New insight in cereal starch degradation: identification and structural characterization of four α-amylases in bread wheat

1 Introduction

With about 10¹⁴ plants grown each year around the world, wheat is the oldest and, along with rice and maize, the most extensively cultivated of all grain crops. Due to its unique properties, wheat (and bread) is deeply embedded into human culture and provides a major portion of caloric intake and nutrients required to support the global population [1].

Starch represents up to 70% of total wheat grain dry mass and is composed of 2 distinct fractions: the amyllopectin (highly branched, 75% of starch dry mass) and the amylose (mostly linear, 25% of dried mass). Both are made of α-1,4-glucosidic bond glucose residues, which can be branched via α-1,6-glucosidic linkages (for a review, see [2]). Starch is accumulated either in leaf during the day for transitory metabolism or during grain and tuber development, to accumulate energy. This polysaccharide must be degraded in order to release simple sugars. In fact, many biological and industrial processes, such as mammalian digestion, plant metabolism, biofuel production, fermentation or malting rely on the hydrolysis of native starch by amylolytic enzymes.

In cereals, the first discovery of an amylase was made by French chemists Payen and Persoz [3]. They isolated a “diastase” from germinating barley, which appeared to be an amylase complex, represented mostly by α-amylases, breaking down the starch macromolecules. α-Amylases are endo-hydrolases belonging to the glycoside hydrolase family GH13. However, no comprehensive study had been done so far to describe and catalogue all the wheat α-amylase isoforms, despite compelling information on the involvement of two α-amylases on economically important issues for the international cereal community, namely pre-harvest sprouting and late maturity α-amylase. This study describes for the first time the genomic localization, nucleotide and amino acid sequences, phylogeny and expression profile of all known α-amylases in wheat, including a hitherto unknown fourth isoform here designated as TaAMY4. Isoform profiling strongly suggested α-amylases to be working in partnership to achieve complete degradation of a starch granule, whereas expression profiling revealed a potential involvement of TaAMY4 in the late maturity α-amylase problem.

Keywords: wheat; starch; α-amylase; late maturity α-amylase; germination; surface binding site.
(GH) family GH13 [4]. This class of α-1,4-glucan hydrolytic enzymes has been thoroughly described from unicellular organisms to plants and humans (for a review, see [5]). In plants, the role and number of α-amylase isoforms vary significantly across monocots and dicots. To date, there are at least four α-amylase categories in barley, from HvAMY1 to HvAMY4, while rice shows at least ten separate genes clustered into three subfamilies (OsAMY1, OsAMY2 and OsAMY3) [6,7]. All subfamilies in barley or rice have been demonstrated to be expressed at different grain developmental stages and in various tissues [8,9]. More importantly, structure comparison of barley α-amylase isoforms 1 and 2 reveals particular features suggesting differences in sugar recognition and processing [10].

In wheat, three isoforms of α-amylase have been described to date. The two major isoforms have been isolated based on their pI since 1972 [11]. Multiple α-amylase 1 (TaAMY1) loci have been localized on the long arm of group 6 chromosomes and produce a high pI α-amylase in the aleurone layer of the mature grain. Its production is initiated at the latest phase of grain development during desiccation and close to maturity. This accumulation of TaAMY1 is a prerequisite in preparation for germination. Once germination is triggered, gibberellic acid hormones released from the embryos will activate the gene leading to occurrence of amylase activity to convert starch into rapidly metabolized sugar allowing coleoptile and root development [12].

Multiple α-amylase 2 (TaAMY2) loci are located on the long arm of the group 7 chromosomes and produce a low pI α-amylase in the pericarp of the developing grain [13]. Accumulation of this particular amylase is thought to increase from anthesis to reach its peak within 10 to 15 days, which coincides with the initiation of the grain filling stage. Then TaAMY2 expression is shown to decline slowly until grain maturity. It has been suggested that this particular α-amylase is involved in the degradation of the starch present in the green pericarp in order to fuel embryonic development and tissue differentiation [14]. However, its strong expression during germination suggested a significant role in grain germination [15].

Finally, a third isofrom α-amylase 3 (TaAMY3) is encoded by a single locus on group 5 chromosomes. The presence of its corresponding mRNA throughout the grain development suggested a role in grain development and maturation [16]. This enzyme indeed mainly appears during grain development in the pericarp, similarly to TaAMY2. Western-blot analysis suggested TaAMY3 is the predominant α-amylase throughout grain development [17]. However, its role during the grain development process remains unclear, as over-expression of this particular TaAMY3 in developing wheat endosperm showed a very limited impact on starch granule integrity despite elevated levels of α-amylase activity.

Unfortunately, wheat α-amylases have also been known among the cereal community for their involvement in two of the current major wheat genetic defects, pre-harvest sprouting (PHS) and late maturity α-amylase (LMA) (for a review, see [18]). LMA is a genetic defect causing expression of TaAMY1 in a subset of cells in the aleurone layer due to an environmental trigger (usually cold shock) during grain development. TaAMY1 protein in LMA induced grain remains present in the aleurone through harvest and negatively affects the falling number test [19], thus reducing the farmers’ income. In addition, breeders’ lines shown to be prone to LMA late in the breeding cycle represent lost investment of millions of dollars annually for breeding companies when either discarded completely or downgraded to feed grain varieties. The multigenic nature of the TaAMY1 locus has made difficult the attempts to generate genetic solutions to this critical issue. Despite some recent efforts to characterize the TaAMY1 locus, it still remains unclear how many TaAMY1 isoforms are actually involved in the LMA phenotype [20]. In an attempt to close this knowledge gap, we undertook a comprehensive analysis of all wheat α-amylase family.

