The contribution of recombinant DNA techniques to reproductive biology

H. J. Stewart, D. S. C. Jones†, J. C. Pascall, R. M. Popkin and
A. P. F. Flint*

*AFRC Institute of Animal Physiology & Genetics Research, Cambridge Research Station,
Babraham, Cambridge CB2 4AT, U.K.*

	<i>Page</i>
Introduction	2
Structures of genes and mRNA molecules	2
Structure of genes	2
Sequences of promoters and enhancers	3
Signals for terminating RNA synthesis	5
Structure of mRNA molecules	5
Structure of genes coding for reproductive polypeptide hormones and hypothalamic releasing factors	7
Structure of gonadotrophin genes	7
Structure of placental lactogen, prolactin and growth hormone genes	10
Structure of the LHRH gene	12
Inhibin, activin and Müllerian duct-inhibiting hormone	13
Localization of gene products by in-situ hybridization	16
Pre- and post-translational processing	18
Pretranslational processing: differential splicing of RNA	18
Post-translational processing	18
Steroidogenic enzymes and steroid sulphatase	20
Steroid hormone receptors	22
Steroid receptor primary structure: amino acid sequences derived from cloned DNA	23
Steroid hormone receptors and oncogenesis	26
Mechanism of binding of receptors to DNA—the finger hypothesis	27
DNA sequences recognized by steroid receptors in eukaryotic cells	28
Fertilization and early development	30
Expression of the embryonic genome	30
Pattern formation in early development	32
Oncogenes and heat-shock proteins	33
Oncogenes	33
Heat-shock proteins	38
The future	39
Appendix 1: recombinant DNA techniques and procedures	40
Appendix 2: glossary of terms	45
Appendix 3: bibliography (textbooks and practical manuals)	47
References	47

*Present address: Institute of Zoology, Regent's Park, London NW1 4RY, U.K.

†Present address: MRC Molecular Genetics Unit, Laboratory of Molecular Biology, MRC Centre, Cambridge CB2 2QH, U.K.

Introduction

What is the structure of inhibin? Did steroid receptors evolve from a common ancestral protein? What is the molecular basis for the parallel between trophoblast growth and oncogenesis? These are questions that might appropriately be asked by a reproductive biologist which have to some extent already been answered through the application of recombinant DNA techniques. The purpose of this review is to outline the ways in which recombinant DNA techniques have contributed to a better understanding of questions in reproductive biology, and to point out the relevance to biology of the kinds of information obtained by applying these techniques to the study of specific genes (and their products) involved in reproduction. Emphasis is put wherever possible on structure/function relationships, especially as they apply to the way in which gene structure determines hormone structure and to evolutionary relationships between polypeptide hormones. Anyone regularly reading *Nature* or *Science* will realize that advances in this field are being made rapidly; therefore this article cannot be expected to be thoroughly up-to-date. However, molecular genetics has already brought to light so much novel information that there is no lack of relevant literature available for review.

The areas in which the more spectacular advances have been made include the structure of genes encoding gonadotrophins, gonadotrophin-releasing hormone and steroidogenic enzymes; the transcription, translation and post-translational processing of reproductive peptide hormones in the hypothalamus, pituitary and ovary; the structure and mechanism of action of steroid hormone receptors, and the activation of specific genes during growth of the trophoblast and embryo. These are the topics we shall cover in most depth. The order in which the topics are dealt with in general reflects the flow of endocrine information from hypothalamus to pituitary to gonad to steroid hormone target organ. The application of recombinant DNA techniques in neuroendocrinology has been reviewed previously by Charlton & Cox (1983), and only more recent findings in this field will be dealt with here.

In this review we have avoided using the term 'molecular biology', preferring instead to refer to 'recombinant DNA techniques'. The term 'molecular biology' was used by William Astbury to refer to the study of the chemical and physical structures of biological macromolecules (see Kendrew, 1970) and therefore it could be (and has been) applied as appropriately to polysaccharides and proteins as to nucleic acids. 'Molecular biology' is therefore too wide a term for the purpose of this review, which principally deals with information gained from studying recombinant DNA molecules. A brief description of some of the methods used is given in Appendix 1 and it may be useful for those unfamiliar with the terminology to read this first. A glossary of terms and list of practical manuals are also appended.

Structures of genes and mRNA molecules

The data reviewed below refer frequently to the structures of genes and mRNA molecules and it is therefore appropriate to review these before describing the information gained from individual sequences of reproductive relevance.

Structure of genes

DNA has several properties without which it could not be the substance of genes: it is chemically stable, it is capable of storing all the genetic information relevant to the species, it can be 'read' by enzymes converting the stored information into chemical structures, it can be replicated and repaired, and it is capable of mutation. DNA contains information on the sequence of amino acids for every protein synthesized by each tissue; signals indicating where on the chromosome the transcription of each protein starts and stops, and sequences capable of interacting with

compounds in the cell which determine how much of each protein is made, and when. The characteristics of each cell imparted during the process of differentiation, beginning at embryogenesis and continuing throughout the life of the organism, are stored in the DNA and reflected in the proteins synthesized from it. Yet despite the vast amount of information contained in the DNA of every mammalian cell, only a small proportion (probably < 10%) of the DNA is transcribed in any one cell; furthermore a large part of the DNA that is transcribed does not code for protein. There is, of course, in addition to DNA coding for proteins that coding for non-translated RNA molecules such as transfer RNA (tRNA) and ribosomal RNA (rRNA).

Eukaryotic genes have a number of features in common. The most striking is the fact that almost all the genes sequenced to date are discontinuous, containing sequences (named intervening sequences or introns; Gilbert, 1978) that are transcribed but not translated. (The histone and interferon genes are notable examples of genes without introns.) Eukaryotic introns were first discovered in the globin, immunoglobulin and chicken ovalbumin genes (see Gilbert, 1978; Crick, 1979) and have since been found in almost every eukaryotic gene examined. Generally the length of the introns exceeds that of the coding sequences (exons; Fig. 1). RNA corresponding to introns is lost during the synthesis of mature mRNA molecules by a process involving formation of a loop (or lariat) in the sequence (Fig. 1; Busch *et al.*, 1982; Sharp, 1987). The purpose of introns is at present uncertain, though it is clear that they perform a useful function, for their sequences are frequently highly conserved (see Crick, 1979; Jeffreys, 1981; Sharp, 1985). It has been suggested that they provide a mechanism for rapid evolution; the shuffling of exons during mutation may represent a mechanism for the conservation of domains in proteins with discrete functions.

One consequence of the existence of introns is that different products can be produced from the same gene by a process of differential mRNA splicing, which is discussed below. An alternative way in which a gene may code for more than one protein involves transcription from both DNA strands. This has been known to occur in prokaryotes for some time, but has only recently been recognized in eukaryotes; the only known mammalian example of this process involves the gonadotrophin-releasing hormone gene (Adelman *et al.*, 1987; see p. 12).

All genes evidently contain signals indicating, to the RNA polymerase molecules transcribing them, where transcription should start and where it should finish. Furthermore, there are sequences involved in controlling the frequency of transcription. Most attention has been given to sequences upstream of the position at which transcription begins; these contain identifiable sequences conserved in all eukaryotic genes—the start-site (or ‘cap’ site) where transcription begins, and promoter and enhancer sequences. In the discussion that follows it should be noted that nucleotide sequences (A, adenine; C, cytosine; G, guanine; T, thymidine; Pu, either purine, A or G; Py, either pyrimidine, C or T) are written in the upstream (5′-) to downstream (3′-) direction, and are given as occurring in the non-transcribed (sense) strand of DNA; RNA polymerase reads the code from the other (antisense or coding) strand, working in the direction 3′- to 5′- on that strand, producing an RNA molecule elongating in the 5′- to 3′- direction (Fig. 1b). As a result, the single stranded mRNA sequences produced are identical to those of the 5′- to 3′- non-transcribed DNA strand (except that T is replaced by U, uridine).

Sequences of promoters and enhancers. Promoters were discovered in bacterial genes, and their existence was proposed on genetic grounds (Jacob *et al.*, 1964) as sequences required for the expression of bacterial structural genes. It was subsequently shown that prokaryotic promoters are regions of the chromosome upstream of (i.e. located 5′- to) genes and that they are composed of a number of characteristic sequences of nucleotides required for the recognition and binding of RNA polymerase. Thus a prokaryotic promoter contains the mRNA start-site (consensus sequence 5′-CAT-3′), the AT-rich Pribnow or Schaller box sequence (consensus 5′-TATAAT-3′) and a ‘recognition sequence’ (consensus 5′-TTGACA-3′). These sequences are arranged in a specific spatial relationship within the DNA 5′- to the gene they control; the Pribnow box is usually about 10 base pairs (bp) upstream (5′-) from the mRNA start-site, and the recognition sequence is 25 bp upstream from the Pribnow box. There may be additional regulatory sequences further upstream.

Fig. 1. Structures of genes, products of gene expression, promoters and enhancers. **(a)** Schematic representation of the structure of a hypothetical eukaryotic gene coding for a secreted protein and the pathway of gene expression. Locations shown in the gene are: exons (hatched); introns (between exons); promoters, pr (GGPyCAA_ACT and TATA_AAA_A) and enhancers (en) in the 5'-untranscribed region; the transcription initiation or cap site (PyAPy); the translation initiation signal (ATG); the translation stop codon (TAG, TAA or TGA) and the polyadenylation signal (AATAAA). Enhancer regions may include those responsive to steroid receptors (e.g. TGTCT) which can be up to 2.6 kbp upstream (i.e. 5'-) of the promoter region or downstream (i.e. 3'-) of the cap site (see Table 2). Subsequent steps illustrate transcription, processing of the pre-mRNA to remove introns, translation and post-translational processing. Points of translation initiation (ti) and stop codons (st) are illustrated in the mature mRNA; the signal peptide (sp) of the preprohormone (necessary for packaging of the mature hormone into secretory granules) is indicated. NH₂ and COOH = amino- and carboxy-terminal ends of the translation product. **(b)** Sense and antisense strands of genomic DNA and the direction of RNA synthesis (transcription).

In contrast to progress with prokaryotes, understanding of the control of gene expression in eukaryotes is at an early stage. However, the development of cell-free systems capable of transcribing viral and cellular genes *in vitro* (Wu, 1978; Weil *et al.*, 1979; Manley *et al.*, 1980), and the use of site-directed mutagenesis to alter DNA sequences in specific ways, coupled with expression of genes in eukaryotic cells (*Xenopus laevis* oocytes or cells in culture), has recently allowed more rapid progress, and a number of 5'-consensus sequences have now been identified in eukaryotic genes (Fig. 1). The mRNA start-site ('cap site') is less well conserved in eukaryotes than in prokaryotes (consensus sequence PyAPy; Muthukrishnan *et al.*, 1976; Shatkin, 1976). About 28 bp upstream of the transcription start-site is a characteristic sequence known as the Hogness-Goldberg box (TATA box; consensus 5'-TATA_AAA_A-3'), which is present in almost all eukaryotic protein-coding genes so far sequenced. It is further from the transcription start-site than the Pribnow box of prokaryotes (3 turns of the double helix, in contrast to 1). A second conserved sequence upstream from

the TATA box at approximately 70–90 bp from the mRNA start-site has the consensus sequence 5'-GGPyCAA_ACT-3'. This has some similarity to the –35 bp region of a number of *E. coli* promoters (Benoist *et al.*, 1980). This and other, further 5'-, upstream sequences are relatively polymorphic, and may be repeated (up to 3 times, as in the rabbit β -globin gene; Dierks *et al.*, 1983). Their position 40–110 bp from the mRNA start-site is variable from one promoter to another, and they bear little relation to the TATA box; they are thought to regulate efficiency of transcription (Dierks *et al.*, 1983; McKnight & Kingsbury, 1982). In addition there also exist other transcription-controlling sequences, termed enhancers, usually located more than 110 bp upstream from the RNA start-site; these were discovered in viral genes but have since been found to be important in eukaryotes (see Rogers & Saunders, 1986). Enhancers occur in both the coding and non-coding DNA strands and may be located downstream as well as upstream of the mRNA start-site (see Table 2); it is uncertain how they act.

Studies with TATA box mutants obtained by site-directed mutagenesis have shown that the TATA box plays both qualitative and quantitative roles in the initiation of transcription. It acts not only as a positioning element directing the transcription machinery to initiate RNA synthesis about 30 bp downstream, at the RNA start-site, thereby determining the accuracy of initiation of transcription (Wasylyk *et al.*, 1980; Grosschedl & Birnstiel, 1982), but it also controls the rate of transcription. In-vitro studies have shown that the TATA box interacts specifically with at least one initiation protein, resulting in the formation of a stable preinitiation complex between the transcription machinery and promoter sequences (Davison *et al.*, 1983). It is speculated that the AT-rich regions 5'- to the transcription start-site represent lengths of DNA readily recognized by proteins; the double helix is presumed to open (melt) preferentially in these regions since A and T hydrogen bond less strongly to each other than do G and C residues. Besides their functional significance in controlling the rate of gene transcription, it should be noted that promoter regions are potentially important, in that mutations in them may lead to the formation of pseudo-genes—copies of genes that are otherwise intact but which are not expressed because their promoters are abnormal.

Promoters and other 5'- sequences are of growing importance in reproductive physiology not least because of their application in the construction of DNA sequences used to produce transgenic embryos. To ensure the expression of a gene transferred to an embryo, a promoter sequence capable of being recognized by eukaryotic RNA polymerase must be incorporated in it. A sequence commonly used for this purpose is the mouse or human metallothionein promoter, which normally controls the expression of the metallothionein gene in response to the presence of a heavy metal, such as zinc or cadmium. When the promoter is ligated to a gene and the resulting 'construct' transferred to an oocyte, it can subsequently be activated by administration of a heavy metal. The 5'-regions are also important in mediating interactions between steroid receptors and DNA (see p. 28).

Signals for terminating RNA synthesis. In contrast to the 5'- controlling regions upstream of the RNA start-site, sequences specifying termination of transcription are not well understood. Although it is clear that such sequences must exist, the complexity of eukaryotic genes and their length have hindered the use of cell-free systems and genes with 3'-deletions. At present no 3'-consensus sequence has been established as a transcription termination sequence in eukaryotes.

Structure of mRNA molecules

In contrast to that of the genes from which they are transcribed, the structure of mRNA molecules is well established (Fig. 1), and something is known of the processes leading to the production of mature mRNA molecules used as templates for the cytoplasmic protein synthetic machinery. The 5'-end of the molecule is 'capped' by addition of a characteristic nucleotide (often 7-methylguanosine) which is linked to the RNA sequence through a triphosphate bond (hence the term 'cap site' in the genomic sequence refers to the point at which this is added). Between the cap

site and the translation initiation codon (AUG), which codes for methionine, there is an untranslated sequence of variable length. Following this is the translated sequence, translation being terminated by one of the codons UAG, UAA or UGA. Loss of either the initiation or termination codons leads to protein synthesis starting or ending at the next appropriate codon downstream; but if the next AUG is not in the same reading frame as that lost, the amino acid sequence produced will be different. Examples of mutations involving loss of stop codons are those leading to the evolution of human chorionic gonadotrophin from luteinizing hormone and causing the failure of vasopressin production in the Brattleboro' rat (see below).

After the termination codon is a second untranslated region of variable length, followed by the polyadenylation signal, AAUAAA (Proudfoot & Brownlee, 1976) which determines the point some distance downstream at which up to 200 adenine residues are added. Polyadenylation to form what is referred to as poly (A⁺)-mRNA, may represent a mechanism for the stabilization of mRNA molecules, or be involved in their transport to the cytoplasm. It is clear that the polyadenylation signal is an important part of the mRNA molecule: in one human thalassaemia the sequence is AAUAAG (rather than AAUAAA), and this reduces the level of α_2 -globin mRNA to 5% that normally present in peripheral reticulocytes (Higgs *et al.*, 1983). Other sequences suggested to be responsible for controlling mRNA stability are also present in the 3'-untranslated end of the molecule, and contain the sequence 5'-AUUUA-3' and related motifs. These sequences decrease the half-life (that is, increase the intracellular clearance) of mRNA molecules to which they are added artificially (Cosman, 1987). The rates of degradation of certain mRNA molecules are controlled by hormones: prolactin reduces the degradation of casein mRNA in the mammary gland, and oestrogen has the same effect on ovalbumin mRNA in the chick oviduct and on induced mRNA in other target tissues (see Knowler & Beaumont, 1985). It is not known why different mRNAs are cleared at different rates. In some cases there is more than one polyadenylation signal in the mRNA sequence; when this occurs it is not uncommon for both to be used, so that there exists more than one mRNA population coding for the same protein. The glucocorticoid receptor appears to be an example of this.

In eukaryotes the mRNA molecules as they are first transcribed are longer than those appearing in the cytoplasm, because they include RNA coded for by the genomic introns. The conversion of the initial or nuclear RNA into mature mRNA molecules, which not only involves capping and polyadenylation, but also the excision of the introns, is termed 'processing', and it occurs in the nucleus. The order of processing events is: capping (which precedes release of the mRNA from the DNA template), polyadenylation, and excision of intervening sequences (splicing). Signals for capping and polyadenylation are, as described above, the 5'-terminal purine and the hexanucleotide AAUAAA respectively; all that is known of the mechanism of removal of introns is that they are frequently flanked by specific nucleotide sequences. Introns therefore generally start with 5'-GT-3' and end with 5'-AG-3' (Breathnach & Chambon, 1981). These sequences may be recognized by enzymes which cleave out the intron RNA and splice the exons together to form the mature mRNA. This occurs in the nucleus and is, compared to poly (A) addition, a relatively slow process (taking about 20 min for adenovirus fibre mRNA; Nevins, 1979). Not all the introns are removed simultaneously, so intermediate partly spliced products can be detected by Northern blotting (Abelson, 1979; Nordstron *et al.*, 1979). Accurate splicing is obviously of great importance in producing the correct mature mRNA, and the process by which this occurs is only beginning to be elucidated, but small nuclear RNAs are almost certainly involved (Busch *et al.*, 1982). At present it seems that the 5'-splice site is cleaved first and a lariat (or loop) formed, followed by excision of the intron as a lariat and ligation of the two exons. For a more detailed account see Reed & Maniatis (1985).

With these processing steps complete the mature mRNA molecule is extruded from the nucleus into the cytoplasm, where it undergoes translation by ribosomes. The process of transport is selective, since partly processed RNA or prematurely terminated RNA is not transported (Ziff & Evans, 1978); there is, however, little information on the mechanism of the selection process. The rate of

translation, and therefore the concentration of protein in the cell, is generally proportional to the concentration of each mRNA molecule, which in turn is controlled by the rates of synthesis and degradation of each mRNA species. Although the structural characteristics of genomic DNA sequences outlined above go some way towards explaining the mechanisms involved in defining RNA splice sites and points of polyadenylation, it is clear that other, as yet undetermined, mechanisms must also be involved. For instance, the Breathnach/Chambon rule describing nucleotide sequences at the boundaries of introns represents only part of the information determining splice sites—it remains to be explained why other potential splice sites are ignored. Furthermore, it is uncertain why in some mRNA molecules with more than one polyadenylation signal, one polyadenylation site appears to be preferred.

With this background in mind, let us investigate the structure of genes of importance in reproductive biology.

Structure of genes coding for reproductive polypeptide hormones and hypothalamic releasing hormones

Genes coding for hormones involved in reproduction were amongst the first mammalian genes to be thoroughly investigated—the human chorionic somatomammotrophin (hCS or hPL) gene was the first human gene to be cloned (Shine *et al.*, 1977)—and they illustrate all the common features of gene structure as well as providing valuable information on the regulation of gene expression. Two families of protein hormones which are excellent examples of advances in the understanding of evolutionary relationships made using recombinant DNA techniques are the glycoprotein hormones (human chorionic gonadotrophin, hCG; luteinizing hormone, LH; follicle-stimulating hormone, FSH; and thyroid-stimulating hormone, TSH) and the prolactin family (prolactin; placental lactogen, PL; and growth hormone, GH). These hormones have largely been characterized by protein chemistry, amino acid sequences revealing that LH, CG, FSH and TSH are homologous proteins consisting of two polypeptide chains, a common α -subunit (of 90–96 amino acids) and different β -subunits (of 118–145 amino acids) the latter showing some degree of sequence homology but conferring biological specificity (Fig. 2; and see Boime *et al.*, 1982).

Structure of gonadotrophin genes

Since hCG is the major protein product of first trimester human placenta it was expected that mRNA isolated from this tissue would yield hCG- α and - β subunits after in-vitro translation. This observation was made in the late 1970s when immunoprecipitation with specific antisera revealed two proteins of M_r 13 000 and 16 000 corresponding to the primary translation products of the α - and β -subunits (Fiddes & Goodman, 1979). Both subunits were produced from larger precursor proteins containing a signal sequence and translated from a large mRNA species. Subsequent preparation of a cDNA library resulted in the isolation of a 621 bp cDNA fragment encoding the human α -subunit, the sequence corresponding to the known 92 amino acid α -subunit, including the expected hydrophobic signal peptide typical of secreted proteins (24 amino acids long in hCG), a 220 base 3'-end (the 3'-untranslated region) and a 50 base 5'-untranslated region (Fiddes & Goodman, 1979). Similar structures have been deduced for rat (Godine *et al.*, 1980, 1982), mouse (Chin *et al.*, 1981), cattle (Nilson *et al.*, 1983), pig (Pascal *et al.*, 1984) and horse (Stewart *et al.*, 1987a, b) α -subunit mRNA sequences.

With the isolation of the cDNA encoding the α -subunit of hCG, which is common to all the human glycoprotein hormones, the way was clear to utilize this as a probe to investigate the α -subunit genomic sequence, the number of genes present in the genome, and their chromosomal location. A human genomic library hybridized with such a probe revealed a single α -subunit recombinant (see Fiddes & Talmadge, 1984), and sequence analysis showed that the gene was 9.4 kb in

Fig. 2. Structural relationships between the genes encoding the α - and β -subunits of the human glycoprotein hormones LH, FSH, TSH and CG, and the evolution of hCG- β from LH- β . **(A)** Organization of 7 hCG- β and 1 LH- β genes on 3 genomic fragments (a,b,c). Cumulative data from several clones; spatial relationships between sequences a, b and c are not determined. Arrows indicate positions of genes and orientation of transcription; only gene 4 codes for LH- β . After Talmadge *et al.* (1984a), although Policastro *et al.* (1986) have suggested that one of these genes may be artefactual. **(B)** Structure of human glycoprotein α -subunit gene. Numbers above horizontal line indicate kilobases, measured from an arbitrary point 5'- to the gene. The 4 exons are thickened; exon 1 is non-coding. **(C)** Structure of the β -genes and evolution of hCG- β from LH- β . Exons are thickened. Single base deletion and insertion of 2 bases near 3'-end (at hyphens in DNA sequence) in hCG- β result in translation from a new reading frame, causing incorporation of 8 different amino acids relative to those in LH- β ; in addition, an extra 24 amino acids are incorporated into the sequence from the 3'-region untranslated in the LH- β gene. Termination of hCG- β translation occurs at the TAA which constitutes part of the polyadenylation sequence in LH- β . The kilobase scale is numbered arbitrarily; bold numbers indicate numbers of amino acids coded for by each exon. ATG = initiation codon; TAA = termination codon; AATAAA = polyadenylation signal (in LH- β only; polyadenylation of hCG- β message occurs at a site further downstream).

length, of which 8.5 kb was located in 3 non-coding regions (introns; Fig. 2). There is only one gene in the human encoding the α -subunit common to all four human glycoprotein hormones (Boothby *et al.*, 1981; Fiddes & Talmadge, 1984), located on chromosome 6 (Naylor *et al.*, 1983; Dracopoli *et al.*, 1986).

The molecular organization of hCG- β coding sequences proved more challenging and yielded unexpected results. From the known amino acid sequence, the predicted DNA sequence should contain a single restriction site recognized by the enzyme *Sau96I*. A cDNA library from first trimester placenta was therefore probed for sequences containing such a site, and two clones were isolated encoding hCG- β (579 bases encoding a 20 amino acid signal peptide and the full 145 amino acid β -chain; see Fiddes & Goodman, 1980). Again a human genomic library was screened using the cDNA probe but in this case the result was surprising. Because of the similarity between the hCG- β and human LH- β amino acid sequences it was predicted that the human LH- β gene would

be detected by the hCG- β probe but the discovery of 16 recombinants in the genomic library was unexpected. Eight genes were thought to be represented (Boorstein *et al.*, 1982; Talmadge *et al.*, 1984a), one encoding LH- β (Talmadge *et al.*, 1984a) and the remaining 7 encoding hCG- β (Fig. 2). However, it has now been suggested that there are only 6 hCG- β genes and 1 hLH- β gene in a 58 kb EcoRI genomic fragment (Policastro *et al.*, 1986). All these genes share a basic structural organization which is more compact than that of the α -gene, consisting of 3 exons and 2 introns.

The hCG- β genes are unlikely to all possess functional promoters and since only 2 or possibly 3 are expressed when injected into COS cells (monkey cells which act as host for SV40 virus), the remainder are likely to be pseudogenes (Talmadge *et al.*, 1984b; Policastro *et al.*, 1986). No differences have been identified in the flanking sequences of expressed or non-expressed genes which might explain why some are transcribed and others not, although differences in the upstream sequences have been demonstrated (Talmadge *et al.*, 1984b; Policastro *et al.*, 1986). In addition, it is still uncertain how tissue-specific regulation of β -subunit genes (hCG- β in placenta and human LH- β in the pituitary) is achieved since both genes have similar sequences upstream of the initiation codon and no obvious sequence changes have been discovered in putative regulatory regions (Fiddes & Talmadge, 1984). Perhaps such regulation occurs as a result of methylation of neighbouring DNA sequences (Felsenfeld & McGhee, 1982) or by other subtle changes in the structure of the chromatin (Weisbrod, 1982; see p. 30). Multiple copies of CG genes also occur in the baboon genome (Crawford *et al.*, 1986), although there are marked differences between nucleotide sequences coding for CG- β in man and the baboon.

The sequence homology between hCG- β and human LH- β , which for hCG- β gene 5 is more than 90% at the nucleotide level, shows that the former evolved from the latter, the suggested mechanism involving a frame shift mutation in the LH- β gene causing loss of the termination codon and read through the 3'-untranslated region (Fig. 2). This creates the 24 amino acid C-terminal extension of hCG- β (Talmadge *et al.*, 1984a). The termination codon utilized in hCG- β is the UAA codon in the polyadenylation signal AAUAAA (Fiddes & Goodman, 1980). Although it has been suggested that there is homology between LH α - and β -subunits, because of a possible common evolutionary precursor, this is now known to be untrue (Stewart & Stewart, 1977).

Clues to the evolution of LH- β and CG- β subunit genes are also provided by studies of their chromosomal distribution. In man these genes are part of a linkage group on chromosome 19, with the FSH- β and TSH- β subunit and the common α -subunit genes being mapped to chromosomes 11, 1 and 6 respectively (Dracopoli *et al.*, 1986; Glaser *et al.*, 1986). In the mouse the LH- β , TSH- β and α -subunit genes are on chromosomes 7, 3 and 4 respectively (Dracopoli *et al.*, 1986), and there is no CG gene. Thus, the duplication of the ancestral β -subunit gene and separation of the LH- β , TSH- β and FSH- β genes to different chromosomes occurred before evolutionary divergence of mice and man and the evolution of hCG- β .

The other group of animals noted for production of a chorionic gonadotrophin, the equids, do not in fact express a characteristically placental gene: the protein synthesized by the chorionic girdle cells has an identical amino acid sequence to horse LH, for both its α - (Stewart *et al.*, 1987a, b) and β - (Bousfield *et al.*, 1987; Sugino *et al.*, 1987) subunits. It is, however, glycosylated in different ways in the pituitary (to produce LH) and placenta (to produce CG or PMSG), and this accounts for differences in the FSH/LH activities of the 2 hormones.

