

 Open access • Journal Article • DOI:10.1007/BF00027337

## **Cloning and sequence analysis of cDNAs encoding the cytosolic precursors of subunits GapA and GapB of chloroplast glyceraldehyde-3-phosphate dehydrogenase from pea and spinach — Source link**

Henner Brinkmann, Rüdiger Cerff, Michael Salomon, Jürgen Soll

**Institutions:** Joseph Fourier University, Ludwig Maximilian University of Munich

**Published on:** 01 Jul 1989 - Plant Molecular Biology (Kluwer Academic Publishers)

**Topics:** Peptide sequence

Related papers:

- [Prokaryotic features of a nucleus-encoded enzyme. cDNA sequences for chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases from mustard \(\*Sinapis alba\*\).](#)
- [Evidence in favor of the symbiotic origin of chloroplasts: Primary structure and evolution of tobacco glyceraldehyde-3-phosphate dehydrogenases](#)
- [Evidence for a chimeric nature of nuclear genomes: eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes.](#)
- [Molecular Cloning: A Laboratory Manual](#)
- [Endosymbiotic origin and codon bias of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/cloning-and-sequence-analysis-of-cdnas-encoding-the-2ictgwgh16>

## Cloning and sequence analysis of cDNAs encoding the cytosolic precursors of subunits GapA and GapB of chloroplast glyceraldehyde-3-phosphate dehydrogenase from pea and spinach

Henner Brinkmann,<sup>1</sup> Rüdiger Cerff,<sup>1\*</sup> Michael Salomon<sup>2</sup> and Jürgen Soll<sup>2</sup>

<sup>1</sup>Laboratoire de Biologie Moléculaire Végétale, CNRS UA 1178, Université Joseph Fourier, B.P. 53 X, F-38041 Grenoble, France (\*author for correspondence); <sup>2</sup>Botanisches Institut, Universität München, Menzinger Str. 67, D-8000 München 19, FRG

Received 23 December 1988; accepted in revised form 4 April 1989

**Key words:** beta-tubulin, gene duplication, isoenzyme interconversion, molecular evolution, phosphate translocator, transit peptides

### Abstract

Chloroplast glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is composed of two different subunits, GapA and GapB. cDNA clones containing the entire coding sequences of the cytosolic precursors for GapA from pea and for GapB from pea and spinach have been identified, sequenced and the derived amino acid sequences have been compared to the corresponding sequences from tobacco, maize and mustard. These comparisons show that GapB differs from GapA in about 20% of its amino acid residues and by the presence of a flexible and negatively charged C-terminal extension, possibly responsible for the observed association of the enzyme with chloroplast envelopes *in vitro*. This C-terminal extension (29 or 30 residues) may be susceptible to proteolytic cleavage thereby leading to a conversion of chloroplast GAPDH isoenzyme I into isoenzyme II. Evolutionary rate comparisons at the amino acid sequence level show that chloroplast GapA and GapB evolve roughly two-fold slower than their cytosolic counterpart GapC. GapA and GapB transit peptides evolve about 10 times faster than the corresponding mature subunits. They are relatively long (68 and 83 residues for pea GapA and spinach GapB respectively) and share a similar amino acid framework with other chloroplast transit peptides.

### Introduction

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH, phosphorylating) is present in both prokaryotes and eukaryotes and is highly conserved across all species with respect both to sequence (for review see [26]) and three-dimen-

sional structures [3, 28, 33]. This indicates that all modern GAPDH variants arose from a single ancestral enzyme without dramatic rearrangements such as deletions and insertions of large polypeptide segments.

Higher plants contain three different phosphorylating GAPDH species: two chloroplast

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers: X15188 for clone pPsGapB1 from pea, X15189 for clone pSoGapB1 from spinach, and X15190 for clone pPsGapA1 from pea.

GAPDH isoenzymes I and II (EC 1.2.1.1.13), for which different subunit structures have been suggested, A<sub>2</sub>B<sub>2</sub> and A<sub>4</sub> respectively, and a single cytosolic GAPDH (EC 1.2.1.12), with the proposed subunit structure C<sub>4</sub> ([6, 8]; for review see [7]). All subunits are encoded in the nucleus [9] and cDNAs have been characterized encoding GapA and GapC from mustard and maize [4, 26], GapC from barley [12], and GapA, GapB and GapC from tobacco [32]. These studies revealed that the sequences of chloroplast GapA and GapB are different from that of GapC and similar to that of the GAPDH from thermophilic eubacteria, suggesting that GapA and GapB are of prokaryotic origin.

The sequences of tobacco GapA and GapB reported by Shih *et al.* [32], although probably incorrect in places (see below), indicate that subunits A and B of the chloroplast enzyme are closely related proteins. This is in basic agreement with recent protein sequencing data on the spinach enzyme by Ferri *et al.* [16]. These findings disagree, however, with our preceding cloning results [10] which seemed to suggest that subunit B of chloroplast GAPDH from pea may be related to beta-tubulin. In an attempt to clarify this discrepancy we tried to identify cDNA clones coding for GapB of angiosperm species other than tobacco. Here we report the cloning and sequencing of cDNAs containing the entire coding sequences of the cytosolic precursors for GapB from pea and spinach and for GapA from pea. The results suggest that GapB is a highly conserved GAPDH protein whose origin antedates angiosperm evolution. It differs from chloroplast GapA in about 20% of its amino acid residues and by the presence of a flexible and negatively charged C-terminal extension probably not necessary for catalytic activity.

## Material and methods

### *Preparation and western blot analysis of stroma and envelope fractions from isolated pea chloroplasts*

Envelopes were isolated from Percoll-purified pea chloroplasts as described by Douce and Joyard

[14]. Stroma fractions correspond to 100 000 g supernatants from hypotonic lysates of isolated chloroplasts [14]. Western blot analysis with protein preparations from total isolated chloroplasts, stroma and envelope fractions respectively was performed according to Beisiegel [1] after electrophoretic separation of proteins on dodecyl sulphate-polyacrylamide gels (12%) according to Laemmli [23]. The antiserum used in this analysis was raised against chloroplast GAPDH from mustard [8] and recognizes both subunits, GapA and GapB, from all higher plants.

### *Construction and immuno-screening of a cDNA expression library from spinach ( $\lambda$ gt11)*

The spinach cDNA expression library used in the present study was a gift from Prof. R.G. Hermann, Botanical Institute of the University of Munich, FRG. For the construction of this library cDNA was synthesized from poly(A)<sup>+</sup> mRNA isolated from etiolated spinach seedlings exposed to continuous white light for 16 h prior to harvest and cloned into the *Eco* RI site of  $\lambda$ gt11 DNA [34]. The library was screened with a rabbit antiserum raised against a mixture of inner and outer envelope membranes from spinach chloroplasts, in essentially the same way as described by Young and Davis [37].

### *Construction and screening of a cDNA library from pea ( $\lambda$ NM1149)*

Total RNA from 50 g light grown pea seedlings (*Pisum sativum* var. 'Rosakrone') was extracted and the poly(A)<sup>+</sup>-containing fraction was purified by adsorption on oligo(dT) cellulose (type 2, Collaborative Research) as described previously [9].

