Crystallization and preliminary X-ray diffraction analysis of proximal thread matrix protein 1 (PTMP1) from *Mytilus galloprovincialis*

Michael H. Suhre,^a Thomas Scheibel,^a Clemens Steegborn^{b*} and Melanie Gertz^{b*}

^aDepartment of Biomaterials, University of Bayreuth, Universitätsstrasse 30, 95447 Bayreuth, Germany, and ^bDepartment of Biochemistry, University of Bayreuth, Universitätsstrasse 30, 95447 Bayreuth, Germany

Correspondence e-mail:
clemens.steegborn@uni-bayreuth.de,
melanie.gertz@uni-bayreuth.de

Received 17 January 2014
Accepted 19 March 2014

In order to deal with the dynamic ocean environment, blue mussels adhere to various surfaces *via* their collagenous byssal threads. PTMP1 (proximal thread matrix protein 1) is one identified matrix protein residing in the proximal thread and is capable of collagen binding. Its sequence comprises two von Willebrand factor type A-like repeats. In order to characterize the structure and domain architecture of PTMP1, recombinant protein was crystallized by vapour diffusion. The obtained crystals diffracted to 1.95 Å resolution and belonged to space group $P2_1$, with unit-cell parameters $a = 62.0$, $b = 62.3$, $c = 122.6$ Å, $\beta = 102.2^\circ$. The Matthews coefficient suggested the presence of two monomers in the asymmetric unit and 48.3% solvent content.

1. Introduction

Sessile marine organisms inhabiting the intertidal zone have to cope with strong forces from waves and currents. Therefore, in blue mussels (*Mytilus* sp.) an intriguing holdfast structure has evolved, the byssus, consisting of several dozens of threads attached to a stem originating from the mussel foot. Each thread represents a mechanically graded material with high stiffness at the distal end and flexibility within its proximal portion (Waite *et al.*, 2004). Byssal threads are almost exclusively of proteinaceous nature, with special collagens being the main load-bearing components (Hagenau *et al.*, 2011). The byssal collagens are arranged in fibrils surrounded by matrix proteins, including proximal thread matrix protein 1 (PTMP1; Sun *et al.*, 2002) and the thread matrix proteins (TMPs; Sagert & Waite, 2009). In contrast to the TMPs, PTMP1 appears to locate exclusively in the proximal thread portion and has been successfully extracted from the mature byssus. Sequence analysis of PTMP1 revealed a two-repeat structure, with each repeat exhibiting homology to von Willebrand factor type A domains (PFAM PF00092). Similar to many of von Willebrand factor type A-like proteins, PTMP1 was also shown to bind to collagen (Sun *et al.*, 2002). Previously, PTMP1 was postulated to inter-bridge collagens in the byssus and therefore influence some of the mechanical characteristics (*e.g.* flexibility) of the respective proximal thread part (Waite *et al.*, 2002). In order to elucidate the structure and the arrangement of its A-domains, we recombinantly produced, purified, refolded and crystallized PTMP1 from *M. galloprovincialis* by vapour diffusion and collected and analyzed X-ray diffraction data from the obtained crystals.

2. Materials and methods

2.1. Cloning and expression

The full-length PTMP1 cDNA sequence (GenBank accession No. AY053391.1) was PCR-amplified from a *M. galloprovincialis* mussel foot cDNA library created with the CloneMiner cDNA Construction Kit (Invitrogen) according to the manufacturer's manual. The respective 5'- and 3'-ends were confirmed by amplification and sequencing of PTMP1 containing pDONR222 vectors using a PTMP1-specific and a vector-specific primer. The sequence of PTMP1 was inserted into a pGEM-T vector (Promega) by TA cloning. From the five isoforms obtained, the one that best matched the published sequence was chosen, with three amino-acid residues differing in the mature protein. Using the *SignalP* 3.0 server (CBS)

