The EMBO Journal vol.3 no.10 pp.2333–2339, 1984

Relaxin gene expression in human ovaries and the predicted structure of a human preprorelaxin by analysis of cDNA clones

Peter Hudson[1], Marie John, Robert Crawford, Jim Haralambidis, Denis Scanlon, Jeffrey Gorman, Geoffrey Tregear, John Shine[2,3] and Hugh Niall

Howard Florey Institute of Experimental Physiology and Medicine, University of Melbourne, Parkville, Victoria, Australia 3052 and [2]Centre for Recombinant DNA Research, Australian National University, Canberra, ACT, Australia 2601

[1]Present address: CSIRO Division of Protein Chemistry, 343 Royal Parade, Parkville, Victoria, Australia 3052
[3]Present address: California Biotechnology Inc., P.O.Box 51650, Palo Alto, CA 94303, USA

Communicated by K.Beyreuther

In earlier studies we identified in a human genomic library a gene (human relaxin gene H1) coding for a relaxin-related peptide. We now have evidence that the human genome possesses an additional relaxin-related gene (designated human relaxin gene H2) which appears to be selectively expressed in the ovary during pregnancy. Nucleotide sequence analysis revealed striking differences in the predicted structures of relaxin encoded by these two genes. Chemical synthesis of biologically active relaxin based on the sequence obtained from ovarian cDNA clones confirmed that the expressed gene (H2) encodes an authentic human relaxin. The expressed gene appears to be transcribed into two different sized mRNAs and preliminary evidence suggests that the mRNA transcripts possess different 3'-untranslated regions. There was no evidence for the expression of human relaxin gene H1 in the ovary and so far it is unclear whether gene H1 is expressed in another tissue or whether it represents a pseudogene. From the sequence data presented here it will now be possible to construct oligonucleotide probes and raise antibodies against synthetic peptides which could then be used to identify sites of relaxin biosynthesis and specifically quantitate the expression from either the H1 or H2 relaxin genes.
Key words: relaxin/DNA sequence/cDNA/human ovaries/gene

Introduction

Relaxin is a peptide hormone which is responsible for remodelling the mammalian reproductive tract before parturition thus facilitating delivery (Hisaw, 1926; Schwabe *et al.*, 1977; James *et al.*, 1977). Elucidation of the amino acid sequence of human relaxin is therefore an initial step in the production of biologically active peptides suitable for physiological and clinical studies. The major site of its biosynthesis is the corpora lutea of ovaries during pregnancy in the rat (Anderson *et al.*, 1975; Hudson *et al.*, 1981a). From protein sequencing data on the purified ovarian hormone (Schwabe *et al.*, 1977; James *et al.*, 1977; John *et al.*, 1981) and nucleotide sequence analysis of cDNA clones (Hudson *et al.*, 1981a; Haley *et al.*, 1982) it appears that both porcine and rat relaxins are expressed as single chain peptide precursors with the overall structure: signal peptide/B chain/C peptide/A chain. No such data were hitherto available for human relaxin due to the scarcity of suitable ovarian tissue. However, we have recently isolated a genomic clone (λH7) which encoded a biologically active human relaxin (Hudson *et al.*, 1983). This gene, which we have designated human relaxin gene I, contains a 3.7-kb intron interrupting the coding region in the C peptide with splicing sites identified by homology to porcine and rat cDNA structures. Regions of λH7 which give specific cDNA probes for relaxin sequences have been used to analyse the genomic organization and expression of the human relaxin gene(s). Our data reported here suggest that two relaxin genes are present in the human genome and the accompanying paper by Crawford *et al.* describes the localization of both genes to chromosome 9. For convenience we have designated the gene structure of λH7 reported earlier as H1 relaxin (Hudson *et al.*, 1983) and the novel gene structure described herein as H2 relaxin.

Results

Gene organization

Southern gel analyses of human genomic DNA hybridized with separate probes corresponding to either exon I or II of the previously reported H1 gene (Hudson *et al.*, 1983) reveals that the human genome contains two relaxin genes. One of these genes possesses a restriction pattern identical to the full-length H1 genomic clone (λH7) reported earlier. The other restriction fragments which hybridize to the relaxin-specific probe correspond to a second gene. Because fragments corresponding to both genes were identified in all the individuals analysed it is probable that the two genes are non-allelic and arose by a gene duplication event (Crawford *et al.*, accompanying paper).

Isolation and sequence analysis of cDNA clones

Samples of human corpus luteum were made available as a result of surgical intervention in ectopic pregnancies or from luteectomy at the time of Caesarian section. From the RNA isolated from a single corpus luteum a cDNA library was constructed in pBR322, providing ~300 unique recombinants. A single cDNA recombinant was isolated using relaxin-specific probes corresponding to exon I and II of the H1 gene reported earlier (Hudson *et al.*, 1983). The nucleotide sequencing strategy on this cDNA clone is shown in Figure 1 and the sequence of the coding region, which differs from the H1 relaxin gene sequence, is presented in Figure 2. We expected that this novel sequence corresponded to the second human relaxin gene (H2) which had been observed in genomic DNA. To increase the total number of recombinants from such small amounts of ovarian tissue we constructed cDNA libraries using the λGT10 cloning system (Huynh *et al.*, personal communication). Twenty three unique cDNA recombinants in λGT10 were isolated using a relaxin-specific probe and of these six were characterized as shown in Figure 1. Nucleotide sequence analysis revealed that all six cDNA recombinants encoded fragments of the same H2 relaxin

