The BURP domain protein family

of *Arabidopsis*: a novel component

related to seed development

Dissertation
zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der

der Mathematisch-Naturwissenschaftlich-Technischen Fakultät

(mathematisch-naturwissenschaftlicher Bereich)

der Martin-Luther-Universität Halle-Wittenberg

vorgelegt von

Van Son Le

geb. 22. Januar 1967 in Hatinh, Vietnam

Gutachterin bzw. Gutachter:

1. Prof. Dr. U. Wobus, Gatersleben

2. Prof. Dr. R. Klösgen, Halle

3. PD Dr. G. Hinz, Göttingen

Tag der Verteidigung: 8. 12. 2005, Halle (Saale)

List of the abbreviations

µ	micro
ALP	Alkaline phosphatase
ATP	Adenosine triphosphate
b, bp	base, base pair
BCIP	5-Bromo-4-chloro-3-indolyl phosphate
BSA	Bovine serum albumin
CaMV	Cauliflower mosaic virus
cDNA	complementary DNA
DEPC	Diethyl pyrocarbonate
DNA	Deoxyribonucleic acid
EDTA	Ethylendiamine tetraacetic acid
et al.	*et al.* (and others)
g	gram
h	hour
IPTG	Isopropyl-ß-D-thiogalactoside
kD	kilo Dalton
L	liter
M	molarity
m	mili
min	minute
MOPS	N- morpholinopropanesulfonic acid
mRNA	messenger RNA
n	nano
NBT	Nitro blue tetrazolium
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	polymerase chain reaction
PEG	Polyethyleneglycol
RNA	ribonucleic acid
SDS	Sodium dodecyl sulfate
T-DNA	transfer DNA
Tris	Tris-hydroxymethylaminomethane
v/v	volume/volume
w/v	weigh/volume
WT	wild type

Table of content

1. Introduction

Plant seeds have evolved to nourish, protect and distribute the next generation and are the basis for the evolutionary success of spermatophytes. Plant seeds have been widely used to study plant specific processes of development and differentiation. Primarily however, they are the main source of human nutrition and animal feeding. With an increasing world population and the critical need for balanced nutrition, the molecular biology and biochemistry of seed storage compounds as well as the cellular and physiological mechanisms regulating their synthesis, deposition and mobilization have become a major focus both of basic and applied research approaches (Shewry *et al*., 1995).

Whereas applied research projects concentrate on crop plants, basic molecular and genetic research is often performed on the model plant *Arabidopsis thaliana* (L.) Heyn., a member of the *Brassicaceae* family. The experimental advantages of this plant species include the small genome size with known sequence, a rapid life cycle compared to most crop plants, space-saving cultivation, self-fertilization, prolific seed production, the availability of a large number of mutant stocks, the possibility to carry out whole mount analysis of developing seeds as well as efficient and well established methods of genetic transformation (Meyerowitz, 1989; Meyerowitz *et al*., 1991; Pyke, 1994).

1.1. Embryogenesis and seed development

Embryogenesis and seed development are initiated by the process of double fertilization. During this process, the pollen tube grows down through the style to enter the ovule through the micropyle and delivers two haploid nuclei. One of them fuses with the nucleus of the egg cell to produce a diploid zygote that will develop into the embryo. The other sperm nucleus fuses with the 2 polar nuclei of the embryo sac, forming a triploid endosperm nucleus. This nucleus divides to form triploid endosperm tissue, which will function as a nutrient source for the developing embryo. The protective seed coat is derived from maternal tissue and surrounds both the embryo and endosperm during embryogenesis.
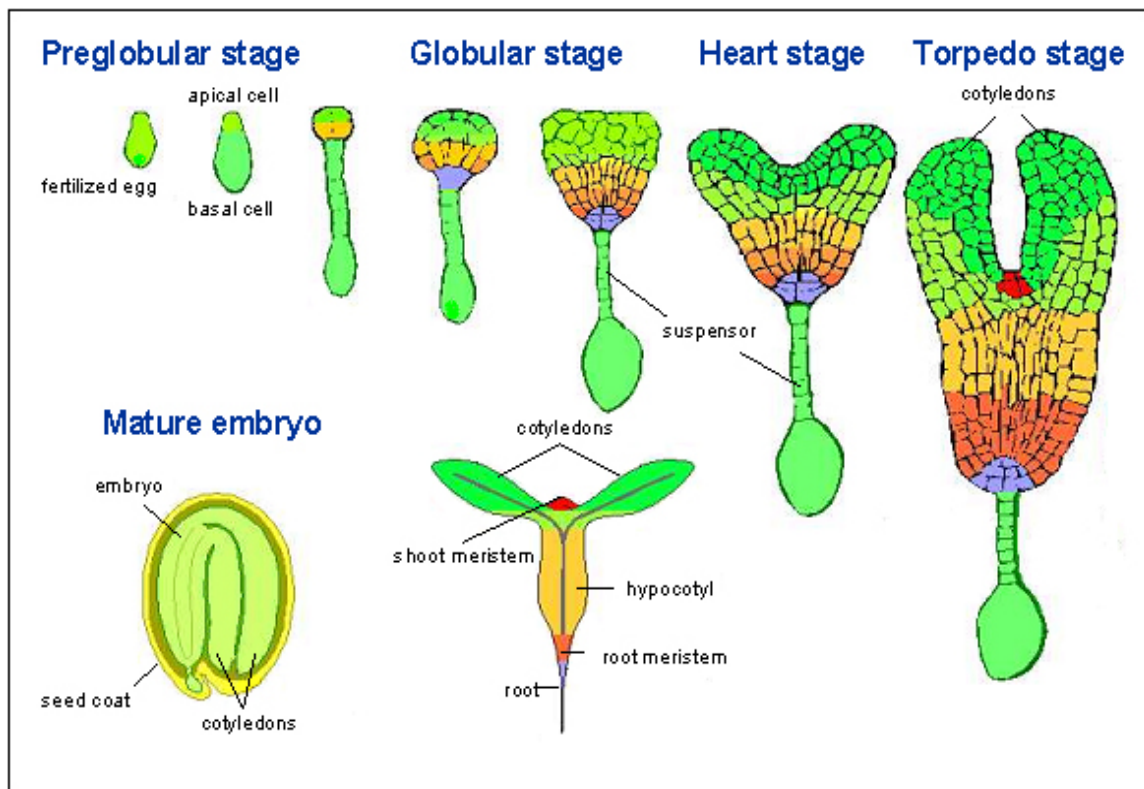
Figure 1.1. A schematic representation of different embryogenesis stages showing a progression from the preglobular stage through maturation (Wolpert, 1998).

As shown in Figure 1.1, the embryo passes through four developmental stages after fertilization. The globular stage is the pattern formation, in which the

axis of the plant body is defined, tissue layers organized, and earliest organs established. The heart stage is the cell diversification and specification phase, in which cell types such as the suspensor, provascular tissue, shoot and root meristems are defined. The torpedo stage is growth and morphogenesis, in which cells become expanded. During the last process, the embryo enters into maturation, in which cell division is completed, embryo storage reserves as proteins, starch, and lipids accumulate and the embryo acquires dormancy and desiccation. The first three stages occur concurrently in the developing embryo and are also known as early embryogenesis, while maturation is a distinct process that begins later in embryogenesis (Goldberg *et al*., 1988; Drews and Goldberg, 1989; Bewley and Black, 1994; Meinke, 1994; Harada, 1998).

During the globular stage, the embryo shows radial symmetry. Through a series of regular cell divisions an outer protoderm layer is produced and two layers of inner cells with distinct developmental fates are established. The apical layer will produce cotyledons and shoot meristem, while the lower layer produces the hypocotyls and root meristem. In the heart stage, the organs such as cotyledons, root meristem and provascular tissue are enlarged. At the end of heart stage, the shoot apical meristem, a highly organized group of quasi-embryonic cells, will give rise to the above-ground structures of the plant after germination. During the torpedo stage of embryogenesis, the embryo completes its growth and morphogenesis, elongating and enlarging to fill the seed. This stage is characterized by greening, rapid cell division, and cotyledon expansion (Goldberg *et al*., 1988; Meinke, 1994;).

Some recent studies subdivided the later stages of embryogenesis into two different phases, including mid-embryogenesis and late embryogenesis. During mid-embryogenesis, the seeds accumulate the macromolecular storage products, including lipids, protein and carbohydrates (Bewley and Black, 1994). Late embryogenesis is characterized by the arrest of tissue growth and development, the induction of dormancy and the acquisition of desiccation tolerance. Seed development and embryogenesis stop as the seed becomes

dormant and loses ~90% of its water. This decrease in water content presumably results from the severing of the vascular connection between the seed and the fruit and evaporative drying (Harada, 1998).

1.2. Accumulation of storage compounds in seed maturation

Accumulation of storage lipids

In most plant seeds, storage lipids are in the form of triacylglycerols (TAG), which is composed of 3 fatty acyl chains connected to a glycerol backbone by ester bonds. Storage lipid is synthesized in two stages in developing seeds, firstly through the production of acyl chains by the plastids, followed by their sequential incorporation into glycerolipids by the acyltransferases of the endoplasmic reticulum (Ohlrogge and Browse, 1995). Most of the biochemical steps are known and many of the genes involved have been identified (Beisson *et al*., 2003). The fatty acid composition of seed oil varies considerably both between species and within species, with fatty acids varying in both chain length and degrees of saturation. Genetic approaches to investigate the regulation of oil content have been performed with limited success. Mutant screening of *Arabidopsis* populations have identified the *triacylglycerol biosynthesis defect 1* (*tag1*) and *wrinkled 1* (*wri1*) loci as causing reduced seed oil content (Katavic *et al.,* 1995; Focks and Benning, 1998).

The storage lipids deposited in discrete organelles called lipid bodies, oleosomes or spherosomes (Herman, 1995). The lipid bodies are coated by a monolayer of phospholipids with embedded specialized proteins called oleosins (Huang, 1994). The oleosin proteins comprise the main component of oil body associated proteins. Although oleosins are involved in oil storage, they are unlikely to play a role in oil synthesis. Since oleosin genes are expressed rather delayed with respect to oil accumulation in seeds (Kater *et al*., 1991) they are thought to prevent the collapse of oil bodies during desiccation (Murphy, 1993). Oleosin molecules contain a lipase-binding site, which might help to position lipases close to its substrate during germination, when the oil bodies serve as an

energy source for the seedling. Thus oleosins are believed to interact with lipases to initiate the breakdown of the triacylglycerols (Huang, 1996).

Carbohydrate accumulation

Starch is the carbohydrate most commonly found in seeds. It is stored in seeds in two related forms, amylose and amylopectin; both are polymers of glucose. Whereas amylose is a straight-chain polymer including some glucose units, amylopectin is a multiple-branched molecule consisting of many amyloses. Starch accumulates in the endosperm, where cells are packed with starch granules that form within the amyloplasts. Starch is also deposited in the embryo and pericarp early in the development. Size and form of starch grains can vary with the species, organ and stage of development (Bewley and Black, 1994; Sivak and Preiss, 1995). The influence of starch synthesis on storage protein accumulation may also operate through changes in osmotic potential, which may influence the production of particular storage proteins (Turner *et al*., 1990). Several mutants with primary effects on starch biosynthesis have multiple effects on other aspects of storage organ development. A mutation at the *r* gene, which encodes a starch-branching enzyme of pea, confers a wrinkled phenotype on the seed and results in the production of relatively more lipids and less storage protein (summarized in Bhattacharyya et al, 1990).

Seed development of *Vicia faba* is under metabolic control and regulated by sugars (Weber *et al*., 1997). Sucrose has a dual function as transport and nutrient sugar and as a signal molecule triggering storage-associated processes. Sucrose acts on transcriptional and posttranscriptional levels, thereby affecting carbon fluxes. It is a key player within the regulatory network controlling seed differentiation (Weber *et al*., 2005). In the cell division phase (IV) the embryo has a high hexose status. During the late stage, hexoses decrease and cotyledonary differentiation and starch accumulation are initiated (Wobus and Weber, 1999). Thus, the initiation of maturation and accumulation of storage products are triggered by a complex regulatory network. This includes transcriptional and

physiological reprogramming mediated by sugar and hormone-responsive pathways (Wobus and Weber, 1999; Weber *et al*., 2005).

Seed storage protein accumulation

Based on the solubility of the proteins, Osborne (1924) classified the seed proteins into 4 groups: 1) albumins are soluble in water; 2) globulins are soluble in salt solution but insoluble in water; 3) prolamins are characterized by solubility in alcohol/water mixtures and 4) glutelins are soluble in diluted acid and alkalis.

Based on molecular and biochemical as well as functional features seed proteins can be classified into at least three groups (Shewry and Casey, 1999).

Firstly, the genuine storage proteins, which are of particular importance because they determine not only the total protein content of the seeds but also its quality as food and provide storage of amino acids as a source of nitrogen and carbon skeletons for use during germination and seedling growth (Higgins, 1984; Shewry *et al.,* 1995). The major groups of storage proteins fall into all four of Osborne's solubility fractions. The expression of storage proteins encoding genes is spatially and temporally regulated, being restricted to a specific storage tissue in the seed, and to a precise stage of development. Seed storage protein gene expression is also regulated by nutrition and environmental conditions as temperature, light, etc. Thus, storage protein synthesis may be for instance restricted under conditions of low sulphur availability (Shewry and Casey, 1999).

Secondly, the structural and metabolic proteins or "housekeeping" proteins are essential for the growth and structure of the seed.

Thirdly, protective proteins may play a role in providing resistance to microbial pathogens, invertebrate pests or desiccation.

1.3. Genetic regulators of seed development

The regulatory mechanisms that coordinate the various developmental events of mid-embryogenesis and late embryogenesis remain largely unknown,

although several regulatory genes have been identified, including for instance the *VIVIPAROUS1 (VP1)* gene in maize (McCarty, 1995) and the genes *ABSCISIC ACID INSENSITIVE3 (ABI3), LEAFY COTYLEDON (LEC1* and *LEC2)*, and *FUSCA3 (FUS3)* in *Arabidopsis* (Bäumlein *et al.*, 1994; Koornneef *et al.*, 1984; Meinke *et al.*, 1994; West *et al.*, 1994; Luerßen *et al.*, 1998; Stone *et al.*, 2001). These genes play an important role in controlling mid- and late embryogenesis, including the regulation of the seed storage protein gene expression in *Arabidopsis* (Nambara *et al.*, 1995; Parcy *et al.*, 1997; Kagaya *et al.*, 2005a, 2005b). In *Arabidopsis*, FUS3 and LEC2 control various processes of seed maturation such as accumulation of storage compounds, developmental arrest of the embryo, dormancy and desiccation tolerance (Bäumlein *et al.*, 1994; Meinke *et al.*, 1994; Keith *et al.*, 1994; West *et al.*, 1994; Parcy *et al.*, 1997; Raz *et al.*, 2001). The ABI3, LEC2 and FUS3 proteins contain a conserved B3 domain, which represents a DNA-binding motif unique to plants. The B3 domain is essential for the regulation of seed maturation genes (Giraudat *et al.*, 1992; Luerßen *et al.*, 1998; Stone *et al.*, 2001). FUS3 and ABI3 bind to the RY elements present in seed protein gene promoters (Ezcurra *et al.*, 2000; Reidt *et al.*, 2000; Kroj *et al.*, 2003; Mönke *et al.*, 2004) as well as in the *AtGA3ox2* gene, encoding an enzyme of the gibberellic acid (GA) biosynthesis (Curaba *et al.*, 2004).

Mutations causing reductions in seed storage protein levels are powerful tools in the effort to understand the network of transcriptional regulation of seed storage protein genes. Accumulation of seed storage proteins and late embryogenesis abundant (LEA) proteins is severely reduced in *lec1*, *fus3* and *abi3* mutants, and the profile of the global pattern of gene expression is also altered during mid- to late-embryogenesis. Mutant *abi3*, *fus3* and *lec1* embryos all accumulate reduced amounts of storage protein, in particular 12S cruciferins (Nambara *et al.*, 1992; Bäumlein *et al.*, 1994; Meinke *et al.*, 1994; Keith *et al.*, 1994). In single mutants of *fus3* and *lec1*, the embryos accumulate high levels of anthocyanins and display leafy traits in the cotyledons (Meinke, 1992; 1994;

Bäumlein *et al.*, 1994; Keith *et al.*, 1994; West *et al.*, 1994). These mutants altered expression of the *Em*-like class I LEA genes (Vincient *et al.*, 2000). Conversely, *abi3* mutated seeds have inhibited responsiveness to ABA and chlorophyll breakdown (Koornneef *et al.*, 1984; Nambara *et al.*, 1992, 1995; Ooms *et al.*, 1993) not affected in *fus3* and *lec1*. Double mutants of *abi3 fus3* or *abi3 lec1* leads to highly pigment and extremely viviparous embryos with dramatically reduced protein content in seeds (Raz *et al.*, 2001). These mutants have stronger phenotypes than their parental single mutants (Bäumlein *et al.*, 1994; Meinke *et al.*, 1994; Keith *et al.*, 1994).

A different class of genes, including *ABSCISIC ACID INSENSITIVE4 (ABI4)* and *ABSCISIC ACID INSENSITIVE5 (ABI5)* also play important roles during the maturation phase of embryogenesis (Giraudat *et al.*, 1992; Finkelstein, 1994; Finkelstein and Lynch, 2000; Finkelstein *et al.*, 1998; Soderman *et al.*, 2000), preparing the embryo for desiccation and postgerminative growth. The *abi4* and *abi5* mutants exhibit altered expression of LEA genes, although they can produce desiccation tolerant seeds. The mutants are defective in gene expression of a subset of LEA genes during late-embryogenesis (Finkelstein, 1994). Although the monogenic mutants of these genes exhibit little or no phenotype on seed storage protein gene expression, other data suggest that ABI5 is involved in the regulation of seed storage protein gene expression (Soderman *et al.*, 2000.)

1.4. The vacuolar protein sorting pathways

Seed storage proteins are synthesized at the endoplasmic reticulum (ER) and stored in two different types of storage organelles, protein bodies (PB) and protein storage vacuoles (PSV).

PBs are defined as organelles containing only storage proteins that are derived by direct budding from the ER and are directly deposited into the vacuole. PBs are osmotically inactive and do not appear to have any active

transport processes into the Golgi. PBs are mainly found in cereals as organelles that contain hydrophobic prolamin-like storage proteins.
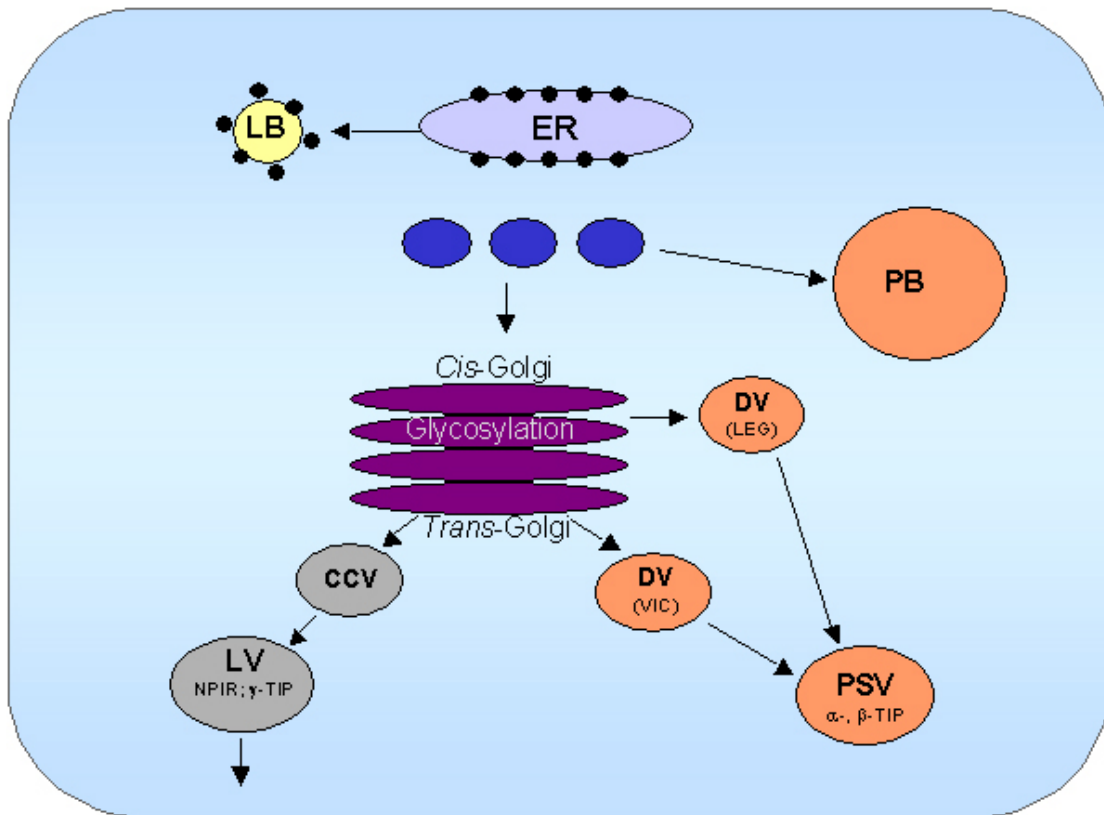
Figure 1.2. A schematic representation of the vacuolar protein sorting via the Golgi apparatus (Hinz and Herman, 2003).

LB, lipid body; ER, endoplasmic reticulum; PB, protein body; DV, dense vesicle; LEG, legume; VIC, Vicia; PSV, protein storage vacuole; TIP, tonoplast intrinsic proteins; CCV, clathrin-coated vesicle, LV, lytic vacuole; NPIR, signal consists of a larger, charged amino acid (N), a non-acidic amino acid (P), a large hydrophobic amino acid (I, L), and an amino acid with a large hydrophobic side chain (L, P).

PSVs represent a different type of organelles and are analogous to the vegetative vacuoles found in most plant cells. The PSVs are surrounded by the tonoplast that contains active transporters for an ATP and pyrophosphate-driven proton transport. Based on the presence of aquaporin-like tonoplast intrinsic proteins (TIP), the PSV can be distinguished from vegetative or lytic vacuoles.

The α- and β-TIP isoforms are characteristic for the tonoplast of the PSV, whereas the tonoplast of the lytic vacuole (LV) contain the γ-TIP isoform (Jauh *et al.*, 1999).

In plant seeds, the vacuolar proteins are sorted *via* the Golgi apparatus, which recognizes the distinct sorting signals present in the different vacuolar proteins and discriminates between lytic and storage vacuoles by two different transport pathways (Di Sansebastiano *et al.*, 2001).

Firstly, the clathrin-coated vesicle (CCV)-dependent pathway is common to all eukaryotic cells and transports vacuolar acid hydrolases into the lytic vacuole (Ghosh *et al.*, 2003). This pathway is used for vacuolar proteins bypassing the Golgi apparatus as well as for proteins that are secreted. The N-terminal NPIR motif containing a larger, charged amino acid (N) at first position, a non-acidic amino acid (P) at the second position, a large hydrophobic amino acid (I, L) at the third position, the fourth position is not strictly conserved, and an amino acid with a large hydrophobic side chain (L, P) at fifth position (Matsuoka and Nakamura, 1999), is a sequence-specific vacuolar sorting sequence (ssVSS) in vacuolar proteins that binds to a vacuolar sorting receptor (VSR) present in clathrin-coated vesicles (Kirsch *et al.*, 1994).

Secondly, the dense vesicle (DV)-mediated pathway seems to be unique to plants and is responsible for the transport of storage proteins into the protein storage vacuole in seeds (Hinz and Herman, 2003). Mature DV do not have a clathrin coat (Hohl *et al.*, 1996), nor do they contain the vacuolar sorting receptor for vacuolar acid hydrolase (Hinz *et al.*, 1999). Sorting of storage proteins into DV starts at the *cis*-region of the Golgi stack (Hillmer *et al.*, 2001) and exits at the *trans*-face of the Golgi apparatus (Figure 1. 2). This pathway processes the primary storage proteins of PSVs, the seed storage globulins.

The precursor polypeptides of the 12S legumin-type globulin form trimers with a molecular mass of 180 kD within the ER. After the import into the PSV,

they are proteolytically processed at a conserved asparagine residue by a vacuolar processing enzyme (Müntz, 1996). Each 60 kD monomer is cleaved into two chains with a molecular mass of 40 kD (α-chain) and 20 kD (β–chain). After this processing, two trimers form a hexamer with a molecular mass of 360 kD. Proteins of the 7S vicilin-type also form trimer oligomers within the ER. After a post-translational glycosylation in the Golgi apparatus they also become deposited in the PSV.

In *Arabidopsis*, the major seed storage proteins are the 12S globulins and 2S albumins, which are named cruciferin and either napin or arabin, respectively. Both 12S globulins and 2S albumins are initially synthesized as a precursor at the rough ER. These preproteins are then processed and sorted further and finally accumulate in the protein storage vacuoles. Like other 12S storage proteins, cruciferin subunits are generated from a single polypeptide-precursor following its specific cleavages by processing enzymes into the α- and β-chains. Both chains remain connected *via* a disulfide bridge. The subunits assemble as hexameric complexes. Similar to cruciferin, the 2S albumin precursor polypeptide is also cleaved into two chains (2S-large and 2S-small), which then accumulate as a heterodimer consisting of two subunits linked by disulfide bridges (Krebbers *et al*., 1988).

The *Arabidopsis* genome (ecotype Col-0) contains a small multi-gene family, which encodes five isoforms of napins, referred to as *at2S1* to *at2S5* (Krebbers *et al*., 1988; van der Klei *et al*., 1993) and three genes encoding 12S globulins (Pang *et al*., 1988).

All seed storage protein genes of *Arabidopsis* are expressed specifically during mid- to late-stages of seed development. The expression control of these genes will be discussed in the following section.

1.5. The BURP-domain protein family

In addition to the seed proteins mentioned above a novel, functionally unknown seed protein has been described and is nowadays designated as BURP protein. The Underline Seed Protein of *Vicia faba* (VfUSP) is the founding member of the BURP domain family (Bassüner *et al*., 1988). This gene encodes a 30 kD primary translation product consisting of distinct domains: (i) a cleavable signal peptide, (ii) a hydrophilic N-terminal stretch possessing two serine clusters, (iii) a valine cluster and a hydrophobic domain in the C-terminal part of the polypeptide. In a cell free translation system the presence of a functional signal peptide was shown, which interacts with the signal recognition particle resulting in a cotranslational translocation across the membrane of the endoplasmic reticulum. If synthesized in *Xenopus* oocytes the translation product of the mRNA was secreted out of the cell. Homologous mRNA was found to be present also in developing cotyledons of pea (*Pisum sativum* L.) and french bean (*Phaseolus vulgaris* L.). The *VfUSP* gene corresponds to the most abundant mRNA present in cotyledons during early seed development; however, the corresponding protein does not accumulate in cotyledons. The characterized *VfUSP* gene with its two introns is one of about 15 members of a gene family. A fragment comprising 637 bp of 5`flanking sequence and the total 5´untranslated region was shown to be sufficient to drive the mainly seed-specific expression of two reporter genes both in transgenic *Arabidopsis thaliana* and *Nicotiana tabaccum* plants. These reporter constructs become active in transgenic tobacco seeds both in embryo and endosperm, whereas its activity in *Arabidopsis* is detectable only in the embryo. Furthermore, it was demonstrated that the gene promoter is transiently active in root tips of both transgenic host species (Bäumlein *et al*., 1991). The gene becomes active during early embryogenesis and previous work has been focused on at least three different aspects.

Firstly, detailed knowledge concerning the tissue specific and developmentally dependent regulation of the *VfUSP* gene promoter has been established. This includes the characterization of important *cis*-elements

including the RY-element and interacting transcription factors like FUS3 and ABI3 (Bäumlein *et al.*, 1991; 1994; Fiedler *et al.*, 1993; Wohlfarth *et al.*, 1998; Reidt *et al.*, 2000; 2001; Mönke *et al.*, 2004).

Secondly, due to its strong activity in seeds the *VfUSP*-gene promoter has been extensively applied to control the expression of various transgenes for gene farming purposes (Czihal *et al.*, 1999). Using transgenic pea seeds as bioreactors, the patented promoter has been proven to work reliably also under field conditions (Giersberg *et al.*, 2004). More detailed experiments demonstrate that the *VfUSP*-gene promoter exhibits a low but detectable activity in pollen (Giersberg *et al.*, 2004). A similar behavior has been described for several other thought to be seed-specific gene promoters (Zakharov *et al.*, 2004).

Thirdly, during *in vitro* embryogenesis a *VfUSP* gene promoter driven marker gene (GFP) construct can be used to label specifically cells with competence for embryogenesis. Cell sorting allows the differential isolation of embryogenesis competent and embryogenesis incompetent cell populations as basis for the identification of genes involved in the control of embryogenesis both in tobacco (Chesnokov *et al.*, 2002), *Arabidopsis* and *Brassica* (Chesnokov and Manteuffel, unpublished).

In strong contrast to these abundant data concerning the gene promoter and its application, the function of the USP-gene product is poorly investigated. A functional characterization of the corresponding protein becomes even more important since sequence comparisons reveal the existence of a rapidly growing, strictly plant specific novel protein family. All proteins of the family share a highly conserved protein domain in their C-termini. This domain has been named U-domain (Wohlfarth *et al.*, 1998) or BURP-domain (Hattori *et al.*, 1998). Since the term BURP has been used in previous databases the domain is consistently named BURP in the following text. BURP stands for BNM2, USP, RD22 and Polygalcturonidase isozyme (see below).

The consensus sequence of the BURP-domain is shown in Figure 1.3.

$$X_4\text{-}\underline{CH}\text{-}X_{10}\text{-}\underline{CH}\text{-}X_{25\text{-}27}\text{-}\underline{CH}\text{-}X_2\text{-}T\text{-}X_{15\text{-}16}\text{-}P\text{-}X_5\text{-}\underline{CH}\text{-}X_8\text{-}W/F$$

Figure 1.3. Consensus sequence of the BURP-domain
The CH motifs with conserved distances are underlined.

Most remarkably are the highly constant positions of the CH-motifs. Current working hypotheses suggest functions of the BURP-domain for metal binding, intramolecular folding processes, protein-protein interactions or cellular targeting. Some members of the protein family exhibit a similar modular structure of tandem repeats in the N-terminal region in addition to the BURP-domain. A putative N-terminal signal peptide has been shown to be cotranslationally processed (Bassüner *et al*., 1988, Zheng *et al*., 1992).

Preliminary experiments with *Vicia faba* and *Pisum sativum* cotyledons did localize the VfUSP in electron dense vesicles 200-400 nm in size, which are distinct from protein bodies. Labeling was also found in dictyosomes and to lesser extent in the endoplasmic reticulum (S. Hillmer, unpublished results).

An amino acid sequence alignment of the CH-pattern within the BURP-domains of selected members of the protein family is shown in Figure 1.4.