A major uncertainty yet to be fully resolved is how the two subunits, which are encoded by separate genes, are co-ordinately expressed and how the hormone-specific β -gene product is controlled to dictate which glycoprotein hormone is produced. Since all the glycoproteins share an α -subunit, a common signal for its production is likely, with tissue-specific production residing in the control of the β -subunit. Although α -subunit mRNA levels are known to correlate with serum and pituitary LH concentrations in cyclic ewes (Landefeld & Kepa, 1984), co-ordinate control of α and β genes is needed, since both subunits are induced in synchrony after gonadectomy (Abbot *et al.*, 1985; Papavasiliou *et al.*, 1986). Increases in both α - and β -subunit mRNAs have been reported in placental tumour cells in response to cyclic AMP analogues, which stimulate hCG production

(Burnside *et al.*, 1985; Jameson *et al.*, 1986), and mRNAs coding for LH- α and LH- β are known to decrease in parallel in ovariectomized ewes after pituitary stalk transection (Hamernik *et al.*, 1986) but it is uncertain how this is achieved. Elucidation of the mechanisms by which the glycoprotein genes are regulated and expressed would yield vital information on the processes involved in tissue-specific gene expression and developmentally induced changes; the demonstration that hCG cDNA can be expressed in isolated eukaryotic cells (Reddy *et al.*, 1985) may provide the means by which such studies can be undertaken.

Information on FSH- β mRNA structure has recently been derived from clones identified in a bovine anterior pituitary cDNA library using synthetic oligonucleotides of 47 and 50 bases (Esch *et al.*, 1986). The data show that FSH- β mRNA is surprisingly large (1.7 kb, compared to 0.7 kb for LH- β and 0.9 kb for hCG- β mRNAs); that there may be multiple forms of mRNA (since 2 clones differing in 3'-sequences were identified), and that, as in hCG- β , the polyadenylation signal (AAUAAA) not only acts as such but also contains the code for the last amino acid (codon GAA for glutamic acid) and the termination codon (UAA). It is uncertain whether FSH- β is encoded by a single-copy gene, although it seems certain that this is the case for TSH- β (Carr *et al.*, 1987).

Structure of placental lactogen, prolactin and growth hormone genes

The other main family of reproductive peptide hormones whose evolutionary relationships have been deduced from amino acid sequences and confirmed and extended by detailed examination of gene structure is the prolactin family, comprising prolactin, human placental lactogen (hPL) and growth hormone (GH). Several excellent reviews are available on prolactin and GH (e.g. Cooke *et al.*, 1981; Maurer, 1983; Miller & Eberhardt, 1983) and so this section will concentrate on information derived from a molecular approach to hPL gene expression. Structural and evolutionary relationships between these genes are illustrated in Fig. 3.

The approach used to investigate hPL gene structure was based on the known abundance of hPL in term placental tissue. Cell-free translation of mRNA isolated from term placenta indicated both that mRNA levels changed in parallel with circulating concentrations of hPL and that hPL represented 50% of total protein synthesis at this time (see Boime *et al.*, 1978). Sucrose gradient sedimentation of mRNA enabled enrichment of mRNA of the size predicted to code for hPL, and hPL was subsequently cloned as a 550-nucleotide cDNA fragment containing hPL sequences (Shine *et al.*, 1977). Both codon usage and coding sequences showed such a striking homology to the hGH gene that it was suggested that the 2 hormones share a common evolutionary origin (Shine *et al.*, 1977; Fiddes *et al.*, 1979; Fig. 3). With the availability of a specific cDNA probe, human genomic libraries were screened. Because of the homology between GH and hPL and cDNA sequences it was expected that the probe would identify both gene sequences, and this was indeed the case, 7 clones being isolated, 5 encoding hPL-like sequences and 2 encoding GH (Kidd & Saunders, 1982; Miller & Eberhardt, 1983, and references therein). Restriction map analysis revealed that these genes share a common nucleotide arrangement with 5 exons separated by 4 introns and, in common with other multigene families with a common evolutionary origin, the genes are linked physically (in a span of 50 kb on chromosome 17; Owerbach *et al.*, 1980; George *et al.*, 1981; Barsh *et al.*, 1983).

As is the case with the hCG- β genes, not all the genes for hPL are transcriptionally active (Barrera-Saldana *et al.*, 1982, 1983; see also Miller & Eberhardt, 1983). Only 2 of the 5 genes appear to be expressed, one providing 60% and one 40% of the hPL produced. The function of the remaining genes, which have not been characterized in detail, is unknown. Only one of the GH genes is expressed. Expression of more than one gene for a circulating hormone may lead to molecular heterogeneity of the hormone extracted from plasma; however, this has not been detected for hPL. Why the 2 hPL genes are transcriptionally active has been elucidated by detailed analysis of cell-free transcription of isolated genes. It is thought that there is more than one transcription initiation site and that the sites are utilized with different efficiencies *in vitro*. As described above, DNA

Fig. 3. Structure of the human prolactin (PRL), placental lactogen (PL) and growth hormone (GH) genes and the evolutionary relationships between them. **(A)** Sequence homologies between the genes suggest that prolactin diverged from growth hormone about 360×10^6 years ago; PL and GH diverged about 65×10^6 years ago. (Palaeontological records indicate that amphibians evolved some 400×10^6 years ago and that mammalian radiation occurred $85\text{--}100 \times 10^6$ years ago; late evolution of PL may explain why some mammalian families, for instance suids, do not have such a hormone.) There are at least 2 GH and at least 5 PL genes (on chromosome 17) but only 1 prolactin gene (on chromosome 6). (In the rat GH and prolactin genes are on chromosomes 10 and 17 respectively.) Nucleotide homology between PL and GH genes (including introns) is approximately 92%; that between GH gene 1 and prolactin (coding regions only) is 42%. **(B)** Arrangement of human PL and GH genes in 2 genomic clones, as determined by restriction mapping; arrows indicate directions of transcription. Other, as yet unidentified, genes may also exist. E indicates position of the enhancer sequence identified by Rogers & Saunders (1986). **(C)** Structures of human prolactin (PRL), PL and GH genes; note similarities in sizes of exons (thickened; lengths indicated above line in bp). Lengths of introns are indicated below line (in bp); introns are considerably longer in the prolactin gene than in the other genes but exon arrangement is similar. All introns in all 3 genes obey the GT-AG rule (Breathnach & Chambon, 1981). *Length of coding region in 5'- and 3'-exons. Initiation (ATG) and stop codons (TAG or TAA), and polyadenylation signals (AATAAA) are indicated. Data from Miller & Eberhardt (1983); Truong *et al.* (1984); Cooke *et al.* (1986).

sequences in the region of the transcription initiation site are likely to be of great interest in unravelling the control features for specific expression of a particular gene. The function of a second initiation site (29 bases downstream from the TATA sequence) in the expressed hPL genes (Selvanayagam *et al.*, 1984) is unknown, although it may regulate efficiency of transcription and may be involved in responding to controlling signals governing the abundance of the hormone in the cell. It remains to be determined how such homologous genes as hGH and hPL are differentially expressed in the pituitary and placenta respectively, despite being interspersed and linked (Barsh *et al.*, 1983). Uncertainty in this area is heightened by the observation that promoter sequences 5'- to the PL gene, when fused to other sequences to form chimaeric genes, will control transcription of foreign DNA in pituitary cells (Cattini & Eberhardt, 1987), while sequences upstream of the GH gene will not induce transcription of other genes in pituitary lactotrophs

(Nelson *et al.*, 1986). On the basis of the latter finding it has been suggested that prolactin- and GH-secreting cells contain discrete proteins interacting with different sequences controlling expression of prolactin and GH genes. An enhancer sequence has been identified in the 3'-flanking region of the hPL-3 gene which is active in placenta-derived cells (Rogers & Saunders, 1986).

Structure of the LHRH gene

With the sequencing of the larger protein and glycoprotein hormone genes largely achieved by the early 1980s, other important reproductive hormones were then investigated by a variety of approaches. The molecular cloning of the decapeptide luteinizing hormone releasing hormone (LHRH; Fig. 4) represents one of the most technically demanding exercises so far because of the low abundance of the peptide in the hypothalamus (e.g. Shivers *et al.*, 1983) and the small proportion (<0.01%) of the total mRNA, produced by hypothalamic neurones, which codes for LHRH. The usual approach for cloning hormone genes relies on the premise that the mRNA for a hormone is likely to be more abundant than the gene sequence. Hence a logical approach is to isolate a cDNA clone and utilize this as a specific probe to find the gene sequence. However, for LHRH the successful group (one of many attempting unsuccessfully to find the sequence in cDNA libraries) used a different approach, screening a human genomic library with synthetic oligonucleotide probes (Seeburg & Adelman, 1984). Since the amino acid sequence of LHRH was known (Schally *et al.*, 1971) it was possible to predict sequences for the gene from the genetic code. The main problem with this approach is the pronounced redundancy inherent in the code; of the 10 amino acids of LHRH only 1 is encoded by a single codon, and 6 are potentially coded for by 4 or more. Clearly, therefore, of prime importance in the cloning of LHRH was the generation of a large library and choice of the optimum sequence for the oligonucleotide probe. The approach of Seeburg & Adelman (1984) was to use pools of oligonucleotides, both short probes (coding for LHRH alone) and longer ones (coding for LHRH plus putative post-translational processing sites). Only one recombinant was found which hybridized to both the short and long probes, and sequencing of the insert within this recombinant indicated that it contained the gene for LHRH. Having located a 600 bp genomic fragment this was utilized as a highly specific probe with which to screen a cDNA library for the corresponding recombinant, from which the structure of the precursor protein could be determined. Again, a large library was prepared, starting with mRNA (10 µg) from human placenta, where it had been found that LHRH was produced (Khodr & Siler-Khodr, 1978; Tan & Rousseau, 1983). Of 1.5×10^6 clones, two, of 1.5 and 0.8 kb, hybridized to the genomic probe. This surprising finding meant that a 10-amino acid peptide originates from a transcript of more than 1 kb. Furthermore, within the reading frame encoding the expected sequence for LHRH are at least two new peptide hormones, the existence of which can be predicted from likely sites of proteolytic post-translational processing (see below); one of these, 56 amino acids in length and representing the C-terminal prohormone fragment is synthesized and released with LHRH (Fig. 4). This protein, termed gonadotrophin-releasing hormone-associated peptide (GAP; Seeburg & Adelman, 1984) has been claimed to inhibit prolactin secretion, and is dealt with in more detail later (see p. 20). A second, shorter fragment corresponding to amino acids 14–26 (Fig. 4), which has been suggested to exist on the basis of post-translational cleavage at a single basic amino acid residue (Lys), has been shown to have gonadotrophin-releasing activity similar to LHRH (Millar *et al.*, 1986).

Further analysis of the LHRH gene from human and rat genomic libraries revealed similar gene arrangements, 4 coding regions separated by 3 introns extending over 4 kb (Fig. 4). The first exon, which was probably incomplete in the clone investigated, contains the 5'-untranslated region of the mRNA; the second encodes LHRH and a processing site separating it from the first 11 residues of GAP, the 3rd and 4th exons encoding GAP and the 3'-untranslated region (Adelman *et al.*, 1986). Genomic analysis indicates that the gene sequence is highly conserved and that it is a single-copy gene in man and rat. It has been localized to chromosome 8 (Yang-Feng *et al.*, 1986).

In addition to peptides produced by post-translational processing of the LHRH prohormone, the LHRH genomic sequence has recently been shown to contain a second gene, which codes for a 106-amino acid protein, the mRNA being transcribed from the opposite strand to that encoding LHRH (Adelman *et al.*, 1987). The sequence coding for this protein includes part of the intron between exons 1 and 2 in the LHRH gene and extends through the exon encoding the LHRH prohormone signal peptide, the LHRH decapeptide and the initial 11 amino acids of GAP. The gene is expressed in the rat heart; its function is not known, although it has been suggested (Adelman *et al.*, 1987) that, being an anti-sense strand, the RNA transcribed from it may control translation of LHRH mRNA by hybridization. This is the only currently known example of 2 genes occupying the same DNA in the mammalian genome.

The abnormality responsible for hereditary hypogonadism in the *hpg* mouse has been identified by genomic sequencing (Mason *et al.*, 1986a). This genetic abnormality, which involves failure to produce LHRH, can be reversed by transplantation of normal hypothalamic tissue to the brain (Cattanach *et al.*, 1977; Kreiger *et al.*, 1982); the lesion has now been shown to be caused by a mutation involving the deletion of a large fragment (at least 33.5 kbp) of the gene (Fig. 4). The part of the gene coding for LHRH, the signal peptide and relevant post-translational processing signals are not lost in the *hpg* mutation, however, and in-situ hybridization indicates that the gene is transcribed; the lack of synthesis of LHRH may, therefore, be due to the major change in structure of the prohormone caused by the deletion. The lesion can be reversed by insertion of a normal LHRH gene into the *hpg* genome using transgenic techniques (Mason *et al.*, 1986b; and see below).

It would be of enormous value to ascertain whether the large number of reports of extra-hypothalamic LHRH-like activity (e.g. Barnea & Porter, 1975; Sharpe *et al.*, 1981; Aten *et al.*, 1987) is due to peripheral expression of the gene expressed in the hypothalamus. Indeed, the role of LHRH in placental function is poorly understood, although recent data support the hypothesis that hCG release may be affected by local production of LHRH with modulatory influences from local steroid hormone levels (Siler-Khodr *et al.*, 1986a). In addition, LHRH may influence the release of steroid hormones from placental culture (Siler-Khodr *et al.*, 1986b). Unfortunately, the extremely low abundance of LHRH mRNA in the hypothalamus makes classical Northern blot and transcription rate analysis unlikely to be successful. Further investigation of the sequences at the 5'-end of the gene would be of interest by providing information on the promoter region and may give insight into how gene expression might be regulated by steroid hormones and other factors.

Inhibin, activin and Müllerian duct-inhibiting hormone

Recombinant DNA technology has provided an insight into the structure of the inhibin molecule (see Fig. 5) and has in addition revealed a number of unexpected findings relating to this molecule (see Tsonis & Sharpe, 1986). Inhibin purified from pig follicular fluid consists of two forms, each comprising one α - and one of two possible β -subunits, β_A or β_B , which differ mainly in their N-terminal sequences. By use of single long oligonucleotides predicted from the N-terminal regions of the α - and β -subunits, the cDNAs encoding all of the subunits were selected from λ gt10 cDNA libraries derived from pig ovarian mRNA (Mason *et al.*, 1985). Sequencing of the cDNA clones predicted 364, 424 and 423 amino acids precursors for the α -, the β_A - and the β_B -subunits respectively. In each case the mature peptide was situated at the C-terminus of the precursor comprising 134 (α -subunit) or 116 (β -subunit) amino acid residues (Mason *et al.*, 1985). Northern blotting indicates that concentrations of α -precursor mRNA are at least 10-fold higher than those of the β_A -precursor mRNA and 20-fold higher than those of the β_B -precursor mRNA.

Recently the story has been taken further. Two groups have independently demonstrated that, although inhibin, consisting of either $\alpha\beta_A$ or $\alpha\beta_B$ chains, inhibits release of FSH from the anterior pituitary, combinations of the two β -chains ($\beta_A\beta_A$ or $\beta_A\beta_B$) induce FSH release. These proteins have been termed FSH-releasing protein ($\beta_A\beta_A$; Vale *et al.*, 1986) or 'activin' ($\beta_A\beta_B$; Ling *et al.*,

(a)	H R	GGGATCTTTTGGCTCTCTGCCTCTAAACAAGA																				65
		TGGTATCCCTTTGGCTTTACATCCAAACAAGA																				
		<div style="display: flex; justify-content: space-between;"> -23 -20 </div> <div style="display: flex; justify-content: space-between;"> MET lys pro ile gln lys leu ala lys leu ATG AAG CCA ATT CAA AAA CTC CTA GCT GGC CTT 65 </div> <div style="display: flex; justify-content: space-between;"> ATG GAA ACC ATC CCC AAA CTG ATG GCT GCT GTT MET glu thr ile pro lys leu met ala ala val </div>																				
		<div style="display: flex; justify-content: space-between;"> -10 1 </div> <div style="display: flex; justify-content: space-between;"> ile leu leu thr trp cys val glu gly cys ser ser gln his trp ser tyr gly leu arg 125 </div> <div style="display: flex; justify-content: space-between;"> ATT CTA CTG ACT TGG TGC GTG GAA GGC TGC TCC AGC CAG CAC TGG TCC TAT GGA CTG CCG </div> <div style="display: flex; justify-content: space-between;"> GTT CTG TTG ACT GTG TGT TTG GAA GGC TGC TCC AGC CAG CAC TGG TCC TAT GGG TTG CCG </div> <div style="display: flex; justify-content: space-between;"> val leu leu thr val cys leu glu gly cys ser ser gln his trp ser tyr gly leu arg </div>																				
		<div style="display: flex; justify-content: space-between;"> 10 20 </div> <div style="display: flex; justify-content: space-between;"> pro gly gly lys arg asp ala glu asn leu ile asp ser phe gln glu ile val lys glu 185 </div> <div style="display: flex; justify-content: space-between;"> CCT GGA GGA AAG AGA GAT GCC GAA AAT TTG ATT GAT TCT TTC CAA GAG ATA GTC AAA GAG </div> <div style="display: flex; justify-content: space-between;"> CCT GGG GGG AAG AGA AAT ACT GAA CAC TTG GTT GAT TCT TTC CAA GAG ATG GGC AAG GAG </div> <div style="display: flex; justify-content: space-between;"> pro gly gly lys arg asn thr glu his leu val asp ser phe gln glu met gly lys glu </div>																				
		<div style="display: flex; justify-content: space-between;"> 30 40 </div> <div style="display: flex; justify-content: space-between;"> val gly gln leu ala glu thr gln arg phe glu cys thr thr his gln pro arg ser pro 245 </div> <div style="display: flex; justify-content: space-between;"> GTT GGT CAA CTG GCA GAA ACC CAA CCG TTC GAA TGC ACC ACG CAC CAG CCA CGT TCT CCC </div> <div style="display: flex; justify-content: space-between;"> GAG GAT CAA CTG GCA GAA ACC CAA AAC TTC GAA TGC ACT GTC CAC TGG CCC CGT TCA CCT </div> <div style="display: flex; justify-content: space-between;"> glu asp gln met ala glu pro gln asn phe glu cys thr val his trp pro arg ser pro </div>																				
		<div style="display: flex; justify-content: space-between;"> 50 60 </div> <div style="display: flex; justify-content: space-between;"> leu arg asp leu lys gly ala leu glu ser leu ile glu glu glu thr gly gln lys lys 305 </div> <div style="display: flex; justify-content: space-between;"> CTC CGA GAC CTG AAA GGA GCT CTG GAA AGT CTG ATT GAA GAG GAA ACT GGG CAG AAG AAG </div> <div style="display: flex; justify-content: space-between;"> CTT CGA GAT CTG CGA GGA GCT CTG GAA CGT CTG ATT GAA GAG GAA GCT GGG CAG AAG AAG </div> <div style="display: flex; justify-content: space-between;"> leu arg asp leu arg gly ala leu glu arg leu ile glu glu glu ala gly gln lys lys </div>																				
		<div style="display: flex; justify-content: space-between;"> 69 </div> <div style="display: flex; justify-content: space-between;"> ile END ATT TAA ATCCATTGGGCCAGAGAAGGAATGACCATTACTACATGACTTAAGTATAATTCGACATTGAAAAATTTATAA 382 </div> <div style="display: flex; justify-content: space-between;"> ATG TAA -TGCACTGGCCAGAGG-AT--CCA---CAACAT---CCGAGTGTGACATTGACGCTGAGA-TCCATGA </div> <div style="display: flex; justify-content: space-between;"> met END </div>																				
		<div style="display: flex; justify-content: space-between;"> 461 </div> <div style="display: flex; justify-content: space-between;"> CCCATTAAATACCTGTAATGGTATGAATTCAGAAATCCTTACACCAAGTTGCACATATTCATAATAAAGTGCTGTG </div> <div style="display: flex; justify-content: space-between;"> CCTGTACAGCTCTGTAATTG-TGTGGGCTCCGCATTTCTATACCC-AGCAGCGTAGATTTTCATAAAGTAATGTG </div>																				
		<div style="display: flex; justify-content: space-between;"> TTGTGAATG-polyA </div> <div style="display: flex; justify-content: space-between;"> TTGTGGATC-polyA </div>																				

1986). Furthermore, homology has been found between the β -subunit sequence and the primary structure of transforming growth factor- β (TGF β), a protein found in transformed fibroblasts that has growth-promoting and -inhibiting effects on different types of cells (see Massagué, 1987) and is a potent stimulator of FSH release.

The mRNA encoding a second protein homologous to TGF β which is involved in reproduction and development has also been cloned recently. This protein, Müllerian duct-inhibiting hormone (MIH), causes regression of the Müllerian duct during development of the male embryo (see Josso, 1986). In the female embryo this structure develops into the uterus, vagina and oviducts. Using oligonucleotides predicted from peptides derived from purified bovine MIH, recombinants were selected and sequenced from a λ gt10 cDNA library prepared from new-born calf testis. Bovine MIH was shown to comprise a 575-amino acid precursor including a leader sequence of 24 amino acids of which the first 16–17 amino acids encode a putative signal sequence (Cate *et al.*, 1986). Cate *et al.* (1986) also reported the isolation of bovine and human genomic clones encoding MIH and confirmed the identity of the latter by introducing the gene in an expression vector into mouse COS cells and demonstrating the presence of MIH bioactivity in the media. The size and distribution of the MIH and mRNA has been confirmed by Picard *et al.* (1986) by screening a λ gt11 cDNA library with antibodies directed against the purified hormone. The homology between TGF β and MIH is of particular interest, as both are capable of inhibiting the cell cycle and normal cellular growth. It may be possible to use MIH to control the growth of tumours derived from tissues related to the Müllerian duct; MIH has been shown to inhibit *in-vitro* growth of a large number of human primary ovarian and endometrial cancers (Fuller *et al.*, 1985). TGF β , despite its name, also blocks cellular division *in vitro* and this may be its normal function (see Massagué, 1985). Structural comparison of the MIH and TGF β genes may provide an insight into the regions of these proteins responsible for their anti-tumour activity and improve our understanding of the regulation of cell growth.

The homologies between TGF β , MIH and the β -subunits of inhibin (Massagué, 1987) have led to speculation that the three proteins may be members of a single gene family. All three have similar distributions of cysteine residues within the region of homology and all are probably active as disulphide bridge-linked dimers rather than monomers. It will be of interest to determine whether inhibin possesses any anti-tumour activity.

Fig. 4. Sequence of the LHRH prohormone and complementary DNA, and the deletion in hypogonadal mice. **(a)** Sequences of human (H) and rat (R) hypothalamic cDNAs encoding the LHRH prohormone, and predicted prohormone sequence. Clones were identified using a human placental LHRH cDNA probe isolated by means of a sequence which was itself obtained by screening a human genomic library with oligonucleotides. Amino acid numbers are given above the sequence (negative numbers correspond to signal peptides); numbers at the ends of each line identify the last nucleotide. The 3'-untranslated regions are aligned to maximize homology (hyphens indicate putative deletions). Arrows show positions of introns in the genomic sequences; the polyadenylation signal is underlined. The LHRH sequences (amino acids 1–10) and putative prolactin-inhibitory factor (GAP: amino acids 14–69) are boxed in solid lines and the GlyLysArg C-terminal post-translational processing signal is boxed in broken lines. (After Seeburg & Adelman, 1984; Adelman *et al.*, 1986.) **(b)** Genomic deletion responsible for hypogonadism in mice; (i) normal mouse LHRH gene, (ii) *hpg* gene. Exons are thickened; figures above exons indicate the number of amino acids for which they code (note exon I is non-coding). Section between arrows is deleted in hypogonadism. ATG and TAG indicate positions of translation initiation and stop codons. The LHRH sequence, contained in exon II as indicated in (ii), is not itself lost in the mutated gene but synthesis of the peptide is much reduced, possibly because of the abnormal structure of the prohormone. (After Mason *et al.*, 1986a).

Fig. 5. Structure of inhibin and related hormones. **(a)** Structures of cloned cDNAs encoding pig inhibin subunits. The 134-amino acid α -subunit represents amino acids 231–364 of a 364-amino acid prohormone; the α -sequence is preceded by an Arg-Arg proteolytic processing signal. The β -subunits (β_A and β_B) are synthesized as related prohormones of 424 and 423 amino acids respectively; both represent the C-terminal end of the prohormones and are preceded by 5 consecutive Arg residues (β_A) or Arg-Lys-Arg (β_B) processing signals. The β_B -subunit is one amino acid shorter than β_A , and is 70% homologous to it. Figures above lines (*italics*) indicate nucleotide numbers (from 5'-ends of cDNA clones sequenced), those below lines show amino acids; boxes indicate translated regions, shaded boxes the α - and β -subunits. (After Mason *et al.*, 1985.) **(b)** Organization of α - and β -subunits in inhibin, FSH releasing protein (FSH-RP) and activin, and the structure of the related homo-dimeric protein, TGF- β . Subunits are linked by disulphide bonds in each dimeric protein. Molecular weights are indicated in brackets and activities of the dimers are given. TGF- β is related to Müllerian duct inhibiting hormone. After Tsonis & Sharpe (1986).

Localization of gene products by in-situ hybridization

In addition to their application to the analysis of gene structure and function, DNA and RNA probes have been used to visualize mRNA expression in cells by the technique of in-situ hybridization (see Fig. 6). One of the first successful demonstrations of the site of production of human placental lactogen mRNA utilized, as the probe, labelled cDNA prepared from enriched mRNA from term placenta (McWilliams & Boime, 1980). As human placental trophoblast tissue undergoes differentiation during pregnancy, cytotrophoblast cells develop into syncytiotrophoblast and the pattern of placental hormone secretion changes: maximum hCG synthesis occurs in the first trimester and maximum hPL synthesis at term. Immunohistochemistry indicated the syncytiotrophoblast was the site of hPL production and this was confirmed by in-situ hybridization; despite criticisms regarding the purity and specificity of a probe derived from heterologous mRNA, the localization of hPL mRNA to the syncytial layer of the villus was later confirmed using nick-translated cloned cDNA probes (Hoshina *et al.*, 1982). Probes for hCG- α and hPL were subsequently used to

Fig. 6. A demonstration of in-situ hybridization. Sections of mouse male submaxillary gland were hybridized with ^{35}S -labelled RNA transcripts encoding a portion of the androgen dependent mouse epidermal growth factor precursor (see Table 1). The sections were washed and treated with liquid emulsion for 13 days before development and were subsequently stained with haematoxylin and eosin: results of hybridization with (a) antisense RNA and (b) sense RNA transcribed from the cloned complementary DNA fragments encoding the EGF precursor (see Fig. 1). Note the presence of dense grains over the granular convoluted tubules when hybridized with the antisense RNA, only background hybridization being visible when the sense RNA was used as probe. $\times 75$. (Unpublished observations of R. M. Popkin & K. D. Brown.)

demonstrate different cellular localizations for these hormones: hCG- α mRNA was detected both in the syncytial layer and in differentiating cytotrophoblast, its concentration decreasing during pregnancy as a result of the progressive decline in cytotrophoblast. In contrast hPL mRNA was detected only in the syncytium, and levels remained constant throughout placental development (Hoshina *et al.*, 1982). Thus absolute levels of hCG α - and β -subunit mRNAs parallel the pattern of secreted hormone (Hoshina *et al.*, 1984, 1985), but changes in circulating concentrations of hPL reflect an increase in the number of cells producing mRNA rather than increased gene transcription within them.

In-situ hybridization has also been used quantitatively to estimate cellular levels of mRNAs for hCG- α , hCG- β and hPL after culture of tissue explants with steroid hormones. Progesterone selectively decreases concentrations of mRNAs for hCG- α and hCG- β (Maruo *et al.*, 1986).

Investigations into the localization of hPL and hCG mRNA species in placental neoplasms with altered trophoblast development suggested that PL mRNA expression was associated with villous development whereas that of hCG did not require such organization (Hoshina *et al.*, 1983). Aberrant expression of hormones produced by placental tumours may correspond to abnormal patterns of differentiation, with PL expression requiring an increased level of tissue differentiation to induce gene expression (see review: Hoshina *et al.*, 1985).

In theory the technique of in-situ hybridization provides a potentially valuable tool for the elucidation of the site of production of specific mRNAs and, in conjunction with immunohistochemistry, can be used to locate sites of gene expression and protein synthesis. In practice, there are relatively few convincing and well-controlled demonstrations of the localization of specific mRNAs by these methods, and those that exist are limited to mRNAs which provide the dominant transcriptional and therefore secretory product of that tissue (for instance, epidermal growth factor in the mouse submaxillary gland; Fig. 6). Further advances are needed to improve the signal-to-noise ratio and enable transcripts of less abundant mRNA to be detected; these advances are likely to involve improvements in the specificity of the probes, possibly by using RNA probes as described in Fig. 6, with increased duplex stability, or high-specific activity labelling of single stranded DNA probes. Encouraging results have been obtained, using a ^{32}P -labelled oligonucleotide from the LHRH precursor, which indicate the presence of cells containing LHRH mRNA in the rat fore-brain (Shivers *et al.*, 1986) in regions similar to those in which immunoreactive LHRH has been detected (Shivers *et al.*, 1983).

