Nucleus-encoded Precursors to Thylakoid Lumen Proteins of Euglena gracilis Possess Tripartite Presequences

Rostislav Vacula,1 Jürgen M. Steiner,2 Juraj Krajcovic,1 Libor Ebringer,1 and Wolfgang Löffelhardt2,*

Institute of Molecular and Subcellular Biology, Comenius University, 81107 Bratislava, Slovak Republic,1 and Institut für Biochemie und Molekulare Zellbiologie der Universität Wien und Ludwig-Boltzmann-Forschungsstelle für Biochemie, Dr. Bohrgasse 9, A-1030 Vienna, Austria2

(Received 14 January 1999)

Abstract

The complete presequences of the nucleus-encoded precursors to two proteins, cytochrome $c_6$ and the 30-kDa protein of the oxygen-evolving complex, that reside in the thylakoid lumen of the chloroplasts of Euglena gracilis are presented. Sorting of these proteins involves translocation across four membranes, the three-membraned chloroplast envelope and the thylakoid membrane. The tripartite presequences show the structure: signal sequence — transit sequence — signal sequence. Three hydrophobic domains become apparent: two of them correspond to signal sequences for translocation across the endoplasmic reticulum (ER) membrane and the thylakoid membrane, respectively, whereas the third constitutes the stop-transfer signal contained in the long stroma-targeting part of the tripartite presequence.

Key words: Euglena gracilis; petJ; protein import; psbO; thylakoid lumen; tripartite presequences
Figure 1. A. Nucleotide and deduced amino acid sequence of the psbO presequence from *Euglena gracilis*. Total RNA was prepared from wild type *E. gracilis*, strain Z, by a single step method of acid guanidinium thiocyanate-phenol-chloroform extraction. mRNA was isolated using the Oligotex™ mRNA isolation kit (Quiagen). For RT-PCRs 10 μg of total RNA were incubated with Superscript II reverse transcriptase (GIBCO/BRL) as described by the supplier. PCR amplification was performed by using Pfu DNA polymerase (Stratagene) as described by the supplier for 30 cycles of 1 min at 94°C, 2 min at 60°C, and 3 min (last cycle 7 min) at 75°C. The part of the 5' untranslated region that is common to most *Euglena* mRNAs and that was used as a PCR primer is shown in a box. Nested primers (3', complementary) on the presequence used for amplification of complete cDNA 5' ends are underlined. PCR products were cloned into plasmid pGEM-T (Promega). Nucleotide sequences were determined in an automated sequencer (LONGREADIR 4200, LI-COR). Nucleotides added to the previously published sequence at the 5' end are given in italic letters, the inserted G residue is bold. The methionine-47 previously assumed to be the starting amino acid is given in italics and the mature part of the protein is indicated by bold letters. B. Nucleotide and deduced amino acid sequence of the petJ presequence from *Euglena gracilis*. A cDNA library of *E. gracilis* was constructed in the vector Lambda Zap II using a cDNA synthesis kit (Stratagene) according to the instructions of the manufacturer. Primers are boxed and underlined, respectively, as described for Fig. 1A. The start of the mature part of the protein is indicated by bold letters.
product of the psbO gene specifying the precursor for the 30-kDa protein of the oxygen-evolving complex (OEC30) of photosystem II. Alignment of this 93-aa presequence to those from Euglena precursors to stromal or thylakoid proteins like Rubisco SSU or LHCP II did not show an additional, third hydrophobic domain. Furthermore, the common 5' end added to all Euglena mRNAs by transsplicing was lacking in the psbO cDNA indicating that it is incomplete. In this case the methionine assumed to be the N-terminal amino acid would be an internal one rather and indeed the translation of the published nucleotide sequence could be continued further upstream for an additional 42 aa. Thus we decided to reinvestigate the presequence of the psbO gene.

We have also been interested for some time in cytochrome c₆, another protein resident in the thylakoid lumen. In this case, only the amino acid sequence of the mature Euglena protein is available and we included the respective presequence that should have a structure analogous to that of pre-OEC30, in this study.

