

Refined Crystal Structure and Mutagenesis of Human Granulocyte-Macrophage Colony-Stimulating Factor

Denise A. Rozwarski,¹ Kay Diederichs,¹ Randy Hecht,² Tom Boone,² and P. Andrew Karplus¹

¹Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca New York; ²AMGEN, Inc., Thousand Oaks, California

ABSTRACT The crystal structure of recombinant human granulocyte-macrophage colony stimulating factor (rhGM-CSF) has been refined against data extending to a resolution of ~ 2.4 Å along a^* and ~ 1.9 Å along b^* and c^* . Anisotropic scale factors of $B_{11} = -20.8$ Å², $B_{22} = 7.4$ Å², $B_{33} = 13.3$ Å² corrected for the more rapid fall of diffraction in the a^* direction. The anisotropy correlates with the weak crystal packing interactions along the a axis. In addition to apolar side chains in the protein core, there are 10 buried hydrogen bonding residues. Those residues involved in intramolecular hydrogen bonding to main chain atoms are better conserved than those hydrogen bonding to other side chain atoms; 24 solvation sites are observed at equivalent positions in the two molecules in the asymmetric unit, and the strongest among these are located in clefts between secondary structural elements. No buried water sites are seen. Two surface clusters of hydrophobic side chains are located near the expected receptor binding regions. Mutagenesis of 11 residues on the helix A/helix C face confirms the importance of Glu-21 and shows that Gly-75 and Gln-86, located on helix C, each cause a greater than fourfold drop in activity. Glu-21 and Gly-75, but not Gln-86, are structurally equivalent to residues involved in the growth hormone binding to its receptor.

© 1996 Wiley-Liss, Inc.

INTRODUCTION

The development of macrophages and granulocytes is governed by protein hormones, called colony-stimulating factors, which function as intercellular messengers that bind and stimulate specific high-affinity receptors on the surface of appropriate progenitor cells.¹ Granulocyte-macrophage colony-stimulating factor (GM-CSF) has a broad range of activities, including participation in the generation of erythrocytes² and dendritic cells,³ as well as the modulation of antibody-mediated neutrophil phagocytosis,⁴ endothelial cell migration,⁵ and alveolar macrophage surfactant processing.⁶ GM-CSF has been approved for medical use and is expected to show the most promise for the reduction of neutropenia, a condition that can follow such procedures as

bone marrow transplantation or cancer chemotherapy. Two independent crystal structures of rhGM-CSF^{7,8} show it folds into an antiparallel four-helix bundle that is slightly open on one side to accommodate a short two-stranded antiparallel β -sheet (Fig. 1). The four helices are arranged in the up-up-down-down topology characteristic of the "short-chain" subgroup within the helical bundle family of cytokines.⁹ We have carried out a detailed comparison of GM-CSF with the other four structurally known short-chain cytokines (interleukin [IL]-2, IL-4, IL-5, and MCSF) to define a common structural core which includes about half of the residues in each structure.¹⁰

The receptors for helical bundle cytokines mostly belong to the hematopoietin receptor superfamily,^{11,12} and the crystal structure of the growth hormone-receptor complex^{13,14} provides a framework for thinking about helical bundle cytokine receptor interactions. In the growth hormone complex, two identical receptor chains bind different sites on the single asymmetric ligand. The first receptor recognizes helices D and A and the loop connecting helices A and B (designated "site 1"), while the second receptor chain to bind recognizes the A and C helices (designated "site 2"). In the case of GM-CSF, two different receptor subunits (α and β) are involved: an α subunit that is capable of low-affinity (nM) binding to GM-CSF and confers cytokine specificity,¹⁵ and a β subunit that by itself is not capable of binding GM-CSF, but, together with the α subunit which builds a high-affinity (pM) receptor. The GM-CSF receptor β subunit can also combine with different α subunits specific for IL-3 and IL-5 to form the high-affinity receptors for those cytokines.¹⁶

As reviewed by Mott and Campbell,¹⁷ most available evidence suggests that GM-CSF is recognized in a similar manner, with the regions on GM-CSF equivalent to site 1 and site 2 of growth hormone binding to the receptor α subunit and β subunits

Received September 28, 1995; accepted April 9, 1996.

Address reprint requests to Dr. P. Andrew Karplus, Section of Biochemistry, Cornell University, Ithaca, NY 14853.

Dr. Diederichs is currently at the Fakultät für Biologie, Universität Konstanz, 78434 Konstanz, Germany.

Dr. Roswarski is currently at the Department of Biochemistry, Albert Einstein College of Medicine, New York, NY 10461.

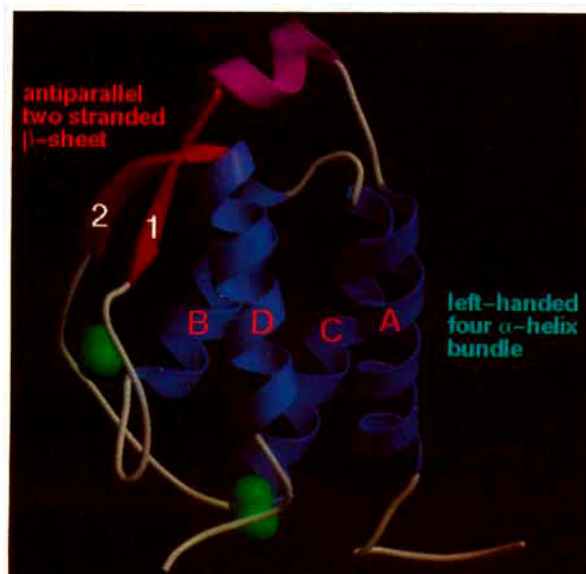


Fig. 1. The structure of human GM-CSF. (A) Ribbon representation³⁸ of rhGM-CSF showing α helices A–D (blue), β -strands $\beta 1$ and $\beta 2$ (red), extra α helix (violet), disulfide bonds (green), and connecting loops (white).

respectively. In particular, Glu-21 has been shown to be crucial for binding to the β subunit,^{18,19} and two Glu-21 mutants have been shown to act as antagonists.²⁰ In another study, a polyclonal antiserum raised against a synthetic peptide corresponding to part of helix D was shown to both bind GM-CSF and block its activity.²¹ In addition, an antiidiotypic antibody, which might resemble the peptide itself, was shown to block low-affinity binding, demonstrating that this region of helix D probably interacts with the α subunit. In a separate study, site-directed mutants with Glu-108 or Asp-112 of helix D substituted by Arg, displayed a reduced signaling ability and reduced affinities for both the $\alpha\beta$ -receptor complex and the α subunit alone.¹⁹ One conflicting study recently reported that an antibody recognizing helix C of GM-CSF interfered with binding to the receptor α subunit suggesting that GM-CSF receptor interactions may not be completely consistent with the growth hormone analogy.²²

