average of -1.48 ± 1.23 ppm indicate extended regions (Spera & Bax, 1991). The $^{13}C\alpha$ secondary shift data for hG-CSF nicely match the four-helix topology. According to the $^{13}C\alpha$ secondary shifts, helix A extends from Gln-11 to Thr-38 with an average secondary shift of 1.92 ± 1.00 ppm. The helical region A′ between Pro-44 and Ser-53 shows positive secondary shifts for Pro-44 and Glu-45, negative shifts for Glu-46 (–0.81 ppm) and Leu-47 (–2.49 ppm), and a return to positive shifts for the remaining residues, Val-48–Ser-53. Helix B encompasses Leu-71–Ala-91 according to these data, with an average secondary shift of 2.52 ± 1.66 ppm. The region of positive $^{13}C\alpha$ secondary shifts corresponding to helix C starts at Pro-101 and terminates at Glu-123 (average secondary shift, 2.26 ± 1.41 ppm). Finally, helix D extends from Ala-143 to His-170 with an average secondary shift of 2.54 ± 0.91 ppm. The regions of negative $^{13}C\alpha$ secondary shifts between Pro-57 and Ala-68 and between Met-126 and Phe-140 correspond to extended regions in the two long loops. Generally, the $^{13}C\alpha$ secondary shifts are consistent with the NOE, coupling constant, and NH exchange data and lend further support to the identification of a helical segment between Pro-44 and Ser-53, with an interruption at

