Volume 304, number 2,3, 119–123 FEBS 11130 June 1992

Characterisation of a chloroplast-encoded *secY* homologue and *atpH* from a chromophytic alga

Evidence for a novel chloroplast genome organisation

Carol D. Scaramuzzi, Harold W. Stokes and Roger G. Hiller

School of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia

Received 17 March 1992; revised version received 21 April 1992

secY is a prokaryotic gene that encodes the SecY protein, an integral membrane component of the prokaryotic protein translocation apparatus. A chloroplast-encoded *secY* homologue has been identified in the unicellular, chromophytic alga, *Pavlova lutherii*. The gene predicts a protein composed of ten membrane-spanning regions, that is approximately 25% homologous and 50% similar to bacterial and plastid SecY proteins. The *secY* gene from *P. lutherii* is independent of the ribosomal protein (*rp*) gene cluster to which it is closely linked in other organisms. In *P. lutherii* *secY* is located 5′ to *atpI* and *atpH*. Since, in higher plants the *atpIHFA* gene cluster and the *rp* gene cluster are separated by approximately 50 kb, we conclude, this indicates a novel chloroplast gene arrangement in *P. lutherii*.

Chloroplast genome; Chromophyta; *Pavlova lutherii*; *secY*; *atpH*; Protein translocation

1. INTRODUCTION

The chloroplast genome of land plants consists of a circular double-stranded molecule of DNA, ranging from 120 to 150 kb [1]. The circular molecule contains an inverted repeat which separates a small copy and a large copy region. This arrangement of the chloroplast genome, thought to confer evolutionary stability, is almost universal among the higher plants [1]. The chloroplast genomes of algae show a greater diversity of size, ranging from 84 kb to 600 kb [2] whereas those of the chromophytic algae for which there is much less information available, fall within the range of 100 to 160 kb [2]. There are indications, based on nucleotide sequence data for the Rubisco genes, *rbcL* and *rbcS*, that the chromophytic alga contain several genes that are similar to prokaryotic genes but which are not present in higher plant and green algal chloroplast DNA [3]. Some genes involved in protein translocation have also been retained on the chloroplast genome of eukaryotic algae outside of the chlorophyta [4–6]. The *secY* gene is found in both *Escherichia coli* [7] and *Bacillus subtilis* [8] where it forms part of the *spc* operon of the ribosomal protein (*rp*) gene cluster and in *Cyanophora paradoxa* [9] where it is located immediately adjacent to the final gene of this operon. Recently, a chloroplast-encoded *secY* homologue from a cryptophytic algae has also been located within the *spc* operon [6]. The bacterial *rp* gene cluster is composed of three operons, the S10, *spc* and alpha operons collectively containing 28 genes. Only 11 of these genes are found in the analogous *rp* cluster of higher plants [10,11]. It is presumed that the remainder have been lost or transferred to the nucleus. Among these is *secY*, of the *spc* operon the product of which is involved in prokaryotic protein translocation [8]. We report here the nucleotide sequence of two chloroplast-encoded genes, *secY* and *atpH*, from the chromophytic alga, *Pavlova lutherii* and show that they are linked on the chloroplast genome but positioned independently of the *rp* cluster. We conclude that the chloroplast genome arrangement in this organism is novel.

Correspondence address: C. Scaramuzzi, School of Biological Sciences, Macquarie University, Sydney, NSW, 2109, Australia. Fax: (61) (2) 8058245.

The nucleotide sequences of the genes reported here have been filed in the EMBL database under the accession numbers X64731 (*atpH*); and X64732 (*secY*).

2. MATERIALS AND METHODS

P. lutherii was cultivated in Provosoli's enrichment media [12]. Harvesting, preparation of chloroplast DNA, construction of clone banks and hybridisation techniques have been previously described [4]. Manipulations of DNA were performed according to standard protocols [13] or, when using DNA-modifying enzymes, to the manufacturers instructions. DNA fragments for subcloning into M13 or for use in hybridisation analysis were gel-purified using GeneClean II (BIO 101, La Jolla, CA, USA). The host strains JM101 [14] and HB101 [13] were used for all manipulations.

