Eur. J. Biochem. 159, 323–331 (1986)

Prokaryotic features of a nucleus-encoded enzyme

cDNA sequences for chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases from mustard (*Sinapis alba*)

William MARTIN and Rüdiger CERFF

Institut für Botanik, Universität Hannover

(Received February 25/June 5, 1986) – EJB 86 0205

Two cDNA clones, encoding cytosolic and chloroplast glyceraldehyde-3-phosphate dehydrogenases (GAPDH) from mustard (*Sinapis alba*), have been identified and sequenced. Comparison of the deduced amino acid sequences with one another and with the GAPDH sequences from animals, yeast and bacteria demonstrates that nucleus-encoded subunit A of chloroplast GAPDH is distinct from its cytosolic counterpart and the other eukaryotic sequences and relatively similar to the GAPDHs of thermophilic bacteria. These results are compatible with the hypothesis that the nuclear gene for subunit A of chloroplast GAPDH is of prokaryotic origin. They are in puzzling contrast with a previous publication demonstrating that *Escherichia coli* GAPDH is relatively similar to the eukaryotic enzymes [*Eur. J. Biochem. 150*, 61–66 (1985)].

It is now well established that plastid and eubacterial genes share a high degree of sequence homology [1–6] in agreement with the endosymbiotic theory of chloroplast evolution (see [7–10] for recent reviews). However, most of the chloroplast components, at least 190 [11] but probably over 300 different proteins [12], are encoded in the nuclear genome. This raises the interesting question of whether or not these nuclear genes are of prokaryotic origin. Nucleus-encoded enzymes of the photosynthetic Calvin cycle are excellent marker molecules with which to investigate this question because their primary structures may be directly compared to the primary structures of glycolytic isoenzymes located in the cytoplasm of the same cell.

Chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases (GAPDH) are especially useful molecular homology criteria for this important aspect of plant cell evolution because of the large data base available, comprising GAPDH sequences from seven eukaryotic and three prokaryotic organisms. Using protein-sequencing techniques Harris et al. established complete amino acid sequences for the GAPDHs from lobster muscle [13], pig muscle [14] and yeast [15] and subsequently also for the GAPDHs from the thermophilic eubacteria *Bacillus stearothermophilus* [16, 17] and *Thermus aquaticus* [18]. More recently cloning and DNA sequencing techniques were used to determine the primary structures of GAPDH enzymes from yeast [19], chicken [20–22], human [23, 24], rat [25, 26] and *Drosophila melanogaster* [27] and *Escherichia coli* [28]. An interspecies comparison of these GAPDH sequences reveals a high degree of sequence conservation ranging from 46% (human/*Thermus*) to 93% (human/pig; see Table 1). X-ray cristallographic analyses of the enzymes from lobster [29–31] and *B. stearothermophilus* [16] demonstrated that the aminoterminal and carboxyterminal moieties of the GAPDH polypeptide are organized as two independent three-dimensional units, which were highly conserved during the divergence of prokaryotes and eukaryotes: the coenzyme-binding domain, which has a similar structure in all NAD-dependent dehydrogenases, and the catalytic domain, which is specific for GAPDH and which differs in different dehydrogenases [32].

In green plants the classic GAPDH of glycolysis has a photosynthetic counterpart, which is NADP-dependent and located inside the chloroplast. Antisera raised against the chloroplast enzyme do not cross-react with the cytosolic GAPDH and fingerprints as well as amino acid compositions are different for the two enzymes [33–35]. The chloroplast GAPDH is also exceptional in that it is composed of two separate subunits A and B, which differ slightly in molecular mass (A $\lesssim$ B [35–38]. Both subunits are encoded by nuclear genes [39] and their identities and presumptive functions have recently been disclosed in our laboratory by molecular cloning and sequencing techniques [40]: while subunit A of chloroplast GAPDH from pea was found to represent the catalytic subunit of the enzyme (a 'true' GAPDH-like structure), the cloned partial sequence of subunit B was recognized to be highly homologous to β-tubulin.

In the present paper we report the molecular cloning and sequence analysis of cDNAs encoding the catalytic subunits of chloroplast and cytosolic GAPDHs from mustard (*Sinapis alba*). The deduced amino acid sequences are compared with one another and with the GAPDH sequences from animals, yeast and bacteria.

Correspondence to R. Cerff, Laboratoire de Biologie Moléculaire Végétale, Université de Grenoble 1, Boîte postale 68, F-38402 St Martin d'Hères Cédex, France

Abbreviations. bp, base pairs; GAPDH, D-glyceraldehyde-3-phosphate dehydrogenase; SDS, sodium dodecyl sulfate.

Enzymes. Cytosolic glyceraldehyde-3-phosphate dehydrogenase, NAD-specific (EC 1.2.1.12); chloroplast glyceraldehyde-3-phosphate dehydrogenase, NADP-dependent (EC 1.2.1.13).

MATERIALS AND METHODS

Purification, fractionation and translation of poly(A)-rich mRNAs was done essentially as described previously [39–41].

Construction of cDNA library

The cDNA was synthesized from size fractionated poly-(A)-rich mRNA from light-grown mustard seedlings essentially according to the protocol of Maniatis et al. [42] with snapback-primed second-strand synthesis, S1 nuclease digestion and C-tailing for annealing into the G-tailed *Pst*I site of pBR322. Chimeric plasmids were transformed into *E. coli* DH1 by the method of Hanahan [43] with an efficiency of 3×10^5 recombinants/μg double-stranded cDNA. Cultures (2 ml) of single colonies were grown overnight and aliquots from these were stored in microtiter plates.

