Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary crystallographic studies on the large extracellular domain of human CD81, a tetraspanin receptor for hepatitis C virus

Kengo Kitadokoro,[a,b] Giuliano Galli,[c] Roberto Petracca,[c] Fabiana Falugi,[c] Guido Grandi[c] and Martino Bolognesi[a]*

[a]Department of Physics – INFM and Advanced Biotechnology Center, University of Genoa, Via Dodecaneso 33, I-16146 Genova, Italy, [b]Research Center for Instrumental Analysis, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan, and [c]Chiron Vaccines Research Center, Via Fiorentina 1, 53100 Siena, Italy

Correspondence e-mail: bolognes@fisica.unige.it

Received 21 September 2000
Accepted 26 October 2000

The large extracellular domain of CD81, a member of the tetraspanin family and a receptor protein for hepatitis C virus envelope E2 glycoprotein, has been expressed, purified and subsequently crystallized using the sitting-drop vapour-diffusion technique. Native diffraction data to 1.6 Å resolution were obtained at the ID14 beamline of the European Synchrotron Radiation Facility from a flash-frozen crystal at 100 K. The crystals belong to space group $P2_1$, with unit-cell parameters $a = 31.5$, $b = 77.2$, $c = 38.5$ Å, $\beta = 107.4^\circ$, and are likely to contain two extracellular domains (2×99 residues) per asymmetric unit.

1. Introduction

Tetraspanins form a recently discovered and growing protein homology family characterized by two main structural features (Levy *et al.*, 1998; Maecker *et al.*, 1997). Firstly, owing to the presence of four hydrophobic transmembrane regions (TM1–TM4) and to their topological organization, they protrude out of the cell with two domains, a 'large extracellular loop' (LEL) and a 'small extracellular loop' (SEL). Secondly, four Cys residues (expected to be arranged into two disulfide bridges) are invariantly present in the tetraspanin LEL. Two of the four cysteines are included in a tetraspanin-conserved Cys-Cys-Gly motif (Levy *et al*., 1998). Tetraspanins are found in a large variety of organisms, being localized in man in different organs and tissues. CD9, CD63, CD81 and CD82 are nearly ubiquitous members of the homology family, whereas other tetraspanins seem to be restricted to specific tissues such as the lymphoid cells (CD53) or the mature B cells (CD37). Remarkably, many tetraspanins are being discovered to be tumour-associated antigens and appear to be involved in tumor growth and proliferation, in line with an emerging body of evidence indicating their involvement in the regulation of cell adhesion and migration processes (Radford *et al.*, 1996; Levy *et al*., 1998). Moreover, tetraspanins are known to interact with surface-associated molecular complexes such as proteins of the β1 integrin family (Berditchevski *et al.*, 1996; Mannion *et al.*, 1996), co-receptors (CD4, CD8 and CD2; Cresswell, 1994) and other tetraspanins (Bradbury *et al*., 1991), promoting their functionality. Despite the recent growth of basic functional data, very little is known about the tetraspanin fold and about structure–function relationships within the tetraspanin family.

HCV is a positive-strand RNA virus of the flaviviridae family (Houghton, 1996) chronically infecting about 170 million persons worldwide. Chronic HCV infection results in severe liver diseases in a sizable fraction of cases (Hoofnagle, 1997). It has recently been shown that the HCV envelope E2 glycoprotein binds with nanomolar affinity to the LEL domain (composed of ~100 amino acids) of the human hepatocyte tetraspanin CD81, a necessary step for infection and virulence (Pileri *et al.*, 1998). It has also been shown that structural integrity of the LEL domain, related to maintenance of the two disulfide bridges in the oxidized state, is a requirement for CD81/E2 binding (Petracca *et al.*, 2000). Moreover, at least two residues mapping in the CD81-LEL domain are capable of affecting HCV binding to hepatocytes. In particular, the Phe196Ala mutation, naturally occurring in the African green monkey, completely abolishes virus binding and therefore infectivity (Higginbottom *et al.*, 2000; Meola *et al.*, 2000; Petracca *et al.*, 2000).

In view of a crystallographic investigation shedding first light on the so far undiscovered structural motifs of the tetraspanin family and on the molecular bases of HCV/CD81 interaction, we present here the crystallization and preliminary X-ray diffraction analysis of the CD81-LEL domain, the receptor region known to bind specifically and with high affinity the viral envelope E2 glycoprotein.

2. Methods and results

2.1. Cloning

His-tagged CD81-LEL was cloned and expressed in *Escherichia coli* as a fusion protein with the IgG binding domain of the *Staphylococcus aureus* Protein A (Nilsson *et*

Table 1
Data-collection statistics for CD81-LEL native data sets.

Values in parentheses refer to the outer resolution shell.