A chloroplast DNA clone bank was generated using *Hin*dIII chloroplast DNA fragments and the plasmid vector, pUC19 [15]. In the course of investigating the gene organisation of the *P. lutherii* chloro-

plast genome we identified a *secY* clone, pMAQ804, containing a 807-bp DNA fragment internal to the *secY* gene (+735 to +1542) (Fig. 1) and an *atpI/H* clone, pMAQ806, containing a 264-bp DNA fragment incorporating the 3′ coding region of *atpI* and the 5′ coding region of *atpH*. Inserts from pMAQ804 and pMAQ806, were used to screen the chloroplast DNA clone bank previously described [4]. Both hybridised to pMAQ803 which contains a 8.0-kb *Bgl*II fragment of chloroplast DNA. DNA sequencing was performed by the chain termination method [16] using a Sequenase kit (United States Biochemical Corporation). Sequencing of both single and double-stranded templates was performed according to the manufacturers protocols. Primers used were either the universal primer supplied or synthetic oligonucleotides synthesized on a Pharmacia Gene Assembler MarkII. Oligonucleotides were designed to provide extensive overlap between gels and, where necessary, compressions were resolved by the substitution of dITP for dGTP. PCR reactions were prepared using a GeneAmp Kit (Perkin Elmer Cetus) in a 50 μl volume using 10 ng of chloroplast DNA and oligonucleotide primers (16-mers) at 200 nM. An initial cycle of 3 min at 94°C, 1 min at 37°C and 3 min at 65°C was followed by 29 cycles of 1 min at 94°C, 1 min at 37°C and 3 min at 65°C.

3. RESULTS

The *secY* gene is composed of 1257 nucleotides (Fig. 1) which predicts a protein of 419 amino acids with a molecular weight of 46,563. The predicted protein sequence of *P. lutherii secY* is shown in Fig. 2 and compared with the predicted *secY* protein of *Cryptomonas Φ* [6] and *C. paradoxa* [9]. The percent identity/similarity of organelle encoded SecY proteins are compared with each other and with SecY from *E. coli* and *B. subtilis* in Table I. The five SecY protein sequences are approximately 50% conserved, with regions of homology confined to ten hydrophobic regions and several charged regions. These domains correspond to the ten putative hydrophobic, membrane-spanning regions (MSR) and four cytosolically exposed charged loops consistent with the structure of other SecY proteins [6,8,9,17]. However, *P. lutherii* SecY shows considerable variation at both termini, in particular at the amino terminus, which contains an additional 7 and 39 amino acid residues in the *E. coli* and *C. paradoxa* SecY proteins respectively.

The chloroplast-encoded *atpH* gene of *P. lutherii* is composed of 249 nucleotides (Fig. 3) and predicts a protein of 83 amino acids, which is highly conserved. *P. lutherii* AtpH is 87% identical to AtpH from the cyanobacteria [18,19], 88% identical to its counterpart in *Odontella sinensus* [20] and spinach [21], 90% identical to the wheat AtpH [22] and 91.3% identical to AtpH of *Marchantia polymorpha* [23].

pMAQ803 DNA was mapped by restriction endonuclease analysis involving digestion with *Xho*II, *Xho*II/*Eco*RI, *Eco*RI, *Pst*I and *Pst*I/*Eco*RI and hybridisation with *secY* and *atpI/H* DNA inserts of pMAQ804 and pMAQ806. Both probes hybridised to a 6.0-kb *Xho*II fragment. From the pattern of fragments in the digests that hybridised to each of the two probes, a physical map of this *Xho*II fragment was constructed (Fig. 4).

The orientation and position of the 2.3-kb *Eco*RI fragment containing the *secY* gene relative to the *atpI/H* genes was deduced from PCR experiments. Outward facing primers were made to the 3′ and 5′ sequenced non-coding regions surrounding *secY* and to the 3′ and 5′ sequenced regions surrounding the *atpH* gene. The relative positions of the PCR primers (A, B, C, D) are indicated in Fig. 4 and the nucleotide sequence of each is overlined in Figs. 1 and 3. All four combinations of outward facing primers were cycled with chloroplast DNA. An amplified product of 1050 bases was generated only with primers B and C (Fig. 4). Combinations without template DNA failed to synthesise a product.