Dot-blot hybridizations

From each of three different overnight cultures 0.5 ml were combined for small-scale plasmid preparation. From each pooled plasmid preparation 5 μl were denatured and dot-blotted to nitrocellulose filters. Filters were then prehybridized for 6 h at 42°C in 50% formamide, 1.0 M NaCl, 50 mM Tris/HCl pH 8.0, 100 μg/ml salmon sperm DNA and 2 μg/ml polyadenylic acid. Hybridization was performed for 48 h at 42°C after adding the nick-translated ($\approx 10^8$ cpm/μg DNA) 246-base-pair(bp) internal *Hin*dIII fragment of pGAP30 from chicken [20] or the 800-bp insert of cDNA clone pP71-11 encoding preA from pea [40]. Filters were subjected to a final wash at 42°C in 0.3 M NaCl, 60 mM Tris/HCl pH 8.0, 1% SDS and exposed on Kodak X-Omat AR film. The same blotting and hybridization procedures were used to identify single positive clones from dots (containing pools of three). Plasmids corresponding to the strongest hybridization signals were analyzed by hybrid-released translation.

Hybrid-released translation

For this procedure 50 μg plasmid DNA was bound to a 0.25-cm nitrocellulose filter, which was then washed for 3 h in 5 × NaCl/Cit and baked for 2 h at 80°C. Prehybridization was performed for 1 h at 52°C in 65% formamide, 0.4 M NaCl, 10 mM Pipes pH 6.4 and 80 μg/ml polyadenylic acid. The prehybridization solution was replaced with 120 μl/filter of 65% formamide, 0.4 M NaCl, 10 mM Pipes pH 6.4, 0.1% SDS and 0.25 mg/ml poly(A)-rich mRNA from light-grown mustard seedlings. After 6-h incubation at 52°C, filters were washed six times for 2 min at 60°C in 2 × NaCl/Cit, 0.5% SDS, twice for 1 min at 20°C in 2 mM EDTA pH 7.9 and once for 5 min at 50°C in 2 mM EDTA pH 7.9. The mRNA was eluted for 80 s at 100°C in H_2O. The eluate was precipitated with ethanol and translated in a cell-free wheat germ system.

Colony hybridizations

Colonies were grown on Schleicher and Schüll BA 85 filters over agar medium with tetracycline (10 μg/ml) for 4–8 h and transferred to agar plates containing chloramphenicol (200 μg/ml) for amplification overnight. Filters were then transferred sequentially to Whatman 3MM paper, soaked with the following solutions, and incubated for the time indicated: 0.5 M NaOH, 10 min; 1.0 M Tris/HCl pH 7.4, 10 min; 1.5 M NaCl, 0.5 M Tris/HCl pH 7.4, 5 min; 0.3 M NaCl, 5 min. Dry filters were baked for 2 h at 80°C and hybridized as described [42] to the *Pst*I inserts of cDNA clones pS6b and pS302b, each nick-translated to a specific activity of approximately 2×10^8 cpm/μg. Filters were given a final wash in 1 × NaCl/Cit, 0.1% SDS at 68°C and autoradiographed for 12 h with Kodak X-Omat AR film.

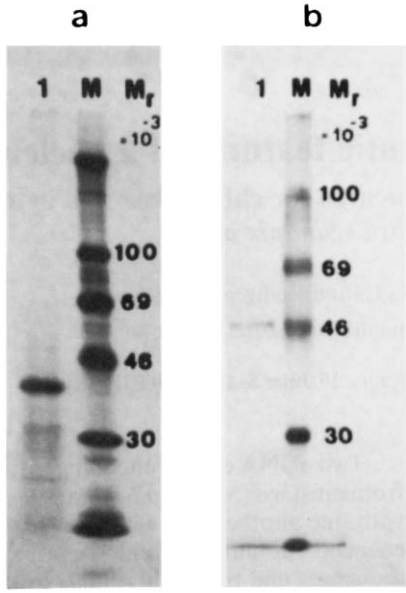

Fig. 1. *Identification by hybrid-released translation of clones pS6b and pS302b encoding cytosolic GAPDH (a, lane 1) and chloroplast GAPDH (b, lane 1) from mustard respectively.* M, marker proteins. The selected mRNAs were translated in 10 μl cell-free wheat germ system with 5 μCi [^{35}S]methionine for 60 min. The translation products were submitted to dodecyl sulfate gel electrophoresis and visualized by fluorography. They have the expected M_r values (39) and can be precipitated with our monospecific antisera (not shown)

DNA sequence determination

Both strands of clones pS198c and pS84b were sequenced according to Maxam and Gilbert [44]. Fragments were labeled at the terminal *Pst*I sites with terminal transferase and [^{32}P]dideoxy-ATP ($\approx$ 5500 Ci/mmol) or at the 3'-recessed termini with Klenow polymerase and [^{32}P]deoxynucleotides ('filling in' reactions).

RESULTS

Cloning of cDNAs and identification of specific clones

The construction of the present cDNA library was performed with pBR322 and size-fractionated poly(A)-rich mRNA from light-grown mustard seedlings [41]. The cDNA synthesis and the construction of the recombinant plasmids (see Materials and Methods) were done in collaboration with H. Sommer in the laboratory of H. Saedler (Max-Planck-Institut für Züchtungsforschung, Cologne). About 1200 transformants were analyzed. Identification of specific clones, corresponding to chloroplast (subunit A) and cytosolic GAPDHs, was achieved in four consecutive steps. (a) Dot blots were first screened with heterologous probes from chicken (cytosolic GAPDH) and pea (chloroplast GAPDH, subunit A) respectively. (b) Single clones from aparent positive dots were then analysed by hybrid-released translation and, if positive by this criterion (see Fig. 1a and b), (c) submitted to partial sequence analysis. (d) Positive clones, as defined by criteria (b) and (c), were then used as homologous probes to

rescreen the library by colony hybridization. The positive clones with the longest inserts were then selected for complete sequence analysis.

We found the cytosolic GAPDH with the heterologous chicken probe pGAP30, kindly provided by H. H. Arnold, Medical School, University of Hamburg.

