Identification of Residues Critical to the Activity of Human Granulocyte Colony-Stimulating Factor

John F. Reidhaar-Olson,* Janet A. De Souza-Hart, and Harold E. Selick

Affymax Research Institute, 3410 Central Expressway, Santa Clara, California 95051

Received November 13, 1995; Revised Manuscript Received April 18, 1996[⊗]

ABSTRACT: Alanine scanning mutagenesis of human granulocyte colony-stimulating factor (G-CSF) was used to identify residues critical for the cell-proliferative activity of the protein. Fifty-eight residues, most of them on the protein surface, were independently mutated to alanine. Most of the variants retained full biological activity; however, 15 mutants were significantly impaired in their ability to stimulate bone marrow cell proliferation *in vitro*. Four of these variants contain mutations at buried residues and two have substitutions at side chains involved in intramolecular hydrogen bonds. The remaining nine down mutations identify two regions on the surface of the molecule important for biological activity. Consistent with these observations, measurements of binding to NFS-60 cells indicate that the residues most important for receptor binding are Lys40 and Phe144 in site 1 and Glu19 in site 2. In addition to these residues, Val48 and Leu49 in site 1 and Leu15, Asp112, and Leu124 in site 2 are also important for biological activity. These results suggest the presence of two binding sites on the cytokine surface required for dimerization of the G-CSF receptor.

Like other cytokines and growth factors, granulocyte colony-stimulating factor exerts its effects by binding to a specific receptor displayed on the surface of target cells (Nagata & Fukunaga, 1991). The interaction of G-CSF[1] with its receptor induces intracellular events that promote cell proliferation and differentiation, leading eventually to the production of mature neutrophils from hematopoietic precursors (Nicola et al., 1985). Because of its ability to replenish neutrophils *in vivo*, G-CSF has become an important therapeutic product of the biotechnology industry (Hammond et al., 1989; Tkatch & Tweardy, 1993). Although the precise mechanisms of receptor activation and signal transduction are unclear, there is evidence that receptor dimerization (Fuh et al., 1992) and interaction with members of the Jak family of protein tyrosine kinases (Nicholson et al., 1994) are involved.

The crystal structure of G-CSF (Hill et al., 1993) indicates that it is a member of the four-helix-bundle structural superfamily of growth factors (Bazan, 1990). The protein has four long helices arranged in the up−up−down−down motif characteristic of members of this superfamily: helix A (residues 11−39), helix B (71−91), helix C (100−123), and helix D (143−172). In addition, there is a short helical segment, helix E (44−53), between helices A and B. Portions of the structure between helices E and B and helices C and D are disordered and not visible in the crystal structure.

The residues important for receptor binding and activation have been identified for several four-helix-bundle growth factors [for review, see Mott and Campbell (1995)]. For factors that induce formation of either homo- or heterodimeric forms of their receptors, such as growth hormone (Wells et al., 1993) and IL-4 (Kruse et al., 1993), mutagenesis data suggest that two spatially separated binding sites exist on the surface of the protein. For growth hormone, this model has been confirmed by X-ray crystallography of the hormone−receptor complex (de Vos et al., 1992).

As a step toward a better understanding of the molecular basis of G-CSF function, we have performed comprehensive alanine scanning mutagenesis (Cunningham & Wells, 1989) of the cytokine surface. Since protein structures are generally tolerant of amino acid substitution at surface positions (Bashford et al., 1987; Reidhaar-Olson & Sauer, 1988), this approach allows identification of residues likely to be directly important for function. We report the identification of two regions on the surface of G-CSF that are critical for receptor binding and biological activity.

EXPERIMENTAL PROCEDURES

Strains and Plasmids. The following *Escherichia coli* strains were used in this work: strain DH10B (Grant et al., 1990) is *araD139* Δ(*ara, leu*)7697 Δ*lacX74 galU galK hsdR hsdM*$^+$ *rpsL deoR* ϕ80d*lacZ*ΔM15 *endA1 nupG recA1 mcrA mcrB*; strain BMH71-18 *mutS* (Zell & Fritz, 1987) is *thi supE* Δ(*lac*−*proAB*)[*mutS*::Tn10] F′ [*proAB lacI*^{q} *lacZ*ΔM15]. Plasmid pBAD-G4 is a derivative of the ampicillin-resistance vector pBAD18 (Guzman et al., 1995) in which a synthetic gene for human G-CSF containing 13 unique restriction sites has been placed under the control of the inducible *araBAD* promoter.

Mutagenesis. Mutations were introduced using either cassette mutagenesis (Reidhaar-Olson et al., 1991; Wells et al., 1985) or primer-directed mutagenesis followed by restriction selection (Deng & Nickoloff, 1992; Wells et al., 1986). In the latter technique, oligonucleotide primers were designed to introduce an alanine mutation at the desired

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

[1] Abbreviations: G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte−macrophage colony-stimulating factor; GH, growth hormone; IL-2, interleukin 2; IL-4, interleukin 4; HPLC, high-performance liquid chromatography; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; MTS, 3-(4,5-dimethylthiazol-2-yl)-5-[3-carboxymethoxy)phenyl]-2-(4-sulfophenyl)-2*H*-tetrazolium; EC_{50}, dose required for 50% maximal cell proliferative response; IC_{50}, dose required for 50% inhibition of binding of radiolabeled ligand; NMR, nuclear magnetic resonance.

S0006-2960(95)02705-X CCC: $12.00

codon and a silent change in a nearby restriction site. Restriction selection was imposed before and after transformation into *E. coli* strain BMH 71-18 *mutS* (Zell & Fritz, 1987). Mutagenized plasmid DNA produced by either technique was introduced into strain DH10B (Gibco BRL) by transformation, with selection for resistance to ampicillin. Mutations were confirmed by dideoxy DNA sequencing (Sanger et al., 1977).

