

THE HUMAN CYTOCHROME P-450 21-HYDROXYLASE GENES



N. R. Rodrigues
Trinity College, Oxford

A thesis submitted for the
Degree of Doctor of Philosophy
October 1987

MRC Immunochemistry Unit
Department of Biochemistry,
University of Oxford

ACKNOWLEDGEMENTS

I would like to thank all the members of the MRC Immunochemistry Unit for their helpful discussions and friendship. I remember the late Prof. Rodney R. Porter with great regret and gratitude, gratitude for accepting me as his student and inspiring me to pursue the path of knowledge, regret for the loss we suffered due to his untimely demise. I am especially grateful to Dr. Duncan Campbell for accepting me as his student for the last two years and for the support and guidance he has given me. I would like to thank Yung Yu for all his help during the last three years.

I would like to thank the Rhodes Trust for the Scholarship and especially The Fletchers, for their understanding and warmth and for treating us Rhodes Scholars as part of their family.

I am grateful to Frank Caddick and Ken Johnson for their help in the preparation of the figures for this thesis, Carolyn Brooks for her patience in teaching me word processing, Animesh Shrivastava for proof reading and Robert Rodrigues for the typing of the thesis.

I would like to thank my parents for their encouragement and love which has the wings to fly across the oceans, and my aunt who has given me a home to run to in England. Finally, I would like to thank my husband Robert for all his love and understanding. But for him the dreaming spires of Oxford would have remained but a dream.

Abbreviations

The nomenclature of complement components is that recommended by the World Health Organization (1968, 1981). Activated components are indicated by bar, e.g. C $\bar{1}$. The one and three letter codes for amino acids are as recommended by the IUPAC - IUB Commission on the Biochemical nomenclature (1969). Restriction enzymes are referred to by the three letter nomenclature of Smith & Nathans (1973). The remaining abbreviations are listed in alphabetical order below, or defined in the text when they are first used.

A ₂₈₀ , A ₆₀₀ , etc.	- Absorbance at 280, 600 nm etc.
ACTH	- Adrenocorticotrophic hormone
B	- Corticosterone
bis-acrylamide	- N', N' -methylene bis-acrylamide
BSA	- Bovine serum albumin
C	- Complement
C - n	- Carbon atom
cDNA	- Complementary DNA
CMO	- Corticosterone methyl oxidase
cdNA	- Complementary DNA
DNA	- Deoxyribonucleic acid
DEAE	- Diethylaminoethyl
DHEA	- Dehydroepiandrosterone
dNTP	- Deoxyribonucleoside triphosphate (A, C, G or T specified instead of N where appropriate)
ddNTP	- Dideoxynucleoside triphosphate (N = A, G, C or T)
DOC	- Deoxycorticosterone

DTT	- Dithiothreitol
EDTA	- Ethylenediaminetetraacetic acid
EtOH	- Ethanol
F	- Cortisol
HBS	- Hepes buffered saline
HLA	- Human leucocyte antigens
Hepes	- N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid
HSD	- Hydroxysteroid dehydrogenase
MHC	- Major Histocompatibility Complex
mRNA	- Messenger RNA
NaOAc	- Sodium acetate
nm	- Nanometre
OD	- Optical density
21-OHase	- 21-Hydroxylase
OH	- Hydroxyl
PBS	- Phosphate buffered saline
rATP	- Adenosine triphosphate
RNA	- Ribonucleic acid
rRNA	- Ribosomal RNA
SLE	- Systemic lupus erythematosus
1 X SSC	- 150 mM NaCl / 15 mM Na citrate, pH 7.0
1 X TBE	- 90 mM Tris / 90 mM Boric Acid / 2.5 mM EDTA, pH 8.3
TEMED	- N, N, N', N' - tetramethylethylenediamine
Tris	- Tris (hydroxymethyl) aminoethane
U	- Units

PUBLICATIONS

Rodrigues, N.R., Dunham, I., Yu, C.Y., Carroll, M.C., Porter, R.R., & Campbell, R.D. Molecular characterisation of the HLA-linked steroid 21-hydroxylase B gene from an individual with congenital adrenal hyperplasia. (1987), EMBO Journal 6, 1653 - 1661

Rodrigues, N.R., Dunham, I., Yu, C.Y., Carroll, M.C., Porter, R.R., & Campbell, R.D. Structure of the HLA linked steroid 21-hydroxylase genes. (1987), Biochemical Society Transactions. 15, 656 - 657

Rodrigues, N.R., Rumsby, G. & Campbell, R.D. Polymorphism of the HLA - linked 21-hydroxylase genes. To be published in Immunobiology of HLA Vol. 2, ed. Dupont, B.

Cotton, R.G.H., Rodrigues, N.R. & Campbell, R.D. Reactivity of cytosine and thymidine in single base pair mismatches with hydroxylamine and osmium tetroxide and application to the study of mutations. Submitted to Proc. Natl. Acad. Sci. USA.

ABSTRACT

The Human Cytochrome P-450 21-Hydroxylase Genes

Nanda R. Rodrigues
Submitted for the degree of D. Phil.

Trinity College
Michaelmas term 1987

Deficiency of the cytochrome P-450 steroid 21-hydroxylase (21-OHase) which causes Congenital Adrenal Hyperplasia (CAH) is a monogenic autosomal recessive disorder which is linked to HLA. There are two 21-OHase genes in man, A and B, and they are mapped to the HLA class III region ~ 3 kb 3' to the complement genes C4A and C4B, respectively. Two genes encoding 21-OHase were isolated, characterized and sequenced. Both 21-OHase genes are ~ 3.3 kb in length and are split into 10 exons by nine introns. Comparison of the two genes showed that although they are highly conserved, there are three deleterious mutations in the 21-OHase A gene which cause frameshifts and introduce in phase premature termination codons. Thus the 21-OHase A gene is a pseudogene.

Comparison of the 21-OHase B gene to the other cytochrome P-450 sequences revealed that although the cysteine-429 was conserved in 21-OHase, there is very little homology with other cytochrome P-450, indicating it belongs to a separate family of genes within the superfamily. Clear evidence of polymorphism in 21-OHase is apparent on comparison with other 21-OHase B sequences. There is a size polymorphism of 494 and 495 amino acids.

The differing severities of 21-OHase deficiency in CAH may be due to allelic variants of the 21-OHase B gene, since in most cases the defect is not due to gene deletion (Rumsby *et al.*, 1986). A 21-OHase B gene from a patient with CAH was characterized and sequenced. There were 13 nucleotide alterations in his single 21-OHase B gene, one of which at codon 269 caused a serine to change to a threonine residue.

The G → C transversion in the 21-OHase B gene from the patient at codon 269 introduced a new NcoI restriction site into the gene. This restriction fragment length polymorphism (RFLP) was used to study other patients with CAH and normal individuals. The NcoI RFLP was found not to be confined to the 21-OHase B gene but was also present in some 21-OHase A genes. It is likely therefore that the mutation occurred in the pseudogene first and then transferred to some 21-OHase B genes.

CONTENTS

Acknowledgements..... i

Abbreviations..... ii

Publications..... iv

Abstract..... v

Contents..... vi

CHAPTER I INTRODUCTION

1.1 General Introduction..... 2

1.2 Steroidogenesis in Man..... 3

 1.2.1 Mineralocorticoids..... 4

 1.2.2 Glucocorticoids..... 5

 1.2.3 Sex Hormones..... 5

 1.2.4 Mechanism of Adrenal Steroid Regulation..... 6

 1.2.5 Fetal Sexual Development..... 7

1.3 Biochemistry of 21-Hydroxylase..... 7

 1.3.1 Cytochrome P-450s..... 8

1.4 21-Hydroxylase Deficiency..... 11

 1.4.1 Simple Virilizing 21-OHase deficiency..... 11

 1.4.2 Salt-wasting 21-OHase deficiency..... 12

 1.4.3 Late Onset 21-OHase deficiency..... 13

 1.4.4 Cryptic 21-OHase deficiency..... 14

1.5 Diagnosis and Therapy of CAH..... 14

1.6 Genetics of CAH..... 15

 1.6.1 HLA Linkage..... 15

 1.6.2 Genetic Linkage Disequilibrium..... 17

1.7 The Major Histocompatibility Complex (MHC)..... 18

 1.7.1 Class I and Class II Antigens..... 18

1.7.2	Class III Proteins.....	20
1.7.3	Molecular Mapping of the HLA region.....	21
1.7.4	The Complement System.....	23
1.7.5	Class III Genes.....	25
 <u>CHAPTER II MATERIALS AND METHODS</u>		
2.1	<u>Materials</u>	36
	Enzymes.....	36
	Chemical Reagents.....	36
	Radioactive Nucleotides.....	36
	Bacterial Strains, Plasmids, Phages and Cell lines.....	37
	Media.....	37
	Buffers.....	38
2.2	<u>Overnight Cultures</u>	39
2.3	<u>Preparation of DNA</u>	39
	2.3.1 Small Scale Isolation of Plasmid DNA.....	39
	2.3.2 Large Scale Isolation of Plasmid DNA.....	40
	2.3.3 Isolation of the Replicative form of Bacteriophage M13mp9.....	41
	2.3.4 Isolation of Bacteriophage by the Plate lysate Method.....	41
	2.3.5 Isolation of Bacteriophage by DE52 Purification Method.....	43
	2.3.6 Isolation of Chromosomal DNA.....	44
2.4	<u>Digestion of DNA by Restriction Endonucleases</u>	45
2.5	<u>Radiolabelling of DNA</u>	45
	2.5.1 "Fill in" Reaction.....	45
	2.5.2 Nick Translation.....	45
	2.5.3 Primer Extension.....	46

2.6	<u>Gel Electrophoresis</u>	46
	2.6.1 Polyacrylamide gel Electrophoresis.....	46
	2.6.2 Autoradiography.....	47
	2.6.3 Agarose Gel Electrophoresis.....	47
2.7	<u>Recovery of DNA from Gels</u>	48
	2.7.1 Elution of DNA from acrylamide gels.....	48
	2.7.2 Recovery of DNA from low gelling temperature Agarose.....	49
2.8	<u>Screening of Cloned DNA</u>	50
	2.8.1 Screening of Plasmid Clones.....	50
	2.8.2 Screening of Bacteriophage plaques.....	51
2.9	<u>Nucleotide Sequencing Techniques</u>	52
	2.9.1 Chemical Degradation.....	52
	2.9.2 Dideoxy Sequencing.....	54
	2.9.3 Analysis of Nucleotide Sequence data.....	58
2.10	<u>Isolation of Total Cellular RNA</u>	58
2.11	<u>DNA Transfer Blot Analysis</u>	59
	2.11.1 Blotting on Nitrocellulose Membranes.....	59
	2.11.2 Blotting on Nylon Membranes.....	61
2.12	<u>Northern Blot Analysis</u>	62
2.13	<u>Tissue Cell Culture</u>	62
	2.13.1 Long term Storage of Cells.....	63
2.14	<u>Transfection of Y1 Cells</u>	63
2.15	<u>Assay of Chloramphenicol Acetyltransferase Activity</u>	64

CHAPTER III STRUCTURE OF THE 21-HYDROXYLASE GENE

3.1	<u>Introduction</u>	66
3.2	<u>The Organization of the 21-OHase Genes in Individual HPFH</u>	67

3.2.1	<u>Cloning of the 21-OHase A Gene.....</u>	68
3.3	<u>The Organization of the 21-OHase Genes</u>	
	<u>in Individual KEM.....</u>	69
3.3.1	<u>Cloning of the 21-OHase B Gene.....</u>	70
3.4	<u>Subcloning of the 21-OHase A Gene into M13mp9....</u>	71
3.5	<u>Subcloning of the 21-OHase B Gene into M13mp9....</u>	72
3.6	<u>Structure of the 21-OHase Genes.....</u>	73
3.7	<u>Comparison of the Two 21-OHase Genes.....</u>	76
3.8	<u>Comparison of the Human 21-OHase Gene with</u>	
	<u>Bovine and Murine 21-OHase Genes.....</u>	78
3.9	<u>Comparison of the 21-OHase Genes with</u>	
	<u>other cytochrome P-450 Genes.....</u>	79
3.10	<u>Polymorphism of the 21-OHase Genes.....</u>	83
3.11	<u>Discussion.....</u>	84

CHAPTER IV CHARACTERIZATION OF A 21-HYDROXYLASE B GENE FROM A PATIENT
WITH CONGENITAL ADRENAL HYPERPLASIA

4.1	<u>Introduction.....</u>	89
4.2	<u>Southern Blot Analysis of Family Q.....</u>	89
4.3	<u>Cloning of the 21-OHase B gene from</u>	
	<u>Individual II.I.....</u>	91
4.4	<u>Subcloning of the 21-OHase B gene</u>	
	<u>from Individual II.I into M13mp9.....</u>	92
4.5	<u>Sequence Analysis of the 21-OHase B</u>	
	<u>gene from Individual II.I.....</u>	93
4.6	<u>NcoI Polymorphism in Individual II.I.....</u>	94
4.7	<u>Expression of the 21-OHase B gene</u>	
	<u>from Individual II.I.....</u>	95
4.8	<u>Discussion.....</u>	97

<u>CHAPTER V RESTRICTION LENGTH POLYMORPHISM STUDIES OF THE</u>	
	<u>21-HYDROXYLASE GENES</u>
5.1	<u>Introduction</u> 102
	5.1.1 TaqI RFLP..... 103
	5.1.2 KpnI RFLP..... 103
	5.1.3 NcoI RFLP..... 104
5.2	<u>NcoI RFLP in Individuals with CAH</u> 104
	5.2.1 Family H..... 105
	5.2.2 Family Bu..... 107
	5.2.3 Family M..... 108
5.3	<u>NcoI RFLP in a Random Population</u> 108
	5.3.1 Family P..... 109
	5.3.2 Family A..... 109
5.4	<u>AvaII Polymorphism</u> 110
	5.4.1 AvaII Band pattern..... 111
	5.4.2 AvaII / NcoI Band pattern..... 111
5.5	<u>Discussion</u> 112
 <u>CHAPTER VI GENERAL DISCUSSION AND SUMMARY</u> 117	
 REFERENCES..... 126	
 APPENDIX I..... 144	

CHAPTER I

INTRODUCTION

1.1 GENERAL INTRODUCTION

Several human metabolic disorders are inherited as autosomal recessive traits. In many of these genetic disorders, translation of the faulty genetic information results either in the production of a defective or abnormal enzyme or no enzyme at all. This is expressed physiologically as the lack of a particular enzymatic activity. The clinical manifestation of the disorders are usually the result of an accumulation of metabolic intermediates that cannot be processed in a normal manner because the appropriate enzyme activity is lacking. Congenital adrenal hyperplasia (CAH) is a family of inherited disorders of adrenal steroidogenesis (Finkelstein and Shaefer, 1979) resulting from deficiency of any of the several enzymes necessary for normal cortisol synthesis.

The first description of the manifestation of this disorder goes as far back as 1865 when Luigi De Crecchio described an autopsy of a cadaver which looked typically male on the outside, but the internal organs were that of a normal female including a vagina, uterus tubes and ovaries.

Over the past 30 years, CAH has been divided into several distinct forms based on deficiencies of discrete enzymatic functions of steroidal biogenesis which involve synthesis of steroid hormones such as cortisol and aldosterone. Cortisol is synthesized from cholesterol in the zona fasciculata of the adrenal cortex in five enzymatic steps. The C-22,27 side chain is cleaved to form pregnenolone, the 3β hydroxyl is dehydrogenated yielding progesterone which is successively hydroxylated at the 17α , 21, 11β positions to yield cortisol (Finkelstein & Shaefer, 1979). All these steps except for 3β -dehydrogenation require specific

cytochrome P-450 enzymes. These are heme containing enzymes with molecular weights of about 50000 which act as terminal oxidases of NADPH dependant electron transport pathways. Impairment of 21-hydroxylation is the most common enzymatic deficiency observed in CAH and the most common cause for ambiguous genitalia and salt-wasting in newborns.

21-hydroxylase (21-OHase) deficiency is linked to the major histocompatibility complex (HLA) on the short arm of human chromosome 6 and is inherited as a monogenic autosomal recessive trait (Dupont et al., 1977). There are two 21-OHase genes in man (White et al., 1984b) and they have been mapped to the HLA class III region together with the genes encoding the serum complement proteins C2, Factor B, C4A and C4B (Carroll et al., 1984b).

Four distinct clinical manifestations of 21-OHase deficiency have been identified; simple virilizing, salt-wasting, late onset and cryptic (New and Levine, 1984). Simple virilizing due to excessive production of fetal androgens and salt-wasting due to aldosterone deficiency are classical forms of 21-OHase and have shown significant association with HLA Bw47 and to a lesser extent with HLA B51 / 5, B53 and B60. Linkage disequilibrium has also been reported for the nonclassical forms of 21-OHase deficiency, with a significantly increased frequency of the HLA haplotype B14, DR1 in both the late onset and cryptic forms. To understand the pathophysiology of 21-OHase deficiency it is essential to know the mechanism and regulation of adrenal steroidogenesis.

1.2 Steroidogenesis in Man

The two primary organs responsible for the secretion of hormonal steroids in man are the adrenal cortex and the gonads. The adrenal synthesizes three main classes of hormones: mineralocorticoids (17-deoxy pathway), glucocorticoids (17-hydroxy pathways), and sex steroids

(Finkelstein & Shaefer, 1979). The primary, if not exclusive, precursor of steroid hormones is cholesterol. The biosynthesis of steroids is depicted in Figure 1.1a,b.

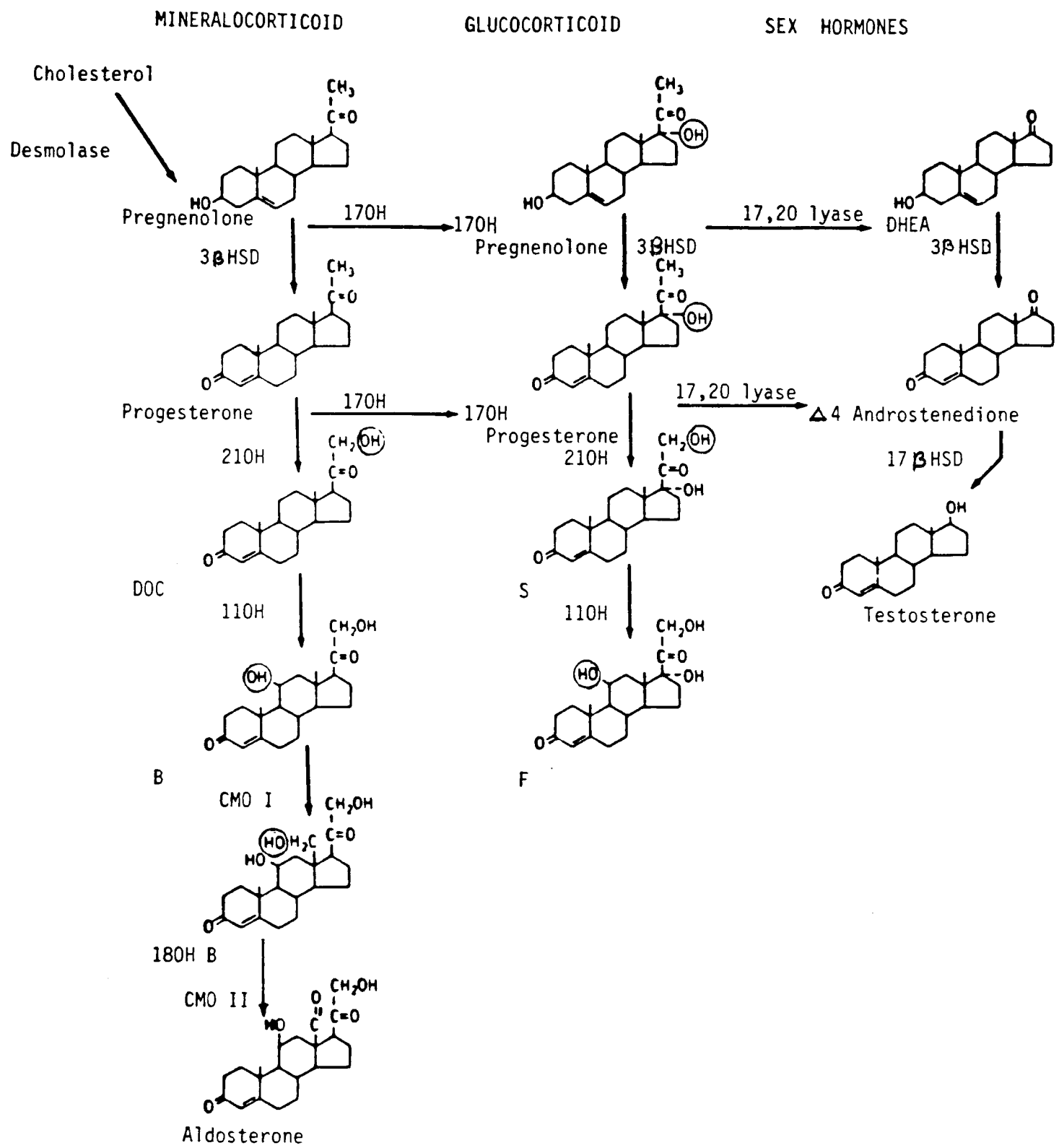
1.2.1 Mineralocorticoids (17 Deoxy Pathway)

The first step in the biosynthesis of steroids is the conversion of cholesterol to pregnenolone (Simpson & Boyd, 1968) (Figure 1.1). Adrenocorticotrophic hormone (ACTH) acts on the adrenals to increase the conversion by binding to the receptors on the external membranes of adrenal cortical cells, stimulating increased synthesis of adenosine - 3',5' - monophosphate (cyclic AMP). Cyclic AMP activates adrenal cellular phosphoprotein kinases which catalyze the side chain cleavage converting cholesterol to pregnenolone (Garren et al., 1971).

Pregnenolone, a Δ^5 - steroid, is converted to a biologically active Δ^4 - steroid, progesterone, by the enzymes 3β - hydroxysteroid dehydrogenase (3β -HSD) and $\Delta^{5,4}$ -oxosteroid isomerase (Koritz, 1964). Progesterone is then hydroxylated at the C-21 position by the enzyme 21-hydroxylase (P450c21) (Ryan & Engel, 1957) to form deoxycorticosterone, an active salt-retaining hormone. When deoxycorticosterone is hydroxylated at the C-11 position by 11β -hydroxylase (P450c11) (Suhara et al., 1978), corticosterone, which is a weak mineralocorticoid, is formed. Corticosterone is the precursor of aldosterone, the most potent salt-retaining hormone. The enzyme system catalysing the conversion of corticosterone to aldosterone is found almost exclusively in the zona glomerulosa (Ulick, 1976), and occurs in two stages. The first step is that of 18-hydroxylation (Pasqualini, 1964) requiring corticosterone methyl oxidase I activity which uses P450c11. This enzyme thus 11-hydroxylates 17-hydroxysteroids and both 11- and 18-hydroxylates 17-deoxysteroids. The final step in aldosterone synthesis is a further oxidation of the

Figure 1.1a

Simplified scheme for adrenal steroidogenesis showing the synthesis of mineralocorticoids, glucocorticoids and sex hormones from cholesterol. Each hydroxylation step is indicated and the newly added hydroxyl group is circled. DOC, deoxycorticosterone; B, corticosterone; S, 11-deoxycortisol; F, Cortisol; DHEA, dihydroepiandrosterone. 17-OH, 21-OH, 11-OH, CMOI, CMOII indicate hydroxylase activities mediated by specific cytochrome P-450s; the cholesterol desmolase (or P-450-side chain cleavage) and 17, 20 lyase steps are also mediated by P-450s. 3 β -HSD is 3 β -hydroxysteroid dehydrogenase, which is not a P-450 enzyme (New & Levine, 1973).



18-hydroxyl group to an aldehyde. The enzyme responsible for this activity, which has been called corticosterone methyl oxidase II again requires P450c11 (Yanagibashi et al., 1986).

1.2.2 Glucocorticoids (17-Hydroxy Pathway)

Glucocorticoid synthesis requires hydroxylation at the C-17 position. The conversion of pregnenolone and progesterone to 17-hydroxy-pregnenolone and 17-hydroxyprogesterone by 17 α -hydroxylase (P450c17) occurs in the inner adrenal zona fasciculata (Kremers, 1976). A virtual absence of 17 α -hydroxylase activity was demonstrated in the bovine zona glomerulosa (Harkins et al., 1974). A similar pattern of 17 α -OHase activity is believed to exist in the human adrenal cortex and it is generally accepted that glucocorticoids and adrenal androgens, both of which depend on 17-hydroxylation for their synthesis, originate predominantly in the zona fasciculata. When 17-hydroxyprogesterone undergoes 21-hydroxylation, 11-deoxycortisol is formed. 11-deoxycortisol is further hydroxylated at C-11 by P450c11 to form cortisol, the most potent glucocorticoid in humans.

1.2.3 Sex Hormones

The main unconjugated C-19 steroid secreted by the adrenal cortex is dehydroepiandrosterone (DHEA). It results from the side-chain cleavage of C-21 steroid, 17-hydroxypregnenolone, by the action of a desmolase enzyme. DHEA, a Δ^5 -steroid with little androgenic activity, is converted to Δ^4 -androstenedione, a moderately active androgen, by the 3 β -hydroxysteroid dehydrogenase and isomerase enzymes. Δ^4 -androstenedione is reduced at the C-17 position to form testosterone. Testosterone is converted peripherally by the 5 α -reductase enzyme to dehydrotestosterone. Both testosterone and dehydrotestosterone have androgenic

activity, affecting different target tissues (Siiteri & Wilson, 1974; Peterson et al., 1977).

1.2.4 Mechanism of Adrenal Steroid Regulation

The circulating level of plasma cortisol mediates the hypothalamic-pituitary-adrenal feedback system. As cortisol is the primary feedback inhibitor of ACTH in man, decreased cortisol secretion results in enhanced ACTH secretion. In CAH where an enzyme deficiency causes impaired cortisol synthesis, there is excessive ACTH secretion and hyperplasia of the adrenal cortex (New et al., 1982) (Figure 1.2).

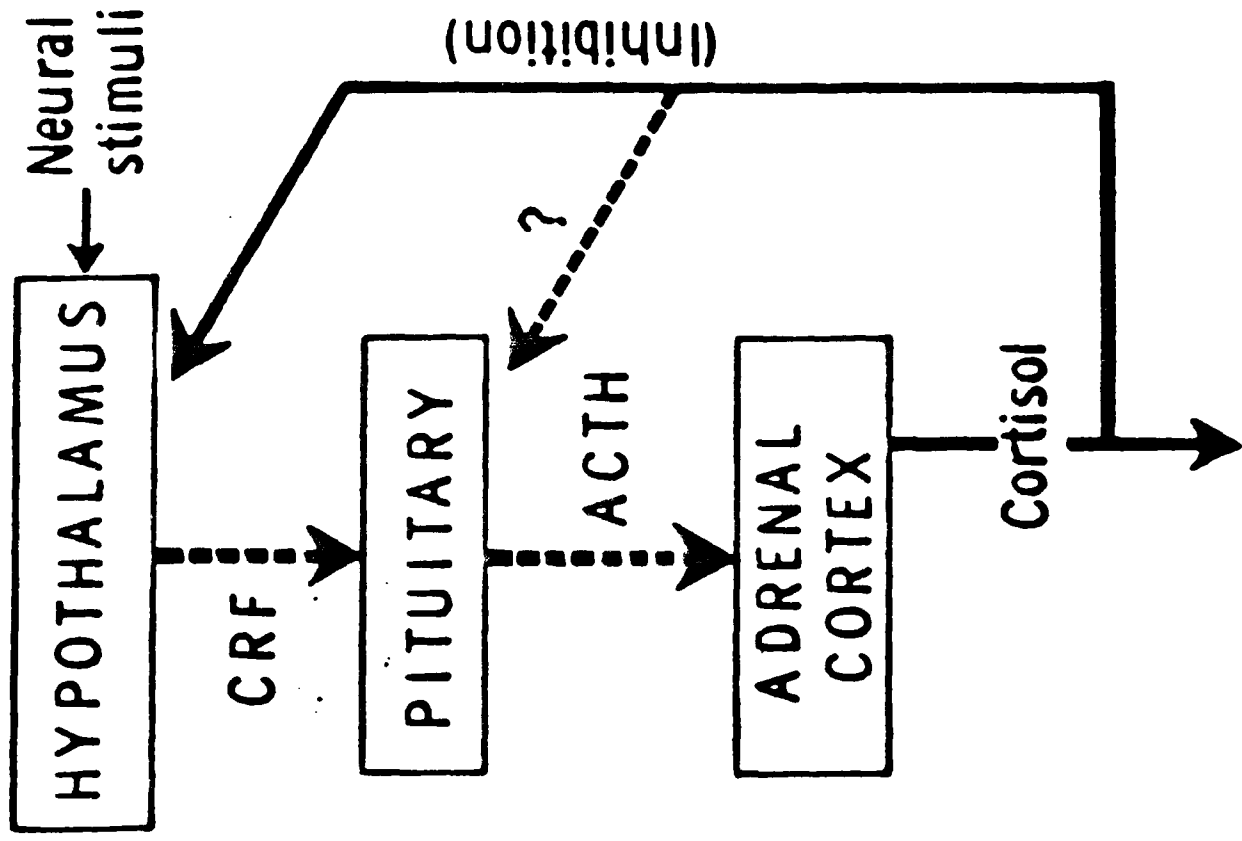
Aldosterone secretion is primarily regulated via the renin-angiotensin system which is responsive to the state of electrolyte balance and plasma volume (New & Peterson, 1966). Aldosterone secretion is also stimulated directly by high serum K^+ concentration.

It has been proposed that the zona glomerulosa and the zona fasciculata behave as two separate glands with respect to regulation and secretion (New & Seaman, 1970). This concept suggests that ACTH regulates steroidogenesis in the fasciculata by stimulating secretion of cortisol, corticosterone and androgens; while that of the glomerulosa is regulated by the renin-angiotensin system where angiotensin stimulates aldosterone secretion, with ACTH presumably exerting only a secondary influence on the secretion of aldosterone by the glomerulosa (Ganong et al., 1974; Ganong et al., 1966; Mason et al., 1979). While the zona fasciculata lacks the enzyme necessary for the terminal step of aldosterone synthesis, the zona glomerulosa lacks the 17α -hydroxylase activity required for the production of 17-hydroxycorticoids and androgens.

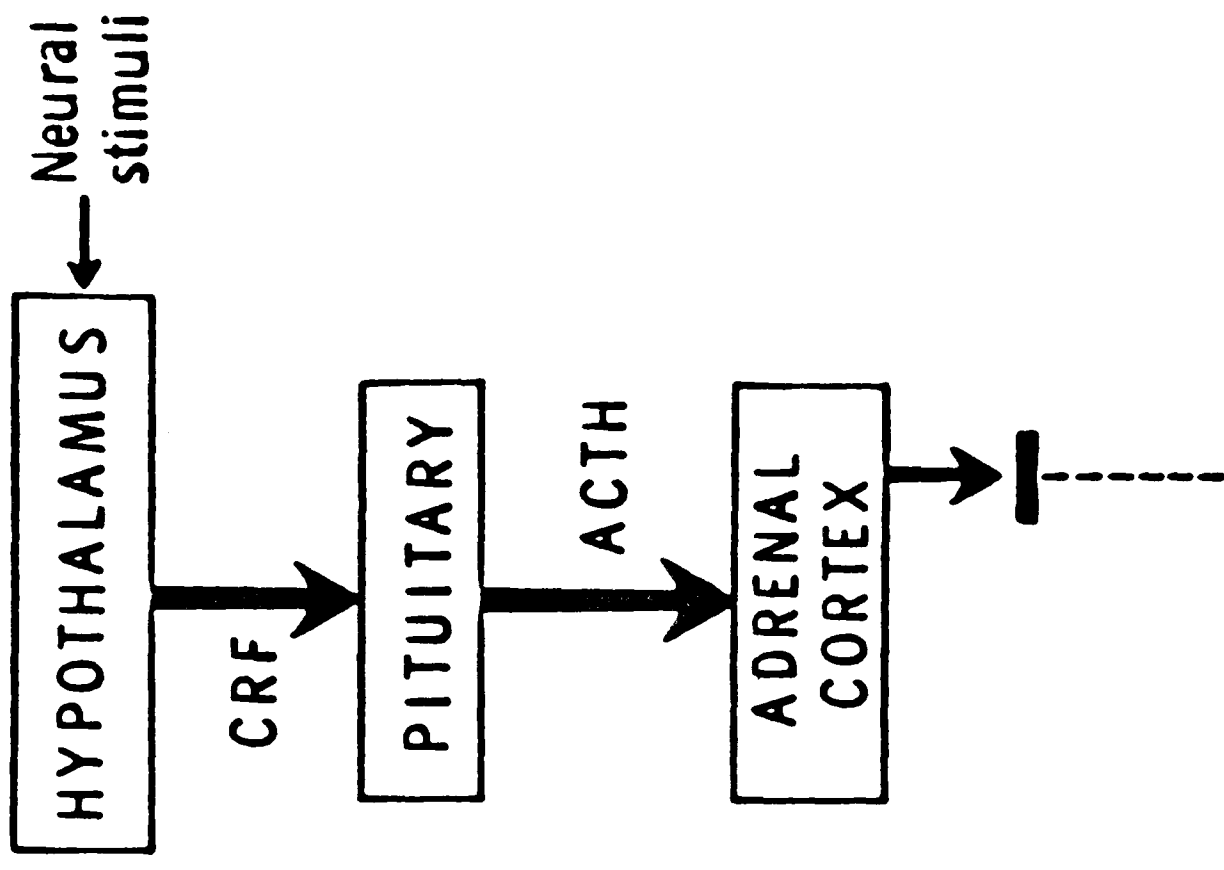
Figure 1.2

The regulation of cortisol secretion in normal subjects and in patients with congenital adrenal hyperplasia. Cortisol being the primary feedback inhibitor of ACTH, decreased cortisol synthesis due to enzymatic deficiency in CAH (indicated by a horizontal bar) causes excessive ACTH secretion and hyperplasia of the adrenal cortex (New & Levine, 1973).

NORMAL



ADRENOGENITAL



1.2.5 Fetal Sexual Development

Normal differentiation of male genitalia is dependent on two functions of the fetal testis (Jost, 1971).

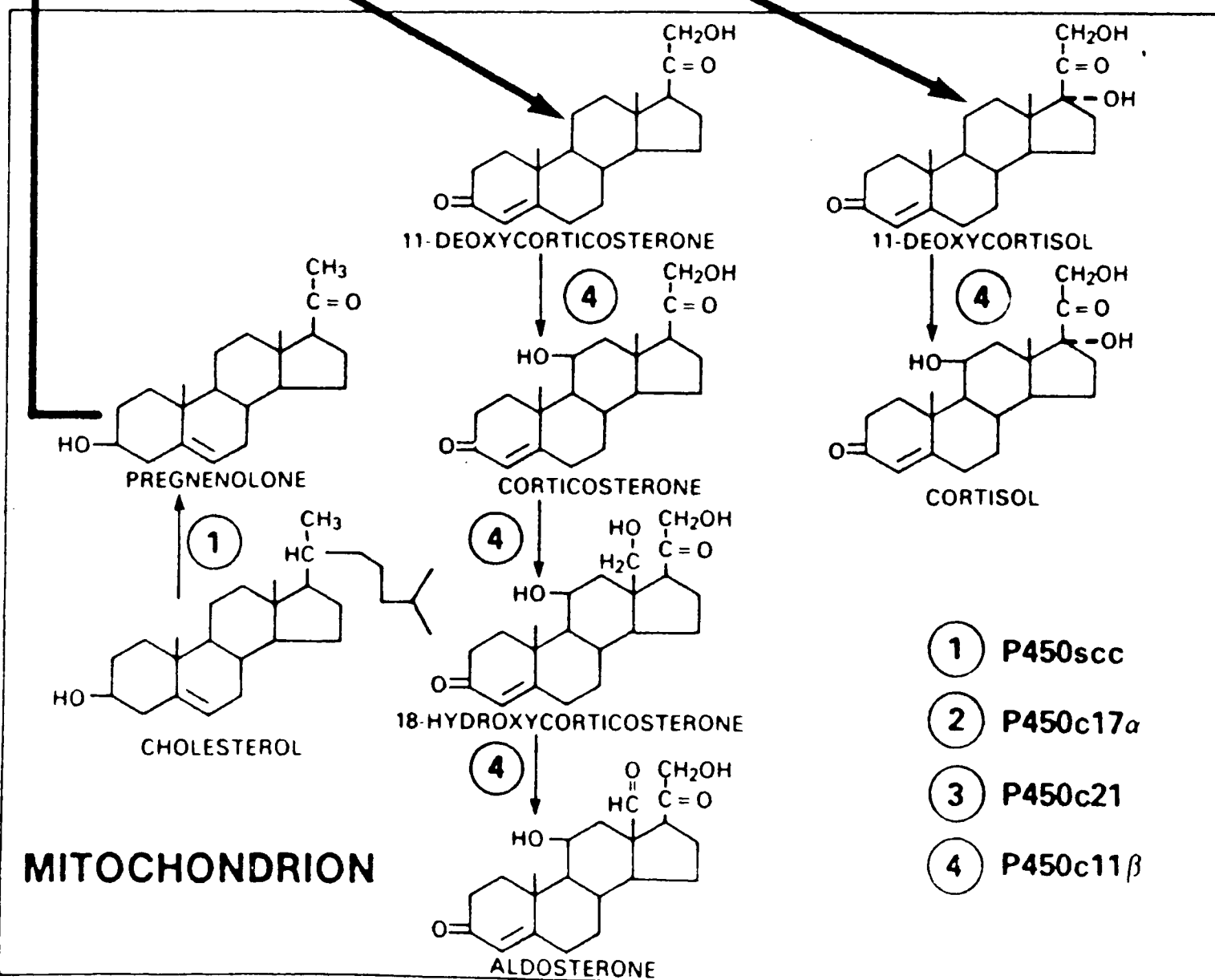
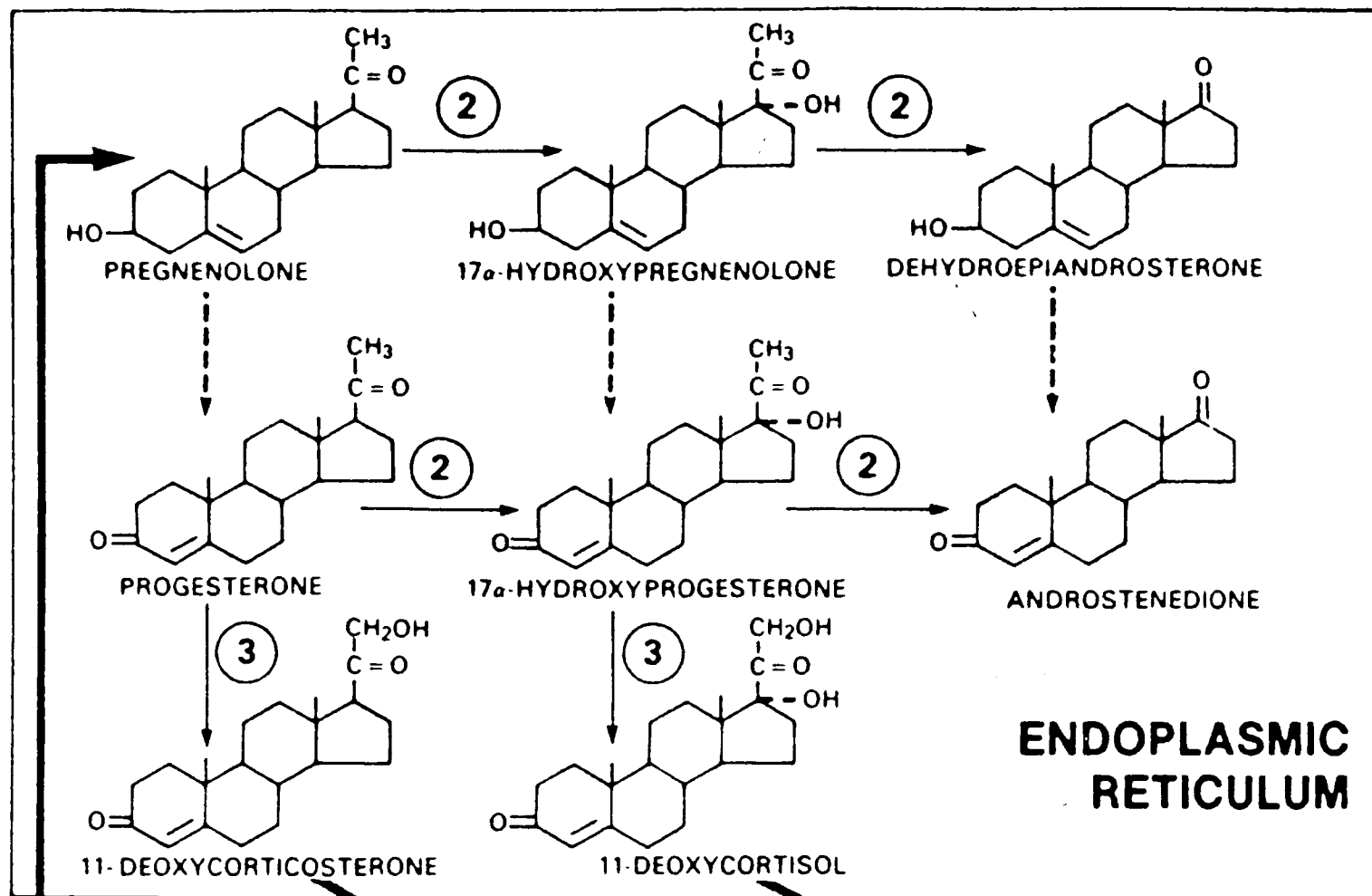
(a) Secretion of the androgen testosterone, which stimulates the Wolffian ducts to develop into the male internal genitalia and which is reduced in the target tissue to dehydrotestosterone. Dehydrotestosterone causes differentiation of the male external genitalia. In the male the normal source of androgen is the fetal testis but androgen from the adrenal or exogenous sources can cause masculinization of the external genitalia.

(b) Secretion of a nonsteroidal substance (Josso, 1977) which inhibits Mullerian duct development such that normal males are born without a uterus. Since the fetal ovary secretes neither testosterone nor the inhibiting factor necessary to inhibit Mullerian structures, the normal female is born without male external genitalia (i.e. with female external genitalia) and without Mullerian repression (i.e. with a uterus and Fallopian tubes). Thus the ovary does not play a determining role in sex differentiation.

Female fetuses exposed to high levels of androgen consequent to CAH, an androgen-producing tumour in the mother, or administration of androgens to the mother, manifest virilization of the external genitalia but have normal internal female genitalia.

1.3 Biochemistry of 21-Hydroxylase

Of all the hydroxylases participating in the biosynthesis of adrenal steroids, the enzyme steroid 21-hydroxylase [21-OHase; steroid hydrogen-donor: oxygen oxidoreductive (21-hydroxylating) EC 1.14.99.19] is the most extensively studied. Unlike most of the other enzymes involved in adrenal steroidogenesis which are found in the gonads as



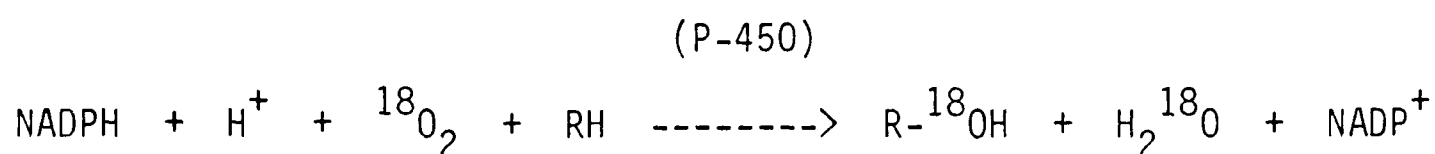
- (1) P450_{scc}
- (2) P450_{c17 α}
- (3) P450_{c21}
- (4) P450_{c11 β}

well as the adrenals, 21-OHase is normally found solely in the adrenal cortex (Tomaoki, 1973). It was shown to be associated with the microsomal fraction and it was established that NADPH and O_2 are necessary for enzyme activity, the source of O_2 being atmospheric oxygen (Ryan and Engel, 1973).

21-OHase is one of the many oxygenases that function in lipid transformation and are recognised from their biochemical properties as falling in the class of red carbon monoxide-binding pigments known as cytochromes P-450 (Coon & Koop, 1983). The enzyme purified from porcine adrenocortical microsomes showed a molecular weight of 54,000 (Yuan et al., 1983). The spectral properties of the enzymes were characteristic of cytochrome P-450s which are described in Section 1.3.1. That this P-450 catalyzed C21 hydroxylation of progesterone and 17-hydroxyprogesterone was shown by the quantitative disappearance of the substrate in both cases and appearance of the corresponding 21-hydroxy steroid (Yuan et al., 1983).

1.3.1 Cytochrome P-450s

Cytochrome P-450s constitute a family of heme proteins that catalyze monooxygenase reactions. These enzymes are reduced by electrons from reduced pyridine nucleotides. The reduced enzyme activates molecular oxygen, and one atom of the oxygen is inserted into the product as a hydroxyl group. The stoichiometry of the hydroxylation reaction being as follows (Nordblom & Coon, 1977):



The substrate is a lipophilic substance that becomes more hydrophilic as a result of the new hydroxyl group and is readily removed from the body by filtration. Lipophilic drugs are thus removed from the body by monooxygenation as well as metabolism of a number of substances made in the body such as steroid hormones and the prostaglandins. The heme-containing monooxygenases vary widely in substrate specificity and organ distribution. Several of the cytochrome P-450s are induced to high levels by xenobiotics while others are expressed constitutively and are involved with steroid or fatty acid metabolism. The extraordinarily broad substrate specificity of the enzyme systems results from the multiplicity of distinct molecular forms of the terminal oxidase cytochrome P-450 which together form the P-450 system. The P-450 system is as important in defending the body against foreign chemicals as the immune system is in dealing with foreign pathogens (Adesnik & Atchison, 1983).

All Cytochrome P-450s have characteristic absolute absorbance spectra. These heme proteins show a pronounced peak at approximately 420nm known as the Soret peak and two peaks in the region of 550nm (Hall, 1983). When heme proteins are reduced, the heme group combines with carbon monoxide (CO) which results in a shift in the Soret peak which in most heme proteins is confined to a few nm. In cytochrome P-450s, the shift in the Soret peak produced by CO is extensive and the peak then appears at 450 nm which gives the enzymes the name " pigment 450 " (P-450). When a substrate binds to an enzyme to form an enzyme substrate complex, the conformation of the enzyme changes, and in case of P-450, this causes a shift in the Soret peak to approximately 390nm. These three spectral properties of cytochrome P-450s are important in studying these enzymes.

While all cytochrome P-450s share some structural features, they have been divided into several families on the basis of their

characterization at the protein level. At least ten gene families are presently known in the P-450 superfamily and the amino acid sequence of a protein from any of the 10 families is $\leq 36\%$ similar to that from any of the other nine families (Nebert et al., 1987). Excluding the LI family in yeast and the CI family in Pseudomonas, the mammalian superfamily consists of at least eight P-450 gene families (Nebert & Gonzalez, 1987). These families include cytochrome P-450s induced by phenobarbital (Atchison and Adesnik, 1983), those induced by polycyclic hydrocarbons (Jaiswal et al., 1985), the P-450 responsible for cholesterol side-chain cleavage (Morohashi et al., 1984) and 21-OHase (White et al., 1984a). Those P-450 genes considered to be in the same subfamily have $\geq 68\%$ similarity to other genes in the same subfamily. Where there exist more than one subfamily in the same family, the amino acid sequence of a protein in any one subfamily is about 40% to 65% similar to that in any of the other subfamilies (Nebert et al., 1987). Thus the cytochrome P-450 genes constitute a superfamily (Dayhoff et al., 1978) which is defined as a collection of genes which are similar in a statistically significant manner, but are less than 50% homologous.

Genetic defects in specific cytochrome P-450 catalytic activities have been implicated as the cause of hereditary deficiency in cortisol biosynthesis by the adrenal gland (New et al., 1982). The conversion of cholesterol to the steroid hormones requires co-ordination of mitochondrial and microsomal enzyme systems. The enzyme reactions involved are that of hydroxylation, dehydrogenation, oxidase, C-C cleavage, aromatization and isomerization. All these, with the exception of dehydrogenation and isomerization require cytochrome P-450 enzymes. To date 13 human P-450 genes have been cloned and the number is increasing rapidly (Nebert et al., 1987). These should prove useful in the diagnosis, treatment and understanding of numerous clinical disorders arising from defects in cytochrome P-450s.

1.4 21-Hydroxylase Deficiency

CAH due to the defective 21-hydroxylation is the most common inborn errors of adrenal steroidogenesis affecting ~ 1 in 5000 to 1 in 15,000 births in a severe form and 0.3% in a milder form in the general Caucasian population. The Yupik-speaking Eskimoes of western Alaska show an extremely high incidence of 1/684 of classical 21-OHase deficiency affecting 1 in 684 births, while the Ashkenazi Jews show a 3.7% prevalence of the nonclassical 21-OHase deficiency. Other populations show varying prevalence of the nonclassical form such as Hispanics, 1.9%; Yugoslavs, 1.6%; and Italians, 0.3% (Speiser et al., 1985). Deficiency of the enzyme 21-hydroxylase results in decreased cortisol synthesis which induces increased ACTH secretion (Ganong et al., 1974). This would result in overproduction of cortisol precursors and sex steroids, the biosynthesis of which does not require 21-OHase. Early studies demonstrated that patients with 21-OHase deficiency have excessive urinary pregnanetriol, the metabolite of 17-hydroxyprogesterone (Bongiovanni et al., 1954, Butler and Marrian, 1937). Urinary 17-ketosteroids, which result from the metabolism of dehydroepiandrosterone, Δ^4 -androstenedione and testosterone are also present in increased amounts. There are four distinct clinical manifestations of 21-OHase deficiencies: simple virilizing, salt-wasting, late onset and cryptic (New and Levine, 1984) (Table 1.1) and these are discussed below.

1.4.1 Simple Virilizing 21-OHase deficiency

The most prominent feature of this form of 21-OHase deficiency is virilization (New et al., 1982). Adrenocortical function begins with the third month of gestation, the fetus is therefore exposed at the critical time of sexual differentiation to the overproduced fetal adrenal androgens resulting from 21-OHase deficiency. The external

TABLE 1.1 21-OHase deficiency syndromes

Onset	Symptoms	Hormonal deficiency	21-hydroxylase genotype
Classical, Congenital	Saltwasting and Virilization	Aldosterone Cortisol	SW*/SW
Classical, Congenital	Virilization	Cortisol	SW/SV** or SV/SV
Non-Classical, Late Onset	Virilization	Cortisol	SW/Non Classical or SV/Non-Classical
Non-Classical, Cryptic	Asymptomatic	Cortisol	Non-Classical/Non-Classical

* SW: Saltwasting 21-OH Allele.
** SV: Simple Virilizing 21-OH allele.

genitalia in the female fetus are masculinized by excessive fetal adrenal androgens, resulting in female pseudohermaphroditism. In rare cases the masculinization may be so profound that the urethra is penile (Wilkins, 1962). Since the female fetus has no testis, Mullerian inhibiting factor is not produced and the female is born with an uterus and Fallopian tubes. The absence of the Wolffian system, despite increased androgen levels, suggests that the level of androgen necessary for Wolffian development is higher than that produced in CAH. Since CAH due to 21-OHase deficiency is the most common cause of ambiguous genitalia in the newborn and because the female pseudohermaphrodite with this disorder has the capacity for an entirely normal female role, including fertility, it is very important to consider this disorder in any inter-sex problem present at birth. Males with this disorder do not manifest genital abnormalities at birth. Without treatment, both males and females manifest rapid somatic growth with advanced bone age, progressive virilization resulting in early fusion of epiphysis and eventual short stature. Other symptoms of androgen excess include premature appearance of sexual hair, acne, an adult body odour and continued excessive growth of the penis or clitoris. Females may have polycystic ovaries.

1.4.2 Salt-wasting 21-OHase deficiency

Salt-wasting 21-OHase deficiency and simple virilizing 21-OHase deficiency together constitute the classical form of CAH. The nature of virilism is the same in both, but in addition there is profound aldosterone deficiency which reduces sodium resorption in the renal tubule, resulting in hyponatremia and hyperkalemia (New et al., 1982). Life threatening adrenal crisis which includes low serum sodium, high serum potassium and vascular collapse may occur in the first few weeks of life

and therefore must be recognised soon after birth in order for treatment to be administered prior to a crisis.

To explain these two forms of 21-OHase deficiency, it has been proposed that an enzyme deficiency may occur in the fasciculata, but not in the glomerulosa. This is in concert with the theory that the zona fasciculata and the zona glomerulosa function as two separate glands under different regulatory control (New and Seaman, 1970). A hypothesis based on the studies of New et al. (1981) and Kuhnle et al. (1981) states that:

- 1) In both simple virilizers and salt-wasters there is a fasciculata defect of 21-hydroxylation in both the 17-hydroxy and 17-deoxy pathways, and
- 2) in the salt-waster there is also a defect in the 21-hydroxylation in the glomerulosa, while in the simple virilizer, the glomerulosa is spared this defect (Figure 1.3).

1.4.3 Late Onset 21-OHase Deficiency

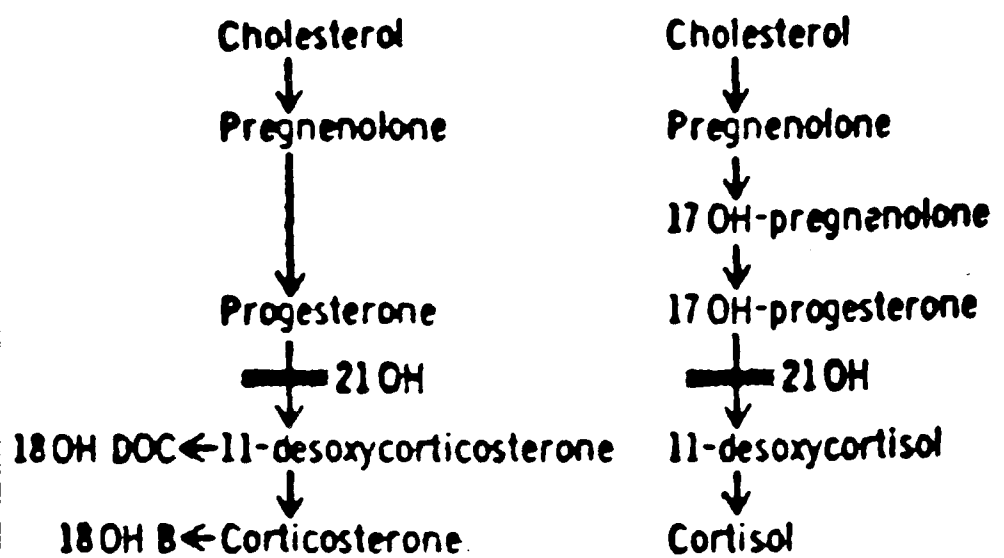
Nonclassical symptomatic (also called late onset, attenuated or acquired) 21-OHase deficiency is an extremely common syndrome compared to classical 21-OHase deficiency and is characterized by virilization, menstrual disturbances and endocrinological features consistent with 21-OHase deficiency that occur in later childhood or adolescence. (Decourt et al., 1957, Kohn et al. 1982). Unlike females with classical 21-OHase deficiency, females with late onset 21-OHase deficiency demonstrate no evidence of in utero virilization and are born with normal vaginal and urethral orifices and no labial fusion. Both types of patients respond similarly to glucocorticoid treatment. This late manifestation of the defect raised the question whether this disorder was inherited or an acquired disorder distinct from CAH.

Figure 1.3

Pathway of adrenal steroidogenesis in the simple virilizing and salt wasting forms of congenital adrenal hyperplasia due to 21-OHase deficiency. There is an enzyme defect in the zona fasciculata in both forms of CAH. In simple virilizers, 21-hydroxylation in the glomerulosa is normal as opposed to defective 21-hydroxylation in the glomerulosa of the salt-wasters (Kunhle et al., 1981). Defective 21-hydroxylation is indicated by a thick horizontal bar across the arrows in the pathways.

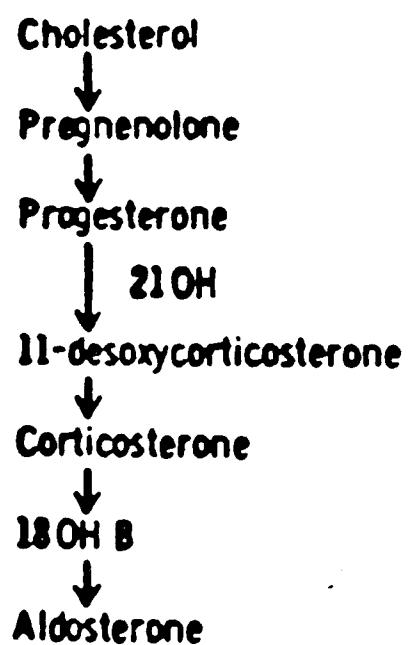
Zona fasciculata
ACTH →

Simple virilizing & salt-wasting CAH:

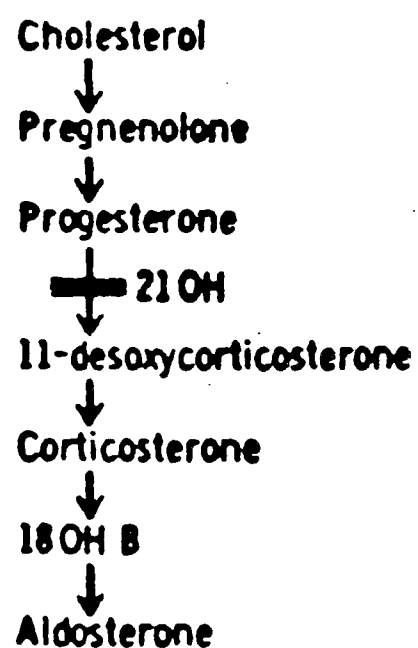


Zona glomerulosa
Renin → Angiotensin II →

Simple virilizing CAH:



Salt-wasting CAH:



1.4.4 Cryptic 21-OHase Deficiency

This form of 21-OHase deficiency is classified as non-classical asymptomatic 21-OHase deficiency and the affected individuals show biochemical profiles characteristic of a mild 21-OHase deficiency (Section 1.5). There are no signs of virilization, abnormal puberty and growth or infertility (Levine et al., 1980), and apart from the biochemical abnormalities discussed in the following section these individuals remain entirely asymptomatic.