Screening of the dot blots was performed with an internal *Hind*III fragment, which is 246 bp long and which comprises codons 225–305 of the catalytic domain [20]. Among the 15–20 possible candidates we found just one clone (pS6b, approx. 350 bp long), which was positive according to criteria (b) and (c).

The heterologous pea probe, clone pP71-11, harbors an 800-bp fragment of preA cDNA previously cloned in our laboratory and identified by hybrid-released translation [40]. For initial isolation of chloroplast GAPDH cDNAs, pea was best suited due to the strong light induction of preA and preB mRNAs in this species [39], providing a simple basis for differential hybridization in screening procedures. But average insert length in the pea library necessitated a second cDNA cloning, for which mustard was chosen since amino acid compositions, fingerprint data and a partial aminoterminal peptide sequence for subunit A of chloroplast GAPDH had previously been determined for the mustard enzymes [34, 41]. With the heterologous probe pP71-11 we identified the mustard clone pS302b, whose insert is about 700 bp long and which was found to be positive according to criteria (b) and (c).

Screening of our cDNA library with the two homologous probes pS6b (cytosolic GAPDH) and pS302b (chloroplast GAPDH) by colony hybridization led to the identification of two separate groups, which each contained seven strongly cross-hybridizing members. From each group the clone with the longest insert was selected, pS198C for cytosolic and pS84b for chloroplast GAPDH, and submitted to nucleotide sequence analysis according to Maxam and Gilbert [44].

Primary structures of the GAPDH messages and amino acid sequences

Both strands of clones pS198C (cytosolic GAPDH) and pS84b (chloroplast GAPDH, subunit A) have been sequenced. In Fig. 2 the coding strands for cytosolic (sequence 1) and chloroplast (sequence 2) GAPDH are aligned according to the standard scheme for GAPDH enzymes [45].

Clone pS198c (cytosolic GAPDH) is 1106 nucleotides long. A 7-nucleotide 5′-non-translated sequence is followed by the complete coding region of 1017 nucleotides, comprising 339 codons including initiation and termination codons. The 3′-non-translated region is 82 nucleotides long. It contains the presumptive polyadenylation signal AATAAG at position +51 and terminates with a poly(A) tail of 15 adenosine nucleotides.

Clone pS84b (chloroplast GAPDH) contains 836 nucleotides and starts at codon 100. The 3′-non-translated region is 132 nucleotides long, contains the polyadenylation signal AATAAA at position +108 and ends with an uninterrupted stretch of 7 adenosine nucleotides, the apparent beginning of the poly(A) tail. The two clones show 55% nucleotide homology with respect to their coding sequences starting at codon 101. They have no homology in their 3′-non-coding regions.

The protein translated from clone pS198c contains 337 residues and presently represents the longest GAPDH known (see also Fig. 3). It starts with the strongly polar peptide Ala-Asp-Lys-Lys in positions −3 to 0. The presence of these four additional aminoterminal residues could not be confirmed directly, because the aminoterminus of the native enzyme is blocked [41]. The enzyme also differs from all other eukaryotic GAPDHs in that it contains two insertions, probably Lys-53A and Glu-68A (see Fig. 3). It has a calculated M_r of 36768, which is somewhat smaller than expected from its electrophoretic mobility in sodium dodecyl sulfate gels, suggesting a M_r of approximately 39000 [34]. The calculated amino acid composition is in rough agreement with the previously determined experimental values [34], except that Cys was overestimated (7 instead of 2 residues) and Val underestimated (26 instead of 37 residues).

The amino acid sequence of mustard chloroplast GAPDH is not yet complete. The first 17 aminoterminal residues have previously been determined by automatic Edman degradation, with two, uncertainties, but probably arginines, in positions 10 and 13 [41]. The native subunit starts with a methionine in position 0 and the subsequent aminoterminal peptide is clearly homologous to the other GAPDH enzymes (see Fig. 3). Clone pS84b translates into a polypeptide comprising 70% (residues 101–333) of chloroplast GAPDH (subunit A) from mustard. Among these 233 carboxyterminal residues there are only 114 (49%) which are identical with the cytosolic GAPDH from the same species (boxed regions in Fig. 2).

DISCUSSION

Plants have parallel pathways of sugar phosphate metabolism located on either side of the chloroplast envelope. As a consequence, many of these metabolic reactions are catalyzed by pairs of plastid/cytoplasmic isoenzymes, which can be separated on the basis of their charge differences [46]. Although it is now generally accepted that the majority of the chloroplast-located isoenzymes are encoded in the nucleus [47, 48], their origin and evolution are not yet understood. There are basically three different possibilities which have been discussed previously by Bogorad [49] and elaborated further more recently by Weeden [47]. (a) The genes for a given pair of plastid/cytoplasmic isoenzymes may be the result of an ancient duplication event in the early plant eukaryotes. (b) A second hypothesis suggests that the genes of plastid isoenzymes may be of prokaryotic origin and that their present nuclear location would be the result of gene transfer from an endosymbiotic plastid ancestor to the nucleus of a primitive eukaryotic 'host'. (c) A third possibility is that plastid isoenzymes are posttranscriptional or posttranslational modifications of the cytoplasmic forms.

Although chloroplast and cytosolic GAPDHs are not isoenzymes in the strict sense of the word, because they differ in their pyridine nucleotide requirements, the two nucleus-encoded enzymes have clearly homologous sequences (Fig. 2). In the following the seven eukaryotic and three prokaryotic sequences published will be used as a data base to analyse the molecular origin of the two plant GAPDHs according to the three alternative possibilities mentioned above.

