J. Cell. Biochem. **18**, 37 (1982), modified]} and sequenced on an Applied Biosystems 477A Sequencer. We obtained a sequence from 27 peptides from both 34-kD bands of the purified BTE, and 24 matched the derived cDNA sequence. No preferential matches with peptides from either member of the protein doublet were apparent. We obtained the eight NH_2-terminal residues of the 34-kD polypeptides.

6. We followed in principle the approach of C. C. Lee *et al.* [*Science* **239**, 1288 (1988)]. Degenerate oligonucleotides were synthesized after reverse-translation of the respective peptide sequences. The sequence of the antisense oligonucleotide 696-2a was: GCCTCGAGCC(IC) CC(TC)TGIAT(AG)TA(AG)TC(IC)GC (I = desoxyinosine, degenerate residues in brackets. A 5′ Xho I site was added). The sequence of the sense oligonucleotide 701-2 was: CTGGATCCGA(CT) AT(ACT)(CT)T(IC)GC(IC)GT(IC)ATGAA (A 5′ Bam HI site was added). RNA from bay cotyledons at the stage of rapid accumulation of bay TE was extracted according to T. H. Turpen and O. M. Griffith [*BioTechniques* **4**, 11 (1986)]. Polyadenylate [poly(A)] RNA, obtained by oligo(dT) cellulose chromatography, was reverse-transcribed, and the single-strand cDNA (1 to 10 ng per 100 μl of reaction) was used for PCR. A Perkin-Elmer thermal cycler was used with primers at 2 μM, 30 cycles of 1 min at 94°C, 1 s at 65°C, 2 min-ramp to 50°C, 1 s at 50°C, and 2 min at 72°C.
7. A cDNA library of the plasmid vector pCGN1703 [D. C. Alexander, *Methods Enzymol.* **154**, 41 (1987)] was made with the use of poly(A) RNA from developing bay cotyledons. The PCR generated a 0.8-kb BTE cDNA fragment (*6*) that was used as a probe. The clone with the longest insert was named pCGN3822. DNA sequencing was according to M. Hattori and Y. Sakaki [*Anal. Biochem.* **152**, 232 (1986)].
8. H. A. Lütcke *et al.*, *EMBO J.* **6**, 44 (1987).
9. H. M. Davies and T. A. Voelker, unpublished results.
10. K. Keegstra, L. J. Olsen, S. M. Theg, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 471 (1989).
11. S. Smeekens, P. Weisbeek, C. Robinson, *Trends Biochem. Sci.* **15**, 73 (1990).
12. J. B. Ohlrogge, D. N. Kuhn, P. K. Stumpf, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1194 (1979).
13. D. S. Knutzon, personal communication.
14. For plant expression, we engineered a Bam HI site 6 bp upstream of the proposed translational start. The BTE cDNA from this engineered Bam HI site to a Pst I site at bp 1361 was inserted into pCGN3223. This plasmid contains 1750 bp of a napin promoter fragment, followed by a napin 3′ untranslated sequence isolated from *Brassica rapa* [J. C. Kridl *et al.*, *Seed Sci. Res.* **1**, 209 (1991)]. The chimeric gene was inserted into a binary vector, pCGN1578 [K. E. McBride and K. R. Summerfelt, *Plant Mol. Biol.* **14**, 269 (1990)], resulting in pCGN3828. *Arabidopsis thaliana*, ecotype No-0 [M. C. Whalen, R. W. Innes, A. F. Bent, B. J. Staskawicz, *Plant Cell* **3**, 49 (1991)], was transformed by *Agrobacterium tumefaciens* cocultivation, and plants were regenerated [D. Valvekens, M. Van Montagu, M. Van Lijsebettens, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5536 (1988), modified].
15. B. Lemieux, M. Miquel, C. Somerville, J. Browse, *Theor. Appl. Genet.* **80**, 234 (1990).
16. D. W. James, Jr., and H. K. Dooner, *ibid.*, p. 241.
17. D. S. Knutzon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2624 (1992).
18. J. Browse, P. J. McCourt, C. R. Somerville, *Anal. Biochem.* **152**, 141 (1986).
19. We thank G. Thompson and V. Knauf for helpful suggestions, S. Becker for DNA sequencing, M. Lassner for preparing the cDNA library, S. Jones and R. Davis for growing the plants, S. Hutchison and J. Byrne for bay seed collection, J. Dimento for FA analysis, L. Comai for providing *Arabidopsis* seeds, and E. Svee for substrate synthesis. C. Shewmaker and B. Martineau critically read the manuscript.