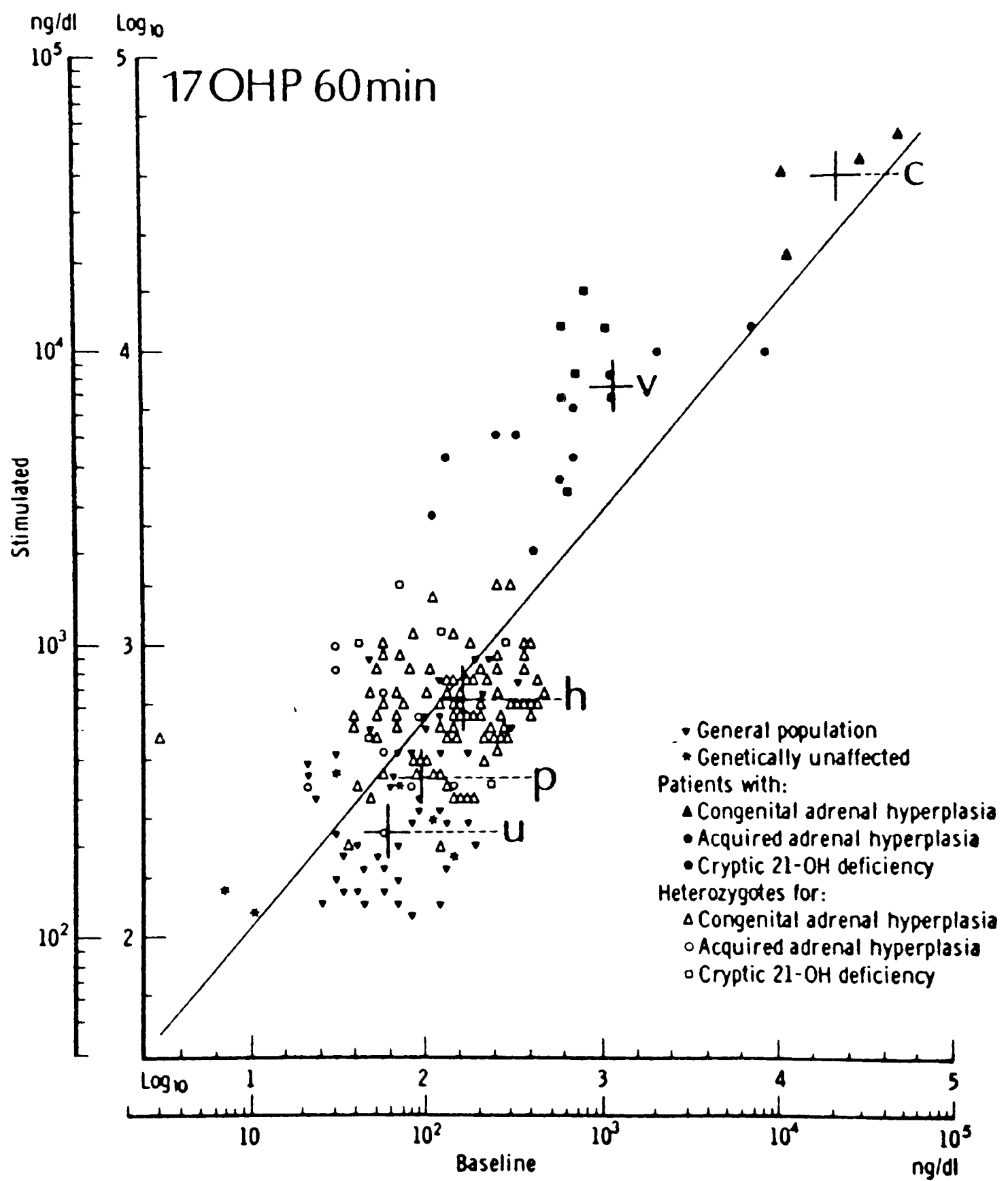
It was proposed that these asymptomatic individuals had 21-OHase deficiency as a result of inheriting two recessive genetic defects, a severe 21-OHase deficiency gene and a mild nonclassical 21-OHase deficiency allele or homozygosity for the mild nonclassical allele (Table 1.1).

1.5 Diagnosis and Therapy of CAH

Untreated 21-OHase deficiency results in elevated levels of cortisol precursors and adrenal androgens in the blood of patients. Cortisol levels fall in the lower limits of detection. The most striking difference in patients and normal individuals is the elevation of 17-hydroxyprogesterone, the immediate substrate for the 21-OHase enzyme. The levels of 17-hydroxyprogesterone in blood can be used for diagnosis of the disorder and for assessing the efficacy of treatment. A series of nomograms created relating the baseline and ACTH stimulated levels of 17-hydroxyprogesterone, Δ^4 androstenedione have provided hormonal standards to use in assignment of the type 21-OHase deficiency (New, 1985) (Figure 1.4). Patients with the classical 21-OHase deficiency, those with the nonclassical 21-OHase deficiency, obligate heterozygotes (parents of a patient) and normal individuals have levels of 17-hydroxyprogesterone that plot on distinct parts of the curve although

Figure 1.4

Nomogram relating base-line to ACTH-stimulated serum concentrations of 17-hydroxyprogesterone. The scales are logarithmic. A regression line for all data points is shown. The mean for each group is indicated by a large cross and adjacent letter: c denotes classic 21-OHase deficiency, v variant or nonclassic 21-OHase deficiency (combined mean of values in patients with cryptic and late-onset disease), h heterozygotes for all forms of 21-OHase deficiency, p general population, and u known unaffected persons (e.g., siblings of patients with 21-OHase deficiency who carry neither affected parental haplotype as determined by HLA typing). OH denotes hydroxyl. (New et al., 1982)



these do not distinguish between salt-wasters and the simple virilizers, or the symptomatic and asymptomatic nonclassical patients, or the heterozygous carriers of all the types of 21-OHase deficiency. Prenatal diagnosis of 21-OHase deficiency is done by detection of elevated concentrations of 17-ketosteroids, pregnanetriol, 17-hydroxyprogesterone and Δ^4 androstenedione in the amniotic fluid of the affected fetus (Pang *et al.*, 1985).

The treatment of 21-OHase deficiency consists mainly of replacement of the deficient steroid hormones. Physiologic doses of hydrocortisone suppress secretion of ACTH causing the levels of 17-hydroxyprogesterone and adrenal androgens to decrease into the normal range (Winter, 1980). In patients with salt-wasting 21-OHase deficiency, mineralocorticoid replacement is usually accomplished with oral fludrocortisone in a dose that suppresses plasma renin activity to normal levels without causing hypertension. Female patients with ambiguous genitalia can, with surgery or early therapeutic intervention, have normal psychosexual development and sexual function.

1.6 Genetics of CAH

Population studies have established that 21-OHase deficiency is transmitted as an autosomal recessive trait (Wilkins, 1962; Childs *et al.*, 1956) with females and males equally at risk. With a few exceptions either the simple virilizing or salt-wasting form is found consistently within a family. Fifty to eighty percent of the patients with 21-OHase deficiency have the salt-wasting form (Cohen, 1969).

1.6.1 HLA Linkage

The genes for HLA (Human leucocyte antigens) which are cell surface antigens important in transplantation, are located in the Major Histo-

compatibility Complex (MHC) on the short arm of chromosome 6. The MHC encodes three classes of proteins. The class I proteins (HLA-A, B, C) are encoded on the telomeric side and the class II proteins (HLA-DP, DZ/DO, DQ, DR) on the centromeric side of the class III proteins which are encoded in the middle.

Each of the loci of the HLA region is polymorphic and multiple alleles have been demonstrated for each locus. Each individual inherits one chromosome 6 from his father and one from his mother. The HLA genotype of an individual is written as follows:

$$\begin{array}{c} \underline{A3, Bw47, (C6), DR7} \\ A2, B44, \quad C5, \quad DR6 \end{array}$$

in which one haplotype (A3, Bw47, (C6), DR7) is inherited from one parent, while the other haplotype (A2, B44, C5, DR6) is inherited from the other parent.

Close genetic linkage between HLA and CAH due to 21-OHase deficiency was first described by Dupont et al. (1977) and substantiated by Levine et al. (1978). HLA genotyping of parents and children in six families with one or more child affected with CAH due to 21-OHase deficiency showed that in five families all the affected offspring were HLA identical and different from their unaffected sibs. More detailed genetic mapping of the specific location of the 21-OHase deficiency gene relative to the different loci that constitute the HLA complex established that the 21-OHase deficiency gene was located between the HLA-A locus and the centromere, but was further away from the centromere than the glyoxylase I (GLO I) locus (Dupont et al., 1980). The gene was thus mapped within 3-4 centimorgans.

1.6.2 Genetic Linkage Disequilibrium

Not only does the 21-OHase deficiency locus show genetic linkage with the HLA loci, but it also shows genetic linkage equilibrium, which is the nonrandom association of particular alleles of different genetic loci. In patients with classical and nonclassical 21-OHase deficiencies, certain HLA antigens appear with either a significantly increased or decreased frequency relative to their frequency in the general population. The most significant association for classical 21-OHase deficiency is with HLA Bw47, where the combined relative risk is 15.4 (Pollack et al., 1979). Slight increases have also been reported for Bw51, Bw53, Bw60, and DR7. The patients with the Bw47 haplotype show predominance of the haplotype A3 Cw6 Bw47 DR7. Several studies have also demonstrated that A1, B8 and DR3 are consistently decreased among 21-OHase deficient patients (Pollack et al., 1979). It was recently reported that some of these HLA antigen associations are characteristic of the different forms of 21-OHase deficiency (Dupont et al., 1984). For example, B5 and B14 antigens were selectively increased among patients with the simple virilizing form; B40 was increased in the salt wasting form and Bw47 was increased in both forms. These data strongly suggest that simple virilizing and salt-wasting 21-OHase deficiency are due to different mutational defects in the 21-OHase deficiency locus.

Genetic linkage disequilibrium has also been reported for the nonclassical form of 21-OHase deficiency (Laron et al., 1980). Initial studies suggested that the late onset 21-OHase deficiency was not HLA linked (New et al., 1979, Morillo & Gardner, 1979). More recent reports have provided evidence that this disorder is in fact genetically linked to HLA (Laron et al., 1980; Pollack et al., 1981). It has been proposed that classical and late onset 21-OHase deficiencies are allelic variants (Pollack et al., 1981). A significantly increased frequency of HLA-B14,

DR1 and complement factor BfS in both late onset and cryptic 21-OHase deficiency has been observed (Kohn et al., 1982; Pollack et al., 1981).

1.7 The Major Histocompatibility Complex (MHC)

The MHC is a complex of genes which control the defence of the immune system against foreign invaders. It is present in all vertebrates and the function of the MHC encoded molecules is regulation of cell-cell interactions during an immune response to antigen (Bodmer and Bodmer, 1978; Hood et al., 1983). The MHC encodes three main classes of proteins.

Class I and Class II genes each encode highly polymorphic families of cell surface glycoproteins concerned in cell-cell recognition and immune regulation. Class III genes code for three complement components, C4, factor B and C2 as well as steroid 21-OHase.

The MHC antigens were first demonstrated in mice and called H2 antigens (Histocompatibility 2) which map to chromosome 17. In man they occur on the short arm of chromosome 6, in the distal position of 6p21.3 band (Lamm and Olaisen, 1985) and are called HLA antigens. MHC molecules were first identified as major obstacles to tissue transplantation (Gorer, 1937). Later the Class I and Class II molecules were distinguished, with the former performing the "transplantation antigen" function, and the latter the immune response (Ir) gene products (Klein et al., 1981).

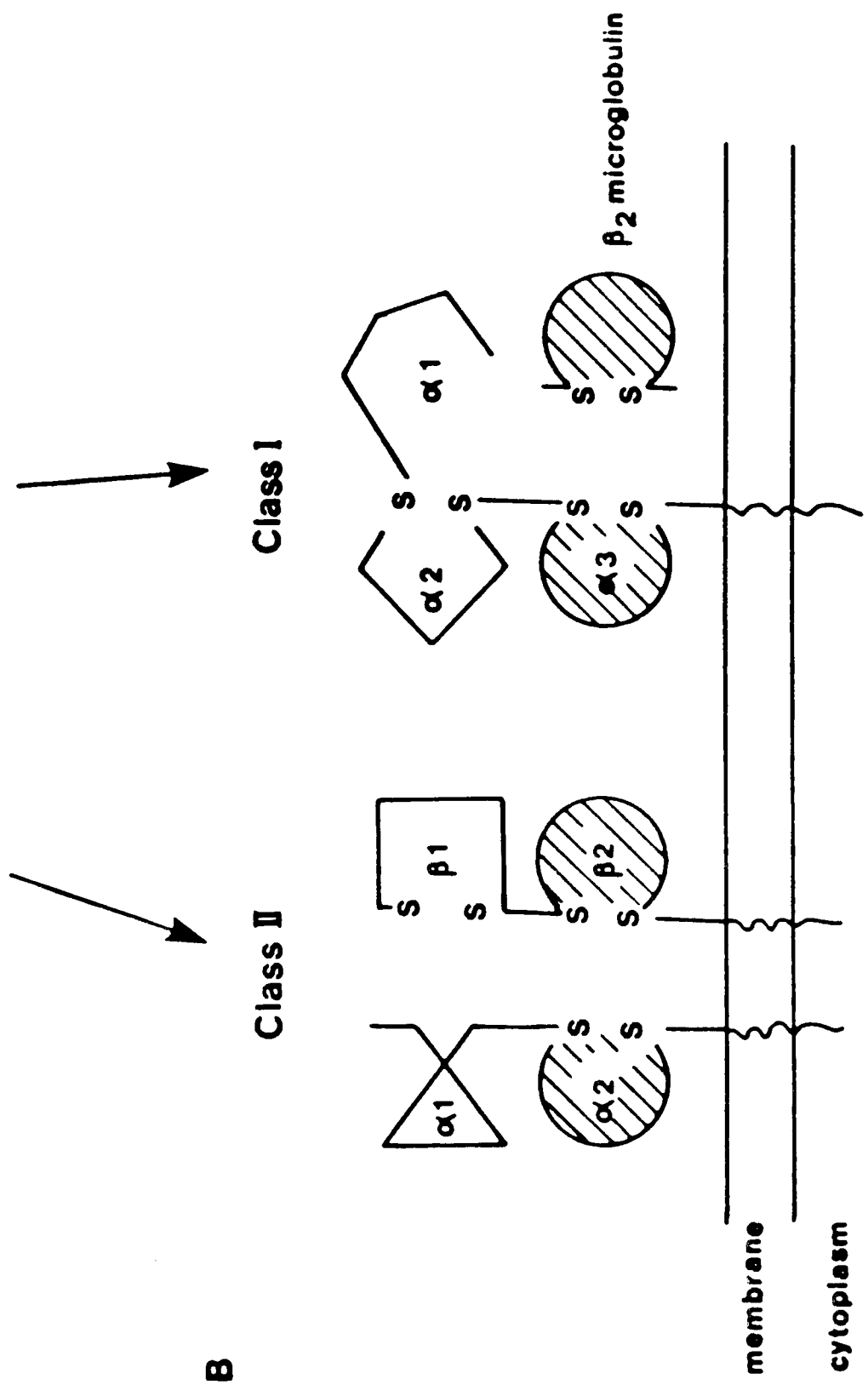
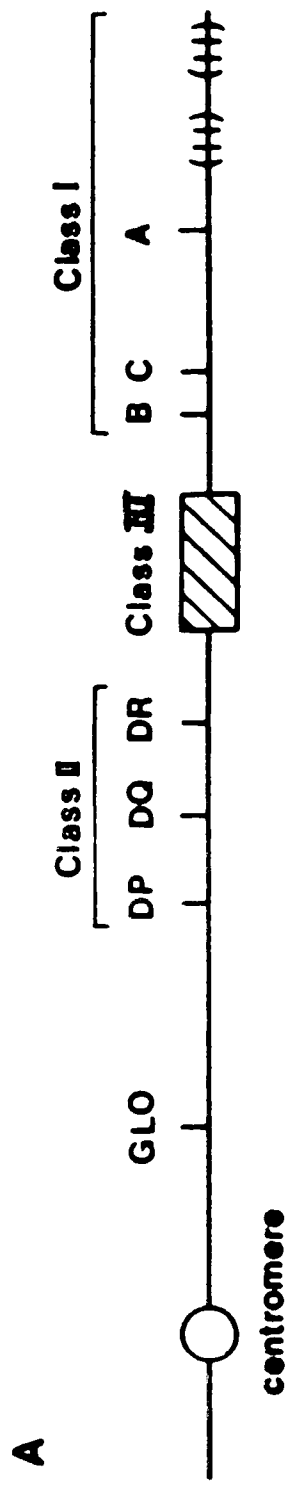
1.7.1 Class I and Class II Antigens

Class I HLA antigens are found on almost all nucleated cells and mediate allogenic killing and T effector cell restriction (reviewed in Strachan, 1987). They are heterodimeric complexes formed by a non-covalent interaction between two polypeptide chains: (1) a polymorphic

Figure 1.5

(A) Shows a portion of the short arm of human chromosome 6 and the location of the class I,II and III genes.

(B) Schematic diagram of the structures of class I and class II molecules. The membrane proximal domains (cross-hatched) shows sequence homology with Ig constant region domains (From Trowsdale, 1987).



heavy chain which is a transmembrane glycoprotein of about 45KDa and is encoded within the HLA complex on chromosome 6, and (2) an invariant light chain, β_2 microglobulin of about 12KDa which is encoded at a single locus on chromosome 15. The α or heavy chain is composed of three external domains termed α_1 , α_2 and α_3 . The α_3 domain, like β_2 microglobulin, resembles the constant domain of immunoglobulins (Figure 1.5). Differential alloantibody recognition has permitted the serological definition of three polymorphic genetic loci which encode the heavy chain, the classical transplantation antigen loci, HLA-A, HLA-B and HLA-C. It is generally concluded from Southern blot analysis and cloning studies that humans have 20 to 40 class I genes (Steinmetz and Hood, 1983).

The exon-intron structure of the class I genes has been established with the available sequence data. Due to the very high level of sequence conservation between non-allelic class I genes, it is possible to describe a generalised organization (Strachan, 1987). There are 8 exons that correlate with the structural domains of the antigens. The first five correspond to a signal peptide, the three extracellular domains α_1 , α_2 and α_3 , and the hydrophobic transmembrane region. However, the cytoplasmic segment is encoded within the exons 6 - 8 for HLA-A2, A3, A24 and Cw3, or exons 6 and 7 for the B-locus genes HLA-B7, B27, and Bw58 in which case exon 8 encodes only 3' untranslated sequence.

The class II antigens are primary determinants in the generation of proliferative responses of lymphocytes in the mixed lymphocyte culture, originally defined as HLA-D determinants. They are heterodimeric proteins like the class I antigens containing four external domains. These domains occur two in each chain, α_1 , α_2 and β_1 , β_2 rather than α_1 , α_2 , α_3 and β_2 as in class I antigens. The two membrane proximal domains α_2 and β_2 , are typical immunoglobulin like domains and are highly

conserved. The two membrane distal domains α_1 and β_1 carry the polymorphic determinants.

Structurally class I and class II antigens may be very similar; both have four domains, of about the same size approximately 90 amino acids each and both contain two immunoglobulin-like constant domains, with a characteristic conserved disulphide bond, next to the cell membrane (Figure 1.5). This feature is shared with a number of other cell surface molecules, belonging to the same supergene family which includes the T cell receptor (Williams, 1985a, 1985b). The spacing of the cysteine residues in this domain is conserved in all of these molecules, as are some of the other key amino acids that presumably contribute to the folding of the immunoglobulin domain (Larhammar *et al.*, 1982a, 1982b; Travers *et al.*, 1984).

The class II antigens are encoded by genes in the HLA-D region. The genes in the HLA-D region are organised into four subregions: DP, DZ/D0, DQ and DR, each subregion containing at least one α/β pair of genes that encode molecules of about 33,000 Mr and 28,000 Mr respectively (Trowsdale, 1987). Another pair of genes DX α and DX β were positioned in the DQ region. Although the DX, D0 and DZ genes seems to be intact and capable of functioning, there is no evidence for the expression of their protein products. In conclusion it appears that the HLA class II region codes for at least six pairs of α and β genes which presumably arose by gene duplication. Despite this complexity there are still only three major expressed products DP, DQ and DR and the contributions of the other genes that have been identified are not yet clear.

1.7.2 Class III Proteins

The first evidence for the existence of complement genes within the HLA complex was provided in 1974 when it was demonstrated in family

studies that the electrophoretic polymorphism of properdin factor B (Bf) segregated with selective HLA haplotypes (Allen, 1974). Further studies of pedigrees of patients with selective deficiencies for the different serum complement components revealed that deficiencies of the second (C2) (Fu et al., 1974) and the fourth (C4) (Rittner et al., 1975; Ochs et al., 1977) complement components were HLA linked. It was then demonstrated that the polymorphic variants of C2 (Meo et al., 1977) and C4 (O'Neill et al., 1978a) were also HLA linked, indicating that the structural genes for C2 and C4 were located within the HLA complex. The gene for 21-OHase was located very near the C4 gene by demonstration that on the HLA-Bw47 haplotype both C4 and 21-OHase genes were affected indicating close proximity (White et al., 1984c, 1985b; Carroll et al., 1985a).

The absence of crossover between the complement genes in the class III region indicated that they were inherited as single haplotypes or "complotypes" (Awdeh et al., 1983). Certain complotypes were found to extend through the HLA-B and D regions and were referred to as "extended" haplotypes or supratypes (Dawkins et al., 1983). One explanation for these "extended" haplotypes being that they were maintained in the population due to a suppression of recombination, as proposed for the murine T/t complex (Awdeh et al., 1983).

1.7.3 Molecular Mapping of the HLA region

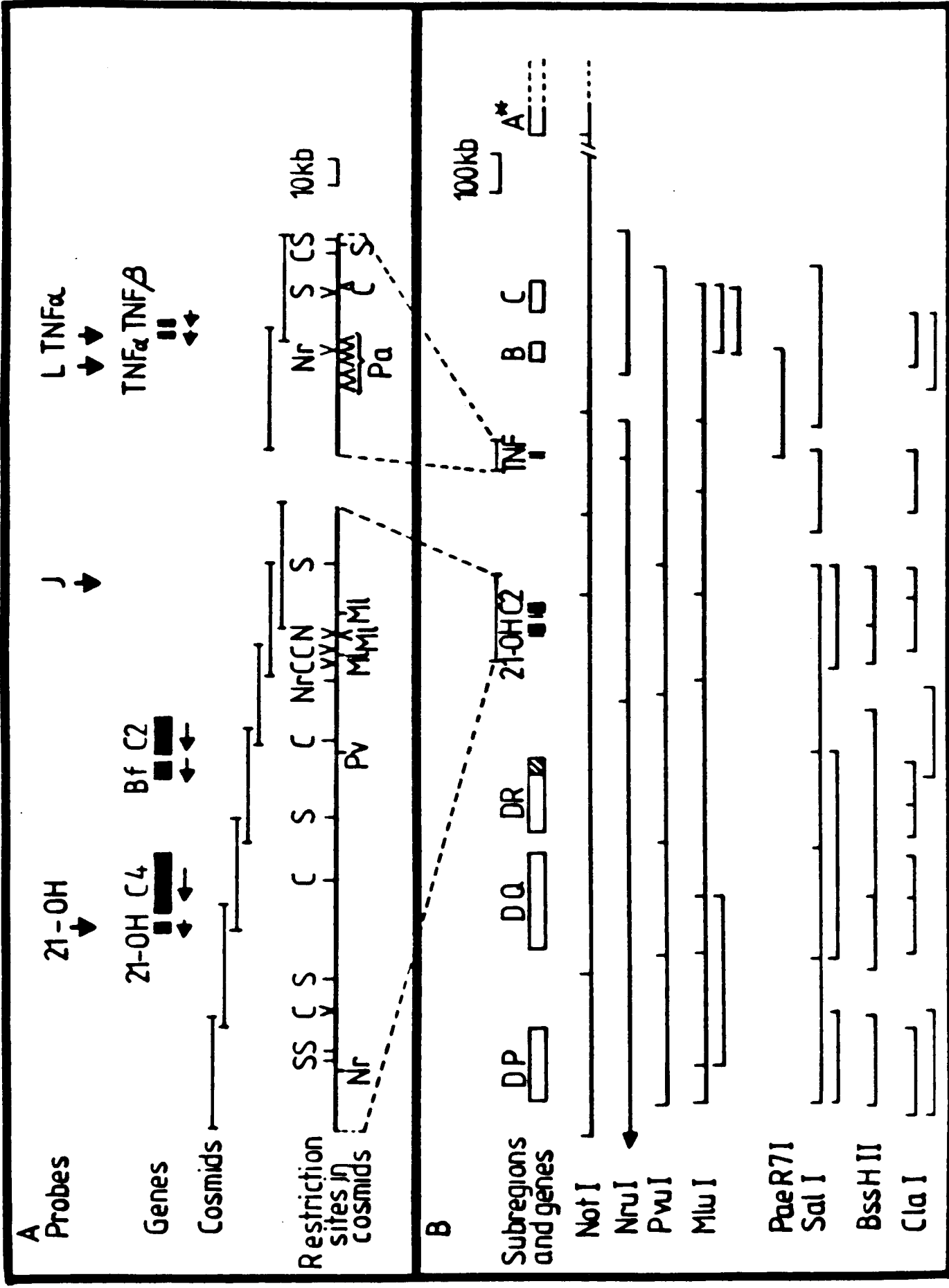
Recent advances in recombinant DNA technology have allowed large regions of both H2 regions (Steinmetz et al., 1982a,b.; Winoto et al., 1983; Muller et al., 1987) and HLA complex (Malissen et al., 1982; Kaufman et al., 1984; Bach, 1985) to be cloned and analysed at the molecular level. Analysis of recombinant HLA haplotypes in family studies had established that the class I loci are telomeric to the class

Figure 1.6

Molecular map of the human MHC.

(A) showing overlapping cosmid clones from complement / 21-0Hase and TNF regions. The 5'- 3' orientation is indicated by horizontal arrows under the genes shown as filled boxes.

(B) Shows the restriction map obtained by Southern blot analysis of genomic DNA separated by PFGE after single and double digests (Dunham et al., 1987).



II genes. Within the class II region the DP subregion maps centromeric to DQ and DR. The class III genes are located between the HLA-DR and HLA-B loci. The molecular map of the class III region linking the complement genes C4, C2 and Factor B was established using overlapping cosmids (Carroll et al., 1984a,b), their order being C2, Factor B, C4A, C4B. The distance between the 3' end of C2 and Factor B was less than 1 kb and the Factor B gene was approximately 30 kb away from the C4A gene. The two C4 genes were separated by 10 kb and 2 kb after each C4 gene was located a 21-OHase gene (Carroll et al., 1985a; White et al., 1985b).

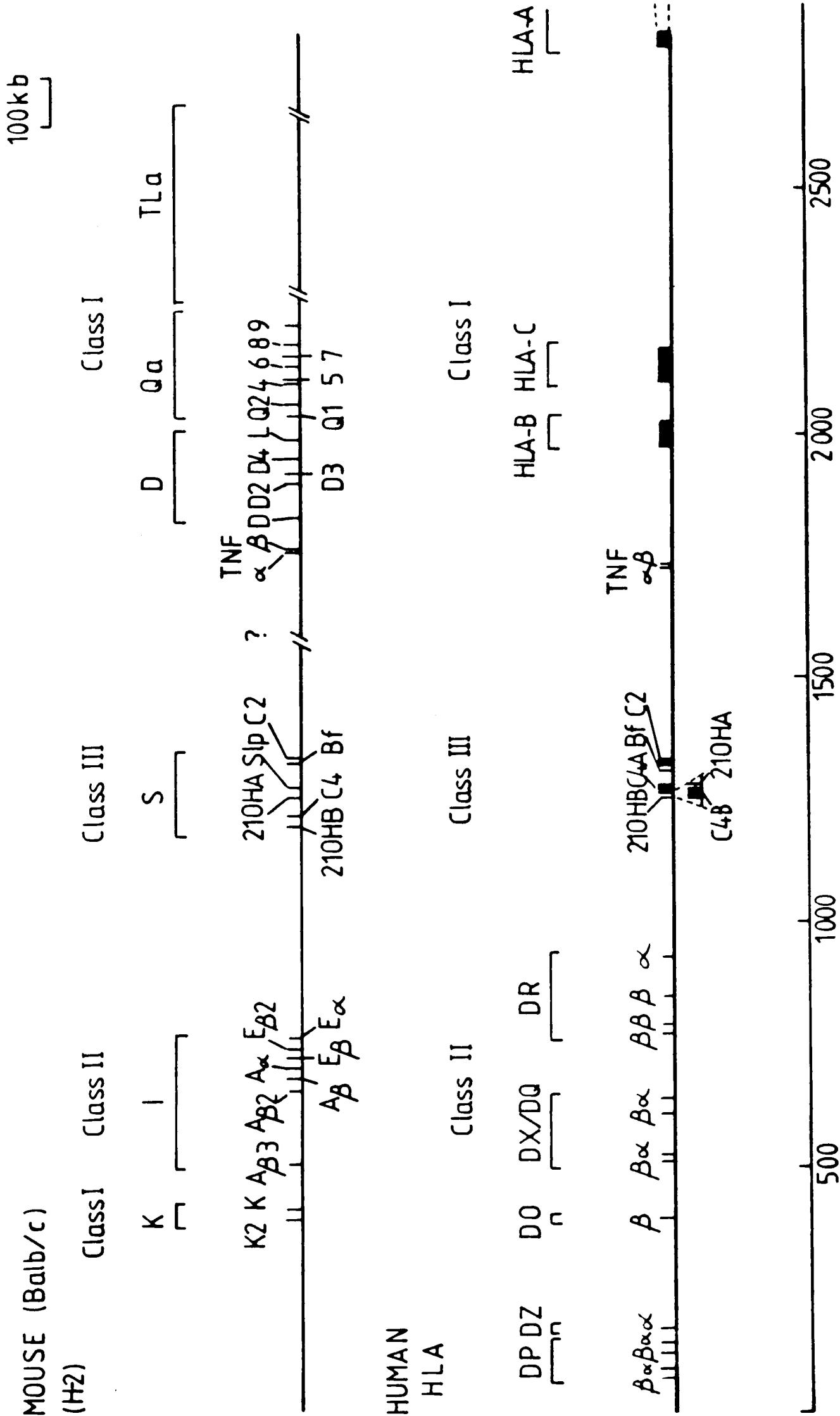
The orientation and distance of the class III genes relative to the class I and class II genes had not been determined due to the absence of informative recombination events within the region. Recently, pulsed field gel electrophoresis and cosmid walking together with restriction enzymes that cut genomic DNA infrequently and Southern blotting have been used to construct a long range genomic restriction map of the MHC and orient the class III gene cluster with respect to the DR subregion (Dunham et al., 1987) (Figure 1.6). The 21-OHase B gene is centromeric to the C2 gene, the distance between the 21-OHase B gene and the DR α locus being 300 - 360 kb, while that between the C2 and HLA-B locus is ~650 kb. The genes for the tumor necrosis factor (TNF α) and lymphotoxin (TNF β) were mapped to the HLA (Spies et al., 1986) and are located between the C2 and HLA-B locus, the TNF α gene lying 390 kb from the C2 gene (Dunham et al., 1987).

The organization of the mouse MHC is very similar to that in man (Muller et al., 1987) (Figure 1.7). The orientation of the class III genes in the S region of the mouse H2 complex is the same as that in man with the 21-OHase B gene being nearer to the class II region and the C2 gene nearer to the class I region. Cloned genomic DNA covering 240 kb have ordered the murine class III genes (Chaplin et al., 1983) with C2

Figure 1.7

Comparison of the molecular maps of the human and murine MHC (Dunham et al., 1987; Muller et al., 1987; Hardy et al., 1987).

COMPARISON OF THE MOLECULAR MAPS OF THE HUMAN AND MURINE MHC



and Factor B genes occupying identical positions as the human genes. The murine Factor B gene is 50 kb away from the two C4-like loci. The two C4-like loci which are separated by about 80 kb encode S1p (Sex limited protein) and C4. S1p is 96% homologous to C4 and shares many structural and biosynthetic properties with C4, but it lacks C4 functional activity. One 21-OHase gene each is located 4 kb after the S1p and C4 genes. There are at least 33 class I genes in the mouse distributed in the K, D, Qa and T1a regions. The class II genes in the I region code for A and E molecules. The TNF_{α} and TNF_{β} genes are closer to the class I region at a distance of about 70 kb than their human counterpart, which is 250 kb away from the HLA-B region. The H2-K locus is centromeric to the H2-I locus and is separated from the other class I genes on the telomeric side of the class II region. Figure 1.7 shows a comparison of the molecular maps of the human and murine MHC.

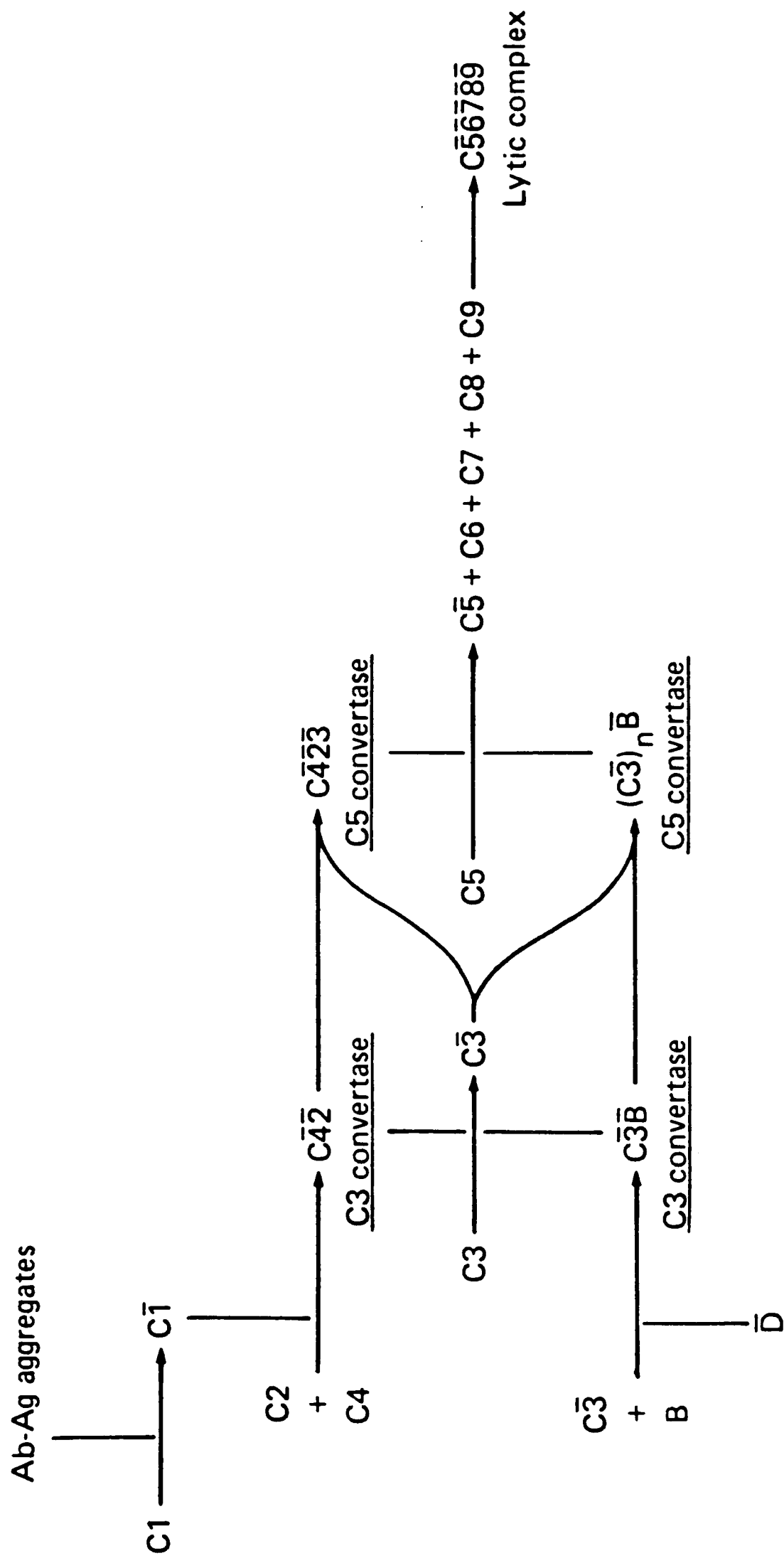
1.7.4 The Complement System

The complement system is a complex system composed of at least 20 plasma glycoproteins and is found in the blood of all vertebrates. The principle biological function of complement is to defend vertebrates against most bacterial infections. Complement can be activated by two separate pathways, the classical and alternative pathways. Both result in the formation of complex proteinases and share in common the terminal or lytic pathway, resulting in lysis of the target cell (Figure 1.8).

In the classical pathway, initiation of activation of complement in blood is caused mainly by formation of antibody-antigen aggregates or by antibody bound to cellular or particulate antigens. Antibody independent activation of this pathway by retroviruses, heart mitochondrial membranes and bacterial lipid A has also been described (Reid &

Figure 1.8

The two pathways of complement activation. The classical pathway is activated by antibody-antigen aggregates. The alternative pathway is activated by antibody-antigen aggregates or bacterial polysaccharides.



Porter, 1981). The alternative pathway too is activated by aggregated antibody, but in addition activation may also be initiated in the absence of antibody by polysaccharides present on the cell envelope of bacteria, yeast and protozoa. It is therefore thought that the alternative pathway provides a first line of defence against infection prior to an immune response.

Figure 1.8 shows the activation scheme of both pathways of complement. In the classical pathway, the first component of complement (C1) binds to antibody-antigen aggregates or antibody bound to cells and is activated through its subcomponents C1q, C1r and C1s. Activated subcomponent C1s is a proteolytic enzyme which hydrolyzes components C2 and C4, yielding the enzyme complex of C4b, C2a comprising the major cleavage fragments of each component. The C4bC2a complex, via its active site in the C2a portion, converts C3 into an activated form that associates with C4bC2a to give a C4bC2aC3 complex that activates C5.

In view of the wide variety of activators of the alternative pathway it is postulated that the activators may all offer 'protected sites' for the deposition of C3b and the formation of a C3b, Bb complex by the action of Factor D. Factor D, a serine proteinase catalyses the conversion of Factor B complexed with C3b and Mg^{2+} to form a C3 convertase C3bBb. The C3 convertase can cleave C3 via the active site in Bb. Any C3b, Bb complex in a 'protected site' would then be expected to avoid rapid inactivation by the control protein Factor I and its cofactor H. In this manner, a high turnover of C3 can be achieved and the activating particle could become coated with many molecules of C3b derived from freshly activated C3. The product of C3 convertase, C3b, then associates with C3bBb to form a C5 convertase $(C3b)_n Bb$. Thus activation of the alternative pathway can be viewed as a disruption of the inhibitory effects of Factors H and I (Reid, 1986). In both pathways the activa-

tion of C5 is the last proteolytic step. Activated $C5\bar{b}$ remains bound to the C5 convertase and has the transient capacity to bind C6 and then C7 to form $C5\bar{b}7$. This complex binds firmly to the membrane close to the site where complement activation was initiated. The addition of $C8$ and $C9$ to this complex are necessary to form the lytic complex which make lesions in plasma membranes, causing the cells to burst (Porter, 1979).

The control proteins of the complement pathways include C1 inhibitor, C4b binding protein (C4b-BP), Factor I, properdin, Factor H, complement receptor type 1 (CR1) and S protein. These play an important role in the control of the classical and alternative pathway complexes to prevent overactivation of the complement system and hence damage to healthy cells.

As mentioned in Section 1.7.2, three of the complement proteins are encoded by genes in the class III region of the MHC. These are discussed in more detail below.

1.7.5 Class III Genes

a) C2

C2 is a single chain glycoprotein of Mr 100,000. In the classical pathway it is cleaved by $C1s$ to yield two non-covalently linked fragments. The C terminal fragment $C2\bar{a}$ of Mr 70,000 contains the catalytic site of the classical pathway C3 and C5 convertases. It is a serine proteinase with a catalytic chain twice the size of other serine proteinases, but with sequence homology around the active site serine residue (Gagnon, 1984). Amino acid sequence analysis of C2 (Kerr and Gagnon, 1982) and nucleotide sequence data (Bentley and Porter, 1984; Bentley, 1986) show that C2 is composed of 732 amino acids, and is encoded by a mRNA species of ~ 2.9 kb.

Like all glycoproteins coded for by the MHC, C2 demonstrates

genetic polymorphism, but to a lesser extent than most other MHC products. This polymorphism in C2 has been defined by isoelectric focusing though most individuals are homozygous for the common variant C2C which has a gene frequency of 0.97. In addition two rare variants, the acidic C2A variant (gene frequency < 1%) and the basic C2B variant (gene frequency 2%) have been described (Alper, 1976). No sequence information is available on the different protein variants.

Deficiency of C2 is the most common of the genetic deficiencies of complement proteins found in man (Agnello, 1978). The disorder is inherited as an autosomal codominant trait found in linkage disequilibrium with certain HLA haplotypes. Recently it was shown by Southern blot analysis that the deficiency was not due to a major gene deletion or rearrangement (Cole et al., 1985). The absence of detectable C2 mRNA in peripheral blood monocytes from C2-deficient individuals was taken to suggest that C2 deficiency results from a defect in transcriptional processing of C2 mRNA.

b) Factor B

Factor B (Bf) is a single chain glycoprotein of Mr 90,000. It is activated by Factor D in the alternative pathway of activation. The C terminal 55 fragment of Mr 60,000 contains the catalytic site of the complex convertases and shows a clear homology with other serine proteinases. The complete amino acid sequence of the zymogen has been determined by a combination of protein (Christie and Gagnon, 1983) and nucleotide sequencing (Morley and Campbell, 1984). Factor B variants are detected by agarose gel electrophoresis followed by immunofixation (Alper et al., 1972). In Caucasians there are four allelic forms at a single autosomal locus; two common variants F and S, and two rare alleles F₁ and S₁. In addition up to 14 very rare variants have also

been described in different studies in different ethnic groups (Mauff et al., 1978; Dykes et al., 1983).

Electrophoretic analysis of the cleavage products of Factor B, Ba and Bb, has suggested that the charge difference between the F and S variants resides in the Ba fragment while those defining F₁ and S₁ and some of the rare variants reside in the Bb fragment (Alper et al., 1972; Mauff et al., 1978). A comparison of the Ba fragment of the gene from both F and S alleles revealed two single point mutations in the coding region, one of which one was silent while the other resulted in an amino acid change from arginine in the S allele to glutamine in the F allele (Campbell, 1987). This would explain the difference in electrophoretic mobility of the two variants with the F allele carrying less positive charge and thus moving more toward the anode.

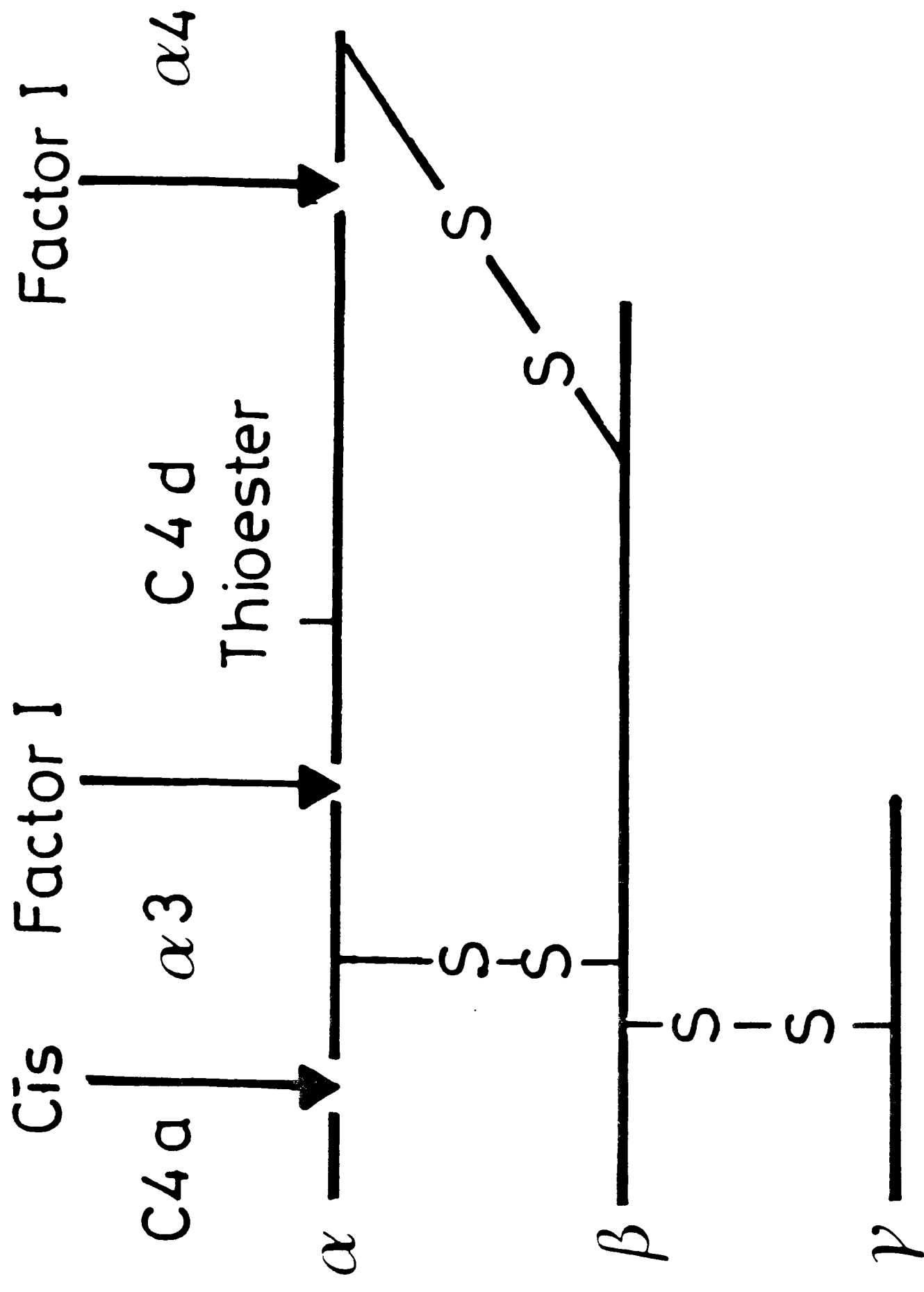
The human Factor B gene structure has been completely determined. It is 6 kb in length and is split into 18 exons (Campbell et al., 1984). The active site residues (His, Asp, Ser) of Factor B are all encoded in separate exons. The 5' end of the Factor B gene lies only 421 nucleotides away from the 3' end of the C2 gene (Wu et al., 1987). The close homology of structure and function, and the chromosomal location of C2 and Factor B suggests that they evolved from a common locus, probably by tandem duplication of the DNA. The C2 gene is 18 kb in length, about 3 times the length of the Factor B gene, though the mRNA species encoding C2 and Factor B are of similar sizes (~2.9 kb and ~2.6 kb). The difference in size of the two genes must therefore be mainly due to the length of the non-coding sequences.

c) C4

Activation of the complement system by the classical pathway is dependent on the presence in the serum of the fourth component (C4).

Figure 1.9

Structure of the C4 molecule. C4 is composed of three disulphide linked polypeptide chains α , β , γ , of Mr 95,000, 70,000 and 33,000 respectively. In the classical pathway of activation C4 is activated by C1s which cleaves the α chain to release C4a. Activated C4 can bind covalently to a variety of surfaces. C4b is inactivated by factor I with C4b-BP as cofactor which cleaves the α' chain in two places to release C4d. The origin of fragments α_2 (C4d), α_3 and α_4 is indicated.



C4 is composed of 3 disulphide linked polypeptide chains α, β, γ of 95,000, 70,000 and 33,000 respectively. It is synthesized as a single chain precursor pro C4 of Mr 200,000 (Hall and Colten, 1977; Gigli, 1978) aligned in the order β - α - γ (Chan and Atkinson, 1983). Prior to secretion, the pro C4 is glycosylated, an internal thiolester bond is formed and the molecule is split to give the major plasma form of the three disulphide linked chains. The internal thiolester on cleavage is transiently capable of covalently acylating hydroxyl, amino and other nucleophilic groups of proteins, carbohydrates and other molecules. In serum, C4 is activated by C1s to produce C4a and C4b which interacts with C2a to form the classical pathway C3 convertase. C4 is inactivated by a second cleavage of the α chain by Factor I with C4b-BP as cofactor. This releases C4d, a peptide of Mr 44,000, from the middle of the α ' chain (see Figure 1.9). Thus C4 is both activated and inactivated by cleavage of the α chain.

C4 was originally mapped to the HLA region from studies of families with hereditary C4 deficiency (Rittner et al., 1975). Electrophoresis identified two variants C4F and C4S as products of two distinct non-allelic loci C4A and C4B (O'Neill et al., 1978a; Olaisen et al., 1979). The genetic studies were complicated by the relatively high (10-15%) gene frequency of null alleles (O'Neill et al., 1978a) and the frequency of duplications (1-2%) at each locus (Raum et al., 1984). Null alleles (Q0) represent absence of detectable levels of C4 protein in the serum and have been linked to susceptibility to SLE (Fielder et al., 1983). Analysis of individuals who had absence of C4A or C4B protein in the serum showed that only 50% of the C4 null alleles are due to deletion of the gene on certain haplotypes (Carroll et al., 1985b). Duplication of either of the C4 locus can occur and this has been confirmed by genomic analysis where three C4 genes were present on one chromosome (Carroll et

al., 1984b).

C4 is the most polymorphic of the class III proteins, comparable with the class I and class II antigens. Both C4 loci show large variation and electrophoresis of neuraminidase treated serum has identified at least 13 C4A alleles and 22 C4B alleles (Mauff et al., 1983). More than 40 alleles have been demonstrated in the two classes since, by differences in electrophoretic mobilities, serological typing (Roos et al., 1984; Giles et al., 1984; Giles, 1987) and direct DNA sequencing (Belt et al., 1984, 1985; Yu et al., 1986). MHC linkage of C4 was also demonstrated when the Chido and Rodgers red cell antigens, which were previously mapped to HLA, were shown to be fragments of the C4 α chain (C4d) (Middleton et al., 1974; Giles et al., 1976; O'Neill et al., 1978b; Tilley et al., 1978). In an haemolytic overlay assay of C4 protein after separation by charge gel electrophoresis (Awdeh and Alper, 1980) C4A is several fold less active than C4B. The difference in haemolytic activity is due to a difference in covalent binding efficiency of C4A and C4B to the antibody coated red blood cell (Law et al., 1984; Isenman and Young, 1984). Results from these studies showed that the C4A protein binds more efficiently to protein antigen and the C4B protein bind more efficiently to carbohydrate antigen.

Using protein (Hellman et al., 1984), cDNA and genomic sequence data (Belt et al., 1984; 1985) it was possible to determine the amino acid differences between C4A and C4B. Out of more than 4,600 nucleotides compared only 14 differences were observed. Of the 14 substitutions, 11 resulted in codon changes and nine of the changes were clustered in the C4d region of the α chain (Belt, et al., 1984). Each of the other two codon changes were located either in the β or the γ subunits. Six of the nine amino acid substitutions, which were derived from the nucleotide sequence, were proposed as isotypic differences

based on the sequence results of Hellman et al., (1984) and on the analysis of an additional four clones that represented C4A or C4B (Belt, et al., 1984, 1985). Recent work has now established that in fact four isotypic amino acid differences in the C4d region of C4A and C4B are the cause of their different chemical reactivities (Yung et al., 1986).

C4 cDNA clones were used as probes to screen human genomic libraries. Clones were isolated and on the basis of hybridization to synthetic oligonucleotides specific for either C4A and C4B, two genes were identified (Carroll et al., 1984a, 1984b). Results indicated that direction of transcription of the two genes was the same; however the C4A gene and some C4B genes are 22 Kb, about 7 kb longer than the remaining C4B genes, due to a 6-7 kb long intron near the 5' end of gene. The near identity of the two genes suggests either that they were duplicated very recently or that there is a mechanism maintaining identity.

d) 21-OHase

As mentioned previously HLA-Bw47 is strongly associated with 21-OHase deficiency. In reviewing the parental HLA haplotypes of patients with the different 21-OHase deficiency syndromes, none of the unaffected haplotypes carry Bw47. It has been suggested that the Bw47 antigen could be a variant HLA-B antigen which only occurs together with 21-OHase deficiency except for the rare cases where Bw47 and the 21-OHase deficiency allele have been separated by genetic recombination. The Bw47 determinant is in positive genetic linkage disequilibrium with A3, Cw6, and DR7 and commonly occurs on the HLA haplotype A3 Cw6 Bw47 C4A1 C4BQ0 BfF C2C DR7 (Fleishnick et al., 1983; O'Neill et al., 1982). This haplotype is very similar to the more common HLA haplotype A3 Cw6 B13 C4A3 C4B1 BfF C2C DR7. The two HLA-B antigens B13 and Bw47 both

carry the public HLA-B determinant Bw4 and both have considerable serological cross-reactivity (Reekers et al., 1984). It was postulated that the Bw47 antigen was the product of a mutated B13 gene which had been modified to express the unique Bw47 determinants as a result of deletion or rearrangement of the adjacent 21-OHase locus (White et al., 1985a). It was suspected that the C4 gene which was postulated to have been changed from a normal gene to a C4 null allele (O'Neill et al., 1982) and the 21-OHase deficiency were caused by a single deletion of adjacent genes (White et al., 1984b) and that comparison of normal DNA and DNA carrying the Bw47 haplotype would allow identification of the deletion. A bovine cDNA clone encoding part of P-450_{C21} (White et al., 1984a) was used to study individuals homozygous for HLA-B13 or HLABw47 (White et al., 1984b).

The DNA was subjected to digestion with various restriction endonucleases. The probe hybridized to two fragments each in DNA from HLA-B13 cell line digested with EcoRI (12 kb and 14 kb) and TaqI (3.7 kb and 3.2 kb). The 12 kb EcoRI and 3.7 kb TaqI band was absent in the digests of DNA from the 21-OHase deficient HLA-Bw47 homozygous cell-line. In DNA from individuals heterozygous for HLA-Bw47, the 3.2 kb TaqI band was twice as intense as the 3.7 kb band. This suggested the deletion of one of two genes but it did not rule out a partial deletion of a single gene or some other type of DNA rearrangement. Analysis of DNA from different patients with CAH suggested that when 21-OHase deficiency is associated with HLA-Bw47, there is a deletion of a 21-OHase structural gene. With other HLA-B antigens, there is occasionally a deletion of at least part of a gene.

As a 21-OHase gene and complement C4 gene were both affected on the Bw47, DR7 haplotype it was likely that these genes were located very near each other on the chromosome. Using the bovine cDNA probe White et

al., (1984c) identified two genes encoding 21-OHase on cosmid clones from a murine genomic library containing S1p and C4. The two 21-OHase genes were adjacent to S1p and C4. In man, a steroid 21-OHase gene was positioned within 2 kb 3' to both C4A and C4B (Carroll et al., 1985a; White et al., 1985b). Carroll et al., (1985a) identified a 2.4 kb fraction of human adrenal RNA that hybridized to a probe prepared from a cosmid clone that included part of the human C4 gene in the Northern blot analysis. The identity of the gene was demonstrated by hybridizing with a 23 base synthetic oligonucleotide complementary to bovine 21-OHase cDNA. Nucleotide sequence of a 0.43 kb fragment that hybridized to the oligonucleotide confirmed that the gene encoded 21-OHase on the basis of its sequence homology with bovine 21-OHase. White et al. (1985b) mapped the two 21-OHase genes on human cosmid clones using the bovine cDNA probe. Both 21-OHase genes named 21-OHase A and 21-OHase B, adjacent to C4A and C4B, respectively, were approximately 4-6 kb in length, similar by restriction map analysis and oriented in the same direction as the two C4 genes with respect to transcription (Carroll et al., 1985a; White et al., 1985b).

As the arrangement of genes in man and mouse is essentially the same, this suggests that the C4 and 21-OHase genes became linked in the major histocompatibility complex before man and mouse diverged during evolution. However, sequence comparison of the two C4 or two 21-OHase genes in each species show that the members of each pair are more homologous to each other than they are to the corresponding genes in the other species (Belt et al., 1985; Chaplin et al., 1985,1986; Rodrigues et al., 1987). This suggests that the duplication in man and mouse are independent, comparatively recent events.

Restriction analysis of the cosmid clones demonstrated that the 21-OHase B gene included the 3.7 kb TaqI fragment and that the 21-OHase

A gene included the 3.2 kb TaqI fragment (White et al., 1985b). Since the deletion of the 3.7 kb TaqI fragment results in the salt-wasting 21-OHase deficiency, it is evident that the 21-OHase B gene is important for steroid biosynthesis in the human adrenal. While it was unclear if the 21-OHase A gene was expressed, deletion of the gene was not linked to CAH. Genetic studies have established a negative correlation between CAH and the HLA haplotype A1 B8 DR3 C4AQ0 C4BQ0 BfS C2C (New et al., 1983). This haplotype has been shown to have a deletion that includes the 21-OHase A gene (White et al., 1985b; Carroll et al., 1985b). DNA from hormonally normal individuals who were homozygous for A1 B8 DR3 showed that the 3.7 kb TaqI band was present at normal intensity, but the 3.2 kb band was absent. These data suggest that the 21-OHase B gene is functional, but the 21-OHase A gene is not (White et al., 1985b).

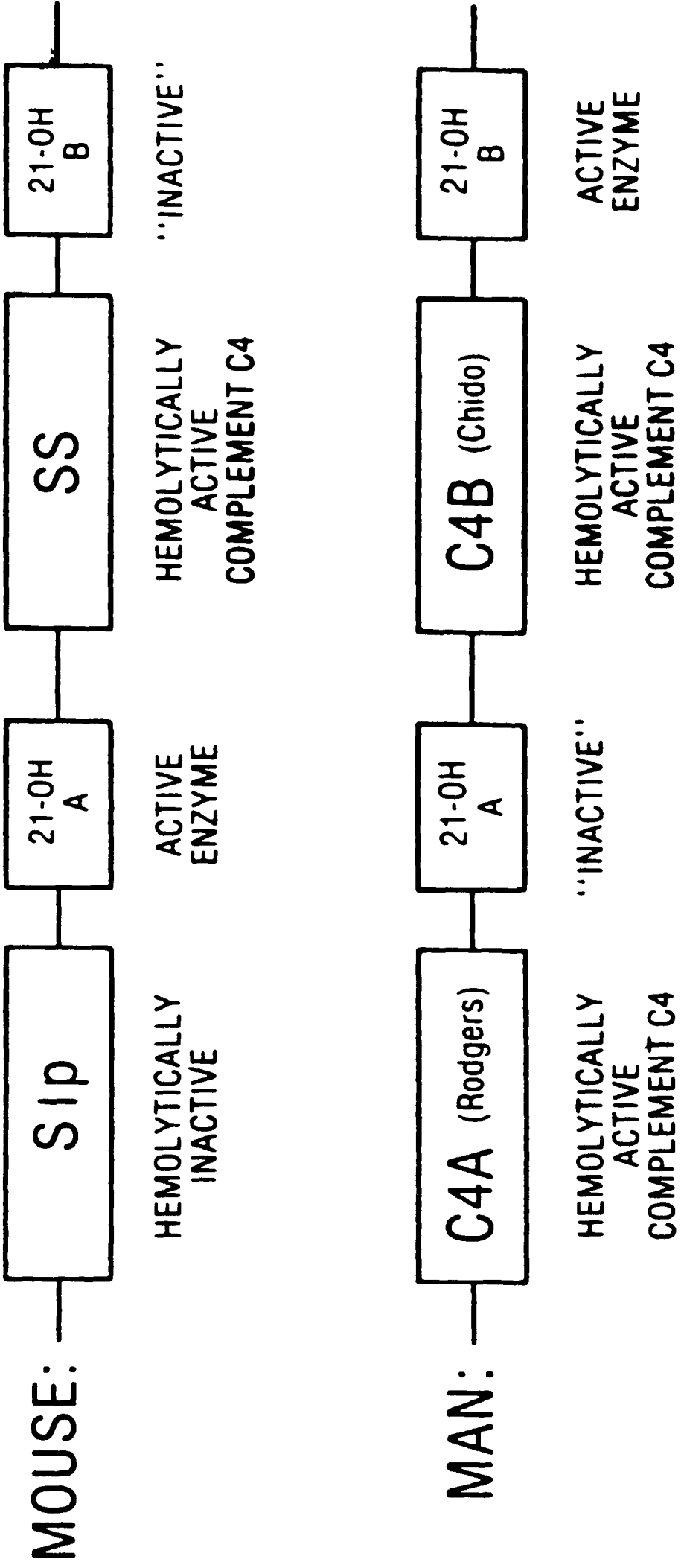
Levels of transcription of the 21-OHase genes in mouse showed that the 21-OHase A gene appears to be the active gene rather than the B gene and this has been confirmed by nucleotide sequencing data (Chaplin et al., 1986). Figure 1.10 shows a schematic comparison of the C4 and 21-OHase genes of man and mouse.

Cloning and nucleotide sequencing of the two human 21-OHase genes has established that the 21-OHase A gene is indeed a pseudogene due to mutations in its coding regions and that only the 21-OHase B gene is the active gene (White et al., 1986; Higashi et al., 1986; Rodrigues et al., 1987). CAH due to 21-OHase deficiency may be caused due to deletion of the 21-OHase B gene or deleterious mutations which may lead to the loss of functional activity of the enzyme.

