New insights into carotenoid accumulation in tomato fruit

Nuevos aspectos sobre acumulación de carotenoides en el fruto del tomate

Karel Willem Frederik De Pourcq

Programa de doctorado de Biotecnología

NEW INSIGHTS INTO CAROTENOID ACCUMULATION IN TOMATO FRUIT

NUEVOS ASPECTOS SOBRE ACUMULACION DE CAROTENOIDES
EN EL FRUTO DEL TOMATE

Memoria presentada por Karel Willem Frederik De Pourcq para optar al grado de doctor por la Universidad de Barcelona.

Este trabajo ha sido realizado en el Departamento de Bioquímica y Biología Molecular de la facultad de Biología de la UB, y en el Centre de Recerca en Agrogenòmica (CRAG)

Directores de tesis   El autor

Albert Boronat Margosa   Santiago Imperial Ródenas   Karel De Pourcq

A Mar.

A mi familia y amigos.

A todo aquello que es bueno y justo.

maestros (en el más amplio de los sentidos), y a toda aquella gente que me ha ayudado en este periplo.

La vida, sin embargo, no va solo de maestros, de ciudadanos belgas, de saltos ni de cambios de acera, y hay que tener tiempo para otras cosas, legibles entre lineas para quien tenga la llave... Me quedo, primero, con esas sesiones, que tanto me aportaron, y que dejaron su huella (plantillas aparte). Cervezas hubo muchas, a patadas y bien repartidas; con frecuencia, regadas de buena conversación, que es como el lúpulo y la malta fermentada saben mejor. Un propósito de vida nueva es recuperar este buen hábito, el de las enriquecedoras conversaciones de cerveza. No puedo dejar de mencionar a mis hermanos gemelos (también a un ya autoproclamado "hermano bastardo") y a los maestros cerveceros europeos, por las cervezas compartidas y elaboradas, y a sus íntimos asociados llamados "los químicos" o "los del abuelo", un grupúsculo recalcitrante que ahí sigue, en tácito culto a Carmen de Mairena (aprovecho la ocasión para volver a solicitar la regencia de un tal "Rey Juan").

Recuerdos compartidos: nitbuses desde o hacia Badalona (con "loco" incluido), salas nutridas de góticos, disfraces de zombie o de "rocker", fiestas de "gomets" y de dispar mezcla de gentes, maratones de cine potable o de cine netamente neurotóxico, sesiones de juegos, Belchicas habidos y por haber, sesiones de música en astrolabis o pastissos, performances teátrico-poéticas de dificil descripción, mudanzas improvisadas, las bravas con allioli bien fuerte, el morro frito siempre pendiente, los buffets japoneses que pudiesen haber puesto carteles negándonos la entrada por ruinosos, las visitas a la arena, las sesiones de filmoteca con películas desconocidas de directores olvidados en los tiempos de los smartphones... las peripecias derivadas de la ausencia de ciertos artefactos de comunicación móvil... el "rocoto", los pimientos africanos y el KFC, los chistes y la jerga compartida (verde, exótica, o simplemente tirando a absurda), el compartir y apoyarse en unos inicios no siempre sencillos... Ciertos gallegos que no están en Galicia (ni en Barcelona) y los perniciosos efectos de su particular cocina (por así llamarla), así como ciertas excursiones de inicio planificado pero de final inopinadamente bastante falto de luz... viajes a países con ciudadanos belgas, franceses y alemanes, incluyendo ciertas islas.

Dedicado por y para...:.por la "repesca" de los masters y el viaje a Ansó que nunca hice, por los conciertos de noche con la copa en el suelo, y las nocturnas vueltas a casa en bicicleta... por los gustos compartidos. Por las conversaciones literario-filosóficas (filosofía generalmente barata), acerca de la naturaleza y el sentido de las cosas, de cómo funciona y de cómo se podría arreglar este mundo de locos. A los que conversan sin tapujos ni prejuicios, siempre con respeto (algunos de voz engolada). A aquellas personas con las que tuve grandes conversaciones, pero de las que luego no fui capaz de recordar el nombre. A las personas cuya paciencia puse involuntariamente a prueba por el mismo motivo.

Por el encontrarse para una cerveza rápida pero acabar a las tres habiendo de trabajar al día siguiente... por las eternamente míticas fiestas de Gracia y los picnics improvisados, por los conciertazos de grupos con estética de "hobbit" (¡viva Puerto de Béjar...!), por el irse a Banyoles pero no acabar de llegar, igual que el estar acabando la tesis pero no acabar de terminarla (la llamada "publicidad engañosa")... pero que finalmente todo llegue; por la paciencia que has demostrado. Por el pensamiento positivo, por la vena artística, la torrencial y la de ayudar a personas ancianas a cruzar la calle; por el tomarse las cosas con calma, que la vida son dos días... ¡así que hay que disfrutarla! Por el plasmar "lo que mola" del mundo (¡o lo que molaría!) en una libreta, por el vivir lo bueno de la vida, y el aprender de lo malo, y por el saber encontrar y sacar lo mejor de sí mismos, y llevarlo a su máxima

expresión. Por los nuevos o ya no tan nuevos amigos. Por los planes que, tarde o temprano, llevaremos a buen puerto. Por el disfrutar del camino juntos y saber encontrar el equilibrio. Por la fortuna de haberte encontrado.

A todas las personas que me hubiese gustado conocer pero no conocí y a aquella gente que conocí lo bastante para atisbar cuán grandes son: os deseo lo mejor.

A todas aquellas personas que saludan con una sonrisa sincera y desinteresada, a las personas que tratan de comprender al prójimo y a ayudarlo cuando se puede; a aquellas personas bienintencionadas que intentan hacer del lugar que los rodea un sitio un poquito mejor... sabéis quiénes sois: muchas, y muchas gracias. No cambiéis.

A los fieles, que siempre estáis ahí. A todos aquellos a los que no veo nunca, pero que seguís en mis pensamientos. Un abrazo fuerte. Aunque últimamente lo disimule bien, no os olvido.

A la gente a la que pese a todo, me haya dejado: no me lo tengáis en cuenta; aunque yo no lo sepa, estoy disimulando.

Finalmente... los esfuerzos tienen recompensa: por el ánimo de superación.

Tenían razón: ¡todo llega!

TABLE OF CONTENTS

INTRODUCTION

1. INTRODUCTION

1.1. Agronomic importance of the tomato fruit

Tomatoes are native to the Andes region of Chile, Colombia, Bolivia, and Peru. It is, however, believed that they were first domesticated in Mexico. It is believed to have been already cultivated as early as 500 BC in southern Mexico and areas of Mesoamerica. After the start of the Spanish colonization of the Americas, the crop was brought to Europe, from where it eventually expanded to the rest of the World.

The earliest written European reports on the tomato appeared in 1544, written by the Italian physician and botanist Pietro Andrea Mattioli, who wrote that a new type of plant had been brought to Italy whose fruits were edible and blood red or golden colour when mature. After ten more years, the same Pietro Mattioli named the tomatoes as "pomi d'oro".

Centuries later, tomato has become one of the most widely cultivated crops worldwide. It ranks to be the tenth most grown individual feed crop in tonnes per year as of 2011. The bulk of the world production is grown in China, India, the United States, and the Mediterranean Basin countries.

Tomato has become one of the crops most widely consumed, being a common ingredient of salads, juices, sauces and diverse food preparations worldwide. The preferred tomato preparation varies depending on the country, and depending on the final use, the tomato fruit has one set of desired characteristics or another. For fresh tomato the main traits pursued are colour, taste, flavour and firmness. For industrial preparations the requirements can differ, and additives can be included to meet the specific requirements.

The main organoleptic and nutritional value and characteristics of tomato depend on sugar content, pigmentation, volatiles influencing flavour, the presence of antioxidants, organic acids, oligo-elements and vitamins, and the proportion of dry matter.

In Western diets the tomato fruit and its derivatives constitute the main source of lycopene, which is a very strong antioxidant, as well as an important source of beta-carotene, which has provitamin-A activity.

Rank	Item	Million Tonnes
1	Sugar cane	1.800,4
2	Maize	885,3
3	Rice (paddy)	722,6
4	Wheat	701,4
5	Potatoes	373,2
6	Sugar beet	273,5
7	Soybeans	262,0
8	Cassava	256,4
9	Palm fruit oil	234,3
10	**Tomatoes**	**159,3**
11	Barley	133,0
12	Bananas	107,1
13	Sweet potatoes	105,0
14	Watermelons	102,9
15	Onions (dry)	86,3
16	Apples	75,5
17	Oranges	69,5
18	Grapes	69,1
19	Rapeseed (canola)	62,5
20	Sorghum	58,6

Table 1. Top 20 individual world feed crops. Production indicated in million tonnes per year. Source: FAOSTAT, 2013.

Rank	Country	Million Tonnes	% World share
1	China	48,6	30,5
2	India	16,8	10,6
3	United States of America	12,5	7,9
4	Turkey	11,0	6,9
5	Egypt	8,1	5,1
6	Iran (Islamic Republic of)	6,8	4,3
7	Italy	6,0	3,7
8	Brazil	4,4	2,8
9	Spain	3,9	2,4
10	Uzbekistan	**2,6**	1,6
11	Mexico	**2,4**	1,5
12	Russian Federation	**2,2**	1,4
13	Ukraine	**2,1**	1,3
14	Nigeria	**1,5**	0,9
15	Tunisia	**1,3**	0,8
16	Portugal	**1,2**	0,8
17	Morocco	**1,2**	0,8
18	Greece	**1,2**	0,7
19	Syrian Arab Republic	**1,2**	0,7
20	Iraq	**1,1**	0,7
	Total top 20	**136,1**	**85,4**

Table 2. Top 20 world tomato producers. Production in million tonnes per year. Sorce: FAOSTAT, 2013.

1.2. Fruit ripening: climacteric fruits

Fruit ripening is a very coordinated process that involves a series of organoleptic changes in the fruit, as well as the maturation of the seeds, all in order to favour reproduction of the plant. The ripening process is influenced by many genes, which modify a number of fruit traits. The affected characteristics include sugar, acids and pigment contents, as well as the release of volatiles. As well, a degradation of chlorophylls and a decrease in the lignification level of the pericarp occurs, finally yielding it softer and more susceptible to pathogen attack (Osorio et al., 2013) (Figure 1).

Tomato is the classic model for the study of ripening of climacteric fruits. Climacteric fruits are characterized by the occurrence of an ethylene peak accompanied by an increased respiration, before being able to complete the maturation process (Nath et al., 2006) (Figure 1). There are three stages in the ripening of climacteric fruits: pre-climacteric stage, central or climacteric stage and post-climacteric stage. In the first stage, the fruit prepares the impending changes accumulating a series of metabolic precursors. The climacteric or central stage is considered to be the actual start of ripening. It is in this stage when the mentioned ethylene and respiration peaks induce irreversible changes that eventually lead to the final maturation. The post-climacteric stage is when the changes in firmness, colour, taste and flavour occur, eventually leading to possible uptake and seed dissemination by herbivores or, eventually, to fruit senescence and destruction, resulting in seed release (Nath et al., 2006).

1.3. Relevance of ethylene in tomato fruit ripening

Even though ethylene has a well established importance during maturation of the tomato fruit, the molecular basis of the response of the fruit to this ethylene formation is not very well-known (Giovannoni, 2007). A number of membrane receptors, among which NR and LeETR4 happen to be expressed at the end of the pre-climacteric stage, and seem to have a critical role in the onset of the ripening cascade of events (Tieman et al., 2000; Klee, 2002).

Ethylene biosynthesis genes are tightly regulated, being those some of the first genes induced during the fruit maturation (Itkin et al., 2009). Ethylene is generated at the so-called Yang cycle, in the final step of which ethylene is released. Different isoforms exist of the involved enzymes, providing part of the framework for the overall regulatory architecture (Nakatsuka et al., 1998; Nath et al., 2006).
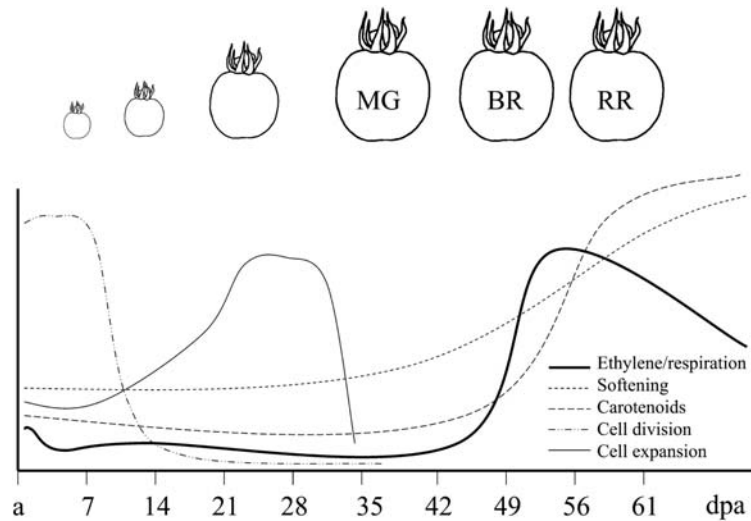
Figure 1. Outline of events during tomato fruit ripening. The variation in ethylene biosynthesis as well as the physic and organoleptic changes that occur during tomato fruit development and maturation are indicated. The coordinate axis indicates the number of days post-anthesis. Reproduced with permission from Giovannoni, 2004 with permission of the American Society of Plant Physiologists. Permission conveyed through Copyright Clearance Center, Inc.
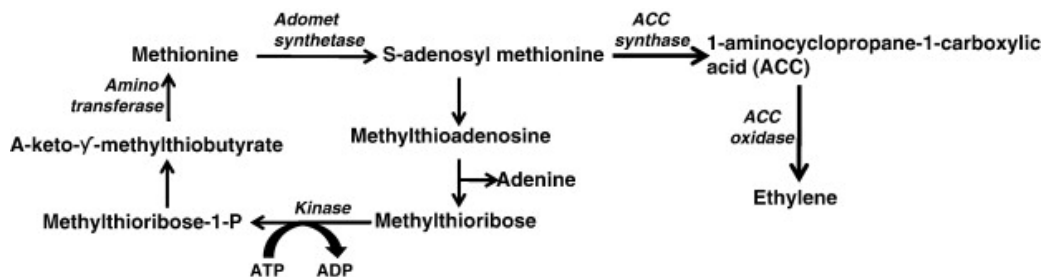
Figure 2. Biosynthesis of ethylene. Republished with permission of Springer Verlag, from Bapat et al., 2010; permission conveyed through Copyright Clearance Center, Inc.

Ethylene biosynthesis, and with it the onset of ripening, is up regulated by the hormone abscisic acid (Zhang et al., 2009). Other regulation of ethylene biosynthesis and climacteric ripening involves a variety of phosphorylation cascades, and allosteric inhibition of the ethylene biosynthetic enzymes (Manjunatha et al., 2010). On their part, plant hormone cytokinines and auxins are known to block fruit ripening. Other factors reported to be playing as well roles in the regulation of ripening are brassinosteroids, polyamines and sugars (Nath et al., 2006; Srivastava and Handa, 2005).
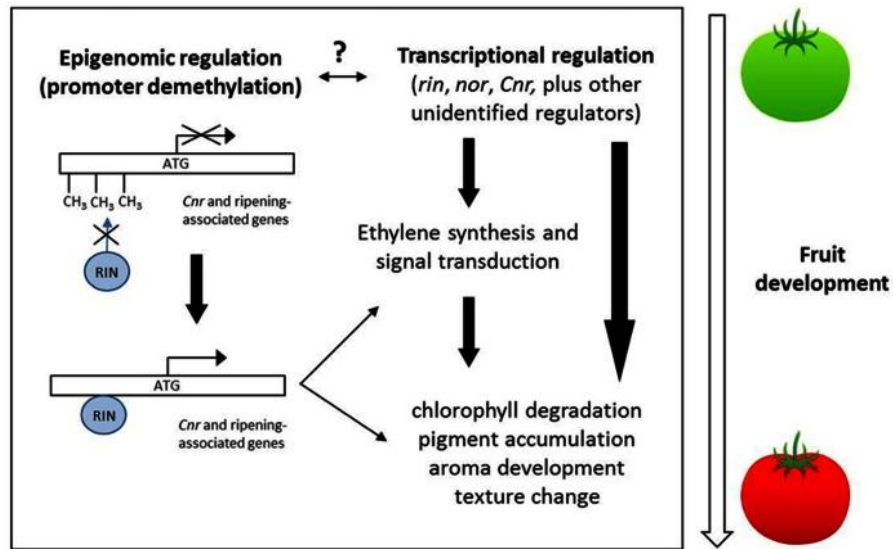
Figure 3. Ripening in tomato, a climacteric fruit. Copyright © 2013 Osorio, Scossa and Fernie.
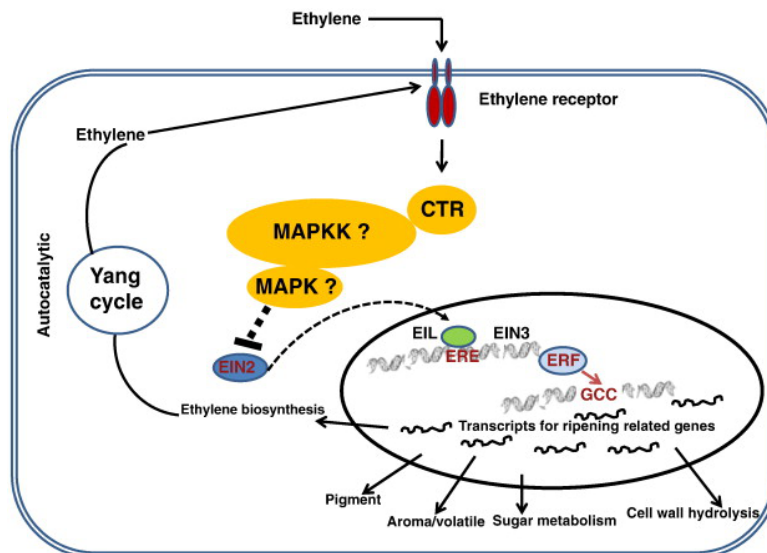
Figure 4. Model for ethylene signal transduction depicting activities of various signal molecules during fruit ripening. CTR; Constitutive triple response, MAPK; Mitogen activated protein kinase, MAPKK; Mitogen activated protein kinase, EIN2; Ethylene insensitive 2, EIN3; Ethylene insensitive 3, EIL; Ethylene insensitive like, ERE; Ethylene-responsive element, ERF; Ethylene response factor. Biosynthesis of ethylene. Republished with permission of Springer Verlag, from Bapat et al., 2010; permission conveyed through Copyright Clearance Center, Inc.

Ethylene triggers its responses in a complex way, and only a few key points are known. Of the mentioned membrane receptors, ETR4 is activated by extracellular ethylene, relaying the signal through a serious of phosphorylation cascades. This eventually results in the onset of the early steps of the maturation process. ETR3, another such membrane receptor is also involved in

the early ripening, participating in the activation of expression of early ripening genes (Leclercq et al., 2005; Lin et al., 2009).

The activated genes in turn activate or inhibit another group of not necessarily ethylene-responsive genes that promote rapid metabolic and biochemical shifts in the fruit. Those changes eventually result in the fruit becoming irreversibly committed to complete the ripening process (Klee and Giovannoni, 2011).

1.4. Carotenoids and their biological relevance

Carotenoids are isoprenoid molecules that are widespread in nature and are typically seen as pigments in fruits, flowers, birds and crustaceans, as well as in some bacteria and in microscopic animals. Over 700 carotenoids have been identified (Britton, 1995; Fraser and Bramley, 2004; DellaPenna and Pogson, 2006; Fraser et al., 2009).

Carotenoids first appeared on Earth, seemingly in archaebacteria, several thousand million years ago and played an important role in the rise of photosynthesis, providing the initially anoxic earth with the oxidant atmosphere it boasts today (McLaren and Kraemer, 2012). Nowadays, mainly photosynthetic organisms produce them, though exceptions as the synthesis by some sporulating *Bacillus* species have been reported (Perez-Fons et al., 2011). The basic structure of carotenoids includes a skeleton of 40 carbon atoms (Figure 5). This basic structure can be modified extensively, mainly by cyclizations at the ends of the molecule or by oxidations introducing hydroxyl groups (Figure 6) (Britton, 1995).
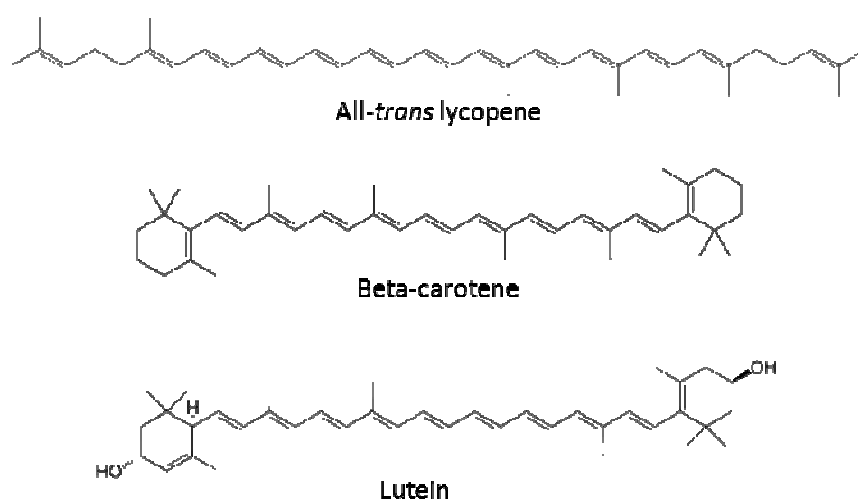
Figure 5. Chemical structures of three carotenoids present in tomato fruit.

In plants, carotenoids are required for photosynthesis-related light-harvesting and for photoprotection against photooxidative damage, as they are efficient singlet molecular oxygen quenchers and peroxyl radical scavengers (Britton, 1995; Stahl and Sies, 2003). Among their other roles, carotenoids are the precursors of ABA and strigolactone and also play an essential role in root-mycorrhiza interactions (Bouvier et al., 2005; Cazzonelli and Pogson, 2010). Further roles are frequently linked to the coloration they confer to flowers and fruits, in relation to plant reproduction.

Animals are unable to synthesize carotenoids, and need to ingest them in the diet in order to obtain vitamin A, which plays a variety of roles in animal physiology and metabolism. The term vitamin A is in fact a generic descriptor for retinoids that exhibit the biological activity of all-trans retinol (Figure 6). Carotenoids with pro-vitamin A thus have the potential for becoming such kind of retinoids. Though both carotenoids and retinoids are found in plants and animals, carotenoids are prevalent in plants, while retinoids are the preponderant forms in animals. The vast majority of existing retinoids derive from carotenoids (McLaren and Kraemer, 2012).
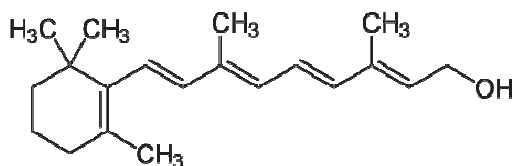
Figure 6. Retinol, a retinoid obtained through breakdown of beta-carotene.

Vitamin A deficiency of staple foods is a problem in some regions of the world, and its importance has ignited biotechnological efforts in order to solve these problems (Howitt and Pogson, 2006). Carotenoids are not only of nutritional interest, but also have industrial applications as natural pigments in the food, feed and cosmetic industries (Gordon et al., 1982; Anunciato and da Rocha Filho, 2012).

1.5. Health benefits of carotenoids

In humans, carotenoids are part of the antioxidant defence system as well. They have been shown to be capable of synergistic interactions with other antioxidants, being mixtures of carotenoids more effective than single compounds. This has been proven for the combination of lutein and lycopene (Heber and Lu, 2002). Evidence suggests that carotenoids also protect human skin, as well as the ocular macula lutea against photooxidative damage. As most

carotenoids exhibit a 450 nm absorption maximum, it has been proposed that blue light filtering could be one of the mechanisms involved (Stahl and Sies, 2003).

The intake of lycopene, the very major carotenoid present in ripe tomato fruit (Figure 5), has been inversely related to the risk of lung and prostate cancer (Arab and Steck, 2000; Holzapfel et al., 2013). A major mechanism of action of lycopene, as for the other carotenoids, is believed to be its antioxidant action, as lycopene can scavenge free radicals and quench singlet oxygen (Rodriguez-Amaya, 2010) and mutagenesis, as measured by the Ames test, is reduced under lycopene treatment (Heber and Lu, 2002). Moreover, at physiological concentration lycopene is capable of interfering with growth factor receptor signalling and cell cycle progression, particularly in prostate cancer cells. As well, the gene connexin 43 is up regulated by lycopene in humans, allowing direct intercellular gap junctional communication restoration or up regulation in tumours (where it is frequently deficient). Both this effects result in a decrease in cell proliferation (Heber and Lu, 2002), showing that anti-oxidant activity is not the only mechanism by which lycopene acts to protect against cancer.

1.6. Carotenoid biosynthesis and accumulation

Carotenoid biosynthesis is a process regulated throughout the life cycle of a plant with dynamic changes in response to developmental requirements and to external environmental stimuli (Cazzonelli and Pogson, 2010). In plants, all stages of carotenoid biosynthesis take place within plastids (Hirschberg, 2001). Carotenoids, as well as all other isoprenoids, are synthesized from the two precursor dimethylallyl diphosphate and isopentenyl diphosphate (DMAPP and IPP) (Bouvier et al., 2005).

Plants contain two metabolic pathways for the synthesis of carotenoid precursors: one located in the cytosol and endoplasmic reticulum and another located in the plastids. They are, respectively, the mevalonate (MVA), and the methylerythritol (MEP) pathway. Moreover, an exchange of the precursor IPP among those two pathways has been described, yielding terpenoid biosynthesis more complex (Bouvier et al., 2005). In the case of carotenoids, these precursors are predominantly obtained from the plastid-localized MEP pathway (Rodríguez-Concepción, 2010). The steps of both the MVA and MEP pathways and their compartmentalization inside the cell are shown in Figure 7.

The biosynthesis of carotenoids in plant plastids follows a sequential addition of precursor molecules over a short pathway until a crossroads is reached. In the case of

carotenoids, this crossroads is geranylgeranyl diphosphate, a metabolite which also serves as a precursor for the synthesis of a myriad of other plastidial isoprenoid compounds including phylloquinones, plastoquinones, chlorophylls, gibberellins and tocopherols (Bouvier et al., 2005; Ruíz-Sola and Rodríguez-Concepcion, 2012), as can be seen in Figure 8.

The first dedicated step of carotenogenesis is the conversion of GGPP into phytoene, a colourless 40-carbon intermediate, in the plastidial stroma (Figure 9). This reaction, catalyzed by the enzyme phytoene synthase (PSY), is regarded as being the main rate-limiting step in the pathway (Hirschberg, 2001). Coloured carotenoids are derived from phytoene by a series of desaturation, cyclization and oxidation reactions catalyzed by membrane bound enzymes. Phytoene is first converted to lycopene by two desaturases (PDS and ZDS) and an isomerase (CRTISO). Lycopene is then converted to the cyclic carotenoids a-carotene and b-carotene by e-lycopene cyclase (LCY-E) and b-lycopene cyclase (LCY-B) (Figure 8). Two beta-ring hydroxylations of beta-carotene lead to the formation of zeaxanthin, whereas hydroxylation of the b- or e-ionone groups of a-carotene leads to the formation of lutein (Hirschberg, 2001). Further modifications lead to the formation of the high diversity of carotenoids found in nature, with more than 700 structures identified (DellaPenna and Pogson, 2006).
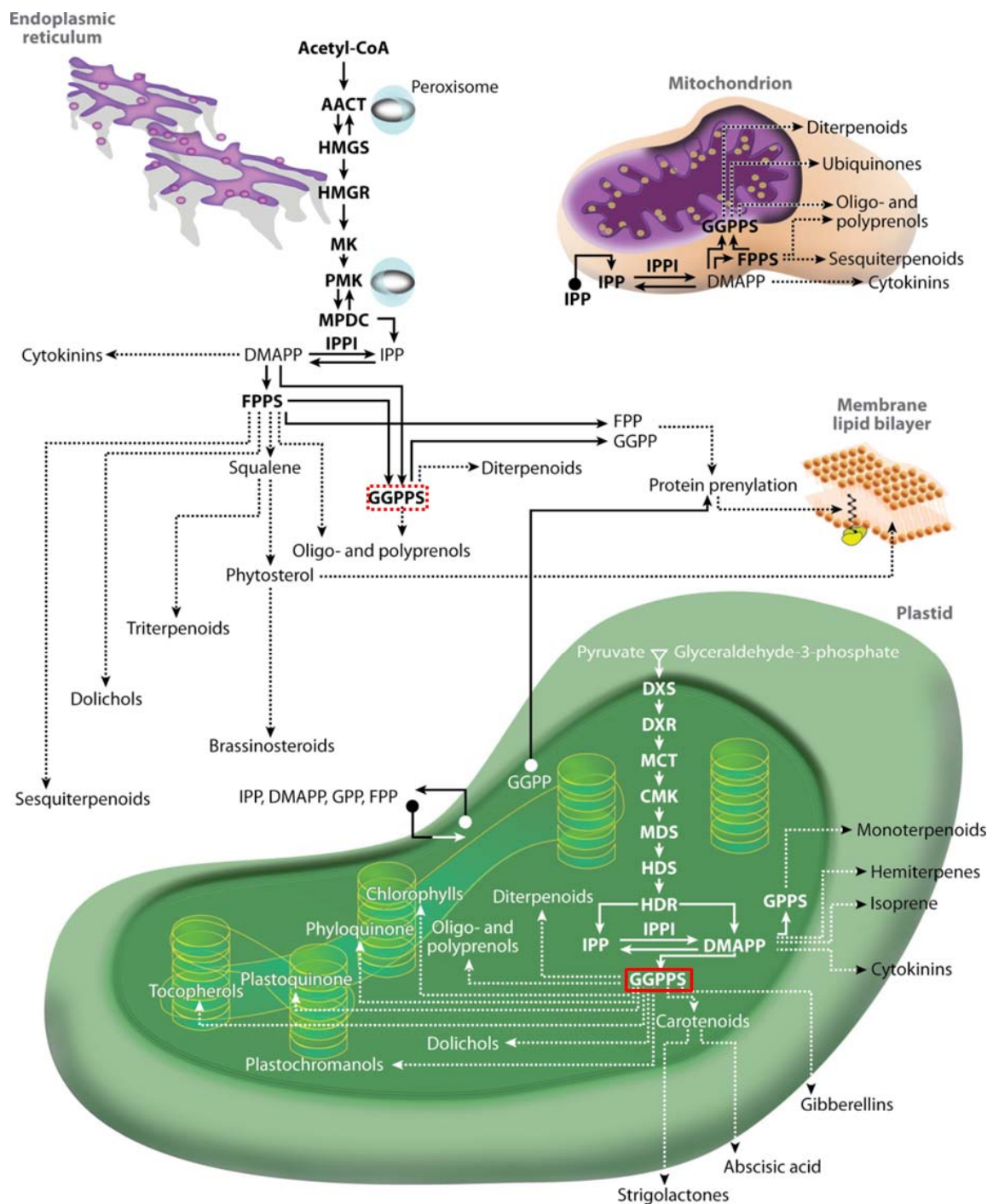
Figure 7. Subcellular compartmentalization of the MVA and MEP pathways in plant cells. GGPPS, a branching point of the MEP pathway, is highlighted in red, from which many different compounds are ultimately synthesised. Republished with permission of Annual Reviews, from Vranová et al., 2013; permission conveyed through Copyright Clearance Center, Inc.
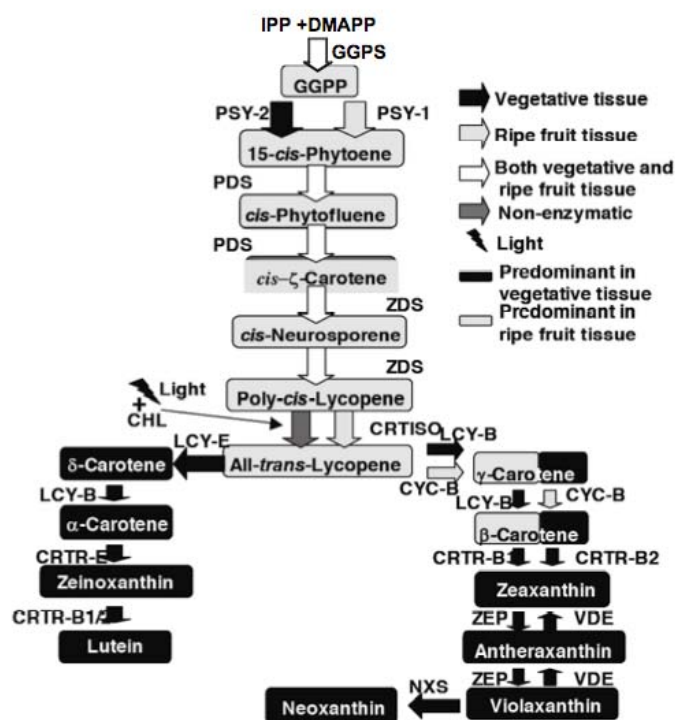
Figure 8. Carotenoid biosynthesis in vegetative and ripe tissue of tomato. Carotenoids present at detectable levels in ripe tomato fruit are shown in grey boxes, and those in black boxes are observed in vegetative tissues. Grey arrows indicate enzymes predominant in ripening tomato fruit; black arrows indicate enzymes that predominate in vegetative tissues; clear arrows are for enzymes present in both vegetative and ripening fruit. Abbreviations: GGPS, geranylgeranyl diphosphate synthase; PDS, phytoene desaturase; ZDS, z-carotene desaturase; CRTISO, carotene isomerase; LCY-B, lycopene b-cyclase; CYC-B, chromoplast-enhanced; LCY-E, Lycopene e-cyclase; CRTR-B1, b-ring hydroxylase-1; CRTR-B2, b-ring hydroxylase-2; CRTR-E, e-ring hydroxylase; NXS, neoxanthin synthase; VDE, violaxanthin de-epoxidase; ZEP, zeaxanthin epoxidase. Adapted from Fraser et al. (2007).

Lycopene begins to accumulate at breaker stage of tomato fruit (when a red/orange coloration becomes apparent) and its concentration increases up to 500-fold in ripe fruits (Fraser et al., 1994). During this process, the transcription of the genes encoding PSY and PDS is up-regulated, whereas the expression of the mRNAs of LYC-B and LCY-E is strongly reduced (Hirschberg 2001) (Figure 8).

Carotenogenesis in ripening tomato fruit is controlled by regulatory mechanisms that are distinct from those of photosynthetic tissues (Hirschberg, 2001). Because of this, the knowledge generated in model plants lacking chromoplasts, like Arabidopsis, is of limited interest in understanding the regulation of carotenoid biosynthesis in tomato fruit. Although the regulation

27

of gene expression at the transcriptional level is a key regulatory mechanism controlling carotenoid biosynthesis in chromoplasts, post-transcriptional regulation of key carotenogenic enzymes involving feed-back mechanisms by end products has also been reported (Fraser and Bramley, 2004).

It is noteworthy that while metabolic engineering efforts of carotenoid biosynthesis in bacteria have been quite successful (Immethun et al., 2013), this is not the case with engineered plant carotenogenesis. Seemingly, a limited knowledge on the interaction with parallel pathways has hindered the obtention of equivalently high carotenoid production levels. For instance, it is known that DXS, DXR and HDR are the enzymes controlling the flux of the MEP pathway providing the bulk of precursors of carotenogenesis (Vranová et al., 2013). The factors regulating flux through the carotenoid pathway are not completely understood, severely burdening metabolic engineering in this field (Lee et al., 2012). Recent works have been addressing how the gene networks of these two pathways are organized and regulated, and how network perturbations impact each pathway and plant development. This will likely help to address the mentioned difficulties in future plant metabolic engineering efforts (Vranová et al., 2013).

In the past, carotenoid sequestration in plants received little attention relative to studies on the biochemistry and molecular biology of carotenoid biosynthesis. In recent years, however, and based mainly on studies of fibrillar and globular chromoplasts, different carotenoid-associated proteins participating in carotenoid-lipoprotein structures have been identified and characterized (Vishnevetski et al., 1999; Simkin et al., 2007). Furthermore, few studies have been addressed to correlate carotenoid biosynthesis and accumulation with the suborganellar structures present in the chromoplast.

1.7. Plastids: general aspects

All existing plastids are derived from the endosymbiosis of a cyanobacterium and a mitochondria-bearing host. This is true also for plastids of non-photosynthetic organism, as is the case of the apicoplast of *Plasmodium falciparum* (Ralph, 2005). Plastids in algae, for their part, stem in the different phyla from multiple independent endosymbiotic events (Chan et al., 2011). Plastids in general have experienced a wide lateral mobility across the branches of evolution (Sheiner and Striepen, 2013).

Plastids have developed during evolution a range of functions inside eukaryotic cells

ranging from photosynthesis, fatty acid, carotenoid, amino acid and chlorophyll biosynthesis, storage functions, nitrogen and sulfur assimilation, and aromatic and terpenoid compound production, among some others (Breuers et al., 2011; Facchinelli and Weber, 2011). Plastids which share other common functions have been grouped into different classes (Wise, 2006): a) chloroplasts, which perform photosynthesis; b) chromoplasts, which accumulate carotenoids; c) proplastids, from which other plastids derive; d) Gerontoplasts, which guide the controlled dismantling in senescing tissues; e) etioplasts, plastids which have not been exposed to light and whose development to chloroplasts has been arrested; f) leucoplasts, colorless plastids which can differentiate into one of the three following other plastids: i) amyloplasts, which accumulate starch; ii) elaioplasts, which store lipids; and iii) aleuroplasts or proteinoplasts, which store and modify proteins.

The events occurring during tomato fruit ripening involve chloroplasts and chromoplasts. While carotenoids are present to some extent in nearly all types of plastids (Howitt and Pogson, 2006), in the case of tomato fruit these two very different examples of plastid exhibit a very different behaviour as far as carotenoid storage is concerned. In chloroplasts, carotenoids are overwhelmingly associated to chlorophyll-binding proteins. In contrast, chromoplasts are characterized by the appearance of specialized carotenoid-accumulating structures (Vishnevetsky et al., 1999).

1.8. Plastid organization and integration into cell functions

Most general information available on plastids is actually referred to chloroplasts, which are by far the most well-known and studied kind of plastid. Although the chloroplast is the photosynthetic organelle of algal and plant cells, it is also involved in a range of other essential functions, like carbon oxidation via photorespiration, the synthesis of carotenoids, α-tocopherol (vitamin E), plastoquinone and phylloquinone (vitamin K), lipids and amino acids, the assimilation of sulfur and nitrogen, and a series of oxidative reactions referred to as chlororespiration (Wise, 2006).

The energy status of the plant cell strongly depends on the energy metabolism in the chloroplasts and mitochondria, which are capable of generating ATP by either photosynthetic or oxidative phosphorylation (Flügge et al., 2011). The metabolism of plastids is very intertwined with their surrounding cytosol. This causes a massive traffic of metabolic precursors,

intermediates and products inwards and outwards, as both cytosolic and plastidic processes are dependent on intermediates synthesized by the other compartment (Breuers et al., 2011; Weber et al., 2005). Additionally, chloroplasts are capable of only a limited amount of semi-autonomous protein synthesis. This leads to more than 90% of their proteins to be nuclear-encoded and to have to be imported from the cytosol (Shi and Theg, 2013). In fact, the plastid is not only the most active site of protein transport in the cell, but also represents the most topologically complex organelle, making the task of protein guidance to its interior very intricate (Bruce, 1998; Lee et al., 2013).

The envelope is the only permanent membrane structure, and the common trait of all the different types of plastids. It plays a central part in the integration and trafficking proteins and metabolites of the plastid with its surroundings. The envelope is also a key player in plastid biogenesis (Block et al., 2007). Being in fact a two-membrane system, the plastid envelope is divided into outer and inner envelope, the latter in turn engulfing the stroma, the interior of the plastid. The outer envelope stems originally from the plasma membrane of a cyanobacterial endosymbiont (Inoue, 2011) and is a very metabolically active compartment of the chloroplast, having a key role in fatty acid metabolism and membrane lipid production. It also seems to have a role as a defence platform against biotic and abiotic stresses (Breuers et al., 2011). The inner envelope membrane pertains to a former endosymbiont cyanobacteria which ultimately became the plastid (Inoue, 2011). It engulfs the interior, playing its role in the second part of the import of proteins inside the plastid.

A third membrane system exists in the stroma of chloroplasts: the thylakoids, which harbour the complexes of the photosynthetic machinery, crucial to plant life. Other plastids might contain or not different membrane systems or components inside the stroma. The kind and function of a plastid will determine which those components are. A number of lipoproteic particles called plastoglobules exist in the stroma, often in close contact with the thylakoidal membrane. These particles have been related to lipid storage and stress response in chloroplasts, as well as to other metabolic processes (Nacir and Bréhélin, 2013).

Phospholipids are the major components of cellular membranes in animal cells and bacteria, as well as non-chloroplastic cellular membranes in plants. The situation in plastids is notably different, as the major glycerolipids of chloroplasts are monogalactosyl diacylglycerol (MGDG), digalactosyl diacylglycerol (DGDG), and sulfoquinovosyl diacylglycerol (SQDG).

Furthermore, two phospholipids, phosphatidyl glycerol (PG) and phosphatidyl choline (PC), are found in chloroplast membranes. The dominant lipids within plastid membranes are MGDG and DGDG, whose synthesis is initiated at one of two possible sites: either in the plastid itself, or at the endoplasmic reticulum. Regardless of that, the final steps of MGDG and DGDG synthesis are necessarily completed at the plastid, particularly in the plastid envelope membranes (Hofmann, 2008b). The lipid composition of thylakoid membranes is highly conserved among oxygenic photosynthetic organisms (Sakurai et al., 2007), and is also found to be predominantly made up of uncharged galactoglycerolipids as MGDG and DGDG as well as the anionic lipids SQDG), and PG (Xu et al., 2008). Apparently, PG is the only phospholipid present in the thylakoid membranes (Sakurai et al., 2007). For their part, the photosystems and light-harvesting complexes in the thylakoids are rich in photosynthesis-related pigments (chlorophylls, carotenes and xanthophylls) and contain a unique set of prenylquinol lipids (tocochromanol/vitamin E, plastoquinol, and phylloquinol/vitamin K1) (Bréhélin and Kessler, 2008)

Inside the plastid stroma also reside multiple copies of the plastid chromosome, folded together with proteins and RNA into structures known as nucleoids. Nucleoids include proteins involved in DNA replication, organization and repair, as well as transcription, mRNA processing, splicing, and editing (Majeran et al., 2012). Much of the regulation occurring inside chloroplasts involves the nucleoid, as plastid regulation has been reported to proceed mainly at the transcriptional level (Choquet and Wollman, 2002; Hofmann, 2008a). It has been found that the plastid genome-encoded genes are expressed at high levels in photosynthetically active chloroplasts, while they are generally very down regulated in other plastids (Kahlau and Bock, 2008).

Plastid biogenesis depends on binary fission through the envelope membranes. As with the rest of plastid processes, the import of nuclear-encoded proteins through the plastid envelope is critical, as only 10 % or less of the plastidic proteins are encoded by the plastid genome (Inoue, 2011; Shi and Theg, 2013). Plastids have in general a very low evolution rate by cause of factors like their uniparental inheritance, very effective repair system, as well as the rarity of plastid-plastid fusion (Wicke et al., 2011).

1.9. Chromoplasts: differentiation and functional roles

Chromoplasts are colored plastids found in flowers, fruits and some roots. Chromoplasts are thought to collaborate in the visually and organoleptically mediated attraction of animals and

insects in order to increase reproductive success of the plant (Wise, 2006).

Mature chromoplasts of different species show high morphological diversity. A classification was created, according to which they were subdivided into five different classes, depending on their morphology (Camara et al., 1995): i) Globular chromoplasts, which are characterized by the accumulation of plastoglobules inside the plastid stroma, and are the most primitive kind of chromoplasts; ii) Crystalline chromoplasts, a widespread chromoplast class, in turn subdivided into a) large beta-carotene crystal forming; b) small beta-carotene crystal forming; and c) lycopene crystal forming; iii) Fibrilar and tubular chromoplasts, where structures of the mentioned shapes are found. This group is subdivided into a) fibrilar organized in bundles; b) fibrilar organized into dispersed substructures; and c) tubular; iv) Membranous chromoplasts, characterized by an extended development of concentric membranes, and low plastoglobule content; and v) Reticulo-tubular chromoplasts, which display a network of twisted fibrils which fill the stroma, in addition to few plastoglobules.

In tomato fruit, the inside of chromoplasts is commanded by a series of membranous structures, large plastoglobules and long crystalline-like structures (Harris and Spurr, 1969). In some plants, plastoglobules have been found to accumulate carotenoids (Vishnevetsky et al., 1999) as well as participating in diverse metabolic processes (Nacir and Bréhélin, 2013).

Chromoplasts develop from preexisting plastids, most usually chloroplasts (Camara et al., 1995; Egea et al., 2010) and, depending on the species and tissue, also from proplastids or elaioplasts (for instance in orange carrot roots) (Wise, 2006). The process of chromoplastogenesis involves a huge accumulation of carotenoids (Leitner-Dagan et al., 2006), as well as an extensive reorganization of the internal membrane system associated to the creation of carotenoid-accumulating structures (Egea et al., 2010).

In tomato fruit, many plastid genes are found to be strongly down regulated if compared to leaf chloroplasts, and during chloroplast to chromoplast transition no important differences in the RNA transcribed inside the plastid occur during ripening. Chloroplast to chromoplast transition is found to rely on changes of the expression of nuclear-encoded genes (Egea et al., 2010; Kahlau and Bock, 2008).
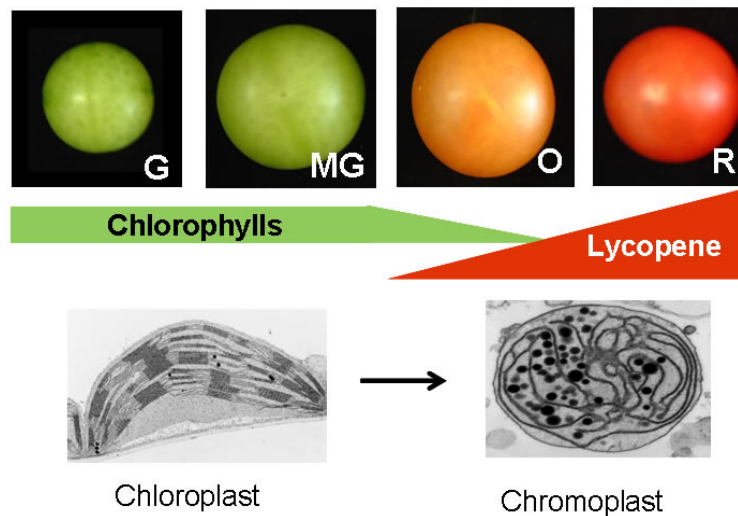
Figure 9. The transition from chloroplast to chromoplast in ripening tomato fruit. G

During chromoplast differentiation the plastid genome remains stable, while there is both transcriptional and translational down-regulation of plastidial genes, more markedly for photosynthesis-related genes (Kahlau and Bock, 2008). While in chloroplasts much of the regulation generally seems to happen at the transcriptional level (Choquet and Wollman, 2002), it is suggested that in the process of chloroplast to chromoplast transition regulation acts mainly at translational level (Hofmann, 2008a). Genes coding for proteins involved in gene expression retain their expression, indicating that a remnant of gene expression capacity must be present (Kahlau and Bock, 2008).

During the structural differentiation of the chloroplast into chromoplast, a small representation of RNA expression machinery is kept active and the only endogenous plastidic gene maintaining a high expression is *accD*, which is the sole plastid-encoded protein involved in fatty acid synthesis (Kahlau and Bock, 2008). More recently it has been confirmed that lipid biosynthesis is indeed a very active process in tomato fruit chromoplast and that metabolic channeling of metabolic precursors to acetyl-CoA is a very efficient process (Angaman et al., 2012). It also has been reported that the major metabolic shifts happening prior to fruit ripening are preceded by the accumulation of the plastid-encoded acetyl-CoA carboxylase D (accD). It has been proposed that accD could account, among other possible roles, for the creation of a storage matrix for carotenoid accumulation (Barsan et al., 2012; Hofmann, 2008a).

Envelope and stroma proteins remain essentially stable, while there is a strong decrease in the abundance of thylakoid proteins, and those involved in light reactions and carbohydrate

metabolism. Thylakoids and the photosystem biogenesis machinery gradually disappear as well as the plastidial division machinery. A strong increase occurs in the abundance of proteins related to isoprenoid biosynthesis and stress response (Barsan et al., 2012).

In the ripe chromoplast a high abundance of proteins related to energy obtention is found in chromoplasts, as well as metabolite import machineries (Barsan et al., 2010). The degradation of the photosynthetic apparatus renders chromoplasts unable to photochemically synthesize ATP. While it is widely accepted that non-photosynthetic plastids obtain ATP via glycolysis or by import from the cytosol, it has been shown that tomato fruit chromoplasts can generate ATP de novo through an alternative respiratory pathway (Pateraki et al., 2013).

A strong representation of proteins of the lipid metabolism and protein trafficking is in display in the mature chromoplast, as well as the proteins of the lipoxygenase pathway, which are necessary for the biosynthesis of lipid-derived aroma volatiles. Proteins related to starch biosynthesis and starch degradation are found to co-exist. Chlorophyll degradation proteins are found, while neither chlorophyll biosynthetic nor proteins related to thylakoid transport machinery are found. Upon completion of the transition, chromoplasts still contain the proteins of the entire Calvin cycle, including RuBisCo, and the entire pentose phosphate pathway (Barsan et al., 2010).

1.10. The plastoglobule: a lipoproteic particle inside plastids

Plastoglobules are lipoproteic particles ubiquitous all over the chloroplast stroma, which are believed to play a role in carotenoid synthesis and sequestration. They are found in most, if not all, kinds of plastids (Bréhélin and Kessler, 2008; Vishnevetsky et al., 1999).

At least in chloroplasts, plastoglobules seem to be formed from outgrowths at the outer rim of thylakoid grana (Austin et al., 2006) (Figure 10). As a consequence they are surrounded by a lipid monolayer. The size and number of plastoglobules varies during plastid development and differentiation, being both strongly increased during luminic stress, senescence or in mutants impaired in thylakoid formation (Bréhélin et al., 2007). The underlying trigger for the increase in plastoglobule number is suggested to be the up regulation of lipid metabolism. The plastoglobule suggestedly takes part in a dynamic transport equilibrium between plastoglobules and thylakoids of the lipids contained in the plastoglobule core, and the interior of the thylakoid membrane. Hence plastoglobules would be functioning as both lipid biosynthesis and storage subcompartments of thylakoids (Austin et al., 2006).
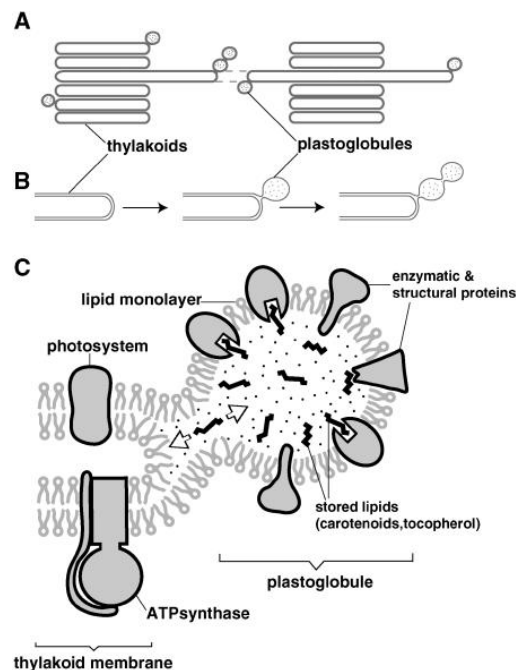
Figure 10. Illustration of the formation of plastoglobules from high-curvature areas at chloroplast thylakoids (A, B). Proteins coat the exterior of thylakoids and newly forming plastoglobules (C). Reproduced with permission, from Austin et al., 2006. Copyright © 2006, American Society of Plant Biologists (www.plantcell.org)

In addition to carotenoids, the hydrophobic core of plastoglobules may contain a range of other neutral lipids including prenylquinones (tocopherols, phylloquinone, plastoquinone), triacylglycerols and fatty acid phytyl esters (Besagni and Kessler, 2013; Lundquist et al., 2012; Piller et al., 2012). It was earlier found that the main components of plastoglobules isolated from chloroplasts of spinach and beech leaves are triacylglycerols and lipophilic prenyl quinones (Steinmüller and Tevini, 1985).

Plastoglobules in chloroplasts play a role in breakdown of carotenoids and in oxidative stress defense, while in red bell pepper chromoplasts they are active sites of carotenoid modifications (Ytterberg et al., 2006). Plastoglobules isolated from broom flowers have also been found to contain carotenoids. In this particular case, plastoglobules have been shown to contain large amounts of carotenoid esters, while triacylglycerols still constitute two-thirds of their content (Steinmüller and Tevini, 1985).

Structural proteins, together with lipid metabolic enzymes, are distributed coating the plastoglobules (Piller et al., 2012). It has been hypothesized that plastoglobules have a role in the channeling and breakdown of lipids, as they contain enzymes involved in their metabolism (Bréhélin et al., 2007). It has been suggested that in chloroplasts plastoglobules form a functional

metabolic link between the inner envelope and the thylakoid membranes (Ytterberg et al., 2006).

Plastoglobules have also been shown to actively participate in prenylquinone and other metabolic pathways (Bréhélin and Kessler, 2008). Proteins found in the plastoglobules include several members of the assumed structural fibrillin family and some metabolic enzymes, among them tocopherol cyclase (Vidi et al., 2006). Of the other proteins known to be present in the plastoglobules, the ABC1K1/3 complex has been shown to contribute in prenyl-lipid metabolism, stress response, and thylakoid remodeling (Besagni and Kessler, 2013; Lundquist et al., 2013).

1.11. Tomato research resources

As a reason for the study of tomato, its commercial importance sums on top of other features as its diploid inheritance, a relatively small genome and the easiness of propagation together with a relatively short time from planting until full-plant development. As a result, tomato has become the main model system in the study of fleshy fruits (Klee and Giovannoni, 2011). Also, tomato and its relatives have a genetic diversity that yields inheritable variation that can be exploited for a better understanding of metabolism (Lee et al., 2012).

The tomato genome has been sequenced and published as of 2012 (The Tomato Genome Consortium, 2012), though to date it is still in a phase of annotation, curation and data re-organization in the form of scaffolds and thereafter, chromosomes.

