Journal of Computer-Aided Molecular Design, **12:** 419–424, 1998.
KLUWER/ESCOM

Homology modeling of the receptor binding domain of human thrombopoietin

Jin-Soo Song, Heungrok Park, Hyo-Jeong Hong, Myeong-Hee Yu & Seong-Eon Ryu*
Protein Engineering Research Division, Korea Research Institute of Bioscience & Biotechnology, KIST, P.O. Box 115, Yusong, Taejon 305-600, South Korea

Received 12 December 1997; Accepted 6 March 1998

Key words: comparative modeling, cytokine, four-helix bundle, platelet, receptor-ligand interaction

Summary

Platelet production in blood is regulated by a lineage specific humoral factor, thrombopoietin (TPO). The amino terminal domain of TPO (TPO-N) is responsible for the signal transduction mediated by the TPO receptor, c-mpl. From the predicted length of helices we found that TPO-N belongs to the long-chain subfamily of the four-helix bundle cytokine family. We built a three dimensional model of TPO-N by a comparative homology modeling procedure. The four helices of TPO-N with an up-up-down-down topology are stabilized by a tightly packed central hydrophobic core and the extended loop AB makes an additional hydrophobic core with helices B and D outside of the four helix bundle scaffold. An interpretation of the previous site directed mutageneses results in light of the model enabled us to identify two isolated receptor binding sites. The surface made of Lys 136, Lys 138 and Lys 140 in helix D, and Pro 42 and Glu 50 in loop AB forms the first receptor binding site, while the surface of Asp 8, Arg 10 and Lys14 in helix A represents the second binding site for the sequential receptor oligomerization.

Introduction

Thrombopoietin (TPO), a recently characterized hematopoietic growth factor, promotes the proliferation of megakaryocytes and the production of platelets in blood [1–3]. The cDNA cloning of the gene revealed that TPO has a characteristic two domain structure with a molecular weight of 35 KDal [4–6]. The amino-terminal domain of TPO, comprised of about 160 amino acids, has a sequence homology to erythropoietin (EPO) and is thought to be a member of the four-helix bundle cytokine family. The rest of the molecule forms the carboxy-terminal domain with six *N*-linked glycosylation sites. The amino-terminal domain has been shown to be responsible for the activation of TPO receptor, c-mpl [4].

The proteins with the four-helix bundle structure are found in many cytokines such as human growth hormone, various interleukines and interferons [7]. They can be divided into three subfamilies according to the length of the helices [8, 9]. The short-chain subfamily consists of relatively short helices (10–20 amino acids) and two β-strands. The helices of the long-chain subfamily proteins are longer (20–30 amino acids) than those of the short-chain subfamily. The interferon-like subfamily, which is named after a member of the family, has an extra helix between helices C and D. Despite the structural similarity, the members of the four-helix bundle cytokine family have very low sequence homology. However, these proteins are thought to have an evolutionary relationships since their roles in cellular communication are similar and they share a similar exon structure [9].

The information on the three dimensional structure of TPO would be important in understanding the atomic interactions of TPO with its receptor as well as the mechanism of the receptor activation. However, no three dimensional structure of TPO has been reported yet. In this study, we constructed a three dimensional model of the amino-terminal domain of TPO (TPO-N) based on its structural similarity to the four-helix bundle cytokine family. The model of TPO-N presented

*To whom correspondence should be addressed.

here provides a basis for understanding molecular interactions of TPO with its receptor and future design of therapeutically useful variants of the molecule.

Secondary structure prediction

The amino-terminal domain (residues 1–156) of human TPO was subjected to the various secondary structure prediction methods. The Holly-Karplus method [10], the program PHD [11] and the program NNPREDICT [12] were used to obtain the most probable boundaries of the helices in TPO-N. The secondary structure prediction of TPO-N by three different prediction methods yielded a consistent pattern of α-helices in the amino acid sequence. We identified the most probable α-helices by taking the regions predicted to be α-helical in at least two of the three prediction methods. Two of helices were predicted to be significantly longer than 20 amino acids. The length of these helices suggested to us that TPO-N belongs to the long-chain subfamily of the four-helix bundle cytokine family.