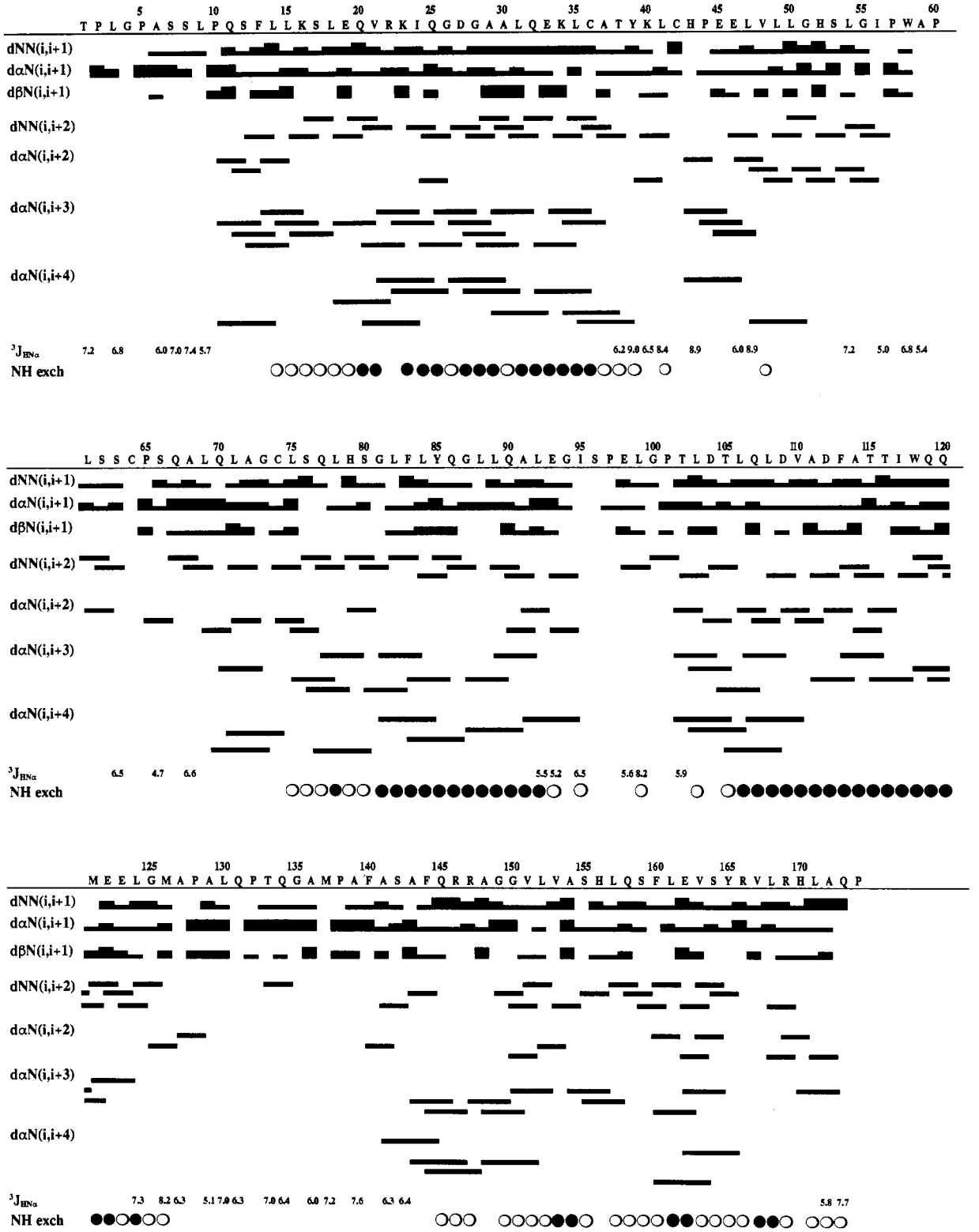

FIGURE 3: Summary of the sequential short- and medium-range NOEs, three-bond $^3J_{NH\alpha}$ couplings, and NH exchange data observed for hG-CSF. Sequential NOEs were classified into strong, medium, or weak, and their approximate strengths are indicated by the width of the bar in each case. The presence of medium-range NOEs, $d_{NN(i,i+2)}$, $d_{\alpha N(i,i+2)}$, $d_{\alpha N(i,i+3)}$, and $d_{\alpha N(i,i+4)}$, is shown with no indication of intensity; $d_{NN(i,i+1)}$ refers to an NOE between the backbone NH resonance of residue i and the NH of residue $i + 1$, while $d_{\alpha N(i,i+1)}$ refers to an NOE between the $C^{\alpha}H$ resonance of residue i and the backbone NH of residue $i + 1$. Other NOE types are represented analogously (Wüthrich, 1986). Fitted three-bond $^3J_{NH\alpha}$ coupling constant values (derived from spectra acquired at 600 and 750 MHz) are shown in hertz. Residues lacking a value either showed correlations that were insufficiently resolved to allow for reliable line-shape fitting or did not show appreciable line broadening due to spin–spin coupling; in the latter case, an upper limit for $^3J_{NH\alpha}$ of about 5 Hz is assumed (see text). NH exchange data (NH exch) are indicated by circles; open circles indicate that the NH resonance in the 2D ^1H–^{15}N HMQC was still present 1 h after transfer of the sample into D_2O, and filled circles indicate continued presence after 2.5 days in D_2O.

Leu-47. The data disagree only in the definition of the end of helix B. Whereas the NOE and NH exchange data suggest that helix B extends up to Ile-95, negative $^{13}C\alpha$ secondary shifts suggest an extended region for residues Leu-92–Ile-95. The definition of the true solution-state conformation of these residues will, however, require a much more comprehensive set of NOE distance restraints. In summary, we conclude that helix A extends from Pro-10 to Thr-38, the helical element

