Characterization of Interaction Between the Nucleus-Encoded TBC2 Protein with the 5'-Untranslated region of *psbC* mRNA and Associated Factors in *Chlamydomonas reinhardtii*

Joanne Wyglinski

A Thesis

In

the Department

of

Biology

Presented in Partial Fulfillment of the Requirements

For the Degree of Master of Science in Biology at

Concordia University

Montréal, Québec, Canada

October 2005

Your file Votre référence
ISBN: 0-494-14227-8
Our file Notre référence
ISBN: 0-494-14227-8

ABSTRACT

Characterization of Interaction Between the Nucleus-Encoded TBC2 Protein with the 5'-Untranslated Region of *psbC* mRNA and Associated Factors in *Chlamydomonas reinhardtii*

Joanne Wyglinski

This thesis has three main objectives, the main focus being a study of the TBC2 translational activation protein and its association with the *psbC* mRNA 5'-untranslated region. Prior experimental evidence using size-exclusion chromatography showed that TBC2p was 400 kDa in mass when its predicted molecular weight is only 114.8 kDa, implying association with additional factors. Bioinformatics analysis was undertaken to determine additional characteristics of the TBC2 protein and the existence of possible paralogs. The translated *TBC2* cDNA sequence had two paralogous protein sequences located within possible nucleotide open reading frames designated *TBC2A* and *TBC2B*, and all contained sequences resembling chloroplast transit peptides. TBC2Ap and TBC2Bp harbored the novel PPPEW repeat originally found in TBC2p. Initial experimental steps such as the preparation of coimmunoprecipitation involving TBC2p were taken.

The second part of this thesis encompasses the theorized existence of a novel thylakoid biogenesis compartment in the *C. reinhardtii* chloroplast. This is based on the theory that light does not control synthesis *per se* but only the transport of proteins from envelope-like membranes to the thylakoids. Pulse-labeling experiments revealed equivalent synthesis of envelope-like membrane proteins in light versus dark growth conditions, providing no evidence supporting the transport theory through a novel thylakoid biogenesis compartment.

The final topic of this thesis addresses the theory of chloroplast genome localization, paralleled to prior findings that the PEND protein in peas binds plastid nucleoids to the envelope membrane. A restriction digest determined that DNA found associated with low density membranes was only nuclear.

Acknowledgements

I would like to express my thanks to my supervisor, Dr. Zerges, as well as to Dr. Gulick, my committee member. Special thanks to committee member Dr. Herrington whom, without her help, this thesis would not have been possible. My thanks also to Dr. Auchincloss for the strains and plasmids used in this thesis, to Daniel Darmon for his help in the bioinformatics segment, and to Dr. Ulycznyj, among others, for use of equipment and technical support.

Thank you to my family for their dedication and love.

List of Figures

List of Tables

1.A. Preface

In order to better appreciate the TBC translational activator protein section, it is necessary to know the background and complexity of knowledge in the area to date, detailed in the following subjects:

-*Chlamydomonas reinhardtii* as an experimental organism

-Molecular aspects of the nuclear and chloroplast genome

-The Z scheme

-Photosystem II and components of the electron transport chain

-Characteristics of *psbC*, its mRNA and its encoded protein, P6

-The TBCps and their mode of interaction with *psbC* mRNA

-Cis-acting mutations in the 5'-UTR of *psbC* mRNA which defined the role of the TBC factors

-5'-UTR deletional mutations affecting translation initiation

1.A.1. *Chlamydomonas reinhardtii* as an experimental organism

According to Harris (1989), Sager and Granick (1953), *Chlamydomonas reinhardtii* is a unicellular, eukaryotic green alga which has the ability to grow heterotrophically through the assimilation of acetate in the glyoxylate cycle. Timko (1998) states that in dark conditions, chlorophyll production rates are adequate enough that the biogenesis of fully functional photosynthetic apparatuses is allowable. *C. reinhardtii* also possesses the capability of growing mixotrophically, that is, in both the presence of acetate and light (Mets and Rochaix 1998). Growth is also possible in autotrophic conditions, with CO_2 being

used as the sole carbon source, but at a much slower rate in comparison to mixotrophic conditions (Mets and Rochaix 1998).

Harris, Rochaix and colleagues found *C. reinhardtii* to be an optimal experimental organism because of its amenability to numerous experimental methods, including genetic, molecular and biochemical approaches (rev. by Zerges 2004). For example, *C. reinhardtii* has been used in chloroplast gene expression studies involving the mutagenesis of nuclear and chloroplast genes encoding subunits of Photosystem II (PSII), one of the main focuses of this thesis (Zerges *et al.* 1994).

Another benefit of using *C. reinhardtii* as an experimental organism is that it is unique in that its sexual life cycle can be controlled, at the same time possessing the ability to grow heterotrophically (Mets and Rochaix 1998). Normally, it exists as a haploid cell; opposite mating types can undergo gametogenesis, fusing to become a diploid zygote which, after maturation and meiosis, releases a tetrad of four vegetative clones (Mets and Rochaix 1998). Tetrad analysis allows for the genetic analysis of nuclear mutations, particularly in assessing the dominance of mutations and complementation analysis (Mets and Rochaix 1998). The generation of crosses will be discussed later in the methods section regarding the creation of new genotypes.

1.A.2. Molecular aspects of the nuclear and chloroplast genome

The nuclear genome contains 264 molecular markers found within 17 known linkage groups, the genome being approximately 1025 centimorgans in Kosambi units or 10^8 bp (Kathir *et al.* 2003). Harris (1989) described the nuclear genome as having an exceptionally high GC content of 62%. The number of repetitive sequences found in the nuclear genome has not been established concretely, however, a review by Harris (1989) discussed an estimation that they constitute up to 30% of the total genome.

C. reinhardtii contains a single chloroplast possessing a 203 kb plastid chromosome which exists in either circular or linear form (Maul *et al.* 2002). The chloroplast genome possesses 99 known genes such as atypical genes encoding RNA polymerase (Maul *et al.* 2002). The chloroplast genome is characterized by an excess of repetitive DNA sequences such as short dispersed repeats (SDRs), comprising more than 20% of the genome (Maul *et al.* 2002). The presence of SDRs may be for structural or evolutionary reasons (Maul *et al.* 2002).

Chloroplast genes have been grouped into three general classes pertaining to function: transcription and translation-related, photosynthesis-related and those encoding products required for the biosynthesis of amino acids, fatty acids and pigments (Kapoor and Sugiura 1998). Genes encoding rRNA species (*rrnS* genes) of the chloroplast genome are clustered and transcribed as an operon (Boudreau *et al.* 1994). Most chloroplast genes in *C. reinhardtii* are transcribed and translated as monocistronic mRNAs and often encode proteins

required for a functional electron transport chain, such as the subunits of PSII (rev. by Sugiura *et al.* 1998).

Aminoacyl tRNA synthetases, 70s ribosomes, initiation factors, elongation factors and release factors comprise chloroplast translational machinery (rev. by Sugiura *et al.* 1998). However, not all components of chloroplast translational machinery in *C. reinhardtii* are encoded by the chloroplast (E. Harris, email comm.). This is attributed to the theory of endosymbiosis whereby chloroplasts were once prokaryotic cyanobacteria (rev. by Timmis *et al.* 2004). The chloroplast genome in *C. reinhardtii* is diminished in size in comparison to the cyanobacterial gen ome due to translocation to the nuclear genome (rev. by Timmis *et al.* 2004). This occurrence was demonstrated by the gene coding for translation initiation factor (IF) 1, *infA*, shown in angiosperms to exist in the nuclear genome when not found in the chloroplast genome, its normal location for this phylum (Millen *et al.* 2001). A difference in interorganellar translocation rate, however, was shown to exist in *C. reinhardtii* in comparison to higher plants, the rate in *C. reinhardtii* being six-fold lower (rev. by Timmis *et al.* 2004). This was theorized to be due to the presence of only one chloroplast in *C. reinhardtii* (rev. by Timmis *et al.* 2004).

Ribosomal RNAs (rRNAs) in *C. reinhardtii*'s chloroplast are encoded by the chloroplast genome (E. Harris, email comm.). For example, 16S rRNA (NCBI [locus CHCR16S, version X03269.1 gi: 11417]), 23S rRNA (NCBI [locus CHCR23S, version X16687.1 gi: 11421]) and 5S rRNA (NCBI [locus CHCR5S, version X03271.1 gi: 11422]) are encoded by this genome.

C. reinhardtii ribosomal proteins required for chloroplast translational processes are encoded by both chloroplast and nuclear genomes (E. Harris, email comm.). Proteins encoded by the chloroplast genome include S3 (NCBI [locus CRPSBFLG, version X66250.1 gi: 393459]), S7 (NCBI [locus CHCRFUDD, version X53977.2 gi: 10178859]), S12 (NCBI [locus CRECPRPS12, version M29284.1 gi: 336681]), L14 (NCBI [locus CHCRL14, version X14062.1 gi: 11432]), L16 (NCBI [locus CHCRL14, version X14062.1 gi: 11432]) and L20 (NCBI [locus CHCRTRNRP, version X62566.1 gi: 11463]).

The *tufA* gene (NCBI [locus CRTUFA, version X52257.1 gi: 14314]) codes for the elongation factor Ef-Tu of the chloroplast of *C. reinhardtii. tufA* is also located on the chloroplast genome (NCBI Database).

Chloroplast transfer RNAs (tRNAs) in *C. reinhardtii* are encoded by the chloroplast genome (E. Harris, email comm.). These include genes *trnT* (NCBI [locus CRDNATATT, version X75037.1 gi: 404181]) which encodes transfer RNA for threonine, and *trnR* (NCBI [locus CRDNATATT, version X75037.1 gi: 404181]) which encodes transfer RNA for arginine.

The remaining components of the chloroplast translational machinery in *C. reinhardtii* are encoded by nuclear genes (E. Harris, email comm.). These include initiation factors 1, 2 and 3, the IF1 factor being coded for by the *infA* gene (JGI_5932, scaffold_178 (-)=121,785-123,160), and release factors (RFs) RF1, RF2 and RF3.

1.A.3. The Z scheme

PSII belongs to a larger system of multisubunit complexes which compose the electron transport chain. The "Z-scheme," discussed by Redding and Peltier (1998), is a model proposed for electron transport whereby electrons originating in PSII are transported to Photosystem I (PSI) via intermediates such as the plastoquinone pool, the cytochrome b_6f complex and plastocyanin.

The following scheme described is considered to be the classic electron transport chain scheme (Redding and Peltier 1998). The first step in this chain entails that the PSII complex oxidizes water and reduces plastoquinone and cytochrome (Redding and Peltier 1998). The scheme initially involves the acceptance of electrons in the PSII complex from water (Redding and Peltier 1998). These electrons are subsequently passed onto the Mn_4 cluster and are then shuttled to tyrosine which donates these electrons to P680 (Redding and Peltier 1998). The electrons are shuttled back to pheophytin, $quinone_A$ and eventually $quinone_B$ (Redding and Peltier 1998). Electrons then flow through plastoquinol to the Rieske iron-sulfur centre of cytochrome b_6f (Redding and Peltier 1998). Plastoquinone is shuttled back to PSII (Redding and Peltier 1998). The electrons continue to plastocyanin and cytochrome C located in the lumen of the thylakoid and are accepted by P700 in PSI, a chlorophyll dimer (Redding and Peltier 1998). P700, along with primary acceptor A_0, A_1 and the 4Fe-4S centre F_x, are bound by PsaB and PsaA proteins (Dauvillée *et al.* 2003). The PsaB subunit of PSI, like many PSII subunits, is translationally activated by nucleus-encoded proteins, in this case Tab2p (Dauvillée *et al.* 2003). Within PSI, electrons are

shifted to a chlorophyll molecule, phylloquinone, a secondary acceptor and a tertiary, non-sulfur acceptor (Redding and Peltier 1998). Electrons are then passed onto F_I, F_{II} in the terminal electron acceptor of PSI (iron-sulfur centres) and subsequently ferrodoxin:$NADP^+$ reductase to convert $NADP^+$ to NADPH (Redding and Peltier 1998). Protons in the lumen are channeled through ATP synthase and used to drive ATP synthesis which, with NADPH, results in carbon fixation (Redding and Peltier 1998).

1.A.4. Photosystem II and components of the electron transport chain

The major core subunits of the dimerized PSII complex are cytochrome b-559, D1, D2, P5 and P6 proteins which are encoded by *psbE*, *psbA*, *psbD*, *psbB* and *psbC* genes, respectively (see Fig. 1) (Rochaix *et al.* 1989; Ruffle and Sayre 1998). De Vitry and colleagues (1991) found that subunits D1, D2, P5 and P6, through their purification and subjection to ion-exchange chromatography, existed in a 1:1:1:1 stoichiometry in the complex. Other subunits of the PSII include the trimeric light-harvesting complex (LHCII) which surrounds the PSII complex, and OEE stabilization proteins coded for by *psbO*, *psbP* and *psbQ*, to name a few (Nield *et al.* 2000; Ruffle and Sayre 1998). Choquet and others stated that translational regulation through a hierarchal order occurs to ascertain the 1:1:1:1 stoichiometric ratio so that proper folding and insertion into the complex occurs, it being hypothesized that chloroplast mRNA-specific translational regulators control this process (rev. by Zerges 2004). This model is known as "Control by Epistasy of Synthesis," or CES (rev. by Zerges 2004). Many translational regulators were determined to be nucleus-encoded, such as

TBC1p and TBC2p, "TBC" being an acronym for (T)ranslation of the mRNA encoding PSII (B) subunit P6 (see Fig. 1) (Zerges *et al.* 1997). Discussed by Chua and Bennoun (1975), Delepelaire (1984), Rochaix and colleagues (1989), recessive nuclear mutations at the *TBC1* and *TBC2* loci would result in photosynthesis deficiency due to an inability to accumulate P6 particles of PSII. TBC1p and TBC2p were proposed to be translational regulators of *psbC* mRNA encoding the P6 protein of PSII (see Figs. 1 and 2) (Zerges and Rochaix 1994). Stoichiometric control, however, also originates within the PSII subcomplex. Ohad, van Wijk, Zhang and colleagues reported that stoichiometric control was theorized to be exerted through D2 (rev. by Zerges 2004). They concluded the role of this subunit to be a stabilizing factor or receptor for nascent D1 intermediates assembled in the PSII core complex (rev. by Zerges 2004). Zhang and colleagues made these conclusions after successful recovery of the D1/D2 complex with antibodies specific for D2 through coimmunoprecipitation experiments (rev. by Zerges 2004). This was reaffirmed in cross-linking experiments with cysteine residues of the D2 protein shown to bond with D1 at specific regions of close proximity (rev. by Zerges 2004). The order of insertion of the subunits into the thylakoid membrane was determined as follows: the cytochrome b-559 α-subunit is co-translationally inserted into the thylakoid membrane and acts as a scaffold for the subsequent insertion of D2 (rev. by Zerges 2004). P6, encoded by *psbC* mRNA, is inserted into this complex and subsequently D1 (rev. by Zerges 2004). *psbB* mRNA translation results in the addition of P5 (rev. by Zerges 2004). After proper scaffolding has been

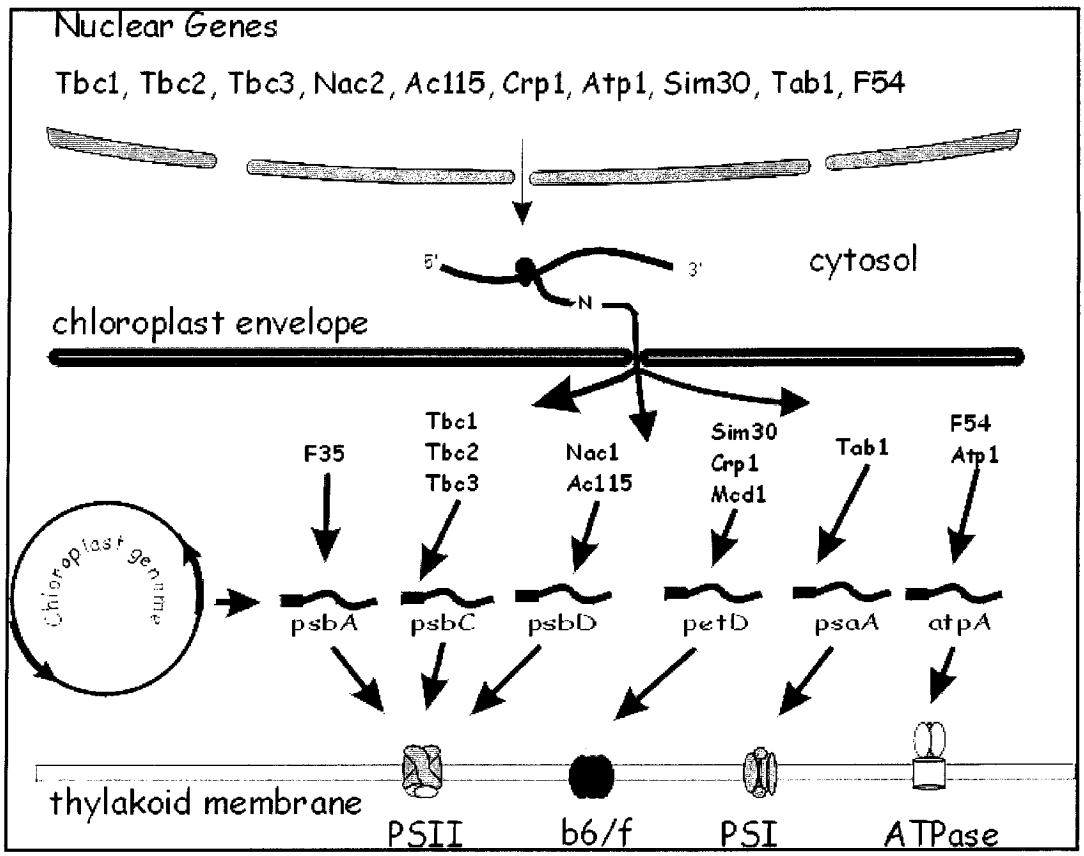

Fig. 1. Nucleus-encoded translational activator proteins target chloroplast mRNAs encoding electron transport chain subunits. mRNAs of nucleus-encoded translational activator proteins are translated by cytosolic ribosomes. Containing transit peptides, these factors are targeted and imported into the chloroplast, translationally activating chloroplast mRNAs expressing subunits of complexes of the electron transport chain (Reprinted from Rev. Biochimie, vol. 82, Zerges, W. "Translation in chloroplasts." p. 583-601, © 2000 with permission from Elsevier).

established, nucleus-encoded proteins (light-harvesting proteins, subunits of the oxygen-evolving complex), which are translated in the cytosol and imported into the chloroplast, are assembled into the complex, resulting in a functional PSII complex (rev. by Zerges 2004). The proteins of the PSII complex will be relevant in the second topic of this thesis, the theory being that these proteins are synthesized in a novel thylakoid biogenesis compartment and transported through vesicles to the thylakoid membranes.

The role of light induction of the complex particles of the electron transport chain has also been explored. Herrin, Malnoe and colleagues found that a transition from dark to light-growth conditions induced PSII complex synthesis through increased levels of D1, D2 and the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (LSU of RuBisCO) (rev. by Zerges 2004). RuBisCO is an enzyme which catalyses 3-phospho-D-glycerate synthesis from ribulose bisphosphate as well as the oxidation of ribulose bisphosphate by O_2 to 3-phospho-D-glycerate and 2-phosphoglycol (ChemiCool). This enzyme requires Mg^{2+} for activity and is coded for by the *rbcL* gene in *C. reinhardtii* (Spreitzer 1998; ChemiCool). The aspect of protein synthesis due to light will be considered when the second topic of this thesis, the possible existence of a novel thylakoid biogenesis compartment, is discussed.

1.A.5. Characteristics of *psbC*, its mRNA and its encoded protein, P6

The P6 protein of PSII is encoded by the *psbC* gene of the chloroplast genome (see Figs. 1 and 2) (Zerges *et al.* 2003). The *psbC* gene was originally found by Rochaix and colleagues (1989) to be located between *Eco*RI fragments

R9 and R23, its genomic sequence later determined through Maxam-Gilbert sequencing. Rochaix and colleagues (1989) initially determined the *psbC* cDNA sequence through the Sanger method of dideoxy sequencing by annealing a 21-mer oligonucleotide 18 nucleotides away from the GTG initiation codon of its mRNA. Rochaix and colleagues (1989) established the GTG initiation codon based on the expected length of the amino acid sequence corresponding to the 43 kDa protein and comparisons made to the P6-equivalent amino acid sequence in spinach determined by Alt and colleagues (1984) (see Fig. 2).

psbC mRNA was found to have an extended 5'-untranslated region (5'-UTR) (see Fig. 2) (Rochaix *et al.* 1989). Zerges and colleagues (2003) reported that the 5'-UTR of the mRNA was 547 nucleotides in length with a region of interaction crucial for initiation of translation (see Fig. 2). Rochaix and colleagues (1989) determined the length of the 5'-UTR through S1 nuclease protection and primer extension experiments. These findings were also comparable to the 5'-UTR of *psbC* mRNA in spinach (Rochaix *et al.* 1989).

Zerges and colleagues (2003) wanted to determine which regions of the 5'-UTR were responsible for transcription, 5' end processing (rev. by Rochaix 1996) or stabilization (Nickelsen 1998) prior to determining regions involved in translation so that phenotypic results of mutations could be attributed properly. This was done in a series of experiments by Zerges and colleagues (1994, 2003) whereby deletions were made in the 5'-UTR of the *psbC* gene. These deletion mutants were ligated into the cg20-atpB-INT plasmid containing the *aadA* coding sequence followed by the 3' end of the *rbcL* gene (Zerges *et al.* 2003). RNA gel

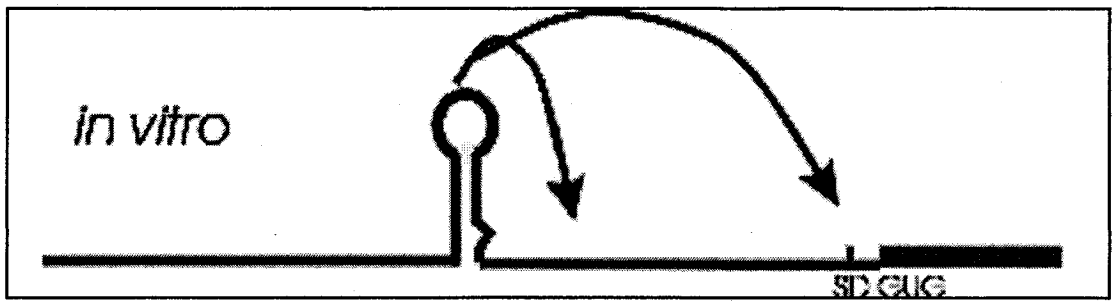

Fig. 2. The 5'-UTR of *psbC* mRNA. This line-drawing shows the 5'-UTR *psbC* mRNA stem loop region (position 223-320 with the 5'-most base being +1). The stem loop region is theorized to be thermodynamically destabilized by translational activator proteins. The SD-like sequence and the GUG initiation codon are also required for translation. (Figure produced by W. Zerges provided for use in this thesis).

blot analysis was performed (Zerges *et al.* 2003). mRNA levels were lowered for sequences with deletions made between bases 3-222 and bases 27-93 in the 5'-UTR (Zerges *et al.* 2003). With this knowledge, Zerges and colleagues proposed that the first 93 bases in the 5'-UTR were required for reasons such as 5' end processing, previously reviewed by Rochaix (1996). An alternate theory proposed by Zerges and colleagues (2003) was that this region was required for mRNA stabilization, as discussed previously by Nickelsen (1998), but not solely for that purpose since overlap for a translational activation requirement existed in region 27-93. The findings above were made for the *psbC-aadA-rbcL* construct only (Zerges *et al.* 2003).

The *psbC-atpA-aadA-rbcL* reporter construct lacking the Shine Dalgarno-like (SD) region (positions 534-538) and the GTG initiation codon was integrated into the *C. reinhardtii* chloroplast genome of *FuD50* and when its RNA was probed in RNA blot analysis, mRNA levels were basically unchanged in the presence of translational activator mutants (Zerges *et al.* 2003). Therefore, the hypothesis that the SD region is not required for transcription, 5' end processing (rev. by Rochaix 1996) or stabilization (Nickelsen 1998) was supported (Zerges *et al.* 2003).

1.A.6. The TBCps and their mode of interaction with *psbC* mRNA

It was mentioned earlier that expression of some chloroplast genes is controlled by translational regulator proteins coded for by nuclear genes (see Fig. 1) (Zerges *et al.* 2003). Three nucleus-encoded factors, TBC1p, TBC2p and TBC3p, interact together and contribute to the regulation of translation of *psbC*

mRNA (see Fig. 3A) (Zerges *et al.* 2003). Rochaix and colleagues (1989) and Zerges and colleagues (1994, 1997) found evidence indicating that these nucleus-encoded factors are translational activators of *psbC* mRNA expression (see Fig. 3). TBC2p will be discussed in detail since the main focus of this thesis is to characterize its interaction with the 5'-UTR of *psbC* mRNA and to determine any additional characteristics of this protein, as well as existing homologs or paralogs.

The *TBC1* locus was discovered in a mutant screen using methyl methane sulfonate resulting in the *tbc1-F34* mutant, selected for based on the mutant phenotype of high fluorescence (Chua and Bennoun 1975). High fluorescence is a phenotypic characteristic of a non-functional PSII complex which yields kinetic data with an initially high fluorescence yield but no eventual variable part (Chua and Bennoun 1975). Joliot and colleagues (1998) stated that fluorescence yield is a result of the redox state of the primary quinone acceptor of Q_A in PSII, together with other reasons, such as the number of pigments bound to the complex. Due to the scaffolding theory (rev. by Zerges 2004), a lack of P6 proteins would result in the inability to add additional subunits to PSII. Therefore, if P6 could not be expressed due to non-functional TBC1p caused by the *tbc1-F34* mutation, then PSII could not be stably formed (see Fig. 3B) (Zerges and Rochaix 1994). Genetic analysis revealed that the *tbc1-F34* mutation is nuclear since it was inherited in a Mendelian fashion (Chua and Bennoun 1975).

To prove the theory that *tbc1-F34* indirectly prevents the assembly of the PSII complex due to a lacking P6 subunit, pulse-labeling experiments using the

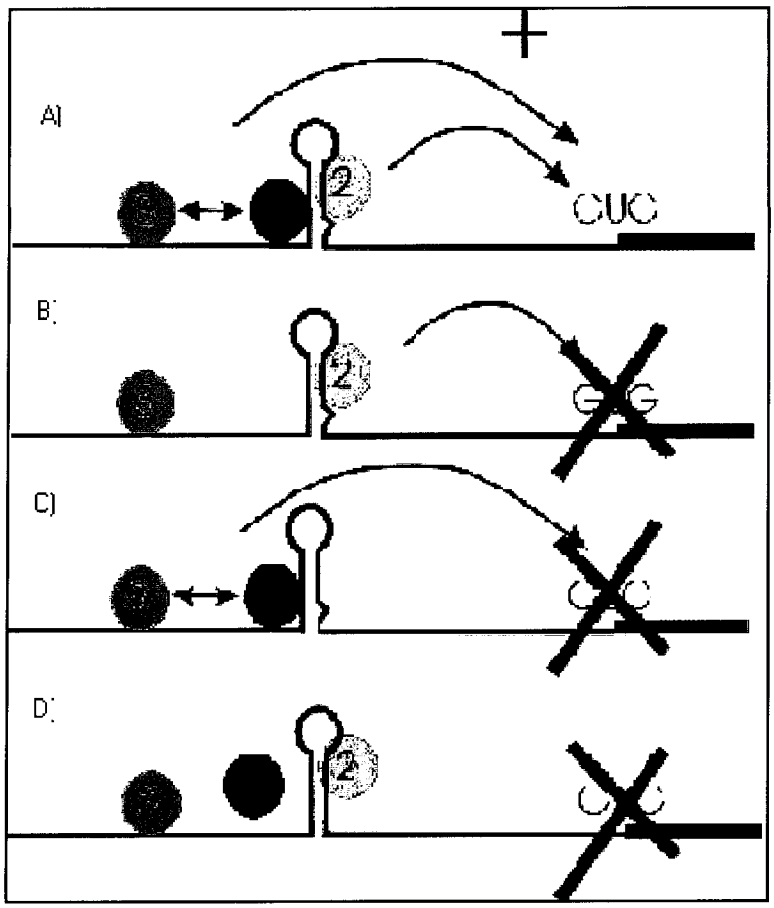

Fig. 3. Hypothetical interactions between the TBC factors and the *psbC* mRNA 5'-UTR. Four interactive combinations of the TBC factors with the *psbC* mRNA 5'-UTR-*aadA* chimeric gene: A, wild-type TBC1p, TBC2p and TBC3p factors successfully destabilize the stem loop, promoting translation; B, the *tbc1-F34* nuclear mutation results in a non-functional TBC1p factor and an inability to activate translation; C, the *tbc2-F64* nuclear mutation results in a non-functional TBC2p factor and an inability to activate translation; D, the *psbC-FuD34* mutation prevents TBC1p interaction resulting in non-activation of translation. (Figure produced by W. Zerges provided for use in this publication.)

tbc1-F34 mutant strain were performed to determine which components of the PSII complex existed (Zerges and Rochaix 1994). It was revealed that the *tbc1-F34* mutant did not express the 43 kDa chlorophyll a-binding PSII core subunit, P6 (see Fig. 3B) (Zerges and Rochaix 1994). Other polypeptide subunits of PSII normally integrated eventually undergo degradation without proper scaffolding (Zerges and Rochaix 1994).

The *TBC2* gene was discovered through a point mutation called *tbc2-F64* induced by ethyl methane sulfonate (EMS) (Bennoun *et al.* 1980; *Chlamydomonas* Genetics Centre, pers. comm.). The *TBC2* locus was found to be unlinked with other loci involved in PSII biogenesis (Auchincloss *et al.* 2002). These include loci involved in post-transcriptional steps of plastid genes such as the *ac114* locus on linkage group III and *ac115* on linkage group I, the *ac* loci known to participate in PSII biogenesis (Auchincloss *et al.* 2002). Core subunits of the PSII complex such as P5, P6 and the three OEE factors were not apparent in the *tbc2-F64* mutant (Rochaix *et al.* 1989). This was explained by the scaffolding theory when the lack of P6 prevented the integration of subsequent subunits resulting in their degradation (see Fig. 3C) (Rochaix *et al.* 1989). The missing P6 protein in the *tbc2-F64* mutant was demonstrated using the same method done for the *tbc1-F34* strain, t hrough 15 minute pulse-labeling experiments with [^{14}C] acetate and cycloheximide during synthesis (Rochaix *et al.* 1989).

The *TBC3* locus was discovered through the spontaneous partial phenotypic revertant of *psbC-FuD34* called RB1.1, shown to grow on minimal

medium (Zerges *et al.* 1997). The dominance of *tbc3-rb1* in its partial suppression of the PSII-deficient phenotype resulting from *psb C-FuD34* was reaffirmed in a strain heterozygous for *TBC3* (*tbc3-rb1/TBC3*) with a *psbC-FuD34* chloroplast mutation (Zerges *et al.* 1997). This strain had lowered PSII activity between that of wild-type and *psbC-FuD34* strains, shown in fluorescent transient data, indicating partial suppression of the *psbC-FuD34* mutation (Zerges *et al.* 1997). Dominance of suppression was also shown in *in vivo* pulse-labeling experiments that the *tbc3-rb1* nuclear mutation, in combination with the *psbC-FuD34* chloroplast mutation, had partial restoration (ca. 10%) of the presence of core subunits relative to the *psbC-FuD34* mutant (Zerges *et al.* 1997).

1.A.7. Cis-acting mutations in the 5'-UTR of *psbC* mRNA which defined the role of the TBC factors

A suppressor mutation was found when *tbc1-F34* was irradiated with UV light (see Fig. 4; Fig. 5C) (Chua and Bennoun 1975). The suppressor allowed growth on minimal medium due to a partially-restored PSII (see Fig. 4; Fig. 5C) (Chua and Bennoun 1975). Rochaix and colleagues (1989) reported similar findings when this mutation suppressed the *tbc1-F34* mutation, shown in pulse-labeling experiments when P6 levels were 20% that of wild-type (see Fig. 4; Fig. 5C). This mutation was named *psbC-F34sul* and was located at position 227 of the 5'-UTR in the region of the stem loop, leading Zerges and colleagues (2003) to conclude that this was the region of interaction with TBC1p (see Fig. 4C; Fig. 5C).

psbC-FuD34 was found through 5-fluorodeoxyuridine induction, its

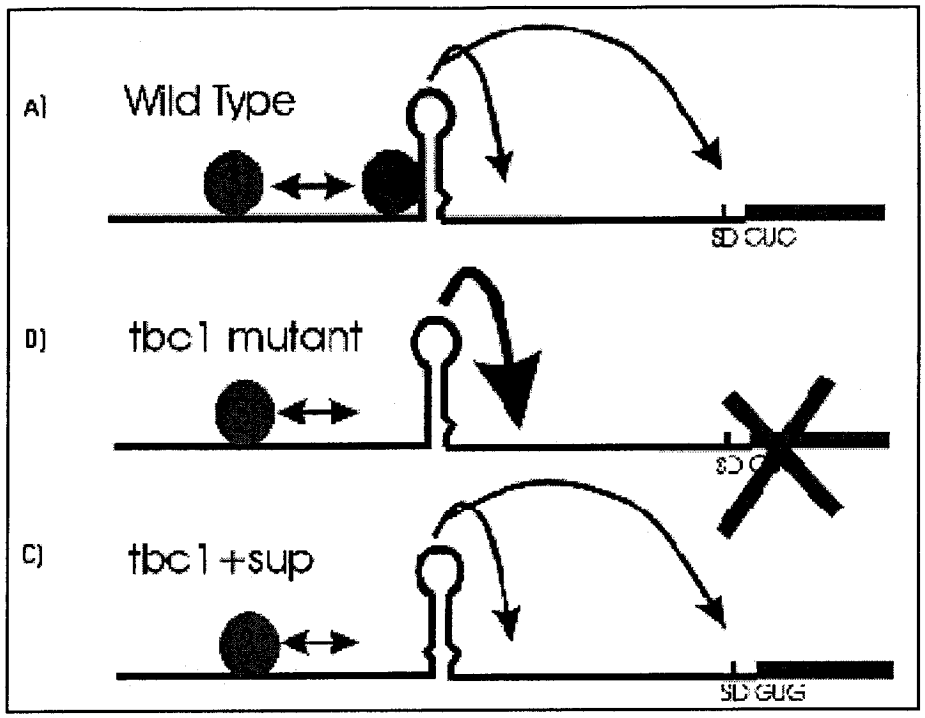

Fig. 4. Theorized interactions of a mutant TBC1p factor, the cis-acting *psbC-F34sul* suppressor mutation and the *psbC* mRNA 5'-UTR. Interactive combinations of the TBC1p factor with the *psbC* mRNA 5'-UTR-*aadA* chimeric gene: A, wild-type TBC1, TBC2 and TBC3 factors successfully destabilize the stem loop, promoting translation; B, the mutated *tbc1-F34* allele causes translational inactivation; C, cis-acting mutation *psbC-F34sul* at position +227 considerably alleviates the requirement for TBC1p at its proposed site of interaction, permitting the *tbc1-F34* mutant to confer lower-than-wild-type levels of spectinomycin resistance. (Figure produced by W. Zerges provided for use in this publication.)