Construction of pea cDNA was performed as described by Gubler and Hoffmann [19] and modified by H. Sommer (Max-Planck-Institut für Züchtungsforschung, Köln, FRG, personal communication). Synthesis of first-strand cDNA was carried out in a reaction volume of 100  $\mu$ l containing 50 mM Tris/HCl (pH 8.3), 50 mM KCl,

8.0 mM MgCl<sub>2</sub>, 5.0 mM DTT, 1.0 mM of each dATP, dGTP and dTTP, 0.3 mM dCTP, 50  $\mu$ Ci of ( $\alpha$ -<sup>32</sup>P) dCTP (3000 Ci/mmol), 30  $\mu$ g/ml oligo(dT), 15  $\mu$ g/ml actinomycin D, 120  $\mu$ g/ml BSA, 0.75 U/ml RNasin, 40  $\mu$ g/ml poly (A)<sup>+</sup>-mRNA and 40 U reverse transcriptase. The mixture was incubated for 5 min at room temperature, then 10 min at 37 °C and subsequently 15 min at 42 °C. Then another 40 units of reverse transcriptase were added and the solution was incubated for final 15 min at 42 °C. The reaction was stopped by adding EDTA to 20 mM, NaCl to 400 mM and SDS to 0.2% final concentrations. The products were extracted once with phenol/chloroform. The single stranded DNA was precipitated with ethanol out of 2 M ammonium acetate as described by Maniatis *et al.* [25]. This procedure removes more than 99% of unreacted deoxyribonucleoside triphosphates.

Synthesis of second-strand cDNA was carried out in a reaction volume of 100  $\mu$ l containing about 500 ng DNA/RNA hybrid, 20 mM Tris/HCl pH 7.5, 5.0 mM MgCl<sub>2</sub>, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 100 mM KCl, 0.5 mM dNTP, 10 U/ml RNase H and 25 U DNA Polymerase I. The mixture was incubated 60 min at 12 °C and then 60 min at 22 °C. The reaction was stopped as described above. The products were extracted with one volume phenol/chloroform. The double-stranded cDNA was fractionated on a 5 ml Sepharose 4BCL column. The peak of radioactivity, comprising the longer cDNAs, was collected and alcohol-precipitated. The cDNA was methylated at internal *Eco* RI restriction sites with *Eco* RI methylase [24]. Blunt ends were produced by incubation with mung bean nuclease (FPLC pure, Pharmacia) and ligated with phosphorylated octameric *Eco* RI linkers [25]. After digestion with *Eco* RI the bulk of non-ligated linkers was separated from the cDNA by gel filtration on Sepharose 4BCL. The double-stranded cDNA was size-fractionated on a 1.5% agarose gel and the size range of 1–10 kb was used for subsequent ligation into the *Eco* RI site of the lambda cloning vector NM1149 [27]. A commercial kit (Stratagene, Gigapack plus) was used for the *in vitro* packaging of the concatemeric recom-

binant lambda DNA. The packaging efficiency is about  $1 \times 10^7$  pfu/ $\mu$ g cDNA.

Recombinant phages are selected by plating the packaged phages onto the *Escherichia coli* host POP13b [30] which is a derivative of the strain POP101 [27]. Plaques were transferred to nitrocellulose filters according to Benton and Davis [2] and the transferred material was denatured and fixed as described by Maniatis *et al.* [25]. The filters were baked for 30 min at 80 °C and hybridized with <sup>32</sup>P-labelled cDNA probes specific for chloroplast GAPDHs. The standard pre-hybridization solution contains 6  $\times$  SSPE, 0.02% PVP 350 and Ficoll 400 (SERVA), 0.1% SDS and 50  $\mu$ g/ml calf thymus DNA. The hybridization solution was identical except that it contained 3  $\times$  SSPE and in addition the radioactive probe. The hybridization was performed for 16 h at 65 °C and at 55 °C for the homologous and heterologous probes respectively. The filters were washed two times for 20 min in 2  $\times$  SSPE, 0.1% SDS at the above temperatures. Positive plaques were rescreened and purified to single plaques by standard procedures.

#### *Subcloning in plasmids and sequence analysis*

All cDNA insertions obtained from the purified recombinant lambda phages after digestion were subcloned into the *Eco* RI site of the plasmid vector 'Bluescript' (Stratagene cloning systems) and submitted to DNA sequence analysis by the dideoxy chain termination method [29]. The spinach clone (pSoGapB1) was sequenced directly in the Bluescript vector as double-stranded DNA by using the Stratagene protocol. The pea clones (pPsGapA1 and pPsGapB1) were sequenced as single-stranded DNA after subcloning of suitable restriction fragments into the vector M13 by following the Amersham protocol.

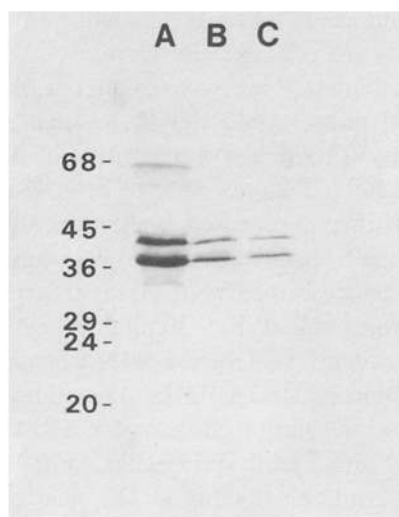


Fig. 1. Western blot analysis showing the distribution of chloroplast GAPDH (GapA and GapB) in fractions containing total purified chloroplasts (lane A), stroma (lane B) and mixed envelope membranes (lane C), all from pea. Equal amounts (25  $\mu$ g) of envelope and stroma proteins were loaded on a dodecyl sulphate polyacrylamide gel [23], while a 5fold excess of purified chloroplast proteins was used. The antibodies recognize both subunits, GapA and GapB [8].

## Results and discussion

### Association of chloroplast GAPDH with chloroplast envelopes *in vitro*

A cDNA clone coding for part of GapB from spinach has been found fortuitously by screening the spinach cDNA expression library (see below) with antibodies raised against total chloroplast envelope proteins. The clone was identified by a computer search using the Protein Sequence Data Bank of the Martinsried Institute (MIPS; Martinsried, FRG). We probed Western blots of pea envelope and stroma fractions with our monospecific antiserum raised against chloroplast GAPDH from mustard [8]. As shown in Fig. 1, GapA and GapB can both be detected in envelope fractions, suggesting that these preparations are associated to some degree with chloroplast GAPDH ( $A_2B_2$ ). Whether this finding is physiologically significant or whether it is due to an artefactual contamination *in vitro* cannot yet be decided (see below). However, one interesting

### A. Transit peptide coding sequences

```

2                                     >ccaaca
3                                     >gc aaaacttgta taactacaaac ttcacottca aaatattcag tgttcottgt tggctcaatc

                                     10                                     20
1                                     >GCT GCT TTA GCT TCT ACA AGA ATC CCA ACA AAC ACA AGG TTT CCA TCT AAG ACC TCT CAC ---
2 ATG GCT TCT CAT GCA GCT CTA GCT CCT TCT AGA ATC CCT GCA AGT ACA AGG CTC GCG TCT AAG GCT TCT CAG CAG
3 ATG GCT TCG --- GCT ACT TTC TCT GTA GCC --- --- --- --- --- --- --- AAA CCA GCT ATT AAG

                                     30                                     40                                     50
1 --- TCT TTC CCA TCT CAA TGT GCC TCA AAG AGA CTT GAG GTA GGT GAA TTC TCT GGA CTC AAA TCA ACT TCA ---
2 TAC TCC TTT CCT ACT CAA TGC TCC TTT AAG AGA CTC GAC GTA GCT GAC TTT TCT GGA TTA CGA TCC AGC AAC ---
3 --- --- --- --- GCA AAT GGG --- --- AAA GGC TTC --- --- TCT GAA TTC TCT GGT CTC CGC AAC TCT TCA AGA

                                     60                                     70
1 TGT ATT TCC TAT GTT --- CAT --- AGT GCT AGA GAT TCT TCT TTT TAT GAT GTT GTA GCT GCT CAA CTC ACT TCC
2 AGC GTG ACA TTC --- --- --- --- --- ACA AGG GAG GCT TCA TTT CAT GAT GTC ATA GCT GCA CAG CTA ACC ACC
3 CAT CTT CCC TTT TCT AGA AAA --- TCT TCA GAT GAT --- --- TTT CAT TCT CTT GTT ACC TTC CAA --- ACC AAT

                                     80                                     90
1 AAG GCA AAT GGA TCA ACT GCT GTG --- --- AAG GGA --- GTG ACT GTG GCT      GapB pea      (80 codons)
2 AAG CCT ACA GGA GCA GCA CCT GTT --- --- AGG GGT --- GAA ACA GTG GCC      GapB spinach (83 codons)
3 --- GCA GTT GGA AGT AGT GGA GGA CAC AAG AAA AGT CTT GTA GTG GAA GCA      GapA pea      (68 codons)