Sequence homologies between the GAPDHs of chloroplasts and thermophilic bacteria

In Fig. 3 all sequences known to the present day are aligned and compared to mustard cytosolic GAPDH, which

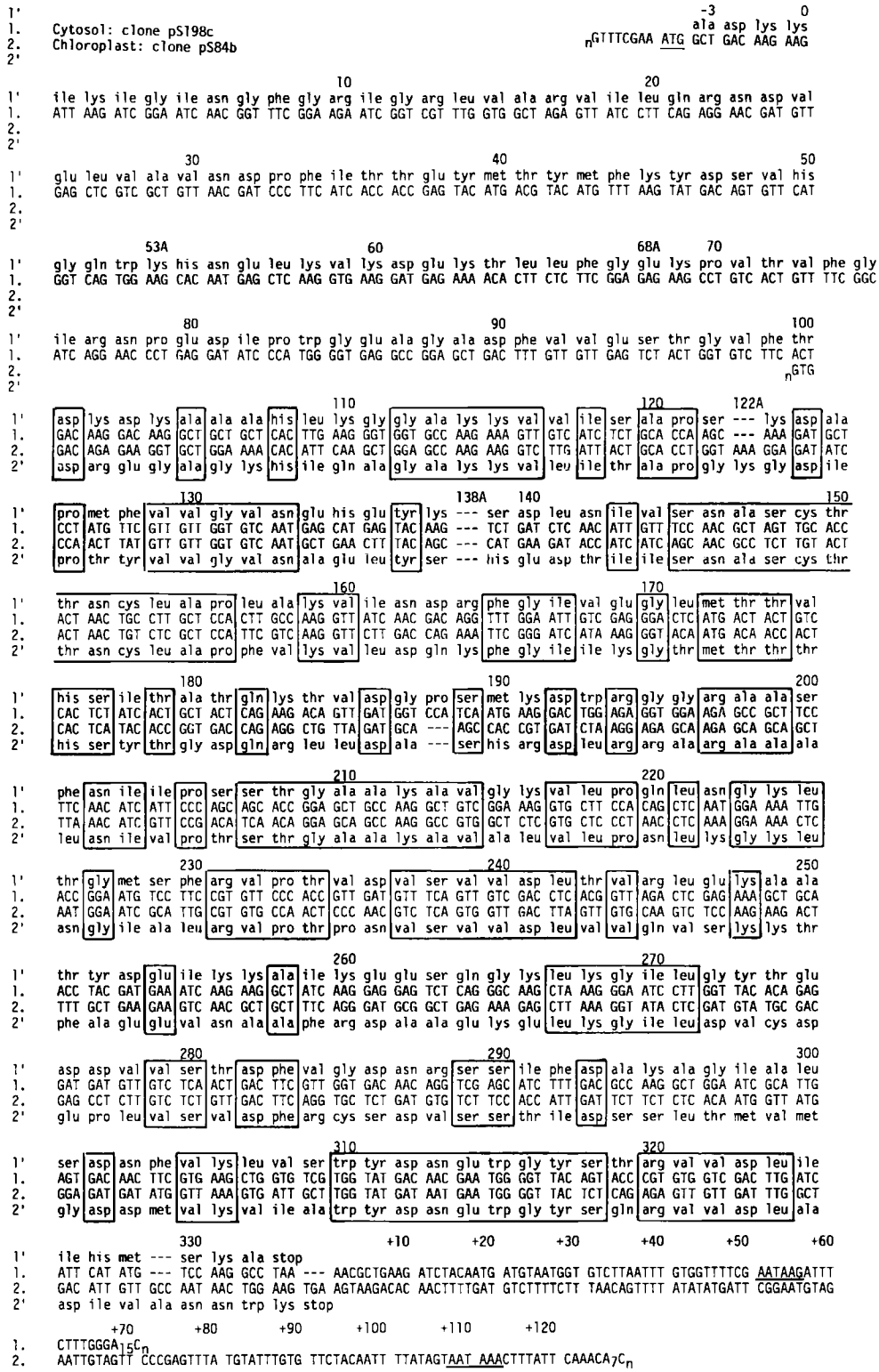

Fig. 2. *Coding sequences and deduced amino acid sequences of cDNAs for cytosolic (1, 1') and chloroplast (2, 2') GAPDH from mustard.* The numbering system used is that of Harris and Waters [45]. Regions of amino acid homologies between the two enzymes are boxed. The initiation codon ATG and the polyadenylation signals AATAAG and AATAAA are underlined. Both strands were sequenced according to Maxam and Gilbert [44]

is the only sequence written in full (line 1). For the other enzymes, including chloroplast GAPDH subunit A (sequence 12), only amino acids not identical to the reference sequence are indicated, thereby creating distinct patterns of mutations and homologies throughout taxonomic groups. Since recent authors [23, 25] have used their own numbering systems, it should be emphasized that in Fig. 3 the numerical order of Harris and Waters [45] is used to permit comparisons with most other papers throughout the literature. At the end of the amino acid sequences in Fig. 3 the homologies between glycolytic and chloroplast GAPDHs and relative to all other sequences have been listed for the polypetide 101 – 334. It can

be seen that the divergence between cytosolic GAPDH from mustard and its photosynthetic counterpart (only 49% homology) is at least as large as between the GAPDH sequences of eukaryotes and thermophilic bacteria in general (see also Table 1). It may be argued that this difference reflects the different metabolic functions of the two plant enzymes in photosynthesis and glycolysis respectively. This, however, is probably not the primary reason, because photosynthetic GAPDH subunit A shows the highest homology, that is 64%, with the enzyme from *B. stearothermophilus*, a non-photosynthetic moderate thermophile bacterium, while the two glycolytic enzymes from *Bacillus* and mustard share only 52% of their residues (50% if the full-length sequences are compared, see Table 1).

Among the 120 residues which are different for the two mustard enzymes there are 47 positions (39%) where the photosynthetic enzyme is identical with *Bacillus* GAPDH. Among the 86 residues which are different for the photosynthetic and the *Bacillus* enzymes there are only 13 positions (15%) where the photosynthetic GAPDH is identical with its cytosolic counterpart. It appears, therefore, that chloroplast and cytosolic GAPDHs from higher plants are orthologous rather than paralogous proteins. Differences in orthologous proteins reflect the phyletic divergence of different organisms, while paralogous proteins originated by gene duplication and subsequent divergent evolution in the same genome [50].