In our initial structural report,⁷ we defined a putative receptor binding site that involved residues on the A and C helices of GM-CSF. Based on the growth hormone analogy, this proposed receptor binding site corresponds largely to site 2. In this report, we describe the fully refined crystal structure of human GM-CSF and use site-directed mu-

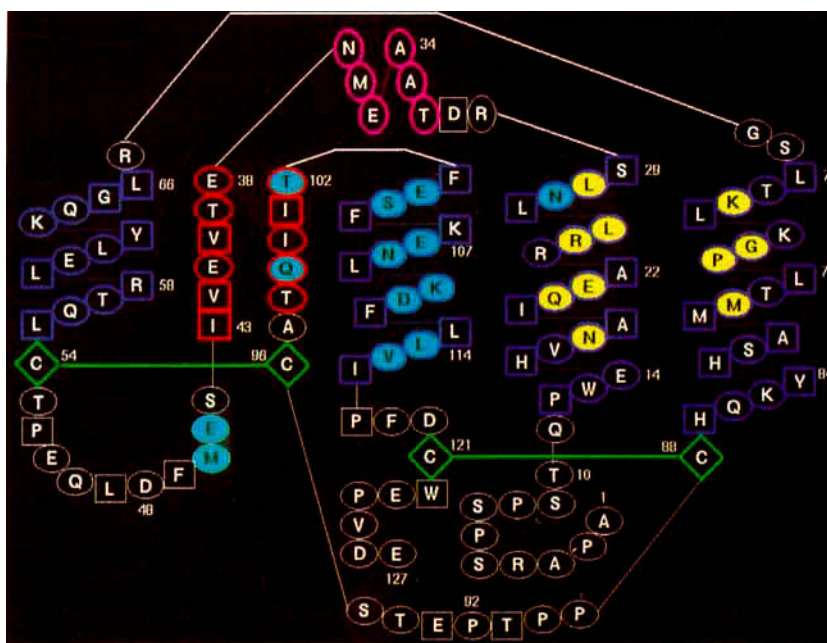


Fig. 1 continued. (B) Secondary structure (colors as in Fig. A) and solvent accessibility (square = $<20\%$ exposed; oval = $>20\%$ exposed) of each residue, and highlighting the residues corresponding to receptor binding site 1 (cyan) and 2 (yellow) of growth hormone. We have included residues 45, 46, 99, and 102 as part of the putative site 1 because the crossover regions of growth hormone involved in receptor binding are close to these residues in the structural superposition. Although the greater length of the growth hormone helices allows for multiple reasonable structural

alignments of GM-CSF, we have shown previously that one alignment (pairing GMCSF residues 20, 60, 70, and 110 with growth hormone residues 11, 77, 114, and 173, respectively) is markedly better than the rest.¹⁰ This best alignment is different than that published by Goodal et al.³⁹ for GM-CSF and growth hormone. In the GM-CSF/growth hormone alignment represented here, the buried GMCSF residues Ala-22, Leu-59, Leu-73, Leu-114, and Phe-113 are aligned with identical residues, but no identities exist among the solvent-exposed residues.

tagenesis of exposed residues on helices A and C to directly document their importance.

METHODS

High-Resolution Data Collection

The protocols for the crystallization of rhGM-CSF and x-ray diffraction data collection at 2.4 Å resolution have been described earlier.^{7,23} The crystals grow in space group $P2_12_12_1$ with unit cell $a = 47.6$, $b = 59.1$, $c = 126.7$ Å and two molecules in the asymmetric unit. The crystals of GM-CSF diffract anisotropically, with diffraction in the a^* direction falling into the noise near 2.7 Å resolution and diffraction in the b^* and c^* directions extending to somewhat beyond 2 Å resolution (Fig. 2). Therefore, to minimize the amount of crystal x-ray beam exposure and maximize the quality of the data, only reflections with high k and l indices, but low h indices, were collected from two additional crystals. These data were then combined with the previous data using the software accompanying the San Diego Multiwire Systems area detector.²⁴ A summary of the completeness and quality of the data is included in Figure 2.

Structure Solution

The structure was solved in a stepwise fashion by multiple isomorphous replacement (MIR) and partial model refinements, aided by noncrystallographical symmetry (NCS) averaging.⁷ Some important details of the structure solution and refinement that have not yet been reported are included here. A total of seven heavy atom derivatives were found, all of which had reasonable phasing power at 6 Å resolution, but poor phasing power beyond 4 Å resolution. We found that with these multiple derivatives, the statistics (phasing power, R_c , figure of merit) reported by the phasing program were rather insensitive to which derivatives were included in the phasing, so we developed an independent probe of phase quality based on the known noncrystallographic symmetry (NCS) of this crystal form. For this purpose, electron density maps in incremental resolution ranges (infinity–6, 6–4, and 4–3.5) were produced, and their NCS correlation coefficient was calculated in order to assess the quality of the phases in the respective resolution ranges. This proved to be a valuable tool to monitor progress during the refinement of heavy atom models and to select which derivatives were making useful contributions. Finally, four heavy atom derivatives were chosen to calculate phases at a resolution of 3.5 Å with an overall figure of merit of 0.60. Using these MIR phases, the NCS correlation coefficients were 0.62, 0.39, and 0.14 in the three resolution ranges specified above. This method has been used to test the performance of various phasing programs independent of the statistics they report.²⁵

