Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Cloning, purification and preliminary crystallographic studies of the 2AB protein from hepatitis A virus

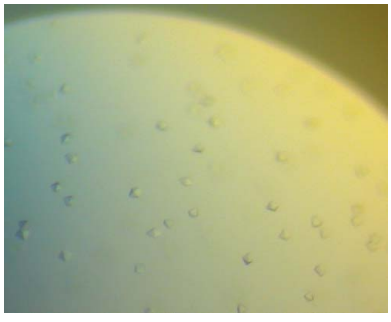
Damià Garriga,[a]‡§ Laia Vives-Adrián,[a]‡ Mònica Buxaderas,[a] Frederico Ferreira-da-Silva,[b] Bruno Almeida,[b] Sandra Macedo-Ribeiro,[b] Pedro José Barbosa Pereira[b] and Núria Verdaguer[a]*

[a]Institut de Biologia Molecular de Barcelona, CSIC, Parc Científic de Barcelona, Baldiri i Reixac 10, 08028 Barcelona, Spain, and [b]IBMC – Instituto de Biologia Molecular e Celular, Universidade do Porto, 4150-180 Porto, Portugal

‡ These authors contributed equally to this work.
§ Present address: Centro Nacional de Biotecnología, CSIC, Darwin 3, 28049 Madrid, Spain.

Correspondence e-mail: nvmcri@ibmb.csic.es

Received 17 May 2011
Accepted 1 July 2011



The *Picornaviridae* family contains a large number of human pathogens such as rhinovirus, poliovirus and hepatitis A virus (HAV). Hepatitis A is an infectious disease that causes liver inflammation. It is highly endemic in developing countries with poor sanitation, where infections often occur in children. As in other picornaviruses, the genome of HAV contains one open reading frame encoding a single polyprotein that is subsequently processed by viral proteinases to originate mature viral proteins during and after the translation process. In the polyprotein, the N-terminal P1 region generates the four capsid proteins, while the C-terminal P2 and P3 regions contain the enzymes, precursors and accessory proteins essential for polyprotein processing and virus replication. Here, the first crystals of protein 2AB of HAV are reported. The crystals belonged to space group $P4_1$ or $P4_3$, with unit-cell parameters $a = b = 90.42$, $c = 73.43$ Å, and contained two molecules in the asymmetric unit. Native and selenomethionine-derivative crystals diffracted to 2.7 and 3.2 Å resolution, respectively.

1. Introduction

Hepatitis A virus (HAV), transmitted by the faecal–oral route, is a common cause of acute hepatitis worldwide. As a member of the *Picornaviridae* family, HAV possesses a nonenveloped icosahedral capsid of 30 nm in diameter which encapsidates a positive-sense single-stranded RNA genome of approximately 7.5 kb covalently linked to a VPg peptide (Weitz *et al.*, 1986). Nevertheless, HAV is biologically distinct from other picornaviruses as the infections that it causes are characterized by slow viral replication and the absence of induced cellular lysis. Another distinct feature of HAV is that it has developed a highly deoptimized codon usage. This strategy is thought to play a role in escaping the host-cell defences, allowing the virus to grow in a quiescent way (Pintó *et al.*, 2007). Thus, HAV presents very low translation and RNA-replication activity (Harmon *et al.*, 1989; Wheeler *et al.*, 1986).

The HAV genome codes for a single polyprotein, which is cleaved into mature proteins by the viral proteinase $3C^{pro}$ and host proteases (Schultheiss *et al.*, 1994). This polyprotein precursor is organized into three regions named P1, P2 and P3. The P2 region comprises proteins 2A, 2B and 2C. HAV 2A (71 amino acids in length) has no sequence similarity to any other picornaviral protein and lacks proteolytic activity (Hollinger & Emerson, 2007). The 2A region within the precursor VP1-2A plays an essential role in early particle assembly as it is required for the formation of the pentamer intermediates (Cohen *et al.*, 2002; Morace *et al.*, 2008). Typical picornavirus 2B proteins (∼28 kDa) are membrane-associated proteins with a hydrophobic region containing a conserved putative amphiphatic α-helix at their C-terminus (de Jong *et al.*, 2008; van Kuppeveld *et al.*, 1996). The HAV 2B protein is approximately 100 amino acids longer than those of other picornaviruses and also appears to be involved in membrane rearrangements during HAV infection (Gosert *et al.*, 2000; Jecht *et al.*, 1998). No structural information is available for any picornaviral 2B protein. In this paper, we report the cloning, expression, purification and preliminary X-ray crystallographic studies of the nonstructural region 2AB of HAV. Native and selenomethionine-derivative crystals diffracted to 2.7 and 3.2 Å resolution, respectively, and were suitable for three-dimensional structure determination.

