Structure of KW-2228, a tailored human granulocyte colony-stimulating factor with enhanced biological activity and stability

Isao Fujii[a], Yoshitomo Nagahara[b], Motoo Yamasaki[b], Yoshiharu Yokoo[b], Seiga Itoh[b], Noriaki Hirayama[a,*]

[a] *Department of Biological Science and Technology, Tokai University, 317 Nishino, Numazu, Shizuoka 410-03, Japan*
[b] *Tokyo Research Laboratories, Kyowa Hakko Kogyo Co. Ltd., 3-6-6 Asahimachi, Machida, Tokyo 194, Japan*

Received 3 March 1997; revised version received 25 March 1997

Abstract KW-2228 is a tailored human granulocyte colony-stimulating factor (hG-CSF) which has more potent granulopoietic activity and is more stable than wild-type hG-CSF. Analysis of the 2.3 Å resolution crystal structure of KW-2228 unambiguously revealed a four-α-helix bundle motif with up-up-down-down connectivity. The structures of long overhand loops connecting the helices and the N-terminus have been definitively determined. The present analysis has clearly revealed that substituted residues play important roles in fastening a long overhand loop to the N- and C-termini to fix the conformation. This conformation should be responsible for a substantial enhancement of the biological activity and stability.

Key words: Granulocyte colony-stimulating factor; Tailored protein; Conformation; X-ray crystallography; Structure-activity relationship; Lymphokine

1. Introduction

Colony-stimulating factors (CSFs) are a family of glycoproteins that stimulate the proliferation and differentiation of hematopoietic cells in vivo. These proteins include granulocyte colony-stimulating factor (G-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), macrophage colony-stimulating factor (M-CSF), interleukins and erythropoietin [1]. Human G-CSF (hG-CSF) is a 19.6-kDa glycoprotein consisting of 174 amino acid residues [2]. One *O*-linked glycosyl group is attached to hG-CSF at Thr^{133}. This group contributes to the stability of the molecule but does not influence its binding to receptors [3]. G-CSF, produced mainly by macrophages, stimulates neutrophilic granulocyte colony formation from bone marrow stem cells and induces the terminal differentiation of leukemic cells [4]. hG-CSF has been isolated from tumor cells [5] and its cDNAs have been cloned [2]. One of the two cDNAs which encodes 174 amino acid residues was isolated from a cDNA library of lipopolysaccharide-stimulating human peripheral blood macrophages [6]. The recombinant hG-CSF has been found to be useful in the treatment of 5-fluorouracil-induced hematopoietic injury in mice [7]. More than 100 novel hG-CSFs have been engineered with the aim of identifying novel proteins with enhanced biological activity and physicochemical properties [8]. One of these engineered proteins, designated KW-2228, has more potent granulopoietic activity than that of wild-type hG-CSF, both in vitro and in vivo [8,9]. KW-2228 (Neu-up) is currently being used clinically and is yielding very satisfactory results. Thr^{1}, Leu^{3}, Gly^{4}, Pro^{5} and Cys^{17} in wild-type hG-CSF are replaced by Ala, Thr, Tyr, Arg and Ser, respectively, in KW-2228. KW-2228 was also found to be more stable both physicochemically and biologically than wild-type hG-CSF, especially in phosphate-buffered saline at 56°C and in human plasma at 37°C [9]. It is essential to determine the three-dimensional structure of KW-2228 to understand the mechanism underlying its high potency and stability, and thus we have undertaken X-ray analysis of KW-2228 [10]. Recently the crystal structures of native canine, bovine and human G-CSFs have been published [11,12]. These three G-CSFs show about 80% sequence similarity and share the same four-α-helix bundle motif with up-up-down-down connectivity. In these molecular models, however, the positions of more than 9% of the amino acid residues could not be located due to poor electron density, in particular the missing residues in hG-CSF range 1–7, 61–69, 126–135 and 172–17. They account for almost 16% of all the residues. Thus the published molecular model of hG-CSF [11] is incomplete, and we need a complete molecular model to be able to fully elucidate the structure–function relationships. In this paper we report on a crystal structure of KW-2228 in which the positions of 98% of the amino acid residues are located. Based on the structure of G-CSF which has been determined, the mechanism underlying the high potency and stability of KW-2228 will be discussed.

