ELSEVIER

Prediction of quaternary assembly of SARS coronavirus peplomer

Andrea Bernini[a], Ottavia Spiga[a], Arianna Ciutti[a], Stefano Chiellini[a], Luisa Bracci[a],
Xiyun Yan[b], Bojian Zheng[c], Jiandong Huang[d], Ming-Liang He[e], Huai-Dong Song[f],
Pei Hao[g], Guoping Zhao[h], Neri Niccolai[a,*]

[a] Department of Molecular Biology, Biomolecular Structure Research Center, University of Siena, I-53100 Siena, Italy
[b] National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China
[c] Department of Microbiology, The University of Hong Kong, Pokfulam, Hong Kong, China
[d] Department of Biochemistry, The University of Hong Kong, Pokfulam, Hong Kong, China
[e] Institute of Molecular Biology, The University of Hong Kong, Pokfulam, Hong Kong, China
[f] State Key Laboratory for Medical Genomics, Shanghai, China
[g] Shanghai Center for Bioinformation Technology, Shanghai, China
[h] Chinese National Human Genome Center at Shanghai, Shanghai, China

Abstract

The tertiary structures of the S1 and S2 domains of the spike protein of the coronavirus which is responsible of the severe acute respiratory syndrome (SARS) have been recently predicted. Here a molecular assembly of SARS coronavirus peplomer which accounts for the available functional data is suggested. The interaction between S1 and S2 appears to be stabilised by a large hydrophobic network of aromatic side chains present in both domains. This feature results to be common to all coronaviruses, suggesting potential targeting for drugs preventing coronavirus-related infections.
© 2004 Elsevier Inc. All rights reserved.

Keywords: SARS coronavirus; Spike glycoprotein; Quaternary structure prediction

We are now in a post-epidemic period of the severe acute respiratory syndrome (SARS), caused by the coronavirus henceforth called SARS-CoV. Nevertheless, since the mode of transmission, spread, and mechanisms of virulence of SARS-CoV are not fully understood, all the possible weapons that Immunology and Pharmacology can provide should be prepared against the virus to defend ourselves better when this virus will rear again its infecting crown [1].

For a pharmacological approach the structural characterisation of the molecular repertoire of the target organism is of fundamental importance . In this respect, not much is available yet for SARS-CoV, as only two crystallographic determinations [2,3] and few predictive models [4–6] are so far available.

Among the above-mentioned structures, the predicted structures of S1 and S2 domains of the viral spike glycoprotein [5] can represent a rational basis to design specific antiviral drugs and diagnostic kits. This protein, indeed, has been found to be the viral membrane protein responsible of SARS-CoV cell entry by interacting with the receptor of the target cell and causing subsequent virus-cell fusion [7].

SARS-CoV ultra-high resolution images have been obtained [8] by scanning electron microscopy, SEM, which indicate that the spike glycoprotein is organised as a trimer. This finding offers a fundamental hint to investigate the overall assembly of the outer viral particles, peplomers, which give that characteristic

---

crown-like aspect to the virion, therefore classified in the *coronaviridae* family [9].

A stable quaternary structure without covalent cross-linking has been proposed, in general, for coronavirus peplomers [7]. This feature is consistent with our previous structural predictions, as no Cys residue without a corresponding cystine-bridged counterpart is present in both models of SARS-CoV spike glycoprotein domains.

The distribution of N-glycosylation and mutation sites has also been considered for a fine-tuning of the peplomer structural features with functional data.

Materials and methods

Peplomer model building has been performed on the basis of the structures of the S1 and S2 domains of the SARS-CoV spike protein which are available in the Protein Data Bank [10]: the structure models 1Q4Z and 1U4K have been used for S1 and S2, respectively. Docking of the two domains has been manually performed and the reliability of each of the possible peplomer assemblies has been discussed according to the ProsaII software package [11]. Accordingly, quaternary structures exhibiting the lowest energies for atom pair and solvent interaction were considered for further optimisation by using molecular dynamics simulations with Gromacs [12]. After a PROCHECK analysis of the final refined peplomer structure it has been deposited in the Protein Data Bank with the ID code 1T7G. All displays of structures, as well as exposed surface area (ESA) calculations, were carried out with the program MOLMOL [13].

Results and discussion

SARS coronavirus peplomer shape and dimensions are now well defined by recent SEM determinations [8], and the club-shaped protrusions of a trimer glycoprotein appear to extend itself approximately 200 Å from the virion envelope membrane with a maximum width of 100–200 Å.

