# The human sperm protein PH-20 has hyaluronidase activity

Michael Gmachl, Sandrine Sagan\*\*, Sigrid Ketter, Günther Kreil\*

Institute of Molecular Biology, Austrian Academy of Sciences, Billrothstrasse 11, A-5020 Salzburg, Austria

Received 22 November 1993

The PH-20 protein present on the membrane of guinea pig sperm was characterized using a monoclonal antibody [(1991) J. Cell Biol. 111, 2939–2949]. We have isolated the cDNA encoding the human PH-20 protein from a testis library. This cDNA was expressed in RK 13 cells using a vaccinia virus expression system. Cells expressing the human PH-20 protein possess hyaluronidase activity. Treatment with PI-PLC releases the hyaluronidase into the the medium with a concomitant large increase in enzymatic activity. These results demonstrate that the human PH-20 protein has hyaluronidase activity.

PH-20 protein; Hyaluronidase; Human sperm

# 1. INTRODUCTION

Hyaluronidases have been isolated from such diverse sources as mammalian testis, the venom of snakes, bees and scorpions and the salivary gland of leeches [1]. These enzymes hydrolyze hyaluronan (hyaluronic acid), a polysaccharide of high molecular mass that is found in the extracellular matrix. We have recently determined the amino acid sequence of bee venom hyaluronidase using recombinant DNA techniques [2]. This represents the first structure of a hyaluronidase of animal origin. Surprisingly, it was found that the sequence of this insect enzyme is homologous to PH-20, a glycoprotein present on the head of guinea pig sperm [3].

PH-20 was identified by monoclonal antibodies and has been shown to play a role in the binding of sperm to the egg zona pellucida. This protein is located at the posterior head surface of guinea pig sperm and it migrates to the acrosomal membrane after the acrosome reaction [3,4]. The PH-20 protein is originally synthesized as a polypeptide with an apparent molecular mass of 64 kDa and it is bound to the membrane via a GPIanchor. During sperm maturation the bulk of PH-20 is cleaved into two domains, an amino-terminal fragment of 41-48 kDa and a carboxy-terminal part of 27 kDa, the two being linked by disulfide bridges [3]. Antibodies against PH-20 have been shown to block fertilization both in vitro and in vivo [5].

Bee venom hyaluronidase and the guinea pig PH-20 protein exhibit 36% identity in a region encompassing

approximately 300 amino acids. This led us to hypothesize that PH-20 might be a hyaluronan binding protein or a hyaluronidase.

Here we describe the isolation of a cDNA encoding the human PH-20 protein. The cDNA was expressed in rabbit kidney cells using a vaccinia virus expression system. It could be shown that cells expressing the PH-20 protein possess hyaluronidase activity.

## 2. EXPERIMENTAL

#### 2.1. Materials

Restriction endonucleases and DNA modifying enzymes were from Boehringer Mannheim, Bethesda Research Lab., New England Biolabds, or Stratagene. All radiochemicals were from NEN Research Products (Boston, USA). Phosphoinositol specific phsopholipase C was obtained from Sigma (St. Louis).

#### 2.2. RNA isolation, cDNA cloning, and library screening

RNA was isolated from guinea pig testis [6] and transcribed into double-stranded cDNA by Moloney murine leukemia virus reverse transcriptase (BRL), and E. coli DNA polymerase using standard procedures. A cDNA fragment encoding guinea pig PH-20 protein was then prepared with two synthetic oligonucleotides taken from the published sequence [3] and the polymerase chain reaction. The forward primer was 5'-GATAAAAGGGCACCACCACTC-3' (positions 460- 480) and the reverse primer was 5'-CAGTTGCCAT-GAGCCTATAGT-3' (position 1958-1938, opposite strand). After 30 cycles of PCR, a cDNA-fragment containing about 1500 basepairs was obtained. Partial sequencing confirmed that this cDNA encoded the mature guinea pig PH-20 protein [3]. This cDNA was labeled by the random priming method and then used to screen a cDNA library in  $\lambda$ gt11 prepared from human testis (Clontech). Hybridization under low-stringency conditions yielded three positive clones. The insert present in clone CM19 was subcloned into the Bluescript plasmid (Stratagene) and sequenced by the enzymatic method [7] with the Sequenase 2.0 sequencing kit (United States Biochemicals).

#### 2.3. Preparation of recombinant vaccinia virus

The cDNA encoding the human PH-20 protein was ligated into the SmaI-site of the plasmid pgpt-ATA-18 [8]. The resulting plasmid was

<sup>\*</sup>Corresponding author. Fax: (43) (662) 624961-29.

<sup>\*\*</sup> Present address: Université Pierre et Marie Curie, Laboratoire de Chimie Organique Biologique, F-75005 Paris, France.

Abbreviations: GPI, glycosyl-phosphatidylinositol; PI-PLC, phosphoinositide-specific phospholipase C.

termed pATA-HPH. Human 143B cells (ATCC CRL8303), an osteosarcoma cell line which lacks thymidine kinase (TK), were infected with 0.1 plaque forming units of wild-type vaccinia virus. After 2 h, a calcium phosphate precipitate of the plasmid pATA-HPH (0.5  $\mu$ g DNA) was added to the cells and these were then incubated for 2 h at 37°C. During this period, recombination between the viral TK gene and the homologous viral TK-sequences of the plasmid takes place. After addition of fresh medium, cells were incubated for 2 days under non-selective conditions. Cells were then lysed by freezing and thawing. Cultures of about 80% confluent human 143B cells were infected with the lysate and incubated under selective conditions, i.e. in the presence of 5-bromodeoxyuridine (100  $\mu$ g/ml), mycophenolic acid (25  $\mu$ g/ml), hypoxanthine (15  $\mu$ g/ml), and xanthine (250  $\mu$ g/ml), for 2 days to propagate the recombinant viruses. By plaque assay purification, single recombinant viruses were obtained [8,9]. RK13 cells were infected with recombinant virus and incubated for 24 h at 37°C. In control experiments, cells were infectes with wild-type vaccinia virus.

#### 2.4. Hyaluronidase assays

Tests for this enzyme were performed as described by Dorfman [10]. Medium was removed from infected cells and these were then washed once with phosphate-buffered saline. To the cells (ca.  $4 \times 10$  cells), 2 ml of a solution of hyaluronan (Sigma, 125  $\mu$ g/ml in 300 mM Naphosphate/35 mM NaCl, pH 5.5, containing 0.005% bovine serum albumin) were added and incubated at 37°C. At different times, aliquots of 0.5 ml were removed, boiled for 5 min, and centrifuged. To the supernatant, 5 volumes of sodium acetate buffer (25 mM, pH 3.75) containing 0.1% bovine serum albumin was added. After 15 min at room temperature, the turbidity was recorded at 600 nm [10].

# 2.5. Digestion with PI-PLC

About  $4 \times 10$  cells were incubated with 0.5 ml phosphate-buffered saline (pH 7.4) containing 1.2 units of PI-PLC for 20 min at 37°C. Subsequently, hyaluronan was added and hyaluronidase activity was determined.

#### 3. RESULTS

#### 3.1. Sequence of the cDNA encoding human PH-20

To test whether the PH-20 protein also possesses hyaluronidase activity, we have first isolated the cDNA encoding the mature guinea pig PH-20 using two oligonucleotides derived from the published sequence [3] and the polymerase chain reaction. Attempts to express this cDNA in E. coli under conditions which yielded the recombinant bee venom enzyme [2] were unsuccessful. We therefore used the guinea pig cDNA as a probe to screen a cDNA library from human testis. The insert of one positive clone, CM19, was sequenced using the enzymatic method [7] and found to contain 1974 nucleotides excluding the poly(A)-part (see Fig. 1). This cDNA starts with 373 nucleotides of untranslated 5'-end. Interestingly, this region contains six ATG initiation codons, each of which is followed by an in-frame stop codon. A homologous sequence was found in the cDNA encoding guinea pig PH-20 [3] which contains five initiation codons. However, in both instances these upstream ATG codons are not in a favorable context, as deduced by Kozak [11], for initiating translation. After the seventh ATG codon of the cDNA isolated from human testis, an open reading frame is present comprising 509 codons, of which the first 35 probably code for a signal peptide (see Fig. 1). As deduced from the cDNA se-

