

Putative structure and function of ORF3 in SARS coronavirus

Xue Wu Zhang*, Yee Leng Yap

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

Received 13 June 2004; accepted 7 October 2004

Available online 19 December 2004

Abstract

Based on molecular modeling techniques we constructed a rational 3D model of ORF3 in SARS coronavirus (SARS-CoV). Our studies suggest that the function of ORF3 could be involved in FAD/NAD binding according to its predicted structure and comparison with other structure neighbors. Furthermore, we identified three pairs of non-canonical N–H $\cdots\pi$ interactions in the structure of ORF3, which can make contributions to the stability of protein structure. These results provide important clues for better understanding of SARS-CoV ORF3 and trying new therapeutic strategies.

© 2004 Elsevier B.V. All rights reserved.

Keywords: SARS-CoV; ORF3; 3D structure; Function; Non-canonical interaction

1. Introduction

Sequence analysis of SARS coronavirus genome reveals that it contains five major open reading frame (ORFs) that encode a polymerase, and S, M, E and N proteins like those of other coronavirus. However, the nine potential ORFs are not found in other coronaviruses [1]. Theoretically all these proteins can be used as targets in drug and vaccine design. However, there are some difficulties in understanding the functions of these unknown proteins due to very poor sequence homology with proteins available in the Protein Data Bank.

The 3D jury system [2] utilizes a global network of independent structure prediction servers to detect patterns of structural similarity between diverse models and select the correct fold from a set of borderline predictions. An exciting finding based on such a method [3] is that the mRNA cap-1 methyltransferase function has been assigned to the nsp13 protein of the SARS coronavirus (3D jury score > 100). In this study, we started with meta-server 3D jury system for fold recognition study of ORF3, constructed rational molecular model, hence to understand the potential function of ORF3 in terms of tertiary structure.

2. Materials and methods

The sequence of ORF3 protein in SARS-CoV was downloaded from GenBank (NP_828851) and used for fold prediction by 3D Jury system [2], it is a comprehensive protein structure prediction servers including more than 10 novel fold recognition methods, which made a dramatic impact on the critical assessment of protein structure prediction (CASP-5) in 2002. The proteins with a sufficiently high 3D score were used as templates to construct 3D models of ORF3 using the MODELLER program [4]. The quality of 3D models was evaluated by ProQ program [5] and the best model was used for further analyses. Specifically, in order to get possible information about the function of ORF3, VAST (<http://www.ncbi.nlm.nih.gov/Structure/VAST/vastsearch.html>), DALI (<http://www.ebi.ac.uk/dali/>) and CE [6] programs were employed to search the structure neighbors of ORF3 protein. The structural comparison was performed by LGA [7]. Finally, NCI program [8] was used to identify non-canonical interactions in protein structures. The visualization of 3D structure was generated by PROTEINEXPLORER (<http://www.proteinexplorer.org>).

3. Results and discussion

The 3D Jury system found three significant hits (3D score > 90) which have a similar fold to ORF3 (threading

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

Table 1
The sequence alignment between ORF3 and 1LVL and the secondary structure

ORF3	MRFFTLRSITAQPVKIDNASPASTVHATATIPLQASLPFGWLIVGV--AFLAVFQSATKI	EEEEEEEEEEEE	HHHHHHHH	HHHHHHHH	EE
1LVL	-----QQTIQTTLLI-----IGGGPGGYVAAIRAG-QLGIPT				
ORF3	IALNKRWQLALYKGFQFICNLLLLFVTIYSHLLLVA-----AGMEAQFLYL	EEE	EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE		HHHHHH
1LVL	VLV--EGQALGGTCLNIGCIPSKALIHVAEQFHQASRFTEPSPLGISVASPRLDIGQSVA				
ORF3	YALIYFLQCINACRIIMRCWLCWKCKSKNPLLYDANYFVCW-HTHNYDYCIPYNS--VTD	HHHHHHHHHHHHHHHHHHHEEE	EEEE	EEEE	E
1LVL	WKDGI VDRLLTTGVAALLKKHGKVVHGWAKVL-DGKQVEVDGQRIQCEHLLLATGSSSVE				EEE
ORF3	TIVVTEGDGISTPKLKEDYQIGGYSEDRHSGVKDYVVVHGYFTEVYYQLESTQITTDGTI	EEEE	HHH	EEEEEEEEEEEEEEEE	EEE
1LVL	LPRRPTKGFNLECLDLKMNGAAIAIDERQCTSMHNVWAI GD--VAGEPMLAHRAMAQG-				
		HHHHHHHHHH			
ORF3	ENATFFIFNKLVK-DPPNVQIHTIDGSSGVANPAMDPIYDEPTTTT SVPL				
1LVL	EMVAEIIAGKARRFE-----				