It has been shown that coronaviruses present the S1 domain as the globular head of the spike with receptor-binding activity and that the S2 domain is present in the stalk portion of the spike [14]. In this respect, the fact that SEM images clearly suggest that in the viral peplomers the spike glycoprotein is present as a trimer [8] results to be a fundamental starting point for our model building procedure. This is also in accord with the general rule that coronavirus spike proteins form three-stranded left-handed coiled-coils. Moreover, the fact that the 320–518 fragment of S1 domain has been identified as the SARS-CoV peplomer binding site to the ACE2 cellular receptor [15] implies that the residues which are the most involved in the interaction with the receptor have to be positioned in the S1 external top side.

These first morphological and functional hints have been coupled to the results of a systematic search for surface hot spots of S1 and S2 SARS-CoV domains, i.e., potential drug binding and/or protein–protein inter-

action sites, to gain structural information on the relative orientations of these S1 and S2 domains. This analysis has been performed on the basis of S1 and S2 molecular models available [5]. Furthermore, a Clustal W [16] analysis of all the coronavirus spike proteins present in the SwissProt protein sequence data bank has been carried out and 236 sequences have been found to be compared with the one of SARS-CoV, SwissProt Accession No. P59594, originally used for our model building of the S1 and S2 domains of the S glycoprotein.

To build the molecular model of the SARS-CoV peplomer, the modelled structures of its S1 and S2 domains have been used together with homology criteria with the quaternary assembly of other viral systems [14,17,18].

In the first step of the model building procedure, the positioning of each of the three S2 in respect to the others was carried out by assembling the long α-helix spanning residues 904–968, constituting the first heptad repeat (HR1, according to the prediction by Multicoil [19]), in a three-stranded, left-handed coiled-coil. The three HR1s were first aligned parallel along the major axis, and then rotated about the center of mass of the same amount to get the amino acids of positions *a* and *d* justaxposed. The structure was refined by minimizaton followed by a simulated annealing dynamics.

In the second step, possible interfaces between the S1 and S2 domains, and among the S1 + S2 components, needed for the assembly of a trimeric structure, have been systematically searched.

Thus, in spite of the limited sequence homology, ranging from 20.39% to 27.63%, found for all these spike glycoproteins, in the S1 hydrophobic pocket delimited by F187, F334, F253, and W423 a high level of residue conservation is present. In this respect SARS-CoV, when compared with all the other coronaviruses, is unique in its W/F swapping between position 253 and 423, see Table 1. From Table 1 it can also be

Table 1
Amino acid occupancy of a possible strong binding site between S1 and S2 domains

Coronavirus	Number of sequenced genomes	187	253	334	423	782	784
Human sars	36	F	F	F	W	L	F
Murine	16	F	W	F,V	Y	L,V	F
Porcine(1)	1	F	W	V	Y	L	F
Porcine(2)	16	F	W	V	Y	L	F
Porcine(3)	30	F	W	V	Y	L	F
Porcine(4)	7	F	W	V	Y	L	F
Bovine	26	Y	W	V	Y	L	F
Human	24	F	W	V	Y	L,I	F
Equine	1	Y	W	V	Y	L	F
Avian	112	F	W	V	Y	L,I,V	F
Feline	54	F	W	V	Y	L	F
Canine	8	Y	W	V	Y	L	F

(1), Hemagglutinating encephalomyelitis virus; (2), transmissible gastroenteritis virus; (3), respiratory virus; and (4), epidemic diarrhea virus.

noticed that in the S2 domain the hydrophobic residues L803 and F805, totally conserved among the SARS-CoV available genomes and fully exposed in the S2 molecular model, are located in a sequence position where only hydrophobic residues are found.

The large difference in the pathologies induced by coronavirus infections suggests that a role for these two hydrophobic moieties of S1 and S2 domains might be attributed to the peplomer assembly rather than to the interaction with the host cell. Residues F187, F334, F253, and W423 could, indeed, form the S1 hydrophobic pocket where S2 puts its hydrophobic finger formed by L803 and F805 residues.

The remarkable agreement between the steric requirements for the S1–S2 interaction and the surface position of the proposed S1 binding site to the ACE2 receptor points towards a finite orientation of these peplomeric domains, see Fig. 1. From this starting point, to recompose the full SARS-CoV peplomer from its components, we used the following simple criteria: (i) orienting each S1 and the S2 stalk domains so that they could dock through the interaction described above, (ii) keeping all the potential N-glycosylation sites as surface exposed as possible, and (iii) positioning the largest hydrophobic surface patches in subunit interfaces. The fact that in the S2 trimer the side chains of L803 and F805 residues are still surface accessible after the coiled coil formation supports the hypothesis that the peplomer reaches its structural stability through the hydrophobic interactions of the S1 pocket with the S2 finger, as depicted in Fig. 2. Then, geometrical and energetic considerations converge towards possible solutions for the structure of the SARS-CoV peplomer. In Fig. 3, three molecules of S1–S2 adducts are positioned after their assembly, in a way which is consistent with the overall size of the pep-