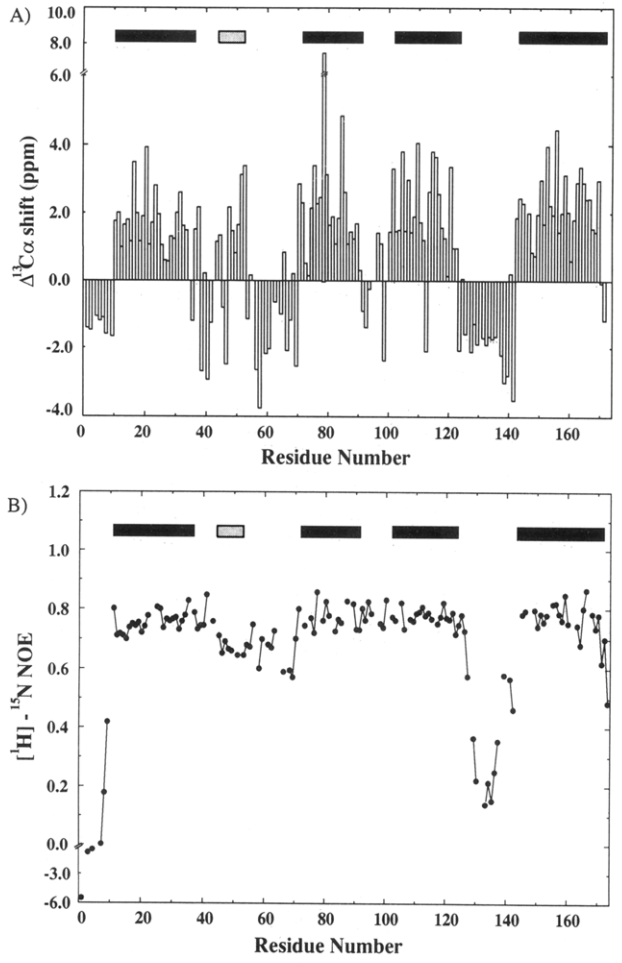

FIGURE 4: $^{13}C\alpha$ secondary shifts and ^1H–^{15}N NOE transfer ratios of hG-CSF. (A) $^{13}C\alpha$ secondary shifts are displayed according to the amino acid sequence. Positive $^{13}C\alpha$ secondary shifts generally correlate with helical secondary structure and negative shifts with extended regions. Helices A–D are denoted by black boxes, and the helical segment between residue Pro-44 and Ser-53 is indicated by a hatched box. (B) ^1H–^{15}N NOE transfer ratios. The ratio of the peak heights in the ^1H–^{15}N NOE experiment with and without presaturation is represented for all resolvable backbone NH resonances by dots for each residue along the amino acid sequence; gaps are due to the presence of overlapped resonances or prolines. The secondary structure is indicated as in A.

A′ from Pro-44 to Ser-53, helix B from Leu-71 to Ala-91, helix C from Thr-102 to Glu-123, and helix D from Ala-143 to Ala-172.

Generally, the definitions of the four core helices deduced from our solution-state NMR data agree well with those found from recent crystallographic studies of hG-CSF and G-CSFs from other species (Hill et al., 1993; Lovejoy et al., 1993). However, there may be differences between solution and crystal in the structure of the helical region A′ from Pro-44 to Ser-53. Our NMR data suggest one turn of loose α-helix, interrupted at Leu-47, and followed by two turns of 3_{10}-helix, whereas the original crystal structure displays a turn of 3_{10}-helix interrupted at Leu-47, leading into two turns of α-helix (Hill et al., 1993). It must be stressed, however, that confirmation of the true structure of this region in solution must await calculation of the 3D structure of hG-CSF based upon a complete set of interproton distance restraints.

The previously reported solution secondary structure of hG-CSF (Zink et al., 1992) proposes the same overall topology, except for the absence of the small A′ helix and some differences in the definitions of the beginning and end of the core helices. In terms of the lengths of the four core helices and the presence of the helical region A′, both the single-crystal and solution-state data show hG-CSF to be structurally more closely related to growth hormone than to GM-CSF and the smaller interleukins IL-2 and IL-4.

Backbone Dynamics of hG-CSF. The magnitudes of ^{15}N NOEs from the directly bonded NH protons were measured by collecting heteronuclear NOE spectra with and without presaturation of ^1H. The resulting spectra were comparatively well resolved and showed good signal-to-noise (data not shown). The set of two experiments was repeated twice with similar results. It was possible to measure the peak height ratios for 132 of 161 backbone amide resonances. The average ^1H–^{15}N heteronuclear NOEs of residues within the well-defined secondary structure elements of helices A–D range between 0.75 and 0.77. For the AB loop region the average ratio is lower at 0.65, while for the CD loop between Pro-128 and Pro-138 the average falls to 0.24, with a minimum of 0.14 at Thr-133. For the 10 N-terminal residues, highly negative ratios were observed, as found for disordered terminal regions in other proteins (Kay et al., 1989b; Redfield et al., 1992; Stone et al., 1992) (Figure 4B).

The ^1H–^{15}N heteronuclear NOE experiment provides information about dynamic processes influencing the reorientation rate in solution of the amide NH bond vector. For a rigid, isotropically tumbling body at a ^1H frequency of 600 MHz (i.e., ^{15}N frequency of 60.8 MHz), the ^1H–^{15}N NOE is a monotonic function of the overall correlation time, τ_m, with a minimum value of –3.5 calculated for the case $\tau_m\omega \ll 1$ and a maximum of 0.83 in the limit $\tau_m\omega \gg 1$ ($\omega = 2\pi \times 600$ MHz) (Abragam, 1961; Kay et al., 1989b). The major influences on the strength of the ^1H–^{15}N NOE are the spectral densities of motion on a sub-nanosecond time scale. However, because quantitative interpretation of the results is complicated by contributions from a number of additional factors, the data are interpreted here in a strictly qualitative manner; a full quantitative characterization of the relaxation properties of the backbone ^{15}N nuclei, leading to the calculation of order parameters, S^2, requires a more rigorous approach involving the measurement of ^{15}N longitudinal (T_1) and transverse (T_2) relaxation times.