Model building and energy minimization

The initial model was constructed by the comparative modeling program MODELLER [13] and manually adjusted by the program QUANTA (Molecular Simulations Inc.) running on a Silicon Graphics Indigo II workstation. To obtain a structural framework for the comparative modeling, we chose three members of the long-chain subfamily among the four-helix bundle cytokines whose three dimensional structures were determined experimentally. By using the protein data bank (PDB) coordinates of the three cytokines (human growth hormone (HGH): 3hhr; granulocyte colony-stimulating factor (G-CSF): 1rhg; and leukemia inhibitory factor (LIF): 1lki), we aligned corresponding helices by using the program TOSS (Hendrickson, W. A., personal communication).

The helical regions of TPO-N identified from the secondary structure prediction were aligned with the framework sequences before running MODELLER (Figure 1). The sequence alignment of TPO-N was manually optimized to match the pattern of hydrophobic residues in the helical regions of TPO-N with that of the framework cytokines. From the model obtained from MODELLER, the register of the helices were adjusted to satisfy the constraint due to the length of connecting loops between helices and the disulfide bonds. Based on the adjusted model, the sequences were aligned again and MODELLER was run with the new sequence alignment. This process was repeated until a satisfactory model of the helices was obtained.

The loops connecting the helices were built to fit the overall context of the molecule mainly by following the loop conformations of the framework cytokine structures with similar loop lengths. There are variations in the length of loops between members of the long-chain cytokine family. The conformation of loop AB was taken from the structures of G-CSF and LIF. Loop CD of TPO-N was modeled following the structure of G-CSF, while Loop BC was an average of the structures of HGH, G-CSF and LIF. After building the connecting loops, the entire model was inspected by using the program QUANTA to adjust unrealistic conformations of side chains and to remove bad contacts.

The energy minimization of the model was carried out by using the program X-PLOR [14]. The empirical energy function included conformation energy terms, non-bonding energy terms with 7.5 Å cutoff, and electrostatic energy terms including partial charge terms to account for the hydrogen bonding energy. The position of polar hydrogens was generated from the heavy atom coordinates by using standard bond lengths and angles, and used in the energy minimization with the empirical force field. The 300 cycles of the Powell conjugate gradient minimization were run at a time. The model was inspected by QUANTA after a minimization and the regions with bad contacts and unfavorable main chain torsion angles were optimized. Several rounds of the minimization and optimization yielded a tightly-packed model with a good geometry.

Overall fold of TPO-N

The four helices of TPO-N are arranged with an up-up-down-down topology as found in other four helix bundle cytokines. The length of the helices of TPO-N ranges from 21 to 23 amino acids, categorizing TPO-N as a member of the long chain subfamily of the four helix bundle cytokines (Table 1). The central hydrophobic core of TPO-N is mainly made of the residues in the hydrophobic face of the amphiphilic helices. Leu 12, Leu 15, Leu 16, Leu 21, Leu 26 of helix A; Ile 63, Val 67 and Val 74 of helix B; Leu 86, Leu 89, Leu 93, Leu 100 and Leu 104 of helix C; Phe 128, Leu 135 and Val 139 of helix D form