In this study we present sequence, structural models, genomic localization and expression data of all known α-amylases in wheat, including a hitherto unknown fourth isoform, designated here as TaAMY4. We report the structural biology of the four wheat α-amylases including key motifs and class specific domains. Expression data were focused on developing and germinating wheat grain, and the role of each isoform in grain development and/or germination is hypothesized. Lastly, a potential role of TaAMY4 in the LMA phenotype is suggested and discussed.

2 Materials and methods

2.1 Plant rearing and sampling

Chara wild-type plants were grown in glasshouses at CSIRO Agriculture and Food, Canberra, Australia, under natural light on a diurnal temperature cycle of 14/20 °C. All leaf, developing and germinating seed samples were taken in biological triplicates (three separate plants) for each sampling point. Developing wheat grains were sampled at midday between 5 and 30 days post anthesis (DPA) at 5 day intervals using different wheat spikes each time. For germinating seeds, 20 dry seeds per plant were
sterilized using chlorine gas and germinated for up to 6 days on soaked tissue paper in plastic tubs in the dark. Samples were snap frozen into liquid nitrogen before being transferred to -80 °C until further processing.

A sub-population of 24 recombinant inbreds lines from the four-tall Multi-parent Advanced Generation Inter-Cross (MAGIC) population, as described previously [21], was used to investigate LMA. Since the Green Revolution and the introduction of the semi-dwarf GA-insensitive Rht1 gene, LMA expression evolved from constitutive to being stochastic and environmentally triggered complicating thus the LMA characterisation. The 24 recombinant inbred lines were selected based on their Tall phenotype to make any potential LMA phenotype constitutive. The population was founded by four Australian commercial cultivars: Baxter, Chara, Westonia and Yitpi. A field trial was conducted in Yanco (New South Wales, Australia) in 2010. RNA was extracted from milled dry seeds (wholemeal flour).

2.2 Sequencing, DNA and RNA extractions, and PCR analyses

Sequencing was performed on TOPO-cloned PCR products from proof-reading PCR reactions (Phusion, NEB) using standard BigDye v3.1 protocols. Chara wild-type and Chinese Spring nullisomic-tetrasomic DNA samples were already available in the laboratory [22]. Primers used for sequencing are listed in Table S1. Sequences were obtained for each isoform of wheat α-amylase in an extensive process of multiple sequencing rounds using a combination of TaAMY isoform and/or homeoform specific and non-specific primers, which were developed using a combination of already publicly available sequences – GenBank ([23]; https://www.ncbi.nlm.nih.gov/genbank/) or Wheat Expression Browser ([24]; http://www.wheat-expression.com/) – and sequences obtained in this study (for details, see Table S2). All obtained full-length gDNA sequences (containing a complete coding DNA sequence) were submitted to GenBank [23].

DNA for copy number analyses was extracted and homeoform-specific primer pairs for TaAMY4, and isoform-specific primers (but not homeoform-specific) pairs for TaAMY1 and 2, were developed and tested as described in Mieog et al. [25]. Plasmids containing clones for each homeoform of TaAMY4 as well as for Epsilon Cyclase genome A, obtained during sequencing procedures, were used to make a mixed plasmid sample, which was used as a calibrator sample for TaAMY4 homeoform copy number determinations. Plasmids containing clones for isoforms TaAMY1 and 2 as well as for Epsilon Cyclase genome A, were used to make a mixed plasmid sample for total TaAMY1 and 2 copy number determinations. Genome non-specific TaAMY1 and 2 primers together with a mixed plasmid calibrator sample were also developed for total copy number analyses for these genes. Copy number assays were performed on a PikoReal real-time PCR machine (ThermoFisher) using Sensifast SYBR kits (Bioline). A typical reaction consisted of 5 µL mastermix, 2.5 µL primer premix (1.4 µM each), and 2.5 µL DNA sample.

RNA was extracted using the protocol described in Whan et al. [17]. For developing seeds, RNA was extracted from frozen halfseeds with the embryo section removed. For germinating seeds, seeds with the embryo removed were freeze-dried and milled before RNA was extracted. Dry seed RNA was extracted using both methods. Seeds from the MAGIC tall LMA lines were milled whole and RNA was extracted from the wholemeal flour.

Amylase-isoform specific primers for real-time quantitative PCR (RT-qPCR) were developed for TaAMY1, 2 and 4 as previously described for qPCR [25]. TaAMY3 primers as well as primers for a reference housekeeping gene (actin) were already available [17]. Annealing sites of the four isoform-specific α-amylase primers pairs are shown in Figure S1. RT-qPCR runs were performed on the MyIQ real-time PCR system (BioRAD) using Sensifast one-step RT-qPCR kits containing a mastermix, reverse transcriptase and RNase inhibitor (Bioline). A typical reaction consisted of 10 µL mastermix, 4.6 µL primer premix (1.6 µM each), 0.2 µL reverse transcriptase, 0.4 µL RNase inhibitor and 5 µL RNA sample. An RNA sample with high levels of TaAMY1, 2 and 4 [5 days after imbibition (DAI)] was used for a dilution series to compare reaction efficiencies (already known for TaAMY3). Plasmids of TaAMY1, 2, 3, 4 and TaActin, obtained during sequencing or from cloned RT-qPCR products, were mixed in equal copy numbers and used as a calibrator sample to be able to directly compare TaAMY1-4 expression levels.