GCT/ ACT/ CAAC TCAT	AT	- 1 - 1						Gar	100		on L		910	• · •	~				_
ACT/ CAAC TCAT	1 mm	auc.	TAG	CCG	GTG	CTC	TAG	GAA	GAC	ATT	GAG.	ACC	AGC	CAA	CTT	CTT	GCC	TIG	ATA
CAAC TCAT	1.1.1	SAA	GAG.	ACA	TTG	GGT	GGC	TGG	ATT	ГTG,	AAA	GCA	GAC	TTC	TGG	TTA	TAG	GTG	ATG
TCAT	TTT	GAA.	AAA	CAA'	TCC	TGA	AAC.	ATG	AAA	CAA	GAA	TAA	TAA	TAT	TTA	AAT	GTA	ACT	TAA
	"TAT	FAC	CTC'	TTT	ATC	CAT	CAA	AGT	GGA	TT	CAT	TCC.	ATT	CCC	TTT	CAT	CTG	TGC	TCA
TAC	ידידי	CA	TCA	GAT	ATT	GGG	TAA	ACC		and	TGT	ACG	AAG		TAA	ATG	I-I-F	TTA	TAG
Tran	-	780	-	101	a 7.	CCa.	-	0778	2 2 2 2	-	830	CAC	3.77.79	THEFT		303	***		
ICHI		uen.	***	n-n	4-5	-	310	~.n			~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	-n	nic.	1 <u>.</u>	110	7000	a.g.	***	311
					met			مد	R.		X,	н	_ <u>+</u>	r	r	<u> </u>	<u>s</u>		V
AAA	CA/	AGT	GGA	GTA	TCC	CAG.	ATA	GTT	TTC.	ICC.	TTC	CTT	CTG.	ATT	CÇA	TGT	TGC	TTG.	ACT
ĸ	s	s	G	v	S	Q	I	v	F	т	F	L	L	I	P	С	С	L	T
CTG	AT?	TTC.	AGA	GCA	CCT	CCT	GTT.	ATT	CCA	ATA.	GTG	GCT	TTC	CTC	TGG	GCC	TGG	AAT	GCC
L	N	F	R	A	P	Р	v	I	P	N	v	À	F	L	W	A	W	N	A
CCAZ	GTY	AA'	ידידי	TGT	CTT	GGA.	-	inter-	GAT	AC	~~ A	CTA	GAT:	a Tra	adr	CTC	1919-0	TOT	TTC
D	e	P.		~	τ.	00.0	¥		<b>n</b>	P	7		- m	Ŷ	a				
30130	0.1	inn			8.00.8	3 3 00	-00	*~~	ñ	· × × /	20.2	~			norde Th		- -	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	
	- -	2000	500	- Gru	-1A	NACI	366		acor		JOAN	91.14	<u>n_n</u>	HIA.	111	TAL	311	GAL	nGH
1	G	2	2	к_	1	R	A	T	G	Q	G	¥	r	1	E.	¥	v	D	ĸ
CTTC	GC	CAC	PAT	CCT.	FAC.	ATA	JAT.	ICA	ATCI	ICA(	GA	GTA.	ACTO	<b>TG</b>	AAT	GGA	GGA.	ATC	CCC
L	G	¥	Y	P	Y	I	D	S	I	т	G	v	т	v	N	G	G	I	P
CAGA	AGA	ATT?	FCC!	TTA	CAA	GAC	CAT	CTG	GAC	AA	SCT	AAG,	AAA	GAC	ATT	ACA	TTT	TAT.	ATG
0	ĸ	I	S	τ.	0	D	H	T.	n	ĸ	A	x	ĸ	D	т	T	F	Y	м
crac	The C	ani	à a mr	TTTC:	cā.	2000	101	200	2 mm	10.00	TOC/		22.81	ñaa	a	crrr	- -	n Cru	CC 8
n	57	50.00	ълд. М	1100	- C	10	8	311.	n <u>,</u> 1/	D.	1000	500.00		10303	-			100	30.0
	×			ц 		M.	~	- <u>*</u> -				E	E.			-r	т.		
AGAA	AC'I	GGA	AAA	cen	AAA	GAT	1.1.	L'AC	AAGA	ATA	٩GG'	rcr.	81-14	an A	FIG	GTT	AG	CAR	ÇAA
R	N	w	ĸ	P	ĸ	D	v	X	ĸ	Ņ	R	S	1	Е	Ľ	v	Q.	Q	Q
AATG	TAC	CAAC	TTT	AGT(	CTC.	ACAG	IAG(	SCC.	ACTO	IAC!	LAA(	GCA	AAA(	CAA	GAA	TTT	SAA.	AAG	GCA
N	v	Q	L	s	L	T	E	А	т	E	ĸ	A	ĸ	0	E	F	Ε	ĸ	A
GGGA	AG	IA TT	PTC	TG	STA	GAG	CT	ATA	AAA	TGO	GAR	-	ኮጥእስ		CGG	CCA	AAT	CAC!	rrG
G	ĸ	n	F	¥.	v	F	-	т	ĸ	τ.	a	R	τ.	τ.	12	P	N	8	τ.
more		10.00	• • • • • • •		-	~~~~	2 8 177	+ n	PACI	200	- NUTV	- 21-17		NAC.			2000		5 3 17
1000	GII	. AL L L	. MIC			CCG1		1.91	TWC	anc.	.ut(	-nu		-			301	100	
						_		~	**					-	~~	-	~		
W	6	*	1	L	£.	P	Ð	С	¥	N	н	H	Y	ĸ	x	P	G	¥	14
W GGAA	G GT	'GC'	rTCI	L AAT(	f GTA	P GAAJ	D Itai	C AAA	y Aga <i>i</i>	N ATC	H SATY	H GAT(	Y CTCI	K AGC	K TGG	P TTG	G FGG.	Y AATI	GAA
W GGAA G	G GTI S	iGC1 C	rtci F	L AAT( N	f GTA V	P GAAJ E	U I I	C AAA K	y Aga <i>i</i> R	N LATC N	H SATY D	H GATV D	Y CTCI L	K AGC S	K TGG W	P TTG L	G IGG. W	X AATI N	GAA E
W GGAA G AGCA	GTI S CTO	iGCI C	rtci F CTTT	L AAT( N FAC(	F GTA V CCA	P GAAI E TCCI	D ATAI I ATT	C AAA K FAT	Y AGA/ R ITTG/	N ATC N ACJ	H BATY D ACT(	H GAT D CAG	Y CTCI L CAG	K AGC S ICT	K TGG W CCT	P TTG L GTA	G FGG. W SCT(	AATI N SCTJ	GAA E ACA
W GGAA G AGCA S	GTI S CTO T	GCT C C C TC A	F F CTTT	L NAT( N FAC( Y	F GTA V CCA P	P GAAI E TCCI S	D ATA I ATT: I	C AAA K FAT Y	Y AGA/ R ITG/ L	N ATC N ACJ N	H SATY D ACT( T	H GAT D CAG	Y CTCI L CAG	K AGC' S ICT' S	K TGG W CCT P	P TTG L GTA V	G IGG W SCT A	X AATI N GCT A	GAA E ACA T
W GGAA G AGCA S CTCT	GTI S CTO T	igen C C C C C C C C C C C C C C C C C C C	F F CTT L	L AAT( N FAC( Y AAT(	F GTA V CCA P CGA	P GAAJ E TCCJ S	D TAI I TT I	С Ада К ГАТ У За ал	Y AGA/ R ITTG/ L	N ATC N ACJ N TCJ	H BATY D ACTO T	H GAT D CAG Q	Y CTCI L CAG: Q	K AGC S ICT S	K FGG W CCT P	P TTG L GTA V CCTV	G FGG W SCT A SAT	X AATI N SCTI A SCAI	GAA E ACA T
W GGAA G AGCA S CTCT	GTI S CTO T ATO	GCT C C C C C C C C C C C C C C C C C C	I F CTT L CGCI	L N TACO Y NATO	F GTA V CCA P CGA	P GAAJ E TCCJ S GTTC	U TA I TT I GG(	C AAA K FAT Y JAA	Y AGA/ R ITG/ L GCC/	N ATC N ACJ N TC T	H BATY D ACTO T AGAO	H GAT D CAG Q STT	Y L CAG: Q FCCI	K AGC S ICT S AAAJ	K FGG W CCT P ATA	P TTG L GTA V CCT	G FGG W SCT A SAT B	X AATI N GCT/ A GCA/ A	GAA E ACA T AAA
W GGAA G AGCA S CTCT L	G GTI S CTO T Y ATO Y	igci C SCTO A TGO V	ITCI F CTT7 L CGCI R	L N TACO Y AATO N	F GTA V CCA P CGA R	P GAAJ E TCCJ S GTTC V	U I I I I CGG( R	C AAA K TAT Y SAA E	Y R R ITG/ GCC/ A	N ACI ACI N TCI I	H GATY D ACT( T AGA( R	H GAT D CAG Q STT V	Y L CAG: Q FCC/ S	K AGC' S ICT' S AAAJ K	K IGG W CCT P ATA	P TTG L GTA V CCT P	G FGG. W SCT A SAT D	X AATI N GCT/ A GCA/ A	GAA E ACA T AAA K
W GGAA G AGCA S CTCT L AGTC	G GTI S CTC T PATC Y CAC	figer C A SCTO A STGO V CTTO	I F CTT L CGC/ R CCGC	L N TACO Y AATO N STTT	F GTA V CCA P CGA R TTT	P GAAJ TCCJ S GTT( V GCA	U I I I CGG( R I I TAT/	C AAA K TAT Y SAA E ACC	Y AGA/ R PTG/ L GCC/ A CGC/	N ATC N ACJ N TCJ I ATA	H D ACTO T AGAO R STT1	H GATY D CAGO Q STT: V FTT/	Y L CAG: Q FCC/ S ACT(	K AGC S ICT S AAAJ K GAT	K IGG W CCT P ATA I CAA	P TTG L GTA GTA CCTX P GTT	G IGG W GCT A BAT D D TG	X AATI N GCT/ A GCA/ A AAA:	GAA E ACA T AAA K TTC
W GGAA G AGCA S CTCT L AGTC S	G GTI S CTO T ATO Y CAO P	rgct C A TGC V TTC L	I F CTT L CGCI R CCGC	L N TACO Y AATO N STTT V	F STA V CCA P CGA R TTT F	P GAAJ TCCJ S GTTC V GCA A	U I I I GGG R I A TA TA Y	C AAA K TAT Y SAA E ACC T	Y AGA/ R PTG/ L GCC/ A CGC/ R	N ACI N TCI I TA( I	H SAT( D ACT( T AGA( R STT)	H GATO D CAGO Q STT V F TT/ F	Y L CAG Q FCC/ S ACT Y	K AGC S ICT S AAA K SAT D	K TGG W CCT P ATA I CAA	P TTG GTA GTA CCTV P GTT V	G FGG W SCT A SAT D TTG L	X AATI N GCT/ A GCA/ A AAA( X	GAA E ACA T AAA K FTC F
W GGAA G AGCA S CTCT L AGTC S CTTI	G GTI S CTO T ATO Y CAO P CTO	rgct C A STGC V TTC L CAAC	I F C C C C C C C C C C C C C C C C C C	L N TACO Y AATO N STTT GAAO	F GTA CCA P CGA R CTT F CTT	P GAAJ E TCCJ S GTTC V GCAS A GTGI	U I I CGG( R TAT/ Y TAT/	C K K TAT Y SAA SAA E ACC T	Y AGA/ R FTG/ L GCC/ A CGC/ R TTTC	N ACI N TCI I GCO	H SAT( D ACT( T AGA( R STT) SAAJ	H GATO D CAGO STT S V F TTI F ACTO	Y L CAG Q FCC/ S ACT T STT	R AGC S ICT S AAA R SAT D SCT	K TGG W CCT P ATA I CAA Q CTG	P TTG GTA GTA CCTX P GTT GTT GGT	G FGG. W SCT A SAT D TTG. L SCT	X AATI N GCT/ A GCA/ A AAA! X FCT(	GAA E ACA T AAA K FTC F GGA
W GGAA G AGCA S CTCT L AGTCT S CTTTI L	G GTT S CTO T ATO Y CAO P CTO S	fgci C C C C C C C C C C C C C C C C C C C	I F CTT L CGC/ R CCGC P SATC D	L NATO Y AATO N STTT V GAAO E	F GTA V CCA P CGA R CGA R TTT F CTT L	P GAAJ E TCCJ S GTTC V GCA GCA A GTG V	U I I I I I I I I I I I I I I I I I I I	C K TAT Y SAA E ACCI T ACA T	Y AGA/ R TTG/ L GCC/ A CGC/ R TTTC F	N ACI N TCI I GCC G	H BAT( D CT( T GA( R STT) SAA/ E	H GAT D CAG Q STT V F TT/ F ACT T	Y L CAG: Q FCC/ S ACTY T STTY V	K AGC' S ICT' S AAAJ K SAT' D SCT' A	K TGG W CCT P ATA I CAA CTG L	P TTG GTA GTA CCTX CCTX GTT GGT GGT G	G FGG. W SCT A SAT D TTG. SCT A	X AAT SCTI A SCAI A SCAI X SCAI X S S	GAA E ACA T AAA K FTC F GGA G
