FEBS Letters 404 (1997) 234–240 FEBS 18260

Tertiary structural models for human interleukin-6 and evaluation by a sequence-structure compatibility method and NMR experimental information

Hiromi Sumikawa[a], Ken-ichi Fukuhara[a], Ei-ichiro Suzuki[a,*], Yo Matsuo[1,b], Ken Nishikawa[2,b]

[a]*Central Research Laboratories, Ajinomoto Co., Inc., 1-1 Suzuki-cho, Kawasaki-ku, Kawasaki 210, Japan*
[b]*Protein Engineering Reseach Institute, 6-2-3, Furuedai, Suita, Osaka 565, Japan*

Received 25 November 1996; revised version received 22 January 1997

Abstract Tertiary structure models of Interleukin-6 were constructed using a routine prediction method based on the X-ray crystal structures of granulocyte colony-stimulating factor (GCSF) and leukemia inhibitory factor (LIF). The models were evaluated with the aid of the sequence-structure compatibility (3D-1D) method program *compass* and NMR experimental information. The model constructed from GCSF gained higher scores on *compass* examination than did that from LIF, and the NOE data [Nishimura et al. (1996) Biochemistry 35, 273–281] also turned to be more consistent with the former model.

Key words: Structure prediction; Interleukin-6; Granulocyte colony-stimulating factor; Leukemia inhibitory factor; 3D-1D method; Nuclear Overhauser effect

1. Introduction

In modern molecular biology and related sciences and technologies, it is a matter of key importance to predict the three-dimensional structure of a protein which is precise enough to be used as a working hypothesis for studying structure-function relationships, and it should be carried out as soon as its one-dimensional sequence has been determined. We have been trying to develop a strategic technique which is sufficiently able to distinguish a better stereo model from the others and which is easy to use for a wide range of proteins.

Interleukin-6 (IL-6) was chosen as a test case in our present study. IL-6 is a soluble protein factor possessing 184 amino acids, which was cloned as a B cell stimulatory factor which induces the final stage of differentiation of B cells to antibody producing cells [1]. Subsequently, it was demonstrated that IL-6 is a multifunctional cytokine which affects not only immune systems but also hematopoietic systems, nervous systems, endocrines, etc. [2]. In particular, recent studies using transgenic and knockout mice suggest that IL-6 plays an important role in various defense systems during the acute phase and its overproduction is concerned with various diseases and a morbid state [3–5]. As IL-6 acts in conjunction with two kinds of receptors for gp80 and gp130 on the target cell surface, knowledge of its three-dimensional structure is of key significance in the development of IL-6 agonists and antagonists. It has been proven, e.g. from CD results, that the secondary structure of IL-6 is all α-helical and it was postulated to possess four helices bundled in an up-up-down-down type of topology [6]. Many cytokines have been demonstrated to adopt this folding, however, there is little sequence identity with IL-6.

Therefore, tertiary structure prediction for IL-6 cannot be achieved with the conventional homology modeling method based on amino acid sequence information alone. Consequently, reference proteins (to be used for modeling) should be chosen for similarity of their biological activity to and structural compatibility with IL-6, and finally two reference proteins were chosen among PDB coordinate sets. It has been proposed, based on cysteine positional conservation and similar biological activity, that the three-dimensional structure of IL-6 closely resembles that of GCSF (granulocyte colony-stimulating factor). Accordingly, rat IL-6 was modeled by Hammacher et al. [7] after human GCSF, and human IL-6 was modeled by Ehlers et al. [8] and Savino et al. [9] after human and bovine GCSF, respectively. Both of these reference GCSF only. We propose here that not only GCSF but also LIF (leukemia inhibitory factor) can be used as a reference protein for IL-6, after searching PDB with the sequence-structure compatibility method (3D/1D) program *compass* [10,11]. In this study, two IL-6 models (one based on GCSF and the other according to LIF) are being constructed and evaluation of these models is being carried out using NMR experimental information (NOE data), a 3D/1D method, and energy calculations.

2. Methods

The alignment shown in Fig. 1 was used. The procedure for our homology modeling is shown in Fig. 2. The software package *BIOCES[E]* was used for the modeling. The modeled structures were evaluated with the 3D/1D method *compass*, which is equipped with a structural library containing 617 structures. They were selected from a total of 3451 proteins in Protein Data Bank release 72 (April, 1995), such that the sequence identity to one another was less than 50% [11]. The *compass* search was performed by adding the two models to the structural library. For optimization calculation, *Kopt* with AMBER's parameter and *CHARMm* with the CHARM parameter were used. *Aro* and *kai* were employed in order to superimpose two proteins. *PROCHECK* [12] was used for checking the accuracy of the coordinates.

