ELSEVIER

Journal of Molecular Structure (Theochem) 677 (2004) 73–76

THEO CHEM

www.elsevier.com/locate/theochem

Structural similarity between HIV-1 gp41 and SARS-CoV S2 proteins suggests an analogous membrane fusion mechanism

Xue Wu Zhang*, Yee Leng Yap

HKU-Pasteur Research Center, Bioinformatics, 8 Sassoon Road, Pokfulam, Hong Kong, China

Received 25 November 2003; accepted 2 February 2004

Abstract

SARS-associated coronavirus (SARS-CoV) has been identified as the causal agent of a new emerging disease: severe acute respiratory syndrome (SARS). Its spike protein S2 is responsible for mediating fusion of viral and cellular membrane. In this study, we modeled the 3D structure of S2 subunit and compared this model with the core structure of gp41 from HIV-1. We found that SARS-CoV S2 and gp41 share the same two α helices, suggesting that the two viruses could follow an analogous membrane fusion mechanism. Further ligand-binding analysis showed that two inhibitors GGL and D-peptide from HIV-1 gp41 may serve as inhibitors for SARS-CoV entry.

Keywords: SARS-CoV; Human immunodeficiency virus type 1; Structural similarity; Inhibitor

1. Introduction

Coronavirus spike protein plays a very important role in virus entry, virus–receptor interaction, variations in host range and tissue tropism. The S proteins of majority of coronaviruses are cleaved into two functional subunits, S1 and S2. Liu et al. [1] indicated that the S protein of SARS-associated coronavirus (SARS-CoV) also forms S1 and S2 domains. The peripheral S1 portion is responsible for cellular receptor recognition, while the membrane-spanning S2 portion mediates the fusion of viral and cellular membrane, hence S protein determines the specificity of host and virulence of coronavirus [2]. Similarly, there are two non-valently associated subunits in the human immunodeficiency virus type 1 (HIV-1) envelope glycoprotein, gp120 and gp41, gp120 directs target-cell recognition and viral tropism through interaction with the cell-surface receptor CD4, while the membrane-spanning gp41 promotes fusion of the viral and cellular membranes so that viral contents are released into the host cell [3].

Sequence analysis revealed that there are some similar motifs in HIV-1 gp41 and SARS-CoV S2 proteins. Gallaher and Garry [4] identified an N-terminal leucine/isoleucine zipper-like sequence and an aromatic-rich region. Kliger and Levanon [5] reported a C-terminal heptad repeat in the upstream of an aromatic-rich region. These discoveries probably lead to development of new therapeutic strategy against SARS-CoV. The goal of this study is to probe whether any similarity exists in the 3D structure of HIV-1 gp41 and SARS-CoV S2 proteins and identification of possible inhibitor-binding sites.

2. Materials and methods

The core structure of HIV-1 gp41 was downloaded from Protein Data Bank (1AIK). The sequence of spike protein was downloaded from GenBank (NP_828851). Liu et al. [1] found that the region 641–1247 of SARS-CoV S protein matches to conserved coronavirus S2 domain PF01601 in HMM database, which is subsequently used in the fold prediction of S2 subunit by 3D Jury meta predictor [6]. The proteins with significant high 3D score were used as templates to construct 3D models of S2 by MODELLER program [7]. The quality of 3D model was evaluated by PROQ program [8] and 'correct' models were chosen for structure comparison with gp41 by LGA program [9]. The visualization of 3D structure was generated by PROTEIN-EXPLORER (http://www.proteinexplorer.org).

