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

Expression, purification, crystallization and preliminary X-ray analysis of the olfactomedin domain from the sea urchin cell-adhesion protein amassin

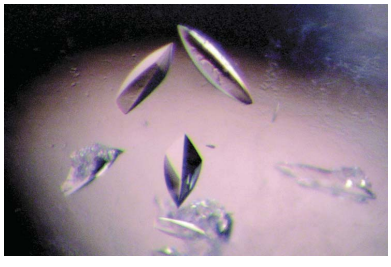
Brian J. Hillier,[a]* Vidyasankar Sundaresan,[b]‡ C. David Stout[b] and Victor D. Vacquier[a]

[a]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0202, USA, and [b]Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

‡ Present address: GE Infrastructure, Water and Process Technologies, Enabling Technologies Group, Trevose, PA 19053-6783, USA.

Correspondence e-mail: bhillier@ucsd.edu

Received 24 October 2005
Accepted 23 November 2005
Online 16 December 2005



A family of animal proteins is emerging which contain a conserved protein motif known as an olfactomedin (OLF) domain. Novel extracellular protein–protein interactions occur through this domain. The OLF-family member amassin, from the sea urchin *Strongylocentrotus purpuratus*, has previously been identified to mediate a rapid cell-adhesion event resulting in a large aggregation of coelomocytes, the circulating immune cells. In this work, heterologous expression and purification of the OLF domain from amassin was carried out and initial crystallization trials were performed. A native data set has been collected, extending to 2.7 Å under preliminary cryoconditions, using an in-house generator. This work leads the way to the determination of the first structure of an OLF domain.

1. Introduction

Originally identified in its namesake protein isolated from the olfactory neuroepithelial mucous layer in the bullfrog *Rana catesbeiana* (Snyder *et al.*, 1991; Yokoe & Anholt, 1993), the number of proteins identified as containing an olfactomedin (OLF) domain continues to grow. Currently, members are found in diverse animal groups including nematodes, arthropods, echinoderms and vertebrates (Bateman *et al.*, 2002). An interest in the OLF domain has arisen in part as a consequence of the identification of the ability of proteins containing this domain to stimulate neurogenic events (Moreno & Bronner-Fraser, 2001; Tsuda *et al.*, 2002) such as neural crest cell formation (Barembaum *et al.*, 2000) and the formation of nodes of Ranvier (Eshed *et al.*, 2005). A majority of research also centers around the finding that mutations in the human protein myocilin, which has been implicated in at least one form of glaucoma, are concentrated within the OLF domain (Ray *et al.*, 2003; Tamm & Russell, 2001; Stone *et al.*, 1997; Adam *et al.*, 1997). The functions of the OLF-family members are often not understood, but recently described activities of these proteins point towards a general structural/organizational role on the extracellular surface. For example, the OLF protein UNC-122 in *Caenorhabditis elegans* appears to organize components at the neuromuscular junction (Loria *et al.*, 2004). Additionally, a unique cell-adhesion activity has been identified for the OLF-family member amassin (Hillier & Vacquier, 2003), the subject of this research. Purified amassin mediates rapid intercellular adhesion of sea urchin (*Strongylocentrotus purpuratus*) coelomocytes.

Amassin is a 495-amino-acid extracellular and glycosylated protein found in the coelomic fluid. Along its length are several structural features: a signal peptide directing its secretion, a predicted short β-region, a region of dimerizing coiled coils and finally the OLF domain at its C-terminus. Previous research utilized various truncated amassin proteins heterologously expressed in the yeast *Pichia pastoris* to find that the OLF domain alone was the portion responsible for coelomocyte-binding activity (Hillier & Vacquier, submitted). While the receptor for amassin's OLF domain on the cell surface has not been identified, other OLF-family members have also been found to bind their targets through this domain. Examples include the protein gliomedin that binds its targets neurofascin and NrCAM through the OLF domain (Eshed *et al.*, 2005) and a

heterotypic OLF–OLF interaction occurring between myocilin and optimedin (Torrado *et al.*, 2002).

