Precursor-product relationship between vitellogenin and the yolk proteins as derived from the complete sequence of a *Xenopus* vitellogenin gene

Susan Gerber-Huber¹, Denise Nardelli, Jacques-Antoine Haefliger², David N.Cooper³, Françoise Givel, Jacques-Edouard Germond⁴, Jürgen Engel*, N.Michael Green⁺ and Walter Wahli

Institut de Biologie Animale, Université de Lausanne, Bâtiment de Biologie, CH-1015 Lausanne,
*Abteilung für Biophysikalische Chemie, Biozentrum der Universität Basel, CH-4056 Basel,
Switzerland and ⁺National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

Received April 23, 1987; Accepted May 26, 1987

Accession no. Y00354

ABSTRACT

In *Xenopus laevis* four estrogen-responsive genes are expressed simultaneously to produce vitellogenin, the precursor of the yolk proteins. One of these four genes, the gene A2, was sequenced completely, as well as cDNAs representing 75% of the coding region of the gene. From this data the exon - intron structure of the gene was established, revealing 35 exons that give a transcript of 5,619 bp without the poly A-tail. This A2 transcript encodes a vitellogenin of 1,807 amino acids, whose structure is discussed with respect to its function. At the nucleic acid as well as at the protein level no extensive homologies with any sequences other than vitellogenin were observed. Comparison of the amino acid sequence of the vitellogenin A2 molecule with biochemical data obtained from the different yolk proteins allowed us to localize the cleavage products on the vitellogenin precursor as follows : NH₂ - lipovitellin I - phosvitin (or phosvette II - phosvette I) - lipovitellin II - COOH.

INTRODUCTION

Synthesis of vitellogenin in the liver of oviparous vertebrates is regulated by estrogen which is produced by the ovary. Vitellogenin, a large lipoglycophosphoprotein, is post-translationally modified already in the liver and then secreted into the bloodstream. It is selectively taken up by growing oocytes where it is cleaved into the yolk proteins, lipovitellin and phosvitin, and stored in the form of yolk platelets or granules. In mature *Xenopus laevis* oocytes the yolk proteins, that are exclusively derived from vitellogenin, represent about 80% of the total protein mass and will serve as a nutrient source during early embryogenesis (1).

The accompanying cellular and molecular events that occur in the liver and ovary provide an ideal system in which to study many aspects of several vital processes. For instance, vitellogenin synthesis in the liver is an attractive model for transcriptional and post-transcriptional regulation of

eukaryotic gene expression by steroids (2). Moreover, the understanding of the mechanisms of vitellogenin post-translational modification, namely glycosylation, phosphorylation, lipidation, and of the secretory pathway is still incomplete. Many other questions remain to be resolved. These include ovary-related events such as passage of macromolecular material through the follicular epithelium, selective binding and uptake of vitellogenin by the growing oocyte, processing of the protein and translocation within the vitellogenic oocyte, as well as formation of the yolk platelets (1).

Clearly, progress in these areas will be very difficult without a thorough understanding of the biochemical characteristics of vitellogenin. This would be facilitated by the determination of the primary sequence of the protein. In *Xenopus laevis*, four genes have been identified that encode distinct, but closely related vitellogenins (3, 4). The shortest of the four is the gene A2, which is about 16 kb long (5, 6, 7). Here we present the complete primary structure of this gene from which we have derived the amino acid sequence of the 200 kD protein A2. The characteristics of the gene and of the protein are discussed.

MATERIAL AND METHODS

DNA clones used and sequencing strategy

The vitellogenin genomic clone λ Xlv 128 that contains the entire gene A2 and its flanking sequences (5, 6) and the two A2 cDNA clones pXlvc 18 and 34 which together represent approximately 75 percent of the A2 mRNA (8, 3), were used as starting material. The genomic DNA sequences were originally cloned in the lambda Charon 4 replacement vector. The cDNA sequences were cloned in the vector pMB9, into which they were inserted by A/T tailing at the Eco RI site.

Both shotgun and directed sequencing were used for these clones. In shotgun sequencing experiments, random restriction fragments, mainly derived from the two cloned cDNAs, were inserted into suitable M13 vectors (9). For directed sequencing, large fragments derived from the genomic or cDNA clones were inserted into the Wayne-Barnes vector (10) as modified by Sahli *et al.* (11). Serial deletions were created either by DNase I (12) or with exonuclease III (13). Sequencing was performed by the dideoxy chain

termination method (14) and in a few cases by the chemical cleavage method (15). The nucleotide sequences were handled on a Norsk Data N540 computer with different analysis programs kindly provided by Rmy Fritz (LGME, Strasbourg).

Determination of the exon-intron structure

The boundaries of the first three exons were determined previously by RNA sequencing (16, 17). The positions of exons 4 to 6 were deduced from the combination of R-loop mapping (6) and the occurrence of consensus splice sites that maintain an open reading frame. The arrangement obtained corresponds to the one found in the Xenopus gene A1 (unpublished) and in the chicken vitellogenin gene (18). The positions of exons 6-9 were determined by comparing the gene sequence with that of a recently obtained A2 cDNA clone (unpublished). The exon-intron structure of the remainder of the gene (exons 9-35) was deduced directly from comparison of the genomic sequences with the cloned mRNA sequences in pXlvc 18 and 34.

Nuclease S1 mapping

Vitellogenin mRNA was purified from the liver of an estrogen-treated Xenopus laevis female as previously described (19). A genomic Sau 3A - Eco RI fragment, containing the putative 3'-end of the vitellogenin gene A2 (position 16,547 to 17,459, see appendix), was labelled at the Sau 3A-site with α -³²P-dGTP using the Klenow fragment of DNA polymerase I, and was used as probe. The S1 nuclease analysis was performed as described by Walker et al. (16) with the exception of the hybridization and S1 nuclease digestion temperatures, which were 30C and 25C, respectively.

RESULTS AND DISCUSSION

Overall sequencing strategy and primary structure of the vitellogenin gene A2

We have sequenced the entire vitellogenin gene A2 and its flanking regions using primarily the chain termination method. Figure 1 illustrates the sequencing strategy and the exon-intron map of the gene and indicates the portion of the coding region contained in the cDNA clones pXlvc 18 and 34 whose primary structure has also been determined. In the appendix we present 18,488 base pairs of genomic DNA representing all of the gene A2,

Figure 1 Exon-intron organization and sequencing strategy of the vitellogenin gene A2 and its cDNAs. The vitellogenin gene A2 locus and the two cDNA clones, pXlvc 18 and 34, are shown with their Eco RI ($\square$) and HindIII (∇) restriction sites. The initiation site of transcription as well as the polyadenylation site are indicated by the big arrow ($\blacktriangleright$) and the numbered exons by grey boxes. The sequenced clones are designated by slim horizontal arrows. Sequence data from 1 kb upstream of the transcription initiation site (position -1,028 in appendix) to intron 3 has been published previously (16, 17). The position of the transcription initiation site of the gene A2 (position +1 in appendix) has also been previously determined by S1 nuclease mapping (16). The 5'-end of the cDNA clone 18 lies in exon 9 and its 3'-end in exon 23. A gap of 207 base pairs in this exon separates the 3'-end of the pXlvc 18 DNA and the 5'-end of the pXlvc 34 DNA. The 3'-end of the latter cDNA clone occurs after the polyadenylation signal close or at the polyadenylation site. The repetitive elements, Rem I and Vi, are marked by a thick line (—).

as well as its 5'- and 3'-flanking regions.