5 March 1992; accepted 15 May 1992

Evidence That Eukaryotes and Eocyte Prokaryotes Are Immediate Relatives

Maria C. Rivera and James A. Lake

The phylogenetic origin of eukaryotes has been unclear because eukaryotic nuclear genes have diverged substantially from prokaryotic ones. The genes coding for elongation factor EF-1α were compared among various organisms. The EF-1α sequences of eukaryotes contained an 11–amino acid segment that was also found in eocytes (extremely thermophilic, sulfur-metabolizing bacteria) but that was absent in all other bacteria. The related (paralogous) genes encoding elongation factor EF-2 and initiation factor IF-2 also lacked the 11–amino acid insert. These data imply that the eocytes are the closest surviving relatives (sister taxon) of the eukaryotes.

The evolutionary antecedents of eukaryotes have been difficult to identify because eukaryotic nuclear gene sequences have diverged extensively from their prokaryotic homologs. Their frequent nucleotide and amino acid changes confound tree reconstruction (*1*) and multiple sequence alignment (*2*) algorithms and cause incorrect trees to be artifactually selected (*3*). The probability of these artifacts could be reduced, however, if slowly changing and well-aligned features, such as inserts that consist of many nucleotides, could be observed and analyzed. We report here the identification, sequencing, and analysis of an 11–amino acid segment within the gene coding for protein synthesis factor EF-1α that occurs in all eukaryotes and eocytes (*4*) but not in other prokaryotes.

Elongation factor EF-1α (termed EF-Tu in eubacteria) is an ubiquitous protein that transports aminoacyl-tRNAs to the ribosome and participates in their selection by the ribosome. The structure of the guanosine diphosphate (GDP)–binding domain of EF-Tu from *Escherichia coli* has been determined by x-ray diffraction (*5*). Within this domain (Fig. 1), sequence $DCPGH_{84}$ (*6*) ends a β strand, forms part of the GDP binding site, and initiates a short α helix. The sequence $KNMITG_{94}$, which is conserved in EF-1α and EF-Tu sequences, terminates this helix. The β strand that follows is terminated by $GPMP_{113}$ at the GDP binding site; $QTREH_{118}$ then starts a 3_{10} helix. The amino acid motifs of the eukaryotic EF-1α are similar, except that the four–amino acid sequence $GPMP_{113}$ is replaced by the 11–amino acid sequence GEFEAGISKDG and its variants.

Unlike that in other prokaryotes, the EF-1α sequence from the eocyte *Sulfolobus acidocaldarius* (*7*) contained an 11–amino acid segment similar to the eukaryotic one. In order to investigate this segment in all three taxonomic subdivisions of the eocytes [the Pyrodictiales, the Thermoproteales, and the Sulfolobales (*8*)], we amplified genomic DNA (*9*) from the eocytes *Pyrodictium occultum*, *Desulfurococcus mucosus*, and *Acidianus infernus* with the polymerase chain reaction (PCR) and primers to the $KNMITG_{94}$ and $QTREH_{118}$ regions (Table 1).