In a recent study it was reported that in the majority of patients with classical CAH, the defect was not due to gene deletion (Rumsby et al., 1986). The 21-OHase B gene was present on at least one of the chromosomes, but obviously nonfunctional, or the expressed product partly

Figure 1.10

Schematic comparison of the C4 and 21-OHase genes of man and mouse. In mouse the 21-OHase B gene is nonfunctional due to a deletion of 215 bp, while in man the 21-OHase A gene is nonfunctional due to deleterious mutations. While both C4 genes in man produce hemolytically active C4, the Slp gene in the mouse produces a hemolytically inactive homolog expressed only in males of certain strains.



or wholly inactive.

Due to the high incidence of this serious inborn error of metabolism it is clearly essential to define molecular markers for the presence of particular disease genes which can be used in prenatal diagnosis and treatment. In order to achieve this goal and to understand the molecular basis of CAH, the present study on 21-OHase genes began in Oxford in 1984. With the use of recombinant DNA techniques, normal and defective 21-OHase genes were studied with the hope that nucleotide sequences would give a clue to the defects observed in 21-OHase deficient patients and would generate DNA markers which would be of use in determining affected haplotypes.

CHAPTER II

MATERIALS & METHODS

2.1 Materials

(a) Enzymes

All enzymes were from Amersham, BRL and New England Biolabs with the exceptions of Lysozyme and RNase which were from Sigma; DNA polymerase I (Klenow fragment) from Pharmacia; T4 polynucleotide kinase and calf intestine alkaline phosphatase from Boehringer; Proteinase K from BDH and Pronase from BCL.

(b) Chemical Reagents

Chemical reagents were obtained from BDH, Sigma, Fisons and Boehringer Mannheim except for:

Agar No. 1 and Yeast extract	-Oxoid
Agarose	-Seakem and Pharmacia
Bactotryptone	-Difco
Dextran Sulphate and Ficoll	-Pharmacia
DE52	-Whatman
β -Mercaptoethanol and Guanidine Thiocyanate	-Fluka
Nitrocellulose	-Amersham
GeneScreen Plus	-Dupont
Sephadex G-100	-Pharmacia

(c) Radioactive Nucleotides

Radioactive Nucleotides were supplied by the Radiochemical Centre, Amersham.

$[\alpha^{32}\text{P}]$ - dNTPs were 10mCi/ml, specific activity 3,000 Ci/mmol.

$[\alpha^{35}\text{S}]$ - dATP was 8 mCi/ml, specific activity 1,000 Ci/mmol.

(d) Bacterial Strains, Plasmids, Phages and Cell lines.

<u>Eschericia coli Strains</u>	MC1061 (Casadaban and Cohen, 1980)
	JM103 (Messing <u>et al.</u> , 1981)
	TG1 (A gift from T. Gibson)
	Q358 (Karn <u>et al.</u> , 1980)
	C600 (Appleyard, R.K., 1954)
Plasmid pAT 153/PvuII/8	(Anson <u>et al.</u> , 1984)
Phage M13 mp9	(Messing and Vieira, 1982)
λ EMBL-3	(Frischauf <u>et al.</u> , 1983)
λ gt10	(Young and Davis, 1983)
Mouse Y1 adrenocortical cell line	(Yasumura <u>et al.</u> , 1982)
Cosmids pTCF	(Grosveld <u>et al.</u> , 1982)

(e) Media

2X TY liquid medium	:15g bactotryptone, 10g yeast extract, 5g NaCl to 1 litre with water. pH adjusted to 7.4 with 2ml 4M NaOH.
L broth liquid medium	:10g bactotryptone, 5g yeast extract, 10g NaCl to 1 litre with water.
LA broth	:Contains 50 μ g/ml ampicillin which is prepared as a 20mg/ml stock solution in 1M NaHCO ₃ .

To make solid media 14g agar/agarose was added per litre of liquid medium. 7g agar/agarose was added per litre of top medium.

Ampicillin to final concentration of 50 μ g/ml (liquid medium) and 100 μ g/ml (solid medium) was added when required. DMEM, fetal calf serum, antibiotics and trypsin were purchased from Sterilin or Flow laboratories.

(f) Buffers

1. Hogness freezing buffer: 4% v/v glycerol / 3.6 mM dipotassium hydrogen phosphate / 1.3 mM potassium dihydrogen phosphate / 2 mM sodium citrate / 1 mM MgSO_4 pH 7.5.
2. Lysis buffer (for plasmid prep) : 50 mM glucose / 10mM EDTA / 25mM Tris-HCl, pH 8.0.
3. SM buffer : 100 mM NaCl / 8mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ / 50 mM Tris-HCl pH 7.5 / 2% w/v gelatin.
4. 10 x Nick translation buffer : 500 mM Tris-HCl pH 7.8 / 70 mM MgCl_2 / 10mM β -mercaptoethanol.
5. 10 x TBE : 900 mM Tris-HCl pH 8.0 / 900 mM boric acid / 25 mM EDTA.
6. 40% bis acrylamide : 38g acrylamide / 2g N, N' - methylene bis-acrylamide to 100 ml with water and deionized by stirring for 30 min at RT with 5g "Amberlite" MB-3 monobed resin followed by filtration and storage in a brown stoppered bottle at 4°C.
7. Hybridization buffer : 50% deionized formamide 1 M NaCl / 50 mM Tris-HCl, pH 7.4 / 10% w/v dextran sulphate / 0.2% w/v Ficoll / 0.2% w/v BSA / 0.2% w/v polyvinyl pyrrolidone / 0.1% w/v sodium pyrophosphate / 0.1% SDS.
8. 20 x SSC : 3M NaCl / 300 mM trisodium citrate.
9. 4M Guanidinium thiocyanate solution : 3.5ml of β -mercaptoethanol (14M) were added to 236.32g of GuScN (BDH) and made up to 500ml with sterile water in an autoclaved bottle.
10. Cesium chloride solution : 96g CsCl , 2ml 0.5M EDTA, 0.735g triNa citrate were made up to 100ml with sterile water in sterile bottle and then filter sterilised.

2.2 Overnight Cultures

Bacteria from a frozen glycerol stock or a single colony from a plate were used to inoculate 1-50 ml liquid media and incubated at 37°C overnight (O/N). E. coli JM103 and TG1 were grown in 2XTY broth. E. coli Q358 and C600 were grown in T broth / 0.2% maltose / 10 mM MgSO₄ and enriched with the same just prior to use.

2.3 Preparation of DNA

2.3.1 Small Scale Isolation of Plasmid DNA (10-100ml)

(Birnboim and Doly, 1979)

Cultures of E. coli were grown in L-broth containing 100µg / ml ampicillin to an A₆₀₀ 0.8 OD units. One ml was removed and centrifuged in a microfuge for 5 min at room temperature (RT). The pelleted cells were resuspended in 180 µl L-broth + 20µl Hogness freezing buffer (Section 2.1 f) and stored at -70°C, as stocks. Chloramphenicol was added at a final concentration of 200 µl / ml to the remainder of the culture, (Clewell and Helsinki, 1972) and the culture incubated at 37°C O/N.

The cells were pelleted at 2,000 rpm for 20 min at 4°C and then resuspended in lysis buffer (200 µl per 10 ml culture) (Section 2.1 f) containing 5mg / ml lysozyme and left on ice for 30 min. Two volumes of 0.2 M NaOH / 1% SDS was added and incubated on ice for 5 min. This was followed by one and a half volumes of 3M NaOAc pH 4.8 which was incubated for further 30 min on ice. This solution was then centrifuged at 11,000 rpm, the clear supernatant transferred to a clean tube and the plasmid DNA precipitated using 1/3 of the total volume of isopropanol. The DNA was pelleted after precipitation at 11,000 rpm for 10 min and resuspended in 0.3 M NaOAc pH 7.4 (300µl / 10ml starting culture).

This was extracted using an equal volume of phenol / chloroform (1:1) and the DNA was again precipitated after the extraction by the addition of 2.5 volumes of absolute ethanol. The pellet was washed twice with 70% ethanol, dried and resuspended in 10mM Tris-HCl / 0.1mM EDTA, pH 8.0 (20µl/10ml starting culture). The plasmid DNA was incubated at 37°C with heat inactivated pancreatic RNase (20µg / ml), and then stored at -20°C.

2.3.2 Large Scale isolation of Plasmid DNA (500-1000ml)

(Radloff et al., 1967; Maniatis et al., 1982).

Cultures of E. coli were grown in L-broth containing ampicillin (50 µg /ml) and treated with chlor amphenicol (Section 2.3.1). After the overnight incubation the cells were harvested and lysed as described previously. The DNA was precipitated with isopropanol and the phenol / chloroform extraction step was omitted. The DNA pellet was washed, dried and resuspended in 15ml of 150mM NaCl / 50mM Tris-HCl / 5mM EDTA, pH 8.0. Ethidium bromide was added to a final concentration of 0.02% and the solution was weighed before adding an equal weight of cesium chloride. The solution was transferred to a 38ml Beckman Quick-Seal ultracentrifuge tube, topped with liquid paraffin, sealed and centrifuged for 17 hr at 35,000 rpm at 18°C in Ti60 rotor.

The centrifuge was stopped without using a brake and the solution transferred from the large RNA pellet to a clean centrifuge tube. This was capped as before and respun for 60 hr at 35,000 rpm at 18°C. Two bands of DNA were visible in the gradient. The lower band of closed circular DNA was collected by puncturing the tube with a needle just below the band and collecting the band as it dripped out through the sterile needle. Ethidium bromide was removed by repeated extractions with ^{equal} volumes of 1-butanol. The DNA was recovered by addition of 3

volumes water and 2 volumes of ethanol. The DNA was reprecipitated once, washed, dried and resuspended in 10mM Tris-HCl / 0.1 mM EDTA, pH 8.0 and stored at 4°C. The DNA concentration was estimated assuming that $A_{260} = 1$ when DNA concentration is 40µg / ml.

[More recent large scale DNA preparations use the programmed Beckman L8 ultracentrifuge to relax the gradient spin after 16 hr (Section 2.3.3) to 45,000 rpm for 45 min. There is no necessity of a respin and the two bands of DNA are very distinct]

2.3.3 Isolation of the Replicative form of Bacteriophage M13 mp9

Ten ml of an overnight culture of E. coli JM103 or TG1 was used to inoculate 1 l of 2XTY liquid broth and incubated at 37°C until $A_{600} = 0.6$ OD units. Ten ml of an overnight culture of E. coli JM103 /TG1, infected with a single viral plaque of bacteriophage M13 mp9, were added to the culture and incubated at 37°C for a further 1½ - 2 hr ($A_{600} \sim 1.0$ OD unit). The cells were then harvested and lysed as described (Section 2.3.1). The bacteriophage DNA was purified by centrifugation to equilibrium in CsCl gradients prepared as described (Section 2.3.2) but spun in the Beckman L8 ultracentrifuge at 55,000 for 16 hr and then at 45,000 for 1 hr. After butanol extraction and ethanol precipitation the DNA was resuspended in 10mM Tris-HCl / 0.1mM EDTA, pH 8.0 and stored at 4°C.

2.3.4 Isolation of Bacteriophage λ DNA by the Plate lysate Method

(Maniatis et al., 1982)

Approximately 10^5 plaque forming units (pfu) of bacteriophage were mixed with 2ml of E. coli Q358 and incubated at 37°C for 30 min. This was then mixed with 9ml of molten T-top agarose at 48°C and poured onto 14cm diameter T-broth agarose plates and incubated at 37°C to give confluent lysis. Twelve ml SM buffer (Section 2.1.f) was added to the

plates and the plates were left at 4°C for atleast 8 hr. Bacterial debris were removed by centrifugation in the Sorvall RC-5B at 8,000 rpm for 10 min at 4°C using SS34 rotor. Pancreatic DNase I and RNase A were added to the supernatant, each at a final concentration of 10µg / ml, and incubated for one hr at RT. Solid NaCl was added to a final concentration of 1M and left on ice for 1 hr at 4°C, after which the sample was centrifuged twice at 8,000 rpm for 20 min. Solid polyethylene glycol (PEG) was added to the clear supernatant to a final concentration of 10% w/v, and dissolved by stirring at 4°C, then left on ice for atleast 3 hr. The bacteriophage was pelleted by centrifugation at 8,000 rpm for 20 min at 4°C. The pellet was resuspended in 2ml SM buffer and spun at 2,000 rpm for 15 min. The supernatant was transferred to a fresh tube and the pellet was extracted with 0.5ml SM buffer. After another spin, the supernatants were pooled and centrifuged for 2½ hr at 25,000 rpm in the Beckman L8 ultracentrifuge. The pellet was resuspended in 1ml SM buffer by leaving on ice O/N and transferred to a 1.5ml eppendorf tube. Debris were removed by spinning at lowspeed in a microfuge for 2 min. The bacteriophage were then incubated at 37°C for 2 hr in 20 mM EDTA / 800µg / ml proteinase K / 0.8 % SDS and transferred to 68°C for 15 min. They were then left at RT for 10 min, then extracted with phenol / chloroform (1:1) and chloroform, separating the two phases by centrifugation in the microfuge for 5 min each time. To the aqueous phase 1 /10th volume of 3M NaOAc pH 6 was added, mixed and 3 volumes of cold ethanol were added. After precipitation at -70°C, the pellet was resuspended in 300 µl 0.3M NaOAc pH 7 and reprecipitated with ethanol. The pellet was washed twice with 70% ethanol, dried, resuspended in 200µl TE pH 8.0 and stored at 4°C.

2.3.5 Isolation of Bacteriophage λ DNA by DE52 Purification Method

(Helms et al., 1985)

The bacteriophage were plated as in Section 2.3.4 to give confluent lysis. SM buffer was added to recover the plate lysate. This was then pooled and spun at 2,500 rpm for 15 min at RT. The supernatant was transferred to a centrifuge tube and spun in the Beckman L8 ultracentrifuge using type 70-T1 rotor at 35,000 for 90 min at 18°C. The supernatant was discarded and the pellet containing the bacteriophage was resuspended O/N in 3ml of λ -diluent (10mM Tris-HCl pH 8 / 2mM $MgCl_2$) on a shaker at 4°C. A DE52 column of volume 5ml was equilibrated with 10mM Tris-HCl pH 8. The bacteriophage suspension was spun at 10,000 rpm for 15 min at RT to spin down bacterial debris, and the supernatant was loaded onto the column. Five ml of chase buffer (10mM Tris-HCl pH 8 / 10mM $MgOAc$ / 60 mM $NaOAc$) was then run through the column. Two ml of elution buffer (10 mM Tris-HCl pH 8, 50mM $MgOAc$) were then added and allowed to run into the column. Another 1ml of elution buffer was added and 1ml of the column runthrough was collected. A 2nd ml of elution buffer was added and 1ml of column runthrough was collected. A 3rd ml of elution buffer was added and the equivalent runthrough collected. To each 1ml fraction are added 25 μ l of 20% SDS and 8.4 μ l 5mg / ml Pronase (BCL), and left at RT for 5 min. 167 μ l of 3M potassium acetate pH 8.5 was then added. A precipitate was formed which disappeared when the samples were heated at 88°C in a dribloc for 20 min. The precipitate reformed when they were transferred on to ice for 5 min. The samples were then spun at RT for 15 min at 10,000 rpm in a microfuge. The supernatants were split into two and to each was added 17 μ l of 1mg / ml mussel glycogen (BCL) and 600 μ l isopropanol. The DNA was precipitated at -70°C for 30 min and spun at RT for 15 min at 10,000 rpm. The

pellets were rinsed with 70% ethanol and briefly dried before resuspending in 20µl TE pH 8. They were then pooled together and stored at 4°C.

2.3.6 Isolation of Chromosomal DNA

(Bell et al., 1981)

Chromosomal DNA was isolated from cultured cells and from white cells ("buffy coat") of blood. The buffy coat is the layer of cells that forms between the red cells and plasma when whole blood is centrifuged at 2,500g for 30 min.

Nuclei were prepared by adding lysis buffer (0.32M sucrose / 10 mM Tris-HCl, pH 7.4/5 mM MgCl₂ /1% Triton X-100) to cells and centrifuging at 2,000 rpm for 15 min at 4°C. The supernatant was carefully decanted, leaving behind the nuclei and cell membranes in the pellet. Lysis was repeated twice more (for blood and once for cultured cells) and the nuclei resuspended in 10ml 150mM NaCl / 5mM EDTA / 50mM Tris-HCl, pH 8.0. Proteinase K was added to 200 µg/ml, SDS to 0.5% and the suspension was incubated at 37°C, until all the protein and membranes were solubilised. The solution was then extracted once with an equal volume of phenol and then subsequently with equal volumes of phenol : chloroform : isopropanol (25:24:1) and once with chloroform : isopropanol (24:1). After each extraction the organic and aqueous layer was separated into two phases by centrifugation at 2,000 rpm for 15 min at 4°C. The organic layer was removed using a syringe, leaving behind the aqueous layer and interface. After the final extraction, the aqueous layer was removed to a clean tube and the chromosomal DNA precipitated with 2.5 volumes of cold ethanol. The DNA precipitated immediately and was removed with a sealed capillary in the form of a hook. It was briefly washed in 70% ethanol and resuspended O/N at 4°C in 200µl 10mM Tris-HCl / 10.1mM EDTA pH 7.4 and stored at 4°C at a concentration of

approximately 1mg / ml.

2.4 Digestion of DNA by Restriction Endonucleases

Restriction enzyme digests were carried out using the conditions recommended by the supplier. Analytical digests used 10-50ng DNA in a final volume of 10-20 μ l if end-labelling, or 0.5 to 1.0 μ g DNA in a similar volume for agarose mini gel. Preparative digests and digests for large agarose gels used 0.5-10 μ g DNA in 50-100 μ l reactions. Routinely the amount of enzyme used was in excess of the amount recommended by the supplier. (1 unit of enzyme is the amount stated by the manufacturer to fully digest 1 μ g of λ DNA in 1 hr). Genomic DNA (5 μ g) was digested for 24 hr with 20U of enzyme. All cloned DNA samples were fully digested after 2 hr.

2.5 Radiolabelling of DNA

2.5.1 "Fill in" reaction

(Wu and Taylor, 1971; Maxam and Gilbert, 1980)

The recessed 3' ends produced by digestion of double stranded DNA with suitable restriction endonucleases were labelled directly by incubating with 20 μ M cold dNTPs / 5 units DNA polymerase I (Klenow fragment) / 1-50 μ Ci[α^{32} P] dNTP for 20 min at RT. To ensure all fragments were "filled in" efficiently the reaction was chased with cold dNTPs at a final concentration of 250 μ M, for a further 10 min at RT. The reaction was terminated by heating at 70°C for 20 min.

2.5.2 Nick Translation

(Rigby et al., 1977)

Ten - 50 μ g DNA in 10mM Tris-HCl / 0.1mM EDTA pH 7.4 was mixed in nick-translation buffer (Section 2.1 f) with 20 μ Ci[α^{32} P]-dATP, 40 μ M

dCTP, dGTP, dTTP and 5µl Amersham nick-translation enzyme (2.5 U DNA polymerase I and 50pg DNase I) in a total volume of 100µl. The reaction was incubated at 15°C for 2 hr, then terminated by addition of 10µl of 100mM EDTA pH 8 and heating at 70°C for 10 min. The DNA was desalted and the unincorporated label was removed from the DNA by desalting on a 1 ml Sephacryl S-300 or Sephadex G50-SF column.

2.5.3 Primer Extension

(Feinberg and Vogelstein, 1983)

Ten-30ng DNA in 10mM Tris-HCl / 0.1mM EDTA pH 7.4 was labelled using the "standard multiprime DNA labelling system" from Amersham. This is based on the use of random sequence hexanucleotides to prime DNA synthesis by the "Klenow" fragment of DNA polymerase I as described by Feinberg and Vogelstein (1983). The DNA sample was denatured by heating for five min in a boiling water bath and then chilled on ice. To the DNA sample was added 10µl of the multiprime buffer containing cold dATP, dGTP and dTTP. Five µl of the primer solution of random hexanucleotides in an aqueous solution containing nuclease-free BSA was added to the reaction. Water was added in an appropriate volume to make the final reaction volume 50µl after addition of 20µCi radiolabelled [$\alpha^{32}\text{P}$]-dCTP and 2U of DNA polymerase I. The reaction was incubated RT for 3 - 5 hr, although it can be left O/N without any ill effects on the incorporation. The reaction was stopped by incubation at 70°C for 10 min and the unincorporated label was removed as described in Section 2.5.2.

2.6 Gel Electrophoresis

2.6.1 Polyacrylamide gel Electrophoresis

(Sanger and Coulson, 1978; Maxam and Gilbert, 1980)

Forty cm by 20cm glass plates were used. The top plate was treated

with a 2% solution of dimethyldichlorosilane in 1,1,1 trichloroethane. The plates were spaced with 0.3mm strips of plastic card and taped together with waterproof electrical tape. Sample wells 2mm - 4cm in width were formed at the top of the gel by insertion of a shaped plastic card "comb" between the glass plates.

Native and denaturing gels:

Four - 20% gels were prepared by mixing the stock acrylamide solution with 5ml 10 x TBE in a total volume of 50ml. For denaturing gels 21g of urea was added. To catalyse the polymerisation, 0.4ml 10% ammonium persulphate and 0.04ml TEMED were added, the solution mixed and the gel poured. Loading dyes were added to each electrophoresis sample. For native gels a glycerol dye mix of 30% glycerol in 5mM EDTA, 0.1% xylene cyanol and bromophenol blue was used. With urea gels a formamide dye mix of 60% formamide in 35mM EDTA, 0.1% dyes were used, and all samples were boiled for 5 min then snap-cooled on ice prior to loading.

Gels were electrophoresed in 1xTBE for 1½ - 4 hr at a constant current of 15 mA for 20% gels and 25-30 mA for lower percentage gels.

2.6.2 Autoradiography

(Laskey and Mills, 1977)

Gels were covered with cling-film and autoradiographed using X-OMAT "S" x-ray film (Kodak) at -70°C with a phosphotungstate intensifying screen (Cronex lightening Plus, Dupont). The films were sensitized by preflashing and developed in an ME-1 X-OMAT automatic processor.

2.6.3 Agarose gel Electrophoresis

(Southern, 1975)

Agarose horizontal slab gels of varying agarose percentages (0.7-1.5%) were run using BRL/Pharmacia apparatus. Agarose in 300ml of 1 x

TBE were poured into the gel tray of size 20 x 24 cm, with a 20 x 60 μ l well-former and left for 1 hr. Two litres of 1 x TBE were used as running buffer. DNA fragments were stained by addition of 0.3 μ g ethidium bromide to the gel before pouring. With genomic DNA the gels were stained after electrophoresis by submerging the gel in 1 x TBE with 2mg/l ethidium bromide for 30 min. Samples were mixed and electrophoresed overnight at 50-70 mA. Minigel apparatus of dimensions 11 x 11 cm (made by Biochem Dept Workshop) were used for rapid restriction mapping and for less accurate resolution, the "Uniscience" minigel apparatus was used (Johnson and Grossman, 1977). Fifty and 30ml agarose was used to form the gels in the respective apparatus and the running buffer was 70ml 1 x TBE / 10 μ l 0.1% ethidium bromide solution. Samples were mixed with glycerol dye mix and electrophoresed at 30-100 mA for 1-3 hr.

The ethidium bromide stained DNA (or RNA) was visualised under short-wave ultra-violet light (Sharp et al., 1973) using a transilluminator, and photographed using a Polaroid MP4 land camera fitted with a red filter. Exposures were for 1 second at f 4.5.

2.7 Recovery of DNA from Gels

2.7.1 Elution of DNA from acrylamide gels

(Maxam and Gilbert, 1980)

The section of gel containing the DNA of interest (identified by superimposition of the autoradiograph on the gel) was excised and placed in a siliconised glass tube containing 0.5ml 0.3M NaOAc pH 7.0. The DNA was eluted at 37°C O/N with constant shaking. After elution the supernatant was removed and the gel slice washed with 100 μ l 0.3M NaOAc pH 7.0, the supernatants were pooled and residual acrylamide removed by centrifugation for 5 min in a microfuge at RT. One and half volumes of

ethanol were added to the supernatant and the DNA pelleted after precipitating at -70°C for 20 min. The pellet was resuspended in 300 μl 0.3M NaOAc, pH 7.0 and ethanol precipitated twice more before washing in 70% ethanol and drying under vacuum. The DNA was resuspended in 25-50 μl of 10mM Tris-HCl / 0.1mM EDTA. pH 8.0.

2.7.2 Recovery of DNA from low gelling temperature Agarose

(Feinberg and Vogelstein, 1984)

A 2% solution of Sigma type VII low gelling temperature agarose in water, produced by heating to above 50°C was mixed with a 5ml packed volume of DE-52 cellulose (Whatman) (pre-equilibrated and stored at 4°C in 5XPBS phosphate buffered saline) and held at 50°C for 30 min. The resin was removed by centrifugation at 2,000 rpm for 5 min and the agarose transferred to a clean tube. This purification was repeated and followed by clarifying spins to remove traces of resin, dipping the agarose at 50°C for 10 min between spins to keep it molten. The molten agarose was then split into 25ml aliquots and 1/10th volume of 10xTBE (Section 2.1 f) added. One melted aliquot of purified agarose was sufficient for one gel and was diluted with 1xTBE for lower percentage gels. The gels were poured and run at 5mA O/N using the minigel apparatus with 4 x 2 cm well-former and loading a maximum of 2 μg total DNA per well. The gel was run until the required band was well separated, this band was then cut out after placing the gel on a UV transilluminator to visualize the DNA. The excised band was placed in a preweighed 1.5ml eppendorf tube and its weight calculated. A volume of water (in ml) equivalent to three times the weight of excised agarose (in g) was added, heated at 70°C for 10 min and cooled to RT. The sample was either aliquoted and frozen at -20°C , or extracted with phenol and phenol / chloroform (1:1), and finally chloroform. The DNA was precip-

itated by addition of 1/10th volume 3M NaOAc pH 7.0 and 2.5 volume absolute ethanol. The pellet was washed extensively with 70% ethanol, dried and resuspended in an appropriate volume in 10mM Tris-HCl /0.1mM EDTA pH 7.4.

2.8 Screening of Cloned DNA

2.8.1 Screening of Plasmid Clones

(Denhardt, 1966; Glegen et al., 1979)

Colonies were plated onto agar/ampicillin plates and grown overnight at 37°C. Bacterial colonies were transferred to Whatman 541 filters by leaving the filter in contact with the colonies for about 10 min at RT. The filters and plates were marked in an identical fashion so they could be matched up later. The plates were sealed and stored at 4°C while the filters were processed. The colonies were lysed in situ by floating the filters colony side up on the following solutions:

0.5M NaOH	- 5 min
0.5M Tris-HCl, pH 7.4	- 5 min
0.5M Tris-HCl, pH 7.4	- 5 min
2 x SSC	- 5 min

The entire cycle was repeated one more time and was followed by a quick dip in ethanol and the filters were then dried at RT.

The filters were moistened in 6 x SSC and prehybridized at 42°C for at least 3 hr in hybridization buffer (Section 2.1 f) containing 50% formamide, 100µg / ml sonicated salmon sperm DNA and 10 µg / ml E. coli pAT 153 / Pvu II/8 DNA which had been boiled for 10 min and snap cooled once before use. The filters were hybridized O/N in the above buffer containing $\sim 5 \times 10^5$ cpm / ml of [$\alpha^{32}\text{P}$] -labelled probe, also boiled and snap cooled before use.

After hybridization the filters were washed as follows:

5 times in 2 x SSC / 0.1% SDS for 10 min at RT.

2 times in 0.2 x SSC for 30 min at 68°C.

The filters were air dried and wrapped in Saran wrap and autoradiographed. The autoradiograph was used to match with the original plates via the markings on the filters.

To re-screen the clones, the colony was scraped from the original plates, diluted in L broth, replated and screened as described above. Single colonies, not contaminated with the neighbouring colonies were picked from the plate and the plasmid DNA isolated (as in Section 2.3.1) for analysis by restriction mapping and nucleotide sequencing.

2.8.2 Screening of Bacteriophage plaques

(Maniatis et al., 1982)

E. coli Q358 cells in T-broth / 0.2% Maltose / 10mM MgSO₄ were grown at 37°C for 20 hr with constant shaking, after which they were again enriched with maltose and MgSO₄. An equivalent of 10⁵ plaque forming units (pfu) was added to 1.5ml of the enriched Q358 culture and incubated at 37°C for 30 min. The mixture was added to 9ml T-top agarose at 48°C, mixed and poured onto 14cm diameter T-agar plates and incubated at 37°C O/N. The plates were left at 4°C for at least 1 hour to allow the top agarose to harden. Dry nitrocellulose filters (Amersham) were placed in direct contact with the plaques. With a sterile needle dipped in cold India Ink, the filter and the solid medium were assymetrically marked. After 1 min the filter was removed and immersed with the DNA side up in the following solutions.

1.5M NaCl / 0.5M NaOH (Denaturing solution)	-60 sec.
1.5M NaCl / 0.5M Tris-HCl pH 7.4 (Neutralising solution)	-60 sec.
2 x SSC	-60 sec.

The filters were then air dried on Whatman 3mm paper. Multiple replicas could be taken from each plate using the original marks on the plate for each subsequent copy. The DNA was fixed on to the nitrocellulose by baking at 80°C for 2 hr before hybridization, washing and autoradiography (Section 2.8.1).

Rescreening of Plaques

Agar plugs were taken from the original plates using the ink marks on the plates to line up the autoradiographs, transferred to eppendorfs containing 500µl SM buffer, 1% chloroform and were stored at 4°C. Various dilutions of this suspension were plated using 0.5ml of ON Q358 culture. The infected cells were plated using 3.5ml of T-top agarose on 7cm diameter T-agar plates and incubated at 37°C. The plaques were screened, hybridized, washed and autoradiographed using the method described above. Single isolated plaques were chosen from the rescreened plates and stocks of infectious particles prepared as above. These were then used to make plate lysates to isolate DNA (Section 2.3.4) for restriction mapping and nucleotide sequencing.

2.9 Nucleotide Sequencing Techniques

2.9.1 Chemical Degradation

(Maxam and Gilbert, 1980)

Single end-labelled DNA fragments were isolated by secondary digestion with restriction enzymes of a fragment labelled at both ends from a "fill-in" reaction (Section 2.5.1) The labelled DNA was eluted from a gel (Section 2.7.1), resuspended in 25µl water + 5µl of 0.75mg / ml sonicated salmon sperm DNA and aliquoted as follows:

10 μ l ; "A + G reaction (partial cleavage of DNA at A and G nucleotides)
10 μ l into 15 μ l H₂O ; "C + T" reaction (partial cleavage at C + T nucleotides).

5 μ l into 20 μ l of a saturated solution of NaCl ; "C" reaction

5 μ l into 200 μ l of 0.05M Sodium cacodylate, pH 7.9 / 10mM MgCl₂ / 1mM EDTA ; "G" reaction.

All samples were stored on ice.

(a) "C + T" and "C" Reactions

30 μ l anhydrous hydrazine was added to each tube, vortexed and incubated at RT, for 7½ mins. The reaction was stopped by the addition of 200 μ l 0.3M NaOAc, pH 7.0 / 0.1mM EDTA / 10 μ g tRNA ("C + T Stop").

(b) "A + G" Reaction

25 μ l of 90% formic acid was added, the tube vortexed and incubated at RT for 4 mins. The reaction was terminated with 200 μ l "C + T Stop".

(c) "G" Reaction

1 μ l dimethyl sulphate was added, mixed and incubated at RT for 4 min. The reaction was terminated with 50 μ l 1.5M NaOAc, pH 7.0 / 1.0M β -mercaptoethanol / 10 μ g tRNA. All samples were precipitated with 750 μ l ethanol and washed with 400 μ l 70% ethanol.

(d) Piperidine Cleavage

The drained pellets were resuspended in 50 μ l 1M piperidine solution (freshly diluted), vortexed and heated for 90°C at 20 min in a Tecam "Dribloc" unit. The samples were centrifuged and the supernatant carefully transferred to fresh tubes containing 50 μ l 0.6M NaOAc, pH 7.0. The samples were precipitated with 3 volumes of ethanol. The pellets were washed in 70% ethanol and resuspended in 10 μ l formamide dye mix (Section 2.6.1). They were then boiled for 5 min and snap-cooled on ice before loading onto gels.

Denaturing gels of 8%, 10%, and 20% polyacrylamide were then run (Section 2.6.1) and autoradiographed (Section 2.6.2).

2.9.2 Dideoxy Sequencing

(Messing, 1983)

(a) Preparation of the Vector

Four μg of the replicative form of bacteriophage M13 mp9 was cut with 20 units of Sma I enzyme at 37°C for 2 hr. After the digest was checked for completion by running an aliquot on a mini gel, it was then incubated with 2 μg of calf intestinal alkaline phosphatase at 37°C for 30 min. The DNA was extracted extensively with phenol / chloroform mixtures before recovery by ethanol precipitation.

(b) Subcloning

DNA restriction fragments were isolated from acrylamide or agarose gels as described (Section 2.7), 20ng cut and phosphatased vector was mixed with 10-50ng DNA in 1mM rATP / 50mM Tris-HCl, pH 7.4 / 10mM MgCl_2 / 20mM DTT / 50 $\mu\text{g}/\text{ml}$ BSA in final volume of 10 μl . This was incubated for 12-16 hr at 16°C with 0.5 μl (approximately 2 units) DNA ligase (Winter et al., 1981).

(c) Transformation of E.coli TG1 cells

The cells were made competent by the following treatment. A 300 μl aliquot of an O/N culture of TG1 was diluted into 30ml 2 x TY broth and grown to an A_{600} of 0.4-0.6 OD units. The cells were pelleted by centrifugation at 1,500 rpm at 4°C for 6 min in the MSE 6L rotor. The cells were resuspended gently in 10ml 50mM CaCl_2 / 10mM Tris-HCl, pH 7.4 and incubated for 30 min on ice. The cells were pelleted again, resuspended in 2ml of the same buffer and stored at 4°C. The cells were used

on the same day to give maximum efficiency.

100-200 μ l competent cells were added to the ligation mix, incubated at 0°C for 45 min and then 42°C for 5 min. The transfected cells were added to 3ml molten L top agar containing 25 μ l of 2.5% aqueous solution of isopropylthiogalactoside (IPTG) and 25 μ l of 2% bromochloroindolyl-galactoside (BCIG) in dimethylformamide, kept at 45°C. Finally, 0.5ml of overnight JM103 cells was added, gently mixed and poured onto 2 x TY agar plates.

Incubation of the plates overnight at 37°C resulted in the production of blue and white plaques; areas of lower density cells caused by M13, a non-lytic phage, reducing the growth rate of infected cells. Blue plaques represent bacteria containing wild type M13 which has β -galactosidase activity due to complementation between the M13 and JM103. IPTG induces enzyme synthesis which hydrolyses BCIG, a substrate analogue, releasing a blue indolyl derivative. The cloning site in M13 is within the β -galactosidase gene so M13 containing insert DNA results in inactivation of the β -galactosidase gene. The M13 can no longer complement the "M15" protein of the host (Messing et al., 1977), no dye is released and white plaques will result.

(d) Preparation of Phage DNA

(Sanger et al., 1980)

Agar plugs of white plaques were grown in 1.5ml of one hundredth dilution of an overnight culture of TG1 in 2 x TY broth at 37°C for 5 hr, with constant shaking. The medium was transferred to an eppendorf tube and spun for 5 min in a microfuge at RT. The supernatant was transferred to a fresh tube and respun for 5 min. Approximately 1ml of this supernatant was transferred to another tube containing 200 μ l of 20% polyethylene glycol / 2.5M NaCl, mixed and incubated for 30 min on ice.

This was then centrifuged for 10 min, the supernatant discarded, and all residual PEG removed. The pellets were resuspended in 100 μ l 10mM Tris-HCl / 0.1mM EDTA pH 8.0, phenol / chloroform extracted and precipitated with 10 μ l 3M NaOAc pH 6.0 and 100 μ l isopropanol for 1 hr at -20°C. The DNA was pelleted, washed in 70% ethanol, dried and resuspended in 50 μ l 10mM Tris-HCl / 0.1mM EDTA, pH 8.0. The single stranded phage DNA was stored at -20°C.

(e) Relative Orientation of Integrated Fragments

(Hermann et al., 1980)

The orientation of the cloned restriction fragments was determined as follows. 2 μ l of each clone was mixed with 2 μ l of a reference clone in a sealed glass capillary in the presence of 4 μ l 600mM NaCl / 70mM MgCl₂ / 60mM Tris-HCl, pH 7.4 and 1 μ l 50% glycerol / 1% SDS / bromophenol blue. The tubes were incubated at 67°C for 1 hr, snap cooled on ice and electrophoresed on a 0.8% agarose minigel. Cloned DNA containing fragments of opposite orientation hybridize with each other and migrate slower due to the increase in size.

(f) Sequencing Reaction

(Sanger et al., 1977; Biggin et al., 1983)

A universal oligonucleotide primer (Duckworth et al., 1981) was used to prime for synthesis using the single-stranded DNA as template. The sequence of the oligomer was:

5' d(G T A A A C G A C G G C C A G T) 3'

The composition of the mixes used was:

"A" mix :- 0.125mM dCTP; 0.125mM dGTP; 0.125mM dTTP; 15 μ M ddATP
 "C" mix :- 6.25 μ M dCTP; 0.125mM dGTP; 0.125mM dTTP; 40 μ M ddCTP
 "G" mix :- 0.125mM dCTP; 6.25 μ M dGTP; 0.125mM dTTP; 80 μ M ddGTP
 "T" mix :- 0.125mM dCTP; 0.125mM dGTP; 6.25 μ M dTTP; 250 μ M ddTTP

Approximately 1.25ng primer was hybridized with 8.5 μ l single stranded DNA in 13.6mM Tris-HCl / 6.8mM MgCl₂ pH 8.5, in a total volume of 11 μ l, for 5 min at 65°C and then kept at RT until ready to use. 2 μ l of primed clone were added to four tubes; T, C, G and A. To this was added 2 μ l of the correct mix and 2 μ l of the enzyme premix: 2 U DNA Polymerase I ("Klenow" fragment) and 2.5 μ Ci[α ³²P]dATP in 11mM DTT / 0.15 μ M dATP / 10mM Tris-HCl, pH 8.0. The reaction was started by mixing the reactants in a capless eppendorf tube in a microfuge. The reaction was incubated at RT for 15 min and then chased for another 15 min with 2 μ l 0.5 mM dNTPs. The reaction was stopped by the addition of 4 μ l formamide dye mix (Section 2.6.1), and the samples boiled for 3 min.

(g) Electrophoresis of sequenced samples

The samples were run on either (a) 6% denaturing gels (Section 2.6.1) at 30mA for 5 hr or (b) buffer gradient gels at 30mA for 1 hr 40 min (Biggin et al., 1983). Prior to preparing both types of gel the top plate was treated with a 2% solution of dimethyl dichlorosilane in 1,1,1 trichloroethane and the back plate was treated with 2.5ml ethanol containing 75 μ l 10% acetic acid and 7.5 μ l "Wackersilicone GF 31" (Wacker Chemie, W. Germany) (Garoff and Ansorge, 1981). The solution was rubbed well onto the glass then washed with ethanol and thoroughly wiped to remove excess silane. The silane binds the gel covalently to the back plate, whereas the solution used on the notched plate repels the gel.

After electrophoresis the gels adhering to the silane treated plates were fixed in 10% methanol/10% acetic acid for 15 min and then

washed in water for 15 min. The urea free gels were dried for about 30 min at 100°C until dry, and then autoradiographed for 1-3 days at RT, without a screen, and the film was not pre-flashed.

Buffer Gradient Gels

0.5 TBE gel mix :- 6% acrylamide / 0.5 x TBE / 7M urea

2.5 TBE gel mix :- 6% acrylamide / 2.5 x TBE / 7M urea / 5% sucrose / 0.05% bromophenol blue.

Seven ml 2.5XTBE mix were mixed with 35µl 10% ammonium persulphate and 14µl TEMED. To 35ml 0.5XTBE mix were added 175µl 10% ammonium persulphate and 70µl TEMED. Six ml 0.5XTBE were taken up into a 10ml pipette, followed by 6ml 2.5XTBE mix, and gradient formed by introducing 5 air bubbles through the interface. This mix was poured between the gel plates and the remaining space filled with the 0.5XTBE mix. Running buffer, 0.5XTBE was used in the top reservoir and 1XTBE in the bottom.

2.9.3 Analysis of Nucleotide Sequence Data

(Staden, 1982a,b,1984)

Sequence data was analysed using a VAX II/780 computer. Sequences were read from the gels and entered into the computer by hand. Sequences were overlapped and aligned using DBAUTO and DBUTIL. Other programmes used in data analysis include DBSTART, CUTSIT, DIAGON, SEARCH and ANALYSEQ.

2.10 Isolation of Total Cellular RNA

(Chirgwin et al., 1979)

Total RNA was isolated from the tissue culture cell lines by the guanidine-cesium chloride method. Tissue culture cells which had been washed in PBS (Section 2.15) were resuspended in 3ml of 4M guanidine

thiocyanate solution (Section 2.1 f). The DNA was sheared with a sterile syringe and needle and until the solution was viscous. Three ml of CsCl solution was added (Section 2.1 f) to make a total of 6ml. Three ml of CsCl solution was put into a clear T.70 centrifuge tube and to it was added the 6ml solution. It was topped with paraffin and sealed. The tubes were spun at 35,000 rpm in the Beckman L8 ultracentrifuge for 16 hr at 18°C, without a brake. The top of the tube was sliced off and the protein layer was fished out with a Pasteur pipette. The supernatant was removed gently with a sterile 1ml Pasteur pipette without disturbing the RNA pellet at the bottom. The pellet was washed with 70% ethanol and dissolved in 200-500µl TE (25mM Tris-HCl / 1mM EDTA pH 7.5) + 0.2% SDS. RNA was precipitated with absolute ethanol at -70°C for 30 mins and resuspended in 100-200µl TES solution. Absorbance was checked at 260 and 280nm with $A_{260} \text{ } 1 = 40\mu\text{g}$ of RNA. RNA is best stored at -20°C in ethanol and precipitated as required.

2.11 DNA Transfer Blot Analysis

(Southern, 1979; Wahl et al., 1979)

2.11.1 Blotting on Nitrocellulose Membranes

Samples in 1/10 volume of glycerol dye mix were electrophoresed in agarose gels / 1 x TBE (Sec. 2.6.3) at 50mA for approximately 16 hr. The gel was washed with water stained with ethidium bromide if necessary, photographed and then washed with the following solutions:

- (a) Twice in 0.25M HCl for 15 min each at RT.
- (b) Twice in 0.5M NaOH / 1M NaCl for 15 min each at RT.
- (c) Twice in 3M NaCl / 0.5 Tris-HCl, pH 7.4 for 15 min each at RT.

The gel was rinsed with water between the washings. Blotting of agarose mini gels required shorter washing times and the acid wash was omitted for transfer of small size fragments.

The gel was placed on two sponges and two pieces of Whatman 3mm paper soaked in 20 x SSC. The area surrounding the gel was covered with cling-film and glass strips placed at the ends of the gel. A piece of nitrocellulose (Amersham) briefly wetted with water and 2 x SSC was placed on top of the gel, this was covered with two sheets of dry Whatman 3mm paper and approximately two inches of paper towels weighted with a glass plate and a weight. The buffer passed through the gel and nitrocellulose carrying the DNA from the agarose to the nitrocellulose by capillary action.

With cloned DNA, blots were taken for 1-2 hr. With genomic DNA blotting was for 24 hr. The sponges were kept moist with additional 20 x SSC and the paper towels were replaced several times. The DNA was baked onto the nitrocellulose for 2-3 hr at 80°C before hybridization.

Hybridization of Nitrocellulose filters

(Bernards and Flavell, 1980)

Filters were floated on 6 x SSC for 5 min to wet them, then washed in 1M NaCl / 1mM EDTA / 50mM Tris-HCl, pH 8.0 / 0.1% SDS for 1 hr at 42°C. Prehybridization was carried out at 42°C for at least 3 hr. Hybridization buffer (Section 2.1 f) containing 100µg / ml sonicated salmon sperm DNA and 10µg / ml E. coli / pAT 153 Pvu II-8 DNA and $\sim 5 \times 10^5$ cpm / ml of [α 32 P] labelled probe was boiled for 5 min and cooled on ice. The filters were hybridized for 16 hr at 42°C. After hybridization the filter were washed as follows:

5 times in 2 x SSC / 0.1% SDS for 10 min each at RT.

2 times in 0.2 x SSC for 30 min each at 68°C.

The filters were air dried and autoradiographed.

2.11.2 Blotting on Nylon Membranes

Nylon membranes (Amersham Hybond-N, Genescreen Plus-Dupont) prove very useful for stripping blots and rescreening. Also they are not as fragile as nitrocellulose. Gels to be blotted onto nylon were treated as follows:

1. Once in 0.25M HCl for 15 min at RT.
2. Twice in 0.4M NaOH / 0.6M NaCl for 15 min each at RT.
3. Twice in 1.5M NaCl / 0.5M Tris-HCl, pH 7.5 for 15 min each at RT.

The Hybond-N membrane did not need to be prewetted before blotting. The Genescreen Plus membrane was blotted on its concave side. It was first wetted with water and equilibrated in 10 x SSC for 30 min. Blotting was done using 10 x SSC. After blotting the Genescreen Plus membrane was treated with 0.4M NaOH for 30-60 sec followed by 0.2M Tris HCl / 2 x SSC for 1 min, then air dried. Hybond-N membranes were washed in 2 x SSC and baked at 80°C for 2 hr. After prewashing, prehybridization was done with 1% SDS in the hybridization buffer. After hybridization the membranes were washed as follows:

Genescreen Plus -

- 5 times in 2 x SSC at RT for 10 min each.
- 2 times in 1 x SSC / 1% SDS at 65°C for 30 min each.

Hybond-N -

- 5 times in 2 x SSC / 0.1% SDS at RT for 10 min each.
- 2 times in 0.2 x SSC at 65°C for 30 min each.

The membranes were not allowed to dry before wrapping in Saran Wrap and autoradiographed for atleast 48 hr. Stripping of the blots was done by washing twice in 2mM Tris HCl pH 7.4 / 1mM EDTA / 1% SDS at 80°C for 1 hr each.

2.12 Northern Blot Analysis

(Maniatis et al., 1982; Thomas, 1980)

For RNA blots, gels were prepared by dissolving 3g agarose in 222ml of water for 20 x 24cm gels. Thirty ml of 10 x MOPS (Section 2.1 f) was added. At 50°C, 48ml of 38% formaldehyde was slowly added in the fumehood along the side of the flask to prevent bubbles and the gel was immediately poured. The sample was prepared as follows: 25µl deionized formamide, 5µl 10 x MOPS buffer, 8µl of formaldehyde and 12µl of RNA solution (20µg) were mixed in a RNase free eppendorf tube and heated to 65°C for 5 min just prior to loading. 5µl of 50% (v/v) glycerol containing 0.1mg / ml bromophenol blue was added. A λ ladder which was used as marker was boiled for 1 min to denature. Running buffer was 1 x MOPS. The gel was run O/N at 45mA until the dye was ~ 10cm from the top. The gel was washed 5-10 min in water. No ethidium bromide was added. The gel was directly blotted onto the nitrocellulose wetted with 2 x SSPE without any treatment. Blotting buffer was 20 x SSPE and the duration was 24 hr. After blotting, the nitrocellulose was washed with 6 x SSC for 5 min and baked for 2 hr at 80°C. It was prewashed in 1M NaCl / 50mM Tris-HCl, pH 7.5 / 5mM EDTA / 0.1% SDS and then prehybridized for atleast 2 hr before hybridization at 42°C for 36-48 hr with labelled DNA probes of specific activity 10^8 cpm μg^{-1} . After hybridization the filters were washed 5 times in 6 x SSC / 0.1% SDS at RT for 15 min each and twice in 2 x SSC at 68°C for 30 min each. They were then air dried and autoradiographed at -70°C.

2.13 Tissue Cell Culture

Mouse Y1 adrenocortical cell line was maintained as monolayers in plastic flasks (Nunc) in Dulbecco's modification of Eagles' medium (DMEM, Flow Laboratories) supplemented with 10% heat inactivated fetal

calf serum / 2mM glutamine / 5µg / ml penicillin - streptomycin / 100µg /ml kanamycin. Confluent cells were replated at 1:3 dilution by the following procedure. Cells were washed with 10ml phosphate buffered saline (PBS Dulbecco's Flow) and incubated with 2ml 0.05% trypsin / 0.02% EDTA for about 5-10 min at RT. The released cells were resuspended in DMEM media, divided into fresh sterile flasks, equilibrated with 5% CO₂ gas and placed in a Hotpack CO₂ incubated at 37°C.

2.13.1 Long term Storage of Cells

Approximately 10⁷ cells obtained by trypsinization were collected by centrifugation in a MSE bench top centrifuge at 2,000 rpm, for 5 min at RT. The cells were resuspended in 1ml of fetal bovine serum / 10% DMSO, slowly frozen at -70°C, and then stored in liquid nitrogen. The cells were recovered by thawing rapidly to 37°C. They were then diluted and washed in DMEM media before incubation in tissue culture flasks.

2.14 Transfection of Y1 cells

(a) The three plasmids used were BglII / BamHI subclones of the 21-OHase A gene, 21-OHase B gene and 21-OHase B gene from individual II.I respectively. The pRSVCAT plasmid was (2µg) used for cotransfection.

(b) 40µg of plasmid DNA in 450µl of TE was coprecipitated with 50µl of 2.5M CaCl₂ by dropwise addition of 500µl of 280mM NaCl / 50mM Hepes / 1.5mM Na₂HPO₄, pH 7.12 (2xHBS) over a period of 1 min with constant agitation. This was then added dropwise to a flask of Y1 cells which had been washed with PBS after removing the media. After 10 min, 10ml media was added to the flask, CO₂ was passed through the media and the flasks were left at 37°C for 6 hr. After 6 hr, the media was changed and ~15 ml fresh media was added, the flasks were gassed and left at

37°C for 48 hr after which the cells were harvested.

2.15 Assay of Chloramphenicol Acetyltransferase Activity

Forty-eight hours after transfection, cells were harvested by trypsinization, washed with PBS and resuspended in 25ml PBS. 20ml of the cell suspension was pelleted at 2,000 rpm for RNA preparation and 5ml of the cell suspension was pelleted for preparation of the cellular extract. The pellet was resuspended in 100µl of 0.25M Tris-HCl pH 7.5. After five cycles of freezing in dry ice - IMS bath and thawing at 37°C, the samples were spun for 10 min at 10,000 rpm in a microfuge and the supernatants were removed to fresh eppendorf tubes. The supernatants were assayed for CAT activity as described by Gorman et al. (1982) and Wu et al. (1987).

Twenty microlitres of the cell extract was assayed in 180µl containing 0.25M Tris-HCl pH 7.5, 1mM acetyl-CoA and 0.5µCi [¹⁴C] chloramphenicol (54mCi / mmol, Amersham). After incubation at 37°C for 30 min, the chloramphenicol was extracted with 500µl of ethylacetate. The organic layer was dried down, dissolved in 10µl of ethylacetate, and spotted onto silica gel thin layer plates. The migration in chloroform-methanol (95:5) was for 10 min. The plate was dried and autoradiographed at RT.

CHAPTER III

STRUCTURE OF THE 21-HYDROXYLASE GENES

3.1 Introduction

CAH due to defective 21-hydroxylation is one of the most common inborn errors of metabolism, second only to Phenylketonuria (PKU) which is an autosomal recessive human genetic disorder caused by a deficiency of hepatic phenylalanine hydroxylase. 21-OHase deficiency had been linked to the MHC in 1977 by Dupont et al., and subsequently mapped to the HLA class III region along with the genes encoding the serum complement proteins C2, Factor B, C4A and C4B (White et al., 1985b; Carroll et al., 1985a). Detailed restriction mapping of uncloned genomic DNA and of cosmid clones have localized two 21-OHase genes, 21-OHase A and 21-OHase B, to 6 kb regions flanking the 3' end of the C4A and the C4B genes, respectively.

The presence of the two 21-OHase genes led to the speculation that there might be two species of 21-OHase encoded by separate genes, as many patients having impaired 21-hydroxylation of 17-hydroxyprogesterone in the simple virilizing form of CAH had little or no impairment in 21-hydroxylation of progesterone as in the case of the salt-wasting form of CAH (Migeon, 1977). This hypothesis fell into disfavour on the demonstration that the homozygous deletion of the 21-OHase A gene in individuals did not affect their adrenal steroidogenesis (White et al., 1985b). A homozygous deletion of the 21-OHase B gene, however, was found in individuals with CAH. This clearly indicated that the 21-OHase A gene did not play an active role in adrenal steroidogenesis while the 21-OHase B gene was essential for normal synthesis of both cortisol and aldosterone.

To define the differences between the two 21-OHase genes which affected their role in adrenal steroidogenesis, it was essential to analyse them at the molecular level. With this in mind, two cosmids that contained the C4A and 21-OHase A genes, and the C4B and 21-OHase B

genes were further characterized. The two 21-OHase genes were subcloned and sequenced. On the basis of their nucleotide sequence, it was then clear that only the 21-OHase B gene encoded an active enzyme while the 21-OHase A gene was a pseudogene due to a number of mutations, three of them preventing the transcription of a functional protein.

3.2 The Organization of the 21-OHase Genes in Individual HPFH

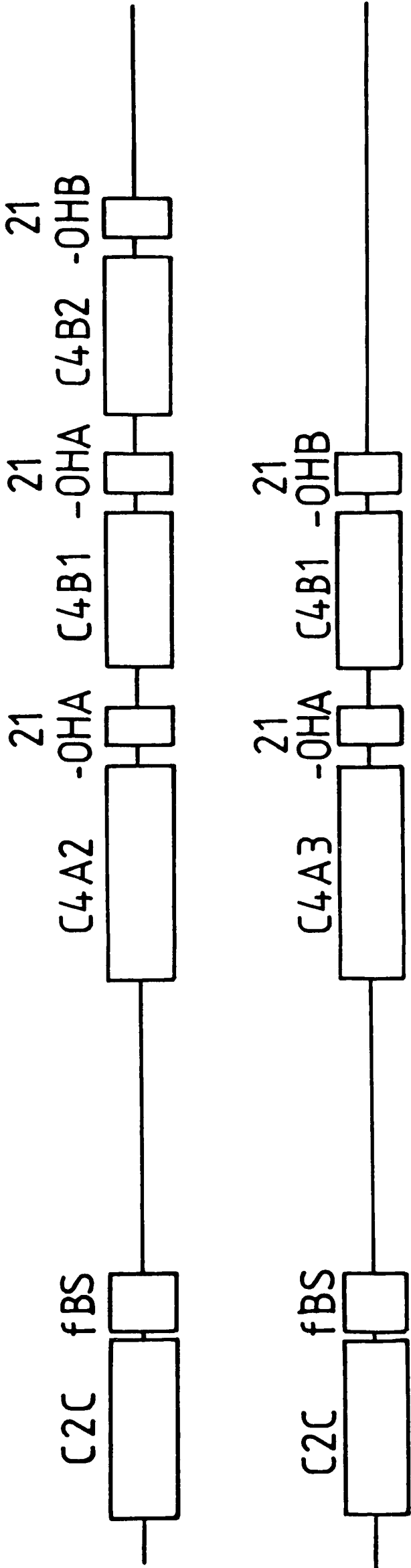
The molecular map of 160 kb of genomic DNA in the HLA class III region containing five complement genes was constructed on the basis of five overlapping cosmid clones isolated from that region from a cosmid library constructed from an individual HPFH typed as HLA A3,31 B14,37 DR1,2 C4A2,3 C4B1,2 BfS C2C, by Grosveld et al. (1982) (Carroll et al., 1984b). The C2 and Factor B genes, separated by less than 1 kb, were 30 kb away from the C4 genes. Analysis of the cosmid clones using the C4A and C4B class specific synthetic oligonucleotides showed that there was a single C4A gene and two C4B genes. These genes were assigned to the haplotype A3, B14, DR1, C4A2, C4B1,B2, BfS, C2C which has been shown to have a duplication at the C4B loci (Raum et al., 1984). The B14, C4A2, C4B1,B2, BfS supertype is associated with late onset congenital adrenal hyperplasia. This haplotype has two 21-OHase A genes and one 21-OHase B gene (Garlepp et al., 1986).

The two proposed haplotypes for HPFH are illustrated in figure 3.1 The B37 DR2 haplotype represents the expected two C4 loci, that is C4A and C4B, as most haplotypes have only two active C4 genes. Alternatively, the B37 haplotype may have a single active C4A3 gene and null at the C4B locus since the complement typing would not identify the number of C4B1 genes expressed. The nature of the 21-OHase genes present on this haplotype is also not known. On the basis of information available, it appears that individual HPFH did not suffer from the nonclass-

Figure 3.1

Proposed haplotypes of HPFH. Chromosome I carries the B14, DR1 haplotype, and has a duplication of the C4B locus and the 21-OHase A locus. The homologous chromosome II carries the B37, DR7 haplotype, and has a single 21-OHase A and 21-OHase B gene.

Organization of the C4 and 21-OHase genes in individual HPFH



ical form of CAH, therefore at least one 21-OHase gene in individual HPFH has to be functional. It would seem that there would be one normal 21-OHase B gene on the B37 haplotype on the assumption that the 21-OHase B gene on the CAH-associated B14 haplotype did not partake in normal adrenal steroidogenesis. Therefore the B37 haplotype could have a gene organization of C4A3 - 21-OHase A - C4B1 - 21-OHase B, or may have C4A3 - 21-OHase B due to a deletion of the 21-OHase A - C4B1 region as in some other characterized haplotypes (Carroll et al., 1985b).

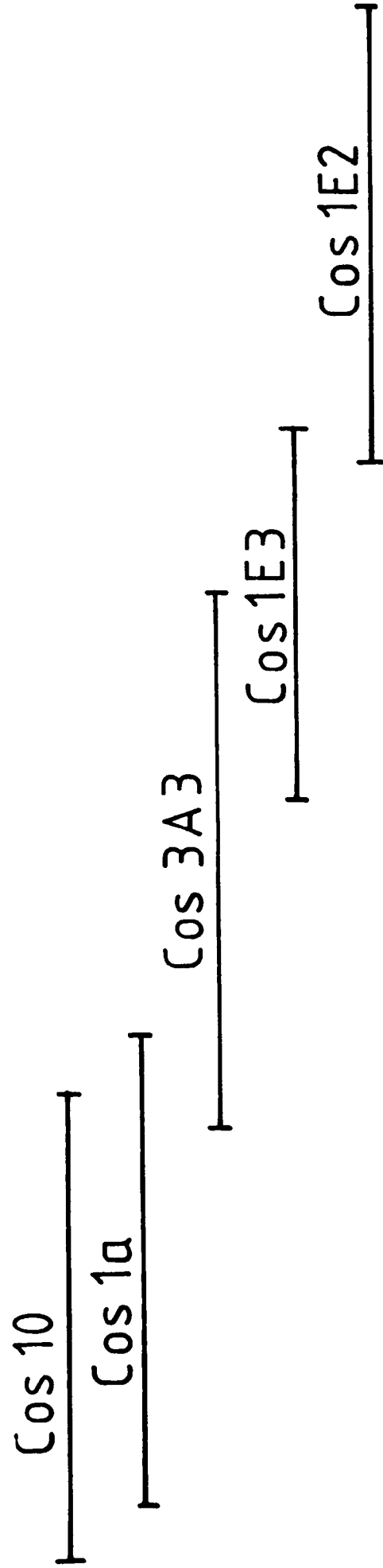
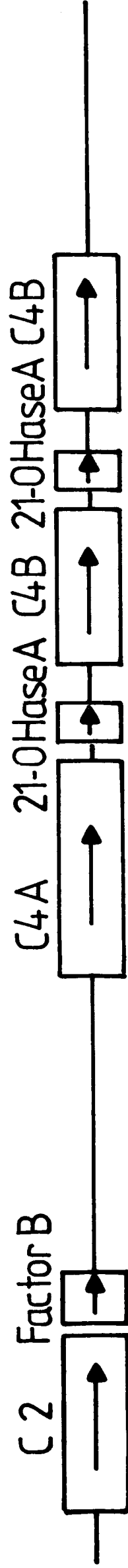
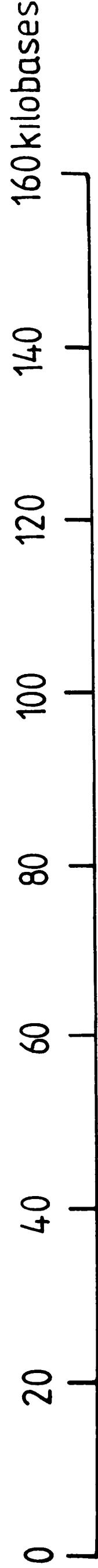
3.2.1 Cloning of the 21-OHase A Gene

Three overlapping cosmids 3A3, 1E3 and 1E2 were isolated from the HPFH cosmid library (Figure 3.1) and three C4 genes that were linked on one chromosome were identified by M.C.Carroll. Cos 3A3 contained a C4A gene, while cos 1E3 and cos 1E2 contained C4B gene sequences. Cos 1E3 was shown to contain a 21-OHase gene by Carroll et al. (1985a). A 2.4 kb band of human adrenal RNA was identified in Northern blot analysis by hybridizing with a probe prepared from cos 1E3. Hybridization to the same 2.4 kb band, of a 23-base synthetic oligonucleotide probe complementary to bovine steroid 21-OHase cDNA, identified the gene as encoding 21-OHase. Restriction digests of cos 1E3 and cos 1E2 indicated the same restriction pattern in both the 21-OHase genes contained in them. An Asp718 digest failed to detect any differences between the two cosmids in the 21-OHase region and hence it was deduced that both cosmids contained 21-OHase A genes, although only one 21-OHase A gene from cos 1E3 was characterized in detail. Since there was no cosmid isolated beyond the limit of the second C4B gene, the 21-OHase B gene which would be there by inference was not isolated.