Pre- and post-translational processing

Physiologically active substances have classically been identified by investigating their effects in animals on administration of tissue extracts, but with the advent of recombinant DNA technology substances are now sometimes identified before their biological effects are known. Several peptides have recently been described which illustrate this, prominent examples being calcitonin gene-related peptide (CGRP) and gonadotrophin-releasing factor gene-associated peptide (GAP). These compounds have been predicted to exist on the basis of potential differential splicing of mRNA molecules and post-translational processing of proteins respectively.

Pre-translational processing: differential splicing of RNA

The discovery of introns in eukaryotic genes in 1978 (see Gilbert, 1978; Crick, 1979) was unexpected, since prokaryotic genes are always contiguous; but further surprises were in store. It soon became clear that mammalian viruses, particularly SV40, with their limited genome size, could produce different protein products from a single gene by differential splicing of the exons (Ziff, 1980). A similar differential use of exons was found in some immunoglobulin genes (Early *et al.*, 1980; Alt *et al.*, 1980), but it was argued that viral and immunoglobulin genes are atypical, and that this phenomenon was not a general property of mammalian genes. However, several examples of other mammalian genes showing tissue-specific alternative splicing have since been discovered (Young *et al.*, 1981; Nawa *et al.*, 1984; Jonas *et al.*, 1985; Takahashi *et al.*, 1985). One of the earliest and clearest demonstrations of alternative splicing was in the calcitonin gene. The two products, calcitonin and calcitonin gene-related peptide (CGRP) are produced in different tissues (Rosenfeld *et al.*, 1981, 1983; Amara *et al.*, 1982); very little calcitonin is produced in the brain, but in thyroid C cells the calcitonin:CGRP ratio is about 95:1. Both peptides are coded for by the same gene, but by using different exons two distinct and physiologically significant products are produced, calcitonin being important in regulation of calcium metabolism (Austin & Heath, 1981) and CGRP as a potent vasodilator (Fisher *et al.*, 1983; Rosenfeld *et al.*, 1984; Brain *et al.*, 1985, 1986). An alternative mechanism for the production of 2 proteins from a single piece of DNA, by transcription from opposite strands (Adelman *et al.*, 1987), has been dealt with above in relation to the LHRH gene.

Post-translational processing

Peptide hormones are generally produced as prohormones (or pre-prohormones) which are then processed proteolytically to yield bioactive peptides (see Douglass *et al.*, 1984 for review). A typical example is oxytocin, which will be described in detail.

Fig. 7. Structures of the arginine vasopressin (AVP) and oxytocin (OT) genes and prohormones. **(a)** Organization of human AVP-neurophysin and OT-neurophysin genes; the genes are transcribed from opposite DNA strands and are separated by approximately 12 kb. (After Sausville *et al.*, 1985.) **(b)** Structures of AVP and OT genes (i) prohormones (ii) and processing products (iii). Both AVP and OT genes contain 3 exons of comparable lengths; intron lengths differ between AVP and OT. The 2nd exon (shaded) is highly conserved, and codes for the major portion of the neurophysins (amino acids 22–88); the AVP or OT moiety (black) of the prohormone is coded for by exon 1. Amino acids 10–12 (Gly-Lys-Arg) constitute a post-translational processing signal. Note presence of a glycoprotein formed by post-translational processing of the AVP prohormone. The putative signal peptide (SP) occupies amino acids –19 to 1. Data for bovine genes redrawn from Ruppert *et al.* (1984).

Oxytocin has long been recognized to be important in reproductive endocrinology by virtue of its roles in stimulating uterine contractions during labour and in the milk ejection reflex during suckling. Recently, however, both oxytocin and the related neurohypophysial peptide, vasopressin, have been assigned new roles in reproduction. In the sheep corpus luteum oxytocin is found in high concentrations (up to 10 µg/g wet weight) and its secretion is stimulated by prostaglandin F-2α (Flint & Sheldrick, 1982). There is evidence that oxytocin plays a significant role in luteolysis in ruminants (Flint *et al.*, 1986) and that the corpus luteum is the source of the peptide; indeed the mRNA for oxytocin-neurophysin has been cloned from corpora lutea of cattle (Ivell & Richter, 1984; Fig. 7). Vasopressin has also been detected (in some cases with the corresponding mRNA) in the corpora lutea and testes of cattle, rats and man (Clements & Funder, 1986).

Oxytocin and vasopressin have been studied for many years and much information about their biosynthesis has been obtained, in particular by pulse-chase experiments (see Breslow, 1979). However, as late as 1980 (Brownstein *et al.*, 1980), there was doubt as to the structure of the vasopressin-neurophysin precursor. Then, in 1982, Land *et al.* published the nucleotide sequence of the vasopressin-neurophysin precursor cDNA. Cloning and sequencing proved that vasopressin

and oxytocin (which was cloned at the same time) were both synthesized as part of larger precursors and demonstrated that both oxytocin and vasopressin display many of the features commonly involved in protein processing (Wold, 1981). At the NH₂-termini of both precursors are signal peptides thought to be involved in the translocation of prohormones into endoplasmic reticulum for subsequent transport to the Golgi apparatus. During this translocation the signal peptides are removed proteolytically (Austen, 1979; Kreil, 1981; Schmale & Richter, 1981). Cleavage seems to take place at an alanine residue and thereby conforms with over 50% of other known signal peptide cleavage sites (Austen, 1979). The cleavage of vasopressin (or oxytocin) from the neurophysin occurs at a pair of basic amino acids Lys-Arg (position 10–11) and again this is a common feature (Nakanishi *et al.*, 1979; Amara *et al.*, 1980; Herbert, 1981; see Figs 4 & 5). At position 10 in the precursors of both oxytocin and vasopressin is a glycine residue, which serves as an amide donor for amidation of the nonapeptide (Suchanek & Kreil, 1977; Nakanishi *et al.*, 1979; Amara *et al.*, 1980). Processing in the corpus luteum is identical to that in the neurohypophysis (Clamagirand *et al.*, 1987). In the case of vasopressin, there is a single basic amino acid (arginine) separating the neurophysin from the glycopeptide. Since both the glycopeptide (Holwerda, 1972; Seidah *et al.*, 1981) and neurophysin (Chauvet *et al.*, 1981, 1983) can be purified from a number of species this is obviously sufficient for cleavage. However, cleavage does not occur at all apparently appropriate sites; there is a pair of arginine residues (position 105/106) which are not cleaved. Obviously our understanding of site selection is incomplete. The precursor for vasopressin is longer than that for oxytocin, reflecting the fact that it contains a glycopeptide of unknown function. The carbohydrate is specifically added to an asparagine residue at position 114 which is part of a well-recognized glycosylation site, Asn-X-^{Ser} - ^{Thr} (Holwerda, 1972; Pless & Lennarz, 1977; Seidah *et al.*, 1981). As indicated above this glycopeptide is of unknown function but cleavage products derived from it have been isolated (Smyth & Massey, 1979).

An example of the use of recombinant DNA techniques to predict the existence of a hormone from a sequence of DNA, and a knowledge of post-translational processing signals, is gonadotrophin-releasing hormone-associated peptide, GAP (see p. 12, Fig. 4). In this case the predicted product has been synthesized in a bacterial expression system and used to investigate the potential physiological role of the peptide (Nikolics *et al.*, 1985). Synthesized in this way GAP has been reported to cause gonadotrophin release from cultured pituitary cells and to inhibit prolactin secretion (Nikolics *et al.*, 1985); it has been suggested that GAP represents a naturally occurring prolactin-inhibiting factor (PIF). However, it must be emphasized that the physiological significance of this peptide is as yet undetermined although its apparent presence together with LHRH in secretory granules in cells of the median eminence suggests these hormones are co-secreted (Phillips *et al.*, 1985).

A further, less well understood, example of the post-translational processing of polypeptide hormones is the removal of C-peptides. These are relatively large peptide sequences separating two polypeptide chains present in the mature hormone. C-peptide removal constitutes an important step in the synthesis of the related hormones relaxin and insulin, and recombinant DNA techniques have been used to determine the sequences of the α - and β -chains of the active form of relaxin, as well as of the C-peptide (Hudson *et al.*, 1983, 1984).

Steroidogenic enzymes and steroid sulphatase

Some of the most striking actions of gonadotrophins are exerted on the synthesis of enzymes catalysing steps in steroid biosynthetic pathways in gonadal tissues, and advances have been made recently in the cloning of cDNA molecules coding for these enzymes. Attention has been directed particularly to the family of related cytochrome P-450 containing enzymes which catalyse cholesterol side-chain cleavage (SCC), 17 α -hydroxylase (17 α) and aromatase, as well as similar enzymes only present in the adrenal cortex (11 β - and 21-hydroxylases).

The first cytochrome P-450 cDNA to be cloned was that coding for the mitochondrial cholesterol side-chain cleavage form of the enzyme (Morohashi *et al.*, 1984). Six cDNA clones were identified in a cDNA library prepared from bovine adrenal poly(A⁺)-RNA using two oligonucleotides synthesized on the basis of a reported partial sequence of cytochrome P-450_{SCC}. Of these, three clones containing long cDNA inserts were sequenced and the identity of the sequences confirmed by comparison with the reported partial amino acid sequence. The cDNAs coded for a pro-protein of M_r 60 300 and predicted the sequence of a 39-amino acid N-terminal signal peptide known to be involved in transporting the protein through the mitochondrial membrane. This peptide is lost in the mature form of the protein (M_r 56 400). Comparison of the predicted amino acid sequence of this mammalian cytochrome P-450_{SCC} with mammalian microsomal P-450s and bacterial cytochrome P-450s indicated a closer homology to the microsomal enzymes, despite the fact that the microsomal P-450s receive electrons from NADPH via a flavoprotein in a manner different from those in mitochondria. (The bacterial enzymes resemble those in mammalian mitochondria in this respect; Sato & Omura, 1978.) This suggests that divergent evolution leading to bacterial cytochrome P-450 and a eukaryotic precursor of both mitochondrial and microsomal P-450s preceded divergence of the mitochondrial and microsomal forms.

Two overlapping cDNA clones specific for bovine cytochrome P-450_{SCC} with inserts of 425 and 950 bp have also been identified by John *et al.* (1984). Northern blots of RNA from bovine corpus luteum and adrenal cortex showed that these clones hybridized with an mRNA molecule of ~2000 bases, synthesis of which was induced by ACTH in adrenal cortical cells in culture. Southern blots of bovine genomic DNA using these probes suggested cytochrome P-450_{SCC} may be encoded by a single gene.

The best characterized microsomal cytochrome P-450 so far studied is that catalysing 17 α -hydroxylation (Zuber *et al.*, 1986). This protein also catalyses C-17,20 lyase activity (Nakajin & Hall, 1981). cDNAs were selected from a bovine adrenal cDNA library using poly(A⁺)-RNA enriched in mRNA coding for cytochrome P-450_{17 α} and P-450_{SCC}. Of 150 such clones, two were found to be specific for cytochrome P-450_{17 α} by in-vitro translation and immunoprecipitation using an antibody directed against purified P-450_{17 α} , and one of these contained the entire sequence coding for the protein of M_r 57 000. The translation initiation codon was assigned by comparison of amino acid compositions of a number of purified cytochrome P-450_{17 α} s with the putative product. The complete 1790 bp sequence consisted of a 46 bp 5'-untranslated region, a 1527 bp coding sequence, 147 bp of 3'-untranslated sequence and a poly(A) tail, and included a 69 bp region conserved in different P-450s from a variety of species and tissues and an N-terminal hydrophobic region characteristic of microsomal enzymes. There is a good deal of homology between cytochrome P-450s from different species and this cloned cDNA has been shown by Northern blotting to detect mRNAs of ~1850 bases in bovine adrenal and human fetal adrenal. Hybridizing mRNA was induced in bovine adrenal cortex cells in culture by both ACTH and dibutyryl cyclic AMP (Zuber *et al.*, 1986). The availability of this cDNA clone potentially provides a means for the antenatal diagnosis of congenital adrenal hypoplasia in man, one form of which is due to 17 α -hydroxylase deficiency. It should also stimulate a more detailed investigation than has previously been possible of the role of 17 α -hydroxylase in the processes of placental preparation for parturition in ruminants and in follicular maturation, in both of which this enzyme plays important roles (Flint & Ricketts, 1979; Bogavich & Richards, 1982).

In addition to these steroidogenic enzymes, a cDNA encoding placental steroid sulphatase has been cloned (Yen *et al.*, 1987). This is of particular interest because congenital absence of this enzyme leads to sulphatase deficiency, a syndrome notable for its effects on circulating oestrogen concentrations in pregnant women carrying affected fetuses. The availability of a cDNA probe has resulted in the chromosomal localization of the gene (confirming its X-linkage) and the identification of genomic deletions in affected subjects.

Table 1. Genes activated by steroid receptors*

Steroid hormone	Target organ	Protein (gene)	Reference
Oestrogen	Chicken oviduct	Egg white proteins†	Moen & Palmiter (1980) and references therein
	Chicken liver	Vitellogenin	Tata & Smith (1979)
	MCF-7 human breast cancer cells	pS2	Chambon <i>et al.</i> (1984)
	Rat and mouse pituitary	Prolactin	Stone <i>et al.</i> (1977) Stanley <i>et al.</i> (1986)
Progesterone	Rabbit endometrium	Uteroglobin	Bullock <i>et al.</i> (1976) Beato <i>et al.</i> (1980)
	Chicken oviduct	Avidin	Chan <i>et al.</i> (1973)
Androgen	Rat prostate gland	Prostate steroid binding protein	Parker <i>et al.</i> (1984)
	Mouse salivary gland	Epidermal growth factor‡	Gubits <i>et al.</i> (1986)
Ecdysone	<i>Drosophila</i> larvae salivary gland	Puff 74F gene product	Möritz <i>et al.</i> (1984)

*This list is not intended to be complete; further examples are given by O'Malley *et al.* (1979).

†The egg white protein genes are also activated by the other steroid hormones to various extents (see Hager *et al.*, 1980), and uteroglobin expression is stimulated by glucocorticoids in the rabbit lung (Cato *et al.*, 1984).

‡See Fig. 6.

Steroid hormone receptors

The fact that gonadal steroid hormones control secondary sexual characteristics such as beard growth, thickening of vocal cords and breast development has been known since the purification of these compounds and their structural elucidation some 50 years ago. The mechanisms by which they act are less certain. What is known is that steroid hormones affect the transcription of certain genes, thereby controlling the rates of synthesis of specific proteins. Table 1 lists some of the proteins, the synthesis of which is controlled by steroid hormones.

In some special cases, steroid hormones in high concentrations can act directly on cell membranes; for instance, progesterone is thought to exert its effect on ovum maturation in this way (see Heap & Flint, 1979). But the mechanism by which steroids usually control gene expression in target organs involves a 2-step process whereby the steroid binds initially to a receptor protein which subsequently interacts with nuclear DNA, increasing or decreasing rates of transcription of specific genes. As a result of the application of immunocytochemical and enucleation techniques it is now known (at least for the oestrogen receptor) that the receptor is primarily located in or on the nucleus before exposure to steroid (King & Greene, 1984; Welshons *et al.*, 1984). Specificity in terms of the identity of the steroids eliciting a response can be accounted for by the specificity of the binding reaction between steroid and receptor. On the other hand, why a steroid stimulates one response in one target organ while having a different effect on another is not established. Why for instance do oestrogens stimulate secretion of egg white protein in the avian oviduct and synthesis of egg yolk protein precursor in the liver? This will probably only be understood as a result of studying the interactions between a receptor and the genes it activates.

Our understanding of steroid hormone receptors and their mechanism of action has recently been advanced by information on their structure derived from the sequencing of cDNA clones and by the elucidation of the sequences of genomic DNA to which they bind. These advances arose from the purification of receptor proteins (which was expedited by the development of labelled photoaffinity ligands, enabling receptors to be labelled covalently with radioactive steroids), and by the subsequent production of anti-receptor antibodies. These achievements led to determination of

partial amino acid sequences (which in turn allowed the production of synthetic oligonucleotide probes) and to the identification of cDNA clones in libraries derived from target cell mRNA in expression vectors, particularly λ gt 11.

Steroid receptor primary structure: amino acid sequences derived from cloned cDNA

There are 4 steroid receptors which mediate the effects of the 4 classes of steroid hormone involved in reproduction: oestrogens, progestagens, androgens and glucocorticoids; glucocorticoids are included as reproductive hormones because of their effects on placental enzymes controlling progesterone and oestrogen secretion at parturition in ruminants (see Flint & Ricketts, 1979). These receptors are generally large proteins (glucocorticoid receptor, 94 000 M_r ; oestrogen receptor, 65 000 M_r) that represent only a small proportion (10^{-4} – 10^{-5}) of cell protein: there are probably only 10^3 – 10^5 molecules of oestrogen receptor in each cell of a responsive tissue. Some steroid receptors are thought to exist in more than one form; a notable example is the progesterone receptor of mammalian and avian uteri (see Lamb *et al.*, 1986, and references therein). The question of whether these multiple forms arise from different genes, or as proteolytic processing products of a single gene, is presently under investigation using recombinant DNA techniques (see Jeltsch *et al.*, 1986). The biology of the glucocorticoid receptor has been reviewed recently by Gustafsson *et al.* (1987).

Amino acid sequences of the oestrogen, progesterone and glucocorticoid receptors have been deduced by sequencing cDNA clones (Hollenberg *et al.*, 1985; Greene *et al.*, 1986; Green *et al.*, 1986; Jeltsch *et al.*, 1986; Conneely *et al.*, 1986) and this has led to a number of remarkable conclusions. Perhaps most significant is the fact that the sequences of these three receptors contain regions of considerable homology, particularly in what is thought to be the DNA binding domain of the molecule. A second surprising finding is that the same region bears a striking resemblance to part of the sequence of an oncogene product, the v-erb A protein, the cellular homologue of which (c-erb A) codes for the thyroxine receptor (Sap *et al.*, 1986; Weinberger *et al.*, 1986). Details of the sequencing of these receptors are as follows.

The sequence of the human oestrogen receptor, which is shown in Fig. 8, was obtained from cDNA clones corresponding to all or part of the mRNA, which were prepared from MCF-7 human breast cancer cells (Green *et al.*, 1986; Greene *et al.*, 1986). Clones were identified by screening with synthetic oligonucleotides derived from peptide sequences obtained from purified MCF-7 oestrogen receptor (in a λ gt 10 cDNA library) or by using monoclonal antibodies against receptor to identify expression products (in λ gt 11 cDNA expression libraries). No single cDNA has been isolated which is sufficiently long to code for the entire human oestrogen receptor mRNA but the whole mRNA sequence has been built up from several overlapping cDNAs (Green *et al.*, 1986). The sequence (Fig. 8) is 6.3 kb long, with a relatively long untranslated 3'-region of 4.3 kb; the 5'-untranslated region has 232 nucleotides. The sequence contains an open reading frame of 1785 nucleotides corresponding to 595 amino acids with a calculated molecular size of 66 182, which is in good agreement with the size of the naturally occurring receptor determined electrophoretically. The translation initiation codon was assigned to an ATG triplet immediately downstream from a TGA terminator at positions –54 to –52, on the basis of the amino acid sequence of a cyanogen bromide fragment of the purified MCF-7 receptor corresponding to amino acid residues 12 to 26. That the proposed primary structure represents the receptor is confirmed by the fact that it contains amino acid sequences corresponding to other fragments of the receptor already established. Furthermore, the cDNA was shown to code for receptor capable of binding [3 H]oestradiol by inserting it into an expression vector and transfecting eukaryotic cells (Green *et al.*, 1986; Greene *et al.*, 1986). A human cDNA clone has subsequently been used to probe a chicken oviduct cDNA library, leading to sequence analysis of the chicken oestrogen receptor (Krust *et al.*, 1986).

Partial sequences for the chicken oviduct progesterone receptor have been obtained from cDNA clones expressed in λ gt 11 and identified using antibodies raised against the purified receptor

[illegible]

Fig. 8. Sequence of the human oestrogen receptor mRNA from MCF-7 breast cancer cells. Indicated by boxes are (from top to bottom): the CAG cap site; the ATG translation initiation codon; the DNA-binding domain (amino acids 179–262); the steroid-binding domain (amino acids 301–552); the TGA termination codon, and three putative polyadenylation signals. Note the short sequence between initiation and termination codons at nucleotides 119 and 179 and the relatively long 3'- non-coding region. This sequence was built up from a number of short cDNA clones. (After Green *et al.* 1986)

(Jeltsch *et al.*, 1986; Conneely *et al.*, 1986). In the study of Conneely *et al.* (1986), a cDNA expression library was constructed from size-fractionated (>2 kb) poly(A⁺)-RNA from the chick oviduct. Screening 3×10^6 recombinants with one of the antibodies led to the identification of 3 positive recombinants of 0.8, 1.1 and 1.3 kb respectively, the largest of which contained a nucleotide sequence coding for a 36-amino acid fragment also identified in the purified receptor. By using the 1.3 kb cDNA fragment as a hybridization probe, mRNA species of sizes 4.5, 4.0 and 3.9 kb were demonstrated by Northern blot analysis in chick oviduct. Hybridizing RNA was undetectable in brain, lung, liver or kidney, and the oviduct mRNAs were inducible with oestrogen pretreatment. Jeltsch *et al.* (1986) used similar methods to isolate short cDNA clones from a chicken oviduct λ gt 11 library and subsequently utilized these to identify overlapping cDNA clones in λ gt 10 covering a total of 1.6 kb of the receptor sequence, including the putative DNA binding site.

The amino acid sequence of the human glucocorticoid receptor has been predicted from the sequence of 4 cDNA clones isolated from fibroblasts and the lymphoid cell line 1M-9; a short cDNA isolated from a 1M-9 cell cDNA library, constructed using the expression vector λ gt 11 and identified by screening with a polyclonal rabbit antiserum against the receptor, was subsequently used to identify longer cDNAs in a fibroblast library in a plasmid vector. These overlapping clones were sequenced to yield the entire 4800-nucleotide mRNA sequence coding for the α -form of the receptor (see below). The amino acid sequence was derived after assigning the positions of the translation initiation site (the methionine codon at positions 133–135, the first ATG downstream from an in-frame terminator TGA at positions 121–123, which must be assumed in the absence of any information on the N-terminal amino acid sequence), and the translation termination codon TGA (positions 2464–2466); the predicted size of the molecule (777 amino acids) accounts for the known molecular size (M_r 94 000). Further evidence that the molecule sequenced represents the appropriate receptor was obtained by synthesizing large quantities of the message using the transcription vector pSP65 and subsequently translating the mRNA molecules formed in a rabbit reticulocyte lysate cell-free translation system in the presence of labelled amino acids; the products obtained were identical on SDS-gel electrophoresis to the 94 000- M_r receptor. Furthermore, the in-vitro translated products bound labelled glucocorticoid with a steroid specificity identical to that of the native receptor. In addition to the α -form of the glucocorticoid receptor, a second, β -form was identified in the studies described above. In-vitro expression of the β -form showed it to lack steroid binding activity (Hollenberg *et al.*, 1985).

Sequence homologies between the human and chicken oestrogen receptor, the chicken progesterone and human glucocorticoid receptors and the v-erb A and c-erb A proteins are illustrated in Fig. 9, and these emphasize the structure/function relationship inherent in the molecules. It is clear that these 3 steroid receptors have evolved from a common ancestral protein. Of particular interest is the similarity between cysteine-lysine-arginine rich areas of the receptors (amino acids 185–250 of the human oestrogen receptor, 421–486 of the glucocorticoid receptor) where all of the 9 cysteine residues are conserved in the four receptors and the oncogene product. If the sequences compared in Fig. 9 are used to align the whole receptor primary structures, it can be seen that the N-terminal part of the glucocorticoid receptor, which is usually considered the immunogenic domain, is absent from the oestrogen receptor; it is possible that this part of the molecule is important in conferring specificity of action. The highly conserved sequences (region C in Fig. 9) are the DNA-binding domains in the receptors and in the c- and v-erb A oncogene products: when this region of the oestrogen receptor is replaced, using site-directed mutagenesis, with that of the glucocorticoid receptor, the resulting chimaeric receptor molecule activates glucocorticoid-responsive genes on exposure to oestradiol (Green & Chambon, 1987). The C-terminal ends of the receptor molecules (region E in Fig. 9), which represent the steroid-binding domains, contain relatively hydrophobic regions, but it is not possible on the basis of information currently available to deduce exactly the position of the steroid binding site.

[illegible]

Fig. 9. Amino acid sequence homologies and domain structures in steroid receptors. **(a)** Comparison of human and chicken oestrogen receptors (hER and cER), human glucocorticoid receptor (hGR), the v-erb A protein (v-erb A) and the chicken progesterone receptor (cPR). Structures have been divided into 6 domains (A–F); domains C and E are responsible for DNA- and steroid-binding respectively. Figures above and below lines indicate amino acids numbered from the N-terminal ends of the proteins. Values between lines are percentage amino acid homologies between pairs of proteins, for each domain. Deletions required for optimal alignment of the sequences are not shown. No calculation was made for the region shown by the broken line for hGR. Homology between cPR and hGR is established for domain C only. **(b)** Alignment of amino acids in the DNA-binding domains of cER, hER, hGR and v-erb A, compared with the cPR sequence, using the single letter code (see Table 5). Asterisks indicate amino acids conserved in all four receptors; hyphens indicate deletions assumed for optimal alignment. Data of Krust *et al.* (1986) and Jeltsch *et al.* (1986).

Steroid hormone receptors and oncogenesis

The involvement of steroid receptors in oncogenesis provides an important incentive to study these compounds, and nowhere is this better illustrated than in breast cancer. Edwards *et al.* (1979) have shown that approximately 50% of human breast cancers contain detectable levels of oestrogen receptor, and that of these about 70% will respond to anti-oestrogen treatment. By comparison only about 5% of tumours lacking oestrogen receptor respond to anti-oestrogens. Oestrogens have also been shown to be mitogenic in MCF-7 cells (Lippman *et al.*, 1976). Clearly, it would be of interest to know whether the homology between steroid receptors and v-erb A has any significance in determining the oncogenic effects of steroids such as oestrogen.

The v-erb A gene is expressed in avian erythroblastosis virus-transformed cells but the gene has no transforming activity alone; none the less it potentiates the transforming activity of v-erb B, and

may be responsible for the early blockade of erythroid cell differentiation (Frykberg *et al.*, 1983). The cellular erb A homologue, c-erb A, codes for a protein apparently identical to the thyroxine receptor (Sap *et al.*, 1986; Weinberger *et al.*, 1986); the v-erb B product is homologous to the epidermal growth factor (EGF) receptor, although it lacks the EGF-binding domain (Ullrich *et al.*, 1984). How v-erb A potentiates the action of v-erb B is unknown, as is the mechanism by which the oestrogen receptor causes increased growth of breast tumours; however, the demonstration (Knabbe *et al.*, 1987) that TGF β (see p. 13) is produced by MCF-7 breast cancer cells in response to anti-oestrogens, and that this factor inhibits cell growth, indicates that a connection exists between steroid hormones and cell surface receptors that may be important in this respect.

Mechanism of binding of receptors to DNA—the finger hypothesis

In assessing the molecular basis for the affinity of steroid hormone receptors for DNA, relevant information can be obtained from the structure of other proteins which bind DNA and regulate transcription. One such protein, transcription factor IIIA (TFIIIA), which has been analysed in detail (Miller *et al.*, 1985), provides an intriguing model. Transcription factor IIIA (also called factor A) is a 40 000- M_r protein required for the correct initiation of transcription (by RNA polymerase III) of the 5S RNA genes of the toad, *Xenopus laevis*. The protein, which has been purified from oocyte extracts (Engelke *et al.*, 1980) interacts with a sequence of approximately 50 nucleotides within the 5S RNA gene called the internal control region (Bogenhagen *et al.*, 1980). The initiation complex is stabilized by the subsequent binding of 2 other protein factors, B and C (Segall *et al.*, 1980; Bieker *et al.*, 1985). TFIIIA also binds to the 5S RNA particle, and in immature oocytes, which contain up to 20 000 5S RNA particles stored in the form of a 7S ribonucleoprotein particle, TFIIIA is a constituent of the 5S RNA particle (Honda & Roeder, 1980; Pelham & Brown, 1980). The suggestion has therefore been made that TFIIIA regulates transcription of the 5S RNA genes, stimulating transcription when not bound to 5S RNA particles.

