DNA RESEARCH 6, 45–49 (1999) Short Communication

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

The chloroplasts of the phytoflagellate Euglena gracilis are surrounded by three membranes. The outermost membrane is thought to be a phagocytic vacuole,9,9 resulting from the uptake of an eukaryote, a green alga (the primary host cell with respect to the endosymbiotic origin of plastids from a cyanobacterial invader) by the secondary host cell, a heterotrophic protist.4,8 Alternatively, the outermost membrane could represent the plasma membrane of the primary host cell.9,13 A clear decision must await purification and analysis of the composition of the individual membranes. During evolution genes are believed to have been transferred from the primary host nucleus to the secondary host nucleus; the primary host cell was degraded with the exception of the chloroplast which finally became a member of the group of complex plastids due to the additional membrane surrounding it. Chloroplast proteins encoded by genes transferred to the secondary host nucleus required additional targeting information to cross the outermost of these three membranes. They acquired a signal sequence fused to the amino-terminal end of their previous transit sequence, which was also increased in length through a hydrophobic domain.6 Investigations in vitro and in vivo6,12,13 revealed an import pathway as follows: partial entry of the precursor into the ER (arrested by the stop-transfer domain which acts as a membrane anchor), cleavage of the signal sequence, transport of the intermediate to the Golgi apparatus, sorting of the respective vesicles to the chloroplast. The last steps are less well understood, but most probably involve vesicle fusion with the outermost chloroplast membrane and then transit sequence-mediated translocation across the two inner membranes.13 The transit sequence is then cleaved and so are eventual polyprotein structures which are typical for some Euglena chloroplast precursor proteins.1,3 Direct in vitro import into Euglena chloroplasts was shown for the precursor of porphobilinogen deaminase.9

With nuclear-encoded proteins of the thylakoid lumen in higher plants we find a different situation. According to the principle of “conservative sorting” a transit sequence was fused to the amino terminus of the signal sequence responsible for translocation across the thylakoid membrane.11 An intermediate form is produced through the action of a stromal processing protease. With luminal proteins of Euglena chloroplasts a tripartite presequence has to be postulated: first, a (eukaryotic) signal sequence; second, a (eukaryotic) transit sequence; third, a (prokaryotic) signal sequence.

To date, only one presequence for a thylakoid luminal protein of E. gracilis has been published, i.e., that for the...
Presequences for Precursors to Thylakoid Lumen Proteins

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. 2 mRNA was isolated using the Oligotex™ mRNA isolation kit (Quiagen). For RT-PCRs 10 fig 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 10 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.
ing regions for both petJ and psbO (Fig. 1). The combination of linker strategy allowed the extension and completion of the codon universal 5' end instead of an anchor primer. This strategy was a modified 5' RACE with a primer directed against the E. gracilis known properties of this gene which was then used as a probe to screen a AZap II library of FADNC i.e., position 4-10 of the mature protein; primer CNGAYAAYTG-3', complementary to the sequence DV-

...the respective presequence that should have a structure constraint in total presequence size. 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 c. 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 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 E. gracilis 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.

Using degenerate primers directed against conserved regions of cytochrome c6, another protein resident in the thylakoid lumen. In this case, only the amino acid sequence of the mature E. gracilis 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'-GAYGTNTYG-CNGAYAAYTG-3', complementary to the sequence DV-FADNC i.e., position 4-10 of the mature protein; primer 2: 5'-ACRTTNGCCCANGCNCCNCC-3', corresponding to the sequence GGAWANVS i.e., position 80-86) we obtained a 250-bp RT-PCR product with total RNA from E. gracilis as a template. Sequencing verified that this PCR product was an internal fragment of the E. gracilis petJ gene which was then used as a probe to screen a AZap II library of E. gracilis 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 E. gracilis 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 E. gracilis 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.

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 c\textsubscript{6} (B) precursors from \textit{E. gracilis}. The window size is 11.

wissenschaftlichen Forschung" (P12078-MOB, to W. L.) and by a grant from the Ministry of Education of the Slovak Republic (1/4234/97, to R. V. and J. K.). Research visits in Vienna (R. V.) were made possible through scholarships from FEMS and from the "Aktion Österreich-Slowakei". The authors thank Prof. Steven Schwartzbach (Lincoln) and Drs. Ruth Mould and Alison Smith (Cambridge) for helpful discussions and reviewing the manuscript.

References

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