*Corresponding author. Fax: (81) (44) 211-7609.

[1]*Present address:* Institute for Social Information Science, Fujitu Laboratories Ltd., 1-9-3 Nakase, Mihama-ku, Chiba 261, Japan.

[2]*Present address:* National Institute of Genetics, 1111, Yata, Mishima, Shizuoka 411, Japan.

Abbreviations: IL-6, interleukin-6; NOE, nuclear Overhauser effect; GCSF, granulocyte colony-stimulating factor; LIF, leukemia inhibitory factor

PII S0014-5793(97)00115-4

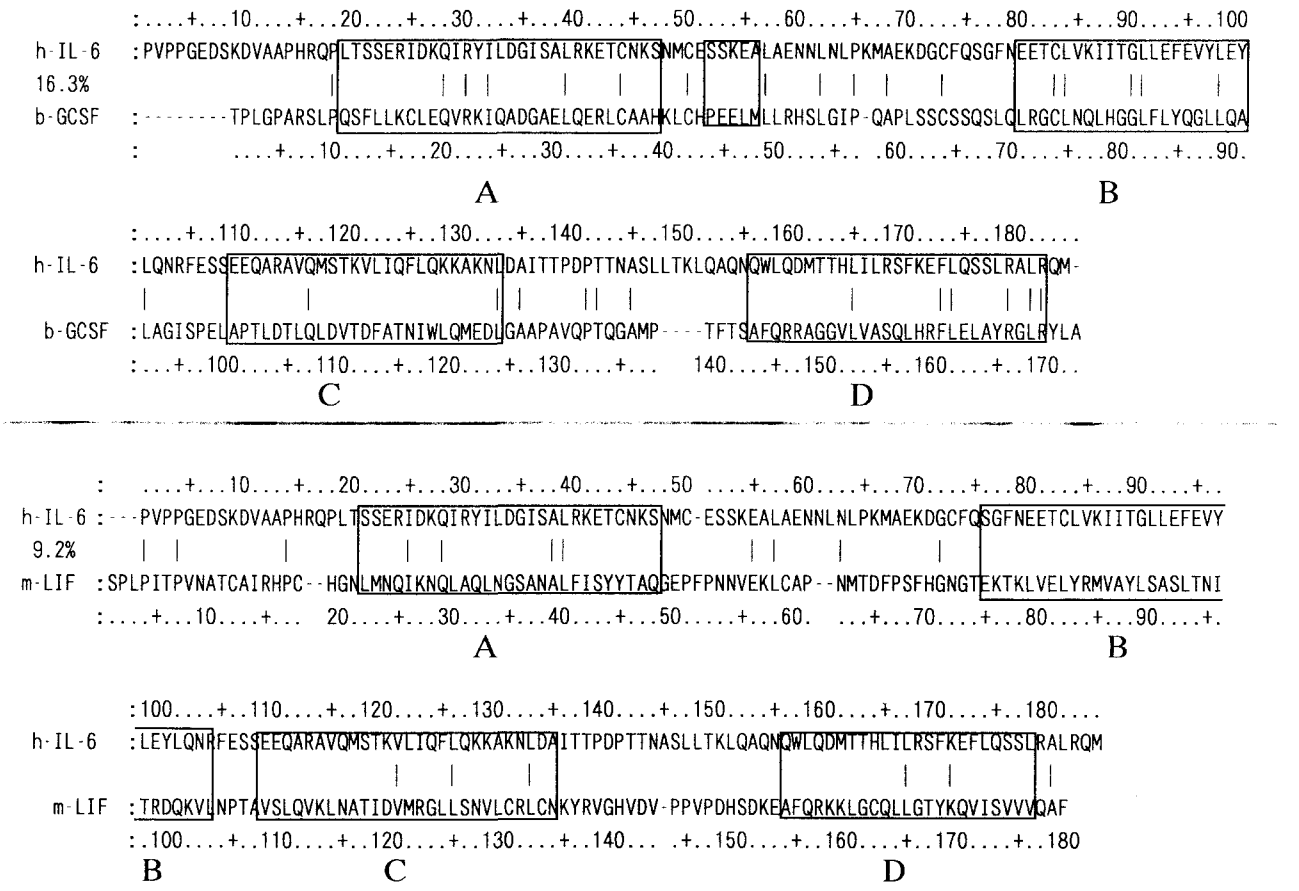

Fig. 1. Sequence alignment for human IL-6 with 184 residues. The single-letter amino acid code has been used. Gaps in the sequences due to the alignment are represented (–). Helical regions are IBGC 18–46 (helix A), 79–100 (helix B), 109–133 (helix C), 156–182 (helix D), 51–55 (helix E). ILKI 21–48 (helix A), 76–104 (helix B), 109–135 (helix C), and 156–178 (helix D). The indicated percentage shows the sequence homology calculated from the number of residues in h-IL-6 that are the same as those in b-GCSF or m-LIF. (Upper) Amino acid sequence of human IL-6 and bovine GCSF; (lower) amino acid sequence of human IL-6 and mouse LIF.

3. Results and discussion

3.1. Choice of reference proteins

Bovine GCSF was chosen, as it was judged from the results with *compass* that its tertiary structure has the best compatibility with human IL-6 as an all α-helical protein. The scores are listed in Table 1, in which the 10 best proteins are listed together with the IL-6 models, IBGC and ILKI. Also shown in Table 1 is one of the reference proteins, 1LKI, which was ranked 100th with a low compatibility