Sequence analysis of TFIIIA reveals 9 internal amino acid repeat sequences, each consisting of approximately 30 amino acids and containing 2 invariant pairs of cysteine and histidine residues (Miller *et al.*, 1985). Since purified TFIIIA contains zinc, and cysteine and histidine bind zinc, it has been proposed that tetrahedral co-ordination complexes formed between these amino acids and Zn^{2+} ions introduce into the secondary structure 9 loops (referred to as fingers) which interact with DNA by binding through basic amino acid side chains (lysine, arginine, histidine, glutamine, threonine; Ohlendorf & Matthews, 1983). The binding may be principally through purine residues; on spatial grounds one would expect each finger to interact with DNA at 5 or 5.5 bp intervals, and there is a purine (particularly guanine) rich periodicity in the internal control region at, on average, 5.5 bp intervals. Binding to DNA presumably contorts or opens the double helix in such a way that it is recognized and more readily bound and transcribed by RNA polymerase.

It has been pointed out (Jeltsch *et al.*, 1986; Green & Chambon, 1987) that the steroid receptors may also have a structure conducive to formation of fingers. The DNA-binding domain of the receptors, which is rich in lysine and arginine, contains cysteine residues in positions appropriate for the formation of at least one, and possibly 2, fingers of a size similar to those of TFIIIA (Fig. 9), and when (in the oestrogen receptor) these cysteines are replaced by histidine residues, using site-directed mutagenesis, the resulting modified receptor molecule no longer stimulates gene transcription (Green & Chambon, 1987). Furthermore oestrogen receptor binding to DNA is reversibly blocked by phenanthroline, which chelates Zn^{2+} (Sabbah *et al.*, 1987).

Since the steroid receptors have evolved from a common ancestral DNA binding protein it might be expected that the 'finger hypothesis' would be more generally applicable. It has been suggested that other polypeptides form such structures, including 2 transcription-regulating proteins in yeast (see Jeltsch *et al.*, 1986). Miller *et al.* (1985) have proposed that the repetitive nature of the fingers has arisen through repeated duplication of an ancestral gene coding for a primitive DNA-binding protein consisting of a single finger structure. Furthermore, Moore *et al.* (1985) have

pointed out that the overall topography of the glucocorticoid receptor–growth hormone gene complex resembles those of the complexes formed between DNA and the prokaryotic repressor proteins *lac*, *λcl* and *λcro*. The processes involved in the interactions between DNA and regulatory proteins such as steroid receptors may, therefore, represent evolutionarily primitive mechanisms for controlling gene activity.

DNA sequences recognized by steroid receptors in eukaryotic cells

As outlined above, rates of gene transcription are controlled by events occurring in 5'-flanking regions ('promoter regions') of genes, and for this reason it is the 5'-region of the DNA which is thought of as the target for the receptor–steroid complex. A consideration of the effects of receptors on gene expression should therefore begin by examining the sequences of 5'-flanking regions of hormone-sensitive genes.

Several different techniques have been used to identify the genomic sequences to which steroid receptors bind (see Ringold *et al.*, 1983; Chambon *et al.*, 1984; Yamamoto, 1985). These include the construction of plasmids with altered or partly deleted 5'-regions from steroid-responsive genes linked to a gene with a readily identifiable product; on transfection into a steroid-responsive cell in culture the ability of steroid receptor to induce transcription can be correlated to promoter sequences included in the construct. A second type of method depends on digesting DNA with endonucleases; regions of genes 'activated' by interaction with receptors are susceptible (or hyper-sensitive) to DNase I, and so can be identified. DNase I can also be used to digest away those sequences of DNA that are not protected when isolated genomic DNA is incubated with receptor; the rationale behind this technique (called nuclease footprinting) is that receptor molecules protect the DNA from digestion by sterically hindering the nuclease. Chemical reagents, such as dimethyl sulphate, which methylates guanine residues, can be used in a similar manner. Thirdly, the binding of labelled receptor to various isolated DNA sequences can be tested by filtering through nitrocellulose filters using conditions under which only protein–DNA complexes bind to nitrocellulose; these can be identified by labelling the DNA or the receptor. The degree to which receptor binds can then be related to the DNA sequences used. Similar information can be obtained by binding calf thymus DNA to cellulose and incubating with labelled receptor; the receptor binds non-specifically to the DNA but binding can be inhibited by adding specific (cloned) genomic sequences. The higher the affinity of the receptor for them, the better the genomic sequences compete.

A method giving information similar to, but more detailed than, the nitrocellulose filter binding assay is the gel-shift assay (see Garner & Revzin, 1986). This technique depends on the altered mobility on electrophoresis of DNA fragments bound to protein, compared to the DNA alone. Labelled cloned DNA fragments are incubated with, for instance, receptor preparations (or crude nuclear extracts containing receptors) and loaded onto non-denaturing polyacrylamide gels. Non-specific binding is prevented by adding a competing nucleic acid such as poly(dI·dC) or salmon sperm DNA. Further competition studies can be carried out by adding unlabelled DNA.

These methods yield two related kinds of information. They can, initially, identify a region of DNA to which the receptor binds, for instance indicating that the responsive region is in a sequence between 100 and 200 bases upstream from the transcription start site (i.e. between bases –100 and –200). When this region is sequenced, and compared with similar data from other steroid responsive genes, it is frequently found that short sequences of 10 or 20 nucleotides are identical or related, in the regions upstream of a number of hormone sensitive genes, and it is thought that these are receptor binding sites.

Examples of some of the results obtained, in roughly chronological order, are as follows (see Table 2). In the early 1980s a number of reports appeared indicating that steroid hormone receptors interacted with 5'-regions upstream of the cap sites of genes. Grez *et al.* (1981) demonstrated that a region of partial homology between the genes coding for the oestrogen-induced egg white proteins, lysozyme, conalbumin and ovalbumin occurred in each case around 140 bp upstream from the

Table 2. DNA sequences recognized by steroid receptors

5'-ATC ^{CC} _{TT} ATT ^{CC} _T TCTG ^{CC} _T TTGTA-3'	Consensus sequence for 5 chick oviduct genes (up to -1 kbp)	(P)	Mulvihill <i>et al.</i> (1982)
TGTTCT	Mouse mammary tumour virus (-50 to -202 bp)	(G)	Scheidereit <i>et al.</i> (1983)
AGA ^{CC} _T CAG ^{CC} _T	Mouse mammary tumour virus (5 regions throughout genome)	(G)	Payvar <i>et al.</i> (1983)
GACCGGAGCTGAAAGAACAC	Chicken vitellogenin (-597 to -620 bp) (SV40 core enhancer)	(E)	Jost <i>et al.</i> (1984)
(GGTGTGGAAG)			
CGGT-ACACTGTGTCCT	Consensus sequence in human metallothionein-IIa gene	(G)	Karin <i>et al.</i> (1984)
TGTTCT	Rabbit uteroglobin (-2.6 kbp)	(G, in lung)	Cato <i>et al.</i> (1984)
GCACAATGTGTCCTG	Human growth hormone (+90-+120 bp)	(G)	Moore <i>et al.</i> (1985)
TGTTCT	Chicken lysozyme (-54 to -63 bp and -179 to -188 bp)	(P & G)	Renkawitz <i>et al.</i> (1984); von der Ahe <i>et al.</i> (1985)
TGTTCT	Rat tyrosine amino transferase (-2.5 kbp)	(G)	Becker <i>et al.</i> (1986)

Sequences present in 5'-regions of genes controlled by progesterone (P), oestrogen (E) and glucocorticoid (G) receptors, as identified in footprinting experiments, which are thought to represent binding sites for the steroid receptors. Vertical lines between sequences indicate homology; note that 6 of the nucleotides in the sequences for the mouse mammary tumour virus proposed by Payvar *et al.* (1983) and the vitellogenin gene (Jost *et al.*, 1984) are complementary to the hexanucleotide 3'-TCTTGT-5' and that the consensus sequence of Mulvihill *et al.* (1982) contains the related sequence, 3'-T^{CC}GTCT-5'. The chicken vitellogenin sequence shows some homology with the SV40 virus enhancer. Distances (in parentheses) are in bp or kbp and are 5'(-) or 3'(+) from the transcription start site. In some cases these sequences (particularly 5'-TGTTCT-3') appear several times in 5'-flanking regions and in both coding and non-coding DNA strands; in the case of the hGH gene the sensitive sequence is 90-120 bp downstream of the start site, in the first intron (see Fig. 3).

transcription start site. At about the same time, receptor-binding regions were identified in the long terminal repeats of mouse mammary tumour virus (MMTV) DNA, by using nitrocellulose filtration and electron microscopy (Govindan *et al.*, 1982), and by DNA/cellulose binding (Pfahl, 1982). This is important because transcription of the viral DNA is activated by glucocorticoid (see Ringold *et al.*, 1983). The progesterone receptor was also shown to bind to cloned 5'-fragments from the ovalbumin gene, by nitrocellulose binding (Hughes *et al.*, 1981; Compton *et al.*, 1982, 1983; Dean *et al.*, 1983). Subsequent experiments have been aimed at identifying specific sequences within these 5'-regions, and have led to the identification of consensus sequences listed in Table 2. In several genes the hexanucleotide 5'-TGTTCT-3' or a complementary sequence has been identified. Characteristics of this and related sequences are that they can occur a number of times in either strand of the double-stranded DNA at variable distances from the transcription start site (2.5 kb and 2.6 kb upstream in the case of the rat tyrosine amino transferase and rabbit uteroglobin genes: Cato *et al.*, 1984; Becker *et al.*, 1986; and in the first intron in the human growth hormone gene: Moore *et al.*, 1985). These characteristics identify the sequences as enhancers.

One surprising finding arising from this work is that genes activated by different receptors appear to share common enhancer sequences, and although this is not unexpected in view of the conservation of structure of the DNA-binding domains of receptor molecules, it does raise the question of the specificity of response to steroid hormones. How is it, for example, that the oestrogen receptor does not activate glucocorticoid-responsive genes? The answer probably lies in differential susceptibility of genes in different target organs. The process of organ differentiation involves the unmasking of certain genes, making them susceptible to transcription, and the blocking of others, preventing their transcription. Evidence that this is so can be obtained by studying the nuclease-sensitivity of genes in different organs—sensitivity to DNase I indicates that

a gene is in a state in which the DNA can interact with proteins (either RNA polymerases or DNase). Thus the ovalbumin gene in the chicken oviduct (and neighbouring sequences in a domain of 100 kbp) can be shown to be at least 2.5 times more susceptible to nuclease attack than the same sequence of DNA in other organs (Lawson *et al.*, 1980) and similar results have been obtained for the *Xenopus* and chicken vitellogenin genes (Gerber-Huber *et al.*, 1981; Williams & Tata, 1983). The ovalbumin gene is also more susceptible than other genes in the oviduct that are not transcribed (e.g. the globin gene). These relative differences in DNA sensitivity are unrelated to hormone treatment or exposure and are present regardless of whether the genes are ever expressed. They may be related to loops of DNA identified as arising from chromosomes (see Gasser & Laemmli, 1987). Such structures are presumably a product of the process of cellular differentiation occurring in the early embryo and remain imprinted in the structure of the chromatin throughout the life of the organism.

One way in which differential susceptibility to transcription may be determined is through DNA methylation. Cytosine residues are frequently methylated, most particularly in the sequence 5'-CG-3'; between 2 and 7% of cytosines in mammalian DNA occur as 5-methylcytosine, and 90% of these are in this dinucleotide. Once CG sequences are methylated the information thereby imparted in them is retained, since mechanisms exist for ensuring that the same bases are methylated following DNA replication (Bird, 1978; Harland, 1982). The extent to which genes are methylated can be assessed by use of the restriction enzymes *Msp* I and *Hpa* II, since whereas both these endonucleases cut C-C-G-G, only one (*Msp* I) cuts at C-MeC-G-G. When this method is used to determine the degree of methylation it is frequently (but not always) found that actively transcribed genes (such as ovalbumin in the oviduct) are under methylated compared to inactive ones (e.g. ovalbumin in liver) (Mandel & Chambon, 1979). The presence of 5-methylcytosine in DNA has been suggested to stabilize the helix, particularly sequences with a high proportion of purine/pyrimidine repeats, in the Z-form; this is more tightly wound than the usual B-form and may represent an alteration to secondary structure which, like nuclease sensitivity, has some significance in relation to the recognition of specific sequences by the gene transcription machinery. The specificity of an organ to a steroid hormone in terms of the genes activated may therefore lie not in the DNA sequence that the receptor recognizes, but in what genes bearing that enhancer sequence are susceptible to recognition or activation by the receptor.

Fertilization and early development

Techniques of molecular genetics have contributed to 2 areas of particular relevance to the study of early development—the questions of the onset of expression of the embryonic genome and the role of homeo-box genes in embryo development. It is appropriate to consider these in the order of their occurrence.

Expression of the embryonic genome

Maternal RNA accumulates in the egg during development and maturation of the oocyte. These RNAs code for proteins involved in the processes of meiosis and fertilization and in the maintenance of the cell during the period before activation of the embryonic genome. Examples of amounts of total RNA, poly(A⁺)-RNA and ribosomal RNA in the mouse oocyte and blastocyst before and after activation of the embryonic genome are given in Table 3; the quantities involved vary markedly between species. Synthesis of poly(A⁺)-RNA ceases in the 1-cell embryo, and the contents of total and poly(A⁺)-RNA in the developing embryo drop sharply by the late 2-cell stage (Pikó & Clegg, 1982; Table 3); it is thought that this is the stage when the majority of maternal RNA is eliminated (Bachvarova & De Leon, 1980). Not all mRNAs are degraded, however; injected haemoglobin message is translated in the 2-cell embryo (Chen & Brinster, 1982). It is at this

Table 3. Changes in total RNA content and its composition from the unfertilized mouse egg to the early blastocyst (data of Pikó & Clegg, 1982)

Stages after hCG injection	Total RNA (ng/cell)	Ribosomal RNA		Poly (A ⁺) RNA	
		ng/cell	% of total	pg/cell	% of total
Preovulatory oocyte	—	—	—	0.95	—
Unfertilized egg	0.35	—	—	0.70	0.20
Fertilized 1-cell	—	0.22	63	0.83	0.24
Late 1-cell to early 2-cell	—	—	—	0.39–0.78	—
Late 2-cell	0.12	0.09	70	0.13	0.11
8- to 16-cell	0.04–0.09	0.03–0.05	58	0.03–0.06	0.06
Early blastocyst (32-cell)	0.05	0.03	68	0.04	0.10

stage that embryos with incorrect numbers of chromosomes die (Gropp, 1974), suggesting that the embryonic genome is now active, and from the 2-cell stage onwards the amount of total RNA and poly(A⁺)-RNA increases rapidly (Pikó & Clegg, 1982; Clegg & Pikó, 1983).

The synthesis of embryo-coded proteins has been detected in mice in two phases (18–21 and 26–29 h after insemination) in which the appearance of new proteins is sensitive to the addition of α -amanitin (an inhibitor of mRNA synthesis; Flach *et al.*, 1982). These proteins, which are detected in the 2-cell embryo and synthesized only in the absence of α -amanitin, were identified as the mouse heat-shock proteins of M_r 70 000 (hsp 70) (Bensaude *et al.*, 1983; Morange *et al.*, 1984). The synthesis of specific isoenzymes, for example glucose phosphate isomerase (GPI) alleles, enables identification of their individual genomic origins. On Days 3 and 4 of mouse embryogenesis the synthesis of maternally and paternally derived GPI alleles occurs simultaneously (West & Green, 1983; Gilbert & Solter, 1985), but thereafter the maternally derived isoenzyme of GPI declines (Gilbert & Solter, 1985). However, paternal genes are activated and expressed in the mouse embryo as early as the 2-cell stage (Sawicki *et al.*, 1981; Surani, 1984) and, since eggs which contain 2 female pronuclei fail to develop after implantation, it appears that the male pronucleus is required for the normal development of extraembryonic membranes and trophoblast (McGrath & Solter, 1984; Surani *et al.*, 1986). Maternal and paternal gene activation may be regulated by methylation; some dispersed repetitive DNA sequences in oocytes are undermethylated while the same sequences are hypermethylated in sperm DNA; but both are less methylated than somatic tissue DNA, so differential methylation of the gametes could be the mechanism of imprinting (Monk *et al.*, 1987). It has been demonstrated that de-novo methylation of specific genes occurs during spermatogenesis, and sites of undermethylation in spermatozoa correspond to hypersensitive chromatin regions in somatic cells (Groudine & Conkin, 1985). Therefore, sites of undermethylation of sperm DNA may regulate paternal gene activation. Indeed patterns of methylation of DNA and chromatin structure (established by histones and other DNA-binding proteins) are stable over many cell generations and have been shown to be stably inherited in transgenic animals (Reik *et al.*, 1987; Sapienza *et al.*, 1987). It is likely that similar mechanisms control X-chromosome inactivation. The X chromosome from either parent is inactivated randomly in embryonic tissues but in extraembryonic tissue the paternal X chromosome is selectively inactivated in many species (Groudine & Conkin, 1985; Surani *et al.*, 1986). The inactivation of the X chromosome is linked to cellular differentiation, cells which differentiate early in development being imprinted with the origin of the X chromosomes. Inactivation of the X chromosome may not be complete in the female mouse embryo until 6–6.5 days of gestation; both X chromosomes are known to be active in the inner cell mass cells but only one is active in epiblast (fetal precursor) cells (Monk & Harper, 1979; Monk *et al.*, 1987).

It is possible to incorporate, into the fertilized eggs of mice, DNA from another species containing a gene not normally expressed in the mouse; e.g. giant transgenic mice that contain the human growth hormone gene have been produced (Palmiter & Brinster, 1985). Other examples are

the demonstration of the involvement of human growth hormone-releasing factor in somatic growth of mice (Hammer *et al.*, 1985), expression of human hypoxanthine phosphoribosyl-transferase in the central nervous system of mice (tissue specificity of expression; Stout *et al.*, 1985), and inherited limb deformity (pattern formation in the embryo; Woychik *et al.*, 1985). The use of this technique may be expected to lead to great insight into the control of gene expression in the embryo.

Pattern formation in early development

Embryo development has been well studied and is best understood in the fruit fly, *Drosophila melanogaster*. These studies were assisted by the large number of limb and other developmental mutants available in this rapidly-reproducing insect. By using these mutants significant progress has been made towards the identification of genes controlling development and pattern formation in *Drosophila* and, as we will see later, this has enabled similar genes to be identified in mammals. There are at least three classes of genes required for embryogenesis in *Drosophila*: (1) maternal effect genes; (2) segmentation genes, and (3) homeotic genes. Maternal effect genes appear to control the spatial co-ordinates of the egg and the deposition of maternal RNA, and will not be discussed in detail here. Segmentation genes control the number and polarity of the body segments.

The *Drosophila* embryo is organized into repeating units, each metameric segment being divided into an anterior and a posterior developmental compartment. These compartment types alternate, as revealed by cell lineage studies (Kornberg *et al.*, 1985) and by in-situ identification of cells expressing the *engrailed* locus, which is only expressed in the posterior compartment cells. The *engrailed* locus is required for segmentation of the embryo to proceed. Another segmentation gene, *fushi tarazu* (*ftz*) is expressed in the young *Drosophila* embryo in 7 evenly spaced bands of cells with two-segment periodicity (Hafen *et al.*, 1984); these loci can be used as molecular markers to follow the process of segmentation. In-situ hybridization using cDNA probes to these genes reveals that subdivision of the embryo proceeds through a sequence of intermediate stages which partition the embryo into progressively smaller units. The pattern of the *ftz* locus matures earlier than does the *engrailed* locus (Weir & Kornberg, 1985). Another gene, *hairy*, functions during segmentation of *Drosophila*, and is first detected throughout the embryo at the syncytial blastoderm stage and later develops a periodic localization. A correct transcription pattern of *hairy* requires maternally acting genes, as its expression is disrupted in maternal-effect mutations, which affect the deposition of maternal RNA in the embryo (Ingham *et al.*, 1985).

The homeotic genes appear to specify segment identity (Lewis, 1978) and occur in two main clusters in *Drosophila*: the bithorax complex, which specifies the thoracic and abdominal segments and the *Antennapedia* complex, which appears to control the development of the head and the anterior thoracic segments. During the mapping of *Antennapedia* (*Antp*—a homeotic mutant) weak cross hybridization was detected between *Antp* cDNA and other sites within the *Antennapedia* complex, revealing a repeated, homologous, DNA sequence of 180 bp, the homeo box. This sequence was mapped to the 3'-exons of the homeotic genes *Antp* and ultrabithorax (*ubx*) and of *ftz*, a segmentation gene (McGinnis *et al.*, 1984a; Gehring, 1985); more than 10 copies of the homeo box have been characterized within the *Drosophila* genome, and in most cases these are in open reading-frame coding for a protein domain in which the amino acid sequence conservation is higher (up to 85%) than at the nucleotide level.

Using *Drosophila* homeo box DNA probes, cross-hybridizing sequences have been identified in frogs (Carrasco *et al.*, 1984), mice (McGinnis *et al.*, 1984b) and man (Levine *et al.*, 1984). There is 70–90% sequence conservation between homeo box sequences isolated from vertebrates and *Drosophila*. These findings suggest that by analogy with *Drosophila* some of the genes that control mammalian development may have been identified. Certainly the mammalian homeo boxes have features in common with *Drosophila* apart from sequence homology. They are present in human (Hauser *et al.*, 1985) and mouse (Colberg-Poley *et al.*, 1985; Hart *et al.*, 1985) DNA in clusters. The

two major loci in mouse are Hox-1 (Duboule *et al.*, 1986) and Hox-2 (Hart *et al.*, 1985), each containing 6 homeo boxes. Each *Drosophila* homeo box specifies only a portion (10–20%) of a larger protein product and, in general, in *Drosophila* and in mammals the bulk of the coding sequences associated with each homeo box are unique and are not conserved. There is one exception, a 63 bp protein-coding region 3' to a mouse homeo box (Mo-en-1) which is 75% homologous with a region immediately 3' to the *engrailed* homeo box and may therefore encode a product that functions in mice in a manner analogous to the *engrailed* gene product of *Drosophila* (Joyner *et al.*, 1985).

All homeo box DNA sequences described to date have been identified as RNA transcripts in embryonic and neonatal tissues, in support of their proposed role in mammalian development. Mouse homeo box genes have been detected as single transcripts and at a constant level in mouse embryos aged between Days 8 and 17 of gestation (Jackson *et al.*, 1985; Hauser *et al.*, 1985). An *engrailed*-like transcript was also expressed in mice over a similar period of gestation but with a peak at Day 23 (Joyner *et al.*, 1985). Using in-situ hybridization, Gaunt *et al.* (1986) have shown that expression of the mouse homeo box gene Mo-10 is spatially restricted in the developing embryo and that localization of expression is already evident within the germ layers (7.5-day embryo) before the body pattern is laid down (9.5-day embryo) (Fig. 10). Northern analysis of a human homeo box gene detects organ- and stage-specific patterns of expression of 5–9-week-old human embryos and fetuses (Mavilio *et al.*, 1986). Surprisingly, sequences of the M6-12 homeo box of mice have also been detected in the majority of adult mouse tissues; however, the pattern of expression in the adult is complex and differs from that in the embryo.

Wolgemuth *et al.* (1986) have isolated a recombinant, pHBT-1 from a mouse testis cDNA library by screening at low stringency with the *Antennapedia* gene of *Drosophila*. This may represent a new homeo box-containing gene expressed only in the testis. There is previous evidence for the expression of homeo box-containing genes in germ cell development, e.g. during oogenesis in lower vertebrates (Muller, M. M. *et al.*, 1984) and in mammalian embryonal carcinoma cells of germ cell origin (Colberg-Poley *et al.*, 1985). Expression of pHBT-1 was limited to male germ cell-containing tissues, since RNAs from juvenile and adult ovaries did not contain detectable amounts of the 1.4 kb transcripts. Expression of pHBT-1 gene was not detected in embryonic testis, or in testes of neonatal animals which contain germ cells up to the Type B stage of spermatogonial development. A role for pHBT-1 expression in the meiotic stages of male germ cell differentiation has been postulated. However, pHBT-1 has been identified as Hox 1–4 (homeo box 4 in the cluster of such sequences at the Hox 1 locus), and its expression observed in whole mouse embryos on Days 9–11 (Duboule *et al.*, 1986). A human homeo box (Hu1) is also implicated in germ cell differentiation as it is only expressed at a high level when undifferentiated teratocarcinoma cells are induced to differentiate (Hauser *et al.*, 1985). The function of the homeo box-containing proteins remains a mystery both for *Drosophila* and mammals. However, sequence homology with DNA-binding proteins has been observed (Laughon & Scott, 1984) and homeo box proteins have now been shown to be DNA-binding (Desplan *et al.*, 1985; Fainsod *et al.*, 1986).

Oncogenes and heat shock proteins

One of the major advantages of recombinant DNA techniques is the ease with which the distribution of a particular gene and gene product within the animal kingdom can be analysed once the gene has been cloned. There are two major groups of genes initially isolated and characterized from non-reproductive, non-mammalian tissues which are of interest in the study of reproduction. These groups are the oncogenes, identified in retroviruses as genes causing neoplastic transformation of infected cells (Bishop, 1978, 1985), and the heat shock proteins, identified in *Drosophila* as proteins which are induced during heat shock (for review: Ashburner & Bonner, 1979).

Oncogenes. Retroviruses (RNA tumour viruses) can induce virtually all major types of malignancy including carcinomas, sarcomas, leukaemias and lymphomas in vertebrates (Klein,

Fig. 10. In-situ hybridization of the mouse homeo box gene Mo-10 to the mouse 9.5-day embryo showing localization of labelling to the nervous system, with the exception of the anterior regions of the brain which were unlabelled. The division between labelled and unlabelled neural tissue was at the junction of two neuromeres in the midbrain. **(a)** Embryo cut in parasagittal section, wider bright-field illumination. hb, hindbrain; mb, midbrain; nt, neural tube; ov, otic vesicle; v2, second visceral arch; s, somite; am, amnion. Scale bar, 0.5 μ m. **(b)** Silver grains in the head region of the same section, viewed with wider dark-field illumination. (Reproduced, with permission, from Gaunt *et al.*, 1986.)

Table 4. Viral oncogenes and their products

Viral oncogene	Species of origin	Tumorigenicity	Protein products of cellular homologues	
			Biochemical function	Subcellular location
v-abl	Mouse and rat	β-Cell lymphoma	PK(tyr)	Plasma membrane
v-erb A	Chicken	(not tumorigenic alone)	Thyroid hormone receptor	Cytoplasm
v-erb B	Chicken	Erythroleukemia and sarcoma	EGF receptor, PK (tyr)	Membranes
v-fms	Cat	Sarcoma	Receptor for macrophage colony stimulating factor (CSF-1), PK(tyr)	Membranes
v-fos	Mouse	Osteosarcoma	Gene activation (?)	Nucleus
v-myc	Chicken	Carcinoma, sarcoma and myelocytoma	DNA binding, regulates transcription (?)	Nucleus
v-ras ^{Ha} } v-ras ^{Ki} }	Mouse and rat	Sarcoma and erythroleukemia	GTPase activity, homology with growth factor signal transducing G proteins	Plasma membrane
v-sis	Woolly monkey and cat	Sarcoma	β-Chain of PDGF	Secreted; cytoplasm (?)
v-src	Chicken	Sarcoma	PK(tyr)	Plasma membrane

PK(tyr), protein-tyrosine kinase; EGF, epidermal growth factor; PDGF, platelet-derived growth factor. (?) indicates uncertainty.

1980). All of the acutely transforming retroviruses are now known to carry specific genes, designated viral oncogenes (v-oncs), which are capable of inducing the malignant phenotype (Table 4). At least 16 such genes have been identified and isolated (Bishop, 1983, 1985). DNA sequences homologous to the v-onc genes have also been identified by Southern blotting in normal nontransformed cells from various uninfected vertebrates; these have been designated cellular oncogenes (c-oncs; Bishop, 1983; Duesberg, 1983). Existing evidence suggests that c-onc genes are the progenitors of v-onc genes, a phenomenon which is believed to have occurred by recombination events (Bishop, 1981; Duesberg, 1983). The fidelity with which the c-onc genes are conserved over a wide span of evolution strongly suggests that these genes serve critical physiological functions. The ability of viral oncogenes to cause uncontrolled cell proliferation has led to the suggestion that the physiological role of their cellular homologues is the control of growth and differentiation (Bishop & Varmus, 1982; Muller, 1986). However, the exact role of the c-onc genes has not been elucidated, although for some the nature of the gene product is known (Bishop, 1983, 1985; Table 4).