Table 1
Data-collection and processing statistics.

Values in parentheses are for the outermost resolution shell.

		Selenomethionine derivative		
	Native	1	2	1 and 2 combined
Wavelength (Å)	0.873	0.873	0.873	0.873
Temperature (K)	100	100	100	100
Space group	$P4_3$	$P4_3$	$P4_3$	$P4_3$
Unit-cell parameters (Å)				
$a = b$	90.4	90.2	90.4	90.3
c	73.4	73.2	73.2	73.2
Resolution range (Å)	57.0–2.70 (2.85–2.70)	63.0–3.30 (3.48–3.30)	56.7–3.20 (3.37–3.20)	63.9–3.20 (3.37–3.20)
Molecules in asymmetric unit	2	2	2	2
Solvent content (%)	59.0	58.7	58.9	58.8
Matthews coefficient (Å^3 Da^{-1})	3.00	2.98	2.99	2.98
$\Delta\varphi$ (°)	1	3	3	3
No. of collected frames	100	300	300	600
Reflections (observed/unique)	54920/13343	609356/11920	436072/11963	1045428/11920
Completeness (%)	81.0 (62.9)	100 (100)	99.9 (100)	100 (100)
$I/\sigma(I)$	4.1 (2.3)	3.9 (1.6)	4.7 (1.8)	3.8 (1.6)
Mean $I/\sigma(I)$	8.9 (2.0)	14.9 (3.0)	19.2 (6.3)	27.3 (6.7)
Multiplicity	2.4 (1.3)	18.3 (5.4)	20.7 (13.1)	38.9 (18.6)
$R_{\rm merge}$†	0.10 (0.27)	0.17 (0.47)	0.14 (0.41)	0.17 (0.45)
$R_{\rm p.i.m.}$‡	0.07 (0.27)	0.06 (0.35)	0.04 (0.16)	0.04 (0.15)
Anomalous completeness (%)	—	100 (99.7)	99.7 (98.7)	100 (100)
Anomalous multiplicity	—	9.3 (2.7)	10.6 (6.6)	19.9 (9.4)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\langle I(hkl)\rangle$ is the average intensity of multiple observations of symmetry-related reflections. ‡ $R_{\rm p.i.m.} = \sum_{hkl}[1/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\langle I(hkl)\rangle$ is the average intensity of multiple observations of symmetry-related reflections.

2. Materials and methods

2.1. Transmembrane-helix prediction

The described interaction of HAV 2B protein with cellular membranes makes its heterologous expression, purification and crystallization considerably more difficult. In order to localize the region(s) responsible for this interaction, the 2B amino-acid sequence was analyzed using different servers for the prediction of transmembrane helices [*TMHMM* v.2.0 (Krogh *et al.*, 2001), *TMpred* (Hofmann & Stoffel, 1993) and *SOSUI* (Hirokawa *et al.*, 1998)]. These servers predicted the existence of a transmembrane helix approximately spanning residues 1012–1031. This information was taken into account in the experimental design, avoiding the inclusion of this putative transmembrane region in the cloned sequence.

2.2. Cloning

HAV codon usage is not appropriate for the heterologous overexpression of its proteins in *Escherichia coli*. To override this problem, a synthetic gene covering 217 residues (23 kDa) of the HAV polyprotein (residues 765–981 from GenBank accession No. M14707.1; Cohen *et al.*, 1987) and including flanking sequences that provide 5′ *Nde*I and 3′ *Sap*I restriction sites was generated. The coding sequence corresponds to the full-length 2A protein (71 residues) and the N-terminal domain of the 2B protein (146 residues; the soluble portion). To this end, 34 DNA oligonucleotides (of between 32 and 47 nucleotides in length) were designed with the program *DNAWorks* (Hoover & Lubkowski, 2002; Supplementary Table 1[1]) and were used in PCR-based gene assembly using *Pfu* polymerase (Promega) essentially according to Stemmer *et al.* (1995). In brief, the oligonucleotide mixture (2 m*M* each) was used in an assembly PCR reaction (heating at 368 K for 3 min prior to 45 cycles of 1 min at 368 K, 1 min at 330 K and 1 min at 345 K) followed by an amplification reaction (30 cycles of 1 min at 368 K, 1 min at 338 K and 1 min at 345 K) using the two outermost oligonucleotides as primers. The resulting PCR product (approximately 680 bp) was cloned into pGEM-T Easy (Promega), sequenced and subcloned into the *Nde*I and *Sap*I sites of expression vector pTYB1 (New England Biolabs), leading to an ORF coding for HAV 2AB with a C-terminal intein tag (55 kDa) which allows its purification by affinity chromatography (IMPACT system; Chong *et al.*, 1998).