```

Fig. 2. A. Nucleotide sequence alignments of the regions encoding the transit peptides of GapB from pea and spinach (lines 1 and 2) and GapA from pea (line 3). Insertions and deletions have been introduced to maximize homology. Codons have been numbered consecutively. The first cloned nucleotide is indicated by an arrowhead. The 5' non-coding leader sequences are shown in lower-case letters. B. Nucleotide sequence alignments of the regions encoding the mature subunits of GapB from pea and spinach (lines 1 and 2) and GapA from pea (line 3). Codons are numbered according to Harris and Waters [20]. Stop codons and polyadenylation signals are underlined. The 3' coding trailer sequences are shown in lower-case letters. The sequences are compared with GapB from pea (line 1), the only sequence written in full. In the two other sequences (lines 2 and 3) only nucleotides not identical to the reference are indicated; identical positions are marked by points and deletions by dashes. The first and the last codon of the mature subunits are shown regardless of homology.

## B. Mature subunit coding sequences

```

0                               10                               20
1 AAG --- TTG AAG GTA GCA ATC AAT GGT TTT GGA CGC ATC GGT AGA AAT TTC CTT CGA TGC TGG CAC GGT CGA AAG
2 AAA --- C.T ... ..G ... ..A ... ..T ... ..G ... ..A ... ..C ... ..C ..A
3 AAA CAA C... ..C ..A ... ..A ... ..A ..T ..A ..G ..C ... T.G A... ..T ... ..T ... ..C ...
  25A B C                               30                               40
1 GAC TCG CCA CTT GAA GTC ATT GTT GTC AAT GAC AGT GGA --- GGT GTC AAG AAT GCT TCA CAT TTG TTG AAA TAT
2 ... ..A ..C ... ..T ..T G... ..C ... ..G ... ..A A... ..C ... C.C ..G ...
3 ... ..T ... ..T ... ..CA A... ..CC ... --- ... ..A ... C.A ... ..T ... C.T C.C ..G ...
  50                               60                               66A                               70
1 GAT TCT ATG CTT GGA ACT TTT AAA GCA GAA GTG AAG ATA CTA AAC AAT GAG ACT ATT ACA GTT GAT GGT AAA CCC
2 ..C ..C ..A ..A ... ..C ..C ... ..C ..T ... ..A ..T A.T G... ..C ... ..C T.. T.C A.C ... ..C ..G ..T
3 ... ..C .CA ... ..TC ... G.T ..T ..T ... CCT G.T GGT .C. ..T GGC ..C T... ..A ... ..A ..G GTT
  80                               80                               90
1 ATC AAG GTT GTC TCT AGC AGA GAT CCT CTT AAG CTT CCT TGG GCC GAA CTT GGA ATT GAC ATT GTT ATT GAG GGA
2 ... ..A ... ..C GA. C.C A.C ... GCC ..C ... ..AAG ..G T.G ..G ..A ... T.G ..G ... ..A ...
3 ... ..A ... ..C GA. C.C A.C ... GCC ..C ... ..AAG ..G T.G ..G ..A ... T.G ..G ... ..A ...
  100                               110                               120
1 ACA GGA GTG TTT GTG GAC GGC CCT GGC GCG GGC AAA CAC ATC CAA GCA GGT GCC AAG AAA GTT ATC ATC ACT GCT
2 ... ..C ... ..T ..T ... ..A ..T ..A ... ..T ... ..T ... ..A ... ..A ... ..C ... ..T ... ..C ...
3 ... ..T ... ..A GAA ..T ..A ..G ..GG ... ..T AC. ... ..G ..T ... ..G ... C... ..C ... ..C
  123 A B                               130                               138 A                               140
1 CCT GCA AAG GGT GCT GAT ATT CCG ACT TAC GTT ATT GGA GTG AAC GAA CAA GAC TAC GGC CAT GAA GTA GCC GAC
2 ..G ... ..C ... ..T ... ..C ... ..C ..T ... G... ..G A... ..G A... ..T ..T ..C ..T ..T ..A A...
3 ... AG. ..A ..A --- ..C ..T ... ..T ..G G... ..T ..C ..T ..CT G.T ..CT ... AC. ..C ..CC ..AC ..A. ---
  150                               160
1 ATC ATA AGC AAT GCT TCT TGC ACC ACA AAC TGT CTT GCT CCC TTT GCT AAG GTC CTG GAT GAA GAG TTC GGA ATC
2 ... ..C ... ..C ... ..C ... ..C ... ..G ... ..A ... ..TG ... ..T ... ..A ... ..A ... ..T
3 ... ..C ... ..C ... ..T ... ..C ... ..TC ... ..T ... C.G A.A ... ..T ...
  170                               180
1 GTT AAG GGA ACC ATG ACA ACC ACA CAT TCC TAC ACC GGA GAC CAG AGG CTT TTG GAT GCT --- TCA CAT AGG GAC
2 ..G ... ..G ..A ... ..C ... ..T ..C ... ..T ..T ... ..G ... ..G ... ..T ..C ... ..C ...
3 A.C ... ..T ... ..T ..T ..T ..C ... ..T ... ..A C... ..C.T ..C ..G --- AGC ..C C.T ...
  200                               210
1 TTG AGA AGA GCT AGA GCT GCA GCA CTG AAC ATT GTT CCG ACC AGC ACA GGA GCA GCC AAG GCT GTA TCT CTA GTG
2 ... ..C ..G ... ..G .. T... ..A ..C ..A ... ..T ..T ..T ... ..G ... ..G ... ..T ..A ..A ..G G.C ..T ..C
3 C.A ..G CGT ..G ... ..A ... ..C ..C ... ..A ..C ..A ..A TCA ... ..T ..A ..A ..G G.C ..T ..C
  220                               230                               240
1 TTG CCA CAG CTC AAG GGA AAG CTC AAC GGA ATC GCC CTC CGT GTG CCT ACG CCT AAT GTT TCA GTT GTT GAC CTT
2 ..A ..C ..A ..T ... ..T ... ..T ... ..T ... ..C ..A ..A ... ..C ..G ..T ... ..T ...
3 C.C ... ACA ... ..A ..C ... ..T ..T ..G ..T ... ..A ..A ..A ..C ... ..G ..G ..G ... ..C
  249 A                               250                               260
1 GTG GTC AAT GTT GCG AAG AAG GGT ATA TCA GCT GAA GAT GTC AAT GCA GCA TTC AGA AAG GCA GCT GAG GGA CCA
2 ..T ... ..C A... ..A... ..GTA ... G.C A... ..G ..C ..A ..C AAT ..C ... ..G ..A ... ..A ..CA ..T ...
3 ..C ..T C.A ..C T.A ... --- ..C ..TT ... ..A ..G ... ..AG ..T ..T ... G... AGT ..A ..CC AAG GAG
  270                               280                               290
1 CTG AAA GGT ATA TTG GAT GTC TGT GAT GTT CCG CTC GTG TCT GTT GAC TTC CGC TGC TCC GAT GTT TCT ACA ACT
2 T... ..G ... G.G ... ..G ..C ..C A.C ..C ... ..A ... ..T ... A.G ... ..T ..T ... T.C ..A T.T ..A
3 ... ..CT ... ..T C.C TCG ... ..C ..C ..A ... ..A ... ..A ..T ..T A.G ... ..A ... ..G ..G T.T ..C
  300                               310
1 ATT GAC TCT TCC TTG ACT ATG GTC ATG GGA GAT GAT ATG GTT AAG GTG GTT GCT TGG TAT GAC AAT GAA TGG GGT
2 ..C ..T ..A ..A C.C ... ..A ... ..T ..GT ... ..C ... ..C ... ..C ... ..G ... ..G ...
3 G... ..T ..G ..A ... ..A ... ..C ... ..C C.A ... ..A ... ..A ... ..A ... ..T ... ..T ... ..G ...
  320                               330                               340
1 TAC AGC CAA AGA GTG GTG GAT TTG GCA CAT CTA GTA GCA AAC AAA TGG CCA GGA ACC CCT AAA --- GTA GGG AGT
2 ... ..C.G ... ..C ... ..G.C T.G ... ..G ... ..G ... TTG GAG GGA TCA ..T ..CA ..C
3 ... TCA ... ..G ..T ..T ... ..T G.C A.T ..T ..C ..T ..C ... AAG TAA
  350                               360
1 GGA GAT CCA TTG GAG GAC TTC TGC GAG ACG AAT CCG GCC GAC GAG GAA TGC AAA GTT TAT GAA TAG
2 ... ..A ..T ... ..AA. GAC ..C ..T ..T ..T ... ..G ... ..C... ..C ..G TAA
  +30                               +60                               +90
1 aggagaatca ttttctgact tgattattac ctctaccagc aagccatttg ttttttccct tttggtatca acgtagatg atttcgttca
2 aaaaagatgt gatcaacttt acacttcttt ttggttattc tgtatttggc ctattacaag gctgtacatt tgattttata gcatcagctt
3 gagccaataa aggtttgta attatatcta tatctatata tatatatata tatatatata tatatatatg tacgtatacc atgtgcatt
  +120                               +150                               +180
1 tatgagaaaa aagaacaaat ttttcagaca aacaagtgtt atttgatgtc tgaggatttg actctaagtt tacatttato tgtggccaat
2 ctcatcttgg aaattgttg caattgaaca aactattagg ctataaatc toagccoacac taccagctct oggctttatt tagcttcgaa
3 gtaattttgt acacttcttg cttatttttt tctttctaaa cttgtattga gaacagaacc atcaatttct tcttgctctc cattgtgcaa
  +210
1 aatgagtttt ggtgtaacaa tgagtcaata gtaaccattat cattatan
2 tttgtgtan
3 taattgtttt taccttctg an