The data in Fig. 3 suggest that cytosolic and chloroplast GAPDH belong to two separate GAPDH subfamilies, the eukaryotic subfamily (sequences 1–9) and the *Thermus*/*Bacillus*/chloroplast subfamily (sequences 10–12) respectively. The surprising fact that *E. coli* GAPDH (sequence 9) belongs to the eukaryotic subfamily will be discussed below. The carboxyterminal part of GAPDH, starting at position 101, contains 124 sites at which all sequences from eukaryotes and *E. coli* are identical. In 46 out of these 124 positions one or both thermobacterial sequences are mutated. Among these 46 'bacteria-specific' mutation sites there are 30 (65%) which are also mutated in chloroplast GAPDH, 20 of which (67%) are homologous to one or both of the thermobacterial sequences. In addition to this the sequences of chloroplast and thermophilic bacteria also share an insertion Lys-122A and a deletion in the boxed S-loop region (Ser-189 for bacteria, Pro-188 for chloroplast), insertion Lys-122A being also present in *E. coli*. There are only 9 mutation sites within sequence 101–334 which are unique for chloroplast GAPDH, that is where all sequences except chloroplast GAPDH are identical (*Thermus*: 12, *Bacillus*: 1). These positions are: 175 (Val→Thr), del-188 (Pro), 191 (Lys→Arg), 204 (Ile→Val), 228 (Met→Ile), 246 (Leu→Val), 265 (Gly→Lys), 273 (Tyr→Val) and 294 (Ala→Ser).

The dichotomy of known GAPDHs into subfamilies is particularly evident in amino acid positions 178–201. These residues comprise the so-called S-loop region of the catalytic domain. The four S-loops form the core of the GAPDH tetramer, most of their residues being internal and making important interactions with the coenzyme and the other subunits [16, 17]. Eukaryotes and thermophilic bacteria have very different S-loop regions, which are highly conserved within the respective subfamilies. It is remarkable to what extent the 'bacterial character' of this functionally and structurally important peptide has been conserved in the chloroplast GAPDH (see boxed region 178–200 in Fig. 3). However, two of the changes unique for chloroplast GAPDH, del-188 (Pro) and 191 (Lys→Arg), occur in this region. Since chloroplast GAPDH is the only GAPDH which uses NADP as coenzyme and which can form a heterotetramer A_2B_2 [38, 40], these changes may be significant.

Fig. 3 also discloses some interesting features of the cytosolic GAPDH from mustard. It can be clearly seen that the NAD-binding domain (residues −3 to 148) as a whole is considerably less conserved than the catalytic domain (residues 149–331) as previously also shown for the *E. coli* enzyme [28]. In Table 1 total homologies and domain homologies for pairwise comparisons of all GAPDH sequences, are given. For example, the GAPDH sequences of mustard cytosol and yeast are in total 68% homologous, while the two domains share 54% and 80% of their residues. Similar differences are found when the enzyme from mustard cytosol, yeast or *E. coli* is compared with any of the other eukaryotic sequences (values boxed by continuous lines), while differences in domain homologies are relatively moderate in comparisons between the non-vertebrate and vertebrate sequences (values boxed by dashed lines) and in all comparisons involving the *Bacillus* enzyme. Differences are small or absent in all comparisons involving *T. aquaticus*. Within the eukaryotic subfamily the non-vertebrate enzymes from lobster and *Drosophila* represent a notable exception to this rule of differential domain conservation: the two enzymes share 78% and 75% sequence homology with respect to their NAD-binding and catalytic domains. To summarize these considerations one can conclude that divergent evolution of the two domains has been especially strong along the lineages leading to the enzymes of yeast, higher plants and *E. coli*.

The region most heavily altered in all GAPDH enzymes so far investigated is the peptide 51–70 of cytosolic GAPDH from mustard, which is part of an external loop of the NAD-binding domain [16]. The two insertions Lys-53A and Glu-68A occur in this extremely polar region, containing 9 charged residues (5 lysines) in mustard and 3–8 in the other 10 sequences. Finally, as might be expected, the longest stretch of fully conserved amino acids is found surrounding the catalytically active Cys-149. This homology block comprises residues 145–156 and is identical in all GAPDH enzymes with the exception of Cys→Ser in position 153 for *T. aquaticus*.

E. coli *and the GAPDH sequence paradox*

While the amino acid sequence of nucleus-encoded chloroplast GAPDH subunit A is related to those of thermophilic bacteria, *E. coli* has recently been shown to contain a GAPDH which is more similar to the eukaryotic enzymes [28]. The homology between *E. coli* GAPDH and the eight eukaryotic sequences, especially with respect to the catalytic domain (residues 149–331), is indeed surprising (see Fig. 3 and Table 1) and Branlant and Branlant [28] interpreted this finding in terms of a divergent evolution of the thermobacterial enzymes as a result of selection for stability and catalytic efficiency of the enzymes in organisms growing under high-temperature conditions. In view of the present data and keeping in mind that higher plants are mesophilic organisms, the data concerning *E. coli* GAPDH require re-evaluation. Sequence similarities, especially those involving hundreds of amino acids as in the present comparisons, probably reflect a common historical background rather than adaptations to similar selection pressures [51]. We, therefore, suggest the alternative possibility that the coding part of the GAPDH gene found in *E. coli* represents, in fact, the descendant of a reverse-transcribed GAPDH mRNA of an ancient and unknown eukaryotic host. There is ample evidence for a reverse-transcriptase-mediated pseudogene pro-

