Circular Permutation of Granulocyte Colony-Stimulating Factor

Yiqing Feng,* John C. Minnerly, Linda L. Zurfluh, William D. Joy, William F. Hood, Ann L. Abegg, Edith S. Grabbe, Jeng-Jong Shieh, Tammy L. Thurman, John P. McKearn, and Charles A. McWherter

Searle Discovery Research, Monsanto Company, St. Louis, Missouri 63198

Received September 15, 1998; Revised Manuscript Received December 16, 1998

ABSTRACT: The sequence of granulocyte colony-stimulating factor (G-CSF) has been circularly permuted by introducing new chain termini into interhelical loops and by constraining the N- and C-terminal helices, either by direct linkage of the termini (L0) or by substitution of the amino-terminal 10-residue segment with a seven-residue linker composed of glycines and serines (L1). All the circularly permuted G-CSFs (cpG-CSFs) were able to fold into biologically active structures that could recognize the G-CSF receptor. CD and NMR spectroscopy demonstrated that all of the cpG-CSFs adopted a fold similar to that of the native molecule, except for one [cpG-CSF(L1)[142/141]] which has the new termini at the end of loop 34 with the shorter L1 linker. All of the cpG-CSFs underwent cooperative unfolding by urea, and a systematically lower free energy change (ΔG_{urea}) was observed for molecules with the shorter L1 linker than for those molecules in which the original termini were directly linked (the L0 linker). The thermodynamic stability of the cpG-CSFs toward urea was found to correlate with their relative ability to stimulate proliferation of G-CSF responsive cells. Taken together, these results indicate that the G-CSF sequence is robust in its ability to undergo linear rearrangement and adopt a biologically active conformation. The choice of linker, with its effect on stability, seems to be important for realizing the full biological activity of the three-dimensional structure. The breakpoint and linker together are the ultimate determinants of the structural and biological profiles of these circularly permuted cytokines. In the following paper [McWherter, C. A., et al. (1999) *Biochemistry 38*, 4564−4571], McWherter and co-workers have used circularly permuted G-CSF sequences to engineer chimeric dual IL-3 and G-CSF receptor agonists in which the relative spatial orientation of the receptor agonist domains is varied. Interpreting the differences in activity for the chimeric molecules in terms of the connectivity between domains depends critically on the results reported here for the isolated cpG-CSF domains.

Granulocyte colony-stimulating factor (G-CSF)[1] is a member of the hematopoietic cytokine superfamily that promotes the growth and differentiation of cells of the myeloid lineage, as well as enhancing neutrophil function (reviewed in ref *1*). It binds to and signals through a transmembrane receptor (G-CSFR), expressed primarily on the surface of cells of the neutrophilic−granulocytic lineage (*1*). G-CSF is produced by a number of cell types, including stromal cells in bone marrow and vascular endothelial cells. In addition to its natural role in constitutive hematopoiesis, G-CSF is used to mitigate neutropenia in patients receiving myelosuppressive cancer chemotherapy (*2*−*5*).

The importance of G-CSF in biology and medicine has led to a number of studies of its structure−function relationships (*6*−*9*). These studies have focused on using the three-dimensional (3D) structure (*10*−*14*) to interpret the results of mutagenesis (*8*, *9*) and antibody blocking (*7*) experiments. In the following paper (*15*), the effect of circular permutation on the biological activity of G-CSF in the context of a family of chimeric growth factors, the myelopoietins (MPO), was examined. The MPOs composed of circularly permuted G-CSFs (cpG-CSFs) exhibited G-CSFR-related potencies ranging from 7 to >100% of that of G-CSF. Using cpG-CSFs to understand the origin of this variation is a major objective of this study.

Compared with classical mutagenesis, circular permutation is an alternative method for protein engineering that entails the manipulation of sequence on a large scale (*16*−*24*). Circular permutation can be conceptually understood in terms of forming a circular protein intermediate by linking the N- and C-terminal ends of a protein together and subsequently breaking the chain at another position in the sequence (*16*). This corresponds to swapping the carboxy- and amino-terminal segments of a sequence, and linking the original termini either directly or with a linker sequence. At a practical level, swapping the sequence is accomplished by transposing the DNA sequence encoding the protein. Circular permutation has been extensively studied for its impact on protein structure, conformational stability, and folding kinetics (*16*−*24*). The requirements for successfully permuting a sequence are simply defined as follows. (i) The N- and C-termini must be capable of being linked, and (ii) the new termini must be introduced at a location that disrupts neither the folding

* To whom correspondence should be addressed: Searle Discovery Research, BB4I, Monsanto Co., 700 Chesterfield Parkway N., St. Louis, MO 63198. Phone: (314) 737-6046. Fax: (314) 737-7425. E-mail: yiqing.g.feng@monsanto.com.

[1] Abbreviations: G-CSF, granulocyte colony-stimulating factor; G-CSFR, G-CSF receptor; IL-3, interleukin-3; IL-3R, IL-3 receptor; MPO, myelopoietin, a chimeric dual IL-3 and G-CSF receptor agonist; cpG-CSF, circularly permuted G-CSF(S17); 3D, three-dimensional.

10.1021/bi982224o CCC: $18.00
Published on Web 03/20/1999

Table 1: Nomenclature and Oligonucleotide Primers Used To Create and Amplify DNA in the Construction of cpG-CSFs

Nomenclature	Breakpoint[a]	New Start[b]	New Stop[c]
cpG-CSF(L1)[39/38]	39/38	GATCGACCATGGCTTACAAGCTGTGCCACCCC	CGATCGAAGCTTATTAGGTGGCACACAGCTTCTCCT
cpG-CSF(L1)[97/96]	97/96	GATCGACCATGGCTCCCGAGTTGGGTCCCACC	CGATCGAAGCTTATTAGGATATCCCTTCCAGGGCCT
cpG-CSF(L1)[126/125]	126/125	GATCGACCATGGCTATGGCCCCTGCCCTGCAG	CGATCGAAGCTTATTATCCCAGTTCTTCCATCTGCT
cpG-CSF(L1)[133/132]	133/132	GATCGACCATGGCTACCCAGGGTGCCATGCCG	CGATCGAAGCTTATTAGGGCTGCAGGGCAGGGGCCA
cpG-CSF(L1)[142/141]	142/141	GATCGACCATGGCTTCTGCTTTCCAGCGCCGG	CGATCGAAGCTTATTAGGCGAAGGCCGGCATGGCAC
cpG-CSF[39/38]	39/38	GATCGACCATGGCTTACAAGCTGTGCCACCCC	CGATCGAAGCTTATTAGGTGGCACACAGCTTCTCCT
cpG-CSF[49/48]	49/48	GATCGACCAT GGCTCTGCTC GGACACTCTC TG	CGATCGAAGC TTATTACACC AGCTCCTCGG GGTGGC
cpG-CSF[97/96]	97/96	GATCGACCATGGCTCCCGAGTTGGGTCCCACC	CGATCGAAGCTTATTAGGATATCCCTTCCAGGGCCT
cpG-CSF[126/125]	126/125	GATCGACCATGGCTATGGCCCCTGCCCTGCAG	CGATCGAAGCTTATTATCCCAGTTCTTCCATCTGCT
cpG-CSF[133/132]	133/132	GATCGACCATGGCTACCCAGGGTGCCATGCCG	CGATCGAAGCTTATTAGGGCTGCAGGGCAGGGGCCA
cpG-CSF[142/141]	142/141	GATCGACCATGGCTTCTGCTTTCCAGCGCCGG	CGATCGAAGCTTATTAGGCGAAGGCCGGCATGGCAC

[a] The breakpoints are the residue numbers in the original sequence which are now the new amino and new carboxy group, respectively. [b] Pairs of primers are used to create and amplify DNA fragments for the new amino-terminal segment. The primer pairs consist of "New Start" and either "L1 Linker Start" (GCTCTGAGAGCCGCCAGAGCCGCCAGAGGGCTGCGCAAGGTGGCGTAGAACGCG) for the L1 Linker or "L0 Linker Start" (GGGCTGCGCAAGGTGGCG) for the L0 linker. [c] Pairs of primers are used to create and amplify DNA fragments for the new carboxy-terminal segment. The primer pairs consist of "New Stop" and either "L1 Linker Stop" (CAGCCCTCTGGCGGCTCTGGCGGCTCT-CAGAGCTTCCTGCTCAAGTCTTTAGAG) for the L1 Linker or "L0 Linker Stop" (ACACCATTGGGCCCTGCCAGC) for the L0 linker.

process, the stability of the fold, nor its biological activity. This has proven to be possible for proteins of a variety of sizes (from 58 to >400 residues), structural classes (α, α and β, and β), and functional classes (enzymes, inhibitors, cytokines, soluble proteins, membrane proteins, and disulfide cross-linked proteins).

This paper complements the previous structure−function studies by examining the effect of circular permutation of the G-CSF sequence on the activity, conformation, and stability of G-CSF. The results reported here provide useful insights into the relationship of stability and activity, the role of the N- and C-terminal helices, and the importance of both the linker and the location of the termini in determining biological and physical properties. In addition, these results provide an interpretation for the results of the following paper in which circularly permuted G-CSF (cpG-CSF) sequences are used to alter the spatial relationship of IL-3 and G-CSF receptor agonist domains in the myelopoietins, a family of novel chimeric IL-3 and G-CSF receptor agonists.

