Evaluation of Homology Modeling of the Severe Acute Respiratory Syndrome (SARS) Coronavirus Main Protease for Structure Based Drug Design

Mayuko TAKEDA-SHITAKA,* Hiroyuki NOJIMA, Daisuke TAKAYA, Kazuhiko KANOU, Mitsuo IWADATE, and Hideaki UMEYAMA

School of Pharmaceutical Sciences, Kitasato University; 5–9–1 Shirokane, Minato-ku, Tokyo 108–8641, Japan.
Received January 26, 2004; accepted March 10, 2004

To accelerate the development of drugs against severe acute respiratory syndrome (SARS), we constructed a homology model of the SARS coronavirus main protease using our modeling software, FAMS Ligand&Complex, and released it before the X-ray structure was solved. The X-ray structure showed our model as accurately predicted and useful for structure based drug design.

Key words homology modeling; severe acute respiratory syndrome (SARS); normal mode analysis; structure based drug design

The first case of severe acute respiratory syndrome (SARS) has been identified as being in China in late 2002, and thereafter SARS rapidly spread to many countries in early 2003. Scientists in all areas are now working to develop effective drugs against SARS, because it is feared that there could be another outbreak of SARS either in this or the next winter season. The cause of SARS was identified as a novel coronavirus; its main protease (SARS-CoV M^{pro}) is the primary target for drugs because it plays an important role in virus replication. Though the three-dimensional (3D) structure of the SARS-CoV M^{pro} is necessary to accelerate the discovery of new drugs, the 3D structure was not solved experimentally when the first complete genome sequences of the SARS-CoV were reported on May 1, 2003.[1,2] Therefore, we constructed a homology model of SARS-CoV M^{pro} using the computer software FAMS Ligand&Complex [Takeda-Shitaka *et al.*, in preparation] and released the model structure at http://www.pd-fams.com/ on May 9, 2003.[3] Many scientists started to carry out structure based drug design using our model without waiting for the X-ray structure to be solved. After our model was released, the X-ray structure of SARS-CoV M^{pro} of Urbani strain was released from the Protein Data Bank (PDB)[4] (PDB ID: 1Q2W) on July 29, 2003. In this paper, we compare our model with the X-ray structure of SARS-CoV M^{pro} to evaluate the usefulness of the model in structure based drug design.

We constructed SARS-CoV M^{pro} model based on the sequence of Urbani strain.[1] Searching for reference protein and sequence alignment was carried out by reverse PSI-BLAST (RPS-BLAST).[5] The X-ray structure of the transmissible gastroenteritis coronavirus (TGEV) M^{pro} (PDB ID: 1LVO) was selected as the reference protein for SARS-CoV M^{pro}. Sequence identity between SARS-CoV M^{pro} and TGEV M^{pro} was 44.5%, and E-value (8×10^{-96}) was low enough that TGEV M^{pro} could be used as the reference protein. There were three 1- or 2-residue insertions in SARS-CoV M^{pro}. Because the reference protein forms a tight dimer in the crystal,[6] we constructed a homodimer model of SARS-CoV M^{pro} based on the alignment using FAMS Ligand&Complex.

First, we checked the quality of the stereochemistry of our model. No unfavorable contacts between the atoms and no unnatural chiral centers were observed, and there were no bad steric clashes that prevented close interaction of the two monomers in the dimer. In the Ramachandran plot of the main-chain ϕ–φ angles made by the program PROCHECK,[7] almost all of the nonglycine residues were in the most favored or allowed regions. All the ω angles were trans-planar.

Second, we superimposed our model structure to the X-ray structure to check the similarity of the structures. Table 1 and Fig. 1 show that our model structure is very similar to the X-ray structure. As we predicted, the X-ray structure of SARS-CoV M^{pro} forms a homodimer. Each monomer molecule has three domains I, II and III (residues 8—101, 102—184 and 201—301, respectively). Domains I and II (active site domain) are β-barrel domains. The active site region is located in a cleft between domains I and II. Domain III is an α-helical domain. As shown in Table 1, root mean square deviations (rmsds) for each domain (active site domain and α-helical domain) are small. Especially in the active site region, which is the most important region for structure based drug design, rmsd is very small. This means that the active site region is quite accurately predicted. Rmsds for all residues (monomer and homodimer) are relatively higher than those of the active site region, and the active site and α-helical domains, probably because the dimer structure between two monomers and the domain structure between the active site and α-helical domains of the X-ray structure (PDB ID: 1Q2W) are slightly different from those of the reference structure (PDB ID: 1LVO).