Purification. G-CSF variants were purified as previously described (Lu et al., 1992; Zsebo et al., 1986), with several modifications. Cells were grown in super broth (Ausubel et al., 1989) to log phase, and expression of the G-CSF gene was induced by addition of L-arabinose to 0.2%. Following overnight growth at 37 °C, another aliquot of L-arabinose was added and growth was continued for 3 h. Cells were harvested by centrifugation, and the insoluble material following lysozyme treatment and sonication was isolated. This inclusion body pellet was washed once with 1% sodium deoxycholate/5 mM EDTA/5 mM dithiothreitol/50 mM Tris-HCl (pH 9.0) and once with water and then solubilized overnight at a total protein concentration of 1−4 mg/mL at room temperature in 2% Sarkosyl/20 μM $CuSO_4$/50 mM Tris-HCl (pH 8.0). G-CSF variants were purified to homogeneity on a reversed-phase C_{18} HPLC column (Waters) run on a BioCAD chromatography workstation (PerSeptive Biosystems). G-CSF variants typically eluted at approximately 56% acetonitrile in a gradient run in 0.1% trifluoroacetic acid. Peak fractions were pooled and lyophilized, and the protein was dissolved in 6 M guanidine hydrochloride/10 mM sodium acetate (pH 4.0) and then dialyzed against 10 mM sodium acetate (pH 4.0). Concentrations were determined by absorbance at 280 nm, using an extinction coefficient of 14 600 M^{-1} cm^{-1}. Typical final yields of purified G-CSF variants were in the range of 5−30 mg/L of culture.

Biological Activity Assay. Bone marrow cells from mouse femurs (Charles River Labs strain SKH-hrBR) were harvested immediately before each cell proliferation assay. For each G-CSF variant, a dose−response curve was generated in triplicate at concentrations between 2 μg/mL and 230 fg/mL. Protein samples were added to 2.5×10^4 cells and incubated for 5 days at 37 °C in RPMI 1640 medium (Gibco/BRL) supplemented with 15% fetal bovine serum (Hyclone) and L-glutamine. Cell growth was quantitated colorimetrically at the end of the assay period after addition of MTS (Promega).

Microphysiometry. Dose−response curves were generated using the Cytosensor microphysiometer (Molecular Devices) (McConnell et al., 1992). Cells of the mouse leukemic line NFS-60 (Holmes et al., 1985), at a density of 2×10^5 cells/chamber, were exposed for 15 min to concentrations of cytokine ranging from 200 ng/mL to 100 fg/mL. A detectable acidification of the medium occurred within 2 min of stimulation and continued for several minutes after removal of the protein. Data were collected until the response returned to near the baseline rate. The maximum acidification rates obtained were plotted as a function of G-CSF concentration to derive EC_{50} values.

Receptor Binding Assay. Measurements of the affinity of G-CSF variants for the G-CSF receptor were carried out as described previously (Nicola & Metcalf, 1984). NFS-60 cells (Holmes et al., 1985) were washed and suspended in RPMI medium (Gibco BRL) containing 10% fetal bovine serum (Hyclone). Cells (2×10^6) were incubated for 2 h at room temperature in a volume of 100 μL with 60 000−80 000 cpm of [^{125}I]G-CSF (Amersham; 800−1000 Ci/mmol) and unlabeled G-CSF variants as competitor at concentrations ranging from 1 μg/mL to 12 pg/mL. Nonspecific binding was determined by addition of a 1000-fold excess of unlabeled ligand. Cells were centrifuged through a 150 μL layer of a 3:2 mixture of dibutyl phthalate−dioctyl phthalate. Centrifuge tubes were frozen on dry ice and then cut to allow determination of the cell-bound radioactivity in a γ counter (Beckman). Counts bound were plotted as a function of ligand concentration to determine IC_{50} values for variants.

RESULTS

In order to facilitate mutagenesis, a synthetic gene for G-CSF containing 13 unique restriction sites was constructed and expressed in *E. coli*. The gene contains an initiation methionine codon followed by the coding sequence of the 174 amino acid human G-CSF sequence.[2] The wild-type G-CSF protein contains two critical disulfide bridges and one free cysteine that is not required for activity (Ishikawa et al., 1992; Kuga et al., 1989; Lu et al., 1992; Wingfield et al., 1988). To avoid potential difficulties caused by the presence of an unpaired cysteine (Arakawa et al., 1993), we have defined the variant bearing a Cys → Ser substitution at position 17 as our pseudo-wild-type molecule and performed all subsequent mutagenesis in this genetic background. In agreement with previous reports (Ishikawa et al., 1992; Lu et al., 1992), the pseudo-wild-type molecule possesses full biological activity (data not shown).

In order to identify important functional residues in G-CSF, we adopted a mutagenesis strategy that involved a thorough investigation of the protein surface. The choice of which positions to mutate was made by examination of the crystal structure of recombinant human G-CSF (Hill et al., 1993). Sixty-one positions, most of them on the protein surface, were chosen as sites for alanine substitution. All but three of these variants (LA41, PA97, and VA163) expressed protein at sufficiently high levels to allow purification. As shown in Figure 1, the mutations are distributed over the surface of the protein in such a way that no large regions of the molecule remain unexplored.

The biological activity of each purified protein was determined by measuring its proliferative effect on mouse bone marrow cells. Dose−response curves were generated in triplicate for each variant, and each protein was tested on at least two, and in most cases three, separate occasions. A range of EC_{50} values was observed (Table 1), with many variants resembling wild type and others exhibiting as little as 5% of wild-type activity.

To determine whether a given mutation leads to a meaningful change in biological activity, the differences between mutant and pseudo-wild-type EC_{50} values were tested for significance at a 99% confidence interval using Student's *t*-test (Brown & Hollander, 1977). The rigorous confidence interval of 99% eliminates mutants that may show small effects on activity but that have standard deviation values large enough to introduce doubt as to the significance

[2] Numbering of amino acids begins with the first residue following the N-terminal methionine. Mutant proteins are identified using the one-letter code for the amino acids: a Phe to Ala substitution at position 144 is indicated by FA144.

