Genes upregulated during flagellar regeneration in the scaly green alga *Scherffelia dubia*: identification of the first scale-associated protein, a novel protein

INAUGURAL-DISSERTATION

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Jean-Michel Doan

aus Paris, France

2007

Berichterstatter:	Prof. Dr. M. Melkonian
	Prof. Dr. S. Waffenschmidt
Tag des mündlichen Prüfung:	27 April 2007

Kurzzusammenfassung

Die Geißeln der einzelligen Grünalge *Scherffelia dubia* sind vollständig mit feinen Strukturen bedeckt, die als Schuppen bezeichnet werden. Diese Schuppen sind in mehreren Schichten angeordnet und stellen eine einfache Form der extrazellulären Matrix dar. Neuere Untersuchungen legen den Schluß nahe, daß Keto-Zuckersäuren beinhaltende beschuppte Algen die Vorfahren der Landpflanzen sein könnten. Die Geißelschuppen von *S. dubia* bestehen hauptsächlich aus selten vorkommenden 2-Keto-Zuckersäuren, die mit ungefähr 20 Glykoproteine (SAPs: scale ssociated proteins) assoziiert sind. Schuppen entstehen im Golgi-Apparat und werden während der Bildung neuer Geißeln synthetisiert. Im Gegensatz zu der gut untersuchten Ultrastruktur und Kohlenhydratzusammensetzung der Geißelschuppen von *S. dubia* ist vergleichsweise sehr wenig über die SAPs bekannt.

Die vorliegenden Arbeit beschreibt die Charakterisierung der SAPs und die Isolierung einer 2,4 kb großen cDNA, die für das c-terminale Ende eines 98 kDa großen SAPs kodiert. Die Sequenz dieser SAP-cDNA (SAP98) zeigt keine Ähnlichkeit zu einem anderen bekannten Protein. Wenn die Geißelbildung experimentell induziert wird beobachtet man eine verstärkte Expression von SAP98. Die für SAP98 entwickelte Methode zur Isolation der SAP98 cDNA konnte nicht auf weiterer SAPs übertragen werden. Um weitere für die Geißelbildung wichtige Gene zu identifizieren wurde eine Subtraktions-cDNA-Bank (Geißelregeneration / Interphase Zellen) mittels "suppression substraction hybridization (SSH)" angelegt.

Die Qualität der für die Geißelregeneration spezifischen c-DNA-Bibliothek wurde durch signifikante Redundanz und Kontaminationen von rRNA beeinträchtigt. Daher konnte ein deutlicher Anteil (55%) der Contigs durch similarity searches nicht identifiziert werden. Nur ca. 5% (35) der erhaltenen Contigs konnten während der Geißelregeneration verstärkt exprimierten Genen zugeordnet werden. 24 der 35 Contigs zeigen keine Ähnlichkeiten zu bekannten Genen. Weitere SAPs konnten in der Subtraktionsbank nicht identifiziert werden, jedoch enthielt die Subtraktionsbank Gene die mit großer wahrscheinlichkeit für die Polysaccharidsynthese der Geißelschuppen.

Abstract

The flagella of the unicellular green alga *Scherffelia dubia* are entirely covered with layers consisting of thousands of discrete structures called scales. This complex covering represent a primitive form of extra cellular matrix which was developed before the land plants evolved from the green algae. Recent studies indicated that 2-keto sugar acid-containing scaly algae might be the ancestors of land plants. The flagellar scales of *S. dubia* are mainly composed of rare 2-keto sugar acids to which about twenty glycoproteins associate. They are produced in the Golgi apparatus and synthesized when new flagella are growing. The ultrastructure and carbohydrate composition of *S. dubia*'s flagellar scales have been studied in great detail but not the scale associated proteins (SAP).

In the present thesis SAPs were characterised and the isolation of a 2,4 kb cDNA corresponding to the C-terminal end of a SAP of 98 kDa was performed successfully. The sequence of the SAP98 cDNA indicated that this protein is not similar to any other known protein. It was also observed that the SAP98 gene was upregulated during experimentally-induced flagellar regeneration. Since the procedure used to isolate the SAP98 cDNA could not be successfully repeated for other SAPs, we isolated genes which are upregulated during flagellar regeneration in *Scherffelia dubia* by constructing a subtracted cDNA library (flagellar regenerating cells – interphase cells) using suppression subtraction hybridization (SSH).

The flagellar regeneration-specific cDNA library suffered from significant redundancy and rRNA contamination. A significant proportion (55%) of the contigs could not be identified by similarity searches. It was estimated than 5% of the contigs of the library corresponded to upregulated genes and 35 contigs were identified as potential candidate for flagellar regeneration-upregulated genes, of which 24 might represent novel genes. The SAPs were not found in this library, but two genes coding for proteins involved in the scale polysaccharides biosynthesis, GDP-mannose-3,5-epimerase and UDP-glucose dehydrogenase, were present.

À mes parents

Table of content

List of tables

List of figures

List of abbreviations

Ø	diameter
µ	micro (10^{-6})
5 OmetKDO	3-deoxy-5-O-methyl-2-octulosonic acid
A	alanine
ADP	adenosine diphosphate
ATP	adenosine triphosphate
BLAST	basic local alignement search tool
bp	base pair
C	Celsius
CCAC	culture collection of algae at the university of Cologne
cDNA	complementary DNA
Ci	Curie units
CMP	cytidine monophosphate
C.r.	*Chlamydomonas reinhardtii*
CsCl	cesium chloride
CTP	cytidine triphosphate
CTAB	hexadecyltrimethylammonium bromide
D	aspartic acid
Da	Dalton
dATP	deoxyadenosine triphosphate
DEPC	diethyl pyrocarbonate
DHA	3-deoxy-2-heptulosaric acid
dH_2O	distilled water
DNA	desoxyribonucleic acid
DTT	dithiothreitol
E	glutamic acid
EC	enzyme commission number
EDTA	ethylenediaminetetraacetic acid
ECM	extra cellular matrix
EGTA	ethylene glycol-bis(2-aminoethylether)-*N,N,N',N'*-tetraacetic acid
EST	expressed sequence tag
Et-Br	ethidium bromide
FAP	flagella-associated protein
FR	flagellar regeneration
G	glycine
GCG	genetics computer group, or Wisconsin package
GDP	guanosine diphosphate
GSP	gene-specific primer
h	hour
HCl	hydrogen chloride

h-scales	hair scales
Hyp	hydroxyproline
I	interphase
IFT	intraflagellar transport
iPCR	inverse PCR
JGI	joint genome institute
k	kilo (10^3)
K	lysine
kb	kilobase
KDO	3-deoxy-2-octulosonic acid
L	liter
LC	liquid chromatography
m	meter
M	molar
min	minute
MOS	mirror oriented selection
MS	mass spectrometry
m-scales	man scales
mRNA	messenger RNA
M.v.	*Mesostigma vriride*
NADH	nicotinamide adenine dinucleotide
ncbi	national center for biotechnology information
NEB	New England biolabs
dNTP	deoxynucleotide
P	phosphate group
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEP	phospho-enol-pyruvate
ppi	pyrophosphate
Pro	proline
p-scales	pentagonal scales
Q	glutamine
R	arginine
RACE	rapid amplification of cDNA ends
Rbcs	ribulose biphosphate carboxydase small unit
RNA	ribonucleic acid
rpm	revolutions per minute
RT	reverse transcriptase
rRNA	ribosomal RNA
S	sulfur
SAP	scale-associated protein
S.d.	*Scherffelia dubia*
SDS	sodium dodecyl sulfate
Ser	serine

SMART	switching mechanism at 5'-end of RNA
SSC	sodium chloride - sodium citrate buffer
SSH	subtraction suppression hybridization
Taq	*Thermus aquaticus*
TBS	trisbuffered saline
TCA	trichloroacetic acid
TEM	transmission electron microscope
Thr	threonine
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol
tRNA	transfer RNA
U	unit
UDP	uridine diphosphate
UTR	untranslated region
UV	ultraviolet
vol:vol	volume per volume
ZMMK	Zentrum für molekulare Medezin der Universität zu Köln

Introduction

1 Background

Scherffelia dubia (Perty) Pascher emend. Melkonian & Preisig (Melkonian and Preisig 1986) is a freshwater, flagellate, green alga found in ponds and pools around Europe. It belongs to the scaly green flagellates (formerly prasynophytes), a group of algae that harbors discrete structures over their cell bodies and/or flagella called scales. These algae represent an important step in the evolution of the Viridiplantae as they are thought to have given rise to all other green algae and land plants.

Scherffelia dubia bears four structurally identical flagella that provide motility to the cell. The cell is oval-shaped, 13-15 μm long, 8-10 μm in width, and 5-6 μm thick (Fig. 1, upper left). Each flagellum is about 12-15 μm long. Two chloroplasts occupy the posterior lumen of the cell, their lobes stretching along the periphery towards the anterior end of the cell where the flagellar apparatus lies. The anterior part of the cell also contains the endomembrane system, in close association with the central nucleus. A unique feature of *Scherffelia,* and its sister taxon *Tetraselmis,* is that the cell body is enclosed in a rigid cell wall called theca. On the anterior end, the theca is interrupted at the flagella groove to let the four flagella emerge. Flagellar scales, organized in precise and complex layers, cover the flagella membrane of each flagellum. They are the focus of this thesis.

Scale coverings are one form of the various extra cellular matrices (ECM) developed in the realm of plants. Thecae, frustules, mucilages, glycoproteins or polysaccharidic frameworks in algae, pecto-cellulosic cell walls in the higher plants ; all these examples represent the diversity and evolution of the ECM to fullfill various essential functions for the cell: protection, adhesion, regulation of growth, signalling, etc.

Scales are essentially found in flagellate algae. Their biochemical constitution, sizes and shapes are often species-specific. They can have various, beautiful, often complex shapes, ranging in size from 30-40 nm up to 1-2 μm. They can be mineralized or non-mineralized, consisting mainly of carbohydrates, and some scale types contain minor amounts of proteins. Scale shapes and organizations are so diverse that they have been commonly used as morphological characters for the identification and classification of algae. Often, scales assemble closely together, creating overlapping layers which cover the surface of the cell. These very dense layers form thick coverings around the cell that have been interpreted as a protection against viruses and pathogens (Becker et al. 1994), while at the same time allowing flagella beating ; however, no function has been characterized yet and, as other types of ECMs, they may well have some other functions. Each scale layer usually consists of a single type of scale and up to five or six layers have been observed on a single

cell. The scales covering the flagella are often different from the scales covering the cell body, which makes it important to distinguish flagellar and cell body scales.

Most scaly algae used to be found in the class *Prasinophyceae* described in Sym & Pienaar (1993), and a few scaly algae also belong to the *Chrysophyceae* and *Prymnesiophyceae* (Melkonian et al. 1991).The class *Prasinophyceae* is however no longer recognised and we shall refer to scaly green flagellates instead. Phylogenetic analyses showed that scaly green flagellates are found at the base of all green algal lineages (Chlorophyceae, Ulvophyceae, Trebouxiophyceae and the Streptophytes). Hence, they are presumed to have included the ancestors of all other green algae and land plants. The ability to synthesize scales is therefore considered a primitive character, which could either have been lost or have led to other types of cell surfaces. The scaly biflagellate *Mesostigma viride* is particularly interesting, as recent molecular and philogenetic data argue for a position within the Streptophyta (Melkonian et al. 1995, Bhattacharya et al. 1998, Marin and Melkonian 1999, Lemieux et al. 2000, Nedelcu et al. 2006, Simon et al. 2006, Lemieux et al. 2007). These studies therefore highlight the possibility of an evolutionary relation between scale coverings and plants' cell wall. In this respect, the presence of 3-deoxy-2-octulosonic acid (KDO), an uncommon sugar typical of gram-negative prokaryote envelopes but also a major constituant of green algae scales (Becker et al. 1991), which is apparently not synthesized in other eukaryotes (Royo et al. 2000b), in the cell wall of land plants is very significant (York et al. 1985).

The complex extracellular structures created by scaly algae naturally raised the interest on the mechanisms by which these molecules are synthesized and assemble. Since scale are synthesized in the Golgi apparatus (Manton 1965, Manton et al. 1965, Brown 1974, Moestrup and Walne 1979), they became an excellent system to study the mechanisms of Golgi apparatus function and exocytosis by electron microscopy (Melkonian et al. 1991, Becker and Melkonian 1996). Because of their thecae, *Scherffelia* and *Tetraselmis* provided model systems for the study of scale biogenesis and Golgi transport (Domozych et al. 1981, McFadden and Melkonian 1986, McFadden et al. 1986, Becker et al. 1995). They have been the strongest support for the cisternal maturation/progression model (Morré and Mollenhauer 1974). However, the answers to such questions as, how are scales composed, synthesized, secreted or assembled, rely on deeper knowledge of scale structure and composition.

2 Flagellar scales of *S. dubia*

Among scaly flagellates, *Scherffelia* and *Tetraselmis* are the only two genera known to synthesize a theca over the cell body in place of scale layers. Together, they form the order *Chlorodendrales* (Melkonian and Preisig 1986, Melkonian 1990). In *Scherffelia*, the theca itself is the result of the extracellular coalescence of two specific types of scales synthesized

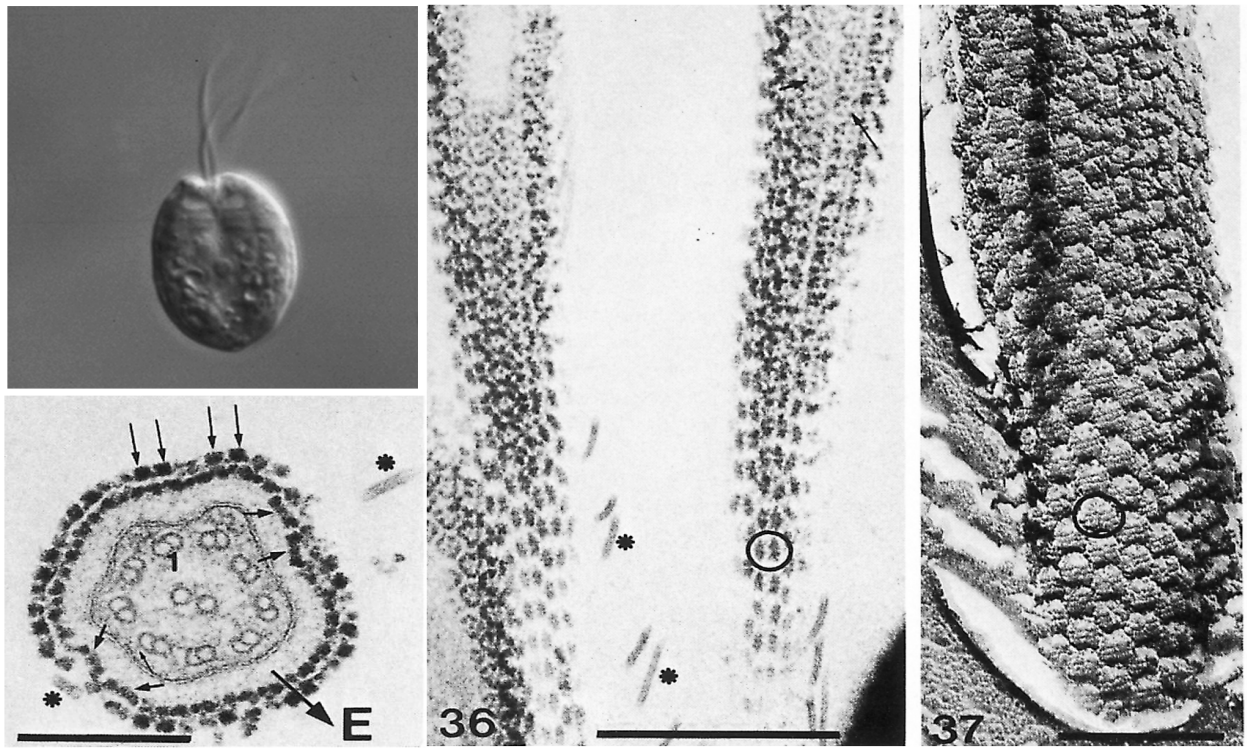

Figure 1. **Structure of flagellar scales of *Scherffelia***. Top left box: upper view of a cell of *S.dubia* at the light microscope. Box E: cross section of a flagellum, 1: first pair of microtubule, thin arrow: β tubule connecting axonemal microtubule with flagella membrane under the site of hair scale attachment, short arrows: square scales, long arrows: cross section through the two subunits of a double scale, large arrow: direction of beating of the flagellum, asterisks: hair scales. Box 36: TEM of *Scherffelia*'s flagella showing the layers of pentagonal scales and double scales (encircled), scale bar: 0,5 μm. Box 37: deep-etch preparation from a flagellum, a double scale is encircled, scale bar:0,25 μm. Figures taken from Melkonian and Preisig 1986.

during cell division (Manton et al. 1965, McFadden et al. 1986, Wustman et al. 2004). The flagella remain covered with layers of individual scales.

This unique character greatly favored the study of flagellar scales in these two organisms because it is associated with a temporal separation between flagellar scales and thecal scales synthesis, and facilitates the purification of flagellar scales.

2.1 Ultrastructure of flagellar scales

Scale ultrastructure was deduced from observations made with the electron microscope ; biochemical analyses provided some data as well. Scales can be observed either *in situ* or after purification (Fig. 1). Crude scale fractions (containing all types of flagellar scales) can be isolated after flagella isolation and solubilization of the flagella membrane with detergents (Becker et al. 1990, Becker et al. 1996). Individual scale types (except hair scales) can be further purified by CsCl or sucrose gradient centrifugation (Becker et al. 1990). These procedures allowed the fine observation of scale structure and the determination of their composition.

Scherffelia dubia and *Tetraslemis striata* Butcher synthesize similar flagellar scales, which I

will describe further. Pentagonal scales (Ø:40 nm) consist of an outer ring of five subunits that form the corners of the scales, to which a sixth subunit is loosely associated in the center of the ring. Man scales (40x25 nm) are asymetrical structures reminiescent of a human figure (hence their name) without discernable subunits. *In situ*, two man scales interlock head-to-tail to form globular double-scales in which individual subunits are not discernable anymore (Melkonian 1982). Hair scales are long hollow tubes (1 μm long) formed by the linear association of numerous, different scale subunits. The arrangement of these subunits define three zones: a proximal filament, a tubular shaft and a distal zone, and is considered species-specific (Marin et al. 1993, Becker et al. 1994). In addition, very thin coiled filaments (about 50 nm long), termed knotted scales, have been identified in *T. striata* (Becker et al. 1990), but their occurrence has not been confirmed in *S. dubia*.

Calcium ions and weak ionic interactions appear to maintain the scale ultrastructure together: EDTA treatment removes the central subunit of pentagonal scales both in isolated scales and *in situ*, and double scales dissociate in single man scales during detergent purification of scales. These interactions might also drive the assembly into the different scales layers.

2.2 Topology of flagellar scales

Flagellar scales of *S. dubia* are organized in two layers that surround the flagella membrane (Fig. 1) (Melkonian and Preisig 1986). The organization is also similar in *T. striata*. An inner layer (or underlayer) is formed by the precise alignment of pentagonal- and square-shaped scales in twenty-four longitudinal compact rows (twenty rows of pentagonal scales and four rows of square scales, in a 10-2-10-2 alternance). Square scales are probably pentagonal scales subjected to contraction forces caused by the density of the assembly (Melkonian 1982). The pentagonal scales inner layer does not lie directly over the flagella membrane (Fig. 1, box E), but is separated from it by a 40 nm region which is stained in tannic acid-prepared electron micrographs, indicative of the presence of protein material (Becker and Melkonian 1992).

Twenty longitudinal rows of man-shaped scales precisely cover the pentagonal scales, forming the second, outer layer. The outer layer is not continuous, as it interrupts itself over the two diametrally-opposed square scales rows (Fig. 1, box E). From this spacing emerge regularly (every 100 nm) the third type of scales, flagellar hairs, creating two diametrally-opposed rows of hairs perpendicular to the plane of beating of the flagellum. These scales are apparently connected to the axoneme through the inner layer and the flagellar membrane, which may be linked to their orientation in relation to the plane of beating. The perpendicular orientation of hair scales increases the contact with the external milieu and may so enhance flagellar beating.

Around 13 600 individual scales cover the surface of each flagellum: 5 000 pentagonal scales (including square scales), 8400 man scales and 200 hair scales (Melkonian and

Preisig 1986). The knotted scales of *T. striata* seem to form a third layer over the man scale layer (Becker et al. 1990). The preceding brief description of scale topography should highlight the precision required for the disposition of scales over the flagellar membrane. The factors that drive such a complex assembly are presently not known but may lie in the physico-chemical properties of their constituents, *i.e.* scale polysaccharides, proteins, and calcium ions.

2.3 Composition of flagellar scales

The major constituents of scales are three unusual 2-keto-sugar acids: 3-deoxy-2-octulosonic acid (KDO), 3-deoxy-5-O-methyl-2-octulosonic acid (5 OMetKDO) and 3-deoxy-2-heptulosaric acid (DHA) (Fig. 2). Arabinose, xylose, gulose, galactose, glucose, and galacturonic acid are minor constituents, as well as calcium and sulfates (Table 2) (Becker et al. 1990, Becker et al. 1994, Becker and Melkonian 1996). The considerable amount of negative charges borne by the acidic sugars of scales may be viewed as having a role in divalent cation metabolism and maintaining salt and water balance in the cell (Aken and Pienaar 1985). Small amounts of proteins are also present in some scale types. They are essentially found in pentagonal scales, a few seem to associate with hair scales, and man scales are completely protein-free (Becker et al. 1990, Becker et al. 1994).

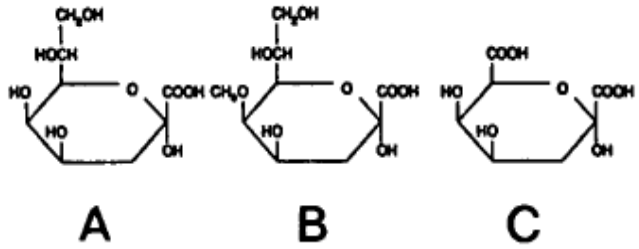

Figure 2. Chemical structure of 2-keto sugar acids found in green algae scales: 3-deoxy-2-octulosonic acid (A), 3-deoxy-5-O-methyl-2-octulosonic acid (B) and 3-deoxy-2-heptulosaric acid (C)

	T. striata			*S. dubia*
	p-scales	m-scales	Crude scale fraction	Crude scale fraction (p,m,h scales)
Protein content (% d.w.)	11,0	-		10
Carbohydrates (% d.w.)	50-70	50-70		70
Monosaccharide composition (mol %)				
Arabinose	+	5,0	+	6,5
Gulose	12,0	9,0	7,5	4,7
Galactose	+	5,0	+	3,9
Glucose	2,0	+	+	2,8
Galacturonic acid	-	-	-	3,5
KDO	23,0	23,0	25	22,5
5OMetKDO	35,0	33,0	37,5	22,5
DHA	28,0	25,0	30	26,7

Table 1.Compositon of different scale types and of crude scales fractions in *Tetraselmis striata* and *Scherffelia dubia* (Becker et al. 1994, Becker and Melkonian. 1996)

Pentagonal and man scales possess a remarkable similar carbohydrate composition (Becker et al. 1990). They differ mostly by the presence of proteins associated to the pentagonal scales. It is therefore likely that the differences of shape between pentagonal and man scales are determined by qualitative differences in the polysaccharide structure. By contrast, the thecal scale composition is much different as they contain as much as 60 % KDO (Becker et al. 1991), and scales synthesized in some other scaly green algae contain the same monosaccharides but in significantly different proportions (Becker et al. 1991). It can also be noted that ultrastructurally identical scales from *T. striata* and *S. dubia* have highly similar but not totally identical compositions. So far almost nothing is known about the scale polysaccharide structures. In higher plants and bacteria, KDO and DHA are only minor components and their rarity suggests that they do not have a structural role. On the contrary, these polysaccharides are abundant in scales. The only data available so far show that thecal scales of *Tetraselmis* contain several polysaccharides of different size built from a KDO backbone on which minor sulfated sugars such as galactose are linked (Becker and Melkonian. 1998).

The three 2-keto sugar acids (Fig. 2) that are the main constituants of flagellar scales are actually quite uncommon in eukaryotes. KDO and DHA have only been found in the pectin rhamnogalacturonan II in the primary cell wall of higher plants (York et al. 1985, Stevenson et al. 1988, Hirano et al. 1994), and in green algae scales (Becker et al. 1991). They occur neither in the volvocalean ECM which is built from glycoproteins (in *Chlamydomonas* or *Volvox*, for instance), nor in the scales of prymnesiophytes (Becker et al. 1991).