The amino acid sequences, translated from DNA, shared the eukaryotic motif rather than that found in methanogens, halobacteria, and eubacteria (Table 2). The longer, 11–amino acid segment, present in eocytes and eukaryotes, shares little obvious similarity with the shorter, four–amino acid segment found in other prokaryotes. In order to ascertain which form of the segment existed first, we compared sequences of the related (paralogous) proteins EF-2 (termed EF-G in eubacteria) and IF-2

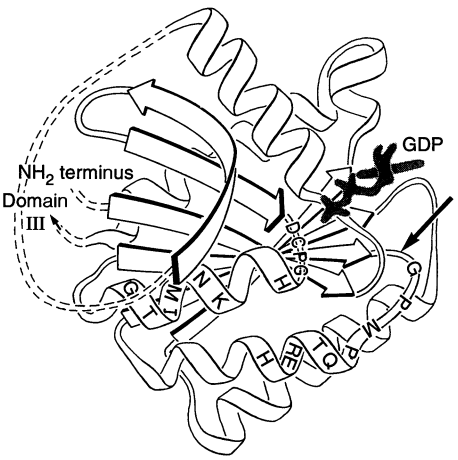

Fig. 1. Structure of the GDP binding domain of trypsin-treated elongation factor EF-Tu from *E. coli* (*5*). Open arrows represent β strands, spirals represent helices, and GDP (labeled) is represented by shading. The dashed line represents a sequence removed by trypsin treatment. The connection to structural Domain III is shown by a small, solid arrow. The large, solid arrow indicates the NH_2-terminus of the amino acid segment ($GPMP_{113}$) analyzed here. The $DCPGH_{84}$, $KNMITG_{94}$, and $QTREH_{118}$ regions are indicated (*6*).

Molecular Biology Institute and Biology Department, University of California, Los Angeles, CA 90024.

because both diverged from EF-1α before the last common ancestor of eukaryotes and prokaryotes (*10*) (Table 2). The sequence $DCPGH_{84}$ (or its variants) was used as the upstream boundary because neither EF-2 (EF-G) nor IF-2 contained the $KNMITG_{94}$ sequence. Sequences from EF-2 and IF-2 all contained the four–amino acid segment, which indicates that $GPMP_{113}$ (or its variants) is the original (plesiomorphic) form of this segment and that the 11–amino acid segment must be the derived (synapomorphic) form.

Two theories have been proposed to explain the origin of eukaryotes: the eocyte theory and the archaebacterial theory (Fig. 2). In the archaebacterial tree (*11*), eukaryotes share a most recent, common ancestor with three groups: the halobacteria, methanogens, and eocytes. In the eocyte tree (*12*), eukaryotes share a most recent, common ancestor only with the eocytes. The tree that requires the fewest changes is most parsimonious and is favored (*13*). Because the 11–amino acid segment in eukaryotes is homologous to that found in eocytes, the observed distribution of the segment among species can be accomplished by one change if the relationships are as shown by the eocyte tree, whereas two changes are required by the archaebacterial tree. Thus, these data support the hypothesis that eocytes are the closest relatives of the eukaryotes.

Several lines of reasoning buttress this interpretation. First, the 11–amino acid segments present in eocytes and eukaryotes are very likely homologous. Eight of eleven amino acids [seven in *Sulfolobus* and *Acidianus*] are identical to the consensus eukaryotic sequence. Amino acid shuffling of the segments (*14*) produces alignments scores that are 6.6 to 6.9 standard deviations lower than those found for the *Saccharomyces cerevisiae–Sulfolobus* alignment, thereby imply-