Anand *et al.* predicted a monomer homology model of SARS-CoV M^{pro} using Insight II (Accelrys Inc.), and reported it in the journal *Science* online on May 13, 2003.[8] Their model was released at PDB (PDB ID: 1P9T) on May 20, 2003. We superimposed their model structure to the X-ray structure and calculated the rmsds (Table 1). The result

Table 1. Root Mean Square Deviations for Superposition between Models and the X-Ray Structure of SARS-CoV M^{pro}

Superimposed residues[a]	rmsd (Å)			
	Cα atom		All atom	
	Our model[b]	1P9T	Our model[b]	1P9T
Active site region[c]	0.78[d]	1.24[e]	1.42[d]	1.66[e]
Active site domain (domains I and II)[f]	1.16[d]	1.99[e]	1.89[d]	2.55[e]
α-Helical domain (domain III)[g]	1.68[d]	2.75[e]	2.67[d]	3.60[e]
All residues (monomer)	2.54[d]	3.48[e]	3.10[d]	3.98[e]
All residues (homodimer)	2.82[h]	—	3.31[h]	—

a) Residues that are used in superposition and rmsd calculations. *b*) We constructed six homodimer models because the simulated annealing in the refinement procedure of FAMS Ligand&Complex gives various solutions. *c*) Catalytic and binding site residues 41, 140, 143—145, 161—168, 172 and 187—191. *d*) Averaged value of twelve rmsds between twelve monomers in six dimer models and the corresponding chain of the X-ray structure. *e*) Averaged value of two rmsds between chain A of 1P9T and chains A and B of 1Q2W. *f*) Residues 8—184. *g*) Residues 201—301. *h*) Averaged value of six rmsds between six dimer models and one dimer X-ray structure.

* To whom correspondence should be addressed. e-mail: shitakam@pharm.kitasato-u.ac.jp

shows that the rmsds of our model are smaller than those of their model; indicating that our model is more accurate than theirs. In homology modeling, the accuracy of the model generally depends on the quality of the alignment and on the model construction procedure. In the case of SARS-CoV M^{pro}, the accuracy of the model depends mainly on the latter, since there is very little error in the alignment because of the high sequence identity between the target and reference proteins. The good result of our modeling study probably comes from the ability of our software FAMS Ligand&Complex to accurately predict structure. FAMS Ligand&Complex was developed by improving the procedures of FAMS,[9] a full automatic modeling system for individual proteins. The high ability of FAMS, and also FAMS Ligand&Complex, are mainly derived from the optimization protocol based on iterative cycles of side chain and main chain optimization. This protocol enables the conservation of the side chain torsional angles and main chain conformation within homologous proteins.

Third, we carried out normal mode analyses of our model,

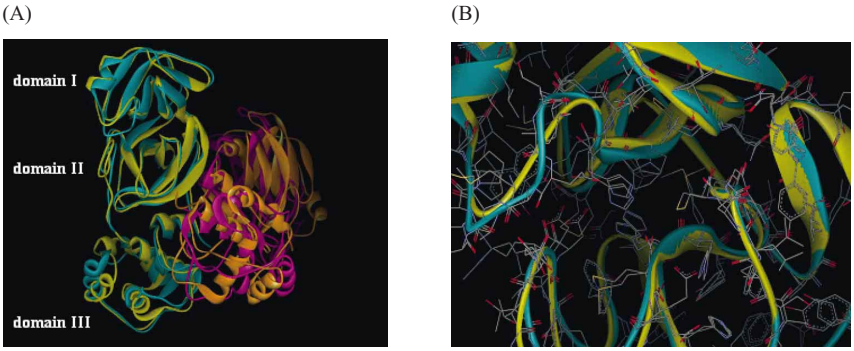

Fig. 1. Superposition of Our Model and the X-Ray Structure of SARS-CoV M^{pro}

Our model is shown in cyan and magenta. The X-ray structure is shown in yellow and orange. (A) Homodimer structure of SARS-CoV M^{pro}. (B) Active site region of SARS-CoV M^{pro}.