The data agree well with the overall secondary structure (Figure 4B). The helices A–D appear to have low mobility, with ^{15}N heteronuclear NOEs close to the theoretical maximum; this, together with the slow NH exchange rates, suggests a relatively rigid and compact protein structure. The remarkably low ^1H–^{15}N heteronuclear NOEs of residues 129–137 in the CD loop are consistent with a substantially enhanced degree of mobility in this region (relative to the tumbling of the whole molecule) on the sub-nanosecond time scale. Interestingly, a number of these residues also show satellites in the ^1H–^{15}N HSQC (Figure 1A), indicating an additional slow conformational exchange, which is tentatively ascribed to cis–trans peptide bond isomerization of one or more of the three prolines in this region. Less pronounced, but still notably lower than those of the helices, are the average NOE ratios for residues in the region between helices A and B. Fast motion in this region may be constrained by the disulfide bridge between Cys-36 and Cys-42 at the end of helix A, the helical segment A′ between Pro-44 and Ser-53, and the Cys-64–Cys-74 disulfide at the beginning of helix B. However, the presence of satellites for residues neighboring the three prolines in this region (Ile-56, Ala-59, and Ser-66) in the ^1H–^{15}N HSQC again suggests slow conformational exchange, possibly due to proline isomerization.

It is of interest to note that relatively high atomic temperature (*B*) factors were found, even for the helical residues in the crystal structure of hG-CSF (Hill et al., 1993). Besides the high mobility of the 10 N-terminal residues, we find the mobility of the long CD loop most pronounced, as noted above. The high mobility on different time scales observed in solution could contribute to the absence of identifiable electron density for residues in the CD loop in the various crystal structures of G-CSF (Hill et al., 1993; Lovejoy et al., 1993). Notably, the CD loop in hG-CSF includes the native glycosylation site at Thr-133 and probably is not directly involved in receptor binding. It is also interesting to note that the CD loop of IL-4 shows a similar increase in mobility relative to the protein as a whole (Redfield et al., 1992). In the AB loop, we see a less pronounced but still notable increase in mobility compared to the protein core, which again is accompanied by the presence of satellites for residues neighboring the three prolines in this region. The region between Cys-64 and the beginning of helix B also lacks observable electron density in the original crystal structure of hG-CSF (Hill et al., 1993), even though Cys-64 is involved in a disulfide bridge to Cys-74 in helix B. In contrast, in bovine and canine G-CSF, where Pro-65 is replaced by Ser, the AB loop can be traced in its entirety, while the refined hG-CSF molecule B displays a quite different conformation and very high atomic *B* factors for the Pro-65–Gly-73 loop region (Lovejoy et al., 1993). The NMR data suggest an explanation for these differences. Intrinsic mobility due to proline isomerization in hG-CSF (but not the bovine or canine proteins) may allow different molecules in the asymmetric unit to adopt different conformations, resulting in the absence of observable electron density in the original hG-CSF model. Since the C-terminal end of the AB loop has been implicated in the specificity of receptor binding (Lovejoy et al., 1993), further studies of the structure and dynamics of this region are expected to contribute to a better understanding of the binding mechanism.

ACKNOWLEDGMENT

We thank Professor D. Eisenberg for providing the human, canine, and bovine G-CSF coordinates and Drs. S. Grzesiek and A. Bax for the gift of the programs type_prob and seq_prob. We are grateful to Drs. P. Driscoll and C. Redfield for the use of their programs, NHfit and peakpick.

SUPPLEMENTARY MATERIAL AVAILABLE

A table with the assignments for the NH, ^{15}N, $^{13}C\alpha$, $C^{\alpha}H$, $^{13}C\beta$, and $C^{\beta}H$ resonances of hG-CSF (3 pages). Ordering information is given on any current masthead page.

REFERENCES

Abdel-Meguid, S. S., Shieh, H.-S., Smith, W. W., Dayringer, H. E., Violand, B. N., & Bentle, L. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6434–6437.

Abragam, A. (1961) *The Principles of Nuclear Magnetism*, Clarendon Press, Oxford, U.K.

Bax, A., Ikura, M., Kay, L. E., Torchia, D. A., & Tschudin, R. (1990) *J. Magn. Reson. 86*, 304–318.

Bazan, J. F. (1990) *Immunol. Today 11*, 350–354.

Bazan, J. F. (1991) *Neuron 7*, 197–208.

Bazan, J. F. (1992) *Science 257*, 410–412.

Bodenhausen, G., & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185–188.

Clark, S. C., & Kamen, R. (1987) *Science 236*, 1229–1237.

de Vos, A. M., Ultsch, M., & Kossiakoff, A. A. (1992) *Science 255*, 306–312.

Diederichs, K., Boone, T., & Karplus, P. A. (1991) *Science 254*, 1779–1782.

