ELSEVIER

Tertiary structure prediction of SARS coronavirus helicase

Andrea Bernini [a,b], Ottavia Spiga [a,b], Vincenzo Venditti [a], Filippo Prischi [a], Luisa Bracci [a],
Jiandong Huang [c], Julian A. Tanner [c], Neri Niccolai [a,b,*]

[a] *Biomolecular Structure Research Center, Department of Molecular Biology, University of Siena, I-53100 Siena, Italy*
[b] *SienaBiografix Srl, Siena, Italy*
[c] *Department of Biochemistry, University of Hong Kong, Pokfulam, Hong Kong, China*

Received 7 March 2006
Available online 23 March 2006

Abstract

SARS coronavirus, SCV, has been recently responsible of a sudden and widespread infection which caused almost 800 victims. The limited amount of SCV protein structural information is partially responsible of the lack of specific drugs against the virus. Coronavirus helicases are very conserved and peculiar proteins which have been proposed as suitable targets for antiviral drugs, such as bananins, which have been recently shown to inhibit the SCV helicase in vitro. Here, the quaternary structure of SCV helicase has been predicted, which will provide a solid foundation for the rational design of other antiviral helicase inhibitors.
© 2006 Elsevier Inc. All rights reserved.

Keywords: SARS coronavirus; Protein structure; Structure prediction; Molecular modeling; Helicase structure; Drug design

Severe acute respiratory syndrome coronavirus (SCV) was the protagonist of a recent severe world-wide spread infection which caused almost 800 deaths within few months between 2002 and 2003 [1]. Genomic investigations have yielded a wealth of information on SCV evolution in terms of sequence mutations [2–4] to understand mechanism of viral transmission between animals and humans. Since SCV is still circulating, so far almost asymptomatically in southern China, our preparedness against new dangerous coronavirus outbreaks should include the development of specific antiviral drugs which, at the moment, remain unavailable. In this respect, structural genomics can play an important role in gaining insight into viral protein targets. Inhibition of these targets may interfere with the metabolism of the infecting virus without strong side effects on patients.

Helicase have been targeted for therapy in other viruses, suggesting that the SCV helicase may be a suitable target for new anti-SCV drugs [5–9]. The SCV helicase tertiary structure, not yet experimentally solved, would be useful for designing specific inhibitors of this SCV enzyme. Here, we report a structure prediction procedure to obtain a reliable molecular model of this viral protein.

Materials and methods

The sequence of SCV helicase (576 aminoacid residues with NCBI Accession No. NP_828870) [10] was aligned with that of representatives of the other groups of coronavirus. Sequence alignment of SCV helicase with other coronavirus helicases were obtained using program ClustalW v.1.8 [11]. The ATP binding site motif was predicted with MotifScan software available at the Expasy web-site [12]. Secondary structure prediction and fold recognition studies were performed by using PsiPred v.2.1 [13] and Pyre v.0.2 [14], respectively. Using the fold recognition method the helicase of SARS Coronavirus (SwissProt Accession ID P59641) [15] identifies the 1UAA pdb entry [16]. Model building of the major domain (494 aminoacids) of SARS helicase was carried out using ClustalW v.1.8 [11] alignment between the target sequence and that of the template structure. Models were subsequently optimized according to secondary structure predictions. Substitution of aminoacid residues and modeling of insertions and deletions in the target structure were performed using DeepView v.3.7 software [17]. The 3D models were optimized by a 900 step minimization run with AMBER [18] and finally validated with the PROCHECK v.3.5.4 procedure [19] and PROSAII v.4.0 software [20].

A model of the metal binding domain (MBD) was obtained using DYANA v.1.5 software [21]. The insertion of zinc atoms was carried out using GROMACS v.3.2 [22] software using the force field ffgmx43a2 that allows a molecular minimization of this domain with the bound metal atoms. To investigate the structure of MBD and helicase domain, a docking simulation was performed using ESCHER software [23]. A first coarse run with a rotation step of 10° was carried out and approximately 500 results were collected. Only those models that were consistent with the available experimental data were subjected to a second run with a rotation step of 2° between ± 20° of the starting position in order to refine the most probable structures. The major side-chain clashes were removed by a 900 cycles minimization in the AMBER force Field [18].

Results and discussion

A set of 35 different amino acid sequences of SCV helicase proteins, collected from the NCBI protein database, was aligned using the ClustalW v.1.8 software [11] and among all the considered sequences pairwise identities of 99% were observed. Thus, a representative consensus sequence was obtained for the SCV helicase, suggesting the presence of two separate domains, i.e., the helicase domain (Hel) and a metal binding domain (MBD). Such a combination of a putative MBD and a Hel domain in the same protein, as in Equine Arterivirus nsp10, has been found in a number of other viral and cellular proteins [24–26]. Accordingly, the nsp13 model building was performed in two separate steps, one for each domain of the enzyme, and a hybrid protein structure was assembled from the two modeled domains.