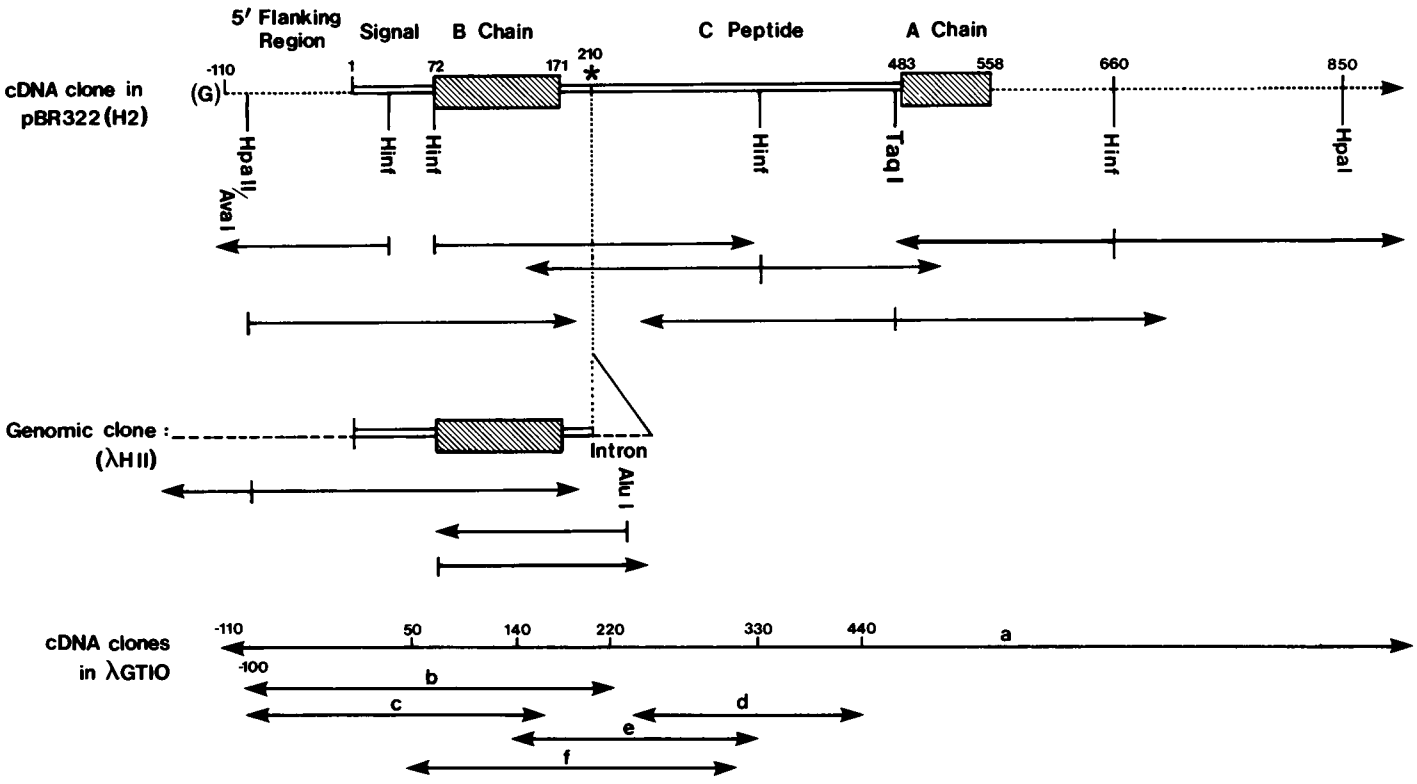

Fig. 1. An abbreviated restriction map and nucleotide sequencing strategy for the cDNA clone in pBR322, genomic clone λH11, and λGT10 cDNA clones a–f. Arrows indicate the direction of sequencing on end-labelled fragments (see Materials and methods). λGT10 clones a–f were sequenced by subcloning into an M13 vector as described in Materials and methods. Nucleotides are numbered from the AUG initiation codon, position 1–3, through the termination codon, position 556–558.

structural gene shown in Figure 2.

Surprisingly, none of the cDNA clones contained a poly-adenosine sequence at the 3′ end and 21 of the 23 λGT10 cDNA clones lacked sequences corresponding to the A chain and 3′-untranslated region. We attributed these truncated cDNA clones either to premature termination of the double-stranding transcription reaction or to excessive S1 nuclease degradation during the cloning procedure. However, the two largest cDNA clones in pBR322 and λGT10 (1800 bp and 1900 bp, respectively) indicate that long transcription products were being synthesized during the cloning procedure. These two large cDNA clones had identical sequences both in the relaxin-coding region and in their overlapping 5′- and 3′-untranslated regions, confirming that they were derived from the same mRNA structure.

A limitation in this work has been the availability of suitable human ovarian tissue. Attempts to construct larger cDNA libraries using modified protocols for the double stranding and S1 nuclease reactions will require further samples of tissue not yet available.

Isolation of a genomic clone corresponding to the second gene

A thorough screen of 10^7 recombinant phage from the human genomic library of Lawn *et al.* (1978) using mixed probes specific for exon I or II of the λH7 relaxin clone, revealed 16 positive phage. Detailed restriction mapping analysis revealed that 14 of these recombinant phage corresponded to the H1 relaxin gene reported earlier [11 were identical to the λH7 genomic clone; three were identical to λH5, a different genomic clone of the H1 gene previously reported (Hudson *et al.*, 1983)]. However, the other two recombinant phage were identical and had a unique restriction pattern characteristic of the H2 relaxin gene whose structure is given in Figure 1. The unexpected ratio of recombinants either may reflect their proportion in the original genomic library or may result from selective growth during amplification. Southern blot analyses of this new recombinant phage (λH11) using separate probes corresponding to either exon I or II of the λH7 clone, revealed that λH11 contained only the exon I coding region. Attempts to find a full length genomic clone corresponding to the H2 relaxin gene either in the library of Lawn *et al.* (1978) or in another library (R.Crawford, unpublished) have so far been unsuccessful.

The nucleotide sequence of the relaxin-coding region of λH11 was found to be identical to that observed in the cDNA clone shown in Figure 2. An intron interrupts the coding region in exactly the same position as in the λH7 genomic clone (Hudson *et al.*, 1983) suggesting that these genes arose by a gene duplication event.

