their scientific use. The cell-free expression of genetic information avoids the many limitations imposed by the cell, so that the preparative biosynthesis of a number of nonstructured polypeptides, as well as unstable and cytotoxic proteins, becomes possible. In particular, this method may provide researchers with the intermediates of protein folding and modification for their biochemical and physical studies. The possibilities of protein engineering can be also significantly extended. We also anticipate new developments in biotechnology based on preparative cell-free translation systems of continuous action.

REFERENCES

1. B. E. Roberts and B. M. Paterson, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 2330 (1973).
2. T. Suryanarayana and A. R. Subramanian, *Biochemistry* **22**, 2715 (1983).
3. D. A. Melton *et al.*, *Nucleic Acids Res.* **12**, 7035 (1984).
4. P. A. Krieg and D. A. Melton, *ibid.*, p. 7057.
5. H. Schagger and G. von Jagow, *Anal. Biochem.* **166**, 368 (1987).
6. J. Chroboczek, *Eur. J. Biochem.* **149**, 565 (1985).

7. We thank J.-P. Ebel and B. Ehresmann at the Institut de Biologie moleculaire et cellulaire, Strasbourg, and M. Grunberg-Manago and her colleagues at the Institut de Biologie physico-chimique, Paris, for fruitful discussions and suggestions. We also thank A. N. Fedorov, S. N. Proshkina, and V. Ya. Pankratova at the Institute of Protein Research for technical assistance and K. G. Skryabin and his colleagues at the Institute of Molecular Biology, Moscow, for the plasmids with the calcitonin gene.

13 June 1988; accepted 17 October 1988

Intron Existence Predated the Divergence of Eukaryotes and Prokaryotes

MING-CHE SHIH,* PETER HEINRICH, HOWARD M. GOODMAN

Nucleotide sequences for the nuclear genes encoding chloroplast (GapA and GapB) and cytosolic (GapC) glyceraldehyde-3-phosphate dehydrogenases (GAPDHs) from *Arabidopsis thaliana* were determined. Comparison of nucleotide sequences indicates that the divergence of chloroplast and cytosolic GAPDH genes preceded the divergence of prokaryotes and eukaryotes. In addition, some intron-exon junctions are conserved among GapB, GapC, and chicken GAPDH genes. These results provide evidence at the molecular level to support the idea that introns existed before the divergence of prokaryotes and eukaryotes.

THREE ALTERNATIVE MODELS CAN BE proposed concerning the time of intron appearance during evolution. First, introns existed in the progenitor genes but were lost from prokaryotes (*1*), and differential intron loss may occur among members of the same gene family of eukaryotes (*2*). Second, introns existed in the progenitor genes, but intron insertion or deletion may occur later during evolution. Third, prokaryotic genes resemble the progenitor genes, and the addition of introns occurred at or after the emergence of eukaryotes (*3*). Current available evidence is more consistent with the first two models (*2*). However, there is no direct evidence as to whether introns existed before the divergence of prokaryotes and eukaryotes, information crucial to distinguishing these models. Current data can only trace the existence of introns up to the divergence of animals and plants (*4–6*).

We have shown previously that the nuclear genes (GapA and GapB) encoding chloroplast glyceraldehyde-3-phosphate dehydrogenase (GAPDH) of higher plants are the direct descendants of the symbiont genes that gave rise to the chloroplast and were subsequently transferred from the chloroplast to the nuclear genome during evolution (*7*). In addition, our data showed that the time of divergence between GapA/B and cytosolic GAPDH genes (~1700 million years) is much earlier than the divergence between plants and animals (~1000 million years). Therefore, comparison of the exon-intron junctions between GapA/B and cytosolic GAPDH genes should provide information on the age of the intron. If introns existed in the progenitor GAPDH gene, conservation between some exon-intron junctions of GapA/B and cytosolic GAPDH genes would be expected.

Genomic DNA blot analyses indicate that there is one copy each of the GapA, GapB, and GapC (cytosolic GAPDH) genes in *Arabidopsis thaliana* (*8*). The complete nucleotide sequences for these three genes and the cDNAs encoded by them have been determined (*8*). When compared to the corresponding cDNA sequences GapA has three, GapB has seven and GapC has eight introns. In contrast, there are 11 introns in the chicken GAPDH gene (*9*). The introns for all three Gap genes are generally very short, consistent with observations on other *Arabidopsis* genes that have been sequenced (*10, 11*).

