The Transcriptional Regulation of the Gene for Human Clotting Factor IX

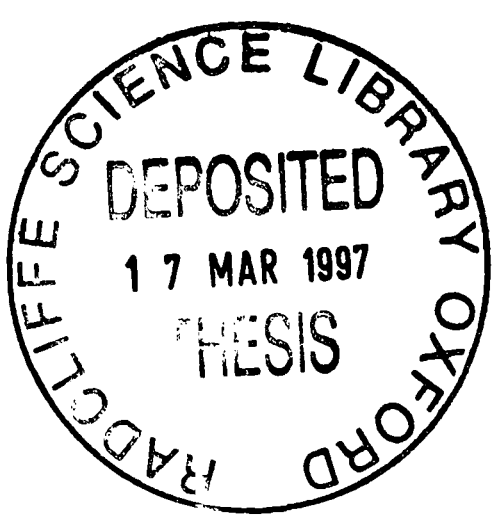
Gareth Edwin Morgan

Sir William Dunn School of Pathology and

Jesus College, Oxford.

A dissertation submitted for the degree of D.Phil.

at the University of Oxford, Trinity Term 1996.



Contents

Acknowledgements

Abstract

Abbreviations

Chapter 1 General Introduction

1.1 Haemophilia B 1

1.1.1 Introduction 1

1.1.2 Therapy for Haemophilia B 1

1.2 The Clotting Cascade 3

1.3 Factor IX 4

1.4 Mutations in the Factor IX Gene 6

1.5 Carrier Detection and Haplotype Analysis of Patients 8

1.6 Eukaryotic Transcription Initiation 9

1.6.1 RNA polymerase II and the General Transcription Factors 9

1.6.2 Initiator Elements and TATA-less Promoters 10

1.6.3 Assembly of the PIC 11

1.6.4 The 'transcriptosome' 12

1.7 Transcription Activation 13

1.7.1 Transcription Factors 13

1.7.2 The Mechanism of Activation 14

1.7.3 Effects of Activators on the PIC 16

1.7.4 Coactivator, Corepressor and Integrator Molecules 17

1.7.5 Chromatin Structure and the Regulation of Transcription 18

1.8 Steroid Hormone Receptors, the Androgen Receptor and AREs 19

1.8.1 An Introduction to Steroid Hormone Receptors 19

1.8.2 Steroid Receptor Associated Proteins (RAPs) 20

1.8.3 The Steroid Hormone Receptor Zinc Finger 22

1.8.4 The Steroid Hormone Receptor Response Element 24

1.8.5 Transcription Activation by Steroid Hormone Receptors 28

1.8.6 The Androgen Receptor 30

1.9 Tissue Specific Gene Expression and the Factor IX Promoter 33

1.10 Haemophilia B Leyden and Suggested Recovery Mechanisms 35

1.10.1 The Leyden Phenotype 35

1.10.2 Haemophilia B Brandenburg and UK232 36

1.10.3 The Overlapping Androgen Response Element 37

1.10.4 The Role of HNF4 37

1.10.5 Site 5 and the PAR Domain Proteins 38

1.10.6 A Possible Indirect Role for AR? 39

1.11 Aims of This Thesis 40

Chapter 2 Materials and Methods

2.1 Genotypes of *E.coli* Bacterial Strains 41

2.2 Preparation of Competent *E.coli* for Transformations 41

2.3 Large Scale Plasmid DNA Preparations 42

2.4 Small Scale Preparations of Plasmid DNA 43

2.5 Electrophoresis of DNA 43

2.6 Plasmid Constructs Used in this Thesis 44

2.6.1 Cloning Vectors 44

2.6.2 Expression Vectors 44

2.6.3 Reporter Gene Vectors 44

2.7 Restriction Digests 45

2.7.1 Restriction Digests and Markers 45

2.7.2 PCR Based Analysis of Polymorphic Loci 45

2.8 Statistical Analysis of the Data from the Haplotype Study 46

2.8.1 The Binomial Distribution 46

2.8.2 χ^2 Goodness of Fit Test (Banic, 1975) 46

2.8.3 The student's t test 47

2.9 Other DNA Manipulations 48

2.9.1 Dephosphorylation of 5' ends 48

2.9.2 Ligation of DNA 48

2.9.3 Preparation of DNA for Sequencing 48

2.10 PCR Reactions 49

2.11 Sequencing of DNA 49

2.12 Protein A-AR DNA Binding Domain Fusion Protein Production and Partial Purification 51

2.13 Determination of Protein Concentration 52

2.14 SDS PAGE of Protein Samples 52

2.15 Western Blots 53

2.16 Synthesis and Purification of Crude Oligonucleotides 53

2.17 Preparation of Radioactive DNA Oligonucleotide Probes 54

2.17.1 Labelling 54

2.17.2 Annealing 54

2.17.3 Purification 55

2.18 Electrophoretic Mobility Shift Assays 55

2.19 Preparation of Double-Stranded Competitor Oligonucleotides for Mobility Shift Assays 56

2.20 Competition Mobility Shift Assays 56

2.21 Phosphorimager Analysis 57

2.22 Mammalian Cell Lines 57

2.23 Construction of ARE Tetramer-CAT Reporter Plasmids 58

2.24 *Sau96*I profiling 58

2.25 Transient Transfections 59

2.26 Preparation of Extracts from Mammalian Cells 60

2.27 Chloramphenicol Acetyl Transferase (CAT) Assays 60

2.28 Luciferase Assays 61

2.29 β-Galactosidase Assays 61

2.30 Charcoal Stripping of Foetal Calf Serum 62

2.30.1. Preparation of Dextran Coated Charcoal 62

2.30.2. Stripping of Foetal Calf Serum 62

Chapter 3: Production and Partial Purification of a Recombinant Protein A-Androgen Receptor Fusion Protein

3.1 Introduction 63

3.2 Production and Partial Purification of the Protein 63

A-AR DNA Binding Domain Fusion Protein

3.3 Characterisation of the Fusion Protein 64

3.4 Functional Assay with the Fusion Protein 65

3.5 Summary 67

Chapter 4: *In vitro* analysis of mutations in the factor IX ARE using a competition gel mobility shift assay.

4.1 Introduction 68

4.2 Mutagenesis Strategy 68

4.3 Optimising the Gel Mobility Shift Assay 70

4.4 The Nature of the Retarded Band 72

4.5 Results of the Competition Gel Mobility Shift Assays 73

4.6 Discussion 76

4.7 Summary 78

Chapter 5: *In vivo* Analysis of Mutations in the Factor IX ARE Using a Transient Transfection Assay.

5.1 Introduction 79

5.2 Mutant Oligonucleotide Cloning 79

5.3 Characterising the Oligonucleotide Reporter Clones 80

5.4 Testing the Oligonucleotide Dimer and Tetramer/Dimer Mixture Reporter Clones *in vivo* 83

5.5 Transfection Control Plasmids 84

5.6 Hormone Response Assays 85

5.7 Results of the Hormone Response Assays 86

5.8 Discussion 88

5.9 Summary 89

Chapter 6: A study of three patients with novel point mutations in the promoter of the factor IX gene.

6.1 Introduction 91

6.2 Haemophilia B Southampton, Patient UK 232. 92

6.2.1 Case History 92

6.2.2 Promoter Transactivation 92

6.2.3 Transcription Factor Binding 94

6.2.4 Transient Transfections to Determine Androgen Receptor/Testosterone Response 95

6.3 HB 388 and HB 396 97

6.3.1 Introduction 97

6.3.2 Promoter Transactivation 98

6.3.3 Transcription Factor Binding 99

6.3.4 Transient Transfections to Determine Androgen Receptor/Testosterone Response 100

6.4 Discussion and Summary 101

Chapter 7 Characterising Patient Haemophilia B Liverpool 2, and the Reason for the High Frequency of Mutations at Nucleotide -6 in the Factor IX Promoter Region

7.1 Introduction 103

7.2 Haemophilia B Liverpool 2 103

7.2.1 Case History 103

7.2.2 Sequence Characterisation and Carrier Detection 104

7.3 Haplotype analysis of patients with -6 G→A mutations 104

7.4 Independent mutation events, or a founder effect? 107

7.5 Discussion and Summary 109

Chapter 8: Summary and General Discussion

8.1 Summary 112

8.2 Discussion Concerning the *in vitro* Study 113

8.2.1 Alternative Data Analysis Techniques 113

8.2.2 The ARE Consensus Sequence 115

8.3 Comparison of the Binding Data with the *in vivo* Transfection Data and Patient Data 116

8.4 Further Evidence for the Importance of an ARE in the Factor IX Promoter in the Leyden Recovery. 119

8.5 Testing the Theory for the Direct Action of AR and Testosterone in the Leyden Recovery 120

8.6 Consequences for Genetic Counselling and Treatment 121

8.7 Analysis of the Overlap Region Between the ARE and the HNF4 Binding Site 122

8.8 Possible Mechanisms for the Recovery of Patients with Mutations Outside the HNF4 Site Spanning Nucleotides -27→-15 123

8.9 Further Work 126

8.9.1 Improvements to the Binding Study 126

8.9.2 Characterisation of AR Action 126

8.9.3 Further *in vivo* Data 127

8.10 Concluding Remarks 128

References 129

Appendix 1 149

Appendix 2 153

Publications 155

Acknowledgements

The work described in this thesis was performed in the laboratory of Professor George Brownlee in the Sir William Dunn School of Pathology. I would like to thank Professor Brownlee for providing me with the facilities necessary for the work and Dr. Penny Handford, Dr. Mauro Figueiredo, and Dr. Peter Winship for their valued encouragement and support through what has been a testing time. I would especially like to thank Dr. Winship for the haplotype data detailed in Appendix 2, and Dr. Figueiredo for his practical assistance throughout. I must also acknowledge the technical support in the department, especially Tim Beesley and Norman Gascoyne, which has been excellent throughout.

Thanks of a different, but nonetheless appreciated, kind go to all of my friends in the Dunn School and Jesus College who have constantly reminded me that man can most certainly not live by science alone.

Abstract

Haemophilia B Leyden is a rare form of haemophilia B in which patients suffer from a bleeding disorder in childhood but completely recover in adulthood. These patients all have point mutations in the gene promoter between nucleotides -21→+13 which cause this altered expression pattern. An unusual patient with a -26 G→C mutation, haemophilia B Brandenburg, did not recover from his bleeding disorder during puberty. It was suggested that this was due to the disruption of the binding of the androgen receptor (AR) to an androgen response element (ARE) in the promoter (-36→-22).

To obtain further evidence for the importance of the ARE in the Leyden recovery, a patient with a -26 G→A mutation, UK 232, was studied here. This mutation, like the -26 G→C, also disrupted the ARE. This patient is now aged 47 and has not recovered from his bleeding disorder, confirming the importance of this nucleotide in the ARE.

To determine the nucleotides important for the response to AR and testosterone in the factor IX ARE, mutations were introduced at each position in the ARE, some spacer positions, and in some flanking positions. These mutant ARE sequences were then tested firstly, *in vitro*, in a competition gel mobility shift assay with an *E.coli* produced protein A-AR DNA binding domain fusion protein, and secondly, *in vivo* in transient transfection assays. In these assays nucleotides -35, -32, -26, -25, and -23 were critical for the response to AR and testosterone, and -36, -31, and -27 impaired this response to a lesser degree. Conflicting evidence was obtained in the two assays concerning the importance of nucleotides -33 and -22.

These data were used to predict the adult phenotypes of two children with mutations in the factor IX promoter (HB 396 and HB 388). It was predicted that one of them, HB 396, would recover from his bleeding disorder as his mutation (-19 G→C) does not disrupt the ARE, whilst the other, HB 388 would not recover as his mutation (-23 C→T) does disrupt this element.

Patient Liverpool 2 with a mild haemophilia B Leyden phenotype was found to have a -6 G→A promoter mutation at a CpG dinucleotide which has been identified in 12 other patients. A haplotype study of 9 of these patients revealed that it is probable both recurrent mutation and a founder effect are responsible for the mutation's high frequency.

Abbreviations

A	adenine
A_{xnm}	absorbance at x nm.
AF1/2	activation function 1/2
AR	androgen receptor
ARE	androgen response element
ARP1	apolipoprotein regulatory protein 1 (also called COUP-TFII)
BHK	baby hamster kidney
bp	base pair
BSA	bovine serum albumin
C	cytosine
CAT	chloramphenicol acetyl transferase
CBP	CREB binding protein
C/EBP	CAAT/enhancer binding protein
CHO	chinese hamster ovary
COUP/Ear3	chicken ovalbumin upstream promoter transcription factor (identical to v-erbA related gene 3.
DBD	DNA binding domain
DBP	albumin D site binding protein
DMEM	Dulbecco's modified Eagle medium
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide
DTT	dithiothreitol
EDTA	ethylene diamine tetra-acetate
EGF-like	epidermal growth factor-like
ER	estrogen receptor
G	guanine
GABPα/β	GA binding protein α/β
Gla	γ-carboxylated glutamic acid
GR	glucocorticoid receptor
GST	glutathione S-transferase
GTF	general transcription factor
Hepes	N-2-hydroxyethylpiperazine-N'-ethanesulphonic acid
HIV	human immunodeficiency virus
HLF	hepatic leukaemia factor
HMG	high mobility group
HNF1	hepatocyte nuclear factor 1
HNF4	hepatocyte nuclear factor 4

hsp	heat shock protein
kb	kilobase
kDa	kilodalton
LEF1	lymphoid enhancer-binding factor 1
LTR	long terminal repeat
MAP	mitogen associated protein
MDa	megadaltons
MMTV	mouse mammary tumour virus
MR	mineralocorticoid receptor
mRNA	messenger RNA
NF1	nuclear factor 1
NMR	nuclear magnetic resonance
ONPG	0-nitrophenyl β-D-galactoside
PAGE	polyacrylamide gel electrophoresis
PBP	prostate binding protein
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
PIC	preinitiation complex
PMSF	phenylmethane sulphonyl flouride
PR	progesterone receptor
PSA	prostate specific antigen
RAP	receptor associated protein
RAR	retinoic acid receptor
RXR	retinoid X receptor
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
rRNA	ribosomal RNA
SDS	sodium dodecyl sulphate
Slp	sex limiting protein
snRNA	small nuclear RNA
SV40	simian virus 40
T	thymine
TAE	Tris-acetate-EDTA electrophoresis buffer
TAF	TBP-associated factor
TB	Tris-borate electrophoresis buffer
TBE	Tris-borate-EDTA electrophoresis buffer
TBP	TATA binding protein
TCRα	T cell receptor α

TFIIA→H	transcription factor IIA→H
tk	thymidine kinase
TR	thyroid hormone receptor
Tris	2-amino-2-hydroxymethylpropane-1,3-diol
tRNA	transfer RNA
VDR	vitamin D receptor

Chapter 1 General Introduction

1.1 Haemophilia B

1.1.1 Introduction

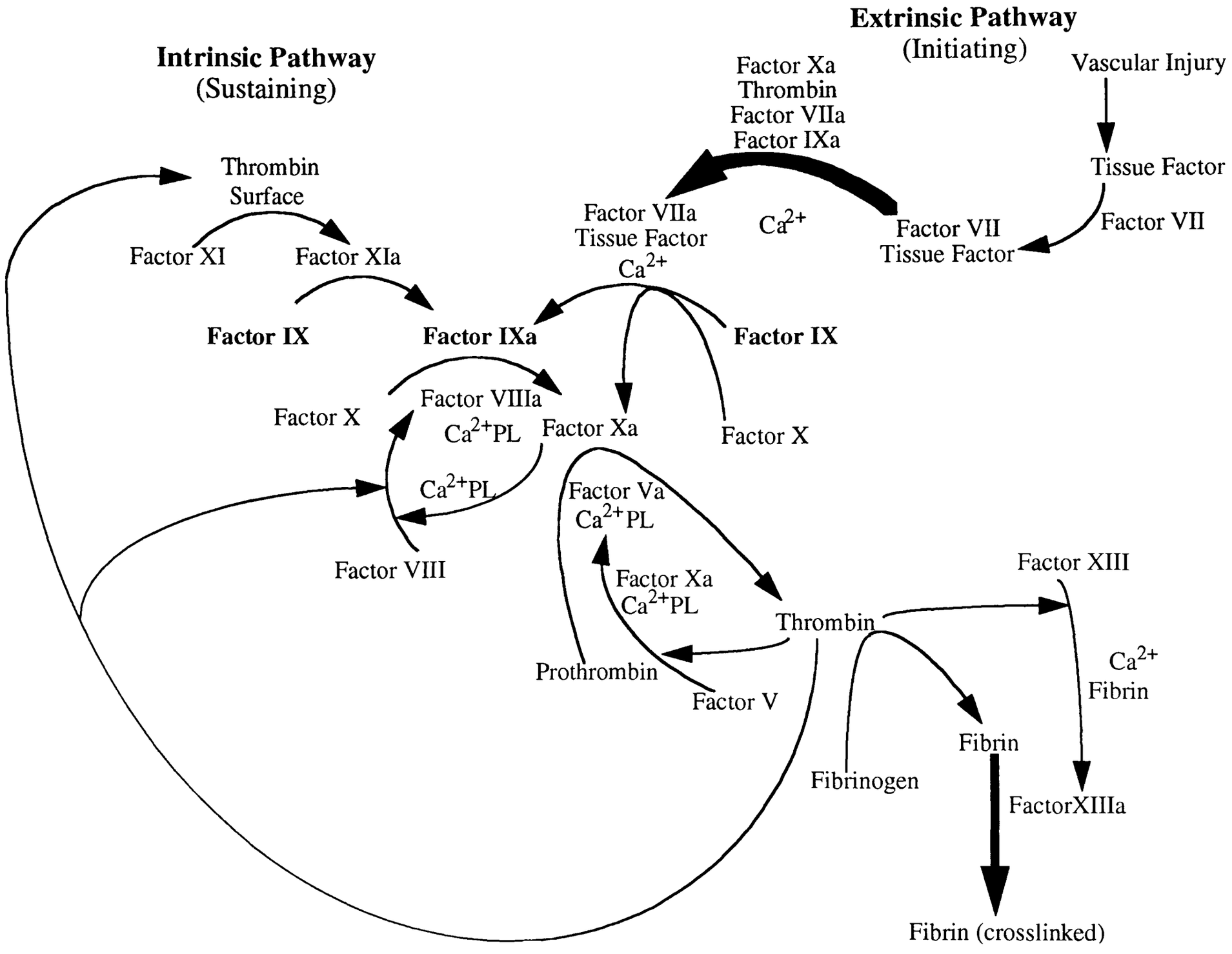
Haemophilia B is an X linked, recessive clotting disorder which is caused by a deficiency or absence of the blood plasma glycoprotein factor IX, or Christmas Factor (see Hedner and Davie, 1989 for review). Haemophilia-like conditions have been reported for over two millennia but it was not until Pavlovsky (1947) found that the plasma from some haemophiliac patients could complement the clotting activity of the plasma from other patients, that the two forms of the disease, haemophilia A and B, were distinguished from one another. Further work allowed the definition of haemophilia B as a separate, clinically indistinguishable disease from haemophilia A, where the clotting defect is in factor VIII (Biggs et al., 1952, Aggeler et al., 1952).

Haemophilia B is the less common form of haemophilia, accounting for about 20% of cases occurring at a frequency of about 1 in every 30,000 males. Females are usually unaffected carriers of the disease, being heterozygous for the defective X chromosome, but can occasionally show symptoms of the disease due to non-random X inactivation (Nisen et al., 1986), or the inheritance of 2 X chromosomes, which are defective at the factor IX locus.

1.1.2 Therapy for Haemophilia B

Factor IX is a critical component of the blood clotting cascade (see Fig 1.1). It occupies a central position in the current model of this cascade, where the sustaining (intrinsic) and initiating (extrinsic) pathways of blood coagulation converge. In each patient with haemophilia B the severity of the disease depends upon the nature of the molecular defect in the factor IX gene causing the haemophilia. The molecular defect determines both the quantity and the activity of factor IX which is synthesised in the liver and secreted into the blood plasma (Giannelli et al., 1996). Approximately 1% of patients with severe haemophilia B also have "inhibitors" (specific anti-factor IX

Figure 1.1. A current model of the blood coagulation cascade, from Davie et al., (1995). Phospholipid is abbreviated to PL, activated clotting factors are denoted by 'a' after the name of the factor. It is not known how factor VII becomes activated, but a number of the candidate proteases are shown. The cascade is initiated when tissue damage occurs, which results in the formation of the activated tissue factor-factor VII complex which can then activate the extrinsic pathway (or initiating pathway). After clot formation is initiated, the maintenance of fibrin is dependent upon the intrinsic pathway (or sustaining pathway). It has been suggested that this pathway can be activated by the action of thrombin on factor XI (Gailani and Broze, 1991).


Intrinsic Pathway
(Sustaining)
Extrinsic Pathway
(Initiating)
Vascular Injury
Tissue Factor
Factor VII
Factor VII
Tissue Factor
Factor Xa
Thrombin
Factor VIIa
Factor IXa
Factor VIIa
Tissue Factor
Ca^{2+}
Ca^{2+}
Thrombin
Surface
Factor XI
Factor XIa
Factor IX
Factor IXa
Factor IX
Factor X
Factor VIIIa
Ca^{2+}PL
Factor Xa
Factor X
Ca^{2+}PL
Factor VIII
Factor Va
Ca^{2+}PL
Factor Xa
Ca^{2+}PL
Prothrombin
Factor V
Thrombin
Factor XIII
Ca^{2+}
Fibrin
Fibrin
Fibrinogen
FactorXIIIa
Fibrin (crosslinked)

antibodies) present in their blood plasma (Thompson, 1987) which complicates the treatment of the patient (see Furie et al., 1994 for review). Thus, haemophilia B is a heterogeneous disease and the treatment given must be tailored to the individual patient requirements. In particular, a patient's factor IX activity level is critical in determining the severity of the disease. Table 1.1 describes how the clinical severity varies with the plasma factor IX activity. One unit (u) of factor IX activity is described as the clotting activity found in 1ml of pooled normal plasma (Furie et al., 1994). In the non-haemophiliac population, factor IX clotting activities vary between individuals from 0.6u to 1.5u.

Table 1.1 Plasma factor IX levels and the clinical severity of haemophilia B.

Factor IX Activity (u)	Nature of Haemophilia B
<0.01	Severe
0.01→0.05	Moderate
0.05→0.5	Mild

There are many clinical problems associated with defective blood coagulation, (see Hedner and Davie, 1989 for a review). The most common problems which can develop in haemophiliacs are associated with spontaneous bleeding episodes. These can occur in the joints, especially the elbow and the knee, and must be treated early as otherwise the joints become chronically deformed and inflamed and this can lead to osteoporosis. Spontaneous bleeding can also occur in tissues (haematoma), in the gastrointestinal tract, and into the cranium, which is obviously particularly dangerous for the patient. All of these spontaneous bleeding episodes require intervention therapy with clotting factor. As well as spontaneous bleeding episodes, any trauma in a haemophiliac must be regarded as serious and treated immediately. Prophylactic treatment of severe haemophiliacs is practised in some cases, as the number of spontaneous bleeding episodes can be dramatically reduced if the factor IX clotting

levels are maintained as low as 0.04u. This greatly reduces the risk of serious joint damage and other long term problems, but it is not used routinely in countries such as the US (see Furie et al., 1994 for review) due to its high cost.

Other problems are associated with the source of clotting factor used for the therapy. Many blood products over the last twenty years have been contaminated with viruses and patients who require long term, intensive therapy have been exposed to diseases such as hepatitis B and human immunodeficiency virus (HIV). Furie et al., (1994) reported that between 1979 and 1984, over half of the haemophiliac population in the US was infected with HIV, and Evatt (1995) reported that the median age of death of haemophiliacs in the US dropped from 57 years between the 1979-1981 period to 40 years between the 1987-1989 period. The development and introduction of recombinant sources of clotting factors is obviously an important step in ensuring that an infection of the haemophiliac population on the scale of that seen in the last 15 years is not repeated. Recombinant factor VIII prepared in CHO (chinese hamster ovary) cells is now available, whilst recombinant factor IX should shortly become available since it is currently in phase III clinical trials.

The ultimate replacement therapy with clotting factor sources is gene replacement therapy. Several approaches have been explored in animal models to date (see Brownlee, 1995 for review) but an acceptable clinical treatment has not yet been devised.

1.2 The Clotting Cascade

A schematic view of the current understanding of the mechanism of blood coagulation is shown in Fig 1.1 (see Davie, 1995 for review). There are two pathways involved in the blood coagulation mechanism, the sustaining (intrinsic) pathway and the initiating (extrinsic) pathway. As is implied from these names, one pathway is thought to initiate the process after tissue damage and the other to sustain the process (Gailani and Broze, 1991). Factor IX can be seen to occupy a position of central importance, where the two pathways converge (see Fig 1.1). It is activated by factor XIa or the

complex of factor VIIa/tissue factor/Ca^{2+}. Once activated, factor IXa is then capable of activating factor X complexed with factor VIIIa, phospholipid and Ca^{2+}. The protease which activates the tissue factor/factor VII complex is not yet known, but it has been suggested (Davie, 1995) that it could be factor Xa, thrombin, factor VIIa, factor IXa, or another cellular or plasma protease. There is no doubting the importance of factor IX in this process as factor IX deficiency causes haemophilia B.

1.3 Factor IX

The sequences of the factor IX cDNA (Anson et al., 1984) and the gene (Yoshitake et al., 1985) were determined more than 10 years ago. The gene is located on the long arm of the X chromosome at q26-q27 (Chance et al., 1983). It encodes a single polypeptide glycoprotein with a molecular mass of approximately 57 kDa (Di Scipio et al., 1977). Factor IX is a vitamin K-dependent protein and circulates in the blood as an inactive zymogen. It is activated by two proteolytic cleavages which cleave amino acids 146→180 (see Fig 1.2). The remaining light and heavy chain of factor IX are held together by a disulphide bond (see Fig 1.2).

The exon/intron structure of factor IX is shown in Fig 1.2. The gene consists of 8 exons and the domain structure of the protein corresponds closely with this exon arrangement (see Fig 1.2). The nascent factor IX polypeptide includes a prepeptide and a propeptide sequence not found in the mature zymogen, which is secreted into the blood. The prepeptide sequence directs the factor IX polypeptide into the endoplasmic reticulum where it is cleaved off. The propeptide sequence has been studied extensively as several naturally occurring mutations in this region cause haemophilia B (Giannelli et al., 1996). This sequence shows a high degree of sequence homology to the propeptides of other vitamin K-dependent blood plasma proteins and it is thought that the sequence represents part of the recognition sequence of the γ-carboxylase (Yamada et al., 1995, Wu et al., 1991).

Exon **b** codes for the first domain of mature factor IX. This domain contains 11 γ-carboxylated glutamic acid residues (Gla residues) and is known as the Gla domain,

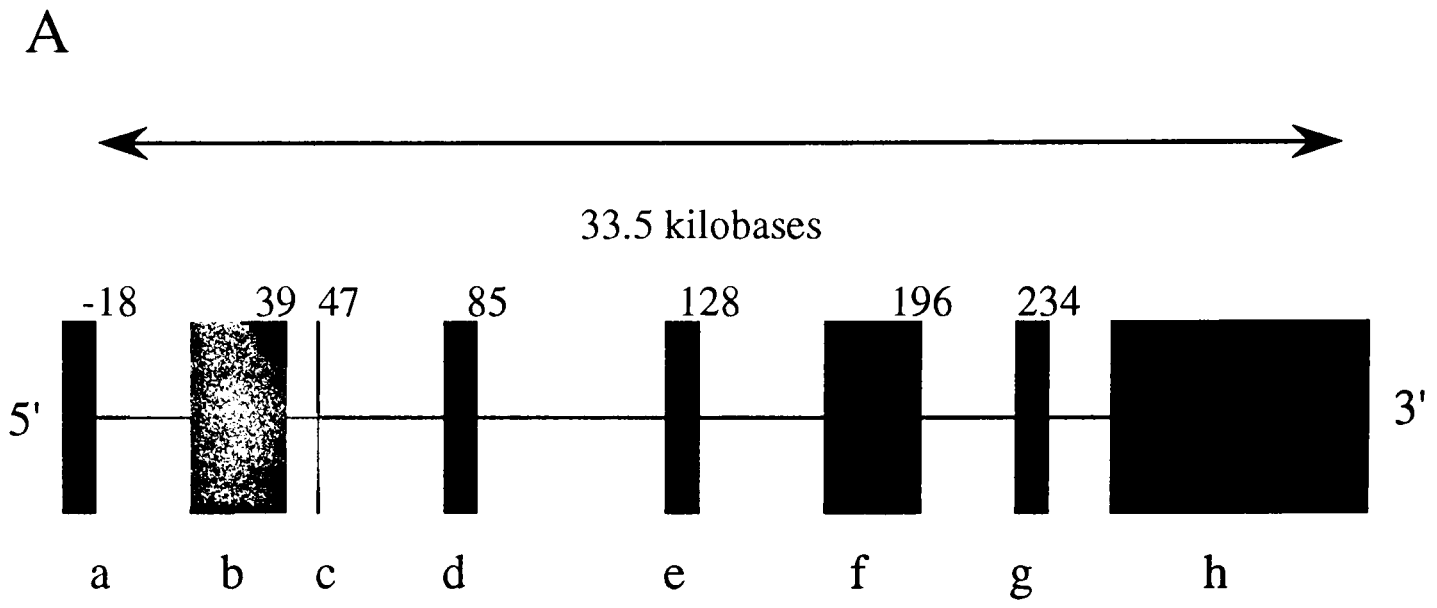


Figure 1.2. A comparison of the exon organisation (A) of the factor IX gene with the domain structure (B) of the factor IX protein. The structure of both the precursor protein (B), and the mature protein (C) are shown. Amino acids 145→181 are cleaved out of the protein upon activation. The protein amino acid number is shown above the end of each exon (A), and the beginning and end of each domain (B). The light (N-terminal) and heavy (C-terminal) chains of factor IX are held together by a single disulphide bond after activation (C). Cat= catalytic domain. For other abbreviations see text.

see Fig 1.2, (Fryckland et al., 1976). The γ-carboxylation of these glutamic acid residues is known to be critical for the function of the vitamin K-dependent plasma proteins (Stenflo and Ganrot, 1973). The next domain in the factor IX protein is coded for by exon **c**. This domain is known as the "hydrophobic stack" domain (H, see Fig 1.2) and is often regarded as being part of the Gla domain. The role of these two domains is closely linked to the binding of Ca^{2+} ions, which allows the protein to adopt the correct conformation (Ohlin et al., 1988, Handford et al., 1990, Jacobs et al., 1994). The crystal structure of the first 148 amino acids of prothrombin (including its Gla domain) also showed a significant conformational change occurring upon binding Ca^{2+} (Soriano-Garcia et al., 1992). Recent studies (Freedman et al. 1995a, 1995b) have described the NMR structures of these domains and it has been shown that the metal ion free structure lacks the necessary structured regions to bind to phospholipids. The phospholipid binding of factor IX/IXa is an essential stage in the blood clotting cascade (Ahmad et al., 1989).

EGF (epidermal growth factor)-like domains are present in a large number of proteins including notch, fibrillin-1, and many of the blood clotting or anticoagulant proteins (e.g. factors VII, IX, X and protein C). The function of the two EGF-like domains in factor IX is not clearly defined at the moment. The N-terminal, B-type EGF-like domain of factor X is known to interact with the Gla domain upon binding Ca^{2+} (Medved et al., 1994), and it is presumed that this interaction plays a major role in defining a conformation needed for interaction with the other components of the clotting system. B-type EGF-like domains are known to assume an ordered structure in the presence of Ca^{2+}, and the crystal structure of this domain from factor IX (Rao et al., 1995) strengthens the idea that this type of domain plays an important role in maintaining the structural integrity of an active factor IX molecule.

The role of the second, type A, C-terminal EGF-like domain is less clear. Mutations in this domain cause haemophilia B (Giannelli et al. 1996), and the antigen levels in such patients are generally low. This suggests that the factor IX protein may

not be in its native conformation in these patients, possibly resulting in an instability of the protein.

The activation domain encompasses amino acids 128→195 and contains the portion of the protein which is cleaved by factor XIa or factor VIIa in order to activate the factor IX zymogen (see Fig 1.2). Finally, the catalytic domain of the factor IX molecule is a classical serine protease domain, showing homology to trypsin and chymotrypsin, as well as the other blood clotting proteins which contain serine protease domains. The catalytic domain also contains a binding site for Ca^{2+} ions and undergoes a conformational change after both activation and Ca^{2+} binding (Bajaj et al., 1992).

Factor IX undergoes many post-translational modifications, some of which are critical for its function, and others which are of unknown significance. Cleavage of the pre- and propeptide are vital for the correct function of factor IX in plasma (Bristol et al., 1994). The γ-carboxylation of the glutamyl residues in the Gla domain is also critical for the binding of Ca^{2+} to the molecule, and for its interaction with phospholipid (Stenflo and Ganrot, 1973, Straight et al., 1985). Factor IX has 0.4 equivalents of hydroxyl modifications per molecule to aspartate 64 (to form β-hydroxyaspartate), in the N-terminal EGF-like domain. The significance of this modification is uncertain since mutations at this position impair the function of the factor IX molecule (Giannelli et al., 1996), but they do not affect Ca^{2+} binding (Sunnerhagen et al., 1993).

Human factor IX is glycosylated at six positions, the oligosaccharide chains making up about 17% of the molecular weight of the protein. It bears N-linked oligosaccharides at residues 157 and 167, and O-linked oligosaccharides at residues 53, 61, 159 and 169 (the last two being in the activation peptide).

1.4 Mutations in the Factor IX Gene

The latest published haemophilia B database of the point mutations and short additions and deletions known to cause haemophilia B (Giannelli et al., 1996) highlights the heterogeneity of the disease (see Fig 1.3). Of the 1380 patients studied to date, there are 552 unique molecular defects. These defects include gross

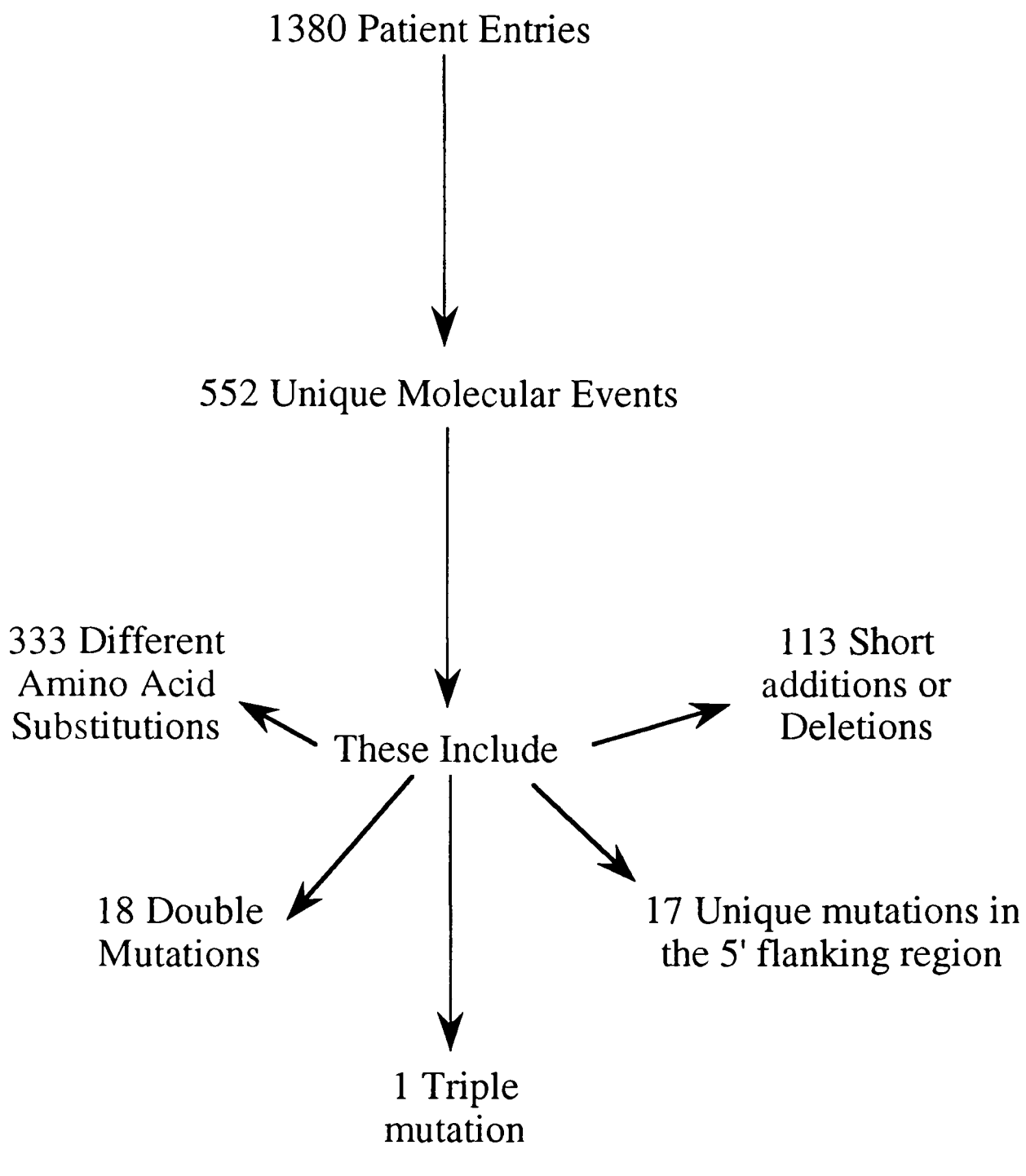


Figure 1.3. A summary of some of the data contained in the most recent haemophilia B database (Giannelli et al., 1996), illustrating the heterogeneity of the mutations which cause the disease.

deletions/additions, splice site mutations and point mutations (nonsense and missense). There are now 333 different amino acid substitutions known to cause varying degrees of haemophilia B in the patient population. These mutations can give vital clues as to the functional importance of particular amino acids in the protein. For example, mutations are known at:

(i) 9 of the 12 γ-carboxyglutamyl residues in the Gla domain,

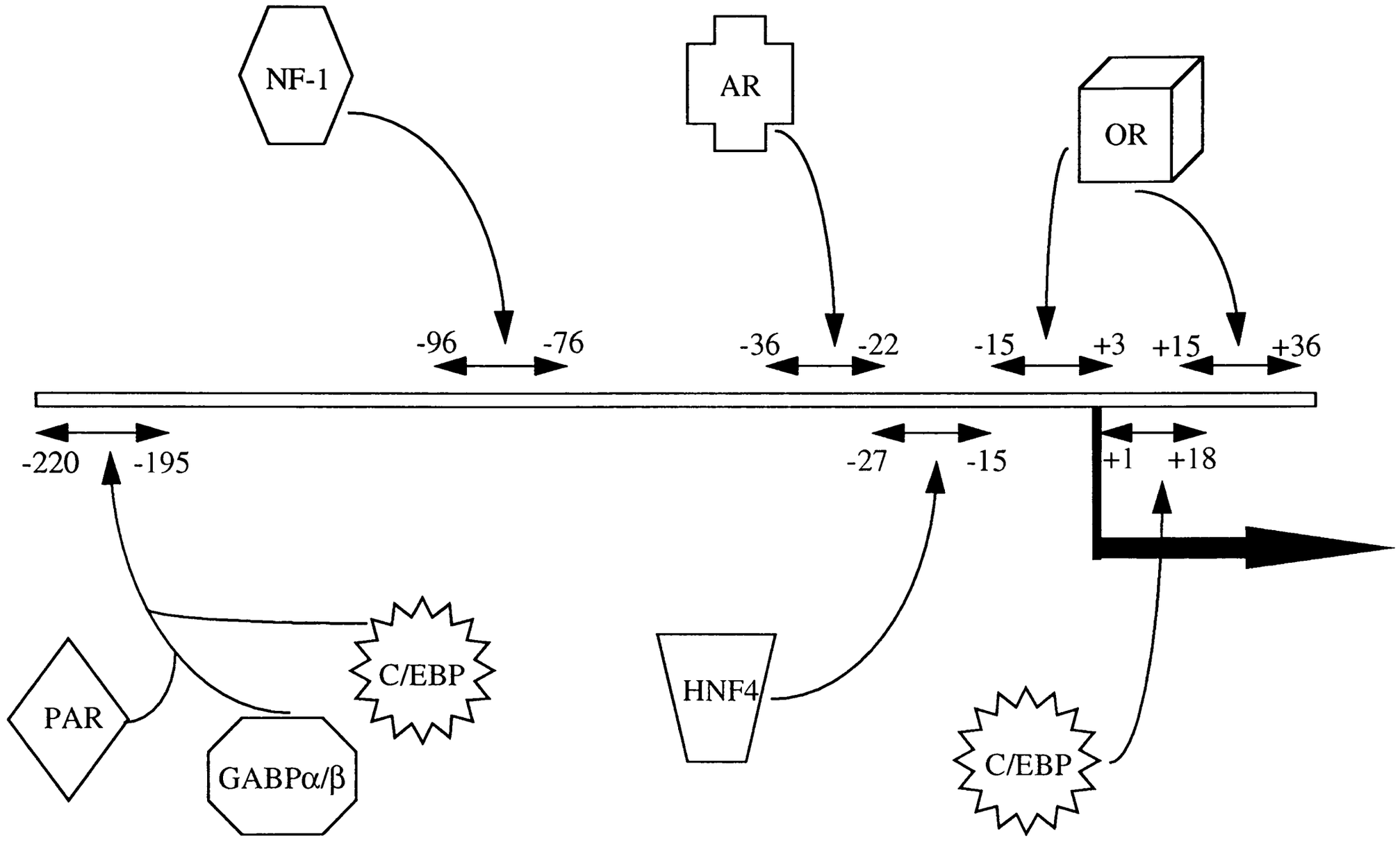
(ii) 2 of the 3 residues (serine 365, and aspartate 269, but not the histidine 221, Giannelli et al., 1996) thought to be important in the active site in the catalytic domain,

(iii) all of the 22 cysteine residues, highlighting the importance of the disulphide bonds in maintaining the structure of factor IX.

This database also lists 17 mutations found in the 5' flanking region of the gene (see Fig 1.3). These are regulatory mutations, and these patients produce normal factor IX, but at much reduced levels (Giannelli et al., 1996) due to the disruption of the factor IX promoter region (Crossley and Brownlee, 1990). These 17 mutations have stimulated numerous studies on the regulation of the factor IX gene (Reitsma et al., 1988, 1989; Crossley and Brownlee, 1990; Salier et al., 1990; Hirosawa et al., 1990; Reijnen et al., 1992, 1993; Crossley et al., 1992; Picketts et al., 1993, 1994; Kurachi et al., 1994; Boccia et al., 1996; Naka and Brownlee, 1996), just as they have in the protein C gene promoter also, Berg et al., (1994). These promoter studies have allowed the identification of several *cis*-acting elements and of the *trans*-acting factors binding to these elements. A summary of these sites and elements can be seen in Fig 1.4, and is listed in Table 1.2.

Table 1.2 The positions of the *cis*-acting elements in the factor IX gene promoter and the *trans*-acting factors reported to act upon them. The nucleotide numbering is with reference to the transcription start site at +1 (Anson et al., 1984). HNF4= hepatic nuclear factor 4, ARP1= apolipoprotein regulatory protein 1 (also called COUP-TFII), COUP/Ear3= chicken ovalbumin upstream promoter transcription factor (identical to v-erbA related gene 3, (see Fig 1.4 for other definitions)

Figure 1.4. A schematic view of the known transcription factor binding sites in the factor IX promoter region (-220→+36), the transcription start site is at +1, Anson et al., (1984). OR= orphan receptors, (HNF4, COUP-Ear3, and ARP1); AR= androgen receptor; NF-1= nuclear factor 1; C/EBP= CAAT enhancer binding protein; HNF4= hepatic nuclear factor 4; GABPα/β= GA binding protein α/β; PAR= PAR domain proteins DBP (D site binding protein), and HLF (hepatic leukaemia factor).


NF-1
AR
OR
-96
-76
-36
-22
-15
+3
+15
+36
-220
-195
-27
-15
+1
+18
C/EBP
PAR
GABPα/β
HNF4
C/EBP

Position	*Trans*-acting factor(s)	Reference
+15→+36	HNF4; ARP1; COUP-Ear3	Naka and Brownlee, 1996
+1→+18	C/EBP	Crossley and Brownlee, 1990
-15→+3	HNF4; ARP1; COUP-Ear3	Naka and Brownlee, 1996
-27→-15	HNF4	Crossley et al., 1992; Reijnen et al., 1992
-36→-22	AR	Crossley et al., 1992
-34→-23	Unidentified protein	Kurachi et al., 1994
-99→-76	NF-1	Crossley and Brownlee 1990
-220→-195	C/EBPα; DBP; HLF; GABPα/β	Boccia et al., 1996

1.5 Carrier Detection and Haplotype Analysis of Patients

The extensive study of mutations and polymorphisms in the factor IX gene (Graham et al., 1991, Sommer et al., 1992, Ketterling et al., 1994, Green et al., 1991) has made factor IX one of the most comprehensively characterised genes in the human genome. In many cases it is desirable to trace defective factor IX genes through pedigrees in order to counsel individuals who may, or may not, be carriers of haemophilia B. In these cases, the method of choice in the analysis may be single strand conformation polymorphisms (SSCP), which depends upon the differential migration of different DNA sequences through native polyacrylamide gels (Sarkar et al., 1992), or the analysis of RFLPs (restriction fragment length polymorphisms) (Morgan et al., 1995). Both methods take advantage of naturally occurring sequence polymorphisms in the human population of factor IX genes. The combination of the sequence at each of the polymorphic sites in a gene will give that gene a "sequence signature" which may allow it to be traced in the pedigree. Restriction fragment length

polymorphisms are analysed by using the polymerase chain reaction (PCR) to amplify regions of the genomic DNA from the patient being analysed. The fragments amplified are then subjected to restriction digest (see 2.7.3 and Chapter 7). The particular combination of the RFLPs present in a particular allele is known as its haplotype.

Apart from the important task of carrier detection in affected pedigrees, analyses of different genes in human populations of different ethnic origin (Figueiredo et al., 1992, 1994) can give clues to the history of population movements. In particular, these data can provide evidence for founder effects being established for certain mutations in the haemophilia B population (Ketterling et al., 1991, Knobloch et al., 1993, Morgan et al., 1995).

1.6 Eukaryotic Transcription Initiation

I will now introduce the processes of transcription initiation/regulation regarding eukaryotic RNA polymerase II. The extensive literature in this field means that this account is necessarily brief.

1.6.1 RNA polymerase II and the General Transcription Factors

In prokaryotes a number of operons are unregulated and at these promoters all that is needed to initiate transcription is an RNA polymerase holoenzyme, consisting of 4 core subunits, and a dissociable σ factor. The σ factor determines the type of promoter to which the RNA polymerase is targeted. However, in eukaryotic cells, the process is more complex. There are three types of RNA polymerase enzyme (I, II, and III). RNA polymerase I copies the RNAs for 18S, 5.8S and 28S rRNA, RNA polymerase III copies all of the tRNAs and 5S rRNA whilst RNA polymerase II copies all of the mRNA precursors and some small RNAs such as U1 snRNA (a component of the splicing apparatus). Of these enzymes RNA polymerase II is the best characterised and so this review will only concern this enzyme. RNA polymerase associates with many general transcription factors (GTFs) before promoter recognition occurs and transcription is initiated (Zawel and Reinberg, 1995 for a review). In effect, these GTFs in eukaryotes act in much the same way as the prokaryotic σ factor.

1.6.2 Initiator Elements and TATA-less Promoters

Two core promoter elements have been identified in eukaryotes. These are the TATA-box and the initiator element (see Fig 1.5). Not all core promoter sequences will contain these two elements, for example the factor IX gene does not contain a TATA-box. The TATA-box of core promoters is recognised by the GTF TFIID. TFIID is made up of at least nine different polypeptides, TBP (TATA binding protein) and TBP-associated factors (TAFs). It is TBP which actually binds to this element and the crystal structure of this protein complexed to DNA has been deduced (Nikolov et al., 1992). It is thought that the other GTFs will then assemble on the promoter to form the preinitiation complex (PIC). In particular, the role of TFIIA and TFIIB has become clearer from recent work by Geiger et al., (1996) and Nikolov et al., (1995) in which the crystal structures of the TFIIA/TBP/DNA and TFIIB/TBP/DNA complexes were determined, and a combination of these structures then modelled (see Fig 1.6). These structural studies shed new light on the stoichiometry of the interactions between GTFs, TBP and DNA which preempt the formation of the PIC, and will contribute towards a precise molecular understanding of the transcription initiation event.

In promoters lacking the TATA-box the initiator element, identified by Smale and Baltimore, (1989), is believed to direct correct transcription initiation. The GTF which interacts with this element was again found to be TFIID, but the fact that TFIID was required for transcription from these TATA-less promoters, and TBP could not substitute for it showed that it was not the TBP component of TFIID which recognised the initiator (see Fig 1.5). Two subunits of the Drosophila TFIID complex, a 60kDa subunit and a 150kDa subunit, were identified as contacting DNA in initiator elements in cross-linking studies (Gilmour et al., 1990). The use of Drosophila partial TFIID complexes containing TBP, TAF150 and TAF250 (Verrijzer et al., 1994, 1995) has allowed transcription from initiator elements to be studied (see Fig 1.5).