Iterative solvent flattening based on the CCP4 li-

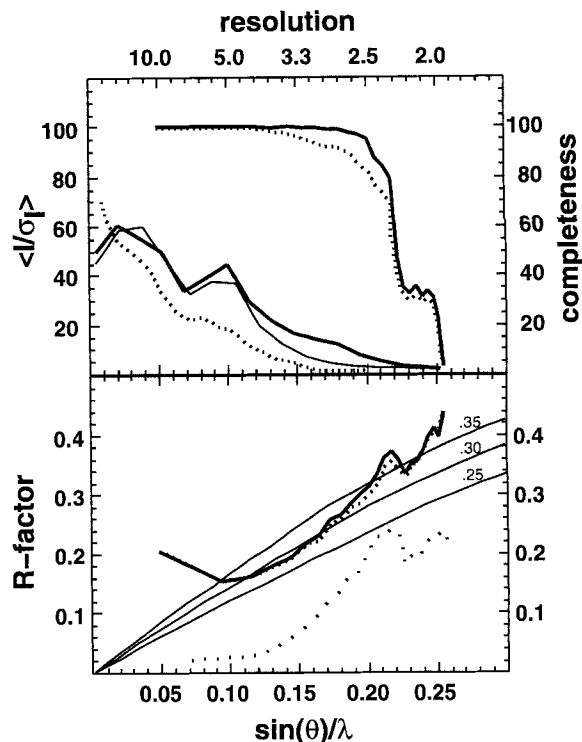


Fig. 2. Diffraction data quality and Luzzati Plot. (A) The signal-to-noise ratios of the diffraction data are separately shown as a function of the h (dotted line), k (thin solid line), and l (thick solid line) indices. To obtain reasonable statistics, each line includes reflections with the other two indices out to $(6 \text{ \AA})^{-1}$ from the axis, but the resolution in each case is an effective resolution based only on the value of the primary index. Also included is the percent completeness as a function of resolution (all data, solid line, data with $F > 2\sigma$, dotted line). The data are >95% complete to 2.4 Å resolution and uniformly near 33% complete between 2.4 and 2.0 Å resolution. In this high-resolution range, the data collected is not a random sample, but includes all of the data with low h (< 8) and high k and/or l . (B) The R factor for the final model (thick line and dotted line for data with $F > 2\sigma$) is shown as a function of resolution. The solid thin lines are the theoretical curves from Luzzati³¹ for 0.25, 0.30, and 0.35 Å residual error in the model. The sparsely dotted line represents the internal agreement (R_{int}) of the data. Note that the precision in the data clearly limits the progress of the refinement, as the R-factor plot parallels the R_{int} plot beyond about 3 Å resolution. The overall R_{int} for the data set is 9.7%. R factor = $\frac{\sum_{hkl} |F_o^{hkl}| - |F_c^{hkl}|}{\sum_{hkl} |F_o^{hkl}|}$, $R_{int} = \frac{\sum_{hkl} |F_o^{hkl}| - |F_o^{-h-k-l}|}{\sum_{hkl} |F_o^{hkl}|}$, where F_o is an observed structure factor and F_c is a calculated structure factor.

brary²⁶ was used both with and without NCS averaging to improve the quality of the map. The electron density map produced when NCS information was included in the phase refinement appeared to be of higher quality and was used for fitting. A comparison with the final calculated phases showed that solvent flattening alone produced a 6° phase improvement, and solvent flattening and NCS together produced a 13° improvement. In the NCS-averaged 3.5 Å resolution map, four helices were clearly visible, but their directions and the connections between them were not clear. 68 residues of poly-Ala, plus residues 54–70 and 99–116 were built into an initial model accounting for these four

helices. In this model only residues 99–116 (helix D) were correctly fit, but the other three helices were built in the wrong direction. After generating the second molecule by the NCS operator, refinement was carried out with the slow cool protocol of X-PLOR²⁷ with NCS restraints. To our surprise, an averaged $2F_o - F_c$ map based on this rather crude model ($R = 37.6\%$ for 7–2.8 Å resolution data) had improved clarity, allowing us to see the correct directions of all four helices. Although sequence could still not be placed, Leu residues were built at positions with visible side chain density, resulting in a model encompassing 85 residues of poly-Ala/poly-Leu and residues 99–116. After the second round of refinement, the sequence of rhGM-CSF could be clearly recognized in several places in the model-phased $2F_o - F_c$ map and the two known disulfides exhibited strong density. Using this information, proper connections were made between the helices to result in a model consisting of residues 8–122. After refinement at 2.4 Å resolution, the last major changes in the model were to shift residues 34–50 (including the residues in strand $\beta 1$) by two residues, and to add residues 5–7 and 123. Two further rounds of refinement without NCS restraints led to the model reported by Diederichs and colleagues.⁷

Higher Resolution Refinement

For further refinement, the higher resolution data were included and an anisotropic scale factor calculated by X-PLOR (version 3.0) was applied to the observed data so that refinement could be carried out against this scaled data. The final anisotropic B factors were $B_{11} = -20.8 \text{ \AA}^2$, $B_{22} = 7.4 \text{ \AA}^2$, $B_{33} = 13.3 \text{ \AA}^2$. Without this scaling, the higher resolution $2F_o - F_c$ electron density map showed “smearing” of the density in the x-direction, reflecting the weakness of the data with higher h indices. Before refinement at 1.9 Å, all ordered water molecules were removed and they were only added back if they showed electron densities $>3 \rho_{\text{rms}}$ in an $F_o - F_c$ map, and $>0.5 \rho_{\text{rms}}$ in the $2F_o - F_c$ map and, for the sites with weaker density, if they had at least one reasonable hydrogen bond partner. Four rounds of conventional positional and individual isotropic temperature factor refinements were then carried out without a sigma cutoff in order to retain the weaker higher resolution data. The final model contained residues 4–124 of molecule A, residues 4–123 of molecule B, and 92 preferred hydration sites modeled as waters. The waters are numbered 401–424 and 501–524 for equivalent positions on molecules A and B, and 601–644 for additional sites. Within each group, the lower numbered water sites have stronger electron density. The final R factor was 23.5% for 19,868 reflections between 10 and 1.92 Å resolution and 22.7% for the 18,384 reflections with $F > 2\sigma(F)$. The rms deviations from ideality are 0.013 Å for bond lengths and 1.6° for bond angles, and the rms devi-

ation of B factors of bonded atoms is 3.0 Å². The average main chain and side chain B factors are 38 and 42 Å² respectively. The somewhat high R factor is probably related to the relatively high disorder in these crystals, which leads both to low-accuracy data and structural features not well accounted for by a single isotropic model with overall anisotropic scaling. We note in this regard that in oscillation photographs taken at the Cornell High Energy Synchrotron Source, there is a great deal of non-Bragg or thermal diffuse scattering. The coordinates are deposited in the protein data bank with access code 2GMF.

Mutant Generation and Analysis

GM-CSF mutants were prepared using oligonucleotide-based site-directed mutagenesis in M13 as described.²⁸ Activity of the mutants was characterized using a cell proliferation assay as described in refs. 29 and 30. The assays for all mutants used purified proteins and the reported numbers are averages of two measurements.

RESULTS AND DISCUSSION

Model Quality

Although the nominal resolution of this analysis is 1.92 Å, the effective resolution is lower because of the weak data resulting from the anisotropic diffraction of the crystals, and the anisotropic scaling, which partially scales the stronger data in the b^* and c^* directions down to match the weak data in the a^* direction (see Methods). Subjective evaluation of the electron density for well-ordered regions of the structure suggests that the effective resolution is closer to 2.2 Å. A Luzzati plot³¹ (Fig. 2), indicates that the average error for the coordinates of the well-defined regions of the model is around 0.35 Å. Among nonglycine residues, only Cys 54, in a left-handed α -helical conformation, and Asp 120, with ϕ, ψ near $-105^\circ, -100^\circ$, lie well outside the β sheet and α -helical regions of the Ramachandran plot. For both residues, the main chain electron densities are well-defined and the conformations are consistent between molecules A and B. Also, Cys-121 in molecule A has ϕ, ψ near $-140^\circ, -150^\circ$, but in molecule B it has a more favorable $\phi, \psi = -140^\circ, -170^\circ$. Interestingly, all three residues are tied to the formation of the disulfides bridging Cys-54: Cys-96 and Cys-121:Cys-88. Overall the model has changed little from the model refined at 2.4 Å resolution, with rms main-chain and side-chain shifts of 0.17 and 0.55 Å. Key changes include correction of a number of side-chain rotamers and more complete placement of solvent.

