



Crystallization of the human tetraspanin protein CD9

Rie Umeda, Tomohiro Nishizawa* and Osamu Nureki*

Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. *Correspondence e-mail: t-2438@bs.s.u-tokyo.ac.jp, nureki@bs.s.u-tokyo.ac.jp

Received 17 September 2018

Accepted 28 December 2018

Edited by A. Nakagawa, Osaka University, Japan

Keywords: crystallization; LCP; membrane proteins; CD9; tetraspanins.

The tetraspanin family of proteins with four membrane-spanning proteins function in a wide range of physiological processes in higher organisms, including cell migration and proliferation, cell fusion, fertilization and virus infection. Although the recently reported structure of CD81 unveiled the basic architecture of this family for the first time, further structural and functional studies are required in order to understand the mechanistic details of the complicated functions of the tetraspanin-family proteins. In this study, attempts were made to crystallize human CD9, a representative member of the tetraspanin family, and it was demonstrated that the truncation of a variable region in the second long extracellular loop significantly improved crystal growth.

1. Introduction

The tetraspanins are a family of four-transmembrane-helix proteins that are widely conserved in eukaryotes, with 33 identified human members (Charrin *et al.*, 2009). The tetraspanin proteins share a common architecture consisting of four transmembrane helices (TM1–TM4) and two extracellular loops, with a short extracellular loop (SEL) between TM1 and TM2, and a large extracellular loop (LEL) between TM3 and TM4 (Zimmerman *et al.*, 2016; Charrin *et al.*, 2014). Numerous studies have suggested that tetraspanin proteins interact with other functional proteins, such as adhesion proteins, cell growth factor receptors and intracellular scaffold proteins, and organize a ‘tetraspanin-enriched microdomain’ (TEM) where tetraspanins and their ‘partner proteins’ form signaling platforms in cells (Hemler, 2005; Yáñez-Mó *et al.*, 2009). Owing to their diverse ‘partner proteins’, tetraspanins are involved in a wide range of cell functions (Hemler, 2003; Levy & Shoham, 2005). Notably, since tetraspanins play important roles in cell proliferation and Hepatitis C virus (HCV) infection (Pileri *et al.*, 1998), they are considered to be drug targets for the treatment of cancer or for protection from HCV infection (Hemler, 2014). The crystal structure of the tetraspanin protein CD81 has recently been reported (Zimmerman *et al.*, 2016), which revealed the basic architecture of this protein family, but further structural and functional studies are required in order to elucidate how tetraspanin proteins function in cells.

CD9 is the best-characterized member of the tetraspanins and is expressed in various tissues and cells (Reyes *et al.*, 2018; Jankovičová *et al.*, 2015). In particular, CD9 plays an essential role in fertilization. CD9-knockout female mice exhibit an infertility phenotype caused by the failure of sperm–egg fusion (Miyado *et al.*, 2000; Le Naour *et al.*, 2000; Kaji *et al.*, 2000). Despite the physiological importance of CD9, the detailed



OPEN ACCESS

Table 1
Macromolecule-production information.

Source organism	<i>H. sapiens</i>
DNA source	cDNA Library, Human Mammary Gland (Takara)
Expression vector	Modified pFastBac1
Expression host	Sf9 insect cells
Complete amino-acid sequence of the construct produced†	MPVKGGTKCIKYLFGFNFI (C) FWLAGIA VLAIGLWLRFDSTKSIQETNNNSS FYTGVIILIGAGALMMLVGFGLGCCAVQ ESQCMLGLFFGFLLVIFAIEIAAAIWGY SHKDEVIKEVQEFYKDTYNKLTKEDEPQ RETLKAIHYALNCCGLAGGVEQFISDIC PKKDVLTEFTVKSCPDAIKEVFDNKFHI IGAVGIGIAVVMIFGMIFSMILCCAIRR NREMV

† The truncated residues are underlined. To prepare mercury-derivatized crystals, we introduced a cysteine residue at Ile20, which is shown in parentheses.

molecular mechanism by which CD9 regulates cell fusion and other signal transduction pathways has remained unclear. To understand this mechanism, the structure of CD9 and structure-based functional analyses have long been awaited. In this study, we designed various constructs of CD9 by modifying the flexible loop region of the LEL, and revealed that truncation of the LEL facilitates the crystallization of CD9 in lipidic cubic phase. This method could be applied to other tetraspanin proteins to facilitate structural studies of this unique protein family.