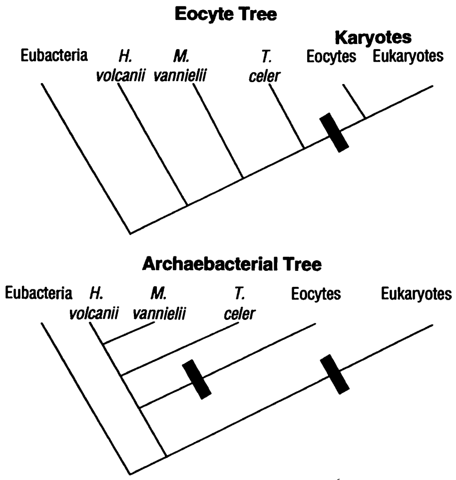

Fig. 2. Rooted tree illustrating two theories proposed to explain the origin of eukaryotes. The trees corresponding to both theories were reconstructed from 16*S* and 18*S* ribosomal RNA sequences (*11, 12*) and rooted in the branch leading to the eubacteria as proposed by others (*26*). Each solid box indicates one change from the 4–amino acid segment to the 11–amino acid form or the reverse. The eocyte tree is favored because it requires only a single change, whereas the archaebacterial tree requires two independent but identical changes. The archaebacterial tree could also be explained by one appearance of the 11–amino acid form and one reappearance of the 4–amino acid form, but even so, two changes would be required. The EF-1α synapomorphism cannot discriminate among the possible branchings below the eocyte-eukaryote node, so that any tree having eocytes and eukaryotes as sister taxa will be consistent with these data. Represented are *Halobacterium volcanii, Methanococcus vannielii,* and *Thermococcus celer.*

Table 1. Nucleotide sequences for *Desulfurococcus mucosus* (*D. muco.*), *Pyrodictium occultum* (*P. occu.*), *Acidianus infernus* (*A. infe.*), and *Sulfolobus acidocaldaritus* (*Su. acid.*) within the $KNMIT_{93}$ to $QTREH_{118}$ segment. The first three nucleotides listed encode the glycine immediately following $KNMIT_{93}$. Lowercase letters indicate primer sequences and uppercase letters indicate nonprimer regions. Triplets corresponding to amino acids are separated by spaces.

Eocytes

P.occu.	ggT GCT AGC CAG GCC	GAC GCC GCC ATC CTA	GTC GTG TCT GCC AGG
D.muco.	... ..A ... .T	..T ..A ..G C	..G ... AGC ..T ...
A.infe.	..A ... .A ..T	..T ..A ..A ..A A..	.CA ..T ... ..T .A.
Su.acid.	... ..A ... .G	... ..T ..A ..T T.G	..A ..A ... ..A .A.
P.occu.	AAG GGT GAG TTC GAG	GCC GGT ATG AGC GCT	GAG GGT
D.muco.	... ...	..A ..G ... ..T ...	C
A.infe.	..A ..A ..A ..T ..A	... ..A ... ... .AA	..A ...
Su.acid.	..A ..A ... .AT ...	... ..A	..A ..A

Table 2. A comparison of elongation factor EF-1α and EF-Tu sequences in, and EF-Tu, EF-G, and IF-2 sequences near, the $KNMIT_{94}$ to $QTREH_{118}$ region. Eleven–amino acid and four–amino acid segments are delimited by solid boxes. The dashed boxes enclose the sequences used as primers. The following designations are used to indicate taxa: eukaryotes are human, *Homo sapiens*; tomato, *Lycopersicon esculentum*; yeast, *Saccharomyces cerevisiae*; and hamster, *Mesocricetus sp.* Eocytes are *D. muco., Desulfurococcus mucosus*; *P. occu., Pyrodictium occultum*; *A. infe., Acidianus infernus*; and *Su. acid., Sulfolobus acidocaldarius.* Methanogens and their relatives are *T. celer, Thermococcus celer* and *Mc.van., Methanococcus vannielii. H. maris.* indicates the halobacterium, *Halobacterium marismortui.* Eubacterial taxa are represented by *Th. mar., Thermotoga maritima; S. plat., Spirulina platensis; E. coli, Escherichia coli*; and Mitoch., yeast mitochondrion. The EF-1α sequences from *Mc.van.* and *Thermoplasma acidophilum* (not shown) contain three and two, respectively, additional amino acids at the beginning of their four–amino acid segments. These extra amino acids do not alter our conclusions because, technically, they represent individual changes (autapomorphies) that cannot serve to distinguish among any trees (*13*). All EF-1α segments (including that from *T. acidophilum*) except those of *Desulfurococcus, Pyrodictium*, and *Acidianus*, which were sequenced by us, were obtained from GenBank.