On the basis of the sequences of the coding regions of the three Gap genes and the information published previously (*7*), an evolutionary tree for the GAPDH genes (Fig. 1), with the number of nonsynonymous substitutions (*12, 13*) used as a measure of the evolutionary distance, was constructed according to the Neighbor Joining method (*14*). This tree topology is consistent with our previous tree that took into account the evolutionary rates of GAPDH and with the biochemical evidence obtained from a comparison of the amino acid sequences of the S loop regions among different GAPDHs (*7*). GAPDHs can be divided into two groups: thermophilic, which includes two bacteria and chloroplast GAPDHs; and mesophilic, which includes *Escherichia coli* and all cytosolic GAPDHs. There is 100% similarity of the S loop sequences within the same group, while there is less than 50% similarity between the two groups. The tree topology indicates that the divergence of GapA/B and cytosolic GAPDH genes occurred before the divergence of prokaryotes and eukaryotes.

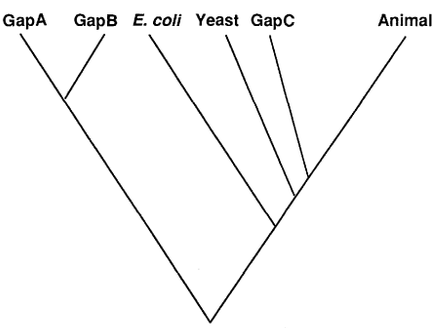

Fig. 1. Evolutionary tree of GAPDH genes. The number of nonsynonymous substitutions between each pairwise combination of GAPDH genes was calculated according to the method described by Li *et al.* (*13*) and used as the measure of the evolutionary distance in constructing the evolution tree. The tree presented here is constructed by the Neighbor Joining method of Saitou and Nei (*14*). Two additional methods, described by Li (*21*) and Dickerson and Geis (*22*), were used in constructing trees. All three trees give identical topology. The *E. coli* data are from Branlant and Branlant (*23*) and the *Arabidopsis* data are from this manuscript. All other data are from Shih *et al.* (*7*).

Department of Genetics, Harvard Medical School, and Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114.

*Present address: Departments of Biology and Botany, University of Iowa, Iowa City, IA 52242.

To compare the positions of the intron-exon junctions in GAPDH genes, the coding regions of *Arabidopsis* GapA, GapB, and GapC were aligned with chicken GAPDH (*9*). The intron-exon junctions that are conserved at least between two genes are illustrated in Fig. 2. Two introns are conserved among GapB, GapC, and chicken GAPDH genes: B2 (displaced by one amino acid, three nucleotides), C3, and K3 (Fig. 2b); and B4 and K6 (Fig. 2d). A third position is probably also conserved between B6 and C7 (Fig. 2f), if one assumes intron displacement of two amino acids (seven nucleotides) is acceptable. These observations can be most easily interpreted by proposing that these introns existed before the divergence of GapA/B and cytosolic GAPDH genes, that is, before the divergence of prokaryotes and eukaryotes, but were subsequently lost from some of the current GAPDH gene descendants. If true, this would argue against the model in which it was proposed that prokaryotic genes resemble the progenitor genes and that insertion of introns occurred after or at the time eukaryotes emerged. More importantly, these results provide direct evidence at the molecular level to indicate that introns did exist in the progenitor GAPDH gene. However, while being consistent with the first two models and differential intron loss, our data cannot distinguish whether intron insertions also occurred, since the introns conserved between GapC and chicken GAPDH (Fig. 2, a and c) could have been either inserted in GapC and chicken GAPDH gene or deleted from GapA/B, and the introns conserved between GapA and GapB (Fig. 2e) could have been either inserted in GapA/B or deleted from GapC and chicken GAPDH gene after the divergence of chloroplast and cytosolic GAPDH genes.