Using degenerate primers directed against conserved regions of cytochrome c₆, which had been sequenced at the protein level (primer 1: 5'-GAYGNTTYG-CNGAYAAYTG-3', complementary to the sequence DVFADNC i.e., position 4–10 of the mature protein; primer 2: 5'-ACRTTNGCCANGCNCCNCC-3', corresponding to the sequence GGAWANVS i.e., position 80–86) we obtained a 250-bp RT-PCR product with total RNA from Euglena as a template. Sequencing verified that this PCR product was an internal fragment of the Euglena petJ gene which was then used as a probe to screen a Zap II library of Euglena cDNAs. The resulting petJ cDNA clone contained part of the targeting signals but appeared to be incomplete, too. Thus we made use of the known properties of Euglena mRNAs and performed a modified 5' RACE with a primer directed against the universal 5' end instead of an anchor primer. This strategy allowed the extension and completion of the coding regions for both petJ and psbO (Fig. 1). The completed and corrected psbO presequence (139 aa) features a 35-aa signal sequence I followed by a 73-aa transit sequence containing a hydrophobic domain and a 31-aa signal sequence II. The published nucleotide sequence was modified by insertion of a G (Fig. 1A) whereby an open reading frame containing the real start codon 135 bp upstream of the second methionine codon was created. The Euglena petJ presequence (129 aa) consists of a 26-aa N-terminal signal sequence I, a 76-aa transit sequence containing a hydrophobic domain, and a 27-aa C-terminal signal sequence II with a signal peptidase cleavage site at the junction presequence/mature protein. In the alignment of both presequences (Fig. 2) the tripartite character and the three stretches of hydrophobic amino acids can be nicely seen. The predicted cleavage sites after signal sequence I and the transit sequence part, respectively, were deduced by comparison with other Euglena nuclear genes for chloroplast proteins. In the hydropathy plots (Fig. 3) three significantly hydrophobic domains are clearly present in both cases. A fourth array of lesser hydrophobicity corresponds to the amino-terminal region of the transit sequence part.

The size difference of 10 aa in the two presequences are mainly due to the hydrophilic domains preceding the hydrophobic cores of both signal sequence I and II which are longer for pre-OEC 30 than for pre-cytochrome c₆. The region between the hydrophobic domains II and III is especially rich in charged amino acids on both sides of the (putative) stromal processing protease cleavage site. The transit sequence parts of the presequences of the luminal proteins are shorter than those of stromal or thylakoid proteins (93 to 106 aa): the stop-transfer domain is somewhat shifted towards the amino-terminus by approximately 10 aa and the adjacent carboxy-terminal part comprises three to five amino acids only, as compared to 18 to 31 aa in the latter cases. This tendency to keep the hydrophobic core of signal sequence II as close as possible to the membrane anchor might indicate the necessity to avoid potential bitopic integration into the vesicle membrane. Alternatively, it may simply reflect a constraint in total presequence size.

These results show that the principles of conservative sorting are also valid for complex plastids. The transit sequence part of the presequence was added to the prokaryotic signal peptide of a thylakoid lumen precursor protein after the primary endosymbiotic event to create a bipartite presequence and the second signal sequence (signal sequence I, when counted from the amino-terminal end of the precursor) was added after the secondary endosymbiotic event resulting in a tripartite presequence to be cleaved stepwise by three different processing proteases. This rather complicated sorting mechanism for nucleus-encoded chloroplast proteins is the price for the "second hand" phototrophy of the Euglena cell.

Acknowledgements: This work was supported by a grant from the Austrian "Fonds zur Förderung der
Figure 3. Kyte-Doolittle plots of the presequences of the OEC30 (A) and cytochrome c₆ (B) precursors from *E. gracilis*. The window size is 11.

References

5. Ishida, K., Cao, Y., Hasegawa, M., Okada, N., and Hara,