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

0166-1280/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.theochem.2004.02.018

Table 1
Alignment between SARS-CoV S2 protein and template 1EZF_A and secondary structure of S2

```
                    EE HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
S2_719-1019         DCNMYICGDSTECANLLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQMYKTPTLKY
1EZF_A_12-312       TCYKYLNQTSRSFAAVIQALDG--EMRNAVCIFYLVLRALDTLEDDMTISVEKKVPLLHN
                     *  *:   * . * ::   ..     * .     . :  :* *    :.   *.* *:
                         EE            HHHHHHH   EEE     HHHHHHH     HHHHHHHHHH
S2_719-1019         FGGFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMKQYGECLGDINARDLICAQKFN
1EZF_A_12-312       FHSFLY-QPDWRFMESKEKDRQVLEDFPTISLEFRNLAEKYQTVIADICRRMGIGMAEFL
                    * .* : *      ::..::.      * .::*   .: ::*   :.**  *  *   :*
                     EEEE     HHHHHHHHHHHHHHHH           EE HHHHHHHHHHHHHHHHHHHH
S2_719-1019         GLTVLPPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVL
1EZF_A_12-312       DKHVTSEQE-WDKYCHYVAGLVGIGLSR---LFSASEFEDPLVGEDTERANSMGLFLQKT
                    .  * .     *  . *.*.**.   :    : :.: :: *:. : : * *.:*:  :
                    HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH     HHHHHHHHHHHHHHHHHHHH
S2_719-1019         YENQKQIANQFNKAISQIQESLTTTSTALGKLQ-----DVVNQNAQALNTLVKQLSSNFG
1EZF_A_12-312       NIIRDYLEDQQGGREFWPQEVWSRYVKKLGDFAKPENIDLAVQCLNELITNALHHIPDVI
                    :. : :* .       **  :   . **.:      *:. *  : * * . :  .:.
                    HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
S2_719-1019         AISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSE
1EZF_A_12-312       TYLSRLRNQSVFNFCAIPQVMAIATLAACYNNQQVFKGAVKIRKGQAVTLMMDATNMPAV
                    :  * *.:       . .:*     ::.  :. *.:     ** .:  :    *:.  :
                    HHH
S2_719-1019         CVLGQS--
1EZF_A_12-312       KAIIYQYM
                    .:  .
```

3. Results and discussion

Meta-predictor (3D Jury) revealed that there are three significant hits (3D score >50) for S2 subunit of SARS-CoV: 1dg3_A (Human guanylate binding Protein (signaling protein), 3D score 117, threading server ShotGun on 3), 1f5n_A (Human guanylate binding Protein (signaling protein), 3D score 95, threading server mGenTHREADER) and 1ezf_A (Human squalene synthase (transferase), 3D score 82, threading server INBGU). Three templates were used to generate 3D model of S2. The quality evaluation of these models was conducted by PROQ program in which two measurements (ProQ-LG and ProQ-MX) were used. The corresponding scores are below: ProQ-LG = 0.583 and ProQ-MX = 0.026 for 1dg3_A, ProQ-LG = 1.409 and ProQ-MX = 0.034 for 1f5n_A, ProQ-LG = 2.503 and ProQ-MX = 0.106 for 1ezf_A. In view of the fact that cutoffs for 'correct model' are ProQ-LG > 1.5 or ProQ-MX > 0.1, so the 'correct' model for S2 is the model built on template 1ezf_A, although other templates obtained marginally higher 3D scores. The alignment between S2 and template 1ezf_A with ClustalW [10] and secondary structure of S2 predicted by PsiPred v2.3 [11] are displayed in Table 1. Fig. 1 shows that the 3D model of S2 subunit is

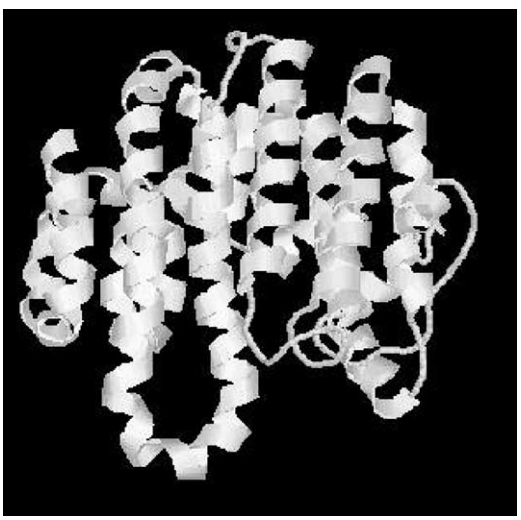

Fig. 1. 3D model of SARS-CoV S2.

