



The *BAM7* gene in *Zea mays* encodes a protein with similar structural and catalytic properties to *Arabidopsis* BAM2

Claire M. Ravenburg,^a McKayla B. Riney,^b Jonathan D. Monroe^a and Christopher E. Berndsen^{b*}

Received 3 November 2021

Accepted 23 February 2022

Edited by M. Czjzek, Station Biologique de Roscoff, France

Keywords: β -amylases; *Zea mays*; *Arabidopsis thaliana*; small-angle X-ray scattering; gene structure; gene evolution; BAM genes; starch metabolism.

SASBDB references: 1bBAM5, SASDM26; ZmBAM7-S, SASBMX5; AtBAM1, SASDMZ5

Supporting information: this article has supporting information at journals.iucr.org/d

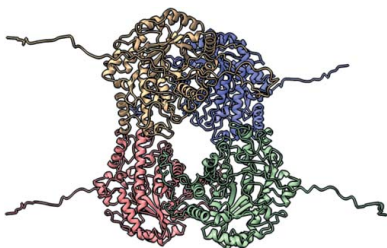
^aDepartment of Biology, James Madison University, 951 Carrier Drive, MSC 7801, Harrisonburg, VA 22807, USA, and ^bDepartment of Chemistry and Biochemistry, James Madison University, 901 Carrier Drive, MSC 4501, Harrisonburg, VA 22807, USA. *Correspondence e-mail: berndsce@jmu.edu

Starch accumulates in the plastids of green plant tissues during the day to provide carbon for metabolism at night. Starch hydrolysis is catalyzed by members of the β -amylase (BAM) family, which in *Arabidopsis thaliana* (At) includes nine structurally and functionally diverse members. One of these enzymes, AtBAM2, is a plastid-localized enzyme that is unique among characterized β -amylases since it is tetrameric and exhibits sigmoidal kinetics. Sequence alignments show that the BAM domains of AtBAM7, a catalytically inactive, nuclear-localized transcription factor with an N-terminal DNA-binding domain, and AtBAM2 are more closely related to each other than they are to any other AtBAM. Since the *BAM2* gene is found in more ancient lineages, it was hypothesized that the *BAM7* gene evolved from *BAM2*. However, analysis of the genomes of 48 flowering plants revealed 12 species that appear to possess a *BAM7* gene but lack a *BAM2* gene. Upon closer inspection, these BAM7 proteins have a greater percent identity to AtBAM2 than to AtBAM7, and they share all of the AtBAM2 functional residues that BAM7 proteins normally lack. It is hypothesized that these genes may encode BAM2-like proteins although they are currently annotated as BAM7-like genes. To test this hypothesis, a cDNA for the short form of corn BAM7 (ZmBAM7-S) was designed for expression in *Escherichia coli*. Small-angle X-ray scattering data indicate that ZmBAM7-S has a tetrameric solution structure that is more similar to that of AtBAM2 than to that of AtBAM1. In addition, partially purified ZmBAM7-S is catalytically active and exhibits sigmoidal kinetics. Together, these data suggest that some *BAM7* genes may encode a functional BAM2. Exploring and understanding the β -amylase gene structure could have an impact on the current annotation of genes.

1. Introduction

In most plants, starch provides the carbon and energy necessary to sustain metabolism at night, when photosynthesis is inactive, or after a long period of dormancy (Zeeman *et al.*, 2010). One group of plant proteins involved in starch metabolism is the β -amylase (BAM) family (Monroe & Storm, 2018; Thalmann *et al.*, 2019). BAM enzymes catalyze the hydrolysis of α -1,4-glycosidic bonds in starch, which releases maltose (Zeeman *et al.*, 2010). In *Arabidopsis thaliana* (At) there are nine members of the BAM family, all of which are encoded by separate genes, have conserved BAM domains and include an N-terminal variable region that is involved in localization of the proteins (Monroe & Storm, 2018; Thalmann *et al.*, 2019).

Five of the *Arabidopsis* BAMs are catalytically active on starch or dextrans (BAM1, BAM2, BAM3, BAM5 and BAM6), while the other four (BAM4, BAM7, BAM8 and BAM9) are



noncatalytic towards these substrates (Monroe & Storm, 2018). The majority of the BAMs in *Arabidopsis* are thought to function as monomers, while some, such as BAM7 and BAM8, are predicted to form dimeric complexes with themselves or each other (Sparla *et al.*, 2006; Reinhold *et al.*, 2011; Soyk *et al.*, 2014; Monroe & Storm, 2018). Additionally, AtBAM2 and BAM5 from *Ipomea batatas* (IbBAM5) were identified to be tetramers (Cheong *et al.*, 1995; Monroe *et al.*, 2017, 2018; Chandrasekharan *et al.*, 2020). In addition to their structural and functional differences, the members of the AtBAM family also vary in their cellular localization; only BAM5 is exclusively cytosolic, while both BAM7 and BAM8 are nuclear. The remaining six AtBAMs, including BAM2, are found in plastids where starch is stored, suggesting that they might be involved in starch metabolism (see Supplementary Fig. S1 for a graphical summary of the structure of the nine *Arabidopsis* proteins; Monroe & Storm, 2018). Although significant attention has been dedicated to some BAM proteins, much remains to be understood about the structures, *in vivo* functions and evolutionary relationships of the BAM proteins.

This work focuses on comparing the BAM2 and BAM7 proteins from *Arabidopsis* and *Zea mays* (Zm; corn). Based on conserved intron positions, BAM2 is the proposed ancestral protein of BAM subfamily 2, which includes BAM4, BAM5, BAM6, BAM7 and BAM8 (Monroe & Storm, 2018). Although we do not yet understand the function of BAM2, it has persisted in nearly all land plants. Additionally, the *BAM7* gene is likely to have arisen from the fusion of a gene encoding a BZR1-like DNA-binding domain to the 5' end of the *BAM2* gene (Reinhold *et al.*, 2011; Soyk *et al.*, 2014; Thalmann *et al.*, 2019). However, the sequences of AtBAM2 and AtBAM7 have significant differences within the catalytic residues of their respective BAM domains (Monroe *et al.*, 2017). While all annotated *BAM2* genes encode the residues necessary for catalytic activity, most *BAM7* genes have mutations in at least three of the 15 catalytic residues, and this is likely to contribute to BAM7 proteins being catalytically inactive (Reinhold *et al.*, 2011; Soyk *et al.*, 2014).

In this work, we have used the conservation of active-site residues to hypothesize that the *BAM7* gene in some species such as corn may contain two transcriptional start sites that encode two different BAM proteins (ZmBAM7-L and ZmBAM7-S). We further characterized the catalytic activity and solution structure of ZmBAM7-S and compared these findings with AtBAM2, IpBAM5 and AtBAM1 (Cheong *et al.*, 1995; Chandrasekharan *et al.*, 2020). This work expands on our current understanding of the variability of both the gene structure and protein form of BAMs. Ultimately, we find that ZmBAM7-S shows sigmoidal saturation kinetics and a tetrameric structure that suggests that it is a BAM2-like β -amylase.

2. Materials and methods

2.1. Protein sequence alignments

Using the AtBAM2 sequence as a reference (NP_191958.3), a *BLAST* search was conducted using the NCBI RefSeq

database to identify *BAM2* and *BAM7* genes in other annotated land plant genomes. Similarly, the sequence of AtBAM1 (NP_189034.1) was used to find *BAM1* genes in land plants for comparison with the *BAM2* and *BAM7* genes. Species names and accession codes are listed in Supplementary Table S1. The FASTA-formatted protein sequences were then downloaded from NCBI and aligned using *Clustal Omega* (Madeira *et al.*, 2019). The full alignment is available as Supplementary File S2.

2.2. Protein start-site predictions

Initial identification of the alternate start site came from expressed sequence tag (EST) and sequenced cDNA information on the UCSC Genome Browser site for corn BAM7 using B73 RefGen_v3 at chr7:62,317,954–62,321,954 (<http://genome.ucsc.edu>; Kent *et al.*, 2002). Nucleotide sequences of *BAM7* genes were manually analyzed for alternative in-frame translational start sites (ATG) in the intron between the BZR1 and BAM domains. After *in silico* translation, cellular locations of predicted full-length BAM7 proteins (BAM7-L) and shorter proteins lacking the BZR1 domain and starting from in-frame Met residues (BAM7-S) were conducted using *LOCALIZER* (Sperschneider *et al.*, 2017).