2. Materials and methods

KW-2228 was crystallized as described elsewhere [10]. We obtained at least three polymorphs, but in the present analysis we used crystals which belong to the monoclinic space group C2 with lattice constants $a = 82.0$, $b = 49.2$, $c = 49.4$ Å, and $\beta = 113.9°$ and diffract to 2.0 Å resolution. The asymmetric unit consists of one monomer with a solvent content of 49%. The 2.0 Å resolution data set for the native crystal was collected at the beamline BL6A of the Photon Factory (Tsukuba, Japan) by use of monochromatized radiation at $\lambda = 1.04$ Å. The intensity data were collected on a screenless Weissenberg camera for macromolecules [13] with Fuji imaging plates (Fuji Film Co. Ltd.). The data were processed with WEIS [14] and scaled with PROTEIN [15]. A 1.7 Å resolution model of bovine G-CSF[11] was positioned in the asymmetric unit by molecular replacement (MR) using X-plor [16]. The top rotation function solution gave a peak which was twice as high as the second-highest one. The model which was constructed based on the top translation function solution was improved by rigid body refinement, and the crystallographic R factor for this model was 46.5% (10-3 Å). The MR phases were further improved by solvent flattening and density histogram matching using SQUASH [17]. X-plor was used to calculate maps, which were fitted using Turbo-FRODO [18]. The initial model gave an R-factor of 48.5%. The standard simulated annealing procedure of X-plor was used for the refinement. After several cycles of model building, water addition, and energy

*Corresponding author. Fax: +81 (462) 53-4850.
e-mail: hirayama@wing.ncc.u-tokai.ac.jp

PII S0014-5793(97)00538-3

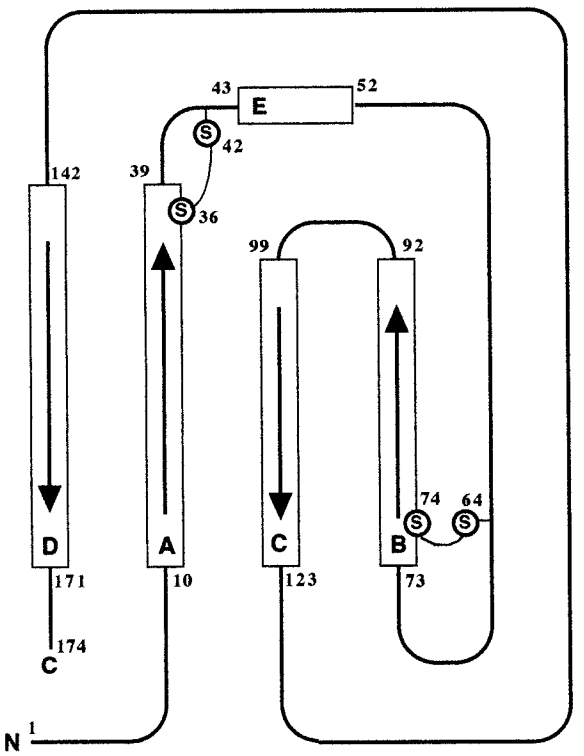

Fig. 1. The schematic drawing of the secondary structure elements of KW-2228. The rectangles mark α-helices. A, B, C, D and E helices are labeled according to Scheme 1.

minimizations, the final model was obtained. The final refined model yielded a crystallographic R-factor of 21.5%, including all data from 8.0–2.3 Å. The r.m.s. deviations from ideal bond lengths and angles are 0.022 Å and 3.70°. Residues Thr^{3}, Pro^{132} and Pro^{174} are flexible and show low electron density. The figures in this paper were created with MOLSCRIPT [19]. Final coordinates will be deposited in the Brookhaven Data Bank.

