FAMS Complex: A Fully Automated Homology Modeling System for Protein Complex Structures

Mayuko Takeda-Shitaka*, Genki Terashi, Chieko Chiba, Daisuke Takaya and Hideaki Umeyama

School of Pharmaceutical Sciences, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan

Abstract: The formation of a protein-protein complex is responsible for many biological functions; therefore, three-dimensional structures of protein complexes are essential for deeper understandings of protein functions and the mechanisms of diseases at the atomic level. However, compared with individual proteins, complex structures are difficult to solve experimentally because of technical limitations. Thus a method that can predict protein complex structures would be invaluable. In this study, we developed new software, FAMS Complex; a fully automated homology modeling system for protein complex structures consisting of two or more molecules. FAMS Complex requires only sequences and alignments of the target protein as input and constructs all molecules simultaneously and automatically. FAMS Complex is likely to become an essential tool for structure-based drug design, such as *in silico* screening to accelerate drug discovery before an experimental structure is solved. Moreover, in this post-genomic era when huge amounts of protein sequence information are available, a major goal is the determination of protein-protein interaction networks on a genomic scale. FAMS Complex will contribute to this goal, because its procedure is fully automated and so is suited for large-scale genome wide modeling.

Key Words: Homology modeling, protein complex structure, protein-protein interaction, structure-based drug design, protein interaction network.

1. INTRODUCTION

Huge amounts of protein sequence information have been generated by genome sequencing projects, and many new protein sequences representing potential therapeutic targets have been found. To efficiently carry out structure based drug design such as *in silico* screening, a precise understanding of the biological function of the target protein based on three-dimensional (3D) structures is indispensable. However, there are many important proteins for which the sequence is available but the 3D structure has not been experimentally solved. In such cases, theoretical methods that can reliably predict 3D protein structure are useful. At present, homology modeling is the most reliable tool for protein structure prediction. Homology modeling constructs 3D model structures of unknown proteins (target proteins) based on known homologous 3D structures (reference proteins) arising from the concept that proteins with similar sequences have similar structures. For homology modeling, various prediction programs have been developed by different laboratories. Our laboratory, for example, developed the CHIMERA modeling system [1-3]. CHIMERA is an interactive homology modeling system that enables human intervention at necessary stages. After the development of CHIMERA, our laboratory also developed FAMS (a fully automated homology modeling system) by automating the procedures of CHIMERA [2-4]. When using homology models in structure based drug design, the accuracy of model structures is important. To understand the quality of homology modeling, the Critical Assessment of Techniques for Protein Structure Prediction (CASP; http://predictioncenter.org/), is a most valuable experiment. CASP is the blind contest of protein structure prediction, started in 1994 as CASP1 and has continued biennially to the most recent CASP6 held in 2004. Our laboratory participated in CASP6 as the group named *CHIMERA* using in-house programs CHIMERA and FAMS. The official results of CASP6 showed that our group was one of the most successful predictors in the homology based modeling category (http://predictioncenter.org/casp6/meeting/presentations/CASP6_Program.doc), which indicates that CHIMERA and FAMS can accurately construct homology models [2]. Moreover, Critical Assessment of Fully Automated Structure Prediction (CAFASP) is a contest to measure the capabilities of programs without any human intervention. Our laboratory, using FAMS, participated in CAFASP2 and CAFASP3 held in 2000 and 2002, respectively. The results of these CAFASPs showed that the best results for evaluations of side-chain $\chi 1$ angles among the private servers were given by our server using FAMS [5,6] (http://www.cs.bgu.ac.il/~dfischer/CAFASP3).

CHIMERA and FAMS are homology modeling systems for individual proteins (single-chain proteins). 3D structures of protein-protein complexes are essential for structure-based drug design, because many biological functions involve formation of protein-protein complexes. Thus prediction methods that reliably predict structures of protein-protein complexes (multi-chain proteins) are invaluable. In this study, we developed new software, FAMS Complex, a fully automated homology modeling system for multi-chain protein structures consisting of two or more chains. FAMS Complex is not docking software that attempts to find the

*Address correspondence to this author at the School of Pharmaceutical Sciences, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan; Tel: +81-3-5791-6331; Fax: +81-3-3446-9553;
E-mail: shitakam@pharm.kitasato-u.ac.jp

1573-4064/06 $50.00+.00

best matching between separate molecules, but is homology modeling software for multi-chain proteins. FAMS Complex was developed to construct multi-chain protein structures by modifying the procedures of FAMS that construct single-chain proteins.

2. MATERIALS AND METHODS

The FAMS procedures are described in detail in a previous paper from our laboratory [4]. FAMS has three steps; 1) construction of Cα atoms, 2) construction of main-chain atoms, and 3) construction of side-chain atoms with main-chain optimization. In this method, main-chain atoms are constructed based on the similarity of environmental residues at topologically equivalent positions in reference proteins (local space homology). Side-chain atoms are constructed by iterative cycles of side-chain generation and main-chain optimization. Side-chain generation is based on the conservation of side-chain conformations for each residue within homologous proteins [7]. Main-chain atoms are optimized with the fixed side-chain conformations. Based on these FMAS procedures, FAMS Complex, which predicts multi-chain protein structures, was developed. In this section, we describe the differences between FAMS Complex and FAMS.

2-1. Input

Inputs for FAMS and FAMS Complex are sequence alignments between target and reference proteins. In the case of single-chain proteins, input is an alignment between target and reference proteins. In the case of multi-chain protein, there are two or more alignments. For example, in the case of modeling a protein consisting of *N* chains, there are *N* alignments obtained by homology search methods such as PSI-BLAST [8] (Fig. (**1**)). In the input file for FAMS Complex, all alignments are combined into one alignment as if it were a single-chain protein in order to simplify the input format (Fig. (**1**)). Because an identification mark, an alphabetic letter like "U" that is not assigned to a one-letter amino acid code is inserted between the end residue of one chain and the first residue of the next chain, FAMS Complex can automatically distinguish individual chains in the following modeling steps. All the combined amino acid residues excepting "U" are sequentially renumbered. For example in Fig. (**1**), the first chain (chain 1) comprises residues 1 – RES_1, the second chain (chain 2) comprises residues RES_1+1 – RES_2, and the last chain N comprises residues RES_{N-1}+1 – RES_N (where RES_n is the residue number of the end residue of chain *n*).

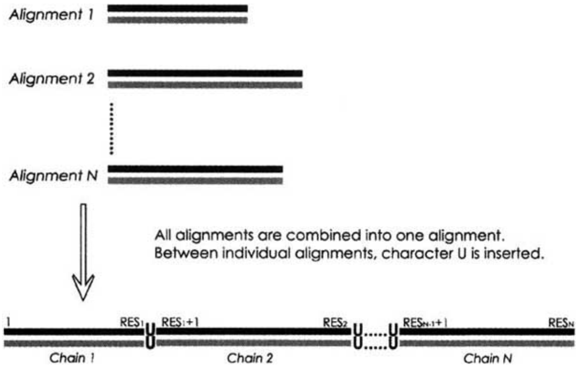

Alignment 1
Alignment 2
Alignment N
All alignments are combined into one alignment.
Between individual alignments, character U is inserted.
1
RES1
RES1+1
RES2
RESN-1+1
RESN
U
U
U
U
U
U
Chain 1
Chain 2
Chain N

Fig. (1). Construction of the input file for FAMS Complex.

This is an example of an input file for a protein consisting of N chains. Black and gray bars denote the sequences of target and reference proteins, respectively. In the input file for FAMS Complex, N alignments are combined into one alignment.

2-2. Construction of Cα atoms

Two separate processes are used to construct coordinates of Cα atoms. One is the assignment of an initial set of coordinates from a reference protein and database searches; the other is optimization with an objective function. These two processes are performed alternatively.

(1) Construction of the Initial Cα Atoms

The process of construction of the initial Cα atoms of FAMS Complex is the same as that of FAMS. In the sequence alignment between the target and reference proteins, in regions where residue pairs between target and reference proteins continue for over three residues, the coordinates of the Cα atoms of the reference protein are assigned to the initial coordinates of the Cα atoms of the target protein. In regions where the coordinates of the Cα atoms are not assigned from the reference protein, the initial coordinates of the Cα atoms are obtained by searching in the fragment database. In FAMS Complex, this process is automatically carried out in all chains.