All known BURP-domain proteins are of plant origin, suggesting that the function of BURP-domain proteins might be plant specific. The following members of the protein family have been previously described in some more detail: The Unknown Seed Protein (USP) of *Vicia faba* (Bassüner *et al*., 1988; Bäumlein *et al*., 1991), the non-catalytic β-subunit of the polygalacturonase isozyme (PG) from ripening tomato (Zheng *et al*., 1992, Watson *et al*., 1994), the two aluminium stress induced proteins SALI 3-2 and SALI 5-4a with similarity to the auxin-downregulated protein ADR6 of soybean (Datta *et al*., 1993; Ragland and Soliman, 1997), the drought stress induced protein AtRD22 of *Arabidopsis* (Yamaguchi-Shinozaki and Shinozaki, 1993), a root cortex parenchyma cell

specific protein of *Zea mays* (Held *et al.*, 1997), the microsporogenesis specific protein BNM2 of *Brassica napus* (Boutilier *et al.*, 1994; Treacy *et al.*, 1997), a female sex-specific gene of *Mercurialis annua* (Yang *et al.*, 1998), an apomixis specific protein of *Panicum maximum* (Chen *et al.*, 1999), further soybean proteins with various expression patterns (Granger *et al.*, 2002) and most recently the sporophytically produced structural protein RAFTIN found in the classical Ubisch bodies with a putative function in pollen development (Wang *et al.*, 2003).

```
                     The CH-pattern in the BURP-domain


  USP    179 NAVMC-RLNFEKVVFNCHQVRDTTAYVVSLVASDGTKT--KA 217
  ADR6   177 KAVMCHRLNFRTAVFYCHEVRETTAFMVPLVAGDGTKT--QA 216
 Raftin  299 VFVACHDEAYPYTVYRCHTTGPSRAYMVDMEGARGGDAV-TI 339
  RD22   309 KSVVCHKQKYPFAVFYCHKAMMTTVYAVPLEGENGMRA--KA 348
  PG     540 KAVSCHQSLYPYLLYYCHSVPKVRVYEADLLELNSKKKINHG 581
  ARO    546 KSVSCHQTLYPSLLYYCHSVPKVRVYEADILDPNSKAKINHG 587
  BNM2   196 KVLGCHRMPYPYVVVYYCHGHKSGTKVFEVNLMSDDGIQLVVG 237
  ASG1   215 DFVGCHPELYPYSVYRCHTSVQTGTYVMEMQSSYNG-GALKL 255
             ...  **.  .    ..  **  .   .  .  .  . ..      ..


  USP    218 LTVCHHDTRGMNP-ELLYEALEVTPGTVPVCHFIGNKAAAWV 258
  ADR6   217 LAICHSNTSGMNH-QMLHQLMGVDPGTNPVCHFLGSKAILWV 257
 Raftin  340 ATVCHTDTSLWNPEHVSFKLLGTKPGGTPVCHLMPYGHIIWA 381
  RD22   349 VAVCHKNTSAWNPNHLAFKVLKVKPGTVPVCHFLPETHVVWF 390
  PG     582 IAICHMDTSSWGPSHGAFLALGSKPGRIEVCHWIFENDMNWA 623
  ARO    588 VAICHVDTSSWGPRHGAFVALGSGPGKIEVCHWIFENDMTWA 629
  BNM2   238 PAVCHMDTSMWNADHVAFKVLKIEPRSAPVCHFFPLDNIVWV 242
  ASG1   256 VAVCHRNTTSWDPEHVSFKVLASKPGGLPICHFVPYGHVIFG 297
             ...**  .*.. .. .    .    *.    .**.. ..   .
```

Figure 1.4. Amino acid sequence alignment of the CH-pattern within BURP-domain of selected members of the BURP-domain protein family

USP (X13242), an abundant seed protein of *Vicia faba; ADR6* (X69639), an auxin down regulated protein of *Glycine m*ax; Raftin (CAE02613), an anther protein of *Triticum aestivum*; *RD22* (D10703), a protein responsive to drought stress of *Arabidopsis thaliana*; *PG* (U63373), the β-subunit of the polygalacturonase isoenzyme 1 of tomato; *ARO* (U64790), an aromatic amino acid rich glycoprotein of tomato; *BNM2* (AF049028), an in vitro-embryogenesis specific protein of *Brassica napus*; *ASG1* (no accession), an apomixis-specific gene of *Panicum maximum.*

With only rare exceptions there are no functional protein data available. For the β-subunit of the polygalacturonase isozyme (PG) it was shown that the signal peptide and the following peptide segment are cleaved during protein processing. A second proteolytic cleavage results in a non-catalytic glycoprotein forming a complex with the catalytically active polygalacturonase (Zheng *et al.*, 1992; Watson *et al.*, 1994). The fate of the C-terminal part of the molecule containing the BURP domain remains unknown. Whether similar processing steps are required for the other proteins mentioned above is also not known.

In *Arabidopsis*, AtRD22 was only characterized on a genetically level. Northern hybridization analysis showed that AtRD22 mRNA is induced by salt stress as well as by water deficiency but not by cold or heat stress. AtRD22 mRNA appeared after the application of ABA, an indication that transcription of AtRD22 mRNA is induced by endogenous ABA, the production of which is triggered by drought and salt stress (Yamaguchi-Shinozaki and Shinozaki, 1993). The pattern of AtRD22 expression is similar to that of VfUSP, AtRD22 mRNA was expressed during the early and middle stages of seed development (Yamaguchi-Shinozaki and Shinozaki, 1993; Genevestigator data).

Several of the BURP-domain proteins have been isolated as seed- or embryogenesis related genes. This is for instance supported by the expression of the VfUSP gene during early stages of zygotic embryogenesis (Bassüner *et al.*, 1988) and very early stages of *in vitro* embryogenesis (Chesnokov *et al.*, 2002), the induced expression of *BNM2* during the microspore derived embryogenesis of *Brassica* (Treacy *et al.*, 1997) as well as the specific expression of ASG1 during the early embryo sac development in apomictic, but not in sexual gynoecia of *Panicum* (Chen *et al.*, 1999). Together these data suggest a putative function of the protein family for embryogenesis and seed development, but does not exclude the involvement in other cellular and developmental processes.

1.6. The aim of this study

It has been estimated that about 30% of the sequenced and annotated *Arabidopsis* genes are functionally completely unknown. Moreover, many functional annotations nowadays depend mainly on sequence similarities to other genes or proteins, which in fact are also only poorly characterized experimentally. This might often result in circular arguments. Without guidance from sequence similarities of functionally better known genes of other organisms like yeast or animal systems - for a strictly plant specific gene/protein family - it is difficult to suggest even an idea about its function. State of the art techniques to solve these questions include the analysis of gain- and loss-of-function, histological and subcellular localization, ultrastructural changes, spatial and temporal expression pattern, interacting proteins or cellular components.

The current study aims to the functional characterization of the BURP-domain protein family and the highly conserved BURP-domain itself. Most of the BURP-proteins known up to now have been isolated from many different plant species and under various experimental conditions without leading to clear functional conclusions. Therefore, the approach described in this thesis will apply the technical advantages of the model plant *Arabidopsis thaliana* for a functional characterization of two members of the BURP-domain protein family.

2. Materials and Methods

2.1. Materials

2.1.1. Plant materials

Table 2.1. Used plant species

Plant species	Cultivars	Source
Arabidopsis thaliana	Columbia (Col-0)	Genebank IPK, Gatersleben
	Wassilewskaja (WS-2)	Genebank IPK, Gatersleben

2.1.2. Bacterial strains

Table 2.2. Used bacterial strains

Microorganism	Strain	Reference
Escherichia coli	DH5α	Sambrook *et al.* (1989)
(*E.coli*)	BL21 (DE3) pLysS	Novagen
Agrobacterium tumefaciens	PGV2260	Deblaere *et al.* (1985)
(*A.tumefaciens*)	PGV3101 (pMB90RK)	Koncz and Schell (1986)

2.1.3. Yeast strains

Table 2.3. Used yeast strains

Strain	Source
AH109	BD Biosciences Clontech
Y187	BD Biosciences Clontech

2.1.4. Plasmids

Table 2.4. Used plasmid vectors

Vector	Characteristic	Approach	Source
pCR2.1	Kan^r, Amp^r	Cloning, sequencing	Invitrogen
pDONR201 (donor vector)	Amp^r, *att*P, ccdB	Cloning, sequencing	Invitrogen
pBIN101	GUS, Kan^r	*AtUSPL1* promoter regulation	Clontech
pBENDER	35S, Kan^r, *att*R, ccdB	Gene overexpression	Marc Jakoby and Weisshaar <mpiz-koeln.mpg.de>
pMDC83	GFP, Kan^r, *att*R, ccdB	Processing of proteins	Curtis et al.(2003) <Zurich Uni.>
pGBKT7	c-Myc, Trp1, Kan^r	Yeast two hybrid	Clontech
pGADT7	HA, Leu2, Amp^r	Yeast two hybrid	Clontech
PGEX-4T-1	GST, Amp^r	Protein expression in *E.coli*	Amersham Bioscienses
pET-23a-d(+)	His, T7, Amp^r	Protein expression in *E.coli*	Novagen

2.1.5. Primers and oligonucleotides

2.1.5.1. PCR primers

Table 2.5. Used PCR primers

Name	Sequence 5'-3'	Note
USP_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATGGCTTCTACTTTTAGGCTCTC	Used for AtUSPL1 overexpresion; GFP fusion with AtUSPL1 or NUSP
USP_R	GGGGACCACTTTGTACAAGAAAGCTGGGTTTACTTTGTTACCCACACAATG	Used for AtUSPL1 overexpresion
USP_R2	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTTTGTTACCCACACAATGT	Used for AtUSPL1-GFP and BURP-GFP fusions

USP-EcoRI	CGGAATTCATGGCTTCTACTTTTAGG	Used for AtUSPL1 expression and two-hybrid
USP-XhoI	CCGCTCGAGCTTTGTTACCCACACAATG	Used for AtUSPL1 or BURP expression
USP-PstI	TGCACTGCAGTTACTTTGTTACCCACACA	Used for AtUSPL1 two-hybrid
NUSP_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCATA CATGTACAACGAAGGGT	Used for NUSP-GFP fusion
NUSP_XhoI	CCGCTCGCGATACATGTACAACGA	Used for NUSP expression
NUSP_EcoRI	CGGAATTCGAGGCACACACGTCTAGA	Used for NUSP expression
CUSP_EcoRI	CGGAATTCACACTCAATGATCTCA	Used for BURP expression
CUSP_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGA AGGAGATAGAACCATGTTCACACTCAATGATCT	Used for BURP-GFP fusion
SPusp_BamHI	CGGGATCCATGGCTTCTACTTTTA	Used for clone AtUSPL1 signal peptide
SPusp_EcoRI	CGGAATTCCTCCACTACCCACAG	Used for clone AtUSPL1 signal peptide
USP-Gusa	GCATGCCCATGGATTGTATTTTCTCTGC	Used for GUS expression
USP-Gusb	AAGCTTCTGTATAAATCATCCATCACAAATA	Used for GUS expression
RD_start	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGA AGGAGATAGAACCATGGCGATTCGTCTTCCTCT	Used for GFP fusion of AtRD22 or NRD
RD_R	GGGACCACTTTGTACAAGAAAGCTGGGTCGTAG CTGAACCACACAAC	Used for GFP fusion of AtRD22
NRD_R	GGGACCACTTTGTACAAGAAAGCTGGGTCGAGA GCCGCGTTAGGATCGTC	Used for GFP fusion of NRD
NRD_EcoRI	CGGAATTCTTAACACCGGAGCGT	Used for AtRD22 or NRD expression and two-hybrid
RD22_HindIII	CCCAAGCTTGTAGCTGAACCACACAA	Used for AtRD22 expression
NRD-XhoI	CCGCTCGAGGAGAGCCGCGTTAG	Used for NRD expression
Pst-RD22	TGCACTGCAGCTAGTAGCTGAAGGACACA	Used for two-hybrid of AtRD22

2.1.5.2. Sequencing oligonucleotides

Table 2.6. Used oligonucleotides for sequencing

Name	Sequence 5'-3'	Note
LBa1	TGGTTCACGTAGTGGGCCATCG	T-DNA Salk confirmation
LBb1	GCGTGGACCGCTTGCTGCAACT	T-DNA Salk confirmation
usp-utr-a	CTCATTTTCTCCGCAGAGA	AtUSPL1 native
usp-utr-b	CACATTATATAGAGCTACGTC	AtUSPL1 native
USP_for	TCAAGGGAAAGCCATAAAGGAGAC	Sequenced AtUSPL1 cDNA
USP_rev	CACCATCTTTGAGCAAATGACTA	Sequenced AtUSPL1 cDNA
KO1	GCCAGTGAATTCCCGATCTAGTAACATA	Clone AtUSPL1 for T-DNA Salk
KO2	TGGTTCACGTAGTGGGCCATCG	Clone AtUSPL1 for T-DNA Salk
USPa	AATTTGATGACCCTTCGTTGT	AtUSPL1 confirmation in Salk and AKF
USPb	GACCTCGGCTCCATCTTCAACACCTTAAA	AtUSPL1 confirmation in Salk and AKF
JL202	CATTTTATAATAACGCTGCGGACATCTAC	T-DNA AKF confirmation
JL270	TTTCTCCATATTGACCATCATACTCATTG	T-DNA AKF confirmation
GUSb	AGTCCCGCTGGTGCCTTGTC	GUS confirmation in transgenic plants
GFP_rev	AAC ATCACCATCTAATTCAACAAGAATTGG	GFP confirmation in transgenic plants
RDa	TGGCGATTCGGCTTCCTCTGA	Confirmation of AtRD22 in Salk
RDb	GCTTTAGCTCGCATCCCGTTCT	Confirmation of AtRD22 in Salk
ATRD22_for	CGGAATTCATGGCGATTCGTCTTCCTCTG	Clone fragment for probe
ATRD22_rev	CCGCTCGAGCTAGTAGCTGAACCACACAA	Clone fragment for probe
M13-21uni	GTA AAA CGA CGG CCA GT	Sequencing primer for pCR2.1 vector
M13-24rev	AAC AGC TAT GAC CAT G	Sequencing primer for pCR2.1 vector
PGAD_F	CTATTCGATGATGAAGATACC	Sequencing primer for pGAD GH vector
PGAD_R	GGGGTTTTTCAGTATCTACG	Sequencing primer for pGAD GH vector

SeqL-A	TCGCGTTAACGCTAGCATGGATCTC	Sequencing primer for pDONR201 vector
SeqL-B	GTAACATCAGAGATTTTGAGACAC	Sequencing primer for pDONR201 vector

2.1.6. Enzymes, antibodies and kits

Table 2.7. Used enzymes, antibodies and kits

Company	Product
Amersham, Braunschweig	GST–tag purification kit, Readiprime II Random prime labeling kit, Restriction endonucleases, ECL kit
Biomol GmbH, Hamburg	Total RNA isolation kit
Fermentas	Restriction endonucleases
Nanoprobes, USA	Gold-labeled goat anti-rabbit IgG (H+L) antibody
Invitrogen, The Netherlands	Gateway clonase enzyme kit, TA cloning kit
Novagen, USA	His-Tag purification kit, T7-Taq purification kit
Qiagen, Hilden	QIAquick agarose gel extraction kit, QIAqiuck PCR purification kit, Plasmid purification kits, Taq PCR polymerase, Taq PCR master mix, DNeasy plant DNA isolation kit, RNeasy plant total RNA isolation kit, RT-PCR kit, Ni-NTA Spin Kit
Roche (Boehringer Mannheim)	Restriction endonucleases, T4 DNA ligase, T4 polynucleotide kinase, Taq DNA polymerase, Expand high fidelity PCR system, Rapid DNA ligation kit, PCR nucleotide mix, RNAse, DNAse, Anti-rabbit IgG alkaline phosphatase conjugate
Sigma, Deisenhofen, Germany	Anti-polyHistidine monoclonal antibody, Anti-mouse IgG alkaline phosphatase conjugate, anti-mouse IgG horse radish peroxidase conjugate
Stratagene, Heidelberg	Restriction endonucleases, Nuc Trap Probe purification column
USB, Cleveland OH, USA	Restriction endonuclease, Klenow enzyme

2.1.7. Chemicals

Table 2.8. Used chemicals

Company	Product
Amersham, Braunschweig	$[\alpha^{32}P]$ dATP, $[\gamma^{32}P]$ ATP, $[\alpha^{32}P]$ dCTP, Hybond-N+ nylon membrane
Amresco, USA	Phenol
Biometra, Göttingen	Chloroform, Phenol, Phenol-Chloroform, ATP, BSA, dNTPs, SDS
Clontech	SD basis medium, DO supplement
Difco, USA	Bacto®-agar, Bacto®-trypton, Yeast extract
Duchefa, The Netherlands	Murashige-Skoog whole medium solid substance, Rifampicin, Kanamycin, Hygromycin, Carbenicillin
Eurogentec, Belgium	Smart Ladder
Fermentas	Prestained protein ladder
Fluka, Schweiz	DEPC, PEG 8000
Gibco-BRL, USA	Agarose, 1Kb DNA ladder, EDTA
Kodak, USA	X-Ray films
Merck, Darmstadt	Ethanol, Ethidium bromide, Formamide, HEPES, Magnesium chloride, Sodium acetate, Sodium hydroxide, Sodium-dihydrogen phosphate, di-Sodium hydrogenphosphate, Trichloroacetate, Tris base
MWG-Biotech AG, Ebersberg	DNA oligonucleotides
National diagnostics, England	Protogel
NEN, USA	GeneScreen Plus hybridization transfer membrane
Roth, Karlsruhe	Phenol, Phenol-chloroform, Chloroform, Formaldehyde, Glycerol, Isopropanol, Lithium chloride, Sodium chloride, TEMED, Amoniumpersulphate
Schleicher&Schuell, Dassel	Blotting paper GB 002, Nitrocellulose membrane BA 85

Serva, Heidelberg	X-Gal, Sodium citrate, Tween$_{20}$, Tetracycline, Coomassie blue, EDTA, X-gal
Sigma-Aldrich	Sodium carbonate, Sodium bicarbonate, Sodium chloride, Tween$_{20}$, Potassium chloride, Sodium monophosphate, Sodium diphosphate, IPTG, X-gal, MOPS

2.1.8. Laboratory tools and equipments

Table 2.9. Used laboratory tools and equipments

Company	Equipment
Ambion	RNAse ZAP cleaning reagent
AGS, Heidelberg	DNA gel-electrophoresis tanks
Berhof GmbH, Eningen	DAP III high pressure block
BioRad, München	Gene-Pulser, Mini Electrophoretic System (Mini-Protean SDS-PAGE running cell, Mini Trans-Blot Electrophoretic transfer cell, Electro Eluter)
Biotec Fischer, Reiskirchen	Phero-stab 200 electrophoresis power supply
CBS, USA	EBS 250 power supply
DuPont, USA	Sorvall centrifuge RC 5C
Eppendorf, Hamburg	Mastercycler®5330 (DNA- thermocycler), Thermomixer 5436 and 5437, Thermomixer compact, Cold centrifuge 5402, BioPhotometer
GFL, Burgwedel	Hybridization oven, Water bath
Heraeus, Osterode	Centrifuges (Biofuge 13, Biofuge 15R), HERASafe laminar boxes
OWL	Agarose gel trays
Perkin-Elmer, USA	GenAmp PCR system 9700 (0,5 and 0,2 mL blocks)
Pharmacia, Freiburg	Photometer, Ultrospec plus
Polaroid, Offenbach	MP-4 camera
Raytest, Straubenhardt	FUJI BAS imager, Imaging plates
Savant	SpeedVac SPD101B

Schutt labotecnik, Germany	Sonicator (Bamdelin sonopuls HD70)
Stratagene, Heidelberg	UV-Stratalinker® 1800, NucTrap® probe purification columns
Varian, Australia	SpectAA 10 plus AAS
Zeiss, Jena, Germany	Stereo Lumar V12 Stereomicroscope; Axioplan Imaging Microscope

2.1.9. Media and buffers

Media for bacteria and yeast were prepared based on Sambrook *et al.* (1989) and Adam *et al.* (1997). Specific media were prepared following instructions of Clontech and Novagen. Plant media were prepared according to Murashige and Scoog (1962), or Ready MS medium (Duchefa, Netherland) was used.

Bacteria medium (details are in appendix)

LB (*E.coli*)

TB (*E.coli*)

YEB (*A. tumefaciens*)

These mediums were supplemented with appropriate antibiotics.

Yeast medium (details are in appendix):

YPD, YPDA

SD with appropriate supplements

Plant growth medium (details are in appendix):

MS medium: for plants growing *in vitro*

GM (½ MS modified medium): for seed germination.

Dehydration mediums:

MS containing 5% -15% PEG 8000 (Van der Weele *et al*., 2000)

MS containing 0.1-0.3 mM NaCl

MS with 0.2-0.4 mM mannitol

All used buffers were prepared according to Sambrook *et al*. (1989) or following product instruction manuals.

2.1.10. Software

DNA and protein sequence data were processed using the program package *Lasergene* version 4 and 5 of DNASTAR Inc., USA and *BLAST* (Basic Local Alignment Search Tool; Altschul *et al.*, 1990).

The autoradiography images were analyzed by TINA 2.09 of Raytest Isotopenmeßgeräte GmbH (Germany).

2.2. Methods

2.2.1. Molecular cloning and sequencing

The Gateway® Cloning Technology is based on the lambda recombination system to facilitate transfer of heterologous DNA sequences (flanked by modified *att* recombination sites) between vectors (Hartley *et al.*, 2000). The cloning method was performed using two recombination reactions, which constitute the basis of the Gateway® Technology as BP and LR reaction. Here, Gateway vectors were used from Zurich University, Max-Planck-Institut für Züchtungsforschung and Invitrogen Company.

The standard molecular cloning methods (e.g. restriction digestion, ligation, DNA and protein gel electrophoresis) were performed according to Sambrook *et al.* (1989). The transformation of *E. coli* was performed using the heat-shock procedure (Cohen *et al.*, 1972). Plasmid DNA extraction and purification was done by using the standard methods described in Sambrook *et al.* (1989), or by using QIAGEN plasmid kit. PCR products were purified with QIAquick PCR purification kit (QIAGEN). DNA fragments were isolated and purified from the agarose gel with the QIAquick kit (QIAGEN).

DNA sequences were determined in the Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) Gatersleben by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977).

2.2.2. Protein expression and purification

2.2.2.1. Cell extract preparation

The genes coding for AtRD22, N-terminal part of AtRD22 (TXV), AtUSPL1, and BURP-domain of AtUSPL1 proteins were cloned into pET23a vector. These polypeptides were expressed in *E. coli* BL21 (DE3) pLysS strain and purified by using the pET system (Clontech). The N-terminal part of AtUSPL1 (P) was cloned into pGEX-4T-1 vector. This polypeptide was expressed as GST fusion protein by using the pGST fusion system of Amersham Bioscienses. Since AtUSPL1 seems to be toxic for *E. coli*, the following protocol was used for its expression:

1. A single colony was inoculated in 2 ml of TB medium containing carbenicillin (200 µg/ml) and glucose (1%). The cells were grown at 37°C.

2. When OD_{600} reached 0.2–0.6, the cells were collected by centrifugation and resuspended in 2 ml fresh TB medium. Then 100 µl of cell culture were added into 250 ml of TB medium containing carbenicillin (500 µg/ml) and glucose (1%) and incubated at 37°C.

3. When OD_{600} reached 0.2–0.6, the supernatant was removed by centrifugation. The cell pellet was resuspended in 500 ml of pre-warmed TB medium containing carbenicillin (500 µg/ml) and IPTG (0.4 mM). This bacteria culture was incubated at 30°C for 3 h, and the bacteria were harvested by centrifugation.

The induced cells were harvested by centrifugation at 5000 rpm and resuspended in buffer containing 5 mM imidazole (ratio 1:100, buffer:bacteria culture). Then the bacteria were destroyed by ultrasonication at 45% power for 45 seconds. The supernatant containing soluble proteins as AtRD22, TXV, BURP and P were separated from bacterial pellet by centrifugation. Inclusion bodies containing aggregated AtUSPL1 were found in the pellet together with cellular debris.

The AtUSPL1 proteins were isolated from the bacterial pellet by suspending in buffer containing 6 M urea and 5mM imidazole. After centrifugation, the supernatant containing AtUSPL1 proteins was filtered through a 0.45-µm membrane.

2.2.2.2. Protein purification

Solubilized proteins containing the C-terminal His-tag such as recombinant AtRD22, BURP and TXV were purified using the Ni-NTA Spin Kit (QIAGEN). The P protein fused to GST was purified by using Glutathione Sepharose 4B (Amersham Bioscienses). AtUSPL1 protein was purified under denaturing conditions by using His Bind$^@$ Column Chromatography (Novagen).

Protocols of protein purification were performed following the producer's instructions.

2.2.2.3. Refolding of insoluble protein

AtUSPL1 eluted with high urea concentration (6 M) was dialyzed by step wise lowering the urea concentration in the dialysis buffer (PBS). The protein concentration of dialyzed AtUSPL1 was measured according to Bradford (1976).

Gel electrophoresis analysis of the AtUSPL1 preparation was performed on 12.5% SDS polyacrylamide gel according to Laemmli (Laemmli, 1970). One of the gels was stained with Coomassie Blue (Sambrook *et al*., 1989); the other was used for Western blot analysis (see 2.2.2.5).

2.2.2.4. Antibody production

Rabbits were injected with 80 µg of purified protein for 4 times. Primary immunization was performed with complete Freund's adjuvant, later on incomplete adjuvant was used. The time between the primary and the secondary injection was 28 days whereas all other injections were given in an interval of 10 days. 10 days after the last injection serum was taken to check the reactivity of the polyclonal antibodies. 10 days after the 4th injection the animals were killed

and their blood was collected. The sera were obtained after agglutination of the blood (4°C, 12h) by centrifugation at 3000 rpm, 4°C for 30 min.

The IgG fractions were purified by affinity chromatography on Protein A-sepharose and applied for immunolocalization. The IgG fractions were also purified according to the following protocol:

1. The recombinant protein was separated on a 12.5% polyacrylamide gel and transferred onto nitrocellulose membrane.

2. A membrane band containing the bound target antigen was cut out and blocked in 3% BSA solution.

3. After 1h, the antigen coated membrane was washed twice in TBS containing 0.05%(v/v) of Tween-20 and 0.2%(v/v) of Triton X-100, and one more time in TBS.

4. The solid fixed antigen was incubated with the appropriate antiserum for 1h. Afterwards the membrane was washed twice in TBS containing 0.05%(v/v) of Tween-20 and 0.2%(v/v) of Triton X-100, and one more time in TBS.

5. The antibody fraction specifically bound to the membrane fixed recombinant protein was eluted 3 times with glycine-HCl, pH 2.2 supplemented with 1% BSA. The membrane was washed and blocked for reusing. After neutralization, the eluted IgG fraction was directly used as primary antibody to label Western blot or histological sections.

The recombinant proteins were detected on Western blots by use of the purified antibodies diluted 1:1000 followed by anti-rabbit IgG conjugated to alkaline phosphatase or conjugated to horseradish peroxidase. Histological sections were labeled by using the primary antibody followed by gold-labeled anti-rabbit IgG or fluorescent-labeled anti-rabbit IgG.

2.2.2.5. Western blot analysis

Proteins were separated on 12.5% SDS polyacrylamide gels according to Laemmli (Laemmli, 1970) and electrophoretically transferred to nitrocellulose membranes (0.4 µm, BA85, Schleicher and Schuell) as described by Borisjuk et al. (Borisjuk *et al*., 1998). After electrotransfer the membranes were washed in TBS and blocked by incubation for 1 h in 3% BSA. The wash step was repeated and the membranes were incubated with the primary antibody in TBS supplemented with 0.5% BSA and 0.05% Tween 20 for 1 h at RT. The membranes were washed 3 times; each for 5 min, in TBS supplemented with 0.5% Triton X-100 and one time without detergent for 10 min.

Membranes were incubated with the secondary antibody conjugated to ALP (color reaction) or to horseradish peroxidase (ECL technique). After additional washing as described above, labeling was detected by staining with NBT/BCIP (color reaction) or by chemiluminescence using the ECL Kit (Amersham Pharmacia Biotech) according to the producer's instructions.

2.2.3. *Agrobacterium tumefaciens* growth and treatment
2.2.3.1. Transformation of A. tumefaciens

The competent cells of *Agrobacterium tumefaciens* pGV3101 and pGV 2260 were prepared using the $CaCl_2$ method (Sambrook *et al*. 1989). The *Agrobacterium* strains were grown in 50 ml of YEB medium at 28^0C until OD_{600}~ 0.5-1.0. The cells were centrifuged at 3000 rpm for 5 min and resuspended in 1 ml of 20 mM $CaCl_2$. 100 µl aliquots of the resuspended cells were dispensed in separate Eppendorf tubes. The transformation with plant expression vectors was done using the thawing-freezing method as described by Höfgen *et al*. (1988).

2.2.3.2. Confirmation of transgenic A. tumefaciens

Colony PCR

Single colonies of transformed *Agrobacterium* cells were plated by toothpicks on agarose plates with YEB medium containing appropriate antibiotic

and incubated for 2 days at 28°C. The bacterial cells were analyzed by PCR amplification with appropriate primers. PCR results were analyzed by gel electrophoresis.

Plasmid preparation from A. tumefaciens

The transgenity of the *A. tumefaciens* was also confirmed by DNA digestion with appropriate restriction enzymes. Recombinant plasmids were prepared using the Low-copy-Number Plasmid protocol (QIAGEN).

2.2.4. *Arabidopsis thaliana* growth and treatment

2.2.4.1. Growth of A. thaliana in soil

Arabidopsis plants were grown in a growth chamber at 22°C with 10000 lux of white light for 16 h. The plants were grown under these conditions until the end of maturation (~22 days after pollination).

2.2.4.2. Isolation of protoplasts from suspension cultures and transient expression assay

A suspension cell culture of *Arabidopsis* was used as material for protoplast isolation (A.Tewes, IPK-Gatersleben). During an overnight incubation of cells in 1% cellulase and 0.5% macerozym solution, cell walls were digested. Cell debris and protoplasts were separated by centrifugation. The protoplasts were washed two times in W5 medium and concentrated in Mg Mannitol to a density of 3.3×10^5 – 3.3×10^6 protoplasts/ml. Protoplast transformation was followed as described by Reidt *et al*. (2000). 50 µl mix of plasmid DNA (5 µg) and carrier DNA (160 µg) were added into 330 µl of protoplast suspension supplemented with 20% PEG. 4 ml of K3 medium were added before transferring them to petri dishes with 10 cm diameter. After overnight incubation in the dark, protoplasts were harvested and the GFP fluorescence of the GFP-fusion protein was determined by an Argon laser– scanning microscopy (excitation wave length of 488nm, detection at 505 - 520 nm).

GUS assays were performed as described by Jefferson *et al*. (1986) and Jefferson *et al.* (1987). The AtUSPL1 gene promoter was cloned into pBIN101 vector (Clontech). The tissues of transgenic plants were collected and vacuum-infiltrated with buffer containing 50 mM sodium phosphates, 1 mM X-Gluc, and 1 mM EDTA, and 0.05% Triton X-100. After completed infiltration, the tissues were incubated at 37°C overnight. Then these tissues were washed with 30, 70, and 100% ethanol. Tissue specific distribution of GUS activity was detected by microscopy.

2.2.4.3. Stable transformation of A. thaliana plants by floral dipping

Stable transformation of *Arabidopsis* plants was performed based on the protocol of Bechthold *et al.* (1993). Plants of *A. thaliana* (ecotype Columbia) were grown for three weeks under short day conditions (8 h light, 16 h dark) and then transferred to long day (16 h light, 8 h dark). After three weeks, the emerging bolts were cut to induce growth of multiple secondary bolts. Vacuum infiltration of plants with *A. tumefaciens* culture was done one week after clipping. Bacteria were grown until $OD_{600} > 2.0$, harvested by centrifugation and resuspended in three volumes of infiltration medium (OD_{600} ~ 0.8). Entire shoots of the plants were submerged into the *A. tumefaciens* suspension. Vacuum was applied by an oil pump for 5 min and then rapidly released. Plants were placed on their side and kept at high humidity under plastic wrap for 24 h, after that they were uncovered and set upright. Seeds were harvested from the siliques, sterilized by rinsing them in 70% ethanol followed by a rinse in 50% sodium hypochloride and 0.05% Tween20 for 10 min. Before plating the seeds on germination medium (GM) containing appropriate antibiotics, they were washed four times with sterilized water. Germinating seedlings were grown for two weeks on GM medium before transferring them to soil. Stable transformation and expression of the constructs were analyzed by PCR.

2.2.4.4. Extraction of genomic DNA

The rapid plant DNA extraction, PCR grade was carried out according to Edwards *et al.* (1991). The leaf tissue (~100 mg) was grinded in liquid nitrogen and then 400 µl of extraction buffer were added and the mixture was shaken for 1 min. The suspension was centrifuged for 5 min at full speed and the supernatant transferred into a new tube containing 300 µl of isopropanol. The DNA was collected by centrifugation for 10 min, washed twice with 70% ethanol, and resuspended in 100µl of H_2O.

2.2.4.5. Extraction of total RNA

The protocol for RNA isolation is based on the Guanidium thiocyanate (GCN) method of Chomczynski and Sacchi (1987). Alternatively, total RNA isolation was performed using the total RNA agent from Biomol or RNeasy kit from QIAGEN.

All the glass- and plastic ware used for RNA isolation were treated with RNAse ZAP cleaning reagent (Ambion) and washed with DEPC treated water.

2.2.4.6. Northern blotting

Transfer of the RNA to the membrane

Total RNA was isolated using the methods described above. For Northern hybridization, 10 µg RNA per lane was separated on a 1.2% agarose gels containing 15% formaldehyde. The RNA was transferred onto a Hybond N^+ membrane (Amersham) using 10X SSC as a transfer buffer on a capillary blot. UV cross-linking of the RNA to the membrane was performed by using the Auto Cross link mode of UV Stratalinker 1800 (Stratagene).

Radioactive labeling of DNA probes and hybridization

Hybridization of the Southern and Northern blot membranes was carried using the method of Church (Church and Gilbert, 1984).