A comparison of the two molecules in the asymmetric unit allows an independent estimate of the coordinate accuracy (Fig. 3). The overall rms deviation in position between the subunits is 0.4 Å for main-chain atoms and 0.7 Å for side-chain atoms,

averaging near 0.3 Å for the best ordered regions of the main chain. This is close to the Luzatti estimate for coordinate accuracy, suggesting that the observed similarity between molecules A and B may be limited by the precision of the analysis. The regions of the molecule which deviate the most are the N and C termini, and the loops connecting helix A to strand $\beta 1$ and strand $\beta 1$ to helix B. The poorer agreement of these four regions correlates with their higher mobility (Fig. 3), emphasizing the relation between mobility and coordinate accuracy. Also, as these four regions are near other molecules in the crystal, they may truly differ in conformation due to their different environments. The significant differences in B factors between molecules A and B occurring near residues 33 and 88 are likely due to unique crystal contacts involving these residues (see Table II). The high average B factors (near 40 Å²) do not necessarily reflect an intrinsic looseness in the GM-CSF core, as they also are affected by the level of disorder in the crystal lattice. For instance, many cases are known where a given protein crystallizes in multiple crystal forms having very different limits of resolution (and thus very different average B factors).

Structural Features

The overall features of the GM-CSF fold (Fig. 1) were described in our earlier work⁷ and in the independent crystallographic analysis of Walter and co-workers.⁸ Here, we describe more details of the structure, in terms of the interactions and roles of specific residues in stabilizing the fold. To guide the discussion, an alignment of GM-CSF sequences from various species is shown in Figure 5, together with a designation of the final secondary structure and which residues are considered as buried or solvent exposed in our analysis.

Among the 44 buried residues, 30 are hydrophobic (11 Leu, 4 Ile, 4 Phe, 4 Pro, 3 Ala, 2 Val, 1 Met and 1 Gly), 4 participate in disulfide bonds (Cys-54: Cys-96 and Cys-88:Cys-121), 3 are hydrophobic yet commonly form hydrogen bonds (Tyr-62, Tyr-84, and Trp-122), and 7 are hydrophilic (His-15, Ser-29, Asp-31, Arg-58, His-83, Thr-91, and Lys-107). The majority of the buried contacts are formed by the interdigitation of hydrophobic residues between adjacent antiparallel helices so that mostly side chains from helices A and B interact with those from helices C and D, while the residues of the β sheet interact with the residues of helices B and D. Interestingly, many of the similar hydrophobic residues are clustered: a large patch of leucines is near the top of the molecule, and the aromatic rings are arranged in a column.

More noteworthy are the buried hydrophilic residues which are involved in side-chain:side-chain and side-chain:main-chain hydrogen bonds that specifically bridge distant parts of the chain (Table I;

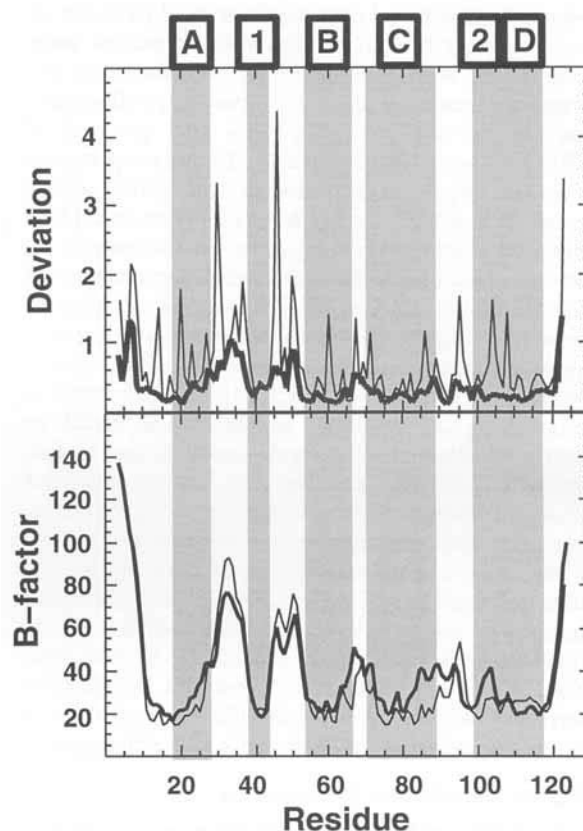


Fig. 3. Comparisons of the coordinates and mobilities of the noncrystallographically related molecules. **Upper panel:** A plot of the rms deviation (in Å) versus residue number based on main-chain atoms (thick line) or side-chain atoms (thin line). **Lower panel:** The average main chain temperature factors (in Å²) for each residue are plotted for molecule A (thick line) and molecule B (thin line). The locations of the secondary structural elements are indicated.

Figure 4). The data in Table I show that those residues involved in side-chain:side-chain interactions are not as well conserved as those which make side-chain:main-chain interactions. This may be a general feature of protein structure as recent mutagenesis experiments on the arc repressor have elegantly documented how complementary substitutions of interacting buried polar side chains can be replaced by hydrophobic side chains with minimal effects on structure and function.³² As we pointed out earlier,¹⁰ this seems to have happened in GM-CSF evolution as the positions of Ser-29, Asp-31, and Lys-107, which form a buried hydrophilic cluster are filled by Met, Val or Ala, and Ile, respectively, in rodent GM-CSFs (Fig. 5). Buried side chains interacting with main-chain atoms may be more conserved because their loss directly affects main-chain conformation and few complementary substitutions may exist. For instance, it is difficult to imagine a substitution for Arg-58 (Fig. 4) and the surrounding residues that could conserve the interactions that stabilize the threading of residues 43–54 inside the long connection between helices C and D. Also, al-

TABLE I. Hydrogen Bonds Made by Polar Atoms of Buried Side Chains

Buried atom	Partner atom	Distance (Å)
*His-15-NE2	Phe-119-O	3.1 (3.4)
Ser-29-OG	Ser-69-OG	3.3 (—)
	Lys-107-NZ	3.1 (3.1)
Asp-31-OD2	Lys-107-NZ	2.9 (—)
*Arg-58-NH2	Glu-45-O	3.4 (3.0)
-NH2	Thr-53-O	3.3 (2.9)
Tyr-62-OH	Asn-109-OD1	2.9 (2.5)
His-83-NE2	Asp-120-O	2.8 (2.8)
*Tyr-84-OH	Pro-118-O	2.6 (3.0)
*Thr-91-OG1	Glu-93-OE1	3.0 (2.7)
Lys-107-NZ	Ser-29-OG	3.1 (3.1)
	Asp-31-OD2	2.9 (—)
Trp-122-NE1	none	