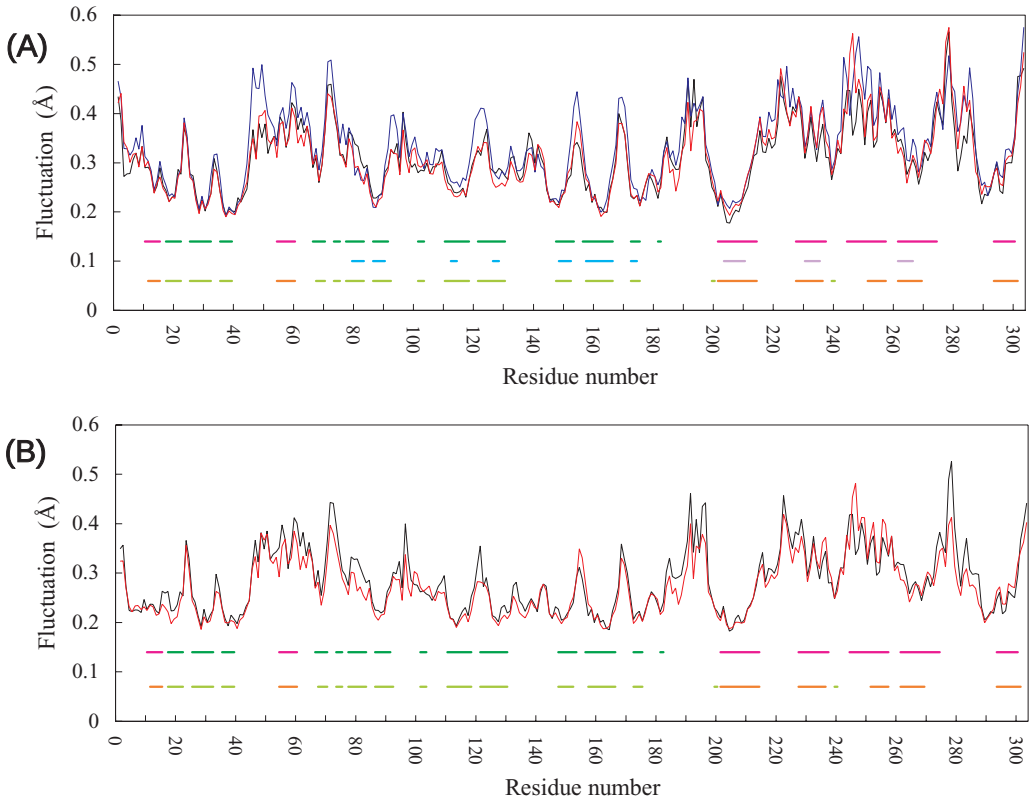

Fig. 2. Fluctuations of Cα Atoms of the Models and the X-Ray Structure of SARS-CoV M^{pro}

The red, blue and black lines denote Cα-atom fluctuations of our model, Anand's model (PDB ID: 1P9T) and the X-ray structure (PDB ID: 1Q2W), respectively. Secondary structure assignment (α-helix or β-sheet) was done by STRIDE.[19] The magenta, pink and orange lines denote the α-helix regions of our model, Anand's model and the X-ray structure, respectively. The green, cyan and yellow green lines denote the β-sheet regions of our model, Anand's model and the X-ray structure, respectively. (A) Fluctuations in the monomer structures of our model, Anand's model and the X-ray structure. In our model and the X-ray structure, the fluctuations were determined by averaging fluctuation-data of six energy-optimized structures of one molecule and those of the other molecule (averaging total twelve energy-optimized structures). In Anand's model, the fluctuations were determined by averaging fluctuation-data of twelve energy-optimized structures of a monomer. (B) Fluctuations in the homodimer structures of our model and the X-ray structure. To examine the inner motion of each molecule, the Eckart's conditions were applied. The fluctuations were determined by averaging fluctuation-data of six energy-optimized structures of one molecule and those of the other molecule (averaging total twelve energy-optimized structures).

Anand's model and the X-ray structure by the method using dihedral angles described in earlier papers from our laboratory[10—15] to evaluate the dynamic characteristics of the models. Before carrying out normal mode analyses, some residues in the X-ray structure, which were disordered and not included in 1Q2W, were modeled using FAMS Ligand&Complex and the CHIMERA modeling system.[16,17] The normal mode analyses were carried out on both homodimer and monomer structures. As shown in Fig. 2A, the results of normal mode analyses for monomer structures of our model, Anand's model and the X-ray structure indicated that the dynamic characteristics of our model were very similar to those of the X-ray structure; but that Anand's model was not as stable as the X-ray structure. Anand's model is very unstable in the loop regions and those regions where the secondary structures (α-helix and β-sheet) are broken. In our model, the secondary structures are accurately predicted. The normal mode analyses for homodimer structures were carried out on our model and the X-ray structure because Anand's model is a monomer. To examine the inner motion of each molecule, the Eckart's conditions[18] were applied. As shown in Fig. 2B, the dynamic characteristics of our model are similar to those of the X-ray structure. These results showing the similarity of our model to the X-ray structure (Figs. 2A and B) are very significant, because proteins must be treated flexible in consideration of the induced fit concept when protein–ligand or protein–protein docking is carried out.

