Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Expression, purification, refolding and crystallization of the carbohydrate-recognition domain of p58/ERGIC-53, an animal C-type lectin involved in export of glycoproteins from the endoplasmic reticulum

Lucas M. Velloso,[a] Kerstin Svensson,[b] Ulla Lahtinen,[b] Gunter Schneider,[a] Ralf F. Pettersson[b] and Ylva Lindqvist[a]*

[a]Molecular Structural Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-17177 Stockholm, Sweden, and [b]Ludwig Institute for Cancer Research, Stockholm Branch, Karolinska Institutet, Box 240, S-17177 Stockholm, Sweden

Correspondence e-mail: ylva@alfa.mbb.ki.se

Received 21 November 2001
Accepted 2 January 2002

p58/ERGIC-53 is a mammalian calcium-dependent lectin that serves as a glycoprotein-sorting receptor between the endoplasmic reticulum (ER) and the Golgi complex. It is a type I transmembrane protein with two lumenal domains, one of which is a carbohydrate-recognition domain (CRD) and homologous to leguminous lectins. The CRD of p58, the rat homologue of human ERGIC-53, was overexpressed in insect cells and *Escherichia coli*, purified and crystallized using Li_2SO_4 as a precipitant. The crystals belong to space group *I*222, with unit-cell parameters $a = 49.6$, $b = 86.1$, $c = 128.1$ Å, and contain one molecule per asymmetric unit, corresponding to a packing density of 2.4 $Å^3\,Da^{-1}$. Knowledge of the structure of p58/ERGIC-53 will provide a starting model for understanding receptor-mediated glycoprotein sorting between the ER and the Golgi.

1. Introduction

Secretory proteins undergo multiple rounds of quality control within the endoplasmic reticulum (ER) to ensure that only correctly folded proteins reach their final destinations in eukaryotic cells (Ellgaard *et al.*, 1999). Transport vesicles that bud from the ER initially fuse or mature into the ER-Golgi intermediate compartment (ERGIC), where segregation between resident and secretory proteins takes place (Aridor *et al.*, 1995; Bannykh & Balch, 1997). Growing evidence suggests that export of secretory proteins from the ER is selective, at least for some proteins (Klumperman, 2001; Mitsuno & Singer, 1993). The exit step therefore appears to require transmembrane cargo receptors. These proteins should have a cytoplasmic tail recognized by COPII coatomer proteins, a transmembrane domain and a cargo-binding domain in the ER lumen (Herrmann *et al.*, 1999). ERGIC-53 (human), and its rat homologue p58 (Saraste *et al.*, 1987; Lahtinen *et al.*, 1996), is a mannose-selective lectin (Itin *et al.*, 1996) and is the most commonly used marker for the ERGIC compartment (Saraste & Svensson, 1991; Hauri & Schweizer, 1992; Hauri *et al.*, 2000). It has been shown to act as a cargo receptor for some glycoproteins (Appenzeller *et al.*, 1999; Hauri *et al.*, 2000). Mutations in ERGIC-53 lead to a bleeding disorder known as combined deficiency of coagulation factors V and VIII (Nichols *et al.*, 1998). p58/ERGIC-53 is an unglycosylated type I transmembrane protein with lumenal, transmembrane and cytoplasmic domains. The lumenal domain consists of a carbohydrate-recognition domain (CRD), which displays weak homology to leguminous lectins, and a putative oligomerization domain, which is responsible for the formation of dimers and hexamers of this protein (Lahtinen *et al.*, 1999; Hauri *et al.*, 2000). Other genes with low but significant sequence homology to p58/ERGIC-53 have recently been identified, suggesting that these proteins define a new family of animal lectins involved in quality control in the secretory pathway (Fiedler & Simons, 1994, 1995). In the following, we use the name p58 as a synonym for ERGIC-53.

Here, we describe the purification, crystallization and preliminary X-ray diffraction analysis of the CRD of p58. Knowledge of the three-dimensional structure of p58 will allow a better understanding of its function in the secretory pathway.

