

**Genetic and molecular dissection of
Photosystem I functions in *Arabidopsis* and
related functional genomics**

Inaugural - Dissertation
zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von

Claudio VAROTTO

aus Verona, Italien
Köln
2002

Die vorliegende Arbeit wurde am Max-Planck-Institut für Züchtungsforschung, Köln-Vogelsang, in der Abteilung Pflanzenzüchtung und Ertragsphysiologie (Prof. Dr. F. Salamini) in der Arbeitsgruppe von Dr. D. Leister angefertigt.

Berichterstatter: Prof. Dr. Francesco Salamini
Prof. Dr. Ulf-Ingo Flügge

Tag der mündlichen Prüfung: 07. 01. 2002

To my brother Marco

ABBREVIATIONS

<i>A. thaliana</i>	<i>Arabidopsis thaliana</i>
AIMS	amplification of insertion mutagenised sites
as	antisense
ATP	Adenosin triphosphate
bp	base pair
cDNA	complementary deoxyribonucleic acid
CO ₂	carbon dioxide
Col-0	Columbia 0
d	day
DNA	deoxyribonucleic acid
EST	expressed sequence tag
g	gram
GST	gene sequence tag
h	hour
HCF	high chlorophyll fluorescence
l	liter
LHC	light harvesting complex
m	meter
min	minute
mol	mole
mRNA	messenger ribonucleic acid
NADP(H ⁺)	nicotinamide adenine dinucleotide phosphate (reduced/oxidised)
NPQ	non photochemical quenching
°C	degree Celsius
P700(⁺)	PSI reaction centre
Pam	photosynthesis affected mutant
PAM	pulse amplitude modulation
PAR	photosynthetic active radiation
PCR	polymerase chain reaction
PSI (II)	photosystem I (II)
QRT-PCR	quantitative reverse transcription-polymerase chain reaction
RNA	ribonucleic acid
s	second
UV	ultra violet
w/v	weight per volume
WT	wild-type
F _{II}	fluorescence yield of photosystem II

CONTENTS

1. INTRODUCTION	1
1.1 OVERVIEW OF PHOTOSYNTHESIS	1
1.2 PSI STRUCTURE AND FUNCTION	2
1.2 PRINCIPLES OF CHLOPHYLL FLUORESCENCE	4
1.3 FUNCTIONAL GENOMICS	6
1.3.1 FORWARD GENETICS	6
a) Mutant screens	6
b) Genes identified	8
1.3.2 REVERSE GENETICS	9
1.3.3 TRANSCRIPTOMICS	10
1.3.4 BIOINFORMATICS	11
1.4 AIM OF THE THESIS	13
2. MATERIALS AND METHODS	14
2.1 Plant propagation and growth measurement	14
2.2 Oligonucleotides and adapter sequences	14
2.3 Nucleic acids preparation	16
2.4 DNA radioactive labelling	16
2.5 Isolation of <i>En</i> - and T-DNA - flanking sequences	16
2.6 Sequence analysis	17
2.7 Isolation of PSI insertion mutants	17
2.8 Chlorophyll fluorescence measurements	18
2.9 P700 absorption measurements	19
2.10 Immunoblot analysis of proteins	20
2.11 Public databases	20
2.12 Programming	20
2.13 Evaluation of secondary structure of oligos and primer-primer interactions	20
3. FORWARD GENETICS	22
RESULTS	22
3.1 An automated PAM fluorometer for the high-throughput screening of photosynthetic mutants	22
3.2 Description of the isolated mutants	23
3.3 Isolation of insertion sites	24
3.4 Identified genes	26
DISCUSSION	27
4. CHARACTERISATION OF A PSAE1 KNOCKOUT	30
RESULTS	30
4.1 <i>En</i> -transposon tagging and cloning of the Pam4 locus	30
4.2 Expression of <i>psaE1</i> and <i>psaE2</i> in wild-type and mutant plants	31
4.3 The abundance of PsaE, C and D proteins is significantly reduced in mutant plants	33
4.4 Alteration of the redox states of PSI and PSII in <i>psae1-1</i> mutants	33
4.5 <i>psae1-1</i> mutants show light green pigmentation and an increased chlorophyll fluorescence phenotype	34

4.6 Decreased growth	35
DISCUSSION	36
5. REVERSE GENETICS FOR PSI SUBUNITS	38
RESULTS	38
5.1 Screening and stabilisation of <i>En</i> -tagged mutations of <i>psaG</i> , <i>psaH2</i> and <i>psaK</i> genes	38
5.2 T-DNA insertion mutants of <i>psaE2</i> , <i>psaN</i> and <i>psaL</i>	38
5.3 Generation of double mutants	39
5.5 Growth behaviour of <i>psag-1.4</i> , <i>psah2-1.4</i> , <i>psag-1.4/psah2</i> mutants and WT plants	39
5.6 Expression of <i>psaG</i> and <i>psaH2</i> in wild-type and mutant plants	41
5.7 PSI polypeptide and thylakoid pigment composition in wild-type and mutant plants	42
5.8 Alteration of the redox states of PSI and PSII in <i>psag-1.4</i> , <i>psah2-1.4</i> and <i>psag-1.4/psah2-1.4</i> mutants	43
DISCUSSION	46
6.THE GST-PRIME SOFTWARE PACKAGE	48
RESULTS	48
6.1 Automatic sequence retrieval and assembly	48
6.2 Automatic primer design	50
6.3 Primer design for <i>Arabidopsis</i> and <i>Drosophila</i> genes	50
6.4 <i>GST-PRIME</i> primer testing by PCR amplification of 1900 GSTs	51
2.14 Operating environments	53
DISCUSSION	56
Summary	59
Zusammenfassung	60
REFERENCES	61
APPENDIX	72
Erklärung	72
Lebenslauf	73
Acknowledgments	74

1. INTRODUCTION

1.1 OVERVIEW OF PHOTOSYNTHESIS

Photosynthesis in higher plants is a series of light-driven reactions responsible for the generation of ATP and reducing power (NADPH) used for the fixation of atmospheric CO₂ in organic compounds. The photosynthetic reactions are performed in the chloroplast, a cellular organelle enclosed by three systems of membranes believed to derive from an ancestral cyanobacterial endosymbiont (Douglas, 1998). Besides photosynthesis, chloroplasts host also the synthesis of several compounds, such as amino acids, nucleotides, fatty acids and lipids, vitamins, plant hormones and secondary metabolites. Moreover, chloroplasts play a central role in the assimilation of sulphur and nitrogen and contain the transcriptional and translational machinery necessary for the expression of their own genome (the plastome). Photons are absorbed mainly by the light-harvesting complexes (LHCs) of two membrane-spanning protein-pigment supercomplexes of the thylakoids, called photosystem I (PSI) and photosystem II (PSII) (see Figure 1.1).

The energy absorbed by the LHCs is transferred to the reaction centres. In a multi-step process, the energy of 4 photons is used to oxidise two molecules of water to molecular oxygen. The resulting electrons are transferred through a series of redox reactions to acceptors at progressively lower electrochemical potential. The first 5 reactions take place among molecules associated to PSII to be finally transferred to plastoquinone (PQ), a small hydrophobic molecule capable to move free in the thylakoid membrane. The plastoquinone transfers the electrons to the cytochrome *b₆f* complex, which is made up of 6 polypeptides not binding any chlorophyll. The electrons are then passed to plastocyanin, the donor of a second charge separation process driven at the reaction centre of PSI by an absorbed photon. This step energises the electrons that are then transferred across PSI to ferredoxin (Fd), and finally to the membrane-bound ferredoxin-NADP⁺-oxidoreductase (FNR). As a net result of the whole process two molecules of water are split, 2 molecules of NADPH are synthesised and a transmembrane pH gradient, used by the plastid ATPase to synthesise ATP, is built up. The ATP and NADPH generated during the light reactions will then be used to fix atmospheric CO₂ in carbohydrates (Calvin cycle).

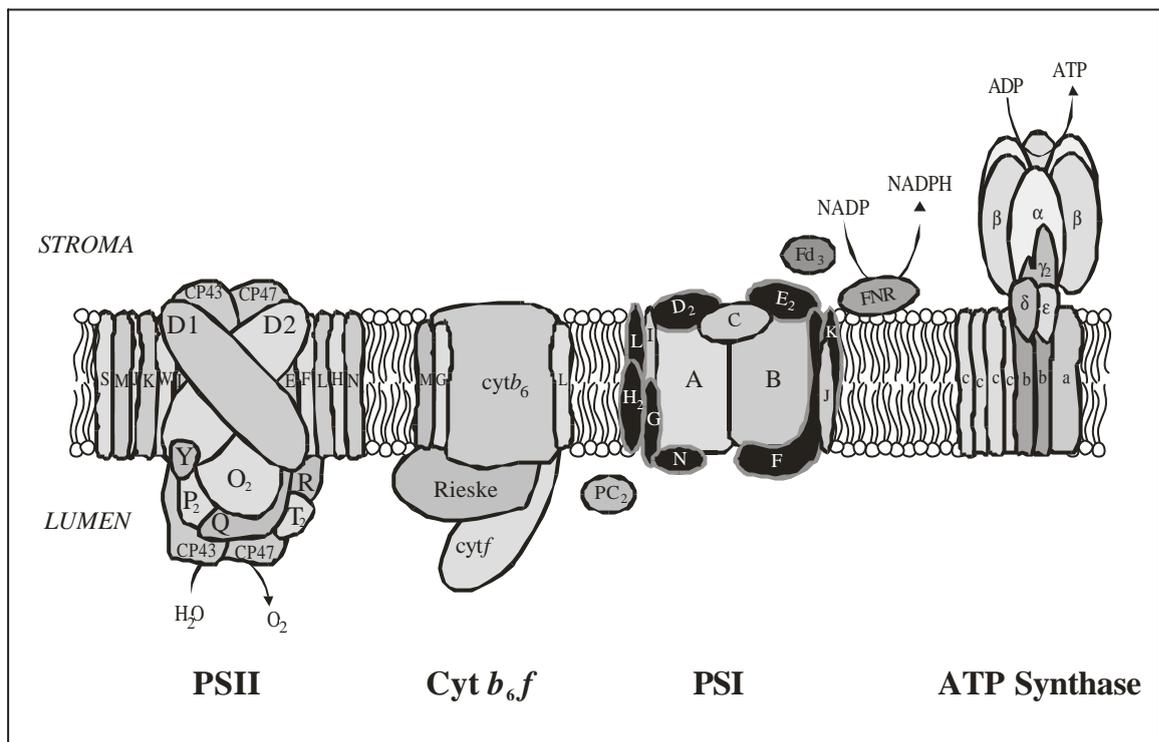


Figure 1.1. Schematic drawing representing the photosynthetic apparatus (from left to right: photosystem II, PSII; cytochrome *b₆f*, Cyt *b₆f*; Photosystem I, PSI; ATP synthase) and the soluble electron carriers (plastocyanin, PC; ferredoxin, Fd) involved in the light-driven production of ATP and NADPH. Nuclear-encoded subunits of PSI are depicted in black.

1.2 PSI STRUCTURE AND FUNCTION

Although the subunits of both photosystems have been studied for the last 30 years (He and Malkin, 1998), many questions concerning their specific functions and contribution to the overall stability of the complexes remain open. Cyanobacterial mutants lacking specific subunits of PSI or PSII cannot always contribute to assign functions to their pendants in plants. In some cases marked functional differences are evident for the same subunit in the two groups of organisms. This can be interpreted in the context of functional changes occurred during the evolution from unicellularity to multicellularity. Although the overall function of PSI is conserved in cyanobacteria and plants, some significant functional and structural differences are evident: (i) the association between plastocyanin and PSI is stabilised in plants by a specific N-terminus extension of PsaF not present in cyanobacteria, where the electron transfer follows

a second order kinetic (Farah et al., 1995; Hippler et al., 1997); (ii) similarly, the N-terminal extension of plant PsaD plays a role in the association with PsaC (Naver et al., 1995); (iii) moreover, the whole stromal ridge of plant PSI (constituted by PsaC, D and E) appears to be associated to the PSI core much tighter than in cyanobacteria (Naver et al., 1998); (iv) in cyanobacteria, PsaL mediates trimerisation of PSI, whereas plant PSI has been found only in a monomeric state (Chitnis and Chitnis, 1993); (v) at the level of polypeptide composition, cyanobacterial PsaM is not present in angiosperms and subunit X of *Synechococcus* is absent in all plants. Furthermore, the subunits PsaG, H, N and O are specific for plants (Jansson et al., 1996; Scheller, personal communication).

In *Arabidopsis thaliana*, PsaD, E and H are coded each by two gene copies, called *psaD1* and *psaD2*, *psaE1* and *psaE2*, *psaH1* and *psaH2*, respectively. Plant PsaG and PsaK share a significant degree of homology (30% aminoacid identity; Okkels et al., 1992) and are equally divergent from the cyanobacterial PsaK. Therefore it appears plausible that the two genes derive from a duplication event of their cyanobacterial progenitor *psaK* (Okkels et al., 1992). Both PsaG and PsaK are integral membrane proteins believed to interact with homodimers of Lhca2 and 3 respectively (Jansson et al., 1996). While mutant analysis of *psaK* revealed a role in LHCI organisation for this subunit, the lack of mutants for *psaG* hampered by now the functional characterisation of PsaG.

Like PsaG and K, also the PsaH subunit is an integral membrane protein. Cross-linking studies suggest physical contact between PsaH and three other subunits, PsaD, PsaI and PsaL (Jansson et al., 1996). PsaH is required for photosynthetic state transitions (Lunde et al., 2000), a process involving the phosphorylation of LHCI and its migration from PSII to PSI (Wollman, 2001). Also a reduction in PSI stability and a decreased NADP⁺ photoreduction in presence of saturating ferredoxin has been noticed in *psaH* cosuppressed plants, probably due to a destabilisation of PsaD and PsaC subunits. Interestingly, PsaH cosuppression lines compensate the PSI instability by increasing the amount of PSI (Naver et al., 1999).

The last characterised plant specific subunit of PSI is PSI-N, a small extrinsic polypeptide located on the luminal side of PSI. *Arabidopsis* plants lacking PsaN are affected in plastocyanin docking to PSI, as demonstrated by the decrease of both second order rate constant of P700⁺ reduction and steady state photoreduction of NADP⁺ (Haldrup et al., 1999). A decreased amount of PsaN has also been observed in plants lacking PsaF, making it difficult to discriminate between effects caused by the single subunits (Haldrup et al., 2000). Similarly

to PsaH cosuppressed lines, plants lacking PsaN compensate the impaired performance of photosystem I by increasing PSI content in the thylakoids.

1.2 PRINCIPLES OF CHLOROPHYLL FLUORESCENCE

Chlorophyll is the main antenna pigment, funnelling the absorbed light energy into the reaction centres, where photochemical conversion of the excitation energy takes place. The absorption of light by chlorophyll causes its conversion to a highly unstable excited state. De-excitation of chlorophyll can take place by means of three different mechanisms (see Figure 1.2; Sauer, 1975):

- 1) dissipation into heat, by an increased molecular vibration of the chlorophyll molecule.
- 2) re-emission as fluorescence, a luminous radiation with a wavelength longer than that of the absorbed light.
- 3) photochemistry, promoting the oxidation of the reaction centres and sustaining the electron transport coupled with ATP and NADPH production in chloroplasts.

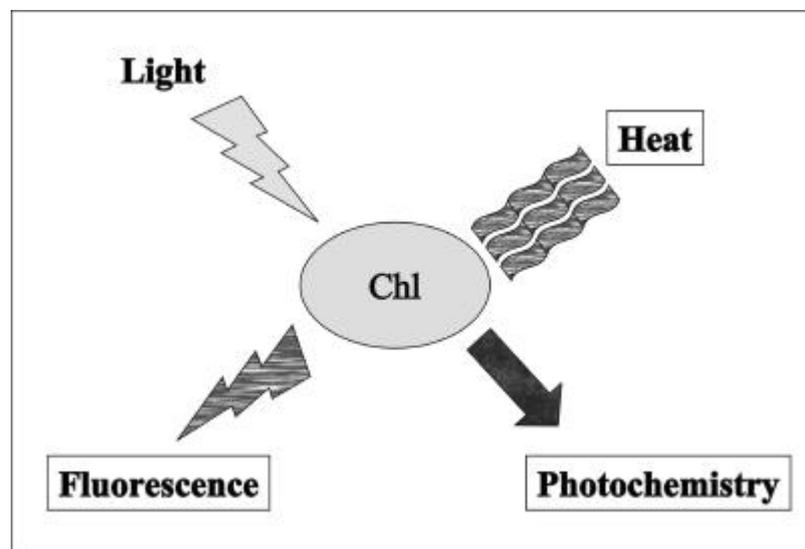


Figure 1.2. Competing mechanisms for de-excitation of chlorophyll molecules. Chl, chlorophyll.

All of these de-excitation pathways are competing processes, but under normal conditions, dissipation as heat represents only a minor contribution to chlorophyll de-excitation. In consequence, fluorescence emission can be regarded as being the only process competing with photochemistry. For this reason, the fluorescence yield is highest when the yield of

photochemistry is lowest and thus chlorophyll fluorescence can be used as a tool to determine both the maximal and the effective efficiency of photon usage in photochemistry (Clayton, 1980). Experimentally it has been observed that the variable part of chlorophyll fluorescence originates mainly in PSII. In particular, in dark-adapted plants, the basal level of fluorescence upon illumination with an amount of light (M.B.) insufficient to activate the oxidation of the reaction centres (Figure 1.3) derives nearly exclusively from the antenna of PSII (F_0 , Figure 1.3).

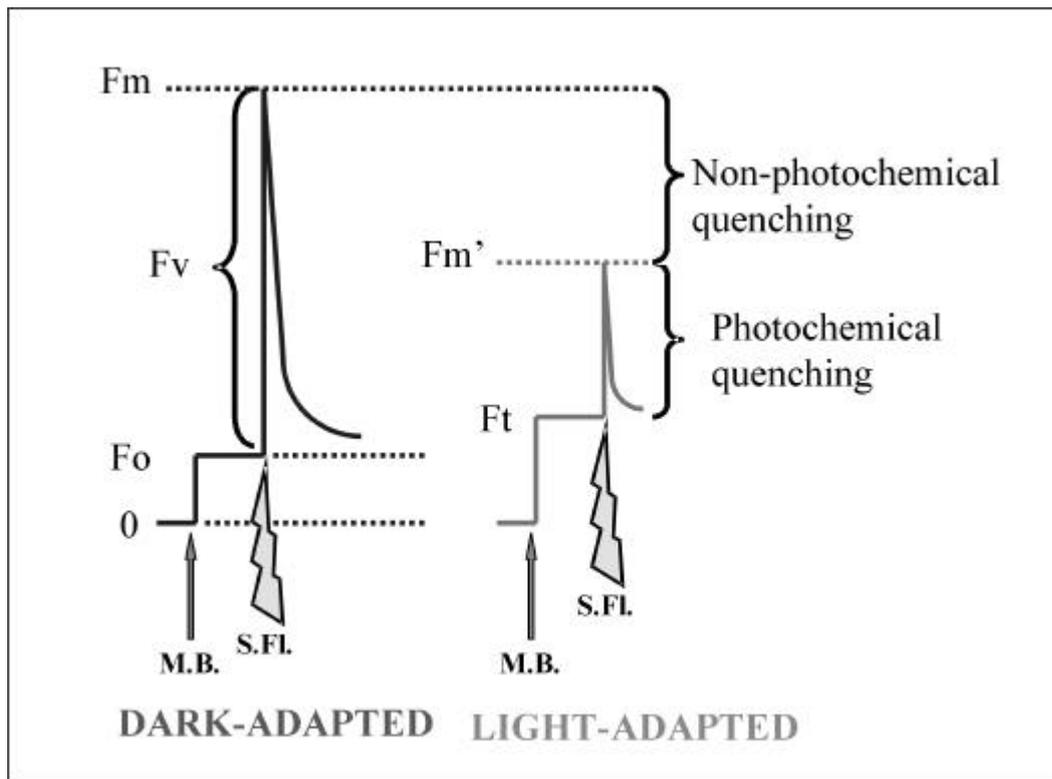


Figure 1.3. Fluorescence parameters obtained by the saturation pulse method. For dark-adapted plants: F_m , maximal fluorescence; F_0 , minimal fluorescence; F_v , variable fluorescence ($=F_m-F_0$). For light-adapted plants: F_m' , light-adapted maximal fluorescence; F_t , steady state fluorescence. M.B., measuring beam; S.Ps., saturating pulse.

On the other hand, when illuminating leaves with a strong and short flash of white light (S.Ps., Figure 1.3), the resulting fluorescence level, F_m , represents the maximal fluorescence emission of the leaf (Krause and Weis, 1991). The ratio $F_v/F_m = (F_m-F_0)/F_m$ is called maximal fluorescence yield of PSII and measures the maximal fraction of absorbed photons that can be used to carry out photochemistry. The fluorescence levels corresponding to F_0 and F_m are called, in light-adapted plants, respectively, F_t and F_m' and the effective (actual) fluorescence yield of PSII is given by $F_{II} = (F_m'-F_t)/F_m'$ (Genty et al. 1989). The last parameter represents

the fraction of photons used to perform photochemistry and can be used as an indicator of the photosynthetic efficiency of the plant (see Figure 1.3).

1.3 FUNCTIONAL GENOMICS

During the last 10 years, several aspects of flowering plants have been elucidated using *Arabidopsis thaliana* as a model organism due to its small size, short life cycle, large number of offsprings and ease of transformation. The genetic approach to dissect its biology has taken big advantage from the identification of mutants isolated by phenotypic screening (“forward genetics”). After the complete sequencing and annotation of its genome, the problem of isolating genes has been shifted to that of assigning a function to them. For this purpose, new technologies have been developed that are complementary to the classical "forward" (phenotype-driven) screening for mutants, collectively summarised under of the term of “genomics”. In particular, tagged mutant populations, already used for forward genetics, have been successfully employed to identify mutations in specific genes predicted in the course of the sequencing of the *Arabidopsis* genome. This has led on one hand to the validation of the computer-assisted predictions of coding sequences and on the other hand to the (*a posteriori*) attribution of functions to genes not homologous to genes of known function. A further step in the understanding of gene functions is provided by the microarray technology, virtually able to determine the changes in expression patterns at the level of the complete *Arabidopsis* genome in response to different experimental conditions. Though commonly used solely for large-scale reverse genetic and tagging approaches, the term “functional genomics” will be used in the following to indicate the whole range of current strategies which can contribute to the assignment of functions to genes at genome level and which cover the fields of genomics (forward and reverse genetics), transcriptomics and bioinformatics. (Richmond and Sommerville, 2000; Sommerville and Sommerville, 1999).

1.3.1 FORWARD GENETICS

a) Mutant screens

The genetic dissection of the mechanisms controlling photosynthesis can be performed by isolating mutants with altered photosynthetic performance. The identification of photosynthetic

mutants has been carried out in several ways. The most commonly used indicators of defects associated with photosynthesis are alterations in pigmentation or chlorophyll fluorescence. In both cases, the screening has been successfully performed visually and the technical simplicity of these approaches largely contributed to their popularity.

In some cases, altered chlorophyll fluorescence and pigmentation are interdependent. Pigmentation mutants like *tha4* and *crp1*, for instance, also show altered chlorophyll fluorescence (Fisk et al., 1999; Walker et al., 1999). In general, every lesion affecting significantly the photosynthetic electron transport chain can be detected as an increase in the fraction of absorbed energy re-emitted as fluorescence. This approach was first applied for the visual identification of *high chlorophyll fluorescence (hcf)* mutants of *Chlamydomonas reinhardtii*, maize and barley (Bennoun and Delepelaire, 1982; Bennoun and Levine, 1962; Miles, 1980; Simpson and von Wettstein, 1980): irradiating plants with UV light, it is possible to visually identify photosynthetic mutants through their increased fluorescence. In most of the mutants identified in this way, due to the relatively low sensitivity of the method, the lesion affecting the photosynthetic electron transport are severe, often causing seedling lethality under photoautotrophic conditions, hampering the analysis of *hcf* mutants. Up to now several *hcf* mutant collections have been described (Dinkins et al., 1994; Meurer et al., 1996; Taylor et al., 1987).

In *Arabidopsis*, Meurer et al. (1996) identified thirty-four *hcf* mutants, of which most were lethal at the seedling stage. Most of them exhibited a reduction in photochemical quenching deriving from a marked decrease in photosynthetic electron transport activity. However, mutations affecting photosynthesis, but not accompanied by severe phenotypes, have been reported. Two laboratories have identified *Arabidopsis* mutants with alterations in non photochemical quenching (NPQ, Bradbury and Baker, 1981) (Niyogi et al., 1998; Shikanai et al., 1999; Table 1.1).

The screenings have been carried out in conditions designed to avoid the survival of seedling lethal mutants (plants were grown either on sucrose-free media (Niyogi et al., 1998) or on soil (Shikanai et al., 1999). Niyogi et al. reported the isolation of 13 NPQ mutants, 3 of which were defective exclusively in the xanthophyll cycle. Using a similar procedure, Shikanai and co-workers found 37 mutants with a primary defect in NPQ, among which 19 showed a reduced quantum yield for both photosystems. Several of the mutants showed also altered pigmentation.

Table 1.1. Overview of screens for chlorophyll fluorescence and pigmentation mutants of *Arabidopsis*.

