Samira Acajjaoui and Chloe
Zubieta*

Structural Biology, European Synchrotron
Radiation Facility, 6 Rue Jules Horowitz,
38000 Grenoble, France

Correspondence e-mail: zubieta@esrf.fr

Received 24 April 2013

Accepted 19 July 2013

Crystallization studies of the keratin-like domain from *Arabidopsis thaliana* SEPALLATA 3

In higher plants, the MADS-box genes encode a large family of transcription factors (TFs) involved in key developmental processes, most notably plant reproduction, flowering and floral organ development. SEPALLATA 3 (SEP3) is a member of the MADS TF family and plays a role in the development of the floral organs through the formation of multiprotein complexes with other MADS-family TFs. SEP3 is divided into four domains: the M (MADS) domain, involved in DNA binding and dimerization, the I (intervening) domain, a short domain involved in dimerization, the K (keratin-like) domain important for multimeric MADS complex formation and the C (C-terminal) domain, a largely unstructured region putatively important for higher-order complex formation. The entire K domain along with a portion of the I and C domains of SEP3 was crystallized using high-throughput robotic screening followed by optimization. The crystals belonged to space group $P2_12_12$, with unit-cell parameters $a = 123.44$, $b = 143.07$, $c = 49.83$ Å, and a complete data set was collected to 2.53 Å resolution.

1. Introduction

The astonishing diversity and complexity of higher organisms can be ascribed to the advent of novel developmental processes and alterations in existing developmental pathways (Theissen *et al.*, 2000). Evolution is often the result of changes in a relatively small set of developmental control genes and the activity of the transcription factors (TFs) that they encode. In plants, the MADS-box family acts as a primary set of developmental control genes for such fundamental developmental processes as seed germination (Chiang *et al.*, 2009; Becker *et al.*, 2000; Liljegren *et al.*, 2000; Bemer *et al.*, 2010), root development (Burgeff *et al.*, 2002), fruit development (Gu *et al.*, 1998; Prasad & Ambrose, 2010; Vrebalov *et al.*, 2002) and, perhaps most notably, flowering and floral organ development (Alvarez-Buylla *et al.*, 2000; Rounsley *et al.*, 1995; Ng & Yanofsky, 2001). Named for MCM1 from *Saccharomyces cerevisiae*, AGAMOUS from *Arabidopsis thaliana*, DEFICIENS from *Antirrhinum majus* and SRF from *Homo sapiens*, the 'MADS-box' motif encodes a ~60-amino-acid DNA-binding domain. While present in fungi and animals, the MADS-box family has undergone extensive duplication and divergence in plants and this increase in the number and diversity of MADS-box genes is thought to be a critical component of the evolutionary leap from nonflowering (gymnosperm) to flowering (angiosperm) plants (Theissen *et al.*, 2000; Becker & Theissen, 2003; Münster *et al.*, 1997).

SEPALLATA 3 (SEP3) is a MADS TF and has been implicated in proper development of the floral organs (Favaro *et al.*, 2003; Pelaz *et al.*, 2000; Immink *et al.*, 2009). As in all MADS TFs implicated in flower development, SEP3 is a member of the MIKC type of MADS TFs. These proteins comprise a highly conserved M (MADS) domain responsible for DNA binding and dimerization, an I (intervening) domain, a short stretch of ~30 amino acids that is likely to be important for dimerization, a K (keratin-like) domain, a predicted coiled-coil domain putatively involved in MADS TF tetramer formation and specificity, and a highly variable C (C-terminal)

domain that is likely to be involved in higher-order complex formation and transcriptional activation (Kaufmann *et al.*, 2005; Kwantes *et al.*, 2012). Isolated MADS domains have been crystallized from human and yeast (He *et al.*, 2011; Mo *et al.*, 2001; Tan & Richmond, 1998; Wu *et al.*, 2010); however, there are currently no structural data for any plant MADS TF and no structural data for the nonhomologous I, K or C domains. In order to investigate the molecular and atomic-level determinants for SEP3 oligomerization, we expressed and purified the complete K domain with portions of the I and C domains.

In this report, we describe the expression, purification and crystallization of a SEP3 construct (SEP3_{74–175}) consisting of amino acids 74–175 from *A. thaliana*. Using high-throughput robotic screening, the protein was crystallized in multiple conditions, yielding small crystals suitable for preliminary X-ray diffraction analysis. This work represents an initial step towards the first structural characterization of a plant MADS TF.

