

Reengineering Granulocyte Colony-stimulating Factor for Enhanced Stability*

Received for publication, May 17, 2001, and in revised form, June 13, 2001
Published, JBC Papers in Press, June 13, 2001, DOI 10.1074/jbc.M104494200

Barney Bishop‡, Debbie C. Koay§, Alan C. Sartorelli§, and Lynne Regan‡¶||

From the Departments of ‡Chemistry and the ¶Molecular Biophysics and Biochemistry, Yale University and the §Department of Pharmacology and Developmental Therapeutics Program, Cancer Center, Yale University School of Medicine, New Haven, Connecticut 06520

Granulocyte colony-stimulating factor is a long-chain cytokine that has both biological and therapeutic applications. It is involved in the production and maturation of neutrophilic progenitor cells and neutrophils and is administered to stimulate the production of white blood cells to reduce the risk of serious infection in immunocompromised patients. We have reengineered granulocyte colony-stimulating factor to improve the thermodynamic stability of the protein, focusing on enhancing the α -helical propensity of residues in the antiparallel 4-helix bundle of the protein. These redesigns resulted in proteins with substantially enhanced stability while retaining wild-type levels of biological activity, measured as the ability of the reengineered proteins to stimulate the proliferation of murine myeloid cells transfected with the granulocyte colony-stimulating factor receptor.

Growth factors and cytokines are potential therapeutic agents because of their key roles in regulating the production, maturation, and activity of blood, muscle, and bone cells. A number of cytokines are currently employed in the treatment of blood deficiencies and cancers. For example, erythropoietin is administered to stimulate red blood cell production in patients suffering from anemia.¹ Human growth hormone is used to stimulate growth in children and adults exhibiting extremely slow or subnormal growth due to hormonal deficiencies,² and granulocyte colony-stimulating factor (G-CSF)³ is employed to treat cancer patients undergoing chemotherapy in order to alleviate the depression of white blood cell levels produced by cytotoxic therapeutic agents.⁴

Although the activity and specificity of function make the cytokines potentially powerful therapeutic agents, they are not ideal drugs. Thus, there is keen interest in developing alternative or improved molecules that demonstrate cytokine function but have superior pharmacological properties (4–10). Protein reengineering offers a means of improving the physical and biological properties of these peptides and proteins to enhance

their efficacy as therapeutics.

Here, we describe the results of reengineering the cytokine G-CSF to increase thermodynamic stability while retaining biological activity. G-CSF plays a critical role in the process of hematopoiesis (11), regulating the proliferation, differentiation, and survival of neutrophils and neutrophilic progenitor cells (Fig. 1). As a therapeutic agent, G-CSF is used to reduce the risks of serious infection in immunocompromised individuals undergoing chemotherapy and thus facilitates recovery. G-CSF cannot be administered orally, and frequent (daily) injections of significant quantities of the cytokine (5–10 $\mu\text{g}/\text{kg}$) are necessary throughout the course of treatment (3). In addition, G-CSF requires stringent formulation and storage conditions such as constant refrigeration; supplies must be discarded if left at room temperature.

G-CSF is a member of the long chain family of cytokines, characterized by an antiparallel 4-helix bundle fold, with long overhand connecting loops (Fig. 2) (14–16). The redesign of G-CSF was facilitated by studies that identified by alanine-scanning mutagenesis the primary binding determinants of the protein for its receptor (17, 18). Key residues that must be conserved in any reengineering study lie on the surfaces of helices A and C (Fig. 2), with additional potentially important contacts in the loop regions. The recently reported crystal structure of G-CSF complexed with the cytokine receptor homologous region of the G-CSF receptor reinforces the importance of these helices and suggests additional contacts by residues in the N-terminal region of G-CSF, which is unstructured in the unbound protein (19).

In reengineering cytokines such as G-CSF, it is far less likely that receptor binding will be disrupted or that mutants will raise an immune response, if the changes to protein sequence and topology are kept to a minimum. Two aspects of G-CSF appear to lend themselves most readily to reengineering; the unusual abundance of glycine residues in the helices and the long overhand loops connecting the helices in the bundle.

The focus of the studies reported here is the helical glycine residues. An abundance of glycine residues in the helical regions of a soluble protein is unusual. It has one of the lowest intrinsic α -helical propensities of all the amino acids (second only to proline). This is manifested in both statistical and thermodynamic studies (20–23). In thermodynamic studies, alanine stabilizes a peptide or protein helix by up to 2 kcal/mol relative to glycine (20). Therefore, glycine to alanine substitutions in the helical regions of proteins would be expected to enhance protein stability. This premise has been investigated in a number of systems (24–29).

We have therefore systematically replaced the helical glycine residues of G-CSF individually and in combination and evaluated both the stability and the biological activity of each mutant. After successive reengineering cycles, we produced three

* This research was funded in part by United States Public Health Service Grants GM49146 and CA02817. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed. Tel.: 203-432-9841; Fax: 203-432-5175; E-mail: lynne@csb.yale.edu.

¹ wwwext.amgen.com/product/epogenHome.html (2001).

² www.genentech.com/products/nutropin/(2001).

³ The abbreviations used are: G-CSF, granulocyte colony-stimulating factor; GdnHCl, guanidine hydrochloride; wt, wild-type; G-CSFR, granulocyte colony-stimulating factor receptor; TEV-p, tobacco-etch-virus protease.

⁴ www.neupogen.com/(2001).

