Molecular Analysis And Physical Mapping Of The Human 3ß-Hydroxysteroid Dehydrogenase Δ5/Δ4 Isomerase Gene Family

A thesis submitted for the degree of
Doctor of Philosophy
at the University of Glasgow

by

Alison Jane McVie

Division of Molecular Genetics
Institute of Biomedical and Life Sciences
Anderson College
University of Glasgow
Glasgow

December 1997

The research reported in this thesis is my own and original work, except where otherwise stated and has not been submitted for any other degree.

This thesis is dedicated to my mum, dad and Les,
with lots of love and thanks.

Table of contents

Chapter 3

Characterisation and sequencing of new members of the 3ß-HSD gene family

Chapter 4

Physical mapping of the 3ß-HSD gene family

Table of figures and tables

Chapter 1

Introduction

Figures:

Tables:

Chapter 3

Characterisation and sequencing of new members of the 3ß-HSD gene family

Figures:

Chapter 4

Physical mapping of the 3ß-HSD gene family

Figures:

Chapter 5

Hybrid selection

Figures:

Chapter 6

Discussion

Figures:

Tables:

Acknowledgements

Firstly, I would like to thank Roger Sutcliffe for his advice and support during the time spent in the lab completing my PhD. Thanks also to Jenny Varley for giving me the chance to spend time in her lab in Manchester, learning how to work with YACs and BACs. Thanks to Richard for being much more than an assessor and helping me understand the complexities of molecular evolution!!

Thanks to Marshall and the rest of the sixth floor for all the consumables I have borrowed over the last three years (I'll give it all back one day!!), and for putting up with me scanning and printing my thesis.

Thanks to everyone in Lab 21 especially Martin, Simon, Sandra, Nicola and Mary. Martin - I couldn't have done this without your endless support and your patience in the early days teaching me to do PCR!!! And Simon, thanks for your artistic impressions on my lab coat and for making the last days in lab 21 lots of fun!!!

A big thanks to all my friends in the genetics department and outwith for all the nights out in the Exchequer, then the Living Room and now Bar Zoo!! Thanks especially to Trish for the cinema nights and take-aways, I would never have seen those film with anyone else!! Thanks also to Kirsty and Karen for taking me away out of the world of science, and to Yvonne for her endless e-mails!!

On a personal note, thanks to my mum, dad and Les. What can I say?? You are the best family that anyone could ever wish for and Les you are my best friend as well as the best sister. I could never have done this without you and I don't really know how to thank you because you've done so much!!!

And Andy, I don't know what to say to you either!! What a time we have had writing this up, I don't think that I could have spent all this time in the department without you. You make me laugh (and cry!!), we've shared the occasional smoked sausage supper, you've given me tons of hugs and you've given me everything to look forward to in the future!!!

Thanks everyone!!!!!

Abstract

3ß-hydroxysteroid dehydrogenase (3ß-HSD) catalyses the conversion of 3ß-hydroxy-Δ5-steroids into the corresponding Δ4-3-ketosteroids and is essential for the biosynthesis of mineralocorticoids, glucocorticoids and sex hormones. Two isoforms encoded by two highly homologous, closely linked 3ß-HSD genes (HSD3B1 and 2) are known to be expressed in humans. However, Southern blot analysis and 3ß-HSD type II mutation screening suggested that there was more than two HSD3B genes in humans; therefore, two human genomic λgem11 libraries were screened with 3ß-HSD type I cDNA and bacteriophage clones containing novel 3ß-HSD sequences were identified. Two of the phage clones were characterised and the segments equivalent to HSD3B1 coding regions were sequenced. The 3ß-HSD coding sequence determined from both of these clones contained frameshift mutations resulting in premature stop codons, and it was concluded that these sequences were unprocessed pseudogene members of the 3ß-HSD gene family.

The library screens generated 5 new members of the 3ß-HSD gene family and these were mapped by fluorescent *in situ* hybridisation (FISH) to chromosome 1p13, the same region of the genome as HSD3B1 and 2. Specific oligonucleotide primer pairs were designed for each gene and using PCR the genes were mapped to a set of 3 overlapping yeast artificial chromosomes (YACs) and 9 overlapping bacterial artificial chromosomes (BACs). The gene order was subsequently confirmed using restriction analysis and 3ß-HSD-specific oligonucleotide probes. The orientation of HSD3B1, 2, ψ1 and ψ4 was determined by extensive restiction analysis of the BACs and the positions of the endpoints in three BAC clones. The estimated length of the entire contig is 500kb with the 3ß-HSD gene cluster over a centrally-based 235kb fragment.

cDNA selection techniques were established to detect expressed sequences from the region of the 3ß-HSD gene cluster. Three BAC clones from the contig were biotinylated and hybridised to a placental PCR-amplifiable cDNA library. The hybrids produced were isolated using streptavidin coated magnetic beads. Many 3ß-HSD transcripts were detected from these experiments indicating that the selection process was successful, however no previously identified genes were detected that could be localised to the 3ß-HSD locus. Although, several unknown sequences were discovered which may belong to unidentified genes present within or close to the 3ß-HSD gene cluster.

CHAPTER 1

Introduction

1.1 The definition of a hormone

The regulation of many biological processes in a multicellular organism requires communication between cells. This cell-cell communication is mediated in part by the endocrine system and molecules known as hormones. The classical definition of a hormone is a chemical substance produced by specialised tissues and secreted into the blood, within which it is carried to target organs. However, it is becoming clear from continued research that this definition is not quite correct (Bern, 1990). Hormones are no longer synthesised only by "specialised tissues". Chemical substances with hormonal activity exist throughout the animal and plant kingdom where discrete endocrine glands are not found, and hormones exist that exert their effects locally and these tend to have almost ubiquitous distribution (Carmeliet and Collen, 1997). "Blood" is not necessarily the only medium that hormones can pass through, local hormones may diffuse through extracellular fluid and some hormones, such as the insect hormones pheromones, are even transmitted via the atmosphere (Donascimento and Morgan, 1996). The term "target organ" is no longer accurate either; local hormones may stimulate the cell that synthesised them and plants produce hormones, known as alarmones, that can be secreted or used internally (Ozeretskovskaya *et al.*, 1994). Therefore, a broader and more accurate definition of a hormone is: a chemical substance that carries information between two or more cells, allowing the chemical co-ordination of bodily functions.

1.2 Hormone-Target relationships

The three main systems of hormone action are the endocrine system, the paracrine system and the autocrine system. In the classical endocrine system, a hormone is made in one part of the body and it reaches its target in another part of the body via the bloodstream. When the hormone remains within the tissue where it is synthesised and reaches nearby cells by diffusion this is known as the paracrine system and if the hormone influences the cell that has secreted it, this is the autocrine system. All of these systems are open, there are no barriers and the selectivity of the target cell is determined by the presence or absence of receptors to the particular hormone.

Other systems have also been determined, including closed systems of hormone action. The first if theses is known as the crytocrine system, where a hormone is secreted into a closed environment. The cells involved here include sertoli cells and nurse cells (Fritz, 1994) or thymus nurse cell and T-lymphocytes (Brelinska and Warchol, 1997), and in both these cases an special intimacy exists between cells. The juxtacrine system involves membrane bound hormone precursors, these may be cleaved to release

an active soluble peptide hormone or remain attached to the plasma membrane retaining their biological activity, affecting cells in close contact only. Finally, in the intracrine system a hormone is synthesised and bound to its receptor internally. This includes the endogenous generation of hormones in peripheral tissues from precursors synthesised in a classical endocrine tissue, for example, 75-100% of oestrogens in postmenopausal woman are generated in peripheral tissues from adrenal precursors (Labrie *et al.*, 1995).

1.3 Principles of hormone action

The chemical structure of hormones is extremely diverse. The most abundant hormones are the peptide and protein hormones. These molecules can range in size from a simple tripeptide to complex glycoproteins with multiple subunits. Modified amino acids can also act as hormones, for example tyrosine is the amino acid precursor for the catecholamines and histimine is the precursor for histamine. Lipids are another source of hormones including steroid hormones, prostaglandins and insect pheromones (fatty acid derivatives). The nucleotides are also represented, these include some insect pheromones, plant hormones (cytokinins; Chernyadev, 1997), and 1-methyladenine has been identified as a hormone in starfish (Mita, 1992). Finally, some oligosaccharides hormones have been recently identified, including plant hormones derived from cell wall breakdown (Darvill *et al.*, 1992); and an aggregation factor identified in sponges that appears to be a glycan (Misevic and Burger, 1990).

Despite the structural diversity of hormones, their mechanism of action generally depends on their solubility in water. Hydrophobic (or lipophillic) hormones, including steroid and thyroid hormones, are difficult to store within a cell as they can pass through the plasma membrane easily and therefore these hormones are synthesised for immediate use. Hydrophobic hormones are not water soluble, and require serum transport molecules to enable them to manoeuvre in the bloodstream. Serum transport molecules contain hydrophobic pockets and this protects the hormones, increasing their half-life. When hydrophobic hormones reach their target they diffuse through the plasma membrane and bind to cytoplasmic or nuclear receptors. Once activated the receptor-hormone complex elicits direct cellular effects (Evans and Bergeron, 1988). In contrast, hydrophilic hormones, mainly peptide and protein hormones, are stored within vesicles until needed, transported free in serum and as a result are eliminated rapidly from the circulation. Also, when hydrophilic hormones reach their target cell they interact with receptors at the cell surface, generating secondary signals to affect cellular processes, i.e. they induce cellular processes indirectly (Dohlman *et al.*, 1987).

1.4 Control of hormone activity

To enable co-ordination of tissues within the body the activity of all hormones must be very tightly regulated. To do this, hormonal regulation conforms to a limited number of mechanisms. The simplest mechanism is negative feedback, where rising levels of a particular hormone will shut off its synthesis to maintain the desired concentration. For example, the hypothalamus releases corticotrophin releasing hormone (CRH) which stimulates the anterior pituitary to release adrenocorticotrophic hormone (ACTH) which stimulates the adrenal gland to synthesise cortisol. Cortisol, in turn, will act on the anterior pituitary and hypothalamus to prevent further release of the stimulatory hormones and therefore reduce further cortisol production (Kemppainen and Behrend, 1997). Positive feedback, where rising levels stimulate further production of the hormone, is less common and tends to occur only where there is a clear termination point, e.g. oxytocin is released to induce smooth muscle contraction in the uterus during the birth process, the uterus will contract harder and continue to stimulate oxytocin secretion until the foetus is expelled (Neumann *et al.*, 1996). The last type of feedback is known as cycle-dependant feedback and is dependant on other physiological parameters, e.g. oestrogen normally has a negative feedback on the hypothalamus, however during mid-cycle this changes to a positive feedback and oestrogen levels will rise until they trigger a surge of luteinizing hormone which leads to ovulation (Hillier, 1985).

1.5 Steroid hormones

The structure of steroid hormones corresponds to that of perhydrocyclopenetenphenanthrene, with 4 rings of C-atoms (3 cyclohexane and 1 cyclopentane) forming the basic molecular skeleton. Adjoining pairs of rings each share 2 carbon atoms at the ring junctions and almost all natural steroids usually possess 2 methyl groups at "angular" or "bridgehead" positions where 2 rings meet (Fig. 1.1). The diversity of biological effects of steroid hormones depends on the nature of the modification of this basic structure. Modifications can include the unsaturation of C-C bonds within the rings or the attachment of hydroxyl, ketone or other groups to specific C-atoms (examples in Fig. 1.3).

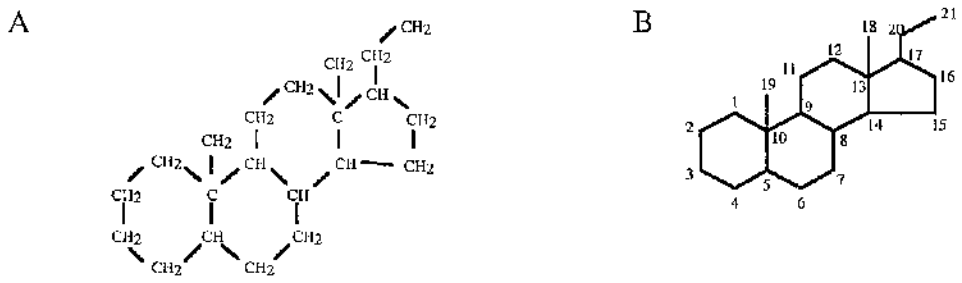


Fig. 1.1 A: Steroid molecular structure
B: Steroid skeleton with locants

Steroid hormones are hydrophobic molecules and can therefore diffuse through the plasma membrane of cells and exert their effects at nuclear receptor sites (Evans and Bergeron, 1988). One consequence of this is that there is a lag time of 30-60 minutes, sometimes longer, between the time of exposure of the target cell and the onset of a biological response. The mechanism of steroid hormone action is described in Fig. 1.2.

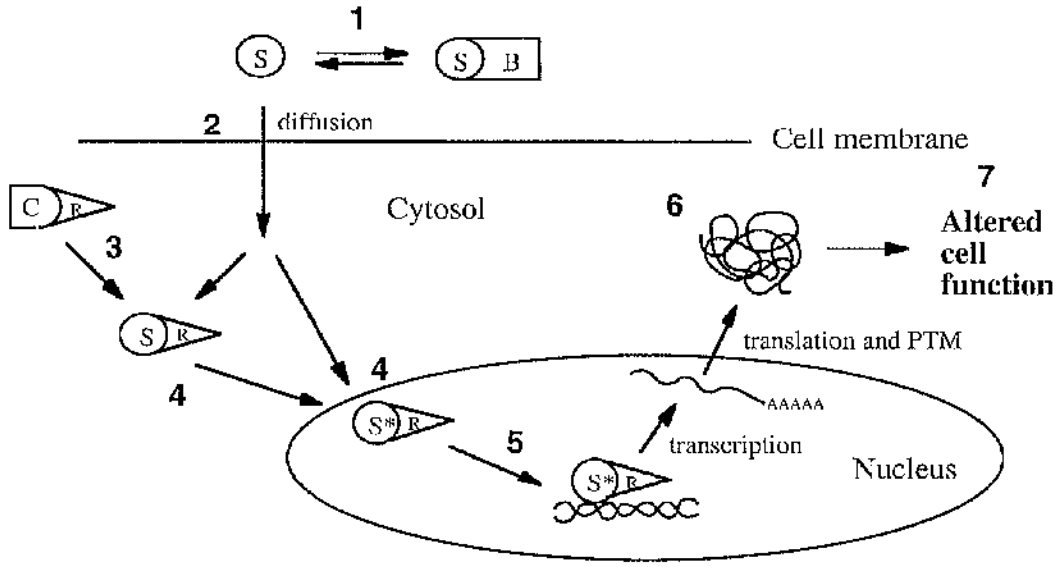


Fig. 1.2 Mechanism of steroid hormone action. Free steroid hormone is in equilibrium with steroid bound to serum transport molecules (1), the free steroid can diffuse across the plasma membrane of the target cell (2). Steroid hormone receptors are associated with chaperones, which maintain the receptors in an inactive but ligand-friendly conformation (3). Binding to the steroid-hormone receptor occurs in the cytoplasm or the nucleus (4). The receptor-hormone complex binds to a hormone responsive element in the promoter region of a specific gene, this induces transcription of that gene (5). The mature RNA leaves the nucleus, translation and post-translational modification (PTM) follows and the final gene-product is obtained (6). The protein will modify the cell function according to the nature of the specific hormone action (7). S=steroid hormone, B=serum transport molecule, R=steroid-hormone receptor, C=chaperone.

Steroid receptor proteins consist of several domains; a DNA binding domain, a nuclear localisation signal, a ligand binding domain and several transcriptional activator functions (Beato *et al.*, 1995). Steroid hormone receptors regulate transcription of specific genes by induction or suppression by binding to palindromic DNA sequences known as hormone responsive elements. After binding to the DNA it is thought that the receptor interacts with components of the basal transcriptional machinery and with sequence-specific transcription factors. Unligated steroid hormone receptors are associated with chaperones which maintain the receptors in an inactive but ligand friendly conformation. The specificity of a single cell for a particular hormone arises because each cell can contain separate receptor proteins that are specific for different steroid hormones.

1.6 Major classes of steroid hormones

There are 5 major classes of steroid hormones; the mineralocorticoids, the glucocorticoids, the androgens, the progestogens and the oestrogens.

Mineralocorticoids:

The mineralocorticoids regulate the water and electrolyte balance in the body, in particular the kidney, the gastrointestinal tract, the salivary glands and the sweat glands. The main mineralocorticoid is aldosterone and its primary site of action is the distal tubule and the collecting ducts of the kidney, where it increases sodium reabsorbtion and excretion of potassium and hydrogen ions. Aldosterone increases the blood pressure, partly by increasing the plasma volume and partly by increasing the sensitivity of the arteriolar muscle to vasoconstrictor agents (Morris, 1981).

Glucocorticoids:

Glucocorticoids maintain critical biological processes at times of stress. The principal action of cortisol, the main glucocorticoid, is to increase the amount of glucose available for energy metabolism. The hormone causes both a decrease in the uptake and utilisation of glucose and an increase in gluconeogenesis. There is decreased protein synthesis and increased protein breakdown, particularly in muscle, this increases the availability of amino acids to be deaminated for *de novo* glucose production. In addition to this, cortisol also has powerful anti-inflammatory and immunosuppressive effects. It affects all types of inflammatory reactions whether caused by invading pathogens, by chemical or physical damage or inappropriate immune responses. Cortisol can also act as a mineralocorticoid, due to its ability to stimulate the mineralocorticoid receptor (MR). The type I MR actually has the same affinity for cortisol and aldosterone, however the renal receptor is protected from

cortisol by the dehydrogenase activity of 11ß-HSD, which converts cortisol into the inactive glucocorticoid cortisone, allowing aldosterone free access to the receptor to regulate sodium homeostasis. (Walker and Edwards, 1991).

Androgens:

The androgens are the principal male hormones, the androgen testosterone is responsible for the initiation and maintenance of spematogenesis, the formation of the male phenotype during sexual differentiation (development of the male internal genitalia - epididymis, vas deferens, seminal vesicles and ejaculatory ducts) and the promotion of sexual maturation during puberty. Testosterone can be converted by 5α-reductase II into the more potent androgen 5α-dihydrotestosterone (Russell and Wilson, 1994). This binds with much greater affinity to the androgen receptor, amplifies androgen activity and promotes male differentiation, including the development of the external genitalia (penis and scrotum) and the prostate gland (Wilson, 1978).

Progestogens:

These hormones are of utmost importance in the ovary. Progesterone is the principal product of the corpus luteum and is required for implantation of the fertilised ovum and maintenance of pregnancy. It also inhibits uterine contractions, increases the viscosity of cervical mucous, promotes development of the breast and increases basal body temperature. 17-hydroxyprogesterone is also secreted by the corpus luteum but has little known biological effect (Kholkute *et al.*, 1995).

Oestrogens:

Like the progestogens, the oestrogens are extremely important in the development, growth and physiology of the female reproductive tract. The principal and most potent oestrogen is oestrodiol-17ß, which is synthesised in the ovary. This hormone promotes development of female secondary sexual characteristics, thickening of the vaginal mucosa, thinning of the cervical mucosa and the development of the ductal system in the breast (Hillier, 1985).

1.7 The major pathways of steroid hormone biosynthesis

Steroid hormone biosynthesis takes place in a wide variety of tissues throughout the body. It is now known that it does not only take place in the classical endocrine tissues, such as the adrenal gland, the gonads and the placenta during pregnancy (Labrie*et al.*, 1992), but also in peripheral tissues like the skin (Dumont *et al.*, 1992), the breast (Lachance *et al.*, 1990), the liver (Zhao *et al.*, 1991), the kidney (Devine *et*

al., 1986) and even the brain (Martel *et al.*, 1994). The peripheral tissues tend to synthesise sex steroids from precursors released from the adrenal gland and this process has been given the term intracrinology (Labrie *et al.*, 1995). Through intracrinology, locally produced sex steroids may exert their action within the same cells that the final steps of their synthesis took place in without the need to be released into the extracellular compartment and the general circulation.

Adrenal Glands:

The adrenal glands are actually two glands in one; the outer layer of an adrenal gland is the adrenal cortex, which can synthesise steroids, and the centre layer, the adrenal medulla synthesises catecholamines (Kemppainen and Behrend, 1997). Each part of the adrenal has a distinct blood supply. The adrenal cortex can be further subdivided into three zones; the zona glomerulosa, containing ball-like clusters of cells and found below the adrenal capsule, the zona fasticulata, the middle layer containing cells in column-like pattern, and the zona reticularis, the innermost layer with cells in a netlike arrangements. Each zone also represents a functional division as the mineralocorticoids are synthesised from the zona glomerulosa, and the sex steroids and the glucocorticoids are synthesised in the zona fasticulata and the zona reticularis. The main reason for the functional zonation is the differential expression of the steroidogenesis enzymes.

The first enzymatic stage in the synthesise of all steroid hormones is the action of P450scc on cholesterol (Fig. 1.3). This enzyme cleaves the side chain of cholesterol to form pregnenolone. This is the rate-limiting enzymatic stage of the process, however the true rate-limiting step is the movement of cholesterol from the outer mitochondrial membrane to the inner membrane. This translocation is acutely regulated by the trophic hormones. It was noted from early studies that hormone production from cells stimulated to synthesise steroids was blocked by inhibiting protein synthesis (Davis *et al*, 1968). This suggested that a newly synthesised, hormone sensitive protein functioned to transfer the cholesterol molecule. There have been several proteins postulated to undertake this role (Stocco and Clark, 1997), one of these is a 30kDa protein studied by Pon *et al.* (1986), Pon and Ormejohnson (1986) and Stocco and Clark (1996a and b); this protein was named the steroidogenic acute regulatory protein (StAR). Although much evidence suggests that this protein is involved in cholesterol transfer, the mechanism of the translocation is still unknown. Stocco and Clarke (1997) have proposed a model whereby StAR may act in the transfer of cholesterol to P450scc. They suggest that a 37kDa precursor of StAR is synthesised in the cytosol in response to hormonal stimulation. This protein is then targeted to the mitochondria and during the processing of this molecule to form the

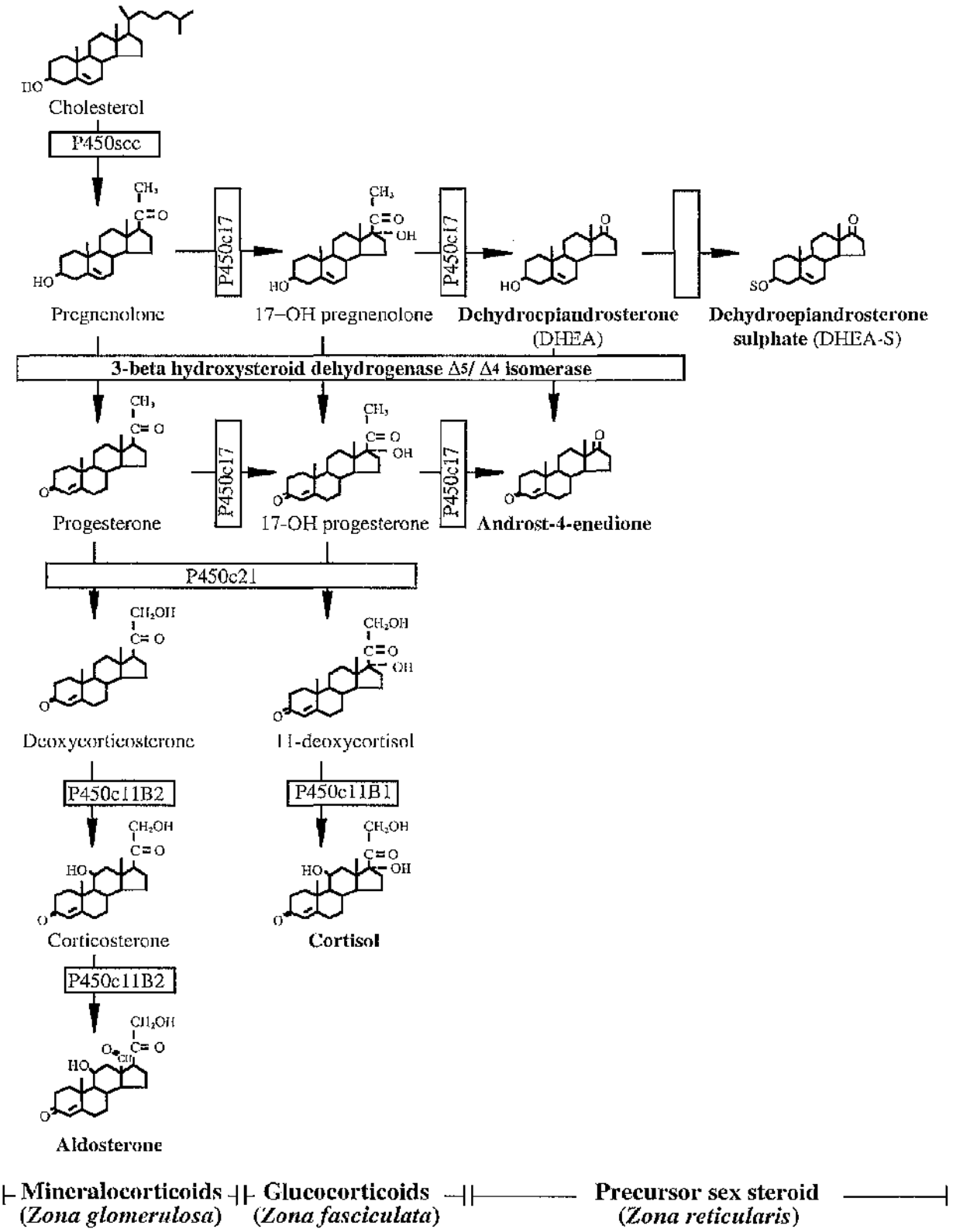


Figure 1.3. Steroid biosynthetic pathways in the three functionally distinct zones of the adrenal cortex. Aldosterone and cortisol are the principal mineralocorticoid and glucocorticoid secreted by the *zona glomerulosa* and *reticularis* . Precursor sex steroids, DHEA, DHEA-S and androstenedione are secreted mainly from the *zona reticularis*. Reproduced by kind permission from Martin McBride.

30kDa StAR protein, "contact sites" are made between the outer and inner mitochondrial membranes. While the protein is being processed, cholesterol is somehow transferred from the outer to the inner membrane towards P450scc. Little is known about this mechanism, but it is possible that the formation of a protein complex or the physical contact between the membranes creates a hydrophobic core through which cholesterol could pass through. Once the mature protein has been synthesised the membranes separate and further cholesterol transfer is prevented.

Once synthesised from cholesterol by P450scc pregnenolone has three options; it can proceed down the mineralocorticoid pathway, it can be 17-hydroxylated and follow the glucocorticoid pathway or it can undergo 17-hydroxylation and cleavage of the C17-20 bond leading to sex hormone production (Fig. 1.3). This is where the functional zonation and the specificity of enzyme expression becomes significant. 3ß-HSD is the next enzyme in the pathway and is expressed in all three zones of the adrenal cortex, however P450c17 is only expressed in the zona fasticulata and the zona reticularis (Sasano *et al.*, 1989). In the zona glomerulosa, pregnenolone is not diverted along the Δ5 pathway but converted into progesterone by 3ß-HSD and onto aldosterone. Within the zona fasticulata and the zona reticularis, pregnenolone is converted into 17-hydroxypregnenolone and either progresses to cortisol or is cleaved at the C17-20 bond and converted to DHEA which is released and converted into sex hormones in the periphery.

Developmental changes can determine the flow of steroids from glucocorticoid to androgen pathways, e.g. puberty, which changes the adrenal cortex capacity for synthesising steroids. Mechanisms which regulate these changes include the dependence of C17-20 cleavage activity on concentrations of P450 reductase which can be altered in different steroidogenic tissues (reviewed by Miller *et al.*, 1997), the regulation of 3ß-HSD and 17α-hydroxylase activities (Bird *et al.*, 1996, see section 1.8.1) and the levels of the substrates and products available (Hanukoglu, 1992).

Progesterone and 17-hydroxylated progesterone are converted into deoxycorticosterone and 11-deoxycortisol by P450c21 in the pathways of mineralocorticoid and glucocorticoid respectively. P450c21 is expressed in all three zones of the adrenal cortex (Sasano *et al.*, 1988). From here, deoxycorticosterone is catalysed to corticosterone, then 18-hydroxycorticosterone which is hydroxylated to produce aldosterone and 11-deoxycortisol is converted into cortisol. Two different but closely related enzymes act to complete the aldosterone and the cortisol synthesis. The genes CYP11B1 and CYP11B2 are located in tandem and encode two enzymes which are 93% identical. CYP11B1 is expressed in the zona fasiculata and the zona reticularis

and encodes P450c11 which catalyses 11-deoxycortisol to cortisone, whereas CYP11B2 is expressed only in the zona glomerulosa and encodes P450c18 which performs the last three catalytic steps in aldosterone production, 11ß-hydroxylation, 18-hydroxylation and 18-methyl oxidation (White and Pascoe, 1992).

Gonads:

The majority of the sex hormones are produced by the testis in the male and the ovaries in the females. Although, the knowledge surrounding the importance of peripheral conversion from adrenal precursors is increasing, and it has been estimated that 30-50% of total androgens in man and possible more oestrogens in woman are synthesised in peripheral intracrine tissues from the adrenal precursors DHEA and DHEA-sulphate (Labrie *et al.*, 1997). However, the testes and the ovaries are still the major tissues involved in sex hormone synthesis.

Testosterone synthesis in males is undertaken in the Leydig cells of the testis. Cholesterol is converted to androstenedione by P450scc, P450c17 and 3ß-HSD and is then catalysed to testosterone by the action of 17ß-HSD. Multiple isoforms of 17ß-HSD exist (Andersson and Moghrabi, 1997), type 3 17ß-HSD is expressed in the testis and promotes the conversion of androstenedione into testosterone. Testosterone can be further converted into the more potent androgen 5α- dihydrotestosterone by 5α-reductase II (Fig. 1.4)

In the ovary, the synthesis of progesterone and androstenedione is identical to that in the Leydig cells in the testis and is undertaken in the theca interna cells. Granulosa cells are the major site of oestrogen production in females. P450arom is expressed in these cells and converts androstendione into oestrone and testosterone (via 17ß-HSD) into oestradiol (Fig. 1.5).

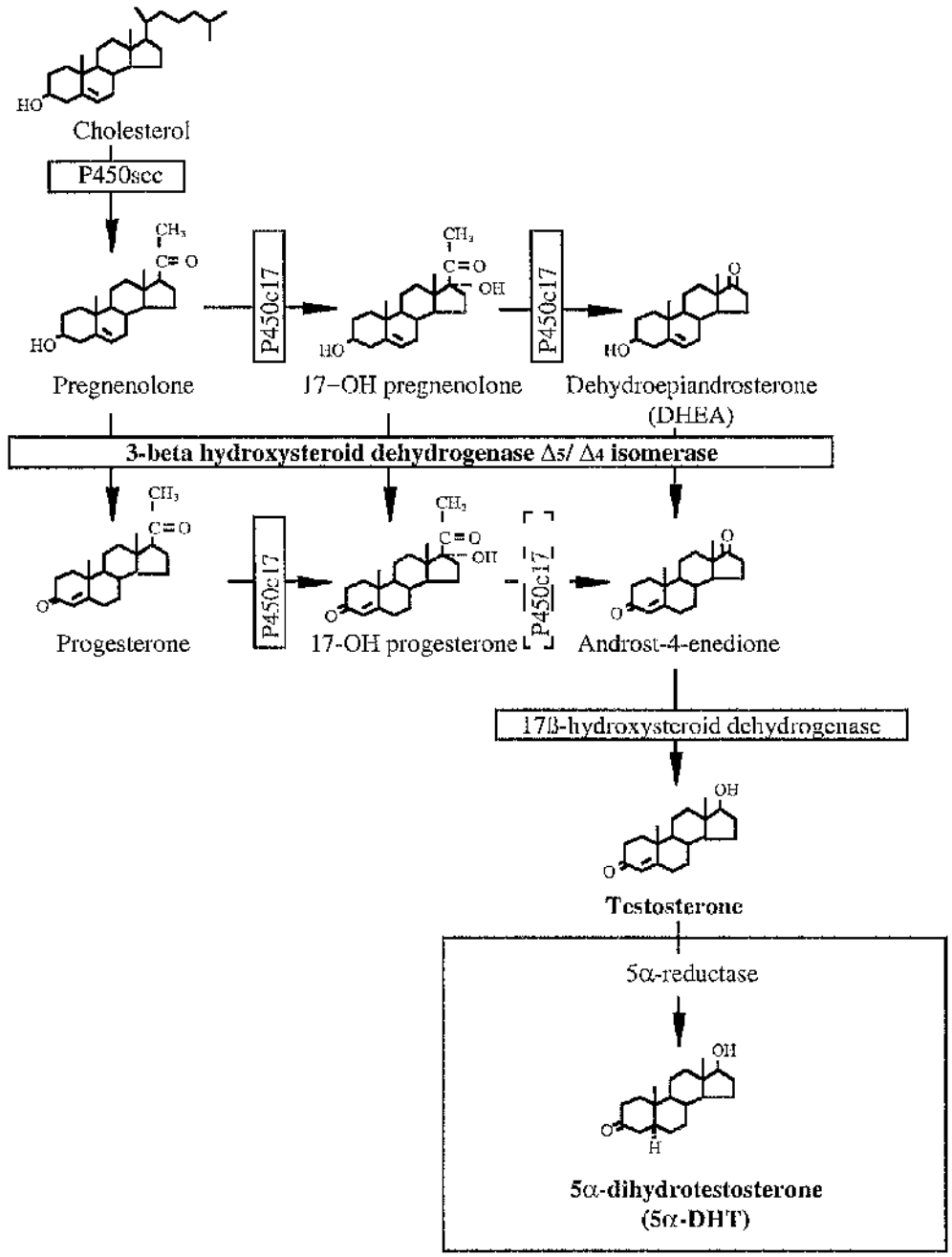


Figure 1.4. Leydig cell steroidogenesis.
Testosterone is the major androgenic steroid hormone secreted by Leydig cells. Leydig cells do not metabolise 17-OH progesterone efficiently. The potent androgenic hormone 5α-DHT is only secreted in small amounts; the major sites of formation and action of 5α-DHT are in peripheral tissues (extraadrenal and extragonadal). Reproduced with kind permission from Martin McBride.

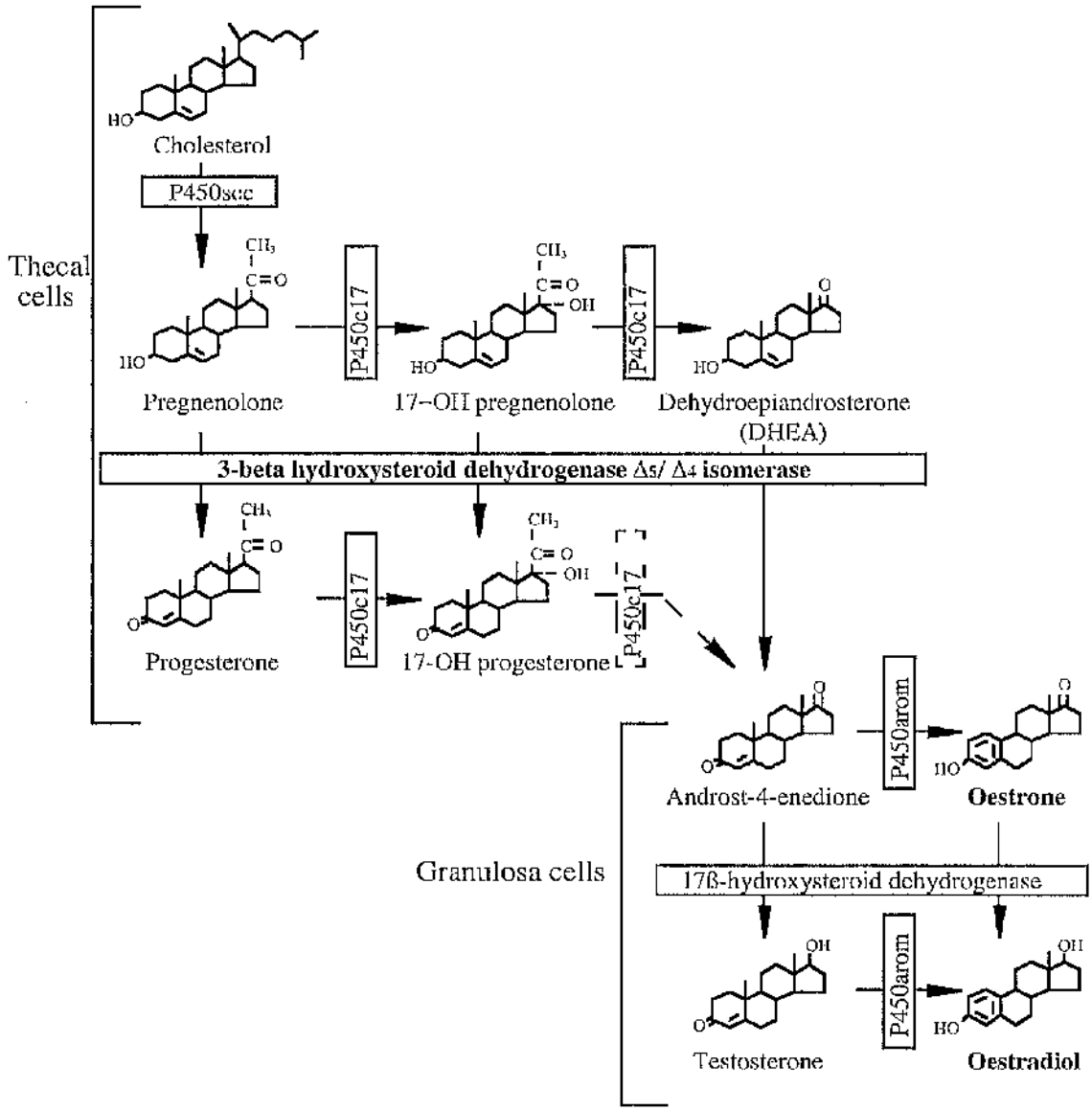


Figure 1.5. Principal pathways of steroid hormone biosynthesis in the human ovary. Each cell type of the ovary contains the complete enzyme complement for the formation of oestradiol from cholesterol; however the concentrations of various enzymes differ amoung the cell types, and consequently hormones are synthesised in thecal and granulosa compartments as indicated . As in Leydig cells ovarian cell types do not metabolise 17-OH progesterone efficiently.
Reproduced with kind permission from Martin McBride

Placenta and human foetal adrenal:

During pregnancy, the placenta must act together with the human foetal adrenal gland to provide the necessary hormones required to maintain the pregnancy and allow the foetus to develop normally. The corpus luteum supplies the progesterone requirement for the first 6 weeks of pregnancy until the placental progesterone synthesis takes over as the primary source after 12 weeks. The placenta does not express P450c17,

therefore no conversion of C21 to C19 steroids can occur. This is overcome by using DHEA-sulphate (DHEA-S) of foetal and maternal adrenal origin as the oestrogen precursor. DHEA-S is converted to 16α-DHEA-S in the foetal liver and then it travels to the placenta where it has the capacity to produce androstenedione, 16-OH androstenedione and testosterone. The high amount of aromatase activity in the placenta means that oestrone, oestriol and oestradiol can be produced from the above substrates respectively (Fig. 1.6; Siteri and MacDonald, 1967).

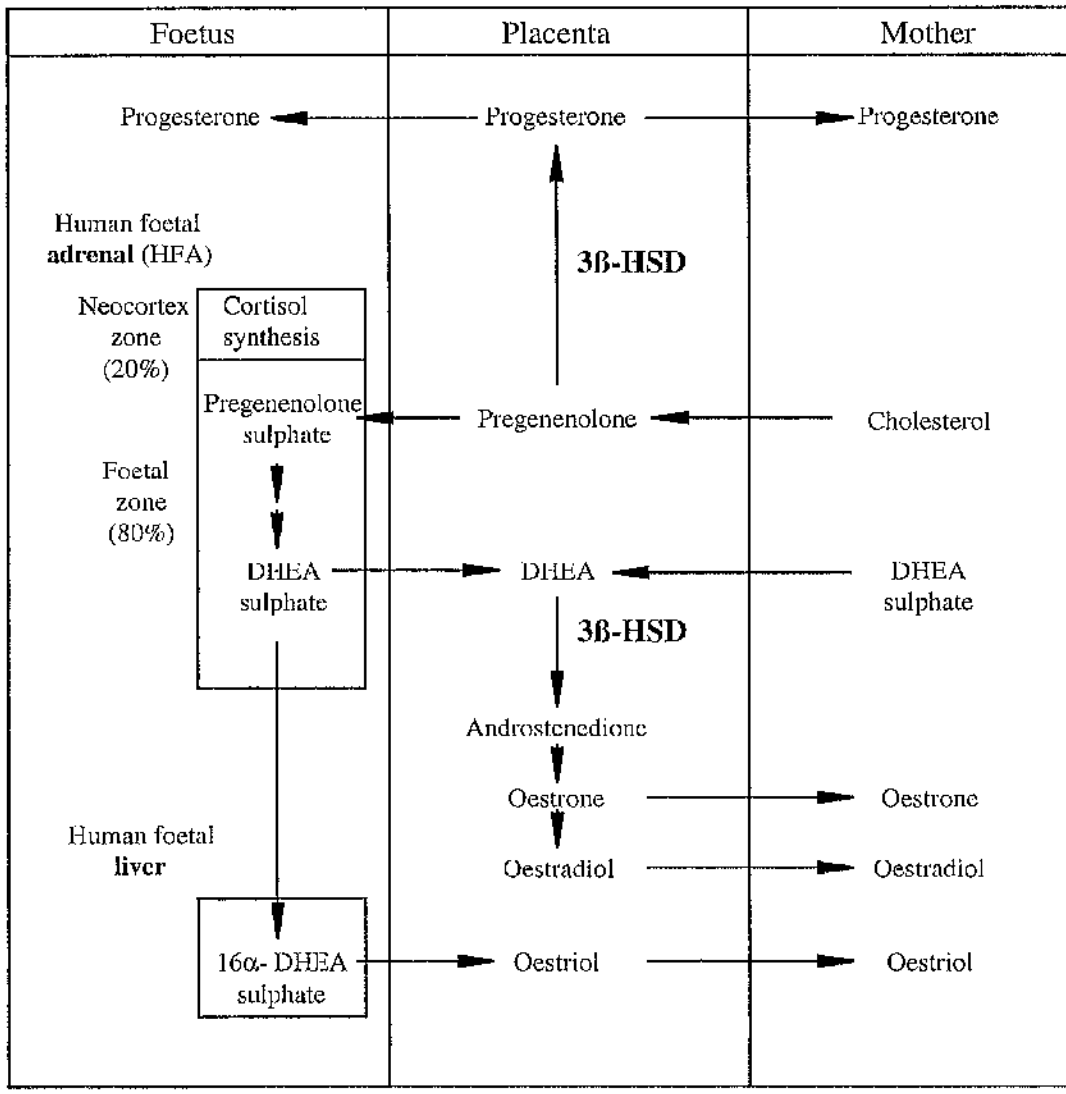


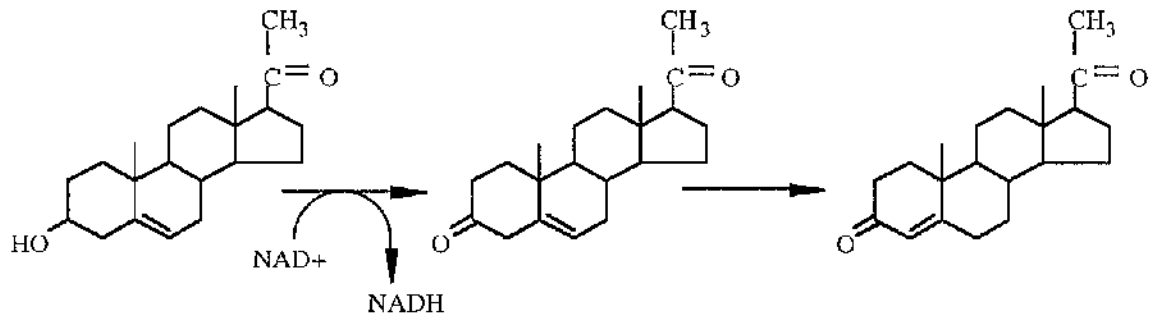
Figure 1.6. Biosynthesis of progesterone and oestrogens (oestrone, oestradiol and oestriol) from the fetal-placental unit. Placental 3ß-HSD enzyme activity is required for the maintenance of pregnancy and normal development of the fetus. Reproduced with kind permission from Martin McBride.

1.8 3ß-hydroxysteroid dehydrogenase

3ß-hydroxysteroid dehydrogenase (3ß-HSD) is an enzyme important in the synthesis of all classes of steroid hormones - the mineralocorticoids, the glucocorticoids and the sex hormones. 3ß-HSD converts 3ß-hydroxy-Δ^5-steroids into the corresponding Δ^4-3-ketosteroids, firstly by 3ß-hydroxysteroid dehydrogenation and then 5-ene-4-ene isomerisation. For example, pregnenolone is converted into progesterone, via 5-pregnene-3,20-dione (Fig. 1.7A). 3ß-HSD also converts 17-hydroxypregnenolone into 17-hydroxyprogesterone and dihydroepiandrosterone (DHEA) into androstenedione. The preferred cofactor for these reactions is NAD+.

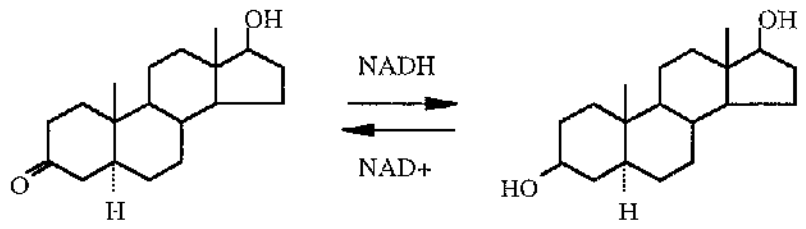
In addition to this activity, 3ß-HSD is responsible for the interconversion of 3-keto-5α-ketosteroids into 3ß-hydroxysteroids, utilising NADH/NAD+ as cofactor. i.e. the interconversion of 5α-dihydrotestosterone and 5α-androstane-3ß,17ß-diol and androstanedione and epiandrostanedione (Fig. 1.7B).

A



5-pregnenolone **5-pregnene-3, 20-dione** **4-progesterone**

B



Dihydrotestosterone **5α-androstane-3ß,17ß-diol**

Fig. 1.7 Enzymatic reactions catalysed by 3ß-HSD.
A: Two step catalytic reaction converting pregnenolone into progesterone. The first step is the dehydrogenase reaction using NAD+ as cofactor,followed by the isomerisation of Δ5 to Δ4.
B: Interconversion of 3-keto-5α-ketosteroids into 3ß-hydroxysteroids using NADH/NAD+ cofactor.

1.8.1 Molecular biology of 3ß-hydroxysteroid dehydrogenase

Characterisation of the human type I and type II 3ß-HSD:

Evidence for a critical role for 3ß-HSD in steroidogenesis was first revealed in the early 1960s. Bongiovanni (1961) described an unusual steroid pattern in congenital adrenal hyperplasia which he determined was caused by a deficiency of 3ß-HSD. The 3ß-hydroxysteroid dehydrogenase-isomerase system was purified by Ford *et al.*(1974) using sheep adrenal cortical microsomes, whilst the activity of human placental 3ß-HSD was described by Ferre *et al.*(1975). Rabe *et al.*(1982) achevied the partial characterisation of the placental 3ß-HSD from mitochondria and Ishiiohba *et al.*(1986) purified the testicular 3ß-hydroxysteroid dehydrogenase/isomerase. More recently, Thomas *et al.*(1988) reported the copurification of the dehydrogenase and isomerases activities from human placental microsomes and described the kinetics of substrate utilisation and product inhibition. Thomas *et al.*(1989) extended his study to human placental mitochondrial 3ß-HSD and compared the properties of the enzymes from the microsomes and the mitochondria. Both enzymes migrated identically on SDS-PAGE producing a single band of protein at approximately 42kDa. Kinetic data showed that NAD+ was the desired cofactor for the substrates pregnenolone and DHEA, with NADP+ utilised 90% less efficiently. Within this report, Thomas published a 29 amino acid sequence that corresponded to the amino terminus of the microsomal and mitochondrial 3ß-HSD.

Polyclonal antibodies were raised to placental 3ß-HSD and used to screen λgt11 expression libraries to isolate cDNA clones corresponding to 3ß-HSD (Luu-The *et al.*, 1989; Lorence *et al.* 1990a). They isolated clones that contained an open reading frame of 372 amino acids with the first 29 amino acids corresponding to the sequence obtained by Thomas *et al.*(1989). In addition, the calculated molecular weight of this peptide was 42,216 (Luu-The *et al.*, 1989) and 42,120 (Lorence *et al.*, 1990a), which also corresponds to the estimated molecular mass of 42kDa by Thomas *et al.* (1989). Lorence *et al.* (1990a) inserted the cDNA into a modified CMV expression vector and expressed it in COS cells (non-steroidogenic monkey kidney tumour cells). The expression protein catalysed the conversion of pregnenolone, 17-hydroxypregnenolone and DHEA into progesterone, 17-hydroxyprogesterone and androstenedione respectively. The reverse reaction was not observed. Cell homogenates of transfected COS cells also oxidised 5α-androstane-3ß,17ß-diol into 5α-dihydrotestosterone when supplied with NAD+ and, upon addition of NADH, reduced 5α-dihydrotestosterone to 5α-androstane-3ß,17ß-diol. Therefore, it was established that a single polypeptide could catalyse both the dehydrogenation and isomerisation steps and could oxidise 4 major substrates, pregnenolone, 17-hydroxypregnenolone,

DHEA and androstane-3ß,17ß-diol. However, Southern analysis of human genomic DNA hybridised with 3ß-HSD sequences suggested that there was more than one isoform of 3ß-HSD, as additional restriction fragments to those predicted were observed (Lachance *et al.*, 1990; Lorence *et al.*, 1990a).

It was known that the adrenal glands and the testis contained immunoreactive 3ß-HSD protein (Doody *et al.*, 1990; Milewich *et al.*, 1991) and with the detection of multiple fragments from the Southern analysis of human genomic DNA it was suspected that there was more than one 3ß-HSD present in humans. In addition, Gibb *et al.*(1985) described different catalytic constants for the human ovarian and human placental 3ß-HSDs. Rheaume *et al.* (1991) screened a human adrenal cDNA library with the placental 3ß-HSD cDNA. They isolated a cDNA with 93% similarity to type I and containing an open reading frame of 371 amino acids, this was designated 3ß-HSD type II. They inserted both the type I and the type II cDNAs into expression vectors to compare the reaction kinetics. Type I possessed a 3ß-HSD activity higher than type II with respective Km values of 0.24µM and 1.2µM for pregnenolone and 0.18µM and 1.6µM for DHEA, while the specific activity (Vmax) was equal for both when standardised for the amount of protein translated. In addition, type I had a higher oxidoreductase activity using DHT as substrate in the presence of NADH, with Km values of 0.26µM and 2.7µM for type I and II respectively. In fact, the relative enzyme activity (Vmax/Km) of type I is 5.9, 4.5 and 2.8-fold higher than type II using pregnenolone, DHEA and DHT as substrates respectively.

Gene structure of the human type I and II 3ß-HSD:

The gene for 3ß-HSD type I (HSD3B1) was characterised first. The type I cDNA was used as a probe to screen a human EMBL 3 genomic library to isolate the genomic copy of type I 3ß-HSD (Lorence *et al* ., 1990b; Lachance *et al.*, 1990). The gene spans approximately 8 kb of DNA and contains 3 exons with sequence identical to that of the cDNA (Fig. 1.7). Examination of the nucleotide sequence revealed a putative TATA element (TATATAA) 26 bp upstream of the exon 2 sequence. Primer extension analysis using poly (A)+ RNA from the placenta and the corpus luteum indicated that the RNA initiated upstream from the TATA box and that there was an untranslated exon 5' to the first coding exon containing 59 bp. Therefore, HSD3B1 actually contained 4 exons and 3 introns. The first exon contains 59bp and is untranslated, the second exon contains 232bp and encodes some 5' untranslated, residues 1-48 and the first nucleotide of residue 49. Exon 3 encodes the second and third nucleotides of residue 49, residues 50-103 and the first nucleotide of residue 104. The last exon contains 1218bp and encodes the second and third nucleotides of residue 104, residues 105-372

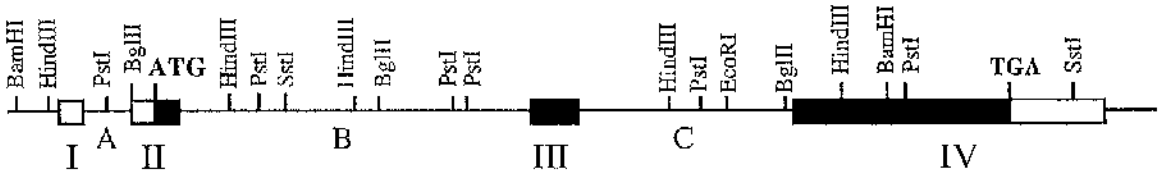
and the whole of the 3' untranslated. The introns separating the exons contain 129, 3383 and 2162bp respectively. A putative CAAT binding sequence is located 57 nucleotides upstream from the TATA box.

The type II gene (IISD3B2) was isolated from an EMBL-3 human genomic library by Lachance *et al.* (1991). HSD3B2 contains 4 exons and 3 introns included within a total length of 7,881bp. Intron/exon junctions were elucidated by comparison of the cDNA, and they are in agreement with type I. The exons contain 57, 231, 165 and 1,214bp respectively and are separated by introns of 128, 3,383 and 2,162 bp. One nucleotide is missing in the 5' untranslated region, as well as four nucleotides in the 3' untranslated compared to HSD3B1. Nucleotide changes at positions -2 and +2 results in the shift of the translational start codon and the loss of one amino acid in type II, compared to type I. This accounts for the difference of one amino acid between type I and II and knocks the residue numbers at the ends of exons out by one. The nucleotide sequence of the type II exons have 77.4, 91.8, 94 and 91% similarity with the corresponding exons of type I with 84, 80.3 and 73.9% similarity for the introns.

Primer extension analysis indicated that the transcription start site is 270 bp upstream from the ATG codon, this is the same position as type I. 1251 bp of the upstream region of type II was sequenced by Lachance to examine the promoter region. TATA elements were detected 28 and 140 nucleotides upstream from the transcription start site and a putative reverse CAAT motif (ATTGG) was located 178bp from start. The difference in TATA and CAAT boxes between type I and type II might suggest that type II has different RNA polymerase binding sites from type I. HSD3B2 also contains a TATA-rich repetitive sequence of 179bp within intron 3 that is absent in type I. Secondary structure analysis of this sequence suggests that it contains 6 hairpin loops which may be involved in transcription termination.

Schematic representation of the structure of HSD3B1 and HSD3B2 are found in Fig. 1.8, and the full sequence of the genes can be found in appendix i.

IISD3B1



HSD3B2

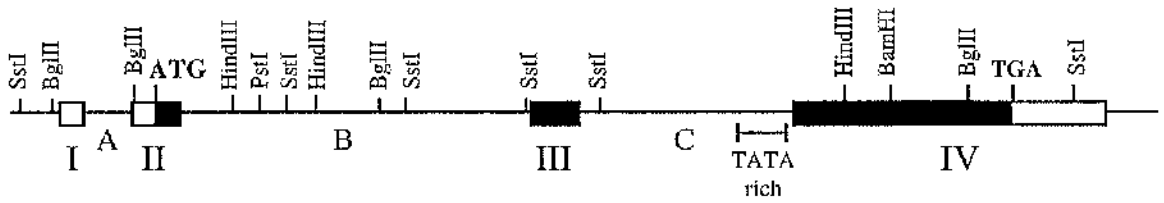


Fig. 1.8 Intron/ exon structure and partial restriction map of the 3ß-HSD genes (not to scale). Exons are represented by boxes numbered I, II, III and IV. Open boxes represent 5' and 3' untranslated sequences, Introns are labelled A, B and C. Common restriction sites, and the translational initiator and stop codons are indicated. Reproduced with kind permission from Martin McBride.

Chromosomal localisation of HSD3B1 and HSD3B2:

Using human-hamster somatic cell hybrids, Lorence *et al* (1990b) localised HSD3B1 to chromosome 1. These experiments also predicted that related genes or pseudogenes were also present on chromosome 1. Berube *et al.* (1989) mapped the genes to chromosome 1p13 and in 1991, Morrison *et al.*(1991) refined the localisation by fluorescent in situ hybridisation (FISH) to chromosome 1p13.1. HSD3B1 and HSD3B2 are contained within a 290kb SacII fragment from human genomic DNA as determined by pulse field gel electrophoresis (Morissette *et al.*, 1995). Morissette and co-workers also reported that the HSD3B loci was closely linked to the Genethon markers D1S514 and the centromeric D1Z5 locus.

Regulation of expression:

Mason *et al.*(1993) examined the regulation of type I 3ß-HSD in the human placenta. 3ß-HSD type I has extremely high levels of expression in the placenta, principally from the multinucleated syncytotrophoblasts. Gibb *et al.*(1985) associated 3ß-IISD activity to the chorion laeve and this was confirmed by Northern and Western analysis, which indicated that it was due to type I 3ß-HSD not type II (Mason *et al.*, 1993). Mason *et al.*(1993) also suggested that the enzyme activities from the above tissues (trophoblasts and chorion leave) were regulated by different mechanisms. In culture, human cytotrophoblast cells initially did not produce 3ß-HSD, although after 48 hours 3ß-HSD production started to increase and continued for up to 6 days. Cyclohexamide,

an inhibitor of protein synthesis, prevented this increase which suggested that the expression of 3ß-HSD required protein synthesis and that it is not a primary response gene. Also, the addition of various cyclic AMP agonists did not alter the already high increases of 3ß-HSD. In contrast, after 24 hours in culture the 3ß-HSD activity of chorion leave cells declined in a time-dependant manner, although similar to 3ß-HSD activity in the cytotrophoblast cells cyclic AMP agonists showed no effect at all. It has been stated that syncytiotrophoblasts differentiate from uninuclear cytotrophoblasts (Kliman *et al.*, 1986), this process is cAMP-promoted and it is possible that this differentiation may be defined by 3ß-HSD, as 3ß-HSD is characteristic of this stage of differentiation. Whereas, the villous cytotrophoblasts initially lacked 3ß-HSD activity, the chorion laeve cells (cytotrophoblasts) already expressed 3ß-HSD. These data suggest that the chorion laeve contains a discrete population of steroidogenically active uninuclear cytotrophoblasts that do not readily differentiate into syncytium, since there was no change in the levels of 3ß-HSD under a variety of culture conditions.

Beaudoin *et al.*(1997a) looked at the regulation of 3ß-HSD type I during syncytium formation by progesterone and oetradiol. 3ß-HSD type I and P450scc reached their maximum steady state mRNA levels and 3ß-IISD specific activity at a time when syncytium formation occurs *in vitro*. By incubating these cells with progesterone or oestradiol, the adundance of the 3ß-HSD mRNA increased but the amount of 3ß-HSD protein stayed the same. This suggests that 3ß-HSD steady state mRNA levels in placenta could be under post-translation regulation. These observations suggest a complex relationship in mechanisms that regulate transcription, mRNA processing and 3ß-HSD type I gene transduction due to the regulation by steroids on 3ß-HSD mRNA levels. This may involve progesterone or oestradiol affecting the level of gene transcription and/or mRNA abundance or stabilisation, rather than affecting the rate of 3ß-HSD protein synthesis.

Beaudoin *et al.*(1997b) also examined the regulation of 3ß-HSD by calcium in human choriocarcinoma cells. In this instance, they found that with 3ß-HSD stimulated by cAMP agonists calcium depleted 3ß-HSD mRNA levels in a dose dependant manner. With PMA-stimulated 3ß-HSD calcium had no effect. In addition, the use of thapsigargin (TG, a cell permeable lactone that releases calcium by inhibiting ER calcium-ATPase) indicated the presence of TG sensitive and TG insensitive calcium ATPases that regulate 3ß-HSD mRNA levels. This data shows the complexity of calcium contribution to the protein kinase A and C pathways in the regulation of 3ß-HSD.

At the level of the gene, Guerin *et al.*(1995) found that the removal of intron I of the type I gene strongly impaired the transcriptional activity directed by the 3ß type I basal promoter, although not cell-specific, as it affected 3ß-HSD activity in placental and adrenal cells similarly. DNAse I and DMS methylation interference footprinting detected a 37kDa nuclear protein bound to a site within intron I, designated the 3ß1-A element. By increasing the protein concentration involved in binding, they reported another 3 nuclear proteins binding to this site, one of which was identified as the transcription factor Sp1. Examination of the DNA sequence revealed the presence of a potential Sp1 binding site that overlapped 3ß1-A. The overlap between these two binding sites may indicate additional levels of control depending on a particular cell concentration of the nuclear proteins. 3ß-HSD type II does not have the same sequence over the 3ß1-A site as type I, although the site of Sp1 binding is identical (see Fig. 1.9). This may indicate that the 37kDa protein does not bind to type II and this may facilitate the binding of Sp1 or other Sp1 family members.

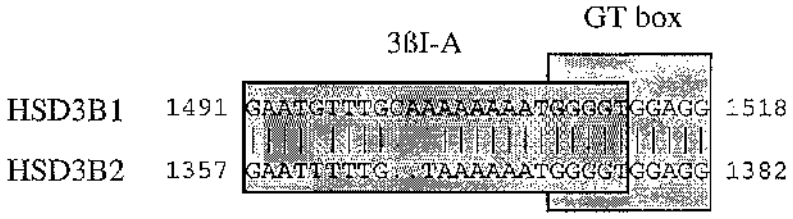


Fig. 1.9 Alignment between 3ß-HSD type I and type II intron I indicating the positions of the 3ßI-A element and the GT box. The coordinates are from the respective genomic sequences and the dots represent gaps in the nucleotide alignment.

The regulation of 3ß-HSD type II in the adrenal cortex is extremely important because of the developmental changes in the activity/expression of 3ß-HSD. In the foetus, the foetal zone of the adrenal produces large amounts of DHEA and very little cortisol. This suggests a block in 3ß-HSD activity or expression. After partuition, cortisol is the major adrenal steroid and this block in 3ß activity/expression must be overcome. At adrenarche, the adrenal cortex starts to secrete large quantities of C19 steroid again and therefore there must be another alteration of regulation of 3ß-HSD in respect to the other enyzmes involved in steroidogenesis (Mason *et al.*, 1997).

In the foetal adrenal, Doody *et al.*(1990) established that 3ß-HSD was not expressed, and that the massive secretion of hydroxysteroids was not because of an inhibition of 3ß-HSD activity. The reason for the lack of expression in the foetal adrenal when the foetal testis express 3ß-HSD is not known at this time, although it has been shown that ACTH can induce 3ß-HSD from foetal adrenal cells in vitro via the protein kinase A pathway (Naville *et al.*, 1991).

Bird *et al.*(1996) examined the differential control of expression of P450c17 and 3ß-HSD in human adrenocortical (H295R) cells. Stimulation of the protein kinase A pathway of signal transduction in H295R cells increased the levels of DHEA and androstenedione produced as well as the levels of 3ß-HSD and P450c17 transcripts. The levels of 3ß-HSD mRNA were increased also by activating the protein kinase C pathway however, the P450c17 mRNA levels were markedly attenuated. This differential response of these two enzymes to the protein kinase C pathway may be critical in the orchestration of adrenal steroidogenesis.

At the gene level of 3ß-HSD type II, Lcers-Sucheta *et al.*(1997) looked at the type II promoter region to identify specific areas that confer basal and phorbol ester induced regulation of transcription in adrenal cortical cells. Phorbol esters (PMA) induces the protein kinase C pathway in adrenocortical cells mimicking the effect of angiotensin II which regulates the production of aldosterone in these cells. To examine the type II promoter, deletion constructs were made with 5' deletions of 1251bps fused to the chloroamphenicol acetyltransferase (CAT) reporter gene. These were transfected into H295R cells and treated with or without PMA. CAT assay data indicated that the region between -101 and 152bp was necessary for the PMA induction. A putative SF-1 (Ad4BP) site was detected within this region (-61 to -56bp) by sequence homology. Steroidogenic factor (SF-1) is an essential transcription factor involved in cAMP transactivation of cytochrome P450 hydroxylases (Morohashi *et al.*, 1995) and it is essential for adrenal and gonadal development and sexual differentiation (Luo *et al.*, 1994). The -101 3ß-HSD-CAT construct was transfected into non-steroidogenic cells with an expression vector for SF-1 and CAT activity increased 49-fold, confirming the functionality and essentiality of this SF-1 site. Unexpectedly, CAT activity increased 540-fold over basal when the transfected cells were treated with PMA, indicating a synergistic activation of type II by PMA and SF-1. Mutation analysis of the SF-1 site abolished SF-1 induced CAT activity and the synergistic activation by PMA. The interaction of SF-1 with the putative element was confirmed by gel shift assays and the presence of SF-1 transcripts in the H295R cells was proved by Northern analysis. These data show that SF-1 is involved in the regulation of type II 3ß-HSD in adrenal cortical cells and this was the first study to show regulation of a non-cytochrome P450 steroidogenic enzyme promoter by SF-1. It is interesting to note that the human type I 3ß-HSD promoter does not contain a SF-1 site. This is consistent with the expression of type I in tissues like the placenta which does not express SF-1 and suggests that other transcriptional factors may be important in 3ß-HSD type I regulation of expression.