No open reading frames (ORFS) greater than 100 bases were identified in either of the 500 bases 5′ or 3′ to the *secY* coding sequence. Nucleic acid searches of each of these regions with sequences in the genbank database failed to reveal any significant matches. As the primers B and C were inset by 93 and 188 bp, respectively, from the end of the sequenced regions, we conclude that the unsequenced region between the *secY Eco*RI fragment and *atpI/H Xho*II/*Eco*RI fragment is 780 bases. As *atpI* is highly conserved and encompasses approximately 725 bp [20] we determine from the nucleotide sequence and PCR data that there is 600 bases of non-coding sequence between the end of the *secY* and the start of *atpI*. The results also show that the *secY* and *atpI/H* genes are transcribed in the same direction.

4. DISCUSSION

The presence of a *secY* gene linked to *atpH*, on the *P. lutherii* chloroplast chromosome indicates that there are obvious differences in the chloroplast genome organisation and structure of this alga when compared with chloroplast genomes from other plant phyla. In all cases, thus far reported, *secY* forms part of or is located

Table I

Pairwise comparison for predicted SecY proteins from *P. lutherii*, *C. paradoxa*, *Cryptomonas*, *E. coli* and *B. subtilis*

	P. lutherii	*C. paradoxa*	*Cryptomonas*	*E. coli*	*B. subtilis*
P. lutherii	–	28/50	32/55	31/49	26/53
C. paradoxa	28/50	–	29/48	25/46	25/46
Cryptomonas	32/55	29/48	–	34/56	33/57

Numbers refer to percent identity/similarity.