KDO is an essential component of the lipopolysaccharide of the outer membrane of Gram-negative bacteria (Unger et al. 1980, Goldman et al. 1987). In *Escherischia coli*, it is synthesized in two steps :

(1) D-Arabinose-5P + PEP → KDO-8P (EC 4.1.2.16)

(2) KDO-8P + H2O → KDO (EC 3.1.3.45)

to which a third reaction is necessary before incorporation into a lipid precursor :

(3) KDO + CTP → CMP-KDO + *PP*i (EC 2.7.7.38)

The genes for CMP-KDO synthetase, which catalyses the third reaction, and KDO-8P-synthase have been identified in some land plants (Royo et al. 2000a, Delmas et al. 2003, Matsuura et al. 2003) and since KDO metabolism enzymes have also been found in *Scherffelia*, *Mesostigma* (Becker et al. 2001, Simon et al. 2006), it is therefore reasonable to believe that the entire pathway has been conserved and is present in algae as well. The fact that the genes coding for the enzymes of KDO metabolism are found in plants, algae and bacteria but not in mammalian cells is causing increased interest in antibacterial production (Goldman et al. 1987, Xu et al. 2003) but also in molecular evolution (Royo et

al. 2000b). Interestingly, at the moment KDO is solely known to be a component of extracellular matrices. The KDO metabolism and cell wall integration in plants may well have been inherited from their scale-possessing algal ancestor, while the diverging chlorophytes lost it.

3 Characterization of scales-associated glycoproteins (SAPs)

The protein content of scales of different algae, as presented in Table 1, was first identified by standard protein determination methods (Kochert 1978, Neuhoff et al. 1979). It is only in *T. striata* and *S. dubia* that these proteins were characterized further (Becker and Melkonian 1992, Becker et al. 1996). SDS-PAGE of isolated scales revealed around 20 protein bands of rather high molecular weight (ranging from 77 to 300 kDa) (Fig. 3). These proteins were termed scale-associated proteins (SAPs) and named according to their apparent molecular weight. In both algae almost all proteins were found to be associated with the pentagonal scale fraction, i.e. the scales which are the closest to the plasma membrane. A few proteins seem to be associated with hair scales, as well as with the knotted scales of *T. striata* (Becker et al. 1990). The man scales, on the other hand, are totally devoid of proteins. Although the flagellar scales of *Tetraselmis* and *Scherffelia* look highly similar, the electrophoretic patterns of scale fractions were not identical, suggesting that the SAPs from these algae differ (Becker and Melkonian 1992, Becker et al. 1996). This conclusion was further corroborated by protein glycosylation analyses which emphasized further differences between *Tetraselmis* and *Scherffelia*'s SAPs. Glycosylation is common among extracellular proteins and all scale-associated proteins harbored N-glycosylated carbohydrate side chains. *Tetraselmis'* SAPs only showed N-glycosylation of the high-mannose type, while *Scherffelia*'s SAPs possessed N-glycans of the complex, hybrid and high-mannose types (Becker and Melkonian 1992, Becker et al. 1996). The existence of O-linked glycans in both organisms has not been reported.

In *Scherffelia dubia*, the majority of SAPs were found to be crosslinked together by disulfide bridges forming very high molecular weight complexes (more than 2.10^6 kDa, Fig. 3). Two of these complexes contain almost all SAPs. However, since they have a very similar composition it is possible that one of these complexes is actually a purification artefact. A third complex of 190 kDa is formed by two proteins of 98 and 96 kDa. Five other SAP do not form any complex. Several lines of evidence suggest that the high molecular weight complexes are connected to the side of the pentagonal scales facing the plasma membrane, thus linking the flagella membrane and the inner layer of scales. Firstly, protein complexes can be isolated and pelleted down with 100 mM EDTA (Becker et al. 1996) or solubilized from p-scales with 8 mM urea(Becker and Melkonian 1992) without disrupting the outer ring ; secondly, pentagonal scales show staining on one of their side when fixed with tannic acid (as does the region which separates the pentagonal scales and the flagellar membrane) (Becker et al. 1996) ; thirdly, an antibody raised against

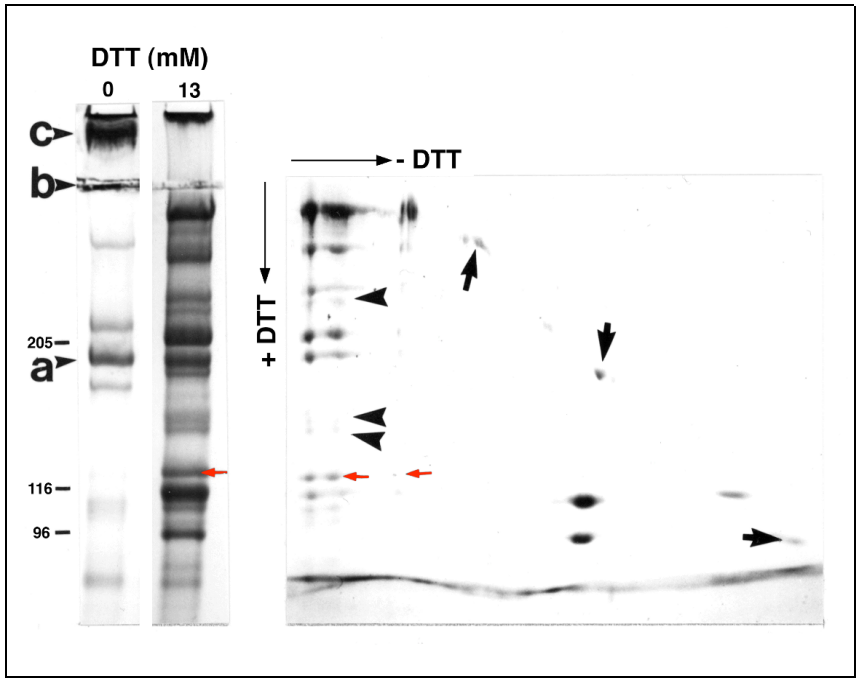

Figure 3. **SDS-PAGE analysis of SAPs**. SAPs from a crude scale fraction of flagellar scales of *S. dubia* were separated by one-dimensional and two dimensional SDS-PAGE, in non-reducing (0 mM DTT) and reducing (13 mM DTT) conditions. a, b and c: SAP complexes, arrowheads: SAPs which belong to complex c only, arrows: SAPs that do not form any complex, small arrow: SAP126. Gel image taken from Becker et al. 1996.

a deglycosylated SAP of 126 kDa present in the high molecular weight complexes labelled the 40 nm region between the pentagonal scales undelayer and the flagellar membrane (Becker et al. 1996). The same antibody also prominently labeled the plasma membrane suggesting the presence of a large pool of SAP there.

The interactions of SAP with scales were also investigated (Becker et al. 1990, Becker and Melkonian 1992). Protease treatment was shown to disrupt the shape of pentagonal scales, releasing the central subunit and opening the ring of outer subunits. These results suggest a role of SAPs in the connection of the rim subunits and of the central subunit with the ring of outer subunits (Becker and Melkonian 1992). Calcium is known to maintain the central subunits in place by ionic interactions and glycoproteins may mediate these interactions. A SAP of 210 kDa which was found to be resistant to protease treatment is therefore a possible candidate for this role. Another SAP of about the same molecular weight was found in *S. dubia* that does not associate in any complex. Since scale subunit integrity was not affected by protease or EGTA treatment, these experiments also indicate that SAPs have no influence on the scales subunits structure, which consequently seem to be dependent on polysaccharide structure only. It would have been interesting to know if any protein remained associated with the outer ring of the pentagonal scales after separation of SAP high molecular weight complexes by EGTA treatment, but this was not investigated.

So far, the precise function of SAPs remain unknown. Structurally, scale associated proteins seem to mediate scale-membrane, scale-subunit and probably scale-scale interactions. They are not involved in maintaining the structure of scale subunits and may therefore not play a role during scale synthesis, but could be involved in maintaining pentagonal scale ultrastructure, organizing scale layers, or connecting hair scales with the

plasma membrane and the axoneme. Other putative roles include sorting of flagellar scales in the trans-Golgi network or specifying flagellar membrane during mitosis (Wustman et al. 2004).

In order to identify the SAPs, individual SAPs were isolated by two-dimensional electrophoresis and analyzed by mass spectrometry (see Results part). The identification and characterization of SAPs was the aim of my project.

4 Flagellar regeneration and scale synthesis in *Scherffelia dubia*

Deflagellation and flagellar regeneration are widely observed in flagellated eukaryotes. These processes use conserved mechanisms. It is well known that flagellar regeneration induces the transcription and translation of a set of genes encoding the proteins which takes part in the synthesis of the new flagella (Keller et al. 1984). Upregulates genes includes axonemal components such as tubulins and dyneins, IFT proteins, and several chaperones (Schloss et al. 1984, Stolc et al. 2005). Gene upregulation is controlled by a deflagellation response element which was found in an intron of the 5'UTR of a flagellar dynein intermediate chain gene (IC70 encoded by the oda6 gene) of *Chlamydomonas* (Lefebvre et al. 1980, Schloss et al. 1984, Kang et al. 1998, Cheshire and Keller 1991, Cheshire et al. 1994). Recently, the availability of the genome sequence of *Chlamydomonas reinhardtii* allowed large scale genomic and proteomic studies to be performed such as the characterization of the proteome of the *Chlamydomonas* flagellum (Pazour et al. 2005), whole-genome transcriptional profiling during flagellar regeneration (Stolc et al. 2005), or the identification of conserved genes in organisms with flagella or cilia, namely the flagellar and basal body proteome (Li et al. 2004, Keller et al. 2005), leading to the identification of genes for human cilia-related diseases (Pan et al. 2005).

In *S. dubia*, experimental amputation of the flagella triggers an immediate response of flagellar regeneration accompanied by the synthesis of a new flagellar scale covering (McFadden and Melkonian 1986). The four flagella regenerate over a period of 180-210 min during which flagellar scales are continuously synthesized in the Golgi apparatus, exocytosed in the region of the flagellar groove and assembled onto the nascent flagella. Therefore, flagellar regeneration in *S. dubia* provides a model system to study flagellar scale biogenesis and endomembrane system flow.

The synthesis of flagellar scales is activated as early as 5 minutes after deflagellation, as observed by the gradual increase of scales in the Golgi cisternae, from the third cis-most cisterna to the trans-most cisterna (Becker et al. 1995), and by the reduction of the number of cisternae per dictyosome from 18-20 down to 17. Scales proceed towards the trans-face of the dictyosomes by cisternal progression (McFadden and Melkonian 1986, Becker et al. 1995) with a particularly fast turn-over rate of one cisterna every 12 seconds in average (McFadden and Melkonian 1986). Thirty minutes after flagellar regeneration, the rate of

scale synthesis is maximal (with around 30-60 scales per cisterna). Scales are eventually released from the Golgi apparatus in large vesicles. Exocytosis takes place at the flagellar groove where scales immediately assemble on the growing flagella. During the 3-4 hours of flagellar regeneration, around 54 000 flagellar scales are synthesized. The activation and specialization of the Golgi apparatus to synthesize flagellar scales strongly suggests the induction of specific genes.

Flagellar regeneration is dependent on *de novo* protein synthesis and N-glycosylation but surprisingly, most SAP are not synthesized during flagellar regeneration as revealed by pulse-chase ^{35}S metabolic labelling (Perasso et al. 2000). The characteristic tubulin bands were also not labelled during these experiments. Immunolocalization of SAP126, which belongs to the high molecular weight complexes linking pentagonal scales to the flagellar membrane, indicates that this protein could rather be recruited from a pool at the plasma membrane by endocytosis, and associated to scales in the secretory vesicles of the trans-Golgi network (Perasso et al. 2000). Still, SAPs are synthesized during a cell cycle (Perasso et al. 2000) and, as suggested by the authors, this apparent uncoupling of SAP and flagellar scale synthesis may be required for the correct formation of scales. In that case, SAP could be synthesized during interphase when only few scales are synthesized, or in a yet unidentified step during mitosis (see below).

The process of scale formation remains a mystery since scales always appear structurally complete. According to the proposed model of KDO incorporation in the higher plant cell wall (Royo et al. 2000a), KDO must be imported into the Golgi apparatus and activated by CMP-KDO synthase before it is used to synthesize the scales. KDO synthesis, import and activation must certainly be regulated with scale synthesis. The three scale types are probably simultaneously synthesized as they all occur in a single cisterna. Membrane proteins or lipids could be involved in pentagonal scales formation since this scale type is always found close to the cisternal membrane. Man scales on the other hand float in the center of cisterna while hair scales are found in all parts of the cisternae, although they could be linked to the membrane by one of their extremities. Scales remain as individual units until they exocytose indicating that linkage-inducing components are either missing or are not activated in the endomembrane compartments. Obviously the regulation of scale synthesis during flagellar regeneration must be coordinated with flagellar synthesis, and the processes which control flagellar length may be involved in regulating scale synthesis as well.

Scale synthesis normally occurs only once during the cell cycle of *S. dubia*, during mitosis (McFadden et al. 1986, Wustman et al. 2004). The regulation of scale synthesis is then much more complex. Cell division occurs inside the parental theca in a non-flagellate state. Before the flagella are even shed, a large pool of flagellar scales is already synthesized. This pool is deposited on the anterior flagellar membrane above the basal bodies after the flagella are shed. During mitosis, the cell synthesizes a new theca, which interrupts flagellar scale synthesis. The Golgi apparatus also functions differently during thecal scale

synthesis: the number of cisternae per Golgi stack is reduced to 10-14, the cisternal turn-over rate is about one every four minutes (twenty times slower than during flagellar scale synthesis), each cisterna contains ca 12 000 thecal scales and entire cisternae are exocytosed at specific sites of the plasma membrane. When daughter cells have nearly completed their theca, axoneme assembly begins while thecal scales are still synthesized. Each growing flagellum is covered by flagellar scales, presumably from the pool synthesized just before mitosis. When the theca is completed, flagellar scale synthesis resumes. During this second flagellar scales synthesis period, hair scales are apparently not synthesized.

Given the complex regulation of scale synthesis during mitosis, there might be a yet unnoticed step during which the SAP complexes are produced and stored for use in the later phases of flagellar scales synthesis. So far, it is not known by which steps SAPs associate to scales. Scale synthesis has not revelealed its mysteries.

5 Objectives

Flagellar scale coverings are the results of several coordinated processes which involve synthesis of abundant complex polysaccharides, endocytosis of a pool of scale associated proteins, and their coordinated assembly and exocytosis on the plasma membrane of a growing flagella. In order to understand further scale biogenesis and assembly, it is necessary to identify its various molecular components. The scale associated proteins are obvious candidates as they are so far unidentified and uncharacterized. These are the objectives of this thesis.

Experimental methods

1 Cell culture

Axenic cultures of *Scherffelia dubia* (Perty) Pascher emend. Melkonian et Preisig (CCAC 0019) were grown photoautotrophically in Waris-H medium (WEES modified by supplementing 1/10th earth extract [Melkonian and Preisig 1986]), at 15°C, under a 70 $m^2.sec^{-1}$ light, with a 14:10 h light:dark cycle. Cells were maintained by transfer into sterile Erlenmeyer flasks containing new autoclaved medium every two weeks.

For DNA and RNA extraction, cells were transfered into 400 mL Waris-H medium inside a 1 L flask, and aerated (0,5 L/min) under gentle agitation. Sterile conditions were observed at all times.

2 Mass spectrometry sequencing

2.1 Isolation of flagellar scales

Flagellar scales were isolated as described in Becker et al. 1996. A one week old 400 mL culture of *Scherffelia dubia* was used to grow 6x10 L culture. After a week, the 60 L culture was concentrated to 1-1,5 L with a Pellicon system (Millipore). Cells were then pelleted 15 min at 700 g (RC-5B, Sorvall). All the following steps were performed on ice. The cells were resuspended in 50 mL Waris-H. Flagella were shed by a salt shock by the additon of 50 mL of 2xTBS (TBS: 150 mM NaCl, 50 mM Tris/HCl, pH 7,5). Cell bodies were pelleted 10 min at 200g (Laborfuge Ae, Heraus), and the flagella were recovered from the supernatent by 20 min centrifugation at 5 000g (RC 28S, Sorvall). Flagella were resuspended in 1 ml 1xTBS to which 70 μL/mL Triton X114 was added. The sample was vortexed and left on ice for 1 h, or overnight at 4°C. Insoluble material was pelleted 15 min at 10 000 rpm in a refrigerated tabletop centrifuge (Sigma 1K15) and incubated 5 min at 37°C. Water phase and detergent phase were separated by 5 min centrifugation in tabletop centrifuge. The upper water phase was extracted once more with Triton X114. The water phase was then centrifuged at maximum speed (14 500 rpm) to pellet any remaining insoluble material and flagellar scales were recovered by 30 min ultracentrifugation at 54 000 rpm (RCM100, Sorvall). The yield of protein was generally about 50 μg per 60 L of culture.

2.2 Isolation and sequencing of SAPs

SAP212, SAP126, SAP116 and SAP98 were isolated by 2D SDS-PAGE in reducing conditions by Dr. Becker (Becker et al. 1996). Trypsin-digested peptides were kindly sequenced by time of flight tandem-MS by Dr. Conradt at the Geselltschaft für Biologie Forschung (Braunschweig, Germany).

SAP98 and SAP96 purification were performed as described (Becker et al. 1996). Isolated flagellar scales were heated 3 min at 100°C in SDS-PAGE sample buffer and separated in a 8-16% SDS-PAGE gel (Anamed). The band corresponding to the SAP98-SAP96 complex was electroeluted. 5% (vol:vol) mercaptoethanol was added to separate the proteins. The sample was run in a second 8-16% SDS-gel. Gel pieces containing SAP98 and SAP96 were cut from the gel and sequenced by LC-MS at the ZMMK service labor (Cologne, Germany).

Peptides identified as trypsin or keratin by MASCOT search were considered as contamination and discarded.

3 Deflagellation and flagellar regeneration

S. dubia cells were harvested by centrifugation at room temperature (200 g, Hettich ROTIA/X) and the cell pellet was immediately resuspended in the desired volume of fresh Waris-H medium. To verify that centrifugation did not induce deflagellation, cells were left at 15°C for 15 min and flagella maintainance was checked by phase-contrast light microscopy (after fixeing with Lugol).

Mechanical-induced deflagellation was obtained by homogenization of 25 mL samples on ice using a glass homogenizer. After 80 strokes (within less than 5 minutes), more than 95% of the cells had shed all four flagella as assessed by light microscopy. The remaining flagellated cells beared 1-4 flagella which sometimes appeared broken.

Flagellar regeneration was allowed by agitating cells in a Petri dish or in a 15 mL polyropylene tubes, at 20°C in (or over) a thermal block. More than 95% of the cells regrew fully functionnal flagella within 3-4 hours.

4 Nucleic acid isolation

4.1 Genomic DNA isolation

Genomic DNA was extracted from 400 mL *S. dubia* cultures grown for 8-14 days (cell density 2-5.10^6 cell/mL). Cells were harvested by centrifugation and grinded into fine

powder using mortar and pestle in liquid nitrogen. The cell powder was processed with lysis buffer from one of the following kits or protocols: DNEasy Plant Minikit (Qiagen), Wizard genomic DNA purification kit (Promega) or hot CTAB plant DNA extraction protocol (Murray and Thompson 1980). The isolated genomic DNA was stored in TE buffer at -20°C, at a concentration ~100 ng/µL. DNA quality was examined by monitoring absorbance at 260 and 280 nm in a UV spectrophotometer (Eppendorf) and ethidium bromide visualization after agarose gel electrophoresis.

1.1 RNA extraction

Total RNA was extracted from 400 mL cultures of *S. dubia* in exponential growth phase (~ 7-10 days old, 0,7-2,5.10^6 cell/mL), before or after mechanical deflagellation, using Trizol reagent (Invitrogen). Cells were pelleted at room temperature in a Hettich ROTIA/X centrifuge at 2000 rpm for 5 min, immediately resuspended in 2 mL Trizol reagent, and grinded into fine powder using mortar and pestle in liquid nitrogen. The cell powder was then collected into a 50 mL polypropylene tube and allowed to thaw at room temperature. The manufacturer's protocol was followed from this point on. Briefly, the Trizol-cell debris mixture was mixed and left at room temperature for at least 5 min. 400 µL chloroform was then added. The solution was vortexed, transfered into 2 mL tubes and left 10 min at room temperature. Aqueous and organic phases were separated by 15 minutes centrifugation at 12000 g at 4°C (Sigma 1K15). The aqueous, RNA-containing phase was recovered and total RNA was precipitated with 1 mL 2-propanol at room temperature. After 10 min centrifugation at 12000 g at 4°C, and a 70% ethanol wash, total RNA was finally resuspended in DEPC-H_2O. Purity and integrity were analysed by UV spectroscopy and agarose gel electrophoresis. Absorbance ratios for A_{260}/A_{280} were ~1,85 ; the expected theoretical ratio between 2,0-2,2 for pure RNA was never obtained (Sambrook et al. 1989). Nevertheless, total RNA could be used for reverse transcription and Northern blot successfully.

4.2 First strand cDNA synthesis

First strand cDNA used in RT-PCR was synthesized by oligodT-primed reverse transcrition performed either with Superscript II first-strand synthesis kit (Invitrogen) or Mu-Mlv reverse transcriptase (New England Biolabs), according to the manufacturers' instructions. 200 ng of total RNA were incubated 50 min with 25 units of Mu-Mlv RT or Superscript II, at 37°C or 42°C respectively in the presence of RNase inhibitors. The reaction was then stopped by heat-inactivation and diluted ten times.

4.3 RACE-compatible cDNA synthesis

2,6 µg of total RNA extracted as described above from *Scherffelia dubia* cells regenerating

their flagella for 30 minutes were used to synthesize RACE-compatible cDNA with the GENERACER kit (Invitrogen) and Superscript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions. The reverse transcription was performed at 50°C for 45 min, in the presence of Rnase H, and using oligodT primers.

5 DNA amplification and sequencing

5.1 Polymerase chain reaction

PCR were run in an MWG Biotech thermal cycler. Gradient PCR were run in an Eppendorf Mastercycler. 25 μL reactions were set up on ice, first by preparing a master mix with all common components, then adding specific components to each PCR tubes. The lid and thermal block were pre-heated before samples were put inside the PCR cycler for semi-hot starts. Unless otherwise specified, Taq polymerase (New England Biolabs) was used with the following standard reaction composition:

Components	Platinium Pfx polymerase	Taq polymerase
Forward primer (10 pmol/µl)	1 µl	
Reverse primer (10 pmol/µl)	1 µl	
DNA template	DNA : 50-150 ng, cDNA : ~25 ng	
Reaction buffer	2,5 µl (10X)	
Polymerase	0,25 µl (2,5 U/µl)	0,3 µl (1,25 U/µl)
dNTP (10 mM)	0,75 µl	0,5 µl
$MgSO_4$ (50 mM)	0,5 µl	-
dH_2O	To 25 µl	To 25 µl

5.2 Degenerate oligonucteotide primed PCR

PCR with standard degenerate primers were performed using Thermozyme polymerase (Invitrogen) in a gradient PCR cycler to test several annealing temperature spanning down to 10°C below the melting temperature of the primers. PCR were generally peformed with 2 min-94°C, followed by 30 cycles [30 sec-94°C, 30 sec-annealing temperature, 3 min-72°C] and a final extension step of 5 min–94°C. Products were visualized with Et-Br staining on agarose gels.

Components	Thermozyme polymerase
Forward primer (10 pmol/µl)	2,5 µl
Reverse primer (10 pmol/µl)	2,5 µl
DNA template	DNA : 50-150 ng, cDNA : 0,5-1 µl
Reaction buffer	5 µl (5X)
Polymerase	1 µl (1 U)
dNTP (10 mM)	0,6 µl
dH_2O	To 25 µl

5.3 Amplification with short degenerate primers

To amplify DNA fragments using short (12-mer) degenerated primers, Taq beads were used (Amersham) to obtain real hot-start conditions. PCR were performed in a gradient PCR cycler (Eppendorf Mastercycler) to find the optimal annealing temperature, with only two steps (denaturation-annealing). The settings were as follow: initial denaturation 94°C-3 min followed by 30 cycles of [denaturation step 94°C-15 sec, annealing step 46°C±5°C-20 sec].

DNA fragments were selected by visualization in agarose gel with ethidium bromide (care was taken to expose the DNA as little as possible) and either ligated with the vector in-gel using low-melting agarose (Promega), or isolated and purified by agarose gel extraction with QIAEXII beads (Qiagen). The DNA fragments were cloned into pGEM-T-easy vector (Promega) and sequenced.

Components	Taq beads
Forward primer (50 pmol/µl)	1 µl
Reverse primer (50 pmol/µl)	1 µl
DNA template	DNA : 50-150 ng, cDNA : 0,5-1 µl
dH_2O	To 25 µl

1.1 Inverse PCR for SAP98

Genomic DNA was prepared for inverse PCR as follow: 400 ng genomic DNA was completly digested in a 10 µL reaction volume (1 h at the optimal temperature) with 10 U of one of the following restriction enzymes: *Rsa*I, *Sau*3AI, *Eco*RI, *Bam*HI, *Bgl*I. The restriction enzymes were heat-inactivated and the digested DNA was circularized by ligation with T4 DNA ligase (2,5 U, Gibco) after a 10X dilution (200 µL reaction volume). The ligase was heat-inactivated and the circularized DNA was purified with ethanol (Sambrook et al. 1989) and resuspended in 20 µL dH_2O (~20ng/µL).