1.8.2 Tissue distribution and subcellular localisation of 3ß-HSD

Tissue distribution:

3ß-HSD is found in the classical steroidogenic tissues, namely the placenta, adrenal cortex, ovary and testis (Labrie *et al.*, 1992). However, 3ß-HSD activity has also been identified in several peripheral tissues including the skin (Dumont *et al.*, 1992), adipose tissue (Labrie *et al.*, 1991), breast (Lachance *et al.*, 1990), lung (Martel *et al.*, 1994), prostate (Abalain *et al.*, 1989), liver (Zhao *et al.*, 1991), kidney (Devine *et al.*, 1986) and brain (Martel *et al.*, 1994; Sanne and Krueger, 1995).

Rheaume *et al.*(1991) used the ribonuclease protection assay to show that 3ß-HSD type I was found in the placenta, the skin and the breast whereas type II was identified in the adrenal, the ovaries and the testis. Although from longer exposures type I was also detected in the ovaries and the testis, and a small amount of type II was found in the breast. No type I was seen at all in the adrenal and no type II was found in the placenta. This data has since been confirmed by RT-PCR in our laboratory where 3ß-HSD type II was identified at low concentrations in the placenta (Sandra Burridge, personal communication).

The widespread distribution of 3ß-HSD suggests that this enzyme plays an important role in the peripheral conversion of adrenal precursor sex steroids into active sex steroids. This extragonadal conversion of sex steroids is particularly important in humans and some primates where the adrenal glands secrete large amounts of precursors of sex steroids such as DHEA and DHEA-sulphate.

Subcellular localisation of 3ß-HSD:

3ß-HSD is a NAD^+-dependant membrane bound enzyme. Lorence *et al.* (1990) showed that the cytosolic fraction of COS cell homogenates transfected with type I had no demonstrable activity, while the particulate fraction retained 3ß-activity. This is consistent with 3ß-HSD membrane-bound localisation. 3ß-HSD activity had been located in both the microsomes and the mitochondria. Studies on 3ß-HSD from each of these organelles revealed that it is the same protein involved (Ferre *et al*, 1974; Thomas *et al.*, 1989). The enzymes had similar kinetic profiles with respect to substrate utilisation, product inhibition and cofactor specificity and similar biophysical properties including molecular weight, subunit composition, pH optimum, temperature optimum, stability in storage and solution and the effects of cation inhibitors. Although, Cherradi *et al.*(1993) stated that the Km values for DHEA and DHT as substrates for both enzymes were similar (1.6 and 2.5 respectively) but the Km for DHEA for the mitochondrial enzyme was much lower than the Km for DHT (4.5

compared with 38). In addition, Cherradi *et al.*(1992) reported that the bovine microsomal enzyme preferred 17-hydroxypregnenolone as a substrate. These data suggest that the microsomal and mitochondrial enzymes are not completely identical. The differences could be due to structural differences, different localisation in the organelles (and hence spatial or availability of substrate constraints) or differential regulation of the enzymes.

Alvarez *et al.*(1994) studied the topology of placental 3ß-HSD in the microsomal membrane. Their methods involved immunoblotting and surface specific proteolysis. Inactivation of activity by proteinase treatment indicated that the crucial domains of the protein are located on the cytosolic side of the endoplasmic reticulum membrane. Sauer *et al.*(1994) also examined this by comparing the access to the active sites of the cofactor and the microsomal inhibitor merasyl, a nonpenetrant organic mercurial anion. Microsomal activity required exogenous NAD+ and was inhibited by merasyl, suggesting that the active site faced the cytoplasm. Merasyl had no effect on the mitochondrial 3ß-HSD unless the permeability of the inner membrane was disrupted. This suggests that in the mitochondria 3ß-HSD is located within the mitochondria and that the active sites face the matrix space.

These observations are supported by Cherradi *et al.*(1993) who found that the microsomal 3ß-activity was abolished when exogenous NAD+ was removed, whereas the activity in the mitochondrial fraction was still 40% of the activity measured with excess NAD+. This suggested that the mitochondrial enzyme utilises the intramitochondrial NAD+ available within the organelle and that the catalytic site was accessible from the matrix space. To confirm these observations, bovine adrenal mitochondria were separated into 3 fractions -the outer membrane (OM), the intermembrane space (IMS) and the inner membrane (IM)- and deduced that the 3ß activity was localised to the inner membrane and the intermembrane contact sites. This study was extended, when it was shown that the 3ß-activity was copurified with P450scc, suggesting a tight association between these proteins (Cherradi *et al.*, 1994). They also reveal that the 3ß-activity was higher in the contact sites, than the IM. This work is consistent with the theory of cholesterol transport in the early stages of steroidogenesis which involves the recently identified StAR protein (Stocco and Clark, 1997). This allows cholesterol to be made available to the enzymes involved in steroid synthesis.

More recently, Cherradi *et al.*(1997) examined the effects of calcium on 3ß-HSD as well as the StAR protein and P450scc. The presence of 3ß-activity in the IM was

confirmed by preparing mitoblasts from bovine glomerulosa cells and mouse Leydig cells. A strong 3ß signal was detected in the mitoplasts confirming that 3ß is present within the mitochondria. Immunogold staining of adrenal fasiculata cells was studied as well. From this, is was observed that 3ß antigenic sites were abundant in the mitochondria of these cells. These experiments also confirmed an earlier statement from Cherradi that 60% of 3ß-HSD was present in the microsomes. The apparent density of immunoreactive 3ß is 56% higher than that of the mitochondria and the enrichment was even higher in stacking of smooth endoplasmic reticulum. The intramitochondrial pattern of 3ß localisation was also examined using high power photomicrographs. The P450scc enzyme antigenic sites were localised exclusively to the crista membranes penetrating into the matrix space whereas 3ß-HSD was found in all compartments examined, although 60% was found in the crista membranes. This distribution also differed slightly from the StAR protein which was confined to the intermembrane space. Thus, it is almost certain that 3ß-HSD is localised inside mitochondria facing the mitochondrial matrix and within the contact sites in the intermembrane space. Although, the N-terminus of 3ß-HSD does not contain a mitochondrial targetting sequence, it has been shown in other genes that the targeting information can be contained internally or at the C-terminus (Court *et al.*, 1996; Folsch *et al.*, 1996).

It is not known why there are two localised sites for 3ß-HSD activiy within the cell, one possible reason may be that 3ß-HSD acts in distinct steroidogenic units (Liebermann *et al.*, 1984) with different regulatory controls. This is supported by Cherradi *et al.*(1997) who has shown tight association and similar localisation for the StAR protein, P450scc and 3ß-HSD. It must be stated here that despite the evidence presented above, some researchers believe that 3ß-HSD is a microsomal enzyme and the activity present in mitochondrial fractions is from microsomal contamination (Morel *et al.*, 1997). Although Cherradi *et al.*(1994) estimated that the microsomal contamination of the mitochondrial fraction was less than 5%, and Pozzi *et al.* (1996) demonstrated mitochondrial localisation of 3ß-HSD in the toad.

1.8.3 Structure-function relationships of 3ß-HSD

Two approaches have been undertaken to help understand the structure-function relationship of 3ß-HSD. The first of these is by examining the protein structure by affinity alkylation of type I 3ß-HSD and comparison of the amino acid sequences of all identified mammalian 3ß-HSDs and the second is by the identification and examining the functional consequences of type II mutations found in patients with 3ß-HSD deficiency.

Studies of the 3ß-HSD protein:

Luuthe *et al.*(1991) provided evidence, using differential inhibition, that there were two distinct active sites present in 3ß-HSD - one site for dehydrogenase activity and another for the isomerase activity. The conversion of dihydroepiandrosterone (DIIEA) into androstenedione proceeds via the intermediate 5-androstene-3,17-dione, the ability to obtain tritiated intermediate gave them the opportunity to investigate differential inhibition of the dehydrogenase and isomerase activities using specific inhibitors. They found that N,N-dimethyl-4-methyl-3-oxo-4-aza-5α-andristane-17ß-carboxy-amide (4-MA) inhibited the dehydrogenase site competitively and specifically, whereas trilostane inhibited the isomerase activity non-competitively. This indicated that the activities belong on separate sites of the protein. He also speculated that because 3ß-HSD did not transform 3α-steroids, then the binding for the dehydrogenase site must be near to the position 3 on the steroid and that the site for the isomerase activity binds to the flattened B or A ring, since a compound without a flattened A or B ring did not bind to the isomerase activity.

Studies have been undertaken to discover where in the amino acid sequence the active sites may be. Thomas *et al.*(1993) identified peptides and amino acids associated with substrate binding in type I 3ß-HSD. This group employed 2α-bromo[2'-^{14}C] acetoprogesterone, an active site-directed alkylator of type I 3ß-HSD, to affinity radiolabel the purified enzyme with or without protection by the substrate, pregnenolone. Trypic peptides of unprotected and substrate-protected radioalklyated were isolated and two of these were protected by the substrate. The primary affinity radioalklated peptide of 3ß-HSD is the Arg-250 trypic peptide containing radiolabelled His262 (^{251}GQFYYISDDTPHQSYDNLNYTLSK^{274}). The presence of pregnenolone substantially reduces the radioalkylation of the histidine residue. Histidine is the residue responsible for the hydride ion transfer during catalysis by lactate dehydrogenase and glyceraldehyde-3-phosphate dehydrogenase, and His residues have been identified in the active sites of bacterial 3α,20ß-HSD, foetal lamb 3ß,20α-HSD and human placental 17ß,20α-HSD. The second peptide Lys-175

(^{175}NGGTLYTCALR^{186}) contains radiolabelled Cys183 which is protected by pregnenolone from radioalklyation to the same extent as His262 in the Arg-250 peptide. This equal protection suggests that these amino acids are both located in the substrate binding region and are located near one another in the tertiary structure of the protein. There is significant homology of the peptides in all identified mammalian forms of 3ß-HSD. 88% homology is detected in the amino acids of the Arg250 peptide between isoforms, with the His262 residue present in the same position throughout, and there is 100% homology of the Lys175 peptide (except for a single substitution in human type II) with Cys183 positioned identically in all forms.

Thomas *et al.*(1992) reported that different protein conformations were responsible for the 2 enzyme activities on 3ß-HSD, rather than the dehydrogenase and isomerase activities being catalysed at separate sites on the protein. This disputed what Luu-The had postulated in 1991. Thomas used a secosteroid, that acted as a site-directed alkylator of steroid Δ-isomerase activity, to evaluate the relationship between the two activities. They found that the isomerase site did not exist at low concentrations of the secosteroid but appeared as the concentration of secosteroid was increased, from this information they proposed that the dehydrogenase site underwent a conformation change to expose the isomerase site. They also predicted that NADH, the product from the dehydrogenation reaction induced this conformational change. Therefore, the protein starts that reaction sequence as a dehydrogenase, the products of this reaction (namely NADH) promote a change in protein conformation that triggers the isomerisation. Further studies have shown that the peptides involved in the inactivation by the secosteroid, and therefore the localisation of the isomerase site, are the same as the peptides identified in their study of the substrate binding site using 5α-BAP, i.e. Arg251-Phe274 with tritiated His-262 and Lys175-Arg186 with tritiated Cys-183 (Thomas *et al.*, 1997; Thomas *et al.*, 1996). This information supports the hypothesis that 3ß-HSD catalyses the two-step reaction at a single region mediated by a NADH-induced conformational change around the bound substrate.

Rutherfurd *et al.*(1991a) obtained 86% of the amino acid sequence of the bovine adrenal 3ß-HSD and found approximately 46% homology when comparing their sequence with the steroid binding region of bacterial isomerases. The conserved amino acids included K216, F271, D241, P242, F252 and Y253. These are found around the same region predicted by Thomas *et al.* (1996, 1997) and provide supporting evidence for the position of the active site.

Thomas and collegues have also examined the 3ß-HSD protein to find the location of the cofactor binding site. NAD+ is the cofactor used in the dehydrogenase activity and it has also been shown to be a potent allosteric activator of the steroid isomerase activity. Thomas *et al.*(1991) used the affinity labelling cofactor analogue 5'-[p-(fluorosulphonyl)-benzoyl] adenosine (FSA) to study the nucleotide binding sites of type I 3ß-HSD. It was known that FSA inhibits both activities of 3ß-HSD in rat adrenal and rat testis, therefore Thomas used FSA to examine the cofactor binding in the human placental isoform. Their data suggested that there was a single cofactor binding region which services both enzyme activities (the enzyme exhibited equal affinity for FSA as either an inhibitor of 3ß or as an activator of isomerase activity, pregnenolone protected both activities from FSA and FSA inactivated both activities similarly). Rutherfurd *et al.*(1991a) proposed the localisation of the NAD+ binding site on the basis of primary structure and comparisons with other nucleotide binding proteins. Some common features of NAD+-binding domains include a "glycine-rich" sequence, GXGXXG, which maintains the appropriate structure for nucleotide binding, and the presence of a ADP-binding ßαß-fold containing specific amino acid identifiers (Wierenga *et al.*, 1986). The segment of residues 5-36 contains a glycine-rich sequence and a high score for a ßαß fold predicts that this region may fold into such a unit, therefore Rutherfurd *et al.*(1991a) proposed that this may at least be part of the cofactor binding site in type I 3ß-HSD. The same team also analysed NAD+ binding to type I 3ß-HSD using FSA inactivation (Rutherfurd *et al.*, 1991b). They reported that specific disulphide bonds were formed in 3ß-HSD due to FSA causing the inactivation of 3ß-HSD activity, the disulphide bridge was found to be between the cysteine residues in peptides 274-282 and 18-26 (Cys20 and Cys278). From this information Rutherford proposed that this putative cofactor binding region was in close proximity to a potential membrane anchor in 3ß-HSD (Trp284-Trp309). This study does not reveal an essential role for these residues (these can be substituted without loss of function), however it does reveal that these cysteine molecules are close to one another in the tertiary structure of the protein. To expand on this, Thomas *et al.*(1993) speculated from the above information that the cofactor bound in the amino terminal may be close to substrate bound at the Arg250 peptide during the 3ß reaction. All amino acids predicted to be important in both the substrate and cofactor binding are represented in Fig. 1.10.

In addition to the above, the question of cofactor specificity sheds more light on this. In the rat, mouse and hamster 3ß-HSD families there has been an isoform identified that can interconvert 3-ketosteroids (e.g. dihydrotestosterone) into 3-hydroxysteroids (e.g. androstane-3ß,17ß-diol) only, these enzymes are known as ketosteroid reductases (see section 1.7.4). These enzymes have a cofactor specificity for NADPH/$NADP^+$.

Within the region suspected of cofactor binding there is a residue that is conserved in all 3ß-HSD isoforms except the ketosteroid reductase members, this residue is an aspartic acid (Asp35 in type II). Wierenga stated that one of the amino acids important in the ßaß fold in the co-factor binding sites of NAD+-requiring dehydrogenases was an acidic amino acid and it has been established that aspartic acid interacts better with NAD^+ compared with $NADP^+$ because a coulombic repulsion occurs between this residue and the 2'-phosphate on NADP+ (Nakanishi *et al.*, 1997; Wierenga *et al.*, 1986). The absence of the aspartic acid in the ketosteroid reductase enzymes (these contain an aromatic residue in its place) allows NADPH and not NADH to interact. This supports the evidence above that this amino terminal region is the cofactor binding site in 3ß-HSD.

```
1                                                          50
*                                                           *
MTGWSCLVTG AGGFLGQRII RLLVKEKELK EIRVLDKAFG PELREEFSKL

51                                                         100
*
QNKTKLTVLE GDILDEPFLK RACQDVSVII HTACIIDVFG VTHRESIMNV

101                                                        150
*
NVKGTQLLLE ACVQASVPVF IYTSSIEVAG PNSYKEIIQN GHEEEPLENT

151                                                        200
*
WPAPYPHSKK LAEKAVLAAN GWNLKNGGTL YTCALRPMYI YGEGSRFLSA

201                                                        250
*
SINEALNNNG ILSSVGKFST VNPVYVGNVA WAHILALRAL QDPKKAPSIR

251                                                        300
*
GQFYYISDDT PHQSYDNLNY TLSKEFGLRL DSRWSFPLSL MYWIGFLLEI

301                                                        350
*
VSFLLRPIYT YRPPFNRHIV TLSNSVFTFS YKKAQRDLAY KPLYSWEEAK

351                    373
*                       *
QKTVEWVGSL VDRHKETLKS KTQ
```

Fig. 1.10 Amino acid sequence of 3ß-HSD type I indicating the residues involved in the substrate and cofactor binding sites. Residues underlined represent the cofactor binding site, residues overlined represent the peptides identified in substrate binding studies (dehydrogenase and isomerase sites) with tritiated residues circled. Residues within boxes are homologous with another isomerase enzymes (Thomas *et al.*, 1992, 1993, 1996 and 1997, Rutherfurd *et al.*, 1991a and b).

3ß-HSD deficiency:

3ß-HSD deficiency is characterised by varying degrees of salt wasting, male pseudohermaphroditism and mild virilisation in females. To date, all mutations responsible for the disorder have been identified in 3ß-HSD type II, no deleterious mutations have been detected in type I. This is consistent with type II expression in the adrenal glands and the gonads and type I in the peripheral tissues.

24 mutations have been identified in 3ß-HSD type II (Morel *et al.*, 1997), the nonsense and frameshift mutations introducing a premature stop codon are all associated with the classical salt-wasting form (Fig. 1.11A). The locations of these mutations suggest that more than the first 318 amino acids out of 371 are essential for 3ß-HSD activity. There have been a number of missense mutations identified some of which correspond to the regions described above for substrate and cofactor binding (Fig. 1.11B) The mutations Y253N (Simard *et al.*, 1993), Y254D (Sanchez *et al.*, 1994) and T259R (Tajima *et al.*, 1995) are all found within the Arg250 peptide identified by Thomas *et al.* (1993) and these mutations result in severe salt-wasting form of 3ß deficiency. A further mutation, G15D, results in the severe salt-wasting form and this is found within the "glycine-rich" segment important for cofactor binding (Rheaume *et al.*, 1995). The activity of this mutant enzyme was established in vitro along with another mutant G15A, in which the glycine has been replaced with alanine (the second smallest amino acid after glycine). There was no detectable activity with G15D in intact cells, although relatively high 3ß-HSD activity was found with G15A. However in homogenates, both mutants had a decreased enzymatic activity compared with type II. This may suggest that the G15D mutation alters the intracellular localisation of 3ß-HSD or that *in vivo* its association with intact membranes modifies its efficient conformation.

Other mutations may be involved in the attachment of 3ß-HSD to the membrane. Two segments have been predicted to be transmembrane domains within 3ß-HSD. the first is from residues 283-306, this is conserved throughout the 3ß family and is reasonably hydrophobic. The second is from residues 75-91 and is identical in all members except bovine, rat II and mouse V. The nonsense and frameshift mutations predict a truncated protein without the C-terminal domain, except 318delC and these result in a severe salt-wasting form of 3ß deficiency as expected. In contrast, A82T mutation is located within the second fragment and is associated with a non salt-wasting form (Mendonca *et al.*, 1994).

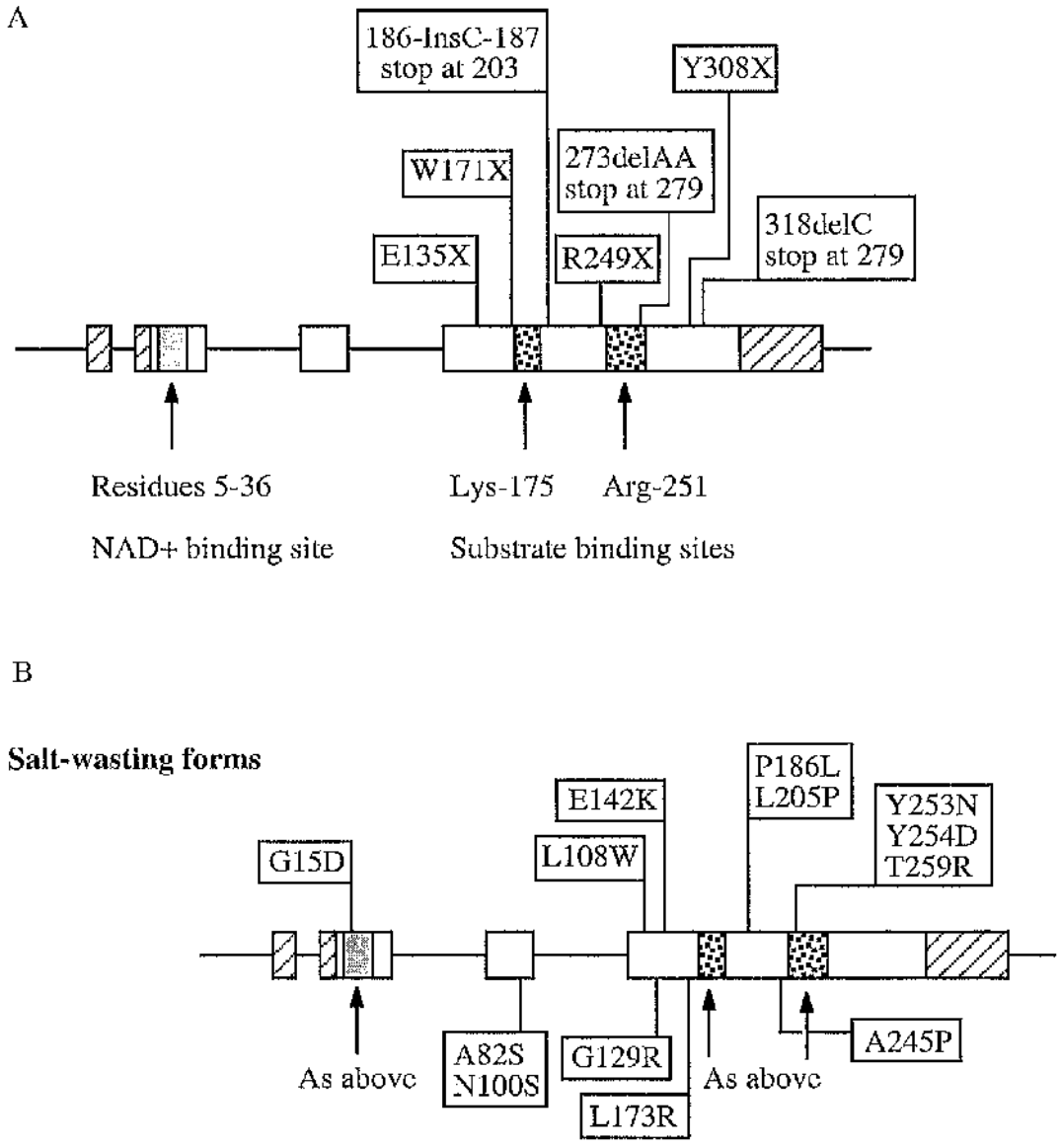


Fig. 1.11 A: Diagram of HSD3B2 showing the location of the nonsense and frameshift mutations. B: Diagram of HSD3B2 showing the location of the missense mutations. Diagonal lines represent untranslated regions, small dots represent the location of the cofactor binding site and large dots represent the putative substrate binding sites. Both diagrams have been modified from Morel *et al* (1997).

1.8.4 Rodent 3ß-HSD gene families

Multiple 3ß-HSD cDNAs have been isolated in rat, mouse and hamster and the genes have been identified by roman numerals according to the order in which they have been isolated. Table 1 summarises some of the characteristics of these rodent 3ß-HSD

isoforms. The rodent isoforms fall into two functionally distinct groups, NAD+-dependant dehydrogenase/isomerases and NADPH-dependant ketosteroid reductases (Abbaszade *et al.*, 1995; Delaunoit *et al.*, 1992). The ketosteroid reductase enzymes include mouse IV and V, rat III and hamster III and catalyse the conversion of ketosteroids into 3ß-hydroxysteroids, i.e. dihydrotestosterone into androstane-3ß,17ß-diol and androstane-3,17-dione (A-dione) into epiandrosterone. It is proposed that these isoforms are involved in the inactivation of active steroid hormones such as the potent androgen DHT. Payne *et al.*(1997) stated that mouse V has a lower Km than mouse IV for DHT and because mouse V is expressed in the male liver with type IV in the kidney, she suggests that mouse V is involved in DHT activation but mouse IV may have a preference for another ketosteroid substrate. This substrate may be dihydroaldosterone, a metabolite of aldosterone in the rat kidney.

The ketosteroid reductase enzymes are dependant on NADPH cofactor and the presence of a hydrophobic residue at residue 36 is predicted to be responsible for this specificity. The NAD^+-dehydrogenase/isomerase isoforms of 3ß-HSD all contain an aspartic acid at residue 36 and it is proposed that this residue interacts more favourably with NAD^+ than NADPH because a coulombic repulsion occurs between the aspartic acid and the 2'-phosphate on NADPH (Wierenga *et al.*, 1986). Thus, the absence of the aspartic residue and its replacement with a hydrophobic residue favours the NADPH specificity.

It is interesting to note that the 3ß-HSD gene family is well-conserved across evolution. Simard *et al.* (1993) stated that from 11 different 3ß-HSD protein sequences (mouse I, II and III, rat I, II, III and IV, human I and II, bovine and macaque) 50% of the amino acids are identical and conservative changes were found at 46 positions which gives an overall similarity of 62.4%. As well as this, the nucleotide sequences of the different mammalian isoforms have homologies ranging from 70% and above, the ketosteroid reductase enzymes are more similar to each other than to the dehydrogenase isomerases.

The mouse gene family have been found to be closely linked within a 3.5cM region of mouse chromosome 3 (Bain *et al.*, 1993). This region is syntenic with the centromeric region of human chromosome 1, where the human 3ß-HSD genes are located (Bain *et al.*, 1993; Morrison *et al.*, 1991). The mouse genes are grouped in a region not larger than 1400kb, determined by analysis of yeast artificial chromosomes containing the mouse 3ß-HSD gene family. However, additional analysis of mouse genomic DNA by pulse field gel electrophoresis indicates that all members of the mouse family are contained within a 400kb fragment (Clarke *et al.*, 1996),

ISOFORM	CLASSIFICATION	TISSUES EXPRESSED	REF
Rat I	A	Adrenal glands, gonads, kidneys	1
Rat II	A	Adrenal glands, fat	2
Rat III	B	Male liver	3
Rat IV	A	Placenta, skin	4
Mouse I	A	Adrenal glands, gonads, neonatal liver	5,6,7
Mouse II	A	Kidney, liver	5,6
Mouse III	A	Liver, kidney	5,6
Mouse IV	B	Kidney	6
Mouse V	B	Male liver	8,7
Mouse VI	A	Placenta, skin, maternal decidua, giant trophoblasts, testis	8*
Hamster I	A	Adrenal glands, male kidney	9
Hamster II	A	Kidney, male liver	9
Hamster III	B	Kidney	9

Table 1.1 Classification of rodent 3ß-HSD gene families.
A=NAD+-dependant dehydrogenase/isomerase, B= NADPH-dependant ketosteroid reductase (1 - Simard *et al.*, 1991, 2 - Zhao *et al.*, 1991, 3 - de Launoit *et al.*, 1992, 4 - Simard *et al.*, 1993, 5 - Bain *et al.*, 1991, 6 - Clarke *et al.*, 1993, 7 - Park *et al.*, 1996, 8 - Abbaszade *et al.*, 1996, 8* - Abbaszade, unpublished observations, 9 - Rogerson *et al.*, 1997)

1.9 Background and aims of project

As described in section 1.8.1, there are two isoforms of 3ß-HSD expressed in humans, known as type I and type II. These enzymes are products of two distinct genes, HSD3B1 and HSD3B2, which have been localised to human chromosome 1p13.1 (Lorence *et al.*, 1990a; 1990b, Morrison *et al.*, 1991). However, several observations were made that suggested more than 2 isoforms of 3ß-HSD existed in the human genome. Southern analysis of human genomic DNA digested with a range of restriction enzymes and hybridised with 3ß-HSD sequences resulted in more restriction fragments than predicted from HSD3B1 and HSD3B2 (Lachance *et al.*, 1990). This could indicate the existance of other functional 3ß-HSD genes, pseudogenes or 3ß-related sequences. Also, during mutation screening experiments in our laboratory novel 3ß-HSD sequences were discovered (Russell *et al.*, 1993). Exon-specific 3ß-HSD PCR amplifications were analysed by denaturing gradient gel electrophoresis (DGGE) and it was noted that some products migrated differently from type I and type II. These products were cloned, sequenced and novel exonic sequences were detected. This evidence led our laboratory to screen two λgem11 human genomic libraries for sequences similar to the known type I and II 3ß-genes (McBride *et al.*, 1995, McBride, 1996).

One library was created from DNA of a member of a family affected with a dominant form of hirsutism (McBride *et al.*, 1995), the other library contained DNA from an unrelated normal individual. These were screened with ^{32}P-labelled 3ß-HSD type I cDNA. From a total of $1.4x10^6$ plaques, 57 positive plaques hybridised reproducibly through to the secondary stage of screening. The DNA from these clones was used as template for PCR amplification with primers designed to amplify exon 4b from 3ß type I. The products were analysed by DGGE and classified according to their mobility which was different from type I exon 4b. In particular, two phage clones were analysed from this screen, clones 19(4) and 24(4), and it was demonstrated that these contained 3ß-HSD sequences similar but not identical to type I or type II (McBride, 1996). For some clones no amplification product was obtained for exon 4b so exon 3 primers were designed and amplification of exon 3 was attempted. Again, the products obtained were analysed by DGGE and several interesting phage clones were identified from these experiments, including clones designated clone 2(7) and 8(3).

The aim of this project was to extend the studies above and investigate the extent of the 3ß-HSD gene family. The primary aims were threefold: Firstly, to characterise two of the phage clones identified from the library screens and demonstrate whether they contain putative new members of the 3ß-HSD gene family (Chapter 3). Secondly, to

establish a genomic contig across the 3ß-HSD gene family using yeast artificial chromsomes (YACs) and bacterial artificial chromosomes (BACs) genomic clones and create a detailed restriction map of the gene family (Chapter 4) and thirdly, to develop hybrid selection experiments to attempt to identify further expressed sequences at the 3ß-HSD locus (Chapter 5).

CHAPTER 2

Materials and Methods

2.1 BACTERIAL STRAINS

Strain	Genotype	Source/Ref.
Nova-Blue	*endA* 1, *hsdR* 17, (r^-_k, m^-_k), *supE* 44, *thi* -1 *recA* 1, *gyrA* 96, *relA* 1, lac[F^- *proAB*, $lacI^qZ\Delta M15$ Tn10 (*tet* ^{r})]	Novagen
E1647	ER1648, *rec* D1014	Biolabs
DS941	*rec* F143, *pro* A7, *str* 31, *thr* 1, *leu* 6, *tsx* 88 *mlt* 12, *his* 4, *arg* E3, *lac* 1q, *lac* ZDM15, *gal* k2 *aro* 14, *supE* 44, *xyl* 5	D.Sherratt
DH5a	F^-, y80dlacZDM15, D(lacZYA-*arg*F), U169, *deo*R, *rec*1A, *end*A1, *hsd*R17(r_K^-, m_K^+), *pho*A, *sup*E44, l^-, *hi*-1, *gyr*A96, *rel*A1.	Gibco BRL

2.2 BACTERIAL MEDIA

L broth	10g bacto-tryptone, 5g bacto-yeast extract, 10g NaCl, 950ml H_2O
L-agar	10g bacto-tryptone, 5g bacto-yeast extract, 10g NaCl, 15g bacto-agar, 950ml H_2O
2xYT	1.6% (w/v) bacto-tryptone, 1% (w/v) bacto-yeast extract, 0.5% (w/v) NaCl pH7.5
NZCYM	10g NZamine, 5g NaCl, 5g Bacto-yeast extract, 1g casamino acids, 2g $MgSO_4$, 1.5% Bacto-agar up to 1 litre
BBL agar	10g trypticase peptone (BBL), 5g sodium citrate, 1.5 % Bacto-agar up to 1 litre
BBL Top agar	10g trypticase peptone, 5g NaCl, 6g Bacto-agar up to 1 litre
BBL Top agarose	As above but with 0.7% agarose

2.3 MANIPULATION OF BACTERIOPHAGE

2.3.1 Preparation of Plating Bacteria (ER1647)

20mls L-broth medium containing 15µg/ml tetracyclin (Sigma) was innoculated with a single colony of ER1647 cells and incubated overnight at 37°C. 2mls of this overnight culture was used to innoculate 100mls L-broth containing tetracyclin as above, and grown at 37°C for approximately 3 hours until the $O.D._{600} = 0.5$. The culture was then centrifuged at 4,000g for 5 minutes, the supernatant was removed and the cells were resuspended in 10mM $MgSO_4$ until the $O.D._{600} = 2.0$.

2.3.2 Plating Bacteriophage

NZCYM plates containing 15µg/ml tetracyclin were poured and dried at 42°C for approximately 1 hour. Tenfold solutions of high titre phage stock (see below) were made, 100µl of each dilution was mixed with 100µl ER1647 cells and incubated at 37°C for 20 minutes. Meanwhile, the top agarose was melted and after 20 minutes 3mls of top agarose was added to each phage/cell mix and quickly poured over NZCYM plates. The plates were incubated at 37°C overnight.

2.3.3 Plaque Lifts

Before starting, the phage plates were placed in the fridge for 30 minutes to harden the top agarose and nylon transfer membrane (Micron Separations Inc.) was cut to the size of the plates. Once the plates had been chilled the membrane was carefully placed on top of the plates avoiding air bubbles. The filters were labelled and orientated by stabbing holes in the membrane through to the agar, these were left on the plates for at least 1 minute before transferring to denaturing solution (1.5M NaCl, 0.5M NaOH, pH 7.4) for 30 seconds, neutralising solution (0.5M Tris, 1.5M NaCl) for 2 minutes and 20xSCC (3M NaCl, 0.3M Tri-sodium citrate, pH 7.0) for a minimum of 10 minutes. The membranes were then blotted onto 3mm paper and the transferred DNA crosslinked to the membrane by exposure to UV radiation or baked at 80°C for 2 hours.

2.3.4 Making high titre phage stocks

Phage were plated as above to produce confluent lysis of the plating bacteria. 1-2mls of phage buffer was added and left for approximately 15 minutes to allow the bacteriophage to diffuse out of the agar. The buffer, containing the phage, was removed from the plate using a pipette and stored at 4°C.

2.3.5 Phage lysis

Dilutions (1/10µl, 1/5µl and 1µl) of high titre phage stock were added to 1ml ER1647 plating cells and incubated at 37°C for 20 minutes. 100mls NZCYM, preheated to 37°C, was added and placed on a shaker at 37°C for 5-6 hours until cell lysis was observed. The media was turbid after 2-3 hours incubation and cleared upon cell lysis. The lysate was centrifuged at 4,000g for 10 minutes to remove the cell debris, the supernatant was transferred to a sterile tube and stored at 4°C.

2.3.6 Purification of Lambda DNA

Wizard Lambda Preps Purification System (Promega). Used as manufacturers recommendations.

2.3.7 Plating phage onto BBL plates

BBL plates contain agar not agarose and so phage can be plated out and used directly in the polymerase chain reaction without interference. BBL plates were poured and dried as NZCYM plates, 100µl top BBL was mixed with 100µl ER1647 plating cells and poured over the plates. Once set, 10µl spots of phage were dropped onto the plate and the plates were facing upwards left overnight at 42°C. Plugs were removed the next day and used directly as template in PCR reactions.

2.4 PURIFICATION OF PLASMID DNA

2.4.1 Isolation of plasmid DNA

Wizard minipreps DNA purification systems (Promega) or Qiagen Plasmid Midi-kit (Qiagen). Both kits used as manufacturers recommendations.

2.4.2 Small-scale isolation of plasmid DNA by phenol/chloroform extraction

1.5-3mls of an overnight culture of E.Coli transformed with the plasmid DNA was centrifuged at maximum speed. The pellet was resuspended in 200µl P1 solution (50mM Tris-HCl (pH 8.0), 10mM EDTA), then 300µl P2 solution (200mM NaCl, 1% SDS) was added gently, followed by 300µl chilled P3 solution (3.0M potassium acetate, pH 5.5), the solution was mixed gently but throughly. It was centrifuged at maximum speed for 10 minutes and the supernatant was transferred into a fresh

eppendorf. 2µl RNAseA (20mg/ml) was added to the supernantent and it was incubated at 37°C for 20 minutes. An equal volume of phenol/chloroform (1:1 ratio) was added, mixed well ,then centrifuged at maximum speed for 1 minute. The aqueous layer was removed and an equal volume of chloroform alone was added to it, it was mixed and centrifuged as before for 1 minute. The aqueous layer was transferred into a fresh eppendorf, 2xvolume of 100% ethanol and 1/10 volume of sodium acetate (3M, pH5.4) was added and the mixture was placed at -20°C for 30 minutes. The DNA was centrifuged at maximum speed for 15 minutes, the pellet was washed with 70% ethanol, dried at 42°C for 10 minutes and resuspended in 20µl sterile water. This prep. gives a yield of 5-15µg DNA depending on the volume of culture used, the copy number of the plasmid and the bacterial strain used.

2.4.3 Large-scale isolation of plasmid DNA by phenol/chloroform extraction

25-100mls overnight culture of E.Coli was centrifuged at 4,000g for 15 minutes. The supernatant was removed and the cells resuspended in 4mls P1 (solutions as above). 4mls P2 was gently added and the mixture left at room temperature for 5 minutes, 4mls of chilled P3 was then added and the mixture placed on ice for 15 minutes. It was then centrifuged at 4°C for 30 minutes at >30,000g. The supernatant was removed and an equal volume of phenol/chloroform was added (ratio1:1), it was mixed well and centrifuged at >20,000g for 10 minutes. The aqueous layer was removed and ethanol precipitated with 2xvolume ethanol and sodium acetate as above. The pellet was washed with 70% ethanol, dried and resuspended in 500-1000µl sterile water, depending on the size of the pellet. The DNA was further treated with RNaseA (2µl 20mg/ml for 20 minutes at 37°C, as above), phenol/chloroform extracted and ethanol precipitated, finally resuspending the pellet in 250µl sterile water. This prep. gives a yield of 50-200µg DNA again dependant on factors above.

2.5 MANIPULATION OF PLASMID DNA

2.5.1 Restriction digests

1µg DNA, 2µl 10x REACT buffer (Gibco BRL), 5-10 units restriction enzyme (Gibco BRL) were mixed carefully and dH2O was added to a final volume of 20µl. The reaction mixes were placed at 37°C for 1-2 hours.

2.5.2 Agarose gel elecrophoresis

Unless otherwise indicated, all agarose gels were run in TBE buffer (89mM Tris, 89mM Boric acid, 2mM EDTA). Ethidium bromide was added to each gel at a concentration of 0.5μg/ml. The sample loading buffer consisted of 0.25% bromophenol blue, 40% (w/v) sucrose in water (6xbuffer). The size marker included was the 1kb ladder from Gibco-BRL unless stated.

2.5.3 Purification of DNA from agarose gel slices

DNA fragments were recovered from agarose gel slices using Spin-X centrifuge filter units (Costar). Briefly, a gel slice containing the fragment of DNA was placed in a costar column, placed at -70°C for 10 minutes, 37°C for 5 minutes then centrifuged at maximum speed for 10 minutes. The eluant containing the purified DNA.

2.5.4 pT7Blue T-vector cloning of PCR products

Taq DNA polymerase leaves single 3' A-nucleotide overhangs on their reaction products, these products can be ligated directly into a vector containing compatible single T-nucleotide overhangs. pT7Blue vector (Novagen) has been constructed for this purpose. For a standard reaction the ligation mix contained 2.0μl x5 ligase buffer (200mM Tris-HCl pH7.6, 50mM Mg Cl_2), 0.5μl 10mM ATP, 1.0μl T-vector (50ng/μl), 2-3 Weiss units T4 DNA ligase, 0.2pmol amplififed product and sterile distilled water to a final volume of 10μl. This was mixed gently and incubated at 16°C overnight.

2.5.5 Transformation of plasmid DNA into Novablue competent cells

20μl of Novablue competent cells (Novagen) was aliquoted into pre-chilled eppendorfs, 1μl of above ligation reaction was added and the tubes were placed on ice for 30 minutes. After 30 minutes, the tubes were heated to 42°C for 40 seconds, then placed on ice for 2 minutes. 80μl of room temperature SOC medium was added to each tube and they were incubated at 37°C for 1 hour. 50μl of each transformation was then spread onto LB agar plates containing 50μg/ml ampicillin and 15μg/ml tetracyclin. For blue/white selection of colonies, 35μl of 50mg/ml X-gal (Gibco BRL) in dimethyl formamide and 20μl 100mM IPTG (Gibco BRL) were pre-spread on the plates and allowed to soak in. The plates were inverted and incubated at 37°C overnight.

2.5.6 Transformation of plasmid DNA into DH5α competent cells

The other strain of competent cells used to transform plasmid DNA were DH5α competent cells (ams Biotechnology). 4µl of the above ligation reactions were added to 20µl of competent cells and incubated on ice for 30 minutes. The transformation mixture was heat shocked at 42°C for 1 minute, 1ml of LB medium was added and the cells were grown at 37°C for 1 hour. The cells were centrifuged at maximum speed for 3 minutes and the pellet was resuspended in 100µl LB medium. 50µl of this was spread on LB agar plates containing ampicillin (50mg/ml) only. Blue/white selection was accomplished as above. The plates were inverted and incubated overnight at 37°C.

2.5.7 Ligation of lambda phage DNA digests with pUC cloning vector

1µg lambda phage DNA was digested with a restriction endonuclease (as above), at the same time 1µg of pUC 18 (Genbank accession number A02710) was digested with the same enzyme. Both digests were purified by a microcon 30 filter (Amicon) to remove the restriction enzyme and allow buffer exchange to the ligation buffer. A molar ratio of 2:1 insert:vector was combined in a reaction tube containing 1µl 10xligation buffer (300mM Tris-HCl (pH 7.6), 100mM MgCl2, 100mM DTT, 10mM ATP) , 1µl 10mM ATP, 1-2 units of T4 DNA ligase (Promega) with a final volume of 10µl made up with sterile water. The reaction was incubated at 16°C overnight. The recombinant clones were transformed into competent cells as above.

2.5.8 Selection of recombinant clones (single colony gel)

Single colonies were picked into 3-5mls LB medium containing the appropriate antibiotic selection and grown overnight at 37°C. 1.5mls of each culture was transferred into an eppendorf and centrifuged at maximum speed for 30 seconds, the supernatant was carefully removed. The pellets were resuspended in 100µl single colony loading buffer (2.5% ficoll, 1.25% SDS, bromophenol blue, made up to final volume with TBE) and the cells were left to lyse for at least 20 minutes. The reactions were then centrifuged at maximum speed for 30 minutes and ~35µl of each supernatant was carefully loaded onto a 1% agarose gel. The gel was ran as described above and then photographed. Supercoiled DNA was visible at the top of the gel and it could be easily determined which clones contained inserts.

2.6 POLYMERASE CHAIN REACTION (PCR)

Unless indicated, PCRs were carried out in a Hybaid "Omnigene" for 30-35 cycles of:

(i) 91°C for 30 seconds

(ii) annealing temperature for 30 seconds

(iii) 72°C for 30 seconds

The reaction mixture contained amplification buffer (50mM KCl, 10mM Tris-HCl (pH 9.0), 0.1% Triton X-100, 15mM $MgCl_2$), 0.2mM of each dNTP, 1μM of each oligonucleotide primer, 0.5-1 unit Taq DNA polymerase (Gibco-BRL) and 10-50ng of DNA template, final volume 50μl made up with sterile distilled water.

2.6.1 Synthesis of oligonucleotides

Oligonucleotides were either synthesised using an Applied Biosystems 392 oligonucleotide synthesiser into ammonium hydroxide or synthesised by Gibco BRL custom oligos.

2.6.2 Primer Sequences

HSD3B1 primers

Name (Exon)	5' coord.	Sequence	3'coord.
A (Ex.1)	1354	AGT GCA TAA AGC TTC AG	1370
B (Ex.2)	1545	TAA CCA TTT GAC ATC TC	1561
C (Ex.2)	1585	CCT AGA ATC AGA TCT GC	1601
*D (Ex.2)	1787	TCT CTC AAT TCT GGT CCG AAG GCC	1764
E (Ex.2)	1855	CCA CAT ACA TGC AGT GT	1839
*F (Ex.3)	5655	GGT TAC TGG ACT GGA CA	5671
G (Ex.3)	5682	TTG AGG TCT TGT TCT GG	5698
H (Ex.3)	5881	CCA CCT TGC TGC ATC TC	5865
I (Ex. 3)	6268	GTG ACT ACA GTG GCA CC	6252
J (Ex.4)	7955	CGT GGT TGG CAC CTC TT	7971
*K (Ex.4)	8031	GAG GCC TGT GTC CA/GG C	8047
L (Ex.4)	8159	GGC CAT GTG TTT TCC AG	8142
*M (Ex.4)	8217	ACC CGT TAG CCG CCA G	8202
N (Ex.4)	8239	CCC TGT ACA CTT GTG CC	8255
O (Ex.4)	8400	CAC ATT CTG GCC TTG AG	8416
P (Ex.4)	8466	AGT AGA ACT GTC CTC GG	8450
Q (Ex.4)	8511	ATT AAG GTT ATC ATA GC	8494

R (Ex.4)	8881	GGA GCT TGA TGA CAT CT	8865
*S (Ex.2)	1671	TCC TGC TCC TGT CAC AAG G	1689

* indicates type II primer also.

HSD3B2 primers

Name	5'coord.	Sequence	3'coord.
T (Ex.1)	1219	AGA GCA TAA AGC TCC AG	1235
U (Ex.2)	1522	ATG GGC TGG AGC TGC CTT GTG AC	1544
V (Ex.2)	1720	CCA CAC CCA TGC AGA GT	1704
W (Ex.3)	5515	AGC TCC AGA ACA GGA CC	5531
X (Ex.3)	5695	CCC CAG GCT ACT GTA CC	5680
Y (Ex.3)	5714	CCA ACT TGT TTT ATC TC	5698
Z (Ex.4)	7849	TGT GGT TGC AGC TCC TT	7865
AA (Ex.4)	8741	GCA CAT CTC TGT CAT CC	8725
BB (Ex.4)	8774	GAG TTT GAT GAC ATT T	8758

Gene-specific primers (HSD3B1 co-ordinates)

Name	5'coord.	Sequence	3'coord.
HSD3B1	1388	AGG TGA GAA GTA CGT CC	1404
HSD3B2	1388	AGG CAG TAA GGA CTT GG	1404
y4 5'	8600	AAT AGT GAG CTT CCT GCC G	8618
y4 3'	8846	CAC ATC TCT GTC ATC CTT CG	8827
y1 5'	5662	CCT GAC CTG TGT TCT TG	5678
y1 3'	5788	AGA GGC GGT GTG GAT GAC T	5770
y2 5'	1391	TGA CCA GCA CCC TGA CTA C	1409
y2 3'	1836	AAG CTG GAG CCA CAC AAC T	1818
y5 5'	1615	CTG TTT CCT GGA GTG TCT TGT	1635
y5 3'	1796	GAA AAT TCC TTC CTC AAT TCC	1776
y3 5'	8221	CAC TGA AAA ATA GCG TAT	8238
y3 3'	8344	CTT GAC AGG ATC CCA TTG TAA	8322

Conserved oligonucleotides (HSD3B1 coordinates)

Name	5'coord.	Sequence	3'coord.
Exon 2 5'	1585	CCT AGA ATC AGA TCT GC	1601
Exon 2 3'	1671	TCC TGC TCC TGT CAC AAG G	1689
Exon 4 5'	8031	GAG GCC TGT GTC CA/GG C	8047
Exon 4 3'	8741	GCA CAT CTC TGT CAT CC	8725

Cloning vector primers

Name	Sequence
M13-40	GTT TTC CCA GTC ACG ACG TTG TA
M13-50	TTG TGA GCG GAT AAC AAT TTC
T7	TAA TAC GAC TCA CTA TAG GG
U19	GTT TTC CAG TCA CGA CGT T

Somatic cell hybrid primers (HSD3B1 coordinates)

Name	5'coord.	Sequence	3'coord.
SCH 5'	8031	GAG GCC TGT GTC CA/GG C	8047
SCH 3'	8890	CAG GAG GGT GGC GCT TGA TGA CAT	8867

YAC primers

Name	Sequence
93096	TTC CTC ATG TCA TCA AAA CCT G
93097	TCA GAA TCA GCT ACT GAG ATT TGG

GAPDH primers

Name	5'coord.	Sequence	3'coord.
5'	78	CGG AGT CAA CGG ATT TGG	95
3'	1051	CAT GTG GGC CAT GAG GT	1035

2.6.3 Primer precipitation from ammonium hydroxide stocks

100µl oligo stock were mixed with 300µl 100%ethanol and 10µl sodium acetate (3M, pH5.2), placed at -70°C for 15 minutes and then centrifuged at 4°C for 30 minutes at maximum speed. The supernatant was removed, 100µl 70% ethanol was added and the mixture was centrifuged at 4°C for 15 minutes at maximum speed. The 70% ethanol was removed and the pellet was dried at 42°C for 10 minutes, then resuspended in 100µl sterile water.

2.7 SOUTHERN BLOTTING

2.7.1 Bidirectional southern blotting

DNA to be blotted was electrophoresed on 1xTBE agarose gels as above. The gel was photographed beside a size marker (usually a ruler) to allow sizing of bands later on. The gels were first depurinated in dilute HCl acid (0.25M) for 20 minutes to increase the efficiency of transfer of big fragments, then washed in denaturing solution (1.5M NaCl, 0.5M NaOH), 2x15 minutes, and followed by neutralising solution (0.5M Tris, 1.5M NaCl, pH 7.4), 2x20 minutes. A sandwich arrangement was set up on the bench as follows:

weight
glass plate
2 inch pile of flat paper towels
3 sheets of 3mm paper
nylon transfer membrane
gel
nylon transfer membrane
3 sheets 3mm paper
2 inch pile of paper towels

This was left for 6 hours to overnight. The DNA was crossed linked to the filters by exposure to UV light. The filters were stored at room temperature until required.

2.7.2 Unidirectional southern blotting

This was carried out as above except the blots were set up as follows:

weight
glass plate
2 inch pile paper towels
3 sheets 3mm paper
nylon transfer membrane
gel
3mm paper whicker into 20xSSC

2.8 PREPARATION AND HYBRIDISATION OF RADIOLABELLED PROBES

2.8.1 Random priming of DNA

20ng of DNA was labelled with a-^{32}P-dCTP (3000Ci/mmol; Amersham) by the random primer method, using a Ready-To-Go DNA labelling kit (Pharmacia). The reaction mixture consisted of 20ng DNA in 46µl water, 4µl a-^{32}P-dCTP (10µCi/µl), reaction mix (dATP, dCTP, dGTP, dTTP, FPLC *pure*^{R} Klenow fragment 4-8units) and random oligodeoxyribonucleotides (primarily 9-mers), the final volume was 50µl. The DNA was heated to 95°C for 3 minutes before being added to the above mix, once added the reaction mix was incubated at 37°C for 45 minutes.

2.8.2 Hybridisation of random-prime probes

For all hybridisations hybaid bottles and hybridisation oven were used. The membranes to be probed were placed in ~40 mls pre-hybridisation solution (5xSSC, 5xDeinhardts, 0.2% SDS, 10µg/ml herring sperm) at 65°C for at least 30 minutes. The labelled DNA (above) was heated to 95°C to denature and added to the membranes and pre-hybridisation solution, this was left overnight at 65°C. The next day the membranes were transferred from the solution containing the probe into wash buffer (1xSSC, 0.1% SDS) and placed at 65°C for 30 minutes, this was repeated twice. The membranes were then placed in bags, sealed and put in an autoradiograph cassette with X-ray film (Fuji Medical X-ray film). The film was developed after 1-2 days.

2.8.3 Preparation of oligo probes

10pmoles of oligo were placed in a reaction mix containing 2.0µl 10xkinase buffer (700mM Tris-HCl (pH 7.6), 100mM MgCl2, 50mM DTT), 3.0µl g-^{32}P-dATP (10µCi/µl, DuPont), 1-2units of polynucleotide kinase (Promega) with a final volume of 20µl made up with sterile water. The reaction was then incubated at 37°C for 45 minutes.

2.8.4 Hybridisation of oligo probes

As with random-prime probes (above) a Hybaid hybridisation oven and bottles were used. The membrane to be probed was placed in ~40mls pre-hybridisation buffer (6xSSC, 0.01M sodium phosphate, 1mM EDTA, 0.5% SDS, 100µg/ml herring sperm (denatured), 0.1% non-fat dried milk) at 42°C for at least 45 minutes. The labelled

oligo (above) was added to the solution and it was incubated at 42°C overnight. 42°C was the hybridisation temperature used for oligos between 20 and 25 nucleotides long. The next day, the membranes were rinsed with wash buffer (6xSSC, 0.1%SDS) for 10 minutes at room temperature followed by 2 washes at 42°C. The membranes were sealed into bags and exposed to X-ray film overnight.

2.8.5 Gene-specific oligo probe sequences

Name	5'coord.	Sequence	3'coord.
HSD3B1	1390	GTG AGA AGT ACG TCC ACT CT	1409
HSD3B2	1387	GAG GCA GTA AGG ACT TGG ACT	1407
HSD3By1	1423	TAA CCA TTT AAT TAA TGG TGA A	1439
HSD3By2	1540	GTG TAT AAA TAT CTG ATC TC	1559
HSD3By3	8315	GGC CCC ATT ACA ATG GGA TC	8336
HSD3By5	1623	TGG AGT GTG TTG TCT GCT ACT	1643

2.9 DIDIEOXY SEQUENCING

2.9.1 Sequencing plasmid DNA

Double-stranded plasmid DNA sequencing was performed with the Sequenase Kit, USB Version 2 (Amersham) or Sequenase Quick-Denature Plasmid Kit (Amersham). Both kits were used as manufacturers recommendations.

2.9.2 Sequencing PCR products

Direct PCR sequencing was carried out using Sequenase PCR sequencing Kit (USB, Amersham). Used as manufacturers recommendations.

2.9.3 Automatic sequencing

Automatic sequencing was performed using ABI automatic sequencer. Protocols used as manufacturers recommendations.

2.10 SYNTHESIS OF FIRST-STRAND CDNA FOR RT-PCR

Total prostate, placental, testicular and ovarian RNA was bought from Clontech and used as template in the reaction (reprecipitated as manufacturers recommendations).

The first-strand cDNA reaction was accomplished using the first-strand cDNA synthesis kit from Clontech. Oligo dT was used as the primer.

2.11 YAC PROTOCOLS

2.11.1 Making YAC plugs

10ml culture (containing the appropriate media and supplements) was innoculated and incubated at 30°C overnight. The yeast cells were harvested at 1500rpm for 5 minutes at room temperature. The pellet was resuspended in 0.5ml YRB (1.2M sorbitol, 10mM Tris, 20mM EDTA, pH 7.5) containing 14mM ß-mercaptoethanol and 50 units of lyticase. 95% spheroblasting was observed under the microscope after 80 minutes at 37°C, then 0.5mls of 1% low melting point agarose in YRB was added. The agarose solution was mixed, poured into taped plug former and placed on ice until set. Once set the plugs were removed into 20mls YLB (1% lithium dodecyl sulphate, 100mM EDTA, 10mM Tris-HCl; filter sterilised), left at RT for 1 hour and then incubated overnight at 50°C. Thereafter, the plugs were placed in 20mls fresh YLB and stored at room temperature for up to one year.

2.11.2 Pre-treatment of YAC plugs for sizing

A whole plug from each YAC to be sized was put into an eppendorf containing TE (10mM Tris-HCl, pH7.5, 1mM EDTA). This was left for 2 hours to overnight at 4°C changing the TE at least twice. Approximately one hour before the YAC plugs were loaded onto the gel, the plugs were pre-equilibriated in 0.5xTAE.

2.11.3 Digestions of YAC plugs

For partial digestions running uncut and four time-points for each enzyme, three YAC plugs were cut in half, incubated in several changes of TE and left overnight at 4°C. The next day the plugs were incubated in 500µl 1xrestriction buffer at 4°C for 1 hour. Fresh buffer was then diluted to give "complete" digestion buffer, 200µl was used for each plug and it was diluted in the following proportions: 100µl 10xbuffer, 10µl 1% gelatin, 2.5µl 1M spermidine, 887.5µl sterile water. 100ml of the above buffer was added to each plug for 20 minutes at 4°C, the buffer was removed and 100µl fresh buffer containing the restriction enzyme (see below) was added. At each time-point (5, 10, 20, 40 and 60 minutes) the reactions were stopped by adding 100µl ice-cold EDTA. This was left for 10 minutes then the plugs were equilibriated in 0.5xTAE with three

changes over 30 minutes. The plugs were then ready to be loaded onto a pulse-field gel.

Amount of enzyme added per plug for complete digestion in one hour:

4.5µl SacII
3.75µl XhoI
7.5µl NotI
2.5µl SmaI

2.11.4 Pulsed field gel electrophoresis (PFGE) for YACS

Pulsed field gels were ran on a CHEF-DR II PFGE system (Biorad). 1.25% Seakem GTG (Flowgen) gels in 0.5xTAE were used for sizing YACS and resolving YAC digests, no ethidium bromide was added to the gel. YAC plugs were pre-treated as above and then cut in half with a single-edged razor-blade. The gel slots were flooded with 0.5xTAE and a half plug was slid into a slot using a microspatula for guidance. Lambda and yeast chromosome size markers (Gibco BRL) were loaded at either side of the gel. The slots were sealed with 1.25% Seakem GTG and once the agarose was set it was placed in the gel support in the CHEF. The running conditions used for the YACs (average insert size of 300kb) were 20-80 seconds for 20 hours at 200V at 14°C. The gels were suitable for southern blotting and hybridisation of radiolabelled probes as described above.

2.12 BAC PROTOCOLS

2.12.1 Isolation of BAC DNA

The isolation of high quantites and high quality BAC DNA was achieved using a modified Qiagen protocol. A 5ml culture containing 12.5mg/ml chloramphenicol was innoculated with a single colony and grown at 37°C overnight, 0.5mls of this was used to innoculate a 100ml culture which was incubated at 37°C overnight also. The next day the culture was split into 2x50mls and centrifuged at 5K for 15 minutes. The Qiagen midi plasmid kit protocol was used with the following additions:

10mls of P1, P2 and P3 were added to ensure complete lysis
The cleared lysates were pooled onto a single column
The DNA was eluted from the column by adding 5 aliquots of 1ml elution buffer pre-heated to 65°C.

The DNA was resuspended in 200µl sterile water, a typical yield from this prep. was 20-30µg DNA. Once the DNA was resuspended, to prevent shearing of the BAC DNA, it was only ever pipetted after the ends of the pipette tips had been cut off.

2.12.2 Digesting BAC DNA

With respect to restriction digesting, BAC DNA was treated as any other plasmid DNA (see above). One difference, however, was that the digests were electrophoresed on a 0.7% 1xTBE agarose gel, rather than a 1% gel. This provided better resolution for the bigger fragments obtained.

2.12.3 Pulsed field gel electrophoresis (PFGE) for BAC DNA

BAC DNA was resolved on pulsed field gel electrophoresis much the same as the YAC DNA (see above). The main differences were that BAC digests were loaded directly into the slots (no need to make plugs), a 1% agarose gel for PFGE running (Sigma) was used and the running conditions varied. The running conditions for resolving BAC DNA were:

1-50kb fragments - 0.1-2.5 switch, 6volts/cm, 8 hours
5-100kb fragments - 0.2-6.0 switch, 6volts/cm, 12 hours
50-400kb fragments - 30-30 switch, 6volts/cm, 16 hours

These gels were suitable for southern blotting and hybridisation of radiolabelled probes as described above.

2.13 HYBRID SELECTION PROTOCOLS

2.13.1 Isolating total RNA from placenta

Total RNA was isolated from placental tissue using TRI REAGENT (Sigma Biosciences). 250µg of total placental RNA was obtained from 300mg tissue. The kit was used as the manufactuers recommendations.

2.13.2 Isolating mRNA from total RNA

mRNA (poly A+) RNA was isolated from total RNA using the FastTrack 2.0 mRNA Isolation Kit from Invitrogen. From 400µg total RNA a yield of 8µg mRNA was obtained. The kit was used as the manufacturers recommendations.

2.13.3 Synthesis of double-stranded blunt-ended cDNA

Double-stranded blunt-ended cDNA ready for ligating to linkers was synthesised using The Copy Kiy cDNA Synthesis System from Invitrogen. 2.8µg of double-stranded cDNA was produced from 5.5µg mRNA using a 50:50 mix of random primers and oligo dT primer. The kit was used as the manufacturers recommendations.

2.13.4 Ligation of oligonucleotide linkers to double-stranded cDNA

5' phosphorylated oligos were synthesised by Gibco-BRL, the sequence of the oligos were:
Oligo 1 - 5' TAG TCC GAA TTC AAG CAA GAG CAC A 3'
Oligo 2 - 5' CTC TTG CTT GAA TTC GGA CTA 3'

The oligos were annealed by heating equimolar amounts of each oligo at 70°C for 10 minutes, to break the secondary structure, then placing the eppendorf into a beaker of water at 65°C and allowing it to cool to room temperature. This permits the oligos to come together gently forming duplex kinase linkers. Before ligating the linkers onto the ends of the cDNA, the cDNA was digested with AluI, producing blunt-end fragments ~250bp in length. 0.1-0.5µg cDNA was then placed in a reaction mix with 1-2µg linkers, 1µl 10mM ATP, 2µl blunt-end ligation buffer (0.66M Tris-Cl (pH 7.6), 50mM $MgCl_2$, 50mM dithiothreitol, 10mM hexaminocobalt chloride, 5mM spermidine-HCl), 1-2 units T4 DNA ligase made up to a final volume of 20µl with sterile water. This was incubated overnight at 16°C. The unincorporated linkers were separated from the linkered cDNA by running the ligation mix on a 2% low melting point (LMP) agarose gel for only 0.5-1cm. Bromophenol blue was not included in the loading buffer as that prevented visualisation of the linkered bands. A gel slice was cut out just above the band of unincorporated linkers. PCR reactions were set up directly from the gel slice, by melting the agarose and transferring an aliquot into pre-heated water, 10µl of the LMP agarose was used as template in a 1ml reaction. The shortest oligo was used as the PCR primer. 10µl of the PCR was electrophoresed on a 1% agarose gel and a smear was observed at ~500bp. The rest of the PCR was purified by Microcon 30 (Amicon), to remove the primer and the nucleotides.

2.13.5 Biotinylation of BAC DNA

1µg of BAC DNA was biotinylated using Nick Translation kit by Gibco-Brl. The kit was used as the manufacturers recommendations with the following modifications:

- a mixture of 0.18M dTTP/0.02M BiodUTP (Boeringer Mannheim) was used.

- various dilutions of DNAse I were used.
- the reaction was stopped with 1µl 0.5M EDTA, 0.5µl 10% SDS and heated to 70°C for 10 minutes.

12.5µl of the reaction was loaded onto a 1% agarose gel looking for a smear ranging from 200bp to 2kb. The rest of the DNA was ethanol precipitated and resuspended in 20µl water.

2.13.6 Blocking cDNA

Before hybridisation the DNA must be pre-blocked with COT-1 DNA to reduce the chances of selecting repeated sequences. 2µg of PCR amplified linkered cDNA and 2µg human COT-1 DNA were made up to 10µl volume with water, heated to 95°C to denature and 10µl 2xhybridisation buffer was added (1.5mM NaCl, 40mM NaPO4, pH 7.4, 10mM EDTA, 10xDenharts, 1.2% SDS). Parafin oil was placed on the top and the mix was incubated at 65°C for 4 hours.

4.13.7 Selection and analysis of retained cDNA

100ng of biotinylated BAC was denatured and added to 20µl blocked cDNA, 5µl 2xhybridisation buffer was also added. This was incubated at 65°C for at least 54 hours, usually 64 hours. 100µl streptavidin beads (Dynal) were washed 3 times with 100µl 2xbinding buffer (10mM Tris-HCl, 1mM EDTA, 2M NaCl), the hybridisation mix was diluted to 100µl volume with water and added to the beads resuspended in 100µl 2xbinding buffer. The beads were incubated with the beads for 15 minutes at room temperature, with rotation. The beads were then washed twice with 500µl 1xSSC, 0.1% SDS at room temperature, then three times with 500µl 0.1xSSC, 0.1% SDS at 65°C. To elute the DNA, the beads were incubated with 50µl 100mM NaOH for 10 minutes at room temperature and neutralised with 50µl 1M Tris, pH 7.4. The DNA was desalted with microcon 30 filter (Amicon) and resuspended in 100µl water.

1µl of eluted cDNA was used as template in 50µl PCR reaction, with 1µl starting library as control. The PCR product was electrophoresed on 1.5% agarose gel and then purified by Microcon 30 filter. 2µg of eluted cDNA was reblocked with COT-1 DNA and hybridised with 100ng biotinylated BAC DNA as before. The hybridisation mix was incubated with the beads, washed, eluted and then re-amplified. This eluted cDNA was the second cycle of selection and was cloned into T-vector.