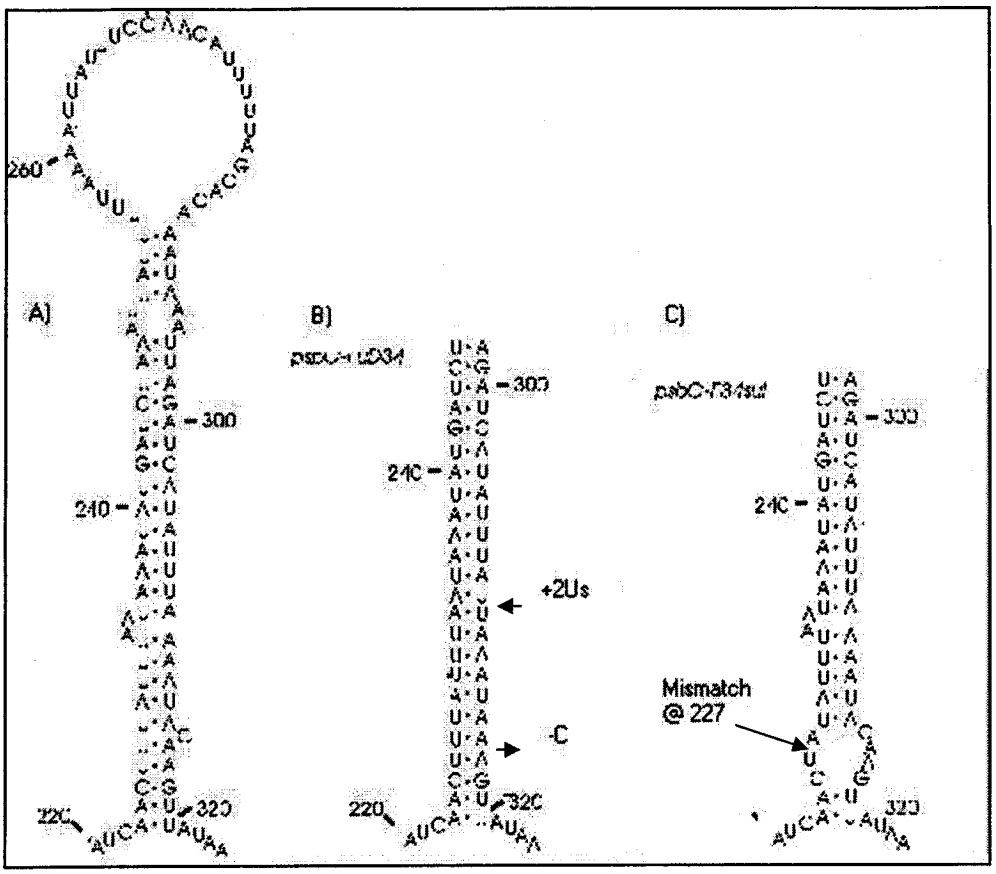

Fig. 5. Cis-acting mutations in the *psbC* mRNA 5'-UTR stem loop region. The *psbC* mRNA 5'-UTR stem loop is destabilized in the presence of wild-type TBCp factors. The following are cis-acting mutations (wild-type excluded) which may alter destabilization: A, the wild-type stem loop with its characteristic bulges; B, the *psbC-FuD34* mutation eliminates bulges at positions +233-234 and +314, adversely affecting destabilization of the stem loop by the TBC factors; C, the *psbC-F34sul* mutation is a result of a mismatch of a U for an A at position +227, creating a bulge and as a result suppressing the *tbc1-F34* nuclear mutation. (Figure produced by W. Zerges provided for use in this publication.)

phenotype PSII-deficient (see Fig. 3D; Fig. 5B) (Rochaix *et al.* 1989). The mutation is a result of an insertion of two Ts between bases +308 and +309 and a deletion of a C at position +314 in the *psbC* mRNA 5'-UTR, +1 being the 5'-most base (see Fig. 3D; Fig. 5B) (Zerges *et al.* 1997).

Alleviating the requirement for TBC3p, *tbc3-rb1* was also shown to suppress the *tbc1-F34* mutation for both PSII-deficiency and synthesis of P6 by alleviating the requirement for TBC1p interaction with the 222-319 stem loop region (Zerges *et al.* 1997). This was shown through fluorescence transient analysis revealing a functional PSII complex, as well as through pulse-labeling experiments which revealed an accumulation of 50% of the wild-type level of P6 (Zerges *et al.* 1997).

1.A.8. 5'-UTR deletional mutations affecting translation initiation

Since cis-acting mutations such as *psbC-FuD34* identified the regions of interaction of the TBC factors with the 5'-UTR of *psbC* mRNA, deletional analysis of the leader region was done to determine additional regions of interaction with translational activators (Zerges *et al.* 2003). Deletions constructed mainly by Zerges and colleagues (2003) were ligated in a vector as a 5'-UTR *psbC-aadA-rbcL* reporter gene containing the 3' *atpB* partial gene which, when transformed into *FuD50*, complements the strain's 5' *atpB* partial gene.

A major region of interaction of the TBC1p factor was within the cis-acting region 321-519, specifically the region 321-378 (Zerges *et al.* 2003). This was shown when spectinomycin resistance levels remained constant in a *tbc1-F34* strain with an *aadA* gene ligated to the 5'-UTR of the *psbC* gene (Zerges *et al.*

2003). Deletion mutations located between positions 379-519 displayed lowered but not abolished levels of spectinomycin resistance in the *tbc1-F34* and *tbc2-F64* mutant, reflecting that this region was less required for TBC1p and TBC2p interaction (Zerges *et al.* 2003). The same region of deletion showed a slight lowering of spectinomycin levels in a wild-type strain (Zerges *et al.* 2003). The region 95-222 deleted (ie. Δ95-222, Δ95-167, and Δ168-222) reflected a requirement for translation initiation factors since spectinomycin resistance levels were severely reduced in comparison to wild-type strains (Zerges *et al.* 2003).

The translation initiation sequence of the *psbC* mRNA 5'-UTR consists of a 14 nucleotide SD-like region as well as a GTG initiation codon (Zerges *et al.* 2003). This translation initiation region was deleted and the remaining sequence of the 5'-UTR was ligated into a vector in the form of 5'-UTR *psbC-atpA-aadA-rbcL*, the *atpA* fragment replacing the initiation region (Zerges *et al.* 2003). Results showed that the mutant *tbc1-F34* strain generated spectinomycin resistance levels comparable to wild-type, indicating that a SD-like sequence and the GTG initiation codon are required for TBC1p interaction (Zerges *et al.* 2003). The *tbc2-F64* strain, on the other hand, had levels which were severely affected, leading to the conclusion that TBC2p does not interact with these regions (Zerges *et al.* 2003). The region neighboring the SD region between positions 519-532 was shown to be required since spectinomycin resistance levels were lowered drastically (Zerges *et al.* 2003).

Considered to be a key region of interaction of the 5'-UTR of *psbC* mRNA, a deletion was made in the stem loop region between positions 223-320,

resulting in complete eradication of translation in wild-type strains since no spectinomycin resistance was conferred (Zerges *et al.* 2003). A deletion mutation between positions 484-532 also eliminated translation (Zerges *et al.* 2003). Because translation is non-existent in these two mutants, the requirement for TBC1p and TBC2p could not be assessed (Zerges *et al.* 2003).

1.B. Background review of the TBC Factors

The main focus of this thesis revolves around the subject of the TBC factors, therefore, different subjects were evaluated and noted to prepare for bioinformatics analysis and experimental objectives of this section. They are listed as follows:

-Previous characterization of TBC2p

-The exclusive PPPEW repeat and other distinguishing characteristics found in TBC2p

-PPR repeats and their relevance to TBC2p

-The TPR repeat and its similarity to the PPR repeat

-Crp1, a protein with significant homology to TBC2p

-TBC2p, like Crp1p, has a transit peptide

-TBC2p pertaining to the aim of this thesis

1.B.1. Previous characterization of TBC2p

TBC2p, which had a triple hemagglutinin epitope at the COOH terminus of the protein, was immunoblotted and found to be 140 kDa in mass and prevalent in whole chloroplasts and chloroplast-derived soluble fractions (Auchincloss *et al.* 2002). The expected molecular weight of TBC2p, determined through its predicted protein sequence using *TBC2* cDNA, is 114.8 kDa with its triple hemagglutinin epitope being 4.2 kDa (Auchincloss *et al.* 2002). Therefore, a 0.5 M $(NH_4)_2SO_4$ wash was done to remove loosely-bound proteins to account for the 21 kDa discrepancy, but to no avail (Auchincloss *et al.* 2002). Since

RuBisCO and thioredoxin protein controls migrated according to their expected weights, speculation still remains as to what accounts for the 21 kDa increase in mass (Auchincloss *et al.* 2002). A possible explanation to this increase in mass could be protein aggregation.

Size-exclusion chromatography of total soluble cell extracts showed that TBC2p eluted as 400 kDa, strongly suggesting that it belongs to a complex (Auchincloss *et al.* 2002). RNAse treatment was done to account for additional mass contributed by RNA but, after subjection to size-exclusion chromatography, the complex eluted again at 400 kDa, indicating that any RNA associated with the complex bound either transiently or not stably (Auchincloss *et al.* 2002). The RB60 protein was used as a control in this experiment since it was considered to possibly interact with the TBC2 protein (Auchincloss *et al.* 2002). The RB60 protein, unlike TBC2p, did shift to a lower molecular weight when subjected to RNAse treatment, verifying that it did not belong to the TBC2p complex (Auchincloss *et al.* 2002). Mg^{2+} was taken away from the surrounding solution of the 400 kDa complex to cause dissociation of TBC2p from the complex if bound transiently (Auchincloss *et al.* 2002). The mass of the complex even surpassed 400 kDa after having undergone this treatment, indicating possible aggregation of proteins (Auchincloss *et al.* 2002).

1.B.2. The exclusive PPPEW repeat and other distinguishing characteristics found in TBC2p

Auchincloss and colleagues (2002) found that certain amino acids were repeated consecutively in strings (hereafter, a string of amino acids is defined as a subsequence of a protein sequence comprised of only one amino acid) of 3-11 residues, such as alanines, glutamines and threonines, especially in the amino (N) and carboxyl (C)-termini of the TBC2 amino acid sequence. Boudreau's (2000) and Vaistij's studies (2000) found that strings of alanines or glutamines in the N and C-termini are characteristic of many nucleus-encoded proteins in *C. reinhardtii.*

The mid-portion of the TBC2 amino acid sequence contains seven copies of a novel degenerate 38-40 amino acid PPPEW repeat with little to no separation by other amino acids (see Fig. 6) (Auchincloss *et al.* 2002). Two other PPPEW amino acid repeats are located within the C-terminal end and flank each other (see Fig. 6) (Auchincloss *et al.* 2002). The amino acids of the consensus were determined by an identity of 44%, the repeat determined using the program GCG (see Fig. 6) (Auchincloss *et al.* 2002; A. Auchincloss, email comm.). It was also noted that the C-terminal region of the PPPEW consensus sequence is more highly conserved than the N-terminal region (see Fig. 6) (Auchincloss *et al.* 2002). The first P (amino acid 33) and W (amino acid 37) of the PPPEW sequence show high conservation of 100% identity (see Fig. 6) (Auchincloss *et al.* 2002). PPPEW repeats are thought to be necessary in a functional role since cDNA containing a 179 nucleotide deletion in a PPPEW-prevalent region could

MLPLEHKASGRVQATGRGVRASVELSSVLPQQRAAQLQHQKCNTGARLGRDPRRGVDAERTLVCTAATTASVP
STSGASPSGSQLSSKALRPRRFSAPVIARLLRSTTTVQELADLVQQQSLYMDSSHVGIAMLHLALLVSRAEQQAA
AQLQLQLAAKQAATRRAGSGASTSGRARGWGSGPGRNGSGSSSVSVNGSGSSSNGSSSSSSSLAMGMQLSM
ASIGDDVVSGVNAGPVPSGGADALLDLEMSSILDDDDGAGARQLQQMSDDLAAGLEAAATTTAAPEAGVAAAGG
TGAGAAADAAASSSAPSLVAAAAAAAAAAASPASSPDVARTLRTLLSRAFSLGLDSLSGPQLAAVFTGLAVLRRP
RQQQQQQQQAGAAGAGANAGAGGVGGVGVSAGDRLVAEQLLAAMGPKLYECRPQDLANTLASVALLGLPPDA
DLRTSFYAAVRQQQRRFGPRELATTLWAYGAMGTYVQEDAVQLVLELSRARLTSFSPLQLAKAVQGLAALRYRP
SPEWVEAYCSVLRPALRRMSSRELCAVLLALASLQVGLDGGTRAALLVHTFSGPLPGMAPGEVALSLWALGRLS
AVDMDLPALIDLDMSGRVLDLTSRLLAAGGFSGGELQQLLEGLTRLALQPPLEWMQAFVAALQPQLDKLDAQQL
AGVLNSLAAQQYRPQPQMQEVVLAATQANMKQLLADTTCSAALLTALRRLNIEPPPGWVGALLEESRSALKNRC
TDLHLANLAGSLAAWGVRPDGRWAARLMWRSQVLMNEDRMSPRALVALLQAMVSLGLSPNPVWTQLCLGAA
VRRASQPAFEPHHYGTLMASLHALGIQPPQEWLTRMLLSTYRCWDRFSVTHWSSLLPALVLLKARPPREWLRR
FEATSAARLADCSALQLLTLAVSLAQLHQLHAAGAVADTPLLLPGAAAAAAAAAPAGASSAAAAGDSPAALSAVPA
AAGDGALVPSFMSIDDDGTAAVAAAATALAAAEPAAHAATSTTTATAVAHPQPQLLPQAQALPQPGPEWQAAW
WAASTRLLLRVRYAPSELVLTAGWLCSLGLRPHPEWLQACAEVAARYSKVMDAAERQQLAAAVAPLALEAVAP
PSAPPAGAASTAH

Fig. 6. PPPEW repeats of the TBC2 amino acid sequence. PPPEW repeats are enclosed in shadowed boxes, repeat sequences flanking each other are highlighted in grey (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]; Auchincloss *et al.* 2002). (These sequence data were produced by the US Department of Energy Joint Genome Institute, http://www.jgi.doe.gov/ and are provided for use in this publication/correspondence only.)

not rescue the *tbc2-F64* mutant strain (Auchincloss *et al.* 2002).

Since the PPPEW consensus characterizes the TBC2 amino acid sequence, it may also be used to identify paralogous sequences which may possess similar functional roles that require the PPPEW repeat (see Fig. 6). The bioinformatics segment will discuss this possibility and the approaches taken to determine this link.

1.B.3. PPR repeats and their relevance to TBC2p

The Crp1 protein, which the TBC2 protein is similar to, contains the 35 amino acid degenerate repeat called the pentatricopeptide repeat (PPR) (see Fig. 7) (Williams and Barkan 2003). This repeat is of great relevance to the topic of TBC2p because Crp1p is known to be involved in protein-protein and protein-RNA interactions in RNA maturation and translational machinery due to the structure that it generates (Yamazaki *et al.* 2004). Since TBC2p, a translational activator itself, has a homologous region to Crp1p in the region of the PPR repeat, then the TBC2 amino acid sequence or even its PPPEW repeat sequence, may generate similar secondary structures required for interaction in the 400 kDa complex found by Auchincloss and colleagues (2002).

PPR repeats, part of the helical repeat family, exist in proteins in copies between 5 and 15 and form helical hairpins in the form of a superhelix (see Fig. 7) (Williams and Barkan 2003). This structure allows the formation of a tunnel with the inner side chains being solely hydrophilic with the bottom of the groove being positively charged (Small and Peeters 2000).

PPR proteins are prevalent in the *A. thaliana* genome (Lurin *et al.* 2004).

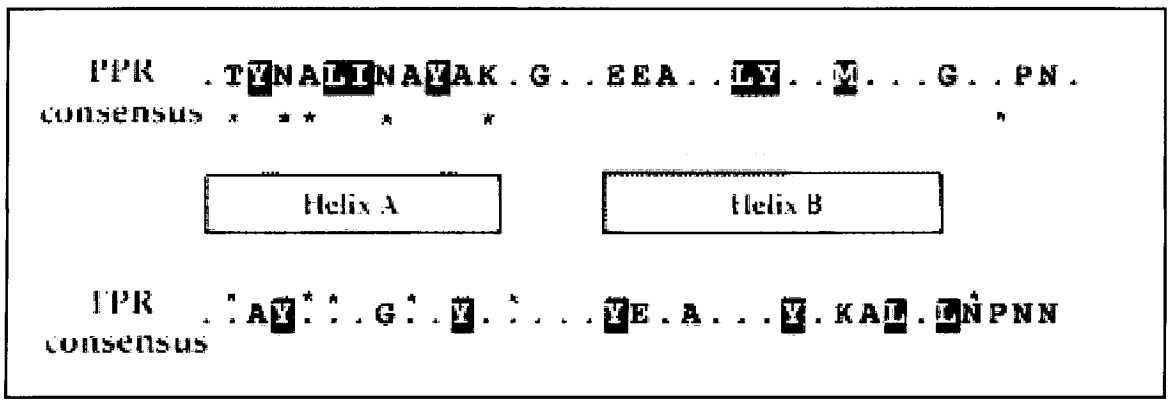

Fig. 7. An alignment of the PPR and TPR consensus sequences. PPR and TPR consensus sequences share very similar secondary structure projections due to the relative number of identities. Characteristics of the amino acids in these sequences: hydrophobic: Y, L, I, M; hydrophilic: T, N, K, E; small side chain: A, G; proline: P. Stars indicate the amino acid side chains predicted to extend into the groove formed by the anti-parallel α-helices. (Reprinted from Trends in Biochemical Sciences, vol. 25, Small, I.D. and N. Peeters. "The PPR motif-a TPR-related motif prevalent in plant organellar proteins." p. 46-47, © 2000 with permission from Elsevier).

6372 PPR repeats within the genome belong to 526 clusters which singularly exist within separate genes (Lurin *et al.* 2004).

The Predator program used showed that the probability that the PPR proteins in *A. thaliana* were targeted to the chloroplast and mitochondria was 19% and 54%, respectively (Lurin *et al.* 2004). Subcellular localization of certain PPR proteins using PPR-green fluorescent protein or PPR DS-Red2 fluorescent fusion protein resulted in 18 PPR proteins visualized as targeted to the mitochondria and 8 were targeted to the plastids (Lurin *et al.* 2004). Similar to Cya5p and PET309p which also function in the mitochondria at the level of mRNA translation (Coffin *et al.* 1997), localization through fluorescent fusion proteins support the theory that PPR proteins have organellar function (Lurin *et al.* 2004).

The findings above are significant considering that PPR proteins appear to target chloroplasts or mitochondria like TBC2p (Lurin *et al.* 2004). Therefore, PPPEW repeats in the TBC2 amino acid sequence may generate secondary structures similar to those generated by PPRs required for similar localization.

1.B.4. The TPR repeat and its similarity to the PPR repeat

Blatch and Lässle (1999) had claimed that proteins involved in post-transcriptional steps such as RNA-editing often contain tandem arrays of degenerate TPR repeats (see Fig. 7). Tetratricopeptide repeats (TPRs) are distinct from PPPEW and PPR repeats, belonging to the helical repeat family (see Fig. 7) (Yamazaki *et al.* 2004). TPRs generate similar secondary structures to the PPR repeat such as two anti-parallel α-helices caused by a β-turn, fulfilling

similar functional roles (see Fig. 7) (Yamazaki *et al.* 2004). Comparing the TPRs to the PPRs gives some insight about how helical repeats are generally structured and the functions they tend to fulfill. α-helices comprise 50% of the TPR structure, determined through circular dichroism studies (see Fig. 7) (Hirano *et al.* 1990). TPR's antiparallel right-handed α-helices called A and B generate a central groove containing varying amino acids per protein, due to the fact that different ligands are bound (Blatch and Lässle 1999; Small and Peeters 2000). α-helix A contains consensus residues 2, 7, 8, 10 and 11, whereas helix B contains consensus residues 20, 24 and 27 (Blatch and Lässle 1999). α-helices A and B form an angle of 24° and, due to their proximity, generate an amphipathic channel composed of amino acids from both helix A and B (Blatch and Lässle 1999). Das and colleagues (1998) wrote that TPR repeats appear in clusters of seven. Das and colleagues (1998) also claimed that the superhelix formed has a pitch of 60Å, a width of 42° with helix A on the inside, helix B on the outside. Blatch and Lässle (1999) proposed that 5-6 TPR motifs are predicted to interact with the target protein.

TPR proteins occur in multiprotein complexes (Blatch and Lässle 1999). They are often required for the proper functioning of chaperone, cell cycle transcription processes, as well as protein transport complexes (Blatch and Lässle 1999).

Boudreau and colleagues (2000) as well as Vaistij and colleagues (2000) found that two nucleus-encoded proteins, Nac2p and Mbb1p, function in stably accumulating their targeted mRNAs, *psbB* and *psbD* respectively, resulting in

large RNA-protein complexes. This reaffirms the role of TPRs in protein-protein interactions, required in multisubunit complexes.

Since TBC2p shows sequence similarity to the Crp1 protein sequence in the region of the PPR repeat, then it indirectly should be similar to the TPR repeat. Therefore, it must be noted that PPR and TPR helical repeats, which share similar functions and protein structure, assign a hypothetical structure and role to the TBC2 protein as a protein which mediates protein-protein or protein-RNA interactions in post-transcriptional processes.

1.B.5. Crp1, a protein with significant homology to TBC2p

Mentioned previously, Auchincloss and colleagues (2002) demonstrated that the mid-portion of the TBC2 amino acid sequence composed of seven PPPEW repeats has slight homology to the PPR-prevalent region of the Crp1 protein sequence. Crp1p, a nucleus-encoded protein in maize, is a translational activator protein for *petA* and *petD* mRNAs, and also processes *petD* mRNA from a polycistronic precursor (Fisk *et al.* 1999). Crp1p shows some similarity to nucleus-encoded proteins of similar function in fungi (Fisk *et al.* 1999).

Manthey and McEwen (1995) and Coffin and colleagues (1997) found other proteins of similarity to Crp1p which, when mutated, resulted in similar phenotypes caused in the mutated *Crp1* gene. They referred to proteins Cya5p in *Neurospora crassa* (37% similarity to Crp1p) and PET309p in *S. cerevisiae* (33% similarity to Crp1p), responsible for mRNA functions and translational activation of the *coxI* mitochondrial gene (Fisk *et al.* 1999). This similarity in mutant phenotypes suggests similar mechanistic roles among all three proteins (Fisk *et*

al. 1999).

Crp1p shares similarities with TBC2p in its functional role and structurally due to the absence of transmembrane domains (Fisk *et al.* 1999). Crp1p, like TBC2p, is localized in the stroma, found to belong to a multisubunit complex not associated with chloroplast membranes or ribosomes (Fisk *et al.* 1999). Similar to TBC2p, Crp1p contains the semblance of a chloroplast transit peptide (cTP), the first 12 amino acids from the N-terminus being uncharged while the next 50 amino acids being predominantly hydroxylated and basic (rev. by Cline and Henry 1996).

Experimental evidence supports the presence of a chloroplast transit peptide within Crp1p (Fisk *et al.* 1999). A 66 kDa mature Crp1 protein in leaf tissue was detected when the protein following *in vitro* translation of *Crp1* cDNA was 5 kDa greater (Fisk *et al.* 1999). This suggests that the N-terminus in the mature protein of Crp1p had been cleaved (Fisk *et al.* 1999).

Because of the similarities listed above to TBC2p, the Crp1 protein sequence in this thesis was again compared to the TBC2p sequence in a ClustalW alignment to determine specific regions of similarity. If paralogs of TBC2p are found, the Crp1 protein sequence will be compared to them as well. If extensive similarities using ClustalW are found between Crp1p and TBC2p paralogs, this could provide some insight on the functions of these paralogs, possibly as translational activator proteins like TBC2p.

1.B.6. TBC2p, like Crp1p, has a transit peptide

Due to the fact that the TBC2 protein is nucleus-encoded and that it functions in the chloroplast, it was hypothesized that TBC2p, as Crp1p, contains a transit peptide (Auchincloss *et al.* 2002). The amino terminus of the TBC2 protein shows characteristics of a transit peptide based on the presence of an abundance of hydroxylated and basic residues (Auchincloss *et al.* 2002). The number of acidic residues was higher than what is normally expected for a cTP and PSORT (http://psort.nibb.ac.jp/) and ChloroP (http://www.cbs.dtu.dk/services/ChloroP/) programs were used to determine its localization (Auchincloss *et al.* 2002). It has been explained that lumen-targeting cTPs in higher plants often contain more than one acidic residue, the reason probably being attributed to targeting of the thylakoid membrane (Franzén *et al.* 1990). cTPs in *C. reinhardtii* resemble mitochondrial transit peptides (mTPs) in composition rather than cTPs in higher plants, observations explained by findings made by Franzén and colleagues (1990). Transit peptides are generally similar in amino acid composition (Franzén *et al.* 1990). For example, characteristics of a high arginine content (14%), the presence of serines (10%) and a lack of acidic residues in the *C. reinhardtii* cTP can also be applied to mTPs in vascular plants (Franzén *et al.* 1990).

cTPs in higher plants often form an amphiphilic β-strand next to the cleavage site of the transit peptide (Franzén *et al.* 1990). cTPs in *C. reinhardtii* are predicted to form α-helices comparable to mitochondrial presequences (Franzén *et al.* 1990).

Cleavage sites of *C. reinhardtii* cTPs have a recognition sequence of VAL-X-ALA between positions -3 to -1 (Franzén *et al.* 1990), a similar cleavage pattern to (VAL/ILE)-X-(ALA/CYS)↓-ALA noted by von Heijne in higher plants (in preparation). It was noted in the RuBisCO small subunit (SSU) protein that neither of these recognition sequences are present, resulting in the conclusion that multiple cleavage sites exist (Franzén *et al.* 1990). It was gathered from 12 cTPs in *C. reinhardtii* that the mean length of a cTP was 29 residues, similar to higher plant mTPs which are, on average, 30 residues, but differ in comparison to a mean length of 60 amino acids in cTPs of higher plants (Franzén *et al.* 1990).

1.B.7. TBC2p pertaining to the aim of this thesis

To explore the characteristics of the TBC2 protein, two approaches were taken, the first being a bioinformatics-based approach whereby various computer programs were used to find possible paralogs of the TBC2 amino acid sequence and, if successful, to analyze common characteristics. These characteristics include the presence of transit peptides, the PPPEW repeat characteristic so far only for TBC2p, similarities to other translational activators and similarities to themselves.

The second approach was to use coimmunoprecipitation to determine if TBC2p or other factors interact with the *psbC* 5'-UTR. This thesis describes the preliminary experiments.

1.C. Basis of hypothesis of the novel thylakoid biogenesis compartment

It was theorized that a novel thylakoid biogenesis compartment existed through co-translational assembly of products into complexes at specific regions of an unknown compartment, gathered from experimental data (rev. by Sato *et al.* 1999; Zerges 2000; Zerges and Rochaix 1998; Fisk *et al.* 1999; Joyard *et al.* 1998). Previous supporting evidence of this theory was that polysomes were visibly associated with inner envelope membranes, achieved through electron microscopy experiments, and that co-sedimentation of polysomes and membranes occurred on sucrose density gradients (Chua *et al.* 1976; Klein *et al.* 1988; Margulies and Michaels 1979; Margulies *et al.* 1975). Additional supporting evidence was obtained when chloroplast ribosomes from *C. reinhardtii* could be isolated along with small patches of inner envelope membrane and that these membranes had a lowered chlorophyll concentration 0.5% that of regular thylakoid membranes (Margulies and Weistrop 1980). Inner envelope membranes also had 50% the number of ribosomes normally associated with thylakoid membranes (Margulies and Weistrop 1980). The requirement for membrane-bound polysomes indicating a necessity for co-translational insertion of thylakoid proteins, possibly into a novel thylakoid biogenesis compartment, was shown in barley (Mühlbauer *et al.* 1999). This was shown when levels of RuBisCO decreased when membrane-bound polysomes were dissociated due to high salt or puromycin concentrations (Mühlbauer *et al.* 1999). A higher percentage of RuBisCO proteins existed in the membrane fraction rather than

soluble fraction, shown by immunoprecipitation (Mühlbauer *et al.* 1999). This reflects the requirement of membrane-bound polysomes in the translation of both soluble and membrane-bound proteins (Mühlbauer *et al.* 1999). Polypeptide composition of these membranes varied from the expected composition of thylakoid membranes (rev. by Zerges 2004). A *psbA* mRNA splicing protein RB47 was found to be associated with thylakoid membranes distinct from regular thylakoid membranes in terms of buoyant density and pigment composition (Perron *et al.* 1999; Zerges and Rochaix 1998). Additional evidence shown in prokaryotic cyanobacteria demonstrated the theory of complex assembly in the plasma membranes (Zak *et al.* 2001). Zak and colleagues (2001) proposed that reaction core assembly of PSI and PSII originates in the plasma membrane and that these membranes fuse with thylakoid membranes through vesicle membrane transport. This was done by phase partitioning of membranes and verification through immunological detection (Zak *et al.* 2001). These findings support the theory that transcriptional and translational events may occur within specialized regions of an unknown compartment responsible for thylakoid biogenesis.

Thylakoid biogenesis requires gene products from two genetic systems, those of the chloroplast and nucleus (rev. by Cline and Henry 1996). This theory includes the role of the TBC factors described in Section 1.B of the Introduction. These nucleus-encoded proteins are imported at specialized entry points by translocation machinery (rev. by Cline and Henry 1996). It is proposed that translation of thylakoid proteins accompanies the integration of chlorophyll

anabolized in these membranes, as well as the integration of nucleus-encoded proteins which are combined synchronously, a mechanism reviewed by Joyard and colleagues (1998).

Westphal and colleagues (2001, 2003) proposed that if a novel thylakoid biogenesis compartment does exist that these membranes and membrane proteins are transported and fused into mature thylakoid membrane proteins through vesicles. This theory in supported in parallel by findings made by Kroll and colleagues (2001) whereby the protein VIPP1 functions in vesicle formation and the maintenance of thylakoid membranes in the stroma of *A. thaliana*. It was found by Hoober and colleagues (1998) through electron microscopy the occurrence of inner envelope membrane budding and the formation of vesicles in developing thylakoid membranes of *C. reinhardtii.*

Based on the above observations and findings, pulse-labeling experiments were done with $[^{35}S]H_2SO_4$ using light and dark-grown cultures to observe whether darkness blocks the maturation of thylakoids in the novel membrane compartment (Zerges, pers. comm.). Novel thylakoid membrane compartment proteins are theorized to appear as aggregates at the top of the gel; an increase in the intensity of the autoradiographs, specifically in aggregate formation, would demonstrate blockage in maturation of the inner envelope membranes in the dark caused by inhibited vesicle formation (Zerges, pers. comm.).

1.D. Prior findings supporting the theory of localization of the chloroplast genome

It is unknown whether or not chloroplast genomic DNA (cpDNA) is localized to some part of the chloroplast or whether it is free in the stroma of the plastid. A review by Kuroiwa (1991) reported that cpDNA in *C. reinhardtii* forms a large protein-DNA complex called the plastid nucleoid, present in 10 or more copies. It had been claimed that the configuration of this plastid nucleoid undergoes changes during development (Miyamura *et al.* 1986; Sato *et al.* 1993, 1997; Sodmergen *et al.* 1989).

It was noted that rapid replication of the plastid nucleoids occurs during plastid development, localizing close to the membrane (Sato *et al.* 859). Herrmann and Kowallik (1970) demonstrated using electron microscopy that cpDNA appears to adhere to the envelope membrane of developing plastids, supporting the cpDNA localization theory. This coincides with experimental data obtained by DuBell and Mullet (1995a, 1995b) that replication and transcription rates peak in immature chloroplasts when the rate of development is at its highest.

The most crucial experimental results suggesting that the plastid nucleoid binds to the envelope membrane has been found in peas (Sato *et al.* 1998). A lithium dodecyl sulfate-PAGE gel determined the PEND protein to be a 130 kDa polypeptide (Sato *et al.* 1998). Sato and colleagues (1993, 1995) found that PEND, an acronym for Plastid Envelope DNA-binding protein, interacts with

cpDNA at specific genomic sequences including midway down the *petA* gene (apocytochrome f), two locations within the *rpoC2* gene (β subunit of RNA polymerase) and to the *psbM* gene which codes for the PSII-M polypeptide. The necessity of PEND was proposed to alter replication, segregation and possibly transcription of cpDNA (Sato *et al.* 1998).

If it can be verified that the PEND protein is a DNA-anchoring protein in the plastid, then localization of the chloroplast genome in *C. reinhardtii* may occur due to a homolog of PEND, resulting in its localization.

In this thesis, an attempt was made to localize the chloroplast genome by determining whether DNA, found in low density membranes (LDM2) dissociated in the absence of Mg^{2+}, was nuclear or chloroplast genomic DNA. This was done by a restriction digest which would result in few distinct fragments if the DNA was chloroplast genomic DNA or many closely resembling fragment sizes due to the relatively large size of nuclear genomic DNA. If cpDNA was found bound to the thylakoid membranes after a Percoll gradient, then this could support the theory that cpDNA is bound and localized to the thylakoid membranes.

2.1. Bioinformatics analysis of TBC2p and its paralogs

2.1.1 Bioinformatics Programs

Various computer programs were used to analyze TBC2p and its paralogous proteins predicted from the *C. reinhardtii* genomic sequence.

Version II of the DOE Joint Genome Institute (JGI) BLAST program (http://genome.jgi-psf.org/cgi-bin/runAlignment?db=chlre2&advanced=1) was used to BLAST TBC2p against JGI's sequenced *C. reinhardtii* nuclear genome. The genome was determined using the whole genome shotgun strategy using paired-end sequencing reads, and, as a result, 1.8 million reads were assembled into 3211 scaffolds comprising 125 Mbp (DOE JGI *C. reinhardtii* V.2.0 DB). A scaffold is defined as a sequence drawn from an arrangement of contigs as they appear in the genome sequence (LaskerFoundation). However, these scaffolds do not necessarily follow each other and may overlap (LaskerFoundation). The draft of this sequence is pre-publication and may contain sequencing errors (L. Peeters, email comm.). The DOE JGI BLAST performed was a translated one of the TBC2 amino acid sequence (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]).