EXPERIMENTAL PROCEDURES

Nomenclature. A cpG-CSF sequence is described by a nomenclature that specifies both the location of the new termini (the breakpoint) and the method of linking the original termini (see Table 1 and Results and Discussion). For example, cpG-CSF(L1)[39/38] specifies a circularly permuted G-CSF(S17) sequence in which the amino- and carboxy-terminal residues correspond to residues 39 and 38, respectively, of the original sequence and with the L1 linker (defined below) connecting the original termini. Whenever no linker designation is given, this corresponds to a peptide bond linker (L0) in which residue 174 is directly linked to residue 1. As summarized in Table 1, a series of nine cpG-CSFs employing six breakpoint locations and two different circularizing linkers (the L0 and L1 linkers) were examined in this study.

Creation of L1 Linker cpG-CSF Genes. The L1 linker corresponds to deleting the amino-terminal 10-residue segment of G-CSF, and replacing it with a seven-residue linker SGGSGGS. To create cpG-CSF(L1) sequences, DNA fragments encoding the amino and carboxy segments of cpG-CSF(L1) were first created, essentially as described by Mullins et al. (*22*). Using the G-CSF(S17) sequence in plasmid pMon13037 (*25*) as a template, the oligonucleotide primer set "New Start" and "Linker Start" (see Table 1) was used to create and amplify the DNA fragment encoding the new N-terminal portion of the cpG-CSF(L1) followed by the L1 linker that connects the original amino- and carboxy-terminal segments. Again using pMon13037 as a template, the primer set "New Stop" and "Linker Stop" (Table 1) was used to create and amplify the segment of DNA encoding the L1 linker followed by the C-terminal portion of the cpG-CSF(L1) sequence. The "New Start" and "New Stop" primers are designed to incorporate an *Nco*I site on the 5′ fragment and a *Hin*dIII site on the 3′ fragment to allow cloning of the cpG-CSF(L1) gene into expression plasmids. The PCRs were performed in a model 480 DNA thermal cycler (Perkin Elmer Corp., Norwalk, CT) using Perkin-Elmer GeneAmp PCR Core Reagents kits.

The fragments encoding the new amino and carboxy regions were joined by taking advantage of their complementary sequences in the linker segment. The two fragments were resolved by electrophoresis on a 1% Tris-acetate buffered gel, stained with ethidium bromide, and isolated using a Qiaex Gel Extraction kit (Qiagen, Chatsworth, CA), and annealed through their complementary sequences in the "Linker Start" and "Linker Stop" segments. Using the annealed fragments as the template DNA in a final PCR, the primers "New Start" and "New Stop" were used to create and amplify the full-length cpG-CSF(L1) gene. The reaction products were purified using Wizard PCR Prep kits (Promega, Madison, WI). After restriction endonuclease digestion, the full-length cpG-CSF(L1) genes were cloned into an *Escherichia coli* vector for expression.

Creation of L0 Linker cpG-CSF Genes. The L0 linker corresponds to a direct linkage of the native N-terminal and C-terminal residues (i.e., residue 174 and residue 1). Similar to the L1 linker molecules, genes encoding cpG-CSF molecules were made using PCR amplification to create the fragments encoding the amino and carboxy fragments. A blunt end ligation was used to join the fragments because there is no overlapping complementary linker sequence.

The primer sets "New Start" and "L0 Linker Start" and "New Stop" and "L0 Linker Stop" (Table 1) were used to

generate DNA fragments encoding the new amino and new carboxy segments, respectively, by PCR amplification using the G-CSF(S17) sequence in pMON13037 as a template. Deep Vent polymerase (New England Biolabs, Beverly, MA) was employed to reduce overhangs using reaction conditions recommended by the manufacturer. The "L0 Linker Start" and "L0 Linker Stop" primers were 5′ phosphorylated to enhance the subsequent blunt end ligation. The PCR products were purified using Wizard PCR Prep kits.

The fragments were digested with restriction endonucleases to generate an *Nco*I overhang on the 5′ fragment and a *Hin*dIII overhang on the 3′ fragment, and isolated by gel electrophoresis as described for the L1 linker molecules. The fragments were combined with and annealed to the ends of the *Nco*I−*Hin*dIII fragment of an expression vector, and then ligated together to create the full-length cpG-CSF gene within the vector.

Optimization of 5′ Terminal Segments for Expression. For all cpG-CSFs except cpG-CSF[39/38] and cpG-CSF(L1)-[39/38], the 5′ ends of the genes were optimized with *E. coli* preferred usage patterns biased toward AT-rich codons to achieve high expression levels. For cpG-CSF[126/125] and cpG-CSF(L1)[126/125], a 92 bp fragment at the 5′ end of the gene was enriched with AT-rich codons. For cpG-CSF[133/132] and cpG-CSF(L1)[133/132], a 71 bp fragment was enriched. A fragment of 44 bp was enriched in the genes encoding cpG-CSF[142/141] and cpG-CSF(L1)[142/141], and a 50 bp fragment in the gene which encodes cpG-CSF-[49/48]. DNA sequence analysis confirmed the modified sequence. The expression vectors were used to transform *E. coli* strain DH5α cells (Life Technologies, Gaithersburg, MD). The plasmid DNA was purified and the sequence confirmed. For protein expression, *E. coli* strain JM101 was transformed with the confirmed plasmid. Despite optimization of the six 5′ codons of cpG-CSF[97/96] and cpG-CSF-(L1)[97/96], they could not be expressed in *E. coli* at sufficiently high levels for further study.

Folding and Purification. G-CSF(S17) and cpG-CSFs were produced in refractile bodies by inducing the expression of the appropriate plasmid in *E. coli* strain JM101. Small-scale cultures of cells were grown overnight and used to inoculate 1−3 L of media at a ratio of 1:40. The protein was labeled with ^{15}N by growing the cells on M9 minimal medium (*26*) supplemented with 1 g/L [U-^{15}N]ammonium chloride (99 at. %) (Cambridge Isotope Laboratories, Woburn, MA) as the sole nitrogen source. The inoculated cultures were grown at 37 °C for approximately 3 h before induction with nalidixic acid (50 μg/mL). Cell pellets were harvested approximately 4 h after induction and frozen at −70 °C until they were needed.

Thawed cells were suspended in 10 mM Tris-HCl (pH 8) containing 1 mM EDTA and sonicated and the refractile bodies isolated by centrifugation. The refractile bodies were washed twice by resuspension in Tris buffer and collected by centrifugation. Refractile bodies were solubilized for 1 h at ambient temperature by suspending in 150 mL (per gram) of 8 M urea and 50 mM Tris-HCl (pH 9.5) containing 5 mM dithiothreitol (DTT). Folding was accomplished by diluting the urea solution to 2 M with deionized water, followed by stirring for 20−50 h at 4 °C. The folding reaction was monitored by reverse-phase HPLC. After the pH was adjusted to 5 with acetic acid, the folding solution was incubated for 1 h, and then clarified by centrifugation at 13000*g* for 30 min. The solution was incubated for another 1 h at 4 °C; the pH was further lowered to 4.0 with acetic acid, and the solution was filtered to remove debris.

Purification was accomplished by cation-exchange chromatography on a CM Fast Flow column (1.5 cm × 5 cm, Pharmacia) equilibrated with 10 mM sodium acetate (pH 4.0). After the protein sample was loaded, the column was eluted with two linear gradients consisting of 0 to 0.5 M NaCl followed by 0.5 to 1.0 M NaCl. Fractions that were >95% pure in the major peak were pooled and dialyzed against 50 mM ammonium acetate (pH 3.9). Samples were characterized by reverse-phase HPLC, SDS−PAGE, electrospray mass spectrometry, N-terminal sequence analysis, and amino acid composition analysis. The protein concentrations were determined by either absorbance at 280 nm or quantitative amino acid analysis.

Cell Surface G-CSF Receptor Binding Assays. The potency of cpG-CSF for cell surface G-CSFR was determined using mouse BaF3 cells transfected to express G-CSFR (BaF3-G; *25*). Binding to G-CSFR was detected by competition with [^{125}I]G-CSF(Y1, Y3, S17) (specific activity of 1000−2000 Ci/mmol). Binding assays were initiated by addition of ice-cold transfected cells to aliquots containing ^{125}I-labeled ligand with and without the competing ligand. Competition was carried out at 0−4 °C until steady state was achieved in approximately 0.5−3 h. The amount of bound ligand was determined by counting the cellular pellet after centrifugation for 1 min at 12000*g*. Nonspecific binding was defined as the residual binding which occurred in the presence of excess G-CSF (400 nM). Data were analyzed using logit-log plots.

Surface Plasmon Resonance (SPR) Measurements of Association and Dissociation Binding Kinetics. The extracellular domain of the G-CSFR was fused to the F_C portion of a mouse IgG domain (sG-CSFR-F_C), expressed in baculovirus-infected insect S9 cells, and collected by concentration of conditioned media. Association and dissociation kinetic rate constants for binding of cpG-CSFs to affinity-captured sG-CSFR-F_C were determined using surface plasmon resonance in a BIAcore model 2000 instrument. Rabbit anti-mouse F_C ($RAMF_C$) served as the capture reagent by covalently coupling it to sensor chips using an ethyldimethylaminocarbodiimide amine coupling kit (BIAcore, Piscataway, NJ). sG-CSFR-F_C was captured by the $RAMF_C$ at a flow rate of 10 μL/min in HBS buffer [10 mM HEPES (pH 7.4), 150 mM NaCl, 3.4 mM EDTA, and 0.005% (v/v) P-20 surfactant].