Recombinants clones were screened by PCR and hybridised to a variety of probes including the BAC used in the experiment, COT-1 DNA and a gene present on the BAC. "Interesting" clones were sequenced.

4.14 DATABASE SEARCHES

Sequence analysis and database searches, using the Fasta and Tfasta programs (Pearson and Lipman 1988), were as implemented in the Wisconsin GCG package release 9.0.

CHAPTER 3

Characterisation and sequencing of new members of the 3ß-HSD gene family

3.1 INTRODUCTION

Although only two isoforms of 3ß-HSD have been described in humans, Southern analysis of human genomic DNA and the identification of novel 3ß-HSD sequences during mutation screening experiments suggested that more than two existed (section 1.9). This information prompted Martin McBride to screen two λgem11 human genomic libraries for sequences similar to 3ß-HSD type I and type II. These screens identified phage clones 2(7) and 8(3) which gave exon 3 amplification products that migrated differently in denaturing gradient gel electrophoresis (DGGE) from 3ß-HSD type I and type II. These fragments were cloned and novel exon 3 sequence was obtained (McBride *et al*, 1996). To determine wheither these phage clones contained putative new members of the 3ß-HSD gene family, they were characterised and the coding regions sequenced.

3.2 RESULTS

3.2.1 Purification of clone 2(7)

Lambda clone 2(7) was plated onto NZCYM medium containing ER1647 cells, the resulting plaques were immobilised onto nylon filters and hybridised with 3ß-HSD typeI full-length cDNA. Plugs were picked from positive plaques and the resultant phages were replated. This was repeated until every plaque on a plate yielded a positive signal and high titre stocks of pure clone 2(7) phage were made.

3.2.2 Does clone 2(7) contain full-length 3ß-HSD gene sequence?

To determine wheither clone 2(7) contained the entire 3ß-HSD coding sequence, the filters of 2(7) plaques prepared above were probed with HSD3B exon 2 and exon 4b amplification products from genomic DNA. Hybridisation was observed with the HSD3B exon 2 probe but not the exon 4b probe (Fig. 3.1). This indicated that clone 2(7) may not contain the entire coding sequence of a 3ß-HSD gene, ending prematurely before exon 4b.

Exon 2 probe Exon 4b probe

Fig. 3.1 Clone 2(7) plaques lifts hybridised with HSD3B exon 2 and exon 4b probe.

3.2.3 PCR amplification of clone 2(7) exons

A range of primers were available to amplify specific 3ß-HSD type I and type II exons and different combinations were chosen to attempt to amplify homologous exons from clone 2(7) (Fig. 3.2; primer sequences can be found in section 2.6.2).

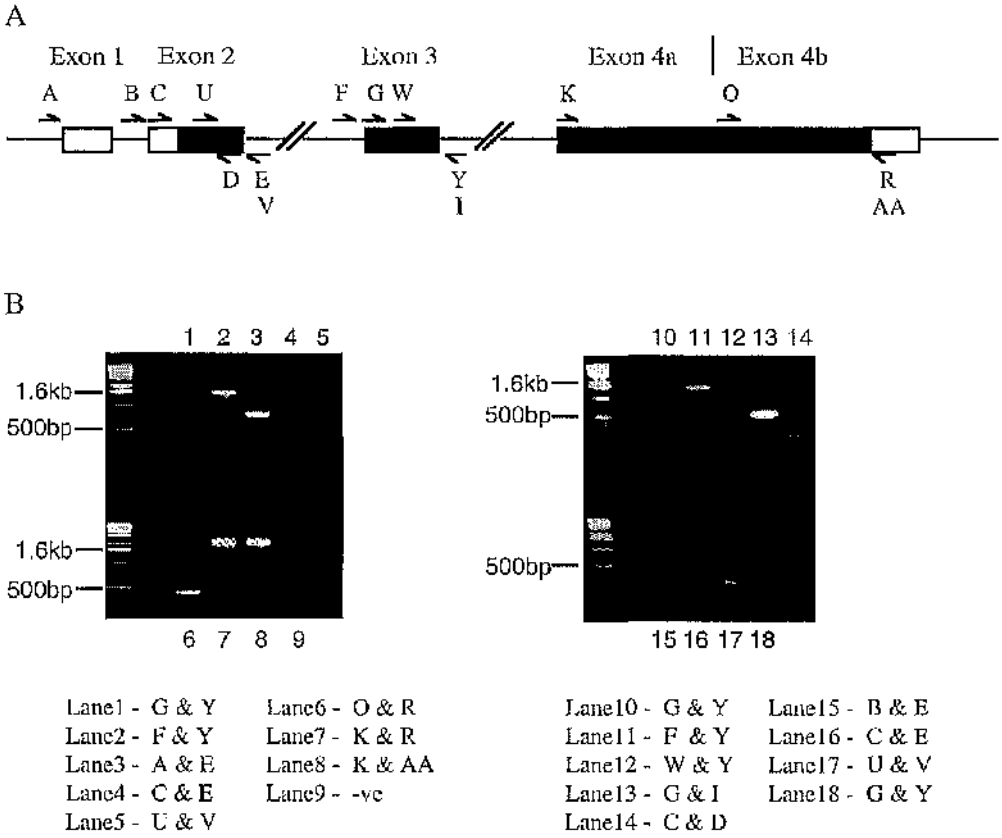


Fig. 3.2 A: Diagram of HSD3B1 indicating the type I and type II primers screened for their ability to amplify homologous exons in clone 2(7). Open boxes represent untranslated regions of HSD3B1 whereas filled boxes represent the translated regions.
B: Agarose gel of PCR amplification products obtained using clone 2(7) DNA as template in the reaction. Primer pairs used as indicated.

Amplifications of the expected size (compared to type I and II) were obtained from:

(i) Lane 6, O and R — Exon 4b, expected size 480bp
(ii) Lane 13, G and I — Exon 3, expected size 590bp
(iii) Lane 14, C and D — Exon 2, expected size 205bp
(iv) Lane 17, U and V — Exon 2, expected size 200bp

Amplification products were observed in lanes 1, 2, 3, 5, 7, 8, 10 and 11, however the fragment sizes indicated were larger than expected from comparisons with type I and II sequence. This suggested that these were spurious amplification products obtained due to sequence divergence between clone 2(7) and type I and II, and therefore these were not analysed any further.

The PCR products from (i), (ii) and (iii) were cloned into T-vector and single colony gels determined which recombinants contained inserts. No recombinants with inserts were identified for exon 2. Clones 1-6, containing exon 3, and clones 7-11, apparently containing exon 4b, were tested by PCR using exon 3 and exon 4b primers (Fig. 3.3A). A strong PCR product was obtained for exon 3, but the exon 4b products were very faint. Clones 4, 6, 7 and 9 were digested with restriction enzymes HincII and SmaI to examine the length of the insert cloned (Fig. 3.3B).

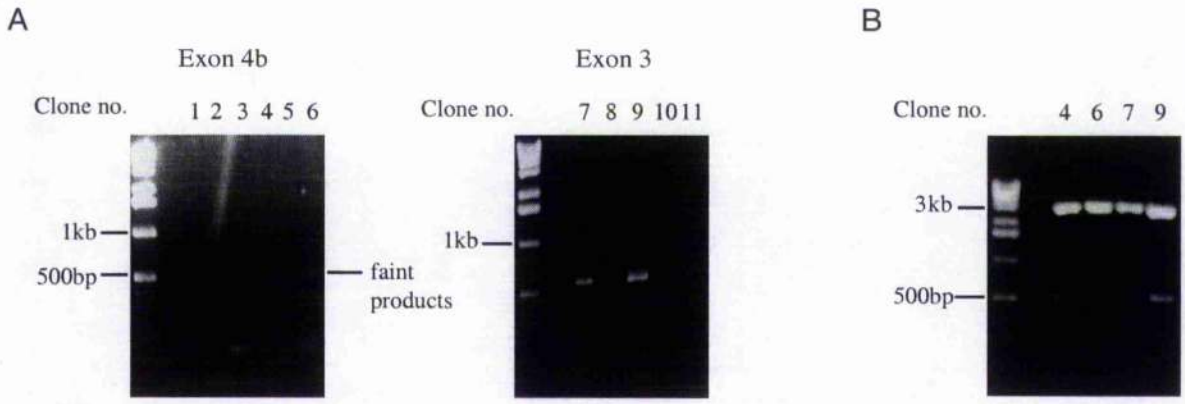


Fig. 3.3 A: PCR amplification of T-vector clones using primer pair O-R (exon 4B, clones 1-6) and primer pair G-I (exon 3, clones 7-11).
B: Restriction digest of positive recombinants clones with HincII and SmaI to excise the insert from vector. These restriction digests confirmed the insert size.

Clones 4 and 7 contained inserts of 460bp and 580bp, respectively, the expected size for exon 4b and exon 3 amplification products. The inserts were sequenced using vector PCR primers T7 and U19. Sequence homologous to HSD3B exon 3 was obtained from clone 7. Surprisingly, sequence from clone 4 was homologous to

HSD3B exon 4a and not exon 4b as expected (Fig. 3.4). This suggested that the conditions used for amplification were not stringent enough, allowing primers to anneal to exon 4a sequence. This would occur if, as suggested from the plaque screening, an exon 4b homologue is not present in clone 2(7). It was not possible to examine clone 2(7) sequence for putative primer sequences because the PCR product sequence from clone 4 contained intron sequence and the intron sequence was not available over the predicted primer site region.

```
Clone4      56 GTGGTTGCAGCTCCTTTGGGATATTTCTTGACACTGTCATCATGCTCTTC 105
               |||||||    |  ||| |.|||||  || |||||  ||  ||  |||||||||
3ß-typeI  7956 GTGGTTGGCACCTCTTAGGGATATATCCTGACAGTGACAATATGCTCTTC 8005

Clone4     106 GTGGGCAGGTACCCGGCT.GAGTTGGAGGCCTGTGTCCAAGCCAGTGTGC 154
               |||  |||||||||  |||   |||  ||||||||||||||||  |||||||
3ß-typeI  8006 ATGGACAGGTACCCAGCTCCTGTTAGAGGCCTGTGTCCAAGCTAGTGTGC 8055

Clone4     155 CAGTCTTCATCTACACCAG 173
               :||||||||||||||||||
3ß-typeI  8056 CAGTCTTCATCTACACCAG 8074
```

Fig. 3.4 Sequence of amplification product from clone 2(7) using primers designed to HSD3B1 exon 4b, aligned with sequence of HSD3B1 exon 4a. Co-ordinates shown are from HSD3B1 (Genbank accession no. m38180).

3.2.4 Subcloning clone 2(7) into pUC18

The attempts to amplify HSD3B exon homologues from clone 2(7) yielded sequence for exon 3, but difficulties were found with cloning exon 4 and exon 2 product (exon 1 and 2 amplifications were repeated several times without success). It was therefore decided to subclone clone 2(7) into pUC18 cloning vector. This would have the additional benefit of overcoming any complications due to Taq Polymerase-induced mutations and the need to sequence over primer sites.

High titre stocks of pure clone 2(7) stock were used to infect established cultures of *E.coli*, after cell lysis the cell debris was removed producing a clear lysate containing infectious phage. Phage DNA was isolated from the lysate using λ phage purification kit (Promega). The DNA was digested with a panel of restriction enzymes and hybridised with exon-specific probes to find suitable restriction bands for cloning (Fig. 3.5). Ideally, bands should contain an entire exon and be small enough to allow trouble-free cloning and sequencing.

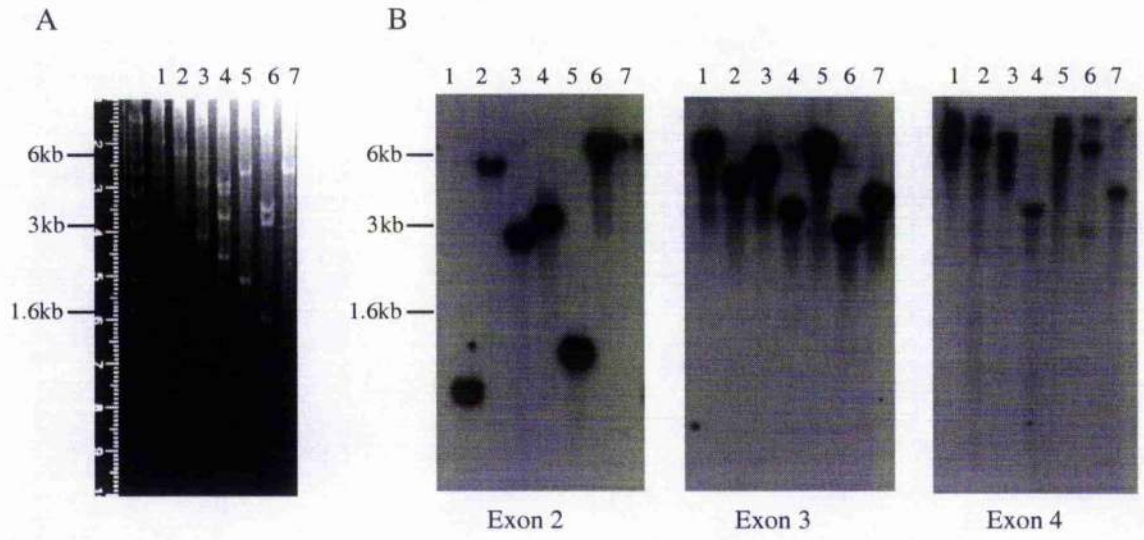


Fig. 3.5 A: Agarose gel showing restriction digests from clone 2(7) using a variety of restriction enzymes.
(Lane1 - BamHI, Lane2 - EcoRI, Lane3 - HindIII, Lane4 - SstI, Lane5 - XbaI, Lane6 - KpnI, Lane7 - SalI)
B: Southern analysis of restriction digests from panel A hybridised to HSD3B exon specific probes. Exon specific probes were synthesised from exon specific PCR amplification products with HSD3B1 template and primer pairs C-D for exon 2, G-I for exon 3 and K-R for exon 4.

The results in Fig. 3.5 indicated that SstI was the best restriction enzyme to use to subclone the exons, as it yielded a band of approximately 3.3kb for exon 2 and either a singlet or doublet band of about 4.0kb for exons 3 and 4. As only 5 fragments are produced from SstI digest of clone 2(7) (Fig. 3.5A) it was possible to "shotgun" clone the entire digest into pUC18 and screen for the subclones of interest. Many recombinants were produced and those with inserts were tested by PCR using exon-specific primers pairs (Fig. 3.6).

Fig. 3.6 Agarose gel of exon-specific amplification of SstI subclones of clone 2(7). Primers used were C-D for exon 2 (expected size 205bp), G-I for exon 3 (expected size 590bp) and K-Q for exon 4 (expected size 490bp). + indicates the positive control reaction using HSD3B1 template, this controls for the efficiency of the reaction as well as providing the expected PCR product size. - indicates the no template negative control reaction.

Clone 11 (containing exon 2) and clone 25 (containing exons 3 and 4) were mapped and sequenced (Fig. 3.7A and B). Exon 2 and exon 3 amplification products were obtained from clone 23, this was unexpected as different sized SstI fragments for these exons were observed in Fig. 3.5. It is possible that this clone represents a fragment from a SstI partial digest, a SstI digest of clone 23 would confirm if this assumption is correct. Restriction analysis of subclone 25 was accomplished using a selection of restriction enzymes and single and double digests were attempted. pUC primers and HSD3B1 and 2 specific primers were used to obtain sequence (Fig. 3.7). The sequence obtained from the pUC primers provided the approximate end-points of the 2(7) subcloned insert.

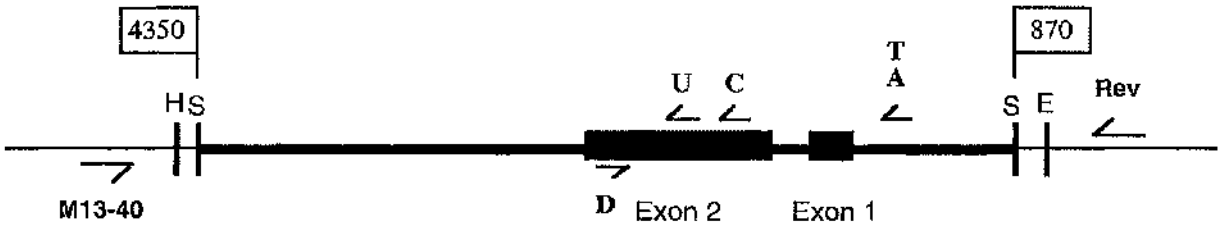


Fig. 3.7 A: Schematic diagram of 2(7) subclone 11 indicating the position of the sequencing primers, approximate coordinates (boxed) are taken from equivalent sequence in HSD3B1 (S=SstI, E=EcoRI, H=HindIII).

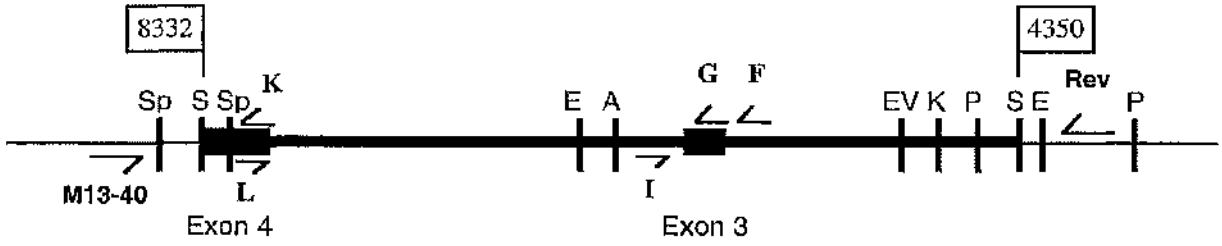


Fig. 3.7 B Restriction map of 2(7) subclone 25 indicating the position of sequencing primers, approximate coordinates (boxed) are taken from equivalent sequence in HSD3B1 (Sp=SphI, S=SstI, E=EcoRI, A=AccI, EV=EcoRV, K=KpnI, P=PstI).

Clone 25 contained HSD3B sequence homologous to HSD3B1 from base 4350 to 8332. The sequence from 2(7) subclone 25 using vector primer M13-40 indicated the end of HSD3B sequence within the phage clone. The phage genomic library was created using human genomic DNA partially cut with Sau3A with the sticky ends of the digest filled in with bases AG to prevent religation of inserts, and vector DNA partially cut with XhoI with the digest ends filled with bases TC. This resulted in complementary pairing for insert and vector to create the library. This cloning site was adjacent to the SstI site which was utilised in subcloning clone 2(7) into the pUC vector. The sequence obtained from clone 25 using primer M13-40 contained the Sau3A restriction site which arises only from the phage cloning and therefore it appears

that 2(7) HSD3B sequence in the phage clone ends at HSD3B1 equivalent base 8332 (Fig. 3.8).

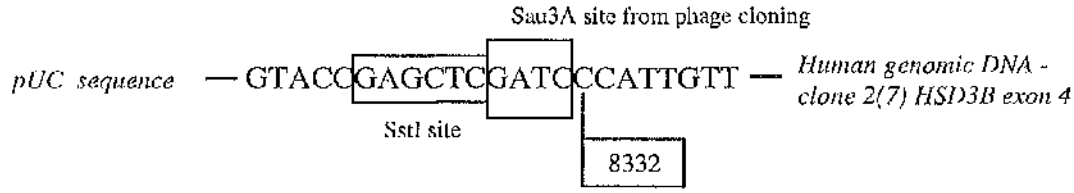


Fig. 3.8 Diagram to indicate the sequence obtained from pUC primer M13-40 and show where the 2(7) HSD3B insert ends within the phage vector.

This information confirms that an exon 4b homologue of HSD3B1 is not present in clone 2(7) since exon 4b sequence starts from approximately base 8400 in HSD3B1. However, coincidently an exon 4 PCR amplification product (primer pair K-R) from another phage clone identified from the library screen (clone 1(2)) produced sequence identical to clone 2(7) exon 4a sequence and extended into exon 4b sequence. Therefore, this clone completed the sequence necessary for clone 2(7) HSD3B exon 4 (Fig. 3.9).

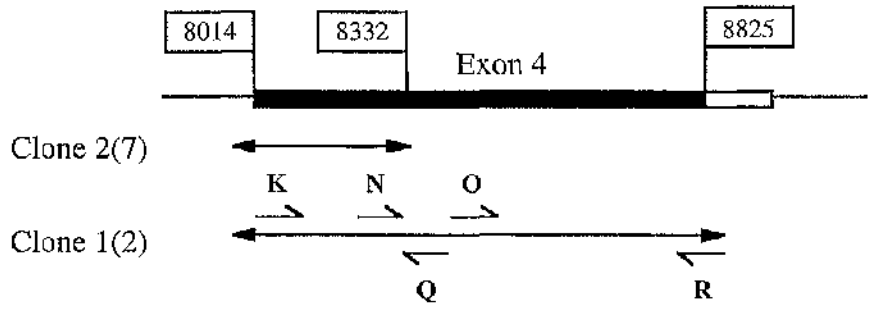


Fig. 3.9 Schematic diagram to represent the sequencing of exon 4 from clone 2(7) and clone 1(2). Primers used for clone 1(2) sequencing are indicated. Approximate coordinates (boxed) are taken from the equivalent sequence in HSD3B1.

A result of sequence divergence between clone 2(7) and 3ß-HSD type I and II was that some of the HSD3B1 and 2 primers did not produce sequence from 2(7) subclone 11, in particular primers A and T (Fig. 3.7A). Consequently, no sequence was obtained from subclone 11 homologous to exon 1 and the 5' end of exon 2. Therefore, to complete the sequencing of clone 2(7) a further exon 1 and 2 subclone was produced. A BamHI fragment from 2(7), approximately 900bp, containing exon 2 (Fig. 3.5) was subcloned into pUC and recombinants were screened by exon 2 specific PCR and

BamHI digests (Fig. 3.10A and B). Clone D yielded a PCR product the correct size (205bp) and contained the appropriate size of insert (900bp) and this was used to obtain exon 1 and 5' exon 2 sequence (Fig. 3.10C).

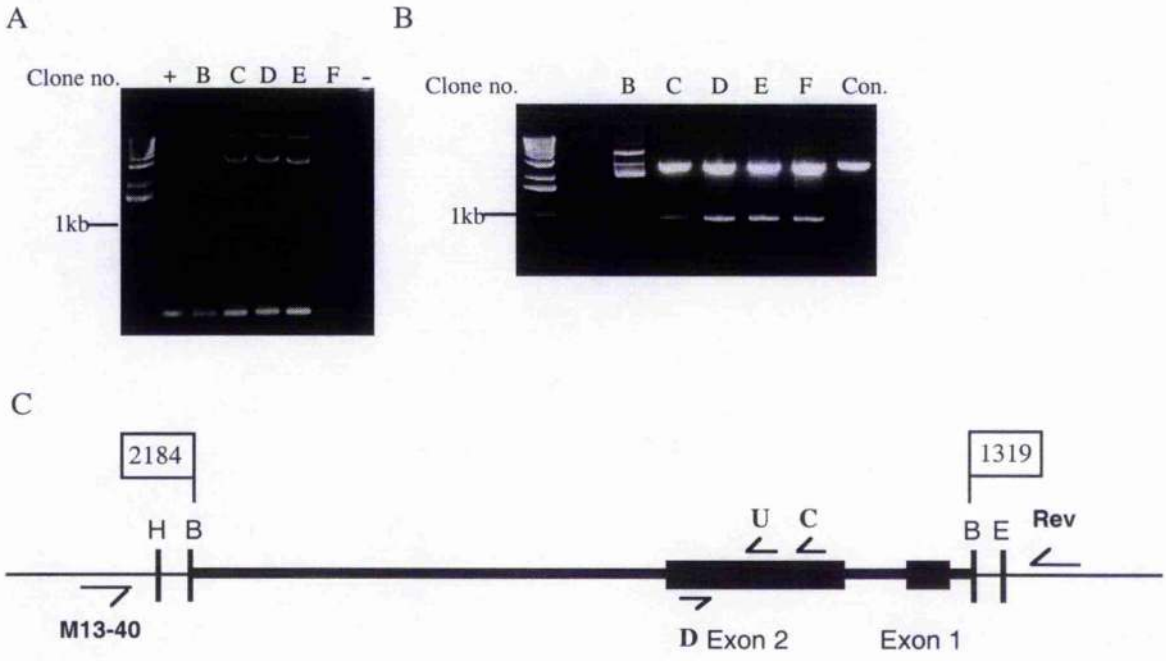


Fig. 3.10 A: Agarose gel of PCR amplification products obtained using recombinant clone template from clone 2(7) BamHI subcloning using exon 2 primer pair C-D, expected size 205bp. + indicates positive control using HSD3B1 template, - indicates no template negative control.

B: Agarose gel of BamHI restriction analysis of recombinant clones from clone 2(7) BamHI subcloning to determine insert size, expected insert size 900bp. Con. indicates control digest using DNA from a blue colony obtained during cloning. This clone does not contain an insert and indicates that blue/white selection was successful.

C: Schematic diagram of subclone D indicating which primers were used to obtain sequence, approximate coordinates (boxed) are taken from equivalent sequence in HSD3B1 (H=HindIII, B=BamHI, E=EcoRI).

Therefore from subclones 11, 25, D and clone 1(2) the entire 3ß-HSD coding sequence from clone 2(7) was obtained (Fig. 3.11).

a) Exons 1 and 2 alignment

```
          1                                                  50
          *                                                  *
   Type2  GAGGCAGTAA GGACTTGGAC TCCTCTGTCC AGCTTTT-AA CAATCTAAGT
   Type1  ....TGAG.. .T..G.CC.. ..T....... .......-.. ........C.
Clone2-7  ....TGACC. .C..CCT... .A........ ....A..-.. T.......C.
```

```
          51                                                  100
          *                                                     *
   Type2  TACGgttaga gctttctcct tttctttcaa ctact---cc tggcagttgt
   Type1  A.T..... .a...t..a. .........g .....---.. ..-....g..
Clone2-7  A.T..... ...a.t..g. .........g .....---t. .....a.g..

          101                                                 150
          *                                                     *
   Type2  ggggtcatgg aatttttgta aaaaa----tg ggggtggagga aaataaggca
   Type1  ....a..ca. ...g....c. .....aa-.. ......... ....g.....
Clone2-7  .a..at.... ........c. .....aaa.. .......a.t ...-..a...

          151                                                 200
          *                                                     *
   Type2  tctg-ctgag tgtataacca ttttacctct tgtttttagC CCTCTTCTGG
   Type1  ....tg.... .a........ ...g..a..- .c........ .....C.A..
Clone2-7  ....tg.... .......at. .c.g.t...- cc........ .-...C.A-.

          201                                                 250
          *                                                     *
   Type2  GTCACGCTAG AATCAGATCT GCTCTCCAGC ATCTTCTGTT TCCTGGCAAG
   Type1  .....C.... .......... ....C..... .......... ......TG..
Clone2-7  .A...C.... ...A...... T...C..... .......... ......AG..

          251                                                 300
          *                                                     *
   Type2  TGGTTTCCTG CTACTTTGGA TTGGCCACGA TGGGCTGGAG CTGCCTTGTG
   Type1  ...A...... .......... -.....ATG. C. ....... ..........
Clone2-7  .......... ....C.C... ......ATG. CA........ ..........

          301                                                 350
          *                                                     *
   Type2  ACAGGAGCAG GAGGGCTTCT GGGTCAGAGG ATCGTCCGCC TGTTGGTGGA
   Type1  .......... .....T.... ...A...... ...A...... .C........
Clone2-7  .......... .....T.... .......... ...A..T... ....A.....

          351                                                 400
          *                                                     *
   Type2  AGAGAAGGAA CTGAAGGAGA TCAGGGCCTT GGACAAGGCC TTCAGACCAG
   Type1  G........G .......... ......T... .......... ...G......
Clone2-7  G....CA..G .......... ..CC...... ....-..... ..........

          401                                 430
          *                                     *
   Type2  AATTGAGAGA GGAATTTTCT Agtaagtaaa
   Type1  .......... .......... ..........
Clone2-7  G..TGA.G.. .......... ..........
```

b) Exon3 alignment

```
          1                                                    50
          *                                                     *
   Type2  ccaatgacct gacctgtgtt cacacagAGC TCCAGAACAG GACCAAGCTG
   Type1  .......... .......... ........A. .........A ..........
Clone2-7  .......... a....cn.c. ........A. .........A. .T.......
```

```
          51                                                        100
          *                                                          *
   Type2  ACTGTACTTG AAGGAGACAT TCTGGATGAG CCATTCCTGA AAAGAGCCTG
   Type1  ..A..G..G. .......... .......... .......... .G........
Clone2-7  ..A..G..G. .......... .......... .......... .G........

          101                                                       150
          *                                                          *
   Type2  CCAGGACGTC TCGGTCGTCA TCCACACCGC CTGTATCATT GATGTCTTTG
   Type1  .......... ......A... .......... .......... ........C.
Clone2-7  .........G .......... .......... .......... ........C.

          151                                                       200
          *                                    (G)                   *
   Type2  GTGTCACTCA CAGAGAGTCC ATCATGAATG TCAATGTGAA AGgtacagta
   Type1  .......... .........T .......... .......... .....tg...
Clone2-7  .A........ ....C....T .......... .......... .....gg...

          201                          230
          *                             *
   Type2  gcctggggag gagataaaac aagttggtt-
   Type1  .g........ .....gc.g. ...g...gg-
Clone2-7  ......nnnn nnnnnnnnnn nnnnnnnnnn
```

c) Exon 4 alignment

```
          1                                                          50
          *                                                          *
   Type2  tcttcgtggg cagGTACCCA GCTACTGTTG GAGGCCTGTG TCCAAGCCAG
   Type1  .....a...a .......... ...C.....A .......... .......T..
Clone2-7  ....tc.... .g........ ...C...... .......... .......T.C

          51                                                        100
          *                                              (A)         *
   Type2  TGTGCCAGTC TTCATCTACA CCAGTAGCAT AGAGGTAGCC GGGCCCAACT
   Type1  .......... .......... .......... .......... ..........
Clone2-7  .......... .......... ......C.C. CC..T..... ..........
          101                                                       150
          *                                                          *
   Type2  CCTACAAGGA AATCATCCAG AACGGCCACG AAGAAGAGCC TCTGGAAAAC
   Type1  .......... .......... ..T.....T. .......... ..........
Clone2-7  .......... ......TG.. ..T.C...T. .......CTT ..........

          151                                                       200
          *                                                          *
   Type2  ACATGGCCCA CTCCATACCC GTACAGCAAA AAGCTTGCTG AGAAGGCCGC
   Type1  .........C .......... AC........ .......... ..........
Clone2-7  ......T.TG .......T.. A......... ...T...... ..........

          201                                                       250
          *                                                          *
   Type2  GCTGGCGGCT AATGGGTGGA ATCTAAAAAA TGGTGATACC TTGTACACTT
   Type1  A......... ..C....... ....G..... C..C.GC... C.........
Clone2-7  ......A... .......... C...G..... .......... ..........

          251                                                       300
          *                                                          *
   Type2  GTGCGTTAAG ACCCACATAT ATCTATGGGG AAGGAGGCCC ATTCCTTTCT
   Type1  ....C...C. .....TG... .......... .....A.... .A........
Clone2-7  ....C..... .....TG... .......... .....A.... ...T......
```

```
          301                                                       350
          *                                                         *
   Type2  GCCAGTATAA ATGAGGCCCT GAACAACAAT GGGATCCTGT CAAGTGTTGG
   Type1  ..T..... .C..... ..... ..... .....
Clone1-2  ....A..... ..... ..... ..... .G.....CA.

          351                                                       400
          *                                                         *
   Type2  AAAGTTCTCT ACAGTCAACC CAGTCTATGT TGGCAACGTG GCCTGGGCCC
   Type1  .........C ..T..T.... ..... ......T... .....
Clone1-2  C........C ....C..... ..... ..... .....

          401                                                       450
          *
   Type2  ACATTCTGGC CTTGAGGGCT CTGCGGGACC CCAAGAAGGC CCCAAGTGTC
   Type1  ..... .........C ....A..... ..... ......CA..
Clone1-2  ..... ..... -..... ..... .....

          451                                                       500
                                                (T)                 *
   Type2  CGAGGTCAAT TCTATTACAT CTCAGATGAC ACGCCTCACC AAAGCTATGA
   Type1  .....A..G. ....C..T.. ..... ..... .....
Clone1-2  .T...A..G. ....C..T.. ..... ..... .....T....

          501                                                       550
          *                                                         *
   Type2  TAACCTTAAT TACATCCTGA GCAAAGAGTT TGGCCTCCGC CTTGATTCCA
   Type1  ..... ....C..... ..... ..... .....
Clone1-2  ..... ......T... ..... C....C.T.. .....

          551                                                       600
          *                                                         *
   Type2  GATGGAGCCT TCCTTTAACC CTGATGTACT GGATTGGCTT CCTGCTGGAA
   Type1  ........T. .......T.. ........T. ..... .....
Clone1-2  ....A..... ..T....TT. ..... ..... ........G.

          601                                                       650
          *                                                         *
   Type2  GTAGTGAGCT TCCTACTCAG CCCAATTTAC TCCTATCAAC CCCCCTTCAA
   Type1  A..... ..... G..... A......G.. .G.....
Clone1-2  A........C ....G..... G...G...G. A..... .G......G.

          651                                                       700
          *                                                         *
   Type2  CCGCCACACA GTCACATTAT CAAATAGTGT GTTCACCTTC TCTTACAAGA
   Type1  ........T. ........G. .......C.. A..... .....T....
Clone1-2  .T........ ..G.....G. .......CA. ..... .....

          701                                                       750
          *                                                         *
   Type2  AGGCTCAGCG AGATCTGGCG TATAAGCCAC TCTACAGCTG GGAGGAAGCC
   Type1  ..... ....T..... ..... ..... .....
Clone1-2  ..... .........A ..... .T..... .....

          751                                                       800
          *                                                         *
   Type2  AAGCAGAAAA CCGTGGAGTG GGTTGGTTCC CTTGTGGACC GGCACAAGGA
   Type1  ..... .G..... ..... ..... .....
Clone1-2  ..... ..A..... A..... ..... .....
```

```
          801                                                       850
          *                                                          *
Type2     GACCCTGAAG TCCAAGACTC AGTGATTTAA GGATGACAGA GATGTGCATG
Type1     ..A....... .......... ..TGA..... .......... ..........
Clone1-2  ..A....... .......... ....C..... .NNNNNNNNN NNNNNNNNNN

          851                                                       900
          *                                                          *
Type2     TGGGTATTGT TAGGAAATGT CATCAAACTC CACCCACCTG GCTTCATACA
Type1     .......... .....G.... ......G... .....T.... ..C.......
Clone1-2  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
```

Fig. 3.11 Sequence obtained from clones 2(7) and 1(2) aligned with homologous sequences from 3ß-HSD type I and II. Triplet nucleotides underlined are translational stop and start codons. Dashes indicate gaps introduced for alignment. Bases corresponding to intron sequence in type I and II are shown in lower case.

3.2.5 Long range PCR to determine the size of HSD3B intron homologues in clone 2(7)

Clone 2(7) contains sequence homologous to HSD3B introns (intron sequence was observed from subclones 25, 11 and D), therefore long range PCR was performed to determine the length of the intron sequence present. PCR amplification was attempted over introns II (exon 2 to 3) and III (exons 3 to 4) using the long template PCR system (Boeringer Mannheim). Primers U and H were used for intron II and primers G and L for intron III, and PCRs were extended at 68°C for 7 minutes (Fig. 3.12). Clone 2(7) yielded amplification products of approximately 4kb for intron II and approximately 2.5kb for intron III, which correspond well with HSD3B1 intron lengths of 4.2kb and 2.6kb for II and III respectively.

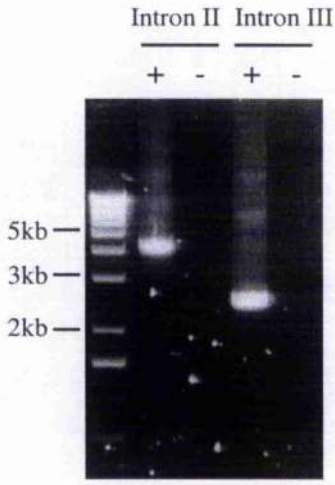


Fig. 3.12 Agarose gel indicating long range PCR amplification of intron II and III from clone 2(7). Primers U and H were used for intron II, G and L were used for intron III.

3.2.6 Identification of clone 2(7) as a pseudogene

The 3ß-HSD coding sequence determined from clone 2(7) contained frameshft mutations resulting in premature stop codons. The first frameshift mutation is the deletion of A producing a stop codon in the predicted polypeptide sequence at residue 42 found at the end of exon 2 (Fig. 3.13). No other significant open reading frames were found within the sequenced fragments. Therefore, it was concluded that clone 2(7) was an unprocessed pseudogene and was thereafter designated 3ß-HSDψ2.

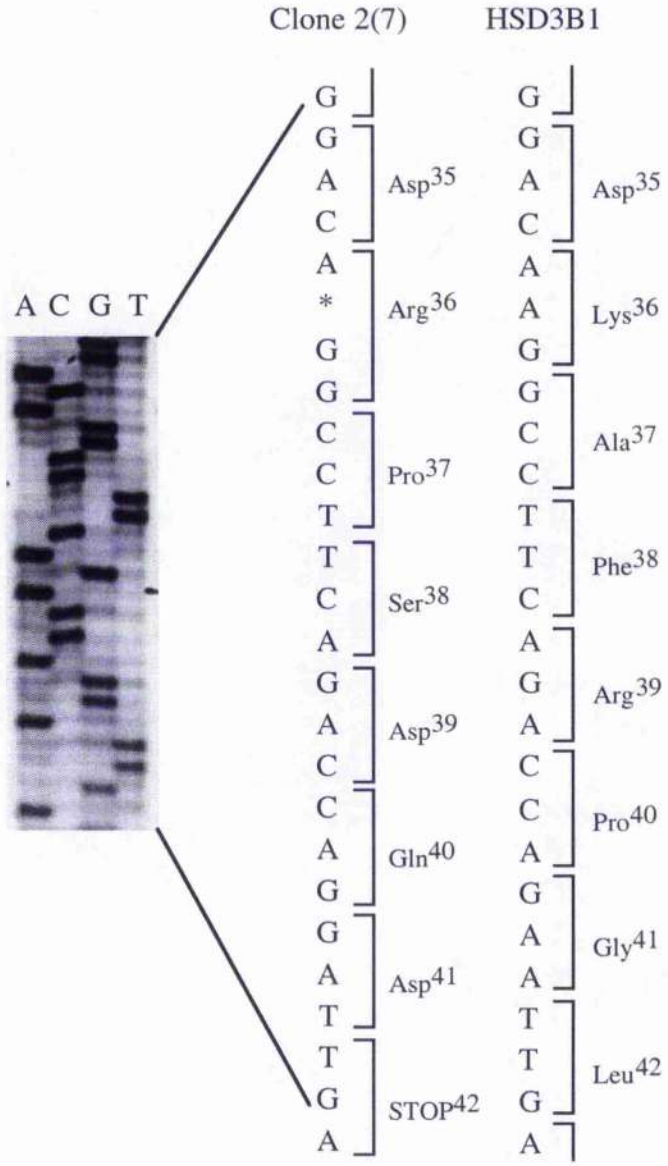


Fig. 3.13 Clone 2(7) sequence indicating deletion of A and compared with homologous HSD3B1 sequence. This delection caused a frameshift and produced a STOP codon at residue 42. Deletion is represented by *.

3.2.7 Expression studies of clone 2(7)

Previously in the laboratory, Sandra Burridge detected RNA transcripts in the placenta from other pseudogenes identified in the λ phage library screen. These transcripts were alternatively spliced. To examine the possibility that 3ß-HSDψ2 may be expressed, primers specific to ψ2 were designed and RT-PCR from the placenta was attempted. The placenta was chosen as it is a major steroidogenic tissue and because the other 3ß-HSD pseudogene transcripts were found here. Term placenta was obtained from Queen Margaret's Maternity Hospital and RNA was isolated using RNAzol B. cDNA was synthesised using the cDNA synthesis kit from Clontech. RT-PCR was carried out using the ψ2 specific primers (2(7) 5' and 2(7) 3'; primer sequences can be found in section 2.6.2), type I primers (type I and D) were used as a positive control and the reactions were carried out at 45°C annealing. The gels was blotted and probed with full-length 3ß-HSD type I cDNA (Fig. 3.14).

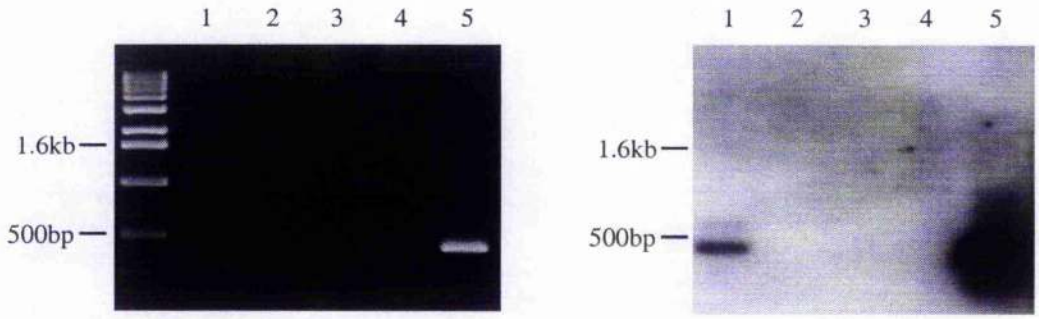


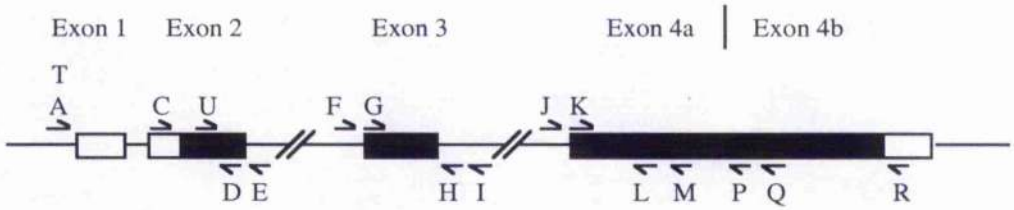
Fig. 3.14 Gel electrophoresis of RT-PCR products using HSD3Bψ2 specific PCR primers with placental cDNA as template. Hybridisation of Southern blot with 3ß-HSD typeI cDNA. Lane 1 - placental cDNA, lane 2 - no reverse transcriptase control, lane 3 - no RNA control, lane 4 - no template PCR control, lane 5 - HSD3Bψ2 template, positive control.

A smear was observed for the RT-PCR using HSD3Bψ2 specific primers, although a distinct band of the expected size (450bp) was detected after hybridisation with type I cDNA (lane 1, Fig. 3.14). The amplification products found in lane 1 were cloned into T-vector, white colonies were screened for inserts by single colony gel and recombinants with inserts were screened by PCR using the ψ2 specific primers. 8 recombinants positive for ψ2 specific PCR were sequenced and from these 3ß-HSD type I sequence only was detected. Therefore, no HSD3Bψ2 transcripts were detected from this experiment and it is likely that this pseudogene is not expressed.

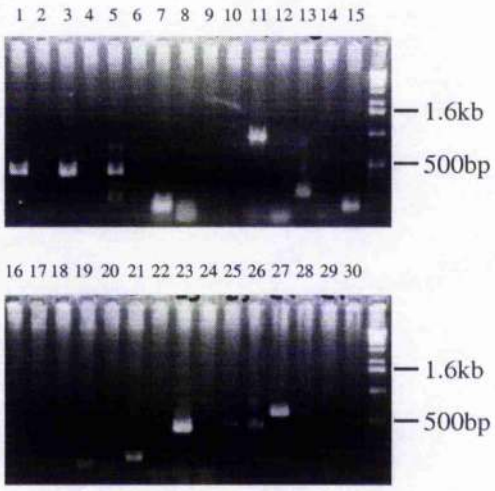
3.2.8 PCR amplification of clone 8(3)

Phage clone 8(3) was another clone discovered from the genomic library screens and novel exon 3 sequence was identified by DGGE analysis and sequencing. Martin McBride initiated sequencing of the coding region of this clone; myself and Tilda Gordon, an undergraduate student, completed the sequencing and subcloned 8(3) exon 4 into pUC 18. The sequencing strategy used was the same as clone 2(7) with exon-specific PCR amplifications and subcloning exons into pUC18. Therefore, to begin with different combinations of typeI and II exon-specific primers were chosen to attempt to amplify exon homologues from clone 8(3) (Fig. 3.15).

A



B



Lane 1 - A & D	Lane 11 - U & E	Lane 21 - K & M
Lane 3 - T & D	Lane 13 - G & H	Lane 23 - K & P
Lane 5 - C & D	Lane 15 - G & I	Lane 25 - K & Q
Lane 7 - U & D	Lane 17 - F & H	Lane 27 - K & R
Lane 9 - C & E	Lane 19 - K & L	Lane 29 - J & L

Fig. 3.15 A: Diagram of HSD3B1 indicating the type I and type II primers used to amplify the homologous exons in clone 8(3). Open boxes represent untranslated regions in HSD3B1 whereas filled boxes represent translated regions.
B: Agarose gel indicating the results from trial amplifications using clone 8(3) as template. Primers used as indicated, even lanes represent no template controls.

Sequence was obtained from exons 1 and 2 (lanes 1, 3 and 7), exon 3 (lane 13) and exon 4 (lanes 19, 21, 23 and 27) products using direct PCR sequencing (Amersham). Direct PCR product sequencing eliminated the need to subclone the products into T-vector. The sequence obtained from the amplification product using primer pair K-R (lane 27) was unexpected. Using the primer R the sequence produced started 200bp from the primer site, even from sequencing reactions containing Mn^{2+} which produces sequence close to the primer sequence (Fig. 3.16). This was repeated several times with the same result. This suggested that primer R was not annealing to the correct sequence and that, similar to clone 2(7), clone 8(3) may not contain the entire coding sequence and may end prematurely before the stop codon.

```
Clone 8(3)      146 CCCAAGAAGGCCCTAAGCATCCGAGGACAGTTCTAANATATCTCAGATGA 97
                    |||||| |||||| |||||||||||||||||||| :|||||||||||||
HSD3B1         8430 CCCAAGAAGGCCCCAAGCATCCGAGGACAGTTCTACTATATCTCAGATGA 8479

Clone 8(3)       96 CATGTCTTACCAAAGCTATGACTACCTTAGTTACACCCTGAACAAAGAAT 47
                    || | || |||||||||||||| |||||| |||||||||| |||||| |
HSD3B1         8480 CACGCCTCACCAAAGCTATGATAACCTTAATTACACCCTGAGCAAAGAGT 8529

Clone 8(3)       46 T.TGCCTCTGCCTTGATTCCAGATGGAGCCTTCCTCTAGCCCTGAT 2
                    |  ||||| |||||||||||| |||||||| ||||| || |||||||
HSD3B1         8530 TCGGCCTCCGCCTTGATTCCAGATGGAGCTTTCCTTTATCCCTGAT 8575
```

Fig. 3.16 Sequence alignment of clone 8(3) sequence from primer R and HSD3B1 sequence. Coordinates from HSD3B1, genbank accession no. m38180.

3.2.9 Subcloning exon 4 from clone 8(3)

To investigate the possibility that clone 8(3) did not containall of HSD3B exon 4 sequence, exon 4 from clone 8(3) was subcloned into pUC18. Exon 1, 2 and 3 sequence was obtained from sequencing PCR products. Clone 8(3) phage DNA was produced and digested with a panel of restriction enzymes, and then hybridised with an exon 4 specific probe to find a suitable restriction band for cloning (Fig. 3.17).

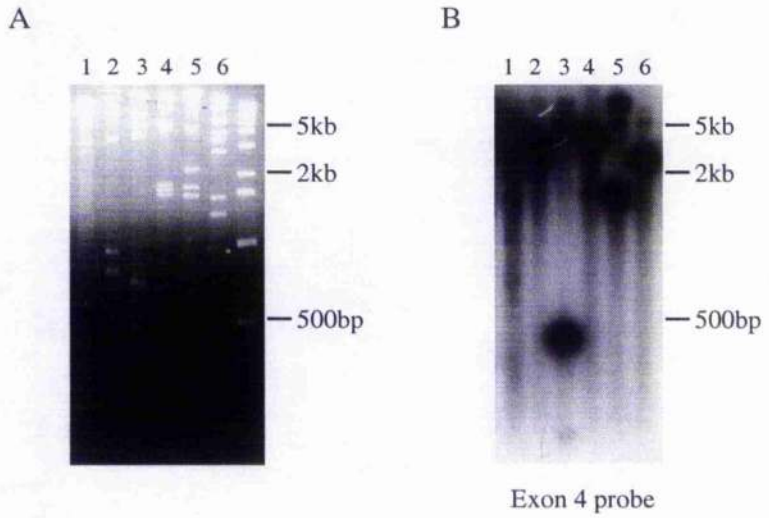


Fig. 3.17 A: Agarose gel indicating the digestion patterns of clone 8(3) with a variety of restriction enzymes. Lane 1 - BamHI, lane 2 - EcoRI, lane 3 - HIndIII, lane 4 - SstI, lane 5 - XbaI, lane 5 - XbaI, lane 6 - KpnI.
B: Southern analysis of restriction digests shown in panel A hybridised to HSD3B exon 4 probe.

The results from Fig. 3.16 suggested that KpnI was the best restriction enzyme to use to subclone exon 4, since a fragment of an ideal size (~3.5kb) containing exon 4 was observed. The KpnI digest was "shotgun" cloned into pUC18, recombinants with inserts were determined by a single colony gel and PCRs with exon 4 specific primer pair identified the clones containing the exon 4 fragment (Fig. 3.18).

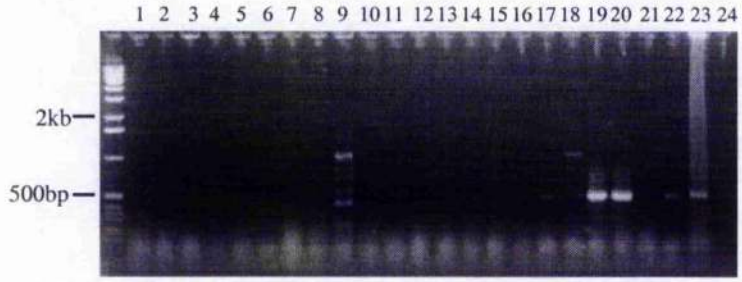


Fig. 3.18 Agarose gel showing e xon 4 specific amplifications of KpnI subclones of clone 8(3), primer pair used was K-Q, expected size 480bp. Lane 24 contains no template control.

Clone 19 was sequenced with pUC specific primers to determine the orientation of the insert (Fig. 3.19).

Fig. 3.19 Schematic diagram of subclone 19 indicating the position of the sequencing primers, coordinates from HSD3B1 (K=KpnI, E=EcoRI, H=HindIII).

Fig. 3.19 indicates that pUC primer M13-50 produced sequence from 5' exon 4 and M13-40 did not produce sequence at all, therefore the precise endpoint of the subclone is unknown. Primer R was used once more to attempt to gain 3' exon 4 sequence. However as with PCR sequencing, the sequence obtained started 200bp from the primer and therefore clone 19 did not provide any further insights into why exon 4b could not be found. It is probable that clone 8(3) does not contain sequence homologous to HSD3B exon 4b, and therefore primer R cannot anneal to the correct sequence and instead produces spurious amplifications and sequence. Exon 4 could be subcloned further until the vector primers read into the 3' end similar to HSD3Bψ2 subclone 25 (Fig. 3.8) which would determine the end of the phage clone. To generate exon 4 sequence it would be possible to use genomic clones BAC 21 or BAC 7 (see Chapter 4) as template for PCR, using an 5' exon 4 8(3) specific primer and primer R.

Direct sequencing of PCR amplification products from clone 8(3) and sequencing of exon 4 from clone 19 determined the sequence of the coding regions homologous to HSD3B except 3' 200bp of exon 4 (Fig. 3.20).

a) Exons 1 and 2 alignment

```
           1                                                  50
           *                                                   *
   Type2   GAGGCAGTAA GGACTTGGAC TCCTCTGTCC AGCTTTT-AA CAATCTAAGT
   Type1   ....TGAG.. .T..G.CC.. ..T....... .......-.. ........C.
Clone8-3   NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN .C..... .. .G......CA

           51                                                 100
           *                                                   *
   Type2   TACGgttaga gctttctcct tttctttcaa ctact---cc tggcagttgt
   Type1   A.T....... .a...t..a. .........g .....---.. ..-....g..
Clone8-3   A.T..c.... .....t..a. ...t...ttt .a...acca. .......g..
```

```
          101                                                       150
          *                                                          *
   Type2  ggggtcatgg aatttttgta aaaaa---tg gggtggagga aaataaggca
   Type1  ....a..ca. ...g....c. .....aa-.. .......... ....g.....
Clone8-3  ....a..ca. ........c. .....a--ga .......... ........g.

          151                                                       200
          *                                                          *
   Type2  tctg-ctgag tgtataacca ttttacctct tgtttttagC CCTCTTCTGG
   Type1  ....tg.... .a........ ...g..a..- .c........ .....C.A..
Clone8-3  ....tg.... .a........ ...g...a.- .c..c.... .....C....

          201                                                       250
          *                                                          *
   Type2  GTCACGCTAG AATCAGATCT GCTCTCCAGC ATCTTCTGTT TCCTGGCAAG
   Type1  .....C.... .......... ....C..... .......... ......TG..
Clone8-3  .....CT... .......... .T..C..... ........C. C......G..

          251                                                       300
          *                                                          *
   Type2  TGGTTTCCTG CTACTTTGGA TTGGCCACGA TGGGCTGGAG CTGCCTTGTG
   Type1  ...A...... .......... -.....ATG. C. ....... ..........
Clone8-3  G......... .......... ..A...ATG. .........A ..T.......

          301                                                       350
          *                         (A)                              *
   Type2  ACAGGAGCAG GAGGCCTTCT GGGTCAGAGG ATCGTCCGCC TGTTGGTGGA
   Type1  .......... .....T.... ...A...... ...A...... .C........
Clone8-3  .......... ..AN.T.... ..A.......A ...A...AT. .C......A.

          351                                                       400
          *                                                          *
   Type2  AGAGAAGGAA CTGAAGGAGA TCAGGGCCTT GGACAAGGCC TTCAGACCAG
   Type1  G........G .......... ......T... .......... ...G......
Clone8-3  G........G .......... .......T.. ...A...... ...G......

          401                            430
          *                               *
   Type2  AATTGAGAGA GGAATTTTCT Agtaagtaaa
   Type1  .......... .......... ..........
Clone8-3  ...TNNNNNN NNNNNNNNNN Nnnnnnnnnn
```

b) Exon3 alignment

```
          1                                                         50
          *                                                          *
   Type2  ccaatgacct gacctgtgtt cacacagAGC TCCAGAACAG GACCAAGCTG
   Type1  .......... .......... ........A. .........A ..........
Clone8-3  .......... .......... nnnnnnnNNN NNNNNNNNNN NNNNNNNNNN

          51                                                        100
          *                                                          *
   Type2  ACTGTACTTG AAGGAGACAT TCTGGATGAG CCATTCCTGA AAAGAGCCTG
   Type1  ..A..G..G. .......... .......... .......... .G........
Clone8-3  NNNNTGT.G. .......... ....A..... T...G...T. ..........
```

```
          101                                                       150
          *                                                           *
   Type2  CCAGGACGTC TCGGTCGTCA TCCACACCGC CTGTATCATT GATGTCTTTG
   Type1  .......... ......A... .......... .......... ........C.
Clone8-3  .......A.. .T....A... .......T.. ..C....... ..C.....C.

          151                                                       200
          *                                                           *
   Type2  GTGTCACTCA CAGAGAGTCC ATCATGAATG TCAATGTGAA AGgtacagta
   Type1  .......... .........T .......... .......... .....tg...
Clone8-3  .......... ....C....T ........A. C......... ..........

          201                           230
          *                               *
   Type2  gcctggggag gagataaaac aagttggtt-
   Type1  .g........ .....gc.g. ...g...gg-
Clone8-3  t......... .....gc.g. ...g...g-c
```

c) Exon 4 alignment

```
          1                                                         50
          *                                                           *
   Type2  tcttcgtggg cagGTACCCA GCTACTGTTG GAGGCCTGTG TCCAAGCCAG
   Type1  .....a...a .......... ...C.....A .......... .......T..
Clone8-3  -......... .......... A..T...... ..-A...... .......T..

          51                                                        100
          *                                                  (A)      *
   Type2  TGTGCCAGTC TTCATCTACA CCAGTAGCAT AGAGGTAGCC GGGCCCAACT
   Type1  .......... .......... .......... .......... ..........
Clone8-3  .A.-....G. ......C... ....C.T... ...A...T.. A..........

          101                                                       150
          *                                                           *
   Type2  CCTACAAGGA AATCATCCAG AACGGCCACG AAGAAGAGCC TCTGGAAAAC
   Type1  .......... .......... ..T.....T. .......... ..........
Clone8-3  ...G...TGA G......... ..T....... ..T....... .T.........

          151                                                       200
          *                                                           *
   Type2  ACATGGCCCA CTCCATACCC GTACAGCAAA AAGCTTGCTG AGAAGGCTGT
   Type1  .........C .......... AC........ .......... ..........
Clone8-3  .T......TG .......... A..G...... .......... .......C..
          201                                                       250
          *                                                           *
   Type2  GCTGGCGGCT AATGGGTGGA ATCTAAAAAA TGGTGATACC TTGTACACTT
   Type1  A......... ..C....... ....G..... C..C.GC... C.........
Clone8-3  ......A... ........A. CA..G..... .A.C.-..T. .......T..

          251                                                       300
          *                                                           *
   Type2  GTGCGTTAAG ACCCACATAT ATCTATGGGG AAGGAGGCCC ATTCCTTTCT
   Type1  ....C...C. .....TG... .......... .....A.... .A.........
Clone8-3  ....C..... CT...TG... .......A.. .....A.... TA.A......
```

```
          301                                                       350
          *                                                          *
   Type2  GCCAGTATAA ATGAGGCCCT GAACAACAAT GGGATCCTGT CAAGTGTTGG
   Type1  ..T.......  .C........ .......... .......... ..........
Clone8-3  ....C..C.. .......... ATT--..... .......... .......CA.

          351                                                       400
          *                                                          *
   Type2  AAAGTTCTCT ACAGTCAACC CAGTCTATGT TGGCAACGTG GCCTGGGCCC
   Type1  .........C ..T..T.... .......... ......T... ..........
Clone8-3  C.....A..C ...-.....A .......... .....T..- ..........

          401                                                       450
          *
   Type2  ACATTCTGGC CTTGAGGGCT CTGCGGGACC CCAAGAAGGC CCCAAGTGTC
   Type1  .......... .........C ....A..... .......... .......CA..
Clone8-3  .........T ....GT.T.C ....A..... .......... ..T...CA..

          451                                                       500
                                                                    *
   Type2  CGAGGTCAAT TCTATTACAT CTCAGATGAC ACGCCTCACC AAAGCTATGA
   Type1  .....A..G. ....C..T.. .......... .......... ..........
Clone8-3  .....A..G. ....C..T.. .......... .T.T..T... ..........

          501                                                       550
          *                                                          *
   Type2  TAACCTTAAT TACATCCTGA GCAAAGACTT TGGCCTCCGC CTTGATTCCA
   Type1  .......... ....C..... .......... .......... ..........
Clone8-3  CA......G. ....C..... A......A.. .-.....T.. ..........

          551                                                       600
          *                                                          *
   Type2  GATGGAGCCT TCCTTTAACC CTGATGTACT GGATTGGCTT CCTGCTGGAA
   Type1  ........T. .......T.. .........T. .......... ..........
Clone8-3  .......... ....C..G.. ....G..... .......... ..........

          601                                                       650
          *                                                          *
   Type2  GTAGTGAGCT TCCTACTCAG CCCAATTTAC TCCCATCAAC CCCCCTTCAA
   Type1  A......... .......... G......... A......G.. .G........
Clone8-3  A.C.....A. ....G..... NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

          651                                                       700
          *                                                          *
   Type2  CCGCCACACA GTCACATTAT CAAATAGTGT GTTCACCTTC TCTTACAAGA
   Type1  ........T. ........G. .......C.. A......... .....T....
Clone8-3  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

          701                                                       750
          *                                                          *
   Type2  AGGCTCAGCG AGATCTGGCG TATAAGCCAC TCTACAGCTG GGAGGAAGCC
   Type1  .......... ....T..... .......... .......... ..........
Clone8-3  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

          751                                                       800
          *                                                          *
   Type2  AAGCAGAAAA CCGTGGAGTG GGTTGGTTCC CTTGTGGACC GGCACAAGGA
   Type1  .......... .G........ .......... .......... ..........
Clone8-3  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
```

```
          801                                                       850
          *                                                          *
Type2     GACCCTGAAG TCCAAGACTC AGTGATTTAA GGATGACAGA GATGTGCATG
Type1     ..A.....  ..........  ..TGA.....  ..........  ..........
Clone8-3  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

          851                                                       900
          *                                                          *
Type2     TGGGTATTGT TAGGAAATGT CATCAAACTC CACCCACCTG GCTTCATACA
Type1     ..........  .....G....  ......G...  .....T....  ..C.......
Clone8-3  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
```

Fig. 3.20 Sequence obtained from clone 8(3) aligned with homologous sequences from 3ß-HSD type I and II. Triplet nucleotides underlined are translational stop and start codons. Dashes indicate gaps introduced for alignment. Bases corresponding to intron sequence in type I and II are shown in lower case.

3.2.10 Long range PCR to determine the size of HSD3B intron homologues in clone 8(3)

Clone 8(3) contains sequence homologous to HSD3B introns (intron sequence was obtained from PCRs using primers A, T, E and I), therefore long range PCR was performed to determine the lengths of the intron sequence present. PCRs were attempted exactly as for clone 2(7) (section 3.2.5). Clone 8(3) yielded amplification products of approximately 4kb for intron II and 2.5kb for intron III (Fig. 3.21), which corresponds well with HSD3B1 intron lengths of 4.2kb and 2.6kb for II and III respectively.

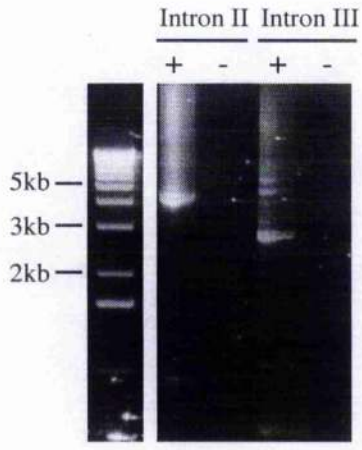


Fig. 3.21 Agarose gel of long range PCR amplification of intron II and III from clone 8(3). Primers U and H were used for intron II, G and L were used for intron III. + indicates clone 2(7) template added, - indicates no template negative contol.

3.2.11 Identification of clone 8(3) as a pseudogene

The 3ß-HSD coding sequence determined from clone 8(3) contained frameshft mutations resulting in premature stop codons. The first stop codon in the predicted polypeptide sequence is at residue 143 (nucleotide 108 in exon 4, Fig. 3.19). No other significant open reading frames were found within the sequenced fragments. Therefore, it was concluded that clone 8(3) was also an unprocessed pseudogene and it was thereafter designated 3ß-HSDψ3.

3.2.12 Searching for a human 3ß-ketosteroid reductase (KSR) enzyme

The members of the 3ß-HSD gene families in rodents fall into two functionally distinct groups. Rat I, II and IV, mouse I, II, III and VI and hamster I and II function as NAD+ dependant dehydrogenases/isomerases whereas rat III, mouse II and IV and hamster III function as NADPH dependant 3-ketosteroid reductases (KSR; section 1.8.4). These enzymes convert 5-keto-5α-androstanes into 3ß-hydroxysteroids, for example, dihydrotestosterone into 5α-androstane-3ß,17ß-diol. Human 3ß-HSD type I and II cDNA have KSR activity when expressed in tissue culture cells (Lorence *et al.*, 1990a, Rheaume *et al.*, 1991). The prefered cofactor for this is NADH, compared to NADPH, the reducing agent favoured by the rodent KSRs. KSR activity has been observed in human tissues, in particular, the prostate gland (Abalain *et al.*, 1989, Tunn *et al.*, 1990, Trapp *et al.*, 1992 and Amet *et al.*, 1992). In the human prostate dihydrotestosterone (DHT) is removed by reduction to 5α-androstane-3ß,17ß-diol and 5α-androstane-3α,17ß-diol by 3ß-HSD and 3α-HSD respectively. Trapp *et al* (1990) showed that the KSR activity in the prostate had a three-fold preference for NADPH than NADH, consistent with the KSR enzymes in the rodent and not the KSR activity of human type I and II 3ß-HSD.