```

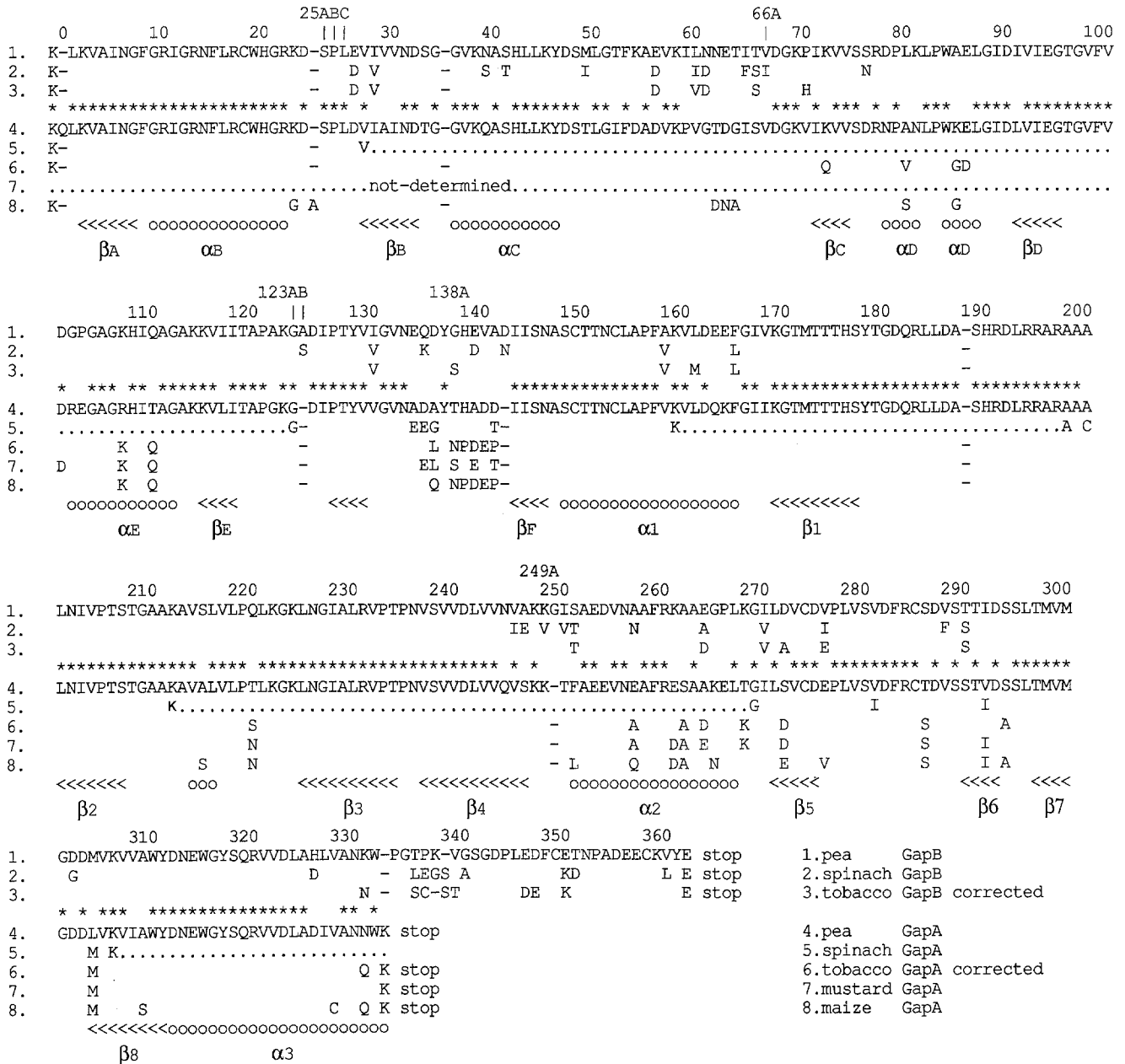


Fig. 3. Amino acid alignment of 3 GapB (lines 1 to 3) and 5 GapA (lines 4 to 8) sequences from five different angiosperms, as specified at the bottom of the figure. The amino acids are aligned to maximize homology and they are numbered according to Harris and Waters [20]. The sequences in each group have been compared to GapB and GapA from pea respectively, the only sequences written in full. Only amino acids not identical to these reference sequences are shown in either group. The first and the last residus of each sequence are indicated irrespective of homology. Sequence regions not determined are indicated by dotted lines. Positions occupied by identical residues in all 8 sequences are indicated by asterisks in the line separating the two sequence blocks. Sequence elements forming helices (o o o o o) and  $\beta$ -structures (<math>\beta</math> <math>\beta</math> <math>\beta</math> <math>\beta</math>) are indicated below the sequences [3]. Sources of sequence informations: GapA and GapB from pea, this paper; GapA and GapB from tobacco, [32] (corrected versions, see Figs. 4A and 4B); spinach GapA, 5 tryptic peptides, [16]; spinach GapB, this paper; mustard GapA, partial sequence, [26]; maize GapA, [4].

possibility is that chloroplast GAPDH binds to the triosephosphate-3-phosphoglycerate-phosphate translocator of the inner membrane of the chloroplast envelope (G. Wildner and R. Hallick, personal communication, see below).