Northern gel analysis

RNA was isolated from samples of human corpora lutea taken from different individuals during surgical intervention for ectopic pregnancy or during Caesarian section operations. Northern gel analysis using probes made from the coding region of either relaxin gene revealed that two major mRNA species of approximate sizes 1000 bp and 2000 bp were present in five human ovarian RNA samples tested (Figure 3a). The smaller mRNA species was estimated from the autoradiographs to be 2- to 3-fold more abundant in the RNA samples tested and this result was independent of whether the

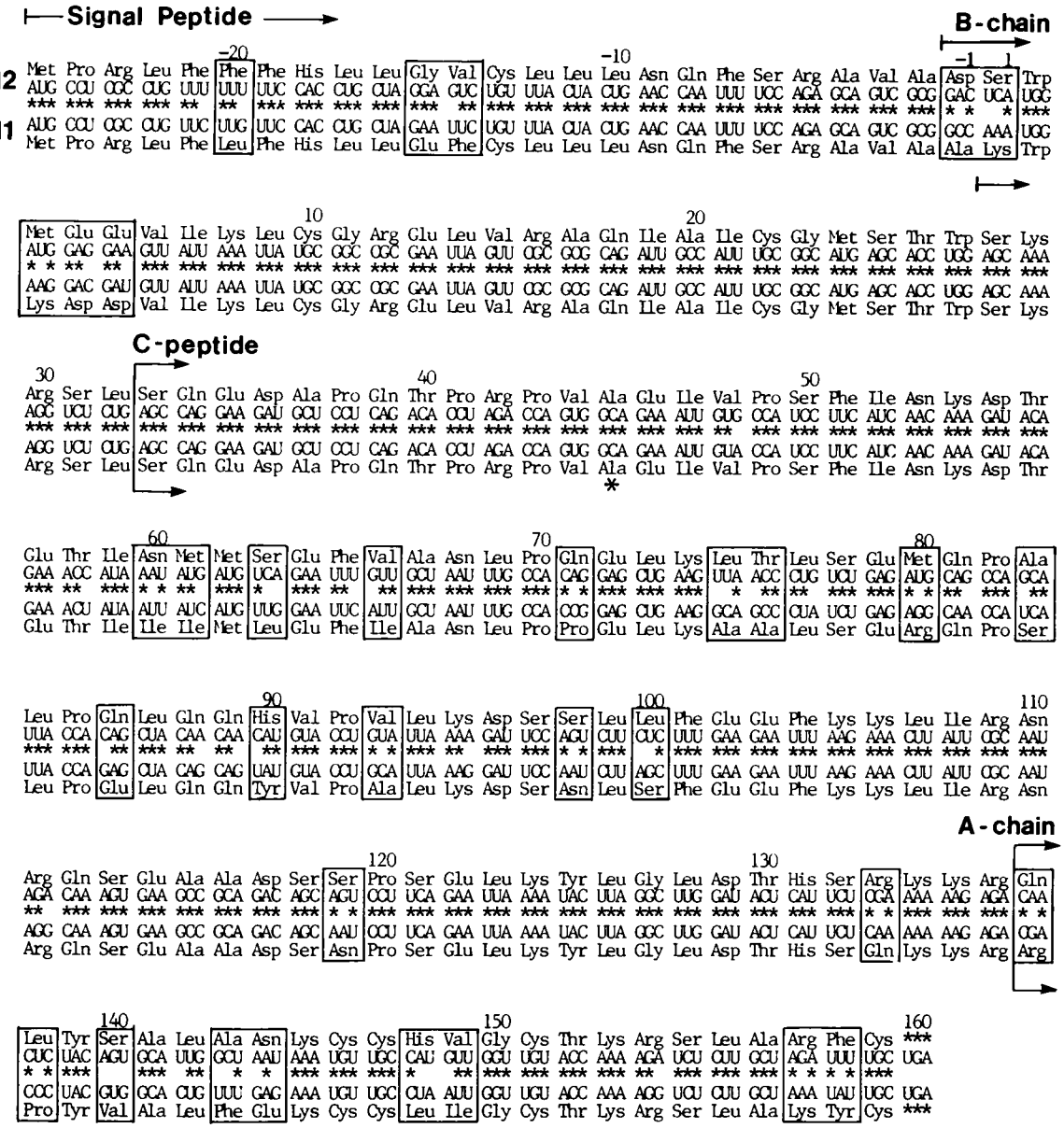

Fig. 2. This compares the amino acid and mRNA sequence of human preprorelaxin H2 (upper) with the corresponding H1 sequence. The sequences have been aligned to maximize homology with nucleotide identities being indicated by asterisks and amino acid differences by boxed-in areas. Amino acids are numbered from the start of the B-chain (H2 gene sequence starting at −1 and H1 sequence at +1) although this position represents only the hypothetical start of the B chain sequence and has been deduced simply from the homology to the related porcine and rat preprorelaxin structures. The asterisk beneath Ala 45 in the C peptide denotes the terminal residue coded by exon-I. The intron interrupts the adjacent G/AA codon in both genes.

probe used in the analysis corresponded to H1 or H2 relaxin indicating that high cross-hybridization rates occur under our experimental conditions. To differentiate whether these two mRNA species might represent the separate products of the H1 and H2 genes, oligonucleotide probes were synthesized over regions of minimum homology between the two relaxin genes. Two separate 25-mers complementary to either the H1 or H2 mRNA sequence were made in the region coding for residues 137−144 (60% homology between the two genes; Figure 2) and one 25-mer was synthesized complementary to the H1 coding region over residues 59−66 (76% homology to the H2 sequence; Figure 2).

These synthetic 25-mers were radiolabelled by kinasing with [γ-^{32}P]ATP and used as hybridization probes under conditions shown to provide specificity for either the H1 or H2 gene (Figure 3b). Northern gel analysis using these radiolabelled probes revealed that both mRNA species corresponded to products of the H2 gene. We could not detect any transcription products from the H1 gene using the specific probes, although low level expression (<5% of the H2 level) would have been difficult to identify (Figure 3a).