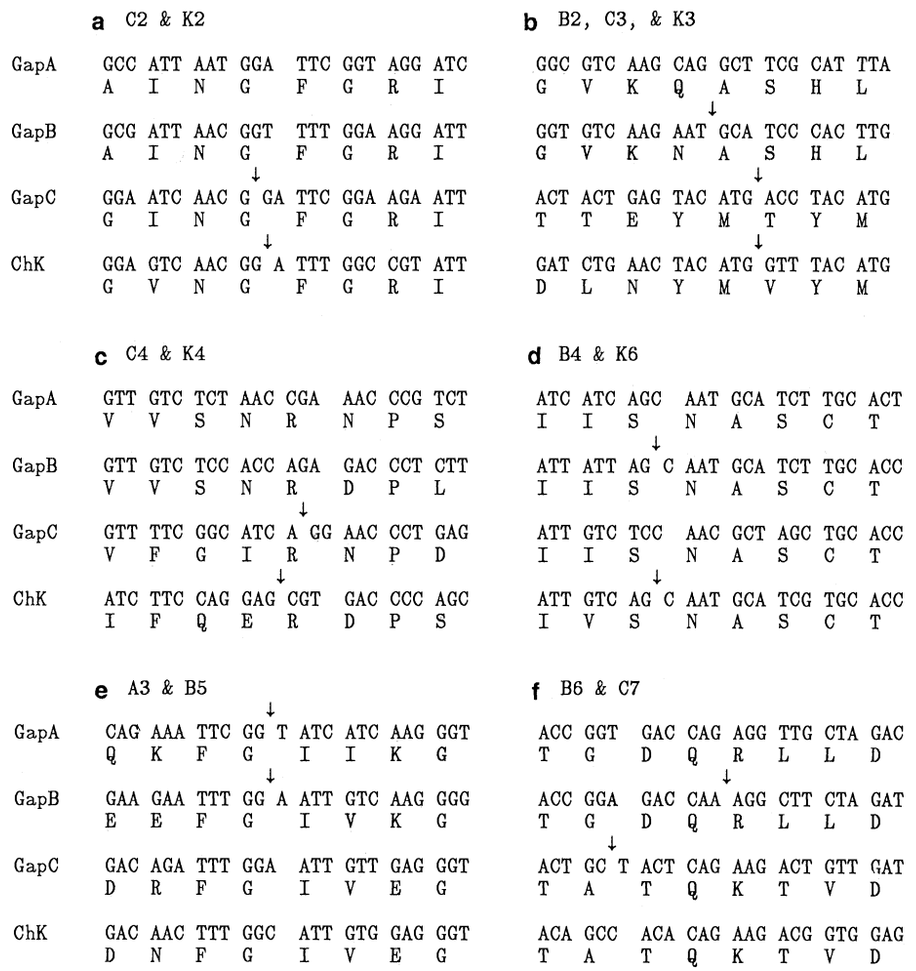

Fig. 2. Conservation of intron-exon junctions among GAPDH genes. The nucleotide sequences of the intron-exon junctions that are conserved at least between two genes from GapA, GapB, GapC, and chicken GAPDH (*9*) are shown. The arrows indicate the positions of the junctions. The nomenclature used to identify the introns is as follows: GapA, A1 to A3; GapB, B1 to B7; GapC, C1 to C8; and chicken, K1 to K11. Four introns—A1 and B1 (located in the 5′ untranslated and the transit peptide regions, respectively); B7 (located in the COOH-terminal coding region of the B peptide); and K1 (located in the 5′ untranslated leader region)—are not included in the comparisons. (**a**) Comparison of C2 and K2; (**b**), B2, C3, and K3; (**c**) C4 and K4; (**d**) B4 and K6; (**e**) A3 and B5; and (**f**) B6 and C7. The most similar sequences of the other genes without intron-exon junctions (no arrows) are also displayed.

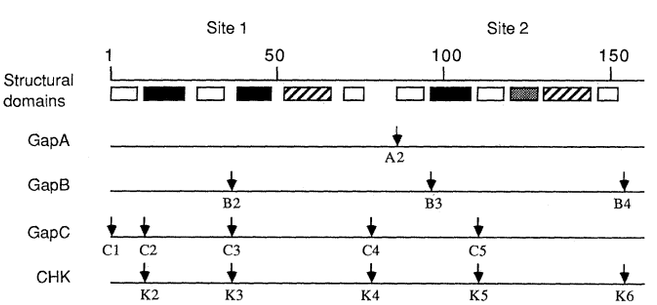

Fig. 3. Relation between structural domains and intron positions of the GAPDH NAD^+ binding domain. The secondary structure of the NAD^+ binding domain of GAPDH is drawn based on the x-ray crystallographic data of lobster and *Bacillus* (*15–18*). The α strand, incomplete α strand, β sheet, and antiparallel sheet are represented by solid, stippled, white, and hatched bars, respectively. The intron positions are indicated by arrows, and the names of the introns that appear in certain positions are underneath that arrow.