score in magnitude. The other reference protein, 1BGC, was found within the best 10. Since IL-6 has been demonstrated, e.g. by using CD data, to possess a secondary structure consisting mainly of α-helices [6], those proteins having the folding pattern of α/β mixtures were excluded. There are two all-α proteins listed in Table 1, i.e. GCSF and endonuclease, which were considered to be appropriate references, and rough models of IL-6 were constructed (data not shown) from the PDB data. Based

Table 1
Compatibility of the human IL-6 sequence with known structures; 617 structures were compared and sorted in order of their compatibility scores

Rank	Structure	PDB code	Compatibility score	Folding type
1	Isocitrate dehydrogenase	7ICD	−2.47	alpha/beta
2	Endo-1,4-β-D-glucanase	1TML	−2.27	alpha/beta
3	Purine nucleoside phosphorylase	1ULA	−2.16	alpha/beta
–	IL-6 model based on b-GCSF	1BGC	−2.14	all alpha
4	Uridylate kinase	1UKZ	−2.02	alpha/beta
5	Granulocyte colony-stimulating factor (GCSF)	1BGC	−1.93	all alpha
6	Pyruvate kinase	1PKN	−1.88	alpha/beta
7	Dethiobiotin synthase	1DTS	−1.86	alpha/beta
8	Type/III chloramphenicol acetyltransferase	3CLA	−1.86	alpha/beta
9	Malate dehydrogenase	1HLP	−1.85	alpha/beta
10	Endonuclease	1ABK	−1.75	all alpha
–	IL-6 model based on m-LIF	1LKI	−1.61	all alpha
100	Leukemia inhibitory factor (LIF) mouse	1LKI	−1.02	all alpha

The best 10 structures are listed together with IL-6's models, IBGC and ILKI, and reference proteins, 1BGC and 1LKI.

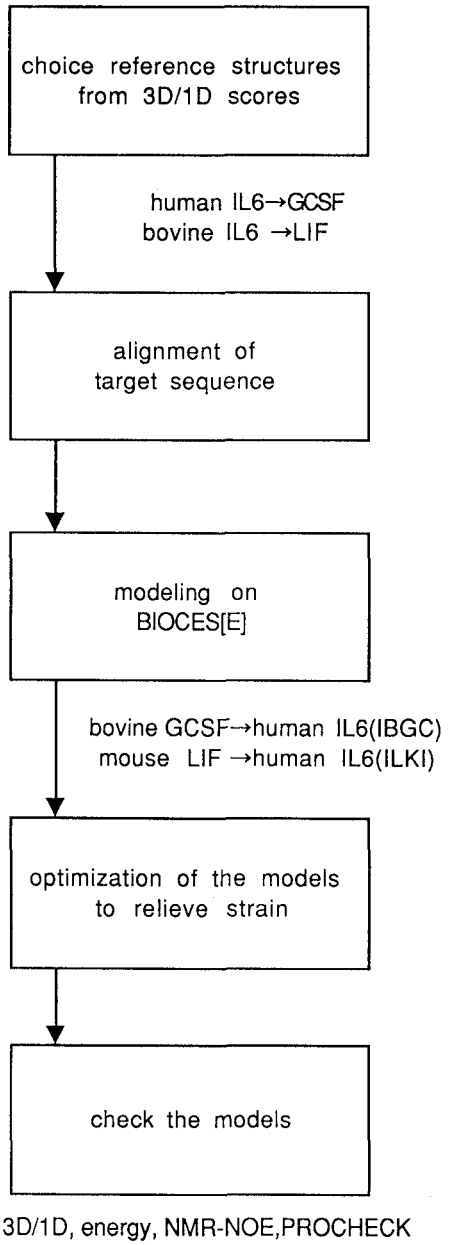

Fig. 2. Flow diagram of our homology modeling procedures.