The circularized DNA was amplified using *Taq* polymerase (NEB) with the following primers were used: iPCR forward: 5'-GACCAGGGSTTTAACTAC-3' , iPCR reverse: 5'-CAGAGATATGTCGTCRGC-3' (Fig. 8B, p 37). PCR settings were: 94°C-3', 35x(94°C-20 sec, 51°C-30 sec, 75°C-270 sec)**.** The PCR fragments were purified with a PCR purification kit (SV kit, Promega) and ligated in the pGem-T-Easy vector (Promega) before sequencing.

Components	Taq polymerase (NEB)
Forward primer (10 pmol/µl)	1 µl
Reverse primer (10 pmol/µl)	1 µl
DNA template (20 ng/µl)	5 µl
10X reaction buffer	2,5 µl
Polymerase (1,25 U/µl)	0,25 µl
dNTP (10 mM)	1 µl
dH_2O	To 25 µl

5.4 RACE PCR for SAP98

RACE PCR was performed according to the GENERACER kit (Invitrogen). RACE-compatible cDNA (see 4.4) was amplified with the generacer primers and the following gene specific primer (GSP):

- forward GSP: 5'-ACRGCRGACGACATATCTCTGGACCAGGGSTTTAACTAC-3'

The optimized PCR settings were: 94°C-3 min, 30x[94°C-20 sec, 65°C-20 sec, 75°C-210 sec], 75°C-5 min. The 3'-RACE product was purified by gel extraction (Wizard SV, Promega), ligated overnight into the pGem-T-easy vector (Promega), and transformed in DH10β electrocompetent cells.

Components	3'RACE
3'Generacer primer (10 pmol/µl)	0,9 µl
Forward GSP (10 pmol/µl)	0,9 µl
RACE cDNA	0,9 µl
10X reaction buffer	2 µl
Polymerase (Taq)	0,25 µl
dNTP (10 mM)	0,4 µl
dH_2O	To 20 µl

5.5 Semi quantitative RT-PCR

a) Standard used

Rbcs and actin were used as internal controls. Actin expression was shown not to be upregulated during flagella regeneration in *Chlamydomonas* (Pazour et al. 2005, Stolc et al. 2005). Rbcs mRNA abundance in *Scherffelia* was verified to be invariant by Northern hybridization and by competitive RT-PCR (Wang et al. 1989, Gilliland et al. 1990) as follow.

Northern hybridization was performed as described (Sambrook et al. 1989). 10 µg total RNA from interphase and flagella regenerating cells were denatured by mixing 1:1 and incubating 15 min at 65°C with a denaturation mix (800 µl formamide, 240 µl formaldehyde, 400 µl 5X TBE running buffer, 10 µl Et-Br) and separated by formamide-agarose gel electrophoresis. Transfer onto nylon membrane (Hybond-NX, Amersham) was performed by capillary transfer as described in Sambrook et al. (1989). The RNA was then crosslinked to the membrane by illumination under UV light (70000 $j.cm^{-2}$, Crosslinker, Stratagene). The blot was hybridized with a 700 bp rbcs probe prepared by amplification of a plasmid insert and labelled by random primers extension with αP^{32}-dATP (3000 Ci/mmol) using Klenow fragment (Gibco). The amount of probe hybdridization was determined by phosphor screen densitometry scanning (Molecular Dynamics) and the ImageQuant software (Molecular Dynamics). This experiment was done once and showed a 5% variation between the two samples.

For competitive RT-PCR, a competitor for rbcs amplification was synthesized as described (Celi et al. 1993) with the rbcs forward primer and rbcs competitor reverse primer. The rbcs competitor was 82 bp shorter than the target and could be easely distinguished from the rbcs . This experiment was done once and showed a 3,9 % variation of the transcripts abundance between interphase and flagellar regeneration.

Standard	Forward primer	Reverse primer
rbcs	CTACGTGATCAAGAACGGCTGG	TCAGGACCTGGCTGGCATC
Rbcs competitor	-	TCAGGACCTGGCTGGCATCCGAACACGTCCTGGCAGTAA
Actin	GCAGCCAAAACGAGAAGAAC	ACATCTGCTGGAAGGTGGAC

The primers chosen to amplify the rbcs mRNA by RT-PCR are separated by a 110 bp intron present in the sequence of the rbcs gene of *Scherffelia dubia* (Simon 2002) so that amplification from genomic DNA can be distinguished from amplification from mRNA.

b) Duplex PCR

The gene of interest and an internal standard (rbcs or actin) were amplified

simultaneously in the same PCR tube. The primers were chosen to amplify a product no longer than 400 bp and of a different length than the standards. The reactions were prepared as described in the table below. The PCR settings were: 94°C-3 min and [94°C-20 sec, 52°C-25 sec, 75°C-40 sec] for the desired number of cycles. A sample (7 µL) of the reaction was removed at three different cycles after the extention step while the temperature of the block was maintained at 75°C. 6 µL of each sample were loaded on a 2% agarose gel soaked with 50 ng/L ethidium bromide. Each experiment was repeated two or three times.

Target gene	Forward primer	Reverse primer
Tubulin A	CGTGAGGTTATCAGCATCCA	AGGCACAGGTCCACAATCTC
Tubulin B	ACTGAGGGTGCTGAGCTGAT	TGTCGTACAGAGCCTCGTTG
Dynein light chain (16 kD)	GCTCAAAGGGCCTAACAGTG	CAGCCTCCCTCTCCTTCTTT
14-3-3 protein	CTGGAAACGAGCGTAAGGAG	TCAAACTGACCAAGCACTCG
Heat Shock Protein 90	TGCGTGTGACGTAGTCCTTC	GGAATTCAAGTCGGTGCAGT
Surface antigen P2	ACGAGGGATGTCTTGAATGG	TCCCAGTAATGTGGCTGTCA
Flagella-associated protein 115 (FAP115)	CTGAAGGGAGGCTGACAGAG	TGTCACCTTCCAGCAGTTCA
SAP98	GCTTCGTATCCATCCGTCAT	GCATTCCATCGTAAGCACTG

Components	Taq
Forward primer A (10 pmol/µl)	1 µl
Reverse primer A (10 pmol/µl)	1 µl
Forward primer B (10 pmol/µl)	1 µl
Reverse primer B (10 pmol/µl)	1 µl
DNA template	DNA : 50-150 ng, cDNA : 0,5-1 µl
Reaction buffer	2,5 µl (10X)
Taq polymerase (NEB)	0,3 µl (0,4 U)
dNTP (10 mM)	0,5 µl
dH_2O	To 25 µl

c) Image analysis

Gels were photographed with a software assisted-CCD camera (Intas GDS Cosmicar, Intas). The intensity of the DNA bands was measured with the public domain software ImageJ 1.33u (NIH). An automatic background subtraction process was run (rolling ball method, radius: 50) and the area under the curve was calculated using the gel macro provided with the software. The linearity of the system was verified for amounts comprised between 10 ng and 1 µg of DNA.

5.6 Sequencing

PCR products were either purified from the reaction components with a PCR purification kit (PCR clean up sytem, Promega) or gel extracted (Wizard SV gel, Promega) and resuspended in dH_2O. The purified products were cloned into the pGEMT-easy vector (Promega) using the standard manufacturer's protocol and transformed in TOP10 bacteria (Invitrogen) by electroporation (1 mm cuvettes, 1,9 kV/cm, Genepulser II, Biorad). Plasmid DNA was isolated with a standard miniprepkit and inserts were sequenced using Big Dye v1.1 (ABI) and an ABI3730 automated sequencer.

6 Subtracted cDNA-library construction and sequencing

Total RNA was prepared in our laboratory, and the cDNA subtracted library was generated by Evrogen (Moscow, Russia).

Two 400 mL, 9-11 days old cultures of *S. dubia* in exponential growth phase, containing >96% of flagellated cells, were mixed and concentrated in 50 mL Waris-H medium by centrifugation at 200 g (Hettich ROTIA/X) at the beginning of the light cycle before the cells enter division (Wustman et al. 2004). The concentrated sample was separated in two samples of equal volume (25 mL).

The first sample was kept on ice for the length of the deflagellation process and is referred to as Interphase cells (I). Cells of the second sample were subjected to mechanically-induced deflagellation as described above. Deflagellated cells were then divided in two sub-samples: 10 mL were processed immediately after deflagellation for total RNA extraction. The remaining 15 mL were allowed to regenerate their flagella at 20°C with agitation for 30 min. 10 mL were then removed and processed for total RNA extraction. The remaining 5 mL were left at 20°C to control that flagella regeneration proceeded correctly. After 3 hours, >95% of the cells regenerated fully functionnal flagella.

Total RNA was extracted from 20 mL of interphase cells, and 10 mL of each deflagellated sample as described above. Concentration of RNA was checked spectroscopically and the quality of the preparation was controlled by formaldehyde-agarose gel electrophoresis and RT-PCR. Equal amounts of flagella-regenerating samples were mixed together to form the FR sample (Flagella Regeneration sample). The I and FR samples were sent in ethanol-precipitated form to the Evrogen company (Moscow, Russia) which carried out full-length cDNA synthesis using the SMART technique (Zhu et al. 2001), suppression subtraction hybridization (SSH) and mirror oriented selection (MOS) (Diatchenko et al. 1996, Diatchenko et al. 1999, Rebrikov et al. 2000).

The following briefly describes the different steps performed at Evrogen (Fig. 4). The full length first strand cDNA was synthesized using SMART Oligo II oligonucleotide (5'-AAGCAGTGGTATCAACGCAGAGTACGCrGrGrG-3') and CDS primer

(5'-AAGCAGTGGTATCAACGCAGAGTAC-d$(T)_{30}$-3'), using 0.3 µg of each total RNA sample. The FR and I first strand cDNA diluted samples were then amplified by 21 PCR cycles with the SMART PCR primer (5'-AAGCAGTGGTATCAACGCAGAGT-3'). The number of PCR cycles has been optimized to preserve cDNA representation. The SMART-amplified cDNA were then digested with *Rsa*I. Two tester populations are created by adapting different suppresion adapters (adapters 1: 5'-ACTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT-3' + 3'-GGCCCGTCCA-5' ; and 2R:

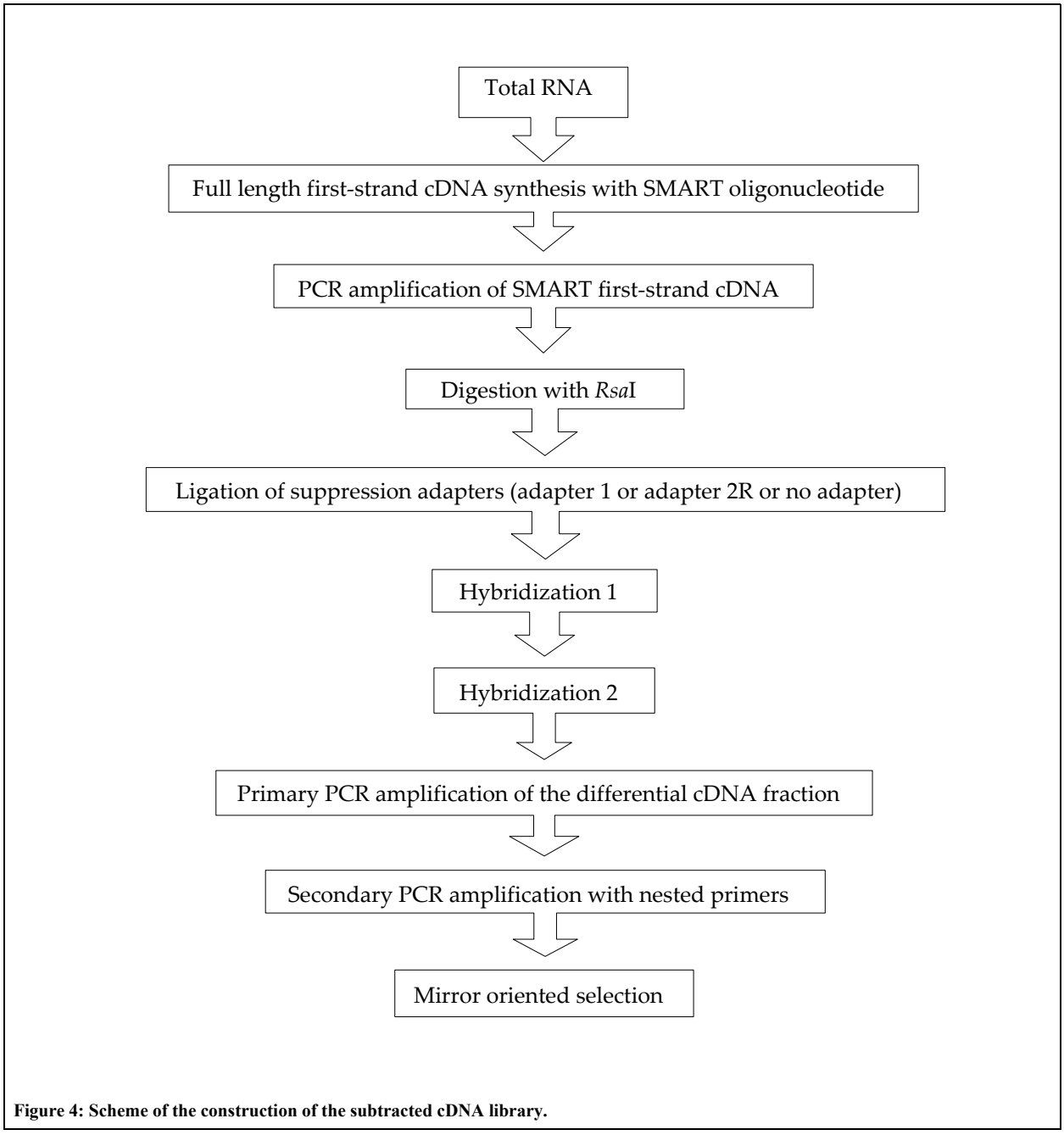

Figure 4: Scheme of the construction of the subtracted cDNA library.

5'-CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGT-3' + 3'-GCCGGCTCCA-5'). The tester populations were mixed with 30X excess driver (no adapter) in seperate tubes, denatured and allowed to renature. After the first hybridization, the two samples were mixed and hybridized together. Subtracted cDNA was then amplified by primary (26 PCR cycles with primer 1: 5'-CTAATACGACTCACTATAGGGC-3') and secondary (nested) PCR (10 PCR cycles with nested primers NP1: 5'-TCGAGCGGCCGCCCGGGCAGGT-3' and NP2R: 5'-AGCGTGGTCGCGGCCGAGGT-3'). The MOS procedure was then performed to eliminate background molecules according to Rebrikov et al. (2000). Adapter 1 was removed by *Xma*I digestion, the samples were denatured and allowed to renature. The duplex DNA were extended and immediately submitted to 20 PCR cycles with primer NP2Rs (5'-GGTCGCGGCCGAGGT-3').

7 Hybridization

7.1 Differential screen

The resulting flagella regeneration-specific cDNA library (FR-I) was cloned in our laboratory in a plasmid vector (pGem-T-easy, Promega) and transformed into TOP10 *E. coli* cells (Invitrogen). Ten colonies were analysed by colony-PCR, each showing an amplification product. The shortest insert had a length of ~350 bp, and the longest insert was ~1.1 kb. The average insert length was 700 bp. 556 clones were randomly selected and plasmid DNA were isolated (Montage kit, Millipore) and sequenced once in each direction using standard sequencing primers and Big Dye sequencing (ABI).

For differential screening, plasmid DNA was denatured under alcaline conditions: 0.5 μL plasmid DNA (100-300 ng.μL^{-1}) were incubated at room temperature for 15 min with 4.5 μ L dH_20 and 5 μL NaOH 0.5 M, and further incubated 5 min at 37°C. The denatured plasmids were dot-blotted in 96 arrays by transfering eight times 1 μL (5-15 ng plasmid DNA) on eight different 12 cm x 8 cm nylon membranes (Hybond-NX, Amersham), neutralized 2 min in 0.5 M Tris-HCl (pH 7.5), and washed 2-4 min with dH_2O. Plasmid DNA was finally crosslinked to the membrane by illumination under UV light (70 000 $j.cm^{-2}$, Stratalinker 1800, Stratagene).

30 ng of subtracted probes (cDNA obtained after SSH subtraction) and 50 ng of unsubtracted probes (SMART amplified first-strand cDNA) were labelled by random priming with αP^{32}-dATP (3000 Ci/mmol) and purified through spin column centrifugation. The specific activities of the probes were determined by TCA precipitation.

Probes	Specific Activity
Flagella regeneration-specific probe (= forward subtracted probe)	1.5 10^7 cpm/μg
Interphase specific probe (= reverse subtracted probe)	1.7 10^7 cpm/μg
Flagella regeneration probe (= forward unsubtracted probe)	0.4 10^7 cpm/μg
Interphase probe (= reverse unsubtratced probe)	0.92 10^7 cpm/μg

Hybridization was performed as described in Sambrook et al. (1989). Each membrane was pre-hybridized 1 hour at 72°C in pre-hybrydization buffer (6xSSC, 0.5% SDS, 5x Denhardt's, 100 mg/mL herring's sperm DNA). The pre-hybridization buffer was exchanged with hybridization buffer (6xSSC, 0.5% SDS, 100 mg/mL herring's sperm DNA), and equal cpm counts for each pair of probes (subtracted and unsubtracted probes) were added to the membranes. The blots were washed with successive solutions of 2xSSC, 0.2xSSC and 0.1xSSC.

7.2 Dot blot

a) Probes synthesis

The β-tubulin, hsp90, rbcs and dynein light chain probes were prepared by PCR amplification of cloned cDNA inserts (Becker et al. 2001) with M13F and T7 sequencing primers, followed by vector's end removal by restriction digestion, and gel purification (GFX gel extraction kit, Amersham). αP^{32}-dATP (3000 Ci/mmol) labelling was done by random primers extension using Klenow fragment (Gibco).

b) Blotting method

Total RNA was diluted to 400 ng/µL and mixed 1:1 (vol:vol) with RNA loading buffer (Fermentas). After denaturion at 65°C for 10 min the RNA was blotted in triplicate onto a nylon membrane (Hybond-NX, Amersham) by depositing five 2 µL drops per dot (1 µg total RNA per µL). Denatured herring's sperm DNA (200 ng/µL) was dot-blotted as well (1 µg DNA per dot) as a negative control. RNA/DNA was finally crosslinked to the membrane. Hybridization and washing were performed as described above.

7.3 Image analysis

Blots were exposed on a phosphor screen and scanned with a densitometer (STORM 860, Molecular Dynamics). Images were analysed using the software ImageQuant 1.2 (Molecular Dynamics).

8 Bioinformatics analyses

8.1 SAP peptide similarity search

To obtain significant results, only peptides longer or equal to 10 amino acids were queried. Each possible sequence for ambigous peptides was queried against the non-redundant GenBank CDS, the *Scherffelia dubia* ,and *Mesostigma vriride* EST databases using BLASTP configured for short, nearly exact matches. The *S. dubia* and *M. vriride* EST were compiled into databases and searched using the GCG Wisconsin package (as implemented at the University of Cologne).

8.2 Analysis of EST clones

Putative cell wall homologs in the *S.dubia* EST database were identified using TBLASTX without the filter for low complexity against the SWISSPROT database. The reverse blast analysis was performed by comparing each cell wall protein that matched an EST clone to the *S. dubia* EST using the TBLASTN algorithm without the low complexity filter and limited by the "Scherffelia dubia[organism]" parameter. This parameter restricts the search to all *S. dubia* EST present in the databases.

8.3 Analysis of clones obtained with Suppression Subtraction Hybridization (SSH)

The sequences obtained from the clones isolated from the flagellar regeneration-specific library were compared with several genomes and EST libraries:

- *Chlamydomonas reinhadrtii*'s genome was searched using the 2nd version of the proteome available at the JGI.
- *Arabidopsis thaliana*'s genome was searched using the TAIR ressources.
- *Mesostigma viride* and *Scherffelia dubia* EST were searched using either a local database with the GCG wisconsin package, or the BLAST algorithm available at the ncbi limited with the "mesostigma viride[organism]" or "Scherffelia dubia[organism]" keys, respectively.

Results

A Isolation of the cDNA for a scale-associated protein of 98 kDa (SAP98)

1 Analysis of scale-associated proteins by sequence comparison

1.1 Peptide sequencing by mass spectrometry and similarity search

The SAPs were sequenced by mass spectrometry (MS) (Table 2). Protein sequencing for SAP212, SAP126, SAP116 and SAP98 was performed by Drs. Becker and Conradt at the Gesellschaft für Biologische Forschung (Braunschweig) after 2D gel electrophoresis isolation (Becker et al. 1996). I sequenced additional peptides for SAP98 and SAP96, two glycoproteins which assemble by disulfide crosslinks to form a 190 kDa complex which has not been localized yet. This 190 kDa complex was first isolated from a crude scale fraction by SDS-PAGE. The subunits were eventually separated by a second electrophoresis in reducing conditions (13 mM DTT). The peptide sequences obtained by LC-tandem-MS performed at the ZMMK (Cologne) are shown in italics in Table 2.

It should be noted that some amino acids in the sequence of most peptide are ambiguous. Three reasons explain why ambiguities exist, i) mass spectrometry does not distinguish between leucine and isoleucine, ii) experimental limits can lead to different possible interpretations of a peptide spectrum. For instance, a glycine-alanine sequence (G, M=57,021 ; A, M=71,037, G+A=128,058) can be misinterpreted as a glutamine (Q, M=128,058), or a hydroxyproline (Hyp) can be misidentified for a leucine/isoleucine (Hyp, M=113,047 ; L/I, M=113,084), iii) the quality of the spectrum. In the Cologne analysis, the peptides are assigned a probability that the interpretation of the sequence is correct to reflect these limits.

MASCOT searches for the SAP96 and SAP98 peptides sequenced in Cologne did not identify any homologous protein. We expanded the search by performing several sequence database similarity searches. To obtain reliable protein identifications, only peptides longer or equal to 10 amino acids were queried, and we requested that several peptides of a SAP matched the same protein. Peptides were compared to the public domain databases available at the ncbi. We used the blastp algorithm configured for short, nearly exact matches.

Each peptide retrieved a list with hundreds of matches apparently unrelated to each other (not shown). Some of the results were consistent with the hypothetical function,

localization or biochemical nature of the SAPs : for instance, one peptide of SAP212 matched a pistil extensin-like protein with 54% identity, another of SAP126 matched a putative pectinesterase with 58% identity. These proteins were however not abundant hits and I therefore concluded that these similarities were not significant.

SAP 212 kDa	VVV(L/I)DSFGQQS(L/I)EAK VVTPW(L/I)S AGAGPPVVANF(L/I)F
SAP 126 kDa	(L/I)AAGFYDVSYGA QSS(L/I)DGYNFFVR EV(L/I)AFGS(L/I)(L/I)SSTPK (L/I)PAFDPD
SAP 116 kDa	E(L/I)T(L/I)PSG (L/I)AFSG(L/I)TDGTHT(L/I)YVSAK QTGASA(L/I)F(L/I)(L/I)TR S(L/I)SPGTHY(L/I)FVK SVDV (L/I)T(L/I)PSPVGV(L/I)W(L/I)EK (L/I)TNPTNQVPVWVFFK AAYTNAAWA(L/I)STSSK EAGQYV(L/I)E(L/I)S(L/I)SDK SGATGVWE(L/I)N(L/I)
SAP 98 kDa	YN(L/I)DSTYQV(L/I)ATK (L/I)TADD(L/I)S(L/I)DQGFNY TYPW *GLLFKEATGANVDIFQEK* (p=1) *TTWDTAGSTTK* (p=0,98) *S(L/I)SVDNDDGR* (p=1) *TFTVYAVDR* (p=1) *PVYAVDR* (p=1) *HPE(L/I)(L/I)NCVKSAR* (p=1) *VVTVSLPR* (p=1) *YGTTCTSPAVVK* (p=0,91)
SAP 96 kDa	*S(L/I)SPGTHV(L/I)MVK* (p=1) *(L/I)SVS(L/I)VDR* (p=1) *VTFGCSEEPCTYK* (p=1)

Table 2. **SAP peptide sequences obtained by mass spectrometry**. Peptides in italic were sequenced in Cologne, other peptides were sequenced in Baunschweig. p : probability, when known, that the sequence is correct.

The peptide sequences were then compared with the translated EST database from interphase and flagellar regenerating cells of *Scherffelia dubia* (Becker et al. 2001). No match with significant similarity was found, indicating that the SAP peptides were not present in this library. SAP peptides were also compared to the *ca*. 3 300 unique contigs obtained for the scaly green alga *Mesostigma viride* (Simon et al. 2006). Again, no contig with significant similarity to at least two peptides of a SAP was identified.