The gene scaffolds consist out of contiguous completely sequenced regions called contigs. These are arranged in an orderly way with the help of the maps of molecular markers previously completed, thus allowing for the mapping of chromosomes. All these data, along with some more preliminary material and tools to aid in their analysis, are available at the website of the International Tomato Sequencing Project (Bombarely et al., 2011). Other resources available for tomato study include the Tomato Functional Genomics Database (http://ted.bti.cornell.edu/) and the Tomato Genetics Resource Center (http://tgrc.ucdavis.edu/).

The Tomato Functional Genomics Database is a project which intends to be the most comprehensive database of tomato expression studies, particularly for micro-array studies. To that aim, it has been designed to make data transfer to the website as straightforward as possible. At its website, a numerous collection of different sets of mRNA expression data from micro-array experiments can be found. A wealth of experiments with different tissues, developmental stages and conditions of the tomato plant or fruit is made available, either in wild-type or mutant

lines (Fei et al., 2011).

Finally, the tomato researcher has the chance to resort to the Tomato Genetics Resource Center, a massive seed bank where one can find many varieties and mutants with interesting characteristics. At its website, a set of recommendations for the tomato propagator and seed harvester and collector are also included (http://tgrc.ucdavis.edu/).

OBJECTIVES

2. OBJECTIVES

The specific location of lycopene storage in tomato fruit has for decades been believed to occur at crystalline structures formed inside chromoplasts during tomato fruit ripening. However, direct proof is largely lacking and neither the definite identity nor the precise mechanism for the formation of these lycopene-accumulating structures has been reported. More recently, plastoglobules have also been proposed to play a role in carotenoid synthesis and storage.

In this framework, the primary objective of this work was aimed at the characterization of plastoglobules and other suborganellar fractions potentially involved in lycopene biosynthesis and/or accumulation in tomato fruit chromoplasts. To this end, starting with the isolation of plastoglobules, the set up of a method for the fractionation of tomato fruit chromoplasts into its components was initially planned. The obtained fractions, could eventually be characterized in detail using metabolomic and proteomic approaches.

During the course of this work it was found that fibrillins are the major proteins present in tomato fruit plastoglobules. Fibrillins belong to a heterogeneous group of largely unknown proteins. However, it has been reported that particular fibrillins may play a relevant role in plastoglobule function and carotenoid storage in chromoplasts. Based on these findings, the work has also focused on the study of the relevance of fibrillins in tomato fruit ripening. To this aim, the elucidation of the characteristics, common traits and particularities relevant to the structure and function of these proteins was also addressed using bioinformatics tools.

MATERIAL AND METHODS

3. MATERIAL AND METHODS

3.1. Isolation of tomato fruit chromoplasts

The chromoplast isolation method used in this work was an adaptation of the protocol previously developed in our group by Angaman et al. (2012). Due to the high amount of tomato fruits which was on a regular basis required for our procedures, commercial tomatoes were used in the majority of cases. We settled on the usage of an ecologic agriculture cherry tomato of the "Piccolo" variety, notorious for exhibiting a very bright and uniform red colour when ripe. The legally limited use of chemicals on ecologic agriculture expectedly would prevent artifactual variations of the fruit composition. The procedure for reproducibly obtaining our starting material (isolated chromoplasts) is reviewed below.

Tomato fruits were cleaned with de-ionized water to ensure removal of external contaminants. Fruits were subsequently sliced open and the inner slime and seeds removed. If due to the amount of material the slicing was to last for more than one hour, the material was divided into batches and every sliced batch was stored in the cold until all fruits were sliced. Sliced tomatoes were homogenized with help of a Waring blender in presence of buffer A. Optimally, for every 100 g of sliced tomato pericarp approximately 200 ml of isolation buffer A (see composition below) was used. To avoid overly plastid rupture, only three short pulses of homogenization at "low" power were applied, with thirty-second pauses between pulses. Failing to do this could result in a decrease in chromoplast yield and in damage to the later isolated chromoplasts. The composition of Buffer A was : 330 mM sorbitol, 100 mM Tris-HCl, pH 2 mM $MgCl_2$, 10 mM KCl, 8 mM EDTA, 10 mM ascorbic acid, 5 mM cysteine, 1 mM PMSF, 1 mM DTT and 1% w/v PVPP. PMSF was added immediately prior to use with strong agitation, as it is sensitive to degradation by water. PMSF was kept in a 1 M stock in DMSO until used.

The homogenate was filtered through a double layer of Miracloth (Merck) to remove debris. As the remaining solids still contained an appreciable amount of chromoplasts, these unfiltered coarse tomato remnants were recovered and re-homogenized in the blender with the addition of one volume buffer A and three pulses at low intensity. This resulted in an increased chromoplast yield. The newly homogenized suspension was filtered on the same two layers of Miracloth used before, unless those had become clogged, in which case they were replaced.

The pooled filtrate was centrifuged at 5,000xg during 10 minutes to remove contaminant

cellular components in either a JLS-16.250 rotor or a JA-10 rotor, depending on the final volume of homogenate. This centrifugation condition must not be varied, as they yield enough chromoplast precipitation and substantial contaminant removal. The obtained pellet was again resuspended, this time in isolation buffer B up to a volume of 35-40 ml, suitable for a smaller-format centrifugation. Buffer B has de composition of buffer A but excluding PVPP. Centrifugation was performed at 5,000xg for 10 minutes in a JA25.50 rotor or equivalent.

The obtained pellet was resuspended in a small volume (4-6 ml) of isolation buffer B and loaded on top of a sucrose gradient previously prepared into a SW-28 ultracentrifuge tube, including 15%, 30 %, 40 %, and 50 % sucrose layers in 50 mM Tris (pH 7.6) with 1 mM freshly added DTT. Note that the thorough resuspension is critical for ensuring purity of the obtained fractions. If resuspension is not thorough, one risks unwanted components adhering to the chromoplasts, either impurifying them or modifying the overall particle density. The latter could as well result in a certain percentage of chromoplasts (or any contaminants) not settling at the right band interphase. Centrifugation was performed at 100,000xg for 1 hour in a suitable ultra-centrifuge swing-out rotor (for instance SW-28). Note: while stopping the centrifugation it has to be made sure that the brake is in a "low" position, not to disturb the gradient while braking. The setup and result of one chromoplast ultra-centrifugation gradient is presented in Figure 12.
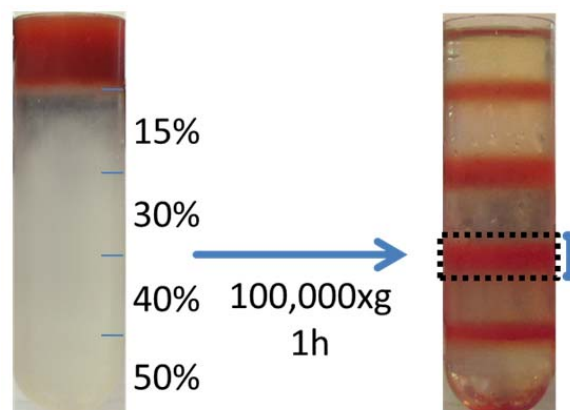
Figure 11. Sucrose gradient setup and standard result of the chromoplast isolation procedure. The concentration of the sucrose gradient blocks is shown.

Upon finishing ultra-centrifugation, the tubes were carefully extracted from the centrifuge baskets and harvesting of the bands was performed. To avoid band intermingling all the bands were harvested in a top-to-bottom way. It was deemed preferable to use a wide-pore

Pasteur pipette to avoid interface over-disturbance and to decrease the chance of inducing premature chromoplast lysis. Generally, the target harvested band was that at the 30-40% interphase (Figure 12). The bands at 40-50% and 15-30% were previously shown to contain chromoplasts in different maturation states and some impurities (Angaman et al., 2012). The harvested bands of interest were placed into clean 35-40 ml Beckman centrifuge tubes, and centrifuged at 5,000xg during 10 minutes. The supernatant was discarded, and the chromoplast pellet was gently resuspended in the appropriate buffer as required for later processing.

3.2. Fractionation of tomato fruit chromoplasts

The method set up to fractionate tomato fruit chromoplasts combined the chromoplast isolation method described above with other methods reported for chloroplast plastoglobule isolation (Vidi et al., 2006; Ytterberg et al., 2006), which were modified according to our needs. As the development of this method is a result in itself, it is described in detail in the Results section.

3.3. Carotenoid and tocopherol analysis

The samples obtained in previous procedures were weighed to assist in later calculations. By visual inspection the samples were labelled as either "concentrate", "dilute" or "intermediate". According to this, a higher or lower volume of sample was extracted, completing the difference by addition of Milli-Q water. This was inevitably subjective but kept extraction efficiencies and peak responses more comparable later on. All the procedure was carried out in dimmed light and with the samples protected from direct light. Failing to do that, reliability of the results would suffer due to an increased occurrence of isomerization and subsequent irreversible degradation of carotenoids.

Samples were put into 2-ml eppendorf tubes, with a maximum volume of 0.5 ml of sample for the more dilute samples and no less than 200 µl for the most concentrated one. All samples were brought to a final volume of 0.5 ml to conduct the extractions. Before proceeding to the extraction an internal standard was added to each extraction aliquot. This was done to prevent any negative bias due to eventual losses in any one of the samples during the extraction. To that aim, 40 µl of a 0.1 mg/ml canthaxantin (Sigma, standard grade) solution in chloroform

was added by means of a Hamilton syringe. The correct completion of this step is critical for any later comparativeness of the samples. Then, 0.5 ml of a 2:1:1 hexane:acetone:methanol solution (HAM) was added to each tube and briefly mixed with a vortex. Samples were subjected to agitation in an eppendorf multitube shaker for 10 minutes in a cold room while protecting the samples from the light. Alternatively a rotary wheel in the cold and in the dark for ten to fifteen minutes was used. Next, centrifugation was performed at 12,000xg for 3 min at 4ºC. The organic supernatant phase was harvested by means of a P200 pipette with scissor-cut shortened tips. Note: It was deemed preferable to leave some unharvested sample in the tube than to carry over any volume of aqueous phase. Also, special care was taken to protect the organic extract from the light.

To the remaining lower aqueous phase 0.5 ml of HAM solution was again added. After mixing, centrifugation was performed at 12,000xg for 3 min at 4ºC. The upper phase was harvested an added to the previously collected organic phase. A final extraction of the remaining aqueous phase with 0,5 ml of HAM was. After centrifugation as indicate above, the supernatant organic phase was harvested and added to the previously collected organic phases. Note: in the cases in which the sample still retained some coloration, additional extractions could be performed. The organic extract was evaporated to dryness, either by means of a nitrogen gas stream or using a "speedvac" fast evaporator. The latter was preferred, as it was considered advantageous for processing many samples, as well as to better protect them from the light. The sample could be stored at -80ºC if the analysis was not to be performed immediately. To store the sample safely, an inert gas atmosphere (nitrogen) was used.

For HPLC analysis, nitrogen-dried samples from the pigment extraction were used. Redissolution of only one sample at a time shortly before its HPLC injection was found to yield optimal results due to minimization of degradation. 50 µl of HPLC-grade dichloromethane were added to the dry pigment samples keeping the tube on ice. The tube was then taken to swirl it shortly and gently before returning it to ice. This was repeated until complete dissolution of the pellet. 150 µl of HPLC-grade ethylacetate containing freshly dissolved pyrogallol with a concentration of up to 0.5% w/v were added. The samples were shortly and gently swirled and put back into ice. The samples were taken with a 1-ml needle-less syringe. Note: when performing this, it is strongly recommended to gradually invert the syringe as it fills to avoid sample drop-out from the syringe. Immediately, a teflon filter was coupled to the inverted

syringe and the samples were filtered into a vial equipped with a suitable insert for accommodating the sample. To avoid teflon filter bursting and sample loss, a firm and steady pressure rather than a quick energic pressure was applied. After filtering, the vial was readily closed with a vial cap crimper and put into the HPLC-analyzer queue. This procedure was also performed in dimmed light until the sample was placed in the HPLC queue. Note: The inclusion of pyrogallol is accessory if one sample at a time is injected. However, it is strongly recommended if two or more resuspended samples are left in queue before being run.

Only freshly made or recent HPLC gradient solutions were used, always after being filtered prior to use. Internal conductions of the instrument were purge-cleaned. An equilibration of system components (column pre-filter, column) was performed and the HPLC sequence was programmed including a blank run.

HPLC program:

Solution A: 1% v/v H_2O, 99% v/v MeOH and 0.01% w/v ammonium acetate. Note: this solution must be filtered prior to use.

Solution B: tert-butyl methyl ether

Step	time	A (%)	B (%)
1	0	100	0
2	12:00	100	0
3	12:10	85	15
4	38:00:00	0	100
5	41:00:00	0	100
6	44:00:00	100	0
7	50:00:00	100	0

Table 3. Gradient program displaying the % of solutions A and B in the mobile phase during the HPLC run.
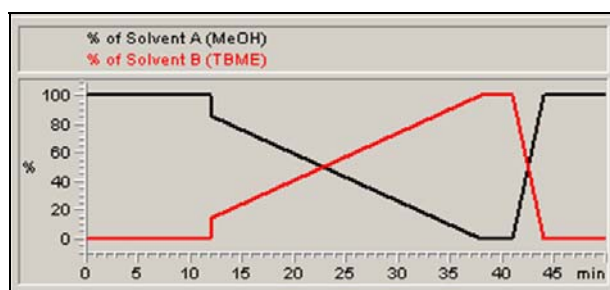
Figure 12. Visualization of the HPLC program gradient as seen in the previous table.

49

The program run utilized three isocratic flow intervals: from time 0 to time 12:00, from time 38:00 to time 41:00 and from time 44:00 to 50:00 (Table 4). The other time intervals involved gradients of varying percentage of solutions A and B. The slope during the gradient steps used was linear. When finished, chromatograms were retrieved and analyzed. Recognition of peaks was made by means of retention times and the associated measured UV-VIS spectra (Fraser et al., 2000). Peak data were manually collected from the spectra after the runs were complete.

3.4. Lipid analysis

Lipid analysis was performed in the frame of a collaboration with the group of Prof. Hubert Schaller (CNRS, Strasbourg). The procedure there followed was the following: Aliquots of 500 µl were taken from each tube of series A and B. A chloroform:methanol:water extraction was performed (Bligh and Dyer, 1959). Organic phases were collected and dried under argon stream. Samples were sealed and stored under argon at -20°C prior to analysis. Relative lipid distribution analysis was performed at their facilities by means of a UPLC/MS approach.

3.5. Electron microscopy

Sample fixation and observation procedures were performed at the Electron Microscopy Service of the PCB-UB under supervision of the resident staff.

For the fixation of samples on copper mesh supports, two kinds of staining were used, both of them involving uranyl acetate as a contrast for visualization. Chromoplast samples were infiltrated into a Spurr resin and ultrasectioned including an osmium tetra-oxide contrast. Chromoplast subfractions were fixed on three different possible kinds of supports: i) copper grid with formvar polymer coating, ii) copper grid with active charcoal coating, and iii) in-resin fixation and ultrathin sectioning (chromoplasts only). Copper mesh supports with active charcoal were predominantly used for fixation of harvested chromoplast subfraction bands. The fixation on copper grid was performed with either aqueous or methanolic uranyl acetate solutions. Resin fixation was used for intact chromoplasts.

3.5.1. Aqueous protocol (negative staining)

The manipulation of the copper grids was done with help of small, fine-tipped grid-grade EM forceps. Filter paper cuttings were used to dry the sample and to help in the manipulation. A Pasteur pipette was used to handle the required solutions, except the sample. The working area was covered with a parafilm coating, the clean side facing up. Before starting the procedure, copper grids were magnetized by means of a Glow discharge unit (BALTEC, CTA005) electro-sputtering device. Note: The magnetization keeps its effect for near 30 minutes, allowing for several grids to be treated at the same time. After that time, grids need to be re-magnetized.

Ten µl of freshly obtained sample were pipetted on the working zone and the coated side of the grid was left to float on top of it, face down, for 1 min. Excess of liquid was removed by marginal contact with Whatman paper. Beforehand, a droplet of 2% w/v uranyl acetate was pipetted on a different spot of the working area. The copper grid was deposited on top of the solution droplet, face down, and left floating for 2 min. Excess solution was removed by marginal contact with Whatman paper. Beforehand, a droplet of water was pipetted on another spot of the working zone. The grid was left to contact with the droplet, face down, for 1 min. Excess water was removed by marginal contact with Whatman paper. The grid was carefully deposited face-up on Whatman paper, in a clearly labelled four-sectored Petri dish.

3.5.2. Methanolic protocol (positive staining)

Manipulation procedures were performed likewise happened in the aqueous protocol. However, accomplishing a good timing was here more crucial for good staining.

Ten µl of freshly obtained sample were pipetted on the working zone, and the grid was left to float on top of it, face down, for 1 min. Excess liquid was removed by marginal contact with Whatman paper. Beforehand, a drop of methanolic 2% w/v uranyl acetate solution was pipetted on top of the working area, and a drop of water beside it. The copper grid was held face down on top of the 2% uranyl acetate droplet during 4 seconds. This timing had to be followed closely. Excess liquid was immediately removed by marginal contact with Whatmann paper. The grid was transferred to the water droplet for 1 min face down. Excess of liquid was removed by marginal contact with Whatmann paper. The grid was carefully deposited face-up on Whatman paper in a clearly labelled four-sectored Petri dish.

3.5.3. Inclusion of chromoplast samples in resin and ultrathin sectioning

An aliquot was taken from freshly obtained tomato fruit chromoplasts and pelleted in an eppendorf at 5,000 g for 5 min. Supernatant was discarded and 1 ml of 2 % p-formaldehyde in 0.1 M cacodylate buffer was added. Sample was gently resuspended by manual agitation on intervals. In between, the sample was kept on ice. The sample was left incubating on an orbital shaker at 4ºC. Ideally, this would last 1-2 hours, but our protocol requirements forced the sample to remain in this solution for 12-14 hours.

After this step, further treatment was performed in order to infiltrate the sample into a Spurr resin to obtain ultrathin sections. The steps followed included: sample washes with 0.1 M cacodylate buffer; post-fixation treatment with 1% w/v osmium tetraoxide in 0.1 M cacodylate buffer; sample washes with 0.1 M cacodylate buffer; sample dehydration by exposure to gradually increasing hydrosoluble organic solvent concentration, resin infiltration, and final ultrathin sectioning with a Leica MZ6 ultracut.

3.5.4. Sample observation

Sample observation was performed by means of a JEOL1010 electron transmission microscope. Images were taken at the discretion of the observer and were conditioned to what was actually found and seen. As such, they represent an image sampling of a wide landscape.

3.6. Protein analysis techniques

3.6.1. Protein quantification

Two methods have been used for protein quantification:

1) Bradford method: Prepare bovine serum albumin standards of known concentration between 0 and 1.5 mg/ml with the same buffer as the samples to be measured. Load 2 μl triplicates of each sample or standard into a 96-well ELISA well. Add 200 μl of BioRad Bradford solution (diluted 1:5 from stock). Read plate at 595 nm. Calculate standard curve and find regression curve. Remove outliers and non-linear responses. Calculate sample concentrations.

2) RC-DC method: This protocol was used for samples containing detergents or reducing agents which interfere with the Bradford reaction. It is a Löwry method-based

procedure which relies on protein precipitation to remove interferences. It is available as a kit distributed by BioRad.

3.6.2. SDS-PAGE

A BioRad casting stand for Mini Protean II mini gels was employed for preparing the SDS-PAGE gels. The needed solution of 10 % APS is prepared freshly before preparing the gel. The ingredients for the resolving gel were pipetted into a disposable falcon tube in the order stated in the table enclosed. TEMED was added only prior to pouring each respective gel, and after a good homogenization of the mixture, avoiding bubble formation. The resolving gel was mixed and poured into the prepared casting stand with the gel glasses. Isopropanol was added on top of the poured gel to ensure upper surface homogeneity, and to improve polymerization. Polymerization was allowed to proceed for over around one and a half hours. Isopropanol was removed when when polymerization was complete. Last remnants of isopropanol can be removed by capillarity using Whatman paper. Preparation of the concentrating gel, with the same precautions as taken with the resolving gel. Pouring of the concentrating gel, and placing of the lane-forming comb. Allow to polymerize for at least 45 minutes.

	Percentage of acrylamide		
	10%	13%	15%
H_2O	2900 µl	2150 µl	1650 µl
Tris-HCl 1.5 M, pH 8.8	1950 µl	1950 µl	1950 µl
SDS 5 %	100 µl	100 µl	100 µl
APS 5 %	50 µl	50 µl	50 µl
Acrylamide/Bisacrylamide 30 %	2500 µl	3250 µl	3750 µl
TEMED	3 µl	3 µl	3 µl

Table 5. Volumes of reactives for preparing a SDS-PAGE resolving gel.

	5%
H_2O	2755 µl
Tris-HCl 1.5 M, pH 8.8	500 µl
SDS 5 %	40 µl
APS 5 %	40 µl
Acrylamide/Bisacrylamide 30 %	665 µl
TEMED	4 µl

Table6. Volumes of reactives for preparing an SDS-PAGE 5% acrylamide stacking gel.

Laemmli 2x sample buffer: 0.5 M Tris-HCl pH6.8; 20 % glycerol; 4% SDS; 0.2% beta-mercaptoethanol (alternatively, 0.2 mM DTT) and a bromophenol blue crystal.

Gels were placed into a Mini protean II electrophoresis cell and fill up with 1x Running Buffer. Sample buffer was added to a final concentration of 1x to each sample and boiled for 5 minutes at 100ºC (alternatively, 10 minutes at 94 ºC). In the case of protein pellets, they were resuspended directly in 1x sample buffer. Loading of up to 15 μl of boiled sample to 0.75 mm gels, or up to 25 μl to 1.5 mm gels was performed. Molecular weight markers were loaded (Benchmark, Life Technologies; PageRuler Plus; Thermo Scientific). Empty lanes were loaded, depending on gel width, with 10-20 μl of 1x sample buffer to avoid run distortion.

Run gel at constant 12 mA amperage per every 0.75 mm of gel width running in the cell (e.g. 2x 0.75 mm gels; 25 mA). Amperage can be raised up to double once proteins have entered the resolving gel. Stop run when bromophenol blue arrives to the inferior extreme of the gel.

3.6.3. Gel staining

Coomassie brilliant blue protein gel staining:

Fixation solution: 25% isopropanol, 10% glacial acetic acid in MQ water. Staining solution: Phastgel blue R, GE Healthcare at 0.1 %. Destaining solution: 5% methanol, 7% glacial acetic acid in MQ water.

Protein was fixed in fixation solution for two hours (optional). Step recommended for thick, high-sized gels. Gel was covered with staining solution and agitated for 30 minutes to overnight. Solution was discarded or recycled, and gel was covered with destaining solution and agitated for 30-60 minutes. Solution was removed, and fresh destaining solution was added. Washes were repeated until bands appeared over a clear background. Destaining solution was replaced with MQ water. Gel could be left for some time in MQ water, preferably protected from the light and in the cold, if it is to be further used.

Silver staining - protocol (described to be compatible with mass spectrometry):

For one gel, 100 ml of each solution was prepared just prior to use. Glassware was preferred instead of plastic cuvettes. All containers were washed prior to use with MQ water and methanol.

Fixation was performed for 30 min. at room temperature or o.n. in cold with a solution containing 40 % v/v MeOH, 10% v/v glacial acetic acid, and completed to 100% with MQ water. Sensitization was performed for 30 minutes with a solution containing: 30 % v/v ethanol; 0.02 %

w/v $Na_2S_2O_3$; 68 g/l of sodium acetate. 3x5 min washed with MQ water. Staining was performed for 20 min. with 2.5 g/l $AgNO_3$. Two 1 min washes with MQ water were done. Development of the staining was done for 3-5 minutes with: 25 g/l Na_2CO_3, 400 µl/l of HCOH 37%. Arrest of the developing was done with 14.6 g/l EDTA-Na_2·2H2O. A final 3x5 min washes with MQ water were done.

3.6.4. Protein solubilization

Samples obtained from the chromoplast subfractionation protocol were used for proteomics studies. The release of proteins from the obtained membranous fractions was performed with Triton X-100, a good membrane-solubilizing detergent (Johnson, 2013).

To prevent any residual protease activity, PMSF was added to a final concentration of 1 mM. From 2% to up to a 5% w/v final concentration of Triton X-100 was added. The sample was incubated on a spinning wheel, orbital shaker or equivalent for 2 hours at 4ºC. Triton X-100 concentration was diluted back to 2% w/v with water or the same buffer as the original sample before proceeding to the de-lipidation. Note: if excessive sample dilution is inconvenient, one can add detergent only to 2% and incubate the sample for somewhat longer to avoid the need for excess detergent dilution in the final step.

3.6.5. Sample delipidization

The proportions that follow are adequate for each 1 ml of final sample volume to de-lipidize, as obtained from the previous pre-treatment. Appropriate scale-up should be applied depending on the volume to be treated.

The sample treated with detergent was transferred into a translucid centrifuge tube. Addition of 4 ml of ACS quality methanol or better was performed. Sample was agitated energically (vortex). Addition of 2 ml of ACS chloroform or better was performed, followed by energic agitation. Upon complete denaturation of protein, from this step on the sample did not require anymore to be kept on ice.

Addition of 3 ml MQ water was performed, followed by a thorough agitation with a vortex. Centrifugation was performed at room temperature for 5 minutes, 10000 g in a JS.13.1 swing-out rotor or equivalent. The aqueous supernatant was removed with a wide-pore Pasteur pipette. 3 ml methanol were added to the centrifuge tube, followed by agitation with a vortex. Centrifugation was performed at 12000 g for 15 minutes in a JA-20 rotor at room temperature.

Upon finishing, the tube was immediately inverted to remove supernatant and avoid pellet loss. Pellet was additionally washed with chilled ACS acetone or better. Centrifugation was performed at 12000 g for 15 minutes in a JA-20 rotor at 4ºC. Upon finishing, acetone was immediately removed by tube inversion. The acetone wash could be repeated if deemed necessary from visual inspection of the pellet. Clearly visible colouring of the pellet would strongly recommend repetition. The obtained pellet was dried under a stream of nitrogen gas and resuspended in 1xSDS-PAGE buffer or equivalent. Note: This protocol was also applied for the preparation of samples for 2D electrophoresis.

3.6.6. 2D-gel electrophoresis

Protein pellets were resuspendend in 8 M urea, 2 M thyourea, 2 % w/v CHAPS, 0.5 % v/v IPG buffer and 80 mM DTT. An isoelectrofocusing was performed on pH gradient strips, for running the first dimension. After that, the second dimension, consisting of an electrophoresis was run and the gel was revealed by means of silver staining. 2D electrophoresis gels were run at the proteomics service of CRAG.

Isoelectrofocusing: 1 mg protein was loaded on one of the strips for 2D electrophoresis. The specific point of placement of the protein is not critical, as the isoelectrofocusing moves the protein to its right position regardless of the starting point. Destreak solution was added (7 μl), as well as IpG buffer (3 μl), and rehydration buffer up to a final volume of 350 μl. Strip holders were assembled. pI separation program is run as follows in next table:

	Length:	Voltage
Re-hydration	0 h	-
Step 1	10 h	steady at 50 V
Step 2	1.5 h	gradient until 500 V
Step 3	1.5 h	gradient until 1000 V
Step 4	1.5 h	gradient until 2000 V
Step 5	1.5 h	gradient until 4000 V
Step 6	2.0 h	gradient until 8000 V
Step 7	6.0 h	steady at 8000 V

Table7. Isoelectrofocusing program used for 2D electrophoresis.

After completion of the program, the strips were frozen at -80ºC during at least one hour.

SDS-PAGE of 2D electrophoresis: The strip-immobilized gradients were equilibrated

with SDS-rich buffers, and were to be run in 12% SDS polyacrylamide gels. First equilibration was done for 15 minutes with a buffer containing an additional 10 mg/ml of DTT (freshly prepared). Second equilibration was performed for 15 minutes in a buffer containing an additional 25 mg/ml iodoacetamide (freshly prepared). Wash of the strips was done with MQ water. Afterwards, strips were assembled with the 12 % SDS-PAGE gels (18x18 cm; x6 gels)

	Volume
MQ water	153 ml
Duracryl	180 ml
Lower buffer	117 ml
APS	2.25 ml
TEMED	380 µl

Table 8. Reagent volumes for the preparation of six 12% SDS-PAGE Duracryl gels for 2D electrophoresis.

5 l of running buffer 1x were prepared; 2 l of running buffer 2x; pouring into the respective running cell chambers.

	Time	Power
Step 1	30 min	15 W
Step 2	4 h	100 w

Table 9. 2D electrophoresis SDS-PAGE program.

Gels were soaked into blocking solution (silver staining protocol) upon completion of run. Gel spots were excised out of the gel by means of scissor-shortened pipette tips. The excised gel spot is pushed out by inserting an intact tip from the rear. The excised fragment is put into an eppendorf containing a small amount of MQ water (50 ul).

The excised fragments were submitted to the Proteomics Service of the PCB-UB. The samples were exposed to in-gel tryptic digestion. A liquid chromatography by means of a nanoAcquity chromatograph was performed on the obtained peptides, which were introduced into an OrbiTrap-LTQ Velos MS/MS analysis instrument. The separated peptides were detected, giving in return the corresponding list of "hits", according to the database by us used.

3.6.7. In-solution protein digestion and analysis by LC-MS/MS

Proteins were digested by means of porcine trypsin. To 5 µg o protein 30µL of 50 mM ammonium bicarbonate were added. Proteins were reduced and alkylated with 50mM DTT incubation for 30 min at 30ºC, and a later treatment with 150 mM iodine acetamide at 30ºC for

30 min. An extra 1μL of the DTT solution was added to inactivate the remaining iodine acetamide (quenching). Proteins were digested overnight at 37ºC with 100 ng porcine trypsin. Sample was desalted with a Proxeon C18 tip. Elution was done by addition of 40 ul of a 70% acetonitrile, 0.1 % trifluoroacetic acid. Sample drying was performed by means of a SpeedVac.

Samples were analyzed by means of a liquid nanoAcquity chromatograph (Waters) coupled to an Orbitrap-Velos mass spectrometer (Thermo Scientific). Aliquots equivalent to 200 ng of protein sample were resuspended in a 1% formic acid solution. Each sample was injected for separation onto a reverse phase C18 column (75 μm Øi, 10 cm, nano Acquity, 1.7μm BEH column, Waters). Samples were injected consecutively with three blank injections between samples. The employed gradient for the separation was of 2% to 40 % of Solution B in 60 minutes,, followed by a 40% to 60% of B in 5 minutes with a flow of 250 nl/min (Solution A: 0.1% formic acid; Solution B: 0.1% formic acid in acetonitrile). Eluted peptides were ionized by means of a metalized silica tip (PicoTip TM, New Objective). The voltage applied to the tip was of approximately 2000V.

3.6.8. Data analysis and database search

The masses of peptides (m/z 350-1700) were measured in "Full Scan MS" at the Orbitrap with a resolution of 60,000 FWHM at 400m/z. Up to 10 of the most abundant peptides (minimum intensity of 500 counts) were selected for each analysis to be fragmented in the ionic trap (CID) with nitrogen as a collision gas with a normalized collision energy of 38%. Data were acquired with the Thermo Xcalibur software (v.2.1.0.1140) in raw data format.

The program Proteome Discoverer (v.1.3.0.339) was used to search from the raw data into the in-house MASCOT search engine. The search parameters were: Database/Taxonomy; MASCOT NCBInr Eukaryota (v. February 2012); Solyc_cRAP; Trembl_SP_viridiplantae_Crap (18/05/2012); missed cleavage: 2; fixed modifications: carbamidomethyl of cystein; variable modifications: oxidation of methionine and pyro-Glu (N-term Glutamine); peptide tolerance: 10 ppm and 0.6 Da (respectively for MS and MS/MS spectra); percolator: FDR strict 0.01; FDR relaxed : 0.05, validation based on q-value. To discriminate between correct and incorrect assignations the Percolator node is used, which allows for the augmentation of the number of correctly identified peptides, if given a specific false discovery rate.

As it was found that the standard available database used for plant proteomics determinations seemed to be poor in tomato proteins, a new database was created, which was

named Solyc_cRAP. This was done on the basis of ITAG 2.3, which contains the proteins inferred from the genomic annotations of the complete Tomato Genome Project, as of may 2011. The ITAG 2.3 FASTA protein list was retrieved from: ftp://ftp.solgenomics.net/tomato_genome/annotation/ITAG2.3_release/ITAG2.3_proteins.fasta

Common contaminant sequences were also added to this database, from: ftp://ftp.thegpm.org/fasta/cRAP.

A second database was created starting from the public data available at Uniprot-SwissProt Plants (18-05-2012), Uniprot-TrEMBL Plants (18-05-2012) and cRAP (ftp://ftp.thegpm.org/fasta/cRAP). This database was called Trembl_SP_viridiplantae_Crap.

3.6.9. Protein "hit" filtering

The accepted identified list of protein hits was subject to several additional filtering levels to discard cross-contamination from other organelles or cell components. Firstly, comparison of the protein SolGenomics identifier with the updated published tomato chromoplast proteome data (Barsan et al., 2012). Secondly, retrieval of the list of tomato plastid-encoded proteins and comparison with our "hits", and identification of "hits" present in the list. Thirdly, systematic BLAST launches against the TAIR database in search for evidence on plastid localization. Additional corroboration performed with SUBA (Heazlewood et al., 2007), PLPROT (Kleffmann et al., 2006), and PPDB databases (Sun et al., 2009) when in doubt. Any provided experimental evidence was given decisory priority. Fourthly, proteins not meeting clearly any of the three previous criteria were subject to seven different subcellular location prediction bioinformatic tools: Bacello, ChloroP, mGOASVM, plantLOC, WOLF pSORT, ESLpred and TargetP (Bhasin and Raghava, 2004; Emanuelsson et al., 1999, 2007; Horton et al., 2007; Pierleoni et al., 2006; Tang et al., 2013; Wan et al., 2012).

If a majority of them would predict the protein to be plastidic, the protein would be salvaged from discarding. If some ambiguity remained, observations from the previous database search would be considered in order to untie the draw in one sense or the other.

3.6.10. Gene ontology (GO) descriptor retrieval and processing

A *.FASTA file was prepared including all of the accepted protein hits. The *.FASTA file was imported into the BLAST2GO online tool v.2.7.0 (Biobam Bioinformatics, S.L.; http://www.blast2go.com/b2ghome). Sequences were sequentially launched in BLAST against

the general NCBI non-redundant (nr) database. From the obtained BLAST hits, we obtained the corresponding GO descriptors from the six most closely related, annotated proteins. Upon final processing and obtention of GO descriptors, the Arabidopsis "GO slim" was applied from inside the BLAST2GO program. This had as a result the removal of redundancies among the descriptors associated to each of query proteins.

A general GO bar diagram was elaborated for each of the heavy fractions III, IV, V and VI, recopilating and describind all their relevant slimmed-down GO descriptors. Subsequently, a Venn diagram was elaborated by means of the VENNY online tool (http://bioinfogp.cnb.csic.es/tools/venny/) to aid in the visualization and identification of common and unique proteins of each of the fractions. The proteins obtained, along with their associated descriptors, were thus classified according to their presence in the different studied fractions III, IV, V or VI.

After the finding of the identity of the common-to-all and the proteins unique to each fraction, graphs were drawn with BLAST2GO for these common proteins, and for the unique proteíns. A manual classification of the "unique" and "common" proteins was as well performed on GO data exported from BLAST2GO, with the help of the VENNY program. Data were graphically represented, according to an arbitrarily decided division of proteins into diverse functional subgroups (from their descriptions). Representations of these data were made for the following: Unique proteins to any one of the fractions III, IV, V or VI; and proteins common to all of the fractions III, IV, V and VI.

The categories or subgroups manually assigned were: ATP/energy-related, stress/defense response, sugar/carbon metabolism, chaperones, protein translation-related proteins, redox homeostasis, signalling/regulation, amino acid biosynthesis, DNA/RNA processing, protease activity, transport activity (excluding ADP/ATP), lipid-related proteins, other proteins (diverse) and proteins of unknown function.

3.7. RNA related techniques

3.7.1. RNA extraction from tomato fruit

All work with RNA was performed with double-autoclaved material to remove risk or RNAse contaminants.

Tomato tissue was frozen with liquid nitrogen and grind with mortar and pestle. 1 g of tissue was reaped in a 13-ml centrifuge tube. It was resuspended in 4 ml of extraction solution (100 mM Tris-HCl pH 9.0, 200 mM NaCl, 15 mM EDTA, 0.5 % sarkosyl in DEPC treated water and 100 μl of 2-beta-mercaptoethanol) followed by a vigorous agitation. Additionally, 4 ml phenol (saturated with Tris-HCl pH 7.9) were added, followed by vigorous shaking. Addittion of 0.8 ml chloroform was performed, followed by vigorous shaking. 280 μl of 3M sodium acetate pH 5.5 were added, followed by vigorous shaking. Incubation on ice for 15 minutes was performed. A centrifugation for 20 min. at 9000 g was performed at 4ºC, using suitable adaptors. The upper supernatant was collected into a 13-ml Sarstedt tube containing phenol:chloroform:isoamyl alcohol (25:24:1), followed by vigorous agitation. The tube was incubated on ice for 15 minutes. Centrifugation for 20 min. at 9000 g at 4ºC was performed. The upper supernatant was transferred to a fresh 13-ml Sarstedt tube, and one volume of isopropanol was added. RNA was allowed to sediment at -20ºC for 30 minutes to overnight. Centrifugation was performed for 20 min. at 9000 g, at 4ºC, and the supernatant was discarded. Resuspension of the sediment was done in 1 ml 80% ethanol in DEPC water, and transferred to an eppendorf tube. Centrifugation was performed for 10 min. at 12000g, discarding the supernatant, and drying the sediment under a nitrogen gas stream. Resuspension of the sediment was performed in 1 ml of DEPC water, followed by centrifugation for 5 min at 12000g to remove insoluble components. The supernatant was transferred to a new eppendorf tube and 0.5 ml of lithium chloride was added. Precipitatation of RNA at 4ºC for 3 h to overnight was performed. Precipitated RNA was centrifuged for 10 minutes at 12000g. Supernatant was discarded and pellet was cleaned with 80% ethanol in DEPC water. Centrifugation was done for 5 min. at 12000g, discarding the supernatant and drying the sediment under a nitrogen gas stream. Final resuspension of the pellet was done by means of 100 μl of DEPC water. The resuspended RNA was quantified in the Nanodrop, and the purity of RNA was confirmed in an agarose gel with buffer prepared in DEPC water, having previously cleaned the cuvette with ethanol.

3.7.2. cDNA generation from RNA

Isolated RNA was processed with the Ambion RETROscript kit. The protocol used involved the de-naturalization and the use of the Oligo dT primer. In case of cDNA purity being critical, it is recommended to perform an additional treatment with DNAse using the Qiagen RNAsse Free DNAsse Set kit.

3.7.3. Primer design for qRT-PCR

Suitable pairs of primers were designed following a series of rules aimed at enhancing replication efficiency of PCR and to avoid parasite amplification reactions. Additionally, an adequate pair of primers for a housekeeping gene was chosen on normalization grounds.

The requirements to be taken into account for the design of good-performing qPCR primers were the following: pairing rather towards the 3' end (often more specific); if possible, pair primers separated by an exon-exon boundary, as it reduces genomic background; amplified region between 80 and 200 nucleotides; GC content: 50-60%; length of the primers should be between 18 and 24 nucleotides, their melting temperature between 58 and 63 ºC; self-complementarity and cross-complementarity should be avoided.

As the sequences to be amplified limit our choice of primers, sometimes a compromise solution was unavoidable. Taking these rules into account, one can use the online website PrimerQuest to obtain an initial set of primers, which can be checked for possible optimization. The PrimerBLAST tool at the NCBI website was also used as an orientation, as well as to identify the possible exon-exon boundaries. After obtaining the primers, their efficiency was tested, and a calibration curve for the qPCR reactions was performed to find out the adequate dilution of cDNA for the reactions. After the primers passing the quality test, independent experimental triplicates of each fibrillin to be tested were performed.

3.7.4. qRT-PCR

The Roche LightCycler 480 instrument was used, under the standard protocol specified by the manufacturer. 1:25 dilutions of the cDNA were used for the qPCR experiments (after calibration curve to check optimal response). Three technical replicas were performed for each sample to be measured. Normalization of measurements was performed with the clathrin adaptor complex (CAC) housekeeping gene (Chandna et al., 2012), either against small green tomato fruit, or tomato leaves. Quantification and data analysis was performed on the same Roche Light Cycler 480 instrument, and exported for further processing in a standard spreadsheet.

3.8. Virus induced gene silencing (VIGS)

A preculture of the Agrobacterium strain was grown for 32-48 hours at 28ºC for each of the constructions to be tested. In parallel, the helper construct pTRV1 was also grown and

follows the same steps. Inoculation of a second preculture (5 ml) was done from the first one, and grown for 16-20 hours at 28ºC. A centrifugation was performed for 15 min. at 5000 g, and agrobacteria were resuspended in agroinfiltration medium supplemented with 200 μM acetosyringone (from a stock in DMSO), to a final DO600 between 0.1 and 0.2. Cells were incubated for 4 hours at room temperature in gentle agitation, and protected from the light. The agrobacteria of the constructs of interest were mixed in a 1:1 proportion with agrobacteria containing the helper vector pTRV1. This implies that the final concentration of bacteria containing the silencing construct was half the previously measured DO600 (thus between 0.05 and 0.1). Finally, by means of a syringe with an insulin needle, the constructs were injected into the carpopodium of mature *DR* green tomato fruits, *in planta*.

3.9. DNA related techniques

3.9.1. Amplification of selected DNA fragments using PCR

During the design of primer pairs for amplification of a gene or gene fragment, a series of factors was needed to be taken into account. It was tried for primers to have as low a secondary structure-forming tendency, and as low a dimer-forming tendency as possible. As well, it was tried for forward and reverse primers of a given PCR reaction to have as similar melting temperatures as possible. The annealing temperatures used were chosen to be 5 degrees lower than the average melting temperatures of the primers used.

The *Taq* polymerase used was the Green Taq Master Mix, provided by Promega. This is a pre-mixed preparation lacking only the primers and the DNA to be amplified. The usual reaction conditions involved: 3 min. initial denaturation; 1 min. denaturation per cycle; 1 min. 30 sec. annealing; 1 min/kB extension time. In the case of using *Pfu* polymerase, the denaturation time per cycle was cut down to only 30 sec.

3.9.2. DNA purification from agarose gels

The bands of interest were excised by means of a scalpel blade, and put into eppendorf tubes. The Macherey-Nagel Nucleospin Extract II kit was used for the purification itself.

3.9.3. DNA digestion with restriction enzymes

The used restriction enzymes were either from Fermentas (ThermoScientific) or

Promega. The usual setup of a digestion involved adding:

1 µg DNA; 2 µl 10x Buffer; 3-5 u of enzyme 1; 3-5 u of enzyme 2; water was added to a final volume of 20 µl. This was incubated for 2-3 hours, generally at 37ºC (except SmaI, at 25ºC). Generated fragments were run and analyzed in agarose gels.

3.9.4. Gateway cloning

BP reactions were performed with the Gateway BP Clonase II mix of Life Technologies LR reactions were performed with the Life Technologies Gateway LR Clonase II Enzyme mix. All of these reactions were performed according to the instructions provided by the manufacturer. The vectors used for VIGS cloning were pDONR207c to generate the entry vector via BP reaction, and the resulting pENTRY vector was ligated with pTRV2, the destination vector.

3.9.5. Transformation of E.coli cells by thermal shock

50 ng of DNA were added to 10 µl of competent cells and kept for 30 minutes on ice. A thermal shock was performed by exposing the cells at 37ºC for 1 minute and 40 seconds in a thermal block. Cells were transferred to ice for ten minutes. 1 ml of LB was added, shaking for 1 h at 37ºC. Cells were plated in an LB/agar plate with the appropriate selection antibiotics. Plates were incubated overnight at 37ºC. Colonies were picked for further processing.

3.9.6. Preparation of electrocompetent *Agrobacterium tumefaciens* competent cells

An Agrobacterium preculture was grown in 4 ml of YEB medium with the adequate antibiotics, preferably in a 50-ml falcon to increase effectiveness of aeration. 100 ml of antibiotic-supplemented YEB was inoculated with 1 ml of the previous freshly grown preculture. Cells were grown for 9-12 h at 28ºC up to an optical density of 0.4-0.5 at 600 nm. Cells were incubated on ice for 15-30 minutes. Centrifugation at 2500 g was performed in a sterile 35-ml Beckman tube for 15 min. at 4ºC, removing supernatant upon finish. Pellet was resuspended in 10 ml of sterile MQ water. Centrifugation at 2500 g for 15 min, at 4ºC was performed, removing the supernatant upon finish. Sediment was resuspended with 1 ml of sterile glycerol 10% v/v. 40 µl-aliquots were frozen into liquid nitrogen and stored at -80ºC until required.

3.9.7. Transformation of competent *Agrobacterium tumefaciens* electrocompetent cells

50 ng of DNA was introduced into a 40 µl aliquot of electrocompetent cells. Note: it is preferrable to avoid adding more than 1 µl DNA, due to the negative effects of high salt presence on electroporation. 1 ml of YEB in ice was pre-cooled on ice. Cells were transferred into the electroporation cuvette, and electroporation was performed at 2kV, with 25 µF capacitance. Note: good electroporation times generally are between 3 and 5 picoseconds. Upon finish, cells are immediately transferred into 1 ml of pre-cooled YEB; and kept on ice for two minutes. Cells were agitated at 28ºC for 2-4 hours. Cells were plated on YEB/agar plates supplemented with the adequate antibiotics, incubating plates at 28ºC for 40 h. Picking of colonies for further processing was performed as required.

3.9.8. MiniPrep protocol for plasmidic DNA preparation

The Macherey-Nagel plasmid purification kit NucleospinPlasmid was employed for plasmid purifications.

3.9.9. Agarose gel electrophoresis

TAE 50x Buffer (1 l): 57.1 ml glacial acetic acid, 242 g Tris, 100 ml EDTA (pH 8), with final adjustment of pH to 8.2.

Nucleic acid 6x loading buffer: 0.25 % bromophenol blue; 30% glycerol in H_2O.

Procedure: 0.5 g agarose was dissolved in 50 ml of 1x TAE buffer heating with consecutive microwave-pulses pausing in-between to avoid over-boiling. Upon complete dissolution, pouring of dissolved agarose was done into a DNA-electrophoresis gel tray. Cooling down was allowed until visible evaporation stopped. 5µl of 10 mg/ml ethidium bromide were added, stirring well and placing a suitable comb over the tray to form the sample loading wells. Gel was allowed to solidify for around 30 minutes. Placing the solidified gel into electrophoresis cell, 1x TAE buffer was poured until the gel was completely covered. Loading of samples, previously having added the suitable volume of 6x loading buffer was performed, loading a suitable molecular weight marker (0.1 Kb ladder, or 1 Kb ladder, Biotools). Running of the gel was performed at 70-100 V.

3.10. Bacterial strains used and growth condition

<u>Bacterial strains used:</u>

Escherichia coli DH5-alpha

Agrobacterium tumefaciens GV3101

<u>*Escherichia coli* growing medium composition:</u>

Liquid, Luria Broth (LB) medium: 5 g of NaCl, 5 g of yeast extract and 10 g of bacteriologic tryptone, dissolved in de-ionized water, and autoclaved at 121 ºC for 20 minutes.

Solid, Luria Broth medium: 1 l of non-autoclaved liquid LB was complemented with 1.5% w/v agar. Autoclave for 20 min. at 121ºC.

Antibiotic concentrations of use: These are the usual concentrations utilized. These antibiotics are added at room temperature for liquid cultures, and at around 50ºC for solid cultures; ampicillin (sodium salt), brought to a final concentration of 100 μg/ml from a stock at 100 mg/ml; kanamycin, brought to a final concentration of 50 μg/ml from a stock of 50 mg/ml; cloranfenicol, brought to a final concentration of 17 μg/ml, from a stock of 34 mg/ml.

<u>*Agrobacterium tumefaciens* growth medium:</u>

Liquid, YEB (Yeast Extract Broth): 5 g of beef extract 1 g of yeast extract, 5 g of bacteriological peptone and 5 g of sucrose in 1 l deionized water. Autoclave at 121ºC for 20 minutes.

Solid, YEB: 1 l of non-autoclaved YEB was supplemented with a 1.5 % w/v of agar. Autoclave at 121ºC for 20 minutes.

Antibiotics used: Ampicillin (sodium salt), brought to 50 μg/ml from a stock of 100 mg/ml; kanamycin, brought to a 100 μg/ml concentration from a stock of 50 mg/ml; rifampicin, brought to a 150 μg/ml concentration from a stock of 100 mg/ml.

3.11. In-silico protein studies

Upon the finding of the relevance of the fibrillin family in the accumulation of carotenoids, and the largely unknownness of this family, it was decided to study them at a functional, regulatory and computational level. The *in-silico* part of the fibrillin characterization

was run mainly with a cluster of different publicly downloadable tools and available online servers and databases.

3.11.1. Databases used

The comparative genomics website CoGe (http://genomevolution.org/CoGe/) had a listing of complete genomes to date, which helped initially find the relevant resources. Databases consulted during the sequence research are listed next. The databases used more relevant to our work were: NCBI (http://www.ncbi.nlm.nih.gov/), EXPASY(http://www.expasy.org/), PHYTOZOME (http://www.phytozome.net/), JGI genome project (Genome Portal of the Department of Energy Joint Genome Institute; http://genome.jgi.doe.gov/) and Solgenomics (http://www.solgenomics.net).

Other resources consulted were: Amborella Gene Project (http://amborella.huck.psu.edu/project), Cannabis sativa Genome Browser Gateway (http://genome.ccbr.utoronto.ca/cgi-bin/hgGateway), Cenicafe bioinformatics (http://bioinformatics.cenicafe.org/), Cocoa Genome Project (http://www.cacaogenomedb.org/), CoGe; Comparative Genomics (http://genomevolution.org/CoGe/), Cucurbit Genomics Database (http://www.icugi.org/cgi-bin/ICuGI/index.cgi), Cyanidioschyzon Merolae Genome Project (http://merolae.biol.s.u-tokyo.ac.jp/), Genome Database for Rosaceae (http://www.rosaceae.org/), Gramene (http://www.gramene.org/), Legume Information System (http://www.comparative-legumes.org/), MaizeGDB (http://www.maizegdb.org/), Medicago Genome Project (http://www.jcvi.org/medicago/), Melonomics (https://melonomics.net/), Nanochloropsis Genome portal (http://www.nannochloropsis.org/), ORCAE (http://bioinformatics.psb.ugent.be/orcae/), PlantGDB (http://www.plantgdb.org/); Rosaceae.org, The banana genome hub (http://www. banana-genome.cirad.fr), Thellungiella.org.

3.11.2. Sequence retrieval

Fibrillin sequences of non-Arabidopsis species were retrieved using as reference the Arabidopsis sequences found in the paper by Singh and coworkers (Singh and McNellis, 2011). The reference sequence accessions are indicated next:

Accessions of the reference sequences:

FBN1a, FBN1b: At4g04020, At4g22240

FBN2: At2g35490

FBN3a, FBN3b: At3g26070, At3g26080

FBN4: At3g23400

FBN5: At5g09820

FBN6: At5g19940

FBN7a, FBN7b: At3g58010, At2g42130

FBN8: At2g46910

FBN9: At4g00030

FBN10: At1g51110

FBN11: At5g53450

Additional, added by us, FBN12/FBN-like protein: AT1G18060

All fibrillin sequences from other species were found using the Arabidopsis sequences in BLAST against the appropriate databases. The search of homologues was hierarchized according to the following database priorization:

NCBI > Expasy > Solgenomics (for tomato) > Phytozome > JGI > Other databases

Whenever the homologue in search was not found for a given species in a given database, it was resorted to the next level of this database hyerarchization.

3.11.3. Analysis of tomato average protein composition

The tomato proteome derived from the translated sequenced genome was taken from the ITAG 2.3 file found at the ftp site of Solgenomics.net was retrieved. It is available at

ftp://ftp.solgenomics.net/tomato_genome/annotation/ITAG2.3_release/ITAG2.3_proteins_full_desc.fasta

This file was stripped bare to leave only the amino acid residues remaining. By means of the Microsoft EXCEL spreadsheet, the number and percentage of each residue was calculated.

3.11.4. Sequence analysis tools used:

3.11.4.1. Alignment tools

The MUSCLE alignment algorithm was used (Edgar, 2004), in its implementation available in the MEGA 5.2 sequence analysis package (Tamura et al., 2011). Small-scale alignments for quick sequence checkings were performed with the online server MULTALIN (Corpet, 1988).

MEGA is available at: ; http://www.megasoftware.net/

MULTALIN is available at: http://www. multalin.toulouse.inra.fr/multalin/

3.11.4.2. Consensus elaboration tools

The ACM Advanced Consensus maker at the HIV Databases website was used to generate a basic scaffold with the more conserved amino acids for each protein alignment.

The VISCOSE consensus making tool (Spitzer et al., 2004) was used to visualize conflictive consensus sites. It was used to help decide how to complete the previously generated consensus amino acid scaffold.

ACM is found at: http://www.hiv.lanl.gov/content/sequence/CONSENSUS/AdvCon.html

VISCOSE is available at: http://bio.math-inf.uni-greifswald.de/viscose/

To reduce as far as possible the impact of variable subjectivities, a number of rules were imposed upon deciding the final consensus of each fibrillin:

1.- The consensus for any aligned column with 80% of equal amino acids is this major amino acid.

2.- The consensus for any aligned column with 80% of amino acids with similar chemical properties (hydrophobicity, polarity, positive charge, negative charge, aromaticity, size) is the major amino acid with that property in that column.

3.- In case of conflict, other conflictive spots are searched in the vicinity of the sequence. If the disagreement is complementary in sequence or in chemical property, decision on one consensus site is considered to be linked to the decision on the other site.

4.- In case of persisting ambiguity, global statistical composition of conflictive amino acids is taken into account. Several criteria would be considered: preferred chemical properties at the conflict site, compared situation of other conflict sites and the global % composition of all amino acids from all involved conflict sites. Calculation of the occurrence of each amino acid in conflict sites is taken into account. Statistically minoritary amino acids are kept out when possible.

5.- An additional guideline used was the calculation of the % presence of each amino acid in the whole conserved zone of the aligned sequences, and comparison with the uncomplete consensus.

The very first preliminary consensus elaboration trials were done submitting the alignments to the BOXSHADE server (http://www.ch.embnet.org/software/BOX_form.html). However, this was not the final tool of choice.

3.11.4.3. Phylogenetic analysis

Phylogenetic trees were drawn by means of the maximum likelihood approach (Goldman, 1990), in the implementation included in the MEGA 5.2 package, by the Jones-Taylor-Thornton model (JTT), using the alignments obtained with the MUSCLE algorythm, using the default parameters.

3.11.4.4. Protein property predictions

ChloroP was used to predict the presence and length of plastid targeting sequences (Emanuelsson et al., 1999).