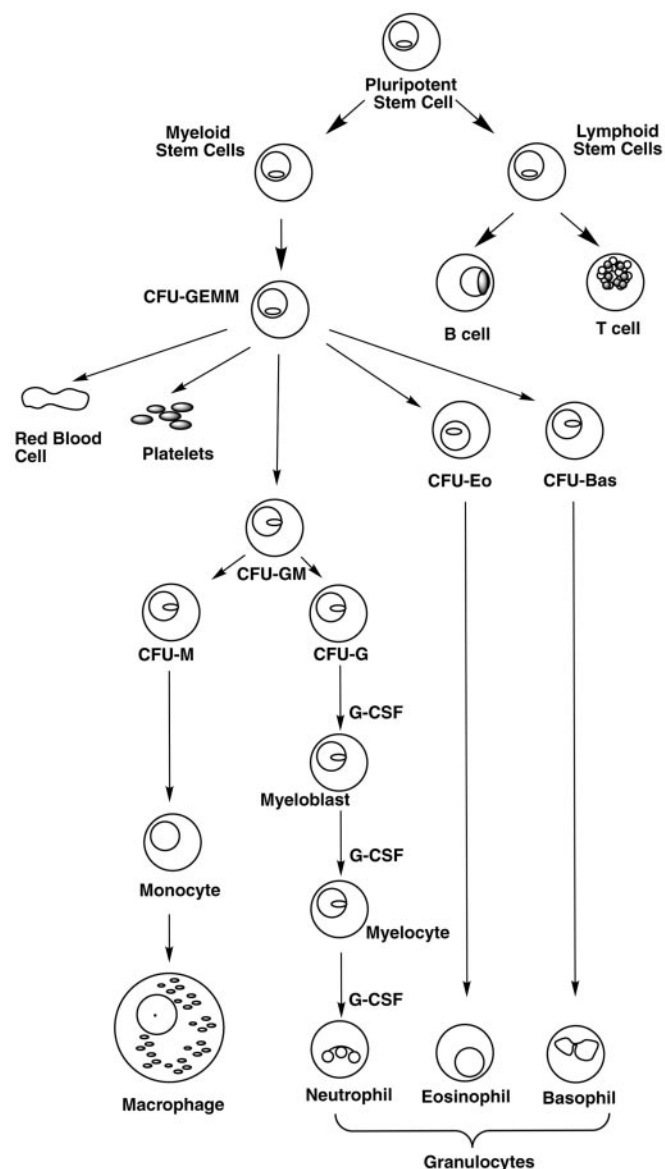


FIG. 1. **Schematic illustration of hematopoiesis.** The sites of action of G-CSF in this complex cascade are indicated. Elements of the diagram were drawn from Ref. 12.

mutants that demonstrated significantly improved stabilities relative to G-CSF and retained the ability to induce cellular proliferation comparable with that of wild-type G-CSF.

EXPERIMENTAL PROCEDURES

Strains and Plasmids

Escherichia coli strain DH10B (30) was used for both cloning and expression purposes. The plasmid pBAD-G4 (17) with an insert for the gene containing the G-CSF wild-type construct (with a C17S substitution to reduce aggregation and the DNA sequence engineered to contain 13 unique restriction sites without changing the amino acid sequence) was derived from the vector pBAD18, Ref. 31; provided by Dr. John Reidhaar-Olson, Affymax Corporation. This vector provides ampicillin resistance and protein expression is under control of the inducible promoter araBAD.

Cloning and Mutagenesis

To aid in purification, a hexahistidine metal affinity tag and a 14-amino acid spacer were incorporated at the N terminus of G-CSF by inserting a DNA cassette defined by the *Nde*I and *Csp*I restriction sites in the original construct. A sequence coding for the tobacco-etch-virus protease (rTEV-p) cleavage site was included in the spacer sequence N-terminal to the threonine (numbered T1) of wt-G-CSF (32). Proteolysis by rTEV-p generates an N-terminal glycine residue (numbered

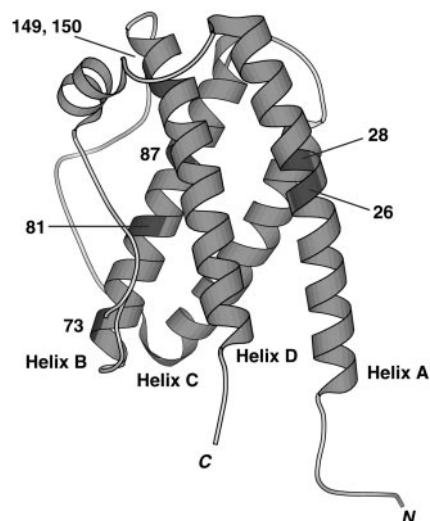


FIG. 2. **Ribbon representation of the structure of G-CSF made in MOLSCRIPT (13) from the NMR structure (14).** A 4-helix bundle motif with long overhand loop connections is shown. The helices are labeled in order from the N terminus: A, B, C, D. The locations of the helical glycine residues are shaded in dark gray and their residue numbers indicated.

G-1) preceding this threonine residue and removes the His₆ tag and linker sequence.

Cassettes encoding the desired mutations were prepared via polymerase chain reaction using synthetic primers incorporating the prescribed change in sequence. The new sequences were then incorporated into the plasmid by cassette mutagenesis utilizing the appropriate restriction sites. Competent DH10B cells were transformed with DNA following ligation and screened for ampicillin resistance. The plasmid DNA of each clone was then sequenced to confirm its identity (HHMI Biopolymer/Keck Foundation Biotechnology Resource Laboratory, Yale University).