2. Experimental methods

2.1. Strains and plasmids

SEP3_{74–175} (AT1G24260, ecotype Columbia) was cloned into the expression vector pESPRIT2 (Hart & Tarendeau, 2006; Guilligay *et al.*, 2008) using the *Aat*II and *Not*I sites. The plasmid contains an N-terminal six-His tag followed by a TEV protease cleavage site. His-tagged SEP3_{74–175} was overproduced in *Escherichia coli* BL21 (DE3) CodonPlus RIL (Agilent Technologies).

2.2. Expression and purification

Cells were grown in LB medium in the presence of 50 µg ml^{−1} kanamycin and 35 µg ml^{−1} chloramphenicol at 310 K and 180 rev min^{−1}. At an OD_{600 nm} of 0.8 the temperature was lowered to 293 K and protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside. After 16 h, the cells were harvested by centrifugation at 6000 rev min^{−1} and 277 K for 15 min. The cells were resuspended in 30 mM Tris pH 8.0, 300 mM NaCl, 5 mM β-mercaptoethanol (BME), 5% (v/v) glycerol, 20% (w/v) sucrose. Cells were lysed by sonication and the cell debris was pelleted at 25 000 rev min^{−1} and 277 K for 30 min. The supernatant containing SEP3_{74–175} was applied onto a 5 ml Ni-NTA column pre-equilibrated with 30 mM Tris pH 8.0, 300 mM NaCl, 5 mM BME, 5% glycerol. The bound protein was washed with 30 mM Tris pH 8.0, 1 M NaCl, 5 mM BME, 5% glycerol and eluted with 30 mM Tris pH 8.0, 300 mM NaCl, 5 mM BME, 5% glycerol, 250 mM imidazole. The fractions of interest were pooled and dialysed against 30 mM Tris pH 8.0, 300 mM NaCl, 5 mM BME, 5% glycerol in the presence of TEV protease overnight at 277 K to remove the polyhistidine tag. After removal of the TEV protease and uncut protein on the same Ni-NTA column, the cleaved protein was concentrated to approximately 5 mg ml^{−1} and applied onto a size-exclusion FPLC column (Superdex 200 10/300 GL, GE Healthcare) pre-equilibrated with 30 mM Tris pH 8.0, 300 mM NaCl, 5 mM BME, 5% glycerol. The fractions of interest were pooled and concentrated to 12 mg ml^{−1} for subsequent crystallization studies.

2.3. Multi-angle laser light scattering (MALLS)

50 µl SEP3_{74–175} at a concentration of 5 mg ml^{−1} was loaded onto an S200 size-exclusion column (Superdex 200 10/300 GL, GE Healthcare) at a flow rate of 0.5 ml min^{−1}. The column was pre-equilibrated with 30 mM Tris pH 8.0, 300 mM NaCl, 5 mM BME, 5% glycerol and connected to a multi-angle laser light-scattering

detector (DAWN HELEOS II, Wyatt Technology Corporation) and a refractive-index detector (Optilab T-REX, Wyatt Technology Corporation). The data were processed with the ASTRA 6.0 software (Wyatt Technology Corporation). A theoretical molecular weight of 12.5 kDa for the monomer was later used as a reference for calculation of the oligomeric state.

2.4. Protein crystallization

Freshly purified SEP3_{74–175} at a concentration of ~10 mg ml^{−1} in 30 mM Tris pH 8.0, 300 mM NaCl, 5 mM BME, 5% glycerol was subjected to high-throughput crystallization trials using the EMBL HTX facility (European Molecular Biology Laboratory, Grenoble, France). Excess protein was stored at 193 K for future use with no apparent impact on crystal quality arising from the storage conditions. Crystallization trials were performed in 200 nl sitting drops using a Cartesian PixSys 4200 crystallization robot (Genomic Solutions, UK) with Greiner CrystalQuick plates (flat bottom, untreated) and imaged with a Formulatrix Rock Imager (Formulatrix Inc., USA) at 277 K (Dimasi *et al.*, 2007). Commercial crystal screens from Hampton Research (Aliso Viejo, California, USA) were used in all robotic screening trials (Crystal Screen, Crystal Screen 2, Crystal Screen Lite, PEG/Ion, MembFac, Natrix, Quik Screen, Grid Screen Ammonium Sulfate, Grid Screen Sodium Malonate, Grid Screen Sodium Formate, Grid Screen PEG 6K, Grid Screen PEG/LiCl, Grid Screen MPD and Index). Crystals were obtained in (i) 0.1 M Tris pH 8.0, 20% ethanol, (ii) 0.1 M Tris pH 8.0, 20% (±)-2-methyl 2,4-pentanediol (MPD), (iii) 0.1 M Tris pH 8.0, 20% ethylene glycol and (iv) 0.1 M sodium/potassium phosphate pH 6.2, 25% 1,2-propanediol, 10% glycerol. Crystals grew to dimensions of ~20 × 20 × 10 µm as clusters over one month. Crystals were harvested and cryocooled in liquid nitrogen without prior cryoprotection for initial screening.