PROTPARAM was used to calculate common physico-chemical properties of the proteins in study which are directly related with the primary sequence composition (Walker, 2005). This tool is available at: http://web.expasy.org/protparam/

PROTSCALE was used to plot hydropathicity graphs, using a window size of 21 amino acids (Walker, 2005). This tool is available at: http://web.expasy.org/protscale/

3.11.4.5. Protein structure prediction and analysis

JPRED3 was used to predict the secondary structure of the fibrillins of interest (Cole et al., 2008). Available at: http://www.compbio.dundee.ac.uk/www-jpred/

TMHMM was used to predict the possibility for transmembrane alpha-helices (Krogh et al., 2001). Available at: http://www.cbs.dtu.dk/services/TMHMM/

HELIQUEST was used to analyze the characteristics of each of the found alpha-helices, using the "full" window size setting (Gautier et al., 2008). Available at: http://heliquest.ipmc.cnrs.fr/

BOCTOPUS was used for analyzing the possibility of transmembrane beta-barrel formation (Hayat and Elofsson, 2012). Available at: http://boctopus.cbr.su.se/

Prediction of intrinsical disorder for the proteins of interest was performed by means of the Genesilico Metadisorder Service (Kozlowski and Bujnicki, 2012). Available at: http://genesilico.pl/metadisorder/

Tertiary protein structure for the consensus sequences was predicted by means of the Robetta Ab-initio online server, versions 3.0 and 3.5 (Kim et al., 2004). Other modeling servers/applications which were tried include: QUARK, I-TASSER, MUSTER, BHAGEERATH, GALAXY, PHYRE2 and CABS-FOLD. All of them were outperformed by the Robetta server in our conditions, found at http://www.robetta.org.

Quality of models obtained was generally checked using the QMEAN online tool (Benkert et al., 2009), found at http://swissmodel.expasy.org/qmean/cgi/index.cgi. Additional model anomaly detection tools checked were the WHAT IF server (Hooft et al., 1996; Vriend, 1990) and WHAT CHECK server (Rodriguez et al., 1998).

Observation of the obtained 3D models was performed using UCSF Chimera version 1.8 (Pettersen et al., 2004)., whose homepage is found here: https://www.cgl.ucsf.edu/Chimera/

3D analysis of distribution of conserved residues for 3D models was performed by means of the ConSurf server (Celniker et al., 2013) located at: http://consurf.tau.ac.il/

Later observation was done either by UCSF Chimera or Prof. Eric Martz's Firstglance in Jmol tool (http://bioinformatics.org/firstglance/fgij/), which is directly linked from ConSurf's result page. The parameters used for ConSurf were: Multiple Sequence Alignment built using MAFFT; homologues collected from CLEAN_UNIPROT database; homolog search algorithm used was BLAST; PSI-BLAST E-value was 0.0001; number of PSI-BLAST Iterations was 5; maximal %ID Between Sequences: 95; minimal %ID For Homologues: 35; max. Number of Homologues: 100.

The retrieval of carotenoid pdb structures for size comparison with predicted fibrillin beta-barrels was done from the GNU-Darwin Carotene Molecules Structural Archive and Gallery (http://molecules.gnu-darwin.org/mod/carotene-more.html). Many of the structures there found were not on-scale, thus it had to be checked if the scale was correct. USCF Chimera was used, and the correct size was assessed by comparison of the size of a tyrosine side-chain with a cycle of six members at the models of interest.

RESULTS

4. RESULTS

4.1. CHROMOPLAST SUBFRACTIONATION

The primary objective of this work, the setup of a method for tomato fruit plastoglobule isolation, had its foundation on the hypothesis that these particular plastidial structures could be playing a relevant role in carotenoid accumulation during tomato fruit ripening (Bréhélin and Kessler, 2008). It was consequently decided to isolate and characterize tomato fruit plastoglobules in order to elucidate their contribution in lycopene accumulation. To this aim, a preexisting method in our group for the isolation of tomato fruit chromoplasts (Angaman et al., 2012) was used as a first step to adapt methods for plastoglobule isolation previously reported in other plant species (Vidi et al., 2006; Ytterberg et al., 2006). During the development of the protocols for tomato fruit plastoglobule isolation it was observed that lycopene was strongly present in other chromoplast subfractions. Because of this, the global objective was broadened by including the isolation and characterization of other chromoplast subfractions, with a special interest on those containing carotenoids. As a result, a wide number of trials were made until the final method was settled upon.

As indicated above, the starting point of the method of chromoplast subfractionation was the availability of a robust method for the isolation of tomato fruit chromoplasts (Angaman et al., 2012). This protocol was to be coupled to the two *Arabidopsis* plastoglobule obtention methods described by Vidi et al. (2006) and Ytterberg et al. (2006). The adaptation of these procedures, based on the use of sucrose gradient ultracentrifugation, suffered from the finding of several hurdles at the technical level. Furthermore, the use of different centrifugation conditions and disruption methods in the protocols taken as reference (Vidi et al., 2006; Ytterberg et al., 2006), led to some initial unexpected results which suggested that chromoplast plastoglobules have different properties than those present in chloroplasts.

Even before solving these problems, and yet during the initial trials of plastoglobule isolation, it strongly drew our attention that a lycopene-rich pellet was obtained and discarded prior to the ultracentrifugation used for plastoglobule isolation. Being one of our aims the identification of the lycopene-storage "sink" elements, it was decided to pursue also the fractionation of the lycopene-rich pellet into its subcomponents. After many trials involving different ultracentrifugation conditions and sucrose gradient setups, the use of different tomato

varieties and fruits at different ripening stages, a near-final protocol was settled for. Under these conditions, plastoglobules were separated along with a series of subfractions of different densities. The final protocol for fractionation of tomato fruit chromoplasts is outlined below. Unless otherwise stated steps were performed at 4ºC.

Experimental protocol used for tomato fruit chromoplast subfractionation:

Step 1. Chromoplasts purification. Chromoplasts were purified as describe by Angaman et al. (2012). Depending on the aim of the extraction, between 0.5 and 2 kg of "Piccolo" or standard Cherry tomato fruit were processed to yield somewhere between 0.3 and 1.2 kg of tomato fruit pericarp. For an experiment aimed at metabolite profiling, 0.5 kg of tomato fruit can suffice, while for proteomic analysis of plastoglobules, no less than 1 kg g should be used. Chromoplasts were centrifuged and resuspended in 5 ml of fractionation buffer (50 mM HEPES pH 8; 5 mM $MgCl_2$) containing 17 % sucrose.

Step 2. Chromoplast lysis. Chromoplasts were lysed using sonication. Osmotic shock and the use of a Dounce homogenizer were assayed as alternative lysis methods without success. Immediately before sonication, PMSF was added to a final concentration of 1 mM. The sonication itself ensures the complete mixing and quick dissolution of the added PMSF. Sonication was performed with a suitable tip applying three pulses of 30 sec, with pauses of 3 sec in between. The tube containing the chromoplasts was kept on ice. Sonication conditions was found to influence the final result obtained upon ultracentrifugation (next step).

Step 3. Ultracentrifugation in sucrose gradients. The sonicated sample was subjected to centrifugation at 5,000xg for 10 minutes to sediment remnant chromoplasts which would interfere with any later interpretations. The supernatant was loaded on top of a gradient containing 20%/30%/40%/50% sucrose blocks in 50 mM HEPES pH8, 5 mM $MgCl_2$ and 1 mM freshly added DTT. Two final layers of 7% and 0% sucrose in the same buffer are added on top of the sonicated fraction. Ultracentrifugation was performed for 3 hours at 220,000xg in a swing-out ultracentrifuge rotor (Beckman SW40). Longer centrifugation times result in poorer band resolution. A shorter centrifugation at higher speed, if local equipment allows it, would be optimal.

An overview of the different steps used tomato fruit chromoplast subfractionation and the typical pattern of bands obtained is shown in figure 13.
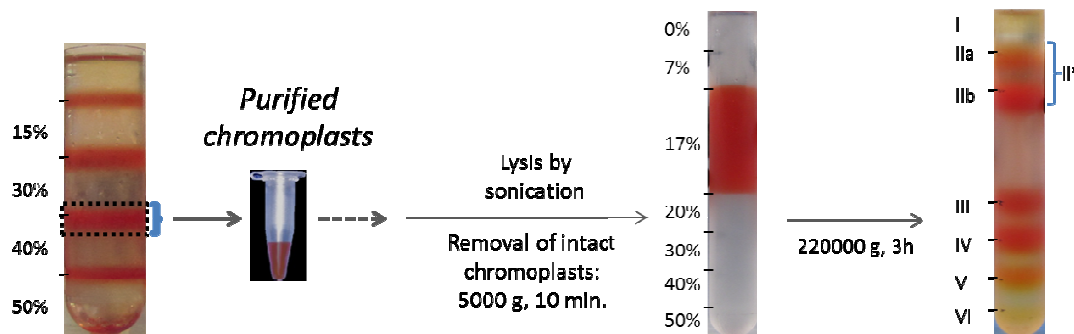
Figure 13. Overall view of the procedure for obtaining tomato fruit chromoplast subfractions. The gradient resulting from a chromoplast purification is shown at the left. The purified chromoplasts are pelleted and resuspended in sub-fraction isolation buffer with 17 % sucrose (50 mM HEPES 8 M, 5 mM MgCl$_2$). Lysis by sonication is performed and intact chromoplasts are removed by low-speed centrifugation. A sucrose gradient as displayed above is elaborated, and ultra-centrifugation is performed at 220000 g for three hours.

As shown in Figure 13 the gradient showed seven bands which were designated as I, IIa, IIb, III, IV, V and VI. The plastoglobule obtention methods adapted made us initially expect to find plastoglobules at fractions I, IIa and IIb. To harvest chromoplast subfractions a 200-microliter pipette was used (tips were widened prior to use by cutting them with scissors 1 mm from their end). The harvesting was performed gently describing horizontal circles inside the band, allowing for an efficient harvesting while not perturbing the interphases. Depending on the later use of the fractions, the samples were transferred into pre-weighed eppendorf tubes. When an increase in fraction purity and concentration was desired and no quantification was required (as in electron microscopy studies) only the central core of the bands was harvested.

Fractions I and VI correspond to the yellowish bands located at the top and the bottom part of the gradient, respectively. The rest of the bands had a varying orange/red coloration, with fraction IIb showing the most intense red colour, followed by fraction IV. A detailed description of the characteristics of each particular fraction is given below:

-Fraction I. It was not visualized unless large amounts of starting material (1 kg or more) were processed. Even so, it was mostly faint-yellowish in colour and of only very little visible entity.

-Fraction IIa. It often appeared weakly, even more so whenever a variety less lycopene-rich than the "Piccolo" cherry tomato was used. Due to its little entity, in some experiments it was decided to use a gradient setup which resulted in the non-existence of this band and its merging with band IIb (by omission of the 7% sucrose layer).

-Fraction IIb. Had a bright red colour and it occasionally displayed a reduced number of

small red particles of subtly elongated shape, of up to 0.5 mm in length. In the mentioned experiments in which the 7% sucrose layer was omitted, band IIb was in practice merged with the material present in band IIa. The resulting merger band was called "band II".

-Fraction III. Visually similar to band IIb, but less intensely coloured. More yellowish than IIa.

-Fraction IV. The most intense band, after band IIb. It displayed an intense red colour with a slightly orange tonality.

-Fraction V. Showed what appeared to be an internal gradient inside the band itself, being the upper part reddish and the lower part yellowish.

-Fraction VI. Was yellow in the upper part but showed the presence of some particulate-like sub-band of brown/orange colour at the bottom. Separate harvesting was only attempted when very good resolution was attained in late extraction attempts.

Some data found later on are referred to either fractions IIa/IIb or to fraction II. Fraction II is a fraction resulting of the summation of the material of IIa and IIb, as a result of a slightly different experimental setup (the omission of the 7 % sucrose layer).

4.2. METABOLITE PROFILING OF CHROMOPLAST FRACTIONS

In the context of our research, it was initially planned to elucidate the contribution of plastoglobules to the accumulation and storage of carotenoid during tomato fruit ripening. To that aim analyses of carotenoids, notably lycopene, were planned. Upon the growing interest in other lycopene-rich fractions the scope of those analyses was broadened to include those new fractions.

4.2.1. Carotenoid and tocopherol profiling

The method described by Fraser et al. (2000) was chosen for carotenoid profiling of the chromoplast subfractions. This method allows not only the quantification of carotenoids but also of their biosynthetic precursors (like phytoene and phytofluene) as well as other isoprenoids (i.e. tocopherols). In spite that this method has been widely used for carotenoid profiling in fruit samples, it was soon realized that its use in the analysis of our samples encountered a series of difficulties. The problems and solutions implemented are summarized below:

-Variability due to cumulative pipetting error during the process seemed to be excessive for some purposes. Because of that, instead of working with volumes, calculations were made

based on weight measurements.

-It seemed that during sample preparation prior to the HPLC analysis itself, part of lycopene could be lost by incomplete redissolution. This was solved by using a different solvent for sample resuspension, switching from ethylacetate to a mixture of ethylacetate and dichloromethane (3:1). This was done by adding dichloromethane first and next the required amount of ethylacetate.

-It was found that carotenoid degradation hindered correct quantification, very particularly that of lycopene. This seemed to happen while samples were in the HPLC queue. It was found in the literature that the major cause for irreversible degradation of lycopene is epoxidation, which happens upon reaction of a "singlet" (excited) oxygen molecule with a *cis* double bond of the carotenoid molecule. For this reason, exposure to light is the first step to approach irreversible degradation as it stimulates the isomerization of the all-trans forms of lycopene to the *cis* forms (Rodríguez-Amaya, 2010). Thus, the main factors involved in lycopene degradation seemed to be light and oxygen, which would be acting synergistically to accomplish irreversible lycopene degradation. As a result, during pigment extraction and all other pre-HPLC run manipulations, extreme caution was taken to prevent any direct light exposure, and storage was performed under inert atmosphere at -80ºC. It has been reported that pyrogallol, which is a cheap antioxidant, would diminish the speed of oxidative degradation by the quenching of singlet oxygen (Parrilla et al., 2007; Nagai et al., 2005). This compound could thus help blocking one of the mechanisms involved in the degradation of lycopene and likely other carotenoids. Even with this finding, to further prevent degradation, samples were kept at -80ºC until immediately before HPLC injection, instead of putting them in a queue at 4ºC for several hours. As well, the length of storage time was minimized.

After tackling the encountered problems, the quantification and estimation of the relative distribution of lycopene, beta-carotene, phytoene, phytofluene, alpha-tocopherol and gamma-tocopherol in the different fractions was undertaken. The results obtained in a typical fractionation process are shown in Table 9.

79

	Lycopene	β-carotene	Phytoene	Phytofluene	α-tocopherol	γ-tocopherol
Fraction I	0	2	12	13	4	2
Fraction IIa	9	15	7	7	3	1
Fraction IIb	**42**	**46**	12	12	6	4
Fraction III	20	16	5	5	8	8
Fraction IV	21	14	20	20	27	31
Fraction V	7	6	**30**	**29**	**38**	**41**
Fraction VI	1	2	15	14	15	13

Table 9. Relative distribution of each analyte among different fractions (expressed as % of that analyte). The fractions containing the major apportation to the total of that analyte in each column are indicated in bold. (indicated as % of total) of lycopene, beta-carotene, phytoene, phytofluene, alpha-tocopherol and gamma-tocopherol in each chromoplast fractions. Highest values for each analyte are indicated in bold.
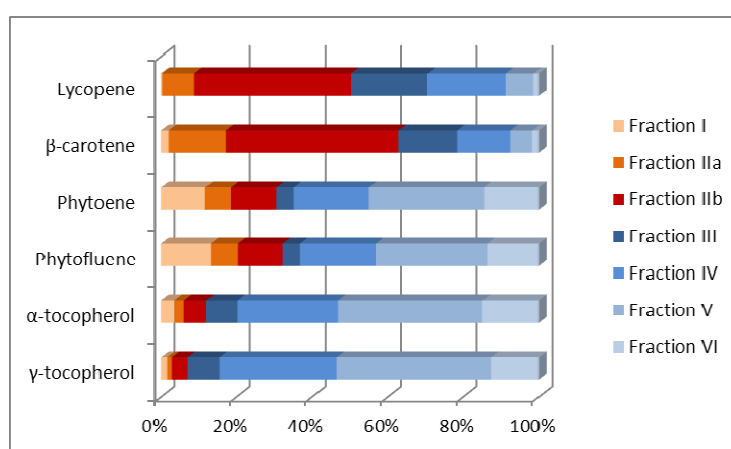
Figure 14. Graphical representation of the data shown in Table 9. Fractions displayed in reddish colours are those where plastoglobules were initially expected upon the adaptation of previous plastoglobule isolation methods. Bluish colours signal additional fractions separed by our extended method.

The following main conclusions may be drawn from the metabolite profiling shown:

-Lycopene was absent in fraction I and is preferentially found in the other light fractions, with over 70% of lycopene present in fractions IIa, IIb and III.

-The majority (near 80 %) of beta-carotene was found in fractions I, IIa, IIb and III. Its distribution shows a similar trend to lycopene. However, more of the total beta-carotene seems to be contained in the lightest fractions, IIa, and even I (where lycopene was absent).

-Phytoene and phytofluene were mainly found in the three heavy fractions IV, V and VI, with over 60% of the total. Interestingly, an important share of these carotenoid precursors are found in fraction I; this is eye-catching if it is taken into account that fraction I contributes very little or nothing to the total of the other analytes measured.

-The majority of tocopherols, over 80 %, were found in the heavier fractions, and only a small percentage was found in the light fractions I, IIa and IIb.

4.2.2. Membrane lipid profiling

The measurement of membrane lipids (galactolipids and phospholipids) was hoped to provide information on the nature of the fractions obtained. Thus, the levels of monogalactosyl diacylglycerol (MGDG), digalactosyl diacylglycerol (DGDG), phosphatidyl ethanolamine (PE) phosphatidyl serine (PS) and phosphatidyl choline (PC) was measured in the different fractions. The results obtained from a typical fractionation process are shown in Table 10.

	MGDG	DGDG	PE	PS	PC
Fraction I	n.d.	n.d.	n.d.	n.d.	n.d.
Fraction II	14	19	24	10	21
Fraction III	5	12	11	4	10
Fraction IV	20	**24**	13	4	19
Fraction V	**39**	**24**	**32**	**74**	**38**
Fraction VI	23	21	20	9	12

Table 10. Relative distribution of monogalactosyl diacylglycerol (MGDG), digalactosyl diacylglycerol (DGDG), phosphatidyl ethanolamine (PE), phosphatidyl serine (PS) and phosphatidyl choline (PC) among different fractions (expressed as % of that analyte). N.d., not determined. The fractions containing the major apportation to the total of that analyte in each column are indicated in bold.
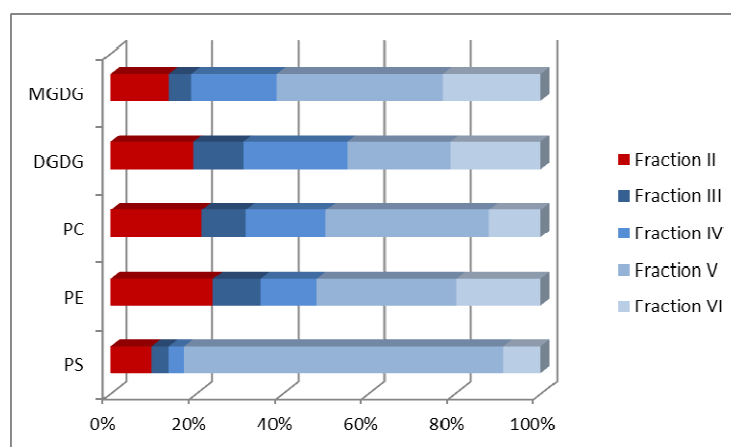
Figure 15. Graphical representation of the data shown in Table 9. Fractions displayed in reddish colours are those where plastoglobules were initially expected upon our adaptation of plastoglobule isolation methods from the literature. Bluish colours signal additional fractions separated by our extended method. Fraction I is not available; fraction IIa and IIb are here merged together into fraction II. Graphical representation of the results shown in Table 10.

According to the lipid composition of the different chromoplast subfractions the following trends can be observed:

-MGDG and DGDG were mainly found in fractions IV, V and VI, where they accounted for over 80% of total MGDG, and near 70% of DGDG, respectively.

-The majority of phospholipids was also found in fractions IV, V and VI (around 60% of PE, 85% of PS and near 70% for PC).

-Fraction V is the highest contributor to total PS found (over 70% of PS is found here).

-PE and PC were distributed roughly in the same way as MGDG.

4.3. ELECTRON MICROSCOPY ANALYSIS OF CHROMOPLAST FRACTIONS

Electron microscopy techniques were used to further characterize the nature of the fractions obtained. To this end, two dedicated electron microscopy techniques were performed using intact chromoplasts and subfractions obtained from different tomato varieties. Samples of the different subfractions were fixed to copper grids with either active charcoal or Formvar polymer coating, and with either methanolic (positive staining) or aqueous (negative staining) uranyl acetate staining. Upon trying a short range of staining supports and conditions it was found that either the Formvar polymer-coated grids or the activated charcoal-coated grids displayed similar performance. However, in the case of methanolic staining a certain degree of distortion was detected on some observed plastoglobules.

Additionally, purified chromoplast samples were fixed, dehydrated and included into resin to generate ultra-thin sections for electron microscope observation. Pictures of intact chromoplasts were intended as references for aiding the interpretation of the fractionated material.

A number of structures in intact chromoplasts from which the structures observed in chromoplast subfractions I to VI (see below) could be identified (Figure 16). They include:

-The outer plastid envelope (OE), and the inner plastid envelope (IE). From the inner plastid envelope, small blebs (B) were apparently forming.

- Plastoglobules (Pg), crystals (Cr), granular bodies (GB), thylakoid remnants (TR) and two different kinds of membrane-enclosed inner bodies (IB) were found inside of the chromoplast.

Apart from intact chromoplasts, a number of chromoplasts would be found in a process of apparently orderly remodeling. In those, chromoplast remnants including inner bodies far larger than those here seen could be found.
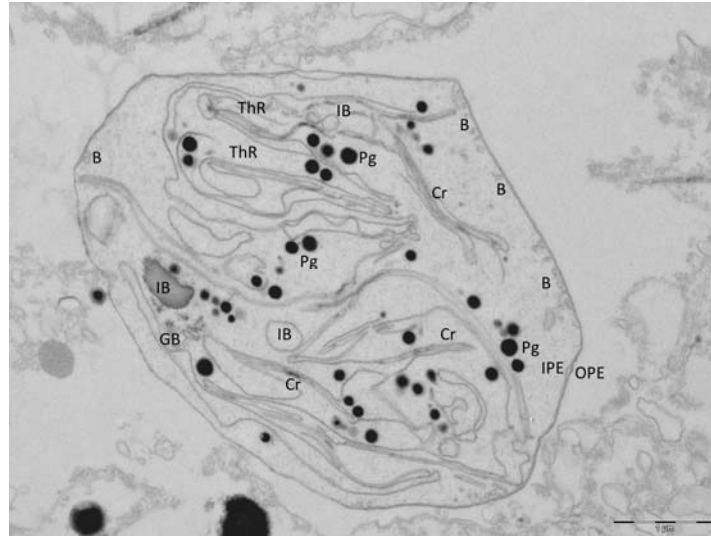
Figure 16. Electron microscopy images of purified chromoplasts. OE, outer plastid envelope; IE, inner plastid envelope; B, blebs; Pg, plastoglobules; Cr, crystals; GB, granular body; IB, inner bodies; ThR, thylakoid remnants.

The pictures corresponding to the different chromoplast subfractions are shown below.

-In fraction I (Figure 17), almost exclusively plastoglobules (Pg) were found, but even those were scarce. Sizes of plastoglobules found ranged from under 50 nm to over 300 nm, being most of them roughly between 100 and 200 nm.
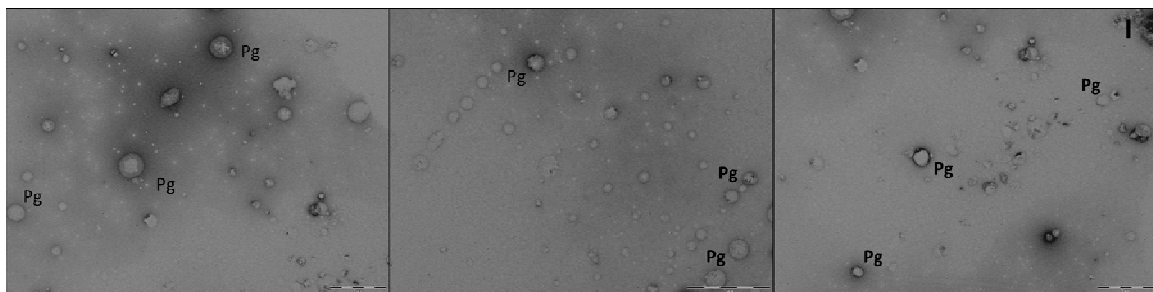
Figure 17. Electron microscopy images derived from fraction I. Pg, plastoglobules. Scale: 1 bar=1 μm.

-In fraction IIa (Figure 18), a mixture of plastoglobules and rod (R) and planar-like (PL) geometric structures are found. It is likely that the rod and rectangular-like structures derive from the lycopene crystals reported in tomato fruit chromoplasts (Spurr and Harris, 1968). Occasionally, on top of the planar structures round-like particles seemed to be found adhering, presumably plastoglobules.
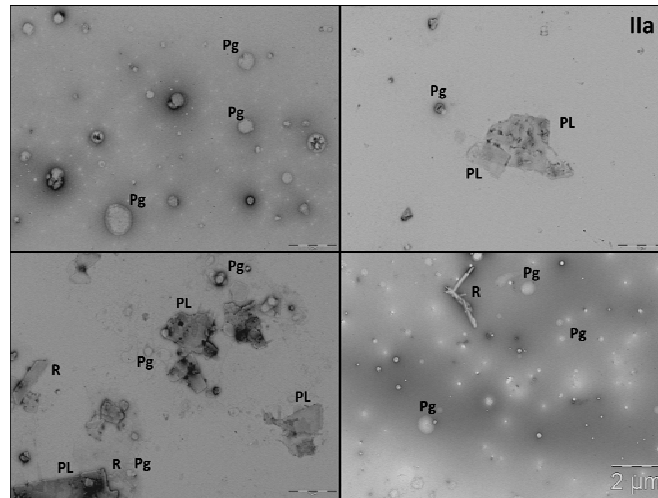
Figure 18. Electron microscopy images derived from fraction IIa. Pg, plastoglobules; R, rod structures; PL, planar-structures. Scale: 1 bar=1µm, except figure at bottom right, 1 bar=2 µm.

-In fraction IIb (Figure 19), a landscape similar to that of fraction IIa was found, with the presence of plastoglobules, rod-like and rectangular-planar structures. Again, in some of the planar structures an apparent adhesion of possible plastoglobules is found.
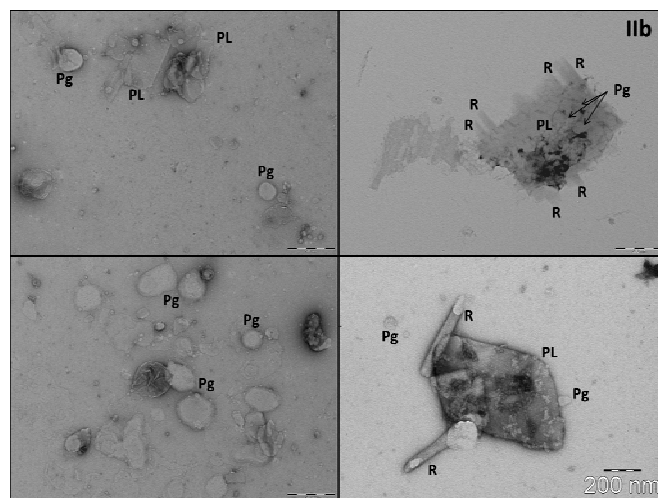
Figure 19. Electron microscopy images derived from fraction IIb. Pg, plastoglobules; R, rod structures; PL, planar-structures. Scale: 1 bar= 500 nm, except bottom right figure, 1 bar= 200 nm.

-In fraction III (Figure 20), particles similar to plastoglobules were found, together with apparent small membrane fragments. A high number of small spots (S) was found, of a size compatible with that of ribosomes (near 20 nm diameter). Prokaryotic and eukaryotic ribosomes are 18 nm, and 25 nm in diameter, respectively (Nelson and Cox, 2008). An agglomeration of fragments was occasionally found, which could be either an accumulation of membranes, ribosomes, or both. A number of vesicle-like structures were also found.
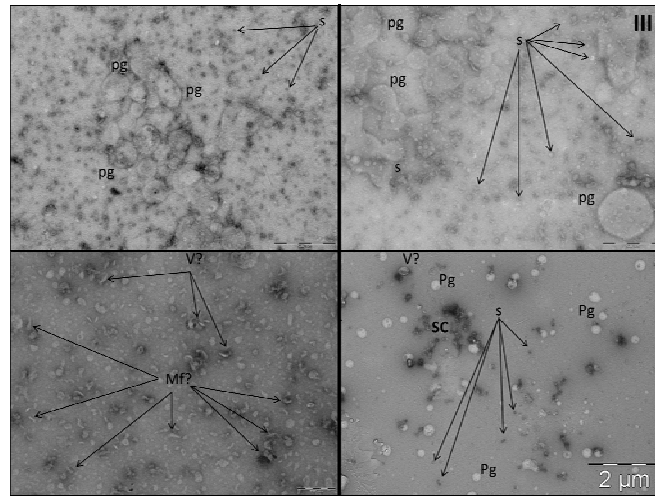
Figure 20. Electron microscopy images derived from fraction III. Mf, small membrane fragments; pg, possible plastoglobules ; s, spot-like particles (possible ribosomes); SC, spot cluster. Scale: top pictures, 1 bar=500nm.; bottom left picture, 1 bar=200nm; bottom right figure, 1 bar=2μm.

-In fraction IV (Figure 21), mainly a display of small vesicle-like particles was observed. Their apparently hollow appearance distinguished them from plastoglobules. A myriad of probable small membrane fragments were also found in this fraction. Possible plastoglobule-like structures were found clustered close together. "Spot" clusters compatible with ribosome-sized particle agglomerations were found.
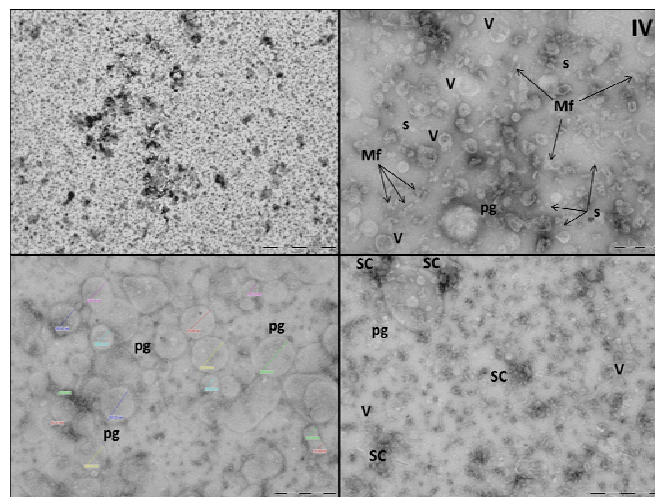
Figure 21. Electron microscopy images derived from fraction IV. Mf, membrane fragments; pg, possible plastoglobules; s, spots; SC, spot cluster; V, vesicle. Scale: top left picture, 1 bar=2μm; top right picture, 1 bar= 200 nm; bottom pictures, 1 bar= 500 nm.

-In fraction V (Figure 22), a variety of small and big membrane fragments were present. Spots similar to those found in fractions III and IV were also present. Large perforated plastid

membrane shells were observed.
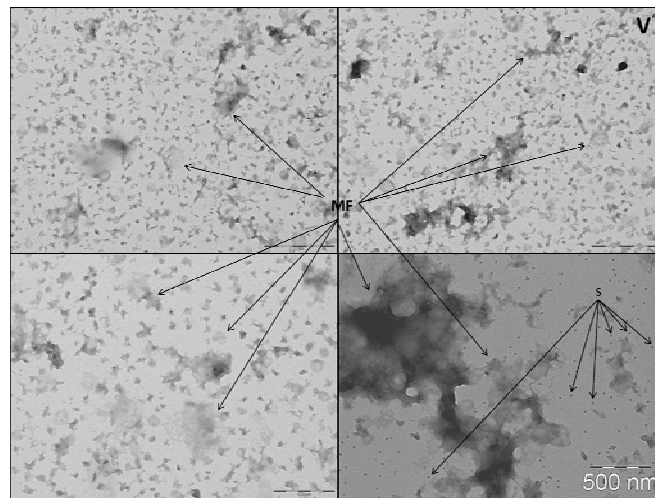
Figure 22. Electron microscopy images derived from fraction V. MF,membrane fragments of different sizes; s, spots (possible ribosomes). Scale: top pictures, 1 bar= 1 µm; bottom pictures, 1 bar= 500 nm.

-In fraction VI (Figure 23), multiple membrane fragments were found. A particularity distinguishes it from fraction V, as some of the membrane fragments were apparently generating vesicles. Sometimes those vesicle formations would take the shape of outgrowths of short elongated finger-like membrane shoots. Those vesicle-generating membranes have been termed by us as vesicular bodies. A number of independent vesicles were also found.
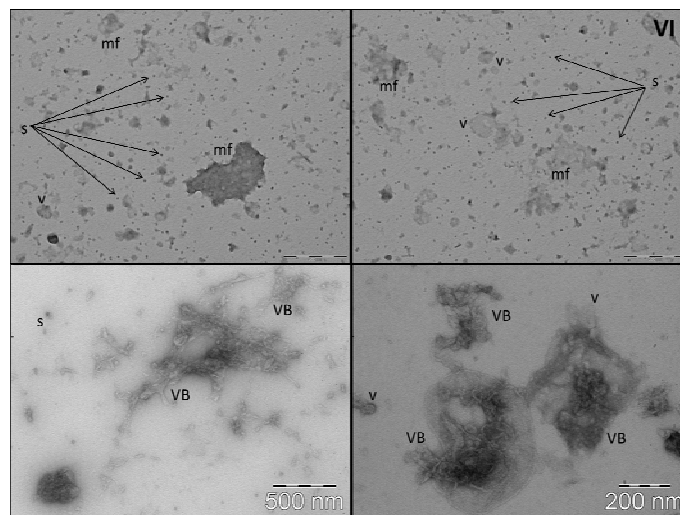
Figure 23. Electron microscopy images derived from fraction VI. mf, membrane fragments; s, spots (possible ribosomes); v, possible vesicles; VB, vesicular body. Scale: 1 bar=500 nm, except bottom right figure, 1 bar=200 nm.

4.4. PROTEOMIC ANALYSIS OF CHROMOPLAST FRACTIONS

Proteomic approaches were used to get a better characterization of the fractions obtained. As well, during the setup of the fractionation method, proteomics was used to identify proteins which could confirm that the strategies undertaken were correct.

During the very initial steps aimed at the isolation of plastoglobules, SDS-PAGE gels were run on the fractions at that moment expectedly containing plastoglobules, which were grossly coincident with the current fractions I, IIa and IIb. It was found that the electrophoretic runs usually became very distorted. This problem was attributed to the high lipid content of this fractions, which was in high contrast with the very low protein content. The implementation of a procedure for the removal of lipids drastically improved band resolution and the overall run performance. A side effect was the decrease of the protein yields, which involved the need to scale-up the later extractions. In the search for markers of correct plastoglobule isolation, the improved SDS-PAGE analysis of the fraction expectedly containing plastoglobules confirmed that we were on the right path (figure 24). Five of the major bands and two minor bands were picked from the gel. The corresponding polyacrylamide gel fragments were exposed to in-gel tryptic digestion and MS/MS identification at the Proteomic Service of PCB-UB. Five out of the seven proteins identified corresponded to proteins previously identified in plastoglobules, and one ATPase subunit was found very closely related to another previously reported protein (Vidi et al., 2006). From the results, as well, it seemed to become evident that the three major proteins found in tomato fruit plastoglobule are proteins of the fibrillin family.
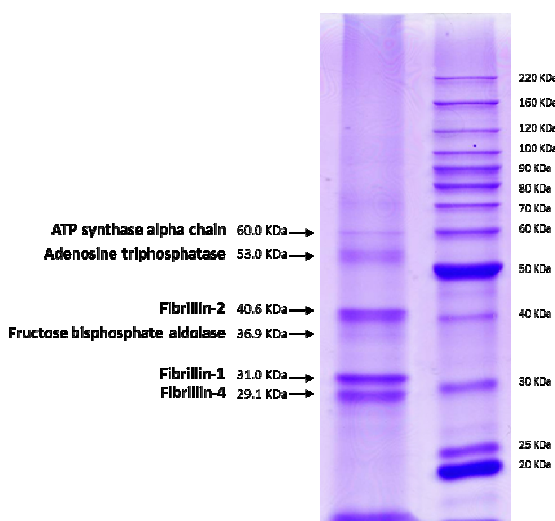
Figure 24. SDS-PAGE of plastoglobule enriched fractions. Spots were picked and identified by tryptic digestion followed by MS/MS. The identified proteins are indicated next to each band.

As the purification method became more robust, further proteomics studies were conducted. However, the relatively trivial task of measuring the protein content in the isolated fractions posed an unexpected complication. The majority of stromal proteins were expected to be lost except, possibly, for the more concentrated proteins or those capable of membrane interactions. In any case, the majority of the proteins present were probably attached or associated to membranes. Thus, the use of a detergent (Triton X-100) was needed for protein solubilisation before protein estimation.

The use of Triton X-100 would yield the standard Bradford protein measurement method useless, and for that reason the search for another protein measurement method was required. A detergent-compatible method was chosen, using a modified Löwry procedure involving protein precipitation (RC-DC protein measurement kit, BioRad). According to its specifications, the system was widely compatible with detergents, reducing agents, and resilient to interference by many soluble contaminants. After establishing the protein solubilization conditions, the content of the protein along the gradient fractions was measured (figure 25).

	Protein %
Fraction I	undetected
Fraction II	4
Fraction III	5
Fraction IV	19
Fraction V	**36**
Fraction VI	**36**
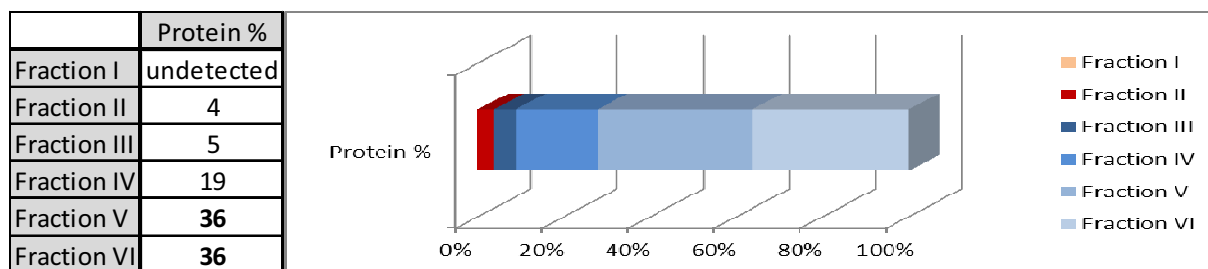
Figure 25. Relative distribution of protein content among different fractions (expressed as % of total protein). Relative protein content (expressed as %) in the different chromoplast subfractions. Fraction II corresponds to the merging of fractions IIa and IIb.

It was found that the protein content of the upper fractions was very low, being the first clear trend to be seen that the light fractions I, II (IIa, IIb) and to a much lesser extent fraction III, were protein-poor fractions. This could reasonably be expected for fractions accumulating high amounts of lipids and pigments. As a result, it was found that the bulk of the protein was found in fractions IV, V and VI, which accounted for close to 90% of the total protein.

This posed us with a technical problem as such a difference in protein content would make some proteomic approaches simultaneously inapplicable to protein-rich and protein-poor fractions. For that reason the strategies for the proteomic studies were different for the light and the heavy fractions.

4.4.1. Proteomic analysis of the light fractions I and II

For the proteomic studies of fractions I and II a collaboration was established with the group of Professor Hans-Peter Mock at the Max Planck Institute at Gatersleben (Germany). Samples initially sent for analysis corresponded to fractions I and II (a merger of IIa and IIb). Because of the low protein yield of fraction I, this fraction was not ready for proteomic analysis until very large amounts of fruit were used. Currently, only data for fraction II (IIa + IIb) are available. The list of confirmed proteins in fraction II are indicated in table 11.

ID:	Description:
Solyc01g014210.1.1	ADP ATP carrier protein-like
Solyc01g096290.2.1	Ubiquitin
Solyc03g095900.2.1	1-aminocyclopropane-1-carboxylate oxidase-like protein
Solyc04g007770.2.1	Major latex-like protein
Solyc04g071610.2.1	Water-stress inducible protein 3
Solyc04g082200.2.1	Dehydrin
Solyc07g049140.2.1	Metallocarboxypeptidase inhibitor
Solyc08g081190.2.1	Aquaporin 1
Solyc09g090330.2.1	Fibrillin 4/Harpin binding protein 1
Solyc11g008990.1.1	Phage shock protein A
Solyc11g012320.1.1	Unknown Protein
Solyc11g062190.1.1	Mitochondrial ADP/ATP carrier proteins
Solyc11g069700.1.1	Elongation factor 1-alpha

Table 11. Proteins identified in fraction II.

Of this list, a few found proteins can be highlighted. Fibrillin-4 (FBN4) is found, which is a major component of plastoglobules and a member of the fibrillin family, which is involved in plastid organization and stress response (Singh et al., 2010). The protein 1-aminocyclopropane-1-carboxylate oxidase-like protein is found, which is involved in the generation of ethylene (Schomburg et al., 2002). The proteins phage shock protein A (PspA), water-stress inducible protein 3 and Dehydrin all are proteins which are synthesized in response to stress.

The proteins identified in fraction II were different from what would be expected from plastoglobules, which were initially anticipated to be found in fraction II. The previous results, as it was mentioned previously, are referred to the merger of fractions IIa and IIb. Thus, these results did not match the expectations about the identification of plastoglobule proteins. However, if we take into account the electron microscopy results, which were not available when these experiments were done, it can be deduced that this light fraction contained a large amount of crystals, as well as a certain amount of plastoglobules whose protein became masked.

Precisely, being the "Piccolo" variety rich in lycopene and crystal-like structures (as seen in the electron microscopy analysis), these crystals may have been obscuring the identification of plastoglobular proteins. Results for fraction I, consisting of "pure" plastoglobules are still pending, and thus can neither be used to support nor to refute this hypothesis.

Luckily, some complementary results were obtained for fraction IIa, using in this case a standard commercial cherry variety, and thus having less crystal structures which could interfere with the identification of plastoglobule proteins. The finding derived from the electron microscopy analysis showing that this variety had scarcity of rod and planar-like crystal structures gave some reassurance that what follows can be related to what "pure" plastoglobules would reflect. Moreover, the relatively similar density expected for fractions I and IIa makes it a plausible assumption that they might have a similar protein composition.

These serendipitously obtained results were obtained during a verification of the major bands of fraction IIa by band-picking after a modification in the fractionation method was implemented. An upgrade in the equipment available at the Proteomics Platform of PCB-UB allowed the identification of a list of identified proteins upon spot-picking, instead of only discrete proteins as the first experiment. The major proteins detected were the same obtained in the first proteomic analysis of plastoglobules (also obtained from standard commercial cherry tomatoes). The IDs and descriptions of the identified proteins are shown in annex I. Among them, the following proteins can be highlighted:

-15-*cis*-zeta-carotene isomerase, carotenoid isomerase and zeta-carotene desaturase (enzymes related to carotenoid biosynthesis)

-Tocopherol cyclase (related to a protein previously reported for plastoglobules by Vidi et al., 2006).

-Cyclopropane-fatty-acyl-phospholipid synthase and lipoxygenase.

-FBN1, FBN2, FBN4, FBN8 and FBN12 (proteins of the fibrillin family - annotated as plastid lipid associated proteins).

4.4.2. Proteomic analysis of fractions III, IV, V and VI

For the proteomic analysis of fractions III, IV, V and VI the classic proteomics approach based on 2D protein gel electrophoresis was initially used. Material was accumulated from five different chromoplast extractions. Even after repeated high-volume extractions, the accumulation of material for fraction III seemed to lag behind the rest, below the amount of protein needed for

an optimal 2D gel electrophoresis analysis. Nevertheless, it was decided to analyse the samples obtained to that point and compensate the low amount of protein by using longer silver staining to increase signal.

After running the 2D gels and staining the gels it was observed that the assayed fractions displayed a similar pattern of spots but also differences in a few of them (as illustrated for fractions V and VI in figure 26). Clearly differential spots were identified and picked up for protein identification after tryptic digestion and MS/MS analysis. Unfortunately, none of them could be identified due to either too low protein concentration or to interference from the silver stain (despite the use of a theoretically MS/MS-compatible silver staining method). Nevertheless, it could be concluded that many similarities were to be expected at the level of protein identity among those heavy fractions. Differences on spot intensities could also be observed in some cases.
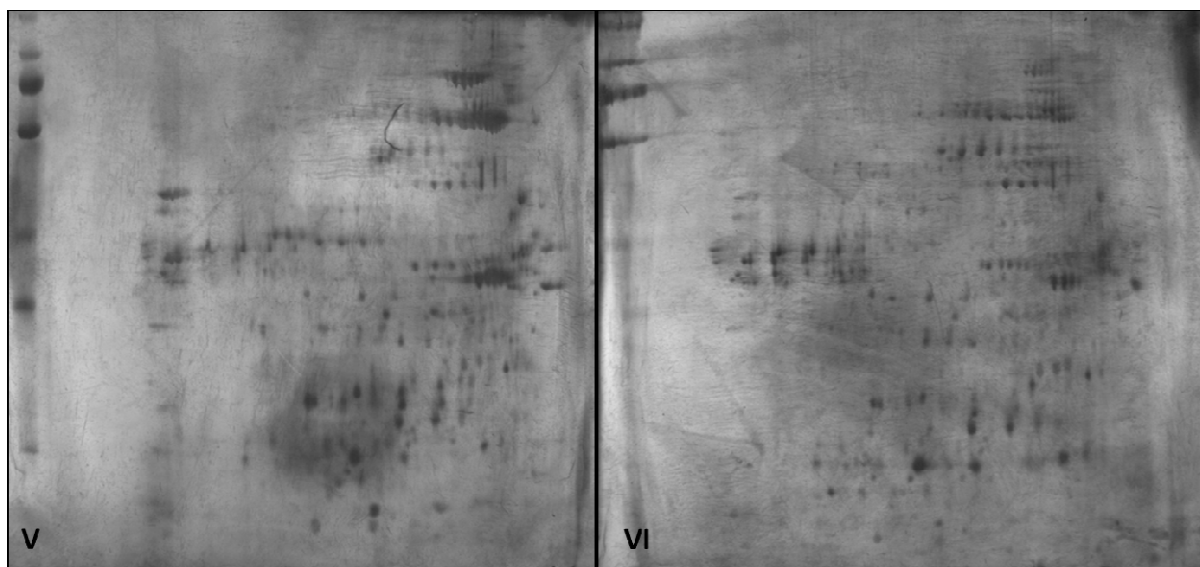
Figure 26. 2D gels of proteins present in fractions V and VI. Proteins were stained using a silver staining protocol.

As identification of the differential spots was unsuccessful it was considered to repeat the analysis using the less sensitive "blue silver" colloidal Coomassie stain, which is known to cause no interference at all with MS/MS protein identification. However, bearing in mind the subtleness of the differences found with silver staining and also the high amount of material necessary for 2D gel assays using a less sensitive staining method, a different approach based on shotgun-proteomics was considered.

To this end, five independent chromoplast fractionations were performed and the fractions were pooled and processed. Sample proteins were pre-treated as normally done (see sections 3.6.4 and 3.6.5), and redissolved by the same procedure as followed for 2D electrophoresis (see section 3.6.6). In-solution tryptic digestion of the protein samples followed by liquid chromatography and MS/MS analysis allowed the identification of a larger number of proteins, with the potential of perceiving more far-reaching differences.

After samples were processed and the first results were available it became evident that the total number of proteins identified was lower than expected. Because of this, it was decided to check if the databases used for protein identification were thorough enough for our purposes. It was realized that the information derived from the sequence of the tomato genome was not fully represented in the general reference databases EXPASY and NCBI. It was, as a result, resorted to the tomato genome project data available at the Solgenomics ftp site (http://solgenomics.net/bulk/input.pl?mode=ftp). There, a file was retrieved containing the protein-translated tomato genome, thus the most extensive listing of proteins of that species available. This listing was added to the local database used for MS/MS identification, with the result that the amount of proteins successfully identified dramatically increased.

After the new database search, a list of 510 identified proteins was generated. This list was cut down to 302 proteins after removing those proteins which could likely represent possible contaminants. Of the filters applied after receiving the identified proteins, around 50% of the proteins were accepted as positive hits as they had been found in the previously published tomato chromoplast proteome (Barsan et al., 2010, 2012). A remaining 15% was added to the final list as a result of checking against plastid-encoded proteins, checking the subcellular location found for protein homologues in other databases and the use of subcellular protein location prediction tools (see Protein "hit" filtering, section 3.6.9).

The proteins validated in each fraction were classified according to their absence or presence in the other fractions. A Venn diagram to visualize the number proteins specific of a particular fraction or common to another fraction(s) is shown in Figure 27.
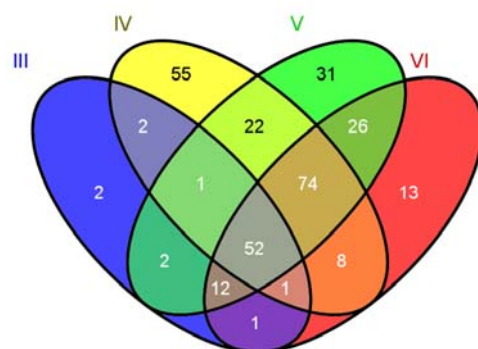
Figure 27. Venn diagram of the proteins identified in fractions III, IV, V and VI. Numbers indicates the number proteins unique to each fraction and common to other fractions.

To obtain a global vision of the significance of the identified proteins they were searched for their associated Gene Ontology (GO) descriptors, according to the Gene Ontology project. The GO project provides a controlled vocabulary of terms for describing gene product characteristics and annotation data, as well as tools to access and process this data, which is useful for standardizing the representation of gene and gene product attributes across species and databases (Ashburner et al., 2000).

The GO project, in association with the NEUROLEX website (http://neurolex.org/wiki/Category:Resource:Gene_Ontology_Tools), provides a set of GO-related tools which were explored to find a suitable aid for GO-retrieval and later graphical representation of the results. Many of the tools found were promptly discarded. Among the few remaining tools, an apparently ideal GO-analysis server devoted to plants was found: AGRIGO (Du et al., 2010). However, support and maintenance apparently had been discontinued. It was thus finally resorted to one generalistic GO-annotation bioinformatic tool: the BLAST2GO program (Conesa and Gotz, 2008), which is capable of retrieving the corresponding GO terms upon successive BLAST searches against GO-annotated protein/DNA databases. This tool provided a list of 1610 GO descriptors for the target proteins.

Upon retrieving the GO descriptors via BLAST2GO data visualization posed us with the problem of having a too large set of descriptors for easy handling. The high number of GO terms, including terms of a handful of different GO hierarchical levels, made our data somewhat unwieldy unless some additional filtering could be applied. Conveniently, BLAST2GO itself has a built-in tool for the application of a "Gene Ontology Slim" (or GO slim), which allows for the pruning of the redundant or excessively specific branches of a GO tree generated from a specific data set. This is useful for the simplification of lists of GO terms in connection to a specific

context of interest as, in our case, plant biology. Users with the required computational skills have the possibility to elaborate their own self-tailored GO slims. Though efforts have been reported to make this task more straightforward (Davis et al., 2010), elaborating a new GO slim usually has been regarded as a time-consuming and non-trivial task, which has remained beyond our plans. As a result, one of the existing GO-slims was thus chosen. The most up-to-date plant GO slim set available belongs to Arabidopsis, and was elaborated by The Arabidopsis Information Resource (TAIR), as is found in the GO project listing http://www.geneontology.org/GO.slims.shtml

This GO slim set, included in the BLAST2GO program suite, was the one finally used in our study. A more general plant slim set was discarded due to its lack of update since the year 2002. After applying the TAIR GO slim, 1423 non-unique GO term descriptors remained for the total of proteins. The remaining GO terms were renamed to their defined upstream equivalents, according to the GO slim devised by TAIR. Of the GO terms obtained, the focus has been kept on the terms related to the "biological process" GO category.