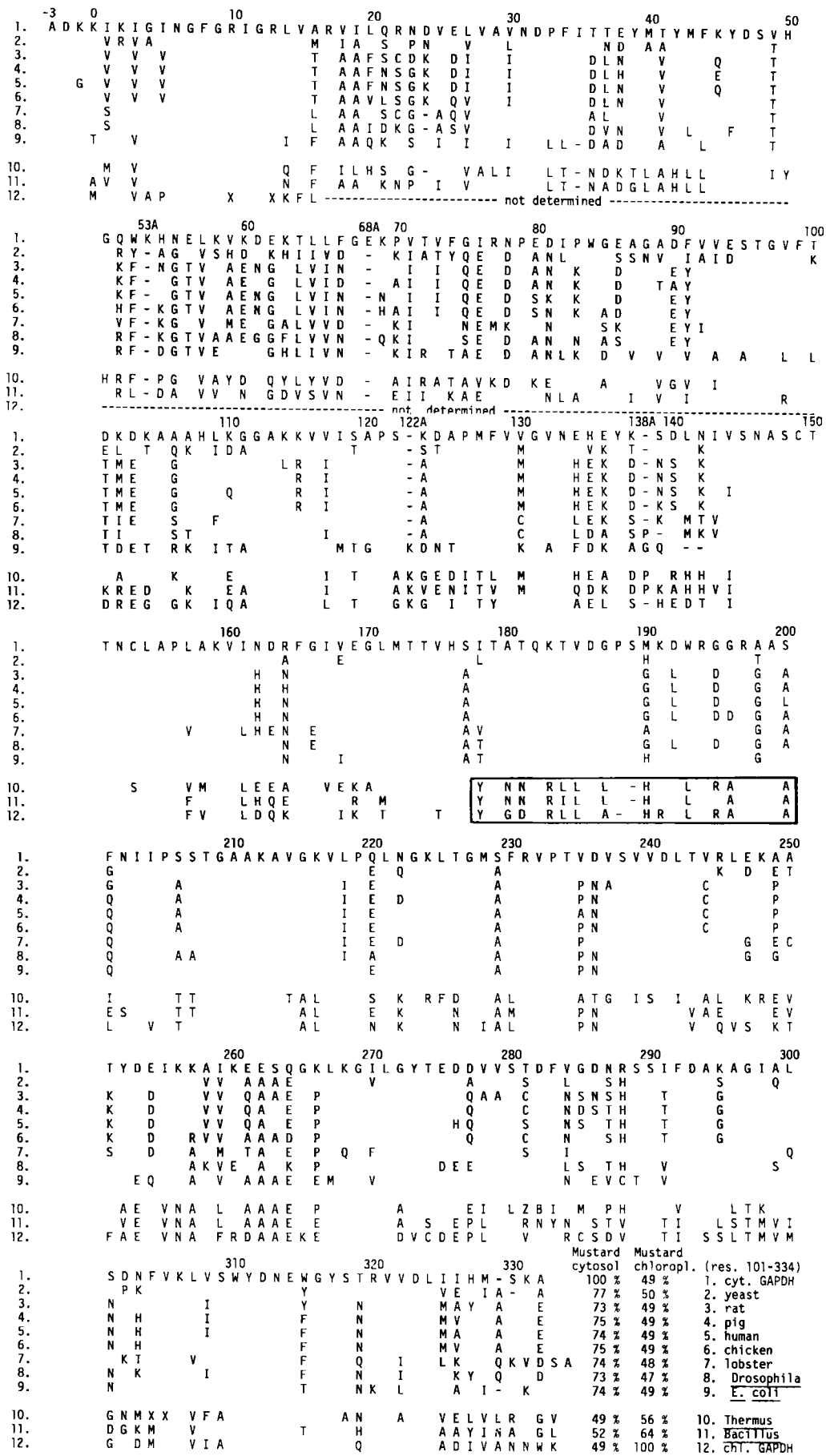
not determined
not determined
not determined
Mustard cytosol
Mustard chloropl.
(res. 101-334)
1. cyt. GAPDH
2. yeast
3. rat
4. pig
5. human
6. chicken
7. lobster
8. Drosophila
9. E. coli
10. Thermus
11. Bacillus
12. chl. GAPDH

Fig. 3

Table 1. *Identities matrix of full-length GAPDH sequences showing the number of sites as the percentage occupied by identical amino acids in pairwise comparisons (see Fig. 3)*
Values above the diagonal: total homologies. Values below the diagonal: homologies between NAD-binding domains ('numerators', residues −3 to 148) and catalytic domains ('denominators', residues 149−334) respectively. Values boxed by continuous lines: cytosolic GAPDHs from mustard, yeast and *E. coli* as compared to one another and relative to the animal sequences. Values boxed by dashed lines: GAPDHs from lobster and *Drosophila* as compared to the vertebrate sequences. All values for the *E. coli* enzyme (except the comparisons of *E. coli* with mustard, rat and *Drosophila*) were taken from [28]. Column 12 and 'numerator' values in row 12 have been omitted because the sequence of chloroplast GAPDH is not full-length. The identities matrix for the partial sequence 101−334 is shown at the bottom of Fig. 3

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. Mustard cytosol	--	68	66	67	67	67	68	66	64	45	50	--
2. Yeast	54/80	--	65	65	63	66	66	65	69	50	53	--
3. Rat	56/74	54/74	--	92	93	92	71	75	65	47	49	--
4. Pig	56/76	55/74	93/92	--	93	92	72	77	68	47	51	--
5. Human	56/76	51/74	92/93	91/95	--	91	70	75	67	46	51	--
6. Chicken	55/77	53/76	90/94	89/94	90/92	--	73	77	68	48	52	--
7. Lobster	60/75	56/71	69/72	68/75	65/74	71/75	--	76	64	48	52	--
8. Drosophila	55/76	56/73	70/80	70/82	67/83	72/82	78/75	--	66	48	53	--
9. E. coli	47/80	57/79	52/76	55/78	55/78	53/80	52/74	54/77	--	49	58	--
10. Thermus	42/48	45/55	46/48	48/46	46/46	48/48	47/48	49/48	49/49	--	60	--
11. Bacillus	45/55	49/57	49/56	46/56	45/55	46/57	48/55	51/54	55/60	57/62	--	--
12. Mustard chloroplast	--/48	--/50	--/49	--/49	--/48	--/48	--/47	--/49	--/48	--/55	--/67	--