2.3. Expression vector construction

DNA sequences for AtBAM2 (NM_116273.5) and ZmBAM7 (NM_001350702.1) were obtained from NCBI. The sequence of ZmBAM7-S was determined based on its in-frame start-codon prediction 42 bases upstream of exon 2. The lengths of the predicted chloroplast transit peptides were determined using *TargetP-2.0* (Almagro Armenteros *et al.*, 2019). Synthesis of the AtBAM2 and ZmBAM7-S coding sequences lacking the predicted 55- and 66-residue chloroplast transit peptides, respectively, was carried out by GenScript (Piscataway, New Jersey, USA) using codon optimization for expression in *E. coli* (sequence available in Supplementary Fig. S2). The cDNAs were then cloned into pET-15b such that the expressed proteins would contain an N-terminal His tag. Transformation of competent *E. coli* DH5 α cells (New England BioLabs, Beverly, Massachusetts, USA) with the plasmid DNA was carried out using the manufacturer's protocol. Plasmids were isolated by miniprep and confirmed after digestion with BamHI and NdeI. Transformation of *E. coli* BL21 cells with each plasmid DNA was carried out using the rapid colony transformation procedure (Micklos & Freyer, 1990). The BAM1 cDNA was a gift from Heike Reinhold and has been described previously (Monroe *et al.*, 2014).

2.4. Protein expression and purification

Cultures of *E. coli* BL21 cells lacking any plasmid (control) or containing one of the previously described recombinant plasmids were grown to an optical density of 0.7 at 600 nm in Luria–Bertani medium with 100 $\mu\text{g ml}^{-1}$ carbenicillin (AtBAM2 and ZmBAM7-S) or in 2 $\times$ YT medium (RPI) along with 1 $\mu\text{l ml}^{-1}$ kanamycin (AtBAM1) at 37°C with shaking at 250 rev min $^{-1}$. Isopropyl β -D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.8 mM and the flasks

were shaken at 250 rev min^{-1} at 20°C overnight. The cells were pelleted by centrifugation and then frozen at -80°C for at least 10 min. The cell pellets were thawed and resuspended in $50 \text{ mM NaH}_2\text{PO}_4$, 0.5 M NaCl , $0.2 \text{ mM tris}(2\text{-carboxyethyl})\text{-phosphine}$ (TCEP), $2 \text{ mM imidazole pH } 8.0$ with EDTA-free protease-inhibitor tablets (Pierce A32965). Cell lysis was completed by sonication in an ice bath (Misonix S-4000; Microtip). After centrifugation of the cell lysate, the supernatants of AtBAM2 and ZmBAM7-S were separately loaded onto a TALON cobalt-affinity column using an ÄKTA start system (Cytiva Life Science, Marlborough, Massachusetts, USA), while AtBAM1 was loaded onto a GE Healthcare nickel-affinity column using the same ÄKTA start system. The bound proteins were eluted from the affinity columns by the stepwise addition of a second buffer consisting of $50 \text{ mM NaH}_2\text{PO}_4$, 0.5 M NaCl , 0.2 mM TCEP , $200 \text{ mM imidazole pH } 8.0$. The four 10 ml elution steps contained 12.5, 50, 125 or 200 mM imidazole mixed using the ÄKTA start system. Fractions were analyzed for purity by SDS-PAGE.

Pure ZmBAM7-S and AtBAM2 were dialyzed overnight using SpectraPor tubing (Spectrum, New Brunswick, New Jersey, USA) with a molecular-weight cutoff of 6–8 kDa. The dialyzed proteins were concentrated in an Amicon Ultra-15 concentrator with a molecular-weight cutoff of 10 kDa at intervals of 30 min at 5000g and 4°C until the desired volume was reached. Protein concentrations were determined using the Bio-Rad Protein Assay Kit with BSA as the standard. AtBAM1 was not dialyzed, but was immediately concentrated in a Spin-X UF concentrator with a PES filter of molecular-weight cutoff 5 kDa for 30 min intervals at $4200 \text{ rev min}^{-1}$ (3215g) until the desired volume of 1 ml was reached.

IbBAM5 from Sigma–Aldrich was resuspended at 7 mg ml^{-1} in $20 \text{ mM HEPES pH } 7.3$, 150 mM NaCl , 0.2 mM TCEP and separated using a HiLoad 16/60 Superdex 200 column equilibrated with SEC buffer ($20 \text{ mM HEPES pH } 7.3$, 150 mM NaCl , 0.2 mM TCEP).

2.5. Size-exclusion chromatography (SEC)

Concentrated ZmBAM7-S and AtBAM1 were further purified using a HiLoad 16/60 Superdex 200 column (Cytiva). Pure ZmBAM7-S and AtBAM1, as confirmed by SDS-PAGE, were concentrated as before, distributed into the plate for SAXS and then flash-frozen in liquid nitrogen. The concentration of ZmBAM7-S was determined by the absorbance at 280 nm using an extinction coefficient of $101\,760 \text{ M}^{-1} \text{ cm}^{-1}$. This value was calculated from the recombinant protein sequence including the His tag using *ProtParam* (Gasteiger *et al.*, 2005). Similarly, the concentration of AtBAM1 was determined at 280 nm using a Synergy H4 Hybrid Reader (BioTek) on a Take3 plate; the path length was 0.05 cm and the extinction coefficient of $59\,511 \text{ M}^{-1} \text{ cm}^{-1}$ was calculated from the sequence using *ProtParam*.

The size-exclusion chromatography data for ZmBAM7-S were used to predict the molecular weight and quaternary structure of this protein. The predicted molecular weight of a ZmBAM7 monomer was calculated using *ProtParam*.

Experimental molecular weights were calculated using calibration standards from the Gel Filtration Molecular Weight Markers Kit for molecular weights 12–200 kDa (Sigma) and the Gel Filtration Calibration Kit for molecular weights 43–669 kDa (Cytiva). These standards were run on the HiLoad 16/60 Superdex 200 column to determine their respective void and elution volumes. The equation of the calibration curve used for molecular-weight calculations was $y = -3.62x + 10.37$ using a void volume of 40.93 ml. The data were analyzed using *Microsoft Excel* (version 16.46), *R* using the *tidyverse* package (version 4.0.3; 10/10/2020; Wickham *et al.*, 2019) and *RStudio* (version 1.3.1093; R Core Team; <https://www.r-project.org/>).

2.6. ZmBAM7-S homology-model construction and quaternary-structure prediction

A homology model of a ZmBAM7-S monomer including the purification tag was produced using *AlphaFold2* and *ColabFold* (Mirdita *et al.*, 2021; Jumper *et al.*, 2021). The sequence alignment was generated using *MMseqs2* (Steinegger & Söding, 2018; Mirdita *et al.*, 2019). The best model based on the pLDDT score was used without further processing for data fitting in *SASREF* and *FoXS*. *SASREF* was used through *ATSAS* online to predict the structure of ZmBAM7-S oligomers (Petoukhov & Svergun, 2005; Petoukhov *et al.*, 2012). A single SAXS curve and a single PDB file of the ZmBAM7-S monomer were used for prediction, with the number of subunits in the final complex controlled by setting the overall symmetry input. No weighting, constraints or contact conditions were added during the fitting process. For comparative fitting of all of the *SASREF* models, the PDB files from *SASREF* were combined into a single zip file and uploaded onto the *FoXS* website (Schneidman-Duhovny *et al.*, 2016). The same SAXS data as used for *SASREF* prediction were provided for data fitting.