3. Results

The main structural feature of KW-2228 is an antiparallel four-α-helix bundle (Figs. 1 and 2) which has a left-handed twist with overall dimensions of approximately 45×25×25 $Å^3$. The α-helices, which are designated A through D, consist of residues 10–39, 73–92, 99–123, and 142–171, respectively. The overall topology of the bundle is very similar to those observed in bovine, canine and wild-type hG-CSFs and very close to that of an ideal left-handed antiparallel four-α-helix bundle. The four helices are connected by three loops designated AB, BC and CD. The AB and CD loops are long overhand connections, and the shortest loop, BC, joins helix B to helix C. In addition to the major helices there is a short extended 3_{10} helix designated E within the AB loop which consists of residues 43–52.

In the X-ray analysis of wild-type hG-CSF, the electron density for the N- and C-termini and the long overhand loops was poorly defined, and the coordinates of residues 1–8, 62–70, 127–136 and 173–174 were not determined [11]. Therefore the topology of the connections between the four helices in hG-CSF is ambiguous. Even in the molecular model of crystal form II of canine G-CSF which is the G-CSF structure most accurately determined so far [11], the coordinates of residues 1–7, 129–135 and 174 are not located clearly and the structures of the N-terminal region and the middle of the CD loop are not definitive. In the structure of KW-2228, however, the electron density is clearly defined and the positions of all of the residues except 1, 2, 130 and 131 are definitive. Although G-CSF is expected to be structurally flexible due to the presence of the remarkably long overhand loops, the substitution of amino acids at residues 1, 3, 4, 5 and 17 in KW-2228 has frozen the conformation and allowed us to determine almost the entire structure of hG-CSF for the first time. In bovine G-CSF Ala^{93}, Gly^{94} and Ile^{95} form almost one turn of a left-handed helix, but in KW-2228 they are not in the region of the left-handed helix on the Ramachandran plot.

There are two disulfide bridges in hG-CSF. The Cys^{64}–

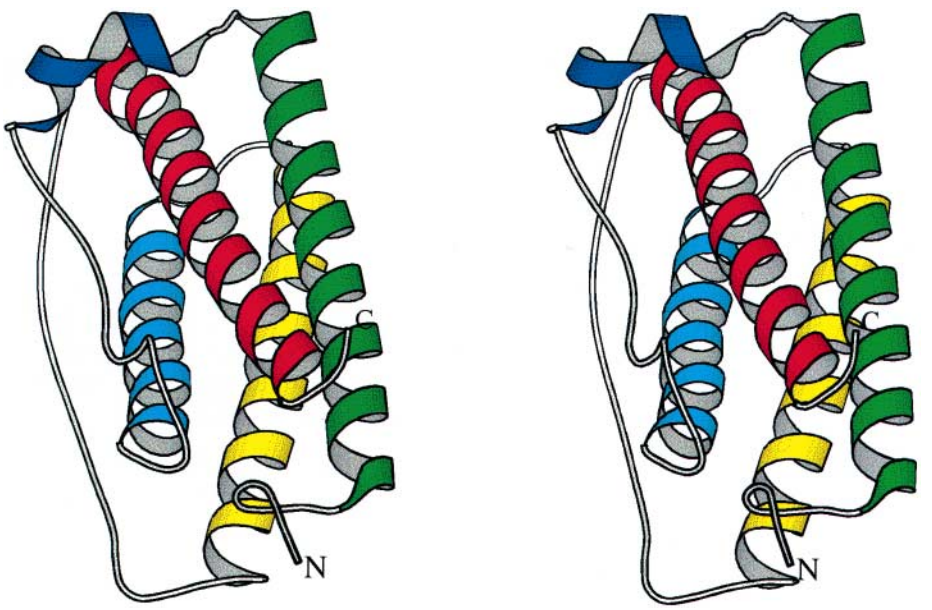

Fig. 2. Stereoscopic ribbon drawing of KW-2228. Helices A, B, C, D and E are shown in green, cyan, yellow, red and blue, respectively.

```
                              L31(A)---L103(C)
                                       |
V21(A)---F113(C)---I24(A)---L106(C)---L89(B)---V153(D)---L88(B)
          |          |          |
        L82(B)   L161(D) L157(D)


          L75(B)
           |
L71(B')---L9(A')---F13(A)---I117(C)
           |
          L124(C")


L18(A)---L168(D)

L35(A)---L99(C)---I95(B")
|
V151(D)

L41(A")---F144(D)

L61(AB)---V163(D)

L92(B)---F140(D')
```