2.3. Expression and purification

E. coli BL21 (DE3) cells were transformed with the expression plasmid and cultures were grown in Luria–Bertani medium supplemented with 100 µg ml^{-1} ampicillin at 303 K until the OD_{600} reached ∼0.3 and then cooled to 285 K. At an OD_{600} of 0.5 the cultures were induced with 0.3 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated for a further 17 h at 285 K, after which the cells were harvested by centrifugation. The pelleted cells were resuspended in cold lysis buffer (20 m*M* Na HEPES pH 8.5, 500 m*M* NaCl, 1 m*M*

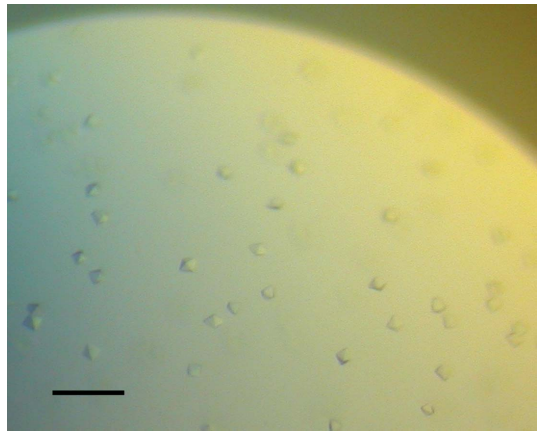


Figure 1
Crystals of HAV 2AB protein. Scale bar: 100 µm.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: PU5340).

EDTA, 1 m*M* PMSF, 20 µg ml^{-1} DNase I) and lysed by sonication on ice. The soluble fraction was then obtained by centrifugation and loaded onto a 5 ml column packed with Chitin Beads (New England Biolabs). After extensive washing with buffer *A* (20 m*M* Na HEPES pH 8.5, 500 m*M* NaCl, 1 m*M* EDTA), the 2AB protein was eluted from the column with buffer *A* supplemented with 50 m*M* DTT. The eluted protein was further purified by size-exclusion chromatography on a HiPrep 16/60 Sephacryl S-100 HR column (GE Healthcare) equilibrated in buffer *A* containing 5 m*M* DTT. Pure 2AB protein was then concentrated to 6.5 mg ml^{-1} using a Vivaspin concentrator (10 000 MWCO PES, Sartorius Stedim Biotech).

To produce selenomethionine-containing 2AB, *E. coli* B834 (DE3) cells were transformed with the expression plasmid and grown overnight in 100 ml minimal medium (SelenoMet, Molecular Dimensions) supplemented with 4 mg l^{-1} L-methionine. Prior to induction, the cells were harvested and washed three times in 100 ml sterile water. The cell pellet was then resuspended in 1 ml sterile water and used to inoculate 2 l prewarmed (285 K) minimal medium containing 1× L-SeMet (prepared according to the manufacturer's instructions from 250× SelenoMethionine Solution, Molecular Dimensions). All other procedures were as described for the nonlabelled protein.

2.4. Crystallization

Preliminary crystallization conditions were explored using the sitting-drop vapour-diffusion method at both 277 and 293 K in 96-well plates (MRC, Swissci AG) using an automated drop dispenser (Phoenix/RE, Rigaku) to mix 150 nl protein solution (6.5 mg ml^{-1}) with an equal volume of reservoir solution. Small crystals of HAV 2AB were observed with reservoir conditions consisting of 0.1 *M* MES pH 6 and 1.26 *M* ammonium sulfate at 293 K.

This condition was refined and crystals were finally grown at 277 K in 24-well VDX plates (Hampton Research) in drops obtained by mixing identical volumes (0.5 or 0.75 µl) protein solution (6.5 mg ml^{-1}) and reservoir solution (0.1 *M* MES pH 5.2–5.8, 1.26 *M* ammonium sulfate, 5 m*M* DTT) and equilibrated against a 750 µl reservoir. Small tetragonal crystals (0.02 mm per edge) were obtained within between one and four weeks (Fig. 1). Selenomethionine-substituted protein crystals were grown using the same conditions as for the native protein.

