importance of this conserved intron sequence has been demonstrated genetically; mutations at positions 1, 2, and 5 significantly impair splicing (22) (mutations at position 4 have not been tested). Furthermore, while few mammalian 5' junctions fit their consensus perfectly (23), Zhuang and Weiner demonstrated that increasing the degree of complementarity from 5 to 6 bp (of 9 bp possible) between U1 and a mammalian 5' splice site resulted in more efficient use of that junction (3).

On the other hand, perfect (nine out of nine) complementarity may be detrimental to optimal splicing efficiency. It is also possible that position 4 of the intron is not needed to form a canonical base pair with snR19, but may be involved in a separate recognition event. In addition, there may be a structural requirement for a U at position 5 (and 10) of snR19, perhaps for tertiary interactions. Construction of mutations in conserved sequences of snR19 should provide a genetic test of their function. Moreover, making the complementary changes in the intron will allow us to assess the specific contribution of Watson-Crick complementarity to 5' splice junction recognition by snR19.

REFERENCES AND NOTES

- 1. T. Maniatis and R. Reed, Nature (London) 325, 673 (1987).
- 2. M. R. Lerner, J. A. Boyle, S. M. Mount, S. L. Wolin, J. A. Steitz, ibid. 283, 220 (1980); J. Rogers and R. Wall, Proc. Natl. Acad. Sci. U.S.A. 77, 1877 (1980). The original model included a more extensive interaction between U1 and the intron, which also included base pairing to the 3' acceptor; the latter interaction has been ruled out on the basis of
- and interaction has been rated out on the basis of phylogenetic comparison (25).
 Y. Zhuang and A. Weiner, *Cell* 46, 827 (1986).
 N. Riedel, J. A. Wise, H. Swerdlow, A. Mak, C. Guthrie, *Proc. Natl. Acad. Sci. U.S.A.* 83, 8097 (1987). (1986).
- 5. J. A. Wise et al., Cell 35, 743 (1983).
- 6. D. Tollervey, J. A. Wise, C. Guthrie, ibid. p. 753; D. Tollervey and C. Guthrie, EMBO J. 4, 3873 (1985);
- R. Parker et al., in preparation.
 M. R. Lerner, J. A. Boyle, S. M. Mount, S. L. Wolin, J. A. Steitz, Nature (London) 283, 220 (1980). 8. N. Riedel, S. Wolin, C. Guthrie, Science 235, 328
- (1987).
- E. M. De Robertis, S. Lienhard, R. F. Parisot, Nature (London) 295, 572 (1982); D. J. Forbes, T. B. Komberg, M. W. Kirschner, J. Cell Biol. 97, 62 (1983).
- C. Branlant *et al.*, *EMBO J.* 1, 1259 (1982); J.-P. Liautard, J. Sri-Wadada, C. Brunel, P. Jeanteur, J. Mol. Biol. 162, 623 (1982); I. Mattaj and E. De Robertis, Cell 40, 111 (1985)
- 11. D. Tollervey and I. Mattaj, EMBO J. 6, 469 (1987).
- 12. B. Patterson and C. Guthrie, Cell 49, 613 (1987). The single copy gene for snR7 encodes two forms of the RNA, designated L (214 nt) and S (179 nt), which differ at their 3' ends.
- 13. P. Siliciano, D. Brow, H. Roiha, C. Guthrie, ibid.,
- in press. 14. C. W. Pikielny, B. C. Rymond, M. Rosbash, Nature (London) 324, 341 (1986); N. Riedel et al., in preparation.
- 15. M. Ares, Jr., Cell 47, 49 (1986).
- 16. D. Tollervey and I. Mattaj (11) have reported anti-Sm precipitation of snR7 and snR14. Precipitation of other, higher molecular weight RNAs was not observed.
- **18 SEPTEMBER 1987**

- 17. C. Branlant, A. Krol, J.-P. Ebel, Nucleic Acids Res. 9, 841 (1981)
- 18. A. Weiner, personal communication.
- 19. J. A. Wise, personal communication.
- 20. B. Rymond, D. Torrey, M. Rosbash, Genes Dev. 1, 238 (1987); N. Riedel et al., in preparation.
 21. J. L. Teem et al., Nucleic Acids Res. 12, 8295 (1984).
 22. R. Parker and C. Guthrie, Cell 41, 107 (1985); A.
- Jacquier, J. R. Rodriguez, M. Rosbash, ibid. 43, 423 (1985); L. A. Fouser and J. D. Friesen, ibid. 45, 81 (1986); U. Vijayraghavan et al., EMBO J. 5, 1683 (1986).
- S. M. Mount, Nucleic Acids Res. 10, 459 (1982). 22 24. S. L. Wolin and J. A. Steitz, Proc. Natl. Acad. Sci.
- U.S.A. 81, 1996 (1984). 25. R. J. Lin, A. J. Newman, S. Cheng, J. Abelson, J.
- Biol. Chem. 260, 14780 (1985) 26. T. E. England and O. C. Uhlenbeck, Nature (Lon-
- don) 275, 560 (1978). S.-C. Cheng and J. Abelson, Proc. Natl. Acad. Sci. U.S.A. 83, 2387 (1986). 27.

- 28. D. McPheeters, A. Christensen, E. Young, G. Stormo, L. Gold, Nucleic Acids Res. 14, 5813 (1986).
- 29. S. M. Mount and J. A. Steitz, Nucleic Acids Res. 9,
- 6351 (1981). 30. The SNR19 gene has also been cloned and sequenced by L. Kretzner, B. Rymond and M. Ros-
- bash (*Cell*, in press).
 31. We thank J. Steitz for giving us antibodies; A. Weiner and J. A. Wise for communication of results prior to publication; J. Tamm for help with the RNA sequencing; N. Riedel and C. Desplan for discussion and advice; and T. Simmons, B. Patterson, and K. Yamamoto for comments on the manuscript. Supported by NIH grant GM21119 and NSF grant DCB8603926 (C.G.), by a postdoctoral fellowship from the Leukemia Society (M.H.J.), and by a Damon Runyon–Walter Winchell Cancer Fund Fellowship, DRG-872 (P.G.S.)

8 July 1987; accepted 14 August 1987

A Sea Urchin Gene Encodes a Polypeptide Homologous to Epidermal Growth Factor

DEBORAH A. HURSH, MARY E. ANDREWS, RUDOLF A. RAFF

A sea urchin DNA clone complementary to an embryonic messenger RNA whose protein product bears striking homology to the epidermal growth factor family of proteins has been identified and characterized. The structure of the protein is similar to that of previously identified regulatory genes in Drosophila and Caenorhabditis. RNA gel blot hybridization showed a unique temporal pattern of expression of this gene during embryogenesis and transcript enrichment in the embryonic ectoderm. These results suggest that this member of the epidermal growth factor gene family plays a role in early development decisions in sea urchin embryos.

HE EPIDERMAL GROWTH FACTOR (EGF) family represents a rapidly expanding collection of proteins that have a variety of cellular functions, including differentiation, cell proliferation, and neoplastic transformation (1). Recently, the products of two developmental switch genes, the notch gene of Drosophila (2, 3) and the lin 12 gene of Caenorhabditis elegans (4), were shown to be members of this gene family, heretofore thought to be exclusively mammalian. Our isolation of a complementary DNA (cDNA) clone encoding a protein resembling the products of these two genes suggests that the EGF peptide domain is widely used in metazoan development.

We isolated a 1.5-kb cDNA clone, uEGF-1, during a screen for cell lineage-specific genes in the sea urchin Strongylocentrotus purpuratus. This cDNA clone recognizes transcripts of 3.0 kb and 4.0 kb on RNA gel blots and is preferentially expressed in embryonic ectoderm, as determined by cell fractionation techniques (5, 6) (Fig. 1).

The sequence of this cDNA revealed a long open reading frame (ORF) of 1447

nucleotides ending with three in-frame stop codons (Fig. 2). Codon bias for this reading frame agrees with that of other sea urchin ORFs (7). A search of the Protein Information Resource (PIR) database by means of



Fig. 1. RNA gel blot analysis of embryonic tissues. (a) Mesenchyme blastula embryos (27 hours) were fractionated into mesenchyme (M) and ectoderm (E) (5). (b) Early pluteus-stage embryos (68 hours) were fractioned into a mesenchyme-endoderm fraction (N) and ectoderm (E) (6). Total RNA was extracted from each fraction, electrophoresed in formaldehyde gels, and transferred to nitrocellulose. Ten micrograms of RNA was loaded from each fraction. Gels were stained with ethidium to monitor concentration. RNA extractions, blots, and hybridizations were as described (26).

Institute for Molecular and Cellular Biology, Department of Biology, Indiana University, Bloomington, Indiana 47405.

the algorithm of Lipman and Pearson (8)indicated strong homology to EGF and other EGF-related proteins. The homology is restricted to an approximately 38-amino acid cysteine-rich repeat common to these proteins. The characteristic feature of this repeat consists of the distribution of six cysteine residues in the pattern CX4CX5CX8CXCX8CX6 (2). One such repeat comprises most of the mature EGF polypeptide and is repeated nine more times in the EGF precursor (9)

Other members of the EGF family contain variable numbers of this repeat. Vaccinia virus protein 19 (VVGF) (10), transforming growth factor- α (TGF- α) (11), and tissue-type plasminogen activator (tPA) (12) each contain one copy, while the mammalian clotting factors IX and X have two (13, 14), the low-density lipoprotein receptor (LDLR) has three (15), and the *notch* gene product has 36 (2).

Figure 2 shows the nucleotide sequence of uEGF-1 with its predicted protein sequence. The 1100 nucleotides nearest the 5' end encode nine 38-amino acid repeats of the EGF-like motif. In addition we have identified one potential site of N-linked glycosylation (Asn-Tyr-Ser) at nucleotide position 803. There are no obvious sites of proteolytic cleavage surrounding any of the repeats, unlike the situation in EGF (9) or TGF- α (11). Within uEGF-1 the repeat structure is highly conserved, the spacing of the cysteine residues being identical in all nine sequenced uEGF-1 repeats. The identities of most of the other repeat residues are also highly conserved. Figure 3 shows the alignment of the nine repeats of uEGF-1. It appears that only three positions in the 38amino acid repeat readily tolerate amino acid substitutions. Overall, very little variation exists among the cysteine repeats of the uEGF-1 protein.