The membranes were pre-hybridized at 64°C with Church buffer containing denaturated carrier DNA (Calf thymus DNA). After 6 hours pre-hybridization, DNA fragments obtained by PCR with specific primers, were used as probes after random prime labeling (Ready Prime Labeling Kit, Pharmacia) with [α-^{32}P]dCTP. Hybridization was carried out overnight at the same temperature as the pre-hybridization. Then, the membrane was washed at high stringency with washing buffer and the signal was detected and quantified with a Bio-Imaging analyzer BSA2000 (Fuji Photo Film Co. Ltd) or X-ray film.

2.2.4.7. Plant protein extraction

Leaves or seeds of plants were extracted with 50 mM Tris-HCl buffer pH 7.6 supplemented with 150 mM NaCl, 5 mM EDTA, 0.1% SDS, and 0.1% β-mercaptoethanol. Protein extracts were centrifuged at 1000 g for 10 min. The protein concentration of the supernatant was measured according to Bradford (1976). Western blot analyzes were carried out as described in 2.2.2.5.

2.2.4.8. Microscopy

For ultrastructural studies seeds were embedded in Spurr's low viscosity resin. In this procedure the seed coat of dry seeds was locally perforated with a sharp needle to facilitate access of fixation and embedding solutions. Treated seeds were transferred into 0.05M-cacodylate buffer pH 7.0 containing 3% formaldehyde (w/v), 2% (v/v) glutaraldehyde, 0.01% Triton X-100. After degassing, fixation was performed overnight in a fridge at 8°C or on a shaker at room temperature. Following, seeds were washed three times 15 min with the same buffer followed by dehydration in a graded ethanol series of 30%, 50%, 60%, 75%, and 90%, each step lasting 15 min. After two more changes with 100% ethanol and two changes with propyleneoxide, each step lasting 60 min, seeds were infiltrated with Spurr's resin: 33% for 12 h, 66% for 6 h and 100% for 12 h. The probes were then transferred to pointed beam capsules filled with Spurr and polymerized in an oven at 70°C. Ultrathin (95 nm) sections were cut on a Reichert Ultracut S and collected on Formvar coated 75 mesh hexagonal grids.

After post-staining with 4% aqueous uranylacetate, grids were examined in a Zeiss 902 electron microscope and digital images taken with a CCD camera.

For immunological studies seeds were embedded in HM20 resin. Here mature dry seeds were first rehydrated on wet tissue paper for 3-4 h to facilitate the removal of the complete seed coat. The embryos obtained this way were degassed and fixed overnight in 0.05 M cacodylate buffer pH 7.0 containing 3% formaldehyde (w/v), 0.5% (v/v) glutaraldehyde and 0.01% Triton X-100. After three times of 15 min washes in the same buffer, embryos were dehydrated in a graded ethanol series under progressive lowering of temperature in a Reichert-Jung AFS. Final dehydration stages, infiltration with HM20 and subsequent polymerization under UV light were performed at -35°C. Ultrathin sections collected on Formvar coated copper grids were blocked with 3% BSA in PBS for 15 min. Sections were incubated with primary antibody for 60 min, washed three times 10 min with 0.01% BSA in PBS and then incubated for 45 min with 1.4 nm gold-labeled goat anti-rabbit IgG (H+L) antibody (Nanoprobes, Yaphank, NY, USA) diluted 1:50 in 0.1% BSA in PBS. After extensive washes in PBS and water, the probes were incubated with gold enhancement solution (Nanoprobes, USA) according to the manufacturer's instruction. After 5 min, the reaction was stopped in water. Sections were contrasted with 4% aqueous uranyl acetate before evaluation in an electron microscope. Controls were prepared avoiding the primary antibody incubation step.

For fluorescence microscopy of GFP, developing seeds were harvested from immature siliques and placed on microscopy slides. The seeds successively where covered with coverslips using an appropriate amount of 0.1 M PBS pH 7.2 as medium. Microscopically investigation was performed using an Axioplan Imaging 2 (Zeiss, Jena, Germany). Pictures where taken by an Axiocam HR and analyzed by Axiovision 4.2 Software package. For GFP documentation the filterset 09 from Zeiss has been applied (excitation 450-490 nm, beamsplitter 510 nm, emission 515 nm).

For Confocal Laser Scanning Microscopy (CLSM), immature seeds of *A. thaliana* were isolated and fixed as above description. By this way the seeds were immobilized without being squashed. Confocal laser scanning microscopy of the whole seeds was performed on a Zeiss 510 Meta CLSM (Carl Zeiss, Jena, Germany). Since the overall fluorescence was relatively weak, a lambda-scan was performed to positively identify the presence of GFP. For this purpose the probes were illuminated with 488 nm laser light. The emission profile between 495 to 580 nm was recorded and compared to that of a known GFP-positive probe. Seeds for which the presence of GFP could be demonstrated unambiguously were further analyzed by single recordings and Z-stacks, measuring the emission between 505 and 537 nm after excitation by 488 nm laser light.

2.2.4.9. Screening and characterization of mutant lines

Genomic DNA was isolated by the methods described above. Transferred genes were confirmed by PCR using appropriate primers. The positive lines were selected for RNA isolation and Northern blotting according to the protocol above. Plant protein extracts were analyzed by Western blotting. Strongly expressing plants were analyzed in detail by electron microscope.

Dehydration treatment of mutant plants:

In vitro:

Four-week-old plants of mutant and wild type lines were transferred to mediums containing 0.1-0.3 mM NaCl or 0.2-0.4 mM mannitol or 5-20% polyethylene glycol (PEG) 8000. Plants were grown at 22°C with 10000 lux of white light for 16 h. Plant development was evaluated everyday.

Greenhouse:

The mutant and wild type lines were seeded and grown on soil under greenhouse condition. After four weeks, these plants were transferred into dehydrated condition with 60% humidity. At different time points, plants were

transferred to rehydrate under saturated water condition. Dehydration response was estimated by counting surviving of plants under condition of rehydration.

2.2.5. Two hybrid system

2.2.5.1. Transformation of yeast

Yeast transformation was followed by the LiAc method (Ito *et al*., 1983), as modified by Schiestl and Gietz (1989), Hill *et al*. (1991), and Gietz *et al.* (1992).

2.2.5.2. Yeast mating

Yeast mating was performed following CLONTECH manual based on the method described by Adams *et al.* (1997). The Y187 strain containing the bait protein was inoculated in SD medium without tryptophan (Trp$^-$) at 30°C overnight. Cells were collected by centrifugation and resuspended in 5 ml of the same medium. This suspension was mixed with a 1 ml aliquot of the AH109 library in a sterile 2-L flask. 45 ml of 2X YPDA medium supplemented with kanamycin was added and the culture incubated at 30°C with gentle swirling. After 24 h of mating, cells were collected by centrifugation and the pellet was washed with 0.5X YPDA. The yeast pellet was resuspended in 10 ml of 0.5X YPDA. Transgenic yeasts were selected on medium without tryptophan and leucine (Leu$^-$) amino acids. Transgenic yeasts containing putatively interacting protein were selected on medium without histidine, leucine, and tryptophan (His$^-$/Leu$^-$/Trp$^-$) or adenine, histidine, leucine, and tryptophan (Ade$^-$/His$^-$/Leu$^-$/Trp$^-$). These yeast colonies were analyzed by PCR and sequencing.

2.2.5.3. Analysis of transgenic yeast

Fresh yeast colonies were picked up and dissolved in 0.2% SDS. This solution was heated at 90°C and centrifuged for 1 min. The crude DNA present at the supernatant was collected and used for PCR with appropriate primers. PCR products were purified by QIAgen Kit and used for sequencing.

3. Results

3.1. The BURP gene family in the *Arabidopsis* genome

Based on data of the *Arabidopsis* Genome Initiative (AGI), the *BURP* domain gene family of *Arabidopsis thaliana* consists of 5 clearly defined members. They include the gene At1g49320, which encodes a protein with strong similarity to the Unknown Seed Protein (USP) of *Vicia faba*, described in the introduction. In this study this gene is named *AtUSPL1*. A second gene, At5g25610 named *AtRD22* encodes the previously described gene product Responsive to Droughtness AtRD22 (Yamaguchi-Shinozaki and Shinozaki, 1993). Three more genes, which are very similar in sequence and size, code for proteins with a high content of aromatic amino acids and are similar to the non-catalytic β-subunit of a polygalacturonase isozyme. They are named *PG1, PG2*, and *ARO*.

Sizes and sequences of these proteins are summarized in Table 3.1 and Figure 3.1.

Table 3.1. Compilation of all BURP domain protein genes in the *Arabidopsis* genome

Gene	Molecular mass	Remarks
At1g49320	32 kD	AtUSP-like1, unknown function
At5g25610	42 kD	AtRD22, Responsive to Droughtness
At1g23760	68 kD	PG1, non catalytic β-subunit of polygalacturonase isozyme 2
At1g60390	68 kD	PG2, non catalytic β-subunit of polygalacturonase isozyme 1
At1g70370	68 kD	ARO, glycoprotein rich in aromatic amino acids, non catalytic β-subunit of polygalacturonase isozyme 1

At1g49320 *(**AtUSPL1***)*

*MASTFRLSISFLTLILFSLWVV*EAHTSRKLISIKEKEGQDISHLLKDGEFDDPSLYMYFTLNDLKLGTKL
LIYFYKNDLQKLPPLLTRQQADLIPFTKSKLDFLLDHFSITKDSPQGKAIKETLGHCDAKAIEGEHKFCG
TSLESLIDLVKKTMGYNVDLKVMTTKVMVPAQNSISYALHNYTFVEAPKELVGIKMLGCHRMPYPYAVYY
CHGHKGGSRVFEVNLVTDDGRQRVVGPAVCHMDTSTWDADHVAFKVLKMEPRSAPVCHFFPLDNIVWVTK

At5g25610 *(**AtRD22***)*

*MAIRLPLICLLGSFMVVAIAAD*LTPERYWSTALPNTPIPNSLHNLLTFDFTDEKSTNVQVGKGGVNVNTH
KGKTGSGTAVNVGKGGVRVDTGKGKPGGGTHVSVGSGKGHGGGVAVHTGKPGKRTDVGVGKGGVTVHTRH
KGRPIYVGVKPGANPFVYNYAAKETQLHDDPNAALFFLEKDLVRGKEMNVRFNAEDGYGGKTAFLPRGEA
ETVPFGSEKFSETLKRFSVEAGSEEAEMMKKTIEECEARKVSGEEKYCATSLESMVDFSVSKLGKYHVRA
VSTEVAKKNAPMQKYKIAAAGVKKLSDDKSVV**CH**KQKYPFAVFY**CH**KAMMTTVYAVPLEGENGMRAKAVA
V**CH**KNTSAWNPNHLAFKVLKVKPGTVPV**CH**FLPETHVVWFSY.

At1g60390 *(**PG1***)*

*MNNIEATLFLCFFCIFSSSNVH*FAGAKQTAGNITPSENPFTPKASLIRYWNNHINGDSPKPSFFLSKASP
LTAVDSTRFASLASNHALNTHHSDFCSAAKLFCFPELAAHSLEKHGDDVNFAAYSGKNFTNYGSDRLSGA
DSFKNYSGGDNIAVDSFRRYSRNSAGHDDGFTNYAGEVNVADQSFTTYATGTTGGSGEFTNYNTDANEPN
GRFTSYSDKANGRSQTFTTYSENGNTGYQSFTSYSKNGNGAPNEFSGYGTGSNVVNTGFTKYGESANGAN
DSFTSYGENGNVPVNEFKGYGDGGNGAVYGFKNYRDQSNIGVDSFSSYAKNSNNEKVNFVNYGKSFNLGS
DNFTGYGQDNVGGNVSFKTYGQGQSFKVYTKDGVVFARYSNNVSSNGKTVNKWVEEGKFFREAMLKEGTL
MQMPDIKDKMPKRTFLPRNIVKNLPFSSSTIGEIWRVFGAGENSSMAGIISSAVSECERPASHGETKRCV
GSAEDMIDFATSVLGRGVVVRTTENVVGSKKKVVIGKVNGINGGDVTRAVS**CH**QSLYPYLLYY**CH**SVPRV
RVYETDLLDPKSLEKINHGVAI**CH**IDTSAWSPSHGAFLALGSGPGQIEV**CH**WIFENDMTWNIID.

At1g70370 *(**PG2***)*

*MLKQFLLLQSFSFFLFNVVIVGG*RTFGGGFSAEENPFTPKASLVRYWNKEIRGQSPRSEFLISKASPLNA
VDSATFSKLAAANSLPTRFPDFCSAANLFCFPDLGASLEKHDDDVKFSVYDQKNFTNYGNARAGGADSFK
NYSKDGNVVTDSFRRYSRNAAGHDDKFTVYGENSNVVEEGFNSYGTFGTGGAGDFTNYQNNVNNPTSRFT
AYSDGGNGRSQTFKTYTHEANAGNGQSFTSYGKNGNGVPNEFTSYGVSSNVIGSGFSNYGESGNAANDTF
TSYGSDGNVPQNNFNNYGASGNAAVDTFANYRDKANVGDDSFSSYAKDSNSEKVNFVNYGQSFNPGSETF
TGYGKGAEGSKLSFKTYTPNSTFKDYAKKGVAFAKYNVSTTTANTVGDGKTVNKWIEPGKFFRESSLKEG
TVIPMPDIKDKMPKRSFLPRSIITKLPFSTSKLGEIKRIFHAVENSTMGGIITDAVTECERPPSVGETKR
CVGSAEDMIDFATSVLGRSVVLRTTENVAGSKEKVVIGKVNGINGGKLTKAVS**CH**QSLYPYLLYY**CH**SVP
KVRVYEADLLELNSKKKINHGIAI**CH**MDTSSWGPSHGAFLALGSKPGRIEV**CH**WIFENDMNWAIAD.

At1g23760 *(**ARO***)*

*MRKQFVFLLPFLSRLYHVVIAGG*ERLTNGGFSPEINPFTPKASLVRYWNKQIHHQSPGSAFLLSKASPLT
AVASATFAKLASENALPDRLPDFCSAANLFCFPDLGPTLEKHSNDVKFSVYDQRNFTNYGTIHPGGADSF
KKYSQNGNAVTDSFRRYSRDAAGHEDKFTGYADNSNVVEEKFNSYGTSATGGSGDFKNYQSGVNNPTTRF
KAYSDEGNGRAQTFKTYTHEANSGPGQSFTSYGKNGNGAPNEFSSYGVSSNVIGSDFSNYGENGNAANDT
FKGYGGDGNVPQNNFKSYGASGNAAVDTFLNYRDNANVGDDSFSSYAKNSNFEKVDFVNYGKSINPGSES
FSGYGKGAEGNKIDFKTYTQNSTFKDYTKTGVEFAKYNRSSLGGGKTVNKWVEPGKFFRESMLKEGTLIW
MPDIKDKMPKRSFLPRSIVSKLPFSTSKIAEIKRVFHANDNSTMEGIITDAVRECERPPTVSETKRCVGS
AEDMIDFATSVLGRSVVLRTTESVAGSKEKVMIGKVNGINGGRVTKSVSCHQSLYPYLLYYCHSVPKVRV
YESDLLDPKSKAKINHGIAICHMDTSAWGANHGAFMLLGSRPGQIEVCHWIFENDMNWAIAD.

Figure 3.1. Amino acid sequences of the five BURP domain proteins encoded by
the *Arabidopsis* genome

The BURP domain of each protein is given in red; the CH motifs are printed
in bold letters. The signal peptides are shown in italic and underline, the N-
terminal part of AtUSPL1 is given in gray, the 4 TXV repeats in AtRD22 are
given in green and the 21 FXXY repeats in PG1, PG2 and ARO are shown
in blue.

All five BURP domain proteins exhibit an overall modular structure consisting of a transient signal peptide and the C-terminal BURP domain, both separated by a variable internal region without repeat (P) in AtUSPL1, and with four repeated peptides (TXV) in AtRD22 and 21 repeated peptides (FXXY) in PG1, PG2 and ARO.
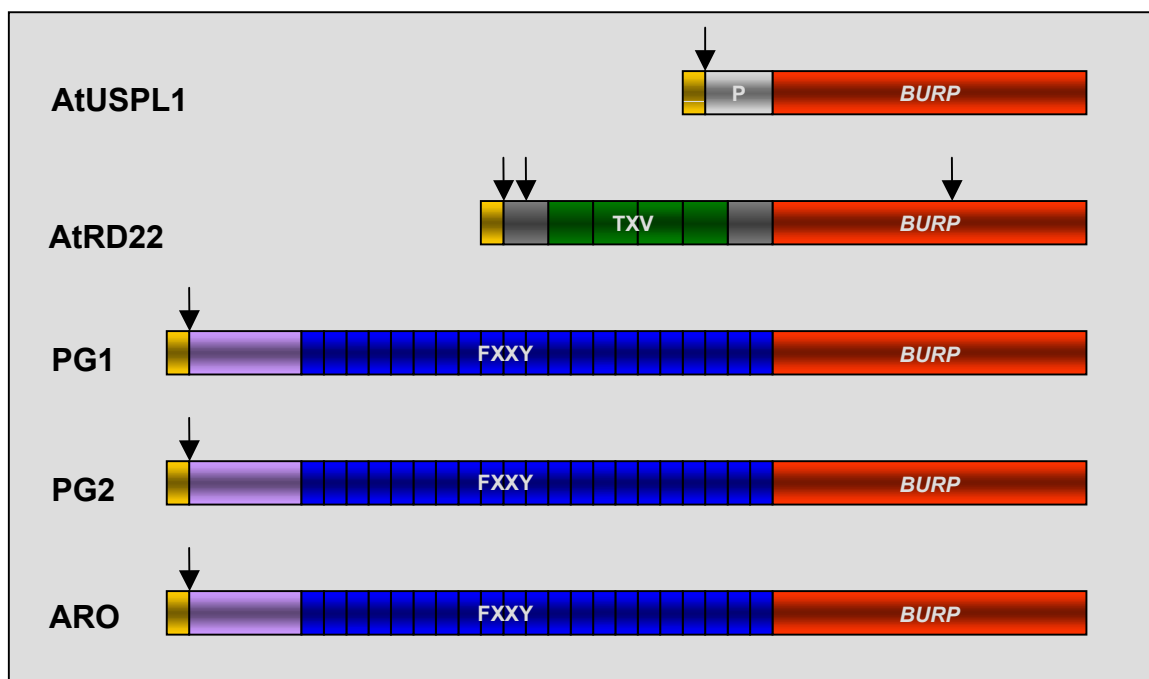

Figure 3.2. Schematic structure of the five members of the BURP domain protein family found in the *Arabidopsis* genome

All proteins include a transient signal peptide (yellow). In the AtUSPL1 protein the BURP domain is preceded by a short non repetitive peptide P (light grey). AtRD22 is characterized by four TXV repeat motifs (green) embedded in short flanking regions (grey). The remaining three proteins (PG1, PG2 and ARO) contain the characteristic 21 fold FXXY repeat motif (blue) upstream of the BURP domain and are separated from the signal peptide by a short region (violet). Based on this totally different N-terminal amino acid sequences, the latter three proteins can be grouped into a subgroup due to their similarity. The BURP domain (red) of all five members is highly conserved. The arrows indicate the position of introns in the genes.

The mentioned subgroup (*PG1, PG2* and *ARO*) has been described in tomato (Zheng *et al*., 1992); therefore, the current studies mainly focuses on the two smaller genes *AtUSPL1* and *AtRD22*.

3.2. Expression of *AtUSPL1* and *AtRD22* based on GENEVESTIGATOR data

GENEVESTIGATOR (Zimmermann *et al*., 2004) is a novel online tool, which was developed to facilitate in silico expression analysis, comprising a gene expression database and a number of querying and functionality analysis. Analysis of gene expression data indicated that the accumulation of *AtUSPL1* mRNA is the highest in hypocotyl, embryo and root. During seed development, *AtUSPL1* is expressed at lower level with a small peak at later stage. In contrast, *AtRD22* transcript accumulates mainly in the petal, pedicel, silique and rosette. *AtRD22* is expressed during the early stages of seed development (https://www.genevestigator.ethz.ch).
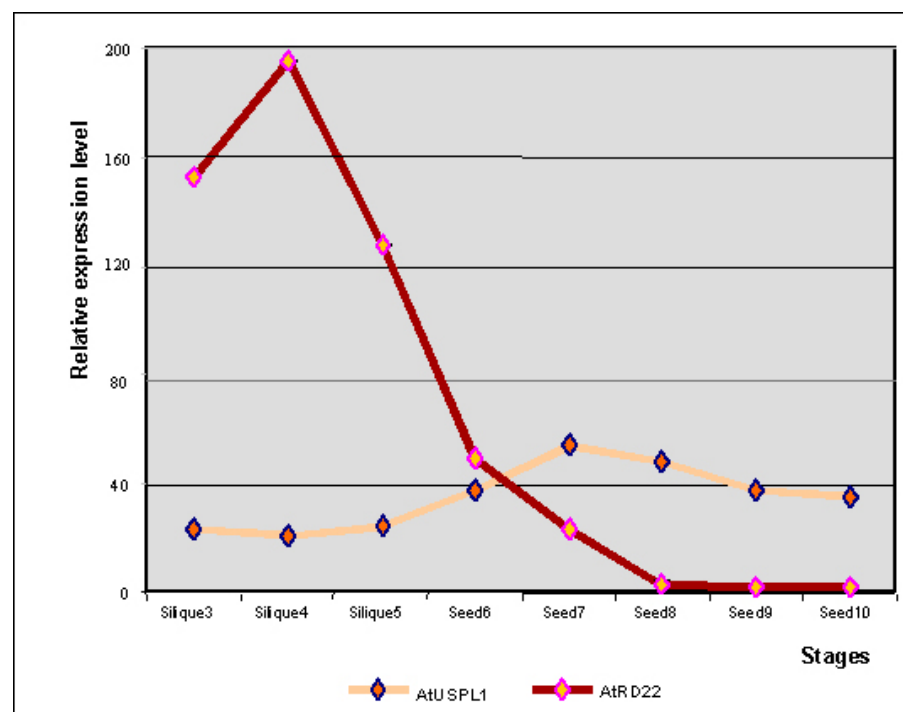
Figure 3.3. *AtUSPL1* and *AtRD22* transcription level during seed development
    AtRD22 is expressed at early stages, and *AtUSPL1* is expressed at later
    stage of seed development (Genevestigator data).

3.3. AtUSPL1 as putative homolog of VfUSP of *Vicia faba* and BnBNM2 of *Brassica napus*

Based on amino acid sequence, the AtUSPL1 is similar to BnBNM2 (Boutilier *et al*., 1994) and VfUSP (Bassüner *et al*., 1988) in length and sequence (Figure 3.4).

```
AtUSPL1          MASTFRLSISF---LTLILFSLWVVEAHTSRKLIS
BnBNM2           MAS-LRFSVTFPALLSLLLLSLWVVEAYTSPKLIS
VfUSP            M-----EFAHLT-VLSLFCLAFVGITATSSGEDYW
                                                ↑

AtUSPL1          IKEKEGQDISHLLKDGEFDD-PSLYM / BURP
BnBNM2           NNEQEGQNISHLFKDGEFED-PTMYM / BURP
VfUSP            QSIWPNTPLPKTFSDLSIPSGKTNSL / BURP
```

Figure 3.4. Amino acid alignment of the N-terminal sequences and the start of the BURP domain of AtUSPL1, BnBNM2 and VfUSP

The arrow labels the position of the signal peptide processing in VfUSP (Bassüner *et al*., 1988).

The structural similarity suggests that *AtUSPL1* might have a homologous function as the mainly embryo-specific *VfUSP*-gene of *Vicia faba* and the *BnBNM2*-gene of *Brassica napus*, known to be specific for microspore derived embryos (Boutilier *et al*., 1994; Hattori *et al*., 1998).

3.4. Expression of an *AtUSPL1* promoter-GUS reporter construct

In order to understand the regulatory mechanism of *AtUSPL1* gene expression, the *AtUSPL1* gene promoter was fused to the β-glucuronidase (*GUS*) reporter gene. The chimeric gene fusion was introduced into the *Arabidopsis* genome *via Agrobacterium*-mediated transformation. The expression of the reporter gene during seed development and plant growth was monitored using a histochemical X-Gluc assay.

In siliques of transgenic plants, GUS enzyme activity was only detected in funiculi of mature seeds. During seed germination and seedling growth, GUS

activity could also be detected in roots and flowers as well as in young leaves, hypocotyls, and stems (Figure 3.5). These findings are similar to analyzes of the *AtRD22* promoter that drives the expression of the *GUS* gene mainly in flowers and bolted stems (Iwasaki *et al*, 1995).
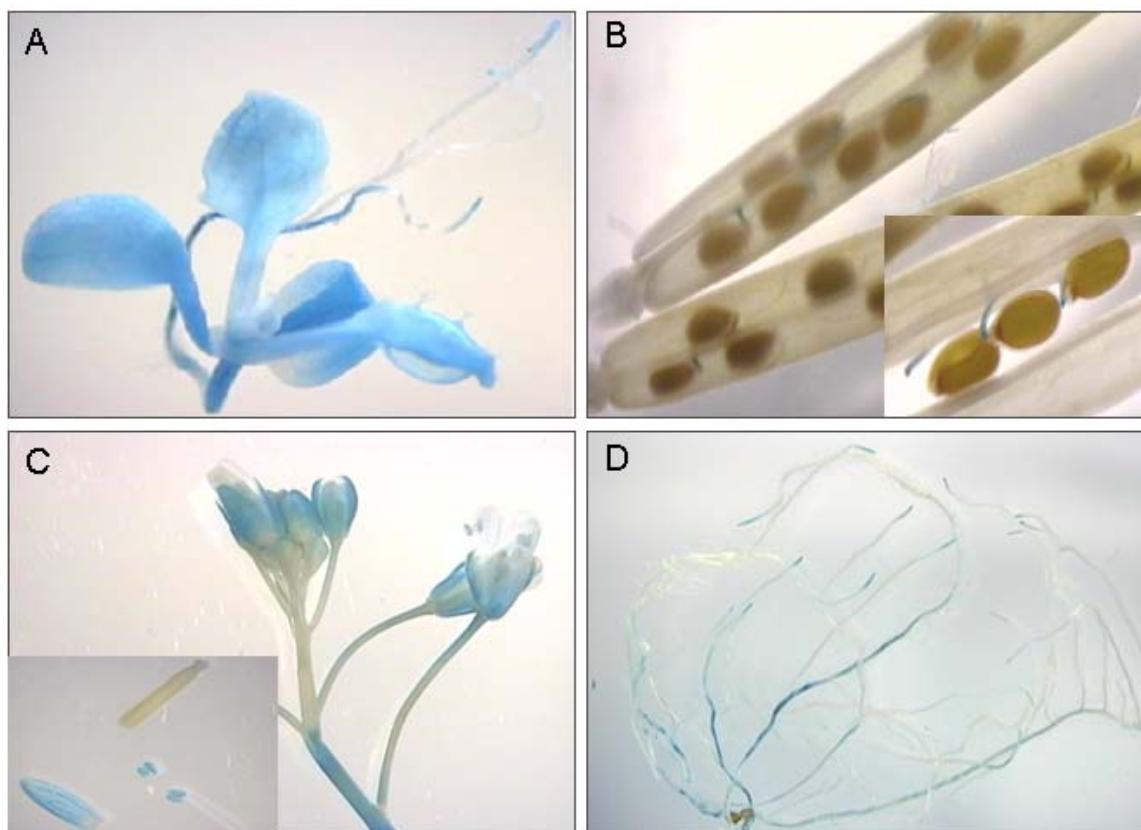
Figure 3.5. Histochemical localization of GUS activity in transgenic plant

AtUSPL1 promoter was regulating the *GUS* gene expression in seedling (A); in funiculus of mature seeds (B); in flowers and stems (C); and in roots (D).

3.5. Expression of the endogenous *AtUSPL1* gene

To characterize the tissue specific expression of the endogenous *AtUSPL1* gene in *Arabidopsis*, Northern hybridization was carried out using *AtUSPL1* cDNA as a specific probe. As shown in Figure 3.6, the *AtUSPL1* transcript is preferentially abundant in flowers and roots. Weak expression has been detected in leaves and other tissues (Figure 3.6).
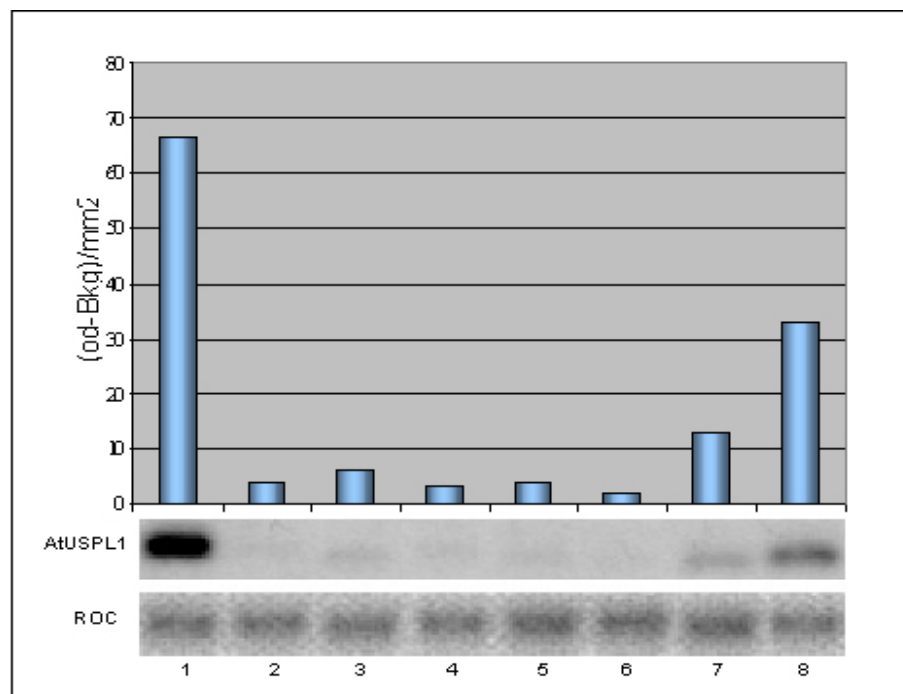
Figure 3.6. Northern blot analysis of *AtUSPL1* gene expression in *Arabidopsis* wild type plant

RNAs isolated from 1) roots, 2) leafs, 3) shoots, 4-6) young siliques, and 7-8) flowers of Columbia plants. They were separated on 1.5% agarose gels and subsequently blotted onto Nylon membranes. One filter was hybridized with ^{32}P labeled *AtUSPL1* cDNA (**upper panel**), and other filter was hybridized with a ^{32}P labeled cDNA of a housekeeping gene (*ROC*) (**lower panel**). Od - optical density; Bkg – background.

3.6. Immunohistochemical localization of endogenous AtUSPL1 in seeds

In order to check the expression of the endogenous *AtUSPL1* gene at protein level, the immunohistological approaches has been applied. The AtUSPL1 protein was expressed in *E. coli* as hexa-histidine-tagged fusion protein using the pET system (Clontech) and purified under denaturing conditions on Ni^{2+}-NTA agarose. After several subcutaneous injection of the purified antigen, specific polyclonal antibodies were generated in rabbits. The activity and specificity of the antisera were tested by Western blot using bacterial extracts

isolated from transformed and non-transformed *E.coli* strains. The produced antisera were also capable to recognize AtUSPL1 in *AtUSPL1* overexpressing *Arabidopsis* lines (Figure 3.7). The anti-AtUSPL1 antiserum was purified and applied for immunolocalization of AtUSPL1 in plant tissues.
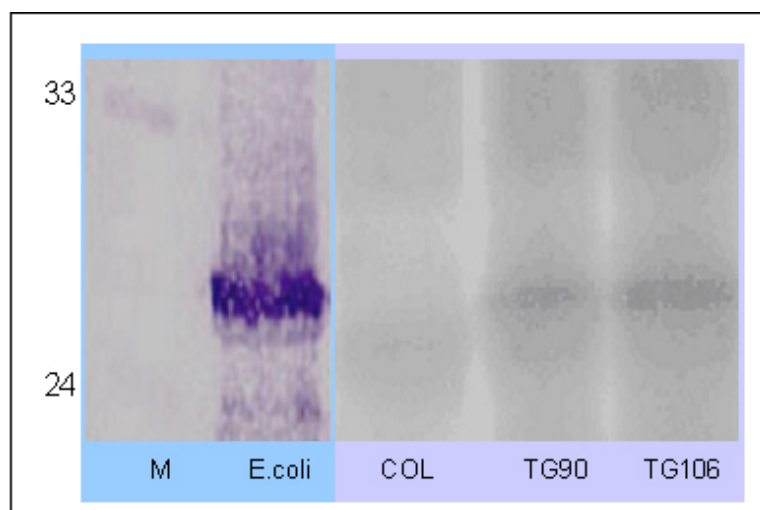
Figure 3.7. Detection of recombinant AtUSPL1 in transgenic *E.coli* and AtUSPL1 overexpressing *Arabidopsis* lines by Western blot analysis

Total protein was extracted from transgenic *E.coli* and overexpressing *Arabidopsis* lines (TG), and separated on 12,5% SDS-PAGE. Extract from Columbia wild type (COL) was used as specificity control. Recombinant AtUSPL1 was detected by anti-AtUSPL1 antibody followed by anti-rabbit IgG antibody conjugated to ALP or HRP. Signals were generated either by color reaction using BCIP/NBT as substrate (*E.coli)* or by chemiluminescence (plants). M - molecular masses of standard proteins in kD are shown on the left.