<i>Mutant phenotype</i>	<i>Population</i>	<i>Mutagen</i>	<i>N^o of mutants</i>	<i>Mutation frequency</i>	<i>Reference</i>
<i>hcf</i>	7,700 M2-families	EMS	34	1/230	Meurer et al., 1996
	11,600 M2-families	T-DNA	33	1/350	K.Meierhoff, personal communication
NPQ	30,000 M2-individuals	EMS, Fast neutron	13	1/2300	Niyogi et al., 1998
	51,500 M2-individuals	EMS	55	1/940	K.Niyogi, personal communication
	44,600 M2-individuals	Fast neutron	6	1/7500	K.Niyogi, personal communication
	43,000 pooled families	T-DNA	8	1/5400	
	21,000 M2-individuals	EMS	37	1/570	Shikanai et al., 1999
Pigmentation	1,900 M1 seed families	<u>EMS</u>	211	1/11	Runge et al., 1995

b) Genes identified

Several mutant genes responsible for changes in chlorophyll fluorescence or pigmentation have been identified. Most of them are coding for proteins imported into the chloroplast. So far, four *Arabidopsis HCF* genes have been isolated (*HCF136*, Meurer et al., 1998; *HCF107*, Felder et al., 2001; *HCF101*, J. Meurer, personal communication; *HCF164*, K. Meierhoff and P. Westhoff, personal communication).

Also several pigmentation mutations are related to proteins imported into the chloroplast. In particular, to this class belong mutations affecting genes coding for enzymes involved in isoprenoid and phytochrome/chromophore biosynthesis, for subunits of the complexes for the import of proteins into the chloroplast and for factors necessary for chloroplast development (*CS*, Koncz et al., 1990; *PAC*, Reiter et al., 1994; *CLA1*, Mandel et al., 1996; *ALBINO3*, Sundberg et al., 1997; *CAO*, Klimyuk et al., 1999; *FFC*, Amin et al., 1999; *IMMUTANS*, Wu et al., 1999; *HO1*, Muramoto et al., 1999; Davis et al., 1999; *VAR2*, Chen et al., 2000). Two NPQ mutants lack enzymes involved in the xanthophyll cycle (zeaxanthin epoxidase and

violaxanthin deepoxidase (Niyogi et al., 1998), while a third NPQ mutant, *npq4*, was deficient for PsbS, an intrinsic subunit of PSII (Li et al., 2000).

The increasing amount of information emerging from the genetic dissection of mutants identified by forward genetic screens indicates the validity of this approach to elucidate the role and relation of the chloroplast in the context of the cell: with the mutant screens available it is possible to identify even those gene products that, though not physically located in the chloroplast, play a role in plastid functions.

1.3.2 REVERSE GENETICS

Although particularly useful to discover new factors affecting the photosynthetic performance of higher plants, the genetic dissection of photosynthesis by forward genetics is a rather inefficient tool for the large-scale analysis of gene functions. Even with the use of tagged populations, the identification of the mutation at molecular level normally represents the bottleneck for this kind of approach, reducing drastically its throughput. Due to the relatively good knowledge on the basic mechanisms underlying photosynthesis, “reverse genetics” represents a more efficient and, in part, complementary tool to the conventional phenotypic screens.

Currently, the method of choice for the identification of loss-of-function mutants for genes of interest is to screen large collections of *Arabidopsis* lines mutagenised by random insertions of transposons or T-DNA (reviewed in Parinov and Sundaresan, 2000). The screenings are based on PCR, in which one of the primers is complementary to the target gene while the other to the insertional mutagen. The realisation of arrays of spotted flanking regions has further simplified the molecular screen, allowing for some of the populations the large-scale identification of insertional mutants by simple hybridisation procedures (Tissier et al., 1999; Steiner-Lange et al., 2001). An alternative strategy involves the systematic sequencing of genomic sequences flanking the insertions; the data obtained are then organised in searchable databases (Parinov and Sundaresan, 2000; Tissier et al., 1999).

The availability of these tools opens the way to the isolation of mutants for all the nuclear genes of *Arabidopsis*. The most recent projections indicate the year 2010 as the deadline for such an achievement. However, current reverse genetics approaches related to photosynthesis or general chloroplast functions are aimed to the identification of mutants for a well-defined and limited subset of genes.

1.3.3 TRANSCRIPTOMICS

In the last few years an additional tool based on the microarray technology has been added to forward and reverse genetics to attribute functions to genes. The function of unknown genes can be inferred by comparing its pattern of expression with that of already characterised genes. The goal of 'transcriptomics' is, hence, to discover how the genes of an organism are expressed during different developmental stages or in response to certain stimuli. The assumption that genes with related expression patterns have related functions represents the basis of transcriptomics.

The availability of the DNA sequences of entire genomes, in combination with the sequencing of a large number of cDNAs in the form of expressed sequence tags (ESTs), facilitates large-scale experiments based on the simultaneous study of a large, or even the entire, set of genes in a given organism (Brown and Botstein, 1999; Duggan et al., 1999; Eisen and Brown, 1999). One of the approaches currently adopted makes use of hybridisation of a labelled, complex cDNA sample to DNA fragments spotted on a solid carrier, either nylon or glass (microarray). The DNA fragments spotted can be cDNAs, genomic clones (DNA array) or oligonucleotides (oligonucleotide array) (reviewed in Schaffer et al., 2000). To date, the most advanced transcriptome analyses have been performed in *Saccharomyces cerevisiae*, human and *Synechocystis* (Kumar and Snyder, 2001; Kao, 1999; Rew, 2001; Hihara et al., 2001). In *Arabidopsis*, the most advanced cDNA and oligonucleotide arrays cover about 11,000 ESTs and 8000 ESTs or genes, respectively (<http://afgc.stanford.edu>; <http://www.affymetrix.com/products/arabidopsis.html>).

Only few experiments were aimed to understand the transcription regulation of genes related to photosynthesis yet. Very dramatic changes associated to etiolation and de-etiolation processes have been shown to significantly affect the transcription levels of about 16 % of a set of approximately 800 cDNAs used in this study (Desprez et al., 1998). Other gene expression studies related to photosynthesis in *Arabidopsis* were performed by the *Arabidopsis* Functional Genomics Consortium (AFGC, <http://genome-www4.stanford.edu/MicroArray/SMD/>); in these experiments the response to changes in either CO₂ concentration or lighting conditions were evaluated with DNA microarrays.

Only recently, DNA microarrays bearing nearly all of the genes of the unicellular cyanobacterium *Synechocystis* sp. PCC 6803 were used to examine the temporal program of gene expression during acclimation from low to high light intensity (Hihara et al., 2001).

1.3.4 BIOINFORMATICS

The complete genome sequences of many prokaryotes and eukaryotes (Goffeau et al., 1996; Kaneko et al., 1996; Blattner et al., 1997; The *C. elegans* Sequencing Consortium, 1998; Adams et al., 2000; The *Arabidopsis* Genome Initiative, 2000) and a draft of the human genome (The International Human Genome Sequencing Consortium, 2001) became recently available, generating a large amount of biological data. In parallel, bioinformatics have undergone a rapid development to provide the tools necessary to analyse the sequencing data. Several public databases, such as GenBank (Wheeler et al., 2001), the EMBL Nucleotide Sequence Database (Stoesser et al., 2001) and the DNA Data Bank of Japan (Tateno et al., 2000) are devoted to the task of systematically collect, analyse and distribute the data obtained.

The genome of *Arabidopsis thaliana* contains around 25,500 genes (The *Arabidopsis* Genome Initiative, 2000) and its sequence is now fully available. Functions have been attributed to about half of the *Arabidopsis* genes, either by sequence homology to genes of known function or by experimental evidence. However, even for the remaining 50% of the genes, it is possible to obtain information from the protein primary sequence, such as the presumable intracellular localisation and post-translational modifications. While only 87 polypeptides are coded by the plastid genome are, the remaining plastid proteins are encoded in the nucleus. These polypeptides are translated in the cytoplasm and then imported into the chloroplast. A proteome-wide search for putative chloroplast transit peptides followed by a homology-based comparison of the predicted chloroplast proteome with the total protein complement of a cyanobacterium (*Synechocystis*) has been performed. By using the ChloroP prediction program (Emanuelsson et al., 1999), between 1,900 proteins and 2,500 proteins with chloroplast localisation were predicted (Abdallah et al., 2000). Of those, at least 31 % are of cyanobacterial origin, indicating a conserved function for these gene products (Figure 1.4).

Thus, the possibility to reliably predict nuclear genes coding for proteins with unknown functions targeted to the chloroplast provides not only a powerful tool to estimate the number of polypeptides necessary for chloroplast functions, but also a useful mean to select the subset of the whole proteome having high potential to be important for photosynthesis. The microarray technology is already taking advantage from the predictive power of bioinformatics: while in the first transcriptomic experiments collections of ESTs have been used, recently the availability of complete genome sequences led to the generation of

microarrays obtained spotting gene sequence tags (GSTs) directly amplified from genomic DNA (Zhang, 1999; Penn et al., 2000).

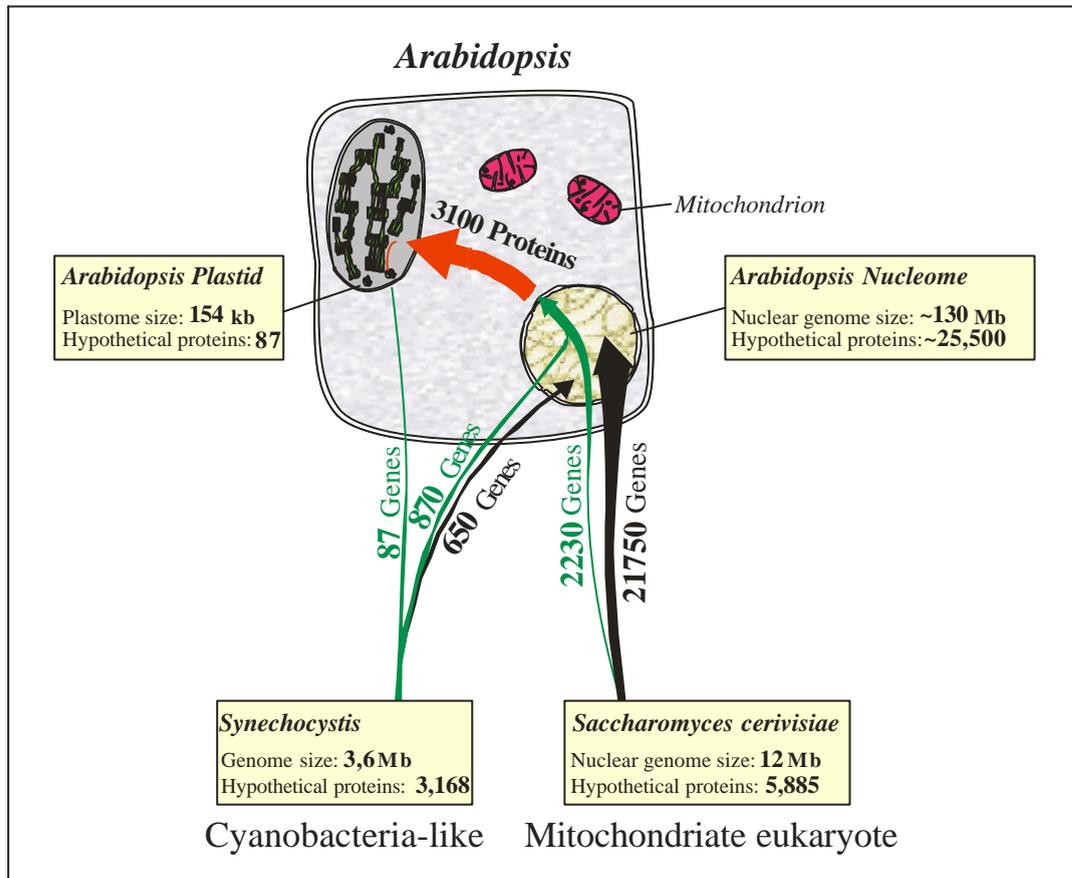


Figure 1.4. Evolution of the chloroplast proteome adapted from Abdallah et al. (2000). Chloroplasts are believed to originate from an endosymbiotic event involving a cyanobacteria-like prokaryote and a mitochondriate eukaryote host. Green lines indicate the origin of chloroplast proteins, whereas black lines represent the descent of non-chloroplast proteins. Widening lines indicate an increase in the size of proteomes as a result of gene acquisition following the symbiotic event, whereas tapering lines reflect gene loss during evolution.

1.4 AIM OF THE THESIS

This thesis had the aim to dissect the function of one of the two multi-subunit complexes involved in the utilisation of light energy to reduce the atmospheric CO₂: the photosystem I. For this purpose three different approaches have been taken:

- 1) the identification of genes relevant for photosynthesis. A novel screening procedure was used to isolate *Arabidopsis* mutants with altered photosynthesis (Pam mutants) (see chapter 3), leading to the identification of the corresponding mutated genes.
- 2) the collection of knock-outs of the other nuclear encoded subunits of PSI. The PSI complex of higher plants is a mosaic of plastid- and nucleus-encoded protein subunits. Furthermore, genes coding for the subunits PsaD, PsaE and PsaH present in the nuclear genome of *Arabidopsis* are redundant (Naver et al., 1999; Obokata et al., 1993). The knockout of the single subunits of PSI could provide insights into both structural and regulative aspects of PSI. In particular an exhaustive mutagenesis of the 11 nuclear genes encoding subunits of PSI could address questions related to:
 - the biological significance of duplication of the genes coding for PsaD, E and H.
 - the function of plant-specific N-terminal extensions of D and E subunits, not present in the cyanobacterial homologues
 - the function of the subunits PsaG, PsaH and PsaN, which are specific for higher plants
- 3) the development of a method suitable for systematic analysis of the *Arabidopsis* transcriptome by DNA-arrays, in particular in relation to genes whose products are putatively imported into the chloroplast. The programs for primer design available (*PRIDE* (Haas et al., 1998), *PRIMER MASTER* (Proutski and Holmes, 1996), *PRIMO* (Li et al., 1997), *PRIMEARRAY* (Raddatz et al., 2001)) are not suitable for this purpose. As a part of this thesis, a program for (i) the automatic retrieval and assembly of large sets of gene sequences, and (ii) the design of primers pairs suitable for the amplification from either the genomic or the corresponding cDNA sequences was developed.

2. MATERIALS AND METHODS

2.1 Plant propagation and growth measurement

The *En*-mutagenised *A. thaliana* (ecotype Columbia 0, Col-0) population comprising 8,000 lines with 48,000 *En*-insertions has been described in Wisman et al, 1998. Additional 8,000 lines mutagenised by T-DNA insertions were obtained from Bernd Reiss (Max-Planck-Institut für Züchtungsforschung). Seeds of *A. thaliana* ecotype Col-0 were sown in plastic trays with "Minitray" soil (Gebr. Patzer GmbH & Co.KG, D-36391 Sinntal-Jossa, Germany) and incubated for 3 d at 2-5 °C in the dark to break dormancy. Plants were grown in a greenhouse under long day conditions (with additional light for a total day length of at least 16 h). Fertilisation with "Osmocote Plus" (15 % N, 11 % P₂O₅, 13 % K₂O, 2 % MgO; Scotts Deutschland GmbH, D-48527 Nordhorn, Germany) was performed according to manufacturer's instructions. For the determination of growth rate, seeds were sown in pots and 1 week after germination individual plants of the same size were transplanted into trays. Growth measurements were performed using a non-invasive imaging system as described in Leister et al. (1999).

For the mutant screening, plant trays were transferred 3-4 weeks after germination into a climate chamber under short day conditions (day period of 10.5 h with 20 °C and constant PAR of 200 $\mu\text{mol sec}^{-1} \text{m}^{-2}$; night period of 13.5 h with 15 °C) and maintained for at least 2 days under these conditions before measuring Φ_{II} , the effective quantum yield of PSII.

2.2 Oligonucleotides and adapter sequences

For the isolation of transposon-flanking regions, the following adapters and primers (5' - 3' orientation) were used: APL1632 (LR32 + APL16), APL1732 (LR32 + APL17), LR32 (ACTCGATTCTCAACCCGAAAGTATAGATCCCA), APL16 (*P*-TATGGGATCACATTAA-NH₂), APL17 (*P*-CGTGGGATCACATTAA-NH₂), LR26 (ACTCGATTCTCAACCCGAAAGTATAG); for *En* 3': En8130s (GAGCGTCGGTCCCCACACTTCTATAC) and En8153s (TACGAATAAGAGCGTCCATTTTAGAGTG); for *En* 5': En249as (GGCAGGGAGAAAGGAGAGAA) and En230as (AGAAGCACGACGGCTGTAGAATAGGA). For the isolation of T-DNA flanking regions, adapters APL1632 and APL1732 and the following

primers were used. Left border: T9750as (ATAATAACGCTGCGGACATCTACATTTT) and T9697as (CTCTTTCTTT TTCTCCATATTGACCAT); right border: T4496s (CAGGGTACCCGGGGATCAGATTGTC) and T4554s (GATCAGATTGTCGTTTCCCG CCTTCAGTTT).

For generation of the *psaE1* Northern probe, primers E1-41s and E1-943as (CCCATTAA GCTGCAACTTCT) were used; for QRT-PCR primers E91/85s (GTGTCTTTCTTGCCG ATGAGAA) and E798/868as (GCGAACCGGACCACAACCGG).

Insertions within *psaG*, *psaK* and *psaH2* (reverse genetic screening of the *En*-mutagenised population) were identified by PCR screening using the gene-specific primers:

psaE1: E91/85s, E798/868as (see above). *psaG*: G-1s (ATGGCCACAAGCGCATCAGCTTT GCTC), G-450as (GGAAGTAGCCAAGATGTAGTAAGCAACG). *psaH2*: H2-234s (AGCTTGCCGCGAGGACCGAGCTTAGG), H2-603as (TGTGGTGGCTAAGTATGGAG ACAAAAGT). *psaK*: K--6s (AAGAAAATGGCTAGCACTATGATGACTA), K-690as (TTCAAATAGCA CCAATGTTTTTAAGGCC).

The primers used for the footprint analysis were:

psaH2: H2-FPs (TACAACCTTCTGCCGCCGTG) and H2-FPas (CTCCATACTTAG CCACCACA). *psaG*: G-s (see above) and G-FPas (GGTTGATAGTTTGGGTAGGG). *psaK*: K-FPs (TTTGTCATCCCAGGCAAGTG) and K-FPas (AACATCAGGGTTCGTCGACGT);

The primers used for the screening of the AFGC alpha and beta populations were:

psae2: E2--923s (AATCCAGGGGAAAGCCAAGCAAACACTAT), E2-1556as (TTAGCC ACTACATTTGCTATGACCATCAC). *psaN*: N-1365as (GTGATAGCAGAAGTGG ATCGTGTTTTGA), N--723s (GACAGCCCAGAGATTTTGATTCCTCGTG). *Actin Interacting Protein (control)*: Con-1A (CGTCTAGGTGGTTCAGTACCTGTTGAATG), Con-1B (TTTATCGAAGAA ACATGTCGTTGAACCAG). *T-DNA left border*: JL-202 (CATTTTATAATAACGCTGCGGACATCTAC).

The primers used for the screening of the Jack lines were: L-301s (CTAGTTGTGTAGATT GGCCATATTCTTT), L-826as (CCGTAGATGGTGAGGCACATGCTGAG), TJ-LB1 (GAACATCGGTCTCAATGCAAAAGGGGAAC).

2.3 Nucleic acids preparation

Isolation of *Arabidopsis* DNA was performed as described in Liu et al., 1995. Total plant RNA was extracted from 100 mg of fresh tissues using the RNeasy Plant System (Qiagen). RNA gel blot analysis was performed under stringent conditions (Sambrook et al., 1989) by using the ³²P-labeled *psaE1*-specific probe E1(41-943). For QRT-PCR analysis first-strand cDNA was synthesised by using the SuperScript Preamplification System (Gibco/BRL). One microliter of first-strand cDNA mixture was used for QPCR amplification in a total volume of 25 µl.

Quantification of RT-PCR and northern signals was performed using a phosphoimager (Storm 860; Molecular Dynamics) and the program *Image Quant for MacIntosh* (version 1.2; Molecular Dynamics).

2.4 DNA radioactive labelling

250 pmol of sense primers used for QRT-PCR or footprint analyses were labelled with 20 µCi of (³³P)dATP by T4 polynucleotide kinase (Pharmacia) in a final volume of 12 µl.

Labelling of probes used for Northern or Southern blot analyses were performed as described in Sambrook et al. (1989).

2.5 Isolation of *En*- and T-DNA - flanking sequences

Sequences flanking the ends of *En* and T-DNA were isolated by PCR amplification of restricted and adapter-ligated plant genomic DNA similar to the procedure described by Frey et al. (1998). 100 ng of genomic DNA were digested with *Csp6I* (*Hin6I*) and ligated overnight at 16 °C to 12.5 pmol of adapter APL1632 (APL1732). Four microlitres of the ligation were used in a linear PCR with primer En8130s (En249as), and subsequently a 1-µl aliquot of the linear PCR was used as the template for an exponential PCR with primers En8153s (En230as) and LR26. For the amplification of T-DNA flanking regions the linear PCR was performed with primer T9750as (T4496s) and the exponential PCR with primers T9697as (T4554s) and LR26. All amplifications were performed using the Advantage[®] 2 PCR Kit (Clontech) and following cycling conditions: initial denaturation for 2 min at 94 °C, followed by 30 cycles of 30 sec denaturation at 94 °C, 1 min annealing at 64 °C and 1 min 30 sec elongation at 73 °C. Products of exponential PCR were separated by electrophoresis on 4.5% (w/v) polyacrylamide gels and bands were visualised by silver staining as described in Sanguinetti et al. (1994). After excision

of the candidate bands, PCR products were eluted in 50 mM KCl, 10 mM Tris-Cl (pH 9.0), 0.1 % Triton-X 100, re-amplified and directly sequenced, after gel-purification, using an ABI prism 377 sequencer.

2.6 Sequence analysis

Sequence data were analysed with the Wisconsin Package Version 10.0, Genetics Computer Group, Madison, Wisconsin (GCG) (Devereux et al. 1984). Chloroplast import sequences prediction was performed using the ChloroP program (version 1.0; <http://www.cbs.dtu.dk/services/ChloroP/#submission>, Emanuelson et al., 1999) and TargetP (version 1.0; <http://www.cbs.dtu.dk/services/TargetP/#submission>, Emanuelson et al., 2000)

2.7 Isolation of PSI insertion mutants

En insertions within the genes *psaE1*, *psaG*, *psaH2* and *psaK* were identified by screening the *En*-tagged population using PCR with the gene-specific primers in combination with *En*-specific primers (listed in §2.2).

Amplifications were performed using *Taq* polymerase (Roche) and the following cycling conditions: initial denaturation for 2 min 45 sec at 94 °C, followed by 40 cycles of 15 sec denaturation at 93 °C, 45 sec annealing at 65 °C and 1 min 30 sec elongation at 72 °C. The PCR products have been displayed on a 1% agarose gel, blotted and hybridised with a gene-specific probe obtained by PCR amplification with the same primers used for the screening. Positive lines were confirmed by sequencing the PCR-amplified insertion sites and homozygous plants were selected by analysing the segregation of the *En* insertion in the progeny. 45 mutants plants for each line were screened by radioactive PCR for footprints disrupting the protein reading frame: two primers designed at about 50 bp on each side of the *En-1* insertion point have been used to amplify the footprint sectors of inflorescence leaves and the resulting products were displayed on a 6% polyacrylamide gel. The primers used are listed above. The PCR conditions are as above with the exception of (i) an annealing temperature of 55°C and (ii) a total of 20 cycles.

The transmission of the stable mutations identified was verified by selfing the positive plants and confirming the presence of the same footprint among 40 individuals in the F1 progeny. 4 of

the F1 positive plants have been selfed and 20 plants from the progeny (F2) of each of them have been analysed for the absence of the transposon. All of the *En* negative F2 plants have been analysed for the presence of the footprint and the corresponding PCR product were sequenced.

The T-DNA insertion lines for *psaE2* and *psaN* were obtained by screening the AFGC alpha and beta populations (*Arabidopsis* Functional Genomics Consortium; <http://afgc.stanford.edu/>), according to the guidelines described at: <http://www.biotech.wisc.edu/NewServicesAndResearch/Arabidopsis/GuidelinesIndex.html>. See §2.2 for a list of the primers used.

The insertion in the *psaL* gene has been obtained by screening the T-DNA promoter trap lines generated by Tom Jack and co-workers (Campisi et al., 1999). The primers used are listed above. The PCR conditions are the ones reported for the reverse screening of the *En*-population, with the only difference of an annealing temperature of 58°C.

2.8 Chlorophyll fluorescence measurements

The screening for mutants with altered effective quantum yield of PSII [$\Phi_{II} = (F_M - F_0)/F_M = \Delta F/F_M$, Genty parameter, Genty et al., 1989] was performed by using an automatic pulse amplitude modulation fluorometer system (J. Kolbowski, D-97422 Schweinfurt, Germany, Figure 2.1). A *Computerised Numerical Control* (CNC) router [Controller C116-4 and flat-bed machine FB1 (1100x750); ISEL automation, Eiterfeld, Germany) was combined with a *Pulse Amplitude Modulation* (PAM) fluorometer (one-channel version of Phyto-PAM, Walz, Effeltrich; Schreiber et al 1986). F_S was measured under PAR of 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ and 500-msec pulses of white saturating light (3,000 $\mu\text{mol m}^{-2} \text{s}^{-1}$) were used to determine F_M and the ratio = $\Delta F/F_M$. The sensor, which provides excitation and measurement of fluorescence, was modified to be movable and be positioned within plant tray dimensions by an automatic steering device. The automatic PAM fluorometer system measured Φ_{II} of *A. thaliana* plants one after the other in a predefined pattern, whereby individual leaves were identified automatically (auto-focus mode) by their optimal F_S ($100 < F_S < 250$). The auto-focus mode comprised a predefined pattern used by the sensor to identify leaves for optimal measurement (centre ? all 4 corners ? centre). The sensor was positioned at a distance of less than 2 cm from leaves

measured to avoid cross-illumination. Up to 1,536 plants grown in trays (128 lines with 12 individuals) were screened per week. To obtain light saturation curves, Φ_{II} was measured using the fluorometer PAM101/103 (Walz, Effeltrich). Plants were illuminated for 10 min with white actinic light (from 6 to 1049 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$).