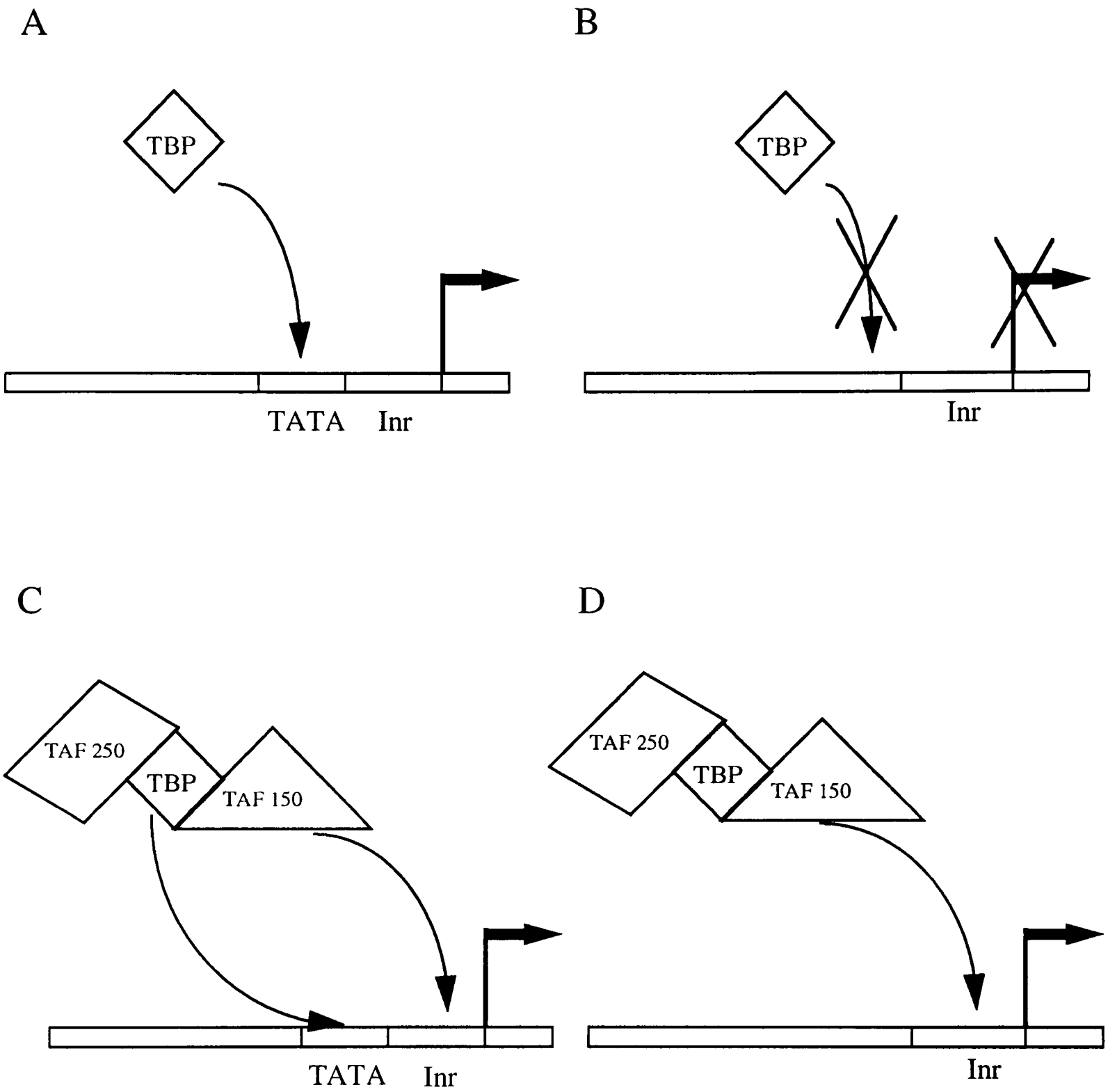


Figure 1.5. TBP can recognise TATA containing promoters and allow transcription to initiate (A), but cannot initiate transcription from TATA-less promoters (B). When TBP has the GTFs TAF 250 and TAF 150 bound it can now initiate transcription from TATA containing and some TATA-less promoters (C and D). Inr= initator element, TATA= TATA box.

Figure 1.6. A model of the complex formed between the transcription factors TBP, TFIIA and TFIIB. This model was constructed by combining the crystal structures of the TBP/TFIIA/DNA (Tan et al., 1996, Geiger et al., 1996), and TBP/TFIID/DNA (Nikolov et al., 1996) complexes. The transcription start site on the DNA (the adenovirus major late promoter) is labelled +1. The colour code for the proteins is as follows:

light blue= core TBP's amino-terminal repeat

dark blue= core TFIIB carboxy-terminal repeat

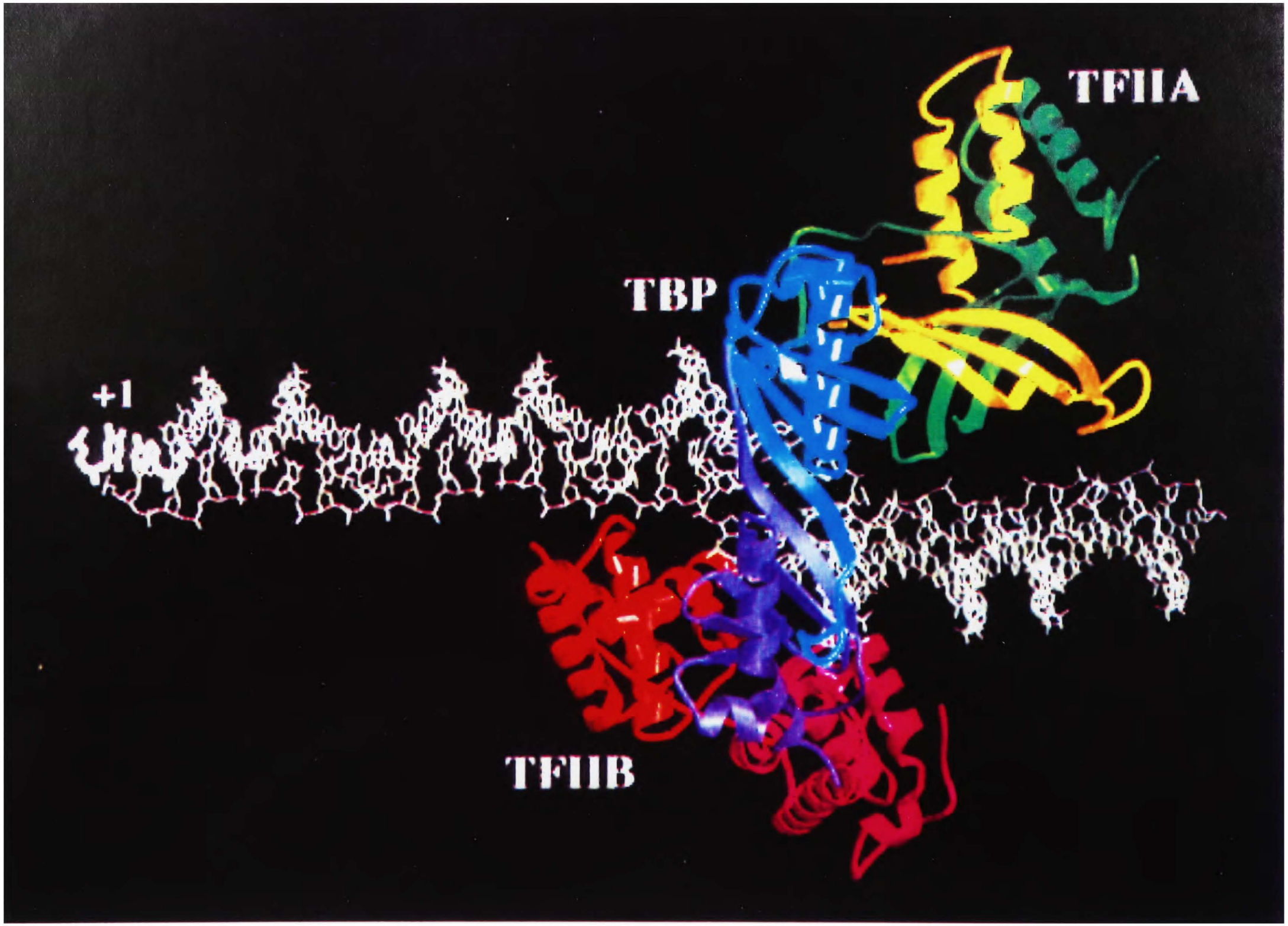
red= core TFIIB amino-terminal domain

magenta= core TFIIB carboxy-terminal domain

yellow= TFIIA small subunit

green= TFIIA truncated large subunit.

The model lacks the divergent amino-terminus of TBP, the amino-terminal 111 amino acids of TFIIB, and a non-conserved portion of the large subunit of TFIIA. From Burley (1996).

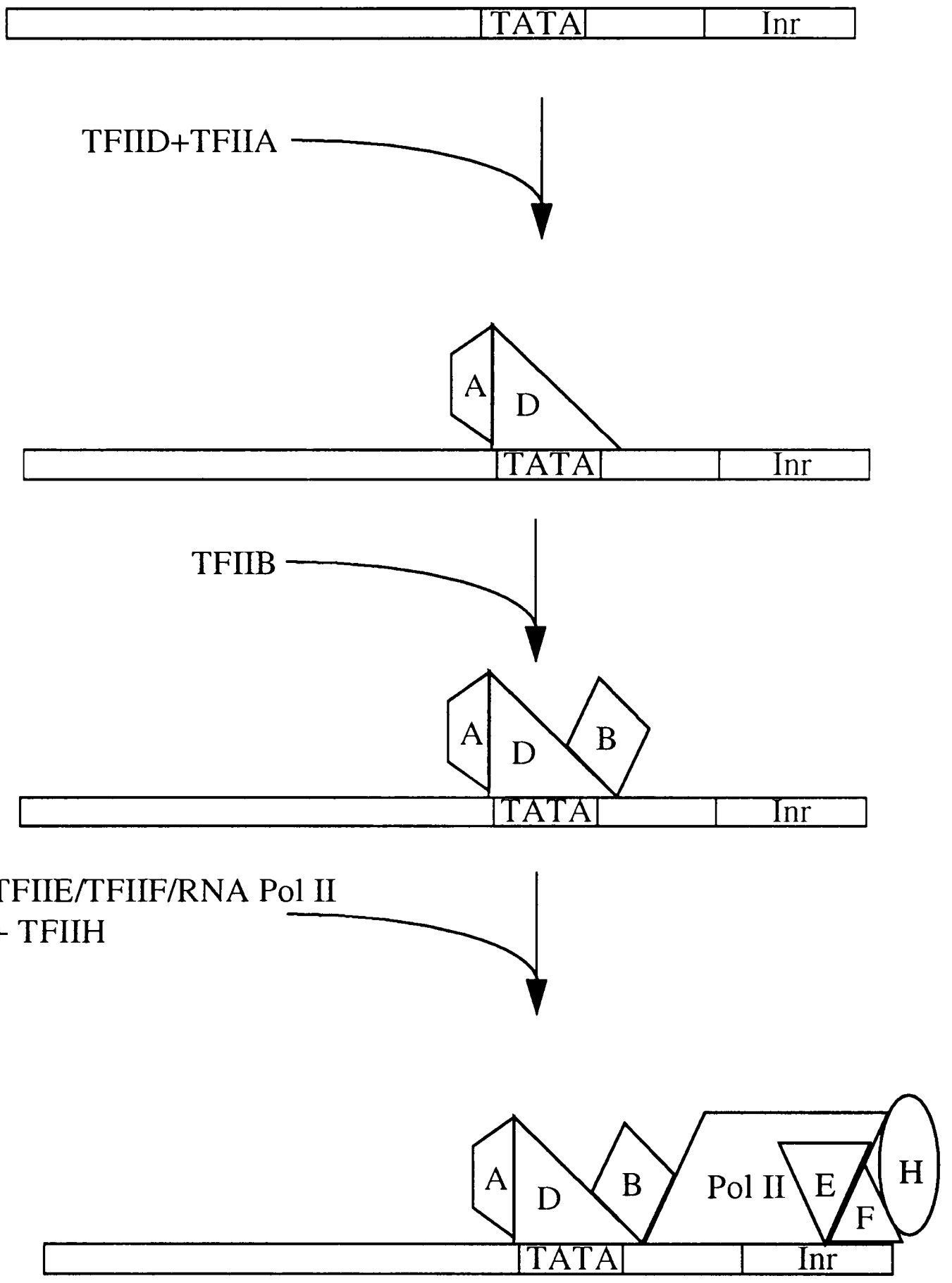

TFIIA
TBP
+1
TFIIB

1.6.3 Assembly of the PIC

TFIID is not the only component of the RNA polymerase/GTF complex which contacts the DNA at the promoter (Goodrich et al., 1996 for review). The polymerase itself is assumed to make contacts with the DNA both upstream and downstream of the transcription start site. TFIIH contains two subunits which are known to have helicase activity making it likely they contact with the DNA at some stage during transcription initiation. Experiments by Coulombe et al., (1994), suggest that subunits of TFIIA, TFIIB, RNA polymerase, the small subunit of TFIIF (RAP30) are within 1nm of the pyrimidine rings in the major groove of the DNA. RAP30 contains a non-specific DNA binding domain that is essential for transcription (Tan et al., 1994).

One of the current models (see Hanna-Rose and Hansen, 1996 for summary) for the assembly of the PIC on the promoter DNA is shown in Fig 1.7. TFIID binds to the DNA first followed by TFIIB. TFIIA is then thought to interact with the TBP subunit of TFIID to stabilise its binding to the promoter and enhance transcription, or to relieve the inhibition of transcription. A complex of RNA Pol II, TFIIE and TFIIF can then be recruited, along with TFIIH. TFIIB acts as a bridging protein between TBP and the RNA polymerase whilst TFIIF consists of two subunits and is capable of associating with RNA polymerase in the absence of other factors and DNA. The recruitment of RNA polymerase is affected by the small subunit of TFIIF (Flores et al., 1991, Tyree et al., 1993) and both subunits of TFIIF are required for proper elongation of transcription by the RNA polymerase (Price et al., 1989). TFIID, TFIIB and TFIIF can recruit the RNA polymerase onto the promoter DNA and TFIIE appears to be necessary for the binding of TFIIH to the complex. TFIIH is a multisubunit GTF and contains a kinase activity that can phosphorylate the C-terminal domain of the RNA polymerase, a DNA-independent ATPase, and DNA helicase activities. Activation of this complex requires the hydrolysis of ATP (in addition to the rNTPs to make the RNA). TFIIH is the only GTF known to possess an ATPase function and so it is probably this GTF which activates the complex. The TFIIE and TFIIH binding is necessary to allow elongation of initiated transcripts. In the absence of these two GTFs

Figure 1.7. One current theory on the mechanism of the assembly of the PIC on a eukaryotic promoter (Hanna-Rose and Harness, 1996). The order of the assembly of the PIC is shown and differs slightly from other models (e.g. Buratowski, 1994). The shape and position of the different factors is meant to be schematic only. The order of assembly of these factors may change under the influence of activator proteins (see Fig 1.9). The RNA polymerase is thought to be in a 3 MDa complex with the SWI/SNF complex, the SRB (suppressors of RNA Pol II) proteins, and the GTFs (Halle and Meisterernst, 1996). This complex has been speculatively termed the "transcriptosome".


TATA
Inr
TFIID+TFIIA
A
D
TATA
Inr
TFIIB
A
D
B
TATA
Inr
TFIIE/TFIIF/RNA Pol II
+ TFIIH
A
D
B
Pol II
E
F
H
TATA
Inr

the RNA polymerase can initiate from some promoters, but cannot elongate efficiently (Goodrich and Tjian, 1994).

1.6.4 The 'transcriptosome'

In addition to the multiple GTFs, another level of regulatory complexity becomes evident if one examines closely the complex formed on promoters at transcription initiation. The core RNA polymerase in eukaryotic cells is approximately 600kDa. This size increases to 1.5-2.0MDa when the GTFs are added and increases even further to about 3MDa if the interaction with the SWI/SNF family of proteins (Winston and Carlson, 1992) is considered. The SWI/SNF family of proteins is known to disrupt chromatin structure, and were identified as activators of transcription (Peterson and Herskowitz, 1992), and as such could be an important component of the transcription complex. The size of this complex has heightened speculation that the polymerase complex may be more immobile in the cell than was previously assumed and raises the possibility that 'transcriptosomes' are present in the eukaryotic cell, each of which could be an immobile transcription factory, which would pull the DNA through it rather than be a mobile polymerase complex which tracks along DNA searching for promoters (Jackson et al., 1993, Halle and Meisterernst, 1996).

Thus, it can be seen that the initiation of transcription from eukaryotic promoters requires the assembly of a multitude of factors onto the DNA, the precise order of whose assembly is still not completely resolved. It is this complex that transcription activators will contact, rather then simply the RNA polymerase, so promoter-specific gene regulation is almost certainly a result of multiple protein-DNA and multiple protein-protein interactions.

1.7 Transcription Activation

1.7.1 Transcription Factors

The regulation of the transcription of any eukaryotic gene requires an array of signals to be integrated with the RNA polymerase complex at the transcription start site. The simplest way in which this is achieved is that *cis*-acting elements in the gene respond to *trans*-acting factors which are capable of interacting with the DNA and altering the activity or stability of the PIC. The *cis*-acting elements in this scheme are sequences of DNA found sometimes proximal to the transcription start site, but often they can be many kilobases distant. These *cis*-acting DNA sequences will bind to *trans*-acting factors called transcription factors. Transcription factors can activate or repress the level of transcription from promoters depending upon the exact nature of the multiprotein complex which assembles on an enhancer or promoter (Diamond et al., 1990).

Transcription factors are modular proteins (see Johnson and McKnight, 1989 for a review), they are composed of domains which allow the protein to perform its function. These domains typically (but not exclusively) include:

(i) a DNA binding domain to tether the protein to the correct place in the promoter region to perform its function,

(ii) a multimerisation domain which allows the protein to interact to form complexes such as homodimers, or heterodimers with other factors,

(iii) an activation/repression domain to contact the PIC and stabilise or destabilise the PIC.

Many transcription factors can act as activators of transcription when the binding sites which mediate their activation are multimerised upstream of heterologous promoter sequences. A number of reporter gene vectors have been developed to exploit this property (Luckow and Schutz, 1987) in order to identify sequences which may act as *cis*-acting elements in promoters.

1.7.2 The Mechanism of Activation

The simplest hypothesis for the regulation of transcription from promoters by transcription factors is that they alter the stability of the PIC. This may be achieved in two ways. First, the structure of the DNA around the transcription start site may be altered when the transcription factor binds, and this conformational change in the DNA could affect the binding of the polymerase complex. Transcription factors which act in this way may be considered to be 'architectural' components of promoters or enhancers. The mouse high mobility group (HMG) domain protein lymphoid enhancer-binding factor 1 (LEF1) binds to one of three sites in the enhancer in the mouse T cell receptor α gene (TCRα). It is a vital component of the enhancer but does not activate transcription independently. Its functions are to bind to the minor groove of the DNA, inducing a sharp bend (Giese et al., 1992), see Fig 1.8, and to mediate interaction between the other proteins in the enhancer complex (Giese and Grosschedl, 1993, Carlsson et al., 1993), see Fig 1.8. The protein HMG I(Y) is also known to order multiple sites in the enhancer of the human interferon β gene (Du et al., 1993) in such a way as to allow highly specific viral activation of the gene.

Secondly, the transcription factor may affect the stability of the PIC by directly contacting it (see Fig 1.9). Transcription factors often contain distinct activation domains and several different types have been identified. These include acidic, proline-rich and glutamine-rich domains. The acidic-rich activation domains are thought to contact TBP (as well as other components of the PIC) itself and this function is seen to be remarkably conserved through evolution, with human TBP being able to respond to yeast acidic activators in a yeast cell (Strubin and Struhl, 1992). In contrast to this, glutamine and proline-rich activation domains do not function in yeast cells, suggesting that the proteins which these domains contact are not highly conserved through evolution. Glutamine-rich domains have been reported to contact the Drosophila protein TAF110 (Hoey et al., 1993), and a yeast homologue of this protein has not yet been identified. The picture may not be as simple as this classification suggests since it is thought that transcription factors with the same type of activation domain may function

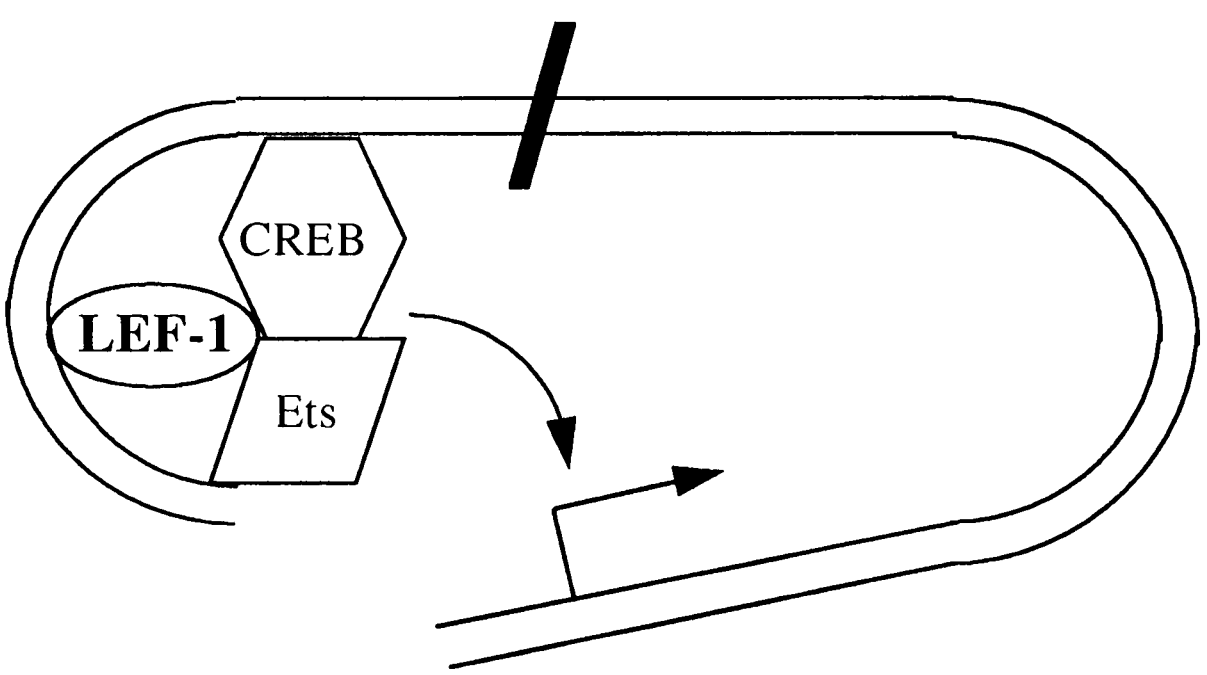


Figure 1.8. The transcription factor LEF-1 is known to bind to the minor groove of DNA and induce a bend of about 130° in the intronic enhancer of the mouse T cell receptor a (Giese et al., 1992). The role of LEF-1 is thus to bend the DNA at the enhancer site and bring together the other two protein binding sites. It is the HMG domain of the LEF-1 which bends the DNA, and another region of the protein is thought to mediate protein-protein interactions between the other two transcription factors (Ets and CREB) which bind to this enhancer region.

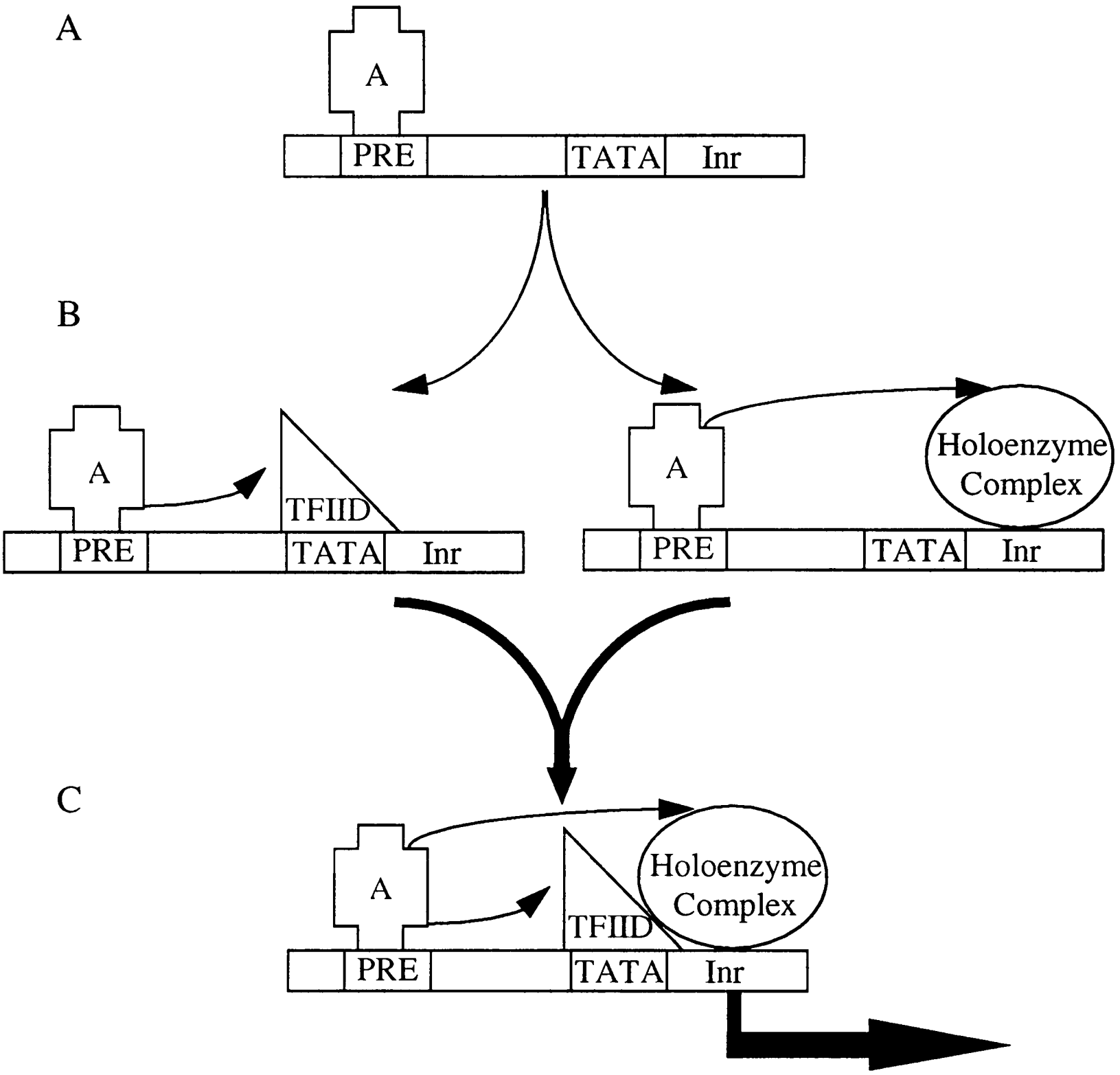


Figure 1.9. One of the current models for the activation of transcription by a transcription factor. The first stage involves the binding of the transcription factor to its binding site in the promoter of the gene (A). The next stage involves the recruitment of either TFIID to the TATA box or the recruitment of the holoenzyme complex. The PIC is then completed by the recruitment of the missing partner, and transcription can be initiated. PRE= proximal response element (to any transcription activator).

in different ways (Gill et al., 1994), and so several different types of glutamine-rich and acidic domains are thought to exist. In these domains mutagenesis studies (Cress and Triezenberg, 1991, Gill et al., 1994) have identified a number of hydrophobic amino acids residues as being important in the activation process. Therefore, it is possible that the glutamine, proline or acidic residues act in an indirect way to organise the hydrophobic surface of the activation domain in such a way as to favour interactions with the GTFs in the PIC. Thus, activation might be driven by the same hydrophobic forces which drive protein folding in general, and it is by multiple, weak interactions that the PIC is stabilised.

Activation domains of transcription factors have now been identified as making direct contacts with several of the GTFs and TAFs known to be present in the PIC. TFIID (TAF110, TAF140), TFIIB, TFIIF and TFIIH have all been identified as being contacted by transcription factors (Gill et al., 1994, Triezenberg, 1995, Zawel and Reinberg, 1995, Nerlov and Ziff, 1995). In particular, one of the transcription factors which is important for the transcription activity of the factor IX promoter in cell culture studies, HNF4, has been shown to interact with the GTF TFIIB (Malik and Karathanasis, 1996). Another of the factors known to be important for the regulation of factor IX transcription, C/EBP, has been demonstrated to bind both to TBP and TFIIB (Nerlov and Ziff, 1995). This feature of transcription factor function parallels the transcription activation process in prokaryotes, during which the targets for the regulatory proteins are the α and σ subunits of the RNA polymerase. Recent studies (Blau et al., 1996) have subdivided activators into three groups which depended upon their ability to bind the GTF TFIIH and their ability to cooperate with the transcription activator Sp1 in an activation assay. It was deduced that those activators which could bind TFIIH (Type II activators) could stimulate elongation of transcription, whereas those that could not only stimulated initiation of transcription and only interacted with TFIID or TFIIB (Type I activators, e.g. Sp1). The type II activators could be subdivided into those which could synergise with Sp1 (Type IIa, e.g. HIV Tat) and those which could not (Type IIb, e.g. p53, VP16). It was found that type I activators

only stimulate initiation of transcription, type IIa predominantly stimulate elongation of transcription, and type IIb can stimulate both steps in the process. It is also thought highly probable that these classes will be subdivided further as factors stimulating individual steps in initiation or elongation are identified.

1.7.3 Effects of Activators on the PIC

The protein-protein contacts which are made during the transcription activation process are thought to have many effects on the nature of the PIC. It is possible that the function of some activators is to induce the recruitment of some GTFs into the PIC (see Fig 1.10A). Much evidence has been obtained that this is certainly true for some promoters (Klein and Struhl, 1994, Sauer et al., 1995). The effect of this increased recruitment of the GTFs would be to increase the stability of the PIC present on that particular promoter.

An activator-GTF interaction may affect the conformation of a PIC already assembled on a promoter (see Fig 1.10B). This conformational change may then have the effect of making the PIC more competent to perform subsequent events in the transcription process, or induce the recruitment of other GTFs. When VP16 activates transcription it is known to alter the conformation of a PIC as it alters the susceptibility of the GTF TFIIB in the PIC to proteolysis (Roberts and Green, 1994).

Another mechanism by which the activator could influence the activity or stability of the PIC is by means of covalent modification (see Fig 1.10C). There are three kinases whose activity is known to be influenced by transcription activators- a subunit of TFIIH, the kinase/cyclin pair associated with RNA polymerase II in yeast and the TAF250 of TFIID (Serizawa et al., 1995, Liao et al., 1995, Dikstein et al., 1996). The TFIIH subunit is known to phosphorylate the C-terminal domain of the large subunit of RNA polymerase, and TAF 250 has a serine kinase activity which can phosphorylate the large subunit of TFIIF (RAP74).

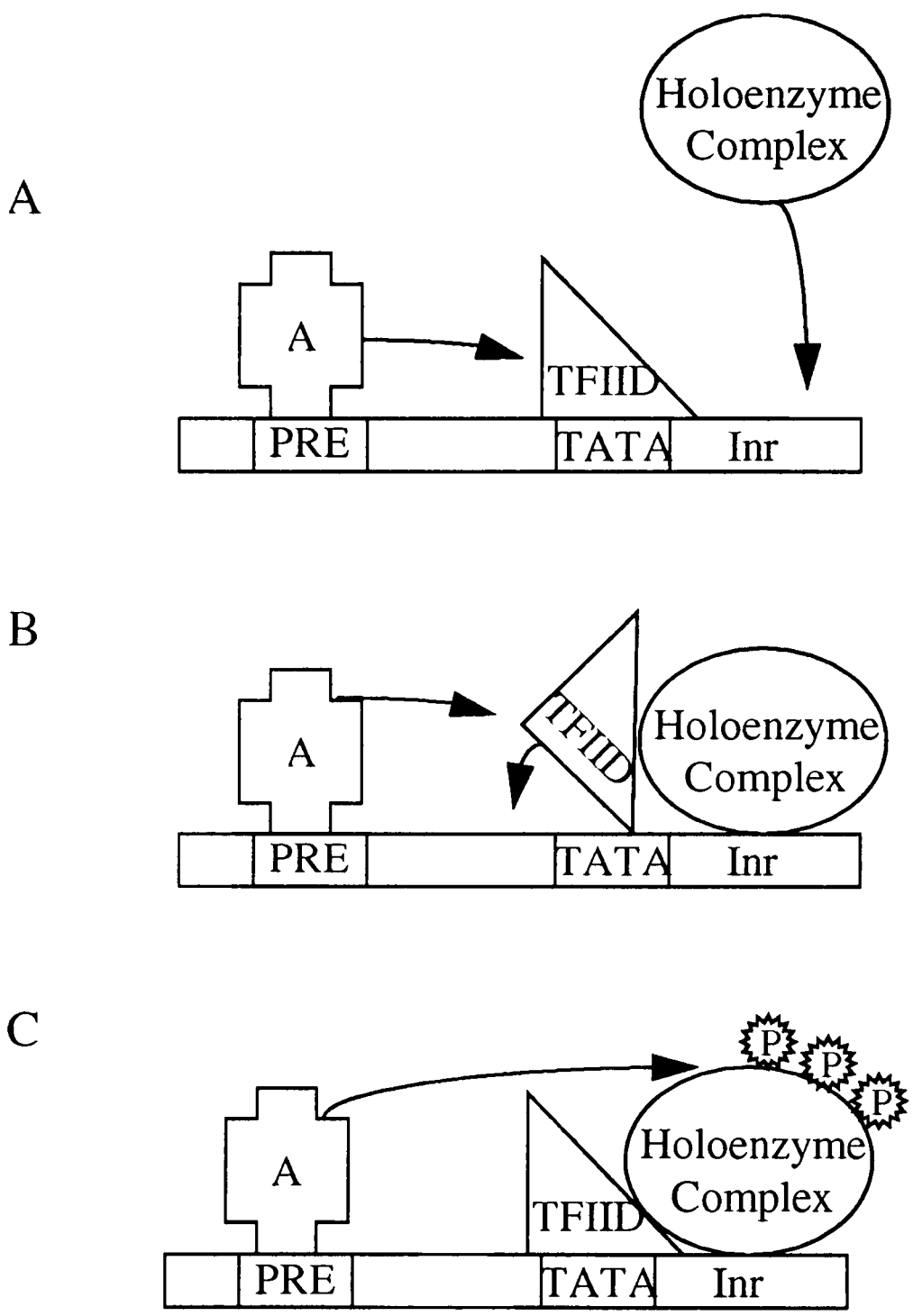


Figure 1.10. Proposed mechanisms for the activation of RNA Pol II transcription by transcription factors. The transcription factors could directly recruit GTFs (A) which would then increase the recruitment of the holoenzyme complex to the promoter; they could induce a conformational change in an inactive or poorly active PIC to convert it to a highly active one (B); they may cause a covalent modification of the PIC (P = phosphorylation) which could increase its activity.

1.7.4 Coactivator, Corepressor and Integrator Molecules

Transcription factors can cooperate with molecules known as cofactors, corepressors or adaptors which assist in the mediation of their action. In particular, the family of nuclear hormone receptors (which includes the androgen receptor (AR), glucocorticoid receptor (GR), progesterone receptor (PR), mineralocorticoid receptor (MR) and the thyroid and retinoic acid receptors (TR, RAR, and RXR) have been identified as interacting with corepressor molecules such as SMRT (Chen and Evans, 1995), and coactivators such as SRC-1 (Onate et al., 1995). These corepressors and coactivators do not directly contact the DNA of the hormone response element in the promoter, but exert their effect through protein-protein interactions with the hormone receptor itself. Some of these molecules, such as RIP140 (Cavailles et al., 1995) associate with the AF2 domain of the nuclear receptors which was known to be essential for hormone dependent transcription activation before the existence of the cofactor molecules was known.

A further level of complexity was added to the mechanism of transcription activation and silencing by the nuclear hormone receptor family when the function of the integrator molecule, CBP (CREB binding protein)/p300, was determined (Kamei et al., 1996). This protein is capable of binding TFIIB and both the nuclear receptor family and nuclear targets of plasma membrane receptors such as Jun, Fos and CREB. Distinct domains of the CBP molecule interact with these different transcription factors and it was also found that the CBP/p300 family is capable of interacting with coactivator molecules such as p160 (variants of the SRC-1 protein, Kamei et al., 1996) see Fig 1.11. These interactions have important implications for the regulation of these pathways in the nucleus. The activating properties of CBP could be due to the presence of a consensus bromodomain in the molecule (Winston and Carlson, 1992). This domain is associated with transcription activation and may have some effect over the local chromatin structure in the region. It has long been known that the nuclear hormone receptors can inhibit transcription activation by the AP-1 family of transcription factors of which Jun and Fos are the best known examples. CBP/p300 is

Figure 1.11 Schematic diagram showing the interaction of several transcription activators and coactivators with the p300/CBP protein. The regions of the protein important for interaction with the nuclear receptors (NR) is shown in green; the region shown to interact with p160, (the nuclear receptor coactivator) is shown in red; the regions which interact with E1A and Fos in orange; and the region which interacts with CREB (cAMP response element-binding protein) and Jun in blue/grey. By interacting with these diverse pathways, and partitioning between the different groups of factors differently, the p300/CBP molecule allows the integration of these responses within the cell at the level of gene expression (Kamei et al., 1996).

Membrane receptors
signalling via PKA

Growth hormones signalling
via membrane receptors

CREB

Jun

Fos

**p300/
CBP**

NR

CREB/Jun

E1A/Fos

p160

p160

Nuclear receptors

Ligand

thought to be at a limiting concentration in the cell nucleus and it now appears that the mechanism of the hormone receptor inhibition of AP-1 involves competition between the 2 groups of transcription factors for the protein CBP/p300, which is important in forming a hormone/coactivator/integrator complex on the DNA that can then activate transcription. The partitioning of the CBP/p300 family of proteins between various transcription factors might be the mechanism whereby the cell can create antagonistic relationships between signalling pathways at the level of transcription activation (see Fig 1.11). Kamei et al., (1996) also suggested that the CBP molecule may be involved in coordinating the response of the nuclear receptors with another new family of transcription factors, the PAS/HLH (PAS domain/ helix-loop-helix family) (Hankinson, 1995), by virtue of the homology of the PAS domain with the region of p160 protein responsible for its interaction with CBP.

1.7.5 Chromatin Structure and the Regulation of Transcription

DNA in a eukaryotic nucleus is not naked, but is packaged into a nucleosomal array by histones to form chromatin. The histones bind to the DNA as an octamer, forming a nucleosome particle. The crystal structure of this particle (Richmond et al., 1984), revealed that this octamer was capable of binding 145bp of DNA into about 1.8 left-handed superhelical turns in which each turn bends the DNA double helix by about 47°. Therefore, all of the mechanisms of transcription regulation discussed so far in this chapter must be able to function in the context of this packaged, distorted DNA. The formation and maintenance of the nucleosomal array is vital to the compaction of DNA within the eukaryotic nucleus, but at the same time this array must be flexible enough to allow *trans*-acting factors access to their *cis*-elements in the transcription control regions of genes.

It is generally thought that the activation of genes packaged in chromatin requires the disruption of the structure of the nucleosomes to free the *cis*-elements in the gene control regions (Kornberg and Lorch, 1991 for review). In particular, the disruption of chromatin structure around the promoters of heat shock genes requires a

specific DNA binding protein (the GAGA factor), see Travers (1994), for review. The GAGA factor has been shown to disrupt chromatin structure in an ATP-dependent fashion *in vitro*. The PIC in eukaryotes has been shown to be associated with the SWI/SNF protein complex see 1.6.4 (Wilson et al., 1996). This protein complex is known to be able to disrupt chromatin structure (see Carlson and Laurent, 1994 for review), and it has been suggested that the initiation of transcription by RNA polymerase II at promoters with nucleosomes in unfavourable positions is dependent upon the ability of this SWI/SNF complex to make the initiation site accessible (Struhl, 1996). The exact positioning of the nucleosomes in gene control regions has been the subject of many investigations (see Pruss et al., 1995 for review), as it is thought that some *trans*-acting factors can bind to their *cis*-elements whilst these elements are still packaged into nucleosomes (see Fig 1.12) (Truss et al., 1995), with the nucleosome structure possibly acting to reduce steric hindrance around adjacent *cis*-elements in the promoter to allow effective activation of transcription. In particular, Truss et al, (1995) suggested that it is the method of binding of the transcription factor to its binding site which will determine whether or not the factor requires the chromatin structure at that site to be disrupted (see Fig 1.12). The steroid hormone receptors are among those factors which, it is suggested, do not necessarily require the disruption of nucleosomes packing the DNA in which their binding sites are found.

1.8 Steroid Hormone Receptors, the Androgen Receptor and AREs

1.8.1 An Introduction to Steroid Hormone Receptors

Steroid hormone receptors are vital components of the endocrine pathway. They are ligand-inducible transcription factors which allow the cell to respond to lipophilic hormones at the level of gene expression. They are modular proteins, and family members show (to varying degrees) conservation of six regions of primary structure (Gronemeyer and Laudet, 1995), regions A→F (some receptors, e.g. the progesterone receptor, lack region F). These regions contain the functional domains of the receptors see Fig 1.13. From Fig 1.13 it can be seen that the steroid hormone family of receptors

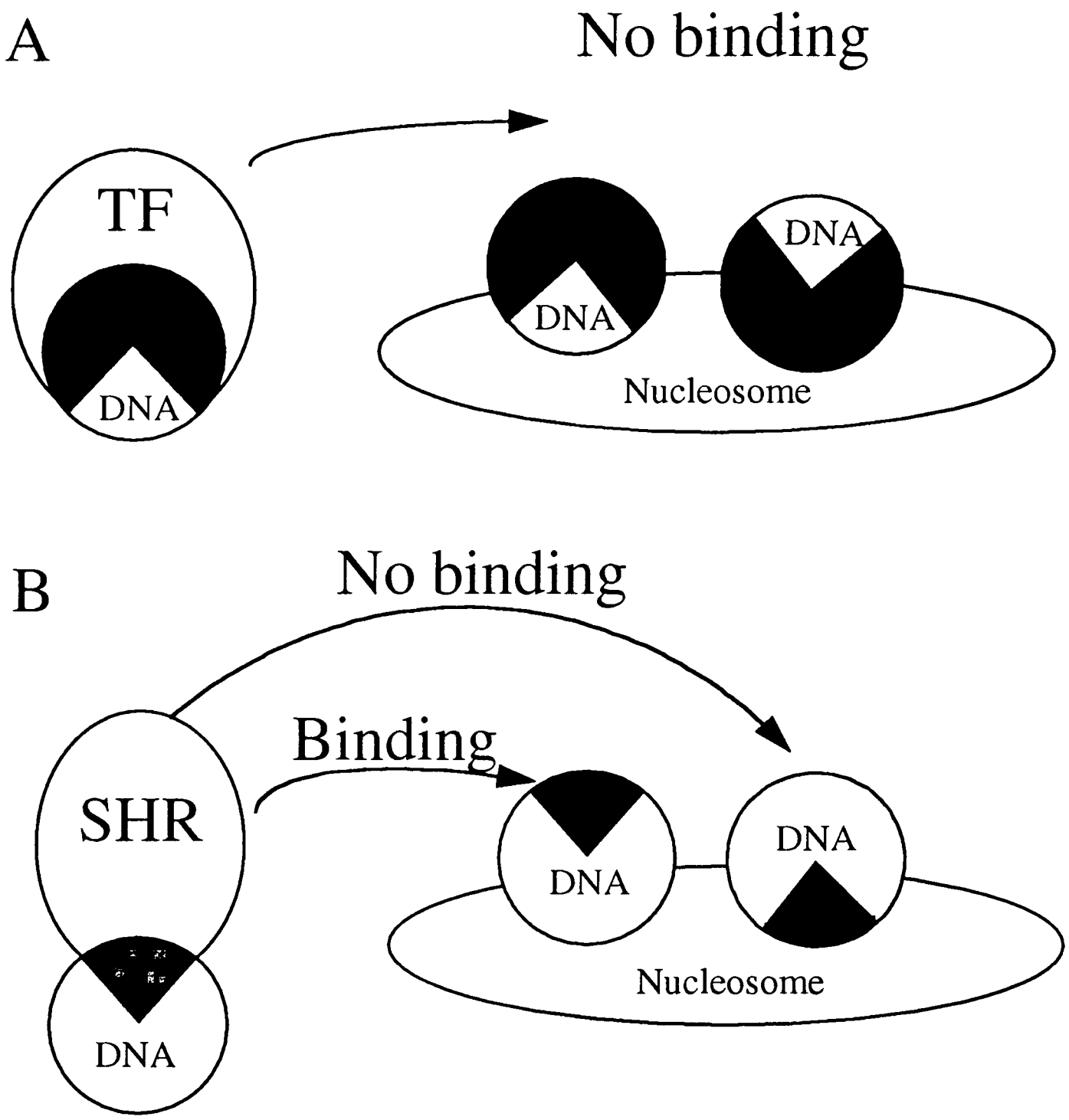


Figure 1.12. Schematic view of (A) a transcription factor such as NF-1 or OTF1 binding to DNA, and (B) a steroid hormone receptor (SHR) binding to its site. As NF-1 or OTF contact more than half of the helix circumference (shown in black), these factors cannot bind to DNA when it is packaged into a nucleosome as their sites will be partially concealed by the histones. The steroid hormone receptors only contact a narrow sector of the major groove (Beato et al., 1995), and so they can bind to their sites even if the nucleosome remains bound, provided that the nucleosome phasing is correct and the entire steroid response element is accessible.

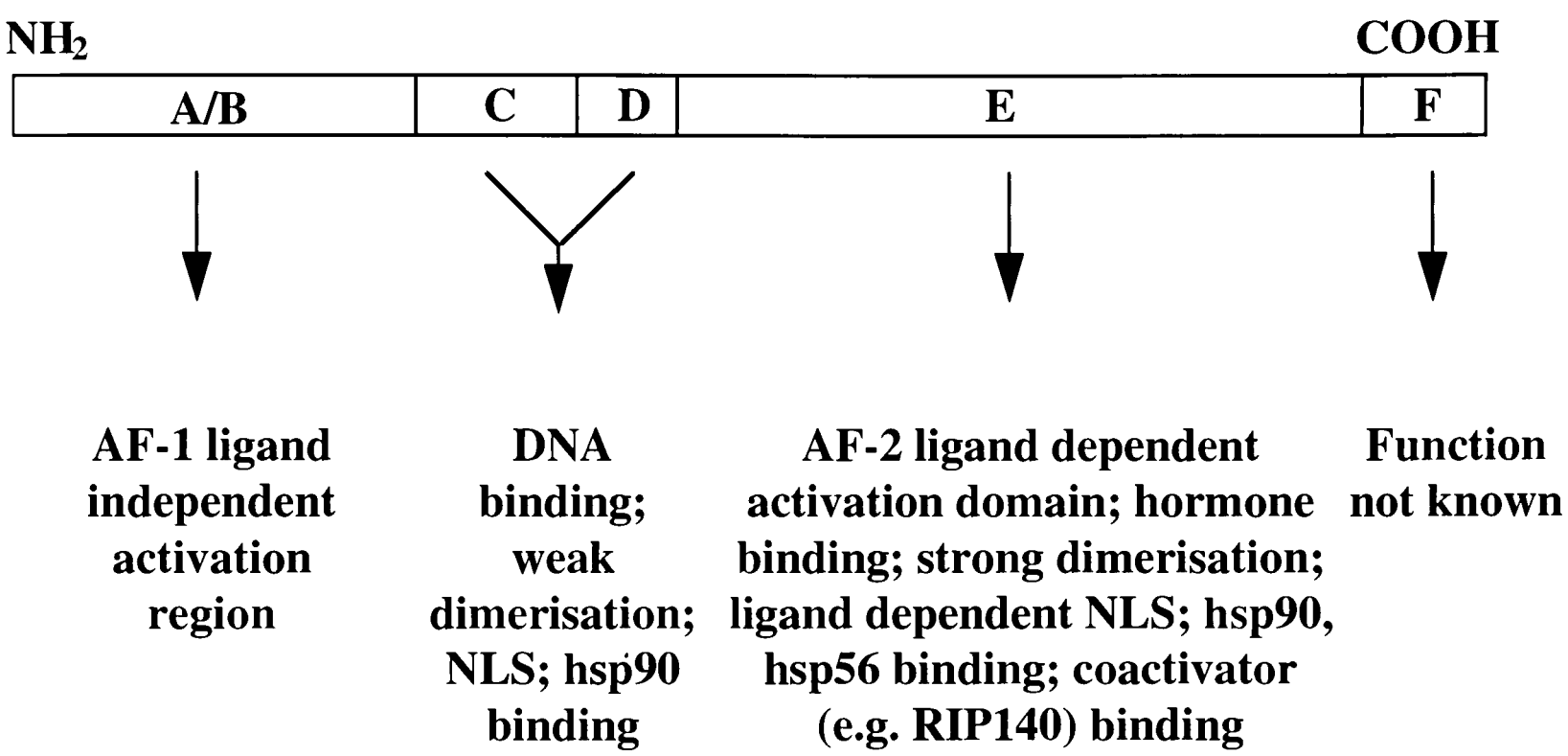


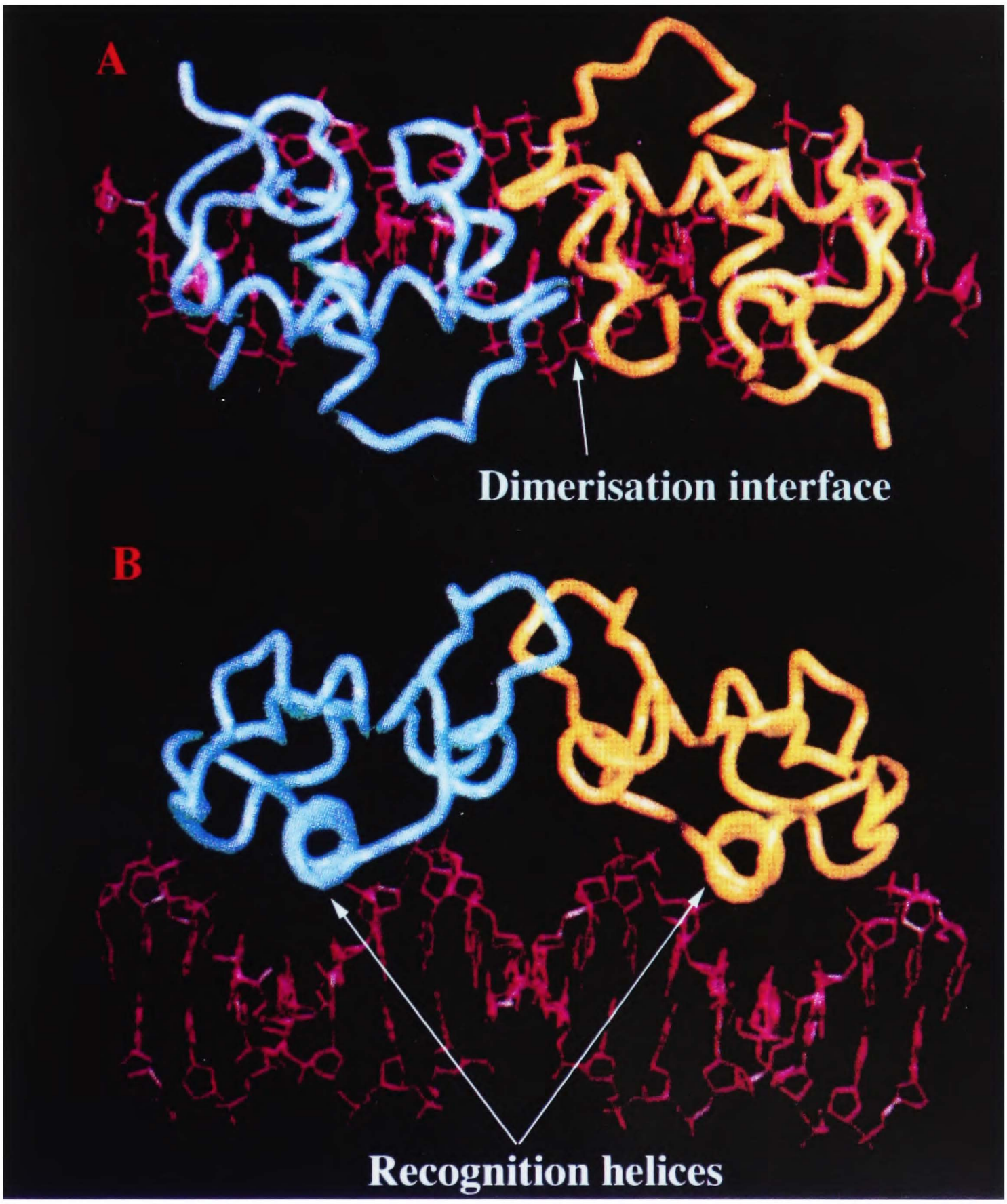
Figure 1.13. Schematic illustration of the structural and functional organisation of the nuclear receptors. Domain C is the DNA binding domain containing the two zinc finger motifs. The regions A/B, D, and F are the most poorly conserved in the family. Domain F is sometimes absent. The functions of each domain are summarised. NLS= nuclear localisation signal.

contain two activation domains, AF1 and AF2 (Nagpal et al., 1993, Danielian et al., 1992). The first of these is hormone-independent and the second is hormone-dependent, and both of these domains from the oestrogen receptor can activate transcription in a promoter and cell-specific manner (Tora et al., 1989, Tzukerman et al., 1994). The most highly conserved region amongst the receptor superfamily is the DNA binding domain. This domain consists of two zinc finger structures (Frankel et al., 1987) in which two zinc atoms are co-ordinated by eight cysteine residues. The zinc fingers in steroid hormone receptors are type II fingers (Harrison, 1991, Klug and Schwabe, 1995). The type I fingers were identified first and co-ordinate the zinc atom with two cysteine residues and two histidine residues per finger. NMR and X-ray structures of the DNA binding domains of the glucocorticoid and estrogen receptors have been solved (Hard et al., 1990a, Schwabe et al., 1990, 1993), and this has improved the understanding of the manner in which these motifs bind to their DNA recognition elements. The oestrogen receptor monomers straddle the major groove of the DNA double helix and make extensive contacts to the phosphate backbone on either side (see Fig 1.14). This arrangement has the effect of orienting the DNA binding domain in such a way that the important sequence recognition elements in the domain enter the major groove, allowing interactions with the DNA (see Klug and Schwabe, 1995 for a review). The other domains in the steroid hormone receptor proteins perform functions such as dimerisation, hormone binding, nuclear localisation, hsp90 and hsp56 binding, and interactions with cofactors and integrator molecules, see Fig 1.13 (see Beato et al., 1995 for review).