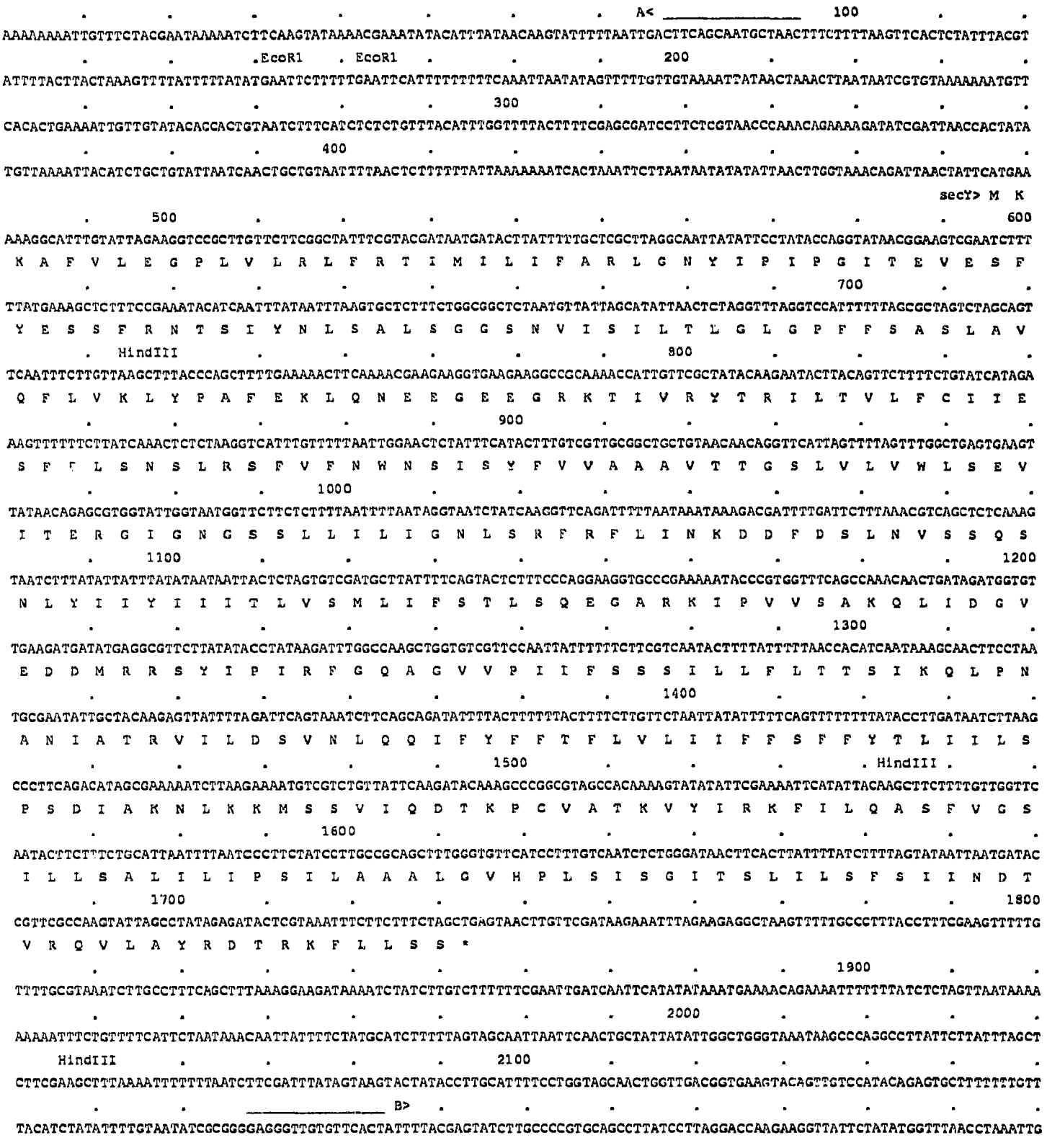

Fig. 1. Nucleotide sequence of the *P. lutherii secY* gene and surrounding sequence. *Hin*dIII restriction sites and the SecY predicted protein sequence are shown. The nucleotide sequences used to construct PCR primers (A and B) are overlined.

within the *rp* gene cluster from which the analogous gene cluster in higher plants appears to be derived [10]. However, the *secY* gene on the *P. lutherii* chloroplast genome is independent of the *rp* gene cluster and is

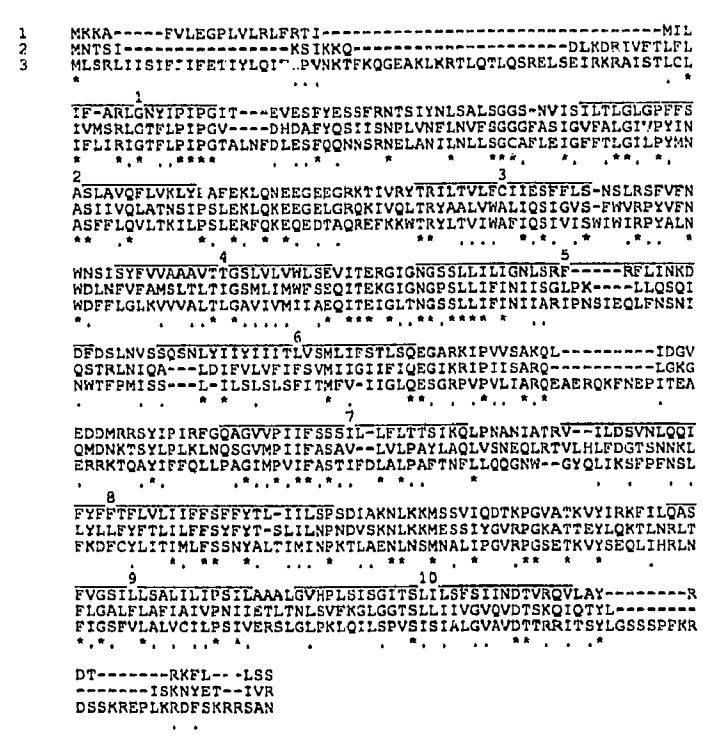

Fig. 2. Comparison of the derived amino acid sequences of *P. lutherii* SecY (1) with SecY proteins of *Cryptomonas* (2) and *C. paradoxa* (3). The predicted protein sequence is shown for all SecY proteins. The ten putative MSR are overlined. Identical amino acid residues are indicated with an asterisk (*); conservative substitutions are indicated with a dot (•). Alignments were carried out using the CLUSTAL program [34]. Substitutions are designated conservative if both amino acids fall within one of the following exchange groups: T, S, A, G, P: R, K, H: F, W, Y: D, E, Q, N: I, L, M, V: and C [35].