server PCONS2): 1LVL (*Pseudomonas putida* lipoamide dehydrogenase, 3D score 102), 3GRS (human glutathione reductase, 3D score 99), 1GES_A (*Escherichia coli* glutathione reductase, 3D score 95.5). Using 1LVL, 3GRS and 1GES_A as templates the corresponding 3D models for ORF3 were generated and the quality of protein model was evaluated by the ProQ program with two measurements. The results are listed below: 1LVL (ProQ-LG=2.826, ProQ-MX=0.249), 3GRS (ProQ-LG=1.498, ProQ-MX=0.146), 1GES (ProQ-LG=1.601, ProQ-MX=0.14). This means all 3D models of ORF3 based on above templates are ‘correct model’, 1LVL is almost ‘good model’ (the cutoff value of two protein quality measurements is: ProQ-LG > 1.5 or ProQ-MX > 0.1 for correct model, and ProQ-LG > 3 or ProQ-MX > 0.5 for good model). So we take the 3D model built on template 1LVL as the 3D model of ORF3. The alignment between ORF3 and 1LVL and secondary

structure are shown in Table 1. Fig. 1 shows the 3D model of ORF3 built on template 1LVL, which consists of 5 α helices and 6 β sheets. Three trans-membrane regions (34–56, 77–99, and 103–125) predicted by Marra et al. [1] correspond to three helices regions in the alignment: 43–58, 71–98, and 103–129. The additional two helices regions are also existed in the structure: 3–12 and 138–147.

Naturally, a question arises: what information about ORF3’s function we can get from its 3D structure? The above three templates (1LVL, 3GRS and 1GES), which belong a FAD/NAD-linked reductase family, lead us to the speculation that ORF3 may be a protein related to FAD/NAD-binding. This seems consistent with the speculation that ORF3 may encode a protein related to

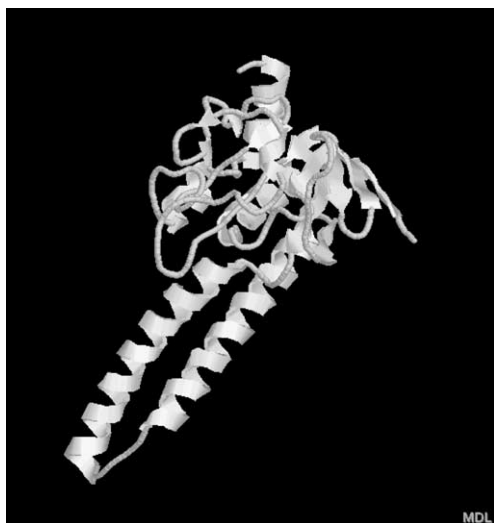


Fig. 1. The predicted 3D model of SARS-CoV ORF3 based on template 1LVL (*Pseudomonas putida* lipoamide dehydrogenase), which consists of 5 α helices and 6 β sheets.

