Acta Cryst. (1998). D**54**, 121–123

Crystallization and preliminary X-ray crystallographic analysis of the helicase domain of hepatitis C virus NS3 protein

LIN-WOO KANG,[a] HYUN-SOO CHO,[a] SUN-SHIN CHA,[a] KYUNG MIN CHUNG,[a] SUNG HOON BACK,[a] SUNG KEY JANG[a] and BYUNG-HA OH[a,b]* *at* [a]*Department of Life Sciences, Pohang University of Science of Technology, Pohang, Kyungbuk, 790-784, S. Korea, and* [b]*School of Environmental Engineering, Pohang University of Science of Technology, Pohang, Kyungbuk, 790-784, S. Korea. E-mail: boh@vision.postech.ac.kr*

(*Received 11 April 1997; accepted 13 June 1997*)

Abstract

The NS3 protein of hepatitis C virus (HCV) is thought to be essential for viral replication. The N-terminal domain of the protein contains protease activity and the C-terminal domain contains nucleotide triphosphatase and RNA helicase activity. The RNA helicase domain of HCV NS3 protein was purified by using affinity-column chromatographic methods, and crystallized by using the microbatch crystallization method under oil at 277 K. The crystals belong to primitive trigonal space group $P3_121$ or $P3_221$ with cell dimensions of $a = b = 93.3$, $c = 104.6$ Å. The asymmetric unit contains one molecule of the helicase domain, with the crystal volume per protein mass (V_m) of 2.50 Å^3 Da^{-1} and solvent content of about 50.8% by volume. A native data set to 2.3 Å resolution was obtained from a frozen crystal indicating that the crystals are quite suitable for structure determination by multiple isomorphous replacement.

1. Introduction

Many proteins exhibiting RNA helicase activity have been found ubiquitously in organisms ranging from viruses to humans. Several examples are pestivirus NS3 (Warrener & Collett, 1995), *Escherichia coli* transcription termination factor rho (Brennan *et al.*, 1987), eukaryotic translation initiation factors (Rozen *et al.*, 1990), and RNA helicase I (Claude *et al.*, 1991). The unwinding of double-stranded RNA is thought to be important in transcription, RNA processing, translation and RNA replication (Gorbalenya *et al.*, 1989; Ray *et al.*, 1985). Most of known RNA helicases are thought to bind to a single strand region and unwind the RNA duplex unidirectionally or bidirectionally (Claude *et al.*, 1991; Rozen *et al.*, 1990). They require NTP and a divalent cation for the helicase activity (Kim *et al.*, 1995). Mutation studies showed that a short stretch of amino-acid sequence, Asp-Glu-Ala-Asp (DEAD box), is essential for the RNA helicase activity. The family of proteins with the DEAD box consists of three sub-families according to the sequence difference in this segment, DEAD, DEAH and DExH protein families (Koonin, 1991).

Hepatitis C virus (HCV) is the major etiologic agent of non-A non-B hepatitis (Choo *et al.*, 1989; Kuo *et al.*, 1989). HCV belongs to *Flaviviridae*, the positive-strand RNA virus family (Inchauspe *et al.*, 1991). Its genome consists of about 9400 nucleotides with gene order of N′-C-E1-E2-NS2-NS3-NS4A-NS4B-NS5A-NS5B-C′ (Grakoui *et al.*, 1993*b*) which encodes a viral polyprotein of about 3010 residues (Kato *et al.*, 1990). The polyprotein is processed into functional proteins by host- and virus-encoded proteases. While the N-terminal portion of the polyprotein consists of structural proteins (core and envelope proteins), the other portion contains non-structural proteins which are the essential catalytic machinery for the life cycle of HCV. The N-terminal one third of NS3 is a serine protease domain (Grakoui *et al.*, 1993*a*) which is believed to cleave the NS3/4A, NS4A/4B, NS4B/5A and NS5A/5B junctions (Bartenschlager *et al.*, 1993; Eckart *et al.*, 1993; Hahm *et al.*, 1995), and, thus, is a potential target for developing therapeutic agents inhibiting the viral replication. Recently, the three-dimensional structure of the protease domain were determined by two groups independently (Kim *et al.*, 1996; Love *et al.*, 1996). The C-terminal two thirds of the HCV NS3 protein is an RNA helicase domain exhibiting nucleotide triphosphatase (NTPase)/RNA helicase activity (Kim *et al.*, 1995). This domain contains a conserved sequence of QR*x*GR*x*GR*xxx*G required for ATP hydrolysis (Gross & Shuman, 1996). The HCV RNA helicase domain also contains DECH sequence, and belongs to the DExH protein family. It unwinds DNA/DNA duplexes and RNA/DNA heteroduplexes as well as RNA/RNA duplexes and translocates only in the 3′ to 5′ direction (Gwack *et al.*, 1996; Tai *et al.*, 1996). The NTPase/RNA helicase activity was found in all the three genera of the positive-strand RNA virus families of *Flaviviridae*, *Flavivirus*, *Pestivirus* and HCV (Suzich *et al.*, 1993). Besides, the RNA helicase domains are always found in relatively large positive-strand RNA viral genomes larger than 5 kb (Koonin, 1991). These reports support the importance of the RNA helicase activity in RNA viral replication, and implicate the RNA helicase as a target for developing new types of antiviral agents. Up to now, there is no report on the three-dimensional structure of an RNA helicase although a DNA helicase structure was recently reported (Subramanya *et al.*, 1996). The structural information of the HCV helicase domain should lead to a better understanding of the catalytic mechanism of all other RNA helicases at the atomic level. Here, we report the crystallization and preliminary X-ray crystallographic analyses of the HCV helicase domain.