Binding of cpG-CSFs to $RAMF_C$-captured sG-CSFR-F_C was assessed by passing various concentrations (0.5−10 × K_d) of the cpG-CSFs over the surface at a flow rate of 20 μL/min in HBS at 25 °C. After a plateau in binding was achieved, the bound cpG-CSF was eluted with HBS buffer to monitor dissociation. The $RAMF_C$ surface was regenerated by washing with 1 N formic acid. Software provided by the manufacturer (BIAcore) was used to simultaneously fit the family of sensorgrams using a nonlinear least-squares global fitting procedure. The association (k_a) and dissociation (k_d) rate constants are treated as adjustable parameters to minimize the root-mean-square deviation of the residual between the experimental and calculated curves.

Cell Proliferation Assays. G-CSF proliferative activity was assessed using BaF3-G cells (*25*). Cells were seeded at 2.5

$\times 10^4$ cells per well in Costar microwell plates with or without cytokines in serum-free assay medium. The assay medium consisted of Iscove's modified Dulbecco's medium (IMDM; Gibco, Grand Island, NY) supplemented with bovine serum albumin (BSA, 100 mg/mL, Boehringer Mannheim, Indianapolis, IN), human transferrin (30 mg/mL, Sigma), 50 μM 2-mercaptoethanol, and a lipid substitute consisting of 2.5 mg of phosphatidylcholine/mL of BSA, resulting in final assay concentrations of 500 μg/mL, 100 μg/mL, 50 μM, and 6.25 μg/mL, respectively. After 48 h, cells were incubated with [*methyl*-^{3}H]thymidine at 0.5 μCi (18.5 kBq) per well for 18−20 h and then harvested onto a glass fiber filter mat for measurement of radioactivity with a β-counter (Wallach, Gaithersburg, MD).

Far-UV Circular Dichroism Spectroscopy. The far-UV CD spectra of G-CSF(S17) and cpG-CSFs were measured at ambient temperature using a JASCO J-500C spectrometer calibrated using a standard sample of *d*-(+)-10-camphorsulfonic acid. Protein samples at 20−50 μM in 10 mM acetate buffer (pH 4.0) were placed in a quartz cell (path length of 2 mm). Spectra were collected by averaging four consecutive scans taken between 260 and 190 nm in 0.2 nm increments. The results are expressed as the molar ellipticity, [θ]. The helical content was calculated according to the methods of Yang (*27*).

NMR Spectroscopy Measurements. All spectra were collected on ^{15}N-labeled G-CSF(S17) and cpG-CSF samples at concentrations ranging from 0.5 to 1.5 mM in 50 mM deuterated acetate at pH 3.9 and 30 °C. Experiments were carried out using a triple-resonance Varian Unity 500 MHz spectrometer equipped with a pulsed field gradient unit and a 5 mm triple-tuned probe. The low-pH condition was used to improve the solubility of these proteins and to match the conditions of the published resonance assignments (*12*, *14*). Solvent suppression was achieved by using gradient pulses, and in all experiments, the enhanced sensitivity method was employed (*28*). All spectra were recorded with spectral widths of 7500 Hz (F_2, ^{1}H) and 1530−1550 Hz (F_1, ^{15}N).

^{15}N HSQC spectra were acquired using the pulse sequence described by Zhang et al. (*29*) with acquisition times of 137 ms (F_2, ^{1}H) and 240 ms (F_1, ^{15}N). Backbone dynamic information was collected using two-dimensional (2D) ^{15}N-{^{1}H} NOE measurements (*30*) with acquisition times of 137 ms (F_2, ^{1}H) and 105 ms (F_1, ^{15}N). Spectra in the presence of a ^{15}N{^{1}H} NOE were generated by saturating protons for 3 s after a 0.6 s delay, while the reference experiment was identical except that it substituted a compensating delay of 3 s for the proton saturation pulses. The data along F_1 were expanded to 160 ms using the linear prediction routine implemented in Triad (Tripos, St. Louis, MO). Squared cosine and Gaussian window functions were applied along F_1 and F_2, respectively. The resonance assignments of the cpG-CSFs were obtained by comparing 2D ^{15}N HSQC, 3D ^{15}N-edited TOCSY (*29*), and NOESY (*29*) spectra with the published assignments of G-CSF (*14*) and G-CSF(S17) (*12*). The 3D spectra were collected with acquisition times of 69 ms (F_3, ^{1}H), 17 ms (F_2, ^{1}H), and 20 ms (F_1, ^{15}N) and were processed using Triad to extend the F_1 dimension to 40 ms and yield matrixes of 256 $\times$ 256 $\times$ 128 real points.

Under the experimental conditions, all cpG-CSF samples were stable beyond the duration of the experiments except for cpG-CSF(L1)[39/38], which exhibited a remarkable decrease of overall signal intensity over a period of only several hours. To compensate for this instability, three data sets rather than two were recorded to estimate the ^{15}N{^{1}H} NOEs in cpG-CSF(L1)[39/38]: the first in the absence of ^{15}N{^{1}H} NOEs, the second in the presence of ^{15}N{^{1}H} NOEs, and the third again in the absence of ^{15}N{^{1}H} NOEs. The average cross-peak intensity in the absence of ^{15}N{^{1}H} NOEs (i.e., from the first and third data sets) was used to calculate the ^{15}N{^{1}H} NOEs presented in Figure 5E.

Urea Denaturation Curves. Urea-induced unfolding of cpG-CSFs was detected by monitoring changes in ellipticity at 222 nm as a function of increasing urea concentrations. Samples (final concentrations of 37−146 μM) in 35 mM sodium acetate (pH 4.0) were prepared in the presence of various concentrations of urea using the method described by Pace (*31*). Each curve was defined using at least 15 concentrations of urea (UltraPure grade, ICN Biochemicals) between 0 and 9 M. After equilibration at ambient temperature for at least 4 h, samples were placed in a 0.5 cm quartz cell and spectra were scanned from 260 to 200 nm in 0.2 nm increments. Values of the molar ellipticity at 222 nm ([θ_{222}]) were calculated after subtraction of an appropriate buffer blank.

Denaturation curves were defined by a four-parameter nonlinear least-squares fitting of the ellipticity data (*32*). The fraction of unfolded molecules, F_u, was calculated using the equation

$$F_u = ([\theta_{222}]_F - [\theta_{222}]_C)/([\theta_{222}]_F - [\theta_{222}]_U) \qquad (1)$$

where $[\theta_{222}]_C$ corresponds to the molar ellipticity of the sample at a given urea concentration and $[\theta_{222}]_F$ and $[\theta_{222}]_U$ are the values for the folded and unfolded forms obtained by extrapolation from linear fits of the $[\theta_{222}]_C$ data in the pre- and post-transition regions, respectively. The urea concentration corresponding to the midpoint of the unfolding transition (C_{mid}) and the free energy change (ΔG_{urea}) for urea-induced denaturation were determined as described by Pace (*31*).

RESULTS AND DISCUSSION

The subject of this paper is the effect of circular permutation on the conformation, stability, and activity of the G-CSF sequence. Circular permutation of protein sequences can be understood in terms of linking N- and C-termini and introducing new termini at locations in the sequence which do not disrupt the chain fold or the biological activity. In the following paper, we have evaluated the effect of circularly permuted G-CSF sequences in the context of myelopoietin (MPO), a dual IL-3R and G-CSFR agonist. In this study, we have investigated the effect of circular permutation on isolated G-CSFR agonist domains. The expression and purification of the isolated circularly permuted molecules allowed for a battery of biological, biochemical, and biophysical assessments.

Selection of the Circularly Permuted Sequences. The location for the new termini (the "breakpoints") in G-CSF and the choice of the linker between the original termini were selected on the basis of the 3D structure of human G-CSF (*10*−*14*). The G-CSF structure consists of four long helices that are approximately 20−30 residues in length packed into an antiparallel bundle. The helices are numbered 1−4 in the

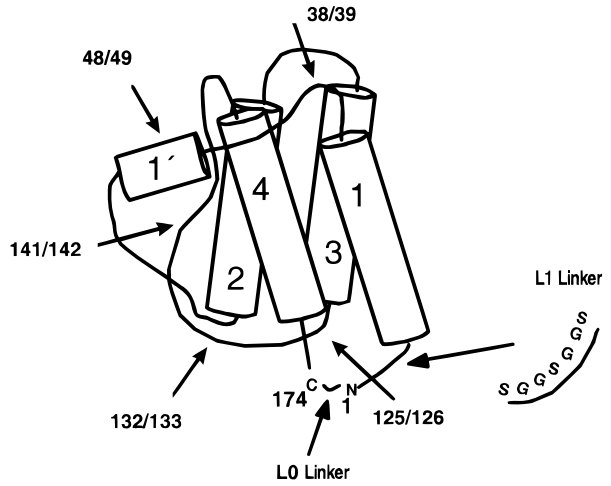

FIGURE 1: Schematic representation of the 3D structure of G-CSF. The locations of new termini (breakpoints) and linkers used in the cpG-CSFs are indicated by arrows and lines, respectively.

order of their occurrence in the sequence (Figure 1). Two long overhand loops (loops 12 and 34) connect helix 1 to 2 and helix 3 to 4, respectively. A two-turn helical structure, referred to as the minihelix, occurs near the beginning of loop 12.