1.8.2 Steroid Receptor Associated Proteins (RAPs)

A number of proteins have been found to interact with both liganded and unliganded steroid hormone receptors (see Smith and Toft, 1993 for review). The best characterised interaction is between the chaperone protein hsp90 (Wiech et al., 1992) and the receptors. This protein associates with unliganded receptor and it is the dissociation of this protein from the receptor which is stimulated by the binding of

Figure 1.14. Illustration from Schwabe et al., (1993) showing the interaction of the ER DNA binding domain dimer with its binding site on the DNA. (A) A view through the proteins toward the DNA, revealing the dimerisation interface of the ER DNA binding domains, which includes the D box (see Fig 1.15). (B) A rotation of 90^{o} of view (A) showing the insertion of the recognition helices (which include the P box, see Fig 1.15) into the major groove of the DNA helix.


A
Dimerisation interface
B
Recognition helices

ligand to the receptor in the cytoplasm of the cell. The action of hsp90 is critical to the functioning of the glucocorticoid receptor. Another chaperone protein, hsp 70 (see Gething and Sambrook, 1992 for review), has been shown to interact with progesterone receptors this protein will also dissociate from the receptor when ligand binds, but its dissociation is much slower than that of hsp90. A protein p60 is known to assist in the assembly of the hsp90/hsp70/hormone receptor complex. It is related to both a human stress protein produced by viral transformation and a yeast stress-induced protein (Smith et al., 1993). Its biochemical role is unknown but it is thought that it cooperates with hsp90 and hsp70 in the assembly of the complex. p59 is a member of the immunophilin protein family (also known as hsp56) which is capable of associating with the steroid hormone receptor complex (Tai et al., 1992). It is a heat shock protein and it is thought that this protein serves a molecular chaperone function in the complex. The other RAPs characterised at the moment include p54 (Smith et al., 1990) and p50 (Whitelaw et al., 1991), whose exact identity and function remain unknown, and p23 (Johnson and Toft, 1994), an acidic phosphoprotein of unknown function with a unique sequence (Johnson et al., 1994).

It is possible that the function of these proteins is merely to maintain the shape of the hormone receptor/RAP complex, but other functions could be envisaged including the maintenance of the redox state of cysteines in the steroid receptor (this is known to be important for both zinc and ligand binding), the mediation of the interaction of the hormone receptor with protein kinases and phosphatases, the translocation of the receptor into the nucleus (hsp90 interacts with microtubules and actin filaments, Pratt, 1993, Miyata and Yahara, 1991), the maintenance of inactive receptor complexes, or assistance in the ligand-induced conformational change in the receptor molecule. It has been suggested that the proteins hsp90, hsp70, and p23 associate into a "foldosome", which can mediate the structural changes required to keep the hormone receptor in its inactive, steroid binding state (Hutchison et al., 1994).

1.8.3 The Steroid Hormone Receptor Zinc Finger

The minimum DNA binding domain of a steroid hormone receptor protein consists of 60-70 amino acids. It is well known that this domain, when expressed in *E.coli*, can bind to DNA (Nagai et al., 1988). This domain consists of two zinc binding motifs as mentioned in 1.8.1. These zinc binding motifs are called zinc fingers and were first recognised in the Xenopus transcription factor TFIIIA (Miller et al., 1985). There are now ten different protein motifs which bind zinc, and have been referred to as zinc fingers (see Klug and Schwabe, 1995 for review). Therefore, it appears as though the stabilisation of a protein motif by binding a zinc ion is a common theme in protein evolution.

The zinc fingers of the steroid hormone receptors lie in the domain most highly conserved between the members of the family. The two zinc fingers appear to fold into one structural domain in the hormone receptor molecule and it was thought that the two fingers were essential for the folding of a zinc finger domain until the DNA binding domain of GATA-1 was shown to possess a single zinc finger (Omichinski et al., 1993), homologous to the hormone receptors but which folded correctly. It is more likely that the two fingers exist due to their different functions in the protein rather than a strict structural interdependence.

The structures of the DNA binding domains of two hormone receptors have been solved by NMR (Hard et al., 1990a, Schwabe et al., 1990). Both DNA binding domains appear to be monomeric in solution and two molecules of the DNA binding domain of the glucocorticoid receptor are known to bind cooperatively to its response element sequence (Hard et al., 1990b). In general, it is considered that the first of the zinc finger motifs determines the sequence specificity of the DNA binding and the second allows homo or heterodimerisation of the receptor with other molecules. The site to which the hormone receptors bind is palindromic. The structures of the protein-DNA complexes determined so far for the hormone receptors demonstrate that dimers of the receptors bind to adjacent major grooves on one side of the DNA helix. The receptor also makes contacts to the phosphate backbone of the DNA. The result of these

interactions is that a "recognition helix" in the DNA binding domain enters the major groove (see Fig 1.14). This then allows the side chains of the amino acids in this helix to contact the bases of the DNA. Several amino acids in the first zinc finger are known to determine the sequence specificity of the binding. These amino acids are known as the P box (see Fig 1.15). If the amino acids in the P box of the oestrogen receptor are swapped for those from the P box of the glucocorticoid receptor, then the specificities of the two proteins are altered (Mader et al., 1989). The ease with which the specificity of some of the receptors may be altered has stimulated speculation about a possible protein-DNA recognition code and this speculation has resulted in a set of rules being suggested to explain the code for some types of DNA binding domain (Susuki and Yagi, 1994). These rules are based on chemical rules which govern the type of amino acid-base interaction which can occur in the receptor-DNA complex, in the context of a set of stereochemical rules depending upon the type of DNA binding motif being examined. Such rules may be useful as they could potentially be used to design proteins with new DNA binding affinities (Kisters-Woike et al., 1991). Despite this, it must be noted that the different contacts which the amino acids of the receptors make to the DNA result in a difference in the arrangement of some of the sidechains of other amino acids in the DNA binding domain. Thus, the phosphate ethylation patterns of the estrogen receptor, and the glucocorticoid receptor are different, as the two DNA binding domains make different phosphate contacts on the DNA backbone, thus protecting them (Schwabe et al., 1993). This implies that the alteration of the binding specificity of the steroid hormone receptors does not just involve the slotting of a different amino acid into the P box of the first zinc finger to allow different amino acid-base contacts, but could also involve a significant alteration of other amino acid-DNA contacts in the domain (Schwabe et al., 1993). This type of gross structural change is much harder to predict than the effects of an amino acid substitution in the recognition helix, and may prove to be a limiting factor in the attempts to predict the protein-DNA recognition code. A successful approach to selecting zinc fingers by screening phage libraries with specific oligonucleotides has also been used (Choo and Klug, 1994).

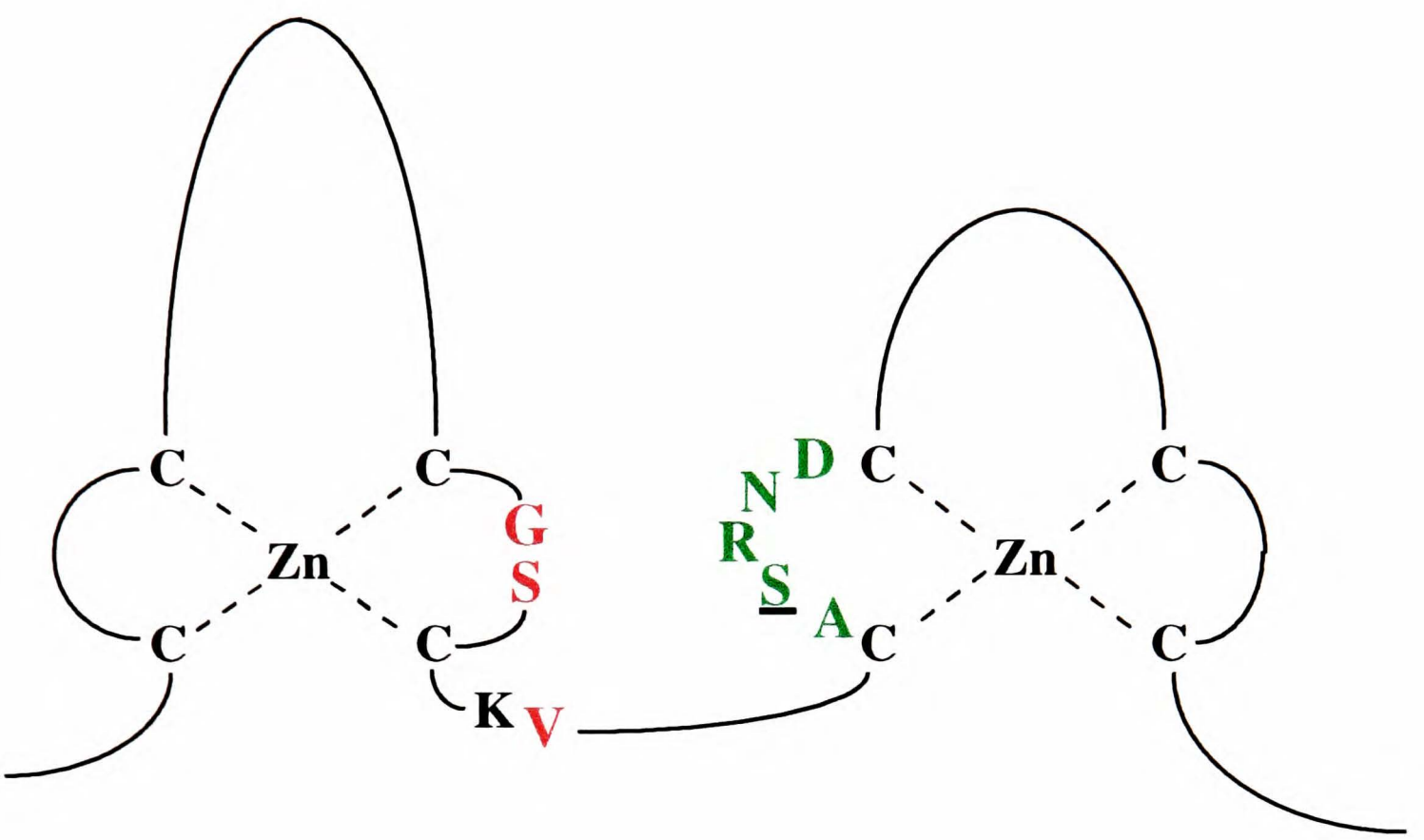


Figure 1.15. A schematic view of the two zinc fingers in the DNA binding domain of the steroid hormone receptors. The domain shown is the human AR DNA binding domain (from Beitel et al., 1993). The amino acids in the P box which are important for determining the sequence specificity of the receptor are shown in red. These amino acids, and several of the adjacent ones, form the recognition helix which enters the major groove of the DNA (see Fig 1.14) upon binding of the receptor to its response element. The amino acids important for the dimerisation of the isolated DNA binding domains (the D box) are in the second zinc finger and are shown in green. These amino acids form part of the dimerisation interface(see Fig 1.14). The only amino acid in the P box, or D box, is specific to the A, and it is underlined.

This approach yielded sequence-specific zinc finger peptides with new DNA binding properties (Choo et al., 1994).

Evidence has been presented which favours a role of the DNA binding domain of some hormone receptors in the modulation of the transactivation domains of the protein (Lefstin et al., 1994). Two point mutations in the DNA binding domain of the glucocorticoid receptor were able to activate transcription, and allow interactions with other activators, without DNA binding. It was suggested that the binding of specific DNA could act as an allosteric activator for the receptor protein, inducing a conformational change in the protein which activates its own transactivation domains and allows it to interact with other activators (e.g. VP16). The two point mutations were thought to mimic this DNA induced conformational change in the receptor structure and make the protein constitutively active.

A group of amino acids in the second finger are known to be concerned with the dimerisation functions of the protein and are known as the D box (see Fig 1.15). These amino acids are not the only dimerisation signals which are found in the protein (see Fig 1.13). Fig 1.15 illustrates that the only amino acid in the androgen receptor which is unique to it in the P box and the D box is a serine residue in the D box. All of the other amino acids in these two important regions can also be found in other members of the steroid hormone receptor superfamily. Deletion of the second zinc finger domain of the androgen receptor leads to total disruption of DNA binding and complete androgen insensitivity (Quigley et al., 1992), highlighting the importance of this domain for the function of the receptor protein.

1.8.4 The Steroid Hormone Receptor Response Element

The response elements of the different hormone receptors are very similar to one another. They are either palindromic sequences or direct repeats of related recognition motifs. Two classes of steroid hormone response elements exist. One class is the estrogen receptor group, which includes the thyroid hormone, the retinoic acid, and the vitamin D receptor. All of these receptors use a similar recognition motif in their

response elements, the sequence TGACCT (5'→3'). The response elements differ, however, in their use of spacer nucleotides between their two recognition motifs, and the arrangement of these motifs (i.e. palindromic, direct repeats etc.). Table 1.3 shows some examples of the types of element which can be bound by this family of receptor proteins, see Gronemyer and Laudet (1995) for full review.

Table 1.3 The difference in the response of DNA elements derived from the common estrogen response element related recognition motif to various steroid hormone receptors (from Gronemeyer, 1992). TR= thyroid hormone receptor, RARα= retinoic acid receptor α, ER= estrogen receptor, VDR= vitamin D receptor, TRβ= thyroid hormone receptor β.

Arrangement of motifs	Spacing	Hormone response
Palindrome	0bp	TR/RARα
Palindrome	1bp	Invertebrate ecdysone
Palindrome	3bp	ER
Everted repeat	3bp	TR, RARα (weakly)
Direct repeat	3bp	VDR
Direct repeat	4bp	TRβ
Direct repeat	5bp	RARα

The other group of response elements which exist is based upon the consensus sequence for the glucocorticoid response element. This sequence is composed of two 6bp asymmetrical elements, forming an imperfect palindrome. These elements are separated by a 3bp spacer. The sequence which allows these hormone receptors to bind to DNA is: GGA/TACANNNTGTTCT (Beato, 1989). A consensus sequence for the androgen receptor was defined in a study by Roche et al., (1992). This study used a gel mobility shift assay with a random pool of oligonucleotides and a protein A-AR DNA binding domain fusion protein, followed by isolation of the oligonucleotides

which bound to the fusion protein, and PCR amplification of the sequences which bound to the fusion protein. Two to four repeats of the procedure produced a group of sequences which could produce a strong retarded band with the protein A-AR DNA binding domain fusion protein. These sequences are shown in Table 1.4.

Table 1.4 The sequences derived by Roche et al., (1992) which bind to the AR *in vitro* in a gel mobility shift assay (from Roche et al., 1992, Table 1C). The position of each base relative to the point of dyad symmetry in the sequences is shown in the extreme left column. The frequency of each base in 24 sequences isolated is shown. The consensus is shown at the extreme right of the table.

Position	G	A	T	C	Consensus
-7	16	4	3	1	G
-6	23	0	1	0	G
-5	4	6	12	2	T
-4	0	19	4	1	A
-3	0	0	0	24	C
-2	7	14	2	1	A
-1	9	3	5	7	N
0	11	6	5	2	N
1	11	10	1	2	G/A
2	0	2	16	6	T
3	24	0	0	0	G
4	1	1	22	0	T
5	5	4	10	2	T
6	0	2	0	22	C
7	1	0	16	7	T

In addition to this consensus the study also identified a preference for A or T nucleotides in the immediate 3' flanking region of the site. The glucocorticoid (GR), progesterone (PR), androgen (AR), and mineralocorticoid (MR) receptors recognise this sequence. Each of the hormone receptors in this group can activate gene transcription through this element in the MMTV (mouse mammary tumour virus) LTR (Arriza et al, 1987, Ham et al., 1988). The ability of each of this group of hormone receptors to bind to this sequence raises a very obvious specificity problem. How does the right receptor activate the right gene? This question cannot be precisely answered, but there are the following possibilities:

(i) cell-restricted expression of the hormone receptors which mediate the response. Therefore, although many hormone response elements may be able to respond to many different hormone receptors *in vitro*, they can only respond to those receptors which are present in the cell *in vivo*. Different cell types may also metabolise the hormone ligand in different ways, ensuring that the hormone receptors in the cell cytoplasm only bind the correct ligand (Katzellenbogen et al., 1996).

(ii) the expression of coactivators or integrator molecules which could be tissue specific and might only interact with particular hormone receptors. This could include interactions with other receptor molecules or other transcription factors. The ligand-dependent activation of some hormone receptors could also lead to the squelching of other receptors if the two types of activated receptors bind a common coactivator or integrator molecule (Kamei et al., 1996).

(iii) the regulation of receptor localisation within the cell. Many transcription factors are known to localise in the nucleus only after they have been phosphorylated (Vandromme et al., 1996). It is well known that many hormone receptors are phosphorylated (Orti et al., 1992), and that the inhibition of phosphorylation can inhibit the trafficking of the protein to the nucleus (DeFranco et al., 1991), so this might be a method of controlling the specificity of the response of certain genes to particular hormone receptors.

(iv) Many genes have been seen to have hormone response elements which do not adhere strictly to the consensus palindromic sequence (Adler et al., 1993, Cleutjens et al., 1996). It is possible that, either by interaction with additional factors binding the local DNA or by multiple hormone receptor-response element interactions, the promoter of a gene may determine the specificity of its response to different hormone receptors.

Some genes have also been identified which have negatively-acting response elements in their promoter regions (Carr and Wong, 1994, Clay et al., 1993, Drouin et al., 1993), but it is not known whether or not repression of transcription by the steroid hormone family of proteins occurs by the same mechanism as activation. The region responsible for the silencing potential of the thyroid hormone receptor is found in the C-terminal ligand binding domain (Tsai and O'Malley, 1994), which has been found to interact with TFIIB (Fondell et al., 1993).

1.8.5 Transcription Activation by Steroid Hormone Receptors

The steroid hormone receptors possess two domains which are known to contribute to the transcription activation function of the molecule. These regions are known as AF-1 and AF-2, see Fig 1.13. AF-1 is the ligand-independent activation region and AF-2 the ligand dependent region. If a receptor is to be maximally active it is thought that both of these domains must be functional in order to synergise in the activation of transcription (Tzukerman et al., 1994). These domains can also function synergistically if expressed as different proteins in a cell (Kraus et al., 1995), highlighting the modular nature of these regions.

The transactivation function of hormone receptors has long been believed to be exclusively dependent upon the binding of ligand to the receptor and the subsequent binding of the receptor-ligand complex to a response element in the promoter of the target gene, which then results in transactivation (see Fig 1.16). However, recent work suggests that the process is more complex. For example, hormone receptors can be activated in a ligand-independent manner also (Bunone et al., 1996, O'Malley et al., 1995), see Fig 1.16. This ligand-independent activation allows the steroid hormone

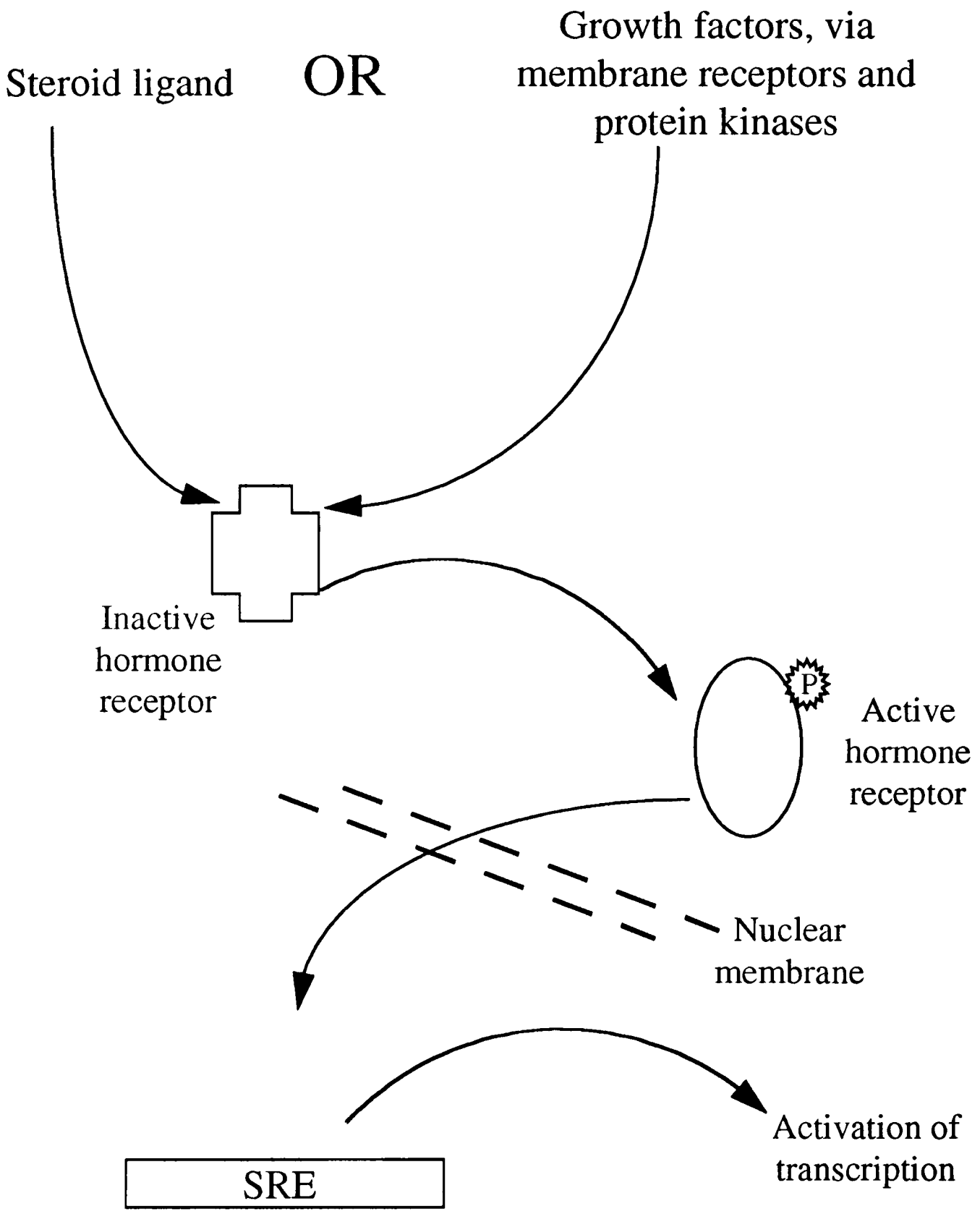


Figure 1.16. A schematic view of the activation of the steroid hormone receptors both by ligand mediated and ligand independent mechanisms. Both of the mechanisms cause the receptor to take on an active conformation and to become phosphorylated (P). The active receptors usually bind to their steroid response elements (SREs) in the genome before they can affect the rate of transcription from a promoter.

receptors to respond to a number of different signal transduction pathways, e.g. the estrogen receptor can be phosphorylated by EGF through the MAP (mitogen-activated protein) kinase pathway. This phosphorylation of serine 118 of the estrogen receptor results in the activation of the hormone receptor. It is not known whether or not the target of the MAP kinase pathway is the naked receptor or the receptor/hsp90 complex. Dopamine is also known to activate the estrogen receptor (Smith et al., 1993) and many other extracellular ligands including insulin, okadaic acid, and vanadate exert an effect on the ER or PR. Interestingly, neither GR, or MR appear to respond to these agonists (Bunone et al., 1996).

Ligand-dependent transcription activation is thought to proceed via protein-protein interactions. The crystal structure of the RAR and TR have been solved (Renaud et al., 1995, Wagner et al., 1995) and they show that hormone binding causes a rearrangement of the hormone binding domain, ordering an amphipathic helix involved in activation. The active hormone receptors can then interact with TFIIB (see Tsai and O'Malley, 1994 for review), and it is possible that TR may interact with TBP (Tsai and O'Malley, 1994) though the latter observation may be an *in vitro* artefact. Thus, a possible mechanism for the action of the receptors is the recruitment of TFIIB into the PIC before transcription initiation. This is one of the rate-determining steps in the assembly of the PIC and its stimulation should increase the rate of transcription. These interactions have all been demonstrated *in vitro*, but their *in vivo* significance remains uncertain. Other proteins can interact with the hormone receptors via the C-terminal AF-2 domain in an agonist-dependent fashion. They include SRC-1, a common coactivator of hormone receptors (Onate et al., 1995), and RIP140 (Cavilles et al., 1995). The importance of the binding of RIP140 to the ER has been demonstrated as ER mutants defective in this interaction, are also incapable of activating transcription (Cavailles et al., 1995). The protein SUG1, which is a component of the RNA polymerase II holoenzyme, also interacts with the AF-2 region of a number of hormone receptors (Mangelsdorf and Evans, 1995). Many other interactions with the AF-2 region of various hormone receptors are known (see Beato et al., 1995 for review) and

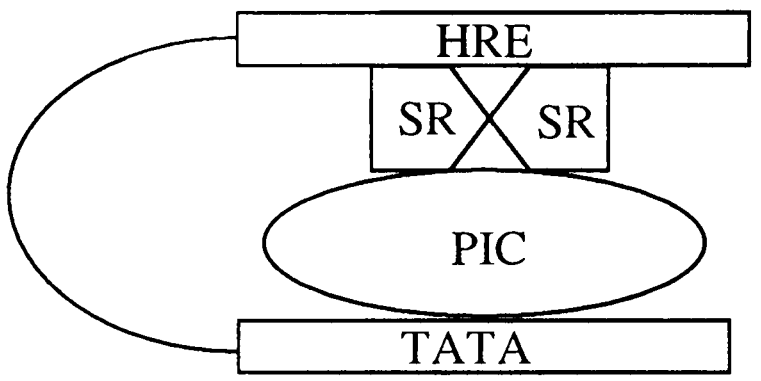
deducing their exact role should lead to a much improved understanding of the precise mechanism of the activation of gene transcription by the hormone receptors.

The binding of the receptor/hormone complex to a response element in a target gene promoter was thought to be a critical event in the activation of transcription by hormone receptors. Again, recent work is beginning to suggest that a hormone/receptor complex can exert its effects upon a target gene indirectly by interacting with the proteins which are bound (or could) bind to the target gene promoter. These types of interactions could provide another way in which the steroid hormone receptors might achieve specificity in their actions (Katzenellenbogen et al., 1996). Three different modes of hormone dependent gene activation by hormone receptors have been suggested (see Fig 1.17, Katzellenbogen et al., 1996). Of these three, one involves the indirect action of a hormone/receptor complex on a target gene promoter by interacting with other factor(s) bound to the promoter. The estrogen receptor has been shown to directly interact with a number of transcription factors including AP-1 (Webb et al., 1995), and possibly Sp1 (Krishnan et al., 1994) in the activation of some genes, though the physiological significance of these interactions is yet to be proved. Repression of gene promoters may also occur by a similar mechanism (Kallio et al., 1995).

1.8.6 The Androgen Receptor

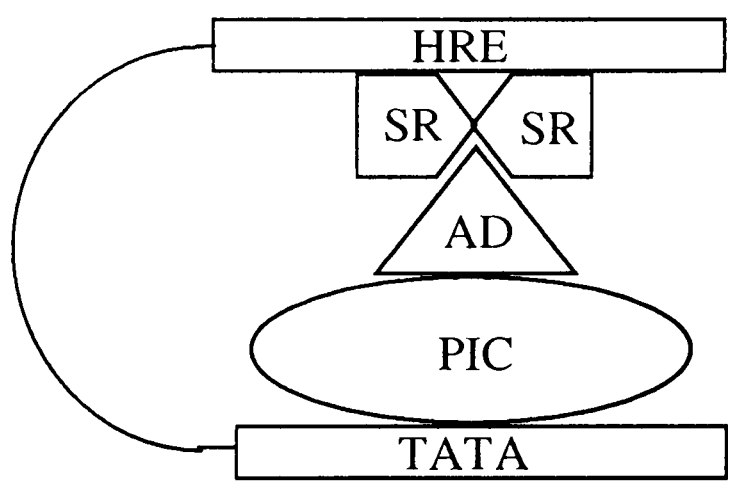
The human androgen receptor gene is found on the X chromosome at position Xq11.2→q12 (Brown et al., 1989). The gene is 90kb long and the protein is encoded by eight exons which produce a 10.6kb mRNA. This mRNA contains both 5' (1.1kb) and 3' (6.8kb) untranslated regions. The open reading frame in the mRNA is 2.7kb long (Faber et al., 1991). The protein which is coded for by the open reading frame is 918 amino acids long and has a molecular weight of 110kDa (Chang et al., 1988a). Two isoforms of the receptor exist, the A form and the B form which differ at the N-terminal region, where the A isoform lacks the normal N-terminus. This reduces the molecular weight of the A isoform to 87kDa. These two forms of the protein are both

A



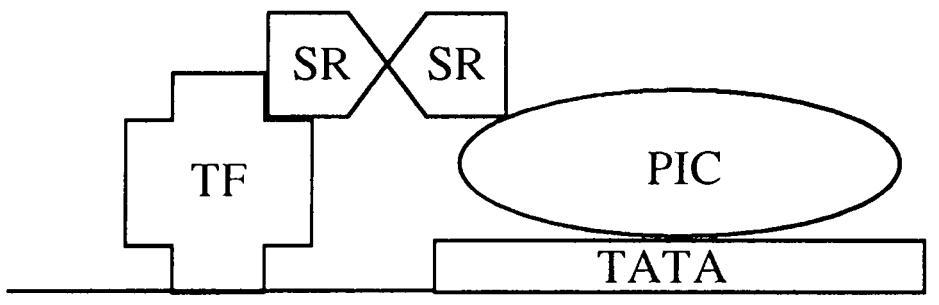
Direct SR/ PIC contact activates transcription after binding to the HRE

B



Indirect contact with the PIC via an adapter molecule after binding to the HRE

C



Direct contact with the PIC mediated by another transcription factor without the requirement for DNA binding

Figure 1.17. Three suggested (Katzenellenbogen et al., 1996) modes of action for the steroid hormone receptors. Each of these modes has different requirements (e.g. DNA binding, interactions with adapters) and each could be stimulated in response to agonists or antagonists, or in response to growth factors acting via protein kinases. HRE= hormone response element; SR= steroid receptor; PIC= preinitiation complex; AD= adapter; TF= transcription factor; TATA= TATA box.

found in genital skin fibroblasts (Wilson and McPhaul, 1994) and are found to be developmentally regulated in Xenopus larynx (Fischer et al., 1993). The A isoform is the minor isoform in normal fibroblasts and is known to be expressed in patients with androgen insensitivity syndrome but the relative function of each isoform remains unclear.

AR is a typical steroid hormone receptor with the typical domain structure Fig 1.13. The androgen receptor binds to ligands such as the physiological androgens testosterone and its analogue 5α-dihydrotestosterone. The receptor has a hormone receptor DNA binding domain arrangement, involving two zinc fingers (Fig 1.15), and binds to the same consensus sequence in DNA as the GR, PR, and MR (Lucas and Granner, 1992, Roche et al., 1992). AR is thought to bind to its DNA element primarily as a homodimer (Wong et al., 1993). Single amino acid substitutions in the D box of the second zinc finger of the DNA binding domain prevent dimerisation, and causes Reifenstein syndrome (partial androgen insensitivity disorders) (Kaspar et al., 1993, Gast et al., 1995). Additional evidence for a critical role for the dimerisation in the activation of transcription by the androgen receptor was obtained by Palvimo et al., (1993) who observed *trans*-dominant repression of AR transactivation by mutant receptors which could dimerise and bind DNA but had the N-terminal domain deleted, this implied that these mutant receptors interfered with the normal receptors by dimerising with them and preventing them from activating transcription. It is one of the least well studied proteins of the hormone receptor family, though recent work is beginning to highlight the diversity of activation mechanisms mediated by AR. The cloning of the gene in a number of species and the expression of both recombinant full length protein (Chang et al., 1992, Xie et al., 1992, De Vos et al., 1994), and fusion proteins (De Vos et al., 1991), has assisted the analysis of the hormone and DNA binding specificities.

The establishment of a database for mutations found in patients with various clinical defects related to mutations in the androgen receptor (Gottlieb et al., 1996) should guide studies of the function of the different domains of the protein, in the same

way that the factor IX database (Giannelli et al., 1996) has provided invaluable resources for delineating the functions of the various domains of the factor IX molecule. In particular, point mutations in the DNA binding domain (Beitel et al., 1994) and deletion of the second zinc finger domain (Quigley et al., 1992) cause androgen insensitivity syndrome, highlighting the importance of a fully functional DNA binding domain. The expansion of a polyglutamine repeat region which starts at amino acid 58 (Perutz et al., 1994) is also of great interest in the study of the transcription activation functions of the protein. This expansion of the repeat causes Kennedy's disease (spinal and bulbar muscular atrophy), which is an inherited, neurodegenerative disease (Perutz et al., 1994). Although no function has yet been ascribed to this polyglutamine repeat it is obviously crucial for the protein to have the correct size of repeat. The disease is caused when the repeat expands from the normal size repeat of 15→31 residues to 40→62 residues. However, an androgen receptor with this expanded polyglutamine repeat region (50 residues) apparently shows no change in its activation potential in some studies (Kallio et al., 1995), though a study by Chamberlain et al., (1994) presented conflicting evidence suggesting that the expansion of the polyglutamine repeat does inhibit the activation of transcription by AR. Some studies (Palvimo et al., 1993, Kallio et al., 1995) have found that the AR possesses two activation domains, one in the N-terminal region of the protein, and the other in the C-terminal region. The C-terminal activation domain has been shown to be ligand dependent (Kallio et al., 1995). The androgen receptor is also apparently capable of inhibiting transcription activation by AP-1 (Kallio et al., 1995) without the requirement for DNA binding, in a ligand-dependent manner, and dependent upon the same N-terminal region which is important for transactivation. When this region is deleted, the truncated receptor then acts as a ligand-dependent activator of AP-1 transcription activation, also without the requirement for DNA binding. However, since it is known that the nuclear receptors can interact with CBP, an integrator protein required for AP-1 transcription activation (Kamei et al., 1996), it has been suggested that this inhibition is due to the competition of both transcription factors for limiting amounts of this crucial

integrator molecule. It is possible that the observations of Kallio et al., (1995) could be due to a similar mechanism, but the specific interaction of CBP with AR has not yet been demonstrated. The AR has also been found to possess three proline-directed phosphorylation sites (Zhou et al., 1995). The inactivation of one of these sites (serine 650 in the hinge region, domain D, of the human androgen receptor) appears to affect the activation of transcription by the receptor.

Despite the definition of a consensus DNA binding site for the receptor (Roche et al., 1992), the androgen response elements in natural target genes deviate considerably from this sequence. The prostate specific antigen (PSA) gene is androgen regulated and the response element which mediates the action of the hormone receptor has recently been characterised (Cleutjens et al., 1996). This response element appears to involve the cooperative action of both a high and a low affinity ARE sequence. Multiple, weak AR binding sites may be able to produce strong responses on genes (see Fig 1.18). The androgen response regions of the Slp gene (Adler et al., 1993), the PBPC3 gene (Celis et al., 1993) and the 20kDa protein (Ho et al., 1993) also show this type of arrangement to respond to AR and show a dependency on the binding of other, non-AR, factors to the gene promoter (see Fig 1.18). The affect of these other proteins may be to relieve inhibitory chromatin structure, or to interact directly with the AR in producing the transcription activation response.

1.9 Tissue Specific Gene Expression and the Factor IX Promoter

Gene expression in eukaryotes can be ubiquitous or confined to distinct tissues i.e. tissue specific. In fact, it is the particular repertoire of genes which a cell activates or represses during its development which will determine the phenotype of that cell and, hence, to which tissue it belongs. Therefore, gene expression must be regulated precisely, e.g. to prevent a cell which should become an adipocyte from becoming a hepatocyte. The control of tissue specific gene expression is centred mainly, but not exclusively, on the transcription of genes. As has been discussed previously, the process of transcription is controlled largely by *trans*-acting factors which bind to *cis*-

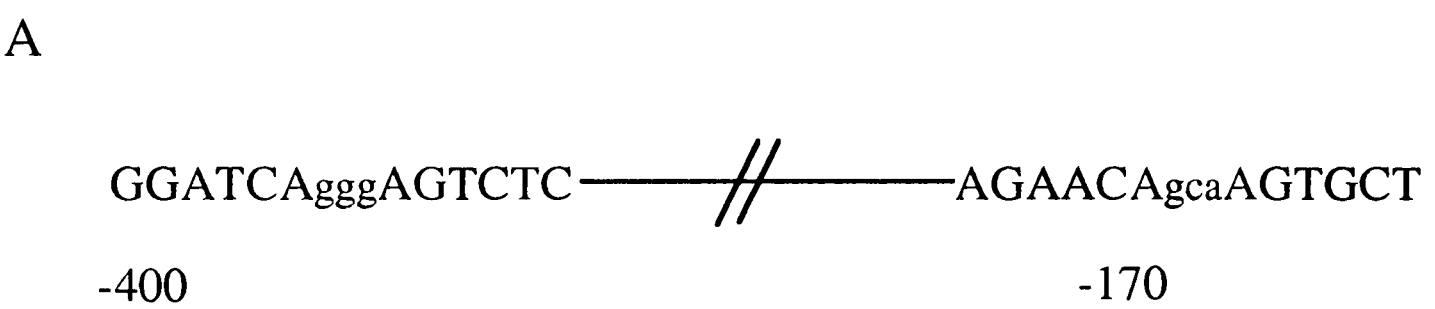


Figure 1.18. The sequences of the regions important for androgen responsiveness in the prostate-specific antigen (PSA), (A), and the sex-limiting protein (Slp), (B) genes. In each case the entire sequence is required for the response to androgens, see Cleutjens et al., 1996, and Adler et al., 1993 respectively. The AREs from the Slp promoter also require a number of as yet unidentified non-receptor factors which bind outside this region for androgen responsiveness.

acting elements in the promoter regions of genes. The simplest way to envisage tissue specificity being regulated is that the promoter regions of tissue specific genes would contain *cis*-acting elements which were recognised only by *trans*-acting elements found in the specific tissue. If the liver, which is one of the most extensively studied tissues in terms of specific gene expression (Lai and Darnell, 1991 for review), is considered this hypothesis does not hold true.

Many of the liver specific gene promoters contain binding sites for liver enriched transcription factors (Crossley and Brownlee, 1990, Greenberg et al., 1995). The difference between liver-specific and liver-enriched also has implications for the regulation of expression of these genes. All of the liver-enriched *trans*-acting factors described to date are not exclusive to the liver, many are expressed in other tissues, e.g. hepatocyte nuclear factor 1 (HNF1) is found also in the lung, kidney and spleen (Xanthopoulos et al., 1991), and hepatocyte nuclear factor 4 (HNF4) is found also in the kidney and intestine (Sladek et al., 1990) . Moreover, the promoters of some liver-specific genes do not contain any binding sites for liver enriched factors (Wang et al., 1994). A number of transcriptional hierarchies and transcriptional cascades have also been found to be involved in the determination of cell fate in hepatocytes (Kuo et al., 1992, Tronche and Yaniv, 1992), see Fig 1.19 and adipocytes (Yeh et al., 1995, MacDougald and Lane, 1995). Therefore, it would appear that many of the tissue-specific transcription factors which have been identified to date are secondary transcriptional messengers controlled by more potent, master genes which have greater control over cell fate. These observations suggest that the regulation of the expression of liver-specific genes is not a simple process, and that it is possible that each gene promoter requires a different combination of liver-enriched and ubiquitous factors to activate transcription. Where no interaction of the promoter with a liver-enriched factor is seen, it is possible that liver-specific coactivator molecules remain to be identified which may confer a liver-specific expression pattern on promoters which bind ubiquitous factors. The mechanism by which cells determine the tissue specificity of a gene is well conserved in evolutionary terms. This is illustrated by the observation that

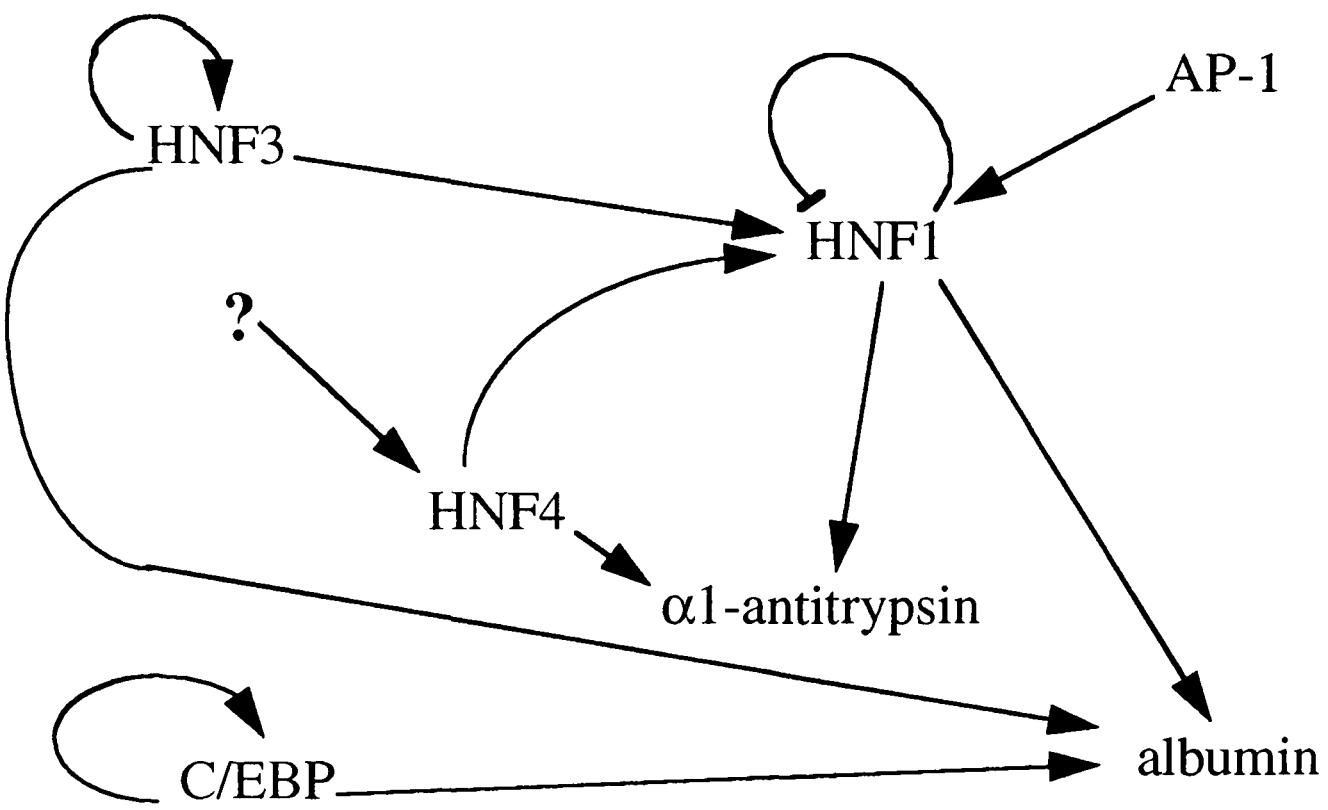


Figure 1.19. A scheme of the known network which exists controlling the expression of the transcription factors HNF1, HNF3, HNF4, and C/EBP, and two liver specific genes, albumin and α1-antitrypsin (adapted from Tronche and Yaniv, 1992). The ? refers to a potential positive developmental regulator of the liver cell phenotype which may upregulate HNF4. The arrows indicate positive regulation and the solid bar represents repression.

human tissue specific promoters can function in mice (Boland et al., 1995) and that rat tissue specific promoters can function in chickens (Cook et al., 1993)

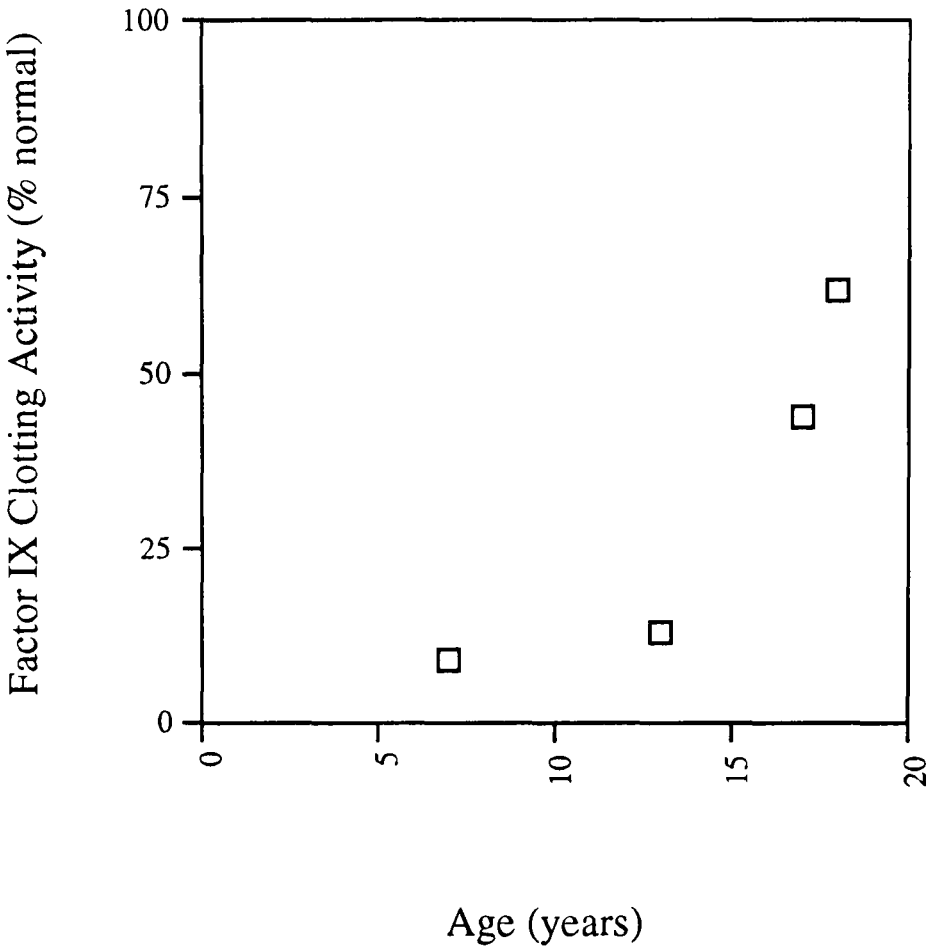
The factor IX gene is expressed in hepatocytes and its transcription has been shown to be highly liver-specific although illegitimate transcription may occur in some cell types (Sarkar and Sommer, 1989, Kaplan et al., 1992). Fig 1.4 shows a schematic representation of the factor IX gene promoter and includes all of the binding sites for transcription factors which have been identified. The factors are a mixture of ubiquitous and liver specific (see Table 1.2 and references therein), which is probably the way in which the tissue-specificity of this gene is controlled. The region of the promoter from -189→+21 has been found to function in a liver-specific manner in transgenic mice (Boland et al., 1995). Several of the factors which bind to this promoter are liver-enriched, HNF4, C/EBP, HLF and DBP. Therefore, it is probable that no one transcription factor confers the tissue specificity upon this promoter region but it is the action of all of them (or perhaps a subset) which controls the liver specific expression. The importance of a number of these binding sites is illustrated by the promoter mutations which have been reported in several patients with haemophilia B (Giannelli et al., 1996, see Fig 1.3). These mutations have been shown to disrupt the binding of one, or more (Crossley et al., 1992), of these transcription factors to the promoter region, thereby reducing the expression of the gene. The large majority of the promoter mutations already identified result in a condition known as the haemophilia B Leyden phenotype.