A 3D model of the SCV helicase Hel domain, spanning residues 80–568, was made on the basis of the crystal structure of the *Escherichia coli* Rep helicase. Both helicases belong to the same Superfamily 1 helicase classification on the basis of seven conserved sequence motifs [27]. Therefore, the Protein Data Bank entry 1UAA [16], corresponding to *E. coli* Rep helicase crystal structure, was used as a suitable template for a model building of the Hel domain using the threading program Pyre v.0.2 [14]. A manual alignment between the secondary structure elements predicted by shuffled PsiPred runs [13] and the ones observed in the model enhanced the correspondence of identical and positively conserved residues, respectively to 11% and 23%.

As far as the MBD structure is concerned, it should be noted that all coronaviruses exhibit very different amino acid sequences. However, coronavirus MBDs, as well as all the other nidoviruses, have a Cys and His rich motif which can orient the overall modeling procedure of this domain. In fact, arterivirus helicases have been shown to possess MBD domains which coordinate four Zn^{2+} [28,29] in a way that could be similarly adopted by coronaviruses. In this Zn^{2+} binding domain, known also as binuclear cluster, a Cys residue coordinates two closely spaced Zn^{2+} ions bridging a Zn_2Cys_6 group, typically found in GAL4-like proteins [30] with $Zn_2Cys_4His_2$, found in RAG1 domain [31]. The SCV MBD structure has been reconstructed using tetrahedral geometry characterization, distance constraints, and orientations typical of Cys and

His residues of the latter clusters, suggested by the crystal structures of GAL4-like and RAG1. The corresponding Protein Data Bank files, 1PYI, and 1RMD, respectively [30,31], provided accurate distance constraints for the two components of the MBD.

The reliability of the predicted tertiary structures of both SCV helicase domains was tested on the basis of simple physico-chemical rules, such as the ones suggested by Ramachandran plots, and the agreement between hydropathy profile and residue accessibility in the two modeled structures. Additional features which support the predicted structure of the Hel domain are given by the fact that 10 out of a total 15 cysteine residues, which are predicted to form cystine bonds, are involved in disulphide bridges. After refinements of loop regions and manual tracing of side-chains for both Hel and MBD models no severe disallowed atomic contact was detected with PROCHECK v.3.5.4 [19], suggesting an essentially good stereochemistry, with 63% and 29% of the residues in the most favored and additional allowed regions, respectively, and with 4.8% and 3.2% residues in generously allowed and disallowed regions of the Ramachandran plot.

Once the models of the Hel and MBD domains had been successfully predicted, a molecular docking simulation was carried out to obtain the complete nsp13 tertiary structure in an unbiased way. Among the lowest energy solutions, only the ones consistent with the overall protein sequence were taken into account and manually refined before a final energy minimization. The theoretical structure of SCV nsp13, shown in Fig. 1, was again confirmed by
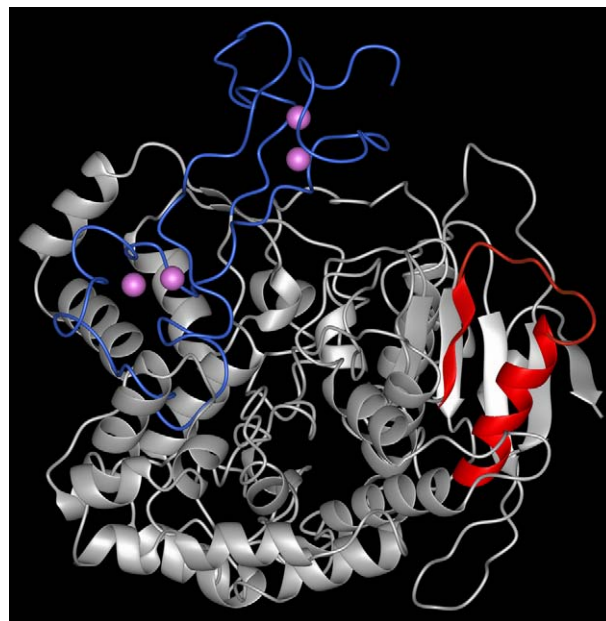