I conclude from these analyses that SAPs are likely to be novel proteins.

1.2 Search for putative cell wall protein homologs among the *S. dubia* EST

The SAP peptides did not allow the identification of SAPs in the *Scherffelia* EST database. This database contained however 10 clones with known cell wall glycoproteins of land plants and green algae as best match (about 30% identity, E values between 10^{-5} and 10^{-14}). The majority of the contigs were similar to the *Chlamydomonas* glycoprotein 1 (GP1) present in the cell wall of the vegetative/gametic cell, and to the higher plant cell wall protein extensin. Two clones were similar to other eukaryotic extracellular glycoproteins: a sulfated surface glycoprotein (SSG185) from *Volvox* and a yeast cell wall mannoprotein (DAN4). To obtain these matches, it was necessary to remove the filter for sequences of low complexity during the blast analysis. The low similarity suggested that the *S. dubia* EST could be distant homologues of these proteins.

All these cell wall glycoproteins are characterized by regions of low complexities. SSG185 possesses a Pro-rich region between residues 210-320. DAN4 is a Ser- and Thr-rich protein with highly repetitive motifs. Extensins are rich in Hyp and Ser and exhibit $Ser(Pro)_4$ repeats and GP1 is divided into domains rich in Hyp and sequence repeats, and domains that lack repetitive motifs and have few Pro. Low complexity regions and highly repetitive motifs are common characters of cell wall proteins that are believed to organize the structural motifs of these proteins (Cassab 1998). Examining the EST clones and their alignments revealed that they actually all lacked these typical characters: they did not contain repetitive motifs or Pro/Thr/Ser-rich regions.

These observations suggested that the alignments were not significant, but driven by the numerous Pro or Thr residues in these EST sequences. To determine whether the alignments could be significant, we performed a reverse blast analysis. Each cell wall protein that matched an EST clone was blasted against the *S. dubia* EST database. In none of the cases were the best matchs reciprocal. Other *S. dubia* EST aligned with the cell wall proteins used as query yielded a higher similarity (but the E value was never below 10^{-19}). These clones were actually unambiguous homologues of other known proteins that matched the cell wall protein queried only because of their high content in Pro, Thre or Ser.

I interpret the results as evidence that the similarities between the EST clones and the cell wall proteins are not significant. Since the peptide sequences did not allow the identification of SAPs, and the EST database did not provide any SAP candidate, we focused on the characterization of SAPs with molecular techniques.

2 PCR with degenerate primers to isolate SAP genes

2.1 Description of the approach

The MS peptides were used to design primers for the polymerase chain reaction. Degeneracy of the genetic code implied that the primers were also degenerate, *i.e*. a single primer contained different oligonuleotide species each coding for the peptide amino acid sequence. The degeneracy value corresponds to the number of oligonucleotides species in the primer mixture, therefore the lower the degeneracy, the higher the specificity of the primer. Amplification with primers over 1 000 fold degeneracy have been successfully used (McPherson et al. 1991).

The primers were used in attempts to amplify the DNA or RNA regions which link the peptides. The major inconvenience of this method is that neither the length of the region to be amplified, nor the arrangement of peptides in the primary structure of the protein are known. Therefore, two primer orientations have to be tested for each peptide and the product would not be identifiable by its size. The apparent molecular weight of the SAP only provides an estimation of the maximum length for the coding sequence.

It is however expected, provided that favorable conditions are met, that the single oligonucleotide pair which amplifies the target product is the most efficient primer pair in the reaction mixture. Therefore, we planned on selecting candidate PCR products based on their amplification efficiencies, and verify that their sequences code for the expected peptide sequences to indicate whether they could belong to the protein of interest.

2.2 Primers design

To achieve the lowest possible degeneracy we avoided the leucine and isoleucine residues, peptide regions which presented some ambiguities, and amino acids encoded by6 codons (Ser, Leu, Ile). On the other hand we tried to select as much as possible the amino acids encoded by fewer codons. The numerous Ser, Leu, and Ile residues present in the SAP peptides greatly reduced the possibilities for primer design. Table 3 contains the primers used. In some cases we took into account the codon usage of *S. dubia* as calculated from 23 EST (3 280 codons) (Becker et al. 2001), or we substituted the bases with Inosine in order to reduce the degeneracy.

Sequence	Forward primer	Deg.	Reverse primer	Deg.
SAP126				
1: L**AAGFYDV**SYGA	GCNGCNGGNTTYTAYGAYGT	512	SACGTCRTARAANCCNGCNGC	512
2: QSSL**DGYNFFV**R	GAYGGNTAYAAYTTYTTYGT	128	NACRAARAAGTTRTANCCRTC	256
3: L**PAFDPD**	CCNGCNTTYGAYCCNGA	256	RTCNGGRTCRAANGCNGG	512
SAP116				
1: LTNP**TNQVPVWV**FFK	ACNAACCAGGTSCCNGTNTGGGT	128	ACCCASACNGGSACYTGRTTNGT	256
1': LTNPTNQV**PVWVFF**K	CCIGTBTGGGTBTTYTTYAA	36	AARAANACCCAVACNGG	192
2: A**AYTNAAWA**LSTSSK	GCNTACACNAACGCNGCNTGGGC	256	GCCCANGCNGCGTTNGTRTANGC	512
3: S**GATGVWE**LNL	GGNGCNACNGGNGTNTGGGA	1024	YTCCCASACNCCNGTNGCNCC	1024
3': S**GATGVWE**LNL	GGSGCNACNGGBGTBTGGGA	288	TCCCAIACVCCNGTNGCSCC	96
SAP98				
1: LMGLL**FKEATGA**NVDLFQEK	TTYAARGARGCNACNGGNGC	512	GCNCCNGTNGCYTCYTTRAA	512
2: **TTWDTAG**STTK	ACNACNTGGGAYACNGCNG	512	CNGCNGTRTCCCANGTNGT	512
3: SLS**VDNDDGR**	GTNGAYAAYGAYGAYGGNCG	256	CGNCCRTCRTCRTTRTCNAC	256
4: T**FTVYAVDR**	TTYACNGTNTAYGCNGTNGA	1024	CGRTCNACNGCRTANACNGT	1024
SAP96				
1: CRV**EAYKFFA**ADVK	gGARGCNTAYAARTTYTTYGC	128	gGCRAARAAYTTRTANGCYTC	128
2: VTFGCS**EEPCTY**K	ttgGARGARCCNTGYACNTA	128	ttgTANGTRCANGGYTCYTC	128
3: SLP**ATFTVEE**ALDTVLTFKR	atgTTYACNGTNGARGARGC	128	ttaGCYTCYTCNACNGTRAA	128

Table 3. Peptides used and sequence of the corresponding primers. The amino acids corresponding to the forward primer are in bold letters, while those correspondings to the reverse primer are underlined. Nucleotides in lower letters are arbitrary additional nucleotides. Deg : degeneracy.

2.3 Degenerated primers amplify reproducable band patterns

We used the SAP primers to amplify *S. dubia* genomic DNA, as well as pSPORT-cloned cDNA isolated from a mixture of flagella-regenerating and interphase cells (Becker et al. 2001), although SAP are not synthesized during flagella regeneration (Perasso et al. 2000). For SAP126 and SAP116 we tried all primer pair combinations. The PCRs for SAP126, SAP98 and SAP96 did not yield any amplification products. The primers for SAP116 yielded considerable background by amplifying reproducable patterns of discernable bands between 0,5 and 2 kb (Fig. 5B). These patterns originate by self-priming, *i.e.* a single primer acts as both forward and reverse primers during amplification. Different optimization attempts failed to reduce this background. Re-designing the primers (primers SAP116-1' and 3') with a lower degeneracy by using the codon usage of *S. dubia* did not result in a significant improvement.

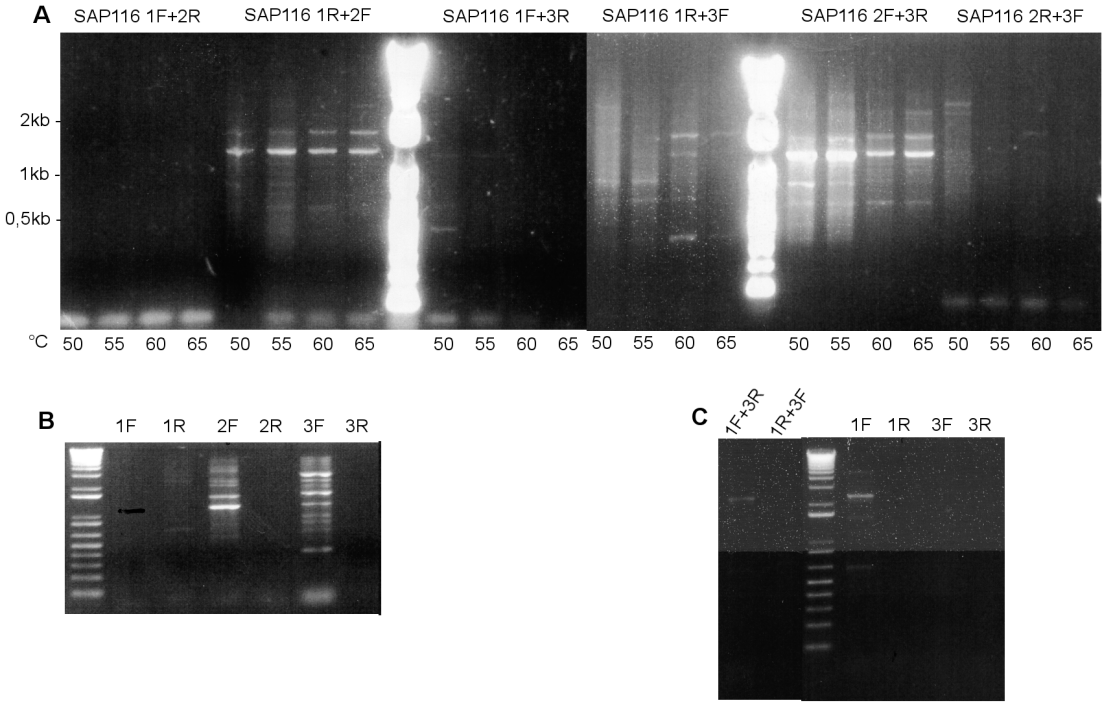

Figure 5. **Exemple of PCR with SAP116 degenerated primers**. **A:** gradient PCR using increasing annealing temperatures on genomic DNA using all SAP116 primers. **B:** Single primer amplification using SAP116 degenerated primers and genomic DNA. **C:** PCR using SAP116 primers on pSPORT-cloned cDNA and single primer amplification.

Figure 5A shows the result when amplifying genomic DNA with the three pairs of forward and reverse primers of SAP116 at increasing annealing temperatures. Many products are amplified within a DNA smear. The patterns of bands that are ampified by self-priming (Fig. 5B) are not eliminated by raising the annealing temperature testifying their strong specificity. Self-priming background was only slightly reduced when amplifying from cDNA (Fig. 5C) suggesting that these background products may be amplified from transcribed sequences. Semi-nested and nested PCR designed to increase the specificity of the amplification only resulted in the reamplification of the same patterns of bands or no amplification at all (not shown).

Our observations suggest that the reproducible bands amplified by some primers represent regions where the primers could bind at several different loci and in different orientations. Most of these regions were GC-rich. It was never possible to amplify a single PCR product. Some of the strong PCR products were isolated and sequenced, which confirmed that they were amplified by different oligonucleotide species that occur in a single degenerate primer. Sequencing also did not indicate that the candidate products could have been part of a SAP gene (the expected amino acids from the peptide sequence, as well as sequences from other peptides were not found within the candidate sequences).

Obviously the large number of products which were generated even at the highest annealing temperatures by self-priming competed efficiently with the amplification of any

other sequence than could have been of interest. Self-priming amplification was favored by the degeneracy of the primers. In the optimistic case where a region of a SAP gene or transcript would have been correctly amplified among the others, identification of this particular product would have remained unlikely. Because of the restricted possibilities for the design of primers, I felt that it was not possible to further optimize this approach. I therefore started a different approach.

3 Amplification of SAP peptide sequences with short primers

3.1 Description

I developed a strategy that takes advantage of knowing the length of the desired PCR product in order to select it in a complex mixture. Primer pairs were designed based on the peptide sequences such that they would amplify the central region of the peptides. The exact length of the products is thus known unless an intron is present between the primer binding sites. The target products would eventually be sequenced, and the sequences used to design non-degenerated primers.

To allow a relatively large region (>4 amino acids) to be amplified the primers had to be quite short (9-11 nucleotides corresponding to 3-4 amino acids). Although unusual, such short primers have already been successfully used in PCR amplification (Afonina et al. 1997). This constraint implied the following aspects:

- low degeneracy of the primers because of the small number of amino acids used to design the primers (degeneracy of SAP short primers was comprised between 8 and 144 fold),
- weaker specificity of the reaction inherent to the short length of the primers,
- low melting temperature for the primers that can allow mismatching more easily,
- short target products (27-62 bp) that can be difficult to purify and ligate.

To provide higher temperature stability and specificity, we tested the effect of additional nucleotides (6-8 nt) at the 5'-end of the primers which enhance stability of annealing during the later rounds of PCR (Rose et al. 1998).

In this strategy, primer extension did not require the polymerization of more than 55 nucleotides to complete a product of interest. This theoretically represents less than 4 seconds of polymerase activity at the optimal temperature (1kb/min <-> 16b/sec). In order to limit the amplification of longer or less specific undesired products, we used a 2-step PCR (annealing-denaturation) instead of the usual three steps (annealing-extension-denaturation). Enough nucleotide polymerization was expected to occur during the 15 sec

ramping time from annealing temperature to denaturation temperature allowing exponential amplification of the PCR products. Restriction-digested genomic DNA was used as template for this purpose as well.

3.2 Experimental validation of the strategy

We first tested the approach with a known gene of *S. dubia*. Eleven-mer degenerated primers were designed to amplify a 16 bp fragment of the rbcs gene (Table 4). Longer primers (17-mer) were also created by arbitrarily adding five or six nucleotides at the 5'-end of the primer. The target product was therefore 38 bp long, or 50 bp when the long primers were used.

A gradient PCR was performed to find the best annealing temperature. The rbcs long primers amplified a product at around 50 bp (arrow, Fig. 6A), while there was no product of the expected size when using the rbcs short primers (not shown). The band around 50 bp was excised from the gel and cloned into a plasmid vector. Few procedures were efficient to clone this 50 bp product : the best results were obtained using QIAEX II gel extraction kit (beads, QIAGEN) or in-gel ligation. Spin column purification kits and electroelution/DNA precipitation did not allow an efficient recovery of the 50 bp product. Sequencing of the inserts finally confirmed that this band was correctly amplified from the rbcs gene.

3.3 Isolation of a short fragment of the SAP98 gene

The strategy was reproduced for the SAP peptides. Some of the primers were divided into pools to account for the degeneracy of serine when it had to be included in the primer sequence (Table 4). Genomic DNA was used as the polymerization template because it is not known when SAP are actually expressed (Perasso et al. 2000). Gradient PCR spanning 10°C were performed as in the rbcs test without affecting the patterns of amplification qualitatively (not shown).

Only two PCR products were observed at around the expected sizes, one for a SAP116 peptide `(L/I)AFSG(L/I)TDGTHT(L/I)YVSAK`, and the other for the SAP98 peptide `(L/I)TADD(L/I)S(L/I)DQGFNY` (arrows, Fig. 6B). The products were cloned and sequenced. The products amplified with the SAP116-1 primer pair did not translate into the 62 bp target sequence. Five clones of the SAP98 primer pair possessed a nucleotidic sequence that could be translated into the SAP98 peptide sequence. Four of them needed a

Rbcs primers	Product length	Forward	Deg.	Tm	Reverse	Deg.	Tm
Short	38 bp	GGNGCNAARGT	32	36°C	ATGGGGTTCCA	8	34°C
Long	50 bp	gaattcGGNGCNAARGT	32	52°C	ctgcagATNGGRTTCCA	8	54°C

SAP peptide	Product length	Forward and reverse primers
SAP212		
VVVLDSFGQQS(L/I)EAK	57 bp	Forward : ggatccGTNGTNGT
		Reverse : gaattcYTTNGCYTC
SAP116		
(L/I)AFSG(L/I)TDGTHT(L/I)YVSAK	62 bp	Forward : atctgcaGCNTTYTCNGG/atactgcaGCNTTYAGYGG
		Reverse : atctcgagGCNGANACRTA/atctcgagGCRCTNACRTA
(L/I)TNPTNQVPVWVFFK	54 bp	Forward : agaattcACNAAYCCNAC
		Reverse : aggatccAARAANACCCA
EAGQYV(L/I)E(L/I)S(L/I)SDK	51 bp	Forward : attctaGARGCNGGNCA
		Reverse : aaggtccYTTRTCNGA/aaggtccYTTRTCRCT
SAP98		
(L/I)TADD(L/I)S(L/I)DQGFNY	53 bp	Forward : agaattcACNAAYCCNAC
		Reverse : aggatccRTARTTRAANCC

Table 4.Primers used for rbcs test and SAP peptide sequencing. Greyed amino acids correspond to the primer sequence in capital letters. Small letters represent arbitrary additional nucleotides. When two primer sequences are shown, they were pooled together. products could be due either to a difficult region to sequence or to mispriming at the 5'-end of the reverse primer. Deg: degeneracy

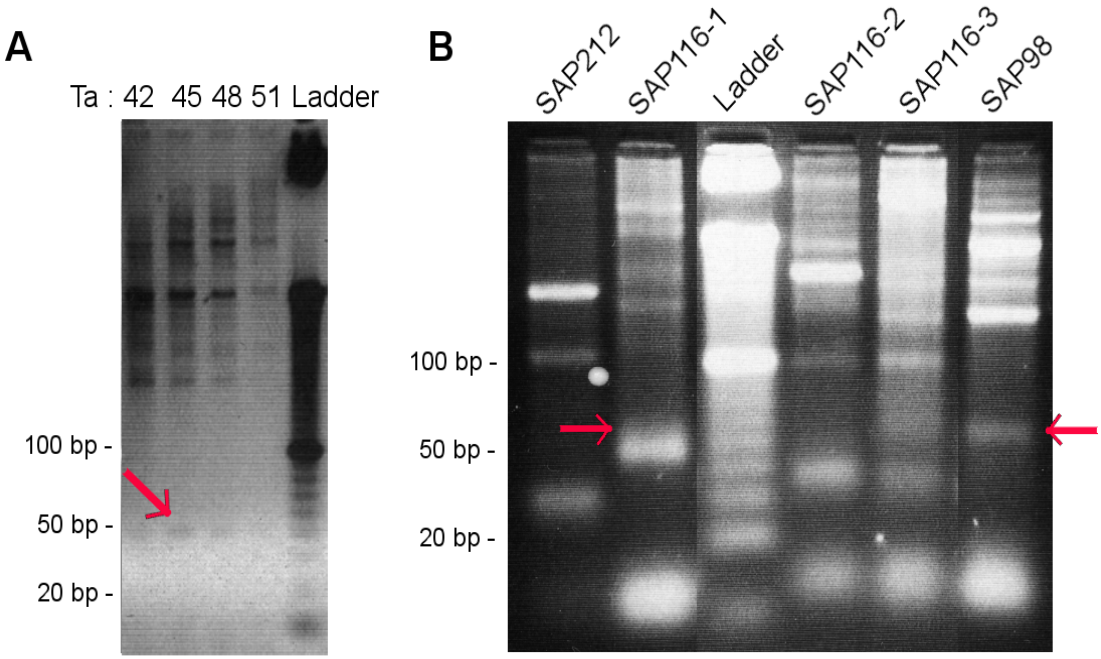

Figure 6. **Amplification of short DNA fragments with short degenerated primers**. **A:** gradient PCR with rbcs short degenerated primers. Arrow: 50 bp target product, Ta: annealing temperature. **B:** PCR with SAP short degenerated primers. Arrows: products at expected sizes.

gap to be added in a GC-rich region, but one was complete (Fig. 7). This suggests that this short amplified genomic sequence could be a piece of the SAP98 gene. The missing pyrimidine in four of the of the products could be due either to a difficult region to sequence or to mispriming at the 5'-end of the reverse primer.

Despite several attempts, this strategy could not be repeated successfully for other SAP peptides, so we focused on using this piece of sequence as an "anchor" to elucidate the SAP98 sequence.

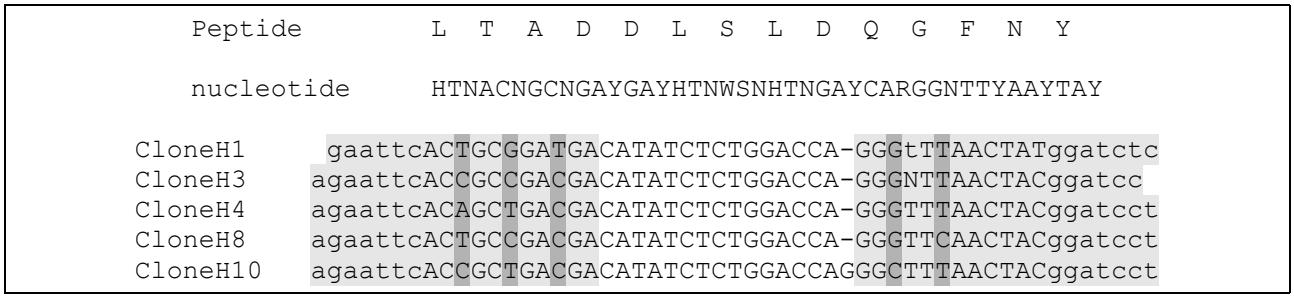

Figure 7. **Nucleotidic sequence of one of the peptides of SAP98 obtained by PCR**. Five clones align with the reverse-translated sequence of a SAP98 peptide. Grey zones highlight priming sites, ambigous bases are highlighted in dark grey. Lower case characters represent arbitrary additional nucleotides at the 5'-end of the primers.

4 Molecular characterization of SAP98

4.1 Extension of SAP98 sequence by inverse PCR

We extended the short piece of DNA using an inverse PCR strategy. Primer pairs for inverse PCR were synthesized and inverse PCR conditions were optimized using plasmids containing this DNA fragment. Then, genomic DNA was prepared with different restriction enzymes to produce suitable inverse PCR templates. Two products of 363 and 913 bp were successfully amplified from *Rsa*I- and *Sau*3AI-digested DNA (Fig. 8A). Sequencing showed that the two products overlapped, confirming that they originate from the same DNA sequence. Since there was no gap between the inverse PCR primers, a continuous piece of 937 nucleotides could be assembled which extended 203 and 699 bp downwards and upwards, respectively, of the starting DNA fragment (Fig. 8B). The continuity between the inverse PCR products was confirmed by PCR amplification with outer primers designed from each inverse PCR product, and sequencing of the amplification product.

The nucleotides extending from the 3'-end of the reverse primer used in inverse PCR (in bold in figure 8B) encode two N-terminal amino acids of the SAP98 peptide. This is a good indication that the sequence amplified by inverse PCR corresponds to the genomic sequence of the SAP98 protein. This sequence does not include any other known peptides

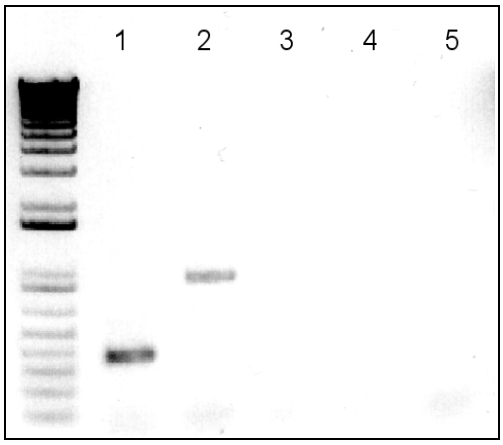

Figure 8A. **Inverse PCR for SAP98**. Two products were amplified from genomic DNA. The DNA was prepared with the following restriction enzymes 1: *Rsa*I, 2: *Sau*3AI, 3: *Eco*RI, 4: *Bam*HI, 5: *Bgl*I. Left-most lane: size markers, from bottom to top: 100, 200, 300, 400, 500, 650, 850, 1000, 1650 and 2000 bp.