To investigate the KSR activity found in the human prostate RT-PCR was utilised to establish what isoform of human 3ß-HSD (if any) was present in this tissue. Prostate cDNA was synthesised from RNA purchased from Clontech and RT-PCR was carried out using 5' primers specific to 3ß-HSD type I (type I with AA; section 2.6.2) and type II (type II with AA; section 2.6.2). Placental and testis cDNA were used as positive controls in these experiments. The RT-PCR amplification products were electrophoresed on an agarose gel, blotted and hybridised with full-length 3ß-HSD type I cDNA (Fig. 3.22). Fig. 3.22 indicates that a product of the expected size (1.3kb) was obtained from the prostate gland using type II specific primers (lane 1, type II),

while no 3ß-HSD type I transcripts were detected. The type II products were cloned into T-vector and sequenced to confirm that the sequence was 3ß-HSD type II.

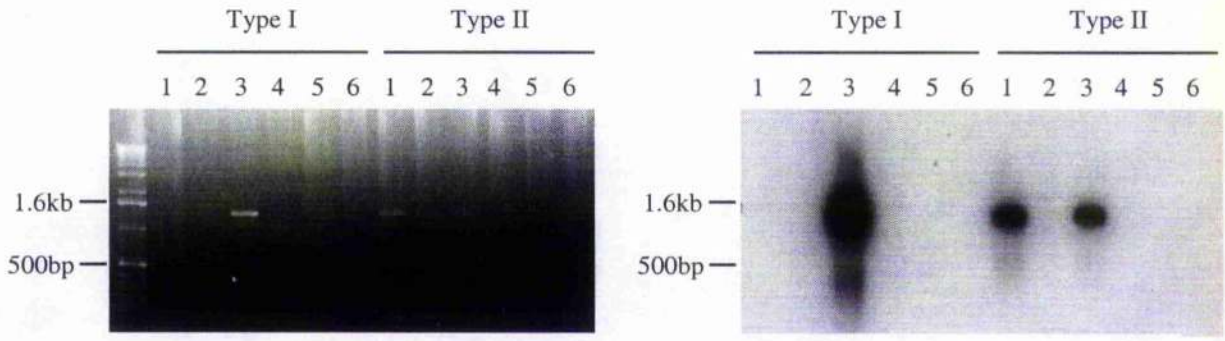


Fig. 3.22 Agarose gel of RT-PCR amplification products from the prostate gland using 3ß-HSD type I and II specific primers and hybridisation with 3ß-HSD type I cDNA. Lane 1 - prostate cDNA, lane 2 - no template control, lane 3 - placental cDNA for type I and testis cDNA for type II, lane 4 - no template control, lane 5 - no RNA control, lane 6 - no RT control.

Therefore, the human prostate gland expresses type II 3ß-HSD and this enzymes does possess KSR enzyme activity. However, it is still unclear as to whether this isoform alone is responsible for the KSR activity detected in the human prostate and the possibility of a specific KSR enzyme in humans cannot be disregarded.

3.3 CONCLUSIONS

This chapter has described the characterisation and sequence analysis of two new members of the 3ß-HSD gene family, HSD3Bψ2 and 3. Both genes contain frameshift mutations which result in premature stop codons in the predicted polypeptide sequence (HSD3Bψ2 at residue 42 and HSD3Bψ3 at residue 143) and therefore we concluded that these genes are unprocessed pseudogenes.

The screen of 2 genomic libraries identified 5 new genes in total; HSD3Bψ1 and 4 were sequenced Martin McBride, HSD3Bψ2 and 3 by myself and HSD3Bψ5 by Nicola Craig. Each gene contains frameshift mutations which result in premature stop codons in their predicted polypeptides and therefore all represent unproccessed pseudogenes. The sequence alignment of the identified 3ß-HSD genes is shown in appendix ii.

CHAPTER 4

Physical mapping of the 3ß-HSD gene family

4.1 INTRODUCTION

To date, the 3ß-HSD gene family in humans contains 7 members; 2 functional HSD3B genes and 5 non-functional pseudogenes (Chapter 3). All of these genes have been localised to chromosome 1p13 by fluorescent in situ hybridisation (FISH) by Norma Morrison from the Dept. of Medical Genetics, Yorkhill hospital (Morrison *et al*, 1995, personnal communication). To analyse the gene family further, it was decided to order the genes at chromosome 1p13 by mapping each sequence to clones from a YAC and BAC genomic library. It was possible that the construction of the 3ß-HSD gene map would provide the answers to the following questions concerning the gene family: (i) Have the mutations identified in HSD3B2 and associated with 3ß-HSD deficiency originated by gene conversion from another 3ß sequence? (ii) How are the 3ß-HSD genes related to one another with respect to the expansion of the gene family? (iii) Are the functional 3ß genes (type I and II) physically close to one another in the genome and could they have shared promoter sequences?

Dr. Jenny Varley at the Paterson Institute of Cancer Research, Manchester screened a human genomic YAC library by PCR using primers specific for HSD3B1. 3 YAC clones containing 3ß-HSD sequences were identified. In Dr. Varley's laboratory I screened a human BAC library for clones containing 3ß-HSD gene family members. The resulting YAC and BAC clones were used to construct a contig over the 3ß-HSD gene cluster. This was investigated by restriction analysis to provide the orientation of the genes with respect to each other, in particular HSD3B1 and HSD3B2, and to establish the physical distances between the genes.

4.2 RESULTS

4.2.1 Designing and testing gene-specific PCR primer pairs and oligo probes.

In order to generate a contig across the 3ß-HSD gene cluster, gene-specific PCR primer pairs and gene-specific oligonucleotide hybridisation probes were designed for each member of the family. To ensure specificity, the PCR primers were designed with at least 2 consecutive base pair differences at the 3' end of the primer. For the same reason, gene-specific oligonucleotide probes were designed with at least four base pair differences located at the centre of the oligonucleotide. As discussed in Chapter three, at the nucleotide level, the members of the gene family are very closely related and it was necessary to search the 5' untranslated region to find sufficient base pair differences. From this area, gene-specific PCR primers were successfully designed

for HSD3B1, 2, ψ2 and ψ5 and oligonucleotide probes were designed for HSD3B 1, 2, ψ1, ψ2 and ψ5. However, the 5' untranslated region did not yield PCR primer sites for HSD3Bψ1, ψ3 and ψ4. Consequently, gene-specific primers were designed to HSD3Bψ1 exon 3 and to exon 4 of HSD3Bψ3 and HSD3Bψ4. A specific oligonucleotide probe for HSD3Bψ3 was designed in exon 4 and no region was identified within known sequence to create a specific probe for HSD3Bψ4 (Fig 4.1A and B and the sequence of the oligonucleotides can be found in section 2.6.2 and 2.8.5). Gene-specific PCR primer pairs were tested for specificity using human genomic DNA as template and the resulting products were sequenced (Fig. 4.1C). Initially, the type II primer pair did not amplify from genomic DNA, however this was successful when repeated (Fig. 4.1D).

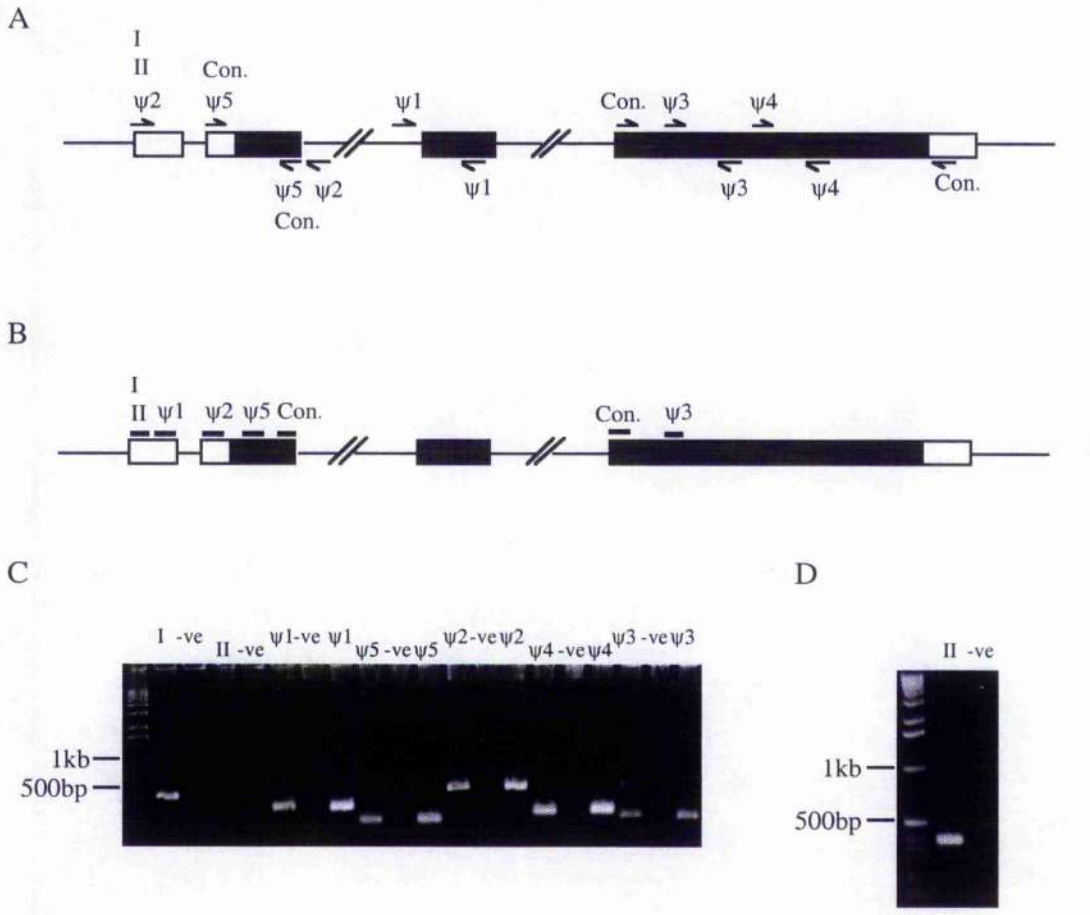


Fig. 4.1 A: Schematic diagram of HSD3B indicating the approximate positions of the gene-specific primers (Con - conserved primer). Not to scale.
B: Schematic diagram of HSD3B indicating the approximate positions of the gene-specific oligonucleotide probes (Con - conserved primer). Not to scale.
C: Agarose gel of test PCRs using gene-specific primers with human genomic DNA template. These products were sequenced to confirm specificity.
D: Agarose gel of test PCR using type II specific primers with human genomic DNA template.

The gene-specific hybridisation probes were tested against the seven members of the HSD3B gene family immobilised on nylon membrane. The original phage clone that genes HSD3Bψ1-5 were identified from were used as control DNA, as well as HSD3B1 and 2 full length RT-PCR products from the placenta and testis respectively (see section 3.2.12). Unfortunately, the HSD3ψ1 specific probe did not anneal to HSD3Bψ1 DNA, and the HSD3Bψ3 probe was not specific for HSD3Bψ3 DNA. Therefore, only probes specific for HSD3B1, 2, ψ2 and ψ5 were used in subsequent experiments (Fig. 4.2).

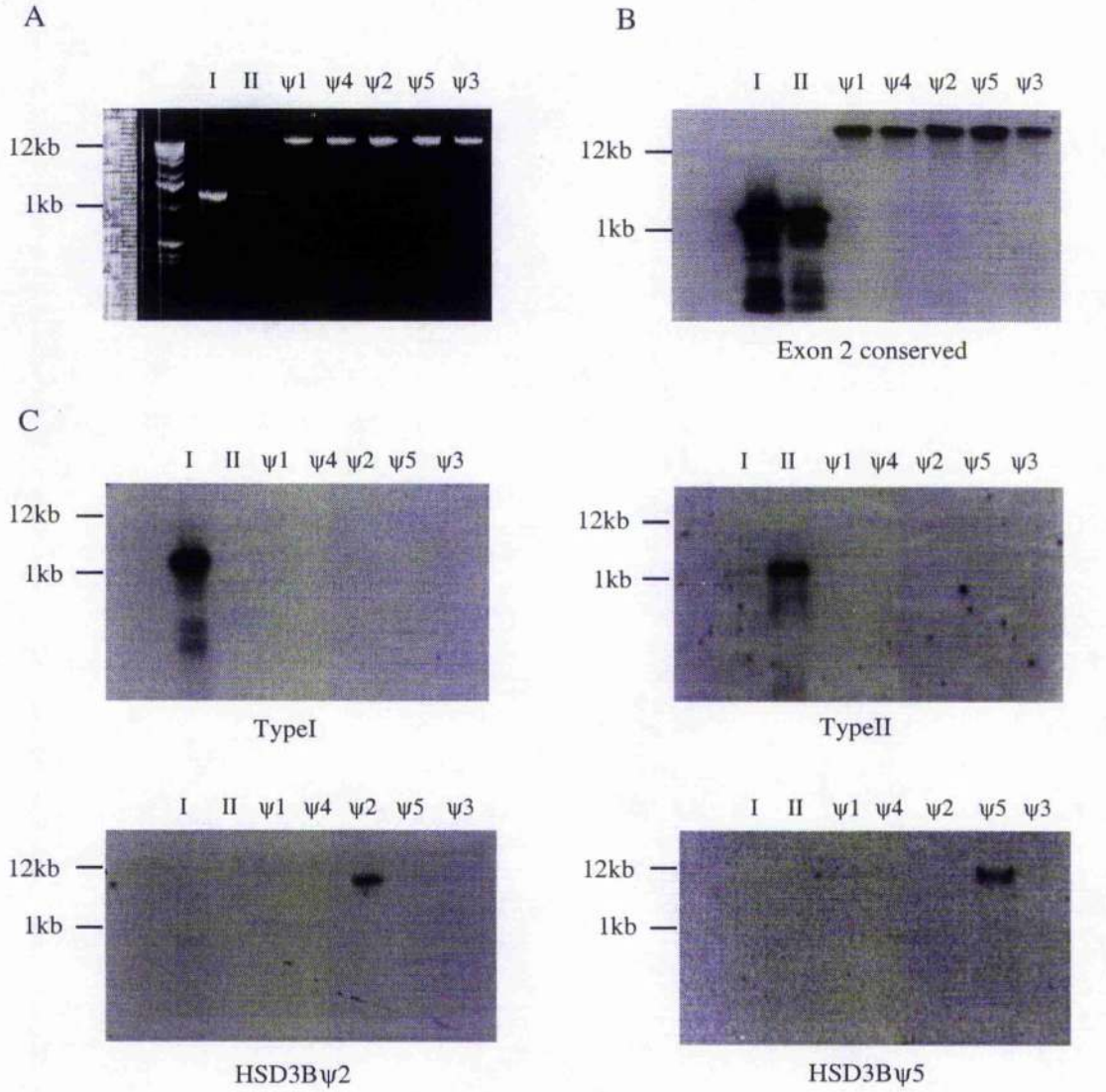


Fig. 4.2 A: 1% agarose gel containing uncut control DNA to test the gene-specific oligonucleotide hybridisation probes. Type I and type II control DNA were full-length RT-PCR products, see section 3.2.11. The phage clone used to identify each pseudogene was used for ψ1 - ψ5.

B: Southern analysis of panel A hybridised with oligonucleotide probe to a conserved region of exon 2 in all 7 family members.

C: Southern analysis of panel A hybridised with gene-specific oligonucleotide probes for HSD3B1, 2, ψ2 and ψ5.

4.2.2 Screen of human YAC library for HSD3B genomic clones.

The YAC library screen was undertaken by Bill Brintnell in Dr. Jenny Varley's laboratory, Manchester. A human YAC library (Zeneca) was screened by PCR with primers to 3ß-HSD (primers 93096 and 93097; section 2.6.2). Three clones were purified: 9GF11, 32EG4, and 1GD11, with genomic inserts of 250, 200, and 220 kb, respectively. YAC DNA from these clones was donated to our laboratory in Glasgow, where each clone was screened with gene-specific PCR primer pairs and a preliminary YAC map was constructed. The PCRs were repeated at least three times and the resulting products were sequenced to check specificity. The PCR products obtained and the YAC map generated is shown in Fig. 4.3. The map contains 2 sets of genes within which the order cannot be deduced, one set contains II and ψ2 and the other I, ψ1, ψ3 and ψ4. It has been established that ψ5 is found at the far left side of the cluster shown, this gene is present on YAC 9GF11 but not YAC 32EG4. The size of the gap between the 2 sets of unordered genes has not been determined.

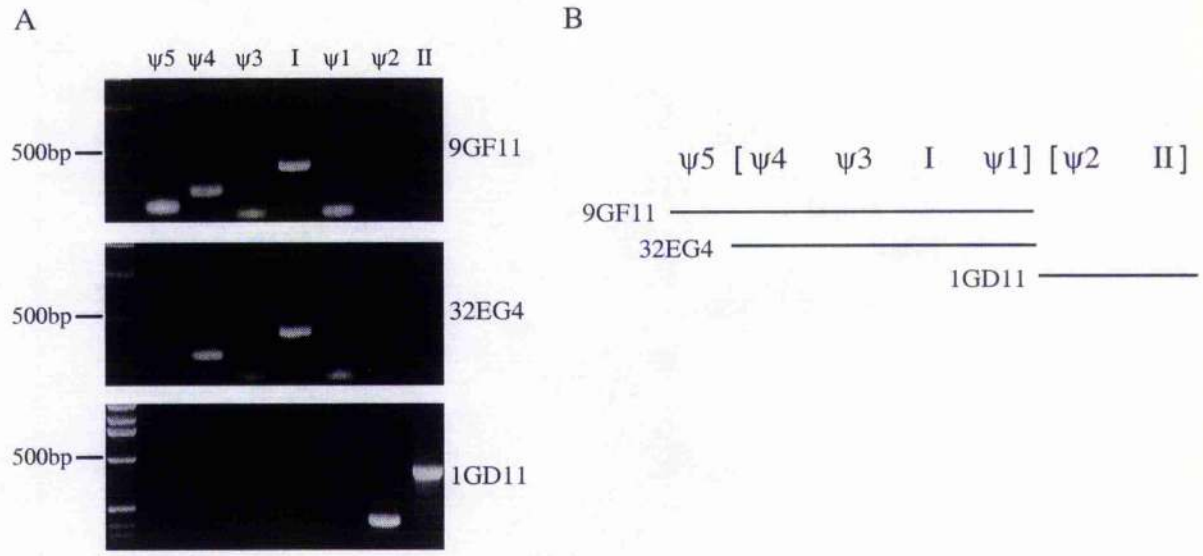


Fig. 4.3 A: Agarose gel of PCR products fromYAC clones 9GF11, 32EG4 and 1GD11 screened using gene-specific primer pairs (Fig. 4.1A).
B: A preliminary genomic map constructed from the YAC PCR results shown in A.

4.2.3 Screen of human BAC library for HSD3B genomic clones.

A human BAC library (Research Genetics Inc.) was screened with primers to sequence that is fully conserved between the seven members of the gene family (primer pairs C-S and K-AA). The BAC library was provided on 192 384-well microtitre plates. Each plate was pooled to produce 192 subpools which were further pooled into 24

superpools. To screen the BAC library, the superpools were screened first, which provided an address of the subpool to screen next, and in turn, gave the address of the plate to be screened. The plate was screened by pooling the rows together and the columns together which led to a row letter and column number address of the positive clone. A representative clone identified from the BAC library screen is shown in Fig. 4.4.

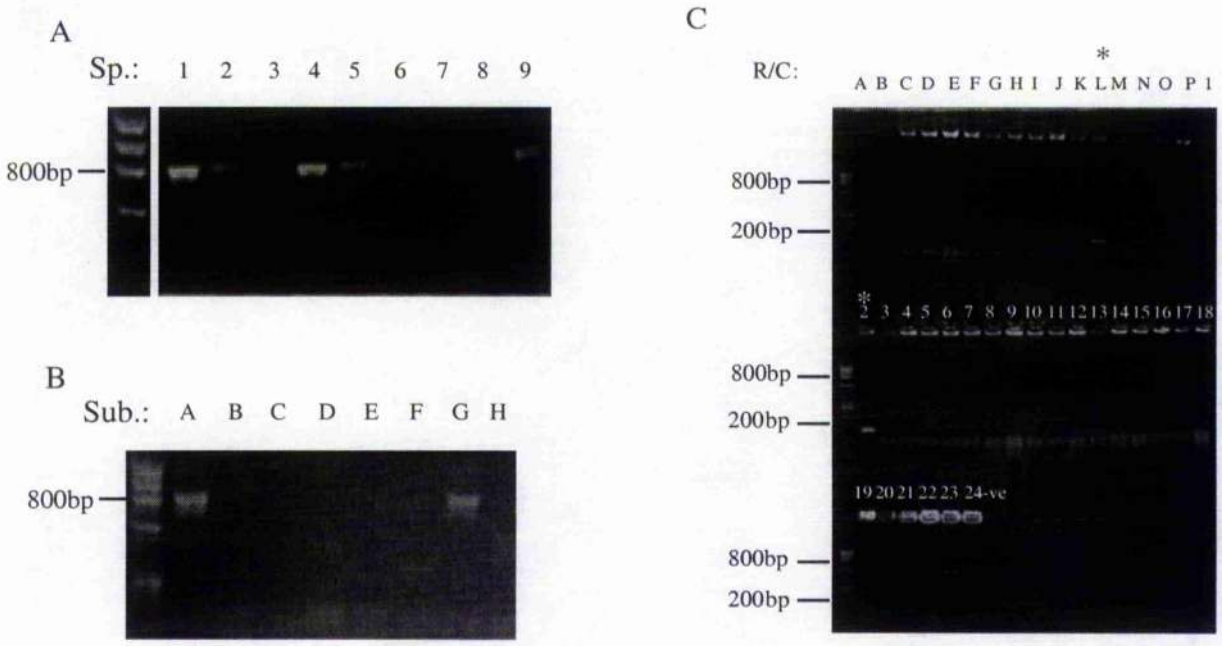
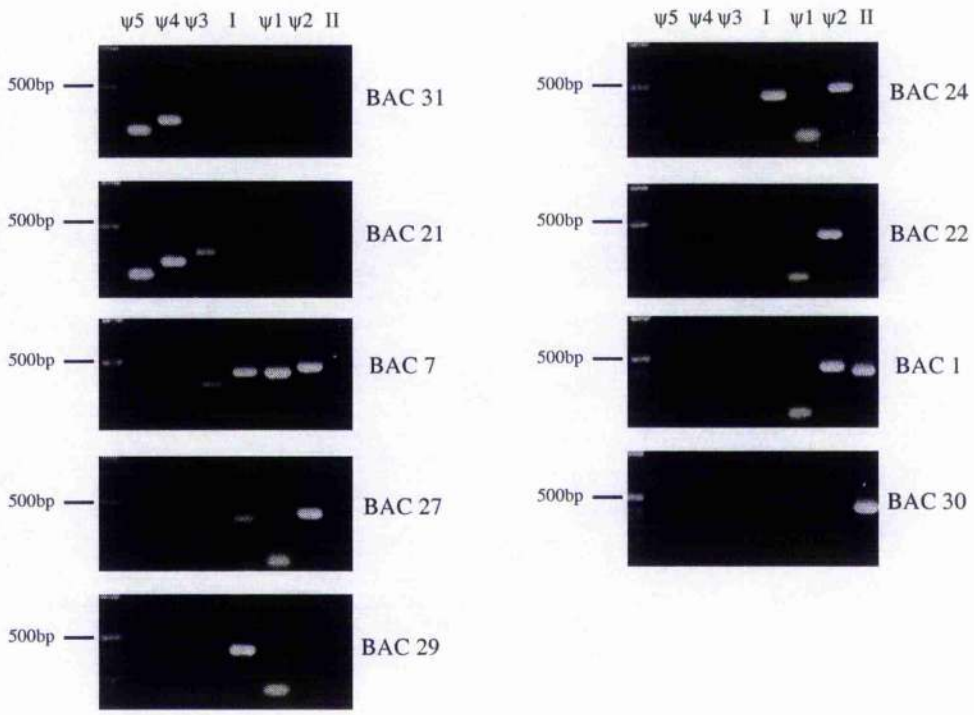


Fig. 4.4 Agarose gels indicating the BAC library screen strategy and the identification of clone L-2
A: Agarose gel of PCR screen of superpools (Sp) 1-9 of BAC library, using primer pair K-AA. Expected product length 830bp.
B: Agarose gel of PCR screen of subpools (Sub) A-H from superpool 1, using primer pair K-AA. Expected product length 830bp.
C: Agarose gel of PCR screen of rows and columns (R/C) from plate 7 which corresponds to subpool G, primer pair used C-S. Expected product length 160bp. Clone L-2 is identified and is shown by *.

32 BAC clones were identified from the library screen. These were tested with the gene-specific primers and 9 of the 32 were shown to contain the 7 members of the gene family. These BACs were used to construct a BAC map (Fig. 4.5). The BAC PCRs were repeated at least 3 times and the products were sequenced to ensure specificity. The BAC PCR data allows the genes within the 2 sets of unordered genes from the YAC PCR data to be ordered (Fig. 4.3B). The only genes that cannot be ordered from the BAC map are ψ4 and ψ5.

A



B

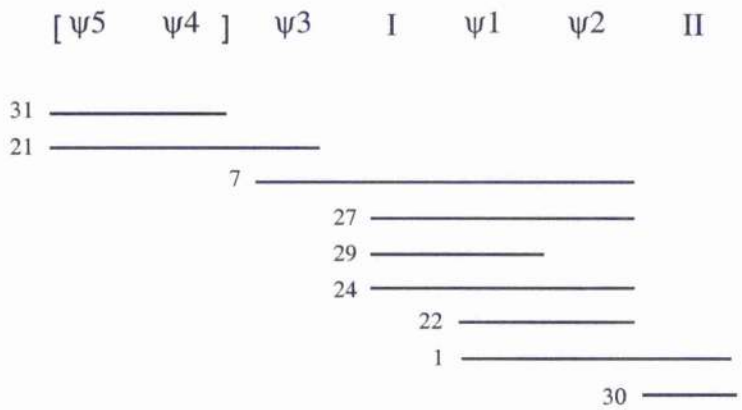


Fig. 4.5 A: Agarose gels of the PCR products from 9 BACs positive with gene-specific primer pairs. Gene-specifc primer pairs as indicated on Fig. 4.1A, except PCR amplification for HSD3Bψ1 using BAC clone 7 template. The PCR primer pair used in this amplification was 5' ψ1 specific with primer I.
B: A genomic map constructed from the BAC PCR results shown in A.

From the other 23 BACs identified, 16 resulted from contamination of the library at source and represented the same clone. The remaining 7 were rescreened with the conserved PCR primers. Some failed to reamplify, which was not unexpected as some of the initial PCR products were extremely faint but were investigated to ensure that nothing was missed. Amplification products were obtained from 3 clones but attempts

to sequence them were unsuccessful. From the gene-specific PCR screen of the BAC clones, the main objective of these experiments was achieved, as a genomic contig was constructed (Fig. 4.5) and because the above observations did not provide any evidence of any other HSD3B genes, these clones were not investigated further.

The BAC data provided the order of all the genes except ψ4 and ψ5, but the YAC data revealed that HSD3Bψ5 was present at the left end of the cluster. Therefore, a combination of the YAC and BAC PCR results provided a genomic contig indicating the gene order of the 7 identified members of the 3ß-HSD gene family (Fig. 4.6).

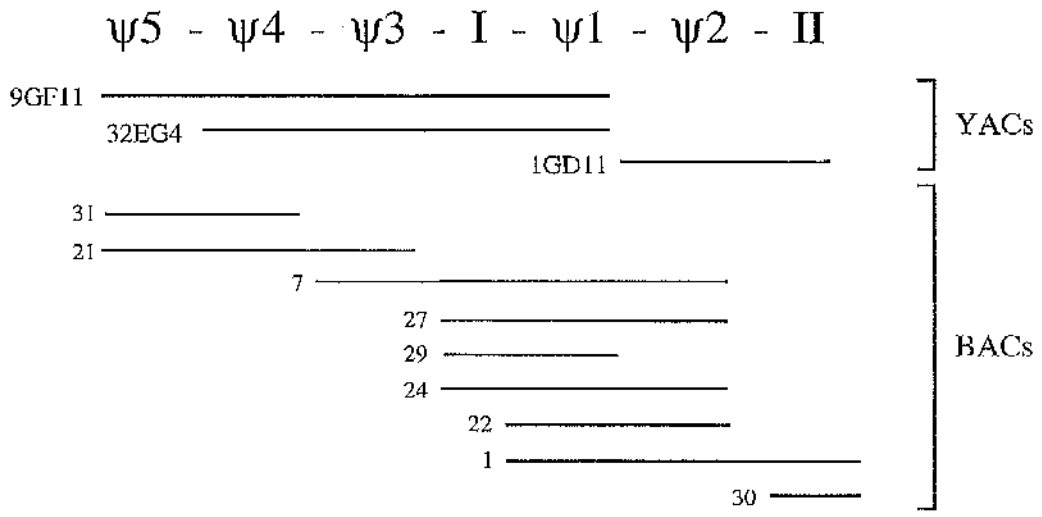


Fig. 4.6 Order of the 3ß-HSD gene family at human chromosome 1p13, deduced from the YAC and BAC contigs. Size of YAC clone genomic inserts are 250kb, 200kb and 220kb for 9GF11, 32EG4 and 1GD11 respectively, therefore the maximum size of the contig is 470kb.

4.2.4 Confirmation of the gene-order across the human 3ß-HSD gene cluster

To confirm the gene order obtained above, the DNA from BAC clones 31, 21, 7, 29, 1 and 30 was isolated. These 6 clones were chosen because together they span the cluster and contain all 7 genes, and the remaining 3 BAC clones are contained within one or more of them (Fig. 4.6). BAC clone DNA was purified using a modified protocol from Qiagen. BAC DNA was manipulated as plasmid DNA, although to prevent shearing of the large BAC molecule the ends were removed from pipette tips. Each BAC clone was digested with BamHI, HindIII and EcoRV and electrophoresed on a

0.7% agarose gel. The gels were blotted as described (section 2.7) with an additional wash with 0.25M HCl to depurinate the larger fragments for ease of transfer. The membranes were probed with the conserved exon 2 oligonucleotide and conserved exon 4 oligonucleotide (Fig. 4.1B).

The HindIII digest probed with conserved exon 2 oligonucleotide yielded a clear pattern which was further investigated.(Fig. 4.7A). From this blot, the number of exon 2 bands present on each BAC and also what gene each band represents could be identified. BAC 30 only contains HSD3B2 (PCR data, Fig. 4.5A), therefore the band observed in this digest must represent HSD3B2 exon 2. This fragment is present in the BAC 1 digest and BAC 1 also contains HSD3Bψ2 and HSD3Bψ1 (PCR data, Fig. 4.5A). HSD3Bψ1 is present in BACs 1, 29 and 7 and therefore HSD3Bψ1 exon 2 can be identified as the 600bp fragment which is common to each of these BACs. HSD3Bψ2 is present on BACs 1 and 7, therefore it is deduced that the 2.8kb fragment represents HSD3Bψ2 exon 2. Moreover, HSD3B1 is found on BACs 29 and 7, therefore if the 600bp fragment from BAC 29 is HSD3Bψ1 exon 2 then the 700bp fragment must be HSD3B1 exon 2. This can also be seen in BAC 1. To confirm these deductions, the membrane was stripped of the conserved exon 2 oligonucleotide and reprobed with HSD3Bψ2 and HSD3B1 specific oligonucleotides, which were designed to exon 2 in both these genes. These hybridisations confirmed that HSD3Bψ2 exon 2 was present on the 2.8kb fragment and that HSD3B1 exon 2 was on the 700bp fragment. (Fig. 4.7B).

BAC 31 contains HSD3Bψ4 and ψ5 and it appears as if the fragments resolve at the same size, which is not surprising considering their sequence similarity. BAC 21 contains these 2 genes also and the 5kb fragment can be observed. BAC 21 also contains HSD3Bψ3 which may be represented by the fragment at 700bp or the 5kb fragment. The 700bp HindIII fragment (indicated as * on Fig. 4.7A) was subcloned into pUC 18 and sequenced to determine if it contained HSD3Bψ3 exon 2. The sequence of the 700bp band was found to be HSD3Bψ3 exon 2. This indicated that HSD3Bψ3 exon 2 HindIII fragment is the same size as the HSD3B1 fragment (seen in BACs 7 and 29) leaving the question, what does the 5kb fragment in BAC 7 contain? This fragment (indicated + on Fig. 4.7A) was also subcloned into pUC and sequenced. The sequence obtained from band + was HSD3Bψ4 exon 2. This confirms the earlier deduction that HSD3Bψ4 is present on the 5kb fragment from BACs 31 and 21, and it reveals some information on the orientation of the genes within the cluster which will be discussed further in section 4.3.7.

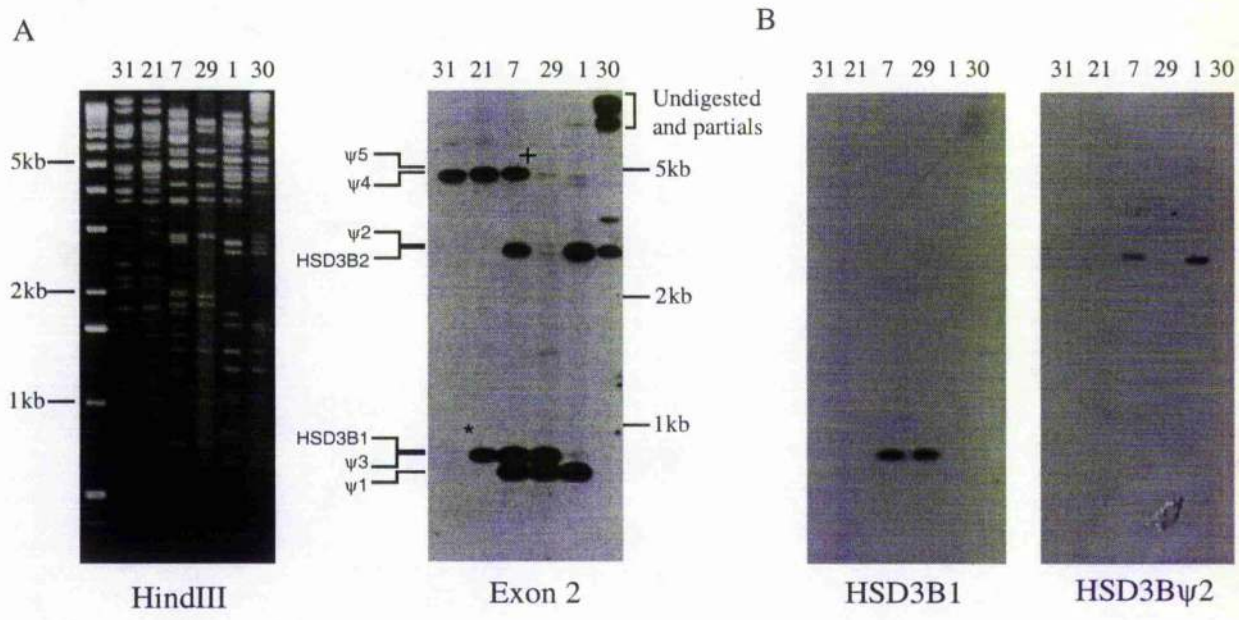


Fig. 4.7 A: Agarose gel of Hind III restriction digests of BACs 31, 21, 7, 29, 1 and 30 and Southern blot hybridised with exon 2 conserved probe. * and + indicate bands that have been subcloned and sequenced.
B: Southern analysis of HindIII digests hybridised with HSD3B1 and HSD3Bψ2 specific probes.

The PCR data indicated that BAC 21 contained the HSD3B genes HSD3Bψ3, 4 and 5 and that BAC 31 contained HSD3Bψ4 and ψ5 (Fig. 4.5). The HindIII digest for BAC 21 probed with conserved exon 2 oligonucleotide only revealed two bands, one at 700bp containing HSD3Bψ3 exon 2 and the other at 5kb containing HSD3Bψ4 and ψ5 exon 2 and the BAC 31 digest revealed one band at 5kb containing HSD3Bψ4 and ψ5 exon 2 (Fig. 4.7A). From this data HSD3Bψ4 and 5 could not be separated. To confirm that ψ4 and ψ5 were both present at this end of the cluster, BAC 21 was digested with a number of restriction digests and hybridised with the conserved exon 2 and exon 4 oligonucleotides. Restriction enzymes XbaI and EcoRV produced 3 bands when hybridised with the conserved exon 2 oligonucleotide (Fig. 4.8A). This membrane was stripped and reprobed with the HSD3Bψ3 and HSD3Bψ5 gene-specific oligonucleotides (Fig. 4.8B), indicating which fragment each gene belonged to. This confirmed that the 3 genes were present on BAC 21, however HSD3Bψ4 and ψ5 could not be separated in BAC 31 (see Fig. 4.9, only one band can be detected from digests hybridised with either conserved exon 2 or exon 4 oligonucleotides).

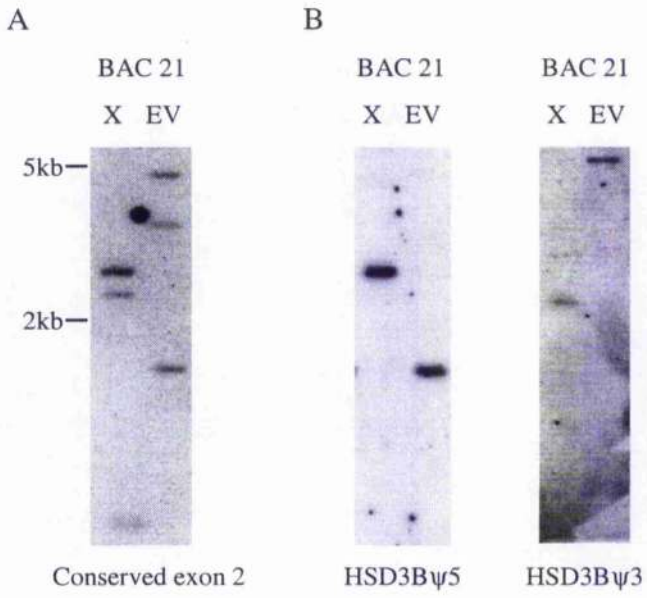
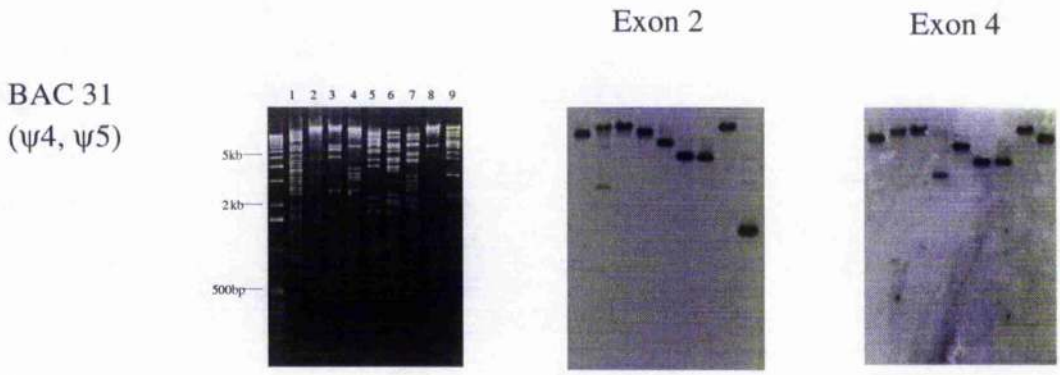


Fig. 4.8 A: Southern analysis of XbaI and EcoRV restriction digests of BAC 21 hybridised with conserved exon 2 oligonucleotide (X=XbaI, EV=EcoRV).
B: Southern analysis of XbaI and EcoRV restriction digests of BAC 21 hybridised with HSD3Bψ5 and ψ3 specific oligonucleotides (X=XbaI, EV=EcoRV).

4.2.5 Searching for unidentified 3ß-HSD genes in the gene cluster

To determine if there were any unidentified 3ß-HSD genes within the cluster, each BAC was digested with a panel of restriction enzymes and probed with the conserved exon 2 and exon 4 oligonucleotide probes. These oligonucleotidess were designed to regions conserved in all 7 identified genes, with the aim that similar genes present in the contig would be detected. From the 6 BACs examined no additional bands were detected. Each BAC had the correct number of bands, or less, predicted from the YAC and BAC contig (Fig. 4.9).



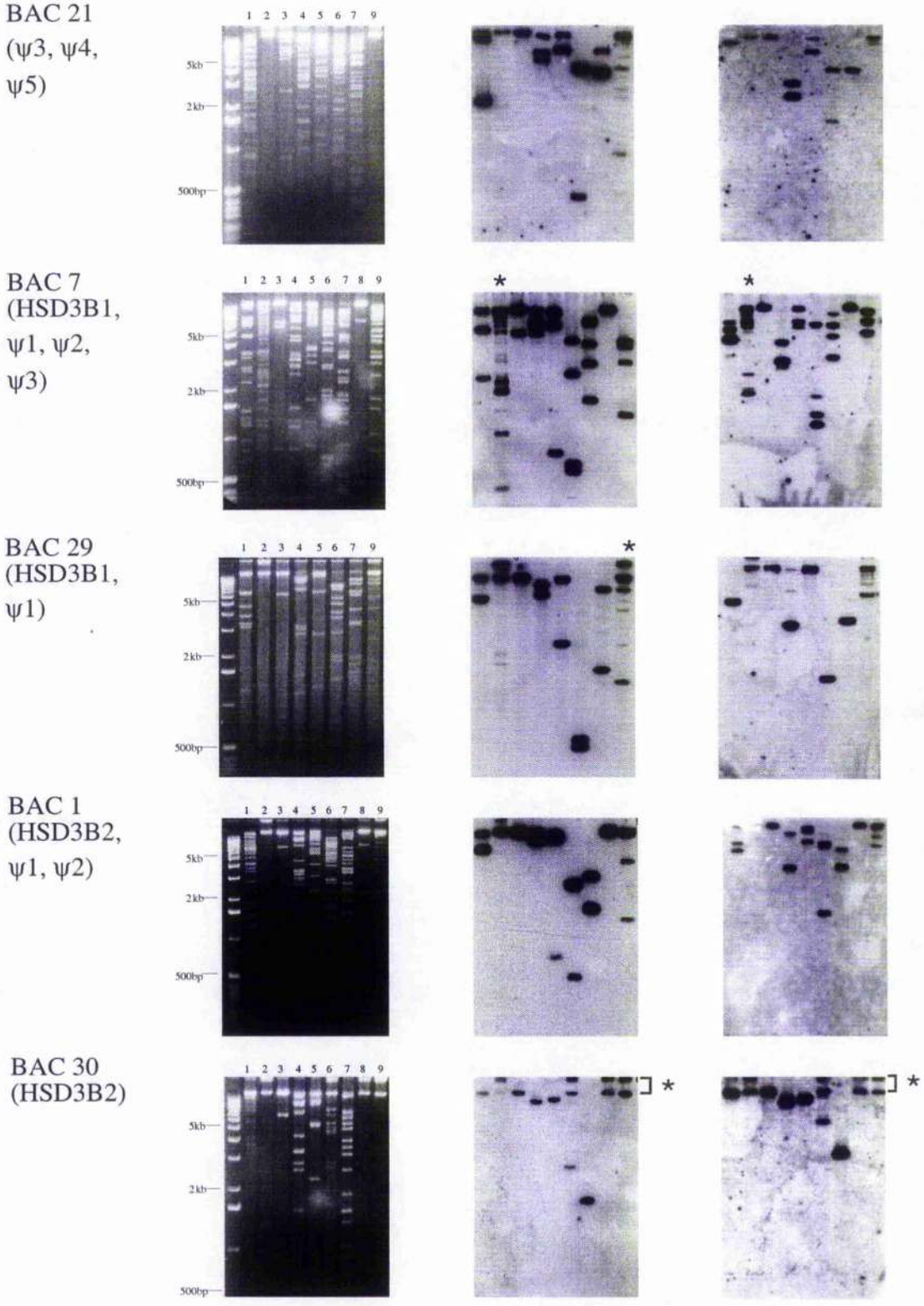


Fig. 4.9 Restriction digests of 6 BACS (left column) probed with exon 2 (central column) and exon 4 (right column) conserved probes to indicate the number of exons present in each BAC. Lane1 - EcoRI; lane2 - XbaI; lane3 - SalI; lane4 - KpnI; lane5 - BamHI; lane6 - HindIII, lane7 - SstI; lane8 - NotI; lane9 - EcoRV. * indicates partial digestions.

4.2.6 Restriction analysis of YAC 1GD11

The YAC and BAC genomic contig suggests that YAC 1GD11 contains HSD3B genes HSD3Bψ2 and 2. The size of this YAC clone is approximately 220kb (Jenny Varley, personnal communication) and it is situated at one end of the contig, therefore, it is conceivable that it may contain more than 2 3ß-HSD sequences. To investigate this possibility, YAC digests were resolved by PFGE and then probed with 3ß-HSD type I cDNA. Agarose plugs containing 1GD11 were prepared and digested as described in section 2.11. The YAC was digested by four enzymes, NotI, SmaI, XbaI and SacII, and the digestion reactions were stopped at various time points to create partial digestions. The running conditions for the digests on PFGE were 20-80 seconds for 20 hours at 200V at 14°C. After electrophoresis the gel was blotted and hybridised with 3ß-HSD type I cDNA as before. The resulting blot (Fig. 4.10) indicated that 3ß-HSD sequences were detectable on YAC clone 1GD11 by Southern analysis, however the limited restriction information obtained does not confirm that two 3ß-HSD genes are present on this clone. Additional restriction analysis and the use of gene-specific probes would be neccessary to verify this.

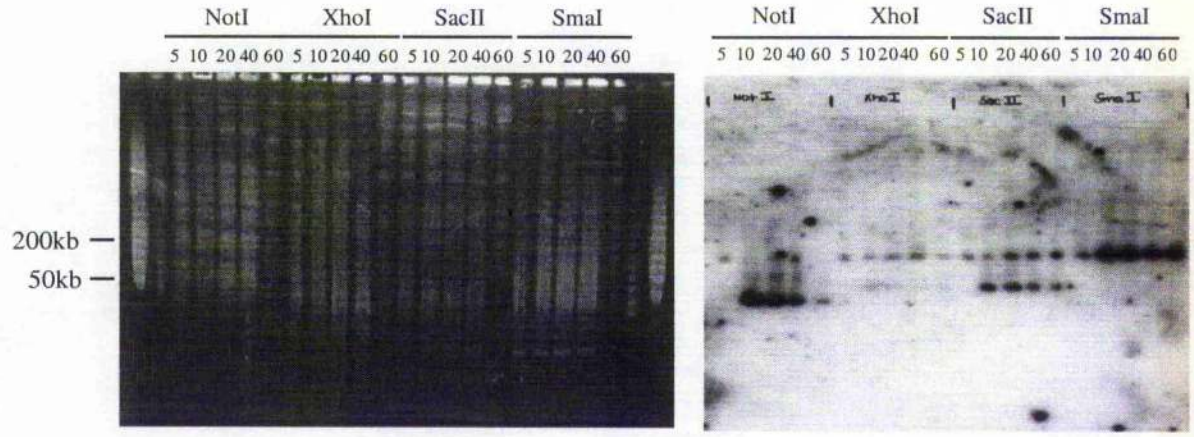


Fig. 4.10 A: PFGE of YAC 1GD11 restriction digests. Four restriction enzymes were used (NotI, XhoI, SacII and SmaI) and partials digests were accomplished at 5 time points (5, 10, 20, 40 and 60 minutes). Size markers are lambda concatamers and yeast chromosome marker. B: Southern hybridisation of PFGE (panel A) with 3ß-HSD type I cDNA.

4.2.7 Defining the physical map of the 3ß-HSD gene cluster

To investigate the 3ß-HSD contig further, restriction digests of the BAC clones were resolved by pulse field gel electrophoresis (PFGE). A detailed map was constructed using conserved oligonucleotide probes to HSD3B exon 2 and exon 4 and the BAC vector arms, T7 and SP6. The final map is presented in Fig. 4.11 and the data that supports it is analysed in the following two sections.

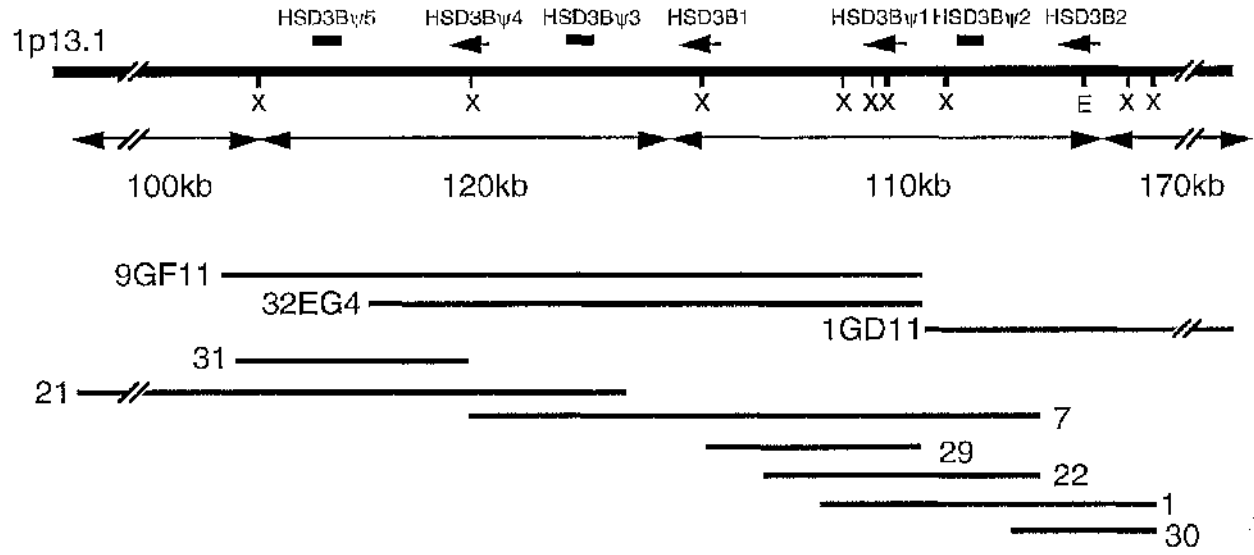


Fig. 4.11 Genomic map of 3ß-HSD gene family at human chromosome 1p13.1, constructed by restriction analysis of overlapping YAC (9GF11, 32EG4 and 1GD11) and BAC clones (31, 21, 7, 29, 22, 1 and 30). X=XhoI, E=EcoRV.

Exon 2 and exon 4 oligonucleotide probes conserved in all 7 members of the gene family (Fig. 4.1) and oligonucleotide probes to the BAC vector arms were hybridised to restriction fragments from each BAC to determine which exons (if any) were present at the ends of the genomic inserts. The pBeloBAC11 vector restriction map is shown in Fig. 4.12. Xho I, EcoRV and NotI cut the BAC vector, releasing the vector arms with the ends of the insert. This information was used to construct the majority of the genomic map. Each blot was handled and analysed with the utmost care, to ensure that all bands identified could be aligned correctly to fragments of the digest (section 2.7). Photo scanning and printing has reproduced the blots quite well, however, the original gels are available for inspection.

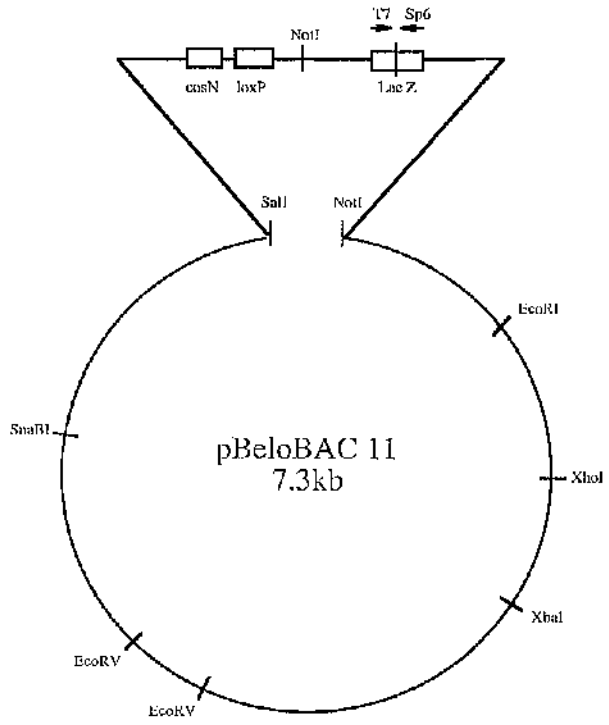


Fig. 4.12 Restriction map of BAC vector - pBeloBAC 11.

RIGHT SIDE OF MAP:

This section is concerned with HSD3B genes HSD3B2, HSD3Bψ2, HSD3Bψ1 and HSD3B1. The data is reasonably complicated and therefore a brief summary of the results which provided the orientation of the genes or helped create the restriction map of the genomic contig is provided below for reference.

GENE:	COMMENTS:
HSD3B2	The orientation of this gene was determined by restriction analysis of BAC clones 30 and 1. An XhoI restriction map was determined for BAC 30 and an EcoRV restriction site orientated HSD3B2 with respect to the BAC vector arms (Fig. 4.12E).
HSD3Bψ2	This gene was not orientated due to the lack of informative restriction sites within the gene itself. However, it was incorporated into an XhoI restriction map of BACs 30, 1, 7 and 22 (Fig. 4.13D).

GENE:	COMMENTS:
HSD3Bψ1	XhoI restriction digests of BAC 29 determined the orientation of this gene. An XhoI restriction site was present between exon 2 and 4 in HSD3Bψ1, and it was established that exon 2 was situated adjacent to one of the BAC vector arms (Fig. 4.14B).
HSD3B1	HSD3B1 was orientated from the position of the breakpoint of BAC clone 29. It was demonstrated that exon 2 but not exon 4 was present in the genomic insert of BAC 29 (Fig. 4.15B)

Orientation and restriction mapping of HSD3B2:

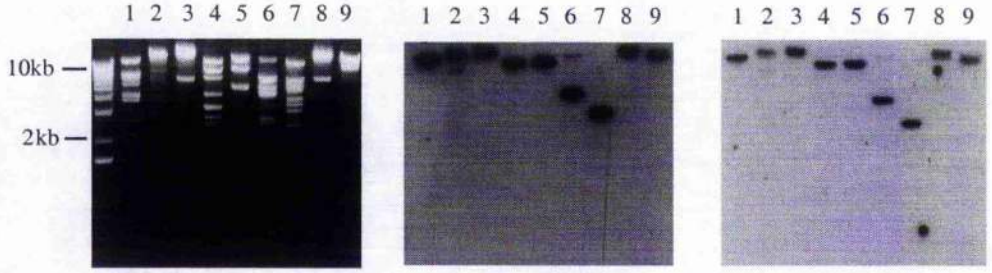
HSD3B2 is found on BAC 30 and BAC 1. Earlier gene-specific PCR data (Fig. 4.5) showed that BAC 30 contains only HSD3B2, and no additional bands were detected by hybridising BAC 30 restriction digests with conserved exon 2 and exon 4 oligonucleotide probes (Fig. 4.9). Further confirmation of this was obtained by probing the same digests with the HSD3B2 specific oligonucleotide probe (Fig. 4.13A), which hybridised to the same fragments as the conserved exon 2 oligonucleotide probe. In addition, HindIII restriction digests of BAC clones 30 and 1 showed that all BAC 30 HindIII fragments were in common with BAC 1 HindIII restriction fragments, indicating that BAC 30 is internal to BAC 1 (Fig. 4.13B).

XhoI restriction fragments for BAC 30 and BAC 1 were resolved by pulse field gel electrophoresis (PFGE), blotted and hybridised with the conserved exon 2, conserved exon 4, T7 and SP6 oligonucleotide probes (Fig. 4.13C). A 50kb fragment from BAC 30 contained HSD3B2 exon 2, HSD3B2 exon 4 (no other HSD3B genes identified on this BAC) and SP6. EcoRV digests of BAC 30 and BAC 1, also probed with conserved exon 2, conserved exon 4, T7 and SP6 oligonucleotides, indicated that exon 4 and SP6 were contained on a 9.5kb fragment from BAC 30 (Fig. 4.13D). The EcoRV digest of BAC 30 produced 6 bands, including 4 bands clustered between approximately 9.5kb to 20kb. The conserved exon 4 and the SP6 oligonucleotide both hybridised to the smallest band (9.5kb) within this cluster, determined by careful measurements during the blotting procedure (section 2.7). This information indicated that HSD3B2 exon 4 was closest the SP6 end of the vector within a 9.5kb fragment.

XhoI restriction digests of BAC 30 produced 3 fragments, 2 of which were in common with XhoI fragments from BAC 1. In particular, both BAC clones had an XhoI fragment, approximately 10.5kb, which hybridised to the T7 oligonucleotide probe.

This suggested that at the T7 end of the vector the genomic inserts of these two BAC clones were created at the same restriction site. The other common band was approximately 7.5kb, leaving a 50 kb fragment from BAC 30 which was equivalent to the 70kb fragment in BAC 1 (since BAC 30 is contained within BAC 1). This is confirmed by the hybridisation of the conserved exon 2 and exon 4 oligonucleotide probes to the 70 kb fragment from BAC 1. Other bands are observed from BAC 1 using these oligonucleotide probes, these are important with respect to HSD3Bψ2 and HSD3Bψ1 and are discussed further on. The XhoI restriction fragments for BAC clone 1 and 30 were used to generate a restriction map for BAC 30 and the left end of BAC 1 (Fig. 4.13E). The EcoRV data for BAC 30 provided an indication of the orientation of HSD3B2. The EcoRV site between HSD3B2 exon 2 and 4 is confirmed in BAC 1. The conserved exon 2 oligonucleotide hybridised to a 14kb fragment which is the same in BAC 30, and the conserved exon 4 oligonucleotide hybridised to a 16kb fragment (these bands are faint on the blot shown), which must correspond to HSD3B2 exon 4 as it is the only band larger than 9.5kb.

A

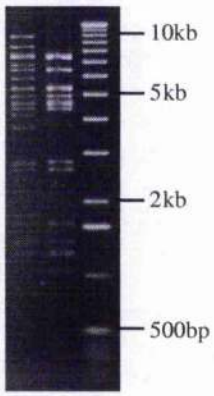


Probe: Exon 2

Probe: HSD3B2

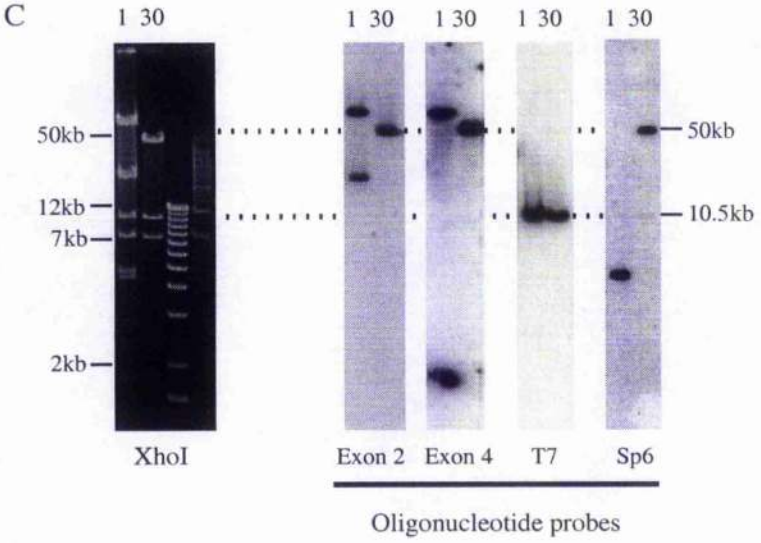
BAC 30

B

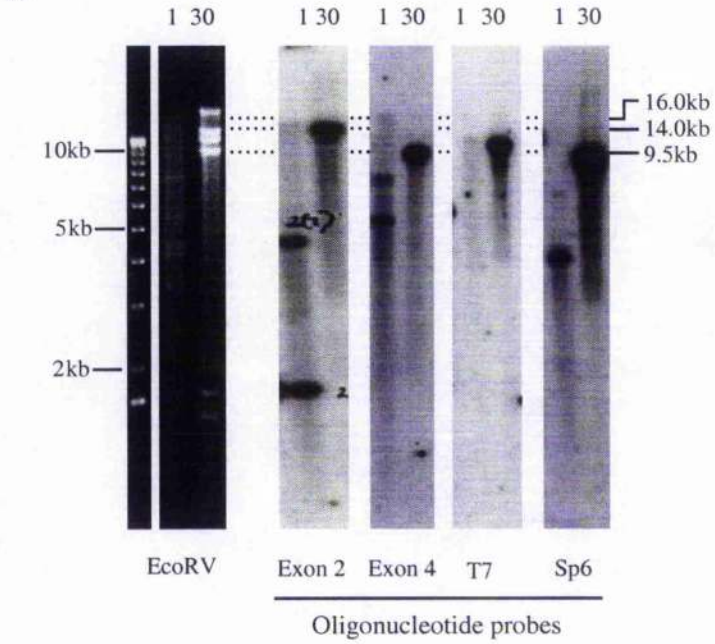


HindIII

C


1 30
1 30
1 30
1 30
1 30
50kb
12kb
7kb
2kb
50kb
10.5kb
XhoI
Exon 2
Exon 4
T7
Sp6
Oligonucleotide probes

D


1 30
1 30
1 30
1 30
1 30
10kb
5kb
2kb
16.0kb
14.0kb
9.5kb
EcoRV
Exon 2
Exon 4
T7
Sp6
Oligonucleotide probes

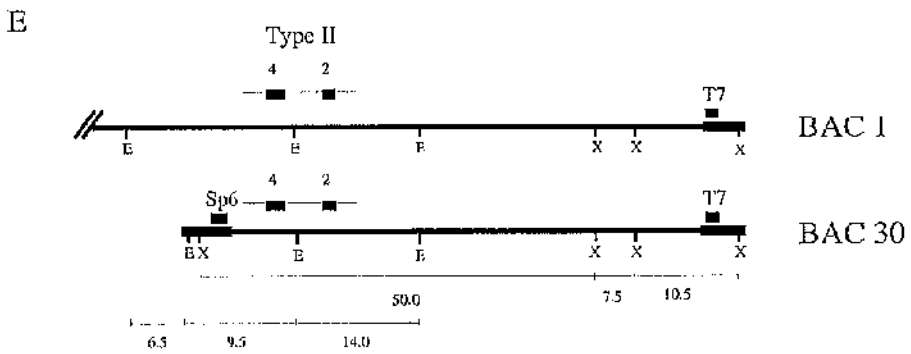


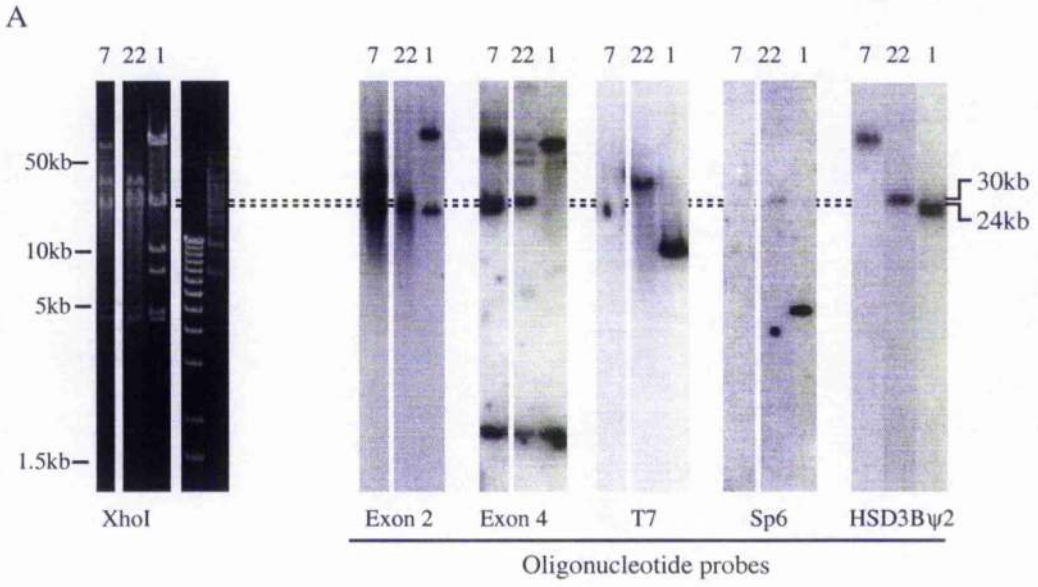
Fig. 4.13 A: Agarose gel indicating restriction digests of BAC 30 and hybridised with exon 2 conserved oligo and HSD3B2 specific oligo. Lane1 - EcoRI; lane2 - XbaI; lane3 - SalI; lane4 - KpnI; lane5 - BamHI; lane6 - HindIII; lane7 - SstI; lane8 - NotI; lane9 - EcoRV.
B: Agarose gel showing Hind III restriction digests of BACs 1 and 30.
C: XhoI restriction digests of BACs 1 and 30, resolved by PFGE and hybridised with conserved exon 2, conserved exon 4, T7 and Sp6 oligonucleotides.
D: EcoRV restriction digests of BACs 1 and 30, resolved by PFGE and hybridised with conserved exon 2, conserved exon 4, T7 and Sp6 oligonucleotides.
E: XhoI and EcoRV restriction map showing the orientation of HSD3B2.

