Tertiary Structural Models of Human Interleukin-6 and Evaluation by Comparison with X-Ray and NMR Structures[1)]

Hiromi SUMIKAWA* and Ei-ichiro SUZUKI

Central Research Laboratories, Ajinomoto Co., Inc. 1–1 Suzuki-cho, Kawasaki-ku, Kawasaki, 210, Japan.
Received September 5, 1997; accepted September 24, 1997

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). Those models were evaluated using a sequence–structure compatibility (3D–1D) method program Compass and a limited amount of NMR distance information when it was concluded that the model based on GCSF (IBGC) was preferable to that from LIF (Sumikawa *et al.*, *FEBS Lett.*, 404, 234 (1997)). We evaluated the quality of this model (IBGC) by comparing with X-ray (Somers *et al.*, *EMBO. J.*, 16, 989 (1997)) and NMR (Xu *et al.*, *J. Mol. Biol.*, 268, 468 (1997)) structures. Consequently, normal mode calculations were carried out for this model, giving conformation fluctuations similar to the C alpha deviation pattern between X-ray and NMR structures.

Key words structure prediction; interleukin-6 (IL-6); nuclear Overhauser effect (NOE); GCSF; 3D–1D method; normal mode

In structure biology and related sciences, it is a matter of key importance to predict the three-dimensional (3D) 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 (1D) sequence has been determined. Interleukin-6 (IL-6) was one of the cytokines chosen as a test case in our present study. Many cytokines have been demonstrated to adopt four helices bundled in an up-up-down-down type of topology, however, there is little sequence identity with IL-6. Therefore, the tertiary structure prediction for IL-6 cannot be achieved with the conventional homology modeling method based on amino acid sequence information alone. We have to be careful when building these models to focus our attention on four helices only.

As reference proteins, GCSF and LIF were chosen for the sequence structure compatibility method (3D–1D).[2a,b)] In our previous study,[3)] two models (one based on GCSF and the other on LIF) were constructed, and evaluation of these models was carried out using NMR experimental information (labeled IL-6 has only 8 pairs of NOE data),[4)] the 3D–1D method, and energy calculations. The IL-6 model based on GCSF is illustrated[5)] in Fig. 1. Furthermore, the model based on GCSF was evaluated crudely as being superior to that based on LIF.

As the models are artificial, it is important to evaluate the quality of this model using experimental structural information. After our models had been constructed, X-ray[6)] and NMR[7)] structures were published, which enabled a reevaluation of our models. In this report, since these experimental coordinate sets are not yet available, only the structural information described in these publications was used for the evaluation of the model based on GCSF, which gained the higher score of the two.

Secondary Structures The four helices in our model were arranged so that helices A and B run in the same direction and C and D in the opposite direction. In the X-ray structure, the helices A, B, C and D extend from 21S to 45A, 80E to 102E, 109E to 129K and 156Q to 182R, respectively. Our model has helix A (19L—47S), helix B (79E—100Y), helix C (109E—133L) and helix D (156Q—182R), the helical regions of which correspond well to those in the X-ray structure. Other than the main four helices, the X-ray structure has a short helix in the CD-loop, which is called helix E (141P—152Q). The GCSF X-ray structure has no short helix in the CD-loop, but it has a short helix in the AB-loop, for which a similar helix was included in our model.

Hydrophobic Core Residues The wheel model for our modeled structure is shown in Fig. 2. The inside region of the wheel model shows the hydrophobic core of the protein. Numbers in circles indicate experimentally reported hydrophobic core residues, which are located

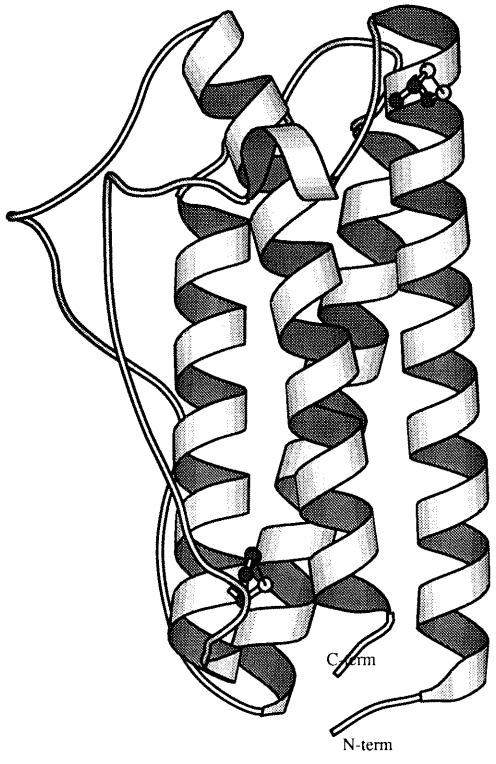

Fig. 1. Homology Model of Human IL-6 Based on Bovine GCSF
A pair of balls and sticks represents the S–S bond (MOLSCRIPT).[5)]

* To whom correspondence should be addressed.