Fig. 2. A detailed representation of the S1 (dark blue) hydrophobic pocket interacting with the S2 (pale blue) finger. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

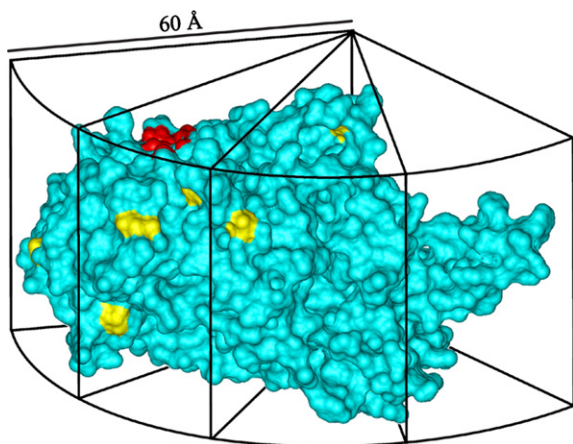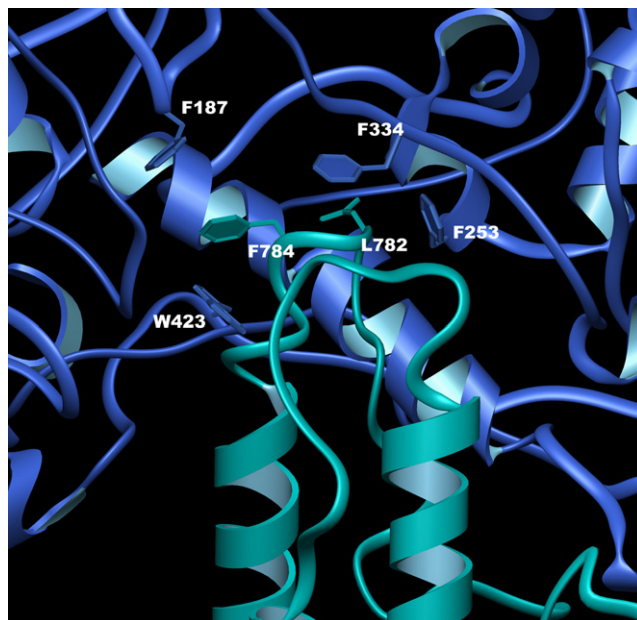Fig. 3. Three spike glycoproteins form the peplomer of SARS-CoV coronavirus.

Fig. 1. The S1 domain oriented to fit the morphological SEM images of [8]. In yellow and in red the potential N-glycosylation sites and the residues involved in the interaction with ACE2 [15] are, respectively, colored. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

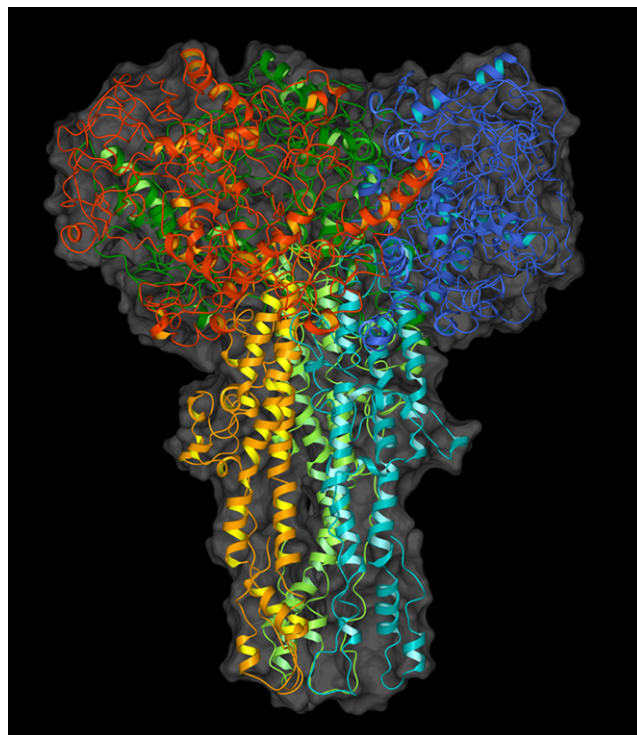lomer [8]. It should be noted also that, among the mutations which have been found in the SARS-CoV spike protein region of all the available genomes, see Table 2, nothing occurs in the S1–S2 interfaces here identified.