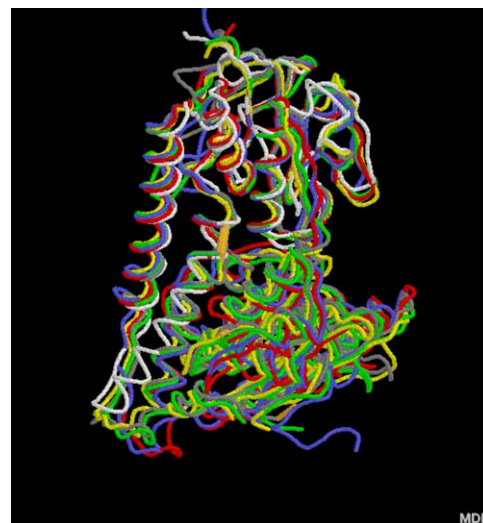


Fig. 2. The superposition of SARS-CoV ORF3 (white) with other dehydrogenases: 1OJT (lipoamide dehydrogenase from bacterium *Neisseria meningitidis*, blue), 1JEH (lipoamide dehydrogenase from yeast *Saccharomyces cerevisiae*, gray), 1LPF (lipoamide dehydrogenase from bacterium *Pseudomonas fluorescens*, green), 1EBD (dehydro-lipoamide dehydrogenase from bacterium *Bacillus stearothermophilus*, red), and 1DXL (lipoamide dehydrogenase from pea *Pisum sativum*, yellow).

ATP-binding [1], because FAD/NAD is generally used as an oxidant to yield ATP [9].

In order to collect as many evidence as possible for such a speculation, we employed VAST, DALI and CE to search the structure neighbors of ORF3. We found that the top hits, except the above templates, focus on the following dehydrogenases (also belong to the FAD/NAD-linked reductase family): 1OJT (lipoamide dehydrogenase from bacterium *Neisseria meningitides*), 1JEH (lipoamide dehydrogenase from yeast *Saccharomyces cerevisiae*), 1LPF (lipoamide dehydrogenase from bacterium *Pseudomonas*

fluorescens), 1EBD (dehydrolipoamide dehydrogenase from bacterium *Bacillus stearothermophilus*), and 1DXL (lipoamide dehydrogenase from pea *Pisum sativum*). The superpositions between these enzymes and ORF3 are shown in Fig. 2. It can be seen from the revised structure alignment (Fig. 3) that there are a number of similar structural patterns showing conservative residues (bold representation). In particular, among them are partially conserved FAD-bind motif (marked as '#') and NAD-binding motif (marked as '\$') [10–14]. This suggests that there could be a link between ORF3 and FAD/NAD-binding protein indeed.

