THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1989 by The American Society for Biochemistry and Molecular Biology, Inc.
Vol. 264, No. 19, Issue of July 5, pp. 10965-10975, 1989
Printed in U.S.A.

The Complete Nucleotide Sequence, Gene Organization, and Genetic Code of the Mitochondrial Genome of *Paracentrotus lividus**

(Received for publication, June 3, 1988)

Palmiro Cantatore, Marina Roberti, Guglielmo Rainaldi, Maria N. Gadaleta, and Cecilia Saccone

From the Department of Biochemistry and Molecular Biology, University of Bari and Centro di Studio sui Mitocondri e Metabolismo Energetico, Consiglio Nazionale delle Ricerche, 70126 Bari, Italy

The 15,697-nucleotide sequence of *Paracentrotus lividus* mitochondrial DNA is reported. This genome codes for 2 rRNAs, 22 tRNAs, and 12 mRNAs which specify 13 subunits of the mitochondrial inner membrane respiratory complexes. The gene arrangement differs from that of other animal species. The two ribosomal genes 16 S and 12 S are separated by a stretch of about 3.3 kilobase pairs which contains the ND1 and ND2 genes and a cluster of 15 tRNA genes. The ND4L coding sequence is not contained in the ND4 mRNA but has its own mRNA which maps between the $tRNA^{Arg}$ and the Co II genes. The main noncoding region, located in the tRNA gene cluster, is only 132 nucleotides long, but contains sequences homologous to the mammalian displacement loop. Other short noncoding sequences are interspersed in the genome: they contain a conserved AT consensus which probably has a role in transcription or RNA processing. As regards the mitochondrial genetic code, the codons AGA and AGG specify serine and are recognized by a tRNA with a GCU anticodon, whereas AUA and AAA code for isoleucine and asparagine rather than for methionine and lysine. Except for ND4L which starts with AUC and ATPase 8 which starts with GUG, AUG is used as the initiation codon. In 11 out of 13 cases the genes terminate with the canonical stop codons UAA or UAG. These observations suggest that during invertebrate evolution each lineage developed its own mechanism of mitochondrial DNA replication and transcription and of RNA processing and translation.

Vertebrate mtDNA,[1] a double stranded circular molecule of 16.3–17.5 kilobase pairs, codes for 2 rRNAs (16 S and 12 S), 22 tRNAs, and 11 mRNAs. These latter specify 13 subunits of the respiratory complexes. These are: the subunits I, II, and III of cytochrome oxidase (CoI, CoII, CoIII); the subunits 6 and 8 of the F_0-ATPase (ATPase 6 and 8); the apoprotein of cytochrome *b*; the seven subunits of the NADH dehydrogenase (ND1, -2, -3, -4, -4L, -5, and -6) (for review see Attardi, 1985 and Cantatore and Saccone, 1987). The rest of the molecule consists of a noncoding region (D-loop region) which varies in size from 879 nt (mouse) (Bibb *et al.*, 1981) to 2134 nt (*Xenopus laevis*) (Roe *et al.*, 1985) and which contains the replication origin of the H-strand and the promoters for H- and L-strand transcription. Vertebrate mt gene organization is a model of economy and compactness: few or no nucleotides separate individual genes, and the tRNA genes are interspersed with almost absolute regularity between mRNA and rRNA coding sequences. This arrangement is consistent with the transcription of the mtDNA H-strand in the form of a single molecule processed by precise endonucleolytic cleavages before and after each tRNA sequence to yield the mature transcripts.

The mt genomes of some invertebrates, such as *Drosophila yakuba* (Clary and Wolstenholme, 1985), *Ascaris suum* (Wolstenholme *et al.*, 1987), and *Caenorhabditis elegans*[2] have also been described. These organisms have an mt gene content similar to vertebrates, but the distribution of the genes between the two strands and the order in which they are arranged are strikingly different. This may be explained by assuming that during animal evolution some characteristics of mtDNA (*e.g.* the compactness of the molecule and the gene content) have been preserved, but extensive alterations of the mt gene order have taken place.

In order to understand the rules which governed animal mtDNA evolution in terms of gene structure and gene order, more information about the mtDNA structure in a variety of vertebrates and invertebrates must be obtained. The sea urchin is an excellent subject for such studies. It is one of the most developed invertebrates and is used as a model in studies of embryonic development and cell differentiation. Moreover, as the Echinodermata phylum contains many groups which cover a time range from 15 to 450 million years (see Smith, 1984) molecular analysis of mtDNA may provide a new tool for solving phylogenetic problems, especially in those cases where the fossil record is poor or incomplete. We have directed our attention to the mtDNA molecule of the most common sea urchin species in the Mediterraneum: the *Paracentrotus lividus* and can now present the complete sequence of its mtDNA. This provides new information about gene organization and transcription and throws light on certain features of the genetic code.

MATERIALS AND METHODS

P. lividus mtDNA, extracted from fresh eggs as described previously (Roberti *et al.*, 1984), was cleaved with *Eco*RI, *Pst*I, and *Bam*HI, producing seven fragments that were cloned in the plasmid vectors pBR322, pUC8, and pEMBL8 (Cantatore *et al.*, 1987a). To determine the sequence of the fragments, bidirectional or unidirectional deletions of predictable sizes were produced. This was done by digesting

* This work has been financed by funds from the Ministero Pubblica Instruzione and from the "Progetto Finalizzato Ingegneria Genetica e Basi Molecolari delle Malattie Ereditarie," Consiglio Nazionale delle Ricerche (National Research Council). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J04815.

[1] The abbreviations used are: mt, mitochondria; nt, nucleotides; D-loop, displacement loop; H-strand, heavy strand; L-strand, light strand.

[2] D. R. Wolstenholme, personal communication.

the fragments *Bam*2, *Bam*1-2, *Bam*1-4, and *Bam*1-5 with the nuclease *Bal*31, and the fragments *Bam*1-1, *Bam*1-3, and *Bam*1-6 with a combination of exonuclease III and mung bean nuclease. The *Bal*31 digestion was performed on isolated fragments, by incubating for varying periods of time at 30 °C 1 pmol of fragment ends with 1 unit of enzyme (New England Biolabs). For the procedure using exonuclease III and mung bean nuclease, linearized recombinant plasmids were incubated for different periods of times at 30 °C with 4 units/μg exonuclease III (Stratagene), followed by 3 units/μg mung bean nuclease (Stratagene) for 30 min at 30 °C. The procedure was carried out from both sides of the fragments. The deletion products were cloned in the plasmid pUC8 and the sequence of the first 300–400 bases of each insert was determined by the dideoxynucleotide chain termination method of Sanger (Sanger *et al.*, 1977), adapted for double stranded templates (Chen and Seeburg, 1985). The sequencing strategy is reported in Fig. 1; at least 80% of the sequence was determined on both strands. Sequence analysis was carried out on a VAX/VMS 4.2 operating system by using the programs ACNUC and GLORIA developed in our laboratory (Gouy *et al.*, 1985).