2.5. Crystal optimization

The best diffracting crystals from the robotic screenings were optimized *via* additive screening using commercially available kits (Additive Screen and Detergent Screen, Hampton Research), fine grid screening of pH and precipitant concentrations, and seeding techniques. Successive rounds of macroseeding yielded the best diffracting single crystals in 0.1 M Tris pH 8.0, 25% ethylene glycol. Large crystal clusters were broken manually and the pieces placed in 2 µl hanging drops consisting of a 1:1 ratio of protein to well solution (25% ethylene glycol, 0.1 M Tris pH 8.0). After 24–48 h the drops were examined and small single crystals were harvested and placed in fresh 2 µl hanging drops, yielding large single crystals. Crystals were harvested and cryocooled in liquid nitrogen for data collection.

2.6. Diffraction data collection and processing

A native diffraction data set from a crystal grown in 0.1 M Tris pH 8.0, 25% ethylene glycol was collected at 100 K on beamline ID14-4 of the ESRF, Grenoble, France. The crystal belonged to space group *P*₂₁₂₁₂ and data were collected at a wavelength of 0.9393 Å. Indexing was performed using *EDNA* (Incardona *et al.*, 2009) and the default optimized oscillation range and collection parameters were used for data collection. The data set was integrated and scaled using the programs *XDS* and *XSCALE* (Kabsch, 2010).

3. Results and discussion

Full-length SEP3 proved to be recalcitrant to soluble recombinant expression. Based on sequence alignments, secondary-structure

predictions and brute-force clone-library screening (Hart & Tarendeau, 2006; Hart & Waldo, 2013; Yumerefendi *et al.*, 2010, 2011), a soluble construct consisting of residues 74–175 spanning a portion of the I domain, the entire K domain and ~18 residues of the C domain was targeted for crystallization (Fig. 1). In order to investigate the oligomeric state of SEP3_{74–175}, the protein was purified *via* affinity chromatography, the polyhistidine tag was cleaved and a final size-exclusion chromatography (SEC) step was performed to obtain a homogeneous sample. SEP3_{74–175} was analysed by analytical SEC–MALLS and gave a monomodal particle-size distribution with a molar mass of 35 460 g mol^{−1} (±1.2%), corresponding to ~2.8 times the molecular weight of the monomer (12 516 Da). Interestingly, the K domain has been described as a tetramerization domain (Honma & Goto, 2001; Theissen & Saedler, 2001; Melzer *et al.*, 2009; Melzer & Theissen, 2009; Kaufmann *et al.*, 2005); however, this construct is predominantly a trimer in solution based on SEC–MALLS experiments. Whether this trimerization is important in the context of the full-length protein under physiological conditions remains to be determined.

High-throughput screening identified multiple conditions which resulted in diffracting crystals of SEP3_{74–175}. Crystals grew in 200 nl sitting drops at 277 K and appeared after one month. Conditions that yielded diffracting crystals contained high concentrations of MPD, propanediol, ethylene glycol or ethanol with pH varying from 6.2 to 8.0. As the protein is small, stable and highly soluble, high precipitant concentrations are likely to be needed for crystal growth, with alcohols clearly favoured based on the screening of over 600 conditions. All initial crystals from the robotic screening were cooled without additional cryoprotectant and were screened for diffraction using a microfocus (10 µm) beam owing to the small size (10–30 µm) and the clustered growth of the crystals. Every crystal tested diffracted, with diffraction ranging from 15 to 3.2 Å resolution. Conditions with high ethylene glycol concentrations routinely gave the best diffraction. Additive screening did not improve the crystal quality; however, extensive macroseeding with the crystal clusters grown in 0.1 M Tris pH 8.0, 20% ethylene glycol yielded single crystals which diffracted to 2.53 Å resolution (Fig. 2 and Table 1).