To analyse the different mRNA transcripts from the H2 gene, we hybridized Northern gel transfers with specific probes made from segments of the H2 cDNA clone corresponding to the coding region and 5′- and 3′-untranslated regions (Figure 4). The larger mRNA transcript (~2 kb in length) hybridized to all probes whereas the smaller mRNA transcript (~1 kb in length) did not hybridize to segments of the 3′ untranslated region from a position ~100 bases 3′ to the termination codon. This selective hybridization with probes from the 3′-untranslated region suggests that the 1800 bp and 1900 bp H2 cDNA clones correspond to the 2-kb mRNA

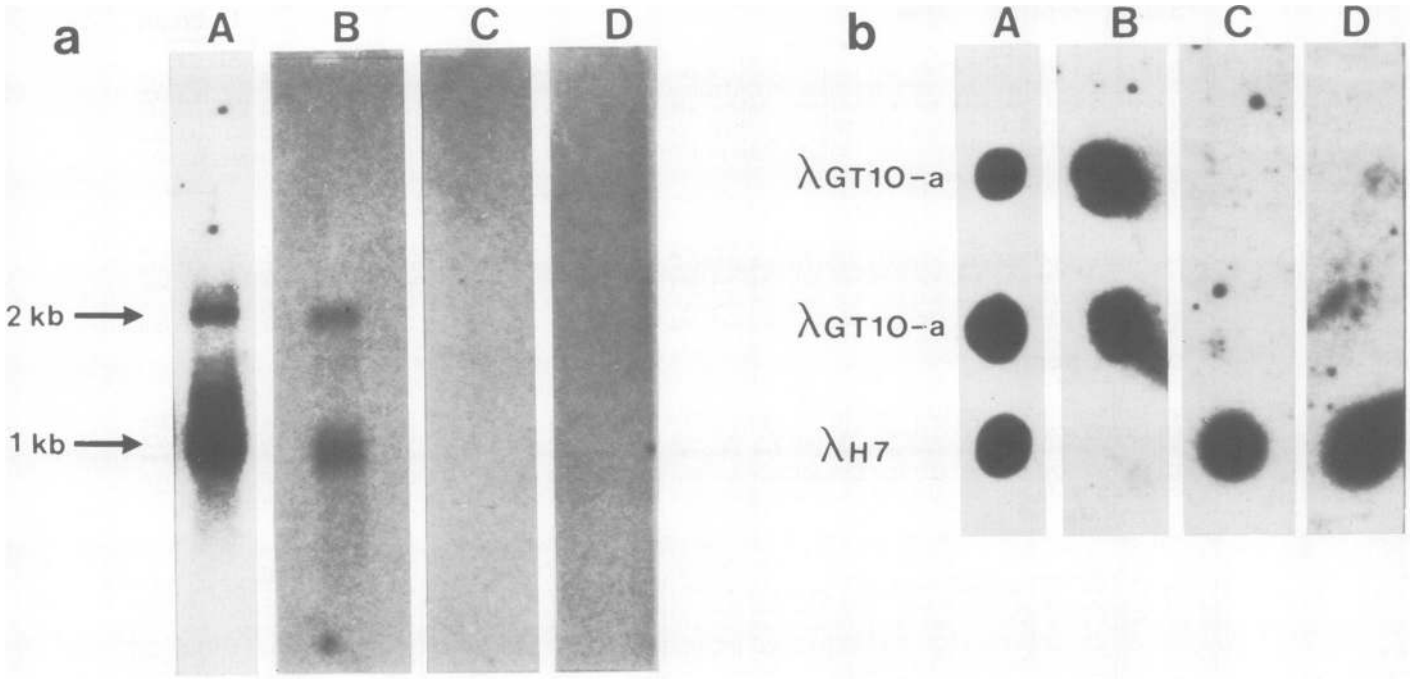

Fig. 3. Autoradiographs of identical nitrocellulose strips taken from either **(a)** Northern gel transfer of human ovarian RNA or **(b)** λ plaques corresponding to the H1 gene (λH7) or H2 gene (λGT10-a) using as hybridization probes **A**: a random primed 590-bp H2 relaxin cDNA fragment (nucleotides 72–660), **B**: H2-specific 25-mer (483–507), **C**: H1-specific 25-mer (483–507), **D**: H1-specific 25-mer (248–272).

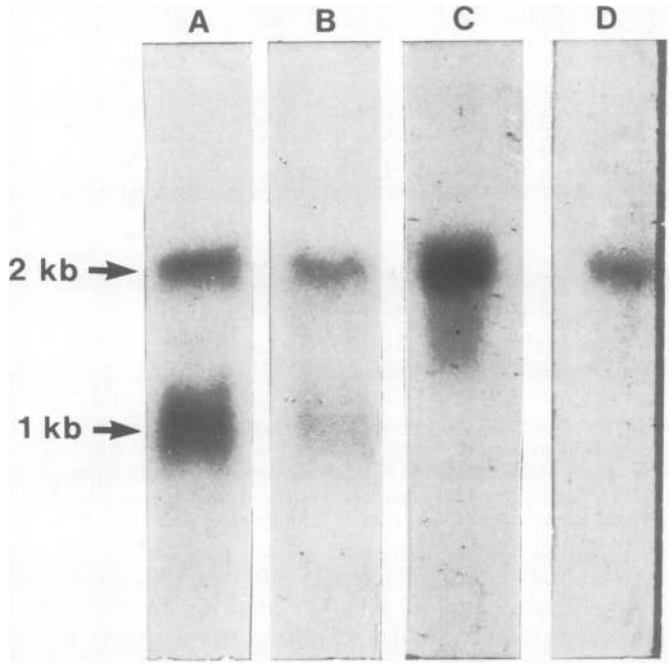

Fig. 4. Autoradiographs of identical nitrocellulose strips following Northern gel transfer of human ovarian RNA using as hybridization probes fragments of the H2 cDNA clone in pBR322 (see Figure 1). **A**: 590-bp fragment (nucleotides 72–660) corresponding to most of the coding region; **B**: 5′-untranslated region (to *Hinf*I site at nucleotide 30); **C**: 3′-untranslated region (from *Hinf*I site at nucleotide 660); **D**: 3′-untranslated region (from *Hpa*I site at nucleotide 850).

species. It is possible that the shorter 1-kb mRNA species could be simply a truncated form of the 2-kb mRNA species and could be generated, like the vimentin and β-microglobulin genes, by cleavage of the primary RNA transcript at alternative polyadenylation sites (Zehner and Paterson, 1983; Parnes *et al.*, 1983). The only evidence for such a mechanism is that a potential polyadenylation signal exists in the nucleotide sequence of the cDNA clones, 140 bases from the termination codon, and this region does have homology to the porcine relaxin polyadenylation site (Haley *et al.*, 1982). On the other hand we cannot rule out the possibility of alternative splicing mechanisms such as occurs in the calcitonin and growth hormone genes (Rosenfeld *et al.*, 1982; DeNoto *et al.*, 1981).

