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

Overproduction, purification and preliminary crystallographic analysis of the carbohydrate-recognition domain of human langerin

Michel Thépaut, Corinne Vivès, Guillaume Pompidor, Richard Kahn and Franck Fieschi*

Laboratoire des Protéines Membranaires, Institut de Biologie Structurale Jean-Pierre Ebel, UMR 5075 CNRS/CEA/Université Joseph Fourier, 41 Rue Jules Horowitz, 38027 Grenoble CEDEX, France

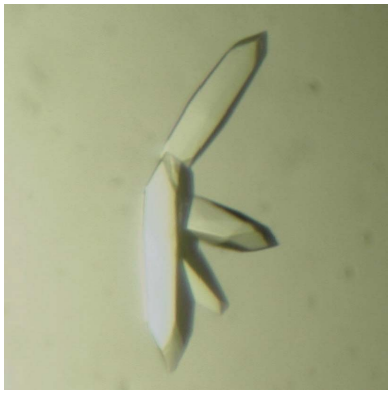
Correspondence e-mail: franck.fieschi@ibs.fr

Received 7 November 2007
Accepted 10 January 2008

Langerin, a lectin that is specific to Langerhans cells, interacts with glycoconjugates through its carbohydrate-recognition domain (CRD). This carbohydrate binding occurs by an avidity-based mechanism that is enabled by the neck domain responsible for trimerization. Langerin binds HIV through its CRD and thus plays a protective role against its propagation by the internalization of virions in Birbeck granules. Here, the overproduction, purification and crystallization of the langerin CRD is reported. Crystals obtained by the hanging-drop vapour-diffusion method allowed the collection of a complete data set to 1.5 Å resolution and belonged to the tetragonal space group $P4_2$, with unit-cell parameters $a = b = 79.55$, $c = 90.14$ Å.

1. Introduction

Langerin (CD 207), a C-type lectin, is a type II transmembrane protein that is located on Langerhans cells (Valladeau *et al.*, 1999, 2000). This protein is composed of a short cytoplasmic domain, a transmembrane spanning segment and an extracellular region divided into an extended neck domain and a C-type carbohydrate-recognition domain (CRD). The neck region allows the oligomerization of langerin, which is liable to an avidity-based mechanism of binding. The avidity compensates for the naturally low affinity of this C-type CRD for mannose-related monosaccharides. Indeed, langerin has been reported to be involved in presenting nonpeptide antigens from *Mycobacterium leprae* to CD1a-restricted T cells (Hunger *et al.*, 2004). Moreover, langerin has been reported to have an ability to bind HIV particles (Turville *et al.*, 2002) and more recently to have a protective role against HIV propagation (de Witte *et al.*, 2007). Such functions have been attributed to its ability to internalize viruses in unique organelles found in the Birbeck granules in Langerhans cells. In addition to glycoconjugate binding, the CRD domain of langerin has been shown to be a crucial player in the formation of Birbeck granules (Verdijk *et al.*, 2005). Thus, structural data on the langerin CRD is awaited in order to highlight its mode of interaction with pathogenic glycoproteins and also to understand the mechanism by which it controls the formation of the unique Birbeck granule. Here, we report the overproduction, purification and crystallization of the langerin CRD. This opens a route to determination of the CRD structure by X-ray crystallography and new insights into the biological mechanisms in which langerin is involved.



2. Materials and methods

2.1. From cloning to purification of langerin CRD

The sequence coding for the carbohydrate-recognition domain of human langerin (Lg S-CRD; amino acids 188–328) was amplified by PCR. The forward-primer sequence 5′-GCATTA*GGTCTC*TGCGC ATG**CAAAATGATATTCTACAGGT**-3′ contains a *Bsa*I restriction site (in italics), a start codon (underlined) and the initial N-terminal amino-acid codons (in bold). The reverse-primer sequence 5′-GCAGCA*GGTCTC*TTATCATCA**CGGTTCTGATGGGACAT**-3′ contains a *Bsa*I restriction site (in italics), a stop codon (underlined) and the last C-terminal amino-acid codons (in bold). The PCR product was inserted into pASK-IBA6 vector (IBA GmbH) at the *Bsa*I sites