2. Methods and results

2.1. Protein production in insect cells

The CRD of p58 (residues 1–285) was defined by a combination of sequence alignment, secondary-structure prediction and limited proteolysis of a longer construct which also contained part of the oligomerization domain and consisted of residues 1–428 of the p58 sequence (Fig. 1). We used the same expression and purification protocol for both the CRD and this longer construct. Residues 1–285 of p58, corresponding to the putative CRD (Fig. 1) and including an insertion of a His_6 tag between residues 34 and 35, were amplified by PCR using the p58 cDNA

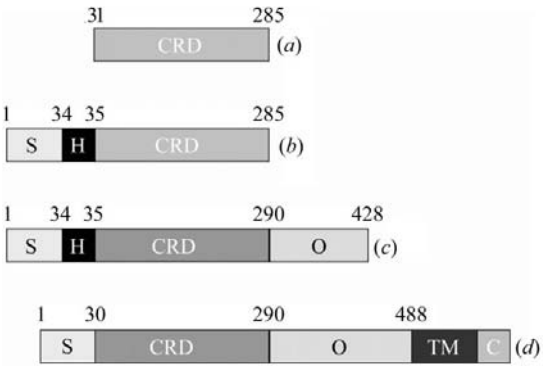

Figure 1
Different constructs of p58 used for protein production/crystallization. Residue numbers are indicated for domain borders. (*a*) CRD expressed in *E. coli*. (*b*) CRD expressed in insect cells. (*c*) The initial construct, containing part of the oligomerization domain, used for crystallization attempts and determination of the CRD boundaries. The wild-type form of the protein (*d*) is shown for comparison. Abbreviations: CRD, carbohydrate-recognition domain; H, His_6 tag; S, signal sequence; O, putative oligomerization domain; TM, transmembrane domain; C, cytoplasmic tail.

(Lahtinen *et al.*, 1996) as a template. Residues 1–30 constitute the native signal sequence for translocation through the ER membrane and are cleaved upon the secretion of the protein into the medium. *Bam*HI and *Hin*dIII sites were introduced upstream and downstream of the fragment, respectively. The PCR-generated fragment was cloned into *Bam*HI/*Hin*dIII-restricted pFASTBac1 transfer plasmid (Life Technologies). The recombinant pFASTBac-plasmid was transformed into DH10Bac cells and recombinant bacmid DNA was isolated. Sf-9 cells (600 ml) at 2×10^6 cells ml^{-1} were infected with high-titre baculovirus stocks. 4 d after infection the cells were removed by centrifugation and the culture medium was concentrated using an Ultrasette ultrafiltration device (Pall Gelman Laboratories, Ann Arbor, MI, USA). The concentrated medium was incubated in the presence of Ni–NTA beads (Qiagen) for 120 min. The beads were washed three times with 20 m*M* Tris–HCl pH 8.0, 0.3 *M* NaCl, 10% glycerol, 5 m*M* $CaCl_2$ and once with 10 m*M* imidazole in the same buffer. The protein was eluted with the same buffer containing 100 m*M* imidazole, concentrated to 10 mg ml^{-1}, dialyzed against 10 m*M* Tris pH 7.5, 2 m*M* $CaCl_2$ and stored at 193 K.

2.2. Protein production in *E. coli*

Residues 31–285 of p58, corresponding to the putative CRD, were amplified by PCR, with the insertion of *Nco*I and *Hin*dIII restriction sites upstream and downstream of the gene, respectively. The resulting fragment was cloned into the corresponding restriction sites of the vector PET-28 (Novagen), which was then transformed into BL-21 Codon Plus *E. coli* cells (Stratagene). Protein production was induced with the addition of 1 m*M* IPTG to cells grown at 310 K when the OD reached a value of 0.6–1.0. The product was purified as inclusion bodies using previously described protocols (Garboczi *et al.*, 1992; Reid *et al.*, 1996), but slightly modified in that the washing buffer contained 1% deoxycholic acid.

The protein was refolded by rapid dilution as described previously (Boyington *et al.*, 2000). Briefly, the inclusion bodies were diluted into 1 *M* L-arginine, 0.5 *M* NaCl, 0.1 *M* Tris–HCl pH 7.5, 1 m*M* $NiCl_2$, 5 m*M* DTT with vigorous stirring and allowed to refold for 72 h. The protein solution was then concentrated 20-fold and dialyzed against 0.1 *M* Tris–HCl pH 7.5, 150 m*M* NaCl for 48 h with three buffer changes, before being concentrated to 4.0 mg ml^{-1} and applied to a Superdex 75 XK16/60 gel-filtration column (Amersham Pharmacia Biotech) equilibrated with 25 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl. Two major peaks were observed in the gel-filtration profile (Fig. 2*a*): one was in the void volume of the column (50 ml), corresponding to aggregated protein, and the other was at around 75 ml, corresponding to the CRD of p58. The fractions corresponding to this peak were pooled, concentrated to 10 mg ml^{-1} and stored at 193 K.