The NEUROLEX website was searched to find any appropriate specialized graphication tools. A tool called REVIGO (Supek et al., 2011) seemed promising at first sight for offering a global overview of the proteins present in each fraction, but was finally discarded. This was due to a limited control on the appearance of the graphic representation, which did not allow to illustrate the conclusions drawn. As well, a problem with the hierarchization of the labels of the represented GO categories further drove us away from this tool. No other options seemed suitable to our goals. Consequently, BLAST2GO was also used to obtain preliminary representations, and the related list of GO descriptors organized according to their relevance "score" based upon GO term occurrence and the number and distance of "child" GO terms. This list was processed to obtain new graphical representations of the fractions III, IV, V and VI involved in the massive protein identification. Next, the contribution to the overall GO terms found in all fractions together can be estimated by dividing the GO terms into different major subcategories after the application of a GO slim and the removal of low-relevance terms. The obtained results are shown in Figures 28 and 29.
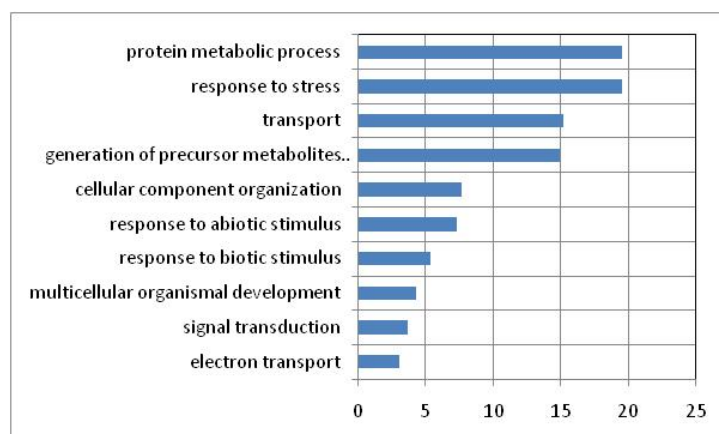
Figure 28. General representation of the percentage of GO terms associated to each category, for the pooled list of all proteins identified in any of chromoplast fractions III-VI, after the application of a GO slim.
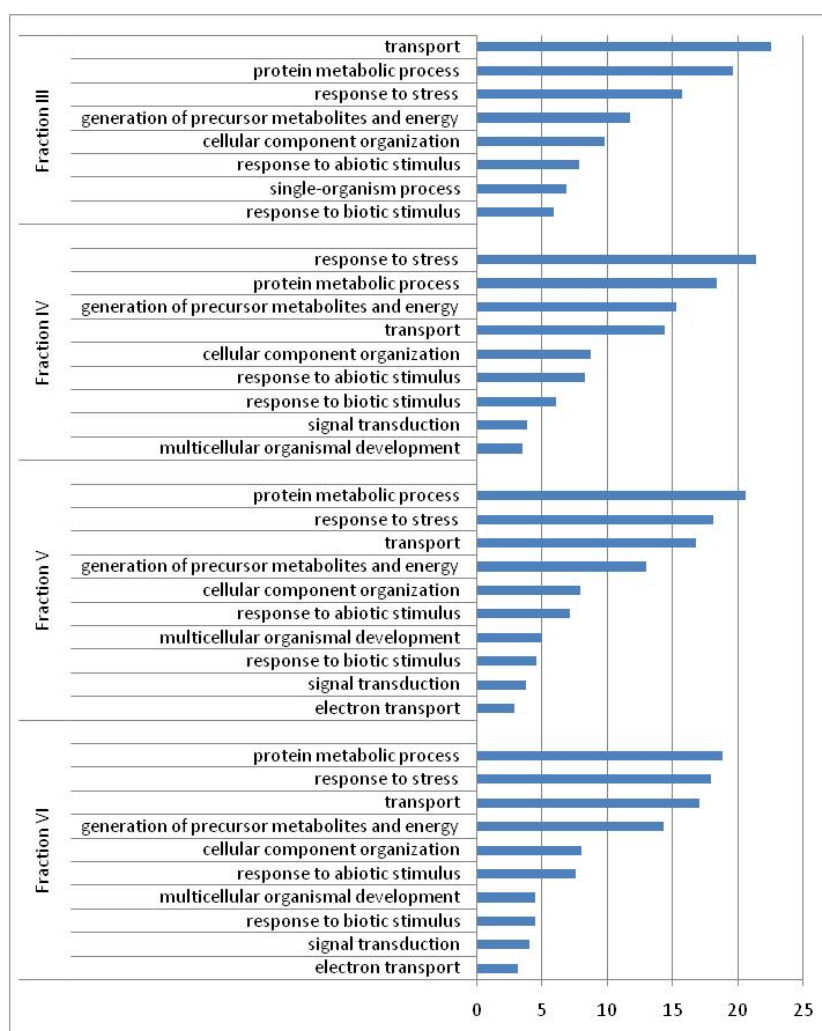
Figure 29. Representation of the contribution (%) of each GO category to the total of GO terms found for the proteins identified in chromoplast fractions III, IV, V and VI.

95

To obtain further insights into the biological activities associated to the different chromoplast fractions, the raw data from BLAST2GO were exported for every one fraction, and the list of unique and common-to-all proteins with their verbose GO descriptors was visually inspected. According to this, they were divided among several different manually designated categories, which were by us chosen according to the abundance of the different proteins and GO descriptors. As frequently proteins had several descriptors, one given protein could belong to more than one category. The selected categories were: ATP/energy-related, stress/defence response, sugar/carbon metabolism, chaperones, protein translation-related proteins, redox homeostasis (or closely related), signalling/regulation, amino acid biosynthesis, DNA/RNA processing, protease activity, transport activity (excluding ADP/ATP), lipid-related proteins, other proteins (diverse), and proteins of unknown function. From this, a graphical representation was as well made.

The GO descriptors were as well captured into a reduced table illustrating the different activities or metabolic involvements detected in the unique proteins, or in those common to all fractions (Table 12, Figure 30). The most generalistic GO descriptors were ignored for the elaboration of this table. The list of proteins unique or common to all fractions can be recovered checking the listing at Annex II.

| | GO descriptors associated to proteins unique to: | | | | |
	Fraction III	Fraction IV	Fraction V	Fraction VI	Common to all
Energy obtention-related	2	2	1	0	10
Stress response-related	0	16	8	3	17
Carbon metabolism	0	13	0	5	4
Chaperones	0	6	0	2	4
Translation-related	0	3	4	0	3
Redox homeostasis and related	0	6	2	2	2
Regulatory proteins	0	5	3	3	7
Amino acid synthesis and related	0	4	3	1	1
Nucleic acid processing	0	4	0	1	1
Proteases	0	5	2	0	2
Transporters	0	2	5	0	4
Lipid-related	0	5	3	2	3
Other	0	6	8	2	6
Unknown	0	1	2	0	3
Total	2	55	31	13	53

Table 12. Table displaying the number unique and common-to-all proteins for each fraction, classified according to selected categories. The raw GO term information was exported from BLAST2GO to manually elaborate this categorization. The localization of the proteins identified can be recovered from the location codes at annex II.
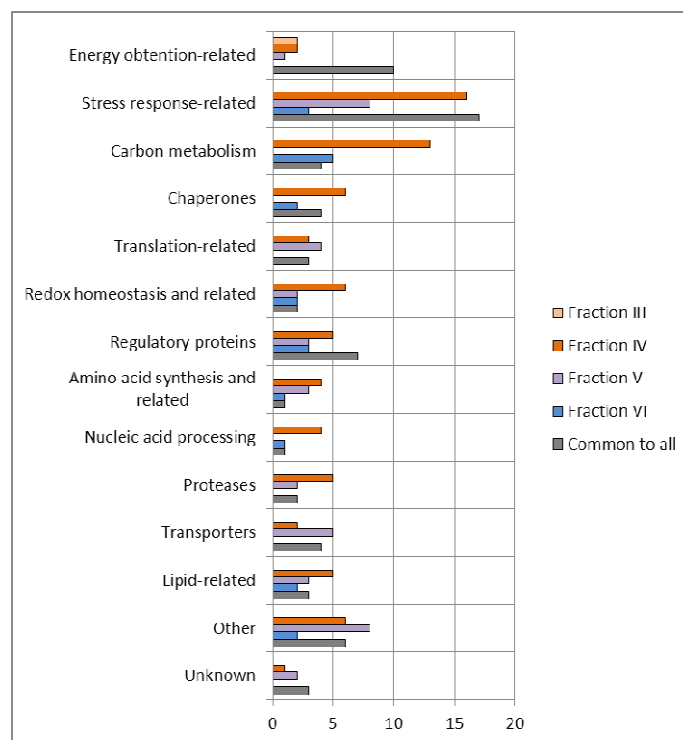
Figure 30. Comparison of the proteins unique to each fraction or common to all fractions. The raw GO term information was exported from BLAST2GO to manually draw this categorization. The coordinate axis represents the number of terms encountered associated to a given function. The localization of the proteins identified can be recovered from the location codes at annex II.

From observing the representation, several assertions can be made regarding the unique proteins of fractions III, IV, V and VI.

Firstly, it can be seen that fraction III has very little specific components of its own. Those are two energy metabolism-related proteins.

Fraction IV had, of all four fractions, the widest variety of unique functionalities found, with 55 proteins. This was the only fraction containing unique proteins for each of the 14 protein categories elected for representation. Of the proteins unique to fraction IV, the most numberous groups were those related to stress response, carbon metabolism, diverse regulation processes, amino acid biosynthesis, nucleic acid processing, proteases and lipid metabolism, as well as the highest number of unique chaperones. The main components unique to this fraction were found to be stress or carbon-metabolism-related, almost accounting for two fifths of the functions found for the unique proteins of this fraction.

Fraction V displayed the second highest number of unique proteins, amounting to 31. Fraction V stands out as possessing the highest number of unique translation-related and

transport-related proteins. Fraction V does not have any unique proteins related to carbon metabolism, nucleic acid processing, nor any specific chaperone. One fourth of the unique descriptors for proteins of this fraction were assigned as "unknown" or "miscellaneous".

Fraction VI, the heaviest of all four, was found to be simpler than IV and V, with only 13 unique proteins. Still, some specificities were found, being a significant number of the annotations of its unique proteins related to lipid metabolism, sugar/carbon metabolism, and to signalling and regulation. No specific transport-related, protease, or ATP/energy metabolism-related proteins were found.

From looking at the function of proteins common to all four fractions, we can find that the majority of them are either related to energy metabolism, or to stress response. These are followed by a number of regulatory proteins, and a variety of other proteins. The unique and common-to-all proteins can be checked in annex II by means of their assigned location codes in the list of identified proteins (see annex II).

4.5. EXPRESSION AND FUNCTIONAL ANALYSIS OF CHROMOPLAST FIBRILLINS

During the initial proteomics studies it was found that the three major proteins found in plastoglobule containing fractions were proteins of the fibrillin family. Fibrillins, also known as plastoglobulins, are evolutionarily conserved plastid lipid-binding proteins reported to be present in the plastoglobule surface, but also to various extents in the thylakoid membrane and stroma (Lundquist et al., 2012; Shanmugabalaji et al., 2012). They are nuclear-encoded and mostly plastidial proteins found associated with chromoplast fibrils, plastoglobules, thylakoids, stroma and photosynthetic antenna complexes. Fibrillins are involved in chromoplast pigment accumulation (Simkin et al., 2007), in photosystem protection from photodamage, the response to hormones and plastoglobule structural development. Coherently with their close relationship with the latter, which are starkly involved in stress response, fibrillins also play an essential role in plant resistance to a wide range of biotic and abiotic stresses (Singh and McNellis, 2011; Yang et al. 2006). Fibrillins are indeed thought to have a structural role in the formation and maintenance of plastoglobules (Shanmugabalaji et al., 2012), but their expression pattern, structure or way of action are largely unknown. With the aim of shedding light on these matters, we addressed the study of these proteins by both experimental and bioinformatics approaches, as

they expectedly would be playing an important role during tomato fruit ripening.

4.5.1. Expression analysis of tomato fruit chromoplast fibrillins

Among the experimental approaches addressing the study of the three major fibrillins found in tomato fruit plastoglobule (FBN1, FBN2 and FBN4), we firstly undertook the study of their gene expression level. Primers for these and the other fibrillins also reported in the literature to be present in the tomato chromoplasts (Barsan et al., 2010) were designed. This led us to design qPCR primers for FBN1, FBN2, FBN3, FBN4, FBN6, FBN8 and FBN9.

From the representation of fibrillin expression levels in different stages of tomato fruit development and ripening it could be observed that FBN1 is the most prominent fibrillin in tomato fruit. FBN4 and FBN6 are the next most expressed fibrillins.
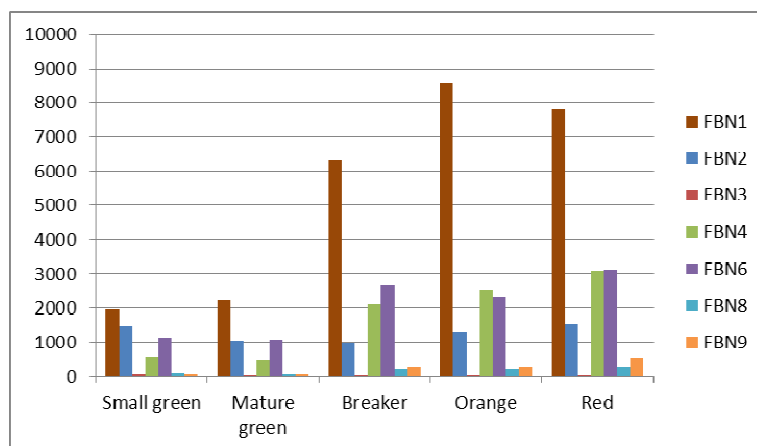
Figure 31. Representation of the estimated copy number of the transcripts corresponding to fibrillins FBN1, FBN2, FBN3, FBN4, FBN6, FBN8 and FBN9 during tomato fruit development and ripening. Expression was normalized using the housekeeping CAC gene.

FBN1, FBN4, FBN6, and more mildly FBN8 and FBN9, showed an overall increase in expression over fruit ripening. The increase in expression was not sustained in all cases throughout ripening, with FBN1 losing ground in the last stage and FBN6 showing an oscillation in the orange stage. Fibrillin transcripts not showing increase during ripening were FBN2 and FBN3, with FBN2 showing a drop in expression which however recovered slowly as ripening proceeded. FBN3, on its part, started from very low expression levels, which decreased to almost zero.

The calculation of the fold increase of FBN1, FBN2, FBN3, FBN4, FBN6, FBN8 and FBN9 expression in red fruit vs. small green fruit (Table 13) offers one additional insight on these results.

	Fold increase ripe red vs. small green
FBN1	4,2
FBN2	1,1
FBN3	0,0
FBN4	5,5
FBN6	2,7
FBN8	3,1
FBN9	6,9

Table 13. Fold increase of fibrillin FBN1, FBN2, FBN3, FBN4, FBN6, FBN8 and FBN9 expression in red tomato fruit vs. small green fruit.

Fibrillins with the highest overall fold-increase during ripening if compared to small green are FBN1, FBN4 and FBN9. FBN6 and FBN8 increase as well, but to a lesser extent. FBN2 does not attain significant overall increases, while FBN3 expression is almost completely lost.

4.5.2. Virus-induced gene silencing of plastoglobule fibrillins

The role of the major plastoglobule fibrillins (FBN1, FBN2 and FBN4) with regard to fruit ripening was evaluated using virus-induced gene silencing (VIGS). VIGS is a reverse genetic tool for the analysis of gene functions which uses viral vectors carrying a gene fragment allowing to form dsRNA inside the targeted cells, which in turn trigger RNA-mediated gene silencing. The advantages of VIGS over other functional genomics approaches is the rapid generation of a phenotype, the lack of need for plant transformation, a relatively low cost, and the possibility for large-scale screenings with a reasonable effort (Unver et al., 2009). A procedure based on a protocol described by Orzaez and coworkers was used (Orzaez et al., 2006, 2009). In that system, an engineered tomato variety is used which expresses the anthocyanin-synthesis Delilah and Rosea genes (*Del*/*Ros*) upon ripening. The genes of interest to be silenced are cloned into suitable vectors containing as well fragments of the genes *Del* and *Ros* in such a way that the gene of interest is in theory silenced together with the *Del* and *Ros* genes, upon tomato fruit agroinjection. Thus, in this system, the regions where the disappearance of the dark-purple anthocyanin happens, are expected to be those where the gene of interest is silenced as well.

VIGS constructs were prepared and experiments were performed in which silencing constructs for FBN1, FBN2 and FBN4 were injected in parallel to controls including either a positive silencing PDS (phytoene desaturase) control, and an empty pTRV2 vector silencing only

the *Del* and *Ros* genes. A clear effect on ripening was found in the fruits agroinfected with the FBN1, FBN2 and FBN4 constructs (Figure 32).
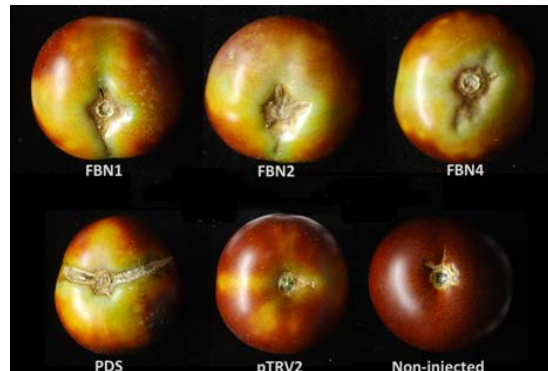
Figure 32. Tomatoes injected with GV3101 agrobacterium and the FBN1, FBN2 or FBN4-silencing constructs. At the bottom, from left to right, a positive control PDS-silencing construct, and two negative controls including an empty pTRV2 injected vector, and a non-injected DR tomato.

The obtained results suggests that FBN1, FBN2 and FBN4 are required for ripening or accumulation of lycopene in tomato fruit. Further studies are needed however to confirm this point and evaluate the mechanism underlying the observed phenotypes.

4.6. STUDIES TO UNRAVEL STRUCTURAL AND FUNCTIONAL FEATURES OF PLANT FIBRILLINS

4.6.1. In silico analysis of the three major tomato fruit plastoglobule fibrillins FBN1, FBN2 and FBN4

As a starting point, it was decided to study the most general characteristics of FBN1, FBN2 and FBN4 such as amino acid composition, isoelectric point and hydrophobicity index. Sequence alignments were also performed in order to find common conserved characteristics. To do this, the tomato FBN1, FBN2 and FBN4 amino acid sequences and those of the Arabidopsis and rice orthologs were retrieved and plastid transit peptides predicted using the ChloroP online server (Emanuelsson et al., 1999).

From the amino acid composition a few traits common to FBN1, FBN2 and FBN4 were found:

-Histidine, tyrosine, methionine and cysteine were clearly underrepresented in all three fibrillins.

-Serine and threonine tended to be overrepresented in FBN1.

-Glutamate, proline and serine tended to be overrepresented in FBN2.

-Aspartate, glycine, leucine and proline tended to be overrepresented in FBN4.

Tables showing the amino acid composition of tomato, Arabidopsis and rice FBN1, FBN2 and FBN4 can be found in annex III.

Upon subsequent analysis using the ProtParam tool, it was found that tomato FBN1, FBN2 and FBN4 globally have a negative charge. While FBN1 and FBN2 has a high global negative charge, FBN4 lags somewhat behind with an overall lower negative charge (Table 14). In all cases, charged residues were distributed along the whole sequence. The high number of charged amino acids, including both negatively and positively charged residues, was striking for proteins reported to be membrane-bound (Singh and McNellis, 2011). However, on the other hand, the net negative charge could be in consonance with the proposed function of fibrillins as surface elements preventing plastoglobule coalescence (Ytterberg et al., 2006).

	Solanum lycopersicum			Arabidopsis thaliana				Oryza sativa		
	FBN1	FBN2	FBN4	FBN1a	FBN1b	FBN2	FBN4	FBN1	FBN2	FBN4
Length	280	297	222	263	231	323	212	265	328	226
Negative charge	36	48	32	36	32	53	28	36	51	30
Positive charge	27	26	25	26	19	23	25	24	21	25
% charged residues	22,5	24,9	30,5	23,6	22,1	23,5	25	22,6	22	24,3
GRAVY	-0,206	-0,332	-0,171	-0,239	-0,074	-0,32	-0,156	-0,103	-0,384	-0,16
pI	4,83	4,37	4,93	4,69	4,44	4,18	5,32	4,64	4,13	5,03

Table 14. Some features of FBN1, FBN2 and FBN4 from tomato, Arabidopsis and rice. The total length is displayed after removal of the transit peptide. Number and overall percentage of charged amino acids is indicated. GRAVY, grand average of hydropathicity (negative values mean hydrophilicity). pI, isoelectric point.

From the analysis and comparison of the physico-chemical properties of the corresponding fibrillin orthologs in Arabidopsis and rice it could be seen that the overall characteristics of FBN1, FBN2 and FBN4 tended to be conserved among species. The main common characteristics were found to be the following:

-FBN1 tends to be negatively charged with a net charge around the 10. The protein is hydrophilic in overall.

-FBN2 tends to have a high net negative charge (around double than FBN1). The protein itself is very hydrophilic, and has the lowest pI of all three.

-FBN4 has a negative charge of slightly variable magnitude, but is closer to neutrality. In consonance, it has the highest pI.

In search for structural common traits among FBN1, FBN2 and FBN4, the corresponding

amino acid sequences were aligned (Figure 33). A number of conserved zones were found, of which two were more closely packed, and more conserved. The first of these was a stretch of 10 amino acids curiously conserved in a discontinuous or "striped" fashion. The second was a 15 amino acid stretch including a RGD sequence. The RGD sequence has been reported as a protein binding motif (Ruoslahti, 1996), though this is found in the extracellular matrix of animal proteins.

```
FBN4_    1 QSGIDEVTF-----------------------IEP--PGSKE----------------
FBN1_    1 AISVKEFTYPKPKFTAQATNYDK----EDEWGPEVEKISPGGVA-VVDEEPP--------
FBN2_    1 --SSNEPEKPTSITPDNDKADDKMVEFVDEWGEKSEP-EPGPVTKLADSDPPEYDDEWGN

FBN4_   18 ---------AEAELIGSLKLKLLSAVSGLNRGLAASEDDLKKADEAAKELESCAGAVDLA
FBN1_   48 ---------KEPSEIELLKKQLADSFYGTNRGLSASSETRAEIVELITQLESKNPNPAPT
FBN2_   58 GSPGVDVSGEEDEKLLELKRCLVDTVYGTDFGFRASSEIRAEALEFVCQLEAANPSPAPT

FBN4_   69 ADLDKLQGRWKLIYSSAFSSRTLGGSRPGPPTGRLLPITLGQVFQRID------------
FBN1_   99 EALTLLNGKWILAYTS------FSGLFPLLSRGNLLLVRVEEISQTIDSESFTVQNSVVF
FBN2_  118 ESPELLDGNWILVFTA------FSELLPLLAVGNIPLLKVEKISQDVNTNSSTIENATTL

FBN4_  117 ---VLSKDFDNIVELELGAP-----------WPFPPVEATATLAHKFELIGS----STIK
FBN1_  153 AGPLATTSISTNAKFEVRSPKRVQIKFEEGIIGTPQLTDSIVLPENVEFLGQKIDLSPFK
FBN2_  172 SSPVATLSFSATATFEVRSPSRIQVEFKEGNFKPPEIKSNINLPETVDLFGQKISLSPVQ

FBN4_  159 IIF---EKTTVKTTGNLSQLPPLEVPRIPDQFRPPSNTGSGEFEVTYIDSDTRVTRGDRG
FBN1_  213 GLITSVQDTASSVAKSISSQPPIKFPI-------SNNNAQSWLLTTYLDDELRISRGDAG
FBN2_  232 QSLGPLENAVAGIARTISGQPPLKIPI-------PGGRTKSWLLTTYLDKDMRISRGD-G

FBN4_  216 ELRVFVIS-------
FBN1_  266 SVFVLIKEGSPLLKP
FBN2_  284 GLFVLVKEGSSLLY-
```

Figure 33. Alignment of tomato FBN1, FBN2 and FBN4.

The most conserved sequence patches are summarized by the sequences:

-LxGxWxLxxyt, the most conserved patch closer to N-terminal.

-TYvDxxxRitRGD, the most conserved patch closer to C-terminal.

4.6.2. Prediction of the secondary structure of FBN1, FBN2 and FBN4

To further characterize FBN1, FBN2 and FBN4 their secondary structure was predicted in the tomato proteins. The obtained results revealed an interesting regularity of the secondary structure elements among the different fibrillins tested. It was found that an obvious common feature was a stretch of two alpha helices close to the N-terminal end. A series of beta chains located downstream the alpha helices were another common trait, though their lengths and

distributions were found to be variable, not allowing to get straightforward conclusions.

The identification of the conserved presence of two alpha helices prompted us to compare the helices found with the corresponding Arabidopsis and rice proteins to check if this feature was conserved. The corresponding sequences were thus aligned.
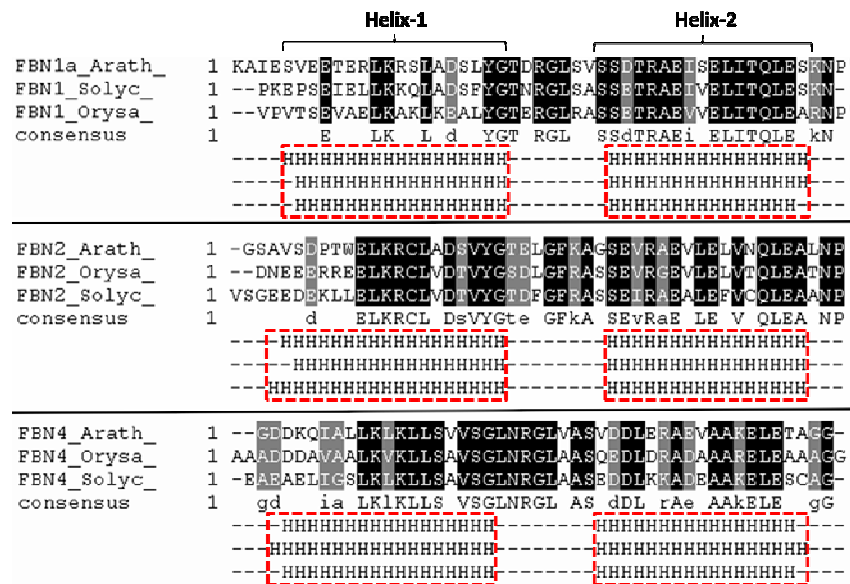
Figure 34. Alignment of the fibrillin-characteristic alpha helices present in Arabidopsis, rice and tomato FBN1, FBN2 and FBN4.

It can be observed in figure 34 that the second alpha helix and the sequence between the two helices are highly conserved in all cases. The first alpha helix of FBN1 seems to have the least conserved overall amino acid sequence, specially at the start the helix. However, the repeated presence of two alpha helices in the three fibrillin types led us to consider those helices as elements possibly playing a role in membrane binding. However, even if this were not the case, they would likely be playing role in the function of these proteins. From here on these alpha helices will be referred to as Helix-1 and Helix-2.

4.6.3. Searching mechanisms for the binding of FBN1, FBN2 and FBN4 to the plastoglobule membrane

The function of FBN1, FBN2 and FBN4 in plastoglobules was expected to be mediated, or at least closely related, to the membrane-binding capacity of these proteins. Several work hypotheses were opened while considering possible mechanisms of membrane-binding of these

fibrillins: i) the involvement of transmembrane alpha helix(es); ii) the involvement of amphipathic alpha-helices capable of membrane interaction, iii) the formation of a beta barrel-like transmembrane or membrane-associated structure, or iv) some kind of membrane association mediated by other proteins.

Regarding the last possibility, the ability of FBN1, FBN2 and FBN4 to bind plastoglobules, being these fibrillins the major protein components of plastoglobules, seemed to make anchoring to membranes via intermediary proteins unlikely. An additional question regarding all these considerations, would be whether the thylakoidal and plastoglobular membrane-binding mechanisms would be similar or different, as the plastoglobule only has an outer hemi-membrane surrounding a central hydrophobic core (Bréhélin et al., 2007).

From hydropathy plots obtained for FBN1, FBN2 and FBN4 (Figure 35) it was found that a 27 amino acid long hydrophobic stretch was present in FBN1, followed by a shorter 17 amino acid long hydrophobic sequence stretch. In FBN2, a similar situation was found. Again, two relatively close hydrophobic sequences were present, this time of 28 and 16 amino acids long. In FBN4, the strongest hydrophobicity peak plotted was found to be coincident with one of the possible amphipathic helices (see below). In addition, a 31 amino acid-long hydrophobic patch is found further downstream of that sequence, which also displays relevant hydrophobicity. However, the irregularity of the plot at this patch suggests that this likely is not a simple hydrophobic membrane anchor. However, it was quickly discarded that any of these sequences could be able to form any transmembrane structure due to the presence and overall distribution of charged and other significantly polar amino-acids. This was further confirmed by the prediction of no transmembrane helix capabilities by the online TMHMM tool on the zones forecast to have any relevant-sized alpha-helix.
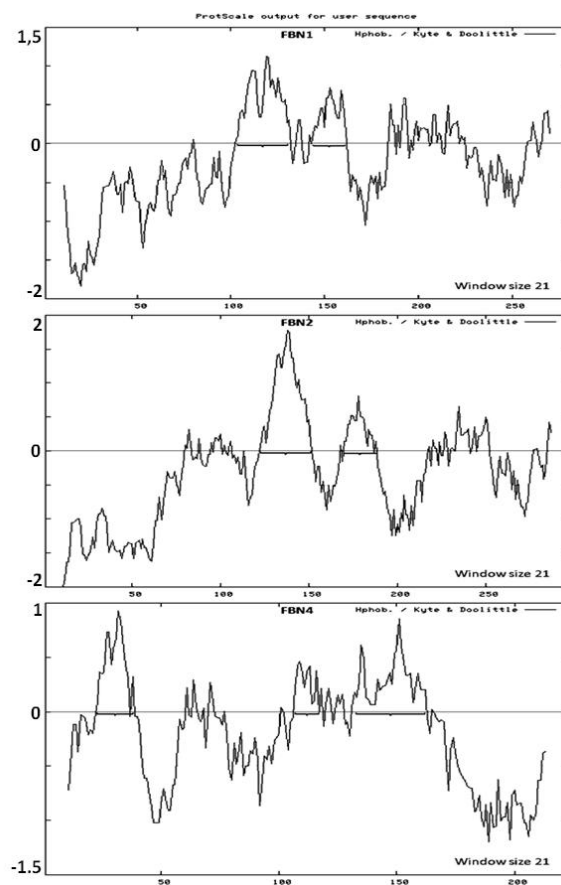
Figure 35. Hydropathy plot of FBN1, FBN2 and FBN4 of tomato. A 21 amino acid window size was used. Vertical axis indicates hydrophobicity score according to Kyte and Doolittle (positive is hydrophobic). Horizontal axis indicates amino acid position. Note the few sustained hydrophobic sequences, as shown underscored.

By means of the HELIQUEST online tool (Gautier et al., 2008) wheel plots were obtained for Helix-1 and Helix-2 (Figure 36). Upon observing the helices, no very long and clean-cut amphipathic helix was found. Nevertheless, in all three proteins at least one helix hinted at the possibility of having an amphipathic behaviour (though of weak entity). In FBN1, amphipathicity seemed to be insinuated in a short stretch of Helix-1. In the case of FBN2, both Helix-1 and Helix-2 appeared to have the possibility for a short length of amphipathicity. Interestingly, the spatial distribution of the charges in Helix-2 of both FBN1 and FBN2 is very similar. FBN4 displayed a possible amphipathic behaviour in both Helix-1 and Helix-2. Helix-1 is the only positively charged helix, has the highest number of hydrophobic amino acids of the six helices studied and displays one sustainedly hydrophobic side. Helix-2 did boast the presence of up to ten charged amino acids out of sixteen, with the highest overall negative charge of all six helices analyzed. These charges were distributed in such a way that this helix had, as well, the

106

strongest charge asymmetry found.

Nonetheless, it cannot conclusively be affirmed that any of these helices will truly display an amphipathic behaviour.

	Net charge	% charged residues
FBN1 Helix 1	-1	29,4
FBN1 Helix 2	-3	31,3
FBN2 Helix 1	-3	47,4
FBN2 Helix 2	-3	31,3
FBN4 Helix 1	1	17,6
FBN4 Helix 2	-4	62,5

Table 15. Net charge and percentage of charged residues in each helix according to HELIQUEST calculation.
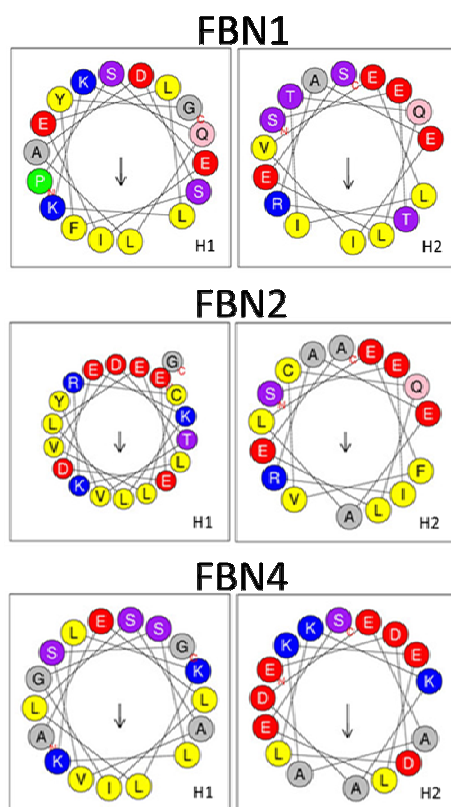
Figure 36. Wheel diagram of Hellix-1 and Helix-2 of FBN1, FBN2 and FBN4 of tomato. Red, negatively charged; blue, positively charged; yellow, hydrophobic; gray, small; purple or orange, polar uncharged; green, proline. Note the charge distribution similitude of the second helix of FBN1 and FBN2.

With the BOCTOPUS online tool it was tested if FBN1, FBN2 and FBN4 could form a beta-barrel transmembrane domain. From the results obtained, it could be concluded that both in FBN1 and in FBN2 only two beta-chains capable of belonging to a trans-membrane beta barrel were present. FBN4 displayed six beta-chains long and hydrophobic enough to belong to a transmembrane beta-barrel. However, as the minimal size of a beta-barrel is of eight beta-chains, a transmembrane beta-barrel was also discarded for FBN4. However, this result does not exclude the possibility for these proteins forming non-transmembrane beta-barrels.

Taking together the results of all the previous predictions, it can be proposed that no transmembrane alpha-helix or transmembrane beta-barrel are predicted either for FBN1, FBN2 nor FBN4. Weak amphipathic helices could be present and, according to hydropathicity plots, a few hydrophobic patches could be found, which could participate in membrane binding. Nevertheless, no conclusive results could be obtained. Other membrane-binding mechanisms could be acting, for instance relatedly to specific 3D protein folding.

4.6.4. Studies involving all members of the fibrillin family

The inconclusiveness of the previous bioinformatic analysis results regarding FBN1, FBN2 and FBN4, linked to the finding that practically no information was available on the structure nor on the way of action of fibrillins, prompted us to study the fibrillin family as a whole. On the other hand, it was intended as well to fill a gap in the knowledge of these proteins, helping to better understand what characterizes them as a family, and what their characteristic structure is like.

To start with, it became necessary to clarify some of the nomenclature used in bibliography and in database registries, an issue largely solved by the first systematic classification of the whole family made by Singh and McNellis (2011). However, generating confusion, other nomenclatures were also encountered and are shown in Table 16. During the repetitive BLAST searches, which helped clarify the equivalences, an unnamed fibrillin was found and termed as FBN12. To date, only in *Triticum aestivum* this protein has been annotated as a fibrillin. Later, this fibrillin was found to be termed as a "fibrillin-like" protein in Arabidopsis (Lundquist et al., 2012).

Fibrillin	Plastid-lipid associated protein	Plastoglobulin	Other
FBN1a - FIB1a	PAP1	AtPGL35	Plastid lipid associated protein CHRC
FBN1b - FIB1b	PAP2		
FBN2 - FIB2	PAP3	AtPGL40	
FBN3a - FIB3a	PAP4	AtPGL25	
FBN3b - FIB3b	PAP5		
FBN4 - FIB4	PAP6	AtPGL30.4	Harpin-binding protein
FBN5 - FIB5	PAP7		
FBN6 - FIB6	PAP8		
FBN7a - FIB7a	PAP9	AtPGL34	
FBN7b - FIB7b	PAP13	AtPGL30	
FBN8 - FIB8	PAP10		
FBN9 - FIB9	PAP11		
FBN10 - FIB10	PAP12		
FBN11 - FIB11	PAP14		OBP3-responsive gene
FBN12 - FIB12	*Proposed: PAP15*		

Table 16. Nomenclature equivalences of Arabidopsis fibrillins. The most systematic and recently accepted nomenclature is that at the left column. The novel FBN12 has been added to this nomenclature.

For the first *in silico* analyses it was decided to retrieve a fibrillin member of every sub-family available for *Arabidopsis thaliana, Solanum lycopersicum, Oryza sativa, Populus trichocarpa and Vitis vinifera*. After removal of the predicted transit peptide the protein sequences were compared against the corresponding Arabidopsis orthologs, showing in all cases a relatively high level of identity/similarity (Table 17).

	Arabidopsis thaliana	*Solanum lycopersicum*	*Oryza sativa*	*Populus trichocarpa*	*Vitis vinifera*
FBN1	100	71 / 82	63 / 71	75 / 88	73 / 85
FBN2	100	60 / 74	61 / 74	65 / 80	72 / 84
FBN3	100	65 / 78	60 / 76	68 / 83	70 / 83
FBN4	100	69 / 81	75 / 87	69 / 83	68 / 80
FBN5	100	62 / 83	54 / 72	61 / 77	58 / 77
FBN6	100	70 / 85	67 / 80	71 / 87	71 / 87
FBN7	100	66 / 82	63 / 81	73 / 85	69 / 84
FBN8	100	71 / 87	69 / 82	59 / 70	75 / 87
FBN9	100	71 / 82	60 / 73	72 / 84	72 / 82
FBN10	100	69 / 83	56 / 72	72 / 85	70 / 83
FBN11	100	66 / 79	59 / 75	73 / 85	70 / 81
FBN12	100	72 / 84	72 / 85	78 / 89	71 / 81

Table 17. Comparison of different plant fibrillins against *Arabidopsis*. Numbers indicate the percentage of identity (left) and similarity (right)

From these data, it can be concluded that the degree of conservation of these proteins is high among each type of fibrillin in the different plant species. In a comparison of fibrillins a diversity trait makes itself evident if we heed attention to their length (Table 18). This length

seems to be relatively conserved in each fibrillin type but may highly differ among particular types.

	Arabidopsis thaliana	Solanum lycopersicum	Oryza sativa
FBN1	263 / 251	280	265
FBN2	323	297	328
FBN3	192 / 189	184	213
FBN4	212	222	226
FBN5	212	202	204
FBN6	188	183	184
FBN7	255 / 251	250	246
FBN8	244	235	239
FBN9	187	182	205
FBN10	354	352	352
FBN11	618	583	611
FBN12	197	205	206

Table 18. Length of a mature fibrillin protein of each type for *Arabidopsis thaliana*, tomato and rice fibrillins.

Further looking into the similarities and differences between the different types of fibrillins, it was decided to compare (using BLAST) all fibrillins present in a particular plant species. Being our plant model, tomato was chosen in this case (Table 19).

	FBN1	FBN2	FBN3	FBN4	FBN5	FBN6	FBN7	FBN8	FBN9	FBN10	FBN11	FBN12
FBN1		45	21	21	23	3	16	17	16	20	7	3
FBN2	-		21	22	17	8	16	14	2	20	5	9
FBN3	-	-		59	29	20	22	11	22	28	19	10
FBN4	-	-	-		22	21	4	10	18	22	13	5
FBN5	-	-	-	-		19	0	2	21	26	18	0
FBN6	-	-	-	-	-		2	14	8	23	15	13
FBN7	-	-	-	-	-	-		4	3	21	10	3
FBN8	-	-	-	-	-	-	-		11	5	5	3
FBN9	-	-	-	-	-	-	-	-		21	5	14
FBN10	-	-	-	-	-	-	-	-	-		11	7
FBN11	-	-	-	-	-	-	-	-	-	-		0
FBN12	-	-	-	-	-	-	-	-	-	-	-	

Table 19. Identities (expressed as %) found among the different tomato fibrillins

Taken together, the results shown in Tables 17, 18 and 19 allow us to conclude that the fibrillin family has a number of sub-types which are conserved among themselves but which differ considerably with regard to each other sub-type. The two examples of most closely related

fibrillins are FBN1-FBN2 and FBN3-FBN4, which have an identity of 45 % and 59 %, respectively. Many other members of the family are highly diverged as far as sequence is concerned. A number of especially dissimilar fibrillins are found in which sequence homology is almost non-existent.

It was pondered if looking deeper into the conservancy of these proteins would yield likewise results for the generality of plants. As the available studies on these proteins did not seem to answer explicitly and definitely what is the characteristic common to all fibrillins, it was decided to bring in-silico research further than planned. A sequence mining was done in search of as many sequences as possible of each fibrillin sub-family. Upon compilation of sequences of each fibrillin type, a global analysis of each could be carried out, and the common traits to each fibrillin subfamily would become clear. From there on, predictions of secondary and, afterwards, tertiary structure would follow.

The reference Arabidopsis sequences were all taken from the review by Singh and McNellis (2011) and the search of other plant or plant-related fibrillins was initially performed with NCBI BLAST and EXPASY blast. Upon the finding of insufficient exhaustiveness of those databases, a handful of different generalistic curated databases, and other specific and possibly less curated databases were used.

The search was conducted prioritizing the inclusion of fibrillins from plants with complete genomes. This would allow to discriminate if all land plants could be expected to contain a representative of every member of the family or not. This approach was made possible by the existence of many genome project websites, and later on made easier by means of several "hub" genomic resource projects among which Phytozome proved to be one of the most valuable.

As far as land plants were concerned, sequences were retrieved for *Amborella trichopoda, Aquilea coerulea, Arabidopsis thaliana, Arabidopsis lyrata, Brachypodium dystakyon, Brassica napus, Brassica rapa, Camelina sativa, Capsella rubella, Capsicum annuum, Carica papaya, Catharantus roseus, Citrus clementina, Citrus sinensis, Coffea canephora, Cucumis melo, Cucumis sativus, Eucaliptus grandis, Fragaria vesca, Glycine max, Gossypium raymondii, Hordeum vulgare, Linum usitatissimum, Malus domestica, Manihot esculenta, Medicago truncatula, Mimulus guttatus, Musa acuminata, Oryza sativa, Oncidium (hybrid), Panicum virgatum, Phaseolus vulgaris, Phyllostachys edulis, Physcomytrella patens,*

Picea glauca, Populus trichocarpa, Prunus persica, Ricinus comunis, Solanum demissum, Solanum phureja, Solanum tuberosum, Sorghum bicolor, Theobroma cacao, Thellungiella parvula, Vitis vinifera and *Zea mays*. Sequences from over forty cyanobacteria, and some algae, diatoms and stramenopiles were also collected during the search. Except for cyanobacteria and higher plants, available completed-genome sequences were not abundant.

During the process of the sequence compilation the following conclusions were reached:

-It was found that plant fibrillins have at least two distant relatives in cyanobacteria, the one being similar to FBN1 and the other, to the previously unnamed FBN12. We have named these cyanobacterial sub-families as cFBN1 and cFBN12.

- With very few exceptions, land plants have at least one representative of every fibrillin type. Of the exceptions initially found, later on, most of them were overruled. Only the absence of FBN5 in the Cucumis species studied and a case of a predecessor of modern wheat, *Aegilops tauschii*, with a particular architecture for FBN11 remained as discordant.

-In many instances more than one sequence of each family member was found, occasionally those replicates being only fragments of the full protein. When multiple family members were found, only the closest relative to one Arabidopsis representative was chosen.

-Arabidopsis possesses three fibrillin subfamilies with two reported representatives: FBN1a/FBN1b, FBN3a/FBN3b and FBN7a/FBN7b (Singh and McNellis, 2011). From our searches it could be concluded that the presence of those specific duplications was not a trend generally followed in other species. This was true even if those other species had several forms of these or of other subfamilies. Taking this into account, to avoid amplification of species-specific bias, during sequence compilation it was decided to select only one sequence of each fibrillin sub-family per species. This was done so even if a plant had multiple different sequences for a specific subfamily. When that situation did happen, the sequence closest to any one Arabidopsis representative was chosen.

-Two of the fibrillin family members had an amino acid sequence notably longer than the rest: FBN10 and FBN11. The difference was found to be explained in FBN10 by this protein having a double-domain fibrillin, as was found during repeated NCBI BLAST searches. During the same BLAST searches, FBN11 was found to have a protein kinase domain at N-terminal, distinguishing it from all the rest of fibrillins.

-It was found that all fully sequenced-genome photosynthetic organisms investigated

possessed at least one fibrillin. Moreover, in the wide majority of cases a multiplicity and variety of them was found. The subfamilies present and the similitude of fibrillins from other organisms, if compared to higher plants, could be variable. A handful of hardly classifiable fibrillins were found with only residual relationship to any known family. As well, a number of incomplete or gapped sequences were found, which apparently stemmed from yet incompletely sequenced genomes.

From the findings during the sequence search, a table of fibrillin family conservation or presence/absence was elaborated (Table 20).

	Higher plants	Mosses	Algae	Cyanobacteria	Diatoms	Cryptophytes	Apicomplexa
FBN1/CFBN1	Yes	Yes	Yes/No	Yes	*Yes/derivative*	*No/derivative*	No
FBN2	Yes	No	No/Yes	No	*No/derivative*	*No/derivative*	No
FBN3	Yes	Yes	Yes	No	*Yes/derivative*	*No*	No
FBN4	Yes	Yes	(Yes)	No	*No/derivative*	*No*	No
FBN5	(Yes)	Yes	(Yes)	No	*No/derivative*	*No*	No
FBN6	Yes	Yes	(Yes)	No	*No/derivative*	*derivative*	No
FBN7	Yes	Yes	No	No	*Yes/derivative*	*Yes*	No
FBN8	Yes	Yes	(Yes)	No	*No/derivative*	*derivative*	No
FBN9	Yes	Yes	(Yes)	No	*No/derivative*	*No*	No
FBN10	Yes	Yes	Yes	No	*Yes (one dom.)*	*Yes*	No
FBN11	Yes	(Yes)	No	No	*No*	*No*	No
FBN12	Yes	Yes	(Yes)	(Yes)	*(Yes)*	*Yes*	No

Table 20. Occurrence of fibrillins in diverse plastid-containing organisms. Parentheses indicate that the specific fibrillin is not found in all species of the category. Derivative indicates the finding of a remote homologue.

In all higher plants with a complete or near-complete sequenced genomes at least one representative of each of the twelve fibrillin subfamilies was found. This suggested that most, if not all of them, had actually a function to fulfil inside the plant. However, ultimately one exception was found: FBN5, found incomplete in more instances than all the rest, and definitely not found in cucurbitaceae. Until the finding of the absence of FBN5 in both *Cucumis melo* and *Cucumis sativus*, it was still considered possible that the incompleteness of the sequencing of some genomes could be hindering our finding of FBN5 sequences in some plant species. However, cucumber and melon have both completed genomes. From the alignments performed, the highest sequence self-conservation degree in the archetypal fibrillins was found in FBN1 and FBN4, with varying, but elevated degrees of conservation in the rest of fibrillins. The modular fibrillins FBN10 and FBN11 displayed a clear partition of conservancy. FBN10 was highly conserved in the first fibrillin module, while the second module was far less conserved among

different FBN10s. FBN11 had a very highly conserved zone in the fibrillin domain, but conservedness was in overall lower at C-terminal; that is, at the fibrillin domain.

A smaller number of the twelve sub-families was represented in algae and mosses. A case of note is that in the mosses *Physcomitrella patens* and *Sellaginella moellendorfii*, FBN2 seemed to be absent. Interestingly, in them, the protein most closely resembling FBN2 was actually more similar to FBN1 than to FBN2. In algae, the number of represented subfamilies is even lower than in mosses, but varies depending on the species. Again, a particularity regarding FBN1/FBN2 is found: in overall, the disappearance of FBN2 is again repeated when a FBN1 relative is present. On occasion, a clear FBN1 was absent, in which case a form closer to FBN2 was found, which suggested some overlapping function of FBN1 and FBN2. Plant FBN4, FBN5, FBN6, FBN8, FBN9 and FBN12 would have relatives in some algal species, while not in others. Depending on the complexity of the organism in question, this could be a result of a simpler regulation network, or that organism having a lack of an unknown need which higher plants have to satisfy for their correct development and/or interaction with the environment.

For cyanobacteria, the case is much simpler, as almost only two fibrillin subfamilies are found. CFBN1 is found in all cyanobacteria analyzed, most likely the closest relative to the ancestral first fibrillin from which all arise. It could however be seen that in some cases FBN1 had possibly evolved to derivative forms far more distant to plant FBN1 and even cyanobacterial cFBN1. Interestingly, a part of the cyanobacteria have a cyanobacterial FBN12-like protein. Finally, as rare exceptions, in three cyanobacterial species fibrillins distantly resembling FBN3 were found.

The conservation of the number of fibrillin sub-families in diatoms *versus* land plants is wider than in cyanobacteria, but in several instances the found fibrillin relatives are very distant to plant fibrillins. As a result, to trace clear relationships is somewhat difficult. Curiously, in diatoms the fibrillin found most resembling any of its plant relatives is by far FBN10, though only the first of its two subdomains is present (FBN10 is a two-module fibrillin). The degree of conservation is very striking, and unexpected if compared to the conservation of all other fibrillins found in diatoms *versus* plant relatives. This is unusual because the most consistently conserved fibrillin found across species was generally FBN1/CFBN1 until this point. FBN1, FBN3 and FBN7 all have in diatoms at least a distant relative to their plant equivalents, but with far less identity than found for the "anomalously over-conserved" FBN10. Nevertheless, the case

of FBN7 could be significant, as this form was shown to be absent in the algae searched. In some cases a conserved FBN12-like protein is found, though not in all species searched. No FBN11-like protein is found. A number of difficult to assign fibrillins are present, possibly in their way towards divergent evolution.

In cryptophytes, the landscape has some similarities to what is found in diatoms. FBN7, FBN10 and FBN12 are found in all cryptophytes searched, as in diatoms, but all in forms pretty distant to plants. FBN1, FBN6 and FBN8 have in some instances forms very slightly related to those of higher plants, which may in fact be becoming already something else.

4.6.5. Peculiarities in the architecture of some fibrillins

During the search and exploration of available fibrillin sequences, a number of fibrillins were found to have double or multiple domains. All of this has given a number of clues on the possible function of several fibrillins.

Of the found modular fibrillins, as has been mentioned, two were standard fibrillins which normally display a composite architecture: FBN10 and FBN11. FBN10 has a concatenation of two fibrillin domains, the first seemingly related to FBN3. FBN11 had a N-terminal kinase domain linked to a fibrillin domain, apparently related to FBN6.

Other than these standard modular fibrillins, a number of rare cases were found combining fibrillin domains with other protein modules. Twenty six unusual modular proteins were found. Of these proteins with extra modules, at least in six cases those modules seemed to potentially have a regulatory or signalling role. Four of the remaining proteins seemed to have protein-protein interaction domains, while other four had an unknown identity or function. The remaining five had a small variety of related functions; among those, two proteins of equivalent architecture related to redox homeostasis. With regard to these modular fibrillins, it was found that FBN12 and FBN3 seemed to be the most frequent partners for modular protein formation. A table displaying the different architectures of modular fibrillins is available in Annex IV.

The finding of twenty six modular fibrillins, several of them in consolidated genome drafts would make these cases rare, but not completely exceptional. Furthermore, upon reviewing the number of different combinations and anomalies found, it seems that FBN3 and FBN12 have the highest number of composite forms, in combinations which could be suggestive or some kind of regulatory role.

Some other cases of unusual fibrillin or fibrillin-related proteins were also found in other

organism. In the photosynthetic marine prokaryote *Prochlorococcus marinus*, a protein termed "pili assembly chaperone" was found to be very distantly related to FBN4. It had unannotated relatives among other unicellular algae, and noteworthily, upon performing a BLAST search, a "fibrillin domain" is detected despite the very weak similitude to any fibrillin, even to FBN4. The last modular *Ostreococcus tauri* fibrillin mentioned previously (outside the table), is one relative of this protein. Thus, both these fibrillins would pertain to a new branch or branches of yet to be classified fibrillins.

The unicellular green alga *Coccomyxa subellipsoidea* seems to have several atypical fibrillins as well: A FBN8 and a fibrillin distantly resembling algal and cyanobacterial FBN1/CFBN1. Both these non-canonical fibrillin forms are lacking the typical two initial alpha-helix stretches at the N-terminal region. Both appear to preserve the number of eight predicted beta chains in the remnant of the secondary structure.

Thellungiella parvula, a plant with known high tolerance for growing in heavy metal-enriched soils, has an unusual FBN11 with a deletion in the middle of the sequence.

One of the ancient progenitors of modern wheat, *Aegilops tauschii*, displays a FBN11 without its commonly associated kinase domain. This kinase domain seems nevertheless to exist, but parted from its usual partner. *Populus trichocarpa* has also this same setup, with the difference that a standard two-domain FBN11 is also present.

4.6.6. Findings derived from the alignments of the different fibrillin types

The members of each fibrillin family were aligned by means of MEGA 5.2, a free sequence analysis package (Tamura et al., 2011). The MUSCLE alignment algorithm was chosen instead of CLUSTAL, due to its reported higher speed and accuracy, specially for difficult alignments (Edgar, 2004). Alignments were conducted for each of the twelve plant fibrillin sub-families, as well as cyanobacterial fibrillin cFBN1. It was decided to circumscribe alignments only to plant and cyanobacterial sequences, as both of those were the most abundantly represented sequences. Consensus sequences were elaborated aiming for elucidating what the average fibrillin would be characterized by. During the alignments, in the occasions when several closely related plant species were found to have identical proteins, the number of representatives from those species to be aligned was trimmed down to only one. This was expected to remove some overly bias from any plant family. Allowing these sequences to be aligned together would possibly have given them too much a specific weight taking into account

the number of fibrillin sequences aligned. This was implemented in several occasions with three Citrus, three Solanum and two Cucumis species when found appropriate. In some cases, these sequences together could have accounted for over 15 % of the sequences aligned if some of them had not been discarded.

It was found that a high conservation within each sub-family was a common trend. During alignment, a handful of genome projects seemed prone to provide sequences with deletions at the start or the end of the sequence, discordant insertions, as well as occasional characters representing unknown amino acids. To remove distortion and facilitate interpretation during the consensus elaboration, such sequences with significant deletions, anomalous insertions or additional modules were deleted from the alignments. A version of the final alignments adapted to printed format is displayed in annex VII.

The obtained alignments were further analyzed to generate reference consensus sequences using two alignment analysis tools: Advanced Consensus Maker (ACM, 2012), and VISCOSE Consensus (Spitzer et al., 2004). First, the alignments were processed through ACM to yield a gapped, partial consensus. ACM was configured to select a consensus amino acid for a given position if 80% of the aligned amino acids in those positions were conserved, as this was deemed reasonably cautious. The alignment was also introduced into the VISCOSE Consensus analyzer for further manual analysis and consensus discrimination. To aid this, VISCOSE was configured to translate each amino acid to a simplified alphabet representing characteristic chemical properties of amino acids. Thus, final consensus sequences were elaborated using the ACM partial consensus as a scaffold whose gaps were manually filled by means of the information visualized with the VISCOSE consensus tool. To reduce as far as possible the impact of variable subjectivities, a number of consensus elaboration rules were imposed, which can be found in the Materials and Methods section (see section 3.9.2).

A critical point happened to be the decision of the N-terminal residue of the mature proteins. It could be argued that sequence conservation would clarify this. Nonetheless, a clear border remained blurred in several cases. It was expected that predicting the transit peptides of the real proteins aligned would give an orientation. This did happen only partly, as transit peptide prediction did yield sometimes quite variable forecasts depending on the species. Upon completion of the consensus sequences, as a means of quality control, the obtained consensus sequences were launched in BLAST against the public databases in the expectation that they

would be very close to many of their sister fibrillins. This was indeed found to be the case.

Thirteen consensus sequences were obtained: twelve corresponding to higher plants and one to cyanobacteria (see annex V). These sequences were primarily analysed using the bioinformatic tools previously used for FBN1, FBN2 and FBN4 (see section 3.9.2). The main physicochemical parameters obtained for each consensus sequence are summarized in Table 21.