A comparison of the cysteine repeat of uEGF-1 with those of other members of the EGF family is shown in Fig. 4. The register of the cysteines in the uEGF-1 repeat is most homologous to notch and coagulation factors IX and X, although amino acid similarities between uEGF-1 and all the listed members can be seen. The similarity to coagulation factors extends to the positioning of the repeats with respect to one another. Two recently reported sequences of the blood coagulation factors thrombomodulin (16) and protein S (17) contain short tandem repeats of the EGF-like domain. Uninterrupted tandem repetition is not seen in any other members of the EGF family from vertebrates, and resembles the structure of the cysteine repeats found in notch, lin 12, and uEGF. The members of the EGF family found in invertebrates form a distinct sub-

Fig. 2. Nucleotide sequence and predicted amino acid sequence of the uEGF-1 cDNA. Nucleotides are numbered above the sequence. A single asterisk denotes each ten amino acids. The termination codons are indicated by a hyphen. The nine repeated units of the EGF-like domain are underlined. A potential N-linked glycosylation site is indicated by the dashed box. The region of 110 residues rich in serines and threonines to which Olinked sugars could be attached is indicated with boxes surrounding those residues. The single letter amino acid codes used are: A, alanine; R, arginine; N, asparagine; D, aspartic acid; C, cysteine; Q, glutamine; E, glutamic acid; G, glycine; H, histidine; Ĭ, isoleucine; K, lysine; L, leucine; M, methionine; F, phenylalanine; P, proline; S, serine; T, threonine; W, tryptophan; Y, tyrosine; and V, valine. The uEGF-1 cDNA clone was isolated from a λgt11 mesenchyme blastula cDNA library by standard protocols. The probe was a cDNA obtained by a differential screen designed to isolate transcripts restricted to the embryonic lineages of the sea urchin (31). The uEGF-1 cDNA clone was subcloned into M13 (32) for sequencing by the dideoxy chain termination method (33, 34) as modified by Biggin et al. (35). Sequences were analyzed by use of the Pustell programs (International Biotechnologies, New Haven, Connecticut).

