# Preliminary crystallographic analysis of the N-terminal domain of FILIA, a protein essential for embryogenesis

**Juke Wang,[a] Tong-Cun Zhang[a] and Xinqi Liu[b]***

[a]College of Bioengineering, Tianjin University of Science and Technology, Tianjin 300457, People's Republic of China, and [b]College of Life Science, Nankai University, Tianjin 300071, People's Republic of China

Correspondence e-mail: liu2008@nankai.edu.cn

FILIA is a component of the subcortical maternal complex that is essential for early stage embryogenesis. Its $6\times$His-tagged N-terminal domain was expressed in *Escherichia coli* and purified to homogeneity. Two types of crystals formed under different crystallization conditions during screening. Orthorhombic crystals appeared in a solution containing 1.4 $M$ ammonium sulfate, 0.1 $M$ Tris pH 8.2 and 12% glycerol, while tetragonal crystals were obtained using 15% PEG 4000 mixed with 0.1 $M$ HEPES pH 7.5 and 15% 2-propanol. High-quality diffraction data were collected from the two crystal forms to resolutions of 1.8 and 2.2 Å, respectively, using synchrotron radiation. The Matthews coefficients indicated that the $P2_12_12_1$ and $P4_12_12$ crystals contained two molecules and one molecule per asymmetric unit, respectively. A selenomethionine-substituted sample failed to crystallize under the native conditions, but another orthorhombic crystal form was obtained under different conditions and anomalous diffraction data were collected.

## 1. Introduction

Maternal effect genes are transcribed in oogenesis and their encoded proteins persist during early embryonic development following fertilization (Christians *et al.*, 2000; Tong *et al.*, 2000). Several of these proteins form a subcortical maternal complex (SCMC) that is essential for cleavage-stage embryogenesis in mammals. The SCMC contains a core formed by FLOPED, MATER and TLE6 and a few attached peripheral components, including FILIA. FILIA has been shown to interact directly with MATER but not with FLOPED and TLE6 (Li *et al.*, 2008). Although a lack of these genes exerts a serious impact on the growing foetus, their respective roles in embryonic development are still unclear (Tong & Nelson, 1999; Tong *et al.*, 2000, 2004; Li *et al.*, 2008, 2010; Ohsugi *et al.*, 2008; Zheng & Dean, 2009).

Two types of *filia* transcripts with lengths of 1.2 and 1.6 kbp coexist in mice oocytes, but only the former is translated into an active protein with 346 residues. *Filia[tm/tm]* mice, although not completely sterile, show a significant delay in oocyte development during the pre-implantation stage (Ohsugi *et al.*, 2008; Zheng & Dean, 2009). FILIA and MATER colocalize at the apical subcortex of 'outer' but not 'inner' cells (Ohsugi *et al.*, 2008). FILIA is important in the regulation of mitotic kinase activity and spindle-assembly checkpoint activation, ensuring chromosome euploidy and normal segregation in mitotic cells, but the relationship between its particular localization and its unique function still requires further exploration at the molecular level (Zheng & Dean, 2009).

*Filia* belongs to a new eutherian gene family that emerged recently during evolution and has several members, such as *Dppa5*, *floped* and *Khdc1*. These proteins are expressed specifically in oocytes and/or embryonic stem cells and are characterized by an N-terminal atypical K homology (KH) domain (Pierre *et al.*, 2007). The KH domain is involved in RNA binding and processing (Grishin, 2001). The N-terminal region of FILIA has been postulated to form a KH domain based on phylogenetic analysis; it is not easily identifiable from the amino-acid sequence (Pierre *et al.*, 2007).

**1111**
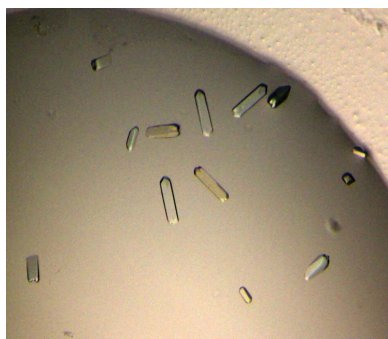
The C-terminus of FILIA has no known function but consists of ten tandem repeats, with each repeat comprising 23 residues (Ohsugi *et al.*, 2008; Li *et al.*, 2010).