OLF domains have been found by secondary-structure prediction (Hillier & Vacquier, 2003) and by circular dichroism (Nagy *et al.*, 2003) to adopt an all-β structure. The domain typically spans ~250 amino acids. No further structural information is known for these domains. To increase our knowledge of this domain, we have set out to obtain the three-dimensional structure of the OLF domain from amassin. To this end, proteins heterologously expressed in *P. pastoris* and *Escherichia coli* have been screened for crystallization ability. Owing to the identification and requirement for proper disulfide bonds in the entire amassin protein, as well as within its OLF domain (Hillier & Vacquier, submitted), expression by secretion in the eukaryotic yeast host was initially utilized. This was followed by prokaryotic expression in a bacterial strain that favors disulfide formation.

2. Protein expression and purification

The region encoding amassin residues 210–495 (NCBI accession No. AAO43562) was inserted into the bacterial expression vector pBH8 (Hillier *et al.*, 1999). This vector codes for the following features: start codon, hexahistidine tag, linker region of amino acids DYDIPTT, tobacco etch viral (TEV) protease cleavage site, the amino acids of interest (preceded by two amino acids GS) and finally stop codons. Upon ligation and subsequent transformation into *E. coli* strain TOP10 (Invitrogen), the entire coding region was verified by DNA sequencing.

For expression, the plasmid was transformed into *E. coli* strain Rosetta-gami B(DE3)pLysS (Novagen) and plated onto LB agar containing carbenicillin, chloramphenicol, tetracycline and kanamycin (100, 34, 12.5 and 15 μg ml^{-1}, respectively). This feature-rich expression strain was chosen mainly owing to the mutations in its thioredoxin reductase and glutathione reductase genes, which greatly enhance the formation of disulfide bonds in the bacterial cytoplasm (Bessette *et al.*, 1999; Prinz *et al.*, 1997). A single fresh transformant colony was used to inoculate a starter culture in LB with the same antibiotics and grown to an OD_{600} of 0.6; it was then stored at 277 K overnight. The starter bacteria were harvested by centrifugation, resuspended in fresh media and used to inoculate five 1 l flasks. On reaching an OD of 0.6, these cultures were induced with 1 m*M* IPTG and incubated at 310 K for 2 h. The bacteria were harvested by centrifugation at 4000*g* for 15 min, resuspended in 160 ml lysis buffer (50 m*M* sodium phosphate pH 8, 500 m*M* NaCl, 10 m*M* imidazole) and frozen at 203 K.

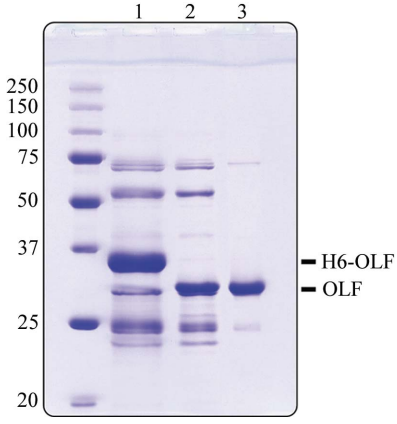


Figure 1
Purification of the bacterially expressed amassin OLF domain. The preparation was fairly impure following elution from the Ni–NTA column (lane 1). The tagged protein ran at a relative mobility of ~35 kDa (band position labeled H6-OLF) and upon tag removal at ~30 kDa (lane 2; band position labeled OLF). Subsequent anion-exchange chromatography improved the purity substantially (lane 3). Proteins were resolved by SDS–PAGE on a 10% gel, followed by staining with Coomassie Blue. Shown on the left are molecular-weight standards labeled in kDa.

Cell suspensions were thawed, lysed by sonication and Triton X-100 was added to 0.5%. The lysate was cleared by centrifugation at 10 000*g*. 1 ml of Ni–NTA Fast Flow (Qiagen) was added to the supernatant and allowed to bind in batch with tumbling for 2 h. The resin was settled and the unbound material removed, then transferred to an empty column. The column was washed with three cycles of 30 ml lysis buffer, followed by 30 ml of lysis buffer with 20 m*M* imidazole and the protein was finally eluted with 10 ml of lysis buffer with 250 m*M* imidazole. The eluate was then dialyzed against cleavage buffer (25 m*M* Tris pH 7.5, 50 m*M* NaCl). The hexahistidine purification tags were removed by incubation with sufficient TEV protease to completely digest within 2 h at 298 K. The protein was then loaded onto a 2 ml HiTrap Q FF anion-exchange column attached to an ÄKTA FPLC system (Amersham Biosciences). An extended elution gradient of 25 m*M* to 1 *M* NaCl over 100 column volumes was necessary to resolve the OLF domain from most contaminants (Fig. 1). Monitoring the elution-profile chromatogram, the core of the protein peak was pooled and dialyzed with three changes against 10 m*M* HEPES pH 7.5. The protein was concentrated by ultrafiltration (Amicon) to 10 mg ml^{-1}, filtered using a 0.22 μm syringe and then frozen in a dry ice–ethanol slurry and stored in aliquots at 203 K. Final yields were typically ~1–3 mg from a 5 l culture.

