

The 3D structure analysis of SARS-CoV S1 protein reveals a link to influenza virus neuraminidase and implications for drug and antibody discovery

Xue Wu Zhang*, Yee Leng Yap

Hong Kong University-Pasteur Research Centre, Dexter H.C. Man Building, 8 Sassoon Road, Pokfulam, HongKong, China

Received 12 March 2004; accepted 19 April 2004

Abstract

The spike protein of SARS-associated coronavirus (SARS-CoV) is an important target for anti-SARS drug discovery. Its S1 domain is responsible for receptor binding and SARS-CoV entry into cells. In this study, we constructed a rational 3D model for S1 domain of SARS-CoV spike protein by fold recognition and molecular modeling techniques. We found that there is a structure similarity between S1 protein and influenza virus neuraminidase. Our analyses suggest that the existing anti-influenza virus inhibitors and anti-neuraminidase antibody could be used as a starting point for designing anti-SARS drugs, vaccines and antibodies. Interestingly, our prediction for antibody is consistent with a recently experimental discovery of anti-SARS antibody.

© 2004 Elsevier B.V. All rights reserved.

Keywords: SARS-CoV; S1 protein; Structure; Influenza virus; Inhibitor; Antibody

1. Introduction

There is an urgent need for effective antiviral therapy when the reemergence of the highly contagious SARS-CoV infection happens. Chou et al. [1] studied the interaction of SARS-associated coronavirus (SARS-CoV) main proteinase with two ligands. Jenwitheesuk and Samudrala [2] identified some inhibitors of the SARS-CoV proteinase. Xiong et al. [3] is screening possible 73 inhibitors of SARS-CoV 3CL proteinase. Zhang and Yap [4] explored the binding mechanism of SARS-CoV main proteinase.

It is known that SARS-CoV has the same structure proteins as three previously known groups of coronaviruses: spike glycoprotein (S), membrane protein (M), envelope protein (E) and nucleocapsid protein (N). All these proteins can be used as targets for anti-SARS drug development in principle. Among these structure proteins, S protein is a type I transmembrane glycoprotein including two functional domains S1 and S2, which are conserved among coronaviruses. S1 is responsible for the binding with its receptor angiotensin-converting enzyme 2 (ACE2) on host cells

and defines the host range of the virus [5]. The goal of this study is to construct a rational 3D model of S1, to identify noncanonical interactions in the structure of S1, possible inhibitors and antibodies, hence to provide important information for anti-SARS drug, vaccine and antibody discovery.

2. Materials and methods

The sequence of spike protein was downloaded from GenBank (NP_828851). Liu et al. [6] found that the region 75-609 of SARS-CoV S protein matches to the conserved coronavirus S1 domain PF01600 in HMM database and the region 641-1247 matches to conserved coronavirus S2 domain PF01601 in HMM database. In previous study, we have predicted the structure of SARS-CoV S2 protein [7]. Here, we used the same method 3D Jury system [8] to predict the 3D structure of SARS-CoV S1 protein based on the domain (residues 75-609) mentioned above. The proteins with sufficiently high 3D score were used as templates to construct 3D models of S1 by MODELLER program [9]. The quality of 3D model was evaluated by ProQ program [10] and finally validated with the PROCHECK

* Corresponding author. Tel.: +852-2816-8407; fax: +852-2872-5782.
E-mail address: xwzhang@hkucc.hku.hk (X.W. Zhang).

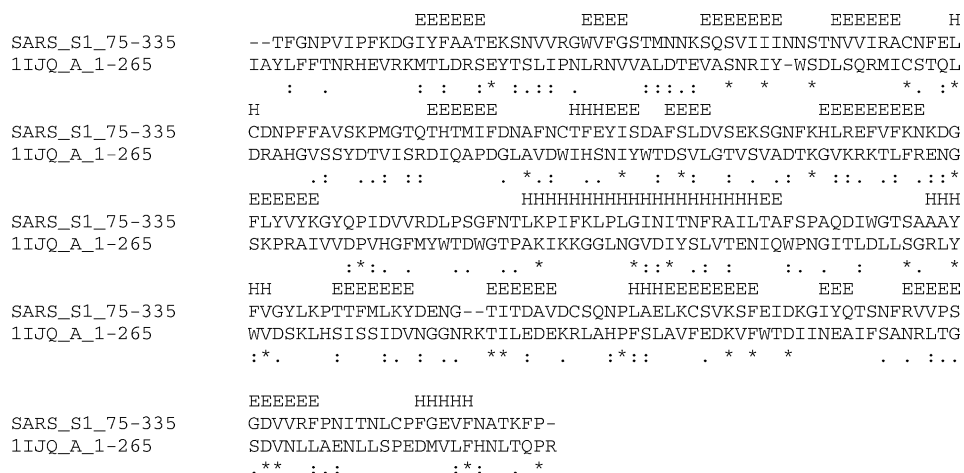


Fig. 1. The sequence alignment between SARS-CoV S1 and template 1ijq with ClustalW. The secondary structure of S1 predicted by PsiPred is also included.