In this study, five breakpoints are placed within the overhand loops of the G-CSF structure on the surface of the protein. The first breakpoint places the new amino- and carboxy-terminal pair at residues 39 and 38 of the original sequence, just after the end of helix 1 in the beginning of loop 12. The second breakpoint occurs at residues 49 and 48 in the minihelix. The third through fifth breakpoints (126 and 125, 133 and 132, and 142 and 141) occur in the beginning, middle, and end of loop 34, respectively. Thus, the breakpoints employed are placed along the loops spanning the helical bundle from top to bottom. An additional cpG-CSF with a breakpoint at residues 97 and 96 in loop 23 was constructed, but could not be studied due to the low level of expression.

In the design of circularly permuted protein sequences, attempts are generally made to choose linkers that accurately span the distance between the termini to avoid the strain caused by linkers that are too short, or the additional entropy that results when linkers are too long. Selection of an accurate linker for the G-CSF structure was hampered by the fact that the conformation of the first 10 residues at the amino terminus is severely disordered in both X-ray (*10*, *11*) and NMR (*12*, *14*) structures. Therefore, the appropriate number of residues for connecting helix 1 and helix 4 was estimated from the distance between residue 11 and residue 171, the first and last residues in the chain that are well-defined in the structure, respectively. Allowing for the uncertainty of this exercise, one of the linkers chosen for this study connects the original amino terminus to the carboxy terminus, i.e., residue 1 to residue 174, providing 13 residues between helix 4 and helix 1 (see Figure 1). This linker is referred to as the L0 linker. The second linker, referred to as the L1 linker, deletes the amino-terminal 10-residue segment and replaces it with a seven-residue sequence consisting of SGGSGGS. The net effect is that the L1 linker connects helix 1 and helix 4 with a 10-residue segment, three residues shorter than the L0 linker.

The unpaired Cys-17 in the wild-type G-CSF was replaced with Ser in all of the cpG-CSFs reported in this study. This mutation has been shown to have no effect on either the biological activity or thermal stability of G-CSF (*6*, *33*). The purified proteins were characterized using a number of analytical methods, including SDS−PAGE, electrospray mass spectrometry, quantitative amino acid analysis, and N-terminal sequencing (data not shown).

Studies of Binding of cpG-CSFs to Cell Surface G-CSFR. A competitive binding assay was used to assess the potencies of the cpG-CSFs for binding to the G-CSFR expressed on the surface of the BaF3-G cells (*25*). The cpG-CSFs inhibited binding to G-CSFR with concentrations that gave 50% displacement of bound [^{125}I]G-CSF(Y1, Y3, S17) between 0.86 and 2.7 nM (see Table 2). These IC_{50}s are comparable to the value of 0.95 nM for wild-type G-CSF. Only three of the cpG-CSFs, namely, cpG-CSF[39/38], cpG-CSF[49/48], and cpG-CSF(L1)[142/141], had small albeit statistically significant ($p < 0.05$) reductions in potency (1.6-, 2.8-, and 2.6-fold, respectively) relative to that of G-CSF. All of the cpG-CSFs reached the same level of maximal binding as the wild-type G-CSF. Thus, all of the circularly permutated G-CSF sequences were able to attain a conformation capable of recognizing the G-CSFR with only a modest variation in their binding affinities.

Surface Plasmon Resonance Studies of G-CSFR Binding. The binding kinetics determined using surface plasmon resonance revealed that while most cpG-CSFs studied have similar association and dissociation rate constants compared with those of G-CSF, two of them exhibited significant differences (Table 2). The dissociation rate constant of cpG-CSF[49/48] is 8-fold greater than that of G-CSF, and the association rate constant of cpG-CSF(L1)[142/141] is ca. 20-fold smaller than that of G-CSF. These findings at ambient temperatures are similar to the results of the competitive binding assay obtained at 4 °C.

Cell Proliferative Activity. The ability of cpG-CSFs to stimulate proliferation was assessed in the BaF3-G cell line. As shown in Table 2, all cpG-CSFs with the L0 linker, except for cpG-CSF[49/48], had similar proliferative effects with concentrations giving rise to a half-maximal proliferation (EC_{50}) of 88−140 pM. The maximal proliferative responses of all of the molecules were the same within experimental error (data not shown), with the exception of that of cpG-CSF(L1)[142/141], for which only a partial dose−response curve could be obtained. Among the molecules employing the L0 linker, cpG-CSF[49/48], with an EC_{50} of 1290 pM, exhibited the greatest decrease in potency (22-fold) compared to G-CSF. In contrast, the proliferative activities of cpG-CSFs with the L1 linker (Table 2) had a broader range of values, ranging from being approximately 10- to 100-fold less potent than G-CSF.

The differences in potencies in the proliferation assay for the L1 linkers are significantly larger than those seen in the receptor binding assay. There are many possible explanations for these differences. Compared to the competitive receptor binding or the surface plasmon resonance experiments, stimulation of proliferation is a significantly more complex process occurring over a much longer period of time. Receptor signaling involves ligand binding, association of subunits into dimers or higher-order oligmers, phosphorylation of cytoplasmic domains, and triggering of signaling

Table 2: Summary of G-CSF Receptor Agonist Properties[a]

growth factor	competitive receptor binding	soluble receptor binding[b]			cell proliferation	thermodynamic parameters	
	IC_{50}(G-CSFR) (nM)	k_a ($\times 10^5$ s^{-1} M^{-1})	k_d ($\times 10^{-4}$ s^{-1})	K_D (nM)	EC_{50}(Baf3-G) (pM)	ΔG (kcal/mol)	urea midpoint (M)
G-CSF	0.95 ± 0.17 (4)	10.0 ± 3.0	2.0 ± 0.4	0.20 ± 0.07 (7)	57 ± 13 (2)	nd	nd
G-CSF(S17)	nd	8.8 ± 0.7	2.6 ± 1.5	0.30 ± 0.18 (9)	51 ± 8 (2)	9.6 ± 0.3 (2)	5.53 ± 0.36
cpG-CSF[39/38]	1.5 ± 0.3 (3)	4.1 ± 0.2	3.8 ± 0.8	0.93 ± 0.15 (3)	140 ± 31 (2)	8.4 ± 0.7 (2)	4.93 ± 0.02
cpG-CSF(L1)[39/38]	1.0 ± 0.2 (3)	6.7 ± 0.4	3.8 ± 0.8	0.56 ± 0.10 (3)	1290 ± 15 (2)	7.0 ± 0.6 (2)	4.45 ± 0.04
cpG-CSF[126/125]	1.5 ± 0.9 (5)	nd	nd	nd	126 ± 25 (2)	8.1 ± 0.2 (3)	4.81 ± 0.31
cpG-CSF(L1)[126/125]	1.1 ± 0.3 (6)	nd	nd	nd	1720 ± 538 (2)	7.2 ± 0.9 (2)	4.43 ± 0.10
cpG-CSF[133/132]	0.86 ± 0.12 (3)	10.0 ± 2.0	1.8 ± 0.6	0.17 ± 0.06 (8)	88 ± 1 (2)	7.9 ± 0.7 (2)	4.70 ± 0.34
cpG-CSF(L1)[133/132]	1.1 ± 0.2 (3)	10.0 ± 1.0	1.2 ± 0.4	0.12 ± 0.04 (9)	565 ± 166 (2)	7.3 ± 0.4 (3)	4.62 ± 0.29
cpG-CSF[142/141]	1.1 ± 0.4 (4)	nd	nd	nd	113 ± 28 (2)	9.1 ± 1.1 (2)	4.85 ± 0.27
cpG-CSF(L1)[142/141]	2.5 ± 0.2 (3)	0.5 ± 0.06	4.4 ± 0.2	8.9 ± 0.1 (3)	5000[c] (2)	5.8 ± 0.3 (2)	4.43 ± 0.14
cpG-CSF[49/48]	2.7 ± 1.3 (3)	6.0 ± 0.5	16.0 ± 2.0	2.6 ± 0.1 (3)	1290 ± 138 (2)	7.3 ± 0.5 (2)	4.46 ± 0.11

[a] Note that the competitive binding experiments were performed at 4 °C while the plasmon resonance experiments were performed at 25 °C. nd means not determined. Numbers in parentheses indicate the number of independent experiments performed. [b] Binding to affinity-captured sG-CSFR-F_C detected by surface plasmon resonance. [c] The EC_{50} was estimated from an incomplete response curve.

mechanisms that lead to gene activation and cell cycling. In addition, there is receptor cycling consisting of internalization, dephosphorylation by phosphatases, and transport back to the cell surface. Thus, the difference between receptor binding and proliferation data could be due to subtle differences in ligand docking, which could in turn lead to conformational changes that are less effective in inducing subunit association and/or phosphorylation. In addition, proliferation is assessed at 37 °C, whereas receptor binding and surface plasmon resonance measurements are taken at 4 and 25 °C, respectively. It might be that the cpG-CSFs are able to efficiently form an initial receptor complex, but are attenuated in their ability to cause the necessary downstream conformational changes and oligomerization of the receptor required for proliferation. Alternatively, given the length and the temperature of the assay, it may be that the cpG-CSFs have variable degrees of stability, and that they may not be maintained in their biologically active conformation sufficiently long to give full stimulation of the proliferative response. A similar discrepancy between the potency for binding to the receptor and proliferative activity has been reported for point mutants of G-CSF (*8*). To assess these possibilities, CD and NMR spectroscopy were used to characterize the conformation of cpG-CSFs and urea-induced denaturation was used to gauge the thermodynamic stability.