The amassin OLF domain was expressed to a sufficient level to allow crystallization trials to be performed. However, its purification was not as simple as would be expected for a histidine-tagged protein. Major contaminants were retained and coeluted from the Ni–NTA column even under more stringent conditions, including the presence of reductants, hydrophobic washes and limiting quantities of chromatography resin (data not shown). The anion-exchange chromatography step was able to remove the majority of the contaminants by careful fractionation during an extended elution. Enhanced purity came at a significant loss, as ~30% of the applied protein was lost in the final purification step.

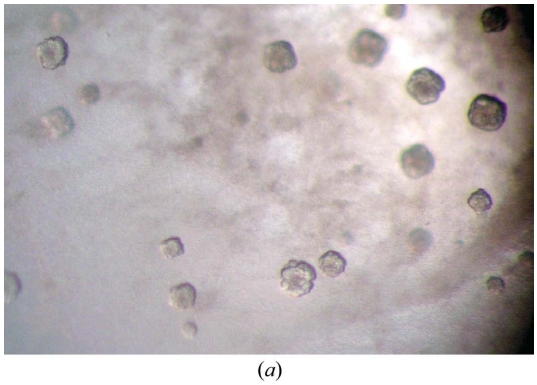
3. Crystallization experiments

Sparse-matrix screening kits Crystal Screens I and II (Hampton Research) were used for all initial crystallization trials using the sitting-drop vapour-diffusion method. 2 μl protein solution was mixed with 2 μl well solution and placed into Cryschem plates (Hampton Research) sealed with tape. The trays were incubated in a laboratory cabinet at ~298 K.

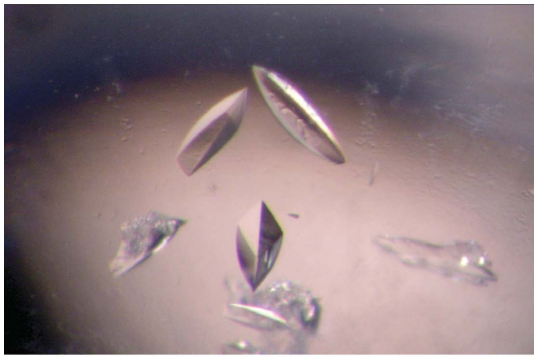
Initial trials used the OLF domain (residues 223–476 and 223–495) of amassin expressed in the eukaryotic host *P. pastoris* (Hillier & Vacquier, submitted). However, no crystals were obtained using these proteins. The proteins were N-glycosylated in the yeast system and it may be that the glycosylation was heterogeneous, inhibiting the formation of crystals. To remedy this, the two N-linked sites were removed by mutating the asparagine residue of the site N-*X*-T/S to aspartate or alanine. These proteins were secreted from *P. pastoris* and the purified proteins indeed appeared to lack glycosylation. This was evidenced by sharp-running bands on SDS–PAGE and no mobility shift upon incubation with PNGaseF (data not shown).

Unfortunately, yields from these mutants were too low to warrant crystallization trials. Additionally, expression in the presence of the glycosylation inhibitor tunicamycin (Sigma) with the wild-type sequence yielded extremely low protein levels. Attempts were also made to deglycosylate the yeast-expressed proteins. However, this was not feasible on the scale required for crystallization because the proteins were fairly resistant, requiring large amounts of enzyme (data not shown).