X-ray crystallographic studies have shown that GAPDH enzymes possess several structurally independent domains that are highly conserved across great evolutionary distance (*15–19*). Therefore, a comparison of the positions of intron-exon and structural domain junctions should provide information as to when introns appeared. The GAPDH subunit consists of two domains: the first domain, residues 1 to 152, is involved in the binding of oxidized nicotinamide-adenine dinucleotide (NAD^+) or oxidized nicotinamide-adenine dinucleotide phosphate ($NADP^+$) and the second domain, residues 153 to 337, in catalysis. It is also believed that the NAD^+ binding domain, which has two nucleotide binding sites, evolved from an early duplication event (*15, 19*).

Two striking observations emerge from a comparison of the structural domains and intron positions within the NAD^+ binding domain (Fig. 3). First, three pairs of introns appear in similar positions of NAD^+ binding sites 1 and 2: C1 and A2, in front of the first β strand; C2/K2 and B3, between the first β strand and first α helix; and C4/K4 and B4/K6, behind the third β strand. This observation provides further support for the ancient origin of the GAPDH introns. Since

prokaryotic and eukaryotic GAPDHs all have the same secondary structure and since the evolutionary distance (amino acid sequence similarity) between site 1 and site 2 of the same GAPDH is much greater than the distance between site 1 of prokaryotic and eukaryotic GAPDHs, this would indicate that the duplication that gave rise to the two NAD^+-binding sites 1 and 2 occurred before the divergence of prokaryotes and eukaryotes. Hence, the occurrence of introns in similar positions in the two sites implies that these introns existed in the progenitor GAPDH gene prior to the divergence of prokaryotes and eukaryotes, but some of these introns were subsequently lost from certain GAPDH species during evolution. The fact that one site has introns in the plant chloroplast (GapB, B4) and animal cytosolic (chicken, K6) genes, while its duplicated site has introns in plant (GapC, C4) and animal (chicken, K4) cytosolic genes gives strong support to this idea. We predict that, as more sequencing data are available, some GAPDH genes should have introns located between residues 20 and 30 in site 1 corresponding to C5/K5 in site 2 and between residues 120 and 125 in site 2 corresponding to B2/C3/K3 in site 1. Second, the introns always appear between, but not within, the structural domains of the NAD^+ binding region. This observation is consistent with the idea, as proposed by Blake (*20*), that introns might play important roles in bringing together small sequence units that encode potentially stabilizing secondary structures in the evolving protein.

The relationships between intron positions and the catalytic domain, which consists basically of nine-stranded antiparallel sheets and a long helical tail (*16*), is more difficult to interpret. However, based on the same lines of reasoning, Stone *et al.* (*19*) have argued that the chicken introns, K7 to K11, should also have existed in the progenitor GAPDH gene. Therefore, among a total of 18 intron positions in four GAPDH genes, 15 of them should have existed in the progenitor GAPDH gene. Whether the remaining three introns (B6, C6, and C8) are the result of deletional or insertional events remains to be elucidated.

In summary, we have provided evidence that some GAPDH introns existed before the divergence of prokaryotes and eukaryotes, consistent with the idea that introns played a role in the assembly of the progenitor GAPDH gene. In addition, a comparison of the intron positions with the structural domains of GAPDH indicates that the majority, if not all, of the introns that exist in the four GAPDH genes whose sequences are known should have existed in the progenitor GAPDH gene.

REFERENCES AND NOTES

1. W. F. Doolittle, *Nature* **272**, 581 (1978).
2. W. Gilbert, M. Marchionni, G. McKnight, *Cell* **46**, 151 (1986).
3. L. E. Orgel and F. H. Crick, *Nature* **284**, 604 (1980).
4. D. M. Shah, R. C. Hightower, R. B. Meagher, *J. Mol. Appl. Genet.* **2**, 111 (1983).
5. M. Marchionni and W. Gilbert, *Cell* **46**, 133 (1986).
6. G. L. McKnight, P. J. O'Hara, M. L. Parker, *ibid.*, p. 143.
7. M-C. Shih, G. Lazar, H. M. Goodman, *ibid.* **47**, 73 (1986).
8. M-C. Shih, P. Heinrich, H. M. Goodman, unpublished data.
9. E. M. Stone, K. N. Rothblum, M. C. Alevy, T. M. Kuo, R. J. Schwartz, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1628 (1985).
10. L. S. Leutwiler, E. M. Meyerowitz, E. M. Tobin, *Nucleic Acids Res.* **10**, 4051 (1986).
11. C. Chang and E. M. Meyerowitz, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1408 (1986).
12. R. Perler *et al.*, *Cell* **20**, 555 (1980).
13. W.-H. Li, C.-I. Wu, C. C. Luo, *Mol. Biol. Evol.* **2** (no. 2), 150 (1985).
14. N. Saitou and M. Nei, *ibid.*, **4** (no. 4), 406 (1986).
15. M. Buehner, G. C. Ford, D. Moras, K. W. Olsen, M. G. Rossmann, *J. Mol. Biol.* **90**, 25 (1974).
16. J. I. Harris and M. Waters, *The Enzymes*, P. Boyer, Ed. (Academic Press, New York, 1976), vol. 13, pp. 1–49.
17. D. Moras *et al.*, *J. Biol. Chem.* **250**, 9137 (1975).
18. G. Biesecker, J. I. Harris, J. C. Thierry, J. E. Walker, A. J. Wonacott, *Nature* **266**, 328 (1977).
19. E. M. Stone, K. N. Rothblum, R. J. Schwartz, *ibid.* **313**, 498 (1985).
20. C. Blake, *ibid.* **306**, 535 (1983).
21. W.-H. Li, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1085 (1981).
22. R. E. Dickerson and I. Geis, *Hemoglobin: Structure, Function, Evolution and Pathology* (Benjamin/Cummings, Menlo Park, CA, 1982), pp. 65–116.
23. G. Branlant and C. Branlant, *Eur. J. Biochem.* **150**, 61 (1985).