In order to study its function on a structural level, we expressed and crystallized the N-terminal domain of FILIA (amino acids 1–124; FILIA-N). Both native and selenomethionine-derivative data sets have been collected for further structure determination.

## 2. Materials and methods

### 2.1. Cloning

The gene encoding full-length mouse *filia* (accession code EDL26288) was provided by Lei Li at the Institute of Zoology, Chinese Academy of Sciences. The primers 5′-TACTTCCAATC-CAATGCCATGGCCTCTCTGAAG-3′ (forward) and 5′-TTATC-CACTTCCAATGCTACTCAACTCCAGCCTC-3′ (reverse) were designed to amplify the sequence coding for amino acids 1–124 (FILIA-N), which was cloned into the pET30-TEV/LIC vector (Novagen) downstream of a 6×His tag by ligation-independent cloning. After 30 cycles of 367 K for 40 s (denaturation), 328 K for 40 s (annealing) and 345 K for 60 s (elongation), the PCR product
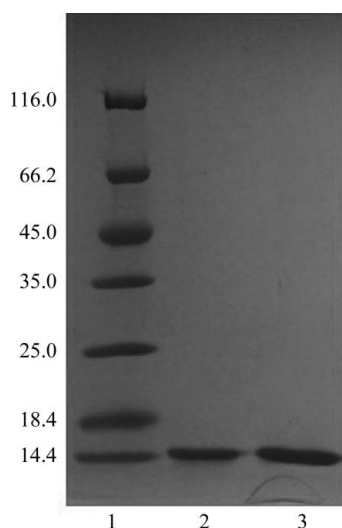


**Figure 1**
SDS–PAGE analysis of FILIA-N after purification. Lane 1, molecular-weight markers (kDa); lane 2, purified FILIA-N protein digested by TEV protease; lane 3, purified selenomethionine-derivatized FILIA-N digested by TEV protease.

was separated on a 1.5% agarose gel and extracted using a Gel Extraction Kit (Tiangen) according to the manufacturer's instructions. The purified PCR product was digested with 0.5 U T4 DNA polymerase (Takara) in a buffer containing 5 m*M* dCTP, 5 m*M* DTT and 1× BSA. After reaction at 295 K for 30 min, the T4 DNA polymerase was inactivated by incubation at 348 K for 20 min.

The pET30-TEV/LIC vector was digested with *Ssp*I and then separated on agarose gel and extracted using a Gel Extraction Kit (Tiangen). The linearized vector was processed with T4 DNA polymerase as described above, except that dGTP was substituted for dCTP. Finally, the vector was annealed with the processed PCR product at 295 K for 5 min, followed by another 5 min in the presence of 5 m*M* EDTA to boost the annealing efficiency.

### 2.2. Expression and purification

The expression vector was transformed into *Escherichia coli* BL21 (DE3). Cells were grown in LB broth medium containing 50 µg ml$^{-1}$ kanamycin at 308 K to an OD$_{600}$ of 0.8; they were then induced with 0.3 m*M* isopropyl $\beta$-D-1-thiogalactopyranoside (IPTG) and cultured for a further 8–12 h at room temperature. The cells were harvested by centrifugation at 6000*g* for 30 min and stored at 243 K until further use. Cell lysis was performed by sonication on ice in Ni–NTA resin binding buffer (50 m*M* Tris, 500 m*M* NaCl, 4 m*M* imidazole pH 8.0). The lysate was clarified by centrifugation at 18 000*g* for 40 min. The supernatant was loaded onto an Ni–NTA resin column (Qiagen) and washed with washing buffer (binding buffer plus 20 m*M* imidazole) to exclude nonspecifically bound contaminants. Finally, 6×His-FILIA-N was eluted in elution buffer (binding buffer plus 500 m*M* imidazole). After concentration using an Amicon Ultra filter (Millipore), the protein was dissolved in low-salt buffer (50 m*M* Tris, 50 m*M* NaCl pH 8.0). The 6×His-tagged protein was then digested by TEV protease at 289 K for 24 h and the cleaved protein was purified on a HiTrap Q HP column (GE Healthcare) using a linear NaCl gradient from 50 to 800 m*M*. The protein peak was pooled and concentrated and loaded onto a HiLoad 16/60 Superdex-200 size-exclusion column (GE Healthcare) for further purification. A sharp peak corresponding to the target protein was pooled and concentrated to 15 mg ml$^{-1}$. On an SDS–PAGE gel, the purified protein had a purity of greater than 95% (Fig. 1).