In order to determine the number of SEP3_{74–175} molecules in the asymmetric unit, the Matthews coefficients (V_M) were calculated (Matthews, 1968), with six being the most probable number of molecules per asymmetric unit. The probabilities were calculated

Table 1

Crystallographic statistics for SEP3_{74–175}.

Values in parentheses are for the outermost shell.

Space group	$P2_12_12$
Unit-cell parameters (Å)	$a = 123.44$, $b = 143.07$, $c = 49.83$
Data collection	
Beamline	ID14-4, ESRF
Wavelength (Å)	1.07
Resolution (Å)	47.0–2.53 (2.60–2.53)
Reflections (total/unique)	146581/30066
Completeness (%)	99.1 (99.3)
$\langle I/\sigma(I) \rangle$	14.8 (1.73)
$R_{\text{merge}}^{\dagger}$ (%)	6.4 (90.2)

$$\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}.$$

Table 2

Matthews coefficient based on trimeric SEP3_{74–175} as the molecular unit.

No. of molecules	Probability	V_M (Å ³ Da ^{−1})	V_s (%)	MW (Da)
3	0.0065	5.91	79.19	37683
6	0.6261	2.96	58.38	75366
9	0.3624	1.97	37.57	113049
12	0.0050	1.48	16.76	150732

according to Kantardjieff & Rupp (2003); four possible solutions were found with a trimer as the basic multimeric unit (Table 2). Based on the total molecular weight of 37.5 kDa for the trimer and the space group and the unit-cell parameters, V_M calculations indicated that the asymmetric unit theoretically contains between three and 12 monomers, with solvent contents ranging from 79 to 17%. Owing to the relatively modest scattering and probability distributions for the number of molecules, the asymmetric unit is likely to contain six to nine molecules, corresponding to two to three trimers.

Based on secondary-structure prediction (Buchan *et al.*, 2010; Jones, 1999), the K domain of SEP3 forms a coiled coil (Pfam 01486) and is implicated in multimer formation. Currently, there is no close structural homologue available in the PDB. Using *BLAST* (*Basic Local Alignment Search Tool*), proteins in the PDB were ranked, with the highest scoring hits giving E values of 0.5, indicative of very weak sequence similarity. The results showed some homology to proteins containing helical domains, with no hits for coiled-coil structures. Thus, it is unlikely that a good search model is available for phasing *via* molecular replacement. We are therefore in the process

Figure 1

Sequence of SEPALLATA 3 with the predicted secondary structure. Helices are shown in red, β -strands in blue and loops in black. The protein is colour-coded by domain with M in green, I in yellow, K in blue and C in purple.

Figure 2

Crystals of SEP3_{74–175}. The crystals were grown from 0.1 M Tris pH 8.0, 25% ethylene glycol.

of producing selenomethionine-derivatized protein. In addition, we are currently working on preparing native crystals for heavy-atom derivatization *via* traditional soaking methods for use in multiple and/or single isomorphous replacement and multi/single-wavelength anomalous dispersion methods.

The authors would like to acknowledge the EMBL HTX facility for crystallization and the EMBL ESPRIT platform for construct-library construction and screening.