The program greenGenie ORF finder (http://www.cse.ucsc.edu/~dkulp/cgi-bin/greenGenie) is a program used to predict possible start and stop codons which encompass the open reading frame (ORF), and exon sequences. greenGenie ORF Finder was used to predict possible start and stop codons, as well as exon sequences of *TBC2*'s putative paralogs, *TBC2A* and *TBC2B* (DOE

JGI tBLASTn).

Version 2.04 of the NCBI Conserved Domain Reverse Position Specific BLAST (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) is comprised of a conserved domain database with multiple sequence alignments for domains and proteins. It was used to determine pFAMs, COGs or SMARTs of similarity (NCBI CDD V.2.04 RPS BLAST). pFAM is an acronym for "**p**rotein **FAM**ilies database of alignments and HMMs". pFAMs, according to the Sanger Institute, are a grouping of multiple sequence alignments and hidden Markov models comprised of numerous protein domains and families. pFAMs are divided into sets A and B, A containing 7868 families and B containing a large number of small families from the PRODOM database (Sanger Institute). COG is an acronym for **C**lusters of **O**rthologous **G**roups of proteins determined from both prokaryotes and eukaryotes and, according to Tatusov and colleagues (2003), is comprised of 138458 proteins forming 4873 COGs which belong to 75% of 185505 predicted proteins. The **S**imple **M**olecular **A**rchitecture **R**esearch **T**ool (SMART) program distinguishes and annotates signaling domain sequences (Shultz *et al.* 1998). SMART is also a predictor of molecular structures of sequences and genomes (Schultz *et al.* 1998).

ClustalW (http://www.ebi.ac.uk/clustalw) performs multiple sequence alignments of DNA and protein sequences and determines any existing homologies. This program aligns these sequences in a hierarchal order so that identities, similarities and differences can be noted. This program was used to align TBC2p (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]), its

paralogs (DOE JGI tBLASTn), and the Crp1 protein (NCBI [accession AAC25599, version AAC25599.1 gi: 3289002]). It was also used to align the PPPEW repeats existing in the TBC2 protein (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) and its paralogs (DOE JGI tBLASTn) to note similarities and identities (ClustalW).

TargetP 1.1 (http://www.cbs.dtu.dk/services/TargetP) is used to predict subcellular location of proteins through comparisons made with known cTPs, signal peptides and mTPs. This program was used to predict the subcellular target of TBC2p (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]), TBC2Ap and TBC2Bp (DOE JGI tBLASTn; TargetP 1.1).

ChloroP 1.1 (http://www.cbs.dtu.dk/services/ChloroP) predicts the existence of cTPs in protein sequences as well as cleavage sequences. This program was used to determine whether or not TBC2p (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) and its paralogs (DOE JGI tBLASTn) had potential targeting domains (ChloroP 1.1).

PROSPECT-PSPP (http://csbl.bmb.uga.edu/protein_pipeline/input_form.php) is a program used to predict protein structure by integrating varying computational tools discovered previously. This program can assess and predict protein characteristics such as targeting peptides (restricted to murine and human proteins only), protein type prediction, the number of membrane domains, secondary structure, folding and atomic structure model generation (Guo *et al.* 2004). PROPSECT-PSPP was used for TBC2p (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]),

TBC2Ap and TBC2Bp (DOE JGI tBLASTn).

Apart from the programs listed above, stretches of amino acids characteristic of targeting proteins were also examined.

2.2. Experimental approach taken towards determining mode of TBC2p interaction: Materials and Methods

2.2.1. Strains (*C. reinhardtii* and *E. coli*) and plasmids

Five *C. reinhardtii* strains were used in this section, one being the *tbc2-F64;cw15;arg7* strain which has a PSII-deficient phenotype, a nuclear genotype of *TBC1*, *tbc2-F64* and *TBC3* and a chloroplast genotype of *psbC-WT*. The second strain used was *CC503;cw92* (mt^+) w hich is cell wall-deficient (*Chlamydomonas* Genetics Centre, <www.chlamy.org/strains/core.html>). Other strains included *F14* (mt^-), mutated at the *Tab2* nuclear locus and is acetate-requiring, as well as the *cw15wt* strain, mutated at the *cw15* nuclear locus causing cell wall-deficiency (Harris 1989). *Raa2:HA* was also used, mutated at the *raa2-A18* nuclear locus and transformed with a construct containing *Raa2* cDNA with a hemagglutinin epitope (Fischer and Rochaix 2001; Perron *et al.* 1999).

The *Escherichia coli* strain *MC1061* has the genotype as follows: *hsdR mcrB araD139 Δ(araABC-leu)7697 ΔlacX74 galU galK rpsL thi supE44 hsdR endA1 pro thi* (Sambrook and Russell 2001). Meissner and colleagues (1987) wrote that it is a *mcrB* strain used for λORF8 primary libraries.

Plasmids obtained by midipreparation are listed in Table 1.

Table 1. Plasmids.

Plasmids	Description	Reference
pARG7.8	Argininosuccinate lyase gene (*ASL*) spanning a region less than 7.8 kb	Debuchy *et al.* 1989
plate 64, well F5	The cosmid which complemented *TBC2*, originates from well F5, plate 64	Auchincloss *et al.* 2002
KBam10 in pKS$^-$ (ie. genomic)	~9.5 kb fragment of genomic DNA which complements the *tbc2-F64* mutant phenotype and is within *Bam*HI cutting sites, cloned in pKS$^-$	A. Auchincloss
KBam10 in pKS$^-$ (delete *Hind*III, *Sal*I, *Not*I and *Sac*I)	~9.5 kb fragment of genomic DNA which complements the *tbc2-F64* mutant phenotype and is within *Bam*HI cutting sites, cloned in pKS$^-$ with *Hind*III, *Sal*I, *Not*I and *Sac*I polylinker sites deleted	Auchincloss *et al.* 2002
*Sal*I site @ stop of 3' end of *TBC2*	3' end of the genomic DNA, from *Not*I ot the *Bam*HI end, with the *Sal*I site present at the stop codon	Auchincloss *et al.* 2002
KBam10 in midi gene in *Sac*I	Contains a "midi-gene" made by fusing the genomic 5' DNA, up until the *Sac*I site, to the cDNA, ie. from *Sac*I to the 3' end it is cDNA	Auchincloss *et al.* 2002
KBam10 in midi gene @ *Hind*III	Contains another "midi-gene" made by fusing the genomic DNA up to the first *Hind*III site to the only *Hind*III site in the cDNA	A. Auchincloss
cDNA 1D in pKS$^-$	cDNA which saves (1D)	Auchincloss *et al.* 2002
cDNA 4K in pKS$^-$	cDNA which saves (4C)	Auchincloss *et al.* 2002
TBC2 cDNA *Sac*I @ stop codon	cDNA from *Sac*I to an internal *Sph*I site cloned into pET16b for over-expression purposes	A. Auchincloss
HA @ stop in KBam10 (ie. genomic)	Full genomic clone with *Sal*I site at the stop codon	Auchincloss *et al.* 2002
HA @ stop at 3' end in *TBC2*	3' end of the genomic DNA, from *Not*I to *Bam*HI end, with the *Sal*I site at the stop codon with an HA epitope at the stop site	A. Auchincloss
KBam10, *Sac*I @ stop codon	Full KBam10 clone with the HA epitope at the stop codon	Auchincloss *et al.* 2002

2.2.2. Media and growth conditions

Luria-Bertani (LB) medium for *E. coli* was used with a concentration of 50 µg/ml ampicillin for ampicillin-resistant mutants (Sambrook and Russell 2001). LB medium was solidified with 15 g/L agar. *E. coli* cells were grown at 37°C with an agitation speed depending on the protocol requirement. Tris-acetate-phosphate medium (TAP) was used to grow *C. reinhardtii* strains, under light (~100 $\mu Em^{-2}s^{-1}$) when required, at 24°C, with agitation at 140 rpm (Gorman and Levine 1965). A final concentration of 55 µg/ml of arginine was used for *arg7* mutants and PSII-deficient strains were grown in reduced-light conditions. High salt minimal medium (HSM) according to Sueoka (1960) was used to select for photosynthetic-deficient (*ac*$^+$) mutants. 15 g/L agar was added to TAP or HSM when required.

2.2.3. Genetic and molecular techniques

pARG7.8 transformation into *MC1061* was done by electroporation (Sambrook and Russell 2001). Plasmid DNA maxipreparations of these transformants were done by alkaline lysis with SDS (Sambrook and Russell 2001). The pARG7.8 preparation was purified by PEG and ethanol precipitated (Sambrook and Russell 2001). The remaining constructs in Table 1, already in *E. coli*, were prepared by midipreparation and ethanol precipitated (Sambrook and Russell 2001).

The "HA @ stop in KBam10 (ie. genomic)" and "KBam10 in pKS$^-$ (ie. genomic)" plasmids used in the *C. reinhardtii* transformation were digested with restriction endonucleases *Nhe*I (Amersham) and *Eco*RV (Amersham) to verify

the presence of a triple hemagglutinin epitope through a 700 bp fragment.

The *tbc2-F64;cw15;arg7 C. reinhardtii* strain was transformed with the "HA @ stop in KBam10 (ie. genomic)" plasmid using the PEG glass bead method (Kindle 1991). Colonies which were saved (ac^+) in light conditions on HSM and arginine plates were used later in a Western blot.

2.2.4. Polymerase Chain Reactions

Genomic DNA from *cw15wt C. reinhardtii* cells was extracted using the CTAB DNA extraction protocol according to G. Glöckner. PCR reactions had a final volume of 50 µl. Table 2 lists the primers designed for *TBC2A* and *TBC2B* amplification.

Table 2. Primers Used in PCR Reactions (Qiagen).

Primer name	Primer sequence
TBC2A forward	5': 5'-GCGTTGGGGGGCTGTGCTAC-3'
TBC2A reverse	3': 5'-ACTGCTACTGCTCCTGCGGC-3'
TBC2B forward	5': 5'-GCTGGTGCTGATGGGTGTGC-3'
TBC2B reverse	3': 5'-CAGGGGTCGTCGGGCAGGTG-3'

The following components listed in Table 3 were added to the PCR reaction amplifying *TBC2A*.

Table 3. PCR Components for *TBC2A* Amplification.

Final conc.	Component added
1 µg	Template DNA (1 µg)
1-fold	10* PCR buffer w/o $MgCl_2$
50 µM each	20 mM solution of 4 dNTPs
1.4 µM	10 µM sense primer (70 pmoles)
1.4 µM	10 µM antisense primer (70 pmoles)
1 U/50 µl	1 unit thermostable DNA polymerase
1.6 mM	20 mM $MgCl_2$
-	dH_20
50 µl	**Final volume**

Table 4 outlines the cycle used on the GeneAmp PCR system 2700.

Table 4. Cycle Used for *TBC2A* Amplification.

Step	Minutes (min)	Temperature (°C)	
1	4	94	
2	1	95	Repeat steps 2-4 30-fold
3	1	60	
4	1.5	72	
5	5	72	
6	Infinite	4	

For the *TBC2B* paralog, Table 5 describes the conditions used:

Table 5. PCR Components for *TBC2B* Amplification.

Final conc.	Component added
1 µg	Template DNA (1 µg)
1-fold	10* PCR buffer w/o $MgCl_2$
50 µM each	20 mM solution of 4 dNTPs
1.4 µM	10 µM sense primer (70 pmoles)
1.4 µM	10 µM antisense primer (70 pmoles)
1 U/50 µl	1 unit thermostable DNA polymerase
1.2 mM	20 mM $MgCl_2$
5% (v/v)	DMSO
-	dH_20
50 µl	**Final volume**

Table 6 describes the cycle used on the GeneAmp PCR system 2700.

Table 6. Cycle Used for *TBC2B* Amplification.

Step	Minutes (min)	Temperature (°C)	
1	4	94	
2	1	95	Repeat steps 2-4 30-fold
3	1	61.7	
4	1	72	
5	5	72	
6	Infinite	4	

The PCR samples were later purified from 0.8% agarose gels using two methods, the first according to the Qiagen Qiaquick purification kit, the second according to the Qiaquick gel extraction kit.

2.2.5. Western blot analysis of TBC2:HA in soluble protein extracts

Seven transformants of the *tbc2-F64;cw15;arg7* strain with the TBC2:HA construct were used for the Western blot experiment. *tbc2-F64;cw15;arg7* and *Raa2:HA* strains were used as negative and positive controls, respectively. Proteins were resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred using 3MM Whatman blotting paper to nitrocellulose for antibody treatment. Covance-purified murine antibody (mono HA.11) was used as a primary antibody, followed by horseradish peroxidase (HRP)-anti-mouse antibody treatment. Full details of the experiment are outlined in Appendix A.1.

2.2.6. Generation of *Tab2;cw92* mutants

Tab2:HA protein was required as a positive control for the coimmunoprecipitation experiment. Therefore, *C. reinhardtii* had to be mutated at both the *Tab2* and *cw92* loci so that a Tab2:HA construct could be transformed into a cell wall-deficient strain and expressed (see Fig. 8). Therefore, a *Tab2* mutant strain, *F14* (mt^-), was crossed with the *CC503;cw92* (mt^+) strain. After the mating period, the total culture was plated onto 4% TAP plates overnight in light conditions and kept in darkness for 5 subsequent days. Plates were unwrapped and allowed to germinate and sporulate so that zygospores could form. One double mutant among the progeny was found after treatment with 4% SDS resulted in its lysis, indicating that it was cell wall-deficient. Preparations were done to transform this strain with the Tab2:HA construct (see Fig. 8).

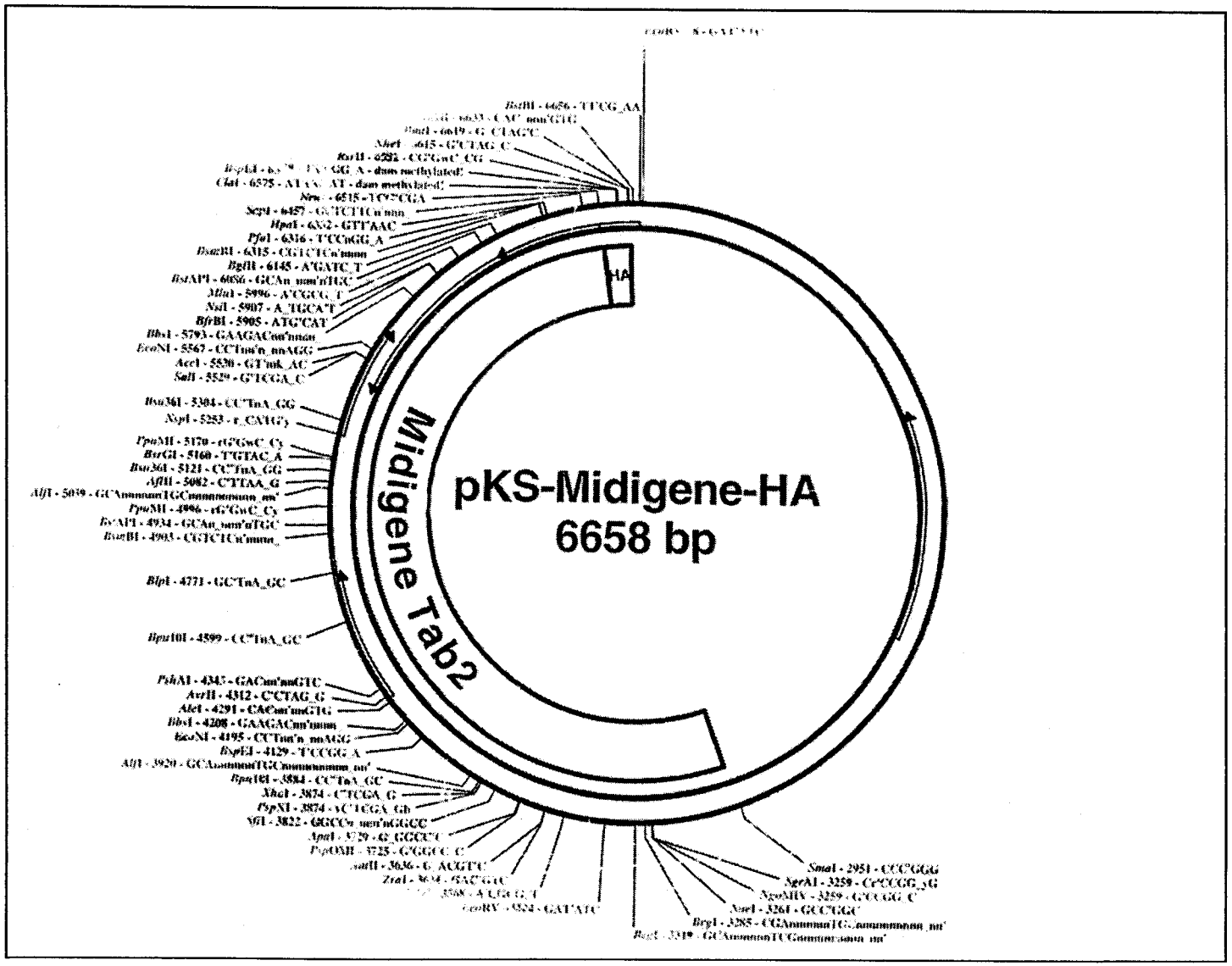

Fig. 8. *Tab2:HA* in the pKS backbone. The Tab2:HA construct used in this thesis had a pBS backbone (not shown). The restriction sites of the *Tab2:HA* midigene, however, are equivalent (AcaClone Software) (Figure kindly provided by A. Auchincloss).

2.2.7. Initial steps taken in coimmunoprecipitation

Starter cultures of *cw15wt*, *tbc2-F64;cw15;arg7* (negative control), *tbc2-F64;cw15;arg7* transformed with the TBC2:HA construct and *Raa2:HA* strains were grown to be sonicated. Full details of the protocol used can be found in Appendix A.2.

2.3. Experimental approach in determining the existence of a novel thylakoid biogenesis compartment

2.3.1. Pulse-labeling and autoradiography

The purpose of this experiment was to analyze differences in levels of newly-synthesized thylakoid proteins of inner envelope membranes grown in light and dark conditions. This experiment was conducted by pulse-labeling using $[^{35}S]H_2SO_4$, SDS-PAGE and autoradiography. Integral membrane proteins of the inner envelope membrane are known to be highly prone to aggregation, while proteins in the thylakoid membranes are less so. A blockage in maturation of inner envelope membranes to thylakoid membranes in the dark would result in an increase in aggregates due to inhibited vesicle formation.

Details of the pulse-labeling experiment are outlined in Appendix A.3.

Table 7 outlines the *C. reinhardtii* strains used in the pulse-labeling experiments along with their phenotypes and nuclear and chloroplast genotypes.

Table 7. *C. reinhardtii* strains used for pulse-labeling with [^{35}S]H_2SO_4.

		Genotype	
Strain	**Phenotype**	**Nuclear**	**Chloroplast**
WT 4A	Wild-type	*WT*	*WT*
F34.3+	PSII deficient	*tbc1-F34, TBC2, TBC3*	*psbC-WT*
CC503;cw92	Cell wall-deficient	*cw92*	*WT*
FuD34.3	PSII-deficient	*TBC1, TBC2, TBC3*	*psbC-FuD34*
F34sul	Partially-active PSII	*tbc1-F34, TBC2, TBC3*	*psbC-F34sul*
RB1.4A	Partially-active PSII	*TBC1, TBC2, tbc3-rb1*	*psbC-FuD34*
CC2518;FuD34	PSII-deficient	*WT*	*psbC-FuD34*
RB1-POAΔ166-222psbCwtTBCG3-tb1	Partially-deficient PSII	*TBC1, TBC2, tbc3-rb1*	(*psbC*(Δ166-222)-*aadA-rbcL*)-atpB-INT

2.4. Experimental approach taken towards determining the localization of the chloroplast genome in *C. reinhardtii*

2.4.1. Restriction digestion of unknown DNA associated with LDM2

LDMs after the second centrifugation in the absence of Mg^{2+} isolated from a Percoll gradient were subjected to DNA extraction to determine the localization of the chloroplast genome and its possible association with the thylakoid membrane. Proteins and lipids were extracted from nucleic acids with phenol:chloroform:isoamyl alcohol twice and nucleic acids were precipitated from the aqueous phase in a standard ethanol precipitation. The DNA sample was digested with *Eco*RI (Amersham) and *Bam*HI (Amersham) restriction endonucleases to test for the pattern of digestion which would determine whether the DNA was nuclear or chloroplast genomic. Samples were run alongside a 1 kb ladder (MBI, Fermentas) in a 0.8% agarose gel. The gel was stained by ethidium bromide and exposed using a UV transilluminator.

3.1. Bioinformatics analysis of TBC2p and its paralogs

Based on the subjects evaluated in the Chapter 1 regarding the TBC factors, it was decided to discuss the following topics determining additional characteristics of TBC2p and its existing paralogs:

-Summary of findings using bioinformatics analysis

-DOE JGI *C. reinhardtii* Genome Database Version 2 Translated BLAST of *TBC2* cDNA

-Comparisons of TBC2p, TBC2Ap, TBC2Bp and Crp1p sequences

-Determination of the presence of transit peptides and organellar localization of TBC2p and its paralogs

-TBC2, TBC2A and TBC2B amino acid sequences, their characteristic PPPEW repeat and the prevalence of amino acid strings

-Predicting the function of TBC2p and its paralogs through analysis of conserved domains

3.1.1. Summary of findings using bioinformatics analysis

The following segment summarizes findings using bioinformatics analysis, which can also be visualized by observing Figures 9-11.

The amino acid sequence of TBC2p used in the bioinformatics segment was obtained from the NCBI database (see Fig. 9) ([locus CAD20887, version CAD20887.1 gi: 22129636]; Auchincloss *et al.* 2002). Auchincloss and colleagues (2002) determined this sequence to be 1115 amino acids in length, predicted from its ORF determined through its cDNA (see Fig. 9). TBC2p

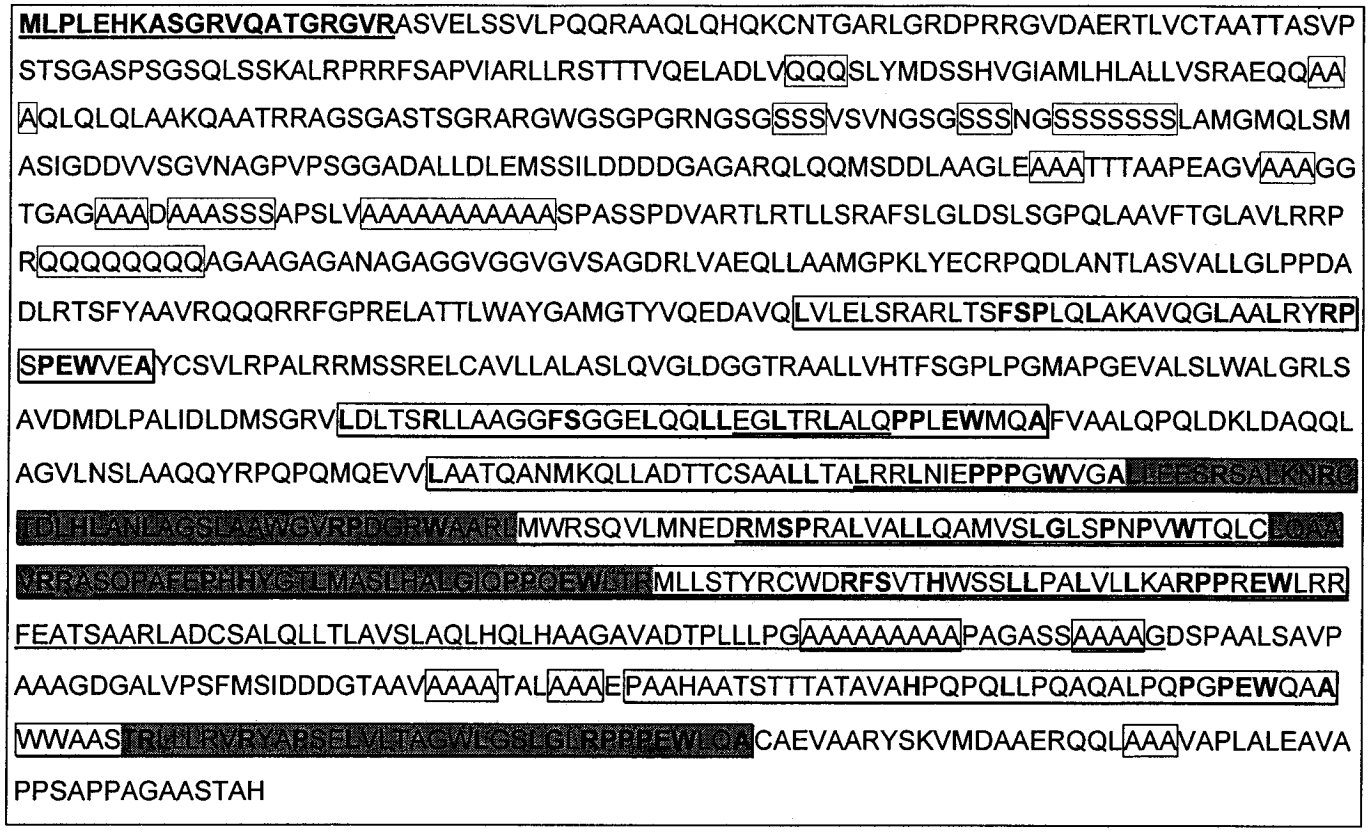

Fig. 9. The TBC2 amino acid sequence. PPPEW repeats are enclosed in shadowed boxes, repeat sequences flanking each other are highlighted in grey. Amino acids matching the consensus are bolded. A, S and Q strings are bordered with a simple box. The N-terminal targeting domain predicted by ChloroP 1.1 is underlined and bolded. The region of shared paralogy with the TBC2A and TBC2B amino acid sequences has a simple underline (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]; Auchincloss *et al.* 2002). (These sequence data were produced by the US Department of Energy Joint Genome Institute, http://www.jgi.doe.gov/ and are provided for use in this publication/correspondence only.)

contains nine novel degenerate PPPEW repeats as well as strings of alanines and glutamines (see Figs. 6 and 9) (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]; Auchincloss *et al.* 2002).

An NCBI pairwise BLAST of the *TBC2* cDNA sequence (NCBI [accession AJ427966, version AJ427966.1 gi: 22129635]) versus its genomic sequence (DOE JGI *C. reinhardtii* V.2.0 Nuclear Genome Database [scaffold 8: nucleotide 1172024-1179154]) with an expect value fixed at e^{-5} resulted in seven regions of homology (see Fig. 12, scaffold 8). This outcome was not comparable to a translated BLAST of the TBC2p sequence (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) in a DOE JGI *C. reinhardtii* V.2.0 BLAST Alignment Search which yielded 16 regions of homology (see Fig. 12, scaffold 8). The additional gaps in homology in the translated BLAST of the TBC2p sequence (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) against the DOE JGI V.2.0 nuclear genome not corresponding to introns can be attributed to errors in the sequencing of the DOE JGI V.2.0 nuclear genome since it is only a pre-publication draft (L. Peeters, email comm.). Additional reasons could be attributed to the differences in stringency of the programs.

A ClustalW alignment of the TBC2p sequence derived from translated cDNA of *TBC2* (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) was made with the TBC2p sequence determined by and registered in the DOE JGI *C. reinhardtii* V.2.0 Nuclear Genome Database. Results showed practically similar regions with 100% homology and high conservation of amino acids with similar properties (ClustalW).

The predicted N-terminal targeting sequence determined using ChloroP 1.1 during this study is specified in Figure 9. Details will be described later on in Section 3.1.4. Regions of paralogy to its newly found paralogs, TBC2Ap and TBC2Bp, can also be seen in this Figure (DOE JGI tBLASTn).

The putative TBC2A amino acid sequence was discovered during this study using the DOE JGI *C. reinhardtii* V.2.0 BLAST Alignment Search program, the cDNA-translated TBC2 amino acid sequence (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) BLASTed against the translated *C. reinhardtii* genome (see Figs. 10 and 12). greenGenie ORF finder predicted the length of the gene to be 4087 nucleotides. The TBC2A amino acid sequence was found to contain the PPPEW consensus sequence, spanning the central region of TBC2Ap (see Fig. 10). PPPEW repeat sequences are identified by shadowed boxes and amino acids with 100% consensus sequence identity are bolded (see Fig. 10). Stretches of amino acids like alanines and serines have also been identified, a common characteristic shared with TBC2p (see Figs. 9 and 10) (Auchincloss *et al.* 2002). The proposed N-terminal targeting sequence predicted during this study usin g ChloroP 1.1 and regions of paralogy to TBC2p and TBC2Bp are also outlined in Figure 10 (DOE JGI tBLASTn).

The TBC2B amino acid sequence, like TBC2Ap, was found by using the DOE JGI *C. reinhardtii* V.2.0 BLAST Alignment Search Program to BLAST the translated *TBC2* cDNA sequence (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) against the database (see Figs. 11 and 12). greenGenie ORF finder was used to predict the length of the expressed sequence, found to be

MELTRQIRWAASWRQVRELLAAAAPGSLDEVHLSTAATRLRAVAPPGQASASGGSAAPSSAVAVEDDPGRGP
GAARSQRPSGGGGRGARSGGVGGGASGSRAAKRERAAFRTFVAGVTELCGHHLPLMSAAPLTGVLYALAMLK
CPPPAPWMEAWLAAAAARMAAQMELVAGATAAGTGEGLAAPGTSSQAGTVSGGRAGGDRRGAGFGPQELS
NCMYALGQLGFDPGNDWMQRYCAAAGPLLLPPAGTAPPPPPPPLPMAATSSHGTSTSRSAFTAQGLSQMAWG
FARLGFVPEPAFLASLLAAVRRALPDFTAQGLANTLWAAASLGLAPPPSWVAAAAAAALRLLPSCSAYDLSVVL
WALLRLGHTPGPDWVSAYLAASYRQLPSASPEQAARMLWCCAAVGLPPPGDWVRRWLGCSYVKLLEAEPAAL
TTMAYALAALGIRPTDRWSDMLLVAAWEAPLRAFPPPDLALLMWGLAHCRVVPEPAWMDEWWAVSYKRLRSF
TPRHLSLLLWSCVTIGQAPSRQWLAGYEVVTLPAMPYMTAQGLSLLSYTYGCLERRPPRVWLAALYGAAAGVPV
EPVAPSPQLVLAEAEVEAEAGHKAEAGHKAEAEGCSGSANVDSVAAAGSSSSSSSSSSSSYEGLSRFTALGL
ERLLWGIAKMEPPAGSLPDDVLPGWTDAFLAAAAARLLPAHTGSSGSGSSSGSSSSGSSSGSSSGSGETGAGA
AGMGEVGLGVEEADGTYGNVANTLYALALLRVRPEAAWLLPVMDRVEALLPDFGHTTLVKVAWALPRLTSGPPT
HRPFTGCTAPAPAQAATPQPAAASVPPHNDAREAGLPSDVGPDASEQRRLRERLVSLQRLVATRMRRTIGGRRA
AGAVAVASGGDADSASDAEAADWDESLEEPGAGVAVFGVAGDVAGSGSATDSDGGDGAQAGVARAPRRHVR
RRAVQD

Fig. 10. The TBC2A amino acid sequence. The TBC2A amino acid sequence possesses TBC2p's novel PPPEW repeat, enclosed in shadowed boxes. The amino acids which are identical to the consensus sequence are bolded. Stretches of A, P and S amino acids are bordered with simple boxes. The N-terminal targeting domain determined by ChloroP 1.1 is bolded and underlined. The region of paralogy to the TBC2 and TBC2B amino acid sequences is defined by a simple underline (DOE JGI tBLASTn). (These sequence data were produced by the US Department of Energy Joint Genome Institute, http://www.jgi.doe.gov/ and are provided for use in this publication/correspondence only.)

3810 nucleotides in length.

Seven PPPEW repeats in the centre of the TBC2B protein sequence are identified by shadowed boxes, whereas 100% identities to the PPPEW amino acid repeat sequence are bolded (see Fig. 11). Glutamine, alanine, serine and glycine stretches are boxed, the predicted N-terminal targeting sequence determined by ChloroP 1.1 is bolded and underlined and the region of paralogy to TBC2p and TBC2Ap is underlined (DOE JGI tBLASTn) (see Fig. 11).

3.1.2. DOE JGI *C. reinhardtii* Genome Database Version 2 Translated BLAST of *TBC2* cDNA

A BLAST of the TBC2p sequence against the translated nuclear genome of *C. reinhardtii* was done in order to determine if any paralogous sequences existed. Retrieval of known sequences could characterize the structural conformation and function of TBC2p to determine similarities. The TBC2 amino acid sequence derived from its translated cDNA sequence (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) was BLASTed against the *C. reinhardtii* Version 2 Genome Database. Four distinct regions on four different scaffolds were found to have significant homology to the TBC2 amino acid sequence (see Fig. 12). The region of homology on scaffold 1767 was considered a segment of the TBC2 amino acid sequence, scaffold 1767 being a relatively miniscule scaffold (see Fig. 12). Regions on scaffold 9 and 10 were assumed to be paralogs of TBC2p because of the presence of TBC2p's novel PPPEW repeat and therefore underwent further analysis (see Figs. 10, 11 and 12). Discussed in subsequent paragraphs, these regions of paralogy were

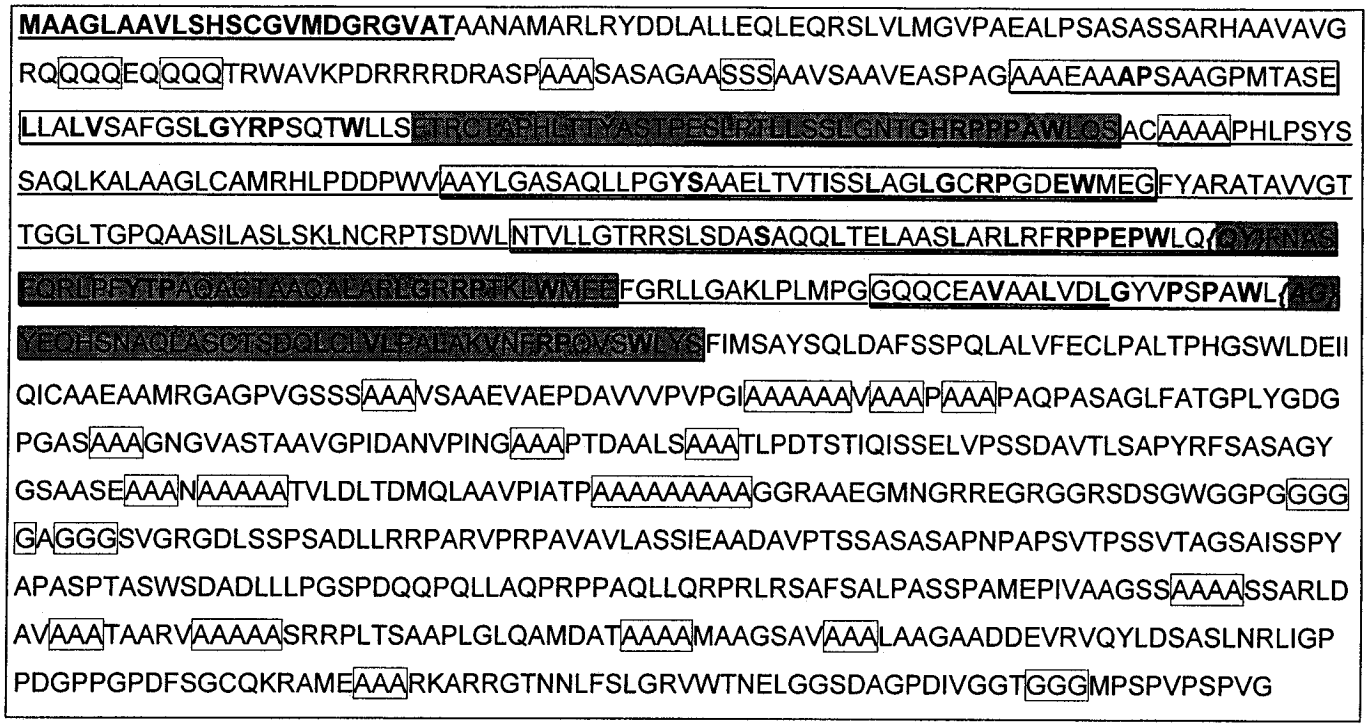

Fig. 11. The TBC2B amino acid sequence. PPPEW consensus sequences are enclosed in shadowed boxes, repeats flanking each other are highlighted in grey. Areas of overlapping consensus sequences are italicized, bolded and bracketed. Amino acids matching the consensus are bolded. Q, A, S and G strings are bordered with simple boxes. The N-terminal targeting domain predicted by ChloroP 1.1 is underlined and bolded. The region of shared paralogy with the TBC2A and TBC2B amino acid sequences is defined by a simple underline (DOE JGI tBLASTn). (These sequence data were produced by the US Department of Energy Joint Genome Institute, http://www.jgi.doe.gov/ and are provided for use in this publication/correspondence only.)