1.10 Haemophilia B Leyden and Suggested Recovery Mechanisms

1.10.1 The Leyden Phenotype

Haemophilia B Leyden is a form of factor IX deficiency which improves through the course of puberty in patients. The patients are typically severe to moderate haemophiliacs in childhood (factor IX activity <0.01u→0.05u) and the level then rises to 0.4u→0.6u in adulthood and the patient recovers from their haemophilia (see Fig 1.20). The phenotype is heterogeneous, with the exact clotting level of each patient

A



B



Figure 1.20. Clinical data from two patients studied in this thesis, Liverpool 2, (A), and UK232, (B). These data show a mild Leyden type haemophilia B phenotype (A), characterised by the rise in the factor IX levels through puberty, and a Brandenburg type haemophilia B phenotype (B), where no recovery is observed.

depending upon the particular molecular defect. In each of the patients exhibiting this phenotype to date, the detected defect has been a mutation in the promoter region of the factor IX gene (Reitsma et al., 1988, 1989, Giannelli et al., 1996). This is a highly unusual condition, as these promoter mutations confer an altered developmental expression pattern upon the factor IX gene. All of the mutations which cause this phenotype have been detected within a short region spanning nucleotides -21→+13 in this region, and have been shown to directly disrupt the binding of a number of different transcription factors (Crossley and Brownlee, 1990, Reijnen et al., 1992, Crossley et al., 1992, Picketts et al., 1993). The mutations which cause this phenotype effectively delay the expression of factor IX until puberty.

1.10.2 Haemophilia B Brandenburg and UK232

A patient with a -26 G→C mutation in the factor IX promoter region has been described (Crossley et al., 1992) which displayed a different phenotype to the usual Leyden phenotype, in that no improvement in the plasma factor IX level was seen at puberty. This patient was known as haemophilia B Brandenburg (Crossley et al., 1992) and remained a haemophiliac until his death aged 21 (see Fig 1.20). This obviously prevented further study of this patient into adulthood, which would have provided crucial data concerning any delayed recovery of factor IX levels associated with this mutation in the factor IX gene. More recently another patient, UK232 (Giannelli et al., 1996), was identified (see Chapter 6) with a G→A mutation at -26 in the promoter region of the factor IX gene. This patient is aged 47 years and is still a severe haemophiliac. This vital clinical data proves that mutations at nucleotide -26 in the factor IX gene promoter prevent any Leyden recovery of factor IX levels in adulthood (see Chapter 6). A number of theories have been proposed to explain the Leyden phenotype and the Brandenburg patient data which will now be discussed (see 1.10.3→1.10.6).

1.10.3 The Overlapping Androgen Response Element

Crossley et al., (1992), proposed that the Leyden recovery was mediated by an androgen response element (ARE) which is present in the factor IX promoter (nucleotides -36→-22). This ARE is normally masked by the binding of the transcription factor HNF4 to a site which overlaps it (-27→-15). When a mutation occurs which prevents the transcription factor HNF4 from binding, the ARE is "unmasked" and is now available to bind to AR (see Fig 1.21). In the Leyden haemophiliacs, AR is not active until testosterone levels rise at puberty, and so significant expression of the factor IX gene is delayed until this time. The mutation found at nucleotide -26 in haemophilia B Brandenburg is in a region of overlap between the ARE and the HNF4 binding site. It was found to disrupt the binding of both AR and HNF4 and so it was able to disrupt the expression of the factor IX gene both in childhood and in adulthood.

This theory explains the Leyden phenotype if the mutations in the HNF4 site are considered. It becomes more difficult to explain the role of the ARE in mediating a recovery of factor IX expression in Leyden haemophiliacs with mutations in other, non-overlapping, transcription factor binding sites (Giannelli et al., 1996) such as the C/EBP site and the HNF4 site spanning nucleotides -15→+3. To explain these mutations, the theory would have to be extended to include possible ways in which the disruption of the binding of any transcription factor to the promoter results in a defective promoter which can be rescued by the action of AR, either through the ARE or by some other mechanism. As the AR is not an extensively studied protein, and the full repertoire of proteins with which it can interact has not yet been characterised, this extension of the theory remains speculation (see Chapter 8).

1.10.4 The Role of HNF4

Reijnen et al., (1993) observed that the G→C mutation at -26 disrupts the binding of HNF4 to a greater extent than other mutations in the HNF4 site such as the -21 T→G mutation which causes haemophilia B Leyden. They proposed that some

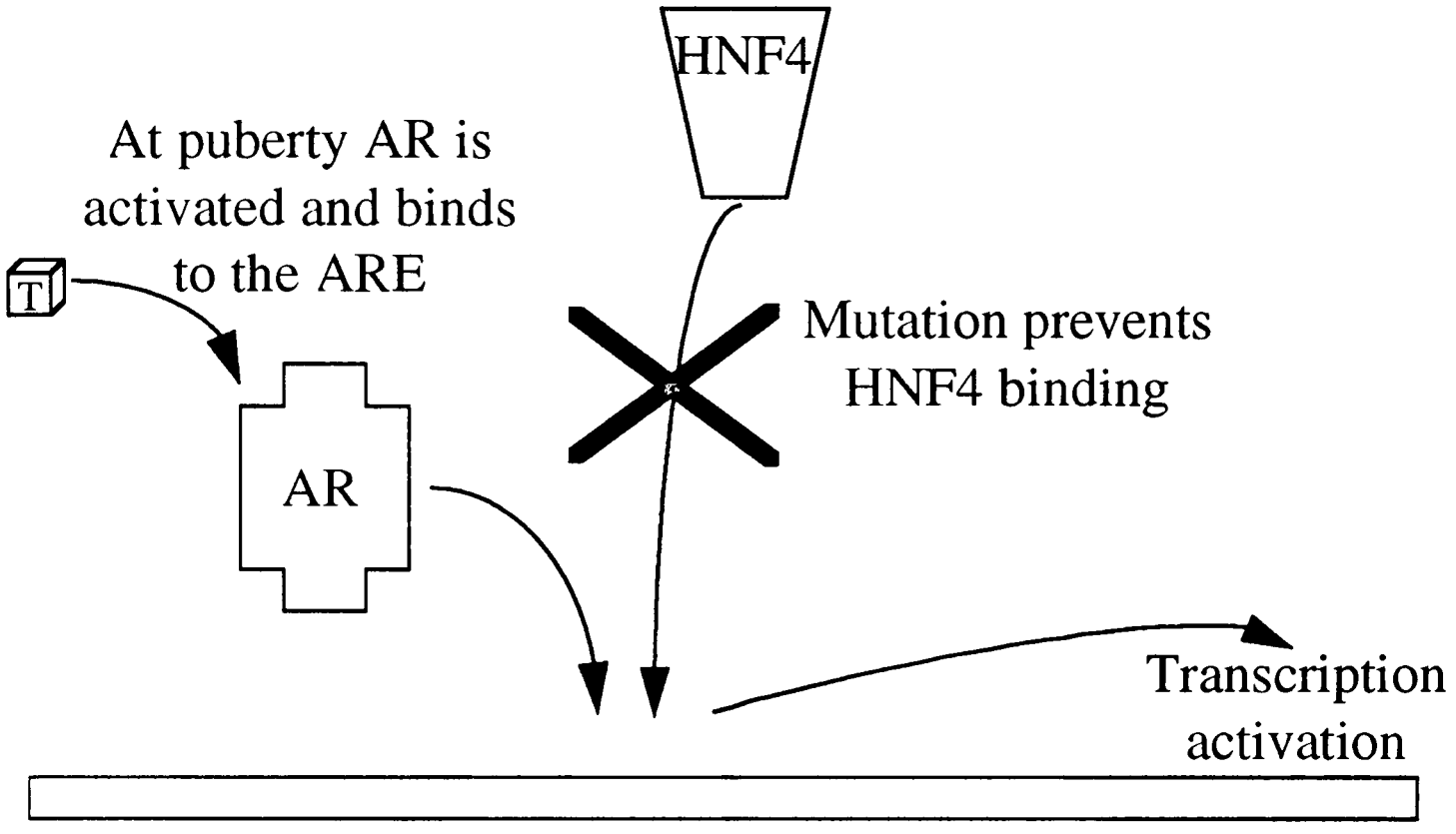


Figure 1.21. The mechanism of the Leyden recovery as described in Crossley et al., (1992). The mutation in the HNF4 site prevents HNF4 binding to the factor IX promoter. When testosterone (T) levels rise at puberty, the AR is activated and it can then bind to the ARE between nucleotides -36 and -22 (overlapping the HNF4 site, see Fig 1.4). This upregulates the promoter and allows expression of the factor IX gene, resulting in an increase in the plasma level of factor IX.

residual binding of HNF4 is required for the Leyden recovery and that it was the complete disruption of the binding of this factor which resulted in the failure of haemophilia B Brandenburg to recover (see Fig 1.22). They also propose that as both HNF4 and the AR are members of the nuclear receptor superfamily, then this might in some way allow them to heterodimerise on the factor IX promoter and activate transcription during puberty in the Leyden haemophiliacs.

A number of reports cast doubt upon this theory. For an ARE mediated mechanism to be operating, the AR must have access to its binding site. If any residual HNF4 binding activity in the Leyden-type factor IX gene is present then the AR would be denied access to the ARE. In such a situation, AR would have to interact with HNF4 if it was to exert any influence on the rate of transcription from the factor IX promoter. HNF4 is known to exist exclusively as homodimers in the nucleus (Jiang et al., 1995) and does not even interact with the retinoid X receptor α (RXRα) with which it shares significant homology. It was also shown not to interact with other related members of the nuclear receptor superfamily including RXRβ, RXRγ, retinoic acid receptor α (RARα), and thyroid hormone receptor α (TRα). This inability to heterodimerise was found to be due to a strong homodimerisation property which prevented any other interactions occurring. It is therefore doubtful whether any heterodimeric complex containing HNF4 and AR would form.

1.10.5 Site 5 and the PAR Domain Proteins

Boccia et al., (1996) describe a complex interplay of factors which occurs at a site which they term site 5 in the factor IX promoter (nucleotides -220→-195).They propose a mechanism which could function in addition to the ARE-mediated recovery of the factor IX gene transcription. They observe that the proteins C/EBPα and GABPα/β are able to bind directly to site 5 in the factor IX promoter (see Fig 1.4) and that maximal transcription activity is not achieved from the promoter unless both factors are bound. The liver enriched PAR domain proteins, DBP and HLF, are then able to further increase the transcription from this promoter by interacting with the

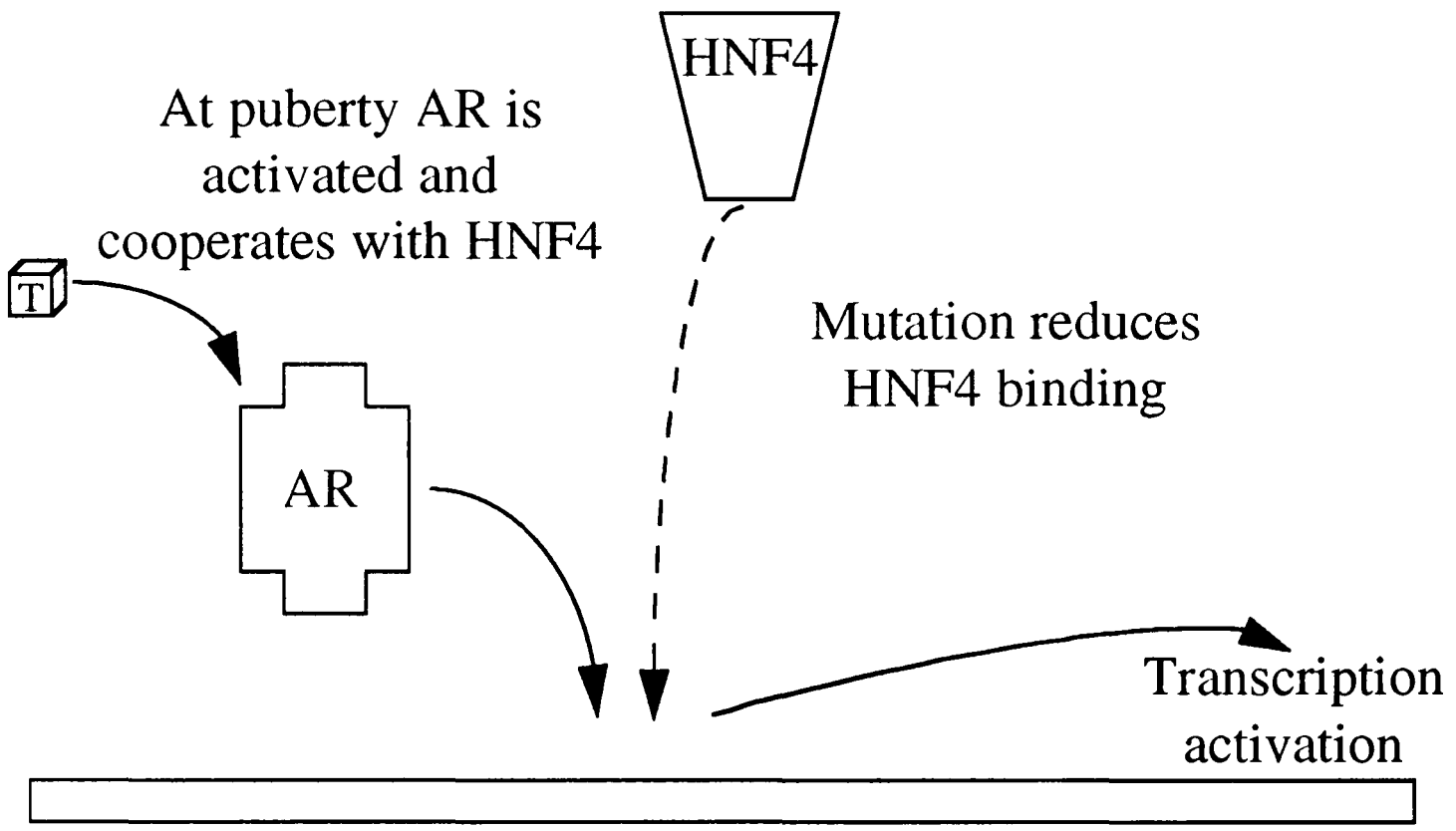


Figure 1.22. The mechanism of the Leyden recovery as described in Reijnen et al., (1993). The mutation in the HNF4 site impairs HNF4 binding to the factor IX promoter, but some residual HNF4 binding remains and when AR is activated it cooperates with HNF4 and restores expression of the factor IX gene.

C/EBPα/GABPα/β complex without binding to the DNA (see Fig 1.23). DBP is developmentally regulated in the rat and its levels increase dramatically around the time of puberty. The action of the proposed DBP/HLF heterodimer was seen to be dependent on the presence of GABPα/β binding to the promoter.

The above theory is attractive as it invokes the action of another transcription factor (DBP), which appears to be upregulated at puberty, in the Leyden recovery. It also provides an additional mechanism which, in theory, could operate alongside the ARE mechanism and account for all of the Leyden mutations, provided all that was needed was one or the other mechanism to be functional. However, this theory does not explain the haemophilia B Brandenburg phenotype, as this recovery mechanism should not be affected by the disruption of the ARE. The most problematic data, which has been reported concerning this mechanism, was the construction of a transgenic mouse bearing a human factor IX Leyden gene promoter fused to a CAT reporter gene (Boland et al., 1995). This promoter conferred tissue-specific expression on the reporter gene which was sex-specific and age-related, mimicking the expression of the factor IX gene in humans with this Leyden-type promoter. The normal factor IX transgenic mouse displayed a normal tissue-specific distribution of transgene expression. The promoter fragment used in the construction of the transgene extended from nucleotide -189→+21. This fragment does not include site 5 of Boccia et al., (1996), and yet still shows a highly specific Leyden type gene activation capability. This proves that site 5 is not crucial for the Leyden type recovery.

1.10.6 A Possible Indirect Role for AR?

Kurachi et al., (1994) describe a footprinting study with HepG2 and baby hamster kidney (BHK) cell extracts which identified a number of footprints not previously described on the factor IX promoter. One of these footprints spanned nucleotides -34→-23, and it was suggested that this footprint was caused by an unidentified, ubiquitous nuclear protein, capable of interacting with AR. They propose that it is the binding of this unidentified protein, not AR, to the factor IX promoter that

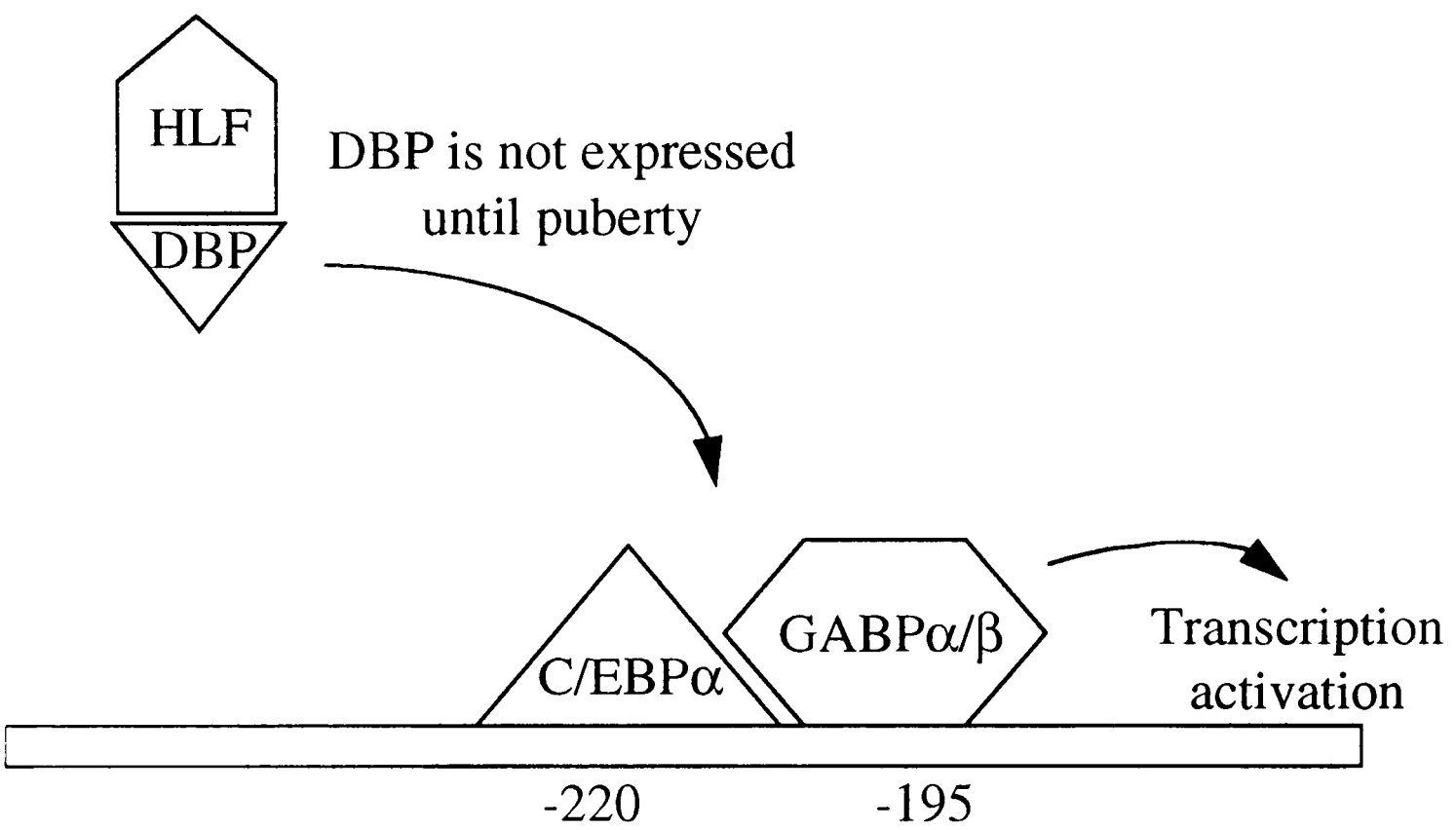


Figure 1.23. The mechanism of the Leyden recovery as described in Boccia et al., (1996). A region of the factor IX promoter between nucleotides -220 and -195 binds the protein GABPα/β and C/EBPα. However, this complex cannot activate the factor IX promoter until it interacts with the PAR domain protein heterodimer formed between HLF (hepatic leukaemia factor) and DBP (D site binding protein). DBP is thought to be upregulated at puberty, so until this protein is expressed, the promoter remains inactive.

is the critical step in the recovery of the factor IX Leyden genes. The interaction of this protein would then allow the AR to interact via protein-protein interactions at puberty, thus allowing the activation of the transcription of the gene (see Fig 1.24). This theory, though possible, requires the full characterisation of the unidentified protein binding to the -34→-23 site, and the definition of the proteins with which it is capable of interacting to activate transcription. However, there is increasing evidence that transcription factors can exert affects on transcription without binding to DNA (Kallio et al., 1996), and the haemophilia B Brandenburg mutation at nucleotide -26 did appear to disrupt the binding of this unidentified protein to the factor IX promoter (Kurachi et al., 1994), so the possibility of AR acting through another protein on this promoter should not be dismissed.

1.11 Aims of This Thesis

The main aim of this thesis was to provide further evidence for the importance of the androgen response element in the factor IX promoter region, originally identified by Crossley et al., (1992), in the recovery of factor IX levels in Leyden haemophiliacs. It was hoped that this study would also provide new information concerning the sequences which can bind AR, both *in vitro* and *in vivo*, and lead to a better definition of a consensus DNA binding sequence for this protein. It was also intended that the mutagenesis study of the region of overlap of the ARE and the HNF4 binding site should identify positions which disrupt both the ARE and the HNF4 site, which would allow one to predict the phenotype of any patient with a mutation in this region of the promoter. This type of information could be of importance when offering genetic counselling to members of affected pedigrees (see Chapter 8). The haplotype study in Chapter 7 was intended to improve the existing knowledge concerning the rate of mutation at CpG dinucleotides in the factor IX gene and the establishment of founder effects in the haemophilia B population.

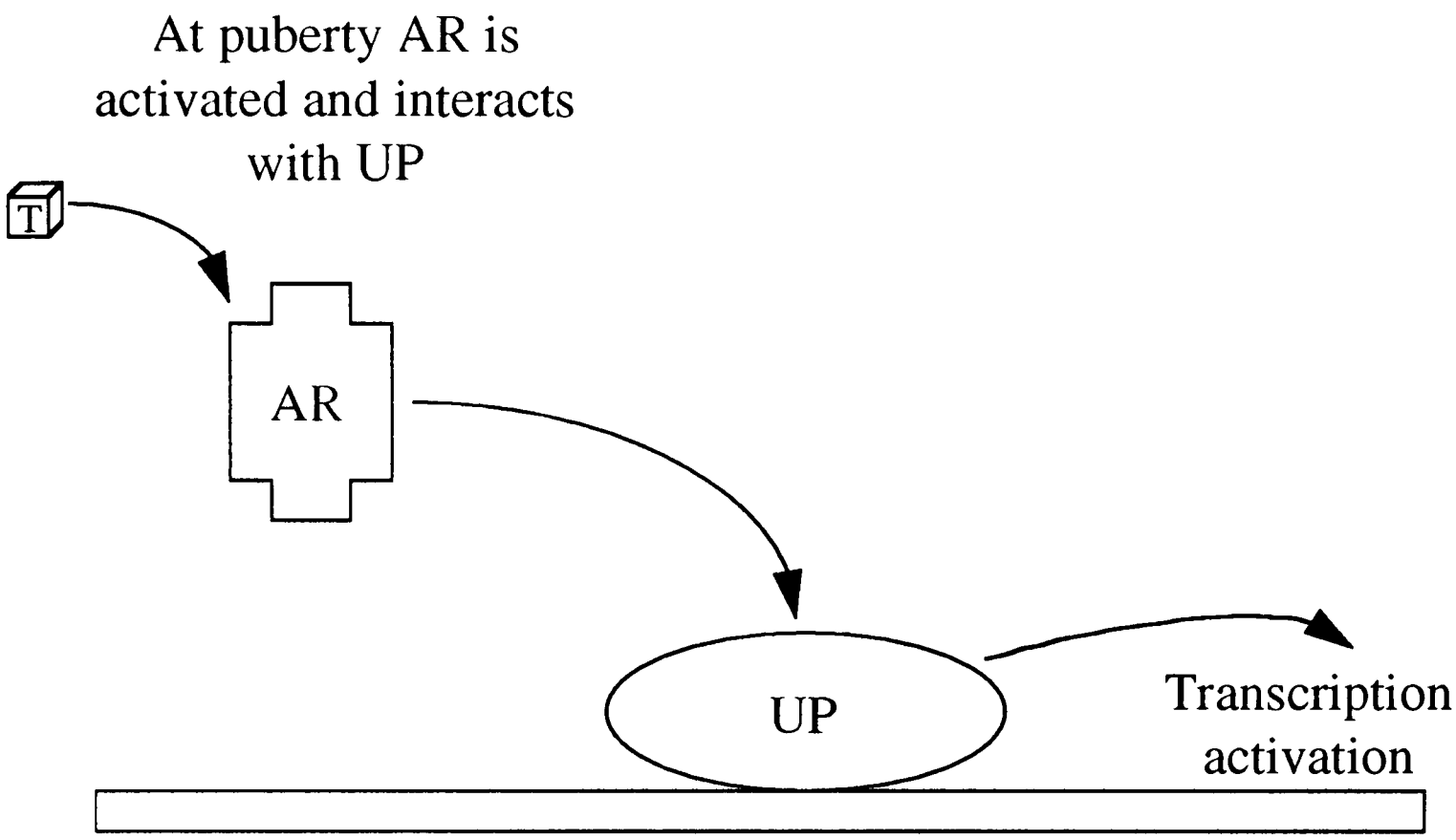


Figure 1.24. The mechanism of the Leyden recovery as described in Kurachi et al., (1994). The mutation in the HNF4 site prevents HNF4 binding to the factor IX promoter, and an as yet unidentified protein (UP) binds to a site between nucleotides -34 and -23. When AR is activated at puberty, it interacts with this unidentified protein and upregulates the factor IX promoter, possibly without the requirement for DNA binding.

Chapter 2 Materials and Methods

All general chemicals and enzymes were obtained only from reputable sources such as Sigma, BDH, Gibco, and NEB. All radionucleotides and dideoxy sequencing reagents were obtained from Amersham/USB. All mammalian cell culture medium and foetal calf serum was obtained from Gibco and Sigma, and the hormone used in the transactivation experiments was obtained from Sigma.

2.1 Genotypes of *E.coli* Bacterial Strains

XL-1 Blue — *end*A1, *hsd* R17, (r_k-,m_k+), *sup*E44, *thi*-, λ-, *rec*A1, *gyr*A96, *rel* A1, Δ(*lac*), [F', *pro* AB, *lac* I^q, *lac* Z Δ M15, Tn 10 Tc]

NM554 — *ara* D139 (*ara*, *leu*) 7697, *str* A, Δ *lac* x74, *gal* U, *gal* K, *hsv* $^-$, *hsm* $^+$, *sup* ^{o} *rec* AB

DL795 — K12 SH28 Δ*hsd*, *mcr*BC, *mrr*; *e*14 (*mcr*A)^o; *sbc*C201, *rec*A::Cm^R; *sup*E44

JC8679 — K12 AB1157 *rec*BC, *sbc*A

SURE — *mcr*A, Δ(*mcr*BC-*hsd*RMS-*mrr*)171, *sup*E44, *thi*-1, λ-, *gyr*A96, *rel*A1, *lac*, *rec*B, *rec*J, *sbc*C, *umu*C::Tn5 (*kan*^{R}), *uvr*C, [F', *pro*AB, *lac*I^qZΔM15, Tn10, (*tet*^{R})]

All of the bacterial strains listed above were grown in 2xTY medium (16g bacto-tryptone, 10g bacto-yeast extract, 5g NaCl in one litre, the pH adjusted to 7.0 with NaOH and sterilised by autoclaving), and stored at -70^{o}C in 15% glycerol.

2.2 Preparation of Competent *E.coli* for Transformations

500mls of 2xTY was inoculated with 1ml of an overnight bacterial culture. This was grown at 37^{o}C until the optical density (OD) of the culture was 0.2-0.4. The cells were then harvested by centrifugation at 5000 rpm for 5 minutes. The cells were then resuspended in 100mls 50mM $CaCl_2$ and incubated on ice for 20 minutes. The cells

were then pelleted, again by centrifugation, and resuspended in 20mls transformation buffer (50mM $CaCl_2$, 15% glycerol, 10mM PIPES pH 6.6). The cells were then aliquoted into microfuge tubes and stored at -70°C.

To transform these cells, 100µl aliquots were incubated with the DNA sample (1→10ng plasmid DNA, or 50% of a ligation reaction, see 2.9.2) on ice for 15 minutes. The cells were then subjected to heatshock at 42°C for 2 minutes, an equal volume of 2xTY was added to the cells, and then they were returned to 37°C for 45 minutes. At the end of this incubation, the cells were then plated onto pre-warmed agar plates and incubated at 37°C overnight in the presence of the appropriate antibiotic (concentrations used were as Maniatis et al., 1982).

2.3 Large Scale Plasmid DNA Preparations

500mls bacterial cultures were grown until saturated(typically inoculate 500mls medium with 500µl of saturated culture and grow overnight). The cells were harvested by centrifugation at 5000rpm in a Beckman J2-21 centrifuge and resuspended in 10mls buffer P1 (50mM glucose, 10mM EDTA, 25mM Tris-Cl pH 8.0). The cultures were incubated on ice for 10 minutes and then the cells were lysed by the addition of 20mls buffer P2 (1% SDS, 0.2M NaOH, freshly made). The cultures were then mixed by inversion and returned to ice for a further 10 minutes. The lysis was stopped by the addition of 15mls of buffer P3 (3M KCH_3COO, 1.8M CH_3COOH). The salt and chromosomal DNA which precipitated was then pelleted by centrifugation at 18000rpm in a Beckman J2-21 for 20 minutes. The nucleic acid still in solution was then precipitated from the supernatant by the addition of 0.6 volumes of isopropanol. The mixtures were left at room temperature for 30 minutes before pelleting the nucleic acid at 10000rpm in a Beckman J2-21 centrifuge. The pellet was air dried and dissolved in 10mls of TE (10mM Tris-Cl, 1mM EDTA pH 7.5).

100µl of a 10mg/ml solution of ethidium bromide was then added to visualise the nucleic acid and caesium chloride solid was added to the solution until the density was between 1.6 and 1.65g/ml. The resulting solution was then spun at 55000rpm

overnight or 77000rpm for 4 hours in a Beckman L8-M ultracentrifuge to set up the caesium chloride gradient and separate the plasmid DNA from RNA and chromosomal DNA.

The plasmid DNA was identified in the ultracentrifuge tube and taken off with a 19 gauge needle. The ethidium bromide was extracted from the solution by repeated mixing with butanol saturated with water. This process was repeated until the butanol showed no trace of ethidium bromide. 4 volumes of water were then added to the solution and then 2 total volumes of ethanol added to precipitate the plasmid DNA.

2.4 Small Scale Preparations of Plasmid DNA

1ml of a saturated bacterial culture was placed into a 1.5 ml microfuge tube and spun in a bench microcentrifuge to pellet the cells. The bacteria were then resuspended in 150µl of buffer P1 and left on ice for 10 minutes. The cells were then lysed by the addition of 300µls buffer P2 and mixed by inversion before returning to ice for 10 minutes. The lysine was terminated by the addition of 225µls buffer P3. The sample was then spun in a bench microcentrifuge to pellet the salt and chromosomal DNA. 0.5mls of the supernatant was taken off into a fresh microfuge tube and an equal volume of isopropanol added to precipitate the nucleic acid. The pellets were then air dried and resuspended in 50µls TE.

If the small scale preparation was to be sequenced directly then 10µls of 5M NaCl and 60µls 13% PEG 8000 were added to the sample to precipitate the DNA. The sample was incubated on ice for 20 minutes and then the DNA pelleted by centrifugation at 4°C in a bench microcentrifuge. The resulting pellet was washed in 70% ethanol, air dried and resuspended in 20µls TE.

2.5 Electrophoresis of DNA

Fragments of DNA up to 500bp were analysed using polyacrylamide gel systems of 5%, or higher, with 1xTBE (89mM Tris-Cl, 89mM boric acid, 2.5mM EDTA) as the electrophoresis buffer. Fragments larger than 500bp were analysed on

agarose gel systems of between 0.6%→3% in 1xTAE (40mM Tris-acetate, 2.5mM EDTA, 1.14mls glacial acetic acid in 1 litre). Samples were typically loaded onto these gel systems in a loading dye consisting of 8% ficoll, 0.05% bromophenol blue, 0.05% xylene cyanol, the samples and the dye were mixed in a 1:1 ratio before loading. After electrophoresis for the appropriate time (depending upon the size of the fragment being analysed), the gels were stained in a solution of ethidium bromide (0.5μg/ml), and the DNA visualised on a UV transilluminator.

DNA was purified from these gels systems in two ways:

1. DNA was extracted from polyacrylamide gels by visualising the DNA band of interest, cutting out the gel slice corresponding to the band, crushing it, and then eluting the DNA from the gel in 3 volumes of 0.5M NH_4CH_3COO, 1mM EDTA, 1% SDS at 60°C for at least 3 hours. The gel debris was then spun down in a microcentrifuge and the supernatant containing the DNA removed. The DNA was then recovered form the supernatant by ethanol precipitation and 5% of the solution containing the DNA was then analysed by gel electrophoresis to determine the yield of the fragment.

2. DNA was purified from agarose gels by using the QAIEX gel extraction kit (QIAGEN Ltd). This purification is based upon the selective adsorption of DNA onto silica gel particles in the presence of a chaotropic salt.

2.6 Plasmid Constructs Used in this Thesis

2.6.1 Cloning Vectors

pBluescript KS+ - Stratagene.

2.6.2 Expression Vectors

pKC3AR- expression vector containing the full length rat AR cDNA under the control of the Simian virus (SV) 40 early promoter, Chang et al., (1988b).

pCB6HNF4- expression vector containing the rat HNF4 cDNA under the control of a cytomeglovirus (CMV) promoter, provided by Dr. J. Talliandis.

2.6.3 Reporter Gene Vectors

pGL2Basic- promoterless luciferase expression vector, Promega.

pGL2Control- luciferase vector with SV40 early promoter and enhancer, Promega.

pGLtk- luciferase expression vector with a Herpes simplex thymidine kinase promoter (same fragment as pBLCAT2) cloned into the *Bgl*II site of pGL2Basic, see Chapter 5.

pGLMIN- luciferase expression vector with the factor IX minimal promoter region, -220→+43, cloned into the *Sst*I and *Hin*dIII sites of pGL2Basic.

pBLCAT2- chloramphenicol acetyl transferase expression vector with the Herpes simplex virus thymidine kinase promoter, Luckow and Schutz, (1987).

pCH110- β-galactosidase expression vector, Pharmacia, Herbomel et al., (1984).

pIRV- β-galactosidase expression vector, Beddington et al., (1989).

2.7 Restriction Digests

2.7.1 Restriction Digests and Markers

Restriction digests were always carried out using the enzyme manufacturers' recommended buffers and conditions. The DNA size markers used on the gels to analyse restriction digest were:

Agarose gels: 1kb ladder (Gibco BRL), 506bp→12216bp.

Acrylamide gels: ϕx174 *Hin*fI (Promega), 24bp→726bp.

2.7.2 PCR Based Analysis of Polymorphic Loci

Taq DNA polymerase was used to amplify the polymorphic loci from several different patient genomic DNA samples. In each case a Perkin Elmer Genamp 9600 system was used and the conditions used were as reported in Stowell et al. (1993); Winship et al., (1993). The sequences of the oligonucleotides used are shown in Appendix 1. The buffer conditions used for the restriction enzyme digests are shown in Table 2.1. No digest was required for the variable size *Dde*I polymorphism, the details of the sizes of the PCR fragments are shown in Table 7.1.

Table 2.1 Buffer conditions for the restriction enzyme digests of the factor IX polymorphic loci PCR fragments. All digests were performed at for 1 hour unless

stated. The final concentration of the acetylated bovine serum albumin (BSA) used in the digests was 1µg/µl. See Appendix 1 for the primers used for the PCR of each locus.

Enzyme	Buffer
*Xmn*I	10mM Tris-Cl pH 7.9, 50mM NaCl, 10mM $MgCl_2$, 1mM DTT + acetylated BSA
*Taq*I	React buffer 2 (Gibco), 65°C, 1 hour.
*Hha*I	NEB buffer 4 + acetylated BSA
*Mnl*I	NEB buffer 2 + acetylated BSA
*Bam*HI	As *Xmn*I but [NaCl]= 150mM
*Msp*I	NEB buffer 2
*Mse*I	NEB buffer 2 + acetylated BSA

2.8 Statistical Analysis of the Data from the Haplotype Study

2.8.1 The Binomial Distribution

For the set of nine patients with -6G→A mutation studied, assuming that the probability of finding the -6G→A mutation in each of the haplotypes I→IV is equal, the probability of finding any number patients with the mutation in haplotype I can be calculated as follows:

The probability of finding A patients (P_A) with haplotype I in a sample size of N, where the frequency of haplotype I is X can be expressed using the binomial distribution.

$$P_A = \frac{N!}{A!\,(N-A)!} (X)^A.(1-X)^{N-A}$$

2.8.2 χ^2 Goodness of Fit Test (Banic, 1975)

The χ^2 test examines the Null Hypothesis that there is no significant difference between groups of data. The data analysed in 7.4 came from small samples and so the

χ^2 test had to be performed incorporating Yates's correction for small sample size. The equation used to determine the value of χ^2 for these two sets of data is shown below.

Table 2.2 The data can be organised into a 2x2 contingency table in the following way:

Group	Haplotype I	Not Haplotype I	Total
Control	A	B	A+B
Experimental	C	D	C+D
Total	A+C	B+D	N

then, $$\chi^2 = \frac{N(|AD-BC|-0.5N)^2}{(A+C)(B+D)(A+B)(C+D)}$$

and for the 2x2 contingency table above, the degrees of freedom = 1.

2.8.3 The student's t test

The students t test was used to examine whether or not the difference between two experimentally derived means was significant or simply due to chance. The t statistic was calculated using the following equations:

$$t = \frac{(m_2 - m_1)\sqrt{N}}{\sqrt{(s_1^2+s_2^2)}}$$

where m_2= the mean of data set 2; m_1= the mean of data set 1; s_1= the standard deviation of data set 1 and s_2= the standard deviation of data set 2. If the numbers in the groups being studied were different, then the equation was modified to:

$$t = \frac{(m_2-m_1)}{s\sqrt{(n_1^{-1}+n_2^{-1})}}$$

where n_1 and n_2= the numbers of samples in data sets 1 and 2. The value of s is defined by the equation:

$$s = \sqrt{\frac{\Sigma y_1^2-(\Sigma y1)^2/n1+\Sigma y2^2-(\Sigma y2)^2/n_2}{n_1+n_2-2}}$$

2.9 Other DNA Manipulations

2.9.1 Dephosphorylation of 5' ends

Calf intestinal alkaline phosphatase was prepared from an ammonium sulphate slurry obtained from Boehringer containing 5mg/ml of the phosphatase. 20µl of this slurry was spun in a microcentrifuge and the ammonium sulphate supernatant removed. Care was taken to ensure no ammonium sulphate solution was carried over at this step. The protein precipitate was dissolved in 100µl of TE containing 50% glycerol and this preparation was stored at -20°C for 1-2 months. To remove 5' terminal phosphates from DNA molecules prior to cloning, 0.5µg of calf intestinal phosphatase was used per µg of DNA termini. The reaction was incubated at 37°C for 30 minutes in the manufacturers' recommended buffer. The reaction was stopped by heating the reaction to 65°C for 20 minutes and the DNA was extracted using QIAGEN nucleotide removal spin columns. The DNA was then eluted from the column in 50µl TE for direct use in other protocols.

2.9.2 Ligation of DNA

Ligation of DNA was routinely carried out in a buffer of 20mM Tris-Cl, pH 7.6, 10mM $MgCl_2$, 10mM DTT, 1mM ATP. Usually, 0.5→2 units of T4 DNA ligase were used in a total reaction volume of 5→20µl. Ligation reactions were typically incubated at room temperature overnight. Blunt ended ligations were incubated at 12°C overnight. Typically 10-50ng of vector DNA was used in a cloning procedure and the amount of insert titrated in the molar ratios 1:1→1:100 vector:insert.

2.9.3 Preparation of DNA for Sequencing

100µg of a large scale preparation of a plasmid was treated with 10µl RNase A (2mg/ml) at 37°C for 20 minutes in the presence of 10mM EDTA (to inhibit any DNase activity present in the RNase preparation). The sample was then extracted twice with chloroform and the DNA isopropanol precipitated to remove the RNase enzyme from it. The DNA pellet was then washed in 70% ethanol and vacuum dried. The pellet was

dissolved in 40µl of TE (10mM Tris-Cl pH 7.4, 1mM EDTA) and to this 50µl of 13% polyethylene glycol (PEG) 8000 (autoclaved) and 10µl 4M NaCl were added and the sample incubated on ice 20 minutes. The DNA was recovered by centrifugation in a microcentrifuge for 15 minutes at 4°C. Care was taken to remove all trace of the PEG by washing three times with 70% ethanol. The pellet was then resuspended in 40µl of TE and used directly in the sequencing protocols after quantitation by reading the A_{260nm} (absorbance at 260nm).

2.10 PCR Reactions

PCR reactions were carried out using Taq DNA polymerase or Vent DNA polymerase. The enzymes were used following the manufacturers' recommended conditions and buffers. All of the genomic DNA samples from the patients studied in this thesis were supplied by colleagues involved in the characterisation, diagnosis and treatment of the patients. The particular conditions used to amplify the minimal factor IX promoter (see Appendix1 for primers) from genomic DNA were as follows:

95°C Denaturation (30 seconds)

55°C Annealing (30 seconds)

72°C Extension (30 seconds)

Repeat for 30 cycles.

2.11 Sequencing of DNA

DNA molecules were sequenced in one of two ways, either using the Amersham/USB Sequenase 2.0 dideoxy sequencing kit, with minor modifications, or using ABI Taq DyeDeoxy sequencing methods. The plasmid DNA for both types of sequencing was prepared as described in section 2.9.3. Typically, 10µg of this plasmid DNA was used for the dideoxy sequencing protocol. The plasmid DNA and primer were mixed in a volume of 16µl to which 4µl of 2M NaOH/1mM EDTA was added and incubated on ice for 5 minutes. 4µl of 2M NH_4CH_3COO was then added and the template DNA and primer precipitated by adding 55µl of ice cold ethanol and this

mixture incubated on dry ice for at least 30 minutes. The DNA was recovered by spinning in a microcentrifuge for 15 minutes and resuspended in 10µl of ice cold Sequenase reaction buffer prior to using the sequencing kit. All of the subsequent reactions were performed on ice and as described in the Sequenase standard protocol. The sequencing reactions were then loaded onto a 6% denaturing polyacrylamide gel in 1xTBE and the electrophoresis performed at 25 watts for 2-5 hours, depending upon the range of sequence which was required. The gels were then dried onto Whatman 3MM paper and exposed to Kodak AR film for at least 24 hours.

Samples were prepared for the ABI DyeDeoxy sequencer by mixing 1 µg of the template DNA with 3.2pmol of primer, and the ABI DyeDeoxy sequencing premix in a total volume of 20µl. This mixture was then subjected to 25 cycles of the following protocol on a Perkin Elmer 9600 GeneAmp system:

96°C Denaturation (15 seconds)

50°C Annealing (15 seconds)

60°C Extension (4 minutes)

After this procedure the DNA in the samples was extracted by phenol/chloroform extraction (twice), chloroform extraction (once), and then ethanol precipitation before they were analysed on the ABI 373 DNA sequencer. If PCR products were to be sequenced directly, then the desired band would be purified as described in section 2.5 and approximately 100ng of this purified sample would be used as the template DNA in the sequencing protocol above.

The oligonucleotide tetramer/dimer reporter construct mixtures posed a problem to these sequencing methods. The ABI Taq DyeDeoxy did not successfully sequence these constructs and so the Sequenase dideoxy sequencing kit was used. As these plasmids were mixtures they had to be sequenced by two primers, one either side of the *Xba*I site into which the oligonucleotides were cloned. As the tetramer insert is, effectively, a dimer of the dimer insert, the dimer construct should have the same sequence as both the 5' two copies of the ARE oligonucleotides, and the 3' two copies

(see Fig 5.7). This then allows the determination of the sequence of both the dimer construct and the tetramer construct in the mixture by using two primers.

2.12 Protein A-AR DNA Binding Domain Fusion Protein Production and Partial Purification

500mls of *E.coli* (strain N4380-1, (Pharmacia)) carrying the expression vector pRIT2TAR (De Vos et al., 1991) was incubated overnight at 30°C in 2xTY medium containing 50µg/ml ampicillin. The production of the fusion protein was induced by the heating of the cultures in a shaking waterbath until the temperature of the cultures was 42°C. The cultures were then incubated at 37°C for 3 hours. The cells were pelleted at 4°C and then resuspended in 20mls buffer A (100mM Tris-Cl pH 7.6, 200mM NaCl, 10mM $MgCl_2$, 0.1mM EDTA, 10% glycerol, 1mM PMSF). Lyzozyme was then added to a final concentration of 2.5mg/ml. The cells were then incubated on ice for 20 minutes. The cell suspension was then sonicated at 6 microns (two bursts of 15 seconds) and the lysate was then cleared by spinning at 48k rpm for 30 minutes at 4°C in a 60Ti rotor and Beckman L8-M ultracentrifuge.

The cleared lysate was then passed over an IgG Sepharose Fast Flow (FF) column (Pharmacia), washed extensively with TST (50mM Tris-Cl pH 7.6, 150mM NaCl, 0.05% Tween 20). The absorbance of the eluent at 280nm was monitored with a Pharmacia LKB uvicorder. When the A_{280nm} of the wash had fallen to baseline levels the fusion protein was eluted with 0.5M acetic acid. The fractions of the eluent which contained the highest A_{280nm} were then pooled and freeze dried in a vacuum drier. The residue was dissolved in deionised water, frozen and dried again to remove any residual acetic acid. The residue remaining was dissolved in 2mls buffer B (25mM Hepes-KOH pH 7.6, 40mM KCl, 0.1mM EDTA, 10% glycerol, 1mM DTT), aliquoted into 100µl aliquots and stored at -70°C.

2.13 Determination of Protein Concentration

Protein concentrations were determined using the Pierce Coomassie Protein Assay which is based on the Bradford assay. 0.5ml of the assay reagent, containing Coomassie brilliant blue G-250, was added to 0.5ml of several dilutions of the protein sample to be tested. The Coomassie brilliant blue changes colour from reddish-brown to blue when added to protein and the A_{595nm} (Absorbance at 595nm) is then determined. The assay had been calibrated with solutions of bovine serum albumin of different, known concentrations prior to assaying the extract dilutions, see Table 2.1. These results were used to construct a standard curve for the determination of total protein concentration in dilutions of the extracts (see Fig 3.4). The average concentration value of any sample could then calculated.

Table 2.3 The construction of a standard curve for the Coomassie protein assay

Sample	A_{595nm} (a)	A_{595nm} (b)	Average	-Blank
BLANK	0.374	0.374	0.374	-
1mg BSA	0.389	0.394	0.3915	0.0175
2.5mg BSA	0.420	0.422	0.421	0.047
5mg BSA	0.475	0.469	0.472	0.098
10mg BSA	0.539	0.547	0.543	0.167
15mg BSA	0.606	0.592	0.599	0.225
25mg BSA	0.682	0.684	0.683	0.309

2.14 SDS PAGE of Protein Samples

Protein samples were analysed on standard Laemmli gradient gels (Laemmli, 1970) systems using a Biorad minigel apparatus. Typically the stacking gel was 4% acrylamide, 0.125M Tris-Cl pH 6.8 and the separating gel was 10% acrylamide, 0.375M Tris-Cl pH 8.8 (the electrophoresis buffer was 3g Tris-Cl, 14.4g glycine, 1g SDS in 1 litre). Samples were diluted at least 1 in 4 with the sample loading buffer

(62.5mM Tris-Cl pH 6.8, 10% glycerol, 2% SDS with 1µl β-mercaptoethanol added to each 20µl buffer immediately prior to use.

The gels were run at 200 volts for 45 minutes and then stained using Coomassie Blue stain (0.25g Coomassie Blue in 45% methanol, 10% acetic acid) and destained (10% methanol, 10% acetic acid) until the protein bands became clear.

2.15 Western Blots

Western blots were performed using the standard Biorad SDS PAGE minigel apparatus. An unstained SDS PAGE protein gel was placed in transfer buffer (2.9g glycine, 5.8g Tris-Cl, 0.37g SDS in 1 litre) for 30 minutes. The gel was then sandwiched between 2 pieces of Whatman 3MM paper and 1 piece of nitrocellulose (Hybond, Amersham). The transfer sandwich was then placed in the electroblotting apparatus (Biorad), ensuring that the nitrocellulose was between the polyacrylamide gel and the positive electrode. The transfer was performed at 100 volts for 1 hour. The transfer could be checked by following the prestained protein markers (Biorad) onto the nitrocellulose.

After transferring the proteins to the nitrocellulose the blot was incubated in blocking buffer (5% skimmed milk powder, 0.1% Triton X-100, in PBS, (phosphate buffered saline 137mM NaCl, 2.7mM KCl, 8mM Na_2HPO_4, 1.5mM KH_2PO_4, pH 7.3)) for 2 hours at 4°C. The antibody was then added, typically diluted 1 in 1000 to 1 in 10000 in blocking buffer, and incubated with the blot for 1 hour at 4°C. The nitrocellulose was then washed 3 times with blocking buffer and once with PBS before the developer was added. The developer used for the blots was the Amersham ECL kit. The standard protocol in the kit was followed and the signal was detected using Kodak S film. Exposures were typically 10 seconds→2 minutes.