Asterisks in the first column denote those residues which are absolutely conserved among the known GM-CSF sequences. The distances for molecule B are given in parentheses. Asp31 and Ser69 are modelled in somewhat different conformations in molecule B, although the density for these residues suggests there is actually a mixture of conformations present. Although His-83 is not conserved, the Tyr replacement seen in murine GMCSF (Figure 5) could conserve the side chain:main chain hydrogen bond. Trp-122 is technically buried, but the NE1 atom is blocked from bulk solvent only by the side chain of Glu-123 which has very high B-factors. The substitution Trp122→Lys seen in murine GMCSF is accompanied by the changes of neighboring residues Pro52→Leu and Leu49→Phe (Figure 5). Since a Lys may not properly fit into this pocket without the complementary substitutions, structural effects may account for the large loss of activity observed for a Trp122→Lys mutation.³⁷

though Glu 93 (31% accessible) is not buried by our criteria, it is conserved and plays a similar structural role, hydrogen bonding the main-chain amides of Gln-56 and Thr-57 at the beginning of helix B (Fig. 4).

Among the 81 solvent exposed residues, 15 are negatively charged (3 Asp and 12 Glu), 11 are positively charged (5 Lys, 5 Arg, and 1 His), 29 are neutral hydrophilic (8 Ser, 10 Thr, 4 Asn, and 7 Gln), and 26 are hydrophobic (7 Pro, 5 Leu, 4 Ala, 3 Met, 2 Val, 2 Gly, 1 Ile, 1 Phe, and 1 Trp). The charged side chains on GM-CSF are distributed with most of the negatively charged residues on the BD face of the molecule, and most of the positively charged residues on the AC face. As has been noted before, this leads to significant electrostatic dipole, which may be important for receptor binding.³³ Two local salt bridges exist, one on helix A (Glu-21 to Arg-24) and one on helix D (Glu-108 to Lys-111). Because exposed hydrophobic residues are often crucial for receptor binding,³⁴ we note that two exposed hydrophobic patches exist: the first spans across the lower portions of helices A and D, and the loop from strand β 1 to helix B and includes Trp 13, Val 16, Met 46, Leu 49, Leu 115, Val 116, and Phe 119; the second spans across the upper portion of helix A and central portion of helix C and includes Leu 25, Leu 28, Gly

75, Pro 76, Met 79, Ala 81. Interestingly, these regions match well with the expected receptor binding sites 1 and 2 (see Fig. 1).

Crystal Contacts

The GM-CSF crystals contain 60% solvent, and there are six types of interfaces between interacting pairs of molecules (Table II). Pair numbers 1 through 4 involve noncrystallographically related molecules, with pair 1 representing the two molecules in the asymmetric unit chosen here and pair 4 corresponding to the asymmetric unit chosen by Walter and coworkers.⁸ Pair 4 has the largest interface by far, suggesting that it is the preferential interaction during crystal formation. The molecules of pair number 4 are related by an approximate two-fold axis (177.4°) and contain numerous hydrophobic and hydrophilic contacts. At the center of the contact area are several hydrophobic residues (Trp 13, Val 16, and Phe 119), which belong to the first solvent-exposed hydrophobic patch mentioned earlier. All of the other pairs of contacting molecules have much smaller interfaces and are much more hydrophilic in character.

The crystal lattice interactions provide a rationale for the anisotropic diffraction of the crystals: contacts 1, 2, 4, and 6 all contribute to the formation of a tightly packed row of "dimers" of molecules along the *b*-axis; interaction 3 butts the well-ordered strand β 2 up against a symmetry version of itself in an antiparallel interaction to establish the spacing along the *c*-axis; and interaction 5, which is small in extent and consists of side-chain interactions involving the very mobile residues 31–33, is the main determinant of the spacing along the *a*-axis. The lack of well-defined interactions along the *a*-axis may also explain the difficulty in getting isomorphous derivatives which have significant phasing power beyond 4 Å resolution. In this regard, the length of the *a*-axis was quite sensitive to heavy atom soaks, and varied in a 2 Å (4%) range around the native value.

Ordered Water Molecules

Using our criteria (see Methods), 92 preferred solvation sites were modeled as waters. These can be grouped into 24 pairs of sites, which are found at equivalent positions (within 1 Å) in molecules A and B, and 44 additional sites. The waters mostly surround the outside of the cytokine fold, and no ordered water sites are found between the helices of the tightly packed bundle. Thus a large fraction of the solvation properties seen can be expected to be crystal packing-dependent. However, one deep surface pocket exists which has a cluster of the three best-defined preferred hydration sites seen in both molecules in the asymmetric unit. The pocket is formed by helix D, the loop connecting strand β 1 to helix B, and helix B (Fig. 4). In this cluster, the

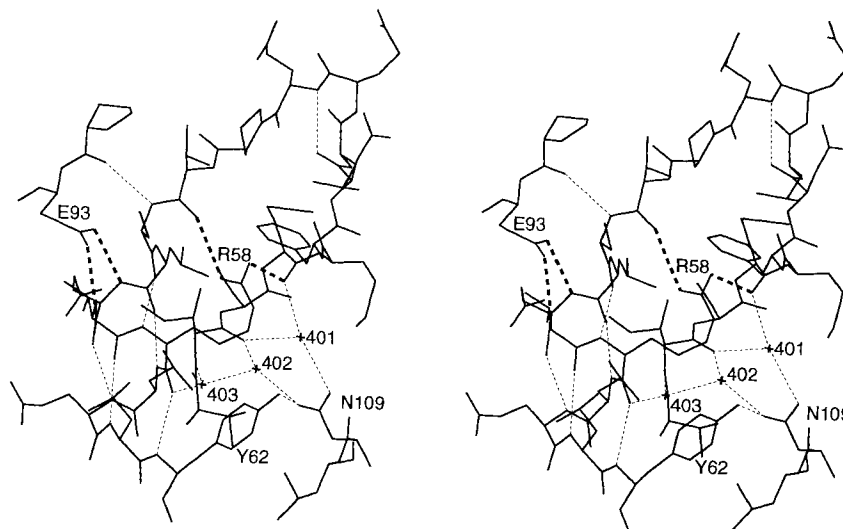


Fig. 4. Interactions of the key residues Arg-58 and Glu-93. Residues Val42-Tyr62, Pro92-Glu93, and Asn109-Leu110 are shown along the three best-defined water sites 401, 402, and 403.

Hydrogen bonds are shown by dashed lines, with the thicker lines used for the side-chain:main-chain hydrogen bonds involving Arg-58 or Glu-93.