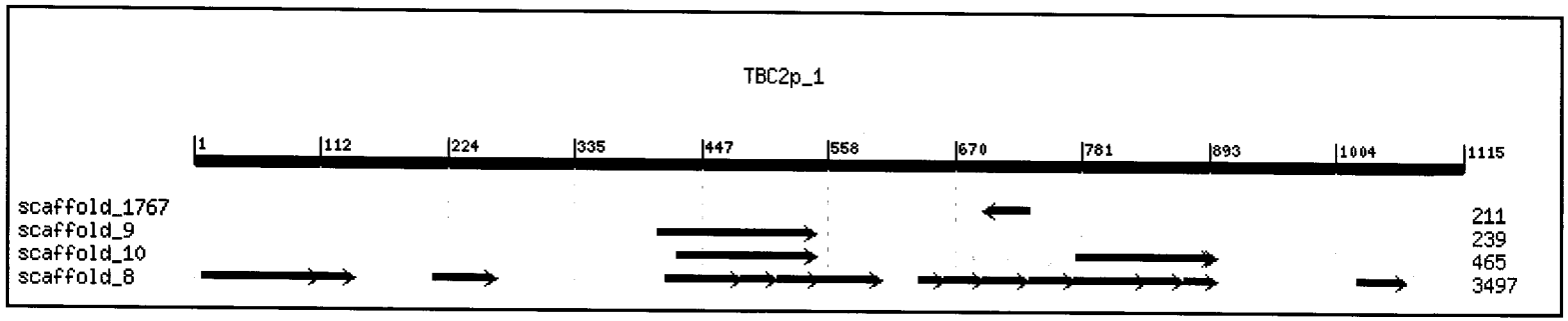

Fig. 12. Results of a translated BLAST of the TBC2 amino acid sequence against the DOE JGI V.2.0 *C. reinhardtii* Genome Database. *From top to bottom:* scaffold 1767 was found to have a homologous amino acid sequence in a negative orientation with 100% homology to TBC2p. This sequence was considered to be a portion of the TBC2 amino acid sequence; scaffolds 9 and 10 contained sequences with the novel PPPEW degenerate consensus found only in TBC2p. Percent identities of these sequences were low (25.81-32.47%) and both were in a positive orientation. The program ORF finder found that these two regions of homology belonged to possible open reading frames and as a result they are theorized to be paralogs; scaffold 8 was previously determined to contain the *TBC2* sequence. (These sequence data were produced by the US Department of Energy Joint Genome Institute, http://www.jgi.doe.gov/ and are provided for use in this publication/correspondence only.)

determined to fall in possible ORFs using greenGenie ORF finder (DOE JGI tBLASTn).

Scaffold 8 contained the TBC2p sequence and was previously determined by and registered in the DOE JGI V.2.0 *C. reinhardtii* Nuclear Genome Database, mapped between nucleotides 1172024 and 1179154. Most of the translated *TBC2* cDNA sequence (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) was homologous to the region of the scaffold up until amino acid 1081 out of 1115 amino acids from the N-terminus. Most of the identities were high, at least 76.60% (see Table 8). Even though identity should theoretically be 100%, it was not, attributed to the fact that the DOE JGI V.2.0 *C. reinhardtii* Nuclear Genome is a pre-publication draft and probably contains errors (L. Peeters, email comm.). Normally, the genome should be sequenced eight to ten times to ensure that no errors exist (Gardner *et al.* 2005). Since this was a translated BLAST, gaps existed in the alignment pertaining to the introns which had been spliced out of the TBC2p sequence. However, the number of regions of homology between the TBC2p sequence (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) versus the DOE JGI V.2.0 *C. reinhardtii* Nuclear Genome sequence did not correspond to a cDNA-to-genomic DNA NCBI pairwise BLAST which showed only seven homologous regions corresponding to exons (see Table 8) (refer to Results 3.1.1.) (DOE JGI tBLASTn). This could be attributed to the fact that the DOE JGI V.2.0 *C. reinhardtii* Nuclear Genome sequence is a pre-publication draft and contains errors which creates many additional regions of non-homology which do not correspond to introns (L.

Peeters, email comm.).

Tables 9A and 9B are the BLAST results obtained for what are considered to be paralogs TBC2Ap and TBC2Bp. Putative ORFs for TBC2Ap and TBC2Bp are located within regions of the nuclear genome which have not been ruled out as containing expressed sequences (ORF Finder). The percent identities achieved for the two paralogous sequences of TBC2Ap were 28.35 and 31.21%, with Hsp scores of 122 and 117, respectively (see Table 9A). As can be seen in this Table, a region of overlap exists in the defined regions of paralogy where the first span extends over region 948788-949210 and the span of lower identity extends over region 948716-949096, overlapping the first region by 308 nucleotides (see Table 9A). The explanation for this occurrence could be due to low identity and that random alignments had occurred (see Table 9A) (DOE JGI tBLASTn).

TBC2Bp, another putative paralog obtained from the DOE JGI V.2.0 *C. reinhardtii* BLAST Alignment, was found on scaffold 10 with a series of five independent paralogous stretches on the scaffold (see Table 9B). Identities for TBC2Bp ranged between 25.81 and 32.44%. Hsp scores ranged from 72 to 109 (see Table 9B) (DOE JGI tBLASTn).

To determine whether the nucleotide sequences of the paralogies designated *TBC2A* and *TBC2B* could be actual expressed sequences, the program greenGenie ORF finder (http://www.cse.ucsc.edu/~dkulp/cgi-bin/greenGenie) was used to predict whether or not these spans existed in ORFs, consisting of both exons and introns between start and stop codons. Sequences

3 kb in length flanking either side of the spanning region of paralogy were used to determine the possible ORFs for *TBC2A* and *TBC2B*. Initially named *H2*, *TBC2A* consists of a predicted ORF of 4087 nucleotides in length, its exons coded for in the positive direction as shown in Table 10A. Regions of paralogy were found within limits of the ORF (see Table 10A) (ORF Finder).

The region of paralogy to TBC2p found on scaffold 10 was initially named *H3* and was renamed *TBC2B* after its putative ORF had been established as shown in Table 10B. The greenGenie ORF Finder program found an ORF containing four exons to be 3810 nucleotides in length and coded for in the positive direction (see Table 10B). Regions of paralogy were within limits of the ORF (see Table 10B) (ORF Finder).

3.1.3. Comparisons of TBC2p, TBC2Ap, TBC2Bp and Crp1p sequences

ClustalW was used to align the amino acid sequence TBC2p (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]), 1115 amino acids in length, with the predicted amino acid sequences for TBC2Ap, which is 950 amino acids in length, and TBC2Bp, 1027 amino acids in length. These protein sequences were determined using the ExPASy Translate Tool (http://www.expasy.ch/tools/dna.html). Regions of paralogy are shown in Figure 13 (ClustalW; DOE JGI tBLASTn; ORF Finder).

According to the Figure, hyphens indicate a discontinuation of the sequence until another region with minimal paralogy has been reached. TBC2Ap and TBC2Bp's N-termini are not paralogous to the TBC2 amino acid sequence (see Fig. 13). PPPEW repeats are found predominantly in the centre

Alignment for TBC2p	Identity (%)	Length in amino acids	Position on scaffold 8	Hsp value
1 (amino acid 1-111)	99.05	104	1172024-1172338	393
2 (amino acid 108-142)	94.29	134	1172678-1172782	155
3 (amino acid 211-266)	100	155	1172987-1173154	275
4 (amino acid 414-483)	88.57	69	1174166-1174375	312
5 (amino acid 481-513)	96.97	32	1174611-1174709	143
6 (amino acid 513-550)	100	131	1175041-1175154	191
7 (amino acid 543-605)	95.24	62	1175349-1175537	297
8 (amino acid 638-661)	87.50	23	1175909-1175980	114
9 (amino acid 659-695)	97.30	36	1176296-1176406	84.0272
10 (amino acid 694-735)	100	41	1176834-1176959	211
11 (amino acid 735-775)	97.56	40	1177091-1177213	215
12 (amino acid 773-839)	97.01	66	1177476-1177676	343
13 (amino acid 824-870)	76.60	46	1177894-1178034	189
14 (amino acid 870-899)	96.67	29	1178288-1178377	149
15 (amino acid 1022-1067)	100	45	1178744-1178881	256
16 (amino acid 1066-1081)	100	15	1179107-1179154	79

Table 8. Positioning of the TBC2p sequence on the nuclear genome resulting from a translated BLAST of the TBC2p sequence (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) versus the DOE JGI V.2.0 *C. reinhardtii* Nuclear Genome Database.

This table describes the identities (identical amino acids), positives (conservative amino acids), and Hsp scores (high scoring pair number; an alignment without gaps resulting in a high score, the score normally taking into account matches, mismatches, and gap penalties) obtained for the TBC2 amino acid sequence (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) against the DOE JGI V.2.0 *C. reinhardtii* Nuclear Genome Database. It is representative of the region of homology on scaffold 8 seen in Figure 12. (These sequence data were produced by the US Department of Energy Joint Genome Institute, http://www.jgi.doe.gov/ and are provided for use in this publication/correspondence only.)

A

Alignment for TBC2Ap	Identity (%)	Length in amino acids	Position on scaffold 9	Hsp value
1 (amino acid 408-548)	31.21	140	948788-949210	122
2 (amino acid 422-548)	28.35	127	948716-949096	117

B

Alignment for TBC2Bp	Identity (%)	Length in amino acids	Position on scaffold 10	Hsp value
1 (amino acid 425-548)	25.81	123	525110-525481	100
2 (amino acid 496-549)	27.78	53	524440-524601	72
3 (amino acid 686-761)	31.58	76	1078508-1078735	78
4 (amino acid 776-897)	26.23	121	524318-524682	109
5 (amino acid 823-899)	32.47	76	525113-525343	106

Table 9. A, Output for the TBC2Ap region of paralogy resulting from a translated BLAST of the TBC2 amino acid sequence (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) versus the DOE JGI V.2.0 *C. reinhardtii* Nuclear Genome Database. The regions of paralogy in the nuclear genome of *C. reinhardtii* showed moderate identity and Hsp values. Existing identity was attributed mainly to alignment with the PPPEW consensus sequence; **B, Output for the TBC2Bp region of paralogy resulting from a translated BLAST of the TBC2 amino acid sequence (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) versus the DOE JGI V.2.0 *C. reinhardtii* Nuclear Genome Database.** Identities and Hsp values of the BLAST were moderate, the span of paralogies were more extensive in comparison to TBC2Ap. TBC2Bp also contained the PPPEW consensus, increasing the value of identity. (These sequence data were produced by the US Department of Energy Joint Genome Institute, http://www.jgi.doe.gov/ and are provided for use in this publication/correspondence only.)

A

Exon	Start of exon (nucleotide position in *TBC2A*)	End of exon (nucleotide position in *TBC2A*)
Initial exon	1	364
Internal exon	518	1480
Internal exon	1843	1995
Internal exon	2291	2398
Internal exon	2518	2911
Internal exon	3053	3230
Final exon	3399	4087

B

Exon	Start of exon (nucleotide position in *TBC2B*)	End of exon (nucleotide position in *TBC2B*)
Initial exon	1	903
Internal exon	1208	1624
Internal exon	1867	1953
Final exon	2137	3810

Table 10. A, The positions of *TBC2A* exons using greenGenie ORF finder. The program predicted an ORF of 4087 nucleotides for the *TBC2A* sequence with 7 exons; **B, The positions of *TBC2B* exons using greenGenie ORF finder.** *TBC2B* was predicted to have an ORF of 3810 nucleotides with four exons. (These sequence data were produced by the US Department of Energy Joint Genome Institute, http://www.jgi.doe.gov/ and are provided for use in this publication/correspondence only.)

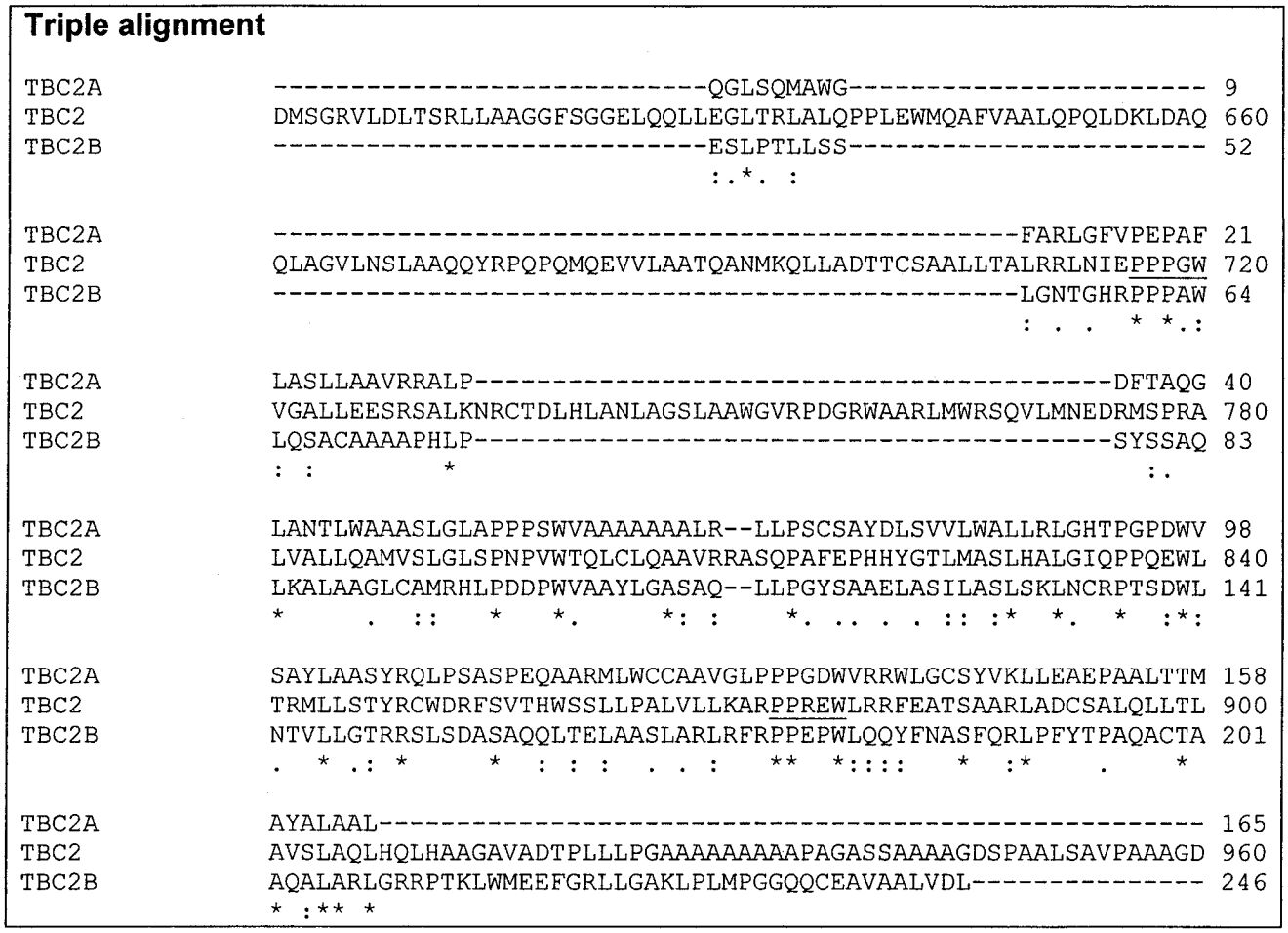

Fig. 13. A ClustalW alignment of the TBC2 protein sequence with TBC2Ap and TBC2Bp. Stars signify that residues in the specific alignment are identical across all sequences in that alignment. A colon indicates that a conserved substitution has been observed and a period indicates that a semi-conserved substitution has been observed, suggesting low homology between sequences. The "PPPEW" string within the PPPEW consensus is underlined in TBC2p (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) where evident. Hyphens substitute the amino acid sequence where no minimal identity has been achieved. (These sequence data for "TBC2Ap" and "TBC2Bp" were produced by the US Department of Energy Joint Genome Institute, http://www.jgi.doe.gov/ and are provided for use in this publication/correspondence only.)

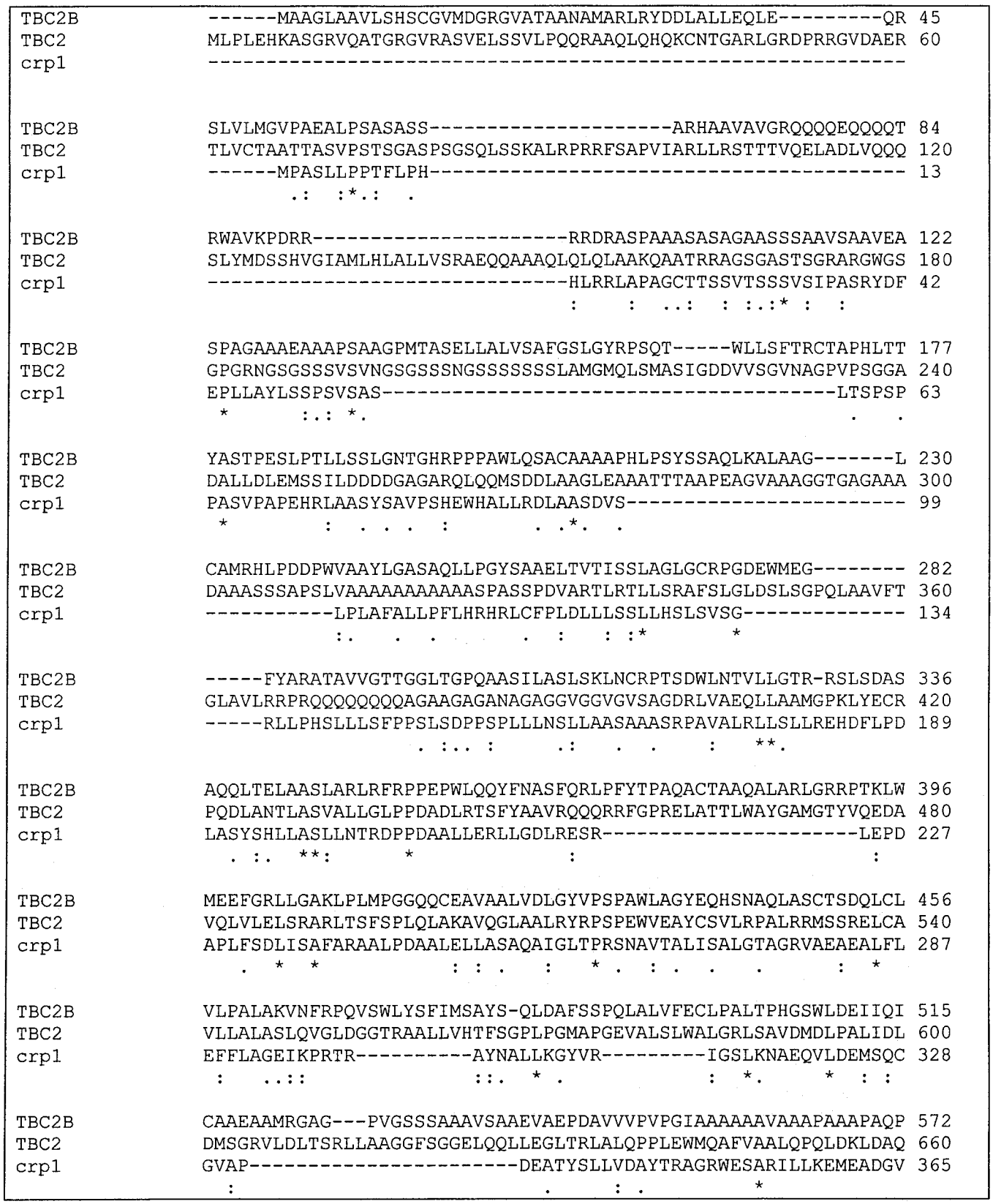

Fig. 14. Legend on page 69.

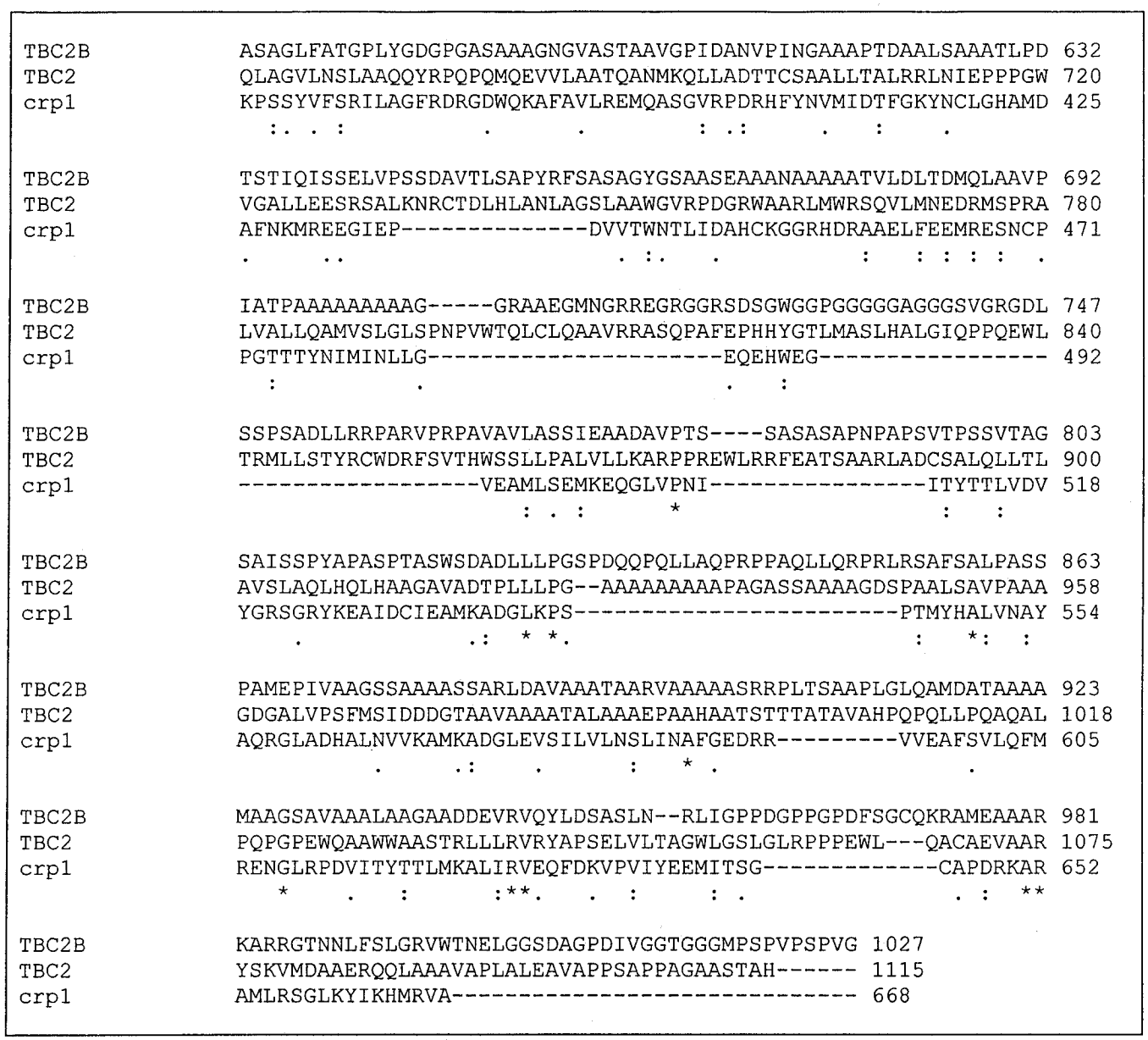

Fig. 14 (cont.). Legend on page 69.

Fig. 14. A 3-sequence ClustalW alignment of TBC2p, Crp1p and the TBC2Bp paralog. This alignment was done in order to demonstrate any homology between the TBC2B protein sequence and the Crp1 protein sequence (NCBI [accession AAC25599, version AAC25599.1 gi: 3289002], with kind permission from A. Barkan) from *Z. mays*. A star indicates 100% identity, a colon indicates that a conserved substitution has occurred and a period indicates that a semi-conserved substitution has taken place (TBC2p was obtained from NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]; these sequence data for "TBC2Bp" were produced by the US Department of Energy Joint Genome Institute, http://www.jgi.doe.gov/ and are provided for use in this publication/correspondence only.)

and at the C-terminus of all three sequences, possibly indicating that these regions are of increased functional significance and that conservation of these regions is of increased importance (see Fig. 13). Certain paralogous spans are shared by all three TBC2 amino acid sequences, indicating that this region may serve a functional purpose (see Fig. 13). Most identities are the result of the PPPEW consensus sequence (see Fig. 13). Outside the paralogy due to the shared PPPEW consensus, weak conservation exists (see Fig. 13). TBC2Bp appears to share greater paralogy with TBC2p than does TBC2Ap (ClustalW; DOE JGI tBLASTn; ORF Finder).

Because TBC2Bp has a higher sequence similarity to TBC2p than TBC2Ap does to TBC2p, the TBC2B protein sequence was included in a ClustalW alignment with Crp1p and TBC2p (see Fig. 14). Auchincloss and colleagues (2002) found that regions of partial similarity exist between TBC2p and Crp1p, especially between the central region of TBC2p containing PPPEW repeats, and the PPR-containing region of Crp1p (see Fig. 14) (see Section 1.B.5. of Introduction) (ClustalW; DOE JGI tBLASTn; ORF Finder). The theory proposed here was that if Crp1p demonstrates similarity to the TBC2 protein sequence, then perhaps the TBC2B protein sequence, as TBC2p, will demonstrate sequence homology and put forth a hypothetical function for TBC2Bp (see Fig. 14).

3.1.4. Determination of the presence of transit peptides and organellar localization of TBC2p and its paralogs

Organellar proteins normally contain a targeting sequence so that they can be imported into the organelle. To determine whether TBC2p and its paralogs contain transit peptides, programs TargetP 1.1 (http://www.cbs.dtu.dk/services/TargetP/) and ChloroP 1.1 (http://www.cbs.dtu.dk/services/ChloroP/) were used which determine whether a putative targeting domain was present.

The TargetP 1.1 program predicts transit peptides on an output scoring system ranging from 1 to 0 whereby scores have been separated into 5 reliability classes, 1.0-0.8, 0.8-0.6, 0.6-0.4, 0.4-0.2 and 0.2-0. TargetP 1.1 predicts chloroplast, mitochondrial and secretory pathway signals in eukaryotes.

The ChloroP 1.1 program determines cTP potential cleavage sites using a MEME-derived scoring matrix, MEME being an automatic motif-finding algorithm. Gavel and von Heijne (1990) discussed the cleavage site appearing as the semi-conserved motif (I/V)-X-(A/C)↓-A. Emanuelsson and colleagues (1999) stated this finding was incorporated into the ChloroP 1.1 program with the knowledge that 1-3 additional residues within the protein sequence are cleaved after the fact. ChloroP 1.1 bases the cleavage site score on the predicted cleavage site located N-terminal to the highest scoring amino acid, the search restricted to the first 40 amino acids of the N-terminus.

Using TargetP 1.1, the N-terminus of TBC2p (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) was found to bear similarity to a mTP in

higher plants with a relatively high prediction score of 0.812. This finding is consistent with the introductory explanation on *C. reinhardtii* cTP properties being interchangeable with mTPs in higher plants (Franzén *et al.* 1990). Using ChloroP 1.1, the prediction score for the N-terminus being a transit peptide was 0.480, its length estimated to be 20 amino acids with a cleavage site score of 7.226.

The TBC2Ap was predicted by TargetP 1.1 to contain an N-terminal transit peptide showing great similarity to a mTP in a plant network. The ChloroP 1.1 program determined the cTP length to be 50 amino acids with a transit peptide score 0.453, similar to the score achieved by the TBC2 protein. TBC2Ap also had a cleavage site score of 5.250 (ChloroP 1.1; DOE JGI tBLASTn).

Using TargetP 1.1, the TBC2Bp sequence had a cTP prediction score of 0.215 and a mTP score of 0.125 for a plant network. In a non-plant network, the N-terminus resembles a mTP with a prediction value of 0.378 and has a slight resemblance to a cTP with a value of 0.142 (TargetP 1.1). When using the ChloroP 1.1 program, the cTP was predicted to be 23 amino acids in length with a prediction score of 0.444, a value similarly achieved for TBC2p and paralog TBC2Ap. TBC2Bp had a cleavage site prediction score of 3.220 (ChloroP 1.1; DOE JGI tBLASTn).

3.1.5. TBC2, TBC2A and TBC2B amino acid sequences, their characteristic PPPEW repeat and the prevalence of amino acid strings

A ClustalW multiple sequence alignment was done for PPPEW repeat sequences found in the TBC2 amino acid sequence (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) as well as TBC2A and TBC2B amino acid sequences. The PPPEW consensus is aligned at the bottom of the Figure 15. Alignments of 100% were only seen for amino acids P and W in positions 33 and 37 of the repeat sequence while other amino acids in the repeat showed weak alignment (see Fig. 15). Amino acids are identified according to hydrophobicity/hydrophilicity/polarity (ClustalW; DOE JGI tBLASTn).

Strings of amino acids in TBC2p, TBC2Ap and TBC2Bp were also found and outlined (see Figs. 9, 10 and 11). TBC2p (NCBI [locus CAD20887, version CAD20887.1 gi: 22139636]) was defined by strings of alanines, serines and glutamines; TBC2Ap also had alanines and serines arranged consecutively but also stretches of prolines as well (see Figs. 9 and 10). TBC2Bp had stretches of alanines, serines and glutamines, as well as glycines (see Fig. 11). These amino acids, due to their unusual succession, are theorized to fulfill a functional or structural role (see Discussion Section 4.1.5) (DOE JGI tBLASTn).

3.1.6. Predicting the function of TBC2p and its paralogs through analysis of conserved domains

Conserved domains were sought to determine the function of TBC2p and its paralogs to a greater extent. Conserved domains are assumed to have maintained the same functional motif throughout molecular evolution and

therefore may retain the same functional purpose. The NCBI Conserved Domain Database (CDD) contains compiled domain sequences from SMART, pFAM and COG databases (see Methods of Analysis for their definitions, p. 41). RPS-BLAST (Reverse Position Specific BLAST) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) uses the PSSM (Position Specific Score Matrix) Database against the query to report significant hits, thereby reversing the BLAST. The expect value of the RPS-BLAST is the "E" value of filtering, while the score of the alignment accounts for mismatches, matches and gap penalties.

The NCBI CDD search was done with the translated cDNA sequence of *TBC2* (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]). The following discussion describes the most significant and pertinent results obtained relative to the function of the translational activator protein.