Table 2
Mutation site topology occurring in all the available strains of SARS-CoV spike glycoprotein

aa	Mutation topology	TOR2	BJ01	HZS2_C	CUHK_LC2	SOD	HZS2_FC	GZ_A	HGZ8L1_A
49	Exposed top	S	S	S	S	S	S	L	S
74	Exposed top	H	H	H	H	H	H	H	F
75	Exposed top	T	T	T	T	T	T	R	T
77	Exposed top	G	D	D	G	G	G	D	D
239	Partially exposed side	S	S	S	S	S	S	S	L
244	Buried	I	T	T	I	I	I	T	T
311	Exposed side	G	G	G	G	G	G	G	R
344	Partially exposed side	K	K	K	K	K	K	K	R
577	Buried, interface S1-S1	A	S	S	S	S	S	S	S
778	Buried	Y	Y	Y	Y	Y	Y	D	D
1148	Exposed down	L	L	L	L	F	F	L	L
1179	Buried	L	L	L	L	L	L	L	F
1208	Non-modelled	A	A	A	V	A	A	A	A

For the S2 moiety of the S glycoprotein, composed by one well-structured moiety containing the HR1 and another subdomain spanning the 1027–1195 segment of S glycoprotein and containing the HR2 (residues 1148–1193), ambiguity remains on the structure and on the location of the second in the peplomer structure. Such subdomain is critical for the interaction with the viral envelope, due to its proximity to the *trans*-membrane region and for the overall structure stability of the peplomer. In fact, a peptide reproducing the C terminal heptad repeat fragment 1161–1187 of S2 exhibits antiviral activity [20].

Thus, 94% of SARS-CoV peplomer structure has been modelled and deposited with the pdb ID code 1T7G. Accordingly, the most interesting regions to be reproduced in synthetic peptides for mimotope design have been found, as well as the hydrophobic sites distributed at the S1/S1, S2/S2, and S1/S2 interfaces for targeting of potential antiviral drugs (patent RM2004A000162). The fact that we could not model the 665–736 sequence of the S glycoprotein does not interfere with the exposed surface on top of the SARS-CoV peplomer, as the missing modelled moiety can be identified in a peplomer lateral region, where a deep groove is found. This peplomer region could be filled by the non-modelled part of the sequence, which consistently exhibits an extensive hydrophobic character [21]. In the present peplomer model, the so-called HR1 and HR2 moieties, i.e., the N and C terminal heptad repeat regions, respectively, spanning the sequences 904–975 and 1148–1193, are not bound together. This feature is consistent with the extensive conformational change, necessary for this and several other viruses for the fusogenic mechanism [22–24].

Then, on the basis of the obtained peplomer structure, correlations can be explored between viral genome mutations and possible interactions with the host cell receptor(s). As reported in Table 2, on the basis of the proposed quaternary assembly a systematic topological analysis of mutation sites occurring in the 36 genomes

so far available for the SARS-CoV spike glycoprotein has been done. It can be observed that (i) most of the mutations are found in exposed sites; (ii) the only mutation involving a relevant position for the receptor interaction, i.e., 344 K/R, is a conservative one; and (iii) non-conservative substitutions are found in the buried positions of residue 778 with Y/D, which do not induce sterical conflicts.

The structural characterisation of SARS-CoV spike glycoprotein domains, here described, suggests also a general scheme for the peplomer assembly of all coronaviruses. In fact, from the sequence alignment of the spike glycoproteins of all known coronaviruses, as shown in Table 1, it appears that the above-described hydrophobic interaction between the S1 pocket and the S2 finger is very conserved. It could, therefore, represent a very critical region for the interaction between S1 and S2 domains for all coronaviruses, opening new perspectives for the design of small molecules that can efficiently interfere with the viral replication. Hence, SARS as well as all the other members of the *coronaviridae* family could put down their infecting crown with the same type of antiviral drug, which could protect from possible transmission of coronavirus infections from wild animals.

References

[1] E.G. Brown, J.A. Tetro, Comparative analysis of the SARS coronavirus genome: a good start to a long journey, Lancet 361 (2003) 1756–1757.
[2] H. Yang, M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G.F. Gao, K. Anand, M. Bartlam, R.

Hilgenfeld, Z. Rao, The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor, Proc. Natl. Acad. Sci. USA 100 (2003) 13190–13195.

[3] M.P. Egloff, F. Ferron, V. Campanacci, S. Longhi, C. Rancurel, H. Dutartre, E.J. Snijder, A.E. Gorbalenya, C. Cambillau, B. Canard, The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world, Proc. Natl. Acad. Sci. USA 101 (2004) 3792–3796.