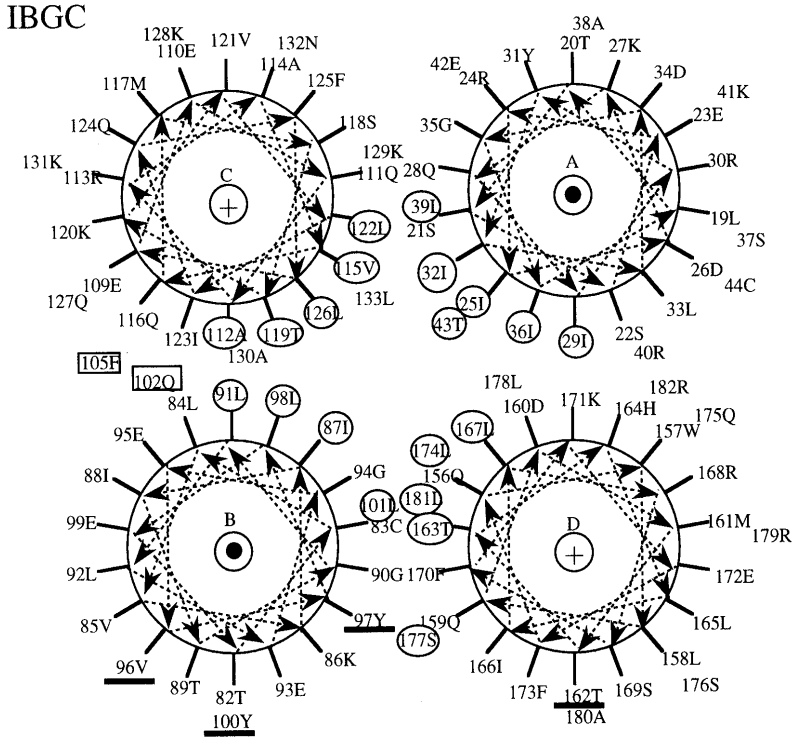

Fig. 2. Wheel Plots Which Show Lateral Information for the Inter-Helical Arrangement

○; hydrophobic core residues (the authors of the X-ray structure wrote). —; residues in the proximity of the E mini-helix.

inside the helices. Therefore, our model is in good agreement with the X-ray study on the hydrophobic core of the protein. Underlined numbers show the residues which are thought to interact with helix E,[6] although it is located beneath the figure.

Packing among Helices A stereo C alpha trace of the four IL-6 main helices is illustrated in the X-ray crystallography literature (Fig. 3 in Ref. 6). Compared with this, the packing condition of the main chain in our model is satisfied.

Helix B in the X-ray structure of IL-6 has a 38 degree bend between 93E and 94F. Since GCSF, which was the reference in our study, has no kink in this helix, our model also has no kink.

Completely Assigned NOE Data Our model was evaluated with recently published assigned NOE data,[8] as shown in Table 1. There were thirty pieces of NOE data, most of which could be reproduced with our model structure, except for the pair of 39L and 105F which was C grade. The amino acid residue of 105F was located in the BC-loop, so it can be adjusted to come closer to 39L in our model.

Structural Fluctuations Recently, biomolecular motion has been described by the linearly independent normal mode.[9] Consequently, we calculated the normal mode of our model[10] to locate any fluctuations. Figure 3 shows fluctuations of C alpha atoms obtained from the normal mode. These C alpha fluctuations have similar features to the C alpha deviations between the X-ray and NMR structures. In other words, our normal mode analysis reproduced well the C alpha deviations between the X-ray and NMR structures.

The solution structure (Fig. 1 and Fig. 2, c; both in Ref. 7) shows that there may be some NMR structures

Table 1. NOE Reproducibility of the Model

Residue pairs		IL-6 model based on GCSF
Leu 39	Leu 98	A
Leu 39	Phe 105	C
Ile 25	Leu 126	A
Gln 28	Leu 122	A
Ile 32	Thr 119	A
Ile 36	Val 115	A
Leu 39	Val 115	A
Leu 39	Ala 114	B
Asp 26	Leu 178	A
Ile 29	Leu 178	A
Ile 29	Leu 174	A
Ile 33	Leu 174	A
Ile 36	Leu 174	A
Ile 36	Phe 170	A
Arg 40	Leu 167	A
Glu 95	Thr 119	A
Leu 98	Thr 119	A
Gln 102	Ala 112	B
Phe 105	Ala 112	B
Cys 83	Met 184	A
Cys 83	Ala 180	A
Ile 87	Ser 177	A
Ile 87	Phe 173	A
Gly 90	Phe 173	A
Phe 94	Phe 173	A
Phe 94	Phe 170	A
Val 96	Asn 144	A
Val 96	Leu 147	A
Glu 99	Leu 148	A
Tyr 100	Leu 115	A

Definition of proximity, in which nearest atom pairs other than hydrogens were used for distance calculations; A, up to 0.60 nm; B, 0.60 to 0.80 nm; C, 0.80 nm and above.