		10			1	20			30			40			5	5Ø		
TCC S	œc	cc G	TTC F	ACT T	ooc G	GTA V	CGA R	TGT C	GAA E	GTC V	AAC N	ATT	GAT D	GAA E	TGT C	GCA A	AGT S	acc A
- 6Ø			- 7Ø			. 8	30	-		90			100	1]	.1ø	
хт	TGT	CAG	* AAT	GGT	GGT	ATC	* TGT	ATT	GAT	* œt	ATC	аат	GGA	TAC	ACC	TGC	* TCA	TGT
P	<u> </u>	Q	N	G	<u>G</u>	1	<u> </u>	1	D	G	I	N	G	Y 160	<u>T</u>	C		<u> </u>
~~	*	~~~~		136	,	C 3 m		*	~		150 *	CAT	CAT	100	' mr	mr	MCC 1	* *
P	L	G	F	S	G	D	N	c	E	N	N	D	D	E	<u>c</u>	S	s	1
		18Ø *			190	J		2	200			21Ø *			220	l.		
P	TGT C	TTA L	AAT N	GGT G	GGA G	ACC T	TGT C	GTG V	GAT D	CTT L	GTT V	AAC N	acc A	TAC Y	ATG M	TGT C	GTC	TGT C
3Ø			240			250	9		2	260			270			280	1	
ŝ	œ	œ	TGG	ACC	œc	oct.	ACC	τœ	GCT	GAC	AAC	ATT	GAC	GAG	тст	GCT	AGT	ccc
2	90	6	~	300	<u> </u>	P	<u>-</u> 316	, <u> </u>	-		120	-		330	<u> </u>		340	
т хт	* TGC	CAG	AAC	* GGA	GGT	GTG	* TGC	ATT	GAC	GGT	* GTG	AAC	GGA	* TAC	ATG	TGT	GAC *	TGT
P	с	Q	N	G	G	V	с	I	D	G	<u>v</u>	N	G	Y	м	С	D	c
	35	5Ø		3	86Ø *			370			38	Ø		3	19Ø			400
Q	P	GGA G	TAC Y	ACC T	G G	ACC T	CAT H	TGC C	GAA	ACT	GAT D	ATC	GAC D	GAG	C	GCA A	AGG R	P
		4	10			420			430	;		4	40			450		
хт Р	тос С		AAT N	GGA G	GCT G	GAC D	TGT C	GTG V	GAT D	GGA G	GTC V	AAC N	GGA G	TAC Y	GTC V	TGC C	ATC I	TGC C
460	1		4	70			480			490	,		9	øø			510	
* 3CT	сст	QGA	TTC	* GAC	QGA	crc	* AAC	TOC	GAG	* AAC	аат	ATT	GAC	* GAA	TGC	œc	* AGC	CGT
<u>A</u>	P	G	F	D	<u> </u>	L	N	<u> </u>	E	N	N		D	<u> </u>	C	<u>A</u>	S	<u>R</u>
~~~	526	, ,	220		*	~~~	<b>m</b> 77	*	CAT	~	308 *	, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	m		*	m	NCC	* *
P	c	Q	N	G	A	v	C	v	D	G	v	N	G	F	V	c	T	c
		580 *	3		ţ	59Ø *			600 *			616	,		6	2Ø		
TCT S	GCT A	GCC	TAC Y	ACA T	GGA G	GTC V		TGT C	GAA E	ACC T	GAT D	ATT I	GAC D	GAG E	TGT C	GCC A	AGT S	GCC
630			640	ð			650			660			670	3			68Ø	
сст Р	TCC C	CAG	AAT N	GGT G	GGT G	GTG V	TGT C	ACT T	GAT D	ACC T	ATT I	AAC N	GGA G	TAT Y	ATC I	TGT C	$\stackrel{\infty}{_{A}}$	TGT C
	690			70	ø			710			720			730	3			740
GTG	œa	GGA	TTC	* ACC	GGA	AGC	AAC	* 100	GAG	ACT	* AAC	ATC	GAC	¢ GAG	TGT	œт	AGC	GAC
GTG V	œA P	GGA G	TTC F	ACC T	GCA G	AQC S	AAC N	700 C	GAG E	ACT T	AAC N	ATC	GAC D	GAG E	TGT C	OCT A	AGC S	GAC D
GTG	œ́A P	GGA G 750	TTC F	ACC T	0GA G 760	AGC S	AAC N		GAG E 770 *	ACT T	AAC N	ATC I 78Ø	GAC	CAG E	TGT C 790	GCT A	AGC S	GAC D
GTG V CCC P	* P TGT C	GGA G 75Ø * CTA L	TTC F AAT N	ACC T GGA G	GGA G 76 GGT G	AGC S Ø ATC I	AAC N TGT C	toc c GTG V	GAG E 770 GAT D	ACT T GGA G	AAC N GTC V	ATC I 78Ø * AAT N	GAC D GGT G	GAG E TTC F	TGT C 790 * GTC V	GCT A J TGC C	AGC S CAG	GAC D TGC C
GTG V CCC P ØØ	¢ P TGT C	GGA G 75Ø * CTA L	TTC F AAT N 810	ACC T GGA G	GGA G 76 GGT G	AGC S ATC I 820	AAC N TGT C		GAG E 770 GAT D	ACT T GGA G	AAC N GTC V	ATC I 78Ø * AAT N	GAC D GGT G 840	GAG E TTC F	TGT C 790 GTC V	GCT A TGC C 850	AGC S CAG	* GAC D TGC C
GTG V CC ØØ *	TGT C C C C	GGA G 750 CTA L AAC	TTC F AAT N 810 TAC Y	ACC T GGA G	GGA G 760 * GGT G GGA G	AGC S ATC I 820 * ACT T	AAC N TGT C J TAT	* TC C GTC > GTC > TC C	GAG E 770 GAT D GAA E	ACT T GGA G B3Ø * ATC I	* AAC N GTC V TCA	ATC I 780 AAT N CTT L	GAC D GGT G 84Ø GAT D	GAG E TTC F GCA A	TGT C 790 * GTC V TGC C	GCT A TGC C 850 AGG R	AGC S CAG Q AGT S	* GAC D TGC C ATG
GTG V QCC P ØØ * CCT P	TGT C C C C C C C C C C C C C C C C C C	GGA G 750 CTA L AAC	TTC F AAT N 810 TAC	ACC T GGA G TCT S 879	GGA 761 * GGT G GGA G	AGC S ATC I 820 * ACT T	AAC N TGT C TAT Y 88		GAG E 770 * GAT D GAA E	ACT T GGA G B3Ø ATC I	* AAC N GTC V TCA S B90	ATC I 780 AAT N CTT L	GAC D GGT G 840 * GAT D	GAG E TTC F GCA A 900	TGT C 790 * GTC V TGC C	GCT A TGC C 850 AGG R	AGC S CAG Q AGT S 910	ATG M
	TGT C P 668 TC	GGA G 750 CTA L AAC	AAT N 810 TAC Y AAT	ACC T GGA G TCT S 878 GC	GCA G G G G G G G G G G G G G G G G G G	AGC S ATC I 820 * ACT T ACG	AAC N TGT C TAT Y 88 TOC		GAG E 770 GAT GAA E GAA	ACT T GGA G 330 * ATC I	* AAC N GTC V TCA S B90 * GGA	ATC I 780 AAT N CTT L	GAC D GGT G S40 GAT D GAC	GAG E TTC F GCA A 900 TAC	TGT 790 790 GTC V TGC GTC GTC	ACT A TGC C 850 AGG R TGC	AGC S CAG Q AGT S 910 GAA	ATGC ATG ATG
	* CCA P TGT C * TGT C * TGT C * TGC	GGA G 750 * CTA L N CAG Q	TTC F AAT N 810 * TAC Y AAT N	* ACC T GGA G S 870 870 870 870	GGA G GGT G GGA G GCC A 930	AGC S ATC I 822( * ACT T ACG T	AAC N TGT C 3 TAT Y 88 * TGC C		GAG E 770 GAT D GAA E AAC N 3	ACT T GGA G 330 * ATC I S T V	* AAC N GTC V TCA S B9Ø * GGA G	ATC I 780 * AAT N CTT L GOC A	GAC D GGT G 840 * GAT D GAC D	* GAG E TTC F GCA A 9000 * TAC Y	TGT 790 * GTC V TGC C C C V 960	GCT A TGC C 85( * AGG R TGC C	AGC S CAG Q AGT S 910 * GAA E	* GAC D TGC C ATG M TGC C 970
	* CA P TGT C TGT C C P 360 * TCC C C C C	GGA G 750 * CTA L AAC Q CAG Q 20 *	TTC F AAT N 810 * TAC Y TAT	* ACC T GGA G TCT S S 870 * CC G CCT	0GA G 761 * GGT G GGA G GGA G GGA G GGA GGA	AGC S ATC I 822 * ACT T ACG T CAA	AAC N TGT C Z TAT Y 88 * TGC C	* TCC C V TCT V TCT C V S TCT V 94i *	GAG E 770 & GAT D GAA E AAC N 0 GAA	ACT T GGA G 330 * ATC I V ATT	* AAC N GTC V TCA S B9Ø * GAC	ATC I 780 * AAT N CTT L GCCC A 950 * ATC	GAC D GGT G 840 * GAT D GAC D	* GAG E TTC F GCA A 9000 * TAC Y GAG	TGT 790 GTC V TGC C GTC V 960 * TGT	GCT A TGC C 850 * AGG R TGC C	AGC S CAG Q AGT S 910 * GAA E	* GAC D TGC C ATG M TGC C 970 * CTT
	TGT C C C C C C C C C C C C C C C C C C	GGA G 750 CTA L N CTA CTA CTA CTA CTA CTA CTA CTA CTA CTA	TTC F AAT N 810 * TAC Y AAT N TAT Y	* ACC T GGA G TCT 55 878 * GC G G CT A	GGA G G G G G G G G G G G G G G G G G G	AGC S ATC I 822 * ACT T ACG T CAA Q	AAC N TGT C Z X X X X X X X X X X X X X X X X X X	* TCC GTC V TGT C Ø GTA V 94 * TGT C	GAG E 770 CAT D GAA E AAC N GAA E GAA E	ACT T GGA G 33Ø * ATC I STT V ATT I	* AAC N GTC V TCA S 89Ø * GGA GGA D	ATC I 780 * AAT N CTT L GCC A OSO * ATC I	GAC D GGT G 840 * GAT D GAC D AAC N	* GAG E TTC F GCA A 9000 * TAC Y GAG E	TGT 790 * GTC V TGC C GTC V 960 * TGT C	GCT A TGC C 850 * A AGG R TGC C C GCT A	AGC S CAG Q AGT S GAA E AGT S	* GAC D TGC C ATG M TGC C D TGC C D TGC C C TGC C L
	TGT P TGT C P 368 TCC C C C C C C C C C C C C C C C C C	GGA G 750 * CTA L N CAG Q 20 * GGA G	TTC F AAT N 810 * TAC Y AAT N TAT Y 980 *	* ACC T GGA G 375 876 876 6 C C C C C	GGA G 766 * GGT G G G G G G G G G G G G G G G G G	AGC S 3 ATC I 820 * ACT T ACG T CAA Q 990	AAC N TGT C 8 8 10 C C AAC N	* TCC GTC V TGT C g GTA V 94* TGT C	GAG E 770 * GAT D GAA E AAC N GAA E 1000	ACT T GGA G 330 ATC I V ATT I 3	* AAC N GTC V TCA S 890 * GAC D	ATC I 780 * AAT N CTT L GCC A 950 * ATC I I I I I I I I I I I I I	GAC D GGT G 840 * GAT D GAC D AAC N 810 *	* GAG E TTC F GCA A 9000 * TAC Y GAG E	TGT C 790 * GTC V TGC C STC C 960 * TGT C	GCT A TGC C 85% * AGG R TGC C C 0CT A	AGC S CAG Q AGT S 910 CAA E AGT S	* GAC D TIGC C ATG M 7 TIGC C 970 970 970 970
GTG V CCP P 00 00 CCA P CCA P CCA P CCA P	* CCA P TGT C C C C C C C C C C C C C	GGA G 750 * CTA L AAC Q 20 * CAG G CAG G CAA Q	TTC F AAT N 810 * TAC Y AAT N TAT Y 980 * AAC N	* ACC T GGA G T T T T T T S S 878 878 878 C G C T A G G C T C T S S S 878 C G C C C C C C C C C C C C C C C C C	GGA G GGT G GGA G G G G G G G G G G G G	AGC S ATC I 822(* * ACT T ACG T CAA Q 990 * CTA L	AAC N TGT C 3 TAT Y 88 8 * TGC C AAC N TGT C	* TCC GTG V GTG V TGT C Ø GTA V 94' * TCT C ATT I	GAG E 770 GAT D GAA E AAC N GAA E 1000 * GAA E	ACT T GGA G 3330 * ATC I S GTT V ATT I 3 GGT G	* AAC N GTC V TCA S B9Ø * GAC D ATT I	ATC I 780 AAT N CTT L GOC A OS 0 S 0 S 0 CTT I I GCT A	GAC D GGT G GAC D GAC D AAAC N 310 * GGA G	* GAG E TTC F GCA A 9000 * TAC Y GAG E TAC Y	TGT 790 * GTC V TGC C GTC V 960 * TGT C AQC T	GCT A TGC C 850 * AGG R TGC C GCT A 1020 * TGT C	AGC S CAG Q AGT S GAA E AGT S CAG Q	* GAC D TIGC C ATIG M 970 * TIGC 970 * TIGC C TIGC C
GTG = CCC = P $000 \times CCT = P$ $000 \times CCT = P$ 000	* CCA P TGT C C P 368 * TCC C C C TCC C C	GGA G 750 CTA L AAC CAG Q 20 * CAG G G CAA G	TTC F AAT N 810 TAT Y AAT N TAT Y 980 AAC N	* ACC T GGA G S S S S S S S S S S S S S S S S S	GGA G GGT G G G G A 930 * G G G G G G G G G G G G G G G G G G	AGC S Ø ATC I S2( * ACT T ACG T CAA Q 9900 * CTA L	AAC N TGT C Ø TAT Y 88 * * C C AAC N TGT C 1050	* TGC C GTG V TGT C GTA V 94. TGT C ATT I	GAG E 778 * GAT D GAA E AAC N 8 6 GAA E 1000 * GAA C AAC N C AAC N C AAC N C AAC N C AAC N C AAC C AAC N C AAC N C AAC C AAC N C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC C AAC AAC C AAC AAC C AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AAC AA	ACT T GGA G 330 * ATC I S GTT V ATT I S GTT G G 10661 *	* AAC N GTC V TCA S B9Ø * GAC D ATT I J	ATC I 780 * AAT L CTT L GOC A ATC I 10 GCT A	GAC D GCT G GAC D GAC D GAC D GAC D GAC D GAC C 10 GAC C 10 GAC C T G G G G C T G C T G C T G C T G C T G C T G C T G C T G C T G C T G C T G C C T G C C T G C C C C	* GAG E TTC F GCA A 9000 * TAC Y GAG E TAC Y 770	TGT 790 * GTC V TGC C GTC V 960 * TGT C T ACC T	GCT A TGC C 85% AGG R TGC C C GCT A LØ2Ø * TGT C	AGC S CAG Q AGT S 910 * GAA E CAG Q LØ8Ø	* GAC D TGC C ATG M 9770 * TGC C TGC C TGC C