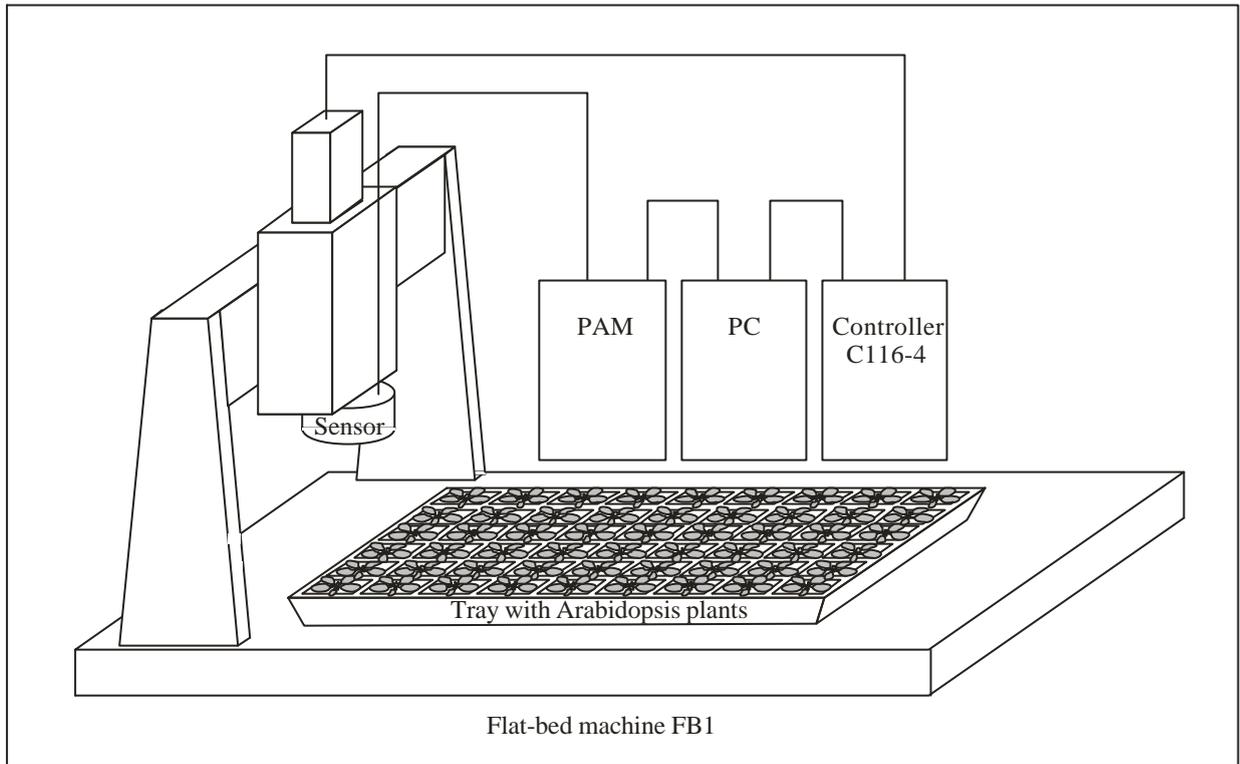


Figure 2.1. Schematic drawing of the robot used in the screening for Φ_{II} mutants.

In vivo Chl a fluorescence of single leaves was excited and detected by a pulse amplitude modulated fluorometer (PAM 101/103; Walz) as described by Meurer et al. (1996). 800-ms white light pulses of 6000 $\mu\text{mol m}^{-2} \text{s}^{-1}$ were used to determine the maximum fluorescence (F_M) and the ratio $(F_M - F_0)/F_M = F_V/F_M$. Actinic light of 80 $\mu\text{mol m}^{-2} \text{s}^{-1}$ intensity was used to drive photosynthesis. In addition, fluorescence quenching parameters qP (photochemical quenching) = $(F_M - F_S)/(F_M - F_0)$, qN (non-photochemical quenching) = $1 - (F_M - F_0)/(F_M - F_0)$ (Schreiber, 1986) and Φ_{II} were recorded.

2.9 P700 absorption measurements

A dual wavelength pulse modulation system (ED-P700DW; Walz) was used to record changes in the absorbance of P700 (Klughammer and Schreiber, 1994). The steady-state redox level of

PSI, $\Delta A/\Delta A_{\max}$ = fraction of oxidised P700/total P700 (Harbinson and Woodward, 1987) was estimated as described by Meurer et al. (1996). Immediately after oxidation of P700 by background far-red light, a saturating blue-light pulse (50 msec) was applied (XMT-103; Walz), which was sufficient to re-reduce P700⁺. The half-life of the reduced state of P700 and the time required for the far red light-driven reoxidation of P700 were recorded.

2.10 Immunoblot analysis of proteins

Membrane proteins of mutant and wild-type plants were isolated as described (Bassi et al., 1985). Membranes were incubated with antibodies specific for individual subunits of PSI (PsaC, D, E, and F) and PSII (PsbO) and signals were detected by using the Enhanced Chemiluminescence Western-Blotting Kit (Amersham).

2.11 Public databases

For the large-scale retrieval of nucleotide sequences and annotations, the NCBI 'Batch Entrez' system (Wheeler et al., 2001; <http://www.ncbi.nlm.nih.gov/Entrez/batch.html>) was employed. Protein sequences used for the design and synthesis of the 1898 primer pairs were selected for the presence of a transit peptide motif as described before (Abdallah et al., 2000).

2.12 Programming

The main body of *GST-PRIME* was written in Visual Basic (v5.0; Microsoft). The subroutine for the sequence download was implemented in Perl (v5.005) and the Windows Executable program file was generated with Perl2Exe (v4.03; IndigoSTAR).

2.13 Evaluation of secondary structure of oligos and primer-primer interactions

Secondary structure formation of oligos was determined using the Mfold program (program parameters: temperature, 37°C; increment, 10; window, 2) of GCG (Zuker, 1989; SantaLucia, 1998; Allawi and SantaLucia, 1997). Prediction of annealing events between forward and reverse primers was performed by using the default scoring matrix of the bestfit program of GCG (Devereux et al., 1984). As a measure for the annealing of forward and reverse primers, the product of the quality score obtained by Bestfit and the relative length of the complementary region was used. For nucleic acids, the Bestfit default scoring matrix has a

match value of 10 for each identical symbol comparison and -9 for each non-identical comparison (not considering nucleotide ambiguity symbols for this example). The quality score for a nucleic acid alignment can, therefore, be determined using the following equation:

$$\text{quality} = (10 \times \text{total matches}) + (-9 \times \text{total mismatches}) - (\text{gap creation penalty} \times \text{gap number}) - (\text{gap extension penalty} \times \text{total length of gaps}).$$

3. FORWARD GENETICS

The first step in the genetic dissection of photosynthetic functions is the identification of corresponding mutations. This chapter describes the set-up of a phenotypic screening system suitable for the identification of *A. thaliana* mutants with an impaired photosynthesis. The system is based on an automated PAM (Pulse Amplitude Modulation) fluorometer able to measure the fluorescence yield of PSII (see Chapter 1; Materials and Methods, Figure 2.1), and led to the identification of 4 mutations having an effect on photosynthesis.

RESULTS

3.1 An automated PAM fluorometer for the high-throughput screening of photosynthetic mutants

The average time for a Φ_{II} measurement of a plant by the automated PAM fluorometer is less than 4 sec, while 10-15 sec are required for moving the emitter/detector unit to the next plant. Consequently, measurement of Φ_{II} in trays containing 54 or 96 *A. thaliana* plants required about 20 min or 30 min per tray, respectively. The accuracy and reproducibility of Φ_{II} measurement were tested by analysing 24 wild-type plants for which an average value of 0.77 with a standard deviation of 0.01 was observed. For 90 % of plants repeated measurements resulted in identical Φ_{II} values, while for 10 % of them the maximum deviations were ± 0.01 .

Twelve sibling plants for each of 2000 lines derived from two different mutagenised populations of *A. thaliana* were analysed for the presence of Φ_{II} mutants. Siblings were sown in duplicate trays in order to minimise random events mimicking lethal phenotypes. For lines with a minimum of two out of 12 individuals having non-wild-type Φ_{II} , 30 additional siblings were analysed to confirm the mutant phenotype and to study its segregation. A total of twelve lines segregating recessive mutations were found. Five Φ_{II} mutants (0.5 %) were identified among the 1100 *En* lines, whereas 7 Φ_{II} mutants were found (0.8 %) among the 900 T-DNA lines.

3.2 Description of the isolated mutants

Nearly all of the Φ_{II} mutants showed significantly altered pigmentation (yellowish or pale green; Table 3.1), except line Pam920 (Photosynthesis Affected Mutant), showing WT pigmentation. All of the mutant lines but three grew, to different extents, slower than WT plants; only lines Pam5, Pam 7 and Pam920 did not show significant reduction in growth. The Φ_{II} values range between 0.40 and 0.83 for all the non-lethal mutations. Among them, only one mutant showed Φ_{II} values comparable to those normally observed for *hcf* mutants (line Pam234).

Table 3.1. Summary of the mutants identified by Φ_{II} forward screening.

LINE	PHENOTYPE	F _{II}	TAG
Pam2	Yellow, smaller than WT, no reversions	0.60	<i>En1</i>
Pam4	Pale green, smaller than WT, reversions	0.65	<i>En1</i>
Pam5	Yellow, reversions	0.83	<i>En1</i>
Pam6	Pale green, very small, early flowering	0.68	<i>En1</i>
Pam7	Yellow, reversions	0.83	<i>En1</i>
Pam234	Yellow, very small (<i>hcf</i>)	0.40	T-DNA
Pam483	Virescent, smaller than WT	0.63	T-DNA
Pam563	Lethal	N.D.	T-DNA
Pam672	Chlorotic	0.58	T-DNA
Pam920	Slightly paler than WT	0.70	T-DNA
Pam1241	Lethal (albino)	N.D.	T-DNA
Pam1263	Virescent, smaller than WT	0.66	T-DNA

For further examination of photoinhibitory effects, Φ_{II} of 3 mutant lines was determined for photosynthetically active radiation (PAR) from 6 to 1049 $\mu\text{mol m}^{-2} \text{sec}^{-1}$ (“light saturation curves”, Figure 3.1). On the basis of the classes of mutants defined by their phenotypic and

spectroscopic characteristics, crosses to test allelism between lines were performed. Lines Pam5 and Pam7 were found to be allelic.

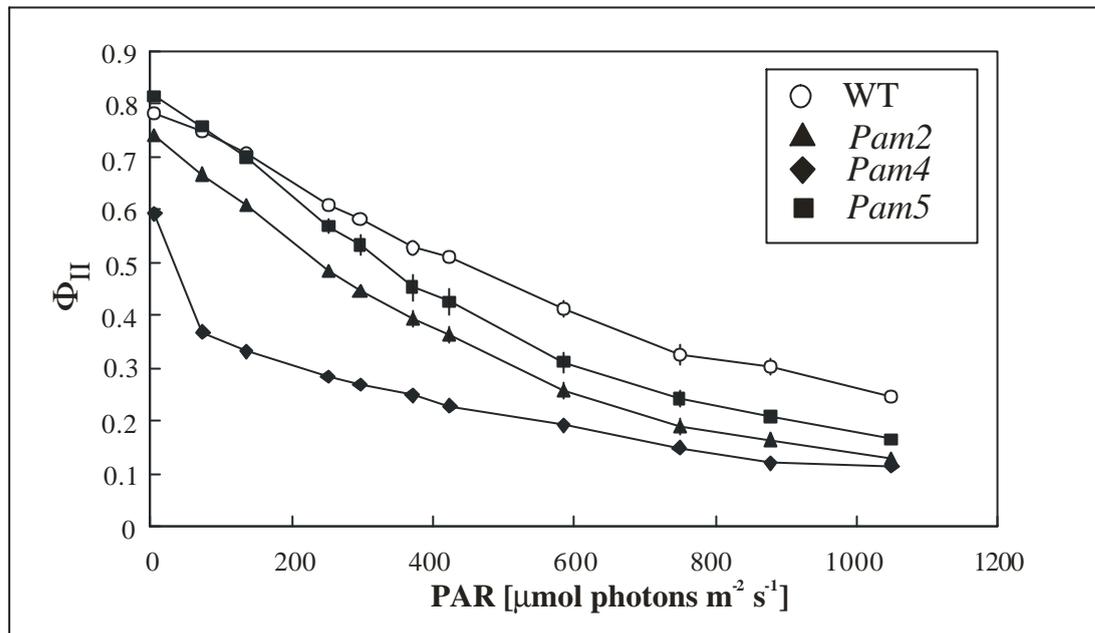


Figure 3.1. Light saturation curves of wild-type (WT), Pam2, Pam4 and Pam5 (Pam: “photosynthesis affected mutant”). Φ_{II} was determined for photosynthetically active radiation (PAR) ranging from 6 to 1049 $\mu\text{mol m}^{-2} \text{sec}^{-1}$. 5 WT plants and 5 plants for each mutant line were measured after incubation for 20 h in the dark followed by 1 h illumination with 160 $\mu\text{mol photons m}^{-2} \text{sec}^{-1}$. Φ_{II} was measured after irradiating the plants for 10 min with every PAR. Error bars indicate standard deviations.

3.3 Isolation of insertion sites

Flanking sequences of T-DNA and *En* insertions were obtained by a PCR-based approach (Table 3.2; Materials and Methods).

Table 3.2. Steps involved in the isolation of flanking regions of *En* or T-DNA insertions

1	Digestion of genomic DNA with suitable restriction enzyme (<i>Csp6I</i> or <i>Hin6I</i>)
2	Ligation of DNA fragments to compatible adaptors (APL1632 or APL 1732)
3	Linear PCR with insertion-specific primers
4	Re-amplification with nested insertion primer and adaptor-specific primer
5	Denaturation and separation on polyacrylamide gel
6	Excision, re-amplification and sequencing of candidate PCR products

For *En*-lines that harbour on average 6 transposon insertions in their genome (Wisman et al., 1998), additional outcrossing and selfing steps were introduced before the isolation of DNA sequences flanking the transposon insertion. For all of the mutant lines, AIMS experiments with 2 different enzymes and both sides of the tag have been performed (see Material and Methods). To identify which insertion caused the mutation resulting in the Φ_{II} -altered phenotype, flanking sequences of sibling plants were displayed on polyacrylamide gels (Figure 3.2). PCR bands co-segregating with 4 mutant phenotypes were excised, extracted from the gel and sequenced after re-amplification. Specific primers complementary to insertion flanking regions were designed and combined with *En* or T-DNA specific primers for segregation analyses.

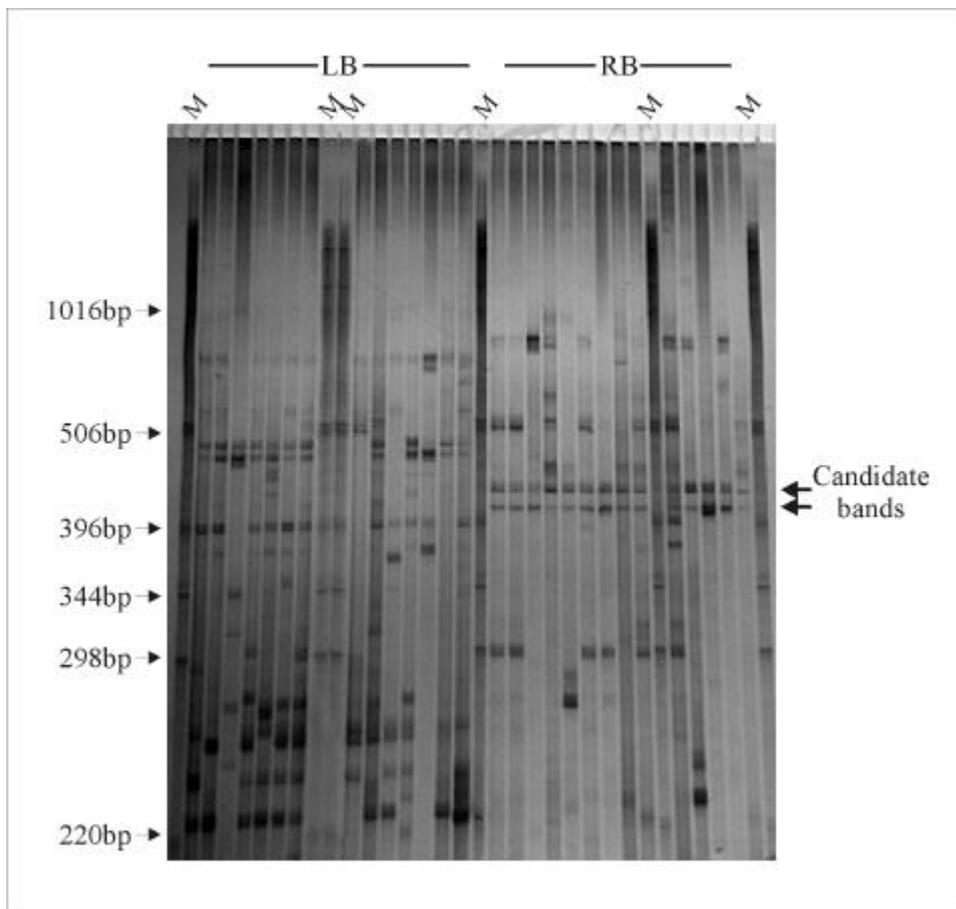


Figure 3.2. Display of genomic sequences flanking *En* insertions on 4.5% polyacrylamide gel. Flanking regions of *En*-transposon insertions from line Pam5 were amplified after *Csp6I* digestion and adapter ligation in mutant and wild-type plants from a segregating F2 population. Two bands were identified co-segregating with all mutant plants (“candidate bands”). DNA sequencing of those bands identified them as part of the CAO gene. M: 1 kb ladder; LB: left border, RB: right border. Silver staining has been performed according to Sanguinetti et al. (1994).

3.4 Identified genes

Of the 12 mutants analysed, four insertions (two *En* and two T-DNA) were identified that inactivate specific genes. Two of the genes code for proteins with a putative N-terminal transit peptide for chloroplast import. The mutation in lines Pam5/7 is caused by an *En*-insertion within the *CAO* gene, coding for a plant-specific component of the chloroplast signal recognition particle pathway. The corresponding *chaos* mutant has been identified and characterised previously (Klimyuk et al. 1999). The second *En*-mutation, identified in the line Pam4, interrupts the *psaE1* gene encoding for the subunit E of photosystem I. The third gene identified contains a T-DNA insertion and encodes a putative mitochondrial protease (line Pam563). The last mutation identified is caused by a T-DNA insertion in the intergenic region between the root-specific iron transporters genes IRT1 and IRT2 (line Pam672; Korshunova et al., 1999; Vert et al., 2001).

DISCUSSION

An automated screening system for the identification of mutants affected in photosynthesis by their altered effective quantum yield of photosystem II (Φ_{II}) has been established. This procedure allows to screen for non-lethal mutants with a maximum throughput of several thousand plants per week. Among 2000 insertionally mutagenised lines (corresponding to 24,000 plants), 12 Φ_{II} mutants were identified, most of them with altered pigmentation and 1 with only minor pigmentation defects and no significant reduction in growth rate. Few other screening procedures leading to the isolation of photosynthetic mutants of *Arabidopsis*, and based on chlorophyll fluorescence measurements, have been described (see Introduction).

The screening procedure described here allowed the identification of mutants not found by *hcf* mutant screenings, as demonstrated by the lack of a visible *hcf* phenotype in most of the Pam mutants. The extent of overlap between the NPQ and our screening approaches has not been determined; it could be defined if either a comprehensive list of mutated genes isolated with the two methods will be available, or when reciprocal tests of NPQ mutants with the Φ_{II} screening and *vice versa* will be performed. Shikanai and co-workers (Shikanai et al. 1999) observed, for about half of their NPQ mutants, significant changes of Φ_{II} that will probably make them detectable by a direct Φ_{II} mutation screen. In the same way, the screen performed by Niyogi and co-workers can identify Φ_{II} mutants, but none of them have been further characterised (K. Niyogi, personal communication).

The advantage of tagged populations is that the presence of a tag sequence at the mutated locus makes the task of isolating the mutated gene easier. Several techniques are available to isolate the DNA regions flanking the insertion sites, the most common of which are plasmid rescue (Perucho et al., 1980), inverse PCR (Ochman et al., 1988), interlaced PCR (Liu et al., 1995) and AIMS (Frey et al., 1998). The prerequisites for an efficient system to identify tagged mutations on a large scale are (i) ease of isolation of regions flanking the tag and (ii) the possibility to isolate directly the insertion associated to the mutant phenotype. In general, plasmid rescue has two major drawbacks with respect to the other techniques. First, it requires the use of molecular tags containing a replication origin and a selectable marker for *E. coli*, thus its use is restricted mainly to T-DNA populations. Second, all the candidate flanking regions have to be checked for segregation, because plasmid rescue is not suited to any kind of display technique helping their selection. On the contrary, all the other techniques mentioned above are based on PCR amplifications. Displaying the amplification products on agarose or

polyacrylamide gels it is thus possible to directly check the segregation of the single bands with the phenotype. We found AIMS in combination with a display on a polyacrylamide (PA) gel an efficient procedure. Polyacrylamide gels offer a high resolution, enabling the separation of bands comigrating in agarose gels. Moreover, excision, reamplification and sequencing of bands from PA gels reduce the risk of contaminating unrelated products.

A total of 12 Φ_{II} mutants have been isolated from the 2000 *En*- and T-DNA mutant lines screened. For four Φ_{II} mutants the corresponding genes were identified; in two cases the products of the identified genes are imported into the chloroplast. One of those proteins (PsaE) is directly involved in photosynthesis, being part of photosystem I (Jansson et al., 1996), while the other is involved in protein import into chloroplasts (CAO, Klimyuk et al., 1999). In the case of the mutation of the gene encoding a putative Fe-transporter, the polypeptide product is not imported into the plastids. However, the effect on photosynthesis can be easily explained since iron is essential for the Fe-S clusters of several electron transport proteins of the thylakoid membrane. For the mutation in the mitochondrial protease a secondary effect on photosynthesis seems probable.

Another important consideration for an efficient analysis of photosynthesis-related mutants is the kind of population to be used (Thorneycroft et al., 2001). One of the advantages of using a transposon-mutagenised population is the possibility to confirm the mutation-phenotype relationship by identifying somatic and germinal revertants, while for T-DNA mutant complementation analysis is necessary. A second advantage of transposon mutagenesis is that transposons are less prone to complex rearrangements than T-DNA, thus allowing an easier isolation of the flanking region by PCR-based techniques. In contrast, clustering of several copies as well as rearrangements and deletions mainly on its right border often hinder the use of T-DNA as efficient tool for identification of mutants.

However, transposons can be present in high copy numbers (Wisman et al., 1998), making several outcrosses necessary before isolation of the flankings. Moreover, their distribution in the genome is normally not random, causing an uneven tagging frequency at different loci (Machida et al., 1997). Transposon-tagged populations also inevitably contain a fraction of non-tagged mutations due to footprints left by former transposition events. The genetic instability of transposon-induced mutations can finally become a problem if the insertion is in a non-coding region of the gene, making it impossible to stabilise such a mutant by footprints.

In our Φ_{II} mutant screen, only knock-out populations were used, whereas for the *hcf* and NPQ screens, chemical or physical mutagens were employed. Although the map-based cloning of chemically or physically-induced mutants of *Arabidopsis* can be accomplished nowadays in a relatively straightforward way, insertion tagging is the method of choice for fast and systematic gene isolation.

Significant drawbacks of current screening strategies performed in plant trays are their demand for extensive greenhouse space and the difficulty to apply stress conditions before fluorescence measurement. These drawbacks can be overcome by adopting fluorescence imaging systems using modulated-light excitation (Nedbal et al. 2000). The main technical problem encountered to make this kind of screening operative is to reach a sufficient intensity and uniformity of illumination on the measuring area to perform the readings. Current imaging systems are suitable for the efficient analysis of areas of about 0.01 m², a size feasible to study objects with dimensions of Petri dishes. This has been demonstrated by Niyogi and co-workers who have applied a custom-made imaging device to screen for NPQ mutants of *Chlamydomonas* and *Arabidopsis* (Niyogi et al. 1997, Niyogi et al. 1998). Imaging systems are not restricted to NPQ measurements but can also be applied to measure a range of chlorophyll fluorescence induction parameters. The major advantage of imaging system based on relatively small-sized culture devices is the possibility to apply various kinds of stresses to the plants to be analysed or to simulate specific environmental conditions. This should enable the development of screening strategies leading to the isolation of conditional mutants identified after specific stress or environmental treatments.