[4] X. Shen, J.H. Xue, C.Y. Yu, H.B. Luo, L. Qin, X.J. Yu, J. Chen, L.L. Chen, B. Xiong, L.D. Yue, J.H. Cai, J.H. Shen, X.M. Luo, K.X. Chen, T.L. Shi, Y.X. Li, G.X. Hu, H.L. Jiang, Small envelope protein E of SARS: cloning, expression, purification, CD determination, and bioinformatics analysis, Acta Pharmacol. Sin. 24 (2003) 505–511.

[5] O. Spiga, A. Bernini, A. Ciutti, S. Chiellini, N. Menciassi, F. Finetti, V. Causarono, F. Anselmi, F. Prischi, N. Niccolai, Molecular modelling of S1 and S2 subunits of SARS coronavirus spike glycoprotein, Biochem. Biophys. Res. Commun. 310 (2003) 78–83.

[6] X. Xu, Y. Liu, S. Weiss, E. Arnold, S.G. Sarafianos, J. Ding, Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design, Nucleic Acids Res. 31 (2003) 7117–7130.

[7] D. Cavanagh, The Coronavirus Surface Glycoprotein, second ed., Plenum Press, New York, 1995.

[8] Y. Lin, X. Yan, W. Cao, C. Wang, J. Feng, J. Duan, S. Xie, Probing the structure of the SARS coronavirus using scanning electron microscopy, Antivir. Ther. 9 (2004) 287–289.

[9] R.J. de Groot, W. Luytjes, M.C. Horzinek, B.A. van der Zeijst, W.J. Spaan, J.A. Lenstra, Evidence for a coiled-coil structure in the spike proteins of coronaviruses, J. Mol. Biol. 196 (1987) 963–966.

[10] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.

[11] M.J. Sippl, Recognition of errors in three-dimensional structures of proteins, Proteins 17 (1993) 355–362.

[12] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, Comp. Phys. Commun. 91 (1995) 43–56.

[13] R. Koradi, M. Billeter, K. Wuthrich, MOLMOL: a program for display and analysis of macromolecular structures, J. Mol. Graph. 14 (1996) 51–55, 29–32.

[14] D.N. Lewicki, T.M. Gallagher, Quaternary structure of corona-virus spikes in complex with carcinoembryonic antigen-related cell adhesion molecule cellular receptors, J. Biol. Chem. 277 (2002) 19727–19734.

[15] S.K. Wong, W. Li, M.J. Moore, H. Choe, M. Farzan, A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2, J. Biol. Chem. 279 (2004) 3197–3201.

[16] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.

[17] N.K. Sauter, J.E. Hanson, G.D. Glick, J.H. Brown, R.L. Crowther, S.J. Park, J.J. Skehel, D.C. Wiley, Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography, Biochemistry 31 (1992) 9609–9621.

[18] B. Delmas, H. Laude, Assembly of coronavirus spike protein into trimers and its role in epitope expression, J. Virol. 64 (1990) 5367–5375.

[19] E. Wolf, P.S. Kim, B. Berger, MultiCoil: a program for predicting two- and three-stranded coiled coils, Protein Sci. 6 (1997) 1179–1189.

[20] B. Zheng, Y. Guan, L. Du, Y. Zheng, Y. Chen, K.L. Wong, H. Sun, J. Huang, N. Niccolai, A. Bernini, O. Spiga, K.Y. Yuen, H.F. Kung, and M.L. He, Protection of SARS-associated coronavirus infection by peptides targeting the viral spike protein (in preparation).

[21] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132.

[22] J. Zhu, G. Xiao, Y. Xu, F. Yuan, C. Zheng, Y. Liu, H. Yan, D.K. Cole, J.I. Bell, Z. Rao, P. Tien, G.F. Gao, Following the rule: formation of the 6-helix bundle of the fusion core from severe acute respiratory syndrome coronavirus spike protein and identification of potent peptide inhibitors, Biochem. Biophys. Res. Commun. 319 (2004) 283–288.

[23] S. Liu, G. Xiao, Y. Chen, Y. He, J. Niu, C.R. Escalante, H. Xiong, J. Farmar, A.K. Debnath, P. Tien, S. Jiang, Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors, Lancet 363 (2004) 938–947.

[24] Y. Xu, Y. Liu, Z. Lou, L. Qin, X. Li, Z. Bai, H. Pang, P. Tien, G.F. Gao, Z. Rao, Structural basis for coronavirus-mediated membrane fusion: crystal structure of MHV spike protein fusion core, J. Biol. Chem. (2004).