*Primary structures of chloroplast GAPDH cDNA clones and the deduced amino acid sequences of their products (cytosolic precursors of GapA and GapB)*

The cDNA libraries have been constructed from total poly(A)<sup>+</sup> mRNA of light-grown seedlings [9] and have been cloned into the lambda vectors gt11 (spinach) and NM1149 (pea) respectively (see Materials and methods). The cDNA clone encoding part of GapB discovered as described above was used to identify longer clones for GapB from both spinach (clone pSoGapB1) and pea (clone pPsGapB1). An almost full-length clone encoding GapA from pea (pPsGapA1) was found by means of the partial cDNA clone pP71-11 previously characterized [10].

For all clones both DNA strands have been sequenced (see Material and methods). The complete nucleotide sequences comprising the regions encoding transit peptides and mature subunits are shown in Figures 2 A and B respectively. Sequences have been aligned to maximize homology and the codons of the mature subunits have been numbered according to Harris and Waters [20]. The corresponding derived amino acid sequences are shown in Fig. 3 (mature subunits) and

Figure 5 (transit peptides) respectively, together with all other chloroplast GAPDH sequences currently known (see below). To facilitate comparisons the standard alignment has been maintained for all figures. Some general informations characterizing the present clones pSoGapB1, pPsGapB1 and pPsGapA1 are summarized in Table 1. The three clones do not show any sequence similarities in their 3' non-coding regions and all contain a poly(A) tract.

*Chloroplast GAPDH and beta-tubulin*

The present results, showing that subunits A and B are closely related GAPDH proteins, are consistent with the findings of Shih *et al.* [32] and Ferri *et al.* [16] (see below). They are in conflict, however, with the interpretation of our previous cloning data [10] which suggested that subunit B may be related in structure to  $\beta$ -tubulin. In a parallel study we have characterized several cDNAs and genomic clones encoding pea  $\beta$ -tubulins (publication in preparation). These clones include an almost full-length version of the partial cDNA clone pP18-12 whose hybrid-released translation product comigrates on dodecylsulfate gels with the precursor of subunit B (GapB) and which, apparently, can be transported into chloroplasts [10]. However, all these clones encode authentic  $\beta$ -tubulins which do neither contain a recognizable transit peptide nor is their synthesis regulated by light as would be

*Table 1.* General characteristics of cDNA clones encoding the cytosolic precursors of chloroplast GapB from pea (pPsGapB1) and spinach (pSoGapB1) and of GapA from pea (pPsGapA1). The clones contain the complete coding sequences except for clone pPsGapB1 which encodes a transit peptide from which the four N-terminal residues are missing (value in parenthesis; see Fig. 2A and Fig. 5).

	Nucleotide sequences					Amino acid sequences		
	total length	5'-leader bp	3'-trailer bp	polyadenylation signal		total length	mature protein	transit peptide
				position	sequence			
pPsGapB1	1567	–	226	+ 178	aataat	447	367	(80)
pSoGapB1	1546	6	187	+ 133	aataaa	451	368	83
pPsGapA1	1477	62	200	+ 179	aataaa	405	337	68



expected for a true subunit of chloroplast GAPDH [9]. In addition, the occurrence of these clones is almost 5-fold lower than those encoding GapA and GapB. Together these results require a reinterpretation of our preceding cloning data. Our previous suggestion that  $\beta$ -tubulin is an integral component of chloroplast GAPDH [10] is probably wrong.

#### *The published sequences for cDNAs encoding tobacco GapA and GapB apparently contain errors*

In Figure 3 all chloroplast GAPDH sequences currently known are aligned. They include those

for the GapBs from pea, spinach (this paper) and tobacco [32] respectively, and the GapAs from pea (this paper), tobacco [32] and maize [4]. In addition to these full-length amino acid sequences we also included the partial GapA sequences from mustard [26] and spinach [16]. The partial spinach GapA sequence covers about one third of the total mature subunit and comprises a mosaic of sequence data from five tryptic peptides sequenced at the protein level.

A number of sequences of tryptic peptides have also been reported for spinach GapB [16]. All GapB peptides are identical with the corresponding regions of our deduced GapB sequence

#### A. Correction of tobacco GapB

	340		350	
1'/2'	S V G/A S G D P L E D F C E/K T/D N P A D E			
1	--- GTA GGG AGT GGA GAT CCA TTG GAG GAC TTC TGC GAG ACG AAT CCG GCC GAC GAG			
2	TCA .T .CA .C ... .. .A .T ... .. A.. GAC .C .T .T .T ...			
3	AGT ACCA..A ... .. .CA... .T ..G ..T ... A.. .A ... ..T ..T ..T ...			
3'	S T G S G D P L D E F C K T N P A D E			
	↑	↑		
	360			
1'/2'	E C K V/L Y E	pea/spinach consensus		
1	GAA TGC AAA GTT TAT GAA TAG	pea		
2	..G ... .. C.. ..C ... TAA	spinach		
3	... .. C ... .. TAA	tobacco published		
3'	E C K V Y E	tobacco corrected		

#### B. Correction of tobacco GapA

	215	220	230	
1	V S/A L V L P Q/T L K G K L N G I A L R V	GapA/GapB consensus		
2	V A L V L P S L K G K L N G I A L R V	tobacco GapA corrected		
3	GTGGCCTT-GTCCCTCCCAAGCCTTAAGGGGAAGCTCAATGGCATTGCCCTCCGTTG	tobacco GapA published		
4	V A L S S Q A L R G S S M A L P L R V	tobacco GapA published		
	↑	↑		

#### C. GapB, carboxyterminal extension

	330	340	350	360	
1	N K W P G T P K - V G S G D P L E D F C E T N P A D E E C K V Y E	stop			pea
2	. . . . L E G S . A . . . . . . . . . K D . . . . . . L . .	stop			spinach
3	. N . . . S C - S T .	stop			tobacco
	Region 1	Region 2a	Region 2b		

*Fig. 4.* A. Proposed correction of the published tobacco GapB sequence. The nucleotide (lines 1, 2, 3) and corresponding deduced amino acid sequences (lines 1'/2', 3') of the GapB carboxy terminal extensions are aligned for pea, spinach and tobacco. The two additional adenine nucleotides in the published tobacco sequence [32] are marked by arrows. Elimination of these nucleotides leads to an amino acid sequence (line 3') closely related to the consensus of pea and spinach (line 1'/2'). B. Proposed correction of the tobacco GapA sequence. Residues 217 to 229 are highly conserved in all GapAs and GapBs but completely different in the published GapA sequence from tobacco [32]. A simple frameshift (+1) together with the elimination of a cytosine nucleotide 40 bases further downstream (see arrows) nullifies this discrepancy suggesting that it may be due to a sequencing error. C. The C-terminal extension of GapB is composed of 2 regions: a flexible 'hinge' region (region 1) of 10 or 11 residues and rich in prolines and glycines, and a second region rich in negatively charged amino acids (Asp and Glu). Region 2 can be subdivided into two similar halves of 9 and 10 residues respectively (2a and 2b) suggesting that it originated by a duplication.

including peptide B<sub>2</sub> corresponding to amino acids 332–350 of the carboxy terminal extension. As shown in Figs. 2 and 3 this C-terminal extension (amino acids 333–362), specific for GapB and absent in GapA (see below), is well conserved in pea and spinach. This is in striking contrast to the finding of Shih *et al.* [32] showing a completely different C-terminal extension for tobacco GapB, which is almost twice as long and without any sequence similarity to that of pea and spinach. A close inspection of this region in tobacco GapB reveals that an amino acid sequence very similar to our consensus sequence is obtained by deleting two single adenine nucleotides between codons 339/340 and 344/345 respectively. The proposed correction and the resulting amino acid sequence for tobacco are shown in Figure 4A. We suggest that the two additional nucleotides shown in the cDNA for tobacco GapB were the result of a sequencing error or a cloning artifact, rather than the consequence of specific frameshift mutations in tobacco.