	10	20	30	40	50	60	70	
human	APARSPSPST	QPWEHVNAIQ	EARRLLNLSR	DTAAEMNETV	EVISEMFDLQ	EPTCLOTRLE	LYKQGLRGSL	
cow	APTRPNTAT	RPWQHVDIAK	EALSLLNHSS	DTAAVMNDT	EVVSEKFDSD	EPTCLOTRLE	LYKQGLRGSL	
sheep	APTRQSPVPT	RPWQHVDIAK	EALSLLNDST	DTAAVMDET	EVVSEKFDSD	EPTCLOTRLE	LYKQGLRGSL	
pig	APTRPSPVPT	RPWQHVDIAK	EALSLLNNSN	DTAAVMNETV	DIVCEMFDPO	EPTCVQTRLN	LYKQGLRGSL	
dog	APTRSPTLVT	RPSQHVDAIQ	EALSLLNNSN	DVTAVMNKAV	KVVSEVDFPE	GPTCLETRLQ	LYKEGLQGS	
rat	APTRSPNPVT	RPWKHVDAIK	EALSLLNMDR	ALENEKNEDV	DIISNEFSIQ	RPTCVQTRLK	LYKQGLRGSL	
mouse	APTRSPIIVT	RPWKHVEAIK	EALNLLDDMP	VT---LNEEV	EVVSNEFSFK	KLTCVQTRLK	LYEQLRGNF	
	** * * *	* ** **	** **		*	** **	** *	
		10	20	30	40	50	60	70
		aaaaaaaa	aaaaaaaa	111	111	bbbbbbb	bbbbbb	c
human	TKLKGPLTMM	ASHYKQHCPP	TPETSCATQI	ITFESFKENL	KDFLLVIPFD	CWEPVQE		
cow	TSLMGSLLTMM	ATHYKHCPP	TPETSCGTQF	ISFKNFKEDL	KEFLFIIPFD	CWEPVQK		
sheep	TSLTGSLTMM	ASHYKHCPP	TQETSCTQI	ITFKSFKENL	KDFLFIIPFD	CWEPVQK		
pig	TRLKSPLTLL	AKHYEQHCPL	TEETSCTQS	ITFKSFKDSL	NKFLFTIPFD	CWEPVQK		
dog	TSLKNPLTMM	ANHYKHCPP	TPESPCATQN	INFKSFKENL	KDFLFNIPFD	CWKPVKK		
rat	TKLNGALTMI	ASHYQTNCP	TPETDCEIEV	ITFEDFIKNL	KGFLFDIPFD	CWKPVKK		
mouse	TKLKGALNMT	ASYYYQTYCPP	TPETDCEIQV	TTYADFIDSL	KTFLTDIPFE	CKKPSQK		
	* * *	* * **	* * *	*	** **	* *		

Fig. 5. Alignment of GM-CSF sequences. The numbering and secondary structural elements seen for rhGM-CSF are indicated in lines 1 and 2. Residues that are buried in the human structure are shaded. Generally, all residues with <20% solvent-accessible surface area in one or both of the GM-CSF molecules are designated as buried. Two exceptions are Cys-121 and Lys-63; Cys-121 has near 33% accessible surface, but as it is involved in a disulfide with the buried Cys-88 we have designated it as buried;

Lys-63 has only 18% solvent exposed surface in one molecule, but as its amino group is well exposed, we have designated it as exposed. References for the sequences follow: human,^{40,41} cow,⁴² sheep,⁴³ dog,⁴⁴ rat,⁴⁵ mouse.⁴⁶ Variants of the human sequence are known with Thr at position 100⁴⁷ and variants of the mouse sequences are known with Ala at position 100 and Val or Gly at position 125.^{40,41}

majority of the contacts made with the water molecules are through main-chain hydrogen bonds. Three other very strong water sites (406, 506, and 602) make no contacts of <3.5 Å with surrounding atoms and are thus more likely to be weakly preferred sulfate or phosphate sites than highly preferred water sites.

Site-Directed Mutants of GM-CSF

To test our proposed receptor binding site,⁷ we have made single site-directed mutants of three sol-

vent exposed residues on helix A and eight of the nine solvent exposed residues on helix C (Table III). According to the growth hormone analogy, these residues are part of or near receptor binding site 2 (Fig. 1B). The Arg and Lys residues near the periphery of the region were mutated to Ala, because of the likelihood that the hydrophobic portions of their side chains were involved. Glu and Glu residues and Arg-24 were mutated to Leu to specifically probe the importance of their hydrogen bonding groups, and the remaining residues (Gly, Ser, Thr, Met) were

TABLE II. Crystal Packing Interactions

Pair	Residues	Partner residues	Area	Contacts
1	A: 24,27-30	B: 41,43,57,60-61,94,98	370 (5)	4 + 7
2&2'	A: 50,124	B: 68,72	135 (2)	0 + 2
3&3'	A: 41,94-95,98,100	B: 37,39,44,98-100	275 (4)	1 + 6
4&4'	A: 4-7,10-13,16-20,23,49,114-120	B: 4-7,10-13,16-20,23,49,114-120	925 (13)	18 + 17
5&5'	A: 31-33,104	A: 41,60,64,74	205 (3)	0 + 7
6&6'	B: 4,24,28,30	B: 7,86-87,94	295 (4)	1 + 3

Contacts made by the crystallographically independent molecules A or B are noted separately. Area gives the total buried surface area (in Å²) between each pair and in parentheses the percent of the surface area of a GM-CSF monomer. Contacts are defined as numbers of atomic approaches within 4 Å, and are reported separately for hydrophobic contacts and hydrophilic contacts. For simplicity, residue ranges are given that may include single residues that do not themselves make contact. The symmetry operations to generate the contacting pairs are as follows: (1) x,y,z ; (2) $x,y-1,z$; (2') $x,y+1,z$; (3) $-1/2+x,1/2-y,-z$; (3') $1/2+x,1/2-y,-z$; (4) $1-x,-1/2+y,-1/2-z$; (4') $1-x,1/2+y,-1/2-z$; (5) $1/2+x,1/2-y,-z$; (5') $-1/2+x,1/2-y,-z$; (6) $1-x,1/2+y,-1/2-z$; (6') $1-x,-1/2+y,-1/2-z$.

each mutated to slightly larger residues with altered hydrogen bonding capabilities, which should cause significant binding decreases if the residues are buried near the center of the interface. All of the mutated side chains are exposed and as expected we observed no properties suggesting the mutants had folding defects. Indeed, although Pro-76 is in the region we targeted for mutation, we did not mutate it because the requisite change in main chain structure would be expected to lead to folding or stability problems.