2. Materials and methods

2.1. Macromolecule production

The *Homo sapiens* CD9 gene (UniProt accession P21926) was cloned into a modified pFastBac1 expression vector (Invitrogen), which includes an N-terminal His₈ tag, a GFP tag and a Tobacco etch virus (TEV) protease cleavage site. The DNA fragment encoding CD9 was PCR-amplified using PrimeSTAR Max DNA Polymerase (Takara). The PCR product was inserted into the KpnI and EcoRI sites of the vector. The five extracellular loop residues and the three C-terminal residues of CD9 were deleted by a PCR-based method. For mercury derivatization, a cysteine mutation was introduced by site-directed mutagenesis (Table 1).

The modified CD9 construct was expressed in Sf9 cells using the Bac-to-Bac baculovirus expression system (Invitrogen). The CD9 expression plasmid was transformed into DH10Bac competent *Escherichia coli* cells (Invitrogen) to generate a recombinant bacmid. The isolated bacmid DNA was transfected into 50 ml of Sf9 cells at a density of approximately 3×10^6 cells ml⁻¹ using FuGENE HD (Promega) and the cells were incubated at 27°C for four days. After incubation, the cells and large debris were removed by centrifugation (2600g, 5 min) and the clarified supernatant was used for the subsequent protein expression.

For protein expression, Sf9 cells cultured in Sf-900 II medium (Invitrogen) were infected at a density of approximately 3×10^6 cells ml⁻¹ and incubated at 27°C for 48 h. The cells were collected by centrifugation (5000g, 12 min) and the

following procedures were performed at 4°C or on ice. The cells were resuspended in 50 mM HEPES pH 7.0, 150 mM NaCl with protease inhibitors (1.7 µg ml⁻¹ aprotinin, 0.6 µg ml⁻¹ leupeptin, 0.5 µg ml⁻¹ pepstatin and 1 mM phenylmethylsulfonyl fluoride). The cells were disrupted by sonication and the cell debris was removed by centrifugation (4000g, 10 min). The supernatant was ultracentrifuged (186 000g, 1 h) and the membrane fraction was collected and resuspended in 50 mM HEPES pH 7.0 containing 150 mM NaCl.

The membrane fraction was solubilized in 10 mM HEPES pH 7.0, 150 mM NaCl, 1.5% (w/v) *n*-dodecyl-β-D-maltoside (DDM), 0.3% (w/v) cholesteryl hemisuccinate (CHS) and the solubilized proteins were purified using the three following chromatographic steps. The insoluble material was removed by ultracentrifugation (186 000g, 30 min). The supernatant was mixed with TALON metal-affinity resin (Clontech) and incubated for 30 min. After incubation, the resin was washed with seven column volumes of 10 mM HEPES pH 7.0, 150 mM NaCl, 0.1% DDM, 0.02% CHS, 20 mM imidazole pH 7.0. The protein sample was then eluted with three column volumes of 10 mM HEPES pH 7.0, 150 mM NaCl, 0.1% DDM, 0.02% CHS, 300 mM imidazole pH 7.0. The eluted sample was mixed with His₈-tagged TEV protease (purified in-house) to cleave the His₈-GFP tag and was dialyzed against 10 mM HEPES pH 7.0, 150 mM NaCl, 0.1% DDM, 0.02% CHS to remove the imidazole. After overnight dialysis, the sample was mixed with 5 ml Ni-NTA Superflow resin (Qiagen) and incubated for 10 min at 4°C to remove the His₈-GFP tag and TEV protease. The collected flowthrough fraction was then concentrated using an Amicon Ultra filter (30 kDa molecular-mass cutoff, Millipore) and further purified by gel filtration (Superdex 200 Increase 10/300 GL, GE Healthcare) in 10 mM HEPES pH 7.0, 150 mM NaCl, 0.05% DDM, 0.01% CHS. The peak fractions were concentrated to approximately 15 mg ml⁻¹ using an Amicon Ultra filter (molecular-mass cutoff 50 kDa, Millipore). The truncated and cysteine mutants were purified using the same procedure as described above.