	Native 1	Native 2
Wavelength (Å)	1.54	0.93
Space group	$P2_1$	$P2_1$
Unit-cell parameters		
a (Å)	31.5	31.5
b (Å)	77.2	77.2
c (Å)	38.4	38.5
β (°)	107.3	107.4
Resolution (Å)	2.4	1.6
Total reflections	35179	36910
Unique reflections	6737	21557
$R_{\rm merge}$ (%)	7.1 (26.7)	3.8 (31.4)
Completeness (%)	97.0 (90.2)	98.0 (93.1)

† $R_{\rm merge} = 100(|I_h - \langle I\rangle_h|)/I_h$, where $\langle I\rangle_h$ is the mean intensity of all symmetry-related reflections I_h.

al., 1987). The fusion protein gene was designed by assembling the DNA fragment coding for a tandem duplication of the Z domain of protein A, a hinge fragment encoding the thrombin cleavage site Leu-Val-Pro-Arg-Gly-Ser, the sequence encoding the LEL domain (residues 112–202 of the CD81 sequence) and a six-histidine tail coding sequence. Briefly, the LEL coding sequence was PCR amplified using the plasmid pCDM8-CD81 as template (Pileri *et al.*, 1998) and the primers 5′-CCTCGAGCTCTCTGGTTCCGCGTGGATCCGGCTTTGTCAACAAGGACC-3′ and 5′-CCCCAAGCTTTCAATGATGATGATGATGATGCAGCTTCCCGGAGAAG-3′. The PCR product was digested with *Sac*I and *Hin*dIII and ligated to the plasmid pEZZ18 (Pharmacia Biotech) digested with the same restriction enzymes. After transformation into *E. coli* DB1035 competent cells, a selected clone was used to produce the recombinant protein.

2.2. Purification of LEL-His

The cell pellet obtained from 2 l of flask culture was subjected to osmotic shock. Briefly, the cells were resuspended in 150 ml of 20 m*M* Tris–HCl, 2.4 m*M* EDTA, 20% sucrose pH 8, incubated for 10 min at 277 K and the suspension clarified by centrifugation for 10 min at 4500*g*. The pellet was then gently resuspended in 150 ml of 20 m*M* Tris–HCl, 2.4 m*M* EDTA pH 8, incubated for 10 min at 277 K and centrifuged for 20 min at 4500*g*. After loading the supernatant onto a 1 × 8 cm IgG Sepharose Fast Flow column (Pharmacia Biotech) equilibrated with TST buffer (50 m*M* Tris–HCl, 150 m*M* NaCl, 0.05% Tween 20 pH 7.6), the column was washed with 40 ml TST buffer followed by 40 ml of 5 m*M* ammonium acetate pH 5.0. The fusion protein was eluted from the column with 0.5 *M* acetic acid pH 3.4 using a FPLC apparatus (Pharmacia) and collecting 2 ml fractions. The peak fractions were pooled and the pH was brought to 7.4 using 1 *M* Tris solution. The purified fusion protein (0.5 mg ml^{-1}) was digested overnight with thrombin (Amersham) at 298 K, using 25 µg of fusion protein per unit of enzyme. The reaction was inactivated with 0.05 m*M* PMSF (Sigma). The reaction mixture was passed through an IgG Sepharose FF column and the non-retained material was further purified on a nickel-activated chelating Sepharose FF column (Pharmacia Biotech) equilibrated in TST buffer. The column was washed with 30 m*M* sodium phosphate pH 7.8 and the protein was eluted using a 0–500 m*M* imidazole gradient pH 7.8. After SDS–PAGE analysis, the fractions (2 ml) containing the purified protein were pooled, dialysed against PBS and stored at 253 K. The correctness of the thrombin cleavage reaction was verified by amino-terminal sequencing using a LF3000 Protein Sequencer apparatus (Beckman).

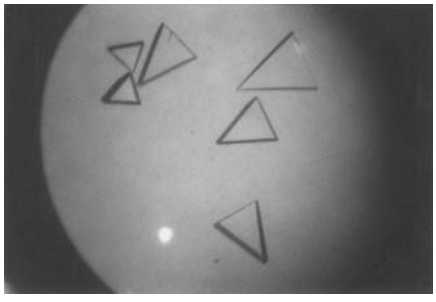

Figure 1
Results of a typical CD81-LEL crystal growth droplet. The longest crystal displayed is approximately 0.3 × 0.3 × 0.1 mm.

2.3. Crystallization

For crystallization experiments, a 10 mg ml^{-1} solution of the recombinant CD81-LEL in 0.01 *M* Tris–HCl pH 7.5 was employed. Crystallization trials were set up at room temperature as sitting-drop vapour-diffusion experiments on Linbro crystallization plates. Initial screening was performed using the sparse-matrix method (Jancarik & Kim, 1991) with commercial crystal screening kits (Hampton Research). Crystals of the recombinant CD81-LEL proteins were readily grown from polyethylene glycol solutions, as well as under different physicochemical conditions (ammonium phosphate, sodium acetate trihydrate, potassium/sodium tartrate), yielding isomorphous crystal forms. The best crystals were obtained through equilibration against a solution containing 15%(*w*/*v*) polyethylene glycol 4000, 0.1 *M* sodium chloride and 0.1 *M* MES buffer pH 6.0 at 293 K in sitting-drop vapour-diffusion setups. The crystallization droplets consisted of 2 µl protein and 2 µl reservoir solutions, with 500 µl of reservoir solution; triangle-shaped tabular crystals appeared within a few days and grew to maximum dimensions of about 0.4 × 0.4 × 0.15 mm (Fig. 1).