>consensus (937bp):

TCGGCGAGACTAGGAGGGTGGGCCCATGCGAGGGGTTGGGGGGCTGGC
GGGACGTTAAGGTTGAGGCCAGGGACGAAGCAAACGGGGCAAAGCAAA
GGGGGAAGGGCAGCTGGAAGGCGACCGAAACGTTATGCCTTAACCCGC
CAGGGGTGTGTGCTGTGGTTGTTTAAGGTTCAGCGGGCGGTGTGCGGT
CCATGACCTACACGCGGGCGATGTGCAGGCAGCTGCAGAAGGGGAGGG
GACAGGGAGGAGCTTTGGCGCGTGTTTGGCGCTTTGCTTCAAACTAGA
TTTAGCTTGGCCTGCACCGTGGCCTGCATACTCTCTCGCTCACCTTCT
GTCTCTTGCGCATCCTTGCTGCTTGCACCTCTGCTGGGTGTTGGCACA
TCGACTGCATACTGCACGCAATCATATATTACACGAAGAACATTGGGC
AACGGTTCAGTCAGAAGTGCACACCACTGCGCAGTGTATTTTCGCTTT
CGGGAATTTATATTGCTGCTGCCGAGCTATCACTGGCACAGCCCGCGT
AGCCTGCTGGCAACCTAACCTCTGCGGGAGATGCCAAGCTGCATGTAC
TCTTTTGACACCCTTTGCTTTTCTGCAGGTGCCACCTGCGGGCTTGCG
GCACTTGCTTCGTATCCATCCGTCGTGACAATAACCAATTGCCAAGCA
TGCAAGCTGACCAAAACTGCA**CTCACT**GCCGACGACATATCTCTGGAC
CAGGGCTTCAACTACCCGACTGCTCTGGGGGGAGAACCAGAAGCGCCTG
TGGTATGCAGTAGACGGCGAGACGGGGTCTACTGGTTGCGGTGCCAGT
GGTTACAGTGGCACGGGTCACTGGACAAAGTTTTTCAACAGTGCTTAC
GATGGAATGCGCCCTGAGAAGGCCATCATCAGCGACACCTACATCAAG
ATCTTCGTCCAGAACGGCGGCACGT

Figure 8B: **Sequence assembly of the two inverse-PCR products**. Light grey: restriction sites for *Rsa*I and *Sau3*AI, dark grey: inverse PCR priming sites, underlined: outer primers binding sites, bold: nucleotides that fit with the remaining amino acids of the SAP98 peptide.

from SAP98, and a blast analysis did not find any similar sequence in the databases. Two stop codons occur in the same frame 3' from the SAP98 sequence, suggesting that either there is an intron upwards, or that the used sequence is close to the N-terminus of SAP98.

Further genomic walking by inverse-PCR was unsuccessful, therefore we turned to a RACE technique to isolate the full length SAP98 transcript.

4.2 Extension of the SAP98 sequence by rapid amplification of cDNA ends (RACE)

We first tried to detect the SAP98 mRNA using the outer primers underlined in figure 8B. RT-PCR successfully amplified the piece of SAP98 in both interphase and flagella regenerating cells (not shown). Full-length cDNA suitable for RACE was then synthesized from mRNA isolated 30 min after the begining of the flagellar regeneration process.

Both 3'- and 5'-RACE were performed with newly synthesized primers based on the inverse PCR result (Fig. 8A). A 2 388 bp 3'-RACE product (excluding polyA tail) was successfully amplified (Fig. 9), but so far we were unable to amplify with certainty the 5'-end of the mRNA.

3'RACE primer : ACRGCRGACGACATATCTCTGGACCAGGGSTTTAACTAC

3'RACE product (2388 bp) :

ACTGCTGACG ACATATCTCT GGACCAGGGG TTCAACTACC CGACTGCTCT GGGGGAGAAC
CAGAAGCGCC TGTGGTATGC AGTAGACGGC GAGACGGGGT CTACTGGTTG CGGTGCCAGT
GGTTACAGTG GCACGGGTCA CTGGACAAAG TTTTTCAACA GTGCTTACGA TGGAATGCGC
CCTGAGAAGG CCATCATCAG CGACACCTAC ATCAAGATCT TCGTCCAGAA CGGCGGCACG
TACGCTGTCG AATACACTTC TTCGCTGAAG ACAACGTTCT ACTATGCACT GTGCAAAGTC
TCGGGCACAA CAGTGACTTG CCCTACCTAC CAGAGCCTGG GCGAGTTCGC CTTCACCGAC
MATCCGTTCA TCCTGATAAA CAAACAT**Y**TG CCCGACGGCG AGTACGTTCT AAAGGTGTAC
TCCAAGGTGA CCACCTGCAC TACCTGCCCA TCGACGCTCA CGGACAGCTC GAGCCCGCTT
CAGATTCCCT TCATTGTGAT GTCCACCGCC CCAGTGGTGT CTATCACCTC AACCACAAGC
GCTTACATGA GCGCGACGAC CAGCACAGGC ACCCTCACCT ACAAGTCCAG CATGAAGAGC
AGGACC**Y**CCT CCAAGAATAC **Y**GACTCCATC ACATACCCCT GGCCTCTCTT CCAGACCTAC
TTTCCAGGTC AAGACCACCT GGGACACTGC TGGCAGCTGG AAGTCGCTGT CTGTGGACAC
GACGATGGCC GGGGGCAGTA TGCGTACTCT CTGTCGTCCG CAGCGCAGGG AACACACACG
TTCAACGTGC TAACAGTGGC AGCATCCAAC TACAGCTTGC CTGGCACAGT GTGCAGCTCT
GTCGCCTGCA GTGCTGCATG CCCGAGCATG TGCCTGTCCG ACTTCACCAA GATCGGCACA
TACAACAAGT ACTCCACCTA CTCTGCAAGC GCTGCCAGCT TCAGCTTCAT CTACGACAGC
ATTGCGCCCC CCTCCCCGCC CGTCAACTGC AAGAGCGGCG AGTGCTCGGG GCTGCTCTCC
ACCAGCACCA AGCACATGAC CTTCGAGTTC ACCTGCAGCG ACGCGAACGC GCCCTGCACC
TTCCACTGCG CGCTGGACGG CAAGCCCACC CTGGACGGCA CCAAGACCAG CAGCCTGAGC
AAGGGCTACA CCACGTGCAC CTCCCCAGCG GTGGTGAAGC CTGCGGCTGC AGGCACCCGC
ACATTCACTG TGTATGCAGT GGACCGAGCT GGGAATGCAG GCGAGCTGTC CATCCCAGTG
AGCTTCTACG TTGACAACAC TGCTCCGCTG GTGTACTTCG CGGGGGTCGC GAAGAGGTGC
CTGATCAACG AGAGGTACTT CGTGCCAGAC ATTGGCAGCT CCACAGCAGG GCCTACTTCC
TCCTCGGTCA CCAACACCGA CGGTGCCCAG GCTGAATTCG CCGCGGGTCG GTGCCTGGAT
GGCACGAATG CGTTCGGCAG CACTGGCTCC ATCACCGAGC CGAACACCTA CATCAGTGCG
GCAACGTACG GCAGCCTGGA CTGCACATGC GGCACCTACC AGATCAAGAC GTCTTCCACC
AGCTCGACGT ACTACACCTA CAACCAGGCC ACCTTCCCTG GGCCGCTCAC TGCCGACTCG
GACTCCTCTC TCAACAGCTA CCTGGGCTAC TACCCGACTG CCATCCCTGA TGTGGGCCTC
TACAAGCTGA CTGAGGGCCC GATTCAGACA CTGACAGACG GCACCACCTC CATCACGGAC
TCTGACGGCT CCTTCGACAT CACGGGGCAG GCGGCAGGGT ACATGATGTA CGACGCGGGC
ACCGCGTTCA ACACCACGGA CGACATCTCC ACCACTGGCG GCATGTACGA CACCGTCATC
CAGGTCCCCC TGGACGTGAC CCGCGAGATC TACAACCTGG ACTCCACCTA CGGCGCGGTG
CTGGCCACCA AGAAGGTGGC GGACAGCTAC TACTACCACG TCCTGCACGC CACCAACTCG
CCGGTGGCCA CCGTCAACCT GGTCTGCAAG CACCCCGAGA TCCTCAACTG CGTGCAGTCC
GCCCGCAACC CCTTCCAGAT GGACTGCGGC GGCATGTACA GCTGAGCGGC GCCTTTGCCC
AGTTGGCTGG GCCGCGGACA GAGGAGGGGC GGCAGGCAGG CTGCCATAGT AGGAGGCTGC
CGGGGCACTC TGCACCTCTG TGGCAGTATC CTTAGGGTCA CAACGCTGGG CTCGCCGGGG
GGACATAAAG CAGCTCCGCT ATCCAAGCGT TGACGCGCTG ATCTGTGCAC CGGCAGCACA
CCAGCCGACT TGTGCGGAAC ACCATGCTGT TCTGGTTTCC CTCGGCATCC AACTGCAAAC
ACTGTTAGAT TTTAAGATAT CAATTGAAGC TGTCACAGCT TACAAACT

Figure 9. **RACE PCR products for SAP98.** The polyA tail was removed. The 3'-UTR is shaded.

4.3 SAP98 sequence analysis

The RACE product was aligned with the genomic sequence obtained by inverse PCR (in Fig 8B). The 3'-end of the cDNA aligns correctly with the 241 last nucleotides of the genomic sequence without interruption. The open reading frame is 2082 bp long (694 aa totalizing 73,9 kDa) and the 3'UTR is 306 bp long, which is about the average length of 3'UTR observed in *S. dubia*. SAP98 migrates in an SDS-polyacrylamide gel as a 98 kDa

```
1   TADDISLDQGFNYPTALGENQKRLWYAVDGETGSTGCGASGYSGTGHWTK     50
    |||||||||||||
   LTADDISLDQGFNY

51  FFNSAYDGMRPEKAIISDTYIKIFVQNGGTYAVEYTSSLKTTFYYALCKV    100


101 SGTTVTCPTYQSLGEFAFTDNPFILINNIADGEYVLKVYSKVTTCTTCPS    150


151 TLTDSSSPLQIPFIVMSTAPVVSITSTTSAYMSATTSTGTLTFKSSMKSR    200


201 TSSKNTDSITYPWPLFQTYFQVKTTWDTAGSW-KSLSVDNDDGRGQYAYSL   250
             ||||          ||||||||. ||||||||||||
            TYPW           TTWDTAGSTTKSLSVDNDDGR

251 SSAAQGTHTFNVLTVAASNYSLPGTVCSSVACSAACPSMCLSDFTKIGTY    300
              |:||     |||.
             VVTV-----SLPR

301 NKYSTYSASAASFSFIYDSIAPPSPPVNCKSGECSGLLSTSTKHMTFEFT    350


351 CSDANAPCTFHCALDGKPTLDGTKTSSLSKGYTTCTSPAVVKPAAAGTRT    400
                                ||||||||||||       |
                               GYTTCTSPAVVK       T

401 FTVYAVDRAGNAGELSIPVSFYVDNTAPLVYFAGVAKRCLINERYFVPDI    450
    ||||||||
    FTVYAVDR

451 GSSTAGPTSSSVTNTDGAQAEFAAGRCLDGTNAFGSTGSITEPNTYISAA    500


501 TYGSLDCACGTYQIKTSSTSSTYYTYNQATFPGPLTADSDSSLNSYLGYY    550


551 PTAIPDVGLYKLTEGPIQTLTDGTTSITDSDGSFDITGQAAGYMMYDAGT    600


601 AFNTTDDISTTGGMYDTVIQVPLDVTREIYNLDSTYGAVLATKKVADSYY    650
                                |||||||  |||||
                               YNLDSTYQ-VLATK

651 YHVLHATNSPVATVNLVCKHPEILNCVQSARNPFQMDCGGMYS           693
                      |||:||||:|||
                      HPELLNCVKSAR
```

Figure 10. **Alignement of the 3'RACE product sequence with the SAP98 peptides obtained by mass spectrometry.**

glycoprotein, thus the 3' RACE product represents at least 75 % of the coding sequence of the protein, but probably much more considering the presence of glycan chains. The stop codon is followed by a 303 nt long 3'-UTR which does not contain the TGTAA

polyadenylation signal commonly found in *S. dubia's* EST. However, two similar sequences are found in the vicinity of the polyA tail (TGTTA at -46 bp, and TTGAA at -25 bp). Eight other peptide sequences obtained by mass spectrometry (Table 2) are included in the 3'-RACE sequence, providing strong evidence that this cDNA codes for the SAP98 protein (Fig. 10).

As figure 10 shows, only four of eight peptides are fully complementary to the translated nine sequences of the SAP98 cDNA. These differences can obviously prevent the success of a PCR-based strategy. They could arise from post-translational modifications or simply from MS sequencing error. It is known that *de novo* sequencing by MS-MS is subject to misinterpretation and it was at least the case once for the YNLDSTYQVLATK peptide where a GA was interpreted as a Q (they have the same monoisotopic mass).

Blast and blastp analyses did not retrieve any similar sequence in the databases indicating that SAP98 is a novel protein. No protein domain or protein motif was recognised using the Interpro search. Analysis of the translated sequence shows that this part of the SAP98 sequence is particulary rich in threonine (13,4%), serine (11,6%) and tyrosine (6,8%). It is also negatively charged (55 E-D vs 37 K-R), all charges being more or less evenly distributed (absence of charge cluster). Two of six tested transmembrane helix prediction programs predict two or three transmembrane helices in the first 300 amino acids (TMpred and Toppred programs were positive, HMM, TMHMM, ConPred II and DAS-TMfilter were negative).

B Identification of putative genes involved in flagellar regeneration from a flagella regeneration-specific cDNA library

Flagellar regeneration is known to trigger the upregulation of a set of genes coding for proteins involved in the assembly of the new flagella in *Chlamydomonas* (Schloss 1984, Stolc et al. 2005). In *Scherffelia dubia*, scale biogenesis is also strongly induced during flagellar regeneration (McFadden and Melkonian 1986, Becker et al. 1995, Perasso et al. 2000), suggesting that the transcription of genes involved in flagellar scale biogenesis might be induced as well. To better understand flagellar regeneration in a scaly green alga, we generated a flagella regeneration-specific cDNA library from *Scherffelia dubia* which could allow the identification of genes involved in scale biosynthesis.

1 Principle of the Suppression Subtraction Hybridization (SSH)

Suppression Subtraction Hybridization (SSH) (Fig. 11) is described as a powerful method to generate subtracted libraries enriched in differentially-expressed cDNA (Diatchenko et al. 1996, Diatchenko et al. 1999, Rebrikov et al. 2004). The subtraction is obtained by two successive hybridizations between a cDNA population (=*tester*) containing the differentially expressed cDNAs (=*target*) and a reference cDNA population which is added in excess (=*driver*). It is important to note that the cDNA populations are digested with the *RsaI* enzyme (`GT|AC`) prior to hybridization. Two tester sub-populations are created by ligating different adapter molecules. Each sub-population is heat-denatured and hybridized a first time with 30X excess driver which has no adapter. During this first hybridization, common cDNA molecules from the tester sample hybridize preferrentially with the driver molecules in excess, while differential cDNA do not hybridize with the driver. The second order kinetics of the hybridization process makes abundant molecules reanneal faster than less abundant molecules. The first hybridization results in normalization and enrichment of differentially expressed cDNA in the single stranded fraction (Gurskaya et al. 1996). The success of the subtraction depends on the efficient isolation of these molecules from the rest of the hybridization mixture. This is obtained by their selective amplification by PCR. Selective amplification is made possible by creating hybrids between the single stranded molecules of each sub-population during a second hybridization. Since these hybrids carry two different adapters, they can be selectively amplified, while undesired products are either non-exponentially amplified or not amplified at all by taking advantage of the "suppression-effect" mediated by the long inverted terminal repeats present in the adapters (Lukyanov et al. 1995) (Fig. 11).

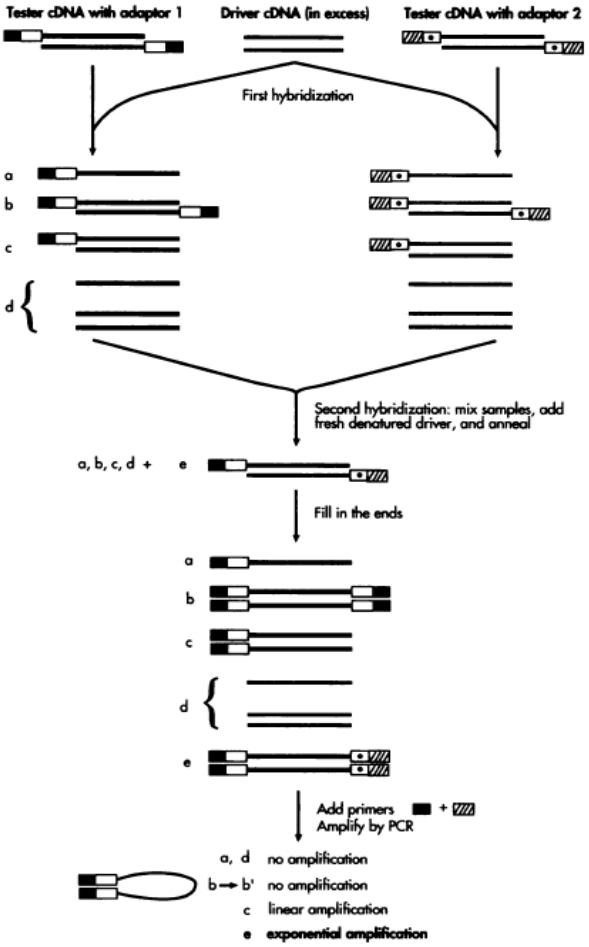

Figure 11. **Flowchart of the SSH method**. Tester and driver full length-cDNA are digested with *RsaI* and two different adapters are ligated. Two successive hybridization are then carried with 30X excess driver (no dapater). In the final reaction mixture there exists five different types of molecules once the adapters are filled: types a, b, c, d, and e. Hybrids which have been formed with driver cDNA (types c and d) bear at most one adapter and therefore cannot be amplified exponentially in subsequent PCR, as well as any remaining single stranded cDNA molecule (types a and d), whereas hybrids formed from tester cDNA only carry either the same adapter (type b) or different adapters (type e, formed during the 2nd hybridization) at each end. The suppression effect prevents type b hybrids from exponential amplification therefore, only type e cDNA are exponentially amplified in subsequent PCR, and they represent a normalized fraction of cDNA enriched in differentially-expressed cDNA.

2 Generation of the subtracted library

Scherffelia dubia cells (1-2.10^6 cells/mL) were harvested during the exponential growth phase by centrifugation at the begining of the light phase before they enter division (Wustman et al. 2004). One third of the cells (interphase sample) was left on ice while the remaining two thirds were deflagellated by mechanical shearing. Because scales are synthesized as soon as 5 minutes after flagellar regeneration begins and flagellar regeneration reaches a maximum at around 30 min, we decided to isolate and mix RNA from the begining of the flagellar regeneration and 30 min after the begining of flagellar regeneration. Therefore, half of the deflagellated cells were processed immediately for

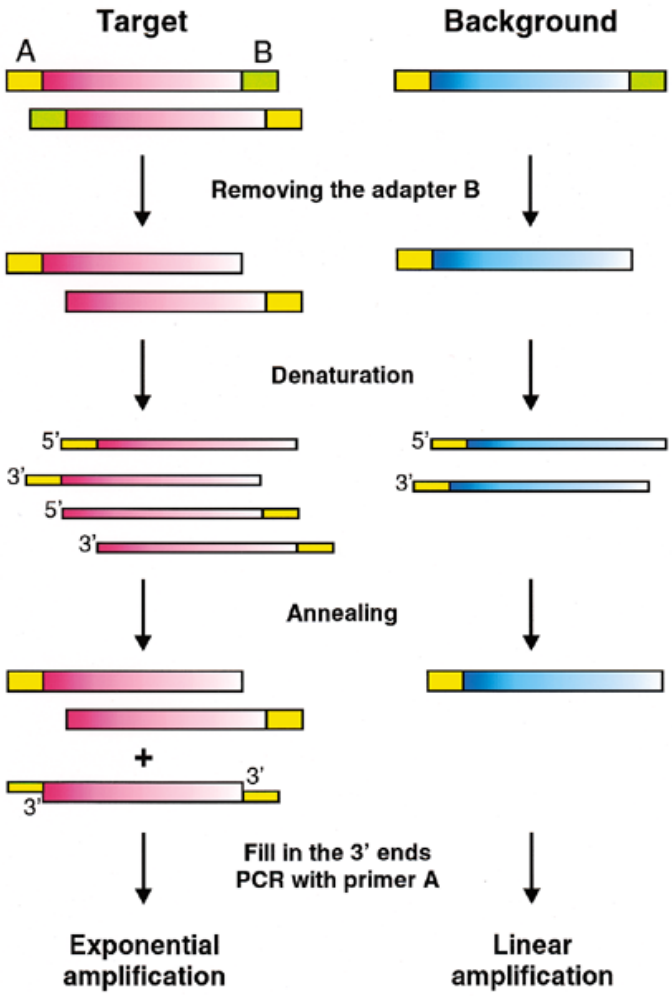

Figure 12. **Principle of miror oriented selection**. MOS procedure reduces the background originating from redundant cDNA molecules having by chance evaded elimination by hybridization. This type of background cannot be revealed by differential screening, and can become really significant in cases when target molecules are rare. MOS takes advantage of the estimation that such events are rare enough so that background molecules have only one orientation with respect to the adapters. MOS is performed on the subtracted cDNA obtained after SSH by removing one adapter by restriction digestion and performing another hybridization, followed by amplification with the remaining adapter. MOS thus reamplifies the differential cDNA without the background molecules.

total RNA extraction, while the other half was allowed to regenerate their flagella for 30 min at 20°C before total RNA was extracted from the cells. The total RNA extraction protocol required a 5 minute centrifugation step performed at room temperature during which it was assumed that flagellar regeneration could proceed before cells were frozen in liquid nitrogen. Equal amounts of the two flagellar regeneration samples were mixed together to form the FR (Flagella Regeneration) sample. Total RNA extracted from the interphase sample is referred to as the I (Interphase) sample.

I and FR total RNA were sent to the EVROGEN company (Moscow, Russia) which performed SSH followed by mirror orientation selection (MOS), a procedure to remove

background molecule (Fig. 12, Rebrikov et al. 2000) to generate the flagellar regeneration-specific cDNA library (FR-I, see Material and Methods for a more detailed protocol).

The resulting subtracted library is composed of *RsaI*-digested cDNA fragments of tester population enriched in differentially expressed molecules with the following consequences: i) full length cDNA information is lost, ii) different clones can represent different, non-overlapping regions of the same mRNA, iii) clones including 5'- or 3'-UTR should be identifiable through their SMART peptide sequence and polyA tail respectively, however long UTR might preclude sequence identification by similarity, iv) 3'-UTR should be under-represented because they can form hybrids through their polyA tail, lastly v) cDNA fragments from differential mRNA which possess a conserved domain could be lost by hybridization with other fragments possessing the same domain, the corrollary is that the fragments which are the most specific to a protein are mostly retained.

3 Outcome of the flagellar regeneration-specific cDNA library

The subtracted cDNA was ligated in plasmid vectors and transformed into bacteria. Plasmid DNA was extracted from 556 randomly selected clones and sequenced, of which 449 sequences were of sufficient quality for further analysis. The average length of the *RsaI*-digested cDNA fragments was about 600 bp. Redundant clones were detected using the program Overlap in the GCG Wisconsin package (threshold: >80% nucleotide identity over a stretch of at least 50 nucleotides) and clustered, resulting in 383 clones (85%) being assembled in 53 contigs (designated with letters) from which a consensus sequence was established. Sixty-six clones (designated with 96-plate coordinates) were unique totalizing 119 different contigs (26% of all clones). Thus, the SSH + MOS cDNA library has a degeneracy 2,4 times higher than the (unsubtracted) interphase and flagellar regeneration EST library (Becker et al. 2001).