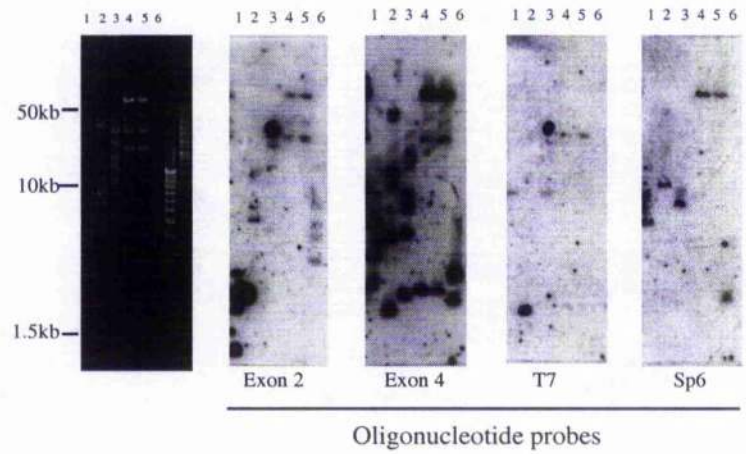
Orientation and restriction mapping of HSD3Bψ2:

HSD3Bψ2 is the next gene in the cluster and gene-specific PCR data (Fig. 4.5) indicates that it is present on BAC clones 1, 22 and 7. XhoI restriction digests were resolved by PFGE and probed with conserved exon 2, conserved exon 4, T7, SP6 and HSD3Bψ2 specific oligonucleotides. HSD3Bψ2 was found within BACs 7, 22 and 1 on fragments of 24kb, 30kb and 70kb respectively (Fig. 4.14A). The 24kb fragment from BAC 7 contained exon 2, exon 4 and the T7 vector end, the 30kb fragment from BAC 22 contained exon 2, exon 4 and the SP6 vector end and the 70kb fragment from BAC 1 contained only exon 2 and exon 4. Other fragments from BACs 1, 22 and 7 hybridised with the conserved exon 2 and exon 4 probes; these have already been discussed in relation to HSD3B2 or will be discussed in relation to the other HSD3B genes.

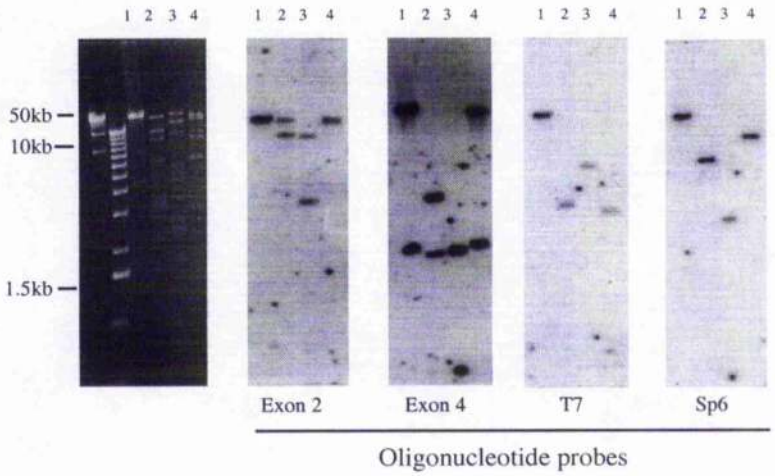
The presence of exon 2 and exon 4 together on the same fragment indicated that HSD3Bψ2 does not contain an XhoI restriction site between these exons. Therefore, other restriction enzymes were used to attempt to digest the 24kb XhoI fragment from BAC 7 and the 30kb XhoI fragment from BAC 22. Ideally, the enzyme would digest between exons 2 and 4 without restricting between the gene and the BAC vector. Although several restriction enzymes were used in double digests with XhoI, none cut between the exons without cutting between the gene and the vector (Fig. 4.14B and C).

It can be observed from Fig. 4.14B and C that none of the fragments containing the BAC vector arms hybridised to either the conserved exon 4 oligonucleotide or the conserved exon 2 oligonucleotide. There was not time to test other restriction enzymes. Therefore, although the orientation of HSD3Bψ2 was not determined, the gene was incorporated into the XhoI restriction map. The left end of BAC 1 had already been mapped in Fig. 4.13E. By examining the restriction fragments and the hybridisation patterns in common between BACs 1, 22 and 7 and analysis of genes further along the cluster, the restriction map of the genomic cluster can be extended to include HSD3Bψ2 (Fig. 4.14D).





C - BAC 22



D

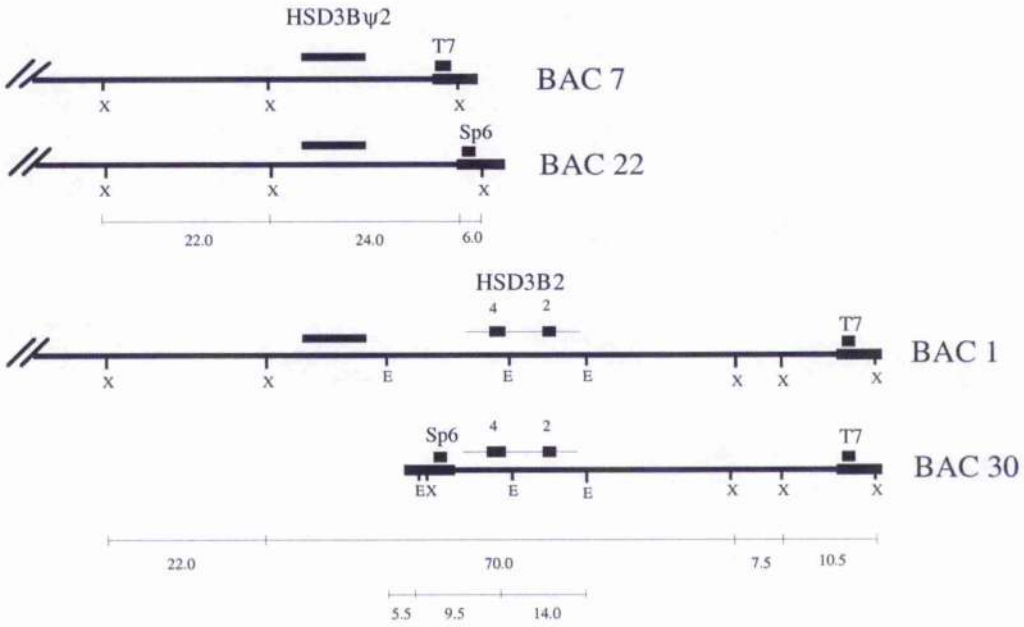


Fig. 4.14 A: PFGE of Xho I restriction digests of BACS 7, 22 and 1, hybridised with exon 2, exon 4, T7 and Sp6 oligos by Southern analysis.

B: Agarose gel showing BAC 7 digested with panel of restriction enzymes, hybridised with exon2, exon 4, T7 nd Sp6 oligos by Southern analysis. Lane1 - XhoI/BstXI; lane2 - XhoI/NdeI; lane3 - XhoI/ SphI; lane4 - XhoI/SstII; lane5 - XhoI/MluI; lane6 - XhoI/HincII.

C: Agarose gel showing BAC 22 digested with a panel of restriction enzymes, hybridised with exon2, exon 4, T7 and Sp6 oligos by Southern analysis. Lane1 - XhoI; lane2 - XhoI/KpnI; lane3 - XhoI/SphI; lane4 - XhoI/PstI.

D: XhoI restriction map including the position of HSD3Bψ2.

Orientation and restriction mapping of HSD3Bψ1:

From the gene-specific PCR data, this gene is found on BACs 1, 22, 29 and 7 (Fig. 4.5). XhoI restriction digests of these clones were probed with conserved exon 2, conserved exon 4, T7 and SP6 oligonucleotides (Fig. 4.15A). The conserved exon 4 oligonucleotide hybridised to a 1.8kb fragment from each BAC. This indicated that these fragments contain HSD3Bψ1 exon 4, since this gene is the only HSD3B gene present on all 4 of these BAC clones. The orientation of HSD3Bψ1 was determined from BAC 29 XhoI restriction digests and hybridisations. From the gene-specific PCR data (Fig. 4.5), BAC 29 contains 2 genes - HSD3Bψ1 and HSD3B1. Two exon 2 fragments are observed for BAC 29 (Fig. 4.15A), one at 12kb and the other at 48kb. The T7 oligonucleotide hybridised to the 12 kb fragment also and the SP6 oligonucleotide hybridised to the 48kb. From this information, it can be concluded that at either end of the BAC 29 genomic insert there is an exon 2 present and one of these must belong to HSD3Bψ1 and the other to HSD3B1. Within BACs 1, 22 and 7 there is an XhoI fragment of 22kb which contains an exon 2. Since HSD3Bψ2 belongs to the left end of BACs 7 and 22 (Fig. 4.14D) and HSD3Bψ1 is the only other gene to be present in all three of these BAC clones, this exon 2 must belong to HSD3Bψ1. Therefore, the 48kb fragment does not fit into the XhoI map as HSD3Bψ1 but the 12kb fragment does.

Only two more bands remain to be incorporated into the BAC 1 XhoI restriction map, one of these is 4.5kb and the other 4.8kb. The 4.5kb fragment is also present in BACs 22, 29 and 7, therefore it can be positioned next to the 1.8kb, and the 4.5kb fragment which contains SP6 vector arm indicating the end of the clone. Again, there are fragments that have not been discussed here. These have either been discussed earlier in connection with HSD3B2 or IISD3Bψ2 or they will be discussed in connection with genes further along the contig.

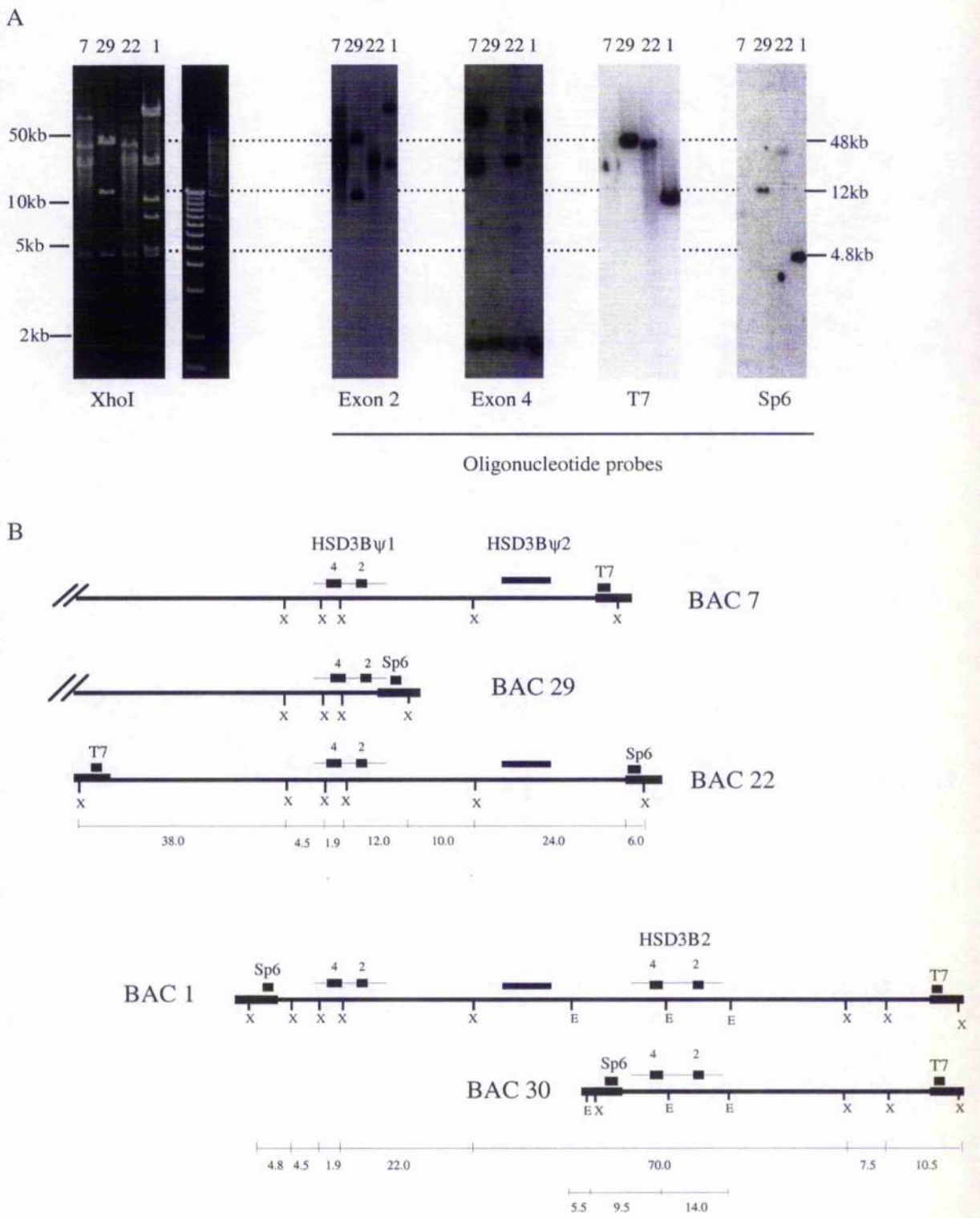
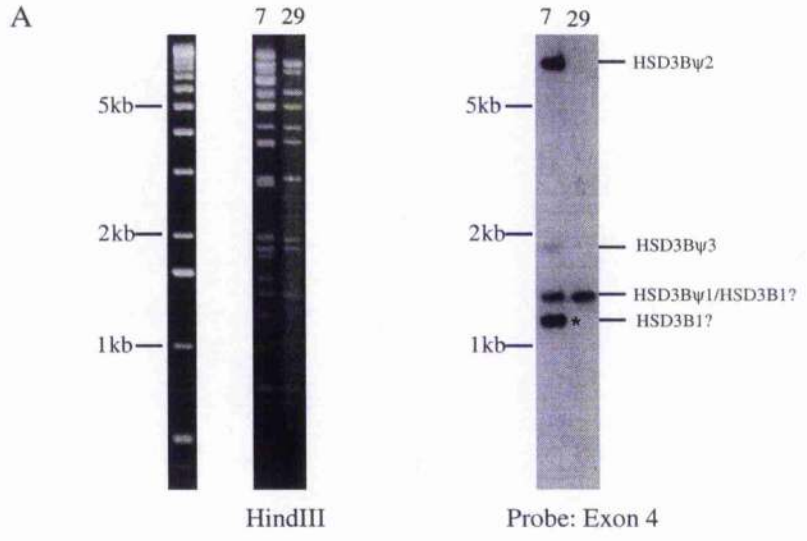


Fig. 4.15 A: PFGE of XhoI restriction digests of BACS 7, 29, 22 and 1, hybridised with exon 2, exon 4, T7 and Sp6 oligos.
B: XhoI restriction map including the orientation of HSD3Bψ1.

Orientation and restriction mapping of HSD3B1:

This gene was found on BAC clones 29 and 7 by gene-specific PCR (Fig. 4.5). When these clones were digested with HindIII and hybridised with the conserved exon 4 oligonucleotide, only one band (1.6kb) was observed from BAC 29, whereas four bands were detected from BAC 7 (Fig. 4.16A). Although four bands were expected from BAC 7 (i.e. HSD3Bψ2, HSD3Bψ1, HSD3B1 and HSD3Bψ3), two were expected from BAC 29 (HSD3Bψ1 and HSD3B1). This raised several questions: Did the 1.6kb fragment from BAC 29 represent both HSD3Bψ1 and HSD3B1 exon 4 fragments and if so, what gene did the 1.4kb fragment from BAC 7 represent, or, was it possible that BAC 29 did not contain HSD3B1 exon 4, only HSD3Bψ1 exon 4? To anwser this, the sequence of the 1.4kb fragment from BAC 7 was determined by subcloning into pUC18. HSD3B1 sequence was obtained from this subclone. Therefore, it was possible that BAC 29 does not contain HSD3B1 exon 4 sequence and that exon 4 sequence of HSD3B1 was only present on BAC 7. This deduction is supported by the restriction digests of BAC 29 hybridised with conserved exon 2 and exon 4 oligonucleotides (Fig. 4.9). Only one exon 4 fragment can be detected in every digests, whereas two exon 2 fragments can be seen in more than one of the digests. Also, as discussed in relation to HSD3Bψ1, the XhoI digest of BAC 29 probed with conserved exon 2, exon 4, T7 and SP6 oligonucleotides (Fig. 4.14A) indicated that BAC 29 contained exon 2 fragments at the ends of the genomic insert. From this information, the XhoI restriction map can be further extended to include the orientation of HSD3B1 (Fig. 4.16B). There is an XhoI site present between exon 2 and exon 4 of HSD3B1, which can be deduced from BAC 7 XhoI restriction map (Fig. 4.15A). There is a fragment of approximately 48 kb in BAC 7 which contains an exon 2, this band represents HSD3B1 exon 2.



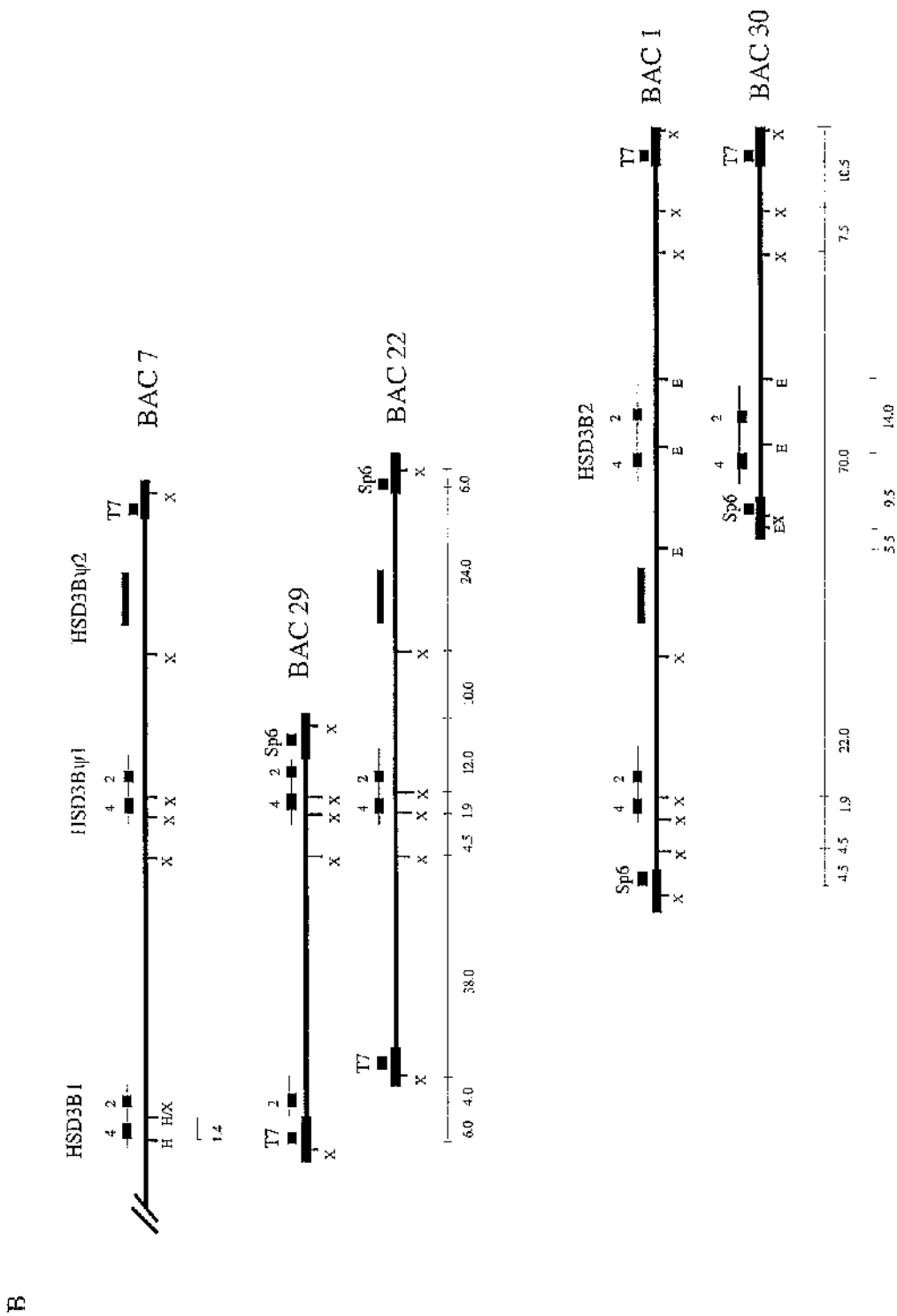
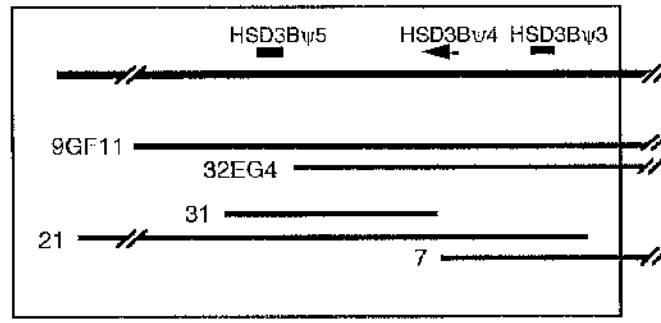


Fig. 4.16 A: Agarose gel indicating HindIII restriction digests of BACS 31, 21, 7, 29, 1 and 30. probed with conserved exon 4 oligonucleotide. * indicates the band that was subcloned.
B: XhoI restriction map including the orientation of HSD3B1. This is the complete XhoI restriction map for section 1.

LEFT SIDE OF MAP



This section considers the HSD3B genes HSD3Bψ3, HSD3Bψ4 and HSD3Bψ5. A brief summary of the significant results concerning the left side of the cluster is presented below.

GENE:	COMMENT:
IISD3Bψ3	This member of the 3ß-HSD gene family was not orientated due to the lack of informative restriction sites (Fig. 4.17B).
HSD3Bψ4	HSD3Bψ4 was orientated from the breakpoint position of BAC 31. BAC 31 does not contain HSD3Bψ4 exon 2 (Fig. 4.18B).
HSD3Bψ5	This gene was not orientated. A possible orientation is proposed (Fig. 4.19B), however the lack of exon 4 sequence for HSD3Bψ5 prevented confirmation of this.

Orientation and mapping of HSD3Bψ3:

HSD3Bψ3 was assigned to BACS 7 and 21 by gene-specific PCR (Fig. 4.5). Xho I digests probed with conserved exon 2, conserved exon 4, Sp6 and T7 oligonucleotides indicated that HSD3Bψ3 was situated within BACs 7 and 21 on fragments of 65kb and 70kb respectively (Fig. 4.17A). This was deduced from the construction of the XhoI maps of BACs 7 and 21, in relation to the other HSD3B genes. These fragments were too large for the vector arm probes to be beneficial and the orientation was not determined. The 65kb fragment from BAC 7 contained exon 2, exon 4 and SP6, the 70kb fragment from BAC 21 contained exon 2, exon 4 and T7. To determine the orientation of IISD3Bψ3 it would require using markers within BAC 7 insert sequence, possibly using HSD3Bψ4, the next gene along, as a starting point. HSD3Bψ3 was incorporated into the XhoI restriction map (Fig. 4.17B).

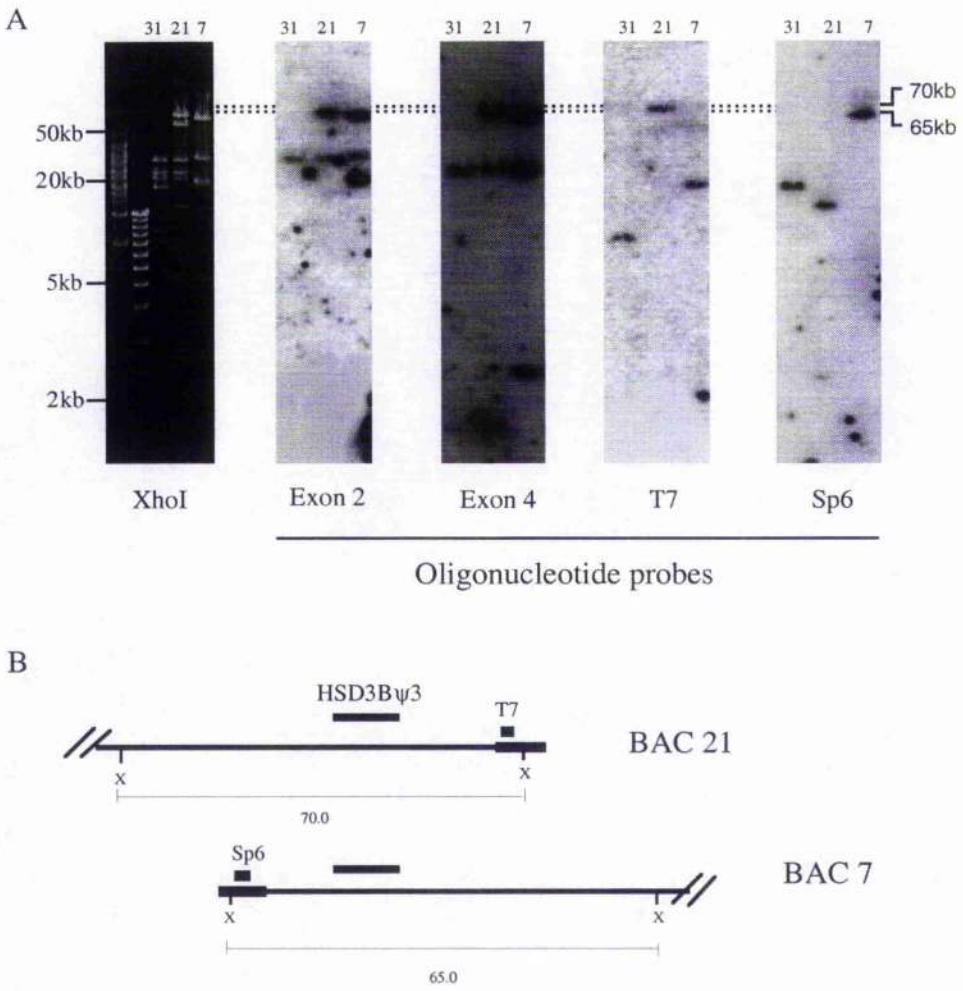


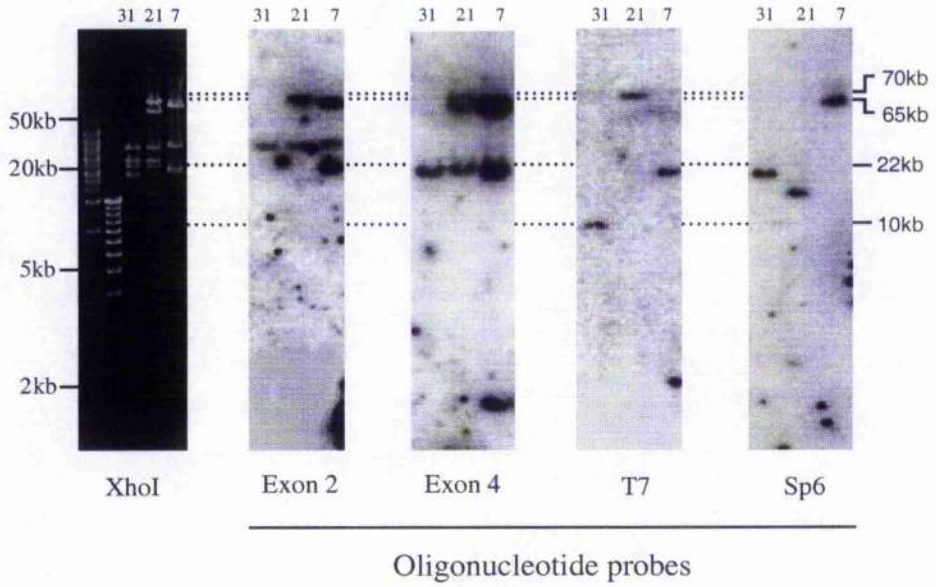
Fig. 4.17 A: PFGE of XhoI restriction digests of BACS 31, 21 and 7 hybridised with conserved exon 2, conserved exon 4, T7 and Sp6 oligonucleotides.
B: XhoI restriction map indicating the position of HSD3Bψ3.

Orientation and mapping of HSD3Bψ4:

The orientation of HSD3Bψ4 was determined by subcloning a fragment from Hind III digest of BAC 7 hybridised with conserved exon 2 oligonucleotide (Fig. 4.7, see section 4.2.4). A 5kb HindIII fragment from BAC 7 was unidentified and was therefore subcloned into pUC. The sequence obtained from the subclone was HSD3Bψ4 exon 2. The gene-specific PCR data indicated that HSD3Bψ4 was present only on BAC 31 and 21, not BAC 7 (Fig. 4.6). However, the gene-specific PCR primer pair for HSD3Bψ4 was designed to exon 4 (Fig. 4.1A) and it was concluded that HSD3Bψ4 exon 2 was present on BAC clones 7 and 21, not BAC 31, whereas HSD3Bψ4 exon 4 was present on BAC clones 21 and 31 and not BAC 7. This is confirmed in Fig. 4.9 where only one exon 2 band can be detected in each lane when BAC 31 is digested with a range of restriction enzymes and hybridised with the

conserved exon 2 oligonucleotide. From Fig. 4.18A, an XhoI restriction map of BAC clones 7, 21 and 31 was constructed and HSD3Bψ4 was incorporated into the map in the correct orientation (Fig. 4.18B). Within the XhoI map, HSD3Bψ4 exon 2 is present on a 65kb fragment from BAC 7 and a 70kb fragment from BAC 21 (the same fragments that contain HSD3Bψ3, Fig. 4.17B). HSD3Bψ4 exon 4 is present on a 22kb fragment on BACs 21 and 31 and the T7 vector arm is found on the 10kb fragment.

A



B

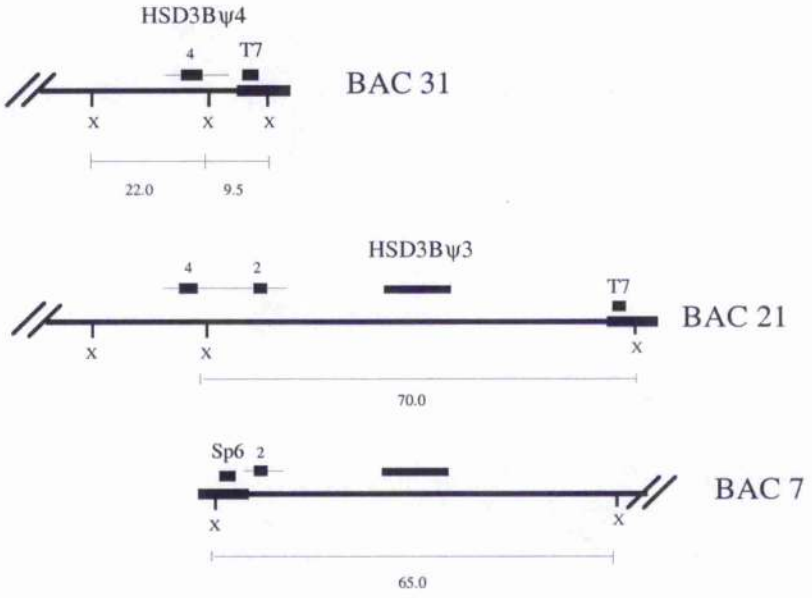


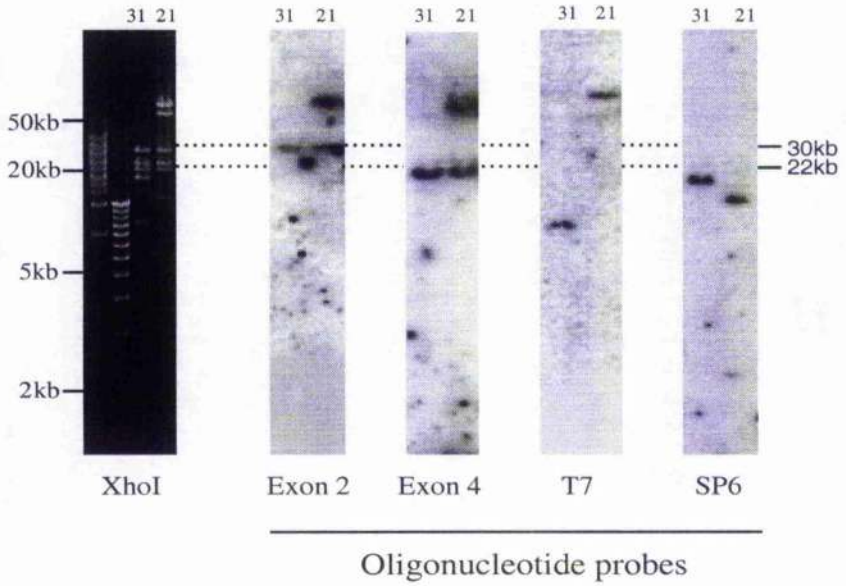
Fig. 4.18 A: PFGE of XhoI restriction digests of BACs 31, 21 and 7 hybridised with conserved exon 2, conserved exon 4, T7 and SP6 oligonucleotides.
B: XhoI restriction map indicating the orientation of HSD3Bψ4.

Orientation and mapping of HSD3Bψ5:

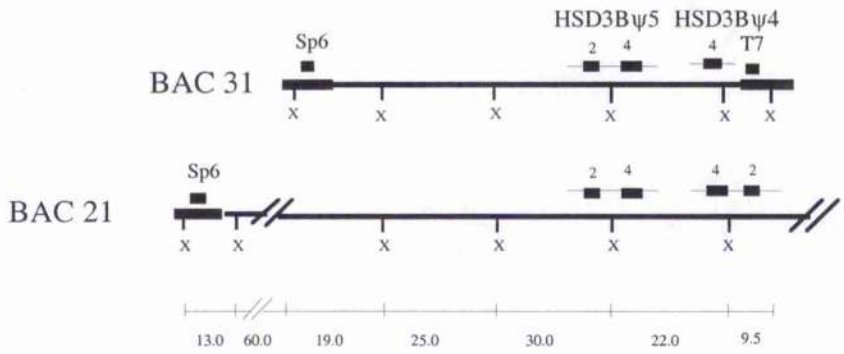
The orientation of HSD3Bψ5 was not determined. It was observed that HSD3Bψ5 exon 2 was present on a 30kb Xho I fragment from BAC 31, however, only one exon 4 fragment (22kb) was detected and it was expected that HSD3Bψ4 exon 4 and ψ5 exon 4 were present on BAC 31 (Fig. 4.19A). It is possible that both HSD3Bψ4 and HSD3Bψ5 exon 4 sequences are present within this fragment and HSD3Bψ5 is orientated in the opposite direction from the other genes (Fig. 4.19B). Although, the intensity of the 22kb fragment does not suggest a doublet. Fig. 4.19B indicates the XhoI restriction maps of BACs 21 and 31 determined from Fig. 4.19A.

HSD3Bψ5 exon 4 sequence was not obtained from the phage clone used to generate HSD3Bψ5 sequence (Appendix ii). Sandra Burridge, a collegue in the lab, attempted to clone HSD3Bψ5 exon 4 from BAC 31 but detected only HSD3Bψ4 sequence from 9 clones. RT-PCR data (produced by Sandra Burridge) supplied 100bp of HSD3Bψ5 exon 4 sequence and within this sequence it was demonstrated that there was an NciI restriction enzyme site difference between ψ4 and ψ5 (Fig. 4.19C). This was used to attempt to distinguish ψ4 and ψ5 exon 4. Unfortunately, only one band was obtained when a NciI/XhoI digest was probed with exon 4 oligonucleotide (Fig. 4.19D). This could be due to the loss of another fragment from the pulse field gel or it is possible that the NciI site was introduced as a Taq polymerase mutation during the RT-PCR. One solution for this problem would be to design an HSD3Bψ5 exon 4 specific probe, if more exon 4 sequence could be acquired, and it could be used to determine which XhoI fragment contains HSD3Bψ5 exon 4 sequence.

A



B



C

```
HSD3Bψ5  acagGTACCCAGCTTCTGTTGGAGGCCTGTGTCCAAGCTACAGTGC
         ||||||||||||||||||||||||||||||||||||||||||||||
HSD3Bψ4  acagGTACCCAGCTTCTGTTGGAGGCCTGTGTCCAAGCTACAGTGC
                    .         .         .
                                        NciI
HSD3Bψ5  CTTCATCTACACCAGTACCCCAGAGGTAGCCGGGCCC
         |||||||||||||||||||||||||||||| ||||| |||||
HSD3Bψ4  CTTCATCTACACCAGTACCCCAGAGGTAGCCAGGCCC
```

D

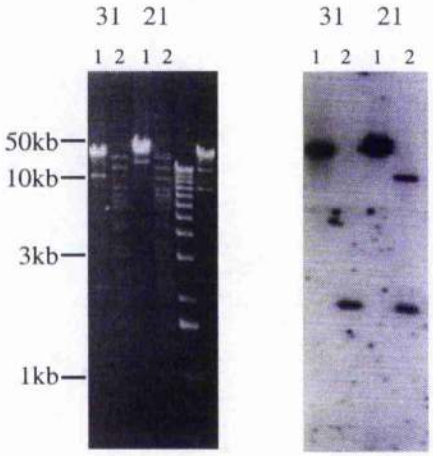


Fig. 4.19 A: PFGE of XhoI restriction digests of BACs 21 and 31 hybridised with conserved exon 2, conserved exon 4, T7 and Sp6 oligonucleotides.
B: XhoI restriction map showing possible orientation of HSD3Bψ5.
C: 5' exon 4 sequence for HSD3Bψ4 and HSD3Bψ5 indicating the NciI site present in HSD3Bψ5. Intronic sequence in lower case.
D: Agarose gel showing BACS 31 and 21 digested with XhoI and NciI, probed with conserved exon 4 oligonucleotide. Lane1 - XhoI; lane2 - XhoI/NciI

4.2.8 Investigating the human genome further for 3ß-HSD sequences using human-mouse somatic cell hybrids

The availability of a panel of human-mouse and human-rat somatic cell hybrids overlapping the entire human genome except chromosome 1, provided an opportunity to determine if there were 3ß-HSD sequences outwith chromosome 1 (appendix iii). PCR primers were designed to amplify all identified human 3ß-HSD genes but not the mouse or rat HSD3B genes (the sequence of these oligonucleotides can be found in section 2.6.2). PCRs were attempted under stringent annealing conditions and one cell hybrid (HA11) gave an amplification product of the expected size (850bp; Fig. 4.20). HA11 contains chromosomes 4, 5, 6, 8, 11, 20 and 21.

Fig. 4.20 Agarose gel indicating PCR amplification products obtained from a panel of human-mouse somatic cell hybrids using primers specific to human 3ß-HSD. Lane 1 - HA11, lane 2 - HA221, lane 3 - HA232, lane 4 - HB29, lane 5 - HB33, lane 6 - HB35, lane 7 - HB111, lane 8 - HB142.2, lane 9 - HB181, lane 10 - JV01, lane 11 - HR40C8, lane 12 - rat genomic DNA control, lane 13 - mouse genomic DNA control, lane 14 - human genomic DNA control, lane 15 - no template -ve control.

The product obtained from HA11 was cloned into T-vector, recombinants were screened for inserts by PCR and positive clones were sequenced. From 9 sequences obtained, 5 were HSD3Bψ4, 1 was HSD3Bψ2, 2 were HSD3B1 and the last was HSD3Bψ1. This suggests that HA11 is contaminated with at least chromosome 1p13 and the above data indicates that no other 3ß-HSD sequences exist outwith chromosome 1.

4.2.9 Attempts to orientate the 3ß-HSD genomic contig with respect to the centromere.

To attempt to orientate the 3ß-HSD gene cluster with respect to the centromere on human chromosome 1, the genetic markers D1S514 and D1S534 were utilised. Microsatellite marker D1S514, also known as AFMa151za5 (Genome database (GDB) accession identification GDB:200157) is a single copy CA dinucleotide repeat sequence

located 177cM from the end of the short arm of chromosome 1 at position 1p13.1. The microsatellite marker D1S534, or CHLC.GATA12A07.778 (GDB:686478) is a single copy GATA tetranucleotide repeat sequence, centromeric to D1S514. Both D1S514 and D1S534 micosatellite markers are closely linked to the 3ß-HSD gene family, therefore PCR primers pairs designed to each marker were used to screen the genomic clones from the 3ß-HSD contig. It was envisaged that one of the markers may be identified on a genomic clone allowing the orientation with respect to the centromere to be determined. Unfortunately, no PCR amplification products were obtained for either of the markers with any of the genomic clones, therefore, the orientation of the 3ß-HSD gene family with respect to the centromere was not obtained from these experiments. It is possible that fluorescent *in situ* hybridisation (FISH) analysis using probes to HSD3B2 and HSD3Bψ5 (HSD3B genes situated at either end of the cluster) may provide the anwser to this orientation question.

CHAPTER 5

Hybrid selection

5.1 INTRODUCTION

By establishing the genomic contig for the 3ß-HSD gene family it was now possible to ask the following questions; what other genes (if any) can be localised to this region and are these genes unidentified members of the 3ß-HSD gene family? Within this chapter, I will briefly describe the methods developed to identify transcribed sequences from particular genomic regions and discuss why I chose to use hybrid selection. I will then present my results from the selection experiments.

One of the earliest techniques devised to identify coding regions within a genomic clone was cross-species hybridisation or "zoo blots". This was based on the observation that coding sequences would be much more conserved between species compared to non-coding sequences. This analysis involved hybridising a genomic clone to DNAs from a variety of species; generally coding sequences would hybridise at higher stringency with the genomic DNAs than non-coding sequences. The distinct disadvantage of this method is that it will not detect genes that have diverged significantly between species. However, this technique was used successfully in the identification of the dystrophin gene involved in X-linked Duchenne Muscular Dystrophy (Monaco *et al.*, 1986).

Another early technique used to identify genes was to use specific detection of human repetitive elements such as Alu repeats from somatic cell hybrids containing a particular chromosome. Approximately 10% of human mRNAs have human specific, highly repeated sequences in their 3' non-coding region. In 1986, a cDNA library constructed from a somatic cell hybrid containing only human chromosome 21 was screened for Alu repeats by Neve *et al.* (1986). However from this screen only 3 clones from a library of ~1 million were identified. A limitation of this type of screen is that it identifies only genes expressed in the hybrid cell, and genes that have repeats in their 3' untranslated regions. A recent modification of this strategy is to use unspliced, heterogeneous nuclear RNA (hnRNA) isolated from the hybrid cells. This removes the requirement of repeat elements in the mature RNA; intron sequences are more likely to contain repeats. This method depends on the half-life of the intermediates, the presence of repeats in the introns and the proximity of the introns to exon sequences.

Some mammalian genes have been isolated by selection for CpG islands. In the mammalian genome cytosines in the CpG dinucleotide are highly methylated. However, a small fraction contain unmethylated cytosines, and these have been shown to occur as discrete "CpG islands" (Larsen *et al.*, 1992). These CpG islands have been associated with expressed sequences, as CpG methylation often inhibits

transcription, and can be detected by cleavage with rare cutting restriction enzymes. HpaII is sensitive to methylation and only cuts CpG doublets that are unmethylated; the fragments obtained are known as HpaII tiny fragments. By looking for these, large stretches of DNA can be checked for the presence of CpG island clusters. A disadvantage of this technique is that the region involved must be analysed by pulse field gel electrophoresis (PFGE), the DNA fragments have then to be cloned and examined for transcribed sequences, and cDNA clones representing these sequences have to be isolated.

Northern blot analysis is another method that can be used to detect expressed sequences contained within a genomic clone. The genomic clone can be used as a probe to screen RNA derived from different tissues. A major disadvantage of this technique is that YACs are too complex to reliably identify transcripts. Phage and cosmids have successsfully detected transcripts, so YAC clones could be subcloned and the subclones used as probes. However, subcloning and subsequent testing of every subclone for hybridisation is quite labour intensive. Another disadvantage is that the transcript must be present in the tissue that the RNA has been derived from to be able to detect it at all.

Screening cDNA libraries using genomic clones as a probe is a technique commonly used to identify new genes. YACs, BACs, P1 clones, phage and cosmid clones can all be directly labelled and used to screen a cDNA library. Using this method multiple transcripts can be identified from a single screen, but there are a number of distinct disadvantages. DNA from the genomic clone of interest must be isolated in sufficient quantity and purity to be used for labelling, and this can be difficult especially with YAC clones. The probe involved will be complex and will contain a large number of repetitive sequences, which could lead to weak signals, increased background and false positives. The problems can be alleviated to a certain extent by pre-hybridising the probe to COT 1 DNA or sheared human DNA to block the repetitive sequences. Clones that contain short exons or clones that are poorly represented in the library may be missed altogether. It is unlikely that a conventional cDNA library of ~1 million clones will adequately represent all the transcripts expressed in the tissue since it will probably not contain the lower abundance transcripts (Morgan *et al.*, 1992), although using a normalised cDNA library (by reassociating the abundant cDNAs and removing them) would help to overcome this problem (Patanjali *et al.*, 1991). Another problem associated with screening a cDNA library is that, as in northern analysis, the gene in question must be expressed in the tissue from which the library is made. More than one library could be screened, but that makes this type of experiment extremely labour intensive, especially when more than one probe is involved. PCR pre-screens have

been used to identify the library with the greatest potential of yielding positive cDNAs with a particular probe, and probe-pooling may be adopted (Harshman *et al.*, 1995). Primers are designed for each probe, and PCR amplification is attempted for each library. Probes that give a faint product or no product can be pooled, i.e. probes are grouped according to their relative representation in the different libraries avoiding problems of pooling well-represented probes. This experiment will also identify which probes give strong signals with a certain library, allowing preferential screening to be undertaken. Although there are certain disadvantages of library screening it will still detect genes with cryptic splice sites, genes that lack CpG islands, genes that are not transcribed in somatic cell hybrids, and genes that have highly diverged between species, all of which may be missed by other reported methods for finding expressed sequences.

A stategy was devised by Duyk *et al.* (1990), and further developed by Buckler *et al.* (1991) that involves the identification of exons from random pieces of genomic DNA, this is known as exon trapping. The stategy is based upon RNA sequences containing functional splice sites. The genomic DNA is "shotgun" cloned into a vector that contains functional splice donor and acceptor sites. The clones are then propagated in *E.coli* and transfected into COS cells. When the cloned DNA contains an exon, splicing ocurrs between the vector and the genomic DNA. RNA is then isolated from the cells, and cDNA is synthesised and amplified using primers to the splicing vector. A unique DNA fragment is obtained where correct splicing has occured (see Figure 5.1). Theoretically this procedure should identify all internal exons from a region of interest and has been applied with great results (Chen *et al.*, 1996). However it is technically complex, and as it simply requires the presence of splice donor and acceptor sites, false positives can be encountered.

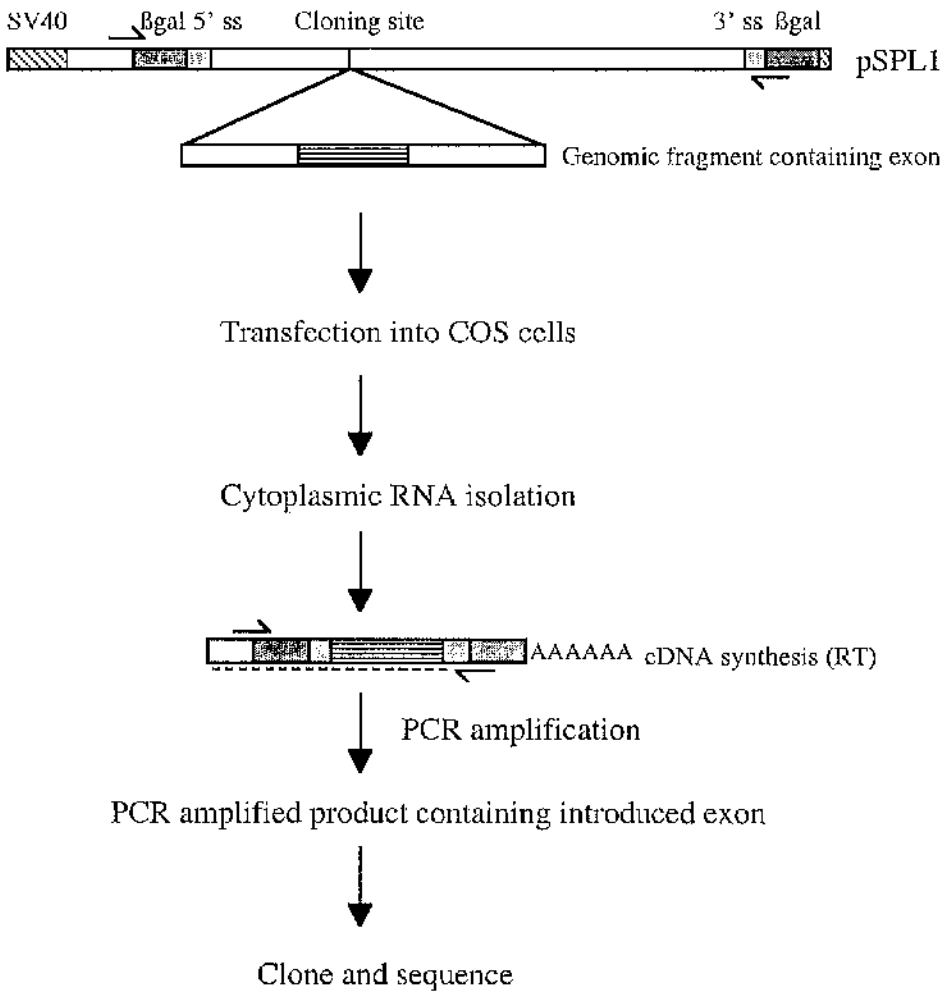


Figure 5.1. Exon Trapping. Genomic fragments are cloned into a pSPL1 plasmid and transfected into cos cells. An exon present in the genomic fragment in the proper orientation will be spliced into the mature RNA and will be present in the amplified product flanked by known sequence (Modified from Buckler *et al.*, 1991).

Recently, direct cDNA selection or "hybrid selection" techniques have been devised. These methods employ the sensitivity of PCR, with specific selection of cDNAs by hybridisation to large genomic regions. The first strategy was developed by (Lovett *et al.*, 1991). The basis of this was the hybridisation of an entire cDNA library to an immobilised genomic clone. YAC DNA was isolated by excision from a contour-clamped homogeneous electric field gel, digested with MboI and immobilised onto nylon filter discs. cDNA inserts were amplified from a oligo d(T)-primed fetal kidney library using primers designed to the library vector. The amplified inserts were preblocked with sheared total human DNA, pBR322 DNA and yeast DNA. This reduced non-specific hybridisation of repetitive elements, vector sequences and yeast sequences. The blocked cDNA was then hybridised in solution with the filter

containing the genomic YAC DNA. Non-specific hybrids were removed, and selected cDNAs were eluted, which were amplified using the primers used to amplify the cDNA at the onset, and either cloned into a phage vector or subjected to further selection (see Figure 5.2). Parimoo *et al.*(1991) developed a similar technique. Their method differed only in that to prevent nonspecific hybidisation the immobilised genomic clone was blocked rather than the cDNA, and they used a highly complex competitive DNA mixture including a human chromosome 15 library, a genomic repetitive sequence library, rRNA-specific clones, Poly(dI)-poly(dC), and yeast DNA. Also, a set of nested primers was used to amplify and reamplify the eluted cDNA to reduce the chance of PCR artefacts.

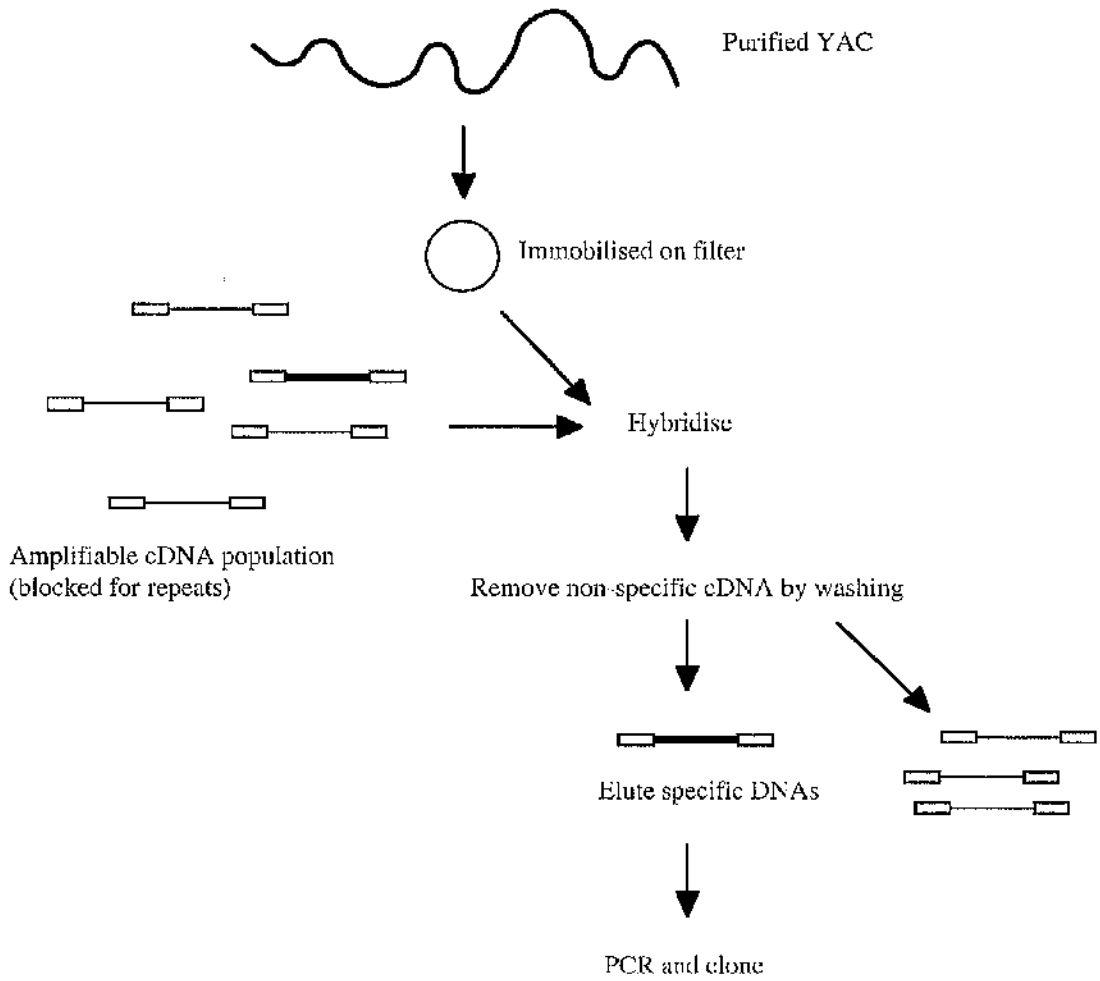


Figure 5.2. Direct selection with genomic clone immobilised onto nylon filter. (Modified from Lovett *et al.*, 1991)

More recent methods maintain both hybridisation partners in solution which allows better control of the hybridisation conditions compared to filter hybridisation (Korn *et al.*, 1992, Morgan *et al.*, 1992). Morgan *et al.* (1992) used this method to isolate novel cDNAs encoded by the regions surrounding the human IL-4 and IL-5 genes. Double stranded cDNA was digested with MboI and ligated to a linker/adaptor oligonucleotide. YAC DNA was also digested with MboI and ligated to a second oligonucleotide linker/adaptor. The YAC DNA and the cDNA were then separately amplified by PCR. A 5' biotinylated primer was used for the YAC amplification. Cot1 DNA was used as a blocking agent by mixing with the starting cDNA (Cot1 DNA is highly enriched in intermediate repeats and does not increase the overall sequence complexity of the mixture compared to total genomic DNA). Blocked cDNA was then hybridised in solution with the biotinylated YAC DNA. YAC DNA plus bound cDNAs were captured by streptavidin coated magnetic beads, and the beads were removed from the solution using a magnet. Selected cDNAs were eluted from the beads, amplified and either cloned or subjected to another round of enrichment (see figure 5.3).

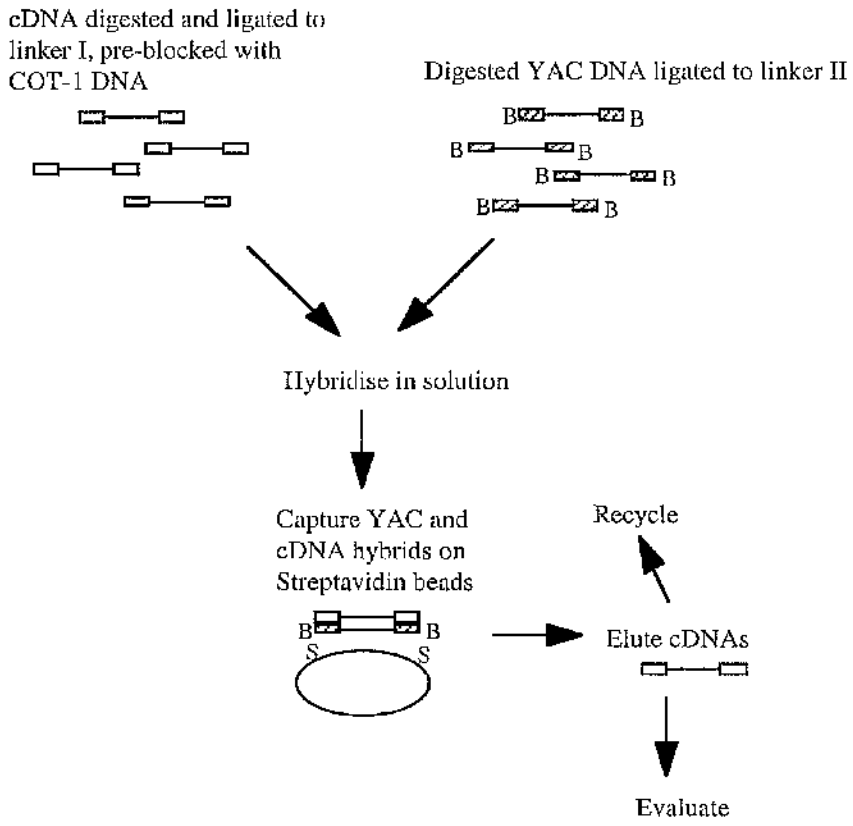


Figure 5.3. A biotin/streptavidin capture system for direct selection. (Modified from Morgan *et al.*, 1992)

Hattier *et al.* (1995) made a few more modifications to this procedure and used P1 clones, BACs and cosmids as genomic probes, rather than YACs. After digestion, the clones were treated with exonuclease III to minimise competitive renaturation of the probe during hybridisation. Also, the genomic clone preparation did not require amplification, as differences in amplication may bias representation of the clone sequences. Finally, cDNA synthesis involved two separate primers for first and second strand synthesis, which improves cloning efficiency and allows some PCR artefacts to be easily identified.

There are several advantages of hybrid selection. First, it is rapidly accomplished. Second, this method tends to normalise the frequency of the transcripts encoded by the genomic clone (especially when it is supplied in limiting amounts) and it has greater enrichment factors for rare transcripts than abundant ones. The abundant cDNAs will quickly saturate their genomic target, leaving the majority of cDNAs in solution. Thus the net result should be an approximate normalisation of the abundant cDNAs downwards and the low abundance transcripts upwards. Since the isolation of rare cDNAs is as important as the isolation of abundant ones, enrichment of these is a great advantage of this technique over screening conventional libraries which may not even contain these transcripts. Third, it is more robust than screening with labelled probes and it has a much higher signal to noise ratio mainly because it is easier to quench sequences by blocking to prevent nonspecific hybridisation. Fourth, it is less sensitive to the size of the clone used for selection. Cosmid, BAC or P1 DNA is more favourable than YAC DNA because it is less prone to deletions and chimerism and is readily separable from *E.coli* DNA, whereas the presence of degraded yeast DNA from higher molecular weight yeast chromosomes, especially ribosomal sequences, can cause contamination problems. These sequences show strong homology to human ribosomal RNA sequence, and so many enriched cDNAs may be of ribosomal origin and not from the region of interest. This problem can be overcome either by screening the enriched sublibrary with a ribosomal probe to identify ribosomal DNA containing clones, or pre-blocking the YAC DNA with total yeast DNA (as above), or by cutting the yeast genomic DNA prior to isolation of the YAC. By employing one of these methods, the selection efficiency of YAC clones compared to cosmid clones is only 2-fold in favour of the cosmids (Korn *et al.*, 1992). Eventually all genomic sources will result in an enrichment of the cDNAs encoded by the insert.

There are a few limitations with hybrid selection. The first is that it does not usually retrieve complete coding sequences for genes; rather, it provides a route to finding expressed sequence tags (ESTs) which can be used as probes to isolate the gene. Thus, screening a library is probably a second step in the process. However, it is

preferable to screening a library directly as abundance normalisation permits rarer transcripts to be found more easily. The second potential problem is the co-selection of pseudogenes which select cDNAs which belong to genes located somewhere else in the genome. This can be sorted by sequencing the cDNA and the corresponding genomic fragment, or using the products of exon-trapping as probes to discriminate against processed pseudogenes (Korn *et al.*, 1992). Thirdly, it is limited when the transcription pattern of the gene of interest is unknown. cDNA from many different tissues or developmental stages may have to be tested. However it is possible to test the cDNA from more than one tissue at a time by ligating on different oligonucleotide linker/adapters and using specific primers for PCR amplification of selected cDNAs from the different tissues (Lovett *et al.*, 1991).

5.2 METHODS AND RESULTS

The purpose of this chapter was to explore the genomic contig generated in Chapter 4 for unknown expressed sequences. To do this, hybrid selection was applied to attempt to isolate expressed sequences from a cDNA library using biotinylated genomic clones.

5.2.1 Preparing the cDNA

The chances of a successful experiment are dramatically increased by creating the best source of cDNA. Any source of cDNA can be used in this experiment as long as it can be amplified by PCR. Oligo (dT) primed libraries, normalised libraries, MboI digested oligo (dT) primed cDNA and a combination of oligo (dT) and random primed libraries have all been used. Different groups have treated the starting cDNA differently, for example, Lovett *et al* . (1991) amplified inserts from an oligo (dT)-primed fetal kidney cDNA library whereas Morgan *et al.* (1992) digested oligo (dT) with MboI and ligated linker oligonucleotides. Using cDNA inserts that have been amplified from the vector results in bias against large cDNAs if the original was full length rather than random primed. Digesting the cDNA with a restriction enzyme fragments the full length cDNAs into smaller fragments and removes an intermediate cloning site and possible loss of some transcripts. However, by digesting the cDNA it probably would not be possible to determine how many individual transcription units the enriched cDNAs are derived from. An additional step to convert each fragment to a full length would be necessary, either by returning to the full length starting cDNA or using random primed cDNA as the starting material which would retain some overlap between clones after selection. One other problem with fragmenting the cDNA would be the loss of genes

because of the lack of internal restriction sites. Therefore using normalised libraries or random-primed libraries would eliminate these problems. However, the quality of a cDNA sub-library created from the enriched cDNAs depends on the size of the inserts. It is advantageous to work with clones that are as long as possible as these would carry more information and be easier to group into genes.

The first decision in making the cDNA library was to identify the tissue of origin of the mRNA. In this case, the placenta was selected as it is a major steroidogenic tissue and it is a complex tissue which expresses a huge variety of genes. It was also available within seconds of delivery which is a feature that applies to few human tissues except cultured cells. In addition, the placenta expresses large amounts of 3ß-HSD type I and therefore would make the ideal control for the selection experiments using the genomic clones isolated in Chapter 4. It was expected that the 3ß-HSD transcripts within the cDNA population would be enriched by the selection process.

Total RNA was isolated from normal full-term placenta (obtained from the Queen Mother's Maternity Hospital, Glasgow) using Tri-reagent (Sigma). Poly (A)+ mRNA was isolated using the FastTrack mRNA isolation kit (InVitrogen) and the InVitrogen Copy kit was utilised to synthesise double-stranded, blunt-ended cDNA. All of the above were used exactly according to the manufacturers recommendations. The cDNA was synthesised using a 1:1 ratio of oligo dT and random primer. This would produce a mixture of full-length and random primed cDNAs. The oligo dT clones may provide more information and the random primed clones may produce overlaps and result in more than just 3' untranslated sequence. However, it is possible that some genomic contamination may be introduced using the random primers. The cDNA synthesised was tested by PCR using primers designed to 3ß-HSD type I (type I and D; section 2.6.2) and glyceraldehyde phosphate dehydrogenase (GAPDH). 3ß-HSD type I transcripts are abundant in the placenta, and GAPDH is a housekeeping gene expressed in every tissue. Both sets of primers resulted in a strong signal from the cDNA produced (Fig. 5.4A).