Driscoll, P. C., Clore, G. M., Marion, D., Wingfield, P. T., & Gronenborn, A. M. (1990) *Biochemistry 29*, 3542–3556.

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

Grzesiek, S., & Bax, A. (1993) *J. Biomol. NMR 3*, 185–204.

Guyer, M. S., Reed, R. R., Seitz, J. A., & Low, K. B. (1981) *Cold Spring Harbor Symp. Quant. Biol. 45*, 135–140.

Hill, C. P., Osslund, T. D., & Eisenberg, D. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 5167–5171.

Kay, L. E., & Bax, A. (1990) *J. Magn. Reson. 86*, 110–126.

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

Kay, L. E., Torchia, D. A., & Bax, A. (1989b) *Biochemistry 28*, 8972–8979.

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

Lovejoy, B., Cascio, D., & Eisenberg, D. (1993) *J. Mol. Biol. 234*, 640–653.

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

McKay, D. B. (1992) *Science 257*, 412–413.

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

Milburn, M. V., Hassell, A. M., Lambert, M. H., Jordan, S. R., Proudfoot, A. E. I., Graber, P., & Wells, T. N. C. (1993) *Nature 363*, 172–176.

Mott, H. R., Driscoll, P. C., Boyd, J., Cooke, R. M., Weir, M. P., & Campbell, I. D. (1992) *Biochemistry 31*, 7741–7744.

Nicola, N. A. (1989) *Annu. Rev. Biochem. 58*, 45–77.

Nirmala, N. R., & Wagner, G. (1989) *J. Magn. Reson. 82*, 659–661.

Norwood, T. J., Boyd, J., Heritage, J. E., Soffe, N., & Campbell, I. D. (1990) *J. Magn. Reson. 87*, 488–501.

Oh-eda, M., Hasegawa, M., Hattori, K., Kuboniwa, H., Kojima, T., Orita, T. Tomonou, K., Yamazaki, T., & Ochi, N. (1990) *J. Biol. Chem. 265*, 11432–11435.

Pandit, J., Bohm, A., Jancarik, J., Halenbeck, R., Koths, K., & Kim, S.-H. (1992) *Science 258*, 1358–1362.

Parry, D. A. D., Minasian, E., & Leach, S. J. (1988) *J. Mol. Recognit. 1*, 107–110.

Plateau, P., & Guéron, M. (1982) *J. Am. Chem. Soc. 104*, 7310–7311.

Powers, R., Garrett, D. S., March, C. J., Frieden, E. A., Gronenborn, A. M., & Clore, G. M. (1992) *Science 256*, 1673–1677.

Redfield, C., & Dobson, C. M. (1990) *Biochemistry 29*, 7201–7214.

Redfield, C., Boyd, J., Smith, L. J., Smith, R. A. G., & Dobson, C. M. (1992) *Biochemistry 31*, 10431–10437.

Senda, T., Matsuda, S., Kurihara, K., Nakamura, K. T., Kawano, G., Shimizu, H., Mizuno, H., & Mitsui, Y. (1990) *Proc. Jpn. Acad. 66B*, 77–80.

Senda, T., Shimazu, T., Matsuda, S., Kawano, G., Shimizu, H., Nakamura, K. T., & Mitsui, Y. (1992) *EMBO J. 11*, 3193–3201.

Smith, L. J., Redfield, C., Boyd, J., Lawrence, G. M. P., Edwards, R. G., Smith, R. A. G., & Dobson, C. M. (1992) *J. Mol. Biol. 224*, 899–904.

Souza, L. M., Boone, T. C., Gabrilove, J., Lai, P. H., Zsebo, K. M., Murdock, D. C., Chazin, V. R., Bruszewski, J., Lu, H., Chen, K. K., Barendt, J., Platzer, E., Moore, M. A. S., Mertelsmann, R., & Welte, K. (1986) *Science 232*, 61–65.

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

Stone, M. J., Fairbrother, W. J., Palmer, A. G., III, Reizer, J., Saier, M. H., Jr., & Wright, P. E. (1992) *Biochemistry 31*, 4394–4406.

Walter, M. R., Cook, W. J., Ealick, S. E., Nagabhushan, T. L., Trotta, P. P., & Bugg, C. E. (1992) *J. Mol. Biol. 224*, 1075–1085.
Wlodawer, A., Pavlovsky, A., & Gutschina, A. (1992) *FEBS Lett. 309*, 59–64.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Zhu, G., & Bax, A. (1990) *J. Magn. Reson. 90*, 405–410.
Zink, T., Ross, A., Ambrosius, D., Rudolph, R., & Holak, T. A. (1992) *FEBS Lett. 314*, 435–439.