Scheme 1. Leucine-rich hydrophobic network. A, B, C, D and AB in parentheses denote the A, B, C and D helices and AB loop, respectively. ′ and ″ indicate positions near the N- and C-termini, respectively: i.e., L9(A′) means that L9 is located near the N-terminus of helix A but not in helix A. Leucine residues are in bold type. Close contacts shorter than the sum of van der Waals radii are shown.
←

Cys^{74} disulfide bridge is buried in the interior of the molecule, but the Cys^{36}–Cys^{42} bridge is on the surface of the molecule. These disulfide bridges restrict the conformation of the long overhand AB connection.

There is a specific hydrophobic network in the interior of the bundle which is markedly rich in leucine residues. Out of 70 least solvent-accessible residues, approximately 33% are leucine residues. Many leucine residues together with some valine, isoleucine and phenylalanine residues constitute a network (hereafter referred to as a 'leucine-rich hydrophobic net-

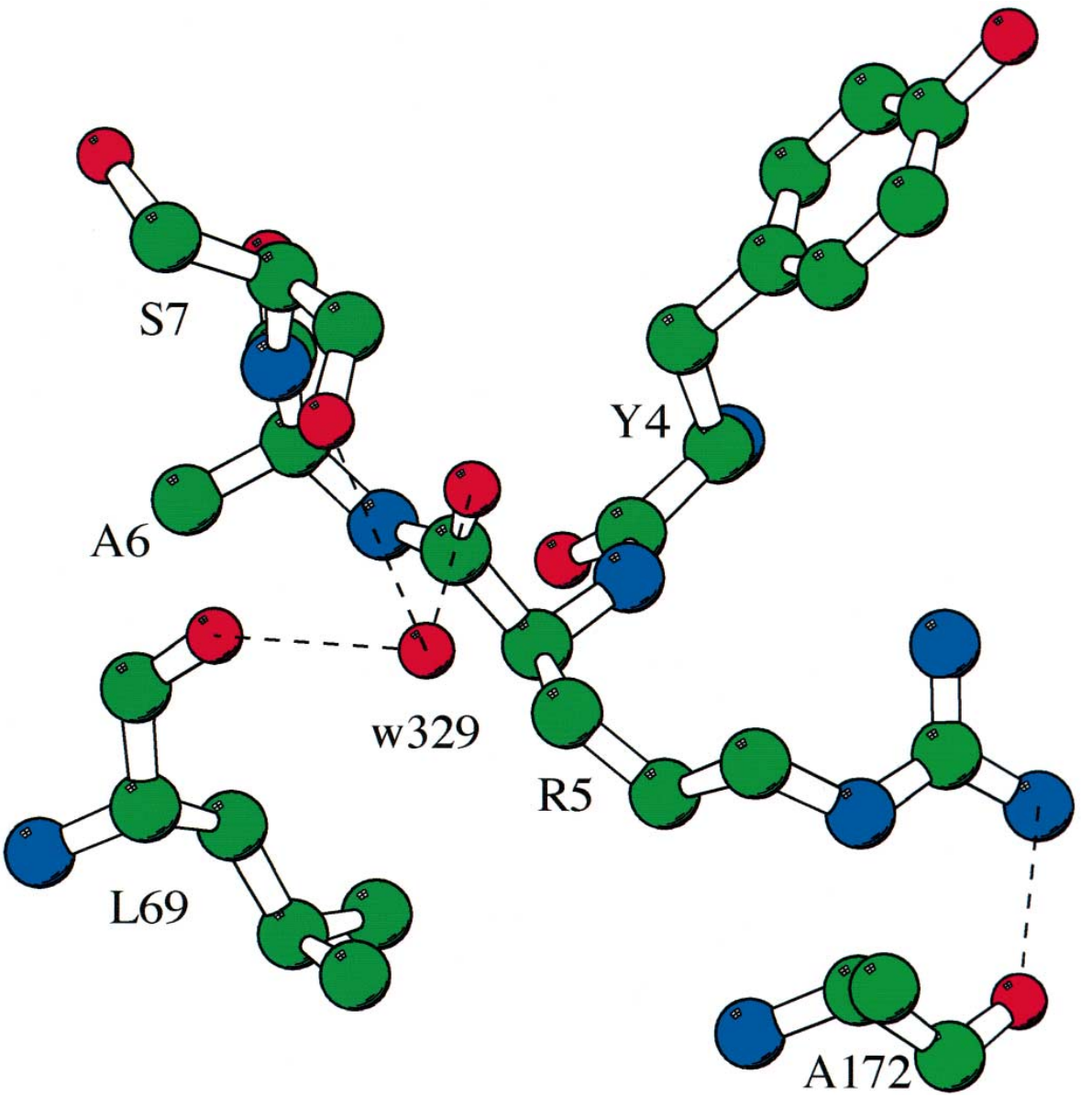

Fig. 3. Structure near the N-terminus displaying Y4, R5, A6,S7, L69, A172 and a water molecule (w329). Green, blue and red circles denote carbon, nitrogen and oxygen atoms, respectively. The broken lines indicate possible hydrogen bonds and an electrostatic interaction.