(2) Refinement of Cα Atoms

The initial coordinates of the Cα atoms are refined using simulated annealing with an empirical objective function. The objective function is set to create a model similar to the native one, and does not have physicochemical meaning. The objective function is defined by the following sum:

$$U_{C\alpha} = U_{len} + U_{ss} + U_{ang} + U_{pos} + U_{vdw} \tag{1}$$

Equation (1) in FAMS Complex is the same as that in FAMS. Though multiple chains are combined into one chain in the input file in FAMS Complex, individual chains must be distinguished in the optimization steps. Therefore, conditions for the calculations for three terms (U_{len}, U_{ang} and U_{vdw}) were modified to distinguish individual chains. The differences of conditions between FAMS Complex and FAMS are as follows.

The U_{len} term in Equation (1) is the function of the distance between the Cα atoms of the sequentially adjacent residues in the same chain and is defined by Equation (2):

$$U_{len} = K_l \sum_i \left(D_i - 3.8\right)^2 \tag{2}$$

where D_i is the distance between the Cα atoms of the residues i and $i+1$ and K_l is a constant value set at 2. Equation (2) is the same as that in FAMS. In order that D_i between a Cα atom of the end residue of one chain and a Cα atom of the head residue of the next chain

($C\alpha_{RES_n} - C\alpha_{RES_n+1}$ where $C\alpha_i$ is the $C\alpha$ atom of residue i) is not calculated, the calculation is carried out under the following condition in FAMS Complex:

$$i \neq RES_n$$

where RES_n is the residue number of the end residue of chain n.

The U_{ang} term in Equation (1) is the function of the angle for the Cα atoms of the three sequentially adjacent residues in the same chain and is defined by Equation (3):

$$U_{ang} = K_a \sum_i \left(\theta_i - \theta_0\right)^2 \qquad (3)$$

where θ_i (rad) is the angle for the residues i, i+1 and i+2. θ_0 (rad) is set at (100/180)π (rad), which was determined from X-ray structures in the Protein Data Bank (PDB) [9]. K_a is the constant value set at 1. Equation (3) is same as that in FAMS. In order that θ_is for Cα atoms of the different chains ($C\alpha_{RES_n-1} - C\alpha_{RES_n} - C\alpha_{RES_n+1}$, $C\alpha_{RES_n} - C\alpha_{RES_n+1} - C\alpha_{RES_n+2}$) is not calculated, the calculation is carried out under the following condition in FAMS Complex:

$$i \neq RES_n - 1,\ RES_n$$

The U_{vdw} term in Equation (1) is the function for the Cα atoms. U_{vdw} is calculated for all Cα atom pairs that are three or more residues apart in the sequence and for all Cα atom pairs between different chains, and is defined by Equation (4):

$$U_{vdw} = K_{vdw} \sum_{i,j} \left\{ \left(\frac{3.8}{D_{i,j}} \right)^{12} - \left(\frac{3.8}{D_{i,j}} \right)^{6} \right\} \qquad (4)$$

where $D_{i,j}$ is the distance between the Cα atoms of the residues i and j. K_{vdw} is the constant value set at 0.01 ($D_{i,j} \leq 3.2$ Å) and at 0.001 ($D_{i,j} > 3.2$ Å). Equation (4) is same as that in FAMS. In order that $D_{i,j}$s between the different chains ($C\alpha_{RES_n-1} - C\alpha_{RES_n+1}$, $C\alpha_{RES_n} - C\alpha_{RES_n+1}$, $C\alpha_{RES_n} - C\alpha_{RES_n+2}$) are calculated, the calculation is carried out under the following conditions in FAMS Complex:

$$\text{if } i = RES_n - 1 \text{ then } j > i + 1$$

$$\text{if } i = RES_n \text{ then } j > i$$

$$\text{else } j > i + 2$$

There is no modification in the conditions of the U_{ss} and U_{pos} terms in Equation (1). The U_{ss} term is the function of the distance between the Cα atoms of the residues forming the disulfide bond, which can be formed both inter and intra chains; therefore, this term is calculated under the same conditions as in FAMS. The U_{pos} term is the function of the positions of Cα atoms, and is defined for the purpose of relatively stably maintaining the positions of Cα atoms in the structurally conserved regions; therefore, this term is calculated under the same conditions as in FAMS. These two terms are described in detail in a previous paper from our laboratory [4].

2-3. Construction of Main-Chain Atoms

Three separate processes are used to construct main-chain atoms including Cβ atoms (see Fig. (**1**) in our previous paper [4]). The first process is the assignment of initial main-chain atoms from the reference proteins and database searches. The second process is the refinement of the initial main-chain coordinates. The third process is the refinement of main-chain coordinates obtained from the second process with conserved side-chain atoms.

(1) Construction of the Initial Main-Chain Atoms

Three main-chain atoms (N, C and O atoms except the Cα atom) and the Cβ atom (except glycine residue) are obtained from the reference proteins or by searching a main-chain database. This process in FAMS Complex is the same as that in FAMS, though in FAMS Complex, it is automatically carried out in all chains.

(2) Refinement of Initial Main-Chain Atoms

The initial coordinates of main-chain atoms are refined using simulated annealing with an empirical objective function. The objective function is set to create a model similar to the native one, and does not have any physicochemical meaning. The objective function is defined by the following sum:

$$U_{main} = U_{bond} + U_{ang} + U_{nonbond} + U_{SS} + U_{pos} + U_{tor} + U_{chi} + U_{hydr} \qquad (5)$$

Equation (5) in FAMS Complex is the same as that in FAMS. Though the multiple chains are combined into one chain in the input file, individual chains must be distinguished in the refinement steps. Therefore, conditions of calculations for four terms (U_{bond}, U_{ang}, $U_{nonbond}$ and U_{tor}) were modified to distinguish individual chains. The differences of conditions between FAMS Complex and FAMS are as follows.

The U_{bond} term in Equation (5) is the function of bond length and is defined by Equation (6):

$$U_{bond} = K_b \sum_i \left(b_i - b_0\right)^2 \qquad (6)$$

where b_i is the bond length of five bonds (N_i–$C\alpha_i$, $C\alpha_i$–$C\beta_i$, $C\alpha_i$–C_i, C_i–O_i and C_i–N_{i+1}). K_b is the constant value set at 225, and b_0 is the standard bond length for each of the five bonds (see Table **A1** in our previous paper [4]). Equation (6) is the same as that in FAMS. In order that b_i for $C_{RES_n} - N_{RES_n+1}$ is not calculated, the calculation of b_i for $C_i - N_{i+1}$ is carried out under the following condition in FAMS Complex:

$$i \neq RES_n$$

Table 1. The Target Proteins for Evaluation of FAMS Complex

PDB ID	Protein name	Quaternary structure [a]	Number of residues [b]	sequence identity (%) [c]
homo-oligomers				
1dxl	Dihydrolipoamide dehydrogenase	dimer (A,B)	934 (467)	48.0
1xso	Cu,Zn superoxide dismutase	dimer (A,B)	296 (148)	68.9
1be4	Nucleoside diphosphate kinase	trimer (A,B,C)	444 (148)	93.9
1dbf	Chorismate mutase	trimer (A,B,C)	354 (118)	48.3
1o8n	Molybdenum cofactor biosynthetic protein domain Cnx1G	trimer (A,B,C)	477 (159)	50.3
1a2z	Pyrrolidone carboxyl peptidase	tetramer (A,B,C,D)	844 (211)	57.6
1d1i	Shiga-like toxin B	pentamer (A,B,C,D,E)	330 (66)	63.6
1f9a	Nicotinamide mononucleotide adenylyltransferase	hexamer (A,B,C,D,E,F)	984 (164)	59.1
hetero-oligomers				
1i7r	Class I MHC A2	dimer (A,B)	373 (274,99)	62.4,69.7
1i9i	Recombinant anti-testosterone Fab fragment	dimer (L,H)	437 (218,219)	77.5,77.6
1phn	Phycocyanin	dimer (A,B)	334 (162,172)	73.5,75.6
1ubo	[NiFe] Hydrogenase	dimer (S,L)	796 (263,533)	71.8,68.3
1a00	Hemoglobin	tetramer (A,B,C,D)	572 (141,145,141,145)	49.6,44.1,49.6,44.1
1apz	Aspartylglucosaminidase	tetramer (A,B,C,D)	555 (138,141,137,139)	44.9,35.5,45.3,36.0
1b2n	Methanol dehydrogenase	tetramer (A,B,C,D)	1246 (565,58,565,58)	65.0,67.2,65.0,67.2
1b8d	Phycourobilin-containing phycoerythrin	tetramer (A,B,K,L)	682 (164,177,164,177)	35.2,56.4,35.2,55.8
1h1l	Nitrogenase Mo-Fe Protein	tetramer (A,B,C,D)	1990 (476,519,476,519)	73.3,67.8,73.3,67.8
1dio	Diol dehydratase	hexamer (A,B,G,L,E,M)	1721 (552,178,130,552,178,131)	70.8,61.8,56.2,70.8,61.8,56.5
1f99	R-Phycocyanin	hexamer (A,B,K,M,N,L)	1002 (162,172,162,162,172,172)	77.2,75.6,77.2,77.2,75.6,75.6

[a] Alphabets in parentheses are the chain codes.

[b] Total number of residues of oligomers. Numbers of residues of individual chains are in parentheses.

[c] Sequence identities between target and reference chains.

The U_{ang} term in Equation (5) is the function of the bond angle and is defined by Equation (7):

$$U_{ang} = K_a \sum_{i,j} \left(\theta_{i,j} - \theta_0\right)^2 \qquad (7)$$

where $\theta_{i,j}$ (rad) is the bond angle of nine types of angles ($C\alpha_i$–C_i–N_{i+1}, O_i–C_i–N_{i+1}, C_i–N_{i+1}–$C\alpha_{i+1}$, $C\alpha_i$–C_i–O_i, N_i–$C\alpha_i$–C_i, N_i–$C\alpha_i$–$C\beta_i$, $C\beta_i$–$C\alpha_i$–C_i, O_i–C_i–OXT_i and S_j–S_i–$C\beta_i$). θ_0 (rad) is the standard bond angle for each of the nine angles (see Table **A1** in our previous paper [4]). K_a is the constant value set at 45. Equation (7) is the same as that in FAMS. In order that $\theta_{i,j}$s for $C\alpha_{RES_n} - C_{RES_n} - N_{RES_n+1}$, $O_{RES_n} - C_{RES_n} - N_{RES_n+1}$ and $C_{RES_n} - N_{RES_n+1} - C\alpha_{RES_n+1}$ are not calculated, the calculations of $\theta_{i,j}$s for $C\alpha_i$–C_i–N_{i+1}, O_i–C_i–N_{i+1} and C_i–N_{i+1}–$C\alpha_{i+1}$ are carried out under the following condition in FAMS Complex:

$$i \neq RES_n$$

The U_{tor} term in Equation (5) is the function of the main-chain torsional angle and is defined by Equation (8):

$$U_{tor} = K_t \sum_i \sqrt{\left(\phi_i - \phi_i^0\right)^2 + \left(\psi_i - \psi_i^0\right)^2} + K_\omega \sum_i \left(\omega_i - \omega_i^0\right)^2 \qquad (8)$$

where ϕ_i, ψ_i and ω_i are ϕ, ψ and ω torsional angles (C_i–N_{i+1}–$C\alpha_{i+1}$–C_{i+1}, N_i–$C\alpha_i$–C_i–N_{i+1} and $C\alpha_i$–C_i–N_{i+1}–$C\alpha_{i+1}$), respectively. ϕ_i^0 and ψ_i^0 are the nearest torsional angles to ϕ and ψ torsional angles, respectively, in the Ramachandran plot (see Fig. **A1** in our previous paper [4]). ω_i^0 is 0 (rad) for cis-proline and is π(rad) for other peptide bonds. K_t and K_ω are constant values set at 10 and 50, respectively. Equation (8) is the same as that in FAMS. Because ϕ_i, ψ_i and ω_i for $\mathrm{C}_{RES_n} - \mathrm{N}_{RES_n+1} - \mathrm{C}\alpha_{RES_n+1} - \mathrm{C}_{RES_n+1}$, $\mathrm{N}_{RES_n} - \mathrm{C}\alpha_{RES_n} - \mathrm{C}_{RES_n} - \mathrm{N}_{RES_n+1}$ and $\mathrm{C}\alpha_{RES_n} - \mathrm{C}_{RES_n} - \mathrm{N}_{RES_n+1} - \mathrm{C}\alpha_{RES_n+1}$ are not calculated, the calcu-lations of ϕ_i, ψ_i and ω_i are carried out under the following condition in FAMS Complex:

$$i \neq RES_n$$

The $U_{nonbond}$ term in Equation (5) is the function of the interactions of nonbonded atoms. $U_{nonbond}$ is calculated for all atom pairs that are four or more bonds apart and for all atom pairs that are between the different chains, and is defined by Equation (9):

$$U_{nonbond} = K_{non} \sum_{i,j} \varepsilon_{i,j} \left\{ \left(\frac{r^*_{i,j}}{r_{i,j}} \right)^{12} - 2 \left(\frac{r^*_{i,j}}{r_{i,j}} \right)^{6} \right\} \qquad (9)$$

where $r_{i,j}$ is the distance between two atoms. $\varepsilon_{i,j}$ and $r^*_{i,j}$ are constant values that are defined according to various atom types (see Table **A2** in our previous paper [4]). K_{non} is a constant value set at 0.25. Equation (9) is the same as that in FAMS. In order that nine $r_{i,j}$s between residue RES_n and residue RES_n+1 ($\mathrm{N}_{RES_n} - \mathrm{N}_{RES_n+1}$, $\mathrm{C}\alpha_{RES_n} - \mathrm{N}_{RES_n+1}$, $\mathrm{C}\alpha_{RES_n} - \mathrm{C}\alpha_{RES_n+1}$, $\mathrm{C}_{RES_n} - \mathrm{N}_{RES_n+1}$, $\mathrm{C}_{RES_n} - \mathrm{C}\alpha_{RES_n+1}$, $\mathrm{C}_{RES_n} - \mathrm{C}_{RES_n+1}$, $\mathrm{C}_{RES_n} - \mathrm{C}\beta_{RES_n+1}$, $\mathrm{O}_{RES_n} - \mathrm{N}_{RES_n+1}$ and $\mathrm{O}_{RES_n} - \mathrm{C}\alpha_{RES_n+1}$) are calculated, these $r_{i,j}$s between residues RES_n and RES_n+1 are added in the $U_{nonbond}$ calculations.

There is no modification in the U_{ss}, U_{pos}, U_{chi} and U_{hydr} terms in Equation (5). The U_{ss} term is the function of distances between the Cα and Cβ atoms of the residues forming the disulfide bond. The U_{pos} term is the function of the positions of the atoms. The U_{chi} term is the function of the chirality of the Cα atoms. The U_{hydr} term is the function of the conservation of hydrogen bonds in the main chain atoms within homologous proteins. These terms are described in detail in our previous paper [4].

(3) Refinement of Main-Chain Atoms with Conserved Side-Chain Atoms

The process of main-chain refinement with conserved side-chain atoms is the same as that of FAMS. For refined main-chain atoms obtained by above process, the coordinates of conserved side chain atoms defined in our previous work [4] are used. Then, main-chain atoms including Cβ atoms are refined using Equation (5) under the same conditions as for the refinement of the initial main-chain atoms in FAMS Complex. In the case of refinement of main-chain atoms with conserved side-chain atoms, the $U_{nonbond}$ term is calculated between main-chain atoms and partially generated side-chain atoms.