There have been several comprehensive analyses of c-onc transcription during mouse embryo development. Generally a cDNA oncogene probe, derived from cloned viral or cellular oncogenes, has been used to analyse dot blots and Northern blots of isolated total and messenger RNA. These studies have shown that the cellular homologues c-fos and c-abl (Muller *et al.*, 1982), c-fms (Muller *et al.*, 1983c), c-src, c-erb and c-sis (Slamon & Cline, 1984) and N-myc (a homologue of c-myc from human neuroblastomas; Jakobovits *et al.*, 1985) exhibit stage- and tissue-specific patterns of expression in mouse embryos, extraembryonic membranes and/or placenta (Figs 11 & 12). In contrast, c-ras^{Ha}, c-ras^{Ki} and c-myc are expressed at high levels in these tissues, but with limited tissue- and stage-specificity (Figs 11 & 12) (Ellis *et al.*, 1980, 1981; Muller *et al.*, 1983c).

Studies of human reproductive tissues have been less comprehensive, but stage-specific expression of c-fos and c-fms (Muller *et al.*, 1983a), c-myc (Goustin *et al.*, 1985; Pfeifer-Ohlsson *et al.*, 1984) and c-sis has been observed in human extraembryonic membranes and placenta. Expression of c-src was observed in specific neural tissues of the human embryo (Levy *et al.*, 1984).

Fig. 11. Relative amounts of c-onc expression during mouse embryonic development. These were determined by soft-laser densitometer scanning of dot-blot autoradiograms of total or poly(A⁺)-RNA derived from whole conceptuses (Days 6–9) or embryos/fetuses (Days 10–18) and hybridized to v-onc probes. Total RNA: c-myc, c-erb, c-src, c-sis and poly(A⁺) RNA: c-fos, c-abl. Values were normalized to 100 for the highest level of expression in the case of each c-onc. After Muller *et al.* (1982) and Slamon & Cline (1984).

Ubiquitous expression of c-ras^{Ki}, c-ras^{Ha} and, in one study, c-myc (Muller *et al.*, 1983a) was again observed.

The most interesting oncogenes from the reproductive viewpoint are c-fms and c-fos. Transcripts of c-fms, about 3.7 kb in size, are detected at high levels in the placenta and extra-embryonic membranes of pregnant mice and at low, constant levels in embryos and fetuses and in all postnatal tissues studied (Muller *et al.*, 1983c). The abundance of c-fms transcripts increases in parallel with placental development reaching a maximum at Day 14–15 of gestation (Fig. 12). At this time the concentration of c-fms transcript is 15-fold higher than in Day 7 to Day 9 conceptuses (which consist largely of tissues participating in the formation of the placenta). The pattern of c-fms expression during development of the extraembryonal membranes closely resembled the stage-specific pattern of expression observed during development of the placenta. Levels of c-fms in extraembryonal membranes at Day 18 were 6-fold higher than at Day 12 of gestation, and were one-half to one-third of levels in placenta at the same stage. It is suggested that c-fms is expressed at a high level in a limited number of cell types present in the placenta and the extraembryonal membranes, whose number increases as gestation proceeds. c-Fms transcripts were detected (Muller *et al.*, 1983a) in human term placenta, in the trophoblastic BeWo cell line (derived from a gestational choriocarcinoma of the human fetal placenta), and term fetal membranes. The concentration of c-fms in fetal membranes was about 20% that in placenta, which in turn was about 20% that in NIH/3T3 cells transformed with v-fms DNA (Donner *et al.*, 1982). The distribution of c-fms is similar in mouse and human, suggesting a crucial role in the nourishment and support of the developing embryo.

Embryonic expression of the oncogene c-fos has also been observed in mice. The 2.2 kb c-fos transcript was identified at high levels in Day 6–9 conceptuses, but in later embryos dissected their membranes it was lower by a factor of 10 (Muller *et al.*, 1982). This distribution is similar to c-fms and suggests that in the early stages c-fos is expressed in a discrete cell population which gives rise to the placenta; indeed on Days 10 and 18 c-fos expression is high in undissected placenta. Further studies investigated the expression of c-fos in the inner and outer portions of the placenta. The level of c-fos transcripts is 15 times higher in the separated outer portion of the mid-gestational

Fig. 12. Relative levels of c-fos (open), c-fms (hatched) and c-ras^{Ha} transcripts in the developing fetus (f), placenta (p) and extra-embryonic membranes (m) of the mouse. (a) Data derived as described in Fig. 11, i.e. total RNA for c-fms and c-fos (fetus and membranes) and poly(A⁺)-RNA for c-fos (placenta). (b) Data derived as described in Fig. 11 but without quantification. Poly(A⁺)-RNA used throughout. nd = not determined. Derived from Muller *et al.* (1983b, c).

placenta (primarily undifferentiated fetus-derived cytotrophoblast or maternal decidua) relative to the inner moiety (predominantly differentiated syncytiotrophoblast). The expression of c-fos was constant in the outer moiety but increased 5- to 10-fold from Day 12 to Day 20 in the inner moiety (Fig. 12). The level of c-fos transcripts in extraembryonal membranes was up to 50-fold greater than in the fetus and increased about 6-fold between Days 12 and 18 of gestation to a level 3-fold higher than that in placenta. No significant differences in c-fms expression could be detected in inner and outer placenta at all stages analysed. In human pregnancy c-fos transcripts were 10–100-fold higher in human term fetal membranes (chorion and amnion) and 3–20-fold greater in human term placenta than in other normal tissues (kidney, liver, lymph nodes, lung; Muller *et al.*, 1983b, c), the levels reached in these amniotic and chorionic cells being similar to those which result in sarcomas in mice (Curran & Teich, 1982). When primary amnion cells are cultured in normal growth medium the expression of c-fos protein rapidly drops to low levels, but expression of c-fos can be induced and maintained in these cells in the presence of dialysed placenta- or embryo-conditioned medium (Muller *et al.*, 1986). This suggests that c-fos expression in primary amnion

cells is regulated by placenta- and embryo-derived factor(s), and in fact high levels of c-fos expression in amniotic cells correlate with a high epidermal growth factor binding capacity (Adamson *et al.*, 1981).

The c-fos RNA transcripts which accumulate in mouse and human extra-embryonic tissues are translated in these tissues to give anti-fos-reactive proteins in the range M_r 55 000–60 000, as shown by protein blot analysis using an antibody prepared to a synthetic fos peptide (Adamson *et al.*, 1985). It is tempting to speculate that c-fos and c-fms proteins may be involved in the 'pseudo-malignant' properties of mouse and human trophoblast (Stromberg *et al.*, 1982). However, the fact that high levels of expression of both c-onc genes occur in late-gestation extra-embryonic membranes argues against such a function, since there is no evidence that constituents of embryonic membranes in late gestation exhibit properties characteristic of malignant cells. Alternatively, a common function of the c-fos and c-fms gene products in processes associated with the transport of nutrients would correlate with the tissue distribution of transcripts from these c-onc genes.

Expression of the oncogene c-myc in response to growth factor stimulation has been briefly mentioned above. There is evidence for an interaction between platelet-derived growth factor (PDGF) and c-myc expression during proliferation of the human placenta. Transcripts of c-myc, 2.4 kb in size, in the human placenta increase to a maximum during the first 28–35 days of pregnancy and then decrease to low levels by the end of the first trimester (Pfeifer-Ohlsson *et al.*, 1984). The distribution of c-myc transcripts in the early placenta determined by in-situ hybridization (see above) correlates with [3 H]thymidine incorporation (a measure of DNA synthesis and cell proliferation); both are most active in the cytotrophoblastic cells. Expression of another oncogene, c-sis, has also been localized by in-situ hybridization to the cytotrophoblast of human placenta, particularly to the highly proliferative and invasive cytotrophoblastic shell (Goustin *et al.*, 1985). The distribution of c-sis parallels that of c-myc in the early human placenta. Transcripts of c-sis are 4.2 kb in size and are homologous to the B chain of PDGF (Doolittle *et al.*, 1983; Waterfield *et al.*, 1983). The gene is actively transcribed and translated, as shown by the release of PDGF-receptor-competing activity into media conditioned by placental explants from the first trimester.

These observations, together with the presence of PDGF receptors on cultured cytotrophoblast cells (Goustin *et al.*, 1985) and the increase in c-myc and c-fos expression in response to stimulation of the cell cycle by growth factors including PDGF (Kelly *et al.*, 1983; Kruijer *et al.*, 1984; Hann *et al.*, 1985; Thompson *et al.*, 1985), led Goustin *et al.* (1985) to postulate an autocrine regulatory model for trophoblast proliferation. In this model, normal placental trophoblast synthesizes a growth factor (PDGF) capable of promoting its own proliferation through the induction of c-myc and c-fos expression. The ultimate signal controlling c-sis expression and hence growth is not known, but could be part of the internal programme of the developing cytotrophoblast or a maternal hormone signal. Evidence also shows that c-sis and c-myc gene expression overlap directly in whole human embryos of first trimester pregnancies (Goustin *et al.*, 1985), suggesting that autocrine regulation of cell growth may be a common control mechanism in early human embryogenesis.

Heat-shock proteins. Heat shock treatment (i.e. a rapid increase in ambient temperature) in *Drosophila* causes the production of a limited number of new proteins, whilst most RNA and protein synthesis stops (for review see Ashburner & Bonner, 1979). The induction of new protein synthesis by heat shock is observed in many organisms including bacteria, yeast, slime mould, protozoans, plants, insects, avian cells and mammalian cells. The pattern of proteins produced in different species when analysed by SDS-polyacrylamide gel electrophoresis varies, but most synthesize heat-shock proteins in the ranges M_r 60 000–70 000 and 80 000–90 000. The structure of the two major heat-shock proteins, M_r 70 000 and 89 000 (designated hsp 70 and hsp 89) were shown, by the use of specific antibodies, to be conserved across a wide range of species (Kelley & Schlesinger, 1982). The DNA sequences are also homologous. The function of these proteins is not clearly understood, but the low molecular weight heat-shock proteins of *Drosophila* are located in

the nucleus and appear bound to the nuclear matrix, as they remain insoluble after DNA and histones have been removed (Ellis *et al.*, 1980, 1981; Slamon & Cline, 1984; Jakobovits *et al.*, 1985), and may be involved in the organization of chromatin. Yeast cells also contain heat-shock proteins which bind DNA (Pfeifer-Ohlsson *et al.*, 1984). While one-third of hsp 70 is located in the nucleus, the rest is in the cytoplasm, where it appears to be a structural protein, as it is not associated with ribosomes or membranes.

Earlier in this review we discussed the activation of the embryonic genome in the late two-cell stage mouse embryo. Activation is preceded by the synthesis of proteins coded from maternal mRNA at the end of the one-cell stage, and the appearance at the early two-cell stage of a set of new polypeptides of M_r 70 000 (Flach *et al.*, 1982). This can be compared with the series of events that occur after hyperthermia in differentiated cells (Ashburner & Bonner, 1979; Schlesinger *et al.*, 1982). Indeed the early 2-cell-specific proteins of M_r 70 000 are identical to two of the mouse heat-shock proteins (hsp 68 and 70) when analysed by 2-dimensional polyacrylamide gel electrophoresis (Bensaude *et al.*, 1983; Wittig *et al.*, 1983). Fertilized oocytes incubated with α -amanitin divide normally into 2-cell embryos. However, hsp 68 and 70 are absent from the range of proteins synthesized by these embryos compared with untreated early 2-cell embryos (Bensaude *et al.*, 1983). As α -amanitin is an inhibitor of mRNA synthesis this suggests that heat-shock genes are amongst the first to be efficiently transcribed after fertilization. Heat-shock proteins may be required for the recovery of gene expression in heat-shocked cells (Arrigo, 1980), but in most cells studied hsp 70 is also synthesized at appreciable levels in the absence of stress (Kelley & Schlesinger, 1982). The only known exceptions are cells in which no transcription occurs, such as the mouse one-cell embryo (Bensaude *et al.*, 1983) and the mammalian erythrocyte (Morimoto & Fodor, 1984), which suggests that a degree of heat-shock protein expression is constitutive whenever transcription occurs. Does the one-cell embryo need to acquire heat-shock proteins to initiate embryonic gene expression?

Although mammalian embryonic heat-shock proteins are homologous to those of *Drosophila*, they do not appear to be true heat-shock proteins because they are not induced by hyperthermia in mouse embryos between the blastocyst stage and the 8-cell embryo (Morange *et al.*, 1984). However, similar genes, which are spontaneously expressed and are not activated by heat shock, have been identified in *Drosophila* and *Saccharomyces* (Ingolia & Craig, 1982; Ingolia *et al.*, 1982) and it is possible that these genes may be involved in the control of embryonic gene expression.

The heat-shock genes appear to be induced by certain gene products which have been implicated both in neoplastic transformation and in normal growth. The c-myc oncogene has limited homology with the adenovirus Ela protein (Ralston & Bishop, 1983); both gene products induce expression from a *Drosophila* hsp 70 promoter (Kingston *et al.*, 1984), although it is not known whether this transcription is stimulated from the normal hsp 70 promoter site. The adenovirus Ela protein also stimulates transcription from a wide variety of cellular and viral promoters including the mammalian hsp 70 gene. Clearly, the interaction between heat-shock proteins, oncogenes and the onset of growth and differentiation in embryos and other tissues is just beginning to be understood and research into the expression of genes which directly control mammalian development is now realistic.

The future

This review has highlighted with respect to a number of proteins how a knowledge of their amino acid sequences, deduced from DNA sequences, may lead to a better understanding of their functions and evolution. Structure-activity studies of this type may prove to generate the most valuable information which recombinant DNA technology can provide for reproductive physiologists. The ability to produce large quantities of a protein at will and to alter the protein sequence by site specific mutation of the gene encoding it will furthermore allow detailed study of how processes such as hormone-receptor interactions occur. Introduction of additional copies of genes or

mutated genes into mammalian embryos will permit the study of their effects on growth and development.

The ability to produce large quantities of polypeptide hormones from genetically engineered bacteria, yeast or cells in culture is potentially of immense value, as illustrated by the commercial production of human insulin and growth hormone. This is underlined by the use of recombinant inhibin A subunit to immunize sheep, thereby raising their ovulation rate (Forage *et al.*, 1987). The significance of this approach is evident from the sources of papers describing cloning of polypeptide hormones in journals such as *Nature* or *Science*, which are frequently commercial concerns. Although this may be frustrating to those in academic departments, it should be remembered that it is to the physiologist that the molecular geneticist will look for the study of the physiological effects of these products.

Recombinant DNA technology also has great potential in the antenatal diagnosis of genetic polymorphisms. This approach (known as restriction fragment length polymorphism) involves studying the distribution between related individuals of differences in the size of restriction fragments obtained from genes and non-coding flanking DNA implicated in disease states. Differences in sequences of identified genes between individuals are demonstrated following endonuclease digestion by Southern blotting. The technique has proved of particular value for a variety of haemoglobinopathies (see Little, 1981; Gosden & Gosden, 1985, for reviews), but probes are being developed every day, and the potential already exists for antenatal diagnosis of α_1 -antitrypsin abnormalities (Cox *et al.*, 1985), Duchenne muscular dystrophy (Monaco *et al.*, 1985) and cystic fibrosis (Knowlton *et al.*, 1985). The demonstration that hypogonadism in the *hpg* mouse can be reversed by using transgenic techniques (Mason *et al.*, 1986b) provides evidence of the therapeutic potential of these procedures. A further use of hybridization techniques is the application of DNA 'finger-printing' in the rapidly developing fields of paternity identification (Jeffreys *et al.*, 1985) and forensic science (Gill *et al.*, 1985).

In this article we have attempted to describe the contributions made by recombinant DNA technology to reproductive biology and to suggest some of the developments likely in the not too distant future. However, speculation is dangerous, and all that can be safely predicted is that whatever the outcome of research over the next 10–15 years it is likely to be an exciting time for reproductive biology.

We thank Dr R. S. Gilmour, Dr R. B. Heap, Dr J. R. Miller and Dr A. Stockell Hartree of the AFRC Institute of Animal Physiology and Genetics Research, Babraham, for commenting on the manuscript.

Appendix 1: Recombinant DNA techniques and procedures

To better understand the means by which the information reviewed here was obtained, it is necessary to describe in the briefest detail some of the procedures used. The reader familiar with the appropriate jargon may wish to leave this section. For those seeking a reminder there is also a glossary at the end of the article (Appendix 2), where words given in this section in *italics* (and elsewhere in the article) are defined. A bibliography is appended of textbooks and laboratory manuals.

To determine the sequence of a DNA or RNA molecule one must first clone it, i.e. make millions of copies of the same molecule, so as to be able to handle a homogeneous preparation of identical sequences. To clone a nucleic acid sequence the foreign DNA is *ligated* into a *vector*. The vector + foreign DNA is called '*recombinant*' (i.e. cut and recombined). Each bacterium *transformed* should only receive a single recombinant molecule, so that each colony derived from a single transformed bacterium represents a source of a single recombinant vector sequence. The recombinant DNA can then be extracted from the total bacterial DNA, purified, and the insert cut

out from it. An *E. coli* bacterium divides every 20 min under optimal conditions and therefore is capable of producing 10^9 progeny in about 20 h. Each progeny bacterium may contain 200 or more copies of a vector (several thousand if host protein synthesis is stopped), and so very large numbers of identical DNA molecules can be produced relatively simply by these means.

The developments in nucleic acid biochemistry and microbial genetics that made this cloning and sequencing procedure possible occurred during the two decades between 1960 and 1980 and sprang largely from the study of the properties of individual enzymes and their roles in natural processes. *Restriction enzymes*, the endonucleases used to cut double stranded DNA at specific sequences, were discovered by Lederberg & Meselson in 1964; they are called 'restriction' enzymes because they 'restrict' the ability of *bacteriophages* to replicate within the particular bacterial strain in which the enzymes are found. They represent part of a natural mechanism by which bacteria defend themselves against bacteriophage infection; so that a bacterium's restriction enzymes do not digest its own DNA, each restriction enzyme has a related methylase which methylates sequences susceptible to cutting in the bacterial genome, thereby protecting them from the cell's own restriction enzyme. Restriction enzymes are named after the bacterial species from which they are obtained, e.g. Eco RI is from *Escherichia coli* (Smith & Nathans, 1973). Our present ability to ligate pieces of DNA together depends on the availability of purified DNA ligase, an enzyme usually extracted from *E. coli* infected with phage T4; the ligase is produced by the bacteriophage as part of the natural mechanism by which its gene products ensure the repair of nicks in its DNA. Sequences of mRNA can be copied to produce *complementary DNA* by using the enzyme *reverse transcriptase* (an RNA-dependent DNA polymerase) extracted from cells infected with avian erythroblastosis virus; this virus is an RNA virus and requires the reverse transcriptase to convert its own genome into DNA, so that it can be incorporated into the host cell's genome and *replicated* with it.

Choice of vectors for DNA cloning is clearly important. To be useful vectors are required to replicate in a bacterial host; they should contain a sequence suitable for cutting, so that foreign DNA can be inserted into them; they should be selectable by some means, so that colonies containing the vectors can be identified; and there should be some means of introducing the vector into the host cell. The most commonly used vectors are *plasmids* and bacteriophages, and genetically engineered (i.e. synthetic) examples of both types of vector are now in widespread use. Both are, of course, based on naturally occurring species, and both meet the criteria listed above for useful vectors. Plasmids are introduced into bacteria (the process of transformation) after the bacterial cell walls are permeabilized by exposure to high concentrations of divalent cations, commonly CaCl_2 . Bacteriophage are introduced by exploiting their ability to infect their hosts. Plasmid and bacteriophage vectors are available containing synthetic sequences representing the sites of action of several restriction enzymes, in order to provide convenient points at which the vectors can be cut, and genes are introduced into them to allow them to be easily detectable; for instance most plasmids in common use contain genes for antibiotic resistance, frequently that coding for a β -lactamase that hydrolyses (and thereby neutralizes) ampicillin. This permits selection of colonies of transformed bacteria. They may also contain the β -galactosidase gene, coding for an enzyme that can readily be detected by a colour reaction, so that colonies of bacteria transformed with non-recombinant vector are blue when grown on a medium containing an appropriate substrate; if the cloning site is located in this gene, the insertion of foreign DNA into the plasmid prevents the production of β -galactosidase and recombinant colonies are white rather than blue, readily allowing their identification. The *packaging* of bacteriophage provides another means of selecting for recombinant molecules, since only phage DNA within a particular size range is an appropriate substrate for the packaging enzymes; genetically engineered phage are available with genomes too small to be packaged unless they contain additional (foreign) DNA, and so non-recombinant phage are not able to transfect bacteria. A wide variety of other techniques is available to allow selection of recombinants, and these are dealt with at length in standard texts (see bibliography).

Recombinant bacteria containing specific DNA sequences are also commonly detected by using DNA or RNA *probes*. These are usually short, sometimes synthetic, oligonucleotides that are

complementary to the sequences being sought, and their use depends on their ability to *hybridize* with (i.e. hydrogen bond to) the required sequences to form double-stranded DNA or DNA/RNA hybrids. If the probes are labelled in some way (most commonly by addition of radioactive nucleotides using the procedure of *nick translation* or by synthesis from radiolabelled nucleotides) the probes can be readily detected by autoradiography. Thus colonies containing the sequences being sought can be detected by transferring material from them onto a membrane layered over the Petri dish in which they are growing, lysing the bacteria (to release the DNA) and incubating with the probe. After washing off unbound probe, that remaining hybridized to material on the membrane can be detected autoradiographically. After detecting an appropriate colony, bacteria from it can be grown in liquid culture and the recombinant plasmid extracted and purified. The same procedure can be used to detect *plaques* formed when bacteriophage infect *lawns* of susceptible bacteria.

The ability of labelled probes to hybridize with complementary sequences can also be used to detect specific sequences of DNA or RNA in mixtures of these molecules. If DNA extracted from a tissue is digested with a restriction enzyme, producing a mixture of sequences of various lengths, the fragments obtained can be separated on the basis of size by electrophoresis in slabs of agarose—small molecules migrate more rapidly than large ones towards the anode (positive electrode), being less hindered by the agarose molecules. After electrophoretic separation the DNA fragments can then be transferred onto a sheet of nitrocellulose (or nylon) membrane by laying the agarose gel on a filter-paper wick dipped in buffer, covering the gel with the membrane and putting a thick layer of blotting paper (commonly 5–10 cm of paper towels) on top. Buffer drawn through the agarose carries the DNA molecules with it; these adhere to the membrane and (if a nylon membrane is used) can be covalently linked to it by exposure to u.v. light. After blotting, the filter can be incubated with labelled probe, unbound probe washed off and hybridized probe located by autoradiography. The probe hybridizes to complementary bands transferred from the original gel and the positions of bands on the filter corresponds to their electrophoretic running positions. This allows complementary sequences in the original DNA to be detected and their size determined. The procedure as described here was first used by Ed Southern (Southern, 1975) and is therefore referred to as a '*Southern blot*'. It can also be applied to the detection of RNA molecules, when the technique is named a '*Northern blot*'. These procedures should not be confused with what is known, by extension of the principle of naming techniques after points of the compass, as the Western blot, which involves the transfer of proteins from electrophoresis gels to membranes for detection using antibodies. The Eastern blot has yet to be devised.

There are a number of other examples of the application of hybridization of single-stranded molecules. For instance, mature mRNA molecules can be purified from preparations of total tissue RNA by virtue of the long sequences of adenosine residues (poly A) they contain at their 3'-ends. These hybridize to poly(dT) sequences attached to cellulose, and can be concentrated chromatographically on columns of oligo(dT) cellulose. Since hybridization is reduced at low salt concentrations, the poly(A⁺)-RNA can be eluted by diluting the buffer. Another example of the exploitation of hybridization is in the use of cDNA probes, and this is dealt with below.

We have discussed how labelled probes can be used to detect desirable cloned DNA sequences, which may themselves subsequently be used as probes, but not how appropriate probes are obtained in the first instance. There are in fact a number of ways in which probes can be obtained from scratch. Two most commonly employed today are the use of expression vectors with immunological detection of products, and synthetic oligonucleotides. In the first of these, fragments of foreign DNA are ligated into a vector containing DNA sequences that cause the protein synthetic machinery of the host cell to synthesize a protein using the insert DNA as template. (This process is called *translation* or *expression* of the inserted material.) By a process analogous to that described above for probing colonies with a labelled DNA or RNA sequence, the proteins synthesized can be transferred to a membrane and located by using an antibody raised against the gene product to be cloned. An expression vector commonly used for this procedure is the synthetic lambda phage, λ gt 11. The second technique, involving synthetic oligonucleotides, requires the synthesis of a probe

composed of perhaps 12–20 nucleotides on the basis of a known amino acid sequence, using the *genetic code* to determine the sequence in which the nucleotides should be linked. This process is complicated by the fact that only a few amino acids are coded for by one codon; because of the *redundancy* of the genetic code some amino acids are coded for by as many as 6 codons. However, with the advent of machines for making oligonucleotides this problem is reduced, as mixtures of all appropriate sequences can be made fairly readily. One of the first applications of the use of synthetic oligonucleotides was in the cloning of the relaxin cDNA (Hudson *et al.*, 1983); the application of the technique to the cloning of LHRH is described (p. 12). The process of probing for desirable DNA sequences has been described in detail above, but little attention has been given to how mixtures of cloned sequences are prepared for probing. Most frequently the mixtures of sequences are in the form of what are known as *libraries*. These are pools of vector molecules containing exogenous DNA sequences; genomic libraries are pools that are so large that all the DNA in a genome (for the purpose of this review, usually a mammalian or avian genome) should be represented in the library. A cDNA library is a pool of recombinant molecules, each containing a cDNA molecule prepared by reverse transcriptase from mRNA extracted from a tissue; to be effective, all mRNA molecules in the tissue should be represented in the library. Thus a genomic library should contain all the genetic information in the cell, including DNA sequences that are not transcribed, whereas a cDNA library consists only of sequences derived from genes that are being expressed in the tissue from which the mRNA was extracted. The size of a genomic library large enough to represent the entire genome can be calculated from the size of the genome involved and the lengths of the cloned DNA fragments. Thus if human genomic DNA was partly digested using a restriction enzyme and fragments 20 kb long were isolated and cloned, there would have to be 1.4×10^5 or more different clones in the library for it to be fully representative, as the human haploid genome contains 2.8×10^6 kb. It would not be worth screening a library any smaller than this, since the desired sequence may not be represented. cDNA libraries are clearly smaller than this, as only a small proportion of the genomic DNA is expressed as mRNA, and a cDNA library may contain more than one copy of a required sequence; in fact the frequency with which a desired sequence is represented will be related to the concentration of its mRNA in the tissue, so that some cDNA libraries contain relatively few different sequences. A cDNA library prepared from human placenta at term, for instance, contains predominantly sequences of hPL.

In many cases the end point for the cloning process, and the purpose for carrying out the procedure, involves the determination of the nucleotide sequence of the nucleic acid clone obtained. There are 2 procedures at present in use for obtaining long DNA sequences, and these are referred to by the names of their originators as the Sanger (or dideoxy) technique and the Maxam and Gilbert method. Both methods involve the production of fragments of DNA differing in length by a single nucleotide which can be separated by high resolution electrophoresis. If all the fragments have one common terminus, then identifying the nucleotide at the other end of each of the fragments will allow the deduction of the nucleotide sequence. The two methods available for producing the fragments and identifying the terminal nucleotide rely on totally different approaches.

The Sanger method (Sanger *et al.*, 1977) depends upon the use of 2',3'-dideoxynucleotides to terminate DNA replication by DNA polymerase and the accuracy with which this enzyme copies single stranded DNA sequences. Four incubations are set up, each containing the DNA to be sequenced, which is normally single stranded and produced by cloning in the bacteriophage M13, annealed (hybridized) to an oligonucleotide primer of 17 bases; since the primer can anneal in one position only and DNA polymerase has an absolute requirement for a primer, the use of a primer ensures that synthesis of the new DNA starts in the same position on every molecule. Synthesis continues until a 2',3'-dideoxy nucleotide is incorporated in the growing chain; since this has no 3'-hydroxyl group, chain elongation stops at this point. In fact, in each incubation elongation will stop on different DNA molecules at each of the potential points of incorporation of a dideoxy residue; thus when the contents of each incubation are separated by electrophoresis, the bands

(radioactive and so detectable by autoradiography) represent chains varying in length by one nucleotide. Each incubation is subjected to electrophoresis in a separate lane, so that by reading off the gel from bottom to top the sequence in which chain elongation terminated can be determined. Since each dideoxy nucleotide is incorporated on the basis of the sequence of the template DNA, the order of bands on the gel reflects the sequence required. This procedure has now become so routine that in some laboratories one person can sequence 2–3 kb of DNA every working day, and kits of reagents are commercially available for those not wishing to establish the conditions anew.