			#####		
ORF3	1	ASTVHATATIPLQASLPFGWLVI	--VAFLAVFQSATKIIALNKR	-----WQLALYK-- 50	
1OJT	2	SADAEYDVVVL	-----GGPGGYSAFAAADE	-GLKVAIVERY-----KTLGGVCLN 47	
1JEH_A	1	TINKSHDVVVI	-----GGGPAGYVAAIKAA	-QLGFNTACVEKR-----GKLGTC-- 44	
1LPF_A	1	--SQKFDVVVI	-----GAGPGGYVAAIRAA	-QLGLKTACIEKYIGKEGKVALGGTCLN 50	
1EBD_A	1	--AIETETLVV	-----GAGPGGYVAAIRAA	-QLGQKVTIVEKG-----NLGGVCL- 42	
1DXL_C	2	SGSDENDVVII	-----GGPGGYVAAIKAA	-QLGFKTTCIEKR-----GALGGTCLN 47	
ORF3	51	G--FQFICNL	---LLLFVTIYS---HLLL	-----V-----AA-----GMEAQFLY 82	
1OJT	48	V-----GCIPSKALLHNA	AAVIDEVRH-----LAANGI	-----KYPEPELDIDMLR 87	
1JEH_A	45	LNV---GCIPSKALLNNSHL	LFHQH---HTEAQKRG-----IDVNGDI	-----KINVANFQ 88	
1LPF_A	51	V-----GCIPSKALLDSS	YKY---HEAKEAFKVGHIEAK	-----GV-----TIDVPAMV 91	
1EBD_A	43	NV---GCIPSKALISASHRYE	QAKH-----SEEMGIKAENV	-----TIDFAKVQ 83	
1DXL_C	48	V-----GCIPSKALLHSSH	MYHEAKH-----S-----FANHGVKVS	SNVEIDLAAAM 88	
ORF3	83	LYALIYF	--LQCINACRIIMRCWLCWKCKSKNPLLYDANYFVC	-----WH 126	
1OJT	88	AYKDGVSRLTGGLAGMAKSRK	VDVIQDGOFL--DPHHLEVSLTAGDAYEQAAPTGEKK 146		
1JEH_A	89	KAKDDAVKQLTGGIELLFPK	KNKVTYYKNGSFED--ETKIRVTPVDGL	-----EGTVKEDH 143	
1LPF_A	92	ARKANIVKNLTGGIATL	FKANGVTSFEGHGKLLAN--KQVEVTGL	-----DGKTQ 140	
1EBD_A	84	EWKASVVKKLTGGVEGLL	KGNKVEIVKGEAYFVD--ANTVRVNVNG	-----DSAQ 131	
1DXL_C	89	GQKDKAVSNLTRGIEGL	FKKNKVTYYKGYGKFV--SPSEISVDTI	-----EGENT 137	
				\$\$\$\$\$\$\$	
ORF3	127	THNYDYCIPY	-NSVTDITIVTE-----GDGISTPKLKE	---- 147	
1OJT	147	IVAFKNCIIAAGSRVTKLP	FPEDPRIIDSSGALALKEVPGKLLI	-IGGGIIGLEMGTVY 206	
1JEH_A	144	ILDVKNIIVATGSEVTPFP	PGIEIDEKIVSSTGALSLSKEIPKRLTI	IGGGIIGLEMGSVY 203	
1LPF_A	141	VLEAENVI	IASGSRPVEIPAPLSDDII	IVDSTGALEFQAVPKKLGVI	GAGVIGLELGSVW 200
1EBD_A	132	TYTFKNAIIATGSRPIEL	PNFKFSNRILDSTGALNLGEVPKSLV	IGGGYIGIELGTAY 191	
1DXL_C	138	VVKGKHII	IATGSDVKSPLPGVTIDEKKIVSSTGAL	LSEIPKKLVVIGAGYIGLEMGSVW 197	
ORF3				-----	
1OJT	207	STLGSRLDVVEMMDGLMQ	GADRDL--VKVWQKQNEYRFDNIMVNTKT	VAVEPKEDGVYV 262	
1JEH_A	204	SRLGSKVTVVEFPQIGAS	MDGEVAKATQKFLKKQGLDFKLSTKVI	SAKRNDKNNVEIV 263	
1LPF_A	201	ARLGAEVTVLEALDKFL	PAADEQ---IAKEALKVLTQKQLNIRL	GARVTAASEVTKQVT 256	
1EBD_A	192	ANFGTKVTILEGAGEIL	SGFEKQM---AAIIKKRLKKKGV	EVVVTNALAKGAEREDGVT 246	
1DXL_C	198	GRIGSEVTVVEFASEI	VPMTDAEI--RKQFQRSLEKQGMKFKL	KTKVVGVDTSGDGVKLT 255	
ORF3	148			-----DYQIGGYSEDRHSGVKDYVVVH 180	
1OJT	263	TFEGANAPKEPQRYDAV	LVAAAGRAPNGKLISAEKAGVA	VTDRGFIEVDKQMRNTNVPHIYA 322	
1JEH_A	264	VEDTKTNQENLEAEVLL	VAVGRRPYIAGLGAEKI	GLEVDKRGLRVIDDQFNKFPHIKV 323	
1LPF_A	257	VFTFDANGEQKETFD	KLIVAVGRRPVTTDLLAADSGVTL	DERGFIVDDHCKTSPVGVFA 316	
1EBD_A	247	VTYEANGETKTI	IDADYVLVTVGRRPNTDELGLEQI	GKMTNRGLIEVDQQCRTSPVNI	FA 306
1DXL_C	256	VEPSAGGEQTI	IEADVLLVSAGRTPFTSGLNLDK	IGVETDKLGRILVNERFSTN	VSQVYA 315
ORF3	181	-GYFTEVYYQLESTQIT	TDTGIENATFFIFNK-- 210		
1OJT	323	IG--DIVGQP	-MLAHKAVHEGHVA	AENCAGHK-- 350	
1JEH_A	324	VGD---VTFGPMLAHKA	EEEGI-AAVEMLKTGH- 351		
1LPF_A	317	IG---DVVRGAMLAHKA	SEEGVMAERIA	AGHK-- 344	
1EBD_A	307	IG--DIVPGPA	-LAHKASYEGKVA	AEIAGHP-- 334	
1DXL_C	316	IGD---VIPGPMLAHKA	EEDGV--ACVEYLAGKV 343		