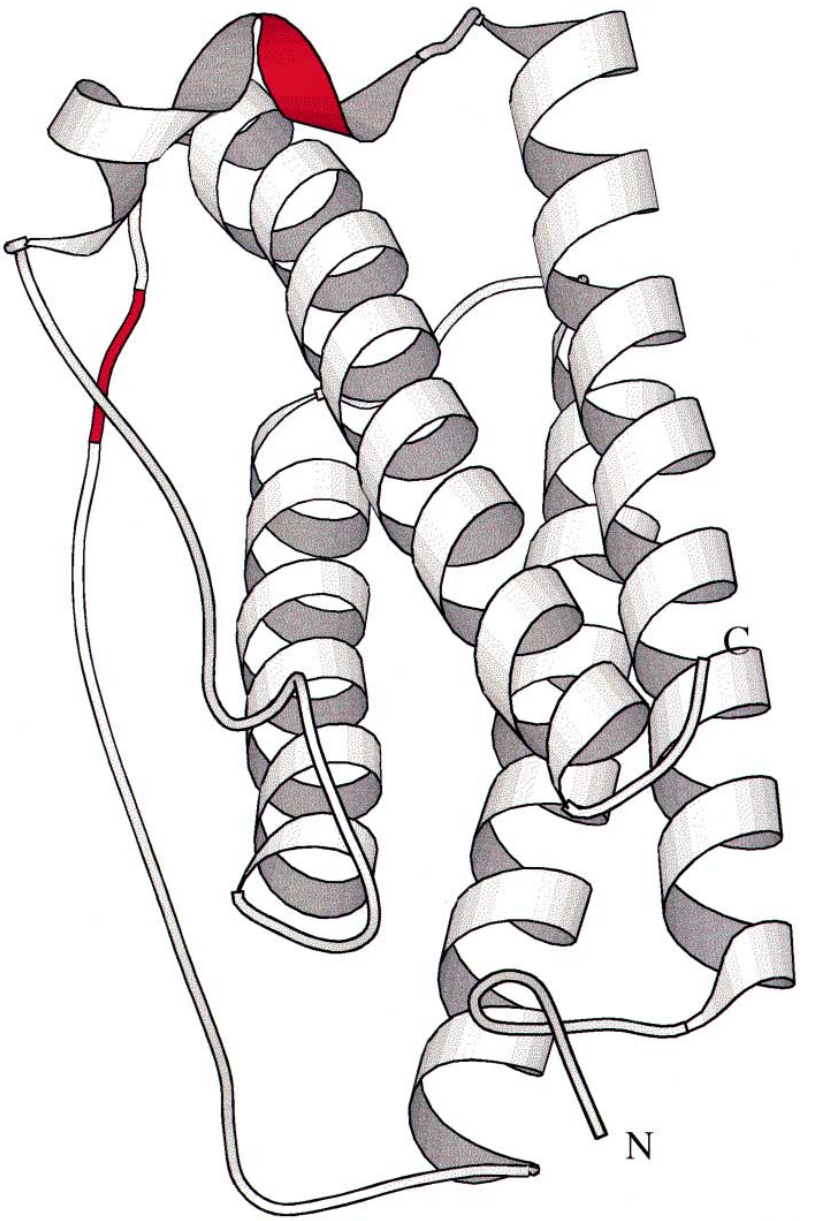

Fig. 4. Predicted receptor binding sites. Regions shown in red are recognized by different monoclonal anti-GCSF antibodies and exposed on the surface of G-CSF.

work') as shown in Scheme 1. Seventeen out of 19 leucine residues in the leucine-rich hydrophobic network are less solvent-accessible. The characteristic leucine-rich hydrophobic network obviously plays a very important role in stabilization of the four helices in the bundle configuration. In G-CSF there are particularly long overhand connections which are potentially flexible. Therefore the leucine-rich hydrophobic network may function as a scaffold to stabilize the bundle and adjust the positions of specific residues favorable for binding of receptors.

4. Discussion

The biological activity and stability of KW-2228 in human serum are markedly enhanced compared with those of wild-type hG-CSF [8,9]. The residues which are engineered in KW-2228 are located near the N-terminus. A water molecule designated w329 is located near the carbonyl oxygens of Ser^{7} and Leu^{69} with O···O distances of 3.11 and 3.10 Å, respectively (Fig. 3). Pro^{5} in wild-type hG-CSF is replaced by Arg in KW-2228 and w329 is located near the carbonyl oxygen of Arg^{5}. The distance between the oxygen atoms of w329 and Arg^{5} is 3.29 Å and it can be a weak hydrogen bond. The positively charged Nη2 atom of Arg^{5} is located near the carbonyl oxygen of Ala^{172} with an N···O distance of 4.08 Å and electrostatic interaction occurring between them. In wild-type hG-CSF the positions of seven residues in the N-terminal region were not determined unequivocally and the position of Pro^{5} is unknown. The structure of the N-terminal region of KW-2228, however, clearly suggests that neither the carbonyl nor the amino group of Pro^{5} in wild-type hG-CSF approaches the water molecule and no interactions occur be-

tween Pro^5 and Ala^{172}. The carbonyl group of Tyr^4 is hydrogen-bonded to the amino group of Ala^6 with an O···N distance of 2.90 Å. The main chain conformation of Tyr^4 could also be influenced significantly by the following residue. The specific network of interactions which involve Arg^5 contribute stabilization of the conformation of the N- and C-terminal regions and fix the middle of the long overhand AB connection. The structure of the sequence of residues 62–70 in wild-type hG-CSF could not be determined, but in KW-2228 they adopt clear conformations.