W GGAA G AGCA S CTCT L AGTC S CTTT L ATTG	G GTI S CTO T ATO Y CTO S TAJ	IGCI C C C C C C C C C C C C C C C C C C	I F CTTT L CGC/ R CCG( P GAT( D IGG(	L NATO Y NATO N STTT SAAO E SGAJ	F GTA V CCA P CGA R TT F CTT L	P GAAJ E TCCJ S GTTC Q GTTC A GTG V CTCJ	U I I CGGC R I ATI Y Y Y AGTI	C AAA K TAT Y SAA SAA E ACC T ACA T ACA	Y R R TTGJ L GCCJ A CGCJ R CGCJ R TTTC F ATGC	N ACI N ACI N TCI I G G G G G A	H SAT( D ACT( T AGA( R STT) SAAJ E AGTJ	H GAT Q Q STT V F T A CT X T A T G	Y L CAG Q FCC/ S ACTY S T V AAS	K AGC S ICT S A A S A T S C T A C T	K IGG W CCT P ATA I CAA CAA CTG L CTG	P TTG GTA GTA CCTX CCTX GTT GGTY GGTY GGTY GGTY	G IGG. W SCT A SAT D ITG. SCT A CTC	X AAT SCTI A SCAI A SCAI X SCAI X SCAI X SCTAI	GAA E ACA T AAA K F GGA G GAC
W GGAA G AGCA S CTCT L AGTC S CTTT L ATTG I	G GTI S CTO T Y CTO S TAM V	IGCI CTC A TGC V TTC L CAAC Q I I	I FTCI F CTTT L CGC/ R CGC/ P SAT( D SAT( D V W	L NAT( Y NAT( STT) V SAA( E SGAJ G	F GTA/ V CCA <sup>+</sup> P CGA <sup>+</sup> F CTT <sup>+</sup> L CC <sup>+</sup> T	P GAAJ E TCCJ S GTTC V GCAT A GTGT V CTCJ L	D I I I CGG( R I ATI Y Y Y AGTI S	C K K TAT SAA SAA E NCCI TA TA TA I	Y R R TTGJ L GCCJ R CGCJ R CGCJ R TTTC F ATGC M	N ACI N TCI TACI G G G R	H SAT( D ACT( T AGA( R STT) SAAJ E SAAJ SAAJ	H GAT D CAG Q STT V T T A CT A T G M	Y L CAG: Q FCC/ S ACT( Y STT( V AAAT K	K AGC S ICT S A K S A D S C T S C T S C T S	X IGG V CCT P ATA I CAA CAA CTG L CC C	P TTG GTA GTA CCTX P GTT GTT GGT GGT GGT L	G IGG. SCT A JAT JAT L SCT A CTC L	X AAT SCTI A SCAI A AAA X FCT S CTA L	GAA E ACA T AAA K F GGA G GAC D
W GGAA G AGCA S CTCI L AGTC S CTTI L ATTG I AATTG	G GTI S CTC T ATC Y CAC P CTC S A TAA	IGCI C C C C C C C C C C C C C C C C C C	I F C C C C C C C C C C C C C C C C C C	L NAT( Y NAT( N STT SAA( E G G	F GTA V CA P CA R TTV F CTTV L CC T	P GAAJ E TCCJ S GTTC V GCA GTG V CTCJ L CTCJ	U I I I I I I I I I I I I I I I I I I I	C K K TAT Y SAA SAA CC TA TA TA TA	Y AGA/ R TTG/ L GCC/ A CGC/ R TTC F ATGC	N ACI N TCI TACI GG GAJ R	H JATY D VCTY T VGA( R STT SAA) SAA) SAA) SAA) SAA)	H GAT D CAG Q STT V T T A C T A C T A C T A C T A C T A C T A C A C	Y L CAG: Q FCC/ S TCC/ S TTX V AAA: K	K NGC' S ICT' S A S A S A S C T' S C T' S C T' S C T' S C T' S C T' S C T' S C T' S S C T' S S S C T' S S S S S S S S S S S S S S S S S S	X IGG V PATA I CAA CAA CTG C C C C C	P TTG GTA GTA C C C C C C C C C C C C C C C C C C C	G IGG. SCT A JAT JAT L SCT A CTC L	X AATI N GCT/ A GCA/ A AAA X FCT( S CTA( L AAA)	GAA E ACA T AAA K F GGA GGA GGA GGA C D
W GGAA G AGCA S CTCI L AGTC S CTTI L ATTG I AATTG	G GTI S CTC T ATC Y CAC P CTC S TAP Y ACP	IGCT C A TGC V TTC L AAC Q I I I I G C TAT	I FTCI F CTTT L CGC/ R CCGC P JATC D IGGC W JAG/ F	L NATO Y AATO STT V SAAO E SGAJ G ACT/	F GTA V CA' P CGA R TT C CA' F C TT L C C T TA TA	P GAAJ E TCCI S GTTC V GCA A GTGI V CTCJ L CTGI	U I I I CGG( R I A TAT/ Y I A TAT/ Y S A TAT/ N S	C K FAT SAA SAA CCI TAI SCA TAI SCT	Y AGA/ R TTG/ L GCC/ A CGC/ R TTC F A TGC M TAC/ Y	N IAT( N IAC) IAC) ITC) ITC) ITC) ITC) ITC) ITC) ITC) IT	H JATC D CTC T GAC STT SAAJ SAAJ SAAJ SAAJ	H GAT( D CAG( Q STT: V FTT/ F T ACT( T ACT( M AAC( N	Y CTCI L CAG Q FCCI S TCCI S TCCI Y AAT K STCI Y	K AGC S ICT S AAA S AT S C T S C T S C T S C T S C T S C T S C T S C T S C T S S C T S S C T S S C T S S C T S S S C T S S S C T S S S S	X IGG W CCT P ATA IGC CTG CC T CCT CCT CCT CCT CCT	P TTG' CCTA CCTA CCTA GTT GTT GGT( GGT( GCA)	G IGG. W JAT JAT L JAT L JCT L JCC L JCC	X AATI N GCT/ A GCA/ A AAA X FCT( S CTA( L AAA/ L	GAA E ACA T AAA K T F GGA GGAC D ATG
W GGAA G AGCA S CTCT L AGTC S CTTT L AGTC I AATTG I AATT N	G GTT S CTC T V CTC S CTC S TAA V V CTC S CTC S CTC S CTC S CTC S C T C T C	IGCT C A TGC V TGC V TTC L A A C A A C A A C A A C A A C A A C A C A C A C A C A C A C A C A C A C A C A C A C C C A C C C A C C C A C	I FTCI F CTTT L CGC/ R CCGC P SATO D SATO D SAG/ E SAG/ E	L NATO Y AATO Y STT7 V SAAO E SGAJ G ACT7 T	F GTA P CCA P CGA R TTA CCA T CTTA L CCA T A CCA T A CCA T A CCA T A CCA T A CCA T A C T T A C T T A C A C	P GAAI E TCCI S GTTC V GTTC A GTGI L CTGI L CTGI	D ITAJ I I I I I I I I I I I I I I I I I I I	C AAA K FAT Y SAA CCI T ACCI T ACCI T ACCI T ACCI T ACCI T ACCI T ACCI T ACCI T ACCI T ACCI T ACCI T ACA	Y AGAN FTGJ L GCCJ A CGCJ R CGCJ R F TTC F A TGCJ Y Y CAC	N AATO N ACJ N TCJ TAO G G G G G A TAJ T T T T T T T T T T T T T T T T T	H JATY D ACTO T AGAO R STT S AAJ E S AAJ S AAJ S A T C I I O T I	H GATO D CAGO STT: V FTT/ F ACTO T ACTO M AACO N	Y CTCI L CAG Q FCCI S TCCI S TTV V AAAT K STCI	K MGC' S NAAJ K SAT' D SCT' A SCT' S CT' S CT' S CT' S CT' S CT' S CT' S CT' S CT' S CT' S CT' S CT' S S CT' S S S S S S S S S S S S S S S S S S S	K FGG W CCT P ATA CAA CTG C C C C C C C C C	P TTG GTA GTA CCTX P GTT GGT GGT GGT GGT GGT GGT C GCA GCA	G IGG. SCT A SAT B SCT A SCT L C C C C C C A	Y AATI N GCT/ A GCA/ A AAA Y TCTO S CTAO L AAA/ K NGA/	GAA E ACA T AAA K T C F GGA GGAC D ATG
W GGAA G AGCA S CTCT L AGTC S CTTTI L ATTG I AATTT N TGTA	G GTT S CTC T NATO Y CAC P CTC S TAA Y CAC P CTC S TAA Y S CC C C C C C C C C C C C C C C C C	I GCTC GCTC A STGC V CTTC L CAAC Q I I TGC M CAAC	I FTCI F CTTT L CGC/ R CCGC P SATO D SATO D SAG/ E STGO	L NATO Y AATO STT V SAAO E GAA G ACT/ T T	F GTAJ V CCA' P CGA R TTV L CTTV L CTTV L CTTV L C CTTV L C C T T L C C C C C C C C C C C C C C	P GAAI E TCCI S GTTC V GTTC A GTGI L CTGI L CTGI	D ATA I I CGGC R ATI Y Y S ATI Y S ATI N S AGC	C AAA K TAT JAA CO T ACA T I I I I CAA	Y AGAN FTGJ L GCCJ A CGCJ R CGCJ R F TTC F A TGC F A TGCJ SGAG	N ACI N ACI N TCI TAC I G G C R TAI I T G T G I T G I T G I I T G I I I I	H BAT( D ACT( T AGA( R STT) SAAJ E SAAJ S S A S T I I G TJ I I G TJ	H GATO D CAGO STT: V F F ACTO T ACTO M ACO N A	Y CTCI L CAG: Q FCCI STCI STTCI V AAAY STCI V GGI	K AGC' S ICT' S AAAJ K BAT' BAT' D CT' S CT' S CT' S CT' S CT' S CT' AAAJ	K IGG W CCT P ATA I CAA CTG C C C C C C C C C C C C C C C C C C	P TTG GTA GTA CCTX P GTT GGT GGT GGT GGT GGA A GCA	G IGG. W SCT A J SCT D TGJ SCT L SCC L SCC A A A TCC	X AATI N GCT/ A GCA/ A AAA/ X S CTA C L AAA/ X I CA/ K	GAA GAA TAAA KTC GAC GAC DG MTG MAG
W GGAA G AGCA S CTCT L AGTCC S CTTTI L AGTCC ATTG I AATTI N TGTA C	G GTT S CTC T NATO Y CAC P CTC S TAL Y CAC P CTC S CAC P CTC S S CAC S S C C C C S S C C C C S S C C C C	IGCT C GCTC A STGC V TTC L SAAC Q I I TGC M A C A A C Q	I F CCGC/ R CCGC/ P JATC D SATC D SAG/ E STGC V	L AAT( N FAC( Y AAT( N STT V GAA( E GGAJ GGAJ GGAJ T T T L	F GTA V CCA P CGA R TTT F CTTN L CCC T T CCC T T C C C C	P GAAJ E TCCJ S GTTC GTCJ GTCJ C C C C C C C C C C C C C C C C C C	D ATA I I CGGC R ATI Y Y S AGTI S AGTI S AGC E	C AAA K TAT J J ACA T T ACA T J I CAA Q	Y AGAI R TTGJ GCCI A GCCI R CGCI R TTGC F A TGC M I A CGCI Y SGAC G	N ACI N ACI N TCI I GCC G G G G G G TAI I T G V	H BATY D ACTY T SAAJ E SAAJ S SAAJ S S ATCJ I I GTJ C	H GAT D CAG Q STT F T T A CT A T M A A CT A T G I N A A C I	Y CTC/ L CAG: Q CC/ S CC/ S TC/ V V AAA1 K S TC/ V V AAA1 K S TC/ V R	K AGC S ICT S AAA K JAA S AT S AT S C T S C T K AAA K	K IGG W CCT P ATA I CAA CTG C C C C C C C C C C N	P TTG' GTA( V CCTV P GTT GGT GGT GGT GGT GGT GGCA( A GGCA( A GGCA( A GGCA( A GGCA( A GGCA( A GGCA( A GGCA( A GGCA( C CCTV) CCTV CCTV CCTV CCTV CCTV CCTV C	G IGG. SCT A JAT D ITG. SCT A CTC L SCC L SCC A N	Y AATI N SCTI A GCAI A AAA: X FCTO S CTAO C CTAO L AAAI X N CAI S	GAA E ACA T AAA K T F G G G G G G G G G G G G G G G G G G