program [11]. The best model was used for further analyses. Specifically, NCI program [12] was used to identify non-canonical interactions in protein structures. VAST (<http://www.ncbi.nlm.nih.gov/Structure/VAST/vastsearch.html>), DALI (<http://www.ebi.ac.uk/dali/>) and CE [13] programs were employed to search the structure neighbors of S1 protein. The structural comparison was performed by LGA [14]. The visualization of 3D structure was generated by PROTEINEXPLORER (<http://www.proteinexplorer.org>).

3. Results and discussion

Fold prediction by meta-server (3D Jury) revealed that top three significant hits (3D score > 50) for S1 protein are as below: 1loq_A (Orotidine monophosphate decarboxylase (lyase), 3D score 154, threading server PCONS2), 1ijq (low-density lipoprotein receptor (lipid transport), 3D score 125, threading server PCONS2) and 2bbk_H (Methylamine dehydrogenase (electron transport), 3D score 118, threading server PCONS2). Using them as templates the corresponding



Fig. 2. 3D model for S1 (75-609), which is basically composed of one long α helix and six three-stranded β -sheets arranged in a propeller fashion.

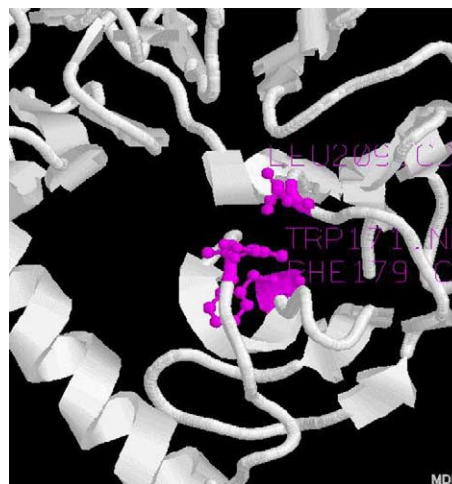


Fig. 3. Non-canonical interactions in the structure of S1 (75-609). The residue pairs involved are: Trp171 and Phe179, Leu209 and Trp171, represented by blue ball-stick (for interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

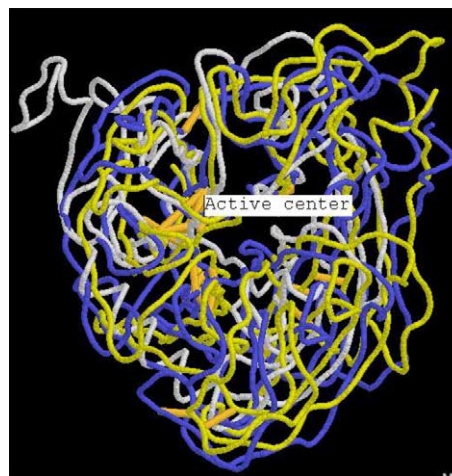


Fig. 4. The superposition of S1 (white) with its structure neighbors: 1INY (neuraminidase from influenza A virus, blue) and 1B9T (neuraminidase from influenza B virus, yellow) (for interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