The transcription initiation site has been mapped previously (16). At the 3'-end of the gene, two hexamers occur corresponding to the polyadenylation signal AATAAA separated by 425 base pairs (positions 16,706 and 17,131, appendix). Multiple polyadenylation signals have been described for various genes (20, 21, 22, 23, 24) and in some cases the signals are used alternatively in mRNA processing. Clearly the hexanucleotide AATAAA is necessary but might not be sufficient for the formation of correct mRNA 3'-termini. Indeed, McLauchlan *et al.* (25) have found that in approximately 70% of about 100 examined mRNA 3'-termini, an additional conserved sequence is located about 30 bp downstream from the polyadenylation signal. This additional sequence, from which the consensus YGTGTTY was derived, is not associated with 3'-end selection but is required for efficient formation of correct mRNA 3'-termini (26). Such a sequence, with one mismatch to the consensus element, is found 31 bp downstream from the first AATAAA signal and, at an unexpected position, 7 bp upstream of the second signal (appendix). S1 nuclease analysis of hybrids between vitellogenin mRNA and an end-labelled genomic fragment comprising both hexamers revealed that only the first one (position 16,706, appendix) serves as polyadenylation signal since the 3'-end of the mRNA maps 15-20 nucleotides downstream of it (Fig. 2). This is also the region where sequencing experiments placed the 3'-end of the cDNA clone pXlvc 34 (Fig. 2) and of two additional cDNA clones sequenced recently (unpublished). However, the observation that the polyadenylation signal utilized is associated, as expected, with the YGTGTTY element, does not prove that this association by itself is sufficient for the recognition of the first site.

From this analysis the total length of the gene A2 can be deduced. From the initiation site of transcription to the polyadenylation site, the gene is 16,731 bp long assuming that the underlined A in the sequence TAGAATG (appendix; complementary sequence in Fig. 2 is ATCTTAC) is the 3' - end of the mRNA.

Exon-Intron structure of the gene A2

The gene A2 comprises 35 exons and 34 introns (Fig.1), whose lengths are given in Table I. The total combined length of the 35 exons is 5,619 bp,

including the 5'- and 3'- nontranslated regions of 13 and 182 nucleotides, respectively. This agrees with our earlier electron microscopic estimations of the mRNA length (19). The mean exon length is 161 bp and there is a single relatively long exon, number 23, with 573 bp in length. The 35 exons of the gene A2 encode a polypeptide of 1,807 amino acids that, without post-translational modifications, has a molecular weight of 201,360.

The genomic exon sequences were compared with the corresponding cDNA sequences over 4,170 bp. The 3'-end of clone pXlvc 18 maps within a cluster of adenines (position 10,230, appendix). We therefore suggest that synthesis of the pXlvc 18 cDNA by oligo dT-primed reverse transcriptase has begun at this internal mRNA position rather than at the 3'-end poly-A sequence of the mRNA. This is consistent with earlier observations in which an oligo(A) region, previously mapped about 3.7 kb downstream from the 5'-end of the mRNA, appears to be a start site for oligo (dT)-primed cDNA synthesis (27). It would also explain the occurrence of several other characterized cloned cDNAs which in all likelihood originated from this internal position (3). Comparing cDNA and genomic sequences five differences were revealed, whose positions are given in Table II. Three of them, with a G in the genomic sequence and an A in the cDNA sequence, lead to a change in an amino acid. It is unclear whether these differences are due to a high level of polymorphism within the gene or to errors inherent in cDNA synthesis and other experimental manipulations.

The 34 introns of the gene A2, which vary in length from 77 bp to 1,105 bp, have a combined length of 11,112 bp. The mean length, 327 bp, is only two times longer than the mean exon length. It is worth noting that half of

Figure 2 Nuclease S1 mapping of the 3'- end of the vitellogenin gene A2. The 3'- end-labelled probe (Sau 3A, position 16,547 in the genomic sequence) was hybridized to poly(A)-containing RNA from the liver of an estrogen-stimulated female (slot no. 1) and to tRNA (slot no. 2). The marker (slot no. 3) is pBR 322 DNA cut with Hinf III. The slots G, G+A, and T+C show a Maxam-Gilbert reaction on the same probe, the sequence (complementary to the RNA strand) is given to the left-hand side. The protected fragment is 185 b long. For comparison, the 3'-end sequence of the cDNA clone pXlvc 34 is shown.

TABLE I Exon and intron lengths of the vitellogenin gene A2

EXONS		INTRONS	
Number	Length (in bp)	Length (in bp)	Number
1 untrl.	13		
1 trl.	40	213	1
2	21	80	2
3	152	714	3
4	252	165	4
5	162	93	5
6	149	563	6
7	151	116	7
8	158	467	8
9	166	521	9
10	126	911	10
11	219	220	11
12	224	308	12
13	98	111	13
14	125	118	14
15	102	85	15
16	120	373	16
17	190	77	17
18	94	156	18
19	186	1105	19
20	148	330	20
21	217	86	21
22	173	90	22
23	573	128	23
24	51	1034	24
25	81	90	25
26	210	96	26
27	143	783	27
28	139	150	28
29	136	80	29
30	187	233	30
31	95	92	31
32	98	646	32
33	160	289	33
34	162	589	34
35 trl.	116		
35 untrl.	183		
Total	5619	11112	Total

Exon 1 and 35 contain each an untranslated (untrl.) and a translated (trl.) portion.

the introns are shorter than 200 bp, and only two are longer than 1 kb. The occurrence of relatively short introns is characteristic of the A2 gene when compared to the other *Xenopus* vitellogenin genes. The latter are up to 5 kb longer but produce mRNAs of the same length (6, 7). The exon-intron junctions are in agreement with the consensus sequence as proposed by Mount (28).

Identification of repetitive DNA elements within introns

Some of the A2 introns carry repetitive DNA elements (29, 30). For example, intron 27 contains a copy of the transposon-like Vi-element (Fig. 1 and appendix) that has also been found elsewhere in the vitellogenin locus and has characteristics of a mobile genetic element (31).

TABLE II Differences between vitellogenin A2 genomic and cDNA sequences

Position in seq.	Genomic sequence		cDNA sequence		clone
	nucleotide	amino acid	amino acid	nucleotide	
6478 (14)	Gtt	Val	Ile	Att	18
9280 (20)	Gct	Ala	Thr	Act	
9732 (21)	ggA	Gly	Gly	ggT	
13989 (30)	Gag	Glu	Lys	Aag	34
15836 (34)	ctG	Leu	Leu	ctC	

The positions indicated refer to the genomic sequence (Appendix). The numbers in brackets refer to the exon number. The nucleotides in capitals are the positions which differ.

In addition, a comparison of the A2 sequence with sequences in the EMBL sequence library (release No.10) revealed a *Xenopus* Repetitive Eco RI Monomer 1 element (Rem 1, ref. 32) within intron 19 (Fig. 1 and appendix). This identifies a repetitive sequence which we had earlier localized at this position, as well as in two other locations, namely the 5'-end region of the gene A2 and near the 3'-end region of gene A1 (29, 30).

Earlier hybridization analysis have revealed the presence of additional repetitive sequences within the A2 introns (29). These have not yet been localized and do not correspond to already characterized elements as revealed by a search in the EMBL sequence library.

Precursor-product relationship between vitellogenin and yolk proteins

The vitellogenin polypeptide is cleaved, within the oocyte (33). With respect to the cleavage pattern, two types of vitellogenin have been recently proposed : in the first, the precursor is cleaved in two places, yielding lipovitellin I, lipovitellin II and phosvitin, whereas in the second type the precursor is cleaved at an additional site within the serine-rich domain, yielding two phosvettes instead of one phosvitin (34).

Sequence determination of the first 13 amino acids at the amino-terminus of secreted vitellogenin (data not shown) revealed that glutamic acid, encoded by the third codon of exon 2, represents the amino end of the circulating protein (Fig. 4). This result strengthens our earlier hypothesis that the very short 21 bp exon 2, which is also found in the chicken vitellogenin gene, has been conserved in both amphibians and birds by positive selection due to the presence of the signal peptidase recognition site (16, 17). This cleavage site (Fig. 4) fulfills the "(-3, -1) rule" of von Heijne (35). Thus, the leader peptide of the vitellogenin A2 is

Figure 3 Location of the yolk proteins on the vitellogenin precursor. The signal peptide of 15 amino acids at the N-terminus of the vitellogenin molecule and the internal serine clusters are given as black boxes. The yolk proteins were placed on their common precursor based upon their molecular weight, their amino acid composition (32) and the primary amino acid sequence of the vitellogenin A2 protein deduced from data presented in this paper. Phosvitin (P) was placed by centering it over the serine-rich domain. Then, amino acids were counted individually and bidirectionally along the primary vitellogenin sequence up to the known content; an amino- and a carboxy-proximal dot bounding this content were placed. Amino acid composition allowed us to give the putative positions of phosvette I and II (PVI + PVII) in the order shown. Lipovitellin I (LVI), the largest yolk subunit, can only be placed starting at the N-terminus of the secreted vitellogenin. This has been confirmed by sequencing the amino-end of LVI which corresponds to the one of secreted vitellogenin (data not shown). The carboxy-proximal boundary of the content for the different amino acids are indicated by dots. Lipovitellin II (LVII) can be placed either adjacent to the putative C-terminus of phosvitin or at the very carboxy-end of vitellogenin. Amino acid content boundaries and sequence conservation compared to other vitellogenins (see text) favour the carboxy-proximal position. Minimal and maximal length with a dashed area inbetween is given for each yolk protein based on their minimal and maximal reported molecular weight.

comprised of the first 15 amino acids at the amino end of the primary translation product (Fig. 4).