2.4. Data collection and processing

X-ray diffraction data from CD81-LEL crystals were collected at 100 K in a nitrogen stream, supplementing the mother liquor solution with 20% glycerol as a cryoprotectant. Initially, diffraction data from the native crystals were collected to 2.4 Å resolution with a MAR Research 345 imaging-plate detector mounted on a Rigaku RU-H3R rotating-anode generator, using monochromated Cu $K\alpha$ radiation. Subsequently, a high-resolution native data set at 1.60 Å was collected from a single crystal on beamline ID14 at the European Synchrotron Radiation Facility, Grenoble, France (Table 1). The X-ray wavelength was 0.93 Å, the angle oscillation range was 1.0° and the crystal-to-detector distance was 110 mm. Analysis of the diffraction pattern and of the systematic absences allowed the assignment of the CD81-LEL crystals to the primitive monoclinic space group $P2_1$, with unit-cell parameters a = 31.5, b = 77.2, c = 38.5 Å, β = 107.4° (see Table 1). The diffracted intensities were processed using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Assuming a MW of 10.8 kDa for the expressed CD81-LEL domain, packing-density calculations indicate the most probable value for $V_{\rm M}$ to be 2.07 Å^3 Da^{-1}, with two CD81-LEL chains per asymmetric unit. This corresponds to a solvent fraction of about 41%, a typical value for protein crystals (Matthews, 1968). Amino-acid sequence searches do not highlight any structural

homology with proteins of known three-dimensional structure. Therefore, the crystallographic analysis of CD81-LEL will be based on the multiple isomorphous replacement method; a search for heavy-atom derivatives is presently in progress.

We are grateful to ESRF for granting access to the ID14 beamline. Part of this work was supported by grants from the National Institute for the Physics of Matter and the Italian Space Agency grant ARS-99-22 (to MB). KK is a member of the Sakabe TARA project of the University of Tsukuba.

References

Berditchevski, F., Zutter, M. & Hemler, M. E. (1996). *Mol. Biol. Cell.* **7**, 193–207.

Bradbury, L. E., Kansas, G. S., Levy, S., Evans, R. L. & Tedder, T. F. (1991). *Immunology*, **149**, 2841- 2850.

Cresswell, P. (1994). *Annu. Rev. Immunol.* **12**, 259–293.

Higginbottom, A., Quinn, E. R., Kuo, C. C., Flint, M., Wilson, L. H., Bianchi, E., Nicosia, A., Monk, P. N., McKeating, J. A. & Levy, S. (2000). *J. Virol.* **74**, 3642–3649.

Hoofnagle, J. H. (1997). *Hepatology*, **26**(Suppl. 1), 15S.

Houghton, M. (1996). *Virology*, edited by B. N. Fields, D. M. Knipe & P. M. Howley, pp. 1035–1058. Philadelphia: Lippincott-Raven.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Levy, S., Todd, S. C. & Maecker, H. T. (1998). *Annu. Rev. Immunol.* **16**, 89–109.

Maecker, H. T., Todd, S. C. & Levy, S. (1997). *FASEB J.* **11**, 428–442.

Mannion, B. A., Berditchevski, F., Kraeft, S. K., Chen, L. B. & Hemler, M. E. (1996). *J. Immunol.* **157**, 2039–2047.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Meola, A., Sbardellati, A., Bruni Ercole, B., Cerretani, M., Pezzanera, M., Ceccacci, A., Vitelli, A., Levy, S., Nicosia, A., Travoni, C., McKeating, J. & Scarselli, E. (2000). *J. Virol.* **74**, 5933–5938.

Nilsson, B., Moks, T., Jansson, B., Abrahmsen, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T. A. & Uhlen, M. (1987). *Protein Eng.* **1**, 107–113.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Petracca, R., Falugi, F., Galli, G., Norais, N., Rosa, D., Campagnoli, S., Bugio, V., Di Stasio, E., Giardina, B., Houghton, M., Abrignani, S. & Grandi, G. (2000). *J. Virol.* **74**, 4824–4830.

Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A. J., Houghton, M., Rosa, D., Grandi, G. & Abrignani, S. (1998). *Science*, **282**, 938–941.

Radford, K. J., Thorne, R. F. & Hersey, P. (1996). *Biochem. Biophys. Res. Commun.* **222**, 3–18.