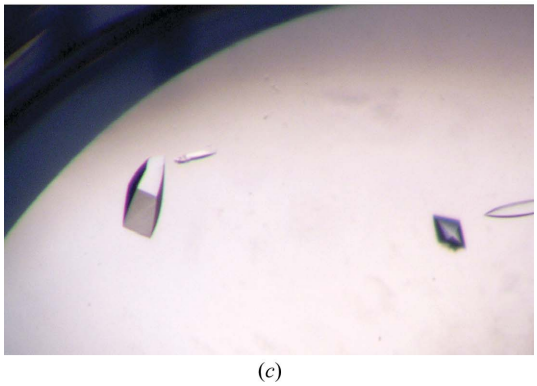
To avoid glycosylation events entirely, an extended residue range (210–495) containing the OLF domain was expressed in the prokaryotic host *Escherichia coli*. The calculated molecular weight after removal of purification tags is 32.8 kDa. Sparse-matrix screens revealed several conditions in which crystals formed within a month. Hampton Screen I condition No. 9 (0.1 *M* sodium citrate pH 5.6, 30% PEG 4000, 0.2 *M* ammonium acetate) was most promising, although its crystals were malformed and walnut-shaped (Fig. 2*a*). A more directed screen surrounding this condition ensued that maintained the buffer composition and substituted PEG 4000 with PEG 3350. We found a marked improvement in crystal appearance coinciding with lower concentrations of ammonium acetate. Additionally, lower PEG concentrations promoted fewer crystal initiating events, leading to larger crystals. The optimized condition was 0.1 *M* sodium citrate pH 5.6 with 20% PEG 3350. After approximately one week, crystals formed from within a light precipitate. At least two gross crystal morphology differences could be observed from within the same drops, one with a tapered appearance and the other more square (Fig. 2*b*). As growth continued, much less precipitant was visible and the size of the crystals gradually increased for 2–3 weeks (Fig. 2*c*).



(*a*)



(*b*)



(*c*)

Figure 2
Walnut-shaped crystals formed in the initial screen with Hampton Crystal Screen condition No. 9 (*a*). Following optimization of conditions, a single drop may yield several crystal morphotypes; the common form was fusiform, while thicker more rectangular forms also were found (*b*). The crystal used for data collection (leftmost crystal) was of the rectangular type with approximate dimensions 0.3 × 0.2 × 0.15 mm (*c*).

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	1.5418
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 66.71, b = 72.28, c = 113.35
Resolution range (Å)	29.9–2.7 (2.80–2.70)
No. of reflections	80424 (7440)
Unique reflections	27497 (2704)
Completeness (%)	94.5 (92.2)
Redundancy	2.92 (2.75)
$R_{\rm merge}$ (%)	10.8 (34.3)
$\langle I/\sigma(I)\rangle$	7.3 (2.9)

4. Data collection and X-ray crystallographic analysis

A cryoprotecting procedure was used which involved first transferring the crystal to a drop containing a slightly higher concentration of precipitant than that used for crystallization, with the addition of glycerol (0.1 *M* sodium citrate pH 5.6, 23% PEG 3350, 15% glycerol). The crystal remained in the cryoprotectant for ∼10 s before removal with a nylon loop and flash-cooling by direct placement into a liquid-nitrogen stream at 110 K. A data set was collected on an in-house R-AXIS IV imaging-plate detector using Cu $K\alpha$ radiation from a Rigaku RU-200 rotating-anode generator equipped with Osmic mirrors. Data were collected, indexed and scaled with the *CrystalClear* (Rigaku) software package. The crystal diffracted to 2.7 Å under this cryoprotectant procedure with 0.3° mosaicity. Systematic absences indicated that the crystal belongs to space group $P2_12_12_1$, with unit-cell parameters a = 66.71, b = 72.28, c = 113.35 Å. Crystal parameters and diffraction statistics are shown in Table 1. Assuming the presence of two OLF domains in the asymmetric unit, the Matthews coefficient is 2.1 Å^3 Da^{-1} and the solvent content is 40.1% (Matthews, 1968).

While no structure has yet been determined of an OLF domain, an effort was made to use the molecular-replacement method using theoretically predicted models. The structure-prediction tool *3D-PSSM* (Kelley *et al.*, 2000; MacCallum *et al.*, 2000), using only the primary sequence as input, predicted by threading onto a library of structures several high-confidence scoring solutions. The top-scoring PDB entries and their confidence scores (E values) were as follows: 1lpx (Springer, 1998) 0.064, 1ndx (Springer, 1998) 0.194 and 1rwi (Good *et al.*, 2004) 0.239. These in turn were used as models, with and without side chains, with the *CCP*4 package component *Phaser* in an attempt to phase the data (Collaborative Computational Project, Number 4, 1994; Read, 2001; Storoni *et al.*, 2004). While a reasonable solution was obtained by this maximum-likelihood technique, the resultant density maps were not sufficiently robust to rebuild or even

properly orient the molecule. This may have been expected, as the highest sequence identity for a search model was merely 16%.