the most probable cleavage site for the secretion signal peptide was predicted, and the signal-peptide-free sequence was amplified using specific primers and cloned into pET28a(+)-vector (Novagen, *Bam*HI and *Not*I restriction sites) additionally encoding for an N-terminal SUMO tag (*Saccharomyces cerevisiae*, located between *Nde*I and *Bam*HI restriction sites). The recombinant PTMP1 was produced in *Escherichia coli* BL21(DE3)-RIPL (Novagen) cultured at 37°C in aerated 5 l bottles in LB medium containing 35 mg l⁻¹ kanamycin and 0.001% (v/v) Breox FMT 30 antifoam (Cognis). Expression was induced by the addition of 0.1 mM IPTG when the optical density reached 0.8; culture continued at 30°C for 3 h and the cells were subsequently harvested by centrifugation.

2.2. Purification

Recombinant PTMP1 was isolated from bacterial inclusion bodies under denaturing conditions according to Rudolph *et al.* (1997) with several adaptations. Cells were disrupted by the addition of lysozyme

and subsequent ultrasonic treatment. Inclusion bodies were isolated by centrifugation and several washing steps with Triton X-100 and Tris-buffered saline (TBS), and nucleic acids were removed by ultrasonic treatment, yielding ~150 mg protein per litre of cell culture. Subsequently, the inclusion bodies were denatured in buffer A (20 mM sodium phosphate pH 8, 150 mM NaCl) supplemented with 8 M urea and 5 mM DTT, and the recombinant protein was isolated by immobilized metal-ion affinity chromatography (IMAC) on a Ni-loaded Chelating Sepharose column (GE Healthcare). Eluted protein was pooled, dialyzed against buffer A supplemented with 1 M urea and 1 mM DTT, and the SUMO tag was cleaved off by digestion with recombinant His-tagged *S. cerevisiae* Ulp1 protease. The digest resulted in PTMP1 carrying an additional N-terminal Ser residue, and the released 118 amino acids of the 6×His-SUMO tag were removed by reverse Ni-IMAC. Refolding of the protein was accomplished by dialysis first against buffer A supplemented with 1 mM DTT and subsequently extensively (several days) against

Figure 1
(a) Size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) analysis of purified PTMP1 exhibiting a high monomer content (96.5%) with only 3.5% residual dimeric protein. The observed monomeric mass is 48.5 kDa, matching the calculated molecular weight of recombinant PTMP1 of 48.9 kDa. Inset, Coomassie R250-stained SDS-PAGE with purified PTMP1. Lane 1, molecular-weight protein standard (New England Biolabs); lane 2, 1.5 µg PTMP1. (b) Optimized crystals of PTMP1. X-ray diffraction patterns at φ angles of 0° (c) and 90° (d). Insets, enlarged views.

Table 1

Diffraction data-collection statistics.

Values in parentheses are for the outermost shell.

Space group	$P2_1$
Unit-cell parameters ($\text{\AA}$, $^\circ$)	$a = 62.0$, $b = 62.3$, $c = 122.6$, $\beta = 102.2$
Resolution [†] ($\text{\AA}$)	55.3–1.95 (2.0–1.95)
Unique reflections	63031 (4590)
Multiplicity	2.0 (1.7)
$\langle I/\sigma(I) \rangle$	5.0 (0.9)
Completeness (%)	94.2 (93.0)
$R_{\text{merge}}^{\ddagger}$ (%)	16.7 (84.4)
$R_{\text{meas}}^{\S}$ (%)	22.0 (118.3)
$CC_{1/2}^{\P}$	96.6 (36.6)