2.7. Small-angle X-ray scattering data collection

Full-length ZmBAM7-S was prepared for SAXS by diluting the SEC-purified protein using SEC buffer to five different concentrations (1.76 , 2.64 , 3.53 , 5.29 and 6.17 mg ml^{-1}) in $35 \mu\text{l}$. Samples in a 96-well sample plate were flash-frozen with liquid nitrogen. Three protein-free controls consisting of SEC buffer alone were included with the samples. Purified IbBAM5 was diluted to concentrations of between 1 and 10 mg ml^{-1} in $20 \text{ mM HEPES pH } 7.3$, 150 mM NaCl , 0.2 mM TCEP and flash-frozen in the plate using liquid nitrogen. The sample plate was shipped overnight on dry ice to the Advanced Light Source at Lawrence Berkeley National Laboratory. Prior to data collection, the plate was spun at $3700 \text{ rev min}^{-1}$ for 10 min by the beamline staff. Scattering data were collected from the samples and controls every 0.3 s for a total of 10 s, resulting in 33 frames of data per sample. The beam energy was 11 keV and the detector was 2 m from the sample holder. Samples were kept at 10°C during data collection. Data from the protein-free buffer were collected before and after ZmBAM7-S data collection to ensure there was no difference in scattering due to contamination of the sample cell.

For AtBAM1, data were collected using size-exclusion chromatography coupled to small-angle X-ray scattering. Samples were shipped overnight at 4°C to the SIBYLS beamline at the Advanced Light Source. The sample buffer was 50 mM MES pH 6.7, 100 mM NaCl, 1 mM DTT. The sample was injected into an Agilent 1260 series HPLC with a Shodex KW-802.5 analytical column at a flow rate of 0.5 ml min⁻¹. Small-angle X-ray scattering (SAXS) data were collected from the eluent as it came off the column. The incident light wavelength was 1.03 Å at a sample-to-detector distance of 1.5 m.

Buffer scattering was subtracted from sample scattering in RAW (version 2.0.3; Hopkins *et al.*, 2017). We used SAXS FrameSlice (version 1.4.13) to identify SAXS data frames that lacked discernable aggregation or radiation-induced degradation of the protein. This setup results in scattering vectors, q , ranging from 0.013 to 0.5 Å⁻¹, where the scattering vector is defined as $q = 4\pi\sin\theta/\lambda$ and 2θ is the measured scattering angle.

2.8. BAM solution-structure comparison

We calculated the radius of gyration and molecular weight from SAXS data for ZmBAM7-S, our previous SAXS data for AtBAM2, the reference data for *I. batatas* BAM from SASBDB entry SASDA62 and data collected for *I. batatas* BAM from Sigma and AtBAM1 using RAW. Molecular weights were determined through Bayesian inference. The R_g value was determined using the Guinier fit function. We then plotted the pair distance distribution function for comparison of the shape and size of all of the proteins. The homology model of ZmBAM7-S was fitted to the intensity data using FoXS (Schneidman-Duhovny *et al.*, 2016). All SAXS data sets have been deposited in the SASBDB as SASDM26 (IbBAM5), SASBMX5 (ZmBAM7-S) and SASDMZ5 (AtBAM1).

2.9. Enzyme assays

Amylase activity assays were conducted using purified protein samples in 0.5 ml containing 50 mM MES pH 6, 100 mM KCl and various concentrations of soluble starch (catalog No. AC424495000, Acros Organics, Morris Plains, New Jersey, USA). After 20 min at 25°C, the reaction tubes were immersed in boiling water for 3 min to stop the reaction. The reducing sugars in each reaction were then measured using the Somogyi–Nelson assay (Nelson, 1944). Data were analyzed using Microsoft Excel (version 16.46) and fitted to the Michaelis–Menten equation (1), which accounts for cooperativity, using Solver,

$$\text{rate} = \frac{V_{\max} \times [S]^n}{K_m^n + [S]^n}. \quad (1)$$

3. Results

Although they are evolutionarily related, BAM2 and BAM7 from *A. thaliana* (At) are functionally and structurally quite different: AtBAM2 is a catalytically active, plastid-localized tetramer, while AtBAM7 is a catalytically inactive, nuclear-

localized transcription factor that probably functions as a dimer (Reinhold *et al.*, 2011; Soyk *et al.*, 2014; Monroe *et al.*, 2017, 2018; Chandrasekharan *et al.*, 2020). Interestingly, some land plants such as corn (*Z. mays*; Zm) do not have an annotated BAM2 gene, but they contain a putative BAM7 gene that appears to share many conserved active-site residues with AtBAM2. Importantly, all of the genomes that contain this unusual BAM7 gene, which we will refer to as dual-function BAM7 or DF-BAM7, also appear to lack a separate, *Arabidopsis*-like BAM2 gene, as described by Monroe *et al.* (2018). Therefore, we wanted to determine whether this interesting BAM7 gene in plants that lack a BAM2 gene encodes a structural and functional ortholog of AtBAM2 within the BAM7 gene. If this is true, the predicted ZmBAM2-like protein, which we hypothesize is encoded within the dual-function ZmBAM7 gene, is initiated from a predicted second transcriptional start site (TSS) and should have catalytic and structural characteristics similar to those of AtBAM2.

3.1. Functional residue alignment of BAM2 and BAM7

A. thaliana BAM7 (AtBAM7) has an N-terminal, BZR1-like DNA-binding domain that contains a bipartite nuclear localization signal (NSL; Reinhold *et al.*, 2011). Although AtBAM7 is reported to be catalytically inactive, the BAM domain of AtBAM7 is necessary for specific DNA binding (Soyk *et al.*, 2014). The ‘active site’ contains four mutations among the 15 residues that form hydrogen bonds to the four glucose residues at the nonreducing end of starch (see Fig. 2b in Monroe & Storm, 2018). These mutations are likely to contribute to the catalytic inactivity of AtBAM7 that was found under certain conditions (Reinhold *et al.*, 2011). In the course of analyzing the BAM proteins from sequenced plant genomes, we noticed that predicted BAM7 genes from several plants contained BAM domains that seemed to be more similar to AtBAM2 than to AtBAM7. We identified 12 BAM7 genes from basal angiosperms, monocots and basal eudicots that lacked a BAM2 gene and compared them with the BAM2 and BAM7 protein sequences from 14 eudicot genomes that contain separate BAM2 and BAM7 genes. We also included BAM1 proteins from the same genomes in our analysis. We used Clustal Omega (Madeira *et al.*, 2019) to align the amino-acid sequences (Supplementary File S2) and we then identified residues that play a role in the specific functions of either BAM7 (Reinhold *et al.*, 2011; Soyk *et al.*, 2014) or BAM2 (Monroe *et al.*, 2017, 2018; Fig. 1). A putative bipartite NLS was found in the BZR1-like domain of each BAM7 gene, with only minor differences. In five of the DF-BAM7 proteins the distance between the two regions of positively charged residues was 13 or 15 residues, as opposed to 14 in all of the BAM7 proteins, while in half of the BAM7 and DF-BAM7 proteins His in the second positive region was substituted with Gln (Fig. 1a; Supplementary Table S2). In addition, a Glu residue that was confirmed to be essential for DNA binding (Glu87 in AtBAM7; Soyk *et al.*, 2014) is perfectly conserved in all BAM7 and DF-BAM7 proteins (data not shown).

The active-site starch-binding residues (Laederach *et al.*, 1999) are perfectly conserved in all but one of the BAM2

proteins and in all of the BAM1 and DF-BAM7 proteins, suggesting that these enzymes are all likely to be catalytically active (Fig. 1*b*). In contrast, all but one of the BAM7

sequences from eudicots we analyzed that also contained a separate BAM2 gene had mutations among the active-site residues (Fig. 1*b*; Supplementary Table S2).