The intracellular localization of AtUSPL1 was performed in cooperation with Dr. T. Rutten (Structural Cell Biology Group, IPK-Gatersleben). Using anti-AtUSPL1 antibodies, an immunolocalization investigation was carried out. Figure 3.8 shows that AtUSPL1 was not detected in young seeds (heart stage), but it appeared in cotyledon cells during later seed development. The endogenous AtUSPL1 protein was detected in the protein storage vacuoles of cotyledon cells (Figure 3.8).
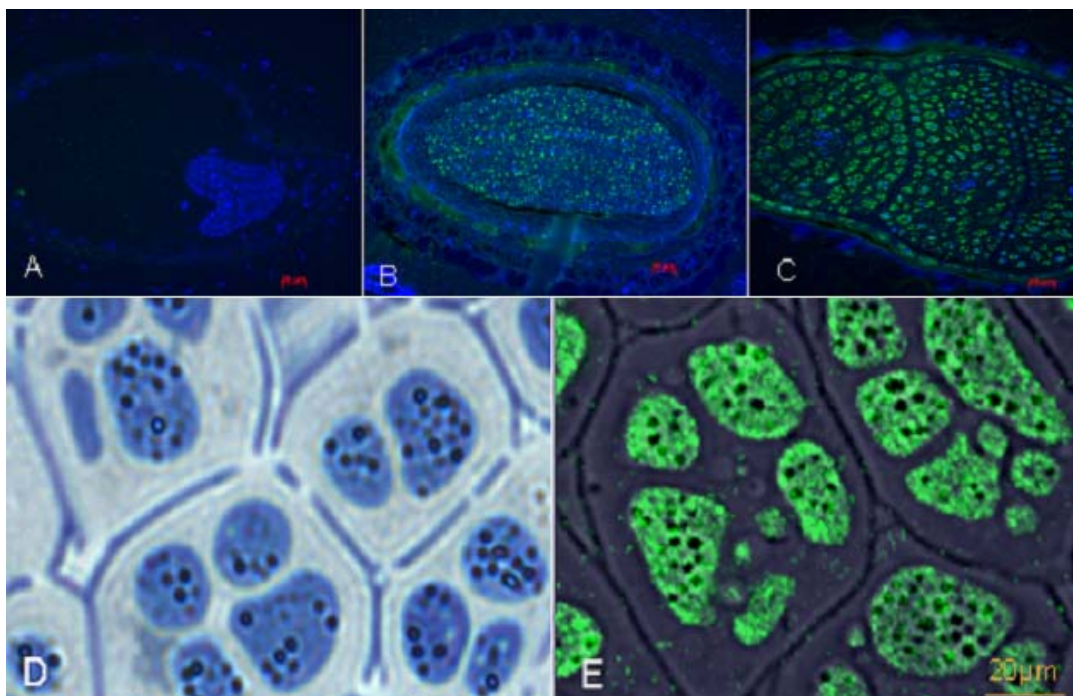
Figure 3.8. Localization of endogenous AtUSPL1 in protein storage vacuoles of cotyledons of Columbia wild type seeds

Immunostaining using anti-AtUSPL1 antibodies and fluorescence labeled secondary antibodies: A- heart stage embryo, B- middle stage seed, C- later stage seed, D- Toluidine blue staining; E- detection of AtUSPL1 in the protein storage vacuoles of cotyledon cells. The black dots within the protein storage vacuoles might represent phytate crystalloids. The red magnification is 20μm.

3.7. Characterization of a gain-of-function mutant of AtUSPL1

To characterize functions of the AtUSPL1 protein, two different approaches have been applied. Firstly, the *AtUSPL1* gene was overexpressed to create a gain-of-function situation, and the phenotype of transgenic plants was characterized. Secondly, the *AtUSPL1* gene was mutated by T-DNA insertions to create a loss-of-function situation to analyze for phenotypic changes.

For the overexpression, the coding sequence of the *AtUSPL1* gene was cloned into the GATEWAY-vector pBENDER (M. Jakoby, B. Weisshaar) to be expressed under the control of the CaMV 35S promoter. This construct was

transformed into *Arabidopsis* Columbia *via Agrobacterium*-mediated transformation. In more than 100 independent transgenic lines generated, the presence of the *AtUSPL1* transformed gene was confirmed by PCR. The transcript levels of these plants have been compared by Northern hybridization (Figure 3.9).

3.7.1. Seed phenotypes of overexpression lines

Based on Northern hybridization data of the transgenic lines, three levels of expression were obtained i) strongly expressing like TG90; ii) middle expressing like TG164, and iii) weakly expressing like TG43.
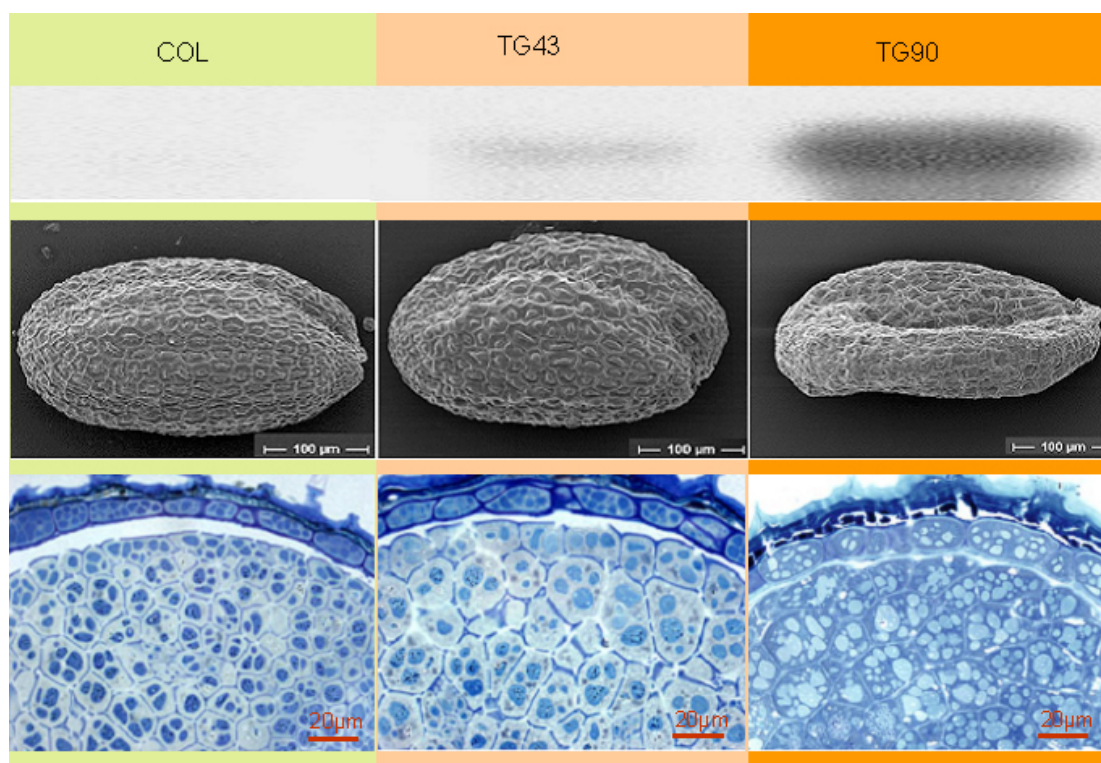
Figure 3.9. Ectopic expression of *AtUSPL1* in transgenic *Arabidopsis* plants

Upper panel- Northern blot analysis of *AtUSPL1* expression in Colombia wild type (COL) and transgenic overexpression lines (TG). The *AtUSPL1* transcript levels were detected by using *AtUSPL1* cDNA as hybridization probe. **Middle panel**- seed phenotype. **Lower panel**- seed histology. The expression level of the transgene in various lines correlates well with the severity of the seed phenotype.

Mature seeds of transgenic *Arabidopsis* plants with *AtUSPL1* overexpression were phenotypically and ultrastructurally analyzed and compared to Columbia wild type seeds using light and electron microscopical investigations. When the *AtUSPL1* transgene is expressed at a low level like in TG43, the seed phenotype is similar to wild type. However, in lines with high transcript level like TG90, the mature seeds are shrunken (Figure 3.9). In all analyzed seeds the *AtUSPL1* transcript levels of transgenic lines correlated well with the severity of the shrunken seed phenotype shown in Figure 3.9.

3.7.2. Ultrastructure of protein storage vacuoles in parenchyma cell seeds

To understand the changes observed in transgenic seeds, a detailed ultrastructural analysis was carried out. Figure 3.10 shows results of the ultrastructure of storage parenchyma cells in mature seeds of transgenic lines and Columbia wild type.
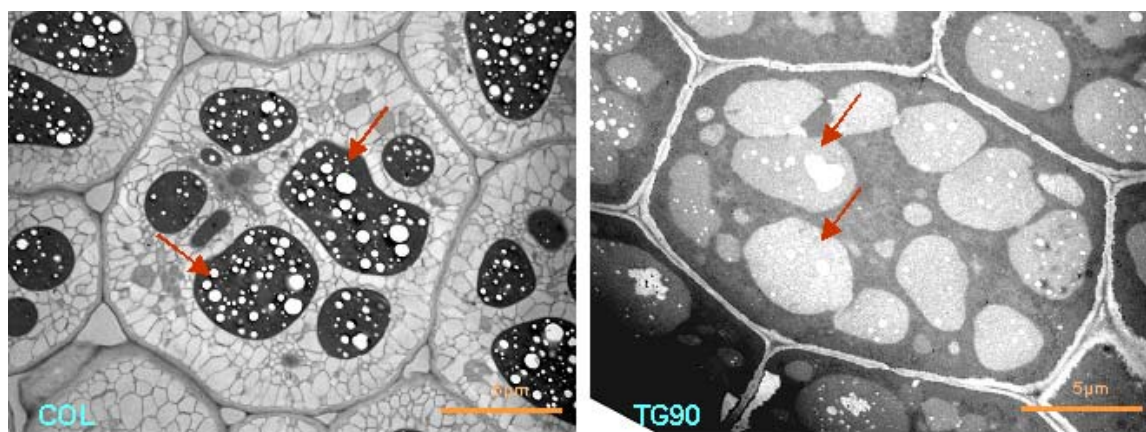
Figure 3.10. Alterations of protein storage vacuoles in storage parenchyma cells of transgenic seeds (TG90) in comparison to the Colombia wild type (COL)

Electronmicroscopical analysis reveals the occurrence of protein storage vacuoles (red arrow) with much lower electron density in TG90 in comparison to wild type.

The protein storage vacuoles of mature wild type seeds are well filled with proteins, whereas transgenic lines like TG90 have more but smaller protein storage vacuoles with strongly reduced protein amounts indicated by their lower contrast.

3.7.3. Immunological detection of cruciferin in plant seeds

The overexpression of *AtUSPL1* leads to serious distortion of the protein storage vacuole structure. Immunolabeling using anti-cruciferin antibodies revealed that the storage protein cruciferin appears as ring-like rim within the protein storage vacuoles of storage parenchyma cells of transgenic seeds, whereas the protein storage vacuoles of Columbia wild type cells are completely filled with cruciferin (Figure 3.11).
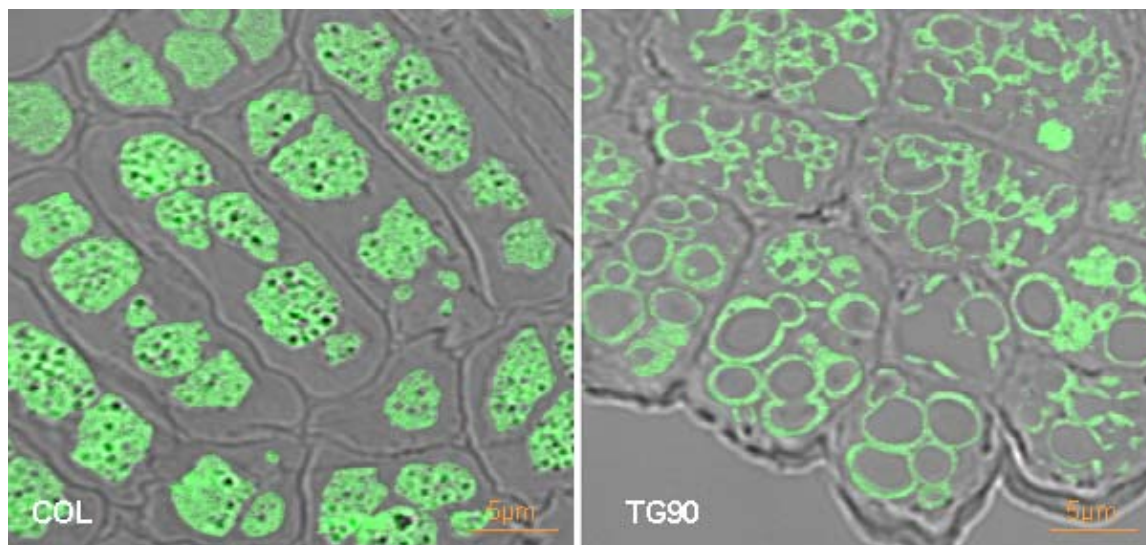
Figure 3.11. Immunohistological detection of the 12S storage globulin, cruciferin in *Arabidopsis* seeds

Using anti-cruciferin antibody followed by indirect immunofluorescence staining with a fluorescence labeled secondary antibody Alexa 488, the major storage globulin, cruciferin, was labeled in protein storage vacuoles of the storage parenchyma cells of Columbia wild type seeds (COL). In overexpressing line (TG90), the storage protein cruciferin was detected at the periphery of the protein storage vacuoles.

3.7.4. Semi- quantitative detection of cruciferin in single seeds

To investigate the quantity and the polypeptide pattern of the major storage protein, cruciferin in transgenic and Columbia wild type seeds, the total salt soluble storage proteins were extracted from single seeds and electrophoretically separated by PAGE.
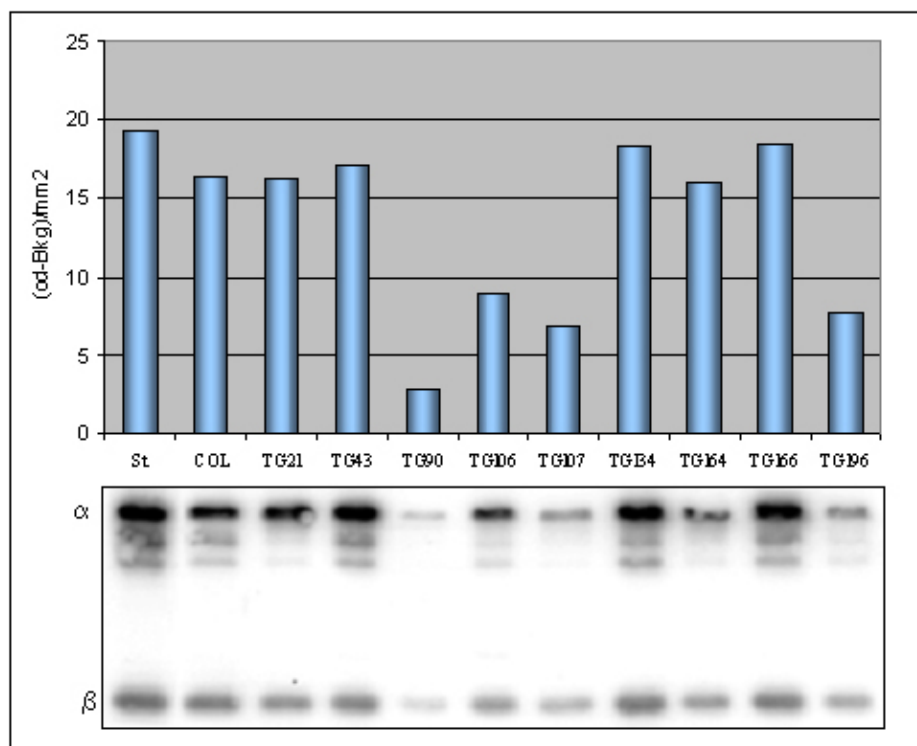
Figure 3.12. Western blot analysis of cruciferin from mature single seeds of transgenic (TG) and Colombia wild type plants (COL)

Total protein was extracted from mature single seeds from individual transgenic and wild type seeds by use of the same volume of extraction buffer. Identical aliquots of the protein extracts were separated under denaturing and reducing conditions on 12.5% polyacrylamide gel. After blotting the cruciferin was recognized by anti-cruciferin antibody (L.Rask University Uppsala, Sweden) and labeled by anti-mouse IgG conjugated to HRP. Signals were generated by chemiluminescence. The α- and β-polypeptides of the globulin cruciferin are indicated on the left. Standard (St): preparation of storage cruciferin as reference. Od - optical density; Bkg – background.

The polypeptide pattern of the storage globulin cruciferin from mature single seeds of transgenic plants showed quantitative differences in comparison with the pattern of the wild type control. The cruciferin amount in the *AtUSPL1* overexpressing (TG) was reduced and correlates well with the level of *AtUSPL1* mRNA transcripts (compare for instance TG43 and TG90 in Figure 3.12).

3.7.5. Accumulation of storage lipids and ribosome arrangement in seeds

In overexpression lines of *AtUSPL1*, distortions in the ribosome arrangement around lipid bodies in storage parenchyma cells were observed in comparison to the wild type.
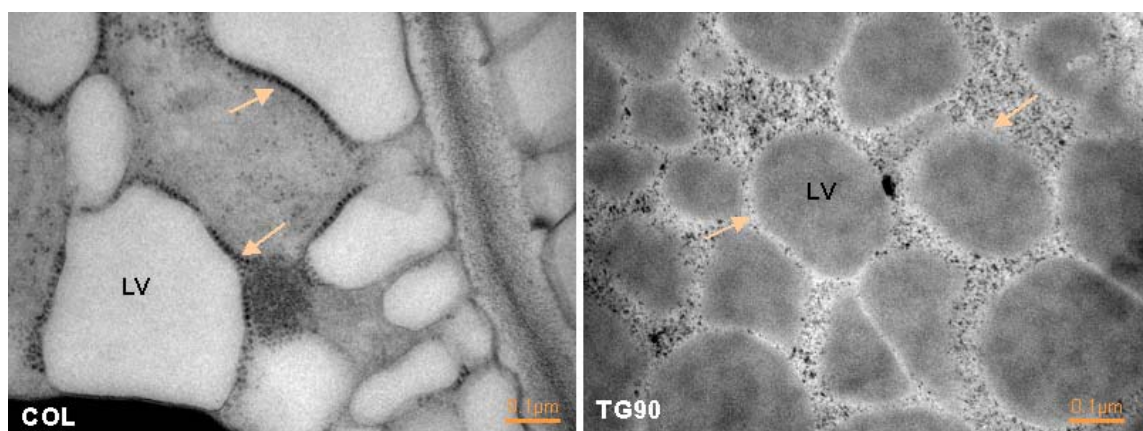
Figure 3.13. Lipid bodies in storage parenchyma cells of *AtUSPL1* overexpression lines

Electron density of lipid bodies (LV) in overexpressing seeds (TG90) was higher than that of Columbia wild type (COL). Note the precise arrangement of ribosomes (yellow arrow) along the lipid body membrane in wild type and the severe distortion of this arrangement in the overexpression line.

In addition, the electron density of lipid bodies in overexpressing seeds was found to be higher than in wild type seeds, indicating differences in the fatty acid composition of transgenic and wild type seeds (Figure 3.13).

3.7.6. Quantification of fatty acids in seeds

The different composition of fatty acids in the seeds of the overexpression lines has been checked by HPLC analysis (in cooperation with S. Knüpfer, IPK-Gatersleben). As shown in Figure 3.14, most weakly and middle expressing transgenic lines exhibit approximately two-fold increase of oleic acid in comparison to non-transformed wild type (COL1, COL2) controls.
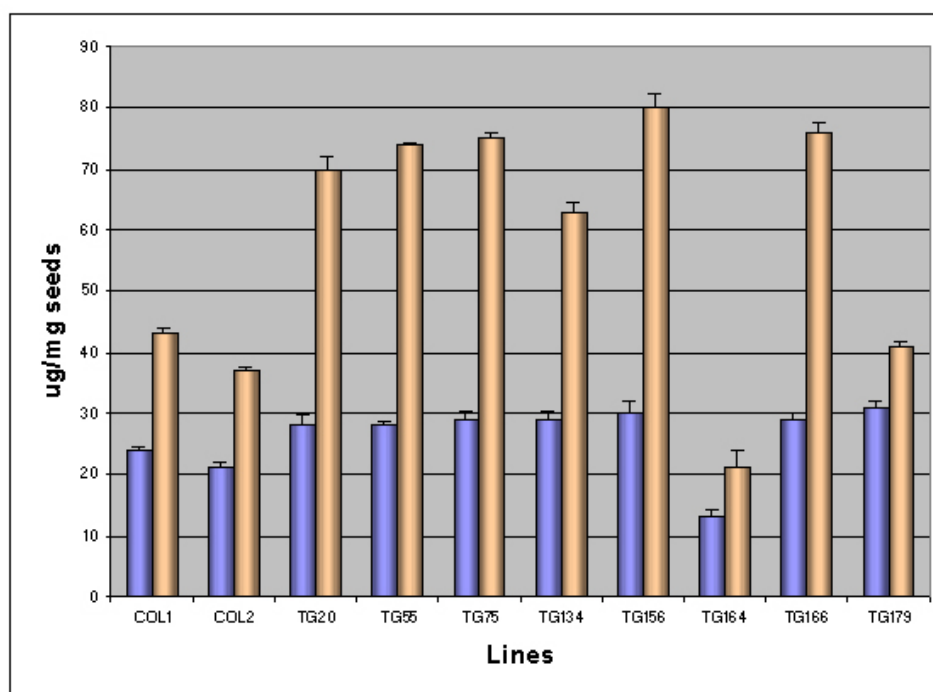
Figure 3.14. HPLC analysis of saturated and unsaturated fatty acids in *AtUSPL1* overexpressing seeds (TG) and Columbia wild type (COL)

The unsaturated fatty acids (yellow) in seeds of transgenic plants exhibit two-fold increase in comparison to wild type controls. The saturated fatty acids (violet) seem to be indifference in seeds of both plant types.

3.7.7. Plant phenotype of overexpression lines

The alterations of the content and structure of protein storage vacuoles, and lipid bodies in transgenic seeds correlate with distortions in seed germination as well as plant development.

Table 3.2. Characterization of transgenic plant phenotype concerning storage product accumulation and seed germination

Line	Protein storage vacuole	Lipid body	Germination (%)
TG90	+++	+++	60
TG106	+++	+++	7
TG107	+++	+++	0
TG134	++	+	45
TG164	++	+	51
TG43	Like COL	Like COL	94
TG166	Like COL	Like COL	90
TG194	Like COL	Like COL	n.d.

COL, Columbia wild type; TG, overexpression lines. Level of phenotypical differences compared to COL: (+) - weak; (++) - medium; (+++) - strong.
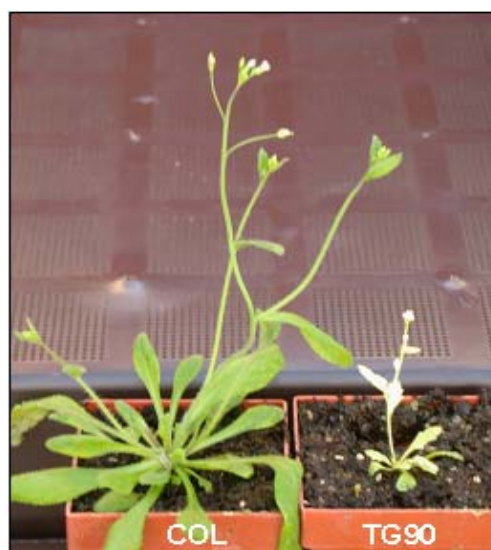
Figure 3.15. Phenotypical differences between wild type and *AtUSPL1* transgenic plants during plant development

A strong overexpression line (TG90) grows weekly compared to Columbia wild type plants (COL).

The seeds of the high expression lines TG90, TG106 and TG107 have low germination rates and grow weakly compared to the control (Figure 3.17). The germination and growth rate of low expression lines like TG43, TG166 and TG194 did not differ from wild type (Table 3.2).

3.8. Characterization of loss-of-function mutants of AtUSPL1 and AtRD22

3.8.1. Screening and molecular characterization

In order to analyze the loss-of-function situation of selected BURP-genes in *Arabidopsis*, two T-DNA-insertion lines of the *AtUSPL1* gene have been isolated from two different T-DNA insertion collections. The first one is a mutant in the Columbia ecotype isolated from the Syngenta *Arabidopsis* Insertion Library (SALK) collection, named *atuspl1-C*. The second T-DNA insertion allele in the Wassilewskaja ecotype was isolated from the *Arabidopsis* Knock-out Facility (AKF), University of Wisconsin Biotechnology Center (UWBC), named *atuspl1-W*. Both mutants were isolated and analyzed by using specific primers (USP-F /LBa1 and USP-F /JL202 for SALK and UWBC lines, respectively).
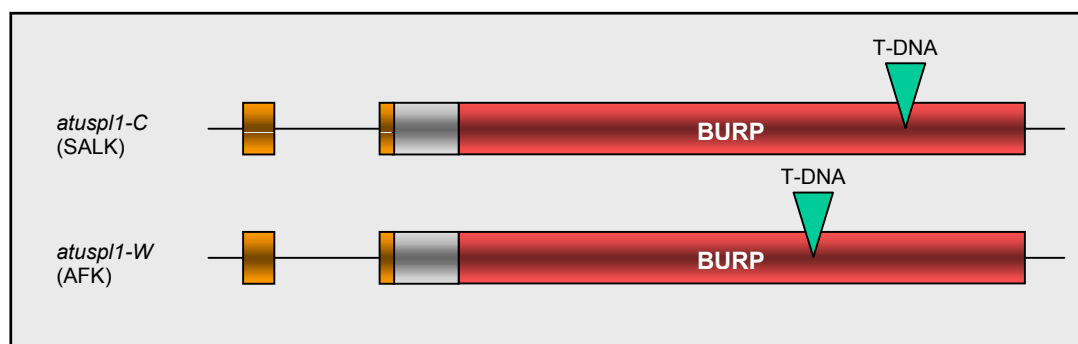
Figure 3.16. Two homozygous T-DNA insertions located in the *AtUSPL1* gene of two different ecotypes Columbia *(atuspl1-C)* and Wassilewskaja *(atuspl1-W)*

The yellow regions indicate the signal peptide; the grey region encodes the N-terminal polypeptide and the BURP domain is given in red. Thin lines represent both untranslated regions and the intron position of the gene.

Based on sequences of the PCR products, the positions of the T-DNA insertions were determined. In the *atuspl1-C* line the T-DNA insertion is located within the BURP domain of the *AtUSPL1* gene. In the *atuspl1-W* line the T-DNA insertion is located slightly upstream in the *AtUSPL1* gene at the first CH-motif of the BURP domain (Figure 3.16).
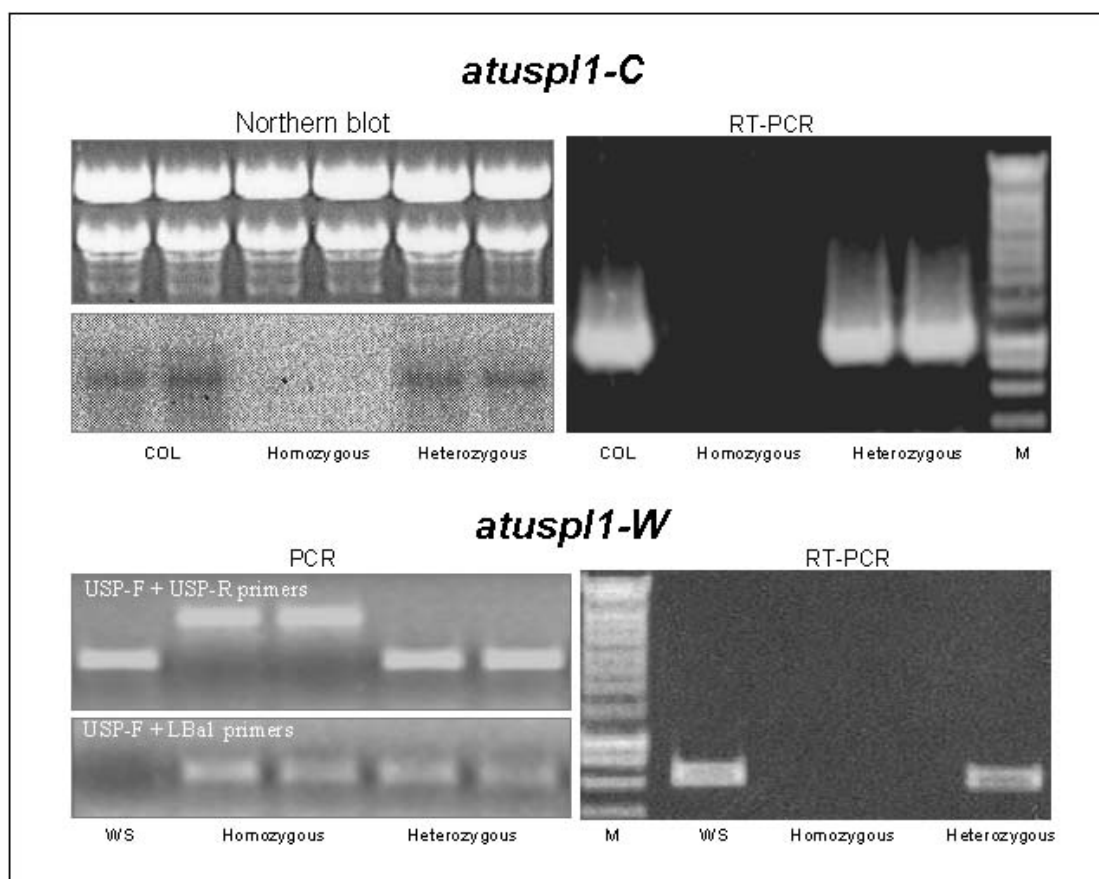
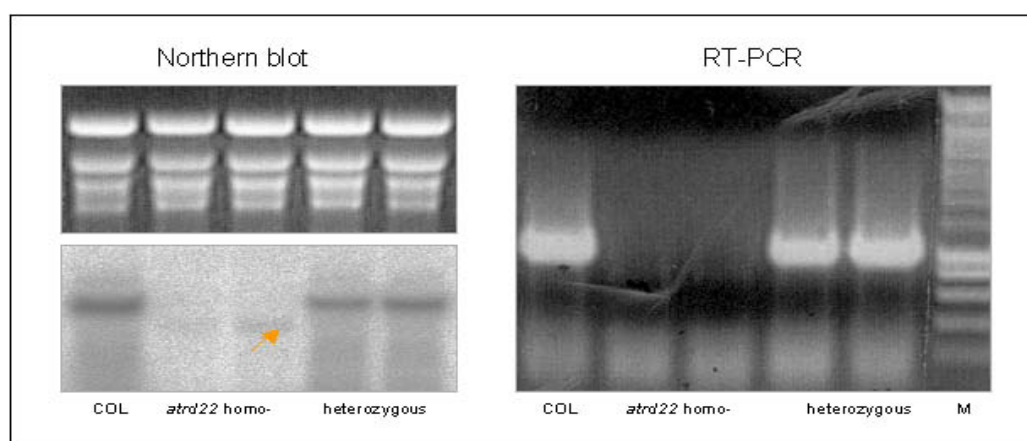**Figure 3.17.** Northern hybridization and RT-PCR analysis of *atuspl1* mutant lines

Total RNA was isolated from *atuspl1* mutant lines and wild type plants (COL and WS). For Northern hybridization, total RNA was separated on a 1.5% agarose gel and subsequently blotted onto a Nylon membrane. The filter was hybridized with ^{32}P labeled *AtUSPL1* cDNA. For RT-PCR, total RNA was used as template. Specific primers were used to amplify *AtUSPL1* cDNA from mRNA. In *atuspl1-W* PCR, the USP-F/LBa1 primers were used for the detection of T-DNA insertion lines. The USP-F/USP-R primers were used for the detection of *atuspl1* homozygous and heterozygous mutant lines. M- DNA smart ladder.