All primers used for qPCR (copy numbers) and/or RT-qPCR (gene expression) are listed in Table S3.

2.3 Sequence collection, evolutionary comparison and structure modelling

Available amino acid sequences were taken from a few previous bioinformatics studies [5,26-30] focusing mainly on the α-amylases from subfamilies GH13_6 (29 plant α-amylases accompanied by 2 bacterial representatives) and GH13_7 (10 archaeal and 2 bacterial sources). These 43 sequences, retrieved from the Uniprot knowledge database ([31]; http://www.uniprot.org/), were completed by the consensus protein sequence for each wheat
isoform. The protein consensus sequences were generated in the programme Geneious v.8.1.3 [32] by translating all available complete coding sequences, creating a Geneious protein alignment per isoform and extracting the consensus sequence. Thus, in total 47 α-amylase protein sequences were compared; the main emphasis being given on plant α-amylases from the subfamily GH13_6 [33].

The alignment was performed for the entire set of sequences using the program Clustal-Omega ([34]; http://www.ebi.ac.uk/Tools/msa/clustalo/) with no additional manual tuning of aligned sequences. The evolutionary tree was based on the alignment (including the gaps) of the sequence segment spanning almost the entire catalytic (β/α)8-barrel domain including the domain B from the beginning of the strand β1 (the region 8_FNW in the barley low pl α-amylase isozyme; UniProt accession No.: P00693; [35]) to the end of the strand β8 (the region 318_GIPCIFYDH in the same enzyme). The tree was calculated as a Phylip-tree type using the neighbour-joining clustering [36] and the bootstrapping procedure – the number of bootstrap trials used was 1,000 [37] implemented on a personal-computer running Clustal-X package [38]. The tree was displayed with the program iTOL ([39]; http://itol.embl.de/).

2.4 Structure modelling

Three-dimensional structure of the barley low pl α-amylase isozyme AMY1 solved as a complex with the substrate analogue, methyl 4I,4II,4III-trithiomaltotetraoside (thio-DP4) was retrieved from the Protein Data Bank (PDB; [40]; http://www.rcsb.org/pdb/) under the PDB code 1P6W [10]. All structural models of four representatives of wheat α-amylases, i.e. TaAMY1_A1, TaAMY2_A1, TaAMY3_A and TaAMY4_A, were created using the Phyre-2 server ([41]; http://www.sbg.bio.ic.ac.uk/phyre2/) employing the “Normal” modelling mode. Structures were superimposed using the program MultiProt ([42]; http://bioinfo3d.cs.tau.ac.il/MultiProt/) and displayed with the WebLab Viewer Lite programme (Molecular Simulations, Inc.). Ligand-binding site interactions were analyzed using 3DLigandSite – predicting ligand-binding sites using similar structures ([43]; http://www.sbg.bio.ic.ac.uk/~3dligandsite/). barley low pl HvAMY1 was used as reference model for the active site comparison.

2.5 Other bioinformatics tools

Amino acid sequence genesis, alignments, annotations and comparisons were performed according to recent study by Sethi et al. [44] as well as using the Geneious v.8.1.8 ([32]; http://www.genious.com/) and Swiss-Model ([45]; http://swissmodel.expasy.org/) programmes. The α-amylases from Arabidopsis thaliana (UniProt accession No.: Q8VZ56), Hordeum vulgare (P00693 and P04063), Oryza sativa (P27939 and A2YGY2), Spirocheta thermophilia (E0RN69) and Stigmatella aurantiaca (Q08YD2) served as the protein sequences used for comparison.

2.6 ELISA test

LMA testing was performed on protein extracted from 10 mg wholemeal flour according to the method described by Verity et al. [46]. All spectrophotometric measurements were performed using a Thermo Multiscan Spectrum plate.

3 Results

3.1 Sequencing of the TaAMY isoforms, gene copy numbers and identification of a fourth isoform of α-amylase in wheat

Full-length sequences were obtained from eight homeoforms of TaAMY1 (Accession Nos.: KY368728-35) and four from TaAMY2 (Accession Nos.: KY368736-39), which were confirmed to be located on chromosome groups 6 and 7, respectively. DNA copy number analysis supported the existence of 10 ± 3 and 6 ± 2 total copies for TaAMY1 and TaAMY2, confirming that these are multi-copy genes as indicated in literature. The results indicated that at least some of the TaAMY2 variants and potentially some TaAMY1 variants as well were missed in our sequencing efforts.

During the sequence alignment process, one set of sequences, aimed to sequence TaAMY1, formed its own cluster separate from any of the three α-amylase family isoforms already (partially) described, but was still clearly an α-amylase. Copy number analysis of this new isoform showed a single copy per genome on chromosome group 5, and after obtaining three full-length gDNA homeoform sequences and determination of a final gene model with an appropriate stop codon, the gene was named TaAMY4 (Accession No.: KY368740-42).

The full-length genomic sequences and coding sequences already publically available allowed us to circumscribe the intron-exon boundaries. TaAMY4 genes showed the presence of three introns. Intron 1 was located between nucleotide 94 and 217 of the genomic sequence, whereas introns 2 and 3 were located from nucleotide 351 to 465 and 1272 to 1372, respectively. The intro-exon structure was very similar to that of TaAMY2 structure,
whereas TaAMY1 and TaAMY3 showed the presence of only two introns (intron 1 and 3 equivalent).