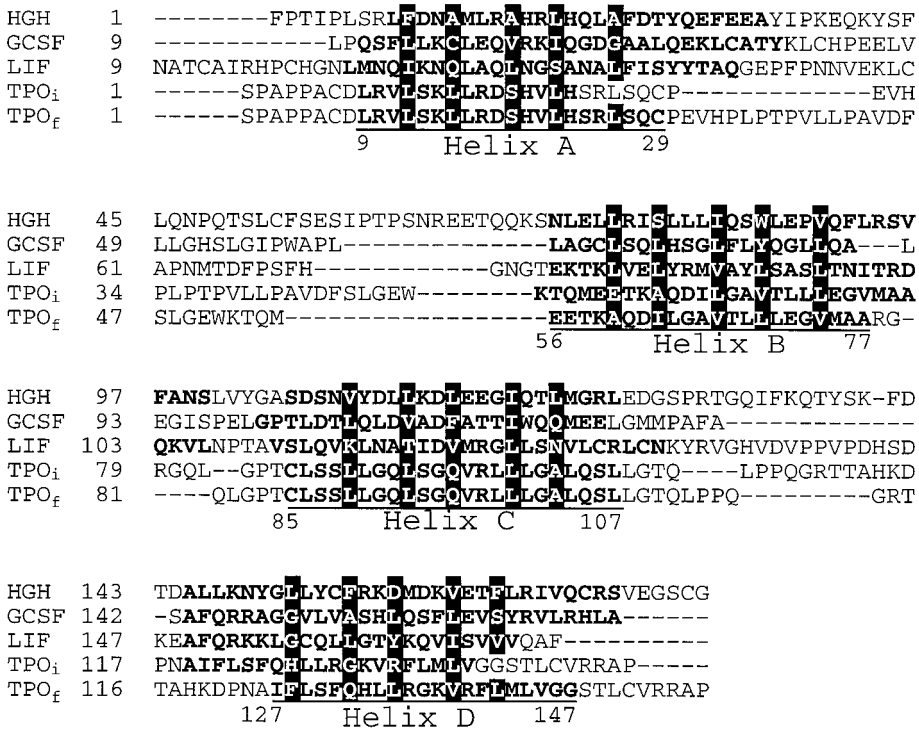

Figure 1. Sequence alignment of TPO-N with members of the long-chain subfamily. Sequences of the framework cytokines (HGH, G-CSF and LIF) were aligned by superposing C_α-carbons of the structures from the Brookhaven protein data bank. The residues in the helices of each structure are presented as bold characters. When only sequence information was used in the sequence alignment with the program BESTFIT (GCG package) the cytokine members yielded low sequence identity scores (10–20%) even though the sequence similarity scores were relatively high (50–70%). The similar scores were found in the sequence alignment between different subfamily members of the four-helix bundle cytokine family. The highest scores in the alignment of TPO-N with the framework cytokines were 17% identity with G-CSF and 66% similarity with LIF. However, the sequence alignment between the framework cytokines did not coincide with the structural alignment and showed gaps in the helical regions. To overcome this difficulty, we used the structural information in the alignment. The structurally aligned framework cytokines as presented in the figure has lower sequence homology scores (10% identity/48% similarity for HGH:G-CSF; 9% identity/52% similarity for HGH:LIF; 11% identity/50% similarity for G-CSF:LIF) than those of the alignment to maximize sequence homology. We determined the helix boundary of TPO-N from the secondary structure analysis (see text) and aligned the helical regions with those of the structurally aligned framework cytokines (TPO_i). The alignment and helix boundary of TPO-N were manually adjusted to optimize the conformation of connecting loops and disulfide bonds during the modeling process. The final alignment of TPO-N (TPO_f), which produced the most reliable model, has sequence homology scores (12% identity/48% similarity with HGH; 11% identity/49% similarity with G-CSF; 9% identity/48% similarity with LIF) comparable to those of the structurally aligned framework cytokines. The helical regions of TPO-N in the final model are underlined. The residues in the hydrophobic side of the amphiphilic helices are indicated as black boxes.

the tightly packed central hydrophobic core of TPO-N. The side chain of a hydrophilic residue (Gln 132) is buried in the hydrophobic core of TPO-N. The NE2 atom of Gln 132 interacts with the carbonyl oxygen of Leu 26 to stabilize the four helix bundle scaffold. The polar interactions in the hydrophobic core stabilizing the overall fold are often found in other cytokines [8].