References

- Alvarez-Buylla, E. R., Liljegen, S. J., Pelaz, S., Gold, S. E., Burgeff, C., Ditta, G. S., Vergara-Silva, F. & Yanofsky, M. F. (2000). *Plant J.* **24**, 457–466.
- Becker, A. & Theissen, G. (2003). *Mol. Phylogenet. Evol.* **29**, 464–489.
- Becker, A., Winter, K.-U., Meyer, B., Saedler, H. & Theissen, G. (2000). *Mol. Biol. Evol.* **17**, 1425–1434.
- Bemer, M., Heijmans, K., Airoidi, C., Davies, B. & Angenent, G. C. (2010). *Plant Physiol.* **154**, 287–300.
- Buchan, D. W., Ward, S. M., Lobley, A. E., Nugent, T. C., Bryson, K. & Jones, D. T. (2010). *Nucleic Acids Res.* **38**, W563–W568.
- Burgeff, C., Liljegen, S. J., Tapia-López, R., Yanofsky, M. F. & Alvarez-Buylla, E. R. (2002). *Planta*, **214**, 365–372.
- Chiang, G. C. K., Barua, D., Kramer, E. M., Amasino, R. M. & Donohue, K. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 11661–11666.
- Dimasi, N., Flot, D., Dupeux, F. & Márquez, J. A. (2007). *Acta Cryst.* **F63**, 204–208.
- Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M. F., Kater, M. M. & Colombo, L. (2003). *Plant Cell*, **15**, 2603–2611.
- Gu, Q., Ferrándiz, C., Yanofsky, M. F. & Martienssen, R. (1998). *Development*, **125**, 1509–1517.
- Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R. W., Ortin, J., Hart, D. J. & Cusack, S. (2008). *Nature Struct. Mol. Biol.* **15**, 500–506.
- Hart, D. J. & Tarendeau, F. (2006). *Acta Cryst.* **D62**, 19–26.
- Hart, D. J. & Waldo, G. S. (2013). *Curr. Opin. Struct. Biol.* **23**, 403–408.
- He, J., Ye, J., Cai, Y., Riquelme, C., Liu, J. O., Liu, X., Han, A. & Chen, L. (2011). *Nucleic Acids Res.* **39**, 4464–4474.
- Honma, T. & Goto, K. (2001). *Nature (London)*, **409**, 525–529.
- Immink, R. G., Tonaco, I. A., de Folter, S., Shchennikova, A., van Dijk, A. D., Busscher-Lange, J., Borst, J. W. & Angenent, G. C. (2009). *Genome Biol.* **10**, R24.
- Incardona, M.-F., Bourenkov, G. P., Levik, K., Pieritz, R. A., Popov, A. N. & Svensson, O. (2009). *J. Synchrotron Rad.* **16**, 872–879.
- Jones, D. T. (1999). *J. Mol. Biol.* **292**, 195–202.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
- Kaufmann, K., Melzer, R. & Theissen, G. (2005). *Gene*, **347**, 183–198.
- Kwantes, M., Liebsch, D. & Verelst, W. (2012). *Mol. Biol. Evol.* **29**, 293–302.
- Liljegen, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowman, J. L. & Yanofsky, M. F. (2000). *Nature (London)*, **404**, 766–770.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Melzer, R. & Theissen, G. (2009). *Nucleic Acids Res.* **37**, 2723–2736.
- Melzer, R., Verelst, W. & Theissen, G. (2009). *Nucleic Acids Res.* **37**, 144–157.
- Mo, Y., Ho, W., Johnston, K. & Marmorstein, R. (2001). *J. Mol. Biol.* **314**, 495–506.
- Münster, T., Pahnke, J., Di Rosa, A., Kim, J. T., Martin, W., Saedler, H. & Theissen, G. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 2415–2420.
- Ng, M. & Yanofsky, M. F. (2001). *Nature Rev. Genet.* **2**, 186–195.
- Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E. & Yanofsky, M. F. (2000). *Nature (London)*, **405**, 200–203.
- Prasad, K. & Ambrose, B. A. (2010). *Plant Signal. Behav.* **5**, 899–902.
- Rounsley, S. D., Ditta, G. S. & Yanofsky, M. F. (1995). *Plant Cell*, **7**, 1259–1269.
- Tan, S. & Richmond, T. J. (1998). *Nature (London)*, **391**, 660–666.
- Theissen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J. T., Münster, T., Winter, K.-U. & Saedler, H. (2000). *Plant Mol. Biol.* **42**, 115–149.
- Theissen, G. & Saedler, H. (2001). *Nature (London)*, **409**, 469–471.
- Vrebalov, J., Ruezinsky, D., Padmanabhan, V., White, R., Medrano, D., Drake, R., Schuch, W. & Giovannoni, J. (2002). *Science*, **296**, 343–346.
- Wu, Y., Dey, R., Han, A., Jayatilaka, N., Phillips, M., Ye, J. & Chen, L. (2010). *J. Mol. Biol.* **397**, 520–533.
- Yumerefendi, H., Desravines, D. C. & Hart, D. J. (2011). *Methods*, **55**, 38–43.
- Yumerefendi, H., Tarendeau, F., Mas, P. J. & Hart, D. J. (2010). *J. Struct. Biol.* **172**, 66–74.