A similar discrepancy was found in the tobacco GapA sequence published by the same authors [32]. Amino acids 217–229, forming a highly conserved block in all GapA and GapB sequences (see Fig. 3), were reported to be completely different in tobacco GapA. However, by introducing a simple frameshift (+ 1) in combination with a deletion of a cytosine nucleotide 40 bases further downstream this difference in tobacco GapA is eliminated (see proposed correction in Fig. 4B),

suggesting that also in this case there was a sequencing error or cloning artifact.

*Chloroplast GapB is a highly conserved GAPDH protein whose origin antedates angiosperm evolution*

As shown in Table 2 GapA and GapB are about 81% similar (underlined values) while the similarity between species in amino acid sequences of GapA and GapB is roughly between 90 and 95% (Table 2, values above and below the diagonal respectively). This clearly indicates that the duplication event leading to GapA and GapB occurred long before the separation of angiosperm species. This conclusion is consistent with our earlier finding [9] that subunits A and B are present in the purified enzymes of 12 different monocot and dicot plant species. An observed difference of 19% corrected for multiple events at identical sites [21] corresponds to a ‘real’ difference of 23%, or 12% of amino acid mutations relative to the common ancestor. Assuming that GapA and GapB change at similar rates as animal GAPDHs (2.2% of amino acid mutations per 100 million years; see [13]), this difference would translate into a divergence time of about 550 million years. However, since the ‘molecular clock’ of chloroplast GapA/GapB is at least 1.5 times slower than that of its cytosolic counterpart CapC (see below), the real divergence date of GapA and GapB may have been considerably earlier.

*Table 2.* Identities matrix of GapA and GapB amino acid sequences and their corresponding transit peptides (values in parenthesis) showing in per cent the number of sites occupied by identical amino acids in pairwise comparisons (see Figs. 3 and 5). Underlined values: intraspecific comparisons between GapA and GapB. Values above the diagonal: GapA comparisons between species. Values below the diagonal: GapB comparisons between species. All GapA and GapB sequences are full-length. For transit peptides partial sequences have been compared starting at position 32 (see Fig. 5). Sources of sequence information see legend of Fig. 3.

	Pea	Tobacco	Maize
Pea	<u>79.4</u> (33.9)	93.2 (53.9)	90.9 (50.9)
Tobacco	94.4 (55.6)	<u>82.1</u> (32.1)	93.8 (41.5)
Spinach	90.6 (49.1)	92.3 (56.6)	–

*Chloroplast GapB has a negatively charged C-terminal extension*

In all higher plants investigated so far the B subunits are of slightly larger size than the A subunits (for a review see [7]). The present results suggest (Figs. 3 and 4C) that this size difference is due to a C-terminal extension in subunit B (GapB) comprising 29 and 30 amino acids in pea/tobacco and spinach respectively. As shown in Fig. 4C, this C-terminal extension of GapB is highly conserved in all three species and can be subdivided into three homology regions. Region 1 is rich in prolines and glycines, residues which are strong disrupters of secondary structure. Hence, region 1 might provide a flexible hinge connecting  $\alpha$ -helix 3 at the end of the catalytic domain (see Fig. 3 and [3, 28, 33]) with the rest of the C-terminal extension. Regions 2a and 2b both start with proline and have an unusually high content of negatively charged residues. They are similar in sequence suggesting that they may have originated by a duplication. These properties of flexibility and charge (6, 7 and 5 negative net charges for pea, spinach and tobacco respectively; see Fig. 4C) suggest that the C-terminal extension of GapB protrudes from the core structure possibly interacting with other (membrane) proteins. As indicated above, one physiologically interesting binding partner would be the phosphate translocator (shortened term for triosephosphate-3-phosphoglycerate-phosphate translocator), a major transport protein of 29 kDa of the inner envelope membrane [18]. A cDNA containing the total coding sequence of this translocator from spinach has recently been cloned by Flügge *et al.* [17]. The derived amino acid sequence reveals a strongly hydrophobic protein with a polar C-terminus of positive net charge (-Lys-Ala-Lys-Met-Glu-Glu-Glu-Lys-Arg-Gln-Met-Lys-Ser-Thr). It is enticing to speculate, therefore, that association of chloroplast GAPDH with chloroplast envelopes (Fig. 1) reflects interaction of this positively charged 'receptor domain' with the negatively charged C-terminal extension of subunit GapB, a hypothesis which can now be tested experimentally. Since the phosphate translocator is a power-

ful transport system of the inner membrane [18], the proposed association with chloroplast GAPDH *in vivo* would seem to be advantageous for the plant cell by facilitating export of photosynthetic glyceraldehyde-3-phosphate into the cytoplasm.

*Is chloroplast GAPDH isoenzyme II a proteolytic product of isoenzyme I?*

As shown by our previous biochemical studies [6, 7] chloroplast GAPDH from many higher plants can be separated by ammonium sulphate fractionation into isoenzymes I and II which have different electrophoretic mobilities in native gels, with isoenzyme I being the more negatively charged species. On dodecyl sulphate gels, the subunits of purified isoenzyme II migrate in a single band, while those of purified isoenzyme I give an additional slower-migrating band (or doublet; see below) indicating that it has a slightly larger size. These data have been interpreted as evidence for different quaternary structures  $A_2B_2$  and  $A_4$  for isoenzymes I and II respectively (for a review see [7]). However, the present results, showing that the size difference between subunits A and B is due to a negatively charged C-terminal extension of subunit B which is probably not necessary for catalytic activity, require a reevaluation of our earlier protein data. In fact, an  $A_2B_2$  species with the B subunits having lost their C-terminal extensions would migrate considerably slower on native gels and the subunits would probably be indistinguishable on dodecyl sulphate gels from those of an  $A_4$  species (compare sequences in Fig. 3). It seems possible, therefore, that isoenzyme II represents a proteolytic product of isoenzyme I ( $A_2B_2$ ) rather than a separate  $A_4$  tetramer. Indeed, in the light of the present results this interpretation would best explain why isoenzymes I and II from mustard and barley are virtually indistinguishable on the basis of tryptic fingerprints, immunological cross reactivity and amino acid compositions [8]. If one assumes that proteolytic cleavage of the C-terminal extension may occur in several steps, this concept would also

explain why subunit B in certain plants or after particular purification procedures gives protein doublets [5, 9]. Experiments are under way to distinguish at the protein level whether isoenzyme II is the proteolytic product of isoenzyme I ( $A_2B_2$ ) or whether it is a separate  $A_4$  tetramer.

*The 'molecular clock' of chloroplast GAPDH is slower than that of its cytosolic counterpart*

There is only about 45% sequence similarity between chloroplast GapA/GapB and its cytosolic counterpart GapC (see underlined values in Table 3) suggesting that the separation of the corresponding genes occurred as early as the divergence of the eubacteria and eukaryotes [4, 11, 26, 32]. If the differences between species are compared, it becomes apparent that chloroplast GapA and GapB are evolving more slowly than GapC. As shown in Tables 2 and 3, the average amino acid differences between the dicot species pea, mustard, tobacco and spinach are 6.4, 7.6 and 11.7% for GapA, GapB and GapC respectively. This corresponds to a 1.8- and 1.5-fold slower rate of change for GapA and GapB respectively relative to GapC. The average differences between maize and dicots are 7.7% for GapA and 16.2% for GapC (see Table 3), suggesting a 2.1-fold slower rate of change in maize GapA relative to maize GapC. We have previously interpreted this as being a consequence of the

extreme codon bias in the maize GapA gene which uses 97% G + C in the triplet third base positions [4]. The above observation of a considerable rate reduction in the non-biased GapA/GapB genes from dicots suggests that codon bias may only be one possible factor among several interacting molecular constraints. Another more specific factor slowing down the rate of chloroplast GAPDH relative to GapC may be molecular coevolution [15] forcing GapA and GapB to evolve in 'mutual consent' and in a cohesive manner. The GapC subunit, which is arranged as symmetrical homotetramer in the native enzyme, would not be subjected to the same constraints.