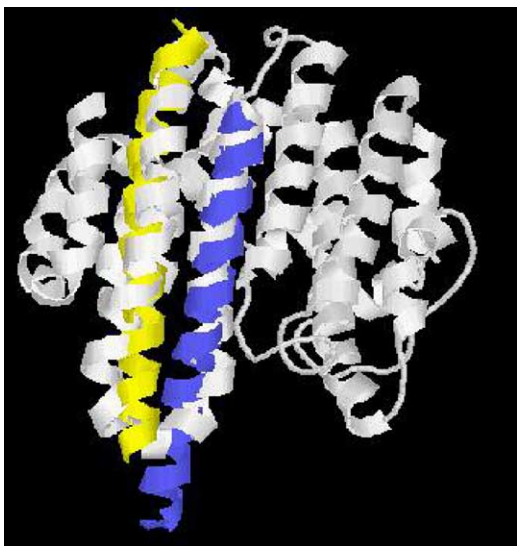

Fig. 2. Comparison between gp41 from HIV-1 and SARS-CoV S2. N and C chains of gp41 are colored blue and yellow, respectively. S2 is colored white.

Table 2
Structural similarity between SARS-CoV S2 and HIV-1 gp41 proteins

SARS-CoV S2	No. of res.	HIV-1 gp41	No. of res.	RMSD	SARS-CoV S2	No. of res.	HIV-1 gp41	No. of res.	RMSD
P	879	S	546_N	1.973	Q	908	W	628_C	2.699
F	880	G	547_N	2.433	F	909	E	630_C	3.28
A	881	I	548_N	1.687	N	910	W	631_C	4.096
M	882	V	549_N	1.176	K	911	D	632_C	3.314
Q	883	Q	550_N	2.052	A	912	R	633_C	4.069
M	884	Q	551_N	2.094	I	913	E	634_C	4.296
A	885	Q	552_N	1.078	S	914	I	635_C	3.357
Y	886	N	553_N	1.738	Q	915	N	636_C	3.839
R	887	N	554_N	2.282	I	916	–	–	–
F	888	L	555_N	1.796	Q	917	N	637_C	2.111
N	889	L	556_N	1.398	E	918	Y	638_C	3.013
G	890	R	557_N	2.154	S	919	T	639_C	2.177
I	891	A	558_N	1.984	L	920	S	640_C	1.469
G	892	I	559_N	1.911	T	921	L	641_C	2.303
V	893	E	560_N	1.962	T	922	I	642_C	2.404
T	894	A	561_N	1.998	T	923	H	643_C	1.588
Q	895	Q	562_N	2.015	S	924	S	644_C	2.077
N	896	Q	563_N	2.436	T	925	L	645_C	3.591
V	897	H	564_N	3.129	A	926	I	646_C	3.826
L	898	L	565_N	2.223	L	927	E	647_C	2.728
Y	899	L	566_N	1.514	G	928	E	648_C	2.996
E	900	Q	567_N	2.452	K	929	S	649_C	3.953
N	901	L	568_N	4.328	L	930	Q	650_C	3.291
Q	902	T	569_N	3.414	Q	931	N	651_C	2.464
K	903	V	570_N	2.291	D	932	Q	652_C	2.637
Q	904	W	571_N	2.691	V	933	Q	653_C	3.039
I	905	–	–	–	V	934	E	654_C	1.927
A	906	–	–	–	N	935	K	655_C	0.506
N	907	G	572_N	3.757	Q	936	N	656_C	0.841
					N	937	E	657_C	1.845
					A	938	Q	658_C	1.872
					Q	939	E	659_C	2.903
					A	940	L	660_C	3.788
					L	941	–	–	–
					N	942	L	661_C	4.77