	CFBN1	FBN1	FBN2	FBN3	FBN4	FBN5	FBN6	FBN7	FBN8	FBN9	FBN10	FBN10(A)	FBN10(B)	FBN11(B)	FBN12
Length	194	259	276	190	203	190	188	252	224	175	338	183	155	259	179
Negative charge	23	37	46	24	26	23	25	33	23	23	41	20	21	27	22
Positive charge	25	22	22	30	24	24	24	26	24	21	38	21	17	36	31
% charged	24,7	22,8	24,6	28,4	24,6	24,7	26,1	23,4	21,0	25,1	23,4	22,4	24,5	24,3	29,6
GRAVY	-0,131	-0,154	-0,212	-0,424	-0,135	-0,183	-0,134	-0,075	0,087	-0,199	-0,216	-0,139	-0,208	-0,306	-0,284
pI	8,6	4,51	4,29	9,49	5,56	7,71	7,84	4,94	8,04	5,5	5,5	7,93	5,07	9,58	9,61
Overall charge:	2	**-15**	**-24**	6	-2	1	-1	**-7**	1	-2	-3	1	-4	**9**	**9**

Table 21. Protparam-calculated physicochemical properties of consensus sequences. GRAVY, grand average index of hydropathicity (negative values mean protein is hydrophylic); pI: isoelectric point.

The obtained results confirm that FBN1 and FBN2 display a high net negative charge, followed by FBN7. FBN4, which in tomato was found to be negatively charged (see section 4.6.1), was here again found to be charged, but closer to neutrality. FBN3 and FBN12 exhibit a clear positive overall charge. The fibrillin domain of FBN11 (FBN11-B) has also a markedly positive charge.

The obtained consensus sequences were subjected to a phylogenetic analysis, giving an overview of the evolutionary distance leading to the diversity of fibrillins currently found in nature. From the results shown in Figure 35 it could be stated that the closest fibrillins among each other were the plant FBN1 and FBN2. The protein least diverged from original ancestor is the cyanobacterial cFBN1. The plant proteins less diverged from the common ancestor are, in this order, FBN5, FBN1 and FBN4.

Several divergent evolutionary branches can be observed. Grossly, we can divide those in at least three distinct evolutionary branches:

-The group of cFBN1, FBN1, FBN2, FBN3, FBN7 and FBN10 (domain1). More distantly, FBN6 and FBN10 (domain 2) also can be included in this broad grouping.

-The group of FBN4 and FBN8.

-The group of FBN5, FBN9, FBN11 and FBN12.
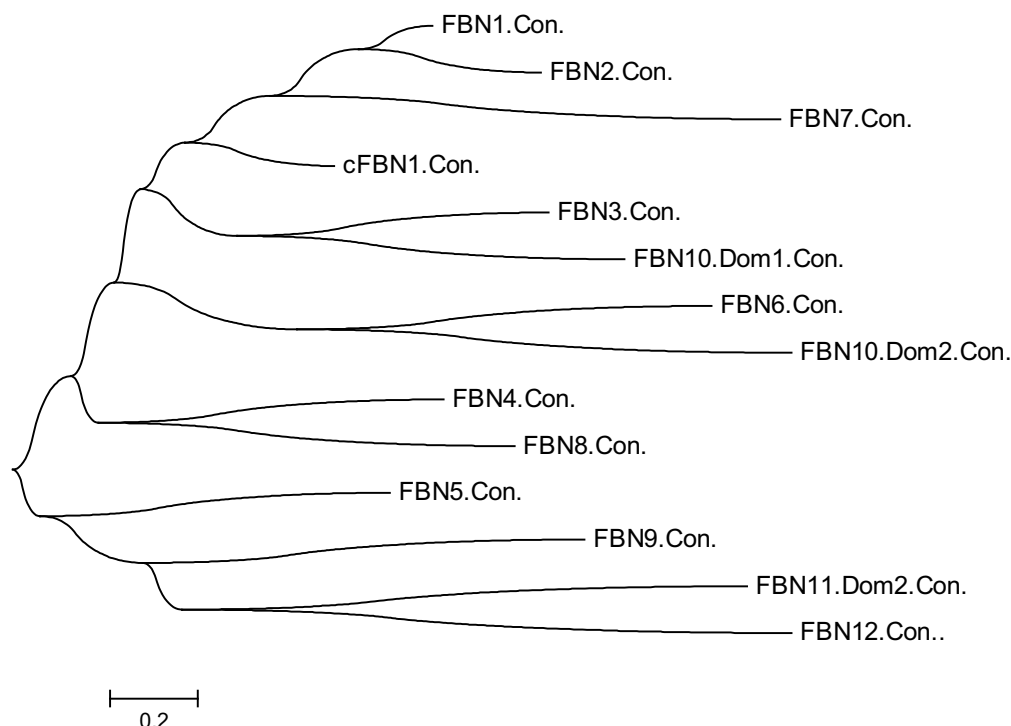
Figure 37. Maximum likelihood phylogenetic tree constructed for the consensus sequences from a default MUSCLE alignment with MEGA 5.2. Modular fibrillins have had their fibrillin subdomains considered individually (FBN10: Dom1, Dom2; FBN11: Dom2). The X axis indicates evolutionary distance, the branching points indicate a common ancestor for its subsidiary branches. The common ancestor is located at the root of the diagram, from where all branches arise.

The amino acid composition, hydropathicity plots and the predictions for transmembrane helices and transmembrane beta-barrels for each one of the thirteen consensus sequences were analyzed as was done previously for the tomato sequences. Amino acid composition, hydropathicity plots and helix wheel diagrams are shown in annex VI. The main conclusions reached from the analysis of these parameters are the following:

-Fibrillins are poor in sulfur-containing amino acids. Some specific fibrillins seems to be specially rich or poor in specific amino acids, but no clear connection can be readily made to function. FBN6 seemed to display several of such imbalances.

-Most fibrillins tend to be in overall hydrophilic. The fibrillins closer to breaking this rule are FBN1, FBN2 and FBN8.

-No transmembrane alpha-helices have been predicted.

-Several fibrillins seem to have a certain probability for the formation of transmembrane

119

beta-barrels. The cases with highest probability are FBN4, FBN5, FBN6, FBN10 and FBN12.

-The alpha-helices detected tend to be very charged. Helix-2 is more charged, as well as more conserved, than the first one.

The consensus sequences were in turn aligned to find common traits among them. The similarities found among the different fibrillin types was limited and circumscribed to two short sequences (Figure 38). One of the conserved sequences (named FCR1, for fibrillin conserved region 1) corresponds to the short stretch of 9 amino acids with a tendency towards alternate or discontinuous conservation of its residues. This motif was already identified when comparing FBN1, FBN2 and FBN4 of tomato, Arabidopsis and rice (see section 4.6.1). The other block of conserved sequences (named FCR2, for fibrillin conserved region 2) extends over about 12 residues located in the C-terminal region of the protein. Alternate conservation of amino acid residues can also be observed in this case.

```
FBN1_Con_2_0_     1 -----IERLKKALVDSFYGTD-RCLSA----SSET--RAELVELITQIEAKNPTPAPTEA
FBN2_Con_2_0_     1 EVDDKLEELKRCLVDTVYGTE-LCFRA----GSEV--RAELLELVNQIEAANPTPAPVEA
FBN3_Con_2_0_     1 ---KDAKQLKEELEAIAPLD-RCAEA----TPED--QQRVDQIARKIEAVNPTKEP-LK
FBN4_Con_2_0_     1 ---DLIASLKLKLLSAVSGLN-RCLAA----SEDD--LQKADAAAKEIEAAGGPVDLSAD
FBN5_Con_2_0_     1 ------SQIKTELYQALQGIN-RCIFG----VPSA-KKSETEGLVKLIESQNPTPEPTLN
FBN6_Con_2_0_     1 ------DDLVASILSKVTGTD-RCVLL----TKEE--HKEVAEVAQELQKYC--VDEPVK
FBN7_Con_2_0_     1 ---MERLSAKESLLLAFKDAG--CFEALVTGKTTDMQRIDVNERITGIERLNPTPRPTTS
FBN8_Con_2_0_     1 ---LELENKKHELLRAVQDTQ-RCLVA----TADQ--RSSIEEALVSVEGYN--AGAPID
FBN9_Con_2_0_     1 -------SAKQHLLNLISDQD-RCLKT----QSDPSKLASTVSAIDALAARG--RDTVTT
FBN10_Con_2_0_A   1 ----SFTEAENSLIEALIGIQGRCRSA----SPQQ--LNEVESAVQVIEGLEGVPDPTSS
FBN10_Con_2_0_B   1 ----KKTEPRQKLLSAISTGT--CV-----------KEAIDEFISSNQNVA-----KDE
FBN11_Con_2_0_B   1 -VVPSNDIIRWGLGSTAVRIA-EEYIY----RQPQ--RSPLAHFIELIENLNPHSKPKNW
FBN12_Con_2_0_    1 -------QPNKEVEESVKVLK-NIAKT---RKVP--AEEVLSALSVIEKAK--LDPSGF
```

FCR1

```
FBN1_Con_2_0_    49 LTLING------KILAY--TSFSGLFPLLSRG-TLPLV---KVEEISQTIDSENFT---
FBN2_Con_2_0_    54 PELLDG------NUVLL--TAFSELLPLLAAG-SLPLL---KVEKISQSIDTSSLT---
FBN3_Con_2_0_    50 SDLING------KEELII--TTSQSILQTQRPK------FLRSIGN-YQAINVDTLR---
FBN4_Con_2_0_    51 LDKIQG------RTKIII-SSAFSSRTLGGSRPGPPTGRLLPITLGQVFQRIDVLSKD---
FBN5_Con_2_0_    49 LDKVGG------CUKII---STITILG-SKRT-KLGLRDFISLGDFFQNIDVAKGK---
FBN6_Con_2_0_    46 CPLIFG------EUDNVI--CSVPTSPGGGYRS-ALGRLFF-KTKEMIQVVEAPDI----
FBN7_Con_2_0_    56 P-FIEG------RINFEI---FGSGSPGLFAAR-FLFERFPSTLAN-LSKMDVLIKDGYA
FBN8_Con_2_0_    49 LVKIDG------TURIQI--TSAPDVLVLFEAA---ARLPFFQVGQIFQKFECRDRSDGG
FBN9_Con_2_0_    47 GDSISA------TURILI--TTEKEQLFIIEKAPLFGT----QAGDVLQVIDVEKKT---
FBN10_Con_2_0_A  51 S-LIEG------RVQLMI--TTRPGTASPIQRT-FVGVDFFSVFQEVYLRTNDPR-----
FBN10_Con_2_0_B  38 LELIEG------EUQAII--SSQTDSWLENAANG-LHGKQIVKKNGQLKFLVD-------
FBN11_Con_2_0_B  53 LELIPG------KIRIL---CTGRHIGLTLRQ-PSVRV---LIGDVHLTVSRASK----
FBN12_Con_2_0_   45 LETIGGTESPGRTWMLIF--TAEKKLKG--------GRYFPVTA---VQRFDAAGKR---
```

```
FBN1_Con_2_0_    94 -VQNSVQF-AG---PLAT------------TSI-STNAKFEVRSPKRI-QIKFEEGI---
FBN2_Con_2_0_    99 -IVNSTTL-SS---PFAT------------FSF-SASASFEVRSPSRH-QV-FKEGT---
FBN3_Con_2_0_    92 -AQNMETW------PFFN---------------QVTADLTPLNARKI-AVKFDYFK---
FBN4_Con_2_0_   102 -FDNIVELELGAPWPLPP----------VEVTATLAHKFELIGSSKI-KITFEKTT---
FBN5_Con_2_0_    95 -AVNVIKF-NARGLNLLS-----------GQL-TIEASFKIASKSRV-DITYESST---
FBN6_Con_2_0_    92 -VRNKVSF---SAFGFLD-----------GEV-SLKGKLKVLDSKWH-QVIFEPPE---
FBN7_Con_2_0_   104 KITANVKLLNSIESKF------------------ILSTKLSVEGPLRI-KEEYVEGI---
FBN8_Con_2_0_    98 IVRNVVRWSIPNLLEEQE-----------GATLLVSAKFSVVSVRNI-YLQFEEIS---
FBN9_Con_2_0_    92 -LNNVITF-------PPD-----------GVF-FVRSSIEIASPQRV-NFRFTSAV---
FBN10_Con_2_0_A  96 -VSNIVKFSDAI-------------------GEL-KVEAAASIKDGKRI-LFRFDRAA---
FBN10_Con_2_0_B  82 -ILPGLRF-------------------------SMTGNFVKSGSNTY-DVTMDDAA---
FBN11_Con_2_0_B  96 -SNTNLSFTSDIGFTVMPGRDWPHDKSGVTGKL-QVNSSFRLTAGRRI-YLKEEKTTSKF
FBN12_Con_2_0_   89 -IENGVYL--------------------GPIGCLTFEGRLSWKKRI-LAFIFERIR---
```

```
FBN1_Con_2_0_   132 --------------IGTPQLTDSIVLPENVEFLGQKIDLSPFKGLLTSVQDTAS-SVAKT
FBN2_Con_2_0_   136 --------------IQPPEIKSSVDLPENVDIFGQKISLSPVQQSLNPLQEAVA-NISRA
FBN3_Con_2_0_   125 --------------IAG-----------------------LIPIKAP--------
FBN4_Con_2_0_   146 --------------VKTTGNLSQ--------------------LPPLEVP--------
FBN5_Con_2_0_   136 --------------ITPDQL----------------------MNVFRKN--------
FBN6_Con_2_0_   131 --------------LKVG-------------------------SLEFQYG--------
FBN7_Con_2_0_   142 --------------IETPTVIEETVPEQLKGALGQAVTTLQQL--PVPIRDA--------
FBN8_Con_2_0_   142 --------------VQNIKISEQLQALIAPAILPRSFLSLQILQFIRTFKAQ--------
FBN9_Con_2_0_   127 --------------IRGKNWEIP--------------------LPPF-----------
FBN10_Con_2_0_A 132 --------------FSFK-------------------------FLPFKVP--------
FBN10_Con_2_0_B 111 --------------IIGG-------------------------PFGYP--------
FBN11_Con_2_0_B 153 SSGQSDAEESLGQKISGRKWRKAIPFKEFPSSLPVAKLVSDEIDVTMSLGDPLNVDSARN
FBN12_Con_2_0_  123 --------------IKVGP-----------------------LNPLEIS--------
```

FCR2

```
FBN1_Con_2_0_   177 ISSQPPLKFSISNSNAQSWLLTTYLDE--ELRISEG-DGGSV--FVLIKEGSPLLNP-
FBN2_Con_2_0_   181 ISGQPPLKVPIPGERTSSWLLTTYLDK--DLRISEG-DGCEL--FVLAKEGSPLLDQ-
FBN3_Con_2_0_   135 --------------GRARGELEITYLDE--ELRISEG-DKFILKMVDPSYRVPL
FBN4_Con_2_0_   162 ---RIPDALRPPSNTGSGEFEVIYLDS--DTRITEG-DRGELRVFVIS----------
FBN5_Con_2_0_   149 ------YDLLLGIFNPEGWLEITYVDE--SIRIGRD-DKGNI--FVLERSEES-----
FBN6_Con_2_0_   142 -------------GESEVKLQITYVDE--KIRIGYG-SRGEL--FVFQRRG-------
FBN7_Con_2_0_   178 -VSSGLKIPLSGSFQRLFMISYLDE--ELLIIRD-TACVP--EVLTRLETPPSSLA
FBN8_Con_2_0_   180 ------IPVSPGRRSVGGLYYLSYLDR--NLLGRAVGGCV--FVFTRAQPLEL---
FBN9_Con_2_0_   140 --------------GQGWFETVYLDD--ELRVVKD-IRGDY--LVVERAPYSWKE--
FBN10_Con_2_0_A 143 ----YPVPFRLLGDEAKGWLDTTYLSPSGNIRISEG-NKCTT--FVLQ----------
FBN10_Con_2_0_B 120 -----------LEMESKFNLELIYTDE--KIRISRG-YNNIL--FVHLRVDG-----
FBN11_Con_2_0_B 213 VLQEVRTQIPPEMFDLSKLVCGTYVDS--RILVLRG-VNCSALLFTRSCVDES-----
FBN12_Con_2_0_  135 -LGQKDDREPSTKDPFFIWF--YVDE--EIAVARGRSGGTA--FWCRCRRVT-----
```

Figure 38. Alignment of all plant consensus fibrillin types and the fibrillin sub-domains found in FBN10 and FBN11. The two regions showing higher conservation are highlighted in red and marked as FCR1 (fibrillin conserved region 1) and FCR2 (fibrillin conserved region2).

A detail of the alignments of the conserved regions FCR1 and FCR2 in the default conditions is shown in Figures 39 and 40.

```
CFBN1_Con_2_0_     1 ELLEG------NWRLLY--TTSKGLL
FBN1_Cons_2_0_     1 TLLNG------KWILAY--TSFSGLF
FBN2_Con_2_0_      1 ELLDG------NWVLLY--TAFSELL
FBN3_Con_2_0_      1 DLLNG------KWELIY--TTSQSIL
FBN4_Cons_2_0_     1 DKLQG------RWKLIY SSAFSSRTL
FBN5_Cons_2_0_     1 DKVGG------CWKLIY--STITILG
FBN6_Cons_2_0_     1 PLIFG------EWDVVY--CSVPTSP
FBN7_Cons_2_0_     1 PFLEG------RWNFEWFGSGSPGLF
FBN8_Con_2_0_      1 VKLDG------TWRLQY--TSAPDVL
FBN9_Con_2_0_      1 DSLSA------TWRLLW--TTEKEQL
FBN10_DOM1_2_0_    1 SLIEG------RWQLMF--TTRPGTA
FBN10_DOM2_2_0_    1 ELLEG------EWQMIW--SSQTDSW
FBN11_DOM2_2_0_    1 ELLPG------KWRLLY--CTGRHIG
FBN12_Con_2_0_     1 ETLGGTESPGRTWMLIF--TAEKKL-
consensus          1      l g        W l y
```

Figure 39. Alignment of the conserved sequence FCR1 of the different fibrillin consensus sequences. A consensus sequence is found at the bottom: LxGxWxLxY. Both fibrillin domains of the modular FBN10 are considered separately, and FBN11 has been aligned considering only its fibrillin (i.e. second) domain. The most conserved residue is W, as found in the alignment; a glycine and a number of other generic hydrophobic amino acids are conserved as well, in an alternate or discontinuous way.

```
CFBN1_Con_2_0_     1 WLDIT------YLDE--DLRIGRGNEGS--VFVLTK
FBN1_Cons_2_0_     1 WLLTT------YLDE--ELRISRGDGGS--VFVLIK
FBN2_Con_2_0_      1 WLLTT------YLDK--DLRISRGDGG---LFVLAK
FBN3_Con_2_0_      1 ELEIT------YLDE--ELRISRGDKGN--LFILKM
FBN4_Cons_2_0_     1 EFEVT------YLDS--DTRITRGDRGELRVFVIS-
FBN5_Cons_2_0_     1 WLEIT------YVDE--SLRIGRDDKGN--IFVLER
FBN6_Cons_2_0_     1 KLQIT------YVDE--KIRIGKGSRGS--LFVFQR
FBN7_Cons_2_0_     1 LFMIS------YLDE--EILIIRDTAGV--PEVLTR
FBN8_Con_2_0_      1 LYYLS------YLDR--NWLGRAVGGGG-VFVFTR
FBN9_Con_2_0_      1 WFETV------YLDD--EIRWVKDIRGD--YLVVER
FBN10_DOM1_2_0_    1 WLDTT------YLSPSGNLRISRGNKGT--TFVLQ-
FBN10_DOM2_2_0_    1 NLELL------YTDE--KIRISRGYNNI--LFVHLR
FBN11_DOM2_2_0_    1 MFDLSKLVCGTYVDS--RLLVLRGVNGS--ALLFTR
FBN12_Con_2_0_     1 WF--------YVDE--EIAWARGRSGG--TAFWCR
consensus          1          Yld    l i r    g      v
```

Figure 40. Alignment of the conserved sequence FCR2 of the different fibrillin consensus sequences. A consensus sequence is found at the bottom: YLDxxLxIxRxxG The most conserved amino acid is Y. An aspartic acid and a miscellaneous positively charged (mostly arginine) amino acid are very conserved as well. A number of sites conserving hydrophobic amino acids are found. Some alternate or discontinuous residue conservation can be observed. FBN10 domains are considered separately) and FBN11 has been aligned considering only its second (fibrillin) domain, removing its kinase initial domain.

A main trait in FCR1 is the conserved presence of a tryptophan residue (W) which expectedly should be playing an essential role for fibrillin structure or function. This residue is usually flanked by a conserved glycine residue located two positions upstream and an aromatic amino acid, generally tyrosine (Y), located four positions dowsnstream. FBN12 has an insertion into FCR1.

FCR2 contains a fully conserved tyrosine (Y) which is generally flanked by an aliphatic amino acid followed by a negatively charged residue (FBN10 being an exception to this). At a

distance of 9-12 amino acids from the mentioned Y residue, a positively charged residue, generally an arginine (R), is present. This positively charged residue is frequently followed by a glycine and a negatively charged or polar amino acid. In several cases in which the conserved positively charged residue is an arginine (R), an RGD protein interaction motif (Ruoslahti et al., 1996) is found. FBN1, FBN2, FBN3, FBN4 and FBN9 exhibit the RGD motif. A conserved aspartic acid residue and a highly conserved glycine residue are also present, with this trend broken, respectively, by one or other of the two domains of FBN10.

4.6.7. Prediction of secondary and tertiary structure of fibrillins

The secondary structures of the consensus fibrillins were also studied. Interestingly it was found that the basic structure of two initial alpha helices followed by several beta-chains is conserved among all fibrillins (Figure 41). In the majority of cases these proteins have eight beta-strands (Figure 42; see also Annex V). The secondary structure differences between fibrillins mainly involve insertions and deletions between those beta-chains. FBN10, FBN11 and FBN12 do not exactly follow the "2helix + 8 beta-chain" trend found in the rest of fibrillins. FBN10 has a dual fibrillin domain, of which the first one complies more closely. The fibrillin domain of FBN11 seems to follow the trend more loosely as far as the two alpha-helices are concerned. The previously outcast FBN12 seems to have nine beta-chains instead of eight. The first of these beta-chains corresponds to FCR1, containing an insertion in this sequence.

In general, each fibrillin (or fibrillin subdomain) shows a marked bias towards having two alpha-helices followed by eight beta-chains which could have an intercalation of coils, occasional short alpha helices or short beta chains between some of the eight beta-chains. In Figures 41 and 42, an alignment of the secondary structures of the fibrillin domains of each fibrillin family is performed. In order to allow for a better view of the common traits in secondary structure the second alignment discards the discordant dual fibrillins FBN10 and FBN11.

As well, it is found that the alternate conservation found in certain amino acids of FCR1 and FCR2 coincides with sequences predicted to belong to beta-chains (see annex V).

Figure content — multiple sequence alignment of secondary structure (H, alpha helix; E, beta chain; X, coils).

Block 1:
```
CFBN1_Con_2_0_     1  ------------------------------------------------XXHHHHHHHHHH----XXXXXXXXHHH
FBN1_Con_2_0_      1  -----------------------XXXXXXXXXXXXXXXXXXXXXXXHHHHHHHHHHHHH----XXXXXXXHHH
FBN2_Con_2_0_      1  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXHHHHHHHHHHHHHH----XXXXXXXHHH
FBN3_Con_2_0_      1  ----------------------XXEEEXXXXXXXHHHHHHHHHHHHHH----XXXXXXXHHH
FBN4_Con_2_0_      1  ----------------------------XHHHHHHHHHHHHHHH----XXXXXXXHHH
FBN5_Con_2_0_      1  ----------------------------XHHHHHHHHHHHEX----XXXXXXXHHH
FBN6_Con_2_0_      1  ----------------------------XXHHHHHHHHHHHH----XXXXXXXHHH
FBN7_Con_2_0_      1  ------------------XXHHHHHXXXHHHHHHHHHHHHHHXXXXXXXXXHHH
FBN8_Con_2_0_      1  ----------------------------XHHHHHHHHHHHHHH----XXXXXXXHHH
FBN9_Con_2_0_      1  ----------------------------XHHHHHHHHHHHHH----XXXXXXXXXH
FBN10_Con_2_0_a_   1  ----------------------------XXHHHHHHHHHH HHHH-----HXXXXXXXH
FBN10_Con_2_0_b_   1  ----------------------------XXXXHHHHHHHHHH HHXX--------------
FBN11_Con_2_0_b_   1  ----------------------------XXHHHHHHHXXXX HHH-----HXXXXXXX
FBN12_Con_2_0_     1  ----------------------------XXXXHHHHHHHH HHH----HXXXXXXXHHH
```

Block 2:
```
CFBN1_Con_2_0_     25  HHHHHHHHHHHXXXXXXXXXXXXXXXXEEEEEE--------EXXXXXXXXXXXXXXXXEEEEEEX---XXXXEEE
FBN1_Con_2_0_      57  HHHHHHHHHHHXXXXXXXXXXXXXXXXEEEEE-----XXXXXXXXXXXXXXXXEEEEEEX---XXXXEEE
FBN2_Con_2_0_      76  HHHHHHHHHHHXXXXXXXXXXXXXXXXEEEEE-----XXXXXXXXXXXXXXXXEEEEEEX---XXXXEEE
FBN3_Con_2_0_      44  HHHHHHHHHHHXXXXXXXXXXXXXXXEEEEEEX------XXXXXXXXXXXXXXXEEEEEX---XXXXEEE
FBN4_Con_2_0_      28  HHHHHHHHHHHXXXXXXXXXXXXXXXXEEEEEXXXXXXXXXXXXXXXXXXXEEEEEEX---XXXXXX
FBN5_Con_2_0_      26  HHHHHHHHHHHXXXXXXXXXXXXXXXEEEEEE-----XXXXXXXXXXXXXXXEEEEEEX---XXXXEEE
FBN6_Con_2_0_      28  HHHHHHHHHHHXXXXXXXXXXXXXXXEEEEEX-----XXXXXXXXXXXXXXXXEEEEEX---XXXXEEE
FBN7_Con_2_0_      50  HHHHHHHHHHHXXXXXXXXXXXXXEEEEEX-XXXXXXXHHXXXXXXXEEEEEEX---XXXXEEE
FBN8_Con_2_0_      28  HHHHHHHHHHHXXXXXXXXXXXXXEEEEEEXX------XXXXXXXXXXXXXXEEEEEEEXXXXXXEEE
FBN9_Con_2_0_      24  HHHHHHHHHHHXXXXXXXXXXXXXEEEEEE-----XXXXXXXXXXXXXXXEEEEEEX---XXXXEEE
FBN10_Con_2_0_a_   26  HHHHHHHHHHHHHHXXXXXXXXXXXXXEEEEEE-----XXXXXXXXXXXXHHHHHEEEX----XXXX
FBN10_Con_2_0_b_   17  XHHHHHHHHHHHXXXXXXXXHHHXX---EEEEEEX-----XXXXXHHHHHHXXXXXEEE---XXXX
FBN11_Con_2_0_b_   28  XHHHHHHHHHHHHXXXXXXXXXXXXEEEEEEE-----XXXXXXXXXXXXXXEEEEEEX---XXXX
FBN12_Con_2_0_     28  HHHHHHHHHHHXXXXHHHHXXXXXXXXXEEEEE-----EEXXXXXXEEEEEEEEEEXX---XXXXEEE
```

Block 3:
```
CFBN1_Con_2_0_     92  EEEEEE-----------XXXXXEEEEEEEEEEEEEXXXXEEEEEEEEEEE-----------------XXX
FBN1_Con_2_0_     127  EEEEEX-----------XXXXXEEEEEEEEEEEEEXXXXEEEEEEEEEEEXXXXXXXXXXXXXXXXXXXXX
FBN2_Con_2_0_     146  EEEEEX-----------XXXXXEEEEEEEEEEEEEXXXXEEEEEEEEEEEXX-XXXXXXXXXXXXXXXXXXX
FBN3_Con_2_0_     110  EEEEEX--------------XXEEEEEEEEEEXXXXEEEEEEEEEEEEX------------------
FBN4_Con_2_0_     104  EEEEEE-------XXXXXXXXEEEEEEEEEEEXXXXEEEEEEEEEEE-------------------
FBN5_Con_2_0_      97  EEEEEE-------XXXXXXXXEEEEEEEEEEEXXXXEEEEEEEEEEE-------------------
FBN6_Con_2_0_      98  EEEXX---------XXXXEEEEEEEEEEEXXXXEEEEEEEEEEEE-------------------
FBN7_Con_2_0_     124  EEEEEX-----------XXXXXEEEEEEEEEEEXXXXEEEEEEEEEE-----------XXXXXXX
FBN8_Con_2_0_     101  EEEEEX-----------XXXXXXEEEEEEEEEEXXXXEEEEEEEEE-------XXXXX-----XXXXXXX
FBN9_Con_2_0_      94  EEEEEX-----------XX------XEEEEEEEEEEEXXXXEEEEEEEEE-------------------
FBN10_Con_2_0_a_   95  EEEEEEEEX----------  ---XXXXEEEEEEEEXXXXEEEEEEEEEEE----------------
FBN10_Con_2_0_b_   76  EEEEEEXXX-----------  -----XEEEEXXXEEXXXXEEEEEEEXXEEE-----------
FBN11_Con_2_0_b_   96  EEEEEEEEXXXXXEEXXXXXX  XXXXXEEEEEEEEEEXXXEEEEEEEEX----------
FBN12_Con_2_0_     91  EEEEEE-------------XXXEEEEXXXXXXXXXXXXXEEEEEX-----------------
```

Block 4:
```
CFBN1_Con_2_0_    134  XXXXXXXXXXXXXXXXXX-----------------------------------XXXXXXXXXXXXXEEEE
FBN1_Con_2_0_     192  XXXXXHHHHHHHXXXXXX-----------------------------------XXXXXXXXXXXXXEEE
FBN2_Con_2_0_     210  XXXXXXXXXXXXXXXXXX-----------------------------------XXXXXXXXXXXXXEEE
FBN3_Con_2_0_     143  --------------------------------------------------XXXXXXXXXXXXXEEE
FBN4_Con_2_0_     148  -----XXXXXXXXXXXXX-----------------------------------XXXXXXXXXXXXXEEE
FBN5_Con_2_0_     138  -----------XXXXXXX-----------------------------------XXXXXXXXXXXXXEEE
FBN6_Con_2_0_     136  -----------------------------------------------------XXXXXXXXXXXXXEEE
FBN7_Con_2_0_     168  XXXXXHHHHHXXXXXXXXX-----------------------------------XXXXHXXXXXXXXXXXXEE
FBN8_Con_2_0_     158  XXXXXXXXXXXXXXXX-----------------------------------XXXXXXXXXXXXXEE
FBN9_Con_2_0_     129  ------------------------------------------------EXXXXXXXXXXXXEE
FBN10_Con_2_0_a_  134  ----  -----XXXXXXXXX-----------------------------  ---XXXXXXXXXX
FBN10_Con_2_0_b_  113  ----  -----------------------------------------  ----EXXEEXXXXX
FBN11_Con_2_0_b_  152  ---X XXXXXHHHHHHXXXXHHHXXXXXXXXXXXXXXXXXXXXXXXXXXXXX XEEXXXXXXXX
FBN12_Con_2_0_    127  -----------------X-----------------------------------XXEEEXXXXXXXXXXXXXXXXE
```

Block 5:
```
CFBN1_Con_2_0_    170  EEEEEXXXEEEEEXXXXXEEEEEX---------------
FBN1_Con_2_0_     228  EEEEEXXXEEEEEXXXXEEEEEEXXXXXXX---------
FBN2_Con_2_0_     246  EEEEEXXXEEEEEXXXXEEEEEEXXXXXXXX--------
FBN3_Con_2_0_     158  EEEEEXXXEEEEEXXXXEEEEEEXXXXXXXX--------
FBN4_Con_2_0_     179  EEEEEXXXEEEEEXXXXEEEEEXX---------------
FBN5_Con_2_0_     163  EEEEEXXXEEEEEXXXXEEEEEEXXXX------------
FBN6_Con_2_0_     152  EEEEEXXXEEEEEXXXXEEEEEX----------------
FBN7_Con_2_0_     212  EEEEEXXXEEEEEXXXXEEEEEXXXXXXXXXXXXXXXXX
FBN8_Con_2_0_     194  EEEEEXXXEEEEEXXXXEEEEEEEXXXXXX---------
FBN9_Con_2_0_     144  EEEEEXXXEEEEEXXXXEEEEEEXXXXXX----------
FBN10_Con_2_0_a_  155  XEEEEEEEXXXXXEEEEXXXXXEEEEE------------
FBN10_Con_2_0_b_  123  XXXEEEEEEXXXEEEEXXXXXEEEEEEEXXX--------
FBN11_Con_2_0_b_  227  XXXXEEEEEXXXEEEXXXXEEEEEEEEEXXXX-----
FBN12_Con_2_0_    151  EEEEEXXXEEEEEXXXXEEEEEXXXX-------------
```

Figure 41. Alignment of the secondary structure of all consensus fibrillin domains (the modular proteins FBN10 and FBN11 have their fibrillin domains considered individually). H, alpha helix; E, beta chain; X, coils.

124

Figure 42. Alignment of the secondary structure of all consensus fibrillin domains (excluding the modular proteins FBN10 and FBN11). H, alpha helix; E, beta chain; X, coils.

The regularity of the secondary structures predicted above made us suspect that this could be related to some kind of three-dimensional structure possibly common to all the family. For that reason, 3D model predictions were tried to be obtained for the consensus protein sequences. Certain proteins exist which by their nature are "intrinsically disordered" and do not have a definite structure on their own and thus cannot be meaningfully modelled. Before proceeding it was checked if fibrillins do belong to that group, finding that they do not, according to the results of the METADISORDER online server. After this checking, two kinds of modeling approaches were assayed: homology modeling, and *ab-initio* or *de novo* modeling.

The first approach encountered the unsurmountable limitation of requiring adequate templates of similar primary structure and known tertiary structure. As soon was found, no

adequate templates were available as no close homologues of known structure exist for the fibrillin family. Because of that, this approach was discarded. The second approach was thus necessarily chosen, despite its drawbacks. The d*e novo* or *ab-initio* approaches require a vast computational capacity, with which a broad but finite collection of protein conformations are sampled and tested for their thermodynamic stability. Additionally, the complexity of the task yields sometimes the problem of protein folding too complex to be solved with any confidence above a certain protein length threshold. Luckily, a number of publicly available servers for *ab-initio* protein modelling exist and most of our proteins of interest have a relatively manageable size.

Of the *ab-initio* modelling servers tried, the Robetta online server was found to be the most powerful and accurately predicting. This server was thus the option of choice. Modelizations were performed for all consensus sequences and provided different kinds of models ranging from disordered structures to beta-sheet and beta-barrel structures, as well as some intermediate models. The obtained models were launched against the QMEAN quality control tool. Almost all the best-scoring models were beta-barrels, while beta-sheets scored low. It has to be mentioned, though, that membrane-binding protein models tend to score low in the QMEAN server, as scoring is made assuming that proteins are soluble.
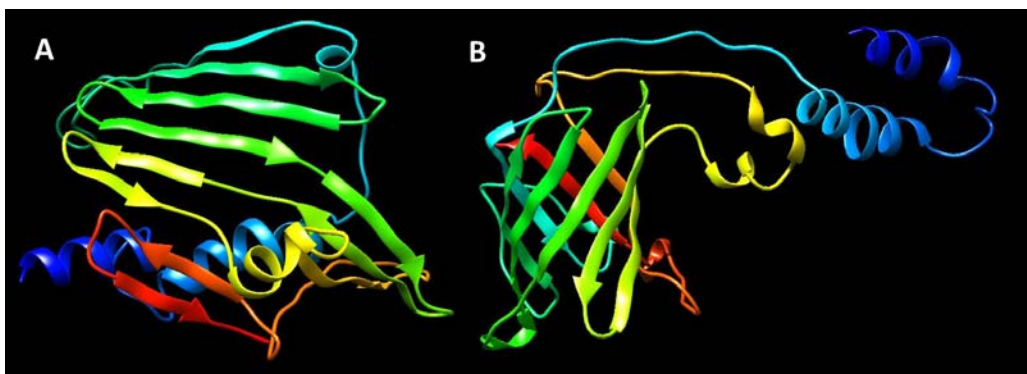

Figure 43. Two distinct predicted 3D structures for the consensus cyanobacterial cFBN1: A) beta-sheet and B) beta-barrel.

The repeated presence of beta-sheet and beta-barrel predictions in our queries could be suggesting the possibility for two different protein conformations. The 3D structures predicted for each fibrillin consensus sequences are indicated in Table 22.

	Barrel-like	Sheet-like
cFBN1	+	+
FBN1	+	+
FBN2	(+)	+
FBN3	+	
FBN4	+	+
FBN5	+	
FBN6	+	+
FBN7	+	+
FBN8	+	+
FBN9	+	+
FBN10	+	
FBN11	(+)	
FBN12		+

Table 22. Occurrence of the prediction of barrel-like and sheet-like structures in each fibrillin type. Symbols between parentheses indicate close-to-call models. Empty boxes indicate a non-prediction of that kind of structure during the modellings performed.

In the cases in which a beta-barrel was predicted, it caught our attention that the most conserved amino acids present in FCR1 (W), and FCR2 (Y), tended to be located at the one and the other end of the beta barrel inner channel. A conserved positively charged residue (R or K) seems to tend to be located in a loop outside the barrel itself, when a beta-barrel is predicted.
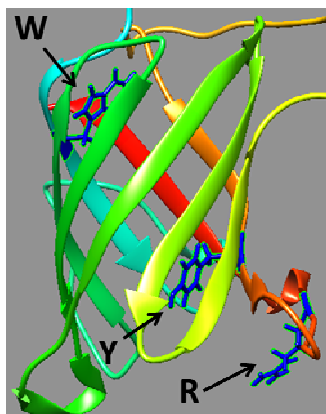
Figure 44. Detail of a predicted beta-barrel of the consensus cyanobacterial cFBN1 indicating the most conserved amino acids W, Y and R highlighted in blue and signalled by arrows.

Regarding the beta-sheet conformation, a tendency is found for aromatic amino acids being located on the edges of the sheet. Higher conformational freedom during model calculation, and with it higher combinatorial possibilities, yield the beta-sheet model much less likely to reflect closely the true structure if it were a beta sheet. As a result, trying to get further

conclusions from these other kind of models is risky.

For the case in which no clear structure could be predicted by modelling, repetition of the prediction can be performed. It has to be borne in mind that any prediction results can vary to some extent due to necessarily finite conformational sampling of the models. However, the repeated finding of highest-scoring models with a given predicted structure (as for instance, the beta barrel) should be reflecting that probably that general structure truly exists, even if the conformational details could very likely vary.

In order to deepen in the study of the prospective 3D structures, the obtained models were submitted to a visual analysis of amino acid conservation using the publicly available webserver ConSurf. This server launches the sequences of the models in PSI-BLAST and then estimates the degree of conservation versus the found homologues within the specified parameters. From this, it generates a graphical representation highlighting the conserved and non-conserved amino acids. The image shown in Figure 45 was elaborated by the mentioned ConSurf server (Celniker et al., 2013). The homolog selection parameters were set in such a way that proteins below 95 % similarity and above 35 % similarity to the model tested were accepted to be used for calculation of the conservation score using the best 100 homologues for the calculation. The conserved amino acids within the compared proteins were by the server highlighted in red, with the most conserved of them having a darker red colour. The homologues found included both cyanobacterial and plant fibrillins.

Figure 45. Three-dimensional structure of modelled cFBN1 beta-barrel seen from opposite sides. Light blue residues are non-conserved; white residues are little conserved; red residues are conserved. Arrows signal residues participating in intermittent or alternate conservation. Note the tendency towards alternate conservancy in several beta-chains. This happens more clearly on one side of the beta-barrel.

The conservation analysis could be performed also with other barrel-like 3D structures as obtained from the modelling in-silico experiments; however, a certain limitation in the sequences available in the ConSurf databases prevented complete well-defined conservation snapshots for each beta-barrel model. In several cases, a number of amino acids remained with unassigned conservation measures, since a significantly lower number of homologues were found in the databases available in the ConSurf server. Nevertheless, the tendency for an alternate conservation seemed to be present in most, if not all cases tested, and in all cases this was found in the beta-barrel itself, and not in the rest of the protein. In all cases, as well, this alternate conservation seemed to be more preponderant on one side of the beta-barrel.

DISCUSSION

5. DISCUSSION

5.1. SUBFRACCIONATION STUDIES OF TOMATO FRUIT CHROMOPLASTS

A number of experimental approaches have been used with the aim of clarifying aspects related with storage of carotenoids, specially lycopene, during tomato fruit ripening. To this aim, different activities were undertaken. They included the setup of a suitable method for chromoplast subfraction and the metabolic and proteomic profiling of the obtained fractions.

5.1.1. Isolation of chromoplast fractions

The method developed for the subfractionation of tomato fruit chromoplast stems mainly from the availability of a method for chromoplast isolation previously set up in our group (Angaman et al., 2012) and two earlier reports where plastoglobules of Arabidopsis chloroplasts were isolated and characterized (Vidi et al., 2006; Ytterberg et al., 2006). Although the primary aim was the isolation of plastoglobules, the method finally set up allows both the isolation of plastoglobules and a number of fractions that globally hold all the membranous components of the chromoplast. At the end, visually distinct and compositionally different fractions were isolated in which all the lycopene and beta-carotene present in the chromoplast has been separated in discrete fractions.

Issues mainly related to adequate resolution of the chromoplast subfractions, as well as to the subjectivity of harvesting, arose during the method development. Inability to address these issues could result in notably increased band cross-contamination, diminishing the informativeness of the results. The main drawbacks found during the development of the fractionation method were: i) the length of the process, which could potentially negatively affect the biological material regardless of the use of protective measures as sample storage in cold, protection from the light and protease inhibition; ii) the relative difficulty, entailing a certain need for training before obtaining repetitive gradient results; iii) the sensitivity to variations in the starting material. To address the first two problems, two solutions could be suggested: i) committing dual labour to complete the whole procedure decreases the time from 12-16 hours to well under 12 hours. If the length of the process is reduced the obtained results would be more representative as any chromoplast degradation would have less chance to take place, and ii) automation of steps potentially introducing variability, like gradient preparation and fraction harvesting.

The sensitivity of the chromoplast fractionation method to external factors was experienced mainly in two ways while working with this method and merits at least in one case being considered as offering an opportunity. On the one hand, the translation of the method to different facilities with a different sonication instrument was found to lead to some variation in the pattern of bands. This could be a result of differences in sample disruption. On the other hand, it was found that different tomato varieties yielded different patterns of bands, which in itself is a reflection of the different tomato composition and likely of the differences in the chromoplasts of different tomato varieties. This, in turn, would be linked to compositional and possibly structural differences related to distribution of lycopene, beta-carotene and tocopherol, as well as other biologically relevant plastid components Significant variations in ripening stage (for instance over-ripening) also resulted in perceptible changes in the final result.

Due to different reasons the first trials of electron microscopy experiments were performed with fruit from a tomato variety different from the "Piccolo" variety generally used throughout this work. Thanks to this circumstance, a clear dissimilarity could be noticed upon observation of the fractions obtained from chromoplast of one or the other variety. One of the tomato varieties used was a standard commercial cherry tomato while the other was the "Piccolo" variety, notoriously having a much more intense red coloration. From the observation of the lightest samples of the gradient using electron microscopy (fractions I, IIa and IIb) plastoglobules (I) or plastoglobules and crystals (IIa, IIb) were found in all cases. However, it became very apparent that in a standard cherry tomato variety fraction IIa was much poorer in crystal-like structures than fraction IIb. In fact, those crystals were encountered only very occasionally in fraction IIa and more frequently in fraction IIb, but their presence in fraction IIb was also relatively scarcer than in the same fraction of the "Piccolo" variety. In contrast, in the "Piccolo" variety both fraction IIa and IIb were rich in crystal-like structures. Thus, different tomato varieties subject to equal extraction conditions showed differences in the subfractions obtained.

The previous observations also led to the simple conclusion that the initial goal of this work, namely the isolation of tomato fruit plastoglobules is not an as straightforward task as in other model plants or tissues as Arabidopsis or tomato leaves, and that the difficulty of this task is directly related to the presence of crystals in the light fractions. The finding that fraction I, the

lightest of our fractions, does not contain any crystals did grant us the opportunity to actually obtain pure tomato fruit plastoglobules. However, the question about what is the compositional difference between these plastoglobules and the heavier plastoglobules found in more dense fractions remains. In any case, it seems likely that lycopene is not accumulated in plastoglobules, as is suggested by the finding that no crystals are found in the pure plastoglobules fraction and that lycopene was not detected in the metabolic profiling of this fraction.

In overall, the attaining of reproducible results by using this method requires strict control of working materials preparation, afterwards yielding the method reproducible yet sensitive to many factors. If reproducibility is attained, the sensitivity of the method could actually become an asset to be used in favour of its wider applicability. For instance, it could be used to investigate subtle ultrastructural, compositional or functional changes of the chromoplast during maturation using tomatoes from different ripening stages. The same approach could be applied, for instance, to the comparison of chromoplasts from different tomato varieties, chromoplasts from different ripening stages or even for the analysis of chromoplasts from different carotenoid producing plants.

The obtained results can be of interest in the frame of the long standing question of where lycopene resides inside tomato fruit chromoplasts. Since long ago it has been proposed that lycopene is stored in the form of crystals (Harris and Spurr, 1969), though proofs were somewhat indirect. Our data sheds some additional light on this. As well, our method tries to further address additionally other related questions like: i) what is the structural, compositional and functional context in which this occurs inside the chromoplast, and ii) what could be the interplay involved and how does this translate to the internal spatial distribution of the chromoplast?. In this respect, it is hoped that our work becomes an outpost for future efforts in the study of this organelle, possibly with a refinement of the gradients used or the use of alternative lysis methods.

5.1.2. Metabolite profiling of chromoplast fractions

In order to characterize the fractions obtained a targeted metabolite profiling focusing on carotenoids, tocopherols and the major polar lipids was performed. It was expected that these data, together with proteomics data, would give us insights for understanding the underlying suborganellar context and to tentatively identify the fractions obtained. Metabolite analyses were divided into carotenoids/tocopherols on the one hand and galactolipids and phospholipids on the

other.

During the metabolite profiling of carotenoids and tocopherols a number of problems with carotenoid analysis appeared. An analytical method (Fraser et al, 2000) which was working well in the group appeared to show discordant results. The apparent lack of traceability of carotenoids during the analytical process, being a significant amount of the analyte either "lost" or "gained " after processing the initial material, hinted us to suspect that something was amiss. Specific assays were conducted to check the recovery and support the measurement method but the unexpected discrepancies forced to revise the whole analytical procedure. This allowed to detect a clear degradation of lycopene in some samples. Furthermore, performing consecutive measurements of one same sample showed that a clear decline of the analyte was happening. The magnitude of this decline was variable, but consistently higher in lycopene than in beta-carotene, where it was however also found. This suggested indeed a degradation of carotenoids which was later found to be supported by the bibliography (Rodriguez-Amaya and Kimura, 2004).

The degradation problem was solved by running samples one by one during the HPLC analyses and retrieving and redissolving them immediately before HPLC injection to prevent any opportunity for degradation while in the HPLC queue. Nevertheless, the antioxidant molecule pyrogallol found later on (Parrilla et al., 2007) could in retrospect have been successfully used to prevent degradation, allowing in this way a less labour-intensive procedure. The fact that no reports existed on the use of pyrogallol in carotenoid analysis had in practice a deterrent effect, though our latest evidence suggested it would have been successful in preventing lycopene degradation during HPLC measurement.

The literature used as a reference for the HPLC method used (Fraser et al., 2000) mentioned no need to add any antioxidant to successfully perform carotenoid analyses and, moreover, suggested that samples could be stored safely at -20ºC for up to a year under inert atmosphere with no further caution. Even, it was reported that the antioxidant molecule BHT (butylated hydroxytoluene) apparently did not influence the degradation of carotenoids. This contrasted with what we had found and it became thus quite possible that our specific procedure had created the conditions for the degradation of carotenoids.

After solving these technical problems it was found that lycopene and beta-carotene were mainly found at the low-density fractions, with beta-carotene having a tendency to be more present in the lighter fractions, if compared to lycopene. An interesting observation is that beta-

carotene is still found in the lightest fraction I, where lycopene was hardly identified. The rest of the measured metabolites (carotenoid precursors and tocopherols) were predominantly found in the heavier fractions. This was also the case for membrane lipids.

To get deeper insights form the data obtained, it was considered to look at them from a different perspective. Proteins, when they are not free in the stroma, lumen or envelope intermembranous space are expected to be interacting with membrane systems. From this, it follows that any non-hollow storage particle of relevant volume, like plastoglobules, will show a disproportion in the distribution of the stored compounds if compared to that of the protein, which will remain at the surface. If this were numerically expressed, it would display which substance or substances are stockpiled in the core of that storage particle. This assertion can be supported if one considers how the volume of a sphere varies if compared to its surface. In the formulae describing this, the volume is cubically dependent on the radius of the sphere. Thus, any component filling the inside of such a sphere would increase exponentially (cubically) as the size of the sphere increases. On the other hand, any component which is circumscribed to the surface, as protein in plastoglobules (Bréhélin and Kessler, 2008), will increase also exponentially, but only on a square-exponential basis, and consequently would quickly be lagging behind as the radius of the sphere increases. Thus, to further assess the significance of the metabolite profiling data, an additional calculation was made to define the relative enrichment of each particular metabolite with respect to the protein level found in the different samples. This was done by dividing the percentage of the total metabolite found in a particular fraction by the percentage of the total protein estimated in that fraction. This allows to clearly see distribution asymmetries of the metabolite of interest with regard to protein, which can expectedly be linked to over accumulation of a given compound inside a non-hollow storage particle which possesses protein only on the surface (see table 23).

Taking the previous arguments into account, any high relative enrichment found should reasonably be related to the presence of components which are accumulated massively into a hydrophobic inner core, whereas protein is kept only on the surface of these particles. From the electron microscopy observations, the sphere-like plastoglobules or the prismoid-like rod or planar crystal structures are candidates for this behaviour.

	Lycopene	β-carotene	Phytoene	Phytofluene	α-tocopherol	γ-tocopherol
Fraction I	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Fraction II	12,8	15,3	4,8	4,8	2,3	1,3
Fraction III	4,0	3,2	1,0	1,0	1,6	1,6
Fraction IV	1,1	0,7	1,1	1,1	1,4	1,6
Fraction V	0,2	0,2	0,8	0,8	1,1	1,1
Fraction VI	0,0	0,1	0,4	0,4	0,4	0,4

Table 23. Relative enrichment of the indicated metabolites with respect to protein content in the different chromoplast subfractions. Fraction II corresponds to the merger of fractions IIa and IIb. n.d., not determined.

A high relative enrichment value for beta-carotene and lycopene was found in fraction II (IIa + IIb) and, to a lesser extent, for phytoene and phytofluene. The specific enrichment of beta-carotene and lycopene in fraction II is thus in consonance with both these carotenoids being stored at the inside of sizeable storage particles at the inside of which no protein resides. It has to be noted that the elements found in fraction II using electron microscopy observations were plastoglobules and rod and planar-like crystals. Fraction III seems to be still relatively enriched in lycopene and beta-carotene, but not in phytoene and phytofluene. In this fraction, possible plastoglobule-like particles were still viewed. As we proceed down the gradient to fractions IV, V and VI, it is found that any relative enrichment disappears, blurred by the abundance of other material. This happens more markedly for lycopene and beta-carotene, whose relative enrichments drop to insignificant levels. This reflects that, even if lycopene is found in fractions IV, V and VI, its presence has a quantitative importance only because of the high quantity of chromoplastic material there found.

Fraction I (first lane of table 23) contains no numerical figures because protein content could not be measured. In a mathematical idealization, we could fill in the gaps with a "tending to infinite" statement, but this would be lacking biological significance and would blur any differences found in that fraction. Instead, if we consider again the percentage of each metabolite found in this fraction, a significant proportion of total phytoene (12 %) and phytofluene (13 %) are present in this fraction (see results, table 9). This can lead to infer that if protein content could be actually measurable, the relative enrichment for phytoene and phytofluene in this fraction would be very relevant. From this inferred specific enrichment, it could be reasonably proposed that phytoene and phytofluene are stored in the hydrophobic core of plastoglobules. If we take into account that phytoene and phytofluene are precursors for carotenoid biosynthesis,

this could reflect the importance of plastoglobules in either the synthesis, storage or distribution of these precursors or their derivative products.

The results obtained in the proteomic analysis of fraction IIa from standard cherry tomato fruit (containing plastoglobules and very few crystal-like structures) (section 4.4.1) could be interesting on this respect. This fraction was shown to contain 15-cis-zeta-carotene isomerase, carotenoid isomerase and zeta-carotene desaturase. Looking further into these proteins, they can univocally be linked to lycopene biosynthesis, as these enzymes catalyse the three steps preceding the formation of lycopene (Figure 46). In support to the assignation of these proteins to plastoglobules, it has been reported that plastoglobules isolated from red bell pepper chromoplast contain two lycopene biosynthesis-related enzymes: lycopene beta-cyclase and a zeta carotene desaturase (Ytterberg et al., 2006). The combination of the high relative enrichment of the carotenoid precursors phytoene and phytofluene and the presence of three enzymes involved in the final steps leading to lycopene biosynthesis suggest that plastoglobules could be involved in the biosynthesis of lycopene during tomato fruit ripening. On the other hand, as also can be noted from the data shown in table 9, lycopene is not detected in plastoglobules. This would indicate that while lycopene could be synthesized in plastoglobules it is apparently not stored there or not at least in the plastoglobules present in fraction I.