2-4. Construction of Side-Chain Atoms

The process of construction of side-chain atoms based on the iterative cycles of side-chain generation and main-chain optimization is the same as that of FAMS (see Fig. (**1**) in our previous paper [4]). First, side-chain conformations are generated based on the conservation of side-chain conformations for each residue within homologous proteins [7]. Then, main-chain atoms including Cβ atoms are optimized keeping torsional angles of the side chain fixed. Next, side-chain atoms are deleted, and side-chain generation is again performed. Then, main-chain optimization is performed again. For the main-chain optimization in the iterative cycles, the objective function defined by Equation (5) is used under the same conditions as for the refinement of the initial main-chain atoms in FAMS Complex. In the case of side-chain construction, the $U_{nonbond}$ term is calculated between all atoms (main and side chains).

3. RESULTS

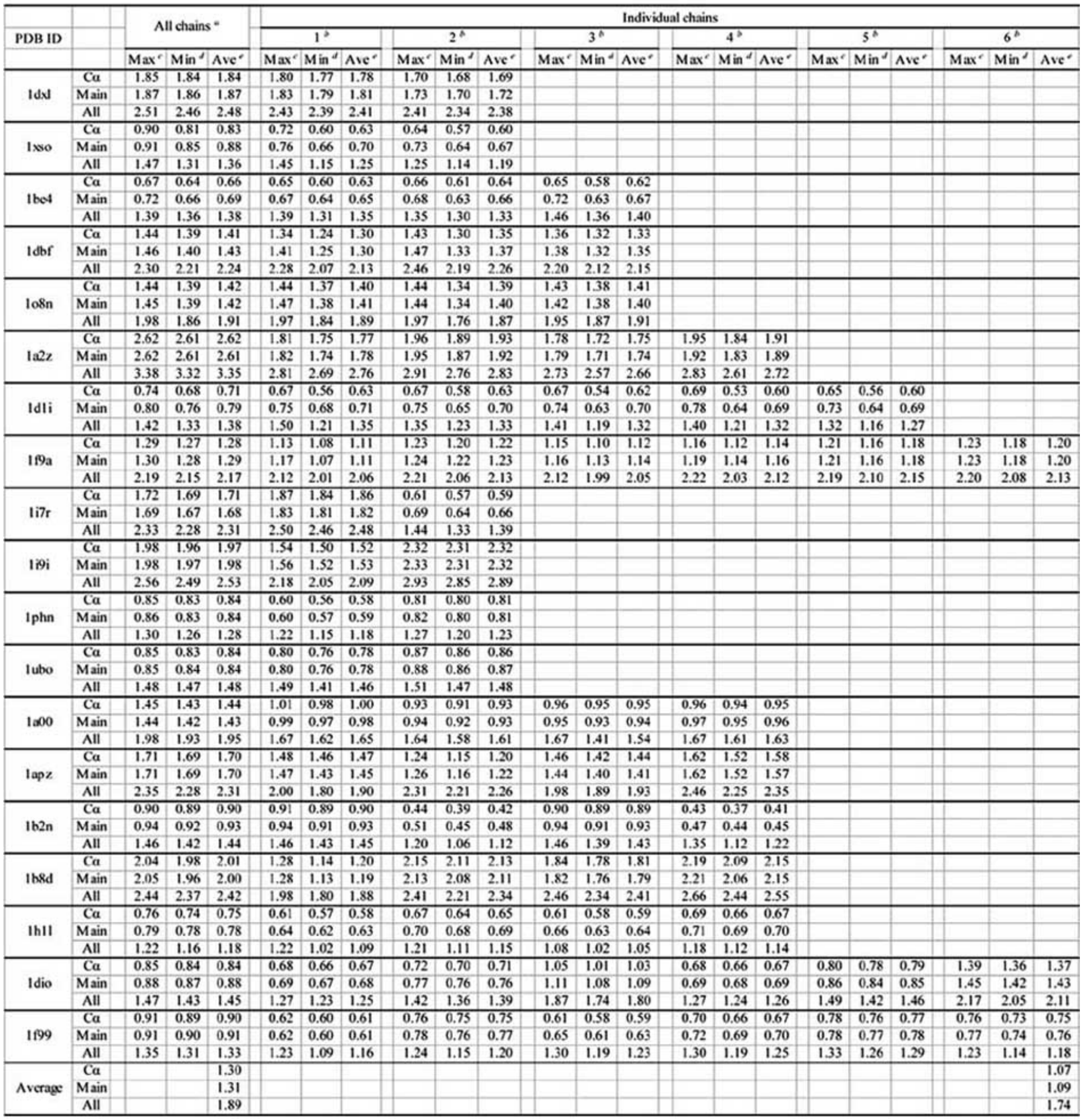
3-1. Evaluation of FAMS Complex

To evaluate the accuracy of FAMS Complex, we blindly constructed protein complex models with 3D structures that were already experimentally known. The test proteins that we used for evaluation of FAMS Complex are listed in Table **1**. The proteins that share more than about 30% sequence identities with the reference proteins were selected as the test proteins from PDB. The proteins whose complex structures were different from those of the reference proteins were not included as test proteins. Searches for reference proteins and generating sequence alignments between target and reference proteins were performed by the Combinatorial Extension (CE) algorithm [10] and PSI-BLAST. We constructed seven models per one target using FAMS Complex because the simulated annealing in the refinement procedure of FAMS Complex gives various solutions. We were unable to compare our results with other methods because no established assessment example for homology modeling of protein complexes exists.

First, we checked the quality of the stereochemistry of the models. No unfavorable contacts between the atoms or unnatural chiral centers were observed, and there were no bad steric clashes between the molecules in the protein complex structures. In the Ramachandran plot of the main-chain φ–ψ angles made by the program PROCHECK [11], almost all of the non-glycine residues were in the most favored or allowed regions. All ω angles were trans-planar.

Second, we superposed the model structure to the corresponding native structure and calculated root mean square deviation (RMSD) between them to check the similarity of the structures. The models were superposed in two ways, one was to superpose all chains (whole protein) and the other was to superpose individual chains. Table **2** shows the RMSD values for Cα atoms, main-chain atoms and all atoms of all chains and individual chains. The average RMSD values for individual chains were 1.07, 1.09 and 1.74 Å for Cα atoms, main-chain atoms and all atoms, respectively, and those for all chains were 1.30, 1.31 and 1.89 Å, respectively. The values for all chains are slightly larger than those for individual chains. This is because the complex structures (orientation of molecules) are slightly

Table 2. Root Mean Square Deviations (Å) for Superposition Between the Models and the X-ray Structures