		#		#####
SARS-COV_S1	1	FGNPVVPFK-----DGIYFAATEK-----	SNVVR-GWVFGSTMN	34
I1NY	46	DECRFYALSQGTTRGKHSNGTIHDRSQYRALISWPLSSPPTVYNSRVECIGWSSTSCHD		105
1B9T_A	50	ECRHFALTHYAAQGGYYN-GTRKDRNKLRLHLSVVKLGKIPTVENSI FHMAAWS-GSACH		107
		# #		#####
SARS-COV_S1	35	NK-SQSVIIINNS--TN--VVIRAC-NFELC----	DNPFFAVSKPMGTQHTMIFDNAFN	83
I1NY	106	G-KTRMSICISGPN--NNASAVIYWN--RRPV----	TEINTWARNILRTOESECVCCH--NG	155
1B9T_A	108	--DGREWYTI GVDGPDNDALVKIKYGEAYTD----	TYHSYAHNILRTOESACNCIGDC	160
		#### #		#####
SARS-COV_S1	84	CTFEYISD--AFSLDVS-----EKS-----	GNFKHLREFFVFNK	115
I1NY	156	VCPVVFDTGSAFGPAETRIYYFKEGKIILK-----	WEPLA-----GTAKHIECSCYCE	204
1B9T_A	161	--YLMITDGSAS--GISKCRFL-----KIREGRIIKEI---	LPTGRVEHTEECTCGFA	206
		#### #		
SARS-COV_S1	116	D-GFLYVYKGYQP---IDVVRDLPS-----	GFNTLKPFIKPLPLGI-----	151
I1NY	205	RAEITCTCRDNWQGSNRPVIRIDPV-----	AMTHTSQYICSPVLTDNPRPNDPTVKGKC--	256
1B9T_A	207	SNKTIECACRDNSTY---AKRPFVKLNVEDTAEIRLMCTKTYLDT	PRPDDGSIAGPCES	263
SARS-COV_S1	152	-----	NITNFRAIL	160
I1NY	257	NDPYPGNNNGVKGFSYLD-GVNTWLGRTISIALRSGYEMLKVPNALTD	DDKSKPTQGGTI	315
1B9T_A	264	NGDKLGGIKGGFVHQRMASKIGRWYSRTMSKTNRMGMELVRYD	GDGPWTDSDALTLGSGV	323
		####		
SARS-COV_S1	161	--TAFSPAQDIWGTSA-----AAYFVGYL-----	K-----PTTFMLKYDE	193
I1NY	316	--VLNTDWSGYSGSFMDYWAEGECYRACFYVELIRGRPKEDKVWWT---	SNSIVSMCSS	369
1B9T_A	324	MVSIIEEPGWYSFGFEIKDKKCDVPCIGIEMVH-----	DGGKDTWHSAAATAIYCL	372

Fig. 5. The structure alignment between S1 and its structure neighbors: I1NY (neuraminidase from influenza A virus) and 1B9T (neuraminidase from influenza B virus). The bold residues indicate conservative residues. The residues marked as '#' make contact with inhibitors.

3D models of S1 were generated and the quality of protein model was evaluated by ProQ program. The results are as follows: 1loq (ProQ-LG = 0.969, ProQ-MX = 0.055), 1ijq (ProQ-LG = 1.955, ProQ-MX = 0.101), 2bbk (ProQ-LG = 0.877, ProQ-MX = 0.057). So the 'correct' model (the cutoffs for 'correct model' are ProQ-LG > 1.5 or ProQ-MX > 0.1) for S1 protein is the model built on template 1ijq. The sequence alignment between template 1ijq and S1 with ClustalW [15] and the secondary structures of S1 predicted by PsiPred v2.3 [16] are displayed in Fig. 1. Thus, the 3D model of S1 is basically composed of one long α helix and six three-stranded β -sheets arranged in a propeller fashion (among them are a couple of small helices with 3–5 residues only) (Fig. 2), comparable to the structure of its template (low-density lipoprotein receptor): consisting of six four-stranded β -sheets arranged in a propeller fashion. It is not difficult to understand such a structure similarity in terms of their functions, the low-density lipoprotein receptor is responsible for the binding and entry into cells of lipoprotein ligands [17], while SARS-CoV S1 is also responsible for receptor binding and SARS-CoV entry into cells [5].

The non-canonical interactions in S1 protein structure were identified by NCI program and the result shows that there are three pairs of main chain-side chain interactions: Trp 171 (donor) and Phe 179 (acceptor), Ala 210 (donor) and Phe 29 (acceptor), and Leu 209 (donor) and Trp 171 (acceptor). Among these interactions, Trp171 accept one N–H $\cdots\pi$ bond from Leu209 and donates one N–H $\cdots\pi$ bond to Phe179, which forms a sandwich-like interaction, as existed in human rac1 [18] and SARS-CoV main protease [4]. These noncanonical bindings fix the helix Trp171 locate to the two β -sheets Phe179 and Leu209 locate, hence stabilizes the structure of S1 protein, but probably

reduces the active site cavity for ligand binding (Fig. 3). Indeed, the non-canonical interactions have been shown to be important for the stability of protein structure [19–21] and ligand recognition [22].