Growth and Purification

Starter cultures in 75 ml of Luria broth with 100 μ g/ml ampicillin were prepared from DH10B cells transformed with the appropriate plasmid and allowed to grow at 37 °C for 4–5 h with agitation (33). Superbroth (1 liter) was then inoculated with 15 ml of the starter culture and allowed to grow at 37 °C with agitation (33). Protein expression was induced by the addition of arabinose to a final concentration of 0.2% once the cell culture had reached an A_{600} of 1–2 absorbance units. Cells were harvested by centrifugation after ~12 h of further growth, and the cell pellets were stored frozen.

The G-CSF variants were purified using a two-step protocol that we devised. This process provided a generic method that could be applied to the purification of G-CSF and all of the mutants. The frozen cell pellets were suspended in 30 ml of lysis buffer (7 M GdnHCl, 10 mM Tris-HCl, pH 8, 100 mM NaCl) per liter of original growth medium. The resuspended cells were lysed by short bursts of sonication. Solubilized protein was then separated from the solids by centrifugation at $13,560 \times g$ for 30 min. The supernatant was collected and loaded onto 2 ml TALON resin (CLONTECH) per liter of original growth medium equilibrated in lysis buffer. The resin was washed with ~10-column volumes of lysis buffer and then washed twice with 2-column volumes of renaturing buffer (10 mM Tris-HCl, pH 8, 100 mM NaCl). The resin was then re-suspended in 5 ml of renaturing buffer and agitated for 30 min before drawing off the aqueous solution. The loaded resin thus washed was suspended in 5 ml of 10 mM Tris-HCl, pH 8, 100 mM NaCl, 500 mM imidazole and agitated at room temperature for 10 min to release the protein from the resin. The eluent, containing the His-tagged G-CSF, was then loaded onto a Superdex G75 column (Amersham Pharmacia Biotech) and eluted with renaturing buffer. By this method, it was possible to isolate ~0.5–2 mg of refolded pure G-CSF per liter of growth medium.

Circular Dichroism Wavelength Scans

Approximately 10 μ M protein in 100 mM sodium phosphate, pH 7, and 200 mM NaCl in a 1-mm path length cell was used to collect spectra from 200–260 nm at 25 °C. Data were collected using an Aviv 62DS Circular Dichroism Spectrometer (Aviv Instruments).

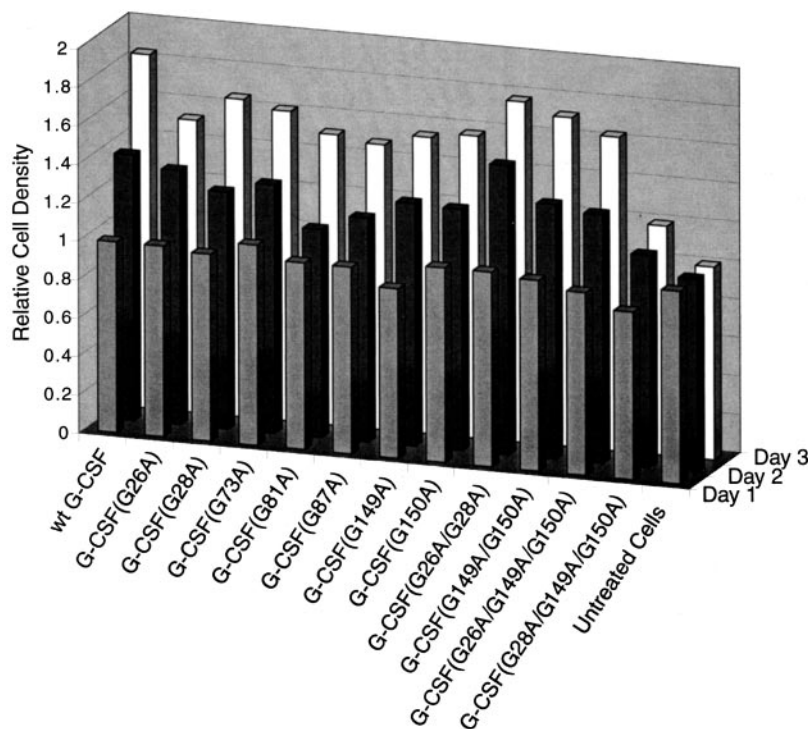


FIG. 3. **Biological activity of G-CSF and the designed mutants.** Here, the G-CSF variants are aligned along the *x* axis, cell density relative to cells not treated with G-CSF is reported on the *y* axis, and the *z* axis indicates the sampling interval.

Chemical Denaturation

Chemical denaturation of wild-type and mutant proteins was followed by monitoring the ellipticity at 222 nm as a function of GdnHCl concentration. Two stock solutions, one renaturing and the other denaturing, were prepared for G-CSF and each mutant. The native solution consisted of $\sim 10 \mu\text{M}$ protein in 100 mM sodium phosphate (pH 7) and 200 mM NaCl. The denaturing solution consisted of $\sim 10 \mu\text{M}$ protein in 100 mM sodium phosphate (pH 7), 200 mM NaCl with 4.5–6 M GdnHCl. The exact concentration of GdnHCl in the denaturing solution was varied for different proteins, reflecting different stabilities. Each point on the denaturation titration was prepared by the serial exchange of equal volumes of buffer between the native and denaturing solutions, followed by thorough mixing. The signal at 222 nm was collected and averaged over a 1-min period following equilibration at 25 °C for 10 min. Guanidine concentrations of the initial and final samples were established within approximately $\pm 1\%$ by refractometry using an Atago R5000 refractometer (Atago, Japan), allowing the calculation of the intermediate concentrations (34). Because the GdnHCl-induced unfolding transition is reversible, this titration method afforded reliable data consuming a minimal amount of protein in the process.