on the consideration that it should have S-S bridges, the distances between the Cys C-α atoms of the corresponding residues were judged to be too long for the endonuclease-based model. Therefore, it is not suitable as a reference protein, leaving only GCSF.

When the *compass* search was performed with the bovine IL-6 sequence, instead of that of human IL-6, the results shown in Table 2 were obtained, where 1LKI (mouse LIF) was found to be the best whereas 1BGC was not good, being ranked 41st. Considering that the sequence identity between human and bovine IL-6 is 50%, the 3D/1D compatibility method appears to be very sensitive to the input sequence. Mouse LIF also combines with gp130 in the same manner as IL-6 does, in other words, gp130 can be shared as a receptor. It is also remarkable that LIF and IL-6 belong to the same long-chain-type classification which was assigned by Sprang and Bazan [13]. Along with these facts, it is reasonable to employ the three-dimensional structure of mouse LIF as another reference structure.

3.2. Modeling

Using the tertiary structures of GCSF and LIF registered in PDB, IL-6 modeling was performed. The alignment used is illustrated in Fig. 1. The alignment of GCSF and IL-6 determined here is consistent with those obtained by others including Bazan [14]. In the alignment of GCSF and IL-6, the positions of Cys residues forming S-S bridges were useful in the arrangement.

For the alignment between human IL-6 and mouse LIF in Fig. 1, the alignment between bovine IL-6 and mouse LIF was first determined using the 3D/1D method, and then bovine IL-6 was used to replace human IL-6 according to sequence homology.

3.3. Features of the models

The two IL-6 models are illustrated in Fig. 3. IBGC is the model of human IL-6 based on bovine GCSF, and ILKI is that based on mouse LIF. Helix *A* of IBGC is straight, however, that of ILKI has a kink, which is because the original structure of LIF has an internal hydrogen bond between Leu-40 (O) and Ser-43 (N). Although IL-6 also has Ser-36 in helix *A*, it was not possible to match it with Ser-43 of LIF in the alignment as the resulting model contains the destroyed hydrophobic core. There is no reason for the presence of a kink in the helix *A* of ILKI as the Ser residue has no possible internal hydrogen bond partner within the helix. The positions

Table 2
Compatibility of the bovine IL-6 sequence with known structures; 617 structures were compared and sorted in order of their compatibility scores

Rank	Structure	PDB code	Compatibility score	Folding type
1	Leukemia inhibitory factor (LIF) mouse	1LKI	−2.23	all alpha
2	Endo-1,4-β-D-glucanase	1TML	−2.20	alpha/beta
3	Isocitrate dehydrogenase	7ICD	−2.19	alpha/beta
4	Annexin V rat	2RAN	−1.92	alpha/beta
5	Methane monooxygenase hydrolase	1MMOB	−1.90	all alpha
6	Deoxyribonuclease I rabbit	1ATNA	−1.83	alpha/beta
7	Cytochrome P450 (BM-3)	2HPDA	−1.79	alpha/beta
8	Renin mouse	1SMRA	−1.78	alpha/beta
9	*Eco*RI endonuclease	1ERLA	−1.76	alpha/beta
10	L-Lactate dehydrogenase	1LDNA	−1.75	alpha/beta
41	Granulocyte colony-stimulating factor (GCSF)	1BGC	−1.39	all alpha

The best 10 structures are listed together with reference protein, 1BGC.