4. CHARACTERISATION OF A PSAE1 KNOCKOUT

One of the mutants isolated by the phenotypic screen (see previous chapter), *psae1-1*, has been characterised in detail from the molecular, physiological and biochemical point of view. The structural and functional alterations of PSI caused by this mutation are described.

RESULTS

4.1 *En*-transposon tagging and cloning of the Pam4 locus

The causal relationship between the knock-out of *psae1* and the observed photosynthesis phenotype was demonstrated by identifying two additional independent mutant alleles of *psae1* that exhibited the same phenotype as *psae1-1*. These alleles, identified by a reverse genetic approach, have been called *psae1-2* and *psae1-3*, respectively (Figure 3.1) and exhibit an Φ_{II} identical to the one of *psae1-1*. Like *psae1-1*, also the *psae1-2* mutation has been caused by an *En* insertion in the first intron, 293 bases downstream with respect to the former one. The *psae1-3* mutation has been caused by an insertion in the second intron.

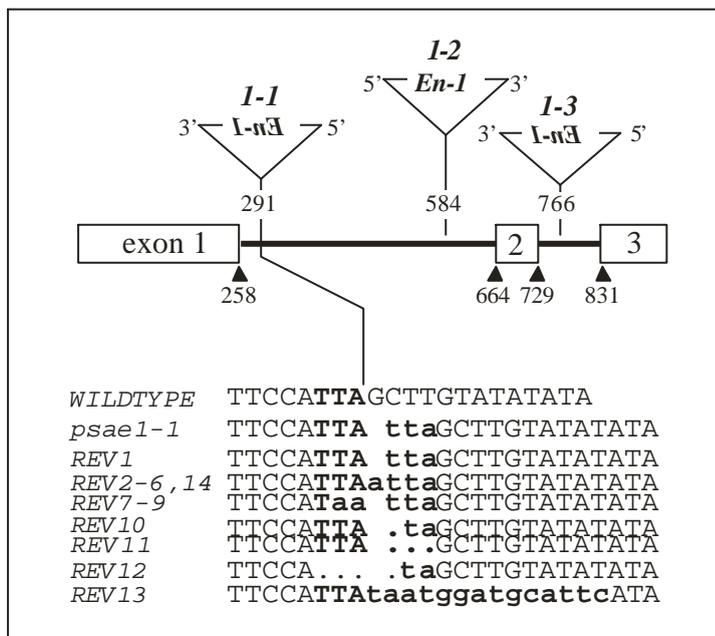


Figure 4.1. Transposon tagging of the *psae1* locus.

DNA sequences of the empty donor sites of DNA from 13 independent germinal revertants of *psae1-1* found among 81 screened plants (REV1 to 13) and from one somatic *psae1-1* revertant sector (REV14) were obtained by PCR amplification. The footprints left at each locus after *En* transposition are indicated by bold lowercase letters, bold uppercase letters indicate the target site in the wild-type

Sequencing of the insertion site amplified by PCR from 13 germinal revertants and one revertant leaf sector showed footprints caused by the excision of the *En* transposon (Figure 4.1).

4.2 Expression of *psaE1* and *psaE2* in wild-type and mutant plants

The PsaE subunit shows a high degree of conservation from cyanobacteria to higher plants. In *Arabidopsis*, the disrupted *psaE1* gene is highly homologous to the *psaE2* gene (Figure 4.2).

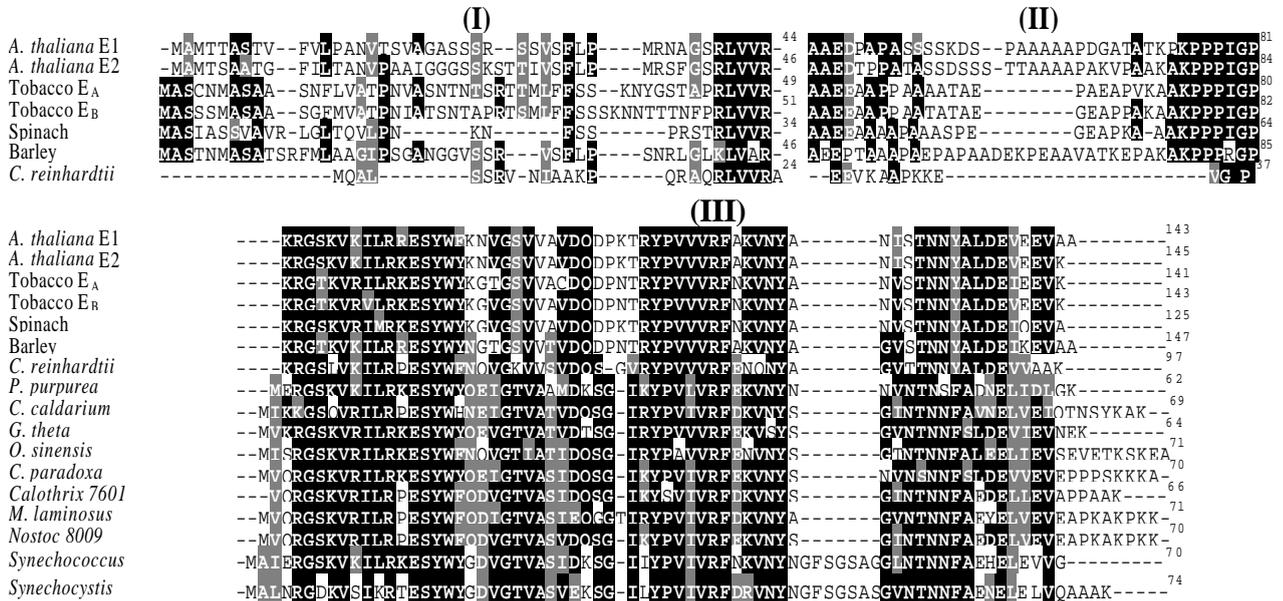


Figure 4.2. Comparison of PsaE proteins from higher plant, algal and cyanobacterial sources.

The amino acid sequences of the *Arabidopsis* proteins PsaE1 and PsaE2 were compared with those of tobacco (Obokata et al, 1994), spinach (Muench et al, 1988), barley (Anandan et al, 1989; Okkels et al, 1988), the green alga *Chlamydomonas reinhardtii* (Franzen et al, 1989), the red algae *Porphyra purpurea* (Reith, 1992) and *Cyanidium caldarium* (GenBank accession no: AF022186), the cryptophyte alga *Guillardia theta* (Douglas and Penny, 1999), the diatom *Odontella sinensis* (Kowallik et al, 1995), the glaucocystophycean alga *Cyanophora paradoxa* (GenBank accession no: U30821) and the cyanobacteria *Calothrix* sp. PCC 7601 (Mann et al, 1991), *Mastigocladus laminosus* (GenBank accession no AF093820), *Nostoc* PCC8009 (GenBank accession no AF148219), *Synechococcus* sp. PCC 7002 (Zhao et al, 1993) and *Synechocystis* sp. PCC 6803 (Chitnis et al, 1989). The three-part alignment shows the N-terminal chloroplast transit peptide of higher plant PsaE (I), the N-terminal extension of mature PsaE proteins specific to higher plants (II) and the C-terminal domain of PsaE from higher plants that is highly similar to the corresponding segment of algal and cyanobacterial PsaE proteins (III). Note that the N-terminal extension of the higher-plant proteins is rich in alanine and proline and highly diversified. The two *Arabidopsis* PsaE proteins show amino acid exchanges at 15 of 38 positions, whereas in tobacco 6 of 31 positions are found to be variable. Strictly conserved amino acids are highlighted by black boxes, whereas grey boxes indicate closely related amino acids.

Northern analysis with a probe derived from the *psaE1* gene showed a strong decrease in the level of the corresponding transcripts in mutant plants (Fig. 4.3a). The residual transcription can be attributed to the presence of somatic reversions, which have been estimated to affect around 10 % of the total plant tissue.

To measure the abundance and relative expression levels of *psaE1* and *psaE2* transcripts, quantitative reverse-transcription PCR (QRT-PCR) was performed on cDNA preparations from WT and *psae1-1* mutant plants (Fig. 4.3b).

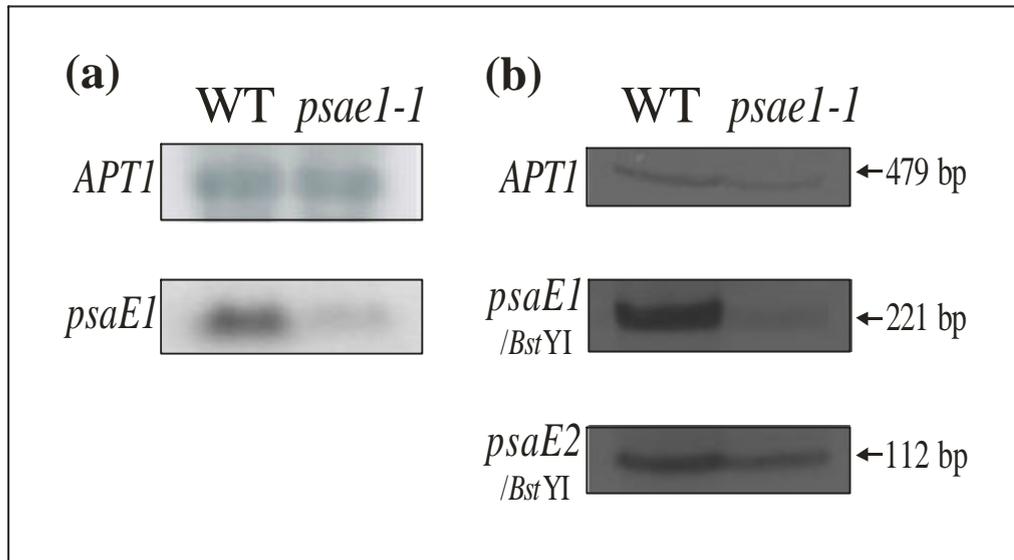


Figure 4.3. Analysis of *psaE* transcripts

(A) Northern analysis of *psaE* transcripts. 30- μ g aliquots of total RNA were analysed using a *psaE1* cDNA fragment. The blots were reprobbed with a cDNA fragment derived from the *APT1* gene to control the loading.

(B) Detection of *psaE1* and *psaE2* transcripts by QRT-PCR. The analysis on 4.5 % (w/v) polyacrylamide gel was performed with ^{33}P -labeled QRT-PCR products obtained after PCR for 16 cycles with *psaE1/psaE2*-specific primers [E(91/85)s and ^{33}P -labeled primer E(798/868)as] and control primers specific for *APT1* (Cowling et al., 1998). For determination of the *psaE1:psaE2* transcript ratio, PCR products were digested with *BstYI* before gel analysis. Note that only the 3' terminal restriction product was labelled by the ^{33}P -labeled primer.

The adenine phosphoribosyltransferase (*APT1*), expressed at a low level in all tissues of *Arabidopsis* (Moffat et al., 1994), was used as a control. To avoid preferential amplification of one template, primers complementary to regions whose sequence is identical between *psaE1* and *psaE2* transcripts were selected. The length of the amplification products from the cDNAs of the two genes differed by only 3 bp. The gene-specific QRT-PCR products could be distinguished by digestion with *BstYI*. In WT plants both *psaE* transcripts were relatively abundant in comparison to the *APT1* gene; the *psaE1* transcript was three times more abundant than that of *psaE2*. This is in accordance with database searches for *Arabidopsis* *psaE* ESTs, which identified 84 ESTs derived from *psaE1* and 15 ESTs derived from *psaE2*. In the *psae1-1* mutant, a strong reduction of *psaE1* RNA was observed, whereas the amount of *psaE2* transcript was not significantly changed.

4.3 The abundance of PsaE, C and D proteins is significantly reduced in mutant plants

Western analyses demonstrated that the PsaE subunit in thylakoids extracted from mutant plants is significantly reduced compared to that from wild-type (Fig. 4.4). Also the other peripheral stromal subunits of PSI, PsaC and PsaD, were significantly reduced in abundance; in contrast, no change in the PSII subunit PsbO or in PsaF, a non-peripheral subunit of PSI, was found in mutant plants.

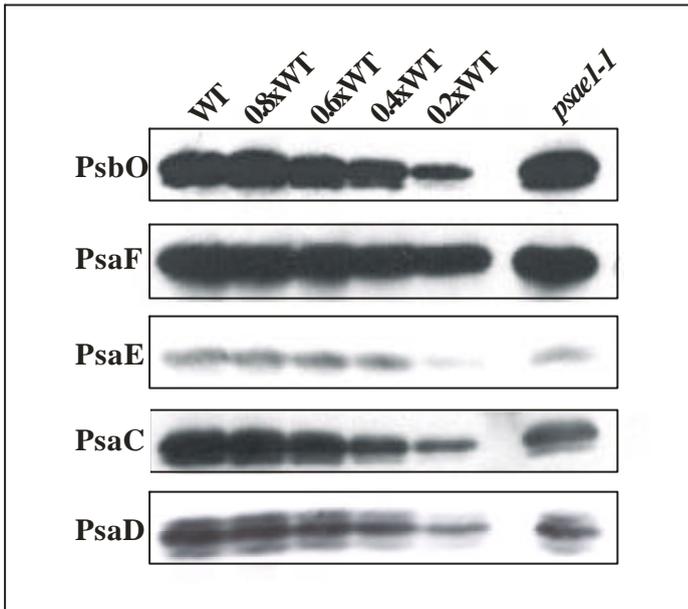


Figure 4.4. Immunoblot analysis of the *psae1-1* mutant and WT plants. Samples of thylakoid membranes equivalent to 15 μg of chlorophyll were loaded in lanes WT and *psae1-1*. Decreasing amounts of WT thylakoid membranes (12, 9, 6, 3 μg of chlorophyll) were loaded in lanes: 0.8x, 0.6x, 0.4x and 0.2xWT. Four replicate filters were immunolabelled with antibodies raised against PsbO, PsaF, E, C, D.

4.4 Alteration of the redox states of PSI and PSII in *psae1-1* mutants

The parameters of chlorophyll fluorescence induction, together with P700 absorbance, were determined to characterise the electron flow in *psae1-1* mutants (Table 4.1). The fraction of Q_A , the primary electron acceptor of PSII, present in the reduced state (1-qP) is increased more than four-fold in *psae1-1* mutants. The effective quantum yield of PSII (F_{II}) is reduced in mutants, whereas q_N is slightly increased in *psae1-1*, indicating photoinhibitory effects. To confirm these photoinhibitory effects, F_{II} was measured in mutant and wild-type plants at different photosynthetically active flux densities. Even at relatively low light intensities a strong decrease in F_{II} was found in mutant plants (Fig. 4.5). Under steady state conditions, *psae1-1* mutant plants showed an increase in the fraction of non-oxidised P700 molecules, as indicated by a significant decrease of the $\Delta A/\Delta A_{\text{max}}$ ratio (see Table 4.1). To confirm this observation, the rates of reduction and re-oxidation were investigated (in terms of $t_{1/2\text{red}}$ and $t_{1/2\text{ox}}$). Interestingly, $t_{1/2\text{red}}$ was only negligibly decreased in the mutant, suggesting normal or even faster reduction rates of P700, whereas $t_{1/2\text{ox}}$ was significantly increased in mutants.

Table 4.1. Spectroscopic characterisation of *psae1-1* mutants and WT plants.

	WT	<i>psae1-1</i>
I-qP	0.06	0.26
qN	0.23	0.26
F_V/F_M	0.83	0.77
Φ_{II}	0.76	0.53
$\Phi_A/\Phi_{a_{max}}$	0.2	0.12
$t_{1/2red}$	78 msec	72 msec
$t_{1/2ox}$	0.64 sec	1.2 sec

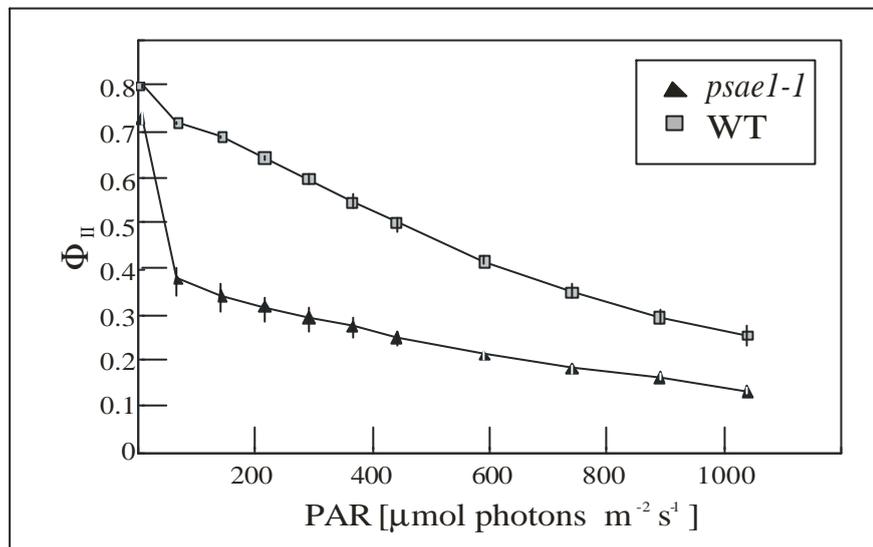


Figure 4.5. Light saturation of the *psae1-1* mutant and WT plants. Φ_{II} was determined for a range of photosynthetically active radiation (PAR) from 1 to 1039 $\mu\text{mol m}^{-2} \text{s}^{-1}$ for 5 *psae1-1* mutant and 5 WT plants, after incubation for 20 h in the dark followed by 4 h illumination with 160 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ and a 10-min incubation in the dark prior to measurement. Error bars indicate standard deviations. For every PAR, Φ_{II} was measured after irradiating for 15 min with the appropriate amount of light.

4.5 *psae1-1* mutants show light green pigmentation and an increased chlorophyll fluorescence phenotype

psae1-1 mutants can be distinguished from WT plants by their light-green pigmentation (Figure 4.6a). Without preincubation in the dark, the levels of chlorophyll fluorescence emitted by wild-type and mutant plants on illumination with UV light are nearly identical. However, after long periods of incubation in the dark, *psae1-1* plants show increased chlorophyll fluorescence compared to WT plants (Figure 4.6b).

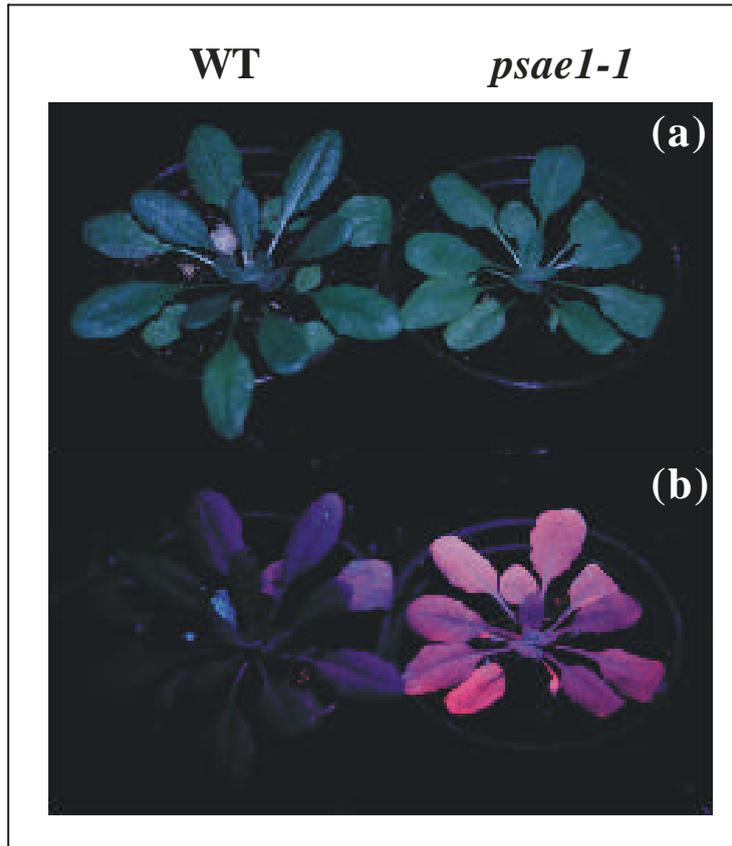


Figure 4.6. WT and *psae1-1* mutant plants illuminated with white light (a) and UV light (B-100AP/R, UVP Inc., California, USA) after a 20-h incubation in the dark (b)

4.6 Decreased growth

The growth of *psae1-1* mutants was compared to that of WT plants by non-invasive image analysis (Leister et al., 1999). Under optimal greenhouse conditions growth was significantly reduced (by up to 50%) in *psae1-1* mutants with regard to both plant size and growth rate (Fig. 4.7).

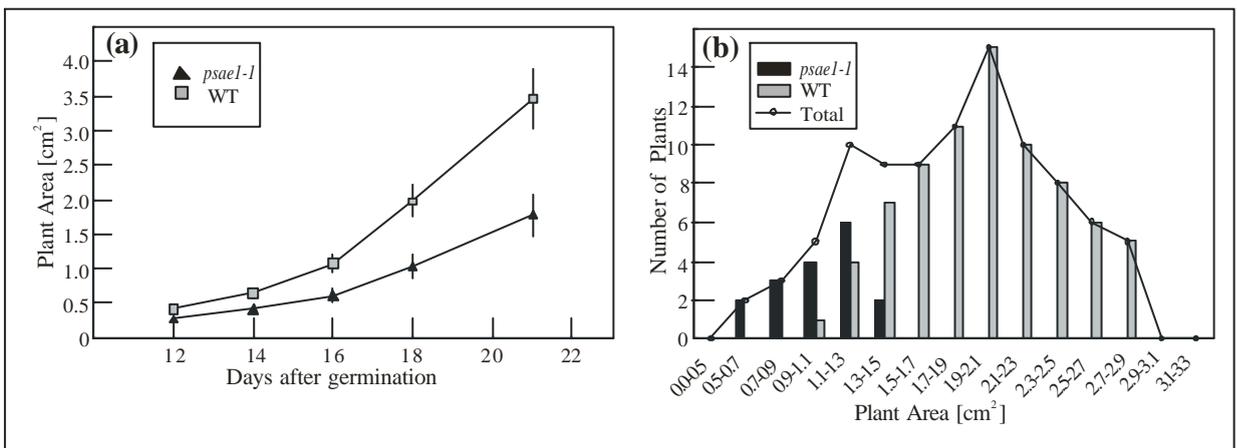


Figure 4.7. Growth curves of *psae1-1* mutants (n=18) compared to WT plants (n=77).

(A) Plant areas measured in the period from 12 to 21 days after germination

(B) Distribution of plant areas on day 18 after germination.

Error bars indicate standard deviations.

DISCUSSION

The phenotype caused by the *psae1* mutation includes a light-green pigmentation, delayed growth and alterations in photosynthetic electron flow. As a consequence of the mutation, the electron flow is altered in both PSII and PSI. The strong decrease in the effective quantum yield of PSII and the increase in the amount of the primary electron acceptor of PSII, Q_A , present in the reduced state (1-qP) both indicate impairment at some point downstream in the electron transport chain. Given the stromal location of PsaE, the stromal side of PSI is the most likely candidate for the observed “bottleneck”. The nearly normal reduction rate and the significant increase of P700 re-oxidation half time strongly support this hypothesis.

According to cross-linking experiments and structural models based on crystallography and NMR analysis in cyanobacteria, PsaE is located in close proximity to PsaC, D and F (Jansson et al., 1996; Oh-oka et al., 1989). The three peripheral subunits PsaC, D and E form a compact and interconnected structure on the stromal side of PSI [the stromal ridge, (Kruip et al., 1997; Klukas et al., 1999)] that is thought to be necessary for electron transfer to ferredoxin (or flavodoxin in cyanobacteria).

For the cyanobacterial PSI it has been demonstrated that PsaE has a stabilising effect on the stromal ridge (Weber and Strotmann, 1993) and that its absence impairs fast electron transfer between PsaC and the soluble electron acceptor (Rousseau et al., 1993). Both observations are supported by crystallographic data showing that one loop of PsaE (about 18 amino acids long) is sandwiched between PsaC and the PSI core (Klukas et al., 1999).

Western analysis was thus performed to confirm the depletion of PsaE in *psae1-1* mutants and to test whether any of the other subunits of the stromal ridge were affected. A significant reduction in the amount of the three extrinsic subunits PsaE, C and D was observed, whereas for PsbO and PsaF, used as controls for PSII and the non-peripheral stromal part of PSI, no differences between mutant and WT plants were found (Fig. 3.4). According to these findings, one can conclude that the depletion of the PsaE subunit leads to a more general decrease in the polypeptide level of the whole stromal ridge of PSI. In particular, the decrease in the level of the PsaC subunit is sufficient to explain the limited electron flow and the altered P700 re-oxidation kinetics in the *psae1-1* mutant, because PsaC itself is directly involved in electron transport, binding the terminal electron acceptors F_A and F_B . The concomitant decrease in the

level of PsaD, which has been demonstrated to play a major role in ferredoxin docking (Lelong et al., 1994; Xu et al., 1994a; Zilber and Malkin, 1988), may further contribute to the limitation in electron flow between P700 and ferredoxin indicated by the spectroscopic data. Whether this destabilisation of the reducing site of PSI is caused, at least in part, by photoinhibitory effects and/or repair mechanisms remains to be investigated.