Besides the central hydrophobic core, TPO-N has a second hydrophobic core between loop AB and helices B and D that is located outside of the four helix bundle scaffold but buried by loop AB. The second hydrophobic core comprises Phe 46 and Leu 48 of loop AB, Ile 127, Phe 131 and Leu 134 of helix D, and Leu 69 and Leu 70 of helix B. The tight hydrophobic interactions between these residues hold the extended loop AB (28 amino acids) to the main body of the four helix bundle. This hydrophobic core is extended upward following the direction of loop AB and Trp 51 of loop AB makes another tight hydrophobic interaction with Phe 141 and Leu 142 of helix D. The second hydrophobic core in a similar location in HGH consists of Phe 44 and Leu 45 of loop AB, and Tyr 160 and Tyr 164 of helix D. In the case of HGH, an extra disulfide bond between Cys 53 of loop AB and Cys 165 of helix D contribute another stabilizing force

Table 1. Helix length of various four-helix bundle cytokines.

Proteins	Length	Helix A	Helix B	Helix C	Helix D	PDB	
Long[1]	HGH	190	26(L9-A34)	21(N72-F92)	23(S106-L128)	30(A155-S184)	3hhr
	G-CSF	174	29(Q11-Y39)	21(L71-A91)	24(G100-E123)	30(A143-A172)	1rhg
	LIF	180	27(L22-Q48)	29(E76-L104)	27(V109-N135)	23(A155-V177)	1lki
	TPO-N	156	21(L9-C29)	22(E56-A77)	23(C85-L107)	21(I127-G147)	this study
Short[2]	IL-4	129	14(15-T18)	18(E41-H58)	25(A70-A94)	18(L109-K126)	2int
	GM-CSF	127	16(W13-L28)	10(L55-Q64)	14(K74-H87)	14(F103-V116)	1gmf
	M-CSF	158	12(S13-D24)	18(P46-D63)	17(P72-K88)	21(P110-K130)	1hmc
IFN[3]	IFN-β	158	18(L6-Q23)	17(S49-V65)	15(E77-T91)	21(S135-L155) E20(S112-M131)	1ifa
	IFN-α	138	12(Y4-F15)	22(E39-F60)	16(Q67-F82)	17(L103-E119) E11(K86-T96)	1hig

The boundary and length of helices in the representative members of each subfamily are presented. All data except for those of TPO-N were obtained from the coordinates of the Brookhaven protein data bank. The interferon-like subfamily has an extra helix that is denoted as E in front of the helix length in the table. The comparison of helix length indicates that TPO-N is a member of the long-chain subfamily as human growth hormone (HGH), granulocyte colony-stimulating factor (G-CSF) and leukemia inhibitory factor (LIF). 1, the long-chain subfamily; 2, the short-chain subfamily; 3, the interferon-like subfamily; IL-4, interleukin-4; GM-CSF, granulocyte-macrophage colony-stimulating factor; M-CSF, macrophage colony-stimulating factor; IFN-β, interferon-β; IFN-α, interferon-α.

to hold the extended loop AB (39 amino acids) to the four helical body.

The Ramachandran plot drawn from the final model by the program PROCHECK [15] shows that there is no residue in the disallowed regions. The program GRASP [16] was used for the calculation of cavities inside of the model to verify the structural integrity. The residues in the hydrophobic core of the final model are very well packed and there is no cavity inside of the molecule when the possible cavities were calculated by GRASP using either Van der Waals surfaces or solvent accessible surfaces with a probe radius of 1.4 Å.

Modeling of disulfide bonds

TPO-N has four cysteines at residues 7, 29, 85 and 151. Among the four cysteines, three (residues 7, 29 and 151) are conserved in EPO. The cysteines of EPO equivalent to the first (Cys 7) and the last (Cys 151) cysteines of TPO-N were demonstrated to form a disulfide bond [17]. By a site directed mutagenesis study, Wada et al. [18] showed that the first and the last cysteines are essential for the activity of TPO. The mutation at the second cysteine (Cys 29) of TPO also exhibited complete loss of the activity. During our modeling process we found that the second and the third cysteines (Cys 29 and Cys 85) became close to each other and had a potential to form the second disulfide bond. Consequently, we built the second disulfide bond between Cys 29 and Cys 85 besides the conserved disulfide bond between Cys 7 and Cys 151. The position of the disulfide bonds and the constraint in the length of the loops between the helices helped us to find an optimal alignment of the helices by sliding each helix relative to the others.