Fig. 3. The structure alignment between SARS-CoV ORF3 and other dehydrogenases: 1OJT (lipoamide dehydrogenase from bacterium *Neisseria meningitides*), 1JEH (lipoamide dehydrogenase from yeast *Saccharomyces cerevisiae*), 1LPF (lipoamide dehydrogenase from bacterium *Pseudomonas fluorescens*), 1EBD (dehydrolipoamide dehydrogenase from bacterium *Bacillus stearothermophilus*), and 1DXL (lipoamide dehydrogenase from pea *Pisum sativum*). The bold representations indicate conservative residues. The FAD-binding motif is marked as '#' and the NAD-binding motif marked as '\$'.



Fig. 4. Non-canonical interactions in the structure of SARS-CoV ORF3. The residue pairs involved are (colored blue): Tyr160 (donor) and Phe43 (acceptor), Phe43 (donor) and Tyr206 (acceptor).

Finally, the non-canonical interactions in ORF3 protein structure were identified by NCI program and the results showed that there are three pairs of main chain–side chain interactions: Tyr160 (donor) and Phe43 (acceptor), Phe43 (donor) and Tyr206 (acceptor), and Ile232 (donor) and Phe231 (acceptor). Among these interactions, Phe43 forms two N–H $\cdots\pi$ bonds in a sandwich fashion: one donates to Tyr206, and one donated by Tyr160, as existed in human rac1 [15] and SARS-CoV main protease [16]. These non-canonical bindings fix the big helix Phe43 locate to the two loops Tyr160 and Tyr206 locate, hence stabilizes the structure of ORF3 protein (Fig. 4). These results can be used for rational design of mutagenesis experiments and analysis of conservation of interactions at functional sites. In recent years, the non-canonical interactions have been shown to be important for the stability of protein structure [17–19] and ligand recognition [20].

Acknowledgements

We wish to thank the Hong Kong Innovation and Technology Fund for supporting the present research.

References

- [1] M.A. Marra, S.J.M. Jones, C.R. Astell, R.A. Holt, A. Brooks-Wilson, Y.S.N. Butterfield, J. Khattri, J.K. Asano, S.A. Barber, S.Y. Chan, A. Cloutier, S.M. Coughlin, D. Freeman, N. Girn, O.L. Griffith, S.R. Leach, M. Mayo, H. McDonald, S.B. Montgomery, P.K. Pandoh, A.S. Petresou, A.G. Robertson, J.E. Schein, A. Siddiqui, D.E. Smailus, J.M. Stott, G.S. Yang, F. Plummer, A. Andonov, H. Artsob, N. Bastien, K. Bernard, T.F. Booth, D. Bowness, M. Czub, M. Drebot, L. Fernando, R. Flick, M. Garbutt, M. Gray, A. Grolla,