As shown in Figure 3, our results allow us to position the mature yolk proteins within the vitellogenin precursor molecule. This is based on the amino acid sequence of vitellogenin A2 presented here and the molecular weights and amino acid compositions of the vitellogenins and the yolk proteins as determined by Wiley and Wallace (34). The serine rich portion of vitellogenin, which encodes phosvitin or phosvettes, occupies an inter-

nal position. Between the amino terminus of this serine-rich domain and the amino terminus of the vitellogenin precursor is the region encoding lipovitellin I. As shown in Figure 3, this is its only possible location due to its size. In this arrangement, the serine stretches (black boxes in Fig. 3) begin near the putative C-terminal region of lipovitellin I, thus indicating that minimal or perhaps no loss of amino acids occurs between lipovitellin I and the serine-rich domain during cleavage of the vitellogenin precursor. Obviously this serine-rich region represents phosvitin or phosvettes. Unfortunately, it is not possible to determine on the basis of the amino acid sequence whether the serine-rich region of the gene A2 gives rise to a phosvitin molecule or, alternatively, to two phosvette molecules. If the latter were the case, then, phosvette II (no met, no thr, etc.) must precede phosvette I, again based on molecular weights and amino acid compositions (Fig. 3). The region from the putative C-terminus of phosvitin or phosvette I to the translation stop codon of vitellogenin could encode a polypeptide of about 50,000 Mr. However, lipovitellin II has a molecular weight of 30,500 to 34,000. Thus there are two theoretically possible positions of lipovitellin II (Fig. 3). The more likely one, based upon its amino acid composition and upon sequence homologies with the chicken vitellogenin gene, is that closer to the carboxy terminus. This leaves enough coding capacity for a putative polypeptide of 15,000 to 20,000 Mr between phosvitin (or the phosvettes) and lipovitellin II. This part of the precursor is either lost during vitellogenin cleavage or has yet to be identified as a component of the egg proteins. A similar arrangement to the one presented in Figure 3 has been proposed by Wiley (36) based on NTCB and CNBr cleavage data. These latter results indicated that phosvitin/-phosvettes were derived from a position in the vitellogenin molecule between the lipovitellin I and lipovitellin II regions.

Sequence comparison with other proteins

Screening of the National Biomedical Research Foundation Protein Sequence Data Base (Dayhoff Library) did not reveal internal homologies and no clear homologies between vitellogenin and other proteins. A possible exception is a 26 amino acid residue long segment (476 to 501, see Fig. 4) which was found to be homologous with segment 394 to 419 of the gene A

4748

protein of bacteriophage G4 (37). Based on identical residues and introducing a single gap the homology was 42% and when leucine was matched with isoleucine, methionine, valine and arginine with lysine it increased to 65%. It is interesting to note that part of the 476 to 501 region is well conserved in the chicken vitellogenin (unpublished). A functional relationship with gene A protein is however not readily apparent. The latter is important in early phases of DNA replication and bacteriophage assembly (37).

In particular we tested for similarities between lipovitellin I and II and other lipoproteins. Nothing was found similar to the regular amphipathic helices of the apo A and apo E proteins (38), nor was there significant sequence homology to the long partial sequence (1,445 amino acids) of human apo B which is quite distinct from the other apolipoproteins (39).

Recently it has been proposed from biochemical analyses that the heavy subunit of Rana pipiens lipovitellin contains domains with chymotrypsin and trypsin inhibitor activities and it has been suggested that such inhibitors might play an active role in the control of early development (40). Unfortunately amino acid sequence data are not yet available for R. pipiens lipovitellin; thus, we compared our amino acid sequence data with known sequences displaying protease inhibitor activity (41, 42). No homologies were found with Xenopus vitellogenin. This leaves open the question of the nature and the origin of amphibian embryo trypsin and chymotrypsin inhibitors found associated with the yolk.

Finally we compared the pattern of cysteine repeats (see Fig. 4) in the C-terminal lipovitellin II domain with that of other cysteine-rich proteins. Interestingly, this motif differs from the known motifs in epidermal growth factor (EGF), fibronectin, prothrombin, laminin, etc. (42,

Figure 4 Some characteristics of the vitellogenin A2 protein. The primary sequence is shown in the one-letter code and the exon boundaries with the exon number are given. Also indicated are : (underlined : ~~~) the signal peptide, the serine clusters (boxed), the position of cysteines (*), the position of doublets or triplets of positively charged residues (hatched) and the position of motifs in which two or more hydrophobic residues precede a positively charged residue (underlined:——). A sequence region with 42% homology to gene A protein of bacteriophage G4 (37) is marked by a dashed line.

43), with exception of the cysteine-glutamic acid-threonine-serine-proline-cysteine (CETSPC) loop at the carboxy terminus (Fig. 4) that also occurs in the factor X protein (44).

Other features of the vitellogenin sequence

Application of the structure prediction algorithm of Chou and Fasman (45) and Garnier *et al.* (46) revealed about equal proportions of α -helix and β -structure arranged in short stretches, often interrupted by conformations predicted as turns or coils in the lipovitellin I region of the sequence. A similar distribution was predicted for the C-terminal segment assigned to lipovitellin II. A stretch of about 100 residues (pos. 1,022 to 1,125) was predicted to be predominantly α -helical but manyfold interrupted by prolines. Clearly a coiled conformation was deduced for the serine-rich segment (pos. 1,126 to 1,320) which is followed by a segment (pos. 1,321 to 1,500) for which predominantly β -structure interrupted by turns and coil regions was predicted. A more detailed analysis by structure prediction methods appears to be justified only when conformation related information on the protein becomes available. Already at this stage it is apparent, however, that a number of domains with different conformation can be distinguished in vitellogenin and that there is a correspondence of these domains with the lipovitellin I, lipovitellin II and phosvitin domains. Another interesting feature of the sequence is the distribution of residues with charged side chains. In total vitellogenin has 1.52 times more positively charged residues (lysine, arginine, histidine) than negatively charged ones (glutamic acid, aspartic acid). The imbalance amounts to more than 2 in the segment at position 990 to 1,500 where 89 positive charges are only partially compensated for by 43 negative charges. The excess of 46 positive charges distributed in sequence regions adjacent to the serine clusters probably compensates for the negatively charged phosphate groups. The about 90 serines in the clusters are assumed to be phosphorylated to a level of about 80% (34). By contrast the C-terminal portion of the molecule starting from residue 1,500 contains a slight excess of negatively charged residues.

Other striking features of the vitellogenin sequence are the frequently occurring pairs or triplets of positively charged residues and the un-

usually frequent motif of two or more hydrophobic residues preceding a positively charged residue (Fig. 4). Whether there is a relation with the lipid binding property of the protein remains to be investigated.

At the protein level, the organization of the vitellogenin molecule is such that the phosvitin/phosvettes region is situated between lipovitellin I and lipovitellin II, which can bind tightly to each other via hydrophobic interactions (33). Thus, as similarly suggested by Wiley (36), the highly charged domain comprising phosphoserine and the arginine-lysine-histidine-clusters would form an exposed, very hydrophilic loop conferring solubility to the molecule. Circulating dimers would then be formed by association of such molecules, possibly in their lipovitellin I domain, thus forming a hydrophobic moiety surrounded by two hydrophilic loops. Removal of these loops by cleavage within the growing oocyte is most likely a major step in the conversion of soluble vitellogenin into insoluble yolk proteins, which are assembled in crystalline yolk platelets.