Confirmation that the gene encoded steroid 21-OHase came by nucleotide sequencing of a 0.43 kb genomic fragment that hybridized with the

Figure 3.2

The three overlapping cosmids from the HPFH cosmid library that contain the C4A and two C4B genes and the two 21-OHase A genes on the B14,DR1 haplotype. No cosmid was isolated beyond the limits of the second C4B gene. Cos 10 and Cos 1a overlap the C2 and Factor B genes.



synthetic probe. Its derived amino acid sequence was more than 70% homologous with bovine and pig steroid 21-OHase sequences, but less than 40% homologous with rat liver P-450 sequence (Carroll et al., 1985a). This gene was later identified to be the 21-OHase A gene on the basis of the 3.2 kb TaqI fragment which is characteristic of the 21-OHase A gene (White et al., 1985b). A 4 kb KpnI fragment which is characteristic of the 21-OHase A gene was subcloned from cos 1E3 in the plasmid vector pBRSV40 (pSV4kpn) by Carroll et al., (1985a). As this fragment did not contain the entire 21-OHase A gene (Figure 3.4), a BglII / BamHI fragment 5.5 kb in length was subcloned into pAT / PvuII / 8 by blunt end ligation of an end filled BglII / BamHI double digest of cos 1E3 (Rodrigues et al., 1987). Positive clones were identified by hybridization to a 600 bp HinfI 21-OHase probe made from the pSV4kpn subclone of the 21-OHase A gene. Plasmid DNA was isolated from the positives and the insert size was estimated after a ClaI / XhoI double digest by comparison with standards. One clone was further characterized by doing several single and double digests with restriction endonucleases such as AccI, Asp718, BamHI, EcoRI, HaeII, NcoI, SmaI, StuI, Asp718 / ClaI, ClaI / NcoI, EcoRI / Asp718, HaeII / ClaI, SmaI / Asp718, SmaI / ClaI, SmaI / NcoI, Asp718 / NcoI, SmaI / NcoI. These digests were subjected to Southern blot analysis using the 600 bp HinfI probe and on the basis of the pattern, the various restriction sites were mapped in the subclone (Figure 3.3). These sites were then used to generate fragments for subcloning into M13mp9 and sequence analysis.

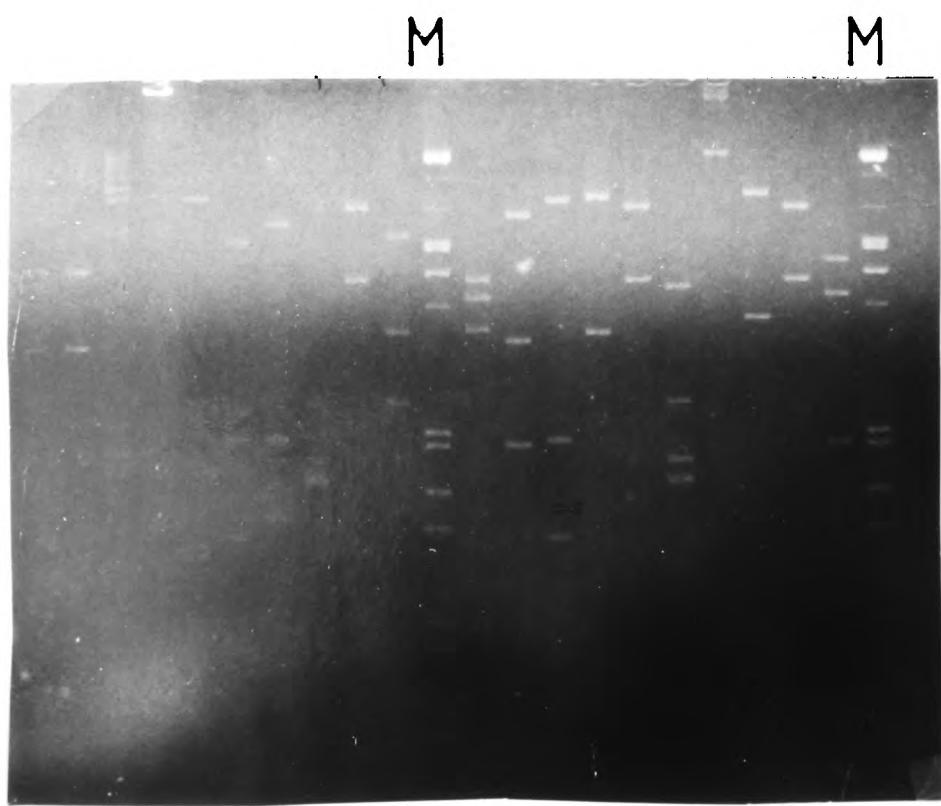
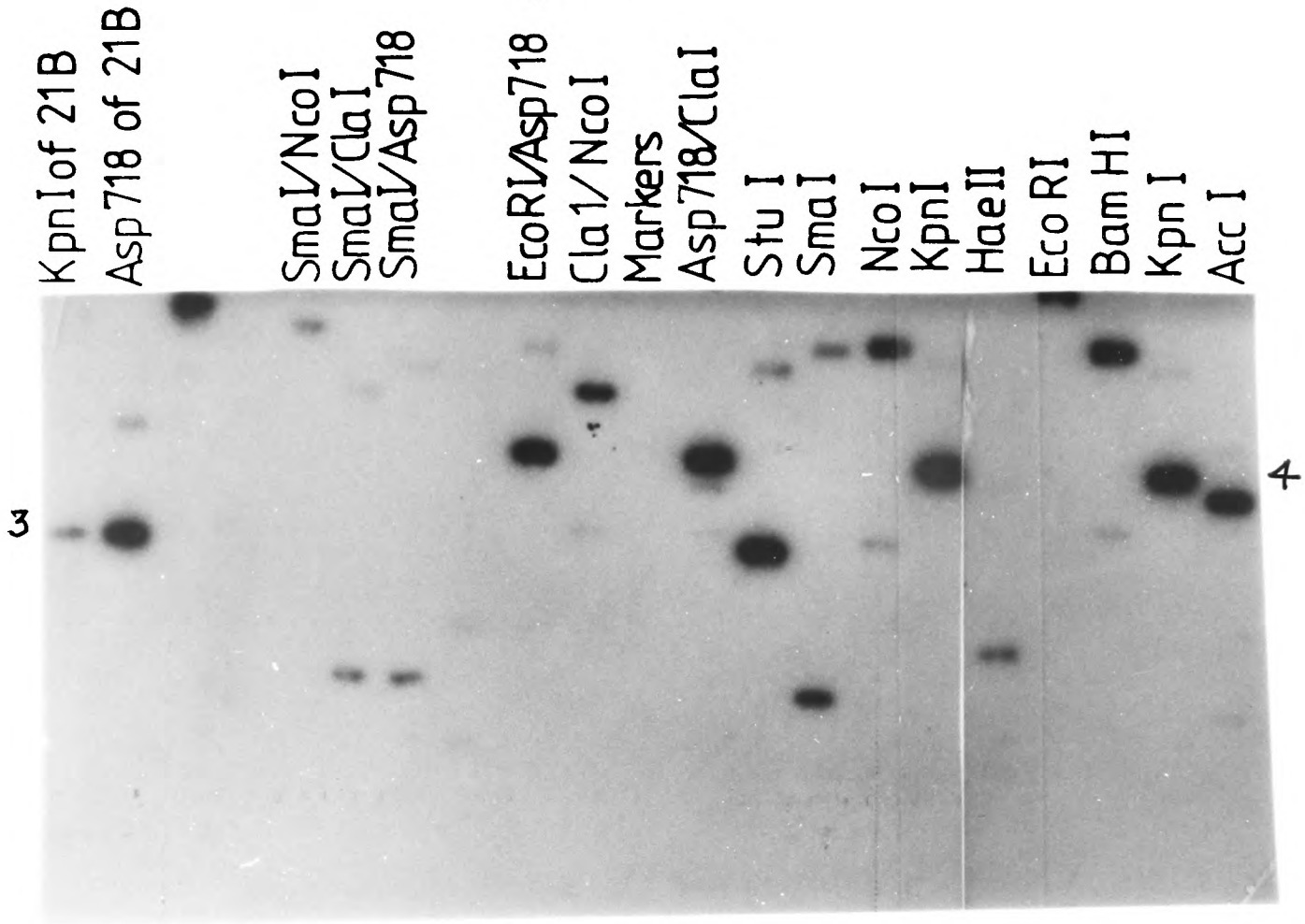
3.3 The Organization of the 21-OHase Genes in Individual KEM

The individual KEM who is homozygous for the deleted C4A null allele, with the HLA haplotypes A1,2 B8,7 DR3,4 C4AQ0 C4B1 BfS C2C, was

Figure 3.3

Southern blot analysis of the BglII / BamHI subclone of the 21-OHase A gene. The digests done were AccI, Asp718, BamHI, EcoRI, HaeII, NcoI, SmaI, StuI, Asp718 / ClaI, ClaI / NcoI, EcoRI / Asp718, HaeII / ClaI, SmaI / Asp718, SmaI / ClaI, SmaI / NcoI, Asp718 / NcoI and SmaI / NcoI. The bottom panel shows the digest run on a agarose gel while the top panel shows the autoradiograph of the Southern blot probed with the 600 bp HinfI probe.

Southern analysis of Bgl II/Bam HI subclone of 21-OHase A gene.



shown to have a deleted C4A and 21-OHase A gene on both his chromosomes by Southern blot analysis of genomic DNA (Carroll et al., 1985b). A molecular map was constructed on the basis of Southern blot analysis after comparing the restriction patterns of the two haplotypes HLA A2, B7, DR4 and A1, B8, DR3 with the deleted C4A null alleles with a normal haplotype C4A3, B1. The restriction patterns of the two haplotypes that had the deleted C4A null allele were identical. Figure 3.4 shows the molecular map comparing the C4 haplotypes C4A3 C4B1 and C4AQ0 C4B1. Based on the results from Southern analysis, the minimum limits of the deletion have been proposed as extending from the 7 kb intron in the C4A gene at the 5' end to the homologous region in the C4B gene. However, localization of the precise site where recombination might have occurred could not be determined at the molecular level. It is interesting that the same region of DNA was deleted on the C4A null haplotypes, which suggests that a common mechanism is involved. KEM has one 21-OHase B gene on each of his haplotypes. A cosmid library of human genomic DNA prepared from white blood cells from individual KEM was constructed by Dr. M.C.Carroll. The cosmid clone KEM-1 isolated from the genomic library contained a complete C4 gene that was identified as C4B by hybridization with a C4B specific probe.

3.3.1 Cloning of the 21-OHase B Gene

The cosmid clone KEM-1 was analysed by digesting it with the restriction endonucleases TaqI and KpnI. Southern blot analysis followed by hybridization to the 600 bp HinfI 21-OHase probe showed the presence of a 3.7 kb TaqI band and a 3 kb KpnI band (Figure 3.5) which is characteristic of the 21-OHase B gene. The 3.7 kb TaqI fragment was subcloned into the plasmid vector pATX (which is a variant of pAT 153 / PvuII / 8) by shotgun cloning of the entire endfilled TaqI digest of cos KEM-1. It

Figure 3.4

Comparison of the molecular maps of haplotypes C4A3, C4B1 and C4AQ0, C4B1 (KEM). Horizontal bar indicates approximate limits of deletion. KEM-1 has one C4B and one 21-OHase B gene on each of his two chromosomes. The C4A and 21-OHase A genes are deleted on both chromosomes. Cos KEM-1 contains the entire C4B and 21-OHase B gene. The restriction map around the region on a normal haplotype is shown on top. Bg, BglII; K, KpnI; T, TaqI and B, BamHI, respectively.

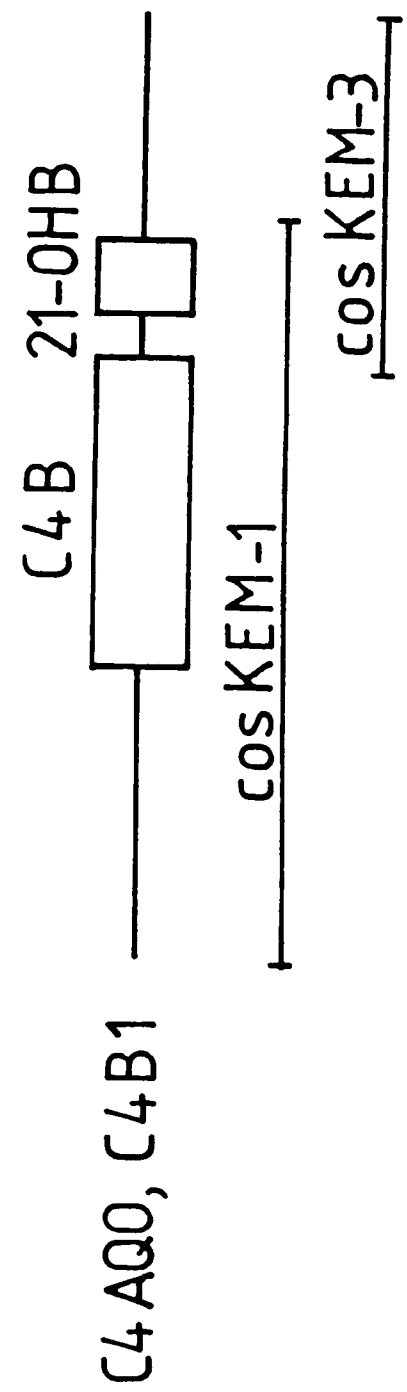
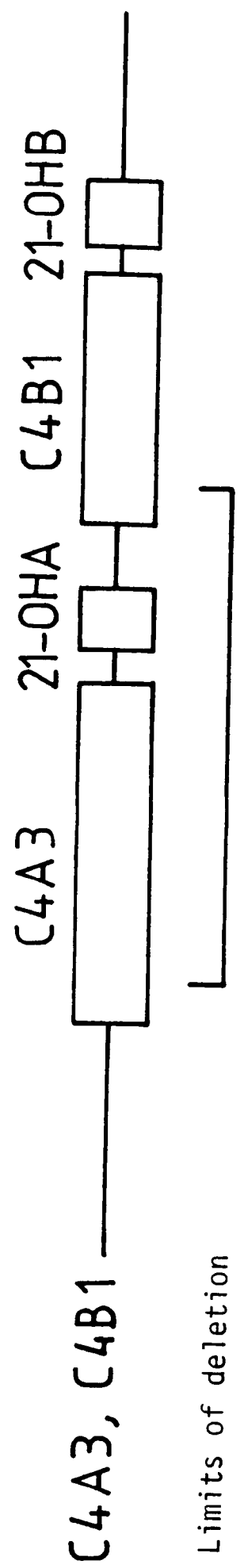
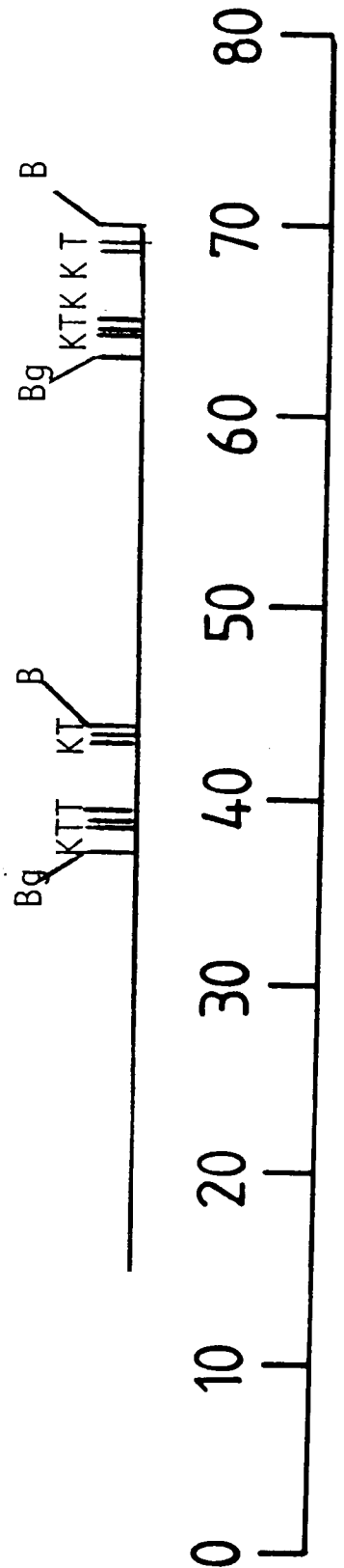
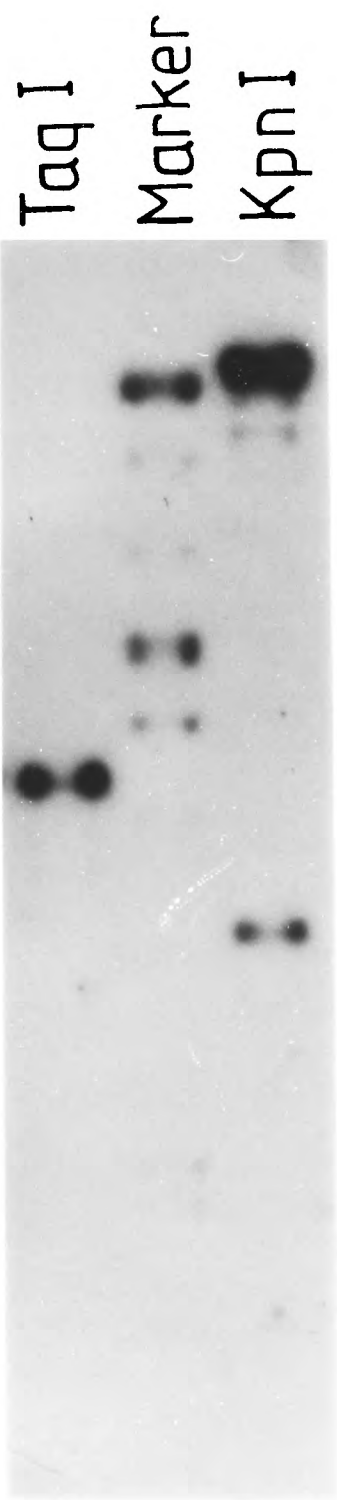


Figure 3.5

Southern blot analysis of cos KEM-1 with TaqI and KpnI. The 3.7 kb TaqI fragment and 2.9 kb KpnI fragment characteristic of the 21-OHase B gene light up with the 600 bp HinfI 21-OHase probe.

Cos KEM-1



3.7 kb

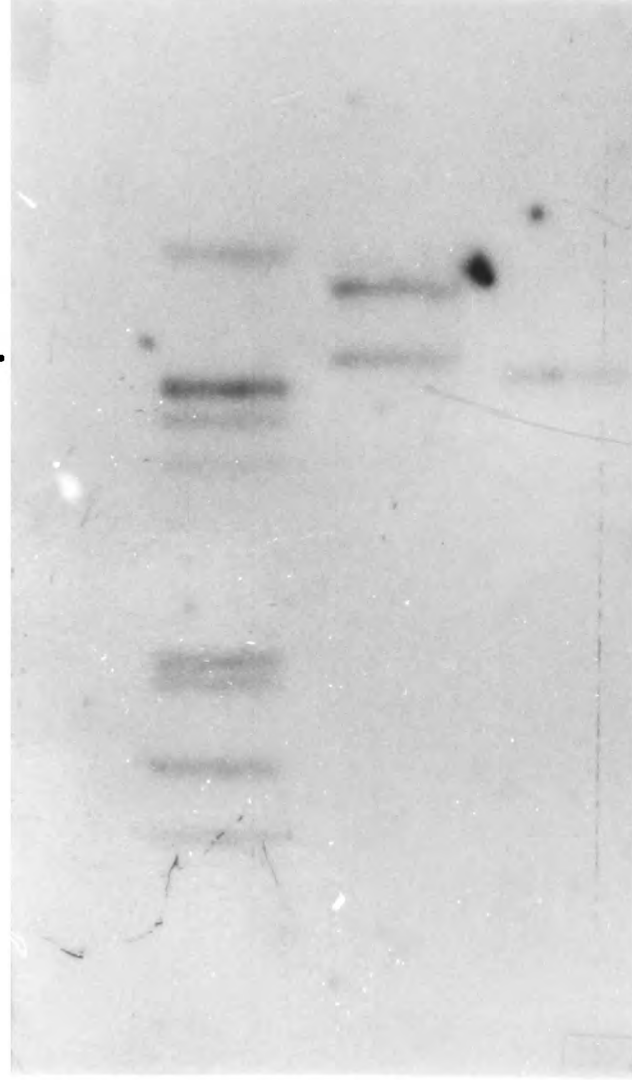
2.9 kb

Figure 3.6

BglII / BamHI digest of cos 1E3 and KEM-1.

BglII / BamHI digest of cos 1E3 gives a band at 5.5 kb that hybridizes to the 21-OHase probe. The BglII / BamHI digest of cos KEM-1 gave a very large fragment which meant that the limits of the clone were short of the BamHI site. The BglII / BamHI digest of cos KEM-1 was further digested with SalI, the SalI site being in the cosmid vector. A 5.2 kb band was seen with the BglII / BamHI / SalI digest which hybridized to the 21-OHase 600 bp HinfI probe.

5.5—



Marker
21-OHase A
21-OHase B

—5.2

was then screened out for positive colonies with the 21-OHase probe. To subclone the entire 21-OHase gene, a BglII / BamHI double digest was done on cos KEM-1. However there was no band at 5.5 kb as predicted from genomic digests. It was likely that the insert stopped just short of the BamHI site in the genomic DNA. The DNA was subjected further to digestion with SalI, the SalI site being that in the cosmid vector. The BglII / BamHI / SalI digest gave a band at 5.2 kb which hybridized to the 21-OHase probe (Figure 3.6). The entire endfilled digest was shotgun cloned into the plasmid vector pATX and screened for positive clones. The 10 clones picked after rescreening for single colonies were subjected to analytical HinfI / KpnI digests and compared to fragments in the 3.7 kb TaqI subclone. One clone was selected on the basis of its near identity to the 3.7 kb TaqI subclone. The vector in this clone was not pATx, instead it was the cosmid vector pTCF which gives a different restriction pattern to that of the plasmid vector. The subclone was then used to generate fragments for sequencing.

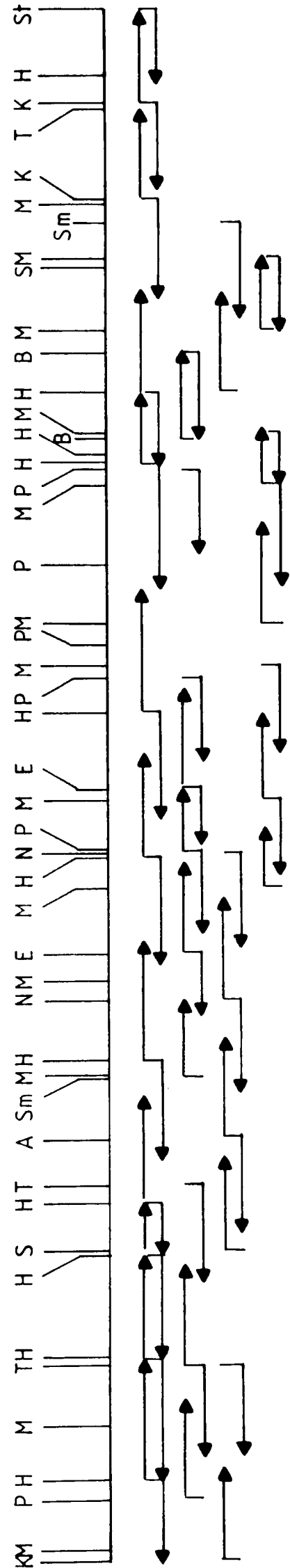
3.4 Subcloning of the 21-OHase A Gene into M13mp9

The 4 kb pSV4kpn subclone was used to generate the required fragments for sequencing. Since a preliminary restriction map on the basis of previous results was known (Carroll et al., 1985a) it was decided to digest the subclone with the enzyme HinfI which gave reasonably sized fragments. The HinfI fragments cloned were of the following lengths : 152 bp, 182 bp, 309 bp, 417 bp, 431 bp, 550 bp, 620 bp and 752 bp (Table 3.1). The HinfI sites were mapped to the pSV4kpn subclone by doing double digests (Figure 3.7). The HinfI digest was then endfilled and blunt end ligated into the vector M13mp9 which had been digested with SmaI and phosphatased (Section 2.9.2a). All the fragments were sequenced by the dideoxy chain termination method (Section 2.9.2f) and in

Figure 3.7

Restriction map of the 21-OHase A gene and the sequencing strategy. The restriction map of the 21-OHase A gene extending from the KpnI site to the StyI site is shown. The sequencing strategy is shown beneath. The arrows indicate the extent and orientation in which each fragment was sequenced A : AccI; B; BanI; E, EcoRI; H, HinfI; K; KpnI; M, MspI; N, NcoI; P, PvuII; S, StuI; Sm, SmaI; St, StyI; T, TaqI.

200 bp



both orientations. On the basis of the restriction map constructed from the results of various digests done on the pSV4kpn subclone and from the nucleotide sequence obtained from sequencing the HinfI fragments, it was then possible to choose restriction sites that would yield suitable fragments which would overlap the HinfI sites. To overlap the HinfI sites, further digests that were done were EcoRI / PvuII to clone 324 bp, 173 bp, 297 bp and 322 bp fragments; SmaI / EcoRI to clone the 374 bp fragment; TaqI / StuI to clone the 341 bp fragment; BanI to clone the 190 bp, and 239 bp fragments; HinfI / SmaI to clone the 507 bp fragment; TaqI / PvuII to clone 390 bp fragment; HinfI / Asp718 to clone the 571 bp fragment; AccI / StuI to clone the 330 bp fragment; AccI / NcoI to clone the 413 bp and 441 bp fragments and finally MspI to clone the 152 bp, 247 bp, 397 bp and 427 bp fragments to complete all overlaps. The 3' end of the gene which is beyond the 4kb KpnI fragment was subcloned from the BglII / BamHI subclone. The fragments cloned were 277 bp and 280 bp from an Asp718 / StyI digest of the subclone. Figure 3.7 shows the restriction map of the 21-OHase A gene and the fragments sequenced are shown below the restriction sites. The region 5' to the TaqI site was also subcloned and sequenced, and is shown in Appendix I.

3.5 Subcloning of the 21-OHase B Gene into M13mp9

The 3.7 kb TaqI subclone was used to generate fragments for sequencing. It was subjected to the same digests as the 21-OHase A gene subclone (Figure 3.8) (Section 3.2.1), the restriction map (Figure 3.9) was constructed and compared to the 21-OHase A gene restriction map. The major differences in the restriction sites between the two genes are with the restriction enzymes TaqI, Asp718 (or KpnI), EcoRI and MspI. The 21-OHase A gene has an additional TaqI site ~ 500 bp away from a second TaqI site in the 5' flanking region which contributes to the

Table 3.2 Fragments used for subcloning the 21-OHase B gene into M13mp9

<u>Restriction enzyme</u>	<u>Positions</u>	<u>Fragment length</u>
HinfI	-720,-412	308
	-412,-260	152
	-260, 156	416
	156, 779	623
	779, 1210	431
	1210, 1961	751
	1983, 2165	182
HinfI / SmaI	2165, 2672	507
MspI	103, 695	592
	695, 950	255
	950, 1347	397
	1465, 1891	426
	1891, 2043	152
	2043, 2251	208
	2362, 2588	226
StuI / ClaI	-403,-743	340
StuI / BamHI	2569, 2999	430
ClaI / Asp718	-743,-111	632
Asp718 / NcoI	111, 339	450
NcoI	339, 783	444
Asp718	2738, 3015	277
Asp718 / StyI	3015, 3295	280

Figure 3.8

Southern blot analysis of the 3.7 kb TaqI subclone of the 21-OHase B gene. The digests done are shown in the figure. The blot was hybridized to the 21-OHase 600bp HinfI probe and the restriction map was constructed on the basis of the fragments that light up.

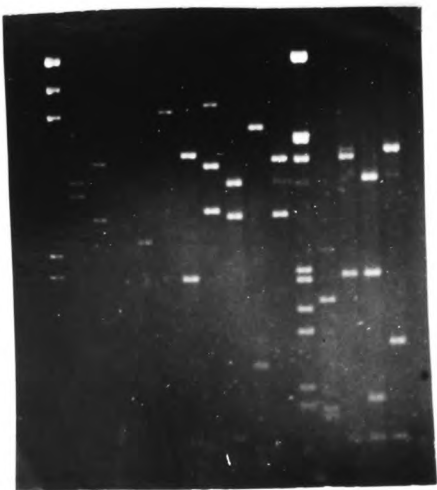
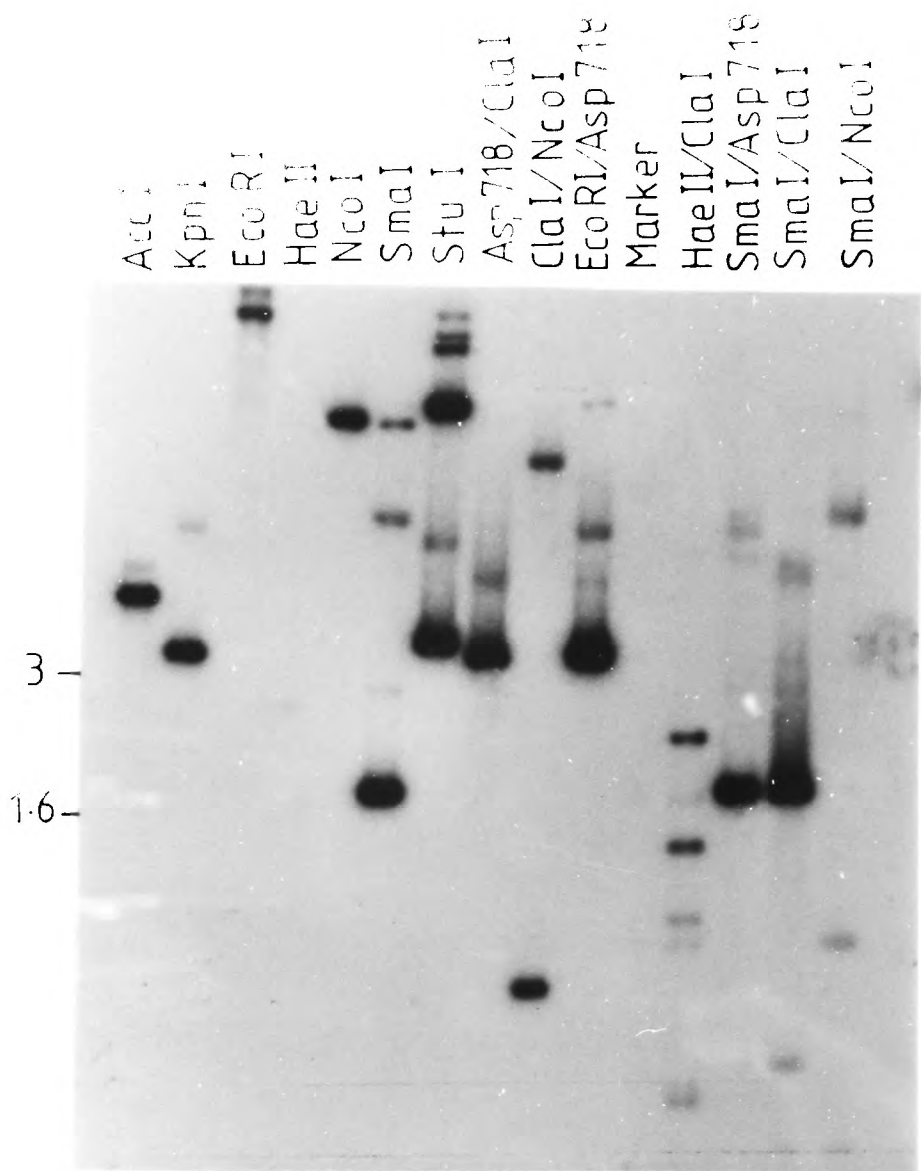


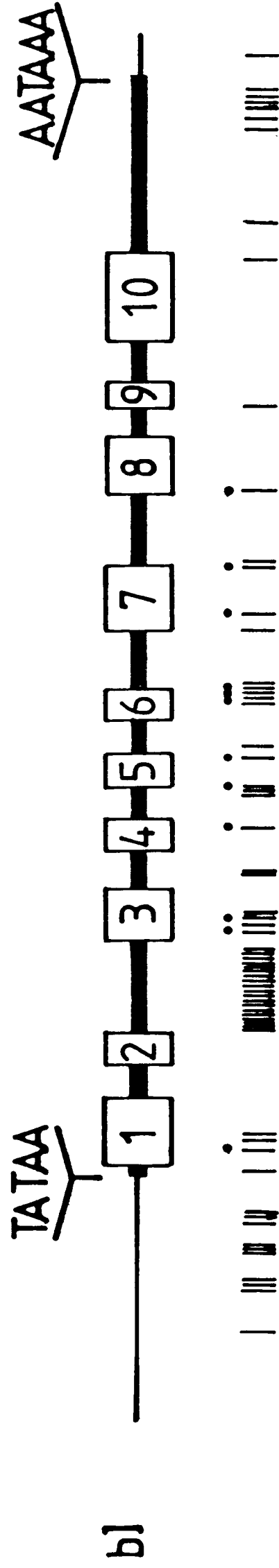
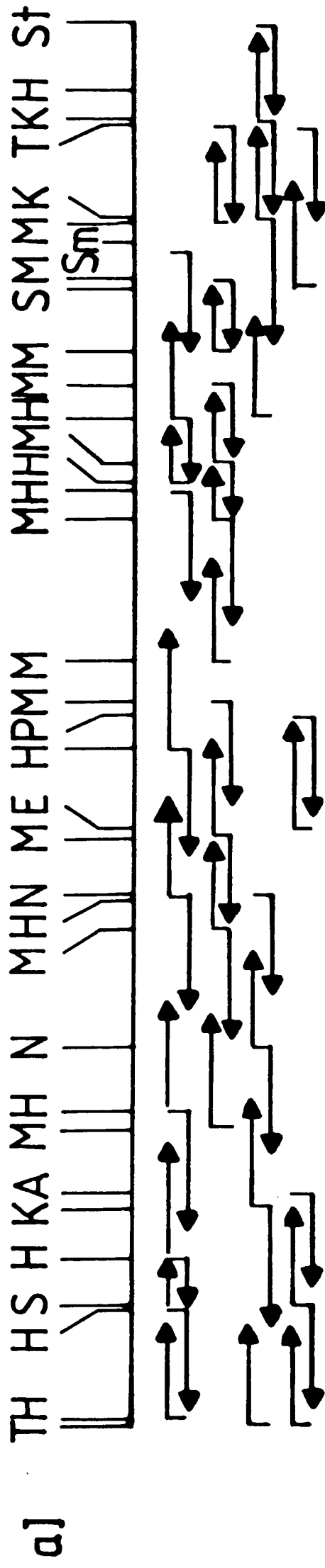
Figure 3.9

(a) Restriction map and sequencing strategy for the 21-OHase B gene is shown beneath the corresponding restriction map. The arrows indicate the extent and orientation in which each fragment was sequenced. The restriction sites are : A : AccI; E, EcoRI; H, HinfI; K; KpnI; M, MspI; N, NcoI; P, PvuII; S, StuI; Sm, SmaI; St, StyI; T, TaqI.

(b) Intron - exon structure of the 21-OHase gene.

The extent of the gene from the putative transcription start site to the polyadenylation site is illustrated by the thick line. Exons are shown boxed and are numbered 1 - 10. The position of the TATAA box and the polydenylation signal AATAAA are shown. The positions at which the 21-OHase A gene differs from the 21-OHase B gene are marked by bars beneath the gene structure. (•) over a bar indicates a codon change due to nucleotide alteration. The three deleterious mutations in the 21-OHase A gene are illustrated below the corresponding bars. The 8-bp deletion and the T insertion, which cause frameshifts, and the C → T transition, which generates a stop codon, are underlined.

200 bp



21-OHase B

21-OHase A

Gly Asp Tyr	Ser Leu Leu Trp	Val Phe Leu Leu	Gln
GGAGAC	TAC TCC CTG CTC TGG	GTT TTT TTT GCTT	CAG
Val Ser Ala	Leu Glu Ser Pro	Val Phe Phe Ala	STOP
GTC TCT GCT	CTG GAA AGC CCA	GTT TTT TTT GCT	TAG

difference in size between the 3.2 kb and 3.7 kb *TaqI* fragments in the 21-0Hase A and 21-0Hase B genes respectively. Similarly an additional *KpnI* site in the 5' flanking region of the 21-0Hase B gene contributes to the differences in size between the 4 kb and 3 kb *KpnI* fragments in the 21-0Hase A and 21-0Hase B genes, respectively. The 21-0Hase A gene has an additional *EcoRI* site ~ 500 bp away from the *EcoRI* site present in both the genes. The 592 bp, 208 bp and 111 bp *MspI* fragments were absent in the *MspI* digests of the 21-0Hase A subclone; instead there were 294 bp, 303 bp and 319 bp fragments present which were probably due to nucleotide differences leading to restriction site changes. A *HinfI* digest was done to yield fragments of the same sizes as those in the 21-0Hase A gene (Section 3.4) (Table 3.2). A *MspI* digest of the subclone overlaps the *HinfI* sites. The *HinfI* fragments were subcloned into M13mp9 and sequenced. The *MspI* fragments were subcloned into M13mp8 and sequenced by I. Dunham. Other fragments that were cloned were *StuI* / *ClaI* (340 bp) and *StuI* / *BamHI* (430 bp) the *ClaI* and *BamHI* sites being in the plasmid vector at the 5' and 3' ends of the insert, respectively. Other digests done were *ClaI* / *Asp718* / *NcoI* to clone 632 bp, 450 bp and 444 bp fragments and *SmaI* / *HinfI* to clone the 507 bp fragment. To clone the 3' end of the gene, an *Asp718* / *StyI* digest was done on the *BglIII* / *BamHI* / *SalI* subclone of the 21-0Hase B gene to clone the 277 bp and 280 bp fragments. All fragments were sequenced in both orientations and at least twice. Figure 3.9a shows the restriction map of 21-0Hase B gene and the fragments sequenced are indicated underneath.

3.6 Structure of the 21-0Hase Genes

The gene structure was initially deduced by comparison with the bovine 21-0Hase cDNA sequence (Yoshioka et al., 1986). There is 78% homology between the bovine and human 21-0Hase amino acid sequences.

Figure 3.10

The 5' flanking region of the 21-OHase gene. The possible transcription initiation signal TATAA is underlined. The three transcription start sites at the -9th, -53rd and -118th nucleotide positions are shown by arrows. The -9th position is probably the major transcription start site.

TCGACAGCTAGATTTCACAGGCTGGAAATCCTGCCCTCCACAAACATCGCGAACAAATACCCGCTGTGCTATAGAGCATGGCTGTGAAGAGTTCACTGAGTGCCCACAAAGCACTTAGACAGCT -625

GTC TGG TAC ATG CTG TAT TTA CT CCG CAG CGG GGA ACC ACT TCT CTT GCT CTG GGC ACT TTT TGT GAG TGA AAG GAG GCA CTA ATA ACAA TCAC ACT GGG ATAC CTG TAT ATAC TGG AA -505

TGCCCCAGGCMAACCAAGGCTTAAACTGTATTACTCTATCTGTAGCTTAAACTAACMAACA-CCCACACAAATCACATTTTCTTCTTCAGGAGGGCTATTAGGCAGGAC -386
A

TGCCAATTTCTCTCTGAGACAAACATCATGCCAGTAACTGGCCACGGTGGGTGGCAGAGGGGAGAGGGGCCAGGTGGGGGGGACACTATTGCCTGCACAGGTGATGTGGAAACCGAA -266

AGCTGACTCTGGATGCAGGAAAAAGCTCAGGGTTCCTTCCTCTCTGCTGATCAATTTTTT-GAAATACGGACGTCCTCCCAAGGCCAATGAGACTGGTGTCTCATTCCTCA -147

[illegible]

CTCAGGGCCCTGACGGGGGTCTCGCCCATGCTGCTCTGGGCCCTGCTGCTCTGCTCTGGCCCCCTGCTGCTGAACTGGTGGAACTGGTGGAACTCCGAGGCCTCCACCTCCC GC 94
↓
I M L L L G L L L L P L L A G A R L L W N W K L R S L H L P T.

The gene structure was confirmed with the published human 21-OHase cDNA sequence which extends from codon 7 to the poly A tail (White et al., 1986). The two 21-OHase genes are ~ 3.3 kb in length and are split into 10 exons by nine introns. All nine introns begin and end with typical splice signals GT-AG. The introns vary from 83 bp to 283 bp in size. Figure 3.9b shows a schematic representation of the intron-exon organization of the genes. As indicated in Table 3.3, the 1st, 2nd and 9th introns occur in codons 68, 98 and 408 after the first nucleotide of the codon triplet, while the 8th intron begins after the 2nd nucleotide of codon 373. The two genes are highly conserved and share 98% homology in the nucleotide sequence of the exons and 96% homology in the introns.

The possible transcription initiation signal TATAA that is postulated to play an important role in determining proper initiation sites for RNA polymerase II (Nussinov, 1986), is located in the 5' region 38 bp upstream of the putative ATG initiation codon in both genes (Figure 3.10). S1 nuclease mapping analysis showed three protected bands of different sizes, each corresponding to the transcription start site at the -9th, -53rd and -118th nucleotide position (Higashi et al., 1986) (Figure 3.11). From the intensity of the protected bands, it appears that the -9th position is the major transcription start site situated 25bp downstream from the TATAA sequence. The other two at -53rd and -118th positions are presumably minor start sites because of the lower intensity of the bands. Upstream from these positions however, there is no apparent TATA sequence or its equivalent.

The polyadenylation signal AATAAA lies 486 bp away from the termination codon in the 10th exon. The cDNA sequence indicates that the polyadenylation signal is 21 bp upstream of the poly(A) tail (White et al., 1986). From these data it appears that the 21-OHase gene codes for a mRNA species of ~ 2 kb and this is similar to that detected by

Table 3.3

<u>Intron No:</u>	<u>Donor Sites</u>	<u>Acceptor Sites</u>	<u>Intron size (bp)</u>
1.	CAAGgtgagagg Q ₆₇ ²⁰³	tctcccagAT D ₆₈ ³⁰⁹	107
2.	ACCTgtaagggc T ₉₇ ³⁹⁰	tcctgcagAC Y ₉₈ ⁶⁷²	283
3.	GAGgtaaggct E ₁₄₉ ⁸²⁷	ctccacagCGC R ₁₅₀ ⁹³³	103
4.	AAGgtgcctca K ₁₈₃ ¹⁰³⁶	tctggcagGAC D ₁₈₄ ¹¹²³	88
5.	AGGgtgaggac R ₂₁₇ ¹²²⁶	tgccgcagTTC F ₂₁₈ ¹³²⁶	101
6.	AAGgtggggac K ₂₄₆ ¹⁴¹⁴	ctctgcagGAG E ₂₄₇ ¹⁵⁸²	169
7.	GAGgtgactcc E ₃₁₃ ¹⁷⁸⁴	ttccccagATT I ₃₁₄ ¹⁹⁸³	200
8.	AGgtgactcc S ₃₇₃ ²¹⁶³	cccttcagCATC I ₃₇₄ ²²⁴⁵	83
9.	CCTGgtatggg P ₄₀₇ ²³⁵⁰	gcccgcagAT D ₄₀₈ ²⁴⁴⁶	97

Capital letters indicate sequences within introns while case letters represent those within exons. Indicated below codons is the encoded amino acid and codon number. The nucleotide number is indicated below the first and last base in the introns.

Figure 3.11

(a) Codon usage and amino acid composition for the 21-OHase protein. The amino acid composition is compared with the expected composition for a protein of this size, based on a number of reference proteins. The value indicates the frequency of presence of a particular amino acid. This data was derived using the ANALYSEQ program (Staden, 1984).

(b) Frequency of GpC residues compared to the frequency of CpG residues occurring in the 21-OHase gene.

a)

F	TTT	19.	S	TCT	12.	Y	TAT	18.	C	TGT	43.
F	TTC	81.	S	TCC	46.	Y	TAC	82.	C	TGC	57.
L	TTA	4.	S	TCA	4.	*	TAA	0.	*	TGA	100.
L	TTG	7.	S	TCG	0.	*	TAG	0.	W	TGG	100.
L	CTT	6.	P	CCT	23.	H	CAT	6.	R	CGT	6.
L	CTC	26.	P	CCC	44.	H	CAC	94.	R	CGC	25.
L	CTA	1.	P	CCA	15.	Q	CAA	19.	R	CGA	6.
L	CTG	56.	P	CCG	18.	Q	CAG	81.	R	CGG	25.
I	ATT	27.	T	ACT	10.	N	AAT	20.	S	AGT	4.
I	ATC	68.	T	ACC	65.	N	AAC	80.	S	AGC	35.
I	ATA	5.	T	ACA	15.	K	AAA	26.	R	AGA	9.
M	ATG	100.	T	ACG	10.	K	AAG	74.	R	AGG	28.
V	GTT	7.	A	GCT	13.	D	GAT	20.	G	GGT	13.
V	GTC	17.	A	GCC	65.	D	GAC	80.	G	GGC	56.
V	GTA	0.	A	GCA	16.	E	GAA	17.	G	GGA	9.
V	GTG	76.	A	GCG	6.	E	GAG	83.	G	GGG	22.

TOTAL CODONS=		496.													
	T	C	A	G											
1	24.65	36.27	37.46	33.10											
2	56.94	21.72	47.00	21.68											
3	18.40	42.01	15.55	45.22											
	-----	-----	-----	-----											
=	100%	100%	100%	100%											
1	14.31	35.69	21.37	28.63	= 100%										
2	33.06	21.37	26.81	18.75	= 100%										
3	10.69	41.33	8.87	39.11	= 100%										
%	19.35	32.80	19.02	28.83	OBSERVED, OVERALL TOTALS										
%	24.79	25.55	24.41	25.24	EXPECTED, EVEN CODONS PER ACID										
	A	C	D	E	F	G	H	I	K	L					
	31.	7.	20.	30.	16.	32.	17.	22.	19.	84.					
O-E %	-27.	-51.	-27.	1.	-10.	-23.	72.	-1.	-42.	129.					
	M	N	P	Q	R	S	T	V	W	Y					
	13.	10.	39.	26.	32.	26.	20.	29.	11.	11.					
O-E %	54.	-53.	52.	35.	32.	-25.	-34.	-11.	71.	-35.					
TOTAL ACIDS=						495.		MOLECULAR WEIGHT=		56028.		HYDROPHOBICITY=		14.4	

b)

GpC
CpG

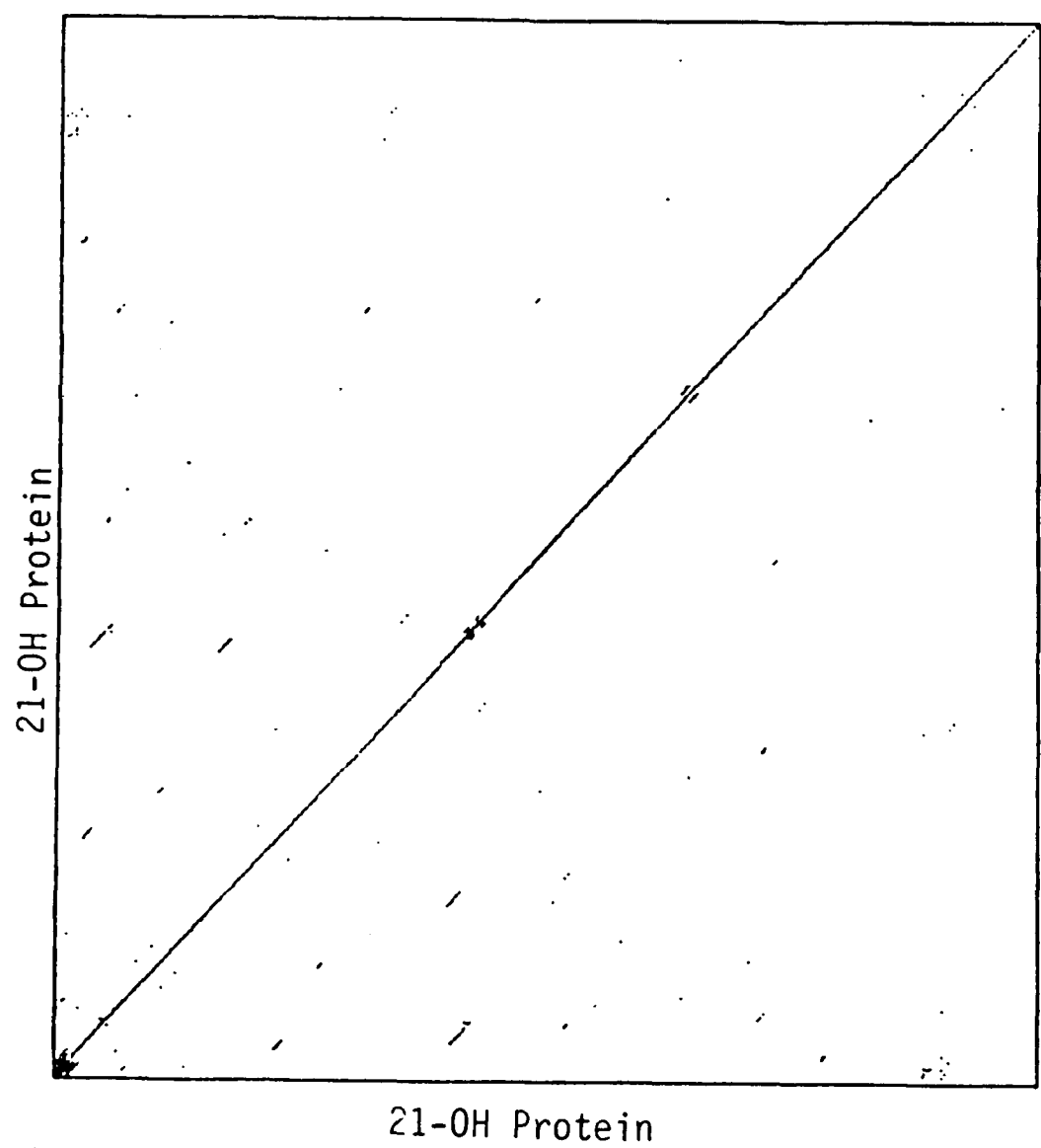


Northern blot analysis of adrenal RNA preparations (Carroll et al., 1985a). The first exon codes for 26 leucines which form the core of a hydrophobic domain that may play a role in anchoring the enzyme to the microsomal membrane. Exon 10 contains the codon for cysteine at position 429 thought to be involved in binding of heme within the enzyme (Morahashi et al., 1984). A search through the nucleotide sequence on the computer programme DIAGON revealed no significant repeats within the sequence. A more detailed search with ANALSEQ which searched for a repeat sequence of 12 bp or more, revealed 8 repeated sequences. The longest repeat is 15 bp and is at positions 534 and 1331 (Table 3.4). Three inverted repeats were found in the sequences (Table 3.4) of 12 bp each.

The derived amino acid sequence of 21-OHase indicated that the proteins is composed of 495 amino acids. This predicted that the molecular weight of human 21-OHase is 56028 which is similar to the porcine (54,000) (Biggin et al., 1983) and bovine (52,000) (White et al., 1984b) 21-OHases, respectively. Computer analysis of codon and amino acid usage for 21-OHase is shown in Figure 3.12. The most striking point is the presence of 84 leucine residues which account for the hydrophobicity of the protein. Other cytochrome P-450s are also rich in leucines (Fujii-Kuriyama et al., 1982). The second observation is that codon usage is not random. For example TTA, CTT and CTA for leucine are rarely used, and the same is true of the TCA and AGT codon for serine. The GTG codon for valine is never used while GTA is most used. There also appears to be a positive discrimination against codons containing CpG in the second and third positions. The TCG (Serine) codon is never used, while CCG (Proline), ACG (Threonine) and GCG (Alanine) are relatively rarely used. Barker et al. (1984) observed that the doublet CpG appears to be a hot spot for mutation. Deamination of 5-methylcytosine

Table 3.4 Inverted Repeats in the 21-OHase nucleotide sequence

(a)	660	2000	12	CTCCTCCTGCAG
	1996	664	12	CTGCAGGAGGAG
	2474	2478	12	CCAGAGCTCTGG
(b)	Long Range Repeats in the 21-OHase nucleotide sequence			
	536	1333	13	CCCAATCCAGGTC
	1285	1806	13	AGGCTCCTTCCCA
	537	1334	12	CCAATCCAGGTC
	1286	1807	12	GGCTCCTTCCCA
	17	20	12	TGCTGCTGCTGC
	534	1331	15	TCCCAATCCAGGTC
	535	1332	14	CCCAATCCAGGTC
	16	19	13	CTGCTGCTGCTGC



leads to the formation of a thymine residue and inefficient correction of this would cause CpG to be a hot spot for mutation. This may act as a selective pressure in the coding sequence, as a mutation in a C residue in the second codon position would alter the amino acid sequence. A search for GpC and CpG strings in the entire 21-OHase gene sequence showed that while it had a high frequency GpC strings, it had relatively fewer CpG strings (Figure 3.11).

3.7 Comparison of the Two 21-OHase Genes

The existence of two 21-OHase genes which were so homologous to each other had led to the speculation that probably there were two separate 21-hydroxylases encoded by separate genes. This was supported by the fact that many patients having impaired 21-hydroxylation of 17-hydroxyprogesterone had little or no impairment in 21-hydroxylation of progesterone (Migeon, 1977; Rosenfield and Miller, 1983). This hypothesis was later disproved on demonstration that a single species of 21-OHase can 21-hydroxylate either progesterone or 17-hydroxyprogesterone in vitro (Kominami et al., 1980). Further proof that only one of the 21-OHase genes in man was vital for 21-hydroxylation was demonstrated in patients with CAH who had a deletion of the 21-OHase B gene while individuals with deleted 21-OHase A genes were hormonally normal (White et al., 1984b).

A comparison of the nucleotide sequences of the 21-OHase A and 21-OHase B genes reveals that of the 4042 nucleotides compared there were 92 nucleotide differences between the two genes. These are depicted in Figures 3.9b and 3.12. The TaqI and KpnI restriction pattern differences were not due to deletions as originally thought (White et al., 1985a), but due to single point mutations in the 5' flanking region of the 21-OHase A gene. Thirteen of the differences in the 21-OHase A

Figure 3.12

Comparison of the nucleotide sequences of the 21-OHase B gene (first line) with a 21-OHase A gene (second line). The complete nucleotide sequence of the 21-OHase B gene and its predicted amino acid sequence are shown. Dashes are introduced to maximise homology. Single nucleotide alteration which cause codon change are marked with an asterisk (*). The premature stop codons that occur in the 21-OHase A gene are underlined. The 8 bp deletion is shown by a dotted line.

TCCTTCCACAGCTGCATTCTCATGCTTCTGCGGCAGTTCTTCCCAATCCAGGCTCCGGAGGCTGAAGCAGGCCNTAGACAGGGATCACATCTGGAGATGCGAGCTGAGGCAGC 1408
 . . . 218 F F P N P G L R R L K Q A I E K R D H I V E M Q L R Q
 C A* A* C A* A*

H K 246 .
ACAAAGTGGGGACTGTATCGTGGACGGCCTCCCTCGGGCCACAGCCAGTGATGCTACCGGCGCTCAGCATTTGCTATGAGCGGGTTCCTTTTGCATACCCACGTTATGGCGCTGTGGCCACT 1528
GT

CTGTACTGCTCTCCCCAGGCCAGCCGCTCAGCCCGCTCCTTTGACCCCTCTGAGAGAGCGCTGTGCGAGGCCCATGCGAGGAGCATGATGCACTACATCTCCAAAGGGGTGGCGCAGCCG 1648

S M E E C S G Q L L E G H V H M A A V D L L I G G T E T T A N T L S W A V V F L
AGCATGGAAGAGGGCTCTGGACAGCTCTGGAGGGCCGTCAGTCAGTGGACCTCTCATGGGTGGCACTGACCAACAGCAACCTCTCTGGGGCCGCTTTT-TT 1767
T*

L H H P E 313 .
GCTTCACCACCCTGAGTGCGTCTCTGGGACACAGCAAGGCTCTTCCACGACCTGGCCAGGGCGGTGGCAGCCTCTACAGCTCTCAGCACTGTCCGCGTGGCGCTGTGCTTGC 1887
C

TCACCGGCACTCAGGCTCACTGGTTGCTGAGGAGCGGCTGGCAGGCTGGCGCTGGGGCAGGACTCCACCCGATCATTTCCCCAGATTTCACGACCGACTCAGGAGGAG 2007
314 I Q Q R L Q E E
T.