Fig. 3. *Amino acid sequence alignment of 12 GAPDH proteins from 11 different organisms, as specified at the bottom of the figure.* Amino acid sequences are aligned to maximize homology according to Harris and Waters [45]. All sequences are compared to cytosolic GAPDH of mustard (sequence 1), which is the only sequence written in full. For the other enzymes, including chloroplast GAPDH from mustard (sequence 12), only amino acids not identical to the reference sequence are indicated. The first and last residues of each sequence are indicated irrespective of homology. Boxed sequence 178−200 designates the S-loop region of bacterial GAPDHs. The two mustard enzymes (sequences 1 and 12) have been compared with one another and relative to all other sequences with respect to residues 101−334 and the homology values are tabulated in two columns after the corresponding C termini (see also Table 1). Sources of sequence information: yeast [19], rat and *Drosophila* [25], pig and lobster [45], human [23], chicken [21], *E. coli* [28], *Thermus* and *Bacillus* [18]

duction in mammals [52, 53]. Therefore, if horizontal gene transfer plays a role in evolution, as there are reasons to believe (see above and [54, 55]), it cannot be seen why a transfer from eukaryotes to bacteria should be a 'forbidden' route of information flow. A gene transfer from eukaryote to prokaryote has previously been suggested to explain the occurrence of the copper + zinc-containing, eukaryotic, superoxide dismutase in *Photobacter leiognathi*, the bioluminescent bacterial symbiont of the ponyfish [56]. Interestingly superoxide dismutase is also the subject of a possible gene transfer in the reverse prokaryote-to-eukaryote direction since the iron-containing, prokaryotic enzyme was found in three higher plant families out of 43 investigated [57].

We do not want to carry this discussion any further at the present time and would rather wait until more GAPDH sequence information from bacteria, especially cyanobacteria, the free-living descendants of the presumptive 'ancestors' of chloroplasts (see [7–10]), is available and until the structure of the GAPDH gene from chloroplasts has been elucidated. For the time being we believe that our data are compatible with the endosymbiotic theory of chloroplast evolution and in particular with the gene-transfer concept, mentioned above as hypothesis (b). Since the glycolytic GAPDH of *B. stearothermophilus* is more closely related to the photosynthetic than to the glycolytic enzyme of higher plants it may finally be hypothesized that the photosynthetic enzyme (subunit A) diverged from the glycolytic enzyme of eubacteria long after the separation of the eubacterial and eukaryotic lineages. Since the Calvin cycle originated more than two billion years ago [58] this separation would have occurred very early in the history of life, in agreement with modern concepts of cell evolution [59, 60].

We thank H. Sommer (Max-Planck-Institut für Züchtungsforschung, Cologne) for teaching us how to clone and sequence; H. H. Arnold (University of Hamburg) for the gift of the GAPDH clone from chicken; J. Collins, H. J. Hauser, H. Blöcker and colleagues (*Gesellschaft für Biotechnologische Forschung*, Braunschweig) for restriction enzymes, radioisotopes and for the use of the computer; Yvette Guyénon for her help in the preparation of the manuscript. The financial support of the *Deutsche Forschungsgemeinschaft* (Ce 1/11-3) is gratefully acknowledged.