In the absence of the complete genomic sequence of the H2 gene or full length cDNA clones corresponding to both mRNA forms we are unable to define the mechanisms leading to the formation of the two mRNA transcripts *in vivo*. We intend to construct larger cDNA libraries for this purpose when further samples of human ovarian tissue become available.

The primary structure of preprorelaxin encoded by the H2 gene

The mode of *in vivo* processing of human preprorelaxin is not yet fully understood and has to be deduced by analogy to the processing of porcine and rat preprorelaxins (Figure 5). The predicted B and A chain structures for the H1 and H2 genes have been aligned with other members of the relaxin family and human insulin in Figure 5.

Cleavage of the signal peptide in H1 preprorelaxin has been predicted (Hudson *et al.*, 1983) to occur after a short side chain residue such as Ala −1, −2 or −4 or after Ser −6. Cleavage after Ala −1 is consistent with the homology to porcine preprorelaxin and human preproinsulin. Similarly by such analogy, cleavage of the H2 signal peptide probably occurs after Ala −2 although cleavage after Ala −4 or Ser −6 are other possibilities.

By analogy to rat and pig prorelaxins, cleavage at the B chain/C peptide junction would occur after Leu 32 in both H1 and H2 precursors. However, both human relaxin B

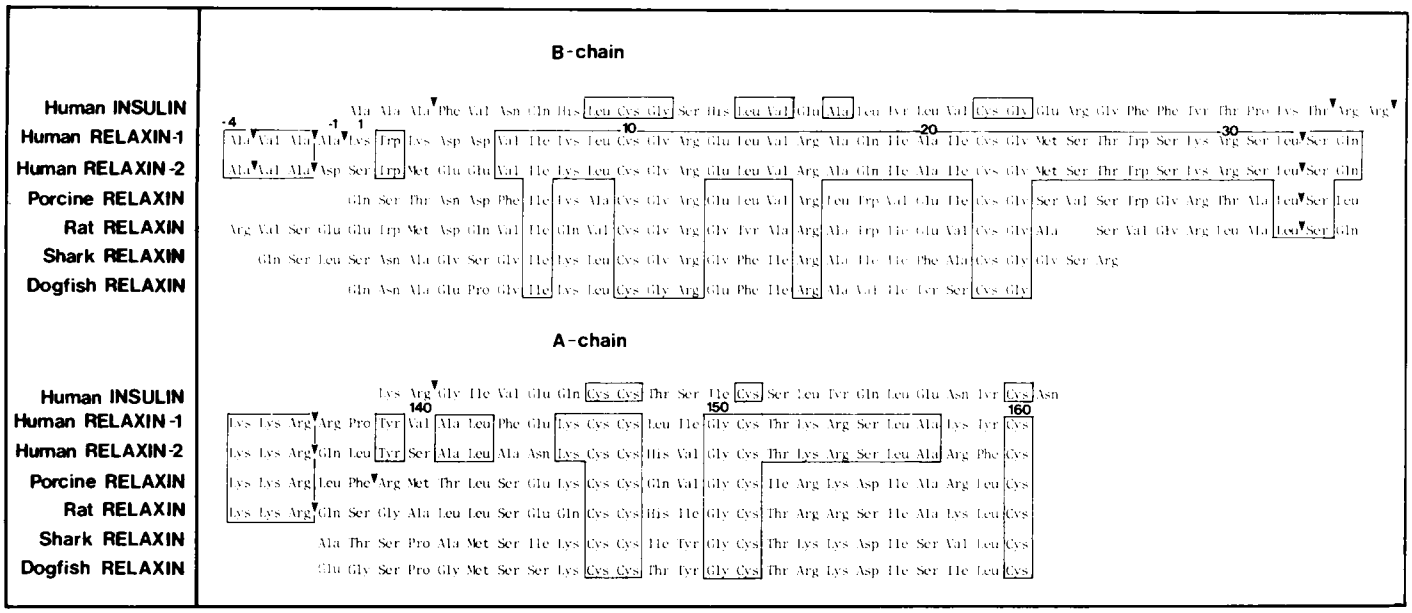

Fig. 5. Comparison of the amino acid sequences of the B and A chains between the two human relaxin genes, human insulin, and other members of the relaxin family. Boxed areas highlight residues which are conserved between the two human relaxin genes and the relaxin family. Arrows indicate probable sites of proteolytic cleavage with confirmation by protein sequencing data of the amino terminal residue of the B and A chains of porcine (Schwabe *et al.*, 1977; James *et al.*, 1977), rat (John *et al.*, 1981), shark (Schwabe *et al.*, 1982) and dogfish relaxins (Schwabe, 1983).

chains possess at positions 29 – 30 the conserved dibasic sequence Lys-Arg, which is a known processing site in other prohormones such as proinsulin, and cleavage here cannot be excluded. Direct amino acid sequence analysis of relaxin isolated from corpora lutea of pregnancy will be required to settle this point. In the meantime it seems that the most likely structure of the H1 B chain would be 32 residues in length (Lys 1 to Leu 32) and the H2 B chain would be 33 residues (Asp – 1 to Leu 32).