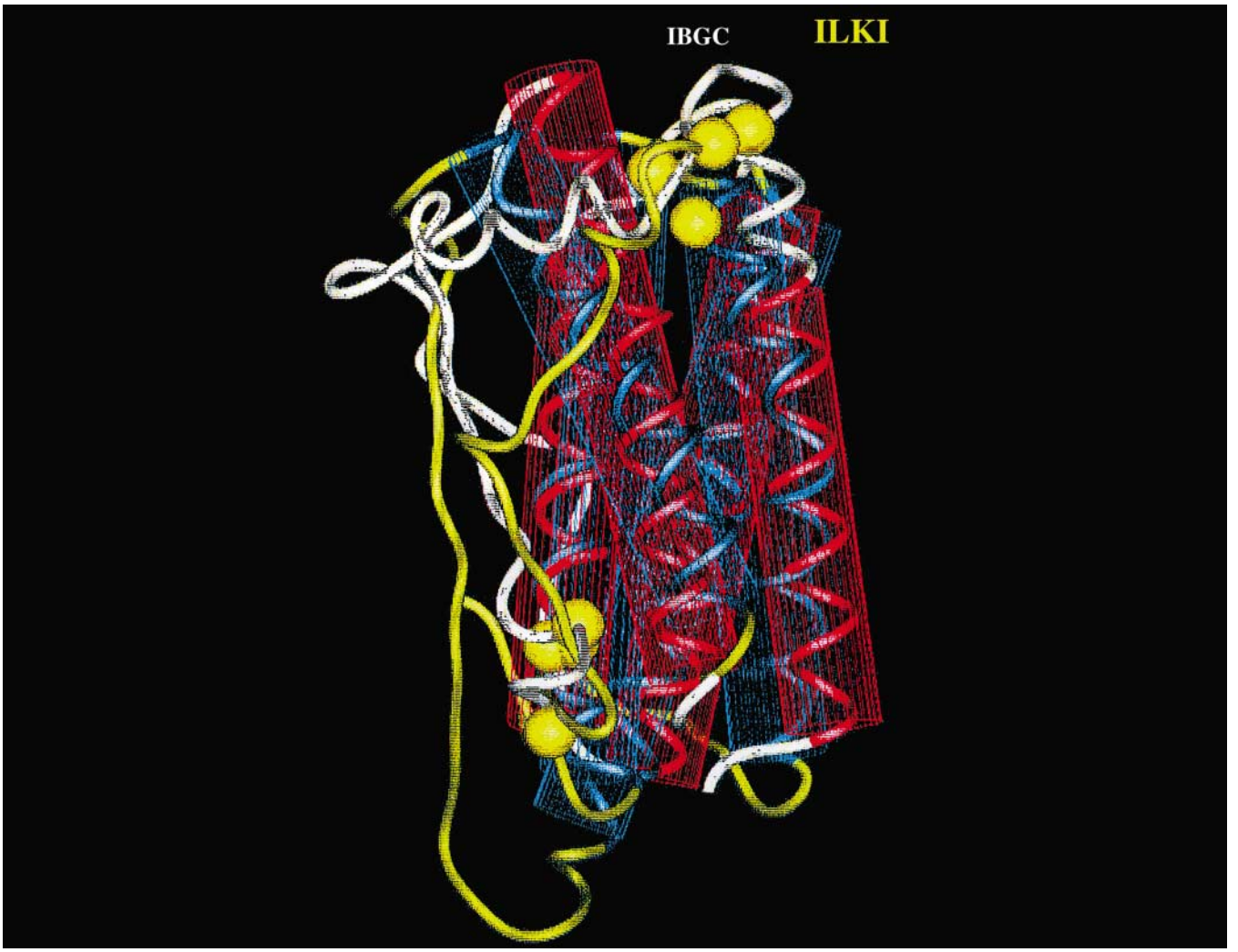

Fig. 3. Homology model of human IL-6 based on bovine GCSF (IBGC, *white*). Homology model of human IL-6 based on mouse LIF (ILKI, *greenish yellow tube*). A pair of yellow balls represents S-S bond (Cys-CA atoms).

of helices *B* and *C* match between the two models. However, the position of helix *D* slips by one roll to each of the others.

The wheel models [15] are shown in Fig. 4, which is of help in gaining knowledge of the inter-helical interactions and nearest neighbors therein. This is lateral information, the vertical relationship among the four helices being shown in Fig. 5. When two models were both superimposed on helices *B* and *C*, it was recognized there is one roll slip in helix *A* between the models. In helix *D*, which has a different relative gradient from the other helices, there is also one roll slip between the two models. No major difference between the models was found for helices *B* and *C*, however, one or two residue slips can be recognized.

As shown here, the two models have different manners of inter-helical packing, so one of those should be incorrect.