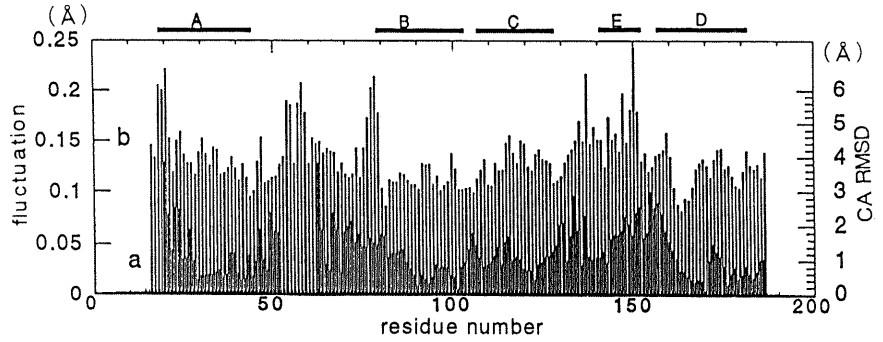

Fig. 3. C Alpha of X-Ray Structure and C Alpha of NMR Structure Distances after the Best-Fit Superposition[7] (a) and Fluctuation of C Alpha for the Normal Mode (b)

Upper; bars show helical regions.

originating from conformational fluctuations. The fluctuation is small in the helices and large in the loops. 50C—55E, at the beginning of the AB-loop, and 68A—80E on the N-terminal side of helix B have large fluctuations. Since 50C—55E are not evident in the X-ray structure (Fig. 2, d in Ref. 7), comparison with our normal mode data has not been carried out (Fig. 3).

The result of the normal mode calculation (Fig. 3) also showed small fluctuations at the helices and large fluctuations at the loops. It can be assumed that there are comparatively large or small fluctuations in the loops, for instance, 68A—80E in the AB-loop and 133—149 in the CD-loop had large fluctuations. The final section of the AB-loop was defined by three structural elements, a type I beta turn (68A—71D), a disulfide (73C—83C) and a type II beta turn (75Q—78F). There were no such turns in our model, resulting in large fluctuations. Because our model had no helix E (141P—152Q), there were large fluctuations in the CD-loop. However, there was a short helix, 52S—56A, in our model, and this region had a larger fluctuation than the four main helices. It was shown that a helix, other than the main helices, had a large fluctuation. Consequently, normal mode calculations were carried out for this model, giving conformation fluctuations similar to the C alpha deviation pattern between the X-ray and NMR structures.

Conclusion

IL-6 has little sequence identity with many cytokines, therefore tertiary structure prediction for IL-6 could not be achieved with the conventional homology modeling method based on amino acid sequence information alone. A better model, based on GCSF using two IL-6 models constructed previously by us, was selectively evaluated, referring to X-ray and NMR structures published recently. Our GCSF-based model was in very good agreement with the experimentally reported structures as far as secondary structures, hydrophobic core residues packing among helices and much of the NOE data were concerned. Accordingly, it is important that we could perform protein modeling which was very difficult using the usual homology modeling. Consequently, normal mode calculations were carried out for this model, giving conformational fluctuations similar to the C alpha deviation pattern between the X-ray and NMR structures. From the results, our model may also be reasonable from a dynamic point of view.

References and Notes

1) Abbreviations: IL-6, interleukin-6; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; GCSF, granulocyte colony-stimulating factor; LIF, leukemia inhibitory factor, IBGC, the model of IL-6 based on bovine GCSF which code is called 1BGC in the protein data bank
2) *a*) Nishikawa K., Matsuo Y., *Protein Engineering*, **6**, 8, 811—820 (1993); *b*) Matsuo Y., Nishikawa K., *Protein Sci.*, **3**, 2055—2063 (1994).
3) Sumikawa H., Fukuhara K., Suzuki E., Matsuo Y., Nishikawa K., *FEBS Lett.*, **404**, 234—240 (1997).
4) Nishimura C., Watanabe A., Gouda H., Shimada I., Arata Y., *Biochemistry*, **35**, 273—281 (1996).
5) Kraulis P. J., *J. App. Crystallography*, **24**, 946—950 (1991).
6) Somers W., Stahl M., Seehra J. S., *EMBO. J.*, **16**, 989—997 (1997).
7) Xu G., Yu H., Hong J., Stahl M., McDonagh T., Kay L. E., Cumming D. A., *J. Mol. Biol.*, **268**, 468—481 (1997).
8) Xu G., Hong J., McDonagh T., Stahl M., Kay L. E., Seehara J., Cumming, D. A., *J. Bio. NMR*, **8**, 123—135 (1996).
9) Quake S., R., Abcock H., Chu S., *Nature* (London), **388**, 151—154 (1997).
10) Kamiya K., Umeyama H., 19th Kozo Kassei Soukan Symposium, 1991, p. 308.