W GGAA G S CTCT L AGTC S CTTTI L ATTG I AATT N TGTA C GACT	G GTT S CTC T CACO P CTC S TAP V CACO Y CACO S CACO S CATC	IGCT CGCTC AGTTCA TTCC CTTCC L CAAC Q I I CAAC Q CTTCC CTTCC CAAC	I TTC: F COTT L CCGC: R CCGCC CCGC: R CCGC: R CCGCCC R CCGCC CCGCCC R CCGCCC R CCGCC R CCGCCC CCGCC R CCGCCC R CCGCCC CCGCCC R CCGCCCCCCCC	L MAT( N FAC( Y AAT( N GTT GAA( E GGAJ G GACTJ T CTT L L CTT I L CTT	F GTA V CCA R TTV F CTTV T CCC T T CCC T T CCC T T CCC T T CCC T T CCC T T CCC T T CCC T T CCC T T CCC T T CCCA T CCCCCCA T CCCCA T CCCCA T CCCCA T CCCCA T CCCCA T CCCCA T CCCCA T C CCCCA T C CCCCA T C CCCCA T C CCCCA C CCCCCA C CCCCA C CCCCA C CCCCA C CCCCA C CCCCA CCCCA C CCCCA CCCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCA CCCA CCCA CCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCA CCCCA CCCCA CCCA CCCCC	P GAAJ E TCCJ S GTTC V GCA C C C C C C C C C C C C C C C C C C	D MTAI I I GGG R TATI Y MGTI S AGTI S AGC E SATI	C AAA FAT JAA E ACCI T ACA T ATAJ Q AAT	Y AGAN R TTGJ GCCH A CGCH R TTGC F A TGC F Y SGAC G G TTCC	N ACI ACI ACI ACI ACI ACI ACI ACI ACI ACI	H GATY D ACTY T GAO STT S SAAJ S SAAJ S S ATCJ I S GTJ C C ATTY	H GAT D CAGO Q STT F F T T T A T G A A T G A A T G A A T G A A C A A C A A C A A C A C C A C C A G C A C A	Y CTC/ L CAG Q PCC/ S TCC/ S TTC/ S TTC/ V AAAY K S TC/ V AGG/ R C TTC/	K AGC' S ICTI S AAJ K D XCTI A XCAI T AAJ K SAG	K IGG W CCT P NTA I CCT C C C C C C C C C C C C C C C C C	P TTG' GTA V CCTY P GGT GGT GGT GGT A GGT M GGT	G IGG. W SCT D BAT D TG. SCT L SCC L SCC L SCC I N SGA	Y AATI N SCTI A CAI S CAI S CTAC S CTAC S CTAC S NAGT	GAA EACA TAA K TF GA GAC BAC BAC SAC SAC SAC SAC SAC SAC SAC SAC SAC S
W GGAA G CTCT L AGTC S CTCT L AGTC I AATTI N TGTA C GACT D	G GTI S CTO T CAC P CTO S TAA V CAC P CTO S TAA V CAC Y GCO S CTO S CTO Y	t GCTC A STGC V TTC L SAAC Q ETAI I L TGC M SAAC Q C TTC L	I TTC: F CTTT: L CCGC: R CCGC: R CCGC: R CCGC: R CCGC: R D CCGC: W SAG: STGC: W SAG: V CACC: H	L MAT( N FAC( Y AAT( N GAA( G GAJ T CTTT L CTTT L L CTTT L	F GTA V CCA P CGA R TTV F C C T TA C C C N N	P GAAJ E TCCJ S GTTC V GCA C C C C C C C C C C C C C C C C C C	D MTAI I I GGG R ATT Y I ATT Y S AGT/ S AGT/ N S AGC E I ATT/ D	C AAA FAT SAA E ACCI T ACA T ATAJ Q AAT N	Y Agai R Itgj L GCC/ A CGC/ R CGC/ R TTGC F SGAC G F TTGC F	N ATCI N TCI TTA G G G G A TTA T G V X T T A	H GAT( D ACT( T AGA( R STT) SAA) S SAA) S S ATC) I C TTC I I TTC I	H GAT D CAG Q STT F F ACT Y T F ACT A T G A ACT A T G A A C A C A C A C Q	Y CTC/ L CAGY Q FTC/ S FTTV V AAAY K S TC/ R S TC/ L	K NGC' S ICTI S A X D X C I S C I X A S C I I S C I I S C I I S A I S A I S A I S A I S A A J I S I S I S I S I S I S I S I S I S I	K IGG W ICT P ATA I C C C C C C C C C C C C C C C C C C	P TTG' GTAC GTAC CCTX P GGTY GGTY GGTY GGTY GGTY G	G IGG. W SCT D BAT D CTC L SCC L SCC L SCC A N SGA G	Y MATH N SCTI A SCAI SCAI S CTAI C L MAAI K S S K K	SAA GAA EA TAA KTF GG GAC GAC SAC SAC SAC F
W GGAA G AGCA S CTCT L AGTC S CTTT L ATTG N TGTA C GACT D ACAG	G GTI S CTC T CTC S CTC S CTC S CTC S CTC S CTC S CTC S CTC S C T A C T C T C T C T C T C T C T C T C	t GCTC C GCTC A STGC V TTCC L CAAC Q L CAAC Q C TTCC L C GTCC C C C C C C C C C C C C C C C C	I TTC: F CCCGC R CCCGC P CCCGC P CCCGC P CCCGC D CCCGC D CCCGC D CCCGC D CCCGC D CCCGC C P CCCGC P CCCGC D CCCCGC D CCCGCC D CCCGCC D CCCGC D CCCGC D CCCGC D CCCGCC D CCCGC D CCCGCC D CCCGCC D CCCGCC D CCCCCGCC D CCCGCC D CCCGCC D CCCGCC D CCCGCCC CCCGCC D CCCCGCCC CCCCGCCCCCCCC	L MATC N FACC Y AATC N STTT V SAAC G G G G ACT/ T T L C TC/ L AAAC	F GTA: V CCA P CGA F CTT L CCA TTA CCA TTA CCA N CCA CCA	P GAAJ E TCCJ S GTTC V GCA C C C C C C C C C C C C C C C C C C	D ATAI I I I I I I I I I I I I I I I I I	C AAA FAT SAA GAA T T CAA T SAA Q N SAA	Y AGAJ R ITGJ L GCCJ A CGCJ R CGCJ R TTC G G G TTCC F SGAC G F TTCC SGAC G SGAC	NATC NATC NACJ ITAC ITAC G G G R TAL ITG V C TA I TG V C TA I TG V	H GAT( D ACT( T AGA( R STT) SAA) SAA) SAA) I SAA) I C I GT) C I SAG( SAG( SAA)	H GAT D D CAG Q STT F V T F ACT C T M A ACT N N A ACT N N A ACT N N A ACT S A C A A C A A C A C A C A C A C A G C A C A	Y CTC/ L CAG' Q CCC/ S TCC/ S TTC S TTC/ R S TC/ L CTTC/ L TTT	K AGC' S ICT' S AAAJ AAAJ AAAJ AAAJ K SACT' T AAAJ K SAGJ E ICT'	K IGG V C C T C C C C C C C C C C C C C C C C	P TTG' GTA V CCTV P GGT GGT GGT GGT GGT GGT GGT GGT MAA	G IGGL SCT J J J J J J J J J J J J J J J J J J J	Y MAT N SCTI A SCAI SCAI SCAI SCAI SCAI SCAI SCAI S	M GAA E ACA T AAA K T F GG G G G G A T G A C A M G G C A T F G A C A T F G A C A T F G A C A T F C A T F A C A T T F A C A T T F A C A T A A C T T F A C A T A A C A T A A C A T T F A C A T T F A C A T T F A C A T T F A C A T T F A C A T T F A C A T T F G G A C T T F G G G A C T T F G G G C A T T F G G G G C A T T F G G G G C A T T F G G G G C C T T F G G G C C T T C T G G G G C C T T C G G G G
W GGAA G CTCI L AGTC S CTTI L AATTG I AATTI N TGTA C GACT D ACAG	G GTI S CTC T ATC Y CTC S TAM Y CTC S TAM S CTC S TAM Y TAC Y TAC Y TAC Y	t GCT C SCTC A STGC V TTC L CAAC Q TTC A A A C TTC C TTC C C TTC C C C C C	I TTC/ F CCCC/ R CCCC/ P CCCC/ P D CCCC/ D D CCCC/ P D CCCC/ D D CCCC/ D D CCCC/ D D CCCC/ D D CCCC/ D D CCCC/ D D CCCCC/ D D D CCCCC/ D D D CCCCC/ D D D D	L MAT( N FAC( Y AAT( N STT) V SAA( E SGAJ) G G SGAJ T T L CTT] L L AA( K	F GTA V CCA P CGA R TTV L CCC TTV L CCC TA CCC N CCG N CCG N CCG	P GAAJ E TCCJ S GTTC V GCAC C C GTCJ L C C C C C C C C C C C C C C C C C C	D ITAL I I I I I I I I I I I I I I I I I I I	C AAA FAT JAA JAA CCI T ACCI T ACCI T ACCI P CAA Q AAT N SAA SAA	Y AGGAI R ITTGJ L GGC/ R CGC/ R CGC/ F CGC/ F GGAC C F TAC/ F SACC D	NATC NATC NACJ ITAC ITAC GAJ TAC TAC TAC TAC TAC TAC TAC TAC TAC TAC	H GATY D ACTY T AGAO R STT S AAAA S ATCA I G TA S AAAA S ATCA I I G TA S A ATTY I I AGAO S A S A A A S A S A S A S A S A S A S	H GAT D CAG Q Q STT F F ACT T F ACT A T G A ACT A T G A A C A A C A A C A A C A A C A A C A A C A C A G A C A G A C A G A C A G A C A G A C A G A C A G A C A G A C A G A C A G A C A G A C A G A C A G A C A G A C A G A C A G A C A C	Y CTC/ L CAG' Q V CC/ S TC/ S TTC/ S TTC/ L S TC/ L TTT F	K MGC' S ICT' S A A BAT' D CT' S A CA' T A A A A CA' T S A CA' T S A CT' S A CT' S A CT' S A C' S A C' S A A C' S S ICT' S A A A S ICT' S A A A A S ICT' S A A A A S ICT' S A A A S ICT' S A A A S ICT' S A A A S ICT' S A A A S ICT' S A A A S ICT' S A A A S ICT' S A A A S ICT' S A A A S ICT' S A A A S ICT' S A A S I CT' S A A S I CT' S A A S I CT' S A A S I CT' S A A S I CT' S A A S I CT' S A A S I CT' S A A S I CT' S A A S I CT' S A S I CT' S A S I C I S A S I S A S I S A S I S A S I S A S I S A S I S S I S A S I S S I S S I S S I S S I S S I S	K IGG W CTP I CAA CTG CTG CTG L CC L CC N K SAA SAA	P TTG' GTA( V CCTV P GGTY GGTY GGTY GGTY GGA( A GGTY GGA( A AAA) K	G IGGL SCT J J J J J J J J J J J J J J J J J J J	Y AATI N GCT/ A GCA/ A GCA/ S CTA( S CTA( S CTA( S CTA( S K K CA/ S K K CA/ S K K TCA/ S K K TCA/ S S CA/ S CA S CA	N GAA E ACA TAA K F GGA G G G G G G G G C M M G C C