3.4. Evaluation by compass and energy calculations

The results on the models obtained by evaluation using

Table 3
Energy values in kJ calculated with CHARMm for 16R-184M of IBGC and ILKI

	IBGC		ILKI	
	Initial	Minimized	Initial	Minimized
Bond energy	713.94	502.95	84 539.96	541.55
Angle energy	4 138.89	4 024.98	4 813.45	2 691.80
Regular dihedral energy	1 894.20	1 836.59	1 981.16	2 232.08
Improper dihedral energy	661.81	875.04	4 436.88	822.36
Lennard-Jones energy	6.355*E*+11	−3 526.01	2.534*E*+11	−1 209.88
Electrostatic energy	−26 723.88	−30 226.51	−21 866.42	−237 729.09
Constraints, other	26 723.85	−0.004186	4 720.57	0.002093
Total energy	6.354*E*+11	−26 512.96	2.534*E*+11	−22 651.18

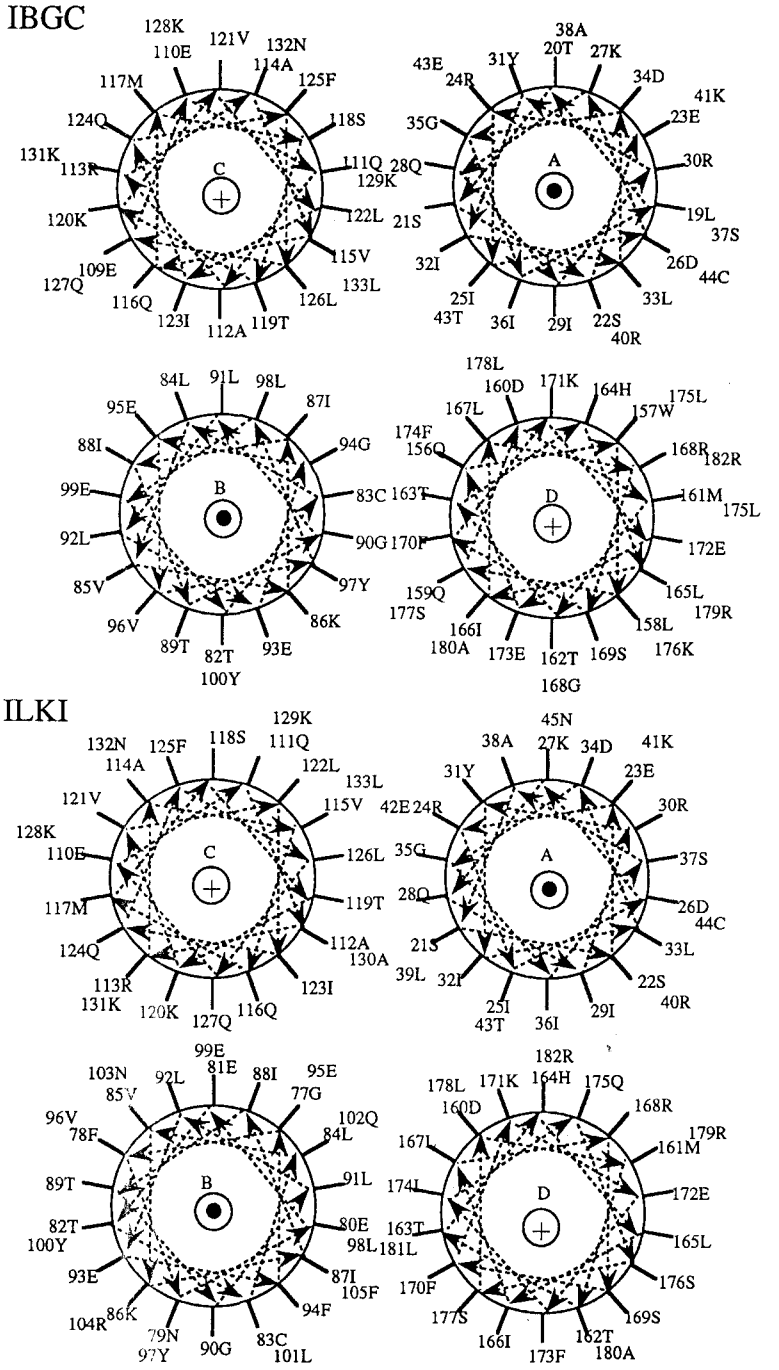

Fig. 4. Wheel plots for models IBGC and ILKI which show lateral information for inter-helical arrangement.

compass are given in Table 1. The data show that both models gained higher scores compared with each reference protein. Furthermore, IBGC was evaluated as being superior to ILKI. The *compass* scores for residues are given separately in Fig. 6, in which it can be seen that the scores of the helices are better than those of the loops. It also shows that the two models have good helical backbones.