The receptor interaction sites of TPO-N

Responses to cytokines are mediated by receptor oligomerizations. Among the members of the long-chain subfamily, IL-6 [19] and LIF [20] transfer the signal by a heterodimerization of different receptor chains, while responses to HGH [21], prolactin [22] and G-CSF [23] are transduced by a homodimerization of two identical receptor molecules. There are indications that the signal transduction by EPO, which has a relatively high homology to TPO-N, is mediated by a receptor homodimerization [24]. We analyzed the putative TPO-receptor interactions in analogy to the growth hormone-receptor and EPO-receptor interactions. HGH [21] and EPO [25] have two separate regions that interact with two identical receptor molecules. The interaction site I resides in helix D and loop AB, as does the interaction site II in helix A. The major parts of the receptor-ligand interactions in both growth hormone and EPO occur with charged residues in helices A and D of the cytokines. Arg 167, Lys 168 and

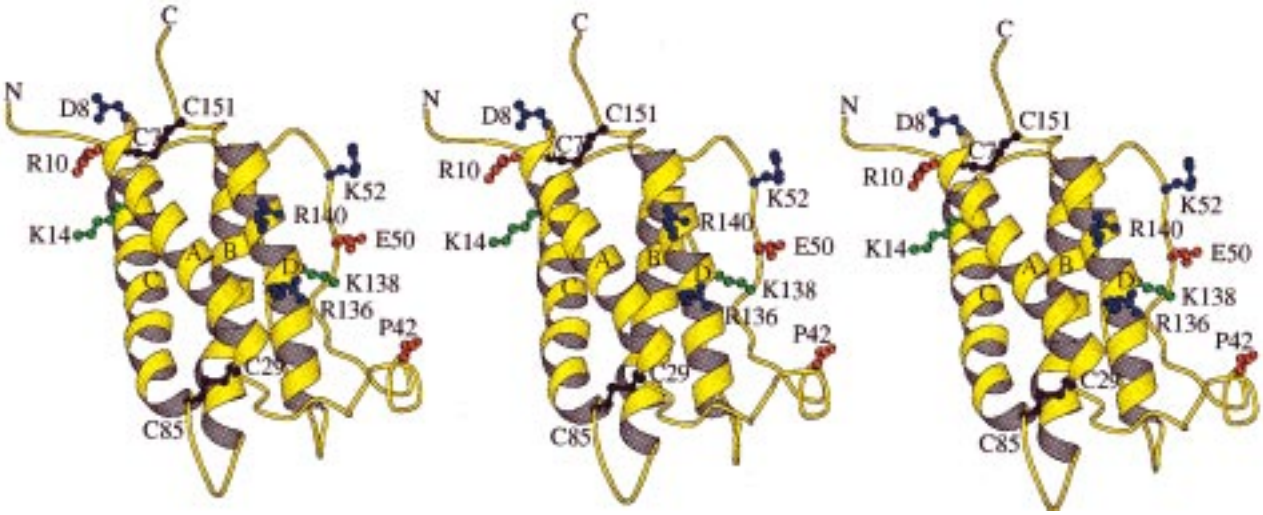

Figure 2. Structure of TPO-N. The two disulfide bonds (Cys 7: Cys 157 and Cys 29: Cys 85) and the side chains of the residues implicated in the receptor interactions are drawn on the schematic ribbon representation of the structure of TPO-N in stereo. The residues found to affect the TPO activity by both the receptor binding affinity assay and the cell proliferation assay are represented in green. The residues identified by one of the two methods are colored blue (the receptor binding affinity assay) and red (the cell proliferation assay). The residues Phe 46 and Phe 128, whose mutational effects are likely to be structural, are not shown for the clarity of the figure. The diagram was drawn using the program MOLSCRIPT [29].