2.16 Synthesis and Purification of Crude Oligonucleotides

Oligonucleotide samples were synthesised using an ABI 394 RNA/DNA synthesiser. Crude, synthetic oligonucleotide samples were subjected to electrophoresis

through a denaturing polyacrylamide gel (20% acrylamide, 7M urea in 1xTBE) to purify the desired oligonucleotide away from the failure sequences generated during its synthesis. An equal volume of formamide was added to the oligonucleotide sample and the mixture heated to 90°C for 3 minutes. The sample was then cooled rapidly on ice before it was loaded onto the gel.

The gel was run at 30 watts for 4-5 hours to allow sufficient separation of the oligonucleotides. The gel was then covered with Saran Wrap and the oligonucleotide visualised by long wave length UV shadowing by placing the gel onto a TLC plate that fluoresces at 254nm and irradiating it with long wave ultra violet light. The desired band was cut out of the gel, and the gel slice covered with 50mM KCl. This was incubated at 37°C for at least 4 hours to elute the oligonucleotide from the gel.

1ml of the eluent from the gel slice was loaded onto an NAP-10 column (Pharmacia) using the manufacturers instructions in order to desalt the sample. The column was washed with distilled water. 500µl fractions were collected from the column and the A_{260nm} measured for each. The fractions which had the highest A_{260nm} were pooled and freeze dried. The residue was dissolved in 100µl deionised water and the concentration of the sample determined by reading the A_{260nm}. The samples were stored at -20°C.

2.17 Preparation of Radioactive DNA Oligonucleotide Probes

2.17.1 Labelling

200ng of purified oligonucleotide was end labelled with $\gamma\,^{32}P$ ATP by incubating it for 30 minutes at 37°C with PNK(polynucleotide kinase) buffer (70mM Tris-Cl, pH 7.6, 10mM $MgCl_2$, 5mM DTT), 10units of PNK (New England Biolabs), and 10µCi $\gamma^{32}P$ ATP in a reaction volume of 10µl.

2.17.2 Annealing

An excess of the complementary oligonucleotide (typically 500ng) was then added to the reaction mix with 2.5µl of 2.5M KCl and the mix was made up to 50µl

with deionised water. This mix was heated to 90°C to completely denature the DNA oligonucleotides and then cooled slowly to room temperature to allow the two oligonucleotides to anneal.

2.17.3 Purification

This reaction mix was then loaded onto a non-denaturing 10% polyacrylamide gel in 1xTBE. The gel was run at 30 watts for 4-5 hours and then covered with Saran Wrap and an autoradiograph was taken to identify the position of the probe. The band which contained the double stranded oligonucleotide probe was then cut out of the gel and the DNA eluted from it by incubating the gel slice in 50mM KCl overnight at 37°C. The radioactivity contained in the probe was checked by counting 1µl of the probe preparation in a scintillation counter using Cherenkov radiation method which allows almost 45% efficiency of counting without the use of liquid scintillant (Beckman Instruments Liquid Scintillation Manual, 1988, T-1648-NUC-88-4)

2.18 Electrophoretic Mobility Shift Assays

1µg of the protein A-AR DNA binding domain fusion protein was incubated with non-specific competitor DNA (poly dI-dC; Pharmacia) on ice for 10 minutes in the reaction buffer (20mM Hepes.KOH pH 7.9, 25mM KCl, 0.1mM EDTA, 10% Ficoll, 0.5mM DTT, 2mM $MgCl_2$).

Then, 10000cpm of the radioactive DNA oligonucleotide probe (<0.1ng) was then added along with 50ng of cold oligonucleotide, making a final amount of approximately 50ng. The protein is in about ten fold molar excess over the oligonucleotide probe in this reaction. The binding reaction was then incubated on ice for 20 minutes and was then loaded onto a 4% native polyacrylamide gel in 0.25x TB (22.3mM Tris-Cl, 22.3mM boric acid) and run at 100 volts for 2 hours.

The gels were then transferred to DE-81 ion exchange paper (Whatman) and dried. The gels were then exposed to Kodak AR film for 12-48 hours or exposed to a phosphorimager screen (Molecular Dynamics). The amount of radioactive

oligonucleotide probe bound to protein relative to the free probe was calculated using a phosphorimager and the ImageQuant software (Molecular Dynamics). C50 values were then calculated from this data.

2.19 Preparation of Double-Stranded Competitor Oligonucleotides for Mobility Shift Assays

Double-stranded DNA oligonucleotides were prepared by mixing together 25µg of each complementary oligonucleotide in 200µl in a final concentration of 0.5M KCl. This mixture was then heated to 95°C and cooled slowly to room temperature to allow annealing to occur. The double stranded oligonucleotide was then precipitated by the addition of 3 total volumes of ethanol and the mixture kept at -70°C for 2 hours. the oligonucleotide was then pelleted by spinning in a bench microcentrifuge and the pellet dried under vacuum. The pellet was then resuspended in 50µls TE and the A_{260nm} read on various dilutions to gain an accurate estimate of the final concentration of the oligonucleotide.

2.20 Competition Mobility Shift Assays

Competition mobility shift assays were performed in an identical manner to the assay described above except that an excess of non-radioactive mutant oligonucleotide was added to the wild type probe in order that it would compete the visible mobility shift on the gel. The competitor oligonucleotide was added to the probe before the probe was added to the protein. Typical levels of competitor used in these assays ranged from 500ng to 4µg of competitor oligonucleotide per reaction (at 4µg the competitor is in about eight fold molar excess over the protein in the reaction).

These reactions were then analysed using a native 4% polyacrylamide gel which was run at 100 volts for 2 hours, then dried and exposed to a phosphorimager screen. The efficacy of a particular competitor oligonucleotide was gauged by calculating a C50 value for the particular mutant oligonucleotide. The C50 value is the molar excess of competitor oligonucleotide which is required to compete out 50% of the control gel

mobility shift. The graphs generated to read off the C50 values can be seen in Chapter 4. Some studies have used the reciprocal of the C50, and defined this value as the relative affinity of the oligonucleotide for the DNA binding protein (Adler et al., 1993).

2.21 Phosphorimager Analysis

A Molecular Dynamics phosphorimager was used to quantitate the results obtained from both the gel mobility shifts and the CAT assays. In both cases, the ImageQuant software package was used to quantitate:

(i) the percentage of the radioactive probe which was bound to the protein in the gel mobility shift assays. After the calculation of this value for each of the lanes in the competition gel mobility shift assays, the percentage of the radioactive oligonucleotide probe found in the retarded band was expressed as a proportion of the percentage of the probe found in the retarded band in the control gel mobility shift which had no unlabelled competitor oligonucleotide added to it.

(ii) the percentage of the chloramphenicol which was acetylated in the CAT assays. The transactivation of the HSV tk promoter in these assays was expressed as the fold increase in the percentage acetylation of the chloramphenicol in the assays containing both hormone and AR expression vector when compared to the background value, which was the average of the result obtained for the two controls, one without hormone, the other without AR expression vector.

2.22 Mammalian Cell Lines

The mammalian cell lines used in this thesis were PLC/PRF5, HepG2 and HeLa . The cells lines were maintained in minimal essential medium supplemented with 10% foetal calf serum, and 1% nonessential amino acids, glutamine was added to a final concentration of 2mM. Cell stocks were stored at -70°C in foetal calf serum with 10% DMSO added. The hormone transactivation assays required the use of DMEM without the phenol red pH indicator (obtained from Gibco), as this is known to act as a

weak estrogen (Welshons et al., 1988). The foetal calf serum used in these experiments was charcoal stripped to remove endogenous steroids (see 2.30).

2.23 Construction of ARE Tetramer-CAT Reporter Plasmids

In order to construct the factor IX ARE tetramer plasmids, the DNA oligonucleotides corresponding to the ARE were first kinased to phosphorylate the 5' nucleotide. Then 100ng of each phosphorylated oligonucleotide was allowed to anneal by mixing the complementary oligonucleotides together, adding KCl to a final concentration of 50mM, heating the mixture to 95°C, and allowing it to cool slowly to room temperature over two hours. These double stranded, phosphorylated oligonucleotides were then multimerised by adding T4 DNA ligase (1 unit), and Gibco Ligase buffer to 1x (50mM Tris-Cl pH 7.6, 10mM $MgCl_2$, 1mM ATP, 5% polyethylene glycol 8000), and incubating the mixture at room temperature for three hours. Then 10ng of phosphatased (see 2.8) *Xba*I cut pBLCAT2 was added to the mixture and the incubation continued overnight. The oligonucleotides were synthesised with four base 5' overhangs complementary to the cutting sites for the restriction enzymes *Xba*I and *Bam*HI. Hence, only those oligonucleotide multimers which had *Xba*I sites at each end would be expected to clone into this vector. The only oligonucleotide multimers with this arrangement of 5' overhangs would be dimers, tetramers, hexamers, octomers etc. After ligation and transformation into *E.coli*, individual clones were analysed by restriction digest after small scale plasmid preparation (see 2.4) and it was found that each clone which possessed a plasmid which contained a tetramer insert, also contained a dimer insert too. The attempts to solve this problem are discussed in Chapter 5.

2.24 *Sau96*I profiling

The enzyme *Sau*96I was incubated with approximately 5ng of the purified DNA fragment at 37°C for 1 hour in React buffer 2 (Gibco). The digest was then added to a Klenow labelling mix with a dNTP (25μM each) mix minus CTP, with added trace

αP^{32} dCTP and incubated at room temperature for 15 minutes. The labelled fragments were then analysed on a 6% denaturing polyacrylamide gel. The gel was run for 2-3 hours at 25 watts and dried down onto Whatman DE81 ion exchange paper. The gel was then exposed to Kodak S film for 12-36 hours and the autoradiograph developed.

2.25 Transient Transfections

All transfections were carried out using a modification of the calcium phosphate method of transfection (Graham and Van der Eb, 1973). To 247.5µl HBS (Hepes buffered saline, 2g Hepes-KOH pH 7.1, 3.2g NaCl in 200mls), 2.5µl of a 70mM solution of Na_2HPO_4 was added and the pH adjusted to 7.05 with NaOH before filter sterilising. To this solution 250µl of a solution containing 20µg DNA and 240mM $CaCl_2$ was added dropwise whilst the solutions were vortexed. The resulting precipitate was left at room temperature for 30 minutes before 5mls of media were added. This mixture was then placed over cells in 60mm dishes which were between 30% and 60% confluent. The optimum amount of DNA for the transfections in these dishes was found to be 20µg and did not vary significantly in the three cell lines used. When determining the plasmid to be used for the internal transfection control, the total DNA amount was set at 20µg for each of the plasmid samples. The typical plasmid mixes used in the transfections were:

HNF4 transactivations: 15µg factor IX minimal promoter-luciferase reporter construct; 1µg HNF4 expression vector (pCB6HNF4, see 2.6); 3µg pBluescript (see 2.6); 1µg pCH110 (see 2.6).

Tetramer construct transactivations: 8µg tetramer reporter vector; 4µg AR expression vector or HNF4 expression vector (pCB6HNF4 see 2.6); 4µg pGLtk internal transfection control (see Fig 5.9); up to 20µg with pBluescript (see 2.6).

The precipitate was left on the cells overnight and then washed off in PBS the following morning. The cells were then incubated for a further 24 hours before harvesting. The hormone transactivation assays required the addition of 5α-dihydrotestosterone to a final concentration of 25nM 16 hours after transfection, as

described in Crossley et al., (1992). The cells were typically harvested 48 hours after the transfection.

2.26 Preparation of Extracts from Mammalian Cells

The method for preparing the cell extracts for reporter gene assays was modified from Crossley (1990), D.Phil thesis. The cell monolayers from 60mm dishes were washed twice with PBS and scraped off into 0.5ml TEN (10mM Tris-Cl pH 7.4, 1mM EDTA, 100mM NaCl). The cells were then pelleted by centrifugation in a bench microcentrifuge for 1 minute. 200 µl of 250mM Tris-Cl pH 7.5 was then added to the cells. The cells were resuspended and lysed by sonication at 6 microns (2 x 5 second bursts). Cell debris was then spun down in a microcentrifuge for 1 minute. The supernatant was then used directly in luciferase, CAT or β-galactosidase reporter gene assays.

2.27 Chloramphenicol Acetyl Transferase (CAT) Assays

The method used was modified from Heard et al., (1987). D-threo-[dichloroacetyl-1-^{14}C] chloramphenicol, 50-60mCi/mmol (Amersham) was made up in 2% ethanol to 0.04µCi/µl. This solution was stored at -70°C. Acetyl coenzyme A (Sigma) was made up to 50mM and stored at -70°C.

Typically 10µl of cell extract was used to perform the CAT assay. The assays were performed in a total volume of 37.5µl and to the cell extract 0.5µl ^{14}C chloramphenicol, 1µl acetyl coenzyme A and 26µl 250mM Tris-Cl pH 7.5 were added. The final concentrations of chloramphenicol and acetyl coenzyme A were approximately 0.01µCi/µl and 1mM respectively. The reaction was incubated at 37°C for 1 hour and then stopped by the addition of 200µl ethyl acetate.

The reactions were then spun for 1 minute in a bench microcentrifuge and the organic phase removed with a pipette. The samples were then dried in a vacuum drier and resuspended in 20µl ethyl acetate. The samples were loaded onto silica gel thin layer chromatography plates (Camlab). The chromatogram was run until the solvent

front reached the top of the plate. The plate was then dried and exposed for 4→12 hours, to a phosphorimaging screen (Molecular Dynamics). The relative amounts of acetylated and non-acetylated chloramphenicol were then calculated using the phosphorimager and the ImageQuant software (Molecular Dynamics) and the level of transactivation calculated by comparing the activity of the extracts from the transactivations (+hormone and AR expression vector) with the background values (-hormone, and - AR expression vector).

2.28 Luciferase Assays

Luciferase assays were performed using the same extracts as were used for the CAT assays. 20µl of the cell extract was mixed with 100µl of the luciferase assay reagent (Promega) at room temperature, and the maximum luminescence was recorded using an LKB 1250 luminometer.

These results were often used as an internal control to correct the CAT assay results for the efficiency of the individual transfections.

2.29 β-Galactosidase Assays

Again, the same cell extracts used for the previous two reporter gene assays were used for the β-galactosidase assays. Typically, 10→40µl of the extract was added to 400µl of Z buffer (60mM Na_2HPO_4, 40mM NaH_2PO_4, 10mM KCl, 1mM $MgCl_2$, which was made up to 60mM β-mercaptoethanol before use), and 80µl of 2mg/ml ONPG (o-nitrophenyl β-D-galactosidase) were added. The reactions were incubated at 37°C for 1-4 hours, until a definite yellow colour was evident. The reaction was stopped by the addition of 400µl of 1M Na_2CO_3 and the A_{420nm} was measured. These measurements were used to correct other reporter assays for transfection efficiency.

2.30 Charcoal Stripping of Foetal Calf Serum

2.30.1. Preparation of Dextran Coated Charcoal

A dextran (Sigma D4751, approximate Mr 74200) solution of 5mg/ml was made up and activated charcoal (Sigma C9157) added to a final concentration of 50mg/ml. This mixture was stirred thoroughly overnight at 4°C. The dextran coated charcoal was pelleted by spinning for 30 minutes at 4000 rpm in a Beckman J2-21 centrifuge and the supernatant removed.

2.30.2. Stripping of Foetal Calf Serum

100mls of dextran/charcoal mixture was spun down to obtain enough dextran coated charcoal for 500mls of foetal calf serum. The pelleted slurry was added to the calf serum and the two were stirred at 4°C overnight. The mixture was then spun at 10000rpm for 1 hour in a Beckman J2-21 centrifuge. The cleared calf serum was taken off and filtered first through a 0.45µm filter and then through a 0.2µm filter. The filtered, stripped serum was then aliquoted into 50ml aliquots and stored at -20°C.

Chapter 3: Production and Partial Purification of a Recombinant Protein A-Androgen Receptor Fusion Protein

3.1 Introduction

Transcription factors are modular proteins (Johnson and McKnight, 1989), and many of their domains are able to function independently after production in recombinant systems. The domain structure of the androgen receptor is shown in Fig 1.13, and the DNA binding domain contains two zinc finger motifs (see 1.8.3). Zinc fingers have been shown to bind to their DNA recognition sequences (Nagai et al., 1988) when produced in isolation in a recombinant system. The aim of the work reported in this chapter was to produce a protein A-AR DNA binding domain fusion protein (De Vos et al., 1991) which could be used in competition gel mobility shift assays with the factor IX wild type and mutated ARE oligonucleotides. This protein A-AR DNA binding domain fusion protein was chosen for this study as it had already been used in similar assays with the factor IX ARE (Crossley et al., 1992). It is a well characterised fusion protein (De Vos et al., 1991) and has been used to determine a consensus binding site for AR (Roche et al., 1992). Another factor influential in the choice of this fusion protein was that, at the time, no full length recombinant AR was available. The construct used to express the fusion protein was kindly provided by Dr. Piet De Vos and is shown in Fig 3.1.

3.2 Production and Partial Purification of the Protein A-AR DNA Binding Domain Fusion Protein

The plasmid pRIT2TAR is shown in Fig 3.1, and is described in De Vos et al., (1991). The fragment of the AR cDNA in the expression vector coded for amino acids 533-637 of the rat AR. This portion of the protein sequence contains the entire DNA binding domain (domain C) and part of the hinge region (domain D). This portion of the AR cDNA is fused to the first 265 amino acids of *S.aureus* protein A. The protein A part of the fusion, being prokaryotic, is suspected to protect the protein from

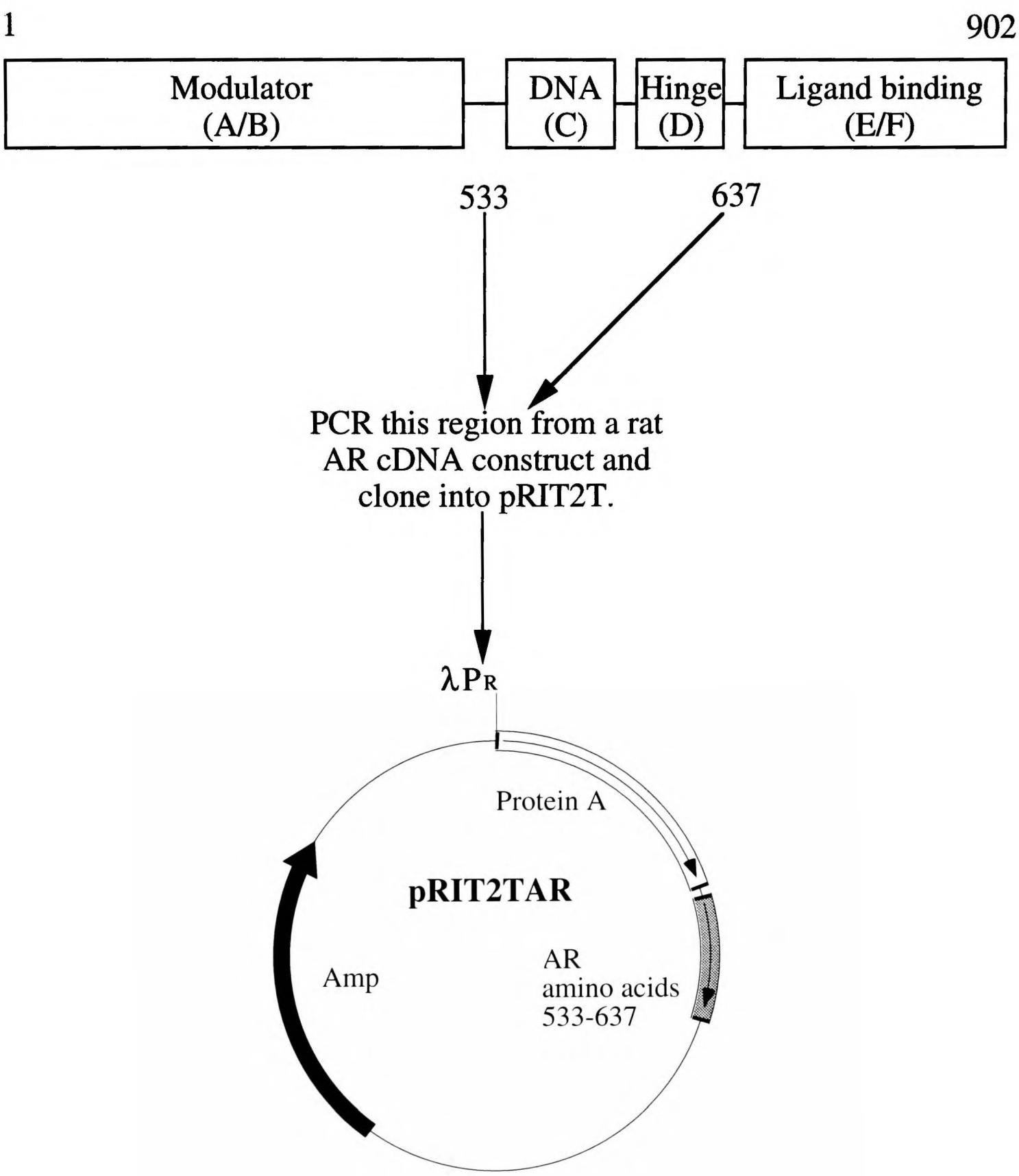


Figure 3.1 The construction of the protein A-AR DNA binding domain fusion protein expression plasmid. Kindly supplied by Dr. P. De Vos. The amino acid numbers from the rat AR are shown above and below the schematic view of the protein. The domains are named and lettered to allow a comparison with Fig 1.13.

degradation in the bacterium, thus allowing high levels of expression to be achieved. The resulting fusion protein is under the control of the bacteriophage λ P_R promoter. The plasmid pRIT2TAR was grown in *E.coli* host N-4380 which harbours a cI857 λ lysogen which produces the temperature sensitive cI protein, see Zabeau and Stanley, (1982). At permissive temperatures (37°C→42°C) the cI mediated repression of the P_R promoter is removed, thus allowing expression of the fusion protein.

After the induction and harvesting of the bacterial cultures, the fusion protein was found predominantly in the soluble fraction of the bacterial extract. This soluble fraction was used directly for the purification of the fusion protein. The protein was partially purified as described in section 2.12 using IgG Sepharose (Pharmacia) affinity chromatography. Eluent from the affinity column was then passed over different ion exchange columns by FPLC (MonoQ and MonoS) to further purify the fusion protein as described in De Vos et al., (1991). The fusion protein was found to be unstable in these procedures; extensive degradation being observed by Coomassie stained SDS-PAGE and Western blot after every FPLC step (results not shown). Due to the instability of the fusion protein in this further purification step, it was decided to characterise the partially pure eluent from the IgG Sepharose affinity column.

3.3 Characterisation of the Fusion Protein

The eluent from the affinity column was analysed by SDS PAGE (see 2.14, and Fig 3.2). The fusion protein migrated with an apparent molecular weight of approximately 45kDa (see arrowed band, Fig 3.2), in agreement with De Vos et al., (1991). Some of the highly purified material obtained from Dr. De Vos was used as a marker for the fusion protein (Fig 3.2, lane 4). From the Coomassie stained gel (Fig 3.2) it can be estimated that approximately 40% of the eluent from the affinity column is protein A-AR DNA binding domain fusion protein.

The eluent and cleared bacterial extract were also analysed by Western blot (see 2.15, and Fig 3.3). The antibody used in the Western blot was a non-specific rabbit anti-goat IgG which bound tightly to the protein A region of the fusion protein. The

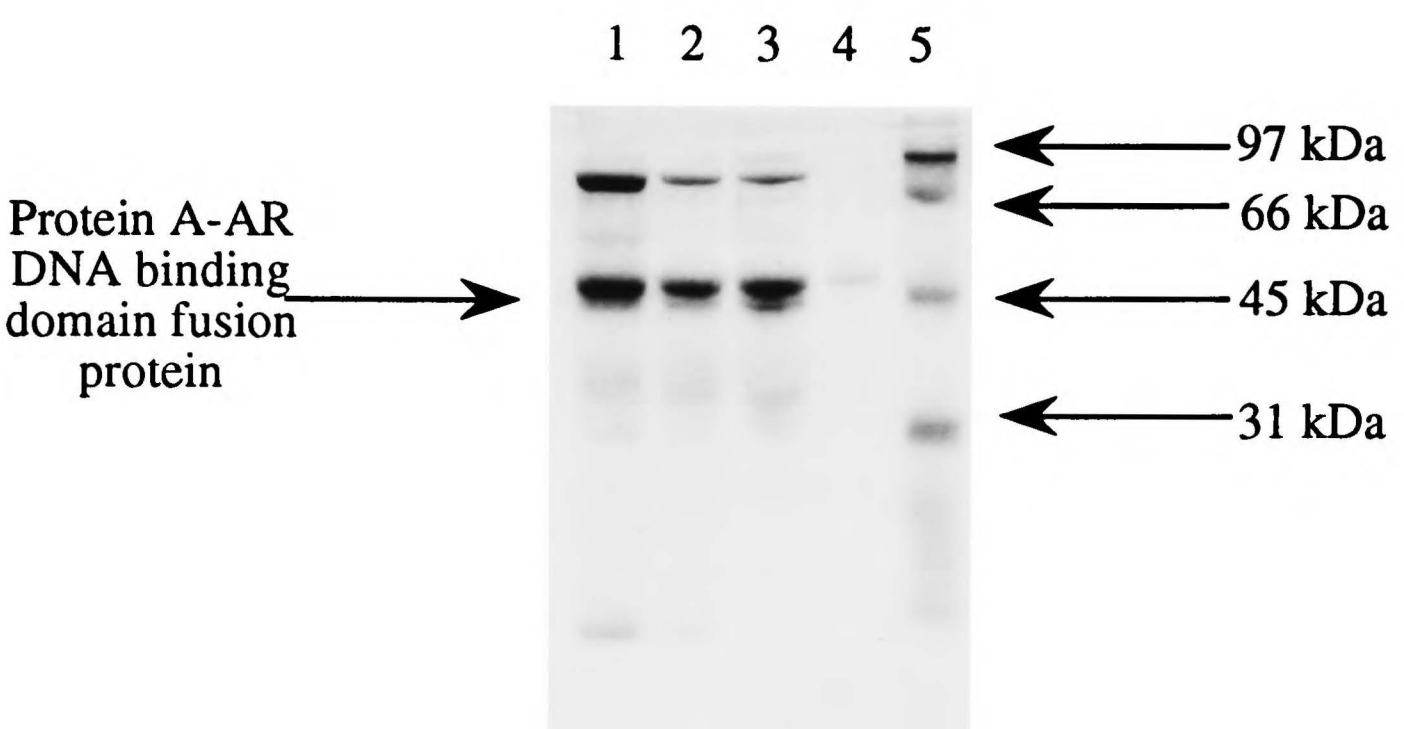


Figure 3.2 Coomassie Blue stained SDS PAGE of 3 fusion protein extracts. The lane order is, lane 1=extract 1; lane 2= extract 2; lane 3= extract 3; lane 4= a sample of purified protein A-AR DNA binding domain obtained from Dr. P. De Vos; lane 5= protein molecular weight markers (Biorad low range molecular weight markers , 97kDa→14kDa). The arrow indicates the protein A-AR DNA binding domain protein band.

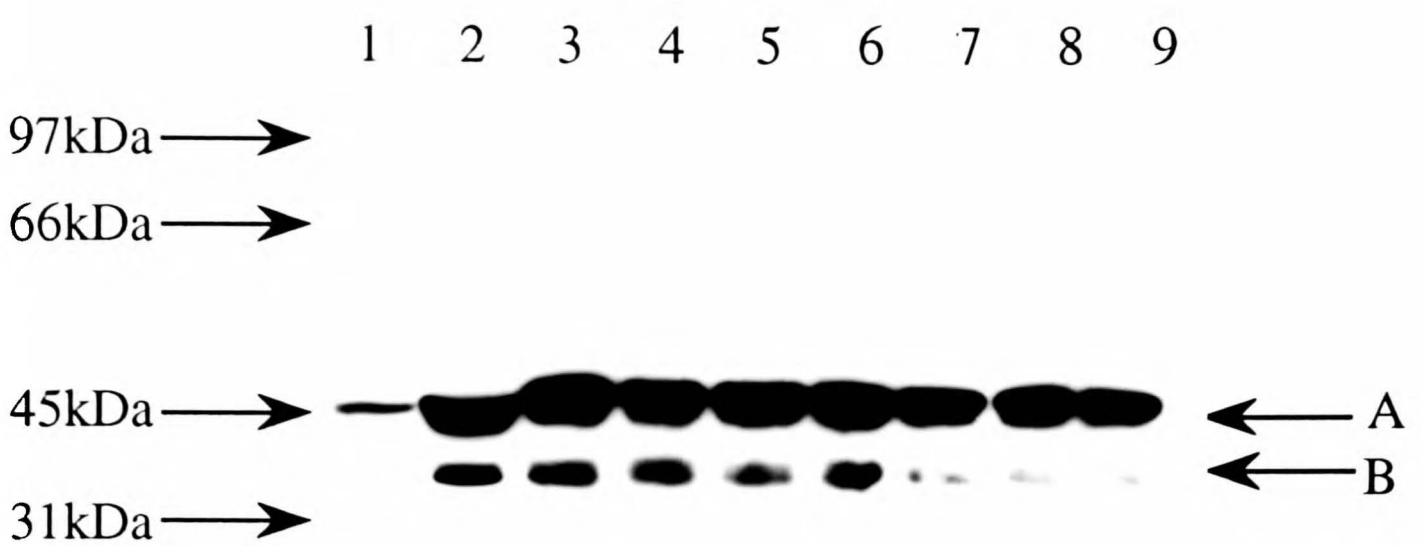


Figure 3.3 Western blot of a number of fractions from the IgG-Sepharose affinity chromatography column. The lane order is, lane 1= cleared extract from *E.coli* grown at 30℃ (non-permissive temperature); lane 2= cleared extract from *E.coli* grown at 42℃ (permissive temperature); lanes 3→9= consecutive fractions from the affinity column. The position of prestained protein molecular weight markers on the blot is shown. The antibody used was a rabbit anit-goat IgG (obtained from Sigma) conjugated to horseradish peroxidase. Two proteins are recognised by this antibody, band A= full length protein A-AR DNA binding domain protein, band B= a degradation product of band A, or a cross-reacting *E.Coli* protein.

antibody was conjugated to the horseradish peroxidase enzyme which allowed the detection of the complex (see 2.15). Fig 3.3 shows that bacteria grown at 30°C (non-permissive temperature) produced small quantities of the fusion protein (Fig 3.3, lane 1, arrow A) but that these quantities increased substantially upon heat induction (Fig 3.3, lane 2, arrow A). Lanes 3-9 are six fractions of the eluent from the affinity column which were subsequently pooled and freeze dried. Fig 3.3 also shows a band running slightly faster than the fusion protein (Fig 3.3, arrow B). This band was either a degradation product of the fusion protein, or an *E.coli* protein which cross reacts with the antibody. This band was estimated to be less than 5% of the intensity of the fusion protein band.

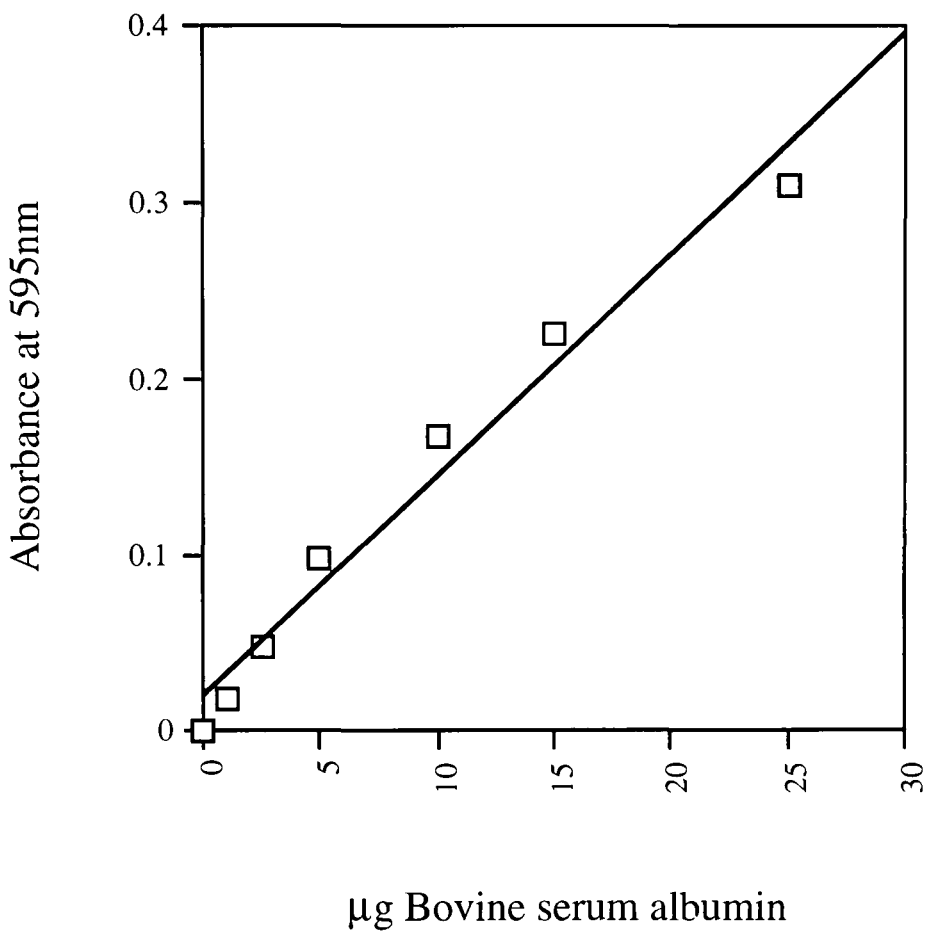
The total protein concentration in the extracts produced was determined by using a Coomassie Protein Assay (see 2.13) with reference to a standard curve constructed from BSA standards of known concentration (see Table 2.3, and Fig 3.4). The concentration of these extracts can be easily calculated from the standard curve. 2-4mg of protein were usually obtained from each 500mls of bacterial culture processed. This was equivalent to approximately 0.8-1.6mg fusion protein.

Table 3.1 Calculating the Concentration of two of the Fusion Protein Extracts

Sample	A_{595nm}	µg/ml	mg/ml (x1)	Average (mg/ml)
Extract1x100	0.272	19.5	1.95	2.25
Extract1x1000	0.052	2.54	2.54	
Extract2x100	0.177	12.2	1.22	1.34
Extract2x1000	0.038	1.46	1.46	

3.4 Functional Assay with the Fusion Protein

The fusion protein extracts were next tested for their ability to bind to DNA in a gel mobility shift assay. This assay is based on the observation that nucleic acid-protein complexes migrate slower through native, low ionic strength polyacrylamide gel



Equation for the line of best fit

$$y = 0.013x + 0.019$$

Figure 3.4 Coomassie protein assay standard plot for the quantitation of fusion protein extracts

systems than the free nucleic acid. If the nucleic acid is labelled (e.g. with a radioactive isotope) then the different complexes can be separated and visualised after a suitable electrophoresis time. This assay was originally developed to study the equilibria and kinetics of lac repressor-operator interactions (Fried and Crothers, 1981) and is now widely used to determine a variety of nucleic acid-protein interactions. The assay is usually performed in low ionic strength buffer as this, together with the supposed "caging effect" of the gel matrix, helps to stabilise the nucleic acid-protein complexes during electrophoresis.

The mutant oligonucleotides synthesised for this study were intended for use in competition gel mobility shift assays (Singh et al., 1986, Carthew et al., 1985). This type of assay involves the addition of a large excess of unlabelled, competitor oligonucleotide to the labelled oligonucleotide probe prior to the binding reaction. The greater the affinity of the competitor oligonucleotide for the protein, the greater will be the reduction of the labelled probe oligonucleotide/protein complex (compare lanes 3→5 with 6→8, Fig 3.5).

The eluent from the affinity column was tested in a competition gel mobility shift assay with a radioactive DNA oligonucleotide probe (see 2.17 and 2.18) which included the sequence of the wild type factor IX ARE (nucleotides -40→-9 from the factor IX promoter region, see Appendix 1). Fig 3.5 shows that a diffuse retarded band (arrowed) was obtained when the partially pure fusion protein was incubated with the wild type factor IX oligonucleotide probe on the gel mobility shift assay (lane 2, Fig 3.5). This diffuse band was competed by an excess of unlabelled factor IX wild type ARE oligonucleotide (lanes 3→5, Fig 3.5), but not by unlabelled -26G→C mutated factor IX ARE oligonucleotide (lanes 6→8, Fig 3.5). The -26G→C mutation has been reported to disrupt the binding of the AR to the factor IX ARE (Crossley et al., 1992). This result suggested that the protein-DNA complex formed in the binding reaction was specific for the sequence of the wild type factor IX ARE.

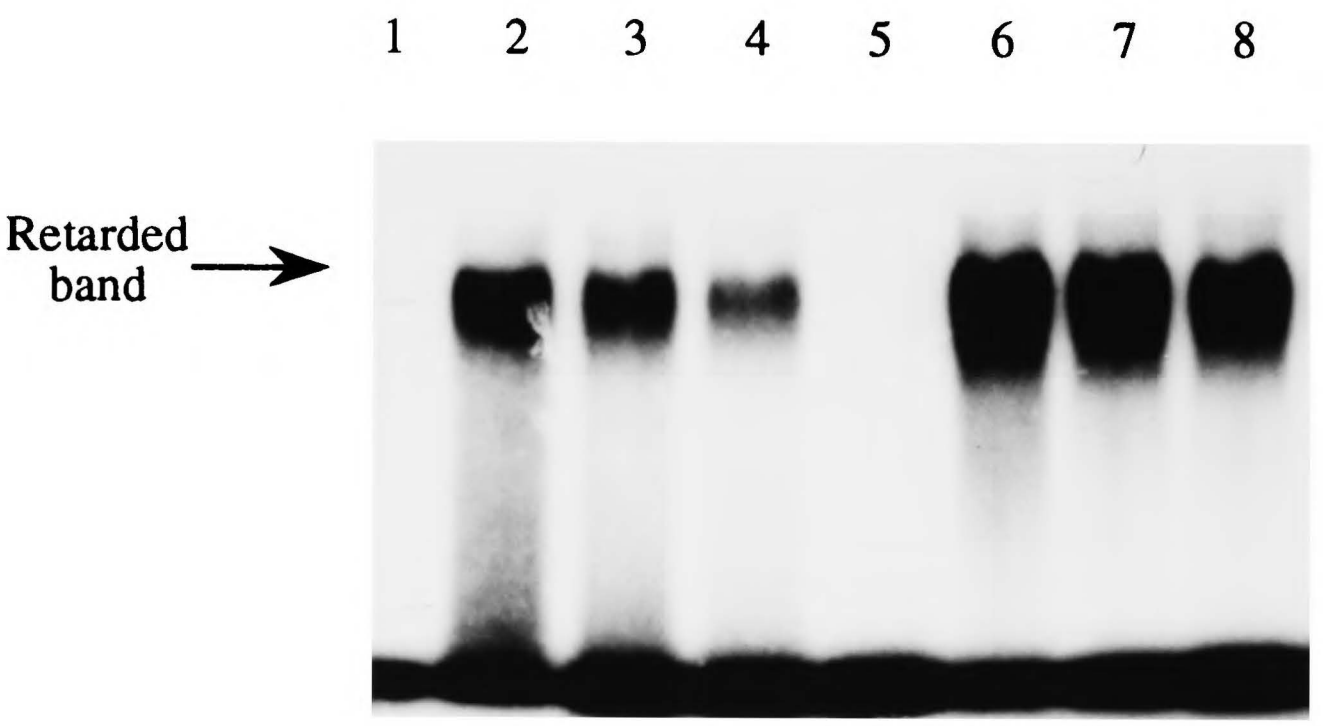


Figure 3.5 Autoradiograph of a competition gel shift assay. Lane order is, lane1= radioactive wild type factor IX ARE (-40→-9) oligonucleotide probe; lane 2= probe with fusion protein extract; lanes 3→5= excess factor IX wild type ARE competitor oligonucleotide; lanes 6→8= excess -26G→C mutated oligonucleotide competitor. The titrations used in both lanes 3→5 and lanes 6→8 was 10 fold, 20 fold, and 40 fold excess of competitor. The diffuse, retarded band is arrowed.

3.5 Summary

A protein A-androgen receptor DNA binding domain fusion protein has been produced and purified in mg quantities. The protein produced has been shown to be the correct molecular weight using SDS PAGE, (Fig 3.2), and to bind to IgG, both on an affinity chromatography column and in a Western blot, (Fig 3.3). In a competition gel mobility shift assay the concentrated eluent from the affinity chromatography column was able to retard a wild type factor IX ARE DNA oligonucleotide in a sequence specific manner (Fig 3.5).

Chapter 4: *In vitro* analysis of mutations in the factor IX ARE using a competition gel mobility shift assay.

4.1 Introduction

Crossley et al., (1992) reported that the disruption of two overlapping binding sites for the AR and HNF4 in the factor IX promoter region of the patient Haemophilia B Brandenburg was the cause of the failure in the recovery of plasma factor IX levels which is typical of the Leyden phenotype (see Fig 1.20). Others have suggested alternative developmental regulatory mechanisms which may function in the Leyden phenotype (Picketts et al., 1993, Kurachi et al., 1994, Reijnen et al., 1993, Boccia et al., 1996). Some of these have questioned the existence and importance the ARE (-36→-22) in the factor IX promoter region. In addition, only one consensus definition study has been performed with the AR (Roche et al., 1992). This study deduced a consensus DNA binding site which compares favourably with the sequence between -36 and -22 (see Fig 4.1).

The aim of the work described in this chapter was to confirm the report that the sequence from -36→-22 in the factor IX promoter could act as an ARE and to add to the present state of knowledge of the different sequences which can bind to AR. A detailed analysis of this sequence was also necessary to fully test the theory of Crossley et al., (1992), as such data may allow one to predict the phenotype of patients found with mutations in the overlapping region of the HNF4 site and the ARE (-27→-22), see Chapter 8. Therefore the sequence between -36→-22 was subjected to a mutagenesis study. This *in vitro* study was to be complemented with an *in vivo* transient transfection study (Chapter 5).

4.2 Mutagenesis Strategy

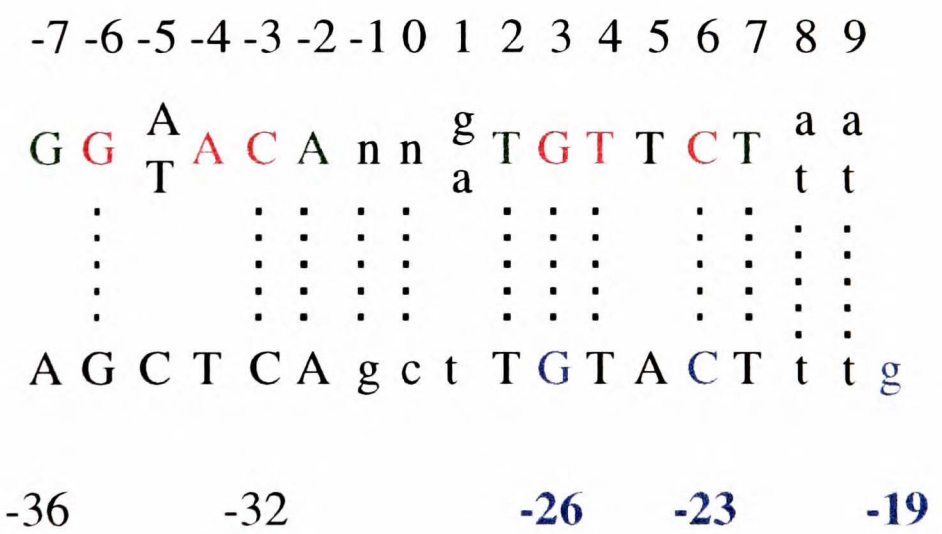
The sequence of the factor IX ARE compares favourably with the consensus sequence derived by Roche et al., (1992) (see Fig 4.1). Five of the six positions in the ARE defined as highly constrained are identical in the two sequences, and three of the

Figure 4.1 (A) A comparison of nine of the known naturally occurring androgen response elements found in the genome. C3= prostatic binding protein C3; SLP= sex limiting protein; PSA= prostate specific antigen; GUS= β-glucuronidase; ODC= ornithine decarboxylase; MEP 24= murine epididymal protein; MVDP= mouse vas deferens protein. The consensus sequence derived by Roche al., (1992) is shown for comparison. (B) A direct comparison of the factor IX ARE (lower sequence) with the consensus sequence of Roche et al., (1992) (upper sequence). The numbering scheme of Roche et al., (1992) is included above the consensus. The study of Roche et al., (1992) defined "highly constrained positions" (shown in red), and "moderately constrained" positions (shown in green). The factor IX sequence is aligned to the consensus and a match is denoted by a vertical dashed line. The positions where mutations have been identified in patients studied in this thesis are highlighted in blue in the sequence.

A

Gene	ARE (5'→3')	Reference
C3 (COREII)	AGTACG tga TGTTCT	Claessens et al., 1989
SLP (HRE 3)	GAAACA gcc TGTTCT	Adler et al., 1991
PSA	AGCACT tgc TGTTCT	Riegman et al., 1991
Probasin	ATAGCA tct TGTTCT	Rennie et al., 1993
GUS (intron9)	AGTACT tgt TGTTCT	Lund et al., 1991
ODC	AGTCCC act TGTTCT	Crozat et al., 1992
MEP 24	AGAACA acc TGTTGA	Ghyselinck et al., 1993
MVDP	TGAAGT tcc TGTTCT	Fabre et al., 1994
Factor IX	AGCTCA gct TGTACT	Crossley et al., 1992
Consensus	$GG^{A}_{T}ACAnn^{g}_{a}TGTTCT^{a\,a}_{t\,t}$	Roche et al., 1992

B



four positions defined as moderately constrained are also identical (see Fig 4.1). Fig 4.1 also highlights the more divergent nature of the 5' half of the palindrome, relative to the 3' half, of most natural ARE sequences.

The mutations were designed by considering the consensus study of Roche et al., (1992) and selecting a mutation which would be expected to disrupt the function of the ARE (see Table 1.4). The mutations tested are shown in Table 4.1.

Table 4.1 The mutations made in the factor IX promoter sequence from -36→-19 to study the binding of AR to this region. ^{P} = a mutation found in a haemophilia B patient; $^{+}$ = a change which makes the sequence a better consensus ARE; ^{D} = a double base pair mutation.

Position	Base Change	Position	Base Change
-36	A→T	-26	G→C^{P}
-35	G→C	-26	G→A^{P}
-34	C→G	-25	T→G
-33	T→C	-24	A→T^{+}
-32	C→A	-24	A→C
-31	A→C	-23	C→T^{P}
-30	G→A	-23	C→A
-28	T→G^{+}	-22	T→A
-27	T→G	-21/-20	T→G^{D}
		-19	G→C^{P}

At some positions, mutations had been identified in patients, so these were included in the study to gauge their effects on AR binding. In some cases, it was decided to create mutations which made the ARE fit the consensus sequence better, in an effort to create stronger AREs. Each of these mutations was designed as a DNA

oligonucleotide sequence. These oligonucleotides spanned the region from position -40→-9 in the factor IX promoter, and also included four base pair overhangs to allow cloning into reporter gene vectors (see Chapter 5). The sequences of the oligonucleotides used in this study are shown in Appendix 1.

These oligonucleotides were used in competition gel shift assays with a protein A-AR DNA binding domain fusion protein , see Fig 3.1 (De Vos et al., 1991). The relative frequency of the bases at each position in the AREs Roche et al., (1992) isolated were used to construct a table (see Table 1.4). This may be used to predict the effects of any mutation, at any position, in any ARE sequence. The predictions based on this pool of sequences identified by Roche et al., (1992) are shown in Table 4.2.

4.3 Optimising the Gel Mobility Shift Assay

To determine the optimum binding conditions for the binding of the factor IX ARE probe to the AR DNA binding domain fusion protein a number of buffers known to be used in gel mobility shift assays (see 2.18) were tested. Fig 4.2 shows the results of the buffer test. The composition of buffers 1→8 is detailed in Fig 4.2. The buffer previously used for the binding reaction (De Vos et al., 1991) corresponds to buffer 1 in the test.

A number of features are immediately evident from the buffer optimisation:

(i) the choice of buffer is critical, the gel mobility shift seen ranges from about 50% of the total radioactivity (lanes 4 and 6), to poor (lanes 5 and 8), or non-existent (lane 2).

(ii) all of the buffers containing 8% ficoll performed considerably better than the equivalent ones containing 20% glycerol (compare lanes 2, 5, 7 with lanes 3, 4, 6, respectively, Fig 4.2) .

(iii) in general, Hepes-KOH based buffers performed marginally better than those based on Tris-Cl.

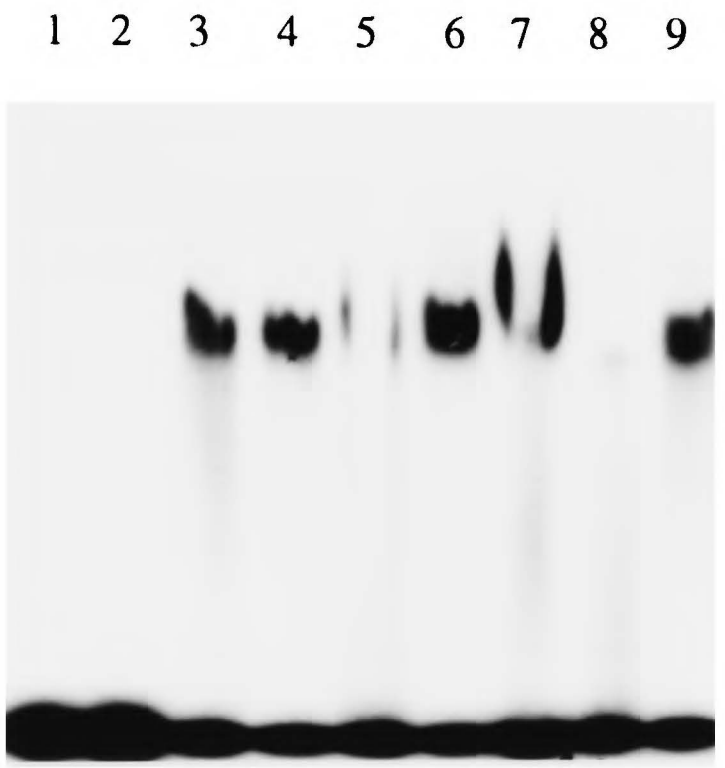


Figure 4.2 Buffer test to optimise the binding reaction of the gel mobility shift assay. Lane order, 1= probe only, lanes 2→9=buffers 1→8. Buffer compositions are given below. All buffers are 2x.