Evaluation of Protein Stability

Determination of C_m —The fraction of folded protein as a function of the concentration of GdnHCl was determined for G-CSF and each mutant protein. The pre- and post-transition regions of the titration curves were fit to Equations 1 and 2, respectively. These describe the signal attributable to the folded and denatured states (Θ_{nat} and Θ_{den} , respectively) as functions of denaturant concentration.

$$\Theta_{\text{nat}} = \Theta_{\text{nat}}^{\circ} + m_{\text{N}}[\text{GdnHCl}] \quad (\text{Eq. 1})$$

$$\Theta_{\text{den}} = \Theta_{\text{den}}^{\circ} + m_{\text{D}}[\text{GdnHCl}] \quad (\text{Eq. 2})$$

In Equations 2 and 3, $\Theta_{\text{nat}}^{\circ}$ and $\Theta_{\text{den}}^{\circ}$ are the ellipticities of the native and denatured states in the absence of denaturant and m_{N} and m_{D} , their respective dependence on GdnHCl concentration. Equation 3 corrects for these baselines and provides the fraction of total protein that is folded.

$$\text{Fraction folded} = (\Theta_{\text{obs}} - \Theta_{\text{nat}}^{\circ}) / (\Theta_{\text{den}}^{\circ} - \Theta_{\text{nat}}^{\circ}) \quad (\text{Eq. 3})$$

Here, Θ_{obs} corresponds to the observed ellipticity. The fraction of folded protein was plotted as a function of GdnHCl concentration (Fig. 3), and the concentration of GdnHCl corresponding to the midpoint of the transition (C_m) was determined. Thus, C_m provides a measure of stability.

Determination of the ΔG of Folding

The free energy (ΔG) of folding was determined using the method of Santoro and Bolen (35). The relationship between the fraction folded and the folding equilibrium (K_{obs}) allows the determination of the ΔG of folding by fitting the chemical denaturation curves of G-CSF and the designed mutants to Equation 4.

$$\Theta_{\text{obs}} = [(\Theta_{\text{nat}}^{\circ} + m_{\text{N}}[\text{GdnHCl}]) + (\Theta_{\text{den}}^{\circ} + m_{\text{D}}[\text{GdnHCl}]) \times \exp(-(\Delta G_{\text{N-D}}/RT + m[\text{GdnHCl}]/RT)) / (1 + \exp(-(\Delta G_{\text{N-D}}/RT + m[\text{GdnHCl}]/RT))] \quad (\text{Eq. 4})$$

Here $\Delta G_{\text{N-D}}$ is defined as the free energy of unfolding in the absence of denaturant; m , the dependence of $\Delta G_{\text{N-D}}$ on the concentration of GdnHCl; R , the universal gas constant; and T , the temperature at which denaturation was performed. In this fashion, the $\Delta G_{\text{N-D}}$ for the wild-type protein and the mutants were determined. The free energy of folding, ΔG ($\Delta G = -\Delta G_{\text{N-D}}$), the corresponding m -value for folding and the degree of stabilization/destabilization relative to wild-type, $\Delta\Delta G$, reported in Table I. Thus, the more negative $\Delta\Delta G$ values correspond to increases in stability over wild-type G-CSF.

Biological Activity

Murine myeloid LGM-1 cells (36) (provided by Dr. Tasuku Honjo, Kyoto University, Japan), and transfectants derived therefrom in our laboratory, were maintained in suspension culture in RPMI 1640 medium (Sigma) supplemented with 10% fetal bovine serum (Life Technologies, Inc.) and 45 units/ml of recombinant murine IL-3 (rmIL-3) (Endogen) at 37 °C in a 95% air, 5% CO₂ humidified incubator. Measurement of cell proliferation in response to G-CSF employed exponentially growing LGM-1 cells transfected with the full-length G-CSFR (37, 38). To measure the growth of the transfected LGM-1 cells in response to wild-type or various G-CSF mutants, cells were washed twice with cold phosphate-buffered saline, and all of the transfected cells were seeded into fresh culture medium at a density of 5×10^4 cells/ml in the presence of 45 units/ml of rmIL-3, 10 pg/ml of wild-type G-CSF, or the various G-CSF mutants. Cell numbers were determined daily for 3 days after exposure to rmIL-3, wild-type G-CSF, or various mutant G-CSFs using a Coulter model ZM particle counter (Coulter Electronics) connected to a Coulter model 256 Channelyzer (Coulter Electronics).

RESULTS

Wild-type G-CSF contains 14 glycine residues, 7 of which lie in central positions within the helices of the 4-helix bundle

(Fig. 2). Two of these glycine residues are located in helix A (Gly-26 and Gly-28), three in helix B (Gly-73, Gly-81, and Gly-87), and two in helix D (Gly-149 and Gly-150). To thoroughly explore alanine replacement as a means of improving stability, we chose to consider all of the helical glycine residues as target replacement sites. Each helical glycine residue was individually replaced by alanine to generate 7 point mutants, identified by the position of the replaced glycine residue; G-CSF(G26A), G-CSF(G28A), G-CSF(G73A), G-CSF(G81A), G-CSF(G87A), G-CSF(G149A), and G-CSF(G150A). G-CSF and the designed mutants were each overexpressed and purified.