Fig. 1. Ribbon representations of SCV nsp13 tertiary structure. In blue the MBD domains is shown and in red and violet the P-loop and zinc atoms are highlighted. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)
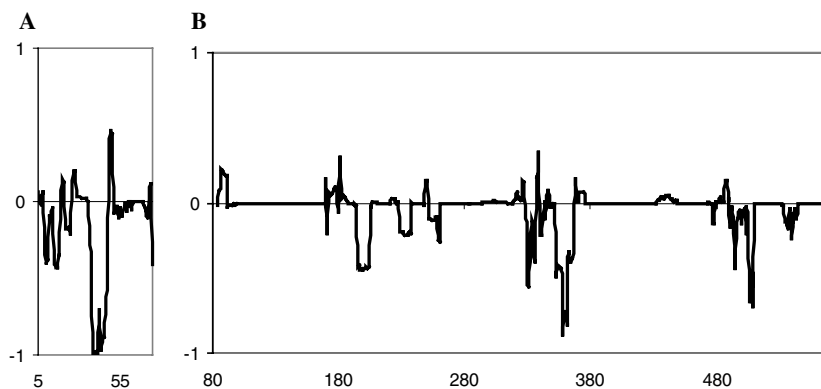
Fig. 2. The difference of the energy profile obtained by using ProSaII program package for the entire nsp13 proteins and MBD is shown in (A). (B) The same difference is calculated for the Hel domain.

PROCHECK v.3.5.4 [19] and deposited in the Protein Data Bank with the ID code 2G1F.

Reliability of protein structures may be conveniently assessed by PROSAII program package [20] and, accordingly, the predicted single components of the SCV helicase as well as the final tertiary structure have been considered by this program package. In Fig. 2, we compared the energy profiles obtained with PROSAII program package [20] for the isolated Hel and MBD domains with the one calculated for the entire SCV helicase structure. From the negative values of the energy, which have been obtained by subtracting the energy profile of the two separate domains from the one calculated for the final protein structure, see Fig. 2, it is apparent that higher stability is achieved when the two domains are assembled. Negative peaks in the latter profiles, indeed, are found for the interface regions of the two domains, i.e., 42–48 residues for MBD and 198–204, 325–335, and 353–360 residues for the Hel domain, respectively.

Since a typical ATP binding site motif between residues 282 and 289 of the helicase sequence may be predicted, it is noteworthy that in the obtained nsp13 model the corresponding fragment, the so-called P-loop, is surface exposed and located between a β strand and a short α helix, exactly as this active site should be [32], see Fig. 1.

We conclude that a theoretical model of SCV helicase is here proposed which will provide a foundation for the rational design of helicase inhibitors that could potentially act as anti-SCV drugs.

References

[1] World Health Organization. Summary of probable SARS cases with onset of illness from 1st November 2002 to 31st July 2003. Based on data as of the 31st December 2003. Available from: http://www.who.int/csr/sars/country/table2004_04_21/en/.

[2] H.D. Song, C.C. Tu, G.W. Zhang, S.Y. Wang, K. Zheng, L.C. Lei, Q.X. Chen, Y.W. Gao, H.Q. Zhou, H. Xiang, H.J. Zheng, S.W. Chern, F. Cheng, C.M. Pan, H. Xuan, S.J. Chen, H.M. Luo, D.H. Zhou, Y.F. Liu, J.F. He, P.Z. Qin, L.H. Li, Y.Q. Ren, W.J. Liang, Y.D. Yu, L. Anderson, M. Wang, R.H. Xu, X.W. Wu, H.Y. Zheng, J.D. Chen, G. Liang, Y. Gao, M. Liao, L. Fang, L.Y. Jiang, H. Li, F. Chen, B. Di, L.J. He, J.Y. Lin, S. Tong, X. Kong, L. Du, P. Hao, H. Tang, A. Bernini, X.J. Yu, O. Spiga, Z.M. Guo, H.Y. Pan, W.Z. He, J.C. Manuguerra, A. Fontanet, A. Danchin, N. Niccolai, Y.X. Li, C.I. Wu, G.P. Zhao, Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human, Proc. Natl. Acad. Sci. USA 102 (2005) 2430–2435.

[3] K.A. Ivanov, V. Thiel, J.C. Dobbe, Y. van der Meer, E.J. Snijder, J. Ziebuhr, Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase, J. Virol. 78 (2004) 5619–5632.

[4] V. Thiel, K.A. Ivanov, A. Putics, T. Hertzig, B. Schelle, S. Bayer, B. Weissbrich, E.J. Snijder, H. Rabenau, H.W. Doerr, A.E. Gorbalenya, J. Ziebuhr, Mechanisms and enzymes involved in SARS coronavirus genome expression, J. Gen. Virol. 84 (2003) 2305–2315.

[5] G. Kleymann, R. Fischer, U.A. Betz, M. Hendrix, W. Bender, U. Schneider, G. Handke, P. Eckenberg, G. Hewlett, V. Pevzner, et al., New helicase-primase inhibitors as drug candidates for the treatment of herpes simplex disease, Nat. Med. 8 (2002) 392–398.

[6] J.J. Crute, C.A. Grygon, K.D. Hargrave, B. Simoneau, A.M. Faucher, G. Bolger, P. Kibler, M. Liuzzi, M.G. Cordingley, Herpes simplex virus helicase-primase inhibitors are active in animal models of human disease, Nat. Med. 8 (2002) 386–391.