Buffer 1= 100mM Tris-Cl pH 7.9, 12.5mM $MgCl_2$, 1mM EDTA, 20% glycerol.

Buffer 2= as buffer 1 but 8% ficoll replaces 20% glycerol.

Buffer 3= 40mM Hepes-KOH pH 7.9, 50mM KCl, 0.2mM EDTA, 1mM DTT 4mM $MgCl_2$, 8% ficoll.

Buffer 4= as buffer 3 but 20% glycerol replaces 8% ficoll.

Buffer5= 40mM Hepes-KOH pH 7.9, 50mM KCl, 4mM Spermidine, 0.2mM EDTA, 0.05% NP-40, 200µg/ml BSA, 4mM $MgCl_2$, 8% ficoll.

Buffer 6= as buffer 5 but 20% glycerol replaces 8% ficoll.

Buffer 7= 40mM Tris-Cl pH 7.6, 100mM KCl, 0.4mM EDTA, 16% ficoll.

Buffer 8= 20mM Hepes-KOH pH 7.9, 80mM KCl, 4mM $MgCl_2$, 1mM DTT, 8% ficoll.

Table 4.2 Predictions on the effect of each base change made in the factor IX ARE based upon the relative base frequencies at each position in the ARE consensus. These values are calculated from Table 1 (C) from the results of the study of Roche et al., (1992) (see also table 1.4). A '+' denotes a predicted increase in AR binding, a '-' denotes a decrease and a'=' denotes no predicted change. Multiple symbols denote the magnitude of the predicted change, i.e. '- ' is a greater decrease than '-'.

Position	Base Change	% frequency change in consensus	Predicted effect on AR binding
-36	A→T	17→13	=/-
-35	G→C	96→0	- - -
-34	C→G	8→17	=/+
-33	T→C	17→4	=/-
-32	C→A	100→0	- - -
-31	A→C	58→4	-
-30	G→A	38→13	-
-28	T→G	4→46	+
-27	T→G	67→0	- -
-26	G→C	100→0	- - -
-26	G→A	100→0	- - -
-25	T→G	92→4	- -
-24	A→T	17→54	+
-24	A→C	17→8	=/-
-23	C→T	92→0	- - -
-23	C→A	92→8	- -
-22	T→A	67→0	- -
-20/-21	T→G	40→17	
-19	G→C	21→8	=/+

In light of the results in Fig 4.2, it was decided that the binding buffer to be used for the competition gel mobility shift assay was buffer 3, which contained both ficoll and Hepes-KOH. This buffer allowed formation of a protein-DNA complex which appeared to be stable under the gel conditions used in this assay.

A convenient way of expressing the relative affinities of the competitor oligonucleotides for the protein is the C50 value. The C50 value is the molar excess of competitor oligonucleotide required to reduce the binding of the protein to the labelled probe by 50% (see 2.20). This type of analysis has been widely used when studying promoter mutations in the factor IX gene (Reijnen et al., 1993, Kurachi et al., 1994). The higher the C50 value of a competitor oligonucleotide, the lower is its affinity for the protein.

4.4 The Nature of the Retarded Band

Fig 4.3 shows a typical competition gel mobility shift assay. When lanes 6→8 of Fig 4.3 are examined closely, see Fig 4.4, then the retarded band can be seen to be made up of at least two complexes of differing mobilities. It is well known that steroid hormone receptors bind to their recognition sites as dimers (Tsai et al., 1988, Luisi et al., 1991), and if it is assumed that this protein A-AR fusion protein binds in a similar fashion, then one would only expect one protein- DNA complex to be formed during the binding reaction.

However, the protein used in these gel mobility shift assays is not the full length androgen receptor molecule, but a small portion of the receptor fused to protein A. The portion of the androgen receptor which is fused to protein A contains the DNA binding domain and a portion of the hinge region (Fig 3.1). This fragment of the AR does possess some of the dimerisation signals of the full length protein, but not all of them (see Fig 1.13 and Fig 3.1). The DNA binding domain of the glucocorticoid receptor is known to exist as a monomer in solution (Hard et al., 1990a), and it is thought that two of these monomers bind cooperatively to the two half sites of the palindromic recognition site (Hard et al., 1990b) with the amino acids of the D box of

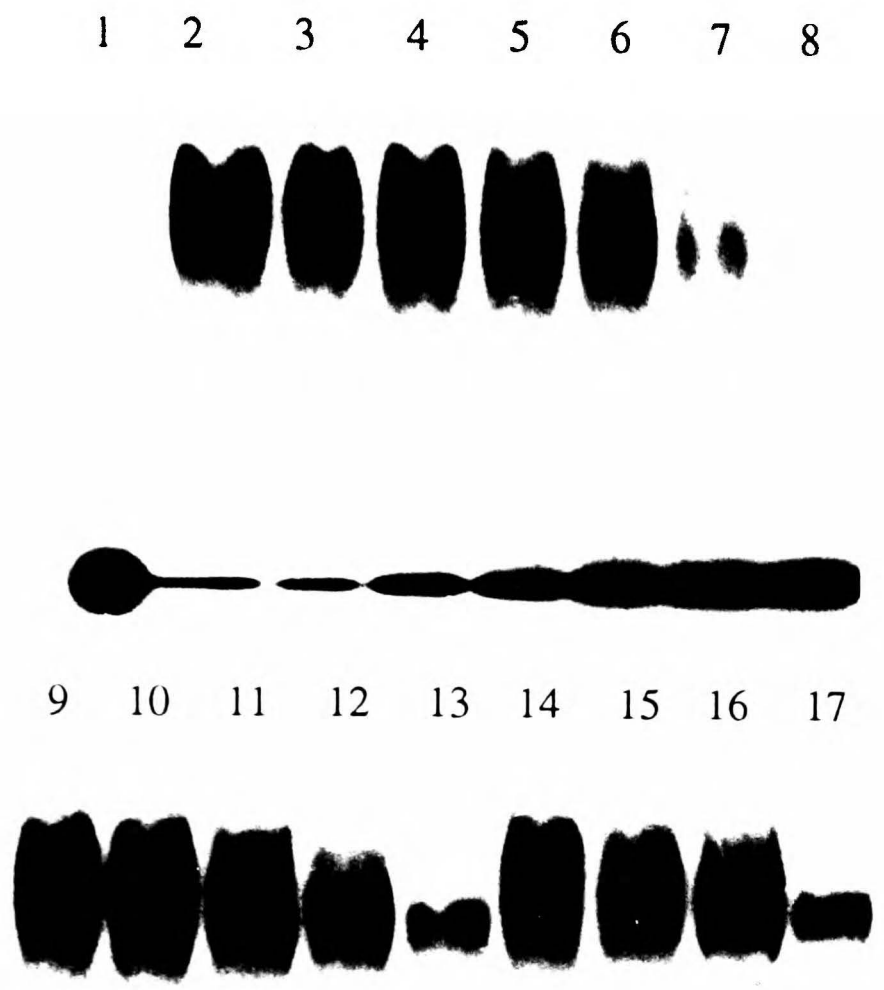


Figure 4.3 Autoradiograph of a typical competition gel mobility shift assay. Lane order, 1= Probe only; 2+3= control gel shift (no competitor); 4→8= wild type oligonucleotide competitor; 9→13= -26 G→C competitor oligonucleotide; 14→17= -36 A→T competitor oligonucleotide. The titrations used in this experiment were as follows, lanes 4→8, and lanes 9→13= 5 fold, 10 fold, 20 fold, 40 fold, 80 fold excess competitor; lanes 14→18= 5 fold, 10 fold, 20 fold, 40 fold excess of competitor.

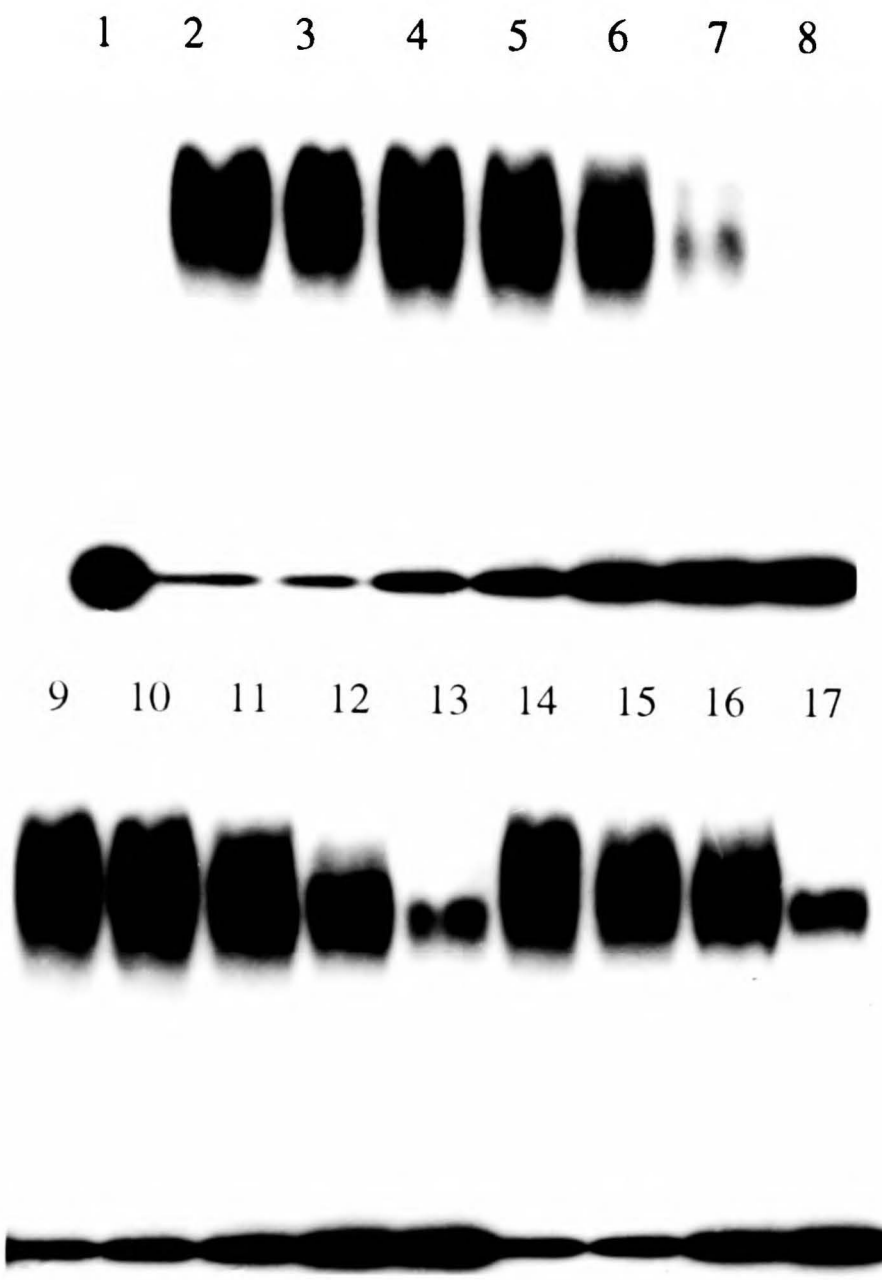


Figure 4.3 Autoradiograph of a typical competition gel mobility shift assay. Lane order, 1= Probe only; 2+3= control gel shift (no competitor); 4→8= wild type oligonucleotide competitor; 9→13= -26 G→C competitor oligonucleotide; 14→17= -36 A→T competitor oligonucleotide. The titrations used in this experiment were as follows, lanes 4→8, and lanes 9→13= 5 fold, 10 fold, 20 fold, 40 fold, 80 fold excess competitor; lanes 14→18= 5 fold, 10 fold, 20 fold, 40 fold excess of competitor.

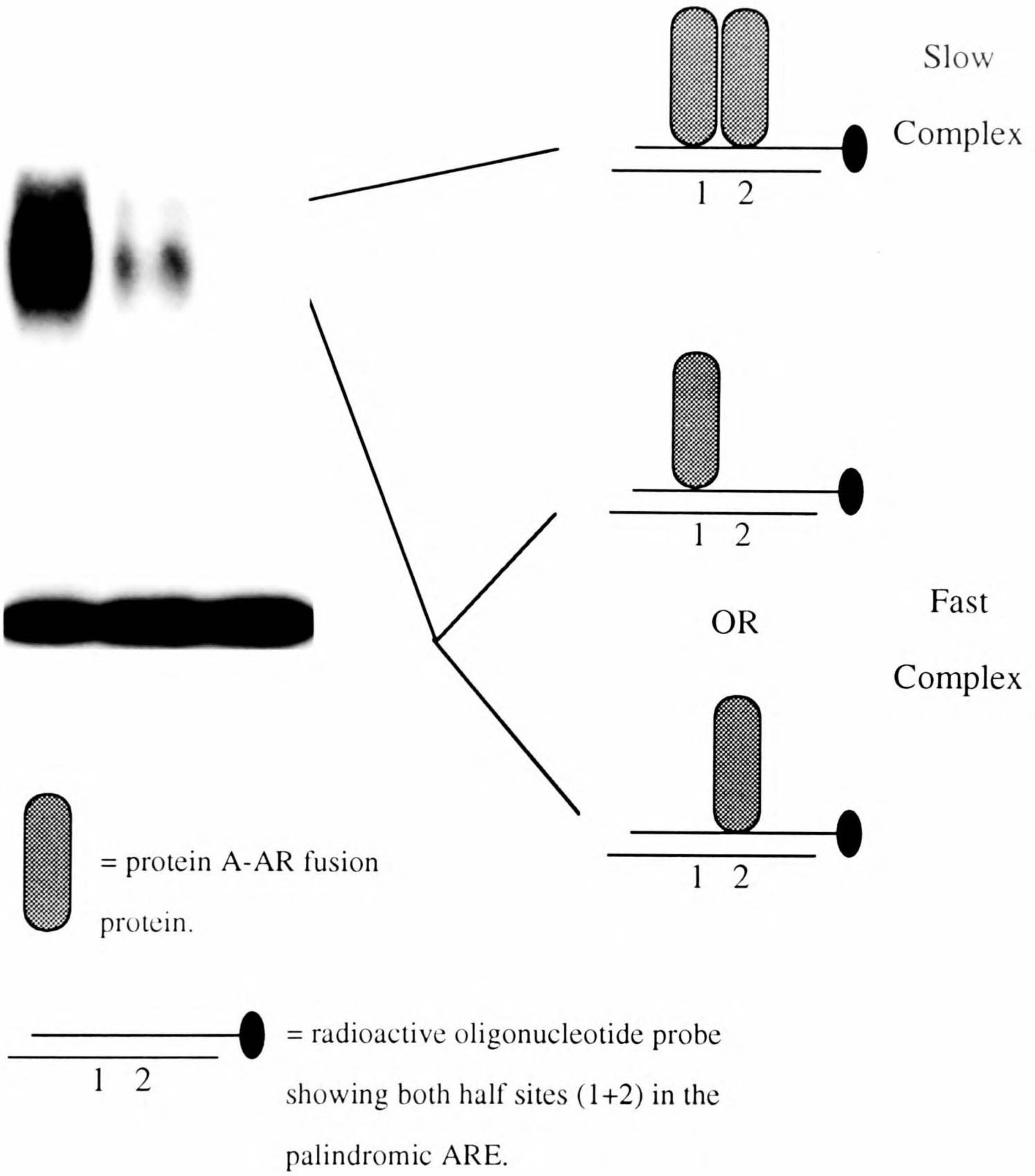


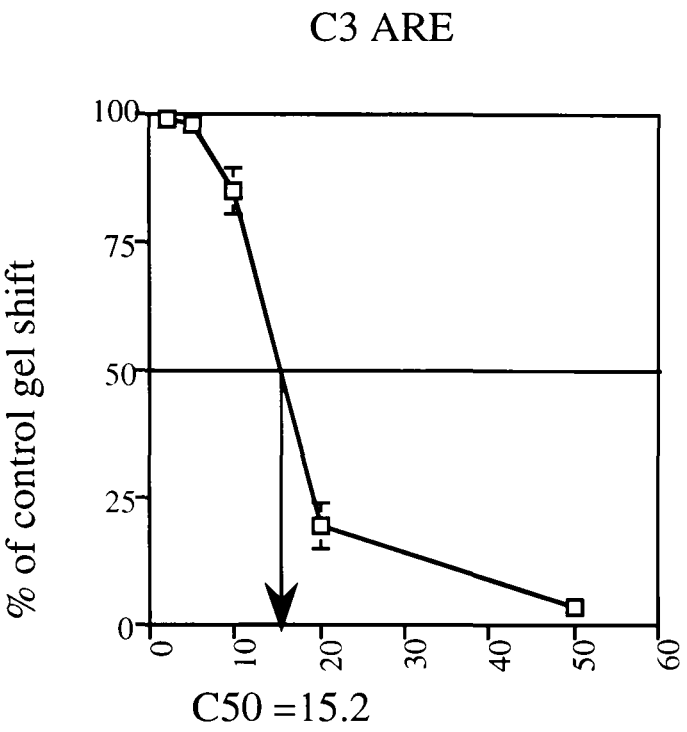
Figure 4.4 Enlarged photograph from lanes 6→8 from Fig 4.3. The retarded complex could be made up of at least two complexes, of different mobilities. The schematic representation of the DNA-protein complexes on the right of the figure suggests the possible identity of these complexes.

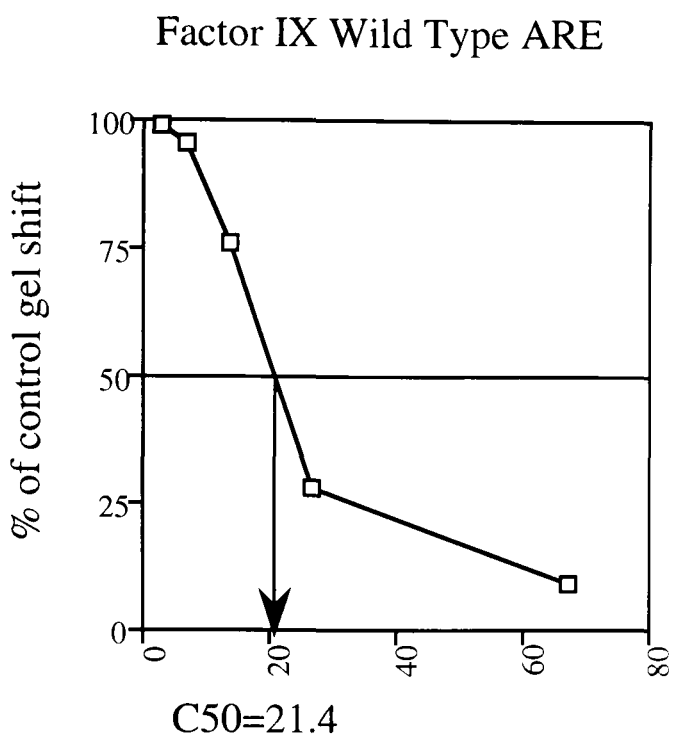
the second zinc finger providing a dimerisation interface between the monomers (see Fig 1.14). The protein A-AR fusion protein is also known to be able to bind to elements which contain only one half site of the palindromic ARE (Roche et al., 1992). Therefore, in this assay two monomers of the protein A-AR fusion protein may be behaving in the same way in this mobility shift assay, and the slow and fast bands in the retarded complex could correspond to two molecules or one molecule of protein A-AR fusion protein binding to the factor IX ARE oligonucleotide, respectively (see Fig 4.4).

4.5 Results of the Competition Gel Mobility Shift Assays

Each oligonucleotide competitor was tested in a competition assay with factor IX wild type sequence as labelled probe. An autoradiograph of a typical competition gel shift assay is shown in Fig 4.3. The competing oligonucleotides in this assay were the factor IX wild type sequence and the -26 G→C, -36 A→T and -35 G→C mutant oligonucleotides. The competition gel mobility shifts were then quantitated using a phosphorimager (see Chapter 2). This analysis allowed the percentage of the radioactivity shifted in each lane to be calculated and then the proportion of the control shift (see lane 2+3, Fig 4.3) that remained could be calculated. These data can then be graphically represented (Figs 4.5→4.10) and the C50 value, the molar excess of competitor which competes out 50% of the control gel shift, can be read off the graph. A comparison of the C50 values for the C3 positive control ARE, wild type factor IX sequence and the -26 G→C mutation is shown in Fig 4.11. The C50 values of the oligonucleotides tested in the gel mobility shift assays are summarised in Table 4.3.

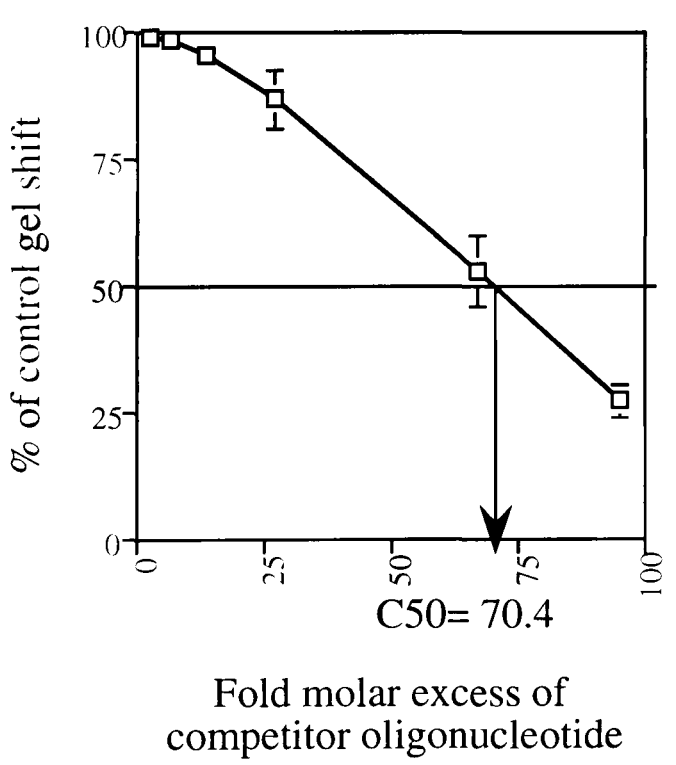
If Tables 4.2 and 4.3 are compared, many similarities between the effects of the mutations predicted from the study of Roche et al., (1992) and the actual results obtained from the competition gel shifts can be seen. Table 4.3 shows firstly, that mutations at positions -6, -3, 3, 4, and 6 (nucleotides -35, -32, -26, -25, and -23, see Fig 4.1) disrupt the binding of the fusion protein to the ARE and secondly, that mutations at positions -7, -2, and 2 (nucleotides -36, -31, and -27, see Fig 4.1) also





-26 G→C Mutation
(Haemophilia B Brandenburg)

-26 G→A Mutation
(UK 232)



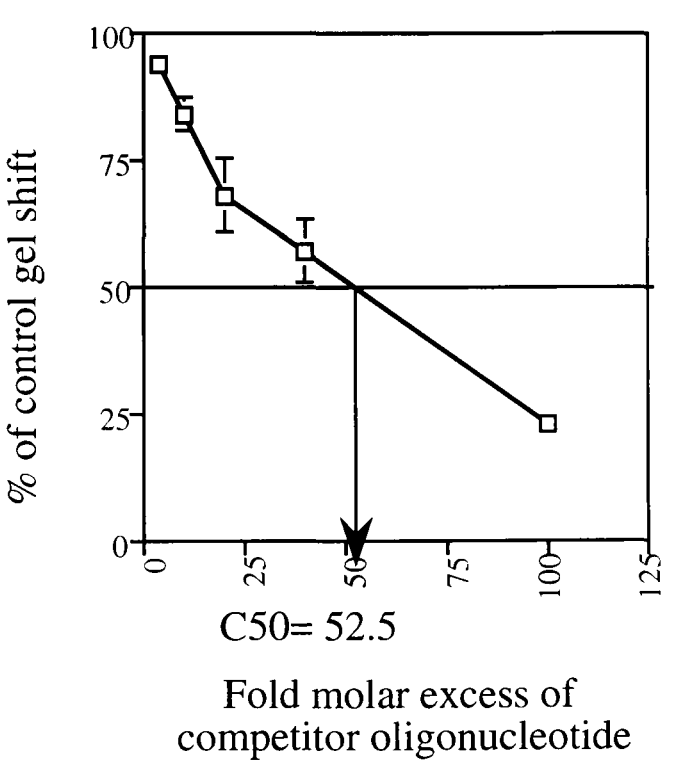
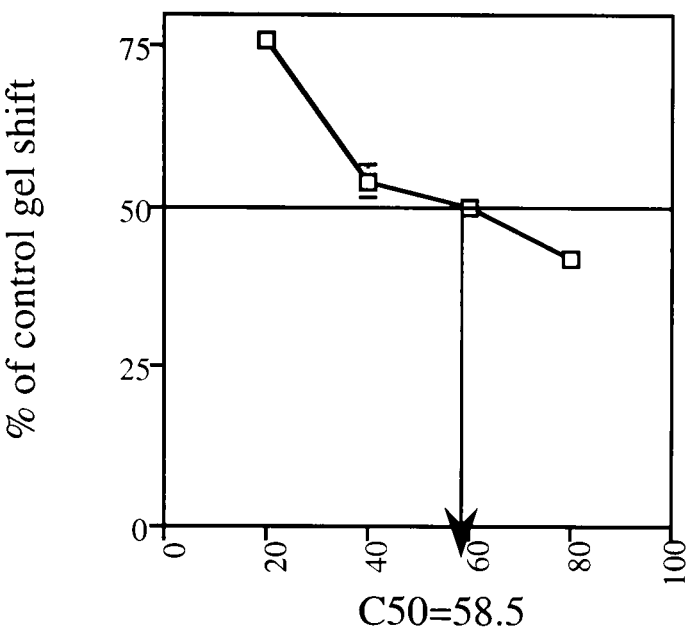


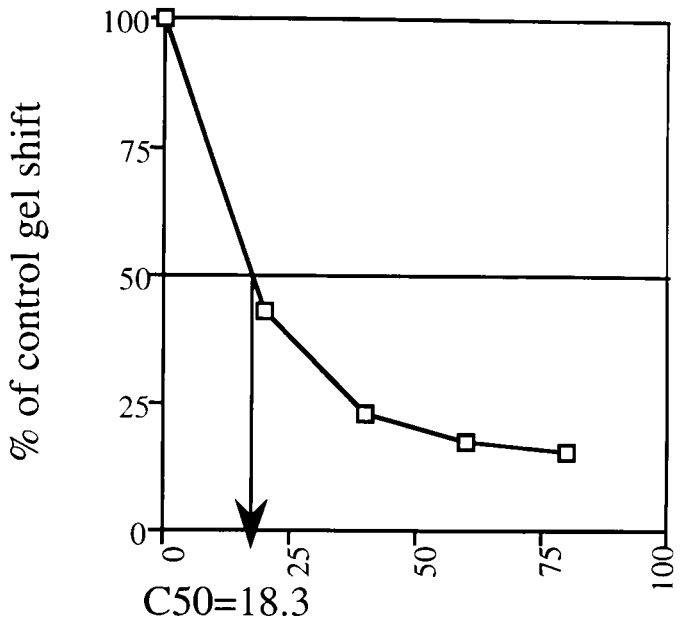
Figure 4.5 Graph of the binding data for factor IX wild type,

C3, -26G→C, and -26G→A AREs.

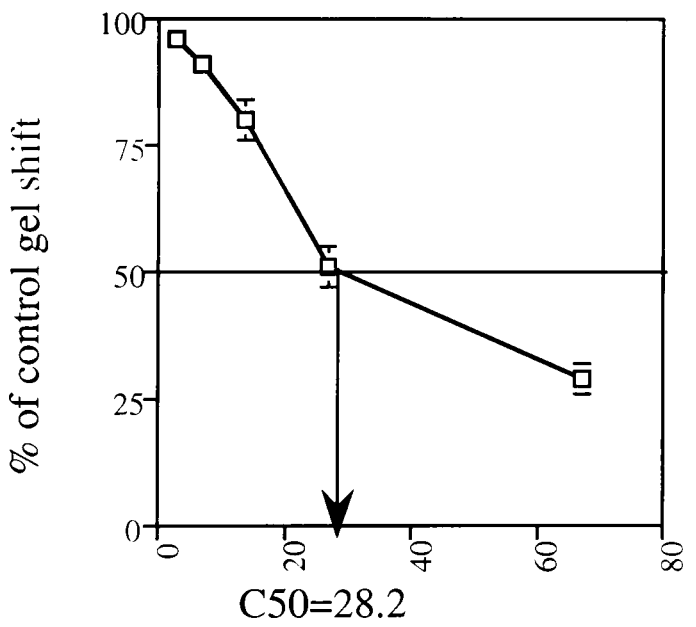
-23 C→T Mutation
(HB 388)



-19 G→C Mutation
(HB396)



-36 A→T Mutation



-35 G→C Mutation

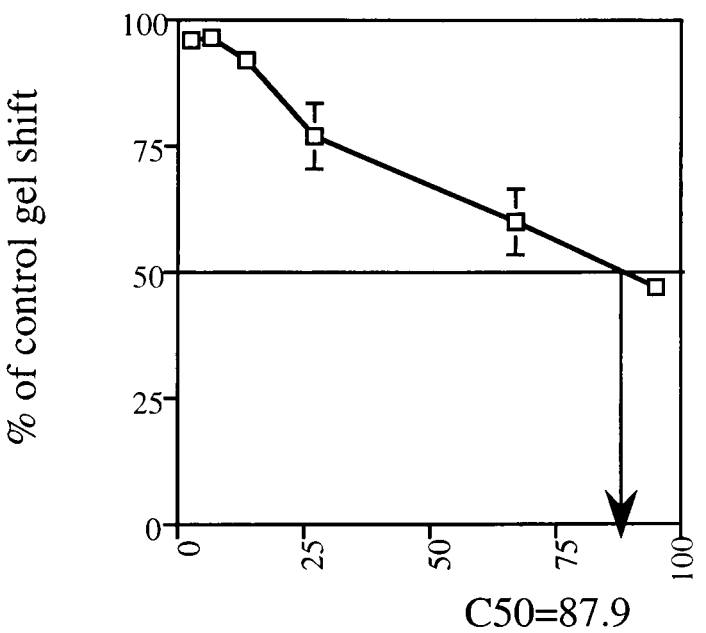
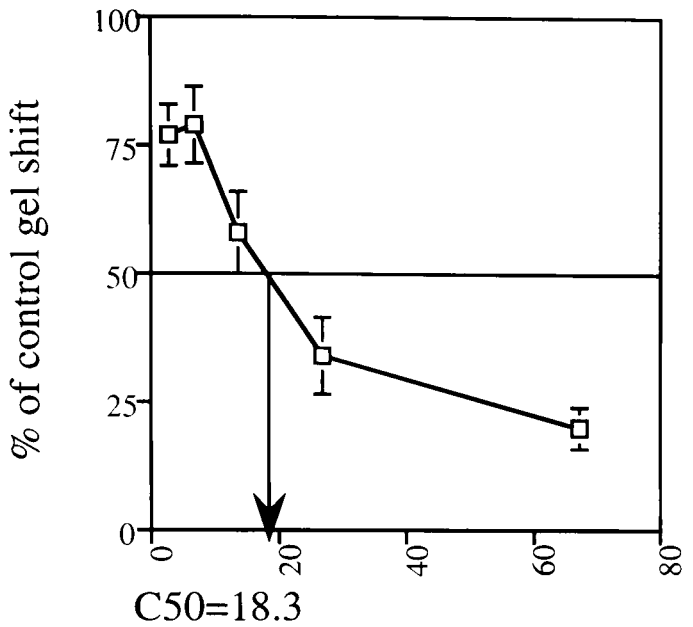


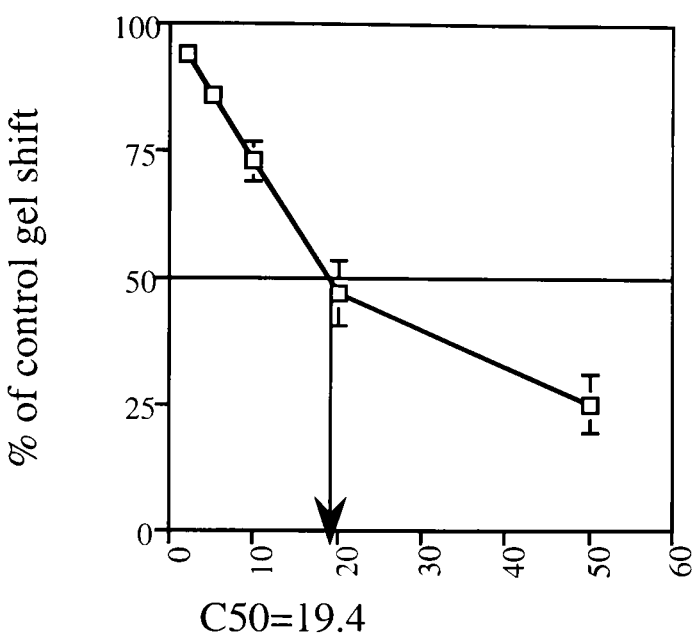
Figure 4.6 Graphs of the binding data for -23 C→T, -19 G→C,

-36 A→T, and the -35 G→C mutant AREs.

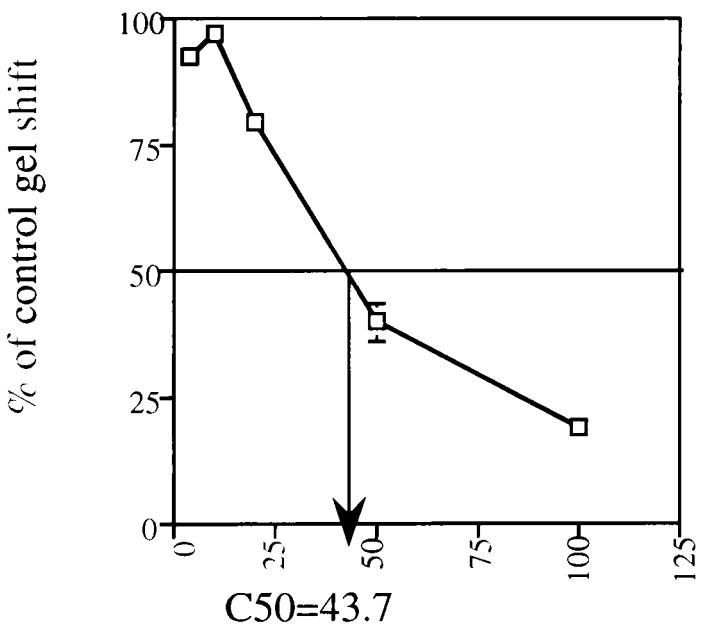
-34 C→G Mutation



-33 T→C Mutation



-32 C→A Mutation



-31 A→C Mutation

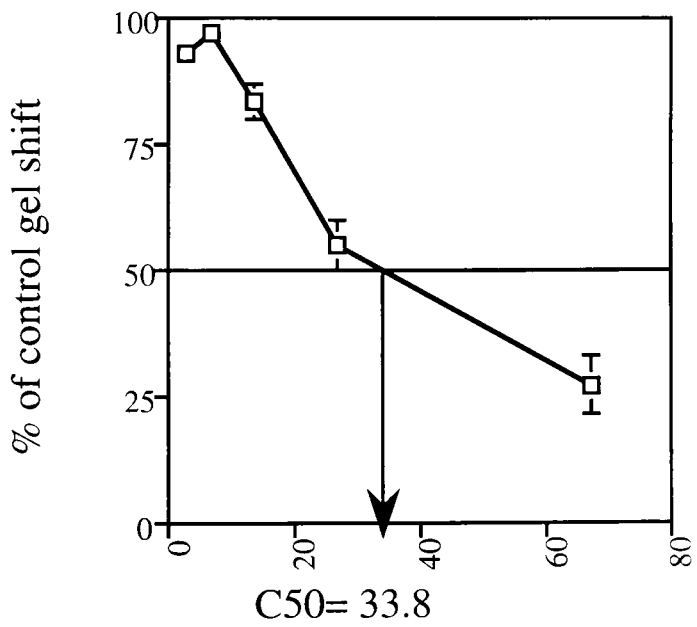
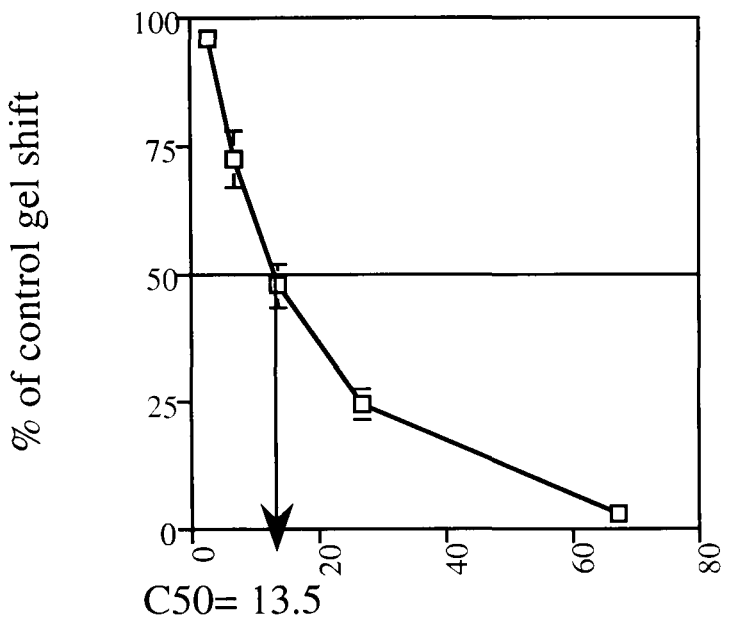


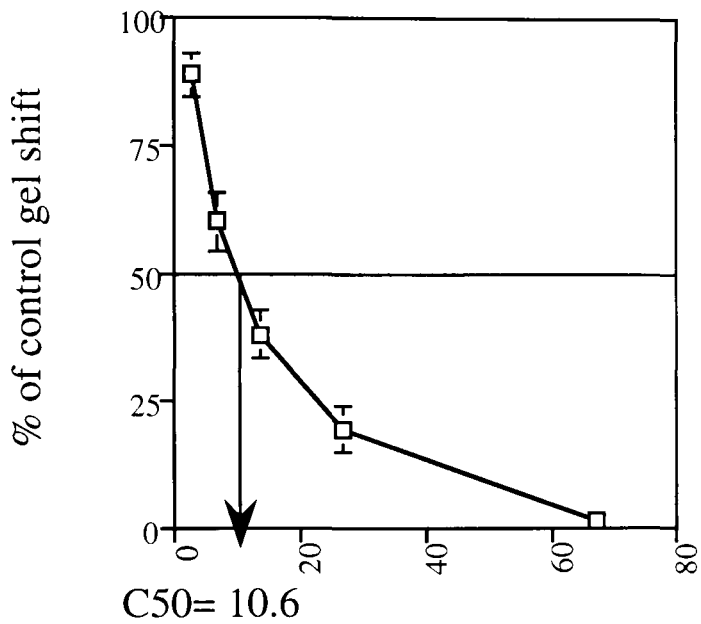
Figure 4.7 Graphs of the binding data for -34 C→G, -33 T→C,

-32 C→A, and the -31 A→C mutant AREs.

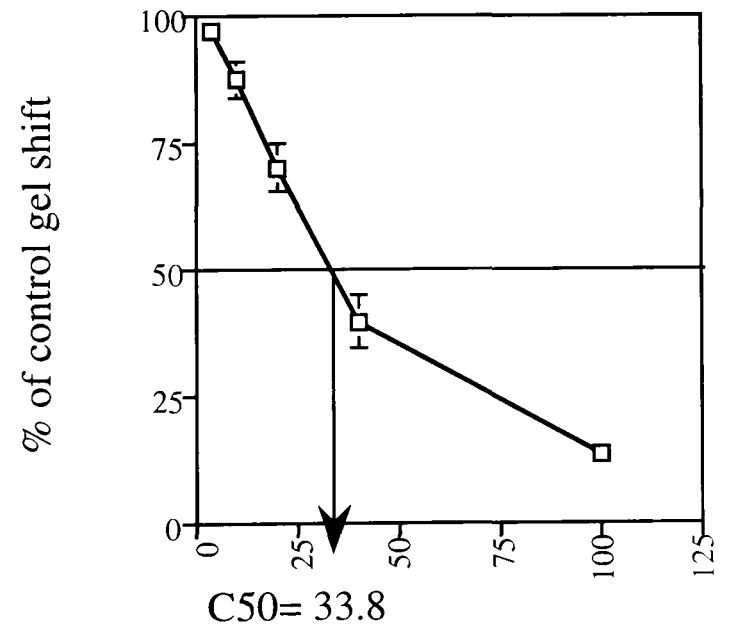
-30 G→A Mutation



-28 T→G Mutation



-27 T→G Mutation



-25 T→G Mutation

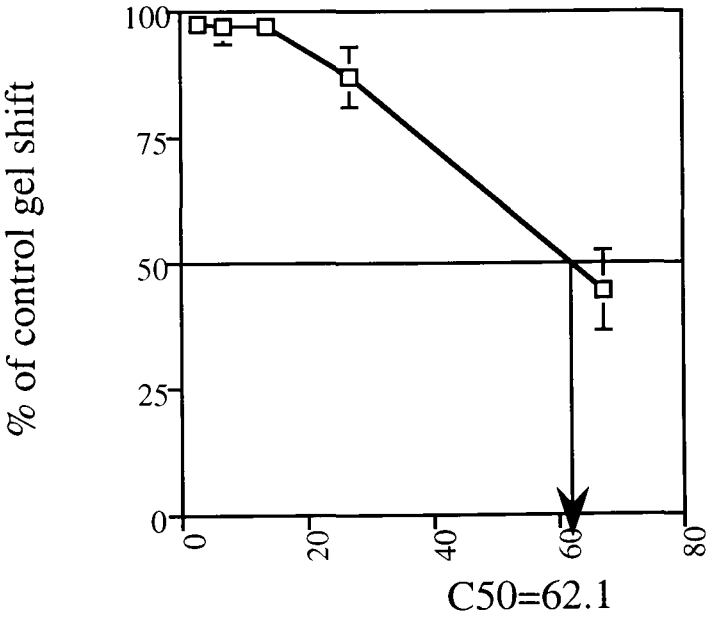
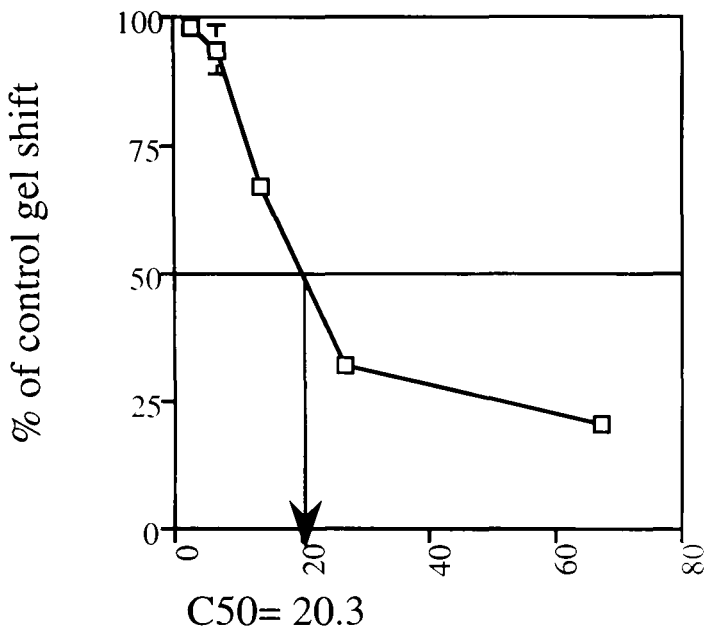
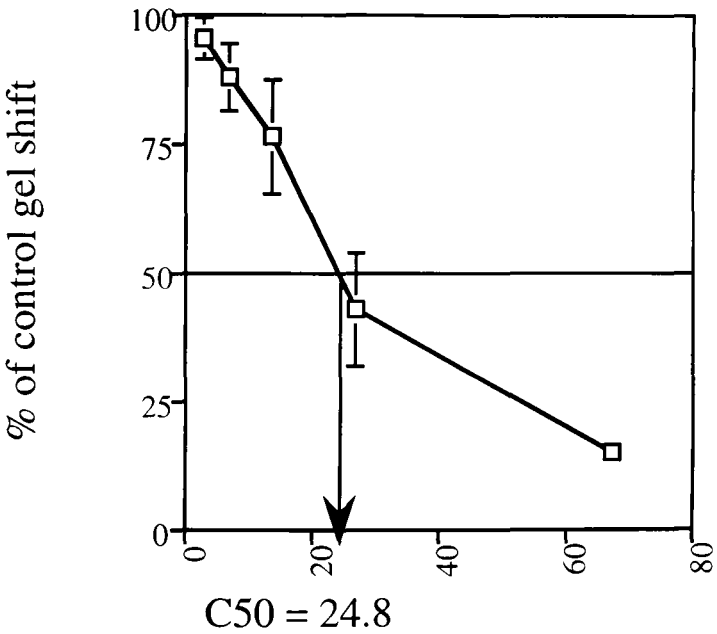


Figure 4.8 Graphs of the binding data for -30 G→A, -28 T→G, -27 T→G, and the -25 T→G mutant AREs.

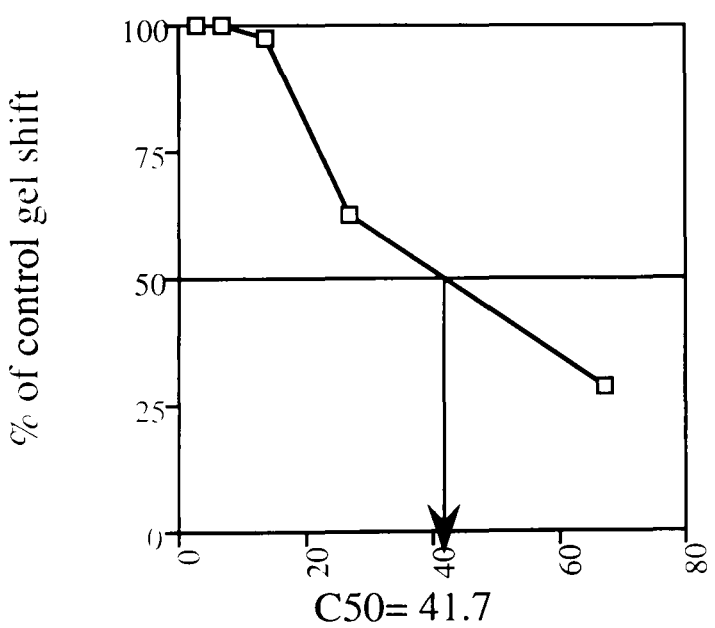
-24 A→T Mutation



-24 A→C Mutation



-23 C→A Mutation



-22 T→A Mutation

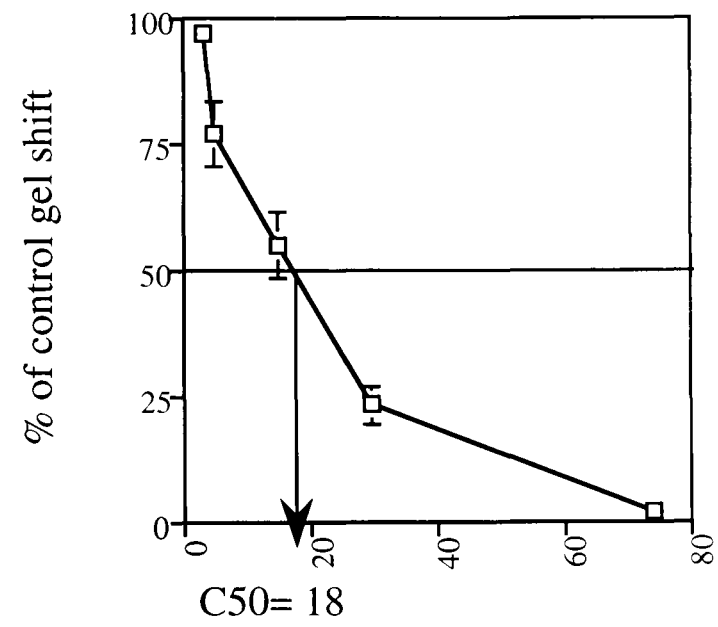


Figure 4.9 Graphs of the binding data for -24 A→T, -24 A→C, -23 C→T, and the -22 T→A mutant AREs.

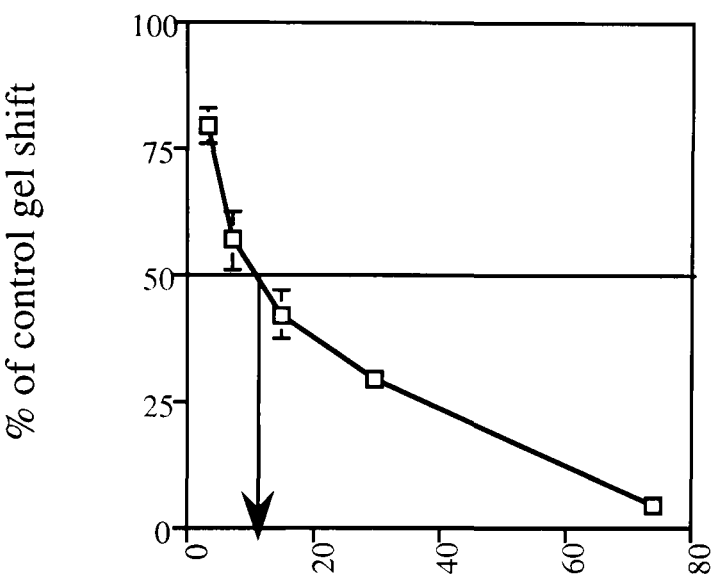


Figure 4.10 Graph of the binding data for -22/-20 T→G mutant ARE.