The only residue with dramatic effects on activity is Glu-21, which has already been shown by others to be a key residue involved in binding to the β -receptor subunit.^{18,35} For the Glu-21→Leu mutant, CD spectra showed the protein had folded properly (data not shown), but crystallization of that mutant has not yet been successful. The other two residues in helix A (Arg-23, Arg-24) which were probed showed about a twofold loss in activity, again consistent with the results of others.^{18,19}

Among the 8 residues in the C helix that were probed, all mutations except that at Met-79 showed some loss in activity, and the mutations at Gly-75 and Gln-86 showed more than a fourfold loss. These changes in activity, although much smaller than observed for Glu-21 mutant, suggest the residues are involved in the receptor binding interface. For comparison, in the systematic mutagenesis of human growth hormone,³⁶ only 14 of the 49 Xaa→Ala mutants tested showed greater than fourfold decreases in binding, and all 14 residues were involved or adjacent to residues involved in the crystallographically observed interface.¹³ We also note that since many residues physically involved in a binding interface do not make large net contributions to binding energy,³⁴ small energetic consequences of mutation cannot be used to exclude a residue from the interface.

Although these mutagenesis results are rather limited, they do provide the first systematic analysis of residues in helix C. The only other GM-CSF mutants with changes in helix C were generated by

TABLE III. Activities of rhGM-CSF Mutants

Mutation	Activity (%)
Glu-21→Leu	0.01
Arg-23→Ala	56
Arg-24→Leu	46
Lys-72→Ala	83
Lys-74→Ala	46
Gly-75→Ser	23
Thr-78→Asp	41
Met-79→Lys	100
Ser-82→Asp	43
Lys-85→Ala	61
Gln-86→Leu	14

Shanafelt and associates³⁷ in their study of human/murine GM-CSF specificity. They showed that a Thr-78→Asn mutant was fully active, and a Met-80→Thr mutant had 34% activity, while the corresponding double mutant had no activity. Although the effects of the Met-80 mutation, which is buried, are likely due to structural disruption, the effects of the Thr-78 mutation in the context of the Met-80 mutation suggest it is at or near the binding site even though the mutation Thr-78→Asn alone had no effect on activity. The importance of Gln-86 and Ser-82 to activity indicate the receptor binding site is further down the C helix than is indicated by strict analogy with growth hormone (Fig. 1B), and is in good agreement with independent mapping done with monoclonal antibodies which implicated the distal two-thirds of the C helix.²²

ACKNOWLEDGMENTS

This work was supported in part by NIH grant GM-43566, and by a Feodor-Lynen Fellowship from the Alexander von Humboldt Foundation to K.D.

REFERENCES

- Nicola, N.A. Hemopoietic cell growth factors and their receptors. *Annu. Rev. Biochem.* 58:45-77, 1989.
- Sieff, C.A., Emerson, S.G., Donahue, R.E., Nathan, D.G., Wang, E.A., Wong, G.G., Clark, S.C. Human recombinant granulocyte-macrophage colony stimulating factor: A multilineage hematopoietin. *Science* 230:1171-1173, 1985.