Cleavage at the C peptide/A chain junction of H1 prorelaxin has been predicted (Hudson *et al.*, 1983) to occur after Arg 136 within a group of four basic residues because the Arg-Pro imide bond at 137 – 138 would be resistant to proteolysis. H2 prorelaxin has the same sequence of four basic residues and a similar processing step after Arg 136 would result in both the H1 and H2 relaxin A chains being 24 residues in length.

Biological activity of the H2 gene

As noted in earlier studies with chemically synthesized pig relaxin peptides there is a core sequence of pig relaxin which contains all the essential elements for biological activity (Tregear *et al.*, 1981). Studies with synthetic peptides based on the H1 relaxin sequence have shown that the combination of a complete H1 A chain (Arg 137-Cys 160) with a shortened form of the H1 B chain (Lys 1-Ser 25) results in a peptide with relaxin-like biological activity in the rat uterine strip assay (Hudson *et al.*, 1983). Similar studies using shortened forms of the H2 relaxin B chain combined with the full length H2 relaxin A chain have also yielded biologically active synthetic relaxin preparations. The synthetic peptides comprising H2 A chain (Arg 137 to Cys 160) coupled to H2 B chain (Asp – 1 to Met 24) and H2 A chain (Arg 137 to Cys 160) coupled to H2 B chain (Asp – 1 to Leu 32) possessed biological activity in the rat uterine strip assay (Wiquist and Paul, 1958). The biological activity of the chain combination mixture comprising the A plus shortened B chain analogue was equivalent to that of the complete H2 relaxin sequence, both synthetic preparations being in the range of 3 – 6% active compared with a porcine relaxin standard. The separate H2 A and B chain peptides were completely devoid of biological activity.

Discussion

The two human relaxin genes show considerable nucleotide and amino acid homology to each other, particularly in the B chain and C peptide where a sequence of 120 nucleotides coding for Val 6 to Ile 47 is exactly conserved and where only 2 nucleotides have diverged in the 90 bases coding for Phe 101 to Ser 132. However, there are some notable regions of sequence divergence particularly in the amino-terminal region of both A and B chains. Earlier investigations (Tregear *et al.*, 1981; Hudson *et al.*, 1983) with synthetic peptides of both pig and human (H1) relaxin show that relaxin activity is present with A chains as short as A (141 – 160) and B chains as short as B (4 – 23). Thus the N-terminal sequence of both A and B chains appear to make little contribution to the biological activity, possibly explaining the apparently high tolerance to sequence divergence in this region throughout the relaxin family. In contrast the homologous insulin family retains a high degree of conservation in both the A and B chains (>90% between rat and human insulins). This striking difference between the insulin and relaxin families might be partly explained by the requirements for a conserved surface structure in insulin necessary for storage as the zinc-complexed hexameric form. Relaxin is stored in an amorphous form and this might reduce the constraints on surface residues.

In comparing the relaxin family (Figure 5) it is obvious that all six cysteine residues have been conserved to maintain the overall disulphide bond configuration. Glycines at positions 11, 23 in the B chain have been retained which supports the proposed similarity between the relaxins and insulin because these residues can have unusual torsional angles essential for the structure and which would be unstable if replaced by

another L-amino acid (Bedarkar *et al.*, 1977; Isaacs *et al.*, 1978). Conserved residues which have been predicted to be on the surface of the molecule (Dodson *et al.*, 1982) are Cys 10, Gly 11, Arg 12 and Arg 16 in the B chain and Cys 147, Gly 150, Lys/Arg 153, and Lys/Arg 154 in the A chain. There are several notable differences between the H1 and H2 A chain structures; such as Glu/Asn at 144 and Leu/His at 148; these might affect receptor binding and so could be the basis of different tissue specificity in the action of these two hormones.

From studies with synthetic peptides we have shown that both the H1 and H2 relaxin genes are capable of expressing peptides which possess relaxin-like biological activity. However, only mRNA transcripts from the H2 gene can be detected in human ovaries taken during early pregnancy or just prior to parturition. The function of the H1 gene is unclear and we cannot exclude the possibility that it is a pseudogene until H1-specific mRNA can be identified in human tissue. Some pseudogenes found in a number of human gene families, such as the tubulins and metallothioneins (Wilde *et al.*, 1982; Karin and Richards, 1982) have lost their intron sequences raising the possibility that they were generated by retroviral insertions. Other pseudogenes appear to have mutations in either the promoter or intron splicing regions which inhibit transcription or processing to a functional mRNA (Knoll *et al.*, 1981; Proudfoot and Maniatis, 1980). From sequencing data the H1 relaxin gene appears to have a normal coding region and intron splice junctions. The promoter regions of the H1 and H2 genes are currently being analyzed and compared in an attempt to clarify whether the H1 gene has a functional promoter, although gene transfer experiments using these promoters to drive transcription in intact cells may be necessary to establish the relative promoter efficiencies. At present, unless mRNA transcripts can be identified, the H1 relaxin gene may fall into the same category as one member of the human growth hormone gene family, i.e., a gene which appears to be functional as judged by the presence of promoters, polyadenylation signals, correct intron/exon splice junctions and absence of premature termination signals yet whose site of tissue specific expression remains unresolved (Seeburg, 1982).

Our results suggest H2 relaxin is involved in the physiology of pregnancy since it is synthesized in the ovary, but we cannot exclude a role for H1 relaxin which may be produced by other reproductive tissues such as decidua or placenta. Preliminary experiments have failed to detect any relaxin biosynthesis in either placenta or decidua using both Northern gel analysis and hybridization histochemistry (Hudson *et al.*, 1981b). This is surprising in view of previous reports of biologically active relaxin in placental and decidual extracts and may indicate that the small quantities of relaxin demonstrated in these studies by immunohistochemistry (Fields and Larkin, 1980) and by bioassay (Bigazzi *et al.*, 1980; Yamamoto *et al.*, 1981) represent corpora luteal relaxin bound to receptors.