2. Protein purification

For overexpression of the helicase domain (471 amino acids) of NS3, a DNA fragment encoding the helicase domain encompassing amino acids 1186–1657 from the cDNA genome of a HCV strain, group I type 1b, was subcloned in the *Nhe*I site of the pRSET-A vector. The helicase domain was fused to 14 extra amino acids (MRGSHHHHHHGMAS) at its N termius for an affinity chromatographic purification of the protein, and six extra amino acids (AAAREA) at its C terminus from the pRSET vector. The vector was introduced into *Escherichia coli* BL21 (DE3) strain harboring pLysS plasmid. The expression of the helicase protein was induced by 1 m*M* isopropyl-D-thiogalactopyranoside. The protein was purified using an

Ni–NTA agarose column (Qiagen, Germany) and a poly-(U) Sepharose column (Pharmacia, Sweden) successively. The purificiation scheme yielded an apparently homogeneous protein with about 55 kDa molecular mass on an SDS–PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis) gel, in accordance with the value calculated from the deduced amino-acid sequence (52 577 Da). The purified protein exhibited RNA helicase activity and NTPase activity which are promoted in the presence of an RNA substrate. By using a vacuum-dialysis membrane (Schleicher & Schuell, Germany), the protein solution was dialyzed and concentrated to 10 mg ml^{-1} against a buffer solution containing 10 m*M* HEPES (pH 7.5), 1 m*M* $MgCl_2$, 5 m*M* β-mercaptoethanol, 10% glycerol, and 0.5 *M* NaCl.

3. Crystallization and X-ray studies

2 µl of final protein sample were mixed with an equal volume of precipitant solution containing 30% polyethylen glycol 4000 (PEG 4000), 0.1 *M* sodium cacodylate (pH 6.5), 0.2 *M* ammonium acetate. The mixture was placed on the microbatch plate under Al's oil (Hampton Research, USA) at 277 K. Crystals appeared after a month and reached a final size of about 0.2 × 0.2 × 0.7 mm in three months (Fig. 1). In order to reduce the crystal growth time, the macroseeding method was employed. After small crystals were placed for 5 h in a precipitant solution containing 15% PEG 4000, 0.1 *M* sodium cacodylate (pH 6.5), and 0.2 *M* ammonium acetate, the crystals were transferred to the mixture on the microbatch plate containing 2 µl of the protein solution and an equal volume of the precipitant solution. The crystals reached a full size in 2 to 3 weeks. The diffraction data were collected using graphite-monochromatized Cu $K\alpha$ radiation from a MacScience M18XAF X-ray generator operated at 50 kV and 90 mA. The crystals were sensitive to radiation damage, and for full data collection, they were frozen at 110 K by using a cryocooling system (Oxford Cryosystems Cryostream). Before X-ray data collection, the crystals were soaked for 2–3 h in a cryoprotectant solution which consisted of 15% PEG 4000, 0.1 *M* sodium cacodylate (pH 6.5) and 20% glycerol. With a frozen crystal, diffraction data to 2.3 Å resolution (93.4% complete, R_{sym} = 4.0%) were obtained. The data were processed and scaled using the programs *DENZO* and *SCALEPACK* (Otwinowski, 1993). The crystals belong to the trigonal space group $P3_121$ or $P3_221$ with unit-cell dimensions of $a = b = 93.3$, $c = 104.6$ Å. We calculated that the crystal volume per molecular weight (V_m) is 2.50 $Å^3\ Da^{-1}$ with a solvent content of 50.8% by volume (Matthews, 1968) when one unit cell was assumed to contain six protomers (one molecule in the asymmetric unit of the crystal). A search for suitable heavy-atom derivative crystals is in progress.