Asp 171 in helix D of HGH, and Arg 143, Arg 150 and Lys 154 in helix D of EPO play major roles in site I, while Arg 16 and Arg 19 in helix A of HGH, and Arg 10 and Arg 14 in helix A of EPO does in site II.

TPO-N has many charged residues on the surface of helices A and D. These include Arg 10, Lys 14, Arg 17, Asp 18, His 20, His 23 and Arg 25 of helix A and His 133, Arg 136, Lys 138 and Arg 140 of helix D. Recently, Pearce *et al.* [26] reported the results of an alanine scanning mutagenesis study that used a receptor-binding assay to define the regions of TPO involved in the activity. In that study, D8A and K138A mutations decreased the activity of TPO by more than 20 fold, and K14A, K52A, R136A and R140A did by 5 to 10 fold. Figure 2 shows the location of these residues in our model. Another mutagenesis study based on a cell proliferation assay identified seven residues whose alanine substitutions significantly reduced the biological activity of TPO [27]. Four of these residues are Arg 10 and Lys 14 in helix A, and Phe 128 and Lys 138 in helix D. The other three are Pro 42, Phe 46 and Glu 50 that are located in loop AB in our model (Figure 2). Among these, Phe 128 is buried in the central hydrophobic core made of four helices and Phe 46 is in the secondary hydrophobic core between loop AB and helices B and D. Thus, the mutational effects for Phe 46 and Phe 128 are likely due to the disruption of the tertiary structure of TPO-N.

Although some of the mutational effects obtained by using the two different assay systems do not match each other [27], the two results seem to complement each other to a certain extent when we consider only the residues with the most pronounced mutational effects as discussed here. In analogy to EPO and human growth hormone, the residues in helix D (Arg 136, Lys 138 and Arg 140) and loop AB (Pro 42, Glu 50 and Lys 52) form one receptor binding surface (site I) and the residues in helix A (Asp 8, Arg 10 and Lys 14) form the other binding surface (site II). Interestingly, the side chain of Glu 50 is close to that of Lys 138 in the model and these two residues are likely to form a salt bridge. The charge interaction between the two residues may be important for maintaining the local conformation of the structure for the interaction with the receptor.

Approximate dimension of the two binding surfaces derived from the location of the residues implicated in the receptor interaction are 25 Å $\times$ 20 Å and 15 Å $\times$ 8 Å for site I and site II, respectively. In the interaction of HGH with its receptors, site I is wider than site II and the first receptor chain binds to site I with high affinity and the next receptor chain is recruited to bind to site II with low affinity [21, 28]. The binding of the second receptor chain is assisted by

extra interactions between the two receptor chains. It is plausible that TPO has a similar sequential mode of receptor interaction. The first receptor chain would come to interact with the wide concave surface composed of Pro 42, Glu 50, Lys 52, Arg 136, Lys 138 and Arg 140 with high affinity. This TPO:receptor complex would recruit the next receptor molecule to form the signal transducing TPO:$(receptor)_2$ complex by interactions between the smaller surface comprising Asp 8, Arg 10 and Lys 14, and the second receptor as well as between the two receptor chains.

The coordinates of the final model will be deposited with the Brookhaven Protein Data Bank.

Acknowledgements

This work was supported by grants from the Ministry of Science and Technology, Korea.