Each energy was calculated using *CHARMm*, where the numbers of residues were adjusted to be equal (16R–184M) for both models, and the values obtained are listed in Table 3, which shows that the total energy of IBGC is lower than that

Table 4
NOE reproducibility of two models

Residue pairs		IBGC	ILKI
Tyr-100	Ile-166	B	C
Val-115	Phe-170	A	C
Ile-29	Leu-174	A	C
Ile-36	Leu-167	A	C
Ile-36	Phe-170	A	A
Phe-94	Phe-170	A	A
Val-96	Leu-151	C	C
Tyr-97	Leu-151	A	C

Definition of proximity, in which nearest atom pairs other than hy-

of ILKI. This is thought to be due mainly to its lower electrostatic and Lennard-Jones energy terms.

3.5. Evaluation by NOE

The models were also evaluated with distance information obtained from NMR investigation [16]. The distances were calculated for the pairs of residues in the models in which NOE cross-peaks were observed in the experiment. The results are shown in Table 4. IBGC showed good agreement with the NOE data, except for the pair consisting of Val-96 and Leu-151. The distance from Val-96 (CA) to Leu-151 (N) in the IBGC model is 1.16 nm, however, as Leu-151 is located on the loop, this inconsistency can be avoided by adjusting the orientation of its side chain. ILKI showed worse agreement with the NOE data; only two pairs could be reproduced (Tyr-94–Phe-170; Ile-36–Phe-170). In their paper, there were two leucines with unknown residue numbers, i.e. Leu-X and Leu-Y. NOE cross-peaks were observed for two residue pairs, Leu-X–Tyr-100 and Leu-X–Phe-105 therein. Both IBGC and ILKI showed that Leu-X was determined to be Leu-101 in the present study. They also showed cross peaks between three residue pairs, Leu-Y–Phe170, Leu-Y–Phe173 and Leu-Y–Leu174. Although IBGC was able to show that Leu-Y was determined to be Leu-91, ILKI could not demonstrate anything.

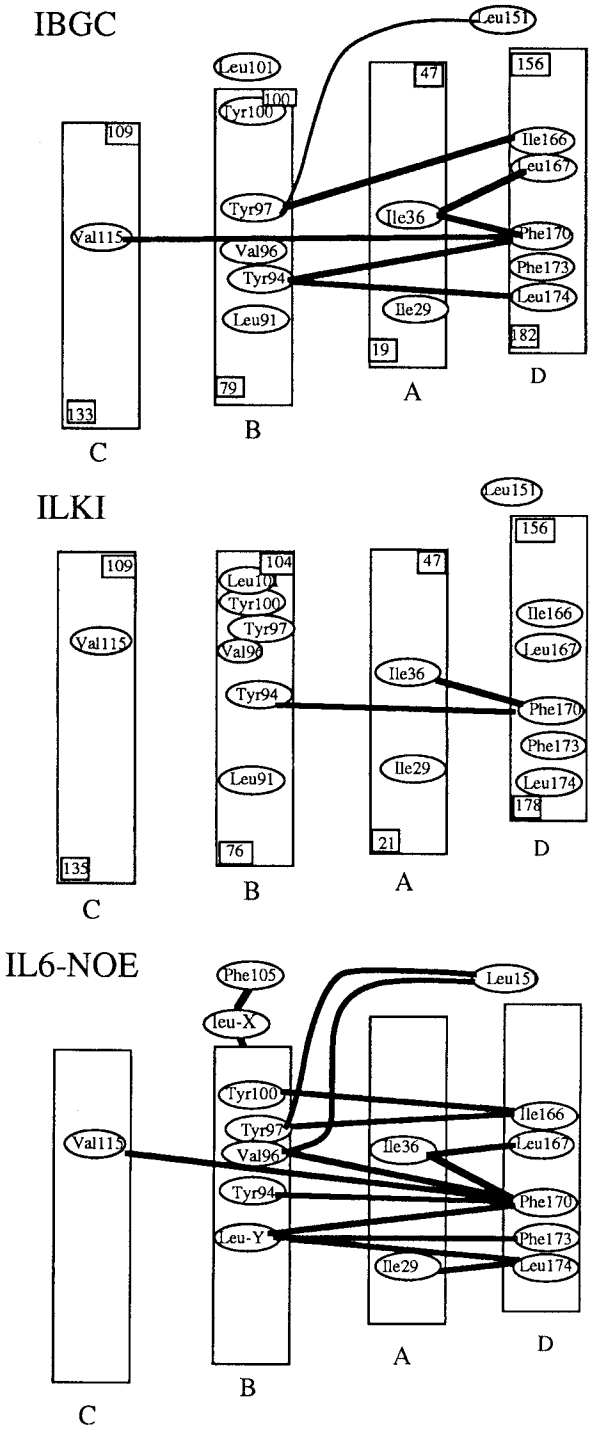

Fig. 5. Vertical relationship among four helices in models IBGC and ILKI. Solid lines define experimental NOE pairs.