$\begin{array}{c} \operatorname{GTG} & V \\ & CCP \\ & gg \\ & gg \\ & scal \\ & CP \\ & gg \\ & scal \\ & CP \\ & gg \\ & GTA \\ & V \\ & CP \\ & GTA \\ & V \\ & CP \\ & GTA \\ & CP \\ & CCA \\ & P \\ & CCA \\ & P \\ & CCA \\ & CC$	* A P TGT C C P B60 * TGC C C C C C C C C C C C C C C C C C C C	GGA G 750 * CTA L AAC O CAG G CAG G CAA G CAA G G G G G G G G G	TTC F AAT N 310 TAC Y AAT N TAT Y 380 * AAC N 10 TAC Y	* ACC T G G G T T T T T T S S 870 8 70 870 870 870 870 870 870 870 87	GGA G G G G G G G G G G G G G G G G G G	AGC S ATC I ACC T T ACG T CAA Q 996 * CTA L GTC V	AAC N N TGT C 3 TAT Y 88% * TGC C AAC N TGT C 1050 * AAC N	* TCC GTC V GTC V GTC V GTA V 94 * TCC ATT I TCC C	GAG E 770 K GAT D GAA E GAA E 1000 K GAG CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAC C CAC C CAC C C C C C C C C C C C C C C C C C C C	ACT T GGA G 330 * ATC I V ATT I G G G G G G G G G G G G G G G G G G	* AAC N TCA S 890 * GAC D ATT I 3 GTT V	ATC I 780 * AAT N CTT L GCC A A CTT I GCC A I I GCC A CT I GCC A GCT G GCT G	GAC D GGT G G G G G G G G G C D G A A C D G A C D G C C G C G C C C G C C C C C C C C	* GAG E TTC F GCA A 9000 * TAC Y GAG E TAC Y TAC Y TAC Y TAC C	TGT 799 *C V TGC C TGC V 960 * TGT C ACC T GAC D	GCT A TGC C A AGG R TGC C GCT A IØ2Ø * TGC C TTG C	AGC S CAG Q AGT S GAA E CAG Q LØ80 K GAG E	* GAC D TGC C ATG C S TGC C 978 * C C TGC C 978 *
$GTG = \frac{V}{CCP}$ $gga = \frac{V}{CCA}$ $gga = \frac{V}{$	* A P TGT C C P B60 * TCC C C C C C C C C C C C C C C C C C C	GGA G 750 * CTA L AAC CAG Q 20 * CAG G G G G G G G G G	TIC F AAT N 810 * TAC Y 380 * AAC N 16 TAC Y	* ACC T G G S S S S S S S S S S S S S S S S S	GGA G G G G G G G G G G G G G G G G G G	AGC S 3 ATC I 82(* * ACT T ACG T CAA Q 996 * CTA L GTC V	AAC N TGT C 8 TAT Y 88* TGC C AAC N TGT C 1050 *	* TCC GTC V TGT C 0 GTA V 944 * TCT C ATT I TCC C 1110	GAG E 770 * GAT D GAA E GAA E 1000 * CAA E GAA E GAA E GAA E GAA E	ACT T GGA G 3330 * ATC I S GTT V V ATT I G G G G G G G G G G G G G G G G G G	* AAC N GTC V TCA S B9Ø * GAC D GAC D GAC D S GTT V 1120	ATC I 780 * AAT N CTT L GCC A CTT I GCC A I I GCT A GGT G	GAC D GGT G 840 8 4 C AAC D AAC D AAC N S 10 5 C GGA D C D C C C C C C C C C C C C C C C C	TTCC F GCA A 9000 * TAC Y GAG E TAC Y TAC Y TAC Y 11	TGT C 79% GTC V TGC C TGC C V 960 * TGT C T T GAC D 130	GCT A 3 TGC C 856 8 AGG R TGC C C A AGG R TGC C C TTGC C TTGC C TTGC C TTGC C	AGC S CAG Q AGT S S S CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG	* GAC D TGC C ATG M M TGC C GGT G GGT G
GTG V CCC P ØØ CCT P GTA V GTA V CCT P GTA V CCT R GTA V ATG	* A P TGT C C P 668 * TGC C C C C C C C C C C C C C C C C C C C	GGA G 750 CTA L CTA L CTA CTA CTA CTA CTA CTA CTA CTA CTA CTA	AAT N 810 * TAT N TAT Y 980 * AAC Y AAC	* ACC T T GGA G TTTT S S S S S S S S S S S S S S S S S	GGA G G G G G G G G G G G G G G G G G G	AGC S ATC I S2( * ACG T ACG T ACG T CAA Q 996 * CTA L GTC V V AAT	AAC N TGT C 3 TAT Y 98% TGC C AAC N TGT C 1050 AAC N GAT	* TCC GTC V TGT C g GTA V 944 * TCC AFT I TCC 1110 CAG	GAG E 770 GAA E GAA E AAC N 0 GAA E GAA E GAA E GAA E GAA E GAA C A GAA E GAA C C C C C C C C C C C C C C C C C C	ACT T GGA G 330 * ATT I G GTT V ATT I G GTT G I 1060 * GAA E AQC	* AAC N GTC V TCA S 890 GAC D GAC D GAC D S GAC D 1120 C T U 1120 C T C ATC	ATC I 784 AAT N CTT L CTT L CCT A 556 * ATC I 10 CCT A CCT A CCT A CCT A CCT A CCT A CCT A CCT A ATC A CCT A CCT A ATC A ATC A ATC A CCT A ATC AT	GAC D GGT G G GAC D GAC D GAC D AAC N SIØ * GGA G G TTC F AAG	AGG E TTC F GCA A 9080 * TAC Y GAG E TAC Y TAC Y TAC Y TAC Y TAC Y II ACC	TGT C 796 * GTC V 76C C 700 * TGC C 960 * TGC C 700 * TGC C 700 * TGC C 700 *	A TOC 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 A C A C A C	AGC S CAG Q AGT S S GAA E CAG Q LØ80 CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG	* GAC D TOC C ATG C 978 * TOC C 978 C TTC C 978 C 1448 ATG
GTG = V CCP = 00 gg = 00 gg = 00 GTA = P GTA = P	* CA P TGT C C P B66 * TGC C C C C C C C C C C C C C C C C C C C	GGA G G CTA L CAG Q 220 * CAA G G G G G G G G G G G G G G G G G G	AAT N 810 * Y AAT N TAT Y 980 * AAC N IG TAC Y	* ACC T GGA G G T T T T T T T T T T T T T T T	GGA G GGT G G G G G G G G G G G G G G G	AGC S S ATC I S2( * * ACG T ACG T CAA Q 996 * CTA L GTC V AAT N	AAC N TGT C TAT Y 88* TGC C AAC N TGT C 10500 AAC N GAT D	* TCC GTG V GTA V 94 * TGC C 1110 CAG Q	GAG E 770 GAA E GAA E GAA E I000 GAA E GAA C A C A A C N GAA E GAA C N GAA C N GAA C N GAA C N GAA C N C S C C N C S C C C C C C C C C C C	ACT T GGA G G 330 * ATC I G G G G G G G G G G G G G G G G G G	AAC N TCA S B9Ø GAC D GAC D GAC D GAC D U 1120 X ATT I S ATT I I S	ATC I 780 AATC N CTT L CTT L CCT A ATC A SOUTH AND	GAC D GGT G GAC D GAC D AAC D AAC F AAG K	GAG E TTC F GCA A 9000 * TAC Y GAG E TAC Y TAC Y TAC Y TAC Y TAC Y TAC Y TAC C II I	TGT C 796 GTC V TGC C GTC V 968 TGT C TGT C GAC D 30 *	A A A A A C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C A A C C A A A C C A A A C C A A A C C A A A C C A A A C C A A A A C A A A A A A A A A A A A A	AGC S CAG Q AGT S 910 * GAA E GAA E CAG Q GAA G GGA G GGA G	* GD TGC C ATGC C TGC C 978 * CTT L TGC C GGT G GGT G GGT 448 *
GTG V GTG P ØØ GTA P GTA P	* CA P TGT C P 368 * TGC C C C C C C C C C C C C C	GGA G G CTA L CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG C CAG G CAG C CAG C CAG C CAG C CAG C CAG C C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C CAC C C CAC C C CAC C C CAC C C CAC C C CAC C C CAC C C CAC C C CAC C C C C C C C C C C C C C C C C C C C	TTC F AAT N 816 * TAC Y AAT N TAT Y 980 * AAC N 10 TAC Y AAC	* ACCC T GGA G G G G G G G G G G G G G G G G G	GGA G GGT G G G G G G G G G G G G G G G	ACC S ACC I B2C ACT T ACG T ACG T CAA Q 996 * CTA L GTC V V AAT N	AAC N TGT C 8 TAT Y 88* TGC C AAC N TGT C 1050 * AAC N TGT C TGT C T	* TCC GTG V TGT C G GTA V 944 * TCC G ATT I TCC C ATT C ATT C ATT C ATT C C ATT C C C C C C C C C C C C C	GAG E 770 GAT D GAA E GAA E GAA E GAA E GAA E GAA E GAA E GAA E GAA C C C C C C C C C C C C C C C C C C	ACT T GGA G G G G G G G G G G G G G G G G	AAC N TCA S B9Ø GGA C D GGA C D S GTT V 1122 ATC I	ATC I 780 * AAT N CTT L CCTT L CCTT L CCTT A ACC I 10 CCT A ACC I 118 CCT A	GAC D GGT G S44 AAC D AAC D S10 S10 F AAG K GGA	TTCC GAG E TTCC F GCA A 9900 S TACC C TACC C TACC TTCC TACC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCC TTCCC TTCC TTCCC TTCCC TTCCC TTCCCCC T	TGT C 796 * GTC C 796 * TGC C 966 * TGT C 966 * ACC T GAC D 36 * TCT S 1: 51 1: 51 51 51 51 51 51 51 51 51 51	A A A C C B S S S S S S S C C C C C C C C C C	AGC AG AGT S 914 GAA E CAG Q Q AGT S CAG Q Q Q AGT S CAG Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q	* GAC GAC D TIGC C ATIG G TIGC G G G G G G G G G G G G G G G G G G
GTG = CCP = 0 GTG = P = 0 GTA = 0	* CA P TGT C C P 568 * TCC C C C C C C C C C C C C C C C C C C	GGA G 750 CTA L CAG Q 20 CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG C CAG G CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C C CAG C C CAG C C CAG C C CAG C C CAG C C C C	AAT N 810 * TAC Y 930 * AAT N TAT Y 930 * AAC N GAT D	* ACCC T GGA G S 870 * C G G G G G G G G G G G G G G G G G G	GGA G G G G G G G G G G G G G G G G G G	ACC S ACC I ACC T ACC T ACC T CAA ACC T CAA C C A C C C A C C C C	AAC N N TGT C 8 8 8 * * 7 C C AAC N TGT C 1050 * AAC N TGT C C TAT Y 8 8 * *	* TCC GTCC V TCTC 0 GTA V 94 * TCC 1110 CAG Q AAT N	GAG E T70 GAT D GAA E GAA E GAA E GAA E GAA E GAA E GAA E GAA E GAA E GAA E GAA E GAA E GAA E GAA E GAA E GAA E C GAA E C GAA E C GAA C C C C A C C C C C C C C C C C C	ACT T GGA G G G G G G G G G G G G G G G G	AAC N TCA S B90 GGA G C GGA G C D N TCA S S S S S S S S S S S S S S S S S S S	ATC I 780 * AAT N CTT L CTT L CTT L CTT A CTT A CTT A CTT L CTT A CTT L CTT L CTT L CTT A CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L CTT L C CTT L C CTT L C CTT L C CTT L C CTT L C CTT L C CTT L C CTT L C CTT L C CTT L C C C C	GAC D GGT G GAC D GAC D GAC D GAC D GAC D GAC C C G G G G G G G G G G G G G G G G	TTCC GCA A 9000 * TAC Y TAC TAC Y TAC Y	TGT C 799 * GTC C V 76C C 796 * TGC C 796 * TGT C 7 966 * TGT C 7 7 7 7 7 7 7 7 7 7 7 7 7	GCT A TCC C A ACC R TCC C C C C C C C C C C C C C C C C	AGC S CAG Q AGT S GAA E CAG CAG CAG GAA G GAA G GAA C CAG C Q Q C CAG C C CAG C C C C C C C C C C C C C	TCC ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG
GTG = V GTG = P gga = CCT =	* CA P TGT C P P S68 * TGC C P TGC C C C C C C C C C C C C C	GGA G 750 * CTA L CAG Q 220 * CAG G CAG G CAA G CAG G CAA G CAG G CAA G CAG G CAA G CAA G CAA G CAA G CAA C C TAC C TA C TA	TTC F AAT N 810 TAC Y AAT N TAT Y S80 AAC N IG TAC Y AAC N C AT C Y	ACC T T G G G S S S S S S S S S S S S S S S	GGA G G G G G G G G G G G G G G G G G G	AGC S ATC I 322 * ACG T ACG T CAA Q 996 * CTA L GTC V V AAT N I 66 * I I I	AAC N TGT C 8 TAT Y 88 * TGC C AAC N TGT C C AAC N TGT C C C AAC N TGT C C TAT Y 28 * * TGC C C TAT Y 28 * * * * * * * * * * * * * * * * * *	* TCC GTG V TGT C g GTA V 94 * TGT C 1110 Q AAT N	GAG E GAA C GAA E GAA E GAA E GAA E GAA C C C C C C C C C C C C C C C C C C	ACT T GGA G G G T V ATT I G G G G G G G G G G G G G G G G G G	AAC N GTCC V TCA S 890 GAC D GAC D GAC D S GAC D V 1122 X TCA S GAC D C C A	ATC I 78.0 AAT N CTTT L CTT L CT	GAC G G G G G G G G G G G G G	TTC F GCA A A 980 * TAC Y GAG E TAC Y TAC Y TAC Y TAC Y TAC Y	TGT 799 GTC V TGC C TGC V 968 TGT C ACC T GAC 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 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 A C A C A C A C A A A A A A A A A A A A A	ACA ACA ACA ACA ACA ACA ACA ACA ACA ACA	AGC S CAG Q Q 3 AGT S CAG Q CAG CAG Q LØ80 * CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG	TGC GGT GGT M ATG M TCC 978 CTT L CC GGT GGT M ACC T
$\begin{array}{c} \operatorname{GTG} V \\ GCC \\ P \\ gg \\ \mathsf$	TGT C C P F C C C P F C C C C C C C C C C C	GGA G 750 * CTA L AAC Q 220 CAG G CAG G CAG G CAG G G TAC Y 1156 GGA G C TAC C Y	AAT N Blø TAT Y Blø TAT Y Blø AAT N TAT Y Blø AAT N TAT Y AAT N TAT Y AAT N II C AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N N AAT N AAT N N AAT N N AAT N N AAT N N AAT N AAT N AAT N AAT N AAT N AAT N AAT N AAT N AAT N AAT N AAT N AAT N AAT N AAT N AAT N AAT N AAT N AAT N AAT N AAT N AAT AAT	ACC T T CCA G G S S S S S S S S S S S S S S S S S	GGA G G G G G G G G G G G G G G G G G G	AGC S	AAC N TGT C 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	* TCC GTCV TCC Ø GTA V 944 * TCC III Ø GTA V 944 * TCC III Ø GTA N N AAC N	GAG E CAT D CAT CAT CAT CAT CAT CAT CAT CAT CAT CAT	ACT T GGA G G G T V ATT I I 3 G G T V V ATT I I 3 G G G G G G G G G G G G G G G G G	AAC N TCA S B9Ø C GAC C C C C C C C C C C C C C C C C C	ATC I 78.6 AAT N CTT L CTT L CCT A ACC A ATC I 10 GCT A ACC C CT A ACC C CT A ACC C CT A ACC C CT A ACC C CT C ACC C ACC C ACC C ACC C CT CT C CT CT C CT CT C CT CT C CT CT C CT CT C	GAC D GGT G G G G G G G G G G G G G G G G G	GAG E F GCA A 9000 TAC Y GAG E TAC Y TAC Y TAC T TAC T TAC T TAC T T TAC T T TAC T	TGT 799 * GTCC V 76C C C 7 960 * TGT C 7 960 * TGT C 7 960 * TGT C 7 960 * TGT C 7 960 * TGC C 7 960 * TGC C 7 960 * TGC C 7 960 * TGC C 7 960 * TGC C 7 960 * TGC C 7 960 * TGC C 7 960 * TGC C 7 960 * TGC C 7 7 960 * TGC C 7 7 7 7 7 7 7 7 7 7 7 7 7	ACA ACA ACA ACA ACA ACA ACA ACA	AGC S CAG Q AGT S S S CAG Q AGT S S CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG	* GC C C C C C C C C C C C C C
$\begin{array}{c} \operatorname{GTG} & \operatorname{CCP} & \operatorname{GTG} \\ & \operatorname{CCP} & \operatorname{gr} \\ & \operatorname{gr} \\ & \operatorname{gr} \\ & \operatorname{GTA} \\ & \operatorname{P} \\ & \operatorname{GTA} \\ & \operatorname{P} \\ & \operatorname{GTA} \\ & \operatorname{GTC} \\ & \operatorname{GTC}$	* CA P TGT C C P B60 * TGC C P B60 * TGC C C C C C C C C C C C C C	GGA G 750 CTA L CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G C TAC Y 1115*	TTC F AAT N 810 * CAT Y AAT N TAT Y AAT N CAT D 1210 GAT G G	ACC T CGA G G G G G G G G G G G G G G G G G G	GGA G G G G G G G G G G G G G G G G G G	AGC S Ø ATC I B22(** * ACT T ACG T CAA C O 9900 * CTA L GTC CTA N I I I ACT T ACG S I ACG S ACG ACG S ACG S ACG ACG S ACG ACG S ACG ACG ACG ACG ACG ACG ACG ACG ACG ACG	AAC N TGT C TAT Y 88* TGC C AAC N TGT C 1050 * AAC N TAC Y 220 * AAC N 1	* TCC_ GTC_ Ø GTA V 944* * TCC_ Ø GTA V 944* * TCC_ ATT I IIIØ AAT N 288	GAG E GAT D GAT E GAA E GAA C GAA C GAA C GAT C C C C C C C C C C C C C C C C C C C	ACT T GGA G GTT V ATT I GGT G G G G G G G G G G G G G G G G	AAC N TCA S B9Ø CGA G G C C C C C C C C C C C C C C C C	ATC I 780 AATC I CTT L CCT L CCT L CCC A ACC C ACC L CCC A ACC L CCC A P	GAC D GGT G G S S GAC D C C D S S S GAC C D S S S S S S S S S S S S S S S S S	TTC F GAG F GAG F GAG C C TAC C C II ACC F TAC C II ACC F II ACC F II ACC F II ACC F II ACC F II ACC F II C C A A A S C A A A S S C A A A S S C A A A S S C A A A S S C A A A S S C A A A S S C A A A S S C A A A S S C A A A S S C A A A S S C A A A S S C A A S S C A A S S C A A S S C A S C A S C A S C A S C A S C A S C C A S C C A S C C A S C C A S C C A S C C S S S S	TGT 796 TGC TGC TGC TGC 796 * TGC 796 * TGC 796 * TGC 796 * TGC 7 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * 7 * * * * * * * * * * * * *	ACT ACC ACC ACC ACC ACC ACC ACC ACC ACC	AGC S CAG Q AGT S S AGT S CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG	* GAC D TOCC ATG C G G G G G G G G G G G C C C C C C C
$\begin{array}{c} \operatorname{GTG} & \operatorname{GTG} \\ & \operatorname{GTC} \\ & \operatorname{GTC}$	* CAA P TGT C P 568 * TGC C P 568 * TGC C P TGC C P TGC C P CTA L S W CTT L GTG V W CTT L C S C C R C C C C C C C C C C C C C C C	GGA G 750 CTA L CAG G CAG G CAA C Q CAA G G TAC Y L CAG G G CAA C Q CAA C C CAA C C C CAA C C C C C	TTC F AAT N 810 * AAT TAT Y 980 * AAC N 10 TAC Y AAC N GAT D 1210 G AAT N	ACC T T CCA G G T T T T T T CCT T S G G G G G C T T C C C A G G G C T T T T C C C A G G C C C C C C C C C C C C C C	GGA G G G G G G G G G G G G G G G G G G	ACC S	AAC N TGT C TAT Y 88 * TCC AAC N TGT C 1055 * AAC N TGT Y 228 AAC N 105 * TGT Y 228 AAC N 105 *	* TCC GTV TTC G GTA V 94 * TTC ATT TCC G ATT AN AGN 288 AGE	GAG E GAT D GAT D GAA E GAA E GAA C GAA C C CAA E C GAC E C C C C C C C C C C C C C C C C C C	ACT T GGA G G G ATT I GGT R GGT R GGAC D TGG G TGG G ACT R GAC T TGG G TGG	AAC N N TCA S B90 CGA C D ATT I S GGC A TTC F C A C C A C C A C C C C C C C C C C	ATC I 780 * AAT N CTT L GCA 550 * ATC I 10 GCT A GGT G ACC L CCA P GGT G	GAC D GGT G G G G G G G G G G G G G G G G G	GAG E TTC F GAA A 9000 * TAC Y GAG E TAC Y TAC Y TAC Y TAC Y TAC Y TAC Y TAC Y TAC Y TAC C II A AC E	TGT C 790 * C GTC V 70 * C TGC C 70 * C * C * C * C * C * C * C * C	GCT A TCC C 85% AGG R TCC C 1020 TCC C AGG R TCC C AGG R AGG R TCC C AGG R AGG R AGG R C C AGG R AGG R	AGC AGT S AGT S AGT S AGT S AGT S AGT S CAG Q AGT S CAG CAG CAG CAG CAG CAG CAG CAG	* GAC D TOC ATC M TOC 97* CL TOC GG GG GAC ATC M ACT GC 97* CL TOC GG GG GAC ATC M ACT GC SIG GAC ATC M