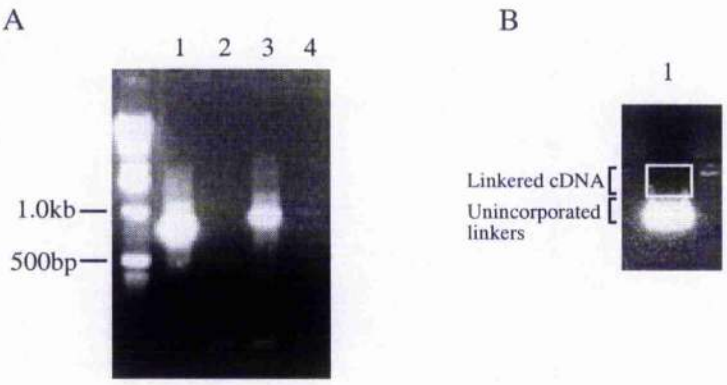
Next, the cDNA was either digested with AluI or ligated to the linkers immediately. This resulted in two cDNA populations. One contained short fragments and eliminated the bias against longer full-length cDNAs in the PCRs, and the other produced longer clones that may contain more information. The linker used was designed by Dr. Gerry Graham from the CRC Beatson Laboratories, Glasgow. The sequences of the oligonucleotides used to produce the linkers were:

A: 5' TAG TCC GAA TTC AAG CAA GAG CAC A 3'
B: 5' CTC TTG CTT GAA TTC GGA CTA 3'

These oligonucleotides produced a linker with sticky overhangs to prevent self-annealing. Their sequence did not result in hairpin loops and the linker contained an EcoRI site to facilitate subsequent subcloning. The oligonucleotides were synthesised and chemically phosphorylated by Gibco-BRL. The linker was generated by mixing equal amounts of A and B, heating to 70 °C and then allowing the oligonucleotides to anneal gently to one another as the solution slowly returned to room temperature, this produced duplex kinase linkers. The linkers were then ligated to the undigested or digested cDNA (with an excess concentration of linkers in the solution) using blunt-ended ligation buffer to increase the efficiency (0.66 M Tris-Cl, pH 7.6, 50 mM MgCl2, 50 mM DTT, 1 mg/ml BSA, 10 mM hexaminocobalt chloride, 5mM spermidine-HCl) and the ligation solution was incubated at 16°C overnight.

The unincorporated linkers were separated from the linker ligated cDNA by gel electrophoresis using low melting point agarose. The linker ligated cDNA was excised from the gel (Fig. 5.4B) and the gel slice was used directly as the template in PCR. The shortest oligonucleotide (B) was used as the PCR primer (Fig. 5.4C). The excess primer and nucleotides were removed from the amplification product using a Microcon 30 column (Amicon) and this PCR amplified linkered cDNA was used as the starting material for hybrid selection.

The linkered undigested cDNA was used as the starting material because, as expected, it produced longer amplification products (smear ranged from 150 to 500bp compared to 150 to 300bp with AluI-digested cDNA) and, because the range was less than 1kb, concern about biased amplification of long transcripts was unnecessary.



C

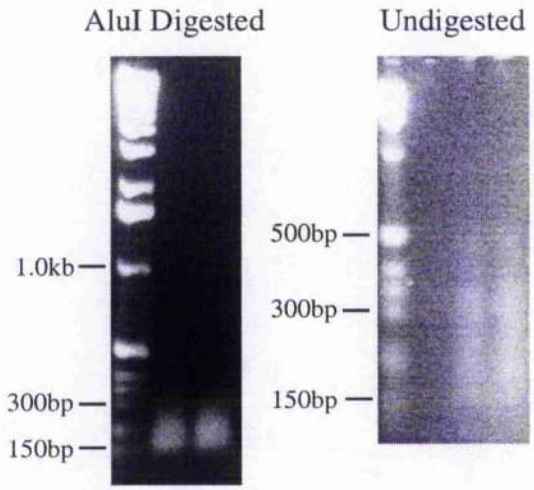


Fig. 5.4 A: Agarose gel indicating PCR amplification products obtained from double-stranded, blunt-ended cDNA (Lane 1 - 3ß-HSD type I primers, lane 2 - 3ß-HSD type I primers, no template control, lane 3 - GAPDH primers, lane 4 - GAPDH primers, no template control).
B: Linkered cDNA electrophoresed to remove unincorporated linkers. White box indicates the area of the gel slice excised containing the linkered cDNA.
C: Agarose gel indicating PCR amplification of AluI digested linkered cDNA and undigested linkered cDNA. Oligonucleotide B used as PCR primer; duplicate PCRs shown on each gel.

5.2.2 Preparing the genomic clone

The major advantage of using biotinylated genomic DNA and streptavidin magnetic beads is that both hybridisation partners are in solution, and this allows better control of the hybridisation kinetics compared to filter hybridisation as mentioned above. Also the magnetic beads (Dynabeads M-280, Dynal) used by Korn *et al.* (1992) and Morgan *et al.* (1994) are monodispersed and thus follow uniform kinetics when subjected to a magnetic field. Different solid supports have been tested for capturing the biotinylated genomic DNA including avidin-agarose, streptavidin-agarose and biotin cellulose but magnetic beads coated with streptavidin showed less non-specific binding of nucleic acids (Korn *et al.*, 1992)

There are different ways that biotin molecules can be introduced into genomic clones: random priming, photobiotinylation, nick translation or PCR amplification of the genomic clone using biotinylated primers. Nick translation has the advantage that the amount of product equals the amount of starting material and it fragments the genomic clone ready for hybridisation. Also, an even incorporation of biotin molecules is more desirable for optimum binding efficiency.

The genomic clone used in the first selection experiments was BAC 21. This is a large BAC clone (235kb), present at the left side of the cluster, containing only three 3ß-HSD genes and a lot of gaps which may contain something unexpected (see section 4.2.6). BAC 21 DNA was nick-translated using the nick translation kit from Gibco-BRL using a mixture of 0.18mM dTTP and 0.02mM Bio16dUTP (Boehringer Mannheim). One quarter of the reaction mix was electrophoresed on an 1% agarose gel, and a faint smear within the appropriate range (500bp to 2kb) was observed (Fig. 5.5). The remaining reaction mix was ethanol precipitated to remove unincorporated biotin molecules, and resuspended in sterile water.

It is interesting to note that although nick translation was the best protocol to use for BAC clones (BAC DNA is easy to isolate in large quantities), if YAC clones were used it would be preferable to digest the YAC DNA, ligate linkers onto the ends and PCR amplify using a biotinylated primer. This method has the advantage of starting from limited amount of DNA, although it may create a bias against longer fragments.

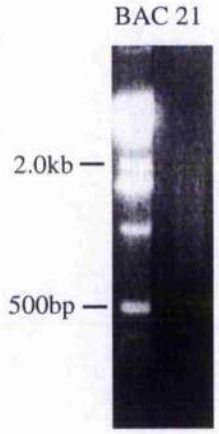


Fig. 5.5 Gel electrophoresis of biotinylated BAC 21 nick translation products. A faint smear can be observed between 2kb and 500bp.

5.2.3 Hybridisation

Before hybridisation, the cDNA was pre-blocked with Cot 1 DNA to reduce the chances of selecting repeated sequence. Equal amounts of PCR-amplified cDNA and COT 1 DNA were mixed, heated to 95°C and incubated with hybridisation buffer (final concentration - 0.75 mM sodium chloride, 20 mM sodium phosphate (pH 7.4), 5 mM EDTA, 5xDenhardt's solution, 0.6% SDS) for 4 hours at 65°C. The blocked cDNA was mixed 20:1 with the denatured biotinylated BAC and hybridisation buffer at 65°C for approximately 65 hours.

After 65 hours, the hybrids generated were isolated by Dynal streptavidin magnetic beads. This procedure was followed exactly as recommended by Dynal. Briefly, the hybridisation mix was incubated with the beads at room temperature for 15 minutes, the beads were subjected to 2 washes of 1xSSC, 1% SDS at room temperature and then a further 3 washes using 1xSSC, 0.1% SDS at 65°C. The cDNA was eluted using sodium hydroxide, neutralised by Tris and finally desalted using a Microcon 30 column (Amicon), then resuspended in sterile water.

The eluted cDNA was PCR-amplified using oligonucleotide B as PCR primer (Fig. 5.6A). The first round of selected amplified cDNA was purified by a Microcon column as above, then re-blocked with Cot 1 DNA and subjected to another cycle of selection with biotinylated BAC 21, exactly as before. The eluted cDNA from the second round of selection was also PCR-amplified, and this time distinct bands were beginning to emerge within the PCR products (Fig. 5.6B).

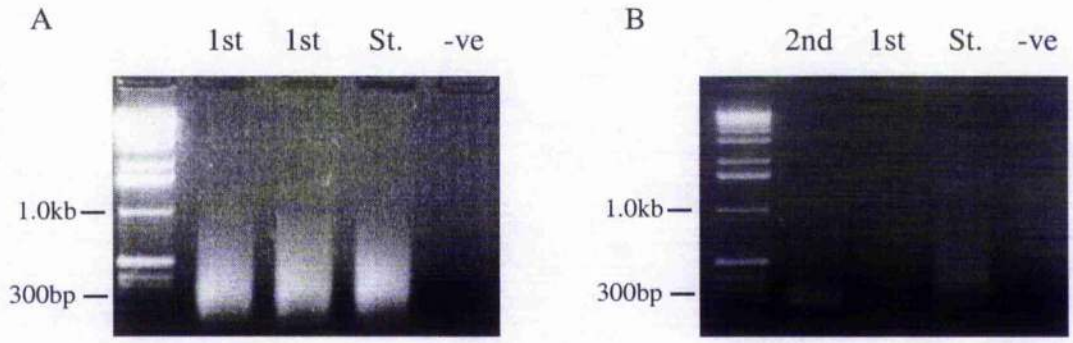


Fig. 5.6 A: Agarose gel indicating PCR amplification of eluted cDNA from first round of selection.. Oligonucleotide B was used as the primer. St. indicates starting cDNA population and -ve indicates no template PCR control.
B: Agarose gel indicating PCR amplification of eluted cDNA from second round of selection. Oligonucleotide B was used as the primer. St. indicates starting cDNA population and -ve indicates no template PCR control.

5.2.4 Analysis of eluted cDNA

The amplified cDNA eluted from the second round of selection was cloned into T-vector. The resulting subclones were examined for inserts by single-colony PCR (Fig. 5.7A). Clones were selected for sequencing by hybridising the single colony PCR products to a 3ß-HSD cDNA probe, a Cot1 probe and a BAC 21 hybridisation probe (Fig. 5.7 B,C and D). Clones that were negative for 3ß-HSD and Cot1 but positive for BAC 21 were deemed "interesting" and were therefore sequenced. Some 3ß-HSD

clones were sequenced to confirm their sequence and also to check for unknown sequences. Sequencing data are discussed in section 5.2.6.

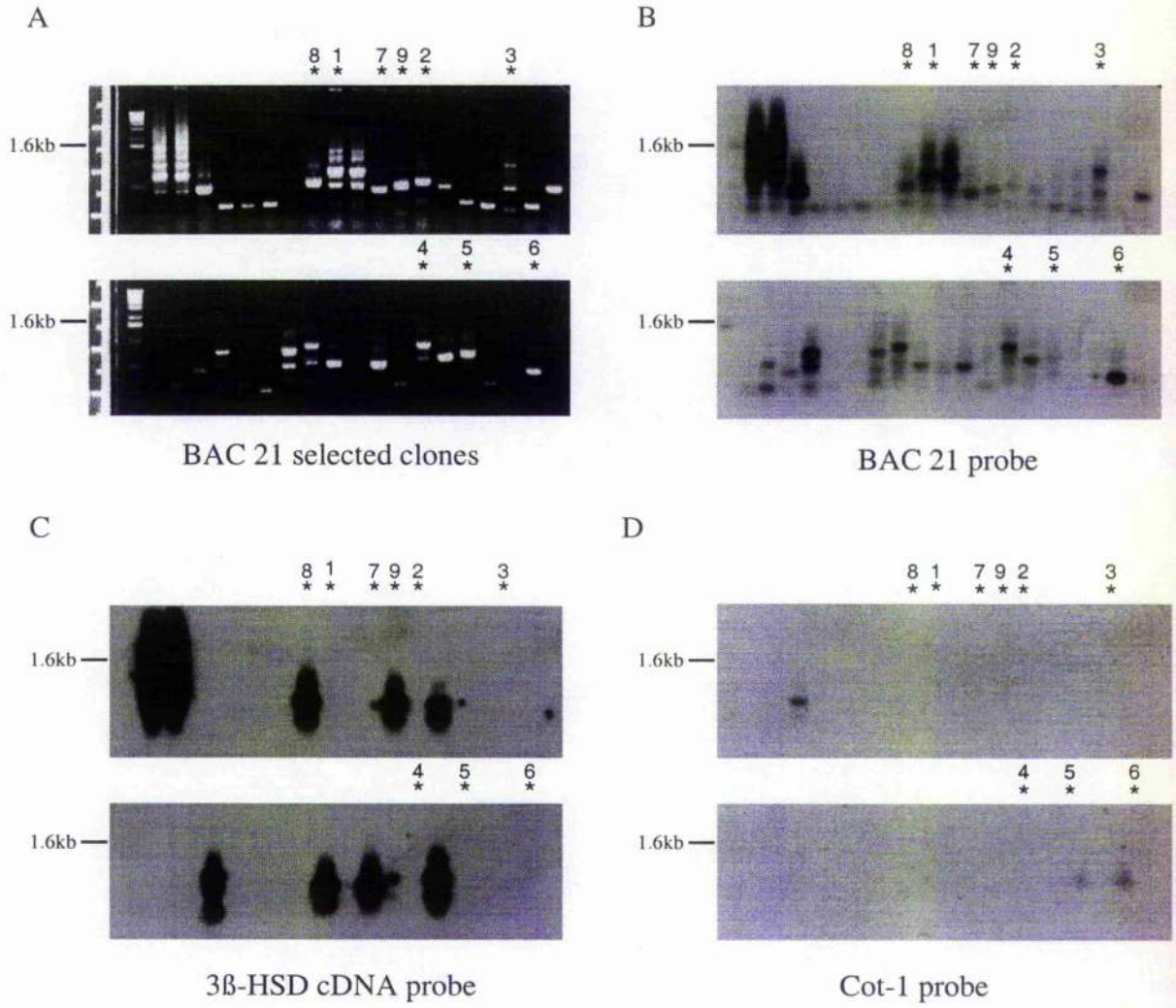
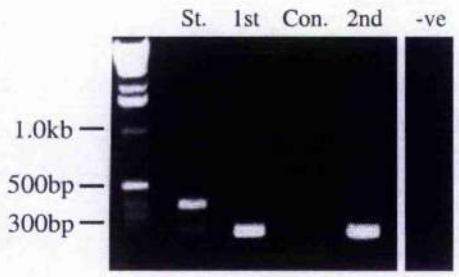


Fig. 5.7 A: Agarose gel of single colony PCR products from the BAC 21 subclones, generated from PCR-amplified second round of selection DNA. This selected for BAC 21 subclones with inserts. Primer used was oligonucleotide B (section 5.2.1).
B: Southern analysis of BAC 21 selected cDNA clones (panel A) hybridised with BAC 21 probe
C: Southern analysis of BAC 21 selected cDNA clones (panel A) hybridised with 3ß-HSD cDNA probe
D: Southern analysis of BAC 21 selected cDNA clones (panel A) hybridised with Cot-1 DNA probe
* represents the clones that were sequenced.

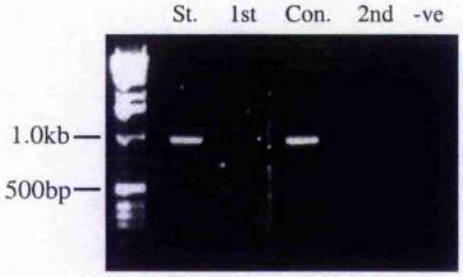
The eluted cDNA was investigated for enrichment of 3ß-HSD sequences and loss of non-specific transcripts by PCR using 3ß-HSD specific primers and GAPDH primers. It was expected that the 3ß-HSD signal would increase and the GAPDH signal would decrease in the first and second rounds of selection, and this is what was observed (Fig. 5.8). The starting cDNA population resulted in 2 products from the 3ß-HSD specific primers, one at 400bp and the other at 270bp (Fig. 5.8A). These represent the genomic and the cDNA copy of 3ß-HSD respectively. The primer pair used (type I specific and D; section 2.6.2) span across intron I which is present in the 400bp genomic product but absent in the 270bp cDNA product. This could suggest genomic contamination in the cDNA library, or it may be due to RNA transcripts which have not been spliced correctly. This is likely to be due to the high nucleus:cytoplasm ratio found in the placenta. The 400bp genomic product disappears in the eluted cDNA from the first and second round of selection. Only the 270bp cDNA product is observed; the reason for this is unknown.

A control experiment was undertaken in which the genomic clone was nick translated as described, but without the addition of biotinylated dUTP. The hybrids created in this experiment should not bind to the magnetic beads; however, an amplification product was observed from the eluted material using oligonucleotide B as primer (Fig. 5.8C). This eluate did not amplify with the 3ß-HSD specific primers (Fig. 5.8A), and therefore the amplification products observed in Fig. 5.8C probably represent non-specific binding of DNA to the magnetic beads. This was supported by a positive amplification using the GAPDH primers, similar to the starting cDNA. No GAPDH product was observed from either of the selected cDNAs (Fig. 5.8B).

A - 3ß-HSD specific primers



B - GAPDH specific primers



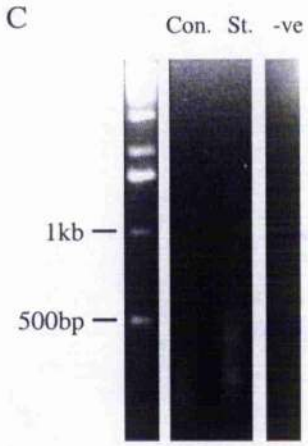


Fig. 5.8 A: Agarose gel showing PCR amplification of cDNA populations using 3ß-HSD specific primers
B: Agarose gel showing PCR amplification of cDNA populations using GAP-DH specific primers
C: Agarose gel showing PCR amplification of control experiment with no biotinlyated dUTP in the nick translation reaction mix. Oligonucleotide B used as primer.
(St. - starting cDNA, 1st - cDNA eluted from 1st round of selection, Con - cDNA eluted from control experiment, 2nd - cDNA eluted from 2nd round of selection, -ve - no template control)

These data indicate that the 3ß-HSD transcripts (specific) were enriched after the first and second round of selections and that the GAPDH transcripts (non-specific) were lost after selection by BAC 21. This is what we would expect, and it provides an initial indication that the hybrid selection experiments have worked.

5.2.5 Further experiments

The hybrid selection protocol was repeated for BAC clones 7 and 30 allowing the entire 3ß-HSD cluster from the BAC contig to be investigated. The placental cDNA library was used as starting material and the procedure was repeated exactly as described for BAC 21. BAC 7 and BAC 30 DNA was biotinylated as before (Fig. 5.9A), first and second rounds of selection were undertaken and the second round of amplified eluted cDNA was cloned into T-vector (Fig. 5.9B). Single colony PCRs indicated subclones with inserts, and these were screened with the specific BAC clone hybridisation probe and the 3ß-HSD cDNA probe (Figs. 5.10 and 5.11). Again, clones negative for 3ß-HSD and positive for BAC clone 7 or 30 were sequenced. The sequence data are discussed in section 5.2.6.

One further experiment was carried out as a control to analyse the enrichment of 3ß-HSD transcripts from these selection experiments. Using O.D. estimates of DNA concentration, equal amounts of starting cDNA, first cycle of selection cDNA and second cycle of selection cDNA were applied in triplicate to a slot blot and hybridised with 3ß-HSD type I cDNA (Fig. 5.12). These clearly showed that 3ß-HSD transcripts were enriched in the first round compared to the starting cDNA, and even more after the second round of selection. The same slot blots were re-probed with a GAPDH hybridisation probe to illustrate a decreased signal corresponding to selection; however the amount of DNA in the slot blots was too low to detect any signals. This is similar to the situation observed with the starting cDNA hybridised with 3ß-HSD cDNA. However, the subsequent enrichment of 3ß-HSD transcripts is clear and provides an indication that the selection process has been successful.

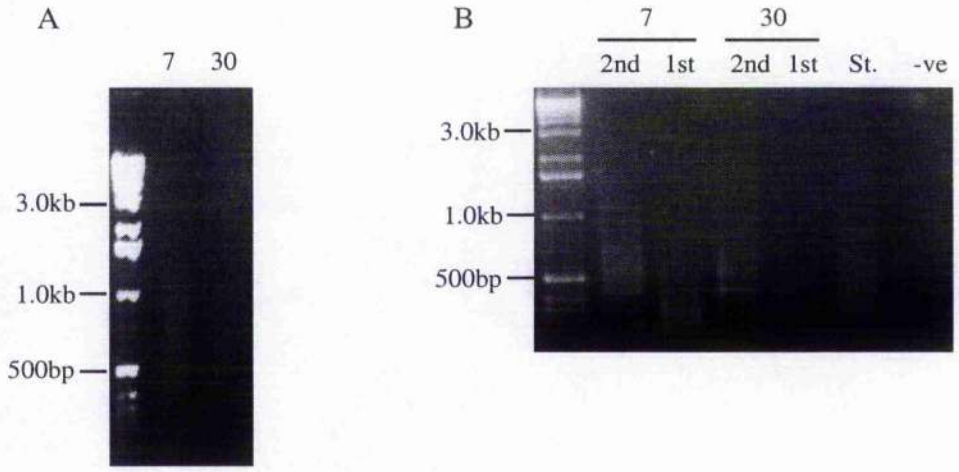


Fig. 5.9 A: Gel electrophoresis of biotinylated BAC 7 and 30 nick translation products. A faint smear can be observed between 2kb and 500bp.
B: Agarose gel of PCR amplified 1st and 2nd round eluted cDNA for BACs 7 and 30 (St. - starting cDNA, -ve - no template control)

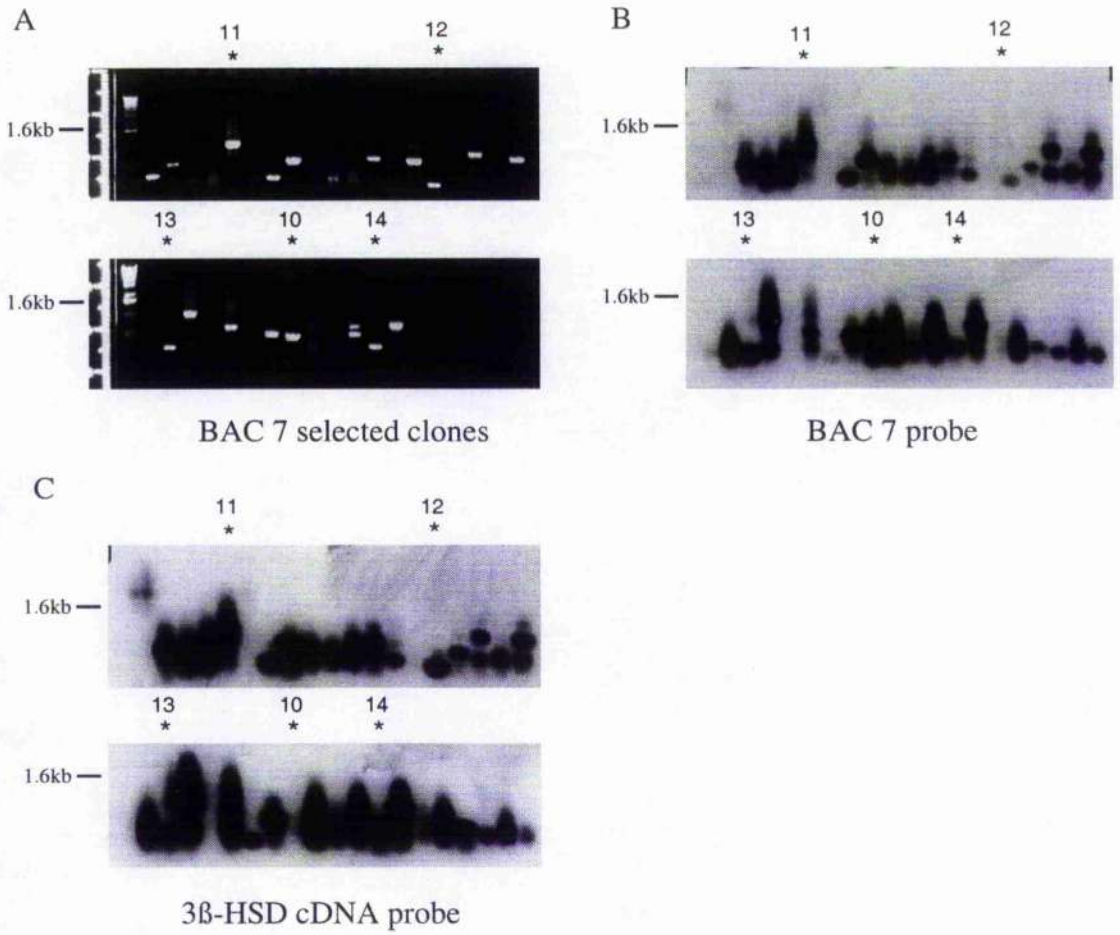


Fig. 5.10 A: Agarose gel showing single colony PCR amplification products from the BAC 7 selected clones , generated from PCR-amplified second round of selection cDNA. Primer used was oligonucleotide B.
B: BAC 7 selected clones (panel A) hybridised with BAC 7 probe
C: BAC 7 selected clones (panel A) hybridised with 3ß-HSD cDNA probe
* represents the clones that were sequenced.

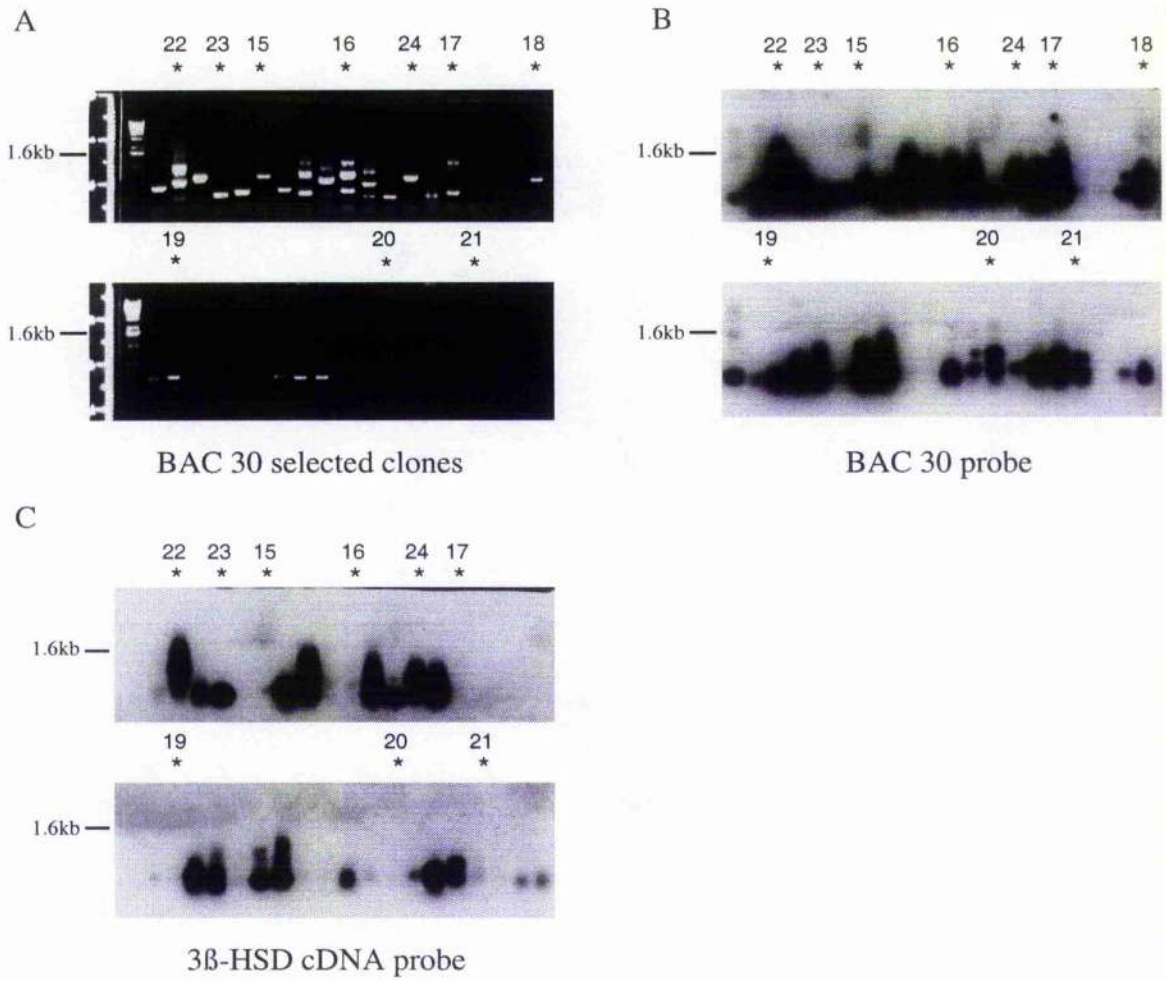


Fig. 5.11 A: Agarose gel of single colony PCR amplification products from the BAC 30 selected clones, generated from PCR-amplified second round of selection cDNA. Primer used was oligonucleotide B.
B: BAC 30 T-vector selected clones hybridised with BAC 30 probe
C: BAC 30 T-vector selected clones hybridised with 3ß-HSD cDNA probe
* represents the clones that were sequenced.

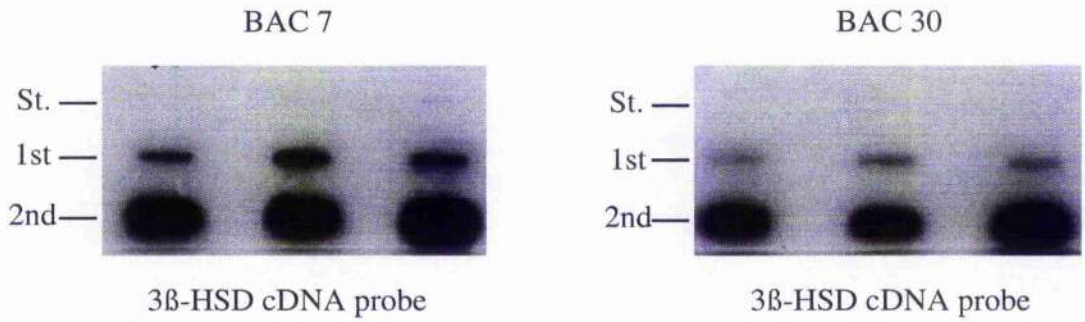


Fig. 5.12 Slot blot (in triplicate) of equal amounts of the starting and the selected populations of cDNA hybridised with 3ß-HSD cDNA probe (St. - starting cDNA, 1st - first round of selected cDNA, 2nd - second round of selected cDNA).

5.2.6 Sequencing data

BAC 21:

From the BAC 21 hybrid selection experiment, 38 T-vector subclones containing products from the second round of selection were analysed (Fig. 5.7). Of these, 9 clones hybridised to the 3ß-HSD cDNA probe, 15 clones were positive for BAC 21 and negative for 3ß-HSD cDNA (although some of these clones only produced a weak signal for BAC 21) and 1 clone contained COT-1 repeated sequences. Some of these clones were sequenced; these are marked * and numbered on Fig. 5.7. Clones 1-7 were "unknown" (i.e. positive for BAC 21 but negative for 3ß-HSD and COT-1) and clones 8 and 9 were 3ß-HSD cDNA positive.

The sequence obtained from clone 8 was 3ß-HSD type I which was expected, however clone 9 contained 3ß-HSD type II sequence. Until now, it was believed that type I was the transcript expressed exclusively in the placenta (Labrie *et al.*, 1991), however type II transcripts were detected at low concentrations from the placenta by RT-PCR (Sandra Burridge, personal communication). Therefore, these data have determined that although type I 3ß-HSD is the major expressed isoform of 3ß-HSD in the placenta, type II transcripts are present and can be detected.

The sequence from clones 1 and 4 resulted in type I 3ß-HSD also. However, the sequence obtained was from intron II, which explains why these clones did not hybridise with the 3ß-HSD cDNA probe. These sequences may have arisen from genomic DNA contamination within the cDNA library or from unspliced RNA as discussed earlier. Sequence from clone 2 aligns precisely with exon 7 of IGF-II nucleotide sequence (Fig. 5.13A), sequence from clone 3 aligns with human mitochondrial DNA for loop attachment (Fig. 5.13B), and sequence from clone 5 is human placental lactogen hormone sequence (Fig. 5.13C).

A:

```
Seq.2      79 ACACACGCATGCACAGCACACATGAACACAGCACACACAAACACACAGCA 128
              ||||||||||||||||||||||||||||||||||||||||||||||||||
IGF-II   1631 ACACACGCATGCACAGCACACATGAACACAGCACACACAAACACACAGCA 1680

Seq.2     129 CACACATGCACACACAGCACACACACTCATGCGCAGCACATACATGAACA 178
              |||||||||||||||||||||||||||||||||||||||||||||||||
IGF-II   1681 CACACATGCACACACAGCACACACACTCATGCGCAGCACATACATGAACA 1730

Seq.2     179 CAGCTCACAGCACACAAACACGCAGCACACACGTTGCACACCCAAGCACC 228
              ||||||||||||||||||||||||||||||||||||||||||||||||||
IGF-II   1731 CAGCTCACAGCACACAAACACGCAGCACACACGTTGCACACGCAAGCACC 1780

Seq.2     229 CACCTGCACACACACATGCGCACACACACGCAC 261
              |||||||||||||||||||||||||||||||||
IGF-II   1781 CACCTGCACACACACATGCGCACACACACGCAC 1813
```

B:

```
Seq.3     27 TANCACGATTAACCCAAGTCAATAGAAGCCGGCGTAAAGAGTGTTTTAGA 76
             | :|||||||||||||||||||||||||||||||||||||||||||||||
Mito.    327 TCACACGATTAACCCAAGTCAATAGAAGCCGGCGTAAAGAGTGTTTTAGA 376

Seq.3     77 TCACCCCCTCCCCAATAAAGCTAAAACTCACCTGAGTTGTAAAAAACTCC 126
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Mito.    377 TCACCCCCTCCCCAATAAAGCTAAAACTCACCTGAGTTGTAAAAAACTCC 426

Seq.3    127 AGTTGACACAAAATAGACTACGAAAGTGGCTTTAACATATCTGAACACAC 176
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Mito.    427 AGTTGACACAAAATAGACTACGAAAGTGGCTTTAACATATCTGAACACAC 476

Seq.3    177 AATAGCTAAGACCCAAACTGGGATTAGATACCCCACTATGCTTAGC 222
             |||||||||||||||||| ||||||||||||||||||||||||||
Mito.    477 AATAGCTAAGACCCAAACTGGGATTAGATACCCCACTATGCTTAGC 522
```

C:

```
Seq.5      1 AGCCGCCGGACTGGGCAGATCCTCAAGCAGACCTACAGCAAGTTTGACNC 50
             ||||||||||||||||||||||||||||||||||||||||||||||:|
HPL      501 AGCCGCCGGACTGGGCAGATCCTCAAGCAGACCTACAGCAAGTTTGACAC 550

Seq.5     51 AAACTCGCACAACCATGACGCACTGCTCAAGAACTACGGGCTGCTCTACT 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||
HPL      551 AAACTCGCACAACCATGACGCACTGCTCAAGAACTACGGGCTGCTCTACT 600

Seq.5    101 GCTTCAGGAAGGACATGGACAAGGTCGAGACATTCCTGCGCATGGTGCAG 150
             ||||||||||||||||||||||||||||||||||||||||||||||||||
HPL      601 GCTTCAGGAAGGACATGGACAAGGTCGAGACATTCCTGCGCATGGTGCAG 650

Seq.5    151 TGCCGCTCTGTGGAGGGCA 169
             ||||||||||||||||||
HPL      651 TGCCGCTCTGTGGAGGGCA 669
```

Fig. 5.13 A: Alignment of clone 2 with exon 7 of IGF-II (Genbank no. - x07868).
B: Alignment of clone 3 with human mitochondrial DNA for loop attachment sequence (Genbank no. - x62996).
C: Alignment of clone 5 with human placental lactogen sequence (Genbank no. j00118).

These sequences probably arose from non-specific binding to the Dynal Streptavidin magnetic beads. We know from the control experiment excluding biotin that some non-specific binding of DNA to the Dynal Streptavidin beads did occur (positive PCR amplification from eluted cDNA, Fig. 5.8C) and IGF-II has been localised to human chromosome 11, mitochondrial DNA is extra-chromosomal and human placental lactogen (HPL) is present on chromosome 17. However, it is not necessarily surprising that HPL was selected, as HPL is the major protein synthesised in term placental tissue. Seeburg *et al.*(1977) estimated that HPL mRNA constitutes 5% of poly (A)+ isolated from the placenta .

It was formally possible that sequences related to IGF-II, mitochondrial DNA and HPL are present on BAC 21. However, this is quite unlikely as the signal of these clones for BAC 21 was quite weak (Clones 2, 3 and 5 in Fig. 5.7B). To determine if this was the case, blots would have to be attempted with larger probes from these genes to

BAC 21 to assess the presence of long stretches of homology. To prevent non-specific binding of DNA to the magnetic beads, Dynal recommend treating the particles with 50 mM of sodium pyrophosphate, either alone or as a component in the binding and / or washing buffer. Therefore, for further experiments this may have to be included.

The sequence obtained from clones 6 and 7 did not match any identified sequences deposited into the databases. The sequence from clone 6 revealed that two cDNA fragments were cloned into T-Vector simultaneously since a perfect inverted repeat of linker sequence was identified within the sequence (Fig. 5.14). Bases 1-139 represent one cDNA frgament (cDNA A) and bases 182-337 represent the other (cDNA B).

```
  1  TCTATCCATC CCATCATCCC ATCCATCCAT CCTATCCATC CATCCCATCC
 51  ATCCTATCCA TCCCATCCAT CCATCCATTC ATCCCATCCA TCCCATCCAT
101  TCATCTATCC ATCCATCCTA TCCATCCCGT CCATTCATCT AGTCCGAATT
151  CAAGCAAGAG CTCTTGCTTG AATTCGGACT ACAGAGGCCC GCTGTTATGC
201  TNNNGCTCAC AGGCTCACCC TGACAGTCCT GGGGCTCCCC CTCGTGGTTG
251  GTTTTAAGAA AGGGCGTGAC ATGGTGAGAC TAGTTTGGAG AGAACTCTGG
301  CAACACTGCT GAGGATAGGT TTCATGGGTG ACTCTCC
```

Fig. 5.14 Sequence obtained from subclone 6 using vector primer T7. Boxed sequence outlines linker sequence. Bases 1-139 represents cDNA fragment A and bases 182-337 represents cDNA fragment B.

cDNA fragment A appears to be a repeated sequence, therefore, it is unlikely that this represents part of an expressed gene. However, additional sequence data would be necessary to confirm this. cDNA fragment B did not match anything in the nucleic acid databases, and before further analysis of this sequence can be undertaken, more sequence data will need to be obtained. Clone 7 sequence is indicated in Fig. 5.15 and similar to cDNA fragment B, additional sequence data is essential before further sequence analysis can take place.

```
  1  TTTAGACAGG ACTGGGTATT GGCACATAAT TCCCAAGCAC ACACTTTCTC
 51  AAGCCAGAAA CCTGGATGGC AGCCCCCCAC TCCGCCCACC TATCATTCAG
101  CGCTACCAAC TTTACCTCCA CAACACCTCT TGAATGTGT
```

Fig. 5.15 Sequence obtained from subclone 7 using vector primer T7.

BAC 7:

For BAC 7, only one "unknown" clone was detected (negative for 3ß-HSD but positive for BAC 7) with 32 3ß-HSD clones from a total of 38 subclones analysed (Fig. 5.10). The "unknown" clone and only 4 3ß-HSD clones were sequenced, clones 10 and 11-14 respectively.

Clone 10 contained 3ß-HSD type I intron sequence, similar to clones 1 and 4 from the BAC 21 selection experiment. Clones 11-14 all contained 3ß-HSD type I as expected. It was not unexpected that BAC 7 has selected only 3ß-HSD sequences. This BAC clone contains 4 3ß-HSD genes and part of HSD3Bψ4 (section 4.2.7) and the restriction mapping for BAC 7 indicated that there were not many gaps in this clone where new sequences could be found.

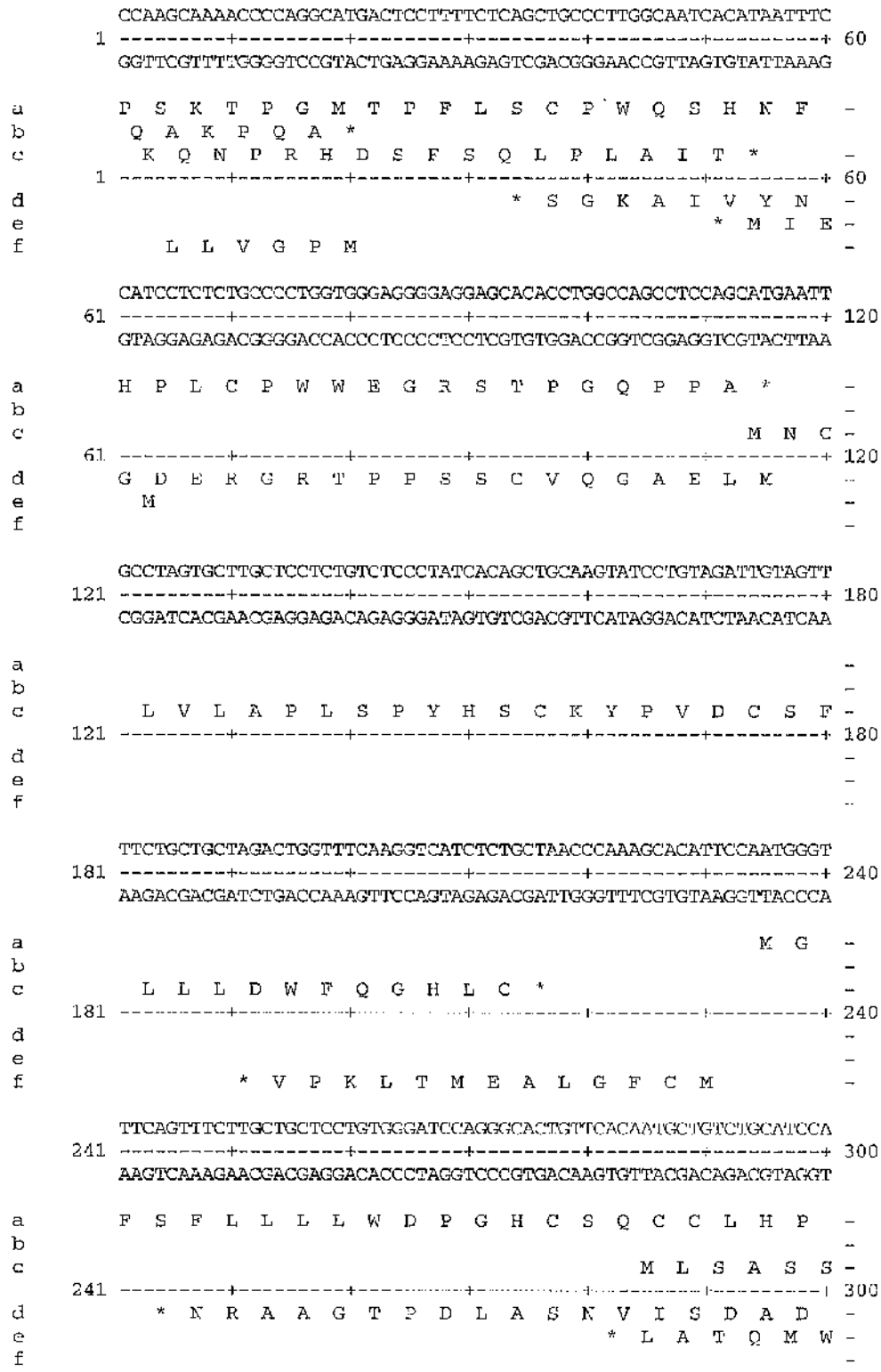
BAC 30:

The BAC 30 hybrid selection experiments produced 12 "unknown" clones (negative for 3ß-HSD and positive for BAC 30) and 19 3ß-HSD clones (Fig. 5.11). 7 "unknown" clones were sequenced, clones 15-21, and 3 3ß-HSD clones, clones 22-24.

The 3 3ß-HSD clones sequenced, clones 22-24, contained 3ß-HSD type I sequence as expected, and clones 18, 19 and 21 contained 3ß-HSD type I intron sequence as described previously for clones 1 and 4. Clones 16 and 17 produced 3ß-HSD intron sequence also, but not from 3ß-HSD type I or type II. It is likely that these clones contain sequence that belongs to one of the identified pseudogenes, however it is impossible to identify which gene as the intron sequence for each of the pseudogenes is not known.

Clone 15 and clone 20 contained the same new sequence, which was not 3ß-HSD sequence and did not match any of the sequences in the nucleotide databases. Fig. 5.16 presents this sequence and indicates the open reading frames present within the fragment. No significant open reading frames (ORF) can be detected, although the ORF in reading frame A contains 48 amino acids and ends only because the sequence

data ends and not because of a stop codon. Similar to subclones 6 and 7 from the BAC 21 selection experiments, additional sequence data will have to be obtained before further analysis can be undertaken. This sequence is different from the sequence of either of the unidentified clones selected by BAC 21 (clone 6 or 7).



```
        GCTGGCTAGT
    301 ---------+ 310
        CGACCGATCA

a       A  G  *    -
b                  -
c         W  L     -
    301 ----------+ 310
d       L  Q  S  T -
e         S  A  L  -
f            *     -
```

Fig. 5.16 Sequence of subclones 15/20 indicating the open reading frames identified from all 6 frames (a- f).

5.3 CONCLUSIONS

Fig. 5.17 summarises the sequence data obtained from the BAC 21, BAC 7 and BAC 30 hybrid selection experiments. The selection experiments did not detect any previously identified genes that could be localised to chromosome 1p13, although sequences were obtained that *may* be part of an unidentified gene. These sequences (from clones 6, 7 and 15/20) do not resemble sequences deposited into the databases. Future studies should examine these in more detail; obtain more sequence, confirm that they originate from their respective BAC clones by hybridising longer stretches of sequence to the genomic clone involved, and then isolate the cDNA or genomic clones if required.

It is clear that the selection experiments have been successful. Enrichment of 3ß-HSD transcripts in the selected cDNA can be detected by PCR (Fig. 5.8A) and in slot blot analysis (Fig. 5.12), and a decrease of GAPDH transcripts was observed from PCR (Fig. 5.8B). However, it could be refined slightly; for example, by adding sodum pyrophosphate to reduce non-specific binding of DNA to the magnetic beads and restricting the priming of cDNA synthesis to oligo dT to decrease the genomic contamination in the cDNA library. In addition, the study of the 3ß-HSD contig should be extended to establish whether other genes are present within or surrounding the 3ß-HSD gene cluster. Additional clones should be sequenced and analysed from the BAC selections and the YAC clones should be utilised, in particular YAC 1GD11 which extends furthest from the cluster.

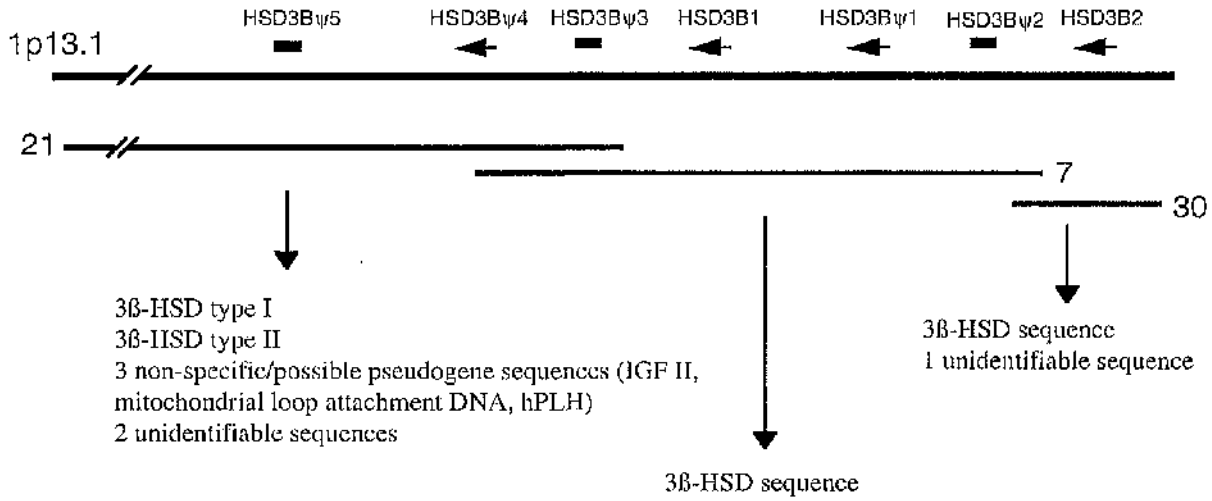


Fig. 5.17 Summary of sequence data obtained from selection experiments involving BAC clones 21, 7 and 30.

6.1 Characterisation of new members of the 3ß-HSD gene family

Two forms of 3ß-hydroxysteroid dehydrogenase Δ5/Δ4 isomerase (3ß-HSD) have been identified in humans and these enzymes are products of two distinct genes, HSD3B1 and HSD3B2 (Lorence *et al.*, 1990a, 1990b, Rheaume *et al.*, 1991, Lachance *et al.*, 1990, 1991). However, several observations suggested that more than two HSD3B genes existed in the human genome. Human genomic DNA digested with a range of restriction enzymes and hybridised with 3ß-HSD type I cDNA probe indicated several hybridisation bands. The pattern and intensity of these bands indicated the existance of additional HSD3B sequences possessing a high degree of similarity to HSD3B1 and 2 (Russell, 1993, Lachance *et al.*, 1990). Furthermore, novel 3ß-HSD sequences were discovered during mutation screening experiments within our laboratory (Russell, 1993). 3ß-HSD exon 3 PCR amplification products were analysed by denaturing gradient gel electrophoresis (DGGE) and products with different migration patterns from type I or type II were detected. These products were subsequently cloned and sequenced, and novel 3ß-HSD sequences were revealed. This information prompted our laboratory to screen two λgem11 human genomic libraries for sequences similar to 3ß-HSD type I and II (McBride *et al.*, 1995, McBride, 1996).

The result of the library screens was the identification of 5 new 3ß-HSD sequences. The library screens were undertaken by Martin McBride, 2 of the sequences identified were characterised by Martin (ψ1 and ψ4), 2 were characterised by myself (ψ2 and ψ3, Chapter 3) and the fifth sequence was analysed by Nicola Craig (ψ5). The 5 sequences identified indicated a 3ß-HSD gene family with 7 members including type I and II, with 84.4% - 94.5% nucleotide identity to the coding sequence of HSD3B1 (Table 6.1). The order of the exons were the same and the lengths of the introns were indistinguishable from HSD3B1 and 2 (sections 3.2.4 and 3.2.9). However, although the sequences contained ATG codons in an equivalent position to the initiator codon of HSD3B1, each sequence had stop codons within the sequences corresponding to the reading frames of HSD3B1 and 2. HSD3Bψ4 and ψ5 had stops at residue 16, ψ2 had a stop at residue 42, ψ3 had a stop at residue 143 and ψ1 had a stop at residue 173 (see appendix ii). RNA products were detected by RT-PCR from ψ4 and ψ5 and the transcripts showed alternative splicing that created frameshifts between exons 3 and 4 and/or inserted intron sequence (Sandra Burridge, personal communication). The location of mutations in HSD3B2 associated with 3ß-HSD deficiency suggested that more than the first 318 amino acids out of 371 are essential for the protein to function as a 3ß-HSD enzyme (Morel *et al.*, 1997, section 1.8.3).

We therefore concluded that the new sequences identified were unprocessed pseudogene members of the 3ß-HSD gene family.

	Type II	Type I	19-4 3ß-HSD ψ4	4-3 3ß-HSD ψ5	2-7 3ß-HSD ψ3	24-4 3ß-HSD ψ1	8-3 3ß-HSD ψ2
Type II	-	93.5%	90.5%	89.4%	91.5%	88.24%	85.4%
Type I		-	91.6%	91.6%	91.5%	88.8%	87.0%
19-4 3ß-HSD ψ4			-	94.5%	91.3%	87.3%	85.4%
4-3 3ß-HSD ψ5				-	89.4%	90.3%	88.3%
2-7 3ß-HSD ψ3					-	88.3%	84.4%
24-4 3ß-HSD ψ1						-	88.7%
8-3 3ß-HSD ψ2							-

Table 6.1. Nucleotide sequence similarity of clone segments homologous to the 3ß-HSD type I and II exons.

6.2 3ß-HSD cluster at chromosome 1p13.1

Each of the pseudogenes were localised to chromosome 1p13 by fluorescent *in situ* hybridisation (FISH) by Norma Morrison from the Medical Genetics department at Yorkhill hospital (McBride *et al.*, 1995, personal communication). Therefore, the next stage in the analysis of the 3ß-HSD gene family was to attempt to order the genes by mapping each sequence to clones isolated from YAC and BAC genomic libraries. Individual HSD3B genes were assigned to YAC and BAC clones using locus specific PCR primers (Fig. 6.1; sections 4.2.2 and 4.2.3). The assignment of the genes to BAC clones was validated by restriction fragment analysis. Each BAC was digested with a variety of restriction enzymes and hybridised to conserved exon 2 and exon 4 oligonucleotides (section 4.2.4). Sequence conservation frequently yielded restriction fragments of identical size. These fragments were detected by the presence of HSD3B bands at identical postions from non-overlapping BAC clones. Their identities were either confirmed by probing with specific oligonucleotides or the bands in question were subcloned and sequenced. The orientation of HSD3B1, 2, ψ1 and ψ4 was determined by the position of restriction sites and the positions of the breakpoints in BAC clones 29, 7 and 31 (section 4.2.7). Lack of time and convenient restriction sites prevented the determination of the orientation of ψ2, ψ3 and ψ5.

The estimated length of the entire contig is 500kb with the 3ß-HSD gene cluster over a centrally-placed 235kb fragment (Fig. 6.1). This is consistent with the previous estimate of 290kb based on SacII restriction analysis (Morrissette *et al.*, 1995). YAC 1GD11, contains HSD3Bψ2 and HSD3B2 and is estimated to extend a further 170kb beyond HSD3B2. Restriction digests of YAC 1GD11 hybridised with full-length type I cDNA showed only two bands, which suggested that only the two 3ß-HSD genes identified were present on this clone (section 4.2.6). Restriction digest analysis and hybridisation of conserved oligonucleotide probes in exon 2 and 4 to the BAC clones contained within the genomic contig failed to detect further HSD3B sequences. Each BAC clone indicated the expected number of bands for the genes already identified (section 4.2.5). These data suggest that the cluster is probably complete as far as sequences homologous to HSD3B are concerned, especially as 7 x genome equivalents were examined during the library screens to find the 5 HSD3B pseudogenes and HSD3B1 and 2.

The functional genes HSD3B1 and 2 are in direct repeat, approximately 90kb apart and are separated by pseudogenes ψ1 and ψ2. This is the first report of the physical distance between HSD3B1 and 2; furthermore, the arrangement of these two genes in direct repeat and separated by a 90kb fragment containing HSD3Bψ1 and 2 precludes these expressed genes from having a shared promoter. However, the present data does not exclude the possibility of a shared enhancer for 3ß-HSD gene expression, similar to the locus control region of the globin genes (Grosveld *et al.*, 1993).

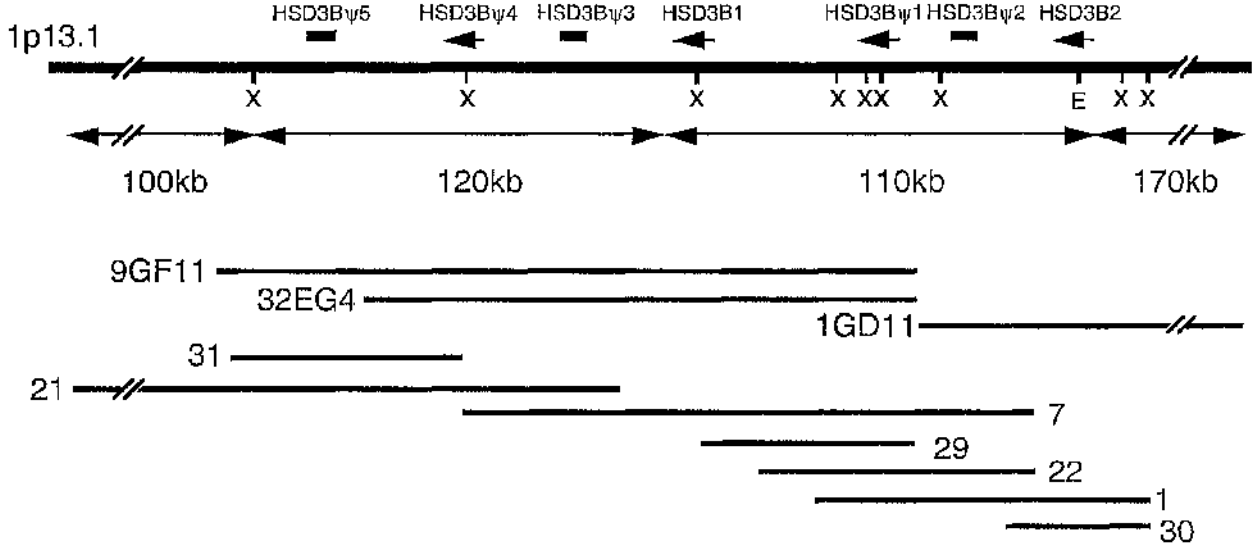


Fig. 6.1 Genomic map of 3ß-HSD gene family at human chromosome 1p13.1, constructed by restriction analysis of overlapping YAC and BAC clones.

6.3 Hybrid selection

Hybrid selection experiments were established to examine the 3ß-HSD gene cluster for novel, unidentified sequences (Chapter 5). These could be further members of the 3ß-HSD gene family or unrelated genes closely linked to the cluster. Briefly, the selection experiments were attempted by generating a placental PCR-amplifiable cDNA library which was hybridised to biotinylated genomic clones. The genomic clone-cDNA hybrids created were isolated by streptavidin magnetic beads, PCR amplified, cloned and sequenced. Three BAC clones were chosen to use in the selection experiments (BACs 21, 7 and 30) to allow the entire 3ß-HSD gene cluster to be investigated. The selection experiments resulted in successful enrichment of HSD3B sequences which was observed by PCR and slot blot analysis, however, no previously identified gene sequences were localised to this region of the HSD3B family. Novel sequences were detected but it was not resolved whether these sequences were from unidentified genes or not. To determine whether these sequences represent unidentified genes further experiments would be essential; these would include additional sequence analysis of the clones obtained, confirmation that the sequences originate from the genomic clone in question and isolation of cDNA or genomic clones. In addition, the hybrid selection experiments could be expanded to include YAC 1GD11, which may detect further sequences at the HSD3B2 end of the cluster; and cDNA libraries from other tissues could be included in the screen, for example, the prostate, the liver and other steroidogenic tissues.

6.4 3-ketosteroid reductase and C_{27}-3ß-HSD

Rodent 3ß-HSD families contain genes which encodes for 3-ketosteroid reductase enzymes (de Launoit *et al.*, 1992, Clarke *et al.*, 1993 and Abbaszade *et al.*, 1995), these enzymes convert 5-keto-5α-androstanes into 3ß-hydroxysteroids. From this study, no evidence of this enzyme in humans was detected. There is considerable data in the literature that the human prostate gland contains 3ß-ketosteroid reductase activity involved in the inactivation of the potent androgen dihydrotestosterone (Amet *et al.*, 1992; Trapp *et al.*, 1992; Abalain *et al.*, 1989; Tunn *et al.*, 1990). The 3ß-HSD enzyme (type I, type II or unidentified isoform) expressed in human prostate was unknown, however, RT-PCR analysis of the prostate gland for 3ß-HSD transcripts resulted in the detection of type II 3ß-HSD (section 3.2.12). 3ß-HSD type II does have KSR activity when expressed from cDNA in transfected tissue culture cells with NADH the preferred cofactor (Lorence *et al.*, 1990, Rheaume *et al.*, 1991). However, the ketosteroid reductase activity in prostate was shown to have a three-fold

preference for NADPH as cofactor (Trapp *et al.*, 1992), which is the preferred cofactor for the rodent KSR enzymes. If there is a human enzyme with exclusive ketosteroid reductase activity, it is reasonable to conclude that it is not a member of the 3ß-HSD family located in the cluster at chromosome 1p13.1 and that it is sufficiently diverged to be undetected by homology probing. This is in contrast to the mouse KSR enzymes which have been shown to be part of the cluster identified on mouse chromosome 3, a region syntenic with human chromosome 1p13 (Bain *et al.*, 1993). Using 3ß-HSD PCR primers conserved throughout the human gene family identified, human-mouse somatic cell hybrids did not reveal any 3ß-HSD sequences outwith chromosome one (section 4.2.8). However, if the KSR is sufficiently diverged then it may not be detected by these experiments.

C_{27}-3ß-HSD is an enzyme involved in bile acid synthesis and a deficiency of this enzyme can lead to abnormal bile acid synthesis and fatal liver damage. Buchmann *et al.* (1990) described familial giant cell hepatitis associated with deficiency of C_{27}-3ß-HSD. C_{27}-3ß-HSD catalyses the same reaction as $C_{19/21}$-3ß-HSD and it is generally believed that the gene for this enzyme is a member of the 3ß-HSD gene family, although the gene has not yet been cloned. However, by studying the segregation of HSD3B1 and 2 in a consanguineous family affected by C_{27}-3ß-HSD deficiency, (Russell *et al.*, 1995) showed that the activities of the C27 and the C19/21 enzymes are encoded by genes that are not in genetic linkage. Therefore, it is probable that this member of the 3ß-HSD gene family is not present within the cluster identified on chromosome 1p13.1.

It is likely that genes for C_{27}-3ß-HSD and the 3-KSR enzyme would be identified by expression cloning rather than experiments involving homology with the known members of the family. This has been described within the 17ß-HSD gene family. To date, five 17ß-HSD enzymes have been discovered (Luu-The *et al.*, 1989; Gast *et al.*, 1989; Wu *et al.*, 1993; Geissler *et al.*, 1994; Adamski *et al.*, 1995 and Zang *et al.*, 1995), type II and type III 17ß-HSD cDNAs were expressed cloned as attempts to isolate the genes using radiolabelled type I were unsuccessful. Indeed, after the characterisation of these genes a pairwise comparison of the amino acid identities between the three isozymes indicated only 23%. In addition, the genes that express the enzymes within the family are not localised to the same chromosome. Type I has been localised to chromosome 7q21 (Luu-The *et al.*, 1989), type II to chromosome 16q24 (Casey *et al.*, 1994), type III to chromosome 9q22 (Geissler *et al.*, 1994), the chromosomal localisation of type IV is unknown and type V has been localised to chromosome 10p14,15 (Zang *et al.*, 1995).

6.5 Comparisons of the human and rodent 3ß-HSD gene families

The characterisation of the human 3ß-HSD gene family enables us to question why rodents have more functional 3ß-HSD enzymes. To date, there have been 6 functional genes identified in mouse (Payne *et al.*, 1997), 4 functional genes in rat (Simard *et al.*, 1993) and 3 in hamster (Rogerson *et al.*, 1997). Both rodents and humans have a peripheral isoform of 3ß-HSD (rat IV, mouse III/VI and human I) and an adrenal/gonadal isoform (rat I, mouse I and human II) and these enzymes seem to play the same role in each mammal. The other enzymes present in rodents are either the KSR specific isoforms or additional dehydrogenases/isomerases which are expressed in the liver and kidney also (Table 1.1). It is unknown why rodents have more functional 3ß-HSD genes than humans. It is possible that the rodent liver and kidneys play a more steroidogenic role than humans. There is a lack of data available on the role of the rodent liver and kidney in steroidogenesis. Little is known about the expression of 3ß-HSD in these tissues except that a number of 3ß-HSD isoforms are expressed and the expression of these isoforms is sexually dimorphic and in some cases sex-specific (Keeney *et al.*, 1993a; Keeney *et al.*, 1993b; Naville *et al.*, 1991). Rat III, mouse V and hamster II are expressed in the male liver only. It is possible to speculate that the rodent liver has developed to metabolise the high amount of exogenous steroids which may be consumed from their diet of plants and seeds.

6.6 The origin of mutations found in HSD3B2 associated with 3ß-HSD deficiency

3ß-HSD deficiency is an autosomal recessive disease resulting in congenital adrenal hyperplasia which impairs steroidogenesis in both the adrenals and the gonads. Patients exhibit varying degrees of adrenal insufficiency associated with pseudohermaphroditism in males and mild virilisation or normal sexual differentiation in females (Simard *et al.*, 1995). The availability of the sequence of the pseudogenes and their map order with respect to the type I and II genes allows us to question whether any of the mutations that have been reported in the type II gene could be due to homologous recombination or gene conversion. These mechanisms of genetic recombination are responsible for mutations observed in 21-hydroxylase (cytochrome P450c21; Collier *et al.*, 1993) and aldosterone synthase (cytochrome P450c11; Lifton *et al.*, 1992a; Lifton *et al.*, 1992b).