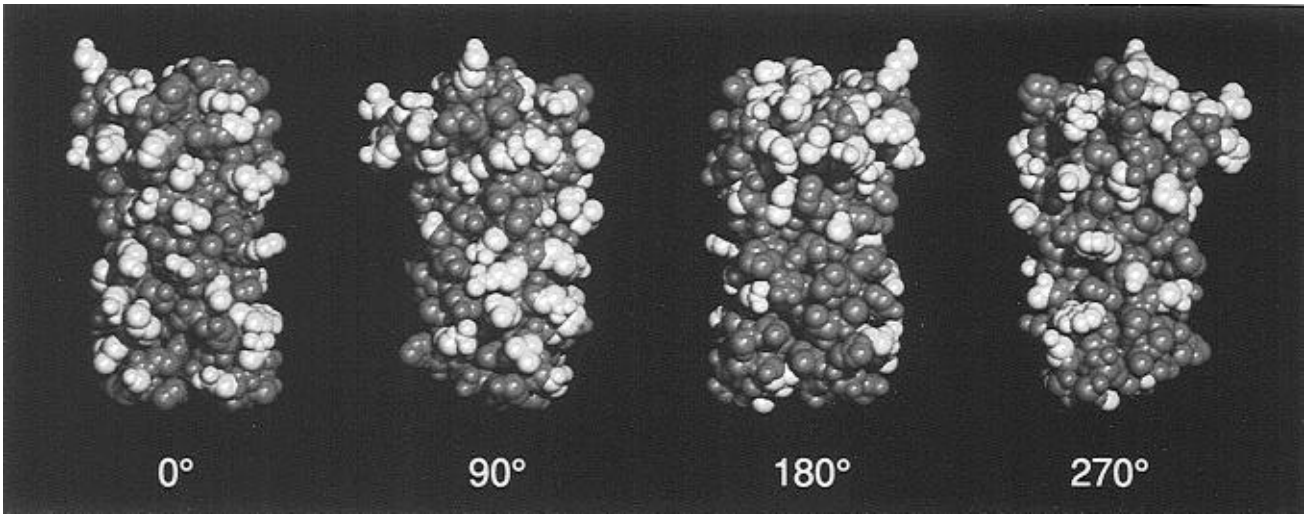

FIGURE 1: G-CSF residues individually mutated to alanine. Four views of G-CSF are shown, each related to its neighbors by 90° rotation. Side chains investigated by substitution with alanine are shown in yellow. Main chain atoms and side chains that were not mutated are shown in blue. Crystal coordinates (Protein Data Bank reference 1RHG) are from Hill et al. (1993).

Table 1: Biological Activity of G-CSF Variants[a]

mutation	EC_{50} (ng/mL)	sd (ng/mL)	*n*	relative activity	mutation	EC_{50} (ng/mL)	sd (ng/mL)	*n*	relative activity
SA12	0.54	0.44	3	0.87	SA62	0.95	1.12	3	0.50
LA15	1.44	0.52	3	0.33	SA66	0.76	0.15	3	0.62
KA16	0.52	0.40	3	0.90	QA67	0.37	0.32	3	1.26
LA18	0.82	0.25	3	0.57	LA71	0.14	0.07	3	3.29
EA19	8.65	4.88	4	0.05	QA77	0.54	0.27	3	0.87
RA22	0.44	0.21	3	1.08	SA80	0.51	0.32	3	0.92
KA23	0.50	0.29	3	0.95	FA83	0.52	0.21	3	0.91
QA25	1.10	0.49	3	0.43	QA90	0.35	0.27	3	1.33
DA27	0.26	0.09	3	1.79	EA93	0.98	0.46	3	0.48
LA31	2.06	0.36	3	0.23	EA98	0.90	0.14	3	0.52
QA32	0.37	0.08	3	1.28	PA101	0.74	0.50	3	0.64
EA33	0.41	0.12	3	1.15	TA102	0.19	0.10	2	2.48
KA34	1.87	1.45	3	0.25	DA104	0.96	0.18	2	0.49
TA38	0.19	0.04	3	2.44	TA105	0.44	0.23	4	1.06
YA39	0.85	0.30	3	0.55	LA108	0.19	0.13	3	2.48
KA40	2.83	3.01	4	0.17	DA112	1.14	0.84	3	0.41
HA43	0.36	0.31	3	1.31	TA115	0.32	0.08	3	1.46
PA44	0.96	0.53	3	0.49	WA118	0.19	0.09	3	2.44
EA45	0.45	0.42	3	1.05	EA122	0.68	0.09	3	0.69
EA46	1.19	1.46	4	0.40	EA123	0.39	0.15	3	1.20
LA47	3.23	0.76	3	0.15	LA124	1.15	0.44	4	0.41
VA48	1.43	0.29	3	0.33	QA134	0.97	0.98	2	0.49
LA49	1.11	0.63	3	0.43	MA137	0.67	0.22	3	0.70
LA50	0.68	0.09	3	0.69	SA142	1.33	0.61	3	0.35
HA52	0.28	0.10	3	1.70	FA144	2.80	1.33	3	0.17
SA53	0.43	0.09	3	1.10	QA145	0.76	0.36	3	0.62
LA54	1.27	0.78	3	0.37	RA146	1.28	0.42	3	0.37
PA57	0.40	0.20	3	1.17	SA159	0.19	0.11	3	2.45
WA58	0.48	0.45	3	0.98	RA169	0.98	0.51	3	0.48

[a] EC_{50} is the concentration of protein giving 50% maximal stimulation in a mouse bone marrow cell proliferation assay; sd is the standard deviation; *n* is the number of separate triplicate determinations; relative activity is the ratio of EC_{50} for the pseudo-wild-type protein (0.47 ng/mL; sd = 0.28 ng/mL; $n = 21$) to EC_{50} for the mutant.

of the change. Using this test, 15 variants were found to be compromised in activity: LA15, EA19, QA25, LA31, KA34, KA40, LA47, VA48, LA49, LA54, DA112, LA124, SA142, FA144, and RA146. A nonparametric statistical approach similarly indicated at 99% confidence that six of these variants are less active than normal: EA19, LA31, LA47, VA48, LA124, and FA144. Conversely, although several mutants initially appear to show substantial increases in biological activity, none of these differences were found to be statistically significant.