2.2. Crystallization

The purified CD9 samples were reconstituted into lipidic cubic phase (LCP) by mixing them with liquefied monoolein (Sigma) in a 2:3 (w:v) protein:lipid ratio using the twin-syringe mixing method (Caffrey & Cherezov, 2009). For sandwich-drop crystallization, aliquots of the protein-LCP mixture (50 nl) were dispensed onto 96-well glass plates and overlaid with the precipitant solution (800 nl) using a Griffin LCP robot (Art Robbins Instruments). Initial crystallization conditions were searched for using screening kits including MemMeso (Molecular Dimensions). The initial hits were optimized by changing the concentration of each component (Table 2).

The crystals were harvested using MicroMounts (MiTeGen) or mesh grid loops (MiTeGen) and were flash-cooled in liquid nitrogen. To prepare mercury-derivative crystals, we co-crystallized CD9 with methylmercury chloride.

Table 2
Crystallization.

	Native CD9 _{cryst}	CD9 ^{120C} _{cryst}
Method	Lipidic cubic phase	Lipidic cubic phase
Plate type	96-well plastic sandwich plate	96-well plastic sandwich plate
Temperature (K)	293	293
Protein concentration (mg ml ⁻¹)	15	15
Composition of reservoir solution	10–50 mM MOPS pH 6.5, 32–38% PEG 200 or 10–50 mM Tris pH 7.5, 32–38% PEG 200	10–50 mM MOPS pH 6.5, 36–38% PEG 200 or 10–100 mM Tris pH 7.5, 32–38% PEG 200
Co-crystallized compound (final concentration in mixture with protein solution)	—	2 mM methylmercury chloride
Volume of drop (nl)	50	50
Volume of reservoir (nl)	800	800

Table 3
Data collection and processing.

Values in parentheses are for the outer shell.

	Native CD9 _{cryst}	Mercury-derivatized CD9 ^{120C} _{cryst}
Diffraction source	BL32XU, SPring-8	BL32XU, SPring-8
Wavelength (Å)	1.0	1.0
Temperature (K)	100	100
Detector	EIGER X 9M	EIGER X 9M
Crystal-to-detector distance (mm)	250	300
Rotation range per image (°)	0.1	0.1
Exposure time per image (s)	0.1	0.1
No. of data sets for merging	25	47
Oscillation range per crystal (°)	10–180	10–180
Space group	C222 ₁	C222 ₁
<i>a</i> , <i>b</i> , <i>c</i> (Å)	45.18, 124.83, 129.23	45.18, 125.21, 129.40
α , β , γ (°)	90, 90, 90	90, 90, 90
Resolution range (Å)	50–2.70 (2.86–2.70)	50–3.17 (3.43–3.17)
Total No. of reflections	604721	397058
No. of unique reflections	10454	17477
Completeness (%)	100.0 (100.0)	100.0 (100.0)
Multiplicity	57.85	22.72
CC _{1/2}	0.998 (0.515)	0.994 (0.740)
$\langle I/\sigma(I) \rangle$	15.79 (1.70)	8.05 (1.70)
<i>R</i> _{meas}	0.526 (3.167)	0.703 (2.651)
Overall <i>B</i> factor from Wilson plot (Å ²)	48.84	51.34

Methylmercury chloride was dissolved in DMSO and added to the protein samples at a final concentration of 2 mM. After incubation for 20 min on ice, the samples were crystallized by the LCP method as described above.

2.3. Data collection and processing

All diffraction data sets were collected using the micro-focused X-ray beam at BL32XU at SPring-8 (Hirata *et al.*, 2013). The microcrystals in the loop were identified by a raster scan and analysis by *SHIKA* (Ueno *et al.*, 2016). Small-wedge data sets, each consisting of 5–30°, were collected from single crystals. The collected data sets were automatically processed with *KAMO* (Yamashita *et al.*, 2018). All data sets were collected at a wavelength of 1.000 Å. Anomalous diffraction data sets were collected from the mercury-derivative crystals at a wavelength of 1.000 Å. Each data set was indexed and integrated using *XDS* (Kabsch, 2010), followed by a hierarchical clustering analysis using the correlation coefficients of the normalized structure-factor amplitudes between data sets.

Finally, a group of outlier-rejected data sets were scaled and merged using *XSCALE* (Kabsch, 2010). The data-processing statistics are summarized in Table 3. One Hg atom site was identified with *SHELXD* (Sheldrick, 2015). The initial phases were calculated with *AutoSHARP* (Vonrhein *et al.*, 2007).