Interestingly, cyanobacterial mutants lacking PsaE do not differ in growth rate from the wild-type when grown in normal photoautotrophic conditions (Chitnis et al., 1989; Zhao et al., 1993). Only under condition of low light or low CO₂ availability, cyanobacterial *psaE* mutants grow more slowly than the wild-type, and do not grow at all under photoheterotrophic conditions (Zhao et al., 1993). Similarly, studies of PsaE-less *Synechocystis* strains showed that PsaE is required for high-affinity binding of flavodoxin semiquinone to PSI but is not essential for the photoreduction of flavodoxin and NADP⁺ (Meimberg et al., 1998; Xu et al., 1994). In contrast, in *Arabidopsis* a knockout mutation in only one of the two functional *psaE* genes of *Arabidopsis* causes a severe phenotypic effect. Despite the high homology between PsaE1 and PsaE2 (Fig. 3.2), this unexpected finding raises the question of whether this is (i) merely due to a dosage effect caused by the observed reduction – by about 75 % - in total *psaE* transcripts or (ii) is due to functional diversification of the two *psaE* gene products. Chromosomal duplication in *Arabidopsis* is a quite common phenomenon, involving about 60% of the genome (The Arabidopsis Genome Initiative, 2000). The functional relevance of duplications varies however from gene to gene. The conservation in higher plants of PSI duplicated genes could reflect the need to activate different signal transduction pathways, thus increasing the flexibility of the plant cell in response to environmental stimuli. This seems to be the case for PsaD, as the genes for PsaD1 and PsaD2 are differentially expressed during leaf development (Yamamoto et al., 1993). Nonetheless, no tissue-specific or developmental differences could be detected between the expression of the two *psaE* genes in *N. sylvestris* at either mRNA or protein level (Obokata et al., 1993; Obokata et al., 1994). A second possibility is that diploid plants that harbour two *psaE* genes can achieve a higher level of expression of this subunit, although the observed severity of the *psae1-1* mutant phenotype seems at odds with the idea that PsaE1 could be functionally replaced, at least in part, by PsaE2. Further studies are in progress in order to understand the biological significance of duplicated PSI subunits. In particular, the identification of a mutant in the *psae2* gene (see chapter 5) should address the question about its role in presence and absence of the PsaE1 polypeptide.

5. REVERSE GENETICS FOR PSI SUBUNITS

To genetically dissect the structural and functional impact of the single subunits of PSI on its overall stability, an extensive reverse genetic approach to collect as many mutations as possible in the genes encoding PSI subunits was started.

RESULTS

5.1 Screening and stabilisation of *En*-tagged mutations of *psaG*, *psaH2* and *psaK* genes

The *En*-mutagenised population at the MPI in Cologne was the first to be screened according to Wisman et al. (1998). Three lines were found to carry a transposon insertion in the genes coding for G, H2 and K subunits of PSI (genes *psaG*, *psaH2* and *psaK* respectively, see Table 5.1, Figure 5.1a,b and c). The insertions in *psaG* and *psaH2* genes were located, respectively, 34 and 60 bp downstream of the start codon. For each of the two mutations several stable mutants were identified as described in materials and methods; the progenies of mutants *psag-1.4* (footprint +4) and *psah2-1.4* (footprint +5) have been used for all the experiments described in the following (see Figure 5.1a, b). In both cases the resulting frameshifts caused the complete disruption of the aminoacid sequence of the corresponding polypeptides.

psak-1 originated from a defective transposition event causing the deletion of more than 7600 bp of the WT *En1* transposon. The *En1* deletion fragment resulted in a stable insertion in the coding region of the *psaK* gene and caused a frameshift disrupting the corresponding reading frame (Fig. 5.1c).

5.2 T-DNA insertion mutants of *psaE2*, *psaN* and *psaL*

To obtain additional PSI knockout mutants, the Tom Jack Lines and the AFGC populations (see material and methods) were screened. Three additional T-DNA insertions were identified for the genes *psaE2* and *psaN* in the AFGC population and for *psaL* in the Jack lines (see table 5.1, Figure 5.1d, e and f). *psae2-1* contains a T-DNA insertion located 44 bp downstream of the start codon, presumably resulting in a complete knockout. In *psan-1* the insertion is located in

the first exon, 186 bp downstream of the ATG. In the case of *psal-1*, the T-DNA insertion point is at position -96 of the *psaL* transcript.

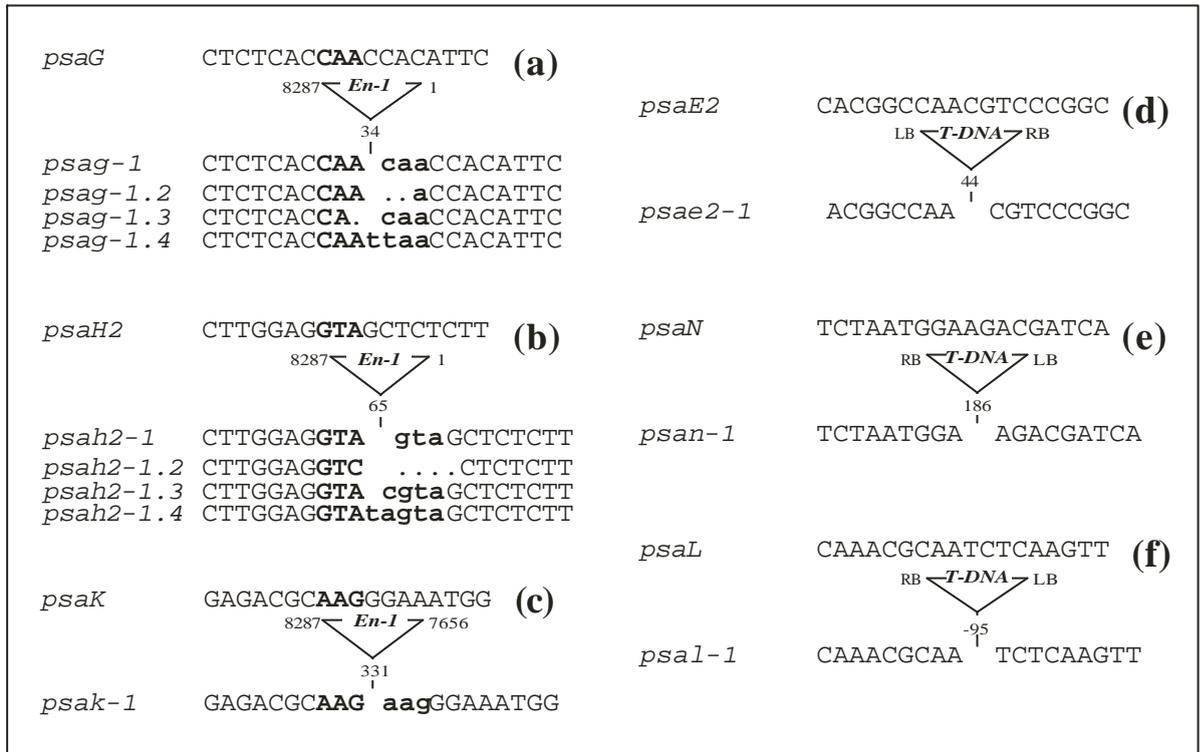


Figure 5.1. Mutations in PSI genes *psaG*, *psaE2*, *psaH2*, *psaK*, *psaL*, *psaN*. The footprints left at each locus after *En* transposition are indicated by bold lowercase letters, bold uppercase letters indicate the target site in the wild-type.

5.3 Generation of double mutants

Several of the mutant lines have been crossed to obtain multiple mutants. The double homozygous mutants G/H2, G/K, G/E1, E1/H2, K/H2, E/1K, E1/E2 have been obtained.

Detailed characterisation of *psag-1.4*, *psah2-1.4* and *psag-1.4/psah2-1.4* mutant lines has been carried out.

5.5 Growth behaviour of *psag-1.4*, *psah2-1.4*, *psag-1.4/psah2* mutants and WT plants

The knockout of subunit G had an effect on mutant growth: plants were significantly reduced in size in comparison to the WT (see Figure 5.2 and 5.3) and they showed a delay in flowering time of about 8 days (figure 5.4). On the other hand, the *psah2-1.4* mutation caused an early flowering (about 3 days before the WT, see Figure 5.4) and, interestingly, a marked increase in plant size under optimal greenhouse conditions (Figure 5.3). The double mutant *psah2-1.4* showed a decrease in growth rate comparable to that of the *psag-1.4* mutant, but a flowering time intermediate between that of the single mutants.

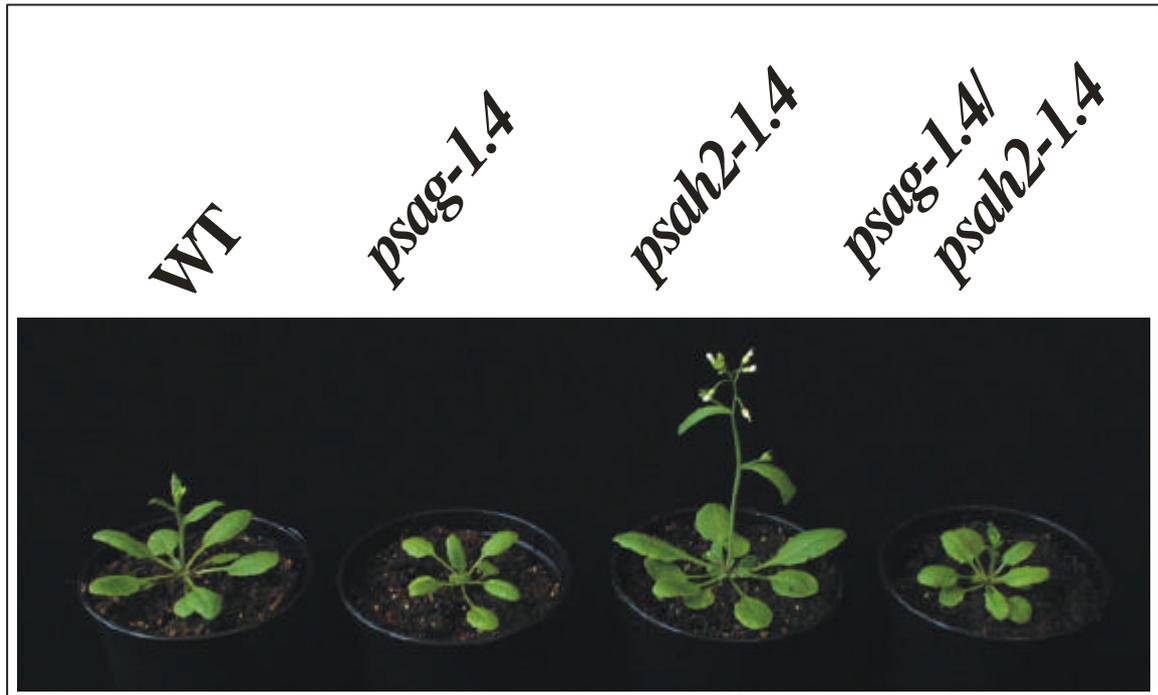


Figure 5.2. WT and *psag-1.4*, *psah2-1.4*, *psag-1.4/psah2-1.4* mutant plants.

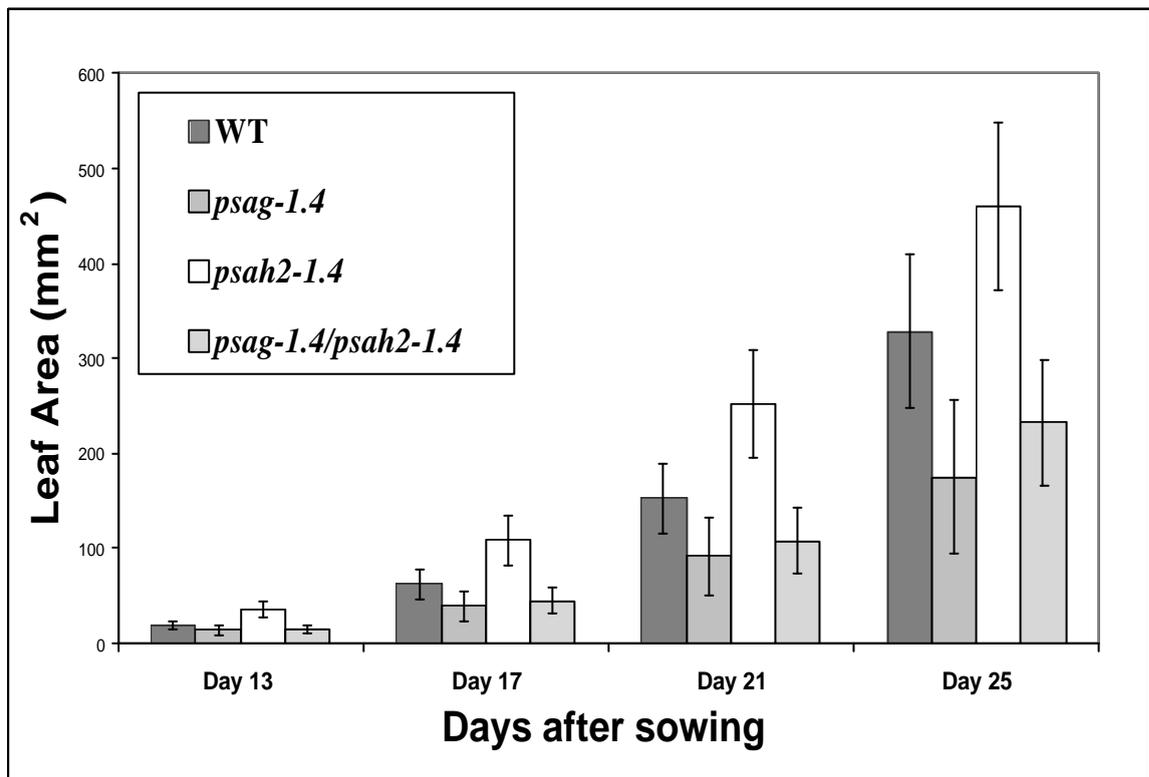


Figure 5.3. Growth curves of *psag-1.4*, *psah2-1.4*, *psag-1.4/psah2-1.4* mutants compared to WT plants (n=54). Leaf area has been measured at day 13, 17, 21, 25 after sowing. Error bars indicate standard deviations.

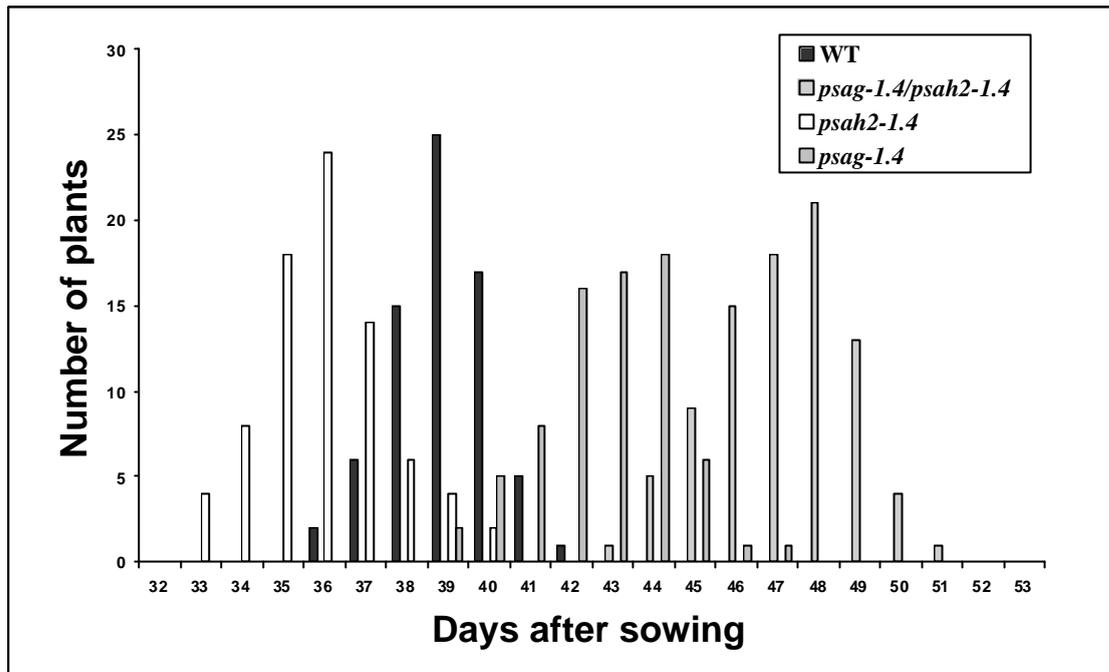


Figure 5.4. Distribution of flowering time (days after sowing) for *psag-1.4*, *psah2-1.4*, *psag-1.4/psah2-1.4* and WT genotypes. A total of 70 plants per genotype were analysed.

5.6 Expression of *psaG* and *psaH2* in wild-type and mutant plants

Both the stable single mutants have been generated through a footprint causing a frame-shift. To test whether this affected the stability of the transcripts, Northern blot analysis was performed: a strong decrease of the mRNA level was observed for both transcripts. In particular, the transcript of *psaG* almost disappears (Figure 5.5, lanes *psag-1.4*, *psah2-1.4*), while the residual signal detected by the *psaH* probe is most probably due to cross-hybridisation to the mRNA of a second gene coding for PsaH, *psaH1*.

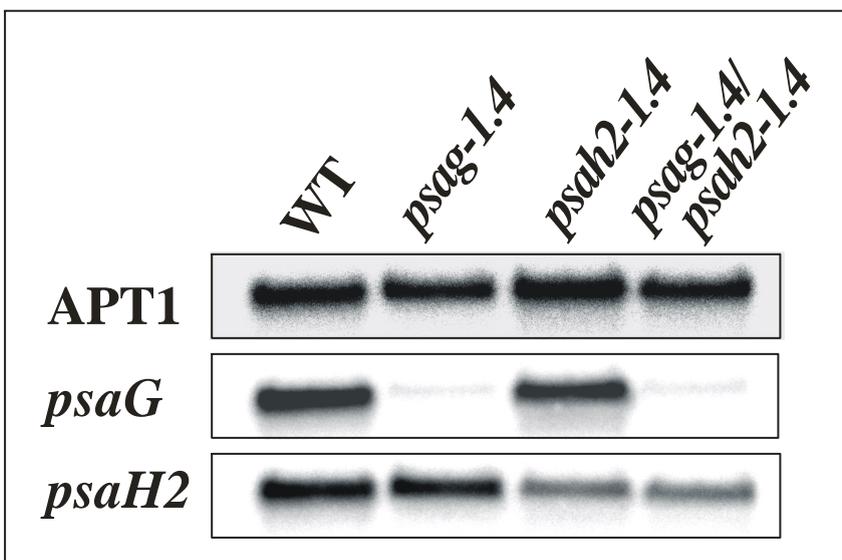


Figure 5.5. Northern blot of WT, *psag-1.4*, *psah2-1.4*, *psag-1.4/psah2-1.4* mutants. APT1, Adenine phosphoribosyltransferase (Moffat et al., 1994). Probes for the coding regions of *psaG* and *psaH2* were obtained by PCR from leaf cDNA with the primer pairs used for the screening (see Materials and Methods).

5.7 PSI polypeptide and thylakoid pigment composition in wild-type and mutant plants

Western analyses demonstrated that the *psag-1.4* mutation leads to complete loss of PsaG in mutant thylakoids, while about 70% of the wild-type level of PsaH is still present in *psah2-1.4* plants (Figure 5.6), due to *psaH1* (Naver et al., 1999). The absence of PsaG also affected the abundance of several other PSI polypeptides, in particular of PsaL (present at about 40% of WT levels). The abundance of the LHCI proteins was also affected: the amount of Lhca2 was about 70%, and of Lhca4 about 110%, that of wild-type.

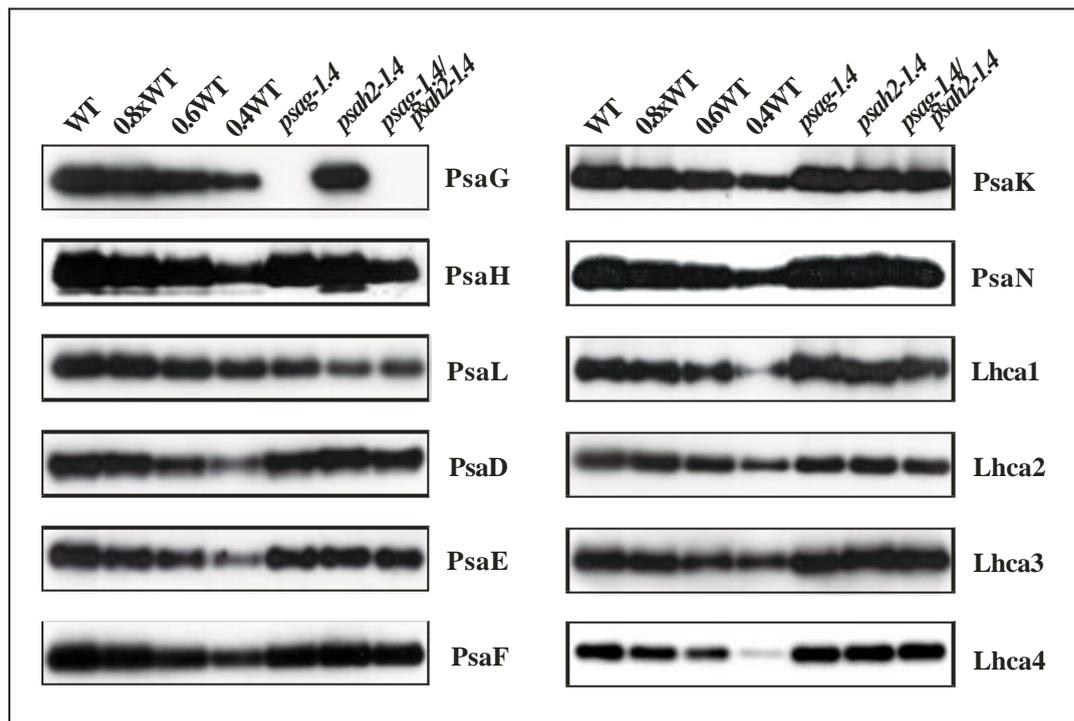


Figure 5.6. Immunoblot analysis of the *psag-1.4*, *psah2-1.4*, *psag-1.4/psah2-1.4* mutants and WT plants. Samples of thylakoid membranes equivalent to 5 μ g of chlorophyll were loaded in lanes *psag-1.4*, *psah2-1.4*, *psag-1.4/psah2-1.4*. Decreasing amounts thylakoid membranes (5, 4, 3, 2 μ g of chlorophyll) were loaded for the WT (lanes 0.8x, 0.6x, 0.4x and 0.2xWT). Replicate filters were immunodecorated with antibodies raised against the PSI subunits whose names appear on the right of each panel.

The partial knockout of PsaH led to a slight decrease in the level of PsaE, F and K. In addition, we detected only 30-40% of the normal amount of PsaL in *psah2-1.4*. This secondary loss of PsaL is more marked than that seen in PsaH-less co-suppression lines, which have 60% of wild-type PsaL levels (Lunde et al., 2000). In the double mutant *psag-1.4/psah2-1.4*, an additive effect of the two mutations on the amount of the H subunit was observed (50% of WT in *psag-1.4/psah2-1.4*, 70% in *psah2-1.4*, and 80% in *psag-1.4*). Weak additive effects of the two mutations were observed for PsaN and Lhca1.

Additional consequences of the mutations were studied by examining leaf pigment composition by HPLC (Table 5.2). In the PsaG knockout lines *psag-1.4* and *psag-1.4/psah2-1.4*, the level of the xanthophyll cycle pigments (VAZ-pool) was higher than in wild-type plants. In *psag-1.4*, the total chlorophyll content (Chl *a+b*) was decreased by about 25%. Interestingly, the drop in total chlorophyll level was not detected in *psag-1.4/psah2-1.4*.

Table 5.2. Pigment composition of wild-type and mutant plants. The pigment content of wild-type (WT, n=5) and mutant plants (n=5 each) was determined by HPLC. The carotenoid content is given in mmol per mol Chl (*a+b*), the Chl content is expressed as nmol Chl (*a+b*) per g fresh weight. Mean values (\pm SD) are shown. Nx = neoxanthin, VAZ = sum of xanthophylls cycle pigments (violaxanthin + antheraxanthin + zeaxanthin), Lu = lutein, β -Car = β -carotene.

	WT	<i>psag-1.4</i>	<i>psah2-1.4</i>	<i>psak-1</i>	<i>psag-1.4/psah2-1.4</i>	<i>psag-1.4/psak-1</i>
Nx	41 \pm 3	43 \pm 2	43 \pm 2	40 \pm 2	42 \pm 3	40 \pm 0
VAZ	42 \pm 2	49 \pm 2	43 \pm 1	40 \pm 0	46 \pm 1	49 \pm 1
Lu	144 \pm 5	155 \pm 2	137 \pm 6	145 \pm 5	139 \pm 7	141 \pm 7
b-Car	61 \pm 2	56 \pm 1	58 \pm 1	61 \pm 0	56 \pm 1	57 \pm 1
Chl <i>a+b</i>	1169 \pm 18	891 \pm 34	1139 \pm 102	1220 \pm 49	1278 \pm 131	1110 \pm 21
Chl <i>a/b</i>	3.18 \pm 0.02	3.11 \pm 0.04	3.21 \pm 0.05	3.61 \pm 0.05	3.19 \pm 0.03	3.61 \pm 0.02

5.8 Alteration of the redox states of PSI and PSII in *psag-1.4*, *psah2-1.4* and *psag-1.4/psah2-1.4* mutants

The parameters of chlorophyll fluorescence induction and the P700 absorbance were determined to characterise the electron flow in the mutants analysed (Table 5.3).