in phase with the OmpA signal peptide sequence (allowing periplasmic expression) and a Strep-Tag II sequence (for purification on a Strep-Tactin column), both of which were located at the N-terminal end of the protein. The resulting plasmid, pASK-IBA6 Lg OmpA-S-CRD, was cleaved using *Xba*I and *Hin*dIII restriction enzymes and the fragment including the Lg OmpA-S-CRD gene was purified using agarose gel. The purified fragment was inserted into a pET-30 vector (Novagen), leading to the pET-30 Lg OmpA-S-CRD plasmid (Fig. 1*a*). This plasmid was checked by sequencing and used to transform calcium-competent *Escherichia coli* BL21 (DE3) cells. Culture was initiated from a 5% dilution of an overnight culture into LB medium with 50 mg l^{-1} ampicillin. Cells were grown for 3 h at 290 K with 200 rev min^{-1} shaking and overnight expression of Lg OmpA-S-CRD was induced by the addition of 100 µ*M* isopropyl β-D-1-thiogalactopyranoside and 25 m*M* $CaCl_2$ at 280 K (Fig. 1*b*). Cells were harvested by centrifugation at 5000*g* for 20 min and the pellet obtained from a 1 l overnight culture was resuspended in 50 ml buffer *A* (150 m*M* NaCl, 25 m*M* Tris pH 8 and 4 m*M* $CaCl_2$). Cells were lysed by freezing at 253 K, thawing and sonication with the addition of one complete EDTA-free tablet (Roche Diagnostics). The total protein extract was recovered by centrifugation at 100 000*g* for 1 h at 277 K and was loaded onto a Strep-Tactin Superflow column (IBA GmbH). Unbound proteins were washed with 100 ml buffer *A* and Lg S-CRD was eluted with 40 ml buffer *B* (150 m*M* NaCl, 25 m*M* Tris pH 8, 4 m*M* $CaCl_2$ and 2.5 m*M* desthiobiotin) and collected in 1 ml fractions (Fig. 1*c*). Fractions containing Lg S-CRD were pooled, salt concentrations were adjusted to buffer *A* by gel filtration on a Superdex 75 column (GE Healthcare) and the protein was eluted in 1 ml fractions and concentrated to 5.6 mg ml^{-1} using a centrifugal Vivaspin 20 10 000 MWCO PES filter device (Vivascience). N-terminal sequencing and mass spectrometry (calculated mass, 17 905.1; experimental mass, 17 908.47) confirmed the integrity of the sample.

2.2. Crystallization

Initial crystallization screening experiments were set up with a Cartesian PixSys 4200 crystallization robot (Genomic Solutions) at the High Throughput Crystallization Laboratory (EMBL, Grenoble) using Greiner Crystal Quick plates and the sitting-drop vapour-diffusion method at 293 K. Drops were prepared by mixing 100 nl reservoir solution with 100 nl concentrated Lg S-CRD. The best positive hits were obtained using Index Screen (Hampton Research) with reagents 84 or 85 (200 m*M* $MgCl_2$, 100 m*M* HEPES pH 7.5 or Tris pH 8.5 and 25% PEG 3350). Conditions were optimized manually using the hanging-drop vapour-diffusion method at 293 K in EasyXtal plates (Qiagen); 1 µl reservoir solution plus 1 µl concentrated Lg S-CRD drops were equilibrated against 1 ml reservoir solution. The best crystals appeared within one to eight weeks with reservoir solution composed of 400 m*M* $MgCl_2$, 25% PEG 3350 and 100 m*M* HEPES pH 7 and reached dimensions of 0.2 × 0.2 × 0.6 mm after two weeks of growth (Fig. 2).