In both homozygous mutants, the transcript levels were analyzed by Northern blot and RT-PCR. As shown in Figure 3.17, *AtUSPL1* mRNA was undetectable in mutant lines by both Northern blot and RT-PCR. Furthermore, no AtUSPL1 protein could be detected in the protein storage vacuoles of *atuspl1-C* mutant seeds using immunohistology (Figure 3.18).
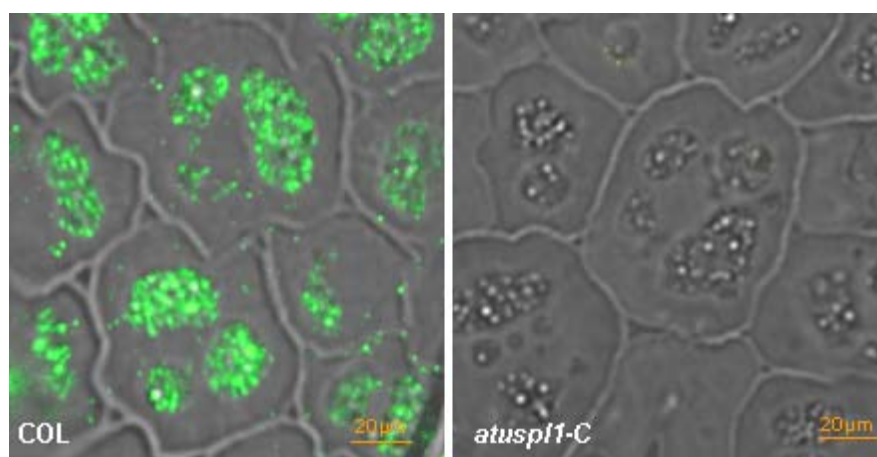
Figure 3.18. Immunohistological detection of AtUSPL1 in seeds of Columbia wild type (COL) and homozygous mutant line (*atuspl1-C*)

Using anti-AtUSPL1 antibody as primary label, the AtUSPL1 protein was only recognized in wild type seeds, but not in seeds of the mutant line.

Similarly, the *atrd22* mutant isolated and characterized from the Columbia SALK collection was analyzed using the primers RDa and LBa1. The resulting PCR product was sequenced and compared to the *AtRD22* gene. The T-DNA insertion was detected in the intron region close to the TXV repeats encoding exon (Figure 3.19).

Northern blot with *AtRD22* cDNA as a probe demonstrated that the *AtRD22* gene was not correctly expressed in the mutant line. Instead of the wild type transcript, another short mRNA was detected as a weak band. This truncated transcript might encode the signal peptide and the N-terminal region of the AtRD22 protein and then terminate within the T-DNA. The absence of an intact *AtRD22* transcript was further demonstrated by RT-PCR with specific primers (Figure 3.20).
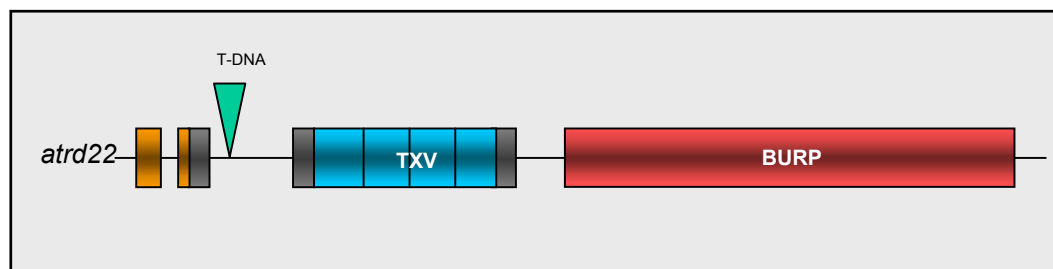
Figure 3.19. Structure of the *AtRD22* gene and localization of the T-DNA insertion in *Arabidopsis* ecotype Columbia

The yellow regions represent the signal peptide; grey labels the N-terminal protein regions, blue label indicates the TXV repeat polypeptide and the BURP domain is given in red. Thin lines represent untranslated and intron regions.

Figure 3.20. Northern hybridization and RT-PCR analysis of *atrd22* mutant lines

Total RNA was isolated from *AtRD22* mutant lines (*atrd22*) and Columbia wild type plants (COL). For Northern hybridization, total RNA was separated on 1.5% agarose gel and subsequently blotted onto a Nylon membrane. The resulting filter was hybridized with ^{32}P labeled *AtRD22* cDNA. For RT-PCR, total RNA was used as template. AtRD22 specific primers were used to amplify *AtRD22* cDNA from mRNA. In *atrd22* homozygous mutant lines, the arrow indicates a faint band of a truncated, most likely nonfunctional transcript. It might encode the signal peptide and the N-terminal region of the AtRD22 protein and then terminate within the T-DNA. The *atrd22* heterozygous mutant lines showing an intact *AtRD22* transcript like wild type plants. M- DNA smart ladder.

Immunolabeling signal could be localized within the protein storage vacuoles of both wild type seeds and mutant seeds (data not shown) by using antibodies directed against AtRD22. However, the strongly reduced level of a truncated transcript suggests that another protein containing AtRD22-like epitopes, most likely AtUSPL1, or a truncated non-functional version of AtRD22 has been synthesized in the mutant.

To analyze this possibility further, crossing has generated a double mutant of *atuspl1-C* and *atrd22* genes. This double mutant was confirmed by PCR using gene specific primers. As shown in Figure 3.21, wild type alleles of *AtUSPL1* and *AtRD22* are present in Columbia wild type and heterozygous lines but not in *atuspl1/atrd22* homozygous double mutant lines.
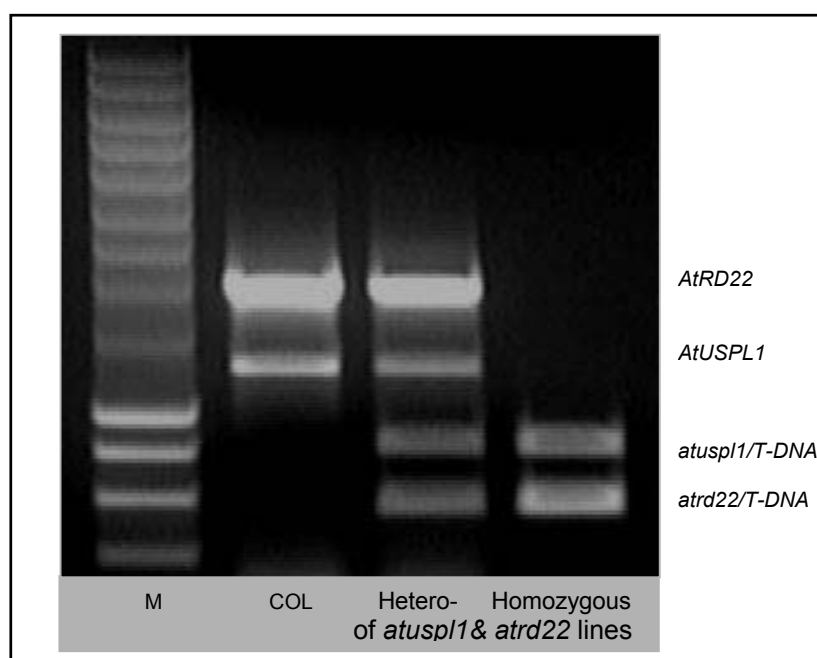
Figure 3.21. PCR analysis of *atuspl1/atrd22* double mutant using genomic DNA Genomic DNA was isolated from leaves of double mutant lines *(atuspl1/atrd22)* and Columbia wild type plants (COL). Using specific primers, *AtUSPL1* and *AtRD22* was amplified from COL and heterozygous line. Both wild type alleles are absent in the homozygous line. T-DNA insertions were detected in both mutant lines by T-DNA and gene specific primers. M- Smart ladder.

3.8.2. Phenotypic and physiological characterization of mutant lines

The double mutant and *atuspl1-C, atrd22* single mutants were analyzed for phenotypic and ultrastructural alterations. The mutants did not exhibit phenotypical alterations in comparison to their wild types (Columbia and Wassilewskaja). Further detailed histological analyzes were carried out. Ultrastructural investigations of mature seeds did not reveal obvious changes between mutant lines and wild type plants (Figure 3.22).
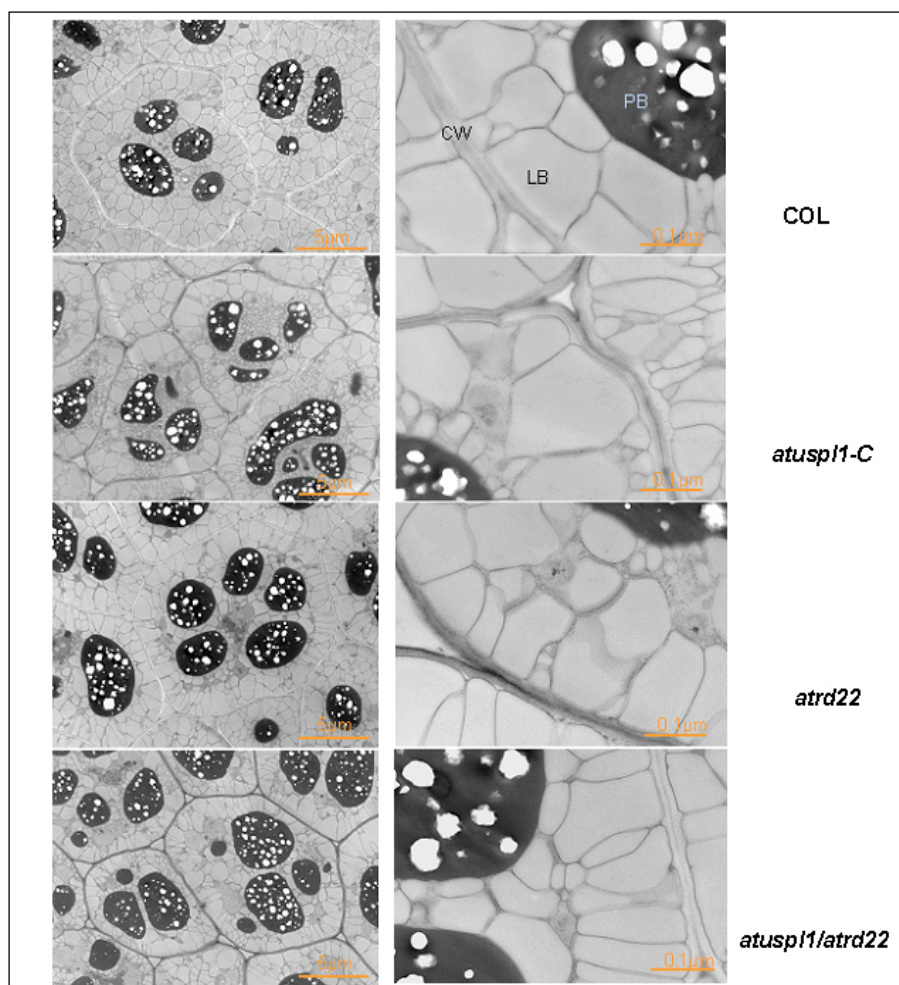
Figure 3.22. Ultrastructure of the storage parenchyma cells of loss-of-function mutant and Columbia wild type seeds

Comparison of mutant lines (*atuspl1-C, atrd22,* and *atuspl1/atrd22*) and Columbia wild type (COL) does not show obvious structural differences in lipid body (LB), protein storage vacuoles (PB) and cell walls (CW).
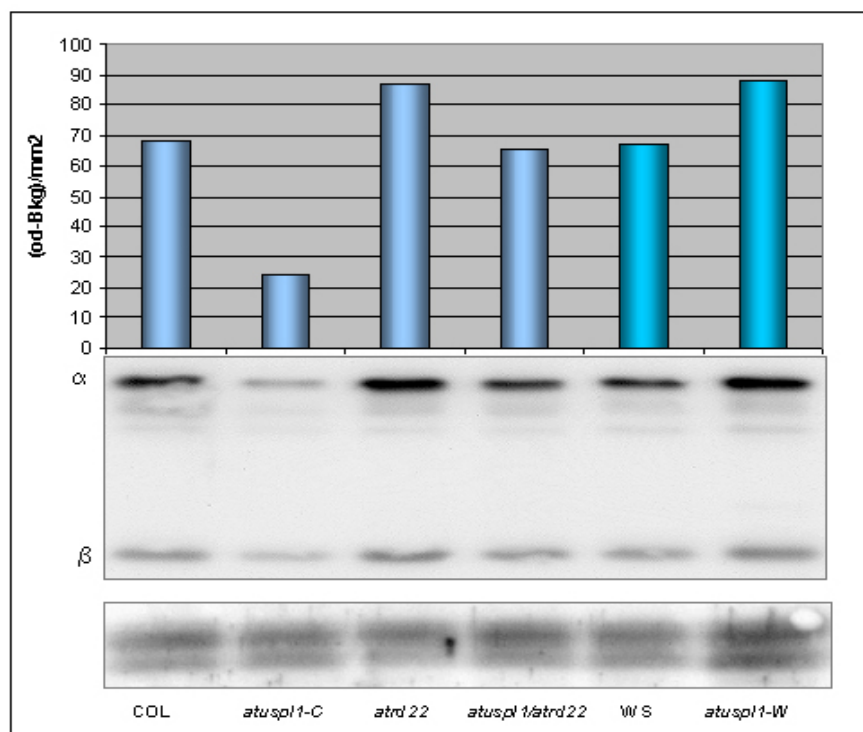
Figure 3.23. Western blot analysis of storage cruciferin and napin from mature single seeds of mutant and wild type plants (COL, WS)

Total protein was extracted from mature single seeds from individual mutants and wild type plants by using identical volumes of extraction buffer. The identical aliquots of protein extracts were separated under denaturing and reducing conditions on 12.5% polyacrylamide gel. After blotting, the storage proteins, cruciferin (**upper panel**) and napin (**lower panel**) were detected by anti-cruciferin or anti-napin antibody and anti-mouse IgG conjugated to HRP. Signals were generated by chemiluminescence. The α- and β-polypeptides of the globulin cruciferin are indicated on the left. Od- optical density; Bkg– background.

To define the levels of storage protein accumulation in loss-of-function mutant lines in comparison to wild type seeds, the cruciferin and napin content in mature single seeds were analyzed. As shown in Figure 3.23, the amounts of cruciferin of *atuspl1-C* and double mutant lines were slightly decreased in comparison to Columbia wild type controls, whereas it was increased in *atuspl1-W* in comparison to Wassilewskaja wild type. The contrary case was observed in

atrd22 mutant lines. The amount of cruciferin was increased in comparison to wild type. The cruciferin reduction is the highest in *AtUSPL1* overexpresion lines (Figure 3.12). The napin amounts showed no differences between wild type and mutant lines (Figure 3.23).

To characterize the fatty acid content of *atuspl*1 and *atrd22* mutant lines, seed extracts were analyzed by HPLC. Preliminary data showed that the total fatty acid content of the *atrd22* mutant was increased by approximately 5% in comparison to Columbia wild type controls. On the contrary, the total fatty acid content was slightly reduced by about 11% and 15% in the *atuspl1-C* and the *atuspl1-W* mutant seeds, respectively (Figure 3.24).
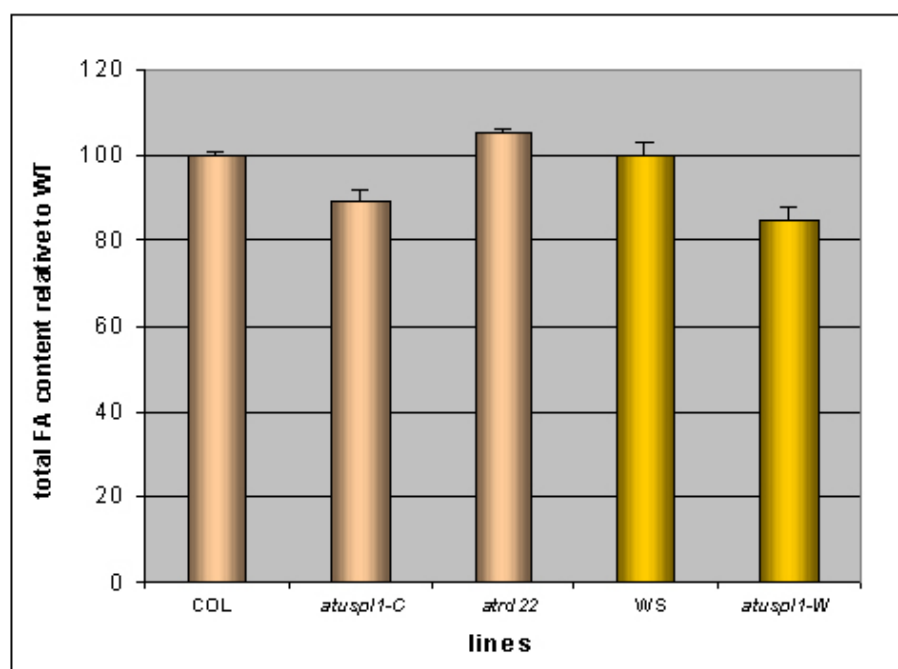
Figure 3.24. HPLC analysis of total fatty acids in mutant seeds and wild type (COL, WS)

Total fatty acid content was slightly reduced in *atuspl1* mutant lines, but slightly increased in the *atrd22* mutant.

Various environmental conditions were considered to detect a putative function of both genes. Therefore dehydration experiments were performed, in which loss-of-function mutant lines and wild type plants were grown on soil for 4

weeks, following growth in a chamber with 60% humidity for different times. After dehydration treatment for 1-5 days, plants were transferred back to the normal culture room. There is an obvious difference between wild type and the loss-of-function mutants after 2-3 days treatment with low humidity. Whereas the Columbia wild type turns brownish due to the accumulation of anthocyanin and retards growth, the three mutants obviously grow better and stay green. These observations suggest that both single mutants as well as the double mutant are more tolerant to low humidity treatment (Figure 3.25).
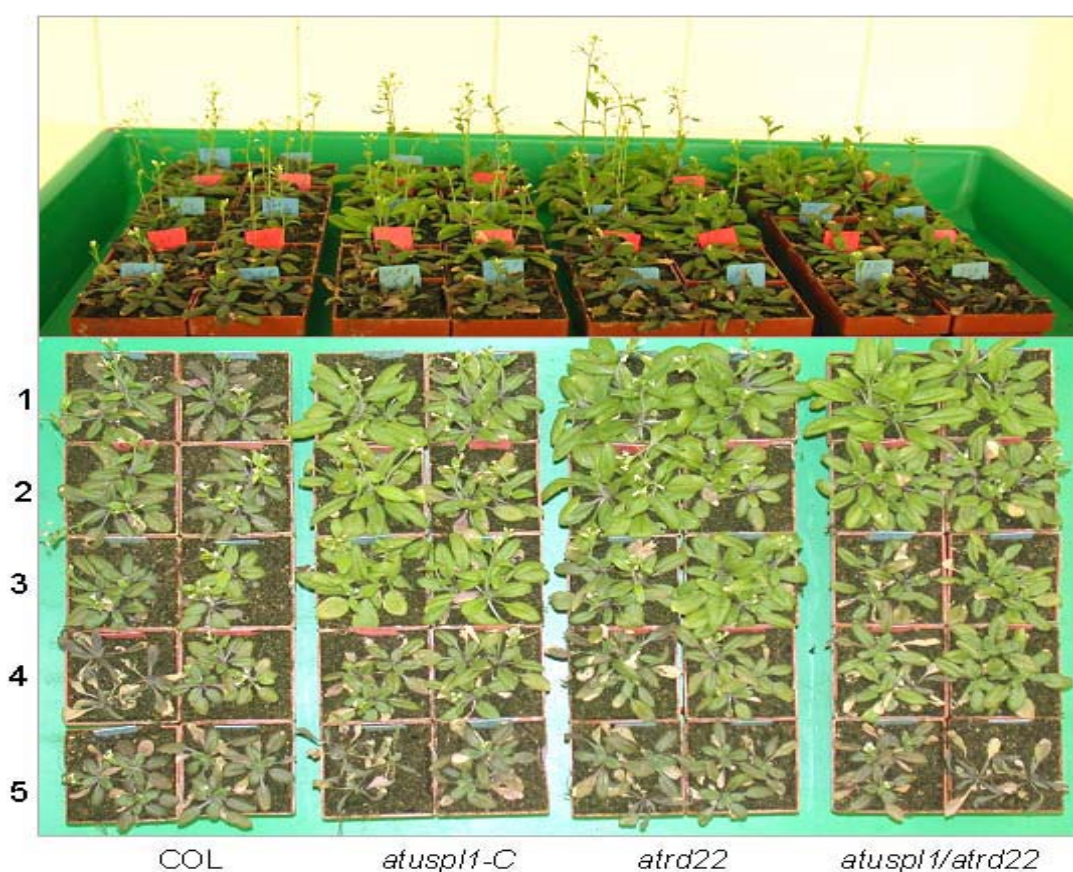
Figure 3.25. Influence of low humidity treatment on single and double loss-of-function mutants

Four-weeks old single and double mutant plants (*atuspl1-C, atrd22*, 2KO) and Columbia wild type plants (COL) were partially dehydrated under 60% humidity conditions for 1-5 days before they were returned to normal humidity conditions. It appears that the mutants exhibit a higher tolerance against drought stress. Time of stress treatment in days is indicated left.

In addition, the influences of various osmolytes have been tested. Two weeks-old loss-of-function mutant and wild type plants were grown on media with different concentrations of osmotically active substances. These included NaCl (0.1-0.3 mM), mannitol (0.2-0.4 mM) and polyethylene glycol (PEG) 8000 (5-20%). No obvious differences could be detected on NaCl and mannitol, however, differences were observed at 10% PEG, in which *atrd22* loss-of-function mutants obviously grew better than the wild type plants (Figure 3.26).
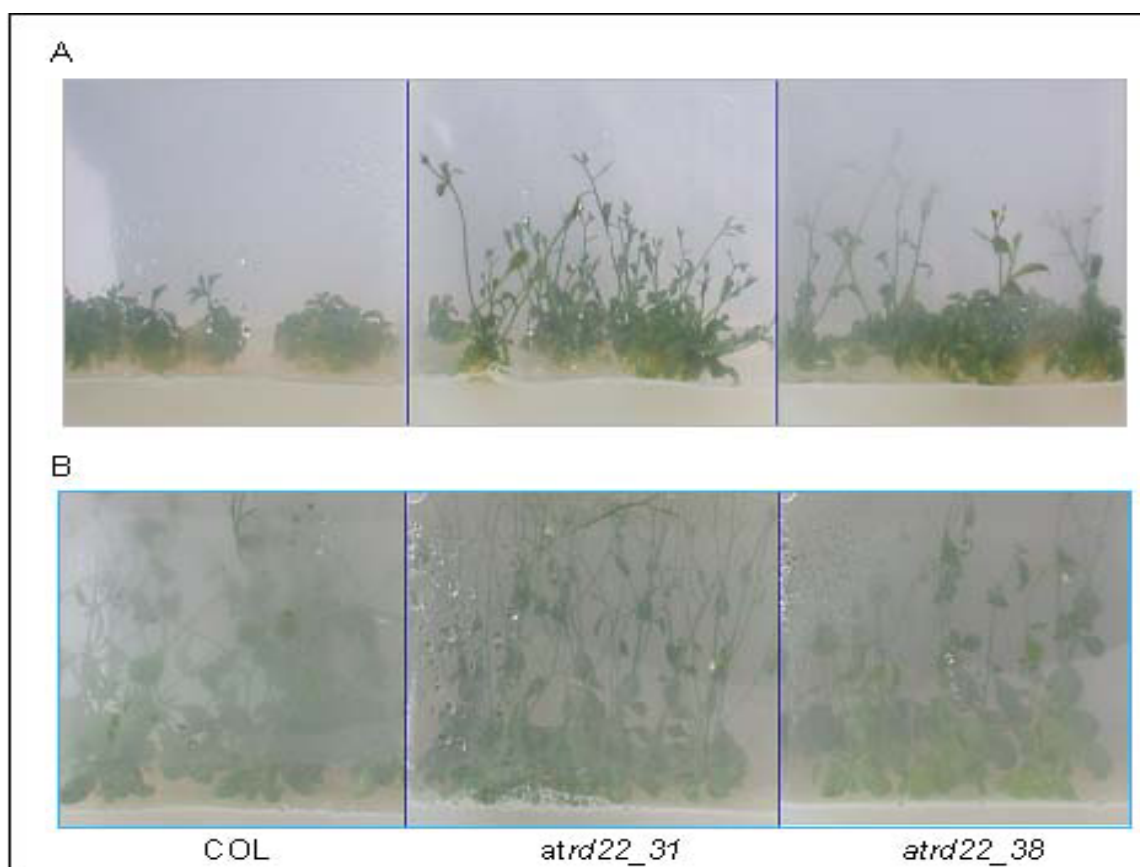
Figure 3.26. Developmental analysis of the *atrd22* loss-of-function mutant under drought stress

Two week-old plants were treated on medium containing 10% PEG 8000 (A). The development of homozygous *atrd*22 mutant lines (*atrd22_31*, *atrd22_38*) was compared with Columbia wild type (COL) and with plants growing under condition without PEG (B). The destruction of the *AtRD22* gene probably leads to a higher tolerance against water deficiency.

To analyze a putative effect of AtUSPL1 loss-of-function mutation on seed germination, mature and immature seeds of the *atuspl1-C* mutant were germinated on MS medium. The germination rate of mature seeds of the *atuspl1-C* mutant reached approximately 95% similar to that of Columbia wild type. As indicated in Section 3.7.7, the AtUSPL1 overexpression line exhibited a very low germination rate of about 7%. A clear difference between the *atuspl1-C* mutant and wild type was observed when examining the precocious germination of immature seeds at heart-torpedo stage. Under these conditions, germination rate of immature *atuspl1-C* seeds was about double (Figure 3.27), but most germinated plants developed abnormally.
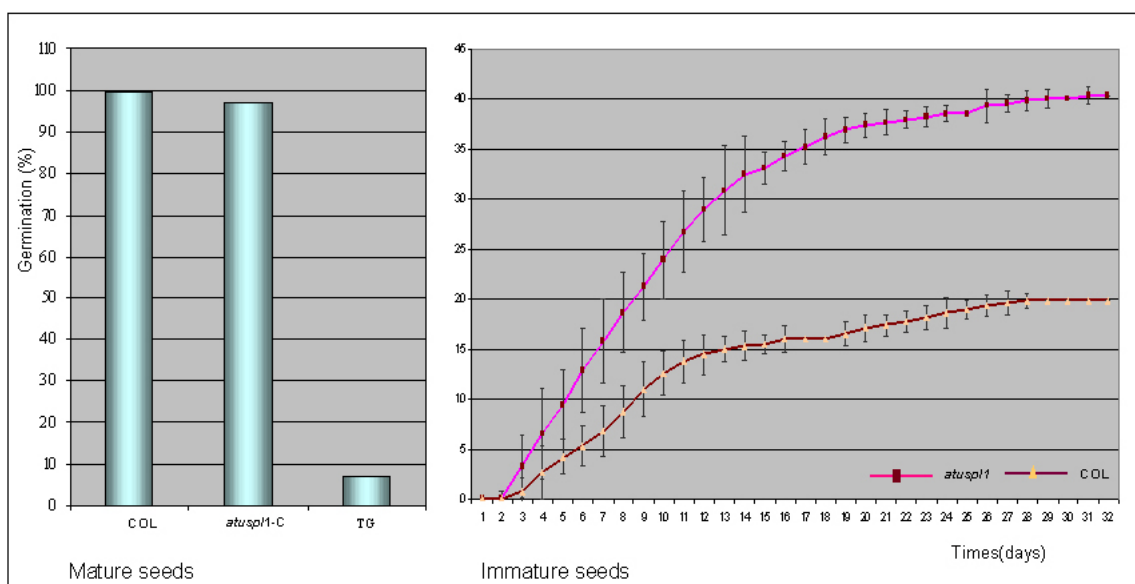
Figure 3.27. Germination analysis of mature and immature seeds of the *atuspl1* mutant and Columbia wild type (COL)

AtUSPL1 overexpression (TG) leads to inhibition of germination. The destruction of the *AtUSPL1* gene (*atuspl1)* leads to increased precocious germination.

3.9. Screening of putative protein interactors of AtUSPL1 and AtRD22

Structural conservation, especially of the CH-motifs, of the BURP domain invited the speculation that it might be involved in protein-protein interaction.

Therefore, *AtUSPL1* and *AtRD22* genes were used as bait to screen for putative interacting proteins using a Yeast Two Hybrid (Y2H) system. An *Arabidopsis* yeast library has been constructed using the BD Matchmaker Kit (Clonetech; Chu and Reidt, Research group of H. Puchta at Karlsruhe University).
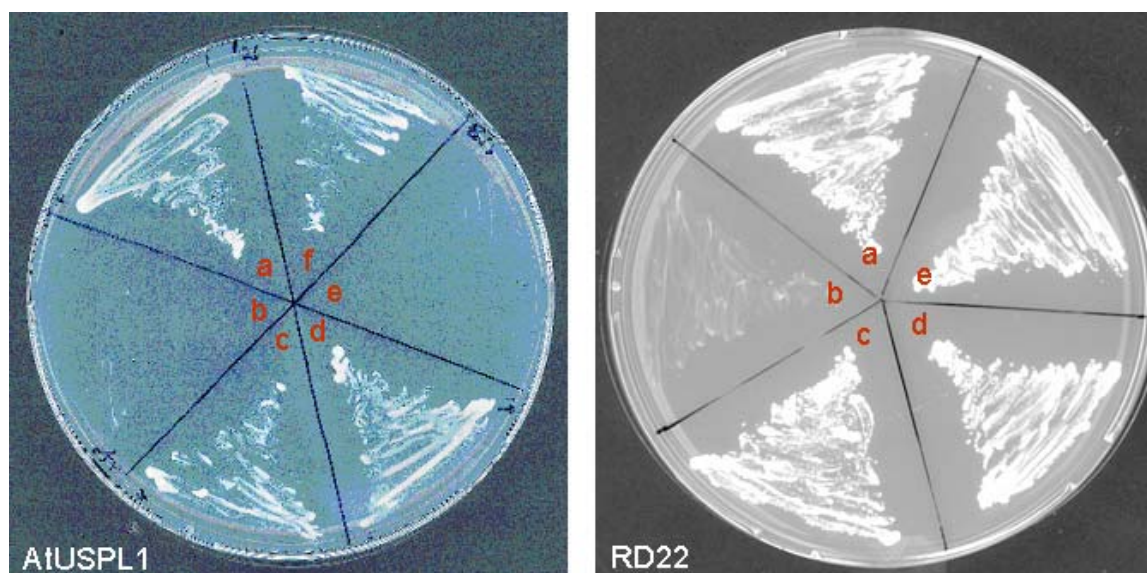
Figure 3.28. Yeast Two Hybrid screen for protein interactions based on growth in the absence of Trp, Leu, His and Ade

AtUSPL1: a) Positive control using RecT and p53 interactors (courtesy HH Chu); b) Negative control without bait protein; c) AtUSPL1 interacting with ribosomal protein L3 At1g43170; d and f) AtUSPL1 interacting with protein At5g26610; e) Negative control with AtUSPL1 only.

AtRD22: a) Positive control using RecT and p53 interactors (courtesy HH Chu); b) Negative control AtRD22 only; c) AtRD22 interacting with subtilase protein similar to subtilisin-type protease precursor At1g20150; d) AtRD22 interacting with aconitate hydratase protein At4g26970; e) AtRD22 interacting with tubulin beta-6 chain (TUB6) protein At5g12250.

Two candidate genes have been selected for sequencing in the AtUSPL1 screen. The first characterized putative interaction is the L3 ribosomal protein At1g43170. The results are shown in Figure 3.28. Further analysis is needed to examine whether the putative interaction of AtUSPL1 with this ribosomal protein might be connected to the distorted ribosome arrangement around the lipid

bodies in storage parenchyma cells of mature seeds (Figure 3.15). The second candidate gene is At5g26610 coding for an unknown protein. Again the observation needs further investigation, since the selected DNA fragment is located in the intron region of the currently annotated gene.

In the AtRD22 screen, three candidate genes have been identified. These are At1g20150 encoding a protein similarity to a subtilisin-type protease precursor, At4g26970 encoding a protein similarity to an aconitate hydratase protein and At5g12250 encoding a protein similarity to a tubulin-beta 6 chain (TUB6). All these candidates need further confirmation by independent approaches like for instance pull down experiments.