Nine surface variants that exhibited significant decreases in biological activity were tested in a receptor binding assay. The ability of these proteins to compete with radiolabeled G-CSF for binding to NFS-60 cells was determined. A range of receptor affinity values was observed, from 38% to almost twice the pseudo-wild-type value (Table 2).

The discrepancy between the 4-fold decrease in activity but nearly 2-fold increase in affinity for the KA34 variant suggested that for this mutant other factors may exert an effect on the observed biological activity. The cell proliferation assay requires a 5-day incubation, whereas the receptor binding assay is complete in 2 h. To examine possible long-term instability, KA34 was compared to the pseudo-wild-type protein in a shorter assay format. Also tested was

Table 2: Receptor Binding Activity of Selected G-CSF Variants[a]

mutation	IC_{50} (ng/mL)	sd (ng/mL)	relative affinity
LA15	8.9	2.9	0.65
EA19	15.0	1.0	0.39
KA34	3.3	0.4	1.8
KA40	11.7	2.9	0.50
VA48	7.4	0.4	0.79
LA49	5.6	2.3	1.0
DA112	6.0	3.2	0.97
LA124	6.0	1.0	0.97
FA144	15.2	8.8	0.38

[a] IC_{50} is the concentration of protein giving 50% inhibition of binding of [^{125}I]G-CSF to NFS-60 cells; relative affinity is the ratio of IC_{50} for the pseudo-wild-type protein (5.8 ng/mL; sd = 2.4 ng/mL; $n = 7$) to IC_{50} for the mutant. Each IC_{50} value is an average of two determinations.

variant EA19, which showed decreases in both activity and receptor binding. The assay uses microphysiometry to measure immediate metabolic changes upon addition of the cytokine to NFS-60 cells (McConnell et al., 1992). In this assay, the average EC_{50} values for the pseudo-wild-type protein and EA19 were 0.9 and 7 ng/mL, respectively, in close agreement with the cell proliferation assay. However, the EC_{50} value for the KA34 variant was 0.8 ng/mL, indicating that in the shorter assay format the protein is fully active. As discussed below, this result suggests that the KA34 mutant is unstable over the course of the 5-day cell proliferation assay.

DISCUSSION

Identification of Critical Residues in G-CSF. From our examination of 58 alanine variants of G-CSF it is clear that, in general, the individual amino acid substitutions have at most small effects on biological activity (Table 1). In fact, only 5 of the 58 mutants tested possess less than 25% of the activity of the pseudo-wild-type protein. The variant with the greatest decrease, EA19, still retains 5% activity. This generally high level of tolerance to amino acid substitution at surface positions suggests that the functional determinants in G-CSF are spread out over several residues, such that no particular residue exerts a dominating effect on activity.

In the variants with diminished activity, the alanine substitutions may influence function either directly or indirectly. Separating direct effects on functionally important residues from indirect effects due to changes in protein structure or stability is not always straightforward. In the case of G-CSF, the tendency of the protein to aggregate upon denaturation precludes simple physical measurements of thermodynamic stability. Thermal melts, monitored by circular dichroism, were found to be irreversible and concentration dependent, suggesting that aggregation of denatured protein obscures the equilibrium between the folded and unfolded forms (data not shown). Similarly, guanidine melts were not feasible because G-CSF precipitates at intermediate concentrations of the denaturant (data not shown).

The general tolerance of protein structures to surface substitution (Bashford et al., 1987; Reidhaar-Olson & Sauer, 1988), as well as the similar purification behavior of all the mutants tested, would suggest that most of the down mutations do not exert their effects by disrupting the global stability of the protein. However, the possibility that some mutations may cause local structural perturbations cannot be rigorously excluded. In particular, substitutions at side chains known from the crystal structure (Hill et al., 1993) to be involved in hydrogen bonds or other potentially stabilizing interactions may show such structural effects. Consequently, in interpreting the functional effect of a particular mutation, the environment occupied by the wild-type residue in the crystal structure must be taken into account.

The set of 58 G-CSF variants can be separated into four structural classes, depending upon whether the substitution replaces a buried, surface, surface hydrogen-bonded, or disordered residue with alanine. These variants can also be classified functionally to give the structure−function matrix shown in Table 3. A total of 32 surface side chains that are not involved in hydrogen bonds were mutated; 23 of these substitutions have no significant effect on activity. The remaining 9 changes cause decreases in activity ranging from 2- to 20-fold and identify residues likely to be directly involved in receptor binding.

These residues are located in two distinct regions of the three-dimensional structure (Figure 2). The first of these regions, designated as site 1 and situated at one end of the four-helix bundle, includes four residues widely separated in the primary sequence but near in the three-dimensional structure: Lys40 (immediately C-terminal to helix A), Val48 and Leu49 (from helix E), and Phe144 (from the N-terminal end of helix D). Lys34, at the C-terminus of helix A, is also near this set of residues but, as discussed below, does not appear to be part of site 1. Site 2 is defined by four residues, also dispersed in sequence but spatially clustered: Leu15 and Glu19 (from helix A), Asp112 (from helix C), and Leu124 (immediately C-terminal to helix C).