Table 1
X-ray data-collection statistics.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	0.979655
Space group	$P4_2$
Unit-cell parameters (Å)	$a = b = 79.55$, $c = 90.14$
Resolution (Å)	50–1.5 (1.59–1.5)
Measured reflections	648915 (99762)
Unique reflections	89721 (14378)
Completeness (%)	99.4 (98.7)
$I/\sigma(I)$	25.42 (6.13)
$R_{\rm merge}$† (%)	4.7 (23.4)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \overline{I(hkl)}|/\sum_{hkl}\sum_i I_i(hkl)$.

2.3. Data collection and processing

A crystal of dimensions 0.2 × 0.2 × 0.4 mm was picked up in a cryoloop, soaked in a solution composed of 400 m*M* $MgCl_2$, 35% PEG 3350 and 100 m*M* HEPES pH 7 for 30 s and flash-frozen in liquid nitrogen. An X-ray diffraction data set was collected on the FIP BM30A beamline at ESRF Grenoble with a Quantum 315r CCD detector (Area Detector Systems Corporation). 360 images were collected at a wavelength of 0.97966 Å (12.656 keV), with an oscillation range of 0.5° per image, an exposure time of 30 s and a crystal-

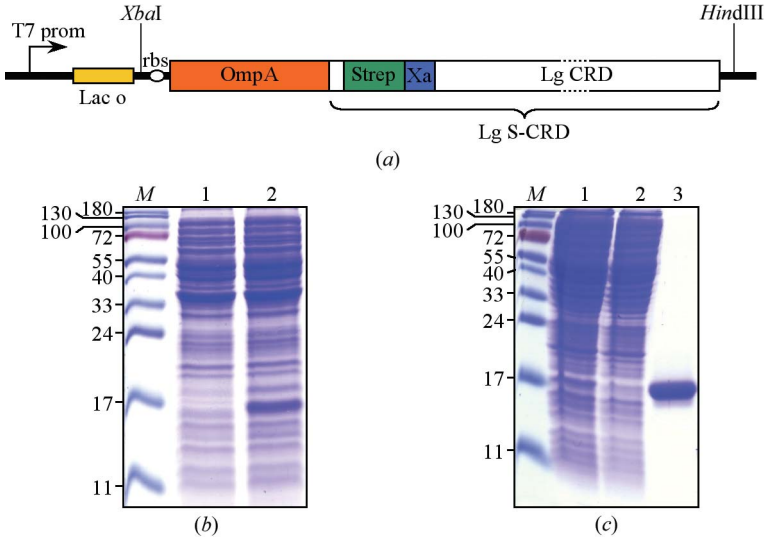


Figure 1
(*a*) Engineered construct for the overproduction and purification of langerin CRD. (*b*) SDS–PAGE analysis of Lg S-CRD overproduction. Lane *M*, PageRuler Prestained Protein Ladder, Fermentas (kDa); lane 1, total proteins before induction; lane 2, total proteins after induction. (*c*) SDS–PAGE of Lg S-CRD. Lane *M*, PageRuler Prestained Protein Ladder, Fermentas (kDa); lane 1, soluble proteins before Strep-Tactin column; lane 2, flowthrough of Strep-Tactin column; lane 3, purified Lg S-CRD eluted from Strep-Tactin column.