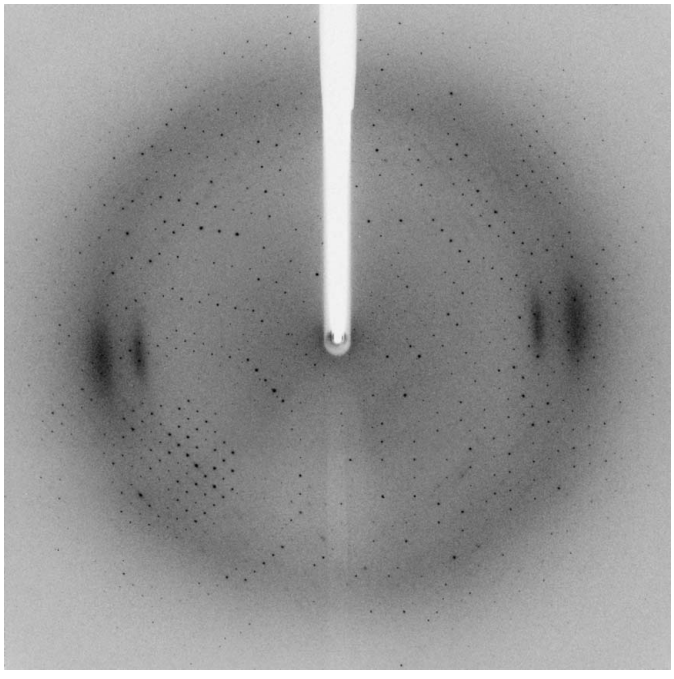


Figure 2
X-ray diffraction pattern of the HAV 2AB crystals. The detector was set to an edge resolution of 2.70 Å (the resolution at the corner was 1.98 Å). Diffraction spots were observed to 2.25 Å resolution.

2.5. X-ray diffraction analysis

Crystals were harvested in CryoLoops (Hampton Research) and soaked for 1 min in the crystallization solution supplemented with 20%(*v*/*v*) glycerol. Mounted crystals were flash-cooled by immersion in liquid nitrogen. X-ray diffraction data sets were collected from both native and selenomethionine-substituted crystals with an ADSC Q315R detector on beamline ID23-EH2 at the European Synchrotron Radiation Facility (ESRF; Grenoble, France) using a 14.20 keV (λ = 0.873 Å) beam focused to 10 µm. A total of 100 frames were recorded at 100 K using an oscillation angle of 1°, 1 s exposure time per frame and a crystal-to-detector distance of 334.07 mm (Fig. 2). Data sets were indexed and integrated using *MOSFLM* (Leslie, 1991) and scaled, merged and reduced with *SCALA* (from *CCP*4; Winn *et al.*, 2011). Although the data sets obtained from selenomethionine-substituted crystals were sufficiently complete, the two best data sets were merged using *CAD* (from *CCP*4; Winn *et al.*, 2011) and scaled together in order to increase the anomalous signal multiplicity at high resolution (Table 1).

3. Results and discussion

The overexpression of HAV 2AB protein for crystallization purposes presented two major complications. Firstly, the HAV 2B protein has been reported to associate with cellular membranes (Gosert *et al.*, 2000; Jecht *et al.*, 1998). Moreover, HAV genes have an unusually high frequency of rare codons, leading to low rates of protein synthesis (Pintó *et al.*, 2007). Sequence analysis allowed the identification of a putative transmembrane helix at the C-terminus of 2B (residues 1012–1031). This region was omitted from the expression construct to avoid the complications associated with the expression and purification of membrane proteins. Furthermore, to overcome the problems derived from HAV codon usage, a synthetic gene was generated. This new sequence codes for the same polypeptide as the region of interest of the HAV genome (residues 765–981 of the P2 polyprotein region) but displays *E. coli* codon usage. Decreasing the *E. coli* growth temperature to 285 K during protein expression was crucial for obtaining soluble recombinant HAV 2AB and ultimately for obtaining crystals (Fig. 1).

