## Isolation and sequence of the gene for actin in Saccharomyces cerevisiae

(DNA sequence/intervening sequence/cloning)

RAY NG AND JOHN ABELSON

Department of Chemistry, University of California at San Diego, La Jolla, California 92093

Communicated by Norman Davidson, April 11, 1980

**ABSTRACT** The yeast Saccharomyces cerevisiae is known to contain the highly conserved and ubiquitous protein actin. We have used cloned actin sequences from Dictyostelium discoideum to identify and clone the actin gene in yeast. Hybridization to genomic fragments of yeast DNA suggest that there is a single actin gene in yeast. We have determined the nucleotide sequence of that gene and its flanking regions. The sequence of the gene reveals an intervening sequence of 309 base pairs in the coding sequences at the 5' end of the gene. The existence and location of the intervening sequence was verified by using the dideoxy chain termination technique to determine the sequence at the 5' terminus of the actin mRNA. The similarity of the splice junction sequences in this gene to those found in higher eukaryotes suggests that yeast must possess a similar splicing enzyme.

Actin, a ubiquitous protein in higher organisms, plays an essential role in cell motility and structure. Although the functions of actin are diverse, ranging from muscle contraction to chromosome movement, the structure of the 42,000-dalton protein is highly conserved, both among the various types of actins found in higher organisms and in evolution (1, 2). The conservation of size and properties of actin has made it possible to identify actin in a number of lower eukaryotes. For example, the actin of *Dictyostelium discoideum* has been studied extensively (3, 4); the amino acid sequence of *Physarum polycephalum* actin is known (5). Recently, two groups have isolated actin in baker's yeast, *Saccharomyces cerevisiae* (ref. 6; R. Scheckman, personal communication).

Because yeast is nonmotile, actin most likely functions in cytoskeletal integrity, in chromosome condensation, or in the process of budding (2, 7, 8). The highly developed genetics and molecular biology of yeast make this a promising system for studying the genetics of actin and the control of its expression

We have taken advantage of actin's highly conserved amino acid sequence to characterize and isolate the yeast actin gene. The *D. discoideum* actin gene hybridizes with a single *Eco*RI fragment in yeast, suggesting that yeast contains only one actin gene. We have cloned this gene and determined its nucleotide sequence. Comparative studies with the nucleotide sequence of yeast actin mRNA establish the existence of a 309-base-pair (bp) intervening sequence at the 5' end of the gene—the first chromosomal gene in yeast found to have an intervening sequence (9). Yeast tRNA genes contain intervening sequences (10, 11), and recent studies indicate that intervening sequences also exist in the yeast mitochondrial genes (12, 13). However, several yeast chromosomal genes that do not have intervening sequences have been characterized (14–16).

The splice junction sequences found in the yeast actin gene bear a strong resemblance to those found in the genes of other

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

eukaryotes (17, 18). This suggests that yeast contains an RNA splicing mechanism like that found in higher organisms. In comparison, there are approximately 17 actin genes in *D. discoideum* which apparently do not have intervening sequences (R. A. Firtel, personal communication). On the other hand, recent findings show that *Drosophila melanogaster* has six actin genes and that the two genes examined contain an intervening sequence at a similar position near the 5' end of the gene (ref. 19; N. Davidson, personal communication).

## MATERIALS AND METHODS

Construction of Plasmids Carrying the Yeast Actin Gene. To detect colonies bearing the yeast actin gene, we made use of the fact that actin is a highly conserved protein. An actin gene from D. discoideum had already been cloned by R. A. Firtel and his colleagues and was made available to us (4). Radioactive cRNA from the actin fragment of the plasmid pcDd actin B was used to screen a collection of colonies bearing yeast HindIII fragments inserted into the plasmid pBR313 (20). One of the colonies was positive, and the plasmid isolated from this clone is designated "pYact II." This plasmid proved to contain only a part of the yeast actin gene. Consequently, a yeast clone bank, consisting of 5000 colonies, in which yeast EcoRI DNA fragments had been inserted into the plasmid pBR322 was constructed. This collection of colonies was screened by the method of Grunstein and Hogness (21) with a radioactive cRNA produced by in vitro transcription of the 1.6-kilobase (kb) HindIII fragment from pYact II. One of the colonies in this collection hybridized and harbored a plasmid carrying a 3.8-kb EcoRI insert which contained the entire yeast actin gene. This plasmid is designated "pYact I."

DNA Preparation. Plasmid DNA was isolated by using a modification of the Clewell and Helinski procedure (22).

Yeast DNA was isolated from S. cerevisiae (2180-1A). Cells were grown at  $30^{\circ}$ C to  $A_{660}$  of 2–3 in YPD medium, and the DNA was extracted and purified by the method of Sherman and Fink (23), with Hoescht dye 33258 in place of 4′, 6-diamidine-2-phenylindole.

RNA Preparation. Total yeast RNA was isolated by using a modification of the Hereford and Rosbach procedure (24). Poly(A)-containing RNA was purified by chromatography of total RNA on poly(U)-Sepharose (4).

Nick Translation. DNA was made radioactive by the nick translation procedure using DNA polymerase I and  $[\alpha^{-32}P]$ -deoxynucleoside triphosphate as described (25).

DNA Sequence Determination. DNA was analyzed according to the method of Maxam and Gilbert (26). DNA copies of actin mRNA were analyzed by the method of Sures *et al.* (27) which was a modification of the Zimmern and Kaesberg procedure (28).

Abbreviations: bp, base pair(s); kb, kilobase(s).

## **RESULTS**

Yeast Contains a Single Actin Gene. In order to identify and quantitate the actin gene(s) in yeast, total yeast DNA was digested with either HindIII or EcoRI. The fragments were separated by electrophoresis in a 1.2% agarose gel. The DNA was denatured in situ and the fragments were transferred to nitrocellulose filter paper by using the methods developed by Southern (29). Fragments containing the actin sequence were identified by hybridization at 60°C to a nick translated, actin-specific probe from D. discoideum. Two HindIII fragments (approximate sizes, 3.7 and 2.5 kb) and a single EcoRI fragment (3.8 kb) hybridized with the actin probe. Additional experiments by Norma Neff (personal communication) showed that yeast BamHI and Pst I fragments hybridized to pYact I DNA, in agreement with our restriction map. As a result, it appears that yeast has a single actin gene, although the unlikely possibility of an identical repeat of the large DNA segment containing the actin gene cannot be ruled out.

In our first attempts to clone the yeast actin gene, a collection of colonies containing yeast *HindIII* fragments inserted into the plasmid pBR313 were screened with the *D. discoideum* probe. This collection contained the plasmid pYact II with the 2.5-kb *HindIII* fragment. Subsequently, an *EcoRI* library was constructed by ligation of yeast *EcoRI* fragments into the vector pBR322. We isolated a plasmid, pYact I, which contained the 3.8-kb *EcoRI* fragment. This fragment hybridized to the actin probe and was shown by DNA sequence analysis to contain the complete actin gene. Restriction endonuclease maps of the plasmids pYact I and pYact II are shown in Fig. 1.

Proof Through DNA Sequence Analysis that pYact I Contains the Actin Gene. The restriction endonuclease analysis of total yeast DNA suggested that there is a single actin gene which can be separated into two fragments by digestion with HindIII. DNA sequence analysis at the HindIII site in pYact I proved that this site is, in fact, located in the actin coding sequence. The DNA sequence in this region predicted an amino acid sequence closely homologous with the amino acid sequence of Physarum actin and with vertebrate actins as well (5, 31). The HindIII site was located at amino acid position 255. Further work led to a nearly complete sequence of the actin gene (Fig. 2).

An Intervening Sequence Separates the First Three Amino Acid Residues from the Rest of the Actin Coding Sequence. Several lines of evidence from the DNA sequence of the yeast actin gene suggested that an intervening sequence occurs in the gene between the third and fourth codons. The DNA sequence of the actin gene predicted an amino acid sequence similar to the Physarum sequence for most of the 374 residues (5). This similarity did not exist, however, at the NH2 terminus of the coding sequence and, most important, a methonine initiator codon in phase with the rest of the coding sequence was lacking. The DNA sequence predicted an NH2-terminal sequence of Ile-Cys-Leu, an amino acid sequence radically different from that of all known actins (31). Further analysis revealed a sequence 309 bp upstream from this region that contained a methonine initiator codon followed by the codons for Asp-Ser-Gly, an NH<sub>2</sub>-terminal sequence compatible with other actins. Although the sequences surrounding this proposed intervening sequence were homologous with the sequences at the splice junctions in higher eukaryotes, the existence of a 309-bp intervening sequence could only be considered plausible from this evidence. To define the limits of the intervening sequence requires a comparison of the gene and mRNA sequences.

Nucleotide Sequence of Actin mRNA in the Region of the Splice Junction Confirms the Existence of the Intervening Sequence. To determine the sequence of the region near the putative splice junction, we used the dideoxy chain-terminating primer-extension technique described by several groups (27, 32). This procedure obviates the need to purify the actin mRNA or to clone its cDNA. A 75-bp Hpa II/HinfI fragment (amino acid positions 22-50) close to the putative splice junction was specifically hybridized to the actin message in poly(A)RNA. Reverse transcriptase was then used to extend the primer toward the 5' end of the mRNA. If an intervening sequence indeed existed, one would expect the mRNA sequence to suddenly deviate from the DNA sequence at the splice point. This is exactly what we found. Using the known DNA sequence for comparison, we were able to deduce a vital portion of the mRNA sequence, despite the appearance of some extraneous bands (Fig. 3). This sequence is colinear with the gene until codon 3. Then, after a hiatus of 309 nucleotides, the proposed

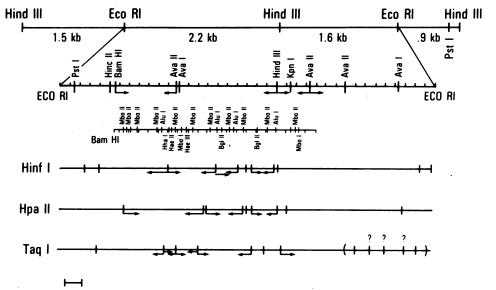


FIG. 1. Genomic organization of the EcoRI and HindIII sites based on hybridization experiments is shown on the top line. The 3.8-kb EcoRI ragment was mapped by the method of Smith and Birnstiel (30). The locations of some of the less-common restriction endonuclease sites appear in the second line; a map of sites derived by DNA sequence determination is given below it. A detailed restriction map for the enzymes Hpa I, HinfI, and Taq I is shown. Arrows indicate the direction and extent of sequence analysis at each site. Scale represents 0.1 kb.

THE COLUMN TAY AND AND A 100 SO SO SO TO 80 SO																							•		,
110 120 130 130 140 150 150 150 160 170 180 190 200 200 201 201 201 201 201 201 201 20			10	20		30			40		5	0		60			<b>7</b> 0			80		9	0		100
GARGARAMA ANA MATCHA TCTITOTTA ADDATAGORT CITCHACTAC ATCACCTIT AGAITITEA COCTACTOC TITTITCTC CCARGATCO AMATTHACTO  MET ASP SER 230 240 250 260 270 280 290 300 310 310 320  AATTHACA ATG OAT ICT 6 GTATOTTCACC GCTICCACCA TCCCATTAA CCTARAMA ANTICACACCO TCCCAATTGC TOGAGAGAT TCTCTTTTAC CTTTTTAC  330 340 350 360 370 380 390 500 510 520 530  TATOTTCACTATA CCCCCATAA CCCCCCATAA CACCCCATATA TOACCGATCT GATAACAC CACATATAT GGAATAAATA GCGCCATGAA ATTTCCATAAA AAAAAAAAAA		CTCTCTTG	IAAT TI	TATATAT	TCAT	TATAT	CACG	CTCT	ст т	ATTT	TCTT	C CT	TTTT	TTCC	TCT	CTCT	TGT	ATTC	TTCC	TT (	CCCT	TTCT	A CT	CAAA	CCAA
NET ASP SER 230 240 250 260 270 280 290 300 310 320  ANTHRACA ATO DAT ICIT G GRAUTICTAGG GCTTGCACCA TCCCANTRAC CTGTAGAAGA BATTGCACCO TCCCANTGC TCGACAGATT TCTCTTTTAC CTTTTTTAC  330 340 350 350 300 370 380 390 400 410 420 330  TATTTTCACC CTCCCCATAC CCCCCTACT TGACTCACCA TCCCANTRAC TGGACTACT TGACTACACA CGCATTATTT TGACTCATTAC CTCCCCATAC CCCCCATACA TTGCACCACACACACACACACACACACACACACACACACA				-																					
ANTIMACA ATTO DAT ICT 0 STATUTICTACC GCTTOCACCA TOCCANTRAC CIGTAGAAGA AND ANTIGCACCO TOCCANTGC TCGAGAGATT TCCTTTTAC CTTTTTTAC  330 340 350 360 370 380 390 400 410 420 430  TATTITICAC CTCCCCATAA CCTCCTATAT TGACTCATCT GTATAACCA CGATATTATT GGATAAATA GOGGCTTGAA ATTTGGAAAA AAAAAAAAAAAAACA CTCAAATATT  440 450 450 460 470 480 490 500 510 520 530  TTCGTGATAA GTGATAGTGA TATTCTTCTT TTATTCCTA CTGTTACTAA GTCTCATOTA CTAACACTGA TTGCTCATT CTTTTTTTTC CTATATATA TOTTTAC  AG GTT GCT GCT TTG GTT ATT GAT AAC GCT TCT GGT ATC TCT GAT AAA GCC GCT TTT GCC GCT GAC GAC GCT CCT GCT GCT GTT CCA TCT GCT VAL ALA ALA LEU VAL ILE NSP ASS GLY SER GLY SER GLY SER GLY MET CTS LIS ALA GLY HE ALA GLY ASP ASP ALA PRO ANG ALA VAL PHE PRO SER  ATC GCT GCT GAC GAC CAA GAC CAA GAT CAT GT GCT GCT ATC ATC GCT AAA GCC GCT TTT GCC GCT GAC GAC GCT CCT GCT GCT CAT CAA GCA CAC CAA GAC ACC CAA GCA TCA TCA GCT ACC ACC GAC GCT CCT GCT GCT GCT GCT CAT GAC ACC ACC GCT CCT GCT GCT GCT GCT GCT GCT GCT G	GAAGAAAAG	AAAAGGTC	A TCTI	TGTTAA	AGAA	TAGGAT	CTTC	TACT	AC A	TCAG	CTTT	T AG	ATTT	TTCA	CGC	TTAC	TGC	TTTT	TTCT	TC (	CAAG	ATCG	A AA	ATTT	ACTG
ANTITIACA ATTO GAT TOT G GYNTOTICTICS COTTOCACCA TECCATITA CTGTARGAG ANTICACOG TECCANTIC TOCACAGAIT TOCOTTIAC CTGTTTITAC CTTTITAC CTTTITAC CTTTITAC CTTTITAC CTGTARGAGC COLOR AND	ME	T ACD CP	230		240		250		26	0		270	)		280		2	90		30	00		310	)	320
THATTITICAC TOTOCCATAM COTOCCTATAN TOACTGATCT GTAATAACCA COATATIATT GGAATAAATA GGGGCTTGAA ATTTGGAAAA AAAAAAAAAA	AATTAACA AT	G GAT TC	G GTA	TGTTCT	AGC G	CTTGCA	CCA 1	CCCA	TTTA	A CT	GTAA	GAAG	AAT	TGCA	CGG	TCCC	AATT	GC T	CGAG	AGA1	TT TC	TCTT	TTAC	CTT	TTTTTAC
THE GLU THE LEU THE LEU AND THE PRO LEG CU DITS CITS CAT CAS ACT CAS A	<b>3</b> 30	34	10	350		360		3	70		38	0		390			400		4	10		42	0		430
TOCOTCATA GEGATACTCA TATTCTCCTT TTATTCCCA CTOTTACTA CTCTCATCA CACACGGA TOCCTCATC CTTTTTGCTG CTATATTATA TOTTTAG  AC GTT GCT GCT TTG GTT ATG GAT AAC GGT TCT GGT ATG TGT AAA GCC GGT TTT GCC GGT GAC GAC GCT CCT CGT GCT GTC TTC CCA CCT CLU VAL ALA ALA LEU VAL ILE ASP ASN GLY SER GLY MET CTS LYS ALA GLY PHE ALA GLY ASP ASP ALA PRO ARG ALA VAL PHE PRO SER  AC GTC GCT AGA CCA AGA CAC CAA GGT ATC ATG GTC GGT ATG GGT CAA AAA GAC CCC TTT GGT AGA CAC CAA GGT ATC ATG GTC AGT AGA GAC CCA AGA CAC CAA GGT ATC ATG GTC ATG GGT CAA AAA GAC CCC AGA GAC CCA AGA CAC CAA GGT ATC ATG GTC AGT ACC CAA GC CAA AAA GAC CCC AAA GCC CAA TCC AGA GAA GGT CTT TC ACC AAC TGC GAC GAT TCC TTT CAC AAC GAA AGA ATC TGC CAT CAT ACC ATT ACC CAC ACT GCT GAT TCC ATT ACC ACC AAC TCC GAC GAC GAT ATC GAA AAA AGA ATC TGC CAT ACC ATT ACC CAC ACT ACC ATC GAC GAC GAT ACC CAC ACT ACC TTC TAC CAC AAC TCC GAC GAT GAC CTT TTA CAC AAC TAC CAC ACC GAT TCC TAC GAC GAC GAT ACC CAA ATC AAC CCC AAT ACC ATC ACC AC	TATTTTTCAC	TCTCCCAT	A CCTC	CTATAT	TGAC	TGATCT	GTA	TAAC	CA C	GATA	TATT	T GG	AATA	AATA	GGG	GCTT	GAA	ATTT	GGAA	AA I	***	AAAA	A CT	GAAA	TATT
AG GTT OCT OCT TITG GTT ATT GAT AAC GGT TCT GGT ATG TGT AAA GCC GGT TTT GCC GGT GAC GAC GCT CCT CGT GCT GTC TTC CAX TCT GU VAL ALA ALA LEU VAL ILE ASP ASN GLY SER GLY MET CIS LYS ALA GLY THE ALA GLY ASP ASP ALA PRO ARG ALA VAL PHE PRO SER CLY VAL GLY ARG PRO ARG HIS GLA GGT ATC ATG GTC GAT AGG GGT CAA AAC GAC GGT AGA CCA AGG CAA GAC CAA GGT ATC ATG GT CAT AGG GGT CAA AAC GGT CTAC GTT GGT GAT GAA GGT CAA TCC AAG GAG GGT LIE WAL GLY ARG PRO ARG HIS GLA GLY ILE MET VAL GLY NET GLY GLU LYS ASP SER TYR VAL GLY ASP GLU ALA GLA SER LYS ARG GLY ATC TTG ACT TTA CGT TAC CAA TC GAC GGT ATT GAA CAC CTT GTT CTT TTT GAT GGT ATT GTT GAT GAA AAC CCT GTT GAA AAC CCT GAA AAA ATT GAA CAC GGT ATT AGT GLU HIS FRO VAL LEU LEU THR GLU ALA PRO MET ASN FRO LIS ASN ARG LIS THR GLI LEU HTH GLU ALA PRO MET ASN FRO LIS ASN ARG LIS THR GLI LEU HTH GLU ALA PRO MET ASN FRO LIS ASN ARG LIS THR GLT LIS TO THE GLU HIS PHE STR LIE GLY ALA CLU LEU SER LEU LYS SER SER GLY ARG THR THR GLY LIE VAL LEU GLY THR PHE ASN VAL PRO ALA PHE TIR VAL SER ILE GIA ALA VAL LEU SER LEU TYR GAG ACT ACT GGT AGA ACT ACT GGT ATT GT GAG AGA SEY SER SER GLY AND GLY VAL THR HIS VAL VAL PRO ILE TYR ALG GLY GAT GAT GAC GAC ATT TTA GAG CTC GAT TTG GCT GAA GAA AGA CTC CC GGT AGA ACT ACT GGT AGA AGA SEY SER SER GLY AND GLY VAL THR HIS VAL VAL PRO LIE TYR ALG GLY GAC GAT GAT GAC GAC ACT GCT GAT GAA GAA AGA AGA ATT GCT GAA GAA AGA CTC GCT GAA GAA CTC	440	49	50	460		470		4	80		49	0		500	ı		510		5	20		53	0		
AG GTT GCT GCT TITG GTT ATT GAT AAC GGT TCT GGT AAT GTT AAA GCC GGT TTT GCC GGT AAC GAC GCT CCT CGT GCT TCT CAC CAT CUL VAL ALA ALA LEU VAL ILE ASP ASN GLY SER GLY HET CIS LYS ALA GLY PHE ALA GLY ASP ASP ALA PRO ARG ALA VAL PHE PRO SER  ATC GTC GGT AGA CAC CAA GGT ATC ATC GTC GGT ATG GGT CAA AAA GAC TCC TAC GTT GGT GAT GAA CAC CAA GGA CAC CAA GGT ATC ATC GTC GGT ATG GGT CAA AAA GAC TCC TAC GTT GGT GAT GAA CAC CAA GAG AGG GTT CAT ACC GTT GGT GAT GAG CAC CAC ATC AGA GAG GGT CAT ACC CAC GCT GAT CAC CAC GGT ATT GAA CAC GGT ATT GCA CAC CAC GGT ATT GCA AAA GAG ACT CAT ACC CAC ATC TAC CAC AAC TCA ACC AAC GGT ATT GAA CAC CAC GGT ATT GCA CAC CAC GGT ATT GAA ACC CAC ATC TAC CAC ACC GGT ATT GCA AAA AAA CCC CAT GAT CAC AAC GAG GAA AAA ATG GAC CAC CAT GTT CAC CAC ACC CAC GAA GAA CAC CAC GAT GAT CAC CAC CAC GAA GAA CAC CAC GGT ATT GAT CAC CAC ACC ATC TAC CAC ACC ATC TAC CAC ACC ATC TAC CAC ACC CAC ATC AAC CAC ACC AC	TTCGTGATAA	GTGATAGT	GA TATI	CTTCTT	TTAT	TTGCTA	CTG1	TACT	AA G	TCTC	ATGT	A CI	TAACA	TCGA	TTG	CTTC	TTA	CTTI	TTGT	TG (	CTATA	TTAT	A TO	TTT	IG.
CLU VAL ALA ALA LEU VAL ILE ASP ASN GLY SER GLY NET CYS LYS ALA GLY PHE PLA GLY ASP ASP ALA PRO ARG ALA VAL PHE PRO SER  ATC GTC GGT AGA CCA AGA CCA CAA GGT ATC ATG GTC GGT ATG GGT CAA AAA GAC TCC TAC GTT GGT GAT GAA GCT CAA TCC AAC AGA GAG GGT ILE VAL GLY ARG PRO ARG HIS GLW GLY ILE MET VAL GLY MET GLY GLN LYS ASP SER TIR VAL GLY ASP GLU ALA GLM SER LYS ARG GLY  ATC TTG ACT TTA CCT TAC CCA ATT GAA CAC GGT ATT GTC ACC AAC TGG GAC GAT ATG GAA AAA ATG CTC AT ACC TTC TAC AAC GAA ILE LEU THR LEU ARG TYR PRO ILE GLU HIS GLY ILE VAL THR ASN TRR ASP ASP MET GLU LYS ILE TRP HIS HIS THR PHE TYR ASH GLU  TTG AGA GTT GCC CCA GAA GAA CAC CCT GTT CTT TTG ACT GAA GCT CCA ATG AAC CT AAA ATCX AAC AGA GXA AAA XTG ACT CAA ATT ATG LEU ARG VAL ALA PRO GLU GLU HIS PRO VAL LEU LEU THR GLU ALA PRO MET ASN PRO LYS  ASN ARG LYS  THR GLM ILE MET  TIT GAA ACT TTC AAC GTT CCA GCC TTC TAC GTT TCC ATC CAA GCC GTT TTG TCC TTC TAC GTT ACC TTC TAC GTT TTC TAC TTT TCC ATC AGA GCT TAT TTG TCC TTC TAC GTT ACT CAA GCT TTT TTC  PHE GLU THR PHE ASN VAL PRO ALA PHE TIR VAL SER ILE GLW ALA VAL LEU SER LEU TIR SER SER GLY ARG THR THR GLT ILE VAL LEU  GAT TCC GGT GAT GGT GTT ACT CAC GTC GTT CCA ATT TAC GCT GGT TTC TCT CTC TCC CAC GCC ATT TTG AGA CTC GAT TTG GCC GGT AGA ASP SER GLY ASP GLY VAL THR HIS VAL VAL PRO ILE TIR ALA GLY PRE SER ILEU PRO HIS ALA ILE LEU ARG LEU ASP LEU ALA GLY ARG  GAT TIG ACT GAC TAC TTG ATG AGG ATC ACT GAA GAG ATG CAT TAC TCT TCC TCC ACC ACT GCT GAA ACA GAA ATT GTC GCT GAC ATC AAC  GLU LYS LEU CYS TIR VAL ALA LEU ASP PHE GLU ARG GLY TRY SER PHE SER THR THR ALA GLU ARG GLU ILE VAL ARG ASP ILE LYS  GAA AAA CTA TGT TAC GTC GCC TTG GAC TTC GAA CAA GAA ATG CAA ACC GCT GCT CAA TTC TTC TCG ACC ACT GCT GAA CAA ATT GTC GCT GAA CTC GAC  GLU LYS LEU CYS TYR VAL ALA LEU ASP PHE GLU GLY GLU HET GLU RET GLY GLU FRE HE HE COLL LYS SER TYR GLU LEU PRO  GAT TIG ACT GAC TAC ACT ACT TG GAC CTC GAA AGA ATC CAC GAC ACT GCT TTT GAA GAA TCC TCC GAC GTC CAC  GAT GAC CAA GTC ATC ACT ACT GAC TCC GAC AACA GAA ATC ACC GCT GTC CAC GAC ACT GCT TTG GGT TCC G	AG GTT GC1	GCT TTG	GTT AT		AAC G	GT TC1	GGT	ATG	TGT	AAA	GCC	GGT		GCC	GGT	GAC	GAC	GCT	CCT	CGT	GCT	GTC		CCA	TCT
ATC GCT GGT AGA CCA GAG CAC AGA GGT ATC ATG GTT CAT GGT CAA AAA GAC TCC TAA CTT GAT CAA CCT CAA TCC AAG GGT LILE VAL GLY ARG PRO ARG HIS GLN GLY ILE MET VAL GLY MET GLY GLN LTS ASP SER TYR VAL GLY ASP GLU ALA GLN SER LTS ARG GLY  TO  ATC TIG ACT TAA CCT TAA CCA ATT GAA CAC GGT ATT GTC ACC AAC TGG GAC GAT ATG GAA AAG ATC TGG CAT CAT ACT TAC TAC TCT TAC AAC GAA ILE LEU THR LEU ARG TYR PRO ILE GLU HIS GLY ILE VAL THR ASN TRP ASP ASP MET GLU LTS ILE TRP HIS HIS THR PHE TYR ASN GLU  TIG AGA GTT GCC CCA GAA GAA CAC CCT GTT CTT TG ACT GAA GCT CCA ATG AAC CCT AAA TCX AAC AGA GAA AAX TG ACT CAA ATT ATG LEU AARG VAL ALA PRO GLU GLU HIS PRO VAL LEU LEU THR GLU ALA PRO MET ASN PRO LTS  TIT GAA ACT TTC AAC GTT CCA GGC TTC TAC GTT TCC ATC CAA GCC GTT TTG TCC TTG TAC TCT TCC GGT AGA ACT ACT GGT ATT GTT TTG PHE GLU THR PHE ASN VAL PRO ALA PHE TYR VAL SER ILE GLN ALA VAL LEU SER GLU TYR SER SER GLY ARG THR THR GLY ILE VAL LEU  GAT TCC GGT GAT GGT GTT ACT CAC GTT CCA ATT TAC GGT GGT TTC TCT CTA CCT CA GCC ATT TTG AGA CTC GAT TTG GGC GT AGA ASP SER GLY ASP GLY VAL THR HIS VAL VAL PRO ILE TYR ALA GLY PHE SER LEU PTO HIS ALA ILE LEU ARG LEU ASP LEU ALA GLY ARG  GAA AAA CTA TGT TAC GTC GAC TTG GAC TTG GAC GAA GAG GGT TAC TCT TCT CC ACC ACT GGT GAA AGA GAT GTC GGT TGT GAC TCC ATT TGA GAC CTC GAT TTG GAC ACC AGT GAT AGA ASP LEU THR ASP TYR LEU MET LYS ILE LEU SER GLU ARG GLY TRY SER PHE SER THR THR HALA LOL ANG GLU ILE VAL ARG ASP ILE LTS  GAA AAA CTA TGT TAC GTC GCC TTG GAC TTC GAA CAA GAA ATG CAA GCC GTC GTT TCT TCT CC CAC CAC GCT GTT GAA AGA CAT CT GGT AGA CTC CAC GCT GTT GAC TCC ATT TGT GAC GCC GTT GAA CTC GAA CTC GAT GTT GAC CTC GAA CTC GAT CTC GAA CTC GAC ACC GCT GTT GAA CTC GAA CTC GAC ACC GCC ATT TTG GAC TCC GAT CTC GAA CTC GAC ACC GCC GAA CTC GCT GAA AGA CAT CTC GTG GAT CAA ASP LEU THR ASP TYR LEU HAT ALE LEU SER GLU ARG GLY TRY SER PHE SER THR THR ALLA GLU ARG GLU LANG GLU LANG GRU LANG ASP ILE LTS  GAA AAA CTA GTT TAC GTC GAC TTG GAC TTC GAA CAA GAA ATG CAA CAC GCT GTT GCC TAA GCT CAA TTT GAA CTC TCT GAA CTT ACT TTG TAC TCT																									
ILE VAL GLY ARG PRO ARG HIS GLN GLY ILE MET VAL GLY MET GLY GLN LYS ASP SER TYR VAL GLY ASP GLU ALA GLN SER LYS ARG GLY  TO  ATC TIG ACT TTA COT TAC CCA ATT GAA CAC GOT ATT GTC ACC AAC TOG GAC GAT ATG GAA AAG ATC TOG CAT CAT ACC TTC TAC AAC GAA ILE LEU THR LEU ARG TYR PRO ILE GLU HIS GLY ILE VAL THR ASN TRP ASP ASP MET GLU LYS ILE TRP HIS HIS THR PHE TYR ASN GLU  TIG AGA GIT GCC CCA GAA GAA CAC CCT GIT CIT TIG ACT GAA GCT CCA ATG AAC CCT AAA TCX AAC AGA GXA AAA XTG ACT CAA ATT ATG LEU ANG VAL ALA PRO GLU GLU HIS PRO VAL LEU LEU THR GLU ALA PRO MET ASN PRO LYS  ASN ARG  LYS  THR GLN ILE MET  TIT GAA ACT TTC AAC GIT CCA GCC ITC TAC GIT CCA ATG CAC CCG GIT TTG TCC TTG TAC TCT TCC GGT AGA ACT ACT GGT ATT GIT TTG PHE GLU THR PHE ASN VAL PRO ALA PHE TYR VAL SER ILE GLN ALA VAL LEU SER LEU TYR SER SER GLY ARG THR THR GLY ILE VAL LEU  GAT TCC GOT GAT GOT GIT ACT CAC GCC GIT CCA ATT TAC GCT GGT TTC TCT CTA CCT CAC GCC ATT TTG AGA CCT GAT TTG OCC GOT AGA ASP SER GLY ASP GLY VAL THR HIS VAL VAL PRO ILE TYR ALA GLY PHE SER LEU PRO HIS ALA ILE LEU ARG LEU ASP LEU ALA GLY ARG  GAA AMA CTA GOT GAT GOT GTC ATG AGA CAT CTC GAA CAT GAG GCT TTC TCT CTA CCT CAC GCC ATT TTG AGA CAT ACT GGT GAA ASP LEU THR ASP TYR LEU MET LYS ILE LEU SER GLU ARG GLY TYR SER PHE SER THR THR ALA GLU ANG GLU ILE VAL ARG ASP ILE LYS  GAA AMA CTA TOT TAC GTC GCC CTO GAC TTC GAA CAG GAA ATC GAT GAT CAA GCT GCT GAA ACC GCT GCT CAA TCT TCT TCT CAC ACT GCT GAA AGA GAA ATT GCC GGT GAA CTT CAC GCU LYS LEU CTS TYR VAL ALA LEU ASP PHE GLU GLN GLU ARG GLY TYR SER PHE SER THR THR ALA GLU ANG GLU ILE VAL ARG ASP ILE LYS  GAA AMA CTA TOT TAC GTC GCC TTO GAC TTC GAA CAG GAA ATC GCA GAT GCT CAA TCT TCT TTG GAA CAT GAT GAT GAA CAT CAT GTC GAA CAT CAT GTT GAA CAT ACT GT AAC GAA ACC GCT GTC CAA TCT TCT TTT GAA ACT ACT GAA CAT CAT GTC GAA CTT CAT GAT GAA CAA CAT GTC GAA GAA ACT CAT TTG GAA CTA CAT GAA CAA GAA TCT GCT GAA CTA CAT GTT GAA CAA ACA GAA ATC GAA ACC GCT GTC TCT TTT TTT TTT GAA CAC ACT ACT GAT GAA CAA ACT TCT GAA CAA ACA GAA ATC GAA ACC GCT GTC GAA CAC ACT CTT T														<b></b>			COT			000		<b>**</b> **********************************			007
ATC TIGA COT TAG COT ACT CAG CAG CATT CAA CAG CAGT ATT GTG ACC AAC TAG CAG ACT ATT GAA AAG ATC TOG CAT CAT ACC TIC TAC CAG CAGA ILE LEU THR LEU AND TYR PRO ILE GLU HIS GLY ILE VAL THR ASN TRY ASP ASP MET GLU LYS ILE TRY HIS HIS THR PHE TYR ASN GLU  TIG AGA GTT CAC CAG GAA GAA CAC CCT GTT CTT TTG ACT GAA GCT CCA ATG AAC CCT AAA TCX AAC AGA GXA AAA XTG ACT CAA ATT ATG LEU AND VAL ALA PRO GLU GLU HIS PRO VAL LEU LEU THR GLU ALA PRO MET ASN PRO LIS ASN AND CLTS THR GLN ILE MET  TIT GAA ACT TTC AAC GTT CAC GCC TTC TAC GTT TCC ATC CAA GCC GTT TTG TCC TTG TAC TCT TCC GGT AGA ACT ACT GGT ATT GTT TTG PHE GLU THR PHE ASN VAL PRO ALA PHE TYR VAL SER ILE GLN ALA VAL LEU SER LEU TYR SER SER GLY AND THR THR GLY ILE VAL LEU  GAT TCC GGT GAT GGT GTT ACT CAC GTC GTT CCA ATT TAC GCT GGT TTC TCT CTA CCT CAC GCC ATT TTG AGA CTC GAT TTG GCC GGT AGA ASP SER GLY ASP GLY VAL THR HIS VAL VAL PRO ILE TYR ALA GLY PHE SER LEU PRO HIS ALA ILE LEU AND LEU ASP LEU ALA GLY AND  GAT TTG ACT GAC TAC TTG ATG AAG ATC TTG AGT GAA CGT GGT TAC TCT TCT TCC CAC CAC CAT GCT GAA AGA AAT GTC CCT GAC ATC AAG ASP LEU THR ASP TYR LEU NET LYS ILE LEU SER GLU ANG GLY TYR SER PHE SER THR THR ALA GLU AND GLU ILE VAL AND ASP ILE LYS  GAA AAA CTA TUT TAC GTC GCC TTG GAC TTC GAA CAA GAA ATG CAA ACC GCT GCT CAA TCT TCT TCT CAT ACT TCT TCA ATT GAT AAA ACT TUT TAC GTC GCC TTG GAC TTC GAA GAA GAA TTC AGA GCC CCA GAA GCT TTC TCA ACT TCT TTT TCA ATT GAT AAA ACT TUT TAC GTC GCC TTG GAA CAA GAA ATG CAA GAA ATG CAA ACC ACT GCT GAT TCT TCT ACT TCT TCA ATT GAT TOT TAC GCC GCT GAC TCC GAC TTC GAA GAA ATG CAA ACA GAA ATG CAA ACC ACT GCT CAA TCT TCT TCA ATT GAT TGA GCC GCC TTG GAC TTC GAC TTC GAC TTC GAC TTC GAC TTC GCC GTC GAA GCT TCT TCC CAT CCT TCT GTT TTC GCT TCT GTT TTC GCT TCC GCT GAC TCC GCC GTC AAC GCT GAA GCT ATC ACT TTC GCC GTG TAC CTT GAA GAA GAT ATC GAC GCC GCA GAA GCT TCT TCC CAT CCT TCT GTT TTC GCT TCT GTT TTC GCC GTG GAC TCC GCC GTG GAC GCC GAA GCT TCT TCC CAT CAT TTC GCT TCT TCC GCT GCT TCC CCC GCT TCC CCC GCT TCC	ILE VAL GLY	AGA CCA ARG PRO	AGA CI	AC CAA IS GLN	GGT A	LE MET	VAL	GLY	MET	GLY	GLN	LYS	ASP	SER	TYR	VAL	GLY	ASP	GLU	ALA	GLN	SER	LYS	ARG	GLY
THE LEU THR LEU ARG TYR PRO ILE GLU HIS GLY ILE VAL THR ASN TRP ASP ASP MET GLU LYS ILE TRP HIS HIS THR PHE TYR ASN GLU  TO 100  TO 110  TO 110  TO 110  TO AGA GTT GCC CCA GAA GAA CAC CCT GTT CTT TTG ACT GAA GGT CCA ATA AAC CCT AAA TCX AAC AGA GAA AAA YTG ACT CAA ATT ATG LEU ANG VAL ALA PRO GLU GLU HIS PRO VAL LEU LEU THR GLU ALA PRO MET ASN PRO LYS ASN ARG LYS THR GLN ILE MET  TIT GAA ACT TTC AAC GTT CCA GCC TTC TAC GTT TCC ATC CAT CAA GCC GTT TTG TAC TCT TCC GGT AGA ACT ACT GGT ATT GTT TTG PPLE GLU THR PHE ASN VAL PRO ALA PHE TIR VAL SER LEE GLN ALA VAL LEU SER LEU TYR SER SER GLY ARG THR THR GLY ILE VAL LEU  GAT TCC GGT GAT GGT GTT ACT CAC GTC GTT CCA ATT TAC GCT GGT TTC TCT TCT CAC GCC ATT TTG AGA CTC GAT TTG GCC GGT AGA ASP SER GLY ASP GLY VAL THR HIS VAL VAL PRO ILE TYR ALA GLY PRE SER LEU PRO HIS ALA ILE LEU ARG LEU ASP LEU ARG GLY ARG  GAT TTG ACT GAC TAC TTG ATG AAG ATC TTG AGT GAA CGT GGT TAC TCT TCT CCA CC ACT GCT GAA AGA GAA ATT GTC CGT GAC TAC AAG  SSP LEU THR ASP TYR LEU MET LYS ILLE LEU SER GLU ARG GLY TYR SER PHE SER THR THR ALA GLU ARG GLU ILE VAL ARG ASP ILLE LYS  GAA AAA CTA TGT TAC GTC GCC TTG GAC TTC GAA CAA GAA ATC CAA ACC GCT GGT GAT TCT TCT TCA ATT GAA AAA TCC TAC GAC TAC TGT TCA AGT GAA AAA TCC TAC GAC TAC TGT TCA ATT GAA AAA TCC TAC GAC TAC TCT TCT TCA ATT GAA AAA TCC TAC GAC TAC GCT GAC TAC ACT ATT GGT GAC TTC GAC TAC GAC TAC ACT ATT GGT GAC TAC GAC TAC GCT GAC TAC ACT ATT GGT GAC TAC ACC GAC GAC GAC GCT GCT GAT TCT TCA ATT GAA AAA TCC TAC GAC TAC GCT GAC TAC ACC GCT GCT GAA GAA ACC GCT GCT GAA GAA ACC GCT GCT GAT GAC GAA AAA TCC TCT TCC GCT GCT TCT GAT TTC GAT TCT TCA ATT GAT AAAA TCC GCT GAC GAC GAC GCT GCT GAT GCT CCT TCT GAT TTC GCT GAT TCT GCT GAT GCT GAC GAC GAC ACC ACC GCT GCT GAC GCC CCA GAA GCT TCT TCC CAT GCT TCT GCT TTT TGT GCT GCT GCT ACC GCT GAC GAT TCT TCC GCT GAT GCT GCT GCT GCT GCT GCT GCT GCT GCT GC				70									80										90		
TITG AGA GIT GCC CCA GAA GAA CAC CCT GIT CIT ITG ACT GAA GGT CCA ATC AAC CCT AAA TCX AAC AGA GAA AAX XIG ACT CAA ATT ATG LEU ANG VAL ALA PRO GLU GLU HIS PRO VAL LEU LEU THR GLU ALA PRO MET ASN PRO LYS ASN ARG LYS THR GLN ILE MET  TIT GAA ACT ITC AAC GIT CCA GCC ITC TAC GIT TCC ATC CAA GCC GIT TIT TCC TTG TAC TCT TCC GGT AGA ACT ACT GGT ATT GIT ITG PHE GLU THR PHE ASN VAL PRO ALA PHE TYR VAL SER ILE GLN ALA VAL LEU SER LEU TYR SER SER GLY ARG THR THR GLY ILE VAL LEU  GAT TCC GGT GAT GGT GIT ACT CCA GTC GTC CCA ATT TAC GCT GGT TTC TCT CTA CCT CAC GCC ATT TIG AGA CTC GAT TTG GCC ASP SER GLY ASP GLY VAL THR HIS VAL VAL PRO ILE TYR ALA GLY PHE SER LEU PRO HIS ALA ILE LEU ARG LEU ASP LEU ALA GLY ARG  GAT TTG ACT GAC TAC TTG ATG AAG ATC TTG AGT GAA CGT GGT TAC TCT TCT CCACC ACT GCT GAA AGA AGA ATT GTC CGT GAC ATC AGG ASP LEU THR ASP TYR LEU MET LIS ILE LEU SER GLU ARG GLY TYR SER PHE SER THR THR ALA GLU ANG GLU ILE VAL ARG ASP ILE LYS  GAA AAA CTA TUT TAC GTC GCC TTG GAC TTC GAA CAA GAA ATC CAA ACC GCT GCT CAA TCT TCT TCA ATT GAA AAA TCC TAC GAC ATC GTT GAA CAA GAC GTC GCT CAA TCT TCT TCA ATT GAA AAA TCC TAC GAC ATC GTT CAA CAA GAC ACT GCT CAA TCT TCT TCA ATT GAA AAA TCC TAC GAC ATC GCT GAA CAA GAC GCT GCT CAA TCT TCT TCA ATT GAA AAA TCC TAC GAA CTT CCA  GAT GGT CAA GTC ATC ATC ATT GGT AAC GAA AGA TTC AGA GCC CCA GAA GCT TTG TTC TCT TCT ATT TG AAAA TCC TAC GAA CTT CCA  GAT GGT CAA GTC ATC ATC ATC ATC AGC GAC TTC GAA CAA GAC ACC CCA GAA GCT TTG TTC CAT CCT TCT GTT TTG GGT TTG GAA TCT CCA  ASP GLY GLN VAL ILE THR ILE GLY ASN GLU ARG PHE ARG ALA PRO GLU ALA LEU PHE HIS PRO SER VAL LEU GLY LEU GLU LEU PRO  ATT GAC CAA ACT ACT TAC AAC TCC ATC ATG AAG GAT TCT AGG AAC GCC CCA GAA GCT TTG TTC CAT CCT TCT GTT TTG GGT TTG GAA TCT CCC GCT GAT ACC ATC ATC ATC ATC ATC ATC ATG AAG TCT CAT CAT GAA GAG ATC ACC GCT TCT GAT TAC CCT CCA TCT TCC ATG AAG ATC ACC GTT TTG TAC TCC GCT GAT TCC TCC ATC ATC ATC ATC ATC ATC ATC A	ATC TTG ACT	TTA CGT	TAC CO	CA ATT	GAA C	CAC GGT	ATT	GTC	ACC	AAC	TGG	GAC	GAT	ATG MET	GAA GLU	AAG LYS	ATC	TGG	CAT	CAT	ACC THR	TTC	TAC	AAC ASN	GAA GLU
TITG ACA GIT GCC CCA GAA GAA CAC CCT GIT CIT TITG ACT GAA GCT CCA ATT AND PRO LYS ASN ARG LYS THR GLN ILE MET  TIT GAA ACT TIC AAC GIT CCA GCC TIC TAC GIT TCC ATC CAA GCC GIT TIC TCC TIC TAC TCT TCC GGT AGA ACT ACT GGT ATT GIT TIC  PHE GLU THR PHE ASN VAL PRO ALA PHE TYR VAL SER ILE GLN ALA VAL LEU SER LEU TYR SER SER GLY ARG THR THR GLY ILE VAL LEU  GAT TCC GGT GAT GGT GTT ACT CAC GTC GTT CCA ATT TAC GCT GGT TTC TCT CTA CCT CAC GCC ATT TTG AGA CTC GAT TTG GCC  ASP SER GLY ASP GLY VAL THR HIS VAL VAL PRO ILE TYR ALA GLY PHE SER LEU PRO HIS ALA ILE LEU ARG LEU ASP LEU ALA GLY ARG  GAT TTG ACT GAC TAC TTG ATG AGG ATC TTG AGT GAC GGT TAC TCT TCT TCC ACC ACT GCT GAA AGA GAA ATT GTC GGT GAC ATC AGG  ASP LEU THR ASP TIR LEU MET L'S ILE LEU SER GLU ARG GLY TIR SER PHE SER THR THR ALA GLU ANG GLU ILE VAL ARG ASP ILE LYS  GAA AAA CTA TUT TAC GTC GCC TTG GAC TTC GAA CAA GAA ATG CAA ACC GCT GCT CAT CTT TCT TC TAC ATT TAC ATT GAA AAA TCC TAC GAA CTT CAG  GLU LYS LEU CYS TYR VAL ALA LEU ASP PHE GLU GLU MET GLN THR ALA ALA GLN SER SER SER ILE GLU LYS SER TYR GLU LEU PRO  GAT GGT CAA GTC ATC ACT ACT ATT GGT AAC GAA AGA TTC AGA GCC CCA GAA GGT TTT TCC ACT CAT CTT TTT TCC GAT CTT TTT TCC GAT TTT GAA TTT GAA TCT GAA TCT GCC  GAT GGT CAA GTC ATC ACT ATT GGT AAC GAA AGA TTC AGA GCC CCA GAA GGT TTT TCC ACT CAT TCT TTT TC GGT TTG GAA TCT GCC GGT  GAT GGT CAA GTC ATC ACT ATT GGT AAC GAA AGA TTC AGA GCC CCA GAA GGT TTT TCC ACT CCT TCT TTT TG GGT TTG GAA TCT GCC GGT  ASP GLY GLN VAL ILE THR ILE GLY ASS GLU ARG FHE ARG ALA PRO GLU ALA LEU PHE HIS PRO SER VAL LEU GLY LEU GLU SER ALA GLY  ATT GAC CAA ACT ACT TAC AAC TCC ATC AAC GAA AGA TTC AGA GCC CCA GAA GCT TTT TCC ACT CCT TCT TTT TTG GGT TTG GAA TCT GCC GGT  THR MET PHE PRO SER ILE ALA GLU ARG MET GLN LYS GLU ILE THR ALA LEU ALA PRO SER SER MET LYS VAL LYS ILE ILE ALA PRO PRO  GAA AGA AGA TAC TCC GTC TGG ATT GGT GGT TCT ATC TTG GCT TCT TTG GGT TCT CAT CAT CAT GAA GAA ATC TCC TCC CCC GT  THR MET PHE PRO SER ILE ALA GLU ARG MET GLN LYS GLU ILE THR ALA LEU ALA PRO SER SER MET LYS VAL LYS	100 000 1111	. LLO ANO			020 1	.10 00		****	••••		••							•••							
TIT GAM ACT TTC AAC GIT CCA GCC TTC TAC GIT TCC ATC CAM GCC GIT TIC TCC TTC TAC TCT TCC GGT AGA ACT ACT GGT ATT GIT TTC PHE GLU THR PHE ASN VAL PRO ALA PHE TYR VAL SER ILE GLN ALA VAL LEU SER LEU TYR SER SER GLY ARG THR THR GLY ILE VAL LEU  GAT TCC GGT GAT GGT GTT ACT CAC GTC GTT CCA ATT TAC GCT GGT TTC TCT CTA CCT CAC GCC ATT TTG AGA CTC GAT TTG GCC GGT AGA ASP SER GLY ASP GLY ASP GLY VAL THR HIS VAL VAL PRO ILE TYR ALA GLY PHE SER LEU PRO HIS ALA ILE LEU ARG LEU ASP LEU ALA GLY ARG  GAT TTG ACT GAC TAC TTG ATG AGA ATC TTG AGT GAA CCT GGT TAC TCT TTC TCC ACC ACT GCT GAA AGA GAA ATT GTC CGT GAC ATC AAG ASP ILE UTHR ASP TYR LEU NET LYS ILE LEU SER GLU ARG GLY TYR SER PHE SER THR THR ALA GLU ARG GLU ILE VAL ARG ASP ILE LYS  GAA AAA CTA TGT TAC GTC GCC TTG GAC TAC GAA CAA GAA ATG CAA ACC GCT GCT CAA TCT TCT TCA ATT GAA AAA TCC TAC GAA CTT CAC GAA CTT GCT GAA CAA GAA CTA TTG GCT GAC ATC ACC GCT GCT CAA TCT TCT TCA ATT GAA AAA TCC TAC GAA CTT CAC GAA CTT GCT GAT CAC GCT GCT CAA TCT TCT TCA ATT GAA AAA TCC TAC GAA CTT CAC GAA CTT GCT GAT CAC GAA GCT GCT CAA GCT GCT CAA TCT TCT TCT TCA ATT GAA AAA TCC TAC GAA CTT CAC GAA CTT GCT GAT GCT GAA CAC GCT GCT CAA TCT TCT TCC ACC ACT TCT GCT TCT GAA CCAA ACC GCT GCT CAA GCT GCT CAC ACT TCT TCT GCT TCT GAA CTA GAA ACC GCT GCT AAC ACC GCT TCT GCT TCT GCT TCT GCT TCT GAA TCT GCC GCT ASP GLY GLN VAL ILE THR ILE GLY ASN GLU ARG PHE ARG ALA PRO GLU ALA LEU PHE HIS PRO SER VAL LEU GLY LEU GLU SER ALA GLY  ATT GAC CAA ACT ACT TAC TAC TAC ACC ATC ATG AGG TCT GAT GCT GAT GCT GAT GCT CCT TCT GCT TCT T				AA CAC									AAC										CAA		
TIT GAA ACT TIC AAC GIT CCA GGC TIC TAC GIT TCC ATC CAT CCA GGC GIT TIG TCC TTO TAC TCT TCC GGT AGA ACT ACT GGT ATT GIT TIG PHE GLU THR PHE ASN VAL PRO ALA PHE TYR VAL SER ILE GLN ALA VAL LEU SER LEU TYR SER SER GLY ARG THR THR GLY ILE VAL LEU  GAT TCC GGT GAT GGT GIT ACT CAC GTC GTT CCA ATT TAC GCT GGT TIC TCT CTA CCT CAC GCC ATT TIG AGA CTC GAT TIG GCC GGT AGA ASP SER GLY ASP GLY VAL THR HIS VAL VAL PRO ILE TYR ALA GLY PHE SER LEU PRO HIS ALA ILE LEU ARG LEU ASP LEU ALA GLY ARG  GAT TIG ACT GAC TAC TIG ATG AGA ATC TTG AGT GAA CGT GGT TAC TCT TCT CTC CACC ACT GCT GAA AGA GAA ATT GTC CGT GAC ATC AAG ASP LEU THR ASP TYR LEU MET LYS ILE LEU SER GLU ARG GLY TYR SER PHE SER THR THR ALA GLU ARG GLU ILE VAL ARG ASP ILE LYS  GAA AAA CTA TGT TAC GTC GCC TTG GAC TTC GAA CAA GAA ATG CAA ACC GCT GCT GAA ACC GCT GAA AGA GAA ATA TCC TAC GAA CTT CCA GLU LYS LEU CYS TYR VAL ALA LEU ASP PHE GLU GLU GLU MET GLN THR ALA ALA GLN SER SER SER ILE GLU LYS SER TYR GLU LEU PRO  GAT GGT CAA GTC ATC ACT ATT GGT AAC GAA AGA TTC AGA GCC CCA GAA GCT TTG TTC CAT CCT TCT GTT TTG GGT TTG GAA TCT GCC GAP GLY GLN VAL ILE THR ILE GLY ASN GLU ARG PHE ARG ALA PRO GLU ALA LEU PHE HIS PRO SER VAL LEU GLY LEU GLU SER ALA GLY  ATT GAC CAA ACT ACT TAC ACT ATT GGT AAC GAA AGA TGC GAT GTC GGT AGA GAA ATC ACC GCT TTG GTT TTG GGT TTG GAT TCT CCG GGT GCC THR MET PHE PRO SER ILE ALA GLU ARG MET GLN LYS GLU ILE THR ALA LEU ALA PRO SER SER MET LYS VAL LYS ILE ILE ALA PRO PRO  GAA AGA AGA AGA TAT CCC GTC TGG ATT GGT GGT TCT ATC TTG GCT TCT TTG ACT TCC ATC AGA AGA GTC ATT GCT CCT CCA THR MET PHE PRO SER ILE ALA GLU ARG MET GLN LYS GLU ILE THR ALA LEU ALA PRO SER SER MET LYS VAL LYS ILE ILE ALA POR PRO  GAA AGA TAG TTC CCC AGT ATT GCC GAA AGA ATG CAA AAG GAA ATC ACC GCT TTG GCT CCC TCT TCC ATC AAG GTC ATC TCC AAA CAA GAA TAC ACC GCT ACC THR MET PHE PRO SER ILE ALA GLU ARG MET GLN LYS GLU ILE THR ALA LEU ALA PRO SER SER MET LYS VAL LYS ILE ILE ALA POR PRO  GAA AGT GGT CCA XCT ATC GTT CAC CAC AAG TGT TCT ATC TTG GCT TCT TTG ACT TCT TTG CAC TCC TCC TATTTTC TATTTTTAA	LEU ANG VAI	L ALA PRO	GLU ĢI	LU HIS	PRO V	AL LE	J LEU	THR	GLU	ALA	PRO	MET	ASN	PRO	LYS		ASN	ARG		LYS		THR	GLN	ILE	MET
PHE GLU THR PHE ASN VAL PRO ALA PHE TYR VAL SER ILE GLN ALA VAL LEU SER LEU TYR SER SER GLY ARG THR THR GLY ILE VAL LEU  160  GAT TCC GGT GAT GGT GTT ACT CAC GTC GTT CCA ATT TAC GCT GGT TTC TCT CTA CCT CAC GCC ATT TTG AGA CTC GAT TTG GCC GGT AGA ASP SER GLY ASP GLY VAL THR HIS VAL VAL PRO ILE TYR ALA GLY PHE SER LEU PRO HIS ALA ILE LEU ARG LEU ASP LEU ALA GLY ARG  GAT TTG ACT GAC TAC TTG ATG AAG ATC TTG AGT GAA CGT GGT TAC TCT TCC ACC ACT GCT GAA AGA GAA ATT GTC CGT GAC ATC AAG ASP LEU THR ASP TYR LEU MET LYS ILE LEU SER GLU ARG GLY TYR SER PHE SER THR THR ALA GLU ARG GLU ILE VAL ARG ASP ILE LYS  GAA AAA CTA TGT TAC GTC GCC TTG GAC TTC GAA CAA GAA ATC CAA ACC GCT GCT CAA TCT TCT TCA ATT GAA AAA TCC TAC GAA CTT CCA GLU LYS LEU CYS TYR VAL ALA LEU ASP PHE GLU GLN MCI UMET GLN THR ALA ALA GLN SER SER ILE GLU LYS SER TYR GLU LEU PRO  GAT GGT CAA GTC ATC ACT ATT GGT AAC GAA AGA TTC AGA GCC CCA GAA GCT TTG TTC CAT CCT TCT TGT TTG GGT TTG GAA TCT GCC GGT ASP GLY GLN VAL ILE THR ILE GLY ASN GLU ARG PHE ARG ALA PRO GLU ALA LEU PHE HIS PRO SER VAL LEU GLY LEU GLU SER ALA GLY  ATT GAC CAA ACT ACT TAC AAC TCC ATC ATC ATG AAG TGT GAT GTC GAT GTC CGT AAG GAA TTA TAC GGT AAC ATC GTT ATG TCC GGT GGT ACC ILE ASP GLN THR THR TYR ASN SER ILE MET LYS CYS ASP VAL ASP VAL ARG LYS GLU LEU TYR GLY ASN ILE VAL MET SER GLY GLY THR  ACC ATG TTC CCA AGT ATT GCC GAA AGA ATG CAA AAG GAA ATC ACC GCT TTG GTC TTG TCC CAT CCT TCC ATG AAG GTC AAG ATC ATT GCC GGT GGT ACC THR MET PHE PRO SER ILE ALA GLU ARG MET GLN LYS GLU ILE THR ALA LEU ALA PRO SER SER MET LYS VAL LYS ILE ILE ALA PRO PRO  GAA AGA AAG TAC TCC GTC TGG ATT GGT GGT TCT ATC TTG GCT TCT TTG ACT TCC CAT CCT TCC CAT CCT TCC AAG AAG ATC ATT GCC GGT GGT ACC GLU ARG LYS TYR SER VAL TRP ILE GLY GGT TCT ATC TTG GCT TCT TTG ACT TCC CAT CCT TCC CAT CCT TCC CAT CCT TCC ATC AT	TTT GAA AC'	TTC AAC	GTT C	130 CA GCC	TTC 1	rac gt	TCC	ATC	CAA	GCC	GTT	TTG	140 TCC	TTG	TAC	TCT	TCC	GGT	AGA	ACT	ACT	GGT		GTT	TTG
GAT TCC QGT GAT GGT GGT GTT ACT CAC GTC GTT CCA ATT TAC GCT GGT TTC TCT CTA CCT CAC GCC ATT TTG AGA CTC GAT TTG GCC GGT AGA ASP SER GLY ASP GLY VAL THR HIS VAL VAL PRO ILE TYR ALA GLY PHE SER LEU PRO HIS ALA ILE LEU ANG LEU ASP LEU ALA GLY ARG ASP SER GLY ASP GLY VAL THR HIS VAL VAL PRO ILE TYR ALA GLY PHE SER LEU PRO HIS ALA ILE LEU ANG LEU ASP LEU ALA GLY ARG GTT TTG ACT GAC ACT GAC ACT GAC ACT GAC AGA ATT GTC CGT GAC ATC AAG ASP LEU THR ASP TYR LEU MET LYS ILE LEU SER GLU ANG GLY TYR SER PHE SER THR THR ALA GLU ANG GLU ILE VAL ANG ASP ILE LYS GLA AAA CTA TGT TAC GTC GCC TTG GAC TTC GAA CAA GAA ATG CAA ACC GCT GCT CAA TCT TCT TCA ATT GAA AAA TCC TAC GAA CTT CCA GCU LYS LEU CYS TYR VAL ALA LEU ASP PHE GLU GLN GLU MET GLN THR ALA ALA GLN SER SER SER ILE GLU LYS SER TYR GLU LEU PRO GASP GLY GLN VAL ILE THR ILE GLY ASN GLU ANG PHE ANG ALA PRO GLU ALA LEU PHE HIS PRO SER VAL LEU GLY LEU GLU SER ALA GLY ASP GLY GLN VAL ILE THR ILE GLY ASN GLU ANG PHE ANG ALA PRO GLU ALA LEU PHE HIS PRO SER VAL LEU GLY LEU GLU SER ALA GLY ANG ASP CLA GAC ACT ACT TAC AAC TCC ATC ATC ATG AAG ATC GTT ATG TCC GAT ACC ILE ASP GLN THR THR TYR ASN SER ILE MET LYS CYS ASP VAL ASP VAL ASP VAL ARG LYS GLU LEU TYR GLY ASN ILE VAL MET SER GLY GLY THR ALA GLY ANG ATC ATT GCC GAA AGA ATG GAA ATG GAC ATC GTT TCC ATC TCC ATC ATC ATG AAG ATC ATT GCC GAA AGA ATG CAA AAG GAA ATC ACC GCT TTG GCC ATC TCC ATC ATG AAG GTC AAG ATC ATT GCC GAA AGA ATG GCA AAG ATG ATG GCC CAA ACC ATC GTT ATG TCC GCT CCA TCT TCC ATG AAG GTC AAG ATC ATT GCC GAT ACC ATC GTT ATC TCC GCT CCA TCT TCC ATG AAG GTC AAG ATC ATT GCC GAA AGA ATG GCA AAG ATG ATC ACC TCT TCC ATC ATC ATG AAG GTC AAG ATC ATT GCC CAC AAG ATG ATG GCC CAC AAG ATC ACC TCC CAC ACC ACC ACC ACC ACC ACC	PHE GLU TH	R PHE ASN	VAL P	RO ALA	PHE 1	TYR VAI	. SER	ILE	GLN	ALA	VAL	LEU	SER	LEU	TYR	SER	SER	GLY	ARG	THR	THR	GLY	ILE	VAL	LEU
GAT TIG ACT GAC TAC TTG ATG AAG ATC TTG AGT GAA CAA GAA ATG CAA AAG GAC CCC GAA GCT TTG TTC CAT CCT TCT GTT TTG GGT TTG GAA TCT GCC GGT ACC GCT GAA AGA ATC ATC GTT ACC GGT GAC ACC ACC ACT CCT TCT GTT TTG GGT TTG GAA TCT TCC AGC GGT GCT CAA ACC ACT GCT GAA AGA AAA TCC TAC GAA CTT CCA GAC ACC ACC ACT GCT GAA AGA AAA TCC TAC GAA CTT CCA GAA CTT CCA GAA ACC GCT GCT CAA TCT TCT TCC ACT GCT GAA AAA ACC TCT CCA GAA CTT CCA GAA CTT CCA GAA CTT CCA GAA ACC GCT GCT CAA TCT TCT TCC ATT GAA AAA TCC TAC GAA CTT CCA GCT GCT CAA TCT TCT TCC ATT GAA AAA TCC TAC GAA CTT CCA GCT GCT CAA TCT TCT TCC ATT GAA AAA TCC TAC GAA CTT CCA GAA CTT TCT TCC ATC ATC GTT TTG GGT TTG GAA TCT GCC GGT ASP GLY GLN VAL ILE THR ILE GLY ASN GLU ARG PHE ARG ALA PRO GLU ALA LEU PHE HIS PRO SER VAL LEU GLY LEU GLU SER ALA GLY ASP GLN THR THR TYR ASN SER ILE MET LYS CYS ASP VAL ASP VAL ARG LYS GLU LEU TYR GLY ASN ILE VAL MET SER GLY GLY THR ACC ATC TCC CCA ACC ACC ACC ACC ACC ACC				160																					
GAT TIG ACT GAC TAC TTO ATG AÃO ATC TTO AGT GAA CGT GGT TAC TCT TTO TCC ACC ACT GCT GAA AGA GAA ATT GTC CGT GĀC ATC AAG ASP LEU THR ASP TYR LEU MET LYS ILE LEU SER GLU ARG GLY TYR SER PHE SER THR THR ALA GLU ARG GLU ILE VAL ARG ASP ILE LYS  220  230  233  240  240  250  260  270  270  280  270  280  280  280  28	ASP SER GL	T GAT GGT Y ASP GLY	VAL T	CT CAC HR HIS	GTC C	STT CC: Val Pro	ATT LLE	TAC TYR	GCT	GCT	PHE	TCT	CTA LEU	PRO	HIS	ALA	ILE	LEU	AGA ARG	LEU	ASP	LEU	ALA	GLY	AGA ARG
GAT TIG ACT GAC TAC TTO ATG AÃO ATC TTO AGT GAA CGT GGT TAC TCT TTO TCC ACC ACT GCT GAA AGA GAA ATT GTC CGT GAC ATC AAG ASP LEU THR ASP TYR LEU MET LYS ILE LEU SER GLU ARG GLY TYR SER PHE SER THR THR ALA GLU ARG GLU ILE VAL ARG ASP ILE LYS  220  GAA AAA CTA TGT TAC GTC GCC TTO GAC TTC GAA CAA GAA ATG CAA ACC GCT GCT CAA TCT TCT TAC ATT GAA AAA TCC TAC GAA CTT CCA GLU LYS LEU CYS TYR VAL ALA LEU ASP PHE GLU GLN GLU MET GLN THR ALA ALA GLN SER SER ILE GLU LYS SER TYR GLU LEU PRO  GAT GGT CAA GTC ATC ACT ATT GGT AAC GAA AGA TTC AGA GCC CCA GAA GCT TTG TTC CAT CCT TCT GTT TTG GGT TTG GAA TCT GCC GGT ASP GLY GLN VAL ILE THR ILE GLY ASN GLU ARG PHE ARG ALA PRO GLU ALA LEU PHE HIS PRO SER VAL LEU GLY LEU GLU SER ALA GLY  ATT GAC CAA ACT ACT TAC AAC TCC ATC ATG AGA GTGT GTG GAT GTC CGT AAG GAA TTA TAC GGT AAC ATC GTT ATG TCC GGT GAT ACC ILE ASP GLN THR THR TYR ASN SER ILE MET LYS CYS ASP VAL ASP VAL ARG LYS GLU LEU TYR GLY ASN ILE VAL MET SER GLY GLY THR  ACC ATG TTC CCA AGT ATT GCC GAA AGA ATG CAA AAG GAA ATC ACC GCT TTG GCT CCA TCT TCC ATG AAG GTC ATG ATT GCT CCT CCA THR MET PHE PRO SER ILE ALA GLU ARG MET GLN LYS GLU ILE THR ALA LEU ALA PRO SER SER MET LYS VAL LYS ILE ILE ALA PRO PRO  GAA AGA AGA AGA TAC TCC GTC TGG ATT GGT GCT TCT TTG ACT ACC TTC TCC AAC AAG ATG TCC AAA CAA AAC AAC AAC AAC AAC AAC AAC				190									200										210		
CAA AAA CTA TGT TAC GTC GCC TTG GAC TTC GAA CAA GAA ATG CAA ACC GCT GCT CAA TCT TCT AAT GAA AAA TCC TAC GAA CTT CCA GLU LYS LEU CYS TYR VAL ALA LEU ASP PHE GLU GLN GLU MET GLN THR ALA ALA GLN SER SER SER ILE GLU LYS SER TYR GLU LEU PRO  CAT GGT CAA GTC ATC ACT ATT GGT AAC GAA AGA TTC AGA GCC CCA GAA GCT TTG TTC CAT CCT TCT GTT TTG GGT TTG GAA TCT GCC GGT ASP GLY GLN VAL ILE THR ILE GLY ASN GLU ARG PHE ARG ALA PRO GLU ALA LEU PHE HIS PRO SER VAL LEU GLY LEU GLU SER ALA GLY  ATT GAC CAA ACT ACT TAC AAC TCC ATC ATG AAG TGT GAT GTC GAT GTC CGT AAG GAA TTA TAC GGT AAC ATC GTT ATG TCC GGT GGT ACC ILE ASP GLN THR THR TYR ASN SER ILE MET LYS CYS ASP VAL ASP VAL ASP VAL ARG LYS GLU LEU TYR GLY ASN ILE VAL MET SER GLY GLY THR  ACC ATG TTC CCA AGT ATT GCC GAA AGA AGA GAA AAG GAA ATC ACC GCT TTG GCT CCA TCT TCC ATG AAG GTC AAG ATC ATT GCT CCA TCT TCC ATG AAG GTC AAG ATC ATT GCT CCA TCT TCC ATG AAG GTC AAG ATC ATT GCT CCA TCT TCC CAA GAA AGA AAG GAA AGA AAG GAA ATC ACC GCT TTG GCT CCA TCT TCC ATG AAG GTC AAG ATC ATT GCT CCA TCT TCC ATG AAG GTC AAG ATC ATT GCT CCA TCT TCC ATG AAG GTC TCC AAAA AAG GAA ATC ACC GCT TTG GCT CCA TCT TCC ATG AAG GTC AAG ATC ATT GCT CCA TCC CCA AGA AGA AAG TCC TCT TCC ATG AAG GTC TCA AAA CAA GAA ATC ATC GCT CCA TCT TCC ATG AAG TCC TCC AAAA AAG AAG AAG AAG TAC TTG GGT GGT TCT ATC TTG GCT TCT TTG ACT ACC TTC CAA CAA ATG TGG ATC TCA AAA CAA GAA TAC GAC GAC ATC GTT GGT GGT TCT ATC TTG GCT TCT TTG ACT ACC TTC CAA CAA ATG TGG ATC TCA AAA CAA GAA TAC GAC GAC ATC GTT GGT GGT TCT ATC TTG GCT TCT TTTG ACT ACC TTC CAA CAA ATG TGG ATC TCA AAA CAA GAA TAC GAC GAC ATC GTT GGT GGT TCT ATC TTG GCT TCT TTTG ACT ACC TTC CAA CAA ATG TGG ATC TCA AAA CAA GAA TAC GAC GAC ATC GTT GGT GGT TCT ATC TTG GCT TCT TTTTTTAAA GGLU SER GLY PRO  1675 1685 1695 1705 1715 1725  GAA AGT GGT CCA XCT ATC GTT CAC CAC AAG TGT TCT TAA TCTCTGCTTT TGTGGCGGTA TGTTTATGTA TGTACCTCTC TCTCTATTTC TATTTTTAAA GGLU SER GLY PRO  1675 1685 1695 1705 1715 1725				TG AAG									TCC										GAC		
GAA AAA CTA TGT TAC GTC GCC TTG GAC TTC GAA CAA GAA ATG CAA ACC GCT GCT CAA TCT TCT TCA ATT GAA AAA TCC TAC GAA CTT CCA GLU LYS LEU CYS TYR VAL ALA LEU ASP PHE GLU GLU GLU MET GLN THR ALA ALA GLN SER SER SER ILE GLU LYS SER TYR GLU LEU PRO  250 GAT GGT CAA GTC ATC ACT ATT GGT AAC GAA AGA TTC AGA GCC CCA GAA GCT TTG TTC CAT CCT TCT GTT TTG GGT TTG GAA TCT GCC GGT ASP GLV GLN VAL ILE THR ILE GLY ASN GLU ARG PHE ARG ALA PRO GLU ALA LEU PHE HIS PRO SER VAL LEU GLY LEU GLU SER ALA GLY  ATT GAC CAA ACT ACT TAC AAC TCC ATC ATG AAG TGT GAT GTC GAT GTC CGT AAG GAA TTA TAC GGT AAC ATC GTT ATG TCC GGT GGT ACC ILE ASP GLN THR THR TYR ASN SER ILE MET LYS CYS ASP VAL ASP VAL ARG LYS GLU LEU TYR GLY ASN ILE VAL MET SER GLY GLY THR  ACC ATG TTC CCA AGT ATT GCC GAA AGA AGA ATG CAA AAG GAA ATC ACC GCT TTG GCT CCA TCT TCC ATG AAG GTC AAG ATC ATT GCT CCC ATG THR MET PHE PRO SER ILE ALA GLU ARG MET GLN LYS GLU ILE THR ALA LEU ALA PRO SER SER MET LYS VAL LYS ILE ILE ALA PRO PRO  GAA AGA AAG TAC TCC GTC TGG ATT GGT GGT TCT ATC TTG GCT TCT TTG ACT ACC TTC CAA CAA ATC TGG ATC TCA AAA CAA GAA TAC GGC GLU ARG LYS TYR SER VAL TRP ILE GLY GLY SER ILE LEU ALA SER LEU THR THR PHE GLN GLN MET TRP ILE SER LYS GLN GLU TYR ASP  GAA AGT GGT CCA XCT ATC GTT CAC CAC AAG TGT TTC TAA TCTCTGCTTT TGTGGGCGTA TGTTTATGTA TGTACCTCTC TCTCTATTC TATTTTTAAA  GAA AGT GGT CCA XCT ATC GTT CAC CAC AAG TGT TTC TAA TCTCTGCTTT TGTGGGCGGTA TGTTTATGTA TGTACCTCTC TCTCTATTC TATTTTTAAA  GAU SER GLY PRO  ILE VAL HIS HIS LYS CYS PHE END	ROI LLO III	N NO. 111	, DEO 11		100 1	- LO - JE	· OLO	AIIU	ULI	***	JEN			••••	••••				000		••••				<b>5.0</b>
GAT GGT CAA GTC ATC ACT ATT GGT AAC GAA AGA TTC AGA GCC CCA GAA GCT TTG TTC CAT CCT TCT GTT TTG GGT TTG GAA TCT GCC GGT ASP GLY GLN VAL ILE THR ILE GLY ASN GLU ARG PHE ARG ALA PRO GLU ALA LEU PHE HIS PRO SER VAL LEU GLY LEU GLU SER ALA GLY  ATT GAC CAA ACT ACT TAC AAC TCC ATC ATG AAG TGT GAT GTC GAT GTC CGT AAG GAA TTA TAC GGT AAC ATC GTT ATG TCC GGT GGT ACC ILE ASP GLN THR THR TYR ASN SER ILE MET LYS CYS ASP VAL ASP VAL ARG LYS GLU LEU TYR GLY ASN ILE VAL MET SER GLY GLY THR  ACC ATG TTC CCA AGT ATT GCC GAA AGA ATG AAG GAA ATC ACC GCT TTG GCT CCA TCT TCC ATG AAG GTC AAG ATC ATT GCT CCT CCA THR MET PHE PRO SER ILE ALA GLU ARG MET GLN LYS GLU ILE THR ALA LEU ALA PRO SER SER MET LYS VAL LYS ILE ILE ALA PRO PRO  GAA AGA AAG TAC TCC GTC TGG ATT GGT GGT TCT ATC TTG GCT TCT TTG ACT ACC TTC CAA CAA ATG TGG ATC TCA AAA CAA GAA TAC GAC GLU ARG LYS TYR SER VAL TRP ILE GLY GLY SER ILE LEU ALA SER LEU THR THR PHE GLN GLN MET TRP ILE SER LYS GLN GLU TYR ASP  GAA AGT GGT CCA XCT ATC GTT CAC CAC AAG TGT TTC TAA TCTCTGCTTT TGTGGCGCGTA TGTTTATGTA TGTACCTCTC TCTCTATTTC TATTTTTAAA  GLU SER GLY PRO  ILE VAL HIS HIS LYS CYS PHE END				CC TTG									GCT			TCT	TCA	ATT						CTT	
GAT GGT CAA GTC ATC ACT ATT GGT AAC GAA AGA TTC AGA GCC CCA GAA GCT TTG TTC CAT CCT TCT GTT TTG GGT TTG GAA TCT GCC GGT ASP GLY GLN VAL ILE THR ILE GLY ASN GLU ARG PHE ARG ALA PRO GLU ALA LEU PHE HIS PRO SER VAL LEU GLY LEU GLU SER ALA GLY  280  ATT GAC CAA ACT ACT TAC AAC TCC ATC ATG AAG TGT GAT GTC GAT GTC CGT AAG GAA TTA TAC GGT AAC ATC GTT ATG TCC GGT GGT ACC ILE ASP GLN THR THR TYR ASN SER ILE MET LYS CYS ASP VAL ASP VAL ARG LYS GLU LEU TYR GLY ASN ILE VAL MET SER GLY GLY THR  310  320  ACC ATG TTC CCA AGT ATT GCC GAA AGA ATG CAA AAG GAA ATC ACC GCT TTG GCT CCA TCT TCC ATG AAG GTC AAG ATC ATT GCT CCT CCA THR MET PHE PRO SER ILE ALA GLU ARG MET GLN LYS GLU ILE THR ALA LEU ALA PRO SER SER MET LYS VAL LYS ILE ILE ALA PRO PRO  GAA AGA AAG TAC TCC GTC TGG ATT GGT GGT TCT ATC TTG GCT TCT TTG ACT ACC TTC CAA CAA ATG TGG ATC TCA AAA CAA GAA TAC GAC GLU ARG LYS TYR SER VAL TRP ILE GLY GLY SER ILE LEU ALA SER LEU THR THR PHE GLN GLN MET TRP ILE SER LYS GLN GLU TYR ASP  370  1675  1685  1695  1705  1715  1725  GAA AGT GGT CCA XCT ATC GTT CAC CAC AAG TGT TTC TAA TCTCTGCTTT TGTGCGCGGTA TGTTTATGTA TGTACCTCTC TCTCTATTTC TATTTTTAAA GLU SER GLU PRO  ILE VAL HIS HIS LYS CYS PHE END	GLU LYS LE	U CYS TY	VAL A	LA LEU	ASP I	PHE GL	U GEN	GLU	MET	GLN	THR	ALA	ALA	GLN	SER	SER	SER	ILE	GLU	LYS	SER	TYR	GLU	LEU	PRO
ASP GLY GLN VAL ILE THR ILE GLY ASN GLU ARG PHE ARG ALA PRO GLU ALA LEU PHE HIS PRO SER VAL LEU GLY LEU GLU SER ALA GLY  280 270 280 270 280 280 280 290 300 300 310 310 320 320 320 320 320 320 320 320 320 32	CAT GGT CA	A GTC ATC	ACTA	TT CCT		GAA AG	<b>ል</b> ጥጥር	AGA	ccc	CCA	GAA	CCT	TTG		CAT	CCT	тст	GTT	TTG	CCT	TTG	GAA	тст		CCT
ATT GAC CAA ACT ACT TAC AAC TCC ATC ATG AAG TGT GAT GTC GAT GTC CGT AAG GAA TTA TAC GGT AAC ATC GTT ATG TCC GGT GGT ACC ILE ASP GLN THR THR TYR ASN SER ILE MET LYS CYS ASP VAL ASP VAL ARG LYS GLU LEU TYR GLY ASN ILE VAL MET SER GLY GLY THR  310  320  330  ACC ATG TTC CCA AGT ATT GCC GAA AGA AGA AGG GAA ATC ACC GCT TTG GCT CCA TCT TCC ATG AAG GTC AAG ATC ATT GCT CCT CCA THR MET PHE PRO SER ILE ALA GLU ARG MET GLN LYS GLU ILE THR ALA LEU ALA PRO SER SER MET LYS VAL LYS ILE ILE ALA PRO PRO  GAA AGA AAG TAC TCC GTC TGG ATT GGT GGT TCT ATC TTG GCT TCT TTG ACT ACC TTC CAA CAA ATG TGG ATC TCA AAA CAA GAA TAC GAC GLU ARG LYS TYR SER VAL TRP ILE GLY GLY SER ILE LEU ALA SER LEU THR THR PHE GLN GLN MET TRP ILE SER LYS GLN GLU TYR ASP  370  1675  1685  1695  1705  1715  1725  GAA AGT GGT CCA XCT ATC GTT CAC CAC AAG TGT TTC TAA TCTTCTGCTTT TGTGCGCGTA TGTTTATGTA TGTACCTCTC TCTCTATTTC TATTTTTAAA GLU SER GLY PRO  ILE VAL HIS HIS LYS CYS PHE END																									
310  ACC ATG TTC CCA AGT ATT GCC GAA AGA ATG CAA AAG GAA ATC ACC GCT TTG GCT CCA TCT TCC ATG AAG GTC AAG ATC ATT GCT CCA THR MET PHE PRO SER ILE ALA GLU ARG MET GLN LYS GLU ILE THR ALA LEU ALA PRO SER SER MET LYS VAL LYS ILE ILE ALA PRO PRO  340  GAA AGA AAG TAC TCC GTC TGG ATT GGT GGT TCT ATC TTG GCT TCT TTG ACT ACC TTC CAA CAA ATG TGG ATC TCA AAA CAA GAA TAC GAC GLU ARG LYS TYR SER VAL TRP ILE GLY GLY SER ILE LEU ALA SER LEU THR THR PHE GLN GLN MET TRP ILE SER LYS GLN GLU TYR ASP  370  1675  1685  1695  1705  1715  1725  GAA AGT GGT CCA XCT ATC GTT CAC CAC AAG TGT TTC TAA TCTCTGCTTT TGTGCGCGTA TGTTTATGTA TGTACCTCTC TCTCTATTTC TATTTTTAAA GLU SER GLY PRO  ILE VAL HIS HIS LYS CYS PHE END					280									290										300	
ACC ATG TTC CCA AGT ATT GCC GAA AGA ATG CAA AAG GAA ATC ACC GCT TTG GCT CCA TCT TCC ATG AAG GTC AAG ATC ATT GCT CCT CCA THR MET PHE PRO SER ILE ALA GLU ARG MET GLN LYS GLU ILE THR ALA LEU ALA PRO SER SER MET LYS VAL LYS ILE ILE ALA PRO PRO  340  350  GAA AGA AAG TAC TCC GTC TGG ATT GGT GGT TCT ATC TTG GCT TCT TTG ACT ACC TTC CAA CAA ATG TGG ATC TCA AAA CAA GAA TAC GAC GLU ARG LYS TYR SER VAL TRP ILE GLY GLY SER ILE LEU ALA SER LEU THR THR PHE GLN GLN MET TRP ILE SER LYS GLN GLU TYR ASP  370  1675  1685  1695  1705  1715  1725  GAA AGT GGT CCA XCT ATC GTT CAC CAC AAG TGT TTC TAA TCTCTGCTTT TGTGCGCGTA TGTTTATGTA TGTACCTCTC TCTCTATTTC TATTTTTAAA GLU SER GLY PRO  ILE VAL HIS HIS LYS CYS PHE END	ATT GAC CA	A ACT AC' N THR TH	TAC ARTYR A	AC TCC SN SER	ATC .	ATG AA Met ly	G TGT S CYS	GAT	GTC	GAT	GTC VAL	CGT	A AG	GAA GLU	TTA Leu	TAC TYR	GCT	AAC	ATC ILE	GTT VAL	MET	TCC	GGT	GGT GLY	ACC THR
ACC ATG TTC CCA AGT ATT GCC GAA AGA ATG CAA AAG GAA ATC ACC GCT TTG GCT CCA TCT TCC ATG AAG GTC AAG ATC ATT GCT CCT CCA THR MET PHE PRO SER ILE ALA GLU ARG MET GLN LYS GLU ILE THR ALA LEU ALA PRO SER SER MET LYS VAL LYS ILE ILE ALA PRO PRO  340  350  GAA AGA AAG TAC TCC GTC TGG ATT GGT GGT TCT ATC TTG GCT TCT TTG ACT ACC TTC CAA CAA ATG TGG ATC TCA AAA CAA GAA TAC GAC GLU ARG LYS TYR SER VAL TRP ILE GLY GLY SER ILE LEU ALA SER LEU THR THR PHE GLN GLN MET TRP ILE SER LYS GLN GLU TYR ASP  370  1675  1685  1695  1705  1715  1725  GAA AGT GGT CCA XCT ATC GTT CAC CAC AAG TGT TTC TAA TCTCTGCTTT TGTGCGCGTA TGTTTATGTA TGTACCTCTC TCTCTATTTC TATTTTTAAA GLU SER GLY PRO  ILE VAL HIS HIS LYS CYS PHE END																									
340  GAA AGA AAG TAC TCC GTC TGG ATT GGT GGT TCT ATC TTG GCT TCT TTG ACT ACC TTC CAA CAA ATG TGG ATC TCA AAA CAA GAA TAC GAC GLU ARG LYS TYR SER VAL TRP ILE GLY GLY SER ILE LEU ALA SER LEU THR THR PHE GLN GLN MET TRP ILE SER LYS GLN GLU TYR ASP  370  1675  1685  1695  1705  1715  1725  GAA AGT GGT CCA XCT ATC GTT CAC CAC AAG TGT TTC TAA TCTCTGCTTT TGTGCGCGTA TGTTTATGTA TGTACCTCTC TCTCTATTTC TATTTTTAAA GLU SER GLY PRO  ILE VAL HIS HIS LYS CYS PHE END					AGA .									CCA										CCT	CCA
GAA AGA AGG TAC TCC GTC TGG ATT GGT GGT TCT ATC TTG GCT TCT TTG ACT ACC TTC CAA CAA ATG TGG ATC TCA AAA CAA GAA TAC GAC GLU ARG LYS TYR SER VAL TRP ILE GLY GLY SER ILE LEU ALA SER LEU THR THR PHE GLN GLN MET TRP ILE SER LYS GLN GLU TYR ASP  370 1675 1685 1695 1705 1715 1725  GAA AGT GGT CCA XCT ATC GTT CAC CAC AAG TGT TTC TAA TCTCTGCTTT TGTGCGCGTA TGTTTATGTA TGTACCTCTC TCTCTATTTC TATTTTTAAA GLU SER GLY PRO ILE VAL HIS HIS LYS CYS PHE END	inn mei Pr	E FRU SE	A TIPE A			mei UL	w ris	GLU	116	IHK	ALA	LEU	ALA		OBK	JER	mc i	LIS	VAL	LIS	166	166	ALA		
370 1675 1685 1695 1705 1715 1725 GAA AGT GGT CCA XCT ATC GTT CAC CAC AAG TGT TTC TAA TCTCTGCTTT TGTGCGCGTA TGTTTATGTA TGTACCTCTC TCTCTATTTC TATTTTTAAA GLU SER GLY PRO ILE VAL HIS HIS LYS CYS PHE END	GAA AGA AA	G TAC TO	C GTC 1	GG ATT		GGT TC	T ATC	TTG	GCT	TCT	TTG	ACT	ACC		CAA	CAA	ATG	TGG	ATC	TCA	AAA	CAA	GAA		
GAA AGT GGT CCA XCT ATC GTT CAC CAC AAG TGT TTC TAA TCTCTGCTTT TGTGCGCGTA TGTTTATGTA TGTACCTCTC TCTCTATTTC TATTTTTAAA GLU SER GLY PRO ILE VAL HIS HIS LYS CYS PHE END	GLU ARG LY	S TYR SE	R VAL 1	RP ILE	GLY	GLY SE	R ILE	LEU	ALA			THR			GLN			TRP					GLU		
GLU SER GLY PRO ILE VAL HIS HIS LYS CYS PHE END	GAA AGT GO	T CCA YO	r Atte	:TT ርልባ	370 CAC	AAG TO	<b>ጥ ጥጥ</b> ር	. TAA	ጥር፣			<b>ፐር</b> ጥር			ሊጥሞ።			TACC					T A THE		
	GLU SER GI	Y PRO	ILE V	AL HIS	HIS	LYS CY	S PHE	E END	101	.0166	***	Tole	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	I M I	.U111	MIUI	A 10	INC	.1010	10	ICIAI	110	1817	AITA	AR

FIG. 2. DNA sequence of yeast actin gene and amino acid sequence of the actin. The first two codons (for Asp and Ser) are separated from the main body of the gene by a 309-bp intervening sequence. Uncertainties exist at codons 113, 116, 118, and 366 (indicated by "X"). Approximately 200 bp upstream from the initiation codon is the putative promoter site, T-A-T-A-T.

intron region, the sequence again becomes colinear. The sequence of the mRNA definitely establishes the existence of a 309-bp intervening sequence in the yeast actin gene. The interruption occurs between codons 2 and 3, giving a predicted NH<sub>2</sub>-terminal sequence for yeast actin of Asp-Ser-Glu-Val-Ala.

CCACCCTCTC AATAAAATAA AAATAATAAA GATTT

Possible Transcription and Translation Signals in the Actin Gene. Comparison of the yeast actin gene sequences to the se-

quences of other eukaryotic genes has allowed the identification of several transcriptional and translational control sequences. If it is assumed that the longest DNA copies in the dideoxy primer-extension experiment are full length, then transcription starts approximately  $140 \pm 30$  bp away from the iniatiation codon (unpublished data). A potential promoter recognition sequence, T-A-T-A-T-A-T, similar to that seen in other eukaryotic genes, is found in a region 30–50 bp upstream from

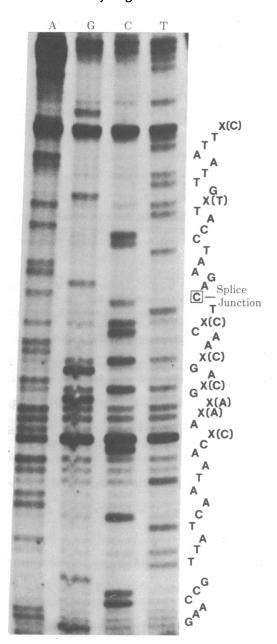


FIG. 3. The sequence of the mRNA was determined using the dideoxy primer-extension technique as reported by Sures et al. (27). Poly(A) RNA (125 µg) was hybridized to 5 pmol of a 75-bp strand-separated Hpa II/HinfI fragment (codons 22-25). Reverse transcriptase was then used to extend the primer, and the DNA products were electrophoresed on a thin sequencing gel (33). Using the known DNA sequence as a guide, we were able to determine that the mRNA sequence was colinear with the gene, except that it lacked the sequence of 309 nucleotides immediately following codon 3. The location of the splice junction is indicated. Xs indicate regions of uncertainties. The letter inside the parentheses is the expected nucleotide.

this site (34) (Fig. 2). The actin gene also uses an A-U-G-G translation start sequence that is found in most higher eukaryotic initiation sites (15, 35). A pentanucleotide—A-A-T-A-A—found near the end of the 3' untranslated region of all eukaryotic mRNAs (36, 37) also appears 70 nucleotides after the termination codon. Poly(A) addition occurs 18 bases after this site in rabbit, human, and mouse  $\beta$ -globin mRNA (36). Further experimentation is needed to prove that the regions discussed have real roles in transcriptional or translational control.

## DISCUSSION

In this paper we report the isolation, identification, and nucleotide sequence of the actin gene from yeast. The nucleotide sequence establishes (with a few remaining uncertainties) the amino acid sequence of yeast actin. The NH2-terminal sequence, Asp-Ser-Glu-Val-Ala, is one amino acid shorter than that in Physarum, Dictyostelium, and rabbit skeletal muscle  $(\alpha)$  actins but is the same length as in bovine brain  $(\beta$  and  $\gamma)$  and chicken gizzard ( $\gamma$ ) actins (5, 31). The amino acid sequence differs from that of Physarum in 43 positions, for an 11.6% variation. By contrast, Physarum actin differs from mammalian cytoplasmic actin in only 12 residues, a variation of 3.2% (5). One of the obvious contributions of the new DNA sequencedetermination technologies will be the detailed study of the changes of gene sequence in evolution. The easy access to actin sequences due to the conserved homologies should make the actin genes among the most studied in this regard.

The sequence of the yeast actin gene reveals an extreme bias in codon usage similar to that found in other yeast genes whose sequences have been determined (14–16). We note, for example, that the 28 glycine residues are all coded for by GGU. There is a similar unique selection of codons for cysteine (UGU), asparagine (AAC), glutamine (CAA), and tyrosine (UAC). Similar cases of extreme preference in the use of codons is seen, for several other amino acids. The significance of this bias is not clear, but it could have practical importance in the design of synthetic DNA sequences for use in the isolation of specific yeast genes (14).

The yeast actin intron resembles other intervening sequences found in eukaryotic organisms; it is extremely A+T rich (68% A+T), and a pyrimidine tract exists near the 3' end. The splice junction sequences are similar to those found at other exonintron regions and conform to rules postulated by several groups (17, 18). Seif et al. (17) found that the 5' junction of most eukaryotic and viral genes are Pu<sup>1</sup>G-T-X-X-G, and that the 3' side has a Py-Py-X-Py-A-G<sup>1</sup> sequence. They also reported that no dinucleotide A-G appears within 13 nucleotides of the terminal A-G of the 3' end of the intervening sequence. The yeast actin junction sequences are as follows: 5' G<sup>1</sup>G-T-A-T-G ··· G-T-T-T-A-G<sup>1</sup>A-G 3'. Because of a terminal repeat at the splice junction, we cannot unambiguously assign the origin of the G in the mRNA (Fig. 4). However, according to the above rules, the G probably comes from the 5' side of the intron.

In comparing the actin genes between species, it is interesting to observe that *Dictyostelium* contains many more actin genes (17) than does yeast, but the genes most likely do not contain intervening sequences. Drosophila has six actin genes, and the two genes examined have an intervening sequence (19). The Drosophila intervening sequence is larger than that of yeast but has a similar location near the 5' end of the gene (19). The highest variation among actin amino acid sequences occurs in two regions (31): between residues 2 and 18 and between residues 259 and 298. The location of the intervening sequence between codons 2 and 3 suggests that the intervening sequence could have a role in promoting functional variations in the actin sequence. This could have been through DNA recombination, a mechanism likely to have been operative in the evolutionary past, or through alternative modes of splicing, as in the case of the adenovirus late mRNAs (38) in the present. The latter mechanism cannot act at a high efficiency in vegetative cells because there appears to be one predominant yeast actin mRNA (Fig. 3).

In our laboratory we have been interested in the mechanism of RNA splicing. We have used the yeast tRNA system to study this reaction (39). Comparison of the splice junctions in tRNA genes with those in nuclear genes coding for proteins has dis-

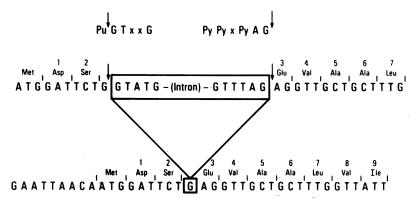


FIG. 4. The 5' end of the actin gene with its intervening sequence is outlined above. The bottom line shows the DNA sequence determined from the mRNA. The mRNA sequence shows that the 5' end of the actin message originates from a region 309 nucleotides upstream from codon 3. The yeast actin splice junction conforms to the canonical splice sequence postulated by Seif et al. (17). According to their rules, one would expect the terminally redundant G to come from the 5' side of the intervening sequence.

couraged the possibility that the two different splicing reactions are carried out by the same enzyme system. The existence of an intervening sequence in the actin gene dictates that yeast must have a mRNA splicing system similar to that found in higher organisms. The yeast system may prove useful in investigation of the mechanism of mRNA splicing.

The results reported here open the way to two avenues of research. The actin gene can be used to study the regulation and expression of a split gene in yeast. The existence of a single actin gene offers the possibility for obtaining conditional lethal mutants of actin. Such mutants would provide us with an opportunity to study the function of actin in yeast.

The data presented in this paper establish the yeast actin gene as a nearly "classic" eukaryotic gene. In evidence are a typical promoter recognition site, an intervening sequence with canonical splice sequences, and a poly(A) addition site near the 3' end of the gene. This evidence adds to a long list of features that establish yeast as an excellent model eukaryote.

We thank Dr. J. Beckmann, Dr. S. Kang, P. Johnson, J. Romeo, J. Rose, and L. Velten for their technical assistance. We are grateful to Dr. R. Firtel for providing us with an actin probe from *Dictyostellum* and to Dr. N. Davidson for unpublished data. We appreciate the efforts of M. Romashko in helping prepare this manuscript. Last, we appreciate the encouragement of the following people: Dr. H. Sakano, E. Smith, and especially, J. Velten and Dr. K. Janssen. This research was supported by Grant CA-10984 from the National Institutes of Health.

- Pollard, T. D. & Weihing, R. R. (1974) in Critical Reviews in Biochemistry, ed. Fasman, G. (CRC, Cleveland, OH), Vol. 2, pp. 1-65.
- Goldman, R., Polland, T. & Rosenbaum, J., eds. (1976) Cell Mottlitty (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), Vols. 1-3.
- 3. Spudich, J. A. (1974) J. Biol. Chem. 249, 6013-6020.
- 4. Kindle, K. L. & Firtel, R. A. (1978) Cell 15, 763-778.
- Vandekerckhove, J. & Weber, K. (1978) Nature (London) 276, 720-721.
- Koteliansky, V. E., Glukhova, M. A., Benjaman, M. V., Surguchov, A. P. & Smirnov, V. N. (1979) FEBS Lett. 102, 55–58.
- Rungger, D., Runggar-Brändle, E., Chaponnier, C. & Gabbiani, G. (1979) Nature (London) 282, 320–321.
- 8. Lazarides, E. (1976) J. Cell Biol. 68, 202-219.
- 9. Crick, F. (1979) Science 204, 264-271.
- Goodman, H. M., Olson, M. V. & Hall, B. D. (1977) Proc. Natl. Acad. Sci. USA 74, 5453-5457.
- Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R. & Rutter, W. J. (1978) Proc. Natl. Acad. Sci. USA 75, 190-194.
- Church, G. M., Slonimski, P. P. & Gilbert, W. (1979) Cell 18, 1209–1215.

- Bos, J. L., Ting, C. H., Borst, P., Arnberg, A. C. & van Bruggen, E. F. J. (1978) Nature (London) 275, 336-338.
- Smith, M., Leung, D. W., Gillam, S., Astell, C. R., Montgomery,
   D. L. & Hall, B. D. (1979) Cell 16, 753-761.
- Holland, J. P. & Holland, M. J. (1979) J. Biol. Chem. 254, 9839-9854.
- Montgomery, D. L., Leung, D. W., Smith, M., Bhalit, P., Faye,
   G. & Hall, B. D. (1980) Proc. Natl. Acad. Sci. USA 77, 541–545
- Seif, I., Khoury, G. & Dhar, R. (1979) Nucleic Acids Res. 6, 3387-3398.
- Breathnach, R., Benoist, C., O'Hara, K., Gannon, F. & Chambon, P. (1978) Proc. Natl. Acad. Sci. USA 75, 4853-4857.
- Fryberg, E. A., Kindle, K. L., Davidson, N. & Sodja, A. (1980) Cell 19, 365–378.
- Beckman, J. S., Johnson, P. F. & Abelson, J. (1977) Science 196, 205–208.
- Grunstein, M. & Hogness, D. S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961–3965.
- 22. Clewell, D. B. & Helinski, D. R. (1972) J. Bacteriol. 110, 1135-1146.
- Sherman, F. & Fink, G. (1974) Cold Spring Harbor Symp. Quant. Biol. 20, 77-80.
- 24. Herford, L. & Rosbach, M. (1977) Cell 10, 453-462.
- Maniatis, R., Jeffrey, A. & Kleid, D. G. (1975) Proc. Natl. Acad. Sci. USA 72, 1184–1188.
- Maxam, A. M. & Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
- Sures, I., Levy, S. & Kedes, L. H. (1980) Proc. Natl. Acad. Sci. USA 77, 1265-1269.
- Zimmern, D. & Kaesberg, P. (1978) Proc. Natl. Acad. Sci. USA 75, 4257-4261.
- 29. Southern, E. (1975) J. Mol. Biol. 98, 503-508.
- Smith, H. O. & Birnstiel, M. L. (1976) Nucleic Acids Res. 3, 2387-2398.
- Vanderkerckhove, J. & Weber, K. (1978) Proc. Natl. Acad. Sci. USA 75, 1106-1110.
- Brownlee, G. G. & Cartwright, E. M. (1977) J. Mol. Biol. 114, 93-117
- 33. Sanger, F. & Coulson, A. R. (1978) FEBS Lett. 87, 107-110.
- Baker, C. C., Herisse, J., Courtois, G., Galibert, F. & Ziff, E. (1979)
   Cell 18, 569-580.
- 35. Kozak, M. & Shatkin, A. J. (1978) Cell 13, 201-212.
- Kondel, D. A., Tilghman, S. M. & Leder, P. (1978) Cell 15, 1125–1132.
- Proudfoot, N. J. & Brownlee, G. G. (1976) Nature (London) 263, 211-214.
- 38. Berk, A. J. & Sharp, P. A. (1978) Cell 14, 695-711.
- Knapp, G., Beckmann, J. S., Johnson, P. F., Fuhrman, S. A. & Abelson, J. (1978) Cell 14, 221–236.