$\begin{array}{c} \operatorname{GTG} & \operatorname{GTG} \\ & & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & &$	* CA P TGT C P GGT C C C P GGT G C C C C C C C C C C C C C	GGA G G CTA L CAG Q 228 CAA G CAA G CAA G CAA G CAA G CAA G CAA G CAA G CAA G CAA G CAA G CAA G CAA C CAA C CAA C CAA C C CAA C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C C CAA C C C C C CAA C C C C C C C C C C C C C C C C C C C C	AAT N 810 TAC Y AAT N TAT Y AAC N GAT D 1216 GAT C G AAT N	ACC T T CCA G G CCT A S T CCT S S T CCT C C G G C C C C C C C C C C C C C	GGA GGT GGT GGT GGT GGT GGT C C A GGT C C A GGT C C A GGT C C A GGT C C A GGT C C A GGT C C A GGT C C C A GGT C C C C C C C C C C C C C C C C C C	ACC S a ATC I a 22 ACT T ACG T CAA Q 996 a CT CAA C Q 996 a CT CAA C CAA C CAA C CT C ACG T I ACG T ACG T AC	AAC N TGT C TAT Y 88* TGC AAC N TGT C 1055 AAC N GAT TAC Y 228 AAC N IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	* TCC GTV TTC G GTV 94 * TTC ATT TCC 111 * GQ ATT AX X 28* QE 1	GAG E 778 * GAT D GAA E GAA E GAA E GAA E GAA E GAA E TAT Y Y AGT TAT Y S 340	ACT T GGA G GTT V ATT I GGT G GGT G GGT R GAC D CGT R GAC D T GCT R GAC D T GGT R GAC G G G G G G G G G G G G G G G G G G	AAC N N TCA S B96 GAC D ATT I I 2 GAC D C A TTC A C C C C C C C C C C C C C C C	ATC I 780 * AAT N CTT L CCA 550 * ATC I 10 CCT A 050 * ATC I 11 CTC L CCA 9 CTC L CCA 9 CTC L CCA 9 CTC L 1350	GAC D GAC D GAC D GAC D GAC D GAC C D GAC C C G G G G G G G G G G G G G G G G	CAG E CAG E TTCC F CA A A 9000 TAC Y TAC Y TAC Y TAC Y TAC Y TAC Y TAC Y TAC Y TAC Y TAC Y TAC Y TAC C I I I I I I I I I I I I I I I I I I	TGT C 790 * C GTC V 796 * C C TGT C 796 * C 70 * C 796 * C 70 * C 796 * C 70 * C * C * C * C * C * C * C * C	CCT A TCC C ACA ACA ACA ACA ACA ACA ACA	AGC CAG Q AGT S GAA AGT S GAA AGT CAG CAG CAG CAG CAG CAG CAG CAG	* GAC D TOC C C TOC C C C C C C C C C C C C C C
$\begin{array}{c} \operatorname{GTG} & \bigvee \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTA} & & & \\ & & & \\ \operatorname{GTA} & & & \\ & & & \\ \operatorname{GTA} & & & \\ & & & \\ \operatorname{GTA} & & & \\ & & & \\ \operatorname{GTA} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & $	* $CAA$ P TGT C C P TGC C C C C C C C	GGA G G CTA CTA CTA CTA CTA CTA CTA CTA CTA CTA	AAT N 810 * TAT Y 980 * AAT N TAT Y 980 * AAC N 10 TAC Y AAC N 1210 * GAT G G T C G G T	ACC ACC T ACC T ACC C C C C C C C C C C C C	GGA GGT GGT GGT GGT GGT GGT GGT GGT C C LI ATG C C A M GGT C C LI ATG C C A TGC C C C C C C C C C C C C C C C C C C	ACC S ATC I B2*ACT I ACG T ACG T ACG T ACG T ACG T ACG T I ACG T I ACG T I ACG T I ACG T I ACG T I ACG T I ACG T I ACG T ACG	AAC N TGT C TAT Y B*TGC C AAC N TGT C TGT C	* TCC GTV TTC G GTV 94* TCC ATT TCC B118 CGQ AATN AANN AGE 1 TSS	GAG E T78 CAT D GAA E GAA E GAA E GAA E GAA E GAA C N CAA C N C GAA E C C C C C C C C C C C C C C C C C C	ACT T GGA G G GT V ATT I GGT C GGT C GGA E GGA C C C R C G C C C C C C C C C C C C C C	AAC N N GTC S B30 GAC D ATT I S GATT F 1290 A TTCC A C ACC A ACC A ACC A	ATC I 788 * AAT N CTT L CCT L CCA ATC C ATC	GAC G G G G G G G G G G G G G	TTC GAG E TTC F GAG A 998* TAC Y GAG C TTC F TAC Y TTC F TAC Y TTC TTC TTC TTC TTC TTC TTC	TGT C 79% GTC C TGC C 79% TGC C 70% TGC C 70%	A TOC A A TOC A A A A A A A A A A A A A A A A A A A	AGC S CAG Q AGT S 911 CAG Q AGT S GAA CAG C CAG C Q 0888 CAG C CAG C C CAG C C CAG C C CAG C C CAG C C C C C C C C C C C C C C	ACC ATC ATC ATC ATC ATC ATC ATC
$\begin{array}{c} \operatorname{GTG} \lor \\ \operatorname{GTC} & \operatorname{CP} \\ \operatorname{GTA} & \operatorname{CP} \\ \operatorname{GTC} & \operatorname{GTC} \\ \operatorname{GTC} \\ \operatorname{GTC} & \operatorname{GTC} \\ \operatorname{GTC} & \operatorname{GTC} \\ \operatorname{GTC} \\ \operatorname{GTC} & \operatorname{GTC} \\ $	* CAA P TGT C C P TGC C P TGC C C P TGC C C C C C C C C C C C C C C C C C C C	GGA G G CTA CTA CTA CTA CTA CTA CTA CTA CTA CTA	TTC F N 816 TAC TAC Y AAAT N TAT Y 980 * AAAT N 10 TAC Y AAAT N 11 TAC Y AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S S 80 * AAAT N S S 80 * AAAT N S S 80 * AAAT N S S 80 * AAAT N S S 80 * AAAT N S S 80 * AAAT N S 80 * AAAT N S S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S 80 * AAAT N S AAAT N S AAT N S AAAT N S AAAT N S AAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N S AAAT N A A A A A A A A A A A A A A A A	ACC T T GGA G G T T T T C C G G G C T T T C C G C C C C	GGA GGT GGT GGT GGT GGT GGT GGT GGT CC C I I I I I I I I I I I I I I I I I	AGC S a ATC I ACT ACG I ACT I ACG I ACT I ACG I ACT I ACG I ACT I ACG	AAC N TGT C & TAY B* TGC AAC N TGT C AAC N TGT C TAY TGT C TAY TAY TGT C TAY TGT C TAY TGT C TAY TGT C TAY TAY TGT C TAY TGT C TAY TGT C TAY TGT C TAY TGT C TGT C TAY TGT C TGT C T	* TCC GTV TTC GTAV 94 * TTCC TTI TCC 11 * CO AT N AC N 28 & ACT 1 TOW	GAG E T70 GAT D GAA E GAA E GAA E GAA E GAA E GAA E GAA E C GAA E C GAA E C GAA E C GAA E C GAA C C T C T C T C T C C A T C C C T C C C C	ACT T GGA G G G G G G G G G G G G G	AAC N N TCA S B90 GAC GAC GAC C ATT I I 290 ACC I I 290 ACC I ACC	ATC I 780 * AAT N CTT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT L CCT CCT	GAC D GGT G GAC D GAC D GAC C D GAC G GAC C D GAC G GAC C C C C C C C C C C C C C C C	TTC GAG E TTC F GAG A A 9800 * TAC Y TAC C TTC F 1300 TTC C GTC V	TGT C 79* GTC C 79* TCC C 70* TCC C	A A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C A A C A A C A A C A A C A A C A A A C A A A C A A A C A A A C A A A A C A A A A A C A A A A A A A A A A A A A	AGC G AGT S AGT S AG	ACC ATS ATS ATS ATS ATS ATS ATS ATS
$\begin{array}{c} \operatorname{GTG} & \bigvee \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & & & \\ \operatorname{GTC} & & & \\ & $	* CCA P TGT C C P TGC C P TGC C P TGC C C C C C C C C C C C C C C	GGA G G CTA CTA CTA CTA CTA CTA CTA CTA CTA CTA	AAT N 816 * TAC Y AAT N TAT Y 986 * AAT N TAT Y 986 * AAT N TAT Y 986 * AAT N TAC Y 126C X AAT N N S S S S S S S S S S S S S	ACC ACC T T CGA G G G G G G G G G G G G G	GGA G G G G G G G G G G G G G G G G G G	AGC S a ATC I B 2 CAA Q 99 CTA L GTV AAT I ACS T CAA Q 99 CTA L GTV AAT I ACS T S CAA C 99 CTA L GTV AAT I ACS T S CAA C 139 CTA L CAA C C C C C C C C C C C C C C	AAC N TOT C & TAY 83* TOC AAC N TOT C 105* AAC N GAT TAY 22* AAC N ACT ACT ACT	* 120 GZ TTC & TZ 94 TTC TT 120 11 CO ATZ AZ 28 QE 1 TZ	GAG E 778 GAT D GAA E GAA E GAA C AAC C C AAC C C AAC C C AAC C C AAC C C AAC C C AAC C C AAC C C C AAC C C C AAC C C C C C C C C C C C C C C C C C C C	ACT T GGA G G G ATT I GGT V ATT I GGT V ATT I GGT C C GGT C C C C C C C C C C C C C C	AAC N AAC N AAC N TCA S B* GGTC V TCA S B* GGTC V TCA S B* GGTC V TCA S C C D ATT I C C A TTC F I 122/0 ACC TCA S C C A C C A TTC S C C A C	ATC I 780 * AAT N CTT L CCT L CCT L CCC A 550 * ATC I 10 CCT A 550 * ATC I 10 CCT A ATC L CCC A ATC I 10 CCT A ATC	GAC D GAC D GAC D GAC D GAC D GAC C D GAC C C GAC C C GAC C C GAC C C C G C C C C	TTC GAG E TTC F GCA A A 9808 TAC TAC TAC TAC TAC TTC GAG E TAC TAC TAC TAC C TTC C GAG GAG GAG GAG GAG GAG GAG	TGT C 79* GIC V 79* GIC V 960 × TCC GIC V 700 × TCC GIC V TCC GIC V TCCC GIC V TCCC GIC V TCCC GIC V TCCC GIC V TCCC GIC V TCCC GIC V TCCC GIC V TCCCC GIC V TCCCCCC GIC V TCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	ACT A TOC C ACC A ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC	AGC S CAG Q AGT S S AGT S AGT S AGT S AGT S CAG Q CAG Q CAG C CAG C C CAG C C CAG C C CAG C C CAG C C CAG C C C C C C C C C C C C C C	AGO TSCC ATS ATS ATS ATS ATS ATS ATS ATS
$\begin{array}{c} \operatorname{GTG} & & \\ \operatorname{GTC} & & $	* CCA P TCT C P B66 * TCC C P B66 * TCC C P B66 * TCC C C P C C C C C C C C C C C C C	GGA G G CTA L CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG	AAT N AAT N AAT N TAT Y AAAT N TAT Y AAAT N TAT Y AAAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT CAT N CAT N CAT N CAT N CAT N CAT N CAT N CAT CAT N CAT N CAT N CAT N CAT CAT CAT CAT CAT CAT CAT CAT	ACT ACC T ACC ACC ACC ACC ACC AC	GGA G G G G G G G G G G G G G G G G G G	AGC S ATC 1 ACT ACT ACT ACT ACT ACT CAA O 99* CTA L GTC V AAT 1 168* ACT 1 168* ACT 1 108 CTS 2 ACT 1 199* ACT 1 100* ACT 1 10	AAC N TOTO & TAY 88* TOO AAN TOO 1050 AAN TOO TAY 88* TOO AAN TOO TAY 20 AAC N ACT	* 12 U G V TT U G TA V 94 * 12 U AT I 12 U III * CG Q AT N AA N 28 * AE I 12 W 63 G	GAG E 778 GAT D GAA E GAT D GAA E GAA E GAG C C C C C C C C C C C C C C C C C C	ACT T GGA G G T I G G T I G G T I G G T I G G T I G G T I G G T I G G T I G G T I G G T I G G T I G G G G	$\begin{array}{c} * \\ AAC \\ N \\ \hline \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\$	ATC I 788 AAT N CTT L CTT L CCA ACC I 10 CCT A ACC I 10 CCT A ACC I 10 CCT A ACC I 10 CCT A ACC I CCT A ACC I ACC A ACC ACC A ACC A ACC ACC	GAC D GAC D GAC D GAC D AAC N GAC D AAC N AAC N AA	TTC F CAG GAG CAG F TTC F C TTC F C TAC TAC TAC TAC TAC TAC TAC	TGT C 799* GTC V TGC GTC V 96* TGC GTC C ACC T GAC A GG CAT H 136* ACC T TAT Y	A TCC STA CC TC A CC A CC A CC A CC A CC	AGC S CAG Q AGT S S AGT S AGT S AGT C CAG Q CAG Q CAG G CAG C CAG C C CAG C C CAG C C CAG C C C C C C C C C C C C C C	ACC ATC ATC 978 TCC 978 CTT C 0 GG GG GG ACC O LL14 ATC M ACC 0 C 0 ACC 0 C 0 CC 0 C 0 C 0 CC 0 C 0 C 0 CC 0 C 0 C 0 C 0 CC 0 C 0 C 0 C 0 C 0 C 0 C 0 C 0 C 0 C