3.10. Processing of the AtUSPL1 and AtRD22 proteins and localization of the processing products

3.10.1. Localization of AtUSPL1 and AtRD22 polypeptides

There is a working hypothesis that the BURP domain could have an essential intracellular targeting-function for the N-terminal, partially repetitive peptides. Therefore, it is essential to analyze the processing and putative localization of the processing products of the BURP proteins. Two parallel experimental set-ups have been established: Firstly, antibodies specifically produced against the N-terminal or C-terminal parts of the gene product AtUSPL1 and AtRD22 were used to investigate their *in vivo* processing as well as their ultrastructural or intracellular distribution. Secondly, GFP-fusion constructs have been created to investigate *in vivo* processing and localization of AtUSPL1 and AtRD22 polypeptides by fluorescence microscopy.

The following polypeptide fragments of AtUSPL1 and AtRD22 were expressed in *E. coli:*

- The AtUSPL1 N-terminal polypeptide (P) fused to a GST tag purified under denaturing condition on Glutathione sepharose 4B.

- The AtUSPL1 C-terminal domain (BURP) as well as the complete AtUSPL1 (USPL1) fused to hexa-histidine tags and purified under denaturing conditions on Ni^{2+}-NTA agarose using immobilized-metal affinity chromatography.

- The AtRD22 N-terminal polypeptide (TXV) and the complete AtRD22 (RD22) fused to hexa-histidine tag purified under denaturing conditions on Ni^{2+}-NTA agarose using immobilized-metal affinity chromatography (Figure 3.29).
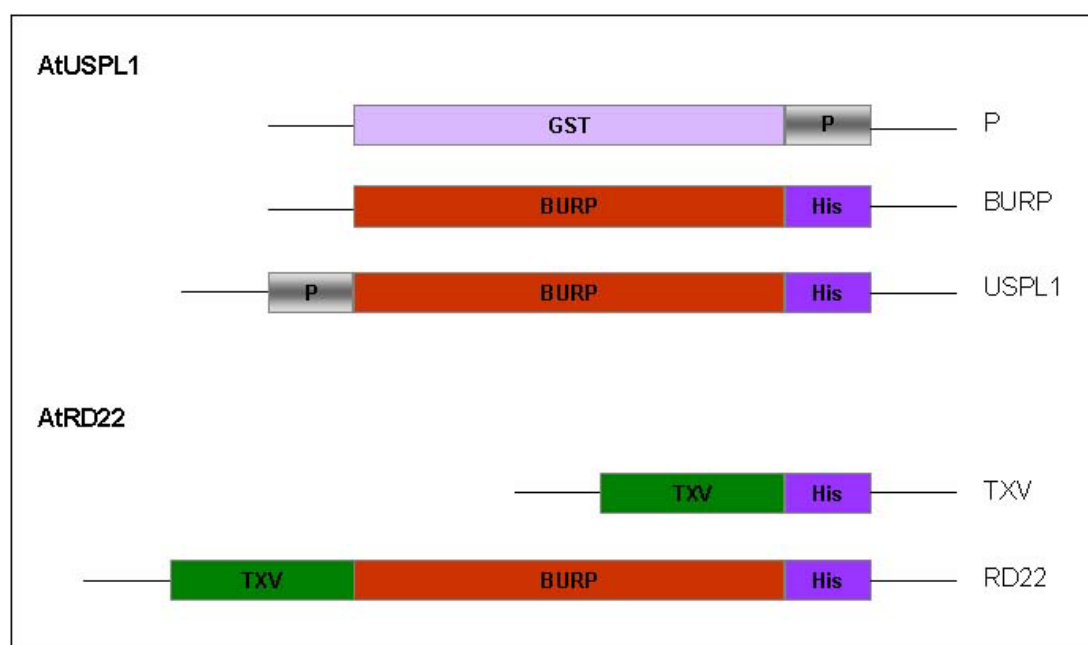
Figure 3.29. Constructs of AtUSPL1 and AtRD22 polypeptides to be expressed in *E.coli*

Grey- AtUSPL1 N-terminal polypeptide (P); Green- AtRD22 N-terminal polypeptide (TXV); red- BURP domains of AtUSPL1 and AtRD22 (BURP); violent- His tag; light violent- GST.

Sizes of expressed recombinant polypeptides P, BURP, USPL1, TXV, and RD22 are 29, 25, 30, 17, and 41.5 kD, respectively. All purified recombinant polypeptides were used to produce antisera in rabbits. Western blotting as shown in Figure 3.30 demonstrates the reactivity of the produced antisera. The positions

of the recombinant polypeptides on the blot were simultaneously verified by using the corresponding anti-His tag antibodies as primary antibody (result not shown).
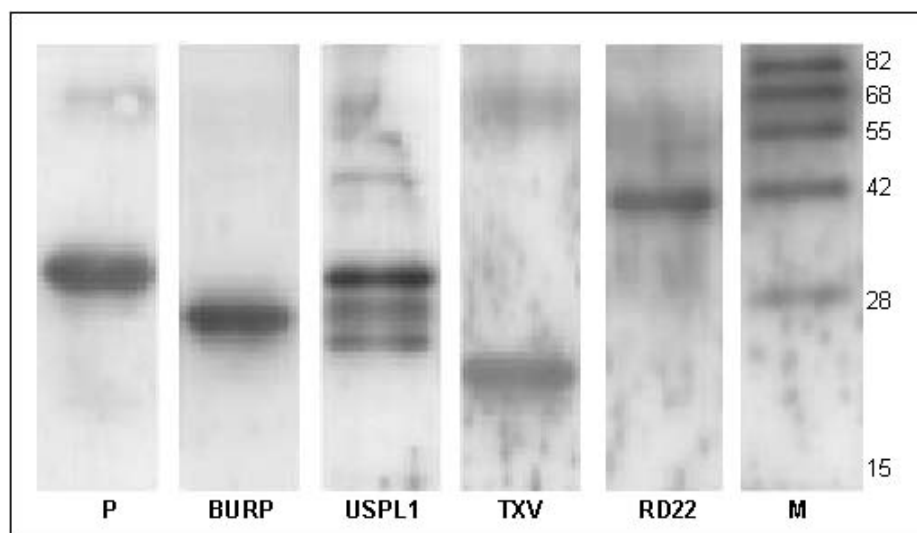
Figure 3.30. Detection of recombinant polypeptides in bacterial extracts by protein fragment specific antisera on Western blot

Total proteins were extracted from transgenic *E.coli* cells and were separated under denaturing and reducing conditions on 12.5% polyacrylamide gel. After blotting the recombinant polypeptides were incubated with their corresponding anti-protein fragment antisera followed by anti-rabbit IgG conjugated to HRP. Signals were generated by chemiluminescence. P, N-terminal polypeptide of AtUSPL1; BURP, C-terminal BURP domain of AtUSPL1; USPL1, AtUSPL1; TXV, N-terminal polypeptide of AtRD22; RD22, AtRD22; M- the molecular masses of the standard protein in kD are shown on the right.

These purified antibodies will use to recognize P, BURP, USPL1, TXV, and RD22 in *Arabidopsis* seed by immunolocalizations.

3.10.2. Expression of GFP fusion constructs

A parallel examination concerning co- and/or posttranslational processing of AtUSPL1 and AtRD22 *in vivo* was carried out by the expression of GFP fusion constructs. In order to synthesize GFP fusion proteins, the DNA fragments of *AtUSPL1* and *AtRD22* were cloned behind a CaMV-35S promoter

and were C-terminally fused with the gene encoding the green fluorescence protein (GFP) as shown in Figure 3.32.
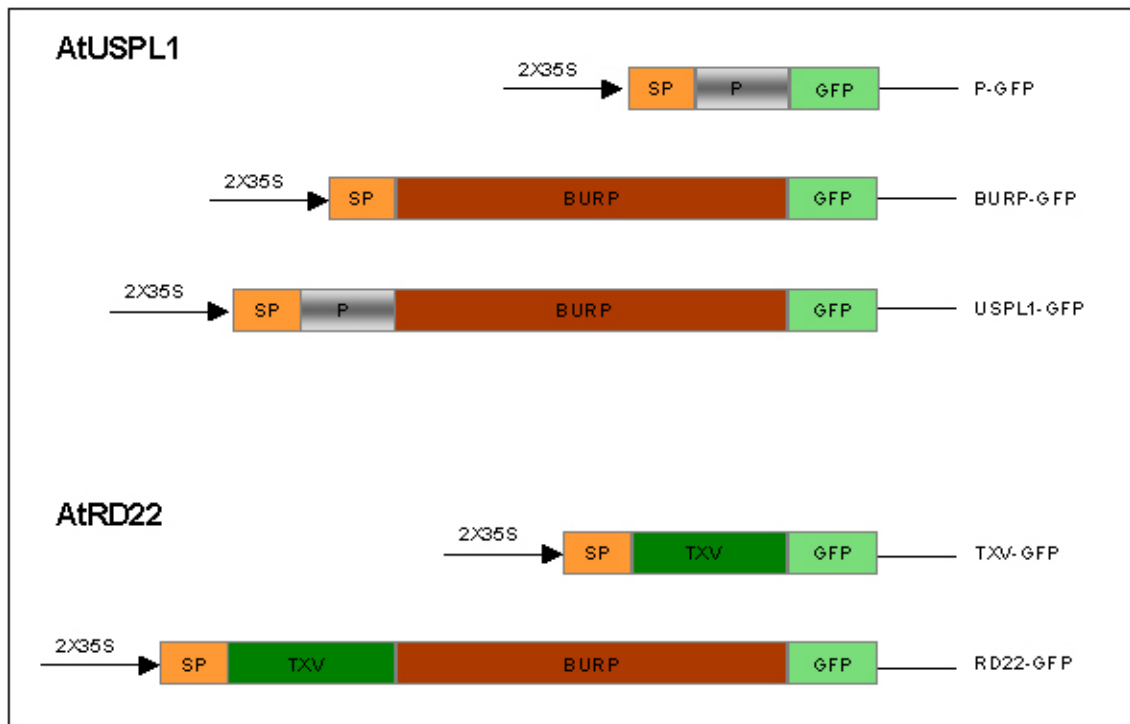
Figure 3.32. GFP-fusion constructs to analyze *in vivo* processing and ultrastructural distribution of the AtUSPL1 and AtRD22 proteins
Yellow- signal peptides (SP); grey- N-terminal fragments of AtUSPL1 (P); dark green- N-terminal fragments of AtRD22 (TXV); red- BURP domain of AtUSPL1 or AtRD22; light green- GFP.

All created constructs were transiently expressed in *Arabidopsis* protoplasts. The results of transient expression assay showed that all constructs expressed the GFP fusion protein, as recognized by GFP fluorecence in the region of the cytoplasm, putatively inside the ER. Figure 3.33 shows an example of the transient expression of the *USPL1-GFP* construct in *Arabidopsis* protoplast.

Furthermore, these constructs have been stably transformed into *Arabidopsis*.
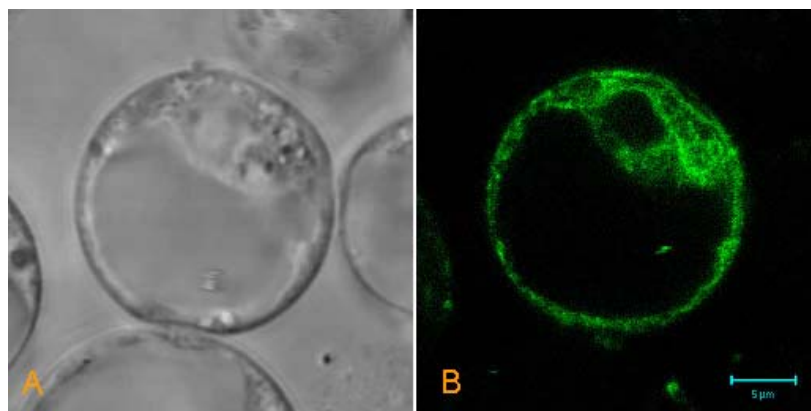
Figure 3.33. Transient expression of the *USPL1-GFP* construct in *Arabidopsis* protoplasts

A- light microscopy image of a transformed protoplast; B- the GFP fluorescence of an *Arabidopsis* protoplast. The fusion protein is found in the region of the cytoplasm, putatively inside the ER.
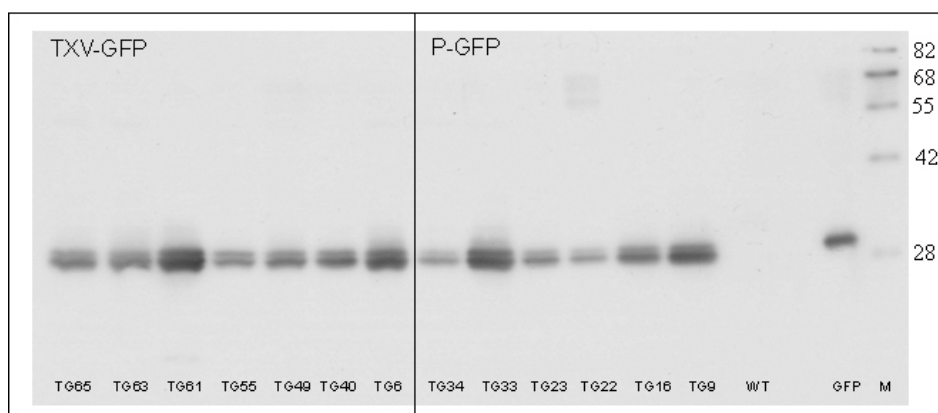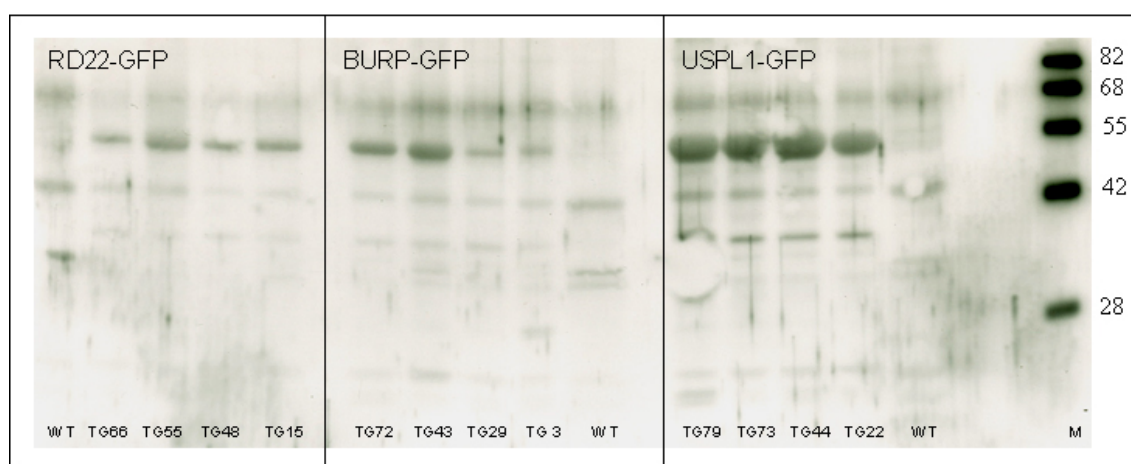

Figure 3.34. Detection of the fusion polypeptides TXV-GFP and P-GFP in transgenic lines by Western blotting

Total soluble protein was extracted from leaves of individual transgenic (TG) and Columbia wild type (WT) plants. Aliquots of the protein extracts were separated under denaturing and reducing conditions on 12.5% polyacrylamide gel. After blotting, the fusion proteins were detected by anti-GFP antibody followed by anti-mouse IgG conjugated to HRP. Signals were generated by chemiluminescence. GFP fusion proteins were detected having the same size than the GFP standard (~28kD); M- the molecular masses of the standard proteins in kD are shown on the right side.

GFP fusion protein overexpressing plant lines were initially identified by PCR and the expression of the fusion proteins was evaluated by Western blotting using antibodies against GFP as well as protein fragment specific antisera mentioned in 3.5.1. Based on the theoretically predicted amino acid sequence of the fusion products, the sizes were calculated to be 34, 47, 58, 54, and 71 kD for P-GFP, TXV-GFP, USPL1-GFP, BURP-GFP, and RD22-GFP, respectively. In practice however, in transgenic plants both the TXV-GFP and the P-GFP protein was found to have the same size as GFP control, 28 kD respectively (Figure 3.34). The size of the remaining three fusion proteins in transgenic plants was identical to the 55 kD BURP-GFP product (Figure 3.35). This suggests that the N-terminal parts are processed off the BURP domain.

Figure 3.35. Detection of the fusion polypeptides RD22-GFP, BURP-GFP and USPL1-GFP in transgenic lines by Western blotting

Total soluble protein was extracted from leaves of individual transgenic (TG) and Columbia wild type (WT) plants. Aliquots of the protein extracts were separated under denaturing and reducing conditions on 12.5% polyacrylamide gel. After blotting, the fusion proteins were detected by anti-GFP antibody followed by anti-mouse IgG conjugated to HRP. Signals were generated by chemiluminescence. GFP fusion proteins were detected having the same size as the BURP-GFP fusion protein (~55kD). M- the molecular masses of the standard proteins in kD are shown on the right side.

3.10.3. Localization of GFP fusion proteins in transgenic seeds

Based on Western blot, the lines with the highest amount of the fusion proteins were analyzed by fluorescence microscopy. The experiments showed that the GFP fluorescence was detectable in the petal veins of transformed plants with the *BURP-GFP* construct (Figure 3.36).
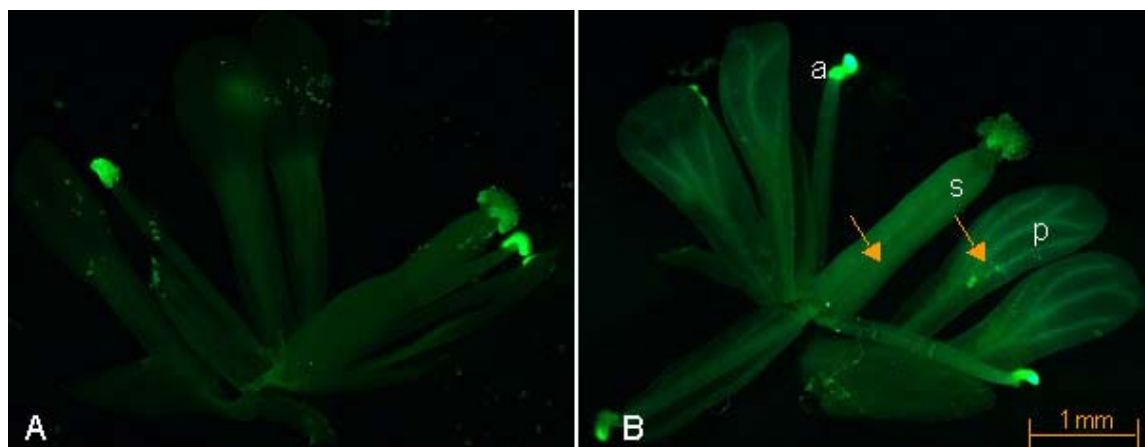
Figure 3.36. Fluorescence-microscopic image of flowers of the BURP-GFP transgenic line in comparison to the wild type

A- Columbia wild type flower; B- BURP-GFP transgenic flower; a- anther, s- silique; p- petal veins. GFP fluorescence expressed in the petal veins (yellow arrow) of the transgenic flowers.

In seeds of *P-GFP, TXV-GFP, USPL1-GFP, RD22-GFP* and *BURP-GFP* transgenic plants, GFP fluorescence signal was detected in the cell wall and the ER- containing cytoplasm of seed coat cells (Figure 3.37).

In connection with the results of the Western blot analysis (Figure 3.35) one can presume for BURP-GFP (E), AtRD22-GFP (C) and AtUSPL1-GFP (F) that the signal represents the localization of the BURP-domain-GFP fusions. In case of TXV-GFP (B) and P-GFP (D) the signal represents GFP alone as shown by the Western blotting (Figure 3.34). Therefore, no detection of P and TXV is possible by use of this method. Differences between the two N-terminal peptide fusions and the three BURP-domain containing fusions are interpreted as slight

differences in seed developmental stages, with younger stages (B and D) showing less red autofluorescence than older stages (C, E, and F).
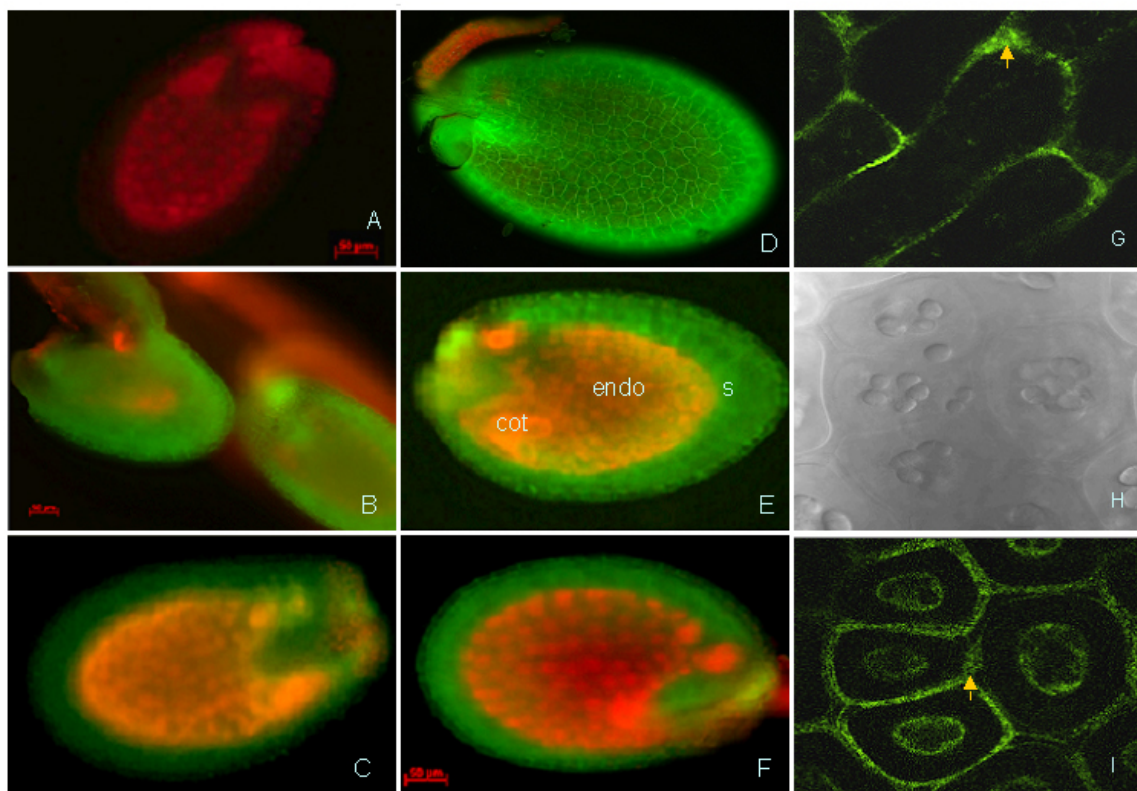
Figure 3.37. Fluorescence-microscopic image of seeds of *P-GFP, TXV-GFP, USPL1-GFP, BURP-GFP* and *RD22-GFP* transgenic lines compared to that of the wild type

A- Columbia wild type seed; B- *TXV-GFP* transgenic seed; C- *RD22-GFP* transgenic seed; D- *P-GFP* transgenic seed; E- *BURP-GFP* transgenic seed; F-I: *USPL1-GFP* transgenic seeds. G-I: CLSM pictures; G- GFP fluorescence signals in young seed; H- seed coat structure in old seed; I- GFP fluorescence signals in old seed; s- seed coat; cot- cotyledon; endo- endosperm. GFP fluorescence was detected in the cell wall and the cytoplast of seed coat cells (yellow arrow).

The GFP fluorescence signal was not detectable in the cotyledons of transgenic seeds due to strong fluorescence crosstalk of autofluorescence overlaying the GFP signals. Using anti-GFP antibody, the GFP fusion protein

was detected in the protein storage vacuoles of cotyledon cells of *USPL1-GFP* transgenic seed (Figure 3.38).
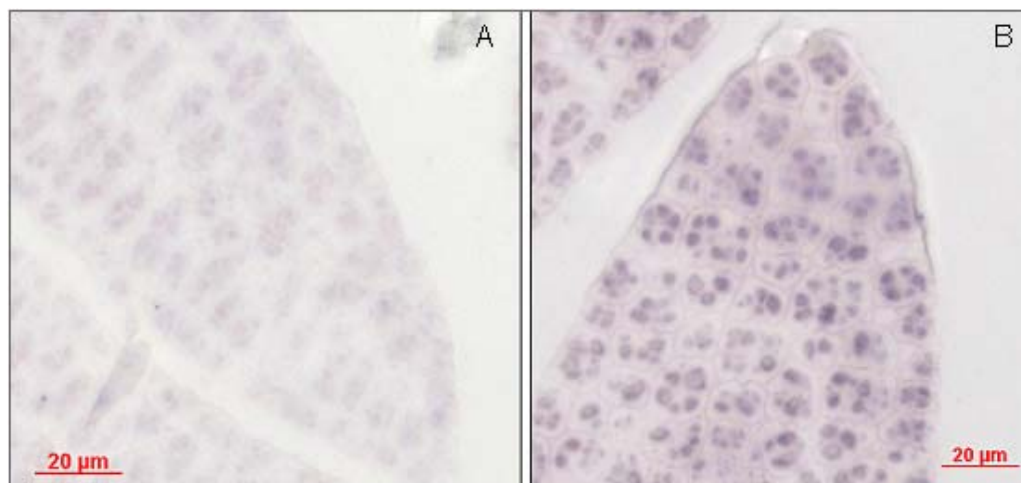
Figure 3.38. Localization of AtUSPL1-GFP fusion in *USPL1-GFP* transgenic seeds

Immunostaining using anti-GFP primary antibodies and anti-mouse IgG conjugated to alkaline phosphatase for signal regeneration. The GFP fusion protein was recognized in the protein storage vacuoles of cotyledon cells of transgenic seed (B). Under the same labeling conditions, the GFP fusion was not detected in Columbia wild type (A).

4. Discussion

To recognize the function of genes is a predominant aim of molecular biology. The *Arabidosis* community has estimated that only 20% of gene functions are known and the function of about 50% of the genes can be predicted only based on sequence comparisons to genes and proteins from other organisms like yeast. Obviously, there are about 30% of genes -some of them are strictly specific for higher plants- without any functional guidance from other organisms. The functional analysis of this kind of genes turned out to be difficult, but necessary to gain a comprehensive understanding of total genomes. State of the art approaches to analyze these genes include techniques like gain-and loss-of-function, localization of expression pattern at the transcriptional level, processing and localization of gene products, search for interacting proteins etc. This study aims for a functional characterization of the plant specific BURP protein family.

The BURP domain is restricted to plants

The BURP domain was firstly identified as part of the functionally unknown USP gene of *Vicia faba* and was originally designated as U-domain (Wohlfarth et al., 1998). In the meantime the domain has been identified in many proteins, including an auxin down regulated protein (ADR6); an aluminium induced protein (SALI3-2); a protein responsive to drought stress (RD22); the non-catalytic β-subunit of the polygalacturonase isoenzyme 1 of tomato (PG); an *in vitro*-embryogenesis specific protein of *Brassica napus* (BNM2); an apomixis-specific gene of *Panicum maximum* (ASG1); seed coat proteins of soybean (SCB1) as well as wheat and rice anther-specific proteins. Thus BURP domain proteins are

present in a wide variety of plant species. BURP protein encoding genes have even been detected in the moss *Physcomitrella patens* EST collection, demonstrating its ancient origin and structural and possibly functional conservation. The fact that BURP protein encoding genes could not be detected in other, non-plant genomes and EST collections suggests that they function in plant specific processes. This corresponds with the observation that most of them seem to be involved in embryogenesis and seed development, suggesting a function of the BURP domain important for cellular processes during seed development.

Structural features of BURP proteins

Based on primary structures, the BURP domain proteins are hydrophilic with the exception of a ~20 amino acid hydrophobic N-terminal region that most likely acts as a signal sequence for transport into the endoplasmic reticulum (Bassüner *et al*., 1988). Members are also strikingly similar in the C-terminal BURP domain. Within this domain there are several highly conserved amino acid positions including four cystein-histidine motifs as well as a threonine and a proline residue. The most obvious structural differences between members of the BURP family occur in the region between the signal sequence and the BURP domain. As shown in the following figure the N-terminal region is highly variable between different members of the family. This region can consist of a rather short non-repetitive sequence of about 40 amino acids (AtUSPL1, BNM2, VfUSP) or can include partially repetitive sequences with the most extensive repetition in the polygalacturonase isozyme (PG). An obvious common feature of several proteins is the abundance of aromatic amino acids like tyrosine (Y) and phenylalanine (F) as well as the frequent presence of proline (P). Other types of repeats include the TXV repeats found for instance in AtRD22 and CFC1. There is growing evidence that these N-terminal regions of BURP domain proteins become processed and cleaved off the BURP domain. With the exception of PG the function of the proposed peptides still remains to be elucidated for all other BURP domain proteins.

Among the five *Arabidopsis* BURP-protein encoding genes *AtUSPL1* is characterized by a short, non-repetitive amino acid sequence. This gene most closely resembles the *VfUSP* and *BNM2* genes of *Vicia faba* and *Brassica napus*, respectively. The short length of the N-terminal peptides of these three proteins might suggest that it has no special function and might be considered simply as a kind of stuffer fragment. Its processing off the BURP domain (as shown for AtUSPL1) therefore suggests that in this case the BURP domain might act alone. An alternative view is to consider also the short and non-repetitive N-terminal peptides as functional entities.