3. Results and discussion

We selected human CD9 as a target for crystallographic analysis and purified the protein by metal-affinity chromatography and size-exclusion chromatography (Fig. 1). We first crystallized wild-type CD9 (full length, residues 1–228) in lipidic cubic phase (Caffrey & Cherezov, 2009), and the initial crystals were obtained in a reservoir solution consisting of 36–42% PEG 200, 10–50 mM Tris–HCl pH 7.5 or a similar solution containing 10–50 mM MOPS pH 6.6 instead of Tris–HCl pH 7.5. Despite repetitive optimizations of the crystallization conditions, these crystals only grew to approximate dimensions of 10 × 10 × 5 μm and diffracted X-rays to a maximum of only 10 Å resolution.

The sequence of the LEL is rather variable among the tetraspanin subtypes, and thus is considered to mediate their

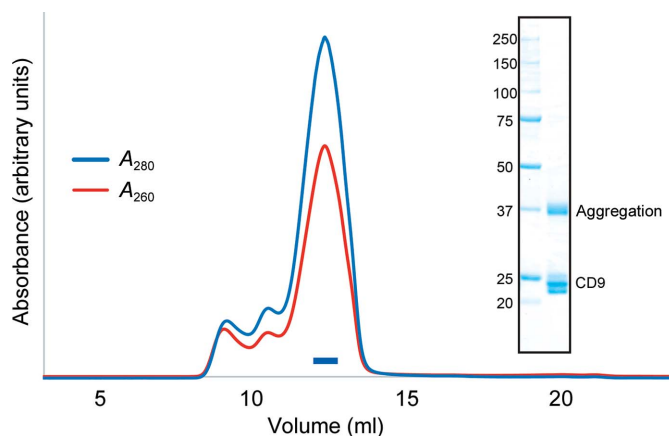


Figure 1
Protein preparation. Size-exclusion chromatogram of wild-type CD9. The blue and red lines indicate the absorbance at 280 and 260 nm, respectively. The blue bar indicates the fractions that were collected and used for crystallization. The inset shows an SDS–PAGE analysis with Coomassie Brilliant Blue staining. Left lane, molecular-mass markers (labeled in kDa); right lane, wild-type CD9. The multiple bands at around 23 kDa represents heterogeneous palmitoylation of the purified CD9 protein, and a higher molecular-weight band at 37 kDa is owing to molecular aggregation during SDS denaturation.

interactions with their partner proteins. Crystal structures of the LEL from CD81, which is closely related to CD9, have been reported (PDB entries 1g8q and liv5) and revealed that the LEL contains five short helical segments stabilized by two conserved disulfide bonds (Kitadokoro *et al.*, 2001, 2002). Among these structures, the third and fourth helical segments between the second and third cysteine residues adopt rather varied conformations (Fig. 2), indicating their intrinsic flexibility. Analysis of a sequence alignment between CD9 and CD81 suggested that the LEL of CD9 probably adopts a similar architecture to that of CD81, including this flexible region. Considering the possibility that this flexibility hinders the tight packing interactions within the crystals, we produced truncated mutants of the corresponding LEL region (Leu155–Glu160 or Lys170–Ser180) and performed crystallization trials. Among the tested constructs, only the construct that lacked Thr175–Lys179 ($\Delta 175$ –179; Table 1) yielded crystals in LCP, and they diffracted X-rays to 3.5 Å resolution.

However, owing to the fragility and low reproducibility of these crystals, we could not collect a complete data set, even though we merged data sets from multiple crystals. To improve the reproducibility of the crystals, we constructed a C-terminally truncated mutant with $\Delta 175$ –179 and $\Delta 226$ –228 deletions, which we hereafter refer to as CD9_{cryst}. Crystals of CD9_{cryst} were obtained in a reservoir solution consisting of 32–38% PEG 200, 10–50 mM MOPS pH 6.5 or a similar solution containing 10–50 mM Tris–HCl pH 7.5 instead of MOPS. The crystals grew to maximum dimensions of $50 \times 20 \times 10 \mu\text{m}$, diffracted X-rays to 2.7 Å resolution (Fig. 3) and belonged to space group $C222_1$, with unit-cell parameters $a = 45.18$, $b = 124.83$, $c = 129.23$ Å. By merging data sets from multiple crystals, we finally collected a complete data set from the CD9_{cryst} crystals.