The results shown in Table 23 indicate as well that by far the highest relative enrichment found was that of lycopene and beta-carotene in fraction II. As had been found in the electron microscope observations, this fraction contains both plastoglobules and lycopene crystals. Taking into account what has been discussed above for fraction I (which contains only plastoglobules), if we consider that plastoglobules present in fraction II possibly have a similar nature, it is likely that the high relative enrichment of beta-carotene and lycopene in fraction II is related the crystal-like structure there found.
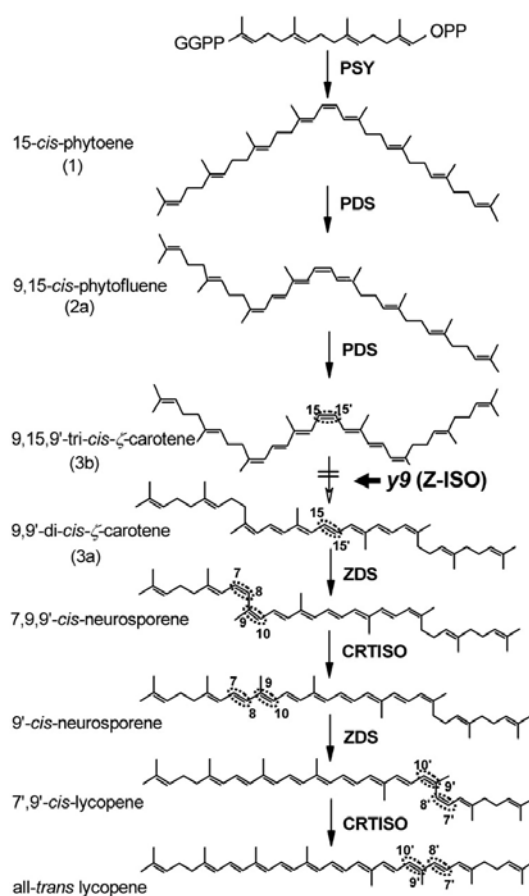
Figure 46. Biosynthesis of lycopene from geranylgeranyl pyrophosphate. PSY, phytoene synthase; PDS, phytoene desaturase; y9 (Z-ISO), 15-cis-zeta-carotene isomerase, ZDS, zeta-carotene desaturase; CRTISO, carotene cis-trans isomerase. Figure reproduced with permission from Li et al., 2007. © 2007 American Society of Plant Biologists

A noticeable enrichment in phytoene and phytofluene can be found in fraction II (Table 23), which can be interpreted as an effect of the presence of plastoglobules (if those are equivalent to those found in fraction I). It is also worth noting that the relative enrichment found for beta-carotene is noticeably higher than that of lycopene. From the finding that plastoglobules in fraction I do contain beta-carotene, plastoglobules could be contributing to the higher relative enrichment in beta-carotene in fraction II. One of the alternatives for the interpretation of these results is that lycopene is stored exclusively in crystals and that phytoene and phytofluene are contributed by plastoglobules. Alternatively, a shared distribution of lycopene, beta-carotene, phytoene and phytofluene between crystals and plastoglobules could be considered. This alternative explanation would entail that different kinds of plastoglobules exist in tomato fruit chromoplast, with some of them including lycopene (in fraction II) and some of them almost

none at all (as in fraction I). The definite answers to these questions require further experiments.

Most of the tocopherols present in the chromoplast were found in the heavy fractions (table 9, figure 14). Interestingly, tocopherol cyclase VTE1, the enzyme involved in the tocopherol pathway found in chloroplast plastoglobules, has also been previously reported by Hussain et al. (2013) and Vidi et al. (2006). This is in agreement with the proposed role of the plastoglobule as a metabolic crossroad from which tocopherols are biosynthesized and redistributed to other locations (Piller et al., 2012).

Further looking into the metabolite profiling data, the same relative enrichment calculations were performed for the membrane lipids monogalactosyl diacylglycerol (MGDG), digalactosyl diacylglycerol (DGDG), phosphatidyl ethanolamine (PE) phosphatidyl serine (PS) and phosphatidyl choline (PC) (Table 24).

	MGDG	DGDG	PE	PS	PC
Fraction I	n.d.	n.d.	n.d.	n.d.	n.d.
Fraction II	3,5	4,8	6,0	2,5	5,3
Fraction III	1,0	2,4	2,2	0,8	2,0
Fraction IV	1,1	1,3	0,7	0,2	1,0
Fraction V	1,1	0,7	0,9	2,1	1,1
Fraction VI	0,6	0,6	0,6	0,3	0,3

Table 24. Relative enrichment of membrane lipids with respect to protein content in the different chromoplast subfractions. Fraction II corresponds to the merger of fractions IIa and IIb. MGDG, monogalactosyl diacylglycerol; DGDG, digalactosyl diacylglycerol; PE, phosphatidyl ethanolamine; PS, phosphatidyl serine; PC, and phosphatidyl choline. N.d., not detected.

Unfortunately, values for fraction I are not currently available. Regarding fraction II (merger of fractions IIa and IIb) some relative enrichment in PE, PC and DGDG was found. This suggests that these polar lipids could be significantly represented in the surface of the elements found in fractions II and III, be those plastoglobules, crystals or any other membranous elements. A lesser enrichment was found for MGDG and PS. PS has a peak of enrichment at fraction V, indicating that it has a strong preference for localization in that fraction. Looking at the results in overall, a similar distribution is found for DGDG, PE and PC among fractions, if compared to that of MGDG and PS.

It has been reported that in chloroplasts the outer membrane and the inner membrane and thylakoid membranes show a distinct lipid composition (Zbierzak et al., 2011). In the case of thylakoid and inner envelope membranes, MGDG is the main lipid found (around 50 %),

followed by DGDG (around 30 %), and a number of minoritary lipids, including phosphatidylglycerol (PG; 10 % of polar lipid, calculated from moles), sulfoquinovosil diacylglycerol (SQD; around 5 %) and phosphatidyl inositol (PI; near 1 %). For their part, outer envelope membranes have a composition which includes MGDG (20 %) and DGDG (30 %), but displays additionally the strong presence of PC (30 %). Minoritary lipids include PG (10%), SQD and PI (near 5% each). It can from this be inferred that an expected difference between plastid inner membranous systems and the plastid outer membrane is the higher presence of phospholipids in the outer plastid membrane.

Considering the lipid composition chloroplast membranes, and also taking into account that fraction V is enriched in PS as compared to the rest of chromoplast fractions, it can be suggested that this fraction is enriched in outer membrane components. The obtained results also suggest that in the case of tomato fruit chromoplast the main outer membrane phospholipid could be PS instead of PC.

5.1.3. Proteomic analysis of chromoplast fractions

The results derived from the proteomic analysis of tomato fruit chromoplast subfractions could be termed to be a novel contribution to the body of knowledge of organelle proteomics. This is one of the few recent plant proteomics works addressing the parallel study of multiple membrane systems of a single organelle, in this case a plastid. It is also currently the only work of its kind in the field of chromoplast studies, preceded however by the milestone that was the recent publication of the tomato chromoplast proteome (Barsan et al., 2010). However, even if we share the same model species, the approach followed in the present work was different from that of Barsan et al. (2010).

Revisiting the results obtained in this part of the work allows to highlight the most interesting aspects related with the proteins present in the different chromoplast subfractions. Firstly, the use of proteomic identification on discrete bands of SDS-PAGE mini-gels resulted in the finding that fibrillins were among the very major proteins found in plastoglobules, even before this was formally reported for plastoglobules isolated from Arabidopsis leaf chloroplasts (Lundquist et al., 2012). Later on, whenever our isolation became in working order, the more extensive proteomic identifications could begin.

Further studies on plastoglobules and the other light fractions yielded some more information about the proteins present. However, most of the data were generated from a purified

plastoglobule fraction which also contains carotenoid crystals. The presence of carotenoid related enzymes in this fraction (section 4.4.1) seems coherent with its carotenoid-rich composition. As well, several fibrillins were found, including FBN1, FBN2, FBN4, FBN6, FBN8 and FBN12. Finally, a tocopherol cyclase (VTE1) responsible for the formation of gamma-tocopherol was found, in consonance with previous findings in Arabidopsis plastoglobules, which yielded the plastoglobules as a place for tocopherol biosynthesis (Vidi et al., 2006).

The analysis of the carotenoid crystal-rich fraction II obtained from fruits of the "Piccolo" variety allowed the identification of the proteins shown in Table 11 (section 4.4.1). The presence of energy metabolism-related proteins could make us think that the process leading to the assembly and growth of lycopene crystals may require for some reason of a steady energy supply. This would, though, contrast with the finding that in red bell pepper chromoplast fibrils had been shown to spontaneously assemble in vitro if the right stoichiometry of protein, lipids and carotenoids was met (Deruère et al., 1994). As well, the presence in fraction II of the fibrillin "Harpin binding protein" (FBN4) instead of FBN1, the latter found by Deruère and coworkers related to fibril formation in pepper, is noteworthy. Taking into account that only FBN4 was detected in this fraction, it is likely that in tomato fruit the formation of lycopene-crystals could be dependent on FBN4, in the same way as the distinct red bell pepper fibrils required FBN1 for their assembly. However, the possibility for other fibrillins being obscured by other stronger signals does not permit to rule out their presence in this fraction.

The analysis of the protein-rich fractions III, IV, V and VI using 2D electrophoresis revealed that they have a lot of protein in common, although the relative intensity of the protein spots varied to some extent among them. In spite that several spots unique to a particular fraction were identified, the identity of the corresponding proteins could not be established. However, the massive shotgun identification of proteins in fractions III, IV, V and VI provided interesting information about the proteins present in each fraction. The overview of the pooled GO descriptors of the overall proteins identified in fractions III, IV, V and VI (Figure 28) yielded a landscape in which roughly 40% of them can be related to either response to stress or to external biotic or abiotic stimuli. Furthermore, over 20% of the proteins are related with targeting or interaction with other proteins either for modification, regulation or protein import. 15 % of the proteins are related to the generation of precursors and metabolites, and another 15% is devoted to diverse transport functions. The rest include proteins related to signal transduction functions

and electron transport.

Whenever one takes a general look of the GO representations of all the proteins identified in each particular fraction (Figure 29) it is clear again that many similarities are found among them. Despite that, a few interesting differences were found. Fraction III differs from the other fractions in that it has relatively short variety of proteins and that among those the transport-related proteins are preponderant. Unlike the other fractions, no significant contribution of signal transduction proteins is found. Fraction IV has the highest contribution of proteins related to stress response. Finally, fractions V and VI display a very high similarity in the distribution of protein functions, and are the only fractions with a significant contribution of proteins related to the electron transport chain.

When considering the proteins specific to each fraction the following aspects can be highlighted. Fraction III seemed to be the simplest fraction of all, having only two detected unique proteins. Both of them were ATP/energy metabolism related (ATPases). In connection with the studies on the light fractions, it could be found that fraction III displayed the presence of only one fibrillin, FBN4, which was also detected in fraction II. This could suggest a connection with the lycopene-accumulating structures in that fraction.

Fraction IV had the widest variety of proteins showing different functionalities. The finding of the highest number of unique carbon metabolism, lipid-related, regulatory, nucleic acid-related and chaperone proteins in this fraction, indicate that this fraction is enriched in inner plastid components. These could be a mixture of inner envelope and other chromoplast internal membranous systems. As fraction IV has the widest variety of unique functionalities, it is reasonable to expect that this is the fraction with the most diverse composition. In other words, the components of this fraction could have a participation of a wider variety of functions than those present in the other fractions.

A closer look at the proteins found in fraction IV reveals that many of them can be related to plastid membranes. This is the case of a dihydrolipoyl dehydrogenase-like protein, a soluble inorganic pyrophosphatase, a transketolase and an enoyl-ACP reductase which have been reported to be localized in the plastid-envelope of Arabidopsis chloroplasts (Ferro et al., 2002; Froehlich et al., 2003). Also, the presence of Tic110 protein suggests the presence of inner membrane components in this fraction (Kovács-Bogdán et al., 2010). An integral membrane ATP-dependent zinc metalloprotease has been described to be present in the plastid envelope,

144

but also in the thylakoids (Froehlich et al., 2003; Rutschow et al., 2008). A fructose-bisphosphate aldolase similar to an Arabidopsis relative found in plastoglobule and thylakoid was also present (Vidi et al., 2006; Ytterberg et al., 2006). A soluble inorganic pyrophosphatase and a transketolase were found, which have been reported to reside in plastid stroma (Ferro et al., 2010; Schulze et al., 2004). Finally, an unknown protein named "chloroplast thylakoid membrane protein isoform 1" was found. In summary, several proteins specific of fraction IV have been found to be related to the plastid envelope, while a few others seem to be associated to thylakoids, plastoglobules or stroma, respectively (see Annex II to retrieve the associated protein ID numbers).

Fraction V was the second richest fraction with regard to the number of unique proteins. It has the highest number of unique protein related with transport-processes. Unique stress proteins were also found in relative abundance, Interestingly, a significant number the functions assigned to the unique proteins of this fraction were assigned as either "unknown" or "miscellaneous". No proteins related with carbon metabolism, chaperones or nucleic acids were found. This could suggest an outer location for the main components of this fraction. It is worth noting. Two proteins annotated as "outer envelope pore protein 16-3 chloroplastic mitochondrial-like protein" and "outer membrane omp85 family protein" were also found. Additionally, a "mitochondrial import inner membrane translocase subunit tim50-like" protein, which was also identified in the updated tomato chromoplast proteome published by Barsan and collaborators (Barsan et al., 2012), was also found. The presence of these proteins might be pointing to the presence of outer envelope membrane components in this fraction. This is in agreement with the findings derived from lipid profiling which also indicated that this fraction was enriched in outer membrane components (section 4.2.2). The fact that this fraction is enriched in transport-related components further seems to supports this observation.

With only 13 identified unique proteins, fraction VI was found to be simpler than fractions IV and V. Still, some specificities were found, being a significant number of annotated proteins related to lipid metabolism, sugar/carbon metabolism and signalling and regulation. No specific transport-related, protease, or ATP/energy metabolism-related proteins were found. As well, a phospholipase D is found with a close relative in Arabidopsis which is located to either the chloroplast envelope, the thylakoids or to stroma (Ferro et al., 2002; Froehlich et al., 2003; Peltier et al., 2004; Rutschow et al., 2008). Related to carotenoid precursor biosynthesis, an

isopentenyl diphosphate delta-isomerase is found in this fraction.

A detailed analysis of the proteins identified in the chromoplast subfractions revealed the presence of many unexpected proteins. Some of them were found to be previously assigned to be mitochondrial, peroxisomal, cytoplasmic, secreted or belonging to endoplasmic reticulum. Some unusual organelle-targeting cases have been reported in the literature. For instance, it has been reported that some proteins are dually-targeted to both mitochondria and plastids (Bahaji et al., 2011; Xu et al., 2013). This dual or even multiple targeting likely shows a differential distribution ratio for each of the proteins displaying it, resulting in different affinities for the destination organelles, or for the localization-related proteins involved. Another possibility for plastid localization of proteins with no apparent transit peptide involves the direct import of mRNA into organelles, as described by Weis et al. (2013). Interestingly, a protein that is well established to be plastidial, as FBN4, has also been found to be extracellular in tobacco (Chen et al., 2012; Wei et al., 1992).

Upon gathering the results obtained from the metabolite profiling, proteomics and electron microscopy analysis obtained in this work, a picture summarizing the composition of the different chromoplast subfractions can be drawn. A schematic representation of the component present in each chromoplast fractions is shown in Figure 47.

This tentative identity assignation offers a general overview of the biological material obtained with the sub-fractionation method by us developed. This has the potential for aiding in the further study of discrete sub-components of the tomato fruit chromoplast. To that aim, a modification of the gradient could be performed in order to adapt it to focus on more detailed study of a particular chromoplast sub-component.
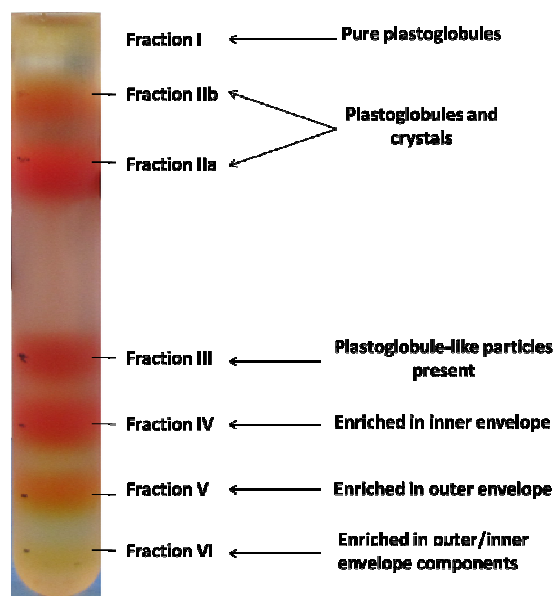
Figure 47. Tentative assignation of the identity of components participating of each fraction. In fraction III identity assignation remains more blurred, though plastoglobule-like particles were observed.

5.2. STUDIES TO UNRAVEL STRUCTURAL AND FUNCTIONAL FEATURES OF FIBRILLINS

Our interest in the study of tomato fruit fibrillins arose from the existence of several reports suggesting the role of these proteins in carotenoid accumulation (Deruère et al, 1994; Simkin et al., 2007). In addition, the results obtained in Section 4.5 further supports the proposed role of fibrillins not only in carotenoid accumulation but also in fruit ripening. At this stage it was realized however that very little was known about the structure and function of these proteins. Only quite recently, some general functional and evolutionary aspects related to plant fibrillins were reviewed by Singh and McNellis (2011), a milestone work in which the first coherent classification of fibrillins was proposed. In the present work, new insights into plant fibrillins have been obtained by using a series of bioinformatic and structural prediction tools.

Regarding the bioinformatic analysis of fibrillins the finding of "hub" genomic websites was particularly useful. One of these, the "Genome Portal of the Department of Energy Joint Genome Institute" aims at unifying many genomics datasets, giving it a present and potential relevance as a "hub" resource for genomic or protein sequence data. Phytozome

(http://www.phytozome.net), another "hub" database, which is devoted to photosynthetic eukaryotes, has been among the most useful sources found, in close rivalry with the generalistic NCBI and EXPASY databases. Still, sometimes it had to be resorted to the primary source: the particular genome project. When this was required, it was encountered that some species-specific genomic databases contained sequences which had to be discarded due to their apparent incompleteness, lack of update or fragmentary data. As a result, some species with only genome drafts available were discarded from our studies.

5.2.1. Main features of the plant fibrillin family

The fact that all photosynthetic organisms always display at least some type of fibrillin and that the other non-photosynthetic plastid-containing organisms do not have this kind of proteins highlights that fibrillins are likely inextricably linked to photosynthesis. The overall conservation of the twelve-member fibrillin family as a whole in the vast majority of higher plants highlights the functional relevance of these proteins. An exception could be FBN5, which is not found in cucurbits (melon and cucumber) and also has a partially conserved sequence in other plants. This suggests that FBN5 could have an accessory role despite its very wide presence in higher plants.

It was found that while some plants had more than one form of each fibrillin type, others had only one copy of each. Of those having multiplicities, handfuls of very similar fibrillin isoforms were sometimes encountered, with somewhat divergent sequences. In some other cases of fibrillin multiplicity, a diversity of close fibrillin relatives were found, displaying a number of uncomplete forms. In some cases, though, there is some chance for this being genome sequencing artifacts. This diversity and multiplicity of fibrillins in some species made our resolve of picking a single representative per species more solid. It is possible that the majority of fibrillin multiplicities are the result of species-specific genome duplications combined with divergent evolution, rather than a response to functional needs. Possibly, a significant part of any multiple fibrillin isoforms are either non-functional, non-expressed, or in any other way, likely non-relevant. The different sub-families of fibrillins found, and the related experimental and bibliographical information available are discussed next.

5.2.2. Functional role of plant fibrillins

FBN1 has been found to be the most highly conserved fibrillin in the whole family, with

relatives present in land plants and in all other photosynthetic organisms. It seems to have evolved from a predecessor in common with the cyanobacterial cFBN1 relative. The finding that in algae and mosses either FBN1 or FBN2 seemed to be alternatively absent in several cases, support previous suggestions for these fibrillins having overlapping functions (Youssef et al., 2010). Another trait which could be supporting this is that both FBN1 and FBN2 are the most (negatively) charged proteins of the family. Thus, of the twelve fibrillin types, FBN1 and FBN2 would be best fitting the possible role of plastoglobule coalescence prevention suggested by Ytterberg et al. (2006) and. This would as well match the generally suggested structural role for fibrillins, even more if we take into account that FBN1 and FBN2 are amongst the most abundant proteins found in plastoglobules.

The widest literature reports on specific fibrillins are those mainly addressing FBN1 and FBN2. It has been reported that FBN1 is involved in the formation of lipoproteic structures as plastoglobules or fibrils in certain chromoplast types. It has been suggested to participate in the generation of a "sink effect" during pigment overaccumulation (Simkin et al., 2007). FBN1 is found in all tissues of the red bell pepper plant, but mainly in the ripe fruit. FBN1 increases during ripening of red bell pepper and is shown to be the major component of plastoglobules and fibrils, where it is localized in the outer surface of these structures (Pozueta-Romero et al., 1997). As well, a direct relationship between FBN1 and flower-specific carotenoid accumulation has also been described (Leitner-Dagan et al., 2006). In high-pigment tomato mutants hp1, hp2 and hp3 an increased abundance of the CHRC (FBN1) is seemingly participating in the increased lycopene levels of these mutants (Kilambi et al., 2013). By expressing FBN1 of red bell pepper into tomato fruit, a two-fold increase of carotenoid content and volatile formation was obtained, though this process was not mediated by the generation of any fibrils, despite this being possibly expected (Simkin et al., 2007).

FBN1-suppressed plants show increased susceptibility to infection, supporting a role for this fibrillin in plant protection (Leitner-Dagan et al., 2006). The expression of pepper FBN1 in tomato fruit resulted as well in a delayed thylakoid degradation in differentiating chromoplasts. This resulted in the creation of a discrete zonification inside the transitional chloro-chromoplast with areas still retaining chloroplastidic structures and others attaining chromoplastic appearance. Thus, it seems reasonable to assume that FBN1 is exerting a membrane-protecting role (Simkin et al., 2007). In Arabidopsis, the simultaneous repression of the expression of

FBN1a, FBN1b and FBN2 by means of RNAi resulted in an impairment in long-term acclimation to environmental constraints, as for instance photooxidative stress imposed by exposure to high irradiance and cold. This same repression led to an inefficient protection of the photosynthetic apparatus, provoked abnormal granal and stromal membrane arrangements, led to a higher photoinhibition of photosystem II under stress conditions and to a slower shoot growth and a deficit in anthocyanin accumulation during stress (Youssef et al., 2010).

FBN1 is expressed and highly conserved in monocotyledonous and dicotyledonous plants, and it seems to be expressed regardless of the kind of plastid (Pozueta-Romero et al., 1997). In cucumber, FBN1 (CHRC) has been suggested to react to gibberellin changes via a MYB-like trans-activator, the cucumber ChrC having been found to have a gibberellin-response element (gacCTCcaa) which mediated its response to gibberelins (Leitner-Dagan et al., 2006). FBN1 (CHRC) is activated in response to biotic and abiotic stresses, as also are other fibrillins (Letiner-Dagan et al., 2006). Plants generate ABA as an endogenous signal when exposed to stress, leading to enhanced tolerance of PS-II to photoinhibition, as found in Arabidopsis. One of these effects is FBN1 accumulation, mediated by the ABI1 and ABI2 abscisic acid response regulators. The increase in fibrillin expression in turn seems to mediate the ABA-induced photoprotection (Yang et al., 2006). The repression of FBN1a, FBN1b, and FBN2 generates abnormal patterns of expression of normally JA-induced genes. A JA-deficient mutant showed similar phenotypic characteristics to the repressed plants. JA treatment reverted all these phenotypic effects. JA could thus be playing an important role in the acclimation of chloroplasts to stress. Light and cold stress-related JA biosynthesis and release could be conditioned by the accumulation of FBN1a, FBN1b and FBN2. If those are less present, the JA-synthesis response is hindered, as well as the response to JA. It is suggested that the dependency between FBN accumulation and JA synthesis could be mediated by plastoglobule and TAG accumulation, being plastoglobules potential sites for initiation of JA-biosynthesis (Youssef et al., 2010).

From de VIGS analysis performed in the present work it has been found that silencing of both FBN1 and FBN2 in tomato fruit results in a clear delay of fruit ripening. As well, RT-PCR expression studies have shown a general trend towards an increase in the expression of FBN1 and, more lightly, of FBN2 during ripening. It is noteworthy that the overwhelmingly highest expression levels for FBN1 in tomato were found in tomato flowers (data not shown).

FBN3 seems to be well conserved among plants, mosses and algae. Its presence in these

organisms suggests that it has an essential function in all of them. As well, the finding of FBN3 as a reiterate member of several modular proteins which include protein-interacting domains (FHA, PB1) (Durocher et al., 2002; Sumimoto et al., 2007) suggests that its spatial coincidence with specific proteins has certain function(s). However, unless these partner protein(s) are known it is difficult to assess what kind of function that would be. The fact that our expression analyses show that FBN3 it is very poorly expressed in tomato fruits and that it even decreases further during fruit ripening suggests that its role is mainly restricted to vegetative tissue.

FBN4, also known as Harpin-binding protein, was another of the major proteins found in tomato fruit plastoglobules during the first proteomic analysis. Its expression seems to increase during the ripening of the fruit. Silencing of FBN4 using VIGS had an effect similar to that of FBN1 or FBN2 .

Reports are available for FBN4 indicating its requirement for plastoglobule development and resistance to multiple stresses in apple tree (Singh et al., 2010). FBN4 is found in the chloroplast associated to plastoglobules and to PS-II light-harvesting complex in thylakoids. It is suggested that FBN4 is involved in regulating plastoglobule content, and that defective regulation of plastoglobule content leads to broad stress sensitivity and altered photosynthetic activity. Curiously, FBN4 knock-down apple trees also had a more efficient carbon fixation at low temperature. However, at high temperatures this seeming advantage shifted in the opposite direction, becoming disadvantageous. These FBN4 knock-down trees evidence an increase of anthocyanin production when transferred from low to high-irradiance, suggesting a more stressed state of the plant. Those trees showed also an increase in sensitivity to methyl viologen (paraquat) (Singh et al., 2012), a herbicide that acts by inhibiting photosynthesis due to generation of reactive oxygen species (Summers, 1980). In the knock-down apple trees, plastoglobule number increased upon exposure to ozone stress. After ozone oxidative stress, osmiophilicity of plastoglobules decreased, while it remained low for knock-down apple trees. A decrease in osmiophilicity can be interpreted as a decrease in unsaturated lipids or hydrophobic compounds in the plastoglobules (Harris, 1970; Singh et al., 2012). Furthermore, FBN4 has been found to be the mediator of the mechanism of action of the antiviral Dufulin (Figure 48), which acts by activating the systemic acquired resistance of plants. Dufulin is also known to activate the salicylic acid (SA) response upon plant leading to resistance to virus infection (Chen et al., 2012).
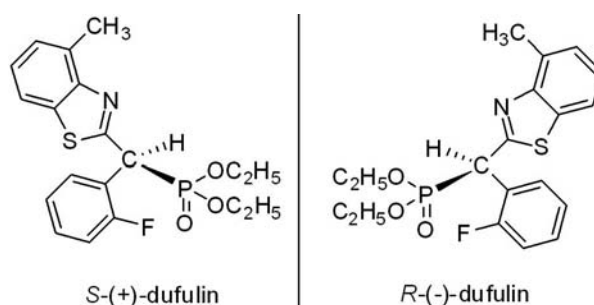
Figure 48. Chemical structure of Dufulin, known to interact with FBN4, leading to SA response. Source: www.alanwood.net/pesticides/dufulin.html

All these observations are compatible with FBN4 having a role helping to quench the effects of different kinds of stresses, and channeling at least some SA-mediated plant responses. The finding that FBN4 knock-out trees maintained osmiophilicity, while FBN4 impairment also increases plant stress, suggests that FBN4 might be mediating the mobilization of the antioxidant reserves present in the plastoglobule, as osmiophilicity is related to the number of unsaturated molecules (potential reducing agents) present at a given structure, in this case plastoglobules. Thus it seems that FBN4 has a role as a mediator of plant stress resistance by allowing for the utilization of the accumulated antioxidants, which could help quenching ROS. The fact that a decreased number of FBN4 protein resulted in a decrease in size of the plastoglobules suggests a possible structural role in addition to the previous, or a different role related to the generation and growth of plastoglobules. On top of all the previous, and taking into account our proteomics results, it can be suggested that FBN4 likely has a role in lycopene crystal formation in tomato fruit chromoplasts. This would be in consonance with our finding that FBN4 steadily increases its expression during tomato fruit ripening.

FBN5 is the least conserved fibrillin in the family, and no information is available whatsoever. However, the finding of several modular FBN5 proteins strengthens the suggestion that it could have a function in at least some organisms. The fact that one of these modular proteins contains a module related to vesicle transport (Vps51,Vps67) (Zaman et al., 2012) suggests that FBN5 can have characteristics which yield it useful in the context of vesicle trafficking.

FBN6 has no literature specifically addressing it. However, and according to TAIR, the computational analysis of a massive data integration of large-scale expression data, functional gene annotations, experimental protein-protein interactions, and transcription factor-target interactions yielded a list of processes which FBN6 could be involved with (Heyndrickx and

Vandepoele, 2012). These processes include: cellular cation homeostasis, divalent metal ion transport, cysteine biosynthesis, glucosinolate biosynthesis, photosynthesis, light reaction, and regulation of protein dephosphorylation. In addition , our expression analysis showed that FBN6 is strongly expressed in tomato leaves, notably less in flowers, and even less in tomato fruit (data not shown), though its expression tends to increase during the ripening process.

FBN7 seems to be absent in algae. This suggests that its function is covered by another fibrillin or that, alternatively, the absence of a specific need which must be satisfied by FBN7 in plants. A localization experiment was reported for Arabidopsis FBN7(a) where it was found that its targeting to the plastoglobule requires their almost complete sequence, except a short stretch at the C-terminal end. It was proposed that targeting of this fibrillin to plastoglobules relies on a global correct folding rather than in a discrete targeting sequence (Shanmugabalaji et al., 2013; Vidi et al., 2006).

FBN8 and FBN9 have no literature associated, and seem to be conserved mainly among land plants, and less so in algae. Our expression analysis showed that both FBN8 and FBN9 increase their expression along the tomato fruit ripening process although their estimated mRNA copy numbers remain relatively low in comparison to FBN2, FBN4 and FBN6, and specially, FBN1.

FBN10 is a special case, being totally conserved in land plants and algae as well as having strikingly close relatives in very distant species as are stramenopiles and diatoms. The fact that it is a dual-module fibrillin distinguishes it from the rest. A remarkable case of a FBN10 including even more additional modules is found in *Fragaria vesca*. The additional domains included seem to confer membrane-anchoring capacity as well as a very-long-fatty-acyl-CoA dehydrogenase activity. The role of a protein having this combination of domains is not clear. It could be speculated that FBN10, as FBN4, may be capable of mobilizing reducing power (for instance from plastoglobules). If this were true, it could make sense that it is cooperating with the fatty acyl-CoA dehydrogenase, as FBN10 could be offering the reducing power needed for the fatty acyl dehydrogenation/desaturation through its fibrillin domains.

The computational study previously mentioned (Heyndrickx and Vandepoele, 2012) also adds information regarding the FBN10 possible function in Arabidopsis, which could be related to the pentose-phosphate shunt, or thylakoid membrane organization. As well, a TAIR database manually-annotated entry suggests that this protein could be related as well to the biosynthetic

process of tryptophan (Mueller et al., 2003) (TAIR Analysis Reference:1445901).

FBN11 is another modular fibrillin, possessing a protein kinase domain at the N-terminal region followed by a fibrillin domain at the C-terminal region. This setup suggests a regulatory role for this protein. The fact that FBN11 is only found in land plants points towards a specific regulatory function found in these organism. A search in NCBI shows that Arabidopsis FBN11 is also called an OBP3-binding protein. OBP3 is a protein which is found at NCBI to be a putatively nuclear-localized protein found to be responsive to SA (which could link its function to FBN4), and which is predominantly expressed in roots. Also according to NCBI, OBP3 has a DNA "Dof" zinc-finger DNA-binding motif. Transgenic overexpressors of OBP3 are reported to have yellow leaves and defective roots (Kang and Singh, 2000). In this context, it seems to make sense that FBN11 is exclusive of land plants, relatedly to their dependence on roots for their development.

FBN12, previously unnamed or simply referred to as a "fibrillin-like protein", also has no specific literature associated to it. However, the computational analysis by Heyndrickx and Vandepoele (2012) suggests an involvement in the following processes: myo-inositol hexakisphosphate biosynthetic process, pentose-phosphate shunt and thylakoid membrane organization. The finding that FBN12 is the fibrillin member with the highest number of FBN12-related modular proteins is deemed to be significant, even more so as two identical setups are found repeated in at least two different organisms, pinpointing its likely usefulness (see Annex IV).

5.2.3. Fibrillins in algae and photosynthetic microorganisms

In algae, FBN3 and FBN10 seem to be conserved in all cases observed, suggesting some essential functionality in these aquatic organisms. In contrast, FBN7 and FBN11 seem to be absent, suggesting that their function is not essential in these organisms.

In cyanobacteria, it could be seen that in some cases cFBN1 had apparently evolved to forms very distantly related to plant FBN1 and even to the consensus cyanobacterial cFBN1 itself. Interestingly, a part of the cyanobacterial species analyzed seemed to have a cyanobacterial FBN12-like protein, we have named cFBN12. This would suggest that in the endosymbiotic event that gave rise to eukaryotic photosynthetic organisms, the endosymbiont belonged to a class of cyanobacteria containing cFBN1 as well as a cFBN12-like protein. Interestingly, in three cyanobacterial species fibrillins distantly resembling FBN3 were found,

raising the possibility that FBN3 could also be of cyanobacterial origin.

In diatoms, the number of fibrillin subfamilies is higher than in cyanobacteria and, in several instances, very distant to plant fibrillins. As a result, it is difficult to trace clear relationships. Curiously, the fibrillin mostly resembling any of its plant relatives is FBN10, though only the first of its two subdomains is present (FBN10 is a two-module protein). The degree of conservation of this fibrillin is very striking, and is unexpected if this degree of sequence conservation is compared to that of any of the other fibrillins found in diatoms versus their plant counterparts. This is even more striking since, up to this point, the most consistently sequence-conserved fibrillin found across species was generally FBN1/cFBN1. From this, it follows that either diatoms do not need FBN1 activity, or that FBN1 is substituted by one of the other fibrillins present in these organisms. FBN1, FBN3 and FBN7 all have in diatoms at least a distant relative to their plant counterparts, but with far less similarity than that found for the surprisingly conserved FBN10. The presence of a FBN7-like protein could also be significant, as this form was shown to be absent in algae. No FBN11-like protein is found, further supporting that this fibrillin is specific of land plants. Finally, in some cases a conserved FBN12-like protein is found, though not in all species searched. Thus, it could be accessory for diatoms, reason by which it could have been dropped, or never acquired in some of the species.

In cryptophytes, the landscape has some similarities to what is found in diatoms, with FBN7, FBN10 and FBN12 all being found in all species searched but being all the forms found quite distant to those of plants. FBN1, FBN6 and FBN8 have in some instances in cryptophytes forms very slightly related to those of higher plants.

A final checking was performed against several apicomplexa, mainly of the *Plasmodium* species. Those microorganisms contain the apicoplast, an organelle resulting from the endosymbiosis of cyanobacteria, as the plastid, but in which photosynthesis has been lost long ago. It was found that none of them had any fibrillin nor any fibrillin-like protein. This further supports the idea that fibrillins are only required in photosynthetic organisms.

5.2.4. Structural modeling of fibrillins

The prediction of the secondary structure of the different fibrillin types yielded what seems to be a relevant finding that was never described before. It refers to the pattern of secondary structure, which was found to be common to all fibrillins. A schematic representation of the secondary structure predicted for fibrillins is shown in Figure 49.

Figure 49. Generalized model of the secondary structure of fibrillins including two initial alpha helices, eight beta chains, and two main variable length regions. The loops drawn represent variable length regions which can in some cases adopt local secondary structures other than random coils.

This general secondary structure outline is well conserved upon all fibrillins of the family from FBN1 to FBN9, and FBN12. The composite fibrillins FBN10 and FBN11 follow this in a more lax way, particularly the second fibrillin domain of FBN10, and the fibrillin domain of FBN11. Nevertheless, the trend can still be found when the secondary structure predictions are carefully analyzed.

This very simple model attempts to unify what this diverse family of proteins have in common in spite of their divergent primary structure. Thus, the strongest unifying structural feature of fibrillins is the presence of two alpha helices followed by eight beta chains with mainly two possible loops of different length (longer linker sequences drawn in the scheme shown in Figure 49). The generalized model of fibrillin secondary structure (two alpha helices + eight beta-chains) is very suggestive of the need for a fibrillin to possess at least eight beta-strands. A tempting plausible explanation is that fibrillins need to form a beta barrel, which is only possible when the protein has eight beta-chains. Regarding the differences in sequence length (Table 18) which are accounted for by the existence of variable-length loops between certain beta-chains, the protein can contain additional small alpha helices, short beta-chains or simply random coils. These loops of variable length may further influence the interaction capacities of these proteins with other elements, like other proteins or membranes.

This common structural trend strongly supports the 3D models predicted for fibrillins (section 4.6.7). During the 3D modelling it was found that mostly barrel-like, planar-like or disorderly models were obtained. However the models offering the best quality scores were those predicting beta-barrel structures. It is to be noted that eight is the number of beta-chains necessary to form a minimal beta-barrel as those predicted for some of the fibrillin consensus sequences.

Several implications need to be considered if likelihood of beta-barrel formation is

accepted. On the one hand, as beta barrels can potentially act as membrane pores usable for the transport of molecules through a membrane, for instance in the thylakoids, fibrillins could have a role in communicating lumen and stroma. On the other hand, beta barrels can act as hydrophobic molecule transporters, as is the case of lipocalins (Grzyb et al., 2006). In this way fibrillins could move from one location to another transporting hydrophobic molecules. Moreover, considering the thylakoid-plastoglobule link in chloroplasts (Austin, 2006) it could be possible that some fibrillins might act as "connectors" in the generation or growth of plastoglobules. Finally, if we consider the possibility for the beta-barrel acting as transport proteins, certain fibrillins could act as "valves" which facilitate the transfer of hydrophobic molecules from one plastidial compartment (i.e. plastoglobule) to another (i.e. thylakoid or crystalloid).

The prediction of 3D structures was repeated to confirm the obtained results. The beta-barrel prediction becomes even more attention-drawing if, after trying to model more distant fibrillins, the beta-barrel structures were shown again. This was found for fibrillins of diatoms *Thalasiosira pseudonana* and *Phaeodactylum tricornutum* and the cryptophyte *Guillardia theta*, despite their big sequence divergence in comparison to their plant fibrillin counterparts. Grossly, though, they shared the commonly predicted two-helix plus eight beta-chain configuration of fibrillins, in this way supporting the beta-barrel structure model. Looking for further hints supporting the proposed model, the analysis of the conservation of the amino acid residues superimposed onto one of the 3D models unveiled the discontinuous or alternate conservation of amino acids in five out of the eight beta chains (Figure 45). This seems related to the need for these specific conserved residues to be oriented in a specific direction; that is, inwards as happens in the beta-barrel. When the results of the ConSurf analysis of the consensus model of cFBN1 are considered, and the conserved amino acids of the beta chains which follow the alternate conservation scheme are highlighted, the image shown in Figure 50 is obtained.
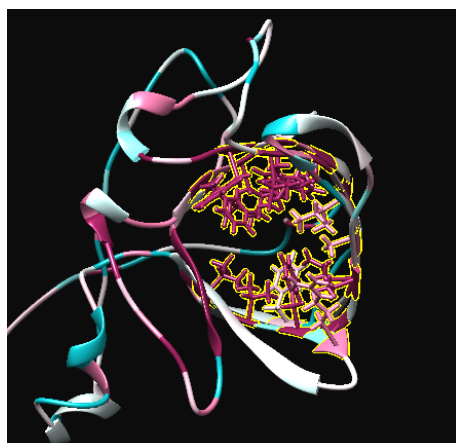
Figure 50. Representation of a modelled cFBN1 beta-barrel displaying the side-chains of the amino acids showing the alternate conservation pattern. Model built using the ConSurf server.

It can be observed that the conserved acids point towards the inside of the beta-barrel. The significance of this could be related with the need to create a specific inner surface (for instance, for interaction with a given molecule) or, alternatively, to help in the formation and/or stabilization of the beta-barrel. This beta-barrel model could be present in higher plant fibrillins, with the exception of FBN12 and perhaps FBN11, which to date have not been successfully modelled into any beta-barrel-like structures. In the case of FBN12, this could be influenced by the presence of a six-residue insertion at FCR1, one of the two most conserved regions in fibrillins. In these conserved sequences, at least one of their respective conserved aromatic amino acids (W, Y) is always found in close vicinity to another aromatic amino acid of another neighbouring beta-chain. This would allow for the formation of π-π interaction among the aromatic rings of these neighbouring aromatic amino acids, likely leading to a stabilization of the beta-barrel structure.

Some data available from the literature actually supports the possibility of the formation of beta-barrels in the fibrillins. Studies on the subcellular localization of the Arabidopsis fibrillin FBN7 concluded that almost the complete protein was required for targeting into plastoglobules (Vidi et al., 2006).Taking a closer look at the experimental design (based on the use of different truncated versions of FBN7) and the predicted secondary structure of FBN7 it can be seen that the only constructs which targeted correctly to the plastoglobule were those containing the eight beta-chains. concluded by the authors that this behaviour would likely be related to the protein requiring a specific correct folding rather than relying on a discrete sequence for correct targeting

to plastoglobule (Vidi et al., 2006). Also in support of a beta-barrel model is the report that FBN4 contains a lipocalin motif (Singh and McNellis, 2011), as lipocalins are small beta-barrel proteins generally related to the binding and transport of hydrophobic molecules, throughout the different kingdoms of life (Grzyb et al., 2006). If we take all the previous into consideration, it follows that the conservation of the beta-chains in fibrillins is very likely related to the need for a specific folding, which most likely would be the formation of a beta-barrel structure.
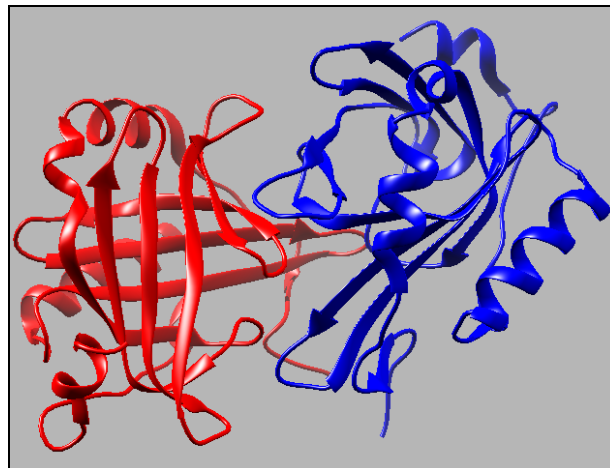
Figure 51. Representation of the 3D structure of the crystallized lobster lipocalin apocrustacyanin C(1) with Pdb code 1OBQ. A dimer is shown, of which the red subunit displays clearly its beta-barrel folding.

In contrast with that assigned role of the beta-chains in the formation of the beta-barrel, the structural and functional role of the two conserved alpha helices is currently unknown. In any case, the alpha helices would be structures complementary to the beta-barrel which would likely be influencing the interaction capabilities of fibrillins with structures other than membranes.

As described above, lipocalins are known to bind small hydrophobic molecules at the inside of their beta-barrel (Grzyb et al., 2006). Thus, considering the structural similarities found between lipocalins and fibrillins, it is reasonable to think that fibrillins could play a similar role concerning the binding of particular plastidial lipids, like carotenoids or tocopherols.The spatial plausibility of this has been checked searching for the structure of an equally scaled carotenoid and comparing with the beta-barrel model obtained (Figures 52 and 53).
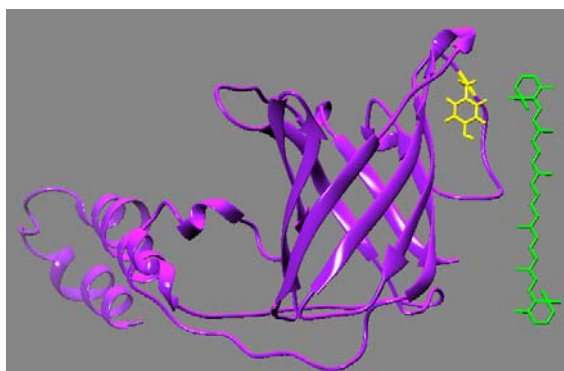
Figure 52. Representation of the predicted 3D structure of cFBN1 and a beta-carotene molecule (green). A tyrosine side-chain of the protein is shown in yellow as a size reference to confirm that the scale used is equivalent.
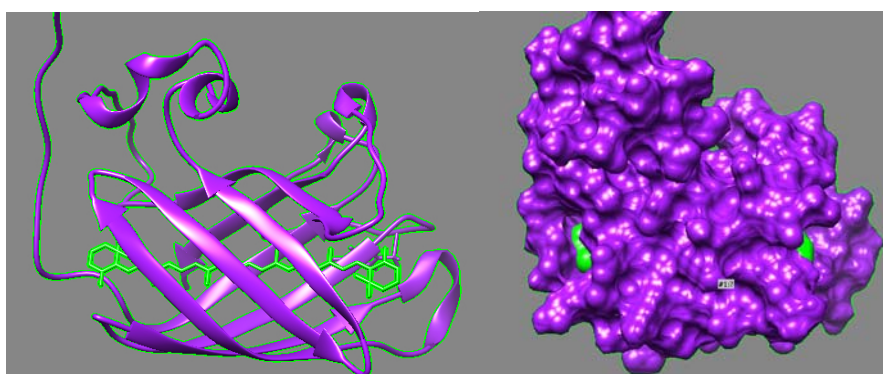
Figure 53. Detail of the beta-barrel, with the carotenoid manually inserted into the beta-barrel. The same configuration is shown with the hypothetical exposed surface calculated.

If only size is considered, the insertion of carotenoids into the beta-barrel seems to be indeed a possibility. Given the variety of plant fibrillins, it can be expected that a variety of different beta-barrels will be able to bind different lipophilic molecules and with different affinities. Digging deeper herein, however, involves experiments which are left for the future.

CONCLUSIONS

6. CONCLUSIONS

-A method for the fractionation of tomato fruit chromoplasts has been set up and used for the isolation of plastoglobules and other chromoplast suborganellar fractions.

-Electron microscopy studies of the isolated chromoplast fractions has allowed the identification of a fraction containing pure plastoglobules and fractions containing both plastoglobules and crystals. The finding of crystals and plastoglobules coexisting in several fractions has uncovered an unknown difficulty inherent to the purification of pure plastoglobules of tomato fruit.

-Metabolite analyses allowed the identification of the chromoplast fractions more largely contributing to the accumulation of lycopene and beta-carotene during ripening. These fractions were coincident with those displaying crystalline structures as well as plastoglobules. Complementary data have also been acquired for other relevant plastid metabolites and membrane lipids.

-The analysis of the relative enrichment of particular metabolites in the isolated chromoplast fractions supports the long-held assertions that lycopene is indeed stored in the form of crystalline structures. Additionally, the obtained results have also revealed that at least the lightest plastoglobules present in tomato fruit chromoplasts do not accumulate lycopene. However, this cannot be discarded for plastoglobules co purifying with lycopene crystals.

-Proteomic studies have contributed to elucidate the identity of the isolated chromoplast fractions. Additionally, they have also offered insights into their general functional context in the chromoplast. During proteomics analysis specific proteins have been identified which are relevant to lycopene accumulation and, particularly, to the formation of lycopene-accumulating structures. The study of these proteins has been further addressed using other approaches.

-The enrichment of the carotenoid precursors phytoene and phytofluene in the fraction containing pure plastoglobules, linked with the finding that plastoglobules contain several

enzymes related to the biosynthesis of lycopene, strongly suggests that plastoglobules could be a main actor at play in the biosynthesis and accumulation of lycopene during tomato fruit ripening.

-The global correlation of the data from metabolite profiling, electron microscopy and proteomics have allowed for the tentative assignation of the identity of the major chromoplast suborganellar fractions. This information can set the basis to undertake further studies in chromoplasts.

-In connection with the finding that fibrillins were the major proteins found in plastoglobules and the proposed relevance of these proteins in carotenoid accumulation, the analysis of the expression of genes encoding chromoplast fibrillins and the silencing of plastoglobule fibrillins were performed. The obtained results corroborated that at least plastoglobule fibrillins have a relevant role in fruit ripening and in the concurrent accumulation of lycopene.

-The bioinformatic analysis of the members of the whole fibrillin family combined with the prediction of their protein structure has allowed to find the underlying common characteristics of this diverse and largely unknown family of proteins. The strongest unifying trait found is at the level of secondary structure, which is very likely to be in connection with the formation of a minimal eight-member beta-barrel structure joined to a double alpha-helix domain. These conclusions provide new insights to further study this protein family in aspects relevant not only in carotenoid accumulation but also in plant development and stress resistance.

BIBLIOGRAPHY

7. BIBLIOGRAPHY

ACM (2012). Advanced consensus maker.

Angaman, D.M., Petrizzo, R., Hernández-Gras, F., Romero-Segura, C., Pateraki, I., Busquets, M., and Boronat, A. (2012). Precursor uptake assays and metabolic analyses in isolated tomato fruit chromoplasts. Plant Methods *8*, 1.

Anunciato, T.P., and da Rocha Filho, P.A. (2012). Carotenoids and polyphenols in nutricosmetics, nutraceuticals, and cosmeceuticals. J. Cosmet. Dermatol. *11*, 51–54.

Arab, L., and Steck, S. (2000). Lycopene and cardiovascular disease. Am. J. Clin. Nutr. *71*, 1691s–1695s.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.

Austin, J.R., Frost, E., Vidi, P.-A., Kessler, F., and Staehelin, L.A. (2006). Plastoglobules Are Lipoprotein Subcompartments of the Chloroplast That Are Permanently Coupled to Thylakoid Membranes and Contain Biosynthetic Enzymes. Plant Cell Online *18*, 1693–1703.

Bahaji, A., Ovecka, M., Bárány, I., Risueño, M.C., Muñoz, F.J., Baroja-Fernández, E., Montero, M., Li, J., Hidalgo, M., Sesma, M.T., et al. (2011). Dual Targeting to Mitochondria and Plastids of AtBT1 and ZmBT1, Two Members of the Mitochondrial Carrier Family. Plant Cell Physiol. *52*, 597–609.

Barsan, C., Sanchez-Bel, P., Rombaldi, C., Egea, I., Rossignol, M., Kuntz, M., Zouine, M., Latché, A., Bouzayen, M., and Pech, J.-C. (2010). Characteristics of the tomato chromoplast revealed by proteomic analysis. J. Exp. Bot. *61*, 2413–2431.

Barsan, C., Zouine, M., Maza, E., Bian, W., Egea, I., Rossignol, M., Bouyssie, D., Pichereaux, C., Purgatto, E., Bouzayen, M., et al. (2012). Proteomic Analysis of Chloroplast-to-Chromoplast Transition in Tomato Reveals Metabolic Shifts Coupled with Disrupted Thylakoid Biogenesis Machinery and Elevated Energy-Production Components. Plant Physiol. *160*, 708–725.

Bathgate, B., Purton, M.E., Grierson, D., and Goodenough, P.W. (1985). Plastid changes during the conversion of chloroplasts to chromoplasts in ripening tomatoes. Planta *165*, 197–204.

Benkert, P., Schwede, T., and Tosatto, S.C. (2009). QMEANclust: estimation of protein model quality by combining a composite scoring function with structural density information. BMC Struct. Biol. *9*, 35.

Besagni, C., and Kessler, F. (2013). A mechanism implicating plastoglobules in thylakoid disassembly during senescence and nitrogen starvation. Planta *237*, 463–470.

Bhasin, M., and Raghava, G.P.S. (2004). ESLpred: SVM-based method for subcellular localization of eukaryotic proteins using dipeptide composition and PSI-BLAST. Nucleic Acids

Res. *32*, W414–W419.

Bick, J.A., and Lange, B.M. (2003). Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. Arch. Biochem. Biophys. *415*, 146–154.

BLIGH, E.G., and DYER, W.J. (1959). A rapid method of total lipid extraction and purification. Can. J. Biochem. Physiol. *37*, 911–917.

Block, M.A., Douce, R., Joyard, J., and Rolland, N. (2007). Chloroplast envelope membranes: a dynamic interface between plastids and the cytosol. Photosynth. Res. *92*, 225–244.

Bombarely, A., Menda, N., Tecle, I.Y., Buels, R.M., Strickler, S., Fischer-York, T., Pujar, A., Leto, J., Gosselin, J., and Mueller, L.A. (2011). The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl. Nucleic Acids Res. *39*, D1149–D1155.

Bouvier, F., Hugueney, P., D'Harlingue, A., Kuntz, M., and Camara, B. (1994). Xanthophyll biosynthesis in chromoplasts: isolation and molecular cloning of an enzyme catalyzing the conversion of 5,6-epoxycarotenoid into ketocarotenoid. Plant J. *6*, 45–54.

Bouvier, F., Rahier, A., and Camara, B. (2005). Biogenesis, molecular regulation and function of plant isoprenoids. Prog. Lipid Res. *44*, 357–429.

Bréhélin, C., and Kessler, F. (2008). The Plastoglobule: A Bag Full of Lipid Biochemistry Tricks†. Photochem. Photobiol. *84*, 1388–1394.

Bréhélin, C., Kessler, F., and van Wijk, K.J. (2007). Plastoglobules: versatile lipoprotein particles in plastids. Trends Plant Sci. *12*, 260–266.

Breuers, F.K.H., Bräutigam, A., and Weber, A.P.M. (2011). The plastid outer envelope – a highly dynamic interface between plastid and cytoplasm. Front. Plant Physiol. *2*, 97.

Britton, G. (1995). Structure and properties of carotenoids in relation to function. FASEB J. *9*, 1551–1558.

Bruce, B.D. (1998). The role of lipids in plastid protein transport. Plant Mol. Biol. *38*, 223–246.

Bugos, R.C., Chiang, V.L., Zhang, X.H., Campbell, E.R., Podila, G.K., and Campbell, W.H. (1995). RNA isolation from plant tissues recalcitrant to extraction in guanidine. BioTechniques *19*, 734–737.

Callister, S.J., McCue, L.A., Turse, J.E., Monroe, M.E., Auberry, K.J., Smith, R.D., Adkins, J.N., and Lipton, M.S. (2008). Comparative Bacterial Proteomics: Analysis of the Core Genome Concept. PLoS ONE *3*, e1542.

Camara, B., Hugueney, P., Bouvier, F., Kuntz, M., and Monéger, R. (1995). Biochemistry and molecular biology of chromoplast development. Int. Rev. Cytol. *163*, 175–247.

Cazzonelli, C.I., and Pogson, B.J. (2010). Source to sink: regulation of carotenoid biosynthesis in plants. Trends Plant Sci. *15*, 266–274.

Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T., and Ben-Tal, N. (2013). ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. Isr. J. Chem. *53*, 199–206.

Chan, C.X., Gross, J., Yoon, H.S., and Bhattacharya, D. (2011). Plastid Origin and Evolution: New Models Provide Insights into Old Problems. Plant Physiol. *155*, 1552–1560.

Chandna, R., Augustine, R., and Bisht, N.C. (2012). Evaluation of Candidate Reference Genes for Gene Expression Normalization in Brassica juncea Using Real Time Quantitative RT-PCR. PLoS ONE 7, e36918.

Chen, Z., Zeng, M., Song, B., Hou, C., Hu, D., Li, X., Wang, Z., Fan, H., Bi, L., Liu, J., et al. (2012). Dufulin Activates HrBP1 to Produce Antiviral Responses in Tobacco. PLoS ONE *7*, e37944.