linked to *atpI/H*. Both these findings support the hypothesis that the chloroplast genome of *P. lutherii* differs from those of higher plants, where the *atpIHFA* and the *rp* gene clusters are separated by approximately 50 kb [11] and *secY* is not present. It also differs from the cryptophytic alga and *C. paradoxa*, where *secY* is present and linked to the *spc* operon, of the *rp* gene cluster [6,9].

The presence of the *secY* gene on the chloroplast genome of *P. lutherii* and its possible role within the chloroplast of this alga is of considerable importance. In *E. coli, secY* is one of six genes that code for protein translocation components. Protein translocation across the prokaryotic membranes involves an array of components (SecA, B, D, E, F and Y), termed 'export apparatus' which include soluble and membrane-associated proteins as well as signal peptides in the translocated protein. These features are common to the translocation process across the eukaryotic rough endoplasmic reticulum [24]. Suh et al. (1990) [8] suggest that signal peptide recognition proteins [25,26] and the soluble cytoplasmic SecA proteins [27,28] are likely candidates for interaction with SecY in order to facilitate bacterial protein translocation.

A chloroplast-encoded gene which predicts a protein related to bacterial SecA [27,28] is also present in *P. lutherii* (Scaramuzzi, PhD Thesis, 1991). The amino terminus of *P. lutherii* SecA is 52% identical to the equivalent region of the *E. coli* and *B. subtilis* SecA proteins. The amino terminus of SecA is essential for its function [27] by coupling ATP hydrolysis to precursor protein transport [29–31]. To date, *secA* has not been reported from any chloroplast genome but its presence, together with *secY* on the chloroplast genome of *P. lutherii* strongly implies that this organism has an alternative mechanism for protein translocation into and within the chloroplast, and which may be similar to bacterial protein translocation. Further support for this is the discovery of a chloroplast-encoded Hsp70 in *P. lutherii* [4].

The chloroplasts of chromophytic algae are surrounded by two extra membranes, the chloroplast endoplasmic reticulum (CER) which bear eukaryotic ribosomes [32]. These features have been implicated in the import of nuclear-encoded, chloroplast light-harvesting proteins in the diatom, *Phaeodactylum tricornutum* [33]. That the chromophytic algae have additional mechanisms accommodating chloroplast protein translocation has also been proposed for another diatom *O. sinensus* [20]. An ORF located upstream of the *atpI* gene predicts a protein that is approximately 20% identical to a prokaryotic periplasmic ATP-binding transport