The contig sequences were eventually analyzed. Similar proteins or EST were searched using the BLAST algorithms in the following databases:

i) non-redundant nucleotide databases at the ncbi website ;

i) the EMBL-SWISSPROT protein databases at the ncbi website ;

ii) the *Chlamydomonas reinhardtii* proteome (chlre2_finalModelsV2) at the JGI website ;

iii) a local database of *Scherffelia dubia* EST containing 714 unique EST (Becker et al. 2001) ;

Blast results with an expectancy value higher than $1.e^{-10}$ were not considered. All the bioinformatic results are presented in tables A, B and C (see annexes pp. 64-66). The synthesis of these analyses is shown in figure 13 and Table 5 and can be summed up as

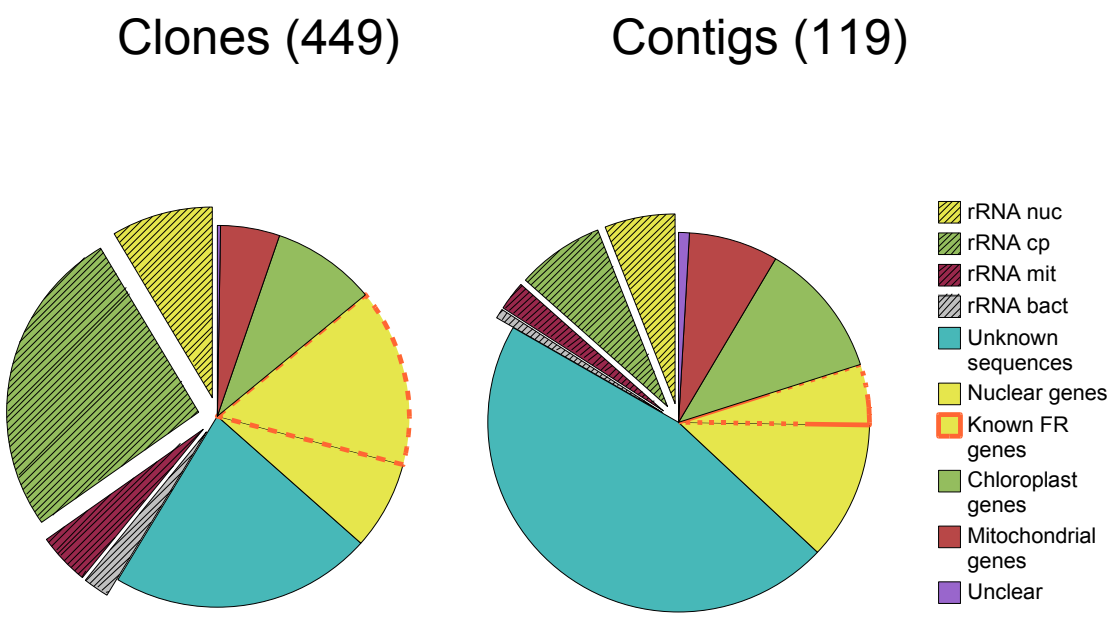

Figure 13. Analysis of the flagella regeneration-specific cDNA library. Repartition of the 449 sequenced clones (left hand diagram) and the 119 contigs assembled from the 449 clones (right hand diagram). FR: flagella regeneration

Inserts	Clones / Contigs
Chloroplastic rRNA	118 (26,3%) / 9 (7,6%)
Cytoplasmic rRNA	38 (8,5%) / 7 (5,8%)
Mitochondrial rRNA	20 (4,4%) / 3 (2,5%)
Bacterial rRNA	10 (2,2%) / 1 (0,8%)
Non RNA clones	263 (59%) / 99 (83,2%)
Total	449 (100%) / 119 (100%)

Putative protein-coding RNA contigs	99
No similarity, or similarity > e^{-10}	55
With similarity < e^{-10} :	
- Nucleus-encoded product	20
- Chloroplast-encoded product	14
- Mitonchondria-encoded product	9
- unclear	1

Table 5: **Characterization of the clones amplified by SSH+MOS**.

follows. The SSH + MOS subtracted cDNA library contains a high proportion of amplified ribosomal RNA fragments (41% of the clones regrouped in 26% chloroplastic rRNA, 8% cytoplasmic rRNA, 4,4% mitochondrial rRNA, 2,6% bacterial rRNA). Although significant in number, these rRNAs accounts for a small part of the diversity of the library (16,8%) as they cluster in only 20 different contigs. The redundancy of the library is therefore largely explained by the abundancy of ribosomal RNA amplified during SSH + MOS. The 99 remaining non-rRNA contigs are mostly unidentified (55%). Among the clones identified by similarity, the majority code for nuclear encoded proteins (20%) but a significant proportion originates from plastidic and mitochondrial transcripts (14% and 9% respectively).

We defined redundant clones as clones which share an overlapping sequence with each other, however this does not take into account the redundancy of transcripts for which different non-overlapping contigs are present in the library, such as transcripts coding for the chloroplastic ribosomal protein L12 (two contigs in the library), the mitochondrial alpha subunit of the ATPase complex (two contigs) and NADH-ubiquinone oxidoreductase (three contigs). Thus the true redundancy of the library is slightly higher.

Clones with unknown sequences are by far the less redundant of the library. These clones were also compared to the 3 300 *Mesostigma viride* EST (Simon et al. 2006) with the reasoning that common unidentified clones in *S. dubia* and *M. viride* may be functionally related to scales, but no homologues were found.

Finally, only twelve non rRNA contigs are also found in the *S. dubia* interphase and flagella regeneration EST library (8 nuclear encoded transcripts and 4 unidentified transcripts), indicating that the majority of the SSH + MOS contigs represent new sequences of *S. dubia*.

4 Validation of the flagella regeneration subtracted library

4.1 Relative enrichment in known flagella-related genes

To estimate the success of the subtraction procedure, I compared the amount of known flagella-related genes in the SSH + MOS cDNA library and the unsubtracted interphase + flagellar regeneration EST library (Becker et al. 2001). Contigs either similar to transcripts induced more than 100% during flagellar regeneration in *Chlamydomonas reinhardtii* (Stolc et al. 2005), or coding for proteins identified by more than five peptides in the flagella proteome (Pazour et al. 2005), or similar to conserved genes in ciliated organisms (Li et al. 2004) were searched in both library. Only six flagella-related contigs were identified in the subtracted library: 14-3-3-like protein (FTT), heat shock protein 90 (hsp90), outer arm dynein light chain 2 (oda-lc2), two flagella-associated proteins of unknown function

(fap184 and fap115) and an open reading frame of unknown function (C_2450124) (Table 7). This is less than the 39 contigs identified in the same way in the *S.dubia* interphase and flagella regeneration EST library. This is however a relative enrichment when considering nuclear transcripts alone: the number of clones containing flagella-related cDNA is 3,5 times more important in the SSH + MOS cDNA library (66,3% vs 18,9%) (Table 6). Qualitatively, this enrichment is less pronounced (1,5 times, 30% vs 20,6%) as the clones are more redundant (especially the 14-3-3 protein) in the SSH + MOS library. Nonetheless, three of these contigs are reported for the first time in *S. dubia*: the outer arm dynein light chain (oda-lc2), and the two unknown proteins present in the flagella proteome of *Chlamydomonas* (fap184, fap115) (Pazour et al. 2005).

	Interphase + FR library		SSH + MOS subtracted library	
Clones		% of flagella-related sequences		% of flagella-related sequences
Flagella-related clones	64		67	
Number of clones sequenced	1032	12,4%	449	14,9%
Nuclear transcripts only	677	18,9%[a]	101	66,3%
Contigs				
Flagella-related contigs	39		6	
Total number of contigs	714	10,9%	119	5,0%
Nuclear contigs only	378	20,6%[a]	20	30,0%

Table 6. **Comparison between subtracted and unsubtracted EST libraries.** a: the proportion of flagella-related sequences in the interphase + FR library has been doubled to reflect the relative abundance of flagella-regenerating cells during the making of this library.

4.2 Experimental validation of induced expression during flagellar regeneration

We verified that the transcripts coding for some of the flagella-related proteins identified above (Table 7) were really induced during flagella-regeneration in *S. dubia* by semi-quantitative (comparative) RT-PCR analysis. The primers were designed based on the sequence of the respective clones of the SSH + MOS library. Each mRNA was simultaneously amplified with an internal standard from oligodT-primed reverse-transcribed total RNA extracted from interphase or 30 min-flagella regenerating cells (Fig. 14). For all transcripts which could be successfully amplified, i.e. 14-3-3 protein (FTT), heat shock protein 90 kDa (HSP90) and flagella associated protein 115 (FAP115), the mRNA were clearly more abundant during flagella regeneration than during interphase confirming that these genes were also upregulated during flagellar regeneration in *S. dubia*. There was no PCR product for oda-lc2 with the primers we used, and FAP184 and contig L were not tested.

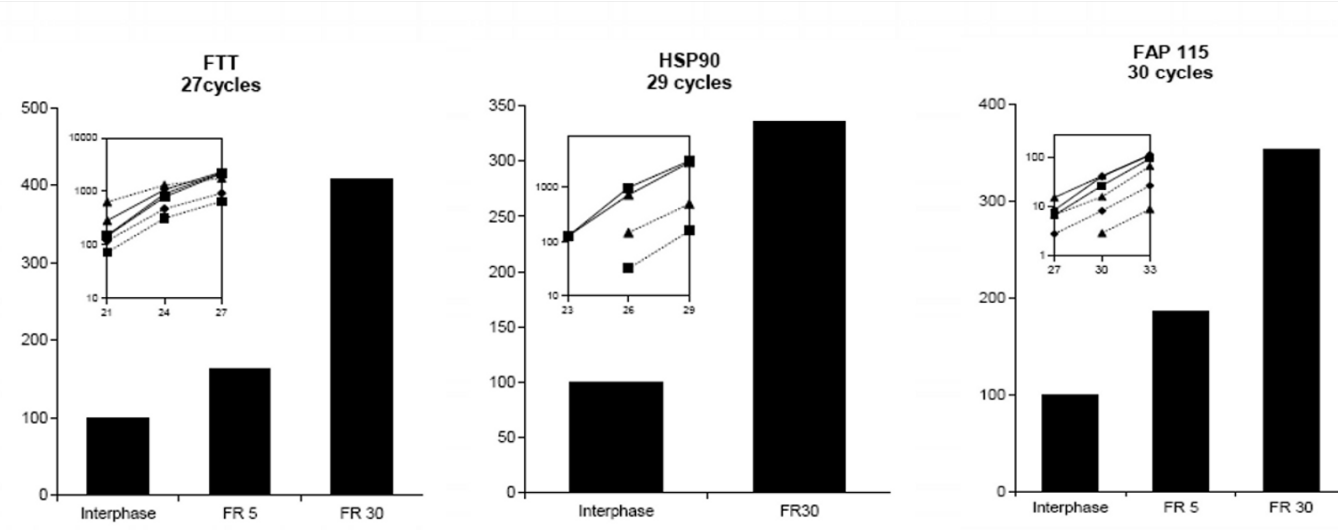

Figure 14. **Abundance change of 14-3-3-like protein (FTT), HSP90 and FAP115 mRNA during flagellar regeneration**. Each mRNA was amplified simultaneously with an endogenous standard (actin for FTT and FAP115, rbcs for hsp90) from total RNA. The efficiency of the reaction was checked at different cycles (inset). Product accumulation was measured by computer assisted imagery. Bars represent fold differences between end-products normalized to the endogenous standard after the indicated number of PCR cycles in one representative experiment. Insets: Product accumulation vs PCR cycle of endogenous control (plain line) or target product (dashed line), from interphase cells (■), cells at the begining of flagella regeneration (♦) and 30 min flagella-regenerating cells (▲).

5 Identification of putative upregulated clones with a differential screen

To identify which among the other known and unknown sequences could be candidate flagellar regeneration-upregulated genes, we used a differential screen (Wang and Brown 1991). Two hundred eighty-eight plasmids representing 82 of the 119 contigs were dot-blotted in 96 arrays and hybridized with equal amounts of four different radioactive probes. Two identical arrays were prepared for each probe to duplicate the experiment. A 'flagella regeneration' probe and an 'interphase' probe were prepared from FR and I SMART-amplified cDNA, respectively (referred to as forward and reverse unsubtracted probes, respectively); and a 'flagella regeneration-specific' and an 'interphase-specific' probes were prepared from FR-I and I-FR subtracted cDNA, respectively (referred to as forward and reverse subtracted probes, respectively). Unsubtracted cDNA probes contain representative amounts of the mRNA present in the cell in the different states (flagella regeneration or interphase), while subtracted probes theoretically represent sets of cDNA which are specific to each state. Normally, only mRNAs which are abundant in the cell produce detectable signals with the unsubtracted probes. Subtracted probes have the advantage of bypassing this problem although they may also result in a higher false-positive rate (Wang and Brown 1991).

Contig	Similar to :	Clones	Protein Id	E	Function
A	14-3-3-like protein (FTT)	42	77%	$3e^{-97}$	Cell wall formation in *C.r.* (Voigt and Frank 2003) Present in proteome of centriole (Keller et al. 2005), in human centrosome proteome (Andersen et al. 2003), and in proteome of the flagella (Pazour et al. 2005)
AP	Heat shock protein 90	2	76%	$4.e^{-55}$	Associates with axonemal proteins (Stephens and Lemieux 1999; Ostrowski, et al., 2002). Locates to the basal body (Keller et al. 2005) Present in proteome of the flagella (Pazour et al. 2005) Upregulated 1800% during FR in *C.r* (Stolc et al. 2005)
01c02	Outer arm dynein light chain 2 (oda-lc2)	1	57%	$4.e^{-20}$	Axoneme structural component Flagellar and basal body proteome (Li et al. 2004) Present in proteome of the flagella (Pazour et al. 2005) Upregulated 160% during FR in *C.r* (Stolc et al. 2005)
02d02	Flagellar associated protein 184 (FAP184)	1	48%	$2.e^{-26}$	Unknown function Flagellar and basal body proteome (Li et al. 2004) Present in proteome of the flagella (Pazour et al. 2005) Upregulated 200% during FR in *C.r* (Stolc et al. 2005)
Y	Flagellar associated protein 115 (FAP115)	10	48%	$6.e^{-13}$	Unknown function Present in proteome of the flagella (Pazour et al. 2005) Upregulated 460% during FR in *C.r* (Stolc et al. 2005)
L	Similar to C_2450124 (probable splicing factor)	3	63%	$8.e^{-40}$	Upregulated 190% during FR in *C.r* (Stolc et al. 2005)

Table 7: **Flagella-related clones found in the SSH + MOS library.**

This comparison of the probe hybridization signals should allow a quick identification of putative differentially expressed mRNAs: FR-upregulated cDNAs should i) produce a stronger hybridization signal with the forward unsubtracted probe than with the reverse unsubtracted probe, ii) produce a signal with the forward subtracted probe, and iii) return no signal at all with the reverse subtracted probe.

In our experiment, almost all clones produced a hybridization signals with all four probes suggesting that either a high level of unspecific binding occured, or that clones were all represented in each of the four probes. The latter is in my opinion the most probable explaination in accordance with the fact that RT-PCR detected the three FR-upregulated mRNAs of FTT, hsp90 and FAP115 in interphase cells as well (Fig. 14). We also observed that the FR subtracted probe bound in average the SSH clones stronger than the I subtracted probe (ratio of total binding: $FR/I_{subtracted}$=2,34) indicating that the subtraction was somehow effective. (Whether these clones represent truly upregulated genes remains to be confirmed.) In contrast, the average binding of the unsubtracted

probes was the same (ratio of total binding: $FR/I_{unsubtracted}$=0,96).

The calculated FR/I ratios for each type of probe (subtracted and unsubtracted) are presented in the annexes in table A, B, and C (pp. 64-66). Since several different clones corresponding to one contig were dot-blotted, we averaged the FR/I ratio for these contigs. The analysis of the FR/I ratios allowed us to select contigs representing candidate upregulated genes. We define as putatively upregulated contigs, all contigs with a differential signal FR/I superior to 2 with any of the probe types. An $FR/I_{(unsubtracted)}$ ratio equal to 2 should indicate a transcript which is two times more abundant in flagella-regenerating cells than in interphase cells, while an $FR/I_{(subtracted)}$ ratio equal to 2 indicates a transcript which is enriched in the FR-specific cDNA library compared with the interphase-specific cDNA library. This arbitrary treshold seemed significant for the following reasons: i) all clones for which induction was confirmed by RT-PCR are included, ii) all but one known flagellar genes (FAP184) are included.

We identified this way 35 putative upregulated contigs among the nuclear and unidentified cDNA (in bold in tables A&B, pp. 64-65), which can be further separated into two groups according to whether they produce a differential signal with both types of probes (11 contigs) or with only one type of probe (24 contigs). 50 % of the screened unknown contigs do not produce a hybridization signal, while only 27 % of the screened nuclear contigs do not. This difference supports our hypothesis that unknown contigs not only contain transcript from nuclear origin, but also chloroplastic or mitochondrial transcripts (of which 66% and 100%, respectively, of the screened contigs show no differential signals).

To confirm that the selected cDNAs corresponded to truly upregulated genes, we started to analyse transcript abundance by semi-quantitative RT-PCR. Besides the known genes induced during flagellar regeneration (Fig. 14), upregulation was confirmed for only one other clone of the library so far, namely clone 04a11, similar to a surface protein of *Leishmania tropica*, which consists of evenly spaced repetitive GTLPQQWSALTSLQ motifs (not shown). Other RT-PCR results are still preliminary.

C Analysis of transcript abundance during flagellar regeneration

1 Flagella structural components

The flagella regeneration library contained very few cDNAs known to be upregulated during flagellar regeneration in *Chlamydomonas*. Stolc et al. (2005) showed that the transcription of more than 200 genes is induced at least by a factor 2 during flagellar regeneration in *Chlamydomonas*. It is especially known for a long time that axonemal proteins such as tubulins are strongly upregulated during flagellar regeneration in *Chlamydomonas* (Lefebvre and al. 1980, Baker et al. 1984, Schloss et al. 1984). Perasso et al (2000) on the other hand noticed that tubulins were apparently not *de novo* synthesized during flagellar regeneration in *Scherffelia*. Therefore, I decided to investigate whether these cDNAs were missing in the library or were simply not upregulated during flagellar regeneration in *S. dubia*.

We compared the mRNA abundance between interphase cells and flagella regenerating cells of axonemal components which are not included in the sample of the SSH + MOS library analysed, but of which an EST was found in the interphase + flagellar regeneration cDNA library. Total RNA from interphase cells, 5 min flagellar regeneration and 200 min flagellar regeneration were hybridized with a β tubulin probe and a probe for the dynein light chain 16 kDa (Fig. 15). As shown, the level of β tubulin mRNA progressively increased, with a maximum after 30 min of flagellar regeneration and remained quite elevated throughout flagellar regeneration. The binding of mRNA for the dynein light chain 16 kDa also suggested a maximal accumulation at 30 min flagella regeneration, but the signal was too low and only binding at FR30 was detectable. Apparently, both mRNAs

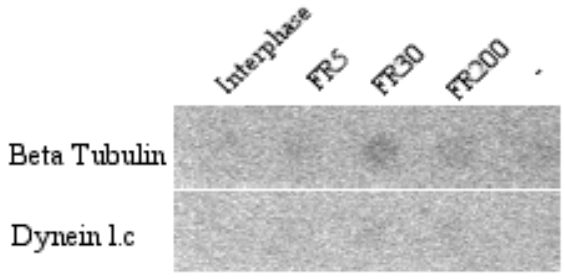

Figure 15. **Abundance change of β tubulin and dynein light chain 16 kDa mRNAs during flagellar regeneration assayed by dot blot**. Total RNA from interphase cells, cells at the begin (FR5) of flagellar regeneration, or after 30 min (FR30) and 200 min (FR3200) of flagellar regeneration were blotted (2 µg per dot) and hybridized with DNA probes for β tubulin or dynein light chain 16 kDa. (-): negative control (2 µg herring's sperm DNA)

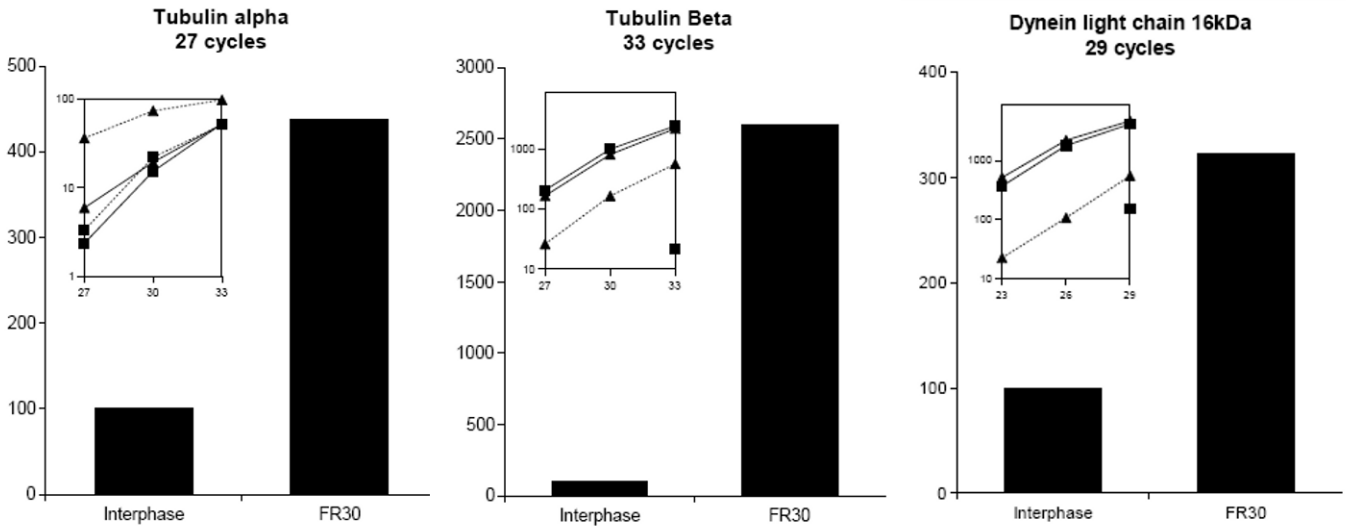

Figure 16. Semi-quantitative RT-PCR for α Tublin, β Tubulin, and outer arm dynein light chain 16 kDa. Targets were amplified simultaneously with an endogenous standard from total RNA (actin for α-tubulin, rbcs for dynein and β-tubulin). The efficiency of the reaction was checked at different cycles (inset). Product accumulation was measured by computer assisted imaging. Bars represent fold differences between end-products normalized to the endogenous standard after the indicated number of PCR cycles in one representative experiment. Insets: Product accumulation vs PCR cycle of endogenous control (plain line) or target product (dashed line), from interphase (■) or flagella-regenerating cells (▲).

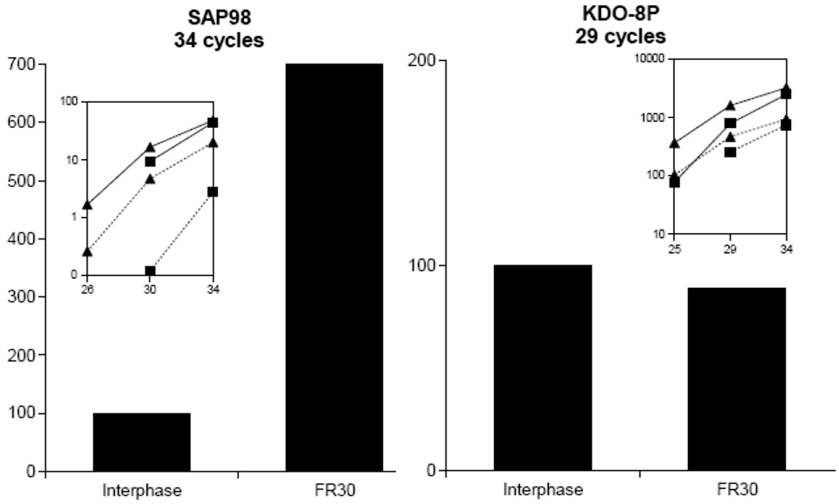

Figure 17. **Semi-quantitative RT-PCR for SAP98 and KDO-8P.** SAP98 and KDO-8P were amplified simultaneously with an endogenous standard (actin) from total RNA. The efficiency of the reaction was checked at different cycles (inset). Product accumulation was measured by computer assisted imagery. Bars represent fold differences between end-products normalized to the endogenous standard after the indicated number of PCR cycles in one representative experiment. Insets: Product accumulation vs PCR cycle of endogenous control (plain line) or target product (dashed line), from interphase (■) or flagella-regenerating cells (▲).

demonstrate the same pattern of upregulation until 30 min flagella regeneration with a return towards interphase level by the end of flagella regeneration.

To obtain more sensitive results, especially for the dynein light chain 16 kDa transcript, we used semi-quantitative RT-PCR. The expression of α tubulin was also studied. Figure 16 shows that the results with semi-quantitative RT-PCR are comparable with the Northern blot analysis. β tubulin strong upregulation during flagellar regeneration is confirmed as its mRNA is barely detectable in interphase cells, but very well amplified in flagella regenerating cells. α Tubulin and Dynein light chain 16 kDa mRNAs are both amplifiable from interphase cells, and become more abundant in flagella-regenerating cells. At 200 min they are slightly above interphase level (not shown). Therefore, the upregulation of the axonemal components during flagellar regeneration is conserved in *S. dubia* and one can wonder why these genes are not represented in the SSH+MOS subtracted library.

2 Proteins involved in scale biogenesis

We also analysed the regulation of genes involved in scale biogenesis for which an EST was available in *Scherffelia dubia* during flagellar regeneration. An EST for KDO-8P synthase was identified in the flagellar-regeneration + interphase library (Becker et al. 2001) and a piece of the cDNAsequence for SAP98 is presented in this thesis. The abundance of these mRNAs was also analysed by semi-quantitative RT-PCR.