The predicted cDNAs of TaAMY4, which corresponds to the protein sequence of 436 amino acid residues, is presented in Figure 1 and Table 1. The TaAMY4 amino acid sequence corresponds to a 47.6 kDa protein mass with a theoretical pI of 6, which classified it as high pI α-amylase according to the ExPASy proteomics server [47]. The overall protein structure is representative of the GH13_6 family corresponding to the α-amylase from the plant kingdom (for a review, see [29]) (Figure 1). TaAMY4 presents a secreting peptide (Gly2-Arg24), a main catalytic domain (domain A including the domain B) with a (β/α)-barrel fold (Leu28-Leu376) and a C-terminal domain C adopting an antiparallel β-sheet structure (Arg366-Lys426). TaAMY4 showed ten residues potentially involved in the active site, three catalytic residues and two calcium binding site residues characteristic of the subfamily GH13_6 (Figure 1).

### 3.2 TaAMY family protein characteristics

Using the ExPASy protein analyzing tools [47] on representative consensus sequences, we compared the protein size, mass and pI of the four TaAmy isoforms. In general, TaAMY1, 2 and 4 gave very similar results (Table 1). The size of the protein was between 436 and 439 amino acids with a molecular mass between 47 and 48 kDa. As mentioned previously, TaAMY4 showed pI values similar to TaAMY1 (pI = 6) and higher than the low pI TaAMY2 (pI = 5.5). All three isoforms also exhibited similar characteristics including 13 functionally important residues and two calcium binding sites (Figure 1).

![Figure 1](image-url)  
*Figure 1. Consensus amino acid sequence of TaAMY4. Genome specific amino acid variations are highlighted in white. Secondary structure elements (α-helix – pink, β-strand – green) are indicated above the consensus sequence, whereas the catalytic triad and other important active-site residues are signified in yellow and blue below the consensus sequence. Secreting peptide, catalytic domain A (including domain B) and domain C are indicated by the red, yellow and blue bar, respectively.*

<table>
<thead>
<tr>
<th>α-Amylase</th>
<th>cDNA length (nt)</th>
<th>Introns (cDNA position)</th>
<th>Amino acid length</th>
<th>Mw (kDa)</th>
<th>pI</th>
</tr>
</thead>
<tbody>
<tr>
<td>TaAmy1_a</td>
<td>1284</td>
<td>2 (87-88) (1032-1033)</td>
<td>437</td>
<td>47.25</td>
<td>6</td>
</tr>
<tr>
<td>TaAmy1_b</td>
<td>1278</td>
<td>2 (81-82) (1026-1027)</td>
<td>437</td>
<td>47.18</td>
<td>5.84</td>
</tr>
<tr>
<td>TaAmy1_d</td>
<td>1284</td>
<td>2 (87-88) (1032-1033)</td>
<td>437</td>
<td>47.32</td>
<td>6</td>
</tr>
<tr>
<td>TaAmy2_a</td>
<td>1314</td>
<td>3 (87-88) (220-221) (1032-1033)</td>
<td>439</td>
<td>47.8</td>
<td>5.6</td>
</tr>
<tr>
<td>TaAmy2_b</td>
<td>1317</td>
<td>3 (90-91) (223-224) (1041-1042)</td>
<td>439</td>
<td>48.02</td>
<td>5.5</td>
</tr>
<tr>
<td>TaAmy2_d</td>
<td>1314</td>
<td>3 (87-88) (220-221) (1032-1033)</td>
<td>439</td>
<td>48.32</td>
<td>5.47</td>
</tr>
<tr>
<td>TaAmy3_a</td>
<td>1242</td>
<td>2 (90-91) (981-982)</td>
<td>385</td>
<td>45.37</td>
<td>8.01</td>
</tr>
<tr>
<td>TaAmy3_b</td>
<td>1242</td>
<td>2 (90-91) (981-982)</td>
<td>385</td>
<td>45.56</td>
<td>8.63</td>
</tr>
<tr>
<td>TaAmy3_d</td>
<td>1242</td>
<td>2 (90-91) (981-982)</td>
<td>385</td>
<td>45.48</td>
<td>8.47</td>
</tr>
<tr>
<td>TaAmy4_a</td>
<td>1311</td>
<td>3 (93-94) (226-227) (1032-1033)</td>
<td>436</td>
<td>47.61</td>
<td>5.95</td>
</tr>
<tr>
<td>TaAmy4_b</td>
<td>1311</td>
<td>3 (93-94) (226-227) (1032-1033)</td>
<td>436</td>
<td>47.57</td>
<td>6.08</td>
</tr>
<tr>
<td>TaAmy4_d</td>
<td>1311</td>
<td>3 (93-94) (226-227) (1032-1033)</td>
<td>436</td>
<td>47.65</td>
<td>6.27</td>
</tr>
</tbody>
</table>
1). Overall, very few differences could be detected among these three isoforms (a single synonymous change) with the noticeable exception of the 11th residue participating in the active site (amino acid residue 231) (Figure 2). TaAMY1 and TaAMY4 displayed a polar neutral side chain amino acid (threonine and asparagine, respectively), whereas TaAMY2 and TaAMY3 both had an acidic amino acid (aspartic acid). In contrast, TaAMY3 showed very different features. The protein was much smaller with only 385 amino acid residues and a 45 kDa molecular mass. The ExPASy protein analyzing tool gave a very high pI (pI = 8.5). These results were confirmed using the Phyre-2 server [41]. This size reduction did not seem to impact on the number of active site residues, nor calcium binding sites but did have an effect on the positioning of the residues along the amino acid sequence including a reduced distance between the 11th and last two active site residues (Figure 2).