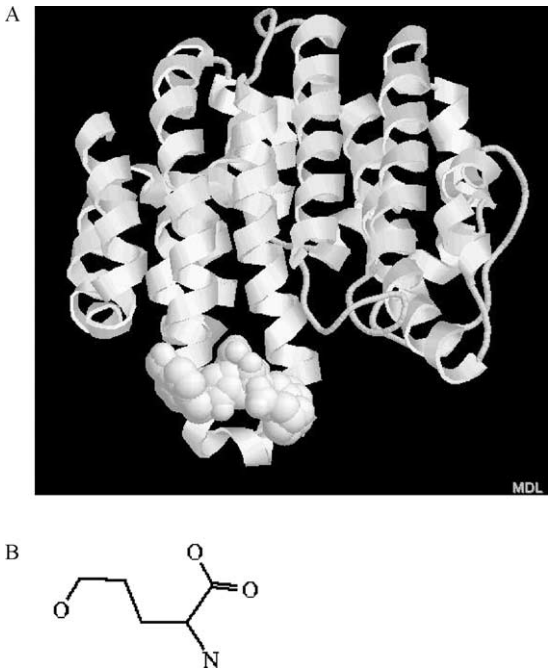

Fig. 3. (A) Binding interaction between S2 and inhibitor GGL (represented by spacefill). (B) Chemical structure of GGL, its formula is: C5H9NO4.

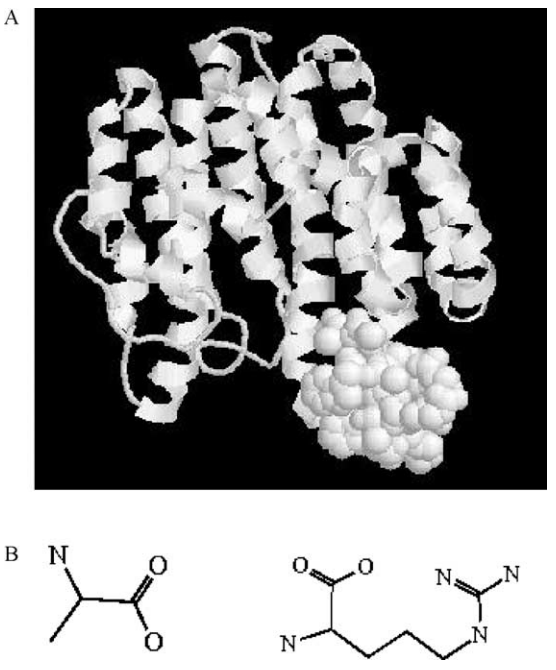

Fig. 4. (A) Binding interaction between S2 and D-peptide inhibitor of HIV-1 gp41 (represented by spacefill. (B) Chemical structure for two molecules of D-peptide DAL (5(C3H7NO2)) and DAR (2(C6H15N4O2)).

primarily consisted of α helices. In fact, Liu et al. [1] predicted the secondary structure of S protein by eight techniques and found that there are nine successive alpha-helices in S2. Based on the presence of hydrophilic and hydrophobic amino acids alternately, they speculated that these helices could be the amphipathic alpha-helices, which collapse into coiled-coils, bring a fusion peptide back toward the transmembrane domain and lead to the fusion of cellular and viral membranes [2].

The LGA method is designed to compare protein structures or fragments of protein structures in sequence dependent and sequence independent modes. We superimposed the protein structure of S2 into the core structure of gp41 (1aik) by LGA. Surprisingly we found that the N chain and C chain of gp41 are well overlapped with two helices in the 3D model of S2 subunit (Fig. 2). For SARS-CoV S2 subunit, the corresponding residues include P879-N942, for HIV-1 gp41, the corresponding residues cover S546-G572 (N chain) and W628-L661 (C chain), each of residue pairs has a distance <5 Å (Table 2).