Genetic deficiency in 21-hydroxylase is responsible for approximately 90% of cases of congenital adrenal hyperplasia (New, 1994) and it is reported that carriers of 21-hydroxylase deficiency attain frequencies of 10% in world populations studies (Witchel *et al.*, 1997). The enzyme converts progesterone into deoxycorticosterone and 17-hydroxyprogesterone into 11-deoxycortisol (see Fig. 1.3). The gene for 21-hydroxylase (CYP21B) has been localised to chromosome 6p21.3 along with a highly homologous pseudogene (CYP21A), both genes contain 10 exons and 9 introns with 98 and 96% homology respectively (Higashi *et al.*, 1986; White *et al.*, 1986). The genes are on tandemly repeated DNA segments approximately 30kb long containing 2 other duplicated genes, namely C4A and C4B, which encode protein components of serum complement (Sinnott *et al.*, 1990). White *et al.*(1994) concluded that greater than 90% of mutant alleles in P450c21 deficiency have arisen by interlocus sequences exchanges such as unequal crossover and gene conversion-like events. 75% of these are mutations copied from deleterious mutations in the pseudogene, suggesting a gene conversion mechanism. Gene conversion describes a non-reciprocal transfer of sequences between a pair of non-allelic or allelic DNA sequences. The mechanism of this transfer of sequence is unknown, but one possible mechanism suggests the formation of a heteroduplex between a DNA strand from the donor and a DNA strand from the acceptor (Fig. 6.2). These sequences will be similar but not identical and DNA repair mechanisms may "correct" the DNA strand of the acceptor to the same sequence as the complementary strand of the donor. Subsequent replication of the acceptor strand and sealing of nicks will complete the conversion. Analysis of one mutation involved in 21-hydroxylase deficiency suggests that the conversion tract may be a maximim of 390bp (Collier *et al.*, 1993).

Homologous recombination events are involved in P450c11 deficiency disease. Deficiency of P450c11 is the second most common cause of congenital adrenal hyperplasia and accounts for 5% of adrenal defects (White *et al.*, 1987). Two homologous P450c11 enzymes are encoded by the CYP11B1 and CYP11B2 genes on chromosome 8q21-22 (Mornet *et al.*, 1989). CYP11B1 encodes P450c11 which is expressed in the zona fasciculata and reticularis and is required for cortisol biosynthesis whereas CYP11B2 encodes P450c18 which possesses three activities required for aldosterone synthesis in the zona glomerulosa (see section 1.7). A dominant form of inherited hypertension (dexamethasone suppressible hypertension) is caused by unequal crossover between the CYP11B1 and CYP11B2 genes (Lifton *et al.*, 1992). This results in a hybrid gene which the 5' end of CYP11B1 fused to the coding region of CYP11B2, placing the P450c11 promoter upstream of P450c18. This results in mineralocorticoid secretion from the zona fasciculata and leads to hypertension.

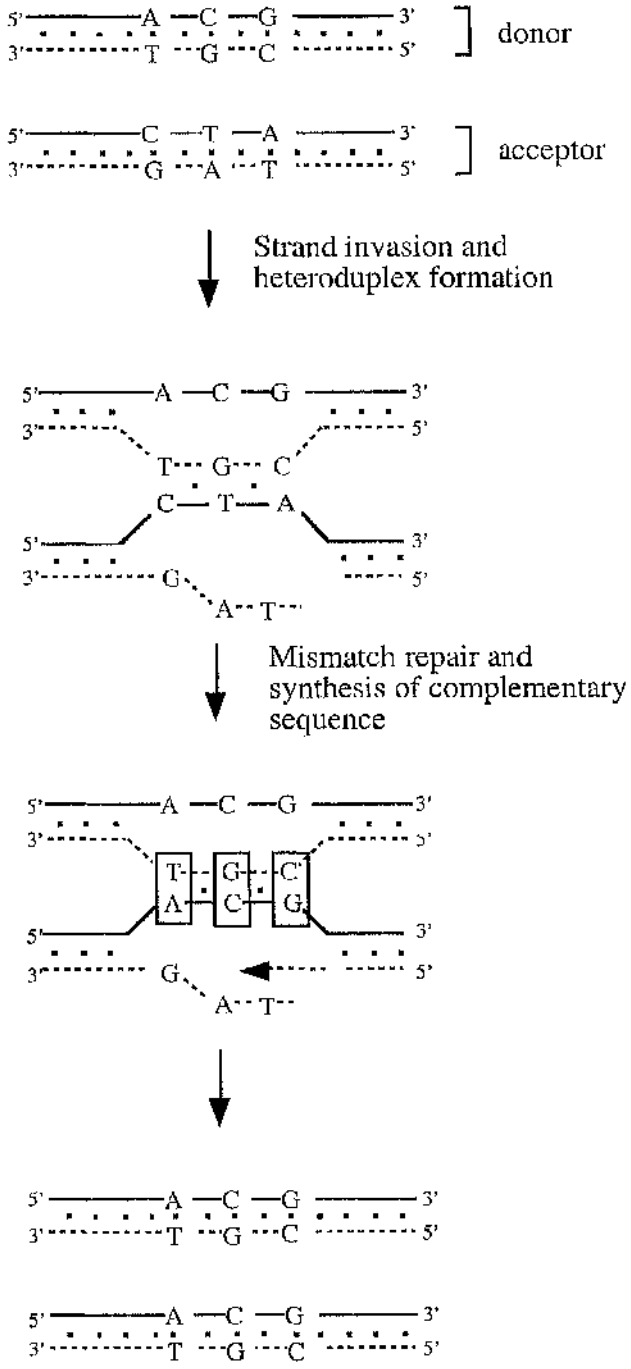


Fig. 6.2 Mismatch repair of a heteroduplex as a possible mechanism for gene conversion. One of the strands from the donor sequence forms a heteroduplex with the complementary strand of the acceptor sequence, displacing the other acceptor strand. Mismatch repair enzymes recognise the mispaired bases and "correct" the sequence so that the acceptor sequence is converted into the complementary strand of the donor. Replication of the acceptor strand results in completion of the conversion. Modified from Strachan and Read, 1996.

There have been 24 deleterious mutations described in 3ß-HSD type II which are responsible for 3ß-HSD deficiency (Morel *et al.*, 1997). 22 of the mutations are indicated on the sequence alignment in appendix ii. From the mutations shown, 6 are found within the pseudogenes (Table 6.2) raising the question whether the mechanisms involved in the origin of the mutations in P450c21 and P450c11 deficiencies are relevant to 3ß-HSD. However, in no cases of 3ß-HSD deficiency is there more than one base change in common with a pseudogene and it is believed that gene conversion results in modification of a segment of sequence from the donor sequence. This does not exclude the possibility that a single base conversion event may take place.

Another aspect of the mutations involved in 3ß-HSD deficiency has been observed that may further argue against conversion. Sequence analysis of the pseudogenes determined that 4 out of 6 mutations result from a CpG dinucleotide present in the normal HSD3B2 sequence and this CpG is conserved in at least 2 of the other members of the gene family (Table 6.2). In contrast, the remaining mutations did not have counterparts in ψ1-5 or HSD3B1, and none were located at CpG sites. CpG dinucleotides are extremely mutable and usually lead to transitions (Cooper and Youssoufian, 1988). Transitions (conversion of a pyrimidine to a pyrimidine or purine to purine) are unusually common in both coding and non-coding DNA. This is partly due to the instability of cytosine residues in the CpG dinucleotide, where the cytosine is often methylated at the 5' carbon atom and 5-methylcytosine is susceptible to spontaneous deamination to thymidine. As a result, the CpG dinucleotide is a hotspot for mutation and it is believed that the mutation rate for the CpG dinucleotide is about 8.5 times higher than that of the average dinucleotide (Cooper *et al.*, 1995). This greatly reduces the probability that a new HSD3B2 mutation at a CpG site will have diverged from a mutation in the CpG site of a pseudogene that has been fixed earlier in the evolution of the cluster. It is noteworthy to add that the CpG sites that are mutated in HSD3B2 are usually conserved in the other family members.

HSD3B2 Exon	Mutation	CpG in normal HSD3B2 allele	Mutation present elsewhere in cluster	base in normal HSD3B allele conserved in other family members
2	G323A	no	ψ3	yes
3	G129A	yes	ψ5	yes
4	G8A	no	ψ4	yes
4	G91A	yes	ψ3, ψ4	HSD3B1, HSD3Bψ2
4	G130A	yes	ψ1	yes
4	C482T	yes	ψ3	yes

Table 6.2: HSD3B2 mutations shared with ψ1-5; location of CpG sites.

A final piece of evidence which supports the argument against conversion was that haplotype studies indicated that mutation G129A (see appendix ii, exon 3) did not arise independently in 3 apparently unconnected families but arose by descent from a single common ancestral mutation (Simon McCartin, personnal communication). This suggests the evolution of a single event throughout the families rather than gene-conversion-like mechanism occuring independently in each family. In conclusion, there is no evidence for gene conversion or unequal crossing over as a mechanism contributing to the known HSD3B2 mutations.

6.7 Evolution of the 3ß-HSD gene family

The evolution of a multicellular complex organism probably involved a gradual increase in genome size and this increase in genome size must have taken place without compromising the functions of the DNA already present. Gene duplication allows sequences to diverge relatively rapidly without affecting the function of the gene product, because one gene at the duplicated locus is surplus to requirement and this can diverge rapidly due to the absence of selective pressure to maintain function. The diverged sequence may acquire a new function but most likely will acquire deleterous mutations and degenerate into a non-functional pseudogene (Strachan and Read, 1996). This mechanism is represented in Fig. 6.3.

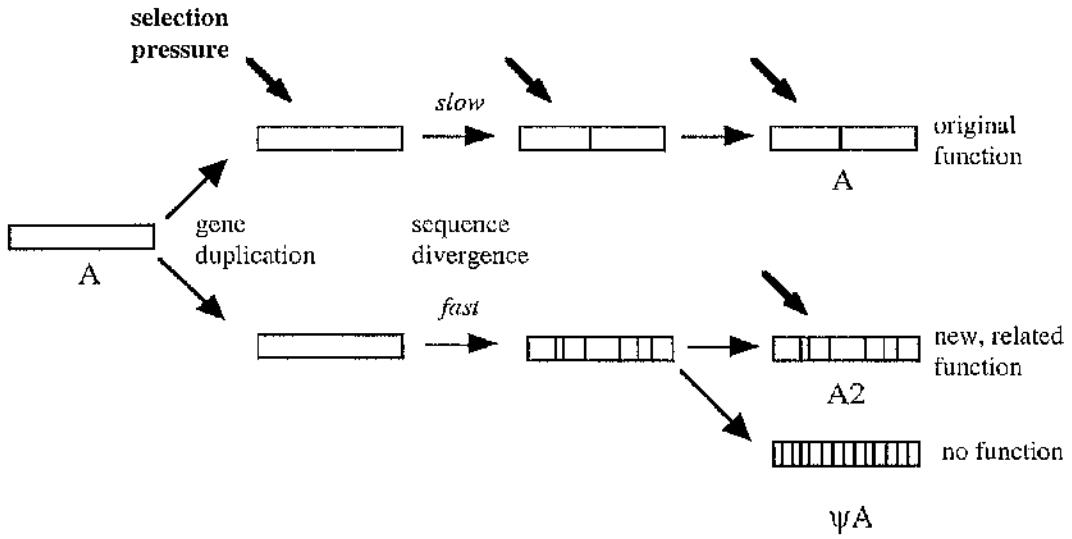


Fig. 6.3 Gene duplication can lead to the formation of genes with a novel function or non-functional pseudogenes (vertical bars within the boxes represent mutations). Modified from Strachan and Read, 1996.

Tandem gene duplication results from unequal crossing over events between similar sequences on different chromosomes or sister chromatids. Clustered gene families often result from tandem gene duplication, a good example of this is the globin gene families (Goodman *et al.*, 1987). The globin genes have duplicated from a single ancestral gene to form a highly-coordinated gene family with developmental and tissue-specific expression. The initial duplication event from a single ancestral gene is rare (Maeda and Smithies, 1986) and it is likely that the unequal crossover or sister chromatid exchange occured at regions of little homology, following mispairing of nonallelic repeats such as Alu repeats (Cross and Renkawitz, 1990; Hu *et al.*, 1991) Once gene families have expanded it is believed that recombination events will become more frequent because of the repetitions in the sequences.

Different mechanisms of recombination can occur within a gene family, these will either allow divergence or homogenisation of the sequences. Concerted evolution occurs by intragenomic sequence exchanges within a DNA sequence family by means of unequal crossover, sister chromatid exchange and gene conversion (Liao *et al.*, 1997). Such non-allelic sequence changes between the members of a gene family will tend to drive the family towards homogeneity. In contrast, the "birth and death" process allows gene families to duplicate with some duplicated genes maintained to create a new or additional function, and others are deleted or become non-functional by deleterious mutations (Nei *et al.*, 1997). This model of gene family evolution produces a gene family with highly homologous genes and a substantial amount of pseudogenes. For example, it is likely that the immunoglobin gene families evolved via the "birth and death" model as these genes strive for divergence, whereas gene families such as the ribosomal genes follow concerted evolution where a large quantity of the same product is produced (Nei *et al.*, 1997).

With respect to the 3ß-HSD gene family, it is likely that the "birth and death" process applies for divergence of the family, in contrast to concerted evolution which creates homogeneous sequences. This gene family contains functional and pseudogene members, which is consistent with the "birth and death" process. With concerted evolution, there is a tendency to see "patching" of the sequence, with some regions highly homologous and other regions more diverged, this does not seem to apply to the sequences identified in the 3ß family. Sequence comparison of the 3ß-HSD gene family indicates that base changes between members are found throughout the sequences and are not restricted to groups within the family.

The relationship between the nucleotide sequences in the 3ß-HSD gene family was determined by weighted parsimony analysis of the alignment between the exon

sequences (Fig. 6.4). Branch lengths indicate the number of informative nucleotide substitutions in the data set. This analysis suggests that seven human sequences are more closely related to one another than to the mouse or the rat which indicates that they expanded after the divergence from the common ancestor with rodents, as previously noted for HSD3B1 and 2 by Abbaszade *et al* .(1995). The position of the ATG codon in type II compared to type I and the pseudogenes (the ATG in HSD3B2 is located one codon downstream from HSD3B1 and ψ1-5) is consistent with the pseudogenes having a more ancient origin than the divergence of HSD3B1 and 2. Pseudogenes ψ4 and 5 show the closest homology to HSD3B1 and 2 and are both expressed as RNA in several tissues. This suggests that either these pseudogenes are more recently duplicated or have been more recently released from the selective pressures of the expression of a functional enzyme.

The phylogenetic tree demonstrates that several duplications took place during the evolution of the 3ß-HSD gene family. One such event gave rise to the KSR enzymes in the rodent family, resulting in the evolution of two distinct functional groups of enzymes, the KSR specific enzymes (rat III, mouse IV and V) and the dehydrogenases/isomerases (rat I, II, IV and mouse I, II, III and VI). Investigations into the macaque 3ß-HSD gene family may yield important information to the expansion of the 3ß-HSD gene family. It is expected that there may be an expanded 3ß-HSD gene family in macaque as there is evidence to suggest the expression of 3ß-HSD in peripheral tissues such as the liver and kidney (Martel *et al.*, 1994).

The precise mechanism of the 3ß-HSD gene family expansion is unknown. We can speculate on possible pathways of duplications and genetic recombinations, however some additional information is required to determine the exact mechanism. For example, comparison of the upstream region of HSD3B1 and 2 shows a marked divergence in upstream sequence (see appendix i, upstream from base 1280, type I) which is consistent with type I and type II having different regulatory elements leading to differential tissue expression. Therefore, sequence data from these upstream regions of the pseudogenes may provide significant information which may allude to the pathway of gene family expansion. Furthermore, by obtaining the intron sequences from the pseudogenes and the orientations of all 7 genes the mechanism of expansion of the 3ß-HSD gene family may become apparent.

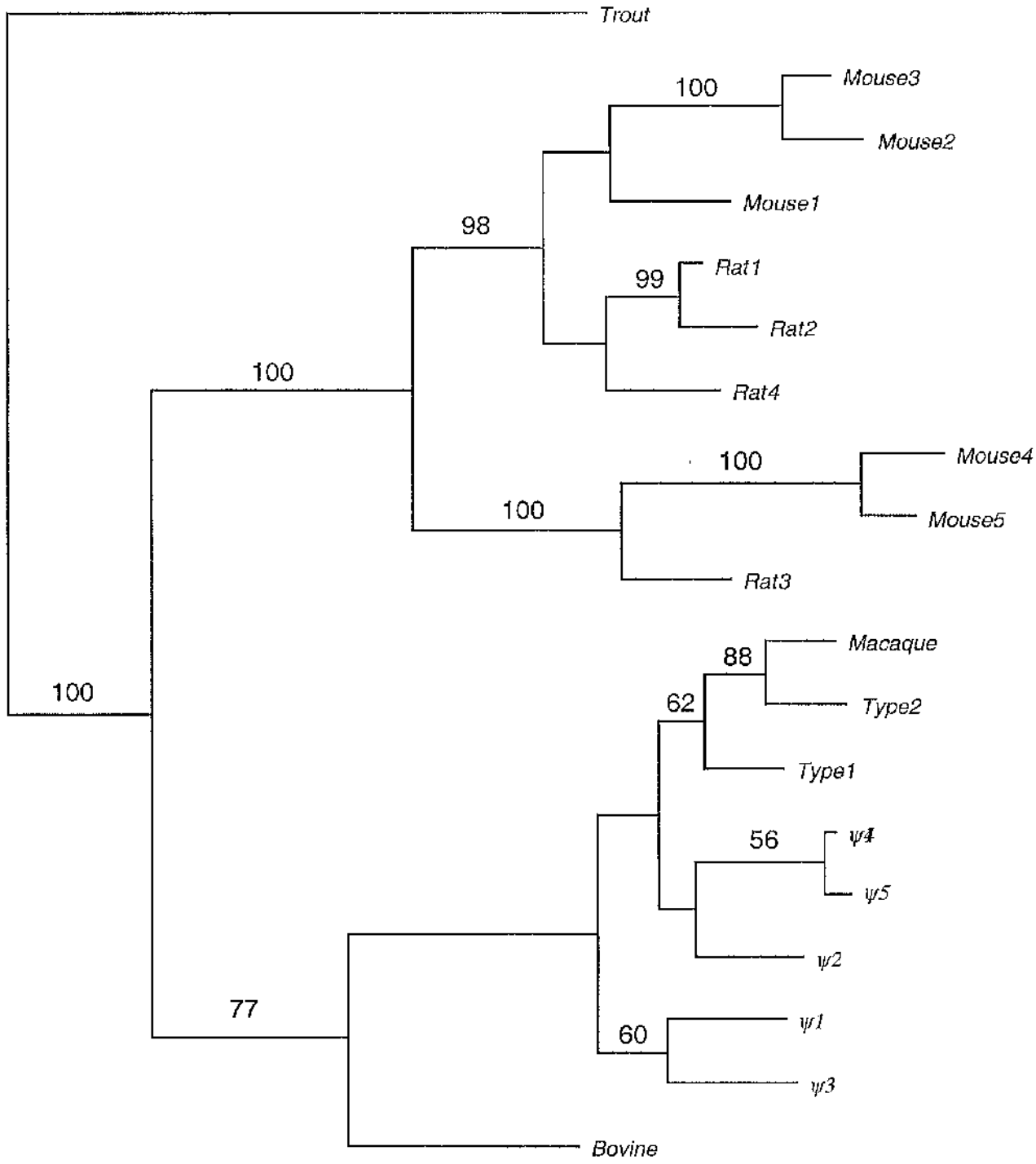


Figure 6.4. Weighted parsimony analysis of an alignment between expressed mammalian cDNAs and human exon homologue DNA sequences (ψ1-ψ5). Branch lengths represent the number of informative nucleotide substitutions in the data set. Bootstrapping percentages greater than 50% are shown on internal nodes (based on 1000 replications with the input order shuffled randomly on each replication). Reproduced with kind permission from Martin McBride.

6.8 Concluding remarks

The 3ß-HSD gene family contains 7 members; 2 previously identified functional genes (HSD3B1 and 2), and 5 novel unprocessed pseudogenes (HSD3Bψ1-5) identified by screening two human genomic libraries with 3ß-HSD type I cDNA. A YAC and BAC genomic contig established the order of the 3ß-HSD genes at chromosome 1p13.1, and the orientation of HSD3B1, 2, ψ1 and ψ4 was determined by restriction analysis and the position of the endpoints of three genomic clones within the contig. The estimated length of the contig is 500kb with the 3ß-HSD gene family over a centrally-based 235kb fragment. No further members of the gene family were detected by extensive restriction analysis of the cluster or from hybrid selection experiments. No member of the 3ß-HSD gene family was discovered that may be a candidate to encode a human 3-ketosteroid reductase or 3ß-C_{27}-HSD. However, it is likely that these genes contain sufficiently diverged sequence from the 7 identified 3ß-HSD genes, and therefore, it is not surprising that experiments involving homology with the other members of the family have not detected these genes. Expression cloning experiments would probably result in the cloning of the 3-ketosteroid reductase and 3ß-C_{27}-HSD.

CHAPTER 7

Bibliography

Andersson, S., and Moghrabi, N. (1997). Physiology and molecular genetics of 17 beta-hydroxysteroid dehydrogenases. Steroids *62*, 143-147.

Abalain, J. H., Quemener, E., Carre, J. L., Simon, B., Amet, Y., Mangin, P., and Floch, H. H. (1989). Metabolism Of Androgens In Human Hyperplastic Prostate - Evidence For a Differential Localization Of the Enzymes Involved In the Metabolism. Journal Of Steroid Biochemistry *34*, 467-471.

Abbaszade, I. G., Clarke, T. R., Park, C. H. J., and Payne, A. H. (1995). The Mouse 3-Beta-Hydroxysteroid Dehydrogenase Multigene Family Includes 2 Functionally Distinct Groups Of Proteins. Molecular Endocrinology *9*, 1214-1222.

Adamski, J., Normand, T., Leenders, F., Monte, D., Begue, A., Stehelin, D., Jungblut, P. W., and Delaunoit, Y. (1995). Molecular-Cloning Of a Novel Widely Expressed Human 80 Kda 17-Beta- Hydroxysteroid Dehydrogenase-Iv. Biochemical Journal *311*, 437-443.

Alvarez, C. L., Gentiraimondi, S., Patrito, L. C., and Flury, A. (1994). Topography Of Human Placental 3-Beta-Hydroxysteroid Dehydrogenase Delta(5-4) Isomerase In Microsomal Membrane - a Study Using Limited Proteolysis and Immunoblotting. Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology *1207*, 102-108.

Amet, Y., Simon, B., Quemener, E., Mangin, P., Floch, H. H., and Abalain, J. H. (1992). Partial-Purification Of 3-Alpha-Hydroxysteroid and 3-Beta- Hydroxysteroid Dehydrogenases From Human Hyperplastic Prostate - Comparison Between the 2 Enzymes. Journal Of Steroid Biochemistry and Molecular Biology *41*, 689-692.

Bain, P. A., Meisler, M. H., Taylor, B. A., and Payne, A. H. (1993). The Genes Encoding Gonadal and Nongonadal Forms Of 3-Beta- Hydroxysteroid Dehydrogenase/Delta(5)-Delta(4) Isomerase Are Closely Linked On Mouse Chromosome-3. Genomics *16*, 219-223.

Beato, M., Herrlich, P., and Schutz, G. (1995). Steroid-Hormone Receptors - Many Actors In Search Of a Plot. Cell *83*, 851-857.

Beaudoin, C., Blomquist, C. H., Bonenfant, M., and Tremblay, Y. (1997a). Expression of the genes for 3 beta-hydroxysteroid dehydrogenase type 1 and cytochrome P450scc during syncytium formation by human placental cytotrophoblast

cells in culture and the regulation by progesterone and estradiol. Journal Of Endocrinology *154*, 379-387.

Beaudoin, C., Bonenfant, M., and Tremblay, Y. (1997b). Regulation of cytochrome P450 cholesterol side-chain cleavage, 3 beta-hydroxysteroid dehydrogenase Delta 5-Delta 4 isomerase type 1 and estradiol 17 beta-hydroxysteroid dehydrogenase mRNA levels by calcium in human choriocarcinoma JEG-3 cells. Molecular and Cellular Endocrinology *133*, 63-71.

Bern, H. A. (1990). The New Endocrinology - Its Scope and Its Impact. American Zoologist *30*, 877-885.

Bird, I. M., Pasquarette, M. M., Rainey, W. E., and Mason, J. I. (1996). Differential Control Of 17-Alpha-Hydroxylase and 3-Beta- Hydroxysteroid Dehydrogenase Expression In Human Adrenocortical H295r Cells. Journal Of Clinical Endocrinology and Metabolism *81*, 2171-2178.

Bird, I. M., Pasquarette, M. M., Rainey, W. E., and Mason, J. I. (1996). Differential Control Of 17-Alpha-Hydroxylase and 3-Beta- Hydroxysteroid Dehydrogenase Expression In Human Adrenocortical H295r Cells. Journal Of Clinical Endocrinology and Metabolism *81*, 2171-2178.

Bongiovanni, A.M. (1961). Journal Of Clinical Endocrinolgy And Metabolism *21* 860-862.

Brelinska, R., and Warchol, J. B. (1997). Thymic nurse cells: Their functional ultrastructure. Microscopy Research and Technique *38*, 250-266.

Buchmann, M. S., Kvittingen, E. A., Nazer, H., Gunasekaran, T., Clayton, P. T., Sjovall, J., and Bjorkhem, I. (1990). Lack Of 3-Beta-Hydroxy-Delta-5-C-27-Steroid Dehydrogenase Isomerase In Fibroblasts From a Child With Urinary-Excretion Of 3-Beta-Hydroxy- Delta-5-C-27-Bile Acids - a New Inborn Error Of Metabolism. Journal Of Clinical Investigation *86*, 2034-2037.

Buckler, A. J., Chang, D. D., Graw, S. L., Brook, J. D., Haber, D. A., Sharp, P. A., and Housman, D. E. (1991). Exon Amplification - a Strategy to Isolate Mammalian Genes Based On Rna Splicing. Proceedings Of the National Academy Of Sciences Of the United States Of America *88*, 4005-4009.

Carmeliet, P., and Collen, D. (1997). Molecular analysis of blood vessel formation and disease. American Journal Of Physiology-Heart and Circulatory Physiology *42*, H2091-H2104.

Casey, M. L., Macdonald, P. C., and Andersson, S. (1994). 17-Beta-Hydroxysteroid Dehydrogenase Type-2 - Chromosomal Assignment and Progestin Regulation Of Gene-Expression In Human Endometrium. Journal Of Clinical Investigation *94*, 2135-2141.

Chen, H., Chrast, R., Rossier, C., Morris, M. A., Lalioti, M. D., and Antonarakis, S. E. (1996). Cloning Of 559 Potential Exons Of Genes Of Human-Chromosome-21 By Exon Trapping. Genome Research *6*, 747-760.

Chernyadev, II (1997). Plant photosynthesis under conditions of water stress and the protective effect of cytokinins: A review. Applied Biochemistry and Microbiology *33*, 1-12.

Cherradi, N., Defaye, G., and Chambaz, E. M. (1993). Dual Subcellular-Localization Of the 3-Beta-Hydroxysteroid Dehydrogenase Isomerase - Characterization Of the Mitochondrial Enzyme In the Bovine Adrenal-Cortex. Journal Of Steroid Biochemistry and Molecular Biology *46*, 773-779.

Cherradi, N., Defaye, G., and Chambaz, E. M. (1994). Characterization Of the 3-Beta-Hydroxysteroid Dehydrogenase-Activity Associated With Bovine Adrenocortical Mitochondria. Endocrinology *134*, 1358-1364.

Cherradi, N., Guidicelli, C., Defaye, G., and Chambaz, E. M. (1992). Purification and Characterization Of 3-Beta-Hydroxysteroid- Dehydrogenase Isomerase From Bovine Adrenal-Cortex. Journal Of Steroid Biochemistry and Molecular Biology *41*, 831-836.

Cherradi, N., Rossier, M. F., Vallotton, M. B., Timberg, R., Friedberg, I., Orly, J., Wang, X. J., Stocco, D. M., and Capponi, A. M. (1997). Submitochondrial distribution of three key steroidogenic proteins (steroidogenic acute regulatory protein and cytochrome p450(scc) and 3 beta-hydroxysteroid dehydrogenase isomerase enzymes) upon stimulation by intracellular calcium in adrenal glomerulosa cells. Journal Of Biological Chemistry *272*, 7899-7907.

Clarke, T. R., Bain, P. A., Burmeister, M., and Payne, A. H. (1996). Isolation and Characterization Of Several Members Of the Murine Hsd3b Gene Family. Dna and Cell Biology *15*, 387-399.

Collier, S., Tassabehji, M., and Strachan, T. (1993). A Denovo Pathological Point Mutation At the 21-Hydroxylase Locus - Implications For Gene Conversion In the Human Genome. Nature Genetics *3*, 260-265.

Cooper, D. N., and Youssoufian, H. (1988). The Cpg Dinucleotide and Human Genetic-Disease. Human Genetics *78*, 151-155.

Cooper, D.N., Krawczak, M. and Antonorakis, S.E. (1995) In: *The Metabolic and Molecular Basis Of Inherited Disease*, 7th Edition (Eds. CR Scriver, AL Beaudet, WS Sly, D Valle) MacGraw-Hill, New York.

Court, D.A., Kleene, R., Neupert, W., Lill, R. (1996) Role Of The N-Termini And C-Termini Of Porin In Import Into The Outer-Membrane Of Neurospora Mitochondria. Febs Letters *390(1)* 73-77

Cross, M., and Renkawitz, R. (1990). Repetitive Sequence Involvement In the Duplication and Divergence Of Mouse Lysozyme Genes. Embo Journal *9*, 1283-1288.

Darvill, A., Augur, C., Bergmann, C., Carlson, R. W., Cheong, J. J., Eberhard, S., Hahn, M. G., Lo, V. M., Marfa, V., Meyer, B., Mohnen, D., Oneill, M. A., Spiro, M. D., Vanhalbeek, H., York, W. S., and Albersheim, P. (1992). Oligosaccharins - Oligosaccharides That Regulate Growth, Development and Defense Responses In Plants. Glycobiology *2*, 181-198.

Davis, W.W. and Garren, L.D. (1968). On The Mechanisms Of Action Of adrenocorticotrophic. The Inhibitory Site Of Cyclohexamide In The Pathway Of Steroid Biosynthesis. Journal Of Biological Chemistry *243*, 5153-5157.

Delaunoit, Y., Zhao, H. F., Belanger, A., Labrie, F., and Simard, J. (1992). Expression Of Liver-Specific Member Of the 3-Beta-Hydroxysteroid Dehydrogenase Family, an Isoform Possessing an Almost Exclusive 3- Ketosteroid Reductase-Activity. Journal Of Biological Chemistry *267*, 4513-4517.

Devine, P. L., Kelly, N. S., and Adams, J. B. (1986). 3-Beta-Hydroxysteroid Isomerase Dehydrogenase In Guinea-Pig Kidney - Possible Involvement In 11-Deoxycorticosterone Formation Insitu. Journal Of Steroid Biochemistry *25*, 265-270.

Dohlman, H. G., Caron, M. G., and Lefkowitz, R. J. (1987). A Family Of Receptors Coupled to Guanine-Nucleotide Regulatory Proteins. Biochemistry *26*, 2657-2664.

Donascimento, R. R., and Morgan, E. D. (1996). Chemicals Involved In the Communication-System Of Social Insects - Their Source and Methods Of Isolation and Identification, With Special Emphasis On Ants. Quimica Nova *19*, 156-165.

Doody, K. M., Carr, B. R., Rainey, W. E., Byrd, W., Murry, B. A., Strickler, R. C., Thomas, J. L., and Mason, J. I. (1990). 3-Beta-Hydroxysteroid Dehydrogenase Isomerase In the Fetal Zone and Neocortex Of the Human Fetal Adrenal-Gland. Endocrinology *126*, 2487-2492.

Dumont, M., Luuthe, V., Dupont, E., Pelletier, G., and Labrie, F. (1992). Characterization, Expression, and Immunohistochemical Localization Of 3-Beta-Hydroxysteroid Dehydrogenase/Delta-5-Delta-4 Isomerase In Human Skin. Journal Of Investigative Dermatology *99*, 415-421.

Duyk, G. M., Kim, S. W., Myers, R. M., and Cox, D. R. (1990). Exon Trapping - a Genetic Screen to Identify Candidate Transcribed Sequences In Cloned Mammalian Genomic Dna. Proceedings Of the National Academy Of Sciences Of the United States Of America *87*, 8995-8999.

Evans, W. H., and Bergeron, J. J. M. (1988). Nuclear Receptors - a Re-Evaluation. Trends In Biochemical Sciences *13*, 7-8.

Ferre et al. (1975) Human Placental Δ5-3ß-Hydroxysteroid Dehydrogenase (D5-HSDH): Intracellular Distribution, Kinetic Properties, Retroinhibition And Influence On Membrane Delipidation. Steroids *26* 551-569.

Folsch, H., Guiard, P., Neupert, W., Stuart, R.A. (1996) Internal Targeting Signal Of The Bcs1 Protein - A Novel Mechanism Of Import Into Mitochondria. Embo Journal *15(3)* 479-487

Fritz, I. B. (1994). Somatic Cell-Germ Cell Relationships In Mammalian Testes During Development and Spermatogenesis. Ciba Foundation Symposia *182*, 271-274.

Gast, M. J., Sims, H. F., Murdock, G. L., Gast, P. M., and Strauss, A. W. (1989). Isolation and Sequencing Of a Complementary Deoxyribonucleic-Acid Clone Encoding Human Placental 17-Beta-Estradiol Dehydrogenase - Identification Of the Putative Cofactor Binding-Site. American Journal Of Obstetrics and Gynecology *161*, 1726-1731.

Geissler, W. M., Davis, D. L., Wu, L., Bradshaw, K. D., Patel, S., Mendonca, B. B., Elliston, K. O., Wilson, J. D., Russell, D. W., and Andersson, S. (1994). Male Pseudohermaphroditism Caused By Mutations Of Testicular 17-Beta- Hydroxysteroid Dehydrogenase-3. Nature Genetics *7*, 34-39.

Gibb, W., Lavoie, J. C., and Moringonthier, M. (1985). Kinetic Comparison Of the 3-Beta-Hydroxysteroid Dehydrogenase- Activity In Human-Placenta, Chorion Laeve, and Ovary. Canadian Journal Of Biochemistry and Cell Biology *63*, 183-186.

Goodman, M., Czelusniak, J., Koop, B. F., Tagle, D. A., and Slightom, J. L. (1987). Globins - a Case-Study In Molecular Phylogeny. Cold Spring Harbor Symposia On Quantitative Biology *52*, 875-890.

Grosveld, F., Dillon, N., Higgs, D. (1993) The Regulation Of Human Globin Gene-Expression Baillieres Clinical Haematology. *6(1)* 31-55

Guerin, S. L., Leclerc, S., Verreault, H., Labrie, F., and Luuthe, V. (1995). Overlapping Cis-Acting Elements Located In the First Intron Of the Gene For Type-I 3-Beta-Hydroxysteroid Dehydrogenase Modulate Its Transcriptional Activity. Molecular Endocrinology *9*, 1583-1597.

Hanukoglu, I. (1992). Steroidogenic Enzymes - Structure, Function, and Role In Regulation Of Steroid-Hormone Biosynthesis. Journal Of Steroid Biochemistry and Molecular Biology *43*, 779-804.

Harshman, K., Bell, R., Rosenthal, J., Katcher, H., Miki, Y., Swenson, J., Gholami, Z., Frye, C., Ding, W., Dayananth, P., Eddington, K., Norris, F. H., Bristow, P. K., Phelps, R., Hattier, T., Stone, S., Shaffer, D., Bayer, S., Hussey, C., Tran, T., Richardson, K., Dehoff, B., Lai, M., Rosteck, P. R., Skolnick, M. H., Shattuckeidens, D., and Kamb, A. (1995). Comparison Of the Positional Cloning Methods Used to Isolate the Brca1 Gene. Human Molecular Genetics *4*, 1259-1266.

Hattier, T., Bell, R., Shaffer, D., Stone, S., Phelps, R. S., Tavtigian, S. V., Skolnick, M. H., Shattuckeidens, D., and Kamb, A. (1995). Monitoring the Efficacy Of Hybrid Selection During Positional Cloning - the Search For Brca1. Mammalian Genome *6*, 873-879.

Higashi, Y., Yoshioka, H., Yamane, M., Gotoh, O., and Fujiikuriyama, Y. (1986). Complete Nucleotide-Sequence Of 2 Steroid 21-Hydroxylase Genes Tandemly Arranged In Human-Chromosome - a Pseudogene and a Genuine Gene. Proceedings Of the National Academy Of Sciences Of the United States Of America *83*, 2841-2845.

Hillier, S. G. (1985). Sex Steroid-Metabolism and Follicular Development In the Ovary. Oxford Reviews Of Reproductive Biology *7*, 168-222.

Hu, X. Y., Ray, P. N., and Worton, R. G. (1991). Mechanisms Of Tandem Duplication In the Duchenne Muscular-Dystrophy Gene Include Both Homologous and Nonhomologous Intrachromosomal Recombination. Embo Journal *10*, 2471-2477.

Ishiiohba, H., Inano, H., and Tamaoki, B. I. (1986). Purification and Properties Of Testicular 3-Beta-Hydroxy-5-Ene- Steroid Dehydrogenase and 5-Ene-4-Ene Isomerase. Journal Of Steroid Biochemistry *25*, 555-560.

Keeney, D. S., Murry, B. A., Bartke, A., Wagner, T. E., and Mason, J. I. (1993a). Growth-Hormone Transgenes Regulate the Expression Of Sex-Specific Isoforms Of 3-Beta-Hydroxysteroid Dehydrogenase/Delta-5-]4-Isomerase In Mouse-Liver and Gonads. Endocrinology *133*, 1131-1138.

Keeney, D. S., Naville, D., Milewich, L., Bartke, A., and Mason, J. I. (1993b). Multiple Isoforms Of 3-Beta-Hydroxysteroid Dehydrogenase/Delta-5-]4- Isomerase In Mouse-Tissues - Male-Specific Isoforms Are Expressed In the Gonads and Liver. Endocrinology *133*, 39-45.

Kemppainen, R. J., and Behrend, E. (1997). Adrenal physiology. Veterinary Clinics Of North America-Small Animal Practice *27*, 173.

Kholkute, S. D., Rodriquez, J., and Dukelow, W. R. (1995). In-Vitro Fertilization and the Effect Of Progesterone and 17-Alpha- Hydroxyprogesterone On Acrosome Reaction Of Mouse Epididymal Spermatozoa. International Journal Of Andrology *18*, 146-150.

Kliman, H. J., Nestler, J. E., Sermasi, E., Sanger, J. M., and Strauss, J. F. (1986). Purification, Characterization, and Invitro Differentiation Of Cytotrophoblasts From Human Term Placentae. Endocrinology *118*, 1567-1582.

Korn, B., Sedlacek, Z. and Poustka, A. (1992) Isolation Of Transcribed and Conserved Sequences From Large Genomic Regions By Magnetic Capture. In: *Advances In Biomagnetic Separation* (Eds. M.Uhlen, E. Hornes and O.Olsvik) Eaton Publishing.

Labrie, F., Belanger, A., Simard, J., Luuthe, V., and Labrie, C. (1995). Intracrinology - Autonomy and Freedom Of Peripheral-Tissues. Annales D Endocrinologie *56*, 23-29.

Labrie, F., LuuThe, V., Lin, S. X., Labrie, C., Simard, J., Breton, R., and Belanger, A. (1997). The key role of 17 beta-hydroxysteroid dehydrogenases in sex steroid biology. Steroids *62*, 148-158.

Labrie, F., Simard, J., Luuthe, V., Pelletier, G., Belanger, A., Lachance, Y., Zhao, H. F., Labrie, C., Breton, N., Delaunoit, Y., Dumont, M., Dupont, E., Rheaume, E., Martel, C., Couet, J., and Trudel, C. (1992). Structure and Tissue-Specific Expression Of 3-Beta-Hydroxysteroid Dehydrogenase/5-Ene-4-Ene Isomerase Genes In Human and Rat Classical and Peripheral Steroidogenic Tissues. Journal Of Steroid Biochemistry and Molecular Biology *41*, 421-435.

Labrie, F., Simard, J., Luuthe, V., Trudel, C., Martel, C., Labrie, C., Zhao, H. F., Rheaume, E., Couet, J., and Breton, N. (1991). Expression Of 3-Beta-Hydroxysteroid Dehydrogenase Delta-5-Delta-4 Isomerase (3-Beta-Hsd) and 17-Beta-Hydroxysteroid Dehydrogenase (17- Beta-Hsd) In Adipose-Tissue. International Journal Of Obesity *15*, 91-99.

Lachance, Y., Luuthe, V., Labrie, C., Simard, J., Dumont, M., Delaunoit, Y., Guerin, S., Leblanc, G., and Labrie, F. (1990). Characterization Of Human 3-Beta-Hydroxysteroid Dehydrogenase Delta- 5-Delta-4-Isomerase Gene and Its Expression In Mammalian-Cells. Journal Of Biological Chemistry *265*, 20469-20475.

Lachance, Y., Luuthe, V., Verreault, H., Dumont, M., Rheaume, E., Leblanc, G., and Labrie, F. (1991). Structure Of the Human Type-II 3-Beta-Hydroxysteroid Dehydrogenase Delta-5-Delta-4 Isomerase (3-Beta-Hsd) Gene - Adrenal and Gonadal Specificity. Dna and Cell Biology *10*, 701-711.

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). Cpg Islands As Gene Markers In the Human Genome. Genomics *13*, 1095-1107.

LeersSuchcta, S., Morohashi, K., Mason, J. I., and Melner, M. H. (1997). Synergistic activation of the human type II 3 beta-hydroxysteroid dehydrogenase/Delta(5)-Delta(4) isomerase promoter by the transcription factor steroidogenic factor-1/adrenal 4-binding protein and phorbol ester. Journal Of Biological Chemistry *272*, 7960-7967.

Liao, D. Q., Pavelitz, T., Kidd, J. R., Kidd, K. K., and Weiner, A. M. (1997). Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the RNU2 locus) involves vapid intrachromosomal homogenization and rare interchromosomal gene conversion. Embo Journal *16*, 588-598.

Lifton, R. P., Dluhy, R. G., Powers, M., Rich, G. M., Cook, S., Ulick, S., and Lalouel, J. M. (1992a). A Chimeric 11-Beta-Hydroxylase Aldosterone Synthase Gene Causes Glucocorticoid-Remediable Aldosteronism and Human Hypertension. Nature *355*, 262-265.

Lifton, R. P., Dluhy, R. G., Powers, M., Rich, G. M., Gutkin, M., Fallo, F., Gill, J. R., Feld, L., Ganguly, A., Laidlaw, J. C., Murnaghan, D. J., Kaufman, C., Stockigt, J. R., Ulick, S., and Lalouel, J. M. (1992b). Hereditary Hypertension Caused By Chimeric Gene Duplications and Ectopic Expression Of Aldosterone Synthase. Nature Genetics *2*, 66-74.

Lorence, M. C., Murry, B. A., Trant, J. M., and Mason, J. I. (1990a). Human 3-Beta-Hydroxysteroid Dehydrogenase Delta-5-]4-Isomerase From Placenta - Expression In Nonsteroidogenic Cells Of a Protein That Catalyzes the Dehydrogenation Isomerization Of C21 and C19 Steroids. Endocrinology *126*, 2493-2498.

Lorence, M.C., Corbin, C.J., Kamimura, N., Mahendroo, M.S., Mason, J.I. (1990b) Structural-Analysis Of The Gene Encoding Human 3-Beta-Hydroxysteroid Dehydrogenase Delta-5-]4-Isomerase Molecular Endocrinology *4 (12)* 1850-1855

Lovett, M., Kere, J. H., and Hinton, L. M. (1991). Direct Selection - a Method For the Isolation Of Cdnas Encoded By Large Genomic Regions. Proceedings Of the National Academy Of Sciences Of the United States Of America *88*, 9628-9632.

Luo, X. R., Ikeda, Y. Y., and Parker, K. L. (1994). A Cell-Specific Nuclear Receptor Is Essential For Adrenal and Gonadal Development and Sexual-Differentiation. Cell *77*, 481-490.

Luu-The, V. L., Labrie, C., Zhao, H. F., Couet, J., Lachance, Y., Simard, J., Leblanc, G., Cote, J., Berube, D., Gagne, R., and Labrie, F. (1989). Characterization Of Cdnas For Human Estradiol 17-Beta-Dehydrogenase and Assignment Of the Gene to Chromosome-17 - Evidence Of 2 Messenger-Rna Species With Distinct 5'-Termini In Human-Placenta. Molecular Endocrinology *3*, 1301-1309.

Luu-The, V., Labrie, C., Zhao, H. F., Couet, J., Lachance, Y., Simard, J., Leblanc, G., Cote, J., Berube, D., Gagne, R., and Labrie, F. (1989). Characterization Of Cdnas For Human Estradiol 17-Beta-Dehydrogenase and Assignment Of the Gene to Chromosome-17 - Evidence Of 2 Messenger-Rna Species With Distinct 5'-Termini In Human-Placenta. Molecular Endocrinology *3*, 1301-1309.

Luu-the, V., Takahashi, M., Delaunoit, Y., Dumont, M., Lachance, Y., and Labrie, F. (1991). Evidence For Distinct Dehydrogenase and Isomerase Sites Within a Single 3-Beta-Hydroxysteroid Dehydrogenase 5-Ene-4-Ene Isomerase Protein. Biochemistry *30*, 8861-8865.

Maeda, N., and Smithies, O. (1986). The Evolution Of Multigene Families - Human Haptoglobin Genes. Annual Review Of Genetics *20*, 81-108.

Martel, C., Melner, M. H., Gagne, D., Simard, J., and Labrie, F. (1994). Widespread Tissue Distribution Of Steroid Sulfatase, 3-Beta- Hydroxysteroid Dehydrogenase Delta(5)-Delta(4) Isomerase (3-Beta- Hsd), 17-Beta-Hsd 5-Alpha-Reductase and Aromatase Activities In the Rhesus-Monkey. Molecular and Cellular Endocrinology *104*, 103-111.

Mason, J. I., Keeney, D. S., Bird, I. M., Rainey, W. E., Morohashi, K. I., LeersSucheta, S., and Melner, M. H. (1997). The regulation of 3 beta-hydroxysteroid dehydrogenase expression. Steroids *62*, 164-168.

Mason, J. I., Ushijima, K., Doody, K. M., Nagai, K., Naville, D., Head, J. R., Milewich, L., Rainey, W. E., and Ralph, M. M. (1993). Regulation Of Expression Of the 3-Beta-Hydroxysteroid Dehydrogenases Of Human Placenta and Fetal Adrenal. Journal Of Steroid Biochemistry and Molecular Biology *47*, 151-159.

McBride, M.W. (1996) Molecular Analysis Of The Human 3ß-Hydroxysteroid Dehydrogenase Δ5/Δ5 Isomerase Gene Family. Thesis.

McBride, M.W., Russell, A.J., Vass, K., Frankraue, K., Craig, N.J., Morrison, N.,Boyd, E., Szpirer, C., Sutcliffe, R.G. (1995) The Human 3-Beta-Hydroxysteroid Dehydrogenase (3-Beta-Hsd) Gene-Cluster On Chromosome 1p13 Contains A Presumptive Pseudogene - 3-Beta-Hsd And Cyp17 Do Not Segregate With Dominantly Inheritedhirsutism. Journal Of Molecular Endocrinology *15(2)* 167-176

Mendonca, B. B., Russell, A. J., Vasconcelosleite, M., Arnhold, I. J. P., Bloise, W., Wajchenberg, B. L., Nicolau, W., Sutcliffe, R. G., and Wallace, A. M. (1994). Mutation Of 3-Beta-Hydroxysteroid Dehydrogenase Type-Ii Associated With Pseudohermaphroditism In Males and Premature Pubarche or Cryptic Expression In Females. Journal Of Molecular Endocrinology *12*, 119-122.

Milewich, L., Shaw, C. E., Doody, K. M., Rainey, W. E., Mason, J. I., and Carr, B. R. (1991). 3-Beta-Hydroxysteroid Dehydrogenase-Activity In Glandular and Extraglandular Human Fetal Tissues. Journal Of Clinical Endocrinology and Metabolism *73*, 1134-1140.

Miller, W. L., Auchus, R. J., and Geller, D. H. (1997). The regulation of 17,20 lyase activity. Steroids *62*, 133-142.

Misevic, G. N., and Burger, M. M. (1990). The Species-Specific Cell-Binding Site Of the Aggregation Factor From the Sponge Microciona-Prolifera Is a Highly Repetitive Novel Glycan Containing Glucuronic-Acid, Fucose, and Mannose. Journal Of Biological Chemistry *265*, 20577-20584.

Mita, M. (1992). Involvement Of Cyclic Adenosine-3',5'-Monophosphate In Methylation During 1-Methyladenine Production By Starfish Ovarian Follicle Cells. Invertebrate Reproduction & Development *22*, 11-16.

Monaco, A. P., Neve, R. L., Colletti-feener, C., Bertelson, C. J., Kurnit, D. M., and Kunkel, L. M. (1986). Isolation Of Candidate Cdnas For Portions Of the Duchenne Muscular- Dystrophy Gene. Nature *323*, 646-650.

Morel, Y., Mebarki, F., Rheaume, E., Sanchez, R., Forest, M. G., and Simard, J. (1997). Structure-function relationships of 3 beta-hydroxysteroid dehydrogenase:

Contribution made by the molecular genetics of 3 beta- hydroxysteroid dehydrogenase deficiency. Steroids *62*, 176-184.

Morgan, J. G., Dolganov, G. M., Robbins, S. E., Hinton, L. M., and Lovett, M. (1992). The Selective Isolation Of Novel Cdnas Encoded By the Regions Surrounding the Human Interleukin-4 and Interleukin-5 Genes. Nucleic Acids Research *20*, 5173-5179.

Morissette, J., Rheaume, E., Leblanc, J. F., Luuthe, V., Labrie, F., and Simard, J. (1995). Genetic-Linkage Mapping Of Hsd3b1 and Hsd3b2 Encoding Human Type-I and Type-Ii 3-Beta-Hydroxysteroid Dehydrogenase/Delta-5-Delta-4- Isomerase Close to D1s514 and the Centromeric D1z5 Locus. Cytogenetics and Cell Genetics *69*, 59-62.

Mornet, E., Dupont, J., Vitek, A., and White, P. C. (1989). Characterization Of 2 Genes Encoding Human Steroid 11-Beta- Hydroxylase (P-45011-Beta). Journal Of Biological Chemistry *264*, 20961-20967.

Morohashi, K., Hatano, O., Nomura, M., Takayama, K., Hara, M., Yoshii, H., Takakusu, A., and Omura, T. (1995). Function and Distribution Of a Steroidogenic Cell-Specific Transcription Factor, Ad4bp. Journal Of Steroid Biochemistry and Molecular Biology *53*, 81-88.

Morris, D. J. (1981). The Metabolism and Mechanism Of Action Of Aldosterone. Endocrine Reviews 2, 234-247.

Morrison, N., Nickson, D. A., McBride, M. W., Mueller, U. W., Boyd, E., and Sutcliffe, R. G. (1991). Regional Chromosomal Assignment Of Human 3-Beta-Hydroxy-5-Ene Steroid Dehydrogenase to 1p13.1 By Nonisotopic Insitu Hybridization. Human Genetics *87*, 223-225.

Nakanishi, M., Matsuura, K., Kaibe, H., Tanaka, N., Nonaka, T., Mitsui, Y., and Hara, A. (1997). Switch of coenzyme specificity of mouse lung carbonyl reductase by substitution of threonine 38 with aspartic acid. Journal Of Biological Chemistry *272*, 2218-2222.

Naville, D., Rainey, W. E., Milewich, L., and Mason, J. I. (1991). Regulation Of 3-Beta-Hydroxysteroid Dehydrogenase/Delta-5-]4- Isomerase Expression By Adrenocorticotropin In Bovine Adrenocortical- Cells. Endocrinology *128*, 139-145.

Naville, D., Rainey, W. E., Milewich, L., and Mason, J. I. (1991). Regulation Of 3-Beta-Hydroxysteroid Dehydrogenase/Delta-5-]4- Isomerase Expression By Adrenocorticotropin In Bovine Adrenocortical- Cells. Endocrinology *128*, 139-145.

Nei, M., Gu, X., and Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proceedings Of the National Academy Of Sciences Of the United States Of America *94*, 7799-7806.

Neumann, I., Douglas, A. J., Pittman, Q. J., Russell, J. A., and Landgraf, R. (1996). Oxytocin Released Within the Supraoptic Nucleus Of the Rat-Brain By Positive Feedback Action Is Involved In Parturition-Related Events. Journal Of Neuroendocrinology *8*, 227-233.

Neve, R. L., Stewart, G. D., Newcomb, P., Vankeuren, M. L., Patterson, D., Drabkin, H. A., and Kurnit, D. M. (1986). Human Chromosome-21-Encoded Cdna Clones. Gene *49*, 361-369.

New, M. I. (1994). 21-Hydroxylase Deficiency Congenital Adrenal-Hyperplasia. Journal Of Steroid Biochemistry and Molecular Biology *48*, 15-22.

Ozeretskovskaya, O. L., Ilinskaya, L. I., and Vasyukova, N. I. (1994). The Mechanisms Of Elicitation Of Plant Systemic Resistance Against Diseases. Russian Journal Of Plant Physiology *41*, 550-556.

Parimoo, S., Patanjali, S. R., Shukla, H., Chaplin, D. D., and Weissman, S. M. (1991). Cdna Selection - Efficient Pcr Approach For the Selection Of Cdnas Encoded In Large Chromosomal Dna Fragments. Proceedings Of the National Academy Of Sciences Of the United States Of America *88*, 9623-9627.

Patanjali, S. R., Parimoo, S., and Weissman, S. M. (1991). Construction Of a Uniform-Abundance (Normalized) Cdna Library. Proceedings Of the National Academy Of Sciences Of the United States Of America *88*, 1943-1947.

Payne, A. H., Abbaszade, I. G., Clarke, T. R., Bain, P. A., and Park, C. H. J. (1997). The multiple murine 3 beta-hydroxysteroid dehydrogenase isoforms: Structure, function, and tissue- and developmentally specific expression. Steroids *62*, 169-175.

Pon, L. A., and Ormejohnson, N. R. (1986). Acute Stimulation Of Steroidogenesis In Corpus-Luteum and Adrenal- Cortex By Peptide-Hormones - Rapid Induction Of a Similar Protein In Both Tissues. Journal Of Biological Chemistry *261*, 6594-6599.

Pon, L. A., Hartigan, J. A., and Ormejohnson, N. R. (1986). Acute Acth Regulation Of Adrenal Corticosteroid Biosynthesis - Rapid Accumulation Of a Phosphoprotein. Journal Of Biological Chemistry *261*, 3309-3316.

Pozzi, A.G., Lantos, C.P. and Ceballos, N.R. (1996) Mitochondrial Localisation of 3ß-Hydroxysteroid Dehydrogenase 5-ene Isomerase In The Interrenals of The Toad *Bufo arenarum H*. General and Comparative Endocrinology *103*, 176-181.

Rabe, T., Brandstetter, K., Kellermann, J., and Runnebaum, B. (1982). Partial Characterization Of Placental 3-Beta-Hydroxysteroid Dehydrogenase (Ec 1.1.1.145), Delta-4-5-Isomerase (Ec 5.3.3.1) In Human Term Placental Mitochondria. Journal Of Steroid Biochemistry *17*, 427-433.

Rheaume, E., Lachance, Y., Zhao, H. F., Breton, N., Dumont, M., Delaunoit, Y., Trudel, C., Luuthe, V., Simard, J., and Labrie, F. (1991). Structure and Expression Of a New Complementary-Dna Encoding the Almost Exclusive 3-Beta-Hydroxysteroid Dehydrogenase Delta-5-Delta-4- Isomerase In Human Adrenals and Gonads. Molecular Endocrinology *5*, 1147-1157.

Rheaume, E., Sanchez, R., Mebarki, F., Gagnon, E., Carel, J. C., Chaussain, J. L., Morel, Y., Labrie, F., and Simard, J. (1995). Identification and Characterization Of the G15d Mutation Found In a Male-Patient With 3-Beta-Hydroxysteroid Dehydrogenase (3-Beta-Hsd) Deficiency - Alteration Of the Putative Nad-Binding Domain Of Type-Ii 3-Beta-Hsd. Biochemistry *34*, 2893-2900.

Russell, A., Nazer, H., Shams, A., Sjovall, J., and Sutcliffe, R. (1995). No Linkage to the 3-Beta-Hsd Gene-Cluster In a Kindred Affected With 3-Beta-Hydroxy-Delta(5)-C-27-Steroid Dehydrogenase-Deficiency and Early-Onset Hepatic-Failure. Human Genetics *95*, 586-588.

Russell, A.J. (1993) Molecular Genetics of Human 3ß-Hydroxysteroid Dehydrogenase. PhD Thesis.

Russell, D. W., and Wilson, J. D. (1994). Steroid 5-Alpha-Reductase - 2 Genes 2 Enzymes. Annual Review Of Biochemistry *63*, 25-61.

Rutherfurd, K. J., Chen, S., and Shively, J. E. (1991a). Isolation and Amino-Acid-Sequence Analysis Of Bovine Adrenal 3-Beta- Hydroxysteroid Dehydrogenase Steroid Isomerase. Biochemistry *30*, 8108-8116.

Rutherfurd, K. J., Chen, S., and Shively, J. E. (1991b). Affinity Labeling Of Bovine Adrenal 3-Beta-Hydroxysteroid Dehydrogenase Steroid Isomerase By 5'-[P-(Fluorosulfonyl)Benzoyl]Adenosine. Biochemistry *30*, 8116-8123.

Sanchez, R., Rheaume, E., Laflamme, N., Rosenfield, R. L., Labrie, F., and Simard, J. (1994). Detection and Functional-Characterization Of the Novel Missense Mutation Y254d In Type-Ii 3-Beta-Hydroxysteroid Dehydrogenase (3- Beta-Hsd) Gene Of a Female-Patient With Nonsalt-Losing 3-Beta-Hsd Deficiency. Journal Of Clinical Endocrinology and Metabolism *78*, 561-567.

Sanne, J. L., and Krueger, K. E. (1995). Expression Of Cytochrome-P450 Side-Chain Cleavage Enzyme and 3-Beta- Hydroxysteroid Dehydrogenase In the Rat Central-Nervous-System - a Study By Polymerase Chain-Reaction and In-Situ Hybridization. Journal Of Neurochemistry *65*, 528-536.

Sasano, H., Mason, J. I., and Sasano, N. (1989). Immunohistochemical Analysis Of Cytochrome-P-450 17-Alpha-Hydroxylase In Pig Adrenal-Cortex, Testis and Ovary. Molecular and Cellular Endocrinology *62*, 197-202.

Sasano, H., White, P. C., New, M. I., and Sasano, N. (1988). Immunohistochemical Localization Of Cytochrome P-450c21 In Human Adrenal-Cortex and Its Relation to Endocrine Function. Human Pathology *19*, 181-185.

Sauer, L. A., Chapman, J. C., and Dauchy, R. T. (1994). Topology Of 3-Beta-Hydroxy-5-Ene-Steroid Dehydrogenase/Delta(5)- Delta(4)-Isomerase In Adrenal-Cortex Mitochondria and Microsomes. Endocrinology *134*, 751-759.

Seeburg, P.H., Shine, J., Martial, J.A., Ullrich, A., Baxter, J.D. and Goodman, H.M. (1977). Nucleotide Sequence Of Part Of The Gene For Human Chorionic Somatomammotrophin: Purification Of DNA Complementary To Predominant mRNA Species. Cell *12*, 157-165.

Simard, J., Rheaume, E., Mebarki, F., Sanchez, R., New, M. I., Morel, Y., and Labrie, F. (1995). Molecular-Basis Of Human 3-Beta-Hydroxysteroid Dehydrogenase-Deficiency. Journal Of Steroid Biochemistry and Molecular Biology *53*, 127-138.

Simard, J., Rheaume, E., Sanchez, R., Laflamme, N., Delaunoit, Y., Luuthe, V., Vanseters, A. P., Gordon, R. D., Bettendorf, M., Heinrich, U., Moshang, T., New, M. I., and Labrie, F. (1993). Molecular-Basis Of Congenital Adrenal-Hyperplasia Due to 3-Beta- Hydroxysteroid Dehydrogenase-Deficiency. Molecular Endocrinology *7*, 716-728.

Sinnott, P., Collier, S., Costigan, C., Dyer, P. A., Harris, R., and Strachan, T. (1990). Genesis By Meiotic Unequal Crossover Of a Denovo Deletion That Contributes to Steroid 21-Hydroxylase Deficiency. Proceedings Of the National Academy Of Sciences Of the United States Of America *87*, 2107-2111.

Siteri, P.K. and MacDonald, P.C. (1967) Oestrogen Biosynthesis During Pregnancy. Journal Of Clinical Endocrinology and Metabolism *26*, 751-761.

Stocco, D. M., and Clark, B. J. (1996). Regulation Of the Acute Production Of Steroids In Steroidogenic Cells. Endocrine Reviews *17*, 221-244.

Stocco, D. M., and Clark, B. J. (1996). Role Of the Steroidogenic Acute Regulatory Protein (Star) In Steroidogenesis. Biochemical Pharmacology *51*, 197-205.

Stocco, D. M., and Clark, B. J. (1997). The role of the steroidogenic acute regulatory protein in steroidogenesis. Steroids *62*, 29-36.

Stocco, D. M., and Clark, B. J. (1997). The role of the steroidogenic acute regulatory protein in steroidogenesis. Steroids *62*, 29-36.

Strachan, T and Read A.P. (1996) In: *Human Molecular Genetics* 1st Edition BIOS Scientific Publishers Limited .

Tajima, T., Fujieda, K., Nakae, J., Shinohara, N., Yoshimoto, M., Baba, T., Kinoshita, E., Igarashi, Y., and Oomura, T. (1995). Molecular Analysis Of Type-Ii 3-Beta-Hydroxysteroid Dehydrogenase Gene In Japanese Patients With Classical 3-Beta-Hydroxysteroid Dehydrogenase-Deficiency. Human Molecular Genetics *4*, 969-971.

The, V. L., Lachance, Y., Labrie, C., Leblanc, G., Thomas, J. L., Strickler, R. C., and Labrie, F. (1989). Full Length Cdna Structure and Deduced Amino-Acid Sequence Of Human 3-Beta-Hydroxy-5-Ene Steroid Dehydrogenase. Molecular Endocrinology *3*, 1310-1312.

Thomas, J. L., Berko, E. A., Faustino, A., Myers, R. P., and Strickler, R. C. (1988). Human Placental 3-Beta-Hydroxy-5-Ene-Steroid Dehydrogenase and Steroid 5-]4-Ene-Isomerase - Purification From Microsomes, Substrate Kinetics, and Inhibition By Product Steroids. Journal Of Steroid Biochemistry *31*, 785-793.

Thomas, J. L., Evans, B. W., and Strickler, R. C. (1997). Affinity radiolabeling identifies peptides associated with the isomerase activity of human type I (placental) 3 beta-hydroxysteroid dehydrogenase isomerase. Biochemistry *36*, 9029-9034.

Thomas, J. L., Myers, R. P., and Strickler, R. C. (1989). Human Placental 3-Beta-Hydroxy-5-Ene-Steroid Dehydrogenase and Steroid 5-]4-Ene-Isomerase - Purification From Mitochondria and Kinetic Profiles, Biophysical Characterization Of the Purified Mitochondrial and Microsomal-Enzymes. Journal Of Steroid Biochemistry *33*, 209-217.

Thomas, J. L., Myers, R. P., and Strickler, R. C. (1991). Analysis Of Coenzyme Binding By Human Placental 3-Beta-Hydroxy-5-Ene- Steroid Dehydrogenase and Steroid 5-]4-Ene-Isomerase Using 5'-[Para- (Fluorosulfonyl)Benzoyl]Adenosine, an Affinity Labeling Cofactor Analog. Journal Of Steroid Biochemistry and Molecular Biology *39*, 471-477.

Thomas, J. L., Nash, W. E., and Strickler, R. C. (1996). Physiological 3-Beta-Hydroxy-5-Ene Steroid Substrates Bind to 3-Beta- Hydroxysteroid Dehydrogenase Without the Prior Binding Of Cofactor. Journal Of Steroid Biochemistry and Molecular Biology *58*, 211-216.

Thomas, J. L., Nash, W. E., Myers, R. P., Crankshaw, M. W., and Strickler, R. C. (1993). Affinity Radiolabeling Identifies Peptides and Amino-Acids Associated With Substrate-Binding In Human Placental 3-Beta-Hydroxy-Delta-5- Steroid Dehydrogenase. Journal Of Biological Chemistry *268*, 18507-18512.

Thomas, J. L., Strickler, R. C., Myers, R. P., and Covey, D. F. (1992). Affinity Labeling Of Human Placental 3-Beta-Hydroxy-Delta-5-Steroid Dehydrogenase and Steroid Delta-Isomerase - Evidence For Bifunctional Catalysis By a Different

Conformation Of the Same Protein For Each Enzyme-Activity. Biochemistry *31*, 5522-5527.

Trapp, T., Tunn, S., and Krieg, M. (1992). Purification and Properties Of the 5-Alpha-Dihydrotestosterone 3- Alpha(Beta)-Hydroxysteroid Dehydrogenase From Human Prostatic Cytosol. Journal Of Steroid Biochemistry and Molecular Biology *42*, 321-327.