Despite the small size of the crystals (∼20 µm per edge), a complete data set was collected to 2.7 Å resolution from a single native crystal using synchrotron radiation on the microfocus beamline ID23-EH2 of the ESRF. The final data-collection and processing statistics are summarized in Table 1. Analysis of systematic absences indicated that the crystal belonged to space group $P4_1$ or $P4_3$. Phase determination (see below) allowed the identification of $P4_3$ as the correct space group. The volume of the crystal asymmetric unit is compatible with the presence of two 2AB molecules, with a Matthews coefficient of 3.0 Å^3 Da^{-1} and a calculated solvent content of 59.0% (Matthews, 1968).

Phase determination by molecular-replacement methods was initially attempted using the available structures of rhinovirus and enterovirus 2A proteins as starting models (PDB entries 2hrv and 1z8r; Petersen *et al.*, 1999; Baxter *et al.*, 2006). This approach was unsuccessful, probably owing to the lack of homology between the

2A region of HAV and the available picornaviral 2A proteinases (Hollinger & Emerson, 2007).

Using the SAD data (Table 1), we have determined the positions of all Se atoms (12, in the asymmetric unit). Model building is ongoing.

References

Baxter, N. J., Roetzer, A., Liebig, H. D., Sedelnikova, S. E., Hounslow, A. M., Skern, T. & Waltho, J. P. (2006). *J. Virol.* **80**, 1451–1462.

Chong, S., Montello, G. E., Zhang, A., Cantor, E. J., Liao, W., Xu, M.-Q. & Benner, J. (1998). *Nucleic Acids Res.* **26**, 5109–5115.

Cohen, J. I., Ticehurst, J. R., Purcell, R. H., Buckler-White, A. & Baroudy, B. M. (1987). *J. Virol.* **61**, 50–59.

Cohen, L., Bénichou, D. & Martin, A. (2002). *J. Virol.* **76**, 7495–7505.

Gosert, R., Egger, D. & Bienz, K. (2000). *Virology*, **266**, 157–169.

Harmon, S. A., Summers, D. F. & Ehrenfeld, E. (1989). *Virus Res.* **12**, 361–369.

Hirokawa, T., Boon-Chieng, S. & Mitaku, S. (1998). *Bioinformatics*, **14**, 378–379.

Hofmann, K. & Stoffel, W. (1993). *Biol. Chem. Hoppe-Seyler*, **374**, 166.

Hollinger, F. B. & Emerson, S. U. (2007). *Fields' Virology*, edited by D. M. Knipe & P. M. Howley, pp. 911–947. Philadelphia: Lippincott, Williams & Wilkins.

Hoover, D. M. & Lubkowski, J. (2002). *Nucleic Acids Res.* **30**, e43.

Jecht, M., Probst, C. & Gauss-Müller, V. (1998). *Virology*, **252**, 218–227.

Jong, A. S. de, de Mattia, F., Van Dommelen, M. M., Lanke, K., Melchers, W. J., Willems, P. H. & van Kuppeveld, F. J. (2008). *J. Virol.* **82**, 3782–3790.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. (2001). *J. Mol. Biol.* **305**, 567–580.

Kuppeveld, F. J. van, Galama, J. M., Zoll, J., van den Hurk, P. J. & Melchers, W. J. (1996). *J. Virol.* **70**, 3876–3886.

Leslie, A. G. W. (1991). *Crystallographic Computing 5: From Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 27–38. Oxford University Press.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 499–501.

Morace, G., Kusov, Y., Dzagurov, G., Beneduce, F. & Gauss-Muller, V. (2008). *BMB Rep.* **41**, 678–683.

Petersen, J. F., Cherney, M. M., Liebig, H. D., Skern, T., Kuechler, E. & James, M. N. (1999). *EMBO J.* **18**, 5463–5475.

Pintó, R. M., Aragonès, L., Costafreda, M. I., Ribes, E. & Bosch, A. (2007). *Virus Res.* **127**, 158–163.

Schultheiss, T., Kusov, Y. Y. & Gauss-Müller, V. (1994). *Virology*, **198**, 275–281.

Stemmer, W. P., Crameri, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. (1995). *Gene*, **164**, 49–53.

Weitz, M., Baroudy, B. M., Maloy, W. L., Ticehurst, J. R. & Purcell, R. H. (1986). *J. Virol.* **60**, 124–130.

Wheeler, C. M., Fields, H. A., Schable, C. A., Meinke, W. J. & Maynard, J. E. (1986). *J. Clin. Microbiol.* **23**, 434–440.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.