Taxon	Organism	Left primer		11–amino acid segment	4–amino acid segment	Right primer
Eukaryotes	Human	KNMITG	TSQADCAVLIVAAGV	GEFEAGISKNG		QTREH
Eukaryotes	Tomato	KNMITG	TSQADCAVLIIDSTT	GGFEAGISKDG		QTREH
Eukaryotes	Yeast	KNMITG	TSQADCAILIIAGGV	GEFEAGISKDG		QTREH
Eocytes	*P. occu.*	KNMITG	ASQADAAILVVSARK	GEFEAGMSAEG		QTREH
Eocytes	*D.muco.*	KNMITG	ASQADAAILVVSARK	GEFEAGMSAEG		QTREH
Eocytes	*A.infe.*	KNMITG	ASQADAAIIAVSAKK	GEFEAGMSEEG		QTREH
Eocytes	*Su.acid.*	KNMITG	ASQADAAILVVSAKK	GEYEAGMSAEG		QTREH
Methanogens	*T.celer*	KNMITG	ASQADAAVLVVAVTD		---GVMP	QTKEH
and relatives	*Mc.van.*	KNMITG	ASQADAAVLVVNVDD		AKSGIQP	QTREH
Halobacteria	*H.maris.*	KNMITG	ASQADNAVLVVAADD		---GVQP	QTQEH
Eubacteria	*Th.mar.*	KNMITG	AAQMDGAILVVAATD		---GPMP	QTREH
Eubacteria	*S.plat.*	KNMITG	AAQMDGAILVVSAAD		---GPMP	QTREH
Eubacteria	Mitoch.	KNMITG	AAQMDGAIIVVAATD		---GQMP	QTREH

Sequences of EF-Tu, EF-G, and IF-2 near the KNMITG-QTREH region

Molecule		Conserved					
EF-Tu	*T.celer*	DAPGH	RDFV	KNMITG	ASQADAAVLVVAVTD	---GVMP	QT
EF-Tu	*Mc.van.*	DCPGH	RDFI	KNMITG	ASQADAAVLVVNVDD	AKSGIQP	QT
EF-Tu	*H.maris.*	DCPGH	RDFV	KNMITG	ASQADNAVLVVAADD	---GVQP	QT
EF-Tu	*Th.mar.*	DCPGH	ADYI	KNMITG	AAQMDGAILVVAATD	---GPMP	QT
EF-Tu	*S.plat.*	DCPGH	ADYV	KNMITG	AAQMDGAILVVSAAD	---GPMP	QT
EF-2	Hamster	DSPGH	VDFS	SEVTAA	LRVTDGALVVVDCVS	---GVCV	QT
EF-2	*Su.acid.*	DTPGH	VDFS	GRVTRS	LRVLDGSIVVIDAVE	---GIMT	QT
EF-2	*Mc.van.*	DTPGH	VDFG	GDVTRA	MRAIDGAVVVCCAVE	---GVMP	QT
EF-2	*S.plat.*	DTPGH	VDFT	IEVERS	MRVLDGVIAVFCSVG	---GVQP	QS
IF-2	*E.coli*	DTPGH	AAFT	SMRARG	AQATDIVVLVVAADD	---GVMP	QT

ing homology (*15*). Second, the alignments are well defined. No gaps are needed to align the eukaryotic and eocytic EF-1α sequences, and no gaps are needed to align the EF-2 and the IF-2 sequences with those EF-1α sequences that contain the four–amino acid segment (except for three amino acids unique to *Methanococcus vannielii*). Third, the sequences encoding EF-1α are not likely to be laterally transferred between organisms. EF-1α is present in all cells and, during protein synthesis, interacts with cellular components encoded by genes dispersed throughout the bacterial genome, including aminoacyl-tRNAs, ribosomal proteins, elongation factor EF-Ts, and 16*S* and 18*S* ribosomal RNAs (*16*).

Other results also support a sister relationship between the eukaryotes and eocytes. For example, the major heat shock protein of *Sulfolobus shibatae* is a molecular chaperone related to a eukaryotic t-complex gene (*17*). Similarly, the eukaryotic ribosomal RNA operons (*18*) are organized like those of *Sulfolobus*, *Desulfurococcus*, *Thermoproteus*, and *Thermococcus*. By contrast, the tRNA-containing ribosomal RNA operons of halobacteria, methanogens, and eubacteria (*19*) share a different pattern.