L D H E L G P G A S S S R V P Y K D R A R L P L L N A T I A E V L R L R P V P
 CTAGACCAAGAACTGGGCGCTGTGCTCCAGCTCCGGGTCCGACCGGTGCAAGGACCGTCTCTCAATGCCACATCGCCGAGTGTCTGGCTCGGGCGGTGTGCCC 2127

L A L P H R T T R P S S 373 . . . 374 . 374
TTAGCCTTGCCCCACCGGACCAACCGGCCCAGGTCCTCCGAGGGTTGGGGATGAGTCAGGMAAGCCCGAGGCCAGGGAGTCTCTACTCCAGGCCCTTCAGCA 2247

I S G Y D I P E G T V I I P N L Q G A H L D E T V W E R P H E F W P 407
 TCTCCGGCTACGACATCCCTGAGGGCAGAGTCATCTCCAGGCCGCCACCTGGGATGACACGGTCTGGGAGAGGCCACATGAGTTCTGGCCTGGTATGTCGGGGGGCCGGG 2367

GCCTGCCGTGAANAATGTGTGGAGGCTGTCCCGCTGMAACGCTCTCCCAACCAAGCTGTCCAGATCGCTTCTCTGGAGCCACGGCAAGAGACTCCAGAGCTCTGGCC 2487

F G C G A R V C L G E P L A R L E L F V V L T R L L Q A F T L L P S G D A L P S
 TTGGCTGGGTCCCGCTGTGCTGGGGGCGCTGGAGCTTGTGTGCTGACCGAGCTGTCAGGCTTCAGCTGCTGCGCTCCGGGAGCGGCTGCGCTCC 2607

L Q P L P H C S V I L K M Q P F Q V R L Q P R G M G A H S P G Q S Q • 495
CTGCAGCCCTGCCCCACTGCAGTGTCTCAAGATGCAGCCCTTTCAGTGGCGGCTGCAGCCCGGGCCACAGCCAGTGTATGGGGCAGGACCGAT 2727

GCCAGCGGGTACCTCAGTTTCTCCTTTATTGCTCTGTACGAACCCCCTCCCCCTCCCTTCCTCCAGATGCC-^CTGCAGATCCGTGGCAGAGAGGGCTTCTCTCCACGCGGTGGGTGGT 2846^C

GAAGGACCCCTGGCTCTTCTCTCTCGGGGGGACCCCTCAGTCCTGGGCAGCATCAAGCTTTCTTGCTC

AGCTTCATTTCCGCTGAAGGGCAACCGAACTCGAAGGCCCTCCAGTGGTACCAGCTCATTCCCTGGGAAAAGGGTGTGTCAAAGA GAGAGAGTCAAGCCGGATGTGCCCATCTGCTCTCTTCCCGT

TGCCCTTAAGGAGGTAGCTGCCAGCACTGACCAAGCTCCCCCGCAGAGCTCCCTTCTCTGACCCCTCGCTGCAGAGGATTGAGGCTTAATTTCTGAGCTGGCCCTTTCCAGCCATAAATCA 3206

ACTCCAGCTCCCTCTCGAGGCTGGCATGATTGTTCCATTTCACCCAGGCCACTCAGTCCTCTGCTGTTACACTCTGCGGGCTGAAACCTTAGG
C

gene lie in the 5' flanking region upstream of the ATG initiation codon, 42 lie in introns and nine are in the 3' untranslated region. Twenty eight differences are in the exons of which eight are silent and 20 cause amino acid changes. Ten of the codon changes which result in an amino acid change would have an unpredictable effect on the gene product while the remaining 10 differences are quite drastic. These include an 8 bp deletion in the third exon of the 21-OHase A gene at codons 111-113 causing a shift in the reading frame as shown in figure 3.9b. This shift introduces a stop signal (TGA) between codons at positions 133 and 134 (Figure 3.12). Another frameshift occurs in the seventh exon of the 21-OHase A gene at codon 307 due to a 1-bp insertion. A third nonsense mutation occurs in the eighth exon in codon 319 due to a C→T transition (Figures 3.9b and 3.12). The changes in reading frame and the introduction of in phase premature termination codons indicate that the 21-OHase A gene would encode a truncated and presumably nonfunctional protein. Northern blot analysis of adrenal RNA using oligonucleotides specific for 21-OHase A and B genes has shown that only the 21-OHase B gene probe hybridized to a mRNA species (Higashi *et al.*, 1986). A C→T transition at -4 position from the initiation codon in the 21-OHase A gene may affect its expression. Kozak (1984) has reported that the sequence CC^A_GCCAUG is a favorable context for initiation by eucaryotic ribosomes, and a change in that sequence may affect the binding of the 40S ribosomal subunit and hence subsequent initiation.

All these differences indicate that the 21-OHase A gene is a pseudogene and non-expression of the putative A gene product may be due to lack of transcription of the gene, incorrect processing or instability of the mRNA. It does not indicate, however, whether the 21-OHase A gene is nonfunctional in all individuals. The sequence of the three 21-OHase A genes reported so far seems to point out that they all carry the three deleterious mutations. The presence of the duplicated and highly homologous sets of C4 and 21-OHase genes in a small stretch of chromosomal DNA may allow for frequent exchange of their DNA

Figure 3.13

The sequence of the second intron in the 21-OHase B gene (first line) and 21-OHase A gene (second line) gene is shown. Out of the 92 differences between the two genes, 45 are centred in this portion of the gene.

sequences by homologous recombination or unequal crossing-over during meiosis. Most of the base changes between the two 21-OHase genes (45 out of 92) are localized in the second intron and the contiguous 5' part of the third exon (Figure 3.13). The presence of a hot spot in this limited region is reminiscent of a recombination point as suggested for the human fetal γ -globin gene (Slightom et al., 1980).

3.8 Comparison of the Human 21-OHase Gene with Bovine and Murine 21-OHase Genes

The presence of two 21-OHase genes have also been reported for the mouse and bovine genome (White et al., 1984c; Chung et al., 1985). Northern blots of bovine RNA show two major species of bovine 21-OHase mRNA of 2.4 kb and 2.2 kb. Thus both bovine 21-OHase genes may be active, or alternatively there might be multiple polyadenylation sites in the 3' untranslated region of the mRNA or alternate splicing resulting in different sized mRNAs arising from a single gene (Chung et al., 1985). In the mouse H2 complex, the two 21-OHase genes, 21-OHase A and 21-OHase B lie approximately 4 kb 3' of the genes for the murine sex-limited protein and fourth component of complement (C4) (Parker et al., 1985). It was shown that only the 21-OHase A gene was expressed in the adrenal glands and transfected Y1 cells whereas no expression of the 21-OHase B gene was seen. Nucleotide sequence analysis confirms this observation. The murine 21-OHase genes have 10 exons and code for 487 amino acids (Chaplin et al., 1986). The murine 21-OHase B gene has a deletion spanning 215 bases which includes the second exon. There are other nucleotide changes that result in frame shifts and premature termination codons. It was demonstrated that the murine 21-OHase B gene promoter was functional by transfecting a hybrid gene containing the 21-OHase B promoter placed 5' to the 21-OHase A structural gene, into Y1

TABLE 3.5
COMPARISON OF THE HUMAN 21-OHASE AMINO ACID SEQUENCE WITH THE MURINE AND BOVINE 21-OHASE SEQUENCE

Hu	M L L L G L L L L L P L L A G A R L L W N W K L R S L H L P P L A P G F L H L L Q P D L P I Y L L G L T Q K
Mu	M L L P G L L L L L L L L A G T R W L W G Q W K L R K L H L P R L A P G F L E F L Q P N L P I Y L L G L T Q K
Bo	M V L A G L L L L L T L L A G A H L L W G R W K L R N L H L P P L V P G F L H L L Q P N L P I H L L S L T Q K
Hu	Δ F G P I Y R L H L G L Q D V V V L N S K R T I E E A M V K K W A D F A G R P E P L T Y K L V S R N Y P D L S
Mu	L G P I Y R I H L G M Q D V V L L N S N R T I E E A L I Q K W V D F A G R P E M L N G K M D L D L S
Bo	L G P V Y R L R L G L Q E V V V L N S K R T I E E A M I R K W V D K A G R P Q I P S Y K L V S Q R C Q D I L S
Hu	L G D Y S L L W K A H K K L T R S A L L A G I R D S M E P V V E Q L S T Q E F C E R M R A Q P Q T P V A I E E
Mu	L G D Y S L M W K A H K L S R S A L M L G M R D S M E P L I E Q L T Q E F C E R M R A Q A G T P V A I H K
Bo	L G D Y S L L W K A H K K L T R S A L L L G T R S S M E P W V D Q L T Q E F C E R M R V Q A G A P V T I Q K
Hu	E F S L L T C S I I C Y L T F G D K I K D D N L M P A Y Y K C I Q E V L K T W S H W S I Q I V D V I P F L R F
Mu	E F S P L T C S I I C L T F G D L N S T L B Q T L H D C V Q D L L Q A W N H W S I Q I L T I I P L L R F
Bo	E F S L L T C S I I C Y L T F G D K E D T L V H A F H D C V Q D L M K T W D H W S I Q I L D M V P F L R F
Hu	F P N P G L R R L K Q A I E K R D H I V E M Q L R Q H K E S L V A G Q W R D M M D Y M L Q G V A Q P S M E E G
Mu	L P N P G L Q K L K Q I Q E S R D H I V K Q Q L K Q H K E S L V A G Q W K D M I D Y M L Q G V E K Q R D G K D
Bo	F P N P G L W R L K Q A I E N R D H M V E K Q L T R H K E S M V A G Q W R D M T D Y M L Q G V G R Q R V E E G
Hu	S G Q L L E G H V H M A A V D L L I G G T E T T A N T L S W A V V F L L H H P E I Q Q R L Q E E L D H E L G P
Mu	E E Q L E E G H V H M S V V D L F I G G T E T T A T T L S W A V A F L L H H P E I Q K R L Q E E L D L K L G P
Bo	P G Q L L E G H V H M S V V D L F I G G T E T T A S T L S W A Y A F L L H H D E I Q R R L Q E E L D R E L G P
Hu	G A S S S R V P Y K D R A R L P L L N A T I A E V L R L R P V V P L A L P H R T T R P S S I S G Y D I P E G T
Mu	G S Q L L T R N R M Q L P L L N A T I A E V L R L R P V V P L A L P H R A T R A S S I S G Y D I P K D M
Bo	G A S C S R V T Y K D R A R L P L L N A T I A E V L R L R P V V P L A L P H R T T R P S S I F G Y D I P E G M
Hu	V I I P N L Q G A H L D E T V W E R P H E F W P D R F L E P G K N S R A L A F G C G A R V C L G E P L A R L E
Mu	V I I P N I Q G A N L D E M V W E L A S K F W P D R F L E P G K N P R T P S F G C G A R V C L G E P L A R L E
Bo	V V I P N L Q G A H L D E L V W E Q P H E F R P D R F L E P G A N P S A L A F G C G A R V C L G E S L A K L E
Hu	L F V V L T R L L Q A F T L L P S G D A L P S L Q P L P H C S V I L K M Q P F Q V R L Q P R G M G A H S P G Q S Q
Mu	L F V V L A R L L Q A F T L L P P P D G T L P S L Q P Q P Y A G I N L P I P P F Q V R L Q P R N L A P Q D Q G E R P
Bo	L F V V L L R L L Q A F T L L P P P V G A L P S L Q P D P Y C G V N L K V Q P F Q V R L Q P R G V G A G A W E S A S A Q

Hu: Human; Mu: Murine; Bo: Bovine.
* : Conserved in all; ♦ : Conserved in Hu & Mu; # : Conserved in Hu & Bo; Δ : Conserved in Mu & Bo; Blank: No Homology.
† : Indicates introns

cells and demonstrating transcripts of 21-OHase mRNA. This clearly indicated that the murine 21-OHase B gene was a pseudogene and lack of its expression resulted from mutations downstream of the promoter.

The sequences of one bovine 21-OHase gene was reported by Chung et al. (1986). It is 3.4 kb in length, has 10 exons and encodes for a protein of 496 amino acids. A comparison of the protein sequence of human, murine and bovine 21-OHase is shown in Table 3.5. The exons in all three species are located in the same position. There is 64% conservation of amino acids in all the three species together. Between human and bovine 21-OHase, the homology is 78% and between human and murine 21-OHase it is 70%, respectively. It is likely that those residues which are not conserved may not be essential for the function of the protein.

The major histocompatibility complexes of humans and mice being remarkably similar, it is possible that the duplication of primordial 21-OHase gene and subsequent inactivation of one gene occurred prior to speciation. The bovine genome also has two 21-OHase genes and they could both be functional, as two distinct 21-OHase mRNA species have been found in the adrenal cortex (Chung et al., 1985). Also sequence analysis has shown that the nature of the deletions that prevent expression is different in the murine and human 21-OHase pseudogenes. While the sequence homology between the two 21-OHase genes in humans is 98% and in mice is 86%, respectively, the homology between the two species is only about 70% at the amino acid level. This suggests that inactivation of one of the 21-OHase genes in mouse and in man must have occurred after mammalian speciation about 85 million years ago.

3.9 Comparison of the 21-OHase Genes with other Cytochrome P-450 Genes

The cytochrome P-450 genes belong to a superfamily of genes which

TABLE 3.6 Comparison of the amino acid sequences of :
 (a) human 21-OHase with other species at the NH₂ terminus.
 (b) human 21-OHase with other cytochrome P-450 at the NH₂

(a)	1				5					10				15
Porcine 21-OHase (Yuan <u>et al.</u> , 1983)	M	V	L	V	W	L	L	L	L	T	T	L	K	A G
Murine 21-OHase (Chaplin <u>et al.</u> , 1986)	M	L	L	P	G	L	L	L	L	L	L	L	L	A G
Bovine 21-OHase (Chung <u>et al.</u> , 1986)	M	V	L	A	G	L	L	L	L	L	T	L	L	A G
Human 21-OHase B (This work)	M	L	L	L	G	L	L	L	L	L	P	L	L	A G
Human 21-OHase B (Higashi <u>et al.</u> , 1986; White <u>et al.</u> , 1986)	M	L	L	L	G	-	L	L	L	L	P	L	L	A G
							↑							
(b)														
Rat pb P-450e	M	E	P	T	I	-	L	L	L	L	A	L	L	V G
Rat mc P-45d	M	A	F	S	Q	Y	I	S	L	A	P	E	L	L A T

A dash (-) indicates absence of a codon.
 ↑ indicates the position at which the 21-OHase cDNA sequence begins.

is speculated to have as many as 10 gene families. Their substrate specificities vary widely and despite certain similarities in function, these enzymes have remarkably little homology to each other (Adesnik and Atchison, 1986). Microsomal monooxygenase systems containing cytochrome P-450 play an important role in the metabolism of endogenous substrates such as steroids and fatty acids, in the detoxification of many drugs and xenobiotics, and in the activation of environmental agents to toxic, mutagenic and carcinogenic forms. The extraordinarily broad substrate specificity of this enzyme system results from the multiplicity of different molecular forms of the terminal oxidase, cytochrome P-450, which have different but overlapping broad substrate specificities. Another important property of several of the forms of cytochrome P-450 is that they are inducible. Many of the well characterized forms of cytochrome P-450 are the ones which are induced by phenobarbital (pb) and 3-methylcholanthrene (mc). These are very different in their enzymatic, structural and immunological properties. The gene structure for two types each of phenobarbital and methylcholanthrene-inducible cytochrome P-450s have been reported (P-450b and -e ; and P-450c and -d respectively) (Suwa et al., 1985; Mizukami et al., 1983; Sogawa et al., 1984, 1985; Gonzalez et al., 1985; Hines et al., 1985). A comparison of the amino acid sequence of 21-OHase with pb inducible P-450e shows 28% homology and a 27% homology with mc inducible P-450d. This finding supports the idea that 21-OHase belongs to the same superfamily of genes but not the same family of genes as the drug inducible cytochrome P-450s (Dayhoff et al., 1978).

There are a few features that are common to the P-450 enzymes. There is a functional conservation of a leucine-rich hydrophobic region at the amino acid terminus that may be involved in the interaction of the enzyme with the endoplasmic reticulum. Table 3.6

shows the amino acid sequences of 21-OHase at the amino terminus of different species. It is evident that the hydrophobic tail is conserved in the 21-OHase of human, bovine, murine and porcine origin (Rodrigues et al., 1987; Yuan et al., 1983; Chaplin et al., 1986; Chung et al., 1985; Higashi et al., 1986; White et al., 1986). The hydrophobic tail is also apparent in the amino terminus of mc P-450d and pb P-450e (Table 3.6). Two other regions are conserved in the cytochrome P-450s. One of them is a conserved tridecapeptide, first described as conserved for phenobarbital-inducible P-450s by Ozols et al. (1981). This tridecapeptide is conserved in cytochrome P-450s belonging to the same subfamily and occurs immediately after a variable region which accounts for the amino acid differences between two closely related cytochrome P-450s. For example the pb-induced cytochrome P-450 2 & 4 differ in 14 nucleotides. Seven of them result in the replacement of 6 amino acids in the variable region just prior to the tridecapeptide which results ^{the} in production of two distinct proteins (Fujii-Kuriyama et al., 1982). A comparison of the sequence of the human 21-OHase with bovine 17 α -hydroxylase shows 50% overall homology which means that they do not belong to the same family of proteins (Dayhoff et al., 1978). However, the homology between 21-OHase and 17 α -OHase in the tridecapeptide is 74% (Zuber et al., 1986) (Table 3.7). The two cytochrome P-450s have the substrate progesterone in common so perhaps this particular region is important in establishing substrate specificity.

The third conserved region in cytochrome P-450s is the possible heme-binding peptide which is common to all cytochrome P-450 enzymes (Morohashi et al., 1984). It is in general agreement that a thiolated cysteine co-ordinates to the heme of cytochrome P-450 as the fifth ligand. Since the optical and magnetic properties of various forms of cytochrome P-450 are very similar, the primary structure as well as the

TABLE 3.7 Comparison of the amino acid sequences of human 21-OHase (I) and bovine 17 α -OHase (II) in the conserved tridecapeptide region.

	339																					361	
I	D	R	A	R	L	P	L	L	N	A	T	I	A	E	V	L	R	L	R	P	V	V	P
II	D	R	N	R	L	V	L	L	E	A	T	I	R	E	V	L	R	I	R	P	V	A	P
	346																						368

| : indicates conserved residues.

conformation around the heme seems to be well conserved among cytochrome P-450s. A comparison of the heme binding peptide sequences in several cytochrome P-450s is shown in Table 3.8. The cysteine residue is conserved in all the forms of cytochrome P-450s and there is a 60% homology between all the forms compared.

Although 21-OHase shares homology with other P-450 enzymes, their gene structures differ grossly. The two types of P-450, phenobarbital- and methylcholanthrene-inducible are supposed from their sequence homology to have originated by divergent evolution from a common ancestor. The pb P-450 genes span 14-23 kb and code for a protein of 491 amino acids. The mc P-450 genes span 6-7 kb and code for a protein of 513-524 amino acids (Sogawa et al., 1984). The pb P-450 gene has 9 exons and the mc P-450 gene has 7 exons, respectively. The intron sizes in the pb P-450 gene in the order of occurrence are 3.2 kb, 0.3 kb, 2.3 kb, 0.8 kb, 0.5 kb, 1.6 kb, 0.3 kb, and 3.2 kb respectively. The intron sizes in the mc P-450 gene are 2.5 kb, 0.5 kb, 84 bp, 93 bp, 0.15 kb and 0.15 kb respectively. When the sites of introns are correlated with 21-OHase which has 10 exons, based on an alignment of amino acid sequences that maximises homology, no intron is located in exactly the same place in any two enzymes (Figure 3.14). If possible shifts of intron locations due to deletion or insertion of coding sequences as reported for α -fetoprotein and albumin genes (Kioussis et al., 1981) are taken into consideration, it would appear that 21-OHase shares two sites of introns with each of the two genes i.e. the 2nd and 5th introns of 21-OHase corresponding to the 2nd and 4th intron of pb P-450e and the 6th and 8th introns of 21-OHase gene to the 2nd and 5th introns in the mc P-450d gene, respectively (Figure 3.14). The finding that 21-OHase gene has possible common sites for introns with either of the two drug-inducible P-450 is consistent with the notion that these cytochrome

TABLE 3.8

1	F	G	C	G	A	R	V	⁴²⁹ C	L	G	E	P	L	A	R	L	E	L	F	V	V	L	T	R	L	L	Q	A	F
2	F	G	A	G	P	R	S	⁴⁴² C	V	G	E	M	L	A	R	Q	E	L	F	L	F	M	S	R	L	L	Q	R	F
3	F	G	W	G	V	R	Q	⁴⁶¹ C	V	G	R	R	I	A	E	L	E	M	T	L	F	L	I	H	I	L	E	N	F
4	F	G	L	G	K	R	R	⁴⁵⁶ C	I	G	E	I	P	A	K	W	E	V	F	L	F	L	A	I	L	L	H	Q	L
5	F	S	T	G	K	R	I	⁴³⁶ C	L	G	E	G	I	A	R	N	E	L	F	L	F	T	T	I	L	Q	N	F	
6	F	S	I	G	Q	P	N	⁴³⁶ C	F	G	E	G	L	A	R	M	E	L	F	L	F	T	T	V	M	Q	N	F	

Comparison of aminoacid sequences of different cytochrome P-450s in the conserved possible heme binding region. The cysteine residue is conserved in all cytochrome P-450s.
1: human 21-OHase (Rodrigues et al.,1987); 2: bovine 17 α -OHase (Zuber et al.,1986);
3: bovine side-chain cleavage cytochrome P-450 (Morohashi et al.,1984); 4: methylcholanthrene inducible rat cytochrome P-450d (Kawajiri et al.,1984); 5: phenobarbital inducible rat cytochrome P-450e (Mizukami et al.,1983); 6: human cytochrome P-450 (Phillips et al.,1985).

Figure 3.14

Comparison of the gene structure of 21-OHase with pb P-450e and mc P-450d. The exons are shown as boxes and the introns as gaps. The size of the introns is indicated under the boxes. The first and the last codon of the exons is indicated on the top of the boxes.



P-450s might have separated from a common ancestral gene. Alternatively, convergent evolution of the cytochrome P-450s from different ancestors cannot be absolutely ruled out. Obviously elucidation of the gene structure of other types of P-450 would enable to understand the origins of the introns and the evolutionary process of diversification in the P-450 genes.

3.10 Polymorphism of the 21-OHase Genes

The presence of allelic variants of 21-OHase at a single locus is strongly suggested by the four different clinical manifestations of CAH. The differential linkage disequilibrium of classical and non-classical CAH to different haplotypes seems to indicate that the less severe form of CAH is an allelic variant of the classical form. The 21-OHase genes lie between a region of the HLA which is highly polymorphic. They lie immediately 3' to the C4A and C4B genes which show a high degree of polymorphism (Yu et al., 1986). More than 40 alleles have been demonstrated in the two classes by differences in electrophoretic mobilities (Mauff et al., 1983), serological typing (Roos et al., 1984 ; Giles et al., 1984 ; Giles, 1987) and direct DNA sequencing (Belt et al., 1984, 1985 ; Yu et al., 1986) It is quite possible that the 21-OHase gene might exhibit a similar polymorphism. Being a microsomal protein, isolation of 21-OHase is difficult and polymorphism can only be inferred from DNA sequencing data. A comparison of 21-OHase gene sequences (Rodrigues et al., 1987) with sequences from other groups (Higashi et al., 1986) suggests polymorphism in 21-OHase genes. As shown in Table 3.9, 77 differences were found in the sequences compared. Most of the differences may represent accumulation of point mutations in the pseudo-gene. Of the 16 differences in the 21-OHase B gene sequence which lie in exons, six cause a codon change. The leucine at codon 6 (CTG) was

TABLE 3.9 Differences in the published 21-OHase nucleotide sequences.

No.	POS.	B.1	Higashi et al.	White et al.	A.1	Higashi et al.	White et al.
			B.2	B.3		A.2	A.3
1	-356	G	T		G	T	
2	-352	G	-		G	G	
3	-335	G	-		G	-	
4	-282	G	-		G	-	
5	-95	T	T	C	T	T	T
6	-8	G	G	G	G	G	T
7	-4	C	C	C	T	T	C
8	28	*C	-	-	C	C	C
9	29	*T	-	-	T	T	T
10	30	*G	-	-	G	G	G
11	118	T	T	C	C	C	C
12	138	A	A	C	C	C	C
13	274	C	C	C	C	C	-
14	422	A	C	C	C	C	C
15	456	C	T	C	T	T	T
16	532	C	C	C	C	C	G
17	533	A	A	A	T	T	CT
18	562	G	G	CG	G	G	G
19	568	G	-	G	-	-	-
20	576	G	G	G	T	T	G
21	577	G	G	G	C	C	T
22	578	T	T	T	A	A	C
23	579	G	G	G	G	G	A
24	581	G	G	G	A	A	G
25	584	G	G	G	A	A	G
26	585	A	A	A	G	G	A
27	587	A	A	A	A	A	AGA
28	589	C	C	C	C	C	A
29	590	A	A	T	A	A	T
30	597	G	G	G	G	G	A
31	598	G	G	G	A	A	G
32	604	C	C	C	C	C	A
33	605	C	C	C	A	A	C
34	653	C	C	C	C	C	G
35	659	C	C	A	G	G	G
36	687	*G	A	A	A	A	A
37	701	C	C	T	C	C	T
38	864	T	C	C	C	C	C
39	865	G	G	G	A	G	G
40	892	C	-	C	C	-	C
41	911	G	G	GG	G	G	G
42	1208	G	G	G	A	G	G
43	1260	A	A	A	G	A	A
44	1423	G	G	C	G	G	C
45	1425	A	A	A	G	G	A
46	1426	C	C	C	T	T	C
47	1647	C	C	C	T	C	C
48	1688	G	G	G	G	T	T

No.	POS.	B.1	Higashi et al.		White et al.		A.1	Higashi et al.		White et al.	
			B.2	B.3				A.2	A.3		
49	1895	C	C	CC			C	C	CC		
50	1902	G	G	G			G	G	GC		
51	1911	G	G	G			G	G	C		
52	2113	C	C	C			C	T	T		
53	2252	C	C	T			T	C	T		
54	2361	G	-	G			G	-	G		
55	2375	G	G	G			G	A	G		
56	2504	G	C	G			G	C	G		
57	2505	*C	G	C			C	G	C		
58	2697	G	A	A			A	A	A		
59	2705	*G	A	A			A	A	A		
60	2764	T	C	C			C	C	C		
61	2801	C	C	C			CC	C	C		
62	2803	G	G	GG			G	G	GG		
63	2911	G	G	G			G	G	-		
64	2913	G	G	-			G	G	-		
65	2914	A	A	-			A	A	-		
66	2939	C	C	-			C	C	-		
67	2940	C	C	-			C	C	-		
68	3119	C	C	-			C	C	-		
69	3120	C	C	-			C	C	-		
70	3121	A	A	-			A	A	-		
71	3122	A	A	-			A	A	-		
72	3150	C	C	C			T	T	C		
73	3177	T	T	T			C	C	T		
74	3187	C	C	C			T	T	C		
75	3197	C	C	T			C	C	T		
76	3198	C	C	T			C	C	T		
77	3288	T	T				C	T			

B and A correspond to the 21-OHase B and 21-OHase A genes, respectively.

B.1, A.1 - the sequences reported here

B.2, A.2 - the sequences reported by Higashi et al. (1986)

B.3, A.3 - the sequences reported by White et al. (1986).

A blank space indicates that sequence data was not available at those positions.

A dash (-) indicates absence of a nucleotide at that position.

An asterisk (*) indicates a codon change.

absent in the sequences from the other two groups. Unfortunately the published cDNA sequence of 21-OHase ends at codon 7 (Table 3.6) (White *et al.*, 1986) and it has not been possible to define whether this amino acid is present in the expressed sequence. However, a comparison of 21-OHase sequences from other species with the human 21-OHase reveals a codon for leucine at the same position (Table 3.6). This may indicate that there is polymorphism in the size of 21-OHase with either 494 or 495 amino acids. Figure 3.15 illustrates the differences observed in the codons. At codon 103 (AGG) an arginine residue was seen in the 21-OHase B gene sequence corresponding to a lysine (AAG) at that position in the other two published sequences. At codon 427 (CGC) an arginine residue in the sequence (Rodrigues *et al.*, 1987) and that of White *et al.* (1986) and Mattesson *et al.* (1987) differs from that of Higashi *et al.* (1986) who identified a proline residue (CCG). At codon 494 (AGC) a serine was observed in 21-OHase B gene sequence corresponding to a asparagine (AAC) in the other two published sequences. Originally this serine was thought to be restricted only to the 21-OHase B gene in individual II.1 and thought to be a deleterious mutation. However, its presence in the normal 21-OHase B gene sequence suggests it is not a mutation but a polymorphism. The differences observed between 21-OHase A genes were 1.4% and between 21-OHase B genes were 0.94%. While sequencing errors cannot be absolutely ruled out, many of the differences appear to be genuine cases of polymorphism arising from genetic drift.

3.11 Discussion

The two sets of C4 and 21-OHase genes probably arose by duplication of a single pair of genes at some time in the past. Although both C4 genes have remained functional in humans, only one 21-OHase gene is

Figure 3.15

21-OHase B present work	6 L <u>CTG</u>	103 R <u>AGC</u>	427 R <u>CGC</u>	494 S <u>AGC</u>
21-OHase B (White et al., 1986)	---	AAG	CGC	AAC
21-OHase B (Higashi et al., 1986)	---	AAG K	CCG P	AAC N
21-OHase A	CTG	AAG	CGC	AAC

Schematic representation of the polymorphism observed in the 21-OHase B genes. Nucleotide differences that result in codon change are underlined. Sequences of three 21-OHase B genes and three 21-OHase A genes are compared, and codons are indicated underneath the sequence observed.

necessary for normal steroidogenesis; thus one of the genes, the 21-OHase A gene in humans and the 21-OHase B gene in mice accumulated deleterious mutations with no apparent ill effects on the organism.

Comparison of the nucleotide sequences of the 21-OHase A and 21-OHase B genes shows that they are highly conserved and share 98% sequence homology in the exons and 96% homology in the introns. Of the two genes, however, only the 21-OHase B gene encodes an active protein. The 21-OHase A gene appears to be nonfunctional as individuals with a homozygous deletion of this gene synthesize cortisol normally whereas individuals with a homozygous deletion of the 21-OHase B gene have classical CAH (White et al., 1984a). The sequence analysis confirms this observation. The 8bp deletion in the third exon, the 1bp insertion in the seventh exon and C→T transition in the eighth exon of the 21-OHase A gene generate frameshift and / or nonsense mutations which would prevent synthesis of a functional protein. That the 21-OHase A gene is a pseudogene is further substantiated by the absence of detectable 21-OHase A mRNA in a Northern blot analysis of adrenal RNA using oligonucleotide probes specific for the A and B gene sequences (Higashi et al., 1986).

Sequence comparison of the 21-OHase B gene with other published 21-OHase B gene sequences suggests that the 21-OHase genes are polymorphic. An extra codon for leucine at position 6 in some 21-OHase B genes would result in two species of 21-OHase of 494 or 495 amino acids with no obvious ill effect on its function. Most of the differences observed were in the 21-OHase A genes which probably indicate accumulation of point mutations in the pseudogenes. The arginine at codon 103 and serine at codon 494 are not conserved in the murine and bovine 21-OHase, therefore any change in those residues may not affect function. The arginine at codon 427, however, is conserved in all three

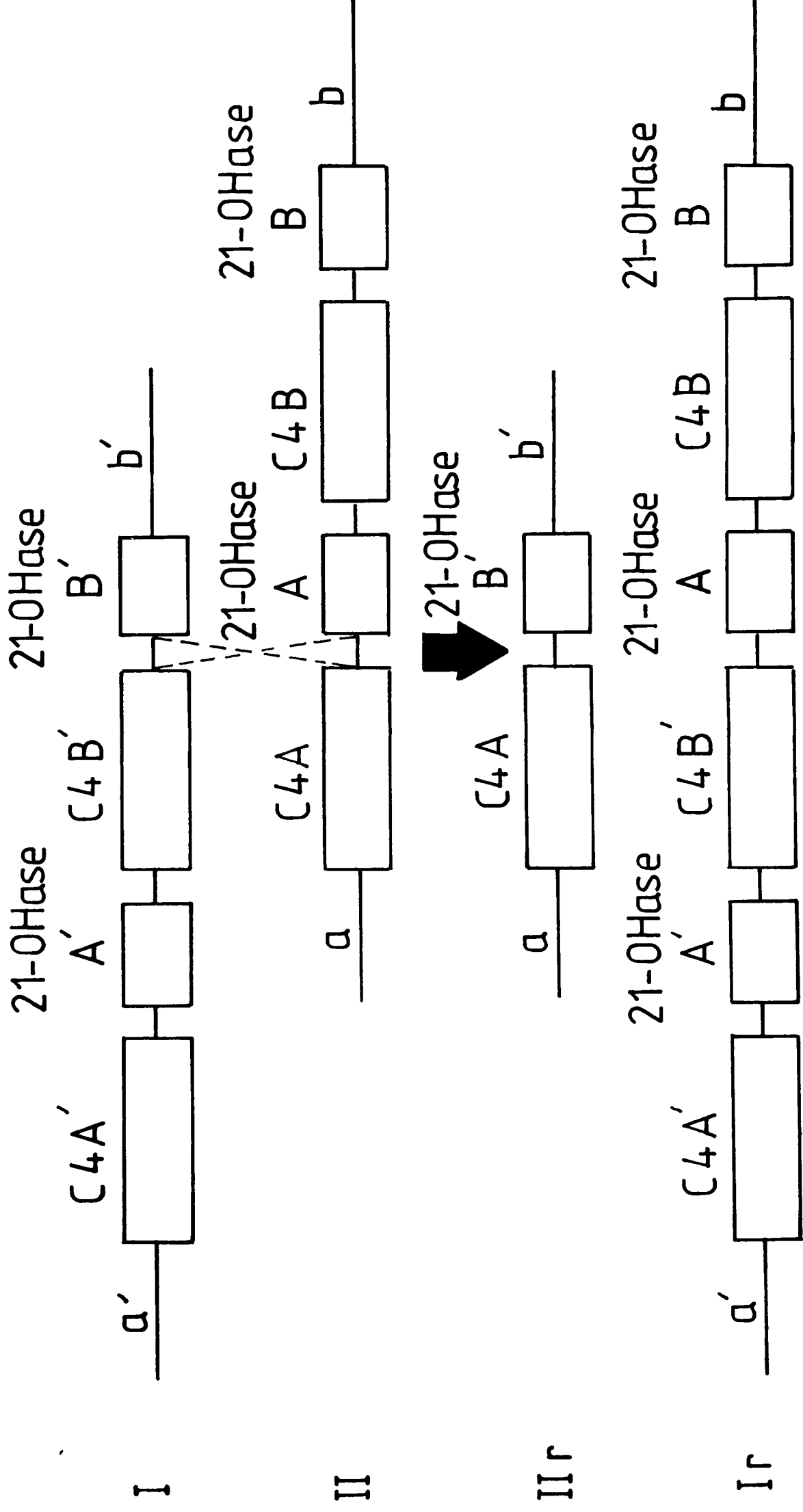
species. It cannot therefore be predicted how the change to proline might affect the function of the protein. It is likely that such polymorphic variations may give rise to allelic variants of 21-OHase which manifest as different forms of CAH.

The possibility of misalignment during meiotic metaphase and unequal crossing over between sister chromatids is increased due to the tandem duplication of the C4 and 21-OHase genes resulting in chromosomes containing one or three sets of C4 and 21-OHase genes. This mechanism presumably is the cause of the rearrangements observed in the HLA-A1 B8 DR3 haplotype in individual KEM and the A3 Bw47 DR7 haplotype that have, respectively, deleted C4A - 21-OHase A and C4B - 21-OHase B pairs of genes (Carroll et al., 1985b). The HLA-B14 DR1 haplotype which was studied has three C4 and three 21-OHase genes and is associated with nonclassical 21-OHase deficiency. The extra C4 gene is a C4B gene and the extra 21-OHase gene is a 21-OHase A gene. If the extra 21-OHase A gene is a pseudogene like the first 21-OHase A gene on the same haplotype, then it obviously does not contribute to the development of the phenotype for nonclassical 21-OHase deficiency, but is merely an associated marker. No clone containing a 21-OHase B gene was isolated from the cosmid library hence the gene could not be characterized.

Variation in gene copy number has been reported in other multiple gene families such as the α -globin genes. Unequal crossing over would explain both duplication and deletion of the C4 and 21-OHase loci as has been suggested for the α -globin genes (Dozy et al., 1979 ; Orkin et al., 1979). By this mechanism, misalignment during meiosis between homologous chromosomes containing C4A - 21-OHase A and C4B - 21-OHase B genes each could result in an exchange of one locus leaving one chromosome with three C4 and three 21-OHase genes and the donor with only one of each (Figure 3.16). The molecular map of B14 is consistent with this

Figure 3.16

Model describing unequal cross-over between two homologous chromosomes resulting in deletion of 21-OHase A and C4B on one chromosome (Iir) and duplication of 21-OHase A and C4B on the other chromosome (Ir). Dashed X indicates the proposed region of cross-over.



type of mechanism. Haplotypes with a null allele, that is only one active C4 gene would represent the donor chromosome. The model does not rule out the possibility of the extra 21-OHase A gene being a recombination between 21-OHase A and 21-OHase B resulting in a protein with reduced 21-OHase activity. It would be interesting and informative to characterize the other two 21-OHase genes on the B14 haplotype to understand the role they play in individuals with late onset CAH.

A comparison of the 21-OHase gene with other cytochrome P-450s shows that it shares all the three characteristic traits of cytochrome P-450s viz. a hydrophobic amino terminus; a conserved tridecapeptide with another cytochrome P-450, 17 α -OHase with whom it shares the common substrate progesterone; and the conserved heme binding peptide containing the cysteine residue which is conserved in all cytochrome P-450s. The structure and size of the 21-OHase gene differs greatly from the other cytochrome P-450s. The number of introns vary in the different forms. Therefore despite the homology shared between the different forms, it is unclear whether they have evolved from a common ancestor or alternatively convergent evolution has taken place.

The DNA sequence obtained from the 21-OHase genes lays a foundation for the study of 21-OHase genes from patients suffering from 21-OHase deficiency. It also shows that 21-OHase belongs to the cytochrome P-450 superfamily, but is in a distinct family of its own.

CHAPTER IV

CHARACTERIZATION OF A 21-HYDROXYLASE B GENE FROM A PATIENT WITH CONGENITAL ADRENAL HYPERPLASIA

4.1 INTRODUCTION :

The four distinct clinical manifestations of 21-OHase deficiency that have been identified; simple virilising, salt-wasting, late onset and cryptic, show linkage disequilibrium to different HLA haplotypes. Although patients homozygous for the HLA haplotype A3, Bw47, DR7 show a deletion of the 21-OHase B gene (White et al., 1984b), the majority of patients show the presence of at least one 21-OHase B gene (Rumsby et al., 1986). Obviously the 21-OHase B gene that is present on at least one chromosome is non-functional or the expressed product partly or wholly inactive. The differences in clinical manifestation of the deficiency strongly suggest the presence of allelic variants of 21-OHase at a single locus. It is important to study this serious inborn error of metabolism to define molecular markers for the presence of particular disease genes which can be used in prenatal diagnosis, since most affected individuals are capable of leading a normal life if detected and treated early. Although it is not practical to analyse the 21-OHase genes from all affected individuals, it is necessary to categorise the different forms of the deficiency and to define a common denominator that can be used in characterization, prognosis and treatment of the disease. To understand the molecular basis of CAH in one such patient with the salt-wasting form of the disease, his single 21-OHase B gene was cloned and characterized. This chapter presents the studies on individual II.I's family and the complete characterization of his 21-OHase B gene.

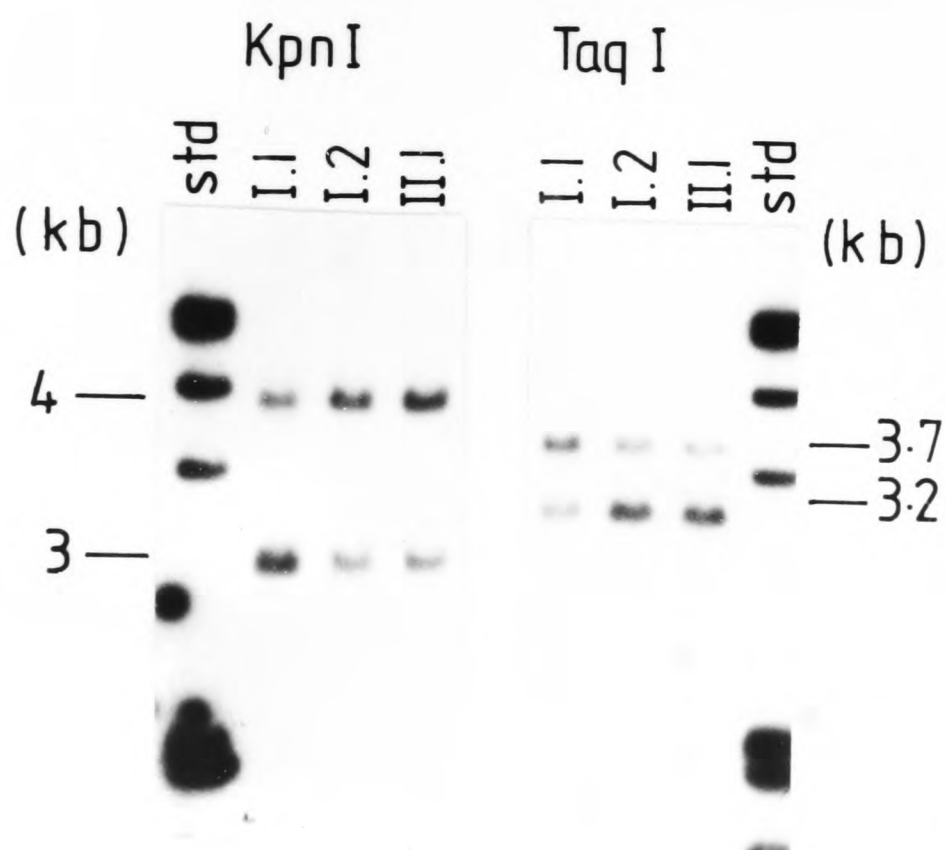
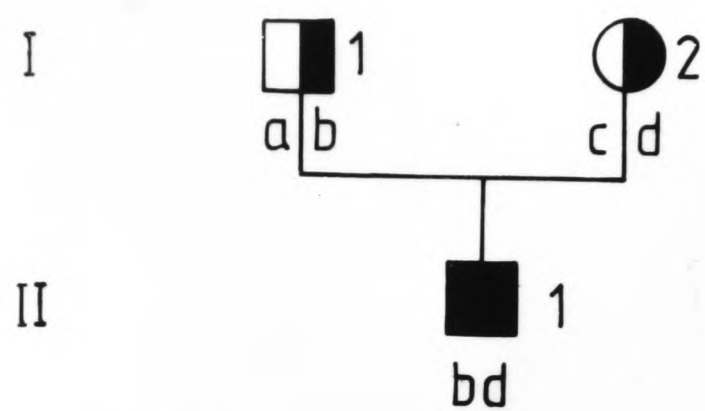
4.2 Southern Blot Analysis of Family Q

In family Q, the individual II.1 suffers from the salt-wasting form of CAH. His HLA haplotypes are A2 B44 C5 DR6 C4 A3 BQ0 / A3 Bw47 (C6)

Figure 4.1

Southern Analysis of Family Q with KpnI and TaqI and NcoI

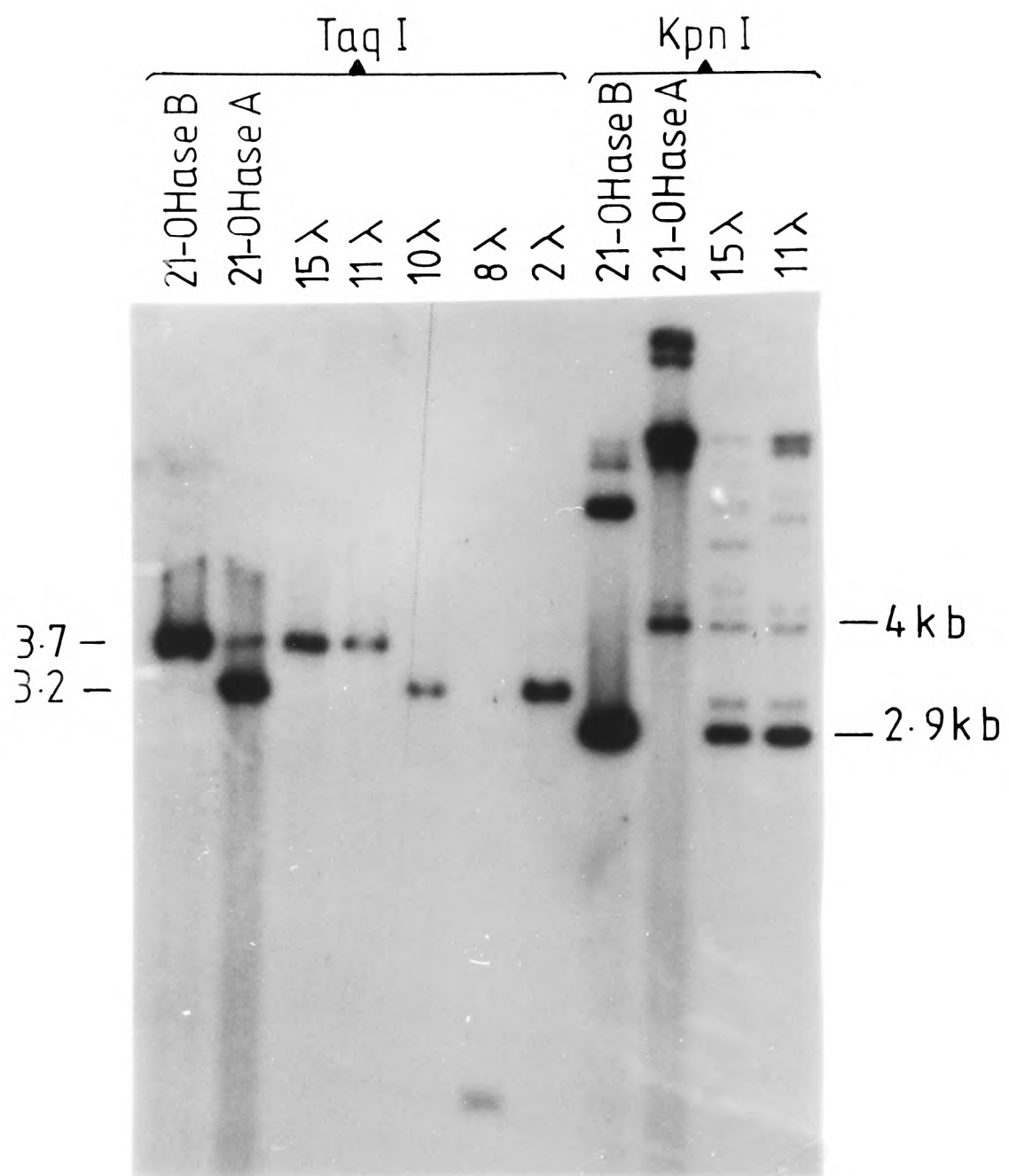
The pedigree of the family Q is shown at the top and HLA/complement haplotypes are given in Table 4.1. □, Male; ○, Female; full shaded, homozygous for 21-OHase deficiency. The 3.7 kb and 3.2 kb fragments in the TaqI digest and the 2.9 kb and 4 kb fragments in the KpnI digest correspond to the 21-OHase B and 21-OHase A genes respectively. The gene dosage shown in Table 4.1 was deduced from the intensity of the fragments by scanning of the autoradiograph.



DR7 C4A1 BQ0. The A3 Bw47 (C6) DR7 haplotype shows a strong genetic linkage disequilibrium with 21-OHase deficiency and there is always a deletion of the 21-OHase B gene along with the adjacent C4B gene. Genomic DNA from the individual II.1 and his parents (family Q) was digested with the restriction enzymes TaqI and KpnI and the Southern blots were probed with the 21-OHase specific probe PA (Probe PA is an 850 bp PvuII / Asp-718 fragment from the 3' end of the gene and covers the 8th, 9th and 10th exons). Two fragments at 3.7 kb and 3.2 kb were seen in the TaqI digest and two fragments at 2.9 kb and 4 kb were seen in the KpnI digest (Figure 4.1). The former fragment in each digest corresponds to the 21-OHase B gene and the latter to the 21-OHase A genes as mentioned previously. The gene dosage was determined by scanning densitometry of the autoradiograph and is listed in Table 4.1. Individual II.1 has a single 21-OHase B gene and two 21-OHase A genes determined on the basis of the intensity of the 3.7 kb TaqI and 2.9 kb KpnI fragments compared to the intensity of the 3.2 kb TaqI and 4 kb KpnI fragments. The father I.1 with the haplotypes A28 B51 C- DR5 C4A3 BQ0 / A2 B44 C5 DR6 C4A3 BQ0, has one 21-OHase A gene and two 21-OHase B genes. The mother I.2 with the haplotypes A2 B62 C3 DR4 C4A3 B3 / A3 Bw47 (C6) DR7 C4A1 BQ0, has two 21-OHase A and one 21-OHase B gene. The individual II.1 has inherited the B44 haplotype from his father and the Bw47 haplotype from his mother. By deduction, he has inherited a single 21-OHase A gene on the A3 Bw47 (C6) DR7 C4A1 BQ0 haplotype from his mother and one 21-OHase A gene and his single 21-OHase B gene on the A2 B44 C5 DR6 C4A3 BQ0 haplotype from his father. Since individual II.1 has CAH, his single 21-OHase B gene is defective. This gene was characterized further by direct cloning (Section 4.3) and sequence analysis (Section 4.5).

Figure 4.2

Southern analysis of the 21-OHase positive clones obtained from the genomic library of individual II.I. The λ clones were digested with restriction enzymes TaqI and KpnI. The controls were 21-OHase A and 21-OHase B gene plasmids, which have the 3.2 kb and 3.7 kb TaqI fragments and the 4 kb and 2.9 kb KpnI fragments, respectively.



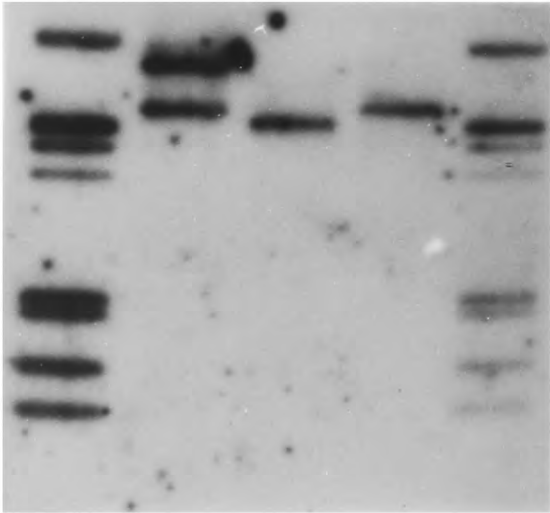
4.3 Cloning of the 21-OHase B Gene from Individual II.I

The Southern blot analysis of family Q established that individual II.I who suffers from the severe salt-wasting form of CAH has a single 21-OHase B gene which may be defective. A genomic library constructed by C.Y.Yu (Yu et al., 1986), of DNA from an EBV-transformed lymphoblastoid cell line derived from lymphocytes of individual II.I was screened with the 21-OHase probe PA . In all, 22 positives were picked and rescreened. DNA was isolated from 7 of the λ clones according to Maniatis et al., (1982). The DNA was subjected to TaqI and KpnI digests and the clones 11 λ 21B and 15 λ 21B were selected for further analysis due to the presence of the 3.7 kb TaqI and 2.9 kb KpnI fragments (Figure 4.2). Since the intensity of the 3.7 kb TaqI fragment in individual II.I indicated the presence of a single 21-OHase gene, both the λ clones with the 3.7 kb TaqI fragment should represent the same 21-OHase B gene while the other λ clones with the 3.2 kb TaqI fragment could represent either of the two 21-OHase A genes in individual II.I. The 15 λ 21B clone was digested with BglII and BamHI in a double digest and compared to the BglII / BamHI digest of cos 1E3 containing the 21-OHase A gene and the BglII / BamHI / SalI digest of cos KEM-1 containing the 21-OHase B gene. Figure 4.3 shows the Southern blot analysis of the gel. Cos 1E3 showed two fragments, one of which was at 5.5 kb and the other at 18 kb which was probably due to the digest not reaching completion. Cos KEM-1 shows a 5.2 kb fragment because the limits of the insert were short of the BamHI site in the genomic DNA, hence the SalI site in the cosmid vector was used to cut out the insert. The 15 λ 21B clone shows a 5.5 kb fragment which is characteristic of the 21-OHase gene. To subclone the single 21-OHase B gene, the BglII / BamHI digest of 15 λ 21B was shotgun cloned into pAT 153 / PvuII / 8 digested with BamHI. The ligation was transformed into MC1061 cells which were plated on LA plates and

Figure 4.3

1. The BglII / BamHI digest of cos 1E3 containing the 21-OHase A gene gives a fragment at 5.5 kb and 18 kb.
2. The BglII / BamHI / SalI digest of cos KEM-1 containing the 21-OHase B gene gives a fragment at 5.2 kb.
3. The BglII / BamHI digest of 15λ21B containing the 21-OHase A gene from the patient II.I gives a fragment at 5.5 kb.

Bgl II / Bam HI
cos IE 3 (21A)
Bgl II / Bam HI / Sal I
cos KEM-1 (21B)
Bgl II / Bam HI
15 x 21B



21

5.2 / 5.0

2.0
1.9
1.6
1.3

Figure 4.4

Comparison of fragments obtained by restriction digests of the 3.7 kb TaqI subclone from II.I (λ) and the 3.7 kb TaqI clone from cos KEM-1 (B). The digests done were:

Asp718

NcoI

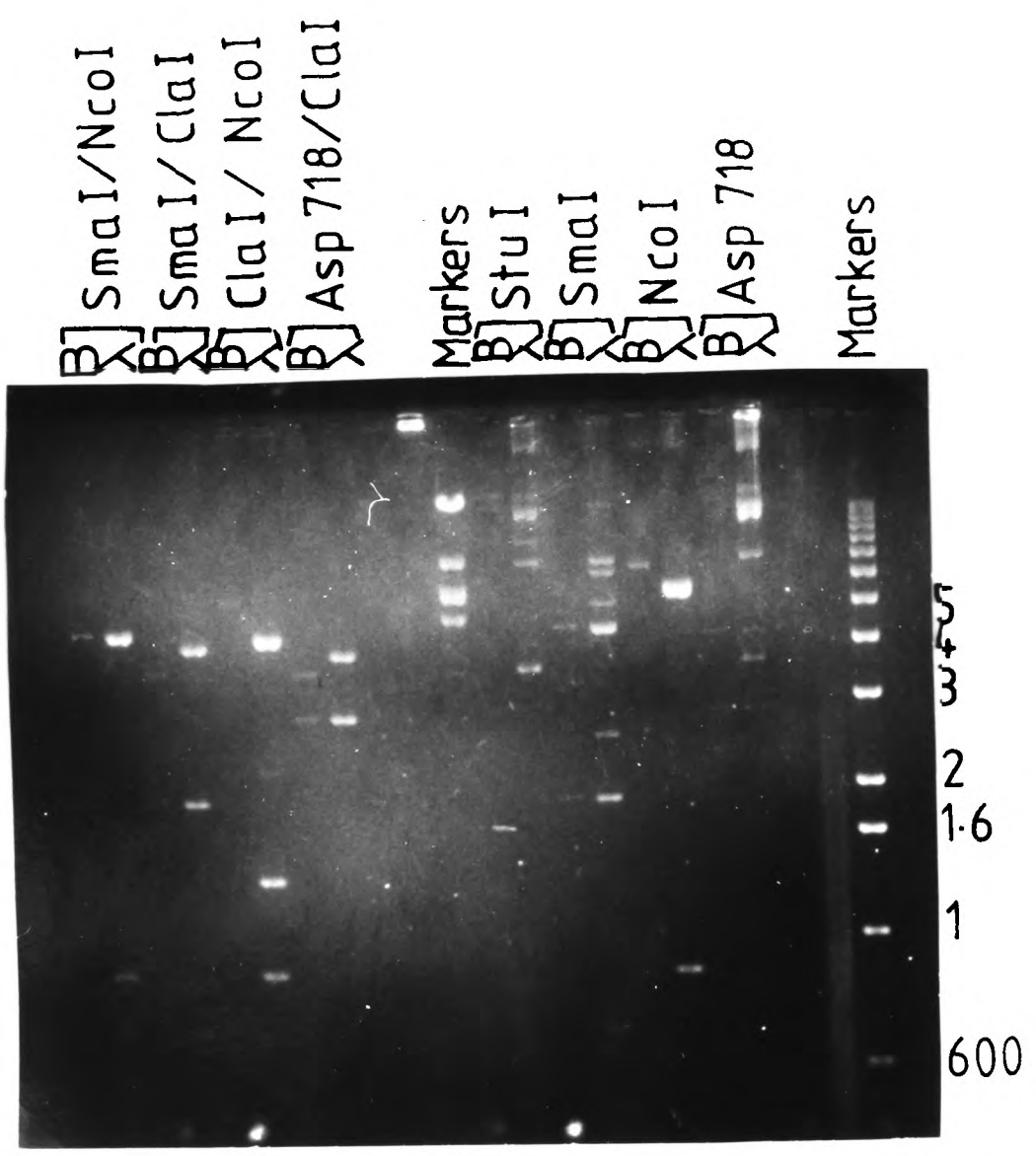
SmaI

Asp718 / ClaI

ClaI / NcoI

SmaI / ClaI

SmaI / NcoI



screened with 21-OHase probes. Plasmid DNA was isolated from positives and the insert site was estimated by digestion with restriction endonucleases ClaI and BamHI. The insert size was 5.5 kb. A 3.7 kb TaqI fragment was also subcloned from clone 11λ21B into pAT153 / PvuII / 8 by blunt end ligation. The 3.7 kb TaqI subclone from 11λ21B was digested with Asp718, NcoI, SmaI, StuI, Asp718 / ClaI, ClaI / NcoI, SmaI / ClaI and SmaI / NcoI, and the restriction patterns were compared with those obtained from the 3.7 kb TaqI subclone from cos KEM-1 (Figure 4.4). The ClaI double digests were used to orient the subclone and it appeared that the subclone from 11λ21B (3.7λ21B) was in the reverse orientation to that of the subclone (3.721B) from cos KEM-1. The Asp718 / ClaI double digest gave fragments at 3 kb and 4 kb with 3.7λ21B which placed the ClaI site at 3' end of the 3.7 kb TaqI insert. In the 3.721B subclone, the Asp718 / ClaI double digest gives fragments at 3.5 kb, 3.0 kb and 0.7 kb and the ClaI site is at the 5' end of the insert (Figure 4.5a). The very apparent restriction site difference between the two subclones was with the enzyme NcoI which gave a 850 bp fragment in the gene from individual II.I that was absent in the gene from KEM-1. The ClaI / NcoI double digest of 3.7λ21B gave fragments at 4.2 kb, 1.4 kb, 0.8 kb and 420 bp while the 3.721B subclone gave fragments at 5.7 kb, 1.1 kb and 420 bp which placed the NcoI site 1.4 kb away from the TaqI site at the 3' end of the insert (Figure 4.5b). The SmaI / NcoI double digest which gave 4.5 kb, 0.85 kb and 0.6 kb fragments with 3.7λ21B and 4.5 kb, 1.25 kb, and 0.6 kb fragments with 3.721B confirmed the same result (Figure 4.5c).

4.4 Subcloning of the 21-OHase B Gene from Individual II.I into M13mp9

The 3.7 kb TaqI subclone, 3.7λ21B was used to subclone most of the fragments of the 21-OHase B gene. Fragments of sizes 416 bp, 431 bp,

Figure 4.5

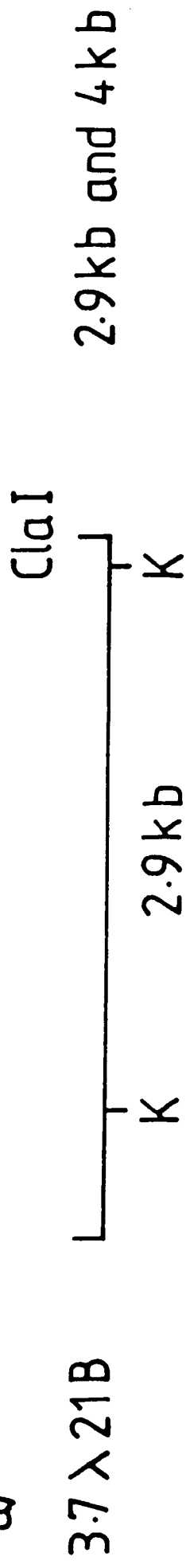
Comparison of the restriction maps of the two 3.7 kb TaqI subclones. 3.7 λ 21B is the subclone of the 21-OHase B gene from individual II.I. 3.721B is the subclone of the 21-OHase B gene from cos KEM-1. The orientation of the inserts in the plasmid vector is opposite in both subclones. 3.7 λ 21B has an extra NcoI site in the insert.