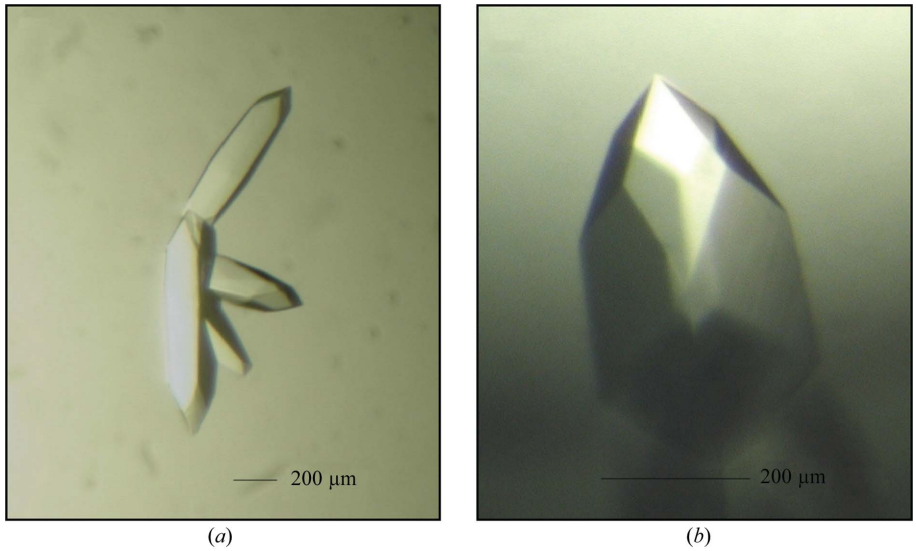
to-detector distance of 200.7 mm. Data were processed using the program *XDS* (Kabsch, 1988). The statistics are presented in Table 1. The Matthews coefficient (V_M) was calculated using the program *MATTHEWS_COEF* (Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

In this article, we first describe an optimization of the classical production protocol for lectin CRDs (Taylor & Drickamer, 2003) applied to langerin CRD. C-type CRDs contain disulfide bridges in their final native form. In order to express the langerin CRD directly in a native form, expression in the oxidative periplasmic environment using an OmpA signal sequence has proven to be successful (Stambach & Taylor, 2003), in contrast to other C-type CRDs, such as DC-SIGN CRD, which were only produced as inclusion bodies. These C-type CRD domains are traditionally purified on monosaccharide-functionalized affinity columns. The naturally low affinity of CRDs for their natural monosaccharide ligands does not allow tight binding and CRDs are merely delayed on such columns. Thus, preparative purification using such media usually implies a prior concentration or lyophilization step before affinity purification in order to minimize the sample volume compared with the column volume. In addition, elution with a retardation profile leads to dilute and spread fractions of the CRD domain, thus requiring an additional long concentration step before any further application to structural experiments. In order to avoid large and tedious concentration steps before and after the affinity-chromatography step and also to improve the purification, we added a Strep-Tag II sequence between the OmpA signal sequence and the DNA sequence encoding the langerin CRD using the appropriate commercial vector (see §2). Additional improvement of the yield obtained was then achieved by transferring this construction in a second cloning step under the control of the strong T7 promoter of the pET vector (Fig. 1*a*). We finally ended up with an expression system that combined the strength of the T7 promoter and the affinity-purification properties provided by the eight-amino-acid Strep-Tag II affinity tag. Indeed, native langerin CRD was overproduced as a soluble protein in the periplasm (Fig. 1*b*) and was purified using Strep-Tag II/Strep-Tactin technology. This method allowed tight binding of the Lg S-CRD on the column, extensive washing and its elution as a sharp concentrated fraction (Fig. 1*c*). This one-step protocol gave us a pure and functional CRD as shown by the classical delayed elution profile obtained from analytical mannose-column chromatography, which was comparable to that obtained using a classical preparative approach (data not shown). As expected from the absence of the neck oligomerization domain in the fragment, Lg S-CRD elutes as a single monomeric peak in gel filtration.