The biological effects of G-CSF are mediated through its interaction with the G-CSF receptor (G-CSFR), a member of the cytokine receptor superfamily. Upon binding of G-CSF to its receptor, the G-CSFR dimerizes and signaling events are initiated from the cytoplasmic domain. To evaluate the biological activity of G-CSF and the reengineered mutants, we have introduced the G-CSF receptor into IL-3-dependent myeloid LGM-1 cells (LGM GR95 cells). LGM-1 cells do not naturally express endogenous G-CSFR (39). Exogenous expression of the full-length G-CSFR makes LGM-1 cells responsive to G-CSF resulting in cellular proliferation upon exposure to G-CSF in the absence of IL-3. Cell proliferation assays using these cells

thus provided a measure of the biological activity of the designed mutants. We stimulated the LGM-1 cells expressing G-CSFR with wild-type G-CSF and the designed mutants (10 pg/ml in each case). Treatment with less than 10 pg/ml wild-type G-CSF results in no discernable proliferation of LGM GR95 cells. The LGM GR95 cells treated with the G-CSF mutants exhibited increases in cellular numbers over a 3-day period in a manner analogous to those treated with the wild-type G-CSF (Fig. 3). Therefore, we conclude that the point mutants retained the ability to induce proliferation comparable with the wild-type protein (Fig. 3).

Chemical denaturation was used to evaluate the stabilities of G-CSF and the designed mutants. G-CSF and the variants undergo reversible denaturation transitions induced by GdnHCl, and it is therefore possible to use this method to derive thermodynamic parameters and assess stability (Fig. 4). By this method, we determined the ΔG and C_m of unfolding (Table I). The thermal denaturation of G-CSF is irreversible at pH 7 (40). Thus, it is not an appropriate method to use to establish the thermodynamic parameters for G-CSF in a context that is relevant to stability at 25–37 °C.

In ranking the mutants, it is necessary to consider both C_m and ΔG . They provide complementary measures of stability relative to the wild-type protein ($\Delta\Delta G$ and ΔC_m) and both can be derived from the same denaturation data (Table I). The ΔG of folding was determined by fitting the denaturation curves to Equation 4. The m value (corresponding to the slope of the transition) reflects the dependence of ΔG on the concentration of GdnHCl is also provided by these fits. The majority of the mutants have m values within $\pm 15\%$ of the mean value. This similarity in the m values suggests that the slopes of the transitions and therefore cooperativity are similar for the mutants. Therefore, the C_m values, which are determined by plotting the fraction of folded protein as a function of GdnHCl concentration, are also indicative of relative stability.

The point mutants G-CSF(G26A), G-CSF(G28A), G-CSF(G149A), and G-CSF(G150A) resulted in the greatest increases in protein stability ($\Delta\Delta G$ -2 to $-3 \pm \sim 0.7$ kcal and ΔC_m 0.1–0.6 M); the mutation at position 73 resulted in a more modest increase in stability ($\Delta\Delta G$ -1.44 ± 0.45 kcal and ΔC_m 0.1 M); and the mutations at 81 and 87 were destabilizing ($\Delta\Delta G$ 1.68 ± 0.51 and 1.86 ± 0.38 kcal, respectively, and ΔC_m values nearly identical to wild-type G-CSF). The mutations at positions 26, 28, 149, and 150 clearly provided the best basis for the following generations of mutants, because they resulted in the most substantial gains in stability.

We combined the most stabilizing point mutations that were in close proximity to each other to create the second generation

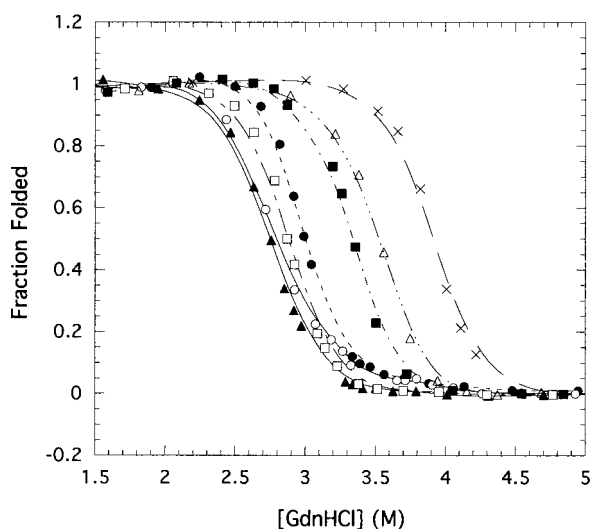


FIG. 4. Fraction of total protein folded (monitored by CD) as a function of denaturant concentration for wild-type G-CSF (\blacktriangle) and the mutants G-CSF(G26A) (\bullet), G-CSF(G81A) (\circ), G-CSF(G149A) (\square), G-CSF(G149A/G150A) (\blacksquare), G-CSF(G26A/G149A/G150A) (\triangle) and G-CSF(G28A/G149A/G150A) (\times).