Although many characters support the eocyte tree (*20*), some do not. First, eubacteria, halobacteria, and eukaryotes share ester-linked fatty acids and functional fatty acid synthetases (*21*). This does not support either the archaebacterial or eocyte tree but does support an alternative topology. Second, halobacteria, methanogens, and eocytes have at least traces of a distinct ether lipid (*22*), which supports the archaebacterial tree and does not support the eocyte tree. If both exceptions were valid, no tree would be acceptable. The exceptions therefore emphasize the need for caution and an appreciation of the chimeric origins (*23*) of some nuclear sequences in an analysis of the phylogenetic relationships of eukaryotes. Reconstruction of the prokaryotic ancestry of eukaryotes requires caution; however, the phylogenetic distribution of the 11–amino acid segment implies that the eocytes are the closest surviving relatives of eukaryotes. This lends support to the proposal (*12*) that the eukaryotes and eocytes comprise a monophyletic superkingdom, the karyotes.

REFERENCES AND NOTES

1. J. Felsenstein, *J. Syst. Zool.* **27**, 401 (1978).
2. J. A. Lake, *Mol. Biol. Evol.* **8**, 378 (1990); D. P. Mindell, in *Phylogenetic Analysis of DNA Sequences*, M. Miyamoto and J. Cracraft, Eds. (Oxford Univ. Press, London, 1991), pp. 119–136.
3. C. Patterson, *Nature* **344**, 199 (1990).
4. J. A. Lake, E. Henderson, M. W. Clark, M. Oakes, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3786 (1984).
5. F. Jurnak, *Science* **230**, 32 (1985); T. F. M. la Cour, J. Nyborg, S. Thirup, B. F. C. Clark, *EMBO J.* **4**, 2385 (1985).
6. The *E. coli* numbering system was used. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
7. A. T. Matheson, J. Auer, C. Ramirez, A. Boeck, in *The Ribosome: Structure, Function, and Evolution*, W. E. Hill *et al.*, Eds. (American Society for Microbiology Press, Washington, DC, 1990), pp. 617–633.
8. K. O. Stetter *et al.*, *FEMS Microbiol. Rev.* **75**, 117 (1990).
9. Genomic DNA was isolated from cell pastes of *Desulfurococcus, Pyrodictium*, and *Acidianus* by SDS–proteinase-K lysis (*24*). Genomic DNA was further purified by gel electrophoresis in the presence of ethidium bromide. The high molecular weight band was excised from the gel and melted at 70°C, and a small aliquot was amplified by PCR (*25*). Degenerate primers designed from the amino acid alignments of known EF-1α sequences for the $KNMITG_{94}$ and the $QTREH_{118}$ regions were used for the PCR reactions. PCR products were analyzed by electrophoresis. A single band of approximately 120 bp was purified from agarose gels and cloned into the pCR1000 vector (Invitrogen) according to the manufacturer's recommendations. Colonies containing the insert were sequenced by the dideoxy chain termination method with Sequenase (U.S. Biochemical Corp.) as recommended by the manufacturer.
10. H. R. Bourne, D. A. Sanders, F. McCormick, *Nature* **349**, 117 (1991).
11. C. R. Woese, *Microbiol. Rev.* **51**, 221 (1987).
12. J. A. Lake, *Nature* **331**, 184 (1988).
13. W. M. Fitch, *Am. Nat.* **111**, 223 (1977); E. O. Wiley, *Phylogenetics* (Wiley, New York, 1981).
14. M. O. Dayhoff, W. C. Barker, L. T. Hunt, *Methods Enzymol.* **91**, 524 (1983).
15. M. Waterman and M. Eggert, *J. Mol. Biol.* **197**, 723 (1987).
16. W. E. Hill *et al.*, Eds., *The Ribosome: Structure, Function, and Evolution* (American Society for Microbiology Press, Washington, DC, 1990).
17. J. D. Trent, E. Nimmesgern, J. S. Wall, F.-U. Hartl, A. L. Horwich, *Nature* **354**, 490 (1991).
18. S. A. Gerbi, in *Molecular Evolutionary Genetics*, R. J. MacIntyre, Ed. (Plenum, New York, 1985), pp. 419–517.
19. N. Larsen, H. Leffers, J. Kjems, R. A. Garrett, *System. Appl. Microbiol.* **7**, 49 (1986).
20. J. A. Lake, *Trends Biochem. Sci.* **16**, 46 (1991).
21. M. Kamekura and M. Kates, "Lipids of halophilic archaebacteria," in *Halophilic Bacteria*, F. Rodrigues-Valera, Ed. (CRC Press, Boca Raton, FL, 1988), pp. 25–54.
22. V. Lanzotti *et al.*, *Biochim. Biophys. Acta* **1004**, 44 (1989).
23. M. W. Gray, *Trends Genet.* **5**, 294; L. Margulis, *Symbiosis in Cell Evolution* (Freeman, San Francisco, 1981); S. L. Baldauf and J. D. Palmer, *Nature* **344**, 262 (1990).
24. B. G. Herrmann and A. Frischauf, *Methods Enzymol.* **152**, 180 (1987).
25. R. K. Saiki, in *PCR Protocols: A Guide to Methods and Applications*, M. A. Innis, D. H. Gelfand, J. J. Sninsky, T. J. White, Eds. (Academic Press, New York, 1990), pp. 13–20.
26. J. P. Gogarten *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6661 (1989); N. Iwabe *et al.*, *ibid.*, p. 9355.