It has been shown that that there are some similar structural motifs in HIV-1 gp41 and SARS-CoV S2 protein [4,5]: (1) N-terminal leucine/isoleucine heptad repeat sequence on residues 913–1000; (2) C-terminal leucine/isoleucine heptad repeat motif on residues 1151–1185. While our results reveal that SARS-CoV S2 and HIV gp41 share very similar helix structure on residues 879–942, these discoveries suggest a similar membrane fusion mechanism for the two viruses.

Naturally, a question arises: could the inhibitors for anti-HIV-1 therapy be used to fight against SARS-CoV? For example, GGL, a HIV-1 specific cell entry inhibitor that can bind to the coiled-coil core of gp41 and efficiently inhibit HIV-1 envelope-mediated cell-cell fusion [12], and another D-peptide inhibitor, which targets the gp41 coiled-coil pocket and inhibits HIV-1 entry [13]. Fig. 3 showed the binding interaction between GGL and S2 protein, the residues involved are: 901–918. Fig. 4 showed the binding interaction between D-peptide and S2 protein, the residues involved are: 899–915. This suggests GGL and D-peptide inhibitors from HIV-1 gp41 could be used as potential inhibitors for SARS-CoV entry.

References

[1] S. Liu, T. Guo, X. Ji, Z. Sun, Bioinformatical study on the proteomics and evolution of SARS-CoV, Chinese Science Bulletin 48 (2003) 1277–1287.
[2] T.M. Gallagher, M.J. Buchmeier, Coronavirus spike proteins in viral entry and pathogenesis, Virology 279 (2001) 371–374.
[3] D.C. Chan, D. Fass, J.M. Berger, P.S. Kim, Core structure of gp41 from the HIV envelope glycoprotein, Cell 89 (1997) 263–273.
[4] W.R. Gallaher, R.F. Garry, Model of the pre-insertion region of the spike (S2) fusion glycoprotein of the human SARS coronavirus: implications for antiviral therapeutics, Virology (2003) http://www.virology.net/sars/s2model.html.
[5] Y. Kliger, E.Y. Levanon, Cloaked similarity between HIV-1 and SARS-CoV suggests an anti-SARS strategy, BMC Microbiology 3 (2003) 20–26.
[6] K. Ginalski, A. Elofsson, D. Fischer, L. Rychlewski, 3D-Jury: a simple approach to improve protein structure predictions, Bioinformatics 19 (2003) 1015–1018.
[7] A. Sali, T.L. Blundell, Comparative protein modeling by satisfaction of spatial restrains, Journal of Molecular Biology 234 (1993) 779–815.
[8] B. Wallner, A. Elofsson, Can correct protein models be identified?, Protein Science 12 (2003) 1073–1086.
[9] A. Zemla, LGA: a method for finding 3D similarities in protein structures, Nucleic Acids Research 31 (2003) 3370–3374.
[10] 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 Research 22 (1994) 4673–4680.
[11] D.T. Jones, Protein secondary structure prediction based on position-specific scoring matrices, Journal of Molecular Biology 292 (1999) 195–202.
[12] G. Zhou, M. Ferrer, R. Chopra, T.M. Kapoor, T. Strassmaier, W. Weissenhorn, J.J. Skehel, D. Oprian, S.L. Schreiber, S.C. Harrison, D.C. Wiley, The structure of an HIV-1 specific cell entry inhibitor in complex with the HIV-1 gp41 trimeric core, Bioorganic and Medicinal Chemistry 8 (9) (2000) 2219–2227.
[13] D.M. Eckert, V.N. Malashkevich, L. Hong, P.A. Carr, P.S. Kim, Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket, Cell 99 (1999) 103–115.