TABLE I
Comparison of the stabilities of wild-type G-CSF and designed mutants

Protein	C_m^a	ΔC_m^b	m-value ^c	ΔG^c	$\Delta\Delta G^d$
	M GdnHCl	M GdnHCl			
Wt-G-CSF	2.75		3.43 ± 0.09	-9.44 ± 0.24	
G-CSF(G26)	2.97	0.22	4.21 ± 0.20	-12.50 ± 0.58	-3.06 ± 0.63
G-CSF(G28)	3.30	0.55	3.64 ± 0.29	-12.10 ± 0.75	-2.66 ± 0.79
G-CSF(G73)	2.84	0.09	3.82 ± 0.13	-10.88 ± 0.38	-1.44 ± 0.45
G-CSF(G81A)	2.79	0.04	2.80 ± 0.19	-7.76 ± 0.45	1.68 ± 0.51
G-CSF(G87A)	2.73	-0.02	3.15 ± 0.12	-7.58 ± 0.30	1.86 ± 0.38
G-CSF(G149A)	2.88	0.13	4.07 ± 0.13	-11.73 ± 0.38	-2.29 ± 0.45
G-CSF(G150A)	3.03	0.28	4.07 ± 0.25	-12.30 ± 0.75	-2.86 ± 0.79
G-CSF(G26A/G28A)	3.65	0.90	3.34 ± 0.27	-12.18 ± 0.96	-2.74 ± 0.99
G-CSF(G149A/G150A)	3.34	0.59	4.17 ± 0.41	-13.64 ± 1.10	-4.20 ± 1.13
G-CSF(G26A/G149A/G150A)	3.50	0.75	3.60 ± 0.19	-12.90 ± 0.48	-3.46 ± 0.54
G-CSF(G28A/G149A/G150A)	3.91	1.16	4.00 ± 0.53	-15.60 ± 2.00	-6.16 ± 2.01

^a C_m was determined from the fraction of folded protein as a function of GdnHCl concentration with an estimated error of ± 0.03 M.

^b $\Delta C_m = C_{m\text{-mutant}} - C_{m\text{-G-CSF}}$.

^c The values and errors reported for m and ΔG were derived by fitting titration data to the Santoro and Bolen equation (Equation 4) (35).

^d $\Delta\Delta G = \Delta G_{\text{mutant}} - \Delta G_{\text{G-CSF}}$.

of mutants, G-CSF(G26A/G28A) and G-CSF(G149A/G150A). Both G-CSF(G26A/G28A) and G-CSF(G149A/G150A) retained the ability to induce proliferation in myeloid LGM-1 cells transfected with the G-CSFR. The double mutant G-CSF(G26A/G28A) exhibited an increase in stability ($\Delta\Delta G -2.74 \pm 0.99$ kcal and $\Delta C_m 0.9$ M) relative to the single individual point mutants G-CSF(G26A) and G-CSF(G28A). G-CSF(G149A/G150A) realized a greater cumulative improvement in relative stability ($\Delta\Delta G -4.20 \pm 1.13$ kcal and $\Delta C_m 0.6$ M).

The third generation of mutants incorporated the stabilizing double substitution G149A and G150A in combination with either G26A or G28A to generate the pair of triple mutants G-CSF(G26A/G149A/G150A) and G-CSF(G28A/G149A/G150A), both of which retained biological activity comparable with G-CSF. By combining the G149A/G150A double mutation and a third stabilizing single mutation, the triple mutants were expected to be more stable than wild-type protein. Chemical denaturation indicated that G-CSF(G26A/G149A/G150A) and G-CSF(G28A/G149A/G150A) demonstrated significantly elevated stability ($\Delta\Delta G -3.46 \pm 0.54$ kcal, $\Delta C_m 0.8$ M and $\Delta\Delta G -6.2 \pm 2.01$ kcal, $\Delta C_m 1.2$ M, respectively). However, the stability gains attained in G-CSF(G26A/G149A/G150A) were comparable with those of the double mutants G-CSF(G149A/G150A) and G-CSF(G26A/G28A). Therefore, incorporation of the substitution at position 26 did not appear to improve upon G-CSF(G149A/G150A). This was not the case for G-CSF(G28A/G149A/G150A), which was notably more stable than G-CSF(G149A/G150A).

DISCUSSION

The objectives of this work were 2-fold. The first was to reengineer a protein of therapeutic interest to enhance its thermodynamic stability. The second was to achieve this goal through minimal changes in protein sequence. The biological function of G-CSF is linked inextricably with the protein fold. Thus, it is critical that alterations in the protein sequence not result in a large perturbation in structure, which would compromise the potential therapeutic utility of the mutant. Therefore, it was essential to assess the ability of each mutant protein to promote cell proliferation, as well as their respective stabilities.

Currently, not all of the factors that contribute to the pharmacokinetic properties of proteins are known, but stability likely contributes significantly, and it represents a tractable target for protein engineering. Some of the basic principles regarding the relationship between protein sequence, structure, and stability have been established (21–23, 26, 41–50, 1–3). Increased stability could translate into enhanced resistance to enzymatic proteolysis or reduce the requirement of refrigeration for storage by improving protein stability at ambient temperatures.

Replacement of helical glycine residues with alanine, an amino acid with a much higher α -helical propensity, potentially provides a relatively simple means of increasing protein stability. The stabilization associated with such glycine to alanine substitutions can be attributed, in part, to the unfavorable entropic cost of constraining the glycine backbone into a helical conformation, relative to alanine. The net effect on protein stability of any amino acid change represents a balance between energetically favorable contributions, for example intrinsic α -helical propensity, and energetically unfavorable interactions, such as steric conflicts. The small increase in sidechain bulk involved in substituting the methyl group of alanine for the hydrogen of glycine helps to minimize the potential for introducing steric conflicts.