Mutations at other non-hydrogen-bonded surface residues near either site 1 or site 2 had little or no effect on activity. In the site 1 region, three residues that are in or immediately adjacent to helix E (His43, Glu45, and His 52) showed normal activity when mutated to alanine. Mutations at four additional residues in this region (Tyr39 at the C-terminus of helix A, Pro44 and Glu46 in helix E, and Glu98 near the N-terminus of helix C) led to decreases in activity that were significant at a 95% confidence limit but not at the more stringent 99% confidence interval. In the site 2 region, mutations were introduced without apparent effect at four residues in helix A (Ser12, Lys16, Leu18, Lys23) and one residue in helix C (Glu123). These noncritical residues near either site may be examples of side chains that physically bind to the receptor but make little or no energetic contribution to the interaction (Clackson & Wells, 1995).

The nine most important surface mutations were further characterized by receptor binding studies (Table 2). As observed previously for G-CSF (Nicola & Metcalf, 1984), half-maximal receptor binding occurred at a concentration approximately 10-fold higher than that required to elicit a half-maximal biological response. In general, biological potency was correlated with receptor affinity: the variants with the greatest decreases in activity tended to exhibit the poorest binding, whereas those with less severe functional effects possessed greater receptor affinities. However, none of the mutants showed decreases in affinity as great as their corresponding losses in activity. This discrepancy may be due to the presence of multiple binding determinants located in two separate sites on the cytokine surface. Removing any one of these determinants will have at most a modest effect

Table 3: Classification of Mutational Effects by Structural Class[a]

structural class	effect of Ala mutation on biological activity	
	no significant effect	significant decrease in activity
surface	Ser12, Lys16, Leu18, Lys23, Gln32, Glu33, His43, Pro44, Glu45, Glu46, His52, Pro57, Trp58, Leu71, Phe83, Gln90, Glu98, Pro101, Asp104, Leu108, Glu123, Met137, Ser159	Leu15, Glu19, Lys34, Lys40, Val48, Leu49, Asp112, Leu124, Phe144
surface H-bonded	Arg22, Thr38, Tyr39, Ser53, Gln77, Ser80, Glu93, Thr105, Thr115, Trp118, Glu122, Gln145, Arg169	Ser142, Arg146
buried	Asp27, Leu50, Thr102	Gln25, Leu31, Leu47, Leu54
disordered	Ser62, Ser66, Gln67, Gln134	

[a] Residues mutated to alanine are classified according to their environment in the crystal structure (Hill et al., 1993): surface, at least 20% side chain solvent accessibility (Eisenberg et al., 1989; Lee & Richards, 1971); surface H-bonded, surface residues whose side chains appear to form hydrogen bonds with other protein atoms; buried, less than 20% side chain solvent accessibility; disordered, not included in the crystal structure due to poor electron density. Mutations classified as showing no effect on activity are those for which EC_{50} values show no statistically significant difference from that of the pseudo-wild-type protein at a confidence interval of 99% by *t*-test analysis.

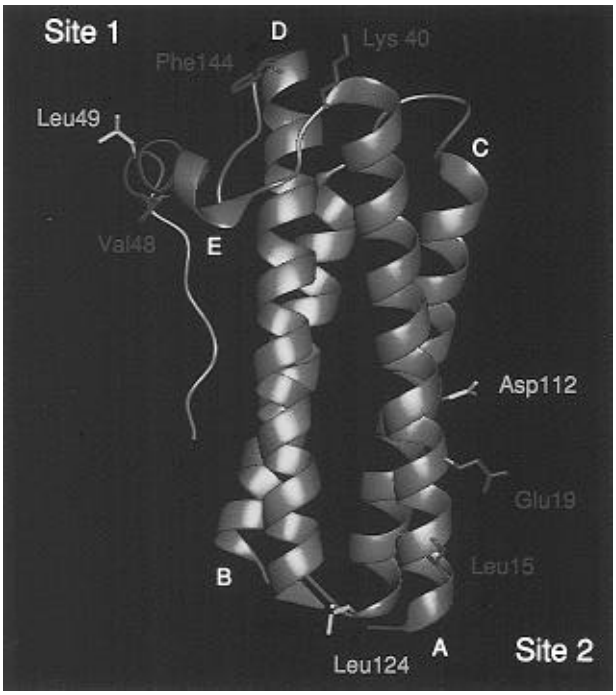

FIGURE 2: G-CSF side chains directly important for biological activity. The side chains shown are surface residues that are required for full biological activity. Side chain colors indicate the effect of mutation: yellow, decrease of less than 3-fold; red, decrease of more than 3-fold. Side chains involved in intramolecular hydrogen bonds are not included. Each of the five helices is labeled near its N-terminus.

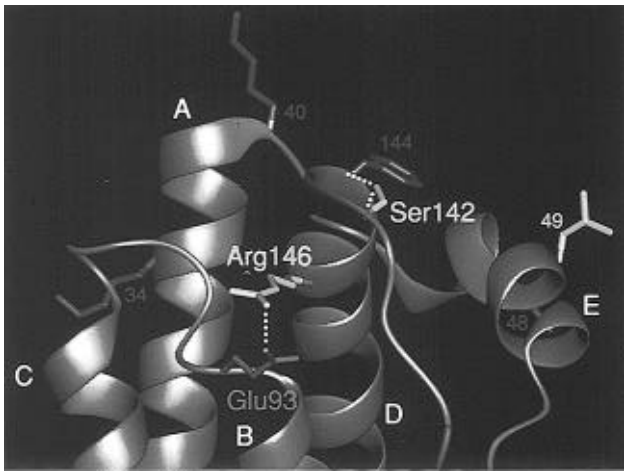

FIGURE 3: Hydrogen-bonding residues important for activity. Side chain colors indicate the effect of mutation: yellow, decrease of less than 3-fold; red, decrease of more than 3-fold; green, no significant effect. The side chain hydroxyl of Ser142 forms hydrogen bonds to the main chain nitrogens of Phe144 and Gln145. The terminal guanido nitrogen of Arg146 hydrogen bonds to the carbonyl oxygen of the Glu93 side chain.

on affinity if the other binding site remains intact; however, a more substantial effect on activity will be observed since both sites are required for receptor dimerization and consequent signal transduction. Moreover, the observation that single mutations in either site cause comparable decreases in affinity suggests that neither site has a decidedly stronger receptor affinity than the other.