[7] P. Borowski, S. Schalinski, H. Schmitz, Nucleotide triphosphatase/helicase of hepatitis C virus as a target for antiviral therapy, Antiviral Res. 55 (2002) 397–412.

[8] J.A. Tanner, R.M. Watt, Y.B. Chai, L.Y. Lu, M.C. Lin, J.S.M. Peiris, L.L.M. Poon, H.F. Kung, J.D. Huang, The severe acute respiratory syndrome (SARS) coronavirus NTP/helicase belongs to a distinct class of 5′ to 3′ viral helicase, J. Biol. Chem. 278 (2003) 39578–39582.

[9] R.Y. Kao, W.H. Tsui, T.S. Lee, J.A. Tanner, R.M. Watt, J.D. Huang, L. Hu, G. Chen, Z. Chen, L. Zhang, T. He, K.H. Chan, H. Tse, A.P. To, L.W. Ng, B.C. Wong, H.W. Tsoi, D. Yang, D.D. Ho, K.Y. Yuen, Identification of novel small-molecule inhibitors of severe acute respiratory syndrome-associated coronavirus by chemical genetics, Chem. Biol. 11 (2004) 1293–1299.

[10] http://www.ncbi.nlm.nih.gov/.

[11] 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.

[12] http://myhits.isb-sib.ch/cgi-bin/motif_scan.

[13] D.T. Jones, Protein secondary structure prediction based on position-specific scoring matrices, J. Mol. Biol. 292 (1999) 195–202.

[14] http://www.sbg.bio.ic.ac.uk/~phyre/.

[15] http://www.expasy.ch/sprot/.

[16] S. Korolev, J. Hsieh, G.H. Gauss, T.M. Lohman, G. Waksman, Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP, Cell 90 (1997) 635–647.

[17] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, Electrophoresis 18 (1997) 2714–2723.

[18] D.A. Case, T.E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K.M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, R. Woods, The Amber biomolecular simulation programs, J. Computat. Chem. 26 (2005) 1668–1688.

[19] 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. Cryst. 26 (1993) 283–291.

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

[21] P. Güntert, K. Wüthrich, Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints, J. Biomol. NMR 1 (4) (1991) 447–456.

[22] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: fast, flexible, and free, J. Comput. Chem. 26 (16) (2005) 1701–1718.

[23] G. Ausiello, G. Cesareni, M. Helmer-Citterich, ESCHER: a new docking procedure applied to the reconstruction of protein tertiary structure, Proteins 28 (4) (1997) 556–567.

[24] S. Dracheva, E.V. Koonin, J.J. Crute, Identification of the primase active site of the herpes simplex virus type 1 helicase-primase, J. Biol. Chem. 270 (23) (1995) 14148–14153.

[25] R.E. Johnson, S.T. Henderson, T.D. Petes, S. Prakash, M. Bankmann, L. Prakash, *Saccharomyces cerevisiae* RAD5-encoded DNA repair protein contains DNA helicase and zinc-binding sequence motifs and affects the stability of simple repetitive sequences in the genome, Mol. Cell. Biol. 12 (9) (1992) 3807–3818.

[26] T. Kusakabe, The role of the zinc motif, J. Biol. Chem. 271 (1996) 19563–19570.

[27] A.E. Gorbalenya, E.V. Koonin, A.P. Donchenko, V.M. Blinov, Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis, Nucleic Acids Res. 17 (1989) 4847–4861.

[28] L.C. Van Dinten, H. van Tol, A.E. Gorbalenya, E.J. Snijder, The predicted metal-binding region of the arterivirus helicase protein is involved in subgenomic mRNA synthesis, genome replication, and virion biogenesis, J. Virol. 74 (11) (2000) 5213–5223.

[29] A. Seybert, C.C. Posthuma, L.C. van Dinten, E.J. Snijder, A.E. Gorbalenya, J. Ziebuhr, A complex zinc finger controls the enzymatic activities of nidovirus helicases, J. Virol. 79 (2) (2005) 696–704.

[30] R. Marmorstein, M. Carey, M. Ptashne, S.C. Harrison, DANN recognition by GAL4: structure of a protein–DNA complex, Nature 356 (1992) 379–380.

[31] S.F. Bellon, K.K. Rodgers, D.G. Schatz, J.E. Coleman, T.A. Steitz, Crystal structure of the RAG1 dimerization domain reveals multiple znc-binding motifs including a novel zinc binuclear cluster, Nat. Struct. Biol. 4 (1997) 586–591.

[32] M. Saraste, P.R. Sibbald, A. Wittinghofer, The P-loop—a common motif in ATP- and GTP-binding proteins, Trends Biochem Sci. 15 (1990) 430–434.