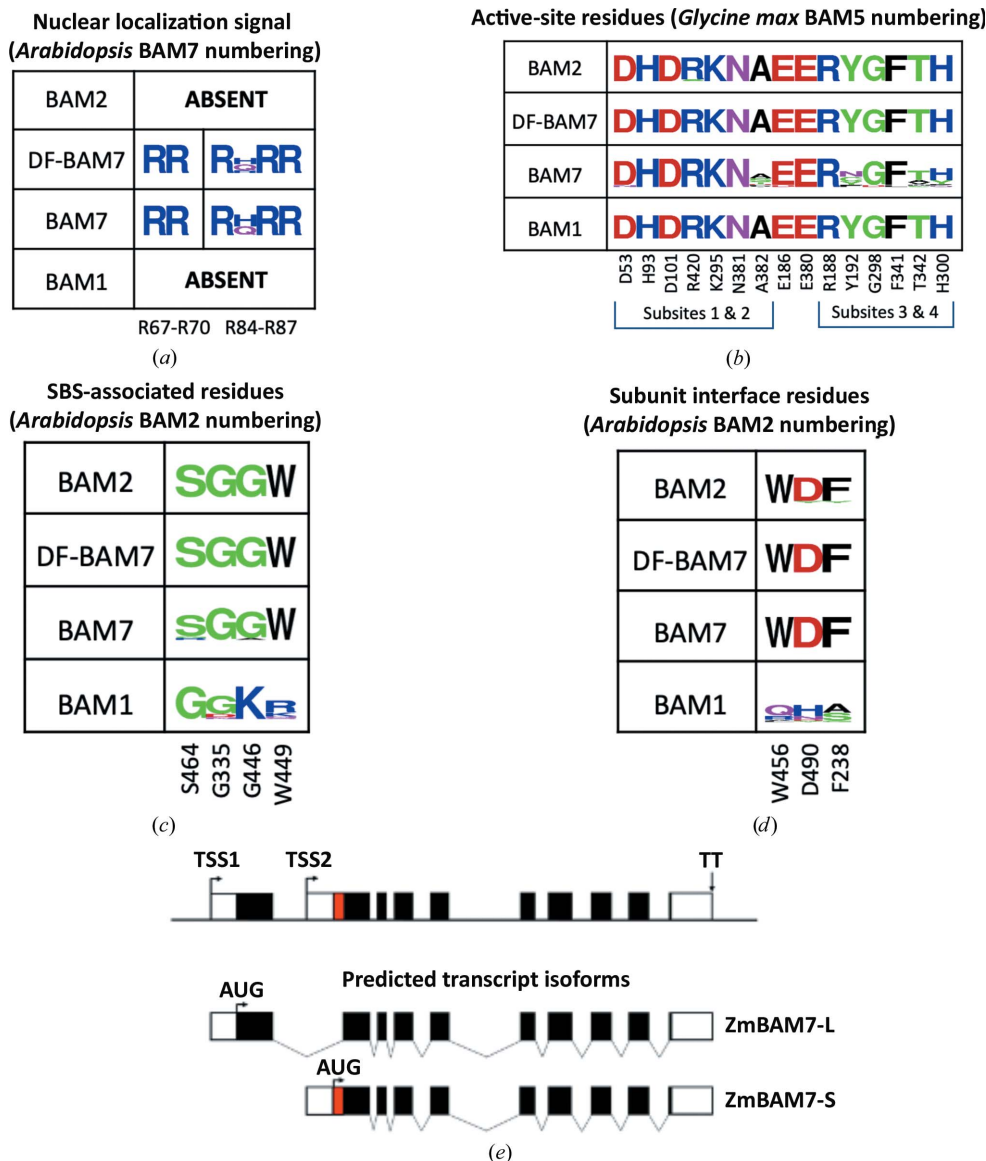


Figure 1

Conservation of key residues in BAM2 and BAM7 from flowering plants and how they compare with the corresponding residues in DF-BAM and BAM1, and a map of the corn *BAM7* gene with its two predicted transcripts. (a) WebLogo illustrating the conservation of residues predicted to be part of the nuclear localization signal in the DNA-binding domain as identified by mutagenesis in *A. thaliana* BAM7 (Reinhold *et al.*, 2011). The alignment included 14 species of flowering plants that contain BAM7 that were selected to represent a diversity of orders and 12 species that contain DF-BAM7 (see Supplementary Table S1 for species and accession numbers). (b) WebLogo illustrating the conservation of 15 residues in the same species as in (a) identified as forming hydrogen bonds to starch in the active site of soybean (*Glycine max*) BAM5 (Laederach *et al.*, 1999). (c) WebLogo illustrating the conservation of residues as in (a) compared with the three residues identified in BAM2 as being involved in allosteric regulation of activity in BAM2 (Ser464) or starch binding to a surface binding site for starch (Gly335, Gly446 and Trp449) (Monroe *et al.*, 2017, 2018). (d) WebLogo illustrating the conservation of three residues in the same species as in (a) that were identified by mutagenesis as being important for tetramer stabilization: Trp456 and Asp590 in interface 'A' and Phe238 in interface 'B' (Monroe *et al.*, 2018). (e) Predicted dual-function ZmBAM7 gene model. Coding regions of exons are colored black, with the exception of a region unique to the N-terminus of ZmBAM7-S that is colored red. The locations of the two putative transcriptional start sites (TSS1 and TSS2) and their respective translational start sites (AUG) in the two predicted transcripts are indicated with arrows. The 5' and 3' untranslated regions (UTR) of both transcripts are colored white.

AtBAM2 is unusual among characterized BAMs in having a sigmoidal substrate-saturation curve, in being tetrameric and in having a secondary starch-binding site (SBS) in a groove between the monomers of each dimer (Monroe *et al.*, 2017, 2018; Chandrasekharan *et al.*, 2020). We next looked for key amino acids within the BAM7 and DF-BAM7 sequences that had previously been identified as functioning in each of these unique aspects of BAM2 (Supplementary Fig. S3). Ser464, Gly335, Gly446 and Trp449, which were previously identified as being associated with the SBS and sigmoidal kinetics in BAM2 (Monroe *et al.*, 2017, 2018), are all perfectly conserved in the BAM2 and DF-BAM7 proteins, and in ten of the 12 BAM7 proteins (Fig. 1*c*, Supplementary Fig. S3). Residues that altered the oligomerization of BAM2 when mutated include Phe238, Trp456 and Asp490 (Monroe *et al.*, 2018). With the exception of two BAM2 proteins in which one of these residues differed, they were conserved in BAM2, BAM7 and DF-BAM7 proteins and were not conserved in the monomeric BAM1 (Fig. 1; Sparla *et al.*, 2006). Together, these results suggest that the DF-BAM7 proteins share most, if not all, of the key residues identified as being important for BAM7 and BAM2, and thus we hypothesize that they may serve both functions.

3.2. Dual-function BAM7 gene structure

Based on the above sequence analysis, we hypothesize that the *BAM7* genes in corn and other land plants that lack a separate *BAM2* gene have alternative transcription start sites (TSSs) so that the first start site leads to a

Table 1
SAXS data-collection parameters and analysis software.

Sample name	Full-length ZmBAM7-S	AtBAM1	IpBAM
Instrument	SIBYLS beamline 12.3.1	SIBYLS beamline 12.3.1	SIBYLS beamline 12.3.1
Wavelength (Å)	1.27	1.03	1.27
q-range (Å ⁻¹)	0–0.5	0–0.5	0–0.5
Concentration (mg ml ⁻¹)	1.76	5.0	7.0
Temperature (K)	283.15	283.15	283.15
I(0) [from p(r)] (AU)	87.1 ± 0.29	16.23 ± 0.056	272.5 ± 0.351
R _g [from p(r)] (Å)	51.6 ± 0.11	27.51 ± 0.12	43.67 ± 0.051
I(0) (from Guinier) (AU)	88.9 ± 0.7	15.68 ± 0.075	272.5 ± 0.632
R _g (from Guinier) (Å)	52.4 ± 0.5	25.75 ± 0.19	43.75 ± 0.127
D _{max} (Å)	163	95	141
Porod volume estimate (V _p) (Å ³)	4.12 × 10 ⁵	6.76 × 10 ⁴	2.51 × 10 ⁵
Molecular weight M _r , Bayes† (Da)	318400; 100% CI 221100–372700	53100; 100% CI 50300–56200	185800; 100% CI 162700–194900
M _r (from Porod volume) (Da)	343800	56100	208400
Calculated monomeric M _r from sequence (Da)	58500	59511	55949
Primary data reduction	RAW	RAW	RAW
Data processing	RAW	RAW	RAW
Computation of model intensities	FoXS	Not used	Not used
Three-dimensional graphical representations	YASARA	Not used	Not used
SASBDB identifier	SASDMX5	SASDMZ5	SASDM26

† CI, confidence interval.