The KA34 variant is anomalous in that it is only 25% active in the cell proliferation assay but shows quite high receptor binding affinity. This variant, when tested for activity in the microphysiometer assay format, however, exhibited normal activity. We conclude that Lys34 is not itself a part of either binding region; rather, the KA34 variant appears to be unstable over the course of the 5-day cell proliferation assay. The mechansim of this instability is not clear. Since Lys34 is over 50% solvent exposed and KA34 is fully active in short-term assays of activity and affinity, it is unlikely that the mutant protein is structurally perturbed or thermodynamically destabilized to any significant extent. The mutation may introduce a protease cleavage site or increase nonspecific binding; either of these mechanisms would lead to protein loss during the cell proliferation assay.

Fifteen alanine substitutions were introduced at surface side chains presumed from the crystal structure to be involved in hydrogen bonds (Table 3). Most of these mutations have no effect, indicating that many of the surface hydrogen bonds are not critical for activity. However, two variants, SA142 and RA146, possess slightly decreased activities and may be indicative of surface hydrogen bonds that are structurally important. Since both of these residues are near residues comprising site 1 (Figure 3), these hydrogen bonds may be important for determining the precise architecture of the site 1 receptor binding region. Alternatively, Ser142 and Arg146 may themselves make receptor contacts. In particular, Arg146 may be directly involved in receptor binding since mutation of its hydrogen bond partner, Glu93, has no significant effect on activity.

Of the seven substitutions at buried residues, four cause significant decreases in biological activity (Table 3). These residues are spatially near the surface residues directly involved in receptor binding: Leu31, Leu47, and Leu54 are

near site 1; Gln25 is near site 2. Mutations at these positions may cause either global or local disruption of the cytokine structure, resulting in loss of activity.

Comparison with Phylogenetically Related Sequences. Given that G-CSF from human (Nagata et al., 1986; Souza et al., 1986), mouse (Tsuchiya et al., 1986), and dog and cow (Lovejoy et al., 1993) share greater than 60% amino acid identity and appear to exhibit complete cross-reactivity among species (Lothrop et al., 1988; Nicola et al., 1985), one might expect most of the conserved residues to be important for activity. However, of the 38 strictly conserved residues included in the alanine scan, only 11 are required for full biological activity. The remaining 27 residues appear to be noncritical. Of the 20 nonconserved residues tested, three (Lys40, Val48, and Asp112) were found to be important. Examination of the amino acids that occur at each of these three positions in the various species reveals common chemical properties of these side chains that may explain why they are conserved. Position 40 is occupied by polar residues (Lys or Gln) in each of the four species, position 48 is hydrophobic (Val or Met), and position 112 is either Asp or Asn.

Outside of this phylogenetic family, G-CSF shows little sequence similarity with other cytokines. However, there is a striking conservation of an acidic residue in the middle of the first helix, occupied by Glu19 in G-CSF (Shanafelt et al., 1991). Asp or Glu is observed at this position in GM-CSF, erythropoietin, and most of the interleukins. Mutation of this Glu residue in G-CSF to Ala results in the largest decrease in activity of any of the mutants tested in this study. As discussed below, the residue at this position also appears to play an important functional role in a number of other cytokines and growth factors.

Comparison with Previous Predictions. Previous studies have attempted to identify receptor binding regions in G-CSF by a variety of means. Layton et al. (1991) mapped the epitopes of a panel of neutralizing monoclonal antibodies and identified the region between amino acids 20 and 46 as containing receptor binding determinants. Our results confirm that this region contains important residues but indicate that additional residues outside of this region are also critical. Lovejoy et al. (1993) examined the crystal structures of human, bovine, and canine G-CSF, the structure and mutagenesis data for growth hormone (Wells et al., 1993), and the pattern of G-CSF sequence conservation across species. They predicted two binding sites: site 1, comprised of nine residues throughout the length of helix D, seven residues in helix E, and five residues in the flexible region connecting helices E and B, and site 2, made up of four residues in helix A and two residues in helix C. Our findings concerning the location of site 2 are in general agreement with their predictions, although there are discrepancies in the importance of particular side chains. For site 1, however, we find no evidence that residues in helix D, other than Phe144 near the N-terminus of the helix, are important for activity. Moreover, none of the mutations (SA62, SA66, and QA67) introduced in the region connecting helices E and B have a significant effect. However, our results do confirm the functional importance of residues in helix E, particularly Val48 and Leu49.

Comparison with Other Cytokines and Growth Factors. A variety of approaches have been applied to several cytokines and growth factors to identify receptor binding regions on the surfaces of these proteins. In general, these studies point to the existence of multiple binding sites on the surfaces of cytokines, required for their interactions with two or more receptor subunits on the cell surface. Thorough mutational analysis of the cytokine surface has been carried out for several of these molecules (Figure 4). For IL-2, random mutagenesis of most residues in the protein sequence indicated that 23 side chains are important for cytokine−receptor interactions. These residues are grouped into three spatially distinct regions of the three-dimensional structure, with each group corresponding to an interaction site for one of the IL-2 receptor subunits. The binding site for the α subunit, shown in red in Figure 4 and situated at the top of the four-helix bundle, is similar in position to G-CSF site 1. This site in IL-2 is comprised largely of residues in helix B and between helices A and B. The binding site for the β subunit, shown in yellow, is analogous to G-CSF site 2, including the highly conserved acidic residue in the middle of helix A. A third binding site, for the γ subunit, was