As shown in figure 17, both mRNAs could be amplified from interphase cells suggesting a constitutive expression of the genes for scale metabolism. However, the expression of SAP98 during interphase seems to be very low as it is barely detectable after PCR (Fig. 17, inset, dashed line with squares). The two genes have different expression patterns: the mRNA of SAP98 becomes more abundant during flagellar regeneration than during interphase, while the mRNA of KDO-8P remains at a similar level. The difference between FR and interphase is about 6-fold for SAP98. It can also be noted that the mRNA is not detectable under 30 PCR amplification cycles in the interphase sample, suggesting that it is not very much abundant at all. Therefore we conclude that SAP98 is strongly upregulated during flagellar regeneration and that KDO-8P is not.

Discussion

SAP98 and scales associated proteins

About 20 glycoproteins can be isolated from the flagellar scales of the unicellular green alga *Scherffelia dubia*. In this thesis, I report the molecular cloning of the first of these scale-associated proteins in *S. dubia*, namely SAP98, a glycoprotein of apparent molecular weight of 98 kDa.

We cloned a cDNA fragment corresponding to the SAP98 protein using peptides sequenced by mass spectrometry. A degenerate oligonucleotide-primed PCR approach was initially used to identify the SAP genes but this approach did not allow the identification of candidate amplification products. There are several reasons why such an approach was not successfull: i) the possible absence or low expression of the SAPs mRNA during interphase and flagellar regeneration, ii) the lack of information on the primary sequence of the SAPs, iii) the background amplified by the degenerate oligonucleotides, iv) the ambiguities and/or errors in the peptides' mass spectrometry sequences. Altogether, these drawbacks make the identification of the target sequences extremely difficult.

Sequencing with Edman's degradation instead, if successfull, could have provided two significant advantages: i) a theoretically more reliable sequence, and ii) a valuable N-terminal peptide. Edman's degradation requires however substantial amount of purified protein that would have needed the harvesting of several hundred of liters of algae culture in the case of SAPs (all SAPs represent only 5% of the scales mass). It is anyway not possible to cope for differences between peptides and coding sequence owing to post-translational modifications or interruptions in the coding sequence when working at the genomic level.

About 80% (74 kDa) of the SAP98 mRNA sequence was finally obtained by RACE PCR after several steps. Although we do not present experimental evidence in this study that the isolated sequence encodes the SAP98 protein, the presence of many peptides from two different MS analyses in the translated sequence of the isolated cDNA is a strong argument in favor of this assumption. Further experiments are currently ongoing to provide more direct evidence: an antibody raised against a 45 kDa C-terminal piece of the cDNA expressed in bacteria recognized the SAP98 protein on an acrylamide gel and labelled the flagella of *Scherffelia dubia* (Becker B., personal communication).

SAP98 is a novel protein that bears no similarity with any other known protein or protein domain. This and the absence of similarities for any other SAP peptide suggest that scale associated proteins may all actually belong to an uncharacterized class of keto sugar acid-binding proteins which share no similarity with other known proteins. Indeed, genomes of

non-scaly, distantly related organisms were searched unsuccessfully (*Chlamydomonas reinhardtii*, *Arabidopsis thaliana*). Thus SAPs may only been found in scale-synthesizing algae and the question whether SAPs are conserved proteins in scaly algae naturally arises. Although SAPs were not found in the ESTs of the scaly green alga *Mesostigma viride*, a unique band was amplified in this alga with primers designed from the SAP98 cDNA (Becker B., personal communication*).* Since *M. viride* synthesizes basket-shaped scales (Domozych et al. 1991), one can conclude that SAPs can share similarities independently from the scale type they associate with. Thus, all SAPs from *S. dubia* may share similar domains which could explain why i) they do not appear in the subtracted library and ii) degenerate primers give several amplification products.

Obviously SAPs are not related with the extracellular proteins that constitute other well known types of ECM, such as the glycoproteins of the Chlorophyceae (*Chlamydomonas*, *Volvox*) or the land plant cell wall proteins. Indeed, SAP98 does not contain the hydroxyproline-rich and repetitive domains typical of these extracellular proteins, which are believed to encode the structural motifs allowing the proteins to form solid matrices (Cassab 1998). This is in agreement with the observation that SAPs do not fulfill the same structural functions as the above-mentioned proteins, rather they mediate scale-membrane and scale-scale interactions and could drive the assembly of scales in layers over the flagellar membrane. Nonetheless, some SAPs may possess inverted repeats that could explain the high amount of self-priming generated with genomic DNA and cDNA by some degenerated primers (SAP116 for instance).

SAP98 possesses a noteworthy high content in serine and threonine residues (25%). The hydroxyl groups that these amino acids carry could be involved in the association with scales by creating hydrogen bonds or divalent cation bridges with the hydroxyl groups of scale polysaccharides. This characteristic could be a common feature among SAPs. The presence of cations also appears necessary in order to negate the repulsion of negative charges between SAP98 and scale polysaccharides.

In vivo, the SAP98 protein forms a complex with SAP96 which remains to be localized. Two of six programs tested predict at least two transmembrane domains for SAP98. These predictions must of course be taken with caution. However, if these predictions are correct, it would suggest that the SAP96-98 complex is involved in the connection between scales and the flagellar membrane. Since the pentagonal scales are linked to the flagellar membrane by the high molecular weight complexes, the SAP98-SAP96 complex may be involved in connecting the hair scales with the flagellar membrane and the axoneme. Immunolocalization with anti-SAP98 should provide some answers regarding the localization and the function of this complex.

Subtracted library

The flagella regeneration-specific library constructed during this study suffers from

significant rRNA contamination, a high level of background cDNA amplification and an elevated redundancy. Moreover, a significant number of non-rRNA contigs could not be identified by similarity, further limiting the scope of the analysis. Nevertheless, potential differential clones were identified by way of a differential screen followed by RT-PCR confirmation for some of them. As a result, 29% of the clones (unknown + nuclear transcripts) analysed gave a differential signal, but beacause of the redundancy, only 8% corresponded to unique contigs. From the 35 putative upregulated contigs identified, semi-quantitative RT-PCR confirmed transcript upregulation for four of them (100 % of the contigs tested), and three others are known to be induced during flagellar regeneration in *Chlamydomonas reinhardtii*, thus so far 35% of the identified nuclear transcripts correspond to flagella regeneration-upregulated genes.

Although the library in flagella regeneration-upregulated genes appears less enriched than expected, it may still provide valuable qualitative information. Indeed, among the thirty five potentially upregulated genes, only five contigs were previously known to be upregulated during flagellar regeneration of which three have been identified for the first time in *S. dubia* (oda-lc2, fap184, fap115). Two thirds (24 contigs) remain unidentified and could therefore reveal novel proteins, such as scale associated proteins. However, as only non-overlapping *Rsa*I fragments of full length cDNA were isolated, one should expect some more redundancy to occur between known and unknown contigs, as they could represent coding and untranslated regions, respectively, of the same transcripts.

The aim of the subtracted library was to investigate flagellar regeneration in a scale-synthesizing alga. The high proportion of unidentified contigs makes it difficult to discuss the physiological state of the cell during flagellar regeneration. The specificity of *S. dubia* when compared to other flagella regeneration model systems is that it builds a new extracellular matrix for the nascent flagella. Accordingly, among the eleven upregulated contigs for which protein identification was available, three encode enzymes involved in sugar metabolisms, and two are similar to cell surface-associated proteins. It is rather expected to observe upregulation of proteins involved in sugar metabolism in a cell that starts to produce large amounts of polysaccharides. Indeed, two of these genes are involved in scale polysaccharides synthesis: GDP-mannose-3,5-epimerase and UDP-glucose dehydrogenase which are enzymes of the biosynthesis pathways of gulose (Wolucka and Van Montagu 2003) and galacturonic acid, respectively, two components of the scale polysaccharides (Becker et al. 1989).

It is difficult to understand why so few conserved flagella regeneration-upregulated genes are present in the library. For instance, my experiments demonstrated that genes encoding three axonemal proteins not found in the analysed sample of the library are actually upregulated during flagellar regeneration in *S. dubia*. I see several likely and non-exclusive explainations for their abscence in the flagella regeneration-specific library:

1) The analysis of transcript abundance and the differential screen revealed that

almost all flagellar regeneration-upregulated genes were expressed during interphase as well. When differences between the two mRNA populations are small (mostly quantitative instead of qualitative), or when complex genomes are hybridized, the amplification of background sequences increases in SSH to the point that differential transcripts can be masked (Rebrikov et al. 2005, Li et al. 2005). For instance, this is the case for IFT proteins which are expressed during interphase, as flagella continuously assemble and disassemble, and are upregulated during flagellar regeneration.

2) The fact that we used a mixture of total RNA extracted at the begining of flagellar regeneration and after 30 min of flagellar regeneration helped reduce the differences flagellar regeneration and interphase, resulting in a lesser entichment in differential cDNAs. Indeed, for all the upregulated genes I tested I found a slight increase of transcript abundance at the beginning of flagellar regeneration and a maximum at 30 min flagellar regeneration (not shown).

1) Transcripts with widely conserved domains or transcripts which belong to families from which members are expressed during interphase may have had their similar coding regions hybridized with driver molecules and be subtracted. Only the more specific untranslated regions of these transcripts would have been conserved during hybridization but since this part of the sequence might be specific to *Scherffelia*, it may not allow their proper identification by similarity search.

A preliminary amplification experiment did detect the β-tubulin and SAP98 cDNAs in the FR-I sample although we did not isolate any clone with these sequences, suggesting that they are apparently not enough represented in the library. We could not determine whether the subtraction actually resulted in an enrichment in those cDNA species.

Besides, the amplification of ribosomal RNA during SSH could have influenced the outcome of the library in three ways: i) competition with mRNA during SMART first-strand cDNA synthesis and amplification thus reducing gene representation, ii) the hybridization conditions were not optimized to normalize such abundant species (Gurskaya et al. 1996) thus they remained abundant, iii) competition with the differential cDNA hybrids during the amplification of the enriched-fraction which followed the second hybridization thus reducing the representation of low abundant transcripts.

The differential screen also suggested that some rRNA species were enriched in the FR-I sample compared with the I-FR sample (see FR/I_{sub} ratios in table C, p66). Whether this enrichment is an experimental artefact or truly represents a physiological difference remains to be verified. A more efficient reverse transcription of rRNA in the FR sample could result for instance from a higher abundance of rRNA species, or a higher polyadenylation rate in chloroplast and mitochondria (which actually leads to their rapid degradation by PNPase, [Herrin and Nickelsen 2004, Holec et al. 2006]).

Differential screen

The differential screen is a rapid method to identify putative differential candidates, however genuine upregulation must be demonstrated individually with direct experiments such as RT-PCR or Northern hybridization. In this study we only confirmed the upregulation of four of the 35 candidates (Tables A&B pp 64-65). The fact that almost all clones returned a hybridization signal with every probe also reflects the possibility that only few qualitative differences existed between the two subtracted samples, consistent with the observation that flagellar regeneration-upregulated mRNAs are also present in inteprhase cells

It should be noted that the interpretation of the differential screen with the unsubtracted probes is an underestimation of the number of differential clones. Indeed, the flagellar regeneration unsubtracted probe was made using a mixture of 5 min and 30 min flagellar regeneration cDNAs. Since my experiments always showed a maximum of upregulation at 30 min (Fig 14, 15, personal observations), had I used a probe only made with 30 min flagellar regeneration cDNA, the FR/I_{unsub} ratios would certainly have been enhanced.

Transcript abundance

The multiplex RT-PCR performed in this study were not designed for the absolute quantification of mRNA in the cell. However, assuming similar amplification efficiencies, we can tentatively group the transcripts in three abundance classes according to the number of PCR cycles needed to detect the transcripts during the different state of the cell (Table 8).

	Higher abundance class	Medium abundance class	Lower abundance class
Invariant	Rbcs HSP90	Actin FTT α-tubulin Dynein 16 kDa	Surface antigen P2 SAP98 KDO-8P synthase
During FR	β-tubulin	FAP115	
During Interphase			β-tubulin FAP115

Table 8. Abundance of transcripts during interphase and flagellar regeneration suggested by RT-PCR product accumulation.

Since the synthesis of SAPs has to be coordinated, the transcript abundance of SAP98 gives a hint concerning the expression of the SAPs: SAPs could be expressed at very low levels during interphase and upregulated during flagellar regeneration or mitosis. The constitutive expression of SAP98 and KDO-8P synthase is consistent with the observation that a few flagellar scales are always seen in the Golgi cisternae during interphase. Actually, one clone containing an insert with the sequence of SAP98 was isolated the interphase + flagella regeneration library (clone 122), but its sequence was not of a

sufficient quality to enable its identification with the MS peptides.

Both SAP98 and KDO-8P synthase mRNAs show different reponses to experimental deflagellation. The abundance level of the SAP98 mRNA shows a significant increase which correlates well with the amount of scales synthesized and the induction of other flagellar components, while the KDO-8P synthase remains at a constant level. Particularly, the upregulation of SAP98 during the first 30 min of flagellar regeneration seems to match the increased rate of scale synthesis (Becker et al. 1995). The SAP98 protein must associate with the scales, and it seems therefore logical that its expression is correlated with the scale biosynthesis activity of the cell. On the other hand KDO-8P synthase is an enzyme which activity may be regulated at the protein level, for instance by phosphorylation or by the product/substrate concentration. Although the synthesis of flagellar scales is induced compared with interphase, the basal level of KDO-8P synthase may cope with the requirement for KDO precursors during flagellar regeneration, thereby not needing a change of concentration. As a comparison, the synthesis of a new theca during mitosis requires 10 times higher KDO synthesis activity than during flagellar regeneration which may then necessitate the upregulation of this gene during mitosis, or imply that the concentration of the enzyme is sufficient during flagellar regeneration.

Our data, combined with the neo-synthesis assay performed during flagellar regeneration of Perasso et al. (2000), indicate however that some transcripts are induced without synthesis of the corresponding proteins (α, β tubulin and SAP98). Why are some transcripts upregulated but not translated ? On the one hand, we know from studies in *Chlamydomonas* that upregulation is a response to an intracellular signal involving calcium that is either generated by deflagellation, or is a branch in the signalling pathway that triggers deflagellation (Quarmby 2004). This signal is transduced to the nucleus where it leads to the activation of the trancription of the genes containing the deflagellation response element (Kang and Mitchell 1998). Several studies showed that the resulting gene upregulation is independent of flagella synthesis (Cheshire and Keller 1991), of the pool of flagella precursor mRNA present in the cell (Baker et al. 1984), or of protein synthesis (Baker et al. 1986). On the other hand, we know that protein synthesis is required for flagellar regeneration in both *C. reinhardtii* and *S. dubia* (Farell 1976, Perasso et al. 2000), and that tubulins are not a limiting factor, at least in *C. Reinhardtii,* because the cell draws from a large cytoplasmic pool (Farell 1976). Similarly, Perasso et al. (2000) showed the existence of a pool of SAP126 localized at the plasma membrane of *S. dubia* and calculated that this pool was large enough to allow several consecutive rounds of regeneration of the flagella. Therefore we hypothesize that the initial burst of mRNA abundance is a systematic response to a signal which follows deflagellation, allowing the cell to rapidly accumulate a set of mRNA necessary for the synthesis of new flagella. The effective translation of these upregulated and stabilized mRNA is regulated by the depletion of the pre-deflagellation flagella precursor pool, so that precursor flagellar proteins in large amounts need not be synthesized. The decay of the accumulated mRNA during flagella

regeneration is then regulated probably by nuclease degradation (Baker et al. 1984, Baker 1993).

Perspectives

It is first necessary to confirm that the cDNA isolated during this study encodes the SAP98 protein. This is now possible with an antibody raised against the C-terminal part of the translated cDNA (Becker, personal communication). Immunolocalization of the SAP98-SAP96 complex in the flagella could then follow and reveal which scale type this protein complex associates with. Additionally, the behavior of the SAP98 and SAP126 proteins (and by extension the SAP98-SAP96 and HMW complexes) during flagellar regeneration could be compared. This would hopefully add new information on how the scales and/or scale layers assemble, and how scale biosynthesis is regulated during flagella synthesis.

The complete sequence of the SAP98 protein still has to be uncovered. This is typically done by 5-'RACE PCR, but this technique has not been successfull yet in the case of SAP98 (this study and Becker, personal communication). Other possible strategies include sequencing of the N-terminal part of the protein with Edman' degradation, inverse PCR or genome walking. This first sequence of a scale associated protein could also be used to verify the existence of homologues in other scaly organisms such as *Tetraselmis*, *Mesostigma* or *Pyramimonas*. PCR with SAP98 specific primers apparently amplified unique products in *Mesostigma* and *Pyramimonas* suggesting that despite their different scale types, scale associated proteins are conserved in scaly green algae (Becker, personal communication). Thus, the comparison of homologous scale associated proteins could indicate if particular features, domains or residues are conserved. Since the SAP98 provides a new and novel example of protein-sugar interactions, it would be especially interesting to know how SAP98 interacts with keto-sugar acids as potential applications may be found in anti-bacterial production for instance.

In this thesis we also present a novel PCR-based approach using short, degenerated primers to characterize new proteins for which only a few peptide sequences are known. It should be noted however that this procedure is lenghty, requiring several PCR optimization steps. Although we were unfortunately not able to repeat it successfully for other SAP proteins, this procedure could in theory be reproduced for other proteins for which traditional techniques are unfruitful.

Besides the sequence of the SAP98, we observed that its mRNA was constitutively expressed and upregulated during flagellar regeneration, and that its cDNA was present in the flagellar regeneration subtracted library. If we assume that the other SAPs follow the same pattern of expression, then the possibility exists that the cDNAs for these SAPs are present in the upregulated unidentified contigs of the FR subtracted library. These contigs could be used as a probes to screen a full length cDNA library. This should allow

the identification of the yet unknown upregulated contigs, and the possible discovery of novel proteins involved in flagellar regeneration in *S. dubia*, from which SAPs could be part.

Furthermore, as we could detect the β tubulin and SAP 98-2 transcripts by PCR amplification from the subtracted cDNA population, sequencing of a larger sample of the flagellar regeneration library should also reveal additional differential clones.

I hope thatthe results provided in this thesis have laid the ground for further research on the characterization of scaly extracellular matrices.

Annexes

Characterization of 449 clones from a flagella regeneration-specific library

Contig	Clones	Length	F/I sub	F/I unsub	Genome	Function	Best hit (<1.e-10)	*	RT
05e01	1	242	**++**	1,17	unknown	unknown	no hit		
04d04	1	340	**12,13**	**4,6**	unknown	unknown	no hit		
04e02	1	592	**5,77**	1,54	unknown	unknown	no hit		
H	2	900	**3,82**	**2,21**	unknown	unknown	no hit		
AT	4	411	**3,7±0,47**	1,48±0,04	unknown	unknown	no hit		
04f01	1	535	**3,22**	0,8	unknown	unknown	no hit		
AY	3	262	**3,19**	**67,38**	unknown	unknown	no hit		
02d03	1	290	**3,15**	**2,11**	unknown	unknown	no hit		
AK	3	521	**2,97**	0,83	unknown	unknown	no hit		
01f05	1	475	**2,82**	**2,03**	unknown	unknown	no hit		
05g09	1	517	**2,8**	1,17	unknown	unknown	no hit		
01d06	1	557	**2,57**	**2,21**	unknown	unknown	no hit		
05b06	1	787	**2,47**	0,58	unknown	unknown	no hit		
AO	2	480	**2,38±0,81**	**4,62±11,23**	unknown	unknown	no hit		
R	2	791	**2,11**	0,61	unknown	unknown	no hit		
04d03	1	422	**2,07**	**2,28**	unknown	unknown	no hit		
02d04	1	519	**2,06**	**3,92**	unknown	unknown	no hit		
05f05	1	379	1,87	0,62	unknown	unknown	no hit		
AF	2	627	1,73	**2,56**	unknown	unknown	no hit		
F	5	748	1,72±0,18	1,91	unknown	unknown	no hit		
AE	2	634	1,72	0,58	unknown	unknown	no hit		
Q	20	832	1,71±0,45	0,89±0,63	unknown	unknown	no hit		
AZ	3	179	1,67	0,96	unknown	unknown	no hit		
AC	2	685	1,55±0,32	1,86±1,03	unknown	unknown	no hit		
02d06	1	553	1,47	**2,79**	unknown	unknown	no hit		
04g09	1	512	1,43	1,08	unknown	unknown	no hit		
01c08	1	442	1,36	1,63	unknown	unknown	no hit		
04b01	1	228	1,31	**2,91**	unknown	unknown	no hit		
04a09	1	470	1,28	**2,2**	unknown	unknown	no hit		
02b12	1	412	1,24	**++**	unknown	unknown	no hit		
04a06	1	269	1,01	**2,27**	unknown	unknown	no hit		
01f10	1	764	0,97	**2,79**	unknown	unknown	no hit		
01a05	1	603	0,81	1,52	unknown	unknown	no hit		
K	3	640	0,76±0,5	0,85±0,76	unknown	unknown	no hit		
05f10	1	439	0,69	0,45	unknown	unknown	no hit		
05e08	1	822	0,51	0,34	unknown	unknown	no hit		
01c11	1	674	0,48	0,64	unknown	unknown	no hit		
01h04	1	592	0,18	1,65	unknown	unknown	no hit		
02g03	1	715			unknown	unknown	no hit		
bish01	1	220			unknown	unknown	no hit		
bisc09	1	435			unknown	unknown	no hit		
02g12	1	409			unknown	unknown	no hit		
02g11	3	728			unknown	unknown	no hit		
bisc01	1	589			unknown	unknown	no hit		
bise11	1	430			unknown	unknown	no hit		
Bise05	1	414			unknown	unknown	no hit		
03d01	1	651			unknown	unknown	no hit		
AU	2	407			unknown	unknown	no hit		
AH	2	588			unknown	unknown	no hit		
03f08	1	598			unknown	unknown	no hit		
02h01	1	687			unknown	unknown	no hit		
04f04	1	875			unknown	unknown	no hit		
01c06	1	612			unknown	unknown	no hit		
03h03	1	476			unknown	unknown	no hit		
02d12	1	262			unknown	unknown	no hit		

Table A. Unidentified contigs isolated from the FR-I subtracted cDNA library

F/I sub: ratio of forward/reverse subtracted probes, F/I unsub: ratio of forward/reverse unsubtracted probes

++: infinity, n.s. : no signal

* : upregulated >100% in Stolc et al. (2005), or identified by >5 peptides in the flagella proteome (Pazour et al. 2005) or present in FABB (Li et al. 2004)

RT : upregulation confirmed by comparative RT-PCR

Contig	Clones	Length	F/I sub	F/I unsub	Genome	Function	Best hit (<1.e-10)	*	RT
04a11	1	876	**16,99**	n.s.	nucleus	Cell wall	**Similar to surface antigen p2**		+
Y	18	737	**10,84±7,85**	1,61±1,38	nucleus	unknown	**FAP115**	*	+
AP	2	657	**4,57±0,21**	**2,31±1,19**	nucleus	transport	**Hsp 90**	*	+
AN	3	488	**3,22**	0,6	nucleus	unknown	**Hypothetical protein**		
L	3	695	**3,29±1,31**	**2,72±1,44**	nucleus	translation	**C_240124 (Probable splicing factor)**	*	
05a01	1	605	**2,44**	0,71	nucleus	Cell wall	**C_470025 (Cell wall protein)**		
W	6	772	**2,15±0,51**	1,03±0,14	nucleus	sugar metabolism	**GDP-mannose-3,5-epimerase**		
A	42	955	**2,0±1,04**	0,81±0,27	nucleus	Cell wall	**14-3-3-like protein**	*	+
AW	2	362	1,62	0,96	nucleus	Protein degradation	26S proteasome rpt6a subunit		
AI	3	550	1,34±0,4	**2,99±2,67**	nucleus	sugar metabolism	**udp-glucose dehydrogenase**		
01e11	1	549	1,27	**2,41**	nucleus	sugar metabolism	**adp-glucose pyrophosphorylase small subunit**		
01c02	1	338	1,24	**++**	nucleus	Flagella	**outer arm dynein light chain 2**	*	
U	9	674	1,2±0,24	0,84±0,34	nucleus	unknown	Hypothetical protein		
D	3	667	1,07	0,97	nucleus	translation	50S rp L16		
02d02	1	739	n.s.	0,67	nucleus	unknown	FAP184	*	
bisf01	1	462			nucleus	Various	thioredoxin-like protein		
bisd06	1	613			nucleus	sugar metabolism	phosphoglucomutase cp precursor		
bise06	1	273			nucleus	translation	40S ribosomal prot S3		
01f07	1	630			nucleus	lipid metabolism	acyl-coa synthetase		
03g07	1	558			nucleus	translation	protein translation factor sui1 homolog		
AG	2	618	2,92	0,94	chloroplast	photosynthesis	PSII 44kDa Rc prot.		
04c08	1	877	2,54	n.s.	chloroplast	photosynthesis	CP43		
AA	9	709	1,82±0,35	2,33±2,57	chloroplast	photosynthesis	PSII D2		
04g07	1	695	1,79	1,21	chloroplast	photosynthesis	Psab		
BA	2	782	0,81±0,05	0,75±0,03	chloroplast	translation	50S rp L12		
O	4	1097	0,71±0,03	0,54±0,17	chloroplast	translation	elongation factor tu		
S	13	858	0,51±0,14	0,4±0,79	chloroplast	photosynthesis	Atp synthase ch a precursor		
AB	2	705	0,41	1,48	chloroplast	translation	translation initiation factor if-1 cp		
01d04	1	750	0,3	1,45	chloroplast	translation	50S rp L12		
bisf09	1	531			chloroplast	photosynthesis	CP-47		
03g04	1	627			chloroplast	photosynthesis	Atp synthase a subunit iv		
01g08	1	540			chloroplast	photosynthesis	photosystem i p700 apoprotein a2		
bisc11	1	608			chloroplast	photosynthesis	li-por subunit b		
bisc04	1	401			chloroplast	photosynthesis	RbcL		
BB	2	763	1,97±0,01	1,08±0,38	mitochondrion	respiration	Nadh-ubiquinone-oxydoreductase		
V	2	776	1,8±1,74	1,74	mitochondrion	respiration	Cytochrome c oxidase subunit i		
AD	4	634	1,65±0,53	0,94±1,46	mitochondrion	respiration	Cytochrome c oxydase polypeptide iii		
05b12	1	875	1,34	0,57	mitochondrion	respiration	nadh dehydrogenase subunit 9		
P	3	763	1,3±0,18	1,05±0,33	mitochondrion	respiration	nadh-ubiquinone oxydoreductase		
J	6	937	1,02±0,91	0,81±0,93	mitochondrion	respiration	f1-atpase alpha subunit		
bisc06	1	571			mitochondrion	respiration	nadh-ubiquinone oxydoreductase 49kDa subunit		
AV	2	776			mitochondrion	respiration	cytochrome oxidase subunit i		
04e07	1	897			mitochondrion	respiration	atpase alpha subunit		