The alternative sequencing method (Maxam & Gilbert, 1977) also allows the sequence to be read off an electrophoresis autoradiograph but in this case specific fragmentation of the sequence is achieved by chemical means, based on the methylation of A and G residues by dimethyl sulphate, removal of the methylated base from the chain by heating, and selective cleavage of the sugar–phosphate backbone with alkali. Hydrazine is used to denature bases adjacent to T and C residues. As with the Sanger technique the molecule is labelled, but in this case labelling is carried out at one end of the chain, rather than throughout. Again a family of molecules, with a common terminus (^{32}P labelled) but differing in lengths by one nucleotide, is produced. Since the base-specific chemical cleavages are carried out in separate reaction vessels and electrophoresed separately, the position of each terminal base within the DNA molecule can be determined from the autoradiogram. Both these methods now contribute approximately equally to the sequence data now being collected at an exponentially accelerating rate (see Schmidtke & Krawczak, 1986).

Several techniques are available by which DNA sequences that interact with proteins can be identified and these provide useful information particularly on the 5'-regions of genes to which steroid receptors bind. Most frequently these techniques depend on the use of enzymes or other chemical agents to modify bases in DNA, those bases remaining unmodified being protected by the steric hindrance effect of bound proteins. Probably the most widely used example of this procedure, which has been applied particularly to the study of interactions between steroid receptors and DNA, is nuclease footprinting. In this technique, a relatively short, double-stranded DNA sequence is labelled at the 5'-end with ^{32}P and incubated with a protein preparation. Exonuclease III from *E. coli* is then added; this progressively digests the DNA from the 3'- to the 5'-end until its activity is blocked by bound protein. (Exonuclease activity may also be transiently slowed or even blocked by loops formed spontaneously in the DNA sequence, but these can be detected in control incubations.) When the digested DNA fragments are analysed on a sequencing gel alongside those generated using the same DNA in a sequencing reaction, the points at which exonuclease is blocked are identified as bands; these correspond in size to bands in the sequencing series and the nucleotides to which protein is bound can be identified by reference to the sequence. A good example of the application of this technique to the identification of DNA sequences that interact with glucocorticoid and progesterone receptors is given by Von der Ahe *et al.* (1985) and these results are described in the section on steroid receptors (Table 2).

Another method for identifying DNA sequences, particularly those hybridizing to mRNA molecules, is S1-nuclease mapping. S1 nuclease, which is isolated from *Aspergillus oryzae*, hydrolyses single stranded regions of DNA or RNA and this can be useful for identifying the cap sites of genes. Thus, if a purified mRNA preparation is incubated with excess cloned, 5'-labelled DNA to form hybrids and S1 nuclease is subsequently added, the DNA is digested until only hybridizing regions remain; the RNA–DNA hybrids are then denatured and the mixture is run on sequencing gels alongside sequencing reactions run on the DNA clone. As with exonuclease footprints, the positions on the sequence at which nuclease activity is blocked can then be identified. This technique has been used, for example, to show that chicken lysozyme mRNAs have at least 3 different 5'-start sites on the lysozyme gene, at positions 29, 31 and 53 nucleotides upstream of the translation initiation codon (Greß *et al.*, 1981).

Table 5. The genetic code and single letter amino acid convention (letters in parentheses)

	U		C		A		G	
U	UUU	Phe(F)	UCU	Ser(S)	UAU	Tyr(Y)	UGU	Cys(C)
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
	UUA	Leu(L)	UCA	Ser	UAA	Stop	UGA	Stop
	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp(W)
C	CUU	Leu	CCU	Pro(P)	CAU	His(H)	CGC	Arg(R)
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
	CUA	Leu	CCA	Pro	CAA	Gln(Q)	CGA	Arg
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
A	AUU	Ile(I)	ACU	Thr(T)	AAU	Asn(N)	AGU	Ser
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
	AUA	Ile	ACA	Thr	AAA	Lys(K)	AGA	Arg
	AUG	Met(M)	ACG	Thr	AAG	Lys	AGG	Arg
G	GUU	Val(V)	GCU	Ala(A)	GAU	Asp(D)	GGU	Gly(G)
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
	GUA	Val	GCA	Ala	GAA	Glu(E)	GGA	Gly
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Appendix 2: Glossary

Bacteriophage: a virus which infects a bacterium.

Complementary DNA (cDNA): a DNA sequence of opposite sense to another sequence (either RNA or DNA) is said to be complementary to it, and will anneal (or *hybridize*) to it to form a double-stranded molecule. In this review cDNA is usually the DNA copied from a mRNA template using reverse transcriptase.

Consensus sequence: an 'average' sequence compiled from a series of related sequences.

Dot blot: similar to a Southern blot (see below), except that the nucleic acid mixture to be studied is spotted onto a filter prior to hybridization, rather than being blotted.

Enhancer: a short sequence of genomic DNA, which interacts with a protein to stimulate gene expression in an undefined way.

Expression: the production of mRNA from DNA (transcription), usually followed by the production of a protein from the mRNA (translation).

Expression vector: a vector engineered to allow expression of inserted DNA.

Genetic code: the arrangement of bases in DNA or RNA which specifies the amino-acids to be incorporated into protein (see Table 5). In practice, the genetic code utilizes triplets of bases for each amino-acid and is read continuously. Every possible triplet is represented in the code (see redundancy).

Hybridize: the annealing (binding) of two complementary nucleic acid molecules by hydrogen bonding.

Infection: viral entry into a cell.

Lawn: a confluent layer of bacteria grown on a solid (usually agar) nutrient base.

Library: a collection of bacterial or viral clones each of which contains a different recombinant DNA molecule representing DNA or copies of RNA (cDNA) from an identified source. A genomic library contains clones of cellular DNA which represent all of the DNA in a genome; a cDNA library contains cDNA copies of mRNA molecules from a particular tissue.

- Ligation:** joining together of two double-stranded DNA molecules; a reaction catalysed by DNA ligase.
- Methylase:** an enzyme which adds -CH₃ groups to specific bases in a double-stranded DNA molecule. The effect of this is often to block the action of a restriction enzyme whose recognition sequence includes the methylated base.
- Nick translation:** a procedure for radiolabelling DNA, using *E. coli* DNA polymerase and deoxyribonuclease I.
- Northern blot:** a procedure whereby RNA is size-fractionated by gel electrophoresis and then transferred ('blotted') onto a nitrocellulose or nylon membrane followed by incubation with a labelled probe which will hybridize to complementary sequences on the membrane.
- Oligonucleotide:** a short sequence of nucleotides, usually complementary to a DNA or RNA sequence of interest and will, therefore, hybridize to it.
- Open reading frame:** a nucleotide sequence between an initiator and a stop codon which is translated into protein.
- Packaging:** the complexing together of viral DNA and viral proteins to form an infectious virus particle.
- Plaque:** a clear circle in a bacterial lawn caused by bacterial lysis, after bacteriophage infection.
- Plasmid:** an episomal DNA molecule capable of replicating independently of the host cell genome. Plasmids are widely used as vectors in recombinant DNA work. They generally carry antibiotic resistance genes to allow selection of transformed bacteria.
- Probe:** a labelled nucleic acid sequence which, by using hybridization techniques, can be used to find complementary sequences.
- Recombinant:** the result of insertion of exogenous DNA into a vector is called a recombinant DNA molecule. A bacterium transformed by a recombinant DNA molecule is termed a recombinant bacterium and the colony it grows into, a recombinant clone.
- Redundancy:** a term used to describe the observation that many amino acids are encoded in the genetic code by more than one codon (i.e. the genetic code is redundant).
- Repressor proteins:** proteins which reduce expression of a particular gene.
- Replication:** the process of copying DNA into a complementary DNA molecule: the reaction is catalysed by a DNA-dependent DNA polymerase.
- Restriction enzyme:** an enzyme which cuts DNA (usually double-stranded) at a specific recognition sequence.
- Reverse transcriptase:** an enzyme isolated from a retrovirus which is an RNA-dependent DNA polymerase, i.e. it copies an RNA template into DNA. It is important in recombinant DNA work as it copies mRNA into cDNA, the first step in cDNA cloning.
- Site-directed mutagenesis:** a procedure which allows alteration of specific bases in a recombinant DNA molecule. The technique is usually used to study the effects of such mutations on biological activity.
- Southern blot:** a similar procedure to a Northern blot (see above), except that DNA rather than RNA is used.
- Transcription:** the process of copying DNA into a complementary RNA molecule; the reaction is catalysed by a DNA-dependent RNA polymerase. The first step in gene expression.
- Transformation:** the process of the introduction of exogenous DNA into a cell, resulting in a changed cellular phenotype.
- Transgenic animal:** an animal bearing in its genome exogenous DNA inserted by artificial techniques.
- Translation:** the conversion of the mRNA code into a protein sequence.
- Vector:** a self-replicating double-stranded DNA molecule into which foreign DNA can be inserted; the foreign DNA is replicated along with the vector DNA in host cells. Expression vectors express the foreign DNA.

Appendix 3: Bibliography

- Nucleic Acid Biochemistry and Molecular Biology**, by W. I. P. Mainwaring, J. H. Parish, J. D. Pickering & N. H. Mann. Blackwell Scientific Publications, Oxford, 1982.
- Molecular Biology: A Comprehensive Introduction to Prokaryotes and Eukaryotes**, by D. Freifelder. Science Books International, Boston, 1983.
- Genetic Engineering**, Vols I–IV, edited by R. Williamson. Academic Press, London, 1981, 1982, 1983.
- Basic Cloning Techniques**, by R. H. Pritchard & I. B. Holland. Blackwell Scientific Publications, Oxford, 1985.
- Principles of Gene Manipulation: An Introduction to Genetic Engineering**, by R. W. Old & S. B. Primrose. Blackwell Scientific Publications, Oxford, 1981.
- Techniques in Molecular Biology**, edited by J. M. Walker & W. Gaastra. Croom Helm, London, 1983.
- IRL Press Practical Approach Series:**
- Gel Electrophoresis of Nucleic Acids**, edited by D. Rickwood & B. D. Hames. IRL Press, Oxford, 1982.
- Transcription and Translation**, edited by B. D. Hames & S. J. Higgins. IRL Press, Oxford, 1984.
- Oligonucleotide Synthesis**, edited by M. J. Gait. IRL Press, Oxford, 1984.
- DNA Cloning**, Vols I & II, edited by D. M. Glover, IRL Press, Oxford, 1985.
- Nucleic Acid Hybridization**, edited by B. D. Hames & S. J. Higgins. IRL Press, Oxford, 1985.
- Molecular Cloning: a Laboratory Manual**, by R. Maniatis, E. F. Fritsch & J. Sambrook. Cold Spring Harbor Laboratory, 1982.

References

- Abbot, S.D., Docherty, K., Roberts, J.L., Tepper, M.A., Chin, W.W. & Clayton, R.N. (1985) Castration increases luteinizing hormone subunit messenger RNA levels in male rat pituitaries. *J. Endocr.* **107**, R1–R4.
- Abelson, J. (1979) RNA processing and the intervening sequence problem. *Ann. Rev. Biochem.* **48**, 1035–1069.
- Adamson, E.D., Deller, M.J. & Warshaw, J.B. (1981) Functional EGF receptors are present on mouse embryo tissues. *Nature, Lond.* **291**, 656–659.
- Adamson, E.D., Meek, J. & Edwards, S.A. (1985) Product of the cellular oncogene, c-fos, observed in mouse and human tissues using an antibody to a synthetic peptide. *EMBO J.* **4**, 941–947.
- Adelman, J.P., Mason, A.J., Hayflick, J.S. & Seeburg, P.H. (1986) Isolation of the gene and hypothalamic cDNA for the common precursor of gonadotropin-releasing hormone and prolactin release-inhibiting factor in human and rat. *Proc. natn. Acad. Sci. U.S.A.* **83**, 179–183.
- Adelman, J.P., Bond, C.T., Douglass, J. & Herbert, E. (1987) Two mammalian genes transcribed from opposite strands of the same DNA locus. *Science, N.Y.* **235**, 1514–1517.
- Alt, F.W., Bothwell, A.L.M., Knapp, M., Siden, E., Mather, E., Koshland, M. & Baltimore, D. (1980) Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. *Cell* **20**, 293–301.
- Amara, S.G., David, D.N., Rosenfeld, M.G., Roos, B.A. & Evans, R.M. (1980) Characterization of rat calcitonin mRNA. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4444–4448.
- Amara, S.G., Jonas, V. & Rosenfeld, M.G. (1982) Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. *Nature, Lond.* **298**, 240–244.
- Arrigo, A.P. (1980) Investigation of the function of the heat shock proteins in *Drosophila melanogaster* tissue culture cells. *Molec. Gen. Genet.* **178**, 517–524.
- Ashburner, M. & Bonner, J.J. (1979) The induction of gene activity by heat shock. *Cell* **17**, 241–254.
- Aten, R., Ireland, J.J., Weems, C.W. & Behrman, H.R. (1987) Presence of gonadotrophin-releasing hormone-like proteins in bovine and ovine ovaries. *Endocrinology* **120**, 1727–1733.
- Austen, B.M. (1979) Predicted secondary structures of amino-terminal extension sequences of secreted proteins. *FEBS Lett.* **103**, 308–313.
- Austin, L.A. & Heath, H. (1981) Calcitonin: physiology and patho-physiology. *New Engl. J. Med.* **304**, 269–278.
- Bachvarova, R. & De Leon, V. (1980) Polyadenylated RNA of mouse ova and loss of maternal RNA in early development. *Devl Biol.* **74**, 1–8.
- Barnea, A. & Porter, J.C. (1975) Demonstration of a macromolecule cross-reacting with antibodies to luteinizing hormone and its tissue distribution. *Biochem. Biophys. Res. Commun.* **67**, 1346–1352.

- Barrera-Saldana, H.A., Robberson, D.L. & Saunders, G.F. (1982) Transcriptional products of the human placental lactogen gene. *J. biol. Chem.* **257**, 12399–12404.
- Barrera-Saldana, H.A., Seeburg, P.H. & Saunders, G.F. (1983) Two structurally different genes produce the same secreted human placental lactogen hormone. *J. biol. Chem.* **258**, 3787–3793.
- Barsh, G.S., Seeburg, P.H. & Gelinas, R.E. (1983) The human growth hormone gene family: structure and evolution of the chromosomal locus. *Nucl. Acids Res.* **11**, 3939–3958.
- Beato, M., Arnemann, J., Heins, B. & Müller, H. (1980) Correlation between uteroglobin synthesis and uteroglobin mRNA content in rabbit endometrium. In *Steroid Induced Uterine Proteins*, pp. 351–368. Ed. M. Beato. Elsevier/North Holland Biomedical Press.
- Becker, P.B., Gloss, B., Schmid, W., Strähle, U. & Schütz, G. (1986) *In vivo* protein-DNA interactions in a glucocorticoid response element require the presence of hormone. *Nature, Lond.* **324**, 686–688.
- Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980) The ovalbumin gene-sequence of putative control regions. *Nucl. Acids Res.* **8**, 127–142.
- Bensaude, O., Babinet, C., Morange, M. & Jacob, F. (1983) Heat shock proteins, first major products of zygotic gene activity in mouse embryo. *Nature, Lond.* **305**, 331–333.
- Bieker, J.J., Martin, P.L. & Roeder, R.G. (1985) Formation of a rate-limiting intermediate in 5S RNA gene transcription. *Cell* **40**, 119–127.
- Bird, A.P. (1978) Use of restriction enzymes to study eukaryotic DNA methylation. II. The symmetry of methylated sites supports semi-conservative copying of the methylation pattern. *J. molec. Biol.* **118**, 49–60.
- Bishop, J.M. (1978) Retroviruses. *Ann. Rev. Biochem.* **47**, 35–88.
- Bishop, J.M. (1981) Enemies within: the genesis of retrovirus oncogenes. *Cell* **23**, 5–6.
- Bishop, J.M. (1983) Cellular oncogenes and retroviruses. *Ann. Rev. Biochem.* **52**, 301–354.
- Bishop, J.M. (1985) Viral oncogenes. *Cell* **42**, 23–38.
- Bishop, J.M. & Varmus, H.E. (1982) Functions and origins of retroviral transforming genes. In *RNA Tumor Viruses, The Molecular Biology of Tumor Viruses*, Pt III, pp. 999–1108. Eds A. Weiss, N. Teich, H. Varmus & J. Coffin. Cold Spring Harbor Laboratory, New York.
- Bogavich, K. & Richards, J.S. (1982) Androgen biosynthesis in developing ovarian follicles: evidence that luteinizing hormone regulates thecal 17 α -hydroxylase and C-17,20 lyase activities. *Endocrinology* **111**, 1201–1208.
- Bogenhagen, D.F., Sakonju, S. & Brown, D.D. (1980) A control region in the center of the 5S RNA gene directs specific initiation of transcription: II. The 3' border of the region. *Cell* **19**, 27–35.
- Boime, I., Landefeld, T., McQueen, S. & McWilliams, D. (1978) The biosynthesis of chorionic gonadotropin and placental lactogen in first and third trimester human placenta. In *Structure and Function of the Gonadotropins*, pp. 235–257. Ed. K. W. McKerns. Plenum, New York.
- Boime, I., Boothby, M., Hoshina, M., Daniels-McQueen, S. & Darnell, R. (1982) Expression and structure of human placental hormone genes as a function of placental development. *Biol. Reprod.* **26**, 73–91.
- Boorstein, W.R., Vamvakopoulos, N.C. & Fiddes, J.C. (1982) Human chorionic gonadotropin β -subunit is encoded by at least eight genes arranged in tandem and inverted pairs. *Nature, Lond.* **300**, 419–422.
- Boothby, M., Ruddon, R.W., Anderson, C., McWilliams, D. & Boime, I. (1981) A single gonadotropin α -subunit gene in normal tissue and tumor-derived cell lines. *J. biol. Chem.* **256**, 5121–5127.
- Bousfield, G.R., Liu, W.-K., Sugino, H. & Ward, D.N. (1987) Structural studies of equine glycoprotein hormones. Amino acid sequence of equine lutropin β -subunit. *J. biol. Chem.* **262**, 8610–8620.
- Brain, S.D., Williams, T.J., Tippins, J.R., Morris, H.R. & MacIntyre, I. (1985) Calcitonin gene-related peptide is a potent vasodilator. *Nature, Lond.* **313**, 54–56.
- Brain, S.D., MacIntyre, I. & Williams, T.J. (1986) A second form of human calcitonin gene-related peptide which is a potent vasodilator. *Eur. J. Pharm.* **124**, 349–352.
- Breathnach, R. & Chambon, P. (1981) Organization and expression of eukaryotic split genes coding for proteins. *Ann. Rev. Biochem.* **50**, 349–383.
- Breslow, E. (1979) Chemistry and biology of the neurophysins. *Ann. Rev. Biochem.* **48**, 251–274.
- Brownstein, M.J., Russell, J.T. & Gainer, M. (1980) Synthesis, transport and release of posterior pituitary hormones. *Science, N.Y.* **207**, 373–378.
- Bullock, D.W., Woo, S.L.C. & O'Malley, B.W. (1976) Uteroglobin messenger RNA: translation *in vitro*. *Biol. Reprod.* **15**, 435–443.
- Burnside, J., Nagelberg, S.B., Lippman, S.S. & Weintraub, B.D. (1985) Differential regulation of hCG- α and - β subunit mRNAs in JEG-3 choriocarcinoma cells by 8-bromo-cAMP. *J. biol. Chem.* **260**, 12705–12709.
- Busch, H., Reddy, R., Rothblum, L. & Choi, Y.C. (1982) SnRNAs, SnRNPs and RNA processing. *Ann. Rev. Biochem.* **51**, 617–654.
- Carr, F.E., Need, L.R. & Chin, W.W. (1987) Isolation and characterization of the rat thyrotropin β -subunit gene. Differential regulation of two transcriptional start sites by thyroid hormone. *J. biol. Chem.* **262**, 981–987.
- Carrasco, A.E., McGinnis, W., Gehring, W.J. & de Robertis, E.M. (1984) Cloning of an *X. laevis* gene expressed during early embryogenesis for a peptide region homologous to *Drosophila* homeotic genes. *Cell* **37**, 409–414.
- Cate, R.L., Mattaliano, R.J., Hession, C., Tizard, R., Farber, N.M., Cheung, A., Ninfa, E.G., Frey, A.Z., Gash, D.J., Chow, E.P., Fisher, R.A., Bertonis, J.M., Torres, G., Wallner, B.P., Ramachandran, K.L., Ragin, R.C., Manganaro, T.F., MacLaughlin, D.T. & Donahue, P.K. (1986) Isolation of the bovine and human genes for Müllerian inhibiting substance and expression of the human gene in animal cells. *Cell* **45**, 685–698.
- Cato, A.C.B., Geisse, S., Wenz, M., Westphal, H.M. & Beato, M. (1984) The nucleotide sequences recognized by the glucocorticoid receptor in the rabbit uteroglobin gene region are located far upstream from the initiation of transcription. *EMBO J.* **3**, 2771–2778.

- Cattanach, B.M., Iddon, C.A., Charlton, H.M., Chiappa, S.A. & Fink, G. (1977) Gonadotrophin-releasing hormone deficiency in a mutant mouse with hypogonadism. *Nature, Lond.* **269**, 338–340.
- Cattini, P.A. & Eberhardt, N.L. (1987) Regulated expression of chimaeric genes containing the 5'-flanking regions of human growth hormone-related genes in transiently transfected rat anterior pituitary tumor cells. *Nucl. Acids Res.* **15**, 1297–1309.
- Chambon, P., Dierich, A., Gaub, M.-P., Jakowlev, S., Jongstra, J., Krust, A., Lepennel, J.-P., Oudet, P. & Reudelhuber, T. (1984) Promoter elements of genes coding for proteins and modulation of transcription of estrogens and progesterone. *Recent Progr. Horm. Res.* **40**, 1–39.
- Chan, L., Means, A.R. & O'Malley, B.W. (1973) Rates of induction of specific translatable messenger RNAs for ovalbumin and avidin by steroid hormones. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1870–1874.
- Charlton, H.M. & Cox, B.S. (1983) The use of recombinant DNA technology in neuroendocrine research. *Q. J. exp. Phys.* **68**, 449–462.
- Chauvet, M.T., Chauvet, J. & Acher, R. (1981) Identification of rat neurophysins: complete amino acid sequences of MSEL- and VLDV- neurophysins. *Biochem. Biophys. Res. Commun.* **103**, 595–603.
- Chauvet, M.T., Hurpet, D., Chauvet, J. & Acher, R. (1983) Identification of human neurophysins: complete amino acid sequences of MSEL and VLDV-neurophysins. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2839–2843.
- Chen, H.Y. & Brinster, R.L. (1982) Stability of rabbit globin and its messenger RNA in the mouse ovum. *Exp. Cell. Res.* **141**, 469–473.
- Chin, W.W., Kronenberg, H.M., Dee, P.C., Maloof, F. & Habener, J.F. (1981) Nucleotide sequence of the mRNA encoding the pre-alpha-subunit of mouse thyrotropin. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5329–5333.
- Clamagirand, C., Camier, M., Fahy, C., Clavreul, C., Creminon, C. & Cohen, P. (1987) C-Terminally extended oxytocin and pro-oxytocin: neurophysin peptide converting enzyme in bovine corpus luteum. *Biochem. Biophys. Res. Commun.* **143**, 789–796.
- Clegg, K.B. & Pikó, L. (1983) Poly(A) length, cytoplasmic adenylation and synthesis of poly(A) and RNA in early mouse embryos. *Dev. Biol.* **95**, 331–341.
- Clements, J.A. & Funder, J.W. (1986) Arginine vasopressin (AVP) and AVP-like immunoreactivity in peripheral tissues. *Endocr. Rev.* **7**, 449–460.
- Colberg-Poley, A.M., Voss, S.D., Chowdhury, K. & Gruss, P. (1985) Structural analysis of murine genes containing homeo box sequences and their expression in embryonal carcinoma cells. *Nature, Lond.* **314**, 713–718.
- Compton, J.G., Schrader, W.T. & O'Malley, B.W. (1982) Selective binding of chicken progesterone receptor A subunit to DNA fragment containing ovalbumin gene sequences. *Biochem. Biophys. Res. Commun.* **105**, 96–104.
- Compton, J.G., Schrader, W.T. & O'Malley, B.W. (1983) DNA sequence preference of the progesterone receptor. *Proc. natn. Acad. Sci. U.S.A.* **80**, 16–20.
- Conneely, O.M., Sullivan, W.P., Toft, D.O., Birnbaumer, M., Cook, R.G., Maxwell, B.L., Zarucki-Schulz, T., Greene, G.L., Schrader, W.T. & O'Malley, B.W. (1986) Molecular cloning of the chicken progesterone receptor. *Science, N.Y.* **233**, 767–770.
- Cooke, N.E., Coit, D., Shine, J., Baxter, J.D. & Martial, J.A. (1981) Human prolactin cDNA structural analysis and evolutionary comparisons. *J. biol. Chem.* **256**, 4007–4016.
- Cooke, N.E., Szpirer, C. & Levan, G. (1986) The related genes encoding growth hormone and prolactin have been dispersed to chromosomes 10 and 17 in the rat. *Endocrinology* **119**, 2451–2454.
- Cosman, D. (1987) Control of messenger RNA stability. *Immunology Today* **8**, 16–17.
- Cox, D.W., Woo, S.L.C. & Mansfield, T. (1985) DNA restriction fragments associated with α_1 -antitrypsin indicate a single origin for deficiency allele PI Z. *Nature, Lond.* **316**, 79–81.
- Crawford, R.J., Tregear, G.W. & Niall, H.D. (1986) The nucleotide sequences of baboon chorionic gonadotropin β -subunit genes have diverged from the human. *Gene* **46**, 161–169.
- Crick, F. (1979) Split genes and RNA splicing. *Science, N.Y.* **204**, 264–271.
- Curran, T. & Teich, N.M. (1982) Identification of a 39,000-dalton protein in cells transformed by the FBJ murine osteosarcoma virus. *Virology* **116**, 221–235.
- Davison, B.L., Egly, J.M., Mulvihill, E.R. & Chambon, P. (1983) Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. *Nature, Lond.* **301**, 680–686.
- Dean, D.C., Knoll, B.J., Riser, M.E. & O'Malley, B.W. (1983) A 5'-flanking sequence essential for progesterone regulation for an ovalbumin fusion gene. *Nature, Lond.* **305**, 551–554.
- Desplan, C., Theis, J. & O'Farrell, P.H. (1985) The *Drosophila* developmental gene, *engrailed*, encodes a sequence-specific DNA binding activity. *Nature, Lond.* **318**, 630–635.
- Dierks, P., Van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. & Weissmann, C. (1983) Three regions upstream from the Cap site are required for efficient and accurate transcription of the rabbit β -globin gene in mouse 3T6 cells. *Cell* **32**, 695–706.
- Donner, L., Fidele, L.A., Garon, C.F., Anderson, S.J. & Sherr, C.J. (1982) McDonough feline sarcoma virus: characterization of molecularly cloned provirus and its feline oncogene (v-fms). *J. Virol.* **41**, 489–500.
- Doolittle, R.F., Hunkapiller, M.W., Hood, L.E., Devare, S.G., Robbins, K.C., Aaronson, S.A. & Antoniades, H.N. (1983) Simian sarcoma virus *onc* gene, *v-sis*, is derived from the gene (or genes) encoding a platelet-derived growth factor. *Science, N.Y.* **221**, 275–277.
- Douglass, J., Civelli, O. & Herbert, E. (1984) Polypeptide gene expression: generation of diversity of neuroendocrine peptides. *Ann. Rev. Biochem.* **53**, 665–715.
- Dracopoli, N.C., Rettig, W.J., Whitfield, G.K., Darlington, G.R., Spengler, B.A., Biedler, J.L., Old, L.J. & Kourides, I.A. (1986) Assignment of the gene for the β subunit of thyroid-stimulating hormone to the short arm of human chromosome 1. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1822–1826.
- Duboule, D., Baron, A., Mähl, P. & Galliot, B. (1986) A new homeo-box is present in overlapping cosmid clones which define the mouse *HOX-1* locus. *EMBO J.* **5**, 1973–1980.