For *TBC2A*, the translated ORF sequence was used, the expect value was fixed at 1 and the filter was taken off low complexity. The most significant result was the third highest in homology to TBC2Ap, gnlCDD|25713 pFAM01213, a catabolite activator protein (CAP) from eukaryotes with a score of 32.7 and an E value of 0.20. Other homologies of significance included gnlCDD|26882 pFAM06743, a Fas-activated serine/threonine (FAST) LEU-rich protein kinase with a score of 32.1 and an E value of 0.31. FAST LEU-rich protein kinase was chosen because it phosphorylates a nuclear RNA-binding protein and could be related functionally to the TBC2A protein. Another was COG5178.1 gnlCDD|14302, a U5 snRNP spliceosome subunit required for RNA processing

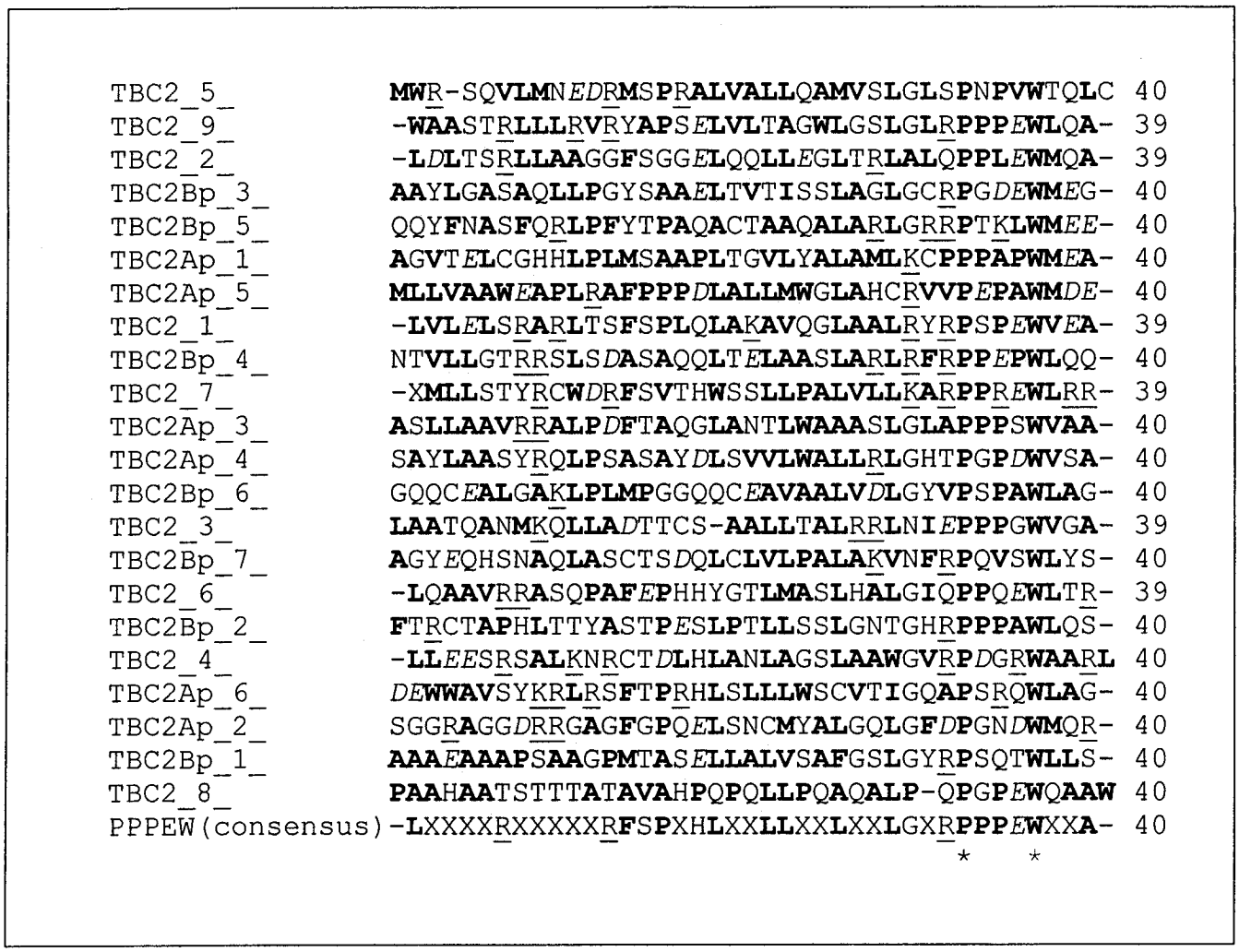

Fig. 15. Legend on next page.

Fig. 15. A ClustalW multiple sequence alignment for the PPPEW repeat sequences of TBC2p, TBC2Ap and TBC2Bp. Stars indicate 100% alignment and appear only for P and W at positions 33 and 37. Differences in font appear according to hydrophobicity/hydrophilicity/polarity which shows some conservation. Sequences are listed from the most highly conserved domains to the lowest. Small hydrophobic or aromatic amino acids such as V, F, P, M, I, L or Y are bolded. Acidic amino acids such as D or E are italicized, basic amino acids such as R, H or K residues are underlined, hydroxyl and amine-containing amino acids such as S, T, Y, H, C, G, Q and N are in regular font style (TBC2p was obtained from NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]; these sequence data for "TBC2Ap" and "TBC2Bp" were produced by the US Department of Energy Joint Genome Institute, http://www.jgi.doe.gov/ and are provided for use in this publication/correspondence only.)

and modification in eukaryotes and is functionally significant considering that TBC2p is a post-transcriptional protein involved in translational activation. Its score was 31.2 and E value 0.58 (DOE JGI tBLASTn; NCBI CDD V.2.04 RPS-BLAST; ORF Finder).

The *TBC2B* translated ORF sequence was entered as a query with an expect value of 1 and the filter removed for low complexity sequences. A finding of pertinence to the function of a translational activator protein was gnlCDD|27279 pFAM 07172, belonging to a glycine-rich protein family (GRP), with a score of 33 and an E value of 0.18. TBC2Bp also showed similarity to gnlCDD|26366 pFAM05236 TAF4, a transcription initiation factor TFIID component in the TAF4 family (DOE JGI tBLASTn; NCBI CDD V.2.04 RPS-BLAST; ORF Finder).

3.1.7. Determining protein characteristics using PROSPECT-PSPP

PROSPECT-PSPP predicted the TBC2 protein (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) to be a soluble one which correlates with findings made by Auchincloss and colleagues (2002). It was predicted to have three domains, all of them mainly α-helical. No coiled-coils were found. Even though most of the PPPEW repeats and strings of amino acids fall into the region of helices, PPPEW repeat sequences are not solely involved in generating them. The model appears to form a crescent shape (see Fig. 16) (PROSPECT-PSPP).

TBC2Ap was also divided into three domains and consisted of mainly α-helices like TBC2p (see Fig. 16). Again, most PPPEW repeats and strings of amino acids were not solely responsible for the generation of α-helices. TBC2Ap

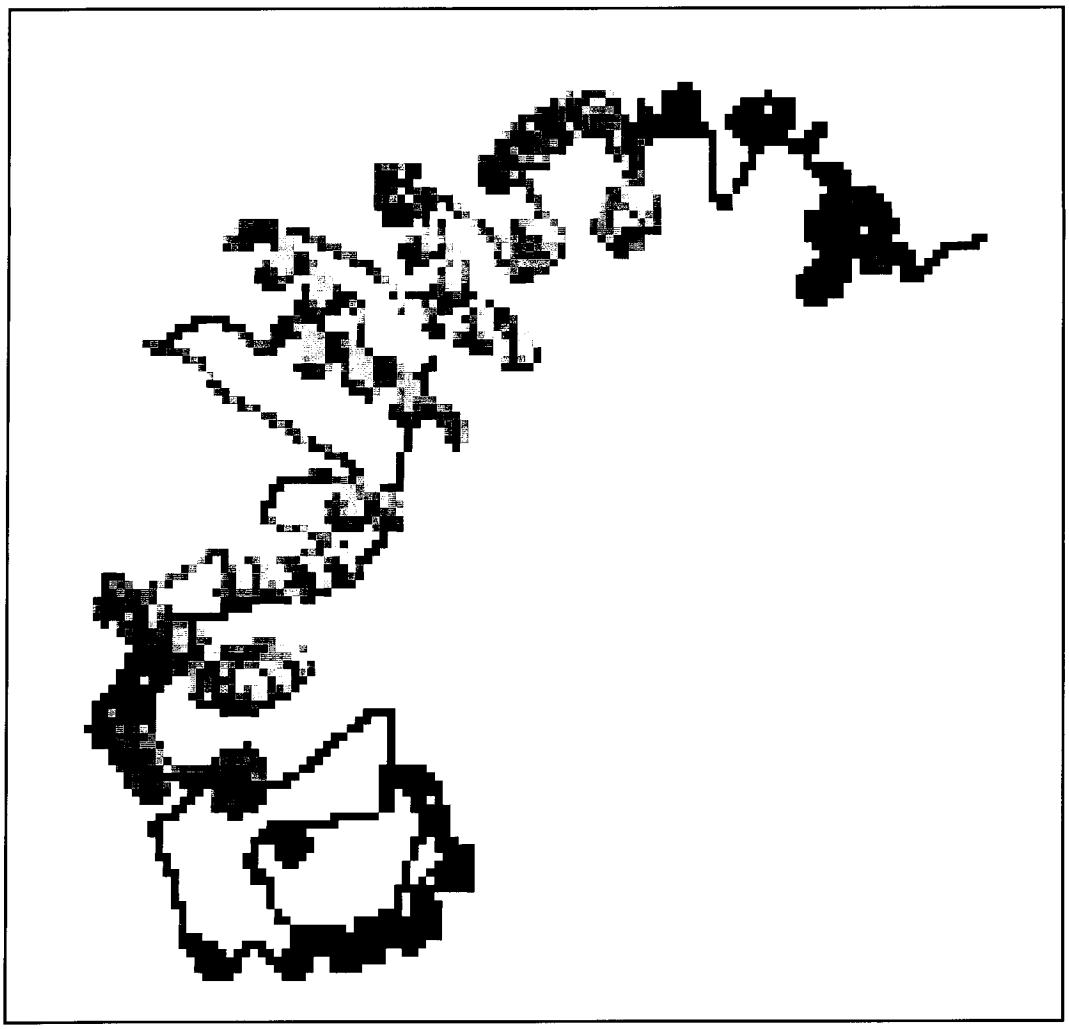

Fig. 16. Structure prediction for TBC2p. PROSPECT-PSPP was used to determine the theoretical structure for TBC2p (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]). Results showed that TBC2p was predicted to be mainly α-helical with three domains and no coiled-coils. TBC2p has a crescent shape and the PPPEW repeats as well as strings of amino acids were shown not to be the sole contributors to α-helical structure (PROSPECT-PSPP) (These sequence data were produced by the US Department of Energy Joint Genome Institute, http://www.jgi.doe.gov/ and are provided for use in this publication/correspondence only.)

was also predicted to be a soluble protein and no coiled-coils were found. Similar results were obtained for TBC2Bp (DOE JGI tBLASTn; PROPSECT-PSPP).

3.2. Experimental approach taken towards determining the mode of TBC2p interaction

Based on Chapter 1.B, this section describes the preparatory steps of a coimmunoprecipitation experiment to determine how TBC2p binds to the 5'-UTR of *psbC* mRNA. The following methods were used in this approach:

-Verification of constructs for Western blots and comimmunoprecipitation experiments

-Transformation of the TBC2:HA construct into *tbc2-F64;cw15;arg7*

-PCR amplification of *TBC2A* and *TBC2B* paralogs

-Western blots of the TBC2:HA protein

-Progeny of *Tab2;cw92* crosses

3.2.1. Verification of constructs for Western blots and coimmunoprecipitation experiments

The agarose gel in Figure 17 shows the plasmid "HA @ stop in KBam10 (ie. genomic)" after cleavage with *Nhe*I and *Eco*RV, including a 700 bp fragment, known to contain the hemagglutinin coding sequence due to restriction mapping of the original plasmid sequence (see Fig. 8). Therefore, verification was successful and positive so that the TBC2:HA construct could be transformed and expressed in *tbc2-F64;cw15;arg7 C. reinhardtii* cells as an autonomously-replicating plasmid. Cells could then be lysed and tested through immunoblotting to determine if they contained the TBC2:HA protein for the

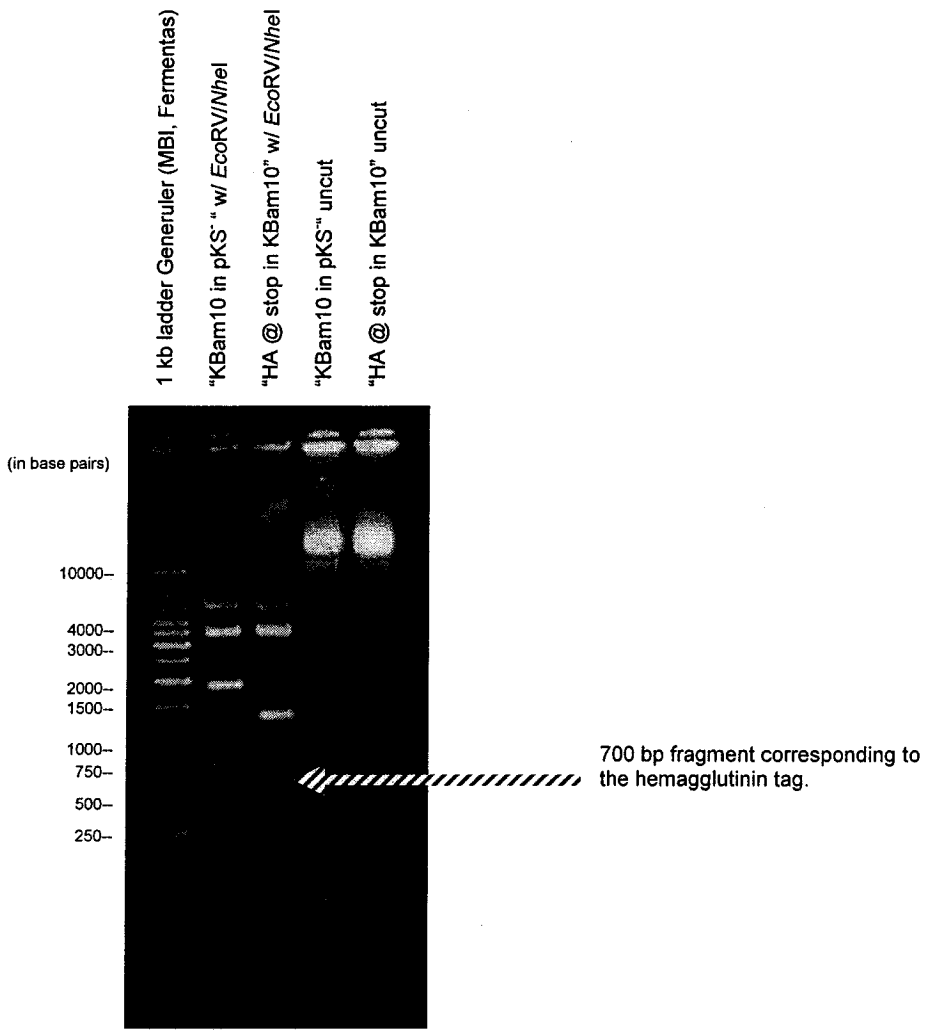

Fig. 17. Verification of the correct plasmids. "HA @ stop in KBam10 (ie. genomic)" was digested with *Nhe*I and *Eco*RV, yielding a 700 bp fragment corresponding to the hemagglutinin DNA sequence.

coimmunoprecipitation experiment.

3.2.2. Transformation of the TBC2:HA construct into *tbc2-F64;cw15;arg7*

"HA @ stop in KBam10 (ie. genomic)" was transformed into *tbc2-F64;cw15;arg7* to express HA-tagged TBC2p. Colonies were selected for an ac^+ phenotype in light conditions on HSM medium containing 55 µg/ml arginine. TBC2:HA protein expression could now be verified using Western blots for coimmunoprecipitation.

3.2.3. PCR amplification of *TBC2A* and *TBC2B* paralogs

PCR was done in order to amplify *TBC2A* and *TBC2B*, paralogs of *TBC2*, which would be labeled with ^{32}P and used to probe a cDNA library. The cDNA would then be cloned into an expression vector and expressed in *C. reinhardtii*. This was in preparation of an RNAi experiment whereby chemically-synthesized siRNAs could be introduced into the cell resulting in degradation of *TBC2A* and *TBC2B* mRNA due to Dicer (Rohr *et al.* 2004). Phenotypic effects would be expected as a result (Rohr *et al.* 2004).

Primers "*TBC2A* forward," *TBC2A* reverse," *TBC2B* forward" and "*TBC2B* reverse" were hybridized to the nuclear genomic DNA sequences shown in Figure 18.

The PCR product achieved for *TBC2A* amplification was 1462 bp, the predicted length as shown in Figure 19. The PCR product obtained for *TBC2B* amplification was 608 bp, also the expected length of amplified product (see Fig. 19). Products were purified using Qiagen Gel Extraction and PCR Purification kits.

3.2.4. Western blots of the TBC2:HA protein

tbc2-F64;cw15;arg7 strains which had been transformed with the TBC2:HA construct had to be verified for TBC2:HA protein expression using Western blots prior to the coimmunoprecipitation experiment. After lysis by sonication, the suspensions and pellets were run separately on 7.5% SDS-PAGE gels, transferred to nitrocellulose paper and tested for transfer using Ponceau S staining (see Figs. 20 and 21). Filters were then treated with purified murine antibody (mono HA.11) as the primary antibody and HRP-anti-mouse antibody.

The TBC:HA protein migrated and was detected at an expected mass of ~119 kDa, calculated according to the lengths of the triple hemagglutinin epitope and the predicted amino acid sequence of TBC2p from its cDNA (see Figs. 22 and 23). The protein was expressed in the cell transformants, especially in the soluble fraction (see Figs. 22 and 23). The Raa2:HA positive protein control did not appear in the gel blot, possibly because of proteolysis or that it ran off the gel, Raa2p being only 44.8 kDa (see Figs. 22 and 23). The *tbc2-F64;cw15;arg7* negative control had no TBC2:HA protein and therefore the lane was blank (see Figs. 22 and 23). These results indicated that these transformants could be used for a coimmunoprecipitation experiment (see Figs. 20-23).

A)

949982 CTTTGACCTTGGCACCCCCACCCGCATGCACACGCGCACAGGGCGGTGTC **950031**
950032 *GTACAAGCGGTTGCGGTCCTTCACCCCGCGACACCTGTCGCTGCTGCTGT* **950081**
950082 *GGTCGTGCGTCACCATTGGCCAGGCGCCCAGTCGGCAGTGGCTGGCGGGG***950131**
950132 TGAGTGCCGCACGCGTTGGGGGGCTGTGCTACATGCCTGACACGCCATAT **950181**

5'-GCGTTGGGGGGCTGTGCTAC-3'

Primer "*TBC2A forward*", non-template underlined

950182 GCCACCCCTTGGGCGGCGCGACCGCCCAACCCTGGACCCATTCAACCTAA **950231**
950232 CACGTACTTCTCCCCCAGGTACGAGGTCGTGACGCTGCCCGCCATGCCGT **950281**
950282 *ACATGACGGCCCAGGGCCTCTCACTGCTGTCGTACACGTACGGCTGCCTG* **950331**
950332*GAGCGCCGCCCGCCACGTGTCTGGCTGGCGGCGCTGTACGGCGCGGCGGC***950381**

B)

951382*GGCTGTACCGCGCCGGCGCCGGCTCAGGCCGCAACACCGCAGCCAGCAGC***951431**
951432 *GGCCTCTGTGCCGCCGCACAACGATGCTCGGGAGGCTGGGTTGCCCTCTG* **951481**
951482*ACGTCGGGCCTGATGCGAGCGAGCAGCGGCGCCTGCGGGAGCGACTGGTC***951531**
951532 *TCGCTGCAGCGGCTGGTGGCGACCCGCATGCGGCGGACCATTGGCGGCCG***951581**

Primer "*TBC2A* reverse", template strand underlined

3'-CGGCGTCCTCGTCATCGTCA-5'

951582 *CCGGGCCGCAGGAGCAGTAGCAGTGGCATCGGGTGGCGACGCAGATTCGG***951631**
951632 *CTTCCGATGCTGAGGCGGCTGACTGGGATGAGAGTCTGGAGGAGCCGGGC* **951681**
951682*GCAGGGGTGGCGGTGTTTGGTGTGGCTGGAGACGTGGCCGGAAGCGGCAG***951731**
951732*TGCCACGGACAGCGACGGAGGTGACGGGGCGCAGGCCGGTGTCGCCCGCG***951781**

C)

523850TGCCGGCGGCAGTGCCGCAGCCGCAGCCCAGCCGCCGGCAGATCACGAGC**523899**
523900 *ATGGCCGCGGGACTGGCGGCGGTGCTGTCGCATTCCTGCGGCGTCATGGA* **523949**
523950 *CGGCCGCGGGGTTGCCACTGCCGCGAACGCCATGGCTCGTTTGCGCTACG* **523999**
524000 *ACGACCTTGCGCTTCTAGAGCAGCTGGAGCAGCGCTCGCTGGTGCTGATG* **524049**

5'-GCTGGTGCTGATG

Primer "*TBC2B forward*", non-template underlined

524050 *GGTGTGCCCGCAGAGGCGCTGCCATCCGCATCCGCTTCGAGCGCACGGCA* **524099**

GGTGTGC-3'

524100 *TGCAGCGGTTGCTGTGGGGCGT* **524121**
524149 *CGCGCTGGGCCGTGAAGCCCGACCGGCGGCGGCGCGACCGGGCAAGTCC***524198**
524199 *TGCCGCGGCGTCCGCTTCCGCCGGGGCTGCCTCCAGCTCGGCCGCGGTGA***524248**

D)

524400 *CACGCGCTGCACCGCCCCCCACCTCACCACCTACGCCTCCACGCCCGAGT* **524449**
524450 *CGCTGCCCACGCTGCTCAGCAGCCTGGGCAACACCGGCCACCGGCCGCCG***524499**
524500*CCGGCATGGCTGCAGTCGGCGTGCGCCGCGGCGGCGCCGCACCTGCCCTC***524549**
524550 *CTACAGCTCGGCGCAGCTGAAGGCGCTGGCGGCGGGCCTGTGCGCGATGC***524599**

Primer "*TBC2B* reverse", template strand underlined

3'-GTGGACGGGCTGCTGGGGAC-5'

524600 *GGCACCTGCCCGACGACCCCTGGGTGGCGGCCTACCTGGGCGCTAGCGCG***524649**
524650 *CAGCTGCTGCCCGGCTACAGCGCGGCGGAGCTAACAGTGACCATCAGCTC* **524699**
524700*GCTGGCGGGGCTGGGCTGCCGGCCGGGCGACGAGTGGATGGAGGGCTTCT***524749**
524750*ACGCCCGCGCCACTGCGGTGGTGGGCACCACCGGCGGGCTCACGGGGCCG***524799**

Fig. 18. Legend on next page.

Fig. 18. Primer hybridization sites on the *C. reinhardtii* nuclear genome. Primers used to amplify the *TBC2* paralogs are hybridized to the following regions on the nuclear genome. Hybridization sites are underlined in the template strand or non-template strand where indicated, whereas actual exon nucleotide sequences are italicized. Scaffold numbers according to the DOE JGI V.2.0 *C. reinhardtii* Nuclear Genome Database are bolded. The primers and their respective regions of hybridization are as follows: A, the "*TBC2A* forward" hybridization site in the non-template strand; B, the "*TBC2A* reverse" hybridization site in the template strand; C, "*TBC2B* forward" hybridization site in the non-template strand; D, "*TBC2B* reverse" hybridization site in the template strand.

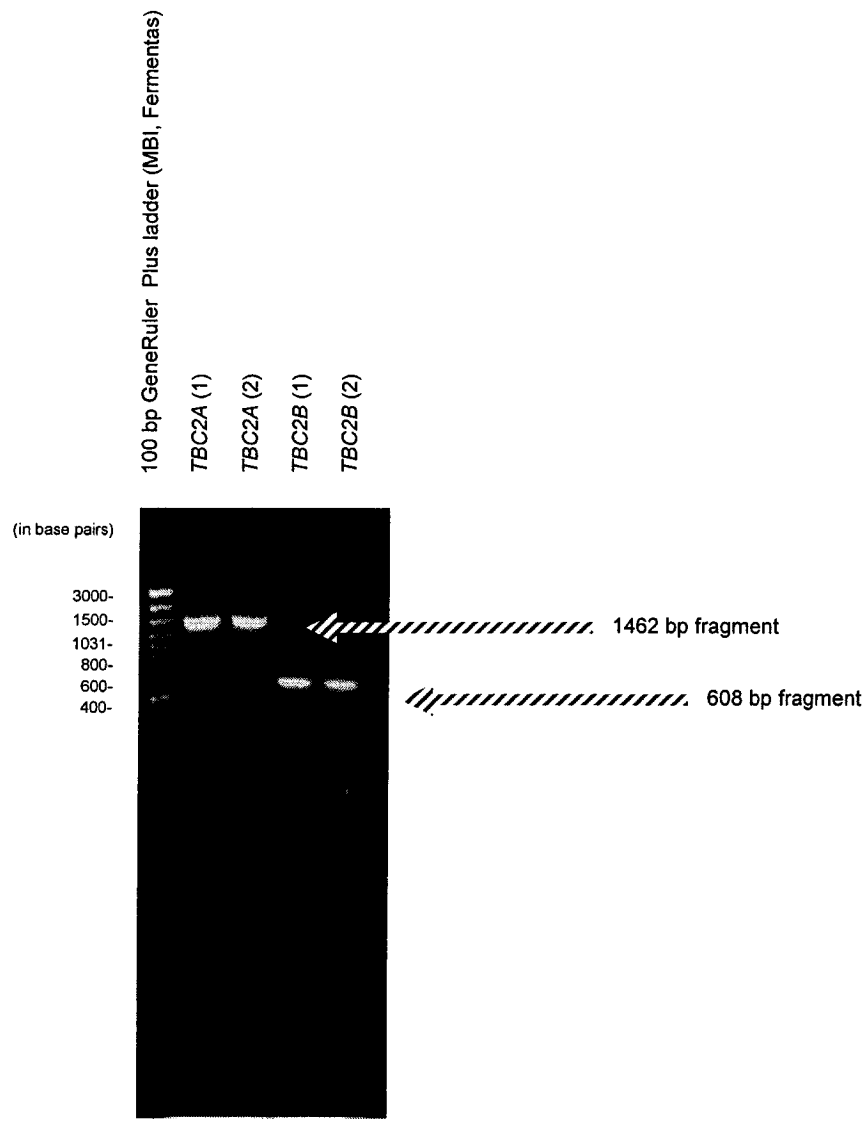

Fig. 19. PCR amplifications of *TBC2A* and *TBC2B*. A 0.8% agarose gel of PCR amplifications of *TBC2A* (1462 bp) and *TBC2B* (608 bp) which were later extracted and gel-purified.

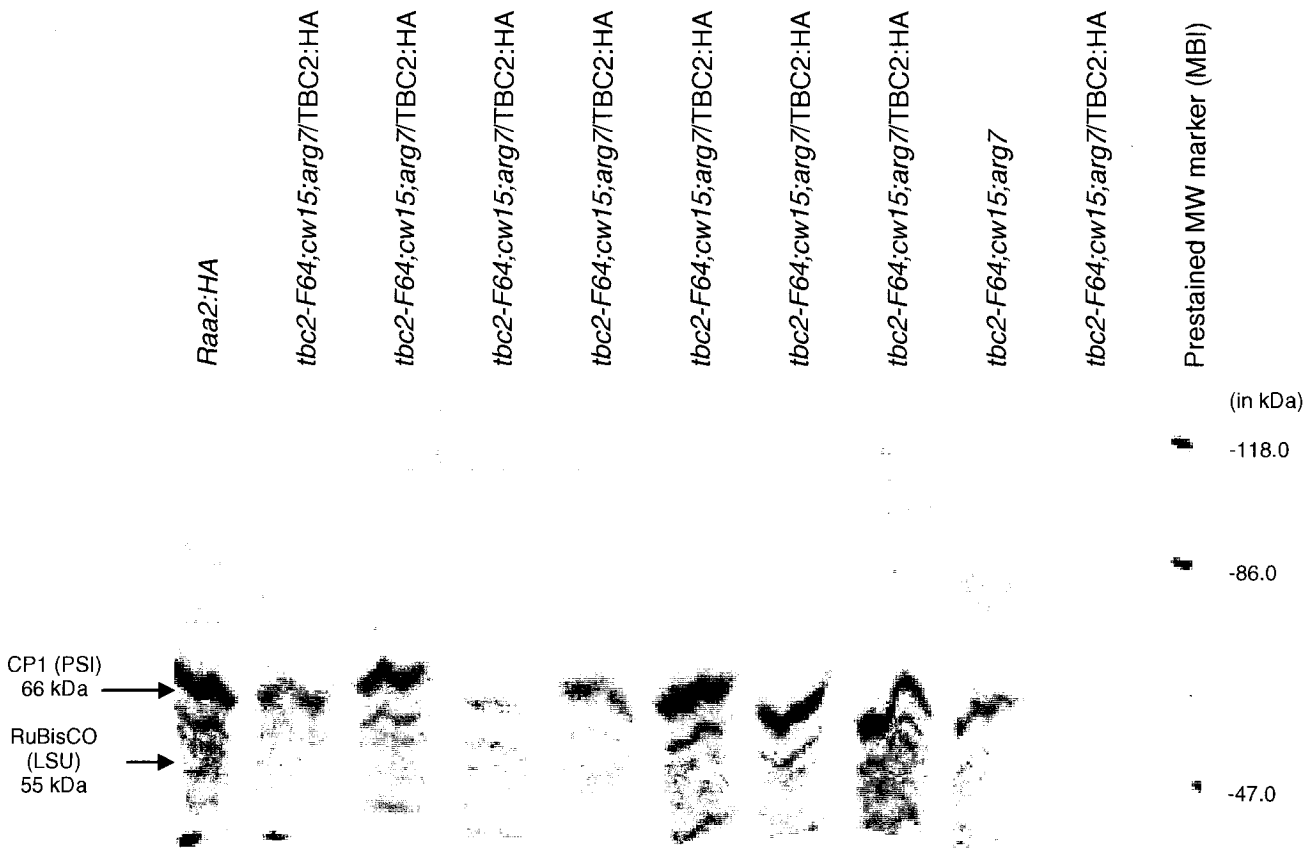

Fig. 20. Ponceau S stain of soluble extract transfer of *tbc2-F64;cw15;arg7* transformants with the TBC2:HA construct.

This Ponceau S stain was done to verify protein transfer. All proteins transferred to nitrocellulose appear pink. All lanes contain samples from different transformants/controls. According to the stain, transfer was successful. The majority of molecular weight markers ran off due to the low percentage of the gel.

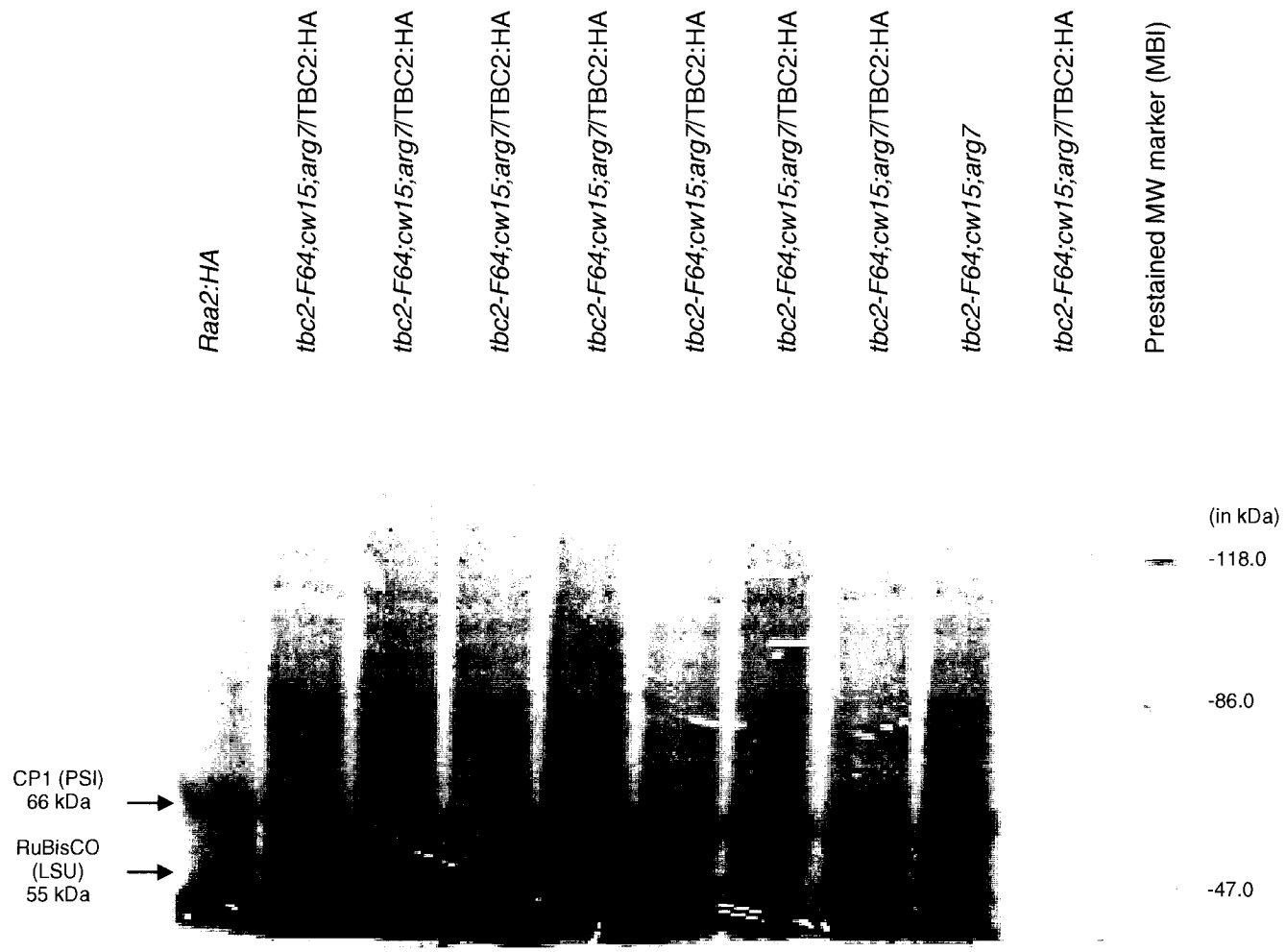

Fig. 21. Ponceau S stain of cell extract transfer of *tbc2-F64;cw15;arg7* transformants with the TBC2:HA construct.

This Ponceau S stain was done to verify protein transfer. All proteins transferred to nitrocellulose appear pink. All lanes contain samples from different transformants/controls. According to the stain, transfer was successful. The majority of molecular weight markers ran off due to the low percentage of the gel.