PG

*MHTKIHLPPCILLLLLFSLPSF*NVVVGGDGESGNPFTPKGYLIRYWKKQISNDLPKPWFLLNK ASPLNAAQYATYTKLVADQNALTTQLHTFCSSANLMCAPDLSPSLEKHSGDIH**FATY**SDKN**FT NY**GTNEPGIGVNT**FKNY**SEGENIPVNS**FRRY**GRGSPRDNK**FDNY**ASDGNVIDQS**FNSY**STSTA GGSGK**FTNY**AANANDPNLH**FTSY**SDQGTGGVQK**FTIY**SQEANAGDQY**FKSY**GKNGNGANGE**FV SY**GNDTNVIGST**FTNY**GQTANGGDQK**FTSY**GFNGNVPENH**FTNY**GAGGNGPSET**FNSY**RDQSN VGDDT**FTTY**VKDANGGEAN**FTNY**GQSFNEGTDV**FTTY**GKGGNDPHIN**FKTY**GVNNT**FKDY**VKD TAT**FSNY**HNKTSQVLASLMEVNGGKKVNNRWVEPGK<u>FFR</u>

BDC1

*MILVYLVLQVAIVVIRCDGAL*SPSERYWKAVLPNSPMPQAVKVLLPTPTGVGVD**AANGRIERH AA**GR**TIYAAAANGKIERHAA**AY**TIYAAAANGRIVRHAA**PII**LIYAAATNGRIER**ANVTGTQLH DDPTASL<u>FFL</u>

TaRAFTIN1a

*MARFLVALLATTLVAVQAG*GQLGHAAPATAEVFWRAVLPHSPLPDAVLRLLKQPAAGVELLTE ATSFVRDAEDRPP**FDYRDYS**RSPPDDEPSKSTGAASGARD**FDYDDYS**GGDKLRGAASGARD**FD YDDYS**GADKLRGATDEYKAPSSSLAGNGASMARGGKAETTTV<u>FFH</u>

OsRAFTIN1

*MARFLLLLVAVAAAAAVLSL*GDAAPSTAEVFWRAVLPESPLPDAFLRLLRPDTSFVVGKAEAA GGAARTGF**PFDY**TDYRGSDSPTTASGLDLAGDFGEPA**PFGYD**YSAQGEGGGGGAAAAAGEQVL AVDA**GFNY**DKYVGARKLRGGSSTAGGENDDE**PFGYD**YKAPSSGSGTAASTTARGVGTGATTTV <u>FFH</u>

AtRD22

*MAIRLPLICLLGSFMVVAI*AADLTPERYWSTALPNTPIPNSLHNLLTFDFTDEKS**TNVQVGKG GVNV**NTHKGKTGSG**TAVNVGKGGVRV**DTGKGKPGGG**THVSVGSGKGHGGGVAVH**TGKPGKR**TD VGVGKGGVTV**HTRHKGRPIYVGVKPGANPFVYNYAAKETQLHDDPNAAL<u>FFL</u>

CFC1

*MKVLSPILACLALAVVVSH*AALSPEQYWSYKLPNTPMPKAVKEILHPELMEEKS**TSVNVGG**GG VNVNTGKGKPGGD**THVNVGG**KGVGVNTGKPGGG**THVNVGD**PFNYLYAASETQIHEDPNVAL<u>FF</u>

OsAnther

*MASLVAIAIAMALMVVQPG*RQMTAFAARTSPAAAAEAFWRAAMPGAPMPDAIVELLHHEHGVA SAGGKANGGGDG**PPPP**MNFNYDDYRALPRSDAPSPDALNRVAAVQNADENGVSS**PPPP**PPTV<u>F FL</u>

```
ASG1
MAFVMGARIAGFMLLLVLMVEAGVYAGTGSQHVDANEWSMVPGVLGEMIQRGMTSRLEDYKYQ
PPAHDRSCTLLYEATARCSSTSMLFFN

SCB1
MEFHCLPIFLYLNLMLMTANAALTPRHYWETMLPRTPLPKAITELLSLESRSIFEYAGNDDQS
ESRSILGYAGYNQDEDDVSKHNIQIFNRLFFL

SALI3-2
MEFRCSVISFTILFSLALAGESHVHASLPEEDYWEAVWPNTPIPTALRDVLKPLPAGVEIDQL
PKQIDDTQYPKTFFY

ADR6
MALRCLVMSLSVLFTLGLARESHARDEDFWHAVWPNTPIPSSLRDLLKPGPASVEIDDHPMQI
EETQYPKTFFY

BNM2
MASLRFSVTFPALLSLLLLSLWVVEAYTSRKLISNNEQEGQNISHLFKDGEFEDPTMYMFFK

VfUSP
MEFAHLTVLSLFCLAFVGITATSSGEDYWQSIWPNTPLPKTFSDLSIPSGKTNSLPIKSEELK
QYSTLFFE

AtUSPL1
MASTFRLSISFLTLILFSLWVVEAHTSRKLISIKEKEGQDISHLLKDGEFDDPSLYMYFT
```

Figure 4.1. N-terminal amino acid sequences of all BURP domain proteins

Italic and underlined- signal peptide; red and underlined- first amino acids of the BURP domain; green- amino acid repeats. Note the partially repetitive structure and the frequent occurrence of tyrosine (Y), phenylalanine (F) and proline (P) residues. PG (U63373), the β-subunit of the polygalacturonase isoenzyme 1 of tomato; BDC1 (AJ843872), a protein from *Plantago*; TaRAPTIN1a (CAE02613), an anther protein of *Triticum aestivum*; OsRAPTIN1 (CAE02618); an anther protein of *Oryza sativa;* AtRD22 (D10703), a protein responsive to drought stress of *Arabidopsis thaliana*; CFC1 (AAL67991), protein from cotton fiber cells; OsAnther (AAC99621), an anther specific protein from *Oryza sativa*; ASG1 (no accession), an apomixis-specific gene of *Panicum maximum;* SCB1 (AAM03361), a soybean seed coats; SALI3-2 (T08896), an aluminium-induced–soybean protein; ADR6 (X69639), an auxin down regulated protein of *Glycine m*ax; BNM2 (AF049028), an *in vitro*- embryogenesis specific protein of *Brassica napus*; VfUSP (X13242), an abundant seed protein of *Vicia faba.*

AtRD22 of *Arabidopsis* is the next larger protein and its N-terminal region consists of four characteristics TXV repeats, with some similarity to that of the protein CFC1 described from cotton. Our results demonstrate that the TXV repeat region of AtRD22 is cleaved off the BURP domain, again suggesting that a function of this domain might be the delivery of the N-terminal peptides to their cellular target compartment, like the protein storage vacuole in case of AtUSPL1 and AtRD22 or the cell wall in case of the polygalacturonase isozyme (Zheng *et al.*, 1992; Watson *et al.*, 1994).

The remaining three BURP-protein encoding genes of *Arabidopsis* closely resemble the polygalacturonase β-subunit of tomato and are highly related to each other in size and sequence, being the result of a recent duplication event. Obviously they form a defined subgroup clearly distinct from AtUSPL1 and AtRD22. The characteristic feature of this subfamily is the occurrence of 21 FXXY repeats in front of the BURP domain. These repeats are known to be cleaved off the BURP domain, their function however still remains to be determined (Zheng *et al.*, 1992; Watson *et al.*, 1994).

The AtUSPL1 accumulates within the protein storage vacuole

The *AtUSPL1* expression pattern examined by Northern hybridization and the expression analysis of an *AtUSPL1* gene promoter-GUS fusion construct showed that the gene becomes more or less expressed in most tissues. Thus, unlike the structurally similar genes *VfUSP* and *BnBNM2* of *Vicia faba* and *Brassica napus*, respectively, (Bäumlein *et al.*, 1991; 1994; Boutilier *et al.*, 1994; Hattori *et al.*, 1998), *AtUSPL1* expression is not only restricted to seeds, but also most abundantly expressed in flowers and roots (Figure 3.6). The obvious difference in expression between the *Arabidopsis AtUSPL1* and the other two genes is unexpected, since these three genes share extensive structural similarities. Thus, the localization of both VfUSP and AtUSPL1 in compartments of the seed storage pathway like dense vesicles and protein storage vacuole suggests that both proteins share seed functions. However, the more ubiquitous

expression of the *AtUSPL1* gene also in non-seed organs like roots indicates that its function exceeds that of VfUSP.

During seed development, AtUSPL1 protein was only detected in late stages. The gene product was found in the protein storage vacuoles of the cotyledon cells of seeds. This rather distinct occurrence inside these organelles strongly suggests that the AtUSPL1 gene product might be essential for the proper accumulation and/or storage of the storage proteins like cruciferins and napins.

Ectopic expression of AtUSPL1 affects seed development

In order to provide a first insight into the function of the *AtUSPL1* gene, transgenic *Arabidopsis* plants were generated in which the *AtUSPL1* cDNA was ectopically overexpressed under the control of the constitutive CaMV*35S* promoter. This promoter was chosen to aim for a strong ubiquitous expression in the plant (Benfey *et al*., 1989). Depending on the expression level in different transgenic lines the phenotype varied from nearly unaffected to heavily distorted seed phenotypes. The strongest *AtUSPL1* overexpression resulted in shrunken seeds accompanied by a strongly reduced amount of storage cruciferin (Figure 3.10; 3.11; and 3.13). These distortions obviously also lead to the observed reduction of the germination rate and further growth retardation during the vegetative development (Figure 3.15). Whether this is strictly due to the reduced amount of storage products or just a consequence of the mechanical distortion of the shrunken seeds cannot be finally decided. All these results are well consistent with the above-mentioned observation that the AtUSPL1 protein is located within the protein storage vacuole and therefore functionally most likely is connected to protein storage processes.

Ectopic expression of AtUSPL1 affects ribosome attachment to lipid body membranes

In addition to the disturbed structure and function of the protein storage vacuole, the ectopic expression of *AtUSPL1* resulted in the distortion of the ribosome alignment along the membranes of the lipid bodies (Figure 3.13). This

observation suggests an additional function of AtUSPL1 in the attachment of ribosomes to the ER membranes. The ectopic overexpression of the *AtUSPL1* gene might lead to an excess of gene product, which could act as a competitor for ribosome binding along the lipid body membrane, thus resulting in the detachment of membrane bound ribosomes and the occurrence of free ribosomes in the cytosol. Remarkably, the search for putative AtUSPL1 interactors using a yeast two-hybrid approach resulted in the isolation of a ribosomal protein (see below), providing independent support for the suggested interpretation.

Based on election microscopic techniques using osmium staining (Figure 3.13), clear differences in staining were observed. Assuming that osmium binds stronger to the double bonds of unsaturated fatty acids, it is suggests that *AtUSPL1* overexpression resulted in changes of the fatty acid composition. The differences in lipid body staining are further supported by a biochemical approach. HPLC analysis of the fatty acids in seeds of wild type and overexpression lines confirmed a higher amount of unsaturated fatty acids when *AtUSPL1* is overexpressed (Figure 3.14).

Together these data suggest a function of the AtUSPL1 within the secretory pathway influencing synthesis and deposition of various storage compounds.

Ribosomal proteins might interact with AtUSPL1 protein

Based on the early assumption that BURP proteins and especially the BURP domain itself might function as part of protein complexes, yeast two hybrid screening (Fields and Song, 1989; Chien *et al*., 1991) has been applied to search for AtUSPL1 interacting proteins. A putative AtUSPL1 interactor was identified as the cytoplasmic ribosomal protein L3. This interactor is vital for the function of the ribosome and has been shown to participate in or even initiate the early steps of the ribosomal assembly, where it binds with high affinity to domain VI of the 23S rRNA (Nowotny and Nierhaus, 1982; Leffers *et al*., 1988; Uchiumi *et al*., 1999).

Thus, it might be that ectopically expressed AtUSPL1 affects ribosome assembly by its interaction with L3.

Obviously, this explanation underestimates the fact that AtUSPL1 and L3 should be localized in different cellular compartments preventing their *in vivo* interaction. The AtUSPL1 -after the removal of the signal peptide- is localized in the ER, whereas L3 contains a nucleolar targeting domain and enters the nucleolus to participate in ribosome formation.

Nevertheless, the Y2H-detected interaction between AtUSPL1 and the ribosomal L3 protein requires further confirmation for instance by pull down experiments.

Loss of AtUSPL1 function causes precocious seed germination

In strong contrast to the dramatic phenotypic effects of the *AtUSPL1* gene overexpression on seed development, a homozygous T-DNA insertion mutant did not exhibit severe phenotypic changes at the morphological, physiological and ultrastructural levels. The accumulation of storage cruciferin and fatty acids showed a moderately reduced level of accumulation. A clear-cut effect of the loss of *AtUSPL1* gene function is the facilitated capability of isolated immature seeds to germinate (Figure 3.29). The precocious germination rate of the mutant is strongly increased in comparison to wild type. Nevertheless these germinated seeds did not develop viable plants as has been reported for other mutants like *fus3* (Mueller and Heidecker, 1968), suggesting that AtUSPL1 might be involved in seed dormancy but in a different functional context. The rather mild phenotype of the loss-of-function mutant might be explained by the partial functional replacement of the AtUSPL1 gene product by the BURP domains of AtRD22 or even of the gene products encoded by the three genes belonging to the PG-like subfamily. To analyze this question further, an AtRD22 mutant as well as a double mutant with both genes, *AtUSPL1* and *AtRD22,* mutated have been generated.

Loss of AtRD22 function affects dehydration response

To address the possible functional redundancy between the genes *AtUSPL1* and *AtRD22*, a T-DNA insertion mutant of the *AtRD22* gene has been isolated. Although the T-DNA is inserted in the first intron, the mutation causes a strongly decreased amount of a truncated transcript but still seems to knock out the gene function. Again the single mutant does not have a strong phenotypic effect on plant growth. However the *atrd22* mutant might be more resistant to dehydration than the wild type (Figure 3.25 and 3.26). This observation is rather unexpected since the *AtRD22* gene is known to be induced under dehydration conditions, suggesting the loss of gene function mutant to exhibit reduced dehydration resistance. Thus, the observations indicate that the relation between AtRD22 gene expression and response to droughtness is more complex than anticipated. Nevertheless, it proposes an additional function for the BURP-domain protein AtRD22 related to water stress.

Although AtUSPL1 and AtRD22 differ greatly in their N-terminal regions, they share the BURP domain and this could be the basis for their mutual replacement. Therefore, a homozygous double mutant with both genes knocked out has been generated. Similar to the *atrd22* mutant also the double mutant exhibits a higher tolerance against desiccation combined with the accumulation of anthocyanins, known to be synthesized under unfavourable conditions. The double mutant represents a valuable tool to analyze both BURP-protein encoding genes further at the cellular and physiological level.

The BURP protein processing

As mention already above, all known BURP domain proteins exhibit a N-terminal signal peptide, suggesting the translocation of the corresponding protein into the ER. With the exception of the non-catalytic β-subunit of the polygalacturonase isozyme (PG) of tomato (Zheng *et al*., 1992; Watson *et al*., 1994) nothing is known about further posttranslational processing. The β-subunit of PG is thought to interact with structural components of the cell wall and with the PG2 catalytic subunit to immobilize or regulate the activity of the

polygalacturonase enzyme complex (Zheng *et al.*, 1992, 1994; Watson *et al.*, 1994). Peptide sequencing of the PG β-subunit has shown that the N-terminal hydrophobic signal peptide and the adjacent peptide, are cleaved off the mature protein (Zheng *et al.*, 1992). Sequentially, the PG polypeptide is cleaved again, between the repetitive region and the BURP domain. Thus the region consisting mostly of the repeated units becomes the functional polygalacturonase β-subunit, which is delivered by a proposed targeting function of the BURP domain. The fate and the possible further function of the cleaved C-terminal fragment of the PG β-subunit, which consists mainly of the BURP domain, remains unknown.
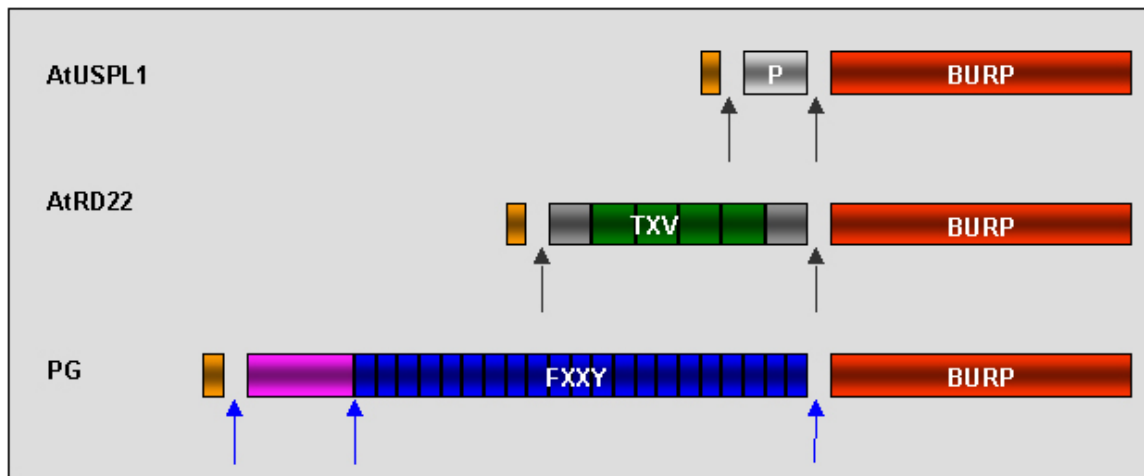
Figure 4.2. Scheme of the putative posttranslational processing steps of the
Arabidopsis BURP domain proteins

The blue upward arrows label the positions of processing sites derived from the non-catalytic β-subunit of polygalacturonase isozyme 1 of tomato generating the FXXY motif containing peptide (Zheng *et al.*, 1992; Watson *et al.*, 1994). The grey upward arrows label the corresponding predicted and partially demonstrated processing sites in AtRD22 and AtUSPL1.

To analyze whether the posttranslational processing scheme described for PG is also valid for other BURP proteins and to understand the functional relevance of the different parts of the *Arabidopsis* gene products we analyzed their posttranslational processing to address the following related questions:

a) Does the processing occur as it is shown in Figure 4.2? b) What are fate and function of the BURP domain and the partially repetitive peptides?

For this purpose, we used translational GFP fusions to the following polypeptides: the intact AtUSPL1 protein, the BURP-domain of AtUSPL1, the N-terminal region of AtUSPL1 as well as the intact AtRD22 protein, and the N-terminal repeat region of AtRD22. All three BURP domain-containing fusions (Figure 3. 35) show the same size in Western blot analysis. This demonstrates that both AtRD22 and AtUSPL1 become cleaved between the N-terminal region and the BURP domain as shown in Figure 4. 2. In connection with the results described for the non-catalytic β-subunit of polygalacturonase isozyme 1 of tomato (PG) (Zheng et al., 1992; Watson et al., 1994) this lead to the conclusion that the posttranslational processing between N-terminal region and BURP domain might be a common feature of this protein family. Thus, the basic function of the BURP domain could be to serve as a domain to target and deliver the N-terminal regions of the various proteins at their cellular site of action. Whereas this interpretation is well conceivable for the highly structured and repetitive TXV- and FXXY- containing N-terminal regions of AtRD22 and the three PG-like proteins, respectively, this might not be true for AtUSPL1 and the related proteins BnBNM2 and VfUSP. The N-terminal regions of these proteins are rather short and not structured as repeats. Thus, their N-terminal regions might simply serve as a structural stuffing fragment between signal peptide and BURP domain. Alternatively, also these short N-terminal regions itself could be of functional importance. Clearly, these questions require further investigations.

Cellular and subcellular localization of the BURP domain

The described experimental data demonstrate that the AtUSPL1 and AtRD22 proteins are localized within the protein storage vacuoles of seeds. In previous experiments performed by S. Hillmer the VfUSP protein has been localized in electron dense vesicles of *Vicia faba* cotyledons. These particles are considered to be different in structure from the final protein bodies, nowadays designated as protein storage vacuole (PSV). Both these observations suggest

that BURP proteins in total or the cleaved processing products might function in relation to storage proteins- for instance they might act as chaperons required for high density packing.

The β-subunit polygalacturonase, which is a BURP-domain protein in tomato, is believed to interact with both the structural components of the cell wall and the polygalacturonase catalytic subunit (Zheng *et al*., 1992). Other studies on the SCB1, an other BURP domain protein of the soybean, indicated that proteins recognized by antibodies generated against the SCB1 BURP domain may be covalently bound to the cell wall matrix of soybean seed coats. It is therefore possible that the BURP domain also represents a general motif for localization of proteins within the cell wall matrix of seed coat cells. The other structural domains associated with the BURP-domain proteins may specify other target sites for intermolecular interactions (Batchelor *et al*., 2002).

The data described above indicate that the BURP domain most likely functions as a new component of the cellular secretion pathway and might be involved in targeting or anchoring of the proteins to subcellular structures. Therefore, functional conclusions are expected from the cellular localization of the various BURP domain proteins. Our studies on GFP fluorescence showed that the BURP domain proteins AtUSPL1 and AtRD22 are localized in the cell wall of seed coat cells.

Thus AtUSPL1 and AtRD22 have been detected on two target sites of the secretory pathway: the cell wall and/or the PSV. The cell wall localization was exclusively restricted to cell types that do not contain PSV, suggesting PSV as the primary target of these gene products. The cell wall localization could be discussed as an artificial effect due to the strong overexpression of the GFP-fusion proteins in cells that do not offer the physiological target for the proteins.

These previous data and the here described results on BURP protein localization are schematically summarized in the following figure, which is based on a recent review by Hinz and Herman (2003).
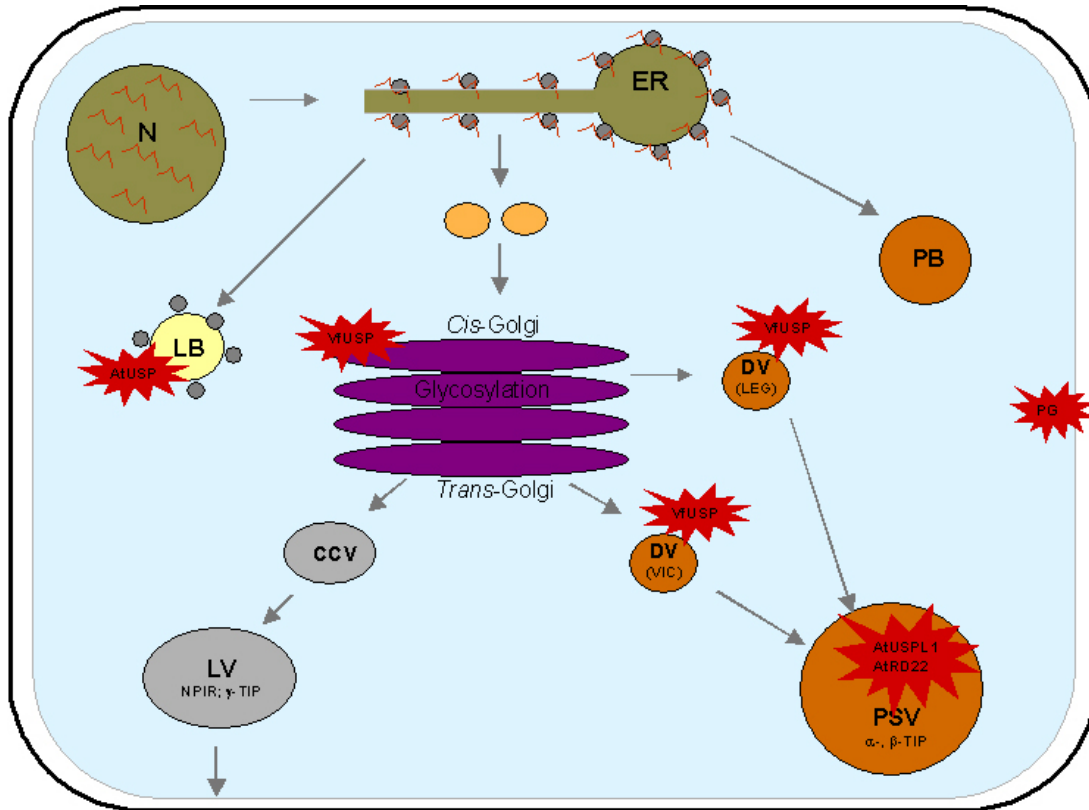
Figure 4.3. Scheme of the secretion pathway in seed storage parenchyma cells
Lipid bodies are considered to be direct derivatives of the ER. In USPL1-overexpresser lines the precise ribosome alignment on the lipid body membrane (Figure. 3.13) is distorted. Other ER derived vesicles are targeted to the *cis*-Golgi. Previous experiments (S. Hillmer, Heidelberg, pers. comm.) on *Vicia* and *Pisum* cotyledons have shown that VfUSP can be localized in the dictyosomes as well as in electron dense vesicles 200-400 nm in size, which are distinct from protein bodies. In *Arabidopsis* the AtUSPL1 and AtRD22 gene products (Figure 3.8) have been detected in the protein storage vacuoles and the non-catalytic β-subunit of polygalacturonase (Zheng *et al*., 1992) has been localized as associated with the cell wall. DV- dense vesicle; LEG- legume; VIC- Vicia; PSV- protein storage vacuole; TIP- tonoplast intrinsic proteins; CCV- clathrin-coated vesicle; LV- lytic vacuole; NPIR- signal consists of a larger, charged amino acid (N), a non-acidic amino acid (P), a large hydrophobic amino acid (I, L), and an amino acid with a large hydrophobic side chain (L, P).

Putative functional activity of the BURP domain

Of special interest is the analysis of the posttranslational processing of the BURP proteins and the suggested cleavage of short, partially repetitive peptides. Possibly, the proper function of BURP proteins is based on the cellular delivery of these peptide regions using the BURP domain as a targeting moiety.

The functional importance of peptides in plants is poorly understood. This is in clear contrast to animal systems, where precursor derived peptides play an important role as peptide hormones for cellular signaling processes. However, only few examples are known for plants. They include the systemins (Ryan and Pearce, 2003), peptides with putative signal functions for cell-cell interactions in the embryo sac (Corts *et al*., 2001), the phytochelatines (Clemens *et al*., 1999) as well as the phytosulfokines with possible functions in cell proliferation and cell differentiation (Lorbiecke and Sauter, 2002).

Preliminary working hypotheses concerning the function of the BURP domain proteins and the BURP domain itself assume:

a) The BURP domain proteins might form polyproteins as basis for functional protein networks.

b) The BURP domain might be involved in targeting or anchoring to/at defined subcelluar structures.

c) The eventually processed peptides might exhibit individual functions.

This is schematically illustrated in Figure 4.3.

In summary, although various state of the art techniques like gain-of-function and loss-of-function, cellular localization, characterization of interacting proteins etc. have been applied in the model plant *Arabidopsis* and several partially surprising conclusions could be drawn, a clear statement about the function of the BURP-protein family could still not be achieved. This again illustrates the current difficulties to assign a clear-cut function to strictly plant

specific functionally unknown gene products. All available data are best consistent with BURP protein functions related to cellular processes during seed storage compound synthesis and/or accumulation as well as seed germination
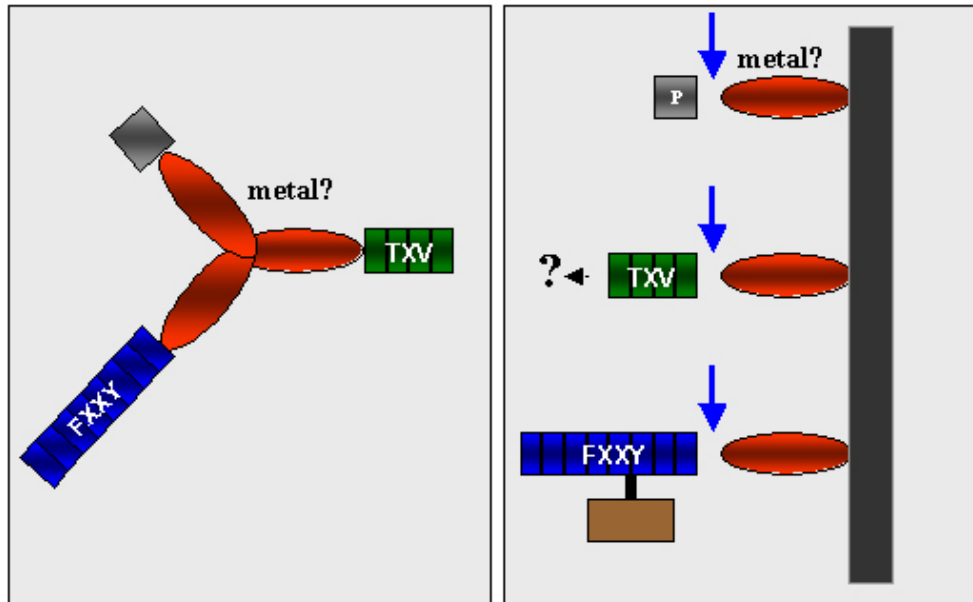
Figure 4.4. Working hypotheses concerning a similar putative function of the BURP domain as a common component of all members of the protein family

The BURP domains (red) might form polyproteins or protein networks (left panel). According to our currently favoured model the BURP domain might be involved in targeting or anchoring of various N-terminal peptides (P grey, TXV green, FXXY blue) to defined subcellular locations (grey) (right panel) but might also interact itself with other proteins. Posttranslational processing (blue arrows) release the N-terminal peptides for further interaction with other cellular proteins (brown) as suggested for the non-catalytical β-subunit of the tomato polygalacturonase. The highly conserved cysteine-histidine pattern of the BURP domain might be involved in these intermolecular interactions or could be a hint for metal binding.

Summary

BURP domain proteins comprise a novel, broadly distributed protein family. All these proteins exhibit a similar modular structure consisting of a signal peptide, a protein specific, more or less expanded, partially repeated middle part and a highly conserved C-terminal BURP domain with a characteristic cystein-histidin pattern. The strictly plant specific BURP domain protein family is growing rapidly with members found in many different plant species. The function of all previously described BURP domain proteins are poorly understood, but thought to be involved in cellular secretion pathway during embryogenesis, seed, fruit and root development.

The family was founded by the Unknown Seed Protein of *Vicia faba* (VfUSP). The *VfUSP* gene is transcribed into the most abundant message present in cotyledons during early seed development; however, the corresponding protein does not accumulate in cotyledons. The *VfUSP* gene promoter has been studied under several aspects, including its tissue specific and development dependent regulation by transcription factors, its application for the control of various transgenes for gene farming purposes and its use for the isolation of embryogenesis competent cell populations. In strong contrast to these abundant data concerning the gene promoter and its application, the function of the USP-gene product and specifically of its BURP-domain is poorly investigated.

The *Arabidopsis* genome contains five BURP-domain protein-encoding genes. Three of them exhibit similarity to the non-catalytic β-subunit of the polygalacturonase of tomato and form a distinct subgroup. A fourth gene, *AtRD22*, has been described as responsive to droughtness. The remaining gene, here designated as *AtUSPL1*, is most similar in size and sequence to the above

mentioned *VfUSP* gene and the *BNM2* gene of *Brassica napus* expressed during microspore-derived embryogenesis. This study aims to the functional characterization of the *Arabidopsis* BURP-protein family with special focus on AtUSPL1 and AtRD22.

Despite partially divergent and complex gene expression patterns, the gene products of both genes co-localise with storage proteins in the protein storage vacuoles of cotyledons. The ectopic expression of *AtUSPL1* leads to a shrunken seed phenotype and distortion in germination as well as to characteristic changes in the ultrastructure of lipid vesicles and protein bodies, which corresponds well with changes in storage fatty acid composition and the reduction of storage protein content. Together these results suggest a critical role of AtUSPL1 for cellular transport and storage processes.

To study loss-of-function conditions, two T-DNA insertion alleles of the *AtUSPL1* gene, a T-DNA-insertion line of the *AtRD22* gene as well as a corresponding double mutant have been isolated. In contrast to the deleterious effects of *AtUSPL1* over-expression, these gene destructions do not result in obvious seed phenotypes, although seeds show precocious germination and slight changes in seed protein content. Furthermore, mutant plants exhibit increased tolerance against dehydration conditions. A more extensive comparative analysis of these mutants is needed and in progress.

To investigate whether the highly conserved BURP domain might be involved in protein-protein interactions, both gene products have been used as bait to screen for putative interacting proteins in Arabidopsis yeast two hybrid libraries. Preliminary results suggest the interaction of AtUSPL1 with the ribosomal protein L3 and another functionally unknown protein. The interaction with a ribosomal protein needs to be proven further for instance by pull down experiments. Moreover, subtilisin- and tubulin-like proteins have been identified as putative interacting partners of AtRD22. All these results also need further confirmation.

To facilitate the cellular localization of the AtUSPL1 and AtRD22 gene products, specific antibodies have been generated and GFP fusion constructs have been transformed into *Arabidospsis.* The AtUSPL1-GFP fusion protein was localised in storage protein vacuoles of seed cotyledon cells. The analysis of the GFP fusions revealed that both primary translation products become posttranslationally processed. After the removal of the signal peptide, a second processing step detaches the N-terminal peptide regions from the BURP-domain. Thus, the BURP domain might act to deliver these peptides to their cellular target place within the secretory pathway.

Although important tools have been developed and applied, no comprehensive functional characterization of the BURP domain proteins could be achieved yet. Nevertheless, the results suggest a function of the AtUSPL1 and AtRD22 gene product and specifically of the BURP domain related to storage compound synthesis, transport and deposition.

Zusammenfassung

BURP-Domänen-Proteine umfassen eine neue, weit verbreitete Proteinfamilie. Alle BURP-Proteine weisen eine ähnliche modulare Struktur auf und bestehen aus einem Signalpeptid, einer Protein-spezifischen, mehr oder weniger ausgeprägten, teilweise repetitiven mittleren Region sowie einer hoch konservierten C-terminalen BURP-Domäne mit einem charakteristischen Cystein-Histidin-Muster. Die Anzahl der ausschließlich in Pflanzen gefundenen BURP-Proteine wächst stetig. Die Funktion aller bisher beschriebenen BURP-Proteine ist weitgehend unverstanden, wird jedoch mit der zellulären Sekretion während der Embryogenese sowie während der Samen-, Frucht- und Wurzelentwicklung in Verbindung gebracht.

Die BURP-Familie wurde durch das Unbekannte Samenprotein von *Vicia faba* (VfUSP) begründet. Das *VfUSP*-Transkript ist der häufigste *messenger* in den Kotyledonen früher Entwicklungsstadien der Samen, ohne dass ein entsprechend abundantes Protein akkumuliert. Der *VfUSP*-Genpromotor wurde unter verschiedenen Aspekten untersucht. Dies betrifft insbesondere seine gewebespezifische und entwicklungsabhängige Regulation durch Transkriptionsfaktoren, seine Anwendung für die Kontrolle der Expression verschiedener Transgene im Rahmen von *gene farming*-Experimenten sowie seine Nutzung für die Isolierung Embryogenese-kompetenter Zellpopulationen. Im Gegensatz zu den umfangreichen Arbeiten am Genpromotor, wurde die Funktion des *VfUSP*-Genproduktes nur unzureichend untersucht.

Im *Arabidopsis*-Genom werden fünf für BURP-Domänen-Proteine kodierende Gene gefunden. Drei davon zeigen ausgeprägte Sequenzähnlichkeit mit der nichtkatalytischen β-Untereinheit der Polygalakturonase aus Tomate und bilden eine klar definierte Subfamilie. Ein viertes Gen, *AtRD22*, wurde bei der Suche nach Trocknungsstress-induzierten Genen gefunden. Das fünfte, im Folgenden als *AtUSPL1* bezeichnete Gen besitzt große Ähnlichkeit in Sequenz und Größe zu dem oben erwähnten *VfUSP*-Gen sowie zu dem Gen *BMN2* aus *Brassica napus*, welches spezifisch während der Mikrosporen-Embryogenese exprimiert wird. Die hier beschriebenen Arbeiten konzentrieren sich auf die funktionelle Analyse der *Arabidopsis* BURP-Protein-Familie unter besonderer Berücksichtigung der Gene *AtUSPL1* und *AtRD22*.