Choquet, Y., and Wollman, F.-A. (2002). Translational regulations as specific traits of chloroplast gene expression. FEBS Lett. *529*, 39–42.

Chrispeels, M.J., and Agre, P. (1994). Aquaporins: water channel proteins of plant and animal cells. Trends Biochem. Sci. *19*, 421–425.

Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Res. *36*, W197–W201.

Conesa, A., and Gotz, S. (2008). Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. Int. J. Plant Genomics *2008*.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. *16*, 10881–10890.

Davis, M.J., Sehgal, M.S.B., and Ragan, M.A. (2010). Automatic, context-specific generation of Gene Ontology slims. BMC Bioinformatics *11*, 498.

Deruère, J., Römer, S., d' Harlingue, A., Backhaus, R.A., Kuntz, M., and Camara, B. (1994). Fibril assembly and carotenoid overaccumulation in chromoplasts: a model for supramolecular lipoprotein structures. Plant Cell Online *6*, 119–133.

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res. *38*, W64–W70.

Durocher, D., and Jackson, S.P. (2002). The FHA domain. FEBS Lett. 513, 58–66.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Egea, I., Barsan, C., Bian, W., Purgatto, E., Latché, A., Chervin, C., Bouzayen, M., and Pech, J.-C. (2010). Chromoplast Differentiation: Current Status and Perspectives. Plant Cell Physiol. *51*, 1601–1611.

Emanuelsson, O., Nielsen, H., and Heijne, G.V. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci. *8*, 978–984.

Emanuelsson, O., Brunak, S., Heijne, G. von, and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. *2*, 953–971.

Facchinelli, F., and Weber, A.P.M. (2011). The Metabolite Transporters of the Plastid Envelope: An Update. Front. Plant Sci. *2*.

Fei, Z., Joung, J.-G., Tang, X., Zheng, Y., Huang, M., Lee, J.M., McQuinn, R., Tieman, D.M., Alba, R., Klee, H.J., et al. (2011). Tomato Functional Genomics Database: a comprehensive resource and analysis package for tomato functional genomics. Nucleic Acids Res. *39*, D1156–D1163.

Ferro, M., Salvi, D., Rivière-Rolland, H., Vermat, T., Seigneurin-Berny, D., Grunwald, D., Garin, J., Joyard, J., and Rolland, N. (2002). Integral membrane proteins of the chloroplast envelope: Identification and subcellular localization of new transporters. Proc. Natl. Acad. Sci. *99*, 11487–11492.

Ferro, M., Brugière, S., Salvi, D., Seigneurin-Berny, D., Court, M., Moyet, L., Ramus, C., Miras, S., Mellal, M., Gall, S.L., et al. (2010). AT_CHLORO, a Comprehensive Chloroplast Proteome Database with Subplastidial Localization and Curated Information on Envelope Proteins. Mol. Cell. Proteomics *9*, 1063–1084.

Fisunov, G.Y., Alexeev, D.G., Bazaleev, N.A., Ladygina, V.G., Galyamina, M.A., Kondratov, I.G., Zhukova, N.A., Serebryakova, M.V., Demina, I.A., and Govorun, V.M. (2011). Core Proteome of the Minimal Cell: Comparative Proteomics of Three Mollicute Species. PLoS ONE *6*, e21964.

Flügge, U.-I., Häusler, R.E., Ludewig, F., and Gierth, M. (2011). The role of transporters in supplying energy to plant plastids. J. Exp. Bot. *62*, 2381–2392.

Fraser, P.D., and Bramley, P.M. (2004). The biosynthesis and nutritional uses of carotenoids. Prog. Lipid Res. *43*, 228–265.

Fraser, P.D., Pinto, M.E.S., Holloway, D.E., and Bramley, P.M. (2000). Application of high-performance liquid chromatography with photodiode array detection to the metabolic profiling of plant isoprenoids. Plant J. *24*, 551–558.

Fraser, P.D., Enfissi, E.M.A., and Bramley, P.M. (2009). Genetic engineering of carotenoid formation in tomato fruit and the potential application of systems and synthetic biology approaches. Arch. Biochem. Biophys. *483*, 196–204.

Friso, G., Giacomelli, L., Ytterberg, A.J., Peltier, J.-B., Rudella, A., Sun, Q., and Wijk, K.J. van

(2004). In-Depth Analysis of the Thylakoid Membrane Proteome of Arabidopsis thaliana Chloroplasts: New Proteins, New Functions, and a Plastid Proteome Database. Plant Cell Online *16*, 478–499.

Froehlich, J.E., Wilkerson, C.G., Ray, W.K., McAndrew, R.S., Osteryoung, K.W., Gage, D.A., and Phinney, B.S. (2003). Proteomic Study of the Arabidopsis thaliana Chloroplastic Envelope Membrane Utilizing Alternatives to Traditional Two-Dimensional Electrophoresis. J. Proteome Res. *2*, 413–425.

Gautier, R., Douguet, D., Antonny, B., and Drin, G. (2008). HELIQUEST: a web server to screen sequences with specific α-helical properties. Bioinformatics *24*, 2101–2102.

Giovannoni, J.J. (2004). Genetic Regulation of Fruit Development and Ripening. Plant Cell Online 16, S170–S180.

Giovannoni, J.J. (2007). Fruit ripening mutants yield insights into ripening control. Curr. Opin. Plant Biol. *10*, 283–289.

Goldman, N. (1990). Maximum Likelihood Inference of Phylogenetic Trees, with Special Reference to a Poisson Process Model of DNA Substitution and to Parsimony Analyses. Syst. Biol. *39*, 345–361.

Gordon, H.T., Bauernfeind, J.C., and Furia, T.E. (1982). Carotenoids as food colorants. C R C Crit. Rev. Food Sci. Nutr. *18*, 59–97.

Grzyb, J., Latowski, D., and Strzałka, K. (2006). Lipocalins – a family portrait. J. Plant Physiol. 163, 895–915.

Harris, W.M. (1970). Chromoplasts of tomato fruits. III. The high-delta tomato. Bot. Gaz. *131*, 163–166.

Harris, W.M., and Spurr, A.R. (1969). Chromoplasts of Tomato Fruits. II. The Red Tomato. Am. J. Bot. *56*, 380.

Hayat, S., and Elofsson, A. (2012). BOCTOPUS: improved topology prediction of transmembrane β barrel proteins. Bioinformatics *28*, 516–522.

Heazlewood, J.L., Verboom, R.E., Tonti-Filippini, J., Small, I., and Millar, A.H. (2007). SUBA: the Arabidopsis Subcellular Database. Nucleic Acids Res. *35*, D213–D218.

Heber, D., and Lu, Q.-Y. (2002). Overview of Mechanisms of Action of Lycopene. Exp. Biol. Med. *227*, 920–923.

Heyndrickx, K.S., and Vandepoele, K. (2012). Systematic Identification of Functional Plant Modules through the Integration of Complementary Data Sources. Plant Physiol. *159*, 884–901.

Hirschberg, J. (2001). Carotenoid biosynthesis in flowering plants. Curr. Opin. Plant Biol. *4*, 210–218.

Hofmann, N.R. (2008a). Regulation of Plastid Gene Expression in the Chloroplast-to-Chromoplast Transition. Plant Cell Online *20*, 823–823.

Hofmann, N.R. (2008b). An Endoplasmic Reticulum Protein Involved in Lipid Transfer to Chloroplasts. Plant Cell Online *20*, 2007–2007.

Holzapfel, N.P., Holzapfel, B.M., Champ, S., Feldthusen, J., Clements, J., and Hutmacher, D.W. (2013). The Potential Role of Lycopene for the Prevention and Therapy of Prostate Cancer: From Molecular Mechanisms to Clinical Evidence. Int. J. Mol. Sci. *14*, 14620–14646.

Hooft, R.W.W., Vriend, G., Sander, C., and Abola, E.E. (1996). Errors in protein structures. Nature *381*, 272–272.

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor. Nucleic Acids Res. *35*, W585–W587.

Howitt, C.A., and Pogson, B.J. (2006). Carotenoid accumulation and function in seeds and non-green tissues. Plant Cell Environ. *29*, 435–445.

Immethun, C.M., Hoynes-O'Connor, A.G., Balassy, A., and Moon, T.S. (2013). Microbial production of isoprenoids enabled by synthetic biology. Front. Microbiotechnology Ecotoxicol. Bioremediation *4*, 75.

Inoue, K. (2011). Emerging roles of the chloroplast outer envelope membrane. Trends Plant Sci. *16*, 550–557.

Itkin, M., Seybold, H., Breitel, D., Rogachev, I., Meir, S., and Aharoni, A. (2009). TOMATO AGAMOUS-LIKE 1 is a component of the fruit ripening regulatory network. Plant J. *60*, 1081–1095.

Johnson, M. (2013). Detergents: Triton X-100, Tween-20, and More. Labome Mater. Methods.

Kang, H.-G., and Singh, K.B. (2000). Characterization of salicylic acid-responsive, Arabidopsis Dof domain proteins: overexpression of OBP3 leads to growth defects. Plant J. 21, 329–339.

Kahlau, S., and Bock, R. (2008). Plastid Transcriptomics and Translatomics of Tomato Fruit Development and Chloroplast-to-Chromoplast Differentiation: Chromoplast Gene Expression Largely Serves the Production of a Single Protein. Plant Cell Online *20*, 856–874.

Kilambi, H.V., Kumar, R., Sharma, R., and Sreelakshmi, Y. (2013). Chromoplast-Specific Carotenoid-Associated Protein Appears to Be Important for Enhanced Accumulation of Carotenoids in hp1 Tomato Fruits. Plant Physiol. *161*, 2085–2101.

Kim, D.E., Chivian, D., and Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. *32*, W526–W531.

Klee, H.J. (2002). Control of ethylene-mediated processes in tomato at the level of receptors. J. Exp. Bot. 53, 2057–2063.

Klee, H.J., and Giovannoni, J.J. (2011). Genetics and Control of Tomato Fruit Ripening and Quality Attributes. Annu. Rev. Genet. *45*, 41–59.

Kleffmann, T., Hirsch-Hoffmann, M., Gruissem, W., and Baginsky, S. (2006). plprot: A Comprehensive Proteome Database for Different Plastid Types. Plant Cell Physiol. *47*, 432–436.

Kovács-Bogdán, E., Soll, J., and Bölter, B. (2010). Protein import into chloroplasts: The Tic complex and its regulation. Biochim. Biophys. Acta BBA - Mol. Cell Res. 1803, 740–747.

Kozlowski, L.P., and Bujnicki, J.M. (2012). MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. BMC Bioinformatics *13*, 111.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L.. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580.

Hussain, N., Irshad, F., Jabeen, Z., Shamsi, I.H., Li, Z., and Jiang, L. (2013). Biosynthesis, Structural, and Functional Attributes of Tocopherols in Planta; Past, Present, and Future Perspectives. J. Agric. Food Chem. 61, 6137–6149.

Leclercq, J., Ranty, B., Sanchez-Ballesta, M.-T., Li, Z., Jones, B., Jauneau, A., Pech, J.-C., Latché, A., Ranjeva, R., and Bouzayen, M. (2005). Molecular and biochemical characterization of LeCRK1, a ripening-associated tomato CDPK-related kinase. J. Exp. Bot. *56*, 25–35.

Lee, D.W., Jung, C., and Hwang, I. (2013). Cytosolic events involved in chloroplast protein targeting. Biochim. Biophys. Acta BBA - Mol. Cell Res. *1833*, 245–252.

Lee, J.M., Joung, J.-G., McQuinn, R., Chung, M.-Y., Fei, Z., Tieman, D., Klee, H., and Giovannoni, J. (2012). Combined transcriptome, genetic diversity and metabolite profiling in tomato fruit reveals that the ethylene response factor SlERF6 plays an important role in ripening and carotenoid accumulation. Plant J. *70*, 191–204.

Leitner-Dagan, Y., Ovadis, M., Shklarman, E., Elad, Y., David, D.R., and Vainstein, A. (2006). Expression and Functional Analyses of the Plastid Lipid-Associated Protein CHRC Suggest Its Role in Chromoplastogenesis and Stress. Plant Physiol. *142*, 233–244.

Lenucci, M.S., Serrone, L., De Caroli, M., Fraser, P.D., Bramley, P.M., Piro, G., and Dalessandro, G. (2012). Isoprenoid, Lipid, and Protein Contents in Intact Plastids Isolated from Mesocarp Cells of Traditional and High-Pigment Tomato Cultivars at Different Ripening Stages. J. Agric. Food Chem. *60*, 1764–1775.

Lin, Z., Zhong, S., and Grierson, D. (2009). Recent advances in ethylene research. J. Exp. Bot. *60*, 3311–3336.

Lu, R., Martin-Hernandez, A.M., Peart, J.R., Malcuit, I., and Baulcombe, D.C. (2003). Virus-induced gene silencing in plants. Methods 30, 296–303.

Lundquist, P.K., Poliakov, A., Bhuiyan, N.H., Zybailov, B., Sun, Q., and Wijk, K.J. van (2012).

The Functional Network of the Arabidopsis Plastoglobule Proteome Based on Quantitative Proteomics and Genome-Wide Coexpression Analysis. Plant Physiol. *158*, 1172–1192.

Lundquist, P.K., Poliakov, A., Giacomelli, L., Friso, G., Appel, M., McQuinn, R.P., Krasnoff, S.B., Rowland, E., Ponnala, L., Sun, Q., et al. (2013). Loss of Plastoglobule Kinases ABC1K1 and ABC1K3 Causes Conditional Degreening, Modified Prenyl-Lipids, and Recruitment of the Jasmonic Acid Pathway. Plant Cell Online *25*, 1818–1839.

Majeran, W., Friso, G., Asakura, Y., Qu, X., Huang, M., Ponnala, L., Watkins, K.P., Barkan, A., and Wijk, K.J. van (2012). Nucleoid-Enriched Proteomes in Developing Plastids and Chloroplasts from Maize Leaves: A New Conceptual Framework for Nucleoid Functions. Plant Physiol. *158*, 156–189.

Manjunatha, G., Lokesh, V., and Neelwarne, B. (2010). Nitric oxide in fruit ripening: Trends and opportunities. Biotechnol. Adv. *28*, 489–499.

McLaren, D., and Kraemer, K. (2012). Vitamin A in nature. In Manual on Vitamin A Deficiency Disorders (VADD), (Basel: Karguer), pp. 7–17.

Mueller, L.A., Zhang, P., and Rhee, S.Y. (2003). AraCyc: A Biochemical Pathway Database for Arabidopsis. Plant Physiol. *132*, 453–460.

Nacir, H., and Bréhélin, C. (2013). When proteomics reveals unsuspected roles: the plastoglobule example. Front. Plant Proteomics *4*, 114.

Nagai, S., Ohara, K., and Mukai, K. (2005). Kinetic Study of the Quenching Reaction of Singlet Oxygen by Flavonoids in Ethanol Solution. J. Phys. Chem. B *109*, 4234–4240.

Nakatsuka, A., Murachi, S., Okunishi, H., Shiomi, S., Nakano, R., Kubo, Y., and Inaba, A. (1998). Differential Expression and Internal Feedback Regulation of 1-Aminocyclopropane-1-Carboxylate Synthase, 1-Aminocyclopropane-1-Carboxylate Oxidase, and Ethylene Receptor Genes in Tomato Fruit during Development and Ripening. Plant Physiol. *118*, 1295–1305.

Nath, P., Trivedi, P.K., Sane, V.A., and Sane, A.P. (2006). Role of Ethylene in Fruit Ripening. In Ethylene Action in Plants, D.N.A. Khan, ed. (Springer Berlin Heidelberg), pp. 151–184.

Nelson, D.L., and Cox, M.M. (2008). Lehninger - Principles of Biochemistry (W.H. Freeman and Company).

Orzaez, D., Mirabel, S., Wieland, W.H., and Granell, A. (2006). Agroinjection of Tomato Fruits. A Tool for Rapid Functional Analysis of Transgenes Directly in Fruit. Plant Physiol. *140*, 3–11.

Orzaez, D., Medina, A., Torre, S., Fernández-Moreno, J.P., Rambla, J.L., Fernández-del-Carmen, A., Butelli, E., Martin, C., and Granell, A. (2009). A Visual Reporter System for Virus-Induced Gene Silencing in Tomato Fruit Based on Anthocyanin Accumulation. Plant Physiol. *150*, 1122–1134.

Osorio, S., Scossa, F., and Fernie, A.R. (2013). Molecular regulation of fruit ripening. Front.

Plant Syst. Biol. *4*, 198.

Parrilla, J., Muñoz, M., and Moreno, F. (2007). Endurance and durability in biodiesel powered engines. In International Conference on Renewable Energies and Power Quality (ICREPQ´07), (Sevilla).

Pateraki I., Renato M., Azcón-Bieto J., and Boronat A. (2013). An ATP synthase harboring an atypical γ-subunit is involved in ATP synthesis in tomato fruit chromoplasts. Plant J. *74(1)*, 74-85.

Peltier, J.-B., Ytterberg, A.J., Sun, Q., and Wijk, K.J. van (2004). New Functions of the Thylakoid Membrane Proteome of Arabidopsis thaliana Revealed by a Simple, Fast, and Versatile Fractionation Strategy. J. Biol. Chem. *279*, 49367–49383.

Perez-Fons, L., Steiger, S., Khaneja, R., Bramley, P.M., Cutting, S.M., Sandmann, G., and Fraser, P.D. (2011). Identification and the developmental formation of carotenoid pigments in the yellow/orange Bacillus spore-formers. Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids *1811*, 177–185.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Pierleoni, A., Martelli, P.L., Fariselli, P., and Casadio, R. (2006). BaCelLo: a balanced subcellular localization predictor. Bioinformatics *22*, e408–e416.

Piller, L.E., Abraham, M., Dörmann, P., Kessler, F., and Besagni, C. (2012). Plastid lipid droplets at the crossroads of prenylquinone metabolism. J. Exp. Bot. *63*, 1609–1618.

Pozueta-Romero, J., Rafia, F., Houlne, G., Cheniclet, C., Carde, J.P., Schantz, M.L., and Schantz, R. (1997). A ubiquitous plant housekeeping gene, PAP, encodes a major protein component of bell pepper chromoplasts. Plant Physiol. *115*, 1185–1194.

Ralph, S.A. (2005). Strange organelles –Plasmodium mitochondria lack a pyruvate dehydrogenase complex. Mol. Microbiol. *55*, 1–4.

Rodriguez, R., Chinea, G., Lopez, N., Pons, T., and Vriend, G. (1998). Homology modeling, model and software evaluation: three related resources. Bioinformatics *14*, 523–528.

Rodriguez-Amaya, D.B. (2010). Quantitative analysis, in vitro assessment of bioavailability and antioxidant activity of food carotenoids—A review. J. Food Compos. Anal. *23*, 726–740.

Rodriguez-Amaya, B.B., and Kimura, M. (2004). HarvesPlus Handbook for Carotenoid Analysis (Washington, DC and Cali: International Food Policy Research Institute; International Center for Tropical Agriculture).

Rodriguez-Concepción, M., and Boronat, A. (2002). Elucidation of the Methylerythritol Phosphate Pathway for Isoprenoid Biosynthesis in Bacteria and Plastids. A Metabolic Milestone

Achieved through Genomics. Plant Physiol. *130*, 1079–1089.

Rodríguez-Concepción, M. (2010). Supply of precursors for carotenoid biosynthesis in plants. Arch. Biochem. Biophys. *504*, 118–122.

Ruoslahti, E. (1996). Rgd and Other Recognition Sequences for Integrins. Annu. Rev. Cell Dev. Biol. *12*, 697–715.

Rutschow, H., Ytterberg, A.J., Friso, G., Nilsson, R., and Wijk, K.J. van (2008). Quantitative Proteomics of a Chloroplast SRP54 Sorting Mutant and Its Genetic Interactions with CLPC1 in Arabidopsis. Plant Physiol. *148*, 156–175.

Sakurai, I., Mizusawa, N., Wada, H., and Sato, N. (2007). Digalactosyldiacylglycerol Is Required for Stabilization of the Oxygen-Evolving Complex in Photosystem II. Plant Physiol. *145*, 1361–1370.

Schomburg, I., Chang, A., and Schomburg, D. (2002). BRENDA, enzyme data and metabolic information. Nucleic Acids Res. *30*, 47–49.

Schulze, S., Mant, A., Kossmann, J., and Lloyd, J.R. (2004). Identification of an Arabidopsis inorganic pyrophosphatase capable of being imported into chloroplasts. FEBS Lett. *565*, 101–105.

Schweiggert, R.M., Steingass, C.B., Heller, A., Esquivel, P., and Carle, R. (2011). Characterization of chromoplasts and carotenoids of red- and yellow-fleshed papaya (Carica papaya L.). Planta *234*, 1031–1044.

Shanmugabalaji, V., Besagni, C., Piller, L.E., Douet, V., Ruf, S., Bock, R., and Kessler, F. (2013). Dual targeting of a mature plastoglobulin/fibrillin fusion protein to chloroplast plastoglobules and thylakoids in transplastomic tobacco plants. Plant Mol. Biol. *81*, 13–25.

Sheiner, L., and Striepen, B. (2013). Protein sorting in complex plastids. Biochim. Biophys. Acta BBA - Mol. Cell Res. *1833*, 352–359.

Shi, L.-X., and Theg, S.M. (2013). The chloroplast protein import system: From algae to trees. Biochim. Biophys. Acta BBA - Mol. Cell Res. *1833*, 314–331.

Simkin, A.J., Gaffé, J., Alcaraz, J.-P., Carde, J.-P., Bramley, P.M., Fraser, P.D., and Kuntz, M. (2007). Fibrillin influence on plastid ultrastructure and pigment content in tomato fruit. Phytochemistry *68*, 1545–1556.

Simm, S., Papasotiriou, D.G., Ibrahim, M., Leisegang, M.S., Müller, B., Schorge, T., Karas, M., Mirus, O., Sommer, M.S., and Schleiff, E. (2013). Defining the core proteome of the chloroplast envelope membranes. Front. Plant Proteomics *4*, 11.

Singh, D.K., and McNellis, T.W. (2011). Fibrillin protein function: the tip of the iceberg? Trends Plant Sci. *16*, 432–441.

Singh, D.K., Maximova, S.N., Jensen, P.J., Lehman, B.L., Ngugi, H.K., and McNellis, T.W. (2010). FIBRILLIN4 Is Required for Plastoglobule Development and Stress Resistance in Apple and Arabidopsis1[W][OA]. Plant Physiol. *154*, 1281–1293.

Singh, D.K., Laremore, T.N., Smith, P.B., Maximova, S.N., and McNellis, T.W. (2012). Knockdown of FIBRILLIN4 Gene Expression in Apple Decreases Plastoglobule Plastoquinone Content. PLoS ONE *7*, e47547.

Spitzer, M., Fuellen, G., Cullen, P., and Lorkowski, S. (2004). VisCoSe: visualization and comparison of consensus sequences. Bioinformatics *20*, 433–435.

Spurr, A.R., and Harris, W.M. (1968). Ultrastructure of Chloroplasts and Chromoplasts in Capsicum annuum. I. Thylakoid Membrane Changes During Fruit Ripening. Am. J. Bot. *55*, 1210.

Srivastava, A., and Handa, A.K. (2005). Hormonal Regulation of Tomato Fruit Development: A Molecular Perspective. J. Plant Growth Regul. *24*, 67–82.

Stahl, W., and Sies, H. (2003). Antioxidant activity of carotenoids. Mol. Aspects Med. *24*, 345–351.

Steinmüller, D., and Tevini, M. (1985). Composition and function of plastoglobuli. Planta *163*, 201–207.

Summers, L.A. (1980). The bypyridinium herbicides (New York: Academic Press).

Sun, Q., Zybailov, B., Majeran, W., Friso, G., Olinares, P.D.B., and Wijk, K.J. van (2009). PPDB, the Plant Proteomics Database at Cornell. Nucleic Acids Res. *37*, D969–D974.

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLoS ONE *6*, e21800.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol. *28*, 2731–2739.

Tang, S., Li, T., Cong, P., Xiong, W., Wang, Z., and Sun, J. (2013). PlantLoc: an accurate web server for predicting plant protein subcellular localization by substantiality motif. Nucleic Acids Res. *41*, W441–W447.

Taylor, M.A., Arif, S.A., Kumar, A., Davies, H.V., Scobie, L.A., Pearce, S.R., and Flavell, A.J. (1992). Expression and sequence analysis of cDNAs induced during the early stages of tuberisation in different organs of the potato plant (Solanum tuberosum L.). Plant Mol. Biol. *20*, 641–651.

The Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature *485*, 635–641.

Tieman, D.M., Taylor, M.G., Ciardi, J.A., and Klee, H.J. (2000). The tomato ethylene receptors NR and LeETR4 are negative regulators of ethylene response and exhibit functional compensation within a multigene family. Proc. Natl. Acad. Sci. *97*, 5663–5668.

Unver, T., and Budak, H. (2009). Virus-Induced Gene Silencing, a Post Transcriptional Gene Silencing Method. Int. J. Plant Genomics 2009.

Vidi, P.-A., Kanwischer, M., Baginsky, S., Austin, J.R., Csucs, G., Dörmann, P., Kessler, F., and Bréhélin, C. (2006). Tocopherol Cyclase (VTE1) Localization and Vitamin E Accumulation in Chloroplast Plastoglobule Lipoprotein Particles. J. Biol. Chem. *281*, 11225–11234.

Vidi, P.-A., Kessler, F., and Bréhélin, C. (2007). Plastoglobules: a new address for targeting recombinant proteins in the chloroplast. BMC Biotechnol. *7*, 4.

Vishnevetsky, M., Ovadis, M., and Vainstein, A. (1999). Carotenoid sequestration in plants: the role of carotenoid-associated proteins. Trends Plant Sci. *4*, 232–235.

Vranová, E., Coman, D., and Gruissem, W. (2013). Network Analysis of the MVA and MEP Pathways for Isoprenoid Synthesis. Annu. Rev. Plant Biol. *64*, 665–700.

Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. J. Mol. Graph. *8*, 52–56, 29.

Walker, J.M. (2005). The proteomics Protocols Handbook (Humana Press).

Wan, S., Mak, M.-W., and Kung, S.-Y. (2012). mGOASVM: Multi-label protein subcellular localization based on gene ontology and support vector machines. BMC Bioinformatics *13*, 290.

Weber, A.P.M., Schwacke, R., and Flügge, U.-I. (2005). Solute Transporters of the Plastid Envelope Membrane. Annu. Rev. Plant Biol. *56*, 133–164.

Wei, Z.M., Laby, R.J., Zumoff, C.H., Bauer, D.W., He, S.Y., Collmer, A., and Beer, S.V. (1992). Harpin, elicitor of the hypersensitive response produced by the plant pathogen Erwinia amylovora. Science *257*, 85–88.

Weis, B.L., Schleiff, E., and Zerges, W. (2013). Protein targeting to subcellular organelles via mRNA localization. Biochim. Biophys. Acta BBA - Mol. Cell Res. *1833*, 260–273.

Wicke, S., Schneeweiss, G.M., dePamphilis, C.W., Müller, K.F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol. Biol. *76*, 273–297.

Wise, P.R.R. (2006). The Diversity of Plastid Form and Function. In The Structure and Function of Plastids, P.R.R. Wise, and J.K. Hoober, eds. (Springer Netherlands), pp. 3–26.

Xu, C., Fan, J., Cornish, A.J., and Benning, C. (2008). Lipid Trafficking between the Endoplasmic Reticulum and the Plastid in Arabidopsis Requires the Extraplastidic TGD4 Protein. Plant Cell Online *20*, 2190–2204.

Xu, L., Carrie, C., Law, S.R., Murcha, M.W., and Whelan, J. (2013). Acquisition, Conservation, and Loss of Dual-Targeted Proteins in Land Plants. Plant Physiol. *161*, 644–662.

Yang, Y., Sulpice, R., Himmelbach, A., Meinhard, M., Christmann, A., and Grill, E. (2006). Fibrillin expression is regulated by abscisic acid response regulators and is involved in abscisic acid-mediated photoprotection. Proc. Natl. Acad. Sci. *103*, 6061–6066.

Youssef, A., Laizet, Y., Block, M.A., Maréchal, E., Alcaraz, J.-P., Larson, T.R., Pontier, D., Gaffé, J., and Kuntz, M. (2010). Plant lipid-associated fibrillin proteins condition jasmonate production under photosynthetic stress. Plant J. *61*, 436–445.

Ytterberg, A.J., Peltier, J.-B., and Wijk, K.J. van (2006). Protein Profiling of Plastoglobules in Chloroplasts and Chromoplasts. A Surprising Site for Differential Accumulation of Metabolic Enzymes. Plant Physiol. *140*, 984–997.

Zbierzak, A.M., Dörmann, P., and Hölzl, G. (2011). Analysis of Lipid Content and Quality in Arabidopsis Plastids. In Chloroplast Research in Arabidopsis, R.P. Jarvis, ed. (Humana Press), pp. 411–426.

Zhang, M., Yuan, B., and Leng, P. (2009). The role of ABA in triggering ethylene biosynthesis and ripening of tomato fruit. J. Exp. Bot. *60*, 1579–1588.

Zaman, A., and Fancy, N.N. (2012). A computational prediction of structure and function of novel homologue of Arabidopsis thaliana Vps51/Vps67 subunit in Corchorus olitorius. Interdiscip. Sci. Comput. Life Sci. 4, 256–267.

ANNEXES

8. Annexes

Annex I

List of proteins identified in fraction IIa (part I):

ID:	Description
Solyc01g005620.2.1	"Mitochondrial 2-oxoglutarate/malate carrier protein
Solyc01g006540.2.1	"Lipoxygenase
Solyc01g007320.2.1	"ATP synthase subunit beta chloroplastic
Solyc01g007330.2.1	"Ribulose bisphosphate carboxylase large chain
Solyc01g007380.1.1	"Apocytochrome f
Solyc01g073690.2.1	"V-type ATP synthase subunit D
Solyc01g079090.2.1	"Protoporphyrinogen oxidase
Solyc01g097810.2.1	Zeta-carotene desaturase
Solyc01g105340.2.1	"Chaperone protein dnaJ
Solyc01g110360.2.1	"Fructose-bisphosphate aldolase
Solyc01g111510.2.1	"Ascorbate peroxidase
Solyc01g111760.2.1	"V-type ATP synthase beta chain
Solyc02g049070.2.1	Genomic DNA chromosome 5 P1 clone MAH20
Solyc02g062340.2.1	"Fructose-bisphosphate aldolase
Solyc02g062600.2.1	"2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase
Solyc02g067460.2.1	"Porin/voltage-dependent anion-selective channel protein
Solyc02g070490.2.1	"Alpha/beta hydrolase fold
Solyc02g070770.2.1	"NAD-dependent epimerase/dehydratase "
Solyc02g080540.1.1	"ATP synthase gamma chain
Solyc02g081170.2.1	"Plastid-lipid-associated protein, chloroplastic
Solyc02g084440.2.1	"Fructose-bisphosphate aldolase
Solyc02g088270.2.1	"Genomic DNA chromosome 5 P1 clone MUL8
Solyc02g092440.2.1	"Mitochondrial porin
Solyc03g005000.2.1	"Peptidase M48 Ste24p
Solyc03g007740.2.1	"Reticulon family protein
Solyc03g025950.2.1	"Membrane-associated progesterone receptor component 1
Solyc03g043750.2.1	"FAD-dependent pyridine nucleotide-disulphide oxidoreductase
Solyc03g062790.2.1	cDNA clone 002-130-C06 full insert sequence
Solyc03g078400.2.1	"Actin
Solyc03g095620.2.1	"ABC-1 domain protein
Solyc03g097440.2.1	"11-beta-hydroxysteroid dehydrogenase-like
Solyc03g112870.2.1	"Glucose-6P/phosphate translocator
Solyc03g113220.2.1	"SPFH domain/band 7 family protein
Solyc03g116110.2.1	Alpha/beta hydrolase fold protein
Solyc04g050470.2.1	Genomic DNA chromosome 5 P1 clone MNJ8
Solyc04g080570.2.1	"2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase
Solyc05g008600.2.1	"Fructose-bisphosphate aldolase
Solyc05g017950.2.1	"Bile acid sodium symporter
Solyc05g032660.2.1	"Dehydrogenase/ reductase 3
Solyc06g009960.1.1	"Elongation factor 1-alpha
Solyc06g034220.2.1	"Vesicle-associated membrane protein-associated protein
Solyc06g059740.2.1	"Alcohol dehydrogenase 2

List of proteins identified in fraction IIa (part II):

ID:	Description
Solyc06g069030.2.1	PAP fibrillin domain containing protein expressed
Solyc07g007300.2.1	"Menaquinone biosynthesis methyltransferase ubiE
Solyc07g008350.2.1	"Porin/voltage-dependent anion-selective channel protein
Solyc07g045420.2.1	"ABC-1 domain protein
Solyc07g049690.2.1	Cytochrome P450
Solyc07g053830.2.1	"Mitochondrial ADP/ATP carrier proteins
Solyc07g063520.2.1	"Transmembrane protein 34
Solyc07g064270.2.1	"Glucose-6-phosphate/phosphate-translocator
Solyc08g006510.2.1	"NAD dependent epimerase/dehydratase family protein
Solyc08g008630.2.1	Homology to unknown gene
Solyc08g062610.2.1	"Cyclopropane-fatty-acyl-phospholipid synthase
Solyc08g068570.2.1	Tocopherol cyclase
Solyc08g068590.2.1	"PAP fibrillin family protein
Solyc08g074560.2.1	"Uncharacterized aarF domain-containing protein kinase 1
Solyc08g075100.2.1	"Alpha/beta superfamily hydrolase
Solyc08g076450.2.1	"3-beta hydroxysteroid dehydrogenase/isomerase family protein
Solyc08g076480.2.1	"Plastid lipid-associated protein 3, chloroplastic
Solyc08g077880.2.1	Light harvesting-like protein 3
Solyc08g080050.2.1	PGR5-like protein 1A, chloroplastic
Solyc08g083010.2.1	Homology to unknown gene
Solyc09g009820.2.1	"Glutathione S-transferase domain protein
Solyc09g011920.2.1	"Mitochondrial uncoupling protein
Solyc09g015650.2.1	"Non-green plastid inner envelope membrane protein
Solyc09g059040.2.1	"Alcohol dehydrogenase zinc-containing
Solyc09g061440.2.1	Os04g0513000 protein
Solyc09g066150.1.1	Cytochrome P450
Solyc09g090330.2.1	"Harpin binding protein 1
Solyc10g005100.2.1	"Salt stress root protein RS1
Solyc10g008980.2.1	"Triose phosphate/phosphate translocator
Solyc10g080900.1.1	"3-oxoacyl-reductase
Solyc10g081650.1.1	Carotenoid isomerase, chloroplastic
Solyc10g083350.1.1	"Soul heme-binding family protein
Solyc11g005620.1.1	"Acetylglutamate kinase-like protein
Solyc11g006970.1.1	Unknown protein DS12 from 2D-PAGE of leaf, chloroplastic
Solyc11g008990.1.1	"Phage shock protein A PspA
Solyc11g010190.1.1	"Prohibitin
Solyc11g013260.1.1	"Prohibitin
Solyc11g018550.2.1	Thylakoid-bound ascorbate peroxidase 6
Solyc11g062190.1.1	"Mitochondrial ADP/ATP carrier proteins
Solyc11g069430.1.1	"Aquaporin 1
Solyc11g069800.1.1	cytochrome P450
Solyc12g098710.1.1	15-cis-zeta-carotene isomerase

Annex II

A compacted list of all proteins by us detected in fractions III, IV, V and VI. Column "A" indicates whether a protein did or did not pass the filters we imposed for accepting a protein for data processing. A binary location code has been created to identify in which fraction each protein was found. each of the ciphers refers to one of the fractions. In that way, codes for unique and common proteins are: a) unique to III: 1000; b) unique to IV: 0100; c) unique to V 0010; d) proteins unique to VI, 0001; e) common to all fractions: 1111. Proteins neither unique to one fraction nor common to all are written in italics but the nomenclature system retains the same logic.

A	Solgenomics ID	Location code	Description
1	Solyc10g055670.1.1	1000	V-type ATP synthase beta chain
0	Solyc03g079900.2.1	1000	Ras-related protein Rab-25
1	Solyc01g110120.2.1	1000	V-type proton ATPase subunit a
1	Solyc10g081030.1.1	0100	Nascent polypeptide-associated complex alpha subunit-like protein
1	Solyc07g062650.2.1	0100	Malate dehydrogenase
1	Solyc07g008720.2.1	0100	Nascent polypeptide-associated complex subunit beta
0	Solyc10g078150.1.1	0100	Nascent polypeptide-associated complex alpha subunit-like protein
1	Solyc01g007740.2.1	0100	Peroxiredoxin
1	Solyc12g056830.1.1	0100	ATP synthase delta subunit
1	Solyc07g066610.2.1	0100	Phosphoglycerate kinase
0	Solyc07g005210.2.1	0100	Outer membrane lipoprotein blc
0	Solyc03g080160.2.1	0100	Nascent polypeptide-associated complex alpha subunit-like protein
1	Solyc03g019880.2.1	0100	UPF0426 protein At1g28150, chloroplastic
0	Solyc01g097870.2.1	0100	40S ribosomal protein S24
0	Solyc02g079750.2.1	0100	Flavoprotein wrbA
0	Solyc04g005340.2.1	0100	Alpha-1 4-glucan protein synthase
1	Solyc04g080850.2.1	0100	Thioredoxin
1	Solyc09g090140.2.1	0100	Malate dehydrogenase
1	Solyc01g105060.2.1	0100	(3R)-hydroxymyristoyl-
0	Solyc02g066930.2.1	0100	RNA-binding protein
1	Solyc08g078700.2.1	0100	Heat shock protein 22
1	Solyc12g042830.1.1	0100	class I heat shock protein
1	Solyc01g100360.2.1	0100	Dihydrolipoyl dehydrogenase
1	Solyc01g096580.2.1	0100	30S ribosomal protein S10
0	Solyc01g059930.2.1	0100	Universal stress protein
1	Solyc12g009250.1.1	0100	chaperonin
0	Solyc04g009410.2.1	0100	Proteasome subunit beta type
0	Solyc09g018750.2.1	0100	Inosine-5&apos-monophosphate dehydrogenase
1	Solyc06g009400.2.1	0100	Nitrogen regulatory protein P-II
0	Solyc07g006650.2.1	0100	Xylose isomerase
0	Solyc08g079020.2.1	0100	Adenine phosphoribosyltransferase-like protein
0	Solyc08g078510.2.1	0100	GRAM-containing/ABA-responsive protein
1	Solyc05g006520.2.1	0100	Pyruvate dehydrogenase E1 component alpha subunit
1	Solyc01g109660.2.1	0100	Glycine-rich RNA-binding protein
1	Solyc05g007100.2.1	0100	DNA-binding protein p24
1	Solyc08g077920.2.1	0100	Isocitrate dehydrogenase
0	Solyc09g082320.2.1	0100	Proteasome subunit beta type
1	Solyc12g044600.2.1	0100	NADP-dependent malic enzyme, chloroplastic
1	Solyc11g005620.1.1	0100	Acetylglutamate kinase-like protein
1	Solyc06g060260.2.1	0100	Stromal ascorbate peroxidase 7
1	Solyc00g006800.2.1	0100	Transaldolase
0	Solyc08g015860.2.1	0100	Rab GDP dissociation inhibitor
1	Solyc08g075750.2.1	0100	ATP-dependent Clp protease proteolytic subunit
1	Solyc10g047950.1.1	0100	Inorganic pyrophosphatase family protein
1	Solyc09g061440.2.1	0100	Os04g0513000 protein
1	Solyc12g056120.1.1	0100	6-phosphogluconate dehydrogenase decarboxylating
0	Solyc01g099630.2.1	0100	Xyloglucan endotransglucosylase/hydrolase 5

0	Solyc12g015880.1.1	0100	Heat shock protein 90
0	Solyc03g097110.2.1	0100	RNA recognition motif-containing protein
1	Solyc08g006720.2.1	0100	Glutathione peroxidase
1	Solyc01g006450.2.1	0100	Enoyl reductase
1	Solyc07g042250.2.1	0100	chaperonin
0	Solyc04g079200.2.1	0100	26S proteasome regulatory subunit
1	Solyc05g008600.2.1	0100	Fructose-bisphosphate aldolase
1	Solyc02g088610.2.1	0100	ATP-dependent chaperone ClpB
0	Solyc06g072580.2.1	0100	Pyruvate dehydrogenase E1 component subunit beta
1	Solyc08g081010.2.1	0100	Glutamate-cysteine ligase
0	Solyc01g107870.2.1	0100	Poly(A) RNA binding protein
1	Solyc07g055320.2.1	0100	ATP-dependent Zn protease cell division protein FtsH homolog
1	Solyc02g083810.2.1	0100	Ferredoxin--NADP reductase
0	Solyc12g042920.1.1	0100	Cytochrome c1
1	Solyc04g053120.2.1	0100	4-alpha-glucanotransferase
1	Solyc10g081240.1.1	0100	Protein grpE
0	Solyc06g083790.2.1	0100	Succinyl-CoA ligase
1	Solyc12g014050.1.1	0100	Zinc-containing alcohol dehydrogenase
1	Solyc02g080540.1.1	0100	ATP synthase gamma chain
1	Solyc02g085730.2.1	0100	Allene oxide cyclase
1	Solyc05g050970.2.1	0100	Transketolase 1
0	Solyc07g043310.2.1	0100	Aminotransferase
0	Solyc04g081740.2.1	0100	Prostaglandin E synthase 2-like
1	Solyc11g022460.1.1	0100	Phosphoribosylanthranilate transferase like protein
1	Solyc07g044860.2.1	0100	Oxygen-evolving enhancer protein 2, chloroplastic
1	Solyc07g066580.2.1	0100	Mercaptopyruvate sulfurtransferase-like protein
0	Solyc06g083730.2.1	0100	RNA binding protein
1	Solyc11g006970.1.1	0100	Unknown protein DS12 from 2D-PAGE of leaf, chloroplastic
0	Solyc08g074840.1.1	0100	Eukaryotic translation initiation factor 3 subunit 4
1	Solyc10g007180.2.1	0100	DAG protein
1	Solyc07g053280.2.1	0100	Ketol-acid reductoisomerase
1	Solyc04g082250.2.1	0100	ATP-dependent zinc metalloprotease FTSH, chloroplastic
0	Solyc09g011030.2.1	0100	Hsp70 nucleotide exchange factor fes1
1	Solyc02g078120.1.1	0100	Eukaryotic translation initiation factor 3 subunit 7
1	Solyc01g080460.2.1	0100	Pyruvate phosphate dikinase
0	Solyc04g082200.2.1	0100	Dehydrin
1	Solyc11g069800.1.1	0100	cytochrome P450
1	Solyc03g025530.2.1	0100	S-layer domain protein
1	Solyc09g031780.2.1	0100	Tic110 family transporter chloroplast inner envelope protein Tic110
1	Solyc02g091580.2.1	0100	Oligopeptidase A
0	Solyc10g005230.2.1	0010	Unknown Protein
1	Solyc01g066720.2.1	0010	Hypoxia induced protein conserved region containing protein
1	Solyc10g078690.1.1	0010	Mitochondrial protein translocase family
0	Solyc08g078070.2.1	0010	Ras-related protein Rab-1A
1	Solyc08g077300.2.1	0010	Adenylate kinase
0	Solyc06g082800.2.1	0010	Mitochondrial import receptor subunit TOM40
0	Solyc11g070030.1.1	0010	NADH-quinone oxidoreductase subunit B
0	Solyc06g067940.2.1	0010	Cytochrome c oxidase subunit Vb
1	Solyc12g057120.1.1	0010	Subunit VIb of cytochrome c oxidase
0	Solyc03g113020.2.1	0010	Vacuolar protein sorting 29
0	Solyc11g011040.1.1	0010	ADP-ribosylation factor-like protein
1	Solyc06g007670.2.1	0010	60S ribosomal protein L5-1
1	Solyc06g005160.2.1	0010	Ascorbate peroxidase
1	Solyc01g079420.2.1	0010	Cytochrome c oxidase subunit VC family protein
1	Solyc01g090770.2.1	0010	Unknown Protein
0	Solyc12g014080.1.1	0010	Ras-related protein Rab-25
1	Solyc07g020860.2.1	0010	Peroxiredoxin
0	Solyc01g110390.2.1	0010	NADH dehydrogenase

0	Solyc09g005720.2.1	0010	60S ribosomal protein L23a
1	Solyc01g098000.2.1	0010	Elongation factor-like protein
0	Solyc09g010370.2.1	0010	Ras-related protein Rab-8A
0	Solyc08g006040.2.1	0010	40S ribosomal protein S6
1	Solyc03g080050.2.1	0010	Band 7 stomatin family protein
0	Solyc01g091530.2.1	0010	Fasciclin-like arabinogalactan protein 13
0	Solyc09g015020.1.1	0010	class I heat shock protein 3
1	Solyc10g085870.1.1	0010	UDP-glucosyltransferase family 1 protein
1	Solyc03g095840.2.1	0010	Eukaryotic translation initiation factor 2 alpha subunit family protein expressed
1	Solyc06g035920.2.1	0010	Remorin 1
0	Solyc03g113220.2.1	0010	SPFH domain/band 7 family protein
0	Solyc07g055290.2.1	0010	Ras-related protein Rab-25
1	Solyc09g011920.2.1	0010	Mitochondrial uncoupling protein
0	Solyc01g111250.2.1	0010	Phosphatidylinositol-specific phospholipase c
0	Solyc04g054470.2.1	0010	Myosin class II heavy chain
1	Solyc06g019170.2.1	0010	Gamma-glutamyl phosphate reductase
0	Solyc09g064370.2.1	0010	Alcohol dehydrogenase
1	Solyc01g099900.2.1	0010	Ribosomal protein L18
1	Solyc04g079270.2.1	0010	Sorting and assembly machinery component 50 homolog
1	Solyc09g007250.2.1	0010	60S ribosomal protein L4/L1
1	Solyc10g006020.2.1	0010	Mitochondrial import inner membrane translocase subunit tim50
0	Solyc10g085230.1.1	0010	UDP-glucosyltransferase
1	Solyc10g081820.1.1	0010	Expp1 protein
0	Solyc12g005330.1.1	0010	50S ribosomal protein L2
1	Solyc07g032740.2.1	0010	Aspartate aminotransferase
0	Solyc10g085430.1.1	0010	Saposin
0	Solyc06g082630.2.1	0010	26S protease regulatory subunit 6B
0	Solyc05g053140.2.1	0010	26S proteasome non-ATPase regulatory subunit 13
1	Solyc06g083620.2.1	0010	26S protease regulatory subunit 4
0	Solyc06g050130.2.1	0010	Alpha-galactosidase-like protein
0	Solyc03g121270.2.1	0010	IAA-amino acid hydrolase
1	Solyc07g065010.2.1	0010	Bactericidal permeability-increasing protein
0	Solyc10g008630.2.1	0010	NADH dehydrogenase like protein
0	Solyc12g021170.1.1	0010	Electron-transfer flavoprotein ubiquinone oxidoreductase
1	Solyc10g084140.1.1	0010	Ferrochelatase
1	Solyc06g064940.2.1	0010	Phosphatidylinositol transfer protein SFH5
0	Solyc11g011020.1.1	0010	Receptor like kinase, RLK
1	Solyc01g087260.2.1	0010	Carotenoid cleavage dioxygenase 1B
0	Solyc07g055060.2.1	0010	Phosphoenolpyruvate carboxylase 1
1	Solyc11g011420.1.1	0010	Peptidyl-prolyl cis-trans isomerase
0	Solyc07g007030.1.1	0010	Transmembrane 9 superfamily protein member 2
1	Solyc08g063050.2.1	0010	Cell division protease ftsH homolog
0	Solyc07g064180.2.1	0010	Pectinesterase
1	Solyc03g118670.2.1	0010	Protein sel-1 homolog 2
0	Solyc07g006130.1.1	0010	Eukaryotic peptide chain release factor subunit 1-1
0	Solyc06g051620.2.1	0010	1-phosphatidylinositol-4 5-bisphosphate phosphodiesterase
1	Solyc11g008770.1.1	0010	LETM1 and EF-hand domain-containing protein 1, mitochondrial
0	Solyc04g014480.2.1	0001	class I heat shock protein 3
0	Solyc09g057650.2.1	0001	40S ribosomal protein S8
0	Solyc03g115490.1.1	0001	LYR motif-containing protein 4
0	Solyc01g104020.2.1	0001	Embryo-specific 3
0	Solyc12g056230.1.1	0001	Glutathione peroxidase
1	Solyc02g063490.2.1	0001	Malate dehydrogenase
1	Solyc09g075010.2.1	0001	Prostaglandin E synthase 3
0	Solyc07g049140.2.1	0001	Metallocarboxypeptidase inhibitor
1	Solyc11g006550.1.1	0001	Uricase
1	Solyc03g083970.2.1	0001	IQ calmodulin-binding motif family protein

1	Solyc02g082760.2.1	0001	Catalase
1	Solyc01g094790.2.1	0001	Cysteine synthase
0	Solyc02g089070.2.1	0001	Eukaryotic translation initiation factor 3 subunit M
1	Solyc11g008720.1.1	0001	Beta-glucosidase G4
1	Solyc01g111710.2.1	0001	26S proteasome non-ATPase regulatory subunit 3
0	Solyc04g078820.2.1	0001	Annexin
1	Solyc04g056390.2.1	0001	Isopentenyl-diphosphate delta-isomerase
0	Solyc07g007790.2.1	0001	Sucrose phosphate synthase
1	Solyc08g013860.2.1	0001	NAD-dependent malic enzyme 2
1	Solyc02g089720.1.1	0001	Endo-beta-1 3-glucanase
0	Solyc07g053650.2.1	0001	26S proteasome regulatory subunit
1	Solyc04g011350.2.1	0001	2-oxoglutarate dehydrogenase E1 component
0	Solyc01g103480.2.1	0001	Coatomer subunit delta
0	Solyc01g081070.2.1	0001	Protein transport protein Sec23
1	Solyc06g068090.2.1	0001	Phospholipase D
0	Solyc03g120270.2.1	0001	Coatomer beta&apos subunit
0	Solyc04g015830.2.1	0001	Actin-binding protein gelsolin
0	Solyc03g096920.2.1	0001	Exportin-1
1	Solyc12g055800.1.1	1111	V-type ATP synthase alpha chain
0	Solyc01g100860.2.1	1111	ADP-ribosylation factor
0	Solyc04g049330.2.1	1111	V-type proton ATPase subunit G 1
1	Solyc01g111760.2.1	1111	V-type ATP synthase beta chain
1	Solyc10g005100.2.1	1111	Salt stress root protein RS1
0	Solyc02g031950.2.1	1111	Pathogenesis-related protein-like protein
1	Solyc10g081530.1.1	1111	V-type proton ATPase subunit d 1
0	Solyc12g010320.1.1	1111	Outer membrane lipoprotein blc
1	Solyc11g039980.1.1	1111	ATP synthase subunit alpha
1	Solyc10g008140.2.1	1111	Prohibitin 1-like protein
1	Solyc03g097790.2.1	1111	V-type proton ATPase subunit C
1	Solyc08g067020.2.1	1111	Ubiquinol-cytochrome c reductase complex protein
1	Solyc11g005330.1.1	1111	Actin
1	Solyc11g072450.1.1	1111	Mitochondrial F0 ATP synthase D chain
1	Solyc04g007550.2.1	1111	ATP synthase subunit beta
1	Solyc07g066600.2.1	1111	Phosphoglycerate kinase
1	Solyc06g059740.2.1	1111	Alcohol dehydrogenase 2
1	Solyc06g073900.2.1	1111	Unknown Protein
1	Solyc11g010190.1.1	1111	Prohibitin
1	Solyc11g011380.1.1	1111	Glutamine synthetase
1	Solyc06g069010.2.1	1111	Guanine nucleotide-binding protein beta subunit-like protein
1	Solyc09g059040.2.1	1111	Alcohol dehydrogenase zinc-containing
1	Solyc01g007320.2.1	1111	ATP synthase subunit beta chloroplastic
1	Solyc11g069790.1.1	1111	chaperonin
1	Solyc03g025850.2.1	1111	Remorin 1
0	Solyc08g081910.2.1	1111	V-type proton ATPase subunit E
0	Solyc04g014600.2.1	1111	Universal stress protein family protein
1	Solyc02g067460.2.1	1111	Porin/voltage-dependent anion-selective channel protein
0	Solyc09g084450.2.1	1111	Chymotrypsin inhibitor-2
0	Solyc08g015690.2.1	1111	Late-embryogenesis abundant protein 2
1	Solyc01g099190.2.1	1111	Lipoxygenase
1	Solyc04g054980.2.1	1111	Lipoxygenase homology domain-containing protein 1
0	Solyc10g006650.2.1	1111	Flavoprotein wrbA
1	Solyc01g087120.2.1	1111	F1-ATP synthase delta subunit
1	Solyc11g062190.1.1	1111	Mitochondrial ADP/ATP carrier proteins
1	Solyc01g111170.2.1	1111	Peptidyl-prolyl cis-trans isomerase
1	Solyc02g063070.2.1	1111	14-3-3 protein beta/alpha-1
1	Solyc01g109520.2.1	1111	Ras-related protein Rab-7
1	Solyc12g008630.1.1	1111	Mitochondrial processing peptidase alpha subunit
1	Solyc06g005060.2.1	1111	Elongation factor 1-alpha