RESULTS AND DISCUSSION

Genome Organization and Genetic Code—Fig. 2 reports the 15,697-nucleotide sequence of the *P. lividus* mtDNA. It contains the same genes found in vertebrate mt genomes. The largest noncoding sequence is 132-nt long and it is placed between the $tRNA^{Thr}$ and the $tRNA^{Pro}$ genes. Because most of the genes have about the same size as in vertebrate mtDNA, the smaller size of *P. lividus* mtDNA can be accounted for mostly by a reduction in size of the noncoding region. The organization of the *P. lividus* genes (Fig. 3) is different from those of vertebrates and other invertebrates. In particular, the two ribosomal genes 12 S and 16 S are not adjacent but are separated by 3,339 base pairs. As opposed to vertebrate mt genomes, where the ND4L gene is located near the ND4 and shares the same mRNA (Cantatore and Saccone, 1987), the ND4L gene of *P. lividus* is located between the $tRNA^{Arg}$

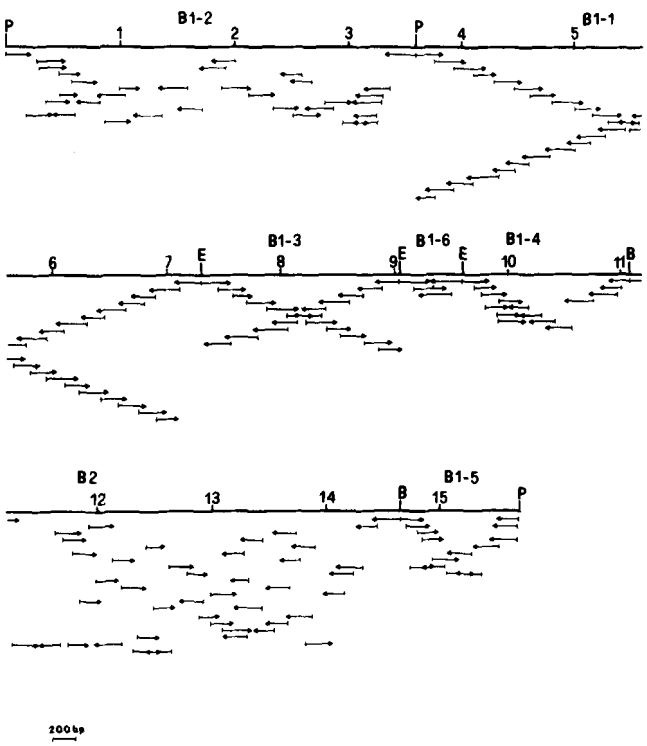

FIG. 1. ***P. lividus* mtDNA sequencing strategy.** The *arrows* show the extent and direction of sequence determined from individual pUC8 subclones by the dideoxynucleotide termination method applied to double stranded templates (Sanger *et al.*, 1977; Chen and Seeburg, 1985). The *numbers* represent the genomic length in kilobases. Restriction sites: *B*, *Bam*HI; *E*, *Eco*RI; *P*, *Pst*I. *bp*, base pairs.

and CoII genes and has its own mRNA (Cantatore *et al.*, 1987a). Therefore, in *P. lividus* the mtDNA codes for 12 mRNA and not for 11 as in vertebrates. Out of 22 tRNA genes, 15 are clustered between the 12 S and ND1 genes, and only 6 ($tRNA^{Phe}$, $tRNA^{Leu}_{UUR}$, $tRNA^{Ile}$, $tRNA^{Lys}$, $tRNA^{His}$, $tRNA^{Ser}_{AGN}$) are in positions corresponding to those of vertebrates. The polarity of the tRNA genes is also unusual: only 15 tRNAs have the same polarity as vertebrate genes, and 17 tRNA genes (not 14 as in vertebrates) are transcribed in the same direction as ribosomal genes. Gene overlaps on the same strand occur in two places. The more remarkable is a 16-nt out-of-phase overlap of the 3′ terminus of the ATPase 8 gene with the 5′ end of the ATPase 6. In mammals these two genes overlap for 40–46 nt (Attardi, 1985), whereas in the *D. yakuba* (Clary and Wolstenholme, 1985) the overlap is restricted to 7 nt. The second case is a 10-nt overlap between the 3′ end of ND4 and the 5′ end of the $tRNA^{His}$. Overlaps of 1 nt can also be noted between the 3′ ends of the genes for $tRNA^{Leu}_{CUN}$ and $tRNA^{Ala}$, $tRNA^{Cys}$ and $tRNA^{Val}$, CoIII, and $tRNA^{Ser}_{UCN}$, encoded by the opposite strands.

The analysis of the codon usage of *P. lividus* mt protein genes (Fig. 4) has revealed an unusual genetic code. It was previously reported (Cantatore *et al.*, 1987a) that the codons AGA and AGG are not used as stop codons but specify serine. Further analysis has shown that the codons AUA (internal) and AAA are not used as methionine and lysine but encode isoleucine and asparagine. This latter observation is supported by several pieces of evidence. When the *P. lividus* ND1, ND4L, ND5, and cytochrome *b* genes are aligned with the corresponding human (Anderson *et al.*, 1981), *X. laevis* (Roe *et al.*, 1985) and *Strongylocentrotus purpuratus* genes (Jacobs *et al.*, 1988), the *P. lividus* triplet AAA aligns with asparagine with a frequency ranging from 42% (vertebrates) to 90% (*S. purpuratus*). Similarly, the *P. lividus* AUA codon corresponds to isoleucine with a frequency of about 50% in vertebrates and 75% in *S. purpuratus*. It may be also observed that in those positions where lysine appears simultaneously in man, rat, and *X. laevis*, the *P. lividus* codon is AAG in 20 cases out of 21. Likewise, in the 14 cases where the methionine is conserved in the three genomes above mentioned, the *P. lividus* codon is always AUG. Finally, a comparison between *P. lividus* and *S. purpuratus* shows a different substitution pattern for AAA and AUA codons with respect to AAG and AUG. In particular, AAA and AUA change frequently to AAY and AUY, respectively, suggesting that these are silent substitutions. On the contrary the codons AAG and AUG are highly conserved (around 80%), as would be expected if they were single codons.

Protein and Ribosomal Genes—Ten of the *P. lividus* protein genes differ in size by less than 3% from the corresponding human genes. In the remaining three greater differences can be noted. The ATPase 8 gene is 20% smaller than its human counterpart, lacking a segment of 14 codons at its 3′ end; the ND6 gene is 42 nt shorter than human ND6, due to deletions at the 3′ end. The ND5 is 105 nt longer than the human ND5, due to a 5′ end extension (Cantatore *et al.*, 1987b). As far as the initiation codons are concerned (see Table I), eight genes initiate with AUG, while in three (CoIII, ND3, ND4) AUG is immediately preceded by AUA or AUU codons, which also function as start codons in other systems (Attardi 1985). The alignment of these genes with the corresponding human (Anderson *et al.*, 1981) and *S. purpuratus* (Jacobs *et al.*, 1988) genes produces a better homology when the genes initiate with AUG, thus suggesting that in these cases the start codon is also AUG.