Most of the parameters studied did not differ significantly; in the light conditions used for the measurements, the *psag-1.4* mutant showed a minor decrease in the maximal fluorescence yield of PSII (F_v/F_m), while the other mutant lines resulted unaffected (Table 5.3).

In order to confirm this observation, Φ_{II} was measured in mutant and wild-type plants at different photosynthetically active flux densities. The *psag-1.4* mutant showed a slightly increased photoinhibition with respect to the WT when grown in normal greenhouse conditions (data not shown); this difference became more pronounced for plants grown at 400 μ E. Both

Table 5.3. Spectroscopic analyses of WT (n=5) and mutant plants (n=5 each). Mean values (\pm SD) are shown.

	WT	<i>psag-1.4</i>	<i>psah2-1.4</i>	<i>psag-1.4</i> <i>/psah2-1</i>
F_V/F_M	0.83 ± 0.005	0.81 ± 0.005	0.82 ± 0.009	0.82 ± 0.001
Φ_{II}	0.77 ± 0.008	0.77 ± 0.006	0.76 ± 0.016	0.76 ± 0.009
1-qP	0.024 ± 0.014	0.029 ± 0.005	0.027 ± 0.014	0.029 ± 0.011
$t_{1/2red}$	54 ± 1 msec	59 ± 1 msec	59 ± 1 msec	60 ± 1 msec
$t_{1/2ox}$	0.39 ± 0.01 sec	0.46 ± 0.02 sec	0.50 ± 0.02 sec	0.51 ± 0.02 sec

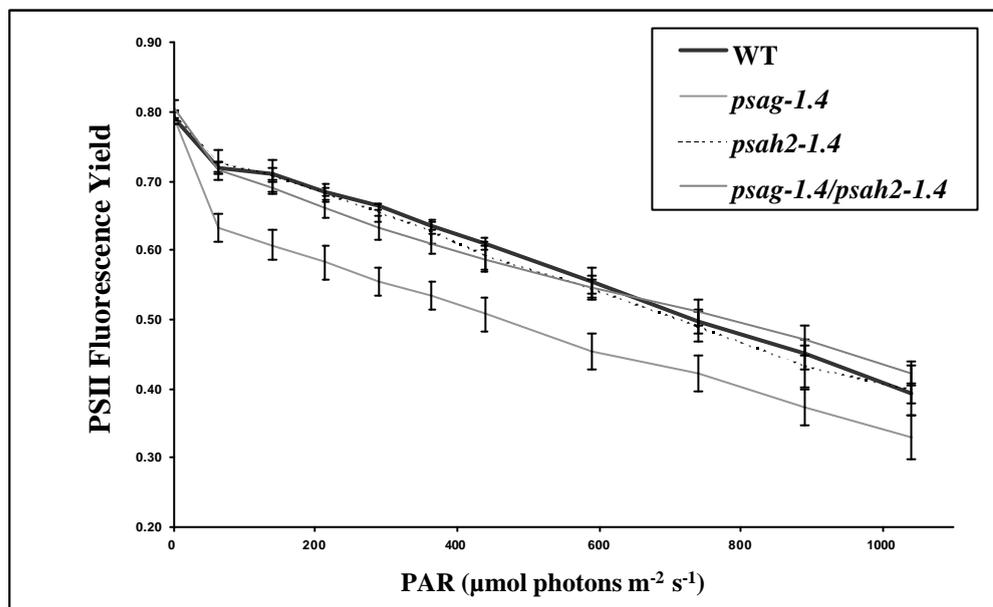


Figure 5.7. Light saturation of the *psag-1.4*, *psah2-1.4* and *psag-1.4/psah2-1.4* mutants and WT plants (n = 4) grown at a light intensity of $400 \mu\text{mol m}^{-2} \text{s}^{-1}$. Φ_{II} was determined for a range of photosynthetically active radiation (PAR) from 1 - $1039 \mu\text{mol m}^{-2} \text{s}^{-1}$ for 5 plants of *psag-1.4*, *psah2-1.4*, *psag-1.4/psah2-1.4* and WT genotypes, after incubation for 20 h in the dark followed by 4 h illumination with $160 \mu\text{mol photons m}^{-2} \text{s}^{-1}$ and a 10-min incubation in the dark prior to measurement. Error bars indicate standard deviations. Prior to Φ_{II} measurement, plants were adapted to every PAR condition for 15 min.

psah2-1.4 and *psag-1.4/psah2-1.4* mutants did not differ significantly from the WT in both growth conditions examined (Figure 5.7).

Under steady state conditions, *psag-1.4* mutants showed a decrease in reduced P700 (plateau in Figure 5.8). This decrease was in good agreement with the 20% reduction of several PSI subunits showed by immunoblot analysis. *psah2-1.4* plants did not show any significant

difference in reduced P700 amount in comparison to the WT, while the double mutant *psag-1.4/psah2-1.4* displayed a plateau level between that of *psag-1.4* and WT. All mutant genotypes showed a slightly altered kinetic of P700⁺ reduction and a marked increase of $t_{1/2ox}$ (in particular the *psah2-1.4* and *psag-1.4/psah2-1.4* mutants, Table 5.3).

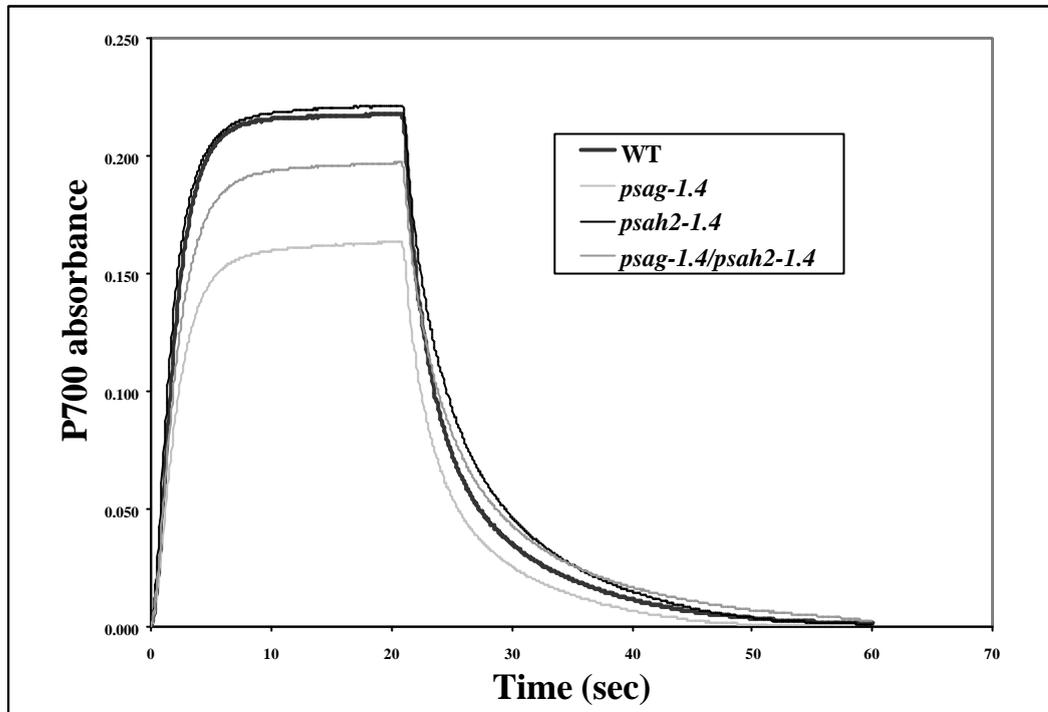


Figure 5.8. Absorption of P700 in whole leaves of *Arabidopsis* WT, *psag-1.4*, *psah2-1.4* and *psag-1.4/psah2-1.4* (n = 4).

DISCUSSION

No knockout for the PsaG subunit of PSI had been described yet, whereas Scheller and co-workers used a cosuppression approach to generate Arabidopsis plants lacking the H (H1 and H2) subunit. Taking advantage of stable and heritable knockouts of the *psaG* and *psaH2* genes, we tried to address the questions concerning the function of the plant-specific G subunit and the relative contribution of the *psaH2* gene to the levels of the H subunit. Due to the very high degree of homology between PsaH1 and PsaH2, this genetic approach represents the only way to determine the contribution of the two polypeptides in the PSI. According to the results presented, we can conclude that the PsaG subunit is necessary to the normal stability of PSI, as indicated by: (i) the altered PSI polypeptide composition in the *psag-1.4* mutant, with most subunits reduced in amount (ii) the decreased amount of active PSI measured spectroscopically.

Another interesting effect produced by the *psaG* knockout is the marked reduction in amount of Lhca2, one of the four polypeptides constituting the LHCI. In previous studies it has been shown that the PsaK subunit stabilises the association of Lhca2 and 3 to PSI, as demonstrated by the fact that PsaK co-suppression plants exhibited a decrease in both Lhca2 and 3 (Jensen et al., 2000); according to cross-linking studies (Jansson et al., 1996), the association between PsaK and Lhca3 had been suggested. Due to its homology to PsaK (Okkels et al., 1992), PsaG had been proposed to be involved in the binding of Lhca2 polypeptides. Our results indicate an effect of PsaG on the amount of Lhca2, suggesting a direct interaction of the two polypeptides. Recently, the interdependency of Lhca2 and 3 has been shown (Ganeteg et al., 2001): antisense plants missing Lhca2 showed a nearly complete lack of Lhca3, while in antisense plants lacking Lhca3 a 70% decrease of Lhca2 was shown, indicating the possibility of heterodimeric association of Lhca2/3. The decrease of Lhca2 caused by the PsaG mutation doesn't significantly affect the amount of Lhca3, thus confirming the homodimeric nature of these polypeptides *in vivo*.

Finally, the alterations in thylakoid pigments composition observed in *psag-1.4* plants seem to reflect two aspects of this mutation: (i) the increase in the xanthophyll cycle pigments (VAZ) could be related to the altered antenna composition in the *psag-1.4* background. (ii) the decrease in total chlorophyll content could, in part, be attributable to a delay in plant development, and therefore reduced chloroplast content in this genotype.

The *psah2* mutant allowed the direct estimation of the relative contribution of the two copies of PsaH present in *Arabidopsis*: from a quantitative point of view, the PsaH1 and PsaH2 polypeptides seem to contribute respectively for the 70% and 30% of the total PsaH subunit amount. Cosuppressed PsaH lines were shown to compensate for the lack of PsaH by increasing the PSI production. Immunoblot analysis of the other PSI subunits and spectroscopic measurements of active P700 exclude this possibility in our line. Particularly interesting features associated to the *psah2* mutation are also the unexpected higher growth rate, the basis of which has still to be investigated. The strong decrease in PsaL content and the increased half time of P700 oxidation could be the starting point to this aim. The understanding of the causes of these phenotypes could contribute to modify the fine-tuning of regulatory mechanisms of photosynthesis, improving the adaptation of plants to different environments.

6. THE GST-PRIME SOFTWARE PACKAGE

Bioinformatics is a rapidly expanding field of biology; due to the huge amount of data generated by the genome projects of several organisms, new tools to handle large sets of nucleic acid and polypeptide sequences are needed. The program described in the following has been conceived to help in the design of large number of primers useful for the creation of genomic DNA microarrays.

RESULTS

GST-PRIME has been conceived to provide a handy tool for the design of large sets of primer pairs to be employed in microarray realisation. For this purpose the program allows to design the primer pairs starting directly from a list of protein accession numbers, without any need to manually download and/or edit the template sequences. The program performs:

- retrieval of DNA sequences corresponding to the selected protein sequences (in combination with the NCBI databases *Batch Entrez* and *GenBank*);
- assembly of DNA sequences into gene sequences with and without introns;
- design of primers that are complementary to assembled gene sequences and suitable for use in PCR and RT-PCR applications (i.e. from both genomic or cDNA).

6.1 Automatic sequence retrieval and assembly

The starting point for *GST-PRIME* is a list of protein (GI) accession numbers. A text file (“annotated list”) containing protein sequences and cross-references to the DNA sequences necessary for extracting the corresponding coding regions is obtained by means of the *Batch Entrez* sequence retrieval system at NCBI (Figure 6.1A: ‘step 1’). From the “annotated list” the accession numbers of the genomic DNA sequences of interest, and the positions and boundaries of the embedded coding sequences are extracted by *GST-PRIME* and saved as a file (“exons list”; Figure 6.1A: ‘step 2’). Nucleotide sequences are then retrieved from the *GenBank* sequence database and the “exons list” is used for their assembly into “gene sequences” (including introns and exons) and “coding regions” (containing only exons). The sequence

download can be delayed (“delayed start option”), allowing it to run during the night or on weekends to avoid overloading the *GenBank* database. Both “gene sequences” and “coding regions” are reformatted into FASTA format (“FASTA files”; Figure 6.1A: ‘step 3’). The orientation of all DNA sequences can be standardised, allowing the conversion of all sequences into either forward (5’ to 3’) or reverse orientation (3’ to 5’).

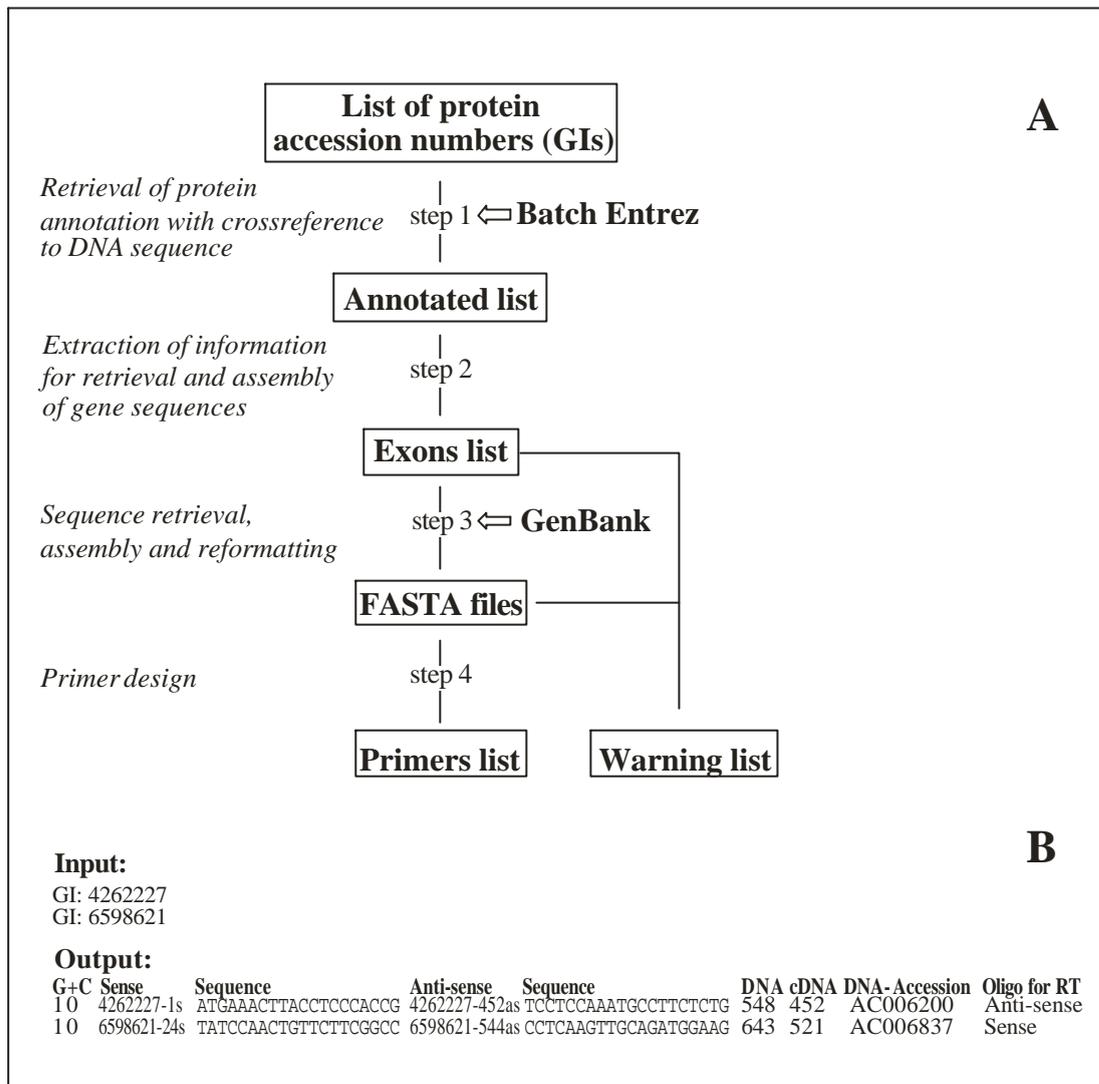


Figure 6.1. Flow diagram for *GST-PRIME* subroutines

(A) The left column indicates the tasks performed during the design of a large set of primer pairs starting from a list of protein accession numbers. The right column explains the files and databases employed.

(B) An example for input and output files of *GST-PRIME*. The input file is a list of protein accession numbers (GIs). In the output file, the following information is given (from left to right): G+C content (‘G+C’); designation and sequences of forward and reverse primers (‘sequence’, ‘sense’, and ‘antisense’, respectively); predicted lengths of amplicons at the genomic DNA (‘DNA’) and cDNA (‘cDNA’) levels; the accession number of the DNA from which the gene sequence was retrieved (‘DNA-Accession’); and the primer to be used for reverse transcription experiments (‘oligo for RT’).

6.2 Automatic primer design

GST-PRIME designs primers suitable for the amplification of both DNA and cDNA. The following default constraints were incorporated into the program:

- annealing sites are located exclusively in exons to allow reverse transcription experiments (such as cDNA first-strand synthesis and RT-PCR);
- standard primer length is 20 nucleotides with a G+C content of 50 %.

GST-PRIME employs the “exons list” and “FASTA files” to design the primer pairs (Figure 6.1A: ‘step 4’). DNA sequences not suitable for primer design (<120 bp) or without start or stop codons are identified and listed in a “warning file”. To obtain amplification products with a preferential size of ~500 bp (unless different specification from the user), forward primers are designed to anneal within the first 180 bp and reverse primers between position 480 and 720. For the forward primer the search begins at position 1 and stops after the first suitable has been identified. For the reverse primer, all suitable primers in the window from position 480 to 720 are identified and, if more than one reverse primer is found, the primer allowing amplification of a GST sized closest to 500 bp is selected. For sequences with lengths between < 720 bp, forward (reverse) primers are designed within the first (last) 50% of the sequence. Forward and reverse primers localised on putative exon/intron borders are rejected. If no primer can be identified based on the criteria listed above, the search is repeated for 21-mer primers with 9 G+C residues or 22-mers with 8 G+C. Independently from its G+C content, the reverse primer will be designed by default in a position allowing, in combination with the forward primer, the amplification of a product of preferentially 500 bp at cDNA level.

For both primers the annealing sites in genomic DNA are determined by using the ‘gene sequence’ files to predict the expected genomic amplicon length. *GST-PRIME* primers are designated according to the protein GI accession number, with a suffix indicating the annealing site relative to the gene sequence. The final output file (“primer list”) contains the primer sequences and the calculated lengths of amplicons at both DNA and cDNA level (Figure 6.1B). The primer-design subroutine of *GST-PRIME* can be run independently from the large-scale sequence retrieval and editing step.

6.3 Primer design for *Arabidopsis* and *Drosophila* genes

Two lists of 2000 randomly selected protein accession numbers each, corresponding to *Arabidopsis* and *Drosophila* genes, were generated and used as input files for *GST-PRIME*.

DNA sequence retrieval was performed using the ‘delayed start option’. For the 2000 *Arabidopsis* and for 1997 *Drosophila* proteins corresponding DNA sequences were retrieved (Table 6.1) and fed into the primer design subroutine of *GST-PRIME*. 1868 primer pairs (94.4%) were designed for *Arabidopsis*, and 1756 (87.8%) for *Drosophila*. Most of the primers discarded were rejected due to inability to meet the length constraints for predicted amplification products. In the case of *Arabidopsis*, 3.4% of predicted products were too short (<150bp at cDNA level), in *Drosophila* 5.7% were too long (>2050 bp at genomic DNA level) (Table 6.1).

Table 6.1. Comparison of the efficiency of primer design for *Arabidopsis* and *Drosophila*

	<i>Arabidopsis</i>	<i>Drosophila</i>
Number of protein accessions	2000 (100%)	2000 (100%)
Number of DNA sequences downloaded	2000 (100%)	1997 (99.9%)
Failure of primer pair design	27 (1.4%)	81 (4.1%)
Predicted amplicon <150 bp cDNA	67 (3.4%)	47 (2.4%)
Predicted amplicon >2050bp DNA	38 (1.9%)	113 (5.7%)
Primer pairs suitable for GST generation	1868 (93.4%)	1756 (87.8%)

Frequency distributions of the predicted cDNA amplicons were similar for the two organisms: more than 70% of amplifications were predicted to amplify cDNA fragments between 451 and 550 bp in length, with the average predicted length for amplified cDNAs being equal in both species (473 bp; Figure 6.2). Distributions of predicted genomic amplicon lengths differed significantly in the two species. The most prominent amplicon class for genomic DNA had again a predicted length between 451 and 550 bp. However, the average predicted genomic amplicon length in *Drosophila* was significantly smaller than in *Arabidopsis* (632 bp versus 739 bp).

6.4 *GST-PRIME* primer testing by PCR amplification of 1900 GSTs

The identification of 2047 GenBank entries for *Arabidopsis* proteins related to chloroplast functions has been described previously (Abdallah et al., 2000). For a total of 149 sequences

(7.5%) it has not been possible to design primer pairs satisfying all the default constraints. The remaining 1898 primer pairs were synthesised as 35-mers, consisting of a gene-specific 20-mer

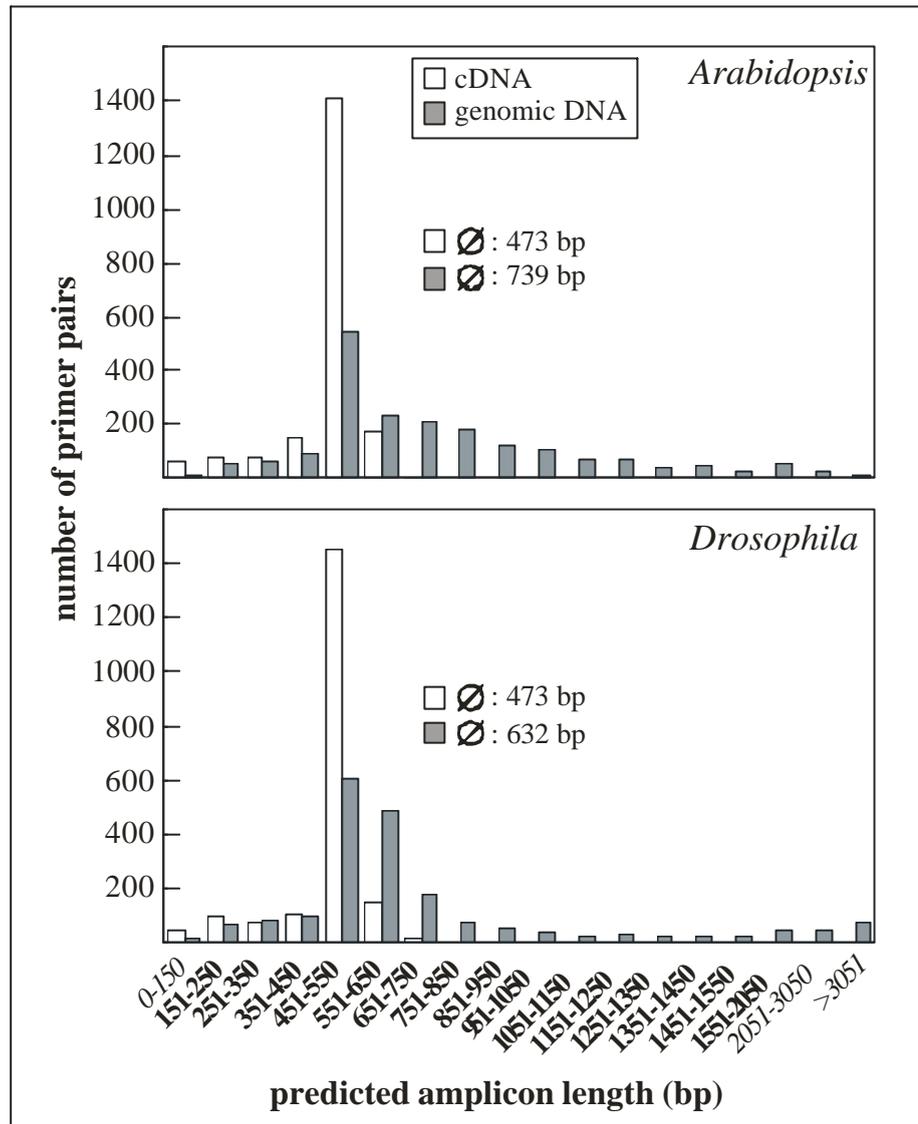


Figure 6.2. Distribution of predicted lengths of amplicons for *Arabidopsis* and *Drosophila*. Lengths suitable for PCR-based amplification of GSTs are indicated in bold, whereas lengths not considered for GST generation are indicated in italics.

sequence provided by *GST-PRIME* and a universal 5' located 15-mer tail sequence suitable for re-amplification with universal tail primers. PCRs with 1804 (95%) of the primer pairs resulted in the amplification from genomic DNA of products that could be identified by agarose gel electrophoresis. The approximate size of 200 PCR products was determined in gels and correlated with the size predicted by *GST-PRIME*: for 190 PCR products the predicted and actual sizes were identical, while 10 PCR products (5%) were larger than predicted. Re-evaluation of the corresponding protein and DNA sequences revealed that the latter were exclusively derived from cDNA sequencing projects and therefore lacked any cross-reference to

the corresponding genomic sequence. Sequence data for these cDNAs compared to those for the PCR products demonstrated that size differences were due to the presence of introns. Sequencing of 50 randomly chosen PCR products confirmed that target gene and PCR product were identical.