Far-UV CD Helical Profile of the cpG-CSFs. Far-UV CD spectra were collected for G-CSF(S17) and each of the cpG-CSFs between 190 and 260 nm to assess the integrity of the secondary structure upon circular permutation. A superposition of the CD spectra for the circularly permuted molecules constructed with the L0 linker is shown in Figure 2A. The double-trough profiles with minima at 207 and 222 nm are characteristic of a helical protein. The spectra and the calculated helical contents (51−58%) of the L0 linker cpG-CSFs are virtually identical to that of G-CSF(S17) (56% helical), with only cpG-CSF[39/38] showing a perceptible difference. An overlay of the CD spectra for the cpG-CSFs which use the shorter L1 linker is shown in Figure 2B. The L1 linker is associated with a much greater range of differences than is seen for the L0 linker molecules, similar to the situation observed in the BaF3-G proliferation assay. In particular, cpG-CSF(L1)[142/141] and cpG-CSF(L1)[126/125] are distinctly different from G-CSF(S17) with 31 and

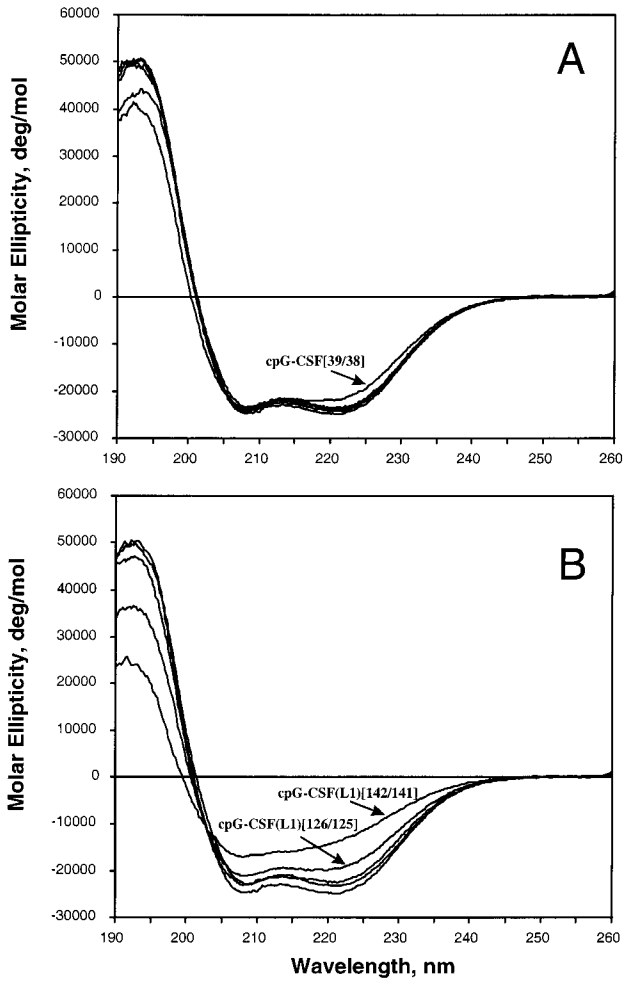

FIGURE 2: Far-UV circular dichroism spectra of (A) cpG-CSF molecules with L0 linkers and (B) cpG-CSF molecules with L1 linkers. Spectra were collected on samples at concentrations between 20 and 50 μM in 10 mM sodium acetate (pH 4.0) at ambient temperature and are presented as the average of four consecutive scans. The spectra for molecules that exhibit significant differences are labeled.

46% helix, respectively. The decrease in the helicity of cpG-CSF(L1)[142/141] to about half of that of G-CSF(S17) is consistent with its decreased potency for receptor binding (2.6-fold by cell surface binding and ca. 20-fold by surface plasmon resonance detection) and proliferative activity (ca.

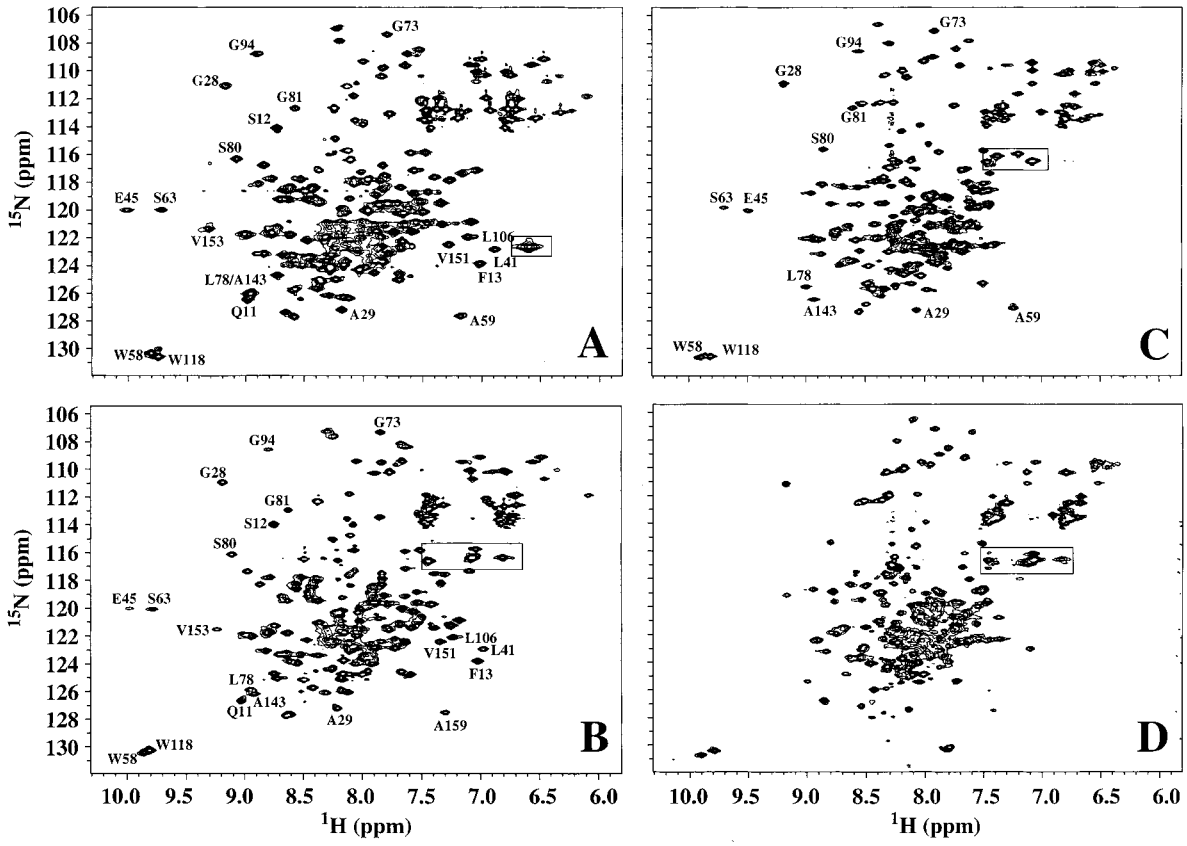

FIGURE 3: ^{1}H−^{15}N HSQC spectra of (A) G-CSF(S17), (B) cpG-CSF[39/38], (C) cpG-CSF(L1)[133/132], and (D) cpG-CSF(L1)[142/141]. All spectra were collected on samples with concentrations between 0.5 and 1.5 mM in 50 mM deuterated acetate at pH 3.9 and 30 °C. A selected set of cross-peaks are labeled with their resonance assignments in panels A−C. No resonances were assigned for cpG-CSF(L1)-[142/141] in panel D. Cross-peaks enclosed in boxes are folded in from outside the spectral window.

100-fold). cpG-CSF(L1)[126/125], with a smaller decrease in helical content, maintained its potency for receptor binding but experienced a significant decrease in activity (30-fold) in the BaF3-G proliferation assay.

Backbone Chemical Shifts of cpG-CSF. The CD results give strong evidence for the preservation of the overall helical content in the cpG-CSFs. To investigate possible structural and dynamical differences in more detail, ^{15}N-labeled samples of G-CSF(S17) and all of the cpG-CSFs were prepared for study by heteronuclear NMR spectroscopy. The pattern of cross-peaks in an ^{1}H−^{15}N HSQC spectrum provides a sensitive fingerprint that reflects the local chemical and conformational environment for amide proton−amide nitrogen pairs along a polypeptide chain in its specific 3D fold (*34*, *35*). Conservation of the chemical shifts of a majority of cross-peaks is a reliable indicator of the conservation of the structure between a pair of related proteins, although chemical shift changes may result from only subtle conformational changes.