ACKNOWLEDGEMENTS

We thank Drs. Johann Schaller and Egon Rickli (Bern) for amino acid sequence determinations, Dr. Keith K. Stanley for discussions, Drs. Bob Hipskind and Riccardo Wittek for comments on the manuscript and H. Pagel for secretarial help. This work was supported by the Etat de Vaud and the Swiss National Science Foundation. D.N.C. was supported by an EMBO Long Term Fellowship.

Present addresses: Institut für Biochemie und Molekularbiologie, Universität Bern, Bülhstrasse 28, CH-3012 Bern, Switzerland¹⁾; Institut de Biochimie, Université de Lausanne, CH-1066 Epalinges, Switzerland²⁾; Haematology Department, King's College Hospital, Denmark Hill, London SE25 8RX, UK³⁾; Research Department, Nestec Ltd, Avenue Nestlé 55, CH-1800 Vevey, Switzerland⁴⁾.

REFERENCES

1. Wallace, R.A. (1985). In *Developmental Biology*, vol. 1, Ed. L.W. Browder, Plenum Publishing Corporation, 127-177.
2. Wahli, W., and Ryffel, G.U. (1985). In *Oxford Surveys on Eukaryotic*

- Genes, vol. 2, Ed. N. Maclean, Oxford University Press, Oxford, 96-120.
3. Wahli, W., Dawid, I.B., Wyler, T., Jaggi, R.B., Weber, R. and Ryffel, G.U. (1979). *Cell* 16, 535-549.
4. Felber, B.K., Maurhofer, S., Jaggi, R.B., Wyler, T., Wahli, W., Ryffel, G.U. and Weber, R. (1980). *Eur. J. Biochem.* 105, 17-24.
5. Wahli, W. and Dawid, I.B. (1980). *Proc. Natl. Acad. Sci. USA* 77, 1437-1441.
6. Wahli, W., Dawid, I.B., Wyler, T., Weber, R. and Ryffel, G.U. (1980). *Cell* 20, 107-117.
7. Germond, J.-E., ten Heggeler, B., Schubiger, J.-L., Walker, P., Westley, B. and Wahli, W. (1983). *Nucleic Acids Res.* 11, 2979-2997.
8. Wahli, W., Wyler, T., Weber, R. and Ryffel, G.U. (1978a). *Eur. J. Biochem.* 86, 225-234.
9. Messing, J., Crea, R. and Seeburg, P.H. (1981). *Nucleic Acids Res.* 9, 309-321.
10. Barnes, W.M. and Bevan, M. (1983). *Nucleic Acids Res.* 11, 349-369.
11. Sahli, R., Mc Master, G.K. and Hirt, B. (1985). *Nucleic Acids Res.* 13, 3617-3633.
12. Hong, G.F. (1982). *J. Mol. Biol.* 158, 539-549.
13. Henikoff, S. (1984). *Gene* 28, 351-359.
14. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
15. Maxam, A.M. and Gilbert, W. (1980). *Methods in Enzymology* 65, 499-560.
16. Walker, P., Brown-Luedi, M., Germond, J.-E., Wahli, W., Meijlink, F.C.P.W., van het Schip, F.D., Roelink, H., Gruber, M. and AB, G. (1983). *EMBO J.* 2, 2271-2279.
17. Germond, J.-E., Walker, P., ten Heggeler, B., Brown-Luedi, M., de Bony, E. and Wahli, W. (1984). *Nucleic Acids Res.* 12, 8595-8609.
18. van het Schip, F.D., Samallo, J., Broos, J., Ophuis, J., Mojjet, M., Gruber, M. and AB, G. (1987). *J. Mol. Biol.*, in press.
19. Wahli, W., Wyler, T., Weber, R. and Ryffel, G.U. (1976). *Eur. J. Biochem.* 66, 457-465.
20. Martens, G.J.M. and Herbert, E. (1984). *Nature* 310, 251-254.
21. Hedén, L.-O., Höög, J.-O., Larsson, K., Lake, M., Lagerholm, E., Holmgren, A., Vallee, B.L., Jörnvall, H. and von Bahr-Lindström, H. (1986). *FEBS LETTRES* 194 (2), 327-332.
22. Lieman-Hurwitz, J., Dafni, N., Lavie, V. and Groner, Y. (1982). *Proc. Natl. Acad. Sci. USA* 79, 2808-2811.
23. Aho, S., Tate, V. and Boedtker, H. (1983). *Nucleic Acids Res.* 11, 5443-5450.
24. Tosi, M., Young, R.A., Hagenbüchle, O. and Schibler, U. (1981). *Nucleic Acids Res.* 9, 2313-2323.
25. McLauchlan, J., Gaffney, D., Whitton, J.L. and Clements, J.B. (1985). *Nucleic Acids Res.* 13, 1347-1368.
26. Zarkower, D. and Wickens, M. (1987). *EMBO J.* 6, 177-186.
27. Wahli, W., Ryffel, G.U., Wyler, T., Jaggi, R.B., Weber, R. and Dawid, I.B. (1978b). *Dev. Biol.* 67, 371-383.
28. Mount, S.M. (1982). *Nucleic Acids Res.* 10, 459-472.
29. Wahli, W., Dawid, I.B., Ryffel, G.U. and Weber, R. (1981). *Science* 212, 298-304.

30. Ryffel, G.U., Muellener, D.B., Gerber-Huber, S., Wyler, T. and Wahli, W. (1983). *Nucleic Acids Res.* 11, 7701-7716.
31. Schubiger, J.-L., Germond, J.-E., ten Heggeler, B. and Wahli, W. (1985). *J. Mol. Biol.* 186, 491-503.
32. Hummel, S., Meyerhof, W., Korge, E. and Knöchel, W. (1984). *Nucleic Acids Res.* 12, 4921-4938.
33. Bergink, E.W. and Wallace, R.A. (1974). *J. Biol. Chem.* 249, 2897-2903.
34. Wiley, H.S. and Wallace, R.A. (1981). *J. Biol. Chem.* 256, 8626-8634.
35. von Heijne, G. (1984). *J. Mol. Biol.* 173, 243-251.
36. Wiley, H.S. (1979). Ph.D. Thesis, The University of Tennessee, Knoxville.
37. Godson, G.N., Banel, B.G., Staden, R. and Fiddes, Y.C. (1978). *Nature* 276, 236-247.
38. Luo, C.-C., Li, W.-H., Moore, M.N. and Chan, L. (1986). *J. Mol. Biol.* 187, 325-340.
39. Knott, T.J., Rall, Jr., S.C., Innerarity, T.L., Jacobson, S.F., Urdea, M.S., Levy-Wilson, B., Powell, L.M., Pease, R.J., Eddy, R., Nakai, H., Byers, M., Priestley, L.M., Robertson, E., Rall, L.B., Betsholtz, C., Shows, T.B., Mahley, R.W. and Scott, J. (1985). *Science* 230, 37-43.
40. Salisbury, N., Calabrice, N. and Triplett, E.L. (1980). *Cell Differentiation* 9, 219-227.
41. Laskowski, Jr., M., Kato, I., Leary, T.R., Schrode, J., and Sealock, R.W. (1974). *Bayer-Symposium V "Proteinase Inhibitors"*, 597-611.
42. Doolittle, R.F. (1985). *Scientific American*, 253, 74-83.
43. Engel, J. and Furthmayr, H. (1986). *Methods in Enzymology*, in press.
44. Fung, M.R., Hay, C.W. and Mac Gillivray, R.T.A. (1985). *Proc. Natl. Acad. Sci. USA* 82, 3591-3595.
45. Chou, P.Y. and Fasman, G.D. (1974). *Biochemistry* 13, 222-244.
46. Garnier, J., Osguthorpe, D.J. and Robson, B. (1978). *J. Mol. Biol.* 120, 97-120.