REFERENCES

1. Bonen, L. & Doolittle, W. F. (1976) *Nature (Lond.) 261*, 669–673.
2. Schwarz, Zs. & Kössel, H. (1980) *Nature (Lond.) 283*, 739–742.
3. Schwartz, R. M. & Dayhoff, M. O. (1981) *Ann. N.Y. Acad. Sci. 361*, 260–272.
4. Edwards, K. & Kössel, H. (1981) *Nucleic Acids Res. 9*, 2853–2869.
5. Bohnert, H. J., Crouse, E. J. & Schmitt, J. M. (1982) in *Encyclopedia of plant physiology*, vol. 14B (Parthier, B. & Boulter, D. eds) pp. 475–530, Springer, Berlin, Heidelberg, New York.
6. Whitfeld, P. R. & Bottomley, W. (1983) *Annu. Rev. Plant Physiol. 34*, 279–310.
7. Taylor, F. J. R. (1979) *Proc. R. Soc., Lond. B Biol. Sci. 204*, 267–286.
8. Margulis, L. (1981) *Symbiosis in cell evolution*, Freeman, San Francisco.
9. Fredrick, J. F., ed. (1981) *Ann. New York Acad. Sci. 361.*
10. Gray, M. W. & Doolittle, W. F. (1982) *Microbiol. Rev. 46*, 1–42.
11. Roscoe, T. J. & Ellis, R. J. (1982) in *Methods in chloroplast molecular biology* (Edelman, M., Hallick, R. B. & Chua, N.-H., eds) pp. 1015–1028, Elsevier/North Holland, Amsterdam.
12. Bottomley, W. & Bohnert, H. J. (1982) in *Encyclopedia of plant physiology*, vol. 14B (Parthier, B. & Boulter, D., eds) pp. 531–596, Springer, Berlin, Heidelberg, New York.
13. Davidson, B. E., Sajgo, M., Noller, H. F. & Harris, J. I. (1967) *Nature (Lond.) 216*, 1181–1185.
14. Harris, J. I. & Perham, R. N. (1968) *Nature (Lond.) 219*, 1025–1028.
15. Jones, G. M. T. & Harris, J. I. (1972) *FEBS Lett. 22*, 185–189.
16. Biesecker, G., Harris, J. I., Thierry, J. C., Walker, J. E. & Wonacott, A. J. (1977) *Nature (Lond.) 266*, 328–333.
17. Walker, J. E., Carne, A. F., Runswick, M. J., Bridgen, J. & Harris, J. I. (1980) *Eur. J. Biochem. 108*, 549–565.
18. Hocking, J. D. & Harris, J. I. (1980) *Eur. J. Biochem. 108*, 567–579.
19. Holland, J. & Holland, M. (1979) *J. Biol. Chem. 254*, 9839–9845.
20. Domdey, H., Wiebauer, K., Klapthor, H. & Arnold, H. H. (1983) *Eur. J. Biochem. 131*, 129–135.
21. Dugaiczyck, A., Haron, J. A., Stone, E. M., Dennison, O. E., Rothblum, K. N. & Schwartz, R. J. (1983) *Biochemistry 22*, 1605–1613.
22. Panabières, F., Piechaczyk, M., Rainer, B., Dani, Ch., Fort, Ph., Riaad, S., Marty, L., Imbach, J. L., Jeanteur, Ph. & Blanchard, J. M. (1984) *Biochem. Biophys. Res. Commun. 3*, 767–773.
23. Hanauer, A. & Mandel, J. L. (1984) *EMBO J. 3*, 2627–2633.
24. Benham, F. J., Hodgkinson, D. & Davies, K. E. (1984) *EMBO J. 3*, 2635–2640.
25. Tso, J. Y., Sun, X.-H., Kao, T., Reece, K. S. & Wu, R. (1985) *Nucleic Acids Res. 7*, 2485–2502.
26. Fort, Ph., Marty, L., Piechaczyk, M., Sabrouty, S. El, Dani, Ch., Jeanteur, Ph. & Blanchard, J. M. (1985) *Nucleic Acids Res. 13*, 1431–1442.
27. Tso, J. Y., Sun, X.-H. & Wu, R. (1985) *J. Biol. Chem. 260*, 8220–8228.
28. Branlant, G. & Branlant, Ch. (1985) *Eur. J. Biochem. 150*, 61–66.
29. Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossmann, M. G. (1973) *Proc. Natl Acad. Sci. USA 70*, 3052–3054.
30. Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossmann, M. G. (1974) *J. Mol. Biol. 90*, 25–40.
31. Moras, D., Olsen, K. W., Sabesan, M. N., Buehner, M., Ford, G. C. & Rossmann, M. G. (1985) *J. Biol. Chem. 250*, 9137–9162.
32. Rossman, M. G., Liljas, A., Brändén, C.-J. & Banaszak, J. (1975) in *The enzymes* (Boyer, P. D., ed.) 3rd edn, vol. 11, pp. 61–102, Academic Press, New York.
33. McGowan, R. E. & Gibbs, M. (1974) *Plant Physiol. 54*, 312–319.
34. Cerff, R. & Chambers, S. E. (1979) *J. Biol. Chem. 254*, 6094–6098.
35. Pupillo, P. & Faggiani, R. (1979) *Arch. Biochem. Biophys. 194*, 581–592.
36. Pawlitzki, K. & Latzko, E. (1974) *FEBS Lett. 42*, 285–288.
37. Ferri, G., Comerio, G., Jadarola, P., Zapponi, M. C. & Speranza, M. L. (1978) *Biochim. Biophys. Acta 522*, 19–31.
38. Cerff, R. (1979) *Eur. J. Biochem. 94*, 243–247.
39. Cerff, R. & Kloppstech, K. (1982) *Proc. Natl Acad. Sci. USA 79*, 7624–7628.
40. Cerff, R., Hundrieser, J. & Friedrich, R. (1986) *Mol. Gen. Genet. 204*, 44–51.
41. Cerff, R. & Witt, J. (1983) in *Endocytobiology II* (Schenk, H. E. A. & Schwemmler, W., eds) pp. 187–193, De Gruyter, Berlin.
42. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular cloning*, Cold Spring Harbor Laboratory.
43. Hanahan, D. (1983) *J. Mol. Biol.* 557–580.
44. Maxam, A. M. & Gilbert, W (1980) *Methods Enzymol. 65*, 499–560.
45. Harris, J. I. & Waters, M. (1976) in *The enzymes* (Bayer, P. D., ed.) 3rd edn, vol. 13, pp. 1–48, Academic Press, New York.
46. Schnarrenberger, C., Herbert, M. & Krüger, J. (1983) in *Isozymes. Current topics in biological and medical research* (Rattazzi, M.

C., Scandalios, J. G. & Whitt, L. S., eds) vol. 8, Liss, New York.
47. Weeden, N. F. (1981) *J. Mol. Evolut. 17*, 133–139.
48. Gottlieb, L. D. (1982) *Science (Wash. DC) 216*, 373–380.
49. Bogorad, L. (1975) *Science (Wash. DC) 188*, 891–898.
50. Fitch, W. M. (1976) in *Molecular evolution* (Ayala, F. J., ed.) pp. 160–178, Sinauer Associates, Sunderland.
51. Schulz, G. E. & Schirmer, R. H. (1979) in *Principles of protein structure*. Springer advanced texts in chemistry (Cantor, C. R., ed.) pp. 167–205, Springer Verlag, New York, Berlin.
52. Sharp, P. A. (1983) *Nature (Lond.) 301*, 471–472.
53. Lewin, R. (1983) *Science (Wash. DC) 219*, 1052–1054.
54. Lewin, R. (1982) *Science (Wash. DC) 217*, 42–43.
55. Syvanen, M. (1986) *Trends Genet. 2*, 63–66.
56. Martin, J. P. & Fridovich, J. (1981) *J. Biol. Chem. 256*, 6080–6089.
57. Bridges, S. M. & Salin, M. L. (1981) *Plant Physiol. 68*, 275–278.
58. Dickerson, R. E. (1980) *Sci. Am. 242*, 98–110.
59. Woese, C. R. & Fox, G. E. (1977) *Proc. Natl Acad. Sci. USA 74*, 5088–5090.
60. Doolittle, W. F. (1980) *Trends Biochem. Sci. 5*, 146–149.