- Duesberg, P.H. (1983) Retroviral transforming genes in normal cells? *Nature, Lond.* **304**, 219–226.
- Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. & Hood, L. (1980) Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. *Cell* **20**, 313–319.
- Edwards, D.P., Chamness, G.C. & McGuire, W.L. (1979) Estrogen and progesterone receptor proteins in breast cancer. *Biochem. Biophys. Acta* **560**, 457–486.
- Ellis, R.W., deFeo, D., Maryak, J.M., Young, H.A., Shih, T.Y., Chang, E.H., Lowy, D.R. & Scolnick, E.M. (1980) Dual evolutionary origin for the rat genetic sequences of Harvey murine sarcoma virus. *J. Virol.* **36**, 408–420.
- Ellis, R.W., deFeo, D., Shih, T.Y., Gonda, M.A., Young, H.A., Tsuchida, N., Lowy, D.R. & Scolnick, E.M. (1981) The p21 src genes of Harvey and Kirsten sarcoma viruses originate from divergent members of a family of normal vertebrate genes. *Nature, Lond.* **292**, 506–511.
- Engelke, D.R., Nig, S.Y., Shastry, B.S. & Roeder, R.G. (1980) Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. *Cell* **19**, 717–728.
- Esch, F.S., Mason, A.J., Cooksey, K., Mercado, M. & Shimasaki, S. (1986) Cloning and DNA sequence analysis of the cDNA for the precursor of the β chain of bovine follicle stimulating hormone. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6618–6621.
- Fainsod, A., Bogarad, L.D., Ruusala, T., Lubin, M., Crothers, D.M. & Ruddle, F.H. (1986) The homeo domain of a murine protein binds 5' to its own homeo box. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9532–9536.
- Felsenfeld, G. & McGhee, J. (1982) Methylation and gene control. *Nature, Lond.* **296**, 602–603.
- Fiddes, J.C. & Goodman, H.M. (1979) Isolation, cloning, and sequence analysis of the cDNA for the α -subunit of human chorionic gonadotropin. *Nature, Lond.* **281**, 351–356.
- Fiddes, J.C. & Goodman, H.M. (1980) The cDNA for the β -subunit of human chorionic gonadotropin suggests evolution of a gene by read through into the 3'-untranslated region. *Nature, Lond.* **286**, 684–687.
- Fiddes, J.C. & Talmadge, K. (1984) Structure, expression and evolution of the genes for the human glycoprotein hormones. *Recent Prog. Horm. Res.* **40**, 43–78.
- Fiddes, J.C., Seeburg, P.H., Denoto, F., Hallelwell, R.A., Baxter, J.D. & Goodman, H.M. (1979) Structure of genes for human growth hormone and chorionic somatomammotropin. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4294–4298.
- Fisher, L.A., Kikkawa, D.O., Rivier, J.E., Amara, S.G., Evans, R.M., Rosenfeld, M.G., Vale, W.W. & Brown, M.R. (1983) Stimulation of noradrenergic sympathetic outflow by calcitonin gene-related peptide. *Nature, Lond.* **305**, 534–536.
- Flach, G., Johnson, M.H., Braude, P.R., Taylor, R.A.S. & Bolton, V.N. (1982) The transition from maternal to embryonic control in the 2-cell mouse embryo. *EMBO J.* **1**, 681–686.
- Flint, A.P.F. & Ricketts, A.P. (1979) Control of placental endocrine function; role of enzyme activation in the onset of labour. *J. Steroid Biochem.* **11**, 493–500.
- Flint, A.P.F. & Sheldrick, E.L. (1982) Ovarian secretion of oxytocin is stimulated by prostaglandin. *Nature, Lond.* **297**, 587–588.
- Flint, A.P.F., Sheldrick, E.L., Theodosios, D.T. & Wooding, F.B.P. (1986) Ovarian peptides: role of luteal oxytocin in the control of estrous cyclicity in ruminants. *J. Anim. Sci.* **62**, (Suppl. 2), 62–71.
- Forage, R.G., Brown, R.W., Oliver, K.J., Altrache, B.T., Devine, P.L., Hudson, G.C., Goss, N.H., Bertram, K.C., Tolstoshev, P., Robertson, D.M., de Kretser, D.M., Doughton, B., Burger, H.G. & Findlay, J.K. (1987) Immunization against an inhibin subunit produced by recombinant DNA techniques results in increased ovulation in sheep. *J. Endocr.* **114**, R1–R4.
- Frykberg, L., Palmieri, S., Beug, H., Graf, T., Hayman, M.J. & Vennström, B. (1983) Transforming capacities of avian erythroblastosis virus mutants deleted in the erbA or erbB oncogenes. *Cell* **32**, 227–238.
- Fuller, A.F., Jr, Krane, I.M., Budzik, G.P. & Donohue, P.K. (1985) Mullerian inhibiting substance reduction of colony growth of human gynaecologic cancers in a stem cell assay. *Gynaecol. Oncol.* **22**, 135–148.
- Garner, M.M. & Revzin, A. (1986) The use of gel electrophoresis to detect and study nucleic acid-protein interactions. *Tr. Biochem. Sci.* **11**, 395–396.
- Gasser, S.M. & Laemmli, U.K. (1987) A glimpse at chromosomal order. *Trends in Genetics* **3**, 16–22.
- Gaunt, S.J., Miller, J.R., Powell, D.J. & Duboule, D. (1986) Homeo-box gene expression in mouse embryos varies with position by the primitive streak stage. *Nature, Lond.* **324**, 662–664.
- Gehring, W.J. (1985) The homeo box: a key to the understanding of development. *Cell* **40**, 3–5.
- George, D.L., Phillips, J.A., Francke, U. & Seeburg, P.H. (1981) The genes for growth hormone and chorionic somatomammotropin are in the long arm of human chromosome 17 in the region q21-qter. *Hum. Genet.* **57**, 138–141.
- Gerber-Huber, S., Felber, B.K., Weber, R. & Ryffel, G.U. (1981) Estrogen induces tissue specific changes in the chromatin conformation of the vitellogenin genes in *Xenopus*. *Nucl. Acids Res.* **9**, 2475–2494.
- Gilbert, S.F. & Solter, D. (1985) Onset of paternal and maternal Gpi-1 expression in preimplantation mouse embryos. *Devl Biol.* **109**, 515–517.
- Gilbert, W. (1978) Why genes in pieces? *Nature, Lond.* **271**, 501.
- Gill, P., Jeffreys, A.J. & Werrett, D.J. (1985) Forensic application of DNA fingerprints. *Nature, Lond.* **318**, 577–578.
- Glaser, T., Lewis, W.H., Bruns, G.A.P., Watkins, P.C., Rogler, C.E., Shows, T.B., Powers, V.E., Willard, H.F., Goguen, J.M., Simola, K.O.J. & Housman, D.E. (1986) The β -subunit of follicle-stimulating hormone is deleted in patients with aniridia and Wilms' tumour, allowing a further definition of the *WAGR* locus. *Nature, Lond.* **321**, 882–887.
- Godine, J.E., Chin, W.W. & Habener, J.F. (1980) Luteinizing and follicle stimulating hormones. *J. biol. Chem.* **255**, 8780–8783.
- Godine, J.E., Chin, W.W. & Habener, J.F. (1982) α -subunit of rat pituitary glycoprotein hormones. *J. biol. Chem.* **257**, 8368–8371.
- Gosden, J.R. & Gosden, C.M. (1985) Recombinant DNA technology in prenatal diagnosis. *Oxford Rev. Reprod. Biol.* **7**, 73–117.

- Goustin, A.S., Betsholtz, C., Pfeifer-Ohlsson, S., Persson, H., Rydner, J., Bywater, M., Holmgren, G., Heldin, C.-H., Westermark, B. & Ohlsson, J. (1985) Coexpression of the *sis* and *myc* proto-oncogenes in developing human placenta suggests autocrine control of trophoblast growth. *Cell* **41**, 301–312.
- Govindan, M.V., Spiess, E. & Majors, J. (1982) Purified glucocorticoid receptor-hormone complex from rat liver cytosol binds specifically to cloned mouse mammary tumor virus long terminal repeats *in vitro*. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5157–5161.
- Green, S. & Chambon, P. (1987) Oestradiol induction of a glucocorticoid-responsive gene by a chimaeric receptor. *Nature, Lond.* **325**, 75–78.
- Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J.-M., Argos, P. & Chambon, P. (1986) Human oestrogen receptor cDNA: Sequence, expression and homology to *v-erb-A*. *Nature, Lond.* **320**, 134–139.
- Greene, G.L., Gilna, P., Waterfield, M., Baker, A., Hort, Y. & Shine, J. (1986) Sequence and expression of human estrogen receptor complementary DNA. *Science, N.Y.* **231**, 1150–1154.
- Grez, M., Land, H., Giesecke, K., Schütz, G., Jung, A. & Sippel, A.E. (1981) Multiple mRNAs are generated from the chicken lysozyme gene. *Cell* **25**, 743–752.
- Gropp, A. (1974) Autosomal trisomies in fetal mice, exencephaly in mice with trisomy 12. *Am. J. Pathol.* **77**, 539–542.
- Grosschedl, R. & Birnstiel, M.L. (1982) Delimitation of far upstream sequences required for maximal *in vitro* transcription of an H2A histone gene. *Proc. natn. Acad. Sci. U.S.A.* **79**, 297–301.
- Groudine, M. & Conkin, K.F. (1985) Chromatin structure and de novo methylation of sperm DNA: implications for activation of the paternal genome. *Science, N.Y.* **228**, 1061–1068.
- Gubits, R.M., Shaw, P.A., Gresik, E.W., Onetti-Muda, A. & Barka, T. (1986) Epidermal growth factor gene expression is regulated differently in mouse kidney and submandibular gland. *Endocrinology* **119**, 1382–1387.
- Gustafsson, J.-A., Carlstedt-Duke, J., Poellinger, L., Okret, S., Wikström, A.-C., Bronnegard, M., Gillner, M., Dong, Y., Fuxe, K., Cintra, A., Harfstrand, A. & Agovati, L. (1987) Biochemistry, molecular biology, and physiology of the glucocorticoid receptor. *Endocrine Rev.* **8**, 185–234.
- Hafen, E., Kuroiwa, A. & Gehring, W.J. (1984) Spatial distribution of transcripts from the segmentation gene *fushi tarazu* during *Drosophila* embryonic development. *Cell* **37**, 833–841.
- Hager, L.J., McKnight, G.S. & Palmiter, R.D. (1980) Glucocorticoid induction of egg white mRNAs in chick oviduct. *J. biol. Chem.* **255**, 7796–7800.
- Hamernik, D.L., Crowder, M.E., Nilson, J.H. & Nett, T.M. (1986) Measurement of messenger ribonucleic acid for luteinizing hormone β -subunit, α -subunit, growth hormone, and prolactin after hypothalamic pituitary disconnection in ovariectomized ewes. *Endocrinology* **119**, 2704–2710.
- Hammer, R.E., Brinster, R.L., Rosenfeld, M.G., Evans, R.M. & Mayo, K.E. (1985) Expression of human growth hormone-releasing factor in transgenic mice results in increased somatic growth. *Nature, Lond.* **315**, 413–416.
- Hann, S.R., Thompson, C.B. & Eisenman, R.N. (1985) C-Myc oncogene protein synthesis is independent of the cell cycle in human and avian cells. *Nature, Lond.* **314**, 366–369.
- Harland, R.M. (1982) Inheritance of DNA methylation in microinjected eggs of *Xenopus laevis*. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2323–2327.
- Hart, C.P., Awgulewitsch, A., Fainsod, A., McGinnis, W. & Ruddle, F.H. (1985) Homeo box gene complex in mouse chromosome 11: molecular cloning, expression in embryogenesis and homology to a human homeo box locus. *Cell* **43**, 9–18.
- Hauser, C.A., Joyner, A.L., Klein, R.D., Learned, T.K., Martin, G.R. & Tjian, R. (1985) Expression of homologous homeo-box-containing genes in differentiated human teratocarcinoma cells and mouse embryos. *Cell* **43**, 19–28.
- Heap, R.B. & Flint, A.P.F. (1979) Progesterone. In *Mechanisms of Hormone Action* (Reproduction in Mammals, 1st edn, vol. 7), pp. 185–232. Eds C.R. Austin & R.V. Short. Cambridge University Press.
- Herbert, E. (1981) Discovery of pro-opiomelanocortin, a cellular polypeptide. *Trends in Biochem. Sci.* **6**, 184–188.
- Higgs, D.R., Goodbourn, S.E.Y., Lamb, J., Clegg, J.B. & Weatherall, D.J. (1983) α -Thalassaemia caused by a polyadenylation signal mutation. *Nature, Lond.* **306**, 398–400.
- Hollenberg, S.M., Weinberger, C., Ong, E.S., Cerelli, G., Oro, A., Lebo, R., Thompson, E.B., Rosenfeld, M.G. & Evans, R.M. (1985) Primary structure and expression of a functional human glucocorticoid receptor cDNA. *Nature, Lond.* **318**, 635–641.
- Holwerda, D.A. (1972) A glycopeptide from the posterior lobe of pig pituitaries. *Eur. J. Biochem.* **28**, 334–339.
- Honda, B.M. & Roeder, R.G. (1980) Association of a 5S gene transcription factor with 5S RNA and altered levels of the factor during cell differentiation. *Cell* **22**, 119–126.
- Hoshina, M., Boothby, M. & Boime, I. (1982) Cytological localization of chorionic gonadotropin- α and placental lactogen mRNAs during development of the human placenta. *J. Cell Biol.* **93**, 190–198.
- Hoshina, M., Hussa, R., Pattillo, R. & Boime, I. (1983) Cytological distribution of chorionic gonadotropin subunit and placental lactogen messenger RNA in neoplasms derived from human placenta. *J. Cell Biol.* **97**, 1200–1206.
- Hoshina, M., Hussa, R. & Pattillo, R. (1984) The role of trophoblast differentiation in the control of the hCG and hPL genes. *Adv. exp. Med. Biol.* **176**, 299–312.
- Hoshina, M., Boothby, M., Hussa, R., Pattillo, R., Camel, H.M. & Boime, I. (1985) Linkage of human chorionic gonadotropin and placental lactogen biosynthesis to trophoblast differentiation and tumorigenesis. *Placenta* **6**, 163–172.
- Hudson, P., Haley, J., Cronk, M., Crawford, R., Haralambidis, J., Tregear, G., Shine, J. & Niall, H. (1983) Structure of a genomic clone encoding biologically active human relaxin. *Nature, Lond.* **301**, 628–631.
- Hudson, P., John, M., Crawford, R., Haralambidis, J., Scanlon, D., Gorman, J., Tregear, G., Shine, J. &

- Niall, H. (1984) Human ovaries and the predicted structure of a human preprorelaxin by analysis of cDNA clones. *EMBO J.* **3**, 2333–2339.
- Hughes, M.R., Compton, J.G., Schrader, W.T. & O'Malley, B.W. (1981) Interaction of the chick oviduct progesterone receptor with deoxyribonucleic acid. *Biochemistry, N.Y.* **20**, 2481–2491.
- Ingham, P.W., Howard, K.R. & Ish-Horowicz, D. (1985) Transcription pattern of the *Drosophila* gene *hairy*. *Nature, Lond.* **318**, 439–445.
- Ingolia, T.D. & Craig, E.A. (1982) *Drosophila* gene related to the major heat shock-induced gene is transcribed at normal temperatures and not induced by heat shock. *Proc. natn. Acad. Sci. U.S.A.* **79**, 525–529.
- Ingolia, T.D., Slater, M.R. & Craig, E.A. (1982) *Saccharomyces cerevisiae* contains a complex multigene family related to the major heat shock-inducible gene of *Drosophila*. *Molec. cell. Biol.* **2**, 1388–1398.
- Ivell, R. & Richter, D. (1984) The gene for the hypothalamic peptide hormone oxytocin is highly expressed in the bovine corpus luteum: biosynthesis, structure and sequence analysis. *EMBO J.* **3**, 2351–2354.
- Jackson, I.J., Schofield, P. & Hogan, B. (1985) A mouse homeo box gene is expressed during embryogenesis and in adult kidney. *Nature, Lond.* **317**, 745–748.
- Jacob, F., Ullman, A. & Monod, J. (1964) Le promoteur, element genetique necessaire a l'expression d'un operon. *C. r. hebdom. Séanc. Acad. Sci., Paris D* **258**, 3125–3128.
- Jakobovits, A., Schwab, M., Bishop, J.M. & Martin, G.R. (1985) Expression of N-myc in teratocarcinoma stem cells and mouse embryos. *Nature, Lond.* **318**, 188–191.
- Jameson, J.L., Jaffe, R.C., Gleason, S.L. & Habener, J.F. (1986) Transcriptional regulation of chorionic gonadotrophin α - and β -subunit gene expression by 8-bromo-adenosine 3', 5'-monophosphate. *Endocrinology* **119**, 2560–2567.
- Jeffreys, A.J. (1981) Recent studies of gene evolution using recombinant DNA. In *Genetic Engineering*, Vol. 2, pp. 1–48. Ed. R. Williamson. Academic Press, London.
- Jeffreys, A.J., Brookfield, J.F.Y. & Semeonoff, R. (1985) Positive identification of an immigration test-case using human DNA fingerprints. *Nature, Lond.* **317**, 818–819.
- Jeltsch, J.M., Krozowski, Z., Quirin-Stricker, C., Gronemeyer, H., Simpson, R.J., Garnier, J.M., Krust, A., Jacob, F. & Chambon, P. (1986) Cloning of the chicken progesterone receptor. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5424–5428.
- John, M.E., John, M.C., Ashley, P., MacDonald, R.J., Simpson, E.R. & Waterman, M.R. (1984) Identification and characterization of cDNA clones specific for cholesterol side-chain cleavage cytochrome P-450. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5628–5632.
- Jonas, V., Lin, C.R., Kawashima, E., Semon, D., Swanson, L.W., Mermod, J.J., Evans, R.M. & Rosenfeld, M.G. (1985) Alternative RNA processing events in human calcitonin/calcitonin gene-related peptide gene expression. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1994–1998.
- Josso, N. (1986) Antimüllerian hormone: new perspectives for a sexist molecule. *Endocr. Rev.* **7**, 421–433.
- Jost, J.-P., Seldran, M. & Geiser, M. (1984) Preferential binding of estrogen-receptor complex to a region containing the estrogen-dependent hypomethylation site preceding the chicken vitellogenin II gene. *Proc. natn. Acad. Sci. U.S.A.* **81**, 429–433.
- Joyner, A.L., Lebo, R.V., Kan, Y.W., Tjian, R., Cox, D.R. & Martin, G.R. (1985) Comparative chromosome mapping of a conserved homeo box region in mouse and human. *Cell* **43**, 173–175.
- Karin, M., Haslinger, A., Holtgreve, H., Richards, R.I., Krauter, P., Westphal, H.M. & Beato, M. (1984) Characterization of the DNA sequences through which cadmium and glucocorticoid hormones induce human metallothionein-II_A gene. *Nature, Lond.* **308**, 513–518.
- Kelley, P.M. & Schlesinger, M.J. (1982) Antibodies to the two major chicken heat shock proteins cross-react with similar proteins in widely divergent species. *Molec. cell. Biol.* **2**, 267–274.
- Kelly, K., Cochran, B.H., Stiles, C.D. & Leder, P. (1983) Cell-specific regulation of the *c-myc* gene by lymphocyte mitogens and platelet-derived growth factor. *Cell* **35**, 603–610.
- Kendrew, J.C. (1970) Some remarks on the history of molecular biology. *Biochem. Soc. Symp.* **30**, 5–10.
- Khodr, G.S. & Siler-Khodr, T.M. (1978) Localization of luteinizing hormone-releasing factor in the human placenta. *Fert. Steril.* **29**, 523–526.
- Kidd, V.J. & Saunders, G.F. (1982) Linkage arrangement of human placental lactogen and growth hormone genes. *J. biol. Chem.* **257**, 10673–10680.
- King, W.J. & Greene, G.L. (1984) Monoclonal antibodies localize oestrogen receptor in the nuclei of target cells. *Nature, Lond.* **307**, 745–747.
- Kingston, R.E., Baldwin, A.S. & Sharp, P.A. (1984) Regulation of heat shock protein 70 gene expression by *c-myc*. *Nature, Lond.* **312**, 280–282.
- Klein, G. (Ed.) (1980) *Viral Oncology*. Raven Press, New York.
- Knabbe, C., Lippman, M.E., Wakefield, L.M., Flanders, K.C., Kasid, A., Derynck, R. & Dickson, R.B. (1987) Evidence that transforming growth factor- β is a hormonally regulated negative growth factor in human breast cancer cells. *Cell* **48**, 417–418.
- Knowler, J.T. & Beaumont, J.M. (1985) The mechanism of action of oestrogens. *Essays in Biochemistry* **20**, 1–39.
- Knowlton, R.G., Cohen-Haguenaer, O., Van Cong, N., Frezal, J., Brown, V.A., Barker, D., Braman, J.C., Schumm, J.W., Tsui, L.-C., Buchwald, M. & Donis-Keller, H. (1985) A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. *Nature, Lond.* **318**, 380–382.
- Kornberg, T., Siden, I., O'Farrell, P. & Simon, M. (1985) The *engrailed* locus of *Drosophila*: *in situ* localization of transcripts reveals compartment-specific expression. *Cell* **40**, 45–53.
- Kreil, G. (1981) Transfer of proteins across membranes. *Ann. Rev. Biochem.* **50**, 317–348.
- Krieger, D.T., Perlow, M.J., Gibson, M.J., Davies, T.F., Zimmerman, E.A., Ferin, M. & Charlton, H.M. (1982) Brain grafts reverse hypogonadism of gonadotropin-releasing hormone deficiency. *Nature, Lond.* **298**, 468–471.

- Kruijer, W., Cooper, J.A., Hunter, T. & Verma, I.M. (1984) Platelet-derived growth factor induces rapid but transient expression of the c-fos gene and protein. *Nature, Lond.* **312**, 711–716.
- Krust, A., Green, S., Argos, P., Kumar, V., Walter, P., Bornert, J.-M. & Chambon, P. (1986) The chicken oestrogen receptor sequence: homology with v-erb-A and the human oestrogen and glucocorticoid receptors. *EMBO J.* **5**, 891–897.
- Lamb, D.J., Kima, P.E. & Bullock, D.W. (1986) Evidence for a single steroid-binding protein in the rabbit progesterone receptor. *Biochemistry, N.Y.* **25**, 6319–6324.
- Land, H., Schütz, G., Schmale, H. & Richter, D. (1982) Nucleotide sequence of cloned cDNA encoding bovine arginine vasopressin-neurophysin II precursor. *Nature, Lond.* **295**, 299–303.
- Landefeld, T.D. & Kepa, J. (1984) Pituitary α -subunit mRNA amounts during the sheep oestrous cycle. *J. biol. Chem.* **259**, 12817–12820.
- Laughon, A. & Scott, M.P. (1984) Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. *Nature, Lond.* **310**, 25–31.
- Lawson, G.M., Tsai, M.-J. & O'Malley, B.W. (1980) Deoxyribonuclease I sensitivity of the nontranscribed sequences flanking the 5' and 3' ends of the ovomucoid gene and the ovalbumin and its related X and Y genes in hen oviduct nuclei. *Biochemistry, N.Y.* **19**, 4403–4411.
- Lederberg, S. & Meselson, M. (1964) Degradation of non-replicating bacteriophage DNA in non-accepting cells. *J. molec. Biol.* **8**, 419–423.
- Levine, M., Rubin, G.M. & Tjian, R. (1984) Human DNA sequences homologous to a protein coding region conserved between homeotic genes in *Drosophila*. *Cell* **38**, 667–673.
- Levy, B.T., Sorge, L.K., Meymandi, A. & Maness, P.F. (1984) pp60c-src kinase is in chick and human embryonic tissues. *Devl Biol.* **104**, 9–17.
- Lewis, E.B. (1978) A gene complex controlling segmentation in *Drosophila*. *Nature, Lond.* **276**, 565–570.
- Ling, N., Ying, S.-Y., Ueno, N., Shimasaki, S., Esch, F., Hotta, M. & Guillemin, R. (1986) Pituitary FSH is released by a heterodimer of the β -subunits from the two forms of inhibin. *Nature, Lond.* **321**, 779–782.
- Lippman, M.E., Bolan, G. & Huff, K. (1976) The effects of androgens and antiandrogens on hormone-responsive human breast cancer in long-term tissue culture. *Cancer Res.* **36**, 4595–4601.
- Little, P.F.R. (1981) DNA analysis and the antenatal diagnosis of haemoglobinopathies. In *Genetic Engineering I*, pp. 62–102. Ed. R. Williamson. Academic Press, London.
- Mandel, J.L. & Chambon, P. (1979) DNA methylation: organ specific variations in the methylation pattern within and around the ovalbumin and other chicken genes. *Nucl. Acids Res.* **7**, 2081–2103.
- Manley, J.L., Fire, A., Cano, A., Sharp, P.A. & Geyer, M.L. (1980) DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3855–3859.
- Maruo, T., Matsuo, H., Ohtani, T., Hoshina, M. & Mochizuki, M. (1986) Differential modulation of chorionic gonadotropin (CG) subunit messenger ribonucleic acid levels and CG secretion by progesterone in normal placenta and choriocarcinoma cultured *in vitro*. *Endocrinology* **119**, 855–864.
- Mason, A.J., Hayflick, J.S., Ling, N., Esch, F., Ueno, N., Ying, S.-Y., Guillemin, R., Niall, H. & Seeburg, P.H. (1985) Complementary DNA sequences of ovarian follicular fluid inhibin show precursor structure and homology with transforming growth factor- β . *Nature, Lond.* **318**, 659–663.
- Mason, A.J., Hayflick, J.S., Zoeller, R.T., Young, W.S., Phillips, H.S., Nikolics, K. & Seeburg, P.H. (1986a) A deletion truncating the gonadotrophin-releasing hormone gene is responsible for hypogonadism in the hpg mouse. *Science N.Y.* **234**, 1366–1371.
- Mason, A.J., Pitts, S.L., Nikolics, K., Szonyi, E., Wilcox, J.N., Seeburg, P.H. & Stewart, T.A. (1986b) The hypogonadal mouse: reproductive functions restored by gene therapy. *Science, N.Y.* **234**, 1372–1378.
- Massagué, J. (1985) The transforming growth factors. *Trends in Biochem. Sci.* **10**, 237–240.
- Massagué, J. (1987) The TGF- β family of growth and differentiation factors. *Cell* **49**, 437–438.
- Maurer, R.A. (1983) Use of recombinant DNA procedures to study mechanisms controlling synthesis of hypophyseal hormones. *Biol. Reprod.* **28**, 27–35.
- Mavilio, F., Simeone, A., Giampaolo, A., Faiella, A., Zappavigna, V., Acampora, D., Poiana, G., Russo, G., Peschle, C. & Boncinelli, E. (1986) Differential and stage-related expression in embryonic tissues of a new human homeo-box gene. *Nature, Lond.* **324**, 664–668.
- Maxam, A.M. & Gilbert, W. (1977) A new method for sequencing DNA. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564.
- McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A. & Gehring, W.J. (1984a) A conserved DNA sequence in homeotic genes of the *Drosophila* Antennapedia and bithorax complexes. *Nature, Lond.* **308**, 428–433.
- McGinnis, W., Hart, C.P., Gehring, W.J. & Ruddle, F.H. (1984b) Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of *Drosophila*. *Cell* **38**, 675–680.
- McGrath, J. & Solter, D. (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37**, 179–184.
- McKnight, S.L. & Kingsbury, R. (1982) Transcriptional control signals of a eukaryotic protein-coding gene. *Science, N.Y.* **217**, 316–324.
- McWilliams, D. & Boime, I. (1980) Cytological localization of placental lactogen messenger ribonucleic acid in syncytiotrophoblast layers of human placenta. *Endocrinology* **107**, 761–765.
- Millar, R.P., Wormald, P.J. & Milton, R.C.L. (1986) Stimulation of gonadotropin release by a non-GnRH peptide sequence of the GnRH precursor. *Science, N.Y.* **232**, 68–70.
- Miller, J., McLachlan, A.D. & Klug, A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* **4**, 1609–1614.
- Miller, W.L. & Eberhardt, N.L. (1983) Structure and evolution of the growth hormone gene family. *Endocr. Rev.* **4**, 97–130.
- Moen, R.C. & Palmiter, R.D. (1980) Changes in hormone responsiveness of chick oviduct during primary stimulation with estrogen. *Devl Biol.* **78**, 450–463.