longer transcript that encodes a BAM7-like protein and contains an NLS. The shorter transcript would be initiated at the second TSS, which we predict is located in the first intron of the DF-BAM7 gene and encodes a BAM2-like protein with an N-terminal chloroplast transit peptide (cTP). We are currently using 5'-RACE to test this prediction, but the hypothesis is supported by known ESTs and cap analysis of gene expression (CAGE) data (Mejía-Guerra *et al.*, 2015). Using the annotated corn BAM7 gene, we created a proposed gene-structure model showing both putative TSSs and their respective translational start sites (Fig. 1e). The longer transcript (BAM7-L) includes the first exon of the coding region which contains the DNA-binding domain and predicted NLS. This gene product would be transcribed from the first TSS, contain all ten exons and function as BAM7. The shorter transcript lacks the first exon and would encode BAM7-S, which we predict is initiated from a cryptic start codon at the N-terminus of a 66-amino-acid putative cTP (Fig. 1e). RNA encoding the N-terminal 14 amino acids of this cTP would be spliced out of the BAM7-L transcript, so it is specific to the BAM7-S transcript. Both transcripts have a common BAM domain with nine exons and a common translational termination site (Fig. 1e). Upon closer inspection of the 12 DF-BAM7 genes that we identified, an in-frame start codon was identified in a similar position within the first intron of each gene. This cryptic translational start site is also predicted to be part of the cTP. Therefore, the shorter versions of the DF-BAM7 genes might be expressed and targeted to the chloroplast if they were translated from a shorter mRNA transcript. In contrast, only four of the 14 BAM7 genes from genomes that also contained a separate BAM2 gene had an in-frame ATG codon in intron 1, and none of these four was predicted to initiate a cTP (data not shown). Importantly, the longer ZmBAM7-L protein would lack a functional cTP because it is not located at the N-terminus, and the shorter ZmBAM7-S protein would lack the NLS because it lacks the DNA-binding domain which contains the NLS.

3.3. Recombinant protein purification

To test the hypothesis that the BAM7 gene encodes both BAM7-like and BAM2-like proteins in some plants that lack a BAM2 gene, we expressed and purified ZmBAM7-S and AtBAM2 in *E. coli*. Both ZmBAM7-S and AtBAM2 were expected to have a molecular weight of 58 kDa including the His tag. The absence of a prominent 58 kDa band in the sonicated supernatant of *E. coli* BL21 cells was used to determine the expression and solubility of the recombinant proteins (Fig. 2a). ZmBAM7-S was expressed (Fig. 2a) and found to be soluble after sonication and centrifugation based on the 58 kDa band in the sonicated supernatant (Fig. 2a). We then purified ZmBAM7-S using TALON cobalt-affinity chromatography followed by size-exclusion chromatography (SEC; Fig. 2a). The pure protein had a molecular weight of about 58 kDa on SDS-PAGE. Following SEC, the molecular weight of ZmBAM7-S using the trend-line equation of the calibration curve was calculated to be 384.8 kDa (Figs. 2b and 2c). The molecular weight of ZmBAM7-S from its sequence should be 58 kDa, suggesting that ZmBAM7-S forms oligomers of up to six subunits in solution.

3.4. ZmBAM7-S homology model

Since there are no experimentally determined structures of ZmBAM7-S, we produced homology models of the recombinant ZmBAM7-S protein sequence using AlphaFold2 and ColabFold (Jumper *et al.*, 2021; Mirdita *et al.*, 2021). The resulting model showed an extended N-terminal region of 77 amino acids followed by a TIM-barrel domain. This latter folded region is consistent with the structure of other BAM enzymes (Monroe & Storm, 2018). To support the model, we then predicted the disorder of ZmBAM7-S, finding that residues 1–90 of the model had the highest probability of being disordered (Fig. 2d; McGuffin, 2008; Mészáros *et al.*, 2018; Erdős & Dosztányi, 2020).

3.5. Small-angle X-ray scattering

We next acquired small-angle X-ray scattering (SAXS) data for ZmBAM7-S using the SIBYLS beamline. The SAXS data-collection parameters and analysis software used are outlined in Table 1. We did not observe a trend in the radius of gyration (R_g) for ZmBAM7-S with concentration, indicating that ZmBAM7-S forms a consistent oligomer (Fig. 3*a*). Because there is limited solution structure information for any BAM, we further collected SAXS data for *A. thaliana* BAM1 (AtBAM1) and *I. batatas* BAM5 (IbBAM5) for comparison with the ZmBAM7-S data. We calculated the R_g for AtBAM1 to be 25.8 ± 0.2 Å with a molecular weight of 53.1 kDa (95% confidence interval) and that for IbBAM5 to be 43.8 ± 0.1 Å with a molecular weight of 185.8 kDa (95% confidence interval) (Table 1). IbBAM5 is generally accepted to be a tetramer in solution, while BAM1 is thought to be monomeric, and these data are consistent with these previous proposals (Cheong *et al.*, 1995; Sparla *et al.*, 2006). When we compared the pair distance distribution function (PDDF) for ZmBAM7-S with the data for AtBAM2, AtBAM1 and IbBAM5, we

observed that the data for ZmBAM7-S showed a similar peak distance value to the data for AtBAM2. However, there was a long tail towards $D_{\max}$ suggestive of a structurally extended region (Fig. 4*c*). We then used *SASREF* to fit the ZmBAM7-S homology model to the data and construct oligomers of the homology model using the SAXS data as a guide. Matching the SEC data, *SASREF* showed that oligomers of ZmBAM7-S fitted the data better than a monomer (χ^2 of 2.25–2.83 for a tetramer versus 1123.26 for a monomer) (Fig. 4*d*). A tetramer with $P222$ symmetry fitted better than a tetramer with $P4$ symmetry (Figs. 4*d* and 4*e*). Collectively, these data support a core structure for ZmBAM7-S that is similar in shape and construction to those of AtBAM2 and IbBAM5.

3.6. Enzyme-activity assays

If ZmBAM7-S functions as a BAM2-like protein, we predict that it would be catalytically active and exhibit sigmoidal kinetics (Monroe *et al.*, 2017). In order to compare the catalytic activity of ZmBAM7-S with that of AtBAM2, both proteins were purified to near-homogeneity (Fig. 4*a*). The