2.3. Crystallization

Initial crystallization attempts used the construct containing residues 1–428 of p58 (Fig. 1), which was expressed in insect cells. However, this construct failed to crystallize and suffered from proteolytic degradation (Fig. 2*a*). Limited proteolysis of this construct, followed by N- and C-terminal peptide sequencing and mass spectrometry, identified a trypsin/chymotrypsin-resistant fragment that basically comprised the sequence of the CRD (data not shown). We therefore chose to continue crystallization attempts with the CRD constructs produced both in *E. coli* (residues 31–285) and in insect cells (residues 1–285). Following overexpression and purification of the CRD (Fig. 2*b*), initial crystallization attempts were carried out using sparse-matrix sampling kits (Hampton Research). Crystals were initially grown by vapor diffusion at 293 K from hanging drops containing 2 µl 10 mg ml^{-1} protein solution in 10 m*M* Tris–HCl pH 7.5, 1 m*M* $CaCl_2$ and 2 µl well solution (100 m*M* Na HEPES pH 7.25 and 1.6 *M* Li_2SO_4) equilibrated against 1 ml of well solution. The first crystals obtained grew within a week and were up to 0.5 mm in the longest dimension, but were polycrystalline. Addition of 10 m*M* EDTA to the crystallization drop resulted in a single crystals with typical dimensions of 0.5 × 0.3 × 0.1 mm (Fig. 3) which diffracted well. Therefore, all subsequent crystallizations employed EDTA as an additive. The CRD domains produced in insect cells and in *E. coli* both crystallized under these conditions.

2.4. Data collection

Crystals grown from the CRD produced both in insect cells and in *E. coli* belonged to the same space group, with similar unit-cell parameters. Since the crystals from the

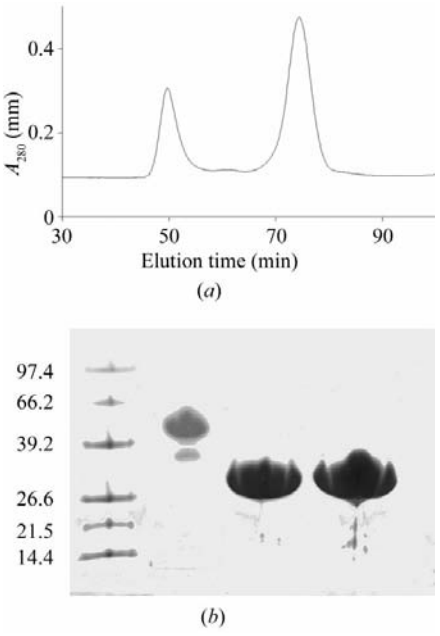

Figure 2
Purification of p58. (*a*) Gel-filtration profile of the refolded CRD on a Superdex 75 XK 16/60 column. The peak at around 50 ml corresponds to aggregated proteins eluting in the void volume of the column, whereas that at around 75 ml contains the refolded CRD. (*b*) SDS–PAGE analysis of purified p58 constructs. Lane 1, molecular-weight marker (sizes in kDa indicated); lane 2, longer construct (residues 31–428) containing part of the putative oligomerization domain and used for initial crystallizations; lane 3, CRD (residues 1–285) expressed in insect cells; lane 4, refolded CRD produced in *E. coli* (residues 31–285).

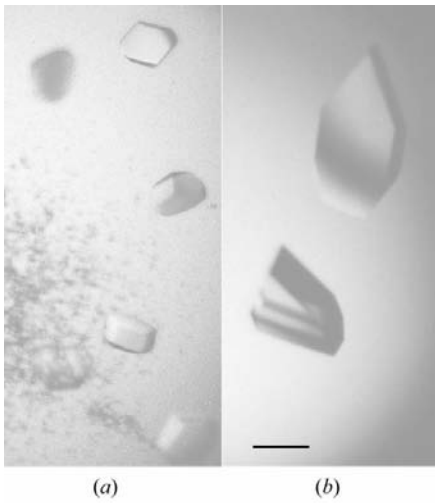

Figure 3
Crystals of the carbohydrate-recognition domain of p58 produced (*a*) in *E. coli* and (*b*) in insect cells. The scale bar is 0.2 mm long.

protein produced in insect cells grew to a larger size, they were chosen for data collection with synchroton radiation. X-ray data were collected at beamline 7-11 at MAX-laboratory (Lund, Sweden). For cryoprotection, crystals were transferred into a solution of 1.6 *M* Li_2SO_4, 0.1 *M* Na HEPES pH 7.25, 20% ethylene glycol for around 10 s and frozen in a liquid-nitrogen stream. The crystals diffracted to better than 1.46 Å. Data were collected on one crystal at a wavelength of $\lambda = 1.0326$ Å on a MAR 345 image-plate detector in two passes, a low-resolution (25–2.44 Å) and a high-resolution (1.46–10 Å) run. Data were processed using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997) (Table 1).