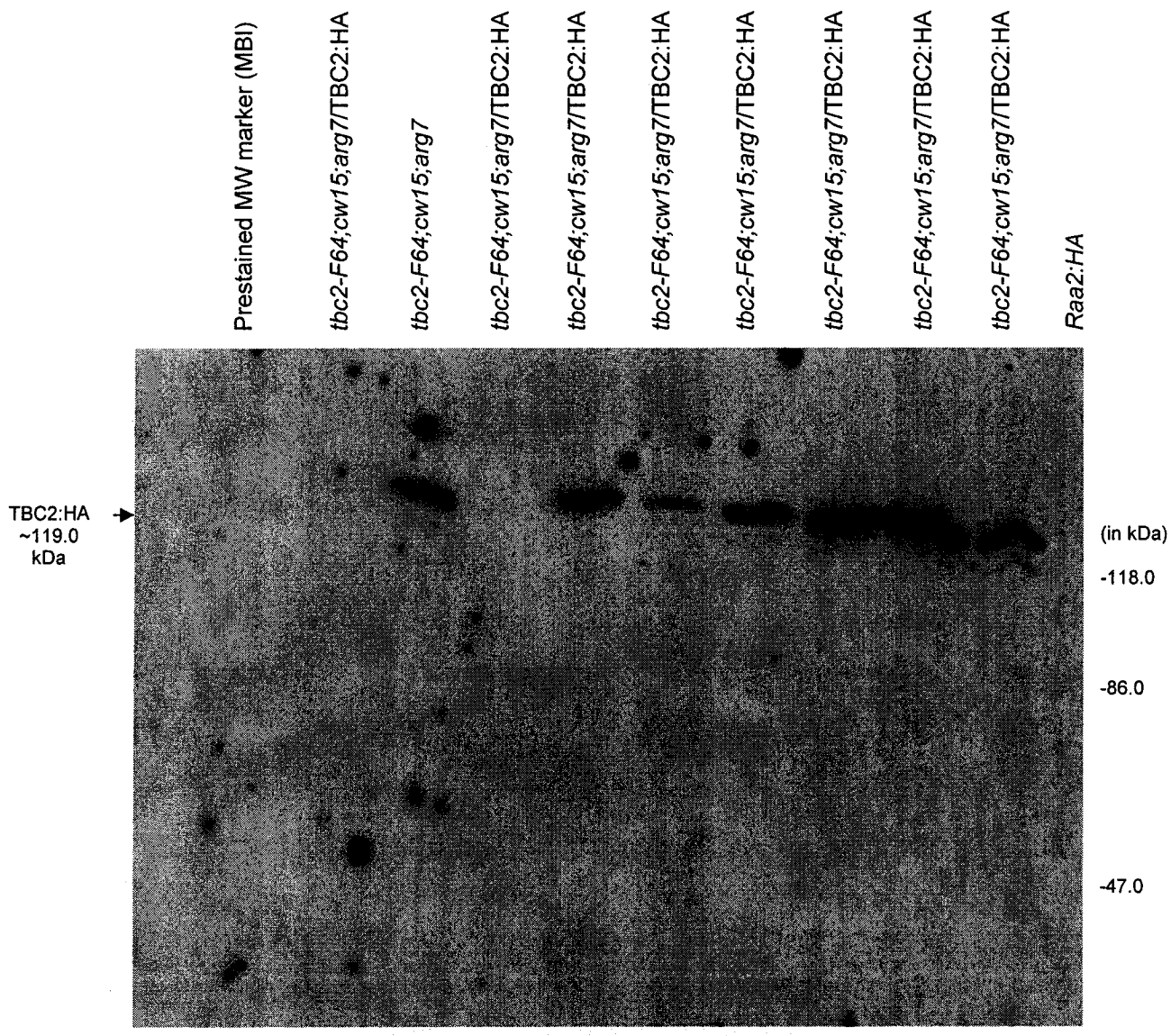

Fig. 22. Soluble protein blot and exposure. The Western blot was exposed to x-ray film by autoradiography according to the ECL protocol. The TBC2:HA protein is evident according to the bands present at ~119.0 kDa, the expected mass of TBC2:HA. The majority of proteins ran off the gel because of its low percentage. The Raa2:HA protein did not appear.

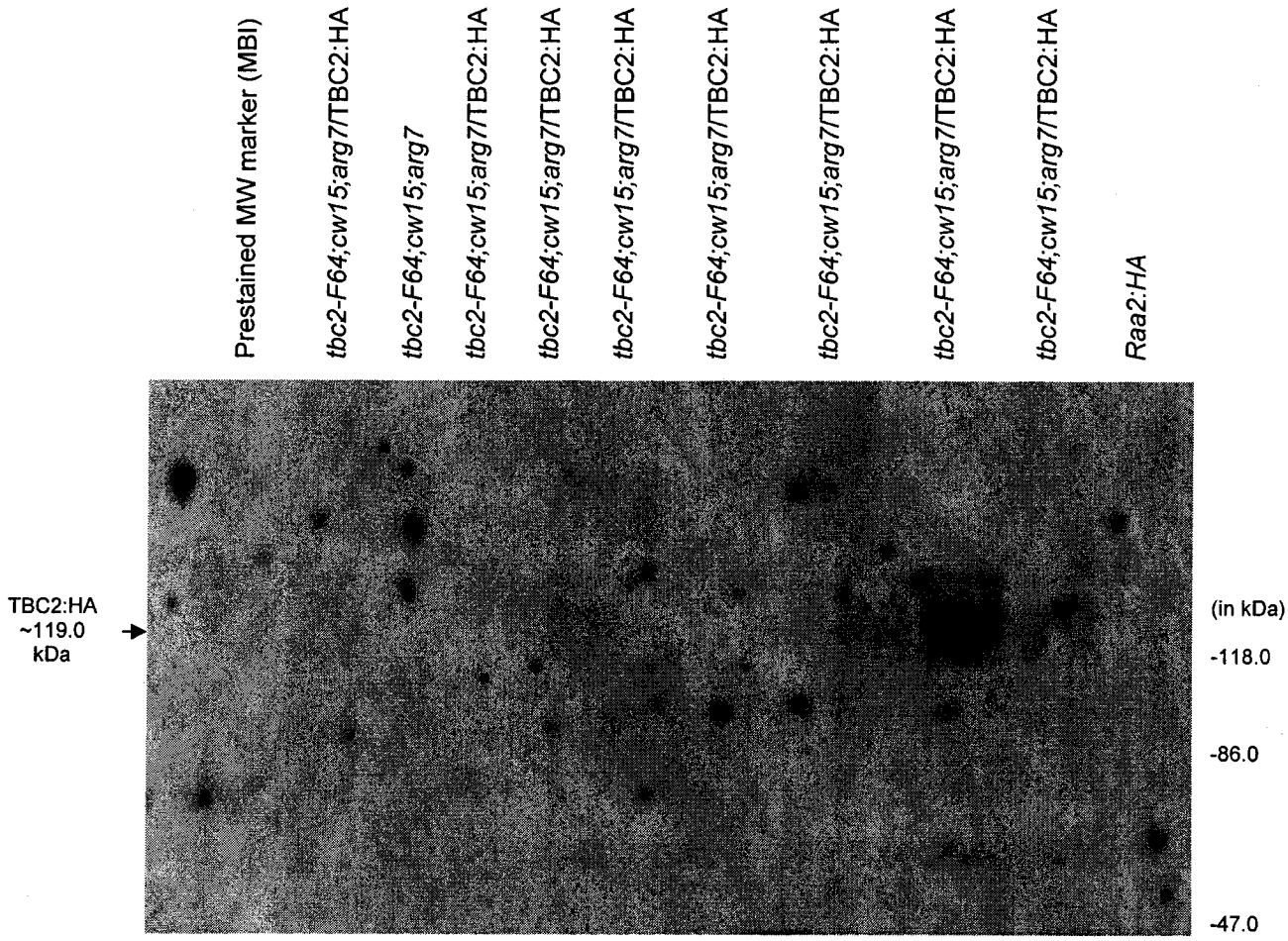

Fig. 23. Cell extract blot and exposure. Cell extract Western blot exposed to x-ray film by autoradiography according to the ECL protocol. The TBC2:HA protein is present in minimal amounts, indicating that TBC2p appears in the soluble fraction due to the nature of the protein. The volume of cell extract may have also played a role in a minimized signal. The majority of proteins ran off the gel because of its low percentage. The Raa2:HA protein did not appear.

3.2.5. Progeny of *Tab2;cw92* Crosses

The *Tab2* mutant, *F14 (mt^-)*, was crossed with *CC503;cw92 (mt^+)*, the possible double mutants allowed to germinate and sporulate. One double mutant strain was found as *Tab2;cw92.* This strain, due to its cell wall-deficiency, could now be transformed with the Tab2:HA construct to be used as a positive control for the coimmunoprecipitation experiment.

3.3. The possible existence of a novel thylakoid biogenesis compartment

Strains were pulsed in $[^{35}S]H_2SO_4$ for different lengths of time (ie. 5 minutes, 10 minutes and 30 minutes) to determine whether dark growth conditions block the maturation of inner envelope membranes. It is characteristic of proteins belonging to low-density envelope-like membranes to be prone to aggregation, unlike proteins of the thylakoid membranes. Dark growth conditions are proposed to block the maturation of low-density envelope-like membranes. This would result in an increase in aggregates in comparison to levels obtained in light-grown samples since maturation of low-density envelope-like membranes is theorized to occur by vesicle formation, resulting in the fusion of these membranes with the thylakoid membranes. Figure 24 shows an autoradiograph of pulse-labeling times in the presence of $[^{35}S]H_2SO_4$ for 5 and 30 minutes. Results show no difference in aggregation between light and dark-grown samples, indicating no blocking of maturation of thylakoid membranes (see Fig. 24). It can also be noted from the gel that D1 as well as RuBisCO proteins in the film have been induced, RuBisCO at a molecular weight of 55 kDa, D1 being 32 kDa. Both proteins have been equally synthesized in light and dark-grown

samples.

Figure 25 shows the same outcome, with the exception that an intermediate pulse-labeling time of 10 minutes was also performed. This was done so that an intermediate signal between that obtained for 5 minutes and 30 minutes could appear in the autoradiograph to show time-sensitive differences in aggregation.

Figure 26 shows an autoradiograph of mutant strains grown in light and dark conditions used to note possible differences in synthesis of thylakoid proteins. Again, no differences appeared in protein synthesis between light and dark-grown samples.

3.4. Localization of the chloroplast genome

Results for the digest of the DNA found in the low density membranes (2) showed a smear in the agarose gel, indicating digested fragments of similar lengths in contrast to what is expected for the chloroplast genome (see Fig. 27). A digest of the chloroplast genome would result in fewer distinct bands due to the relatively small size of the genome in comparison to the nuclear genome (see Discussion, Section 4.4). These results define expectations of a digest of nuclear genomic DNA and it was therefore concluded that this DNA was not chloroplast genomic DNA (see Fig. 27).

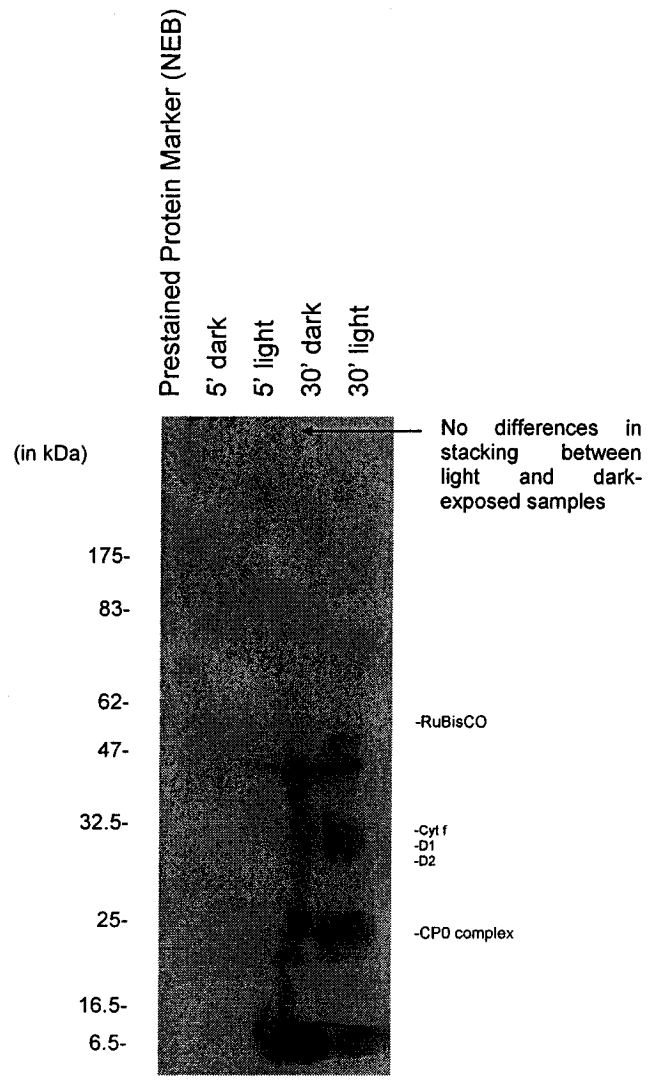

Fig. 24. Autoradiograph of $[^{35}S]H_2SO_4$ pulse-labeling gel shows no difference in protein synthesis between light and dark-grown samples of the *cw15wt* strain. It can be seen that no notable differences in protein synthesis have occurred between light and dark-grown samples exemplified in equal stacking of aggregates at the top of the gel (both at 5 minutes and 30 minutes). The heavy band at approximately 6.5 kDa is unincorporated $[^{35}S]H_2SO_4$.

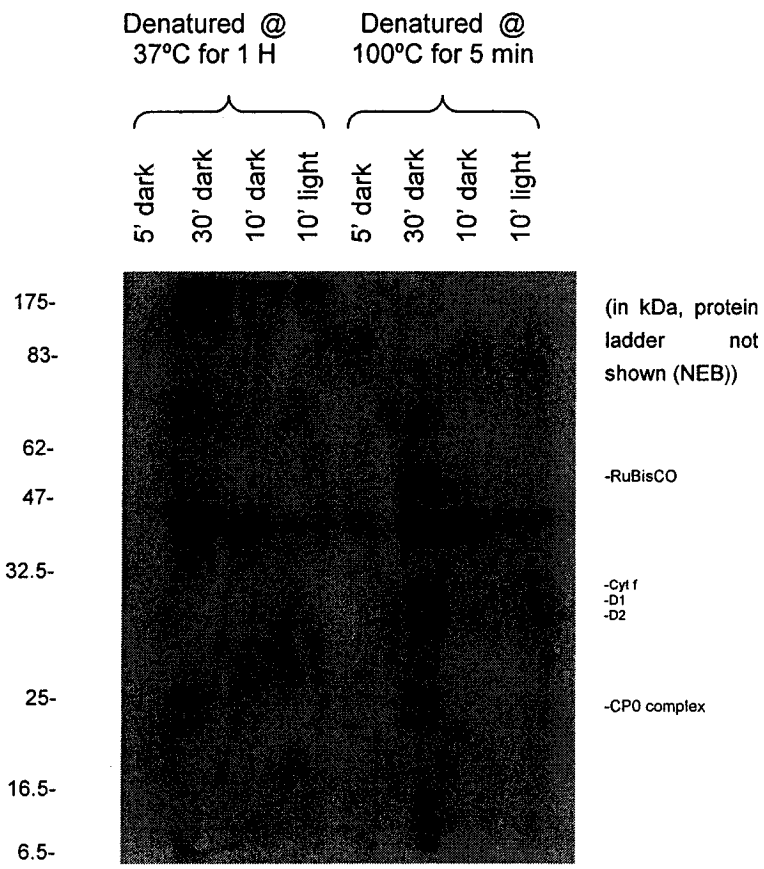

Fig. 25. Autoradiograph of [^{35}S]H_2SO_4 pulse-labeling gel shows no difference in protein synthesis between light and dark-grown samples of the *cw15wt* strain. It can be seen that no notable differences in protein synthesis have occurred between light and dark-grown samples for the *cw15wt* strain, apparent in the relative stacking of aggregates at the top of the gel (applies to 5 minute, 10 minute and 30 minute pulse-labeling samples).

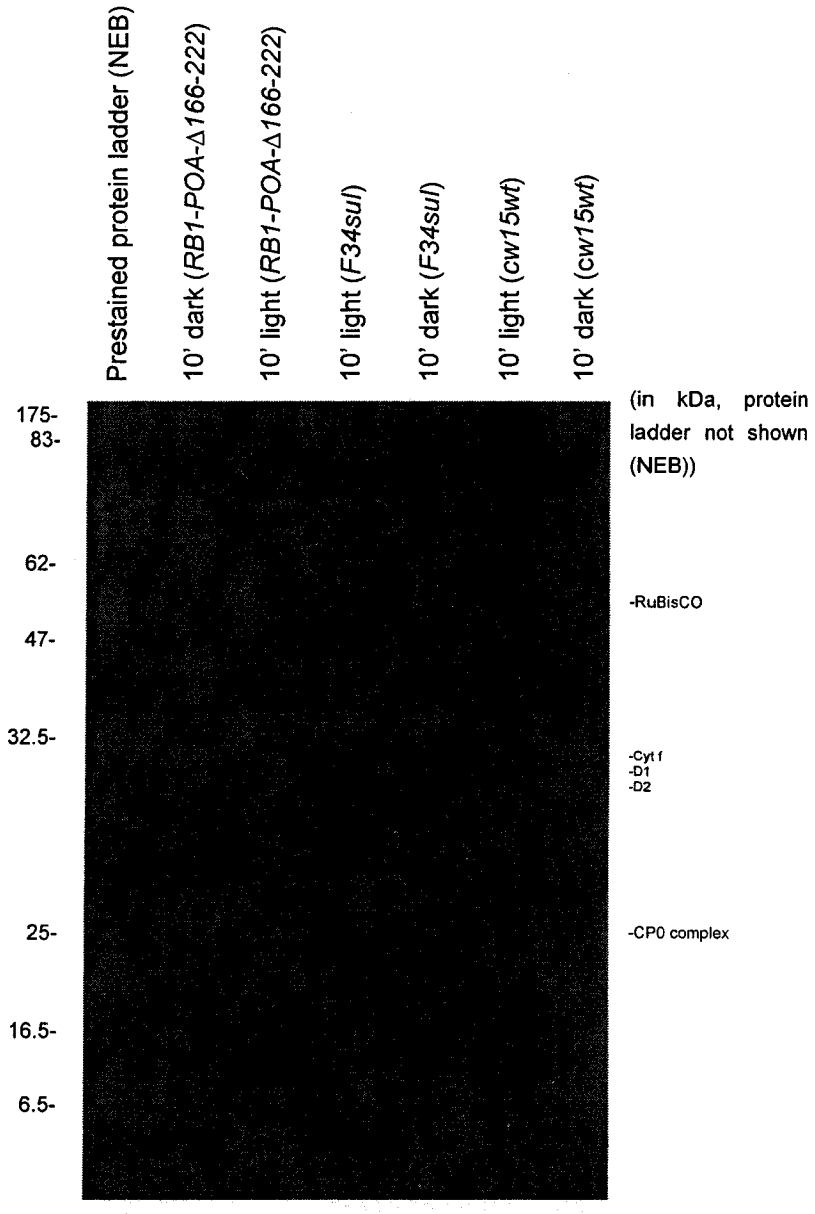

Fig. 26. Autoradiograph of [^{35}S]H_2SO_4 pulse-labeling gel shows no difference in protein synthesis between light and dark-grown samples in mutant strains. It can be seen that no notable differences in protein synthesis have occurred between light and dark-grown samples for different genotypes observed in relative stacking levels of aggregates for each sample.

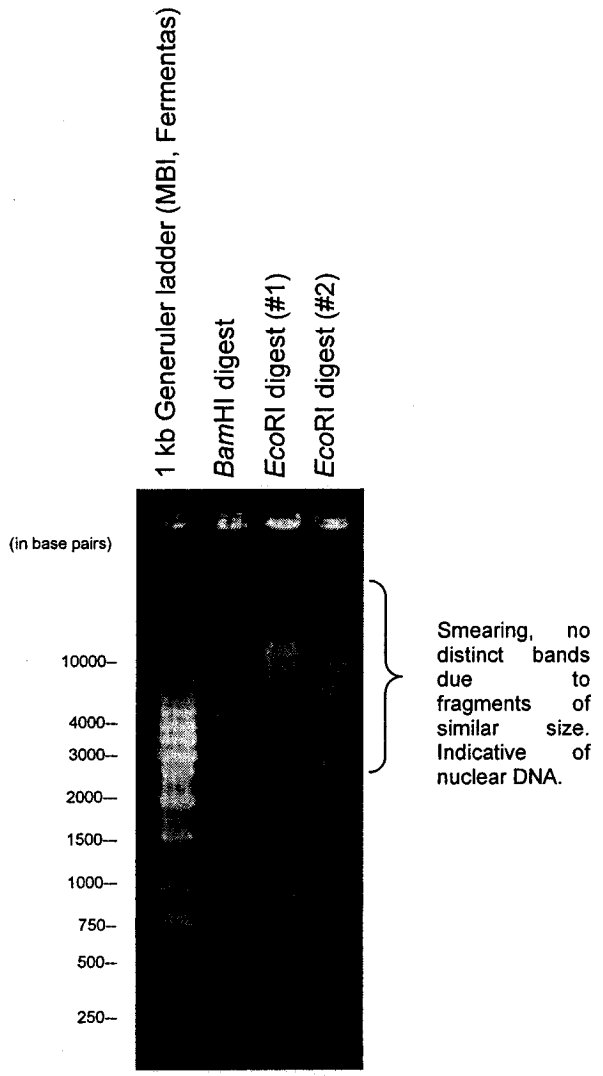

Fig. 27. Agarose gel of possible chloroplast genomic DNA digest. Digestion of DNA found in low density membranes by *Bam*HI (lane 2) and *Eco*RI (lanes 3 and 4). Smears are seen with no distinct fragments indicating that this is nuclear genomic DNA and not chloroplast genomic DNA.

4.1 Bioinformatics analysis of TBC2p and its paralogs

4.1.1. Summary of findings using bioinformatics analysis

A bioinformatics-based approach was used to determine existing paralogs of *TBC2* in the nuclear genome of *C. reinhardtii*. The purpose was also to determine unknown defining features of TBC2p and its paralogs.

The results obtained in the last section will be discussed with regards to their pertinence to the subject area and future directions which should be taken. They are as follows:

-Summary of findings using bioinformatics analysis

-DOE JGI *C. reinhardtii* Genome Database Version 2 Translated BLAST of *TBC2* cDNA

-Comparisons of TBC2p, TBC2Ap, TBC2Bp and Crp1p sequences

-Determination of the presence of transit peptides and organellar localization of TBC2p and its paralogs

-TBC2, TBC2A and TBC2B amino acid sequences, their characteristic PPPEW repeat and the prevalence of amino acid strings

-Predicting the function of TBC2p and its paralogs through analysis of conserved domains

4.1.2. DOE JGI *C. reinhardtii* Genome Database Version 2 Translated BLAST of *TBC2* cDNA

TBC2A and *TBC2B*, located on scaffolds 9 and 10 of the *C. reinhardtii* nuclear genome, were predicted to be paralogs of *TBC2* because their paralogous spans were within range of their respective predicted ORFs (see Table 10) (DOE JGI tBLASTn; ORF Finder). Paralogy was also attributed to the presence of a novel degenerate PPPEW repeat found, until now, only in the TBC2p sequence (see Figs. 10, 11 and 12, Table 9) (DOE JGI tBLASTn; NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]). The TBC2p sequence was previously registered in the DOE JGI V.2.0 *C. reinhardtii* Nuclear Genome Database to be located on scaffold 8 (see Figs. 9 and 12, Table 8). Sequences with significant alignment, especially caused by a novel degenerate repeat, can be assumed to be paralogs.

Scaffold 1767 was previously determined to contain a segment of the TBC2 amino acid sequence. There are many reasons as to why this conclusion was drawn, the first being that the homologous span is 100% identical to the corresponding region in the TBC2p sequence. This fragment is exactly the same length as the fragment in the TBC2p sequence between nucleotides 1176834 and 1176959 of scaffold 8 of the second version of the DOE JGI *C. reinhardtii* Nuclear Genome. This information suggests that this fragment was sequenced as scaffold 1767 due to error in assembly, that is, this fragment was not assembled in correct order according to the actual genomic sequence. Assembly error in this case can easily be explained by the fact that this second version of

the DOE JGI *C. reinhardtii* Nuclear Genome is a pre-publication draft, probably containing errors (L. Peeters, email comm.). The only way of circumventing this problem is by sequencing the genome eight to ten times, which is what is normally done before publishing a sequence to exclude the possibility of errors like redundancy and gaps between contigs (Gardner *et al.* 2005). The theory of error in assembly is also supported by the fact that scaffold 1767 contains no previously determined loci to date and therefore cannot be a paralog (DOE JGI V.2.0 *C. reinhardtii* Nuclear Genome Database). EBI's InterProScan (http://www.ebi.ac.uk/InterProScan/) did not indicate any conserved domains within this fragment.

4.1.3. Comparisons of TBC2p, TBC2Ap, TBC2Bp and Crp1p sequences

In terms of comparative genomics, analysis was done through alignments of the TBC2p sequence with the TBC2Ap and TBC2Bp sequences. Considerable identity was revealed in Figure 13 in a ClustalW alignment, suggesting that TBC2Ap and TBC2Bp, coded for by putative ORFs shown in Table 10, may be evolutionarily divergent proteins of the TBC2p protein (ORF Finder). The regions of paralogy of the TBC2p sequence and its paralogs are mostly in the regions of the PPPEW repeat and it can be concluded that the PPPEW repeat contributes to paralogy between sequences (see Fig. 13, Table 9) (ClustalW; DOE JGI tBLASTn; NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]).

Extensive regions of paralogy with low identity were found between the Crp1 protein (NCBI [accession AAC25599, version AAC25599.1 gi: 3289002]) and the TBC2Bp sequence (see Fig. 14) (ClustalW; DOE JGI tBLASTn). This

may suggest that the functional role of TBC2Bp is similar to Crp1p's role as a translational activator protein (see Fig. 14).

4.1.4. Determination of the presence of transit peptides and organellar localization of TBC2p and its paralogs

Since Auchincloss and colleagues (2002) had previously discovered that the TBC2 amino acid sequence contained a possible transit peptide targeting the chloroplast using ChloroP 1.1, the same method was applied again to the TBC2p sequence as well as its paralogs. Results again confirmed using ChloroP 1.1 and TargetP 1.1 that TBC2p (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) contains a transit peptide resembling a mTP from higher plants and that it is 20 amino acids long. TBC2Ap was also shown to resemble a cTP by ChloroP 1.1 and TargetP 1.1 because of its similarity to a mTP in a higher plant. TBC2Bp had a slightly lower prediction for a mTP in a plant network and a cTP in a non-plant network. Figures 9, 10 and 11 highlight the transit peptide sequences discussed above (ChloroP 1.1; DOE JGI tBLASTn; TargetP 1.1).

The next step in determining whether these are actual transit peptides would be experimentally since ChloroP 1.1 and TargetP 1.1 are computer programs well suited to determine transit peptides within algal nuclear sequences. This could be done by performing protein sequencing by Edman degradation according to the method of Coligan and colleagues (2001) on *in vitro* translated and *in vivo* translated TBC2p, TBC2Ap and TBC2Bp to determine what amino acids were cleaved off after transit. However, Auchincloss and colleagues (2002) already discussed problems with Western blotting the TBC2:HA protein since

results showed a protein 21 kDa higher in mass than expected. Therefore, until it can be determined what accounts for the increase in mass, protein sequencing may lead to unexpected and false results since minimal changes in mass are being measured.

4.1.5. TBC2, TBC2A and TBC2B amino acid sequences, their characteristic PPPEW repeat and the prevalence of amino acid strings

A ClustalW alignment was done of the PPPEW repeat sequences within the TBC2p (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) sequence as well as those in the TBC2Ap and TBC2Bp sequences (DOE JGI tBLASTn) (see Fig. 15). The objective was to determine the relative number of identities and paralogy of these sequences so that their contribution to the paralogy of the full-length protein sequences could be determined. ClustalW alignments are automatically arranged in a hierarchal order according to the sequences of highest homology. As can be seen in Figure 15, the three-most conserved repeats are from the TBC2p sequence (NCBI [locus CAD20887, version CAD20887.1 gi: 2219636]). None of the amino acids in any of the repeats show 100% identity except for P and W in positions 33 and 37 which were strictly conserved. Generally, minor conservation existed regarding amino acid properties (see Fig. 15) (ClustalW). The consensus sequence of PPPEW itself is poorly conserved and was derived from an alignment of the TBC2p PPPEW repeats where, again, the 33rd and 37th amino acids were the only residues which showed 100% identity (Auchincloss *et al.* 2002). Perhaps the 33rd and 37th amino acids are conserved because they are crucial in the structure

which PPPEW may propagate and their substitution would disrupt the structure/function of the protein. Proline is required for β-turns and the lack thereof may prevent domains from forming a compact structure or, if required for tunnel structure, may prevent specific interaction of the protein with mRNA (Creighton 1993). Tryptophan has a bulky side chain and may therefore disrupt protein conformation if replaced since the volume it would occupy would be great (Creighton 1993). Similarity in this situation could be drawn to the PPR repeat in *A. thaliana* whereby the "A" helix is conserved, implying that it interacts or binds to a specific substrate (Williams and Barkan 2003). This example supports the hypothesis that the 33rd and 37th amino acid of the PPPEW sequence in TBC2p (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) as well as TBC2Ap and TBC2Bp may also fulfill a specific function since identity has been conserved (DOE JGI tBLASTn).

Since the PPPEW repeat is novel, little is known about it except that it runs in the region of TBC2p homologous to the PPR-prevalent region in Crp1p, suggesting functional similarity (see Introduction, Section 1.B.3, Discussion, Section 4.1.3.) (see Figs. 6, 9, 10, 11 and 14) (Auchincloss *et al.* 2002). Existing homology between the PPR region of the Crp1 protein sequence and TBC2 and TBC2B protein sequences was reaffirmed in this study (see Fig. 14) (ClustalW; DOE JGI tBLASTn; NCBI [accession AAC25599, version AAC25599.1 gi: 3289002]; NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]). PPR proteins, especially prevalent in *A. thaliana*, form a channel by amphipathic α-helices (see Fig. 7) (see Introduction, Section 1.B.4.) (Blatch and Lässle 1999;

Small and Peeters 2000).

Mili and Pinol-Roma (2003) showed that LRP130, a human PPR-containing protein, binds to mRNA and that when these PPR motifs were removed in deletional analysis in the region of the C-terminus, RNA-binding activity was nearly abolished. Due to this lowering of binding specificity caused by deletion mutations, it was concluded that PPR repeats in the C-terminus constitute and are required for an independent RNA-binding domain (Mili and Pinol-Roma 2003). Nakamura and colleagues (2004) wrote that the HCF152 protein in *A. thaliana*, a nucleus-encoded stromal protein responsible for *petB* splicing, contains 12 PPR repeats, two of which are required for RNA-binding. Nakamura and colleagues (2004) proposed that the PPR repeats contribute to the crescent shape characteristic of helical repeat proteins that the HCF152 protein may possess.

The Puf motif in Pumilio forms a concave surface required for protein-RNA interactions, the same concave surface which is theorized to be required in most helical repeat proteins for protein-protein interactions (Nakamura *et al.* 2004). It was mentioned that the Puf repeat, a helical repeat like the PPR and TPR repeat, appears throughout the Pumilio protein sequence and that every Puf repeat recognizes its own RNA base, defining the specificity of Pumilio towards certain RNAs (Nakamura *et al.* 2004). This was determined by mutating the three side chains within a Puf repeat which bind to an mRNA base (Wang *et al.* 2003). It is mere speculation but the PPPEW repeat, if part of the helical repeat family, may also have structure and function similar to that of Puf, involved in specific

interaction with a singular RNA base in the 5'-UTR of *psbC* mRNA (see Fig. 16).

Strings of alanines, glutamines and serines were previously found to exist in the TBC2 amino acid sequence by Auchincloss and colleagues (2002) (see Introduction, 1.B.2.) (see Fig. 9). Boudreau and colleagues (2000) as well as Vaistij and colleagues (2000) found this to be a characteristic of many nucleus-encoded proteins that function in chloroplast gene expression in *C. reinhardtii.* Often, this abundance appears in the first thirty amino acids from the N-terminus, characterizing the protein as containing a cTP, according to PSORT's criteria (http://psort.nibb.ca.jp). In the present analysis of the TBC2 amino acid sequence, strings of alanines, serines and glutamines were found and outlined in the TBC2 amino acid sequence (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) shown in Figure 9, reaffirming previous findings made by Auchincloss and colleagues (2002).

There are many reasons as to why strings of amino acids like alanines, serines and glutamines exist. These strings are fairly common, for example, a string of eleven serine amino acids in an NCBI protein BLAST for short, nearly exact matches (http://www.ncbi.nlm.nih.gov/BLAST) revealed 587 matches, a string of glutamines revealed 1427 matches, a string of prolines revealed 730 matches, a string of alanines revealed 652 matches and a string of glycines revealed 730 matches in Viridiplantae. The most likely purpose that these amino acids serve is that they are structural spacers between domains, required for separation so that the domains could either interact with each other or a substrate (K. Gehring, pers. comm.). Prolines and glycines achieve opposite

effects, prolines creating a rigid backbone minimizing movement whereas glycines create a flexible backbone allowing movement (K. Gehring, pers. comm.). Alanines and serines have small side chains and therefore do not sterically hinder each other, allowing for flexibility as well (K. Gehring, pers. comm.). Howard and colleagues (2004) stated that polyserine linkers often occur in eukaryotic protein sequences and that, along with polythreonine, polyproline and polyglycine linkers, are hypothesized to serve as structural spacers. These linkers are especially important if the translational activator protein interacts directly with *psbC* mRNA as the substrate and that the protein is required to bind specifically and tightly.

Depending on the location of the polyamino acid linker in the protein, the linker may generate varying secondary structures according to the hydrophobicity of its surroundings (Nguyen *et al.* 2004). For example, it was demonstrated using an off-lattice intermediate-resolution protein model and discontinuous molecular dynamics simulation that varying the hydrophobicity surrounding a polyalanine stretch caused it to generate either β-turns, coiled-coils, β-sheets or α-helices (Nguyen *et al.* 2004).

If not for the purpose of a linker, there are other reasons as to why these strings of amino acids exist. Results in this present study confirming the presence of strings of serines, alanines and glutamines in the translational activator protein TBC2 (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) support similar findings with the hypothesis being that these strings are often a necessity in nucleus-encoded translational activators. Vaistij and

colleagues (2000) discussed that conserved alanines exist in the Mbb1 nucleus-encoded protein required for mRNA accumulation. These alanines are present in Mbb1p's TPR repeats, previously described in Section 1.B.4 (see Fig. 7) (Vaistij *et al.* 2000). Boudreau and colleagues (2000) explained that TPR alanines are essential for the proper folding of the protein, demonstrated when a substitution mutation made at ALA 1038 for GLU within a TPR repeat destabilized the protein. These findings reflect the importance of alanines in protein conformation, especially when alanines are hydrophobic and are often found in the interior of proteins because of their small volume (Creighton 1993).

Auchincloss and colleagues (2002) discovered through size-exclusion chromatography that TBC2p associates with other proteins in a 400 kDa complex (see Introduction, Section 1.B.1.). Therefore, certain domains of the TBC2 protein may be reserved for protein-protein interactions, and alanine, because of its advantageous properties, may fulfill a functional role in these domains. Rattanachaikunsopon and colleagues (1999) found that the *ac115* nuclear gene (see Introduction, Section 1.A.6. for explanation on the *ac* loci) codes for hydrophobic amino acids at its C-terminus such as alanines, proposed to constitute a membrane-spanning domain or be required for protein-protein interaction (Rattanachaikunsopon *et al.* 1999).

Funes and colleagues (2002) also discussed amino acids like alanines and serines comprising secondary structure domains. Funes and colleagues (2002) reported that the *atp6* gene codes for the F_1F_0-ATP synthase protein which has invariant amino acids exemplified by SER^{141} and ALA^{152}. Funes and

colleagues (2002) proposed that these amino acids are required in a translocation domain helix.