3.3 TaAMY family protein structures

The tertiary structure of four representatives of each wheat α-amylase isoform were modelled and the results confirmed both the fold and domain arrangement typical for family GH13 members [5]: (i) the catalytic (β/α)8-barrel domain; (ii) domain B protruding out from the barrel between the strand β3 and helix α3; and (iii) domain C succeeding the barrel (Figure 3). Although they all most resemble the structures of both barley α-amylase isozymes [10,48], the structure of the low pI isozyme (Figure 3a) was used for comparison. At a more detailed view of the overall structures, it is evident that the most striking difference concerns the helix α6, where one of the two thio-DP4 fragments is bound (Figure 3b). There is a tryptophan pair responsible for this binding in both barley α-amylases [10]. Interestingly, of the four wheat α-amylase isoforms, the corresponding tryptophan pair is present only in members of the TaAMY1 and TaAMY2 isoforms; the members of the isoform TaAMY3 possess only the first of the two adjacent tryptophans, whereas the members of the isoform TaAMY4 lack both tryptophans (Figure S2). This presence or absence of one or both tryptophans may be used as a feature to potentially catalogue the individual cereal α-amylase isoforms. In a wider sense, this feature can also indicate different functionality due to a different surface-binding site (SBS) (i.e. the ability or inability to bind a saccharide) among plant α-amylases [49].

The further point of interest was domain C (Figure 3c), where a very specific difference between the two barley α-amylase isozymes was revealed as the so-called ‘pair of sugar tongs’ binding site [10]. In addition to Tyr380, which is the crucial residue responsible for binding in this SBS, Ser378 was proposed as the major determinant of binding by virtue of keeping the relevant loop flexible in the low pI isozyme AMY1 (Figure 3d). In the high pI isozyme AMY2, despite the presence of the corresponding tyrosine (Tyr378), the serine was replaced by a proline (Pro376) that contributes to a more restrained loop and consequently may prevent the binding [10]. Of the four α-amylase isoforms from wheat, it is very probable that isoforms TaAMY2 and TaAMY3 can undertake this “sugar-tongs”-like surface binding activity, whereas for the isoforms TaAMY1 and TaAMY4 the binding would be hindered due to serine to proline replacement (Figure 3d,e), as observed in the high pI isozyme of barley α-amylase [50].

3.4 Evolutionary relationships

The evolutionary relatedness of wheat α-amylases determined in the present study to other representatives of plant α-amylases from the subfamily GH13_6 as well as their mutual relationships to each other are depicted in the evolutionary tree (Figure 4). For comparison and in order to get a wider picture, a few bacterial
Characterisation of the bread wheat α-amylase family

but indicative differences within their conserved sequence regions [4,26,29].

In the evolutionary tree (Figure 4), the four consensus wheat α-amylases from the present study reflect their assignment to four distinct wheat α-amylase isoforms. Based on the observed relationships, the members of isoforms TaAMY1 and TaAMY2 may very probably represent high and low pl α-amylase isozymes, respectively, since they are grouped with well-characterized α-amylase isozymes from barley, i.e. TaAMY1 with high pl isozyme representatives previously classified in the original plant subfamily GH13_6 were added together with their prokaryotic counterparts from the most closely related subfamily GH13_7 [26], covering only the α-amylases from archaeons and bacteria [29]. In agreement with the CAZy classification [51], the two GH13 subfamilies are clearly separated from each other. Within the two subfamilies also the α-amylases originating from two different taxa are clustered separately (plant vs. bacteria for GH13_6 and archaeons vs. bacteria for GH13_7) reflecting also subtle