GTIS GTA CCCC P P CCCC P P F F CCCC P P F CCCC CCC P F CCCC CCCC P F CCCC CCCC P CCCC CCCC P F CCCC CCCC P CCCC CCCC P CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCCC CCCC CCCCC CCCC CCCCC CCCC CCCCC CCCC CCCCC CCCC CCCCC CCCC CCCCC CCCC CCCCC CCCC CCCCC CCCC CCCCCCCC	* CCA P TCT C C C TCC C C TCC C C C C C C C C C C C C	GGA G G CTA L C CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G CAG G C CAG G C CAG G C CAG G C CAG C C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C C CAG C C CAG C C CAG C C CAG C C CAG C C CAG C C CAG C C CAG C C CAG C C C C	AAT N 810 TAC Y 380 AAT N TAT Y 380 AAC N 10 TAC Y AAC N CAT D 1210 K GTT V 1380 K AAA K	* ACT ACT ACC C C C A C C C C C C C C C C C C C	GGA G G G G G G G G G G G G G G G G G G	AGC S T ACT ACT ACT ACT ACT ACT ACT	AAC N TTC & TAY 88 * TCC AAN TTCC 1050 AAC N ACT	* TCC GT> TTC & GT> 9* TCC AT I TCC 10* CC AT AT AC 29* QE 1 TCW CC 8	GAG E 776 GAT D GAA E GAA E GAA E GAA E GAA E GAA E GAA E TAT Y Y ACT I I 340 CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C CAA C C CAA C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C C CAA C C C C CAA C C C C CAA C C C C CAA C C C C C C C C C C C C C C C C C C C C	ACT T GGA G G G G G G G G G G G G G G G G	* ANC V ANC V TO'S B90 GGA C GGA C GGA C CGA C I 1290 ACT F 200 ACT ANA K 460 ACT ANA K 460	ATC I 780 AAT N CTT L CTT L CCA 50 ATC I 10 CCT A ACC I 10 CCT A ACC C A CCT C CCA ACC C A CCT C CCA ACC C A CCT C CCA ACC C A CCT C CCA ACC C A CCT C CCA ACC C ACC	GAC D GGA GAC D GAC D	TTC F GAG E TTC F GAG E TAC F TAC C TAC C TA	TGT C 799* GTC V TCC C C 795* GTC V 96* TCC C C 795* GTC V 96* TCC C GTC V 96* TCC C GTC V 96* TCC C GTC V 13 GTC V 13	CCT A TCC S5*CA ACC TCC TCC TCC ACA ACC ACC TCC ACA ACT C C A	AGC CAG AGT AGT S CAG CAG CAG CAG CAG CAG CAG CAG	$\begin{array}{c} \mathbf{C} \\ $

group, as judged by the lengths of tandemly repeated EGF-like domains. The notch gene contains 36 repeats, lin 12 contains 11, and a partial sequence of a longer uEGF cDNA clone indicates that it repeats this pattern for at least another 1 kb, suggesting that the uEGF protein may contain a minimum of 18 contiguous repeats. In comparison with notch and lin 12, however, the uEGF-1 repeats are more rigidly conserved. Each uEGF-1 repeat contains exactly 38 amino acids. Furthermore, the register of the six cysteines within the repeat remains perfect throughout the nine sequenced repeats, unlike the situation in notch and lin 12. Similarly the uEGF-1 repeats are less divergent in the noncysteine residues. The basis of such constraint is unclear, although Stenflo *et al.* (18) suggested that the post-translational hydroxylation of aspartic acid and asparagine residues within the EGF-like repeat of various plasma proteins may require a specific cysteine register and the conservation of particular amino acids around the hydroxylated residues. Figure 3 shows that uEGF-1 contains many conserved aspartic acid and asparagine residues that might be sites for hydroxylation.

The 120 amino acids at the carboxyl terminal of the uEGF-1 ORF represent a distinct domain of the protein. This region is rich in threonine; 45% of the threonine

Demost 1	26		le.		м		n	-	lc l	١.	c	٨	Ы	c	1	N	്റ	° c		lc l	1	D	G	1	N	G	v	τl	lc	s	lc]	Р	1	G	F	s	G	D	Ν
кереат т	20	Ľ	C	v	N	1	U	-	Ľ	^	3	~	1	Č	Q			ŭ	•	Ŭ	Ľ	Ŭ	ŭ	•		ŭ	•		ľ	Ĩ	Ŭ	ľ	-		•	Č			
Repeat 2	138	с	Ε	N	N	D	D	E	с	s	S	ł	Ρ	С	٤	N	G	G	т	с	۷	D	L	۷	N	A	Y	м	С	۷	С	Α	Ρ	G	W	т	G	Ρ	т
Repeat 3	252	с	А	D	N	ł	D	E	с	A	S	A	Ρ	С	Q	Ň	G	G	۷	с	ł	D	G	۷	Ν	G	Y	м	С	D	С	Q	Ρ	G	Y	т	G	т	н
Repeat 4	366	с	ε	т	D	1	D	Ε	с	A	R	Ρ	Ρ	с	Q	N	G	G	D	с	٧	D	G	۷	N	G	Y	۷	с	I.	с	А	Ρ	G	F	D	G	L	Ν
Repeat 5	480	c	E	N	N	1	D	Ε	с	A	S	R	D	с	Q	N	I G	A	v	с	٧	D	G	۷	N	G	F	۷	с	т	С	s	A	G	۷	т	G	۷	L
Repeat 6	594	c	E	т	D	I	D	Ε	c	A	S	A	Ρ	с	Q	N	1 G	G	v	c	т	D	т	1	N	G	Y	I	с	A	с	٧	Ρ	G	F	т	G	S	N
Repeat 7	708	с	E	т	N	١	D	E	с	A	S	D	Ρ	с	L	N	I G	G	1	с	٧	D	G	۷	N	G	F	۷	с	Q	с	Ρ	Ρ	Ν	Y	S	G	т	Y
Repeat 8	822	с	E	1	S	L	D	A	с	R	S	N	Ρ	с	Q	N	I G	A	τ	c	٧	N	۷	G	A	D	Y	۷	c	E	С	۷	Ρ	G	Y	A	G	Q	N
Repeat 9	936	c	Ε	I	D	I	N	Ε	с	A	s	L	Ρ	с	Q	t	I G	G	i L	с	ŀ	D	G	ł	A	G	Y	т	с	Q	С	R	L	G	Y	ł	G	۷	N

Consensus CEXNIDECASXPCQNCGXCVDGXNGYVCXCXPG^FXGXN

**Fig. 3.** Alignment of repeated sequence within uEGF-1. The cysteine repeat register is indicated by heavy bars surrounding those residues. Invariant noncysteine amino acids are shaded. A consensus sequence is presented at the bottom and is based on the presence of an amino acid in  $\geq$ 50% of the sequenced repeats. Boxed X's indicate substitutions limited to nonpolar amino acids. X's without boxes indicate nonconservative substitution at that position. The numbers at the beginning of each repeat indicate the first nucleotide of that repeat.

uEGF-1	123	C	E	т	-	D	Ĩ.	D	E	C	-	-	-	A	R	Ρ	P	C	Q	N	G	G	D	C	-	-	v	D	G	v	N	G	Y	۷	c	1	C	Α	Ρ	G	F	D	G	L
notch	215	c	s	Y	-	Ð	1	E	ε	с	-	-	-	Q	s	N	p	с	к	Y	G	G	5000 1	с	-	-	v	N	т	н	G	S	Y	Q	с	м	с	Ρ	т	6	Y	т	G	к
lin 12	327	к	s	s	L	s	E	N	L	с	-	-	-	L	s	D	p	c	м	N	N	A	т	с	1	D	v	D	A	н	I	G	Y	A	с	1	с	ĸ	Q	G	F	Ë	G	D
EGF human	2					s	D	s	E	с	Ρ	L	s	н	D	G	Y	c	L	. н	D	G	۷	с	м	Y	I	E	A	L	D	ĸ	Y	A	с	N	с	v	۷	G	Y	I	G	E
EGF prec. mouse	361	c	E	-	-	D	V	N	E	с	-	A	т	Q	N	н	G	с	-	т	L	G	-	c	-	-	E	N	Y	Ρ	G	s	Y	н	с	т	с	Ρ	т	G	v	L	L	Ρ
Plasmin ACT human	79	c	н	s	v	Ρ	v	к	s	с	-	-	~	s	E	Ρ	т	c	F	N	G	G	т	c	Q	Q	A	L	Y	F	s	D	F	v	с	Q	с	Ρ	E	6	F	A	G	к
Factor IX bovine	45		Q	Y	V	D	G	D	Q	с	-	-	-	E	s	N	P	c	L	N	G	G	M	c	-	-	к	D	D	I	N	s	Ŷ	E	с	w	с	Q	A	G	F	E	G	τ
Factor X human	71	с	к	Y	к	D	G	D	Q	с	-	-	-	E	т	s	P	c	Q	N	Q	G	ĸ	c	-	-	к	D	G	L	G	E	Y	т	с	т	с	L	E	G	F	E	G	к
LDL recept.	353	с	E	-	-	D	1	D	E	с	Q	D	p	D	-	-	т	c	s	Q	L	-	-	c	-	-	۷	N	L	E	G	G	Y	к	с	Q	с	E	E	G	F	Q	L	D

**Fig. 4.** Alignment of uEGF-1 with other EGF-like proteins. With the exception of *notch* and *lin 12*, all protein sequences depicted were obtained by searching the PIR database (fastP program) with the fourth repeat of uEGF-1. The fourth repeat was chosen as a representative repeat to avoid using unspecified amino acids in the analysis. A *notch* repeat was also chosen because of its similarity to the *notch* consensus (2). The *lin 12* repeat used here was the repeat analyzed (4). Hyphens indicate gaps introduced to maximize homology. Cysteines are indicated with heavy bars. Amino acids which occur in  $\geq 50\%$  of the sequences are shaded. Numbers at the beginning of every sequence indicate the number of that amino acid residue within the total protein sequence.

18 SEPTEMBER 1987

residues of the predicted uEGF-1 protein appear in this second domain, which comprises only 25% of the ORF. The threonine residues are found in clusters of two or more, often interspersed with serine residues. Clusters such as this are thought to be sites of O-linked glycosylation (13, 19). A scan of the PIR database with this domain of uEGF-1 did not reveal any significant homologies to known proteins.

The members of the EGF family are thought to execute their function extracellularly either as membrane-bound proteins (EGF, TGF- $\alpha$ , VVGF, LDLR, and notch) or as secreted proteins (factor X and mature EGF). A computer-aided analysis of uEGF-1, in which the hydropathy algorithm of Kyte and Doolittle (20) was performed by the programs of Pustell (International Biotechnologies), does not reveal a classic transmembrane domain as defined by those workers. Because we do not have complete cDNA clones for the uEGF-1 protein, we do not know whether the uEGF protein contains a signal sequence. The presence of so many potential O-linked glycosylation sites clustered in a discrete region is suggestive of an extracellular protein, but the cellular location of uEGF is currently unknown.