De Vitry and colleagues (1996) explained the importance of alanines and serines in cTPs using the nucleus-encoded *PetM* gene coding for the cytochrome b_6f subunit in *C. reinhardtii*. Their existence in this cTP supported the observation that their presence is a trademark of many *C. reinhardtii* cTPs (Franzén *et al.* 1990). Krimm and colleagues (1999) also showed the requirement of alanines in α-helical cTP domains. This was done using a synthetic transit peptide 32 amino acids in length of the RuBisCO protein which formed amphipathic α-helices; alanines were prevalent in this peptide, comprising the hydrophobic side of the helix (Krimm *et al.* 1999). Two specific alanines shown to be part of this side were ALA^{21} and ALA^{28} (Krimm *et al.* 1999). Franzén and colleagues (1990) wrote that alanines are more prevalent in *C. reinhardtii* cTPs compared to cTPs of other species and mTPs from higher plants. Even though alanine, serine and glutamine strings (ie. more than three consecutive amino acids) are not present in the defined transit peptide stretches of TBC2p, they may still fulfill an unknown role in proteins containing a cTP.

Krimm and colleagues (1999) also demonstrated that cTPs in *C. reinhardtii* undergo a helix-coil to coil-helix transition in 2,2,2-trifluoroethanol. Hydrophobic amino acids such as alanine comprised the domain of the coiled-coil which converted to a helix, demonstrated by UV-CD and NMR, possibly being essential to this transition (Krimm *et al.* 1999). This transition may be functionally required to lodge the cTP peptide into the lipid phase so that it can

interact with translocation machinery (Krimm *et al.* 1999). Alanines may demonstrate a similar unknown function in the remainder of the cTP-containing protein in *C. reinhardtii*, considering that the peptide is translocated through the lipid phase as a whole.

If TBC2p interacts directly with the 5'-UTR of *psbC* mRNA, then it is possible that the strings of amino acids confirmed to exist in this study are required for proper RNA interaction. This was shown by Barnes and colleagues (2004) who demonstrated that the RB38 RNA-binding protein is comprised of basic amino acids (18.2%) as well as alanines which comprise a large proportion compared to all amino acids present (12.4%). If not relevant to chloroplast import proteins, amino acids such as alanines may be required to generate structures, possibly for specific interaction with RNA bases.

Valuable predictions made by Treger and Westhof (2001) determined amino acids which were structurally and property-wise the most likely to interact with RNA. Taking into account atomic contact, serines and alanines belonged to the second most favorable group of amino acids to interact at the RNA-protein interface (Treger and Westhof 2001).

The TBC2Ap paralog, theorized to possess a similar functional role to the TBC2 protein because it contains TBC2p's novel PPPEW repeat, was also shown to contain strings of alanines and serines as well, possibly fulfilling a similar functional or structural role to the same amino acid strings present in TBC2p (NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]) (see Fig. 10). A heptapeptide of seven prolines between amino acids 251 and 257 (from

the N-terminus) is also present in the TBC2Ap sequence (see Fig. 10) (DOE JGI tBLASTn). Prolines have a cyclic five-membered ring which causes kinks in secondary structures such as α-helices and also demonstrate a high propensity towards the generation of β-turns (Creighton 1993). Prolines within the α-helix, when substituted for another amino acid, results in severe effects on the stability of the structure because of proline's rigid backbone, demonstrating the crucial structural role of prolines (Creighton 1993). Strings of prolines can also form what is known as the polyproline helix, which can occur as either right-handed (I) or left-handed (II) helices, depending on the surrounding solvent (Creighton 1993). Prolines, because of their discussed molecular structure, are required for specific secondary structure conformation.

Taylor and colleagues (2001) wrote that prolines in cyanobacteria and plants are essential and conserved in the RuBisCO protein in cyanobacteria and plants. Larson and colleagues (1997) also describe the importance of prolines in RuBisCO after a substitution of proline at position 89 for an arginine affected the specificity of activase in RuBisCO's large subunit. Prolines are present in PPRs and TPRs, this being significant due to the fact that the PPR-prevalent region in the Crp1 protein sequence is slightly homologous to TBC2p (see Figs. 7 and 14) (Auchincloss *et al.* 2002).

The TBC2Bp sequence, also theorized to possess a similar function to TBC2p, is similar to TBC2p in terms of the presence of glutamines, alanines and serines. However, it also possesses strings of glycines (see Fig. 11) (DOE JGI tBLASTn). Because of glycine's small side chain (H), they are required for tight

turns, for example, in the β-turn (Creighton 1993). Glycines tend to create flexibility in the peptide backbone and secondary structures containing glycines are disrupted when this amino acid is replaced (Creighton 1993).

Bowers and colleagues (2003) showed that glycines line the exit tunnel of the chloroplast ribosome. Therefore, glycines may comprise a fraction of amino acids forming tunnels in translational activator proteins, a role theorized in this study for TBC2Bp. De Vitry and colleagues (1999) discovered that the Rieske iron-sulfur protein 2Fe-2S subunit encoded by the *PETC* gene contains six glycines in a string responsible for domain movement. Therefore, reiterating previous the previous theory of polyamino acid linkers, a string of glycines could constitute a hinge required for folded structure, especially when interacting with small molecules like mRNAs.

4.1.6. Predicting the function of TBC2p and its paralogs through analysis of conserved domains

An NCBI Conserved Domain Search using a Reverse Specific Position BLAST was performed on TBC2p, TBC2Ap and TBC2Bp. Even though the following proteins may only have amino acid strings homologous to conserved domains in other proteins, these proteins must be taken into consideration if they are functionally similar to provide insight on the mechanism of the TBC2 proteins. The most significant results pertaining to similar function were obtained with the paralogs (DOE JGI tBLASTn; NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]).

For TBC2Ap, the CAP translational activator in eukaryotes was shown to

be homologous to the TBC2Ap sequence. The CAP sequence in *Schizosaccharomyces pombe* has two domains, the N-terminus which is a domain required for cellular responsiveness to Ras and the C-terminus which, when mutated, can cause pleiotropic effects, including those which are morphological (Kawamukai *et al.* 1992). This indicates that additional functions of the C-terminus exist unrelated to the function of adenylyl cyclase (Kawamukai *et al.* 1992). Unfortunately, the only alignment achieved between CAP and TBC2Ap was with the string of prolines defined earlier in the discussion to exist in TBC2Ap, as well as in some surrounding residues. Nevertheless, this may be a conserved domain of significance for translational activator proteins. This region of homology is located within the centre of the CAP protein sequence. CAP, like TBC2Ap, is also an adenine and serine-rich protein and binds indirectly to mRNA, suggesting that TBC2Ap, a putative translational activator protein like CAP, would bind indirectly to mRNA in a theoretical complex (DOE JGI tBLASTn; NCBI CDD V.2.04 RPS-BLAST).

Another protein of homology found of relevance to the role of TBC2Ap was FAST-LEU-rich protein kinase with a score of 32.1. The reason why this protein is pertinent to the subject is because FAST-LEU-rich protein kinase phosphorylates the nuclear RNA-binding protein TIA-1, previously discovered to trigger apoptosis (NCBI CDD V.2.04 RPS-BLAST). After phosphorylation of the TIA-1 protein through the serine/threonine kinase, TIA-1, in turn, performs DNA fragmentation (Tian *et al.* 1995). "FAST" is an autophosphorylated protein and has an affinity for SH3 domains through its proline strings (Tian *et al.* 1995).

However, recent evidence suggests that phosphorylation is done through an intermediate protein and not the FAST-LEU-rich protein kinase domain (P. Anderson, pers. comm.). The region of homology within the FAST protein shared with TBC2Ap contains two dispersed prolines which is relatively abundant considering that alignment is low. At present, the function of the FAST-LEU-rich protein kinase domain is not known for certain (P. Anderson, pers. comm.). However, FAST-LEU-rich protein kinases have been shown to resemble translational activator proteins, this finding being of relevance if TBC2Ap is an actual translational activator (P. Anderson, pers. comm.) (see Fig. 10) (DOE JGI tBLASTn; NCBI CDD V.2.04 RPS-BLAST).

The third homologous protein to TBC2Ap of functional significance was the U5snRNP spliceosome unit which is responsible for RNA processing and modification in eukaryotes. Even though it is not a translational activator protein, the fact that U5snRNP is a post-transcriptional protein and binds RNA indicates that U5snRNP may be structurally related to TBC2Ap (DOE JGI tBLASTn; NCBI CDD V.2.04 RPS-BLAST).

As for TBC2Bp, a protein of relevance in terms of functionality is from a glycine-rich protein (GRP) family. This could explain the prevalence of glycines in the TBC2Bp amino acid sequence (see Fig. 11). Previous studies have shown that glycine-rich proteins bind single-stranded DNA and have been experimentally shown to be phosphorylated (rev. by Sachetto-Martins *et al.* 2000). It was demonstrated that the glycine-rich domain in the RZ-1 protein is required for RNA-binding activity, and, due to immunolocalization of GRPs, was

hypothesized to be required in RNA processing, maturation and control of gene expression (DOE JGI tBLASTn; NCBI CDD V.2.04 RPS-BLAST; rev. by Sachetto-Martins *et al.* 2000).

Another protein of homology to TBC2Bp was the transcription initiation factor TFIID belonging to the TAF4 family. Human TAFs were shown to form pairs through histones, the histone controlling their interactions (Birck *et al.* 1998). If TBC2Bp is functionally similar to TBC2p in a complex, then this similarity supports this theory (DOE JGI tBLASTn; NCBI CDD V.2.04 RPS-BLAST).

4.1.7. Determining protein characteristics using PROSPECT-PSPP

PROPECT-PSPP predicted the structure of the TBC2p to be somewhat crescent-shaped. This can be expected for translational activator proteins, predicted by Nakamura and colleagues (2004) to occur in PPR helical repeat proteins such as Crp1p, a protein homologous to TBC2p (see Fig. 16). The TBC2 protein and its paralogs are predicted to consist of mostly α-helices and possess three domains. TBC2p has been reaffirmed as being a soluble protein, and TBC2Ap and TBC2Bp have been predicted to be soluble proteins. This information suggests that the paralogs fulfill similar functions to TBC2p (DOE JGI tBLASTn; NCBI [locus CAD20887, version CAD20887.1 gi: 22129636]).

4.2. Experimental approach taken towards determining the mode of TBC2p interaction

Initial steps were taken in a coimmunoprecipitation experiment to verify the expression of the *TBC2* gene with its triple hemagglutinin epitope in the "HA @ stop in KBam10 (ie. genomic)" plasmid (see Fig. 17). This was done by transforming the plasmid into the *tbc2-F64;cw15;arg7* strain resulting in the strain's complementation so that TBC2:HA protein could be expressed and verified in a Western blot.

Results of the Western blot showed a detected protein of 119 kDa (see Figs. 22 and 23). Ponceau S stains of soluble and membrane proteins are shown in Figures 20 and 21.

The coimmunoprecipitation experiment is to be carried out as follows. The affinity matrix is blocked by *cw15wt* total protein extract, the beads washed and the protein extract from the TBC2:HA transformants incubated with the beads, which are then washed and resuspended in buffer for immunoblotting (Dauvillée *et al.* 2003). Raa2:HA protein would be used as a positive control (see Results, Section 3.2.4.) RNA could then be extracted and used for slot blot analysis to determine whether *psbC* mRNA was bound to TBC2p or its complex (Dauvillée *et al.* 2003).

If the coimmunoprecipitation experiment does not reveal any additional information about the 400 kDa TBC2p complex, then additional experiments would have to be done. This could be accomplished by isolating the 400 kDa complex using size-exclusion chromatography outlined by Auchincloss and

colleagues (2002) and the complex characterized by a 2-dimensional SDS-PAGE denaturing gel (Ausubel *et al.* 2001). This will determine the isoelectric points and molecular weights of the subunits if they dissociate at the same time separate them. Dissociation could be maximized using the reducing agent dithiothreitol (Coligan *et al.* 2001). If these proteins could be separated and identified based on these characteristics, they can be verified by undergoing purification and sequenced using automated degradation and a commercially available protein sequencer (Ausubel *et al.* 2001).

To determine how these subunits interact with the 400 kDa complex, x-ray crystallography could be performed if the complex can be crystallized. The protein complex must be isolated by size-exclusion chromatography with a Sephadex G-10 gel (Coligan *et al.* 2001). Then one can perform x-ray crystallography by resuspending the protein complex in crystallization solution and subjecting it to vapor-diffusion crystallization (Coligan *et al.* 2001).

To determine how the TBC2p complex binds to the 5'-UTR of *psbC* mRNA, RNA-binding experiments could be done. Initially, TBC2p could be ruled out or determined as binding directly with the 5'-UTR of *psbC* mRNA. This could be done by transcribing *psbC* mRNA *in vitro* and purifying it on an agarose gel according to the method of DiCello and colleagues (2005). TBC2p would then be allowed to bind to *psbC* mRNA and the RNA could then be treated with dimethylsulfate chemical modifier, the reaction terminated by adding β-mercaptoethanol, and sodium acetate added to dislodge TBC2p so that RNA could be isolated (Hwang *et al.* 1989). If results reveal no binding, then the

experiment could be repeated with the 400 kDa complex, individual subunits, or a combination of subunits to determine which subunits bind directly with the 5'-UTR of *psbC* mRNA.

PCR amplification of the *TBC2A* and *TBC2B* nuclear genomic sequences was done so that these amplifications could be labeled to probe a cDNA library (see Fig. 18 and 19). cDNAs probed would be cloned into high copy expression vectors so that RNAi experiments could be done according to Rohr and colleagues (2004). If phenotypic differences result, then this could reveal the function of the TBC2p paralogs.

4.3. The possible existence of a novel thylakoid biogenesis compartment

It is theorized that when *C. reinhardtii* cells are grown in the dark that the maturation of inner thylakoid membranes is blocked (see Introduction, Section 1.C). Proteins of the novel thylakoid biogenesis compartment were predicted to appear as aggregates at the top of the gel and dark-grown samples would have an increased amount of aggregates relative to light-grown samples. This was theorized to occur because a block in maturation would result in halted vesicle formation. Figures 24 through 26 show autoradiographs of the pulses performed with $[^{35}S]H_2SO_4$ for either 5, 10 or 30 minutes. All autoradiographs showed no difference in synthesis of these aggregates at the top of the gel, and D1 (32 kDa) and RuBisCO (55 kDa) did not differ in synthesis in light-grown samples compared to dark-grown samples.

Different approaches could be used to discount the theory of the novel thylakoid biogenesis compartment further. Fusion proteins comprised of a green

fluorescent protein and a thylakoid membrane protein could be used to visualize the localization of thylakoid membrane proteins and whether they initially exist in a novel thylakoid biogenesis compartment. This would be done by using a chloroplast transformation vector such as p72B which contains the chloroplast genomic DNA span harboring, for example, *psbA* (Bateman *et al.* 2000). A selectable marker should be used to select for successfully transformed cells (Bateman *et al.* 2000). The *GFP* gene could then be amplified using PCR and cloned into the transformation vector controlled by the 5' and 3' UTR of a gene matching *psbA* codon usage. (Franklin *et al.* 2002). Chloroplast transformation would be performed using a Biolistic particle delivery system (Bateman *et al.* 2000). GFP fusion proteins could be visualized using a confocal laser scanning microscope (Franklin *et al.* 2002). Additional fusion proteins using spectral variants such as yellow fluorescent and cyan fluorescent proteins fused to other genes encoding the thylakoid membrane could also define the initial and subsequent localization of these proteins (Veening *et al.* 2004). If a thylakoid biogenesis compartment does exist and vesicle transport and fusion of the inner envelope membranes into the thylakoid membranes does occur like Westphal and colleagues claimed (2001, 2003), then one could observe this using confocal laser scanning electron microscopy as well.

Another approach to be done conjointly with GFP fusion proteins would be to probe mRNAs co-translated into the inner thylakoid membrane with ^{3}H-labeled riboprobes synthesized *in vitro* in the presence of ^{3}H-UTP and ^{3}H-CTP (Drews *et al.* 1991). This would be performed by the method of *in situ* hybridization and

autoradiography according to Drews and colleagues (1991).

4.4. Localization of the chloroplast genome

DNA found associated with low density membranes after a second centrifugation in the absence of Mg^{2+} to chelate membranes together was subjected to restriction digestion to determine whether the DNA was chloroplast or nuclear genomic. *Bam*HI and *Eco*RI sites in the *C. reinhardtii* chloroplast genome are sparsely scattered and would result in relatively few distinct fragments of varying lengths considering that the length of the chloroplast genome is only 203 kb (see Figs. 27 and 28).

On the contrary, nuclear DNA in *C. reinhardtii* (~10^8 bp) is approximately 500-fold larger than chloroplast DNA and therefore occurs as a smear due to the relative number of DNA fragments after digestion with *Bam*HI and *Eco*RI, hence the results obtained in this section. If the experiment were to be redone, one would have to ensure that the chloroplasts isolated prior to lysis were free of extraneous matter such as nuclear DNA. Percoll gradients (45%-75%) alone are unable to separate chloroplasts from impurities such as cytosolic components completely (Cline lab, pers. comm.). Therefore, it is possible to wash them three times with import buffer (50 mM HEPES/KOH pH 8, 0.33 M sorbitol), centrifuging them at 1500 *g for 5 minutes between each wash according to the ClineWeb protocol for pea chloroplast isolation (http://www.hos.utl.edu/clineweb/protocols/Peaisol.htm). Seigueurin-Berny and colleagues (2000) mentioned that one could wash the chloroplasts in 10 mM MOPS-NaOH pH 7.8, 0.33 M sucrose in the absence of EDTA or Ca^{2+}.

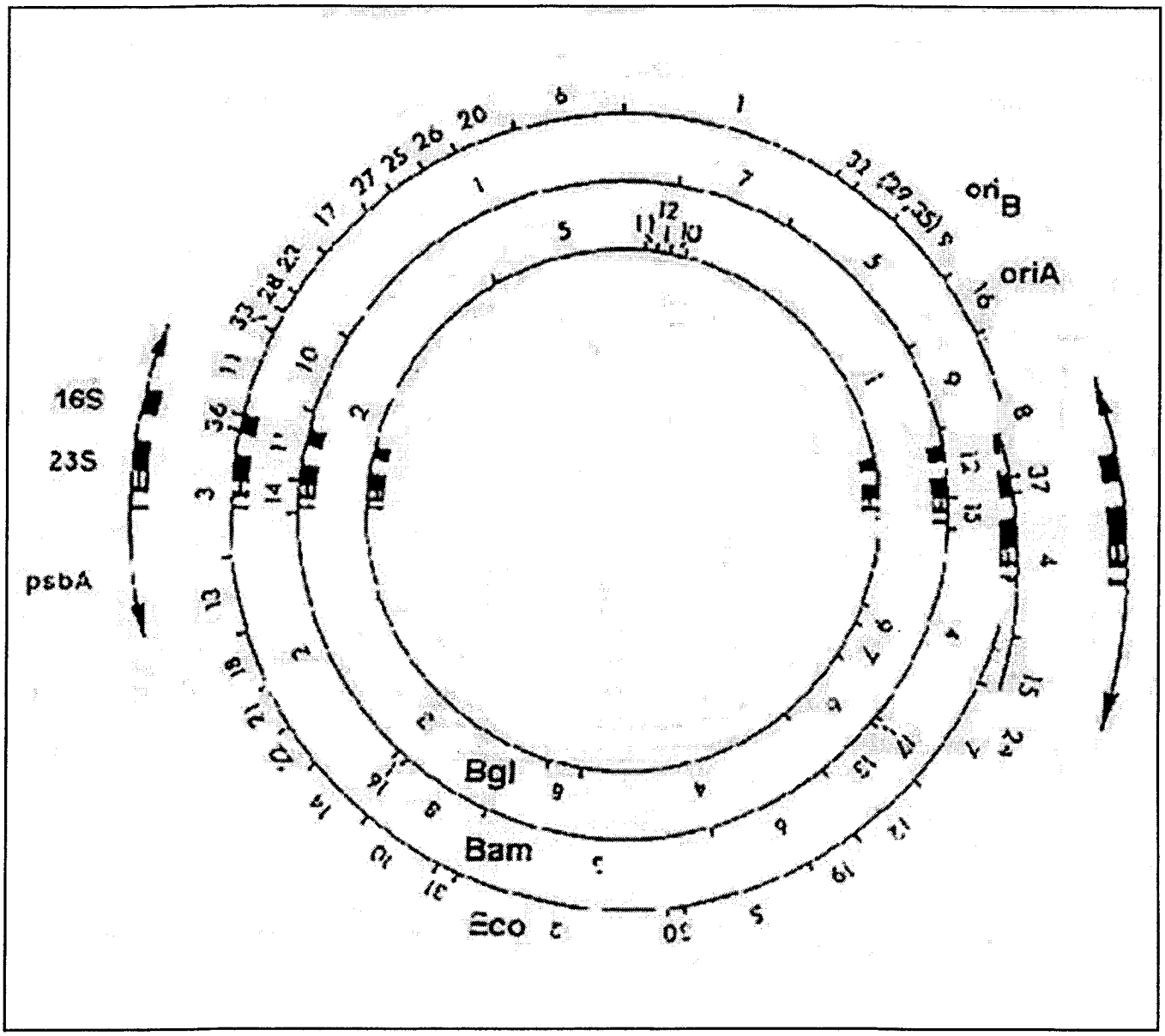

Fig. 28. *Bam*HI and *Eco*RI recognition sequences in the chloroplast genome of *C. reinhardtii.* The restriction map shows the interdistances between restriction enzyme recognition sequences. Run on a gel after a digest with *Bam*HI or *Eco*RI, distinct fragments would result (Sears 2004). (Permission to use of this figure kindly granted by B. Sears and Springer Academic Publishers).

Chloroplasts in *C. reinhardtii* are difficult to isolate because the chloroplast is large, cup-shaped and encompasses the nucleus (E. Harris, email comm.). The above washes described more often than not disrupt the chloroplasts (E. Harris, email comm.). DNA found associated with the second set of low density membranes run the risk of contamination with nuclear genomic DNA. In order to isolate chloroplast DNA from this preparation, sodium iodide gradients could be used to separate chloroplast genomic DNA from nuclear genomic DNA since nuclear genomic DNA differs in density due to a high GC content (E. Harris, email comm.). A sodium chloride gradient could be used instead of a cesium chloride gradient to achieve a better separation (E. Harris, email comm.). A second sodium iodide gradient centrifugation could be done on chloroplast genomic DNA from the original sodium iodide gradient centrifugation to ensure no contaminating genomic DNA is present before restriction digestion (E. Harris, email comm.). The protocol which should be used is according to Grant, Gillham and Boynton (1980) (E. Harris, email comm.).

Since it is theorized that the chloroplast genome is localized, another way of possibly determining this is by performing fluorescence *in situ* hybridization, counter-staining the nucleoid using 4, 6-diamidino-2-phenylindole (DAPI). Armbrust (1998) discussed that 8-10 chloroplast nucleoids exist in the organelle which can be identified using DAPI staining. It was found in bacteria *Vibrio cholerae* through DAPI staining that chromosome I and II were localized near the cell pole and the cell centre, respectively (Fogel and Waldor 2005). This may apply to the chloroplast nucleoids in *C. reinhardtii.* If the nucleoids consistently

demonstrate localization at certain regions of the chloroplast, then this would support the theory that the chloroplast genome is localized, combined with the data obtained from restriction digest of purified genomic DNA from the sodium iodide gradient.

Adair, W. S. and K. E. Apt. "Cell wall regeneration in *Chlamydomonas*: Accumulation of mRNAs encoding cell wall hydroxyproline-rich glycoproteins." *Proceedings of the National Academy of Sciences of the United States of America* 87, no. 19 (1990): 7355-7359.

Alt *et al.* "Nucleotide sequence of the clustered genes for the 44 kD chlorophyll a apoprotein and the '32 kD'-like protein of the photosystem II reaction center in the spinach plastid chromosome." *Current Genetics* 8 (1984): 597-606.

Anderson, P. Brigham and Women's Hospital. Harvard Medical School. FAST-LEU-rich protein kinase and its current function. Pers. comm.

Armbrust, E. V. "Uniparental Inheritance of Chloroplast Genomes." Chap. 6 in *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas*. Edited by J.D. Rochaix, M. Goldschmidt-Clermont and S. Merchant. Vol. 7. The Netherlands: Kluwer Academic Publishers, 1998.

Auchincloss *et al.* "Characterization of Tbc2, a nucleus-encoded factor specifically required for the translation of the chloroplast *psbC* mRNA in *Chlamydomonas reinhardtii*." *The Journal of Cell Biology* 157, no. 6 (June 2002): 953-962.

Ausubel *et al.* Eds. *Current Protocols in Molecular Biology*. 3 Vols. U.S.A.: John Wiley and Sons, 2001.

Barnes *et al.* "Identification and Characterization of a Novel RNA Binding Protein That Associates with the 5'-Untranslated Region of the Chloroplast *psbA* mRNA." *Biochemistry* 43, no. 26 (July 2004): 8541-8550.

Bateman, J. M. and S. Purton. "Tools for chloroplast transformation in *Chlamydomonas*: expression vectors and a new dominant selectable marker." *Molecular and General Genetics* 263, no. 3 (April 2000): 404-410.

Bennoun *et al.* "A METHOD FOR COMPLEMENTATION ANALYSIS OF NUCLEAR AND CHLOROPLAST MUTANTS OF PHOTOSYNTHESIS IN CHLAMYDOMONAS." *Genetics* 95, no.1 (May 1980): 39-47.

Birck *et al.* "Human TAF_{II} 28 and TAF_{II}18 Interact through a Histone Fold Encoded by Atypical Evolutionary Conserved Motifs Also Found in the SPT3 Family." *Cell* 94, no. 2 (July 1998): 239-249.

Blatch, G. L., and M. Lässle. "The tetratricopeptide repeat: a structural motif mediating protein-protein interactions." *Bioessays* 21, no. 11 (November 1999): 932-939.

Boudreau *et al.* "Conserved Gene Clusters in the Highly Rearranged Chloroplast Genomes of *Chlamydomonas moewusii* and *Chlamydomonas reinhardtii*." *Plant Molecular Biology* 24, no. 4 (February 1994): 585-602.

Boudreau *et al.* "The *Nac2* gene of *Chlamydomonas* encodes a chloroplast TPR-like protein involved in *psbD* mRNA stability." *The EMBO Journal* 19, no. 13 (July 2000): 3366-3376.

Bowers *et al.* "Molecular Markers for Rapidly Identifying Candidate Genes in *Chlamydomonas reinhardtii*: *ERY1* and *ERY2* Encode Chloroplast Ribosomal Proteins." *Genetics* 164, no. 4 (August 2003): 1345-1353.

ChemiCool. D. Hsu. © 2005. MIT. <http://www.chemicool.com/definition/ribulose_1_5_bisphosphate_carboxylase_oxygenase_rubisco.html> (RuBisCO's definition; accessed 2005).

Chlamydomonas Genetics Centre Personnel. Discovery of the *tbc2-F64* mutation and method of induction. Pers. comm.

ChloroP 1.1 Prediction Server. Eds. Emanuelsson, O and K. Rapacki. © 2005. CBS, DTU. <www.cbs.dtu.dk/services/ChloroP> (for determining chloroplast transit peptides; accessed 2004).

Chua, N. H., and P. Bennoun. "Thylakoid Membrane Polypeptides of *Chlamydomonas reinhardtii*: Wild-Type and Mutant Strains Deficient in Photosystem II Reaction Center." *Proceedings of the National Academy of Sciences of the United States of America* 72, no. 6 (June 1975): 2175-2179.

Chua *et al.* "PERIODIC VARIATIONS IN THE RATIO OF FREE TO THYLAKOID-BOUND CHLOROPLAST RIBOSOMES DURING THE CELL CYCLE OF *CHLAMYDOMONAS REINHARDTII*." *The Journal of Cell Biology* 71, no. 2 (November 1976): 497-514.

Cline, K., and R. Henry. "IMPORT AND ROUTING OF NUCLEUS-ENCODED CHLOROPLAST PROTEINS." *Annual Review of Cell and Developmental Biology* 12, no. 1 (1996): 1-26.

ClineWeb. K. Cline. © 2005. Univ. of Florida. <http://www.hos.ufl.edu/clineweb/Protocols/Pealsol.htm> (methods for pure chloroplast isolation; accessed 2005).

ClustalW. Rodrigo Lopez. © 2005. EMBL-EBI. <http://www.ebi.ac.uk/clustalw> (alignment of protein sequences and PPPEW repeats; accessed 2004).

Coffin *et al.* "The *Neurospora crassa cya-5* nuclear gene encodes a protein with a region of homology to the *Saccharomyces cerevisiae* PET309 protein and is required in a post-transcriptional step for the expression of the mitochondrially encoded COXI protein." *Current Genetics* 32, no. 4 (October 1997): 273-280.

Coligan *et al.* Eds. *Current Protocols in Protein Science*. 3 Vols. U.S.A.: John Wiley and Sons, 2001.

Creighton, T. *Proteins: Structures and Molecular Properties*. 2nd ed. New York: W. H. Freeman and co., 1993.

Das *et al.* "The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions." *The EMBO Journal* 17, no. 5 (March 1998): 1192-1199.

Dauvillée *et al.* "Tab2 is a novel conserved RNA binding protein required for translation of the chloroplast *psaB* mRNA." *The EMBO Journal* 22, no. 23 (December 2003): 6378-6388.

Day, R. A. *How to Write and Publish a Research Paper.* 5th ed. U.S.A.: Orynx Press, 1998.

Debuchy, R., Purton, S., and J. D. Rochaix. "The argininosuccinate lyase gene of *Chlamydomonas reinhardtii*: an important tool for nuclear transformation and for correlating the genetic and molecular maps of the *ARG7* locus." *The EMBO Journal* 8, no. 10 (October 1989): 2803-2809.

Delepelaire, P. "Partial characterization of the biosynthesis and integration of the Photosystem II reaction Centers in the thylakoid membrane of *Chlamydomonas reinhardtii.*" *The EMBO Journal* 3, no. 4 (April 1984): 701-706.

De Vitry *et al.* "Photosystem II Particles from *Chlamydomonas reinhardtii.* PURIFICATION, MOLECULAR WEIGHT, SMALL SUBUNIT COMPOSITION, AND PROTEIN PHOSPHORYLATION." *The Journal of Biological Chemistry* 266, no. 25 (September 1991): 16614-16621.

De Vitry *et al.* "The 4 kDa Nuclear-encoded PetM Polypeptide of the Chloroplast Cytochrome b_6f Complex. NUCLEIC ACID AND PROTEIN SEQUENCES, TARGETING SIGNALS, TRANSMEMBRANE TOPOLOGY." *The Journal of Biological Chemistry* 271, no. 18 (May 1996): 10667-10671.

De Vitry *et al.* "Analysis of the Nucleus-Encoded and Chloroplast-Targeted Rieske Protein by Classic and Site-Directed Mutagenesis of Chlamydomonas." *The Plant Cell* 11, no. 10 (October 1999): 2031-2044.

Di Cello *et al.* "Approaches to Bacterial RNA Isolation and Purification for Microarray Analysis of *Escherichia coli* K1 Interaction with Human Brain Microvascular Endothelial Cells." *Journal of Clinical Microbiology* 43, no. 8 (August 2005): 4197-4199.

DOE Joint Genome Institute (JGI) *C. reinhardtii* BLAST Program V.2.0. U.S. Department of Energy, U. of California. © 2005. <http://genome.jgi-psf.org/cgi-bin/runAlignment?db=chlre2&advanced=1> (to determine paralogs of TBC2p; accessed 2004).

Drews *et al.* "Negative Regulation of the Arabidopsis Homeotic Gene *AGAMOUS* by the *APETALA2* Product." *Cell* 65, no. 6 (June 1991): 991-1002.

DuBell, A. N. and J. E. Mullet. "Continuous Far-Red Light Activates Plastid DNA Synthesis in Pea Leaves but Not Full Cell Enlargement or an Increase in Plastid Number per Cell." *Plant Physiology* 109, no. 1 (September 1995a): 95-103.

DuBell, A. N. and J. E. Mullet. "Differential Transcription of Pea Chloroplast Genes during Light-Induced Leaf Development (Continuous Far-Red Light Activates Chloroplast Transcription)." *Plant Physiology* 109, no. 1 (September 1995b): 105-112.

Emanuelsson *et al.* "ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites." *Protein Science* 8, no. 5 (May 1999): 978-984.

ExPASy Translate Tool. Swiss Institute of Bioinformatics. © 2005. Canadian Bioinformatics Resource. <http://ca.expasy.org/tools/dna.html> (to translate putative nucleotide ORFs of *TBC2A* and *TBC2B*; accessed 2004).

Fischer, N. and J. D. Rochaix. "The flanking regions of *PsaD* drive efficient gene expression in the nucleus of the green alga *Chlamydomonas reinhardtii*." *Molecular Genetics and Genomics* 265, no. 5 (July 2001): 888-894.

Fisk *et al.* "Molecular cloning of the maize gene *crp1* reveals similarity between regulators of mitochondrial and chloroplast gene expression." *The EMBO Journal* 18, no. 9 (May 1999): 2621-2630.

Fogel, M. A. and M. K. Waldor. "Distinct segregation dynamics of the two *Vibrio cholerae* chromosomes." *Molecular Microbiology* 55, no. 1 (January 2005): 125-136.

Franklin *et al.* "Development of a GFP reporter gene for *Chlamydomonas reinhardtii* chloroplast." *The Plant Journal* 30, no. 6 (June 2002): 733-744.

Franzén *et al.* "Chloroplast transit peptides from the green alga *Chlamydomonas reinhardtii* share features with both mitochondrial and higher plant chloroplast presequences." *FEBS Letters* 260, no. 2 (January 1990): 165-168.

Funes *et al.* "The Typically Mitochondrial DNA-encoded ATP6 Subunit of the F_1F_0-ATPase Is Encoded by a Nuclear Gene in *Chlamydomonas reinhardtii.*" *The Journal of Biological Chemistry* 277, no. 8 (February 2002): 6051-6058.

Gardner *et al.* "Draft versus finished sequence data for DNA and protein diagnostic signature development." *Nucleic Acids Research* 38, no. 18 (October 2005): 5838-5850.

Gavel, Y. and G. von Heijne. "A conserved cleavage-site motif in chloroplast transit peptides." *FEBS Letters* 261, no. 2 (February 1990): 455-8.

Gehring, K. NMR Laboratory, McGill University. Determination of amino acid strings in theTBC2p sequences. Pers. comm.

Glöckner, G. CTAB Extraction Protocol. © 2005. Duke University. <http://jupiter.biology.duke.edu/methods/dna.html> (CTAB extraction protocol; accessed 2005).

Gorman, D. S. and R. P. Levine. "*CYTOCHROME F AND PLASTOCYANIN: THEIR SEQUENCE IN THE PHOTOSYNTHETIC ELECTRON TRANSPORT CHAIN OF CHLAMYDOMONAS REINHARDTI.*" *Proceedings of the National Academy of Sciences of the United States of America* 54, no. 6 (December 1965): 1665-1669.