- S. Jones, H. Feldmann, A. Meyers, A. Kabani, Y. Li, S. Normand, U. Stroher, G.A. Tripples, S. Tyler, R. Vogrig, D. Ward, B. Watson, R.C. Brunhman, M. Krajdjen, M. Petric, D.M. Skowronski, C. Upton, R.L. Roper, The genome sequence of the SARS-associated coronavirus, *Science* 300 (2003) 1399–1404.
- [2] K. Ginalski, A. Elofsson, D. Fischer, L. Rychlewski, 3D-Jury: a simple approach to improve protein structure predictions, *Bioinformatics* 19 (2003) 1015–1018.
- [3] M. Grotthuss, L.S. Wyrwicz, L. Rychlewski, mRNA cap-1 methyltransferase in the SARS genome, *Cell* 113 (2003) 701–702.
- [4] A. Sali, T.L. Blundell, Comparative protein modeling by satisfaction of spatial restraints, *J. Mol. Biol.* 234 (1993) 779–815.
- [5] B. Wallner, A. Elofsson, Can correct protein models be identified?, *Protein Sci.* 12 (2003) 1073–1086.
- [6] I.N. Shindyalov, P.E. Bourne, Protein structure alignment by incremental combinatorial extension (CE) of the optimal path, *Protein Eng.* 11 (1998) 739–747.
- [7] A. Zemla, LGA: a method for finding 3D similarities in protein structures, *Nucleic Acids Res.* 31 (2003) 3370–3374.
- [8] M.M. Babu, NCI: a server to identify non-canonical interactions in protein structures, *Nucleic Acids Res.* 31 (2003) 3345–3348.
- [9] O. Carugo, P. Argos, NADP-dependent enzymes. II: evolution of the mono- and dinucleotide binding domains, *Proteins* 28 (1997) 29–40.
- [10] I. Li de la Sierra, L. Pernot, T. Prange, P. Saludjian, M. Schiltz, R. Fourme, G. Padron, Molecular structure of the lipoamide dehydrogenase domain of a surface antigen from *Neisseria meningitidis*, *J. Mol. Biol.* 269 (1997) 129–141.
- [11] T. Toyoda, K. Suzuki, T. Sekigushi, J. Reed, A. Takenaka, Crystal structure of eucaryotic E3, lipoamide dehydrogenase from yeast, *J. Biochem. (Tokyo)* 123 (1998) 668–674.
- [12] A. Mattevi, G. Obmolova, K.H. Kalk, W.J. van Berkel, W.G. Hol, Three-dimensional structure of lipoamide dehydrogenase from *Pseudomonas fluorescens* at 2.8 Å resolution. Analysis of redox and thermostability properties, *J. Mol. Biol.* 230 (1993) 1200–1215.
- [13] S.S. Mande, S. Sarfaty, M.D. Allen, R.N. Perham, W.G. Hol, Protein–protein interactions in the pyruvate dehydrogenase multienzyme complex: dihydrolipoamide dehydrogenase complexed with the binding domain of dihydrolipoamide acetyltransferase, *Structure* 4 (1996) 277–286.
- [14] M. Faure, J. Bourguignon, M. Neuburger, D. Machereel, L. Sieker, R. Ober, R. Kahn, C. Cohen-Addad, R. Douce, Interaction between the lipoamide-containing H-protein and the lipoamide dehydrogenase L-protein of the glycine decarboxylase multienzyme system. 2. Crystal Structure of H-and L-Proteins, *Eur. J. Biochem.* 267 (2000) 2882–2889.
- [15] 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.
- [16] 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.
- [17] G.F. Fabiola, S. Krishnaswamy, V. Nagarajan, V. Pattabhi, C–H \cdots O hydrogen bonds in beta sheets, *Acta Crystallog. Sec. D.* 53 (1997) 316–320.
- [18] A. Senes, I. Ubarretxena-Belandia, D.M. Engelman, The C–H \cdots O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions, *Proc. Natl. Acad. Sci. USA* 98 (2001) 9056–9061.
- [19] M.M. Babu, S. Singh, P. Balam, A C–H \cdots 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. 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.