APPENDIX

Nucleotide sequence of the vitellogenin gene A2 and prediction of the amino acid sequence of the protein.

Nucleotide sequences of the exons are given in capitals and are overled by the amino acid sequence deduced from them; gene flanking and intron sequences are given in small letters. The transcription initiation site (Exon 1, +1) and the poly(A)-site (last capital underlined in exon 35) are indicated; TATA-box, poly(A)-signals, as well as YGTGTTY elements in the 3'- end region are underlined. Stars indicate nucleotide differences between cDNA and genomic DNA (see also table II). The REM I element (32) and the Vi-element (31) in intron 19 and 27, respectively, are shaded. The first and last nucleotide contained in cDNA 18 (▽ ▲) and in cDNA 34 (▼ ▲), respectively, are indicated.

gaattctgagttttctggataacggatttccggatattggatccctacctgtatttccccctgtgattttgtgggtttt	-950
ctccagggaagtctgttttccctgacacactccaaaagtatacagacaggtgacatggttcccttataaagaatgacaaaaac	-870
gtgtgtgtgaattgtgataaggagcttagattgaagcttcgcccagagacaagaactgatgtaaatgatataataatctgt	-790
aaagcactgcagaatgtgttaggcctgtataaagattgatctgaacaacatgctttgtcagattcttaacaactgt	-710
gtctgtcacattcaaaagtatttagtaccccaacccattttatattatataacctactgccagagatgcctctgaacctg	-630
aaagtctcttactatgtaggcatgttctatctgggaaaaataaatagggttatctttttatttgaacattttatatagtc	-550
ttattatagctattattttatcttaagattaaaaaagctgtcaattctcttattcttattgtattttaaattcatttttag	-470
agaatgaaaaataagtgattcatgttgaaataaaaaacaaacttttttggacatcaaaatgtgtttttactgggactt	-390
ttgttcataaaaatagttttctgcatagcctcttttgattagattaactgtccaaagtcaggtcacagtgacctgatcaaa	-310
ttaatgtaacctcaacctggacaaagctactagtttgaaatcagaaataaataaataaataattttattagacat	-230
attacattttttttacattttatttggttaaaattcattgtcatataaataaagcaaaaattcacatcaaatataaatct	-150
gtttatgctgtttgatgcgaatatagattggtgaggttaattgtttacacaacctgataaacagtaactggaatgagccaaaa	-70
Exon 1 +1	
tcacacatctaagatcacgtgttacagattttctctgcaataaaatagggcgttttctgggttcagtgTTCACCATCA	10
MetLysGlyIleValLeuAlaLeuLeuLeuAlaLeuAlaG	
CCATGAAGGGAACTCGCTTAGCATTGTCGCAATTAGCGGGtaagtacagaaagtcgagagccgagattacacgccat	13
cagaactctgccaactttgaaacctcaaggataaaaactctcctgagctctaccatatagacatgccttgatttaaatat	90
aatagcaaatgcgaatgagcactgctttgataatatagttttcttcagcagaaactttcattatgtcacattttgcttt	170
	250
Exon 2 lySerGluArgThrHisIleG	
ttctttccttccagGCTCTGAAAGAACTCATAGGtaagtgtctgttttagttaccaaatgcaatttttaattctgca	20
	330
Exon 3 luProValPheSerGluSerLysIleSerValTyrAsnTyrGlu	
tgtgtactgaccattctgtgcatttgtctatttttagAGCCTGTGTTTCAGTGAAGCAAGATATCTGTGTATAACTATGAA	35
	410
AlaValIleLeuAsnGlyPheProGluSerGlyLeuSerArgAlaGlyIleLysIleAsnCysLysValGluIleSerAl	
GCTGTCTACTGAATGGATTTCCTGAAAGTGGTTGTCCCGGGTGGTATTAATAAATACTGCAAGTTGAGATCAGCGC	62
	490
aTyrAlaGlnArgSerTyrPheLeuLys	
CTATGCTCAGAGGTCATACTTCCTAAAGGtaagtaccatttggccctatttttaaaaaagaaaaaaaccacatggtc	71
accctcaattagtgatgtaaacacttaaaattgattagctgttcgttattcatatgatagcaccataaatctgaaacat	570
acaaaaagggcagtcgagcacagctacaatgtatgaagccgatatttttaatttcagtttaacatgaaatctaggaactgg	650
gcacatatacagtaattgtttatatttaatacagcatttctatatgatcatactctttttatttaaaaggaaaaatacatt	730
atttttaaaaaataatgtgggaatgcaaacatcaaccatgaatgcactccctttgtacatagctcatatatttaaaaaatgg	810
gcacatacaaaagcatgaagtgaagcaagcaagtagctatacgtattacctttacaatgtaattagaacaactttgtttaa	890
gttcagtttaatacgtgcgcaccgattgggactgttactgttaccagtgctcagtgatgataataacataaataatgattt	970
tgttgcatatgcgaatgttttaggcagccaatagtttaattgtgtttgtgtttgaagtagaacataattgtctggg	1050
tttttttgagataataactactgttttagggcatatttcacctaaattccccaccaaaattacatgatacaataacat	1130
	1210
Exon 4 IleGlnSerProGluIleLysGluTyrAsnGlyValTrpProLysAspProPheThrA	
gatcatatttatattctccttttagATCCAGTCCCCAGAAATAAAGAGGTACAATGGTGTCTGGCCCCAAGACCCGTTCACTC	90
	1290
rgSerSerLysLeuThrGlnAlaLeuAlaGluGlnLeuThrLysProAlaArgPheGluTyrSerAsnGlyArgValGly	
GATCCTCAAGAGTCACTCAAGCTCTTGCAGAACAGCTCACCAAAACAGCAAGGTTGAATACAGCAATGGCGAGTTGGA	117
	1370
AspIlePheValAlaAspAspValSerAspThrValAlaAsnIleTyrArgGlyIleLeuAsnLeuLeuGlnValThrIl	
GACATTTTGTGTGTATGACGTTTCTGATCTGTCGCAAAATATCTATAGAGGAATTCTGAATCTGCTGCAGGTCAACAT	144
	1450
eLysLysSerGlnAspValTyrAspLeuGlnGlu	
AAAGAAATCCCAGGATGTGTATGACCTTCAAGAGGttagtgtaaacatctaaatttgtaaatgatattactgaaattctg	155
cactcctgctgctagccagaaaacctgtaaaaacacaaattaaaaatgaaatttactcttcatctatgctctgcttcaaca	1530
	1610
Exon 5 SerSerValGlyGlyIleCysHisThrArgTyrValIleGl	
agaatccttacagttgtgtattgtactctgttccacagTCTAGTGTGGTGGCATCTGTACACACAGATATGTGATTCA	169
	1690
nGluAspLysArgGlyAspGlnIleArgIleIleLysSerThrAspPheAsnAsnCysGlnAspLysValSerLysThrI	
AGAAGACAAGAGAGAGACCAGATTAGAATAATCAAGCTACAGACTTCAATAATTGTCAAGACAAAGTGTCAAAACCA	195
	1770
leGlyLeuGluLeuAlaGluPheCysHisSerCysLysGln	
TTGGTCTTGAGTTGGCGCAATTCTGCCACTCTTGTAACAGgtatccaatttttcaaagacttttgctaattttcacaag	209
	1850
Exon 6 LeuAsnArgValIleGlnGlyAlaAl	
aacacctaaaaaacattctaaaaatgcaatgcataataattttctgtgcagCTGAATAGAGTCATACAGGGGCTGC	218
	1930
aThrTyrThrTyrLysLeuLysGlyArgAspGlnGlyThrValIleMetGluValThrAlaArgGlnValLeuGlnValT	
AACCTACACATATAAATGAAGGCGAGATCAGGCTACTGTTATCATGGAAGTCACAGCTCGACAGGTCTCCCAAGTTA	244
	2010