Table B. Contigs of the FR-I subtracted cDNA library identified by similarity ordered by genome origin

F/I sub: ratio of forward/reverse subtracted probes, F/I unsub: ratio of forward/reverse unsubtracted probes

++: infinity, n.s. : no signal

* : upregulated >100% in Stolc et al. (2005), or identified by >5 peptides in the flagella proteome (Pazour et al. 2005) or present in FABB (Li et al. 2004)

RT : upregulation confirmed by comparative RT-PCR

Contig	Clones	Length	F/I sub	F/I unsub	Genome	Function	Best hit (<1.e-10)	*	RT
AJ	3	538	n.s.	0,18	rRNA nuc	rrna	LSU rRNA 25S 2620-3150		
AQ	12	419	7,13±2,98	2,71±7,67	rRNA nuc	rrna	28S rRNA 246-721		
G	15	1045	2,78±1,19	1,02±0,55	rRNA nuc	rrna	28S rRNA 725-1750		
N	3	873	1,02±0,97	0,94±0,78	rRNA nuc	rrna	16S rRNA 679-1493		
AS	2	414			rRNA nuc	rrna	16S rRNA 1271-1610		
AX	2	294			rRNA nuc	rrna	28S rRNA		
02h10	1	307			rRNA nuc	rrna	16S rRNA 198-498		
AL	11	677	7,22±3,52	2,07±1,3	rRNA cp	rrna	cp rrnL 920-1421		
01d09	1	369	5,97	1,16	rRNA cp	rrna	meso2a25c10		
T	58	783	4,74±1,59	1,16±0,74	rRNA cp	rrna	cp 16S rRNA		
M	9	686	2,44±0,79	1,22±1,4	rRNA cp	rrna	cp rrnL 410-4...		
B	27	907	2,29±0,87	1,24±0,87	rRNA cp	rrna	RrnL 1627-2532 23S RNA		
01e02	1	394	1,4	6,2	rRNA cp	rrna	meso2a25c10		
E	5	779	0,91±0,57	1,6±1,19	rRNA cp	rrna	cp rrnL 2594-3330		
X	5	834			rRNA cp	rrna	SSU rRNA		
02f03	1	183			rRNA cp	rrna	Meso2a25c10 (chl DNA)		
C	16	1005	3,56±0,63	0,98±0,49	rRNA mit	rrna	mit rRNA		
01h12	1	497	2,11	1,86	rRNA mit	rrna	mit rRNA		
AM	3	428			rRNA mit	rrna	Meso2a19h05 (mit DNA)		
Z	10	718	2,35±0,66	0,92±0,16	rRNA bact	rrna	16S rRNA		
03g02	1	609			unclear	unclear	putative DNA endonuclease or RNA or unknown		

Table C. Contigs containing rRNA or with unclear identification ordered by genome origin

F/I sub: ratio of forward/reverse subtracted probes, F/I unsub: ratio of forward/reverse unsubtracted probes

++: infinity, n.s. : no signal

* : upregulated >100% in Stolc et al. (2005), or identified by >5 peptides in the flagella proteome (Pazour et al. 2005) or present in FABB (Li et al. 2004)

RT : upregulation confirmed by comparative RT-PCR

References

Afonina I, Zivarts M, Kutyavin I, Lukhtanov E, Gamper H et al. (1997) Efficient priming of PCR with short oligonucleotides conjugated to a minor groove binder. Nucleic Acids Res 25: 2657-2660.

Aken E, Pienaar RN (1985) Preliminary investigations on the chemical composition of the scale boundary and cyst wall of *Pyramimonas pseudoparkeae* (Prasinophyceae). S Afr J Bot 51: 408-416.

Andersen JS, Wilkinson CJ, Mayor T, Mortensen P, Nigg EA et al. (2003) Proteomic characterization of the human centrosome by protein correlation profiling. Nature 426: 570-574.

Baker EJ, Schloss JA, Rosenbaum JL (1984) Rapid changes in tubulin rna synthesis and stability induced by deflagellation in *Chlamydomonas*. J Cell Biol 99: 2074-2081.

Baker EJ, Keller LR, Schloss JA, Rosenbaum JL (1986) Protein synthesis is required for rapid degradation of tubulin mRNA and other deflagellation-induced RNAs in *Chlamydomonas reinhardtii*. Mol Cell Biol 6: 54-61.

Baker EJ (1993) Tubulin mrna instability and stabilization by protein synthesis inhibitors are reproducible in nontranslating extracts from *Chlamydomonas*. Dev Genet 14: 460-470.

Baker EJ, Schloss JA, Rosenbaum JL (1984) Rapid changes in tubulin RNA synthesis and stability induced by deflagellation in *Chlamydomonas*. J Cell Biol 99: 2074-2081.

Becker B, Hard K, Melkonian M, Kamerling JP, Vliegenthart JF (1989) Identification of 3-deoxy-*manno*-2-octulosonic acid, 3-deoxy-5-o-methyl-*manno*-2-octulosonic acid and 3-deoxy-*lyxo*-2-heptulosaric acid in the cell wall (theca) of the green alga *Tetraselmis striata* Butcher (prasinophyceae). Eur J Biochem 182: 153-160.

Becker B, Becker D, Kamerling J P, Melkonian M (1991) 2-keto-sugar acids in green flagellates: a chemical marker for prasinophycean scales. J Phycol 27: 498-504.

Becker B, Marin B, Melkonian M (1994) Structure, composition and biogenesis of prasinophyte cell coverings. Protoplasma 181: 233-244.

Becker B, Bölinger B, Melkonian M (1995) Anterograde transport of algal scales through the Golgi complex is not mediated by vesicles. Trends Cell Biol 5: 305-308.

Becker B, Melkonian M (1996) The secretory pathway of protists: spatial and functional organization – evolution. Microbiol. Rev 60: 697-721.

Becker B, Perasso L, Kammann A, , Salzburg, M, Melkonian M (1996) Scale-associated glycoproteins of *Scherffelia dubia* (chlorophyta) form high-molecular-weight complexes between the scale layers and the flagellar membrane. Planta 199: 503-510.

Becker B, Melkonian M (1998) The cell wall (theca) of *Tetraselmis striata* (Chlorophyta) : macromolecular composition and structural elements of the complex polysaccharides. J Phycol 34: 779-787.

Becker B, Feja N, Melkonian M (2001) Analysis of expressed sequence tags (ESTs) from the scaly green flagellate *Scherffelia dubia* Pascher emend. Melkonian et Preisig. Protist 152: 139-147.

Becker D, Becker B, Satir P, Melkonian M (1990) Isolation, purification, and characterization of flagellar scales from the green flagellate *Tetraselmis striata* (Prasinophyceae). Protoplasma 156: 103-112.

Becker D, Melkonian M (1992) N-linked glycoproteins associated with flagellar scales in a flagellate green alga: characterization of interactions. Eur J Cell Biol 57: 109-116.

Bhattacharya D, Weber K, An SS, Berning-Koch W (1998) Actin phylogeny identifies *Mesostigma* viride as a flagellate ancestor of the land plants. J Mol Evol 47: 544-550.

Brown RM (1974) The golgi apparatus and endomembrane system: its role in the biosynthesis, transport, and secretion of cell wall constituents in *Pleurochrysis*. Port Acta Biol 14: 369-384.

Cassab GI (1998) Plant cell wall proteins. Annu Rev Plant Physiol Plant Mol Biol 49: 281-309.

Celi FS, Zenilman ME, Shuldiner AR (1993) A rapid and versatile method to synthesize internal standards for competitive PCR. Nucleic Acids Res 21: 1047.

Cheshire JL, Keller LR (1991) Uncoupling of *Chlamydomonas* flagellar gene expression and outgrowth from flagellar excision by manipulation of ca2+. J Cell Biol 115: 1651-1659.

Cheshire JL, Evans JH, Keller LR (1994) Ca2+ signaling in the *Chlamydomonas* flagellar regeneration system: cellular and molecular responses. J Cell Sci 107 (Pt 9): 2491-2498.

Delmas F, Petit J, Joubes J, Seveno M, Paccalet T et al. (2003) The gene expression and enzyme activity of plant 3-deoxy-d-*manno*-2-octulosonic acid-8-phosphate synthase are preferentially associated with cell division in a cell cycle-dependent manner. Plant Physiol 133: 348-360.

Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F et al. (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci U S A 93: 6025-6030.

Diatchenko L, Lukyanov S, Lau YF, Siebert PD (1999) Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes. Methods Enzymol 303: 349-380.

Domozych DS, Stewart KD, Mattox KR (1981) Development of the cell wall in *Tetraselmis*: role of the Golgi apparatus and extracellular wall assembly. J Cell Sci 52: 351-371.

Domozych DS, Wells B, Shaw PJ (1991) Basket scales of the green alga, *Mesostigma viride*: chemistry and ultrastructure. Journal of Cell Science 100: 397-407.

Farrell KW (1976) Flagellar regeneration in *Chlamydomonas reinhardtii*: evidence that cycloheximide pulses induce a delay in morphogenesis. J Cell Sci 20: 639-654.

Gilliland G, Perrin S, Blanchard K, Bunn HF (1990) Analysis of cytokine mRNA and DNA: detection and quantitation by competitive polymerase chain reaction. Proc Natl Acad Sci U S A 87: 2725-2729.

Goldman R, Kohlbrenner W, Lartey P, Pernet A (1987) Antibacterial agents specifically inhibiting lipopolysaccharide synthesis. Nature 329: 162-164.

Gurskaya NG, Diatchenko L, Chenchik A, Siebert PD, Khaspekov GL et al. (1996) Equalizing cDNA subtraction based on selective suppression of polymerase chain reaction: cloning of jurkat cell transcripts induced by phytohemaglutinin and phorbol 12-myristate 13-acetate. Anal Biochem 240: 90-97.

Herrin D, Nickelsen J (2004) Chloroplast RNA processing and stability. Photosynth Res 82: 301-314.

Hirano M, Kiyohara H, Yamada H (1994) Existence of a rhamnogalacturonan II-like region in bioactive pectins from medicinal herbs. Planta Med 60: 450-454.

Holec S, Lange H, Kuhn K, Alioua M, Borner T et al. (2006) Relaxed transcription in arabidopsis mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase. Mol Cell Biol 26: 2869-2876.

Kang Y, Mitchell DR (1998) An intronic enhancer is required for deflagellation-induced transcriptional regulation of a *Chlamydomonas reinhardtii* dynein gene. Mol Biol Cell 9: 3085-3094.

Keller LR, Schloss JA, Silflow CD, Rosenbaum JL (1984) Transcription of alpha- and beta-tubulin genes in vitro in isolated *Chlamydomonas reinhardtii* nuclei. J Cell Biol 98: 1138-1143.

Keller LC, Romijn EP, Zamora I, Yates JR3, Marshall WF (2005) Proteomic analysis of isolated *Chlamydomonas* centrioles reveals orthologs of ciliary-disease genes. Curr Biol 15: 1090-1098.

Kochert G (1978) Protein determination by dye binding.. In: Hellebust JA, Craigie JS, editors. Handbook of physiological methods. Physiological and biochemical methods. Cambridge University Press. pp. 91-93.

Lefebvre PA, Silflow CD, Wieben ED, Rosenbaum JL (1980) Increased levels of mRNAs for tubulin and other flagellar proteins after amputation or shortening of *Chlamydomonas* flagella. Cell 20: 469-477.

Lemieux C, Otis C, Turmel M (2000) Ancestral chloroplast genome in I reveals an early branch of green plant evolution. Nature 403: 649-652.

Lemieux C, Otis C, Turmel M (2007) A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the streptophyta in chloroplast genome-based phylogenies. BMC Biol 5: 2.

Li JB, Gerdes JM, Haycraft CJ, Fan Y, Teslovich TM et al. (2004) Comparative genomics identifies a flagellar and basal body proteome that includes the bbs5 human disease gene. Cell 117: 541-552.

Li L, Techel D, Gretz N, Hildebrandt A (2005) A novel transcriptome subtraction method for the detection of differentially expressed genes in highly complex eukaryotes. Nucleic Acids Res 33: e136.

Lukyanov KA, Launer GA, Tarabykin VS, Zaraisky AG, Lukyanov SA (1995) Inverted terminal repeats permit the average length of amplified DNA fragments to be regulated during preparation of cDNA libraries by polymerase chain reaction. Anal Biochem 229: 198-202.

Manton I (1965) Observations on scale production in *Pyramimonas amylifera* conrad. J Cell Sciences 1: 429-438.

Manton I, Parke M (1965) Observations on the fine structure of two species of *Platymonas* with special reference to flagellar scales and the mode of origin of the theca. J mar Biol Ass UK 45: 743-754.

Marin B, Matzke C, Melkonian M (1993) Flagellar hairs of *Tetraselmis* (Prasinophyceae): ultrastructural types and intrageneric variation. Phycologya 32: 213-222.

Marin B, Melkonian M (1999) Mesostigmatophyceae, a new class of streptophyte green algae revealed by SSU rRNA sequence comparisons. Protist 150: 399-417.

Matsuura K, Miyagawa I, Kobayashi M, Ohta D, Matoh T (2003) *Arabidopsis* 3-deoxy-d-*manno*-oct-2-ulosonate-8-phosphate synthase: cDNA cloning and expression analyses. J Exp Bot 54: 1785-1787.

McFadden GI, Melkonian M (1986) Golgi apparatus activity and membrane flow during scale biogenesis in the green flagellate *Scherffelia dubia* (Prasinophyceae) i. flagellar regeneration. Protoplasma 130: 186-198.

McFadden GI, Preisig HR, Melkonian M (1986) Golgi apparatus activity and membrane flow during scale biogenesis in the green flagellate *Scherffelia dubia* (Prasinophyceae) ii. cell wall secretion and assembly. Protoplasma 131: 174-184.

McPherson MJ, Jones KM, Gurr SJ (1991) PCR with highly degenerate primers. In: McPherson MJ, Quirke P, Taylor GR, editors. PCR : a practical approach. Oxford University Press. pp. 171-186.

Melkonian M (1982) Effect of divalent cations on flagellar scales in the green flagellate *Tetraselmis cordiformis*. Protoplasma 111: 221-233.

Melkonian M, Preisig HR (1986) A light and electron microscopic study of *Scherffelia dubia*, a new member of the scaly green flagellates (Prasinophyceae). Nord. J. Bot 6: 235-256.

Melkonian M (1990) Class Prasynophyceae.. In: Margulis L, Corliss JO, Melkonian M, Chapman DJ, editors. Handbook of protoctista. Jones and Barlett Publishers, Boston. pp. 600-607.

Melkonian M, Becker B, Becker D (1991) Scale formation in algae. J Electron Microsc Tech 17: 165-178.

Melkonian M, Marin B, Surek B (1995) Phylogeny and evolution of the algae. In: Arai R, Kato M, Doi Y, editors. Biodiversity and evolution. Tokyo: The National Science Museum Foundation. pp. 153-176.

Moestrup O, Walne PL (1979) Studies on scale morphogenesis in the golgi apparatus of *Pyramimonas tetrarhynchus* (Prasinophyceae). J Cell Sci 36: 437-459.

Morré DJ, Mollenhauer HH (1974) The endomembrane concept: a functional integration of endoplasmic reticulum and Golgi apparatus. In: Robards AW, editor. Dynamic aspects of plant ultrastructure. McGraw-Hill Book Co., London. pp. 84-137.

Murray, M.G. and Thompson, W.F (1980) Rapid isolation of high molecular weight plant DNA. Nucl. Acids Res 8: 4321-4325.

Nedelcu A, Borza T, Lee R (2006) A land plant-specific multigene family in the unicellular *Mesostigma* argues for its close relationship to streptophyta. Mol Biol Evol 23: 1011-1015.

Neuhoff V, Phillip K, Zimmer HG, Mesecke S (1979) A simple, versatile, sensitive and volume-independent method for quantitative protein determination which is independent of other external influences. Hoppe Seylers Z. Physiol. Chem. 360: 1657-1670.

Pan J, Wang Q, Snell WJ (2005) Cilium-generated signaling and cilia-related disorders. Lab Invest 85: 452-463.

Pazour GJ, Agrin N, Leszyk J, Witman GB (2005) Proteomic analysis of a eukaryotic cilium. J Cell Biol 170: 103-113.

Perasso L, Grunow A, Bruntrup IM, Bolinger B, Melkonian M et al. (2000) The Golgi apparatus of the scaly green flagellate *Scherffelia dubia*: uncoupling of glycoprotein and polysaccharide synthesis during flagellar regeneration. Planta 210: 551-562.

Quarmby LM (2004) Cellular deflagellation. Int Rev Cytol 233: 47-91.

Rebrikov DV, Britanova OV, Gurskaya NG, Lukyanov KA, Tarabykin VS et al. (2000) Mirror orientation selection (MOS): a method for eliminating false positive clones from libraries generated by suppression subtractive hybridization. Nucleic Acids Res 28: E90.

Rebrikov DV, Desai SM, Siebert PD, Lukyanov SA (2004) Suppression subtractive hybridization. Methods Mol Biol 258: 107-134.

Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM et al. (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. Nucleic Acids Res 26: 1628-1635.

Royo J, Gomez E, Hueros G (2000a) A maize homologue of the bacterial cmp-3-deoxy-d-*manno*-2-octulosonate (kdo) synthetases. similar pathways operate in plants and bacteria for the activation of kdo prior to its incorporation into outer cellular envelopes. J Biol Chem 275: 24993-24999.

Royo J, Gimez E, Hueros G (2000b) CMP-KDO synthetase: a plant gene borrowed from gram-negative eubacteria. Trends Genet 16: 432-433.

Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning 2nd edition. Cold Spring Harbor Laboratory Press.

Schloss JA (1984) Subroutines in the programme of *Chlamydomonas* gene expression induced by flagellar regeneration. J Embryol Exp Morphol 83 Suppl: 89-101.

Schloss JA, Silflow CD, Rosenbaum JL (1984) mRNA abundance changes during flagellar regeneration in *Chlamydomonas reinhardtii*. Mol Cell Biol 4: 424-434.

Simon A. (2002) Entwicklung eines Transformationssystems für *Scherffelia dubia* (Chlorophyta). Diplomarbeit. University of Cologne.

Simon A, Glockner G, Felder M, Melkonian M, Becker B (2006) Est analysis of the scaly green flagellate *Mesostigma viride* (streptophyta): implications for the evolution of green plants (viridiplantae). BMC Plant Biol 6: 2.

Stephens RE, Lemieux NA (1999) Molecular chaperones in cilia and flagella: implications for protein turnover. Cell Motil Cytoskeleton 44: 274-283.

Stevenson TT, Darvill AG, Albersheim P (1988) 3-deoxy-d-*lyxo*-2-heptulosaric acid, a component of the plant cell-wall polysaccharide rhamnogalacturonan-II,. Carbohydr Res 179: 269-288.

Stolc V, Samanta MP, Tongprasit W, Marshall WF (2005) Genome-wide transcriptional analysis of flagellar regeneration in *Chlamydomonas reinhardtii* identifies orthologs of ciliary disease genes. Proc Natl Acad Sci U S A 102: 3703-3707.

Unger FM, Stix D, Schulz G (1980) The anomeric configurations of the two ammonium (methyl 3-deoxy-d-*manno*-2-octulopyranosid)onate salts (methyl α- and β-ketopyranosides of KDO). Carbohydr Res 80: 191-195 .

Voigt J, Frank R (2003) 14-3-3 proteins are constituents of the insoluble glycoprotein framework of the *Chlamydomonas* cell wall. Plant Cell 15: 1399-1413.

Wang AM, Doyle MV, Mark DF (1989) Quantitation of mRNA by the polymerase chain reaction. Proc Natl Acad Sci U S A 86: 9717-9721.

Wang Z, Brown DD (1991) A gene expression screen. Proc Natl Acad Sci USA 88: 11505–11509.

Wolucka BA, Van Montagu M (2003) GDP-mannose 3',5'-epimerase forms GDP-l-gulose, a putative intermediate for the *de novo* biosynthesis of vitamin C in plants. J Biol Chem 278: 47483-47490.

Wustman BA, Melkonian M, Becker B (2004) A study of cell wall and flagella formation during cell division in the scaly green alga, *Scherffelia dubia* (chlorophyta). J Phycol 40: 895-910.

Xu X, Wang J, Grison C, Petek S, Coutrot P et al. (2003) Structure-based design of novel inhibitors of 3-deoxy-d-*manno*-octulosonate 8-phosphate synthase. Drug Des Discov 18: 91-99.

York WS, Darvill AG, McNeil M, Albersheim P (1985) 3-deoxy-d-*manno*-2-octulosonic acid (KDO) is a component of rhamnogalacturonan II, a pectic polysaccharide in the primary cell walls of plants. Carbohydr Res 138: 109-126.

Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD (2001) Reverse transcriptase template switching: a smart approach for full-length cDNA library construction. Biotechniques 30: 892-897.

Acknowledgements

This thesis would never have been possible without the help of other people.

My first thoughts are for Dr. Cornelia Spetea and Prof. Anne-Lise Etienne who introduced me to the molecular world of plants and algae. Prof. Dr. Michael Melkonian, on top of that, offered me the opportunity to do a PhD in his laboratory. I want to thank him for this, but also for his limitless curiosity for algae. Michael Melkonian is especially to be praised for the truly open-minded atmosphere he is able to create in his laboratory.

I also want to thank Dr. Becker for his continuous supervision, support, criticism and advice, as well as his enthousiasm and his numerous ideas which turn every dead end into a new challenge.

Many thanks must go to my coworkers in the Arbeitgroup as well, for both their scientific and friendly companionship. Among them, I especially want to thank Nicole Feja who made everything feasible in the lab, Nayma, Katrin, Sophie and Andreas for their exceptional welcome, and the biotechnology group for their sincere and lasting friendship.

It is not enough to thank only the people which were involved in this work, as I shared this experience with many other people which I cannot all cite here.

Particularly, I could not have done this without the tremendous help, support and affection I received from my family and from my companion. I simply cannot find words which are significant enough to express all my gratitude to them.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertig thabe, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie abgesehen von unten angegebenen Teilpublikationen noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. M. Melkonian betreut worden.

Lebenslauf

Name:	Jean-Michel Doan
Wohnort:	50 rue de Strasbourg
	94300 Vincennes
	France
geboren:	08-12-1973 im Paris 10
Familienstand:	ledig

Studium

1992:	Baccalauréat Serie D, im Lycee Maurice Ravel, Paris *(gleichbedeutend Abitur)*
1996:	DEUG Science Nature et Vie, im Université Pierre & Marie Curie, Paris
1998:	Licence de biologie cellulaire et moléculaire, im Université Pierre & Marie Curie, Paris
1999:	Maîtrise de biologie cellulaire et moléculaire, im Université Pierre et Marie Curie, Paris

Promotionsstudium

Jün. 2001 – Dez. 2005:	am Botanischen Institut der Universität zu Köln unter Anleitung von Prof. Dr. M. Melkonian

Anstellungen

Nov. 1999 – Jün. 2001:	Wissenschafliche Angestellte, laboratoire dynamique des membranes végétales, École normale supérieure, Paris.
Aug. 2004 – Nov. 2004:	Wissenschafliche Hilfskraft, Botanisches Institut