PDB ID		All chains[a]			Individual chains																	
					1[b]			2[b]			3[b]			4[b]			5[b]			6[b]		
		Max[c]	Min[d]	Ave[e]	Max[c]	Min[d]	Ave[e]	Max[c]	Min[d]	Ave[e]	Max[c]	Min[d]	Ave[e]	Max[c]	Min[d]	Ave[e]	Max[c]	Min[d]	Ave[e]	Max[c]	Min[d]	Ave[e]
1dxl	Cα	1.85	1.84	1.84	1.80	1.77	1.78	1.70	1.68	1.69												
	Main	1.87	1.86	1.87	1.83	1.79	1.81	1.73	1.70	1.72												
	All	2.51	2.46	2.48	2.43	2.39	2.41	2.41	2.34	2.38												
1xso	Cα	0.90	0.81	0.83	0.72	0.60	0.63	0.64	0.57	0.60												
	Main	0.91	0.85	0.88	0.76	0.66	0.70	0.73	0.64	0.67												
	All	1.47	1.31	1.36	1.45	1.15	1.25	1.25	1.14	1.19												
1be4	Cα	0.67	0.64	0.66	0.65	0.60	0.63	0.66	0.61	0.64	0.65	0.58	0.62									
	Main	0.72	0.66	0.69	0.67	0.64	0.65	0.68	0.63	0.66	0.72	0.63	0.67									
	All	1.39	1.36	1.38	1.39	1.31	1.35	1.35	1.30	1.33	1.46	1.36	1.40									
1dbf	Cα	1.44	1.39	1.41	1.34	1.24	1.30	1.43	1.30	1.35	1.36	1.32	1.33									
	Main	1.46	1.40	1.43	1.41	1.25	1.30	1.47	1.33	1.37	1.38	1.32	1.35									
	All	2.30	2.21	2.24	2.28	2.07	2.13	2.46	2.19	2.26	2.20	2.12	2.15									
1o8n	Cα	1.44	1.39	1.42	1.44	1.37	1.40	1.44	1.34	1.39	1.43	1.38	1.41									
	Main	1.45	1.39	1.42	1.47	1.38	1.41	1.44	1.34	1.40	1.42	1.38	1.40									
	All	1.98	1.86	1.91	1.97	1.84	1.89	1.97	1.76	1.87	1.95	1.87	1.91									
1a2z	Cα	2.62	2.61	2.62	1.81	1.75	1.77	1.96	1.89	1.93	1.78	1.72	1.75	1.95	1.84	1.91						
	Main	2.62	2.61	2.61	1.82	1.74	1.78	1.95	1.87	1.92	1.79	1.71	1.74	1.92	1.83	1.89						
	All	3.38	3.32	3.35	2.81	2.69	2.76	2.91	2.76	2.83	2.73	2.57	2.66	2.83	2.61	2.72						
1d1i	Cα	0.74	0.68	0.71	0.67	0.56	0.63	0.67	0.58	0.63	0.67	0.54	0.62	0.69	0.53	0.60	0.65	0.56	0.60			
	Main	0.80	0.76	0.79	0.75	0.68	0.71	0.75	0.65	0.70	0.74	0.63	0.70	0.78	0.64	0.69	0.73	0.64	0.69			
	All	1.42	1.33	1.38	1.50	1.21	1.35	1.35	1.23	1.33	1.41	1.19	1.32	1.40	1.21	1.32	1.32	1.16	1.27			
1f9a	Cα	1.29	1.27	1.28	1.13	1.08	1.11	1.23	1.20	1.22	1.15	1.10	1.12	1.16	1.12	1.14	1.21	1.16	1.18	1.23	1.18	1.20
	Main	1.30	1.28	1.29	1.17	1.07	1.11	1.24	1.22	1.23	1.16	1.13	1.14	1.19	1.14	1.16	1.21	1.16	1.18	1.23	1.18	1.20
	All	2.19	2.15	2.17	2.12	2.01	2.06	2.21	2.06	2.13	2.12	1.99	2.05	2.22	2.03	2.12	2.19	2.10	2.15	2.20	2.08	2.13
1i7r	Cα	1.72	1.69	1.71	1.87	1.84	1.86	0.61	0.57	0.59												
	Main	1.69	1.67	1.68	1.83	1.81	1.82	0.69	0.64	0.66												
	All	2.33	2.28	2.31	2.50	2.46	2.48	1.44	1.33	1.39												
1i9i	Cα	1.98	1.96	1.97	1.54	1.50	1.52	2.32	2.31	2.32												
	Main	1.98	1.97	1.98	1.56	1.52	1.53	2.33	2.31	2.32												
	All	2.56	2.49	2.53	2.18	2.05	2.09	2.93	2.85	2.89												
1phn	Cα	0.85	0.83	0.84	0.60	0.56	0.58	0.81	0.80	0.81												
	Main	0.86	0.83	0.84	0.60	0.57	0.59	0.82	0.80	0.81												
	All	1.30	1.26	1.28	1.22	1.15	1.18	1.27	1.20	1.23												
1ubo	Cα	0.85	0.83	0.84	0.80	0.76	0.78	0.87	0.86	0.86												
	Main	0.85	0.84	0.84	0.80	0.76	0.78	0.88	0.86	0.87												
	All	1.48	1.47	1.48	1.49	1.41	1.46	1.51	1.47	1.48												
1a00	Cα	1.45	1.43	1.44	1.01	0.98	1.00	0.93	0.91	0.93	0.96	0.95	0.95	0.96	0.94	0.95						
	Main	1.44	1.42	1.43	0.99	0.97	0.98	0.94	0.92	0.93	0.95	0.93	0.94	0.97	0.95	0.96						
	All	1.98	1.93	1.95	1.67	1.62	1.65	1.64	1.58	1.61	1.67	1.41	1.54	1.67	1.61	1.63						
1apz	Cα	1.71	1.69	1.70	1.48	1.46	1.47	1.24	1.15	1.20	1.46	1.42	1.44	1.62	1.52	1.58						
	Main	1.71	1.69	1.70	1.47	1.43	1.45	1.26	1.16	1.22	1.44	1.40	1.41	1.62	1.52	1.57						
	All	2.35	2.28	2.31	2.00	1.80	1.90	2.31	2.21	2.26	1.98	1.89	1.93	2.46	2.25	2.35						
1b2n	Cα	0.90	0.89	0.90	0.91	0.89	0.90	0.44	0.39	0.42	0.90	0.89	0.89	0.43	0.37	0.41						
	Main	0.94	0.92	0.93	0.94	0.91	0.93	0.51	0.45	0.48	0.94	0.91	0.93	0.47	0.44	0.45						
	All	1.46	1.42	1.44	1.46	1.43	1.45	1.20	1.06	1.12	1.46	1.39	1.43	1.35	1.12	1.22						
1b8d	Cα	2.04	1.98	2.01	1.28	1.14	1.20	2.15	2.11	2.13	1.84	1.78	1.81	2.19	2.09	2.15						
	Main	2.05	1.96	2.00	1.28	1.13	1.19	2.13	2.08	2.11	1.82	1.76	1.79	2.21	2.06	2.15						
	All	2.44	2.37	2.42	1.98	1.80	1.88	2.41	2.21	2.34	2.46	2.34	2.41	2.66	2.44	2.55						
1h11	Cα	0.76	0.74	0.75	0.61	0.57	0.58	0.67	0.64	0.65	0.61	0.58	0.59	0.69	0.66	0.67						
	Main	0.79	0.78	0.78	0.64	0.62	0.63	0.70	0.68	0.69	0.66	0.63	0.64	0.71	0.69	0.70						
	All	1.22	1.16	1.18	1.22	1.02	1.09	1.21	1.11	1.15	1.08	1.02	1.05	1.18	1.12	1.14						
1dio	Cα	0.85	0.84	0.84	0.68	0.66	0.67	0.72	0.70	0.71	1.05	1.01	1.03	0.68	0.66	0.67	0.80	0.78	0.79	1.39	1.36	1.37
	Main	0.88	0.87	0.88	0.69	0.67	0.68	0.77	0.76	0.76	1.11	1.08	1.09	0.69	0.68	0.69	0.86	0.84	0.85	1.45	1.42	1.43
	All	1.47	1.43	1.45	1.27	1.23	1.25	1.42	1.36	1.39	1.87	1.74	1.80	1.27	1.24	1.26	1.49	1.42	1.46	2.17	2.05	2.11
1f99	Cα	0.91	0.89	0.90	0.62	0.60	0.61	0.76	0.75	0.75	0.61	0.58	0.59	0.70	0.66	0.67	0.78	0.76	0.77	0.76	0.73	0.75
	Main	0.91	0.90	0.91	0.62	0.60	0.61	0.78	0.76	0.77	0.65	0.61	0.63	0.72	0.69	0.70	0.78	0.77	0.78	0.77	0.74	0.76
	All	1.35	1.31	1.33	1.23	1.09	1.16	1.24	1.15	1.20	1.30	1.19	1.23	1.30	1.19	1.25	1.33	1.26	1.29	1.23	1.14	1.18
Average	Cα			1.30																		1.07
	Main			1.31																		1.09
	All			1.89																		1.74

[a] All chains are used in superposition and rmsd calculations.

[b] Chain number used in superposition and rmsd calculations. Chain number is equal to that in Table 1.

[c] Maximum of rmsd values of seven structures.

[d] Minimum of rmsd values of seven structures.