Grant *et al.* "Inheritance of chloroplast DNA in *Chlamydomonas reinhardtii.*" *Proceedings of the National Academy of Sciences of the United States of America* 77, no. 10 (October 1980): 6067-6071.

greenGenie ORF Finder. D. Kulp. Affymetrix Inc. © 2005. University of California. <http://www.cse.ucsc.edu/~dkulp/cgi-bin/greenGenie> (to determine possible ORFs containing paralogous sequences to *TBC2*; accessed 2004).

Guo *et al.* "PROSPECT-PSPP: an automatic computational pipeline for protein structure prediction." *Nucleic Acids Research* 34 (2004): W522-W525. <http://nar.oxfordjournals.org/> (accessed 2005).

Harlow, E. and D. Lane. *Antibodies: A Laboratory Manual.* U.S.A.: Cold Spring Harbor Laboratory Press, 1988.

Harris, E. "Re: *Chlamy* questions." Email to Joanne Wyglinski. November 20, 2005.

Harris, E. "Re: one last question (I hope)." Email to Joanne Wyglinski. July 26, 2005.

Harris, E. *The Chlamydomonas Sourcebook: A Comprehensive Guide to Biology and Laboratory Use.* U.S.A.: Academic Press, 1989.

Herrmann, R. G. and K. V. Kowallik. "Multiple Amounts of DNA Related to the Size of Chloroplasts. II. Comparison of Electron-microscopic and Autoradiographic Data." *Protoplasma* 69, no. 3 (1970): 365-372.

Hirano *et al.* "Snap Helix with Knob and Hole: Essential Repeats in S. pombe Nuclear Protein *nuc2*$^{+}$." *Cell* 60, no. 2 (January 1990): 319-328.

Hoober *et al.* "Assembly of Light-Harvesting Systems." Chap. 19 in *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas.* Edited by J. D. Rochaix, M. Goldschmidt-Clermont and S Merchant. Vol. 7. The Netherlands: Kluwer Academic Publishers, 1998.

Howard *et al.* "Identification and analysis of polyserine linker domains in prokaryotic proteins with emphasis on the marine bacterium *Microbulbifer degradans.*" *Protein Science* 13, no. 5 (May 2004): 1422-1425.

Hwang *et al.* "Predicted Structures of Apolipoprotein II mRNA Constrained by Nuclease and Dimethyl Sulfate Reactivity: Stable Secondary Structures Occur Predominantly in Local Domains via Intraexonic Base Pairing." *The Journal of Biological Chemistry* 264, no. 14 (May 1989): 8410-8418.

InterProScan. R. Apweiler and colleagues. © 2005. EMBL-EBI. <http://www.ebi.ac.uk/InterProScan/> (to determine if region of homology to *TBC2* on scaffold 1767 is a domain; accessed 2005).

Joliot *et al.* "In vivo Measurements of Photosynthetic Activity: Methods." Chap. 22 in *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas.* Edited by J. D. Rochaix, M. Goldschmidt-Clermont and S. Merchant. Vol. 7. The Netherlands: Kluwer Academic Publishers, 1998.

Joyard *et al.* "The Biochemical Machinery of Plastid Envelope Membranes." *Plant Physiology* 118, no. 3 (November 1998): 715-723.

Kapoor, S., and M. Sugiura. "Expression and Regulation of Plastid Genes." Chap. 6 in *Photosynthesis: A Comprehensive Treatise.* Edited by A. S. Raghavendra. U.K.: Cambridge University Press, 1998.

Kathir *et al.* "Molecular Map of the *Chlamydomonas reinhardtii* Nuclear Genome." *Eukaryotic Cell* 2, no. 2 (April 2003): 362-379.

Kawamukai *et al.* "Genetic and Biochemical Analysis of Adenylyl Cyclase-Associated Protein, cap, in *Schizosaccharomyces pombe.*" *Molecular Biology of the Cell* 3, no. 2 (February 1992): 167-180.

Kindle *et al.* "Engineering the chloroplast genome: Techniques and capabilities for chloroplast transformation in *Chlamydomonas reinhardtii.*" *Proceedings of the National Academy of Sciences of the United States of America* 88, no. 5 (March 1991): 1721-1725.

Klein *et al.* "Light-regulated Translation of Chloroplast Proteins. I. Transcripts of PsaA-PsaB, PsbA, and RbcL Are Associated with Polysomes in Dark-grown and Illuminated Barley Seedlings." *The Journal of Cell Biology* 106, no. 2 (February 1988): 289-301.

Krimm *et al.* "A coil-helix instead of a helix-coil motif can be induced in a chloroplast transit peptide from *Chlamydomonas reinhardtii.*" *European Journal of Biochemistry (FEBS)* 265, no. 1 (October 1999): 171-180.

Kroll *et al.* "*VIPP1*, a nuclear gene of *Arabidopsis thaliana* essential for thylakoid membrane formation." *Proceedings of the National Academy of Sciences of the United States of America* 98, no. 7 (March 2001): 4238-4242.

Kuroiwa, T. "The Replication, Differentiation, and Inheritance of Plastids with Emphasis on the Concept of Organelle Nuclei." *International Review of Cytology* 128 (1991): 1-62.

Larson *et al.* "Specificity for Activase is Changed by a Pro-89 to Arg Substitution in the Large Subunit of Ribulose-1,5,-bisphosphate Carboxylase/Oxygenase." *The Journal of Biological Chemistry* 272, no. 27 (July 1997): 17033-17037.

LaskerFoundation Genetic Dictionary. Lasker Medical Research Network. © 2002. <http://www.laskerfoundation.org/news/weis/g_dictionary.html> (scaffold definition; accessed 2005).

Leong, M. M. L. and G. R. Fox. "Luminescent Detection of Immunodot and Western Blots." *Methods in Enzymology.* 184 (1990): 442-451.

Lurin *et al.* "Genome-Wide Analysis of Arabidposis Pentatricopeptide Repeat Proteins Reveals Their Essential Role in Organelle Biogenesis." *The Plant Cell* 16, no. 8 (August 2004): 2089-2103.

Manthey, G. M. and J. E. McEwen. "The product of the nuclear gene *PET309* is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial *COX1* locus of *Saccharomyces cerevisiae*." *The EMBO Journal* 14, no. 16 (August 1995): 4031-4043.

Margulies, M. M. and A. Michaels. "RIBOSOMES BOUND TO CHLOROPLAST MEMBRANES IN *CHLAMYDOMONAS REINHARDTII*." *The Journal of Cell Biology* 60, no. 1 (January 1974): 65-77.

Margulies, M. M. and A. Michaels. "FREE AND MEMBRANE-BOUND CHLOROPLAST POLYRIBOSOMES *CHLAMYDOMONAS REINHARDTII*." *Biochimica et Biophysica Acta* 402, no. 3 (September 1975): 297-308.

Margulies, M. M. and J. S. Weistrop. "SUB-THYLAKOID FRACTIONS CONTAINING RIBOSOMES." *Biochimica et Biophysica Acta* 606, no. 1 (January 1980): 20-33.

Maul *et al.* "The *Chlamydomonas reinhardtii* Plastid Chromosome: Islands of Genes in a Sea of Repeats." *The Plant Cell* 14, no. 11 (November 2002): 2659-2679.

Meissner *et al.* "Bacteriophage λ cloning system for the construction of directional cDNA libraries." *Proceedings of the National Academy of Sciences of the United States of America* 84, no. 12 (June 1987): 4171-4175.

Mets, L., and J. D. Rochaix. "Perspectives." Chap. 36 in *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas.* Edited by J. D. Rochaix, M. Goldschmidt-Clermont and S. Merchant. Vol. 7. The Netherlands: Kluwer Academic Publishers, 1998.

Mili, S. and S. Pinol-Roma. "LRP130, a Pentatricopeptide Motif Protein with a Noncanonical RNA-Binding Domain, Is Bound In Vivo to Mitochondrial and Nuclear RNAs." *Molecular and Cellular Biology* 23, no. 14 (July 2003): 4972-4982.

Millen *et al.* "Many Parallel Losses of *infA* from Chloroplast DNA during Angiosperm Evolution with Multiple Independent Transfers to the Nucleus." *The Plant Cell* 13, no. 3 (March 2001): 645-658.

Miyamura *et al.* "Quantitative Fluorescence Microscopy on Dynamic Changes of Plastid Nucleoids During Wheat Development." *Protoplasma* 133, no. 1 (1986): 66-72.

Mühlbauer, S. K. and L. A. Eichacker. "The stromal protein large subunit of ribulose-1,5-bisphosphate carboxylase is translated by membrane-bound ribosomes." *The FEBS Journal* 261, no. 3 (May 1999): 784-788.

Nakamura *et al.* "Chloroplast RNA-binding and pentatricopeptide repeat proteins." *Biochemical Society Transactions* 32, pt. 4 (August 2004): 571-574.

NCBI Conserved Domains. NCBI. © 2005. NIH. <http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml> (determine whether conserved domains exist in TBC2p, TBC2Ap and TBC2Bp; accessed 2005).

NCBI Conserved Domain Search v.2.04, RPS-BLAST (Reverse-Position Specific BLAST). NCBI. © 2005. NIH. < http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi> (to determine whether conserved domains exist in TBC2p, TBC2Ap and TBC2Bp; accessed 2005).

NCBI Pairwise BLAST. NCBI. © 2005. NIH. <http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi> (to determine exon and intron sequences of the *TBC2* sequence by BLASTing *TBC2* cDNA with genomic; accessed 2005).

NCBI *protein-protein* BLAST for short, nearly exact matches. NCBI. © 2005. NIH. <http://www.ncbi.nlm.nih.gov/BLAST/> (to determine commonality of amino acid strings in proteins; accessed 2005).

Nguyen *et al.* "Solvent effects on the conformational transition of a model polyalanine peptide." *Protein Science* 13, no. 11 (November 2004): 2909-2924.

Nickelsen, J. "Chloroplast RNA Stability." Chap. 9 in *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas.* Edited by J. D. Rochaix, M. Goldschmidt-Clermont and S. Merchant. Vol. 7. The Netherlands: Kluwer Academic Publishers, 1998.

Nield *et al.* "Three-dimensional Structure of *Chlamydomonas reinhardtii* and *Synechococcus elongatus* Photosystem II Complexes Allows for Comparison of Their Oxygen-evolving Complex Organization." *The Journal of Biological Chemistry* 275, no. 36 (September 2000): 27940-27946.

pDRAW32 DNA Analysis Software, Acaclone Software. VB 5.0 beta, revision 1.0.32. K. Oleson. © 2005. < http://www.acaclone.com/> (borrowed map of the Tab2:HA vector from A. Auchincloss; accessed 2004).

Peeters, L. "Re: [jgi-pst.org#1431] (chlre2) copyright clearance." Email to Joanne Wyglinski. July 13, 2005.

Perron *et al.* "A factor related to pseudouridine synthases is required for chloroplast group II intron *trans*-splicing in *Chlamydomonas reinhardtii*." *EMBO Journal* 18, no. 22 (November 1999): 6481-6490.

pFAM Home Page. Sanger Institute. © 2005. Wellcome Trust. <http://www.sanger.ac.uk/software/Pfam> (overview of the pFAM database; accessed 2005).

Porra *et al.* "Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls *a* and *b* extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy." *Biochimica et Biophysica Acta* 975, no. 3 (August 1989): 384-394.

PROSPECT-PSPP. J. Guo © 2005. Computational Systems Biology Labs. <http://csbl.bmb.uga.edu/protein_pipeline/login.php> (to determine putative characteristics of TBC2p and hypothetical proteins TBC2Ap and TBC2Bp; accessed 2005).

PSORT User's Manual. Kenta Nakai. © 1990. Human Genome Centre, IMS, Tokyo. <http://psort.nibb.ac.jp/helpwww.html#chl> (characterizing the presence of amino acids strings in the TBC2p sequences; accessed 2005).

Rattanachaikunsopon *et al.* "Cloning and characterization of the nuclear AC115 gene of *Chlamydomonas reinhardtii*." *Plant Molecular Biology* 39, no. 1 (January 1999): 1-10.

Redding, K. and G. Peltier. "Reexamining the Validity of the Z-Scheme: Is Photosystem I Required for Oxygenic Photosynthesis in *Chlamydomonas*?" Chap. 18 in *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas.* Edited by J. D. Rochaix, M. Goldschmidt-Clermont and S. Merchant. Vol. 7. The Netherlands: Kluwer Academic Publishers, 1998.

Rochaix *et al.* "Nuclear and chloroplast mutations affect the synthesis or stability of the chloroplast *psbC* gene product in *Chlamydomonas reinhardtii.*" *The EMBO Journal* 8, no. 4 (April 1989): 1013-1021.

Rochaix, J. D. "Post-transcriptional regulation of chloroplast gene expression in *Chlamydomonas reinhardtii.*" Rev. *Plant Molecular Biology* 32, no. 1-2 (October 1996): 327-341.

Rohr *et al.* "Tandem inverted repeat system for selection of effective transgenic RNAi strains in *Chlamydomonas.*" *The Plant Journal* 40, no. 4 (November 2004): 611-621.

Roswell, D. F. and E. H. White. "The Chemiluminescence of Luminol and Related Hydrazides." *Methods in Enzymology.* 57 (1978): 409-423.

Ruffle, S. V. and R. Sayre. "Functional Analysis of Photosystem II." Chap. 16 in *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas.* Edited by J. D. Rochaix, M. Goldschmidt-Clermont and S. Merchant. Vol. 7. The Netherlands: Kluwer Academic Publishers, 1998.

Sachetto-Martins *et al.* "Plant glycine-rich proteins: a family or just proteins with a common motif?" Rev. *Biochimica et Biophysica Acta-Gene Structure and Expression* 1492, no. 1 (June 2000): 1-14.

Sager, R. and S. Granick. "Nutritional Studies with *Chlamydomonas reinhardti.*" *Annals of the New York Academy of Sciences* 56, no 5 (October 1953): 831-838.

Sambrook *et al. Molecular Cloning: A Laboratory Manual.* 2nd ed. 3 vols. U.S.A.: Cold Spring Harbor Laboratory Press, 1989.

Sambrook, J., and D. W. Russell. *Molecular Cloning: A Laboratory Manual.* 3rd ed. 3 vols. New York: Cold Bring Harbor Laboratory Press, 2001.

Sato *et al.* "Detection and characterization of a plastid envelope DNA-binding protein which may anchor plastid nucleoids." *The EMBO Journal* 12, no. 2 (February 1993): 555-561.

Sato *et al.* "DNA-binding proteins mediate interaction of nucleoids with envelope membrane in developing plastids." In *Photosynthesis: From Light to Biosphere.* Edited by P. Mathis. Vol. 3. The Netherlands: Kluwer Academic Publishers, 1998.

Sato *et al.* "Dynamics of localization and protein composition of plastid nucleoids in light-grown pea seedlings." *Protoplasma* 200, nos. 3-4 (September 1997): 163-173.

Sato *et al.* "Molecular Characterization of the PEND Protein, a Novel bZIP Protein Present in the Envelope Membrane That Is the Site of Nucleoid Replication in Developing Plastids." *The Plant Cell* 10, no. 5 (May 1998): 859-872.

Sato *et al.* "Do plastid envelope membranes play a role in the expression of the plastid genome?" Rev. *Biochimie* 81, no. 6 (June 1999): 619-629.

Schultz *et al.* "SMART, a simple modular architecture research tool: Identification of signaling domains." *Proceedings of the National Academy of Sciences of the United States of America* 95, no. 11 (May 1998): 5857-5864.

Sears, B. "Replication, Recombination, and Repair in the Chloroplast Genetic System of Chlamydomonas." Chap. 7 in *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas.* Edited by J. D. Rochaix, M. Goldschmidt-Clermont and S. Merchant. Vol. 7. The Netherlands: Kluwer Academic Publishers, 1998.

Seigneurin-Berny *et al.* "Sulfolipid Is a Potential Candidate for Annexin Binding to the Outer Surface of Chloroplast." *Biochemical and Biophysical Research Communications* 272, no. 2 (June 2000): 519-524.

Silflow, C. D. "Organization of the Nuclear Genome." Chap. 3 in *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas.* Edited by J. D. Rochaix, M. Goldschmidt-Clermont and S. Merchant. Vol. 7. The Netherlands: Kluwer Academic Publishers, 1998.

Simpson *et al.* "A stable chemiluminescent-labelled antibody for immunological assays." *Nature* 279, no. 5714 (June 1979): 646-47.

Small, I. D. and N. Peeters. "The PPR motif-a TPR-related motif prevalent in plant organellar proteins." *Trends in Biochemical Sciences* 25, no. 2 (February 2000): 46-47.

Sodmergen *et al.* "Preferential digestion of chloroplast nuclei (nucleoids) during senescence of the coleoptile of *Oryza sativa*." *Protoplasma* 152, nos. 2-3 (June 1989): 65-68.

Spreitzer, R. J. "Genetic Engineering of Rubisco" Chap. 27 in *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas.* Edited by J. D. Rochaix, M. Goldschmidt-Clermont and S. Merchant. Vol. 7. The Netherlands: Kluwer Academic Publishers, 1998.

Stoscheck, C. M. "Quantitation of Protein." *Methods in Enzymology.* 182 (1990): 50-68.

Sueoka, N. "*MITOTIC REPLICATION OF DEOXYRIBONUCLEIC ACID IN CHLAMYDOMONAS REINHARDTI.*" *Proceedings of the National Academy of Sciences of the United States of America* 46, no. 1 (January 1960): 83-91.

Sugiura *et al.* "EVOLUTION AND MECHANISM OF TRANSLATION IN CHLOROPLASTS." Rev. *Annual Review of Genetics* 32 (December 1998): 437-459.

TAIR. Carnegie Institution of Washington. © 2005. Stanford University. <http://www.arabidposis.org/info/genefamily/> (determining divergent gene families; accessed 2005).

TargetP 1.1 Server. Eds. Emanuelsson, O and K. Rapacki. © 2005. CBS, DTU. <http://www.cbs.dtu.dk/services/TargetP> (determine targeting domains of the TBC2p, TBC2Ap and TBC2Bp; accessed 2004).

Tatusov *et al.* "The COG database: an updated version includes eukaryotes." *BMC Bioinformatics* 4 (September 2003): 41-54. (<http://www.biomedcentral.com/1471-2105/4/41>; accessed 2005)

Taylor *et al.* "First Crystal Structure of Rubisco from a Green Alga, *Chlamydomonas reinhardtii.*" *The Journal of Biological Chemistry* 276, no. 51 (December 2001): 48159-48164.

Tian *et al.* "Fas-activated Serine/Threonine Kinase (FAST) Phosphorylates TIA-1 during Fas-mediated Apoptosis." *The Journal of Experimental Medicine* 182, no. 3 (September 1995): 865-874.

Timko, M. "Pigment Biosynthesis: Chlorophylls, Heme, and Carotenoids." Chap. 20 in *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas.* Edited by J. D. Rochaix, M. Goldschmidt-Clermont and S. Merchant. Vol. 7. The Netherlands: Kluwer Academic Publishers, 1998.

Timmis *et al.* "Endosymbiotic Gene Transfer: Organelle Genomes Forge Eukaryotic Genomes." Rev. *Nature Reviews. Genetics.* 5, no. 2 (February 2004): 123-135.

Treger, M and E. Westhof. "Statistical analysis of atomic contacts at RNA-protein interfaces." *Journal of Molecular Recognition* 14, no. 4 (July-August 2001): 199-214.

Turabian, K. L. *A Manual for Writers of Term Papers, Theses, and Dissertations.* 6th ed. U.S.: The University of Chicago Press, 1996.

University of Chicago Press. *The Chicago Manual of Style*, 15th ed. Chicago: University of Chicago Press, 2003.

Vaistij *et al.* "Characterization of Mbb1, a nucleus-encoded tetratricopeptide-like repeat protein required for expression of the chloroplast *psbB/psbT/psbH* gene cluster in *Chlamydomonas reinhardtii*." *Proceedings of the National Academy of Sciences of the United States of America* 97, no. 26 (December 2000): 14813-14818.

Veening *et al.* "Visualization of Differential Gene Expression by Improved Cyan Fluorescent Protein and Yellow Fluorescent Protein Production in *Bacillus subtilis*." *Applied and Environmental Microbiology* 70, no. 11 (November 2004): 6809-6815.

Wang *et al.* "Modular Recognition of RNA by a Human-Pumilio Homology Domain." *Cell* 110, no. 4 (August 2002): 501-512.

Westphal *et al.* "A vesicle transport system inside chloroplasts." *FEBS Letters* 506, no. 3 (October 2001): 257-261.

Westphal *et al.* "Evolution of Chloroplast Vesicle Transport." *Plant Cell Physiology* 44, no. 2 (February 2003): 217-222.

Williams, P. M. and A. Barkan. "A chloroplast-localized PPR protein required for plastid ribosome accumulation." *The Plant Journal* 36, no. 5 (December 2003): 675-686.

Yamazaki *et al.* "PPR motifs of the nucleus-encoded factor, PGR3, function in the selective and distinct steps of chloroplast gene expression in *Arabidopsis*." *The Plant Journal* 38, no. 1 (April 2004): 152-163.

Zak *et al.* "The initial steps of biogenesis of cyanobacterial photosystems occur in plasma membranes." *Proceedings of the National Academy of Sciences of the United States of America* 98, no. 23 (November 2001): 13443-13448.

Zerges, W. "Translation in chloroplasts." Rev. *Biochimie* 82, nos. 6-7 (June-July 2000): 583-601.

Zerges, W. "Regulation of Translation in Chloroplasts by Light and for Protein Complex Assembly." In *Molecular Biology and Biotechnology of Plant Organelles.* Edited by H. Daniell and C. D. Chase. The Netherlands: Kluwer Academic Publishers, 2004.

Zerges, W., and J. D. Rochaix. "The 5' Leader of a Chloroplast mRNA Mediates the Translational Requirements for Two Nucleus-Encoded Functions in *Chlamydomonas reinhardtii*." *Molecular and Cellular Biology* 14, no. 8 (1994): 5268-5277.

Zerges *et al.* "Translation of the Chloroplast *psbC* mRNA Is Controlled by Interactions between Its 5' Leader and the Nuclear Loci *TBC1* and *TBC3* in *Chlamydomonas reinhardtii*." *Molecular and Cellular Biology* 17, no. 6 (June 1997): 3440-3448.

Zerges, W. and J. D. Rochaix. "Low Density Membranes Are Associated with RNA-binding Proteins and Thylakoids in the Chloroplast of *Chlamydomonas* reinhardtii." *The Journal of Cell Biology* 140, no. 1 (January 1998): 101-110.

Zerges *et al.* "Multiple Translational Control Sequences in the 5' Leader of the Chloroplast *psbC* mRNA Interact With Nuclear Gene Products in *Chlamydomonas reinhardtii*." *Genetics* 163, no. 3 (March 2003): 895-904.

Culture and Sonication

-cells are grown at 24°C, 140 rpm in 10 ml TAP with 5.5% arginine if required until post-log phase.

-cells are centrifuged at room temperature for 2 minutes at 5 K and pellets are resuspended in 200 µl TE with protease inhibitor cocktail and 50 µl of 10 mg/ml PMSF in isopropanol.

-cells are sonicated on the Fisher Scientific 550 Sonic Dismembrator with a 1/8 inch tip.

-sonication time set at 5 seconds with 15 second intervals during a period of 1 minute.

-sonication success is verified using microscopy.

-suspensions are centrifuged on a tabletop centrifuge at 4°C for 15 minutes at 14 K.

-both pellet and supernatant fractions are run on different SDS-PAGE gels for the Western blot.

For the soluble proteins (supernatant):

-to run an appropriate volume of supernatant, the amount of chlorophyll in the supernatant is determined in methanol using the equation $22.12(OD_{652}) + 2.71(OD_{665})$ so that loaded volumes can be calculated (Porra *et al.* 1989).

-to the cell extract, sucrose loading dye is added to the supernatant and DTT to make a 1-fold final concentration of 0.1 M (50 µl run).

For the membrane proteins (pellet):

-a Pierce BCA Protein Assay is performed on the pellet samples to determine the amount of protein in the sample (Stoscheck 1990).

-to the soluble proteins, protein buffer is added to a 1-fold final concentration and DTT to a final concentration of 0.1 M (ranging between 8.27-26.07 µl).

-samples are boiled for 2 minutes at 100°C before loading.

Gel (12% SDS polyacrylamide) (modified from Sambrook and Russell 2001)

Components	Resolving Gel	Stacking Gel
30% acrylamide	4 ml	1 ml
1.5 M Tris-Cl, pH 8.8	2.5 ml	1.25 ml (0.5 M Tris-Cl, pH 6.8)
20% SDS	50 µl	25 µl
10% APS	23 µl	30 µl
TEMED	7 µl	7 µl
H_2O	3.42 ml	2.66 ml
Final volume	10 ml	5 ml

5-Fold Protein Loading Dye (modified from Sambrook, Fritsch and Maniatis 1989)

Component	Volume/mass
250 mM Tris pH 6.8	2.5 ml
SDS	0.5 g
Glycerol (Biorad)	2.5 ml
Bromophenol blue	0.025 g
Final volume	5 ml

Electrode Buffer Protein Gel (10-Fold) (Sambrook, Fritsch and Maniatis 1989)

Component	Volume/mass
Glycine	188 g
Tris (Sigma)	30.2 g
SDS	10 g
dH_2O	Up until 100 ml total

30% Acrylamide (modified from Sambrook and Russell 2001)

Component	Volume/mass
Acrylamide	28.9 g
N'-N'-bis-methylene-acrylamide	1.08 g
dH_2O	add until 100 ml

*filter sterilize.

1.5 M Tris-Cl (pH 8.8) (Sambrook and Russell 2001)

Component	Volume/mass
Tris (Sigma)	18.16 g
dH_2O	dissolve with 80 ml, then adjust to 100 ml after pH is set and autoclave
HCl	adjust to pH 8.8

*autoclave.

0.5 M Tris-Cl (pH 6.8) (Sambrook and Russell 2001)

Component	Volume/mass
Tris base	6 g
dH_2O	dissolve in 80 ml, then adjust to 100 ml after pH is set
HCl	adjust to pH 6.8

*autoclave.

20% SDS (Sambrook and Russell 2001)

Component	Volume/mass
Sodium Dodecyl Sulfate	20 g
dH_2O	dissolve in 90 ml, heat to 68°C with stirring until dissolved and add dH_2O until 100 ml.

10% Ammonium persulfate (Sambrook and Russell 2001)

Component	Volume/mass
Ammonium persulfate	1 g
dH_2O	add until 10 ml

*store at 4°C.

Prestained Molecular Weight Marker (MBI, Fermentas)

Protein	Source	MW, kDa
β-galactosidase	*E. coli*	118.0 kDa
Bovine Serum Albumin	Bovine plasma	86.0 kDa
Ovalbumin	Chicken egg white	47.0 kDa
Carbonic Anhydrase	Bovine erythrocytes	34.0 kDa
β-lactoglobulin	Bovine	26.0 kDa
Lysozyme	Chicken egg white	19.0 kDa

The Transfer Step:

-all filters should be wet before transfer is set up.

-the transfer is set up in a sandwich-like fashion as follows:

-two pieces of filter paper the exact size of the gel.

-roll out the gel onto the filter.

-roll on nitrocellulose paper.

-add two more pieces of filter.

-sponges are put on either side of the gel which faces the anode of the Biorad Power Pac 200.

-transfer performed at 0.05 A at 4°C.

Transfer Buffer, pH 8.3 (Sambrook, Fritsch and Maniatis 1989)

Component	Concentration	Volume/mass
Glycine	39 mM	2.9 g
Tris (Sigma)	48 mM	5.8 g
SDS (electrophoresis grade)	0.037%	0.37 g
Methanol	20%	200 ml
dH_2O	-	up until 1 L

-to verify if protein transfer is successful, staining with Ponceau S is done.

Ponceau S (Sambrook, Fritsch and Maniatis 1989)

Component	Volume/mass
Ponceau S	2 g
Trichloroacetic acid	30 g
Sulfosalicyclic acid	30 g
dH_2O	up until 100 ml total

-the nitrocellulose filter is then blocked with blocking solution before being incubated with primary antibody.

Blocking Solution (Sambrook, Fritsch and Maniatis 1989)

Component and Final Conc.	Volume/mass
Nonfat dried milk 5% (w/v)	5 g
Antifoam A (Sigma) 0.01%	0.01 g
Sodium azide in phosphate-buffered saline (PBS) (Bioshop) 0.02%	0.02 g
Tween 0.02%	0.02 g
dH_2O	until 100 ml

-after incubation at room temperature, blocking solution (1 ml/cm^2 of nitrocellulose filter) is added to a final volume of 4 ml with a concentration of murine monoclonal antibody (HA.11) (Covance) of 1:1000.

-the nitrocellulose filter is then washed 3 times in PBS buffer.

10-fold PBS Buffer (Sambrook, Fritsch and Maniatis 1989)

Component	Volume/mass
NaCl	8 g
KCl	0.2 g
Na_2HPO_4	1.44 g
KH_2PO_4	0.24 g
dH_2O	add 80 ml, then adjust to 100 ml after pH is increased to 7.4 with NaOH

*the solution is autoclaved for 20 minutes at 121°C.

-the nitrocellulose filter is then transferred to a tray of 200 ml TBS and incubated for 10 minutes at room temperature with agitation.

TBS (Sambrook, Fritsch and Maniatis 1989)

Component	Volume/mass
NaCl	4 g
KCl	0.1 g
Tris base	1.5 g
Phenol red	$7.5*10^{-3}$ g
HCl	until pH 7.4
dH_2O	up until 500 ml

-the nitrocellulose filter is then exposed to phosphate-free, azide-free blocking solution with a 1:2000 concentration of 2° antibody and incubated at room temperature for 1.5 H with agitation.

Phosphate-free, azide free blocking solution (Sambrook, Fritsch and Maniatis 1989)

Component	Final Concentration
Non-fat dried milk	5% w/v
NaCl	150 mM
Tris-Cl (pH 7.5)	20 mM
dH_2O	added until 100 ml

-2° horseradish peroxidase was reacted according to the ECL protocol and then exposed to Fuji x-ray film (Simpson *et al.* 1979; Leong and Fox 1990; Roswell and White 1978).

ECL Protocol

Solution A: P-coumaric acid (Sigma) 90 mM in DMSO

Solution B: Luminol 3-aminophalhydrazin (Fluka) 250 mM in DMSO

ECL Protocol: Solution 1

Component	Volume
Tris 100 mM pH 8.5	3 ml
Solution A	13.3 µl
Hydrogen peroxide	1.66 µl

ECL Protocol: Solution 2

Component	Volume
Tris 100 mM	3 ml
Solution B	30 µl

-the nitrocellulose filter is cut up according to lanes and soaked in a combination of solution "1" and "2" for 1 minute.

-the filters are then exposed for 45 minutes to x-ray film.

-300 ml TAP cultures containing 5.5% arginine and 1% sorbitol (exception: 1.5 L for *cw15wt*).

-cells are grown at 24°C, 140 rpm in appropriate light conditions of approximately 100 $\mu Em^{-2}s^{-1}$ unless PSII-deficient such as the *tbc2-F64;cw15;arg7* strain.

-cultures are grown until log phase.

-cultures are centrifuged at the following settings:

Condition	Setting
Rotor	Ja-14
Speed	5000 rpm, 3856 *G
Time	7 minutes
Temp.	21°C, Max. temp.: 23°C
Automatic shutdown	None
Acceleration	Maximum
Deceleration	Maximum

-the centrifuge which was used in this thesis was the Beckman-Coulter J30-I Avanti centrifuge.

-cultures are then resuspended in 3 ml of PMSF/heparin buffer which acts as a protease inhibitor.

PMSF/heparin solution (Zerges pers. comm.)

Component	Final Concentration
Tris	50 mM
KCl	50 mM
$MgCl_2$	10 mM
Heparin (Sigma)	1 mg/ml
PMSF	20 µg/ml

-cells are sonicated for 2 minutes on ice with a pulsar of five seconds and intervals of 15 seconds with a 1/8 inch sonicator tip.

-cells are verified by microscopy for successfulness of sonication

-sonicated cultures are then spun on a pre-refrigerated centrifuge at the following settings:

Condition	Setting
Rotor	Ja-30.50
Speed	15000 rpm or 27216 *G
Time	30 minutes
Temp.	4°C (pre-cooled), Max. temp.: 6°C
Automatic shutdown	None
Acceleration	Maximum
Deceleration	Maximum

-the centrifuge used in this experiment was the Beckman-Coulter Avanti J30-I centrifuge.

-the supernatant contains the translational protein complexes and RNA if bound.

-the supernatant is stored in 50 ml Falcon tubes at -80°C.

-strains are grown in 50 ml TAP containing 1% sorbitol and 5.5% arginine with shaking at 140 rpm under light conditions (100 $\mu Em^{-2}s^{-1}$) unless Photosystem II-deficient.

-cells are grown until a density of 5*10exp^5 cells/ml.

-cell cultures are centrifuged, resuspended in 50 ml TAP-RS and divided into two 25 ml cultures, one grown in light (100 $\mu Em^{-2}s^{-1}$), the other grown in darkness.

-cells are pelleted in SS34 tubes at 5 K on a benchtop centrifuge for 5 minutes.

-cells are washed once with TAP-RS and resuspended in 305 µl TAP-RS and 100 µl of sample is aliquoted into screw-cap vials.

-1.2 µl of a 3 mg/ml cycloheximide solution is added to each vial.

-samples are shaken at 300 rpm for 5 minutes, followed by the addition of 100 µCi of $[^{35}S]H_2SO_4$.

-samples are then incubated for either 5, 10 or 30 minutes, depending on the pulse-labeling time required.

-samples are then pelleted on a tabletop centrifuge (in this thesis, the Biofuge was used) and pellets were resuspended in TAP at a temperature of 4°C.

-samples are pelleted again and resuspended in TE and protease inhibitors (Sigma), were frozen at -80°C and thawed with 5-fold loading dye.

-a 2% final concentration of β-mercaptoethanol was added resulting in a total volume of 50 µl.

-samples were denatured for 1 hour at 37°C except for the protein ladder (see A.1 for components) which was denatured for 5 minutes at 100°C.

-samples are run on a regular 12% SDS-PAGE with the appropriate amount of loading dye (see A.1 for components for the 12% SDS-PAGE, the electrode buffer, and the loading dye used).

Coomassie Blue Staining Solution (Sambrook, Fritsch and Maniatis 3: 18.55)

Component	Mass/volume
Coomassie Blue	1.25 g
Methanol	225 ml
Acetic acid	50 ml
dH_2O	added until 500 ml

Destaining Solution (Sambrook, Fritsch and Maniatis 3: 18.55)

Component	Concentration
Methanol	30%
Acetic acid	10%

Shrinking Solution (Zerges pers. comm.)

Component	Concentration
Methanol	50%

-the gel is run at 25 mA.

-after the gel is completed, stain in the Coomassie staining solution for 45 minutes and destain twice for 15 minutes with destaining solution.

-dry the gel on an apparatus like a Speed Gel 562100 (Savant) and expose to Fuji x-ray film overnight.