4755

euAlaLeuLysAlaLeuGlyAsnAlaGlyGlnProGluSerIleLysArgIleGlnLysPheLeuProGlyPheSerSer	505
TGGCTCTGAAAGCTCTAGTGAATGCTGGACGCTGAAAGTATTAAGCGGATCCAGAAATCTCTGCCCGGATTTCTCTCC	5370
SerAlaAspGlnLeuProValArgIleGlnThrAspAlaValMetAlaLeuArgAsnIleAlaLysGluAspProArgLy	532
AGTGTGTATCAACTTCTGTGAGAAATTCAGACTGATGCTGTGATGGCCCTGAGGAACATAGCCAAGGAGACCCGCGTAA	5450
s	
Agtaagatactctattctattcactttacagatacacgcttttagttttactttattaaagtagatttagaaaccttccccaa	5530
tgtttcttactttttgatccacactaaaaagatttatgagtgggtatcataggaataaacatttgtatgttgagaaat	5610
Exon 12 ValGlnGluIleLeuLeuG	538
gaagtagattttgaaaccttctcctaattgttataaaaccacatttgtgaaaacttattttcagGTTCAAGAAATCCTTCTTC	5690
lnIlePheMetAspArgAspValArgThrGluValArgMetMetAlaCysLeuAlaLeuPheGluThrArgProGlyLeu	565
AGATTTTCATGGACCGCATGTTGCGACTGAAGTTCGATATGATGGCATGTTTGGCTTTGTTGAGACTAGGCCCTGGTCTG	5770
AlaThrValThrAlaIleAlaAsnValAlaAlaArgGluSerLysThrAsnLeuGlnLeuAlaSerPheThrPheSerGl	592
GCAACAGTGACTGCCATTGCAAAATGTGGCAGCAAGAGAGACAAACCTTCAGCTTGCAAGCTTCACCTTCTCCCA	5850
nMetLysAlaLeuSerLysSerSerValProHisLeuGluProLe	607
AATGAAGGCATATCAAAAAGCAGTGTACCTCACCTGGAACCACTGtaagtacaaagcaacaatgaaagcagttttattgg	5930
taggcatttttagtgcaatttttcatctccatcatcatttttccaagttttagttcttcataatttccaataacatgtcagt	6010
gaatgtctcaaaatgacttaattgttataaatgaagccctgattcttgcagttattccatgggagttggtatttgggctcagat	6090
taactttttgttttctaataattaagaatggaaatgtaaaatgatttttttaacctacatgctttatatcataatatata	6170
Exon 13 uAlaAlaAlaCysSerValAlaLeuLysIleLeuAsnProSerLeuA	622
cagtagtataatttagaattttcttgcaatttagGGGGCGGCATGCAGTGTGCTCTTAAGATTTTAAATCCAAGCCTTG	6250
spAsnLeuGlyTyrArgTyrSerLysValMetArgValAspThrPheLysT	639
ACAACCTGGGATACCGATACAGCAAAGTAATGCGGTAGATACTTTCAAATGtaagtaaatgctaattgttttgattgc	6330
Exon 14	
ataatcagtagttataaaaaatagccctgtattgcgacacatgactgcacccttaatagacaaagtttttctgtttcttc	6410
yrAsnLeuMetAlaGlyAlaAlaAlaLysValPheIleMetAsnSerAlaAsnThrMetPheProValPheIleLeuA	665
agACAATCTAATGGCTGGTGTGCTGCTAAAGTCTTCATTATGAACAGTGCCAATACAATGTTCACAGTTTTCATTCTCG	6490
laLysPheArgGluTyrThrSerLeuValGluAsnAspAspIleGlu	681
CTAAATTCGAGAGTACAGCTCTTTAGTAGAGAATGACGACATAGAGgtaaactttatgctgcttttattgctaccttca	6670
Exon 15	
aacagatgaacagaatgattcattgaatatctgtgtcagacatacaacatttcactttattttatatattttttttgttcacc	6650
IleGlyIleArgGlyGluGlyIleGluGluPheLeuArgLysGlnAsnIleGlnPheAlaAsnPheProMetArg	706
tctagATTGGAATAAGGGGAGAAGGAATAGAAGAATTTCTTAGAAAACAAAACATTTCAGTTTGCTAATTTCCCCATCGCC	6730
LysLysIleSerGlnIleValLysSer	715
AAGAAATCTCACAAATTTGTAATCCgttaagtatttacatttttgcccaatgatctgttttttattatgccaaaatgaat	6810
Exon 16 LeuLeuGlyPheLysGlyLeuProSerGlnValProLeuIleSerGly	731
ttggtgacctaaattcttattttcttccagCTGCTGGGATTAAAGGGCTGCCTTCTCAAGTGCCTCTGATATCCGGC	6890
TyrIleLysLeuPheGlyGlnGluIleAlaPheThrGluLeuAsnLysGluValIleGlnAsnThrIleGln	755
TATATCAAACCTTTTGGCCAAGAATTTGCCTTCACTGAACCTCAATAAAGAAGTCATTCAAAATACTATCCAGgttagttaa	6970
tgtaataatttatcttcagtatataacctgatttttttaattgtgttactttgttaataaacttacagtgaaatataca	7050
acttattttaccaagctttttacttattgtcatcacactgtctctttctattataatagcatttagttttcacattaa	7130
ataatttggtactgctttgttaattgggaacaatttgacccagtcacaaatgctggagcactaagggtgcccctttgtgt	7210
tcattcaagaaatataatattattattattcttaataaacaactggttaccctctttattaatgtattacaatgaaagtct	7290
Exon 17 AlaLeuAsnGlnProAlaGluArgHisThrMetIl	767
aaaactatgcatgctctgttttaaaaaacttttttttacacaccagGCATGAATCAGCCAGCAGAGACACATATGAT	7370
eArgAsnValLeuAsnLysLeuLeuAsnGlyValValGlyGlnTyrAlaArgArgTrpMetThrTrpGluTyrArgHisI	793
TAGAAATGTGCTGAACAAGCTTCTCAATGGAGTAGTAGGACAATATGCCCGACGGTGGATGACTTGGGAGTATCGACACA	7450
leIleProThrThrValGlyLeuProAlaGluLeuSerLeuTyrGlnSerAlaIleValHisAlaAlaValAsnS	818
TCATTCCAACGACTGTAGTTTGGCAGCTGAGTCAGTTTATACCAGTCCGCCATTGTCCATGCAGCTGTCAACgtaag	7530