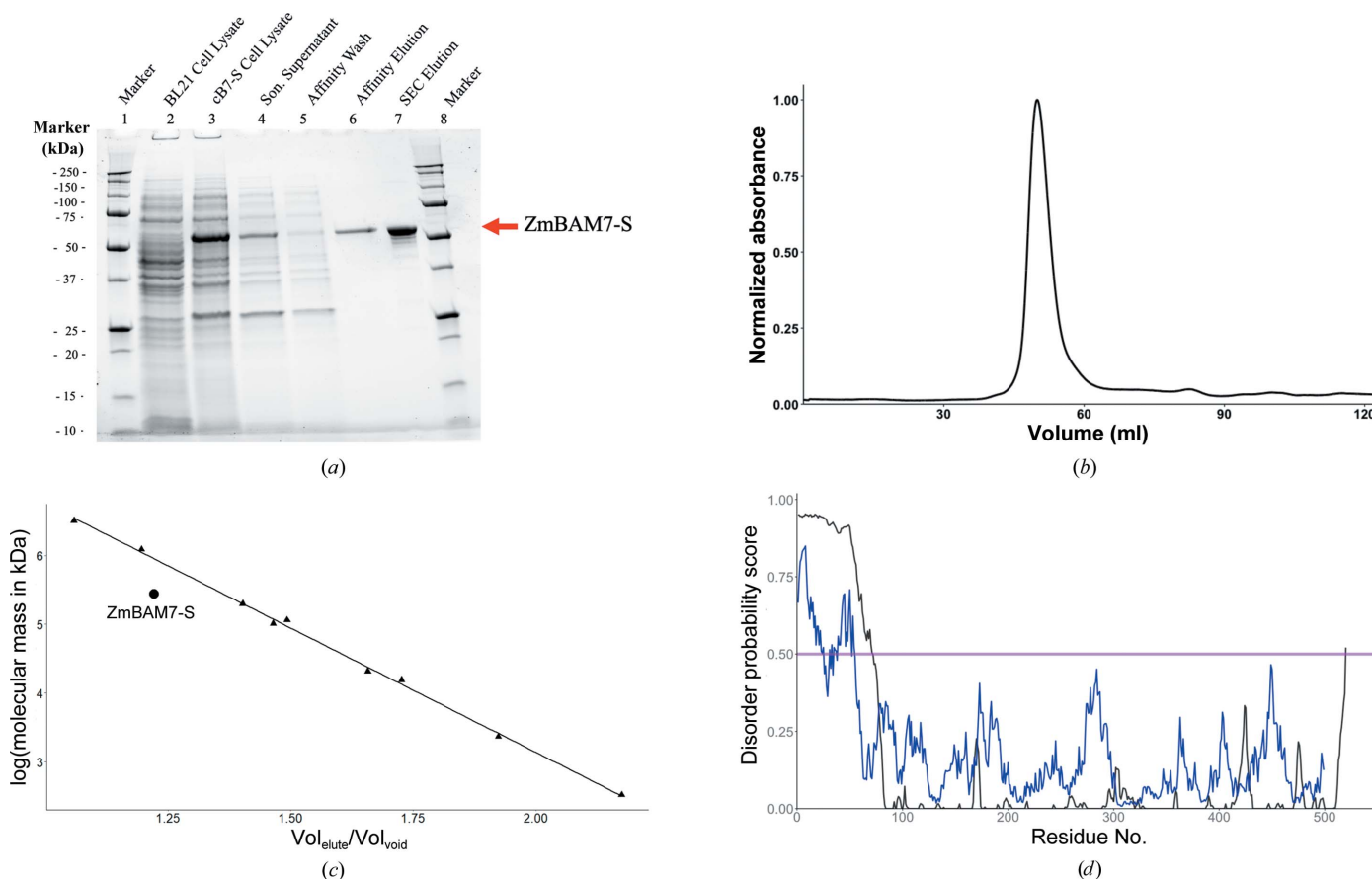


Figure 2

Purification of recombinant ZmBAM7-S. (*a*) Pure protein was eluted in 10 ml fractions at increasing imidazole concentrations during affinity chromatography; only one elution fraction is shown for ZmBAM7-S (lane 6). A wash fraction with less than 12.5 mM imidazole is also shown (lane 5). Size-exclusion chromatography (SEC) was also performed in preparation for small-angle X-ray scattering analysis (lane 7). Markers are present in lanes 1 and 8. (*b*) SEC of ZmBAM7-S. Absorbance data were normalized to the largest value. The peak elution volume for ZmBAM7-S was 50.0 ml. (*c*) SEC molecular-weight calibration curve. The black line with the equation $y = -3.62x + 10.37$ was created from nine different calibration standards (gray points). The expected ZmBAM7-S tetrameric molecular weight (black) calculated from the ZmBAM7-S sequence (232 kDa) is shown. (*d*) Disorder predictions from *IntFOLD* and *IUPred2A* for ZmBAM7-S. Using a disorder probability score cutoff of 0.5 (pink line), the probability of being disordered was predicted for each residue in the ZmBAM7-S homology model (black line) or from the sequence of ZmBAM7-S using *IUPred2A* (blue line).

purified corn protein was catalytically active, and this activity did exhibit sigmoidal kinetics, but because soluble starch at or above 100 mg ml^{-1} rapidly retrogrades at 25°C , we were

unable to conduct assays at higher levels of substrate to reach saturation (Fig. 4*b*). From fitting of the data to a cooperative Michaelis–Menten equation, we calculate the maximum

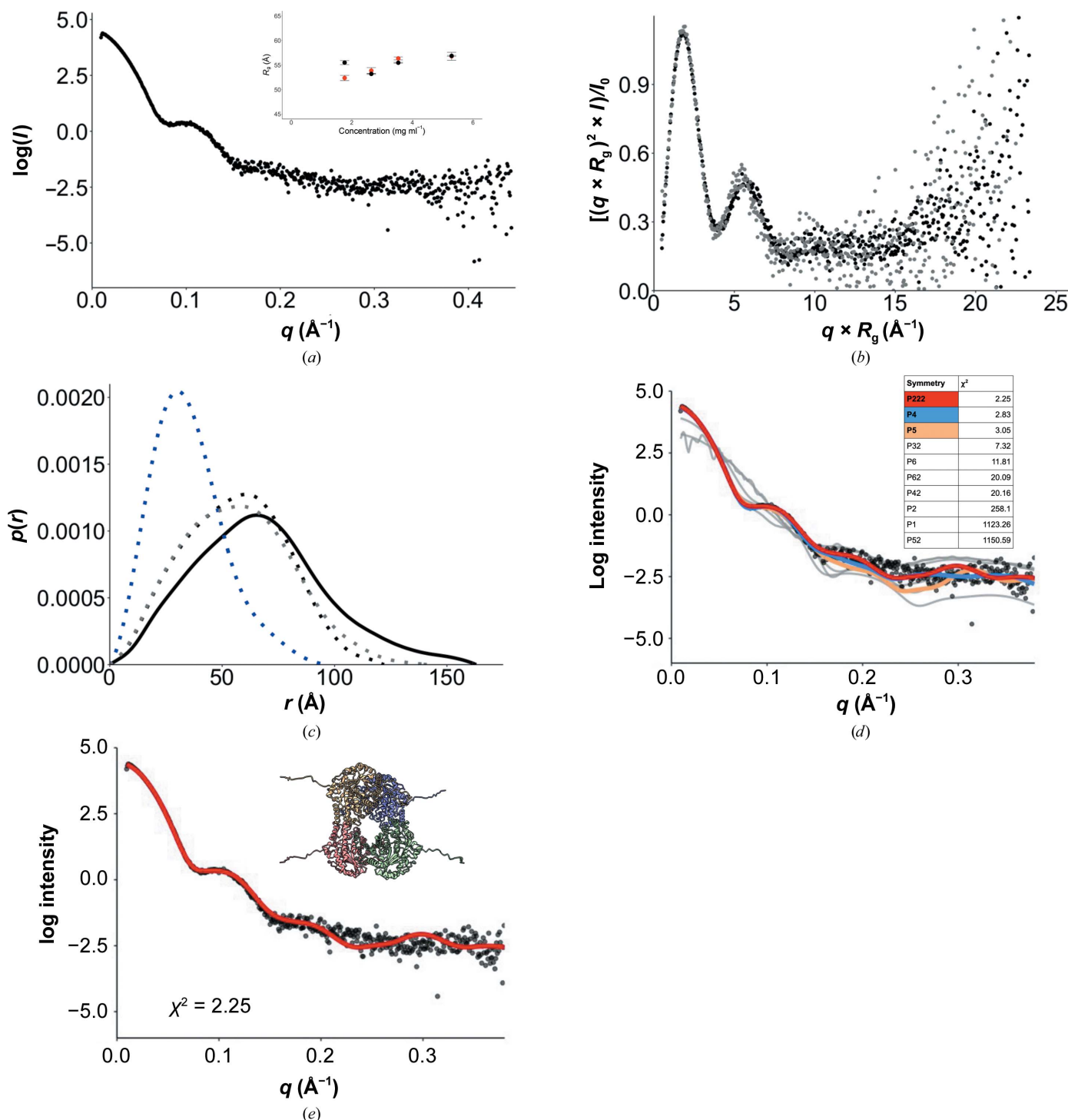


Figure 3

SAXS data for ZmBAM7-S and truncated ZmBAM7-S. (a) Log of intensity versus momentum transfer for ZmBAM7-S (black) and truncated ZmBAM7-S (gray). The inset plot shows the radius of gyration (R_g) versus ZmBAM7-S concentration (mg ml^{-1}). The R_g values of ZmBAM7-S at four different concentrations submitted to SAXS analysis were calculated from the Guinier plot (red data points) and from the $P(r)$ plot (black data points). (b) Kratky plot of ZmBAM7-S (black) and truncated ZmBAM7-S (gray). (c) Pair distance distribution function plot comparing ZmBAM7-S with other BAMs. ZmBAM7-S is shown as a solid black line, truncated ZmBAM7-S is shown as a solid gray line, AtBAM1 is shown as a dotted blue line, IbBAM5 is shown as a dotted gray line and AtBAM2 is shown as a dotted black line. (d) SASREF fits of ZmBAM7-S monomer and oligomers to the SAXS data. SASREF models were combined into a single file and fitted to the SAXS data using FoXS. The χ^2 values were 1123.26 for P1, 258.1 for P2, 2.83 for P4, 2.25 for P222, 11.81 for P6, 7.32 for P32, 20.16 for P42 and 1150.59 for P52. (e) FoXS fit of the P222 model from SASREF to the SAXS data.