The 94 primer pairs failing to give PCR amplifications were compared to a set of 94 randomly selected amplifying primer pairs for primer-primer interaction or self-complementarity using the Mfold and Bestfit programs of the GCG package (Devereux et al., 1984). No significant difference between the two groups was predicted for both the minimal energy of self-annealing for each primer pair (Figure 6.3) and the normalised quality parameter for primer-dimers formation (Figure 6.4).

To rule out the possibility that the unsuccessful PCR amplification could derive from the combination of both an increased tendency to self-annealing with a higher probability of primer-dimers formation, the two parameters were combined to derive a “overall quality” parameter for the single primer pairs (Figure 6.5). Also in this case, the distributions resulting for the working and non-working primers were not significantly different.

2.14 Operating environments

GST-PRIME should be able to run under any windows environment and has been successfully tested under Windows NT 4.0 (service pack 6), windows 2000 (service pack 2) and windows 98 SE.

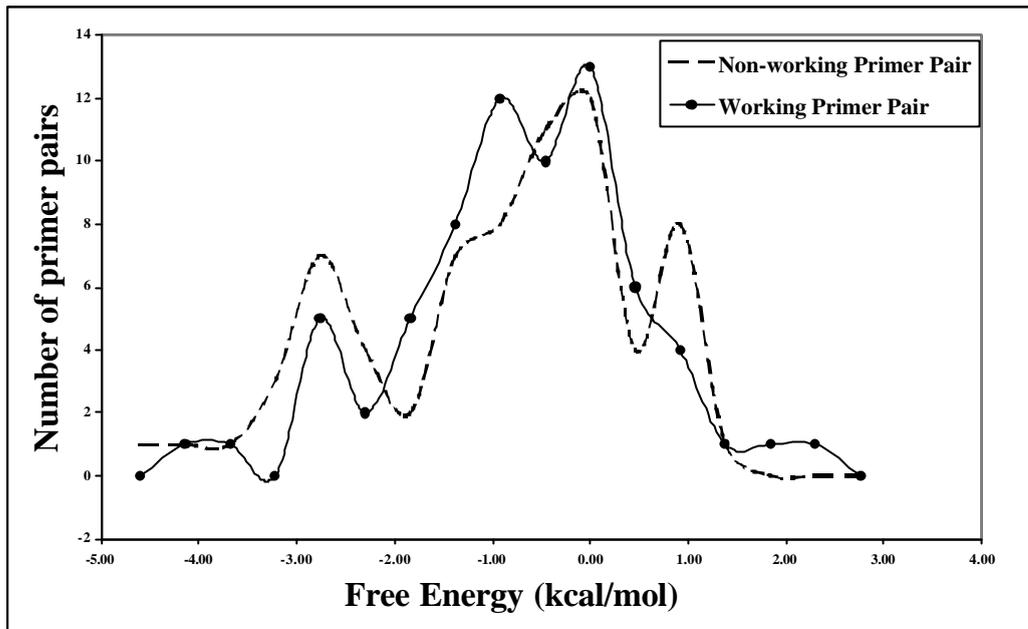


Figure 6.3. Distribution of minimal energy of self annealing within pair of primers. The energy of self-annealing for the primers of each pair was calculated by means of the program Mfold of the GCG package. The value represents the free energy change associated to the most stable secondary structure of each primer molecule. The lower this value, the more stable the secondary structure is. The minimum between the two energies of each primer pair has been calculated and the resulting distribution for amplifying (working primer pair) and not amplifying (non-working primer pair) primer pairs is shown (to prevent amplification it's sufficient that only one of the primers is not working, so the lowest value of the pair, corresponding to the worse primer, has been chosen). Note that the distribution is not gaussian, due to the fact that different genes have different codon compositions.

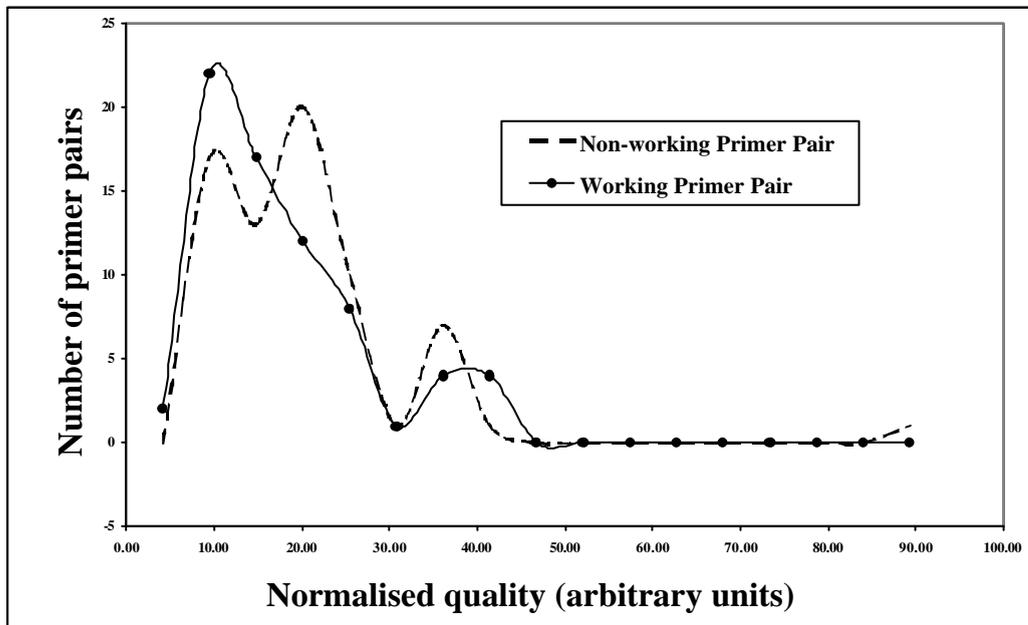


Figure 6.4. Distribution of normalized quality of primer-dimers formation. The value is the product of the “quality” parameter calculated by GCG, program Bestfit (see material and methods), times the fraction of the longest calculated annealed region with respect to complete annealing. The higher the value, the higher the tendency of a primer pair to form dimers.

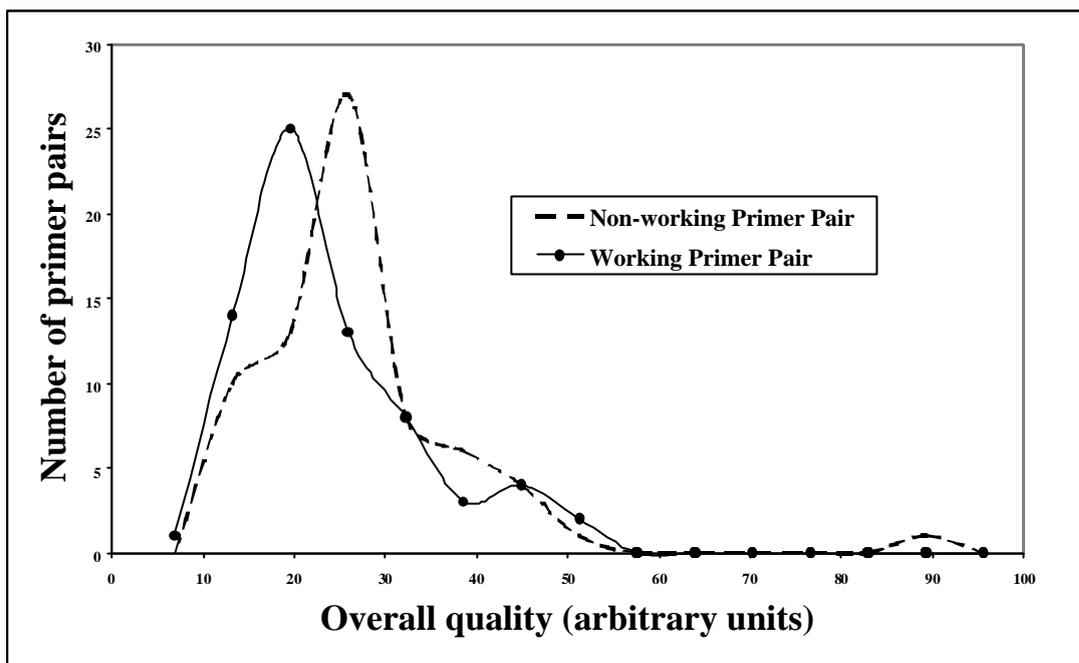


Figure 6.5. Distribution of overall quality per primer pair.

To derive a parameter taking into account of both self-annealing and dimers formation, the minimal energies of self-annealing were shifted to positive values subtracting their values from the arbitrary value of 5 (see Fig1). In this way all the calculated energies become positive. The value of the “overall quality” is given by the sum of the “shifted” minimal self annealing energy associated to each primer pair and the pair annealing energy. Also in this case, the higher the value, the worse the primer pair.

DISCUSSION

A pre-requisite for the large-scale generation of GSTs from DNA or cDNA is the establishment of a standard procedure based on

- (i) selection of a set of genes/proteins of interest;
- (ii) retrieval of the corresponding gene sequences, including genomic sequence and (predicted) transcribed regions; and
- (iii) design of primers that can be used for the amplification of genes from DNA or cDNA

Several programs are currently available to design primers for large-scale sequencing projects (Haas et al., 1998; Proutski and Holmes, 1996; Li et al., 1997) or primer pairs for DNA microarray construction (Raddatz et al., 2001), but they cannot be used for automated and genome-wide primer design starting from protein sequence collections. Moreover, none of these programs allows the design of primer pairs suitable for GST amplification from intron-containing genomic DNA. Both these limitations are overcome by the *GST-PRIME* program by means of an automated sequence retrieval and assembly function and the avoidance of splicing sites during primer design, thus allowing the amplification of GSTs either from DNA or cDNA.

The systematic classification of genes according to the biological function of their products is based on the analysis of protein sequences rather than DNA sequences. For this reason we have established a routine allowing the automatic retrieval of DNA sequences that encode the specified protein sequences. The use of GSTs instead of ESTs arrays has the advantage of allowing maximal coverage for the genomes already sequenced: the amplification from genomic DNA solves the problem of obtaining ESTs for genes expressed at a low level or in tissues from which it is difficult to obtain RNA. In this way, virtually all the genes of an organism whose genome sequencing has been completed can be represented without redundancy on a microarray. Moreover, the ease and low cost of genomic versus cDNA isolation render even more attractive the use of GSTs arrays. On the other hand, the design of GSTs to be spotted on arrays raises the problem of including in each GST a fraction of coding sequence sufficient to hybridise to the labelled cDNA. For this reason, the more stringent constraint during the design of a GST is the minimal amplification length at cDNA level,

making it suitable for transcriptomics experiments. This requirement, on its turn, generates the need to check the amplification length from genomic DNA: if the predicted amplification length exceed the range of amplification of a normal Taq polymerase, it will be necessary to use a long-range PCR or perform the amplification directly from cDNA. Taking into account these constraints, *GST-PRIME* has been projected to design primers that can be employed for amplification from either cDNA or genomic DNA. The download and sequence assembly subroutine of the *GST-PRIME* program has been tested on two model organisms for which complete genome sequences are available. Both for *Arabidopsis* and *Drosophila*, sequence retrieval and assembly were performed successfully. The sequence retrieval step of *GST-PRIME* can be carried out overnight using the 'delay start' function of the program, which relieves the pressure on public databases and improves the performance of the download process. The automatic assembly of the genomic sequences solves the problem of extracting and editing the parts of the downloaded sequences to be used for primer design.

The constraints employed for the primer-design subroutine led in 87.8% and 93.4% of cases, respectively, for *Drosophila* and *Arabidopsis*, to primer pairs providing predicted GSTs with coding regions longer than 150 bp and a maximal genomic length of 2050 bp. Surprisingly, predicted genomic amplicons were significantly smaller in *Drosophila* compared to *Arabidopsis*, although fruit fly genes are in general larger than *Arabidopsis* genes (Adams et al., 2000; The Arabidopsis Genome Initiative, 2000). This discrepancy can be attributed to the high frequency of introns of 59 to 63 bp in *Drosophila* genes (Adams et al., 2000); the finding of a prominent class of genomic amplicons with a size between 551-650 bp (Figure 6.2) may reflect the presence on the average of one or two of such relatively short introns in the first 500 bp of the coding region.

The suitability of *GST-PRIME* primers for generating amplicons from genomic DNA was demonstrated in *Arabidopsis*. A total of 1900 primer pairs were tested, and 95% of reactions resulted in PCR products. That these were derived from the appropriate genomic sequences was verified by size and sequence analysis. The sizes of a few PCR products were not correctly predicted by *GST-PRIME*. However, this could be attributed to sequence annotations that lacked cross-references to the genomic DNA including the introns. The 5% of the primer pairs not leading to PCR amplification were demonstrated not to significantly differ in comparison to functioning primer pairs for their tendency to self-anneal or to form primer-dimers. The failure could therefore due to regions difficult to amplify within the intervening region between the 2

primers of a pair (e.g. GC-rich regions, loops and hairpins, etc.) or to incorrect annotation of the genomic DNA resulting in design of primers annealing on splicing sites.

Extension of the *GST-PRIME* program to other species depends on the state of genome research in each organism. For yeast, generation of amplicons of all 6000 genes has been accomplished successfully (DeRisi et al., 1997). This was made easier by the low frequency of intron-containing genes in this species, which allowed the use of a relatively unsophisticated primer-design software. Higher organisms, like *H. sapiens*, *D. melanogaster* or *A. thaliana* represent a more challenging task for systematic primer design. The different G + C content, average introns number and length, abundance and size of gene families present in different organisms will, by consequence, require several improvements to the actual program implementation. Future upgrades of *GST-PRIME* should therefore calculate the G+C content and the presence of secondary structure within the GSTs, in order to help the optimisation of the PCR conditions for the single amplifications. A second major improvement will be the possibility for the program to automatically select for the GST design those regions with the lowest degree of homology to other genes, in order to reduce cross-hybridisation between different members of multigenic families.

SUMMARY / ZUSAMMENFASSUNG

SUMMARY

Photosynthesis is a characteristic trait that distinguishes algae and plants from other eukaryotic organisms. This thesis focussed on the genetic and molecular analysis of photosystem I (PSI) functions in *Arabidopsis*. For this purpose, a novel screening procedure for the identification of mutants affected in photosynthesis based on the automated measurement of the effective quantum yield of PSII (F_{II}) was applied. During the screening of 2000 *En*- or T-DNA-tagged lines, a total of 12 mutants was found and for four of them the mutated genes were identified. One of the mutants, a line with a knockout in the *psaE1* gene, was characterised in detail. The analyses performed showed that a decrease in PsaE level leads to a reduced stability of the stromal ridge of PSI, affecting concomitantly the electron transfer to ferredoxin. Furthermore, stable mutant alleles for the genes *psaE2*, *psaG*, *psaH2*, *psaK*, *psaL* and *psaN* coding for the PSI subunits PsaE2, G, H2, K, L and N were obtained by a reverse genetics approach. It could be shown that the PsaG subunit stabilises the PSI core and interacts with the light harvesting protein Lhca2. Furthermore, the contribution of *psaH2* to the total expression of PsaH was analysed.

In order to lay the foundation for the genome-wide isolation of gene sequence tags (GSTs) relevant for photosynthesis, the computer program package *GST-PRIME* was developed. This program allows automated sequence retrieval, followed by the automated design of primer pairs which can be used for PCR amplification from genomic DNA or cDNA. *GST-PRIME* primer pairs were tested on a set of 1900 *Arabidopsis* genes coding for chloroplast-targeted proteins and 95% of the primer pairs resulted in correct amplicons.

ZUSAMMENFASSUNG

Photosynthese ist ein Prozess, der bei Eukaryoten charakteristisch für Algen und Pflanzen ist. Im Rahmen dieser Doktorarbeit wurde eine genetische und molekularbiologische Analyse des Photosystems I (PSI) durchgeführt. Dafür wurde ein Durchmusterungsverfahren für die Identifizierung von Mutanten mit einem Defekt in der Photosynthese entwickelt, der auf einer automatisierten Messung der effektiven Quantenausbeute von Photosystem II (F_{II}) basiert. Während der Durchmusterung von 2000 *En*- oder T-DNA Linien wurden 12 Mutanten identifiziert. Für vier dieser Mutanten wurden die mutierten Gene isoliert. Eine dieser Mutanten wies einen knock-out im *psaE1* Gen auf und wurde im Detail charakterisiert. Diese Charakterisierung zeigte, daß eine Reduktion der PsaE1 Expression zu einer Destabilisierung der Stromaseite von PSI führt und gleichzeitig den Elektronentransfer zu Ferredoxin beeinflusst. Desweiteren wurden stabile Mutantenallele für die Gene *psaE2*, *psaG*, *psaH2*, *psaK*, *psaL* und *psaN*, die für die PSI-Untereinheiten PsaE2, G, H2, K, L and N kodieren, mit Hilfe von reverser Genetik isoliert. Es konnte gezeigt werden, daß PsaG den PSI-Komplex stabilisiert und mit dem Lichtsammlerprotein Lhca2 interagiert. Desweiteren wurde der Anteil von *psaH2* an der Gesamtexpression von PsaH untersucht.

Für die genomweite Isolierung von, für die Photosynthese relevanten, "Gene Sequence Tags" (GSTs) wurde das Computer Programm *GST-PRIME* entwickelt. Dieses Programm ermöglicht eine Kombination von automatisierter Sequenzabfrage und automatischen Design von Primerpaaren, mit welchen eine PCR-Amplifizierung von genomischer DNA oder cDNA erfolgen kann. Für 1900 Arabidopsis-Gene, die für Chloroplastenproteinen kodieren, wurden *GST-PRIME* Primerpaare entworfen und 95% der getesteten Primerpaare ergaben korrekte Amplifikationsprodukte.

REFERENCES

A

- Abdallah F., Salamini F., Leister D., A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*, *Trends Plant Sci.* 5 (2000) 141-142.
- Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F. et al., The genome sequence of *Drosophila melanogaster*. *Science* 287 (2000) 2185-2195.
- Allawi H.T., SantaLucia J.Jr., Thermodynamics and NMR of internal G.T mismatches in DNA, *Biochemistry* 36 (1997) 10581-94.
- Amin P., Sy D.A., Pilgrim M.L., Parry D.H., Nussaume L., Hoffman N.E., *Arabidopsis* mutants lacking the 43- and 54-kilodalton subunits of the chloroplast signal recognition particle have distinct phenotypes, *Plant Physiol.* 121 (1999) 61-70.

B

- Bassi R., dal Belin Peruffo A., Barbato R. and Ghisi R., Differences in chlorophyll-protein complexes and composition of polypeptides between thylakoids from bundle sheaths and mesophyll cells in maize, *Eur. J. Biochem.* 146 (1985) 589-595.
- Bennoun P., Delepelaire P., Isolation of photosynthesis mutants in *Chlamydomonas*, in: Edelman M., Hallick R.B., Chua N.H. (Eds.), *Methods in Chloroplast Molecular Biology*, Elsevier Biomedical Press, Amsterdam (1982) pp. 25-38.
- Bennoun P., Levine R.P., Detecting mutants that have impaired photosynthesis by their increased level of fluorescence, *Plant Physiol.* 42 (1967) 1284-1287.
- Blattner F.R., Plunkett G.3rd, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F. et al., The complete genome sequence of *Escherichia coli* K-12, *Science* 277 (1997) 1453-1474.
- Bradbury M., Baker N.R., Analysis of the slow phases of the in vivo chlorophyll fluorescence induction curve. Changes in the redox state of photosystem II electron acceptors and fluorescence emission from photosystems I and II, *Biochim. Biophys. Acta* 635 (1981) 542-551.
- Brown P.O., Botstein D., Exploring the new world of the genome with DNA microarrays, *Nat. Genet.* 21 (1999) 33-37.

C

- Campisi L., Yang Y., Yi Y., Heilig E., Herman B., Cassista A.J., Allen D.W., Xiang H., Jack T., Generation of enhancer trap lines in *Arabidopsis* and characterization of expression patterns in the inflorescence, *Plant J.* 17 (1999) 699-707.
- Carol P., Stevenson D., Bisanz C., Breitenbach J., Sandmann G., Mache R., Coupland G., Kuntz M., Mutations in the *Arabidopsis* gene *IMMUTANS* cause a variegated phenotype by inactivating a chloroplast terminal oxidase associated with phytoene desaturation, *Plant Cell* 11 (1999) 57-68.
- Chen M., Choi Y., Voytas D.F., Rodermel S., Mutations in the *Arabidopsis* *VAR2* locus cause leaf variegation due to the loss of a chloroplast FtsH protease, *Plant J.* 22 (2000) 303-313.
- Chitnis V.P. and Chitnis P.R., PsaL subunit is required for the formation of photosystem I trimers in the cyanobacterium *Synechocystis* sp. PCC 6803, *FEBS Lett.* 336 (1993) 330-4.
- Chitnis P.R., Reilly P.A., Miedel M.C. and Nelson N., Structure and targeted mutagenesis of the gene encoding 8kDa subunit of photosystem I from the cyanobacterium *Synechocystis* sp. PCC 6803, *J. Biol. Chem.* 264 (1989) 18374-18380.
- Clayton R.K., *Photosynthesis: Physical mechanisms and chemical patterns*. Cambridge University Press, Cambridge, England (1980).

D

- Davis S.J., Kurepa J., Vierstra R.D., The *Arabidopsis thaliana* *HY1* locus, required for phytochrome-chromophore biosynthesis, encodes a protein related to heme oxygenases, *Proc. Natl. Acad. Sci. USA* 96 (1999) 6541-6546.
- Desprez T., Amselem J., Caboche M., Hofte H., Differential gene expression in *Arabidopsis* monitored using cDNA arrays, *Plant J.* 14 (1998) 643-652.
- Devereux, J., Haeberli, P., Smithies, O., A comprehensive set of sequence analysis programs for the VAX, *Nucleic Acids Res.* 12 (1984) 387-395.
- DeRisi J.L., Iyer V.R. and Brown P.O., Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278 (1997) 680-686.
- Dinkins R.D., Bandaranayake H., Green B.R., Griffiths A.J., A nuclear photosynthetic electron transport mutant of *Arabidopsis thaliana* with altered expression of the chloroplast *petA* gene, *Curr. Genet.* 25 (1994) 282-288.
- Douglas S.E., Plastid evolution: origins, diversity, trends, *Curr. Opin. Genet. Dev.* 8 (1998) 655-661.

-
- Duggan D.J., Bittner M., Chen Y., Meltzer P., Trent J.M., Expression profiling using cDNA microarrays, *Nat. Genet.* 21 (1999) 10-14.

E

- Eisen M.B., Brown P.O., DNA arrays for analysis of gene expression, *Methods Enzymol.* 303 (1999) 179-205.
- Emanuelsson O., Nielsen H., Brunak S., von Heijne G., Predicting subcellular localization of proteins based on their N-terminal amino acid sequence, *J. Mol. Biol.* 300 (2000) 1005-1016.
- Emanuelsson O., Nielsen H., von Heijne G., ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites, *Prot. Sci.* 8 (1999) 978-984.

F

- Farah J., Rappaport F., Choquet Y., Joliot P., Rochaix J.D., Isolation of a psaF-deficient mutant of *Chlamydomonas reinhardtii*: efficient interaction of plastocyanin with the photosystem I reaction center is mediated by the PsaF subunit, *EMBO J.* 14 (1995) 4976-84.
- Felder S., Meierhoff K., Sane A.P., Meurer J., Driemel C., Plucken H., Klaff P., Stein B., Bechtold N., Westhoff P., The nucleus-encoded HCF107 gene of *Arabidopsis* provides a link between intercistronic RNA processing and the accumulation of translation-competent psbH transcripts in chloroplasts. *Plant Cell* 13 (2001) 2127-41.
- Fisk D.G., Walker M.B., Barkan A, Molecular cloning of the maize gene *crp1* reveals similarity between regulators of mitochondrial and chloroplast gene expression, *Embo J.* 18 (1999) 2621-2630.
- Frey M., Stettner C., Gierl A., A general method for gene isolation in tagging approaches: Amplification of insertion mutagenised sites (AIMS), *Plant J.* 13 (1998) 717-721.

G

- Ganeteg U., Strand A., Gustafsson P., Jansson S., The Properties of the Chlorophyll a/b-Binding Proteins Lhca2 and Lhca3 Studied in Vivo Using Antisense Inhibition, *Plant Physiol.* 127 (2001) 150-8.
- Genty B., Briantais J.M., Baker N.R., The relationship between the quantum yield of photosynthetic electron-transport and quenching of chlorophyll fluorescence, *Biochim. Biophys. Acta* 990 (1989) 87-92.
- Goffeau A., Barrell B.G., Bussey H., Davis R.W., Dujon B., Feldmann H., Galibert F, Hoheisel J.D., Jacq C., Johnston M. et al., Life with 6000 genes, *Science* 274 (1996) 563-567.