[e] Average value of rmsd values of seven structures.

different between target and reference proteins in some test cases, and RMSD values for all chains include these differences. We previously investi-gated the average RMSD values among the experimental structures of highly similar proteins having sequence identity greater than 95 % [4]; being 0.51, 0.53 and 0.99 Å for Cα atoms, main-chain atoms and all atoms, respectively. The average RMSD values for individual chains of FAMS Complex in Table **2** are the double of those for highly similar proteins. The RMSD calculations of the test proteins in this study show that the accuracy for the FAMS Complex models is generally related to the sequence identity between target and reference proteins. Among the test proteins, 1xso, 1be4, 1d1i, 1phn, 1ubo, 1b2n, 1h1l and 1f99 had very small RMSD values (less than 1 Å for the main-chain atoms) both for individual chains and all chains. These proteins had high sequence identities to the reference proteins (60% or more). The resulting models were very similar to their native structures (Fig. (**2A**)). However, in the case of 1i9i (recombinant monoclonal wild type anti-testosterone Fab fragment), RMSD values of both all and individual chains were not small in spite of the high sequence identity (>70%). The model structures were constructed based on the X-ray structure of the esterase-like catalytic antibody Fab fragment (PDB ID: 1kno) that formed hetero dimer of the light and heavy chains, each of which consisted of variable and constant domains (VL, CL, VH and CH) (Fig. (**2B**)). When the four domains of the model were individually superposed to the native structures, the averaged RMSD values for the Cα atoms of VL, CL, VH and CH were 1.57, 0.63, 2.00 and 1.63 Å, respectively. RMSD values for variable domains were larger than those for constant domains. Because the structures of the variable domains (especially in the complementarity determining regions) were different between 1i9i and 1kno, RMSD values were large in this test case. In the case of 1a2z (pyrrolidone carboxyl peptidase from *Thermococcus litoralis*), RMSD values of all chains were larger than those of individual chains. The models were constructed based on the X-ray structure of pyrrolidone carboxyl peptidase from *Pyrococus horikoshii* (PDB ID: 1iu8) (Fig. (**2C**)). 1a2z and 1iu8 were similar in monomer structure, but were slightly different in complex structure, therefore RMSD values of all chains were higher. The active site region including the catalytic triad of Glu80, Cys143 and His167 was accurately modeled (Fig. (**2D**)).

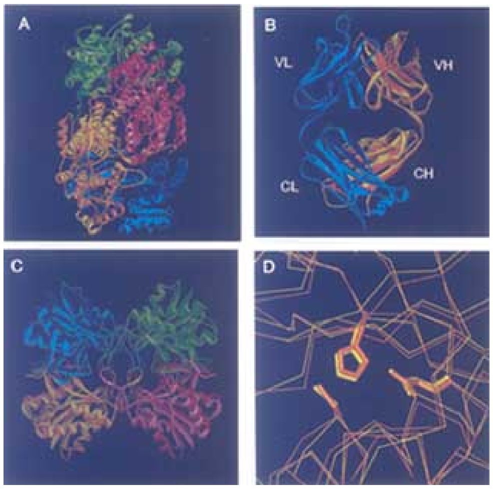


Fig. (2). Superposition of models and their corresponding native structures.

The chains of models are shown in cyan, orange, yellow green and red, and the chains of the experimental structures are shown in blue, yellow, green and pink.

(A) Nitrogenase MoFe protein. (B) Recombinant anti-testosterone Fab fragment. (C) Pyrrolidone carboxyl peptidase. (D) Active site region of pyrrolidone carboxyl peptidase. Catalytic triad is shown in bold sticks.

Third, we checked the main-chain structures. Table **3** shows the percentage of correctly modeled hydrogen bonds between the main-chain atoms in comparison with the native structures. In the present paper, the hydrogen bond was defined by the methods of Kabasch and Sander [12]. The percentages were calculated for the individual chains and all chains. In the calculation for all chains, the hydrogen bonds of inter chains were included. The average percentages of correctly modeled hydrogen bonds were 84.4 and 83.9 % for individual and all chains, respectively. These values indicate that most of the hydrogen bonds were correctly predicted both intra and inter chains.

Fourth, side-chain conformations were evaluated by comparing the side-chain $\chi 1$ and $\chi 2$ torsional angles with those in the native structures for the residues within 1.0 and 3.5 Å in the MaxSub structure alignment [5,13,14]. Side-chain conformations were considered correct if $\chi 1$ and $\chi 2$ were within 30° of the experimental structure values. Table **4** shows high percentages of correct side-chain conformations in the model structures. The average percentages for the residues within 1.0 Å were slightly higher than those for the residues within 3.5 Å. This indicates that the side-chain conformations in the structurally conserved region can be more accurately predicted. The important regions, such as active sites, tend to be in the structurally conserved regions, so that side chains in important regions tend to be modeled more accurately.

3-2. Application of FAMS Complex

As an application of FAMS Complex to predict a protein complex structure whose 3D structure had not been solved experimentally, we constructed human RNA polymerase II using FAMS Complex. Searching for reference proteins and sequence alignments was performed using PSI-BLAST. The reference structure was the X-ray structure of 10-subunit yeast RNA polymerase II (PDB ID: 1i5o). The sequence identities of ten subunits of RNA polymerase II (Rbp1, Rbp2, Rbp3, Rbp5, Rbp6, Rbp8, Rbp9, Rbp10, Rbp11 and Rbp12) between human and yeast were 50, 56, 45, 44, 71, 34, 43, 70, 50 and 52 %, respectively. These high sequence identities suggested that the human RNA polymerase II could be constructed based on the X-ray structure of yeast RNA polymerase II and that the alignments of the subunits were reliable. Therefore, we constructed 10-subunit human RNA polymerase II using FAMS Complex. The model structure is shown in Fig. (**3**). The model comprises of 3656 amino acid residues. In the model structure, no unfavorable contacts between the atoms and no unnatural chiral centers were observed, and there were no bad steric clashes between the ten subunits in the model. Main-chain torsionalangles were evaluated by the program PROCHECK. This homology modeling is a good example showing FAMS Complex to be useful in the case of large protein-protein complex structures.

Another application for FAMS Complex is a homology modeling of the severe acute respiratory syndrome corona-

Table 3. Percentages (%) of the Correctly Modeled Hydrogen Bonds Between Main-Chain Atoms in the Models

PDB ID	All chains [a]–			Individual chains																	
				1 [b]			2 [b]			3 [b]			4 [b]			5 [b]			6 [b]		
	Max [c]	Min [d]	Ave [e]	Max [c]	Min [d]	Ave [e]	Max [c]	Min [d]	Ave [e]	Max [c]	Min [d]	Ave [e]	Max [c]	Min [d]	Ave [e]	Max [c]	Min [d]	Ave [e]	Max [c]	Min [d]	Ave [e]
1dxl	77.0	73.8	75.7	77.9	73.9	76.3	77.7	72.1	75.5												
1xso	80.7	71.7	77.2	82.9	72.0	77.2	86.3	70.0	77.7												
1be4	83.1	78.4	80.6	88.0	77.1	83.3	86.6	78.0	82.4	81.1	72.2	76.5									
1dbf	80.1	76.2	77.5	84.6	63.1	78.2	88.1	70.1	77.2	83.8	73.5	79.0									
1o8n	82.2	73.6	78.8	84.5	73.8	81.4	82.7	71.2	77.9	81.0	77.1	78.5									
1a2z	80.0	74.0	77.8	81.6	74.4	78.4	81.1	73.2	76.9	81.0	70.6	76.8	82.5	75.4	79.1						
1d1i	86.0	78.3	82.2	89.7	74.4	83.2	90.2	73.2	81.2	85.4	73.2	80.8	82.9	73.2	77.0	92.3	74.4	82.8			
1f9a	77.6	72.6	75.8	80.8	71.7	75.9	80.8	70.7	77.8	79.2	71.3	74.5	75.2	66.3	71.3	83.0	71.0	76.4	82.0	75.0	78.1
1i7r	93.2	90.5	91.8	92.5	89.0	90.8	95.7	93.6	95.4												
1i9i	87.2	79.6	82.4	88.5	80.3	83.7	85.8	77.9	80.9												
1phn	93.4	90.5	92.4	96.6	92.2	94.1	92.1	89.0	90.9												
1ubo	88.1	85.8	86.9	87.2	83.0	85.7	88.8	85.6	87.3												
1a00	91.9	89.5	90.5	96.1	92.2	94.0	91.6	85.0	88.1	96.1	92.2	93.8	87.9	84.1	86.2						
1apz	82.3	80.1	81.3	85.0	80.0	82.3	83.1	75.3	79.0	86.3	81.3	84.1	79.5	73.1	77.3						
1b2n	85.1	81.0	82.8	85.8	78.2	82.3	96.3	92.6	93.1	83.7	76.8	80.9	96.2	88.5	93.4						
1b8d	90.8	89.2	89.9	90.1	86.5	88.5	92.5	90.0	91.2	90.2	87.5	88.8	91.8	89.3	90.7						
1h1l	89.0	87.9	88.6	89.1	87.7	88.3	90.4	88.2	89.1	89.3	86.5	88.0	90.2	88.0	88.8						
1dio	90.9	88.7	89.6	92.6	87.9	90.0	92.3	90.4	90.9	86.2	82.8	85.1	92.3	89.9	91.0	90.3	85.4	87.7	90.8	85.1	87.5
1f99	92.9	91.7	92.4	97.5	91.5	94.2	96.0	92.7	94.7	94.9	92.4	93.3	95.8	93.2	94.2	91.4	87.5	89.6	91.3	87.3	89.8
Average			83.9																		84.4