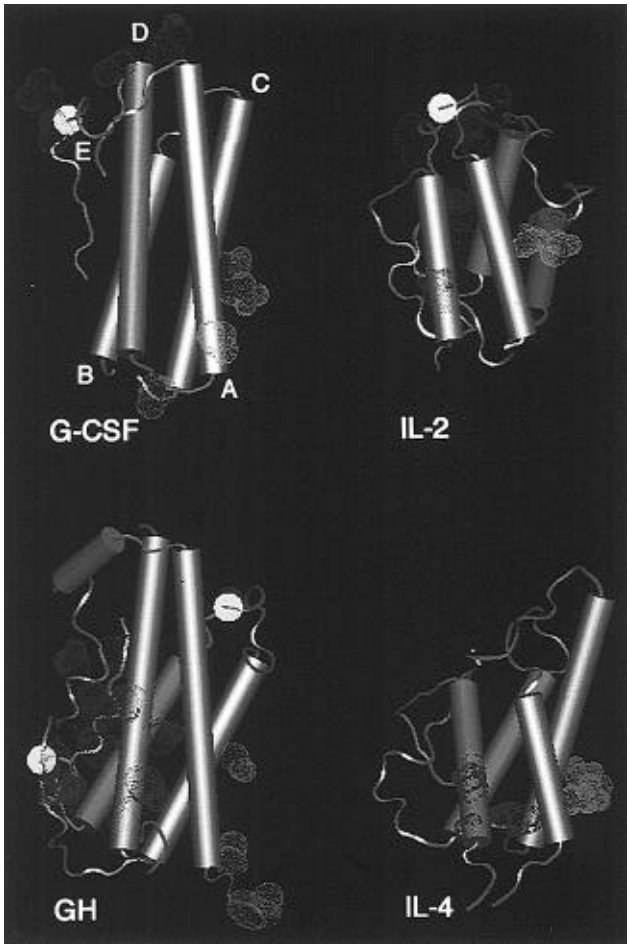

FIGURE 4: Receptor binding residues in four cytokines studied by extensive mutational analysis. All four proteins are shown in the same general orientation; the N-terminal ends of each helix are labeled in the G-CSF structure. Highlighted residues are those showing the largest effects on binding or activity when mutated. Different colors indicate binding sites for different receptor subunits. Mutational data are from Zurawski et al. (1993) (mouse IL-2), Wells et al. (1993) (growth hormone), Kruse et al. (1993) (IL-4), and this study (G-CSF). IL-2 is shown as a model (Peitsch & Jongeneel, 1993) of the mouse structure based on the crystal coordinates of the human protein (McKay, 1992). The IL-4 structure is a minimized average NMR structure (Powers et al., 1992). The growth hormone (de Vos et al., 1992) and G-CSF (Hill et al., 1993) structures are based on X-ray crystallography.

observed along the face of helix D; this site has no analogy with G-CSF receptor binding regions.

A somewhat different spatial organization of receptor binding sites has been observed in growth hormone and IL-4. For growth hormone, scanning alanine mutagenesis (Wells et al., 1993) and X-ray crystallography of the ligand−receptor complex (de Vos et al., 1992) have been used to identify important residues. For IL-4, binding residues were identified by mutagenesis of conserved charged residues (Kruse et al., 1993). In both of these proteins, one receptor binding site, shown in red in Figure 4, is comprised primarily of residues from helix D and, in contrast to site 1 in G-CSF, lies along the face of the four-helix bundle rather than at its end. Since mutations at several residues (Leu71, Gln77, Ser159, and Arg169) in the analogous region of G-CSF had no significant effect on activity, it is unlikely that a binding site exists in this region of the molecule. However, Val163, which also lies in this region, could not be tested because the Ala mutant at this position expressed poorly. Thus, if there is a receptor binding site along the face of helix D in G-CSF, it must be considerably smaller than the corresponding site in these other proteins. In contrast, the second binding site, made up of side chains from helices A and C, is in essentially the same position in growth hormone and IL-4 as the analogous site in G-CSF and IL-2. Since G-CSF and growth hormone both appear to signal via a simple receptor homodimerization mechanism, the different binding site locations on the two molecules suggest that the ligand−receptor binding geometries are substantially different.

Other studies, relying upon less extensive mutational analysis, homolog scanning, or neutralizing antibody epitope mapping, have identified some of the important side chains in a number of other cytokines. Again, a common feature is the importance of the conserved acidic residue in helix A: in IL-3, Asp21 binds the α chain of the receptor (Barry et al., 1994); in GM-CSF, Glu21 is involved in high-affinity binding (Lopez et al., 1992); and in IL-5, Glu12 binds to the β chain of the receptor (Graber et al., 1995).

The results presented here provide the first detailed study of functionally important residues in G-CSF. The presence of two receptor binding sites on the surface of G-CSF is consistent with the dimerization model for receptor activation and lends support to the generality of this mechanism for cytokine signal transduction. Further mutagnesis of residues within these binding sites may lead to the discovery of more potent analogs of G-CSF.

ACKNOWLEDGMENT

The authors thank Robyn Davis and Lina Patel for assistance with mutagenesis and protein purification; Caroline Davies, Maria Schaefer, and Robyn Davis for oligonucleotide synthesis and DNA sequencing support; James Ihle for the gift of NFS-60 cells; and Doug Laurents for assistance with CD measurements.

REFERENCES

Arakawa, T., Prestrelski, S. J., Narhi, L. O., Boone, T. C., & Kenney, W. C. (1993) *J. Protein Chem. 12*, 525−531.

Barry, S. C., Bagley, C. J., Phillips, J., Dottore, M., Cambareri, B., Moretti, P., D'Andrea, R., Goodall, G. J., Shannon, M. F., Vadas, M. A., & Lopez, A. F. (1994) *J. Biol. Chem. 269*, 8488−8492.

Bashford, D., Chothia, C., & Lesk, A. M. (1987) *J. Mol. Biol. 196*, 199−216.

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

Clackson, T., & Wells, J. A. (1995) *Science 267*, 383−386.

Cunningham, B. C., & Wells, J. A. (1989) *Science 244*, 1081−1085.

Deng, W. P., & Nickoloff, J. A. (1992) *Anal. Biochem. 200*, 81−88.