W GGAA G CTCT L AGTC S CTTTI L AATTG I AATTI N TGTA C GACT D ACAG T C	G G G S C T A T C A C A C A C A C A C A C A C A	TGCT CGCT CGCTC CGCTC CGCTC CGCC CGCC C	I FTC/ F SCTTY L CGC/ R CGC/ P SAC/ D SAC/ B SAC/ B STGC X C C C C C C C C C C C C C C C C C C	L MATC N FACC Y AATC N STTY V SAAC G SGAC TY T C CTTY L C TC L L AAAC K	F GTA V CCA P CGA R TTV L CCA TTV L CCA TTV L CCA TTV L CCA N CCG N CCG N CCG N CCG N CCG N TTV	P GAAJ E TCCJ S GTTC V GCAC V CTCJ L CAGC Q CCAC C CAGC Q CCAC C CAGC C CAGC C CAGC C CAGC C C CAGC C C C	D ITAL I I I I I I I I I I I I I I I I I I I	C AAA K IAT JAA T JAA T JAA T JAA N SAA K SAA K SAA	Y AGGAI R ITTGJ GCC/ A CGC/ R TTTC F GGAC G F SACC D D SACI D	NATCHAR	H BATY D ACTY T AGAO R STTT SAAJ SAAJ SAAJ SAAJ SAAJ SAAJ SAAJ	H GAT D D CAG Q Q TT F T T T T T T A CT A C A C A C A C A C A	Y LCTCJ LCAG: QQ FICCJ S TTC Y V AAATY K STTCL L TTTT F F TTTL	K MGC' S ICT' S A A BAT' D CT' S A CA' T S A CA' T S A CA' T S A CA' S A CA' S A CA' S A CA' S A CA' S A C C S S ICT' S A A J S C C S S ICT' S A A J S C C S S ICT' S A A J S C C S S A A J S C T' S A A J S C T' S A A J S C T' S A A J S C T' S A A J S C T' S A A J S C T' S A A J S C T' S A A J S C T' S A A J S C T' S A A J S C T' S A A J S C T' S A A J S C T' S A A J S C T' S A A J S C T' S A S C T' S A S C T' S A S S C T' S A S S S S S S S S S S S S S S S S S	K FGG W CCT P ATA CAA CCT CC CC CC CC CC AC NAA K SAA SAA SAA SAA	P TTG' GTA CCTV CCTV P GGTY GGTY GGTY GGTY GGTY GGTY GGTY GG	G IGGL SCT SAT SAT SAT SCT A SCC SCC SCC SCC SCC SCC SCC SCC SCC S	Y AAT N GCT/ A GCA/ A GCA/ S GCA/ S CTA( S CTA( S CTA( S CTA( S CTA( S CTA( S CTA( S CTA( S CA/ S CA S CA	A GAA E ACA TAA K TTC F GGA GAC BACG SAC SAC F F GCC TT F CCC TTC CCCTT
W GGAA G S CTCL AGTC S CTTI L AGTC S CTTI I AATTG I AATTG G ACT D G ACAG T AGCC	G G S C T C A C C C C C C C C C C C C C C C C	TGCT C SCTC A STGC Z Z A A C A A C C C C C C C C C C C C	I TTC: F F CCCC R CCCC P CCCC P SAC: D D SAC: SAC: STCC V CACC H G G STCC V CACC STCC V CACC STCC V CCCC V CCC V CCC V TTY R R CCCC P CCC V TTY R R CCCC V TTY R R D CCCC V TTY R R D CCCC V TTY R CCCC V TTY R D D CCCC V TTY R D D CCCC V TTY R D D CCCC V TTY R D D CCCC V D D CCCC V TTY R D D CCCC V D D CCCC V TTY R D D CCCC V D D C CCCC V D D C CCCC V D D C CCCC V D D C CCCC V S CCCC V D D C CCCC V S C CCCC V D D C CCCC V C C C CCCC V D D C CCCC V C C C C	L MAT( N FAC() Y M GAT() SAT() SAA( C GGAJ G G G C T T L C T C T T L C T C T T L C C T T C T T C T C	F GTAJ V CCA P CGA R F T C CA C C C C C C C C C C C C C C C C	P GAAJ E TCCJ S GTTCJ GGTCJ GGCA' A GTGJ CCCA' CCCA' CCCA' CCCA' CCCA' CCCA' CCCA' CCCA' CCCA' CCCA' CCCA' CCCA' CCCA' CCCA' CCCA' CCCA' CCCJ CCCJ	D MTAL I I CGG( R TAT/ Y Y MGT/ S AGT/ D S AGT/ D CTTX L MGT/	C AAA FAT JAA JAA T JAA T J JAA CA T J JAA CA SAA C A SAA SAA SAA	Y AGGAI R GGCI A CGCI R TTTCC F ATGC G G TTTCC G G G C TTTCC G G AGI G AGI G AGI G C G AGI G C C G C C C C C C C C C C C C C C C	NATCHAR	H GATY D ACTY T AGAO R STT S AGTI S ACTO I GTI I GTI I S ACTO A	H GAT D D CAG Q Q STT F V T T T S ACT Q C ACT Q Q C ACT Q Q S AT C D	Y CTCJ L CAG: Q FCCJ S S TTC V S TTC V AAA S TTV K S TTV F F TT J V	K MGC' S ICT' S A A A A A A A A A A A A A A A A A A	K FGG W CCT P ATA CAA CCT CCT CCT CCT CCT CCT	P TTG' GTA: GTA: V CCTX GGT GGT GGT GGT GGT GGG GGT GGG GGT K CTG GGT CA CTG GGT CA CTG CA CTG CA CTG CA CTG CA CTA CCTX CCTX CCTX CCTX CCTX CCTX CC	G IGG. W JAT( D JAT( D CTC: A CTC: A CTC: C C C C C C C C C C C C C C C C C	Y MATH N GCTI A GCAI S GCAI S CTAI S CTAI S CTAI K N AAAI K S CTAI X S CTAI X S CTAI X S CTAI X S CTAI X S CTAI X S CTAI X S C S C A S C S C	MA GAA E CA T AAAA KITC F GGA G G G G G G C T F G C C T T F C C T T
W GGAA G S CTCT L AGTC S CTTT L AATTG I AATTG GACT GACT GACT S C C C C C C C C C C C C C C C C C C	G G T S C T C A C C C C C C C C C C C C C C C C	TGCTI CGCTICA STGCY TTCC LCAAC QCTTCC QCTATI I STCC QCTCC R STCC R STCC QCTCC R STCC QCTCC R STCC QCTC	I TTCI F COTTI L CGCJ/R CCGC/ BAT(C D GGC/ BAC/ BAC/ BAC/ BAC/ BAC/ BAC/ BAC/ S	L MAT( N FAC() Y STTT V SAA() E SGAJ C TTT L C TTT L C TTT L C C TT T	F GTA V CCA P CGA R TTF T T T C C C C C C C C C C C C C C C	P GAAJ E TCCJ S GTTC GCA: A GTG? CTCJ L CTCJ L CAGC Q CCAC P CCAC T T S	D MTAL I MTT I CGGC R R MTT Y Y MGT/ Y SAGC E MTT L C C	C AAA FAT JAA JAA JAA T JCT T JCA T JCA T JCA N SAA K SAA K	Y AGAN R TTGJ L GCCA A CGCA F CGCA F CGCA F SGAC D G G G AGA F G SAGA E E	N IATO NACI NACI NACI NACI NACI NACI NACI NACI	H JATY D ACTY T NGA R S STTY JAA S S A M TC/ I S S C NTC/ I JAG C A	H GAT( D CAG( Q STT) V F F T T T T T T T A T G A T G A T G A T G A T G A T G A T G A C T C T C T T T T T T T T T T T T T T	Y CTC/ L CAG: Q FCC/ S TTC/ S TTC/ S TTC/ R S TC/ R S TC/ L TTT F S TA/ Y	K MGC' S ICT' S AAA B T S ACA T S ACA T S ACA S ACA K S ACA K S ACA K S ACA K S ACA K S ACA K S ACA S S ICT' S S A A A S S S ICT' S S A A A S S S S S S S S S S S S S S	K FGG W CCT P ATA I CAA CCT CCT L CCT CCT AA CCT SAA SAA SAA CCT D	P TTG' GTA V CCTX P GGTV V GGTV GGTV GGTV GGTV GGTV GGTV	G IGG. W JCT D JAT D C TG L C C C L C C C C L C C C C C C C C C	Y AATI N SCTI A SCAI A SCAI A AAAI SCTA K SCTA K SCAI SCTA X SCTA X SCTA X SCTA	N GAA E AACA T AAAA K T F GGA G GGA S GGA S GGA S GGA S GGC T F I GC S T T V
W GGAA G S CTCT L AGTC S CTTTI L AGTC S GACT D ACAG T C ACAG T S GATG	G GTI S CTO Y CTO Y CTO S CACO P CTO S CACO Y CTO S CACO Y CTO S CACO Y CTO C TGTI C TGTI	t GCTIC A STGC TTC C TTC C C TTC C C C C C C C C C	ITCH FTCH FTCH FCTTF L CCGC/ P CCGC/ P CCGC/ P CCGC/ P SATO D GGG/ W SAG/ S TGC/ S TGC/ S TGC/ S TGC/ S S TGC/ S S TGC/ S S TCT S C TTF S L CCGC/ P S S T C S S S S S S S S S S S S S S S S	L MAT( N FAC() Y M GTT/ V GAA( E GGAJ G G CTT/ T L C TTT/ L L AA( K K CC1 T GCT( C	F GTAJ CCA P CCA R TT CCA R TT L CCA TT L CCA N CCA N CCA N CCA N CCA N CCA N CCA N CCA N CCA N CCA N CCA N CCA N CCA N CCA N P CCA N R TT L CCA N CCA N R CCA N R CCA N R CCA N R CCA N R CCA N R CCA N R CCA N R CCA N R CCA N R CCA N R CCA N R CCA N R CCA N R CCA N R CCA N R CCA N R CCA N R CCA N C CCA N R CCA N C CCCA N C CCA N C CCA N C CCA N C CCA N C CCA N C CCA N C CCA N C CCA N C CCA N C CCA N C CCA N C CCA N C CCA N C C C C	P GAAJ E TCCJ S GTTC GCA: GTCJ L CTCJ L CAGC Q CCAC P CCAC P CCAC S GGTC	D MTAL I MTT: I CGGU R TATI Y Y MGTI S AGTI S AGTI D TTX L MGTI C C TCI	CAAA KIAT SAA SAA CCI T XCA T XCA T XCA T XCA T XCA N SAA N SAA KIT X SAA	Y AGGAI R TTGJ L GCCJ GCCJ GCCJ GCCJ GCCJ F CCCC F GCCJ F GCCD F F GCCD F F GCCD F GCCD F F GCCD F F GCCD F F GCCD F F GCCD F F GCCD F F GCCD F GCCD F GCCD F GCCD F GCCD F GCCD F GCCD F GCCD F GCCD F GCCD F GCCD F GCCD F GCCD GCCD	NATO NATO NACINITA INTO INTO INTO INTO INTO INTO INTO INTO	H BATY D ACTY T AGA STT S AAA S S AAA S S AAA S C V T C S AGT S AGT S AGT S AGT S AGT S AGT S AGT S AGT S AGT S AGT S AGT S AGT S AGT S AGT S A S T S S T S S T S S S T S S S T S S S T S S S S T S	H GATY D CAGGO Q STTY F V V T T A T C A C C A C C A C C A C C A C C A C C C A C	Y CTC/ L CAG' Q CCZ S STC/ S TTC/ V AAAT R S TTC/ P S TA/ S TA/ S TA/ S TA/	K MGC' S NAAJ K BAT' D K C T' S AAJ K S AGJ S C T' S AAJ K K AAJ	K FGG CTP ATA CAA CTL CCC CTA AAA SAA SAA CCT CTA	P TTG' GTA V CCTV GGTY GGTY GGTY GGTY GGTY GGTY ACTY T CCC/	G GIGG W SCT D SAT L SCT L SCT L SCT L SCT L SCT L SCA T T T C D SCA T T T D SCA T C D SCA T C D SCT L S C T C D S C T C D S C T S S T T G S S A T S S S S	Y AATI N SCTI A SCAI A AAA SCAI SCTAC SCTAC SCAI SCAI SCAI SCAI SCAI SCAI SCAI SC	A GAA E GAA MAAA K TTC F GGG G GAC D ATG SAC STT F GC STT V ACA