We next attempted to prepare mercury-derivatized crystals for experimental phase determination. CD9 contains ten cysteine residues, all of which are predicted to undergo post-translational modifications: four in the large extracellular loop

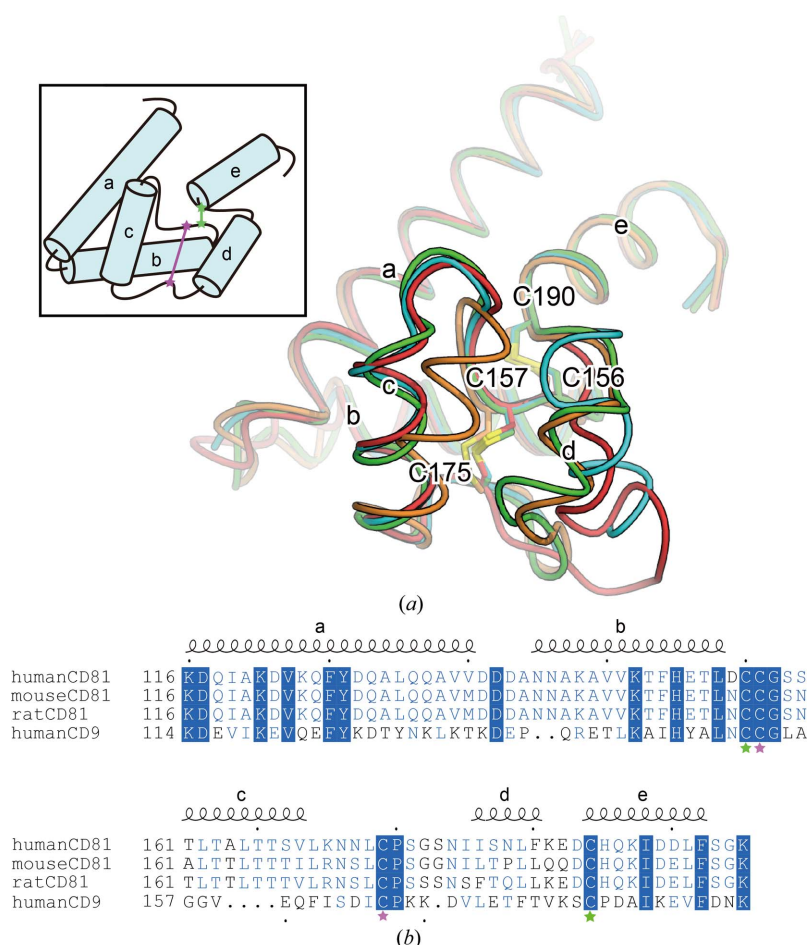


Figure 2

Flexibility of the large extracellular loop. (a) Superimposition of the large extracellular loops of the human CD81 structures [PDB entries 1g8q (green and cyan) and liv5 (orange and red)], viewed from the extracellular side. The inset shows a schematic diagram, with green and pink stars representing cysteine residues forming intramolecular disulfide bonds. (b) Amino-acid sequence alignment of the large extracellular loops of human CD81, mouse CD81, rat CD81 and human CD9. Fully and partially conserved residues are highlighted by blue panels and blue letters, respectively. The secondary structure of CD81 is indicated above the alignment. Residues are numbered according to human CD81. Green and pink stars indicate the cysteine residues shown in the diagram in (a).