K : KpnI

N : NcoI

S : SmaI

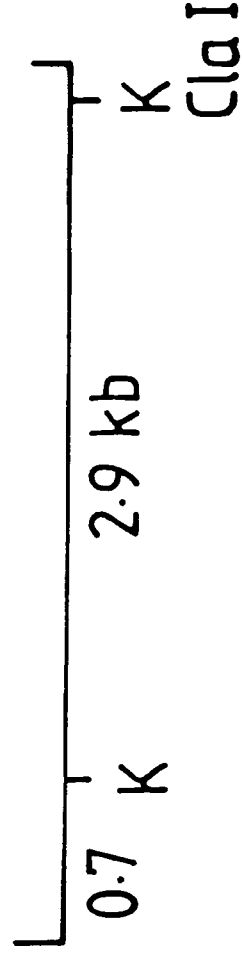
a)



2.9 kb and 4 kb

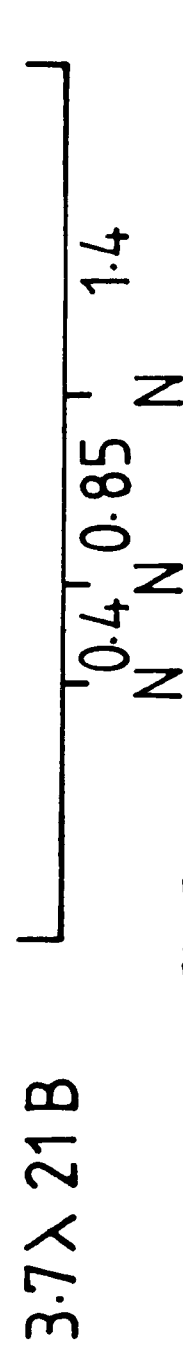
ClaI

3.721B



0.7 kb, 2.9 kb and 3.5 kb

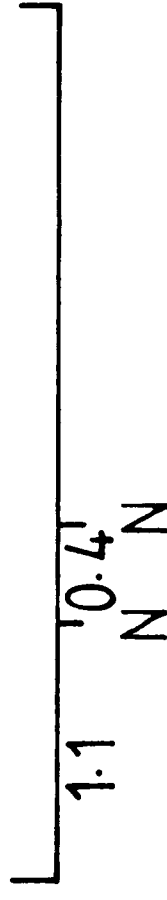
b)



420 bp, 850 bp, 1.4 kb
and 4.2 kb

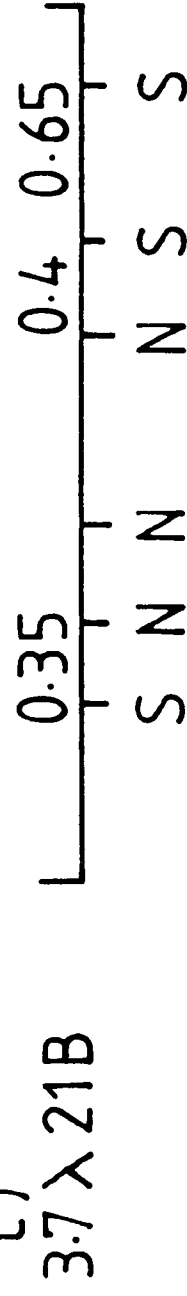
ClaI

3.721B

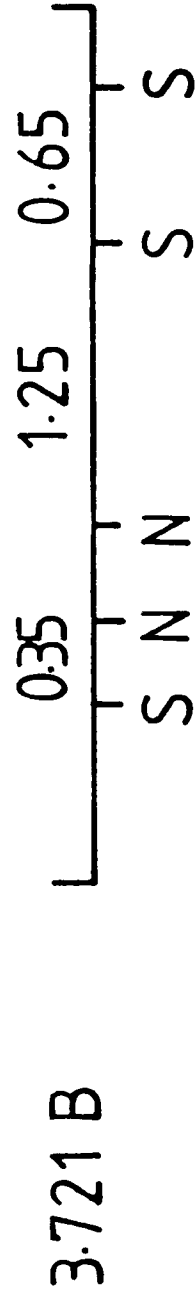


420 bp, 1.1 kb and
5.7 kb

c)



350 bp, 400 bp, 420 bp,
650 bp, 850 bp and 4.5 kb



350 bp, 420 bp, 650 bp,
1.25 kb and 4.5 kb

TABLE 4.2 Fragments used for Subcloning the 21-0Hase B Gene from II.I into
M13 mp9

<u>Restriction Enzyme</u>	<u>Positions</u>	<u>Fragment Length</u>
HinfI	-260, 156	416
	156, 776	620
	776, 1207	431
	1207, 1959	752
	1981, 2163	182
MspI	103, 695	592
	695, 950	255
	950, 1347	397
	1465, 1891	426
	1891, 2043	152
	2043, 2252	209
	2363, 2589	226
HinfI/SmaI	2166, 2674	508
AccI/StuI	-403,-73	330
AccI/NcoI	-73, 340	413
NcoI	340, 784	444
StuI/BamHI	-303,-743	340
StuI/ClaI	2569, 2999	430
Asp718 (KpnI)	2738, 3015	277
Asp718/StyI	3015, 3295	280

751 bp, 182 bp obtained by HinfI digest were subcloned into M13mp9. In order to overlap the HinfI sites fragments of sizes 592 bp, 255 bp, 397 bp, 426 bp, 152 bp, 209 bp, 226 bp were obtained by MspI digest (Figure 4.4) and were subcloned into M13mp9. In addition, HinfI / SmaI (508 bp), AccI / StuI (330 bp), AccI / NcoI (413 bp and 444 bp), StuI / BamHI (340 bp) and StuI / ClaI (430 bp) double digests were also done, BamHI and ClaI sites being those in the vector pAT153 / PvuII / 8. The 3' untranslated region was subcloned using Asp718 (277 bp) and Asp718 / StyI (280 bp) digests on the 5.5 kb BglII / BamHI clone (Table 4.2). The sequencing strategy used was the same as that for the 21-OHase B gene (Section 3.5, Figure 3.10) and hence is not repeated here. All restriction sites were overlapped at least once and the fragments were sequenced by the dideoxy termination method (Section 2.9.2f) and in both orientations.

4.5 Sequence Analysis of the 21-OHase B Gene from Individual II.I

A comparison of the DNA sequences of the 21-OHase B genes from individual II.I with a normal 21-OHase B gene revealed 13 nucleotides alterations which are shown in figure 4.6 and figure 4.7. Of these differences two are in the 5' flanking region upstream of the ATG initiation codon and are due to an A → G transition at nucleotide -551 and an G → T transversion at -282. These two changes may affect promoter activity causing a decrease in the expression of the gene. Six differences are in introns, two are in the 3' untranslated region and three are in exons. Eight of the differences (two in exon 1, two in intron 2, one in intron 3, two in intron 6, and one in the 3' untranslated region) change the nucleotides at these positions to that of the 21-OHase A gene sequence (Figure 4.7). It is possible that these 21-OHase A-like changes were introduced into the 21-OHase B gene by gene

Figure 4.6

Schematic representation of differences between a normal 21-OHase B gene and a 21-OHase B gene from a patient (II.1)

The structure of the 21-OHase B gene is shown. Boxes indicate exons. Vertical bars indicate the position of nucleotide differences between the normal 21-OHase B gene and the defective 21-OHase B gene sequences. (•) indicates a change in codon due to a nucleotide alteration. One nucleotide alteration in the seventh exon at codon 269 causes a codon change from a Ser→Thr. The G→C transversion also introduces a NcoI site. Two of the other differences in exons are silent mutations. Six differences are in the introns, 2 in the 5' flanking region and 2 in the 3' untranslated region. Other than the 13 differences, the 21-OHase B gene from the individual with CAH is the same as the normal 21-OHase B gene.

200 bp

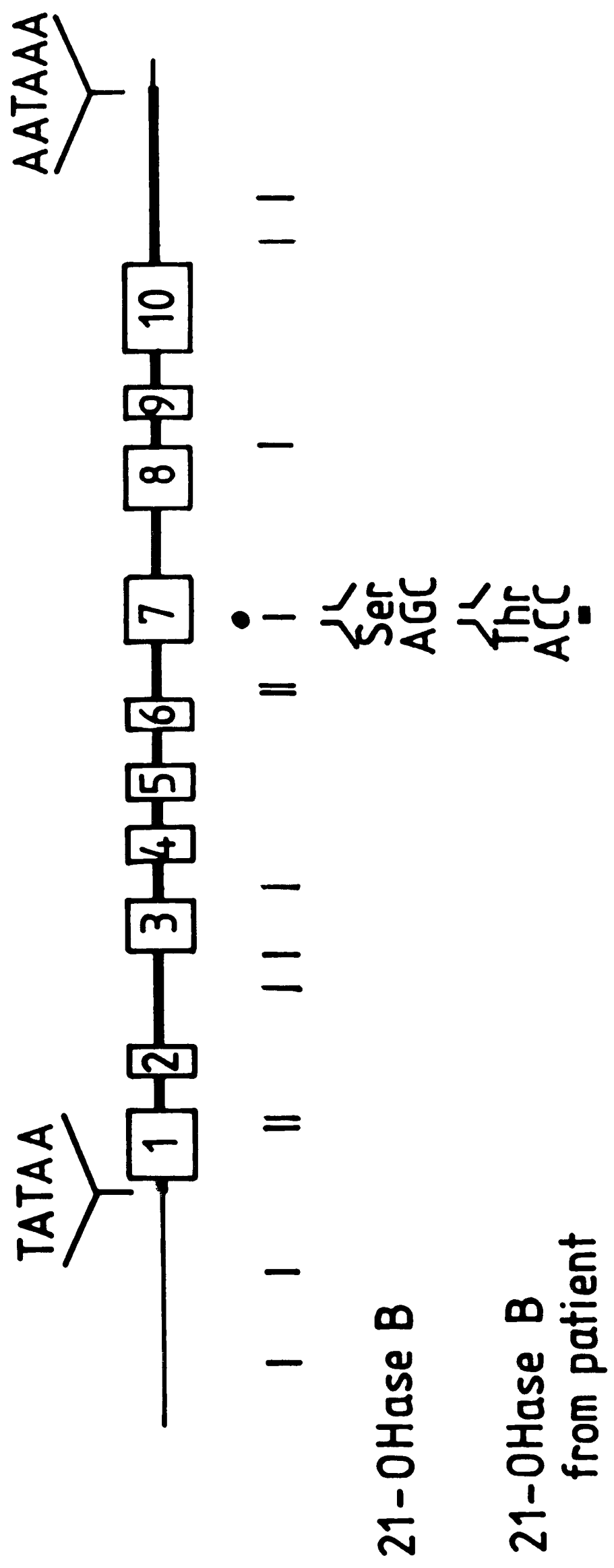


Figure 4.7

Sequence comparison of a 21-OHase B Gene from a patient with CAH (II.1) with a normal 21-OHase B Gene

The complete nucleotide sequence of the 3.7 kb TaqI fragment containing the 21-OHase B gene from a normal individual is shown. The nucleotide differences in the 21-OHase A gene are indicated under the corresponding nucleotides in the 21-OHase B gene and the differences in the 21-OHase B gene from the patient with CAH are indicated in the third line under the corresponding nucleotides of the normal 21-OHase B gene and are highlighted. (Δ) indicates an insertion mutation. A single nucleotide alteration which causes a codon change is marked with an asterisk (*). The putative TATAA transcription signal and the polyadenylation signal (AATAAA) is underlined.

[illegible]

TCCTTCCACAGCTGCATTCTCATGCTTCTCTGCGCAGATTCTTCCCCAATCCAGGTCTCCGAGGCTGAAGCAGGCCATAGAGAGAGGATCACATCGTGGAGATCGATCGAGCTGAGGCAGC 1408
C A* A* A*
H K 246
ACAAGTGGGACTCTGTACGTGGAGCGCCTCCCTCGGCCACAGCCAGTGAATGCTACCGGCTCAGCAATGCTATAGCGGGTTCTTTTGCATACCCAGTTATGGGCCTGTGCCACT 1528
GT
GT
CTGTACTCCTCTCCCCAGGCCAGCGCTCAGCCCGCTCCTTTTACACCTCTCTGCAGGAGAGCCCTGCTGGCAGGCCAGTGGAGGACATGATGGACTACATGCTCCAAAGGGTGGCGAGCCG 1648
T*
S M E E G S G Q L L E G H V H M A A V D L L I G G T E T T A N T L S W A V V F L
AGCATGGAAGAGGGCTCTGGACAGCTCTCTGGAAGGCCACGTGCACATGGCTGCAGTGGACCTCTGATCGGTGGCACTGAGACCACAGAAACACCCCTCTCTGGGCCGTGTTT-TT 1767
T*
C*
L H H P E 313
GCTTACACCCCTGAGGTGCTCTGGGGACAAGCAAAAGGCTCCTTCCAGAACCTGGCCAGGGCGGTGGCACCCCTCACTCAGCTCTGAGCACTGTGCGGCTGTGGGGCTGTGCTTGCC 1887
C
TACCCGCACTCAGGCTCAGTGGTTGCTGAGGGAGCGGCTGGAGCTGGGCTGCTGGGAGGAGTCTCCAGGAGGAGTCTCCCGATCATTTCCCCAGATTTCAGCAGGACTGCAGAGGAG 2007
T*
314 I Q O R L O E E
L D H E L G P G A S S S R V P Y K D R A R L P L L N A T I A E V L R L R P V V P
CTAGACCAAGAACTGGGCCCTGGTGCTCCAGCTCCGGGTCCCTACAAGGACCGTGCACGGTGCCTTCTCAATGCCACCATCCCGAGGTGCTGCGCCCTGGGCCCTGTGTGCC 2127
L A L P H R T T R P S S 373
TTAGCCTTGGCCCCACCGCACACAGCGCCAGCAGGTGACTCCCGAGGGTTGGGATGAGTCAGGAAGCCCCGAGCCAGGGAGTCTCTGGCCAGCTTAACTCCAGCCCCCTTCAGCA 2247
A
T
I S G Y D I P E G T V I I P N L Q G A H L D E T V W E R P H E F W P 407
TCTCCGGCTACGACATCCCTGAGGGCACAGTCATCATTCGAACTCCAGGCCCCCACTGGATGAGACGGTCTGGAGAGGCCACATGAGTTCTGGCCTGGTATGTGGGGGCCGGG 2367
T
GCCTGCCGTGAAATGTGGTGGAGGCTGCTCCCGCTGCGGCTGAAGCCCTCCCAACCCCTGTCTCCACCGCCCGCAGATCGCTTCTTGGAGCCAGGCAAGAACTCCAGAGCTGTGGCC 2487
408 D R F L E P G K N S R A L A
F G C G A R V C L G E P L A R L E L F V V L T R L L Q A F T L L P S G D A L P S
TTCGGCTGGGTGCCCGTGTGCTGGGGAGCCGCTGGCGGCTGGAGCTCTGTGTGTGTGCTGACCCACTGTGAGGCTTTCAGGCTGTGCTGCTCCGGGAGCGCCCTGCCCTCC 2607
L Q P L P H C S V I L K M Q P F Q V R L Q P R G M G A H S P G Q S Q * 495
CTGCAGCCCTGCCCACTGCAGTGTCTCATCTCAAGATGCAGCCTTTTCCAAGTGGGCTGCAGCCCGGGGATGGGGCCACAGCCCGGCCAGAGCCAGTGTATGGGCGAGGACCGAT 2727
A
GCCAGCCGGTACCTCAGTTTCTCTTTATTGCTCTCTAGGAACCCCTCCCTCCCTCTGAAACACACTGC-TGCCAGATCGCTGGCAGAGAGGCTTCTCCAGCGGCTGGGTGGT 2846
C
C
GAAGGACCCCTGGCTCTTCTCTGGGGGACCCCTCAGTGTCTGGCACTCATACTTGGGTGCGAGAGAGGTGGGAGCAGCTCAGCCTCCCGCCGCTGGGAGCGAAAGTTTCTTGGTCTC 2966
T
AGCTTCATTTTCGTGAAGGACCGAGAACTCGAAGCCCTTCCAGTGTGTACCTCAGCTCAGTCCCTGGGAAAGGGTTGTCAAGAGAGAGTCAAAAGCCGATGTCCCATCTCTTCCCGT 3086
T
TCCCCTTAAGGAGGTAGTCCCAGCACTCAACCACTGCCCGCAGAGCTCCCTTCTGTGACCCCTGGCTGCGAGAGGATTGAGGCTTAATTCTGAGCTGGCCCTTTCCAGCCCAATTAATCA 3206
G
T
C
T
ACTCCAGCTCCCTCTGCGAGGCTGGCATGATTGTTCCATTTCACCCAGCCACTCAGTCCCTTGTACTGTGGGGCTGAACCTAGG 3298
C

conversion-like events, although the remainder of the nucleotides are entirely 21-OHase B like. The two 21-OHase A-like mutations in intron 6 at positions 1425 and 1426 introduce the nucleotides GT at that position. The possibility of an alteration of a splice donor site occurring due to these two mutations cannot be ruled out. Two of the three differences in the exons (an T $\rightarrow$ C transition in codon 40 and an A $\rightarrow$ C transition in codon 46) are silent mutations while the third causes a codon change (Figure 4.6). A transversion from G $\rightarrow$ C in the seventh exon at codon 269 causes a serine to change to threonine. None of the above alterations cause a change in reading frame nor introduce stop codons. Another nucleotide alteration thought to be at codon 494 in the 10th exon which would have been an A $\rightarrow$ G transition changing an asparagine to serine was later found in the sequence of the normal 21-OHase B gene but was absent from other reported sequences (White *et al.*, 1986; Higashi *et al.*, 1986). Therefore it is likely that it is another example of polymorphism and not a deleterious mutation in the 21-OHase B gene from individual II.I.

The sequence of the 21-OHase B gene from the patient suggests that several factors may be responsible for the defect. The gene may not be expressed due to the mutation in its promoter or the gene may be expressed but due to a faulty splice junction, the message may be aberrant. Since the functions of introns have not been clearly defined, it is not possible to predict the effect that the changes in the introns might have on the gene. If the gene is expressed the functional activity of the gene product may be altered due to the single amino acid change at position 269.

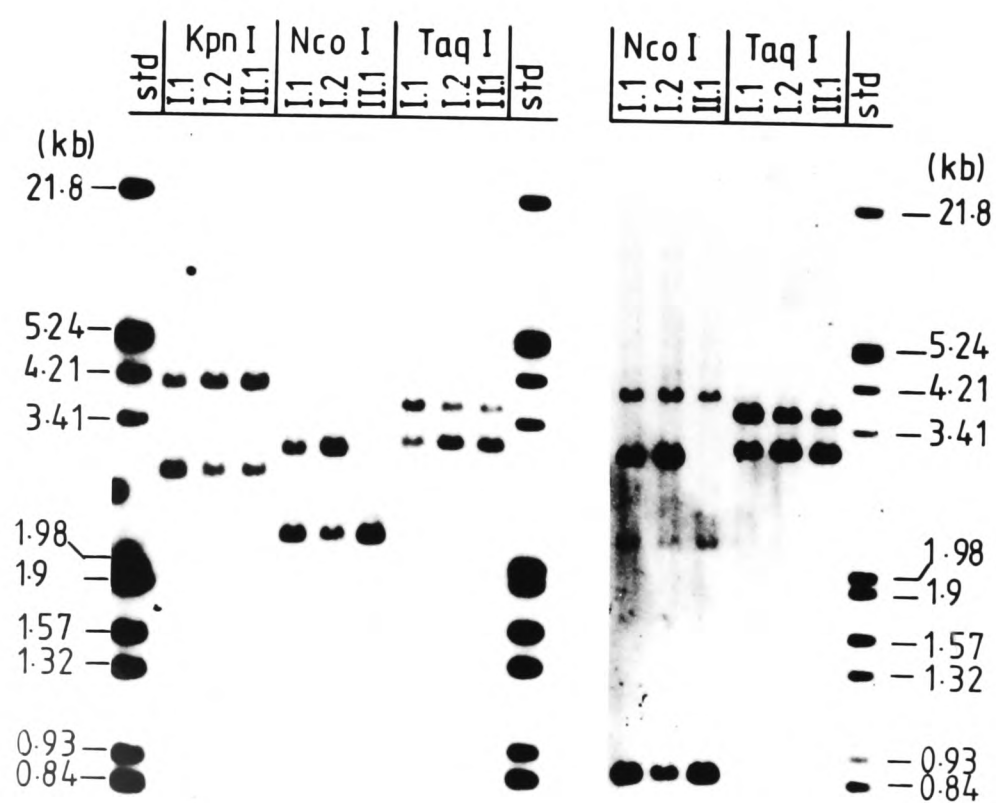
4.6 NcoI Polymorphism in Individual II.I

The transversion from G $\rightarrow$ C in the seventh exon of the 21-OHase B

Figure 4.8

Southern Analysis of KpnI, TaqI and NcoI polymorphisms in Family Q

Unclassified genomic DNA from individuals of family Q was digested with the restriction endonucleases KpnI, TaqI and NcoI. The left panel shows the Southern blot hybridized with probe PA. In the NcoI digest individual II.1 has a single fragment at 2.3 kb while his parents have fragments at 2.3 kb and 3.1 kb. The right panel shows the same blot hybridised with probe H which is 5' to PA. The TaqI and KpnI (not shown) patterns remain unchanged. However, in the NcoI digest two new fragments at 4 kb and 0.85 kb light up.



gene in individual II.I which changes a serine to threonine at codon 269 generates a new NcoI site which is absent in the 21-OHase B gene studied from the normal individual. To confirm the NcoI site at the genomic level, DNA from individual II.I and his parents was digested with NcoI and probed with probes PA and H (H consists of the 410 bp and 430 bp HinfI fragments from the 21-OHase gene), respectively. The NcoI digest pattern with probe PA is shown in figure 4.8. Individual II.I has only one fragment of 2.3 kb while his parents have fragments of 3.1 kb and 2.3 kb. In order to map precisely the position of the NcoI fragments within the gene the blot was reprobed with probe H which is 5' to probe PA. Individual II.I showed two fragments of 4 kb and 0.85 kb, while his parents each showed three fragments of 4 kb, 3.1 kb and 0.85 kb (Figure 4.8).

Hybridization of the blot with a probe specific for the 3' end of C4 revealed that the 4 kb NcoI fragment extends onto the 3' region of the neighbouring C4 gene (not shown). On the basis of these results and scanning of the autoradiographs to determine band intensities, the NcoI restriction map was constructed (Figure 4.9). This confirmed the polymorphic NcoI site within the seventh exon of the 21-OHase B gene in individual II.I. This NcoI polymorphism is not restricted to the 21-OHase B gene as it is also found in the 21-OHase A genes of the B44 and Bw47 haplotypes in this family. Nucleotide sequences from random unrelated individuals do not show this site in either of their 21-OHase genes (Higashi et al., 1986 ; White et al., 1986 ; Rodrigues et al., 1987). In this family, however, both affected haplotypes exhibit this additional NcoI site.

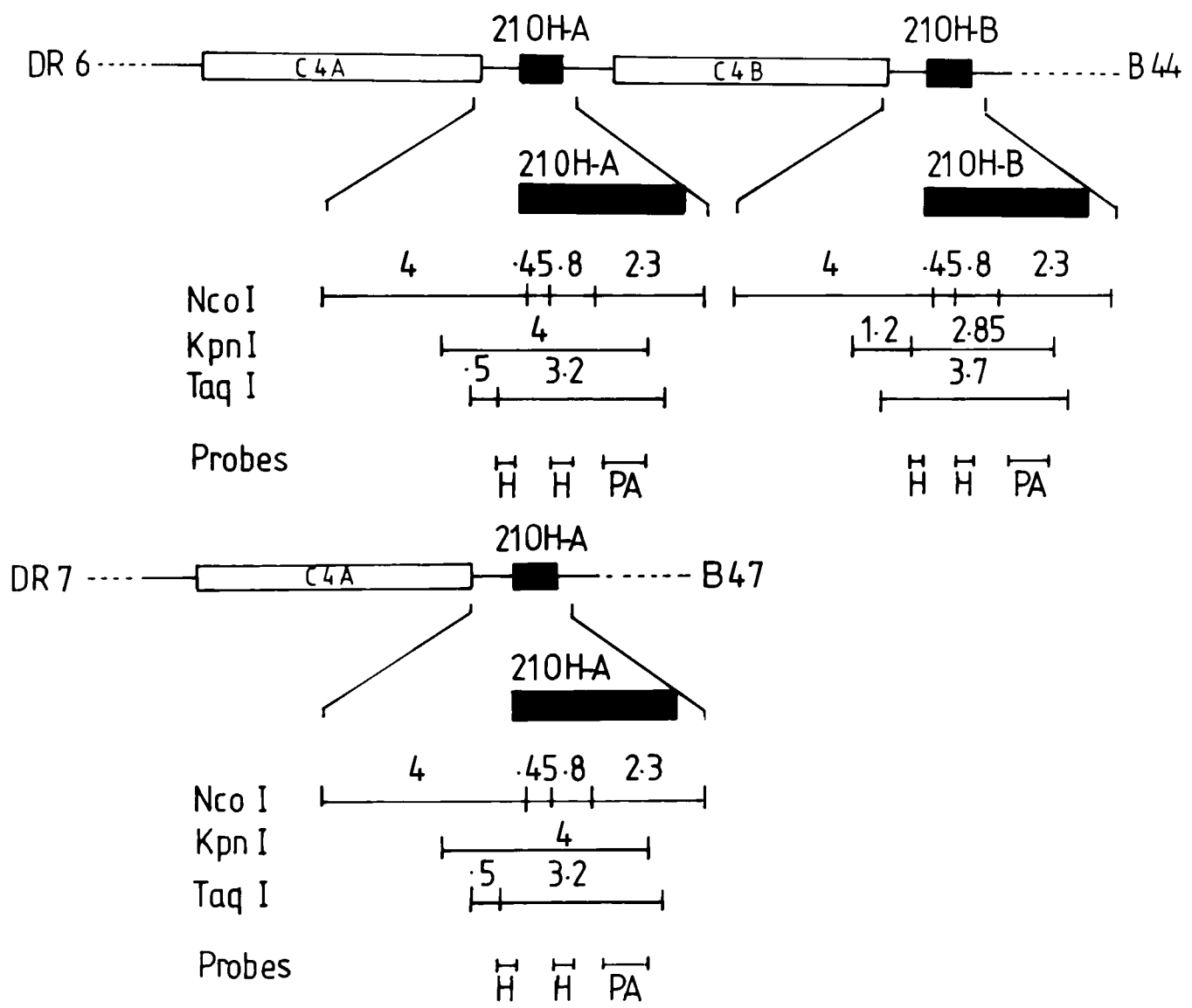
4.7 Expression of the 21-OHase B Gene from Individuals II.I

Expression of 21-OHase is restricted to the adrenal and no 21-OHase

Figure 4.9

Restriction Maps of the 21-OHase Genes in Individual II.1

Restriction maps of the 21-OHase genes on the B44 and B47 haplotypes of individuals II.1 are shown for enzymes NcoI, KpnI and TaqI. The 4 kb KpnI and 3.2 kb TaqI fragments correspond to the 21-OHase A gene, and the 2.85 kb KpnI and 3.7 kb TaqI fragments correspond to the 21-OHase B gene. The polymorphic NcoI site which cleaves the 3.1 kb fragment into two fragments of 2.3 kb and 0.85 kb is present in all the three 21-OHase genes of individual II.1. The probes used are indicated below the maps.



RNA was detected in kidneys, liver, testis and ovaries (Parker et al., 1985). The Y1 adrenocortical cell line expresses many of the enzymes involved in adrenal steroidogenesis, but does not normally express 21-OHase despite the presence of the 21-OHase genes in the genome. Transfection of the Y1 cell line with cosmid clones containing murine 21-OHase A and 21-OHase B genes demonstrated 21-OHase mRNA in those cells transfected with the 21-OHase A genes, but no 21-OHase mRNA in the nontransfected Y1 cells or Y1 cells transfected with the 21-OHase B gene (Parker et al., 1985). Subsequent nucleotide analysis has shown that the murine 21-OHase A gene is the active gene while the murine 21-OHase B gene is a pseudogene due to a deletion of 215 bases that include the second exon as well as several single base changes that result in premature termination codons (Chaplin et al., 1986). In the human 21-OHase genes, it is evident that the 21-OHase A gene is a pseudogene due to three deleterious mutations which give rise to premature termination codons. Northern blot analysis of adrenal RNA using oligonucleotides specific for the 21-OHase A and B genes has shown that only the 21-OHase B gene specific probe hybridized to an mRNA species (Higashi et al., 1986). Although the changes in the 5' non transcribed region of the 21-OHase B gene from individual II.I are relatively distant from the cap site, they may affect expression of the gene by altering tissue specific regulation sites which may reside in that region.

In preliminary experiments carried out to demonstrate whether the 21-OHase B gene in individual II.I was expressed, his 21-OHase B gene, the 21-OHase A gene from cos 1E3 and 21-OHase B gene from cos KEM-1 were cotransfected with the plasmid pRSVCAT containing the bacterial chloramphenicol acetyltransferase (CAT) gene into Y1 adrenocortical cells. Cells were harvested after 48 hr without being subjected to G418 selection and the total cellular extract was checked for CAT activity to check the efficiency of transfection of the CAT gene. The results of

Figure 4.10

CAT Assay of cellular extracts from transfected Y1 adrenocortical cells

+ : positive control showing 95% CAT activity

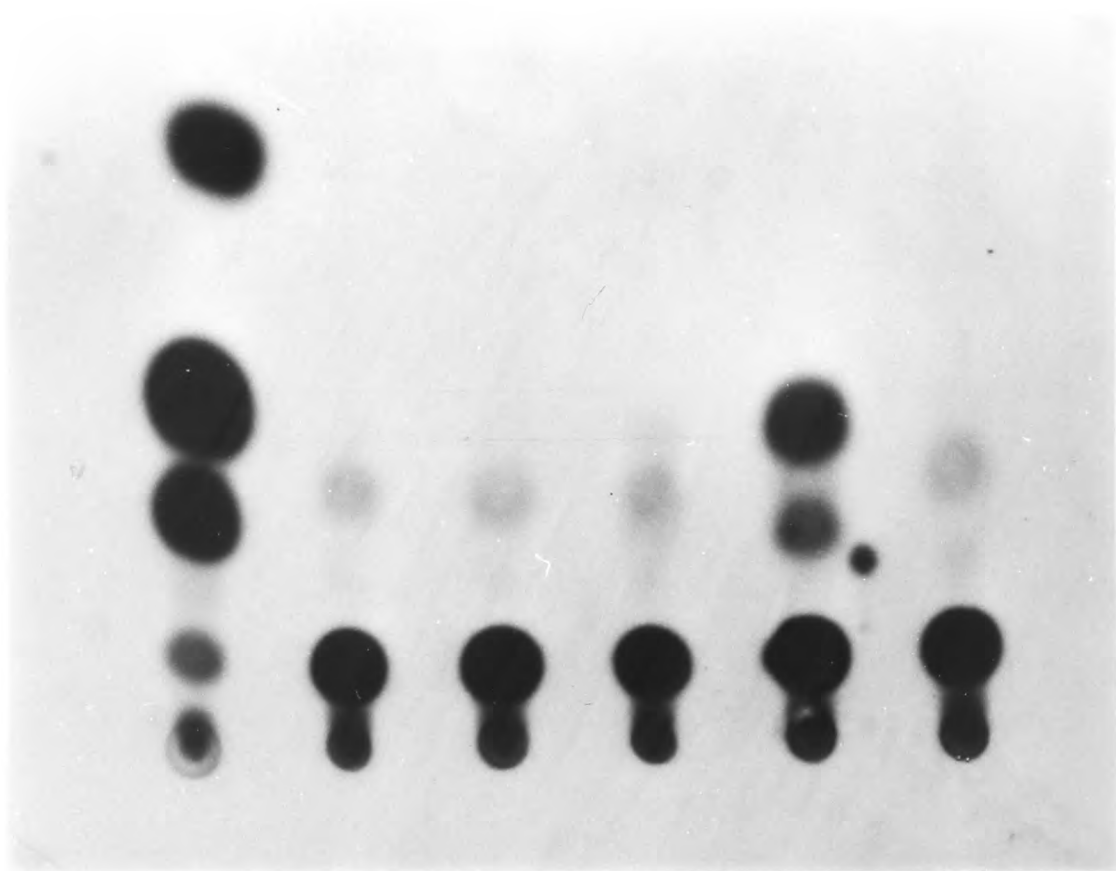
B : extract from Y1 cells transfected with the 21-OHase B gene

M : extract from Y1 cells transfected with the 21-OHase B gene from individual II.I

A : extract from Y1 cells transfected with the 21-OHase A gene

C : extract from Y1 cells transfected only with the pRSCVAT plasmid

T : Tris-HCl buffer



+ B M A C T

Figure 4.11

Northern blot of RNA isolated from transfected Y1 adrenal cells.

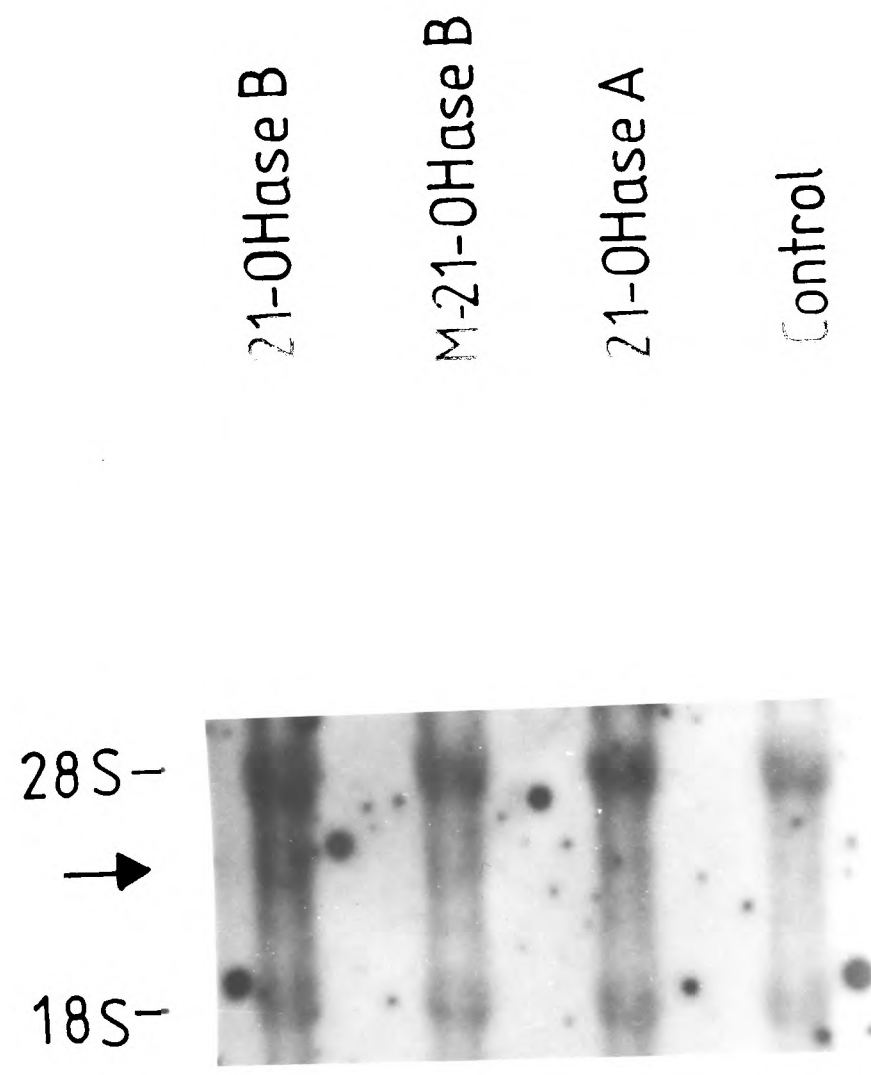
C : RNA from cells not transfected with the 21-OHase genes

A : RNA from cells transfected with the 21-OHase A gene

M : RNA from cells transfected with the 21-OHase B gene from
individual II.I

B : RNA from cells transfected with the normal 21-OHase B gene

Probe : PA



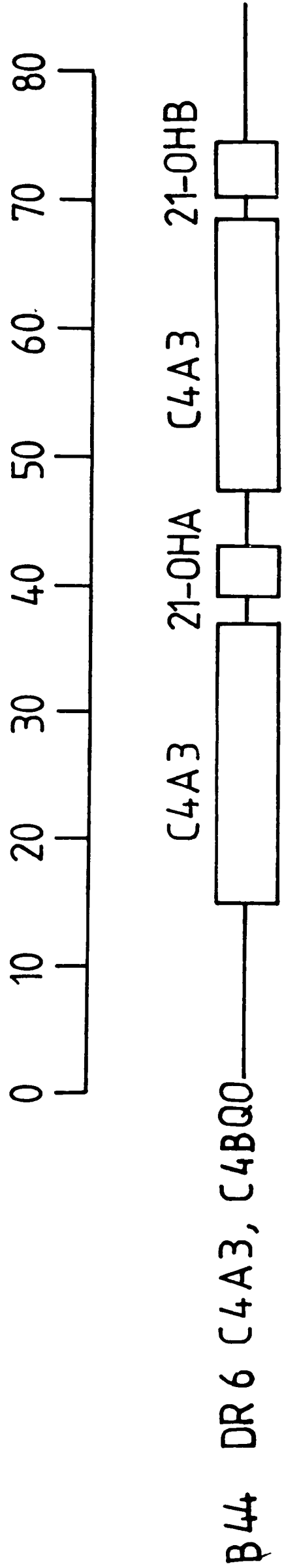
the CAT assay are shown in Figure 4.10. The positive control is an extract from the HepG2 cell line which shows 95% CAT activity. Of the other samples, only C shows some activity when compared to the blank T. C is the extract from cells which had been transfected only with pRSVCAT. All others show the same activity as blank T. It meant that either the plasmids containing the 21-OHase genes had competed with the pRSVCAT plasmid in transfection, or the transfection had not been successful. Northern blot analysis was performed on the total RNA isolated from cells. The probe used was PA (Section 4.2). Although the 28S and 18S bands light up in all the tracks, the autoradiograph was examined for differences between the tracks. Faint bands could be seen between the 28S and 18S bands in the Y1 cells transfected with normal 21-OHase B gene (B) and 21-OHase B gene from individual II.I (M) which were fainter in the Y1 cells transfected with 21-OHase A gene (A) and absent in the nontransfected Y1 cells (C) (Figure 4.11). These experiments were preliminary investigations, and therefore the results were difficult to interpret. Obviously long term transfection has to be performed before identifying the basis of the defect in the 21-OHase B gene from individual II.I.

4.8 Discussion

Individual II.I with the salt-wasting form of CAH was chosen for the study because he had one 21-OHase B gene in his genome which was clearly nonfunctional at some level. Studying the single 21-OHase B gene at the molecular level would provide some clue to the nature of the defect. DNA sequencing of the gene revealed 13 nucleotide differences when compared with the normal 21-OHase B gene. Apart from one amino acid substitution at codon 269 there were no differences that were obviously deleterious. Six differences were located in introns, two in the 5' flanking region of the gene, two in the 3' untranslated region and three in exons. Eight of the differences correspond to nucleotides

Figure 4.12

Organization of the C4 and 21-OHase genes on the A2 B44 C5 DR6 C2C BfF C4A3 BQ0 haplotype. There are two C4 loci and two 21-OHase loci present. The first C4 locus is C4A3 and the first 21-OHase locus is a 21-OHase A gene. The second C4 locus is not a C4B allele although it has C4B the characteristic 6 kb TaqI fragment. It has the C4A characteristic 276 bp and 197 bp NlaIV fragments and probably expresses the C4A3 allotype. The second 21-OHase locus is a 21-OHase B gene which is defective and causes CAH in the patient.



in the 21-OHase A gene at those positions. Studies of the adjoining C4 gene have indicated that the second C4 locus on the A2 B44 C5 DR6 C2C fBF C4A3 BQ0 is not a C4B allele but probably expresses the C4A3 allotype which is identical to that of the first C4 locus on the same chromosome (Yu et al., 1987, Figure 4.12). Although it has a 6.0 kb TaqI fragment characteristic of a C4B gene, it lacks the 467 bp NlaIV fragment specific to the C4B gene. Instead it has the 276 bp and 191 bp NlaIV fragments which are characteristic of the C4A gene. There is likelihood that gene conversion-like or other recombinational events may have resulted in the second C4 locus expressing C4A. Gene conversion has been proposed to be the cause of CAH in one patient where the deleterious mutations from the 21-OHase A gene are substituted for the 21-OHase B gene sequence (Donohoue et al., 1986) although in individual II.I, the 21-OHase A like changes in the exons are silent mutations. It is likely that 21-OHase deficiency in some cases will be due to the introduction of the deleterious mutations found in the 21-OHase A gene into the 21-OHase B gene presumably by gene conversion-like events.

The precise effect of these nucleotide alterations on the expression of the 21-OHase in individual II.I is not clear. The two nucleotide alterations in the 5' flanking region, or even those differences in the introns may lead to no or low expression of the gene. Faulty splicing in intron 6 due to a GT splice site being introduced may produce an aberrant message. Figure 4.13 shows the possible effect of a faulty splicing in intron 6 on the protein. If the splicing occurs at position 1425 instead of 1414, the resultant protein sequence would be as shown in the figure. A premature termination codon is also introduced due to a frame shift. Both these changes would lead to an altered and truncated protein product. Alternatively, the gene may be fully expressed but the protein product due to the single amino acid substitu-

Figure 4.13

The protein sequence of the normal 21-OHase B gene is shown from amino acid 218 to amino acid 268 above the corresponding nucleotide sequence. Intron 6 begins after codon 246 in the normal 21-OHase B gene (†). Instead if the intron begins 12 bp later due to faulty splicing in the 21-OHase B gene from individual II.I, the protein sequence is altered as shown. A premature termination codon (*) is also introduced due to a shift in the reading frame.

Normal 21-0Hase Sequence

218
F F P N P G L R R L K Q A I E K R D H I
TTCTTCCCCAATCCAGGTCTCCGGAGGCTGAAGCAGGCCATAGAGAAGAGGGGATCACATC
10 20 30 40 50 60

V E M Q L R H K E S L V A G Q W R D M M
GTGGAGATGCAGCTGAGGGCACAAGGAGAGCCTCGTGGCAGGCCAGTGGAGGGGACATGATG
70 80 90 100 110 120
Intron6

D Y M L Q G V A Q F 268
GACTACATGCTCCAAGGGGTGGCGCAGCCG
130 140 150

21-0Hase Sequence due to faulty splicing in intron 6

218
F F P N P G L R R L K Q A I E K R D H I
TTCTTCCCCAATCCAGGTCTCCGGAGGCTGAAGCAGGCCATAGAGAAGAGGGGATCACATC
10 20 30 40 50 60

V E M Q L R H K V G T V R A S W Q A S G
GTGGAGATGCAGCTGAGGGCACAAGGTGGGGACTGTGAGAGCCTCGTGGCAGGCCAGTGGGA
70 80 90 100 110 120
Intron6

G T * W T T C S K G W R S
GGGACATGATGGACTACATGCTCCAAGGGGTGGCGCAGCCG
130 140 150 160

tion may be inactive. Single amino acid changes have been known to drastically alter the function of a protein as in sickle cell anaemia (Ingram 1957), where an A $\rightarrow$ T transversion in codon 6 of the β -globin chain of hemoglobin changes a glutamic acid residue to a valine residue. As a result sickle hemoglobin tends to crystallize in red blood cells, the cells become less flexible and are removed by the spleen, and anaemia occurs.

The active site of a cytochrome P-450 contains iron protoporphyrin IX presumably in a hydrophobic cleft in the protein (White and Coon, 1980) (Figure 4.14). The heme which is not covalently bound to the protein is always penta or hexa coordinated, four of the ligands provided by the N atoms of the porphyrin ring and the fifth ligand is believed to be a specific cysteine residue of the polypeptide chain. This cysteine is conserved in all cell cytochrome P-450 enzymes (Morohashi et al., 1984), corresponding to cys-429 in 21-OHase. The sixth coordinated position of the iron is occupied by an easily exchangeable oxygen ligand either from water or the hydroxyl group of a serine or threonine residue (Ullrich et al., 1979). If the serine at codon position 269 is indeed involved in the sixth coordination position, a change in it would affect the functioning of the protein, although a threonine could provide the necessary group. A comparison with the bovine and murine 21-OHase genes shows that at that position there is no homology between the two sequences although the flanking regions are very homologous (Table 4.3). It remains to be seen whether individuals having a single 21-OHase B gene with the NcoI polymorphism are normal. Since the 21-OHase genes also exhibit polymorphism, it is difficult to predict whether the single amino acid change renders the protein wholly nonfunctional. Preliminary work done on the expression of the 21-OHase B gene from the patient and a normal 21-OHase B gene suggests that the

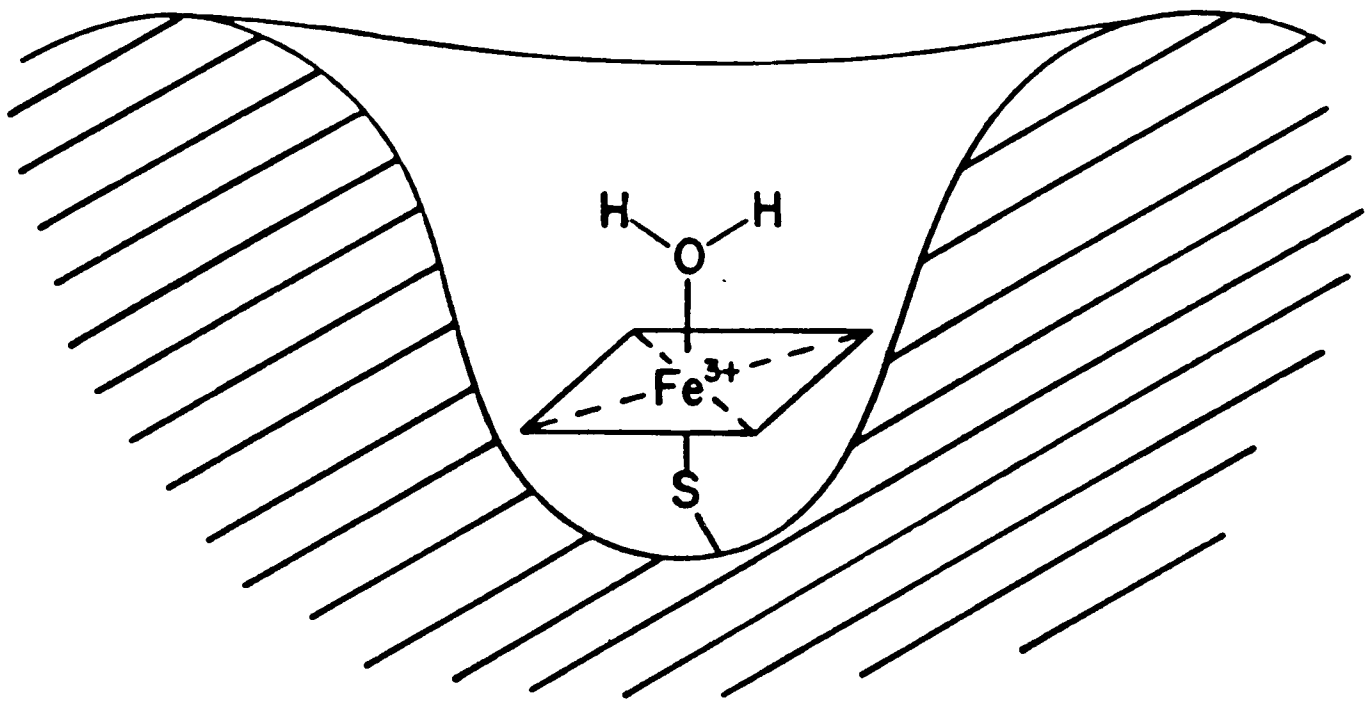
TABLE 4.3 Comparison of human 21-OHase from a normal person and a patient with porcine and bovine 21-OHase around codon 269

Porcine 21-OHase		M	L	Q	E	A	G	R	Q	R	V	E	E	G	Q	G	
Bovine 21-OHase	259	M	L	Q	G	V	G	R	Q	R	V	E	E	G	P	G	273
Human 21-OHase B	261	M	L	Q	G	V	A	Q	P	S	M	E	E	G	S	G	275
Human 21-OHase B from CAH patient	261	M	L	Q	G	V	A	Q	P	T	M	E	E	G	S	G	275
										*							

* indicates the Ser to Thr change in the 21-OHase B gene from the CAH patient.
| indicates conservation of residues among species.

Figure 4.14

The model of the active site of cytochrome P-450 (White and Coon, 1980).



Model of the active site of cytochrome P-450.

21-OHase B gene in the patient may be expressed. A bioassay of the 21-OHase enzymatic activity in G418 resistant transformants should provide an answer as to whether the enzyme has any functional activity compared to normal individuals or is totally inactive.

In conclusion the characterization of the patient with CAH has allowed some insight into the molecular basis of 21-OHase deficiency. It is likely that 21-OHase deficiency occurs due to a heterogenous collection of defects and is not due to an isolated mutation which is common to all affected genes as in the case of sickle cell anaemia. These point mutations may lead to an aberration in any one of the many steps in the pathway of gene expression: transcription, RNA processing, mRNA stability, or translation. This has been the case in β -thalassemias, the most extensively studied inherited disorders of hemoglobin synthesis where 23 different mutations have been reported any of which can cause the defect (Weatherall, 1985). Frequency of specific mutant alleles may be high in certain regions or ethnic groups. Characterization of 21-OHase B genes from patients with CAH is therefore crucial in defining different mutations in order to use them in diagnosis.

CHAPTER V

RESTRICTION FRAGMENT LENGTH POLYMORPHISM STUDIES OF THE 21-HYDROXYLASE GENES

5.1 INTRODUCTION

Restriction fragment length polymorphisms (RFLP) have been used to identify alleles or mutations which have single base differences leading to changes in restriction sites giving rise to different restriction patterns (Caskey, 1987). The two 21-OHase loci can be distinguished by the differences in patterns they yield with the restriction enzymes TaqI and KpnI. Although there are other enzymes which can be used to differentiate the two loci, TaqI and KpnI yield fragments which are easily detected and differentiated on genomic Southern blot analysis. The two ~~enzymes~~ can be used to identify the presence or absence of the 21-OHase loci on a given haplotype, and have been used to diagnose the deletion of the 21-OHase B gene on some haplotypes leading to CAH. The study of patients is complicated by the fact that in only one of twenty was CAH due to homozygous deletion of the 21-OHase B gene (Rumsby et al., 1986). In the majority of the patients, therefore, TaqI and KpnI ~~digests~~ can be used only to determine the gene dosage of the 21-OHase A and B genes, respectively. To study the patients who have the 21-OHase B genes present, it is necessary to use other restriction enzymes that would detect point mutations or deletions within the gene. The previous chapter described the NcoI RFLP which was found in individual II.I who has CAH. Since it was not confined to one gene, but was present on both his haplotypes, the study was extended to other patients to determine the frequency of the NcoI polymorphism. Also, being a recessive disorder, random individuals from a normal population were examined for the presence of the RFLP. Another ~~digest~~ with the restriction enzyme AvaII which differentiates the two 21-OHase genes was used in conjunction with NcoI to allocate the NcoI polymorphism to the gene it belonged to.

5.1.1 TaqI

As described previously, the TaqI ^{size} difference in the two 21-OHase genes occurs due to a T → C base substitution at nucleotide -211 in the 5' flanking region of the 21-OHase A gene leading to an extra TaqI site (TCGA) yielding a 3.2 kb fragment as compared to the 3.7 kb fragment characteristic of the 21-OHase B gene. The difference in the sizes of the TaqI fragments is therefore not due to a 500 bp deletion in the 21-OHase A gene as originally suggested (White et al., 1985a). An absence of either of the two fragments in Southern blot analysis indicates a homozygous deletion of that gene, while decreased intensity of the fragments as compared to those of the normal haplotypes with both genes present would indicate a heterozygous deletion (White et al., 1984b). Increased intensity of the fragments would indicate presence of more than one copy of the gene on one or both of the haplotypes. Alternatively, one of the copies may be a hybrid of both the 21-OHase loci which the TaqI ~~digest~~ by itself cannot differentiate. A recent publication by Matteson et al. (1987) has shown that some patients lacking the 3.7 kb TaqI fragment have a normal sized 21-OHase B gene by the restriction pattern obtained with at least two other enzymes. The enzymes used in their study were BglII, EcoRI, KpnI and XbaI. They have suggested that the absence of the 3.7 kb TaqI fragment may be due to RFLPs, gene conversions, crossovers and / or polymorphism. Therefore, although the TaqI enzyme is used to detect gene deletions and the number of copies of the 21-OHase genes on the haplotypes being analysed, it is not conclusive. Apart from the absence of the TaqI fragments, no other abnormalities in the TaqI restriction pattern have been reported.

5.1.2 KpnI

The KpnI ~~digest~~ can also be used to differentiate between the two

21-OHase genes. The 21-OHase A gene yields a 4 kb fragment with KpnI while the 21-OHase B gene yields a 3 kb fragment. This difference in sizes occurs due to a single T → C base substitution at nucleotide-120 in the 5' flanking region of the 21-OHase A gene which loses a KpnI site (GGTACC) present in the 21-OHase B gene. Since the position of the KpnI site is very close to where the TaqI site occurs, both serve the same function in detecting gene deletions and duplications, but fail to detect any other defect in the gene. No deviations from the normal 4 kb and 3 kb restriction pattern have been reported with KpnI.

5.1.3 NcoI RFLP

The NcoI RFLP was first detected in the 21-OHase B gene of individual II.I. A G → C transversion at position 1650 introduced the new NcoI restriction (CCATGG) site into the gene. This new NcoI site cuts down a 3.1 kb fragment to a 2.3 kb fragment and a 0.8 kb fragment. However, this polymorphism was also found in the 21-OHase A genes of individual II.I (Figure 4.9). The NcoI RFLP was used to study 20 other patients with CAH and a number of normal individuals from a random population. The results obtained are discussed in the sections below.

5.2 NcoI RFLP in Individuals with CAH

Twenty patients were studied for the NcoI polymorphism. Eighteen of the samples were obtained from G. Rumsby of The Institute of Child Health, London. The data on the HLA haplotypes, and the TaqI and KpnI banding patterns of these patients were also supplied by G. Rumsby. Table 5.1 shows the haplotypes and the band intensities with TaqI, KpnI and NcoI of the patients. Only two of the 20 patients showed a homozygous deletion of the 21-OHase B gene. Patient JB having a homozygous deletion the of 21-OHase B gene has the haplotypes A9,32 B44,13 Cw4

TABLE 5.1

No	Proband	HLA Haplotype				KpnI		TaqI		NcoI		kb
						4	3	3.2	3.7	3.1	2.3	
1	JB	A9,32	B44,13	Cw4	DR1,7	+	-	+	-	$\frac{1}{2}$	$\frac{1}{2}$	
2	CB	A3	Bw47,62	C6,1	DR7,5	+	$\frac{1}{2}$	+	$\frac{1}{2}$	+	$\frac{1}{2}$	
3	MP	A2,3	B40,w3	Cw6	DR?	+	$\frac{1}{2}$	+	$\frac{1}{2}$	+	$\frac{1}{2}$	
4	BB	A3,23	B39,44		DR2,7	+	-	+	-	-	+	
5	RW	A1,29	B44,40/8	Cw5	DR3,7 DRw52	$\frac{1}{2}$	+	$\frac{1}{2}$	+	+	$\frac{1}{2}$	
6	DM	A3,2	B7,44	C5				+	$\frac{1}{2}$	+	$\frac{1}{2}$	
7	SM	A3,2	B7,44	C5				+	$\frac{1}{2}$	+	$\frac{1}{2}$	
8	CBu	A2,11	Bw47,55	C6,3	DR7,4	+	$\frac{1}{2}$	+	$\frac{1}{2}$	$\frac{1}{2}$	+	
9	ZT	A2,3	B7,35	Cw3	DR6	+	$\frac{1}{2}$	+	$\frac{1}{2}$	+	-	
10	LR	A3	B7		DR2,?5	+	+	+	+	+	-	
11	AR	A3	B7		DR2,?3,?5,?7	+	+	+	+	+	-	
12	A1B	A1,30	B8,13	C6,7	DR3,7	$\frac{1}{2}$	+	$\frac{1}{2}$	+	+	-	
13	CO	A1,32	B39,14	C8	DR1,7	+	$\frac{1}{2}$	+	$\frac{1}{2}$	+	-	
14	DW	A24,1	B14,57	C6,8	DR1	+	$\frac{1}{2}$	+	$\frac{1}{2}$	+	-	
15	SG	A2	B22,27	Cw2?	DR4 DQ3 DRw53	+	$\frac{1}{2}$	+	$\frac{1}{2}$	+	-	
16	C1B	NOT AVAILABLE				+	+	+	+	+	-	
17	LM	NOT AVAILABLE				+	$\frac{1}{2}$			+	-	
18	TC	NOT AVAILABLE				+	$\frac{1}{2}$	+	$\frac{1}{2}$	+	-	
19	LiM	A3,24	B7,14	C7	DR1,2	+	+	+	+	+	-	
20	FH	A2	B51	Cw6	DR2,4	+	$\frac{1}{2}$	+	$\frac{1}{2}$	+	-	

+ indicates the presence of a band

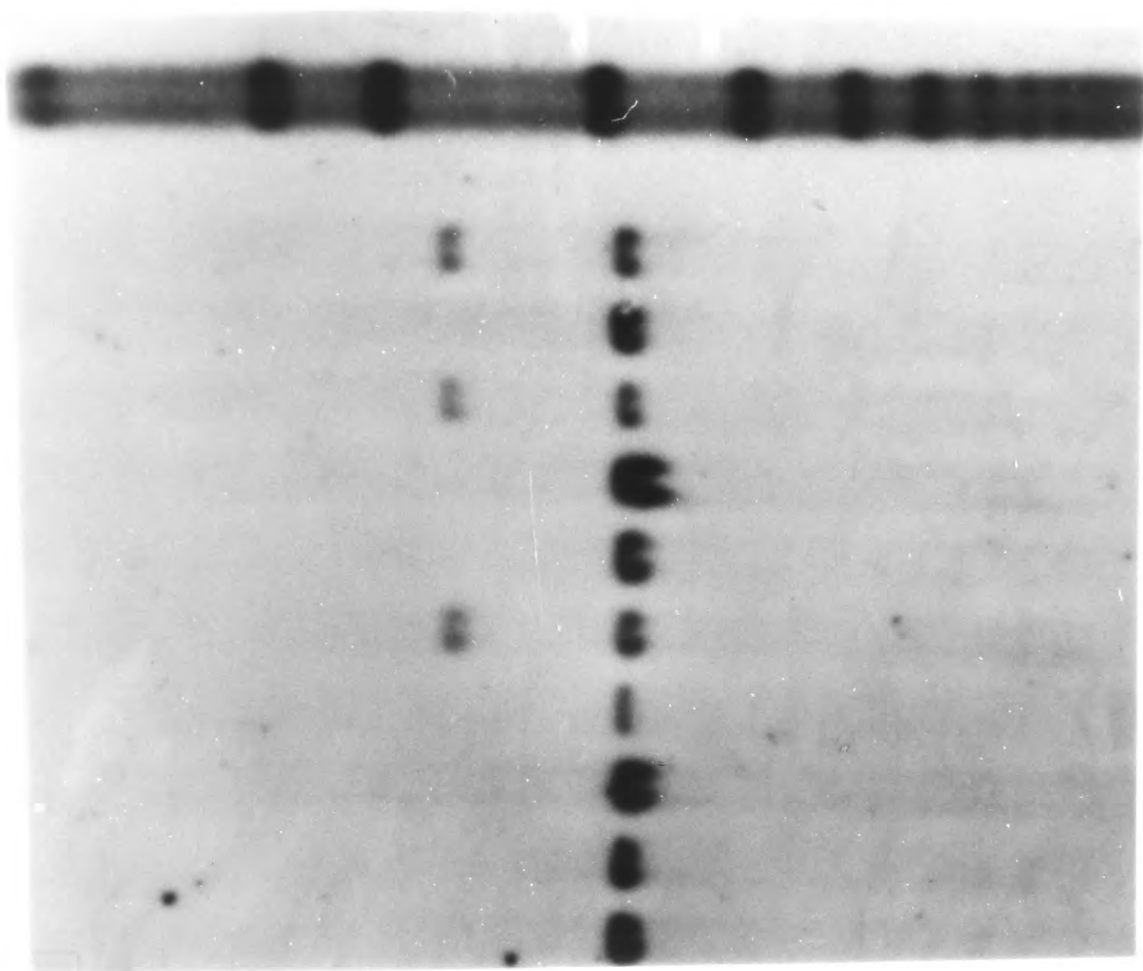
- indicates the absence of a band

$\frac{1}{2}$ indicates half intensity of a band compared to +

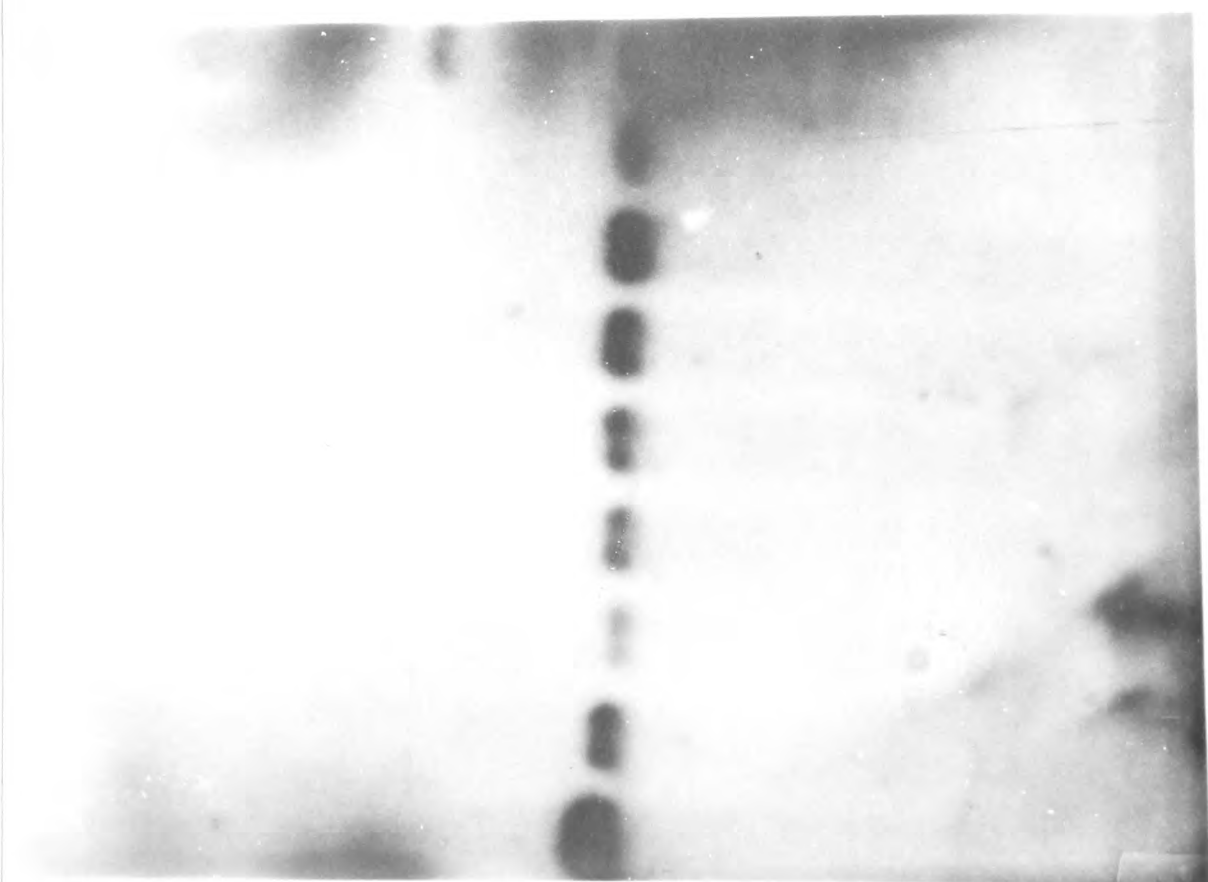
? indicates that the precise allele could not be assigned to the locus

Figure 5.1

Southern analysis of 15 patients with the restriction enzyme NcoI. The two fragments observed are at 3.1 kb and 2.3 kb. The patients who have the 2.3 kb fragment are RW, MP, CB and JB. AR is the sibling of proband LR, Mr. and Mrs. R are the parents. B is the father of proband JB and does not have the 2.3 kb fragment.