Structure solution of the OLF domain will most likely require the acquisition of experimental phases. The OLF domain expressed and crystallized here contains six methionine residues out of a total of 288 residues. Since it can be expressed in bacteria, the substitution of those residues for selenomethionine is possible, leading the way to structure determination by multiwavelength anomalous diffraction. With the first OLF domain structure in hand, we hope to gain valuable insight into how this domain of emerging importance is able to function and bind its extracellular targets.

We thank Brad Tebo and Hope Johnson for the use of their FPLC and Russell Doolittle for all his advice. This work was supported by the Graduate Department at Scripps Institution of Oceanography and National Institutes of Health grant GM48870 to CDS.

References

Adam, M. F., Belmouden, A., Binisti, P., Brezin, A. P., Valtot, F., Bechetoille, A., Dascotte, J. C., Copin, B., Gomez, L., Chaventre, A., Bach, J. F. & Garchon, H. J. (1997). *Hum. Mol. Genet.* **6**, 2091–2097.

Barembaum, M., Moreno, T. A., LaBonne, C., Sechrist, J. & Bronner-Fraser, M. (2000). *Nature Cell Biol.* **2**, 219–225.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. & Sonnhammer, E. L. (2002). *Nucleic Acids Res.* **30**, 276–280.

Bessette, P. H., Aslund, F., Beckwith, J. & Georgiou, G. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 13703–13708.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Eshed, Y., Feinberg, K., Poliak, S., Sabanay, H., Sarig-Nadir, O., Spiegel, I., Bermingham, J. R. Jr & Peles, E. (2005). *Neuron*, **47**, 215–229.

Good, M. C., Greenstein, A. E., Young, T. A., Ng, H. L. & Alber, T. (2004). *J. Mol. Biol.* **339**, 459–469.

Hillier, B. J., Christopherson, K. S., Prehoda, K. E., Bredt, D. S. & Lim, W. A. (1999). *Science*, **284**, 812–815.

Hillier, B. J. & Vacquier, V. D. (2003). *J. Cell Biol.* **160**, 597–604.

Kelley, L. A., MacCallum, R. M. & Sternberg, M. J. (2000). *J. Mol. Biol.* **299**, 499–520.

Loria, P. M., Hodgkin, J. & Hobert, O. (2004). *J. Neurosci.* **24**, 2191–2201.

MacCallum, R. M., Kelley, L. A. & Sternberg, M. J. (2000). *Bioinformatics*, **16**, 125–129.

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

Moreno, T. A. & Bronner-Fraser, M. (2001). *Dev. Biol.* **240**, 340–360.

Nagy, I., Trexler, M. & Patthy, L. (2003). *Biochem. Biophys. Res. Commun.* **302**, 554–561.

Prinz, W. A., Aslund, F., Holmgren, A. & Beckwith, J. (1997). *J. Biol. Chem.* **272**, 15661–15667.

Ray, K., Mukhopadhyay, A. & Acharya, M. (2003). *Mol. Cell. Biochem.* **253**, 223–231.

Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.

Snyder, D. A., Rivers, A. M., Yokoe, H., Menco, B. P. & Anholt, R. R. (1991). *Biochemistry*, **30**, 9143–9153.

Springer, T. A. (1998). *J. Mol. Biol.* **283**, 837–862.

Stone, E. M., Fingert, J. H., Alward, W. L., Nguyen, T. D., Polansky, J. R., Sunden, S. L., Nishimura, D., Clark, A. F., Nystuen, A., Nichols, B. E., Mackey, D. A., Ritch, R., Kalenak, J. W., Craven, E. R. & Sheffield, V. C. (1997). *Science*, **275**, 668–670.

Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.

Tamm, E. R. & Russell, P. (2001). *J. Glaucoma*, **10**, 329–339.

Torrado, M., Trivedi, R., Zinovieva, R., Karavanova, I. & Tomarev, S. I. (2002). *Hum. Mol. Genet.* **11**, 1291–1301.

Tsuda, H., Sasai, N., Matsuo-Takasaki, M., Sakuragi, M., Murakami, Y. & Sasai, Y. (2002). *Neuron*, **33**, 515–528.

Yokoe, H. & Anholt, R. R. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 4655–4659.