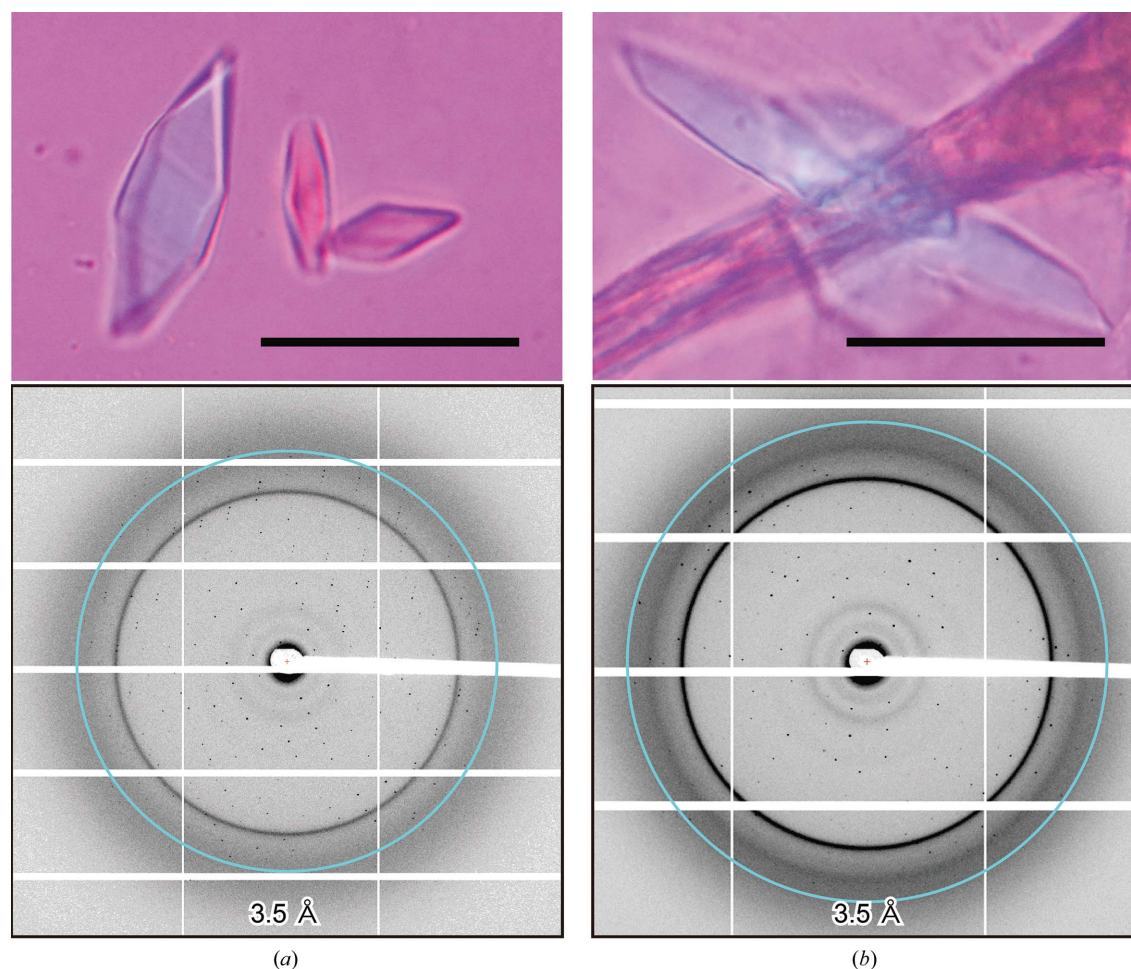


Figure 3
 CD9 crystals and X-ray diffraction. Crystals and X-ray diffraction patterns of native CD9_{cryst} (a) and mercury-derivatized crystals of CD9^{120C}_{cryst} (b). Scale bars represent 30 μm . The rings indicate 3.5 \AA resolution.

form disulfide bonds and the remaining six at the intracellular ends of the transmembrane helices are heterogeneously palmitoylated (Charrin *et al.*, 2002). Therefore, to uniformly label the protein we introduced an additional cysteine residue at Ile20, which is predicted to be located at the intracellular end of TM1. This construct, termed CD9^{120C}_{cryst}, was co-crystallized with methylmercury chloride, and after optimization crystals of CD9^{120C}_{cryst} grew to maximum dimensions of $50 \times 20 \times 10 \mu\text{m}$ under reservoir conditions similar to those used for the native protein. The crystals belonged to the same space group as the native crystals, with similar unit-cell parameters ($C222_1$, $a = 45.18$, $b = 125.21$, $c = 129.40 \text{ \AA}$; Table 2), and diffracted X-rays to 3.2 \AA resolution (Fig. 3).

The data-processing statistics are summarized in Table 1. One Hg-atom site was identified with *SHELXD*, and the initial phases were calculated with *SHARP* by the single isomorphous replacement with anomalous scattering (SIRAS) method, which clearly visualized four transmembrane helices and the extracellular loops. Overall, the current results demonstrate a large improvement in the crystal quality on the modification of the LEL in CD9. While the architecture of the four transmembrane domains is conserved among tetraspanin-family members, the LEL sequences have diverged among the

subtypes, suggesting their high flexibility. Notably, the crystallized construct (CD9_{cryst}) could rescue the sperm-fusing ability of the eggs from CD9-knockout mice, suggesting that it still retains the critical function of CD9 (manuscript in preparation). Therefore, modification (truncation and/or mutation) of the LEL could be exploited for the crystallization of other tetraspanin proteins, and thus will promote structural studies of the tetraspanin-family proteins.