We sought to improve the helical propensity of residues within the 4-helix bundle region of G-CSF by systematically replacing the 7 helical glycine residues with alanines. By generating and evaluating single point mutants, we were able to es-

tablish the effects of each change and then group them based on whether they were stabilizing or destabilizing relative to wild-type G-CSF. Alanine replacements at positions 73, 81, and 87 resulted in either a smaller increase or even a decrease in stability and for this reason were not utilized in later redesigns. The smaller increase in stability associated with G-CSF(G73A) can be explained in part by its localization at the N terminus of helix-B, adjacent to the disulfide bridging Cys-64 and Cys-74. The destabilizing effects of Gly/Ala replacements at positions 81 and 87 may be attributable to the close proximity of the AB and CD loops along much of helix-B. They could partially occlude residues on the surface of helix-B, and thus the methyl sidechain of alanine may introduce steric clashes. Additionally, the helix unwinds slightly in the turn containing position 81 allowing Leu-82 to pack into the core of the 4-helix bundle. Alanine replacement could interfere with this unwinding. Replacing glycine residues with alanine at positions 26, 28, 149, and 150 were found to result in the greatest increases in stability, and therefore, were used as the basis for future generations of mutants. Presumably this increased stabilization results from fewer deleterious interactions associated with the substituted alanine residue.

Two successive reengineering cycles, combining favorable mutations, resulted in four versions of G-CSF with significant stability enhancements over the wild-type protein, G-CSF(G26A/G28A), G-CSF(G149A/G150A), G-CSF(G26A/G149A/G150A) and G-CSF(G28A/G149A/G150A). These mutants provide an opportunity to explore the correlation between the factors contributing to thermodynamic stability and those influencing pharmacokinetic properties. They also allow us to begin considering how protein stability may directly affect receptor-binding kinetics and thus biological activity.

Variants of G-CSF and other cytokines with substantially enhanced stabilities have additional practical application potential. Many cell lines that are used in the laboratory and maintained in tissue culture are dependent on exogenously added cytokines for their survival and proliferation. The increased stability of the reengineered G-CSF variants described here could translate into extended effective lifetimes in tissue culture, thus reducing the amount and frequency of the cytokine supplement that is required. In addition, reengineered variants of G-CSF may also have potential therapeutic applications with immunosuppressed patients, where G-CSF is used to expedite the recovery of neutrophil counts that have been depleted as a result of chemotherapy.

The dramatic improvements in the stability of G-CSF arising from double and triple Gly/Ala substitutions demonstrate the importance of intrinsic α -helix propensities to overall protein stability. These gains in stability were achieved without compromising the biological activity of G-CSF. Similar approaches can be used to rationally manipulate the physicochemical properties of other long chain cytokines, to optimize their preparation, efficacy, and storage properties.

Acknowledgments—We thank Dr. John Reidhaar-Olson (Affymax Corporation) for fruitful conversations and for generously providing the pBAD-G4 plasmid. We are grateful to Dr. Tasuku Honjo (Kyoto University, Japan) for supplying the murine myeloid LGM-1 cells. We would also like to thank members of the Regan laboratory for invaluable contributions in preparing this paper.

REFERENCES

1. Predki, P. F., Agrawal, V., Brunger, A. T., and Regan, L. (1996) *Nat. Struct. Biol.* **3**, 54–58
2. Nagi, A. D., and Regan, L. (1997) *Fold Des.* **2**, 67–75
3. Baldwin, E., Xu, J., Hajiseyedjavadi, O., Baase, W. A., and Matthews, B. W. (1996) *J. Mol. Biol.* **259**, 542–559
4. Chaiken, I. M., and Williams, W. V. (1996) *Trends BioTechnol.* **14**, 369–375
5. Lauffenburger, D. A., Fallon, E. M., and Haugh, J. M. (1998) *Chem. Biol.* **5**, R257–R263
6. Quan, C., Skelton, N. J., Clark, K., Jackson, D. Y., Renz, M. E., Chiu, H. H., Keating, S. M., Beresini, M. H., Fong, S., and Artis, D. R. (1998) *Biopolymer*