The ND4L gene has a AUC codon located either immedi-

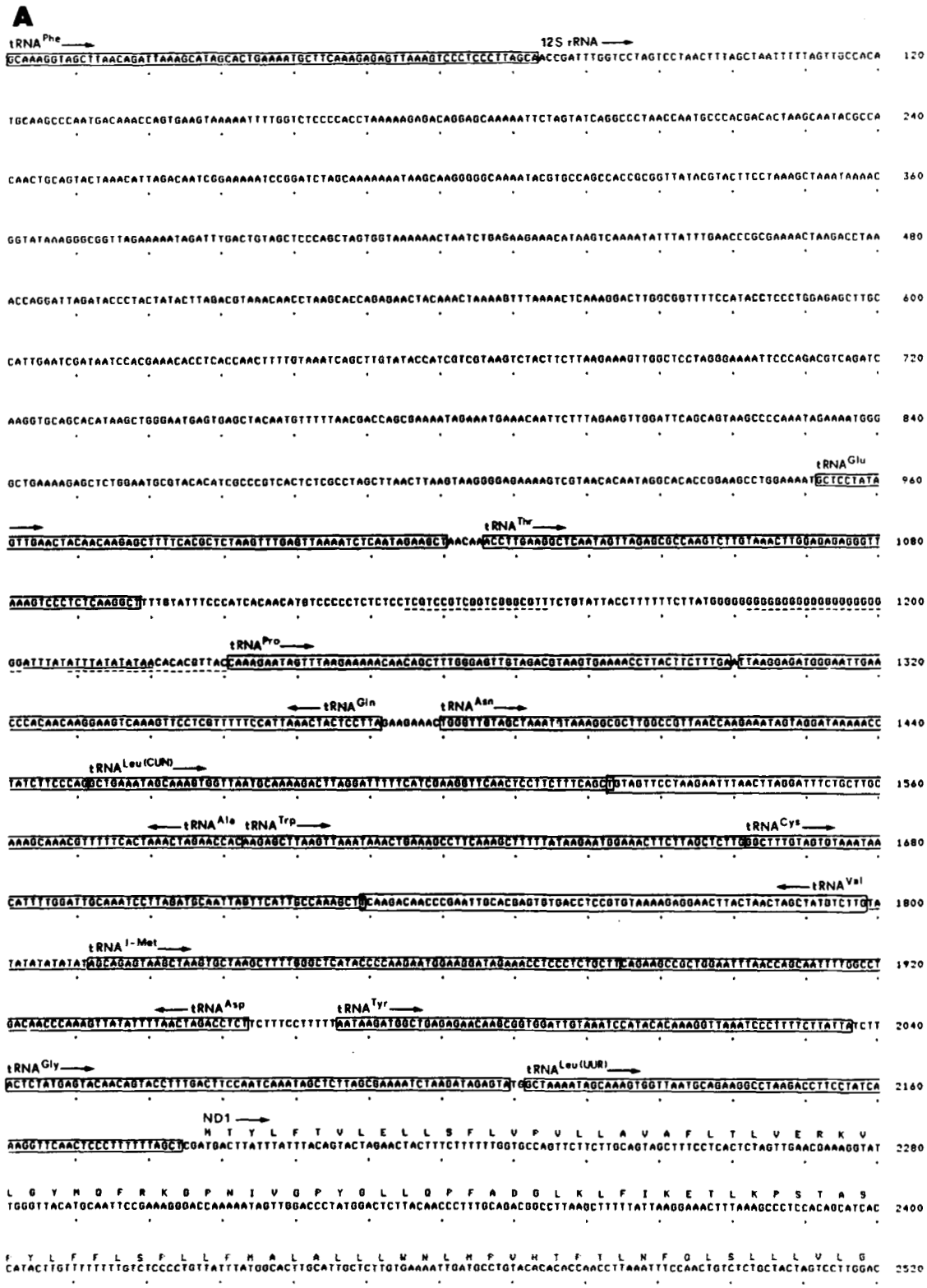

FIG. 2. **Sequence of *P. lividus* mtDNA.** Nucleotide 1 is the first nucleotide of the tRNA^{Phe}. All coding regions read from left to right except for ND6 and five tRNA genes encoded by the L-strand. The amino acid translation for the 13 mt proteins is indicated by one-letter amino acid abbreviation written above the nucleotide sequence. The regions encoding tRNA genes are *boxed*. The conserved sequence block-like regions at nucleotides 1135–1154, 1182–1202, 1209–1230 are *underlined* with *dashed lines*. The AT-rich regions, which have been aligned to generate the consensus sequence, are *underlined* with *solid lines*. Only some sections of the DNA (less than 20%) have been sequenced on just one strand.

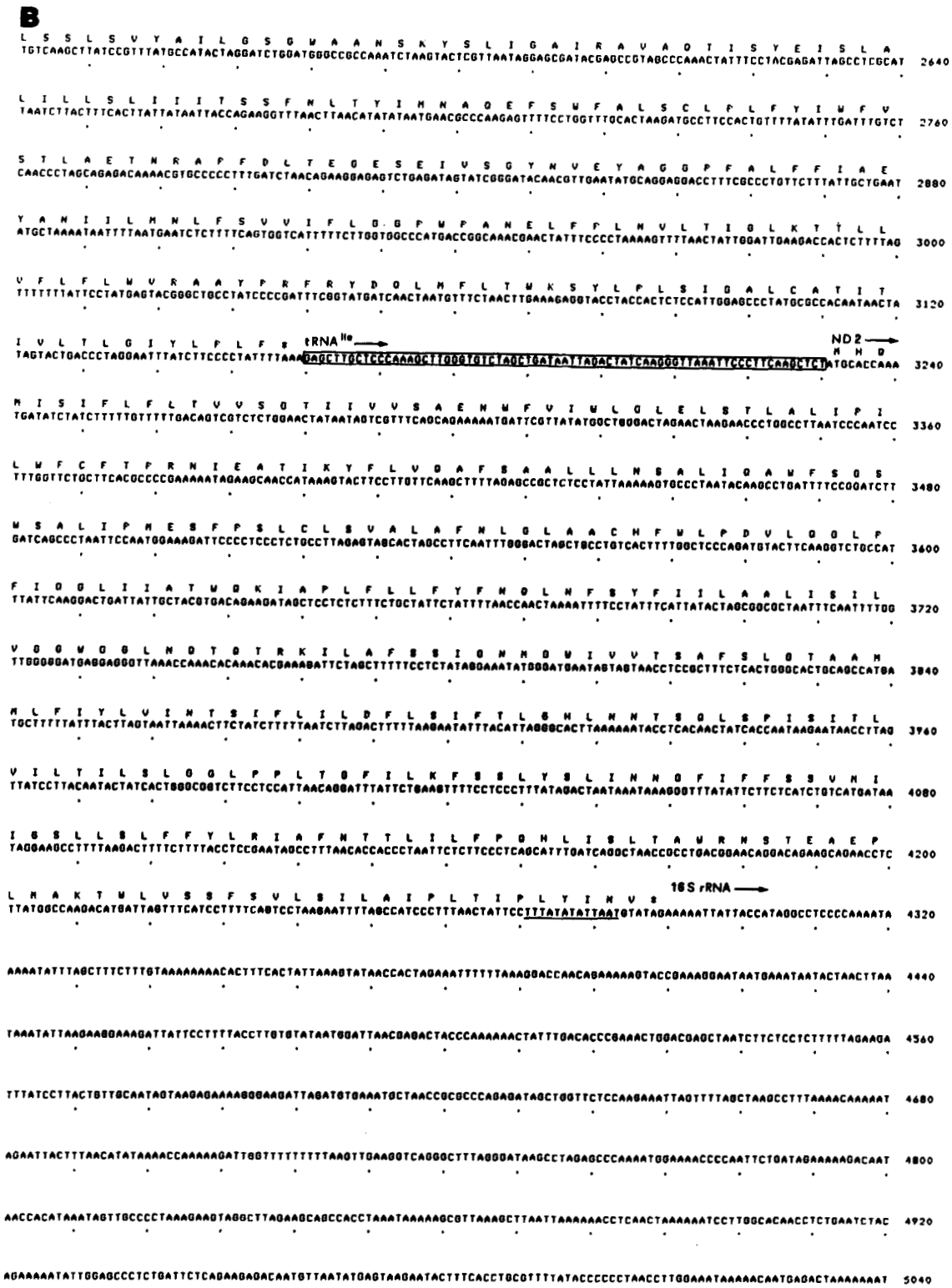

FIG. 2—*continued*

ately after the tRNA^{Arg} or after 1 nt, depending on the structure selected for the aminoacyl stem of this tRNA (see Fig. 5). The first AUG codon is separated from the 3′ end of the tRNA^{Arg} by 28 nt. Were this the start codon it would make the sea urchin ND4L gene about 10% smaller than that of mammals, erasing a region which retains a high degree of homology (64%) with its human counterpart (Cantatore *et al.*, 1987a). This suggests that the gene really starts with AUC. The ATPase 8 gene probably initiates with the codon GUG which is situated just after the 3′ end of tRNA^{Lys}. If the gene initiated from the first AUN in frame codon, located 27 nt downstream the 3′ end of the tRNA^{Lys}, the gene would be shortened, thus losing the first 4 amino acids (nucleotides 8526–8537), which are universally conserved in vertebrate, invertebrate, and fungal proteins (Attardi, 1985). The GUG codon has been reported to occur as the first in frame triplet both in prokaryotic genes (Stormo *et al.*, 1982) and in a number of mitochondrial genes such as the ND5 gene of *D.*

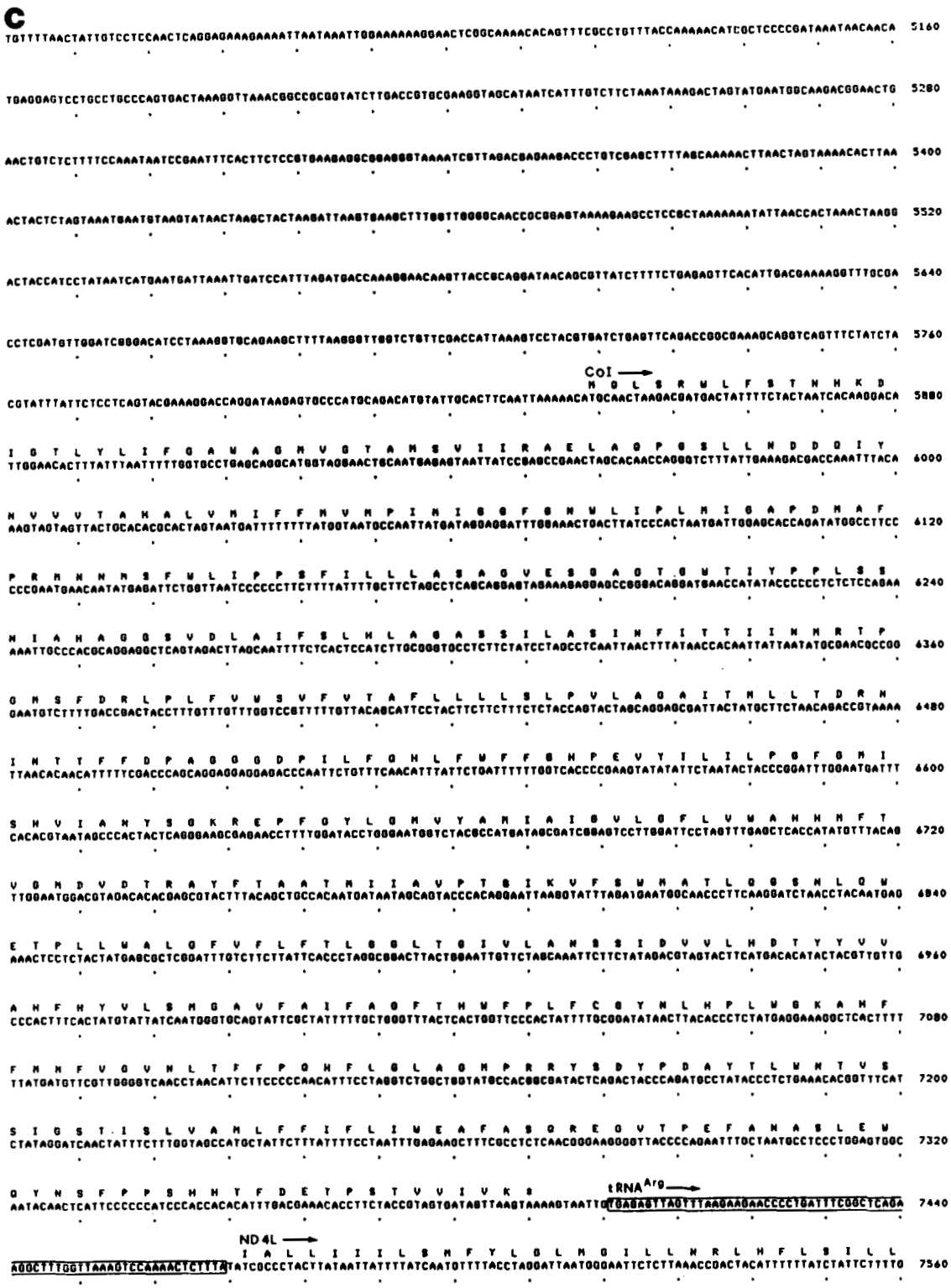

FIG. 2—*continued*

yakuba (Clary and Wolstenholme, 1985) and the cytochrome *b* of sunflower.[3] Recently Gadaleta *et al.* (1989) have presented evidence for the initiation role of the GUG codon in the ND1 gene of both mouse and rat.

The distribution of the initiation codons in vertebrate mtDNAs has been recently investigated by some of us (Gadaleta *et al.*, 1989). It has been observed that the AUA, AUU, AUC, and GUG codons occur only in those cases where there are no spacer nucleotides between two adjacent genes, otherwise AUG codons are used. This regularity has also been found in *P. lividus* where the AUC and GUG codons found in the ND4L and ATPase 8 are immediately preceded by the end of the previous gene. It may be that when the translation of mRNAs starts from the first base, any AUN codon and possibly even GUG can act as initiator, whereas AUG is present where there are 5′-noncoding nucleotides and a more

[3] R. Gallerani, personal communication.

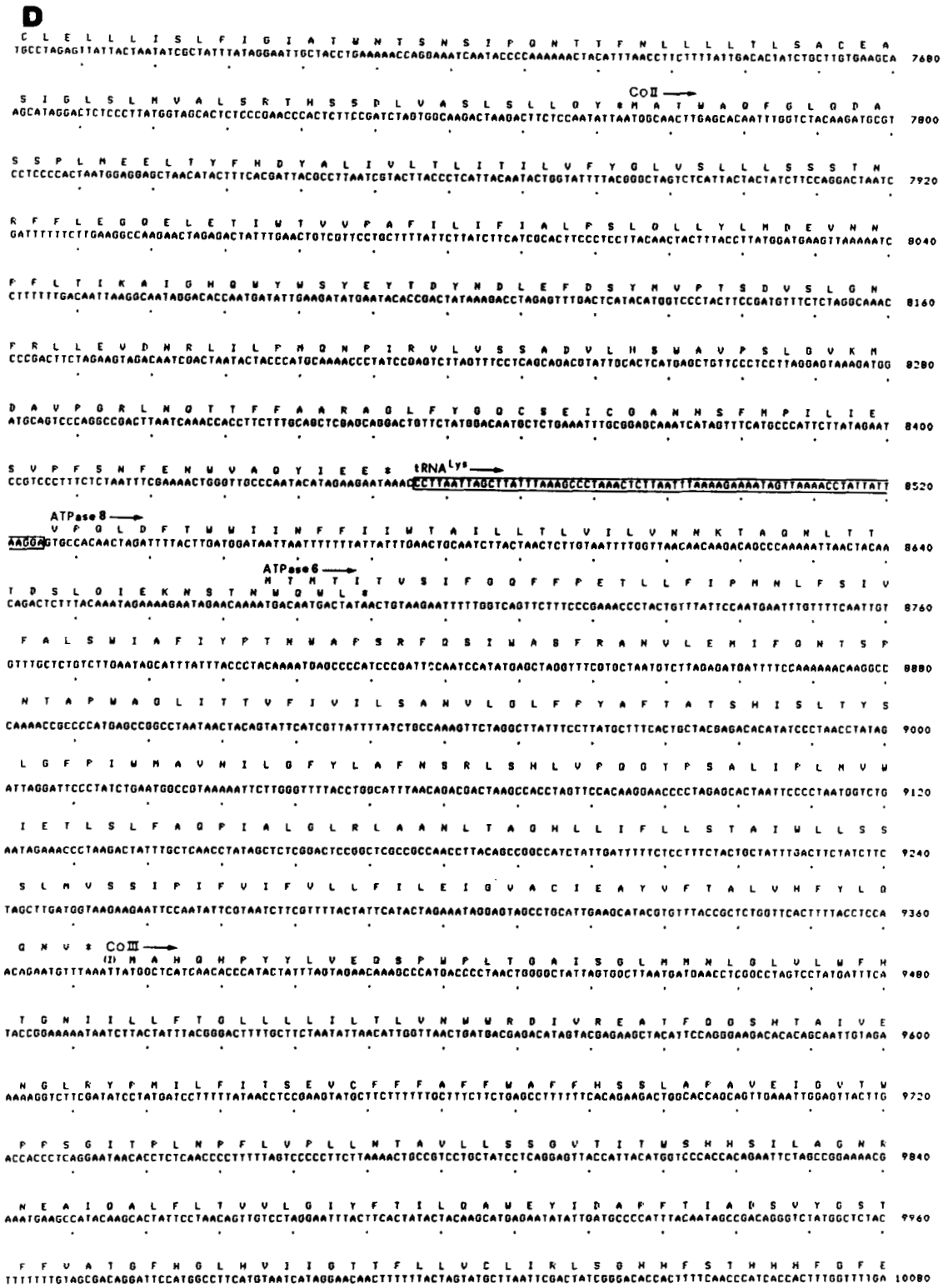

FIG. 2—*continued*

accurate choice of the correct start codon is necessary. As for the stop codons, 2 out of the 13 *P. lividus* mt genes lack a complete termination codon, but end with TA. Completion of the stop codon probably occurs after transcription by addition of a poly(A) tail at the mRNA 3′ end (Ojala *et al.*, 1981).

The two *P. lividus* 12 S and 16 S rRNA genes are 883 and 1549 nt long as results from the mapping of the 5′ and 3′ termini of the corresponding transcripts,[4] whereas the 5′ ends of both genes border without any spacer nucleotide on the $tRNA^{Phe}$ and ND2 genes, respectively, and their 3′ ends overlap by few nucleotides with the downstream genes ($tRNA^{Glu}$ and CoI). This may be due to the presence of transcription

[4] P. Cantatore, unpublished observations.

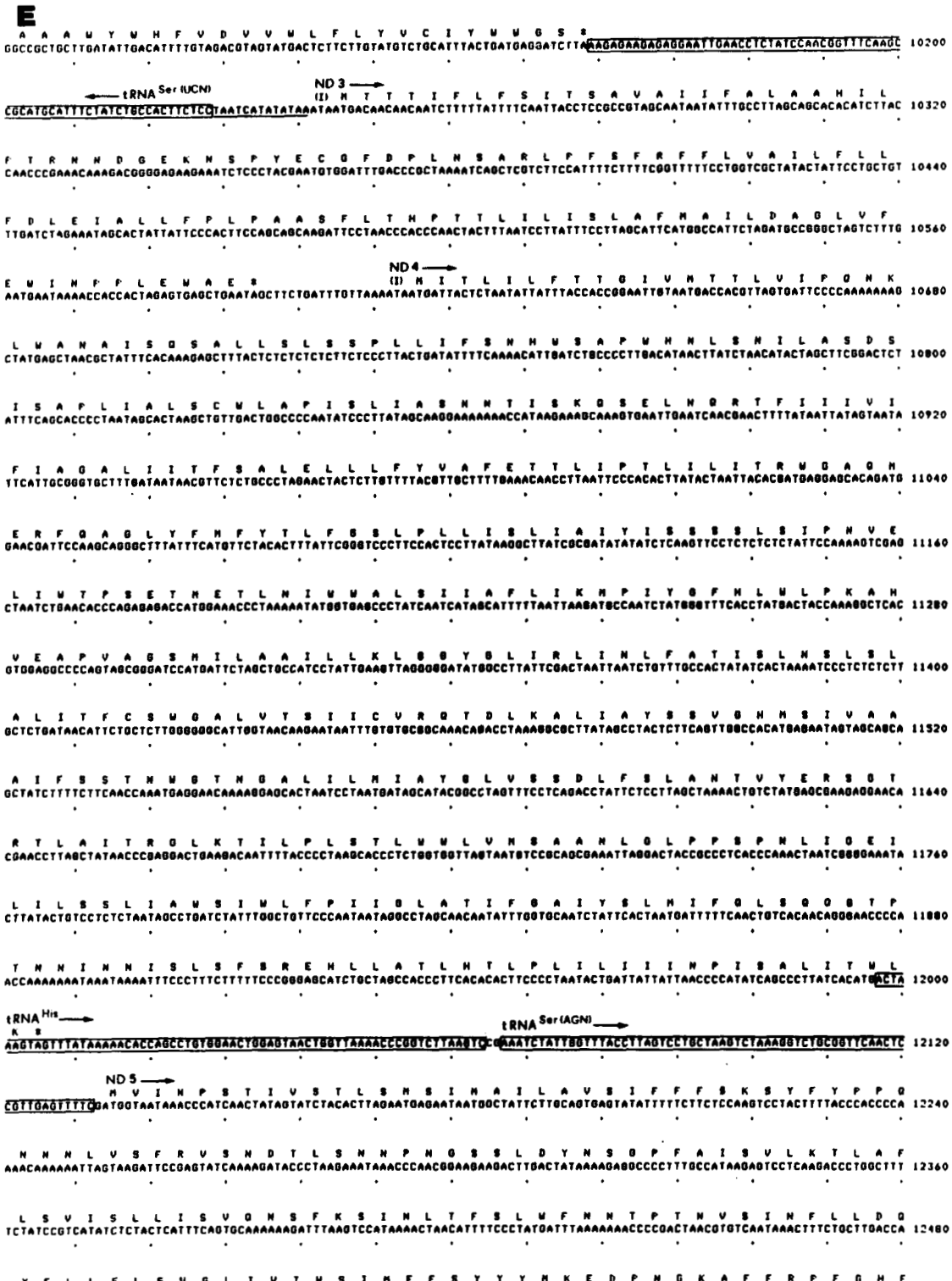

FIG. 2—*continued*

termination sites at the end of the two rRNA genes or to an alternative processing. This observation, together with the other unusual aspects of the gene organization of *P. lividus*, like the unusual location of the two rRNA genes and the absence of tRNAs at several junctions between mRNA genes, raises a series of questions about the mechanism of RNA synthesis and processing in sea urchins. To provide the answers, exact determination of the 5′ and 3′ ends of the mRNAs and studies of the transcription mechanism are in progress.

Sea urchin mtrRNAs can be folded in secondary structures analogous to those proposed for mammalian, *Drosophila*, and *Xenopus* mtrRNAs (Glotz *et al.*, 1981; Zweib *et al.*, 1981; Clary and Wolstenholme, 1985; Roe *et al.*, 1985). The differences

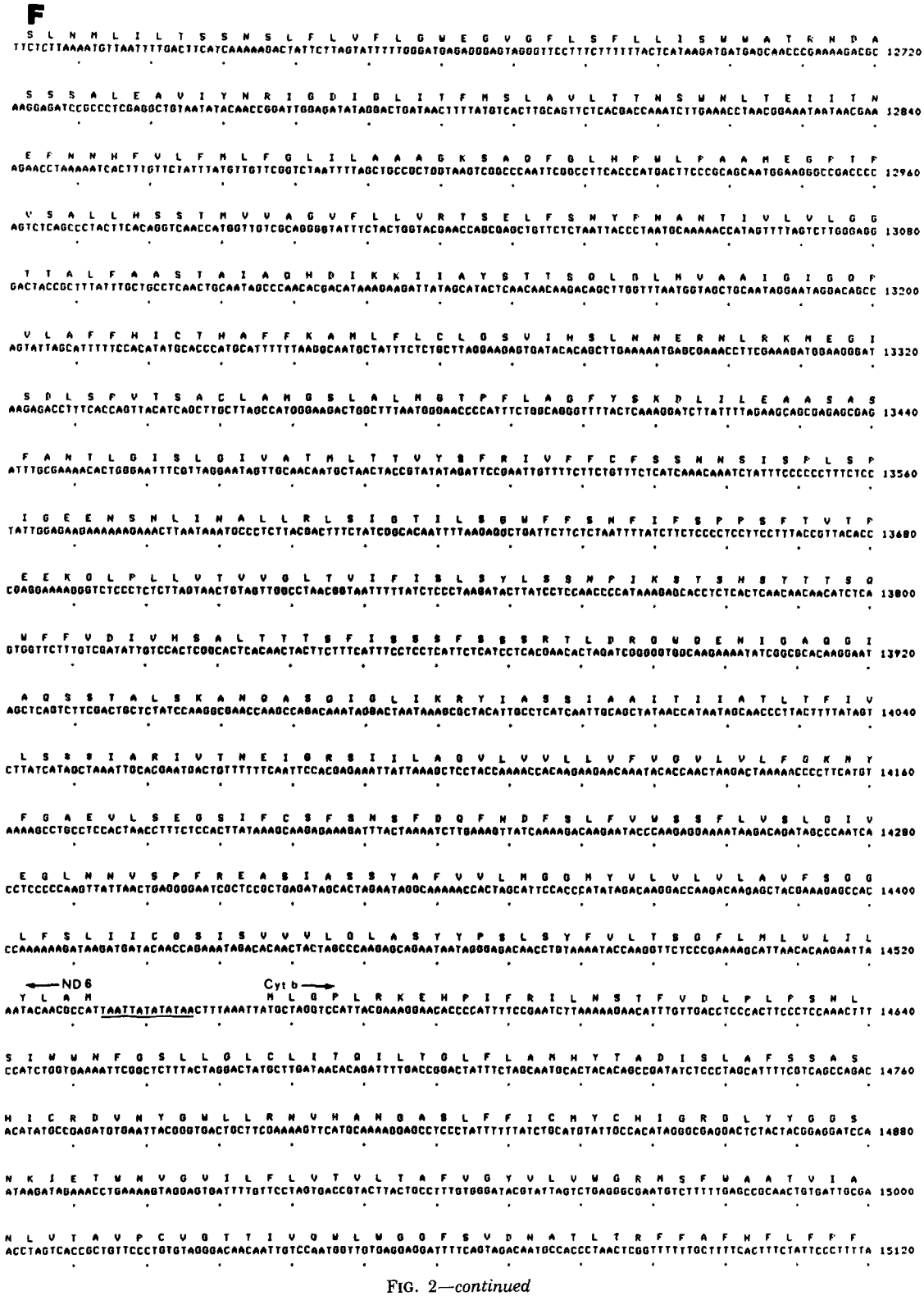

FIG. 2—*continued*

consist in the variation in size of some loops or stems and in the addition or deletion of some domains. This situation agrees with the view of Gray *et al.* (1984) who consider addition or deletion of rRNA pieces a common feature of rRNA evolution.

tRNA Genes—Fig. 5 shows the cloverleaf structures of the 22 *P. lividus* mt tRNAs. Like other animal mt tRNAs, sea urchin tRNAs are of a reduced size (from 66 to 73 nt), lack several universal bases, and have some mismatchings and a high proportion of AT. The comparison of the *P. lividus* mt tRNAs with the corresponding tRNA genes of *D. yakuba*, *X. laevis*, and man reveals that the anticodon loop and stems and the D-stems are the most conserved regions, whereas the most variable ones are the T-stems and the amino acid stems. In general, the degree of conservation of the 5′ half is higher than that of the 3′ half. Some sea urchin mt tRNAs have a

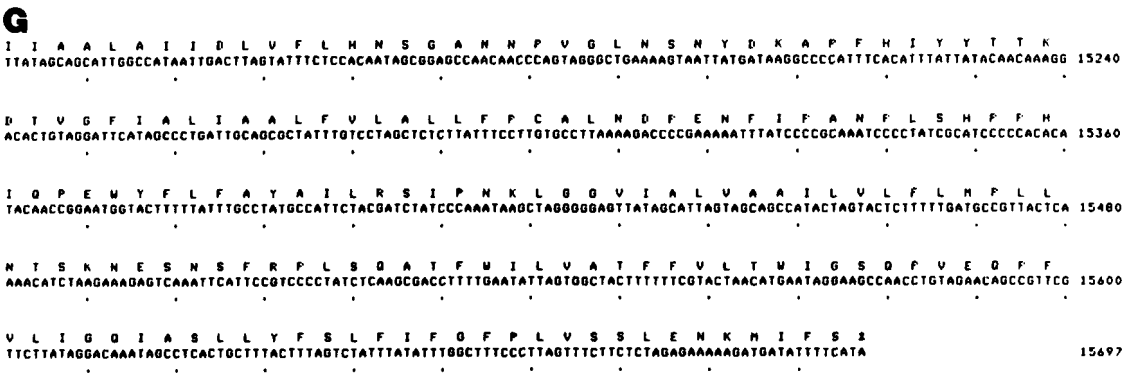

FIG. 2—*continued*

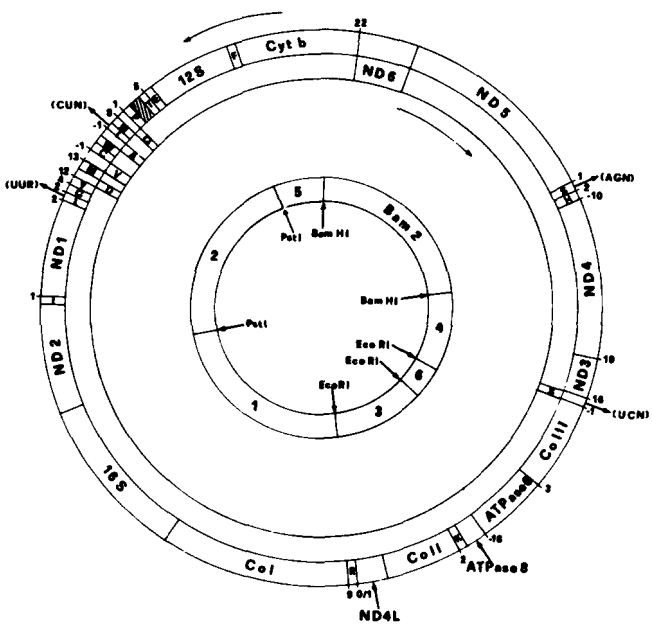

FIG. 3. **Overall gene organization of the *P. lividus* mtDNA.** The *inner circle* shows the *Bam*HI, *Eco*RI, and *Pst*I restriction sites. The two outer circles report the gene order. The *arrows inside* and *outside* these circles show the direction of transcription of the respective genes. Each tRNA is identified by the one-letter amino acid code and individual serine and leucine tRNA genes are identified by the codon family (in *parentheses*). The number of apparently noncoding nucleotides situated between the different genes is shown at the gene boundaries on the outer edge of the map. *Negative numbers* indicate overlapping nucleotides of adjacent genes. The *dashed area* between $tRNA^{Thr}$ and $tRNA^{Pro}$ genes is the main noncoding region.

unique codon-anticodon recognition pattern. The $tRNA^{Ser}{}_{AGN}$ (which as in mammals lacks the D-arm) is able to recognize both AGY and AGR codons, although its anticodon (GCU) does not initiate with U, as occurs in those mt tRNAs which recognize four codon families. $tRNA^{Met}$ and $tRNA^{Lys}$, which bind only one type of codon (AUG and AAG), are the only tRNAs having a C at the first anticodon position. This could be one of the reasons why the recognition of the AUA codon as methionine and the AAA codon as lysine is prevented. These triplets specify asparagine and isoleucine; their binding with the respective tRNAs probably occurs by interaction with modified bases. For example the $tRNA^{Asn}$ or $tRNA^{Ile}$ may have the anticodon sequence G*AU or G*UU where G* may be an inosine-like base, able to form a base pair with either C, U, or A. The determination of the nucleotide sequences of the most important tRNAs is crucial to verify these hypotheses.

Phe	UUU	208	(GAA)	Ser	UCU	94		Tyr	UAU	51	(GUA)
	UUC	109			UCC	70	(UGA)		UAC	60	
Leu	UUA	126	(UAA)		UCA	83		Ter	UAA	6	
	UUG	52			UCG	11			UAG	5	
Leu	CUU	111		Pro	CCU	39		His	CAU	22	(GUG)
	CUC	52	(UAG)		CCC	60	(UGG)		CAC	56	
	CUA	197			CCA	66		Gln	CAA	70	(UUG)
	CUG	52			CCG	9			CAG	14	
Ile	AUU	146		Thr	ACU	65		Asn	AAU	44	
	AUC	53	(GAU)		ACC	79	(UGU)		AAC	50	(GUU)
	AUA	163			ACA	94			AAA	88	
Met	AUG	102	(CAU)		ACG	12		Lys	AAG	54	(CUU)
Val	GUU	69		Ala	GCU	75		Asp	GAU	25	(GUC)
	GUC	48	(UAC)		GCC	102	(UGC)		GAC	41	
	GUA	98			GCA	107		Glu	GAA	72	(UUC)
	GUG	26			GCG	24			GAG	31	

Cys	UGU	9	(GCA)				
	UGC	21					
Trp	UGA	83	(UCA)				
	UGG	23					
Arg	CGU	7					
	CGC	1	(UCG)				
	CGA	54					
	CGG	13					
Ser	AGU	11					
	AGC	14	(GCU)				
	AGA	66					
	AGG	15					
Gly	GGU	25					
	GGC	32	(UCC)				
	GGA	128					
	GGG	39					

FIG. 4. **Codon usage in the 13 protein genes of the *P. lividus* mtDNA.** The total number of codons in the *P. lividus* protein genes is 3831 including 6 UAA and 5 UAG codons. The anticodons corresponding to each codon family are shown in *parentheses*.