For the expression of selenomethionine-derivatized protein, the pET30-FILIA-N vector was transformed into the methionine-auxotroph strain B834 (DE3). Cells were cultured with minimal medium M9 supplemented with 50 µg ml$^{-1}$ L-methionine to an OD$_{600}$ of 1.0. The cells were then transferred to medium without
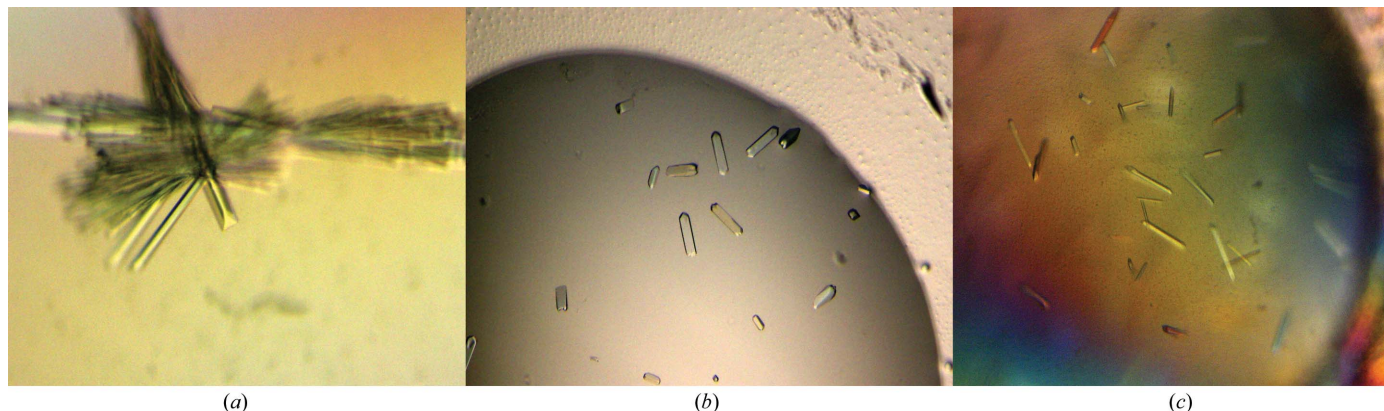


**Figure 2**
Crystals of FILIA-N. (*a*) Crystals grown in 1.4 *M* ammonium sulfate, 0.1 *M* Tris pH 8.2, 12%(*v/v*) glycerol. (*b*) Crystals grown in 0.1 *M* Na HEPES pH 7.5, 15%(*v/v*) 2-propanol, 15%(*w/v*) PEG 4000. (*c*) Selenomethionine-derivative crystals.

**Table 1**
Data-collection and processing statistics.

Values in parentheses are for the outer resolution shell.