Acknowledgements

We thank Dr Reiya Taniguchi for assistance in manuscript preparation and fruitful discussions, and the beamline staff at BL32XU of SPring-8, Hyogo, Japan. The diffraction experiments were performed on BL23XU at SPring-8 (proposals 2015A1057, 2015B2057, 2016B2717 and 2017A2717) with the approval of RIKEN.

Funding information

This work was supported by a MEXT Grant-in-Aid for Specially Promoted Research (grant 16H06294) to ON. This work was also supported by JSPS KAKENHI (grant

17H05000 to TN). TN was funded by JST, PRESTO (grant JPMJPR14L8).

References

- Caffrey, M. & Cherezov, V. (2009). *Nature Protoc.* **4**, 706–731.
- Charrin, S., Jouannet, S., Boucheix, C. & Rubinstein, E. (2014). *J. Cell Sci.* **127**, 3641–3648.
- Charrin, S., le Naour, F., Silvie, O., Milhiet, P.-E., Boucheix, C. & Rubinstein, E. (2009). *Biochem. J.* **420**, 133–154.
- Charrin, S., Manié, S., Oualid, M., Billard, M., Boucheix, C. & Rubinstein, E. (2002). *FEBS Lett.* **516**, 139–144.
- Hemler, M. E. (2003). *Annu. Rev. Cell Dev. Biol.* **19**, 397–422.
- Hemler, M. E. (2005). *Nature Rev. Mol. Cell Biol.* **6**, 801–811.
- Hemler, M. E. (2014). *Nature Rev. Cancer*, **14**, 49–60.
- Hirata, K., Kawano, Y., Ueno, G., Hashimoto, K., Murakami, H., Hasegawa, K., Hikima, T., Kumasaka, T. & Yamamoto, M. (2013). *J. Phys. Conf. Ser.* **425**, 012002.
- Jankovičová, J., Simon, M., Antalíková, J., Cupperová, P. & Michalková, K. (2015). *Physiol. Res.* **64**, 279–293.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kaji, K., Oda, S., Shikano, T., Ohnuki, T., Uematsu, Y., Sakagami, J., Tada, N., Miyazaki, S. & Kudo, A. (2000). *Nat. Genet.* **24**, 279–282.
- Kitadokoro, K., Bordo, D., Galli, G., Petracca, R., Falugi, F., Abrignani, S., Grandi, G. & Bolognesi, M. (2001). *EMBO J.* **20**, 12–18.
- Kitadokoro, K., Ponassi, M., Galli, G., Petracca, R., Falugi, F., Grandi, G. & Bolognesi, M. (2002). *Biol. Chem.* **383**, 1447–1452.
- Le Naour, F., Rubinstein, E., Jasmin, C., Prenant, M. & Boucheix, C. (2000). *Science*, **287**, 319–321.
- Levy, S. & Shoham, T. (2005). *Nature Rev. Immunol.* **5**, 136–148.
- Miyado, K., Yamada, G., Yamada, S., Hasuwa, H., Nakamura, Y., Ryu, F., Suzuki, K., Kosai, K., Inoue, K., Ogura, A., Okabe, M. & Mekada, E. (2000). *Science*, **287**, 321–324.
- 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.
- Reyes, R., Cardeñes, B., Machado-Pineda, Y. & Cabañas, C. (2018). *Front. Immunol.* **9**, 863.
- Sheldrick, G. M. (2015). *Acta Cryst.* **C71**, 3–8.
- Ueno, G., Hikima, T., Yamashita, K., Hirata, K., Hasegawa, K., Murakami, H., Furukawa, Y., Mizuno, N., Kumasaka, T. & Yamamoto, M. (2016). *AIP Conf. Proc.* **1741**, 050021.
- Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.
- Yamashita, K., Hirata, K. & Yamamoto, M. (2018). *Acta Cryst.* **D74**, 441–449.
- Yáñez-Mó, M., Barreiro, O., Gordon-Alonso, M., Sala-Valdés, M. & Sánchez-Madrid, F. (2009). *Trends Cell Biol.* **19**, 434–446.
- Zimmerman, B., Kelly, B., McMillan, B. J., Seegar, T. C. M., Dror, R. O., Kruse, A. C. & Blacklow, S. C. (2016). *Cell*, **167**, 1041–1051.