RW *
 AIB*
 MP *
 ZT *
 CO *
 CB *
 AR *
 LR *
 Mrs. R
 Mr. R



JB *
 Father B
 DW*
 SG *
 TC *
 KC *
 LM*
 CIB*
 LIM*

DR1,7 and patient BB with the deleted 21-OHase B genes has the haplotypes A3,23 B39,44 DR2,7. Twelve of the other patients have at least one 21-OHase B gene corresponding to the 3.7 kb TaqI and 3 kb KpnI fragment. The remainder of the patients have at least two 21-OHase B genes present. The NcoI polymorphism was found in 8 patients.

It was observed that the Bw47, DR7 haplotype in the patients always consistently had the NcoI RFLP of 2.3 kb as seen from the banding patterns of the patients CB, AB, CBu and II.I. Since the Bw47 haplotype has a deleted 21-OHase B gene (White et al., 1984b) the NcoI RFLP is in the 21-OHase A gene. The patient JB who has a homozygous deletion of the 21-OHase B genes also exhibits the NcoI RFLP in one of the two 21-OHase A genes. His father AB with the HLA haplotype A9,3? B7,44 Cw4,8 DR1,4 DRw57 (? = not known precisely) does not show the NcoI RFLP (Figure 5.1). Therefore the haplotype which carries the RFLP in its 21-OHase A gene is A32 B13 DR7.

The patient MP having the NcoI RFLP in one of her genes has the HLA haplotypes A2,3 B40,Bw3 Cw6 DR? with two 21-OHase A genes and one 21-OHase B gene. However, without any family data it is not possible to determine on which gene the RFLP lies. RW with the HLA haplotypes A1,29 B44,40/8 Cw5 DR3,7 DRw52 has one 21-OHase A and two 21-OHase B genes and has the NcoI RFLP in one of her genes, but it cannot be precisely located on the basis of the data available.

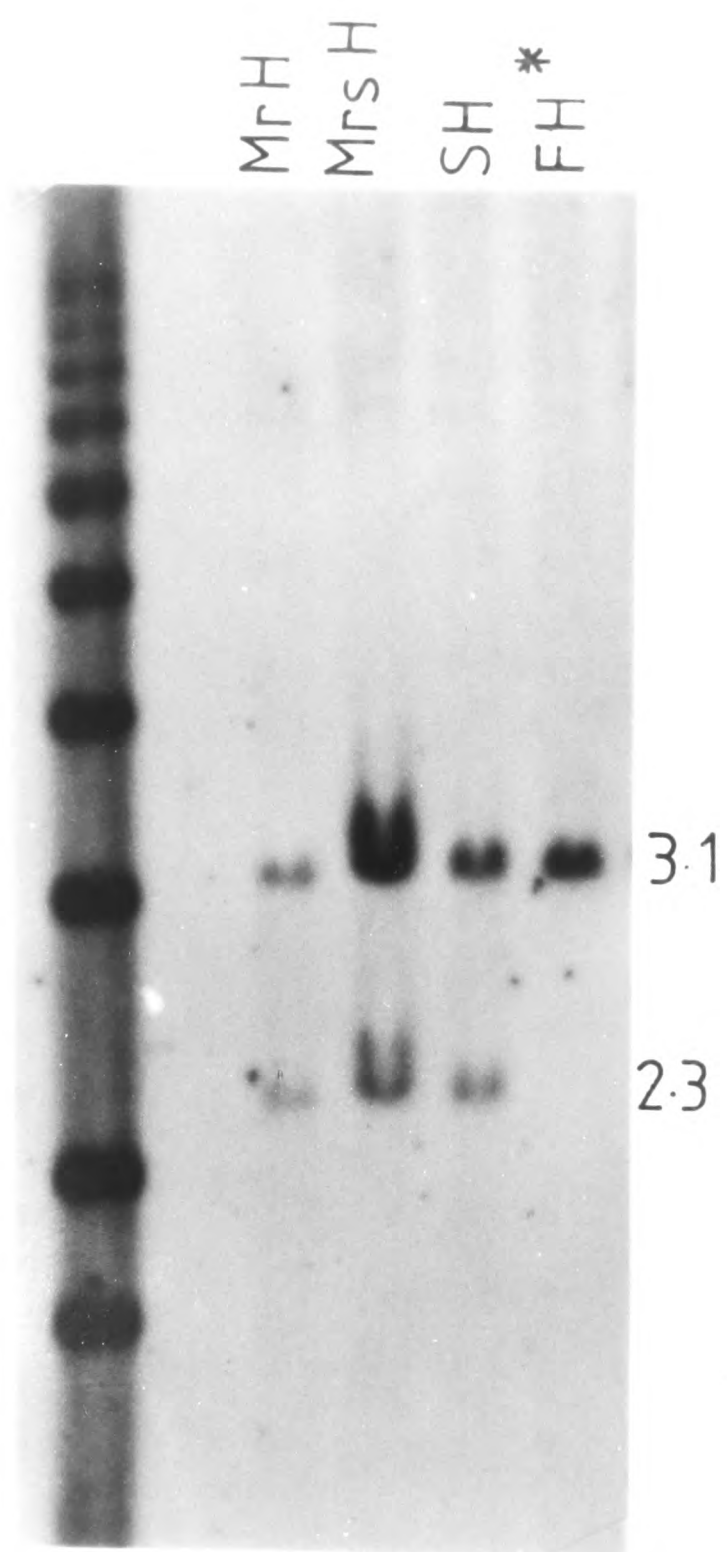
5.2.1 Family H

Individual FH from family H has the non salt losing form of CAH and the HLA haplotypes A2 B51 Cw6 DR2,4. She has two 21-OHase A genes and one 21-OHase B gene. Southern blot analysis of the family showed that both parents and the normal sibling SH had the NcoI polymorphism while FH did not (Figure 5.2). Their TaqI and KpnI banding patterns are

Figure 5.2

Southern analysis of family H with NcoI.

The parents Mr. H and Mrs. H and the sibling SH have both fragments of 3.1 kb and 2.3 kb. The proband FH is indicated by an asterisk (*). The proband has a single fragment of 3.1 kb. The 3.1 kb fragment in Mrs. H is four times as intense as the 2.3 kb fragment.

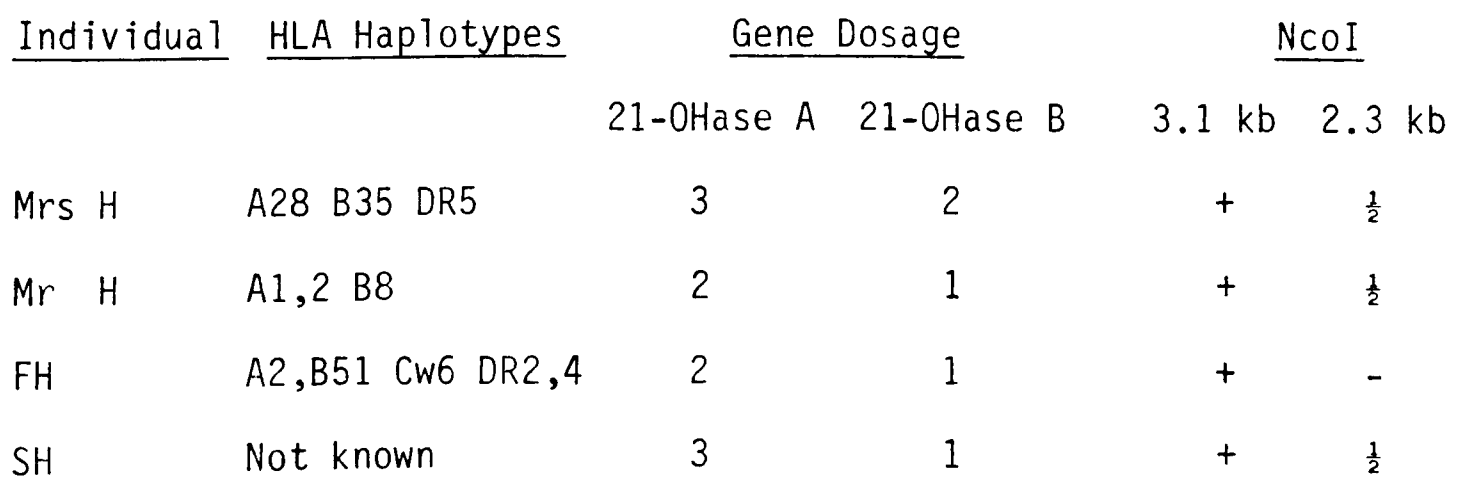


Nco I

indicated in figure 5.3. As the HLA haplotypes of the sibling were not available the NcoI polymorphism could be allocated only by deduction. Since FH has two 21-OHase A genes and one 21-OHase B gene, none of which have the NcoI RFLP, she has inherited one 21-OHase A gene from one parent and the remaining two loci from the second parent. The father has two 21-OHase A genes and one 21-OHase B gene, hence his 21-OHase B gene has to be normal. The NcoI RFLP is presumably in the 21-OHase A gene on the same haplotype as the normal 21-OHase B gene since it is not present in the single 21-OHase A gene on the haplotype which he has passed to his daughter. The single 21-OHase B gene in FH has come from her mother along with one 21-OHase A gene both of which do not have the NcoI RFLP. The mother though would have to have another 21-OHase B gene for her to be normal. The TaqI banding pattern indicates + and 1/2 intensities for the 3.2 kb and 3.7 kb fragments respectively. As the mother has two 21-OHase B genes which show about half the intensity of the 21-OHase A genes in Southern analysis it is possible that she has three 21-OHase A genes. The NcoI banding pattern suggests that this is likely since the 3.1 kb fragment is four times as intense as the 2.3 kb fragment. As the mother has one normal 21-OHase B gene and one 21-OHase B gene which is associated with CAH on the A2, B51, Cw6, DR2 haplotype, but does not have the NcoI RFLP, the RFLP must be in one of the two 21-OHase A genes on the same haplotype (A28 B35 DR5) as the normal 21-OHase B gene. The sibling SH could have inherited the NcoI RFLP from either or both of her parents. Her NcoI pattern however suggests that only one of her genes has the NcoI RFLP, therefore she has inherited one haplotype with a normal 21-OHase B gene and a 21-OHase A gene with the NcoI RFLP. Her other haplotype must be an affected haplotype and she is a carrier of the 21-OHase deficiency. Figure 5.3 illustrates the pattern of inheritance of the 21-OHase genes in family H.

Figure 5.3

The organization of the 21-OHase genes in the family H. The mother has two 21-OHase A genes and one 21-OHase B gene on one of her haplotypes (A28, B35, DR5). The NcoI RFLP is carried by one of the 21-OHase A genes. The 21-OHase B gene on this haplotype is normal. The A2, B51, DR2 haplotype which is affected has one 21-OHase A and one 21-OHase B gene each. The father has one 21-OHase A and one 21-OHase B gene on his normal haplotype. The NcoI RFLP is on the 21-OHase A gene of that haplotype (shown by *). The second haplotype which has a single 21-OHase A gene is the affected haplotype (affected haplotypes are shown by #). Both parents are carriers shown as half shaded and the affected proband (fully shaded) who has CAH has inherited the affected haplotypes without the NcoI RFLP. The unaffected SH sibling has the RFLP in one of 21-OHase her genes she has inherited an affected haplotype from her father and she is a carrier.



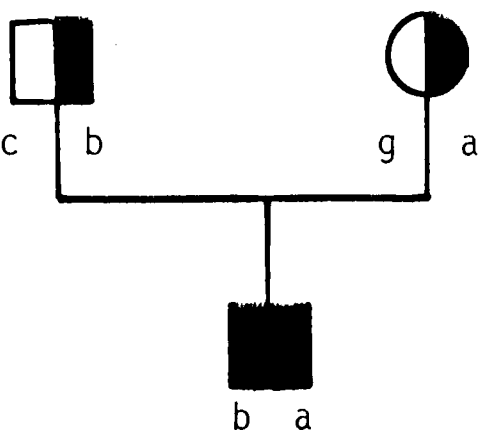
5.2.2 Family Bu

The pedigree of family Bu and its HLA and complement haplotypes are shown in Figure 5.4. The patient CBu has inherited the 21-OHase deficiency on haplotype (b) A2 Bw47 C6 DR7 from his father and haplotype (a) A11 B55 C3 DR4 from his mother. The 21-OHase gene dosage was analysed by Southern blot analysis using TaqI (Palsdottir, 1986). On the basis of the band intensity it was deduced that the patient had two 21-OHase A genes but only one 21-OHase B gene. His mother has two 21-OHase A and two 21-OHase B genes. Although no DNA was available from the father, his HLA haplotypes A2,1 Bw47,8 C6 DR7,3 suggest the presence of one 21-OHase A gene and one 21-OHase B gene in total. This was confirmed by the Southern blot analysis of the grand parents of CBu (Palsdottir, 1986). The Bw47, DR7 haplotype always has a 21-OHase B gene deletion and the B8 DR3 haplotype, a 21-OHase A gene deletion. The single 21-OHase B gene in patient CBu is inherited from his mother on the A11, B55 C3, DR4 haplotype. The NcoI digest of CBu and his mother are shown in Figure 5.5. The mother has a single 3.1 kb fragment while CBu has both 3.1 kb and 2.3 kb fragments, the 2.3 kb fragment being twice as intense as the 3.1 kb fragment. Since the 2.3 kb fragment is absent in the mother, CBu could have inherited it only from his father. The Bw47 haplotype which he has inherited from his father indicates the presence of a single 21-OHase A gene, yet the NcoI RFLP is present in two 21-OHase genes. The presence of the NcoI RFLP in two of the three 21-OHase genes in individual CBu suggests that one of his genes may have acquired the RFLP and not inherited it. This result does not indicate which of the 21-OHase genes the NcoI RFLP is present in.

Figure 5.4

Pedigree and HLA haplotype of family Bu.

Proband CBu has inherited the 21-OHase deficiency on haplotypes a and b, from the mother and the father respectively.



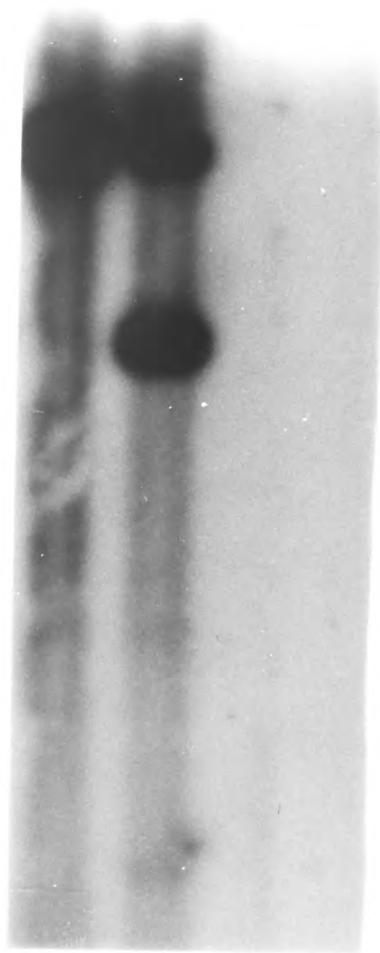
HLA / Complement Haplotypes of family Bu

							Gene Dosage		NcoI	
							21-OHase A	21-OHase B	3.1	2.3
MOTHER	a)	A11	B55	Cw3	DR4	C4A3,2 BQ0	1	1	+	-
	g)	A2	B51	Cw2	DR4	C4A3 BQ0	1	1		
FATHER	b)	A2	Bw47	Cw6	DR7	C4A1 (BQ0)	1	-		
	c)	A1	B8	Cw6	DR3	C4AQ0 B1	-	1		
CBu	a)	A11	B55	Cw3	DR4	C4A3,2 BQ0	1	1	$\frac{1}{2}$	+
	b)	A2	Bw47	Cw6	DR7	C4A1 (BQ0)	1	-		

Figure 5.5

Southern analysis of CBU and his mother with NcoI. The mother has one fragment at 3.1 kb while CBU has two fragments at 3.1 kb and 2.3 kb. The 2.3 kb fragment is twice as intense as the fragment 3.1 kb.

Mother
CBu*



-3.1

-2.3

Nco I

5.2.3 Family M

The pedigree and HLA and complement haplotypes of family M are shown in Figure 5.6. Two of the children D & S are adrenogenics and have identical haplotypes while the third child is normal. All the members of the family with the exception of the mother have two 21-OHase A genes and one 21-OHase B gene, as can be seen from the TaqI digest pattern (Figure 5.7). The NcoI pattern shows all the members of the family other than the mother having the 3.1 kb fragment twice as intense as the 2.3 kb fragment while the 2.3 kb fragment is absent in the mother (Figure 5.7). Both affected probands have the haplotypes A3,2 B7,44 C5 and are homozygous for the syndrome. The A2, B44 haplotype from the mother has both 21-OHase loci present and obviously does not carry the NcoI RFLP. Therefore all three children who have the NcoI polymorphism have inherited it from the A3, B7 haplotype from their father which has only one 21-OHase A gene. The NcoI RFLP in family M is, therefore, in the 21-OHase A gene on the A3,B7 haplotype.

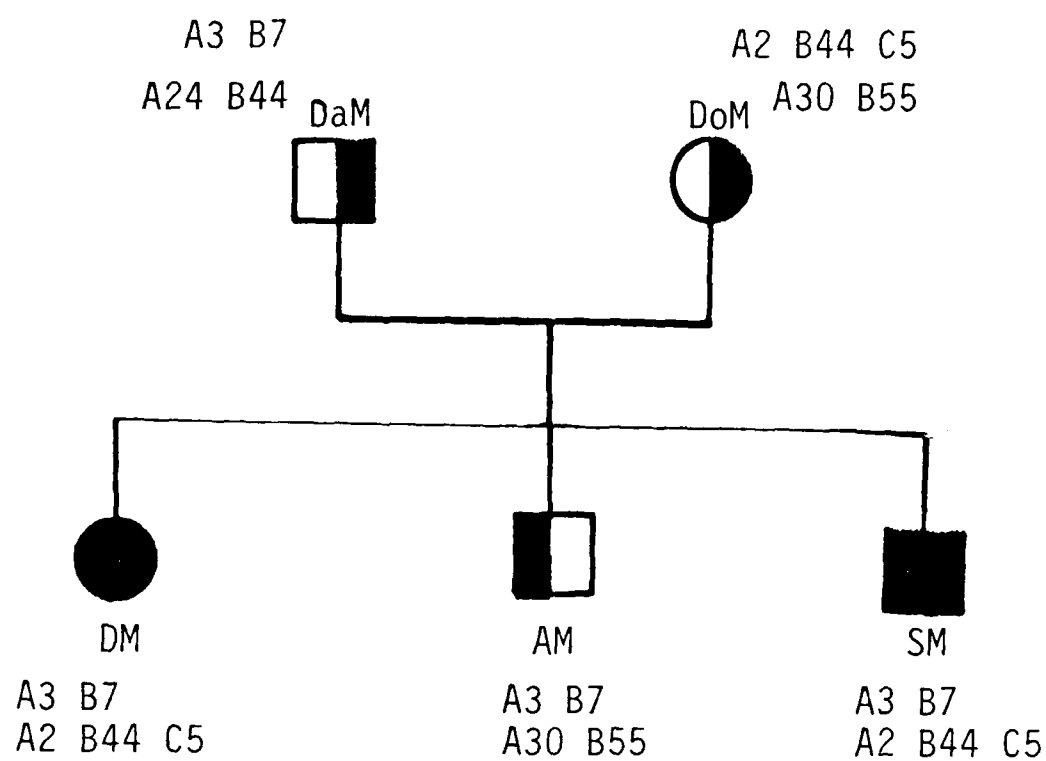
5.3 NcoI RFLP in a Random Population

A study of individuals in a random population was carried out to find the frequency of occurrence of the NcoI polymorphism. Out of 17 individuals studied, 6 individuals showed the NcoI polymorphism (Figure 5.8). Table 5.2 shows the haplotypes of the individuals studied. This suggested that the NcoI RFLP was present at a substantially high frequency of 9% in a random population although it did not give any information as to which of the haplotypes and genes had the RFLP. Siblings of some of the individuals were also analysed for the RFLP. JH who has the NcoI RFLP has the haplotype A29, 3 B44,7 Cw- DR5,2, his sibling SJ who does not have the NcoI RFLP has the haplotypes A2,3 B51,7. Therefore the NcoI RFLP is contained in the A29, B44 haplotype of JH. NJ with

Figure 5.6

Pedigree and HLA haplotypes of family M.

D and S are adrenogenics (shown fully shaded) and have inherited the defect from the A2, B44 haplotypes from the mother and A3, B7 haplotype from their father. The sibling shares the A3, B7 haplotype with the affected probands and is a carrier (half shaded).



HLA Haplotypes of family M

				Gene Dosage		TaqI		NcoI	
				21-OHase A	21-OHase B	3.2	3.7	3.1	2.3
MOTHER	DoM	A30,2	B55,44 C5	2	2	+	+	+	-
FATHER	DaM	A3,24	B7,44	2	1	+	½	+	½
PROBAND	DM	A3,2	B7,44 C5	2	1	+	½	+	½
PROBAND	SM	A3,2	B7,44 C5	2	1	+	½	+	½
SIBLING	AM	A3,30	B7,55	2	1	+	½	+	½

Figure 5.7

Southern analysis of family M with TaqI and NcoI. The father and three children all have two fragments at 3.1 kb and 2.3 kb with NcoI. The mother has a single fragment at 3.1 kb. The TaqI pattern shows two fragments at 3.7 kb and 3.2 kb in all individuals. The affected probands are shown with (x).

Family M

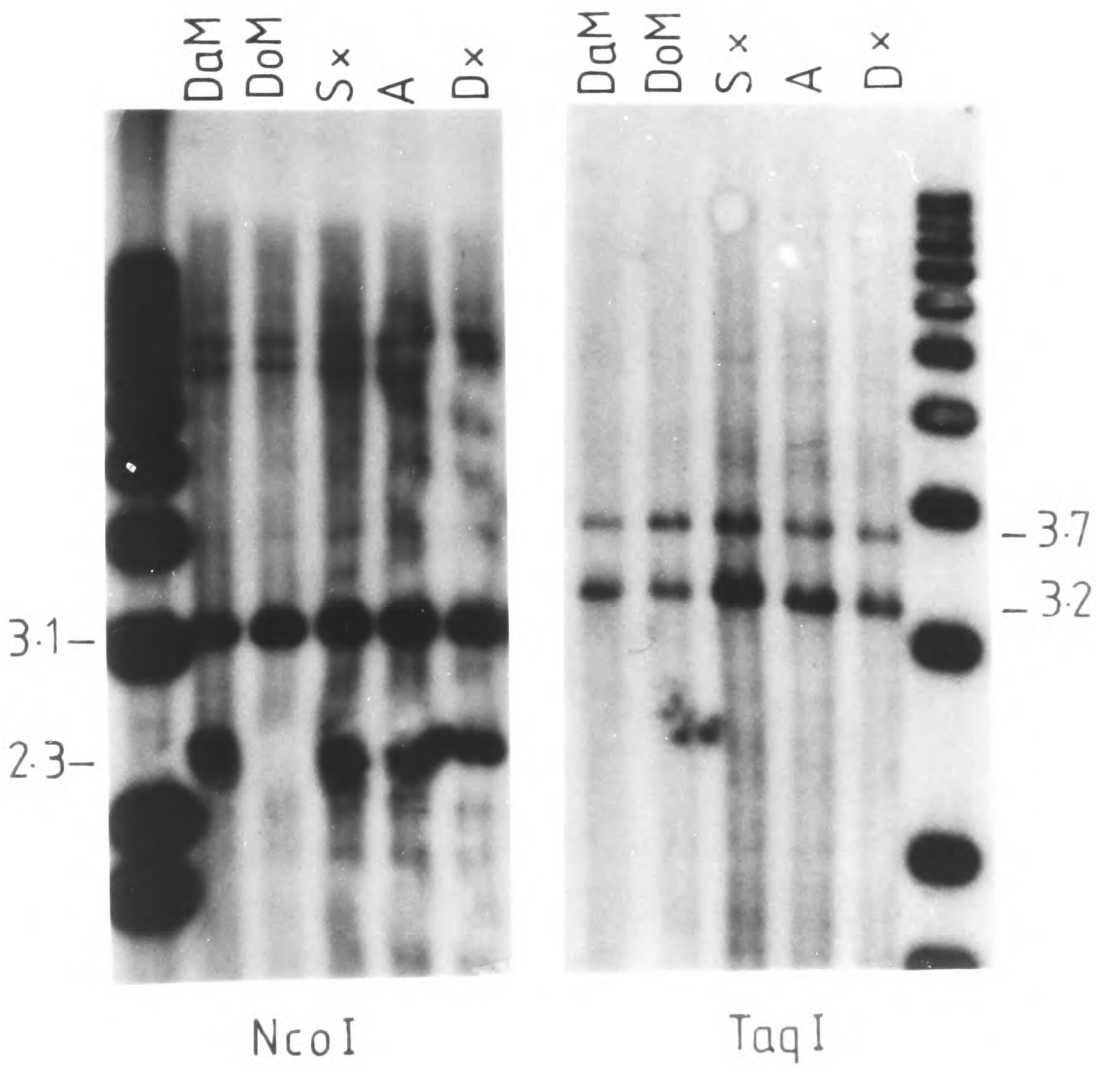


TABLE 5.2 Individuals from a Random Population

<u>No.</u>	<u>Name</u>	<u>HLA Haplotype</u>				<u>NcoI</u>	
						3.1	2.3 kb
1	AC	A2,3	B8,7	Cw-	DR3,2	+	-
2	DC	A2,1	B8,16(39)	Cw-	DR3	+	-
3	JD	A32,38	B14,40(60)	Cw-	DR2,4	+	-
4	DoC	A1,28	B17,40	Cw6,w3	DR7,4	+	-
5	SJ	A2,3	B51,7	Cw5	DR2	+	-
6	JH	A29,3	B44,7	Cw-	DR5,2	+	$\frac{1}{2}$
7	RH	A9,32	B12,55(22)	Cw-	DR4/61	+	-
8	PH	A32	B21?,55(22)	Cw4	DR5,4/61	+	-
9	NJ	A2,11	B51(or blank),51		DR7	+	$\frac{1}{2}$
10	PJ	A11,31	B51,62		DR7	+	-
11	WG	A28,3	B35,14	Cw3,4	DRw52,53	+	$\frac{1}{2}$
12	CG	NOT AVAILABLE				+	$\frac{1}{2}$
13	AP	A1	B17,7	Cw6	DR7,2	+	-
14	MeP	A29,(29)	B44,(44)		DR7,2	+	$\frac{1}{2}$
15	MBP	A1,29	B17,44		DR7,2	+	-
16	AA					+	$\frac{1}{2}$
17	BA					+	-

+ indicates the presence of a band

- indicates the absence of a band

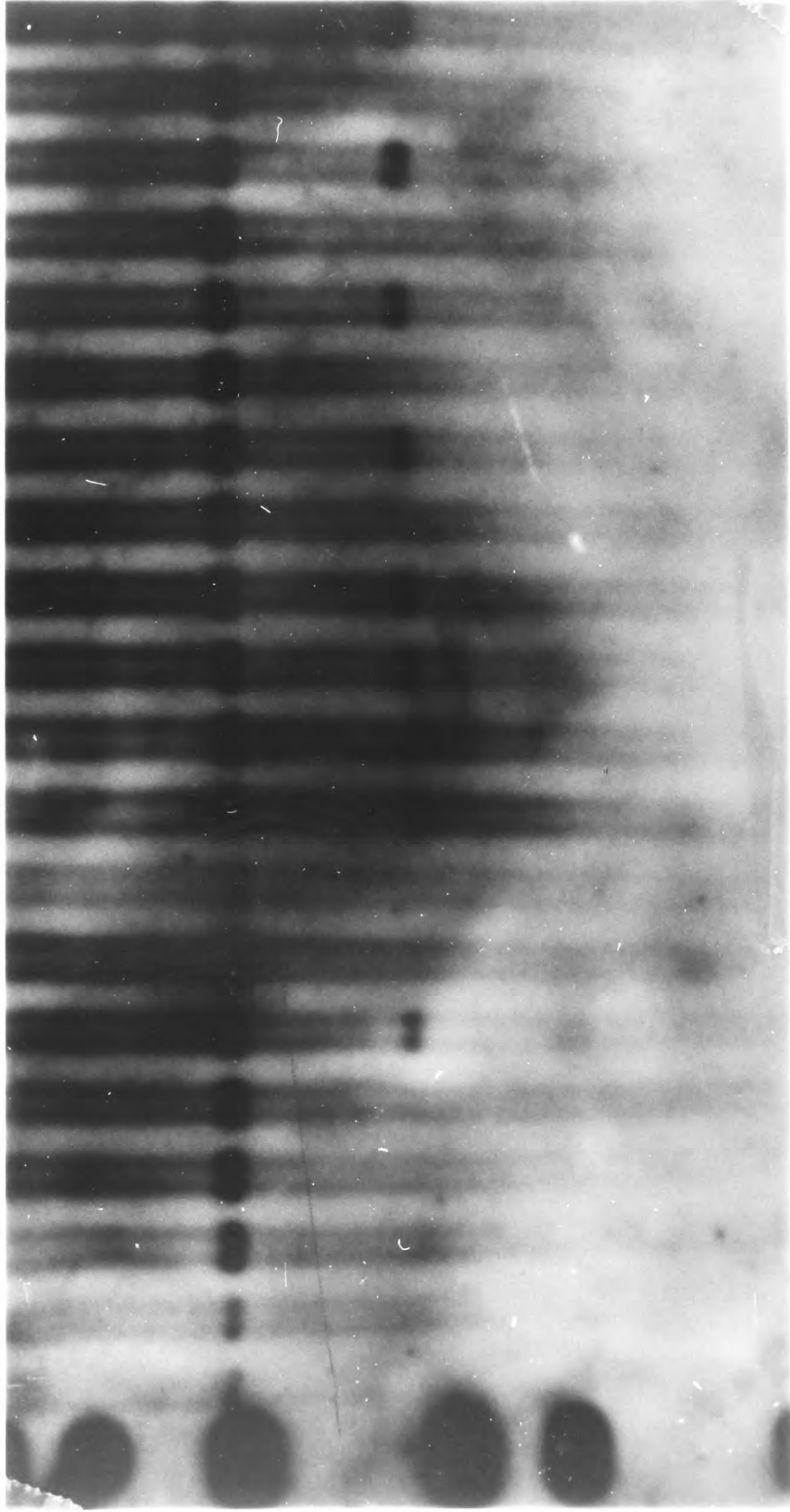
$\frac{1}{2}$ indicates half intensity of a band compared to +

? indicates that the precise allele could not be assigned to the locus

Figure 5.8

NcoI polymorphism in a random population. Lanes 1 - 17 are normal individuals. Lanes 18, 19 and 20 are individuals from a family with adrenogenics. All individuals have the 3.1 kb fragment, while six individuals from the unaffected population (Lanes 6, 9, 11, 12, 14, 16) also have the 2.3 kb fragment. Individuals 18 and 20 from the adrenogenic family also have the 2.3 kb fragment.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20



Nco I

haplotypes A2,11 B51 (or blank), 51 DR7 has the RFLP while sibling PJ who does not have the NcoI RFLP has haplotypes A11,31 B51,62 DR7. Hence the NcoI RFLP can be located to the A2,B51 (or blank) haplotype. WG and her son CG both have the NcoI RFLP but since only the mother's haplotypes A28,3 B35,14 Cw52,53 are known, the NcoI RFLP cannot be ascribed to any particular haplotype.

5.3.1 Family P

The pedigree of family P and their band patterns with TaqI and NcoI are shown in figure 5.9. All the members have both 21-OHase genes present in their haplotypes. The father and son does not show the RFLP while the mother and daughter do. The mother and daughter both have the haplotype, A29 B44 DR7 in common which must carry the NcoI RFLP. Since both 21-OHase genes are present on this haplotype, the NcoI RFLP could be on either of the two genes and this analysis would be unable to detect which gene it is present in.

5.3.2 Family A

In the family A, the father and all the children have the NcoI RFLP. The pedigree and the HLA haplotypes are indicated in figure 5.10. The haplotype which is shared in common by all the individuals with the NcoI RFLP, is A19,29 B12 Bw44 C- DR7 but since both 21-OHase genes are present as indicated by the TaqI digest, the RFLP could be on either of the genes.

The A29 B44 DR7 haplotype is common to both families presented here and the NcoI polymorphism is presented on this haplotype in both families. Since the A29 B44 DR7 haplotype exists as an extended haplotype in the population (Dawkins et al., 1983), the NcoI RFLP is probably always likely to occur on the haplotype.

Figure 5.9

Southern analysis of family P with TaqI and NcoI. The mother and daughter exhibit both the 3.1 kb and 2.3 kb NcoI fragment, while the father and sons only have the 3.1 kb fragment.

HLA Type						Gene Dosage		NcoI	
						21-OHase A	21-OHase B	3.1	2.3
FATHER	AP	a)	A1	Cw6	B17 DR7 GL02	1	1	+	-
		b)	A	C	B7 DR2 GL01	1	1		
MOTHER	MeP	c)	A29	C	B44 DR7 GL01	1	1	+	$\frac{1}{2}$
		d)	A29	C	B44 DR2 GL01	1	1		
SON	MBP	a / d				2	2	+	-
SON	PP	Not available				2	2	+	-
SIBLING	FP	a / c				2	2	+	$\frac{1}{2}$

Family P

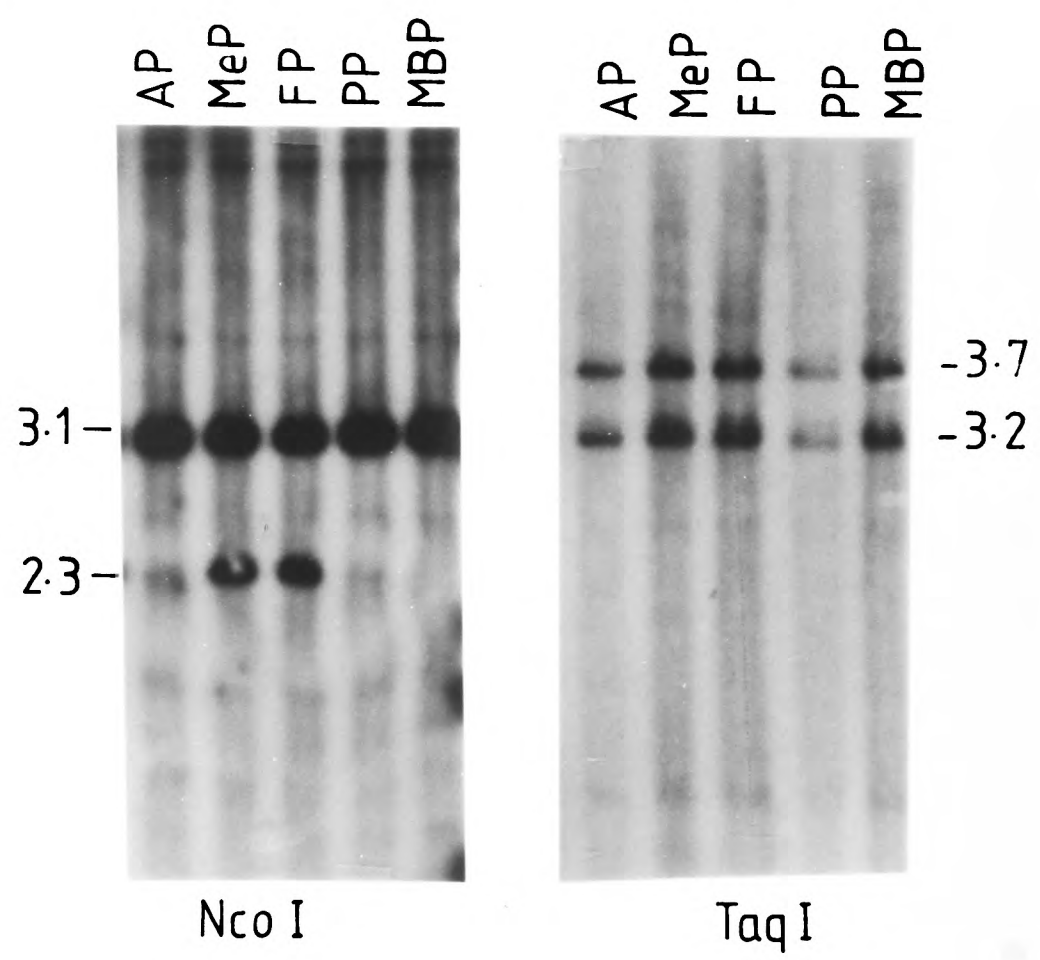
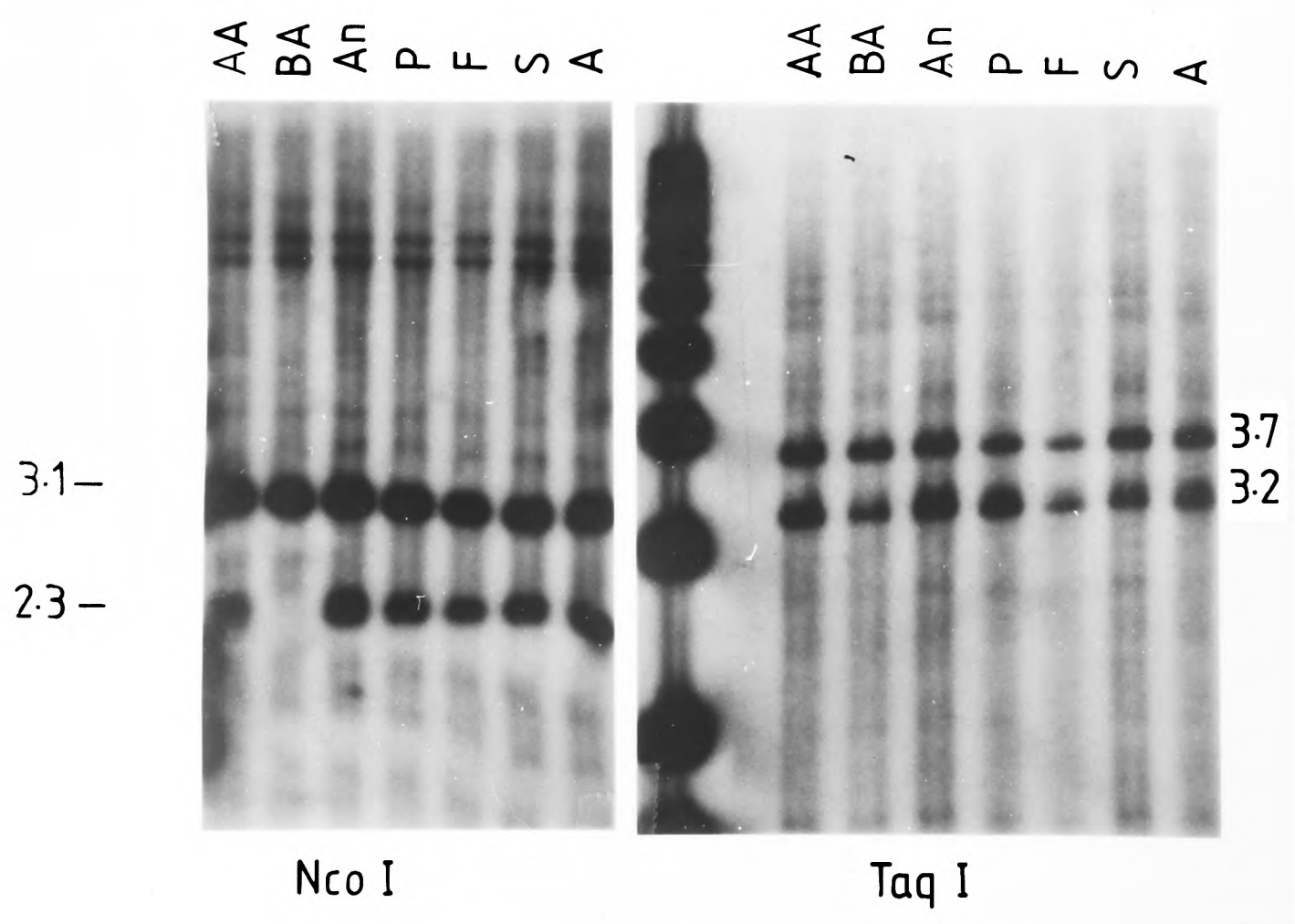


Figure 5.10

Southern analysis of family A with NcoI and TaqI. The father AA and all the children show the presence of the 2.3 kb NcoI fragment with the 3.1 kb fragment. The mother BA has only the 3.1 kb fragment present. The TaqI pattern indicates that all the genes are present as in a normal haplotype.

HLA Type			Gene Dosage		TaqI		NcoI	
			21A	21B	3.2	3.7	3.1	2.3
FATHER AA	a) A2 Bw48 C- DR8		1	1	+	+	+	$\frac{1}{2}$
	b) A19 29 B12 Bw44 C- DR7		1	1				
MOTHER BA	c) A1 B8 Cw7 DR3		-	1	$\frac{1}{2}$	+	+	-
	d) A2 B12 Bw44 Cw5 DR4		1	1				
SON An		b / d	2	2	+	+	+	$\frac{1}{2}$
SON P		b / d	2	2	+	+	+	$\frac{1}{2}$
SIBLING F		b / d	2	2	+	+	+	$\frac{1}{2}$
SIBLING S		b / c	1	2	$\frac{1}{2}$	+	+	$\frac{1}{2}$
SON A		b / c	1	2	$\frac{1}{2}$	+	+	$\frac{1}{2}$



5.4 AvaII Polymorphism

To differentiate between the 21-OHase A and B genes around the NcoI RFLP, a restriction endonuclease that gives distinct unique bands was selected. On examination of the sequences around the NcoI restriction site at position 1653 it was found that AvaII sites at nucleotide positions 1233 and 1708 in the 21-OHase B gene yielded a 475 bp fragment. A single nucleotide substitution at position 1380 in the 21-OHase A gene introduced a new AvaII site into the 21-OHase A gene which would yield two fragments of 147 bp and 328 bp instead of a single fragment of 475 bp (Figure 5.11). Southern blot analysis of plasmid DNA containing 21-OHase A, 21-OHase B and the 21-OHase B gene from the individual II.I gave the expected bands with the AvaII digest (Figure 5.12). The probe used was AN which was obtained by isolating a 420 bp fragment from an AvaII / NcoI digest of the 3.7 kb TaqI subclone of the defective 21-OHase B gene from the individual II.I. Thus on the basis of the AvaII digest, 21-OHase A and 21-OHase B genes can be distinguished.

If there is a NcoI RFLP in the gene, the AvaII digest followed by a NcoI digest would yield fragments that a normal gene would not have. Table 5.3 shows the pattern expected in a normal haplotype and in haplotypes with the NcoI RFLP present on either of the genes or both of the genes. Thus a normal haplotype, that is both genes without the NcoI, RFLP would give the following pattern of fragments; 147 bp and 328 bp corresponding to the 21-OHase A gene and 475 bp corresponding to the 21-OHase B gene respectively. A 21-OHase A gene which has the NcoI RFLP would give 147 bp and 274 bp fragments, while a 21-OHase B gene with the NcoI RFLP would give a 420 bp fragment, respectively. Theoretically the AvaII and AvaII / NcoI digests should be able to define which gene has the NcoI RFLP.

Figure 5.11

AvaII restriction sites in the 21-OHase genes between nucleotide position 1233 and 1708. There is a single base pair substitution T → C at position 1380 in the 21-OHase A gene which introduces a new AvaII site. The AvaII sites are marked in blue colour and the NcoI site is marked in red colour.

L K T W S H W S I Q I V D V I P F L R 217.
TTAAAAACCTGGAGCCACTGGTCCATCCAAATTGTGGACGTGATTCCTTCTCAGGGTGAGGACCTGGAGCCTAGACACCCCTGGGTTGTAGGGGAGAGGCTGGGGTGGAGGGAGAGGC 1288
G
A*

TCCTTCCCACAGCTGCATTCTCATGTCTTCCTGCCGCAGTTCTTCCCCAATCCAGGCTCTCCGAGGCTGAAGCAGGCCCATAGAGAAGAGGGATCACATCGTGGAGATGCAGCTGAGGCAGC 1408
C A* A* A* A*

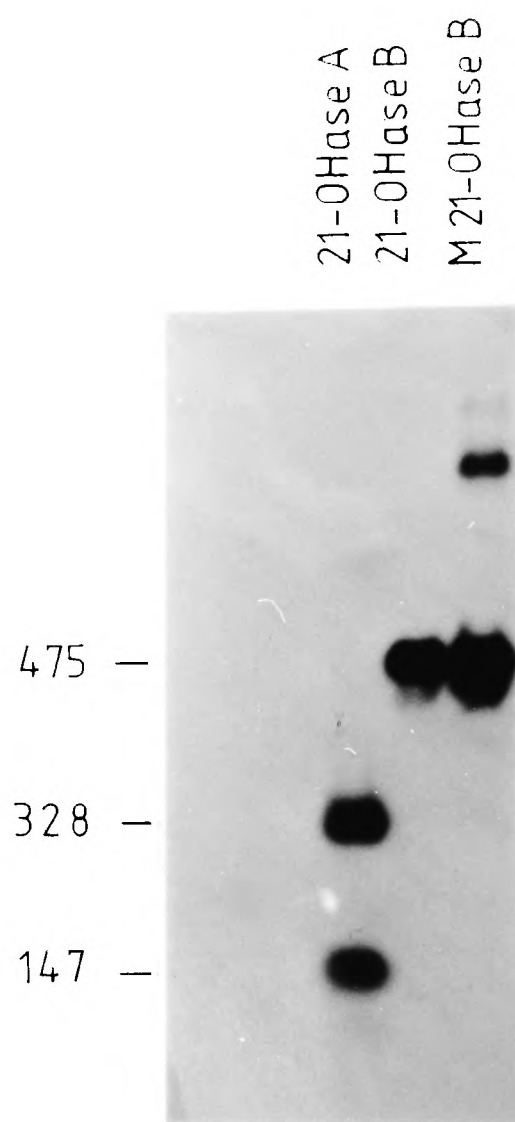
H K 246 .
ACAAAGGTGGGACTGTACGTGGACGGCCCTCCCTCGGGCCCCACAGCCAGTGATGCTACCGGCCCTCAGCATTTGCTATGAGGGGGTCTCTTTTGGCATACCCAGTTATGGGCTGTGGCCACT 1528
GT
GT

CTGTACTCCTCTCCCCAGGCCCGCTCAGCCCCGCTCCTTTCACCCCTCTGCAGGAGAGCCTCGTGGCAGGCCAGTGAGGACATGATGGACTACATGCTCCAAGGGGTGGCGCAGCCG 1648
G
T*

S M E E G S G Q L L E G H V H M A A V D L L I G G T E T A N T L S W A V V F L
AGCATGGAAGAGGGCTCTGGACAGCTCCTGGAAGGGCACGTCACATGGCTGCAGTGGACCTCCTGATCGGTGGCACTGAGACCACAGCAAAACACCCCTCTCCTGGGCCGTGGTTT-TT 1767
T*

Figure 5.12

AvaII digest of the BglII / BamHI subclones of 21-OHase A gene, 21-OHase B gene and 21-OHase B gene (M21-OHase B) from individual II.I. The 21-OHase A gene has two fragments at 328 bp and 147 bp while both 21-OHase B genes have one fragment at 475 bp which lights up with probe AN (AvaII / NcoI - 420 bp fragment from 3.7 λ 21B). A partially digested fragment is also seen in M21-OHase digest at 1 kb.



Ava II

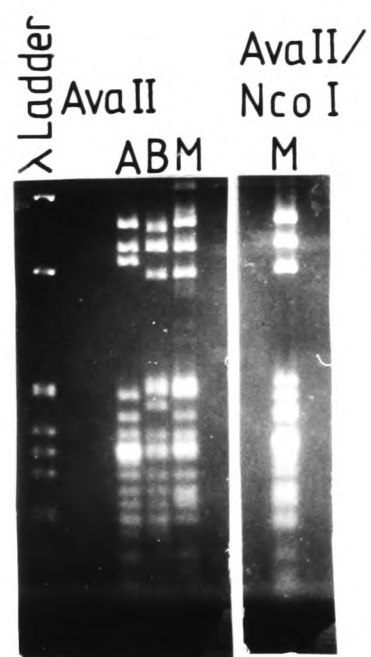


TABLE 5.3 Pattern of AvaII / NcoI Fragments on Haplotypes with the NcoI RFLP

<u>Gene Type</u>	<u>Bands</u>	<u>Observed</u>
Normal 21-OHase A, and 21-OHase B	A - 147 bp A - 328 bp	B - 475 bp
Mutant 21-OHase A, and normal 21-OHase B	A - 147 bp A - 274 bp	B - 475 bp
Normal 21-OHase A, and mutant 21-OHase B	A - 147 bp A - 328 bp	B - 420 bp
Mutant 21-OHase A, and mutant 21-OHase B	A - 147 bp A - 274 bp	B - 420 bp

To study the *Ava*II / *Nco*I polymorphism individuals BB, Ice5, CBu, II.I and KEM were selected (Table 5.4). BB shows no 3.7 kb *Taq*I fragment and hence was presumed to have a homozygous deletion of the 21-OHase B genes. Ice5 has only the 21-OHase B genes present due to homozygous deletion of the 21-OHase A genes; CBu has two 21-OHase A genes and one 21-OHase B gene; similarly II.I has two 21-OHase A genes and one 21-OHase B gene and KEM has two 21-OHase B genes.

5.4.1 AvaII Band pattern

The *Ava*II digest of BB gave fragments at 475, 328 and 147 bp (Lane 1, Figure 5.13). The 475 bp fragment corresponds to a 21-OHase B gene, the 328 bp and 147 bp fragments correspond to the 21-OHase A gene. This was an unexpected result as BB does not have a 3.7 kb *Taq*I fragment characteristic of a 21-OHase B gene. Ice5 showed a single fragment at 475 bp corresponding to the 21-OHase B genes (Lane 2). CBu and II.I showed a single fragment at 328 bp, the 147 bp fragment being very faint (Lane 3 and 4), corresponding to the 21-OHase A gene. Both showed a 475 bp fragment corresponding to the 21-OHase B gene. In addition to these expected fragments, CBu had a fragment at 520 bp which was twice as intense as the fragment at 475 bp (Lane 3). Individual II.I had a fragment at 520 bp of the same intensity as the fragment at 475 bp (Lane 4). KEM showed the presence of two fragments, one at 475 bp which was twice as intense as the one at 520 bp. There was no fragment at 328 bp (Lane 5) which supports the absence of the 21-OHase A genes in KEM. The presence of the 520 bp fragment was unexpected and could not be explained except as a partial digest.

5.4.2 AvaII / NcoI Band pattern

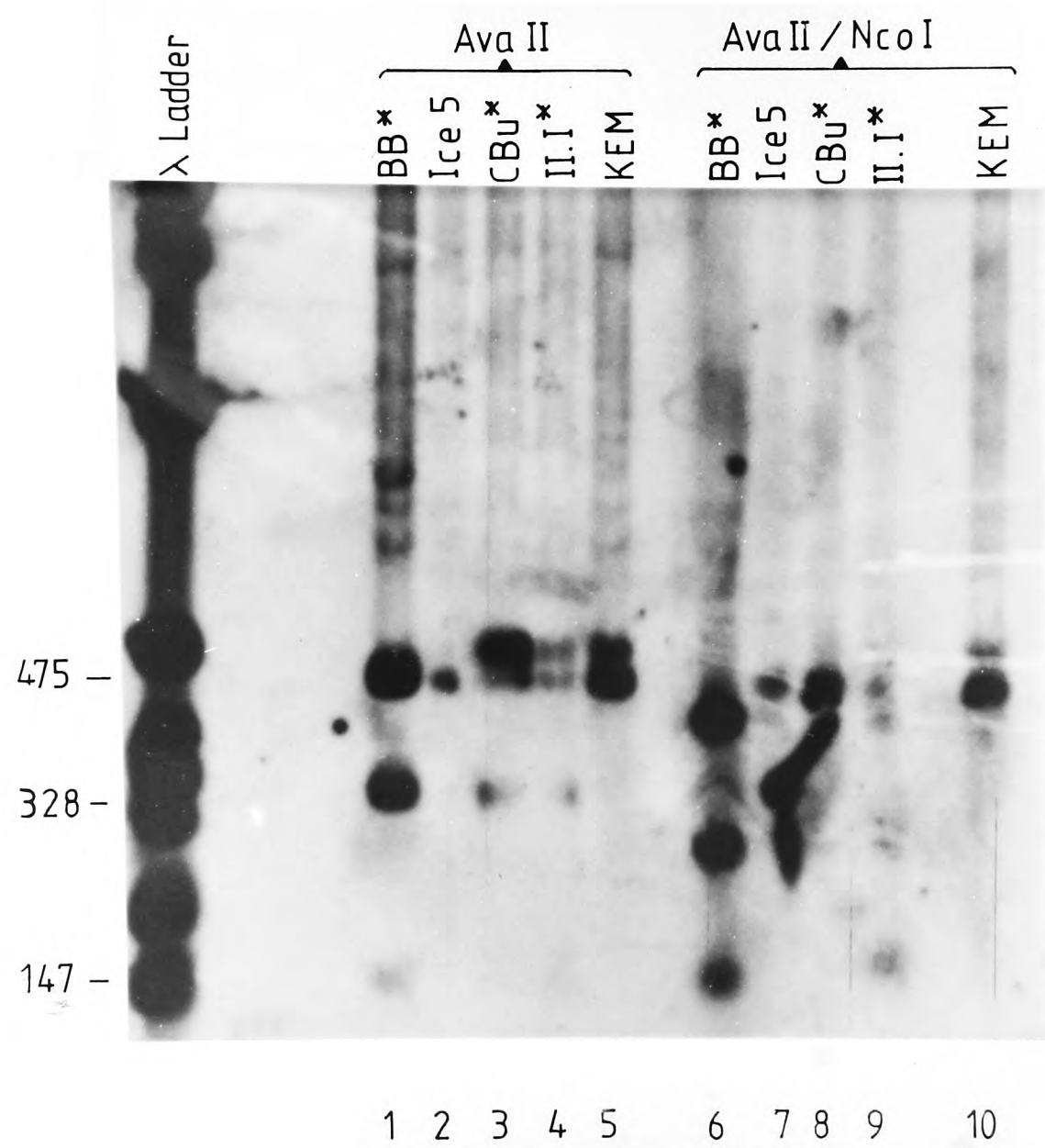
In the *Ava*II / *Nco*I digest, both fragments of 475 bp and 328 bp in

Figure 5.13

AvaII and AvaII / NcoI digests of genomic DNA from affected individuals BB, CBu and II.I and normal individuals Ice5 and KEM. Lanes 1 - 5 are AvaII digests and lanes 6 - 10 are AvaII / NcoI double digests. The 475 bp fragment in the AvaII digest is characteristic of the 21-OHase B gene. The 328 bp and 147 bp fragments are characteristic of the 21-OHase A gene. In the AvaII / NcoI digest, the presence of the NcoI RFLP cuts the 475 bp fragment to 420 bp and 328 bp fragment to 274 bp. There is an additional fragment of 520 bp in some individuals. The explanation is given in the text.

Table 5.4

<u>Individual</u>	<u>Haplotypes</u>	<u>TaqI Intensity</u>		<u>Gene Dosage</u>	
		3.2 kb	3.7 kb	21-OHase A	21-OHase B
BB	A3,23 B39,44 DR2,7	+	-	2	-
Ice5	A2 B7 DR2	-	+	-	2
CBu	A2,11 Bw47,55 C6,3 DR7,4	+	$\frac{1}{2}$	2	1
II.I	A2,3 B44,w47 C5,6 DR6,7	+	$\frac{1}{2}$	2	1
KEM	A1,2 B7,8 DR3,4	-	+	-	2



BB are cut down to 420 bp and 274 bp respectively. (Lane 6, Figure 5.13). Thus both 21-OHase genes in BB, characterized by a 3.2 kb TaqI fragment have the NcoI RFLP. The single 475 bp fragment in Ice5 was not cut by NcoI (Lane 7) indicating the absence of the NcoI RFLP in both the 21-OHase B genes in Ice5. The 328 bp fragment in CBU and II.I was cut down to 274 bp (Lane 8 and 9) confirming the presence of the NcoI RFLP in the 21-OHase A gene in the two patients. The 520 bp fragment in CBU was cut down to 475 bp while the 475 bp fragment remained uncut, the resultant fragment therefore being thrice as intense as the original 475 bp fragment (Lane 3 and 8). The NcoI RFLP was therefore absent in the 475 bp fragment corresponding to the 21-OHase B gene but present in a the 520 bp fragment. The two fragments of 520 and 475 bp in Individual II.I (Lane 4) in the AvaII digest were seen at 475 bp and 420 bp in the AvaII / NcoI digest (Lane 9, Figure). Since the NcoI RFLP is present in all the 21-OHase genes in individual II.I (Chapter IV), two of which are 21-OHase A genes, the 520 bp fragment in Individual II.I corresponds to a 21-OHase A gene. It is either a partial digest or a polymorphism, but the study done does not provide an answer to that. The bands in KEM remained unchanged (Lane 5 and 10). The 520 bp fragment in KEM who has no 21-OHase A genes very likely corresponds to a 21-OHase B gene.