Relaxin may have a possible therapeutic role in human obstetrics and there is a need for measurement of hormone levels in human pregnancy to understand the pattern of secretion in normal and abnormal circumstances. Specific radioimmunoassays could now be developed to measure H1 and H2 relaxin levels independently in the circulation during pregnancy. In view of the wide distribution of many peptide hormones it is possible that either or both the H1 and H2 relaxin genes are expressed in non-reproductive tissues, including brain and the gastrointestinal tract. Relaxins are members of the family of insulin-like growth factors and are capable of modifying connective tissue structure and influencing smooth muscle contraction. It is possible, therefore, that relaxin peptides expressed from these genes play an important physiological role in addition to their well documented hormonal function during reproduction.

Materials and methods

Messenger RNA isolation and cDNA cloning

Human ovarian tissue (corpora lutea) obtained during surgery for the treatment of an ectopic pregnancy or during Caesarian section was quickly frozen on dry ice and the RNA isolated in 5 M guanidinium thiocyanate (Merck) according to the method of Chirgwin *et al.* (1979). Poly(A)$^+$ RNA was converted into double-stranded DNA and cloned either by the homopolymeric G/C tailing method into a pBR322 plasmid vector (Chang *et al.*, 1978) or by the lambda packaging method using the λGT10 vector (Huynh *et al.*, personal communication). The efficiency of transformation with the pBR322 method (10^4 recombinants/μg of cDNA) was far less efficient than the lambda technque (up to 10^6 recombinants/μg of cDNA).

Selection of specific cDNA clones

To screen the human ovarian cDNA clone bank for relaxin-specific sequences we used as a probe a DNA fragment of the previously identified human H1 gene corresponding to a 400 nucleotide segment coding for the C peptide and A chain from amino acid 64, through the termination codon and including 80 bases of the 3′-untranslated region. Radiolabelled probes were prepared by primed synthesis using 100–200 ng of the DNA fragment and 1 μg of denatured random primers of calf thymus DNA (Hudson *et al.*, 1983; Taylor *et al.*, 1976). A single positive cDNA clone from the pBR322 library was isolated using the screening method of Grunstein and Wallis (1979). Twenty three unique recombinants were isolated from the λGT10 libraries, but of these only six were subjected to nucleotide sequence analysis.

DNA sequence analysis

The sequencing strategy and an abbreviated restriction map of the cDNA clone are summarized in Figure 1. The recombinant plasmid pBR322 was digested with restriction enzymes *Hpa*II (P), *Hinf*I (F) or *Taq*I (T) and end-labelled using reverse transcriptase and the appropriate α-labelled deoxynucleotide triphosphate (dCTP for *Hpa*II, and *Taq*I, dATP for *Hinf*I). Fragments were cleaved internally with a second restriction endonuclease and then separated by electrophoresis on 8% polyacrylamide gels prior to sequencing by the chemical degradation method of Maxam and Gilbert (1977).

cDNA clones in λGT10 were sequenced by subcloning *Eco*RI restriction fragments into M13mp9 and employing the techniques described by Sanger *et al.* (1980).

Southern and Northern gel analyses

These were performed on purified genomic DNA after restriction endonuclease cleavage by the method of Southern (1975), or on purified RNA. The DNA fragments which were used as probes were found to be specific for either exon I or exon II of the H1 genomic clone despite having a small amount of flanking sequences. These fragments were generated by subcloning into M13mp8 a 500-bp *Alu*I fragment of the λH7 clone in the case of the exon I probe, or a 400-bp *Eco*RI-*Ava*II fragment for the exon II probe. A probe from the H2 cDNA clone was generated by digesting with *Hinf*I and isolating a 300-bp doublet corresponding to the coding region from Asp 1 to the termination codon and including 110 bases of the 3′-untranslated region (Figure 1). Oligonucleotide probes were synthesized by the phosphite chemistry method of Beaucage and Caruthers (1981) and were end-labelled with [γ-^{32}P]-ATP using T4 polynucleotide kinase. Hybridization conditions were calculated on the basis of the nucleotide composition (Smith, 1982).

Isolation and nucleotide sequence analysis of the H2 genomic clone

The human genomic lambda library of Lawn *et al.* (1978) was screened by methods described earlier (Hudson *et al.*, 1983) except that a mixture of DNA fragments corresponding to exons I and II of the H1 genomic clone was used for the probe as described above. Positive phage were grown in litre scale liquid cultures, the DNA isolated and digested with restriction endonucleases prior to mapping with the exon I and II probes. A 4-kb *Eco*RI fragment was found to contain the entire exon I coding region which differentiated this clone from the homologous H1 gene structure. This fragment was subcloned into M13mp8 and sequenced by the technique of Maxam and Gilbert (1977). After digesting with *Ava*I, fragments spanning the coding region were end-labelled and cleaved internally by a second restriction enzyme (*Hpa*II or *Hinf*I) to generate fragments suitable for sequence analysis.

Chemical synthesis of H2 relaxin peptides

Synthetic peptides were prepared by the step-wise solid-phase procedure using phenylacetamidomethyl-polystyrene resins (Kent and Merrifield, 1981). The A and B chains were prepared separately and purified as their S-sulfonated cysteine derivatives using gel filtration ion-exchange chromatography and preparative h.p.l.c. Purification of the synthetic peptides was monitored by amino acid analysis and analytical h.p.l.c. Chain combination was effected by reduction and oxidation of a mixture of the S-sulfonated A and B chain peptides in a single step process using dithiothreitol and air-oxidation (Chance *et al.*, 1981). Biological activity was assessed by the rat uterine contractility assay (Wiquist and Paul, 1958).

Acknowledgements

We thank John Haley, Peter Roche and Peter Aldred for valuable suggestions on the DNA sequencing and λGT10 cloning strategies and Anne Casey for technical assistance. Rat uterine contractility assays were performed by Ms. S.Vienet and Dr. B.Kemp. Computer programmes were written by Dr. Roger Staden and modified by Dr. Tony Kyne. Human ovarian tissue was generously collected by Dr. Alastair Maclennan, Mr. John Walsh and Dr. Robin Bell. The project was carried out under NIH/ASCORD containment guidelines and supported by grants from the National Health and Medical Research Council of Australia, the Myer Family Trusts, the Ian Potter Foundation, the Howard Florey Biomedical Foundation and by NIH grant HD11908.