[†] The high-resolution limit was chosen according to the $CC_{1/2}$ value described by Karplus & Diederichs (2012). [‡] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an individual measurement and $\langle I(hkl) \rangle$ is the corresponding mean value. [§] $R_{\text{meas}} = \sum_{hkl} \{N(hkl) / [N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an individual measurement, $\langle I(hkl) \rangle$ is the corresponding mean value, and hkl and $N(hkl)$ are the indices and redundancies of the reflections. For details, see Diederichs & Karplus (1997). [¶] $CC_{1/2}$ is the Pearson correlation coefficient between two parts of the data set, each containing a random half of the measurements for each unique reflection (Karplus & Diederichs, 2012).

buffer A lacking any reducing agent to achieve intramolecular disulfide-bond formation [controlled by analysis with Ellman's reagent 5,5'-dithiobis-(2-nitrobenzoic acid); Ellman, 1958]. Finally, monomeric PTMP1 was isolated by size-exclusion chromatography on a Superdex 200 column (GE Healthcare) and the protein was dialyzed against 10 mM ammonium bicarbonate and lyophilized. Purity was estimated from SDS-PAGE and size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) using a DAWN EOS MALS detector (Wyatt Technology) to be higher than 95% (Fig. 1a) and the identity of the protein was confirmed by nano-LC ESI mass spectrometry (sequence coverage 35.9%, score 183.7). Protein concentration was determined by UV spectroscopy, using a calculated extinction coefficient of $15\,275\text{ M}^{-1}\text{ cm}^{-1}$ (ProtParam; <http://www.expasy.org>) which was validated by gravimetric analysis. The yield of purified PTMP1 was $\sim 10\text{ mg}$ per litre of cell culture.

2.3. Crystallization

For crystallization experiments lyophilized PTMP1 was dissolved in 20 mM Tris-HCl pH 8, 100 mM NaCl, ultracentrifuged (30 min, 186 000g) and the concentration was adjusted to 7.5 mg ml^{-1} . Initial crystallization trials were performed using a Phoenix Liquid Handling System (Art Robbins Instruments) and The JCSG Core Suites I–IV crystallization screens (Qiagen). 0.2 μl protein solution were mixed with 0.2 μl reservoir in sitting drops in 96-well MRC 2 well crystallization plates (Molecular Dimensions) and equilibrated against 50 μl reservoir at 20°C . A single box-shaped crystal appeared after two weeks with a reservoir consisting of 30%(w/v) PEG 2000 MME, 0.2 M ammonium sulfate, 0.1 M sodium acetate pH 4.6. For crystal reproduction and improvement PEG and salt concentrations were varied against the buffer pH. Optimal crystal growth was obtained by mixing 1 μl protein solution with 1 μl reservoir consisting of 28%(w/v) PEG 2000 MME, 0.2 M ammonium sulfate, 0.1 M sodium acetate pH 4.6 in sitting drops and equilibrating against 750 μl reservoir in 24-well crystallization plates (Hampton Research) at 20°C . The crystals grew reproducibly as thin plates of 300–400 μm in length (Fig. 1b). Interestingly, these crystals belonged to space group $P2_1$ (see §3), while the initial box-shaped crystal was found to belong to space group C2 (Suhre *et al.*, 2014).

2.4. Data collection and processing

For data collection, optimized crystals were flash-cooled in liquid nitrogen after incubation in cryoprotectant solution [22.5%(w/v) PEG 2000 MME, 0.2 M ammonium sulfate, 0.1 M sodium acetate pH 4.6, 25%(v/v) glycerol]. X-ray diffraction data were collected on MX beamline BL14.1 operated by the Helmholtz-Zentrum Berlin (HZB) at the BESSY II electron-storage ring (Berlin-Adlershof, Germany; Mueller *et al.*, 2012) with an MX-225 CCD detector (Rayonix, Evanston, USA). At a wavelength of 0.918 $\text{\AA}$ and a crystal-to-detector distance of 198 mm a 130° data set was collected with 1° rotation per frame at -173°C . Indexing, scaling and merging were performed with XDS (Kabsch, 2010).