Crystallization trials with Lg S-CRD were conducted without cleavage of the Strep-Tag II. The tag did not hinder the crystallization of the protein and the reproducible conditions led to single crystals. These crystals were robust and diffracted to 1.5 Å resolution (Table 1). Three acceptable values of the Matthews coefficient were calculated using the *CCP*4 package: 4.0, 2.7 or 2.0 Å^3 Da^{-1}, corresponding to solvent contents of 68.9, 53.3 or 37.8%, respectively (Matthews, 1968). Protein crystals that diffract to higher resolution tend to correspond to lower Matthews coefficients (Kantardjieff & Rupp, 2003), which suggests that the value of 4.0 Å^3 Da^{-1} for the Matthews coefficient should be discarded. This suggests the presence of either three or four molecules in the asymmetric unit. A total of 180° of data collected in space group $P4_2$ led to a complete data set with a redundancy higher than 7. Molecular replacement was attempted using *MOLREP* and *AMoRe* from the *CCP*4 package with two existing models, DC-SIGN CRD and DC-SIGNR CRD, and two built models. The first built model was an Lg CRD model generated with *SWISS-MODEL* (Schwede *et al.*, 2003) from the DC-SIGN CRD. The second built model was a mean structure generated with the *DALI* server (Holm & Sander, 1993) from 16 known similar structures (PDB codes 1b6e, 1byf, 1dv8, 1fm5, 1g1t, 1h8u, 1htn, 1hup, 1jzn, 1k9i, 1k9j, 1kg0, 1pw9, 1qdd, 1wmz and 2cs6). To date, no satisfactory solution has been found by molecular replacement. We are currently testing various heavy-atom crystallization strategies in order to solve the phase problem, which is the last pitfall for structure determination.

Insight into the langerin CRD structure will provide information in relation to its role in Birbeck granule formation and its interactions with the HIV envelope glycoprotein and other pathogens. Structural information on this protein will also provide additional information to unravel the selectivity-filter specificity for sugar binding of the C-type CRDs compared with other lectins.



(*a*) (*b*)

Figure 2
Crystals of Lg S-CRD. (*a*) 'Swiss knife-like' rod cluster; (*b*) optimized single crystal.

We thank the B. & M. Gates Foundation for a postdoctoral grant to MT. Sidaction–Ensemble contre le Sida are also acknowledged for their financial support during the initial stages of this work. We also wish to thank E. Girard and L. Serre for their advice in the use of some of the crystallography computing programs, and A. Amara for access to the cDNA encoding the extracellular part of the langerin. J. P. Andrieu and B. Dublet are acknowledged for the N-terminal sequencing and mass-spectrometric quality control of the protein, respectively.

References

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Hunger, R. E., Sieling, P. A., Ochoa, M. T., Sugaya, M., Burdick, A. E., Rea, T. H., Brennan, P. J., Belisle, J. T., Blauvelt, A., Porcelli, S. A. & Modlin, R. L. (2004). *J. Clin. Invest.* **113**, 701–708.
Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.
Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. (2003). *Nucleic Acids Res.* **31**, 3381–3385.
Stambach, N. S. & Taylor, M. E. (2003). *Glycobiology*, **13**, 401–410.
Taylor, M. E. & Drickamer, K. (2003). *Methods Enzymol.* **363**, 3–16.
Turville, S. G., Cameron, P. U., Handley, A., Lin, G., Pöhlmann, S., Doms, R. W. & Cunningham, A. L. (2002). *Nature Immunol.* **3**, 975–983.
Valladeau, J., Duvert-Frances, V., Pin, J. J., Dezutter-Dambuyant, C., Vincent, C., Massacrier, C., Vincent, J., Yoneda, K., Banchereau, J., Caux, C., Davoust, J. & Saeland, S. (1999). *Eur. J. Immunol.* **29**, 2695–2704.
Valladeau, J., Ravel, O., Dezutter-Dambuyant, C., Moore, K., Kleijmeer, M., Liu, Y., Duvert-Frances, V., Vincent, C., Schmitt, D., Davoust, J., Caux, C., Lebecque, S. & Saeland, S. (2000). *Immunity*, **12**, 71–81.
Verdijk, P., Dijkman, R., Plasmeijer, E. I., Mulder, A. A., Zoutman, W. H., Mieke Mommaas, A. & Tensen, C. P. (2005). *J. Invest. Dermatol.* **124**, 714–717.
Witte, L. de, Nabatov, A., Pion, M., Fluitsma, D., de Jong, M. A., de Gruijl, T., Piguet, V., van Kooyk, Y. & Geijtenbeek, T. B. (2007). *Nature Med.* **13**, 367–371.