27. We thank J. Arrington for technical assistance; A. Scheinman, G. Shankweiler, T. Atha, and A. Aguinaldo for discussions, suggestions and advice; and K. Stetter for cells of eocytes, methanogens, and their relatives. Supported by grants from the NSF, NIH, and the Alfred P. Sloan Foundation to J.A.L.

27 February 1992; accepted 7 May 1992

Conformation of the TAR RNA-Arginine Complex by NMR Spectroscopy

Joseph D. Puglisi, Ruoying Tan, Barbara J. Calnan, Alan D. Frankel, James R. Williamson*

The messenger RNAs of human immunodeficiency virus–1 (HIV-1) have an RNA hairpin structure, TAR, at their 5′ ends that contains a six-nucleotide loop and a three-nucleotide bulge. The conformations of TAR RNA and of TAR with an arginine analog specifically bound at the binding site for the viral protein, Tat, were characterized by nuclear magnetic resonance (NMR) spectroscopy. Upon arginine binding, the bulge changes conformation, and essential nucleotides for binding, U23 and A27·U38, form a base-triple interaction that stabilizes arginine hydrogen bonding to G26 and phosphates. Specificity in the arginine-TAR interaction appears to be derived largely from the structure of the RNA.

The diverse structures formed by RNA molecules contribute to their specific recognition by proteins (*1*, *2*). The interaction of the HIV Tat protein with TAR, an RNA hairpin located at the 5′ end of the viral mRNAs, provides a well-characterized system for the study of RNA-protein recognition. The binding of Tat to TAR is essential for Tat to function as a transcriptional activator (*3–6*). The predicted secondary structure of TAR consists of two stem regions separated by three unpaired nucleotides (a bulge) and a loop of six nucleotides (Fig. 1A). Many mutational studies have identified nucleotides in and near the bulge that are important for specific binding of Tat (*4*, *7–10*). The loop region is not involved in Tat binding but is important for activation of transcription (*4–6*, *10*, *11*). Specific binding of Tat to TAR is mediated by a single arginine (*12*) within a nine-residue stretch of basic amino acids, as shown by specific binding of model peptides

J. D. Puglisi and J. R. Williamson, Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139.
R. Tan, B. J. Calnan, A. D. Frankel, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142.

*To whom correspondence should be addressed.