24. We thank K. Peterman for the cDNA library, E. Meyerowitz and B. Hauge for genomic libraries, and C.-I. Wu and M. Nei for the computer programs used in the evolutionary calculations. Supported by a grant from Hoechst AG.

25 April 1988; accepted 20 September 1988

HIV-Infected Cells Are Killed by rCD4-Ricin A Chain

MARK A. TILL, VICTOR GHETIE, TIMOTHY GREGORY, ERIC J. PATZER, JAMES P. PORTER, JONATHAN W. UHR, DANIEL J. CAPON, ELLEN S. VITETTA*

The gp120 envelope glycoprotein of the human immunodeficiency virus (HIV), which is expressed on the surface of many HIV-infected cells, binds to the cell surface molecule CD4. Soluble derivatives of recombinant CD4 (rCD4) that bind gp120 with high affinity are attractive vehicles for targeting a cytotoxic reagent to HIV-infected cells. Soluble rCD4 was conjugated to the active subunit of the toxin ricin. This conjugate killed HIV-infected H9 cells but was 1/1000 as toxic to uninfected H9 cells (which do not express gp120) and was not toxic to Daudi cells (which express major histocompatibility class II antigens, the putative natural ligand for cell surface CD4). Specific killing of infected cells can be blocked by rgp120, rCD4, or a monoclonal antibody to the gp120 binding site on CD4.

MOST INDIVIDUALS INFECTED with the human immunodeficiency virus (HIV) develop acquired immunodeficiency syndrome (AIDS) (*1*), which is characterized by the progressive depletion of T cells expressing CD4, the cellular receptor for HIV (*2*). A potential approach for preventing or delaying the onset of AIDS is to eliminate cells producing viral proteins early in the course of the disease. This may prevent the spread of infection and the release of viral proteins that may participate in the pathogenesis of the disease (*3*). HIV-infected cells could be eliminated with a toxic agent coupled to a targeting molecule that would bind only to cells expressing HIV-encoded proteins. An attractive targeting entity is recombinant soluble CD4 (rCD4) (*4, 5*) that binds to gp120, the envelope glycoprotein of HIV, with an affinity comparable to that of cell surface CD4 (*4*). Although gp120, which is expressed on the surface of many HIV-infected cells, shows extensive variability among different strains of HIV, its CD4 binding site is highly conserved (*6*).

Conjugates of toxins and cell-reactive ligands can specifically delete cells in vitro and in vivo (*7*). We and others have used the A chain of the plant toxin, ricin, conjugated to cell-reactive antibodies. Such conjugates kill cells after endocytosis of the conjugate-antigen complex and translocation of the A chain into the cytosol where it inhibits protein synthesis (*7*). Below we describe the coupling of soluble rCD4 to deglycosylated

M. A. Till, V. Ghetie, J. W. Uhr, E. S. Vitetta, Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, TX 75235.
T. Gregory, Department of Process Sciences, Genentech, Inc., South San Francisco, CA 94080.
E. J. Patzer and J. P. Porter, Department of Medicinal and Analytical Chemistry, Genentech, Inc., South San Francisco, CA 94080.
D. J. Capon, Department of Molecular Biology, Genentech, Inc., South San Francisco, CA 94080.

*To whom correspondence should be addressed.