Cys^{17} is substituted by Ser in KW-2228. In wild-type hG-CSF Cys^{17} is located in the interior of the molecule and surrounded by hydrophobic residues but no specific interactions occur between Cys^{17} and surrounding residues. In KW-2228, however, the hydroxyl group of Ser^{17} forms a hydrogen bond with the carbonyl group of Leu^{14} (O···O distance of 3.10 Å). The substitution may contribute to the stability of the protein.

The present analysis has revealed that the mutations introduced in KW-2228 contribute to the stabilization of the conformation of the molecule including the two long overhand connections. The significant enhancement of the physicochemical and biological stability and biological activity of KW-2228 is undoubtedly due to its stabilized active conformation.

The binding of G-CSF to its receptor on pluripotent stem cells triggers terminal differentiation of the cells. Understanding the binding mechanism is a prerequisite to designing an improved G-CSF. First we must determine the regions of G-CSF that bind to the receptor. It is expected that residues that contribute to G-CSF receptor binding are located on the surface of G-CSF. In KW-2228 the particularly solvent-accessible residues are present in the AB and CD loops and the E short helix. Most of the solvent-accessible residues are located on the same side of the molecule. Around the E short helix 8 of the 13 residues 43–55 are highly solvent-accessible. In the CD loop 6 of the 10 residues 131–140 are also highly solvent-accessible. These 14 residues are on the same side of the molecule. Layton et al. prepared monoclonal anti-G-CSF antibodies and mapped the regions of G-CSF recognized by different antibody groups [20]. They predicted that residues 20–46 and the C-terminus bind to the G-CSF receptor. Residues 138–142 are recognized by common antibodies. Based on the structural and biological information obtained to date, we predict that residues 43–46 and 138–140 and/or those near them are involved in G-CSF receptor binding (Fig. 4).