*Structure and evolution of the transit peptides of GapA and GapB*

The first 30 amino-terminal residues of purified spinach GapA and GapB have been determined by Ferri *et al.* [16]. According to this analysis and the sequence alignments shown in Figs. 3 and 5 the GapA and GapB precursors (preGapA and preGapB) of angiosperms have a conserved processing site alanine/lysine. With threonine or valine and alanine in positions -3 and -1 respectively (see Fig. 5, positions 90 and 92) it conforms well to the (-3, -1) rule established for the processing site of signal sequences of secretory proteins by von Heijne [35].

*Table 3.* Identities matrix of GapA and GapC amino acid sequences showing the number of sites in percent occupied by identical amino acids in pairwise comparisons. Underlined values: intraspecific comparisons between GapA and GapC. Values above the diagonal: interspecific comparisons of GapA polypeptides. Values below the diagonal: interspecific comparisons of GapC polypeptides. Values in parenthesis: comparisons of partial sequences starting at amino acid 100. Sources of sequence information: pea, this paper (GapA) and unpublished (GapC); mustard GapA (partial sequence) and GapC, [26]; tobacco GapA (corrected version) and GapC, [23]; maize GapA and GapC, [4].

GapC	Pea	Mustard	Tobacco	Maize
GapA				
Pea	<u>44.4</u>	(92.7)	93.2 (92.3)	90.9 (90.2)
Mustard	88.1	-	(94.9)	(92.3)
Tobacco	87.4	89.5	<u>44.7</u>	93.8 (94.4)
Maize	82.7	85.4	83.4	<u>45.2</u>

Block I	Block II	Block III	
1. MA+SM+SS	PF+G+K P	S	1. general consensus
2. MASSM+SS	PF+GLK P	S	2. SSU consensus
3. >AALASTRIPTNTRFPSKTS--SFPSQCASKRLEVGFSGLKSTS-CISYV-H-SARDSSFYDVVAAQLTSKANGSTAV--KG-VTVA			3. GapB pea
4. MASH...PS...AS...LA...A.QQY...T...SF...D.AD...R.SN-SVTF-----T.EA..H..I....T.PT.AAP.--R.-E..A			4. GapB spinach
5. >CL..KF..A..A..R.SG..VTF-----NKE...F...S....P.TTR..P.--E..A			5. GapB tobacco
6. > K EFSGLRSS F D F VV Q T G G T A			6. GapA/GapB consensus
7. MAS--ATFSVA-----KPAIK---ANG--KGF--SEFSGLRNRRHLPFSRK-SSDD--FHSLVTFQ-TN-AVGSSGGHKKSLVVEA			7. GapA pea
8. >NSSLQ---VSN-----T...-AI..G..TN...-LL.V.A...-S-VI.GGNS..RG...A			8. GapA tobacco
9. M...SML.ATTVP-----L---QQ---GG---L-----S.A-S..MR.NAT...-M.A.S.R-H-...T...PRRAP-T.A			9. GapA maize
I I I I I I I I I I			
10 20 30 40 50 60 70 80 90			

Fig. 5. Amino acid sequence alignment of transit peptides from 3 GapB (lines 3 to 5) and 3 GapA (lines 7 to 9) precursor proteins of 4 different angiosperms (pea, spinach, tobacco, maize). The alignment is based on maximal sequence homology at the nucleotide level (nucleotide sequence alignments not shown). The transit sequences of spinach, tobacco and maize are compared to those of pea GapB (line 3) and pea GapA (line 7) respectively. Arrow heads indicate the first nucleotides of the three partial sequences (lines 3, 5, 6, 8). The three homology blocks postulated to be important for protein transport are indicated in lines 1 and 2 as general consensus and small subunit (SSU) consensus respectively [22]. The GapA/GapB consensus (line 6) shows residues which are identical in at least 4 of the 6 transit peptides. Sources of sequence information; see legend of Fig. 3.

In a previous study [9] we compared the molecular weights of preGapA and preGapB polypeptides synthesized *in vitro* (of 12 different angiosperms) with the sizes of their corresponding mature subunits synthesized and processed *in vivo*. These comparisons demonstrated that transit peptides of GapB are always longer than those of GapA. As shown in Fig. 5 the full-length transit peptides of spinach GapB, pea GapA and maize GapA are 88, 68 and 66 amino acids long respectively, and the length differences are mainly due to large deletions/insertions in the N-terminal part of the transit peptides.

In Table 2 the preGapA and preGapB sequences are compared separately for mature subunits and for the corresponding transit peptides (values in brackets; and Fig. 5) respectively. These comparisons suggest that, after correction for multiple mutations at identical sites [21], transit sequences change about 10-fold faster than the mature GapA/GapB subunits. This clearly demonstrates that the selective constraints controlling transit peptide evolution are much less stringent than those conserving catalytic structure and function in GAPDH subunits. However, in spite of this high evolutionary rate, Gap transit peptides are clearly homologous proteins not only within a given family (GapA and GapB) but also between GapA and GapB proteins as shown by a sequence conservation of over 30% in the part starting at position 32 (see Fig. 5 and underlined values shown in parenthesis in Table 2).

Gap transit peptides share a number of common features with the transit peptides from other nuclear encoded chloroplast proteins and this observation may be interpreted in terms of the 'framework concept' [22, 31]. The peptides are rich in the hydroxy amino acids serine and threonine and contain a positive net charge. The following values are obtained for the amino acid charge ratio  $(K + R)/(D + E)$ :  $9 + /4 -$  for pea GapB,  $9 + /5 -$  for spinach GapB,  $9 + /4 -$  for pea GapA and  $6 + /4 -$  for maize GapA. The relatively low positive net charge (+2) of the maize GapA transit peptide is due to the complete absence of charged residues in the 20 N-terminal amino acids. The essential features of the three homology blocks postulated to be important for protein transport [22] are more or less conserved (see Fig. 5). The amino termini of all full-length sequences start with the sequence MAS and contain at least 9 uncharged residues (Block I). Block II is reasonably well conserved and represented by the GapA/GapB consensus sequence EFSGLRSS (see Fig. 5). Blocks I and II have been proposed [22] to be essential and perhaps sufficient for mediating recognition, binding and uptake of precursors into the chloroplast. Block III seems to be specific for vascular plants and is thought to be important for proteolytic cleavage of the transit peptide at the nearby processing site [22]. Although not so well conserved at the sequence level, there is a clear preference in this region for glycine, proline, hydroxy amino

acids and positively charged residues, especially in the GapA transit peptides (see Fig. 5). This composition would probably be sufficient to keep the nearby cleavage site on the surface and in a random coil configuration.

The structure of the Gap transit peptides seem to conform less well to the model of von Heijne *et al.* [36] which is based on a comparison of 26 chloroplast transit peptides from stromal and thylakoid proteins. This model distinguishes three different regions: a non charged N-terminal region (10 residues), a central region of variable length containing few (if any) acidic residues, and a C-terminal region (8 to 10 residues) with a relatively high proportion of arginine residues and with a tendency to form an amphiphilic  $\beta$ -strand. Contrary to these predictions, most acidic residues of Gap transit peptides occur in the central region, and the C-terminal region, which contains as many lysine as arginine residues, does not seem to display a preference for the formation of amphiphilic  $\beta$ -structures [36]. These discrepancies show that a unifying concept relating structure and function of chloroplast transit peptides does not yet exist. This may be due to the relatively limited amount of sequence information available or to the existence of several chloroplast specific transport mechanisms.