References

1. Gordon, M.S. and Hoffman, R., Blood, 80 (1992) 302.
2. Shreiner, D.P., Weinberg, J. and Enoch, D., Blood, 56 (1980) 183.
3. Tsukada, J., Misago, M., Ogawa, R., Oda, S., Morimoto, I., Eto, S. and Kikuchi, M., Blood, 81 (1993) 866.
4. de Sauvage, F.J., Hass, P.E., Spencer, S.D., Malloy, B.E., Gurney, A.L., Spencer, S.A., Darboune, W.C., Henzel, W.J., Wong, S.C., Kuang, W.-J., Oles, K.J., Hultgren, B., Solberg, L.A. Jr., Goeddel, D.V. and Eaton, L.D., Nature, 369 (1994) 533.
5. Kaushansky, K., Lok, S., Holly, R.D., Broudy, V.C., Lin, N., Bailey, M.C., Forstrom, J.W., Buddle, M.M., Oort, P.J., Hagen, F.S., Roth, G.J., Papayannopoulous, T. and Foster, D.C., Nature, 369 (1994) 568.
6. Wendling, F., Maraskovsky, E., Debill, N., Florindo, C., Teepe, M., Titeux, M., Methia, N., Breton-Gorius, J., Cosman, D. and Vainchenker, W., Nature, 369 (1994) 571.
7. Vigon, I., Mornon, J.-P., Cocault, L., Mitjiavila, M.-T., Tambourin, P., Gisselbrecht, S. and Souyri, M., Proc. Natl. Acad. Sci. USA, 89 (1992) 5640.
8. Rozwarski, D.A., Gronenborn, A.M., Clore, G.M., Bazan, J.F., Bohm, A., Wlodawer, A., Hatada, M. and Karplus, P.A., Structure, 2 (1994) 159.
9. Bazan, J.F., Proc. Natl. Acad. Sci. USA, 87 (1990) 6934.
10. Holley, L.H. and Karplus, M., Proc. Natl. Acad. Sci. USA, 86 (1989) 152.
11. Sander, C. and Schneider, R., Proteins, 9 (1991) 56.
12. Kneller, D.G., Cohen, F.E. and Langridge, R., J. Mol. Biol., 214 (1990) 171.
13. Sali, A. and Blundell, T.L., J. Mol. Biol., 234 (1993) 779.
14. Brunger, A.T., Kuriyan, J. and Karplus, M., Science, 235 (1987) 458.
15. Morris, A.L., MacArther, M.W. and Thornton, J.M., Proteins, 12 (1992) 345.
16. Nicholls, A., Sharp, K. and Honig, B., Proteins, 11 (1991) 281.
17. Wang, F.F., Kung, C.-F. and Goldwasser, E., Endocrinology, 116 (1985) 2286.
18. Wada, T., Nagata, Y., Nagashisa, H., Okunomi, K., Ha, S.H., Ohnuki, T., Kanaya, T., Matsumura, M. and Todokoro, K., Biochem. Biophys. Res. Commun., 213 (1995) 1091.
19. Hibi, M., Murakami, M., Saito, M., Hirano, T., Taga, T. and Kishimoto, T., Cell, 63 (1990) 1149.
20. Robinson, R.C., Grey, L.M., Staunton, D., Vankelecom, H., Vernallis, A.B., Moreau, J.-F., Stuart, D.I., Heath, J.K. and Jones, E.Y., Cell, 77 (1994) 1101.
21. de Vos, A.M., Ultsch, M. and Kossiakof, A.A., Science, 255 (1992) 306.
22. Elberg, G., Kelly, P.A., Djiane, J., Binder, L., Gertler, A., J. Biol. Chem., 265 (1990) 14770.
23. Fuknunaga, R., Ishizaka-Ikeda, E. and Nagata, S., J. Biol. Chem., 265 (1990) 14008.
24. Watowich, S.S., Hilton, D.J. and Lodish, H.F., Mol. Cell. Biol., 14 (1994) 3535.
25. Caravella, J.A., Lyne, P.D. and Richards, W.G., Proteins, 24 (1996) 394.
26. Pearce, K.H. Jr., Potts, B.J., Presta, L., Bald, L.N., Fendly, B.M. and Wells, J.A., J. Biol. Chem., 272 (1997) 20595.
27. Park, H., Park, S., Song, J.-S., Jin, E.-H., Ryu, S.-E., Yu, M.-H. and Hong, H.-J., J. Biol. Chem., 273 (1998) 256.
28. Clackson, T. and Wells, J.A., Science, 267 (1995) 383.
29. Kraulis, P.J., J. Appl. Crystallogr., 24 (1991) 946.