The temporal pattern of expression of transcripts recognized by the uEGF-1 cDNA clone is consistent with a function during early sea urchin development, a time when cell fates are known to be assigned (21). Figure 5a represents an RNA gel blot of staged embryonic RNAs probed with the uEGF-1 cDNA clone. Restriction fragments from all portions of the clone recognize both the 3.0- and the 4.0-kb transcripts. Whether these transcripts emanate from different genes or are alternatively spliced messages from a single gene cannot be determined at present. Figure 5b shows a quantitation of the RNA gel blot data. The transcripts recognized by uEGF-1 are at their highest level in the unfertilized egg, decline in prevalence during early cleavage, and increase in prevalence at the blastula stage before tapering off after gastrulation. This is an unusual transcript accumulation pattern for a sea urchin messenger RNA (mRNA). Studies of randomly chosen cDNAs have shown that many prevalent messages accumulate early, decay, and are replaced later in development (22), but none have been reported to do so within the first 6 hours of development. The rate of transcription per cell in sea urchins is virtually unchanged throughout embryogenesis (23). Thus uEGF transcript accumulation cannot be explained by changes in the transcriptional ability of zygotic nuclei, as would be the case in Xenopus or Drosophila. The accumulation of histone message is also shown in Fig. 5b. The



Fig. 5. Time course of uEGF transcript accumulation. (a) RNA gel blot analysis, carried out as described in Fig. 1. Numbers above the lanes indicate hours of development at 13°C. These times represent the following developmental stages; 0 hours, unfertilized egg; 5.5 hours, 8-cell stage; 6.3 hours, 16-cell stage; 10 hours, latecleavage stage; 14 hours, prehatching blastula; 17 hours, late-hatched blastula; 27 hours, mesenchyme blastula; 41 hours, early gastrula; 53 hours, late gastrula. (b) Quantification by densitometry of RNA gel blot data. The ordinate represents hours of development at 13°C. The abscissa represents arbitrary values relative to the maximal accumulation level.  $(\bullet)$  The behavior of the uEGF transcript; the results shown are the average of two blots, one of which is shown in (a). The dashed line represents the accumulation of the  $\alpha$  histories and was plotted from (36).

synthesis of histones in this embryo correlates with the rate of cell division (24). Because the second period of uEGF message accumulation occurs after the rapid cell divisions of early cleavage, we do not expect uEGF to be operating solely as a stimulator of cell proliferation in this system. The transcripts of many lineage-specific genes in sea urchins first appear at the blastula stage; they accumulate asynchronously, but their midpoints of accumulation all occur after the midpoint of accumulation of the uEGF transcript (25-27).

Some mammalian members of the EGF family are thought to act as regulators of cell proliferation and differentiation (1, 28). The EGF-like molecules encoded by notch and lin 12 are involved in the specification of embryonic cell fates (29, 30). The behavior of the uEGF transcript, whose accumulation precedes known molecular indices of cell differentiation, as well as the structural similarity of the uEGF gene to other invertebrate binary switch genes, suggests that it

too may play an important role in embryogenesis. Whatever its function, the existence of this highly conserved sequence in the echinoderms indicates that the EGF peptide domain predates the radiation of coelomate animals and has been retained in a number of diverse evolutionary lineages.

**REFERENCES AND NOTES** 

- 1. G. Carpenter and S. Cohen, Annu. Rev. Biochem. 48, 193 (1979).
- K. A. Wharton, K. M. Johansen, T. Xu, S. Artavanis-Tsakonas, Cell 43, 567 (1985).
   S. Kidd, M. R. Kelley, M. W. Young, Mol. Cell Biol.
- 6, 3094 (1986).
- I. Greenwald, Cell 43, 583 (1985).
- 5. M. A. Harkey and A. H. Whiteley, Wilhelm Rouse Arch. Dev. Biol. 189, 111 (1980).
- 6. D. McClay and A. F. Chambers, Dev. Biol. 63, 179 (1978).
- 7. T. Maruyama, T. Gojobori, S.-I. Aota, T. Ikemura,
- Nucleic Acids Res. 14 (suppl.) r151 (1986).
   D. J. Lipman and W. R. Pearson, Science 227, 1435 (1985)
- 9 R. F. Doolittle, D. F. Feng, M. S. Johnson, Nature (London) 307, 558 (1984). J. P. Brown, D. R. Twardzik, H. Marquardt, G. J.
- 10. Todaro, ibid. 313, 491 (1985).
- H. Marquardt, M. W. Hunkapiller, L. E. Hood, G. 11. J. Todaro, Science 223, 1079 (1984).
- D. Pennica et al., Nature (London) 301, 214 (1983).
   B. A. McMullen et al., Biochemistry 22, 2875 (1983).
- D. L. Enfield et al., ibid. 19, 659 (1980).
   D. W. Russell et al., Cell 37, 577 (1985)
- R. W. Jackman et al., Proc. Natl. Acad. Sci. U.S.A. 16. 83, 8834 (1986). 17. J. Hoskins, D. K. Norman, R. J. Beckman, G. L.
- Long, ibid. 84, 349 (1987).

- 18. J. Stenflo, A. Lundwall, B. Dahlback, *ibid.*, p. 368. 19. T. Nikaido et al., Nature (London) 311, 631 (1985).
- 20. J. Kyte and R. F. Doolittle, J. Mol. Biol. 157, 105 (1982).
- 21. R. C. Angerer and E. H. Davidson, Science 226, 1153 (1984).
- 22. C. N. Flytzanis, B. P. Brandhorst, R. J. Britten, E. H. Davidson, Dev. Biol. 91, 27 (1982).
- 23. E. H. Davidson, Gene Activity in Early Development (Academic Press, New York, 1976).
- 24. B. Moav and M. Nemer, Biochemistry 10, 881 (1971)
- 25. W. H. Klein et al., in Time, Space, and Pattern in Embryonic Development, W. R. Jeffery and R. A. Raff, Eds. (Liss, New York, 1983), pp. 87–100. 26. R. J. Shott *et al.*, *Dev. Biol.* 101, 295 (1984)
- 27. D. S. Leaf et al., ibid. 121, 29 (1987) 28. D. R. Twardzik, J. E. Ranchalis, G. J. Todaro,
- Cancer Res. 42, 590 (1982). 29. I. S. Greenwald, P. W. Sternberg, H. R. Horvitz,
- Cell 34, 435 (1983). 30. B. Yedvobnik et al., Cold Spring Harbor Symp.
- Quant. Biol. 50, 841 (1985).
- 31. D. A. Hursh and R. A. Raff, in preparation. 32. J. Messing, Methods Enzymol. 101, 20 (1983).
- J. Hecking, Hermitic Lingmer. 101, 20 (1900).
   F. Sanger, S. Nicklen, A. R. Coulson, Proc. Natl. Acad. Sci. U.S.A. 74, 5463 (1977).
- 34. A. J. H. Smith, Methods Enzymol. 65, 560 (1980).
- M. D. Biggin, T. J. Gibson, G. F. Hong, Proc. Natl. Acad. Sci. U.S.A. 80, 3963 (1983).
   D. E. Wells et al., Wilhelm Roux Arch. Dev. Biol. 195, 252 (1986).
- We thank T. Thomas for providing the cDNA libraries used in this study. We thank E. Raff, B. 37. Parr, D. Leaf, P. Cherbas, and M. Muskavitch for critical comments regarding the manuscript. Supported by NIH grant R01 HD21986. D.A.H. was supported by an NIH predoctoral training grant.

23 March 1987; accepted 26 June 1987

## Acantharian Fluxes and Strontium to Chlorinity **Ratios in the North Pacific Ocean**

R. E. BERNSTEIN, P. R. BETZER, R. A. FEELY, R. H. BYRNE, M. F. LAMB, A. F. MICHAELS

Data on particulate strontium sulfate fluxes and strontium to chlorinity ratios were compared to provide insights into the strontium cycle of the North Pacific. Freedrifting sediment traps were used to derive large particle fluxes between depths of 100 and 3500 meters in the eastern and western North Pacific Ocean. Flux data revealed substantial quantities of acantharian skeletons and cysts (both made of strontium sulfate) settling through the upper kilometer of the water column. The greatest fluxes of celestite were detected at 400 meters. Minimal to nondetectable fluxes noted at and below 900 meters provide evidence that by this horizon, the majority of acantharian specimens had dissolved, thereby contributing to the pool of dissolved strontium. Growth and subsequent dissolution of acantharians in the upper kilometer are qualitatively consistent with the well-developed minimum and maximum strontium to chlorinity ratios that are consistently noted in these waters. These fluxes of particulate strontium and model calculations for fluxes of dissolved strontium indicate that acantharians play an important role in the ocean's strontium budget.

HE GLOBAL OCEANIC STRONTIUM budget may be dramatically affected

by the only marine organisms to use Sr as a major skeletal component. These abundant organisms, the acantharians, are marine planktonic protists that secrete a

R. E. Bernstein, P. R. Betzer, R. H. Byrne, Department of Marine Science, University of South Florida, St. Petersburg, FL 33701. R. A. Feely and M. F. Lamb, Pacific Marine Environ-

Administration, Seattle, WA 98115. A. F. Michaels, Institute of Marine Sciences, University

of California, Santa Cruz, CA 95060.