In order to extract more information from the predicted structure of S1 protein, we employed VAST, DALI and CE

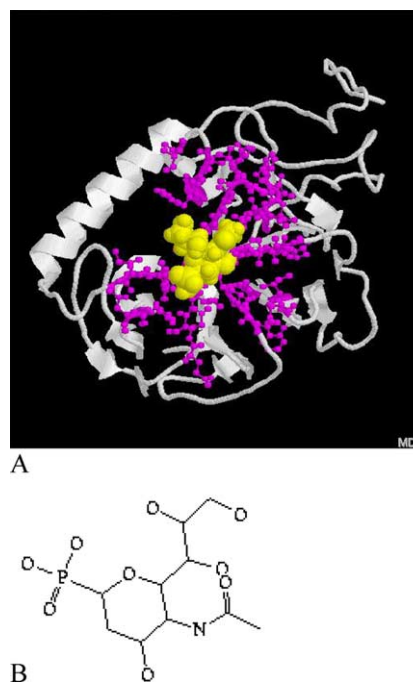


Fig. 6. (A) The interactions between S1 (white cartoon) and Epana inhibitor (yellow spacefill) from influenza A virus neuraminidase, the binding pockets are represented by blue ball-stick. (B) The chemical structure of Epana inhibitor (for interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

to SARS-CoV S1 protein. It can be seen that these compounds can bind to the active center of S1 in the same way as they did in neuraminidase, the residues that make contact with the inhibitors ($\text{RMSD} \leq 5$) are marked as ‘#’ in Fig. 5 and represented by blue ball-stick in Figs. 6 and 7, among them are 10 conservative residues.

Furthermore, we also speculate that influenza virus anti-neuraminidase antibody could be applied to SARS-CoV S1 protein after appropriate modification, such as, Nc10 and Nc41 anti-influenza virus neuraminidase antibodies [25,26]. Indeed, a very recent publication confirmed our speculation, Sui et al. [27] identified an anti-S1 human monoclonal antibody 80R, which potently neutralizes SARS-CoV infection and efficiently inhibits syncytia formation through blocking of S1 binding to its receptor angiotensin-converting enzyme 2 (ACE2) on host cells. The sequence alignments between Nc10/Nc41 and 80R show more than 50% identity (Fig. 8).

In summary, our modeling exercise on S1 protein find a structure similarity between SARS-CoV S1 protein and influenza virus neuraminidase. In fact, there are clinically similar symptoms for influenza virus and SARS-CoV, both can cause a fever, cough, pains, pneumonia and death [28]. Moreover, our study suggests that the influenza virus neuraminidase inhibitors and anti-neuraminidase antibody could be a starting point for screening anti-SARS drug, vaccine and antibody.