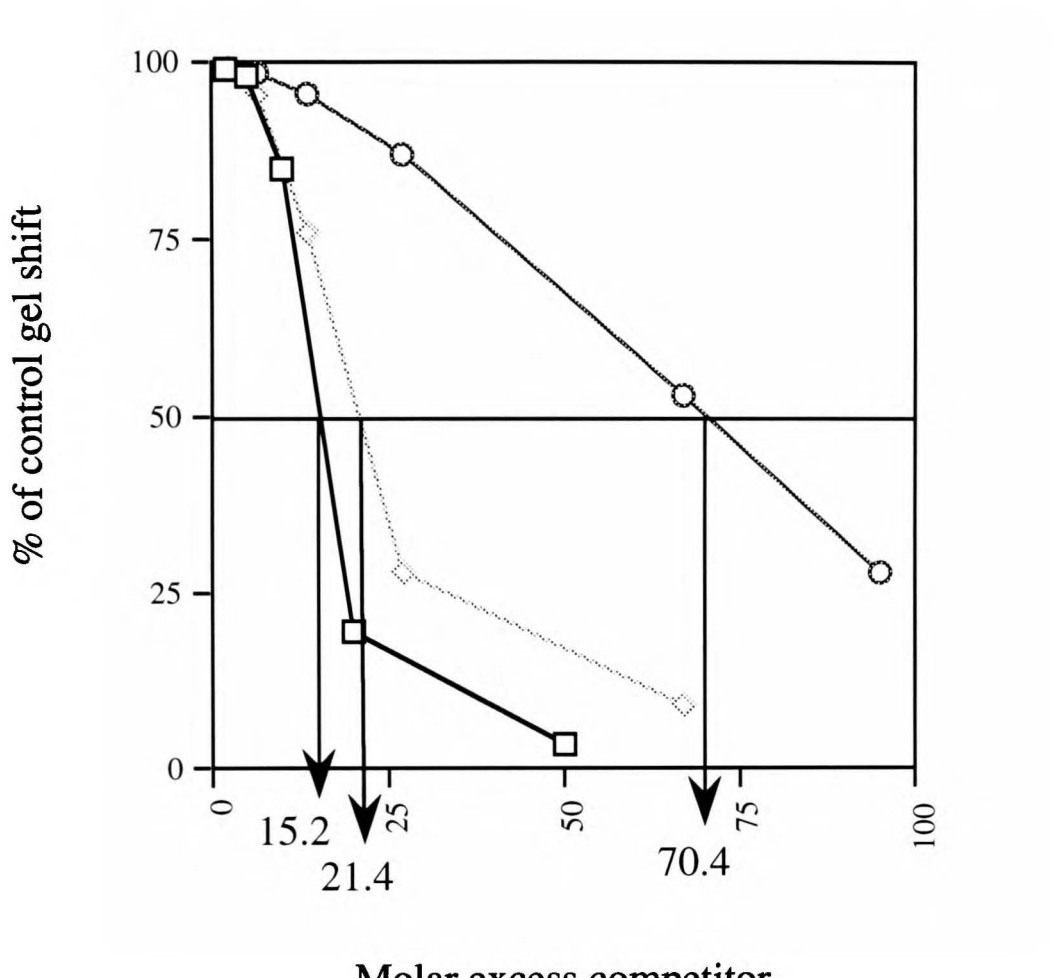


Figure 4.11 A direct comparison of the C50 values for three oligonucleotides: wild type (◇), -26 G→C (O), C3 ARE (□).

Table 4.3 The C50 values for the oligonucleotides used in the competition gel shift assays. The symbols used in the table are the same as those used in Table 4.2. For convenience, the predicted results from Table 4.2 are included.

Oligonucleotide	C50	Effect on AR binding	Predicted effect
C3	15.2		
Wild Type	21.4		
-36 A→T	28.2	-	=/-
-35 G→C	87.9	- - -	- - -
-34 C→G	18.3	=/+	=/+
-33 T→C	19.4	=/+	=/-
-32 C→A	43.7	- -	- -
-31 A→C	33.8	-	
-30 G→A	13.5	+	-
-28 T→G	10.6	+	+
-27 T→G	33.8	- -	- -
-26 G→C	70.4	- - -	- - -
-26 G→A	52.5	- - -	- - -
-25 T→G	62.1	- - -	- -
-24 A→T	20.3	=	+
-24 A→C	24.8	=/-	=/-
-23 C→T	58.8	- -	- - -
-23 C→A	41.7	-	- -
-22 T→A	18	=/+	- -
-21/-20 T→G	11.3	+	-
-19 G→C	18.3	=/+	=/+

impair the binding of the fusion protein to the ARE, but to a lesser degree, see Table 4.3.

Possibly the most striking differences are seen when the flanking positions are examined. In these positions the mutations are predicted to have a deleterious effect to AR binding at positions , -20/-21, and -19. All of these mutations proved to be either beneficial or neutral. However, the prediction for the spacer mutation at -28 did correlate with the C50 result, suggesting that position 1 (see Fig 4.1) in the ARE influences AR binding. The mutation at -22 was predicted to have a detrimental effect (Table 4.2) but did not show any disruption in the C50 values for the ARE oligonucleotide (Table 4.3). It is possible that in the context of the factor IX ARE this position is not important for the binding of AR as the study of Roche et al., (1992) only defined this position as "moderately well conserved" (Fig 4.1). Other mutations in the factor IX ARE did not show any disruption of AR binding e.g. positions -5, -4 and 5, nucleotides -34, -33, and -24). This is due to the fact that the wild type ARE is imperfect, i.e. there are a number of positions which do not adhere to the consensus defined by Roche et al., (1992) (Fig 4.1). Therefore, the mutation of the ARE at these non-consensus positions would not be expected to disrupt the binding of the fusion protein to the ARE. Of the two mutations made at nucleotide -24, one was intended to improve AR binding (A→T), the other to marginally disrupt (A→C), see Tables 4.1 and 4.2. The A→T mutation at this site does not show any significant improvement, and the A→C shows only a small disruption. Therefore, it appears that nucleotide -24 (position 5, see Fig 4.1) plays no significant role in the recognition of this element by AR.

Of nineteen mutant ARE oligonucleotides tested, only three deviated substantially from the values predicted for them using the consensus study of Roche et al., (1992), which illustrates how powerful consensus studies are. All three mutations in this region which have been found in patients with haemophilia B also showed a disruption of the factor IX ARE (-26 G→C, -26 G→A, -23 C→T), and the one patient

mutation tested which lay outside the ARE failed to disrupt the binding of the protein A-AR fusion protein (-19 G→C).

4.6 Discussion

The ARE consensus element consists of two six base pair half sites which form an imperfect palindrome (Roche et al., 1992). When the C50 values of the factor IX mutants in equivalent positions in each half site are overlaid (Fig 4.12), it can be seen that the half sites do appear to behave in an equivalent fashion, with the exception of position 4/-4 (positions -25 and -33). As mentioned earlier, section 4.5, the factor IX ARE does not adhere to the consensus ARE in some positions, and position 4 (nucleotide -33), see Fig 4.13, is one of those positions. Mutating this nucleotide will not disrupt the binding of the fusion protein to the oligonucleotide as the base change is not making the sequence a worse fit to the consensus.

From Table 4.3 the bases which seem critically important for the contact of the protein A-AR DBD fusion protein with the factor IX ARE, and which increase the C50 more than 100%, are in positions -6, -3, 3, 4, and 6 (see Fig 4.13). These positions correspond to nucleotides -35, -32, -26, -25, and -23 in the factor IX promoter region (Fig 4.13). The consensus study by Roche et al., (1992) identified the positions 3,-3; 4,-4; and 6,-6 as the most highly constrained. The factor IX ARE has bases critical for the contact of the protein A-AR fusion protein at five of these six positions (Fig 4.13). Mutations at 2 of these positions (-26 and -23, see Chapter 6) are known to lower factor IX plasma levels in patients and, in the case of the two mutations at -26, to prevent the recovery of these factor IX levels typically shown by patients with mutations in the promoter of the factor IX gene (Giannelli et al., 1996, see Fig 1.20). The A→T mutation at position -7 (nucleotide -36) disrupts the binding of AR to the factor IX ARE despite the fact that the factor IX ARE does not match the consensus nucleotide for this position, see Fig 4.1. It is possible that either G or A is allowable at this position and that any deviation from that sequence will disrupt the binding of AR to the sequence. Roche et al., (1992) found that adenine was the second most common

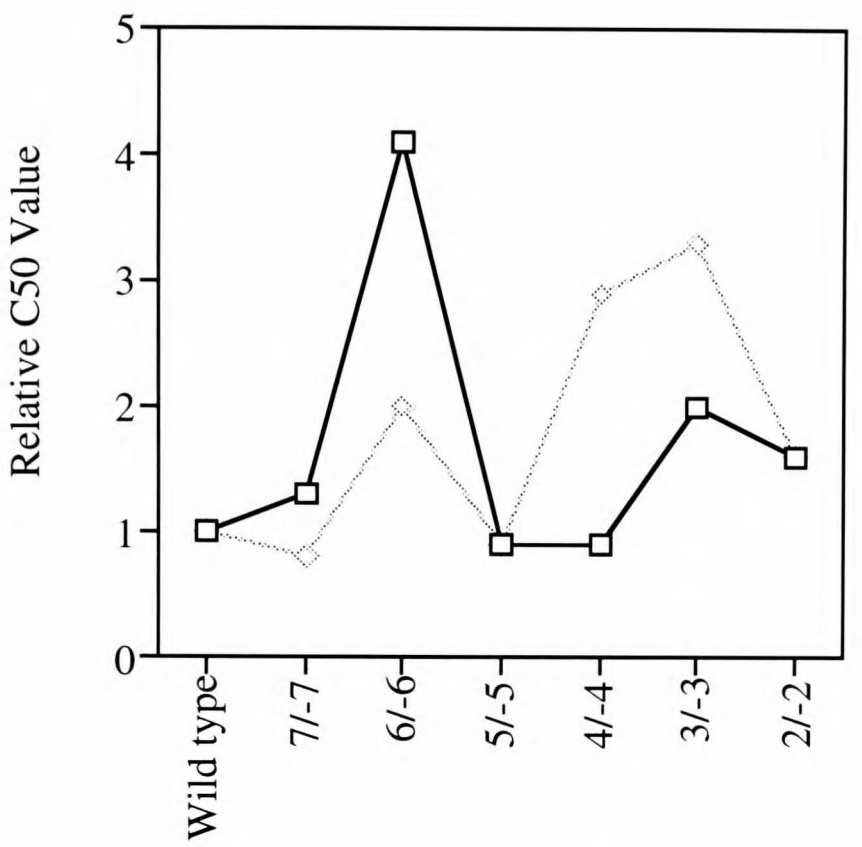


Figure 4.12 Graph of the relative C50 values for each equivalent position in the half sites of the factor IX ARE. The wild type factor IX ARE is given an arbitary value of 1. The numbering used is that from Roche et al., (1992). 5' half site = (□), 3' half site= (◇), see Fig 4.1 and Fig 4.13.

base found at this position in their studies, and Fig 4.1 illustrates that seven of nine naturally occurring ARE sequences compared have A in this position (only one has G in this position).

In this study positions at which a mutation does impair binding but does not increase the C50 more than 100% are at positions -7, -2 and 2 (nucleotides -36, -31, and -27, see Fig 4.13). Roche et al., (1992) also identified positions 2/-2 and 7/-7 as having an effect on AR binding to the ARE (Fig 4.13). Mutations at 3 of these four positions also result in an increase of the C50 value for the factor IX ARE. A mutation at position 7 (nucleotide -22, see Fig 4.13) in the factor IX ARE does not show an increased C50 value in the study even though it is predicted to by the consensus study. This could be due to the nature of this particular sequence somehow compensating for the mutation at this position, or could merely serve to highlight the fact that most natural ARE sequences identified so far do not adhere to the consensus sequence perfectly, and in some cases can deviate from it quite substantially (see Fig 4.1, and Cleutjens et al., (1996). All of the other mutations in the factor IX ARE were either neutral, or appeared to improve the binding of the protein A-AR fusion protein to the region, i.e. the C50 was lowered. The T→G mutation in the spacer region of the factor IX ARE at position 1 (nucleotide-28, see Fig 4.13) was predicted to increase the binding of the protein A-AR fusion protein and the C50 value indicates that it does (Table 4.3). Therefore, the presence of a G or A residue at position 1 in the ARE element may promote AR binding.

Where this study cannot support the previous study is in the effect of the 3' flanking nucleotides in the binding of AR to the element. The results summarised in Table 4.3 suggest that the mutation of the 3' flanking nucleotides (-21/-20, and -19) has no detrimental effect on AR binding. Moreover, this result does support the patient data which is available concerning the 3' flanking positions in the factor IX ARE. Mutations have been identified in patients at all three positions -21, -20 and -19 in the factor IX promoter (Reitsma et al., 1988, Reijnen et al., 1993), but the patient with the -19 mutation is still in childhood (Ketterling et al., 1995, see Chapter 6). The mutation

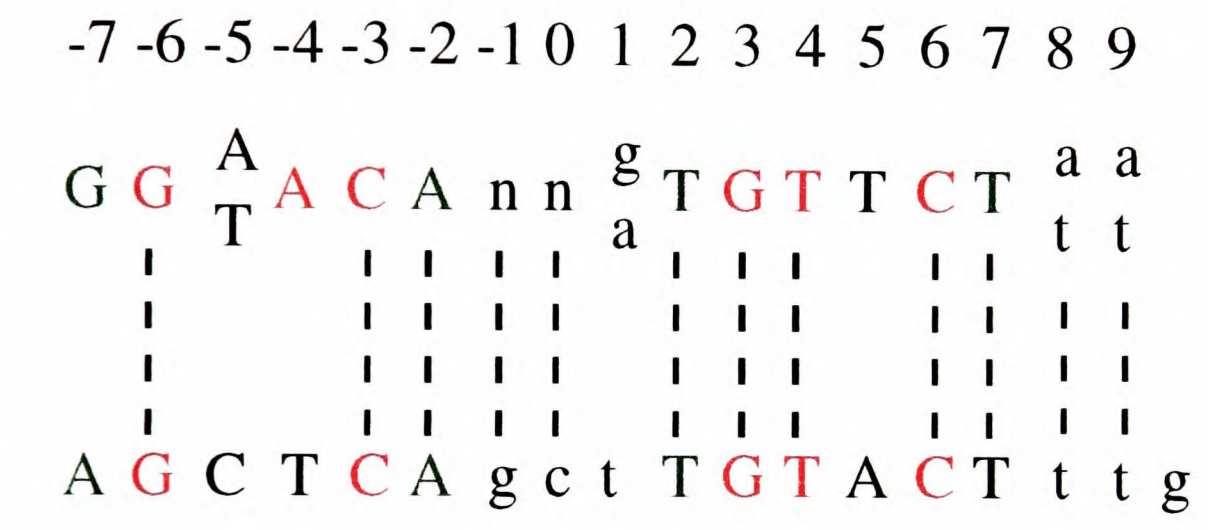


Figure 4.13 A summary of the binding data obtained with the factor IX ARE mutant oligonucleotides and a comparison of the factor IX related results (lower sequence) with the consensus sequence of Roche et al., 1992 (upper sequence) is included. Mutations at the positions in red badly impair the binding of the protein A-AR fusion protein to the ARE (see Table 4.3) and mutations at the positions in green impair the binding of the fusion protein to a lesser extent (see Table 4.3).

T→G at -21 might be expected to disrupt the binding of the AR, whereas the T→A mutation at -20 would probably not be expected to do so (see Table 4.2, and Roche et al., 1992). Both of these patients show a clear increase in their factor IX clotting levels through puberty (Giannelli et al., 1996), their adult clotting levels are 14% (and still rising) in the case of the -21 T→G mutation and 37% in the case of the -20 T→A patient. These levels suggest that the ARE sequences proposed to be involved in the recovery of these patients are active.

4.7 Summary

In summary, the binding study of nineteen mutations in the factor IX ARE largely supports the ARE consensus sequence derived by Roche et al., (1992). The data presented in this chapter has highlighted the positions at which a mutation is likely to result in the disruption of the ARE in the factor IX promoter region and has also brought to light some differences between the behaviour of this sequence and its predicted behaviour from the consensus study . This data is of critical importance when considering the genetic counselling of families and patients with these mutations in their pedigree given the contrasting clinical prognosis of the Leyden and Brandenburg phenotypes (see Fig 1.20), see Chapters 6 and 8.

Chapter 5: *In vivo* Analysis of Mutations in the Factor IX ARE Using a Transient Transfection Assay.

5.1 Introduction

It has already been shown (Crossley et al., 1992) that tetramers of the factor IX ARE (-42→-17) could confer androgen responsiveness upon a Herpes simplex virus thymidine kinase promoter (HSV tk) in the CAT reporter plasmid pBLCAT2 (Luckow and Schutz, 1987). These reporter plasmids were cotransfected along with an AR expression vector (pKC3AR, see section 2.6) into HeLa cells cultivated in the presence of 5α-dihydrotestosterone. Tetramers of this element were used in these assays as single factor IX ARE elements do not allow measurable levels of activation of transcription and dimers were only found to activate transcription three fold in these assays (Crossley et al., 1992). In the only previous study on the ARE consensus, Roche et al., (1992) found that even the sequence which they identified as having the highest affinity for AR only activated the transcription of the tk promoter 3.7 fold as a monomer in a similar assay to that used by Crossley et al., (1992). Therefore, it has now become accepted practice to multimerise such elements when studying mutations in them (Cleutjens et al., 1996). Adler et al., (1993) and Cleutjens et al., (1996) show that not all AREs are simple consensus binding sequences alone, but are more complex constructions which require multiple androgen response elements or other transcription factor binding sites for full activity. Hence, even though a mutation may reduce the transactivation of transcription in the assay system used in this chapter, one must remember that non-receptor factors may also influence the action of AR on this element.

5.2 Mutant Oligonucleotide Cloning

The mutant oligonucleotides used in the competition gel mobility shift assays were synthesised with four base pair overhangs at their ends (see Appendix 1) corresponding to the restriction cutting sites for the enzymes *Xba*I and *Bam*HI. These sites allowed the oligonucleotides to multimerise in a ligation reaction (see 2.9.2 and

2.23, Fig 5.1). The oligonucleotide ligation mixture produced a ladder of products (Fig 5.1) which could then be cloned into the reporter gene vector pBLCAT2 (Luckow and Schutz, 1987) cut with *Xba*I and phosphatased (see 2.9.1). In the mixture of oligonucleotide multimers, the only fragments which would be expected to clone into the *Xba*I cut vector were dimers, tetramers, hexamers etc. which had two *Xba*I-compatible overhangs at their 3' and 5' termini (see Fig 5.2). After each cloning experiment, the candidate clones were analysed by restriction digestion with *Pst*I. This enzyme cuts either side of the *Xba*I cloning site in pBLCAT2 (see Fig 5.3A), and so will produce a restriction fragment which contains any insert which has been cloned. The digests were analysed on 6% polyacrylamide gels as detailed in 2.5.

Fig 5.3B shows an ethidium bromide stained gel of restriction digests of minprep DNA (see section 2.4) from three clones from a cloning experiment with the wild type oligonucleotide. Lane 1 is a digest of the parent plasmid, pBLCAT2, which was used as a negative control size marker (133 bp), lane 2 shows a positive control digest which was also used as the size marker for the gel. The plasmid used in this digest was the wild type tetramer construct used in the previous study of Crossley et al., (1992). It can be clearly seen that lanes 4 and 5 contain two bands, as does lane 2 but in lane 3 only the lower band of the two is visible. It was suspected that the lower of the two bands corresponded to an oligonucleotide dimer construct, and the upper to a tetramer oligonucleotide construct. In order to determine more clearly what fragments form these bands, it was decided to characterise them by purifying them, and performing further restriction analyses with the purified fragments.

5.3 Characterising the Oligonucleotide Reporter Clones

The two bands present in lanes 2 and 5, Fig 5.3B, were characterised by isolation and purification of the two restriction fragments from the gel, digestion of the purified fragment with *Sau*96I, Klenow labelling, and the separation of the labelled fragments on a 6% denaturing polyacrylamide gel (see 2.24 and Fig 5.4). A known DNA dideoxy sequencing reaction was run alongside to act as an exact size marker for

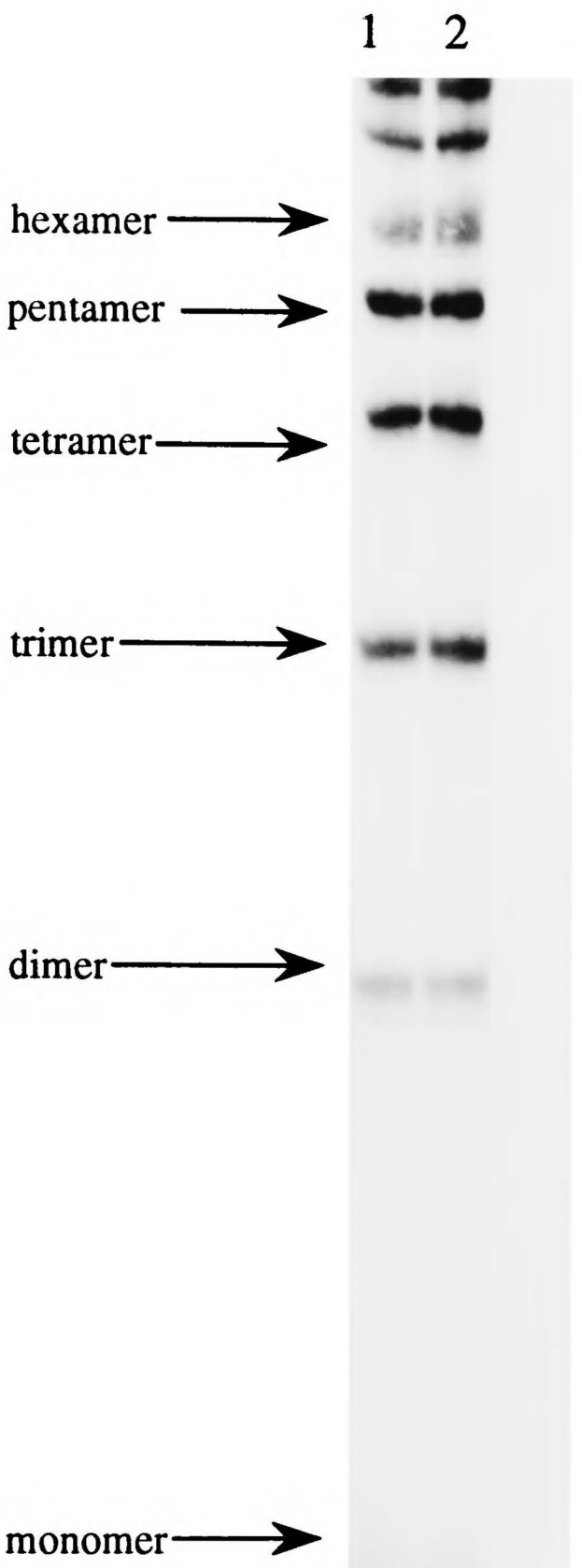


Figure 5.1 Autoradiograph of a 10% native polyacrylamide gel showing the multimerisation of the wild type factor IX ARE (lane 1) and -26G→C mutated (lane 2) oligonucleotides in a ligation reaction (see Methods for details). The monomer band was identified by its comigration with unligated, labelled double stranded oligonucleotide (results not shown).

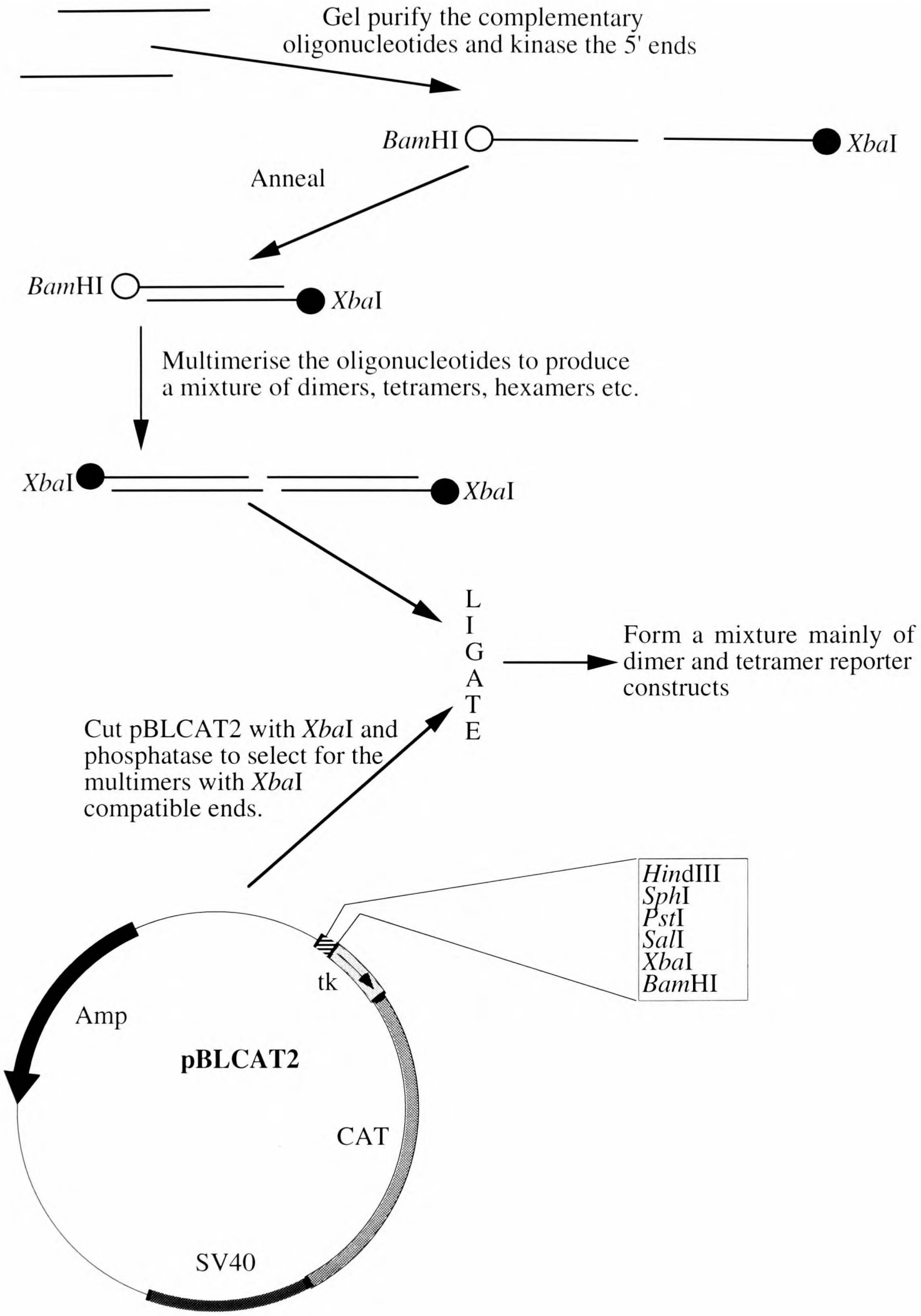
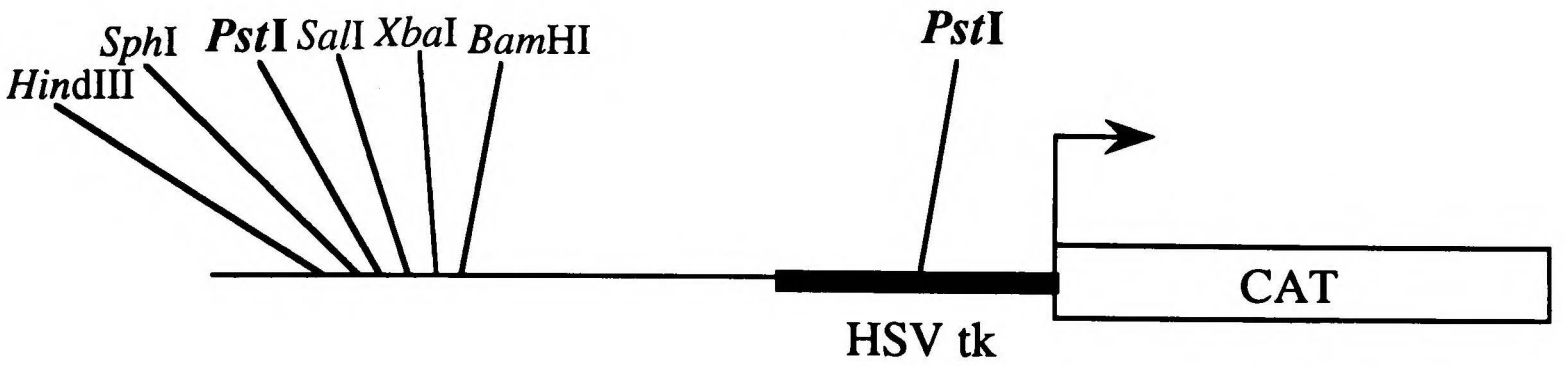


Figure 5.2 Formation of dimer and tetramer oligonucleotide reporter constructs in pBLCAT2.

A



B

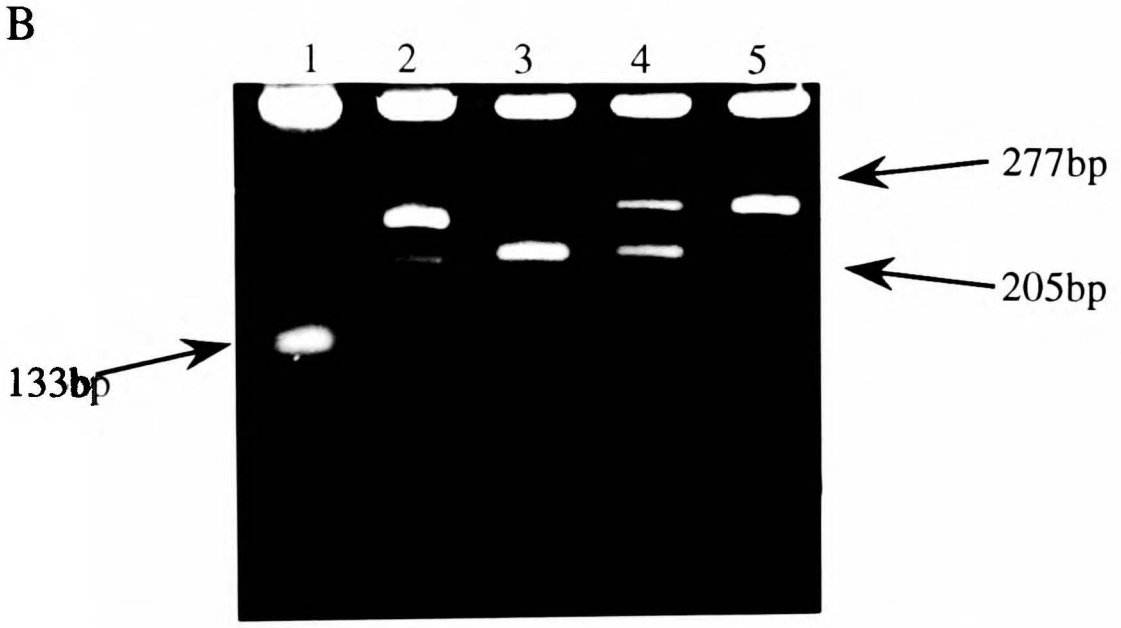


Figure 5.3 (A) The multiple cloning site of the plasmid pBLCAT2. The oligonucleotide multimers are cloned into the *Xba*I site. *Pst*I cuts either side of the *Xba*I site and releases a fragment containing the insert. Bands of 133bp, 205bp, and 277bp are produced by the vector with no insert, the dimer reporter plasmid and the tetramer, respectively. (B) An ethidium bromide stained, 6% native polyacrylamide gel showing *Pst*I digests of pBLCAT2 (lane 1), pARE4 (lane 2), and 3 clones (lanes 3→5). Lane 1 is the negative control, lane 2 is the positive control. The bands of 133, 205, and 277bp are arrowed.

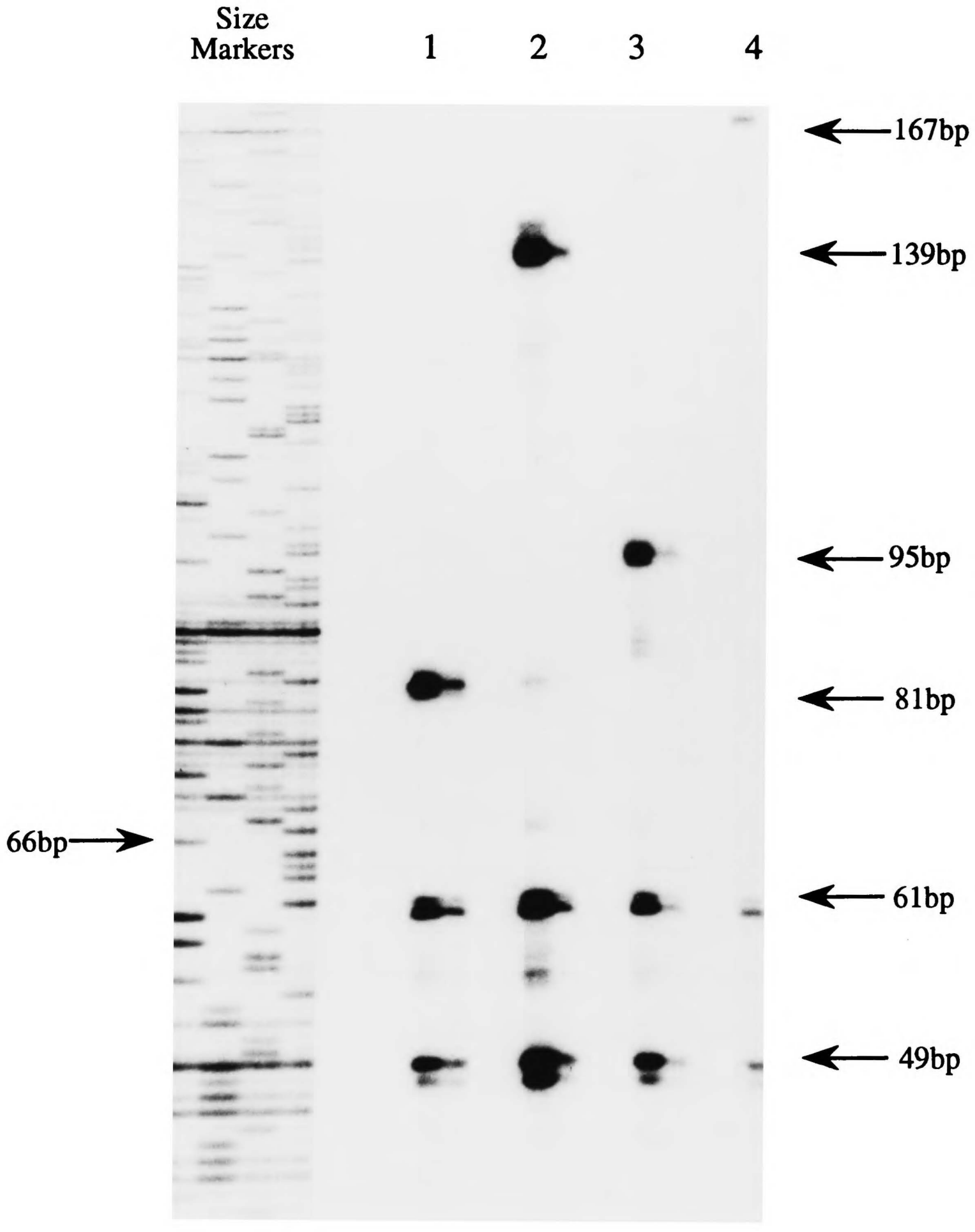
the fragments. The two bands from the restriction digests of both the previously constructed wild type oligonucleotide (-42→-17) tetramer (Crossley et al., 1992) and the newly constructed wild type oligonucleotide (-40→-9) tetramer were purified and analysed in this way. The dimer oligonucleotide fragment and the tetramer oligonucleotide fragment would be expected to give different patterns of labelled restriction fragments on this gel (see Table 5.1).

Table 5.1 The fragments expected from the *Sau*96I digest of *Pst*I restriction fragments purified from restriction digests of dimer and tetramer reporter constructs. The sizes for the dimer and tetramer fragments from the plasmid ARE4, containing copies of -42→-17 (*). from Crossley et al., (1992), and the wild type tetramer construct with the -40→-9 oligonucleotide (§), used in this thesis, are shown.

Construct	Fragment size (bp)	Nature
Dimer	49	Vector derived
	61	Vector derived
	81*/95§	Oligonucleotide dimer
Tetramer	49	Vector derived
	61	Vector derived
	139*/167§	Oligonucleotide tetramer

Fig 5.4 shows the gel of the *Sau*96I restriction profile of the *Pst*I restriction fragments isolated from the two wild type oligonucleotide tetramer reporter plasmids. On the dideoxy sequencing reactions the band corresponding to 66bp is arrowed. All four fragments show the invariant 61 and 49bp bands which belong to parent vector sequence. The fragments digested in lanes 1 and 3 were suspected of containing oligonucleotide dimers and those digested in lanes 2 and 4 tetramers. In Fig 5.4 the 81bp band in lane 1, and the 95bp band in lane 3 correspond to the dimer oligonucleotide fragment present in this *Pst*I restriction fragment. The 139bp band in

Figure 5.4 6% denaturing polyacrylamide gel of the *Sau*96I profiles of the purified *Pst*I restriction fragments. The size markers used were DNA dideoxy sequencing reactions of known size. Lane 1= profile of the smaller fragment from lane 2, Fig 5.3B, lane 2= larger fragment from lane 2, Fig 5.3B, lane 3= smaller fragment from lane 5, Fig 5.3B, lane 4= larger fragment from lane 5, Fig 5.3B. The sizes of all of the *Sau*96I restriction fragments are indicated and these concur with the fragments expected in Table 5.1.

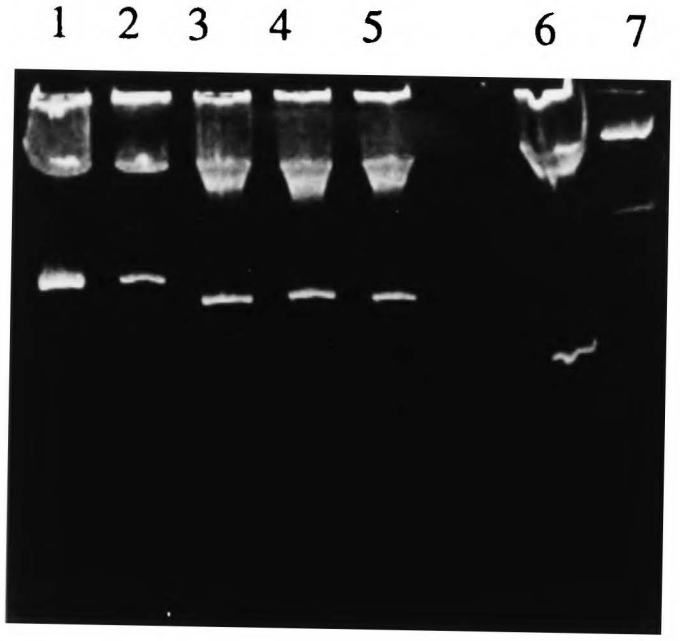

Size Markers
1
2
3
4
167bp
139bp
95bp
81bp
66bp
61bp
49bp

lane 2 and the 167bp band in lane 4 correspond to the oligonucleotide tetramer fragment in this *Pst*I restriction fragment. As the oligonucleotides used to make these constructs were different sizes, the dimer and tetramer bands in the previously constructed plasmids were smaller by 12bp and 24bp, respectively.

This experiment strongly suggests that both the previously used wild type oligonucleotide tetramer clone, and the newly constructed clone are mixtures and contain both dimer and tetramer oligonucleotide inserts in each clonal population. It was possible that the sample of tetramer plasmid could have been contaminated with dimer constructs early in their production, so the plasmid mixture obtained from the miniprep of the wild type clones was retransformed into the *E.coli* strain NM554 (see 2.2). When clones were isolated from this experiment, and plasmid DNA prepared from them, it was found that they, too, contained mixtures of the dimer and tetramer fragments in their plasmid populations (see Fig 5.5A). This result suggested that the four repeated copies of the factor IX ARE (which is an imperfect palindrome) sequence were inherently unstable in the bacterial strain used.

In a further effort to isolate pure tetramer clones, the plasmid mixtures were transformed into the *E.coli* strains DL795, JC8679 and SURE host strains which may have been able to propagate the tetramer construct without also producing the dimer construct. Fig 5.5B shows a *Pst*I restriction digest of plasmid DNA isolated from minipreps in the *E.coli* strains JC8679 (lanes 1→4), DL795 (lanes 5→7), and SURE (lanes 8 and 9). Lane 1 contains the positive control digest of the original plasmid construct. The digests show that the tetramer plasmids appear to be unstable in these three host strains too. The only clones which show the tetramer restriction band also possess the dimer band. In an effort to find stable reporter constructs, the oligonucleotides were also cloned into pGLtk (see 2.6), a luciferase reporter gene vector. The same type of instability of the tetramer reporter constructs was observed (data not shown).

A



B

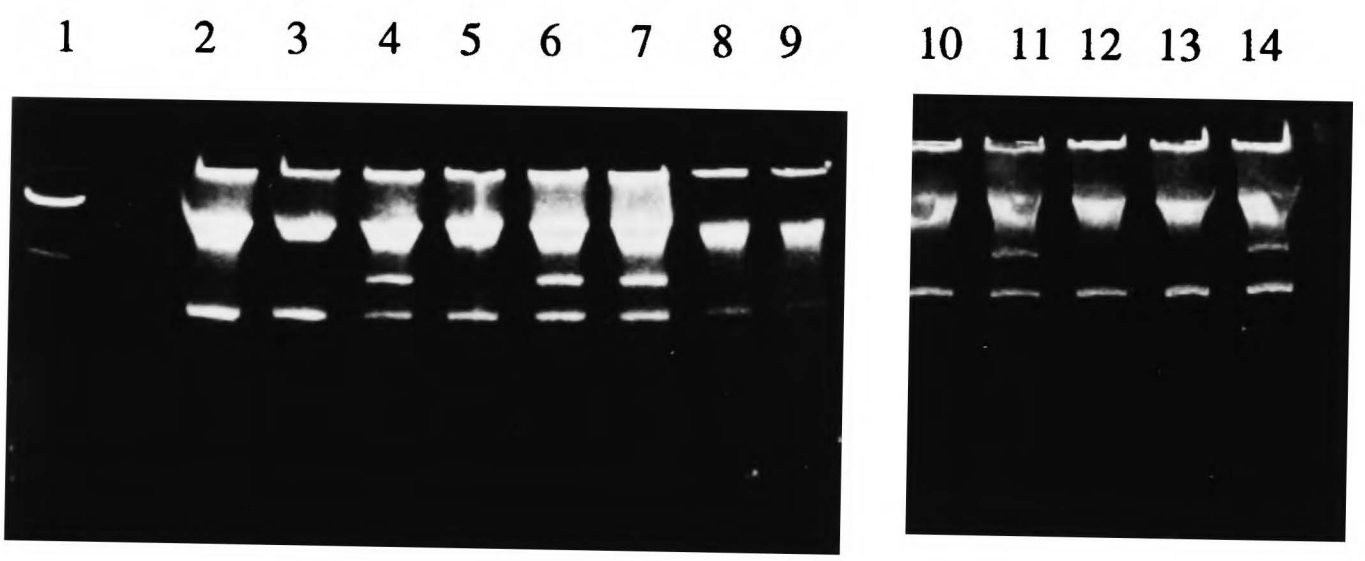


Figure 5.5 (A) Ethidium bromide stained gel of *Pst*I digests of clones from the retransformation of NM554 with a tetramer/dimer mixture of plasmids. Lanes 1→5= individual clones, lane 6=negative control (pBLCAT2), lane 7=positive control (original plasmid mixture). (B) Ethidium bromide stained gels of *Pst*I digests of plasmid DNA from clones isolated from the transformation of *E.coli* DL795 (lanes 2→5), JC8679 (lanes 6→9), and SURE (lanes 10→14), with the tetramer/dimer reporter construct. Lane 1= positive control (original plasmid stock).

5.4 Testing the Oligonucleotide Dimer and Tetramer/Dimer Mixture Reporter Clones *in vivo*

As the tetramer oligonucleotide clones could not be separated from the contaminating dimer oligonucleotide clones, it was decided to compare the activities of the dimer and dimer/tetramer mixture constructs in an *in vivo* transient transfection assay in HeLa cells. The assay system used was the same as that employed by Crossley et al., (1992) (see 2.25). The oligonucleotide-reporter constructs were tested for their ability to confer androgen responsiveness on a heterologous promoter in a reporter gene vector. The promoter used was the HSV tk promoter, which controlled the expression of a CAT reporter gene in the plasmid pBLCAT2 (Luckow and Schutz, 1987). Constructs which were known to contain only dimer oligonucleotide inserts and clones which were mixtures of tetramer and dimer oligonucleotide inserts were tested in these transient transfection assays. The constructs tested were the factor IX wild type construct and the -26 G→C Brandenburg mutation, a mutation known to disrupt the factor IX ARE (Crossley et al. 1992). These two constructs would probably illustrate the full range of hormone response which would be observed using this type of construct in this assay- the wild type being the most active and the Brandenburg mutation being the most inactive. The results for each of the constructs were expressed relative to the background reporter gene activity in the absence of AR and 5α-dihydrotestosterone. Fig 5.6 shows that the AR and 5α-dihydrotestosterone can transactivate the wild type factor IX ARE dimer oligonucleotide-reporter construct 3.5 fold above the background reporter activity. The activity of the -26 G→C mutated dimer oligonucleotide in the presence of AR and 5α-dihydrotestosterone was less than the background reporter activity, an overall transactivation of 0.75 fold. This range of activity for the plasmid constructs was considered unsuitable for the *in vivo* analysis of the mutant ARE oligonucleotides studied in Chapter 4.

The tetramer clones which were studied in the transient transactivation assays were selected from gels of the *Pst*I restriction digests which indicated that they consisted of 60→80% tetramer and only 20→40% dimer. Fig 5.6 shows that the

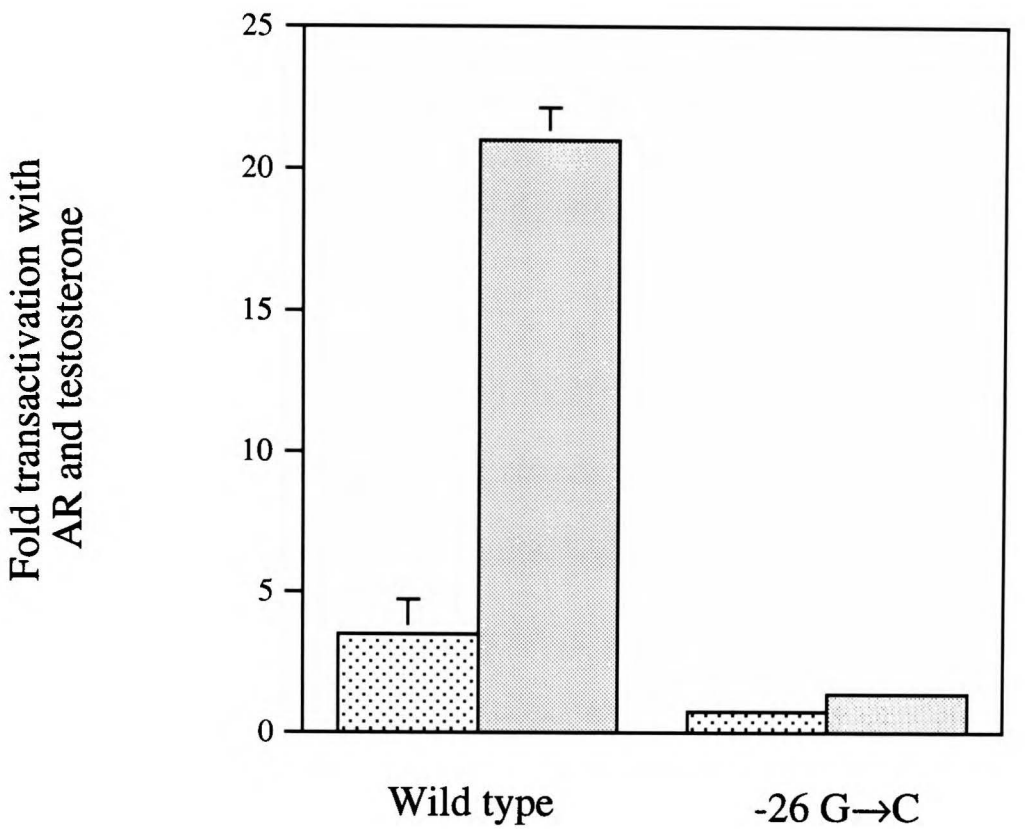


Figure 5.6 Transactivation of the Herpes simplex virus thymidine kinase promoter in the plasmid pBLCAT2 in HeLa cells with two copies ( ), or four copies ( ) of the factor IX wild type and -26 G→C Brandenburg ARE elements cloned immediately upstream (see Fig 5.2). The results are the average of at least three, independent transfections. Standard errors are shown where large enough to be visible.

hormone response conferred on the tk promoter by four copies of the wild type factor IX ARE was approximately 20 fold and that of the -26 G→C Brandenburg construct was 1.3 fold. This differential of response was considered suitable to detect a range of mutations in the factor IX ARE and so tetramer/dimer mixture constructs were obtained for each oligonucleotide mutation studied in the competition gel mobility shift assays, with the exception of the -21/-20 mutation as the data for one of these positions was already known (Crossley et al., 1992). The activity of the dimer constructs in the assay was low enough that any contaminating dimer construct in the tetramer clones was assumed to have an insignificant effect on the final data.