Figure 3. Comparison of tertiary structures of plant α-amylases from the subfamily GH13_6 with focus on new α-amylases from wheat. (a,b) Overall structures shown as solid ribbon models signifying catalytic (β/α)8-barrel domain (blue), domain B (red) and domain C (green) protruding out of and succeeding, respectively, the barrel domain. (a) Barley low pl α-amylase isozyme AMY1 complexed with its substrate analogue thio-DP4 (PDB code: 1P6W; [10]). Two thio-DP4 fragments are seen at the surface of catalytic domain, whereas the entire thio-DP4 molecule is bound to the domain C. The catalytic triad is indicated inside the barrel domain together with one calcium ion (blue sphere) as well as three additional calcium ions present within the domain B (red spheres). (b) Structural models of representatives of four wheat α-amylase families shown in a similar orientation as that of the barley α-amylase structure. All four models were obtained at the fold recognition server Phyre-2 [41] using the low pl α-amylase isozyme AMY1 (PDB code: 2QPU; [50]) as the best template in each case at the confidence equal to 100, sequence identity no lower than 64% and at least 91% alignment coverage. (c-e) Structural overlay of domain C emphasizing the ‘pair of sugar tongs’ site identified in the barley α-amylase AMY1 and corresponding residues in the models of four wheat α-amylases. The models of separated domain C were obtained at the fold recognition server Phyre-2 [41] using the template of barley α-amylase high pl isozyme AMY2 (PDB code: 1AVA; [71]) for the wheat α-amylase from the family Amy1, whereas for wheat α-amylases from families Amy2, Amy3 and Amy4, the barley α-amylase low pl isozyme AMY1 (PDB code: 1HT6; [10]) was used. The individual α-amylases are coloured in black, red, magenta, blue and cyan for barley and wheat Amy1, Amy2, Amy3 and Amy4, respectively. (c) Superimposed domains C of barley α-amylase AMY1 (1P6W) and the counterparts from four wheat families (shown as solid ribbon models). (d) The same picture of domains C (shown as line ribbon models) illustrating the ‘pair of sugar tongs site’ in the barely α-amylase with displayed thio-DP4 (scaled balls and sticks, coloured as elements) and the crucial binding residue Tyr380 supported by Ser378 (both shown as sticks, coloured black). Corresponding residues present in the four wheat α-amylases are shown for comparison. (e) The same picture displayed in stereo and seen from a different perspective.
AMY2 (UniProt: P04063; [52]) and TaAMY2 with low pI isozyme (UniProt: P00693; [35]). Similar feature can probably be assigned to the two α-amylases from rice (UniProt accession Nos.: P17654 and A2YGY2) positioned near both TaAMY1 and TaAMY2, respectively (Figure 4). The members of the isoform TaAMY3 form also a separate cluster together with their counterparts from rice and maize; the members of the totally novel isoform TaAMY4 being positioned on the adjacent branch (Figure 4). These results clearly show that the four wheat α-amylase isoforms can be divided into two, evolutionary separate, groups, with the two multi-copy genes (TaAMY1 and
TaAMY2) on one side and the two single-copy genes (TaAMY3 and TaAMY4) on the other side.

The present phylogenetic analysis, however, delivers an interesting observation concerning the α-amylases from Arabidopsis. There are three α-amylase isoforms (AtAMY1-AtAMY3) in the Arabidopsis thaliana [28,53,54] forming obviously their own clusters in the evolutionary tree (Figure 4) together with their homologues from, e.g., Vigna angularis, Vitis vinifera and Camelia sinensis. Of the three Arabidopsis α-amylase isoforms (AtAMY1 – UniProt accession No.: Q8VZ56; AtAMY2 – Q8LFG1; AtAMY3 – Q94A41), just the AtAMY1 with its cluster seems to be a homologue of all four wheat α-amylase isoforms, while the two mutually neighbouring clusters of both AtAMY2 and AtAMY3 are placed unambiguously far away (Figure 4).

### 3.5 TaAMY4 co-expresses with TaAMY1 including within LMA affected grains

Expression of the 4 types of α-amylases were monitored during development from five DPA up to six DAI, as shown in Figure 5. Each TaAMY isoform shows a distinct expression profile. During grain development, TaAMY2 was the most expressed isoform, especially around 10 days following anthesis (the cellular differentiation stage), while TaAMY3 exhibited its highest expression (which is still low compared to the other isoforms) between 10 and 30 DPA (but note a dip at 25 DPA). Both isoforms decreased at the final seed maturation stage (grain desiccation). In contrast, TaAMY1 had very low expression levels, while TaAMY4 showed a low but stable expression during grain development.

During germination, TaAMY1 was quick to react to the imbibition and showed a strong and rapid increase in expression from 1 DAI onwards. TaAMY2 showed a strong but delayed increase in expression compared to TaAMY1, only showing a strong increase from 3 DAI onwards. TaAMY3 did not seem to be involved in the germination process as the (relative) transcript level dropped dramatically after imbibition. The transcript level of TaAMY4 increased during germination as early as 1 DAI but remained at a lower level than either TaAMY1 or 2.

A set of 24 “Tall” genotypes were selected from the MAGIC population to study the link between the four α-amylase isoforms and LMA. The rationale of using tall genotypes was to avoid the impediment of semi-dwarf GA-insensitive Rht1 gene on LMA expression and to make

![Figure 5](image-url)
any potential LMA phenotype constitutive. The presence of LMA phenotype was determined using the ELISA based high pI α-amylase detection kit [46]. LMA values in the wholemeal flour ranged from 0.17 to 0.87. These ELISA values correspond to a falling number equivalent of 350 to 250 according to Verity et al. [46] with 300 being the threshold for quality. The TaAMY1 expression level exhibited the highest correlation (0.65) with the LMA ELISA test values (Figure 6). Additionally, a weaker correlation was found between TaAMY4 and LMA (0.35) and between TaAMY1 and TaAMY4 (0.61). Expression levels of TaAMY2 and 3 were much lower than TaAMY1 and 4, and no meaningful correlation could be found between these and the LMA signals.

4 Discussion

The α-amylase family has always shown a high level of complexity and multiplicity throughout the plant kingdom from the Prasinophyceae to the Poaceae [5]. Even in the case of unicellular green algae, like Ostreococcus tauri, the presence of three different isoforms of α-amylase despite reductions occurring within other important metabolic processes, such as cell cycle control, illustrated this high level of complexity [55,56]. In our analysis, we were able to inventory four isoforms, on each of the three wheat genomes and a total of at least 21 homeoforms, although some variants are likely to have been missed. Previous work already suggested such a level of complexity among the α-amylase family. Cheng et al. [20] suggested the presence of at least 36 TaAMY1 expressed genes in wheat using the comparison with barley sequence, and relative expression abundance of high pI amylase transcript level. Our genomic characterization described the presence of at least nine different TaAMY1 homeoforms among the three wheat genomes confirming a high level of complexity, however, our data have also suggested that the estimation of Cheng et al. [20] may be too high.