H

- Haas S., Vingron M., Poustka A., Wiemann S., Primer design for large scale sequencing, *Nucleic Acids Res.* 26 (1998) 3006-12.
- Haldrup A., Naver H., Scheller H.V., The interaction between plastocyanin and photosystem I is inefficient in transgenic *Arabidopsis* plants lacking the PSI-N subunit of photosystem I, *Plant J.* 17 (1999) 689-698.
- Haldrup A., Simpson D.J., Scheller H.V., Down-regulation of the PSI-F subunit of photosystem I in *Arabidopsis thaliana*. The PSI-F subunit is essential for photoautotrophic growth and antenna function, *J. Biol. Chem.* 275 (2000), 31211-31218.
- Harbinson, J. and Woodward, F.I., The use of light-induced absorbency changes at 820 nm to monitor the oxidation-state of P-700 in leaves, *Plant Cell Environ.* 10 (1987) 131-140.
- He W.Z., Malkin R., Photosystems I and II, in: Raghavendra A.S. (Ed.), *Photosynthesis: a comprehensive treatise*, University Press, Cambridge, 1998, pp. 29-43.
- Hihara Y., Kamei A., Kanehisa M., Kaplan A., Ikeuchi M., DNA microarray analysis of cyanobacterial gene expression during acclimation to high light, *Plant Cell* 13 (2001) 793-806.
- Hippler M., Drepper F., Farah J., Rochaix J.D., Fast electron transfer from cytochrome c6 and plastocyanin to photosystem I of *Chlamydomonas reinhardtii* requires PsaF, *Biochemistry* 36 (1997) 6343-9.

I

J

- Jansson, S., Andersen, B. and Scheller, H. V., Nearest-neighbor analysis of higher-plant photosystem I holocomplex, *Plant Physiol.* 112 (1996) 409-420.
- Jensen P.E., Gilpin M., Knoetzel J., Scheller H.V., The PSI-K subunit of photosystem I is involved in the interaction between light-harvesting complex I and the photosystem I reaction center core, *J. Biol. Chem.* 275 (2000) 24701-24708.

K

- Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y., Miyajima N., Hirosawa M., Sugiura M., Sasamoto S. et al., Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.

-
- strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* 3 (1996) 109-136.
- Klimyuk V.I., Persello-Cartieaux F., Havaux M., Contard-David P., Schuenemann D., Meierhoff K., Gouet P., Jones J.D., Hoffman N.E., Nussaume L., A chromodomain protein encoded by the *Arabidopsis* CAO gene is a plant-specific component of the chloroplast signal recognition particle pathway that is involved in LHCP targeting, *Plant Cell* 11 (1999) 87-100.
 - Klughammer, C. and Schreiber, U., An improved method, using saturating light pulses, for the determination of photosystem I quantum yield via P700+ absorbance changes at 830 nm, *Planta* 192 (1994) 261-268.
 - Klukas O., Schubert W.D., Jordan P., Krauß N., Fromme P., Witt H.T. and Saenger W., Photosystem I, an improved model of the stromal subunits PsaC, PsaD, and PsaE, *J. Biol. Chem.* 274 (1999) 7351-7360.
 - Koncz C., Mayerhofer R., Koncz-Kalman Z., Nawrath C., Reiss B., Redei G.P., Schell J., Isolation of a gene encoding a novel chloroplast protein by T-DNA tagging in *Arabidopsis thaliana*, *Embo J.* 9 (1990) 1337-1346.
 - Korshunova Y.O., Eide D., Clark W.G., Guerinot M.L., Pakrasi H.B., The IRT1 protein from *Arabidopsis thaliana* is a metal transporter with a broad substrate range, *Plant. Mol. Biol.* 40 (1999) 37-44.
 - Krause G.H. and Weis E., Chlorophyll fluorescence and photosynthesis: the basics, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 42 (1991) 313-349.
 - Kruip J., Chitnis P.R., Lagoutte B., Rogner M. and Boekema E.J., Structural organization of the major subunits in cyanobacterial photosystem I. Localization of subunits PsaC, -D, -E, -F, and -J, *J. Biol. Chem.* 272 (1997) 17061-17069.
 - Kumar A. and Snyder M., Emerging technologies in yeast, *Nat. Rev. Genet.* 2 (2001) 302-12.
- L**
- Leister D., Varotto C., Pesaresi P., Niwergall A., Salamini F., Large-scale evaluation of plant growth in *Arabidopsis thaliana* by non-invasive image analysis, *Plant Physiol. Biochem.* 37 (1999) 671-678.
 - Lelong C., Setif P., Lagoutte B. and Bottin H., Identification of the amino acids involved in the functional interaction between photosystem I and ferredoxin from *Synechocystis* sp. PCC 6803 by chemical cross-linking, *J. Biol. Chem.* 269 (1994) 10034-10039.
 - Li P., Kupfer K.C., Davies C.J., Burbee D., Evans G.A., Garner H.R., PRIMO: A primer design program that applies base quality statistics for automated large-scale DNA sequencing, *Genomics* 40 (1997) 476-85
 - Li X.P., Bjorkman O., Shih C., Grossman A.R., Rosenquist M., Jansson S., Niyogi K.K., A pigment-binding protein essential for regulation of photosynthetic light harvesting, *Nature* 403 (2000) 391-395.
-

-
- Liu Y.G., Mitsukawa N., Oosumi T., Whittier R.F., Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR, *Plant J.* 8 (1995) 457-463.
 - Lunde C., Jensen P.E., Haldrup A., Knoetzel J., Scheller H.V., The PSI-H subunit of photosystem I is essential for state transitions in plant photosynthesis, *Nature* 408 (2000) 613-615.

M

- Machida C., Onouchi H., Koizumi J., Hamada S., Semiarti E., Torikai S., Machida Y., Characterization of the transposition pattern of the Ac element in *Arabidopsis thaliana* using endonuclease I-SceI, *Proc. Natl. Acad. Sci. USA* 94 (1997) 8675-8680.
- Mandel M.A., Feldmann K.A., Herrera-Estrella L., Rocha-Sosa M., Leon P., *CLA1*, a novel gene required for chloroplast development, is highly conserved in evolution, *Plant J.* 9 (1996) 649-658.
- Meimberg K., Lagoutte B., Bottin H. and Mühlhoff U., The PsaE subunit is required for complex formation between photosystem I and flavodoxin from the cyanobacterium *Synechocystis* sp. PCC 6803, *Biochemistry* 37 (1998) 9759-9767.
- Meurer J., Meierhoff K., Westhoff P., Isolation of high-chlorophyll-fluorescence mutants of *Arabidopsis thaliana* and their characterisation by spectroscopy, immunoblotting and northern hybridisation, *Planta* 198 (1996) 385-396.
- Meurer J., Plucken H., Kowallik K.V., Westhoff P., A nuclear-encoded protein of prokaryotic origin is essential for the stability of photosystem II in *Arabidopsis thaliana*, *Embo J.* 17 (1998) 5286-5297.
- Miles D., Mutants of higher plants: maize, *Method. Enzymol.* 69 (1980) 3-23.
- Moffatt B.A., McWhinnie E.A., Agarwal S.K. and Schaff D.A., The adenine phosphoribosyltransferase-encoding gene of *Arabidopsis thaliana*, *Gene* 143 (1994) 211-216.
- Muramoto T., Kohchi T., Yokota A., Hwang I., Goodman H.M., The *Arabidopsis* photomorphogenic mutant *hyl* is deficient in phytochrome chromophore biosynthesis as a result of a mutation in a plastid heme oxygenase, *Plant Cell* 11 (1999) 335-348.

N

- Naver H., Haldrup A., Scheller H.V., Cosuppression of photosystem I subunit PSI-H in *Arabidopsis thaliana*. Efficient electron transfer and stability of photosystem I is dependent upon the PSI-H subunit, *J. Biol. Chem.* 274 (1999) 10784-10789.
-

-
- Naver H., Scott M.P., Adersen B., Moller B.L. and Scheller H.V., Reconstitution of barley photosystem-I reveals that the N-terminus of the PSI-D subunit is essential for tight binding of PSI-C, *Physiol. Plant.* 95 (1995) 19-26.
 - Naver H., Scott M.P., Golbeck J.H., Olsen C.E., Scheller H.V., The eight-amino acid internal loop of PSI-C mediates association of low molecular mass iron-sulfur proteins with the P700-FX core in photosystem I, *J Biol Chem.* 273 (1998) 18778-83.
 - Nedbal L., Soukupova J., Kaftan D., Whitmarsh J., Trtilek M., Kinetic imaging of chlorophyll fluorescence using modulated light, *Photosynth. Res.* 66 (2000) 3-12.
 - Niyogi K.K., Bjorkman O., Grossman A.R., Chlamydomonas xanthophyll cycle mutants identified by video imaging of chlorophyll fluorescence quenching, *Plant Cell* 9 (1997) 1369-1380.
 - Niyogi K.K., Grossman A.R., Bjorkman O., *Arabidopsis* mutants define a central role for the xanthophyll cycle in the regulation of photosynthetic energy conversion, *Plant Cell* 10 (1998) 1121-1134.

O

- Obokata J., Mikami K., Hayashida N., Nakamura M., Sugiura M., Molecular heterogeneity of photosystem I. psaD, psaE, psaF, psaH, and psaL are all present in isoforms in *Nicotiana* spp, *Plant Physiol.* 102 (1993) 1259-1267.
- Obokata J., Mikami K., Yamamoto Y. and Hayashida N., Microheterogeneity of PSI-E subunit of photosystem I in *Nicotiana sylvestris*, *Plant Cell Physiol.* 35 (1994) 203-209.
- Ochman H., Gerber A.S., Hartl D.L., Genetic applications of an inverse polymerase chain reaction, *Genetics* 120 (1988) 621-3.
- Oh-oka H., Takahashi Y. and Matsubara H., Topological considerations of the 9KDa polypeptide which contains centers A and B, associated with the 14- and 19-KDa polypeptides in the photosystem I complex of spinach, *Plant Cell Physiol.* 30 (1989) 869-875.
- Okkels J.S., Nielsen V.S., Scheller H.V., Moller B.L., A cDNA clone from barley encoding the precursor from the photosystem I polypeptide PSI-G: sequence similarity to PSI-K, *Plant Mol Biol.* 18 (1992) 989-94.

P

- Parinov S., Sevugan M., De Y., Yang W.C., Kumaran M., Sundaresan V., Analysis of flanking sequences from *Dissociation* insertion lines. A database for reverse genetics in *Arabidopsis*, *Plant Cell* 11 (1999) 2263-2270.

-
- Parinov S., Sundaresan V., Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome sequencing project, *Curr. Opin. Biotechnol.* 11 (2000) 157-161.
 - Perucho M., Hanahan D., Lipsich L., Wigler M., Isolation of the chicken thymidine kinase gene by plasmid rescue, *Nature* 285 (1980) 207-10
 - Proutski V. and Holmes E.C., Primer Master: a new program for the design and analysis of PCR primers, *Comput. Appl. Biosci.* 12 (1996) 253-5.

Q

R

- Raddatz G., Dehio M., Meyer T.F., Dehio C., PrimeArray: genome-scale primer design for DNA-microarray construction, *Bioinformatics* 17 (2001) 98-9.
- Reiter R.S., Coomber S.A., Bourett T.M., Bartley G.E., Scolnik P.A., Control of leaf and chloroplast development by the *Arabidopsis* gene pale cress, *Plant Cell* 6 (1994) 1253-64.
- Rew D.A., DNA microarray technology in cancer research, *Eur. J. Surg. Oncol.* 27 (2001) 504-8.
- Richmond T., Somerville S., Chasing the dream: plant EST microarrays, *Curr. Opin. Plant. Biol.* 3 (2000) 108-116.
- Rousseau F., Setif P. and Lagoutte B. Evidence for the involvement of PSI-E subunit in the reduction of ferredoxin by photosystem I, *Embo J.* 12 (1993) 1755-1765.
- Runge S., van Cleve B., Lebedev N., Armstrong G., Apel K., Isolation and classification of chlorophyll-deficient *xantha* mutants of *Arabidopsis thaliana*, *Planta* 197 (1995) 490-500.

S

- Sambrook J., Fritsch E.F. and Maniatis T., *Molecular cloning: A laboratory manual*. 2nd Ed. Plainview, NY: Cold Spring Harbor Laboratory Press (1989).
- Sanguinetti C.J., Dias Neto E., Simpson A.J., Rapid silver staining and recovery of PCR products separated on polyacrylamide gels, *Biotechniques* (1994) 914-21.
- SantaLucia J. Jr., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, *Proc. Natl. Acad. Sci. USA* 95 (1998) 1460-5.

-
- Sauer K., Primary events and the trapping of energy. In: Bioenergetics of Photosynthesis, Govindjee, ed., 116-181. Academic Press, New York (1975).
 - Schreiber U., Schliwa U. and Bilger W., Continuous recording of photochemical and non-photochemical chlorophyll fluorescence quenching with a new type of modulation fluorometer, *Photosynth. Res.* 10 (1986) 51-62.
 - Shikanai T., Munekage Y., Shimizu K., Endo T., Hashimoto T., Identification and characterization of *Arabidopsis* mutants with reduced quenching of chlorophyll fluorescence, *Plant Cell Physiol.* 40 (1999) 1134-1142.
 - Simpson D.J., von Wettstein D., Macromolecular physiology of plastids. XIV. *Viridis* mutants in barley: Genetic, fluoroscopic and ultrastructural characterisation, *Carlsberg Res. Commun.* 45 (1980) 283-314.
 - Somerville C., Somerville S., Plant functional genomics, *Science* 285 (1999) 380-383.
 - Somerville C.R., Analysis of photosynthesis with mutants of higher plants and algae, *Annu. Rev. Plant. Physiol.* 37 (1986) 467-507.
 - Steiner-Lange S., Gremse M., Kuckenberger M., Nissing E., Schachtele D., Spenrath N., Wolff M., Saedler H., Dekker K., Efficient identification of *Arabidopsis* knock-out mutants using DNA-arrays of transposon flanking sequences, *Plant Biology* 3 (2001) 391-397.
 - Stoesser G., Baker W., van Den Broek A., Camon E., Garcia-Pastor M., Kanz C., Kulikova T., Lombard V., Lopez R., Parkinson H. et al., The EMBL nucleotide sequence database, *Nucleic Acids Res.* 29 (2001), 17-21.
 - Sundberg E., Slagter J.G., Fridborg I., Cleary S.P., Robinson C., Coupland G., *ALBINO3*, an *Arabidopsis* nuclear gene essential for chloroplast differentiation, encodes a chloroplast protein that shows homology to proteins present in bacterial membranes and yeast mitochondria, *Plant Cell* 9 (1997) 717-730.

T

- Taiz L., Zeiger E., Photosynthesis: the light reactions. In *Plant Physiology* (Beard Brady E. and Donohoe L. eds.) The Benjamin / Cummings Publishing Company (1991) pp. 179-218.
 - Tateno Y., Miyazaki S., Ota M., Sugawara H. and Gojobori T., DNA data bank of Japan (DDBJ) in collaboration with mass sequencing teams, *Nucleic Acids Res.* 28 (2000) 24-26.
 - 2:
 - Taylor W.C., Barkan A., Martienssen R.A., Use of nuclear mutants in the analysis of chloroplast development, *Dev Genet.* 8 (1987) 305-20.
 - The *Arabidopsis* Genome Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, *Nature* 408 (2000) 796-815.
-

-
- The *C. elegans* Sequencing Consortium, Genome sequence of the nematode *C. elegans*: a platform for investigating biology, *Science* 282 (1998) 2012-2018.
 - The International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome, *Nature* 409 (2001) 860-921.
 - Thompson J.D., Higgins D.G. and Gibson T.J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* 22 (1994) 4673-4680.
 - Thorneycroft D., Sherson S.M., Smith S.M., Using gene knockouts to investigate plant metabolism, *J. Exp. Bot.* 52 (2001) 1593-601.
 - Tissier A.F., Marillonnet S., Klimyuk V., Patel K., Torres M.A., Murphy G., Jones J.D., Multiple independent defective *suppressor-mutator* transposon insertions in *Arabidopsis*: a tool for functional genomics, *Plant Cell* 11 (1999) 1841-1852.

U

V

- Vert G., Briat J.F., Curie C., Arabidopsis IRT2 gene encodes a root-periphery iron transporter, *Plant J.* 26 (2001) 181-9.

W

- Walker M.B., Roy L.M., Coleman E., Voelker R., Barkan A., The maize *tha4* gene functions in sec-independent protein transport in chloroplasts and is related to *hcf106*, *tatA*, and *tatB*, *J. Cell Biol.* 147 (1999) 267-276.
- Weber N. and Strotmann H., On the function of subunit PsaE in chloroplast Photosystem I, *Biochim. Biophys. Acta* 1143 (1993) 204-210.
- Wheeler D.L., Church D.M., Lash A.E., Leipe D.D., Madden T.L., Pontius J.U., Schuler G.D., Schriml L.M., Tatusova T.A., Wagner L. and Rapp B.A., Database resources of the National Center for Biotechnology Information, *Nucleic Acids Res.* 29 (2001) 11-16.
- Wisman E., Hartmann U., Sagasser M., Baumann E., Palme K., Hahlbrock K., Saedler H., Weisshaar B., Knock-out mutants from an En-1 mutagenized *Arabidopsis thaliana* population generate phenylpropanoid biosynthesis phenotypes, *Proc. Natl. Acad. Sci. USA* 95 (1998) 12432-12437.

- Wollman F.A., State transitions reveal the dynamics and flexibility of the photosynthetic apparatus, *EMBO J.* 20 (2001) 3623-30.
- Wu D., Wright D.A., Wetzelschloffer C., Voytas D.F., Rodermel S., The *IMMUTANS* variegation locus of *Arabidopsis* defines a mitochondrial alternative oxidase homolog that functions during early chloroplast biogenesis, *Plant Cell* 11 (1999) 43-55.

X

- Xu Q., Yu L., Chitnis V.P. and Chitnis P.R., Function and organization of photosystem I in a cyanobacterial mutant strain that lacks PsaF and PsaJ subunits, *J. Biol. Chem.* 269 (1994) 3205-3211.

Y

- Yamamoto Y., Tsuji H. and Obokata J., Structure and expression of a nuclear gene for the PSI-D subunit of photosystem I in *Nicotiana glauca*, *Plant Mol. Biol.* 22 (1993) 985-994.
- Yu L., Zhao J., Mühlenhoff U., Bryant D.A. and Golbeck J.H., PsaE is required for in vivo cyclic electron flow around photosystem I in the cyanobacterium *Synechococcus* sp. PCC 7002, *Plant Physiol.* 103 (1993) 171-180.

Z

- Zhang M.Q., Large-scale gene expression data analysis: a new challenge to computational biologists, *Genome Res.* 9 (1999) 681-8.
- Zhao J., Snyder W.B., Mühlenhoff U., Rhiel E., Warren P.V., Golbeck J.H. and Bryant D.A., Cloning and characterization of the *psaE* gene of the cyanobacterium *Synechococcus* sp. PCC 7002: characterization of a *psaE* mutant and overproduction of the protein in *Escherichia coli*, *Mol. Microbiol.* 9 (1993) 183-194.
- Zilber A.L. and Malkin R., Ferredoxin cross-links to a 22 kD subunit of photosystem I, *Plant Physiol.* 88 (1988) 810-814.
- Zuker M., On finding all suboptimal foldings of an RNA molecule, *Science* 244 (1989) 48-52.

APPENDIX

ERKLÄRUNG

“Ich versichere, daß ich die von mir vorgelegte Dissertation selbständig angefertigt, 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. Francesco Salamini betreut worden.”

Köln, den 25.10.2001

Claudio Varotto

Teilpublikationen:

Varotto C., Pesaresi P., Maiwald D., Kurth J., Salamini F., Leister D.,
Identification of photosynthetic mutants of Arabidopsis by automatic screening for altered effective quantum yield of photosystem II.
Photosynthetica 2000; 38: 497-504.

Varotto C., Pesaresi P., Meurer J., Oelmüller R., Steiner-Lange S., Salamini F.,
Leister D.
Disruption of the Arabidopsis photosystem I gene *psaE1* affects photosynthesis and impairs growth.
Plant J. 2000, 22: 115-24.

Pesaresi P., Varotto C., Richly E., Kurth J., Salamini F., Leister D.
Functional genomics of Arabidopsis photosynthesis
Plant Physiol. Bioch. 2001, 39: 285-294

Varotto C., Richly E., Salamini F., Leister D.
“*GST-PRIME*” – a software package for genome-wide primer design suitable for the generation of gene sequence tags in *Arabidopsis* and other species.
Nucleic Acids Res. (*In press*)

LEBENS LAUF

Name

Varotto Claudio

Adresse

Max-Planck-Institut für Züchtungsforschung
 Carl-von-Linne Weg 10
 D-50829 Köln
 Tel: +49-221-5062-416
 Fax: +49-221-5062-413
 e-mail: varotto@mpiz-koeln.mpg.de

Geburtsdatum und -ort

12.12.1973 Negrar (Verona)
 Italien

Nationalität

italienisch

Ausbildung

1979-1984	Grundschule
1984-1992	Gymnasium
1992	Abitur
1992-1998	Studium der Biotechnologie an der Universität Verona, Italien
09/1996-03/1998	Diplomarbeit an der Universität Verona, Prof. R. Bassi, Thema: "Point mutation analysis of LHCII, the major antenna complex of Photosystem II."
05.03.1998	Diplomprüfung
03-07/1998	Betreuung von Praktika an der Universität in Verona als Assistent
08/1998	Beginn der Doktorarbeit am Max-Planck Institut für Züchtungs-forschung, Köln, Abt. Prof. F. Salamini, Arbeitsgruppe Dr. D. Leister

Publikationen

- 1 De Luca C., Varotto C., Svendsen I., Polverino De Laureto P., Bassi R. Multiple light-harvesting II polypeptides from maize mesophyll chloroplasts are distinct gene products. *J Photochem Photobiol B*. 1999 Mar; 49(1): 50-60.
- 2 Croce R., Remelli R., Varotto C., Breton J., Bassi R. The neoxanthin binding site of the major light harvesting complex (LHCII) from higher plants. *FEBS Lett*. 1999 Jul 30; 456(1): 1-6.
- 3 Leister D., Varotto C., Pesaresi P., Niwergall A., Salamini F. Large-scale evaluation of plant growth in *Arabidopsis thaliana* by non-invasive image analysis. *Plant Physiol. Bioch*. 1999 Sep; 37(9): 671-678
- 4 Remelli R., Varotto C., Sandona D., Croce R., Bassi R. Chlorophyll binding to monomeric light-harvesting complex. A mutation analysis of chromophore-binding residues. *J Biol Chem*. 1999 Nov 19; 274(47): 33510-21.
- 5 Varotto C., Pesaresi P., Maiwald D., Kurth J., Salamini F., Leister D. Identification of photosynthetic mutants of *Arabidopsis* by automatic screening for altered effective quantum yield of photosystem II. *Photosynthetica* 2000; 38: 497-504.
- 6 Varotto C., Pesaresi P., Meurer J., Oelmüller R., Steiner-Lange S., Salamini F., Leister D. Disruption of the *Arabidopsis* photosystem I gene *psaE1* affects photosynthesis and impairs growth. *Plant J*. 2000 Apr; 22(2): 115-24.
- 7 Pesaresi P., Varotto C., Richly E., Kurth J., Salamini F., Leister D. Functional genomics of *Arabidopsis* photosynthesis. *Plant Physiol. Bioch*. 2001 Mar-Apr; 39(3-4): 285-294
- 8 Pesaresi P., Varotto C., Meurer J., Jahns P., Salamini F., Leister D. Knock-out of the plastid ribosomal protein L11 in *Arabidopsis*: effects on mRNA translation and photosynthesis. *Plant J*. 2001 Aug; 27(3): 179-189
- 9 Varotto C., Richly E., Salamini F., Leister D. 'GST-PRIME' – a software package for genome-wide primer design suitable for the generation of gene sequence tags in *Arabidopsis* and other species. *Nucleic Acids Res.* (*In press*)

ACKNOWLEDGMENTS

I would like to thank Prof. Dr. F. Salamini for providing me the opportunity to work in his department at the Max-Planck-Institut für Züchtungsforschung and Prof. Dr. H. Saedler for providing the fellowship for this Ph.D.

I wish to express my gratitude to Dr. Dario Leister for taking the responsibility of being my supervisor and for suggestions, discussions and encouragement during my Ph.D.

I am grateful to Paolo Pesaresi for his friendship and valuable help in every situation. I thank him also for the immunoblottings performed on the *psae1-1* mutant showed in this thesis.

Many thanks to all the colleagues and friends of my group, Alexander, Alexandra, Angela, Daniela, Erik, Fabio, Joachim, Marco and Martin. A special mention is due to the help given me by Alexandra, Daniela and Angela to solve small and big bureaucratic problems; moreover, Erik Richly is acknowledged for the design of the GST-PRIME module for sequence download and critical reading of this thesis, Marco Tizzano e Fabio Schiavon for contributing to the reverse screening of PSI subunits and Angela Lessnick for helping with immunoblotting of *psaG*, *psaH2* and *psaG/psaH2* knockouts.

I'd like to thank also external collaborators for their precious help (Dr. Peter Jahns for HPLC analysis, Dr. Jörg Meurer for the PSI absorption measurements of the *psae1-1* mutant, Dr. Jörg Kolbowski for technical assistance with the screening robot).

Finally, I would like to express my deep gratitude to my brother Marco, without whom this Ph.D. would not have been possible, and to my parents, Rita and my grand mother for their support and encouragement; *jiejie* to Mingai for her invaluable help.

*„Gutta cavat lapidem non vi,
sed saepe cadendo“*