4757

		△		
Exon 23	AspThrAsnGluThrAlaLeuTyrArgSerLysGlnLysLysLysAsnLysIleHisAsnArg	1112		
tttttcggtttccacagGACACAAATGAGACTGCACCTTTATCGCTCAAAACAAAAGAAAAACAAATACACACAGA		10250		
ArgLeuAspAlaGluValValGluAlaArgLysGlnGlnSerSerLeuSerSerSerSerSerSerSerSerSerSe		1139		
CGACTTGATGCTGAAGTGTGAGAAGCTAGAAAGCAACAATCGTCATTATCTAGTTCTTCATCATCTTCTTCATCATCTTC		10330		
rSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerProSerSerSerSerSerSerSerTyrSerLysArgS		1165		
ATCCTCTTCTTCGTCTTCTTCATCATCTTCATCTCATCTTCCCCATCATCTCTCTCGTCTCTTCATACAGTAAAGAT		10410		
		▼		
erLysArgArgGluHisAsnProHisHisGlnArgGluSerSerSerSerSerSerSerGlnGluGlnAsnLysLysArgAsn		1192		
CTAAAGGAGAGAGCATAATCTCATCATCAGAGGAGTCCAGTAGCAGTAGCAGTCAGGAGCAGAACAAAAAGGAAT		10490		
LeuGlnGluAsnArgLysHisGlyGlnLysGlyMetSerSerSerSerSerSerSerSerSerSerSerSerSerSe		1219		
TTCAGGAGAATAGAAAGCATGGCCAAAAGGGAATGAGTAGCAGCAGTAGCAGTAGCAGCAGTAGCAGCAGTAGTAG		10570		
rSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerGluGluAsnArgProHisLysAsnArgGlnHisAspAsnL		1245		
CAGCAGCAGTAGCAGCAGCAGCAGTAGCAGCAGTAGCAGTAGGAAAAATAGACCCACAGAAATAGACAGCATGATAATA		10650		
ysGlnAlaLysMetGlnSerAsnGlnHisGlnGlnLysLysAsnLysPheSerGluSerSerSerSerSerSerSerSer		1272		
AACAAGCAAAAATGCAGCTCTAATCAGCATCAACAGAAGAACAATTTTCTGAAAGCAGCAGTAGTAGCAGTAGC		10730		
SerSerSerGluMetTrpAsnLysLysLys		1282		
AGCAGCAGTGAAATGTGGAACAAGAAGAGtacagttaaaaaaacttctaccaataaacattttaccaaccaacttgta		10810		
			Exon 24	Hi
gttatgtgaatgtatgtctatctgatatactaaaccttccatgatgtttaatgaagcatttttatcccttcagCA		1283		
		10890		
sHisArgAsnPheTyrAspLeuAsnPheArgArgThrAlaArgThrLys		1299		
TCATCGAAACTTCTATGACCTCAATTTTCAGGAGAACAGCAAGGACAAAAGtgagtttatttaattagtgtctattgtcagt		10970		
taaactgaagaaaatcccatagttctacaagttaaactgaaaacccaccatgctaggagtgcaatgcagggttttcaaatt		11050		
agcatatgtagagcgtggaagtggaccactagaaaaatgcattatggttgaggagaatgtcctcattcatatgcaaacg		11130		
agaaggaaattctgggcagaaatcccaggtatagctcataaaatttccatttactggatcacgcttgacactgagctcaccc		11210		
acaagatgagccacaatatagggcaaaactggtttgcatttgggaatcttgatcatcatacagccattagctgcactta		11290		
aaaaatcaagtgggtgagttctctgtctaaagggtattacatgtaaatggaaaaatggttggaacaatcctgtgcttccctccc		11370		
agaattccttctgctttgcataagaaaaaggatgtctctattcaacaaaaatgcttttctagtgtgccactatcgtaac		11450		
actgaatcagacaaatttgcctcatttgggtgacctctccaaaaagcacatttgcctcatggatgggtatatttctgaccag		11530		
catttttaagcgagctgagaccctcttgaacctcctgtcatttgcctttaaattaaaattacctatcaccacaaagaaaaac		11610		
aaacctaccaaaagttagagaatgaagcaaatacctgattgggtgctactggcaactgcacttgtgtaaatataacacat		11690		
atgataataaatgagcacttgtgccgtgacctgatgtgtctattcttgcatataatagaagacctaatgcataaaaaag		11770		
acccttatttgcacaaaacccctgggcttcaagtaggttagaaaaataggccctctacgtgtaaccttccatactcttaaga		11850		
tgtaaaaataatggctaagactcagtgtagtctcagctttgttactgatgtgtgactataattaaaatgtagtgatata		11930		
			Exon 25	GlyThrGluHisArgGlySerArgLeuSerSerSerS
accattctcatttttaataccacattttaacaaatgctttcagGGTACTGAACACAGAGGGTACGTTTGTCTATCTTCTT		1311		
		12010		
erGluSerSerSerSerSerSerSerGluSerAlaTyrArgHisLys		1326		
CAGAAAGTTCTTCATCAAGCAGTGAAAGTGCTTACAGGCACAAAgtagtctttcatcgcaatgatatcagaatgtctca		12090		
			Exon 26	AlaLysPheLeuGlyAspLysGluPr
cttggttactttggggaaattcataactaaaaatgtttttctgtcctgtatttttagGCAAAATCTTGGGTGACAAAGAGCC		1335		
		12170		
oProValLeuValValThrPheLysAlaValArgAsnAspAsnThrLysGlnGlyTyrGlnMetValValTyrGlnGluT		1361		
CCCAGTGCTGGTAGTTACTTTTAAGGCTGTGCGGAATGATAACACCAAGCAAGGATACCGATGGTTGTCTATCAGGAGT		12250		
yrHisSerSerLysGlnGlnIleGlnAlaTyrValMetAspIleSerLysThrArgTrpAlaAlaCysPheAspAlaVal		1388		
ACCATTCTTCTAAGCAACAAATTCAGGCTTATGTTATGGACATATCCAAAAGTATAGTGGCAGCTTGTGTTGATGCTGTT		12330		
ValValAsnProHisGluAlaGln		1396		
GTAGTTAACCACATGAAGCTCAGGtaatgtctatatgacattaattcttattcattaataaacaataaaccaatttagg		12410		
			Exon 27	AlaSerLeuLysTrpGlyGlnAsnCysGlnAspTyrLysI
tgtgtacagtataaggttttactgtcatccatttatatagGCAAGTCTTAAATGGGGCCAGAATTGCCAAGATTACAAGA		1409		
		12490		
leAsnMetLysAlaGluThrGlyAsnPheGlyAsnGlnProAlaLeuArgValThrAlaAsnTrpProLysIleProSer		1436		
TCAATATGAAGGCTGAAACTGGTAATTTGGAAATCAACCAGCACTGAGGGTGACTGCTAACTGGCCTAAAATTCCTATCC		12570		

LysTrpLysSerThrGlyLysVa 1444
 AAATGGAAGTCAACTGGCAAGTgttaagtttataaaatgttcatactatgttcatactgtgcacccctagtaaaaaagata 12650
 aggggacacaaaatgttaactaacctatcaaacatagtcctgtaaggcttcaattaccagtaagtcctttgtgcaaaagcta 12730
 catttcagcaagaagaagcaaaacaaattgtattttgcaacttatttgaataatgtcctaatgtccattttgtttacataag 12810
 aagctgaaaaacacaaaaataatccattcacagtcctcaatgtcaggtcaaaattcataaatttttataaaatgggtgact 12890
 taaaccagccatattattttaaataagatactaccaatttaggttcaaattattagtatcaacacataggggacacattacc 12970
 tagctggagtgaaaggttcasatgcacaaacttcgaatttcgaagtgttttttggctacttccaccatcgactacgac 13050
 ttgcattcaaacgatttagactaaaactcattcgactattgacacattctttaaaaaaactttgactacacttttgc 13130
 aactaaaactcccaactcactgttagctatggggaaggtcccataggcttttccttgcttttttggctgaagaaa 13210
 aatggttcgattgatggattaaaaactttgaatcattcaattgggtaaatcctataacagtcgaatatcgaaggtt 13290

Exon 28
 aattaaactcgtatttcgacctatgttaatttgcacctaatgatataagattggatttgcttacagtgatctttttca 13370

lValGlyGluTyrValProGlyAlaMetTyrMetMetGlyPheGlnGlyGluTyrLysArgAsnSerGlnArgG 1468
 attcagGGTGGTGAATATGTCCTGGAGCAATGTATATGATGGGATTCCAAGGTGAATACAAAGGAACCTCTCAACGTC 13450

lnValLysLeuValPheAlaLeuSerSerProArgThrCysAspValValIleArgIleProArg 1490
 AAGTGAAACTTGTCTTTGGCCCTATCTTCACCCAGAACATGTGACGTTGTAATCCGCATACCCAGGtaattacataatat 13530
 ttgctatcaaatattatcatgttaagggctcgatctattaactggacagctatttatcattggattgtaaaatgaagaaca 13610

Exon 29 LeuThrValTyrTyrArgAlaLeuA 1498
 ttttaaagcaattgtgtttttattactttattatacttggtgttttaattgtagCTTACAGTGTACTATAGGGCCTTGA 13690

rgLeuProValProIleProValGlyHisHisAlaLysGluAsnValLeuGlnThrProThrTrpAsnIlePheAlaGlu 1525
 GACTTCAGTGCCAAATCTCTGTCGGACACCATGCAAAAGAAAATGTATTACAGACCCCAACATGGAATATATTGCTGAA 13770

AlaProLysLeuIleMetAspSerIleGlnG 1535
 GCTCCCAAAATTAATAATGGACTCCATACAAGgtgagccaagtgtattaccgatcattatgtattcattaatatagatca 13850