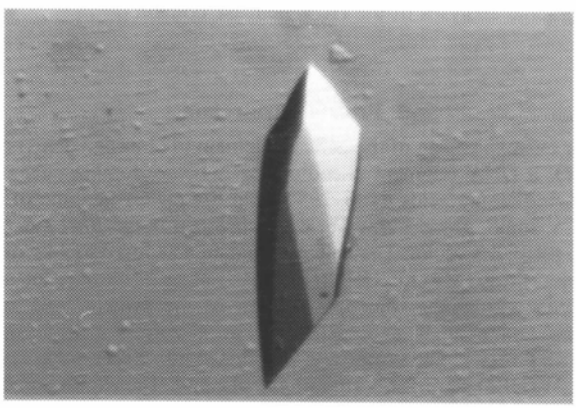

Fig. 1. Native crystals of the HCV RNA helicase domain (0.2 × 0.2 × 0.7 mm).

This study made use of the X-ray facility at Pohang Light Source, and was supported by the Basic Science Research Institute Program, Ministry of Education (BSRI-96-4434) and in part by KOSEF through the Research Center for New Bio-Materials in Agriculture at Seoul National University.

References

Bartenschlager, R., Ahlborn-Laake, L., Mous, J. & Jacobsen, H. (1993). *J. Virol.* **67**, 3835–3844.
Brennan, C. A., Dombroski, A. J. & Platt, T. (1987). *Cell*, **48**, 945–952.
Choo, Q., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W. & Houghton, M. (1989). *Science*, **244**, 359–361.
Claude, A., Arenas, J. & Hurwitz, J. (1991). *J. Biol. Chem.* **266**, 10358–10367.
Eckart, M. R., Selby, M., Masiarz, F., Lee, C., Berger, K., Crawford, K., Kuo, C., Kuo, G., Houghton, M. & Choo, Q.-L. (1993). *Biochem. Biophys. Res. Commun.* **192**, 399–406.
Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1989). *Nucleic Acids Res.* **17**, 4713–4730.
Grakoui, A., McCourt, D. W., Wychowcki, C., Feinstone, S. M. & Rice, C. M. (1993*a*). *Proc. Natl Acad. Sci. USA*, **90**, 10583–10587.
Grakoui, A., McCourt, D. W., Wychowski, C., Feinstone, S. M. & Rice, C. M. (1993*b*). *J. Virol.* **67**, 2832–2843.
Gross, C. H. & Shuman, S. (1996). *J. Virol.* **70**, 1706–1713.
Gwack, Y., Kim, D. W., Han, J. H. & Choe, J. (1996). *Biochem. Biophys. Res. Commun.* **225**, 654–659.
Hahm, B., Han, D. S., Back, S. H., Song, O.-K., Cho, M.-J., Kim, C.-J., Shimotohno, K. & Jang, S. K. (1995). *J. Virol.* **69**, 2534–2539.
Inchauspe, G., Zebedee, S., Lee, D.-H., Sugitani, M., Nasoff, M. & Prince, A. M. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 10292–10296.
Kato, N., Hijikatu, M., Ostsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T. & Shimotohno, K. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 9524–9528.
Kim, D. W., Gwack, Y., Han, J. H. & Choe, J. (1995). *Biochem. Biophys. Res. Commun.* **215**, 160–166.
Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R. & Thomson, J. A. (1996). *Cell*, **87**, 343–355.
Koonin, E. V. (1991). *Nature (London)*, **352**, 290.
Kuo, G., Choo, Q.-L., Alter, H. L., Gitnick, G. I., Redeker, A. G., Purcell, R. H., Miyamura, T., Dienstag, J. L., Alter, M. J., Stevens, C. E., Tegtmeier, G. E., Bonino, F., Colombo, M., Lee, W. S., Kuo, C., Berger, K., Shuster, J. R., Overby, L. R., Bradley, D. W. & Houghton, M. (1989). *Science*, **244**, 362–364.
Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W., Adachi, T. & Hostomska, Z. (1996). *Cell*, **87**, 331–342.
Matthews, B. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Ray, B. K., Lawson, T. G., Kramer, J. C., Cladaras, M. H., Grifo, J. A., Abramson, R. D., Merrick, W. C. & Thach, R. E. (1985). *J. Biol. Chem.* **260**, 7651–7658.
Rozen, F., Edery, I., Meerovich, K., Dever, T. E., Merrick, W. C. & Sonenberg, N. (1990). *Mol. Cell. Biol.* **10**, 1134–1144.
Subramanya, H. S., Bird, L. E., Brannigan, J. A. & Wigley, D. B. (1996). *Nature (London)*, **384**, 379–383.

Suzich, J. A., Tamura, J. K., Palmer-Hill, F., Warrener, P., Grakoui, A., Rice, C. M., Feinstone, S. M. & Collett, M. S. (1993). *J. Virol.* **67**, 6152–6158.

Tai, C.-L., Chi, W.-K., Chen, D.-S. & Hwang, L.-H. (1996). *J. Virol.* **70**, 8477–8484.

Warrener, P. & Collett, M. S. (1995). *J. Virol.* **69**, 1720–1726.