References

Anderson,M.L., Long,J.A. and Hayashida,T. (1975) *Biol. Reprod.*, **13**, 499-504.

Beaucage,S.L. and Caruthers,M.H. (1981) *Tetrahedron Lett.*, **22**, 1859-1862.

Bedarkar,S., Turnell,W.G., Blundell,T.L. and Schwabe,C. (1977) *Nature*, **270**, 449-451.

Bigazzi,M., Nardi,E., Bruni,P. and Petrucci,F. (1980) *J. Clin. Endocrinol. Metab.*, **51**, 939-947.

Chance,R.E., Hoffmann,J.A., Kroeff,E.P., Johnson,M.G., Schirmer,E.W., Bromer,W.W., Ross,M.J. and Wetzel,R. (1981) in Rich,D.H. and Gross, E. (eds.), *Peptides, Synthesis-Structure-Function*, Rockford; Pierce Chem. Co. Press, pp. 721-728.

Chang,A.C.Y., Nunberg,J.H., Kaufman,R.J., Erlich,H.A., Schimke,R.T. and Cohen,S.N. (1978) *Nature*, **275**, 617-624.

Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry (Wash.)*, **18**, 5294-5299.

DeNoto,F.M., Moore,D.A. and Goodman,H.M. (1981) *Nucleic Acids Res.*, **9**, 3719-3730.

Dodson,G.G., Eliopoulos,E.E., Isaacs,N.W., McCall,M.J., Niall,H.D. and North,A.C.T. (1982) *Int. J. Biol. Macromol.*, **4**, 399-405.

Fields,P.A. and Larkin,L.H. (1980) *J. Clin. Endocrinol. Metab.*, **52**, 79-90.

Grunstein,M. and Wallis,J. (1979) *Methods Enzymol.*, **68**, 379-389.

Haley,J., Hudson,P., Scanlon,D., John,M., Cronk,M., Shine,J., Tregear,G. and Niall,H. (1982) *DNA*, **1**, 155-162.

Hisaw,F.L. (1926) *Proc. Soc. Exp. Biol. Med.*, **23**, 661-663.

Hudson,P., Haley,J., Cronk,M., Shine,J. and Niall,H. (1981a) *Nature*, **291**, 127-131.

Hudson,P., Penschow,J., Shine,J., Ryan,G., Niall,H. and Coghlan,J. (1981b) *Endocrinology*, **108**, 353-356.

Hudson,P., Haley,J., John,M., Cronk,M., Crawford,R., Haralambidis,J., Tregear,G., Shine,J. and Niall,H. (1983) *Nature*, **301**, 628-631.

Isaacs,N., James,R., Niall,H., Bryant-Greenwood,G., Dodson,G., Evans,A. and North,A.C.T. (1978) *Nature*, **271**, 278-281.

James,R., Niall,H., Kwok,S. and Bryant-Greenwood,G. (1977) *Nature*, **267**, 544-546.

John,M.J., Walsh,J.R., Borjesson,B.W. and Niall,H.D. (1981) *Endocrinology*, **108**, 726-729.

Karin,M. and Richards,R.I. (1982) *Nature*, **299**, 797-802.

Kent,S.B.H. and Merrifield,R.B. (1981) in Brunfeldt,K. (ed.), *Peptides 1980*, Scripton, Copenhagen, pp. 328-333.

Knoll,B.J., Woo,S.L.C., Beattie,W. and O'Malley,B.W. (1981) *J. Biol. Chem.*, **256**, 7949-7953.

Lawn,R.M., Fritsch,E.F., Parker,R.C., Blake,G. and Maniatis,T. (1978) *Cell*, **15**, 1157-1174.

Maxam,A.M. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560-564.

Morrison,D.A. (1979) *Methods Enzymol.*, 326-331.

Parnes,J.R., Robinson,R.R. and Seidman,J.G. (1983) *Nature*, **302**, 449-452.

Proudfoot,N.J. and Maniatis,T. (1980) *Cell*, **21**, 537-544.

Rosenfeld,M.G., Lin,C.R., Amara,S.G., Stolarsky,L., Roos,B.A., Ong,E.S. and Evans,R.M. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 1717-1721.

Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161-178.

Schwabe,C., McDonald,J.K. and Steinetz,B.C. (1977) *Biochem. Biophys. Res. Commun.*, **75**, 503-510.

Schwabe,C., Gowan,L.K. and Reinig,J.W. (1982) *Ann. N.Y. Acad. Sci.*, **380**, 6-12.

Schwabe,C. (1983) in Bigazzi,M., Gasparri,E. and Greenwood,F. (eds.), *Relaxin and its Role in the Human: Proceedings of the First International Conference*, Elsevier, Amsterdam/NY, pp. 22-31.

Seeburg,P.H. (1982) *DNA*, **1**, 239-250.

Smith,M. (1982) in Weissman,S.M. (ed.), *Methods of RNA Sequencing*, Praeger Scientific, NY, pp. 182-187.

Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.

Taylor,J.M., Illmersee,R. and Summers,J. (1976) *Biochim. Biophys. Acta*, **442**, 324-330.

Tregear,G.W., Fagan,C., Reynolds,H., Scanlon,D., Jones,P., Kemp,B., Niall,H.D. and Du,Y.C. (1981) in Rich,D.H. and Gross,E. (eds.), *Peptides: Synthesis-Structure-Function*, Rochford: Pierce Chem. Co. Press, pp. 249-252.

Wilde,C.D., Crowther,C.E. and Cowan,N.J. (1982) *Science (Wash.)*, **217**, 549-555.

Wiquist,N. and Paul,K.G. (1958) *Acta Endocrinol.*, **29**, 135-146.

Yamamoto,S., Kwok,S.C.M., Greenwood,F.C. and Bryant-Greenwood,G. (1981) *J. Clin. Endocrinol. Metab.*, **52**, 601-615.

Zehner,Z.E. and Paterson,B.M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 911-915.

Received on 15 June 1984