W GGAA G S CTCI L AGTC S CTTI L CTTI AGTC AATT N AATT D AATT D AATT D AATT S C GACT S GACT C C T C C C C C C C C C C C C C C C	G GTT S CACO Y CACO P CACO S TAL S CACO S TACO Y GCT C S TACO Y GCT C TACO Y C TACO Y C TACO Y C TACO Y C TAC Y C TAC Y C TAC Y C TATO Y S C ATO Y S C ATO Y S C ATO Y S C ATO Y S C ATO Y S C ATO Y S C ATO Y S C ATO Y S C ATO Y S C ATO Y S C ATO Y S C ATO Y S C ATO Y S C ATO S S C ATO Y S C ATO S S C ATO S S S C ATO S S S C ATO S S S C ATO S S S C ATO S S S S C ATO S S S S C ATO S S S S S S S S S S S S S S S S S S S	ticctic CCTCCA STGC STGCA STGC	ITCH FTCH FTCH CCTTF L CCGC/P	L AAT( N FAC() Y STTT V SAA( E SGAJ G G A T T L L A A C CT C K C CT C A	F GTAJ CCA P CCA P CCA T T C CCA C C C C C C C C C C C C C	P GAAI E TCCIS GTTC GCA: CGCA: CTCI L CAGC Q CCA: P ACAC T CAGC CCA: SGTC G	D MTAL I MTT: I CGGC R MATI S AGTI S AGTI S AGTI C TTC V V	CAAA KIIAT SAA SAA CCI TAI SAA NCCI TAI I CCT P CAA N SAA KII C SAA C	Y AGAJ R ITGJJ L GCCJ R CGCJ R CGCJ R CGCJ R CCC R TTTC G GAGJ SAGJ SAGJ SAGJ I I I I I I I I I I I I I I I I I I I	NATO NATO NACINIACI INTA GGAINI GGAI ITA TA TA TA TA TA TA TA TA TA TA TA TA T	H JATY D VCT T VGAU R S V V S AAJ S S V T C N T C N T C N T C N T C N T C N T C N T C N C N	H GAT( D CAG() Q STT' F F ACT( T T ATG/ M ACC( N N AAC() N AAC() N AAC() S AT() C AAC Q C AAC Q C AAC Q C AAC C C C C C	Y CTC/ L CAG' Q CC2/ S TCC/ S TCC/ S TCC/ S TCC/ S TCC/ S TCC/ S TCC/ S TCC/ S TCC/ S TCC/ S TCC/ S TCC/ S S S TCC/ S S TCC/ S S TCC/ S S TCC/ S S TCC/ S S TCC/ S S TCC/ S S TCC/ S S TCC/ S S TCC/ S S TCC/ S S TCC/ S S TCC/ S S TCC/ S S S S S S S S S S S S S S S S S S	K MGC' S MAAJ K BAT' D MAAJ K MAAJ K K MAAJ K K K K K K K	K FGG CTP ATA CAA CTL CCTA CTL CCTA AA SAA SAA CTN P	P TTG' GTA( V CCTV GGTY GGTY GGTY GGTY GGTY GGTY GGTY GG	G GIGG W SCT A SAT D TTG L SCT L SCT L SCT L SCT M SGAI SGAI M M	Y MATH N GCTI A GCAI A GCAI S CTAC S CTAC S CTAC S CTAC S CTAC S A S CTAC S A S CTAC S A S CTAC S A S C C C C C C C C C C C C C C C C	A GAA GAA TAAA KTF FGG GGGC DGGC MAGS TFF GC STV ACA
W GGAA G CTCT L CTCT C CTTT L AGTC S CTTT I AATT N TGTA C GACT GACT S GATG GAAG	GTI S CTC T ATC T ATC T C C C C C C C C C C C	TGCTI CGCTICA CCTICA CTTCCA CTTCC CCTCCA CCT	ITCL FTCL CCCC R CCCC P CCCC P CCCC P CCCC P CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC	L AAT( N FAC() Y AAT( N STT7) V SAA( E SGAJ STT7) L SGAJ L STT7) L SCT() A A ATT7)	F GTAJ CCA P CCA P CCA F CCA TTA I CCA N CCA P TCA I CCA N CCA P TCA I CCA TTA I CCA TTA I CCA TTA I CCA TTA I CCA TTA I CCA P CCA P CCA P CCA P CCA P CCA P CCA P CCA I CCA P CCA P CCA I CCA P CCA I CCA P CCA I C C C C	P GAAJ E GTTCCJ S GTTC V GCA: A GTCJ L CTCJ L CTCJ L CTCJ CTCJ S GTC Q CCA: Q CCA: Q CCA: S GTC C CACA C CACA C CACA C C CACA S C C C C	D MTAL I MTT: I GGC RATH Y MGTH Y MGTH SAGK E MGTH C C TTY L ATC V MGTH	CAA KTY JAAU EGCT CAAU CAAU TAA I CAAU EGTI CAAU EGTI CCT	Y AGAJ R ITGJJ L GGCJA CGCJ R CGCJ R CGCJ R TTTC GGCJ R TTTC GGCJ S GGCJ S GGCJ S GGCJ S GGCJ S G G S G S G G S G G S G G S S G C J S S G C J S S C S S S S S S S S S S S S S S S	N TO NATO NATO NACINA IN G G R I I TA I C C I I I C I I C I I I I C I	H JATY D ACT T JGAU R STT J SAAJ S S T J J C C T C J AGC T C A C C A C C C A C C C C C C C C C	H GAT( D CAG( Q STT: F T T T T S T T S T S T S T S T S T S T	Y CTCI L CAG' S CCI S S CCI S S T CCI S T CCI S T CCI S T CCI S T CCI S T CCI S S T CCI S S T CCI S S CCI S S CCI S S CCI S S CCI S S CCI S S CCI S S CCI S S CCI S S CCI S S CCI S S CCI S S CCI S S CCI S S CCI S S CCI S S T C CI S S S T C CI S S S T C CI S S S T C CI S S T C CI S S T C CI S S T C CI S S T C CI S S T C CI S S T C CI S S S T C CI S S S T C CI S S T C CI S S S T C CI S S S S S S S S S S S S S S S S S	K MGC' S S A C T S A C T S A C T S A C T S A C T S A C T S A C T S A C T S A C T S A C T S A C T S C T S C T S C T S C T S S A D S C T S S A D S C T S S A D S C T S S A D S C T S S A D S C T S S A D S C T S S A D S C T S S A S S S S	K IGG CT P ATA I CA C C C C C C C C C C C C C C C C C	P TTG' GTA CCTV CCTV GGTV GGTV GGTV GGTV GGTV GGTV	G G G G G G G C T G G C C C C C C C C C	Y AATI N SCTI A GCAI A SCAI S CTAC S CTAC S CTAC S S CTAC S S CTAC S S CTAC S S CTAC S S CTAC S S CTAC S S C T T C S C T T C S C T T C S C S	A GAA E GAA A CA T A A K TC F GGG G GAC G G G G G G G G G G G G G G G G G G G
W GGAA G AGCA S CTCTI L AATTG T AAGTC I AATTI N AATTI N AATTI D GAATG GAATG GAATG GAATG GAATG GAATG GAATG C	GTI SGTI SGTO T CAC P CAC P CAC S ATO S ATO S C T ATO C T G T G T G T G T G T G T G T G T G T	t GCTC C C C C C C C C C C C C C C C C C	ITCH FTCH FCH CCTT RCCGCC RCCGCC RCCCGCC RCCCGCC RCCCGCC RCCCGCC RCCCGCC RCCGCC RCCGCC RCCCGCCC	L AAT( N Y VAT( N JTT V JAA( E GGAJ T T L JAA( K T T L AAA( K T T I I I I I I I I I I I I I I I I I	F GTA: CCA P CCA F CCA F CCA T T CCA T T CCA T T CCA T T CCA T T CCA T T CCA T T CCA T T CCA T T CCA T C CCA T C CCA T C CCA T C CCA T C CCA T C CCA T C CCA T C CCA T C CCA T C CCA T C CCA T C CCA T C CCA T C CCA T C CCA T C CCA T C CCA T C CCA T C CCA T C C C C	P GAAI E GTTC S GTTC V GCA: C C C C C C C C C C C C C C C C C C	D VITAL I VITI I CGGK VITI V V V V V V V V V V V V V V V V V	CAA KTY SAAU TAS CCT TAS CCT NAAU CCT CAAU CCT CAAU CCT	Y AGGAJ R ITGJJ L CGC/J R CGC/J R CGC/J R TYTC G G G G G G G G G G G G G G G G G G G	NATO NATO NATI IN I TAI I GCO GAJ I TGI V CA I AN KATO C C I	H JATY D ACTY T JAAA STT JAAA STT JAAA STT JAAA STT JAGY JAAA SCT A SCT S	H GAT( D CAG() Q STT' F T T T T T S T T S T T S T T S T T T T	Y CTC/ L CAG' Q Q CS STC/ STTV STTV STTV STTV STTV L CTAJ L CTAJ L	K MGC' SAAJ K JAT' SAT' SAT' SAT' SAT' SAT' SAA' K SAG' SAA' K K AA' K K SACT' S	K G G G C C C C C C C C C C C C C	P TTG' GTA' V CCTV GGT GGT GGT GGT GGG GGT GGG MAAA' K CCC' P ACAJ T	G IGG. W SCT A SAT C SCT L SCT L SCT C A T C C C A SGA I T C C M M SGA M M T G M M M C M M	Y AATI NGCTI SCAI SCAI SCAI SCAI SCAI SCAI SCAI SCA	A GAA E ACA T A X T F GG G G G G G C T T F G C S T T F G C S T T F G C A T T A X T F G A S A C A T A X T F G A S A C A T A X T A X T F G A S G G G G G G G G G G G G G G G G G