3. Inaba, K., Inaba, M., Deguchi, M., Hagi, K., Yasumizu, R., Ikehara, S., Muramatsu, S., Steinman, R.M. Granulocytes, macrophages, and dendritic cells arise from a common major histocompatibility complex class II-negative progenitor in mouse bone marrow. *Proc. Natl. Acad. Sci.* 90:3038–3042, 1993.
4. Weisbart, R.H., Kacena, A., Schuh, A., Golde D.W. GM-CSF induces human neutrophil IgA-mediated phagocytosis by an IgA Fc receptor activation mechanism. *Nature* 332:647–648, 1988.
5. Bussolino, F., Wang, J.M., Defilippi, P., Turrini, F., Sanavio, F., Edgell, C.J., Anglietta, M., Arese, P., Mantovani, A. Granulocyte- and granulocyte-macrophage colony stimulating factors induced human endothelial cells to migrate and proliferate. *Nature* 337:471–473, 1989.
6. Dranoff, G., Crawford, A.D., Sadelein, M., Ream, B., Rashid, A., Bronson, R.T., Dickersin, G.R., Bachurski, C.J., Mark, E.L., Whitset, J.A., Mulligan, R.C. Involvement of granulocyte-macrophage colony stimulating factor in pulmonary homeostasis. *Science* 264:713–716, 1994.
7. Diederichs, K., Boone, T., Karplus, P.A. Novel fold and putative receptor binding site of granulocyte-macrophage colony stimulating factor. *Science* 254:1779–1782, 1991.
8. Walter, M.R., Cook, W.J., Ealick, S.E., Nagabhushan, T.L., Trotta, P.P., Bugg, C.E. Three-dimensional structure of recombinant human granulocyte-macrophage colony stimulating factor. *J. Mol. Biol.* 224:1075–1085, 1992.
9. Sprang, S.R., Bazan, J.F. Cytokine structural taxonomy and mechanisms of receptor engagement. *Curr. Opin. Struct. Biol.* 3:815–827, 1993.
10. Rozwarski, D.A., Gronenborn, A.M., Clore, G.M., Bazan, J.F., Bohm, A., Wlodawer, A., Hatada, M., Karplus, P.A. Structural comparisons among the short-chain helical cytokines. *Structure* 2:159–173, 1994.
11. Cosman, D., Lyman, S.D., Idzerda, R.L., Beckman, M.P., Park, L.S., Goodwin, R.G., March, C.J. A new cytokine receptor superfamily. *Trends Biochem. Sci.* 15:265–270, 1990.
12. Bazan, J.F. Structural design and molecular evolution of a cytokine receptor superfamily. *Proc. Natl. Acad. Sci.* 87:6934–6938, 1990.
13. de Vos, A.M., Ultsch, M., Kossiakoff, A.A. Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* 255:306–312, 1992.
14. Wells, J.A., de Vos, A.M. Structure and function of human growth hormone: Implications for the hematopoietins. *Annu. Rev. Biophys. Biomol. Struct.* 22:329–351, 1993.
15. Gearing, D.P., King, J.A., Gough, N.M., Nicola, N.A. Expression cloning of a receptor for human granulocyte-macrophage colony stimulating factor. *EMBO J.* 8:3667–3676, 1989.
16. Nicola, N.A., Metcalf, D. Subunit promiscuity among hemopoietic growth factor receptors. *Cell* 67:1–4, 1991.
17. Mott, H.R., Campbell, I.D. Four-helix bundle growth factors and their receptors: protein-protein interactions. *Curr. Opin. Struct. Biol.* 5:114–121, 1995.
18. Lopez, A.F., Shannon, M.F., Hercus, T., Nicola, M.A., Cambareri, B., Dottore, M., Layton, M.J., Eglinton, L., Vadas, M.A. Residue 21 of human granulocyte-macrophage colony stimulating factor is critical for biological activity and for high but not low affinity binding. *EMBO J.* 11:909–916, 1992.
19. Hercus, T.R., Cambareri, B., Dottore, M., Woodcock, J.M., Bagley, C.J., Vads, M.A., Shannon, M.F., Lopez, A.F. Identification of residues in the first and fourth helices of human granulocyte-macrophage colony stimulating factor involved in biologic activity and in binding to the alpha and beta chains of its receptor. *Blood* 83:3500–3508, 1994.
20. Hercus, T.R., Bagley, C.J., Cambareri, B., Dottore, M., Woodcock, J.M., Vadas, M.A., Shannon, M.F., Lopez, A.F. Specific human granulocyte-macrophage colony stimulating factor antagonists. *Proc. Natl. Acad. Sci.* 91:5838–5842, 1994.
21. Seelig, G.F., Prosisse, W.W., Scheffler, J.E. A role for the carboxyl terminus of human granulocyte-macrophage colony stimulating factor in the binding of ligand to the alpha-subunit of the high affinity receptor. *J. Biol. Chem.* 269:5548–5553, 1994.
22. Brown, C.B., Pihl, C.E., Kaushansky, K. Mapping of human granulocyte-macrophage-colony-stimulating-factor domains interacting with the human granulocyte-macrophage-colony-stimulating-factor-receptor α -subunit. *Eur. J. Biochem.* 225:873–880, 1994.
23. Diederichs, K., Jacques, S., Boone, T., Karplus, P.A. Low resolution structure of recombinant human granulocyte-macrophage colony stimulating factor. *J. Mol. Biol.* 221:55–60, 1991.
24. Howard, A.J., Nielson, C., Xuong, N.H. Software for a diffractometer with multiwire area detector. *Methods Enzymol.* 114:452–472, 1985.
25. Diederichs, K. A comparison of some heavy-atom refinement and phasing programs. *Joint CCP4 and ESF-EACBM newsletter on protein crystallography.* 31:23–30, 1994.
26. Collaborative Computational Project, N. 4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D* 50:760–763, 1994.
27. Brünger, A.T., Kuriyan, J., Karplus, M. Crystallographic R-factor refinement by molecular dynamics. *Science* 235:458–460, 1987.
28. Boone, T., Chazin, V., Kenney, W., Swanson, E., Altrock, B. Construction, purification and biological activities of recombinant human interleukin-2 analogs. *Dev. Biol. Std.* 69:157–168, 1988.
29. Smith, K.A., Zsebo, K.M. Basic protocols UT-7 cell proliferation assay for murine and human SCF activity. *Curr. Protocol. Immunol.* 1:17-1, 1992.
30. Komatsu, N., Nakauchi, H., Miwa, A., Ishihara, T., Eguchi, M., Moroi, M., Okada M., Sato, Y., Wada, H., Yawata, Y., Suda, T., Miura, Y. Establishment and characterization of human leukemic cell line with megakaryocytic features: dependency on GM-CSF, IL-3, or EPO for growth and survival. *Cancer Res.* 51:341–348, 1991.
31. Luzzati, P.V. Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Crystallogr.* 5:802–810, 1952.
32. Waldburger, C.D., Schildbach, J.F., Sauer, R.T. Are buried salt bridges important for protein stability and conformational specificity? *Nat. Struct. Biol.* 2:122–128, 1995.
33. Demchuk, E., Mueller, T., Oshikinat, H., Sebald, W., Wade, R.C. Receptor binding properties of four-helix bundle growth factors deduced from electrostatic analysis. *Protein Sci.* 3:920–935, 1994.
34. Clackson, T., Wells, J.A. A hot spot of binding energy in a hormone receptor interface. *Science* 267:383–386, 1995.
35. Shanafelt, A.B., Miyajima, A., Kitamura, T., Kastelein, R.A. The amino-terminal helix of GM-CSF and IL-5 governs high affinity binding to their receptors. *EMBO J.* 10:4105–4112, 1991.
36. Cunningham, B.C., Wells, J.A. High-resolution epitope mapping of hGH-receptor interactions by alanine scanning mutagenesis. *Science* 244:1082–1085, 1989.
37. Shanafelt, A.B., Johnson, K.E., Kastelein, R.A. Identification of critical amino acid residues in human and mouse granulocyte-macrophage colony stimulating factor and their involvement in species specificity. *J. Biol. Chem.* 266:13904–13810, 1991.
38. Kraulis, P.J. "MOLSCRIPT": a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24:946–950, 1991.
39. Goodall, C.J., Bagley, C.J., Vadas, M., Lopez, A.F. A model for the interaction of the GM-CSF, IL-3 and IL-5 receptors with their ligands. *Growth Factors* 8:87–97, 1993.
40. Miyatake, S., Otsuka, T., Yokota, T., Lee, F., Arai, K. Structure of the chromosomal gene for granulocyte-macrophage colony stimulating factor: comparison of the mouse and human genes. *EMBO J.* 4:2561–2568, 1985.
41. Cantrell, M.A., Anderson, D., Cerretti, D.P., Price, V., McKereghan, K., Tushinski, R.J., Mochizuki, D.Y., Larson, A., Grabstein, K., Gillis, S., Cosman, D. Cloning, sequence, and expression of a human granulocyte-macrophage colony stimulating factor. *Proc. Natl. Acad. Sci.* 82:6250–6254, 1985.
42. Maleszewski, C.R., Schoenborn, M.A., Cerretti, D.P., Wignall, J.M., Picha, K.S., Cosman, D., Tuschinski, R.J., Gillis, S., Baker, P.E. GM-CSF: Molecular cloning and biological activity of the recombinant protein. *Mol. Immunol.* 25:843–850, 1988.
43. McInnes, C.J., Haig, D.M. Cloning and expression of a

- cDNA encoding ovine granulocyte-macrophage colony-stimulating factor. *Gene* 105:275-279, 1991.
44. Nash, R.A., Schuening, F., Appelbaum, F., Hammond, W.P., Boone, T., Morris, C.F., Schlichter, S.J., Storb, R. Molecular cloning and in vivo evaluation of canine granulocyte-macrophage colony-stimulating factor. *Blood* 78: 930-937, 1991.
45. Smith, L.R., Lundeen, K.A., Dively, J.P., Carlo, D.J., Brostoff, S.W. Nucleotide sequence of the Lewis rat granulocyte-macrophage colony stimulating factor. *Immunogenetics* 39:80, 1994.
46. Gough, N.M., Gough, J., Metcalf, D., Kelso, A., Grail, D., Nicola, N.A., Burgess, A.W., Dunn, A.R. Molecular cloning of cDNA encoding a murine haematopoietic growth regulator, granulocyte-macrophage colony stimulating factor. *Nature* 309:763-767, 1984.
47. Wong, G.G., Witek, J.S., Temple, P.A., Wilkens, K.M., Leary, A.C., Luxenberg, D.P., Jones, S.S., Brown, E.L., Kay, R.M., Orr, E.C., Shoemaker, C., Golde, D.W., Kaufman, R.J., Hewick, R.M., Wang, E.A., Clark, S.C. Human GM-CSF: Molecular cloning of the complementary DNA and purification of the natural and recombinant proteins. *Science* 228:810-815, 1985.