0	Solyc08g008210.2.1	1111	V-type proton ATPase subunit E
1	Solyc04g007770.2.1	1111	Major latex-like protein
1	Solyc04g054990.2.1	1111	Lipoxygenase homology domain-containing protein 1
0	Solyc09g089580.2.1	1111	1-aminocyclopropane-1-carboxylate oxidase-like protein
0	Solyc08g080630.2.1	1111	Chymotrypsin inhibitor-2
1	Solyc12g013690.1.1	1111	Monooxygenase FAD-binding protein
1	Solyc01g102660.2.1	1111	Maleylacetoacetate isomerase / glutathione S-transferase
1	Solyc02g069100.2.1	1111	Cathepsin B
0	Solyc01g104680.2.1	1111	GTP-binding nuclear protein Ran-A1
1	Solyc01g010750.2.1	1111	Stress responsive protein
1	Solyc11g056680.1.1	1111	LRR receptor-like serine/threonine-protein kinase, RLP
1	Solyc03g025950.2.1	1111	Membrane-associated progesterone receptor component 1
1	Solyc06g008260.2.1	1111	60 ribosomal protein L14
0	Solyc07g051850.2.1	1111	Aspartic proteinase
1	Solyc06g005360.2.1	1111	Actin depolymerizing factor 3
1	Solyc01g010760.2.1	1111	Porin/voltage-dependent anion-selective channel protein
0	Solyc01g057000.2.1	1111	Universal stress protein family protein
0	Solyc01g073690.2.1	1111	V-type ATP synthase subunit D
1	Solyc03g058920.2.1	1111	Porin/voltage-dependent anion-selective channel protein
0	Solyc08g078250.2.1	1111	Palmitoyl-protein thioesterase 1
0	Solyc09g066430.2.1	1111	60S ribosomal protein L14
1	Solyc07g053830.2.1	1111	Mitochondrial ADP/ATP carrier proteins
0	Solyc06g011280.2.1	1111	Elongation factor 1-gamma
1	Solyc06g072120.2.1	1111	40S ribosomal protein SA
0	Solyc04g063290.2.1	1111	30S ribosomal protein S5
1	Solyc01g103450.2.1	1111	Chaperone DnaK
1	Solyc02g094470.2.1	1111	Mitochondrial phosphate carrier protein
1	Solyc03g093360.2.1	1111	Wound/stress protein
1	Solyc04g081570.2.1	1111	Chaperone protein htpG
1	Solyc03g118040.2.1	1111	Calcium-binding protein Calnexin
1	Solyc11g020040.1.1	1111	Chaperone DnaK
0	Solyc04g080960.2.1	1111	Cysteine proteinase cathepsin F
0	Solyc03g083910.2.1	1111	Acid beta-fructofuranosidase
1	Solyc11g065490.1.1	1111	Homology to unknown gene
0	Solyc02g068450.2.1	1110	Vacuolar ATPase F subunit
0	Solyc03g097800.2.1	1110	V-type proton ATPase subunit C
0	Solyc03g082720.2.1	1110	Yippee zinc-binding-like protein
0	Solyc10g085040.1.1	1110	Soul heme-binding family protein
1	Solyc07g005940.2.1	1110	Vacuolar ATPase subunit H protein
0	Solyc06g076520.1.1	1101	class I heat shock protein
0	Solyc02g021400.1.1	1101	40S ribosomal protein S28
0	Solyc12g005080.1.1	1101	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex
1	Solyc12g098940.1.1	1101	Ubiquitin
1	Solyc05g008460.2.1	1011	ATP synthase subunit beta
1	Solyc01g079540.2.1	1011	ADP-ribosylation factor-like protein 3
1	Solyc11g068400.1.1	1011	Ubiquinol-cytochrome C reductase
0	Solyc03g116590.2.1	1011	Embryo-specific 3
1	Solyc11g068510.1.1	1011	F1F0-ATPase inhibitor protein
1	Solyc06g073310.2.1	1011	Ribosomal L9-like protein
1	Solyc11g013260.1.1	1011	Prohibitin
1	Solyc08g066160.2.1	1011	Enod93 protein
0	Solyc07g007750.2.1	1011	Defensin protein
1	Solyc05g012480.2.1	1011	Mitochondrial processing peptidase beta subunit
1	Solyc06g075810.2.1	1011	NADH dehydrogenase
0	Solyc01g099830.2.1	1011	60S ribosomal protein L22-2
0	Solyc03g043850.2.1	1011	Cytochrome c oxidase subunit VIa family protein expressed
0	Solyc07g007760.2.1	1011	Defensin protein

1	Solyc10g078620.1.1	1011	Ribosomal protein S5
1	Solyc02g088700.2.1	1011	Mitochondrial processing peptidase beta subunit
0	Solyc01g103380.2.1	1011	Ras-related protein Rab-25
1	Solyc02g092440.2.1	1011	Mitochondrial porin
0	Solyc04g080880.2.1	1011	Cathepsin B-like cysteine proteinase
0	Solyc10g080210.1.1	1011	Polygalacturonase A
1	Solyc01g110540.2.1	0111	Unknown Protein
1	Solyc11g072190.1.1	0111	Elongation factor beta-1
1	Solyc02g081170.2.1	0111	Plastid-lipid-associated protein, chloroplastic
0	Solyc11g067100.1.1	0111	60s acidic ribosomal protein-like protein
1	Solyc10g080900.1.1	0111	3-oxoacyl-reductase
1	Solyc06g071920.2.1	0111	Glyceraldehyde-3-phosphate dehydrogenase
1	Solyc06g009020.2.1	0111	Glutathione S-transferase
1	Solyc09g064450.2.1	0111	NADH dehydrogenase
1	Solyc04g076060.2.1	0111	14-3-3 protein beta/alpha-B
1	Solyc02g086240.2.1	0111	50S ribosomal protein L5
1	Solyc00g009020.2.1	0111	Mitochondrial ATP synthase
0	Solyc11g012080.1.1	0111	Brain protein 44
1	Solyc04g011390.1.1	0111	Histone H4
0	Solyc01g100370.2.1	0111	Universal stress protein
0	Solyc07g016150.2.1	0111	Elongation factor 1-beta
1	Solyc02g086880.2.1	0111	Formate dehydrogenase
1	Solyc05g005490.2.1	0111	Carbonic anhydrase
0	Solyc09g090980.2.1	0111	Major allergen Mal d 1
1	Solyc05g052470.2.1	0111	Ferritin
1	Solyc05g014470.2.1	0111	Glyceraldehyde 3-phosphate dehydrogenase
1	Solyc06g050980.2.1	0111	Ferritin
1	Solyc03g115110.2.1	0111	ATP synthase gamma chain
1	Solyc06g007710.2.1	0111	Mitochondrial import receptor subunit TOM40
1	Solyc12g094620.1.1	0111	Catalase
1	Solyc04g011440.2.1	0111	heat shock protein
1	Solyc02g062500.2.1	0111	2-oxoglutarate-dependent dioxygenase
0	Solyc11g069150.1.1	0111	Proteasome subunit beta type
1	Solyc09g073000.2.1	0111	Elongation factor Tu
1	Solyc07g065840.2.1	0111	Heat shock protein 90
0	Solyc01g099670.2.1	0111	40S ribosomal protein S10-like
0	Solyc11g066060.1.1	0111	heat shock protein
0	Solyc00g019950.1.1	0111	NADH-quinone oxidoreductase subunit C
1	Solyc07g047800.2.1	0111	Short-chain dehydrogenase/reductase family protein
0	Solyc09g066150.1.1	0111	Cytochrome P450
0	Solyc06g069090.2.1	0111	40S ribosomal protein S7-like protein
0	Solyc06g084230.2.1	0111	40S ribosomal protein S24
1	Solyc03g111200.2.1	0111	Adenylate kinase
1	Solyc07g053260.2.1	0111	14-3-3 protein sigma gamma zeta beta/alpha
0	Solyc06g076970.2.1	0111	Peptidyl-prolyl cis-trans isomerase
1	Solyc06g049080.2.1	0111	Superoxide dismutase
1	Solyc08g006510.2.1	0111	NAD dependent epimerase/dehydratase family protein
0	Solyc08g066240.2.1	0111	Decarboxylase family protein
0	Solyc10g006070.2.1	0111	40S ribosomal protein S8
1	Solyc03g082600.2.1	0111	Cytochrome b5
0	Solyc10g085880.1.1	0111	UDP-glucosyltransferase family 1 protein
1	Solyc01g028810.2.1	0111	chaperonin
0	Solyc04g071610.2.1	0111	Water-stress inducible protein 3
0	Solyc06g073370.2.1	0111	40S ribosomal protein S18
0	Solyc07g062970.2.1	0111	Serine/threonine phosphatase family protein
0	Solyc09g007270.2.1	0111	Ascorbate peroxidase
1	Solyc04g073990.2.1	0111	Annexin
1	Solyc06g036290.2.1	0111	Heat shock protein 90

0	Solyc12g096000.1.1	0111	ATP-dependent RNA helicase eIF4A
0	Solyc06g073060.2.1	0111	Iaa-amino acid hydrolase 6
0	Solyc01g101240.2.1	0111	Aspartic proteinase
1	Solyc05g007060.2.1	0111	Bifunctional N-succinyldiaminopimelate-aminotransferase/acetylornithine transaminase protein
0	Solyc07g064800.2.1	0111	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex
1	Solyc01g108560.2.1	0111	Acetyl esterase
0	Solyc01g010540.2.1	0111	Ribosomal protein
1	Solyc11g051160.1.1	0111	Phosphatidylinositol transfer protein SFH5
1	Solyc09g092380.2.1	0111	Adenosylhomocysteinase
1	Solyc12g005860.1.1	0111	3-isopropylmalate dehydratase large subunit
0	Solyc03g119080.2.1	0111	Beta-glucosidase
1	Solyc09g009390.2.1	0111	Monodehydroascorbate reductase
0	Solyc11g022590.1.1	0111	Kunitz trypsin inhibitor 4
1	Solyc01g100320.2.1	0111	Thioredoxin/protein disulfide isomerase
0	Solyc02g093680.2.1	0111	Succinate dehydrogenase iron-sulfur protein
1	Solyc05g050120.2.1	0111	Malic enzyme
1	Solyc12g042060.1.1	0111	ATP-dependent clp protease ATP-binding subunit
1	Solyc06g035970.2.1	0111	Tubulin beta chain
1	Solyc05g053470.2.1	0111	chaperonin
1	Solyc08g014130.2.1	0111	2-isopropylmalate synthase 1
0	Solyc02g070320.2.1	0111	40S ribosomal protein S4-like protein
1	Solyc09g091470.2.1	0111	3-ketoacyl CoA thiolase 2
1	Solyc07g044840.2.1	0111	2 3-bisphosphoglycerate-independent phosphoglycerate mutase
1	Solyc01g009990.2.1	0111	Peptidyl-prolyl cis-trans isomerase
1	Solyc05g009530.2.1	0111	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex
0	Solyc01g008370.2.1	0111	26S proteasome regulatory subunit
0	Solyc01g109410.2.1	0111	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit
1	Solyc11g065620.1.1	0111	Sulfite reductase
1	Solyc04g045340.2.1	0111	Phosphoglucomutase
1	Solyc03g082920.2.1	0111	Heat shock protein
1	Solyc08g061960.2.1	0111	Ribosomal protein
1	Solyc09g082650.2.1	0111	Acireductone dioxygenase
1	Solyc03g115990.1.1	0111	Malate dehydrogenase
1	Solyc03g115230.2.1	0111	ClpB chaperone
0	Solyc05g054350.2.1	0111	Epoxide hydrolase
0	Solyc01g110450.2.1	0111	NADP dependent sorbitol 6-phosphate dehydrogenase
1	Solyc01g006540.2.1	0111	Lipoxygenase
1	Solyc06g005940.2.1	0111	Protein disulfide isomerase
1	Solyc03g097900.2.1	0111	40S ribosomal protein S1
1	Solyc04g077020.2.1	0111	Tubulin alpha-3 chain
0	Solyc10g083720.1.1	0111	Pyruvate kinase
1	Solyc01g109940.2.1	0111	26S protease regulatory subunit
0	Solyc12g015860.1.1	0111	Farnesyl pyrophosphate synthase
1	Solyc01g005560.2.1	0111	Isocitrate dehydrogenase
1	Solyc01g106210.2.1	0111	Chaperone DnaK
1	Solyc01g007330.2.1	0111	Ribulose bisphosphate carboxylase large chain
1	Solyc11g013440.1.1	0111	Eukaryotic translation initiation factor 3 subunit 5
1	Solyc03g112150.1.1	0111	Elongation factor Tu
1	Solyc01g105340.2.1	0111	Chaperone protein dnaJ
1	Solyc06g074000.1.1	0111	Aspartic proteinase nepenthesin-2
1	Solyc06g076360.2.1	0111	Outer envelope membrane protein
1	Solyc04g076790.2.1	0111	Serine hydroxymethyltransferase
1	Solyc03g079940.2.1	0111	Mitochondrial import inner membrane translocase
1	Solyc05g054580.2.1	0111	60S acidic ribosomal protein P0
0	Solyc07g009140.2.1	0111	26S proteasome non-ATPase regulatory subunit 6

1	Solyc05g013030.1.1	0111	26S proteasome regulatory subunit
1	Solyc04g008740.2.1	0111	Pyruvate kinase
0	Solyc05g054640.2.1	0111	2-oxoglutarate dehydrogenase E1 component
0	Solyc07g062530.2.1	0111	Phosphoenolpyruvate carboxylase 2
1	Solyc03g052980.2.1	0111	ATP dependent RNA helicase
0	Solyc04g080980.2.1	0111	Coatomer alpha subunit-like protein
0	Solyc05g052510.2.1	0111	Clathrin heavy chain
1	Solyc09g090330.2.1	1100	Harpin binding protein 1
1	Solyc11g008990.1.1	1100	Phage shock protein A PspA
0	Solyc12g006830.1.1	1010	Histone H2A
0	Solyc02g038690.1.1	1010	Histone H2B
1	Solyc03g044010.2.1	1010	Mitochondrial porin
0	Solyc05g053940.2.1	1010	Ras-related protein Rab-1A
0	Solyc07g006380.2.1	1010	Defensin-like protein
1	Solyc07g008350.2.1	1010	Porin/voltage-dependent anion-selective channel protein
0	Solyc03g025340.1.1	1010	C2 domain-containing protein
1	Solyc12g096700.1.1	1001	Ribosomal L9-like protein
0	Solyc01g100350.2.1	0110	ADP-ribosylation factor-like protein 3
1	Solyc06g072540.1.1	0110	ATP synthase subunit alpha chloroplastic
0	Solyc06g073460.2.1	0110	Glutathione peroxidase
1	Solyc10g082030.1.1	0110	Peroxiredoxin
1	Solyc09g082720.2.1	0110	Aldo/keto reductase family protein
1	Solyc02g093530.2.1	0110	Ras-related protein Rab-2-A
1	Solyc08g062450.1.1	0110	class II heat shock protein
0	Solyc06g007570.2.1	0110	40S ribosomal protein S8
0	Solyc06g009050.2.1	0110	Universal stress protein
1	Solyc02g070310.2.1	0110	Ribosomal protein L32
1	Solyc02g094440.2.1	0110	Translocon-associated protein subunit beta
1	Solyc01g100380.2.1	0110	Calreticulin 2 calcium-binding protein
0	Solyc05g015390.2.1	0110	REF-like stress related protein 1
1	Solyc01g099770.2.1	0110	Translationally-controlled tumor protein homolog
0	Solyc03g117470.2.1	0110	Calcineurin subunit B
1	Solyc01g111450.2.1	0110	Proteasome subunit alpha type
1	Solyc03g116110.2.1	0110	Alpha/beta hydrolase fold protein
1	Solyc04g079180.2.1	0110	Unknown Protein
1	Solyc05g053810.2.1	0110	Serine hydroxymethyltransferase
1	Solyc05g018570.2.1	0110	26S protease regulatory subunit 8 homolog A
1	Solyc09g009020.2.1	0110	Enolase
0	Solyc01g112280.2.1	0110	Succinyl-diaminopimelate desuccinylase
1	Solyc07g016200.2.1	0110	Proteasome subunit beta type
1	Solyc04g008520.2.1	0110	Outer envelope protein
1	Solyc05g056270.2.1	0110	Isocitrate lyase
1	Solyc07g043320.2.1	0110	Unknown Protein
1	Solyc02g093830.2.1	0110	Glucose-6-phosphate 1-dehydrogenase
0	Solyc07g065120.2.1	0110	Glycerophosphoryl diester phosphodiesterase family protein
1	Solyc08g081250.2.1	0110	Aminopeptidase N
1	Solyc05g056400.2.1	0110	Protein disulfide isomerase
0	Solyc09g091180.2.1	0101	chaperonin
1	Solyc01g089970.2.1	0101	Nucleoside diphosphate kinase
0	Solyc03g007890.2.1	0101	Heat shock protein 90
0	Solyc06g008120.1.1	0101	Mitochondrial import receptor subunit TOM22
1	Solyc06g048410.2.1	0101	Superoxide dismutase
1	Solyc02g067750.2.1	0101	Carbonic anhydrase
1	Solyc08g041870.2.1	0101	Aspartate aminotransferase
0	Solyc04g054310.2.1	0101	Alanine-glyoxylate aminotransferase
1	Solyc04g082460.2.1	0101	Catalase
1	Solyc03g117430.2.1	0101	Cobalamin synthesis protein P
1	Solyc01g100520.2.1	0101	ATP-dependent Clp protease proteolytic subunit

1	Solyc04g011360.2.1	0101	Ras-related protein Rab-8A
0	Solyc12g009140.1.1	0101	Proteasome subunit alpha type
0	Solyc07g055210.2.1	0101	Aspartate aminotransferase
0	Solyc08g079260.2.1	0101	Serine/threonine-protein phosphatase
0	Solyc02g085350.2.1	0101	Succinate dehydrogenase flavoprotein subunit
0	Solyc09g010100.2.1	0011	30S ribosomal protein S11
0	Solyc06g007540.2.1	0011	Cytochrome b-c1 complex subunit 8
0	Solyc07g065280.2.1	0011	Os08g0431500 protein
1	Solyc03g119360.2.1	0011	40S ribosomal protein S7-like protein
1	Solyc07g062570.2.1	0011	Ubiquitin-conjugating enzyme E2 N
1	Solyc02g093900.2.1	0011	Cytochrome c1
0	Solyc04g012120.2.1	0011	14-3-3 protein beta/alpha-1
1	Solyc07g008370.2.1	0011	60S ribosomal protein L7
1	Solyc02g062460.2.1	0011	2-oxoglutarate-dependent dioxygenase
0	Solyc06g075180.1.1	0011	Ribosomal protein L12
1	Solyc01g022750.2.1	0011	NADH dehydrogenase
0	Solyc04g074230.2.1	0011	14-3-3 protein sigma gamma zeta beta/alpha
1	Solyc03g019720.2.1	0011	Carnitine operon protein caiE
1	Solyc04g080570.2.1	0011	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase
0	Solyc10g081440.1.1	0011	NADH cytochrome b5 reductase
1	Solyc01g088510.2.1	0011	Dynamin like protein
1	Solyc11g011470.1.1	0011	NADH-ubiquinone oxidoreductase subunit
0	Solyc02g062600.2.1	0011	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase
1	Solyc06g036350.2.1	0011	NADH dehydrogenase
1	Solyc09g065830.2.1	0011	NADH ubiquinone oxidoreductase subunit
0	Solyc02g087240.2.1	0011	NADH-quinone oxidoreductase F subunit family protein
0	Solyc00g060810.2.1	0011	Major latex-like protein
0	Solyc02g085040.2.1	0011	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase
1	Solyc04g026100.1.1	0011	30S ribosomal protein S9
1	Solyc01g088560.2.1	0011	Ras-related protein Rab-6A
0	Solyc01g094690.2.1	0011	Aquaporin
1	Solyc05g050600.2.1	0011	Dynamin 2
0	Solyc00g014830.2.1	0011	NADH-quinone oxidoreductase subunit D
1	Solyc01g005620.2.1	0011	Mitochondrial 2-oxoglutarate/malate carrier protein
1	Solyc01g109620.2.1	0011	NADH-quinone oxidoreductase subunit I
0	Solyc12g039120.1.1	0011	40S ribosomal protein S19-like
0	Solyc03g111690.2.1	0011	Pectate lyase
1	Solyc11g039840.1.1	0011	Ubiquinol-cytochrome c reductase iron-sulfur subunit
1	Solyc01g099760.2.1	0011	26S protease regulatory subunit 6A homolog
0	Solyc01g108580.2.1	0011	Gibberellin receptor GID1L2
0	Solyc11g039650.1.1	0011	Dynamin-2A
1	Solyc01g099100.2.1	0011	Long-chain-fatty-acid coa ligase
0	Solyc11g005150.1.1	0011	Leucine-rich repeat family protein
0	Solyc04g008460.2.1	0011	Ribosomal protein L15
1	Solyc07g009320.2.1	0011	Metaxin 2
0	Solyc01g104590.2.1	0011	Ribosomal protein L3
0	Solyc03g025270.2.1	0011	rRNA 2&apos-O-methyltransferase fibrillarin
0	Solyc02g070640.2.1	0011	60S ribosomal protein L18a
1	Solyc01g087540.2.1	0011	Os01g0611000 protein
1	Solyc06g074980.2.1	0011	26S protease regulatory subunit 6B homolog
1	Solyc12g043110.1.1	0011	Heat shock protein 4
1	Solyc08g062920.2.1	0011	Elongation factor EF-2
0	Solyc09g098150.2.1	0011	Metacaspase 7
0	Solyc02g070020.1.1	0011	UDP-glucosyltransferase
0	Solyc12g099440.1.1	0011	Fatty acid oxidation complex subunit alpha
1	Solyc01g106050.2.1	0011	Dynamin-2
0	Solyc08g081320.2.1	0011	AP-1 complex subunit beta-1

Annex III

Percentual amino acid composition of several plant fibrillins. As a reference, the an average calculated tomato protein composition is indicated at the right. Over-represented and under-represented amino acids are highlighted in blue and in red, respectively.

FBN1 Sol.ly.		FBN1a Arabidopsis		FBN1b Arabidopsis		FBN1 Oryza sativa		Avg. Tom. protein %
Ala (A) 17	6.1%	Ala (A) 13	4.9%	Ala (A) 16	6.4%	Ala (A) 27	10.2%	6,2
Arg (R) 8	2.9%	Arg (R) 13	4.9%	Arg (R) 12	4.8%	Arg (R) 10	3.8%	5,1
Asn (N) 12	4.3%	Asn (N) 9	3.4%	Asn (N) 10	4.0%	Asn (N) 8	3.0%	4,8
Asp (D) 11	3.9%	Asp (D) 18	6.8%	Asp (D) 15	6.0%	Asp (D) 12	4.5%	5,3
Cys (C) 0	0.0%	Cys (C) 0	0.0%	Cys (C) 0	0.0%	Cys (C) 1	0.4%	1,9
Gln (Q) 11	3.9%	Gln (Q) 11	4.2%	Gln (Q) 10	4.0%	Gln (Q) 10	3.8%	3,7
Glu (E) 25	8.9%	Glu (E) 18	6.8%	Glu (E) 18	7.2%	Glu (E) 24	9.1%	6,4
Gly (G) 16	5.7%	Gly (G) 15	5.7%	Gly (G) 17	6.8%	Gly (G) 18	6.8%	6,3
His (H) 0	0.0%	His (H) 0	0.0%	His (H) 0	0.0%	His (H) 0	0.0%	2,4
Ile (I) 20	7.1%	Ile (I) 18	6.8%	Ile (I) 21	8.4%	Ile (I) 17	6.4%	5,8
Leu (L) 29	10.4%	Leu (L) 28	10.6%	Leu (L) 30	12.0%	Leu (L) 29	10.9%	9,5
Lys (K) 19	6.8%	Lys (K) 13	4.9%	Lys (K) 11	4.4%	Lys (K) 14	5.3%	6,4
Met (M) 0	0.0%	Met (M) 0	0.0%	Met (M) 0	0.0%	Met (M) 0	0.0%	2,5
Phe (F) 13	4.6%	Phe (F) 11	4.2%	Phe (F) 9	3.6%	Phe (F) 11	4.2%	4,3
Pro (P) 21	7.5%	Pro (P) 14	5.3%	Pro (P) 15	6.0%	Pro (P) 16	6.0%	4,7
Ser (S) 31	11.1%	Ser (S) 35	13.3%	Ser (S) 26	10.4%	Ser (S) 24	9.1%	8,8
Thr (T) 20	7.1%	Thr (T) 21	8.0%	Thr (T) 20	8.0%	Thr (T) 21	7.9%	5,0
Trp (W) 3	1.1%	Trp (W) 3	1.1%	Trp (W) 2	0.8%	Trp (W) 3	1.1%	1,3
Tyr (Y) 5	1.8%	Tyr (Y) 4	1.5%	Tyr (Y) 4	1.6%	Tyr (Y) 4	1.5%	3,0
Val (V) 19	6.8%	Val (V) 19	7.2%	Val (V) 15	6.0%	Val (V) 16	6.0%	6,6
Pyl (O) 0	0.0%	Pyl (O) 0	0.0%	Pyl (O) 0	0.0%	Pyl (O) 0	0.0%	0,0
Sec (U) 0	0.0%	Sec (U) 0	0.0%	Sec (U) 0	0.0%	Sec (U) 0	0.0%	0,0

FBN2 Sol.ly.		FBN2 Arabidopsis		FBN2 Oryza sativa		Avg. Tom. protein %
Ala (A) 17	5.7%	Ala (A) 21	6.5%	Ala (A) 26	7.9%	6,2
Arg (R) 9	3.0%	Arg (R) 11	3.4%	Arg (R) 13	4.0%	5,1
Asn (N) 13	4.4%	Asn (N) 15	4.6%	Asn (N) 10	3.0%	4,8
Asp (D) 19	6.4%	Asp (D) 21	6.5%	Asp (D) 25	7.6%	5,3
Cys (C) 2	0.7%	Cys (C) 1	0.3%	Cys (C) 1	0.3%	1,9
Gln (Q) 7	2.4%	Gln (Q) 8	2.5%	Gln (Q) 15	4.6%	3,7
Glu (E) 29	9.8%	Glu (E) 32	9.9%	Glu (E) 26	7.9%	6,4
Gly (G) 21	7.1%	Gly (G) 26	8.0%	Gly (G) 23	7.0%	6,3
His (H) 0	0.0%	His (H) 0	0.0%	His (H) 1	0.3%	2,4
Ile (I) 15	5.1%	Ile (I) 12	3.7%	Ile (I) 17	5.2%	5,8
Leu (L) 31	10.4%	Leu (L) 41	12.7%	Leu (L) 29	8.8%	9,5
Lys (K) 17	5.7%	Lys (K) 12	3.7%	Lys (K) 8	2.4%	6,4
Met (M) 2	0.7%	Met (M) 1	0.3%	Met (M) 1	0.3%	2,5
Phe (F) 12	4.0%	Phe (F) 11	3.4%	Phe (F) 11	3.4%	4,3
Pro (P) 26	8.8%	Pro (P) 24	7.4%	Pro (P) 41	12.5%	4,7
Ser (S) 30	10.1%	Ser (S) 42	13.0%	Ser (S) 35	10.7%	8,8
Thr (T) 19	6.4%	Thr (T) 17	5.3%	Thr (T) 15	4.6%	5,0
Trp (W) 4	1.3%	Trp (W) 6	1.9%	Trp (W) 6	1.8%	1,3
Tyr (Y) 4	1.3%	Tyr (Y) 3	0.9%	Tyr (Y) 4	1.2%	3,0
Val (V) 20	6.7%	Val (V) 19	5.9%	Val (V) 21	6.4%	6,6
Pyl (O) 0	0.0%	Pyl (O) 0	0.0%	Pyl (O) 0	0.0%	0,0
Sec (U) 0	0.0%	Sec (U) 0	0.0%	Sec (U) 0	0.0%	0,0

FBN4 Sol.ly.		FBN4 Arabidopsis		FBN4 Oryza sativa		Avg. Tom. protein %
Ala (A) 17	7.7%	Ala (A) 13	6.1%	Ala (A) 26	11.5%	6,2
Arg (R) 12	5.4%	Arg (R) 13	6.1%	Arg (R) 16	7.1%	5,1
Asn (N) 4	1.8%	Asn (N) 4	1.9%	Asn (N) 5	2.2%	4,8
Asp (D) 14	6.3%	Asp (D) 17	8.0%	Asp (D) 19	8.4%	5,3
Cys (C) 1	0.5%	Cys (C) 1	0.5%	Cys (C) 0	0.0%	1,9
Gln (Q) 5	2.3%	Gln (Q) 5	2.4%	Gln (Q) 5	2.2%	3,7
Glu (E) 18	8.1%	Glu (E) 11	5.2%	Glu (E) 11	4.9%	6,4
Gly (G) 19	8.6%	Gly (G) 20	9.4%	Gly (G) 22	9.7%	6,3
His (H) 1	0.5%	His (H) 1	0.5%	His (H) 1	0.4%	2,4
Ile (I) 14	6.3%	Ile (I) 10	4.7%	Ile (I) 10	4.4%	5,8
Leu (L) 26	11.7%	Leu (L) 26	12.3%	Leu (L) 24	10.6%	9,5
Lys (K) 13	5.9%	Lys (K) 12	5.7%	Lys (K) 9	4.0%	6,4
Met (M) 0	0.0%	Met (M) 1	0.5%	Met (M) 0	0.0%	2,5
Phe (F) 10	4.5%	Phe (F) 11	5.2%	Phe (F) 7	3.1%	4,3
Pro (P) 16	7.2%	Pro (P) 14	6.6%	Pro (P) 16	7.1%	4,7
Ser (S) 20	9.0%	Ser (S) 19	9.0%	Ser (S) 16	7.1%	8,8
Thr (T) 15	6.8%	Thr (T) 15	7.1%	Thr (T) 16	7.1%	5,0
Trp (W) 2	0.9%	Trp (W) 2	0.9%	Trp (W) 2	0.9%	1,3
Tyr (Y) 2	0.9%	Tyr (Y) 2	0.9%	Tyr (Y) 3	1.3%	3,0
Val (V) 13	5.9%	Val (V) 15	7.1%	Val (V) 18	8.0%	6,6
Pyl (O) 0	0.0%	Pyl (O) 0	0.0%	Pyl (O) 0	0.0%	0,0
Sec (U) 0	0.0%	Sec (U) 0	0.0%	Sec (U) 0	0.0%	0,0

Annex IV

Table of non-canonical modular fibrillins.

	Species	Additional domain(s)
FBN1	Nanochoropsis gaditana	Phenylpropionate dioxygenase
FBN3	Malus domestica	Additional FBN3 domain
	Vitis vinifera	tRNA synthetase, partial
		Plant transposase
		PB1 domain
	Ostreococcus tauri	2x forkhead-associated domains (FHA)
	Micromonas pusilla	2x forkhead-associated domains (FHA)
	Medicago truncatula	Retrotransposon-related UBN2_3
FBN5	Triticum urartu	Vps51/Vps67
	Malus domestica	Unknown
	Musa acuminata	Isy1 domain
	Solanum phureja	Unknown
FBN6	Volvox carteri	Unknown
FBN7	Thalassiosira pseudonana	Phospholipid-methylating domain
	Eucaliptus grandis	DUF1240 (domain of unknown function)
FBN9	Sellaginella moellendorfii	Unknown
	Thellungiella halophila	Chalcone and stilbene synthase
FBN10	Fragaria vesca	very long chain acyl-CoA dehydrogenase
		2x C2 repeats
		Antagonist of mitotic exit network;Amn1
		Phosphoribosyltransferase domain
	Nanochloropsis gaditana	Zn-dependent amino peptidase
FBN11	Triticum urartu	DUF (domain of unknown function)
		Plant knuckle domain
FBN12	Brassica rapa	Elongation factor 1-alpha
	Capsella rubella	Unknown
	Carica papaya	Serine/threonine kinase
	Cicer arietinum	Serine/threonine kinase
	Bathycoccus prasinos	Thyoredoxin/hydroperoxide reductase
	Ostreococcus tauri	Thyoredoxin/hydroperoxide reductase
	Aureococcus phagenes	Unknown

A hard-to-classify example of another of these 26 modular fibrillins was an Ostreococcus tauri protein which had no clear fibrillin equivalent. Being distantly related to both FBN1, and FBN3, and distant relatives annotated as "pili assembly chaperones" it was left out of this list. This twenty-sixth modular protein was found to be annotated as a putative serine palmitoyltransferase, and contained, in addition to the undefined fibrillin domain, a serine palmitoyltransferase domain, an aspartate aminotransferase domain, and a "PAP2-like" (FBN1b) domain related to a protein superfamily containing both phosphatases and haloperoxidases.

Annex V

List of fibrillin consensus sequences and predicted secondary structure:

>CFBN1.Con.2.0. (cyanobacterial)

MLKKATLLEAIAGKNRGLLATETDKQAILAAIAQLEDRNPTPRPLEASELLEGNWRLLYTTSKGLLNLDRF

PLLKLGQIYQCIRVETAKVYNIAEIYGLPYLEGLVSVAAKFEPVSERRVQVKFERSIIGLQRLIGYQSPASFIQQIE

SGKKFTAIDFPINSREQQGWLDITYLDEDLRIGRGNEGSVFVLTKV

```
    --HHHHHHHHHHH--------HHHHHHHHHHHHHHHH---------------EEEEEEEE----------
------EEEEEEE----EEEEEEEEE-----EEEEEEEEEEEE----EEEEEEEEEEEE-----------------
----------------EEEEEEEEE---EEEEE-----EEEEEE-
```

>FBN1.Con.2.0.

AADADDEWGPEKEEGGAAVAVAEEEPAEVSEIERLKKALVDSFYGTDRGLSASSETRAEIVELITQLEAKN

PTPAPTEALTLLNGKWILAYTSFSGLFPLLSRGTLPLVKVEEISQTIDSENFTVQNSVQFAGPLATTSISTNAKFEV

RSPKRVQIKFEEGIIGTPQLTDSIVLPENVEFLGQKIDLSPFKGLLTSVQDTASSVAKTISSQPPLKFSISNSNAQS

WLLTTYLDEELRISRGDGGSVFVLIKEGSPLLNP

```
    --------------------------HHHHHHHHHHHHHHHHHH--------HHHHHHHHHHHHHHHH--
--------------EEEEEEE-------------------EEEEEEE----EEEEEEEE------EEEEEEEEEE
----EEEEEEEEEE-----------------------------HHHHHHH--------------------
EEEEEEE---EEEEE-----EEEEEE--------
```

>FBN2.Con.2.0.

ATDEWGEKSEPEEPESKLSDSDPPKNEDEWGGEGNGTPAEAGEGEEVDDKLEELKRCLVDTVYGTELGFRA

GSEVRAEVLELVNQLEAANPTPAPVEAPELLDGNWVLLYTAFSELLPLLAAGSLPLLKVEKISQSIDTSSLTIVNST

TLSSPFATFSFSASASFEVRSPSRIQVFKEGTLQPPEIKSSVDLPENVDIFGQKISLSPVQQSLNPLQEAVANISRA

ISGQPPLKVPIPGERTSSWLLTTYLDKDLRISRGDGGLFVLAKEGSPLLDQ

```
    -----------------------------------------HHHHHHHHHHHHHHHHHHHH-------
-HHHHHHHHHHHHHHHH---------------EEEEEE-------------------EEEEEEE----EEEEEE
EE------EEEEEEEEEE----EEEEEEEEEE---------------------------------------------
----------------EEEEEEE---EEEEE----EEEEEE--------
```

FBN3.Con.2.0.

KWRARVSFFPAFLTKAKDAKQLKEELLEAIAPLDRGAEATPEDQQRVDQIARKLEAVNPTKEPLKSDLLNG

KWELIYTTSQSILQTQRPKFLRSIGNYQAINVDTLRAQNMETWPFFNQVTADLTPLNARKVAVKFDYFKIAGLIPIK

APGRARGELEITYLDEELRISRGDKGNLFILKMVDPSYRVPL

```
    ---EE----------HHHHHHHHHHHHHHHH--------HHHHHHHHHHHHHHH--------------
EEEEEEE----------------EEEEE----EEEEEEEE---EEEEEEEEE----EEEEEEEEEEE-------
-----EEEEEEEEE---EEEEE-----EEEEE---------
```

```
>FBN4.Con.2.0.
DLIASLKLKLLSAVSGLNRGLAASEDDLQKADAAAKELEAAGGPVDLSADLDKLQGRWKLIYSSAFSSRTL
GGSRPGPPTGRLLPITLGQVFQRIDVLSKDFDNIVELELGAPWPLPPVEVTATLAHKFELIGSSKIKITFEKTTVKT
TGNLSQLPPLEVPRIPDALRPPSNTGSGEFEVTYLDSDTRITRGDRGELRVFVIS
     -HHHHHHHHHHHHHHH--------HHHHHHHHHHHHHHHH---------------EEEEEEE--------
----------------EEEEEEE-----EEEEEEEE----------EEEEEEEEEEEE----EEEEEEEEEEEE-
------------------------EEEEEEEE---EEEEE-----EEEEE--

>FBN5.Con.2.0.
SQIKTELYQALQGINRGIFGVPSAKKSEIEGLVKLLESQNPTPEPTLNLDKVGGCWKLIYSTITILGSKRT
KLGLRDFISLGDFFQNIDVAKGKAVNVIKFNARGLNLLSGQLTIEASFKIASKSRVDITYESSTITPDQLMNVFRKN
YDLLLGIFNPEGWLEITYVDESLRIGRDDKGNIFVLERSEES
     -HHHHHHHHHHHH---------HHHHHHHHHHHHHHHH----------------EEEEEEE----------
-----------EEEEEE----EEEEEEEE-------EEEEEEEEEEEE----EEEEEEEEEEEE-----------
-----------EEEEEEEE---EEEEE-----EEEEE----

>FBN6.Con.2.0.
TGPDDLVASILSKVTGTDRGVLLTKEEHKEVAEVAQELQKYCVDEPVKCPLIFGEWDVVYCSVPTSPGGGY
RSALGRLFFKTKEMIQVVEAPDIVRNKVSFSAFGFLDGEVSLKGKLKVLDSKWIQVIFEPPELKVGSLEFQYGGESE
VKLQITYVDEKIRLGKGSRGSLFVFQRRG
     ---HHHHHHHHHHHHH--------HHHHHHHHHHHHHHHH--------------EEEEEEEE---------
-------------EEEEE----EEEEEEE------EEEEEEEEEEEE----EEEEEEEEEEEE-------------
EEEEEEEE---EEEEE-----EEEEEE--

>FBN7.Con.2.0.
AMVQQAVQGAPAAYAKEMERLSAKESLLLAFKDAGGFEALVTGKTTDMQRIDVNERITGLERLNPTPRPTT
SPFLEGRWNFEWFGSGSPGLFAARFLFERFPSTLANLSKMDVLIKDGYAKITANVKLLNSIESKFILSTKLSVEGPL
RMKEEYVEGILETPTVIEETVPEQLKGALGQAVTTLQQLPVPIRDAVSSGLKIPLSGSFQRLFMISYLDEEILIIRD
TAGVPEVLTRLETPPSSLAEPTVEYES
     --HHHHH----HHHHHHHHHHHHHHHHHHHHHH-------------HHHHHHHHHHHHHHH---------
------EEEEEEE------------H------------EEEEEE----EEEEEEEE------EEEEEEEEEE----
EEEEEEEEEEEE-----------HHHHH--------------H---------------EEEEEEE---EEEEE-
----EEEEEE----------------

>FBN8.Con.2.0.
LELENKKHELLRAVQDTQRGLVATADQRSSIEEALVSVEGYNAGAPIDLVKLDGTWRLQYTSAPDVLVLFE
AAARLPFFQVGQIFQKFECRDRSDGGIVRNVVRWSIPNLLEEQEGATLLVSAKFSVVSVRNIYLQFEEISVQNIKIS
EQLQALIAPAILPRSFLSLQILQFIRTFKAQIPVSPGRRSVGGLYYLSYLDRNMLLGRAVGGGGVFVFTRAQPLEL
```

197

```
        --HHHHHHHHHHHHHH--------HHHHHHHHHHHHHHHH--------------EEEEEE----------
----------EEEEEE--------EEEEEEE----------EEEEEEEEEE----EEEEEEEEEEE-----
-----------------------------------------EEEEEEE---EEEEE------EEEEE------
```

>FBN9.Con.2.0.

```
SAKQHLLNLISDQDRGLKTQSDPSKLASIVSAIDALAARGRDTVTTGDSLSATWRLLWTTEKEQLFIIEKA
PLFGTQAGDVLQVIDVEKKTLNNVITFPPDGVFFVRSSIEIASPQRVNFRFTSAVLRGKNWEIPLPPFGQGWFETVY
LDDEIRVVKDIRGDYLVVERAPYSWKE
```

```
       -HHHHHHHHHHH----------HHHHHHHHHHHHHH---------------EEEEEEE------------
--------EEEEEEE----EEEEEEE----EEEEEEEEEE----EEEEEEEEEEEE-----------EEEEEEE
---EEEEE-----EEEEE--------
```


>FBN10.Full.Con.2.0.

```
SFTEAENSLIEALIGIQGRGRSASPQQLNEVESAVQVLEGLEGVPDPTSSSLIEGRWQLMFTTRPGTASPI
QRTFVGVDFFSVFQEVYLRTNDPRVSNIVKFSDAIGELKVEAAASIKDGKRILFRFDRAAFSFKFLPFKVPYPVPFR
LLGDEAKGWLDTTYLSPSGNLRISRGNKGTTFVLQKKTEPRQKLLSAISTGTGVKEAIDEFISSNQNVAKDELELLE
GEWQMIWSSQTDSWLENAANGLMGKQIVKKNGQLKFLVDILPGLRFSMTGNFVKSGSNTYDVTMDDAAIIGGPFGYP
LEMESKFNLELLYTDEKIRISRGYNNILFVHLRVDG
```

```
       --HHHHHHHHHHHHHHH-------HHHHHHHHHHHHHHHHH--------------EEEEEEEE--------
---------HHHHEEEE-----EEEEEEEE-----EEEEEEEEE-----EEEEEEEEEEEE---------------
-------EEEEEEEE-----EEEEE-----EEEEE----HHHHHHHHHH---HHHHHHHHHHHH---------HHH-
-EEEEE------HHHHHH------EEEE----EEEEEE----EEEEE---EE-----EEEEEEE--EEEE--EE--
------EEEEEEEE---EEEEE-----EEEEEE---
```


>FBN11.Dom.2.Con.2.0.

```
VVPSMDIIRWGLGSTAVRIAEEYIYRQPQRSRLAHFIELMEMLNPHSKPKNWLELLPGKWRLLYCTGRHIG
LTLRQPSVRVLIGDVHLTVSRASKSNTNLSFTSDIGFTVMPGRDWPHDKSGVTGKLQVNSSFRLTAGRRLYLKEEKT
TSKFSSGQSDAEESLGQKLSGRKWRKAIPFKEFPSSLPVAKLVSDEIDVTMSLGDPLNVDSARNVLQEVRTQIPPEM
FDLSKLVCGTYVDSRLLVLRGVNGSALLFTRSCVDES
```

```
       ---HHHHHHH----HHHHHH--------HHHHHHHHHHHHHH--------------EEEEEEE------
----------EE-EEEEEE-----EEEEEEEEE-----EE-----------EEEEEEEEEEEE----EEEEEEEEE
EE-------HHHHHHH------HHH-------------------------------------EE----------
-----EEEEE----EEEE-----EEEEEEE-----
```

```
>FBN12.Con.2.0.
QPNKEVEESVKVLKNAAKTRKVPAEEVLSALSVIEKAKLDPSGFLETLGGTESPGRTWMLIFTAEKKLKGG
RYFPVTAVQRFDAAGKRIENGVYLGPIGCLTFEGRLSWKKRILAFIFERIRIKVGPLNPLEISLGQKDDREPSTKDP
FFIWFYVDEEIAVARGRSGGTAFWCRCRRVT
     -----HHHHHHHHHHHH-------HHHHHHHHHHHHHHH----HHHH----------EEEEEEEE------
-EEEEEEEE------EEEEEEEE---EEEEE--------------EEEEEE----EEE---------------
-EEEEEE---EEEEE-----EEEEEE-----
```

Kyte-Doolittle hydropathicity plots for each consensus fibrillin or fibrillin domain (positive indicates hydrophobic). Wheel diagrams for Helix-1 and Helix-2. Code: red, negative; blue, positive; yellow, hydrophobic; gray, small; green, proline; purple and orange, non-charged polar amino acids. Amino acid composition of consensus fibrillins.
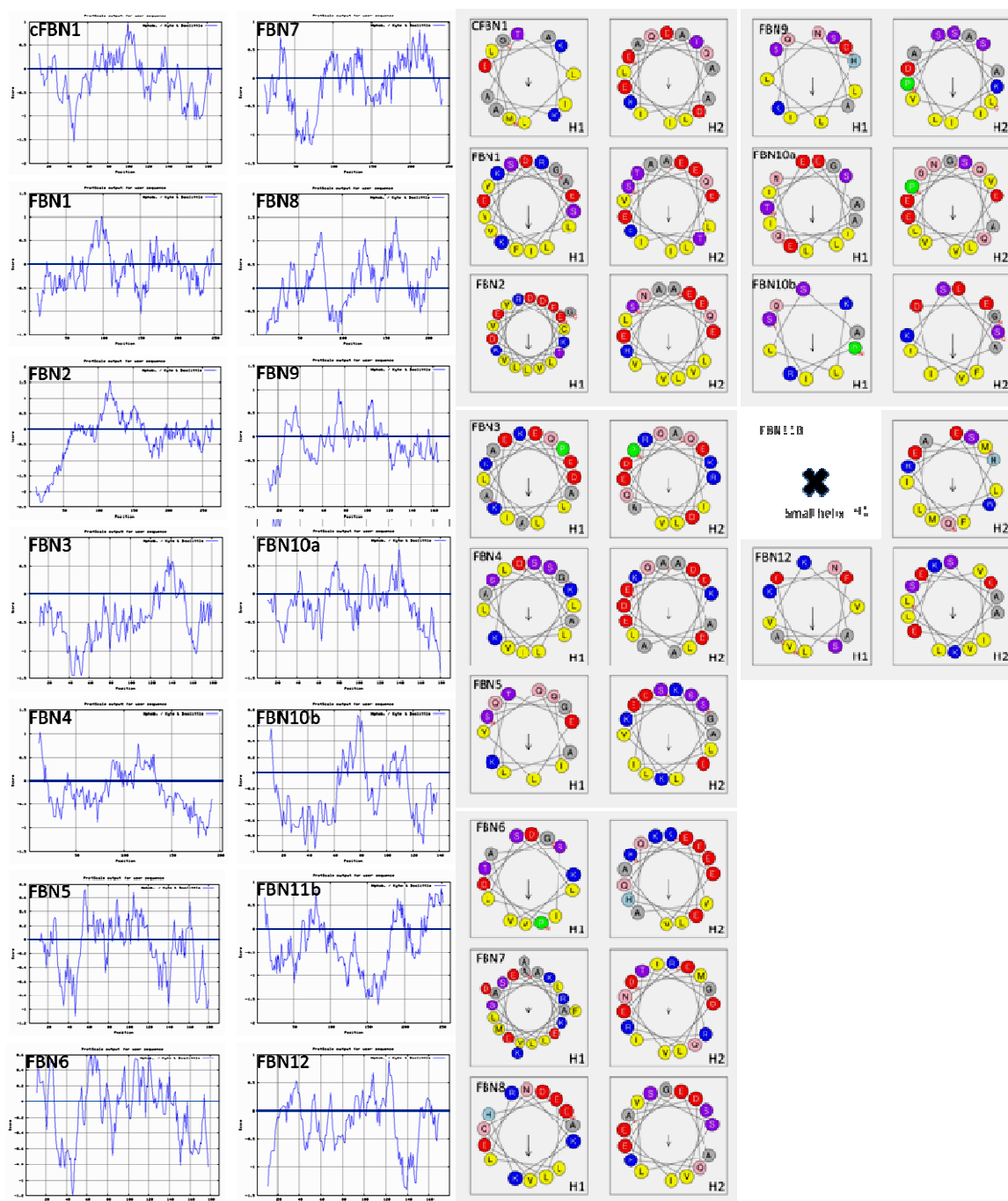
% amino acid composition of consensus fibrillins vs. average tomato protein

Amino acid	CFBN1.2.0	FBN1.2.0	FBN2.2.0	FBN3.2.0	FBN4.2.0	FBN5.2.0	FBN6.2.0	FBN7.2.0	FBN8.2.0	FBN9.2.0	FBN10.2.0	FBN10(A).2.0	FBN10(B).2.0	FBN11(B).2.0	FBN12.2.0	Avg. tomato protein
Ala (A)	15 7.7	21 8.1	19 6.9	18 9.5	16 7.9	7 3.7	8 4.3	20 7.9	18 8.0	12 6.9	19 5.6	12 6.6	7 4.5	10 3.9	13 7.3	6,2
Arg (R)	13 6.7	8 3.1	9 3.3	14 7.4	12 5.9	9 4.7	10 5.3	13 5.2	17 7.6	11 6.3	18 5.3	13 7.1	5 3.2	21 8.1	15 8.4	5,1
Asn (N)	7 3.6	9 3.5	9 3.3	8 4.2	4 2.0	11 5.8	1 0.5	6 2.4	7 3.1	5 2.9	16 4.7	6 3.3	10 6.5	8 3.1	4 2.2	4,8
Asp (D)	7 3.6	11 4.2	15 5.4	11 5.8	14 6.9	10 5.3	9 4.8	8 3.2	8 3.6	13 7.4	17 5.0	8 4.4	9 5.8	13 5.0	6 3.4	5,3
Cys (C)	1 0.5	0 0.0	1 0.4	0 0.0	0 0.0	1 0.5	3 1.6	0 0.0	1 0.4	0 0.0	0 0.0	0 0.0	0 0.0	3 1.2	3 1.7	1,9
Gln (Q)	11 5.7	10 3.9	10 3.6	10 5.3	5 2.5	7 3.7	7 3.7	9 3.6	16 7.1	8 4.6	14 4.1	8 4.4	6 3.9	8 3.1	3 1.7	3,7
Glu (E)	16 8.2	26 10.0	31 11.2	13 6.8	12 5.9	13 6.8	16 8.5	25 9.9	15 6.7	10 5.7	24 7.1	12 6.6	12 7.7	14 5.4	16 8.9	6,4
Gly (G)	14 7.2	17 6.6	20 7.2	8 4.2	17 8.4	16 8.4	19 10.1	17 6.7	16 7.1	10 5.7	29 8.6	15 8.2	14 9.0	18 6.9	16 8.9	6,3
His (H)	0 0.0	0 0.0	0 0.0	0 0.0	1 0.5	0 0.0	1 0.5	0 0.0	1 0.4	1 0.6	1 0.3	0 0.0	1 0.6	5 1.9	0 0.0	2,4
Ile (I)	16 8.2	16 6.2	11 4.0	12 6.3	11 5.4	18 9.5	8 4.3	14 5.6	14 6.2	12 6.9	21 6.2	10 5.5	11 7.1	11 4.2	11 6.1	5,8
Leu (L)	27 13.9	29 11.2	35 12.7	22 11.6	29 14.3	23 12.1	17 9.0	31 12.3	30 13.4	17 9.7	34 10.1	17 9.3	17 11.0	32 12.4	15 8.4	9,5
Lys (K)	12 6.2	14 5.4	13 4.7	16 8.4	12 5.9	15 7.9	16 8.5	13 5.2	7 3.1	10 5.7	20 5.9	8 4.4	12 7.7	15 5.8	16 8.9	6,4
Met (M)	1 0.5	0 0.0	0 0.0	2 1.1	0 0.0	1 0.5	1 0.5	6 2.4	1 0.4	0 0.0	6 1.8	1 0.5	5 3.2	6 2.3	1 0.6	2,5
Phe (F)	7 3.6	11 4.2	9 3.3	9 4.7	7 3.4	9 4.7	10 5.3	12 4.8	12 5.4	9 5.1	22 6.5	15 8.2	7 4.5	9 3.5	11 6.1	4,3
Pro (P)	8 4.1	16 6.2	23 8.3	12 6.3	15 7.4	6 3.2	9 4.8	17 6.7	9 4.0	9 5.1	15 4.4	11 6.0	4 2.6	13 5.0	10 5.6	4,7
Ser (S)	10 5.2	28 10.8	31 11.2	7 3.7	18 8.9	16 8.4	16 8.5	19 7.5	16 7.1	13 7.4	29 8.6	17 9.3	12 7.7	29 11.2	9 5.0	8,8
Thr (T)	10 5.2	20 7.7	15 5.4	11 5.8	14 6.9	11 5.8	7 3.7	19 7.5	7 3.1	12 6.9	19 5.6	11 6.0	8 5.2	15 5.8	10 5.6	5,0
Trp (W)	2 1.0	3 1.2	4 1.4	3 1.6	2 1.0	2 1.1	2 1.1	2 0.8	2 0.9	5 2.9	5 1.5	2 1.1	3 1.9	5 1.9	4 2.2	1,3
Tyr (Y)	7 3.6	3 1.2	3 1.1	5 2.6	2 1.0	5 2.6	5 2.7	5 2.0	6 2.7	3 1.7	7 2.1	3 1.6	4 2.6	5 1.9	3 1.7	3,0
Val (V)	10 5.2	17 6.6	18 6.5	9 4.7	12 5.9	10 5.3	23 12.2	16 6.3	21 9.4	15 8.6	22 6.5	14 7.7	8 5.2	19 7.3	13 7.3	6,6

201

Annex VII: Abbreviatures used

Table of the used abbreviatures			
ABA	abscissic acid	HPLC	high performance liquid chromatography
ACC	1-aminocyclopropane-1-carboxylic acid	JA	jasmonic acid
accD	acetyl coA carboxylase D	LB	luria broth
ADP	adenosine diphosphate	LCY-B	lycopene b-cyclase
APS	ammonium persulphate	LCY-E	Lycopene e-cyclase
ATP	adenosine triphosphate	MAPK	Mitogen activated protein kinase
BHT	(butylated hydroxytoluene)	MAPKK	Mitogen activated protein kinase
CAC	clathrin adaptor complex (housekeeping gene)	MGDG	monogalactosyl diacylglycerol
CRTISO	carotene isomerase	NXS	neoxanthin synthase
CRTR-B1	b-ring hydroxylase-1	PC	phosphatidylcholine
CRTR-B2	b-ring hydroxylase-2	PCR	polymerase chain reaction
CRTR-E	e-ring hydroxylase	PDS	phytoene desaturase
CTR	Constitutive triple response	PE	phosphatidylethanolamine
CYC-B	chromoplast-enhanced	Pg	plastoglobule
DGDG	dygalactosyl diacylglycerol	PG	phosphatidyl glycerol
DMSO	dimathyl sulfoxide	PI	phosphatidyl inositol
DO600	optical density at 600 nm	pI	isoelectric point
DTT	dithiothreitol	PMSF	phenylmethylsulfonyl fluoride
EDTA	ethylenediaminetetraacetic acid	PS	phosphatidyl serine
EIL	Ethylene insensitive like	PVPP	polyvinyl polypyrrolidone
EIN2	Ethylene insensitive 2	qRT-PCR	quantitative real-time PCR
EIN3	Ethylene insensitive 3	ROS	reactive oxygen species
ELISA	enzyme-linked immunosorbent assay	SDS	sodium dodecylsulphate
ERE	Ethylene-responsive element	SQDG	sulphoquinovosyl diacylglycerol
ERF	Ethylene response factor	TAE	tris acetate EDTA buffer
FBN	fibrillin	TAG	triacylgliceride
FCR	fibrillin conserved region	TEMED	tetramethylethylenediamine
FFA	free fatty acid	VDE	violaxanthin de-epoxidase
GGPS	geranylgeranyl diphosphate synthase	VIGS	virus-induced gene silencing
GRAVY	grand average of hydropathicity	YEB	yeast extract broth
HAM	hexane:acetone:methanol 2:1:1	ZDS	z-carotene desaturase
HAM	hexane:acetone:methanol mixture 2:1:1 v/v	ZEP	zeaxanthin epoxidase

Number, description and location of tables displayed: