Characterization of unr; a gene closely linked to N-ras

M.Jeffers, R.Paciucci and A.Pellicer

Department of Pathology and Kaplan Cancer Center, New York University Medical Center, New York, NY 10016, USA

Received February 14, 1990; Revised and Accepted June 4, 1990

EMBL accession no. X52311

ABSTRACT

The mammalian N-ras gene is believed to play a role in cellular proliferation, differentiation, and transformation. While investigating N-ras, we isolated cDNA's that originate from a closely linked upstream gene. RNase protection assays reveal that this gene, unr, is transcribed in the same direction as N-ras and that its 3' end is located just 130 base pairs away from the point at which N-ras transcription begins. The close spatial relationship between the two genes is conserved in all species from which the N-ras gene has been isolated. An open reading frame, potentially encoding a 798 amino acid protein, is contained within the unr cDNA. Neither the primary protein structure nor the nucleic acid sequence of unr is homologous to any other known gene, including N-ras. Unr transcripts are detected in mouse, rat and human cells, and Southern analysis indicates that the unr locus found immediately upstream of the N-ras gene is transcriptionaly active in the mouse since only a single copy of unr is detected in this species. Unr produces multiple transcripts that differ in their 3' ends and are apparently created through the differential use of multiple polyadenylation sites located in the 3' untranslated region of the gene. Both unr and N-ras are expressed in all tissues examined. In the testis, both genes are developmentally regulated, with an increase in expression occurring upon testicular maturation. Thus the two genes may be coordinately regulated, at least in certain circumstances. Our findings suggest that a thorough analysis of the relationship that exists between the two genes could potentially provide insights into the regulation and/or function of N-ras.

INTRODUCTION

Ras genes comprise a gene family that is represented in all eukaryotic organisms from yeast to man, and whose members exhibit a high degree of interspecies homology at both the nucleic acid and protein levels (1, 2 for reviews). This finding suggests that *ras* proteins perform an essential cellular function. Although it is known that *ras* proteins localize to the inner surface of the plasma membrane (3) and are able to bind, exchange, and hydrolyze guanine nucleotides (4), their precise function remains

unknown. However, similarities in structure, biochemical properties, and intracellular location between *ras* and G proteins have been noted (5). G proteins are a class of proteins known to play a role in signal transduction and thus *ras* too is thought to be involved in this process whereby cells can undergo specific intracellular modifications as orchestrated by extracellular signals.

In mammals, three different functional ras genes (N-ras, c-Ki-ras-2, c-Ha-ras-1), and a number of ras-related genes have been identified (reviewed in 2). The protein products of the ras genes have been implicated in the processes of cellular proliferation, differentiation, and transformation. The proteins encoded by each of the three mammalian ras genes are very similar, but not identical, and thus are likely to exhibit some functional differences. Differences between the ras genes also exist at the level of gene expression. Although all mammalian ras genes are expressed in a wide variety of tissues and have promoters characteristic of housekeeping genes (6, 7, 8), they exhibit different patterns of expression during pre and postnatal development, and certain adult tissues preferentially express one of the ras genes over the others (9). The observation that the ras genes are differentially expressed is consistent with the hypothesis that the different genes may perform specialized functions and thus may be needed in different amounts at different times. It is important to investigate the mechanisms that regulate the expression of each of the ras genes, to define the factors responsible for their differential regulation, and to identify other functionally related proteins since such studies could provide insights that may lead to a more thorough understanding of the functions of the different ras genes. The study of ras gene expression is also important from the standpoint of cancer research since an increase in the expression of normal ras proteins has been shown to transform cells in culture (10, 11).

Our lab has been studying the N-ras promoter with the hope of identifying the elements that are important for its expression. Recently, an active transcription unit called *unr* has been discovered in the 5' flanking region of the N-ras gene (12). In this report, we describe the initial characterization of this upstream gene. The analysis of the relationship that exists between *unr* and N-ras may ultimately lead to a better understanding of the way in which the N-ras gene is regulated. In addition, the observation that these two genes are found closely linked in a variety of mammals (guinea pig, mouse, rat, human) raises the possibility that their protein products may be functionally related, as has been shown for the products of other linked genes (13).

MATERIALS AND METHODS

cDNA cloning

An adult rat testis λ gt11 cDNA library (generously supplied by A.R. Means, Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030), was initially screened by standard methods (14) using as the probe a mouse genomic N-ras 1.5 kilobase (kb) Pst-I/Sac-I fragment that spanned from the 5' flanking region of the gene to the beginning of the second intron. In the subsequent screen, the probe was a rat unr 1.5 kb cDNA fragment that had been obtained in the initial screen. cDNA inserts were removed from the phage by EcoRI digestion and run on 1% agarose gels. The gels were either subjected to Southern analysis, or were used to obtain cDNA fragments for subcloning into the PGEM-3Z plasmid (Promega).

Southern blotting

Southern blotting of high molecular weight DNA was performed as described by Southern (15). Electrophoresis was performed on a 1% agarose gel using 10 μ g of HindIII digested DNA/lane. Probes were labeled with [α -³²P] dCTP by random priming (Boehringer Mannheim). The blot was hybridized at 37°C and washed at 55°C to .2× SSC (1× SSC contains 150 mM NaCl and 15 mM sodium citrate at pH 7.2).

Northern blotting

Northern blotting was performed using total cellular RNA (14). Total RNA was isolated from the cell lines (figure 5) and from the mouse testis (figure 8) by guanidinium isothiocyanate lysis followed by centrifugation through a CsCl cushion (16). All other samples were isolated by the acid guanidinium thiocyanate-phenol-chloroform extraction method (17). Electrophoresis was performed on a 1.2% agarose-formaldehyde gel using 15 μ g of total RNA/lane. Probes were labeled by random priming. Blots were hybridized at 37°C and washed at 55°C to .2× SSC.

Primer extension

Primer extension analysis was performed following established protocols (14) using two oligonucleotides, one a 30 mer (5'-GGA-TCAAAGCTCATCTCGCAGTGATAATCG-3') which is the inverted complement of nucleotides (nt) 96–125 of the *unr* cDNA, and the other a 17 mer (5'-ATATTGCACTTT-CAGTA-3') which is the inverted complement of nt 78–94 of the *unr* cDNA. Each primer was end labeled with [γ -³²P] ATP, annealed to 5 μ g of poly(A)⁺ RNA at 30°C, and extended with Moloney murine leukaemia virus reverse transcriptase. Poly(A) ⁺ RNA was isolated by oligo(dT)-cellulose affinity chromatography (14) from total RNA that had been obtained by guanidinium isothiocyanate lysis followed by centrifugation through a CsCl cushion (16). The samples were analyzed on a 6% denaturing polyacrylamide gel.

RNase protection

To produce internally labeled antisense RNA probes, cloned fragments of the mouse N-*ras* gene and its 5' flanking sequence were inserted into vectors containing an Sp6 promoter (18). The 442 nt Ava-II/Ava-I probe complementary to the 3' region of *unr* and the 5' region of N-*ras* was produced from an Sp6-5 vector (Promega) that contained the Pst-I/Ava-I fragment (see figure 4B). This vector was linearized at the Ava-II site prior to transcription. The 291 nt Ava-II/Hinf-I probe complementary to the 3' region of *unr* was made from a PGEM vector that

contained the Ava-II/Hinf-I fragment. This vector was linearized at the polylinker BamHI site prior to transcription. The assays were performed according to the Promega protocol using total RNA that had been isolated by guanidinium isothiocyanate lysis followed by centrifugation through a CsCl cushion (16). Hybridization temperatures of $50-55^{\circ}$ C were chosen and digestion with RNase A and T1 was performed at $30-37^{\circ}$ C. The samples were analyzed on an 8% denaturing polyacrylamide gel.

Sequencing

cDNA fragments were subcloned into the PGEM-3Z plasmid and both strands of DNA were sequenced entirely. The chain termination method of sequencing (19) using the sequenase enzyme (United States Biochemical Corporation) was performed according to the manufacturer's protocol. Samples were analyzed on a 6% denaturing polyacrylamide gel. To determine the orientation of the 3 fragments produced by EcoRI digestion of the 2.5 kb unr cDNA, the phage DNA was digested with KpnI/Sac-I and a 3.4 kb fragment that contained nearly the entire intact cDNA insert was released. This fragment was either sequenced directly or subcloned into the Bluescript plasmid (Stratagene) for subsequent sequencing. By using appropriately selected synthetic oligonucleotides as primers, it was possible to sequence through the two internal EcoRl sites, and in doing so the orientation of the fragments that are produced by EcoRI digestion became apparent.

Computer analysis

The SOAP program, which is based on the method of Kyte and Doolittle (20), was used to plot the hydrophobicity along the protein sequence. The PSIGNAL program, which is based on the method of von Heijne (21), was used to locate possible secretory signal sequences in the protein. Both of these programs are licensed by Intelligenetics, Inc. Access to nucleic acid (Genetic Sequence Databank—Genbank; European Molecular Biology Laboratories—EMBL) and to protein (Protein Identification Research—PIR; Swiss-Prot.) data banks was obtained through the BIONET national computer resource for molecular biology.

RESULTS

Isolation and initial characterization of unr cDNA's

In the course of screening a rat testis cDNA library to obtain N-ras clones, we isolated a clone that contained a 1.5 kb insert that did not react with probes specific for either N-ras exon (-1)(which represents a noncoding exon that lies at the most 5' end of the N-ras mRNA) or N-ras exon (1) (which lies adjacent to exon (-1) and represents the first coding exon of the N-ras mRNA). This was puzzling since the probe that we used to screen the library spanned the area from the 5' flanking region to the beginning of the second intron of the murine N-ras gene, and thus all of the N-ras clones that we isolated should have contained either one or both of these exons. We subcloned the cDNA into PGEM-3Z and sequenced the insert. To our surprise, the 3' end of the polyadenylated clone was highly homologous (90-95%)over 450 bases) to the 5' flanking regions of the N-ras genes that had previously been isolated from the mouse (22), the guinea pig (23), and the human (6) (see figure 1). A polyadenylation consensus sequence (AATAAA) was present 16 nucleotides (nt)



Figure 1. Arrangement of rat *unr* cDNA's relative to each other and to N-*ras*. The 1.5 kb *unr* cDNA was isolated using a probe that contained mouse N-*ras* 5' flanking sequences. The last 450 bp of this cDNA (shaded region) are 90-95% homologous to the 5' flanking sequences of the N-*ras* genes isolated from several mammalian species (also shaded). The 2.5 kb *unr* cDNA was isolated using the 1.5 kb cDNA as a probe. This clone contains two internal EcoRI (Eco) sites which cause it to be released in three fragments from its phage vector upon EcoRI digestion. The arrow in the N-*ras* gene represents the location at which transcription begins in the mouse (see figure 4) and in the guinea pig (12). Although it is known that the last 450 bp of the *unr* gene and its cDNA are collinear (i.e. contains no introns), the intron/exon structure of the remainder of the *unr* gene is unknown. Only the cDNA's have been drawn to scale.

upstream of the poly(A) tail. Thus it seemed likely that an active transcription unit was positioned just upstream of the N-ras gene. By aligning the sequence of the cDNA with that of the 5' flanking region of the N-ras gene, it was evident that transcription from the upstream gene took place on the same strand of DNA as the N-ras gene, and gave rise to transcripts whose 3' ends terminated very close to the point at which N-ras transcription began. This transcription unit has also been detected by others and designated *unr* for upstream of N-ras (12).

The sequence of the 1.5 kb unr insert indicated that this clone represented only a partial cDNA since the 5' end of the clone contained an unfinished open reading frame. To obtain a longer clone, the rat testis cDNA library was rescreened using the 1.5 kb unr cDNA insert as a probe. From the screening of 375,000 plaques, about 100 positive clones (.03%) were identified. Some of the clones were chosen for further analysis and the most interesting one contained an insert of 2.5 kb which was released from the phage in three fragments of 1.5, .8, and .2 kb upon EcoRI digestion. Hybridization studies revealed that only the .2 kb fragment reacted with the original 1.5 kb unr cDNA probe. Sequencing revealed that the .2 kb fragment was completely homologous to the 5' end of the original 1.5 kb unr cDNA, while the other two fragments represented new information on the unr gene (see figure 1). The orientation of the fragments relative to each other was determined by sequencing through the junctions of the EcoRI sites present in the 2.5 kb insert (see sequencing section of materials and methods).

Analysis of the 5' ends of unr transcripts

In order to verify that our overlapping *unr* cDNA's represented a full length cDNA, primer extension analysis was performed using two independent primers (a 30mer and a 17mer) that were specific for regions close to the 5' end of the *unr* cDNA that we had isolated (see figure 2). Products of 118 nt using the 30mer and 87 nt using the 17mer were expected if the cDNA was indeed full length. The results obtained correspond quite closely with these predictions, with the 30mer yielding a product of 120-127nt and the 17mer giving rise to a product of 85-92 nt. Thus, we believe that the overlapping cDNA's that we have isolated represent a full length clone.

Analysis of unr cDNA sequence information

The combined sequence (3755 nt) of the two overlapping cDNA's that we have isolated is presented in figure 3. Studies on the



Figure 2. 5' end analysis of *unr* transcripts. Primer extension using two separate oligonucleotides was performed. One primer was 30 nt long and it corresponded to the inverse complement of nt 96–125 of the *unr* cDNA, while the other primer, a 17mer, corresponded to the inverse complement of nt 78–94 of the *unr* cDNA (see figure 3 for the location of the primers in the cDNA). The primers were end labeled, annealed to 5 μ g of Fisher rat fibroblast poly(A)⁺ RNA, extended with reverse transcriptase, and analyzed on a 6% denaturing polyacrylamide gel. A sequencing reaction run on the same gel provided molecular weight markers.

initiation of protein synthesis have revealed that translation begins at the most upstream AUG in 90–95% of all mRNA's examined, although sequences in the immediate vicinity of this AUG can influence it's translational strength (see 24 for review). From a survey of vertebrate mRNA's, a consensus sequence (GCCGCC(A/G)CC<u>AUG</u>G) for translation initiation has emerged. A purine located three nucleotides upstream of the AUG is the most highly conserved nucleotide in the vicinity of the start codon, and mutational analysis has shown that as long as there is a purine in this position, deviations from the remainder of the consensus sequence only marginally impair initiation. The first ATG in the *unr* cDNA is at nt 112, and continuing from this

		CAATTCCTTTTTTTAATACTAGCAACTTTCATTCTTTGGAACTTGTGCTGAAAAAAAGAAGAGTCAGCAAATACTACTGAAAGTGCAATATTCGATTATCACTGC														GAG	111																
set as pho app pro an iso jue his an angly his an git pt pro an git the res of a low or git to the git will be given that and the given of the given	ATG .	AGC	TTT	GAT	CCA	ANC	CTT	CTC	CAC	AAC	λλΤ	GGA	CAC	AAC	GGG	TAC	ccc	ANT	GGT	ACT	TCT	GCA	GCA	CTC	CGT	GAA	ACT	GGG	GTT	ATT	GAA	222	207
	met	392	phe	asp	pro		leu	leu	his	asn	asn	aly	his	asn	aly	tyr	Dro	asn	alv	thr	ser	ala	ala	leu	ara	alu	thr	alv	val	110	alu	178	
<pre>int let the ref tyr ety pair lie glin orginer glu erg glin ell erg lin pair pair glin erg glin gyr eng tyr eng tyr eng lin yr eng tyr eng</pre>	CTG	CTG	ACC	TCT	-	663	***	ATT	CAG	TOT	TCA	GAA	caa	CAA	GCT	ACC	CTT	TTC	TTC	CAC	TOT	TCA	CAG	TAT	AAT	660	AAC	CTC	CAG	arc	TTA		303
	leu	leu	thr	ser	tvr	aiv	phe	110	aln	CVS	302	alu	ara	aln	ala	Arg	leu	nhe	nhe	his	CVR	30T	aln	tvr		alv	A 80	leu	aln	4.80	leu	178	
<pre>vul by sep sep vul gil pies glu vul ser ser sep org seg the gly vup por lis til is zep in vul jer ling per grup ut gil is per grup sen glu lis per grup sen glu lis per sep set set sep org sen glu lis per vul sep sen glu sep seq glu glu vul sep set sen glu glu vul sep set and vul glu sep set and vul glu sep set sen glu glu vul sep set sen glu glu vul glu sep set y ut glu sep set sen glu glu vul sep set sen glu glu vul set org set sen glu glu vul glu sep set sen glu glu vul sen glu sep vul glu glu sep set sen glu glu vul sen glu sep vul glu glu sep set sen glu glu vul sen glu sep vul glu glu sep set sen glu glu vul set set sen glu sep vul glu glu sep set sen glu glu y lis sen glu sep vul glu glu sep sen glu sep vul glu glu sep sen glu sep vul glu glu sep sen sen glu glu vul set set set glu glu y glu sep set sen glu glu y vul sen glu sep vul glu glu glu glu glu glu sep sen glu sep vul glu glu sep sen sen glu glu vul set set sen glu glu y glu glu glu glu glu glu glu glu glu glu</pre>	CT1	663	GAT	GAT	CTC	61)	***	GAN	GTA	TCA	TCT	GAC	C00	ACC	ACT	000		CCR	1.00	CCT.	1.44		TTO	070	110	1-3 171		CC3	Ch h	1.42			300
ON DA COA NT AN T GU CAA OT THE TO COT OT COT CAC LAC THA GUA OT AN TOT COA OT COT COT COT COT CAC AND GUT OT THA 455 GU CU CAA ME HAN ON COT COT THA COT ANT THA COA CAT GUA AND ANT TOT COA COT GUA AND ANT THA COT AND THA 455 GU CU	val	alv	4.80	4.80	val	alu	nhe	alu	val	10n	807	4.80	ara	ara	thr	alv	178	nro	110	ala	110	178	200	va]	178	11.	1 1 2	nro	a) u	110	his	Dro	333
<pre>und und und und und und und und und und</pre>	Chh .		CO2	100		002	C	077	070	ROC	007			C) C		9-3 971	010	200			CC3		000		-2-	Ch0	100	003	202	000	ACT	CT.)	405
<pre>tec the def of the set is th</pre>	alu	ann. alu	ara	mat			<i>a</i> la	Va7	010 VA]			VA]	Dro	his		1.0	alu		100	101	Dro	a1a	a1a	Dro		ala		nro	thr		ngi ser	VA)	433
The first out of we show out with the first first first intervent is and any output set of a show of the set o	W-U	910 810				9-3 CCC	()) ())	CT.)			CRC	ACT	PLC.	200		Ch1	() T				110		C20	CTC.	9-J	100	CC3	chc		y-7			601
<pre>Gr U ye true the set is any and the true of the true of the true true true true true true of true</pre>	100	tur	a)	470		000	a1.	01A	nhe	+ + +	200	thr	tur	thr	101	alu		U 1 0	alu		A	v al		1.00		thr		Aen	100	114		nhe	331
USA AT GAT AAC AAT AAA CAC ACT GAT CAT AAG CUT CAT AAC TAT ING CUT TA AAA AAC AAA CAC AAT CAT AT CAT TAT CAA AAC AAT CAC TAT CAT AAG CAT TAT CAA AAA CAC TAT CAT AAA AAA CAT CAA AAT CAC TAT CAT C			910			9-3	y10									y10		~	y 20	y-3			y		920		y-y			~~~~		200	
Next step isn isn ip in isn ip in the optimal part of the step isn ip isn is its isn isn isn isn isn isn isn isn isn is	GTA .	ATT	GAT	AAC	AAT	2.00	CAC	ACT	GGT	GCT	GTA	AGT	GUT	CGT	AAC	ATT	ATG	LTG	TTG	1	ANG	7.00	-1-	GCT	CGC	TAT		GLA	GTA	GTT	TGT	GCC	98/
Are AM GA GA GA TH TAT GAN AND GAT GAN GT TH TAT AM AND AND TAT THE THE TAT AND GAN TAT TAT AND GAT GAN TAT AND AND AND TAT THE TAT AND GAN TAT TAT AND AND AND TAT THE GAN TAT AND AND TAT THE GAN AND TAT THE GAN AND TAT THE GAN TAT AND AND TAT THE GAN TAT AND AND TAT THE GAN AND AND TAT THE GAN AND AND TAT THE GAN TAT AND AND TAT THE GAN TAT AND AND TAT THE GAN TAT AND TAT AND AND TAT THE GAN TAT AND AND TAT THE GAN TAT AND AND TAT THE GAN TAT THE GAN TAT AND AND TAT THE GAN AND TAT THE GAN TAT AND AND TAT THE GAN TAT AND ANT THE GAN TAT AND AND THE GAN THE GAN TAT AND AND THE GAN THE GA			asp			170			9±9						• • • •	114		100	140			170	y III			- 71	910 910	913			cys		
<pre>min 1 ye ulu at part of y the life full any dy day at ye unit 1 ye ulu at a per part at by any lar ye dy at part of ulu at any dy any lar ye ulu at a per of ye ulu at any dy any lar ye any any any dy any lar ye ulu at any any lar any lar any lar any dy any lar ye ulu at any dy any lar ye ul</pre>	ATG .	AAG	GAG	GCG	TTT	GGC	TTT	ATT	GAA	AGA	GGT	GAT	GTT	GTA	1	GAG	ATA	TTC	TTT	CAC	TAT	AGT	GAA	TTT		GGT	GAT	CTA	GAA	ACC	CTG	CAG	783
Cer Ge AN GL GTG GAL THE ARA HE GLE ANA AN GET GAL ANA AN GET GAL ANA GHT GEA HAA GAL GTE ANT GET CET CAA GLA HEA GTE CET THA GAL GTF FO GU SA BAP VIS JOU DEN LET ILS 1/28 AND GAL ANA GTA ANC ANA GTA GEA ANA GHT GEA HAA GHT GTA HEAL GAL CAA THE ANA GAL CAA THE ANA JOU DEN LET IN THE GAL GU SA THA ANA THA ANA GTA ANA THA GAL CAA THE GAL CAA THE GAL ANA TTA GAL GAL THA ANA GAL ANA TTA GAL GAL ATA THA ANA GAL ANA TTA GAL GAL ANA TTA GAL GAL ANA GTA ANA THA GAL CAA ANA GTA ANA GAL GAL ANA TTA GAL GAL ANA TTA GAL GAL ANA TTA GAL GAL ANA GTA GAL ANA TTA GAL GAL ANA TTA GAL GAL ANA TTA GAL ANA GTA CAA ANA TTA GAL GAL ANA GTA GAL GAL GAL ANA TTA GAL GAL GAL ANA TTA GAL GAL ANA GTA GAL GAL GAL ANA TTA GAL GAL GAL ANA TTA GAL GAL GAL CAA CAA CAA CAA CAA CAA CAA CAA CAA C	Me C	1.7.2	gru	414	pne	g1y	pne	110	g t u	arg	gry	asp	Vel	Vel	173	ĝ. u	114	pne	pne	NT 8	cyr	361	gru	pne	198	₫1¥	asp	160	g1u	cnr	160	g1n	
pro gly app app val glu pae tar lie jy app arg ang gly jy glu val als tar app val arg tael garg of gly glu glu tar val lie pae glu app lie est lie glu his phe glu gly thr val thr jy val ite pro jy val pro est jy ang lie ang pro gly arg lie jy of AK TT GG ATT GCT AM GA GAT TT GT TA AG CA AM AK CC AM TA AC CGT GT GA AG GA CAT TA TA TT 1071. Val app phe val lie pro jy glu lew pro phe gly app jy app ite yai thr lew lew glu gly app his val arg phe san lie TA AC TH GG CA GC GA AG AM GA CTT GC TT GA AG AM AG CA AM AM CC AG TG GT GA AG GA AG CA AG GA AG GG GT GT 1167. TA GT GG CA GC GA AG AG AG GT TT AGC AG AG AG CA AT AM TA TA TG TA TA GA TC AG TA AG AG CA AG GA AG GA 1163 als als est arg apg jy lew glu berg als tar an lie glu val app arg ang at arg wei pe his yai arg glu est gly val 116 als als als est arg ang jy lew glu berg als tar an lie glu val app arg ang at arg wei pe his phe his phe san glu als arg glu ga glu ast gly val 116 als als als est arg ang jy luw glu berg als arg ang pe als arg wei pe his phe his phe san glu als arg glu pe his yai arg pro his yai arg als y glu ast gly ang glu 116 als als als apg luw glu phe thr val zer ga pe tis ure tar glu glu glu and thr AG ATT AG GA TT AC GA GA GA TT ACT GG GT CT GA AG AAA GA AG CA ATT AG ATT AGT ATT AG	CCT	GGA	GAT	GYC	GTG	GAA	TTC	ACA	ATC	AAG	GAC	λGλ	AAT	GGT	***	GVY	GTT	GCA	ACA	GAT	GTC	AGA	CTA	TTG	CCT	CVY	GGA	YCY	GTC	ATT	TTT	GAA	879
CAN TAY GAC ANT TAY GAN GAN ACT GTA ACC ANA GTA TAY CAC ANA GTA ACC ANA TAY GAN CAC AS TAY GAN CAC CAN THE GAN GAN GAN ATT TAY ANA CAC AGA TAY GAN GAN GAN CAC AGA AGA CAC AGA TAY GAN CAC AGA CAT AGA GAN AGA CAC AGA AGA CAC AGA TAY GAN CAC AGA CAGA CAGA CAG CAG	pro	дту	asp	asp	VAI	gru	pne	thr	116	178	asp	arg	a sn	đ 1À	178	g t u	VAI	818	thr	asp	VAI	arg	160	Ten	pro	đτυ	G 1À	CUL	VAI	116	pne	g 10	
<pre>ap lis er lis gu nis phe glu gy thr val the jys val lis pro jys val pro er jys an gap to lea pro gly arg lis jys fan cht tit ga ant cot AA GA CT CC TT GG GA AA AA CA AA AA AA CC AA TT CAA CT GA CT GA CAA CT AA CT AA CT GA CAA CAA CAA CT GA CT GA AA CAA CAA CAA CT CAA TAA CC TG GA AA CT AA CT CAA CT GA CT GA AA CAA CAA CAA CAA CAA CAA CAA CAA C</pre>	GAT	ATC	AGC	ATT	GAA	CAT	TTT	GAA	GGA	ACT	GTA	ACC	***	GTT	ATC	CCA	***	GTA	ccc	AGT	***	YYC	CAG	λλT	GYC	CCA	TTG	CCY	GGA	CGA	ATC	***	975
GTT GA ATT CCT MA GA CTT CCT TT GGA GAC AAA GAC AAA TCC MA GTG GAC GT GAC GT GAC GTT GT GAG GTT TA ATT INT 101 Val sep phe val ile projegius pictury property sep ile val the ise ile val glu gy sep his val ise phe sen ile 111 TCA ACA CAC CA GT GAC AAA TTA GAA CGA ACC AAC ATA GAG GTT CTA TCA ATA ACA TTT CGA TTC ACT AAT GAA GGA GAG GAG GTG II 112 TCA ACA CAC CAG CTA GAC AAA TTA GAA CGA ACC AAC ATA GAG GTT CTA TCA ATA ACA TTT CGA TTC ACT AAT GAA GGA GAC AGA GAG GAG GAG II 112 Ser Tha rap ser gar gas plu yie lau glu gal at far anni is glu wal is ser sen the phe phe his phe ser glu glu glu ser glu val i 113 ATT GCA CAT GAA GAA GGA GAT TAT GG GT CCT GAT ATG CTC TCT GC CAA AAA AAT CAT GCA TTT AGA ATT AAAA TCT CCAAT GGA GAT GAA GGA GAA CAAA 113 Sec GTT CAT TCC CAA TCA GA GT TTT CTG GG CT CCT GAT ATG CT CT GC CAA AAA AAA CAAT GCA TTT TCT AAT CCT AAA CCAAT GAA GC CAA ATA AAA AAA GA GCA TAT GCT TT CTA GAA CAAT GCA TT TAT GAA ACCAT GAA CTA AAA CCAAT GCAAT TAT ACT CAAT CAA	a sp	110	ser	110	g 1u	D18	pne	đτn	g 1 y	thr	VAI	CNL	178	VAI	116	pro	178	Val	pro	80 T	178	aan	g 10	aan	asp	pro	leu	pro	gly	arg	110	1ys	
<pre>val asp he val is pro jys glu aup pro he gly sep ifx iys sep ifx iys ser jys val ifr ieu aug du gly sep his val arg phe sen ile TCA ACA ACC CAC CA CTA ACA NT TA GAA CAC ACC ACA YA ACA ACA GAT CTA ATA ACA TTA TA TT CAA TTA CAA TCA ACA A</pre>	GTT	GAC	TTT	GTG	ATT	CCT	***	GAA	CTT	ccc	TTT	GGA	GAC	***	GAC	YCY	***	TCC	***	GTG	YCC	CTG	CTG	CYY	GGT	GAC	CAT	GTT	λgg	TTT	алт	ATT	1071
TCA ACK GAC GEA GET GAC AAA TTA GAA GEA ACC AAC ATA GAG GTT CTA TCA AAT ACA TTT CAC TAAT GAA GEC GAA GAG ATG GTT GTG IS ser tha sap ser arg sarg sarg sarg sarg lea burg sail a the sam is lie glu vail lie sarg and the part of the sam glu als arg glu sarg	val	a sp	phe	val	110	pro	1 ys	glu	leu	pro	phe	gly	a s p	1 ys	asp	thr	1 ys	80I	178	Val	thr	leu	leu	glu	gly	a sp	his	val	arg	phe	asn	11e	
<pre>ser thr ap arg arg arg is jug jug jug in gin arg als thr asn jie gin wil jug ars an thr phe gin phe thr asn gin als arg gin ast giy val Art GCT GCA KG AKG AKG AKG ATT GGC TT KGC TH ACK AKG ATG GG GA KG CC GA KGA AKT CAT GCA KTA GAA CCA AGA GA Jie als als ast arg arp giy phe giy phe jie jyr gy val asp arg arp als arg ast phe phe his phe ser gin jie bu arg giy asn gin CTC CAC ATT GCA GAT GAA GTA GAG TTA ACT GTG GTC CC GA ATAG CC CTA GGA AMA CAT GTC ATT GAG ATT AMA ACT CCC AMA GAA CTC CAC ATT GCA TTC CAT TCA CAT CAT CGT GTG GTC CC GA GAA AMA GAA GCC ACT TTT CTA ATC CT AMA ACT ACA AGC CCA AT AMA ACG ATT TCA TTC CAT TCC CAT TCA CAT CAT CGT TTT CTG GGC CAC GTA GAA AMA GAA GCC ACT TTT CTA ATC CT AMA ACT ACA AGC CCA AT AMA ACG AMA GCA AMA GCA GAA GAA GGC ATT ATT GCT TAT CAT GCC GTA GAA AMA GAA GCC ACT TTT CAT ACC AMA CAT ACA AGC CAA GAA CAT Jiy Jyr asp jyr giu als giu phe jis is als tyr asp asp gry giy val jyr ul yr ul thr lie als phe jin als jir asp val giu giy gir att TCT CTC CAA ATA GGG AGT AMA GTT GAA TTT AGT ATT ACT GAC AGG CCT GGA CAG CAG ATT GCT TG GTG GAA CAT TTA GGT CGT 1647 Thr ser pro gin lie giy asp jyr val giu phe ser lie ser asp jyr gin arg pro giy gin gil alls als thr gry val arg lee lee giy arg att TCT CTC CAA ATA GGG AGT ATA GCT GGA ACT CG GAA ACT CGG AGA CAT GTT ATT ATA AA ACAG CAG CAA TTA CTC ATA AGG AGA ATT TTC CAA AGT AAA CTC AAA GGT GAA TTT AGT ATT AGT AGT GGA ATT TTG GAT AGT ATA ATT AGA AGT TTA TTA</pre>	TCA .	ACA	GAC	CGA	CGT	GAC	YYY	TTA	GAA	CGA	GCA	YCC	AAC	λтλ	GAG	GTT	ста	TCA	AAT	YCY	TTT	CAG	TTC	ACT	λλт	GAA	GCC	AGA	GAG	ATG	GGT	GTG	1167
ATT GCT GCC ATG AGA GAT GGT TT GCC TTC ATC AGG TO GTG GAC GGT GAT GCT COT ATG TTC CAC TTC AGT GAG ATT CTC GAT GGG AAC CAG 1261 is als als asset arg asg gip yph egy phe ile jys gr val asg parg parg at arg met phe phe hip he ser giu lie leu asg gip yan gin CTC CAC ATT GCA GAT GAA GTG GAG TTT ACT GTG GTC CCT GAT ATG CTC TCT GCC CAA MGA AAT CAT GCT ATT AGG ATT AAA AAA CTT CCCC AAA GGG GA CTC CAC ATT GCA TTC CAT TCA GAT CAT GT GTC CCT GAT ATG CTC TCT GCC CAA MGA AAT CAT GCT AAT AGG ATT AAA AAA CTT CCCC AAA GGG CA GGG TT CAT TC CAT TCA GAT CAT CA GAT CAT CT GTT GTG GGG CAC GAA GGA AAT GAA AAT GAT CAT AAA AATA TAA CAAA GGA GA GAT GAT GGC AAA GAA GGA GAA GGA GAA GGA TA ATT AGT TAT GAT CAT CT GGG GTG AAA CTG ACT ATT GCT AAAA GGA ACT TTA GGT GAT GGC AAA GAA GGA GAA GGA GAT GAA TTA GAT TA GAT GAC CAT GGG GTG AAA CTG ACT TTT GCT AGA GAT TGT GTO AGA CTT TTA GGT GCT AATT CT CT CT CAA ATA GGA GAT GGA GTT GAA TTA ATT AGT GAC AAT GAA GAA GGA GAA TTG GCA AAT GAA AAG GA TT GT GTA GAA CTT TTA GGT GCA AAT GAA GAA GGC AATT GTT GTA GAA CTT TTA GGT GGA ATT GTT GCT AAAA GGA GTT GAT ATG GCA ACT CTG GGA GAA ATG GAA ATG AAT GAA GGA GTT GAA GGA ATT GTT G	80 <i>2</i>	thr	asp	arg	arg	asp	1 ys	leu	glu	arg	ala	thr	a sn	110	glu	val	leu	30 2	asn	thr	phe	gln	phe	thr	asn	glu	ala	arg	glu	met	gly	val	
<pre>ile als als est arg arg dy phe dig phe ile lys cys val arg arg arg arg arg arg met phe phe his phe met glu ile leu arg dy arm din CC CAC ATG GCA ATG AGT GCAC TTA CTG GTG CC GT AGT AGT CTG CTG CCAA AGA ATTA CGA ATA AGA ATA CCCCAAG GGC labbis ile als arg glu val glu phe thr val val pro arg met leu ser als glu arg arm his als ile arg ile lys lys leu pro lyg gly AGG GTT TCA TTC CAT TCG CAT TCA GAT CAT CGT TTT CTG GGC ACC GTA GAA AMA GCA ACT TTT TCT TAAT CCT AMA ACT ACA AGC CCAA TAA AA List thr val ser phe his ser arg his arg phe leu gly thr val glu lyg glu als thr phe sers and pro lys thr thr ser pro arm lys GGC AMA GCA GGA GGA GAA GAT GGT TTT GTG GGA CAC GTA GAA ATA GTG ACT ATT GT TTT CTAAT CCT TTAA GCC AMA GGT GGA ATT AGT GTA ATT AGT GAC AMA CGA GGA CTT GTG GTG GAA ATA GGG ATT ATT AGT TAAT AGT GAC AMA CGA GAC TTA GTT GTA CAA TTA GGT GGA ATT AGT GGA ATT AGT GGA AMA CTG ACT TTT GAT ACA CT AG GCT ATT AGT GGA ATT AGT GGA AMA CTG ACT TTT AGA CCA AGT ATA GGG GAT TTT AGT GGG AAA CTG GGA ATT ATT GAA ACA CGA GAT ATT AGT GGA ATT ATT AGT GGA ATT ATT</pre>	ATT ·	GCT	GCC	ATG	λgλ	GAT	GGT	TTT	GGC	TTC	ATC	λλg	TGT	GTG	GAC	CGT	GAT	GCT	CGT	ATG	TTC	TTC	CAC	TTC	AGT	GAG	ATT	CTC	GAT	GGG	AAC	CAG	1263
CTC CAC ATT GCA GAT GAA GTG GAA TTT ACT GTG GTC CCT GAT ATG GTC TT GTC CAA GAA ATA CTT GCA GTT GAA ATA ATA ATA CTT GCC AAG GGC 1359 leu his lie als asp glu val glu phe thr val val pro sapp set lau set als gin arg san his sla lie arg lie y lau pro lys gly AG GTT TCA TC CCAT TCA GAT CAT CGT TT TGT GGC ACC GTA GAA AN GAA GCC ACT TTT TGT ANT CCT AAA ATA ACA ACC CCA ATA AAA ATG GTT CAT TCC ATA TCA CAT CGT TTT GTG GGC ACC GTA GAA AN GAA GCC ACT TTT TGT AAT CCT AAA ATA ACA ACC CAAA ATA AAA AG GTA TCA TCC CAT TCA GAT GTA CAT TGT GTT TA GAT ACT TGT GGG GG GAAA ACT ACT ATT GGT TTT CAA GCA AG AT GTG AAG GAA TGT GGC ANA GAC GAG GAA GAA GGA ATT ATT GCT TA TGT TA TCA TCA TCG GTG GGG GTA AAA CGT ACT TTT GTG TGG GAGA ATT TTA CAC AAC GGT GTA ATT GCT CCT CAA ATA GGA GAT GAG GTT GAA TTT AGT ACT ACT GGG GG GAAA GTG ACT ATT GTG TGT GTG GA GAA CTT TTA GGT CAA ATT GCT CAA ATA GGA GAT GAG GGT GAA TTT AGT ACT AGT GAC AAA GAG AG GTT GAA CTA GCT ATT AG GGA CAT TTT TCC ATT TCC TAAA AGG CTC TG GGT TAA GTG GCA ACT CTG AAA GAT AAT TTT GGA TTT ATT GGA ATT GAT GA	11e	ala	ala	met	arg	asp	gly	phe	gly	phe	11e	1 ys	cys	val	asp	arg	asp	ala	arg	met	phe	phe	hi s	phe	ser	glu	<i>ile</i>	leu	a sp	gly	a sn	gln	
<pre>ieu his ile ala asp glu val glu phe thr val val pro asp met leu met ala gin arg ann his ala ile arg file lys lys leu pro lys gly AGG GTT TCA TTC CAT TCC GAT CAT CGT GTT TCG GGC ACC GTA GAA AAA GAA GCC ACT TTT TCT AAA CCT AAA AGA AGA AGA ATA TCA TTCA TTC CAT TCC CAT TCA GAT CAT CGT TTT CGT GGC ACC GTA GAA AAA GAA GCC ACT TTT TCA AGC AAA AGA AGA ATA AAA Isis GGC AAA GAC GAG GCA GAA GAT GGC ATT ATT GCT TAT GAT GGC GGG GGTA AAA CTA ATT GCT TTT CAA GCC AAG GAT GTG GAA GGA TCT Isig Jys asp lys glu ala glu asp gly lie lie ala is tyr asp asp gry gly val ys leu thr ile alp be gln ala lys asp val glu gly asr ACT TCT CCT CAA ATA GGA GAT AAT GGT GAA TTT AGT ATT AGT ATT AGT GAA GAG AAA CAG AGG CCT GGA CAG GAG ATT GCA GAA GCT TTA GGT GAA TTT CAA TTC CAA AGG CTT GTG GTT AAT TA GAT ATT GAA CAAA CA</pre>	CTC	CAC	λTT	GCA	GAT	GAA	GTG	GAG	TTT	ACT	GTG	GTC	CCT	GAT	ATG	CTC	TCT	GCC	CAA	λGA	AAT	CAT	GCT	λTT	AGG	ATT	***	***	CTT	ccc	λλg	GGC	1359
ACG GTT TCA TTC CAT TCA GAT CAT CTT TTT GTG GGA ACC GTA GAA ANA GAA GCA CTT TTT TCT AAT CCT AAA ACT ACA AGC CCA AAT AAA L455 Chr wal ser phe his ser his ser ap his arg phe leu gly thr wal glu lys glu als thr phe ser aan pro lys thr thr ser pro an lys GCC AAA GAA GAA GAA GAT GAA TTT GCT TTT CTA GAC GAT GTG GGA GTA AC TGA ACT TTT CAA GCC AAA GAA CTA GAA GAA CTA GJy Jys sep lys glu als glu als glu ap gly lie lie als tyr sep sep cys gly val lys leu thr ile als phe gln als lys sep val glu gly ser ACT TCT CTC CTA AATA GGA GAT AGG GTT GAA TTT AGT ATT AGT GAC GAA GAA CGA GAG GAT GGA GAA TTTA GGA GAA TTTT Har ser pro gln lie gly ap lys val glu phe ser lie ser sep lys gln arg pro gly gln gln lie als thr cys val arg leu leu gly arg AAT TCT AAT TCC AAA AGG CTC TTG GGT TAT GTG GCA ACT CTG AAA GAT AAT TTT GGA TTA ATT GAA CAA GTA AAT GAA TAA GGA GAA TAA GTA GAT TTTT CT ATA GTA GTA GTC TGT GGT GAT GTT GAT GAC ACT G GAA GAC CTG GAA GAC ATT GTA GAA CAA GCA TTG CTA AAA GGA GAA TAAA GTC ATT GT GGA AAA GTA ACT CAA CTA CAC CTG GAA CTT GAAA CTG GAA GAC AAT CCC ATT GT CT GTA AGC AAAA GGA AAA AAA GCA GAA TAAA GTC AGT GAA GGA AAA GTA AAA TAA CAAA GGA ATG AATT GAA GAA GAA GAA GAA ATA TTT GGA TTA ATT GGA TTA ATT GCT AAAA GCA AAA GCA CAAA CAC ACT GAA GAA GCA AAA CCC ATT GT GT GAA GTC ATA GTA GAA GTA AAA GAA GAA AAA CAA CAA GAA GTA GAA GTA GAA GTA GAAA CCC AAT CTA CTA GT GAA GGC AAAA GCA AAAA GAA GAA TAAA TTA GAA GAA GTA AAA GTA GAA TTAAAA GAA G	leu.	his	110	ala	asp	glu	val	glu	phe	thr	val	val	pro	asp	met	leu	ser	ala	gln	arg	aan	his	ala	110	arg	11e	1 ys	1 ys	leu	pro	1 ys	gly	
thr val ser phe his ser his ser asp his arg phe leu gly thr val glu lys glu als thr phe ser asm pro lys thr thr ser pro asm lys GGC AMA GGC AGG GGC GAG GGC ATT ATT GGC TAT GAT GGC GTG GAG GGG GAG ACG CAT ATT GGT TTT CAA GGC AMA GGA AGG ACT IJ31 GGC JU gly gly alg lu alg glu apg lys ell is lis als tyr asp asp cyg gly val lys leu thr lis als phe gla als lys asp val glu gly ser ACT TGT CGT CAA ATA GGA GAT AMG GTT GAA TT AGT ATT ATT ATT ATT AGT GAC AMA CAG AGG CGT GGA CAG CAG ATT GGC AGT GGT GAA GGA ATT TTT GGA ATT ATG GAC AAAT AGT GAG CAT TAG GG CAG CAG ATT GGA AGT GT TAG GG CAT TAG TG CGAA CT CTG AMA GTA AGT TAT TTG GAA CAG CAG CAG CAG ATT CGA AT CGA ATA GGA GAA TTT TTC 1743 sen ser asm ser lys arg leu leu gly tyr val als thr leu lys asp asm phe gly phe lie glu thr als asm his asp lys glu als phe phe 1753 AT TAT ATA GGA TG TGT CGT GAT GT GGA CAT CAT GAA GGA GAC TAT GAA TAC ACT CAT AGG AAA AGG AAA TTT TCC 1743 sen ser asm ser lys arg leu leu gly tyr val als thr leu lys asp asm phe gly phe lie glu thr als asm his asp lys glu asm lys val ser als 1755 his tyr ser glu phe ser gly asp val asp ger leu glu leu gly asp set val glu tyr ser lau ser lys gly lys gly sam lys val ser als 1753 1754 1755 1755 1757 175	ACG	GTT	тсл	TTC	CAT	TCC	CAT	TCA	GAT	CAT	CGT	TTT	CTG	GGC	ACC	GTA	GAA	***	GAA	GCC	ACT	TTT	TCT	AAT	CCT	***	ACT	усу	AGC	CCA	AAT	777	1455
GGC ANA GAC ANG GAC GCA GAA GAT GGC ATT ATT GCT TAT GAT GAC TGT GGG GTG ANA CTG ACT ATT GCT TTT CAA GCC ANG GAT GTG GAA GGA TGT gly jys asp jys glu ala glu asp gly ile ile ala tyr asp asp cyg gly val jys let the file ala phe gln ala jys asp val glu gly asr GTT CTT CCT CAA ATA GG ATT AGT TA ATT ATT ATT GAT CAA AN CAA GGC CTT GAA CAA CAA TT GCA CAT TGT GTG AAA CTT TAT GGT thr ser pro gln ile gly asp jys val glu phe ser ile ser asp jys gln arg pro gly gln gln ile ala thr cys val arg jeu leu gly arg AAT TCT AAA TCC AAA AGG CTT GTG GT TAT GTG GCA ACT CTG GAA ACT ATA GGA CAA TTT TGCA TTA ATT CC AAA AGG CTC TG GGT TAT GTG GCA ACT CTG GAA ACT ATA GGA CAA TTT TGCA TTA ATA CAA GCA ATA CGC AAA CGC AAA Sen ser asn ser jys arg jus leu jeu jyt yr val ala thr leu jys asp asn phe gly phe ile glu the ala asn lis asp jys glu is phe phe CAT TAT ACT GAA TC CCT GGT GAT GTT GAT AGC CTG GAA CTC ATG GGA GCA ATT CCT CAA AAA CGC AAA GGC AAT AAA CTC AG GCA AGG AAA GTA AAC AAA ACT CAC CAC GTG AAT GCC ATC CAT GAG GAA CAC TTG TGCA ATA CAC TTG TGC TAAA GGC AAT AAA CTC AG CAC TT glu jys val asn jys thr his ser val asn gjy ile thr glu glu asp ser her jus gly glu val lis arg pro isu arg gly val GAT CAA CA CAA CA CAA CG CAC GTG AAT GGC ATT GTG GAA GAG GGG GAA ATG TAA GAG GAA ATT GCA CAA CT CAA GGC ATA CAT GAA GGA ATG GAA GGA ATG GAA GTG GAA GTG TAA TGT GGA ATG GCC 2031 asp pro thr gln ile glu yr gln gly set ile glu ile val glu glu asp set ily gly glu val its arg pro isu arg gly val GAT CAA CAA CAA AAA GG GG GGA GTG GTG AAA CAC ATT GAA TTG CAA TTT GAA TAA CAA GAA AGA AAA CT AAA CAAC CCC CTT CGT AGG GCA TAC GC TAA GGA GAG GTG GAA TCC CAA TTG GT TAA TGAA GAA AGA AGA AGA AGA A	thr	val	ser	phe	his	ser	his	ser	asp	his	arg	phe	leu	gly	thr	val	glu	lys	glu	ala	thr	phe	ser	aan	pro	lys	thr	thr	80 T	pro	aan	lys	
gly lys asp lys glu als glu asp gly ile ile als tyr asp asp cys gly val lys leu thr ile als phe gln als lys asp val glu gly asr ACT TCT CCT CAA ATA GGA GAT AMAG GTT GAA TTT AGT ATT AGT GGA CMA CAG AGG CCT GGA CMG CMA GCT GAT GCA ACT TGT GGC AGA CTT TA GGT CCT The ser pro gln ile gly asp lys uglu phe ser ile ser ile ser als ag in arg pro gly gln gln ile als the rgy val arg leu leu gly arg AAT TCT AAT TCC AMA AGG CTC TTG GGT TAT GTG GCA ACT CTG GMA GAT AAT TTT GGA TTT ATT GAA ACA GCT AAT CAT GAT AMA GMG ATA TTT TTC TAT TCT AAT TCC AMA AGG CTC TTG GGT TAT GTG GCA ACT CTG GMA GMA GAT AAT TTT GGA TTT ATT GAA ACA GCT AAT CAT GAT ATA TTT GGA TAT TTT TC CAT TAT ATG GMA TCT CTG GGT TTT GAT AGC CTG GAA CT GGA GA CA CT GTT AA TA TA GAA CAC AGC AAT CGT CTA TCA TCA TCA GAT AAT TTT AGA TAT ATA TC GAA TCA CTG GT AAT TT AGT GA ACC GGA GA CA CT GTT AA TA CAA GCT AAT CAT CGT CTG TCA AAA GCC AT TAA TT GAA ACC AAA ACT CAC CAC AT GAA CTG GGA GAA CT CTG TTA ATA CAC AAA GCT AAT TCA TCA CTCA GGA AAA CCC ACA AAA CTCA CAC AGGA ATT GAA TGC ATA CAC AGGA ATT CAT TAC TTT GGT AAA GTC ATT CAC TCA CTG AGA GGC ATT GAG ACT ACT GGG GAA ATT CAT TAC TTT GGT AAA GTC AAT TGA GGA ATT GAA AGC AAAT GCC ACA ACA CAG AAAT CAC AAA ACT CAC ACA ACA ATT CAC AGGA ATT GAA AGGA ATT GAA AAA GTC AAT AAA ACT AAA ACT AAAA ACT AAAA AAT CAC AAA ACT AAAA ACT AAAA AAT CAC AAA ACT AAAA ACT AAAA AAT AAT	GGC	AAA	GAC	- AAG	GAG	GCA	GAA	GAT	GGC	ATT	ATT	GCT	TAT	GAT	GAC	TGT	GGG	GTG	λλλ	CTG	ACT	ATT	GCT	TTT	CAA	GCC	λλG	GAT	GTG	GAA	GGA	TCT	1551
ACT TOT COT CAA ATA GGA GAT AAG GTT GAA TIT AGT ATT AGT GAC ATA CAG AGG CCT GGA CAG CAG CAG ATT GCA ACT TGT GTG AGA CTT TTA GGT CGT thr ser pro gln ile gly asp lys wil glu phe ser ile ser asp lyg gln arg pro gly gln gln ile als the cys wel arg leu leu gly arg AAT TCT AAT TCC AAA AGG CT TTG GGT TAT GTG GCA ACT CTG AAA GAT AAT TTT GGA TTA ATT GAA ACA GCT AAT CAT GAT AAA GGA ATA TTT TCC TAT AGT GAA TTC CTA AA GGC CT TG GGT TAT GTG GCA ACT CTG GAA GAC ATG GTT ATA TGCA AACA GCC AAA CAT GAG GAA AAA TTT TCC AAT TCT AAT TCC AAA AGC CTC GTG GAT GTT GAA AGC CTG GAA CTG GGA AC TG GGA GAA CTG GTT GTA ATA CACA CTT TCC AAA GGC AAA AGC ATA AGT CGT Is tyr ser glu phe ser gly asp val asp ser leu glu leu gly asp met val glu tyr ser leu ser lys gly lyg dly ann lys val ser ala GGA AAA GTC ATC CAC GTG AAT GGC ACA CTG GGA GAA CTG GGA GGA GAA ATC CAC CTA CT CTG GT AAA GTC ATT GGT CAT CTG MAA GGC ATA AGT CAT GT GGA GAA GCA ATA CCA ACC ATG CGT CT CT AGA GAA GGC ATA AGT CAT GT GGA GAA GAA ATT GGA GAA GCA ATA CCA ACC TAC GTG GAA GTG ATA CCA ACC AAA GGA ATA ATA CAA GAA GAA TA TA GGA ATT GTA GGA GTG GAA CTG GTG CTA CAT GGA GGA ATT GTA GGA GGA GGA GAA GTC ATA CCT GCA CAA CAA CAA GAA AGT GAT GCA ACA CAAC ACA CAA GAA GTG TT ATA GAA GTG GAT GCC TAC AACA ATC ACA CAA CAA GAA AGG GAA AGT GTT AMA GTC CAA CTT GGA GGA GAA GAT GTT AMA GTC CAA CAA CAA CAA GAA GAA GAA GTT TA MA TTC CAA GTA GGA ATT GTA AMA TTC CAA GTA GGA TAT CAT TTA AGT GGA ATA GAA GTT TA MA GTA GAA TAT CAA TTA GGA ATT GAA GAA CAA TT GGA GT GTT ATA AGT CAA GAA CAA CT CT AGA GTA AGA CT GC TAA GAA TT GTA AMA GT GAA TT GAA GAA CAA TT GGA GT GTT ATA AGT CAA GAA CGA ATA CAAT TGGA GAA AGA GTC TT ATA AGT GGA GTT GTT AMA GAA GAA GAA TT AAG GAA AGA GTC CTT TTA GAA GAA CGA ATA CAA TT GGA GT GTT ATA AGT GGA GTT GTT AMA GAA GAA ATT TACA GAA AGA ATT GCA CTA AGA GTA GCA ATA GAA TT GGA GTA GT GAA GAA CAA GT GTT ATA AGT GGA GTT GTT ATA AGT GGA GTA GAA TT CAA GTA GGA GAA AAA GTA GAA ATA CAA CAA AGA GTC GTT TA AGA CAA GGA ATA GAA AGA GT GT GTT ATA AGA GAA GGA ATA GAA GA	aly	175	4.8D	175	alu	ala	alu	asp	aly	110	ile	ala	tyr	ASP	asp	cys	gly	val	175	leu	thr	110	ala	phe	gln	ala	175	asp	val	glu	aly	ser	
The ser pro gln lie gly app lys val glu phe ser lie ser app lys gln arg pro gly gln gln lie als the gys val arg lue u gly arg AAT TCT AAT TCC AAA AGG CTC TTG GGT TAT GTG GCA ACT CTG AAA GAT AAT TTT GGA ACA GCT AAT CAT GAT AAG GGA ATA ATT TTC asn ser asn ser lys arg leu leu gly tyr val als thr leu lys asp asn phe gly phe ile glu thr als asn his asp lys glu lie phe phe CAT TAT AGT GAG TTC TCT GGT GAT TG ATA GC CTG GAA CTG GGA GAC ATC GTT GAA TAC AGC TTG TCC AAA GGC AAA GGC AAA AGC CAT AT GT GAT AGC CTG GAA CTG GGA GAC ATC GTT GAA ACA GCT AAA GCC AAA GCC AAA GCC GAT AGT GAT AGC CTG TG GAA AGC GAAA GCC AAA GCC AAA GCC AAA GCC AAA GCC GAT AGT GAT AGC CTG CAC CAC GTT GAC AAC CTG GTC GAA GGC GAA TG GC ACT CAC TG GC AAA GTC ATT GGC ATC GTC GAG AGC GTT 1935 glu lys val asn lys thr his ser val asn gly lie thr glu glu glu als an pro thr lie tyr ser jlu ser glu glu gar gdly val GAT CCA ACA CAG ATT GAA AGC ATA GTT GGA GGA GGA GGA GAT GTC AGC TAT CCT GTT GGC ATA GTT GGA GAG GGA TG GCC 2011 asp pro thr gln ile glu tyr gin gly set lie glu ile val glu glu glu sap set lyg gly glu val tyr pro phe gly lie val gly set als ACC AAA GGG GAT TGC CTA CAG AAA GGC AAA GTT GGA GTA GAT TGT GGA CAT GAT GAA AGC AAA TGC CAA ATA CTA CAA ATA CAA CAA CTA CAA CA	ACT	TCT	CCT	CAA	- 	CC.	GAT		CTT	GAA	TTT	AGT	- 377	AGT	GAC	***	CAG	AGG	CCT	GGA	CAG	CAG	ATT	GCA	ACT	TGT	- 676	AGA	CTT	TTA	GOT	CGT	1647
ANT TCT ANT TCC ANA AGG CTC TTG GGT TAT GTG GCA ACT CTG ANA GAT ANT TTT GGA TTA ATT GAA CTA GCT ANT CAT GAT ANG GAG ATA TIT TTC asn ser asn ser lys arg leu leu gly tyr val ale thr leu lys asp asn phe gly phe lie glu thr ale sen his asp lys glu lie phe phe his tyr ser glu phe ser gly asy val asp ser leu glu gly asp met val glu tyr ser leu ser lys gly gly sgly asn lys val asr as leu glu ugly asp met val glu tyr ser leu ser lys gly gly sgly asn lys val asp asr leu glu gly asp met val glu tyr ser gly asp yer leu glu gly asp met val glu gly asp met val glu tyr ser gly asp yer leu gly gly yer last as gly lie thr glu gly asp met val glu gly asp met val glu gly asp met val glu gly asp met ale glu gly asp met val asn pro thr gln lie glu tyr gln gly met lie glu glu glu glu gly asp met lys gly glu val las asp gro thr gln lie glu tyr gln gly met lie glu glu glu glu glu gly asp met lys gly glu val tyr pro phe gly lie val gly met ale and CAA ACA GAG AT GAA TAC CAA GAG GAA GT GTA AG GTG GTG GTG GTG GTA GT GT GGC TAT GGC TAA CAAC ACC ACC ACC ACC ACC ACC ACC A	thr	 	nro	aln	110	alv	4.80	178	val	alu	nhe	307	110	ser		lvs	aln	ATG	pro	alv	aln	aln	110	ala	thr	CVS	val	Arg	leu	leu	alv	ara	
And her and her arg lev lev gly tyr val ala the lev lys ap and her her her her and her		TOT		TCC		3-3	CTC		COT	TAT	GTG	601	ACT	CTG		CAT			663	7-7 777	177	GAA	101	CCT		CAT	GAT	ANG	GAG	171			1743
CAT TAT AGT GAG TT CTT GGT GAT GT GAT AGC CTG GAA CTG GGA ACTG GTA GAT GAT GAT AAA GTC TAT GTC AAA GGC AAA GGC AAT AAA GTC AGT GAT GAT GAT GAT GAT GAT GAT GAT GAT		ser	480	200	100	ara	100	100	alv	tvr	va)	ala	thr	leu	100	4.50	880	nhe		nhe	110	alu	thr			his	4.80	140	alu	110	nhe	nhe	1/45
Chi in An die hie fei dei dei dei dei dei dei dei dei dei d	C)		100	C10			007	C. 1	9-3	017	100	CTG	()))	000	602	010	ATC.	077	())	***	100	990	-		000		000			080	207	007	1
AG AM GTA AM ACT CA AT CA GTA AT GGC ATC ACT GAG GA GGC ATT GTG GAG ATA CCC ACC ATC TAC TGT GGT GAA GTG TAT CGT CCT CTG AAA GGT GT ATT GAG ATA GGC ATT GTG GAG ATG GTA ATG AAA GGT GTA CTT TT GGC ATA GTT GGG ATG GGC ATT GTG GAG GGG GAT ATG AAA GGT GAT GTG TAT CAT GAG ATG GGC CCC 2031 asp pro thr gln ile glu tyr gln gly met ile glu ile val glu glu glu gly asp met lys gly glu val tyr pro phe gly ile val gly met ala AMC AM GGG GAT TGC CTA CAG AMA GGG GAG ATT GTG GAG GAG GGG GAT ATG AMA GGT GAA CTA ATG GGC TAA ATC AAC ACC CC 2127 asp pro thr gln ile glu tyr gln gly met ile glu ily val gly glu glu glu gly gap met lys gly glu val tyr pro phe gly ile val gly met ala AMC AM GGG GAT TGC CTA CAG AMA GGG GAG ATG GTT GAG TTC CAG TTG TGT GTA CTT GGC CAA ATT GAC AAC ATC AAC ACC CC 2127 cTT CGT AGG GAT TGC CTA CAG GAG GGT GTG AMA GAT CAG TTT GGC TTT ATT AAC TAT GAA GAA GAT AGC AAA AAG GCC TTT TTC CAT GTG AAA GAA GAA GAA GAA GAA GAA GAA CAC CAG GAA ATC AAC ACC CC 2139 cTT CGT AGG CAT CTG GG GAG GTG GTG GAA ATC CAG TTT GGC TTT ATT AAC TAT GAA GTA AGA AAA CAG CAC TTT TTC CAT GTG AAA GAA GAA GAA GAA GGC ATT GAG GCA GGA ATT GAG GGA AT CAG GGA ATT GAC GGA ATT GAC GCA GGA GAA GGA GAT CAG GGA ATT CTC GG GAA TCT CA GG ATT CAG CGC ACT GGA ATG AAC ACC CC GAA AGA AG GCC CTT ATT GGC CAAT GTT GG 2319 glu asp gly lie glu leu gln ala gly asp glu val glu wal glu asn arg leu lys asn ile thr leu asp asp ala ser ala pro cGT CTA ATG GTT CTT CGT CGA GCA GGA GAA CAA GAA ACT CGA CTG GAA AAA AAA AAA CAA CTA ACC CTG GAA ATT CAC TAA CCA 2511 arg leu asrg glu gly pro lys ala val ala ala pro arg pro asp arg leu val asn arg leu lys asn ile thr leu asp asp ala ser ala pro CGTCCAAAAGCACATCATTATACCACTAGGTCAATTGGGGGGAATTCGGGGAACACATTTAAAATTATCCCCTCGCAGAAAAATTATCTCGCCTGGAAAAATATTATCCACTGGGACATTATAACACTAGGCAATTATATCTCGCCTGGAAAAAATATATTCGCACAGAATTATATTCGCACAGAATTATATCTCGCCTGGAAAAAAATATATTCGCACAGAATTATATCCACTGGGCCATATGAAAAAAAA	hie	tur	807	ດໃນ	nhe	101		480	VA 7	4.80	ser	100	alu	1.00	alv	4.80	met	val	alu	tur	807	100	100	1 1 1	alv	148			140	va)		ala	1033
Gun And Bin Act and Act and Act and Act and act and been will als ash pro the file by set als of a set als are pro lew arg giv valGUN Jy val ash lys the his set val ash giv file the glu glu als ash pro the file by set glu ye val its arg pro lew arg giv valGAT CCA ACA CAG ATT GAA TAC CAA GGA ATG ATT GAG ATT GTG GAG GAG GGG GAT ATG ANA GGT GAA GTG TAT CCT TTT GGC ATA GTT GGG ATT GCC CAA AAC ACA CCA ACA ACA ACA ACA ACA	~~~		002			100	420		080		000	100	100	c>c	C	001			100	100				-30			3-3	000		202			1075
(a) (a) (b) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c		100	VAI	~~~	100	thr	bia	104	010			17.	the		~)		~~	nro	thr	110	tur	101	alu	100	UIC VA	110		Dro	200	A.F.G.		GTT VAJ	1322
AN COA ACA CAC ACT THE CAN THE CAN GRE ATT THE ATT THE GRE ATT ONE ATT	910					~~~					9-3			910	910	010		200	100		- 31		9-3					20			y-7		
asp pro thr gin file gil tyr gin giy met file gil file var gil gil giy asp met sys gil gil var fyr pro pro gil gil var gil gil met andAAC AAA GGG GAT TGC CTA CAG AAA GGG GAG AGT GTT AAG TTC CAG TTG GTG TGT GTG TGT GTC GCA AAAT GCA CAA ACT ATC GCC TAC AAC ACC ACC 2127asn lys gil asp cys leu gin lys gil gil user val lys phe gin leu cys val leu gil gil asn ala gin thr met ala tyr asn ile thr proCTT CGT AGG GCT ACT GTG GAG TGT GTG AAA GAT CAG TTT GGC TTT ATT AAC TAT GAA GTA GGA GAT AGC AAG AAG CTC TTT TTC CAT GTG AAG GAAleu arg arg ala thr val gilu cys val lys asp gin phe gily phe ile asn tyr gilu val gily asp ser lys lys leu phe phe his val lys giluGTT CAG GAT GGC ATT GAG CTA CAG GCA GAT GAG GTG GAA TTC TCA GTG AAC CTT AAT CAG CCC ACT GGG AAA TGT ACC GCC TGT AAT GTT FGGCGA GTT GTA AGC CCC AAA GCT GTT GCA GCT CCT CGA CCT GAT CGG TTG GTT AAT CAG CTT AAG AAC ACC ACT GAT GCC AGT GCT CAT CGA CCT GAA CTC ATG GGA TTT GGT GAA AGA AAG AAG AAC ACC ACC GGT GTC ATT GAC CAA GCT GCT CAA GCT CTT CGT CAG GGA CCA GAT AAC TCA ATG GGA TTT GGT GCA GAA AGA AAG AAG AAC ACC AGT GTC CAT GAC CAA CAA ACC ACC 2511arg val cys gilu gily pro lys ala val ala ala pro arg gily pro asp asn gar met gily ala gilu arg lys ile arg gin ala gily val ile aspCGT CTA ATG GTT CTT CGT CAG CCA AGG GGA CCA GAT AAC TCA ATG GGA TTT GGT GCA GAA AGA AAG AAG ATC CGT CAT AGC GGT GTC ATT GAC TAA CCAarg val cys gilu gily pro arg gily pro asp asn aser met gily phe gily ala gilu arg lys ile arg gin ala gily val ile aspCGT CTA ATG GTT CTT CGT CAG CCA AGG GGA CCA GAT AAC TCA ATG GGA TTT GGT GCA GAA AGA AAG AAG AAC ATC GGT GTC ATT GAC TAAC CAA 2511arg val cys gil no arg gily pro asp asn aser met gily phe gily ala gilu arg lys ile arg gin ala gily val ile aspCGTCTCACTACGCGTGTATGCTATATGCACTATGCACTTGGCCTGTGTAAATTATGCACTATAGGAGGGCCATTTAGGTGGAAAAAAAA	GAT	CCA	ACA	CAG	ATT	GAA	TAC	CAA	GGA	ATG	ATT	GAG	ATT	GTG	GAG	GAG	GGG	GAT	ATG	7.10	GGT	GAA	GTG	TAT	CCT	TTT	GGC	ATA	GTT	GGG	ATG	GCC	2031
ARC ARA GGG GAT TGC CTA CAG ARA GGG GAG AGT GTT TAG TTC CAG TTG TGT GTA CTT GGC CAA AAT GGA CAA ACT ATG GCC TAC AGC ACC CC 2127 asn lys gly asp cys leu gln lys gly glu ser val lys phe gln leu cys val leu gly gln asn ala gln thr met ala tyr asn ile thr pro CTT CGT AGG GCT ACT GTG GAG TGT GTA AG GTA CAG TTT GGC TTT ATT AAC TAT GAA GTA GGA GAT AGC ANA ACT ATG GCC TTT TTC CAT GTG AAG GAA 2223 GTT CAG GAT GGC ACT GTG GAA GTG TG GAA GAT CAG TTT GGC TTT ATT AAC TAT GAA GTA GGA GAT AGC AAG AGC TC TTT TTC CAT GTG AAG GAA GTT CAG GAT GGC ACT GAG CTA CAG GCA GAA GAG GTG GAA TTC TCA GTG ATC CTT AAT CAG CGC ACT GGG AAA TGT AGC GCC TGT AAT GTT TGG 2319 val gln asp gly ile glu leu gln ala gly asp glu val glu phe ser val ile leu asn gln arg thr gly lys cys ser ala cys asn val trp CGA GTC TGT GAA GGC CCC AAA GCT GTT GCA GCT CCT CGA CCT GAT CGG TTG GTT AAT CGC TTA AAG AAC ATC ACC CTG GAT GAT GCC AGT GCT CCA 2415 arg val cys glu gly pro lys ala val ala ala pro arg pro asp arg leu val asn arg leu lys asn lie thr leu asp asp ala ser ala pro CGT CTA ATG GTT CTT CGT CAG CCA AGG GGA CCA GAT AAC TCA ATG GGA TTT GGT GCA GAA AGA AAG AAG ATC CGT CAA GCT GGT GTC ATT GAC TAA CCA 2511 arg leu met val leu arg gln pro arg gly no asp asn ser met gly phe gly ala glu arg lys ile arg gln ala gly val ile asp CGTCCACATATAATCACACTATTAGTCACATTGGCGGAATTCAGGTGAAGGGTTCTGATAATATTATCTCCCCTCCCAAAATCTAATAGTGCAAAATTAAGTCAAATTAACTCAACCAT 2765 TATTTGGTGCAATTAACCACATGGTAATTGCAATTGCACTTGGTGAAGGGTTCTGGTGAAGGAGTACATTAAAGGTAGAATATTATATCTGAGACAAATTAATT	asp	pro	LAF	gin	110	gru	cyr	gin	gry		110	gru	114	V#1	910	910	919	asp	MUL.	178	919	910	Vel	Cyr ann	pro	pne	919	114	Vel	81¥			
221 If y fy sp cy sec cy is be of it iy gif gu ser val iy phe git he be cy val ieb gy git as all git the act of as it cy as it cy for an it could be cy val ieb gy git as an it git the act of the be	AAC	~~~	GGG	GAT	TGC	CTA	CAG		GGG	GAG	AGT	GTT	AAG	TTC	CAG	TTG	TGT	GTA	CTT	GGC	CAA	AAT	GCA	CAA	ACT	ATG	GCC	TAC	AAC	ATC	ACA	CCC	2127
CHT CCT AGG GCT ACT GTG GAG TGT GTG AAA GAT CAG TTT GCC TTT ATT TA AC TAT GAA GTA GGA GAT AGC AAA AAG AAG CTC TTT TTC CAT GTG AAG GAA 2223 leu arg arg ala thr val glu cys val lys asp gln phe gly phe ile asn tyr glu val gly asp ser lys lys leu phe phe his val lys glu GTT CAG GAT GGC ATT GAG CTA CAG GGA GGA GAT GAG GTG GAA TTC TA GTG ATC CTT AAT CAG CGC ACT GGG AAA CGC AGC GCC TGT AAT GTT TGG CAG GTC GAA GGC CCC AAA GCT GTT GCA GCT CCT CGA CCT GAT CGG TG GTT AAT CGC TTA AAG AAC ATC ACC CTG GAT GAT GCC AGT GCC AGT GCT CCA GCT CCA GCT GAT GGC TG GTT GAT GCC ATT GAC GTT GCA GCT CCT CGA CCT GGA TG GTG GTT AAT GGC TA AAG AAC ATC ACC CTG GAT GAT GCC AGT GCT CCA CCA GCT CCA ATG GTT CTT CGT CAA GCT GTT GCA GCT CCT CAA CCG GTG GTT AAT GGC GAA AGA AAG AAG AAG AAC ATC ACC CGG GT GTC ATT GAC TAA CCA 2511 arg leu met val leu arg gln pro arg gly pro asp asn ser met gly phe gly ala glu arg lys ile arg gln ala gly val ile asp CGTCCACAAAGCAACTCATTAAACCAACTGGGCAGGTTCTGGGGAAGGGTTCTGAAAGCAAATTATGCACCAGTTGGCGCCTTATAATGGCAAATTAAGGAACCAATTTAAAGAACCAATTTAAGAAACAATTTAAGGAACCGATTGGTGCACCTATGGTGAAATTGGCAACTTAAGGTGGCGCCATTATAAGGCAAAAGAAAAAAAA	asn	172	g1y	asp	cys	160	g 1U	172	g⊥y	gru	361	Val	173	pne	gin	Ten	cys	Vel	Ten	gry	g TU	asn		gin	LAF			cyr	ean	110	CAF	pro	
Ieu arg arg arg the val glu cys val lys apg gin phe gly phe lie ash cyf glu val gly apg ser lys lys leu phe phe his val lys glu GTT CAG GAT GGC ATT GAG CTA CAG GCA GGA GAT GAG GTG GAA TTC TCA GTG ATC CTT AAT CAG CGC ACT GGG AAA TGT AGC GCC TGT AAT GTT TGG Val gin asp gly ile glu leu gin ala gly asp glu val glu phe ser val ile leu asn gin arg the gly lys cys ser ala cys asn val trp CGA GTC TGT GAA GGC CCC AAA GCT GTT GCA GCT CCT CGA CCT GAT CGG TG GTT AAT CGC TTA AAG AAC ATC ACC CTG GAT GAT GCC AGT GCT CCA arg val cys glu gly pro lys ala val ala ala pro arg pro asp arg leu val asn arg leu lys asn ile the leu asp asp ala ser ala pro CGT CTA ATG GTT CTT CGT CAG CCA AGG GGA CCA GAT AAC TCA ATG GGA TTT GGT GCA GAT ACC CGT GTA AAG AAC AAC CGT GTG GTC AAT GCT AA CCA arg val eys glu gly pro lys ala val ala ala pro arg gly pro asp asn ser met gly phe gly ala glu arg lys ile arg gin ala gly val ile asp CGTCCACAAAGCACATCATTAAACACATTGGTCAAGTTGGGGGGGATTCTGGTGAAGGGTTCTGAATATATTCTCCACTCATCGCCTCCCAAAATTAAACTTAAAAAAAA	CTT	CGT	AGG	GCT	ACT	GTG	GAG	TGT	GTG	***	GAT	CAG	TTT	GGC	TIT	ATT	AYC	TAT	GAA	GTA	GGA	GAT	AGC	AAG	AAG	CTC	TTT	TTC	CAT	GTG	ANG	GAA	2223
GTT CAG GAT GGC ATT GAG CTA CAG GCA GAT GAG GTG GAA TTC TCA GTG ATC CTT AAT CAG CGC ACT GGG AAA TGT AGC GCC TGT AAT GTT TGG 2319 val gin asp gly ile glu leu gin ala gly asp glu val glu phe aer val ile leu asn gin arg thr gly lys cys aer ala cys asn val trp CGA GTC TGT GAA GGC CCC AAA GCT GTT GCA GCT CCT CGA CCT GAT CGG TG GTT AAT CGC TTA AAG AAC ATC ACC CTG GAT GAT GCC AGT GCT CCA arg val cys glu gly pro lys ala val ala ala pro arg pro asp arg leu val asn arg leu lys asn ile thr leu asp asp ala aer ala pro CGT CTA ATG GTT CGT CAG CCA AGG GGA CCA GAT AAC TCA ATG GGA TTT GGT GCA GAA AGA AGA ACT CGT CAA GCT GGT GTC ATT GAC TAA CCA 2511 arg leu aet val leu arg gin pro arg gly pro asp asn aer aet gly phe gly ala glu arg lys ile arg gin ala gly val ile asp CGTCCACAAAGCACATCATTAATGCCAATGGTCAAGTTGGGGGGGATTCTGGTGAAGGGTTCTGAATATATTATGCACAGCAGTTGGAGGCGGCCTTATAGGTGCAATTTAAGAAAAACAAAAACAAAAACAGACTGTTACTTGTGGCCCAAAATTTAAAGAAAATTAAATTTAAGAAAAACAAAAACAGAACTGTTACTTGTTGGCCCAAAATTGGAAGCAGTTATATAACCTATTAGGAAGTTTTAAAAAAAA	Ten	arg	arg	a1a	thr	VAI	giu	cys	val	178	asp	дти	pne	g1y	pne	116	asn	cyr	g 1u	Val	g 1y	asp	30 I	173	178	100	pne	pne	n1 8	VAL	198	giu	
Val gin sep gly ile glu leu gin ale gly sep glu val glu phe ser val ile leu sen gin arg thr gly lys cys ser ale cys aan val trp CGA GGC TGT GGAA GGC CCC AAA GGCT GTT GGA GGT CCT CGA CGG AT CGG TG GGT TGAT CGG TTA AAG AAC ATC ACC CTG GAT GAT GGC AGC GGT CCC CA arg val cys glu gly pro lys ale val ale ale pro arg pro asp arg leu val asn arg leu lys asn ile thr leu asp asp ale ser ale pro CGT CTA ATG GTT CTT CGT CAG CCA AGG GGA CCA GAT AAC TCA ATG GGA TTT GGT GCA GAA AGA AAG ATC CGT CAA GCT GGT GTC ATT GAC TAA CCA 2511 arg val cys glu gly pro lys ale val ale ale pro arg pro asp arg leu val asn arg leu lys asn ile thr leu asp asp ale ser ale pro CGT CTA ATG GTT CTT CGT CAG CCA AGG GGA CCA GAT AAC TCA ATG GGA TTT GGT GCA GAA AGA AAG ATC CGT CAA GCT GGT GTC ATT GAC TAA CCA 2511 arg leu set val leu arg gln pro arg gly pro asp asn asr set gly phe gly ale glu arg lys ile arg gln ale gly val ile asp CGTCCACAAAGCACATCATTATATCCACTATGGTGAAGGGGGATTCTGGTGAAGGGGTTCTGATATATAT	GTT	CAG	GAT	GGC	ATT	GAG	CTA	CAG	GCA	GGA	GAT	GAG	GTG	GAA	TTC	TCA	GTG	ATC	CTT	AAT	CAG	CGC	ACT	GGG	***	TGT	AGC	GCC	TGT	AAT	GTT	TGG	2319
CGA GTC TGT GAA GGC CCC AAA GCT GTT GCA GCT CCT CGA CCT GAT CGG TTG GT AAT CGC TTA AAG AAC ATC ACC CTG GAT GAT GCC AGT GCT CCA 2415 arg val cys glu gly pro lys ala val ala ala pro arg pro arg pro arg arg leu val asn arg leu lys asn ile thr leu asp arp ala ser ala pro CGT CTA ATG GTT CTT CGT CAG CCA AGG GGA CCA GAT AAC TCA ATG GGA TTT GGT GCA GAA AGA AAG AAG ATC CGT CAA GCT GGT GTC ATT GAC TAA CCA arg leu met val leu arg gin pro arg gly pro arg gly pro asn ser met gly phe gly ala glu arg lys ile arg gin ala gly val ile asp CGTCCACAAGCACATGATTATATCACTGATGGTGCAAGTGGGGGGATTCTGGTGAAGGGTTCTGAATATCTCTCTC	VAL	gin	asp	gly	110	glu	leu	gin	a]a	gly	asp	glu	val	glu	phe	ser	val	110	leu	asn	gin	arg	thr	gly	178	cys	80T	ala	cys	aan	Val	trp	
arg val cys glu gly pro lys ala val ala ala pro arg pro asp arg leu val asn arg leu lys asn ile thr leu asp asp ala eer ala pro CGT CTA ATG GTT CTT CGT CAG CCA AGG GGA CCA GAT MAC TCA ATG GGA TTT GGT GCA GAA AGA AGG ATC CGT CAA GCT GGT GT ATT GAC TAA CCA arg leu met val leu arg gln pro arg gly pro asp asn ser met gly phe gly ala glu arg lys ile arg gln ala gly val ile asp CGTCCACAMAGCACATCATTATATCCACTATGGTCAAGGTGGTGGAGGGTTCTGGAAGAGGTTCTGAATATCTCCCCCCCACAAACCTGAAATATCTAATTATATATA	CGA	GTC	TGT	GAA	GGC	CCC	YYY	GCT	GTT	GCA	GCT	CCT	CGA	ССТ	GAT	CGG	TTG	GTT	ANT	CGC	TTA	AAG	AAC	ATC	ACC	CTG	GAT	GAT	CCC	AGT	GCT	CCA	2415
CGT CTA ATG GTT CTT CGT CAG CCA AGG GGA CCA GAT AAC TCA ATG GGA TTT GGT GCA GAA AGA AAG ATC CGT CAA GCT GGT GTC ATT GAC TAA CCA 2511 arg leu met val leu arg gin pro arg giy pro asp asn ser met giy phe giy als giu arg lys ile arg gin als giy val ile asp CGTCCACAAAGCACATCATTAATCCACATAAGGTCAAGTGGGGGGGATCTGGTGAAGGGTTCTGAATAATCTTCCTCTTCATCCCCCAAAAATCTGGGAAAACTTAATCTAATCAAAGGGTAATAACACAAG TTTTAACACCTTCCCTGGTGATATGGCAAAGTGGGGGGGATTCTGGTGAAGGGTTCTGAATAATCTTCCTCTTCATCCCCCCAAAAATCTAGGGCCTAATAGGTGCCAAATTAAGGAACCAATTTAAGGAACCAATTAAGGAACCAATTAAGGAACCAATTAAGGAACCAATTAAGGAACCAATTAAGGAACCAATTAAGGAACCAATTAAGGAACCAATTAAGGAACCAATTAAGGAACCAATTAAGGAAGCCAAAAGGAAGG	arg	val	cy s	glu	gly	pro	1 ys	ala	val	ala	ala	pro	arg	pro	asp	arg	leu	val	aan	arg	leu	178	aan	110	thr	leu	a sp	asp	ala	80 <u>r</u>	ala	pro	
arg leu met val leu arg gin pro arg giy pro asp asn ser met giy pho giy als giu arg iys ile arg gin als giy val ile asp CGTCCACAAAGCAACATCATTAATCCACTATGGTCAAGTGGGGGGATTCTGGTGAAGGGTTCTGAATATCTTCCTCTCATCCACCCCCCAAAAATCTGGAATACTTAATGCACATTTAAGGTGGCCACTTATAGGTGCACCTTAAGTGTCCAATTAGGTGCCACTTAAGTGTCCAATTAGGTGCCACTTAAGTGTCCAATTAGGTGCCCCTAAAAAAAA	CGT	стл	λtg	GTT	CTT	CGT	CAG	CCA	AGG	GGA	CCA	GAT	уус	TCA	ATG	GGA	TTT	GGT	GCA	GYY	λGλ	ANG	ATC	CGT	CYY	GCT	GGT	GTC	ATT	GAC	тал	CCA	2511
CGTCCACAAAGCAACATCATTAATCCACTATGGTCAAGTTGGGGGGATTCTGGTGAAGGGTTCTGAATATCTTCCTCTTCATCCACCCCCAAAAATCTGGAATATCTTAATTAA	arg	leu	met	val	leu	arg	gln	pro	arg	gly	pro	asp	asn	80 Z	met	gly	phe	gly	ala	glu	arg	1 ys	11e	arg	gln	ala	gly	val	110	asp			
ITTTAACACCTTCCTCGGGTTAAGTTAAAGAAAAAAAAAA	CGTC	CAC	AAAG	CACA	TCAT	TAAT	CCAC	TATG	GTCA	AGTT	GGGG	GGAT	TCTG	GTGA	AGGG	TTCT	GAATI	ATCT:	TCCT	CTTC	ATCC	CTCC	CVVV	ATCT	GGAN	INCT	TATT	CTAT	FGAG	CTAT:	ГЛСЛ	CAG	2638
TATTTGGTGCATTTAACGCAACTGGTAATTGCAAATTCCACTTGGCCTGTGTAAGTGAACAAAACAGACTGTTACTTTGTTGGCCCTATGAAATTCTGCACTCTAGTCAGTATATATA	TTTT	AAC	ACCT	TCCT	CGTG	TTAT	GTTT	AAAG	,,,,,	лата	ATA	<u>AA</u> TT	TAAG	AAAC	CATT	ттал	NATT/	ATGC	ACAG	TTGC	NGCC	TGGA	ллас	ттал	GGTG	CCC	CTTA	TAGT	BTCA	ATTI	AGGA	CTT	2765
CATGACTTAATACTCAGCTGCATTCTTTTCCTTTTTCCTTTATACCCTCTGTTCATTATGCTTTAATTCTGAGGACCATATGAAGGTAGAATATATTACCTTTAAAAATTGCAAAATTAATAAGGAAACGAATT 3019 TTCTTAAGTTGATGGCCCAAATGTTAATTTCATATATTCATATAAACTTGTCACAATCCACTTAACGAACG	TATT	TGG	TGCA	TTTA	ACGC	AACT	GGTA	ATTG	CYYY	TTCC	ACTT	GGCC	TGTG	TAAG	FGAN	CAAA	ACAG	ACTG	TTAC	TTTG	TTGG	CCCT	ATGA	AATT	CTGC	ACTC	TAGT	CAGT	TAT	ATAC?	ICTN	CTT	2892
TTCTTAAGTTGATGGCCAAATGTTAAATGTTAATTTCATATCATTAATAATCTTGTCACAATCCACTTAACGAAGTTTGATTACATTCAGTGAAAAATATCTTCCATAAAAAAGGGTTTTTATTAAAAAATGTGCACAAATGAAATGCACAAATGCACAAATGAAAAAAAA	CATG	ACT	TAAT	ACTC	AGCT	GCAT	TCTT	TTCC	TITT	TTCC	CTCT	GTTC	ATTA	IGCT	ГТАЛ!	TTCT	GAGG	ACCAS	TATG	AAGG	TAGA	ATAT	ATTA	CCTT	LYYY	AATT	GCAN	MTT	ENTN:	IVOCI	AAC	IATT	3019
TTCCTGGGATGGGGGGCCAAATACTTTACTACAATTAGTATGTAT	TTCT	TAN	GTTG	ATGG	CCAA	ATGT	ТЛЛЛ	ATGT	TATT	TTCA	TATC.	ATTT	ATAA'	ICTT	GTCA	CANT	CCAC	ГТАЛ	CGAN	GTTT	GATT	ACAT	TTCA	GTGA	AAAT	TATC	TTCC	ATAA	LYCC	FTTT?	TTT	TTCC	3146
GGTICLIGUTIANINGLALIULAUCTICTAUGACUCAUTITIACTUAUTITAAAAAAAAAAAAAAAAAAAAA	TTCC	TGG	GATG	GGGG	GCAA	ATAN ANAC	CTTT.	ACTA	CAAT	TAGT.	ATGT.	ATGG	TGCA	GAAT	TTCA	TGCA	ATG	AGT	GTGC	CAGC	AGTG	TGAT	AATT	TAAA	CATA	TTTN	ACA	AAAG	ACGA	ATGC	ACAN	NATT .	3273
GGCAAGTGGAAGGCTTTTTGCTTTTTGCTTAAAGAAGGCGAAGAAGGGGGGGG	GCTG	CTU	CTTA CTTA	WATC CORO	RUTG PPO>	CAGC ATLA	ACCC	7.66A	CLCCA	GTTT CTTC	GTTT GTOO	IACT	WATT'	****	ACR	8682		2002	ACCO	ATAA		GITG	TGCC	1 GAA	nTGA PP/	ATCC	CAR	TTT	TAT	ANGT/	NGCC	CCT	3400
TTTGTAAACACAAAACATTTTGCCTTTCTCCCGGTTCAATGGCGAAAAGAAGGCGAAAGAAGGCGAATAAAATTTTCCTGAATTTCCAAAAAAAA	GCC1		CC70 C	uu tu Tall				NUTT NN10	ATCA	CANG		ACCT	CACC	 20 0 00	CT CC	LOIA GGC1			TCAT	61 I UI	CCCC	WATT TTAT	LYCE TOLL	996CC 7620		TGAP		106C		ACOC	- Cas	PATC	3327
	TTTG	TAN	ACAC	~~~~	CATT	TTTG	CTTT	CTCC	GGTT	TCAT	GTTA	ATGG	CGAN	AGAA	IGGA	AGCG	AATA	AGT	TTTA	CTGA	ITTT	TGAA	~~~	AAAG	GAAT	TC							3755

Figure 3. The nucleotide and deduced amino acid sequence of the rat *unr* cDNA. Nt 2263-3755 are derived from the original 1.5 kb cDNA and the remainder of the sequence from the 2.5 kb cDNA that was subsequently isolated. The EcoRI sites (GAATTC) at the ends of the sequence are from a synthetic polylinker that was ligated to the cDNA's during cloning. The three polyadenylation sites in the 3' untranslated region of the gene that match the (AATAAA) consensus signal are underlined. Polyadenylated cDNA's during the signals at nt 3348 and nt 3718 have been isolated. For the cDNA shown, the signal at nt 3718 was used and the tract of (A's) from nt 3741-3748 represents a short poly(A) tail. Another cDNA utilizing the nt 3718 polyadenylation signal but differing in the location at which the poly(A) tract was added has also been isolated. This cDNA ended with the sequence (GATTTTTGAGACACTAAAAAAAGGAATTC). The locations of the primers that were used for primer extension analysis are indicated with arrows.

point is the largest open reading frame in the *unr* cDNA, extending to nt 2505. This ATG in the *unr* cDNA contains a purine (guanosine) at the appropriate upstream location, and thus is likely to be competent for translation initiation. An in frame stop codon (beginning at nt 52) exists upstream of this ATG.

If, as would be predicted, the first ATG is used to initiate translation in the *unr* cDNA, then the cDNA sequence predicts an 111 nt 5' untranslated region followed by a 239 nt open reading frame, and a 1250 nt 3' untranslated region. The protein encoded by the open reading frame is 798 amino acids long with a



Figure 4. Determination of the *unr/N-ras* boundary. An RNase protection assay using two different probes was performed in (A) and the probes used are shown in (B). The probes were derived from the mouse N-ras gene, and its 5' flanking region (which contains part of the *unr* gene). The AvaII/AvaI probe is a 442 nt fragment covering the 3' region of *unr* and the 5' region of the N-ras gene. The AvaII/Hinfl is a 291 nt fragment covering the 3' region of the N-ras gene. Total RNA obtained from a C57BL/10 thymoma cell line was used with the AvaII/AvaI probe and normal murine thymocytes from a CD-1 hybrid mouse were the source of the RNA used with the AvaII/Hinfl probe. The products were analyzed on an 8% denaturing polyacrylamide gel.

predicted molecular weight of 88,894. Overexpression of the *unr* cDNA in COS cells yields a protein product of the predicted size upon immunoprecipitation with anti-*unr* antibodies (data not shown). Computer assisted analysis has indicated that the protein does not possess any long stretches of hydrophobic amino acids and thus is probably not secreted or integrally associated with a membrane. A search of nucleic acid and protein data banks revealed no significant homology between *unr* and other known genes or their protein products.

The unr/N-ras spatial relationship

In order to accurately define the spatial relationship that exists between *unr* and N-*ras*, we performed an RNase protection assay using fragments of the mouse N-*ras* gene and its 5' flanking region as a probe to analyze *unr* and N-*ras* transcripts from mouse cells (figure 4). This technique allowed us to delineate the 3' end of the *unr* transcription unit and the 5' end of the N-*ras* gene in the mouse. The first probe that was used covered both genes, and two major groups of protected fragments were detected (see figure 4A; AvaII/AvaI probe). We believed that the lower molecular weight group of fragments (80–100 nt) was being protected by N-*ras* transcripts, and that the larger molecular weight group of fragments (214–235 nt) represented *unr*-



Figure 5. Unr expression in mouse, rat, and human cells. 15 μ g of total RNA from each sample were electrophoresed on a 1.2% agarose-formaldehyde gel, transferred to nitrocellulose, and hybridized to a *unr* probe which contained the entire *unr* cDNA depicted in figure 3. The blot was washed to .2 × SSC at 55°C and exposed for 7 hours. The thymus sample is from a CD-1 hybrid mouse. Mouse 3T3 is a fibroblast cell line; NRK is a normal rat kidney fibroblast cell line; HeLa is a human epithelial-like tumor cell line.

protected fragments. To verify this hypothesis, the probe was shortened at its N-ras proximal end such that it should no longer be protected by N-ras transcripts. As predicted, the lower molecular weight group of protected fragments was not detected with this probe, while the higher molecular weight group of fragments was still detected (see figure 4A; AvaII/HinfI probe). Similar results were obtained when this experiment was repeated on mouse fibroblast RNA (not shown). When the 442 nt AvaII/AvaI probe was used, the major unr-protected fragment was 230 nt in length and the major N-ras protected fragment was 82 nt in length. Thus in the mouse, the 3' end of unr is located 130 nt (442 - (230 + 82)) away from the point at which N-ras transcription begins. This experiment also verified that unr and N-ras are transcribed from the same strand of DNA since transcripts from both genes protected the same single stranded probe.

Unr expression in different mammalian species

Our *unr* clones had been isolated from a rat testis cDNA library so we were initially interested in seeing whether *unr* expression was limited to that particular species or tissue. The Northern blot presented in figure 5 represents RNA from mouse, rat, and human cells that was probed with the *unr* cDNA. From this blot, it is evident that different species and cell types express multiple, and similarly sized, *unr* transcripts. These transcripts are approximately 4.2, 3.8, and 3.2 kb in size, but the two largest transcripts do not always resolve well.

Analysis of the 3' end of unr transcripts

We were interested in understanding the mechanism responsible for producing the three transcripts from the *unr* gene. A number of genes have been shown to produce multiple transcripts through the differential use of multiple polyadenylation sites located in the 3' untranslated region of the gene (25 for review). We therefore searched for such sites in the 3' untranslated region of *unr* and located four sites that matched the consensus sequence



Figure 6. 3' end analysis of *unr* transcripts. The probes used for the analysis are shown in (B). In this figure, the *unr* cDNA that was shown in figure 3 is now schematically represented. 5' and 3' untranslated regions are shown as thin lines and the *unr* protein coding region is shown as an open box. The three (A's) depicted in the 3' untranslated region represent the polyadenylation consensus sequences located in this region of the cDNA. From left to right, their precise locations are nt 2674, nt 3348, and nt 3718 (also see figure 3). *Probe X* is a 324 nt Sau3A fragment extending from nt 3431 to nt 3755. *Probe Y* is a 1017 nt NarI fragment from nt 2738 to nt 3755. *Probe Z* is a 1492 nt EcoRI fragment from nt 2263 to nt 3755. In (A), 15 μ g of total RNA from an adult RF/J testis were electrophoresed on a 1.2% agarose-formaldehyde gel, transferred to nitrocellulose, and sequentially hybridized to probe X, Y, and Z. After each hybridization, the blot was washed to .2× SSC at 55°C and exposed for 1–2 days.



Figure 7. Unr and N-ras expression in mouse tissues. 15μ g of total RNA from each tissue sample were electrophoresed on a 1.2% agarose-formaldehyde gel and transferred to nitrocellulose. In (A), the blot was hybridized to a unr probe which consisted of the entire unr cDNA depicted in figure 3 and exposed for seven hours. In (B), the same blot was stripped and rehybridized to an N-ras probe (9) and exposed for four days. Blots were washed to .2× SSC at 55°C. Samples from the ovary, spleen, and gut were derived from a female CD-1 hybrid mouse, and all other samples from a male CD-1 hybrid mouse.

for polyadenylation (AATAAA). One signal at nt 3718 and nt 3348, and two overlapping signals beginning with nt 2674 were located (see figure 3 and 6B). The appropriate nucleic acid sequence data was not available to check for the conservation of the signal at nt 2674 between different organisms, but the other two polyadenylation signals that we detected were found to be

conserved in the mouse (22), guinea pig (23), and human (6). If these signals were in fact used to generate transcripts, then the differences in the sizes of the transcripts would be about .4 kb between the largest and the midsized transcript, and about 1 kb between the largest and the smallest transcript. This was in very close agreement with the different sized *unr* transcripts



Figure 8. Unr expression during murine testicular development. Total RNA was isolated from the testis of 13 day old immature (I) and 49 day old mature (M) B10T6R mice. 15 μ g of each RNA were electrophoresed on a 1.2% agarose-formaldehyde gel and transferred to nitrocellulose. In (A), the blot was hybridized to a *unr* probe that consisted of nt 1–804 of the cDNA shown in figure 3 and exposed for 2 hours. In (B) the same blot was stripped, rehybridized to a N-ras probe (9), and exposed for 8 hours. In (C) the same blot was stripped again, rehybridized to a K-ras probe (9), and exposed for 20 hours. Blots were washed to .2× SSC at 55°C. The ethidium bromide stained RNA prior to transfer is presented in (D).

we had in fact observed via Northern blotting (4.2, 3.8, and 3.2 kb).

To gain evidence in support of our hypothesis, we generated three overlapping probes corresponding to the 3' untranslated region of *unr* (see figure 6B). All of the probes went to the very end of the *unr* cDNA, but each began at a different location in the 3' untranslated region of the gene. *Probe X* contained within it the polyadenylation signal at nt 3718 but neither of the other two signals and thus, if our hypothesis is correct, should detect only the largest *unr* transcript. *Probe Y* included the polyadenylation signals at nt 3718 and nt 3348, but not the one at nt 2674 and thus should detect the largest and the midsized *unr* transcripts. *Probe Z* included all three polyadenylation signals and thus should detect all three *unr* transcripts. The Northern blot in figure 6A represents mouse testis RNA that was sequentially hybridized to probes X, then Y, and then Z. The results obtained are consistent with our predictions.

Additional support for this hypothesis has been obtained from the sequencing of *unr* cDNA clones which has allowed us to directly identify polyadenylated clones utilizing the signals at nt 3348 and nt 3718.

Unr and N-ras expression in various tissues

We were next interested in examining the tissue distribution of unr more closely and comparing it with the tissue distribution of N-ras. For this experiment, a Northern blot containing RNA from a variety of mouse tissues was probed with unr, and then stripped and rehybridized with an N-ras specific probe (see figure 7). With regard to qualitative unr expression, it is clear that all 3 transcripts (4.2, 3.8, 3.2 kb) can be detected in all of the tissues examined, with a small amount of a 3.0 kb transcript appearing only in the testis. Differences in the level of unr expression among the tissues examined are evident. Within a particular tissue, an equal ratio of the two largest transcripts is always seen and the amount of the smallest transcript is usually less than that of either of the larger transcripts (the testis represents an exception).

When the blot was hybridized with the N-ras probe, expression

A. MOUSE

B. HUMAN



Figure 9. Unr and N-ras gene copy number and linkage in murine and human genomes. 10 μ g of DNA from a B10 mouse (A) and a single human subject (B) were digested with HindIII, electrophoresed on a 1% agarose gel, and transferred to nitrocellulose. The blot was hybridized with a unr 3' probe and then stripped and rehybridized with an N-ras 5' probe (the probes are described in detail in the text). After the unr hybridization, the blot was washed to .2× SSC at 55°C and exposed for 1–2 days. After the N-ras hybridization, the blot was washed to .5× SSC at 37°C and exposed for 1–2 weeks.

was detected in all tissues (see figure 7B). Consistent with our previous report (9), transcripts of 5.0, 2.4, and 1.3 kb were observed. Thus, both *unr* and N-*ras* are expressed in all of the tissues that we investigated. The overall expression of both genes was found to be the highest in the testis, and the smallest transcript of each gene was more pronounced in this tissue than in any other.

Unr expression during testicular development

Previous reports have shown N-ras to be developmentally regulated in the testis, with an increase in expression being observed upon testicular maturation (9, 26). We therefore were interested in seeing whether the same pattern of expression existed for the unr gene. To investigate this possibility, RNA was isolated from sexually immature (day 13) and mature (day 49) mouse testis. A Northern blot was prepared with these RNA samples and hybridized first with a unr probe and then stripped and rehybridized with an N-ras probe (see figure 8). The results of this experiment indicate that both genes are indeed developmentally regulated in the testis, with an increase in expression occurring upon maturation. If a general upregulation of all genes occurs upon testicular maturation, then the coordinate upregulation of unr and N-ras would be of little significance. To dismiss this possibility, we stripped the blot, rehybridized it with a K-ras probe, and found the expression of this gene to be lower in mature than in immature mouse testis (see figure 8C). This represented a pattern of expression opposite to that of unr and N-ras. Thus not all genes, and in particular not all ras genes, exhibit upregulation during testicular development, as do both N-ras and its upstream neighbor unr. A photograph of the ethidium bromide stained gel is included to demonstrate that RNA of comparable quality and quantity was used for each sample (see figure 8D).

Analysis of unr gene copy number and linkage to N-ras

We wanted to determine the number of unr genes present in human and murine genomes, and to find out whether each unr gene detected was linked to N-ras. To this end, a Southern blot containing mouse and human HindIII digested DNA was first hybridized with a unr probe, and then stripped and rehybridized with an N-ras probe. The probes used corresponded to the 3' end of the unr gene and to the 5' end of the N-ras gene. Specifically, the unr probe contained the last 324 bp of the rat unr cDNA (see probe X, figure 6B). Sequence comparisons revealed that this probe was highly conserved (>90%) among the species being studied. The N-ras probe was a 151 bp Hinf-I/Ava-I fragment derived from a mouse genomic clone (see figure **4B**) and it contained most of exon (-1) of N-ras. This stretch of DNA is 80-85% conserved between humans and mice but is not well conserved between the different ras genes, and thus should be N-ras specific. Using published sequence information (6, 22), a restriction enzyme (HindIII) that did not cut within or between the region of the genome covered by the probes was selected. Thus, each band detected on a Southern blot by either probe should represent a distinct locus.

In the mouse, a single band of 8.5 kb was detected by the unr probe and this same band hybridized to the N-ras probe as well (figure 9A). This finding has been verified in another strain of mice (RF/J) and with a different restriction enzyme (EcoRI) (data not shown). Thus, the mouse most likely contains a single copy of unr per haploid genome, and it is linked to N-ras. The 5.5 kb band recognized by N-ras but not by the unr probe represents a previously characterized N-ras cDNA-like pseudo gene that lacks unr sequences in its 5' region (22). In the human, the unr probe recognized 2 bands, only 1 of which hybridized to the Nras probe (figure 9B). The size of the band that was recognized by both probes (3.9 kb) is consistent with expectations based on the restriction map of the human N-ras gene (27). These results were verified with different human DNA samples and with a different restriction enzyme (EcoRI) (data not shown). Thus, it is likely that there are two copies of unr per haploid human genome, only one of which is linked to the solitary N-ras gene. It remains to be determined whether both of these genes are transcriptionally active.

DISCUSSION

We have identified and characterized a gene, *unr*, that is located in the immediate upstream region of N-*ras*. The close association between these two genes is conserved in all of the species from which N-*ras* has been isolated. A number of other genes have previously been mapped to the region of the chromosome in the immediate vicinity of N-*ras*, and these genes, along with N-*ras*, form a linkage group that has been conserved between man (chromosome 1) and mouse (chromosome 3) (28). The *unr* gene can be considered a new member of this linkage group that is structurally unrelated to the other members. The area of the genome in which this linkage group is located is of special interest since abnormalities in this region have been reported in certain human cancers (29).

The sequence of the *unr* cDNA indicates that the gene is transcribed from the same strand of DNA as N-*ras*, and that the 3' end of *unr* lies very close to the 5' end of N-*ras*. RNase protection data verified these assertions and indicated that the genes were a mere 130 bp apart. This is one of the most closely

apposed pair of mammalian genes reported to date. The RNase protection data also revealed microheterogeneity with respect to unr 3' and N-ras 5' end formation. The unr 3' end was formed at different locations covering a 21 bp stretch of DNA. A certain degree of microheterogeneity at the unr 3' end was also detected directly via cDNA sequencing (see legend to figure 3). This finding is somewhat unusual since most mRNA cleavage and polyadenylation occurs at a single site downstream of any given polyadenylation signal (30 for review). The significance of this observation is, however, unknown. The 5' end of the N-ras transcripts was formed at different locations covering a 20 bp stretch of DNA. This is not an unusual finding since N-ras possesses a GC-rich housekeeping type of promoter (6, 12, R.P. and A.P. manuscript in preparation) and this class of promoter often produces transcripts that exhibit 5' end microheterogeneity (6, 7, 8, 12, 31, 32). Unr apparently also produces transcripts that exhibit 5' end microheterogeneity since a broad band spanning 8 nt was detected when primer extension analysis was performed on the 5' end of unr transcripts.

Sequencing of cDNA's has shown that a protein of 798 amino acids, that is not structurally related to any known protein, can potentially be produced by *unr* transcripts. Although the function of this protein remains unknown, it will be of interest to search for a functional relationship between the N-*ras* and the *unr* proteins in future investigations. Other linked genes have been found to be functionally related (13) and thus the physical proximity of *unr* and N-*ras* may be suggestive of a functional relationship. In addition, our finding that the expression of *unr* and N-*ras* is coordinately regulated, at least in the developing testis (see below discussion), may also be suggestive of a functional relationship between the products of these two genes, since other coordinately regulated genes have been found to share a functional relationship (33).

Unr produces three transcripts of 4.2, 3.8, and 3.2 kb in rat, mouse, and human cells. Based on an analysis of the 5' and 3' ends of unr transcripts, we believe that the cDNA that we have isolated is a full length representative of the largest unr transcript. The finding that our cDNA of 3755 nt seems to be shorter that the longest unr transcript is most likely due to differences in the length of the poly (A) tail that is present on the unr transcript as compared to that found on our cDNA (which is only 8 A's in length) and/or minor inaccuracies in our determination of the unr transcript size.

All three unr transcripts were detected in all mouse tissues examined (an additional 3.0 kb species was detected only in the testis) and thus unr, like N-ras (9), is apparently ubiquitously expressed. We have determined that the unr transcripts differ in the length of their 3' untranslated region and are apparently created via the differential use of multiple polyadenylation signals that are present in the gene. The ratio of the two largest unr transcripts is approximately equal in all tissues, while the smallest transcript is usually at a lower level than the larger species. A notable exception is in the testis where the smallest unr transcript reaches a level of expression equal to or greater than that of either of the two larger transcripts. Curiously, N-ras exhibits a similar testis associated expression of its smallest (1.3 kb) transcript. Tissue associated changes in the ratios of different transcripts have been reported by others (34). One possible explanation for our finding is that the polyadenylation signals which give rise to the smallest unr and N-ras transcripts share some feature that enables them to be more efficiently used in the testis than in other tissues. It is also possible that the smallest unr and N-ras transcripts are somehow preferentially stabilized in the testis.

As has been previously noted, a small amount of a 3.0 kb *unr* transcript is detected in the testis, but not in any other tissue. Other genes that have been reported to produce testis specific transcripts include *c-abl* (35) and *pim-1* (26). The mechanism responsible for the generation of testis specific transcripts from *unr* or these other genes is not known.

It is also notable that the overall expression of both unr and N-ras is higher in the testis than in any other tissue, and that the expression of both genes is subject to developmental upregulation in this tissue while other ubiquitously expressed genes, for example H-ras (26) and K-ras (26, this paper) are not. These findings suggest that unr and N-ras may be coordinately regulated, at least in this tissue. Genes that are not physically linked may accomplish the task of coordinate regulation by possessing similar transcriptional regulatory elements in their promoters (36). However, the close proximity of unr and N-ras could conceivably allow a single regulatory element, such as a testis responsive enhancer, to be shared between the two genes. The utilization of a single regulatory element by a set of linked genes has been described for the β -globin locus, where a single cis-acting control region is believed to play a role in the tissue specific expression of all five of the genes that reside in the cluster (37). It is possible that cis-acting elements that play a role in the regulation of N-ras reside within unr. Situations have previously been described in which sequences involved in the regulation of a downstream gene have been found in the 3' region of an upstream gene (38), and we have recently obtained evidence that such an arrangement exists within the unr/N-ras gene cluster (R.P. and A.P. manuscript in preparation).

Southern blotting has shown that *unr* is probably a single copy gene in mice, and that this copy is linked to the murine N-ras gene. This indicates that in mice the unr locus in close association with N-ras is transcriptionally active. Transcription from the unr locus lying immediately upstream of N-ras presents an interesting transcriptional scenario since available evidence in eukaryotes indicates that transcription will generally proceed through functional polyadenylation signals and terminate anywhere from hundreds to thousands of nucleotides downstream of the last polyadenylation signal (39 for review). If these finding hold true for unr, then transcription will continue past the polyadenylation signals present in the 3' untranslated region of the gene, and terminate either somewhere in the short segment of DNA (130 bp) that exists between unr and N-ras, or somewhere within the N-ras gene. If transcription from unr does proceed into N-ras, it could potentially have some effect on N-ras transcription. Although one may expect that the most likely effect of unr transcription would be to decrease N-ras transcription via interference (40), our expression studies do not support this expectation since we have found that both unr and N-ras may be expressed at relatively high levels in the same tissue, for example the testis, and in fact may be coordinately regulated in this tissue.

Since *unr* is not found in the 5' flanking region of H-*ras* or K-*ras*, any role that *unr* plays in the regulation of N-*ras* is likely to be specific for this member of the *ras* family. Thus, an examination of the transcriptional relationship that exists between *unr* and N-*ras* may advance our understanding of how, and perhaps why, the *ras* genes are differentially regulated. In addition, it is important to understand the role that *unr* may play in the regulation of N-*ras* since a perturbation of the transcriptional relationship that may exist between these two genes could potentially lead to deleterious alterations in N-*ras* gene expression.

ACKNOWLEDGEMENTS

Special thanks to Timothy Thomson for isolating the initial 1.5 kb *unr* cDNA clone (depicted in figure 1) and to Ramon Mangues for supplying RNA samples from various mouse tissues. This work was supported by NIH grants CA36327 and CA16239. A.P. is a Leukemia Society Scholar.

REFERENCES

- 1. Barbacid, M. (1987) Ann. Rev. Biochem. 56, 779-827.
- 2. Santos, E. and Nebreda, A.R. (1989) FASEB J. 3, 2151-2163.
- Willingham, M.C., Pastan, I., Shih, T.Y. and Scolnick, E.M. (1980) Cell 19, 1005-1014.
- Sweet, R.W., Yokoyama, S., Kamata, T., Feramisco, J.R., Rosenberg, M. and Gross, M. (1984) Nature 311, 273-275.
- 5. Gilman, A.G. (1984) Cell 36, 577-579.
- 6. Hall, A. and Brown, R. (1985) Nucleic Acids Res. 13, 5255-5268.
- 7. Ishii, S., Merlino, G.T. and Pastan, I. (1985) Science 230, 1378-1381.
- 8. Yamamoto, F. and Perucho, M. (1988) Oncogene Res. 3, 125-138.
- 9. Leon, J., Guerrero, I. and Pellicer, A. (1987) *Mol. Cell. Biol.* 7, 1535-1540.
- Chang, E.H., Furth, M.E., Scolnick, E.M. and Lowy, D.R. (1982) Nature 297, 479-483.
- 11. McKay, I.A., Marshall, C.J., Cales, C. and Hall, A. (1986) EMBO J. 5, 2617-2621.
- 12. Doniger, J. and DiPaolo, J.A. (1988) Nucleic Acids Res. 16, 969-980.
- Bray, P.F., Barsh, G., Rosa, J-P., Luo, X.Y. and Magenis, E. (1988) Proc. Natl. Acad. Sci. USA 85, 8683-8687.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- 15. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
- Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299.
- 17. Chomczynski, P. and Sacchi, N. (1987) Analytical Biochem. 162, 156-159.
- Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) Nucleic Acids Res. 12, 7035-7056.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- 20. Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.
- 21. von Heijne, G. (1986) Nucleic Acids Res. 14, 4683-4690.
- Chang,H.-Y., Guerrero,I., Lake,R., Pellicer,A. and D'Eustachio (1987) Oncogene Research 1, 129-136.
- 23. Doniger, J. (1987) Oncogene 1, 331-334.
- 24. Kozak, M. (1989) J. Cell. Biol. 108, 229-241.
- Leff,S.E., Rosenfeld,M.G. and Evans,R.E. (1986) Ann. Rev. Biochem. 55, 1091-1117.
- Sorrentino, V., McKinney, M.D., Giorgi, M., Geremia, R. and Fleissner, E. (1988) Proc. Natl. Acad. Sci. USA 85, 2192-2195.
- Brown, R., Marshall, C.J., Pennie, S.G. and Hall, A. (1984) EMBO J. 3, 1321-1326.
- Povey, S., Morton, N.E. and Sherman, S.L. (1985) Cytogenet. Cell Genet. 40, 67-106.
- 29. Povey, S. and Parrington, J.M. (1986) J. Med. Genet. 23, 107-115.
- 30. Manley, J.L. (1988) Biochemica et Biophysica Acta 950, 1-12.
- Ishii, S., Xu, Y.-H., Stratton, R.H., Roe, B.A., Merlino, G.T. and Pastan, I. (1985) Proc. Natl. Acad. Sci. USA 82, 4920-4924.
- Reynolds, G.A., Goldstein, J.L. and Brown, M.S. (1985) J. Biol. Chem. 260, 10369-10377.
- Tollefsen, S.E., Sadow, J.L. and Rotwein, P. (1989) Proc. Natl. Acad. Sci. USA 86, 1543-1547.
- Marie, J., Simon, M.-P., Lone, Y.-C., Cognet, M. and Kahn, A. (1986) Eur. J. Biochem. 158, 33-41.
- 35. Ponzetto, C. and Wolgemuth, D.J. (1985) Mol. Cell. Biol. 5, 1791-1794.
- 36. Chang, S.C., Erwin, A.E. and Lee, A.S. (1989) Mol. Cell. Biol. 9,
- 2153-2162. 37. Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, G. (1987) Cell
- **51**, 975–985.
- 38. Wu,L.-C., Morley,B.J. and Campbell,R.D. (1987) Cell 48, 331-342.
- 39. Proudfoot, N.J. (1989) TIBS 14, 105-110.
- 40. Proudfoot, N.J. (1986) Nature 322, 562-565.