W GGAA G GACA S CTTTI L AGCT S CTTTI N TGTA C GACT D ACAG GACA G GAAG G GAAG C CTTA	G GT G GT S CTC T CAC P CAC P CTC S TAP V CTC S TAP V CTC S TAP V CTC S TAP V CTC C TAP V CTC C T C C T C C C C T C C C C C C C C C C C C C	" GCTC CCTC ATGC TTC LAC Q TTC LAC Q TTC LAC Q TTC LAC Q TTC CTC P TTT	I I I I I I I I I I I I I I I I I I I	L AAT( N FAC() Y AAT( N GTT) V GAA( E GGAJ T CTT) L CTT) L CTT C A CTT T T T T T T T T T T T	F GTA V CCA P CCA R TTV L CC TTV L CC TTV I CC N CCGA N CCGA N CCGA TTC I CC CCGA N CCGA TTV I CCA N TTV I CCA N C CCA N CCA N C CCA N C CCA N C CCA N C CCA N C CCA N C CCA N C CCA N C CCA N C C CCA N C C CCA N C C CCA N C C C C	P GAAJ E GAA S GTTC V GCAT C C C C C C C C C C C C C C C C C C	D VTAI I CGGU VATA I CGGU VATA V SATA V SATA V SATA V SATA V V V ATCI V V TTT	CAAA KTY JAAU JAAU TAJ ITAJ ICTY CAAU Q VATU N SAAU SAAU CCTY CCTY	Y AGAN R TTGN GCCN GCCN CGCN F CGCN F CGC G G G G G G G G G G G G C C C C C	NATCANAL AND	H BATY D ACTY T AGA( R FTT) V A A B T C A C C A C C A C C A C C A C C A C C A C C A C C A C C C A C T C C C C	H GAT( D CAG() Q STTT F F T T T T S ACT( N A ACC( N A T A CAAC Q C S AT( D F T C C A C T T C T C T C T C C C C C C C	Y CTC/ L CAG' Q CCC/ STTC/ STTC/ STTC/ STTC/ STTC/ L CTA/ L CTA/ L CTA/ L CTC/ L CTC/ L CTC/ L CTC/ L CAG' CCC/ STTC/ L CAG' CCC/ STTC/ L STC/ STTC/ L STC/ STTC/ L STC/ STTC/ L STC/ STTC/ STTC/ L STC/ STTC/ STTC/ L STC/ STTC/ L STC/ STTC/ L STC/ STTC/ L STC/ STTC/ L STC/ STTC/ L STC/ STTC/ L STC/ STTC/ L STC/ STTC/ L STTC/ STC/ STC/ STC/ STC/ STC/ STC/ STC	K MGC' SAAJ K JAAJ JAT' SAT' SAT' SA	KGG UCP PALA CONTRACTOR CONTRACTOR CONTRACTOR CONTRACTOR CONTRACTOR CONTRACTOR CONTRACTO	P TTG' GTA' CCTY CCTY GGT GGT GGT GGG GGA GGT GGG A GGT CCC/ P A CC/ P CCC/ P CCC/ CCC/ CCC/ CC	G IGG. W SCT A SAT C SCT L SCT L SCT L SCT L SCT C A TC M SGAI SGAI M GC M M GC	X AATI N GCTI A GCAI A CTAC S CTAC S CTAC S CTAC S CTAC S A A GCTC A A A GCTC A C A C C T C C T C C T C C T C C T C C T C	A GAA E ACA TAA KTC F GG G G G S A C T F G G C T F G C C T V A C T I C C T I C C T I C C T C C T C C T C A S A C A S A C A S A C A S A C A S A C A S A C A S C C S C C S C C T A S C S C C S C C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S
W GGAA G G AGCCTCT L AATTG AATTG AATTG AATTG N TGTA C GACTG GACTG G GACG G GATG G GATG G G GTTA S C GATG G G G C T C C C C C C C C C C C C C C	GTI SGTI ATC YATC YATC YCTC SATC YCTAL YCC SATC YCTAL YCC SATC YCC TAL YCC YCC SATC YCC TAL YCC YCC SATC YCC C TAL YCC YCC SATC YCC YCC SATC YCC YCC SATC YCC YCC SATC YCC YCC SATC YCC YCC YCC SATC YCC YCC YCC SATC YCC YCC SATC YCC YCC YCC SATC YCC YCC YCC SATC YCC YCC YCC SATC YCC YCC YCC SATC YCC YCC SATC YCC YCC YCC SATC YCC YCC YCC SATC YCC YCC SATC YCC YCC YCC YCC SATC YCC YCC YCC SATC YCC YCC SATC YCC YCC YCC SATC YCC YCC YCC YCC YCC YCC SATC YCC YCC YCC SATC YCC YCC YCC YCC YCC YCC YCC YCC SATC YCC YCC YCC YCC SATC YCC YCC YCC SATC YCC YCC SATC YCC YCC YCC SATC YCC YCC YCC SATC YCC YCC YCC SATC YCC YCC SATC YCC YCC SATC YCC YCC SATC YCC YCC SATC YCC YCC SATC YCC YCC SATC YCC YCC YCC SATC YCC YCC SATC YCC YCC YCC SATC YCC YCC YCC YCC YCC YCC YCC YCC YCC Y	t GCTC C C C C C C C C C C C C C C C C C	I TTTT F TTTT LCGC/ RCGC/ BAT(C BAT(C BAT(C BAT(C)	L AAT( N FAC( N STT SAA( SGA) T SGA) CTT L CTT L CTT L CTT L CTT I CCT C A T SCT C A T T SCT C A T T I C TT I I C CTT I I C C T T I C C T T T T	F GTA VCA P CCA R TTV L CCA TTV L CCT N CCA N CCA F TTV L CCA N CCA F TTV L CCA N CCA TTV L CCA TTV L CCA R TTV L CCA N C CCA N C CCA N CCA N C CCA N C C C C	P GAAJ E GAA S GTTC V GCA V C C C C C C C C C C C C C C C C C C	D VTAI I VTTI I CGGG R R I V I V V V V V V V V V V V V V V V	CAA KIAT JAA TAA TAA TAA ICT JAA SAA KIT SAA KIT SAA KIT SAA KIT SAA KIT SAA KIT SAA KIT SAA KIT SAA	Y AGAJ R ITGJ GCCJ GCCJ R CCCJ R CCCJ R F CCCJ R F CCCJ R F GAGG D GAGJ E GAGJ E CCCJ C S C C C C C C C C C C C C C C C	NATO NATO NATITA ITA GG GAI ITA TA GG GAI ITA ITA ITA ITA ITA ITA ITA ITA ITA I	H BATY D ACTY T A R BTTY S AAAA S BTTY S AAAA S C T T C C T T C C T T C C T C C C C C	H GAT( D CAG() STT F F MCT() T MAC() N MAC() N MAC() D SAT() D T T CAA() C A C C C C C C C C C C C C C C C C C	Y CTCI L CAG: Q CCC STTC T STTC T STTC V AAAN STTCI V STTCI F STAI CTAN L CTAN L CTAN	K NGC' S S A A B A C T A A C A C A C A C A C C A A C C C A A C	K IGG W CCT P IA CCT I CCCC I CCCC I CCCCCCC I CCCCC I CCCCC I CCCCC I CCCCCCC I CCCCCCCC	P TTG GTA V CCTV P GGTV GGTV GGTV GGTV GGTV GGTV GGTV	G IGG. W SCT' A J CTC' L CTC' L CTC' L C C C C L C C C C L C C C C C C C C	Y AATH NGCTIA SCAI SCAI SCAI SCAI SCAI SCAI SCAI SC	A GAA E AACA T AAA K ITTC F GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC
W GGAA G GAA CTCT L CTCT L CTTT L AAGTC S CTTT N TGTA C GATG G GATG D GAAG C C GATG C CTTA S GATG C C CATA N C C CTCT C CTTT I C CTCT C C C CTCT C C C C	GTT SCTC TATC TATC CTC CTC TATC CTC CTC SCT V CTC SCT V CTC TGT V CC C C TGT V C C C C C C C C C C C C C C C C C C	"GCTC CCTCL STTCL STTCL SAAC QCTAJ I I SAAC QCTCC P TTT I I ACA	I TTTT F F CCGC/ R CCGC/ P SAT(C P SAT(C P SAT(C P SAT(C SAT(C) S S S S CCGC/ P S S S S CCGC/ P S S S T G C S C C C C C C C C C C C C C C C C C	L AAT( N FAC( Y AAT( N STTT) SAA( STTT) L SAA( SCTV L SCTV L X C T T SCTC A T T T SCTC A T T T SCTC T T T T T T T T T T T T T T	F GTAN GTAN CCA' P CCA' F TT L CCI TTA I CCC N CCA P TT L CCI N CCA P TT L CCI N CCA P TT L CCI N CCA P TT L CCI N CCA' N CCA' N CCA' CCA' P CCA' P C	P GAAJ E GTATC S GTTC V GCAT C CCAT C C C C C C C C C C C C C C C	UTAL ITAL ICGGG RATI YATY YATY YATY SAAN BAGG ETA VATO NGTI VATO NTTI I	CAA KIT JAA TAA TAA TAA TAA TAA TAA TAA SAA SAA S	Y AGAN R TTGJ GCC/ A CGC/ F CGC/ F CGC/ F CGC/ F S GCC F S CGC/ F S S CGC/ F S S CGC/ F S CGC/ F C S CGC/ F C CGC/ F C CGC/ F C CGC/ F C CGC/ F C CGC/ F C CGC/ F C C CGC/ F C C C C C C C C C C C C C C C C C C	N LACIN NACING GAINS CONTRACTOR C	H BAT( D CT( T NGA( R TT) S A S TT) S A S S TC) I S C T S A S C T S C C S C C S C C S C C S C C S C T S S C T C T	H GATY D CAGU Q STTY F F ACTY T F ACTY T AACCO D S ATGJ M AACCO D S ATGJ D F F CAC CAC T T S S C CAC C CAC C C C C C C C C C C C C	Y CTCJ L CAG: STTC T STTC T V AAAN K STCJ V V AAAN K STCJ L TTTT L TTAN L C TAJ L C TAJ	K S S S A K S A C T S A C S C A C T S A C T S A C T S A C T S A A C T S A A A C T S A A A A C T S A A A C T S A A A A C T S A A A A C T S A A A A A A A A A A A A A	KGG UCT PAN I A QUARTER CONTANT A QUARTER CONTANT A QUARTER CONTANT A A A A A A A A A A A A A A A A A	P TTG C GTA C C C C C C C C C C C C C C C C C C C	G IGG. W SCT' J J TCC L GCT' L GCC' N J GCC' N SGAI U D J GCC' N M GC M M GC' N M GC' SAT'	Y AATH NGCTI SCTI SCTI SCTI SCTI SCTI SCTI SCTI S	A GAA E GAA T AAA K FTC F GGG GAC G GGG GAC MAGT F GGC GTT V ACA T I GCT