References

- [1] K.C. Chou, D.Q. Wei, W.Z. Zhong, Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS, *Biochem. Biophys. Res. Commun.* 308 (2003) 148–151.
- [2] E. Jenwithesuk, R. Samudrala, Identifying inhibitors of the SARS coronavirus proteinase, *Bioorg. Med. Chem. Lett.* 13 (22) (2003) 3989–3992.
- [3] B. Xiong, C.S. Gui, X.Y. Xu, C. Luo, J. Chen, H.B. Luo, L.L. Chen, G.W. Li, T. Sun, C.Y. Yu, L.D. Yue, W.H. Duan, J.K. Shen, L. Qin, T.L. Shi, Y.X. Li, K.X. Chen, X.M. Luo, X. Shen, J.H. Shen, H.L. Jiang, A 3D model of SARS-CoV 3CL proteinase and its inhibitors design by virtual screening, *Acta Pharmacol. Sinica* 24 (2003) 497–504.
- [4] X.W. Zhang, Y.L. Yap, Exploring the binding mechanism of the main proteinase in SARS-associated coronavirus and its implication to anti-SARS drug design, *Bioorg. Med. Chem.* 12 (2004) 2219–2223.
- [5] W. Li, M.J. Moore, N. Vasileva, J. Sui, S.K. Wong, M.A. Berne, M. Somasundaran, J.L. Sullivan, K. Luzuriaga, T.C. Greenough, H. Choe, M. Farzan, Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus, *Nature* 426 (6965) (2003) 450–454.
- [6] S. Liu, T. Guo, X. Ji, Z. Sun, Bioinformatical study on the proteomics and evolution of SARS-CoV, *Chin. Sci. Bull.* 48 (2003) 1277–1287.
- [7] X.W. Zhang, Y.L. Yap, Structural similarity between HIV-1 gp41 and SARS-CoV S2 proteins suggests an analogous membrane fusion mechanism, *J. Mol. Struct. THEOCHEM* 677 (2004) 73–76.
- [8] K. Ginalski, A. Elofsson, D. Fischer, L. Rychlewski, 3D-Jury: a simple approach to improve protein structure predictions, *Bioinformatics* 19 (2003) 1015–1018.
- [9] A. Sali, T.L. Blundell, Comparative protein modeling by satisfaction of spatial restraints, *J. Mol. Biol.* 234 (1993) 779–815.
- [10] B. Wallner, A. Elofsson, Can correct protein models be identified?, *Protein Sci.* 12 (2003) 1073–1086.
- [11] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* 26 (1993) 283–291.
- [12] M.M. Babu, NCI: a server to identify non-canonical interactions in protein structures, *Nucleic Acids Res.* 31 (2003) 3345–3348.
- [13] I.N. Shindyalov, P.E. Bourne, Protein structure alignment by incremental combinatorial extension (CE) of the optimal path, *Protein Eng.* 11 (1998) 739–747.
- [14] A. Zemla, LGA: a method for finding 3D similarities in protein structures, *Nucleic Acids Res.* 31 (2003) 3370–3374.
- [15] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* 22 (1994) 4673–4680.
- [16] D.T. Jones, Protein secondary structure prediction based on position-specific scoring matrices, *J. Mol. Biol.* 292 (1999) 195–202.
- [17] H. Jeon, W. Meng, J. Takagi, M.J. Eck, T.A. Springer, S.C. Blacklow, Implications for familial hypercholesterolemia from the structure of the LDL receptor YWTD-EGF domain pair, *Nat. Struct. Biol.* 8 (6) (2001) 499–504.
- [18] M. Hirshberg, R.W. Stockley, G. Dodson, M.R. Webb, The crystal structure of human rac1, a member of the rho-family complexed with a GTP analogue, *Nat. Struct. Biol.* 4 (1997) 147–152.
- [19] M.M. Babu, S. Singh, P. Balaram, A C–H···O hydrogen bond stabilized polypeptide chain reversal motif at the C terminus of helices in proteins, *J. Mol. Biol.* 322 (2002) 871–880.
- [20] G.F. Fabiola, S. Krishnaswamy, V. Nagarajan, V. Pattabhi, C–H···O hydrogen bonds in beta sheets, *Acta Crystallogr. Sec. D.* 53 (1997) 316–320.
- [21] A. Senes, I. Ubarretxena-Belandia, D.M. Engelman, The C–H···O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions, *Proc. Natl Acad. Sci. USA* 98 (2001) 9056–9061.
- [22] G. Kryger, I. Silman, J.L. Sussman, Structure of acetylcholinesterase complexed with E2020: implications for the design of new anti-alzheimer drugs, *Structure* 7 (1999) 297–307.
- [23] J.B. Finley, V.R. Atigadda, F. Duarte, J.J. Zhao, W.J. Brouillette, G.M. Air, M. Luo, Novel aromatic inhibitors of influenza virus neuraminidase make selective interactions with conserved residues and water molecules in the active site, *J. Mol. Biol.* 293 (5) (1999) 1107–1119.
- [24] C.L. White, M.N. Janakiraman, W.G. Laver, C. Philippon, A. Vasella, G.M. Air, M. Luo, A sialic acid-derived phosphonate analog inhibits different strains of influenza virus neuraminidase with different efficiencies, *J. Mol. Biol.* 245 (1995) 623–634.
- [25] R.L. Malby, A.J. McCoy, A.A. Kortt, P.J. Hudson, P.M. Colman, Three-dimensional structures of single-chain Fv-neuraminidase complexes, *J. Mol. Biol.* 279 (4) (1998) 901–910.
- [26] W.R. Tulip, J.N. Varghese, W.G. Laver, R.G. Webster, P.M. Colman, Refined crystal structure of the influenza virus N9 neuraminidase-NC41 Fab complex, *J. Mol. Biol.* 227 (1992) 122–148.
- [27] J. Sui, W. Li, A. Murakami, A. Tamin, L.J. Matthews, S.K. Wong, M.J. Moore, A.S. Tallarico, M. Olurinde, H. Choe, L.J. Anderson, W.J. Bellini, M. Farzan, W.A. Marasco, Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association, *Proc. Natl Acad. Sci. USA* 101 (8) (2004) 2536–2541.
- [28] J. Oxford, S. Balasingam, R. Lambkin, A new millennium conundrum: how to use a powerful class of influenza anti-neuraminidase drugs (NAIs) in the community, *J. Antimicrob. Chemother.* 53 (2) (2004) 133–136.