The spectra of cpG-CSF(L1)[39/38], cpG-CSF(L1)[133/132], cpG-CSF[39/38], cpG-CSF[133/132], and cpG-CSF-[49/48] were subjected to detailed 3D NMR analysis. Many of the amide proton and amide nitrogen resonance assignments can be transferred from G-CSF(S17) (*12*) to the cpG-CSFs and confirmed using a combination of ^{15}N-edited TOCSY and NOESY spectra. Characteristic 2D ^{1}H−^{15}N HSQC spectra are shown in Figure 3A−D. The amide fingerprint region of G-CSF(S17) (Figure 3A) is well dispersed with a pattern of cross-peaks in good agreement with previously published assignments (*12*). Similarly, the patterns of cross-peaks for cpG-CSF[39/38] and cpG-CSF-(L1)[133/132] shown in panels B and C of Figure 3, respectively, are quite similar to Figure 3A, reflecting the correspondence in the backbone 3D fold of these molecules with that of the native fold.

For those cpG-CSFs studied in detail, the major differences in the C_αH chemical shifts from those of G-CSF(S17) (Figure 4) are found at residues that are either near the new termini or near the circularizing linker, thus reflecting localized sequence and conformational changes in these regions. The degree of perturbation in the G-CSF fold that results from circular permutation depends on the spatial location of the breakpoints as well as the type of linker used. These perturbations are reflected by the extent of C_αH chemical shift changes. For cpG-CSF[133/132] (Figure 4C), the perturbation is, in general, smaller and highly localized to residues immediately adjacent to the new termini and the linker. For cpG-CSF[39/38] (Figure 4A), in which the breakpoint is located between helix 1 and the minihelix, the perturbation extends to both the minihelix and the N-terminus of helix 4, the latter being in close apposition to residues 38 and 39. Similarly, the perturbation in cpG-CSF[49/48] (Figure 4B) can be seen in helix 1, helix 3, helix 4, and the portion of loop 12 following the minihelix, where residues 48 and 49 are located.

Additional perturbations were observed for the corresponding cpG-CSFs with the L1 linker [i.e., cpG-CSF(L1)-[39/38] in Figure 4D and cpG-CSF(L1)[133/132] in Figure 4E]. The cross-peaks corresponding to the segment between residues 11 and 13 are perturbed to such an extent in the

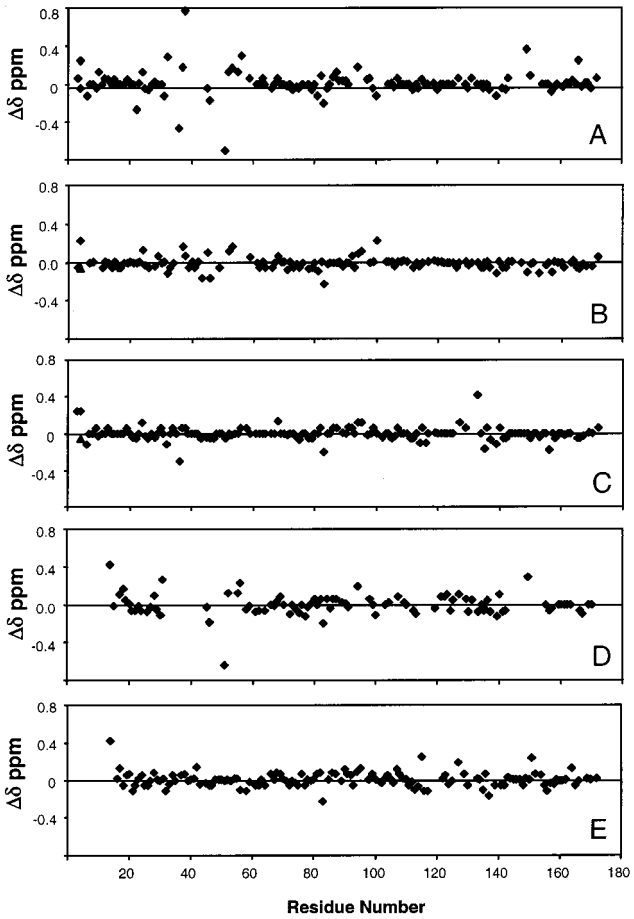

FIGURE 4: $C_\alpha H$ chemical shift differences between cpG-CSFs and G-CSF(S17). All spectra were collected on samples with concentrations between 1.0 and 1.5 mM in 50 mM deuterated acetate at pH 3.9 and 30 °C: (A) cpG-CSF[39/38], (B) cpG-CSF[49/48], (C) cpG-CSF[133/132], (D) cpG-CSF(L1)[39/38], and (E) cpG-CSF(L1)-[133/132].

spectra of both of these molecules (e.g., compare panels A and B of Figure 3, as well as data not shown) that they could not be identified by comparison with those of G-CSF(S17). A partial unwinding of the N-terminal portion of helix 1 of the L1 linker cpG-CSFs was suggested by the absence of a H^N-H^N NOE between F13 and L14 in the spectra of either cpG-CSF(L1)[39/38] or cpG-CSF(L1)[133/132]. In contrast, the $C_\alpha H$ resonances from the C-terminus of helix 4 for the L1 molecules are nearly identical to those in G-CSF(S17) (panels D and E of Figure 4) and the pattern of helical NOE connectivities (data not shown) was observed up to residue 172, thus documenting the preservation of helix 4. The chemical shift profile (panels D and E of Figure 4) also indicates that the C-terminus of helix 3, which is spatially adjacent to the L1 linker, is also perturbed in these cpG-CSFs.

The example spectra shown in panels B and C of Figure 3 are representative of those of all of the cpG-CSFs except for that of cpG-CSF(L1)[142/141]. The $^1H-^{15}N$ HSQC spectrum of the latter molecule (Figure 3D) lacks the dispersion and resolution evident in the spectra of the other cpG-CSFs. This result is consistent with the low helicity observed by CD spectroscopy, and suggests a large-scale conformational change which contributes to its low proliferative activity.

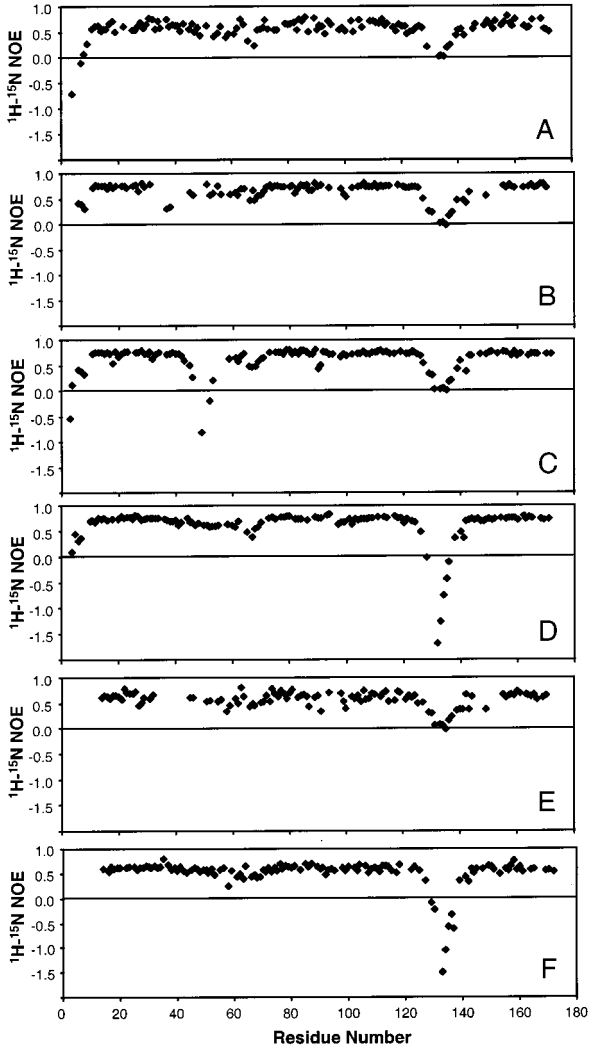

FIGURE 5: Profile of backbone $^{15}N\{^1H\}$ NOEs for G-CSF(S17) and cpG-CSFs. All solution conditions are the same as described in the legend of Figure 4: (A) G-CSF(S17), (B) cpG-CSF[39/38], (C) cpG-CSF[49/48], (D) cpG-CSF[133/132], (E) cpG-CSF(L1)-[39/38], and (F) cpG-CSF(L1)[133/132].

Pattern of Backbone $^{15}N\{^1H\}$ NOEs for the cpG-CSFs. The $^{15}N\{^1H\}$ NOE at a backbone amide nitrogen is a quantity that reflects the amplitude and frequency of motion at the polypeptide backbone (*36*). By measurement and comparison of this quantity in the cpG-CSFs with that of G-CSF(S17), a pattern reflecting changes in the relative motion was characterized.

Figure 5A depicts the pattern of $^{15}N\{^1H\}$ NOEs along the peptide backbone for G-CSF(S17). This pattern is essentially identical to that reported previously (*12*). The strongly negative NOEs seen at the extreme amino terminus reflect the highly flexible nature of this segment of the peptide, a result which is consistent with the crystallographic studies (*10*). A pattern of NOEs with values of >0.6 is observed for long stretches of the backbone, reflecting the relatively rigid helical core of the protein. A segment corresponding to loop 34 shows a dip with the NOEs approaching zero, indicating the flexibility of this loop in the native sequence.