Acknowledgements: We thank N. Yasuoka for valuable discussions.

References

[1] Klingemann, H.-G. (1989) Can. Med. Assoc. J. 140, 137–142.
[2] Souza, L.M., Boone, T.C., Gabriloe, 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. and Welte, K. (1986) Science 232, 61–65.
[3] Oh-eda, M., Hasegawa, M., Hattori, K., Kuboniwa, H., Kojima, T., Orita, T., Tomonou, K., Yamazaki, T. and Ochi, N. (1990) J. Biol. Chem. 265, 11432–11435.
[4] Nocola, N.A., Metcalf, D., Matsumoto, M. and Johnson, G.R. (1983) J. Biol. Chem. 258, 9017–9023.
[5] Nagata, S., Tsuchiya, M., Asano, S., Kaziro, Y., Yamazaki, T., Yamamoto, O., Hirata, Y., Kubota, N., Oheda, M., Nomura, H. and Ono, M. (1986) Nature 319, 415–418.
[6] Komatsu, Y., Matsumoto, T., Kuga, T., Nishi, T., Sekine, S., Saito, A., Okabe, M., Morimoto, M., Itoh, S., Okabe, T. and Takaku, F. (1987) Jpn. J. Cancer Res. 78, 1179–1181.
[7] Shimamura, M., Kobayashi, Y., Yuo, A., Urabe, A., Okabe, T., Komatsu, Y., Itoh, S. and Takaku, F. (1987) Blood 69, 353–355.
[8] Kuga, T., Komatsu, Y., Yamasaki, M., Sekine, S., Miyaji, H., Nishi, T., Sato, M., Yokoo, Y., Asano, M., Okabe, M., Morimoto, M. and Itoh, S. (1989) Biochem. Biophys. Res. Commun. 159, 103–111.
[9] Okabe, M., Asano, M., Kuga, T., Komatsu, Y., Yamasaki, M., Yokoo, Y., Itoh, S., Morimoto, M. and Oka, T. (1990) Blood 75, 1788–1793.
[10] Nagahara, Y., Konishi, N., Yokoo, Y. and Hirayama, N. (1990) J. Mol. Biol. 214, 25–26.
[11] Lovejoy, B., Cascio, D. and Eisenberg, D. (1993) J. Mol. Biol. 234, 640–653.
[12] Hill, C.P., Osslund, T.D. and Eisenberg, D. (1993) Proc. Natl. Acad. Sci. USA 90, 5167–5171.
[13] Sakabe, N. (1991) Nucl. Instrum. Methods Phys. Res. Ser. A 303, 448–463.
[14] Higashi, T. (1989) J. Appl. Crystallogr. 22, 9–18.
[15] Steigemann, W. (1989) A Program System for the Crystal Analysis of Proteins. Max-Planck Institut für Biochemie, Martensreid, Germany.
[16] Brunger, A.T., Kuriyan, J. and Karplus, M. (1987) Science 235, 458–460.
[17] CCP4 (1979) The SERC(UK) Collaborative Computing Project No. 4. A suite of programs for protein crystallography, distributed from Daresbury Laboratory, Warrington WA4 4AD, UK.
[18] Roussel, A. and Cambillau, C. (1991) Silicon Graphics Directory. Silicon Graphics, Mountain View, CA.
[19] Kraulis, P.J. (1991) J. Appl. Crystallogr. 24, 946–950.
[20] Layton, J.E., Morstyn, G., Fabri, L.J., Reid, G.E., Burgess, A.W., Simpson, R.J. and Nice, E.C. (1991) J. Biol. Chem. 266, 23815–23823.