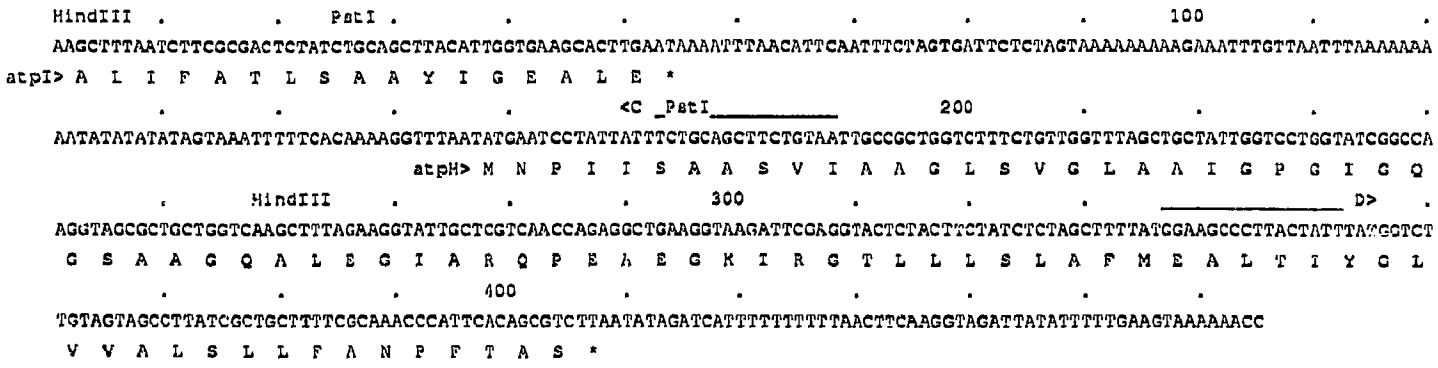

Fig. 3. Nucleotide sequence of the *P. lutherii, atpH* gene and the 3′ region of the preceding *atpI* gene. Nucleotide sequences used to construct PCR primers (C and D) are overlined. The predicted amino acid sequences of AtpH and the carboxy terminus of AtpI are shown.

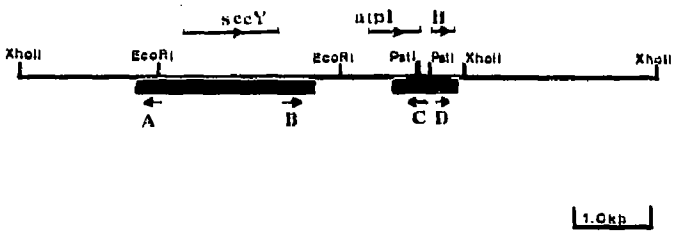

Fig. 4. Map of pMAQ803 which contains an 8.0-kb fragment of *P. lutherii* chloroplast DNA showing the positions of *secY*, *atpI* and *atpH*. The *secY* gene is located on a 2.3-kb *Eco*RI fragment and the *atp* genes are located on a 1.6-kb *Eco*RI/*Xho*II fragment. The two DNA fragments containing the genes are contiguous. The bold bars represent the regions of DNA sequenced. The length and position of the top arrows correspond to the location of the respective genes. The direction of presumed transcription is indicated. Bottom arrows indicate the position and direction of the PCR primers (A, B, C, D).

protein but does not display any similarity to the SecY protein reported here. The authors [20] conclude that the 'peculiar organisation of the chromophytic plastids including four additional membranes requires additional transport systems'. Movement of proteins within the chloroplast across thylakoid membranes may also be facilitated via these proteins.

The discovery of chloroplast-encoded *secY*, *secA* and *hsp70* genes and the linkage of *secY* to the *atpI* and *atpH* genes reveal that the chloroplast genome of *P. lutherii* differs substantially from that of other plant phyla. It may also be noted that the percent identity/similarity of *P. lutherii* SecY with *Cryptomonas Φ* SecY is similar to that with *C. paradoxa* cyanelle SecY and bacterial SecY and that there is only a low percent identity/similarity in the C-terminal region of chloroplast-encoded Hsp70 proteins (27% identity/48% similarity in 128 C-terminal residues) from *P. lutherii* and the red algae *Porphyra umbilicalis* [5]. We tentatively suggest that the chromophytic algae form a lineage distinct from the Chlorophyta, Cryptophyta, Rhodophyta and the cyanelle containing group of algae.

Acknowledgements: This research was supported, in part, by Macquarie University Research Grants to H.W. Stokes and R.G. Hiller.