The pattern of $^{15}N\{^1H\}$ NOEs for the helical core of the cpG-CSFs is quite similar to that of G-CSF(S17). Important differences are observed where the original termini have become linked and where new termini have been introduced

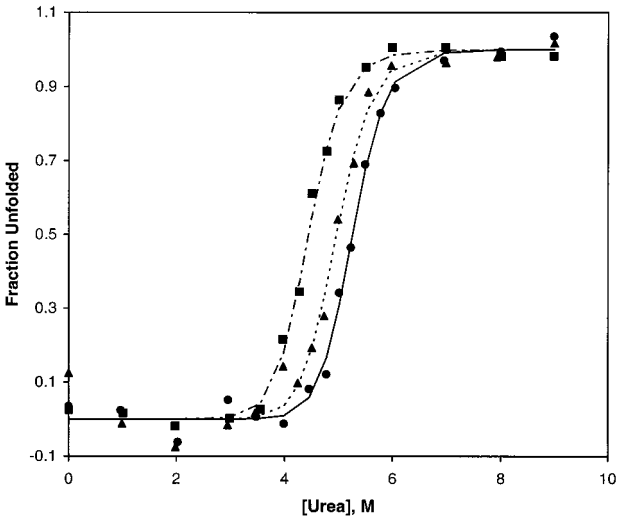

FIGURE 6: Representative urea-induced denaturation curves of cpG-CSFs. Samples were at concentrations between between 37 and 146 μM in 35 mM sodium acetate at pH 4 and 20 °C: (●) G-CSF, (▲) cpG-CSF[39/38], and (■) cpG-CSF(L1)[39/38].

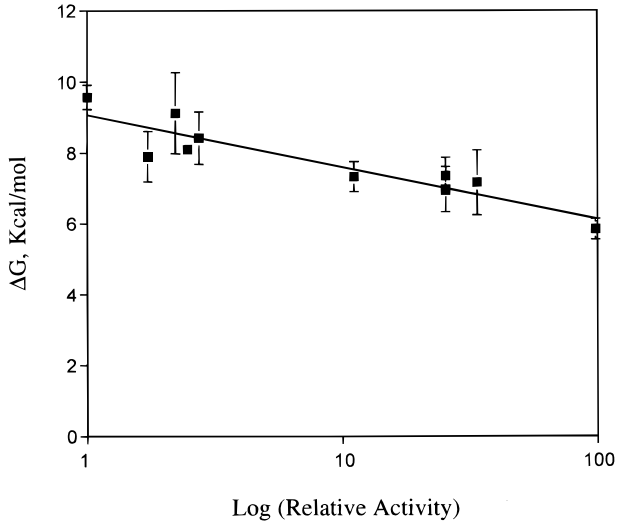

FIGURE 7: Scatter plot of ΔG_{urea} vs log relative activity in the BaF3-G cell proliferation assay for the cpG-CSFs and G-CSF(S17). The concentration of native G-CSF that gives half the maximal response is used to calculate the relative activity. The line is a linear least-squares fit of the data points with an r^2 value of 0.84.

within the sequence. For example, in panels B and D of Figure 5, corresponding to the NOE patterns for cpG-CSF-[39/38] and cpG-CSF[133/132], the NOEs at the amino-terminal segment are found to be close to zero, rather than the highly negative values seen for the native sequence (Figure 5A). This finding indicates an increase in the rigidity resulting from the linkage. Introduction of the new termini at residues 39 and 38 leads to a noticeable dip in the NOE near that position in the sequence. A similar decrease in NOEs is seen when new termini are introduced at residues 133 and 132 within loop 34 (Figure 5D), where the NOEs change from being close to zero to becoming strongly negative. Similar trends relating changes in mobility to the effects of linkers and breakpoints can be seen in the corresponding amide backbone NOE profiles for cpG-CSF-[49/48], cpG-CSF(L1)[39/38], and cpG-CSF(L1)[133/132] in panels C, E, and F of Figure 5, respectively.

Urea-Induced Cooperative Unfolding of cpG-CSFs. NMR and CD spectroscopy give a consistent picture in which all of the cpG-CSFs, with the exception of cpG-CSF(L1)[142/141], and to a lesser extent cpG-CSF(L1)[126/125], are able to fold into active 3D structures that strongly resemble that of native G-CSF in terms of both structure and backbone dynamics. Thus, conformational differences are unable to provide an explanation for why some of the cpG-CSFs have a discrepancy between the potency for binding to the receptor and proliferative activity. To assess the possibility that it is stability and not conformation per se which is responsible for these observed differences, the thermodynamic stability of the cpG-CSFs was investigated using CD-detected urea-induced denaturation.

The effect of 0−9 M urea on the helical content of G-CSF-(S17) and the cpG-CSFs was assessed by monitoring the circular dichroism at 222 nm. All of the cpG-CSFs underwent cooperative unfolding transitions as reflected in the representative examples shown in Figure 6. For example, cpG-CSF[39/38] had a transition that is nearly indistinguishable from that of G-CSF(S17), while the midpoint of the denaturation curve for cpG-CSF(L1)[39/38] was shifted by about 0.5 M to ca. 4.5 M urea. At the other extreme, cpG-CSF(L1)[142/141], which has been shown to have a severely perturbed structure compared with the other cpG-CSFs, nevertheless underwent a cooperative unfolding, although it did so with a ΔG_{urea} lower than those of the other cpG-CSFs (see below).

A summary of the thermodynamic parameters derived by analysis of the urea denaturation curves (*31*) is given in Table 2. G-CSF(S17) has a ΔG_{urea} of 9.3 kcal/mol, while the ΔG_{urea} values for the cpG-CSFs range from 4.8 to 10.1 kcal/mol. A pattern that emerges from Table 2 is that the ΔG_{urea} values for the molecules with the L1 linkers are generally lower than those with the L0 linkers. These findings indicate that the shorter L1 linker destabilizes the circularly permuted G-CSF sequences, perhaps because of strain introduced by forcing the termini to be closer than is optimal and/or the perturbation to the stability of helix 1 and helix 4 by the L1 linker.

There is no direct method for assessing the role of thermodynamic stability in the proliferative activity of the cpG-CSFs. However, a relationship between these parameters can be explored by plotting the ΔG_{urea} values against the relative proliferative activity (1.0 for G-CSF), as shown in Figure 7. It is seen that there is in fact a correlation between ΔG_{urea} and the proliferative activity with an r^2 value of 0.84. Although the complexity of events that determine proliferation makes it difficult to directly relate it to the thermodynamic analysis, the similarity of the conformation of the cpG-CSFs (with the exception of cpG-CSF(L1)[142/141]) together with the correlation suggests that stability and its implicit connection with ligation and receptor signaling may be important in determining the bioactivity of the cpG-CSFs.

Relationship between Breakpoint Location and the Receptor Binding Sites. Regions or residues of hG-CSF important for receptor binding and proliferation activity have been identified using neutralizing monoclonal antibodies and site-specific mutagenesis (*6*−*9*). The interpretation of these studies has been guided by the 3D structure of G-CSF (*10*−*14*).

The N-terminal segment which precedes helix 1 is not required for the biological activity of G-CSF (*6*), although

mutations in this region resulted in a more potent molecule. Epitope mapping of neutralizing monoclonal antibodies pointed to the functional importance of residues 20−46 (helix 1 and a portion of loop 12) (*7*). Systematic substitution with alanines revealed a number of residues that are likely to be involved in receptor binding. One study proposed that L15, E19, K40, V48, L49, D112, L124, and F144 contribute to receptor binding sites (*8*). In another study (*9*), alanine mutants K16A, E19A, K23A, D109A, and D112A were reported to cause significant decreases in both proliferation activity (>7-fold) and receptor binding affinity (>4-fold), therefore forming the receptor contact surface.

Four of the five breakpoints investigated in this study are implicated as being either at or near the receptor binding site as described above, i.e., breakpoints 39 and 38, 49 and 48, 126 and 125, and 142 and 141. The new termini introduce a pair of charges at a location where a covalent linkage exists in the native structure. The introduction of ionizable groups might be expected to introduce dramatic changes in the local environment. The dynamic behavior of the backbone nitrogen atoms near the breakpoints seems to rule out the possibility of the new termini forming a stable salt bridge. Despite the proximity of the breakpoints to the putative receptor binding sites, all the L0 linker cpG-CSFs, with the sole exception of cpG-CSF[49/48], retained receptor binding activity and cell proliferative activity similar to those of the wild-type G-CSF. Apparently, the new terminal residues are able to maintain a conformation that does not seriously affect the ligand−receptor interaction. The structural perturbations introduced by the breakpoints are highly localized, a conclusion supported by the results of the heteronuclear NMR studies. These findings suggest that small displacements at the new termini can accommodate the new charges and thereby permit them to be well-solvated by surrounding water. In light of the diminished affinity for the receptor and the increased dissociation rate, the decrease in the potency for cpG-CSF-[49/48] appears to arise because the perturbation to the conformation of the minihelix is detrimental to the optimal fit between the receptor and the ligand. If the surface-exposed hydrophobic side chains of V48 and L49 are receptor contact residues in the native structure, even small distortions of the minihelix in cpG-CSF[49/48] may prevent them from being held in the optimal orientation for engaging the receptor. The fifth breakpoint (residues 133 and 132) does not appear to be near any proposed receptor binding site. It resides in the most flexible internal sequence of the native protein and is the single glycosylation site in G-CSF (*37*). Our results agree with a previous report on circularly permuted IL-4, in which the breakpoint at a glycosylation site gave a permuted molecule with the highest affinity for the IL-4 receptor (*38*).