### Acknowledgements

We like to thank R.G. Herrmann for supplying the *λgt11* cDNA library from spinach and the members of his group for their continuous help and advice. This work was funded by the Centre National de la Recherche Scientifique (UA 1178), the Ministère de l'Éducation Nationale (programme 'Essor de Biotechnologies') and the Deutsche Forschungsgemeinschaft (SFB 184/B8 and Ce 1/11-4).

### References

1. Beisiegel, U: Protein blotting. *Electrophoresis* 7: 1–18 (1986).
2. Benton WD, Davis RW: Screening  $\lambda$ gt recombinant clones by hybridization to single plaques in situ. *Science* 196: 180–182 (1977).
3. Biesecker G, Harris JL, Thierry JC, Walker JE, Wonacott AJ: Sequence and structure of D-glyceraldehyde 3-phosphate dehydrogenase from *Bacillus stearothermophilus*. *Nature* 266: 328–333 (1977).
4. Brinkmann H, Martinez P, Quigley F, Martin W, Cerff R: Endosymbiotic origin and codon bias of the nuclear gene for chloroplast glyceraldehyde 3-phosphate dehydrogenase from maize. *J Mol Evol* 26: 320–328 (1987).
5. Cerff R: Glyceraldehyde-3-phosphate dehydrogenase (NADP) from *Sinapis alba*. NAD(P)-induced conformation changes of the enzyme. *Eur J Biochem* 82, 45–53 (1978).
6. Cerff R: Quaternary structure of higher plant glyceraldehyde-3-phosphate dehydrogenases. *Eur J Biochem* 94: 243–247 (1979).
7. Cerff R: Separation and purification of NAD- and NADP-linked glyceraldehyde-3-phosphate dehydrogenases from higher plants. In: Edelman M, Hallick RB, Chua NH (eds) *Methods in Chloroplast Molecular Biology*, pp. 683–694. Elsevier Biomedical Press, Amsterdam (1982).
8. Cerff R, Chambers SE: Subunit structure of higher plant glyceraldehyde-3-phosphate dehydrogenases (EC 1.2.1.12 and EC 1.2.1.13). *J Biol Chem* 254: 6094–6098 (1979).
9. Cerff R, Klopstsch K: Structural diversity and differential light control of mRNAs coding for angiosperm glyceraldehyde-3-phosphate dehydrogenases. *Proc Natl Acad Sci USA* 79: 7624–7628 (1982).
10. Cerff R, Hundrieser J, Friedrich R: Subunit B of chloroplast glyceraldehyde-3-phosphate dehydrogenase is related to beta-tubulin. *Mol Gen Genet* 204: 44–51 (1986).
11. Conway T, Sewell GW, Ingram LO: Glyceraldehyde-3-phosphate dehydrogenase gene from *Zymomonas mobilis*: cloning sequencing and identification of promoter region. *J Bacteriol* 169: 5653–5662 (1987).
12. Chojecki J: Identification and characterization of a cDNA clone for cytosolic glyceraldehyde-3-phosphate dehydrogenase in barley. *Carlsberg Res Commun* 51: 203–210 (1986).
13. Dayhoff MO: *Atlas of Protein Sequence and Structure*, vol. 5, suppl. 3, p. 3. National Biochemical Research Foundation, Washington, DC (1978).
14. Douce R, Joyard J: Structure and function of the plastid envelope. *Adv Bot Res* 7: 1–116 (1979).
15. Dover GA, Flavell RB: Molecular coevolution: DNA divergence and the maintenance of function. *Cell* 38: 622–623 (1984).
16. Ferri G, Stoppini M, Iadarola P, Zapponi MC, Galliano M, Minchiotti L: Structural characterisation of the subunits of spinach chloroplast glyceraldehyde-3-phosphate dehydrogenase (NADP). *Biochim Biophys Acta* 915: 149–156 (1987).

17. Flügge UI, Fischer K, Gross A, Sebald W, Lottspeich F, Eckerskorn C: The triose phosphate-3-phosphoglycerate-phosphate translocator from spinach chloroplasts: nucleotide sequence of a full-length cDNA clone and import of the *in vitro* synthesized precursor protein into chloroplasts. *EMBO J* 8: 39–46 (1989).
18. Flügge UI, Heldt W: The phosphate-triosephosphate-phosphoglycerate translocator of the chloroplast. *Trends Biochem Sci* 9: 530–533 (1984).
19. Gubler U, Hoffman BJ: A simple and very efficient method for generating cDNA libraries. *Gene* 25: 263–269 (1983).
20. Harris JI, Waters M: Glyceraldehyde-3-phosphate dehydrogenase. In: Boyer PD (ed) *The Enzymes*, 3rd ed, vol. 13, pp. 1–48. Academic Press, New York (1976).
21. Jukes TH, Cantor CR: In: Munro HN (ed) *Mammalian Protein Metabolism*, pp. 21–123. Academic Press, New York (1969).
22. Karlin-Neumann G, Tobin EM: Transit peptides of nuclear encoded chloroplast proteins share a common amino acid framework. *EMBO J* 5: 9–13 (1986).
23. Laemmli OK: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680–685 (1970).
24. Maniatis T, Hardison RC, Lacy E, Lauer J, O'Connell C, Quon D, Sim GK, Efstratiados A: The isolation of structural genes from libraries of eukaryotic DNA. *Cell* 15: 687–701 (1978).
25. Maniatis T, Fritsch EF, Sambrook G: Synthesis of cDNA. In: *Molecular cloning; A laboratory manual*, pp. 211–246. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1982).
26. Martin W, Cerff R: Prokaryotic features of a nucleus-encoded enzyme. *Eur J Biochem* 159: 323–331 (1986).
27. Murray NE: Phage lambda and molecular cloning. In: Hendrix RW, Roberts JW, Stahl FW, Weisberg RA (eds) *Lambda II*, pp. 395–432. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1983).
28. Rossmann MG, Liljas A, Bränden CJ, Banaszak J: Evolutionary and structural relationships among dehydrogenases. In: Boyer PD (ed) *The Enzymes*, 3rd edn, vol. 11, pp. 61–102. Academic Press, New York (1975).
29. Sanger F, Nickler S, Coulson AR: DNA sequencing with chain-termination inhibitors. *Proc Natl Acad Sci USA* 74: 5463–5467 (1977).
30. Scherer G, Telford J, Baldari C, Pirrotta V: *Dev Biol* 86: 438–447 (1981).
31. Schmidt GW, Mishkind ML: The transport of proteins into chloroplasts. *Ann Rev Biochem* 55: 879–912 (1986).
32. Shih MC, Lazar G, Goodman HM: Evidence in favor of the symbiotic origin of chloroplasts: primary structure and evolution of tobacco glyceraldehyde-3-phosphate dehydrogenases. *Cell* 47: 73–80 (1986).
33. Skarzynski T, Moody PCE, Wonacott AJ: Structure of the holo-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1,8 Å resolution. *J Mol Biol* 193: 171–187 (1987).
34. Tittgen J, Hermans J, Steppuhn J, Jansen T, Jansson C, Andersson B, Nechushtai R, Nelson N, Herrmann G: Isolation of cDNA clones for fourteen nuclear-encoded hylakoid membrane proteins. *Mol Gen Genet* 204: 258–265 (1986).
35. Von Heijne G: A new method for predicting signal sequence cleavage sites. *Nucl Acids Res* 14, 4683–4690 (1986).
36. Von Heijne G, Steppuhn J, Herrmann RG: Domain structure of mitochondrial and chloroplast targeting peptides. *Eur J Biochem*, in press.
37. Young RA, Davis RW: Yeast RNA polymerase II genes: isolation with antibody probes. *Science* 222: 778–782 (1983).