activity of ZmBAM7-S to be 160.7 U mg^{-1} (95% confidence interval $137\text{--}185 \text{ U mg}^{-1}$), while the value for AtBAM2 was 367 U mg^{-1} (95% confidence interval $353\text{--}381 \text{ U mg}^{-1}$). ZmBAM7-S also shows an apparently weaker affinity for soluble starch, with a K_m of 86.9 mg ml^{-1} (95% confidence interval $83\text{--}91 \text{ mg ml}^{-1}$) compared with a value of 59.4 mg ml^{-1} (95% confidence interval $58\text{--}61 \text{ mg ml}^{-1}$) for AtBAM2 (Supplementary Fig. S4). These data show that ZmBAM7-S is apparently a less efficient enzyme; however, the inability to saturate the enzyme with substrate means that these values are estimates and should be taken as a preliminary study of ZmBAM7-S activity.

4. Discussion

While analyzing the genomes of land plants for the presence of *BAM2* genes, we observed that some plants appeared to lack a *BAM2* gene but contained a *BAM7* gene. This was puzzling because *BAM2* is more ancient and *BAM7* is likely to have arisen from a fusion of a BZR1 DNA-binding domain to the 5' end of *BAM2* (Thalman *et al.*, 2019). However, upon closer inspection of these *BAM7* genes we discovered that the BAM domains of some genes have a greater percentage identity to AtBAM2 than to AtBAM7, and these *BAM7* genes have all of the catalytic and starch-binding residues that *BAM7* genes sometimes lack (Figs. 1*a–d*). In addition, in-frame start codons and cryptic chloroplast transit peptides were predicted within the first introns in each of the annotated *BAM7* genes in 12 genomes that lack a separate *BAM2* gene (Fig. 1*e*). These observations led us to hypothesize that the *BAM7* gene in these plants that lack a separate *BAM2* gene, such as corn, is a dual-function gene that encodes two structurally and func-

tionally different proteins, BAM7-L and BAM7-S, by alternative transcriptional start sites, forming functional BAM7-like and BAM2-like proteins, respectively. To test this hypothesis, we designed a clone of ZmBAM7-S for expression in *E. coli* to compare its catalytic properties with those of AtBAM2. We also purified ZmBAM7-S using size-exclusion chromatography for small-angle X-ray scattering analysis and showed its oligomeric structure to be tetrameric (Fig. 4).

AtBAM7 is a nuclear-localized transcription factor that is likely to function as a dimer and has no previously observed catalytic activity on starch, while AtBAM2 is a catalytically active tetramer (Reinhold *et al.*, 2011; Soyk *et al.*, 2014; Monroe & Storm, 2018). If the predicted BAM7-S proteins function like AtBAM2 proteins then they should have properties that are more similar to AtBAM2 than to AtBAM7. To test this, we began by analyzing the protein sequences encoded by *BAM7* genes found in some genomes that lack a separate *BAM2* gene. The catalytic activity of AtBAM2 has been attributed to the active-site residues that it shares with soybean BAM5; these residues include those that are necessary for starch binding and catalytic activity (Laederach *et al.*, 1999; Kang *et al.*, 2005; Monroe *et al.*, 2017). In our sequence alignment, all of the putative dual-function *BAM7* genes that we analyzed have perfectly conserved active-site residues like nearly all *BAM2* sequences but unlike most *BAM7* sequences from genomes that contain a separate *BAM2* gene (Fig. 1 and Supplementary File S2). The BAM domain of BAM7 was found to be necessary for the transcription-factor activity of BAM7, but it does not apparently catalyze a reaction like the BAM domain of BAM2 (Soyk *et al.*, 2014). Our active-site residue analysis indicated that BAM7 sequences showed conservation with other BAMs only in subsite 1 and part of

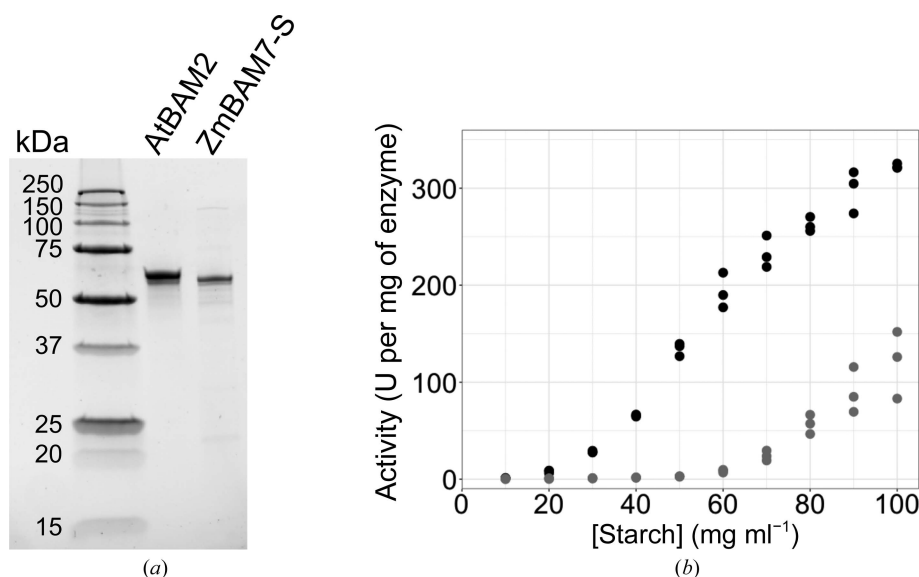


Figure 4

Activity of ZmBAM7-S. (a) Gel showing the purity and amount of AtBAM2 and ZmBAM7-S used in activity assays. (b) Effect of substrate concentration on ZmBAM7-S (gray) and AtBAM2 (black) activity on a per milligram of protein basis. Points are shown for each of three replicate assays. Data were fitted to the Michaelis–Menten equation for cooperative enzymes. For AtBAM2 the K_m was 59.4 mg ml^{-1} (95% confidence interval $58\text{--}61 \text{ mg ml}^{-1}$) and for ZmBAM7-S the K_m was 86.9 mg ml^{-1} (95% confidence interval $83\text{--}91 \text{ mg ml}^{-1}$). The fitted data and all kinetic values are shown in Supplementary Fig. S4.

subsite 2. This may indicate that BAM7 binds maltose in the deep pocket of the BAM domain but does not bind starch like BAM2-like proteins. In addition, when the full sequences of AtBAM2, AtBAM7 and ZmBAM7 were aligned, ZmBAM7 and AtBAM7 had a conserved N-terminal sequence, indicating that they potentially have similar predicted DNA-binding properties. However, within the BAM domain of all three proteins, ZmBAM7 is more like AtBAM2 than AtBAM7 (Figs. 1a–1d).

In addition, ZmBAM7 shares many of the interface residues found in AtBAM2, which suggests it may form a similar tetramer (Fig. 1d). When we analyzed the SAXS data and the PDDF derived from these data, we observed that the data for ZmBAM7-S aligned better with the data for AtBAM2 and IbBAM5, both of which are tetrameric (Fig. 3c). In comparison, the data for BAM1, which we showed to be monomeric in solution, did not align with the data for ZmBAM7-S (Fig. 3c). The main difference between the SAXS data for AtBAM2 and ZmBAM7-S was the long tail in the data towards $D_{\max}$, suggesting an extended region (Fig. 4c). While we are not certain of the exact conformation of the ZmBAM7-S structure, we are confident that the protein is tetrameric. Most BAM proteins are thought to be monomeric or a mixture of monomers and tetramers in solution, but AtBAM2 is constitutively a tetramer (Chandrasekharan *et al.*, 2020). While the physiological functions of AtBAM2 and ZmBAM7-S have not been determined, it is known that AtBAM2 is active on soluble starch and exhibits sigmoidal kinetics (Monroe *et al.*, 2017). Similarly, ZmBAM7-S was active and showed sigmoidal substrate-saturation kinetics (Fig. 4b and Supplementary Fig. S4). Together with the sequence and structure information, these data support our hypothesis that ZmBAM7-S is an AtBAM2-like β -amylase. It is likely that the long form, ZmBAM7-L, would also be catalytically active, but this would be of no functional consequence because it is located in the nuclei, which lack starch.