REFERENCES

[1] Palmer, J.D. (1985) Annu. Rev. Genet. 19, 325–354.
[2] Cattolico, R.A. (1986) TREE 1, 64–67.
[3] Reith, M. and Cattolico, R.A. (1986) Proc. Natl. Acad. Sci. USA 83, 8599–8603.
[4] Scaramuzzi, C.D., Stokes, H.W. and Hiller, R.G. (1992) Plant Mol. Biol. 18, 467–476.
[5] Reith, M. and Mulholland, J. (1991) FEBS Lett. 294, 116–120.
[6] Douglas, S. (1992) FEBS Lett. 298, 93–96.
[7] Cerretti, D.P., Dean, D., Davis, G.R., Bedwell, D.M. and Nomura, M. (1983) Nucleic Acids Res. 11, 2599–2616.
[8] Suh, J.W., Boylan, S.A., Thomas, S.M., Dolan, K.M., Oliver, D.B. and Price, C.W. (1990) Mol. Microbiol. 4, 305–314.
[9] Michalowski, C.B., Pfanzagl, B., Loffelhardt, W. and Bohnert, H.J. (1990) Mol. Gen. Genet. 224, 222–231.
[10] Fukuzawa, H., Kohchi, T., Sano, T., Shirai, H., Umesono, K., Inokuchi, H., Ozeki, H. and Ohyama, K. (1988) J. Mol. Biol. 203, 333–351.
[11] Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, T., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) EMBO J. 5, 2043–2049.
[12] Provosoli, L. (1963) Growing marine seaweeds, in: Proc. Int. Seaweed Symp. (D. DeVirville and J. Feldman, eds.) vol. 4, Permagon, Oxford, pp. 9–17.
[13] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
[14] Yanisch-Perron, C., Viera, J. and Messing, J. (1985) Gene 33, 103–119.
[15] Vieira, J. and Messing, J. (1982) Gene 19, 259–268.
[16] Sanger, F., Coulson, A.R., Barell, B.G., Smith, A.J. and Roe, B.A. (1980) J. Mol. Biol. 143, 161–178.
[17] Akiyama, Y. and Ito, K. (1987) EMBO J. 6, 3465–3470.
[18] McCarn, D.F., Whitaker, J.A., Vrba, J.M. and Curtis, S.E. (1988) J. Bacteriol. 170, 3448–3458.
[19] Cozens, A.L. and Walker, J.E. (1987) J. Mol. Biol. 194, 359–383.
[20] Pancic, P.G., Strotmann, H. and Kowallik, K.V. (1992) J. Mol. Biol., submitted.
[21] Hennig, J. and Herrmann, R.G. (1986) Mol. Gen. Genet. 203, 117–128.
[22] Howe, C.J., Auffret, A.D., Doherty, A., Bowman, C.M., Dyer, T.A. and Gray, J.C. (1982) Proc. Natl. Acad. Sci. USA 79, 6903–6907.
[23] Umesono, K., Inokuchi, H., Shiki, Y., Takeuchi, M., Chang, Z., Fukuzawa, H., Kohchi, T., Shirai, H., Ohyama, K. and Ozeki, H. (1988) J. Mol. Biol.
[24] Saier Jr., M.H., Werner, P.K. and Muller, M. (1989) Microbiol. Rev. 53, 333–366.
[25] Stader, J., Bensom, S.A. and Silhavy, T.J. (1986) J. Biol. Chem. 261, 15075–15080.
[26] Emr, S.D., Hanley-Way, S. and Silhavy, T.J. (1981) Cell 23, 79–88.
[27] Sadaie, Y., Takamatsu, H., Nakamura, K. and Yamane, K. (1991) Gene 98, 101–105.
[28] Schmidt, M.G., Rollo, E.E., Grodberg, J. and Oliver, D.B. (1988) J. Bacteriol. 170, 3404–3414.
[29] Austen, B.M. and Westwood, O.M.R. (1991) Protein targeting and secretion, IRL Press, Oxford.
[30] Lill, R., Cunningham, K., Brundage, L.A., Ito, K., Oliver, D. and Wickner, W. (1989) EMBO J. 8, 961–966.
[31] Matsuyama, S., Kimura, E. and Mizushima, S. (1990) J. Biol. Chem. 265, 8760–8765.
[32] Gibbs, S.P. (1981) Int. Rev. Cytol. 72, 49–99.
[33] Grossman, A.R., Manodori, A. and Snyder, D. (1990) Mol. Gen. Genet. 224, 91–100.
[34] Higgins, D.G. and Sharpe, P.M. (1989) CABIOS 5, 151–153.
[35] Dayhoff, M.O., Schwartz, R.M. and Orcott, B.L. (1978) in: Atlas of Protein Sequence and Structure (M.O. Dayhoff, ed.) Vol. 5, suppl. 3, Washington DC, National Biomedical Research Foundation, pp. 345–352.