TABLE I

Characteristics of P. lividus ribosomal and protein mitochondrial genes

Gene	Size[a]	% homology with human[b]		Initiation codon[c]	Termination codon[d]
		Nucleotide	Amino acid		
	nucleotides				
16 S	1549	46.9			
12 S	883	57.7			
CoI	1551	68.3	75.8	AUG	UAA
CoII	687	62.7	59.1	AUG	UAA
CoIII	780	64.8	67.8	AUU/AUG	UAA
ATPase 6	696	46.1	40.3	AUG	UAA
ATPase 8	162	28.5	34.5	GUG	UAA
ND1	969	55.0	55.2	AUG	UAA
ND2	1056	44.6	36.2	AUG	UAG
ND3	348	51.7	47.0	AUA/AUG	UAG
ND4	1389	50.0	44.4	AUA/AUG	UAG
ND4L	291	48.3	36.7	AUC?	UA(A)
ND5	1914	42.9	39.9	AUG	UAG
ND6	480	32.9	38.6	AUG	UAG
Cytochrome *b*	1140	59.4	60.4	AUG	UA(G)

[a] Excluded the nucleotides involved in the stop codon.

[b] The human data are from Anderson *et al.* (1981). The percentage of homology has been derived using a computer program developed in our laboratory, which gives the best alignment between two sequences.

[c] In the case of CoIII, ND3, and ND4 the two possible initiation codons are reported (see also the text).

[d] The nucleotide in parentheses represents the first base of the following gene.

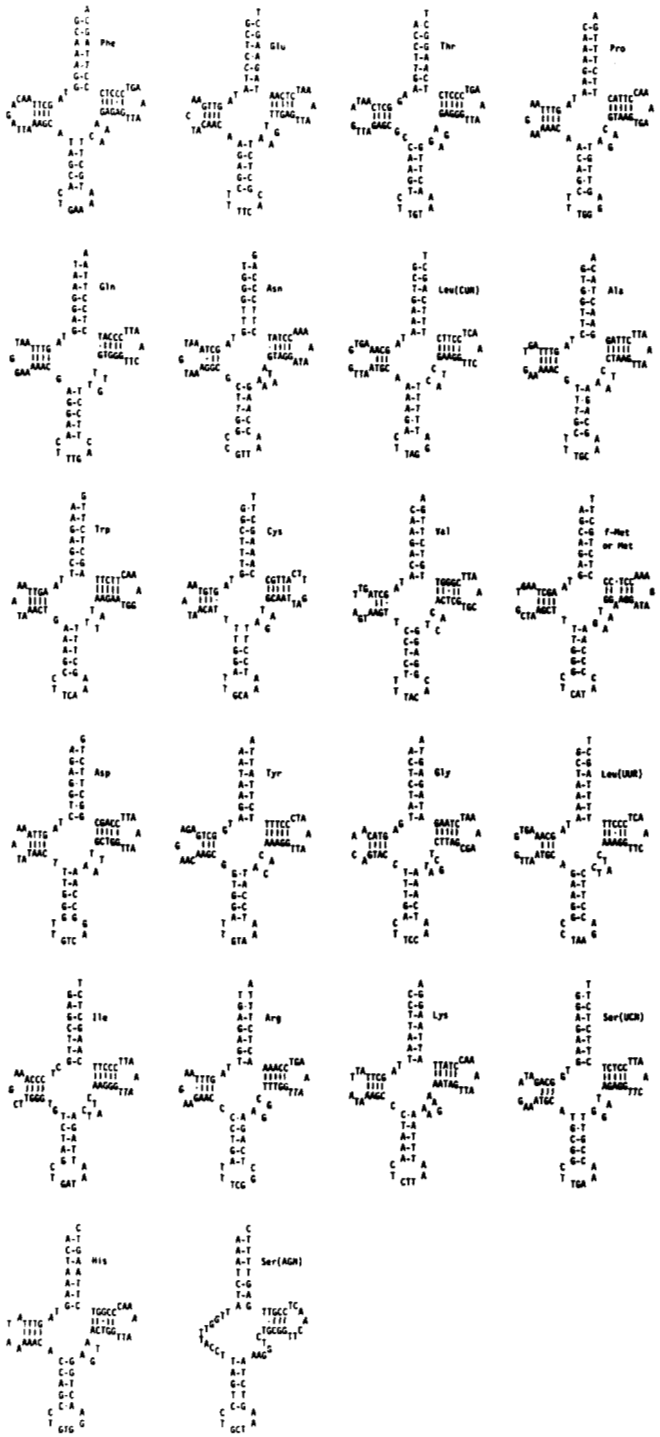

FIG. 5. **Sequence of *P. lividus* mt tRNA genes represented in the cloverleaf form.** A Watson-Crick base pair in the stem is indicated by a *dash*. A *dot* marks G·T or C·A base pairs. In the case of the $tRNA^{Arg}$ an alternative configuration of the amino acid stem, with a stability similar to that reported in the figure, may be constructed. The stem should have the following structure:

5′ TG AGAGT.....
3′TATTTCTCA.....

In this case (see also the text) the 3′ end of the $tRNA^{Arg}$ would be jointed at the ends to the ND4L initiation codon.

The comparison of the organization of mt tRNA genes in invertebrate genomes shows a much higher variability compared to that observed for rRNA or protein genes. While a few intergenic rearrangements are sufficient to explain the different positions of the rRNA and the mRNA genes of vertebrates and invertebrates (Cantatore *et al.*, 1987b), it is very difficult to find a single explanation for the variability in position of the tRNA genes. It has been recently demonstrated (Cantatore *et al.*, 1987b) that the sea urchin mt tRNAs may evolve by gene duplication; this is the case for the $tRNA^{Leu}{}_{CUN}$ which arose by duplication and subsequent divergence from the $tRNA^{Leu}{}_{UUR}$ gene. Moreover, tRNAs or tRNA like structures have been found near the origin of mtDNA replication in several systems, including mammals (Brown *et al.*, 1986). If tRNAs act as primers for mtDNA synthesis and sometimes fail to be removed, they may function as templates in the next round of replication and become incorporated at the replication origin. These observations and the recent finding of tRNAs at the end of duplications in animal mt DNAs (Moritz and Brown, 1987), support the view that tRNA genes could represent mobilization elements in the rearrangement of mt genomes and that they may evolve under a different evolutionary pressure from that operating on mRNA and rRNA genes.

The Noncoding Nucleotides—The *P. lividus* mtDNA contains 255 noncoding base pairs, 132 of which are located between the $tRNA^{Thr}$ and the $tRNA^{Pro}$ genes. The remainder, which in most cases has a high AT content, is spread over the genome. The AT-rich sequences present in five locations (main noncoding region, nt 1,203–1,220; $tRNA^{Val}$-$tRNA^{Met}$ junction, nt 1,799–1,811; 3′ end of ND2, nt 4271–4283; $tRNA^{Ser}{}_{UCN}$-ND3 junction, nt 10,229–10,241; ND6-cytochrome *b* junction, nt 14,534–14,546) may be aligned generating the very conserved consensus sequence TATATATAA. This motif, which has also been found in similar position in *S. purpuratus* (Jacobs *et al.*, 1988) could represent a recognition signal for enzymes involved in transcription or processing.