5.5 Discussion

The NcoI polymorphism in individual II.I due to single nucleotide change is not unique to II.I. This was predicted from the fact that all his genes had the NcoI RFLP and that he had inherited it from both his parents. NcoI was used to study other patients with CAH. Eight out of the 20 other patients exhibited the RFLP, but since the RFLP could be present on any one of the four 21-OHase genes, the frequency of the NcoI RFLP in the patients would be 8 out of 80 genes, that is 10%. The Bw13,

DR7 haplotype of patient JB also had the RFLP. The HLA-Bw47 (w4) antigen is serologically very similar to the HLA B13 (w4) antigen and the Bw47 haplotype has been implied to have originated from the B13 haplotype (White et al., 1984b). Although the HLA B13, DR7 haplotype has both 21-OHase loci present in most individuals, in patient JB it has only one 21-OHase A gene present. CAH is not associated with the B13 haplotype as it is with the Bw47 haplotype. This result suggests that the NcoI polymorphism must have originated in the 21-OHase A gene and subsequently passed on from the pseudogene to some 21-OHase B genes by gene conversion-like events. It is possible however that the mutation might arise in a 21-OHase B gene in the same way as it did in the 21-OHase A gene. The A29, B44, DR7 haplotype also showed consistency in inheriting the NcoI RFLP, but it was not possible to identify the gene on which it was present. It was clear from the study that in most cases the NcoI RFLP was present in the pseudogene.

In patient CBu for example, the inheritance pattern suggests that he should have had the NcoI RFLP in only one of his three genes, two of which have been inherited from his mother who does not show the NcoI RFLP. Yet his NcoI pattern suggests that the RFLP is in two of his three genes. This indicates that one of his genes has acquired the NcoI RFLP and not inherited it. The AvaII / NcoI digest should be able to identify which gene has the NcoI RFLP. The interpretation of the result is complicated by the extra fragment of 520 bp which is cut down by NcoI. Since the 328 bp fragment is cut, one of the NcoI RFLP is in his 21-OHase A gene. The 475 bp fragment remains uncut, suggesting that the 21-OHase B gene does not have the NcoI RFLP. The 520 bp fragment which is twice as intense as the 475 bp fragment and is cut by NcoI should therefore represent the 21-OHase A genes. Figure 5.11 illustrates the AvaII sites around the 475 bp fragment. There is an AvaII

site at position 1233. In the 21-OHase A gene there is an additional *Ava*II site at 1380 which cuts the 475 bp fragment into 147 bp and 328 bp fragments. If both the *Ava*II sites at 1233 and 1380 were not cut for some reason by *Ava*II, the fragment which would result would be 520 bp in size. Therefore the possibilities that arise are that *Ava*II did not cut at those sites due to methylation at C residues, or that the enzyme reaction was incomplete or that those *Ava*II sites did not exist and it was infact a polymorphism. Individual II.I also shows a fragment at 520 bp along with one at 475 bp and 328 bp. He has two 21-OHase A genes and one 21-OHase B gene. In the *Ava*II / *Nco*I digest he shows a pattern of 475, 420 and 274 bp. Since it has been conclusively shown in chapter IV that all his three genes have the *Nco*I RFLP, it indicates that the 520, 475 and 328 bp fragments have been cut down to 475, 420 and 275 bp fragments. The 520 bp fragment corresponds to a partially digested 21-OHase A gene in individual II.I.

The digest of BB shows that she has two 21-OHase genes. The *Taq*I digest had indicated that the 3.7 kb *Taq*I fragment was absent in BB but the *Ava*II digest gives two fragments of 475 and 328 bp. This suggests that one of the 21-OHase genes is a 21-OHase A gene while the other is a hybrid of a 21-OHase A and 21-OHase B gene. In the *Taq*I digest, it gives a 3.2 kb fragment due to the 5' region being 21-OHase A-like. As the 3' region is 21-OHase B-like, the *Ava*II fragment is 475 bp in length. BB does not have a 520 bp fragment. Both fragments in BB were cut down by *Nco*I to give the expected 420 bp and 274 bp fragments. Thus the *Ava*II / *Nco*I digest was able to distinguish 21-OHase A and 21-OHase B genes in BB, but did not identify to which haplotype it belonged to. If the DNA and HLA haplotype data of the family had been available, this could have been resolved. Ice5 had a complete digestion leading to a single fragment of 475 bp which was not cut by *Nco*I. KEM however,

showed two fragments of 520 bp and 475 bp with *Ava*II, the fragment of 520 bp being half as intense as the one at 475 bp. Both fragments were unaffected by *Nco*I. KEM has no 3.2 kb *Taq*I fragment and has two 21-OHase B genes. This suggested that the 520 bp fragment was a partial digest of a 21-OHase B gene. Since DNA samples were prepared independent of each other, it is possible that the *Ava*II digest was complete in some samples while in others it was incomplete. Family studies are required to analyse the inheritance of the *Ava*II / *Nco*I restriction pattern so that the full implication of these results can be understood and the genes can be mapped precisely. The *Ava*II polymorphism can thus be used to study haplotypes in greater detail and extends the study to detect limits of deletion of the 21-OHase genes. The *Nco*I RFLP is interesting because of the substantially high frequency at which it occurs in a random population and methods of detecting its presence in affected 21-OHase B genes would provide a useful tool for the detection of carriers.

In conclusion, it can be said that the study of the 21-OHase deficiency is not as straightforward as other recessive disorders. It is complicated by the fact that (a) there are different forms of 21-OHase deficiencies, (b) by the presence of two highly homologous genes one of which is a pseudogene and (c) because of the polymorphism exhibited by the genes.

The use of RFLPs to study this disorder should prove immensely valuable because their inheritance in a family can be followed quite precisely. Matteson et al. (1987) have reported in a restriction polymorphism analysis of patients with CAH, that the lack of the 3.7 kb *Taq*I fragments does not indicate 21-OHase B gene deletion, as other restriction sites characteristic to the 21-OHase B gene are present.

Results obtained from the restriction analysis of individual BB in the present study support the same hypothesis. Thus more than one RFLP will have to be used in the study of patients to identify the genes on the affected haplotypes.

CHAPTER VI

GENERAL DISCUSSION AND SUMMARY

General Discussion and Summary

The mapping of the two 21-OHase genes to the MHC along with the complement genes Bf, C2 and C4A and C4B in man and mice has led to several speculations. The organization of the class III region is very similar in both species. It has been suggested that a single duplication event has resulted in the tandem orientation of both the complement C4 / 21-OHase gene units in the two species. A comparison of the amino acid sequence indicates that the interspecies homology (human 21-OHase B / murine 21-OHase A, 69%) is significantly less than the intraspecies homology (human 21-OHase A / 21-OHase B 98%, murine 21-OHase A / 21-OHase B 86%), calculated at the nucleotide level. Similarly the interspecies homology of C4A calculated by Nonaka et al. (1985a,b) is (C4A / C4 79%, C4A / S1p 78%), which is significantly less than the intraspecies homology (C4A / C4B 99.7%, C4 / S1p 96%). This suggests that while the ancestral 21-OHase gene became linked to the ancestral C4 gene prior to speciation, the duplication event could have occurred after the divergence of man and mouse. Also, the 21-OHase B gene is active in man while the 21-OHase A gene is active in mice and the nature of mutations observed in the pseudogenes are different in both species. Both C4 genes are functional in man, while in mouse only the C4 gene has C4 functional activity while S1p has no known function. It is surprising therefore that the inactive 21-OHase genes and S1p have not diverged significantly from their active counterpart due to lack of selective pressure. Alternatively, the genes may have duplicated before the divergence of man and mouse. The high degree of intraspecies homology would be maintained by concerted evolution operating in the form of gene conversion and / or unequal crossing over as with the class I and class II gene loci.

The biological significance of the presence of the complement genes in the MHC is not very clear, but it has been suggested that the different polymorphic forms of C4 could be important in individual susceptibility to autoimmune diseases due to varying efficiencies of complement activation by C4 (Porter, 1983). The presence of C4 in the MHC may therefore be due to some selective advantage conferred on the species in evolution by this association, for example in disease resistance. The close proximity of 21-OHase suggests that the gene was carried with the piece of C4- containing DNA when it was being moved to the MHC, and the association was not disadvantageous (Bodmer, 1984).

The 21-OHase genes belong to the cytochrome P-450 supergene family. The enzyme it encodes is of 56,028 Daltons and it plays a crucial role in the synthesis of cortisol from cholesterol in the human adrenal cortex. Sequencing studies have revealed that the two 21-OHase genes are 3.3 kb in length and are split into 10 exons. They are highly conserved with an overall sequence homology of 98%. There are 92 nucleotide differences between the two genes, 27 of which are in exons. Of these eight are silent mutations while 19 cause codon changes. Nine of the codon changes have an unpredictable effect on the gene product while the remaining differences are quite drastic. An 8 bp deletion in the third exon of the 21-OHase A gene, a 1 bp insertion in the seventh exon and a third nonsense mutation in the eighth exon change the reading frame and introduce premature termination codons in the 21-OHase A gene. These indicate that the 21-OHase A gene would encode a truncated and presumably non functional protein. Absence of detectable 21-OHase A mRNA in an adrenal RNA preparation using oligonucleotides specific for the A and B gene sequences have been shown by Higashi *et al.* (1986). In addition, homozygous deletion of the 21-OHase A gene in individuals does not affect normal cortisol synthesis while individuals with a homozygous

deletion of the 21-OHase B genes have classical CAH (White et al., 1984b). It can thus be concluded that the 21-OHase A gene is a pseudo-gene due to the deleterious mutations which prevent synthesis of a functional protein.

Deletion and homoduplication of the 21-OHase genes and the adjacent complement C4 genes have been reported in man (Garlepp et al., 1986; Carroll et al., 1985b; Schneider et al., 1986). This suggests that the class III region is not genetically stable, but is still undergoing gene expansion and contraction. Both duplication and deletion of the class III genes have been explained by a mechanism of unequal crossing over. However no direct evidence of a recombination event has been found upon comparison of variant and a normal MHC haplotype (Carroll et al., 1985).

The presence of the duplicated and highly homologous sets of the C4 and 21-OHase genes in the relatively small stretch of the chromosomal DNA may allow for frequent exchange of their DNA sequences by homologous recombination or unequal crossing over during meiosis. It is interesting to note that 45 of 92 nucleotides differences between the two 21-OHase genes are localized in the second intron and the third exon. This is reminiscent of a recombination point suggested for the human fetal γ -globin gene (Slightom et al., 1980). The presence of recombinational hot spots in the mouse MHC has already been described (Steinmetz et al., 1986).

The 21-OHase genes lie between a region of the HLA which is highly polymorphic. The neighbouring C4 genes have 40 or more alleles that have been demonstrated by differences in electrophoretic mobilities, serological typing, and direct DNA sequencing. It is not surprising therefore that the 21-OHase genes exhibit polymorphism. A comparison of the published nucleotide sequences of 21-OHase genes (Higashi et al., 1986; White et al., 1986; Rodrigues et al., 1987) showed 77 differences.

Most of the differences may represent accumulation of point mutations in the pseudogene. Six differences caused a codon change. Three of these (CTG), which code for a leucine at codon 6 were absent in the 21-OHase B gene sequences of the other groups, but present in their 21-OHase A gene sequences. However, the codon for leucine was present in both the 21-OHase B genes and the 21-OHase A gene sequences presented here. Thus there is a strong indication that there are two allelic variants of the 21-OHase gene product, of 494 or 495 amino acids. Since the codon for leucine is present in all the three reported 21-OHase A gene sequences it is possible that it was previously ubiquitously present in all 21-OHase genes before being deleted in some of the 21-OHase B gene sequences with no obvious ill effect on the gene product.

The presence of allelic variants of 21-OHase at a single locus is supported by the four different clinical manifestations of CAH. The differential linkage disequilibrium of classical and nonclassical CAH to different haplotypes seems to indicate that the less severe form of CAH is an allelic variant of the classical form. Although homozygous deletion of the 21-OHase B gene always causes CAH, only one in 20 patients had a homozygous deletion of the 21-OHase B gene (Rumsby et al., 1986). Of the 20 patients selected at random in the present study, two had a homozygous deletion of the 21-OHase B gene, twelve were heterozygous for the 21-OHase B gene deletion, while the remaining patients showed the presence of two 21-OHase B genes (that is 16 out of 40 alleles were gene deletions). Similar reports have also been published by Werkmeister et al. (1986) and Matteson et al. (1987), which support the theory that in most cases of CAH, the defect is due to point mutations within the gene. To characterize one such defective gene from a patient with CAH, individual II.I was chosen. Since he had a single defective 21-OHase B gene, cloning and characterization of that gene

would give some understanding as to the nature of the mutations involved. Sequence analysis of the gene revealed 13 nucleotide alterations. Only one of the alterations causes a codon change at codon 269 and results in a serine being replaced by a threonine. Single amino acid changes have been known to drastically alter the function of a protein. Sick cell anaemia for example, results from a mutation that changes a glutamic acid residue (GAG) to a valine residue (GTG) at position 6 in the β -globin chain of hemoglobin.

Since none of the nucleotide alterations in the 21-OHase B gene from the patient have any evident effect on the expression of the gene, the gene is probably expressed, but the protein product is nonfunctional. Expression studies indicated the presence of a message in Y1 cells transfected with the defective gene. The functional activity of the gene product can be studied by doing bioassays of 21-OHase activity from long term transfected Y1 cells. Further studies can also be done in the expression system by doing invitro mutagenesis. The normal 21-OHase B gene sequence can be substituted by the defective gene sequences and the effect on the expression of the gene and gene product can be studied. In their work with the murine 21-OHase genes, Chaplin et al. (1986) have shown that the promoter of the inactive 21-OHase B gene in mice is active by placing it 5' to the 21-OHase A structural gene, transfecting the hybrid gene into Y1 cells, and showing the presence of transcripts in Northern blot analysis.

Restriction fragment length polymorphisms that occur due to point mutations provide a valuable tool for diagnostic methods. Prenatal diagnosis of both sick cell anaemia and α_1 -antitrypsin deficiency has been made possible by means of RFLP associations with the disease alleles (Philips III et al., 1980; Hejtmancik et al., 1986). Mornet et al. (1986) have identified parental chromosomes carrying alleles for

21-OHase deficiency in prenatal diagnosis using probes corresponding to the closely-linked HLA class I and class II genes which flank the 21-OHase genes. Since these genes are highly polymorphic, the parental chromosomes carrying the alleles for 21-OHase deficiency can be readily identified by examination of the hybridization pattern in the proband. The single codon change in the patient's 21-OHase B gene also gave rise to a RFLP with NcoI. On examination of the patient's DNA, the NcoI RFLP was also found to be in his 21-OHase A genes. This suggested that the NcoI RFLP originated in a 21-OHase A gene and was transferred to some 21-OHase B genes by gene conversion like events. When other patients were tested for the NcoI polymorphism, it was found to be present in eight of them. As it could be contained in any of the four 21-OHase genes normally present, it was essential to identify which gene contained the NcoI RFLP. Using AvaII in conjunction with NcoI it is possible to identify the gene in which the RFLP is present. This result also showed that absence of the 3.7 kb TaqI fragment did not indicate a gene deletion in all cases. One of the individuals who did not have the 3.7 kb TaqI fragment had the AvaII / NcoI fragment of 420 bp corresponding to the 21-OHase B gene. This suggested that the 21-OHase B gene was a hybrid gene due to a recombination event or a gene conversion-like event. Gene conversion is a well established phenomenon in fungi and yeast (Petes and Fink, 1982; Hamza et al., 1986), but appears not to be restricted to these organisms. Gene conversion-like events have been shown to take place in the class I and class II genes in the MHC (Steinmetz and Hood, 1983; Gorski and Mach, 1986).

The NcoI RFLP was also found in a substantially high frequency in a random selection of normal individuals. This was not surprising as it is found in both 21-OHase genes and in more 21-OHase A genes than 21-OHase B genes. This supports the theory that it is a mutation in the

pseudogene which when carried into a 21-OHase B gene may cause a defect in the gene product. In one patient CBu, the NcoI RFLP is on two of his genes. One of the 21-OHase A genes with the NcoI RFLP has been inherited from the father. The other 21-OHase A gene and the single 21-OHase B gene inherited from his mother should not have the NcoI RFLP since his mother does not have it. Yet two of his three genes show the NcoI RFLP. Therefore either the second gene which appears to be a 21-OHase A gene from the AvaII / NcoI pattern has acquired the NcoI RFLP due to a spontaneous mutation or has acquired it during mitotic cell division.

Not all point mutations give rise to RFLPs. Synthetic oligodeoxynucleotides have provided the next technologic refinement for diagnostic accuracy. By means of allele specific oligonucleotides that matched the nucleotides flanking and including the point mutations and using appropriate hybridization conditions that permitted stability of only perfectly matched oligonucleotide / DNA duplexes, it is possible to identify the DNA containing that point mutation (Conner et al., 1983). Selective amplification of the gene sequences in vitro with the polymerase chain reaction using two priming oligodeoxynucleotides complementary to opposite strands that flank the mutation, eliminates the need of having to clone the gene to study it in detail (Saiki et al., 1986). These methods have been used for detection of the 23 different mutations in β thalassemia (Weatherall, 1985) and the point mutation in the 5' splice donor site of intron 12 of the phenylalanine hydroxylase gene which accounts for 30% of the cases of phenylketonuria (PKU) (DiLella et al., 1986). More recently a second most common PKU allele which accounts for 20% of the PKU alleles and results in an amino acid substitution was identified. Using a pair of synthetic oligonucleotides complementary to the normal and mutant sequences, it was shown that the mutation was in linkage disequilibrium with haplotype-2 (DiLella et al.,

1987).

An alternative approach to the detection of single base substitutions has been the use of Ribonuclease cleavage at mismatches in RNA:DNA duplexes using a labelled RNA probe synthesized in vitro (Myers et al., 1985). This method was used to detect single base mutations in individuals with β -thalassemia.

In 21-OHase deficiency, prompt accurate prenatal diagnosis would improve prenatal therapy which at present is only possible at 16 weeks of fetal age, at which time masculinization of the genitalia in affected females has already occurred. Chorionic villus sampling can provide material for analysis before the end of 10 weeks of gestation. DNA can be extracted from the sample for diagnostic purposes. It is essential therefore to have the right probes and assays to identify the defects in the genes.

The characterization of both 21-OHase genes and polymorphism studies provide the elementary tools to study defective genes. It has been used as a model system to detect point mutations using chemical modifications of mismatched bases (Cotton et al., submitted). One patient who has been characterized in detail gives a guideline as to which of the residues may be important for functional activity. The same mutation was also found in a subset of 21-OHase A and 21-OHase B genes. Thus 21-OHase deficiency in patients is due to heterogenous defects, and analysis of more defective genes will provide some understanding as to whether the defects can be categorised like other well characterized disorders. Identification of defective genes is of prime importance for successful therapy in affected individuals.

REFERENCES

- Adesnik, M. and Atchison, M. (1986) *CRC. Crit. Rev. Biochem.* 19,
247 - 305
- Agnello, V. (1978) *Medicine* 57, 1 - 23
- Allen, F.H. (1974) *Vox Sang* 27, 382 - 384
- Alper, C.A., Boenisch, T. & Watson, L. (1972) *J. Exp. Med.* 135, 68 - 80
- Alper, C.A., (1976) *J. Exp. Med.* 144, 1111 - 1115
- Anson, D.S. Choo, K.H. Rees, D.J.G., Giannelli, F., Gould, K.,
Huddleston, J.A. & Brownlee, G.G. (1984) *EMBO J.* 3, 1053 - 1060
- Appleyard, R.K. (1954) *Genetics* 39, 440
- Atchison, M. & Adesnik, M. (1983) *Biol. Chem.* 258, 11285 - 11295
- Awdeh, Z.L. & Alper, C.A. (1980) *Proc. Natl. Acad. Sci. USA* 77,
3576 - 3580
- Awdeh, Z.L., Raum, D., Yunis, E.J. & Alper, C.A. (1983) *Proc. Natl.*
Acad. Sci. USA 80, 259 - 263
- Bach, F.H. (1985) *Immunol. Today* 6, 89 - 94
- Barker, D., Schafer, M. & White, R. (1984) *Cell* 36, 131 - 138
- Bell, G.I., Karam, J.H. & Rutter, W.J. (1981) *Proc. Natl. Acad. Sci.*
USA 78, 5759 - 5763
- Belt, K.T., Carroll, M.C. & Porter R.R. (1984) *Cell*, 36, 907 - 914
- Belt, K.T., Yu, C.Y., Carroll, M.C. & Porter, R.R. (1985)
Immunogenetics 21, 173 - 180
- Bentley, D.R., & Porter, R.R. (1984) *Proc. Natl. Acad. Sci. USA* 81,
1212 - 1215
- Bentley, D.R., Campbell, R.D., Cross, S.J. & Edwards, J.E. (1985)
Immunogenetics 22, 377 - 390
- Bentley, D.R. (1986) *Biochem. J.* 239, 339 - 345
- Bernards, R. & Flavell, R.A. (1980) *Nucleic Acid Res.* 8, 1521 - 1534
- Biggin, M.D., Gibson, T.J. & Hong, G.F. (1983) *Proc. Natl. Acad. Sci.*
USA 83, 3963 - 3965

- Birnboim, H.C. & Doly, J. (1979) *Nucleic Acids Res.* 7, 1513 - 1523
- Bodmer, W.F., & Bodmer, J.G. (1978) *Brit. Med. Bull.* 34, 309 - 316
- Bodmer, W.F. (1984) In 'Histocompatibility Testing' ed. Albert, E .D.
et al., pp 11 - 22
- Bongiovanni, A.M., Eberlain, W.R. & Cara, J. (1954) *J. Clin. Endocrinol.*
Metab. 14, 409
- Butler. G.C., & Marrian G.F., (1937) *J. Biol. Chem.* 119, 565
- Campbell, R.D., Bentley, D.R. & Morley, B.J. (1984) *Phil. Trans. R.*
Soc. Lond. B. 306, 367 - 378
- Campbell, R.D., (1987) *Brit. Med. Bulletin* 43, 37 - 49
- Carroll, M.C., Campbell, R.D., Bentley, D.R., Porter, R.R. (1984a)
Nature 307, 237 - 241
- Carroll, M.C., Belt, K.T., Palsdottir, A. & Porter, R.R. (1984b) *Phil.*
Trans. R. Soc. Lond. B. 306, 379 - 388
- Carroll, M.C., Campbell, R.D., & Porter, R.R. (1985a) *Proc. Natl. Acad.*
Sci. USA 82, 521 - 525
- Carroll, M.C., Palsdottir, A., Belt, K.T. & Porter, R.R. (1985b) *EMBO*
J., 4, 2547 - 2552
- Carroll, M.C., Belt, K.T., Palsdottir, A. & Yu, C.Y. (1985c) *Imm. Rev.*
87, 39 - 60
- Casadaban, M.J. & Cohen, S.N. (1980) *J. Mol. Biol.* 138, 179 - 207
- Caskey, C.T. (1987) *Science* 236, 1223 - 1229
- Chan, A.C. & Atkinson, J.P. (1983) *J. Clin. Invest.* 72, 1639 - 1649
- Chaplin, D.D., Woods, D.E., Whitehead, A.S., Goldberger, G., Colten,
H.R. & Seidman, J.G. (1983) *Proc. Natl. Acad. Sci. USA* 80,
6947 - 6951
- Chaplin, D.D., Sackstein, R., Perlmutter, D.H., Weis, J.H.,
Kruse, T.A., Coligan, J., Colten, H.R., & Seidman, J.G. (1984)
Cell 37, 569 - 576

- Chaplin, D.D., Parker, K., White, P. & Seidman, J. (1985) Fed. Proc. 44, 783
- Chaplin, D.D., Galbraith, L.J., Seidman, J.G., White, P.C. & Parker, K.L., (1986) Proc. Natl. Acad. Sci. USA 83, 9601 - 9605
- Childs, B., Grumbach M.M. & Van Wyk, J.J. (1956) J. Clin. Invest. 35, 213 - 222
- Chirgwin, J.M., Przybyla, A.E., Macdonald, R.J. & Rutter W.J. (1979) Biochemistry 18, 5294 - 5299
- Christie, D.L. & Gagnon, J. (1983) Biochem. J. 209, 61 - 70
- Chung, B., Matteson, K.J. & Miller, W.L. (1985) DNA 4, 211 - 220
- Chung, B., Matteson, K.J. & Miller, W.L. (1986) Proc. Natl. Acad. Sci. USA 83, 4243 - 4247
- Clewell, D.B., & Helsinki, D.R. (1972) J. Bact. 100, 1135 - 1146
- Cohen, J.M. (1969) Pediatrics 44, 621 - 622
- Cole, F.S., Whitehead, A.S., Auerbech, H.S., Lint, T., Zeity, H.J., Kilbridge, P., Colten, H.R., (1985) New Engl. J. Med. 313, 11 - 16
- Conner, B.J. Reyes, A.A., Morin, C., Itakura, K., Teplitz, R.L. & Wallace, R.B. (1983) Proc. Natl. Acad. Sci. USA 80, 278 - 282
- Coon, M.J & Koop, D.R. (1983) The Enzymes Vol. XVI. P.D. Boyer Ed. 645 - 677 Academic Press New York
- Cooper, D.Y., Levin, S., Narasimhulu, S., Rosenthal, O. & Estabrook, R.W. (1965) Science 147, 400 - 402
- Dawkins, R.L., Christiansen, F.T., Kay, P.H., Garlepp, M., MuCluskey, J., Hollingsworth, P.N. & Zilko, P.J. (1983) Immunol. Rev. 70, 1 - 22
- Dayhoff, M.O., Barker, W.C., Hunt, L.T. & Schwartz, R.M. (1978) In Atlas of Protein Sequence and Structure ed. Dayhoff, M.O., (Nat. Biomed. Research Foundation, Washington DC) 9 - 10
- De Crecchio, L. (1865) Morgagni 7, 154 - 188

- Decourt, J., Jayle, M.F. & Baulieu, E. (1957) *Ann. Endocrinol.* 18, 416 - 422
- Denhardt, D.T. (1966) *Biochem. Biophys. Res. Commun.* 23, 641 - 644
- DiLella, A.G., Marvit, J., Lidsky, A. S., Guttler, F. & Woo, S.L.C. (1986) *Nature* 322, 799 - 803
- DiLella, A.G., Marvit, J., Brayton, K. & Woo, S.L.C. (1987) *Nature* 327, 333 - 336
- Donohue, P.A., Van, Dop, C., McLean, R.H., White, P.C., Jospe, N. & Migeon, C.J. (1986) *J. Clin. Endocrinol. Metab.* 62, 995 - 1002
- Dozy, A.M., Kan, Y.W., Embury, S.H., Mentzer, W.C., Wang, W.C., Lubin, B., Davis, J.R. & Koenig, H.M. (1979) *Nature* 280, 605 - 607
- Duckworth, M.L., Gait, M.J., Goelet, P., Hong, G.F., Singh, M. & Titmas, R.C. (1981) *Nucleic. Acid Res.* 9, 1691 - 1706
- Dunham, I., Sargent, C.A., Trowsdale, J. and Campbell, R.D. (1987) *Proc. Natl. Acad. Sci. USA* In Press
- Dupont, B., Oberfield, S.E. Smithwick, E.M., Lee, T.D. & Levine, L.S. (1977) *Lancet* 2, 1309 - 1311
- Dupont, B., Pollack, M.S., Levine, L.S., O'Neill, G.J., Hawkins, B.R. & New, M.I. (1980) *Histocompatibility Testing 1980*, ed. Terasaki, P.I. 693 - 706 UCLA Press
- Dupont, B., Virdis, R., Lerner, A.J., Pollack, M.S., Nelson, C. & New, M.I. (1984) *Quebec Excerpta Med. Int. Congr. Sec.* 652, 543 - 566
- Dykes, D.P., Defurio, M. & Polesky, H.F. (1983) *Am. J. Hum. Genet.* 35, 652 - 655
- Estabrook, R.W., Cooper, D.Y., & Rosenthal, O. (1963) *Biochem. Z.* 338, 741 - 755
- Federman, D.D. (1968) Philadelphia, W.B. Saunders
- Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.* 132, 6 - 13
- Feinberg, A.P. and Vogelstein, B. (1984) *Anal. Biochem.* 137, 266 - 267

- Ferreira, A., Nussenzweig, V. & Gigli, I. (1978) *J. Exp. Med.* 148, 1186 - 1197
- Fielder, A.H.L., Walport, M.J., Batchelor, J.R., Rynes, R.I., Black, C.M., Dodi, I.A. & Hughes, G.R.V. (1983) *Brit. Med. J.* 286, 425 - 428
- Finkelstein, M. & Shaefer, J.M. (1979) *Physiol. Rev.* 59, 353 - 406
- Fleischnick, E., Awdeh, Z.L., Raum, D., Granados, J., Alosco, S.M., Crigler, J.F., Gerald, P.S., Giles, C.M., Yunis, E.J. & Alper, C.A. (1983) *Lancet* 1, 152 - 156
- Frischauf, A.M., Lehrach, H., Poustka, A. & Murray, N. (1983) *J. Mol. Biol.* 170, 827 - 842
- Fu, S.M., Kunkel, H.G., Brusman, H.P., Allen, F.H. & Fotino, M. (1974) *J. Exp. Med.* 140, 1108 - 1111
- Fujii-Kuriyama, Y., Mizukami, Y., Kawajiri, K., Sogawa, K. & Muramatsu, M. (1982) *Proc. Natl. Acad. Sci. USA* 79, 2793 - 2797
- Gagnon, J. (1984) *Phil. Trans. R. Soc. Lond. B.* 306, 301 - 309
- Ganong, W.F., Biglieri, E.G., Mulrow, P.J. (1966) *Rec. Prog. Horm. Res.* 22 - 381 - 430
- Ganong, W.F., Alpert, L.C., Lee, T.C. (1974) *N. Engl. J. Med.* 290, 1006
- Garlepp, M.J., Wilton, A.N., Dawkins, R.L. & White, P.C. (1986) *Immunogenetics* 23, 100 - 105
- Garren, L.D., Gill, G.N., Masui, H. & Walton, G.M. (1971) *Rec. Prog. Horm. Res.* 27, 433 - 478
- Gergen, J.P., Stern, R.H. & Wensink, P.C. (1979) *Nucleic Acid Res.* 7, 2115 - 2136
- Gigli, I. (1978) *Nature* 272, 836 - 837
- Giles, C.M., Gedde - Bahl, T., Robson, E.B., Thorsby, E., Olaisen, B., Arnason, A., Kissmeyer - Neilsen, F., & Schreuder, I. (1976) *Tissue Antigens* 8, 143 - 149

- Giles, C.M., Batchelor, J.R., Dodi, I.A., Feilder, A.H.L., Rither, C.,
Mauff, G., Bender, K., Levene, C., Schreuder, G.M.T. and Well, L.J.
(1984) *J. Immunogenet.* 11, 305 - 317
- Giles, C.M. (1987) *Human Immunol.*, In Press
- Gonzalez, F.J., Kimura, S. and Nebert, D.W. (1985) *J. Biolchem.* 260,
5040 - 5049
- Gorer, P.A. (1937) *J. Pathol. Bacteriol.* 44, 691 - 697
- Gorman, C.M., Moffat, L.F. & Howard, B.H. (1982) *Mol. Cell. Biol.*
2, 1044 - 1051
- Gorski, J. & Mach, B. (1986) *Nature* 322, 67
- Grosveld, F.G., Lund, T., Murray, E.J., Mellor, A.L., Dahl, H. &
Flavell, R.A. (1982) *Nucleic Acid Res.* 10, 6715 - 6731
- Grumbach, M.M., Van Wyk, J.J. (1974) *Textbook of Endocrinology*, 5th. ed.
Philadelphia, W.B. Saunders, (1974) 423
- Hall, P.F. (1985) *Congenital Adrenal Hyperplasia in Annals of New York
Academy of Sciences ed. New, M.I.* 458, 203 - 215
- Hall, R.E. & Colten, H.R. (1977) *Proc. Natl. Acad. Sci. USA* 74,
1707 - 1710
- Hamza, H., Kalogeropoulos, A., Nicolas, A. & Rossignol, J.L. (1986)
Proc. Natl. Acad. Sci. USA 83, 7386 - 7390
- Hann, J.E., Hodgson, K.O., Anderson, L.A. & Dawson, J.H. (1982) *J. Biol.
Chem.* 257, 10934 - 10941
- Hardy, D.A., Bell, J.I., Long, E.O., Lindsten, T. & McDevitt, H.O.
(1986) *Nature* 323, 453 - 455
- Harkins, J.B., Nelson, E.B., Masters, B.S.S., Bryan, G.T. (1974)
Endocrinology 94, 897 - 902
- Hejtmancik, J.F., Sifers, R.N., Ward, P.A., Harris, S., Mansfield, T. &
Cox, D.W., (1986) *Lancet* 2, 767 - 770
- Hellman, U., Eggertsen, G., Lundwall, A., Engstrom, I. & Sjoquist, J.

- (1984) FEBS Letts. 170, 254 - 258
- Helms, C., Graham, M.Y., Dutchik, J.E. & Olson M.V. (1985)
DNA 4, 39 - 49
- Herrmann, R., Neugebauer, K., Prikl, E., Zentgraf, H. & Schaller, H.
(1980) Mol. Genet. 177, 231 - 242
- Higashi, Y., Yoshioka, H., Gotoh, O. & Fujii-Kuriyama, Y., (1986)
Proc. Natl. Acad. Sci. USA 83, 2841 - 2845
- Hines, R.N., Levy, J.B., Shen, M.L., Renli, A.M., & Bresnik, E. (1985)
Arch. Biochem. Biophys. 237, 465 - 476
- Holn, B., & Collins, J. (1980) Gene 11, 291 - 298
- Hood, L., Steinmetz, M. & Malissen, B. (1983) Ann. Rev. Immunol. 1,
529 - 568
- Ingram, V. (1957) Nature 180, 326 - 328
- Isenman, D.E. & Young, J.R. (1984) J. Immunol. 132, 3019 - 3027
- IUPAC - IUB Commission on Biochemical Nomenclature (1969) Biochem.
J. 113, 1 - 4
- Jaiswal, A.K., Gonzalez, F. J. & Nebert, D.W. (1985) Science 228,
80 - 83
- Johnson, P.H. & Grossman, L.I. (1977) Biochemistry 16, 4217 - 4224
- Josso, N., Picard, J.Y. & Tran, D. (1977) Recent Prog. Horm.Res. 33,
117 - 168
- Jost, A. Embryonic Sexual Differentiation. In Jones, H.W. and Scott,
W.W. (eds) 2nd. ed. Baltimore, Williams & Wilkins, 1971, 17 - 64
- Karn, J., Brenner, S., Barnett, L. & Cesareni, G. (1980) Proc. Natl.
Acad. Sci. USA 77, 5172 - 5176
- Karp, D.R., Capra, D., Atkinson, J.P. & Shreffler, D.C. (1982)
J. Immunol. 128, 2336 - 2341
- Kaufman, J.F., Auffray, C., Korman, A.J., Shackelford, D.A. &
Strominger, J. (1984) Cell 36, 1 - 13
- Kozak, M. (1984) Nuc. Acids Res. 12 3873 - 3893

- Kawajiri, K., Gotah, O., Sogawa, K., Tagashira, T., Muramatsu, M., & Fujii-Kuriyama, Y. (1984) *Proc. Natl. Acad. Sci. USA* 81, 1649 - 1653
- Kerr, M.A. & Gagnon, J. (1982) *Biochem. J.* 205, 59 - 67
- Kioussis, D., Eiferman, F., Ver de Rijn, P., Gorin, M.B., Ingram, R.S. & Tilghman, S.M., (1981) *J. Biol. Chem.* 256, 1960 - 1967
- Klein, J., Juretic, A., Baxevasis, C.N. & Nagy, Z.A. (1981) *Nature* 291, 455 - 460
- Kohn, B., Levine, L.S., Pollack, M.S., Pang, S., Lorenzen, F., Levy, D., Lerner, A., Rondanini, G.F., Dupont, B. & New, M.I. (1982) *J. Clin. Endocrinol. Metab.* 55, 817 - 827
- Kominami, S., Ochi, H., Kobayashi, Y. & Takemori, S. (1980) *J. Biol. Chem.* 255, 3386 - 3394
- Koritz, S.B. (1964) *Biochemistry* 3, 1098 - 1102
- Kremers, P., (1976) *J. Steroid Biochem.* 7, 571 - 575
- Kuhnle, U., Chow, D., Rapaport, R., Pang, S., Levine, L.S. & New, M.I. (1981) *J. Clin. Endocrinol. Metab.* 52, 534 - 544
- Lamm, L.U. and Olaisen, B. (1985) *Cell Genet.* 40, 128 - 155
- Laragh, J.H., Sealey, J.E. (1973) *Handbook of Physiology*, Orloff, J., Berliner, R.W. (eds) *Renal Physiology*, Baltimore, Maryland, Waverly Press p831
- Larhammar, D., Schemming, L. Gustafsson, K. *Proc. Natl. Acad. Sci. USA* (1982a) 79, 3687 - 3691
- Larhammar, D., Schemming, L. Claesson, L. (1982b) *Cell* 30, 153 - 161
- Laron, Z., Pollack, M.S. Zamir, R., Roitman, A., Dickerman, Z., Levine, L.S., Lorenzen, F., O'Neill, G.J., Pang, S., New, M.I. and Dupont, B. (1980) *Hum. Immunol.* 1, 55 - 66
- Laskey, R.A. & Mills, A.D. (1977) *FEBS Letts.* 82, 314 - 316
- Law, S.K.A., Dodds, A.W. & Porter, R.R. (1984) *EMBO J.* 3, 1819 - 1823

- Levine, L.S., Zachmann, M., New, M.I., Prader, A., Pollack, M.S.,
O'Neill, G.J., Yang, S.Y., Oberfeild, S.E., Dupont, B. (1978)
N.Engl. J. Med. 299, 911 - 915
- Levine, L.S., Dupont, B., Lorenzen, F., Pang, S., Pollack, S.M,
Oberfield, S.E., Kohn, B., Lerner, A., Cacciari, E., Mantero, F.,
Cassio, A., Scaroni, C., Chiumello, G., Rondanini, G.F.,
Gargantini, L., Giovannelli, G., Virdis, R., Bartolotta, E.,
Migliori, C., Pintor, C., Tato, L., Barboni, F. & New, M.I. (1980)
J. Clin. Endocrinol. Metab. 51, 1316 - 1324
- Levi-Strauss, M., Tosi, M., Steinmetz, M., Klein, J. & Meo, T. (1985)
Proc. Natl. Acad. Sci. USA 82, 1746 - 1750
- Malissen, M., Malissen, B. & Jordan, B.R. (1982) Proc. Natl. Acad. Sci.
USA 79, 893 - 897
- Maniatis, T., Frisch, E.F. and Sambrook, J. (1982) A Laboratory Manual,
Cold Spring Harbor Laboratory Press, N.Y.
- Mason, P.A., Fraser, R., Semple, P.F. & Morton, J.J. (1979) J. Steroid
Biochem. 10, 235
- Matteson, K.J., Phillips III, J.A., Miller, W.L., Chung, B., Orlando,
P.J., Frisch, H., Ferrandez, A. & Burr, I.M (1987) Proc. Natl.
Acad. Sci. USA 84, 5858 - 5862
- Mauff, G., Hauptmann, G., Hitzeroth, H.W., Gauchel, F., & Scherz, R.
(1978) Z. Immun - Forsch. 154, 115 - 120
- Mauff, G., Alper, C.A., Awdeh, Z., Batchelor, J.R., Bertrams, T., Broon
- Peterson, G., Dawkins, R.L., Denant, P., Edwards, J., Gross -
Wilde, H., Hauptmann, G., Klouda, P., Lamm, L., Mollenhaver, E.,
Nerl, C., Olaisen, B., O'Neill, G., Rither, C., Ross, M.H., Skanes,
V., Teisberg, P. & Wells, L. (1983) Immunobiology, 164, 184 - 191
- Maxam, A.M. & Gilbert, W. (1980) Methods Enzymol. 65, 499 - 560
- Meo, T., Atkinson, J.P., M. Bernoco, D. & Ceppellini, R. (1977) Proc.

- Natl. Acad. Sci. USA 74, 1672 - 1675
- Messing, J., Crea, R. & Seeburg, P.H. (1981) Nucleic Acid Res. 9, 309 - 321
- Messing, J. & Vieira, J. (1982) Gene, 19, 269 - 276
- Messing, J. (1983) Methods in Enzymol. 101, 20 - 78
- Middleton, J., Croodston, M.C., Falk, J.A., Robson, E.B., Cook, P.L.J., Batchelor, J.R., Bodmer, J., Ferrara, G.B., Festenstein, H., Harris, R., Kissmeyer-Nielsen, F., Lawler, S.D., Sacks, J.A. & Wolf, E (1974) Tissue Antigens 4, 366 - 373
- Migeon, C.J. (1977) Adrenal Cortex. In Pediatrics, 16th ed. Rudolph, A.M. ed. (Appleton - Century - Crofts, N.Y.) 1619 - 1651
- Mizukami, Y., Sogawa, K., Suwa, Y., Muramatsu, M. & Fujii-Kuriyama, Y. (1983) Proc. Natl. Acad. Sci. USA 80, 3958 - 3962
- Morillo, E. & L.I. Gardner (1979) Lancet 2, 202 - 203
- Morley, B.J. & Campbell, R.D. (1984) EMBO J. 3, 153 - 157
- Mornet, E. Boué, J., Raux-Demay, M., Couillin, P., Oury, J.F., Dumez, Y., Dausset, J., Cohen, D. & Boüe (1986) Hum. Genet. 73, 358 - 364
- Morohashi, K., Fujii-Kuriyama, Y., Okada, Y., Sogawa, K., Hirose, T., Inayama, S. & Omura, T. (1984) Proc. Natl. Acad. Sci. USA 81, 4647 - 4651
- Muller, U., Jongeneel, C.V., Nedospasov, S.A., Lindahl, K.F. & Steinmetz, M. (1987) Nature (London) 325, 265 - 267
- Myers, R.M., Larin, Z. & Maniatis, T. (1985) Science 230, 1242 - 1246
- Nebert, D.W. & Gonzalez, F.J. (1987) Annu. Rev. Biochem. 56, 945 - 993
- Nebert, D.W., Adesnik, M., Coon, M.J., Estabrook, R.W., Gonzalez, F.J., Guengerich, F.P., Gunsalus, F.P., Johnson, E.F., Kemper, B., Levin, W., Phillips, I.R. & Waterman, M.R. (1987) DNA 6, 1 - 11
- New, M.I. & Peterson, R.E. (1966) Pediatr. Clin. North. Am. 13, 43 - 58
- Nordblom, G.D. & Coon, M.J. (1977) Arch. Biochem. Biophys. 180, 343 - 347

- New, M.I., & Seaman, M.P. (1970) J. Clin. Endocrinol. Metab. 30, 361
- New, M.I., F. Lorenzen, S. Pang, P. Gunczler, B. Dupont. & Levine (1979)
J. Clin. Endocrinol. Metab. 48, 356 - 359
- New, M.I. & Levine, L.S. (1973) Congenital adrenal hyperplasia. In
Advances in Human Genetics ed. Harris, H. & Hirschhorn. Plenum
Press London. 4, 251 - 326
- New, M.I., Dupont, B., Pang, S., Pollack, M.S. & Levine, L.S (1981)
Recent Prog. Horm. Res. 37, 105 - 181
- New. M.I., Dupont, B., Grumbach, K. & Levine, L.S. (1982)
Metabolic Basis of Inherited Disease, eds. Stanbury, J.B.,
Wyngarden, J.B., Fredrickson, D.S., Goldstein, J.L. & Brown, M.S.
(Mcgraw - Hill, New York) 973 - 1000
- New, M.I. Grumbach, K. & Levine, L.S.(1983) Principles & Practise of
Medical Genetics, 61, 36 - 38
- New, M.I. & Levine, L.S.(1984) Ann. Rev. Med. 35, 649 - 663
- New, M.I. (1985) Ann. of N.Y. Acad. of Sci. 458, 1 - 27
- Nonaka, M., Nakayama, K., Yeul, Y.D., Shimuzu, A. & Takahashi, M.
(1985a) Immunol. Rev. 87, 81 - 99
- Nonaka, M., Nakayama, K., Yeul, Y.D., & Takahashi, M. (1985b) J. Imm.
136, 2989 - 2993 Nordblom, G.D. & Coon, M.J. (1977) Arch. Biochem.
Biophys. 180, 343 - 347
- Nussinov, R. (1986) Biochimica et Biophysica Acta 866, 93 - 108
- Ochs, H.D., Rosenfeld, S.I., Thomas, E.D., Giblett, E.R., Alper, C.A.,
Dupont. B., Schaller, J.G., Gilliland, B.C., Hansen, J.A., &
Wedgewood, R.J. (1977) New Engl. J. Med. 296, 470
- Olaisen, B., Teisberg, P., Nordhagen, R., Michaelson, T. & Gedde - Dahl,
T. (1979) Nature 279, 736 - 737
- O'Neill, G.J., Yang, S.Y. & Dupont, B. (1978a) Proc. Natl. Acad. Sci.
USA 75, 5165 - 5169

- O'Neill, G.J., Yang, S.Y., Tegoli, J., Berger, R. & Dupont, B. (1978b)
Nature 273, 668 - 670
- O'Neill, G.J., Dupont, B., Pollack, M.S., Levine, L.S. & New, M.I.
(1982) Clin. Immunopathol. 23, 312 - 382
- Orkin, S.H., Old, J., Lazarus, H., Atlay, C., Gurgey, A., Weatherall,
D.J. Nathan, D.G. (1979) Cell 17, 33 -42
- Ozols, J., Heinemann, F.S. & Johnson, E.F. (1981) J. Biol. Chem.
256, 11405 - 11408
- Palsdottir, A. (1986) D. Phil. Thesis, Oxford University
- Pang, S., Pollack, M.S., Loo, M., Green, O., Nussbaum, R., Clayton, G.,
Dupont, B. & New, M.I. (1985) Ann. of N.Y. Acad. of Sci. 458,
111 - 129
- Parker, K.L., Chaplin, D.D., Wong, M., Seidman, J.G., Smith, J.A. &
Schimmer, B.P. (1985) Proc. Natl. Acad. Sci. USA 82, 7860 - 7864
- Pasqualini, J.R. (1964) Nature 201, 501
- Peterson, R.E., Imperato - McGinley, J., Gautier, T. & Starla, E.
(1977) Ann. J. Med. 62, 170
- Petes, T. & Fink, G.R. (1982) Nature 300, 216 - 217
- Phillips III, J.A., Panny, S.R., Kazazion, H.H., Boehm, C.O., Scott,
A.F. & Smith, K.D. (1980) Proc. Natl. Acad. Sci. USA 77,
2853 - 2856
- Phillips, I.R., Shephard, E.A., Ashworth, A. & Rabin, B.R. (1985) Proc.
Natl. Acad. Sci. USA 82, 983 - 987
- Pollack, M.S., Levine, L.S., Zachmann, M., Prader, A., New, M.I.,
Oberfield, S.B. & Dupont, B. (1979) Transplant Proc. XI,
1315 - 1316
- Pollack, M.S., Levine, L.S., O'Neill, G.J., Pang, S., Lorenzen, F.,
Kohn, B., Rondanini, G.F., Chiumello, G., New, M.I. & Dupont, B.
(1981) Am. J. Hum. Genet. 33, 540 - 550

- Porter, R.R. (1979) Complement. In Defense and Recognition IIB. Structural Aspects (MTP Int. Rev. Sci. Ser.-Biochem) ed. E.S.Lennox, 23, 177 - 212 Baltimore: Univ. Park Press
- Porter, R.R., (1983) Mol. Biol. Med. 1, 161 - 168
- Radloff, R., Bauer, W., & Vinograd, J. (1967) Proc. Natl. Acad. Sci. USA 57, 1514 - 1521
- Raum, D., Awdeh, Z., Anderson, J., Strong, L., Grandos, J., Teran, L., Giblett, E., Yunis, E.J. & Alper, C.A. (1984) Am. J. Hum. Genet. 36, 72 - 79
- Reekers, P., Kennedy, L.J. & Schreuder, G.M.T. (1984) Antigen Report: HLA-Bw47. In Histocompatibility Testing (1984), ed. Albert, E.D., 154 - 155 Springer Verlag, Berlin, Heidelberg.
- Reid, K.B.M. & Porter, R.R. (1981) Ann. Rev. Biochem. 50, 433 - 464
- Reid, K.B.M. (1986) Essays in Biochem. 22, 27 - 68
- Rigby, P.W.J., Dieckmann, M., Rhodes, C., Berg, P. (1977) J. Mol. Biol., 113, 237 - 251
- Rittner, C., Hauptmann, G., Gross - Wilde, H., Grosshaus, E., Tongio, M.M. & Mayer. S. (1975) Histocompatibility Testing 1975 ed. Kissmeyer - Nielson, F. 945 - 954 Munksgaard, Copenhagen.
- Rodrigues, N. R., Dunham, I., Yu, C. Y., Carroll, M. C., Porter. R. R. & Campbell, R. D. (1987) EMBO J. 6, 1653 - 1661
- Roos, M.H., Giles, C.M. Demant, P., Mollenhaver, E. and Rother, C. (1984) J. Immunol., 133, 2634 - 2640
- Rosenfield, R.G. & Miller, W.L. (1983) Congenital Adrenal Hyperplasia. In Hirsutism and Virilism. V.B. Mahesh & R.B. Greenblatt, Eds(John Wright - PSG Inc., Boston) 87 - 119
- Ruf, H.H., Wende, P. & Ullrich, V. (1979) J. Inorg. Biochem. 11, 189 - 204
- Rumsby, G., Carroll, M.C., Porter, R.R., Grant, D.B. & Hjelm, M. (1986)

- J. Med. Genetics 23, 204 - 209
- Ryan, K.J., & L.L. Engel. (1957) J. Biol. Chem. 225, 103 - 114
- Sanger, F., Nicklen, S. & Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463 - 5468
- Sanger, F. & Coulson, A.R. (1978) FEBS Letts. 87, 107 - 110
- Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. & Roe, B. (1980) J. Mol. Biol. 143, 161 - 178
- Saiki, R. K., Bugawan, T. L., Horn, G.T., Mullis, K. B. & Erlich, H. A. (1986) Nature (London) 324, 163 - 166
- Schneider, P.M., Carroll, M.C., Alper, C.A., Rittner, C., Whitehead, A.S Yunis, E.J. & Colten, H.R. (1986) J. Clin. Invest. 78, 650 - 657
- Sharp, P.A., Sugden, B. & Sambrook, J. (1973) Biochemistry 12, 3055 - 3063
- Shiroishi, T., Sagai, T., Natsuume-Sakai, S. & Moriwaki, K. (1987) Proc. Natl. Acad. Sci. USA 84, 2819 - 2823
- Siiteri, P.K. & Wilson, J.D. (1977) J. Clin. Endocrinol. Metab. 38, 113
- Simpson, E.R. & Boyd, G.S. In : Functions of the adrenal cortex, ed. K.W. Mckerns, N.Y. Appleton - Century - Crofts (1968) 1, 49 - 76
- Slightom, J.L., Blechl, A.E. & Smithies, O. (1980) Cell 21, 627 - 638
- Smith, H.O. & Nathan, D. (1973) J. Mol. Biol. 81, 419 - 423
- Speiser, P.W., Dupont, B., Rubenstein, P., Piazza, A., Kastelan, A. & New, M.I. (1985) Am. J. Hum. Genetics. 37, 650 - 667
- Spies, T., Morton, C.C., Nedospasov, S.A., Fiers, W., Pious, D. & Strominger, J.L. (1986) Proc. Natl. Acad. Sci. USA 83, 8699 - 8702
- Sogawa, K., Gotah, O., Kawajiri, K. & Fujii-Kuriyama, Y. (1984) Proc. Natl. Acad. Sci. USA 81, 5066 - 5071
- Sogawa, K., Gotah, O., Kawajiri, K., Harada, T. & Fujii-Kuriyama, Y. (1985) J. Biol. Chem. 260, 5026 - 5032
- Southern, E.M. (1975) J. Mol. Biol. 98, 503 - 517

- Southern, E.M. (1979) *Methods in Enzymol.* 68, 152 - 176
- Staden, R. (1982a) *Nucleic Acid Res.* 10, 4731 - 4751
- Staden, R. (1982b) *Nucleic Acid Res.* 10, 2951 - 2961
- Staden, R. (1984) *Nucleic Acid Res.* 12, 521 - 538
- Steinmetz, M., Minard, K., Horvath, S., McNicholas, J., Srelinger, J.,
Wake, C., Long, E., Mach, B. & Hood, L. (1982a) *Nature*,
300, 35 - 42
- Steinmetz, M., Winoto, A., Minard, K. & Hood, L. (1982b) *Cell*, 28,
489 - 498
- Steinmetz, M. & Hood, L. (1983) *Science (Wash.)* 222, 727 - 733
- Steinmetz, M., Stephen, D. & Fischer Linddahl, K. (1986) *Cell*
44, 895 - 904
- Strachan, T., *British Medical Bulletin* (1987) 43, 1 - 14
- Suhara, K., Gomi, T., Sato, H., Itagaki, E., Takemori, S. & Katagiri, M.
(1978) *Arch. Biochem Biophys.* 190, 290 - 299
- Suwa, Y., Mizukami, Y., Sogawa, K. & Fujii-Kuriyama, Y. (1985)
J. Biol. Chem. 260, 7980 - 7984
- Thomas, P.C. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5201 - 5200
- Tilley, C.A., Romans, D.G. & Crookson, M.C., (1978) *Nature*, 276, 713 -
715
- Tomaoki, B. (1973) *J. Steroid Biochem.* 4, 89 - 118
- Travers, P., Blundell, T.L., Sternberg, M.J.E., Bodmer, W.F. (1984)
Nature, 310, 235 - 238
- Trowsdale, J. (1987) *British Medical Bulletin Vol.* 43, 15 - 36
- Ulick, S. (1976) *J. Clin. Endocrinol. Metab.* 43, 92 - 96
- Ullrich, V., Sakurai, H. & Ruf, H.H. (1979) *Acta. Biol. Med. Ger.* 31,
287 - 297
- Wahl, G.M., Stern, M. & Stark, G.R. (1979) *Proc. Natl. Acad. Sci. USA*
76, 3686 - 3687
- Takemori, S. & Kominami, S. 1984 *TIBS* 393 - 396

- Weatherall, D.J. (1985) *The New Genetics and Clinical Practice* (Oxford University Press, N.Y.)
- Werkmeister, J. W., New, M.I., Dupont, B. & White, P.C. (1986) *Am. J. Hum. Genet.* 39, 461 - 469
- White, P.C., New, M.I. & Dupont, B. (1984a) *Proc. Natl. Acad. Sci. USA* 81, 1986 - 1990
- White, P.C., New, M.I. & Dupont, B. (1984b) *Proc. Natl. Acad. Sci. USA* 81, 7505 - 7509
- White, P.C., Chaplin, D., Weis, J., Dupont, B., New, M.I & Seidman, J.G. (1984c) *Nature*, 312, 465 - 467
- White, P.C., New, M.I. & Dupont, B. (1985a) *Immunol. Reviews* 87, 123 - 150
- White, P.C., Grossberger, D., Onufer, B.J., Chaplin, D.D., New, M.I., Dupont, B. & Strominger, J.L. (1985b) *Proc. Natl. Acad. Sci. USA* 82, 1089 - 1093
- White, P.C., New, M.I. & Dupont, B. (1986) *Proc. Natl. Acad. Sci. USA* 83, 5111 - 5115
- White, R.E. & Coon, M.J. (1980) *Annu. Rev. Biochem.* 49, 315 - 356
- Wigler, M., Silverstein, S., Lee, L.S., Pellicer, A., Cheng, Y.C. & Axel, R. (1977) *Cell*, 11, 223 - 232
- Wilkins, L. (1962) *Arch. Dis. Child.* 37, 231 - 241
- Williams, A.F. (1985a) *Nature*, 308, 12 - 13
- Williams, A.F. (1985b) *Nature*, 314, 579 - 580
- Winoto, A., Steinmetz, M. & Hood, L. (1983) *Proc. Natl. Acad. Sci. USA* 80, 3425 - 3429
- Winter, J.S.B. (1980) *J.Pediatr.* 97, 81 - 82
- Winter, G., Fields, S., Gait, M.J. & Brownlee, G.G. (1981) *Nucleic Acid Res.* 9, 237 - 245
- World Health Organisation (1968) *W.H.O. Bull.* 39, 935 - 938

- World Health Organisation (1981) W.H.O. Bull. 59, 489 - 491
- Wu, L.C., Morley, B.J. & Campbell, R.D. (1987) Cell, 48, 331 - 342
- Wu, R. & Taylor, E. (1971) J. Mol. Biol. 57, 491 - 511
- Yanagibashi, K., Haniu, M., Shirely, J.E., Shen, W.H. & Hall, P. (1986)
J. Biol. Chem. 261, 3556 - 3562
- Yasumura, Y., Buonassisi, V. & Sato, G.H., (1966) Cancer Res. 26,
529 - 535
- Yoshioka, H., Morohashi, K., Sogawa, K., Yamane, M., Kominami, S.,
Takemori, S., Okada, Y., Omura, T. & Fujii-Kuriyama, Y. (1986)
J. of Biol. Chem. 261, 4106 - 4109
- Young, R.A. & Davis, R.W. (1983) Proc. Natl. Acad. Sci. USA 80,
1194 - 1198
- Yu, C.Y. & Campbell, R.D. (1987) Immunogenetics 25, 383 - 390
- Yu, C.Y., Belt, K.T., Giles, C.M., Campbell, R.D. & Porter, R.R. (1986)
EMBO J. 5, 2873 - 2881
- Yuan, P., Nakajin, S., Haniu, M., Shinoda, M., Hall, P.F. &
Shively, J.E. (1983) Biochemistry 22, 143 - 149
- Zuber, M.X., John, M.E., Okamura, T., Simpson, E.R. & Waterman, M.R.
(1986) J. Biol. Chem. 261, 2475 - 2482

APPENDIX I

GGTACCATCCAGGAGGCCCTGGCCTCTCACTGAACCCGGCCACTCCTCTTTGGCATGGCC
TCTTCCCAAATCCCCAAACTGCCTCCTTACTCACAAAAGTGGTCTCTGAGTGTCAGTCCA
GTGGGACCCCCACCCCTTATGGCTTCAGTTCCCCAAATAGGGCTGGACCCTTGATCCTGA
TCCAGCTGTGGCTATCCAGCCCCCTTCCTGGGGACTTTGGACTTTGAGCGGGGGCATGCCC
AGTTGTGCTGGGAATCCATACTTTCCCTGGCTGGAGTAGAACCTGTGGACTGTAGTCCTG
AGGGCAGTCATGTTCTGCCTGTGCCTGGAAACACAAGAACTTGACTGCAGAGAGAAGAA
AGAGGAGAGAGGAACAGAGCGAGGAAACTGCCCCGTCTCCGGGGCTTTTTCTGTTCCCTAT
CCTTGGCTTTCTAAGACCAGTGGGGTCCCCTCCTCTGCTTCTTTTTCTGAGTTCTGTGA
AATTCCCCAATCCTTACTTTTTGTCTCAAACCAGCTCAAGGTGGGCTGTTTTCTTTCAA
CCAAAGAAAGGTGCTCCTGGTGGCTAAAGGTACATATTCGA

Nucleotide sequence of the 21-OHase A gene from the KpnI site
in the 5' region to the TaqI site shown in figure 3.12.