Trotz der teilweise unterschiedlichen, komplexen Expressionsmuster beider Gene werden die Genprodukte zusammen mit den Speicherproteinen in den Protein-Speichervakuolen der Kotyledonen gefunden. Die ektopische Expression von *AtUSPL1* führt zu geschrumpften Samen und Störungen in der Keimung sowie zu charakteristischen Veränderungen in der Ultrastruktur der Lipidvesikel und Speichervakuolen. Dies geht einher mit Veränderungen in der Zusammensetzung der Fettsäuren und der Reduktion des Gehaltes an Speicherproteinen. Zusammen genommen belegen diese Befunde eine wesentliche Rolle von AtUSPL1 für zelluläre Transport- und Speicherungsprozesse.

Für das Studium der *loss-of-function*-Situation wurden zwei T-DNA-Insertionsallele für das Gen *AtUSPL1*, eine T-DNA-Insertionslinie für das Gen *AtRD22* sowie eine entsprechende Doppelmutante isoliert. Im Gegensatz zu den starken Effekten der *AtUSPL1*-Überexpression zeigen die Insertionsmutanten keine offensichtlichen Samen-Phänotypen, wenngleich schwache Effekte wie vorzeitige Keimung und geringe Unterschiede im Gehalt an Samenproteinen gefunden werden. Darüber hinaus zeigen die Mutanten eine erhöhte Toleranz gegen Austrocknung. Offensichtlich ist eine detaillierte vergleichende Analyse der Mutanten erforderlich. Entsprechende Versuche sind in Arbeit.

Um die Beteiligung der hoch konservierten BURP-Domäne für Protein-Protein-Interaktionen zu untersuchen, wurden beide Genprodukte für Yeast Two Hybrid-Versuche benutzt. Vorläufige Ergebnisse belegen die Interaktion von AtUSPL1 mit dem ribosomalen Protein L3 sowie einem weiteren, funktionell unbekannten Protein. Für AtRD22 wurden Subtilisin- und Tubulin-ähnliche Proteine als putative Interaktionspartner gefunden. Alle diese Yeast Two Hybrid-Ergebnisse erfordern eine Überprüfung beispielsweise durch pull down-Experimente.

Für die zelluläre Lokalisation der AtUSPL1 und AtRD22 Genprodukte wurden spezifische Antikörper gegen die Gesamtproteine sowie gegen Teile davon erzeugt und eingesetzt. Schließlich wurden GFP-Fusionskonstrukte für beide Proteine in *Arabidopsis* transformiert. Das AtUSPL1-GFP-Fusionsprotein wurde in den Protein-Speichervakuolen lokalisiert. Die Befunde zeigen, dass beide primären Translationsprodukte posttranslational an zumindest zwei Positionen prozessiert werden. Nach der Abspaltung des Signalpeptides erfolgt eine zweite Prozessierung, die zur Abspaltung der N-terminalen Peptidregion von der BURP-Domäne führt. Diese Befunde unterstützen die Vermutung, dass die BURP-Domäne am *targeting* der N-terminalen Peptidregion beteiligt ist.

Wenn gleich eine Reihe wichtiger experimenteller Voraussetzungen geschaffen und genutzt wurden, ist eine umfassende Aussage zur Funktion der BURP Proteine zur Zeit nicht möglich. Dennoch legen die Ergebnisse eine Funktion der beiden Proteine AtUSPL1 und AtRD22 sowie deren BURP-Domäne für Synthese, zellulärem Transport und Akkumulation von Samen-Speicherkomponenten nahe.

References

Adams A, Gottschling DE, Kaiser CA, and Stearns T (1997). Methords in Yeast genetics. *Cold Spring Harbor Laboratory.*

Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ (1990). Basic local alignment search tool. *J Mol. Biol*. **215**: 403-440.

Bassüner R, Bäumlein H, Huth A, Jung R, Wobus U, Rapoport TA, Saalbach G, and Müntz K (1988). Abundant embryonic mRNA in field bean (*Vicia faba* L.) codes for a new class of seed proteins: cDNA cloning and characterization of the primary translation product. *Plant Mol. Biol.* **11**: 321-334.

Batchelor AK, Boutilier K, Miller SS, Hattori J, Bowman LA, Hu M, Lantin S, Johnson DA and Miki BLA (2002). SCB1, a BURP-domain protein gene, from developing soybean seed coats. *Planta* **215**: 523-532.

Bäumlein H, Boerjan W, Nagy I, Bassüner R, Van Montagu M, Inzé D, and Wobus U (1991). A novel seed protein gene from *Vicia faba* is developmentally regulated in transgenic tobacco and *Arabidopsis* plants. *Mol. Gen. Genet*. **225**: 459-467.

Bäumlein H, Miséra S, Luerßen H, Kölle K, Horstmann C, Wobus U, and Müller AJ (1994). The *FUS3* gene of *Arabidopsis thaliana* is a regulator of gene expression during late embryogenesis. *Plant J*. **6**: 379-387.

Bechthold N, Ellis J, and Pelletier G (1993). *In planta Agrobacterium* mediated transfer by infiltration of adult *Arabidopsis thaliana* plants. *CR Academy of Sciences Paris* **316**, 1194-1199.

Beisson F, Koo A, Ruuska S, Schwender J, Pollard M, Thelen J, and Paddock T, Salas J, Savage L, Milcamps A, et al (2003) Arabidopsis genes involved in acyl lipid metabolism. A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a Web-based database. *Plant Physiol*. **132:** 681–697.

Benfey PN, Ren L, and Chua N-H (1989). The CaMV 35S enhancer cotains at least two domains which can confer different developmental and tissue-specific expression patterns. *EMBO J*. **8**: 2195-2202.

Bewley JD and Black M (1994). Seeds physiology of development and germination. *Plenum press*, New York.

Bhattacharyya MK, Smith AM, Ellis THL, Hedley C, and Martin C (1990). The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell* **60**: 115-122

Borisjuk N, Sitailo L, Adler K, Malysheva L, Tewes A, Borisjuk L, and Manteuffel R (1998). Calreticulin expression in plant cells: developmental regulation, tissue specificity and intracellular distribution. *Planta* **206**: 504-514

Boutilier K, Gines MH, DeMoor JM, Huang B, Baszcynski CL, Iyer VN, and Miki BL (1994). Expression of the BnNAP subfamily of napin genes coincides with the induction of *Brassica* microspore embryogenesis. *Plant Mol. Biol.* **26**: 1711-1723.

Bradford MM (1976). Rapid and quantitative method for quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**: 248-252.

Chen L, Miyazaki C, Kojima A, Saito A, and Adachi A (1999). Isolation and characterisation of a gene expressed during early embryo sac development in apomictic guinea grass (*Panicum maximum*). *Plant Physiol.* **154**: 55-62.

Chesnokov Y, Meister A, and Manteuffel R (2002). A chimeric green fluorescent protein gene as an embryonic marker in transgenic cell culture of *Nicotiana plumbaginifolia* Viv. *Plant Sci.* **162**: 59-77.

Chien CT, Bartel PL, Sternglanz R, and Fields S (1991). The two-hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Nat .Acad. Sci. USA* **88**: 9578–9582.

Chomczynski P and Sacchi N (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analyt. Biochem.* **162**: 156-159.

Church GM and Gilbert W (1984). Genomic Sequencing. *Proc. Nat. Acad. Sci. USA* **81**: 1991-1995.

Clemens S, Kim EJ, Neumann D, and Schroeder J (1999). Tolerance to toxic metals by a gene family of phytochelatin synthases from plants and yeast. *EMBO J.* **18**:3325-3333.

Cohen SN, Chang ACY, and Hsu L (1972). Nonchromosomal antibiotic resistance in bacteria- genetic transformation of *Escherichia coli* by R-factor DNA. *Proc. Nat. Acad. Sci. USA* **69**: 2110-2114.

Cordts S, Bantin J, Wittich PE, Kranz E, Lorz H, and Dresselhaus T (2001). ZmES genes encode peptides with structural homology to defensins and are specifically expressed in the female gametophyte of maize. *Plant J.* **25**:103-114.

Curaba J, Moritz T, Blervaque R, Parcy F, Raz V, Herzog M, and Vachon G (2004). AtGA3ox3, a key gene responsible for bioactive Gibberellin biosynthesis, is regulated during embryogenesis by LEC2 and FUS3. *Plant Physiol* **136**: 3660-3669.

Curtis MD and Grossniklaus U (2003). A Gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol*. **133**: 462-469.

Czihal A, Conrad B, Buchner P, Brevis R, Farouk AA, Manteuffel R, Adler K, Wobus U, Hofemeister J, and Bäumlein H (1999). Gene farming in plants: Expression of a heatstable *Bacillus* amylase in transgenic legume seeds. *Plant Physiol.* **155**: 183-189.

Datta N, LaFayette PR, Kroner PA, Nagao RT, and Key JL (1993). Isolation and characterisation of three families of auxin down-regulated cDNA clones. *Plant Mol. Biol.* **21**: 859-869.

Deblaere R, Bitenier B, Degreve H, Deboeck F, Schell J, Vanmontagu M, and Leemans J (1985). Efficient octopine Ti plasmid-derived vectors for *agrobacterium*-mediated gene-transfer to plants. *Nucleic Acids Res.* **13**: 4777-4788.

Di Sansebastiano GP, Paris N, Marc-Martin S, and Neuhaus JM (2001). Regeneration of a lytic central vacuole and of neutral peripheral vacuoles can be visualized by green fluorescent proteins targeted to either type of vacuoles. *Plant Physiol*. **126**: 78-86.

Drews DG and Goldberg RB (1989). Genetic control of flower development. *Trends Genetics* **5**: 256-261.

Edwards K, Johnstone C, and Thompson C (1991). A simple and rapid method for the preparation of genomic plant DNA for PCR analysis. *Nucleic Acids Res.* **19**:1349.

Ezcurra I, Wycliffe P, Nehlin L, Ellerström M, and Rask L (2000). Transactivation of the *Brassica napus* napin promoter by ABI3 requires interaction of the conserved B2 and B3 domains of ABI3 with different *cis*-elements: B2 mediates activation through an ABRE, whereas B3 interacts with an RY/G box. *Plant J.* **24**: 57-66.

Fiedler U, Filistein R, Wobus U, and Bäumlein H (1993). A complex ensemble of *cis* regulatory elements controls the expression of a *Vicia faba* non-storage seed protein gene. *Plant Mol. Biol.* **22**: 669-679.

Fields S and Song O 1989). A novel genetic system to detect protein-protein interactions. *Nature* **340**: 245–247.

Finkelstein RR (1994). Mutation at two new *Arabidopsis* ABA response loci are similar to the abi3 mutations. *Plant J*. **5**: 765-711.

Finkelstein RR and Lynch T (2000). The *Arabidopsis* ABA response gene ABI5 encodes a basic leucine zipper transcription factor. *Plant Cell* **12**: 599-609.

Finkelstein RR, Wang ML, Lynch TJ, Rao S, and Goodman HM (1998). The *Arabidopsis* ABA response locus ABI4 encodes an APETALA2 domain protein. *Plant Cell* **10**: 1043-1054.

Focks N and Benning C (1998) *wrinkled*1: a novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. *Plant Physiol*. **118**: 91–101.

Ghosh P, Dahms NM, and Kornfeld S (2003). Mannose 6-phosphate receptors: new twists in the tale. *Nat. Rev. Mol. Cell Biol.* **4**: 202-212.

Giersberg M, Saalbach I, and Bäumlein H (2004). Gene farming in pea under field conditions: Seed specific synthesis of a bacterial exoenzyme. (Kluwer Book review).

Gietz D, St. Jean A, Woods RA, and Schiestl RH (1992). Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res*. **20**:1425.

Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, and Goodman HM (1992). Isolation of the *Arabidopsis AB13* gene by positional cloning. *Plant Cell* **4**: 1251-1261.

Goldberg RB, de Paiva G, and Yadegari R (1988). Plant embryogenesis, body plan elaboration and preparation for life after the seed. *Science* **240**: 1460-1466.

Granger C, Coryell V, Khanna A, Keim P, Vodkin L, and Shoemaker RC (2002). Identification, structure, and differential expression of members of a BURP domain containing protein family in soybean. *Genome* **45**: 693-701.

Harada JJ (1998). Seed maturation and control of germination. In: *Cellular and Molecular Biology of Plant Seed Development*, Larkins B.A. and Vasil I. K. (eds).

Hartley JL, Temple GF, and Brasch MA (2000). DNA Cloning Using *in vitro* Site-Specific Recombination. *Genome Res.* **10**: 1788-1795.

Hattori J, Boutilier KA, van Lookeren Campagne MM, and Miki BL (1998). A conserved BURP domain defines a novel group of plant proteins with unusual primary structures. *Mol. Gen. Genet*. **259**: 424-428.

Held BM, John J, Wang H, Mora-Goda L, Tirimanne TS, Wurtle ES, and Colbert JT (1997). Zpr2: a novel maize gene whose mRNA accumulates in the root cortex and mature stems. *Plant Mol. Biol.* **35**: 367-375.

Herman EM (1995). Cell and molecular biology of seed oil bodies. In: *Seed development and Germination*. Kigel J. and Galili G. (eds.). M. Dekker Inc. New York, pp. 195-214.

Higgins TJV (1984). Synthesis and regulation of major proteins in seeds. *Annu. Rev. Plant Physiol.* **35**: 191-221.

Hill J, Donald KA, and Griffiths DE (1991). DMSO-enhanced whole cell yeast transformation. *Nucleic Acids Res*. **19**: 5791.

Hillmer S, Movafeghi A, Robinson DG, and Hinz G (2001). Vacuolar storage proteins are sorted in the cis-cisternae of the pea cotyledon Golgi apparatus. *J. Cell Biol*. **152**: 41-50.

Hinz G and Herman EM (2003). Sorting of storage proteins in the plant Golgi apparatus. in: The Golgi Apparatus and the plant secretory pathway. *Annual Plant Reviews*, Vol.9, Hrsg.: D.G. Robinson, Blackwell Publishing and CRC Press, Oxford. pp: 141-164.

Hinz G, Hillmer S, Baumer M, and Hohl I (1999). Vacuolar storage proteins and the putative vacuolar sorting receptor BP-80 exit the Golgi apparatus of developing pea cotyledons in different transport vesicles. *Plant Cell* **11**: 1509-1524.

Höfgen R and Willmitzer L (1988). Storage of competent cells for *Agrobacterium* transformation. *Nucleic Acids Res.* **16**: 9877.

Hohl I, Robinson DG, Chrispeels MJ, and Hinz G (1996). Transport of storage proteins to the vacuole is mediated by vesicles without a clathrin coat. *J. Cell Sci.* **109**: 2539-2550.

Huang AHC (1994). Structure of plant seed oil bodies. *Cur. Opin. Structural Biol.* **4**: 493-498.

Huang AHC (1996). Oleosin and oil bodies in seeds and other organs. *Plant Physiol* **110**: 1055–1061.

Ito H, Fukada Y, Murata K, and Kimura A (1983). Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol*. **153**:163–168.

Iwasaki T, Yamaguchi-Shinozaki K, and Shinozaki K (1995). Identification of a cis-regulatory region of a gene in Arabidopsis thaliana whose induction by dehydration is mediated by abscisic acid and requires protein synthesis. *Mol Gen Genet*. **247**: 391-398.

Jauh GY, Phillips TE, and Roger JC (1999). Tonoplast intrinsic proteins isoforms as markers for vacuolar functions. *Plant Cell* **11**:1867-1882.

Jefferson RA, Burgess SM, and Hirsh D (1986). β-glucuronidase from *E.coli* as a gene- fusion marker. *Proc. Nat. Acad. Sci. USA* **83**: 8447-8451.

Jefferson RA, Kavanagh TA, and Bevan MW (1987). GUS fusion: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J*. **6**, 3901-3907.

Kagaya Y, Okuda R, Ban A, Toyoshima R, Tsutsumida K, Usui H, Yamamoto A, and Hattori T (2005b). Indirect ABA-dependent regulation of seed storage protein genes FUSCA3 transcription factor in *Arabidopsis. Plant Cell Physiol*. **46**: 300-311.

Kagaya Y, Toyoshima R, Okuda R, Usui H, Yamamoto A, and Hattori T (2005a). Leafy cotyledon 1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INTENSITIVE3. *Plant Cell Physiol*. **46**: 399-406.

Katavic V, Reed DW, Taylor DC, Giblin EM, Barton DL, Zou J, Mackenzie SL, Covello PS, and Kunst L (1995) Alteration of seed fatty acid composition by an ethyl methanesulfonate-induced mutation in Arabidopsis thaliana affecting diacylglycerol acyltransferase activity. *Plant Physiol*. **108**:399-409.

Kater MM, Konigstein GM, Nijkamp HJJ, and Stuitje AR (1991). cDNA cloning and expression of *Brassica napus* enoyl-acyl carrier protein reductase in *Escheria coli*. *Plant Mol. Biol.* **17:** 895-909.

Keith K, Kraml M, Dengler NG, and McCourt P (1994). *Fusca3*: A heterochronic mutation affecting late embryo development in *Arabidopsis*. *Plant Cell* **6**: 589-600.

Kirsch T, Paris N, Butler JM, Beevers L, and Rogers JC (1994). Purification and initial characterization of a potential plant vacuolar targeting receptor. *Proc. Natl. Acad. Sci. USA* **91**: 3403-3407.

Koncz C and Schell J (1986). The promoter of T$_L$-DNA gene *5* controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol. Gen. Genet.* **204**: 383-396.

Koornneef M, Reuling G, and Karssen CM (1984). The isolation and characterization of abscisic acid-insensitive mutants of *Arabidopsis thaliana*. *Plant Physiol.* **90**: 463-469.

Krebbers E, Herdies L, De Clercq A, Seurinck J, Leemans J, Van Damme J, Segura M, Gheysen G, Van Montagu M, and Vandekerckhove J (1988). Determination of the processing sites of an *Arabidopsis* 2S albumin and characterisation of the complete gene family. *Plant Physiol.* **87**: 859-866.

Kroj T, Savino G, Valon C, Giraudat J, and Parcy F (2003). Regulation of storage protein gene expression in *Arabidopsis. Development* **130**: 6065-6073.

Laemmli MK (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227:** 680-685.

Leffers H, Egebjerg J, Andersen A, Christensen T, and Garrett RA (1988) Domain VI of *Escherichia coli* 23S ribosomal RNA. Structure, assembly and function. *J. Mol. Biol.* **204**, 507–522.

Lorbiecke R and Sauter M (2002). Comparative analysis of PSK peptide growth factor precursor homologs. *Plant Science* 163, 321-332

Luerßen H, Kirik V, and Miséra S (1998). The *FUSCA3* gene of *Arabidopsis thaliana* encodes a product with partial homology to Vp1/ABI3-like regulatory proteins. *Plant J.* **15**: 755-764.

Matsuoka K and Nakamura K (1999). Large alkyl side-chains of isoleucine and leucine in the NPIRL region constitute the core of the vacuolar sorting determinant of sporamin precursor. *Plant Mol.Biol*. **41**: 825 - 835

McCarty DR (1995). Genetic control and integration of maturation and germination pathways in seed development. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. **46**: 71-93.

Meinke DW (1992). A homeotic mutant of *Arabidopsis thaliana* with leafy cotyledons. *Science* **258**: 1647-1 650.

Meinke DW (1994). Seed development in *Arabidopsis thaliana*: In *Arabidopsis*. E. Meyerowitz and C. Sommerville (eds). Cold Spring Harbor, NY. Cold Spring Harbor Laboratory Pres, pp. 253-295.

Meinke DW, Franzmann LH, Nickle TC, and Yeung EC (1994). Leafy cotyledon mutants of *Arabidopsis. Plant Cell* **6**: 1049-1064.

Meyerowitz E (1989). *Arabidopsis*, a really useful weed. *Cell* **56**: 263-269.

Meyerowitz EM, Bowman JL, Brockman LL, Drews GN, Jack T, Sieburth LE, and Weigel D (1991). A genetic and molecular model for flower development in *A.thaliana. Development (supplement)* **1**: 157-167.

Moreland RB, Nam HG, Hereford LM, and Fried H (1985) Tdentification of a nuclear localization signal of a yeast ribosomal protein. *Proc. Natl Acad. Sci. USA* **82**, 6561-6565.

Mönke G, Altschmied L, Tewes A, Reidt W, Mock HP, Bäumlein H, and Conrad U (2004). Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA. *Planta* **219**: 158-166.

Müntz K (1996). Protease and proteolytic cleavage of stored proteins in developing and germinating dicotyledonous seeds. *J. Exp. Bot*. **47**: 605-622.

Mueller AJ and Heidecker U (1968) Lebensfaehige und letale *fusca* mutanten bei *Arabidopsis thaliana*. *Arabidopsis Inf. Serv*. **5**: 54-55.

Murashige T and Skoog F (1962). A revised medium for rapid growth and bioassays with tobacco tissue culture*. J. Physiol. Plant* **15**: 473-497.

Murphy DJ (1993). Structure, function, and biogenesis of storage lipid bodies and oleosins in plants. *Pro. Lipid Res.* **32**: 247-280.

Nambara E, Keith K, McCourt P, and Naito S (1995). A regulatory role for the A513 gene in the establishment of embryo maturation in *Arabidopsis thaliana*. *Development* **121**: 629-636.

Nambara E, Naito S, and McCourt P (1992). A mutant of *Arabidopsis* which is defective in seed development and storage protein accumulation is a new abi3 allele. *Plant J*. **2**: 435-441.

Nowotny V and Nierhaus KH (1982) Initiator proteins for the assembly of the 50S subunit from *Escherichia coli* ribosomes. *Proc. Natl Acad. Sci. USA* **79**, 7238–7242.

Ohlrogge J and Browse J (1995) Lipid biosynthesis. *Plant Cell* **7:** 957–970.

Ooms, JJJ, Léon-Kloosterziel KM, Bartels D, Koornneef M, and Karssen CM (1993). Acquisition of desiccation tolerance and longevity in seeds of *Arabidopsis thaliana*. A comparative study using abscisic acibinsensitive *abi3* mutants. *Plant Physiol.* **102**: 1185-1191.

Osborn TB (1924). The vegetable proteins (London: Longmans, Green).

Pang PP, Pruitt RE, and Meyerowitz EM (1988). Molecular cloning, genomic organization, expression and evolution of 12S seed storage protein genes of *Arabidopsis thaliana. Plant Mol. Biol*. **11**: 805-820.

Parcy F, Valon C, Kohara A, Miséra S, and Giraudatag J (1997). The *ABSCISIC ACID-INSENSITIVE3, FUSCA3,* and *LEAFY COTYLEDONf* loci act in concert to control multiple aspects of *Arabidopsis* seed development. *Plant Cell*. **9**: 1265-1277.

Pyke K (1994). *Arabidopsis* - its use in the genetic and molecular analysis of plant morphogenesis: *New Phytologist*. **128**: 19–37.

Ragland M and Soliman KM (1997). Two genes induced by aluminium in soybean roots. *Plant Physiol*. **114**: 395.

Raz V, Bergervoet JH, and Koornneef M (2001). Sequential steps for developmental arrest in *Arabidopsis* seeds. *Development* **128**: 243-252.

Reidt W, **Ellerström M, Kölle K, Tewes A, Tiedemann J, Altschmied L, and Bäumlein** (2001). FUS3-dependent gene regulation during late-embryogenesis. *Plant Physiol.* **158**: 411-418.

Reidt W, Wohlfarth T, Ellerström M, Czihal A, Tewes A, Ezcurra I, Rask L, and Bäumlein H (2000). Gene regulation during late embryogenesis: the RY motiv of maturation-specific gene promoters is a direct target of the FUS3 gene product. *Plant J.* **21**: 401-408.

Ryan CA and Pearce G (2003). Systemins: a functionally defined family of peptide signals that regulate defensive genes in Solanaceae species. *Proc Natl Acad Sci U S A*. **100**: 14577-14580.

Sambrook J, Fritsch EF, and Maniatis T (1989). *Molecular Cloning: A Laboratory Manual.* New York: Cold Spring Harbor Laboratory Press.

Sanger F, Nicklen S, and Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Nat. Acad. Sci. USA* **74**: 5463-5467.

Schiestl RH and Gietz RD (1989). High efficiency transformation of intact cells using single stranded nucleic acids as a carrier. *Curr. Genet.* **16**: 339–346.

Shewry PR and Casey R (1999). Seed proteins. In Seed proteins. *Kluwer Academic Publishers*.

Shewry PR, Napier JA, and Tatham AS (1995). Seed storage proteins: structures and biosynthesis. *Plant Cell* **7**: 945-956.

Sivak MN and Preiss J (1995). Starch synthesis in seed. In Seed development and germination. Edited by J.Kigel and G.Galili. Marcel Dekker*, Inc. New York*. pp.139-169.

Soderman E, Brocard I, Lynch T, and Finkelstein RR (2000). Regulation and function of the *Arabidopsis* ABA-insensitive4 gene in seed and ABA response signalling networks. *Plant Physiol.* **124**:1752-1765.

Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, Goldberg RB, and Harada JJ (2001). LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc. Nat. Acad. Sci. USA* **98**: 11806-11811.

Treacy BK, Hattori J, Prud'homme I, Barbour E, Boutilier K, Baszczynski CL, Huang B, Johnson DA, and Miki BL (1997). Bnm1, a *Brassica* pollen-specific gene. *Plant Mol. Biol.* **34**: 603-611.

Turner SR, Barrett DHP, and Casey R (1990). The effect of different alleles at r locus on the synthesis of seed storage proteins in *Pisum sativum. Plant Mol. Biol.* **14**: 793-803.

Uchiumi T, Sato N, Wada A, and Hachimori A (1999) Interaction of the sarcin/ricin domain of 23S ribosomal RNA with proteins L3 and L6. *J. Biol. Chem.* **274**, 681–686

Van der Klei H, Van Damme J, Casteels P, and Krebbers E (1993). A fifth 2S albumin isoform is present in *Arabidopsis thaliana*. *Plant Physiol.* **101**: 1415-1416.

Van der Weele CM, Spollen WG, Sharp RE, and Baskin TI (2000). Growth of *Arabidopsis thaliana* seedlings under water deficit studied by control of water potential in nutrient-agar media. *J. Exp. Botany* **51**: 1555-1562.

Vicient CM, Bies-Etheve N, and Delseny M (2000). Changes in gene expression in the leafy cotyledon 1(*lec1*) and fusca (*fus3*) mutants of *Arabidopsis thaliana* L. *J. Exp. Botany* **51**: 995-1003.

Wang A, Xia Q, Xie W, Datla R, and Selvaraj G (2003). The classical Ubisch bodies carry a sporophytically produced structural protein (RAFTIN) that is essential for pollen development. *Proc. Nat. Acad. Sci. USA*. **100**: 14487-14492.

Watson CF, Zheng L, and DellaPenna D (1994). Reduction in tomato polygalacturonase β-subunit expression affects pectin solubilisation and degradation during fruit ripening. *Plant Cell* **6**: 1623-1634.

Weber H, Borisjuk L, and Wobus U (1997). Sugar import and metabolism during seed development. *Trends Plant Sci.* **2**, 169–174.

Weber H, Borisjuk L, and Wobus U (2005). Molecular physiology of legume seed development. *Annu. Rev. Plant Biol.* **56**, 253-279.

West MAL, Yee KM, Danao J, Zimmerman JL, Fischer RL, Goldberg RB, and Harada JJ (1994). *LEAFY COTYLEDON1* is an essential regulator of late embryogenesis and cotyledon identity in *Arabidopsis. Plant Cell* **6**: 1731 1745.

Wobus U and Weber H (1999). Sugars as signal molecules in plant seed development. *Biol. Chem.* **380**, 937–944.

Wohlfarth T, Braun H, Kirik V, Kölle K, Czihal A, Tewes A, Luerssen H, Miséra S, Shutov A, and Bäumlein H (1998). Regulation and Evolution of seed globlulin genes. *Plant Physiol.* **152**: 600-606.

Wolpert L (1998). Principles of Development. Oxford University Press, NY.

Yamaguchi-Shinozaki K and Shinozaki K (1993). The plant hormone abscisic acid mediates the drought-induced expression but not the seed-specific expression of rd22, a gene responsive to dehydration stress in *Arabidopsis thaliana*. *Mol. Gen. Genet*. **238**: 17-25.

Yang Z, Aidi JE, Ait-Ali T, Augur C, Teller G, Schoentgen F, Durand R, and Durand B (1998). Sex-specific marker and trans-zeatin ribosidase in female annual mercury. *Plant Sci.* **139**: 93-103.

Zakharov A, Giersberg M, Hosein F, Melzer M, Müntz K, and Saalbach I (2004). Seed-specific promoters direct gene expression in non-seed tissue. *J. Exp. Bot.* **55**, 1463-1471.

Zheng L, Heupel RC, and DellaPenna D (1992). The β-subunit of tomato fruit polygalacturonase isoenzyme1: isolation, characterisation and identification of unique structural features. *Plant Cell* **4**: 1147-1156.

Zimmermann P, Hirsch-Hoffmann M, Hennig L, and Gruissem W (2004). GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol*. **136**: 2621-2632.

Table of figures

Table of growth media

LB
 10.0 g/L NaCl

 10.0 g/L tryptone,

  5.0 g/L yeast extract

pH 7.4

Sterilized by autoclaving.

TB
 12.0 g/L tryptone,

 24.0 g/L yeast extract

 4.0 ml glycerol

 0.017 M KH_2PO_4

 0.072 M K_2HPO_4

pH 7,4

Sterilized by autoclaving.

YEB
 0.5 g/L $MgSO_4.7H_2O$

 5.0 g/L beef extract

 5.0 g/L peptone

 5.0 g/L saccharose

 1.0 g/L yeast extract

pH 7.0

Sterilized by autoclaving.

MS
 4.49 g/L MS medium basal salt mixture including vitamins and microelements.

 30 g/L sucrose

 1% Difco agar for solid medium

 pH 5.8

Sterilized by autoclaving.

YPD
 20 g/L Difco peptone

 10 g/L yeast extract

 20 g/L agar (for plates only)

pH to 5.8

Sterilized by autoclaving.

YPDA
 Sterilized YPD

 0.003% adenine hemisulfate

SD
 6.7 g/L yeast nitrogen base without amino acids

 20.0 g/L agar (for plates only)

100 ml of the appropriate sterile 10X Dropout Solution

pH to 5.8

Sterilized by autoclaving.

Acknowledgments

Declaration

Hereby I declare that all the work presented in this manuscript is my own, carried out solely with the help of the literature and aid citied.

Furthermore I declare that I never before submitted this thesis to gain a PhD degree at any other university.

Gatersleben, July 2005

Posters

Le VS, T Rutten, J Tiedemann, R Manteuffel, and H Bäumlein (2004). The BURP/U domain protein family of Arabidopsis: a novel component of the embryogenesis related to secretion pathway. 9th *International Symposium on Plant seeds: Seeds in the –omics Era. Chateau Meisdorf /IPK Gatersleben*, *15-19 May 2004:* P43.

Le VS, T Rutten, J Tiedemann, R Manteuffel, and H Bäumlein (2003). U-domain protein family., *Institutstag IPK-Gatersleben*, *9 October 2003*: P95.

Bäumlein H, VS Le, J Tiedemann, T Rutten, and R Manteuffel (2003). Die Familie der U-Domänen-Proteine. *Programm 16. Tagung "Molekularbiologie der Pflanzen". 25-28 Februar 2003, Dabringhausen:* P66.

Curriculum vitae

Personal data

Name: Van Son Le

Date of birth: 22 January 1967

Place of birth: Hatinh, Vietnam

Nationality: Vietnamese

Education and employment

Nov. 2001 - Present: PhD. fellow

IPK-Gatersleben, Germany.

1993 - 2001:   Assistant researcher

Institute of Biotechnology (IBT), National Center for

Natural Sciences and Technology, Vietnam (NCST)

1998 - 1999:   Master of Science in Molecular Biology and

Biotechnology

Vrjie University Brussels (VUB), Belgium

Title of thesis: Cloning and characterization of seed gene

promoters of rice

1987-1992:   Diploma of Genetics

State University of Hanoi, Vietnam

Title of thesis: Structural analysis of the chromosomes of

Anopheles species in North Vietnam

1982-1985:   High school in Hatinh, Vietnam