The broad range of proliferation activity resulting from using the L1 linker is intriguing. While cpG-CSF(L1)[133/132] is 10-fold less potent than G-CSF, the other L1 linker cpG-CSFs had more pronounced potency losses (Table 2). The most striking example is cpG-CSF(L1)[142/141]. While the new termini in cpG-CSF(L1)[142/141] are located far away from the L1 linker (ca. 40 Å), and the corresponding molecule with the L0 linker exhibited a potency similar to that of native G-CSF, cpG-CSF(L1)[142/141] nevertheless exhibited a nearly 100-fold decrease in its ability to stimulate proliferation. The observed effect of the L1 linker on bioactivity may reflect the strain exerted on helix 1, which is proposed to contact the receptor, as the glycine-rich L1 linker could destabilize the helix. Indeed, this possibility is supported by the NMR results which provided evidence of unwinding of the first turn in helix 1 in cpG-CSFs.

The location of the breakpoints may add to the destabilizing effects of the L1 linker in a variety of ways. For example, being juxtaposed to the N-terminus of helix 1, the breakpoint at residues 126 and 125 could destabilize helix 1 by weakening the packing between helix 1 and 3. The breakpoint at residues 39 and 38 resides at the C-terminal end of helix 1, resulting in significant chemical shift changes in helix 1 extending through A30 (Figure 4) and thus possibly distorting helix 1 at its C-terminus. Residue F140 packs against a hydrophobic cluster formed by L50 and L54 of the minihelix, I56 following the minihelix in loop 12, L88 and L92 in helix 2, and L152 in helix 4. It is plausible that this packing pattern may become deranged when the new termini are introduced at residues 141 and 142. In the NMR spectrum of cpG-CSF(L1)[142/141], the upfield changes in amide proton chemical shifts for residues 45 and 63 [>9.5 ppm in the G-CSF(S17) spectrum] suggest that the environment of these residues has undergone a significant change. Indeed, the decreased receptor affinity and the 20-fold slower association rate imply that this ligand experiences some difficulty in engaging the receptor. These results suggest that both the linker and breakpoint are important in the design of circularly permuted molecules, and that their impact may be rationalized by the way they interact with each other to impact the structure, stability, and complementarity of the receptor binding surface.

ACKNOWLEDGMENT

We gratefully acknowledge Rodney Combs, Kevin Duffin, Jim DuMontell, Cindy Gottschalk, Sherry Guo, Nancee Kimack, Barbara Klein, Jon Klover, Nathan Pack, Kumnan Paik, Lyle Pegg, Joseph Polazzi, Nick Staten, Scott Vaccaro, Jennifer Wang, Andrea Watson, and Jim Zobel for their technical assistance and Keats Nelms and Joseph Welply for critical reading of the manuscript.

REFERENCES

1. Metcalf, D., and Nicola, N. A. (1995) *The Hemopoietic Colony-Stimulating Factors. From Biology to Clinical Applications*, Cambridge University Press, Cambridge, U.K.
2. Bronchud, M. H., Scarffe, J. H., Thatcher, N., Crowther, D., Souza, L. M., Alton, N. K., Testa, N. G., and Dexter, T. M. (1987) *Br. J. Cancer 56*, 809−813.
3. Gabrilove, J. L., Jakubowski, A., Scher, H., Sternberg, C., Wong, G., Grous, J., Yagoda, A., Fain, K., Moore, M. A., Clarkson, B., Oettgen, H. F., Alton, K., Welte, K., and Souza, L. (1988) *N. Engl. J. Med. 318*, 1414−1422.
4. Morstyn, G., Campbell, L., Souza, L. M., Alton, N. K., Keech, J., Green, M., Sheridan, W., Metcalf, D., and Fox, R. (1988) *Lancet 1*, 667−672.
5. Gabrilove, J. L., Jakubowski, A., Fain, K., Grous, J., Scher, H., Sternberg, C., Yagoda, A., Clarkson, B., Bonilla, M. A., Oettgen, H. F., Alton, K., Boone, T., Testa, N. G., and Dexter, T. M. (1988) *J. Clin. Invest. 82*, 1454−1461.
6. Kuga, T., Komatsu, Y., Yamasaki, M., Sekine, S., Miyaji, H., Nishi, T., Sato, M., Yokoo, Y., and Asano, M. (1989) *Biochem. Biophys. Res. Commun. 159*, 103−111.
7. Layton, J. E., Morstyn, G., Fabri, L. J., Reid, G. E., Burgess, A. W., Simpson, R. J., and Nice, E. C. (1991) *J. Biol. Chem. 266*, 23815−23823.

8. Reidhaar-Olson, J. F., De Souza-Hart, J. A., and Selick, H. E. (1996) *Biochemistry 35*, 9034−9041.
9. Young, D. C., Zhan, H., Cheng, Q.-L., Hou, J., and Matthews, D. J. (1997) *Protein Sci. 6*, 1228−1236.
10. Hill, C. P., Osslund, T. D., and Eisenberg, D. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 5167−5171.
11. Lovejoy, B., Cascio, D., and Eisenberg, D. (1993) *J. Mol. Biol. 234*, 640−653.
12. Werner, J. M., Breeze, A. L., Kara, B., Rosenbrock, G., Boyd, J., Soffe, N., and Campbell, I. D. (1994) *Biochemistry 33*, 7184−7192.
13. Zink, T., Ross, A., Ambrosius, D., Rudolph, R., and Holak, T. A. (1992) *FEBS Lett. 314*, 435−439.
14. Zink, T., Ross, A., Lüers, K., Cieslar, C., Rudolph, R., and Holak, T. A. (1994) *Biochemistry 33*, 8453−8463.
15. McWherter, C. A., Feng, Y., Zurfluh, L. L., Klein, B. K., Baganoff, M. P., Polazzi, J. O., Hood, W. F., Paik, K., Abegg, A. L., Grabbe, E. S., Shieh, J.-J., Donnelly, A. M., and McKearn, J. P. (1999) *Biochemistry 38*, 4564−4571.
16. Goldenberg, D. P., and Creighton, T. E. (1983) *J. Mol. Biol. 165*, 407−413.
17. Luger, K., Hommel, U., Herold, M., Hofsteenge, J., and Kirschner, K. (1989) *Science 243*, 206−210.
18. Zhang, T., Bertelsen, E., Benvegnu, D., and Alber, T. (1993) *Biochemistry 32*, 12311−12318.
19. Buchwalder, A., Szadkowski, H., and Kirschner, K. (1992) *Biochemistry 31*, 1621−1630.
20. Yang, Y. R., and Schachman, H. K. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 11980−11984.
21. Protasova, N. Y., Kireeva, M. L., Murzina, N. V., Murzin, A. G., Uversky, V. N. U., Gryaznova, O. I., and Gudkov, A. T. (1994) *Protein Eng. 7*, 1373−1377.
22. Mullins, L. S., Wesseling, K., Kuo, J. M., Garrett, J. B., and Raushel, F. M. (1994) *J. Am. Chem. Soc. 116*, 5529−5533.
23. Hahn, M., Piotukh, K., Borriss, R., and Heinemann, U. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10417−10421.
24. Kreitman, R. J., Puri, R. K., and Pastan, I. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 6889−6893.
25. Giri, J. G., Abegg, A., Abrams, M., Aykent, S., Caparon, M., Doshi, P., Donnelly, A., Favara, J., Goldberg, S. B., Hood, W., Joy, W., Kahn, L., Keith, R., Monahan, J., Paik, K., Smith, W., Thurman, T., and McKearn, J. P. (1998) *Blood* (submitted for publication).
26. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
27. Yang, J. T., Wu, C.-S. C., and Martinez, H. M. (1986) in *Methods in Enzymology*, Vol. 130, Part K, pp 208−269, Academic Press, New York.
28. Kay, L. E., Keifer, P., and Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10664.
29. Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1994) *J. Biomol. NMR 4*, 845−858.
30. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.
31. Pace, C. N., Shirley, B. A., and Thomson, J. A. (1989) in *Protein Structure* (Creighton, T. E., Ed.) pp 311−330, IRL Press, New York.
32. Rodbard, D., Munson, P. J., and de Lean, A. (1977) *I. A. E. A. Bull.*, 469.
33. Ishikawa, M., Iijima, H., Satake-Ishikawa, R., Tsumura, H., Iwamatsu, A., Kadoya, T., Shimada, Y., Fukamachi, H., Kobayashi, K., et al. (1992) *Cell Struct. Funct. 17*, 61−65.
34. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
35. Neri, D., Wider, G., and Wuethrich, K. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 4397−4401.
36. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546−4559.
37. Oheda, M., Hase, S., Ono, M., and Ikenaka, T. (1988) *J. Biochem. (Tokyo) 103*, 544−546.
38. Kreitman, R. J., Puri, R. K., McPhie, P., and Pastan, I. (1995) *Cytokine 7*, 311−318.

BI982224O