Exon 30 lyGluCysLysValAlaGlnAspGlnIleThrThrPheAsnGlyValAs 1552
 gtattttatcgtttttatgccttcatttcagGTGAATGCAAAGTAGCCCAAGATCAATCACAACCTTCAATGGTGTAGA 13930

pLeuAlaSerAlaLeuProGluAsnCysTyrAsnValLeuAlaGlnAspCysSerProGluMetLysPheMetValLeuM 1578
 TTTGGCTTCTGCTTTGGCTGAAAACGTCTACAATGTCCTGGCCCAAGACTGCAGCCCTAGATGAAATTTATGGTTTTGA 14010

etArgAsnSerLysGluSerProAsnHisLysAspIleAsnValLysLeuGlyGluTy 1598
 TGAGAACTCAAAGAGTCTCCTAATCACAAAGATATTAATGTCAAGCTTGGCGAATagtaagtgggaaataagctactgt 14090
 actgtacagtatagctactgaattgtatcttttaacacactgtgctagttaaattttacatttgacatttctggatgtgt 14170
 gtacaaaattttgtgtacataaaaaataaagtttatgtatatacatttagtcagcatttctattccagcttctacca 14250

Exon 31 rAspIleAspMetTyrTyrSerAlaAspA 1607
 atattttttttggctacagaatttaaacatgctttctaatccttaaacagTGATATAGATATGTACTACTCAGTGCAGC 14330

laPheLysMetLysIleAsnAsnLeuGluValSerGluGluHisLeuProTyrLysSerPheAsnT 1629
 CCTTTAAGATGAAGATCAACAACCTTAGAAGTATCAGAAGAATCTCCCATACAAATCATTTAACTgtaagagcattcat 14410

Exon 32 yr 1630
 ttattttgttcaactaaagaaagtggtaattattctcaaacagtttaaaatatcatgctctttatttccctgggttcagAT 14490

ProThrValGluIleLysLysLysGlyAsnGlyValSerLeuSerAlaSerGluTyrGlyIleAspSerLeuAspTyrAs 1657
 CCTACTGTTGAGATTAAGAAGAAAGCAATGGCGTGCATTATTCGCATCGGAATATGGCATTGATTCACTTGACTATGA 14570

pGlyLeuThrPheLys 1662
 TGGACTGACATTCAAGgtattttgtgattaagtaagtacagagtcagtagaatgtcatttagtgccaggatattctca 14650
 gccctaggatatagaaccactagctattaaattatattttttgttttgggggggtgaaaaatctacaattgtttgggtgt 14730
 ctgaagctcttgggtcatgatgcaaaatttttcccttaactttaaatccactgtgcttcaaataggagagaaaacctata 14810
 aataactgggtatagaacacatttggctcatattatctctcacaacattctaagagtgcaaacacacataaccttg 14890
 cattagtagtgcataatagctattcagattttatcacattgttttaaacatcattaaaacttttccacaggttatta 14970
 cacaattgcattacaatgattttgtgaaattatttaccactgcacattgggaataactagaaggattataaagaattat 15050
 tattattgttattgttagtattacaatttagaccattattttacaggtggaagacacaaagtcaaaattttacaaaatg 15130
 actgtccttttcttcagccaatatgcatagagagactttatatagtttcattataggtgacagtagtaatcactctac 15210

Exon 33 PheArgProThrIleTrpMetLysGlyLysThrCysGlyIleCysGlyHisAsnAspA 1681
 aagatccttttattcttttaccagtTCAGACCAACAATATGGATGAAAGGCAAGACTTGTGGAATCTGTGGACACAATGATG 15290

spGluSerGluLysGluLeuGlnMetProAspGlySerValAlaLysAspGlnMetArgPheIleHisSerTrpIleLeu 1708
 ATGAAAGTGAAGAAGAACTCAGATGCCAGATGTTCTGTGGCCAAAGATCAAATGAGATTTATTCATTTCATGGATTCTT 15370

ProAlaGluSerCysSerGluG 1715
 CCTGCAGAGTCTGCTCTGAAGgtattttatcaaattaaggtattatattatggattataattgaataaacggagcaccgaa 15450
 aaggacatccattagttaaaaatcacactttattcccaaacattaaaaacatctctgcatacaggacatttggccatcct 15530
 cgtttcattggcatttcgccacttctgacaaagttaaagaaagcattttgtaaagattttgtatacagacatttcagaaaaatg 15610

Exon 34 lyCysAsnL 1718
 caagaatactcacttgtaataatattttatttggccacatttttttttttacaataaacaactattttatttagGTTGCAACC 15690

euLysHisThrLeuValLysLeuGluLysAlaIleAlaThrAspGlyAlaLysAlaLysCysTyrSerValGlnProVal 1745
 TCAAAACACACTCTTGTGAAACTTGAGAAAGCAATTGCAACTGATGGAGCAAAGGCCAAGTGCATTTCAGTTCACGCCTGTT 15770

LeuArgCysAlaLysGlyCysSerProValLysThrValGluValSerThrGlyPheHisCysLeuProSerA 1769
 TTGCGTTGTGCCAAAGGATGCTCTCCAGTTAAGACTGTTGAAGTGCTACTGGTTTCCACTGTCTGCCCTCTGtgtagta 15850
 tgtatgttgttattgatcacatgaattgatgcaactgtcagtttattacaagcaacttatgttacgggcatagctgaga 15930
 acaatagttattagacttttataataaactgaccctaactaactttttctagcactgcctatctgaatgtttatttatggt 16010
 acagccatgctttgattctacgtgggttttggcagtagttaagtgagccaactcttaataatagaatacaagatactaaagaat 16090
 attgtgggtgttaattatacattgtgaattatacacatacacatacaagtgtaaaacttgccatattattattggaccaggc 16170
 agcagggtggggtctgtgtggaatacagcagaagctgttcaaaatcatccagagtttagaccataatagtaaaatagcatag 16250
 gaagcaaaactgcctttttgtataaaggaactgtttctgtctcaaacatttcagtgagtagatgatgtttattttgccatc 16330
 tgtaaagaggttaattgttccataaattactttacaaggactcttcaataccaggctataagacatacagtagtaagattat 16410

Exon 35 spValSerLeuAspLeuProGluGlyGlnIleArgLeuGluLysSerGluAspPheSe 1789
 aatgtatttttttaattacagACGTATCTCTGGATCTTCTGAAGGACAAATAAGACTGGAAAAGTCTGAAGATTCTC 16490

rGluLysValGluAlaHisThrAlaCysSerCysGluThrSerProCysAlaAla*** 1807
 TGAAGAGTTGAAGCCACACCGCTTGTTCATGTGAGACATCTCCATGTGCAGCATAGATCAAGCAATCTTCTACCTTA 16570
 TTTTCTGGCCTGATAATTAAGTCTTCAGATTTCATGAAGATAATTAATGTGTAACCTATAACCAACAATAACAAATGCAT 16650
 GAATTATGTTTGAGAATTAACCTTTATATAATCTGTATATTGTAATACATTTCGAATAAAACTGTGCATTCAAATAGAA 16730
 tgtttgacctgttctgtttctttaaataattgatgttgtacattctgtttagaacatttaatttcacgcactgtgacactc 16810
 ctatatacagcagtcacagccattagtttacaactacagacaactctgagatattcctatagtttttgataatttactcoat 16890
 caaattcggttcttctgaaacctctctggggtttacaaaactgtggagcagtagatactgctttaaatacataattccagtt 16970
 tagaaaatgatttactgacatactgcagcttctcctctcctaccttagagtaaatggtcagtggtccgcctcatttccatct 17050
 ttatttcttcaagtagaaaactaactaaataatatttttaaatgatattaattatattcctctgcctgtttcttttac 17130
 aataaaaaatgtctcctttgaaagacatctaggtcttgaacctaaagtaatttccactacatctatgccagttttgcacaa 17210
 gcatttttgcaaatattttaacatgtagaaactaaaaattgactcaaacacggataaagggttttcaagcatttcaataaat 17290
 aatggcagtagaaagtgcaccaaattcatatggtataacatgatggcacagaagtttattagaagttgcctatgaacttg 17370
 tcacactgagcagagacttgtagttttataataatacttctgtttatttgtatactggtaatttaatacaagttatgtggtc 17450
 agggaattc 17459