- Monaco, A.P., Bertelson, C.J., Middlesworth, W., Colletti, C.-A., Aldridge, J., Fischbeck, K.H., Bartlett, R., Pericak-Vance, M.A., Roses, A.D. & Kunkel, L.M. (1985) Detection of deletions spanning the Duchenne muscular dystrophy locus using a tightly linked DNA segment. *Nature, Lond.* **316**, 842–845.
- Monk, M. & Harper, M.I. (1979) Sequential X-chromosome inactivation coupled with cellular differentiation in early mouse embryos. *Nature, Lond.* **281**, 311–313.
- Monk, M., Boubelik, M. & Lehnert, S. (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**, 371–382.
- Moore, D.D., Marks, A.R., Buckley, D.I., Kapler, G., Payvar, F. & Goodman, H.M. (1985) The first intron of the human growth hormone gene contains a binding site for glucocorticoid receptor. *Proc. natn. Acad. Sci. U.S.A.* **82**, 699–702.
- Morange, M., Diu, A., Bensaude, O. & Babinet, C. (1984) Altered expression of heat shock proteins in embryonal carcinoma and mouse early embryonic cells. *Molec. cell. Biol.* **4**, 730–735.
- Morimoto, R. & Fodor, E. (1984) Cell-specific expression of heat shock proteins in chicken reticulocytes and lymphocytes. *J. Cell Biol.* **99**, 1316–1323.
- Möritz, Th., Edström, J.E. & Pongs, O. (1984) Cloning of a gene localized and expressed at the ecdysteroid regulated puff 74EF in salivary glands of *Drosophila* larvae. *EMBO J.* **3**, 289–295.
- Morohashi, K., Fujii-Kuriyama, Y., Okada, Y., Sogawa, K., Hirose, T., Inayama, S. & Omura, T. (1984) Molecular cloning and nucleotide sequence of cDNA for mRNA of mitochondrial cytochrome P-450 (SCC) of bovine adrenal cortex. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4647–4651.
- Muller, M.M., Carrasco, A.E. & de Robertis, E.M. (1984) A homeo-box-containing gene expressed during oogenesis in *Xenopus*. *Cell* **39**, 157–162.
- Muller, R. (1986) Cellular and viral fos genes: structure, regulation of expression and biological properties of their encoded products. *Biochim. Biophys. Acta* **823**, 207–225.
- Muller, R., Slamon, D.J., Tremblay, J.M., Cline, M.J. & Verma, I.M. (1982) Differential expression of cellular oncogenes during pre- and post-natal development of the mouse. *Nature, Lond.* **299**, 640–644.
- Muller, R., Tremblay, J.M., Adamson, E.D. & Verma, I.M. (1983a) Tissue and cell type-specific expression of two human c-onc genes. *Nature, Lond.* **304**, 454–456.
- Muller, R., Verma, I.M. & Adamson, E.D. (1983b) Expression of c-onc genes: c-fos transcripts accumulate to high levels during development of mouse placenta, yolk sac and amnion. *EMBO J.* **2**, 679–684.
- Muller, R., Slamon, D.J., Adamson, E.D., Tremblay, J.M., Muller, D., Cline, M.J. & Verma, I.M. (1983c) Transcription of c-onc genes c-ras^{ki} and c-fms during mouse development. *Molec. cell. Biol.* **3**, 1062–1069.
- Muller, R., Muller, D., Verrier, B., Bravo, R. & Herbst, H. (1986) Evidence that expression of c-fos protein in amnion cells is regulated by external signals. *EMBO J.* **5**, 311–316.
- Mulvihill, E.R., Lepennec, J.-P. & Chambon, P. (1982) Chicken oviduct progesterone receptor: location of specific regions of high-affinity binding in cloned DNA fragments of hormone-responsive genes. *Cell* **24**, 621–632.
- Muthukrishnan, S., Furuichi, Y. & Both, G.W. (1976) Effect of 5'-terminal structures on the binding of ribopolymers to eukaryotic ribosomes. *Prog. Nucl. Acid Res. Mol. Biol.* **19**, 473–476.
- Nakajin, S. & Hall, P.F. (1981) Mitochondrial cytochrome P-450 from neonatal pig testis. Purification and properties of a C₂₁ steroid side-chain cleavage system (17 α -hydroxylase–C_{17,20} lyase). *J. biol. Chem.* **256**, 3871–3876.
- Nakanishi, S., Inoue, A., Kita, T., Nakamura, M., Chang, A.C.Y., Cohen, S.N. & Numa, S. (1979) Nucleotide sequence of cloned cDNA for bovine corticotropin- β -lipoprotein precursor. *Nature, Lond.* **278**, 423–427.
- Nawa, H., Kotani, H. & Nakanishi, S. (1984) Tissue-specific generation of two preprotachykinin mRNAs from one gene by alternative RNA splicing. *Nature, Lond.* **312**, 729–734.
- Naylor, S.L., Chin, W.W., Goodman, H.M., Lalley, P.A., Grzeschik, K.H. & Sakuguchi, A.Y. (1983) Chromosome assignment of genes encoding the alpha and beta subunits of glycoprotein hormones in man and mouse. *Somat. Cell Genet.* **9**, 757–770.
- Nelson, C., Grenshaw, E.B., 3rd., Franco, R., Lira, S.A., Albert, V.R., Evans, R.M. & Rosenfeld, M.G. (1986) Discrete cis-acting genomic sequences dictate the pituitary cell type-specific expression of rat prolactin and growth hormone genes. *Nature, Lond.* **322**, 557–562.
- Nevins, J.R. (1979) Processing of late adenovirus nuclear RNA to mRNA. Kinetics of formation of intermediates and demonstration that all events are nuclear. *J. molec. Biol.* **130**, 493–506.
- Nikolics, K., Mason, A.J., Szönyi, E., Ramachandran, J. & Seeburg, P.H. (1985) A prolactin-inhibiting factor within the precursor for human gonadotropin-releasing hormone. *Nature, Lond.* **316**, 511–517.
- Nilson, J.H., Thomason, A.R., Cserbak, M.T., Moncman, C.L. & Woychik, R.P. (1983) Nucleotide sequence of a cDNA for the common alpha subunit of the bovine pituitary glycoprotein hormones. Conservation of nucleotides in the 3'-untranslated region of bovine and human pre-alpha subunit mRNAs. *J. biol. Chem.* **258**, 4679–4682.
- Nordström, J.L., Roop, D.R., Tsai, M.-J. & O'Malley, B.W. (1979) Identification of potential ovomucoid mRNA precursors in chick oviduct nuclei. *Nature, Lond.* **278**, 328–331.
- Ohlendorf, D.H. & Matthews, B.W. (1983) Structural studies of protein-nucleic acid interactions. *Ann. Rev. Biophys. Bioeng.* **12**, 259–284.
- O'Malley, B.W., Roop, D.R., Lai, E.C., Nordström, J.L., Catterall, J.F., Swaneck, G.E., Colbert, D.A., Tsai, M.-J., Dugaiczky, A. & Woo, S.L.C. (1979) The ovalbumin gene: organization, structure, transcription, and regulation. *Recent Progr. Horm. Res.* **35**, 1–42.
- Owerbach, D., Rutter, W.J., Martial, J.A., Baxter, J.D. & Shows, T.B. (1980) Genes for growth hormone chorionic somatomammotropin and a growth hormone-like gene are located on chromosome 17 in humans. *Science, N. Y.* **209**, 289–292.

- Palmiter, R.D. & Brinster, R.L.** (1985) Transgenic mice. *Cell* **41**, 343–345.
- Papavasiliou, S.S., Zmeili, S., Herbon, L., Duncan-Weldon, J., Marshall, J.C. & Landefeld, T.D.** (1986) α and luteinizing hormone- β messenger ribonucleic acid (RNA) of male and female rats after castration: quantitation using an optimized RNA dot blot hybridization assay. *Endocrinology* **119**, 691–698.
- Parker, M., Hurst, H. & Page, M.** (1984) Organization and expression of prostatic steroid binding protein genes. *J. Steroid Biochem.* **20**, 67–71.
- Pascal, O., Morel, Y., Lejeune, H. & Andre, J.** (1984) Further characterization of the porcine LH subunit precursors in the translation products of pituitary RNAs. *Biochem. Biophys. Res. Commun.* **124**, 383–387.
- Payvar, F., DeFranco, D., Firestone, G.L., Edgar, B., Wrangé, O., Okret, S., Gustafsson, J.-Å. & Yamamoto, K.R.** (1983) Sequence-specific binding of glucocorticoid receptor to MTV DNA at sites within and upstream of the transcribed region. *Cell* **35**, 381–392.
- Pelham, H.R.B. & Brown, D.D.** (1980) A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4170–4174.
- Pfahl, M.** (1982) Specific binding of the glucocorticoid-receptor complex to the mouse mammary tumor provirus promoter region. *Cell* **31**, 475–482.
- Pfeifer-Ohlsson, S., Goustin, A.S., Rydnert, J., Wahlstrom, T., Bjersing, L., Stehelin, D. & Ohlsson, R.** (1984) Spatial and temporal pattern of cellular myc oncogene expression in developing human placenta: implications for embryonic cell proliferation. *Cell* **38**, 585–596.
- Phillips, H.S., Nikolics, K., Branton, D. & Seeburg, P.H.** (1985) Immunocytochemical localization in rat brain of a prolactin release inhibiting sequence of gonadotropin-releasing hormone prohormone. *Nature, Lond.* **316**, 542–545.
- Picard, J.-Y., Benarous, R., Guerrier, D., Josso, N. & Kahn, A.** (1986) Cloning and expression of cDNA for antimüllerian hormone. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5464–5468.
- Pikó, L. & Clegg, K.B.** (1982) Quantitative changes in total RNA, total poly(A) and ribosomes in early mouse embryos. *Devl Biol.* **89**, 362–378.
- Pless, D.D. & Lennarz, W.J.** (1977) Enzymatic conversion of proteins to glycoproteins. *Proc. natn. Acad. Sci. U.S.A.* **74**, 134–138.
- Policastro, P.F., Daniels-McQueen, S., Carle, G. & Boime, I.** (1986) A map of the hCG β -LH β gene cluster. *J. biol. Chem.* **261**, 5907–5916.
- Proudfoot, N.J. & Brownlee, G.G.** (1976) 3'-non-coding region sequence in eukaryotic messenger RNA. *Nature, Lond.* **263**, 211–214.
- Ralston, R. & Bishop, J.M.** (1983) The protein products of the myc and myb oncogenes and adenovirus Ela are structurally related. *Nature, Lond.* **306**, 803–806.
- Reddy, V.B., Beck, A.K., Garromone, A.J., Vellucci, V., Lustbader, J. & Bernstein, E.G.** (1985) Expression of human choriogonadotropin in monkey cells using a single simian virus 40 vector. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3644–3648.
- Reed, R. & Maniatis, T.** (1985) Intron sequences involved in lariat formation during pre-mRNA splicing. *Cell* **41**, 95–105.
- Reik, W., Collick, A., Norris, M.L., Barton, S.C. & Surani, M.A.** (1987) Genomic imprinting determines methylation of parental alleles in transgenic mice. *Nature, Lond.* **328**, 248–251.
- Renkawitz, R., Schütz, G., von der Ahe, D. & Beato, M.** (1984) Sequences in the promoter region of the chicken lysozyme gene required for steroid regulation and receptor binding. *Cell* **37**, 503–510.
- Ringold, G.M., Dobson, D.E., Grove, J.R., Hall, C.V., Lee, F. & Vannice, J.L.** (1983) Glucocorticoid regulation of gene expression: mouse mammary tumour virus as a model system. *Recent Progr. Horm. Res.* **39**, 387–421.
- Rogers, B.L. & Saunders, G.F.** (1986) Transcriptional enhancers play a major role in gene expression. *Bioessays* **4**, 62–65.
- Rosenfeld, M.G., Amara, S.G., Roos, B.A., Ong, E.S. & Evans, R.M.** (1981) Altered expression of the calcitonin gene associated with RNA polymorphism. *Nature, Lond.* **290**, 63–65.
- Rosenfeld, M.G., Mermod, J.-J., Amara, S.G., Swanson, L.W., Sawchenko, P.E., Rivier, J., Vale, W.W. & Evans, R.M.** (1983) Production of a novel neuropeptide encoded by the calcitonin gene via tissue-specific RNA processing. *Nature, Lond.* **304**, 129–135.
- Rosenfeld, M.G., Amara, S.G. & Evans, R.M.** (1984) Alternative RNA processing events as a critical developmental regulatory strategy in neuroendocrine gene expression. *Biochem. Soc. Symp.* **49**, 27–44.
- Ruppert, S., Scherer, G. & Schütz, G.** (1984) Recent gene conversion involving bovine vasopressin and oxytocin precursor genes suggested by nucleotide sequence. *Nature, Lond.* **308**, 554–557.
- Sabbah, M., Redeuilh, G., Secco, C. & Baulieu, E.-E.** (1987) The binding activity of estrogen receptor to DNA and heat shock protein (Mr 90 000) is dependent on receptor-bound metal. *J. biol. Chem.* **262**, 8631–8635.
- Sanger, F., Nicklen, S. & Coulson, A.R.** (1977) DNA sequencing with chain-terminating inhibitors. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467.
- Sap, J., Munoz, A., Damm, K., Goldberg, Y., Ghysdael, J., Leutz, A., Beug, H. & Vennstrom, B.** (1986) The c-erb-A protein is a high-affinity receptor for thyroid hormone. *Nature, Lond.* **324**, 635–640.
- Sapienza, C., Peterson, A.C., Rossant, J. & Balling, R.** (1987) Degree of methylation of transgenes is dependent on gamete of origin. *Nature, Lond.* **328**, 251–254.
- Sato, R. & Omura, T.** (Eds) (1978) *Cytochrome P-450*. Academic Press, New York.
- Sausville, F., Carney, D. & Battey, J.** (1985) The human vasopressin gene is linked to the oxytocin gene and is selectively expressed in a cultured lung cancer cell line. *J. biol. Chem.* **260**, 10236–10241.
- Sawicki, J.A., Magnusson, T. & Epstein, C.J.** (1981) Evidence for expression of the paternal genome in the two-cell mouse embryo. *Nature, Lond.* **294**, 450–451.
- Schally, A.V., Arimura, A., Kastin, A.J., Matsuo, H., Baba, Y., Redding, T.W., Nair, R.M.G., Debeljuk, L. & White, W.F.** (1971) Gonadotropin releasing hormone: one polypeptide regulates secretion of

- luteinizing and follicle-stimulating hormone. *Science*, N.Y. **173**, 1036–1038.
- Scheidereit, C., Geisse, S., Westphal, H.M. & Beato, M. (1983) The glucocorticoid receptor binds to defined nucleotide sequences near the promoter of mouse mammary tumour virus. *Nature, Lond.* **304**, 749–752.
- Schlesinger, M.J., Aliperti, G. & Kelley, P.M. (1982) The response of cells to heat shock. *Trends in Biochem. Sci.* **7**, 222–225.
- Schmale, H. & Richter, D. (1981) Immunological identification of a common precursor to arginine vasopressin and neurophysin II synthesized by *in vitro* translation of bovine hypothalamic mRNA. *Proc. natn. Acad. Sci. U.S.A.* **78**, 766–771.
- Schmidtke, J. & Krawczak, M. (1986) Human gene cloning: the storm before the lull? *Nature, Lond.* **322**, 119.
- Seeburg, P.H. & Adelman, J.P. (1984) Characterization of cDNA for precursor of human luteinizing hormone releasing hormone. *Nature, Lond.* **311**, 666–668.
- Segall, J., Matsui, T. & Roeder, R.G. (1980) Multiple factors are required for the accurate transcription of purified genes by RNA polymerase III. *J. biol. Chem.* **255**, 11986–11991.
- Seidah, N.G., Benjannet, S. & Chretien, M. (1981) The complete sequence of a novel human pituitary glycopeptide homologous to pig posterior pituitary glycopeptide. *Biochem. Biophys. Res. Commun.* **100**, 901–907.
- Selvanayagam, C.S., Tsai, S.Y., Tsai, M.J., Selvanayagam, P. & Saunders, G.F. (1984) Multiple origins of transcription for the human placental lactogen genes. *J. biol. Chem.* **259**, 14642–14646.
- Sharp, P.A. (1985) On the origin of RNA splicing and introns. *Cell* **42**, 397–400.
- Sharp, P.A. (1987) Splicing of messenger RNA precursors. *Science, N.Y.* **235**, 766–771.
- Sharpe, R.M., Fraser, H.M., Cooper, I. & Rommerts, F.F.G. (1981) Sertoli-Leydig cell communication via an LHRH-like factor. *Nature, Lond.* **290**, 785–787.
- Shatkin, A.J. (1976) Capping of eucaryotic mRNAs. *Cell* **9**, 645–653.
- Shine, J., Seeburg, P.H., Martial, J.A., Baxter, J.D. & Goodman, H. (1977) Construction and analysis of recombinant DNA for human chorionic somatomammotropin. *Nature, Lond.* **270**, 494–499.
- Shivers, B.D., Harlan, R.E., Morrell, J.I. & Pfaff, D.W. (1983) Immunocytochemical localization of luteinizing hormone-releasing hormone in male and female rat brains. *Neuroendocrinology* **36**, 1–12.
- Shivers, B.D., Harlan, R.E., Hejtmancik, J.F., Lonn, P.M. & Pfaff, D.W. (1986) Localization of cells containing LHRH-like mRNA in rat forebrain using *in situ* hybridization. *Endocrinology* **118**, 883–885.
- Siler-Khodr, T.M., Khodr, G.S., Valenzuela, G. & Rhode, J. (1986a) Gonadotropin-releasing hormone effects on placental hormones during gestation: I. α -human chorionic gonadotropin, human chorionic gonadotropin and human chorionic somatomammotropin. *Biol. Reprod.* **34**, 245–254.
- Siler-Khodr, T.M., Khodr, G.S., Valenzuela, G. & Rhode, J. (1986b) Gonadotropin-releasing hormone effects on placental hormones during gestation: II. Progesterone, estrone, estradiol and estriol. *Biol. Reprod.* **34**, 255–264.
- Slamon, D.J. & Cline, M.J. (1984) Expression of cellular oncogenes during embryonic and fetal development of the mouse. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7141–7145.
- Smith, H.O. & Nathans, S.D. (1973) A suggested nomenclature for bacterial host modification and restriction systems and their enzymes. *J. molec. Biol.* **81**, 419–423.
- Smyth, D.G. & Massey, D.E. (1979) A new glycopeptide in pig, ox and sheep pituitary. *Biochem. Biophys. Res. Commun.* **87**, 1006–1010.
- Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. molec. Biol.* **98**, 503–517.
- Stanley, H.F., Curtis, A., Sheward, W.J., Roberts, J.L. & Fink, G. (1986) Prolactin messenger ribonucleic acid levels in the normal and hypogonadal mouse pituitary gland. *Endocrinology* **119**, 2422–2426.
- Stewart, F., Thomson, J.A., Leigh, S.E.A. & Warwick, J.M. (1987a) Nucleotide (cDNA) sequence encoding the horse gonadotrophin α -subunit. *J. Endocr.* **115**, 341–346.
- Stewart, F., Leigh, S.E.A. & Thomson, J.A. (1987b) Application of recombinant DNA techniques to structure–function studies of equine protein hormones. *J. Reprod. Fert., Suppl.* **35**, 1–8.
- Stewart, M. & Stewart, F. (1977) Constant and variable regions in glycoprotein hormone beta subunit sequences: implications for receptor binding specificity. *J. molec. Biol.* **116**, 175–179.
- Stone, R.T., Maurer, R.A. & Gorski, J. (1977) Effect of estradiol-17 β on preprolactin messenger ribonucleic acid activity in the rat pituitary gland. *Biochemistry, N.Y.* **16**, 4915–4921.
- Stout, J.T., Chen, H.Y., Brennand, J., Caskey, C.T. & Brinster, R.L. (1985) Expression of human HPRT in the central nervous system. *Nature, Lond.* **317**, 250–252.
- Stromberg, K., Pigott, D.A., Ranchalis, J.E. & Twardzik, D.R. (1982) Human term placenta contains transforming growth factors. *Biochem. Biophys. Res. Commun.* **106**, 354–361.
- Suchanek, G. & Kreil, G. (1977) Translation of melittin messenger RNA *in vitro* yields a product terminating with glutaminyglycine rather than with glutaminamide. *Proc. natn. Acad. Sci. U.S.A.* **74**, 975–978.
- Sugino, H., Bousfield, G.R., Moore, W.T. & Ward, D.N. (1987) Structural studies on equine glycoprotein hormones. Amino acid sequence of equine chorionic gonadotropin β -subunit. *J. biol. Chem.* **262**, 8603–8609.
- Surani, M.A.H. (1984) Differential roles of paternal and maternal genomes during embryogenesis in the mouse. *Bioessays* **1**, 224–226.
- Surani, M.A.H., Reik, W., Norris, M.L. & Barton, S.C. (1986) Influence of germline modifications of homologous chromosomes on mouse development. *J. Embryol. exp. Morph.* **97** (Suppl.) 123–136.
- Takahashi, N., Roach, A., Teplow, D.B., Prusiner, S.B. & Hood, L. (1985) Cloning and characterization of the myelin basic protein gene from mouse: one gene can encode both 14 kd and 18.5 kd MBPs by alternate use of exons. *Cell* **42**, 139–148.
- Talmadge, K., Vamvakopoulos, N.C. & Fiddes, J.C. (1984a) Evolution of the genes for the β -subunits of human chorionic gonadotropin and luteinizing hormone. *Nature, Lond.* **307**, 37–40.

- Talmadge, K., Boorstein, W.R., Vamvakopoulos, N.C., Gething, M.J. & Fiddes, J.C. (1984b) Only three of the seven human chorionic gonadotropin β -subunit genes can be expressed in the placenta. *Nucl. Acids Res.* **12**, 8414–8436.
- Tan, L. & Rousseau, P. (1983) The chemical identity of the immunoreactive LHRH-like peptide biosynthesized in the human placenta. *Biochem. Biophys. Res. Commun.* **109**, 1061–1071.
- Tata, J.R. & Smith, D.F. (1979) Vitellogenesis: a versatile model for hormonal regulation of gene expression. *Recent Progr. Horm. Res.* **35**, 47–90.
- Thompson, C.B., Challoner, P.B., Neiman, P.E. & Groudine, M. (1985) Levels of c-myc oncogene mRNA are invariant throughout the cell cycle. *Nature, Lond.* **314**, 363–366.
- Truong, A.T., Duez, C., Belayew, A., Renard, A., Pictet, R., Bell, G.I. & Martial, J.A. (1984) Isolation and characterization of the human prolactin gene. *EMBO J.* **3**, 429–437.
- Tsonis, C.G. & Sharpe, R.M. (1986) Dual gonadal control of follicle-stimulating hormone. *Nature, Lond.* **321**, 724–725.
- Ullrich, A., Coussens, L., Hayflick, J.S., Dull, T.J., Gray, A., Tam, A.W., Lee, J., Yarden, Y., Libermann, T.A., Schlessinger, J., Downward, J., Mayes, E.L.V., Whittle, N., Waterfield, M.D. & Seeburg, P.H. (1984) Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature, Lond.* **309**, 418–425.
- Vale, W., Rivier, J., Vaughan, J., McClintock, R., Corrigan, A., Woo, W., Karr, D. & Spiess, J. (1986) Complementary DNA sequences of ovarian follicular fluid inhibin show precursor structure and homology with transforming growth factor- β . *Nature, Lond.* **321**, 776–779.
- Von der Ahe, D., Janich, S., Scheidereit, C., Renkawitz, R., Schütz, G. & Beato, M. (1985) Glucocorticoid and progesterone receptors bind to the same sites in two hormonally regulated promoters. *Nature, Lond.* **313**, 706–709.
- Wasylyk, B., Kédinge, C., Corden, J., Brison, O. & Chambon, P. (1980) Specific *in vitro* initiation of transcription on conalbumin and ovalbumin genes and comparison with adenovirus-2 early and late genes. *Nature, Lond.* **285**, 367–373.
- Waterfield, M.D., Sprace, G.T., Whittle, N., Stroobant, P., Johnsson, A., Wasteson, A., Westermark, B., Heldin, C.-H., Huang, J.S. & Deuel, T.F. (1983) Platelet-derived growth factor is structurally related to the putative transforming protein of p28^{sis} of Simian sarcoma virus. *Nature, Lond.* **304**, 35–39.
- Weil, P.A., Luse, D.S., Segall, J. & Roeder, R.G. (1979) Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. *Cell* **18**, 469–484.
- Weinberger, C., Thompson, C.C., Ong, E.S., Lebo, R., Gruol, D.J. & Evans, R.M. (1986) The c-erb-A gene encodes a thyroid hormone receptor. *Nature, Lond.* **324**, 641–646.
- Weir, M.P. & Kornberg, T. (1985) Patterns of *engrailed* and *fushi tarazu* transcripts reveal novel intermediate stages in *Drosophila* segmentation. *Nature, Lond.* **318**, 433–439.
- Weisbrod, S. (1982) Active chromatin. *Nature, Lond.* **297**, 289–295.
- Welshons, W.V., Lieberman, M.E. & Gorski, J. (1984) Nuclear localization of unoccupied oestrogen receptors. *Nature, Lond.* **307**, 747–749.
- West, J.D. & Green, J.F. (1983) The transition from oocyte-coded to embryo-coded glucose phosphate isomerase in the early mouse embryo. *J. Embryol. exp. Morph.* **78**, 127–140.
- Williams, J.L. & Tata, J.R. (1983) Simultaneous analysis of conformation and transcription of A and B groups of vitellogenin genes in male and female *Xenopus* during primary and secondary activation by oestrogen. *Nucl. Acids Res.* **11**, 1151–1166.
- Wittig, S., Hense, S., Keitel, C., Elsner, C. & Wittig, B. (1983) Heat shock gene expression is regulated during teratocarcinoma cell differentiation and early embryonic development. *Dev. Biol.* **96**, 507–514.
- Wold, F. (1981) *In vivo* chemical modification of proteins (post-translation modification). *Ann. Rev. Biochem.* **50**, 783–814.
- Wolgemuth, D.J., Engelmyer, E., Duggal, R.N., Gizang-Ginsberg, E., Mutter, G.L., Ponzetto, C., Viviano, C. & Zakeri, F. (1986) Isolation of a mouse complementary DNA coding for a developmentally regulated testis-specific transcript containing homeo box homology. *EMBO J.* **5**, 1229–1236.
- Woychik, R.P., Stewart, T.A., Davis, L.G., D'Eustachio, P. & Leder, P. (1985) An inherited limb deformity created by insertional mutagenesis in a transgenic mouse. *Nature, Lond.* **318**, 36–40.
- Wu, G.-J. (1978) Adenovirus DNA-directed transcription of 5-S RNA *in vitro*. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2175–2179.
- Yamamoto, K.R. (1985) Steroid receptor regulated transcription of specific genes and gene networks. *Ann. Rev. Genet.* **19**, 209–252.
- Yang-Feng, T.L., Seeburg, P.H. & Francke, U. (1986) Human luteinizing hormone-releasing hormone gene (LHRH) is located on short arm of chromosome 8 (Region 8p11.2-p21). *Somat. Cell. molec. Genet.* **12**, 95–100.
- Yen, P.H., Allen, E., Marsh, B., Mohandas, T., Wang, N., Taggart, R.T. & Shapiro, L.J. (1987) Cloning and expression of steroid sulphatase cDNA and the frequent occurrence of deletions in STS deficiency: implications for X-Y interchange. *Cell* **49**, 443–454.
- Young, R.A., Hagenbuehle, O. & Schibler, U. (1981) A single mouse alpha-amylase gene specifies two different tissue-specific mRNAs. *Cell* **23**, 451–458.
- Ziff, E.B. (1980) Transcription and RNA processing by the DNA tumour viruses. *Nature, Lond.* **287**, 491–499.
- Ziff, E.B. & Evans, R.M. (1978) Coincidence of the promoter and capped 5'-terminus of RNA from the adenovirus 2 major late transcription unit. *Cell* **15**, 1463–1476.
- Zuber, M.X., John, M.E., Okamura, T., Simpson, E.R. & Waterfield, M.R. (1986) Bovine adrenocortical cytochrome P-450_{17 α} . Regulation of gene expression by ACTH and elucidation of primary sequence. *J. biol. Chem.* **261**, 2475–2482.

Received 20 March 1987