[a] All chains are used in calculation of percentage.

[b] Chain number used in calculation of percentage. Chain number is equal to that in Table 1.

[c] Maximum of percentage of seven structures.

[d] Minimum of percentage of seven structures.

[e] Average value of percentage of seven structures.

virus main protease (SARS-CoV M^{pro}). From late 2002 to early 2003, SARS spread to several countries. Though the 3D structure of the SARS-CoV M^{pro} was needed to accelerate the discovery of new drugs, it was not solved at the time when the first complete genome sequences of the SARS-CoV were reported on May 1, 2003 [15,16]. Therefore, we constructed a homodimer model of SARS-CoV M^{pro} using FAMS Complex and released the model structure at http://www.pd-fams.com/ on May 9, 2003. Our model was, to our knowledge, the first structure that was opened to public access. Many scientists started to carry out structure-based drug design using our model without waiting for the X-ray structure to be solved. After the X-ray structure was solved, our model proved to be an accurate prediction [17]. This homology modeling is a good example that FAMS Complex contributed to accelerating drug discovery.

4. DISCUSSION

In this study, FAMS Complex was developed by improving the procedures of FAMS, the accuracy of which was demonstrated in the recent CASP and CAFASP blind contests of protein structure prediction. The evaluation of FAMS Complex showed that FAMS Complex could construct multi-chain protein models as accurately as FAMS constructs single-chain protein models. The 3D structural data of protein-protein interactions predicted by FAMS Complex will provide a deeper understanding of protein functions and the mechanisms of diseases at the atomic level.

Table 4. Percentages (%) of Correct Side-Chain Conformations in the Models

PDB ID		3.5 Å [a]			1.0 Å [b]		
		Max [c]	Min [d]	Ave [e]	Max [c]	Min [d]	Ave [e]
1dxl	$\chi1$	56.8	54.4	55.7	63.2	58.6	60.8
	$\chi1,\chi2$	56.0	49.8	53.1	61.9	54.3	58.3
1xso	$\chi1$	82.3	78.2	80.6	84.4	80.1	82.4
	$\chi1,\chi2$	78.5	73.6	76.9	80.5	78.4	79.6
1be4	$\chi1$	66.7	64.6	65.5	67.0	65.0	65.9
	$\chi1,\chi2$	64.8	61.6	63.1	65.9	62.7	64.4
1dbf	$\chi1$	67.3	63.8	65.6	69.1	66.8	67.8
	$\chi1,\chi2$	66.9	62.9	65.2	70.2	65.5	67.0
1o8n	$\chi1$	70.5	66.8	68.8	78.0	74.9	76.1
	$\chi1,\chi2$	68.2	64.1	65.8	70.6	67.5	69.0
1a2z	$\chi1$	69.6	66.9	68.3	74.8	72.4	73.3
	$\chi1,\chi2$	67.6	63.3	65.4	70.1	64.2	68.2
1d1i	$\chi1$	73.7	69.1	71.8	76.3	71.2	74.2
	$\chi1,\chi2$	63.0	56.4	59.7	64.1	56.5	60.9
1f9a	$\chi1$	70.1	67.5	69.2	74.1	71.1	73.0
	$\chi1,\chi2$	71.6	67.2	69.3	72.2	67.4	69.8
1i7r	$\chi1$	68.2	64.8	65.9	71.0	67.4	68.9
	$\chi1,\chi2$	66.3	63.8	65.4	68.4	65.9	67.2
1i9i	$\chi1$	66.1	64.7	65.4	70.1	65.8	67.9
	$\chi1,\chi2$	66.9	59.1	63.5	69.0	59.2	62.6
1phn	$\chi1$	78.5	76.7	77.6	79.3	76.5	78.0
	$\chi1,\chi2$	85.5	80.7	83.3	86.7	82.7	84.8
1ubo	$\chi1$	79.0	76.9	77.9	80.7	78.8	79.8
	$\chi1,\chi2$	69.6	66.6	68.5	70.1	67.1	69.0
1a00	$\chi1$	69.7	65.6	67.4	74.5	70.2	71.6
	$\chi1,\chi2$	70.8	64.2	67.6	76.2	68.5	72.5
1apz	$\chi1$	60.1	57.3	58.7	66.2	60.9	63.3
	$\chi1,\chi2$	66.9	61.9	64.6	69.8	65.0	66.6
1b2n	$\chi1$	77.8	75.4	76.8	79.7	77.4	78.4
	$\chi1,\chi2$	64.7	62.6	63.4	65.7	63.5	64.4
1b8d	$\chi1$	71.7	67.0	68.4	75.6	71.1	72.7
	$\chi1,\chi2$	66.8	62.2	64.6	66.2	59.2	63.8

(Table 4. Contd....)

PDB ID		3.5 Å [a]			1.0 Å [b]		
		Max [c]	Min [d]	Ave [e]	Max [c]	Min [d]	Ave [e]
1h1l	$\chi1$	83.9	82.5	83.3	86.7	84.6	85.7
	$\chi1,\chi2$	85.1	82.6	84.2	85.9	83.7	85.0
1dio	$\chi1$	71.1	69.0	69.9	75.1	72.9	73.8
	$\chi1,\chi2$	70.5	68.3	69.3	72.1	70.4	71.3
1f99	$\chi1$	71.9	69.8	70.6	74.4	71.8	72.8
	$\chi1,\chi2$	70.4	67.8	69.1	71.2	68.3	69.8
Average	$\chi1$			71.1			74.9
	$\chi1,\chi2$			69.4			71.4

[a] Value of the distance threshold of Maxsub is 3.5 Å.

[b] Value of the distance threshold of Maxsub is 1.0 Å.

[c] Maximum of percentages of seven structures.

[d] Minimum of percentages of seven structures.

[e] Averaged value of percentages of seven structures.

FAMS Complex is a fully automated system, requires only sequences and alignments of target protein as input, and so constructs all components simultaneously and automatically. Before FAMS Complex was developed, the procedure of homology modeling of protein-protein complex was complicated: first, individual molecules were modeled one by one, then individual modeled molecules were superimposed to their equivalents in the reference complex structures, and finally superimposed molecules were optimized to delete any bad inter molecular steric clashes [18]. FAMS Complex will make it possible for a wider audience of non-experts to easily use homology modeling of protein complexes.

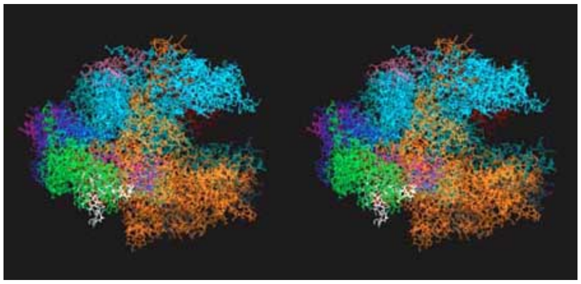


Fig. (3). Stereo view of human RNA polymerase II.