Alternative transcriptional start sites are an underappreciated mechanism of gene and protein regulation compared with alternative splicing and translational regulation. These alternative transcriptional start sites and promoters have been hypothesized to regulate gene expression, alter mRNA stability or produce two proteins with different N-terminal regions (Ayoubi & Van De Ven, 1996; Mejía-Guerra *et al.*, 2015). However, others found that alternative transcriptional initiation was likely to be due to molecular errors and was not adaptive (Xu *et al.*, 2019). Genome-wide transcriptional start-site (TSS) determination in corn identified about 1500 genes that have multiple transcriptional start sites (Mejía-Guerra *et al.*, 2015). Sequenced cDNAs in the Maize Genome Database from corn locus Zm00001d019756 appear to encode both long and short BAM7 proteins (Portwood *et al.*, 2019). In addition, our preliminary analysis of ZmBAM7 transcripts using 5'-RACE also supports the existence of long and short transcripts of this gene *in vivo* (K. Ozcan & J. Monroe, unpublished data). If the corn DF-BAM7 gene indeed encodes two functionally distinct proteins orthologous to AtBAM7 and AtBAM2, then it is conceivable

that other unrecognized functional genes reside within annotated genomes. Techniques for identifying alternative transcriptional start sites within coding regions would be useful in correcting these oversights and preventing more in the future.

5. Related literature

The following reference is cited in the supporting information for this article: Pettersen *et al.* (2021).

Funding information

This work was supported by a National Science Foundation Research at Undergraduate Institutions grant to JDM and CEB (MCB-1932755), a Summer Undergraduate Research Fellowship from the American Society of Plant Biologists to CMR, National Science Foundation Research Experience for Undergraduates (CHE-1757874), infrastructure provided by a grant from the Thomas F. and Kate Miller Jeffress Memorial Trust to CEB and a grant from the 4-VA organization to CEB. Additional support comes from the National Institutes of Health project ALS-ENABLE (P30 GM124169) and a HighEnd Instrumentation Grant (S10 OD018483).

References

- Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G. & Nielsen, H. (2019). *Nat. Biotechnol.* **37**, 420–423.
- Ayoubi, T. A. & Van De Ven, W. J. M. (1996). *FASEB J.* **10**, 453–460.
- Chandrasekharan, N. P., Ravenburg, C. M., Roy, I. R., Monroe, J. D. & Berndsen, C. E. (2020). *Acta Cryst. D* **76**, 357–365.
- Cheong, C. G., Eom, S. H., Chang, C., Shin, D. H., Song, H. K., Min, K., Moon, J. H., Kim, K. K., Hwang, K. Y. & Suh, S. W. (1995). *Proteins*, **21**, 105–117.
- Erdős, G. & Dosztányi, Z. (2020). *Curr. Protoc. Bioinform.* **70**, e99.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa: Humana Press.
- Hopkins, J. B., Gillilan, R. E. & Skou, S. (2017). *J. Appl. Cryst.* **50**, 1545–1553.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. (2021). *Nature*, **596**, 583–589.
- Kang, Y.-N., Tanabe, A., Adachi, M., Utsumi, S. & Mikami, B. (2005). *Biochemistry*, **44**, 5106–5116.
- Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. & Haussler, D. (2002). *Genome Res.* **12**, 996–1006.
- Laederach, A., Dowd, M. K., Coutinho, P. M. & Reilly, P. J. (1999). *Proteins*, **37**, 166–175.
- Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D. & Lopez, R. (2019). *Nucleic Acids Res.* **47**, W636–W641.
- McGuffin, L. J. (2008). *Bioinformatics*, **24**, 1798–1804.
- Mejía-Guerra, M. K., Li, W., Galeano, N. F., Vidal, M., Gray, J., Doseff, A. I. & Grotewold, E. (2015). *Plant Cell*, **27**, 3309–3320.

- Mészáros, B., Erdős, G. & Dosztányi, Z. (2018). *Nucleic Acids Res.* **46**, W329–W337.
- Micklos, D. A. & Freyer, G. A. (1990). *DNA Science: A First Course in Recombinant DNA Technology*. New York: Cold Spring Harbor Laboratory Press.
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. & Steinegger, M. (2021). *bioRxiv*, 2021.08.15.456425.
- Mirdita, M., Steinegger, M. & Söding, J. (2019). *Bioinformatics*, **35**, 2856–2858.
- Monroe, J. D., Breault, J. S., Pope, L. E., Torres, C. E., Gebrejesus, T. B., Berndsen, C. E. & Storm, A. R. (2017). *Plant Physiol.* **175**, 1525–1535.
- Monroe, J. D., Pope, L. E., Breault, J. S., Berndsen, C. E. & Storm, A. R. (2018). *Front. Plant Sci.* **9**, 1176.
- Monroe, J. D. & Storm, A. R. (2018). *Plant Sci.* **276**, 163–170.
- Monroe, J. D., Storm, A. R., Badley, E. M., Lehman, M. D., Platt, S. M., Saunders, L. K., Schmitz, J. M. & Torres, C. E. (2014). *Plant Physiol.* **166**, 1748–1763.
- Nelson, N. (1944). *J. Biol. Chem.* **153**, 375–380.
- Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V. & Svergun, D. I. (2012). *J. Appl. Cryst.* **45**, 342–350.
- Petoukhov, M. V. & Svergun, D. I. (2005). *Biophys. J.* **89**, 1237–1250.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. & Ferrin, T. E. (2021). *Protein Sci.* **30**, 70–82.
- Portwood, J. L., Woodhouse, M. R., Cannon, E. K., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., Walsh, J. R., Sen, T. Z., Cho, K. T., Schott, D. A., Braun, B. L., Dietze, M., Dunfee, B., Elsik, C. G., Manchanda, N., Coe, E., Sachs, M., Stinard, P., Tolbert, J., Zimmerman, S. & Andorf, C. M. (2019). *Nucleic Acids Res.* **47**, D1146–D1154.
- Reinhold, H., Soyk, S., Šimková, K., Hostettler, C., Marafino, J., Mainiero, S., Vaughan, C. K., Monroe, J. D. & Zeeman, S. C. (2011). *Plant Cell*, **23**, 1391–1403.
- Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. (2016). *Nucleic Acids Res.* **44**, W424–W429.
- Soyk, S., Šimková, K., Zürcher, E., Luginbühl, L., Brand, L. H., Vaughan, C. K., Wanke, D. & Zeeman, S. C. (2014). *Plant Cell*, **26**, 1746–1763.
- Sparla, F., Costa, A., Lo Schiavo, F., Pupillo, P. & Trost, P. (2006). *Plant Physiol.* **141**, 840–850.
- Sperschneider, J., Catanzariti, A.-M., DeBoer, K., Petre, B., Gardiner, D. M., Singh, K. B., Dodds, P. N. & Taylor, J. M. (2017). *Sci. Rep.* **7**, 44598.
- Steinegger, M. & Söding, J. (2018). *Nat. Commun.* **9**, 2542.
- Thalmann, M., Coiro, M., Meier, T., Wicker, T., Zeeman, S. C. & Santelia, D. (2019). *BMC Evol. Biol.* **19**, 66.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. & Yutani, H. (2019). *J. Open Source Softw.* **4**, 1686.
- Xu, C., Park, J.-K. & Zhang, J. (2019). *PLoS Biol.* **17**, e3000197.
- Zeeman, S. C., Kossmann, J. & Smith, A. M. (2010). *Annu. Rev. Plant Biol.* **61**, 209–234.