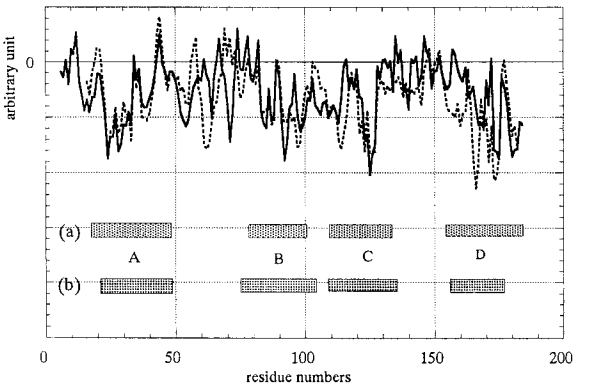

Fig. 6. 3D/1D score for each residue. (- - -) IBGC, (—) ILKI. Box shows helical regions; (a) IBGC, (b) ILKI.

As shown above, by constructing two models of IL-6 based on two known protein structures of GCSF and LIF, it was judged from *compass* scores, energy values, and NOE reproducibility that the GCSF model is superior to the LIF model.

Acknowledgements: We thank Dr. Chiaki Nishimura and Prof. Ichio Shimada for the use of their NOE data and helpful discussions.

References

[1] Hirano, T. and Yasukawa, K. (1986) Nature 324, 73–76.
[2] Kishimoto, T. (1989) Blood 74, 1–10.
[3] Suematsu, S., Matsuda, T., Aozasa, K., Akira, S., Nakano, N., Ohno, S., Miyazaki, J., Yamamura, K., Hirano, T. and Kishimoto, T. (1989) Proc. Natl. Acad. Sci. 86, 7547–7557.
[4] Poli, V., Balena, R., Fattori, E., Markatos, A., Yamamoto, M., Tanaka, H., Cliiberto, G., Roda, G.A. and Costantini, F. (1994) EMBO J. 13, 1189–1196.
[5] Kopf, M., Baumann, H., Freer, G., Freudenberg, M., Lamers, M., Kishimoto, T., Zinkernagel, R., Bluethmann, H. and Koehler, G. (1994) Nature 368, 339–342.
[6] Kruettgen, A., Stefan, R.-J., Moeller, C., Wroblowski, B., Wollmer, A., Muellberg, J., Hirano, T., Kishimoto T. and Heinrich, P.C. (1990) FEBS Lett. 262, 323–326.
[7] Hammacher, A., Ward, L., Weinstock, J., Treutlein, H., Yasukawa, K. and Simpson, R.J. (1994) Protein Sci. 3, 2280–2293.
[8] Ehlers, M., Grotzinger, J., deHon, F.D., Mullberg, J., Brakenhoff, J.P., Liu, J., Wollmer, A. and Rose-John, S. (1994) J. Immunol. 153, 1744–1753.
[9] Savino, R., Lahm, A., Salvati, A.L., Ciapponi, L., Sporeno, E., Altamura, S., Paonessa, G., Toniatti, C. and Ciliberto, G. (1994) EMBO J. 13, 1357–1367.

[10] Nishikawa, K. and Matsuo, Y. (1993) Protein Eng. 6, 811–820.
[11] Matsuo, Y. and Nishikawa, K. (1994) Protein Sci. 3, 2055–2063.
[12] Laskowski, R.A., MacArthur, M.W., Moss, D.P. and Thornton, J.M. (1993) J. Appl. Crystallogr 26, 282–291.
[13] Sprang, S.R. and Bazan, J.F. (1991) Curr. Opin. Struct. Biol. 3, 815–827.
[14] Bazan, J.F. (1991) Neuron 7, 197–208.
[15] Schiffer, M. and Edmundson, A.B. (1967) Biophys. J. 7, 121–135.
[16] Nishimura, C., Watanabe, A., Gouda, H., Shimada, I. and Arata, Y. (1996) Biochemistry, 35, 273–281.