The main noncoding sequence has a high GC content (51.5%) and may be divided into three domains. Going from 5′ to 3′ there is a polypyrimidine tract of 23 nt (bases 1122–1143) and then a run of 26 consecutive G (bases 1177–1202) which is immediately followed by an AT-rich segment of 18 nt (bases 1203–1220). Despite the considerable size difference compared to the corresponding mammalian noncoding sequence, we have been able to align the *P. lividus* sequence with the mouse D-loop region (Bibb *et al.*, 1981) and to identify three areas of homology with the conserved sequence blocks found in the mammals. These blocks, which are also conserved in other echinoids (Jacobs *et al.*, 1988), are situated in the opposite position from those in mammals (Fig. 6). Given the suggestion that the sea urchin arrangement arose from that present in a common ancestor (Cantatore *et al.*, 1987b), the rearrangement of the noncoding region might have occurred first by the insertion near the 3′ end of the 12 S rRNA, of the $tRNA^{Glu}$ and the $tRNA^{Thr}$, and then of the D-loop region and of the $tRNA^{Pro}$. The inversion of this last piece of DNA is also demonstrated by the change in polarity of the $tRNA^{Pro}$ gene.

The presence in the noncoding region of sequences homologous to those (conserved sequence blocks) which in mammals mark the transition from RNA to heavy strand DNA (Chang and Clayton 1985), probably means that this region contains the replication origin of *P. lividus* mtDNA. In the mammals the replication origins are situated one (O_H) in the D-loop and the other (O_L) in a T-rich loop of a hairpin structure located between the $tRNA^{Cys}$ and the $tRNA^{Asn}$ genes (Clayton, 1982). If the *P. lividus* mtDNA replicates according the same pathway existing in the mammals, the replication of the first strand might initiate from the main noncoding region,

```
CSB1
          16035                                   16056
mouse        T T C A T G C T T G T T A G A C A T A A A T
               *   *   *     * * * * *   *   * * * * * *
P. lividus   G T A A C G T G T G T T A T A T A T A A A T
            1230                                     1209

CSB2
          16083                                   16102
mouse        C C A A A C C C C C C A C C C C C T C C
             * *       * * * * * *   * * * * *   * *
P. lividus   C C C C C C C C C C C C C C C C C C C C
            1202                                     1182

CSB3
          16107                                   16126
mouse        A A T G C C A A A C C C C C A A A A A C - A
             * *   * * *     * * *     *       * *   *
P. lividus   A A C G C C C G A C C G - A C G G A C G A
            1154                                     1135
```

FIG. 6. **Conserved sequence block-like regions of the *P. lividus* mtDNA.** Comparison of conserved sequence blocks (*CSBs*) in *P. lividus* and mouse mtDNA. The *P. lividus* sequence has been complemented and inverted. An *asterisk* shows identical nucleotides. The numeration refers to the genomic sequences (*P. lividus*: see Fig. 2; mouse: see Bibb *et al.*, 1981).

whereas the second strand should duplicate only after the newly synthesized DNA has displaced a substantial portion of the parental template strand. Alternatively, and in agreement with the lack of any convincing stem-loop structure both around the 3′ end of the $tRNA^{Cys}$ gene and in the short unassigned sequences in the genome, the second origin could be situated in the noncoding region itself. The second strand should then be synthesized only after the first strand synthesis is complete. In this case a possible stem-and-loop structure would involve the polypyrimidine tract, base-paired with the G string located 35 nt downstream of it. Early electron microscopic analysis of *S. purpuratus* mtDNA extracted by immature oocytes (Matsumoto *et al.*, 1974) suggests a third possibility. According to these authors the second strand synthesis may start in multiple and perhaps random sites on the displaced strand after the passage of the replication fork.

The knowledge of the base sequence of *P. lividus* mtDNA has not only provided information about a novel gene organization, gene transcription, and genetic code but has also raised several interesting questions about the mtDNA gene expression in invertebrates. Further studies will be focused on the clarification of aspects such as: (*a*) mtDNA replication and transcription; (*b*) RNA processing; (*c*) codon-anticodon interactions concerning $tRNA^{Ser}_{AGN}$, $tRNA^{Asn}$, and $tRNA^{Ile}$; (*d*) determination of divergence times between *P. lividus* and other echinoids, such as *S. purpuratus*, *Strongylocentrotus franciscanus*, and *Arbacia lixula*, whose mtDNA sequence is under study in our own and other laboratories.

Acknowledgments—We wish to thank Dr. H. Jacobs (Glasgow, Scotland) for providing us with *S. purpuratus* mtDNA sequencing data, D. Munno for excellent technical assistance, and E. Lampugnani for typing the manuscript.

REFERENCES

Anderson, S., Bankier, A. T., Barrel, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R., and Young, I. G. (1981) *Nature* **280,** 457–465

Attardi, G. (1985) *Int. Rev. Cytol.* **93,** 93–145

Bibb, M. J., Van Etten, R. A., Wright, C. T., Walberg, M. W., and Clayton, D. A. (1981) *Cell* **26,** 167–180

Brown, G. G., Gadaleta, G., Pepe, G., Saccone, C., and Sbisá, E. (1986) *J. Mol. Biol.* **192,** 503–511

Cantatore, P., and Saccone, C. (1987) *Int. Rev. Cytol.* **108,** 149–208

Cantatore, P., Roberti, M., Morisco, P., Rainaldi, G., Gadaleta, M. N., and Saccone, C. (1987a) *Gene (Amst.)* **53,** 41–54

Cantatore, P., Gadaleta, M. N., Roberti, M., Saccone, C., and Wilson, A. C. (1987b) *Nature* **329,** 853–855

Chang, D. D., and Clayton, D. A. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 351–355

Chen, E. J., and Seeburg, P. H. (1985) *DNA (N. Y.)* **4,** 165–170

Clary, D. O., and Wolstenholme, D. R. (1985) *J. Mol. Evol.* **22,** 252–271

Clayton, D. A. (1982) *Cell* **28,** 693–705

Gadaleta, G., Pepe, G., De Candia, G., Quagliariello, C., Sbisá, E., and Saccone, C. (1989) *J. Mol. Evol.*, in press

Glotz, C., Zweib, C., Brimacombe, R., Edwards, K., and Kossel, H. (1981) *Nucleic Acids Res.* **9,** 3287–3306

Gouy, M., Gautier, C., Attimonelli, M., Lanave, C., and Di Paola, G. (1985) *CABIOS* **1,** 167–172

Gray, M. W., Sankoff, D., and Cedergreen, R. I. (1984) *Nucleic Acids Res.* **12,** 5837–5852

Jacobs, H. T., Elliott, D. J., Math, V. B., and Farquharson, A. (1988) *J. Mol. Biol.* **202,** 185–217

Matsumoto, L., Kasamatsu, H., Piko, L., and Vinograd, J. (1974) *J. Cell Biol.* **63,** 146–155

Moritz, C., and Brown, W. M. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 7183–7187

Ojala, D., Montoya, J., and Attardi, G. (1981) *Nature* **290,** 470–474

Roberti, M., Cantatore, P., Morisco, P., Rainaldi, G., Santarella, G., Gadaleta, M. N., and Saccone, C. (1984) *Boll. Soc. It. Biol. Sper.* **60,** 1325–1331

Roe, B. A., Ma, D. P., Wilson, R. K., and Wong, J. F. H. (1985) *J. Biol. Chem.* **260,** 9759–9774

Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467

Smith, A. B. (1984) in *Echinoid Paleobiology* (Allen, G., ed) Unwin, London

Stormo, G. D., Schneider, T. D., and Gold, L. M. (1982) *Nucleic Acids Res.* **10,** 2971–2996

Wolstenholme, D. R., Macfarlane, J. L., Okimoto, R., Clary, D. O., and Wahleithner, J. A. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 1324–1328

Zweib, C., Glotz, C., and Brimacombe, R. (1981) *Nucleic Acids Res.* **9,** 3621–3640