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

Eisenberg, D., Wesson, M., & Yamashita, M. (1989) *Chem. Scripta 29A*, 217−221.

Fuh, G., Cunningham, B. C., Fukunaga, R., Nagata, S., Goeddel, D. V., & Wells, J. A. (1992) *Science 256*, 1677−1680.

Graber, P., Proudfoot, A. E. I., Talabot, F., Bernard, A., McKinnon, M., Banks, M., Fattah, D., Solari, R., Peitsch, M. C., & Wells, T. N. C. (1995) *J. Biol. Chem. 270*, 15762−15769.

Grant, S. G. N., Jessee, J., Bloom, F. R., & Hanahan, D. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4645−4649.

Guzman, L.-M., Belin, D., Carson, M. J., & Beckwith, J. (1995) *J. Bacteriol. 177*, 4121−4130.

Hammond, W. P., IV, Price, T. H., Souza, L. M., & Dale, D. C. (1989) *New Engl. J. Med. 320*, 1306−1311.

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

Holmes, K. L., Palaszynski, E., Fredrickson, T. N., Morse, H. C., III, & Ihle, J. N. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6687−6691.

Ishikawa, M., Iijima, H., Satake-Ishikawa, R., Tsumura, H., Iwamatsu, A., Kadoya, T., Shimada, Y., Fukamachi, H., Kobayashi, K., Matsuki, S., & Asano, K. (1992) *Cell Struct. Funct. 17*, 61−65.

Kruse, N., Shen, B.-J., Arnold, S., Tony, H.-P., Müller, T., & Sebald, W. (1993) *EMBO J. 12*, 5121−5129.

Kuga, T., Komatsu, Y., Yamasaki, M., Sekine, S., Miyaji, H., Nishi, T., Sato, M., Yokoo, Y., Asano, M., Okabe, M., Morimoto, M., & Itoh, S. (1989) *Biochem. Biophys. Res. Commun. 159*, 103−111.

Layton, J. E., Morstyn, G., Fabri, L. J., Reid, G. E., Burgess, A. W., Simpson, R. J., & Nice, E. C. (1991) *J. Biol. Chem. 266*, 23815−23823.

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

Lopez, A. F., Shannon, M. F., Hercus, T., Nicola, N. A., Cambareri, B., Dottore, M., Layton, M. J., Eglinton, L., & Vadas, M. A. (1992) *EMBO J. 11*, 909−916.

Lothrop, C. D., Jr., Warren, D. J., Souza, L. M., Jones, J. B., & Moore, M. A. (1988) *Blood 72*, 1324−1328.

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

Lu, H. S., Clogston, C. L., Narhi, L. O., Merewether, L. A., Pearl, W. R., & Boone, T. C. (1992) *J. Biol. Chem. 267*, 8770−8777.

McConnell, H. M., Owicki, J. C., Parce, J. W., Miller, D. L., Baxter, G. T., Wada, H. G., & Pitchford, S. (1992) *Science 257*, 1906−1912.

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

Mott, H. R., & Campbell, I. D. (1995) *Curr. Opin. Struct. Biol. 5*, 114−121.

Nagata, S., & Fukunaga, R. (1991) *Prog. Growth Factor Res. 3*, 131−141.

Nagata, S., Tsuchiya, M., Asano, S., Kaziro, Y., Yamazaki, T., Yamamoto, O., Hirata, Y., Kubota, N., Oheda, M., Nomura, H., & Ono, M. (1986) *Nature 319*, 415−418.

Nicholson, S., Oates, A. C., Harpur, A. G., Ziemiecki, A., Wilks, A. F., & Layton, J. E. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2985−2988.

Nicola, N. A., & Metcalf, D. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3765−3769.

Nicola, N. A., Begley, C. G., & Metcalf, D. (1985) *Nature 314*, 625−628.

Peitsch, M. C., & Jongeneel, C. V. (1993) *Int. Immunol. 5*, 233−238.

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

Reidhaar-Olson, J. F., & Sauer, R. T. (1988) *Science 241*, 53−57.

Reidhaar-Olson, J. F., Bowie, J. U., Breyer, R. M., Hu, J. C., Knight, K. L., Lim, W. A., Mossing, M. C., Parsell, D. A., Shoemaker, K. R., & Sauer, R. T. (1991) *Methods Enzymol. 208*, 564−586.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.

Shanafelt, A. B., Miyajima, A., Toshio, K., & Kastelein, R. A. (1991) *EMBO J. 10*, 4105−4112.

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

Tkatch, L., & Tweardy, D. J. (1993) *Lymphokine Cytokine Res. 12*, 477−488.

Tsuchiya, M., Asano, S., Kaziro, Y., & Nagata, S. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7633−7637.

Wells, J. A., Vasser, M., & Powers, D. B. (1985) *Gene 34*, 315−323.

Wells, J. A., Cunningham, B. C., Graycar, T. P., & Estell, D. A. (1986) *Philos. Trans. R. Soc. London A 317*, 415−423.

Wells, J. A., Cunningham, B. C., Fuh, G., Lowman, H. B., Bass, S. H., Mulkerrin, M. G., Ultsch, M., & De Vos, A. M. (1993) *Rec. Prog. Horm. Res. 48*, 253−275.

Wingfield, P., Benedict, R., Turcatti, G., Allet, B., Mermod, J.-J., DeLamarter, J., Simona, M. G., & Rose, K. (1988) *Biochem. J. 256*, 213−218.

Zell, R., & Fritz, H.-J. (1987) *EMBO J. 6*, 1809−1815.

Zsebo, K. M., Cohen, A. M., Murdock, D. C., Boone, T. C., Inoue, H., Chazin, V. R., Hines, D., & Souza, L. M. (1986) *Immunobiology 172*, 175−184.

Zurawski, S. M., Vega, F., Jr., Doyle, E. L., Huyghe, B., Flaherty, K., McKay, D. B., & Zurawski, G. (1993) *EMBO J. 12*, 5113−5119.

BI952705X