Tunn, S., Haumann, R., Hey, J., Fluchter, S. H., and Krieg, M. (1990). Effect Of Aging On Kinetic-Parameters Of 3-Alpha(Beta)-Hydroxysteroid Oxidoreductases In Epithelium and Stroma Of Human Normal and Hyperplastic Prostate. Journal Of Clinical Endocrinology and Metabolism *71*, 732-739.

Walker, B. R., and Edwards, C. R. W. (1991). 11-Beta-Hydroxysteroid Dehydrogenase and Enzyme-Mediated Receptor Protection - Life After Licorice. Clinical Endocrinology *35*, 281-289.

White, P. C., and Pascoe, L. (1992). Disorders Of Steroid 11-Beta-Hydroxylase Isozymes. Trends In Endocrinology and Metabolism *3*, 229-234.

White, P. C., New, M. I., and Dupont, B. (1986). Structure Of Human Steroid 21-Hydroxylase Genes. Proceedings Of the National Academy Of Sciences Of the United States Of America *83*, 5111-5115.

White, P. C., New, M. I., and Dupont, B. (1987). Congenital Adrenal-Hyperplasia .1. New England Journal Of Medicine *316*, 1519-1524.

White, P. C., Tusieluna, M. T., New, M. I., and Speiser, P. W. (1994). Mutations In Steroid 21-Hydroxylase (Cyp21). Human Mutation *3*, 373-378.

Wierenga, R. K., Terpstra, P., and Hol, W. G. J. (1986). Prediction Of the Occurrence Of the Adp-Binding Beta-Alpha-Beta-Fold In Proteins, Using an Amino-Acid Sequence Fingerprint. Journal Of Molecular Biology *187*, 101-107.

Wilson JD and Siiteri PK (1973) Developmantal pattern of testosterone synthesis in the fetal gonad of the rabbit. Endocrinology 92:1182-1191

Witchel, S. F., Nayak, S., SudaHartman, M., and Lee, P. A. (1997). Newborn screening for 21-hydroxylase deficiency: Results of CYP21 molecular genetic analysis. Journal Of Pediatrics *131*, 328-331.

Wu, L., Einstein, M., Geissler, W. M., Chan, H. K., Elliston, K. O., and Andersson, S. (1993). Expression Cloning and Characterization Of Human 17-Beta-Hydroxysteroid Dehydrogenase Type-2, a Microsomal-Enzyme Possessing 20-Alpha-Hydroxysteroid Dehydrogenase-Activity. Journal Of Biological Chemistry *268*, 12964-12969.

Zang, Y., Dufort, I., Soucy, P., Labrie, F. and Luu-The, V. (1995). Cloning And Expression Of Human Type V 17ß-Hydroxysteroid Dehydrogenase. 77th Annual Meeting Of The Endocrine Society. *622* (P3-614) Abstract.

Zhao, H. F., Labrie, C., Simard, J., Delaunoit, Y., Trudel, C., Martel, C., Rheaume, E., Dupont, E., Luuthe, V., Pelletier, G., and Labrie, F. (1991). Characterization Of Rat 3-Beta-Hydroxysteroid Dehydrogenase Delta-5- Delta-4 Isomerase Cdnas and Differential Tissue-Specific Expression Of the Corresponding Messenger-Rnas In Steroidogenic and Peripheral- Tissues. Journal Of Biological Chemistry *266*, 583-593.

CHAPTER 8

Appendices

Appendix i

Type I (genbank accession number m38180; exons are boxed):

```
   1  TCTGTGAACT TAAACTAGAC CAGGAGAAAT CTGCAAGTGC AATTTGAATT
  51  GTGCCCTTGT TCAGTAAATC CATCTGGAGG GAAATCCATC TAAGCCCATT
 101  ATCTCCTCTC TCAGACTATC ATTTGAAATG TTGCTGCTGT CTTACAAAAT
 151  TTAGAAATTT GCAATGGTGA CAGCTCCAAT CACTTAGTTT TATGTGAGGT
 201  GACTTTCTTT CCTACCCAAT ATATCTGAAA CCTCAATGGT CATTTGCCCC
 251  TGGAGAGCCC TGGGTTCCTA GATGAACAAC ATCATTGAAG GAAAGCTATC
 301  TGGTGATAAC CACTTAATGA TATGGCCTGT ATCACAAGAG TGGGGATTTG
 351  TGGTTGTATG CCCCAATGAC TCTTATCATA GGTGTTCGAA CTTATCCATA
 401  TCTTTGGCTG CAAAACGATT CTCAGCAAAT TCTTGAACTG GACATACTGT
 451  GAATGCTTCT GAAGCCTATG CCCCAGAACT GCCCAAATAT CACCGCCATA
 501  AGCTCCCCAA GACACTTTAC TCAAAAAAAT TGAAATATAA TTTACATGCA
 551  TTAAAATGCC CAGGTCTTCA GTGTTCAGTT CACTCGAACT TCACAGTTAT
 601  ATATGCCCAT GTAAGTAGCA CCCCAAAAAA GATATAGACC ATTTCCGTCA
 651  CTCCAGAAAG TTGTTTTCTG CCTGTTCCCT GTTAATTCCC TGACCTTCCC
 701  CCAATCTACC TCATAGGCAA CTACTTCCTG ATTTTTATTC CCACAGTTCT
 751  CATTTTATTT CTATATTTCT CTAATTTTTT CTTCTAAGAT CTCTGTGGTT
 801  CTGGATATAT ATTTAAATAT AAAATTATTT TGAAATGAAG TTTTGTGCAG
 851  GGAGTACAGA AAAATTGACC GTTGATTGTC TCTGTTGTTT ATTTTTCTAT
 901  TTCATGATTT TCATTATTTT TATTATTGGC TTACATTTAT TAATTTTGTT
 951  TTTTTTTTTC ACTCGTTTTT CACATTTTTT TTTTTTTTTT TTGAGATGGA
1001  GTCTTGCTCT GTCACCCAGG CTGAATGGCA GGGGCACCAT CTCAGCTCAC
1051  TGCAACCTCT GCTCCCGGGT TCAAGCAATT CTCCCACCTC AGCCTCCGAA
1101  GTAGCTGGGA TTACAGGTGC CCGCCACTGC ACCGGGCTAA TTTTTGTATT
1151  TTCAGTAGAG GTGGGGTTTC ATCATGTTGA CCAGCTGGTA TCAAACTCCT
1201  GACTTTGTGA TCTGCACGCC TCAGCCTCCC AAAGTGCTGC GATTACAGGC
1251  GTGAGCCACT GCACCCAGCC CCGTTTTTCA GGATCCCATA GGAGGAGAGA
1301  GCAATGAGTA CATGGCCAGA GATCAAAGTG ATAAGGGTTG GGCCAGAAGC
1351  CACAGTGCAT AAAGCTTCAG ACTTGCCACA GGAAATGAGG TGAGAAGTAC
1401  GTCCACTCTT CTGTCCAGCT TTTAACAATC TAACTAATGG TTAGAGATTT
```

1451 TTCATTTTCT TTCAGCTACT CCTGCAGTGG TGGGGACACA GAATGTTTGC

1501 AAAAAAAATG GGGTGGAGGA AAATGAGGCA TCTGTGTGAG TATATAACCA

1551 TTTGACATCT CTTTTTAGCC CTCTCCAGGG TCACCCTAGA ATCAGATCTG

1601 CTCCCCAGCA TCTTCTGTTT CCTGGTCAGT GATTCCTGCT ACTTTGGATG

1651 GCCATGACGG GCTGGAGCTG CCTTGTGACA GGAGCAGGAG GGTTTCTGGG

1701 ACAGAGGATC ATCCGCCTCT TGGTGAAGGA GAAGGAGCTG AAGGAGATCA

1751 GGGTCTTGGA CAAGGCCTTC GGACCAGAAT TGAGAGAGGA ATTTTCTAGT

1801 AAGTAAACTT GGGTCATGGG TGTGTCGTTC CATCTTAAAC ACTGCATGTA

1851 TGTGGGGGGA GATGGACCTT GTCTAGCAAG TTATTGAAAT TTGTAGCCAA

1901 ATCTAAGCCA ATCCCACATC CAAAGTCATC AAGAAATAAA TATTAAATAG

1951 TATAAAATGG CATAGTATGA AAGATACTGG GTGGGATTTC CAGAGACTAG

2001 ATTCTGGCCC TGACCCAGAA CTTGAGAAGC AGCCACCTCA GCCTCCAGGC

2051 CTCTTTTCTT CTCATCTAGA AAATCCCACA TCAGTTCTTT GTCTTGTCTG

2101 CAACTCTTTT ATGTTCTGAA GCTTTTGTCT TGGCAATTGC TGTACAACAT

2151 TCATAAAGGA CACTATTACC CTGGAGACCT CACCAATGGG TCATTGTCTC

2201 TCTAGAACGT GCCAGGCTGT GTTTTGTTTT GCAGCTTCCT TAGTATCCTT

2251 TGTGAGGGCA AAATAAAAGG AAGTTAAGGT AAAGATGAGT GTATAAGTGG

2301 GTGTGTGTGA CAGAGAGAGA GGAAAACAAT AAGGGGGAGA GAGAGTCGGG

2351 GAGAGAGAGT CGGGGAGAGA GACTCAGAGA GAGAGACAAT GAGAAAGACA

2401 GAGAAATAAT GAAAAATATA TAGTGAGAGA GAGAGCATCA GTGAGAGAAT

2451 GTGTGTGCAC AGTGCACAGC ACAGACCAGA GACAGAGTAA GAGGCAGTAT

2501 AAGGCCAGAC ATGCCTCATT TAGATTTTGC ATATATGGCT TTTTTTTAAA

2551 AAAAAAAAAA CAAAACATTC ACCTCTGTTG CTCATCATCA AAAAGAAATT

2601 GACTGCAGAA TTTTCTGTGG CAACTCCTTG TACAAATGGC CCCAAGAGGC

2651 AGGTAGAAAA GGCATCACAT CCAACAGGAC AATGAACGCA TCAAGGTCCG

2701 GAATCACAAA GCAACATGCT AAGGCCATCT GATTGATGCG TAGCAGAGCT

2751 CAGACCAGAG CTTTATCCAT AGGCTATTTC ACCTGGCCCC AGAGCTGATG

2801 CTTGGCAAAG TGGATCTTCA GTCTCTCCCT TTGCTCCTTG GTGTGCCCCA

2851 AAACAATTCC CTGTCCTCCT CCCCAACCCA GGGGATTTGG GGGACTGTGT

2901 GTCTACATGT GGGCTCTGGC TGCCCAGAGG TCCTGCCCTG CCCACAGAGA

2951 ATCCCCCACT TGACAGCACT CAGGGACCAC ACCCAACACT CTACTTTTGC

3001 TCTTCTCTGA CCTCTTGAAA TGTTCGATTT CTTCCACCCT CTGTTTTAAT

3051 GTTGTCTTTG ATAAACAAAT ATGAAATGCT TATAATCCTT GCCTGTGTCT

3101 CAGCAATATT TGCAATTTAA CATCTGAAGC CAGAAAAAAG CAATGCTCAG

3151 CCTAGAGGTA TTGTTCCCCT ATGGTAAAAA TGCAGATCAT ATAATAGACA

3201 GGCAGGGATC TGTATGAAGA TAAAAACAGA GAAAGCATTC AGTGTTCTTT

3251 TTGTGCCAGG CACTTGGCCA AAGGTTTGCC AACTCATTTG CTCCTCATAA

3301 CAACCCTATG AGATCGGTAA TATTATCCTC CTTTTTTAAA GATGAGGAGA

3351 TAGAAACAGA GAGGTTAGTT AACTTTTTGA AGGTCCCACA GTTAGCATAT

3401 GGCAAAGCTA GATTTGAACC CAGACAGACT AGTGCAGCAT CTGTGCTCTT

3451 AAGCAGTAGT TTGTCTTACA GCTCCTTTAG GCCTTCCTCT AGCCCAACTT

3501 ACACCATAAA AGCAACTCCA TTTAGTCTAC ATAAAACACT TGCCAAGTCT

3551 ACAGTGCCTT GGAGAAGAAG GGAGCCATGA CTTCATCAAC AAAGTACCTT

3601 TTCCTAGTCC CTCTAGATCC CTGGGAACTC ACTGGGGAAG AGTCCTTTTC

3651 TCTCAGCTTT GGACTTCTTC CACTTTGACA CTCCTTTGAG CCAGTCTGTC

3701 TGAAGTGGCA GCAAAGAAGA AGCAGAGGAA GGCAGCAGGA AGGGCAGAGG

3751 TGGAACTAGA ACCAGAGAAC ATCCCACTCT CAGGCTGTTA CCCACATCAC

3801 CACCAGCAGC CTCTTCCCTG ACTCCCCCAC TCCAGGACCT TCTTGGAGTT

3851 GCCTGACAAC GGGAAGAGGT TGGGAGTAGA CAAAACATGG CAGAGGCATC

3901 AGGAGAGCAA GTGGTCCTGC TTGTACCCCA GGAAGGAATT TCTCCAGCTT

3951 CGAATTTTTC ACTCTGAATT TTGCTCTCAC TGGTCTTTGG TCTTTGTTGC

4001 CAATTTCAGT TAAACTGGCT TATTCTTTGG AAAAGAATGA TATTAAACAA

4051 CAGAAAAGCT GATATTAAAC AGCCAGTCCT GCACCTGGTG TACAGGGTAC

4101 AAGCTTGGAG GCAGGGTGGG CGTGTTTAAG CCTTCACTAG AAACGTGACC

4151 TTAGATCTGT CCCTTAACCT CTCTGAGACT CAGTTTCCTC ACTGATAAAG

4201 TGGGACTCAG AATACCTACC TCATGAGATA GCTGTAAAAA GAAATAAATA

4251 TCCTGTAAGT TGCATACAGA TGATAAAGCA CAAACAAGTG TACAGTATTT

4301 ATCCAAGCTC CATGTAGCAT CGTGTTCATT TTGTGGGCTA ATTCCAGGTG

4351 ACAACCTTAC TGGCAGCGTC ACTTCCTGTG GATTTCAGAG GAACCTGAAC

4401 ACATTGATGT TGTAACCACT GCCTCCAGCT AGACAGCTGC GATGTTAGTT

4451 CATTGGTGGT TAACTCATTT TCAGCATTGC TACCTGCTTT TGATGATATC

4501 TGATAATAGT TTTCCAGGGA AATTCATCAT TGAAGCCCAC CCACAATTCC

4551 TGAGTAATGA CTCCACTCTC CTCCCAGTTG CCCAAGCAGG AAACTGTAGG

4601 GGTGATTTTA TCTTCTCCTT TTCCTATCAG CTCAGATCCA ATAGGTACCA

4651 GGACCTTCAA GGTAAGGTCT GAAATTCAAA CCTGGCATAT CTCTCCCCTC

4701 CTCCGCTTCC ACTGCTGGAC CCTGCTTCCA GTCCTCATCG AAACTTTGGC

4751 ACAGACATTC TGATTAACCA CTTGATCATC AGTCCTTCTC AAATCCCCTA

4801 TACCAGTCCA TTTCATATCC TCAAGCCAAG ATTAACTTCC TACACCACAA

4851 CCCTGCTTTT GTTCCTTTTT TTGCTGTTGC TTGTGTGGTT TTCTTTTTTT

4901 TTTTTCTTTT TTTTTTCTTT TTTTTTCTTT TTTGAGACAA GGTGGTCTTG

4951 CTCTACCACC CAGGCTAGAG TGCAGAGGTG TGATCTTGGC TCACTGCCAG

5001 GAACCAGAGA CCCCAGCCTC CATTTATCCT CCCACCTCAG TCTGCAGAGT

5051 AGCTGGGACT GCAGGCAGGC ACATGCCACC ATGCCCAGCT ATTTCATGTA

5101 TTTTAATAGA AATGGGATTT CCCCATGTTG TCCAAGCTGG TCTCAACCTC

5151 CTGCGCTCCA GTGATCCATC CATCACAGCC TCCCAAAGTG CTGGAATAAC

5201 AGGCATGAGC CACGGGGCTT GGCCTTTTTG TCATTTTATT CCAAATATAA

5251 ATAAGGGGTT TTCTTTAACA TGATCCTGGA GTCCCAAAAT TATGATTCTA

5301 TACATGATTT CTCATTACTT ACAGAATTAA CTTCAAACTT TTCCCTCATA

5351 CTCTCTACAA TCACACTACT CTCCTTAGAA AATTTGATGG GCTTCCCATG

5401 ACTCTAGCAA TCCTTATATT CCAGGCATAC CTGACAGCTC ACTGCTCAGA

5451 TTACTTTTCT AAAATGAGAT GGGAGAGAAA CAGATGTTTG CTCTTTCCAA

5501 GAAATTGATA CACATGCTGA TTTCCTGTGCC TTTTTTACTT GTTCTTTCCG

5551 TAGAATGTAC ACCCTCCACT CTAACACCCT ACTCTAACCA TCCTTGAACA

5601 CCTATGTAAC ATCACCTTTA TCAGAAAACT TCTCAGCCAG ATACAGAAAT

5651 CATTCCAATG ACCTGACCTG TGTTCACACA GAACTCCAGA ACAAGACCAA

5701 GCCTGACAGTG CTGGAAGGAG ACATTCTGGA TGAGCCATTC CTGAAGAGAG

5751 CCTGCCAGGA CGTCTCGGTC ATCATCCACA CCGCCTGTAT CATTGATGTC

5801 TTCGGTGTCA CTCACAGAGA GTCTATCATG AATGTCAATG TGAAAGGTAT

5851 GGTAGGCTGG GGAGGAGATG CAGCAAGGTG GGGAATTAAG GATCACAAAG

5901 AAGGGCCAGG AAGGGAAGAG AAGTCCCTCC ACTCAACACC TGCTGTGCTC

5951 TGGGCCAAGT GCCTTTGCTG ATCACTACGA ATAGGAGAGT TCAAGACTGC

6001 TAACTTTAGT TTTTTAGATG ATAAAACTAG GACTGAGAGA GGGCAAGTAA

6051 CTTGTCCAAG GTCCCCCAGG TAAGTAAGCA GGTAGGAGAG TTAGACTTTA

6101 AACTCATCCC TGTGTGACTC CAAAGGCTCT TTCTACTGTG ACTCCAAAAG

6151 CTCTTTCTAC TGTGGTTCCA ATCAAAAGTC AACTAATTTC TGACTTCAGA

6201 CTCTTGATAC CCAGACACCC CTTGCCTCCC AGGCCAGAAG CAGACCTTGC

6251 AGGTGCCACT GTAGTCACCA TTTTGAACCT TGTGTGTAGG CTGATGAAAA

6301 CATTCAGAGC CCTCCTGCCC ACCTCAAACA AAAAGTCCTC TCGAGAGAAC

6351 TAGCAAAGCT GGCTCACAAG GTCTGTCAGG ACAGAATTAT CCAGCACATG

6401 CCTTCCCACA ATATCTTTTT AACAGTGAGA TTTTTCCAGT ATCAAGAATA

6451 CAGTCTCTTC AGCTCACCAC AGGGCCTACT ATTCCAGAGC CGTTTTCAGC

6501 CAGGTGCCCA AAGTGCACCC TCATTTAATC TTCATCCCTT AAATGCTTGG

6551 CGTTTCCTTT CAAAGTATGG CTTTCCAACT ACCTTAAATT CAGACACATT

6601 TCCTTAAGGA ATCTTCTCAG AAATCTCACC CTCTGATCTC CAGAGTCCAG

6651 TCTCTTCAGC ACCATAGCAC CCCAAACCCC GGCCTGGCCC TCCGCATTTA

6701 GTCTACTCTG ACTCTTTTTC CAAAGCAACC AATTGCTTTT GACTCTCCAG

7751 GTACACACAC ACACACATCC ATGAAACCCA AAAGTCTAGA AATTCTTACT

7801 CTCCAGAACC CACATGTCAG TTTCTCTTCA TTTTGGTACT TTTCTTATGG

7851 CTGTAGTACG ACCAAATCTC AGACAGAACC ACAGAAGAAT GTACCCTGAG

7901 TCTGTTACAA CCACCATATT TGGGAGTGGG GGGTGGGGCA CATAGATCTG

7951 TGTTCGTGGT TGGCACCTCT TAGGGATATA TCCTGACAGT GACAATATGC

8001 TCTTCATGGA CAGGTACCCA GCTCCTGTTA GAGGCCTGTG TCCAAGCTAG

8051 TGTGCCAGTC TTCATCTACA CCAGTAGCAT AGAGGTAGCC GGGCCCAACT

8101 CCTACAAGGA AATCATCCAG AATGGCCATG AAGAAGAGCC TCTGGAAAAC

8151 ACATGGCCCG CTCCATACCC ACACAGCAAA AAGCTTGCTG AGAAGGCAGT

8201 ACTGGCGGCT AACGGGTGGA ATCTGAAAAA CGGCGGCACC CTGTACACTT

8251 GTGCCTTACG ACCCATGTAT ATCTATGGGG AAGCAAGCCG ATCCCTTTCT

8301 GCTAGTATAA ACGAGGCCCT GAACAACAAT GGGATCCTGT CAAGTGTTGG

8351 AAAGTTCTCC ACTGTTAACC CAGTCTATGT TGGCAATGTG CCCTGGGCCC

8401 ACATTCTGGC CTTGAGGGCC CTGCAGGACC CCAAGAAGGC CCCAAGCATC

8451 CGAGGACAGT TCTACTATAT CTCAGATGAC ACGCCTCACC AAAGCTATGA

8501 TAACCTTAAT TACACCCTGA GCAAAGAGTT CGGCCTCCGC CTTGATTCCA

8551 GATGGAGCTT TCCTTTATCC CTGATGTATT GGATTGGCTT CCTGCTGGAA

8601 ATAGTGAGCT TCCTACTCAG GCCAATTTAC ACCTATCGAC CGCCCTTCAA

8651 CCGCCACATA GTCACATTGT CAAATAGCGT ATTCACCTTC TCTTATAAGA

8701 AGGCTCAGCG AGATTTGGCG TATAAGCCAC TCTACAGCCG GGAGGAAGCC

8751 AAGCAGAAAA CGCCGGAGTG GGTTGGTTCC CTTGTGGACC GGCACAAGGA

8801 GAACCTGAAG TCCAAGACTC AGTGATTTAA GGATGACAGA GATGTGCATG

8851 TGGGTATTGT TAGGAGATGT CATCAAGCTC CACCCTCCTG GCCTCATACA

8901 GAAAGTGACA AGGGCACAAG CTCAGGTCCT GCTGCCTCCC TTTCATACAA

8951 TGGCCAACTT ATGTTATTCC TCATGTCATC AAAACCTGCG CAGTCATTGG

9001 CCCAACAAGA AGGTTTCTGT CCTAATCATA TACCAGAGGA AAGACCATGT

9051 GGTTTGCTGT TACCAAATCT CAGTAGCTGA TTCTGAACAA TTTAGGGACT

9101 CTTTTAACTT GAGGGTCGTT TTGACTACTA GAGCTCCATT TCTACTCTTA

9151 AATGAGAAAG GATTTCCTTT CTTTTTAATC TTCCATTCCT TCACATAGTT

9201 TGATAAAAAG ATCAATAAAT GTTTGAATGT TTAATGTGGA GGGAGGTGTG

9251 AGTTATGTTA ATACATACTT TCCTCCTTTC TCCTTCTTTT CCAGGTATTG

9301 CCATCTCCAA TTTAGAAGAG TAGCATAAAA CCTGGGTTGG GGAGAGCCCA

9351 GAGTAGTGGG AGGGCCAGGA AGGGCGATGA AGACTGAATA AGCTCTACTA

9401 CCTT

Type II (genbank accession number m77144; exons are boxed):

```
   1  TTAATAAACA TTTAAGCCAA TAATAAAAAT AATGAAAATC ATGAAATAGA
  51  AAAATGAACA AACAATAGAG ACAATCAGTG GTCAATTTTT CTGTACTCCA
 101  TGAACAAAAC TTCAGTTCAG AATGATTTTA TATTTAAATA TATATCCAGA
 151  ATCACAAAGA TCTTAGAAGA AAAAATTAGG AAAATATAGG AATAAAGTGG
 201  GAACTCTGTG GGAATAAAAA TCAGAAAGTA ATTGCCTACC AGGTAAACTC
 251  GGGGGAGGGC AGGCAATTGA CTAAGAACAG GCAGATAAAA CTTTCTGGAG
 301  TGAAGGAAAT GGTGGTCTTC TTTTCGTTGG TACTTACATG GACATATATA
 351  ACTGTCAAGG TTCATTGACA ATGAACACTT AAGTGTACAC TGAACGCTTA
 401  AGATCTGGGC CTTGTAATGC CCAGATTACA TCTACTTAAA AAAACTACAT
 451  CTCTATTTTT TTAAGTAGAG AAAAGTGTCT TGGGGAGTTT ATGGCAGTGA
 501  TGTTTGGGCA GACTGGGGCA TAGGTTTCAG AAGCATTCAT AGCATCTTCA
 551  GTTCAACAAT TTGCTAAGAA TGGTTTTGCA GCCAAAAATG TAAGATGAGT
 601  GCAAATATCT ATGATTAGAG TCATTGGGCA TACAACTACA AATCCCCACA
 651  CTTGTATTAC AGGCCATATG ATTAATTCGT TATCACTAGA TACCTTTCCT
 701  TCAATGAGGT TCTTCCATTA GGAACCCAGA GCTCTCCAGG GAAAAATTGC
 751  AATTGAGGTT TTGGATATAT TGGGTGGAAA AGAAAGTCAC CTCACATAAA
 801  ACTTAGTGAT TGGAGCTGTC ACCATTGAAA ATTTCTAAAC TTTGCAAGAC
 851  AGCAGTGACG TTTCAAATGA CACTCTCAGA GAGTTGATAA TGGGCTAAAA
 901  TAGATCTCCC TCCAGGTGGA TTTACTGTAC AAGGACAACA TTTACATTGC
 951  ACTTGGAGAC TTCTCCCAGT TTGGTTTAAG TTCACAGATT GCAGATCCCA
1001  GACAGCTGGT ATCAACTGAC CAGTGTTCTG TTAAGGCTAA AGCCAAGACT
1051  CTTTATCACA CTGTGGCCTT AAGATTGGAT TTCTCTTCCT GTTCCTGGGA
1101  AGAATTAGAG ATATAACCTA AAGGTCACTA TTATTCTGAG AAAAGGGATT
1151  CTGGAGGAGG AGGGAGCAAT GAGTATGTGG CAGGAGTTCA AGGTAATAAG
1201  GGCTGAGACA CAAGCCACAG AGCATAAAGC TCCAGTCCTT CCTCCAGGGA
1251  TGAGGCAGTA AGGACTTGCA CTCCTCTGTC CAGCTTTTAA CAATCTAAGT
1301  TACGGTTAGA GCTTTCTCCT TTTCTTTCAA CTACTCCTGG CAGTTGTGGG
1351  GTCATGGAAT TTTTGTAAAA AATGGGGTGG AGGAAAATAA GGCATCTGCT
1401  GAGTGTATAA CCATTTTACC TCTTGTTTTT AGCCCTCTTC TGGGTCACGC
1451  TAGAATCAGA TCTGCTCTCC AGCATCTTCT GTTTCCTGGC AAGTGTTTCC
```

1501 TGCTACTTTG GATTGGCCAC GATGGGCTGG AGCTGCCTTG TGACAGGAGC

1551 AGGAGGGCTT CTGGGTCAGA GGATCGTCCG CCTGTTGGTG GAAGAGAAGG

1601 AACTGAAGGA GATCAGGGCC TTCGACAAGG CCTTCAGACC AGAATTGAGA

1651 GAGGAATTTT CTAGTAAGTA AACTTGAGTC ATGGGTCTGT GGCTCCATCT

1701 TAAACTCTGC ATGGGTGTGG GGAGGTTGAC CTTGTCTAGC AAGTTAAGGA

1751 AAGTTGTAGC CAAATGAAAG CCAGTCACAC ATCTAAAGTC ATCAGAAAGG

1801 AAATAGAATA AAATGGTATA GTGTGAAAGA TACTGGATGG GGTGTCCAGA

1851 GACTGGATTC TGGCCCTGAC GCAGAACTTG AGAGGCAGCC ATGTCAGCCT

1901 CAAGGCCCCT TTTCTTCTCT TCTAGAGAGT CACACATGAG TTCTCCTTCT

1951 TGTCTACAAC TCTTATGTTC TGAAGCTTTT TGTCTTGGCG ATTGCTGTGC

2001 GACATTCACA AAGGACATCA TTTACCTGGA GACCTCACCA GTGGGTCCTG

2051 CCTGTCTAGA CTGCCCCAGG CTTCATTTCA TTTTATAGCT TCCTTAGTAT

2101 CTTTAGTGAG GGCAAAATAA AAGGAAGTGA AGGTACGGAT GTGTGTGTAA

2151 TTGTGTGTGT GTGTCAGACA GAGAAAGAAA ATGAGGGGGA AAGAAGGTCA

2201 GAGACAGAGA ATGAGAAAGA CAGAGAATGA ATGAGAAAGA TAGTGTGAGG

2251 GAGAGCATCA GCAAGAGAGA GAGAGTGAAA GTGAGAGAGA TCATCAGTGA

2301 GAGAGAGAGA AAGAGAGAAA GAGAGCATGC ACAAGTGCAC AGCACAGAGC

2351 AGAGACAGAG TAAAAGCCAC TATGAGGCCA ACATGCCTCA TTTATATTTC

2401 TCATATACAG CCTTTTTAAA AAAAAACCAT TTGTATCTGT TCTTCAATAC

2451 TCCAAATAAA TTGACTGCAG AATTTGCTGT AGTAACTCTT TGTGCAAATA

2501 ACCCTGTGAG GAAGGTAGAA AAGGGGTCAC CTCCAACTGG ACAATGAACC

2551 CATCAAGTTC CAGAAACATA AAGTTACTTG CCAAGAACAG CTGATTAATG

2601 TGTAGCAGAG CTCAGACCAG AGCTTCATTC CAGAAGCTTC ATCACTTGAC

2651 CCCCGGGCTA ATCCTTGGTG GGTCTTCAGT CTCTCCCTTT GCCTCTCTGT

2701 GTGCCCCAAA CCAATTCCAT GTTTTCCTCC CAACCCAGGG CACTGGGGGA

2751 ACTATGTATC TTCTTCATGC AGGTTCTGGC TGCCTAGAAG GTCCTGCCCT

2801 GTTCACACAG AAGCTCCCCT TGACACTCTC CTCTTCTTTC ACTCTTCTCT

2851 GACCTCTTTA AATGTCCAAT TTCCTCCACT CTCTGTTCTA ATAGTGAACA

2901 GATATTAACA CATGACTGAT ATTAAATGTT TGTAATCCTT ATGTCTCAGC

2951 AATATTTACA ATCTAACATC TGATACCAGG AAAAAGCAAT GCTCAGTCTA

3001 GAGGAAGAGT TCTGCTATGG TGAAAATGCA GGGAGCATAA TGGATGGGCA

3051 GGGCCTGTAT GAAGATAAAA ACAGAGAATC CTTTCAGTGT TCTTTCTGTG

3101 CCAGGCACTG TGCTAAAGGT TTTGCATGTG CTGACTCATT TACTTCCCAT

3151 AACAACCCTA TGAGATGGGT AATACTATTA TCCTCCTTTT TTATAGATGA

3201 GGAGACAGAC ACAGAGAGGT TAATCAACTT GTCGAAGGTC CCAAGTTAGA

3251 TTACGGCAAA GCTGGATTTG AATCCAGGCA GTCTAGTGCA GCTTCTGTGC

3301 TCTTAAGTAG TAGTTTGTCT AGCAGCTTCT CTGAGCCTCC CTCCAGCCCA

3351 ACACACAACA CAAAAGCAAC TGAATTTAGT CTACATGAAA CACTTGCCAA

3401 GTCCACAGTG CCTTGGAGAA GGAGGGAGCT AAGACTCCAT CAAAAATTGC

3451 GCTTTTCTGG TCTCTCTAGA TCCATGGAAA CTCACCGGGG AAGAGTCATT

3501 TTCTATCGGC TTTGGACTTC TGCTATGAAA TCCCAATTCC TTTGAGCCAG

3551 TCTGTCTAAA GTCGGCAGCA AAGGAGAAGC AGATGAAGGG AGTAGAAAAG

3601 GTAGAGGTGG AAATACAACC AGAGAAAATT CCACTCTCGG GGCTGACACC

3651 CACCTCCTCC CCAGCAGCCT CTTCCCTGCC TCCACCATTC CAGGGCCTTC

3701 TTGGAGTTGC CTGACAGTGG GAAGGGGCTG GGAGTAGACA AAACAAGGCA

3751 GAGGCATCAG GAGAGCAAAT GGTCCTGCCT GTAACCCAGG AAGGAATTTC

3801 TCCAGCCTTG AATTTTTCAT TCTGAATTTC GCTCTCACTG TTCTTTGGCC

3851 TTTGTGGGCA TTTTCAGTTA AACTGGCTTA TTCTTACTCT TTCTCAGAAA

3901 AGTTGATTCT AAACAGCTAT TCCTGCACCT GGGGTACAGG GTACAAAAAT

3951 GGCAGGTGGA ATGGGCAGGT TTGAGCCTTC ACTAGAAAAT TGACTTTAGA

4001 TCTGTCCCTT AACCTCTCTG AGCCTCAGTT TCCTCACCAA TAAACTCCGA

4051 CTCAGAATAC CTACCTCACG AGATAGCTGT AAAAGGAAAT AAGCAAGCTG

4101 TGAGCCGTAT GCAGATTATA AAGCACAAAT GAGTGTAGAG TATTTATCTG

4151 AGCTCCATGT AGCATGGTGT TCATTTTGTG GGCTAATTCC AGGTGACAAC

4201 TTTACCGGCA GCGTCACTAC CTGTGGATTT CAGAGGAGCC TGAAGATATG

4251 GATGTTACAT CCGCTGCCCT TGGCTAGACA TTAGACAGCT GAGATGTTAG

4301 TTCATTGGTG GCTAACTCAT TTTCAGCATT GCTACCTGCT TTTTATGATA

4351 TCTGATAAAA GTTTCTCAAT GAAATTAATC ATTTAAGCCC ATCTGCAATT

4401 CCTGAGTAAT GGCTCCACTC TCCTCCCAGT TGCCCAAGCA GGAAACTGTA

4451 AGGGTGATTT TGTCTCCTTT TCCTCTCATC CCAGCTCCAG TAGGTACCAG

4501 GACCTTCAAG GTGAGGTCTG AAATTCAAAC CTGGAGTGCA TCTCTCCCCT

4551 CCTCTGCTTC CACTGCTGGA CCCTGGTTCC AGTCCTCATC TAAACTTTGG

4601 CACAGACATT CTGATTAATC TCTTGATCAC CTGGCTTTCT CATTTCCCCT

4651 ATACCAGTCT GTTTTATATC CTCAAGCCAA GGTAAACTTC CTACGCCACA

4701 ACTCTGCGTC TGTTTATCTT TTATTATTAT TACTATTATT ATTATTATTA

4751 TTATTATTAT TATTTTGAGA CAGGGTCTTT CTCTGCCACT GAGGCTAGCA

4801 TGCAGTGGCA TGATCTCAAT TCACTGCAAC CTCCACCTCC CAGACTCCAT

4851 CCATCCTCCT GCTTCAGGTT TGTGATTAGC TGGGACTATA GGCACATACC

4901 ATCACCCCCA GCTAATTTTT GTAATTTTAG TAAAGACAGG GTTTTGCCAT

4951 GTTTTCCAGC CAGACCTCCA CCTCCACAAA TCAAGTGATT CACCACCTCA

5001 GAGTCCCAAA GAGCTGGAAT TACAGGAATG AGCCACTGCA CCCGGCCTGT

5051 TTCTCTACTT ATTCCAACAT AAATAAGGGT TTTTCTTTGA CATGGTGCTG

5101 GAGACCCAAA ATCATAATTC TATACATGAT TCCTCATCGC CTACAGAATT

5151 AAATTCAAAC TTTCCCTCTT ATACCCAACT CTCCACAATC ACTCTATTCT

5201 CCATAAAAAA TTTGGTGAGT CTCCCATGAC TCTAGCAATG CTTATATTTC

5251 ACGGATGTGT GACAATTCAC TGCTCACATT ACTTTTCTGG AATAAGATGG

5301 GGTAGAAACA GATGTTTGCT CTTTCCCAAT AAACTGCTAC ACATGCTGAT

5351 TTCTGTGCTT TTGTTTACTT GTTCCTTTTA TGGAATGTAG TACACCCTCC

5401 ACTCTAATAC CCACACTCTA ATCTCTTTGA GCACCTATGT AACATCGCCT

5451 TTATCAGAAA ACTTCCCAGC CAGATCCAGA AATCTTTCCA ATGACCTGAC

5501 CTGTGTTCAC ACAGAGCTCC AGAACAGGAC CAAGCTGACT GTACTTGAAG

5551 GAGACATTCT GGATGAGCCA TTCCTGAAAA GAGCCTGCCA GGACGTCTCG

5601 GTCGTCATCC ACACCGCCTG TATCATTGAT GTCTTTGGTG TCACTCACAG

5651 AGAGTCCATC ATGAATGTCA ATGTGAAAGG TACAGTAGCC TGGGGAGGAG

5701 ATAAAACAAG TTGGTTAAAT GAGGATCAGA AAGAAGGACA AGAAAGGGAA

5751 GAGAAGTCAC TCCATTGAAC ACCTGCTGAG CTCTTGGCCA AGTGCCTCTG

5801 CTGATCACTA CTGACTGGGG AGTTCAAGGC TGGTAACCTC AGTTTTTTAG

5851 ATGAGAAAAC TAGGGCTGAG AGACAGCAAG TAACTTGTCC AAGGCCCCAA

5901 AAGTAAGTAA GTAAGTAGGA GAGTTAGACT TTAAACTCAC TCCTGTGTGA

5951 CTCCAAAGGC TGTGGAAGCT CTTTCTACTG TGGCCCCAAT CAAAAGTCAA

6001 CTAAACTCCA ACTTCAAATT CACGATATCC AGCTACTCCT TGGCTCCCCG

6051 GGCCAGAATC AGACCTTCCA GGTGCCACCA TAGTCATCAT TTTGAACCTT

6101 GTGTGTAGGC TGATGAGAAC ATTCACACTC TTCCTGCCCA CCTCAAAGAA

6151 GTCCTCTCAA GAGAACTAGC AAAGCTGGTT CACAGAGGTC TGTCAGGACA

6201 GAATTATCCA GCACATCCCT TCCCATAATA TTTTCTTAAC AAAAAGAGTT

6251 TCCCAGTGTC CAGAATACAA TCTCTTCAGC TCACCACAGG GTCTACTATT

6301 ACAGAGCCAT TTCCTGCCAG GTGCCCAAAG TGCACCCTCA TTTGATCCAC

6351 ATTTTTTAAA GGAATCTTGC CATGCAGTCT CACCCTCTGA TTCCCAGAGC

6401 CCAGTCTCTT CAAAACACAG CTCCCCACTG CTGGCCTGAC CCTCCACATT

6451 CAGACTACCC AAACTCCTTT TATAAAGCAA CTAATTGCTT TTGAGTCTCC

6501 AGCACTGTCT AAAAAAACAA ACAAAAAAAA AACCTTTCCA GTTTCTACAT

6551 AGCCATACCC CTAAGTACCG CTATAGTTCA AGCAGTGTGC TAAGAGTGAA

6601 TAGTTTTTTG TTTTAATTTT GACTTTCCAC AGCTCCCAAA AACAAGGGAT

6651 AAGAAATGAC AAGGGGAGTG TTAGAGGCGT CTGTTTCTTA AAACCATTAT

6701 CACTCCCGGG CTTAGAAAAT ACTTCACAAT GATGGCCATT TCATGAGGCT

6751 GTCTCTCCAG AAACTCAAAT TGCACAGAGA CTTTAAAAGT GGCTAAGAAT

6801 ACAAAATAAG TAAATAAAAA TGATCCTTTG CCTAGGAAAT TCCTTCCTTC

6851 TTCAAGGGAA GACACTCCAG ATACCTCCAT GAGCTTTGCT TCCACACAGC

6901 TCAGGTGATG CTGACAGGTC TTCCCAGTGC AGGTGTTATC AGCACACGGC

6951 ACACTCCCTC CCCAGTCCTC CACCAGGCTC CACAGGAAAT GCCAGGGCAG

7001 GATTTAAAGG AAGGTTTATC ACCCCCACTT TATATAACTG GGTAACAGAG

7051 TCATAGCCAG AATTGGAAGC AGCTTTTCCA GGGAACACAC AATCAGGAAA

7101 GGAGTGTGGG TTTCCAGCAT CTCCCCACAA CCCACTGTTT GCTCGATGAC

7151 TCCAGGGATG GTCTTTGTGA CAAAGTTTTA TTAATCCATT GCAAAATGAG

7201 AAAAATACCA GCAATTCTAT GTATATCTAA GAGTGTATAC TGCCTATATT

7251 CAACTGTGAA TTTGTGTGTG TAACACAAAC TCTTCCACTG CCCTTTCTTG

7301 CCTAAGGACT GATTGGGACT GTTCATATTC TCAGAGTATA GCCTCCTGAT

7351 TTTTTAGATT TTAAATGATC TGTCTTTAAT TGTATTGATG ACAAATCATA

7401 CTTTTTCAAT ATCCTTACTT GGCAAAAAAT AAAGTGATTA CCCTAGGTCC

7451 TTTTCCAAAA TTAAAGTGTG TGTGTATATA TATATATGTA TATATATATG

7501 TATATATATA TATGTATATA TATGTATATA TATATATGTA TATATATATG

7551 TATATATATA TCTATATATA TATGTATATA TATATATTTT ATAATATATA

7601 TATATATATA CACACACGTA TACATACACA CACACACACA CACACCACAC

7651 ACAAACATAG GTCCATGAAA CCCAAAAGTC TAGAAATTCT TGCTTTCTAG

7701 AACCCAAATG TCAGTTTCTT TTTATTTTTG TGTTTTTCCT ACGGCTGTAT

7751 CATGACCCAA TCTCAGTCAG AGCCACAGAA GAATGCACCC TGAGTCTGTT

7801 ATAACCACTG CACTTGGCAG TGGGGAGTGG GGCACATGGA TCTGTGCATG

7851 TGGTTGCAGC TCCTTTGGGA TATTTCCTGA CACTGTCATC ATGCTCTTCG

7901 TGGGCAGGTA CCCAGCTACT GTTGGAGGCC TGTGTCCAAG CCAGTGTGCC

7951 AGTCTTCATC TACACCAGTA GCATAGAGGT AGCCGGGCCC AACTCCTACA

8001 AGGAAATCAT CCAGAACGGC CACGAAGAAG AGCCTCTGGA AAACACATGG

8051 CCCACTCCAT ACCCGTACAG CAAAAAGCTT GCTGAGAAGG CTGTGCTGGC

8101 GGCTAATGGG TGGAATCTAA AAAATGGTGA TACCTTGTAC ACTTGTGCGT

8151 TAAGACCCAC ATATATCTAT GGGGAAGGAG GCCCATTCCT TTCTGCCAGT

8201 ATAAATGAGG CCCTGAACAA CAATGGGATC CTGTCAAGTG TTGGAAAGTT

8251 CTCTACAGTC AACCCAGTCT ATGTTCGCAA CGTGGCCTGG GCCCACATTC

8301 TGGCCTTGAG GGCTCTCCGG GACCCCAAGA AGGCCCCAAG TGTCCGAGGT

8351 CAATTCTATT ACATCTCAGA TGACACGCCT CACCAAAGCT ATGATAACCT

8401 TAATTACATC CTGAGCAAAG AGTTTGGCCT CCGCCTTGAT TCCAGATGGA

8451 GCCTTCCTTT AACCCTGATG TACTGGATTG GCTTCCTGCT GGAAGTAGTG

8501 AGCTTCCTAC TCAGCCCAAT TTACTCCTAT CAACCCCCCT TCAACCGCCA

8551 CACAGTCACA TTATCAAATA GTGTGCTCAC CTTCTCCTAC AAGAAGGCTC

8601 AGCGAGATCT GGCGTATAAG CCACTCTACA GCTGGGAGGA AGCCAAGCAG

8651 AAAACCGTGG AGTGGGTTGG TTCCCTTGTG GACCGGCACA AGCAGACCCT

8701 CAAGTCCAAG ACTCAGTGAT TTAAGGATGA CAGAGATGTG CACGTGGGTA

8751 TTGTTAGGAA ATGTCATCAA ACTCCACCCA CCTGGCTTCA TACAGAAGGC

8801 AACAGGGGCA CAAGCCCAGG TCCTGCTGCC TCTCTTTCAC ACAATGCCCA

8851 ACTTACTGTC TTCTTCATGT CATCAAAATC TGCACAGTCA CTGGCCCAAC

8901 CAGAACTTTC TGTCCTAATC ATACACCAGA AGACAAACAA TATGATTTGC

8951 TGTTACCAAA TCTCAGTGGC TGATTCTGAA CAATTGTGGT CTCTCTTAAC

9001 TTGAGGTTCT CTTTTGACTA ATAGAGCTCC ATTTCCCCTC TTAAATGAGA

9051 AAGCATTTCT TTTCTCTTTA ATCTCCTATT CCTTCACACA GTTCAACATA

9101 AAGAGCAATA AATGTTTTAA TGCTTAA

Appendix ii

Sequences obtained from clones isolated from two human genomic libraries, aligned with homologous sequences from 3ß-HSD type I and type II (exons 1-4). Triplet nucleotides underlined are translational start and stop codons. 3ß-HSD type II missense and nonsense mutations are shown above the aligned sequences in brackets. Corresponding single nucleotides in the clones HSD3Bψ4, ψ5, ψ1 and ψ3 are highlighted in bold and underlined. Dashes indicate gaps introduced for alignment. Bases corresponding to intron sequence in 3ß-HSD types I and II are shown in lower case.

a) Exons 1 and 2

```
          1                                                          50
          *                                                           *
  Type2   GAGGCAGTAA GGACTTGGAC TCCTCTGTCC AGCTTTT-AA CAATCTAAGT
  Type1   ....TGAG.. .T..G.CC.. ..T..... .......-.. ........C.
HSD3Bψ4   ..A.TGACC. .C...CT... ..T...A... .......T.. T.......CA
HSD3Bψ5   ..ATGGAC.. .C..CCT... ..T.T..... ......--.. ........CA
HSD3Bψ2   ....TGACC. .C..CCT... .A........ ....A...-. T.......C.
HSD3Bψ1   ..A.TG.CC. .C..-CT... ..T....... ...C...-.. .C..T...T.
HSD3Bψ3   NNNNNNNNNN NNNNNNNNNN NNNNNNNNN. .C.....-.. .G......CA


          51                                                        100
          *                                                           *
  Type2   TACGgttaga gctttctcct tttctttcaa ctact---cc tggcagttgt
  Type1   A.T..... .a...t..a. .........g .....---.. ..-....g..
HSD3Bψ4   ----...g.. .t..ct.ta. .......t.g .....---t. .....a.g..
HSD3Bψ5   ----..a... ....ct..a. .........g .....---.. .....a.g..
HSD3Bψ2   A.T..... ...a.t..g. .........g .....---t. .....a.g..
HSD3Bψ1   A.T...g.a. .....t..a. .........g g.....---.t .......g..
HSD3Bψ3   A.T..c.... .....t..a. ...t...ttt .a...acca. .......g..


          101                                                       150
          *                                                           *
  Type2   ggggtcatgg aatttttgta aaaaa---tg gggtggagga aaataaggca
  Type1   ....a..ca. ...g....c. .....aa-.. .......... ....g.....
HSD3Bψ4   .aa.a..... ........cc .....aaa.. .......... ......a...
HSD3Bψ5   ..a.a..... ........c. .....aa-.. ....a..... .......a..
HSD3Bψ2   .a..at.... ........c. .....aaa.. .......a.t ...-..a...
HSD3Bψ1   ...-a..... ........c. ..c..---.. t......... ..........
HSD3Bψ3   ....a..ca. ........c. .....a--ga .......... ........g.


          151                                                       200
          *                                                           *
  Type2   tctg-ctgag tgtataacca ttttacctct tgtttttagC CCTCTTCTGG
  Type1   ....tg.... .a........ ...g..a..- .c........ .....C.A..
HSD3Bψ4   .t..tg.... .a........ ...g.t...- -.....c... .....C.A..
HSD3Bψ5   ....tg.... .a........ .-.g.t...- .t........ .-...C.A..
HSD3Bψ2   ....tg.... .......at. .c.g.t...- cc........ .-...C.A..
HSD3Bψ1   ....tgca.. .c........ c........- .........T ...G.C.AT.
HSD3Bψ3   ....tg.... .a........ ...g...a.- .c..c..... .....C....
```

```
          201                                                       250
          *                                                         *
  Type2   GTCACGCTAG AATCAGATCT GCTCTCCAGC ATCTTCTGTT TCCTGGCAAG
  Type1   .....C.... .......... ....C..... .......... ......TG..
HSD3Bψ4   ....AT.... .......... ....C..... .......... C.....G...
HSD3Bψ5   .....C.... .......... ....C..... .......... ......AGT.
HSD3Bψ2   .A...C.... ...A...... T...C..... .......... ......AG..
HSD3Bψ1   .T...C.... ...A...... ....C....G .......A.. .......G..
HSD3Bψ3   .....CT... .......... .T..C..... ........C. C......G..

          251                                                       300
          *                                                         *
  Type2   TGGTTTCCTG CTACTTTGGA TTGGCCACGA TGGGCTGGAG CTGCCTTCTG
  Type1   ...A...... .......... -.....ATG. C. ....... ..........
HSD3Bψ4   .......... .A.TA..... .....AATG. CA........ ..........
HSD3Bψ5   .....GT... .......... ......ATG. CA........ ..........
HSD3Bψ2   .......... ....C.C... ......ATG. CA........ ..........
HSD3Bψ1   .......... .A.TC..... -.....ATG. .......... ..........
HSD3Bψ3   C......... .......... ..A...ATG. .........A ..T.......

          301                                                       350
          *                         (A)                             *
  Type2   ACAGGAGCAG GAGGGCTTCT GGGTCAGAGG ATCGTCCGCC TGTTGGTGGA
  Type1   .......... .....T.... ...A...... ...A...... .C........
HSD3Bψ4   .......... .....T.... ....TAG... ...A...... .C........
HSD3Bψ5   .......... .....T.... ....TAG..T ...A...... .C......A.
HSD3Bψ2   .......... .....T.... .......... ...A..T... ....A.....
HSD3Bψ1   .......... .....T...C .......... ..T...AA.. .C......A.
HSD3Bψ3   .......... ..AN.T.... ..A......A ...A...AT. .C......A.

          351                                                       400
          *                                                         *
  Type2   AGAGAAGGAA CTGAAGGAGA TCAGGGCCTT GGACAAGGCC TTCAGACCAGA
  Type1   G........G .......... ......T... .......... ...G.......
HSD3Bψ4   G........G .......... .......... .......... .........A.
HSD3Bψ5   G........G .......... .......... .......... ........G..
HSD3Bψ2   G....CA..G .......... ..CC...... ....-..... ..........G
HSD3Bψ1   G........G ..A....... .......... .......... ......T...G
HSD3Bψ3   G........G .......... .......T.. ...A...... ...G.......

          401                                 430
          *                                   *
  Type2   AATTGAGAGA GGAATTTTCT Agtaagtaaa
  Type1   .......... .......... ..........
HSD3Bψ4   .GC....G.. .......... G.........
HSD3Bψ5   .......GA. .......... G.........
HSD3Bψ2   G..TGA.G.. .......... ..........
HSD3Bψ1   G......... .A........ ..........
HSD3Bψ3   ...TNNNNNN NNNNNNNNNN Nnnnnnnnnn
```

b) Exon3

```
          1                                                  50
          *                                                   *
  Type2   ccaatgacct gacctgtgtt cacacagAGC TCCAGAACAG GACCAAGCTG
  Type1   .......... .......... ........A. .........A ..........
HSD3Bψ4   .......... .......... nnnnnn.... .........A ..........
HSD3Bψ5   .......... .......... nnnnnn.... .........A ..........
HSD3Bψ2   .......... a....cn.c. ........A. .........A. .T.......
HSD3Bψ1   ..-....... .......... .ttg...... .........A .......A..
HSD3Bψ3   .......... .......... nnnnnnnnNNN NNNNNNNNNN NNNNNNNNNN

          51                                                 100
          *                                                   *
  Type2   ACTGTACTTG AAGGAGACAT TCTGGATGAG CCATTCCTGA AAAGAGCCTC
  Type1   ..A..G..G. .......... .......... .......... .G........
HSD3Bψ4   ..A..G..G. .......... .......... .......... .G........
HSD3Bψ5   ..G..G..G. .......... .......... T...G..... .G........
HSD3Bψ2   ..A..G..G. .......... .......... .......... .G........
HSD3Bψ1   ..A..G..A. .......... .......... T...G..... .G........
HSD3Bψ3   NNNNTGT.G. .......... ....A..... T...G...T. ..........

          101                                                150
          *                              (A)                  *
  Type2   CCAGGACGTC CCGGTCGTCA TCCACACCGC CTGTATCATT GATGTCTTTG
  Type1   .......... ......A... .......... .......... ........C.
HSD3Bψ4   .......A.G ......A... .......T.. ..C......A T.....A.C.
HSD3Bψ5   .......A.G ......A... ........A. ..CC.....A ..CA..A.C.
HSD3Bψ2   .........G .......... .......... .......... .........C.
HSD3Bψ1   .......A.. .....A.... .......... ..C....... ..CA....C.
HSD3Bψ3   .......A.. .T....A... .......T.. ..C....... ..C.....C.

          151                                                200
          *                                   (G)              *
  Type2   GTGTCACTCA CAGAGAGTCC ATCATGAATG TCAATGTGAA AGgtacagta
  Type1   .......... .........T .......... .......... .....tg...
HSD3Bψ4   .......... ...--..... .......... .......... ..........
HSD3Bψ5   .......... .......... ........CA .......... .C........
HSD3Bψ2   .A........ ....C....T .......... .......... .....gg...
HSD3Bψ1   .......... .........T ........CT ....C..... ..........
HSD3Bψ3   .......... ....C....T ........A. C......... ..........

          201                          230
          *                             *
  Type2   gcctggggag gagataaaac aagttggtt-
  Type1   .g........ .....gc.g. ...g...gg-
HSD3Bψ4   a......... ......a.g. .......gg-
HSD3Bψ5   .......... .....g..g. ...g...gg-
HSD3Bψ2   ......nnnn nnnnnnnnnn nnnnnnnnnn
HSD3Bψ1   .....a.... .....gg.g. ...g...agc
HSD3Bψ3   t......... .....gc.g. ...g...g-c
```

c) Exon 4

```
          1                                                                50
          *      (a)                          (G)                          *
Type2     tcttcgtggg cagGTACCCA GCTACTGTTG GAGGCCTGTG TCCAAGCCAG
Type1     .....a...a .......... ...C.....A .......... .......T..
HSD3Bψ4   .t.....a.. t.......... ...T...... .......... .......T.C
HSD3Bψ2   ....tc.... .g........ ...C...... .......... .......T.C
HSD3Bψ1   .....a.... .......... A.AG...... .......... .......T..
HSD3Bψ3   -.......... .......... A..T...... ..-A...... .......T..

          51                                                               100
          *                                                          (A)   *
Type2     TGTGCCAGTC TTCATCTACA CCAGTAGCAT AGAGGTAGCC GGGCCCAACT
Type1     .......... .......... .......... .......... ..........
HSD3Bψ4   A.......A.. .......... .......C.C. .......... A.......T.
HSD3Bψ2   .......... .......... .......C.C. CC..T..... ..........
HSD3Bψ1   .....T.... ......C.T. ....C..... .C........ T.........
HSD3Bψ3   .A.-....G. ......C... ....C.T... ...A...T.. A.........

          101                                                              150
          *                                   (A)                          *
Type2     CCTACAAGGA AATCATCCAG AACGGCCACG AAGAAGAGCC TCTGGAAAAC
Type1     .......... .......... ..T.....T. .......... ..........
HSD3Bψ4   .......... ......T... .....T..T. .......... ..........
HSD3Bψ2   .......... ......TG.. ..T.C...T. .......CTT ..........
HSD3Bψ1   .......A.. G..TT..... ..TC.A...A ..C......A ..........
HSD3Bψ3   ...G...TGA G......... ..T....... ..T....... .T.......

          151                                                              200
          *                                                                *
Type2     ACATGGCCCA CTCCATACCC GTACAGCAAA AAGCTTGCTG AGAAGGCTGT
Type1     .........C .......... AC........ .......... ..........
HSD3Bψ4   ......TA.G .......... AC........ .......... ..........
HSD3Bψ2   ......T.TG .......T.. A......... ...T...... ..........
HSD3Bψ1   ......T.TG .......T.. A......... .......... ..........
HSD3Bψ3   .T......TG .......... A..G...... .......... .......C..

          201                                                              250
          *                    (A)     (G)                                 *
Type2     GCTGGCGGCT AATGGGTGGA ATCTAAAAAA TGGTGATACC TTCTACACTT
Type1     A......... ..C....... ....G..... C..C.GC... C.........
HSD3Bψ4   .......... ......CT.. C...G..... C....GC... ..........
HSD3Bψ2   ......A... .......... C...G..... .......... ..........
HSD3Bψ1   .....T.... ....T..... C..TGA.... ..C..GC.TG ..........
HSD3Bψ3   ......A... ........A. CA..G..... .A.C.-..C. .......T..
```

```
          251                                                        300
          *            (T)                                            *
  Type2   GTGCGTTAAG ACCCACATAT ATCTATGGGG AAGGAGGCCC ATTCCTTTCT
  Type1   ....C...C. .....TG... .......... .....A.... .A........
HSD3Bψ4   ....CG.... ...A.TG.T. .......... .....A.... ..........
HSD3Bψ2   ....C..... .....TG... .......... .....A.... ...T......
HSD3Bψ1   ....C..... CT.A.TGC.. .......... .....A.... ...T...A..
HSD3Bψ3   ....C..... CT...TG... .......A.. .....A.... TA.A......

          301                                                        350
          *                    (C)                          (G)      *
  Type2   GCCAGTATAA ATGAGGCCCT GAACAACAAT GGGATCCTGT CAAGTGTTGG
  Type1   ..T....... .C........ .......... .......... ..........
HSD3Bψ4   ...G...... .......... A......C.. ..A....... .....T.CA.
HSD3Bψ2   ....A..... .......... .......... .......... .G.....CA.
HSD3Bψ1   .A..A..... ..A.....T. ...A...... .C........ ...C......
HSD3Bψ3   ....C..C.. .......... ATT ------ .......... .......CA.

          351                                                        400
          *G)                                                        *
  Type2   AAAGTTCTCT ACAGTCAACC CAGTCTATGT TGGCAACGTG GCCTGGGCCC
  Type1   .........C ..T..T.... .......... ......T... ..........
HSD3Bψ4   C........C .G........ .......... .......A.A ..........
HSD3Bψ2   C........C ....C..... .......... .......... ..........
HSD3Bψ1   .........C ...AC..... .......... G..------- --........
HSD3Bψ3   C.....A..C ...-.....A .......... ......T..- ..........

          401                                                        450
          *                                                 (C)      (AA)
  Type2   ACATTCTGGC CTTGAGGGCT CTGCGGGACC CCAAGAAGGC CCCAAGTGTC
  Type1   .......... .........C ....A..... .......... ......CA..
HSD3Bψ4   .......... .........C ........G. .......... ..........
HSD3Bψ2   .......... .......... -......... .......... ..........
HSD3Bψ1   .......... .....-...C ...T...G.. .........T ......CA..
HSD3Bψ3   .........T ....GT.T.C ....A..... .......... ..T...CA..

          451                                                        500
         (TAG)          (A) (G)                  (T)                *
  Type2   CGAGGTCAAT TCTATTACAT CTCAGATGAC ACGCCTCACC AAAGCTATGA
  Type1   .....A..G. ....C..T.. .......... .......... ..........
HSD3Bψ4   .....A..G. ....C..T.. ...C...... .......... ..........
HSD3Bψ2   .T...A..G. ....C..T.. .......... .......... .....T....
HSD3Bψ1   T....A..G. ....C..... .......... .......... ..........
HSD3Bψ3   .....A..G. ....C..T.. .......... .T.T..T... ..........
```

```
          501                                                       550
          *                                                         *
Type2     TAACCTTAAT TACATCCTGA GCAAAGAGTT TGGCCTCCGC CTTGATTCCA
Type1     .......... ....C..... .......... .......... ..........
HSD3Bψ4   ........G. ....C.T... .......... .......T.. ..........
HSD3Bψ2   .......... ......C... .......... C....C.T.. ..........
HSD3Bψ1   .......... ...GC..... .......... C......T.. ..........
HSD3Bψ3   CA......G. ....C..... A......A.. .-.....T.. ..........

          551                                                       600
          *                                                         *
Type2     GATGGAGCCT TCCTTTAACC CTCATGTACT GCATTGGCTT CCTGCTGGAA
Type1     ........T. .......T.. ........T. .......... ..........
HSD3Bψ4   .T........ G......T.. ....C..... .......... ..........
HSD3Bψ2   ....A..... ..T....TT. .......... .......... ........G.
HSD3Bψ1   .......G.. .......G.. ....G..... .......... ..........
HSD3Bψ3   .......... ....C..G.. ....G..... .......... ..........

          601                                                       650
          *                                                         *
Type2     GTAGTGAGCT TCCTACTCAG CCCAATTTAC TCCTATCAAC CCCCCTTCAA
Type1     A......... .......... G......... A......G.. .G........
HSD3Bψ4   A......... ....G.CG.. G...G..... A...G..G.. .G........
HSD3Bψ2   A........C ....G..... G...G...G. A......... .G......G.
HSD3Bψ1   A......... ....G..... GT.......T A.....TG.. .T........
HSD3Bψ3   A.C.....A. ....G..... NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

          651                                                       700
          *                                                         *
Type2     CCGCCACACA GTCACATTAT CAAATAGTGT GTTCACCTTC TCTTACAAGA
Type1     ........T. ........G. .......C.. A......... .....T....
HSD3Bψ4   ..A.....G. ..G.....G. .......C.. .......... ..........
HSD3Bψ2   .T........ ..G.....G. .......CA. .......... ..........
HSD3Bψ1   .T........ ..G.....G. .C..C..CA. .G........ ..........
HSD3Bψ3   NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

          701                                                       750
          *                                                         *
Type2     AGGCTCAGCG AGATCTGGCG TATAAGCCAC TCTACAGCTG GGAGGAAGCC
Type1     .......... ....T..... .......... .......... ..........
HSD3Bψ4   .........A .........A ......C... .T........ ..........
HSD3Bψ2   .......... .........A .......... .T........ ..........
HSD3Bψ1   .......... .......... .........A .T........ ..........
HSD3Bψ3   NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
```

```
          751                                                       800
          *                                                         *
Type2     AAGCAGAAAA CCGTGGAGTG GGTTGGTTCC CTTGTGGACC GGCACAAGGA
Type1     .......... .G........ .......... .......... ..........
HSD3Bψ4   .......... ..A..... .......... .........T ..........
HSD3Bψ2   .......... ..A..... A......... .......... ..........
HSD3Bψ1   C......... ..A..... .......... .........T ..A.......
HSD3Bψ3   NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

          801                                                       850
          *                                                         *
Type2     GACCCTGAAG TCCAAGACTC AGTGATTTAA GGATGACAGA GATGTGCATG
Type1     ..A....... .......... ..TGA..... .......... ..........
HSD3Bψ4   .......... .......... .......CG.. .......... ..........
HSD3Bψ2   ..A....... .......... ....C..... .NNNNNNNNN NNNNNNNNNN
HSD3Bψ1   ......A... .......... .......... .......... ..........
HSD3Bψ3   NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

          851                                                       900
          *                                                         *
Type2     TGGGTATTGT TAGGAAATGT CATCAAACTC CACCCACCTG GCTTCATACA
Type1     .......... .....G.... ......G... .....T.... ..C.......
HSD3Bψ4   .......... .....G.... .C....G... ..T..T.... ..........-
HSD3Bψ2   NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
HSD3Bψ1   .......... ....CG.... ......G... ..T..T.... .TC.......
HSD3Bψ3   NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
```

Appendix iii

Hybrid name:	Chromosomes contained within:	Mouse or rat:
HA11	4, 5, 6, 8, 11, 20, 21	mouse
HA221	1, 4, 6, 7, 11, 15, 21	mouse
HA232	3, 4, 6, 11, 14, 15, 16, 21, 22	mouse
HB29	3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22	mouse
HB33	6, 8, 14, 22	mouse
HB35	3, 4, 8, 14, 17	mouse
HB111	3, 11, 12, 14, 17, 21	mouse
HB142-2	4, 14, 15, 17, 20	mouse
HB181	2, 3, 6, 7, 8, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22	mouse
JV01	3, 5, 7, 10, 11, 13, 14, 17, Y	rat
HR40C8	X, 7, 10, 12, 13, 14, 15, 17, 18, 19, 21	rat