| Data sets | Native, orthorhombic | Native, tetragonal | SeMet derivative |
|---|---|---|---|
| X-ray source | KEK synchrotron, Japan | KEK synchrotron, Japan | SSRF synchrotron, China |
| Wavelength (Å) | 0.9798 | 0.9798 | 0.9795 |
| Detector | ADSC Quantum 270 | ADSC Quantum 270 | MAR CCD 245 |
| Crystal parameters | | | |
| Space group | $P2_12_12_1$ | $P4_12_12$ | $P2_12_12_1$ |
| Unit-cell parameters | | | |
| $a$ (Å) | 38.30 | 61.61 | 56.54 |
| $b$ (Å) | 73.96 | 61.61 | 59.16 |
| $c$ (Å) | 89.96 | 83.04 | 89.62 |
| $\alpha$ (°) | 90 | 90 | 90 |
| $\beta$ (°) | 90 | 90 | 90 |
| $\gamma$ (°) | 90 | 90 | 90 |
| Resolution (Å) | 1.8 | 2.2 | 2.3 |
| Mosaicity (°) | 0.45 | 0.74 | 0.83 |
| Solvent content (%) | 46.17 | 56.48 | 54.24 |
| Molecules per asymmetric unit | 2 | 1 | 2 |
| Data processing | | | |
| No. of observed reflections | 118062 (4132) | 48785 (2892) | 67764 (2694) |
| No. of unique reflections | 20859 (1011) | 8353 (638) | 11624 (824) |
| Completeness (%) | 82.4 (41.3) | 93.2 (82.2) | 89.6 (69.7) |
| Redundancy | 4.4 (0.6) | 6.7 (3.8) | 5.4 (1.9) |
| $R_{merge}$ (%) | 9.2 (62.4) | 11.4 (52.3) | 7.4 (29.9) |
| Average $I/\sigma(I)$ | 16.2 (2.7) | 19.5 (2.1) | 17.4 (2.6) |

L-methionine and cultured for 6 h to deplete the residual methionine. After depletion of L-methionine, 50 µg ml$^{-1}$ selenomethionine was provided and 0.2 m$M$ IPTG was used to induce protein expression. The isolation and purification procedures were the same as those for the native protein.

## 2.3. Crystallization

Crystals of FILIA-N were grown using the hanging-drop vapour-diffusion method at 277 K. Initial screening using Index, Crystal Screen and Crystal Screen 2 (Hampton Research) produced two promising crystallization conditions. Needle-like crystals appeared in Crystal Screen 2 condition No. 42 [1.5 $M$ ammonium sulfate, 0.1 $M$ Tris pH 8.5, 12%($v/v$) glycerol], whereas rod-like single crystals were obtained a few days later in Crystal Screen condition No. 41 [0.1 $M$ Na HEPES pH 7.5, 10%($v/v$) 2-propanol, 20%($w/v$) PEG 4000]. Initial crystallization conditions were optimized by changing the concentrations of the precipitants and the pH and by the use of additives. Finally, a 2 µl droplet of protein solution (15 mg ml$^{-1}$) mixed with an equal amount of reservoir solution was equilibrated against 500 µl reservoir solution [1.4 $M$ ammonium sulfate, 0.1 $M$ Tris 8.2, 12%($v/v$) glycerol for the first condition and 0.1 $M$ Na HEPES pH 7.5, 15%($v/v$) 2-propanol, 15%($w/v$) PEG 4000 for the second] to yield FILIA-N crystals that were suitable for data collection (Fig. 2). Surprisingly, while the native protein crystallized without much difficulty, the selenomethionine-derivatized protein was refractory to crystallization under either of the above crystallization conditions. After screening with the derivatized protein, we obtained a new crystallization condition: condition No. 7 from the JCSG Plus kit [Molecular Dimensions; 0.2 $M$ ammonium formate, 20%($w/v$) PEG 3350]. After a slight optimization, single crystals for anomalous data collection were grown from 0.2 $M$ ammonium formate, 16%($w/v$) PEG 3350.

## 2.4. Data collection

Data collection from native crystals was performed at 100 K using a wavelength of 0.9798 Å at the Photon Factory (KEK), Tsukuba, Japan. Selenomethionine-derivative data were collected at the Shanghai Synchrotron Radiation Facility (SSRF), Shanghai, People's Republic of China. For data collection under cryogenic conditions, crystals were soaked for a few seconds in mother liquor containing 20%($v/v$) glycerol. Crystals were then mounted on the beamline in a nylon loop and flash-cooled in a nitrogen-gas stream at 100 K. Native data sets were collected by rotation through 180° with an increment of 1° per frame at a crystal-to-detector distance of 300 mm. For the derivative crystal, only the peak absorption data set was collected at a wavelength of 0.9795 Å over 180° at a crystal-to-detector distance of 280 mm. All data sets were processed using the *HKL*-2000 package (Otwinowski & Minor, 1997). The diffraction amplitudes were converted from the intensities using the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) for further use in structure determination. Data-collection and processing results are summarized in Table 1.