Fig. 1. Nucleotide sequence and deduced amino acid sequence of the insert present in clone CM19 isolated from a human testis library (Clontech). Initiation and in-frame stop codons present in the 5'-untranslated region and the modified polyadenylation signal AT-TAAA close to the 3'-end are underlined, the predicted cleavage site of signal pertidace (score 6 746 see [21]) is marked (1)

site of signal peptidase (score 6.746, see [21]) is marked ().

quence, the human PH-20 protein contains 474 amino acids. Excluding the signal sequence, the human and the guinea pig PH-20 proteins are about 60% identical in amino acid sequence up to residue 435. However, the structure of the two proteins differs markedly in the segment encompassing the carboxy-terminal 40 residues; moreover, the guinea pig PH-20 protein contains 20 additional amino acids at this end [3].

# 3.2. Expression of human PH-20 cDNA in RK-13 cells using recombinant vaccinia virus

The cDNA encoding the precursor of the human PH-20 protein was ligated into *SmaI* site of the plasmid pgpt-ATA-18 [8] to yield plasmid pATA-HPH. This vector contains, between two segments of the vaccinia virus thymidine kinase gene, the gene for the bacterial xanthine-guanine phosphoribosyl-transferase [12] under the control of the early/intermediate I3 viral pro-



Fig. 2. Hyaluronidase activity of RK13 cells infected with recombinant vaccinia virus (**■**). As control, cells infected with wild-type virus (-□-) were used. Cells were washed as described in section 2.3 and then incubated with hyaluronan for different periods of time. The decrease of OD-600/ca. 10 cells is plotted against the time of incubation.

moter, and a multiple cloning site downstream of the 11K late promoter.

Recombinant vaccinia virus containing the cDNA encoding the human PH-20 protein was prepared by homologous recombination with the plasmid pATA-HPH. Rabbit kidney RK13 cells were infected with recombinant virus and after 24 h, cells were incubated in the presence of hyaluronan to test for hyaluronidase acitivity [10]. The results of such an experiment are shown in Fig. 2. Hyaluronidase activity was detected on cells infected with recombinant vaccinia virus but not on cells infected with wild-type virus. Essentially the same results were obtained in experiments with crude membranes prepared from cells lysed by freezing and thawing (data not shown).

# 3.3. Solubilization of recombinant human PH-20 by PI-PLC

Proteins bound to membranes via a GPI anchor can be selectively solubilized by incubation with PI-PLC. Upon incubation of RK13 cells expressing the human PH-20 protein with this enzyme, all the hyaluronidase activity was rapidly released from the cells into the medium. Surprisingly, the total amount of activity recovered in the medium was up to ten times higher than that originally detected on the cell surface (see Fig. 3). Again, identical, results were obtained when crude membranes of these infected cells were treated with PI-PLC. In control experiments, no detectable hyaluronidase activity was released under these conditions from cells infected with wild-type vaccinia virus (data not shown).

# 4. DISCUSSION

Fertilization is a complex process which involves a specific interaction between sperm and egg cells mediated by cell adhesion proteins and the subsequent action of hydrolytic enzymes. PH-20 isolated from guinea pig sperm by using a monoclonal antibody has been shown to mediate binding of sperm to the zona pellucida of eggs. The results presented in this communication demonstrate that human PH-20 also possesses hyaluronidase activity. Similar results have recently been obtained with monkey and mouse PH-20 cDNA expressed in insect cells (P. Primakoff, personal communication). Based on the homology with the bee venom enzyme we assume that the hyaluronidase represents a domain of the PH-20 protein encompassing 350-400 amino acids including 4 conserved cysteine residues [2]. This domain may correspond to the 41-48 kDa fragment of PH-20 present on guinea pig sperm [3]. The PH-20 polypeptide apparently contains a second domain, the 27 kDa carboxy-terminal fragment. There is no counterpart for this part in the soluble bee venom enzyme; moreover, the human and the guinea pig proteins differ both in length and sequence in this region. It is currently not known which parts of the PH-20 glycoprotein are involved in sperm-egg adhesion. One could assume that the variable carboxy-terminal segment is involved in this species-specific interaction.

Like the guinea pig PH-20 protein, its human homologue is also bound to the membrane by a GPI-anchor. Surprisingly, after treatment of RK13 cells expressing the human PH-20 protein with PI-PLC, a 5-10-fold increase in hyaluronidase activity was observed. The reason for this increase in enzymatic activity is presently unknown. A simple assumption would be that the sub-



Fig. 3. Release of hyaluronidase by treatment with PI-PLC Cells infected with recombinant vaccinia virus were treated with PI-PLC (see section 2.5). Enzymatic activity, i.e. decrease in OD 600/10 cells, is plotted against time. (-▲-): Hyaluronidase activity after treatment with PI-PLC; (-■-): Control without PI-PLC.

strate, a polysaccharide with a molecular mass up to several million [13], may have only limited access to the active site of the membrane-bound hyaluronidase due to steric hindrance. In addition, dense clustering of the PH-20 protein molecules on the cell surface could contribute to this effect. It has been shown that in epithelial cells, GPI-anchored proteins and glycosphingo-lipids tend to form tight complexes that are insoluble in certain detergents [14-16]. A well-studied example is the folate receptor which occurs in dense clusters on the cell surface in association with caveolae [17]. In mature sperm. PH-20 is also clustered in the posterior head region, but it diffuses freely after the acrosome reaction [4.18]. It is currently not known whether this also leads to an increase in hyaluronidase activity of the PH-20 protein.

Many years ago it was shown that mammalian testis is a rich source for soluble hyaluronidases [1]. At present, nothing is known about the structure and biological function of these enzymes. However, hyaluronidases from bovine or ovine testis have, for example, been used for in vitro fertilization to digest the outer layer of cumulus cells which surround the eggs of various mammals. These cells are embedded in a matrix rich in hvaluronan. The present results suggest that upon fertilization in vivo, the hyaluronidase activity of the PH-20 protein enables the sperm to penetrate this outer layer. After this early step, interaction between sperm head proteins, possibly including PH-20, and constituents of the zona pellucida triggers the acrosome reaction [18,19] and fertilization ensues. It has been shown previously that female guinea pigs immunized with the PH-20 protein are sterile [5]. In conjunction with the present results it seems likely that the action of PH-20 as a hyaluronidase is essential for fertilization. This raises the interesting possibility that hyaluronidase inhibitors might also block fertilization and thus function as contraceptives.

Acknowledgements: We thank Dr. G. Seethaler for help and advice with the vaccinia virus expression system and Dr. H.G. Stunnenberg for supplying the pgpt-ATA-18 plasmid used in this study. S.S. was the recipient of a postdoctoral fellowship of the French Ministry for Research and Space.

### REFERENCES

- Meyer, K. (1971) in: The Enzymes, 3rd edn. (Boyer, P.D. ed.) vol. V, pp. 307-320, Academic Press, New York.
- [2] Gmachl M. and Kreil, G. (1993) Proc. Natl. Acad. Sci. USA 90, 3569–3573.
- [3] Lathrop, W.F., Carmichael, E.P., Myles, D.G. and Primakoff, P. (1990) J. Cell Biol. 111, 2939–2949.
- [4] Primakoff, P., Hyatt, H. and Myles, D.G. (1985) J. Cell Biol. 101, 2239-2244.
- [5] Primakoff, P., Lathrop. W., Woolman, L., Cowan, A. and Myles, D.G. (1988) Nature 335, 543-546.
- [6] Chirgwin, J.M., Przybyla, A.B., MacDonald, L.J. and Rutter, W.S. (1979) Biochemistry 18, 5294–5299.
- [7] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
- [8] Stunnenberg, H.G., Lange, H., Philipson, L., van Miltenburg, R.T. and van der Vliet, P.C. (1988) Nucleic Acids Res. 16, 2431– 2444.
- [9] Brent, R. (1991) in: Current Protocols in Molecular Biology (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. eds.) chapter 16.17 and 16.18, Wiley, New York).
- [10] Dorfman, A. (1955) Methods Enzymol. vol. 1, 166-173.
- [11] Kozak, M. (1991) J. Cell Biol. 115, 887-903.
- [12] Mulligan, R.C. and Berg, P. (1981) Proc. Natl. Acad. Sci. USA 78, 2072–2076.
- [13] Laurent, T.C. and Fraser, J.R. (1992) FASEB J. 7, 2397-2404.
- [14] Simons, K. and Wandinger-Ness, A. (1990) Cell 62, 207-210.
- [15] Brown, D.A. and Rose, J.K. (1992) Cell 67, 533-544.
- [16] Fiedler, K., Kobayashi, T., Kurzchalia, T.V. and Simons, K. (1993) Biochemistry 32, 6365–6373.
- [17] Rothberg, K.G., Ying, Y.-S., Kamen, B.A. and Anderson, R.G.W. (1990) J. Cell Biol. 111, 2931–2938.
- [18] Phelps, B.M., Primakoff, P., Koppel, D.E., Low, M.G. and Myles, D.G. (1988) Science 240, 1780–1782.
- [19] Wassarman, P.M. (1990) Development 108, 1-17.
- [20] Miller, D.J., Macek, M.B. and Shur, B.D. (1992) Nature 357, 589–593.
- [21] von Heijne, G. (1986) Nucleic Acids Res. 14, 4683-4690.

#### NOTE ADDED IN PROOF

After submission of this manuscript, the following paper was published: Lin, L., Kimmel, L.H., Myles, D.G. and Primakoff, P.: 'Molecular cloning of the human and monkey sperm surface protein PH-20' (1993) Proc. Natl. Acad. Sci. USA 90, 10071-10075.