The question of the absolute requirement of such level of complexity needs to be raised. All α-amylases predominantly hydrolyze in an endo-fashion the α-1,4-glucosidic linkages in starch or related carbohydrates into smaller polysaccharides and oligosaccharides, so why are so many isoforms present to fulfill the same role [5,14]? This multiplicity could imply subtly different functions among the α-amylase isoforms. Evolutionary relationships clustered closely TaAMY1 and TaAMY2 with barley or rice homologues AMY1 and AMY2. Both TaAMY1 and TaAMY2 share over 75% identity between their amino acid sequences and over 90% with their barley counterparts. TaAMY4, and to a lesser extent, TaAMY3

![Figure 6. Correlation between TaAMY1-4 gene expressions (RT-qPCR) expressed in log scale and LMA scoring (ELISA) using wholemeal flour samples from selected MAGIC constitutive LMA Tall lines based on LMA responses.](image-url)
are closely related to rice OsAMY3 or Arabidopsis AtAMY1. Interestingly, OsAMY1 and OsAMY3 showed clear distinct roles in rice grain germination [9]. While OsAMY1 seems to be constitutively expressed during grain germination in both embryo and aleurone layer of the rice grain, OsAMY3 expression is enhanced during germination mainly under anaerobic conditions. Additional studies confirmed OsAMY3 to be involved in numerous stress response mechanisms including anaerobic but also sugar starvation [57]. AtAMY1 has also been described as being expressed during biotic or abiotic stress conditions and during leaf senescence. Doyle et al. [58] suggested a recycling role of AtAMY1 to ensure all remaining starch is scavenged rather than wasted during biotic stress or leaf cell death. All 4 TaAMYs were clustered within a group including AtAMY1 and clearly separated from a second group involving AtAMY2 and AtAMY3 (Figure 4). Tyler et al. [28] pointed out that the AtAMY gene expansion was the result of a recent evolutionary duplication. However, the observed absence of AtAMY2 and AtAMY3 like α-amylases in wheat might suggest an even more recent loss of specific α-amylases and should deserve further attention.

TaAMY1 and TaAMY2 appeared to share common specific structural features with the barley HvAMY2 and HvAMY1, respectively, but only TaAMY2 (and TaAMY3) have indications (Figure 3d,e) of the presence of an extra SBS, i.e. SBS2, called the pair of sugar tongs [10]. Neither TaAMY1, nor TaAMY4 displayed the presence of this sugar tongue site suggesting a different level of efficiency or a different target. This particular SBS has been extensively characterized in barley [10]. It seems to be instrumental in lowering the enzyme $K_m$ by anchoring the glucan substrate near the catalytic site thus enhancing enzymatic activity [59]. Nielsen et al. [59] also suggested a second, minor role of this SBS in binding chaperone molecules to stabilize the enzyme folding. Very recently, Cockburn et al. [49] proposed a hypothesis that SBS1 and SBS2 had synergetic roles in binding A-type starch glucan chains with SBS1 being more effective at binding small chain motifs at the surface of the granule, while the SBS2 binds glycosyl residues near branch points to direct and feed long glucan chains to the catalytic site. Glucan chains are classified by their level of branched points and position within the crystalline structure [60]. To summarize – C type chains hold the unique reducing end of the starch molecule; B type chains contain several branch points and can be subclassified based on their glucan chain length; and A types are the small and single branched chains localized on the outskirt of the carbohydrate molecule. A type chains usually contains 12-16 glucose residues, whereas B type chains are known to consist of at least 20 glucose residues [61]. It is therefore tempting to speculate on the roles of each of the α-amylase clusters. TaAMY1 and TaAMY4 could potentially be more efficient on numerous, small and easily accessible short glucan chains at the surface of the granule, while TaAMY2 and TaAMY3 could present a higher efficiency on B-type chains allowing entry points to their counterpart in order to complete the starch digestion process. This hypothesis correlates with the results obtained with our TaAMY3 overexpression study, where only slight signs of degradation could be seen on the granule despite a 30-fold increase in total amylase activity in dry grain associated with a very rapid degradation of starch during rheological analysis when alternative α-amylases are activated. Another hypothesis would involve some sort of synergism, in which TaAMY3 and TaAMY4 scavenge partial degradation products produced by TaAMY1 and TaAMY2 to ensure complete degradation of the carbohydrate reserve. In Arabidopsis, AtAMY3 has been identified releasing small branched glucan chains that, due to the presence of α-1,6-glucosidic bonds nearby the α-1,4-glucosidic bonds, are undigested by β-amylase, but can further be degraded by limit dextrinase or isoamylase [62,63].