## 3. Results and discussion

FILIA is a recently identified protein that is specifically expressed in the growing oocytes of adult animals. It participates in the subcortical maternal complex and is colocalized with MATER. The N-terminus of FILIA (FILIA-N) is assumed to form a relatively independent domain based on secondary-structure prediction. Native FILIA-N protein could be readily crystallized in two different space groups, $P2_12_12_1$ and $P4_12_12$, which contained two molecules and one molecule per asymmetric unit, respectively, based on solvent-content calculations (Matthews, 1968). The screw-axis symmetry was elucidated from systematic absences in the data sets. For the tetragonal crystal form, space group $P4_12_12$ was discriminated from $P4_32_12$ during subsequent molecular replacement using the model built from the seleno-methionine-derivative data. Interestingly, the selenomethionine-derivatized protein was resistant to crystallization under either of the crystallization conditions used for the native protein, but a completely different crystallization condition was obtained in a new screen with selenomethionine-derivatized protein. Since FILIA-N contains six methionines in 124 amino acids, it is possible that these seleno-methionine substitutions resulted in this different crystal-growth behaviour. The selenomethionine-derivative crystal also has a distinct

crystal packing, although one axis (axis *c*) is nearly identical to those of the other crystal forms. Owing to the high data quality, we could find the positions of nine Se atoms from the 12 selenomethionines contained in the asymmetric unit using only the peak-wavelength derivative data set. A preliminary model was readily built from the single-wavelength anomalous diffraction (SAD) phased density using the *PHENIX* package (Adams *et al.*, 2010). Therefore, we stopped further data collection at the inflection and remote wavelengths. Approximately 70% of the structure, including two helices and six β-strands within a dimer, were discernable at the end of autobuilding. Further structural refinement is ongoing.

Further structure determination of FILIA-N will demonstrate whether FILIA is a KH-domain-containing protein. It will provide insight into the function of FILIA in RNA processing and contribute to the understanding of the underlying molecular mechanism of the phenotype of *filia* knockout mice.

## References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.

Christians, E., Davis, A. A., Thomas, S. D. & Benjamin, I. J. (2000). *Nature (London)*, **407**, 693–694.

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

Grishin, N. V. (2001). *Nucleic Acids Res.* **29**, 638–643.

Li, L., Baibakov, B. & Dean, J. (2008). *Dev. Cell*, **15**, 416–425.

Li, L., Zheng, P. & Dean, J. (2010). *Development*, **137**, 859–870.

Matthews, B. W. (1968). *J. Mol. Biol.* **28**, 491–497.

Ohsugi, M., Zheng, P., Baibakov, B., Li, L. & Dean, J. (2008). *Development*, **135**, 259–269.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pierre, A., Gautier, M., Callebaut, I., Bontoux, M., Jeanpierre, E., Pontarotti, P. & Monget, P. (2007). *Genomics*, **90**, 583–594.

Tong, Z.-B., Gold, L., De Pol, A., Vanevski, K., Dorward, H., Sena, P., Palumbo, C., Bondy, C. A. & Nelson, L. M. (2004). *Endocrinology*, **145**, 1427–1434.

Tong, Z.-B., Gold, L., Pfeifer, K. E., Dorward, H., Lee, E., Bondy, C. A., Dean, J. & Nelson, L. M. (2000). *Nature Genet.* **26**, 267–268.

Tong, Z.-B. & Nelson, L. M. (1999). *Endocrinology*, **140**, 3720–3726.

Zheng, P. & Dean, J. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 7473–7478.