3. Results and discussion

In this study, we successfully cloned and produced full-length PTMP1 protein from *M. galloprovincialis* in *E. coli* and established purification and crystallization of the recombinant protein. Initial diffraction analysis of the optimized crystals showed that they belonged to space group $P2_1$, with unit-cell parameters $a = 62.0$, $b = 62.3$, $c = 122.6\text{ \AA}$, $\beta = 102.2^\circ$. A complete X-ray diffraction data set was collected over 130° with a resolution of 1.95 $\text{\AA}$ (Table 1). The diffraction showed a significantly anisotropic behaviour during rotation of the φ angle, with good data quality only for the flat crystal side perpendicular to the beam and elongated ellipsoid reflections after 90° rotation (Figs. 1c and 1d). In order to ensure good overall data quality, 35 frames of the worst diffracting orientation had to be removed from the data set. Therefore, the diffraction data were processed and scaled in two separate sets and truncated individually (data set 1, frames 1–75 at 1.95 $\text{\AA}$ resolution; data set 2, frames 110–130 at 2.1 $\text{\AA}$ resolution) at the resolution indicated by the respective $CC_{1/2}$ value. Merging the separate sets into one data set resulted in slightly reduced overall completeness compared with processing frames 1–130 in one run. However, the merged data set featured a good compromise with sufficient completeness ($>94\%$) and redundancy (2.0) yet significantly improved $CC_{1/2}$ and R_{meas} values, indicating better data quality from this processing strategy and thus allowing the inclusion of higher resolution data based on $CC_{1/2}$.

Analysis of the obtained diffraction data revealed a Matthews coefficient of $2.38\text{ \AA}^3\text{ Da}^{-1}$, indicating two molecules in the asymmetric unit with a solvent content of 48.3% (Matthews, 1968). The obtained diffraction data set eventually provided the basis for solution of the structure of PTMP1. Despite the high structural similarity – at low sequence identity – of von Willebrand factor A (vWA) domain proteins, automatic molecular-replacement searches failed, but manual selection of a search-model combination from the large number of vWA domain structures in the Protein Data Bank ultimately enabled molecular-replacement phasing. Phasing details, structure refinement and structure analysis are described elsewhere (Suhre *et al.*, 2014).

We thank Lukas Eisoldt for providing Ulp1 protease, Dr Frank Fischer and C. Synatschke for mass-spectrometric analysis and Dr Michael Weyand and Dr Sebastien Moniot, as well as the BESSY (Berlin) beamline staff, for technical assistance. Financial support from DFG SCHE 603/7-1 is gratefully acknowledged.

References

- Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
- Ellman, G. L. (1958). *Arch. Biochem. Biophys.* **74**, 443–450.
- Hagenau, A., Papadopoulos, P., Kremer, F. & Scheibel, T. (2011). *J. Struct. Biol.* **175**, 339–347.

- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Mueller, U., Darowski, N., Fuchs, M. R., Förster, R., Hellmig, M., Paithankar, K. S., Pühringer, S., Steffien, M., Zocher, G. & Weiss, M. S. (2012). *J. Synchrotron Rad.* **19**, 442–449.
- Rudolph, R., Böhm, G., Lilie, H. & Jaenicke, R. (1997). *Protein Function: a Practical Approach*, edited by T. E. Creighton, pp. 57–99. Oxford University Press.
- Sagert, J. & Waite, J. H. (2009). *J. Exp. Biol.* **212**, 2224–2236.
- Suhre, M. H., Gertz, M., Steegborn, C. & Scheibel, T. (2014). *Nature Commun.* **5**, 3392.
- Sun, C., Lucas, J. M. & Waite, J. H. (2002). *Biomacromolecules*, **3**, 1240–1248.
- Waite, J. H., Lichtenegger, H. C., Stucky, G. D. & Hansma, P. (2004). *Biochemistry*, **43**, 7653–7662.
- Waite, J. H., Vaccaro, E., Sun, C. & Lucas, J. M. (2002). *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **357**, 143–153.