Temporal profiling of transcript levels suggested differences in roles for each of these catalytic enzymes during grain development and/or grain germination. The apparent leading roles of TaAMY2 and TaAMY3 during grain development is consistent with previous work described in barley, where two α-amylase type enzymes were also described in the pericarp of developing grain, the low pl HvAMY1 and HvAMY4. It was suggested that, similarly to transitory starch metabolism, starch turnover was required in the green tissue of developing grain to avoid carbon starvation and to optimize the energy supply of developing tissues [64]. The indicative presence of a sugar tongue-like SBS in both TaAMY2 and TaAMY3 may suggest some differences in the carbohydrate crystal structure generated during the early phase of grain development compared to the starch structure present in desiccated endosperm. Transitory starch molecules present in green tissue contain generally shorter average glucan chains associated with a low amylase ratio and differ from reserve starch, where longer glucan chains and higher amylose contents are present [56]. We can speculate that with the presence of the SBS2, TaAMY2 and TaAMY3 would be more prone to digest transitory type starch from green scutellum, while an extra sugar tongue-like SBS would not be required for the degradation of reserve starch occurring at the early stage of germination.

Three of the four α-amylase isoforms strongly increase in expression during germination with TaAMY1
and TaAMY2 being the most abundant and TaAMY1 and TaAMY4 the most prone to respond to imbibition. The involvement of the majority of the α-amylase isoforms in the germination illustrated the absolute requirement for the embryo to access energy stored within the starchy endosperm to allow rapid metabolic functions, such as root and coleoptile growth. Within the different isoforms and especially for TaAMY1 and TaAMY2, the high number of homeoforms suggests a critical function for grain germination and development. At the physiological aspect, this could indicate a potential redundancy in the function of each of these α-amylases. Such redundancy is not uncommon in starch metabolism among the plant kingdom. α-1,4-Glucanotransferases or starch synthases have exhibited partial redundancies between isoforms allowing to partial complementation of mutations and therefore maintaining starch crystal integrity to some extents [65]. In barley, similar partial redundancy was described for the starch branching enzyme class 2, SBEIIb mutation being remained asymptomatic due to partial complementation by SBEIIa [66].

During PHS, the full suite of amylolytic enzymes including TaAMY1 and TaAMY2, and proteolytic enzymes are produced to mobilize both C and N reserves for the germinating embryo. LMA, on the other hand, is a genetic defect causing stochastic expression of the high pI α-amylase (TaAMY1) in a subset of cells in the aleurone layer due to a cold shock during grain development. Recent results from our laboratory have shown that over-expression of TaAMY3 to levels of up to 30-fold above those present in the endosperm of the wild-type wheat actually enhanced loaf volume and colour in small scale baking tests [67]. Our current results suggest that not only TaAMY1 but also TaAMY4 may be involved in the LMA issue. As expression profiling throughout the grain lifecycle suggested, both TaAMY1 and TaAMY4 are mostly expressed during germination. Using a subpopulation of wheat tall (non rht mutant) lines, we could find a clear correlation between LMA detection and TaAMY1 expression. This is very much in accordance with results found by Mrva and Mares [68]. According to our data, TaAMY4 was the only other isoform co-expressing with TaAMY1 and increasing with LMA expression. If, as we hypothesized, α-amylase isoforms have complementary functions, it is logical to assume that both TaAMY1 and TaAMY4 would also be co-expressed within the LMA context to achieve the so far non-elucidated metabolic role.

The current strategies to mitigate and, if possible, eliminate LMA revolve around efforts to understand the genetic basis of LMA. This is achieved through identification of quantitative trait loci (QTL) – defined as stretches of DNA linked to the genes that underlie a quantitative trait – that control the expression of LMA. The aim of this work is to develop markers that allow breeders to make selections earlier in the breeding cycle of potentially LMA-susceptible lines and remove them, allowing thus resources to be focused on lines that have a higher probability of passing the LMA screen. To date, one major QTL has been found on chromosome 7B accounting for 31% of the variation, and minor QTLs were detected on 3B, 3A, 2D and 6B with 3B being the most significant accounting together for 13% of the variation [69]. Although very useful in providing short-term solution to circumscribe the LMA issue, QTL identification is unlikely to provide a long-term solution to eradicate this issue, especially if numerous genes including various α-amylases are involved.

If LMA/PHS is detected using the falling number test, there is a potential $20-50/t penalty to growers due to downgrading superior milling wheat classes to feed grade [70]. Reputation can also be damaged if grains affected by LMA are traded in domestic and international markets. Breeders are in a race to introduce new high yielding varieties in order to maintain competitiveness in the international export market. As a consequence, growers might not have access to potentially high yielding lines that are not released due to LMA risk. The elimination of high yielding varieties represents a huge economic loss to breeders and growers. Understanding the physiology behind the trait will therefore provide the necessary information required to develop the appropriate tools to solve this economically important issue. While the importance of TaAMY4 in the LMA phenotype remains to be confirmed, this new information represents an additional step toward the complete elucidation of the LMA issue.

Acknowledgments: The authors acknowledge Andrew Warden and Phil Larkin for their fruitful discussions and recommendation on protein structure. SJ thanks the Slovak Grant Agency VEGA for the grant No. 2/0146/17.

Conflict of interest: The authors declare no conflict of interest.
References


Radchuk V.V., Borisjuk L., Sreenivasulu N., Merx K., Mock H.-P., Rolletschek H., et al., Spatiotemporal profiling of starch biosynthesis and degradation in the developing barley grain, *Plant Physiol.*, 2009, 150, 190–204.


Mrvka K., Mares D.J., Quantitative trait locus analysis of late maturity α-amylase in wheat using the doubled haploid


**Supplemental Material:** The online version of this article (DOI: 10.1515/amylase-2017-0004) offers supplementary material.