Space-group determination was performed using the autoindexing routine of *DENZO* (Otwinowski & Minor, 1997) and pseudo-precession photographs generated with the program *PATTERN* (Lu, 1999). The crystals belong to the orthorhombic space group *I*222, with unit-cell parameters $a = 49.6$, $b = 86.1$, $c = 128.1$ Å. The presence of a monomer in the asymmetric unit gives a crystal packing parameter of 2.4 $Å^3\ Da^{-1}$ and a solvent content of 48%.

Since p58 displays only very limited sequence homology to other proteins, including leguminous lectins, we are currently screening for heavy-atom derivatives for structure determination by MIR. The structure of the CRD of p58 will provide insights into how this protein interacts with its ligands and thereby facilitates glycoprotein trafficking from the ER to the Golgi.

Table 1
Data-collection statistics for p58.

Values in parentheses refer to the outer resolution shell (1.49–1.46 Å).

Resolution (Å)	1.46
Unique reflections	44706
Redundancy	4.5
$I/\sigma(I)$	22.2 (7.3)
$R_{\rm merge}$†	4.1 (11.9)
Completeness	95.6 (84.6)

† $R_{\rm merge} = \sum_h \sum_i |I_{h,i} - \langle I \rangle| / \sum_h \sum_i I_{h,i}$.

We thank Adnane Achour and Jackob Michaelson for help and advice with refolding experiments, Tatyana Sandalova for help with data collection and the MAX-laboratory, Lund, Sweden for access to synchrotron radiation. This work was supported by grants from the Swedish Research Council.

References

Appenzeller, C., Andersson, H., Kappeler, F. & Hauri, H.-P. (1999). *Nature Cell. Biol.* **1**, 330–334.
Aridor, M., Bannykh, S. I., Rowe, T. & Balch, W. E. (1995). *J. Cell Biol.* **131**, 875–893.
Bannykh, S. I. & Balch, W. E. (1997). *J. Cell Biol.* **138**, 1–4.
Boyington, J. C., Rais, A. N., Brooks, A. G., Patamawenu, A. & Sun, P. D. (2000). *Protein Expr. Purif.* **18**, 235–241.
Ellgaard, L., Molinari, M. & Helenius, A. (1999). *Science*, **286**, 1882–1888.
Fiedler, K. & Simons, K. (1994). *Cell*, **77**, 625–626.
Fiedler, K. & Simons, K. (1995). *J. Cell Sci.* **109**, 271–276.
Garboczi, D. N., Hung, D. T. & Wiley, D. C. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 3429–3433.
Hauri, H.-P., Kappeler, F., Andersson, H. & Appenzeller, C. (2000). *J. Cell Sci.* **113**, 587–596.
Hauri, H.-P. & Schweizer, A. (1992). *Curr. Opin. Cell. Biol.* **4**, 600–608.
Herrmann, J. M., Malkus, P. & Schekman, R. (1999). *Trends Cell Biol.* **9**, 5–7.
Itin, C., Roche, A.-C., Monsigny, M. & Hauri, H.-P. (1996). *Mol. Biol. Cell*, **7**, 483–493.
Klumperman, J. (2001). *Curr. Opin. Cell Biol.* **12**, 445–449.
Lahtinen, U., Hellman, U., Wernstedt, C., Saraste, J. & Pettersson, R. F. (1996). *J. Biol. Chem.* **271**, 4031–4937.
Lahtinen, U., Svensson, K. & Pettersson, R. F. (1999). *Eur. J. Biochem.* **260**, 392–397.
Lu, G. (1999). *J. Appl. Cryst.* **32**, 375–376.
Mitsuno, M. & Singer, S. J. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 5732–5736.
Nichols, W. C., Seligsohn, U., Zivelin, A., Terry, V. H., Hertel, C. E., Weatley, M. A., Moussalli, M., Hauri, H. P., Ciavarella, N., Kaufman, R. J. & Ginsburg, D. (1998). *Cell*, **93**, 61–70.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Reid, S. W., Smith, K. J., Jakobsen, B. K., O'Callaghan, C. A., Reyburn, H., Harlos, K., Stuart, D. I., McMichael, A. J., Bell, J. I. & Jones, E. Y. (1996). *FEBS Lett.* **383**, 119–123.
Saraste, J., Palade, G. E. & Farquhar, M. G. (1987). *J. Cell Biol.* **105**, 2021–2029.
Saraste, J. & Svensson, K. (1991). *J. Cell Sci.* **100**, 415–430.