In conclusion, our model proved to be accurately predicted. Moreover, our model is, to our knowledge, the first structure that was opened to public access. It is sure that structural information obtained from our model was much more useful than sequence information alone before the X-ray structure was solved. There are many target proteins like SARS-CoV M^{pro} for which the sequence is available but the 3D structure has not been solved; therefore, to accelerate drug discoveries, homology modeling methods that can construct models accurately (and quickly if possible) are indispensable for structure based drug design.

References

1) Rota P. A., Oberste M. S., Monroe S. S., Nix W. A., Campagnoli R., Icenogle J. P., Penaranda 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., Gunther S., Osterhaus A. D. M. E., Drosten C., Pallansch M. A., Anderson L. J., Bellini W. J., *Science*, **300**, 1394—1399 (2003); published online 1 May 2003 (10.1126/Science.1085952).
2) Marra M. A., Jones S. J. M., Astell C. R., Holt R. A., Brooks-Wilson A., Butterfield Y. S. N., 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*, **300**, 1399—1404 (2003); published online 1 May 2003 (10.1126/Science.1085953).
3) Takeda-Shitaka M., Takaya D., Chiba C., Tanaka H., Umeyama H., *Curr. Med. Chem.*, **11**, 551—558 (2004).
4) Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., Bourne P. E., *Nucleic Acids Res.*, **28**, 235—242 (2000).
5) Marchler-Bauer A., Panchenko A. R., Shoemaker B. A., Thiessen P. A., Geer L. Y., Bryant S. H., *Nucleic Acids Res.*, **30**, 281—283 (2002).
6) Anand K., Palm G. J., Mesters J. R., Siddell S. G., Ziebuhr J., Hilgenfeld R., *EMBO J.*, **21**, 3213—3224 (2002).
7) Laskowski R. A., McArthur M. W., Moss D. S., Thornton J. M., *J. Appl. Crystallogr.*, **26**, 283—291 (1993).
8) Anand K., Ziebuhr J., Wadhwani P., Mesters J. R., Hilgenfeld R., *Science*, **300**, 1763—1767 (2003); published online 13 May 2003 (10.1126/Science.1085658).
9) Ogata K., Umeyama H., *J. Mol. Graphics. Mod.*, **18**, 258—272 (2000).
10) Takeda-Shitaka M., Kamiya K., Miyata T., Ohkura N., Madoiwa S., Sakata Y., Umeyama H., *Chem. Pharm. Bull.*, **47**, 322—328 (1999).
11) Nojima H., Takeda-Shitaka M., Kurihara Y., Adachi M., Yoneda S., Kamiya K., Umeyama H., *Chem. Pharm. Bull.*, **50**, 1209—1214 (2002).
12) Kamiya K., Sugawara Y., Umeyama H., *J. Comput. Chem.*, **24**, 826—841 (2003).
13) Nojima H., Takeda-Shitaka M., Kurihara Y., Kamiya K., Umeyama H., *Chem. Pharm. Bull.*, **51**, 923—928 (2003).
14) Kurihara Y., Watanabe T., Nojima H., Takeda-Shitaka M., Sumikawa H., Kamiya K., Umeyama H., *Chem. Pharm. Bull.*, **51**, 754—758 (2003).
15) Adachi M., Kurihara Y., Nojima H., Takeda-Shitaka M., Kamiya K., Umeyama H., *Protein Sci.*, **12**, 2125—2131 (2003).
16) Yoneda T., Komooka H., Umeyama H., *J. Protein. Chem.*, **16**, 597—605 (1997).
17) Takeda-Shitaka M., Umeyama H., *FEBS Lett.*, **425**, 448—452 (1998).
18) Eckart C., *Phys. Rev.*, **47**, 552—558 (1935).
19) Frishman D., Argos P., *Proteins Struct. Func. Genet.*, **23**, 566—579 (1995).