Rbp1, Rbp2, Rbp3, Rbp5, Rbp6, Rbp8, Rbp9, Rbp10, Rbp11 and Rbp12 are shown in cyan, orange, green, red, pink, magenta, gray, white, blue and purple, respectively.

The accuracy of models depends not only on the abilities of the modeling software but also on sequence identity between target and reference proteins. As for modeling individual proteins, Baker and Sali indicated that generally errors in alignment or template selection increase for low sequence identities (less than about 30 %), and the quality of the model becomes poor [17]. In the case of modeling protein complexes, similarity of the orientations of molecules between the target and reference proteins is an additional important factor for model accuracy. Aloy *et al*. investigated the relationship between sequence and interaction divergence in proteins [18,19]. Their results indicated that interactions between proteins tend to be similar when the sequence identity is above approximately 30-40%. These results of individual and complex structures suggest that the accuracy of the model is related to the sequence identity between target and reference proteins in both cases. When sequence identity is low, evaluation of the resulting model is necessary, and refinement of the orientations of molecules using a method such as docking must be performed using the homology model for the initial coordinates.

Because FAMS Complex is a fully automated system and so suited for large-scale genome wide modeling, it can make a great contribution to proteomics in the post-genomic era. A major goal in the post-genomic era is to determine protein-protein interaction networks on a genomic scale. Computational methods that can identify protein interactions and their associated networks through known 3D complex structures have been developed [21-23]. These methods provide the putative interaction of sequences; thus given putative interaction networks of sequences on known 3D complex structures, FAMS Complex can automatically construct protein-protein complex structures. It is clear that structural information obtained by FAMS Complex is much more useful than sequence information alone to understand protein-protein interaction networks. FAMS Complex is undoubtedly an essential tool in this post-genomic era in which the focus has been shifted from the structures of individual proteins to the structures of large assemblies. Because the number of known protein-protein complex structures (reference structures) is increasing [20], FAMS Complex is destined to become a even more powerful and useful tool.

ACKNOWLEDGEMENTS

We thank Drs. Kazuyoshi Ebisawa, Mitsuo Iwadate, Hirokazu Tanaka and Kazuhiko Kanou for helpful discussions. This work was supported by Kitasato University Research Grant for Young Researchers.

REFERENCES

[1] Yoneda, T.; Komooka, H.; Umeyama, H. *J. Protein Chem.* **1997**, *16*, 597.

[2] Takeda-Shitaka, M.; Terashi, G.; Takaya, D.; Kanou, K.; Iwadate, M.; Umeyama, H. *Proteins*, in press.

[3] Takeda-Shitaka, M.; Takaya, D.; Chiba, C.; Tanaka, H.; Umeyama, H. *Curr. Med. Chem.*, **2004**, *11*, 551.

[4] Ogata, K.; Umeyama, H. *J. Mol. Graphics Mod.*, **2000**, *18*, 258.

[5] Fischer, D.; Elofsson, A.; Rychlewski, L.; Pazos, F.; Valencia, A.; Rost, B.; Ortiz, A.R.; Dunbrack R.L. Jr. *Proteins*, **2001**, (Suppl. 5), 171.

[6] Iwadate, M.; Ebisawa, K.; Umeyama, H. *Chem-Bio. Info. J.*, **2001**, *4*, 136.

[7] Ogata, K.; Umeyama, H. *Proteins*, **1998**, *31*, 355.

[8] Altchul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. *Nucleic Acid Res.*, **1997**, *25,* 3389.

[9] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. *Nucleic. Acids. Res.*, **2000**, *28*, 235.

[10] Shindyalov, I.N.; Bourne, P.E. *Protein Eng.*, **1998**, *11*, 739.

[11] Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. *J. Appl. Cryst.*, **1993**, *26*, 283.

[12] Kabsch, W.; Sander, C. *Biopolymers*, **1983**, *22*, 2577.

[13] Siew, N.; Elosson, A.; Rychlewski, L.; Fischer, D. *Bioinformatics*, **2000**, *16*, 776.

Received: 31 March, 2005 Revised: 30 September, 2005 Accepted: 01 October, 2005

[14] Fischer, D.; Rychlewski, L.; Dunbrack, R.L. Jr.; Ortiz, A.R.; Elofsson, A. *Proteins*, **2003**, *53*, 503.

[15] Rota, P.A.; Oberste, M.S.; Monroe, S.S.; Nix, W.A.; Campagnoli, R.; Icenogle, J.P.; Peñaranda, S.; Bankamp, B.; Maher, K.; Chen, M.; Tong, S.; Tamin, A.; Lowe, L.; Frace, M.; DeRisi, J.L.; Chen, Q.; Wang, D.; Erdman, D.D.; Peret, T.C.T.; Burns, C.; Ksiazek, T.G.; Rollin, P.E.; Sanchez, A.; Liffick, S.; Holloway, B.; Limor, J.; McCaustland, K.; Olsen-Rasmussen, M.; Fouchier, R.; Günther, S.; Osterhaus, A.D.M.E.; Drosten, C.; Pallansch, M.A.; Anderson, L.J.; Bellini, W.J. *Science*, **2003**, *300*, 1394; published online 1 May 2003 (10.1126/Science.1085952).

[16] Marra, M.A.; Jones, S.J.; Astell, C.R.; Holt, R.A.; Brooks-Wilson, A.; Butterfield, Y.S.; Khattra, J.; Asano, J.K.; Barber, S.A.; Chan, S.Y.; Cloutier, A.; Coughlin, S.M.; Freeman, D.; Girn, N.; Griffith, O.L.; Leach, S.R.; Mayo, M.; McDonald, H.; Montgomery, S.B.; Pandoh, P.K.; Petrescu, A.S.; Robertson, A.G.; Schein, J.E.; Siddiqui, A.; Smailus, D.E.; Stott, J.M.; Yang, G.S.; Plummer, F.; Andonov, A.; Artsob, H.; Bastien, N.; Bernard, K.; Booth, T.F.; Bowness, D.; Czub, M.; Drebot, M.; Fernando, L.; Flick, R.; Garbutt, M.; Gray, M.; Grolla, A.; Jones, S.; Feldmann, H.; Meyers, A.; Kabani, A.; Li, Y.; Normand, S.; Stroher, U.; Tipples, G.A.; Tyler, S.; Vogrig, R.; Ward, D.; Watson, B.; Brunham, R.C.; Krajden, M.; Petric, M.; Skowronski, D.M.; Upton, C.; Roper, R.L. *Science*, **2003**, *300*, 1399; published online 1 May 2003 (10.1126/Science.1085953).

[17] Takeda-Shitaka M.; Nojima H., Takaya D.; Kanou K.; Iwadate M.; Umeyama H. *Chem. Pharm. Bull.*, **2004**, *52*, 643.

[18] Yoneda, T.; Yoneda, S.; Takayama, N.; Kitazawa, M.; Umeyama, H. *J. Mol. Graphics Mod.*, **1999**, *17*, 114.

[19] Baker, D.; Sali, A. *Science*, **2001**, *294*, 93.

[20] Aloy, P.; Ceulemans, H.; Stark, A.; Russell, R.B. *J. Mol. Biol.*, **2003**, *332*, 989.

[21] Aloy, P.; Pichaud, M.; Russell, R.B. *Curr. Opin. Struct. Biol.*, **2005**, *15*, 15.

[22] Aloy, P.; Russell, R.B. *Bioinformatics*, **2003**, *19*, 161.

[23] Lu, L.; Lu, H.; Skolnick, J. *Proteins*, **2002**, *49*, 350.

[24] Russell, R.B.; Alber, F.; Aloy, P.; Davis, F.P.; Korkin, D.; Pichaud, M.; Topf, M.; Sali, A. *Curr. Opin. Struct. Biol.*, **2004**, *14*, 313.

[25] Aloy, P.; Russell, R.B. *Nat. Biotechnol.*, **2004**, *22*, 1317.