- (*Pept. Sci.*) **47**, 265–275
7. Domingues, H., Cregut, D., Sebald, W., Oschkinat, H., and Serrano, L. (1999) *Nat. Struct. Biol.* **6**, 652–656
 8. Ballinger, M. D., Shyamala, V., Forrest, L. D., Deuter-Reinhard, M., Doyle, L. V., Wang, J.-x., Panganiban-Lustan, L., Stratton, J. R., Apell, G., Winter, J. A., Doyle, M. V., Rosenberg, S., and Kavanaugh, W. M. (1999) *Nat. Biotechnol.* **17**, 1199–1204
 9. Livnah, O., Stura, E. A., Johnson, D. L., Middleton, S. A., Mulcahy, L. S., Wrighton, N. C., Dower, W. J., Jolliffe, L. K., and Wilson, I. A. (1996) *Science* **273**, 464–470
 10. Livnah, O., Johnson, D. L., Stura, E. A., Farrell, F. X., Barbone, F. P., You, Y., Liu, K. D., Goldsmith, M. A., He, W., Krause, C. D., Pestka, S., Jolliffe, L. K., and Wilson, I. A. (1998) *Nat. Struct. Biol.* **5**, 993–1003
 11. Demetri, G. D., and Griffin, J. D. (1991) *Blood* **78**, 2791–2808
 12. Babior, B. M., and Stossel, T. P. (1994) *Hematology: A Pathophysiological Approach* 3rd Ed, pp. 279–281, Churchill Livingstone, New York
 13. Kraulis, P. M. (1991) *J. Appl. Crystallogr.* **24**, 946–950
 14. Zink, T., Ross, A., Luers, K., Cieslar, C., Rudolph, R., and Holak, T. A. (1994) *Biochemistry* **33**, 8453–8463
 15. Mott, H. R., and Campbell, I. D. (1995) *Curr. Opin. Struct. Biol.* **5**, 114–121
 16. Wells, J. A., and de Vos, A. M. (1996) *Annu. Rev. Biochem.* **65**, 609–634
 17. Reidhaar-Olson, J. F., De Souza-Hart, J. A., and Selick, H. E. (1996) *Biochemistry* **35**, 9034–9041
 18. Young, D. C., Zhan, H., Cheng, Q. L., Hou, J., and Matthews, D. J. (1997) *Protein Sci.* **6**, 1228–1236
 19. Aritomi, M., Kunishima, N., Okamoto, T., Kuroki, R., Ota, Y., and Morikawa, K. (1999) *Nature* **401**, 713–717
 20. Chakrabartty, A., and Baldwin, R. L. (1995) *Adv. Protein Chem.* **46**, 141–176
 21. Chakrabartty, A., Schellman, J. A., and Baldwin, R. L. (1991) *Nature* **351**, 586–588
 22. Munoz, V., and Serrano, L. (1994) *Proteins* **20**, 301–311
 23. Pace, C. N., and Scholtz, J. M. (1998) *Biophys. J.* **75**, 422–427
 24. Hecht, M. H., Sturtevant, J. M., and Sauer, R. T. (1986) *Proteins* **1**, 43–46
 25. Ganter, C., and Pluckthun, A. (1990) *Biochemistry* **29**, 9395–9402
 26. Marqusee, S., and Regan, L. (1991) *Curr. Biol.* **1**, 207–209
 27. Margarit, I., Campagnoli, S., Frigerio, F., Grandi, G., De Filippis, V., and Fontana, A. (1992) *Protein Eng.* **5**, 543–550
 28. Blaber, M., Baase, W. A., Gassner, N., and Matthews, B. W. (1995) *J. Mol. Biol.* **246**, 317–330
 29. Predki, P. F., Nayak, L. M., Gottlieb, M. B., and Regan, L. (1995) *Cell* **80**, 41–50
 30. Grant, S. G., Jessee, J., Bloom, F. R., and Hanahan, D. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 4645–4649
 31. Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995) *J. Bacteriol.* **177**, 4121–4130
 32. Lucast, L. J., Batey, R. T., and Doudna, J. A. (2001) *BioTechniques* **30**, 544–554
 33. Ausubel, F. M. (1989) *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*, pp. 3–4, Greene Pub. Associates and Wiley-Interscience, New York
 34. Kita, Y., Arakawa, T., Lin, T. Y., and Timasheff, S. N. (1994) *Biochemistry* **33**, 15178–15189
 35. Santoro, M. M., and Bolen, D. W. (1988) *Biochemistry* **27**, 8063–8068
 36. Lee, K. H., Kinashi, T., Tohyama, K., Tashiro, K., Funato, N., Hama, K., and Honjo, T. (1991) *J. Exp. Med.* **173**, 1257–1266
 37. Koay, D. C., and Sartorelli, A. C. (1999) *Blood* **93**, 3774–3784
 38. Li, J., Koay, D. C., Xiao, H., and Sartorelli, A. C. (1993) *J. Cell Biol.* **120**, 1481–1489
 39. Dong, F., van Buitenen, C., Pouwels, K., Hoefsloot, L. H., Lowenberg, B., and Touw, I. P. (1993) *Mol. Cell. Biol.* **13**, 7774–7781
 40. Kolvenbach, C. G., Narhi, L. O., Philo, J. S., Li, T., Zhang, M., and Arakawa, T. (1997) *J. Peptide Res.* **50**, 310–318
 41. Rohl, C. A., Fiori, W., and Baldwin, R. L. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 3682–3687
 42. Munson, M., O'Brien, R., Sturtevant, J. M., and Regan, L. (1994) *Protein Sci.* **3**, 2015–2022
 43. Munson, M., Balasubramanian, S., Fleming, K. G., Nagi, A. D., O'Brien, R., Sturtevant, J. M., and Regan, L. (1996) *Protein Sci.* **5**, 1584–1593
 44. Lim, W. A., and Sauer, R. T. (1989) *Nature* **339**, 31–36
 45. Lim, W. A., and Sauer, R. T. (1991) *J. Mol. Biol.* **219**, 359–376
 46. Lim, W. A., Farruggio, D. C., and Sauer, R. T. (1992) *Biochemistry* **31**, 4324–4333
 47. Lim, W. A., Hodel, A., Sauer, R. T., and Richards, F. M. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 423–427
 48. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) *Science* **262**, 1401–1407
 49. Harbury, P. B., Tidor, B., and Kim, P. S. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 8408–8412
 50. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T., and Kim, P. S. (1998) *Science* **282**, 1462–1467