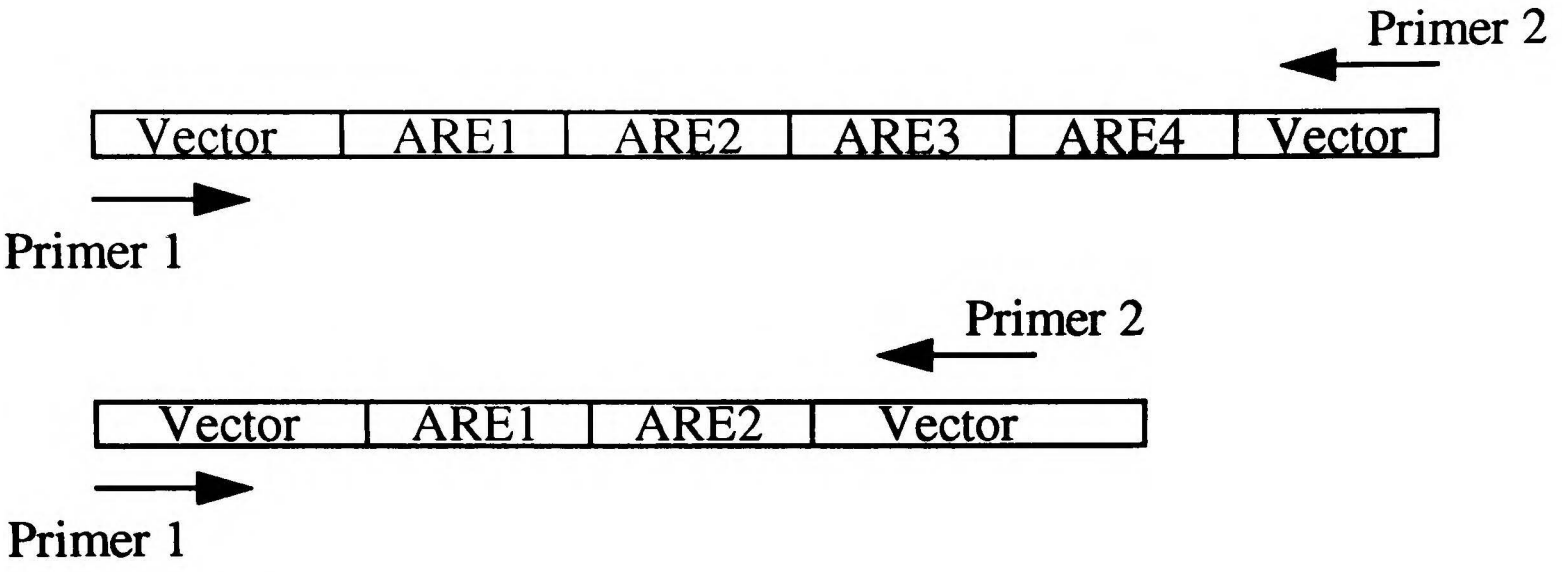
To ensure that no synthesis errors were present in each of the oligonucleotide tetramer/dimer clones, each plasmid mixture was sequenced as shown in Fig 5.7A. Two sequencing primers were designed to obtain sequence from either side of the oligonucleotide cloning site. This allows the sequence of both the tetramer and the dimer construct in the mixtures to be determined (see 2.11 and Fig 5.7A). Fig 5.7B shows a typical dideoxy sequencing gel of a set of reactions with the -25 T→G clone. Lanes 1→4 contain the set of reactions from primer 1, and lanes 5→8 the set of reactions from primer 2 (see Fig 5.7A). All of the oligonucleotide sequence read from the two primers is identical as expected (see 2.11) and the mutations are arrowed (see Fig 5.7B). Only two copies of the oligonucleotide can be read from each set of reactions as the dimer in the plasmid mixture then begins to interfere with the sequencing (see bracketed region, Fig 5.7B). Only clones sequenced in this way, and found to be correct, were used in the transient transfections.

5.5 Transfection Control Plasmids

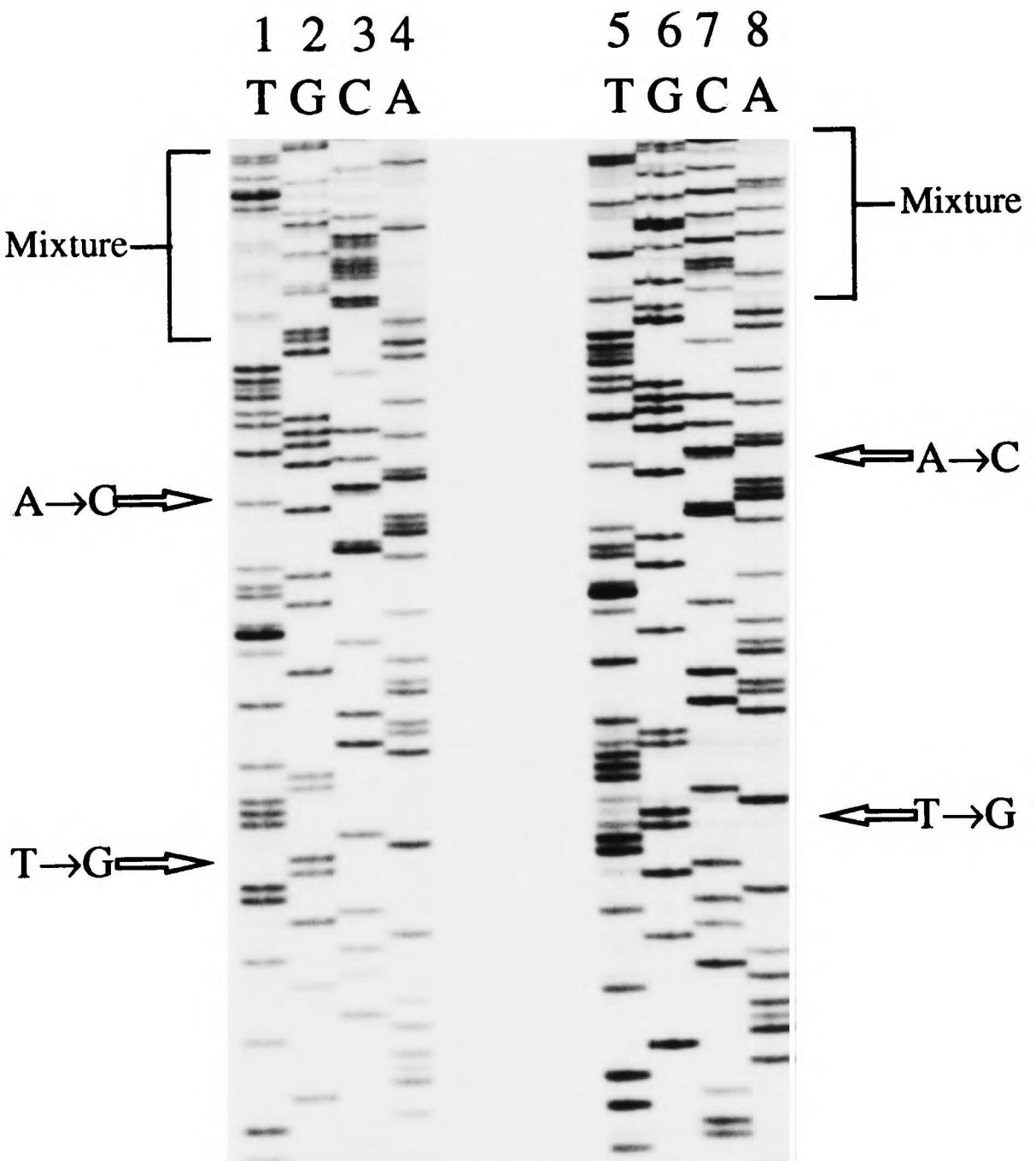
Transfection experiments are routinely controlled for transfection efficiency by the cotransfection of an additional, unrelated reporter gene construct. One such plasmid is the β-galactosidase expression vector pCH110 (see 2.6). This vector uses a simian virus 40 early promoter to drive the expression of the lacZ gene. In initial experiments, obtaining reliable reporter activity with this plasmid proved difficult in HeLa cells and

Figure 5.7 (A) A schematic view of the sequencing strategy used for the tetramer/dimer oligonucleotide mixture clones. The four copies of the oligonucleotides in the tetramer plasmid construct can only be sequenced properly by priming on both sides of the insert cloning point. (B) A typical dideoxy sequencing gel showing the mutations (arrowed) in this clone (position -25). Lanes 1→4 are from primer 1 and lanes 5→8 from primer 2. The bracketed area shows the region where the dimer clone in the mixture interferes with the sequencing.

A


Primer 2
Vector
ARE1
ARE2
ARE3
ARE4
Vector
Primer 1
Primer 2
Vector
ARE1
ARE2
Vector
Primer 1

B


1 2 3 4
T G C A
5 6 7 8
T G C A
Mixture
Mixture
A→C
A→C
T→G
T→G

so three other candidate expression vectors were tested as internal transfection control plasmids (see 2.6): the β-galactosidase expression vector pIRV, and the luciferase expression vectors pGL2Control, and pGL2MIN. Again, none of these plasmids provided suitable reporter activity (see Fig 5.8) to use in the transfection experiments, it was decided to construct a new expression vector to act as an internal transfection control.

It was known from the initial experiments that the Herpes simplex thymidine kinase promoter in the pBLCAT2 plasmid was able to drive the expression of the CAT gene in this cell line. The CAT gene was already being used in the experiments to determine the hormone responsiveness of the factor IX related oligonucleotides, so it was decided that a suitable construct to test would be this promoter driving expression of a different reporter gene. The reporter gene chosen was the luciferase gene, due to the ease of detection of its activity and the sensitivity of the assay. The plasmid pGL2tk was constructed by cloning the tk promoter fragment from pBLCAT2 into the vector pGL2Basic (see 2.6 and Fig 5.9). This vector was also tested in HeLa cells and the results can be seen in Fig 5.8. This plasmid was able to direct an acceptable, reliable level of reporter expression under the conditions of the transfection (see Fig 5.8), and so was chosen as the internal transfection control to use in these experiments.

5.6 Hormone Response Assays

The CAT assays for all of the mutant oligonucleotide reporter constructs were conducted as described in 2.27. They were quantitated by exposing each thin layer chromatogram to a phosphorimaging screen and then using the ImageQuant software (see 2.21) to quantitate the percentage acetylation of the chloramphenicol in the assay. An autoradiograph of a typical hormone response CAT assay can be seen in Fig 5.10. The results were expressed as the increase in expression of the reporter gene over the background values in each experiment, so lanes 3 and 4 should be compared to lanes 1 and 2 (wild type), and lanes 7 and 8 to lanes 5 and 6 (-26 G→C). The background was defined as the average of the activity in the two negative controls included with each

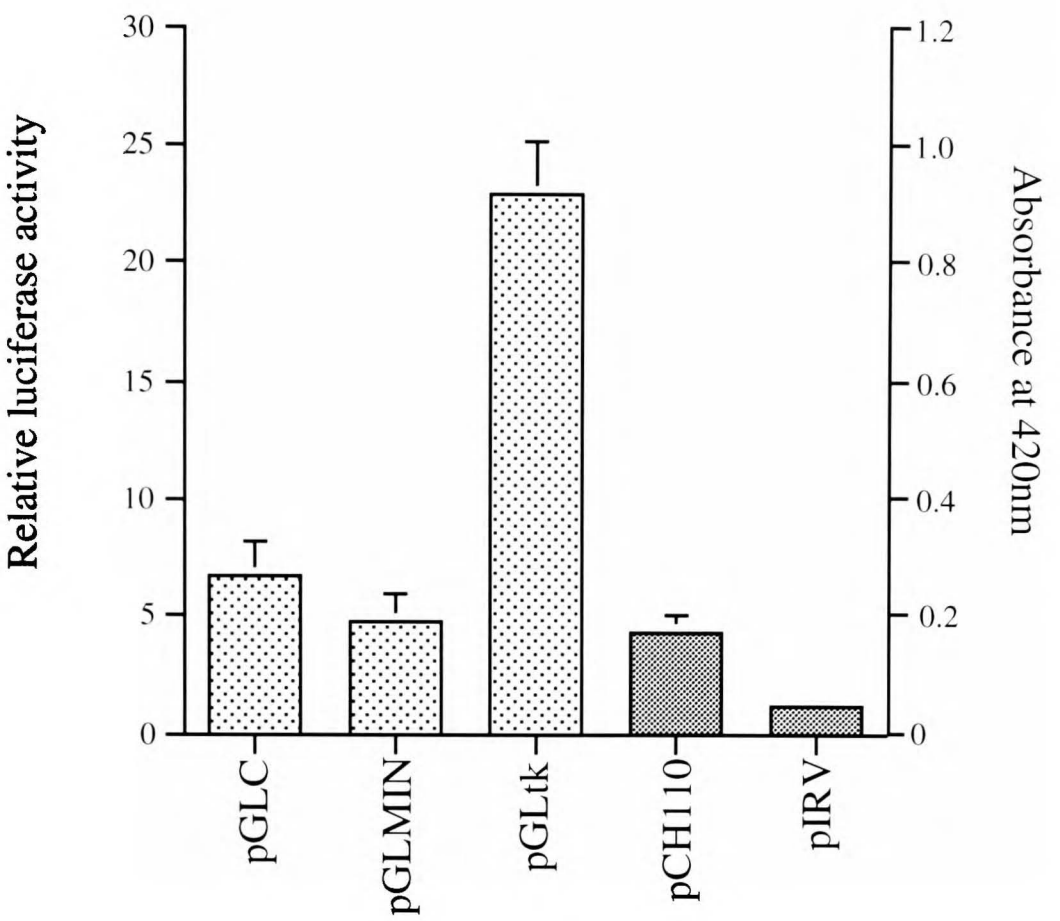


Figure 5.8 A comparison of several candidate internal control plasmids for the transient transfections in HeLa cells. The plasmids are expression vectors for the luciferase (⊡), or β-galactosidase (▩) reporter genes. The experiments were repeated a minimum of three times and the standard errors are shown where large enough to be visible. See 2.6 for details concerning the plasmids.

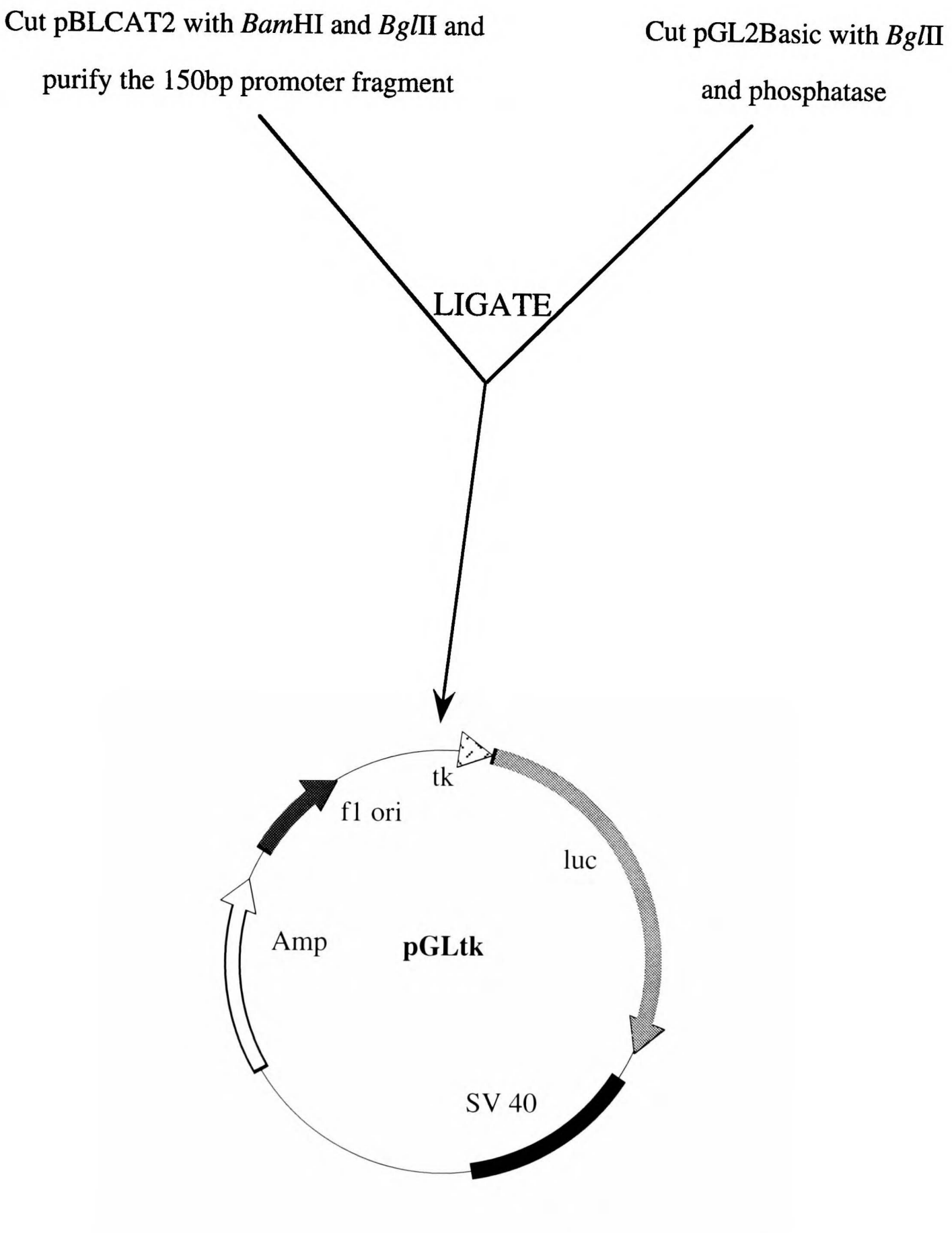


Figure 5.9 Construction of the new luciferase expression vector pGLtk for use as an internal control in the transient transfections.

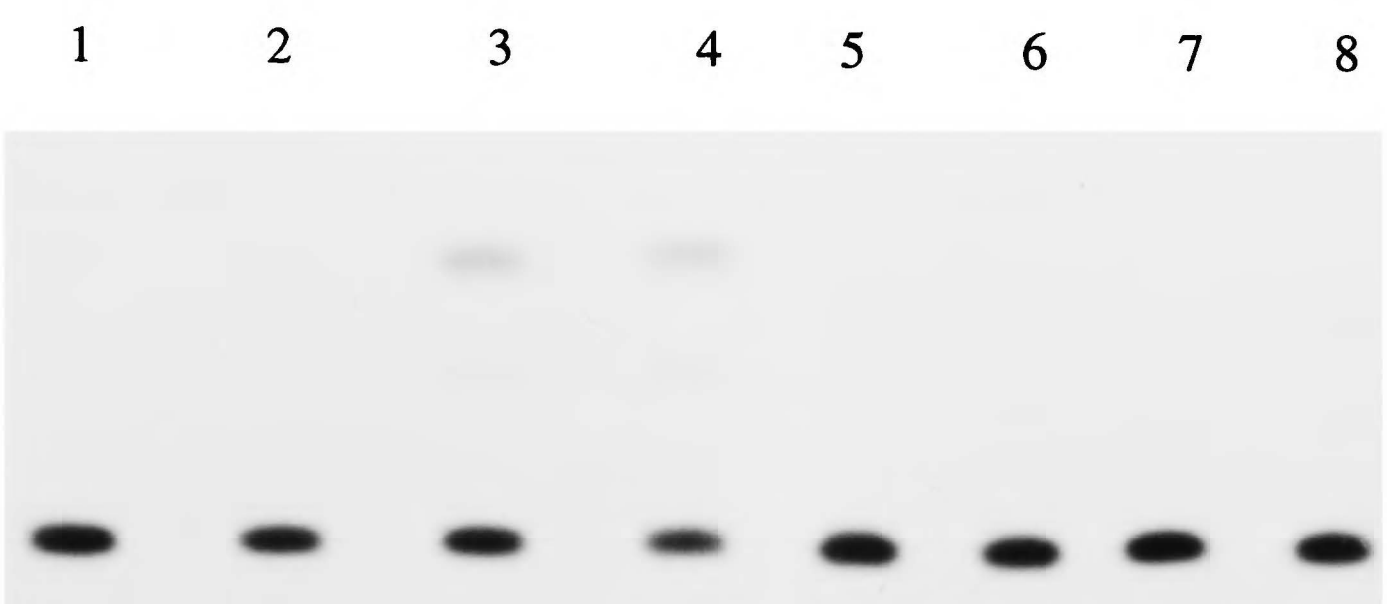


Figure 5.10. Autoradiograph of a typical hormone response CAT assay. Lane 1= wild type factor IX ARE tetramer construct -AR, + 5α-dihydrotestosterone; lane 2= wild type factor IX ARE tetramer construct +AR, -5α-dihydrotestosterone; lanes 3+4= wild type factor IX ARE +AR, +5α-dihydrotestosterone; lane 5= -26G→C tetramer construct -AR, +5α-dihydrotestosterone; lane 6= -26G→C tetramer construct +AR, -5α-dihydrotestosterone; lanes 7+8= -26G→C tetramer construct +AR, +5α-dihydrotestosterone.

construct (i.e. lanes 1+2 for the wild type, and lanes 5+6 for the -26 G→C). The first of these involved the absence of the AR expression vector from the experiment, and the second involved removing the 5α-dihydrotestosterone from the assay. The results obtained with each construct can be seen in Table 5.2. Each value is an average of at least three independent hormone response assays and the standard error of the mean is shown for each construct tested.

5.7 Results of the Hormone Response Assays

Table 5.2 shows the results of the hormone response assays with the mutant factor IX ARE oligonucleotide reporter constructs. Fig 5.11 highlights the positions at which the hormone response assay identified mutations which disrupt the ARE. Mutations at the 5 of the 6 positions defined as being highly conserved in the study of Roche et al., (1992) disrupt ARE function, these positions are -6, -4, -3, 2, 3, 4, and 6 (nucleotides -35, -33, -32, -27, -26, -25, and -23, see Fig 5.11). Mutations at positions -7, -2, and 7 (nucleotides -36, -31, and -22, see Fig 5.11) all appear to impair the ARE function, but not disrupt it completely. These results concur with the data from the consensus study of Roche et al., (1992) for twelve of the eighteen constructs tested. The mutations which did not give the predicted results were:

(i) the mutation at position -4 (nucleotide -33, see Fig 5.11) disrupted ARE function,

(ii) the mutations in the spacer region at positions -1 and 1 (nucleotides -30 and -28, see Fig 5.11) did not alter, and impaired ARE function, respectively,

(iii) the mutations at position 5 (nucleotide -24) did not create a stronger ARE (A→T), or disrupt the ARE function A→C),

(iv) the mutation at position 10 (nucleotide -19) did appear to impair the ARE function in this assay.

The positive control construct containing 4 copies of the C3 ARE showed a poorer response to hormone in these assays than the factor IX wild type ARE, see Table 5.2, the transactivation only being 50% of the factor IX wild type levels.

Table 5.2 The results of the hormone response assays with the oligonucleotide tetramer reporter constructs. ^{P}= mutation found in a patient, += mutation designed to make the ARE a better consensus. The standard error of the mean is shown. Each result is the average of at least three independent experiments.

Construct	Activation in Response to AR and Testosterone	Standard Error of the Mean
Wild type	20.9	1.2
C3	10.9	1.7
-36 A→T	4.6	1
-35 G→C	1.6	0.8
-34 C→G	17.8	1.9
-33 T→C	1.7	0.3
-32 C→A	1.3	0.1
-31 A→C	3.9	1.3
-30 G→A	27	6.3
-28 T→G^{+}	11.3	1.7
-27 T→G	1.5	0.5
-26 G→C^{P}	1.3	0.2
-26 G→A^{P}	2.8	0.4
-25 T→G	1.4	0.5
-24 A→T^{+}	14.7	2.5
-24 A→C	22.7	2.4
-23 C→T^{P}	1.2	0.3
-23 C→A	7.4	0.8
-22 T→A	9.1	2.3
-19 G→C^{P}	10.2	1

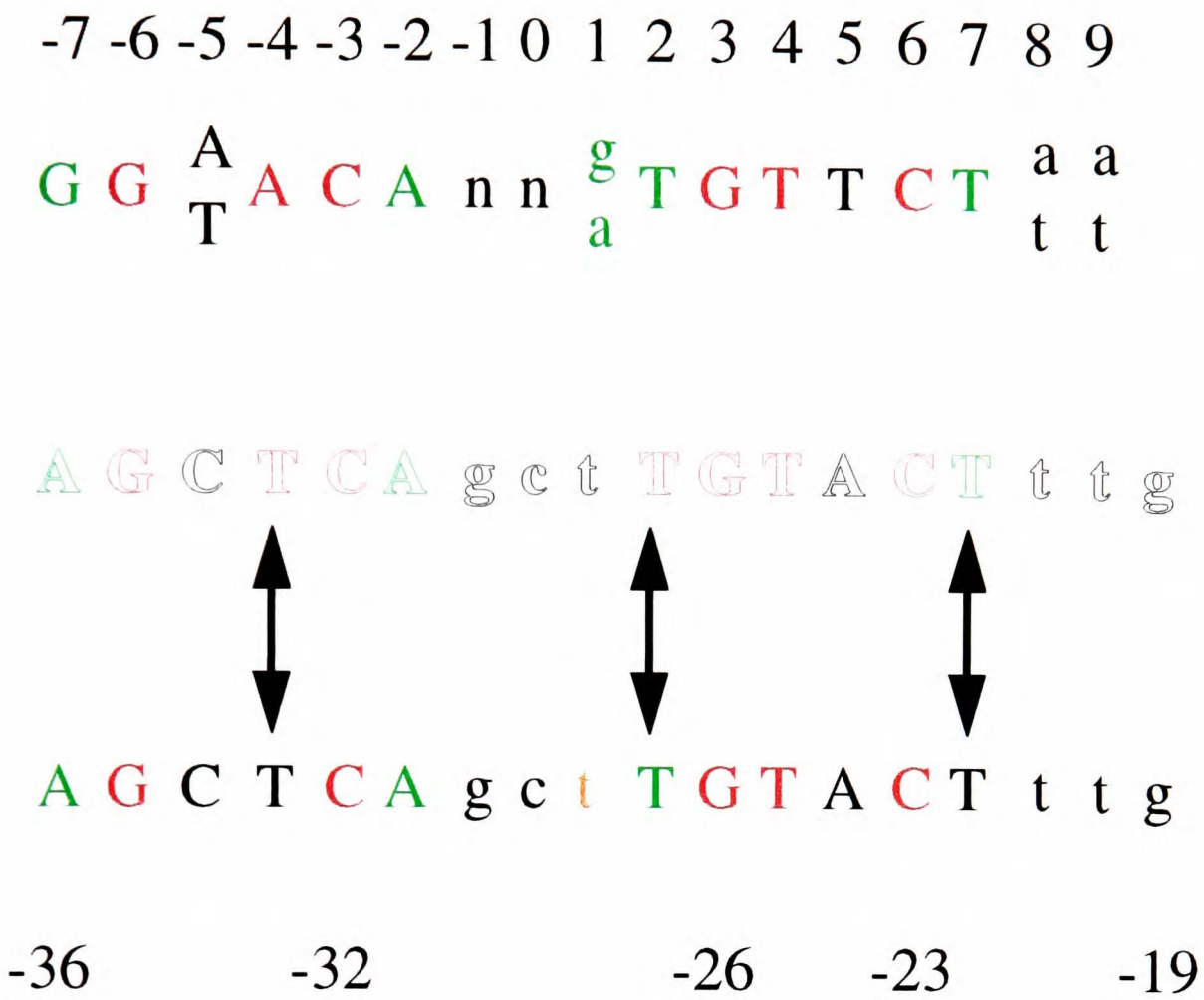


Figure 5.11 A comparison of the results obtained from the gel mobility shift assay (plain type, lower sequence), the transient transfection assays (outline type, middle sequence), and the consensus study (plain type, upper sequence). The colour key is the same as that in Fig 4.13, red indicates a mutation which severely disrupts the ARE function, and green a position where a mutation partially disrupts ARE function. Position +1 is coloured orange in the lower sequence denoting that a mutation of the thymine to a consensus guanine improved binding of the fusion protein to the ARE. The three differences between the gel shift data and the transfection data are highlighted with double headed vertical arrows.

5.8 Discussion

The results of these hormone response assays follow the predictions made from the consensus study of Roche et al., (1992) in five of the six positions defined as "highly constrained", and all four of the positions defined as "moderately constrained" (see Fig 5.11). The results from the mutations in the spacer region at positions -1 and 1 (nucleotides -30 and -28) do suggest that this region does not play a significant role in determining the response to AR and hormone. The result for mutation at -1 (-30) is not significantly different from the wild type ARE if the standard errors are taken into account and the result for the mutation at 1 (-28) does not follow the prediction made for it from the consensus study (see Table 4.2). The sequences found in natural AREs do not strictly conform to the consensus derived by Roche et al., (1992) (see Fig 4.1) at position 1, also implying that this position is not critical for the function of the ARE element.

The mutation at position -4 (nucleotide -33) is in a position which is defined as "highly constrained", but the factor IX ARE does not adhere to the consensus in this position, so the mutation was only expected to have a minor inhibitory effect (see Table 4.2). This suggests that this position, though a non-consensus nucleotide, is important for the binding of the AR homodimer to the ARE between -36 and -22 in this promoter. The possible reasons for this result will be discussed more extensively together with the *in vitro* data for this position in Chapter 8.

The mutations at position 5 (nucleotide -24) were designed to have differing effects (see Table 4.2). The A→T mutation at this position was intended to create a stronger ARE element than the factor IX ARE, and the A→C mutation was designed to marginally disrupt the ARE (see Table 4.2). Table 5.2 shows that the A→T mutation did not increase the transactivation in this assay, and the A→C mutation does not disrupt the transactivation, in fact the opposite is true for both mutations. These results suggest that position 5 plays no significant part in the response of this element to AR and testosterone, which concurs with the data reported by Roche et al., (1992).

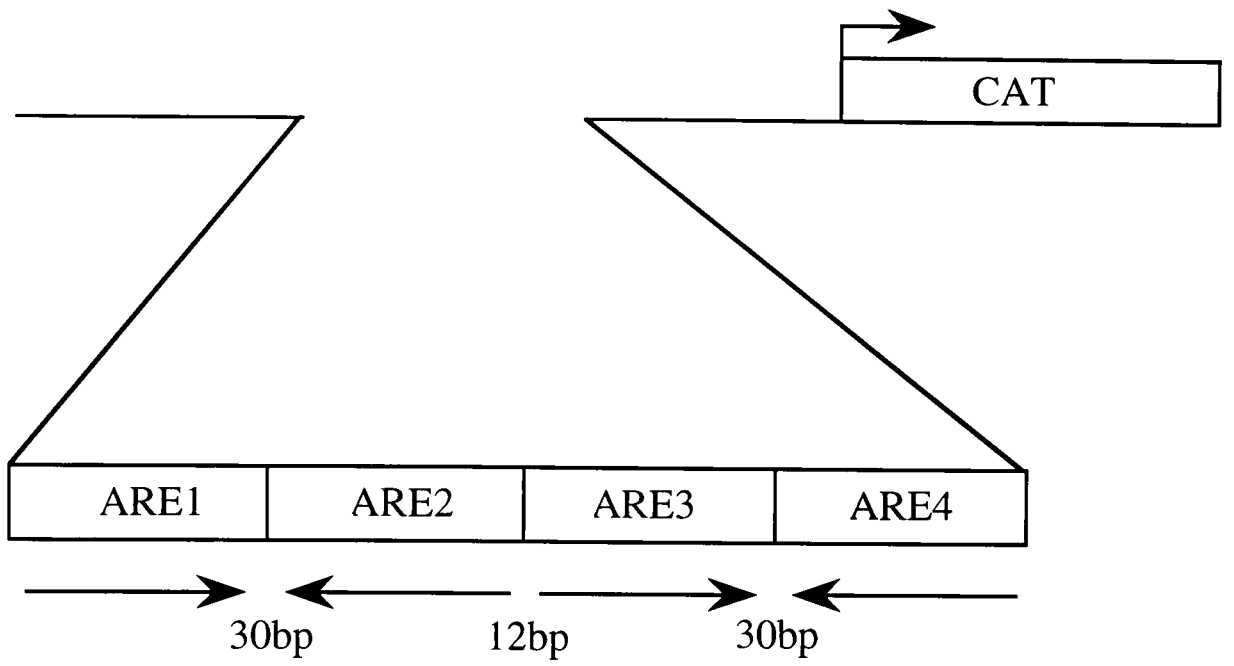
Roche et al., (1992) reported that the consensus ARE has a preference for A/T nucleotides in the immediate 3' flanking sequence. The *in vivo* data reported here suggests that mutations at these positions might effect the activity of the ARE (see Table 5.2). However, the existence of three patients with mutations in this region (at nucleotides -21→-19), of whom two patients' factor IX levels have risen through puberty (the other has not yet reached puberty, see Chapter 6) does suggest that these positions are not vital for the functioning of the factor IX ARE.

Table 5.2 shows that the transactivation of the factor IX wild type sequence is greater than that of the positive control oligonucleotide (C3 ARE, Claessens et al., 1989). If the gel mobility shift data is analysed then the C50 value of the C3 oligonucleotide is less than that of the factor IX wild type sequence (see Table 4.3) suggesting that the former sequence binds to the protein A-AR DNA binding domain fusion protein with higher affinity than the latter. It is possible that the differences in arrangements of the two tetramer reporter construct could be the causative factor here (see Fig 5.12). The spacing between the C3 ARE elements in the tetramer construct was 16 base pairs and 14 base pairs whereas the spacing in the factor IX related oligonucleotide tetramer constructs was 12 base pairs and 30 base pairs. The DNA double helix has about 10 nucleotides per turn and so these distances between ARE elements would place all four factor IX elements on roughly the same face of the helix, possibly allowing each AR homodimer to cooperate in the transactivation of the tk promoter. In contrast, the C3 ARE elements are arranged such that elements 1 and 3 are on the same face of the helix and so are elements 2 and 4, but these two pairs of elements are on opposite sides of the helix. These arrangements may have reduced the transactivation of the C3 ARE tetramer relative to the factor IX related constructs in a sequence independent manner.

5.9 Summary

To summarise, this chapter has described the cloning of factor IX ARE oligonucleotides into the reporter gene vector pBLCAT2 to test the effects of mutations

Factor IX related ARE oligonucleotide tetramers



C3 ARE oligonucleotide tetramer

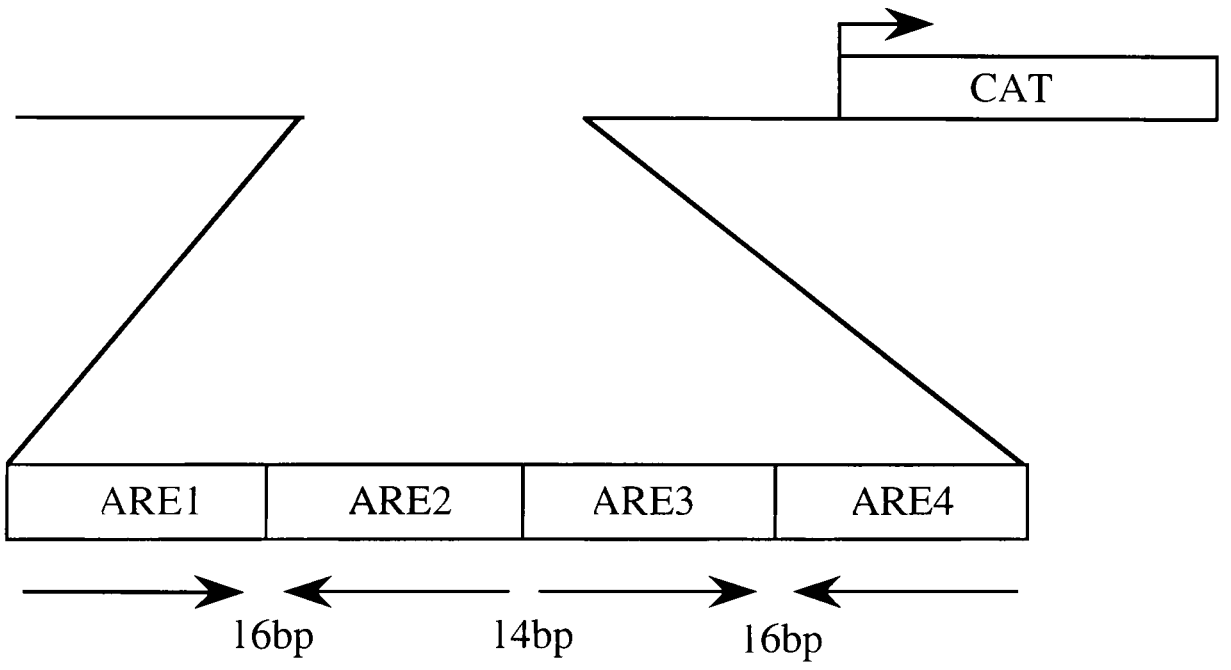


Figure 5.12 Representation of the factor IX related oligonucleotide tetramers and the C3 ARE oligonucleotide tetramer constructs. The orientation (5'→3') of the ARE is denoted by an arrow and the spacing between adjacent AREs is marked.

at each position in the ARE. The oligonucleotides were not stable as tetramers in these constructs and the plasmid samples used for the transfections were tetramer/dimer oligonucleotide mixtures. The transactivation assays identified a number of positions which were critical for the androgen responsiveness of this sequence and other positions which impaired the function of the ARE without abolishing activity. The positions identified were those which had previously been defined as important for the binding of AR to DNA by Roche et al., (1992). Some differences between this mutagenesis study and the consensus study did appear though, the presence of a G at position -7 (nucleotide -36) did not seem to be important for the androgen responsiveness. The spacer nucleotides did not seem to effect the response to AR and testosterone by this element. Position 4 (nucleotide -33) seemed critical for the AR/testosterone response even though it was not a consensus nucleotide. Hence, it can be seen that though the consensus study has proved to be highly accurate, it is still necessary to perform this type of analysis of naturally occurring ARE sequences to be certain which positions are critical for their action.

Chapter 6: A study of three patients with novel point mutations in the promoter of the factor IX gene.

6.1 Introduction

An interesting subset of patients suffering from haemophilia B recover from their disease through puberty. This phenotype is known as the Leyden phenotype (see Fig 1.20). Patients displaying this phenotype typically have factor IX clotting activities of 1-3% in childhood and these levels rise throughout puberty to 30-60% in adulthood. All these patients have mutations in the promoter region of the factor IX gene between nucleotides -21 and +13 (Giannelli et al., 1996). The study of these patients has been invaluable in the functional characterisation of the factor IX promoter and in the identification of transcription factor binding sites in the promoter region (see Fig 1.4, and Table 1.2).

A -26 G→C mutation in the promoter of haemophilia B Brandenburg (Crossley et al., 1992) was found to disrupt overlapping binding sites for AR (-36→-22) and HNF4 (-27→-15). This patient did not show the typical Leyden recovery of factor IX plasma levels through puberty and died aged 21 years (see Fig 1.20). All of the mutations which result in the Leyden phenotype are found between -21→+13 in the promoter region, and so do not disrupt this ARE (-36→-22). Therefore, Crossley et al., (1992) suggested that the action of AR and its ligand testosterone, acting through this ARE could upregulate the factor IX promoter throughout the course of puberty. The -26 G→C mutation in the ARE disrupted the binding of AR and so prevented any such rise in factor IX levels.

Many alternative mechanisms have been suggested (see sections 1.10.2→1.10.6). These reports have cast doubt upon the importance, and existence, of this ARE in the factor IX promoter. The study of new patients, and the identification of the molecular effects of any promoter mutations in and around this ARE will be important further evidence, required to further characterise the role of the ARE in the Leyden recovery mechanism.

In this chapter the study of three patients with unique molecular defects in the factor IX promoter region is reported. The effects of these mutations on the transcription factor binding and transactivation of both the minimal gene promoter, and of the isolated -40→-9 region, are characterised. As two of the patients are still in childhood, the implications of these results for their adult phenotype are discussed.

6.2 Haemophilia B Southampton, Patient UK 232.

6.2.1 Case History

Patient UK 232 (Giannelli et al., 1996) was born in 1949 and diagnosed as suffering from severe haemophilia B in childhood. Details of the plasma factor IX activity before the age of 24 years are unfortunately not available (Giannelli, personal communication), but at 24 years of age his factor IX level was undetectable. The two most recent checks on this level were in 1993 and 1994, aged 44 and 45 years of age, respectively. The levels were 3% (0.03u) and 2% (0.02u), respectively in a one stage clotting assay (Austen and Rhymes, 1975). From the age of 30 years to the present the patient has had approximately 130 bleeding episodes in various joints. These episodes always required therapeutic intervention.

The patient was referred to Professor Brownlee after a novel -26 G→A mutation was identified in the promoter region of the factor IX gene (Giannelli, personal communication). No other mutations were detected.

6.2.2 Promoter Transactivation

Nucleotide -26 in the factor IX promoter region lies in a region of overlap in the binding sites for the two transcription factors AR and HNF4 (see Fig 1.4), and previous work on the -26 G→C mutation (Crossley et al., 1992) predicted that this -26 G→A mutation would be expected to disrupt both of these binding sites. In order to test the functional effects of this mutation, the minimal promoter region, (-221→+43), was PCR amplified from samples of wild type and UK232 total genomic DNA (see 2.10). These PCR fragments were cloned upstream of the luciferase reporter gene in

the vector pGL2Basic (see 2.6). The constructs were sequenced to ensure no errors had been incorporated during the PCR and then transiently transfected into HeLa cells and the liver-derived cell lines PLC/PRF5 and HepG2, with or without cotransfection of the HNF4 expression vector pCB6HNF4 (see 2.6). The results were expressed as the fold increase in activity of the reporter construct in the cotransfection with HNF4 relative to the activity without the HNF4 expression vector present.

The optimal amount of reporter construct to be used in the transactivations was determined by titrating the wild type minimal promoter-luciferase reporter construct in the presence of 1μg of HNF4 expression vector, and 4μg pBluescript. Fig 6.1 shows the results obtained with the HeLa cell line. Similar results were obtained with the other cell lines (results not shown). The optimal amount of reporter plasmid was 15μg per transfection (the total amount of DNA being 20μg).

The transactivation of the wild type factor IX promoter region was more than 4 fold greater than the mutated promoter in HeLa cells (Fig 6.2). The same experiment in PLC/PRF5 cells showed more than a 3 fold decrease of transactivation when the -26 G→A mutation was present (Fig 6.2), and the mutation reduced the transactivation of the construct to less than 40% of the wild type value in HepG2 cells (Fig 6.2). The significance of the results of the HNF4 transactivations in each of the cell lines was tested using the student's t test. Table 6.1 shows the results of those tests. In every case the probability that the difference in the results is not significant and simply due to chance is <1%, and in 2 of the three tests performed it was <0.1%. These results strongly suggest that the differences observed in the results of the HNF4 transactivations are highly significant and hence, that the -26 G→A mutation disrupts this transactivation.

Table 6.1 Statistical analysis of the data from the transactivation of the wild type and -26 G→A mutated promoter constructs using the student's t test (see 2.8.3 for equations used)

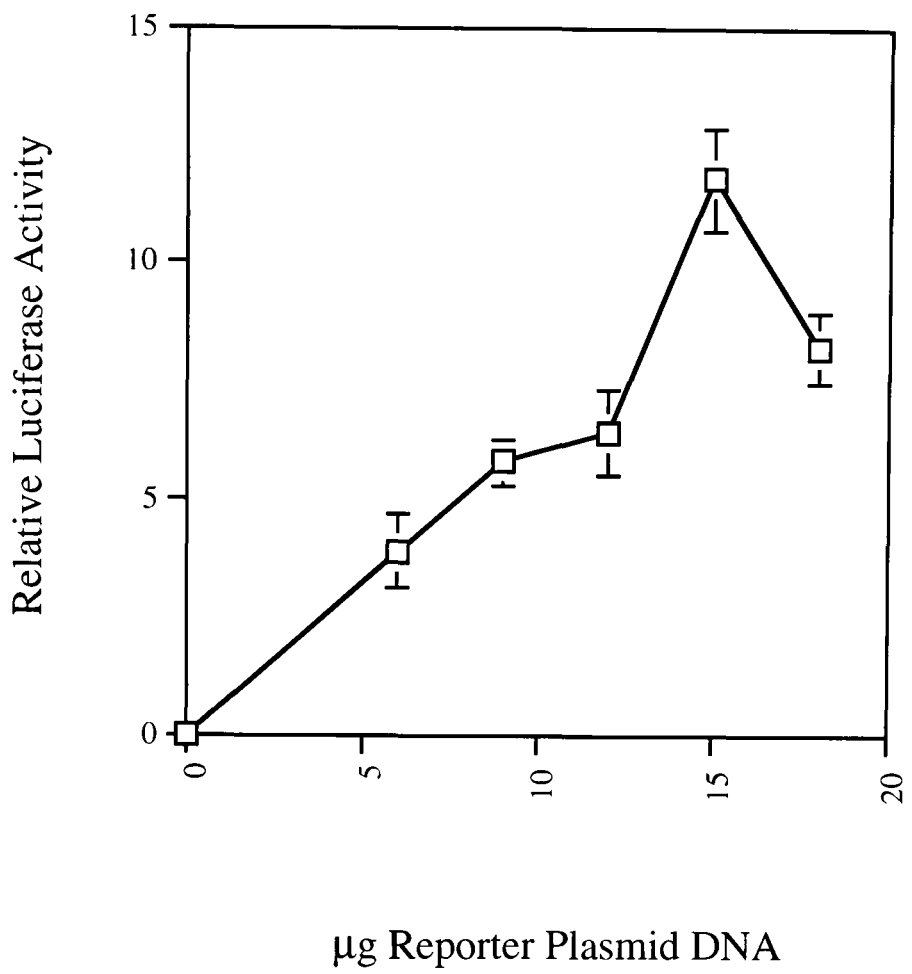


Figure 6.1 Optimisation of the amount of reporter plasmid DNA for use in the transactivation assays with HNF4 in HeLa cells. The factor IX wild type minimal promoter-luciferase construct was titrated in the presence of 1µg of HNF4 expression vector, and 4µg pBluescript. The transfections were repeated a minimum of 3 times and the standard errors of the results are shown. The relative luciferase activities were measured using a Wallac LKB 1250 luminometer.

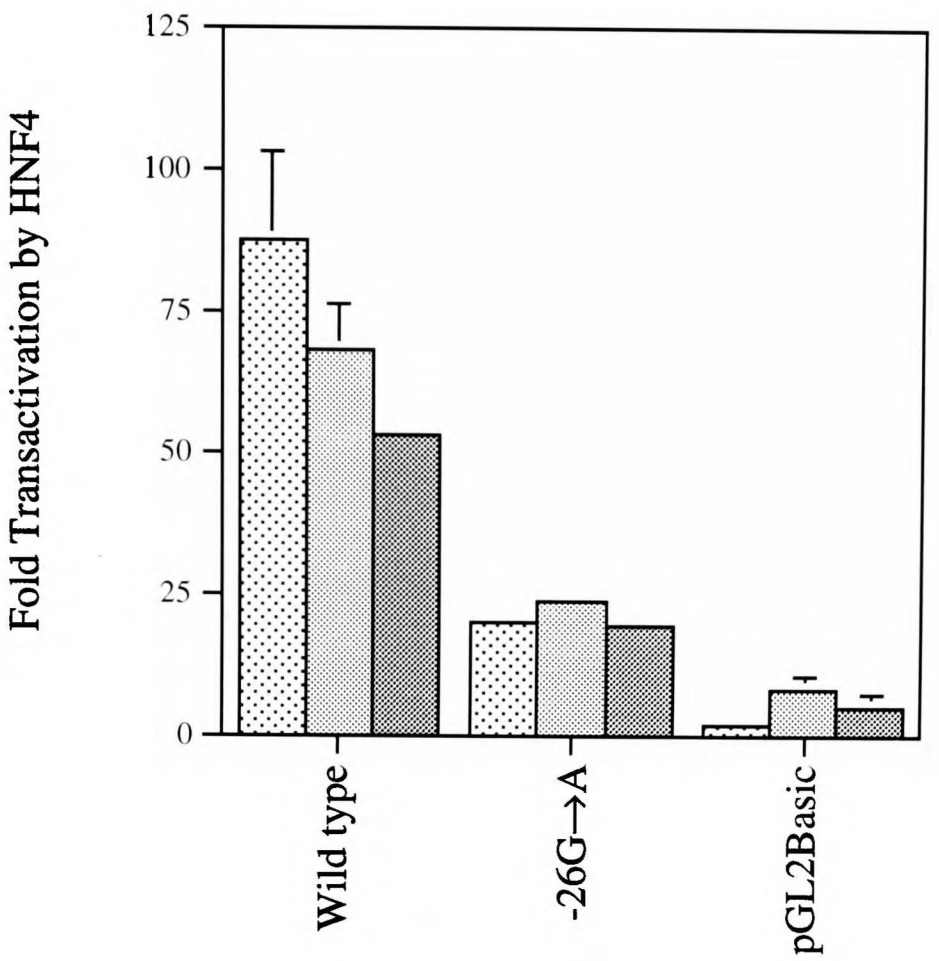


Figure 6.2. Fold transactivation of the factor IX minimal promoter region constructs by 1μg of an HNF4 expression vector. Each construct was transiently transfected into HeLa (▨), PLC/PRF/5 (□), and HepG2 (▩) cells. The activity of the construct with cotransfection of 1μg pCB6HNF4 (see 2.6) was expressed relative to the activity of the construct without this cotransfection.

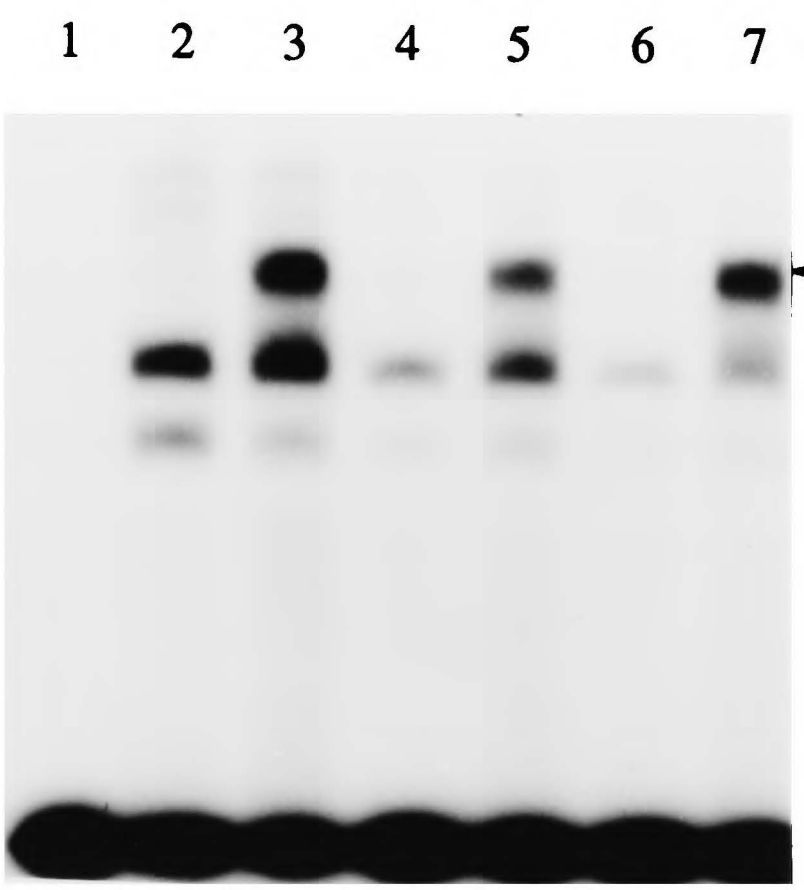
Cell Line	Construct	t	p difference is not significant
HeLa	-26 G→A	4.19	<1%
PLC/PRF5	-26 G→A	5.27	<0.1%
HepG2	-26 G→A	15.83	<0.1%

6.2.3 Transcription Factor Binding

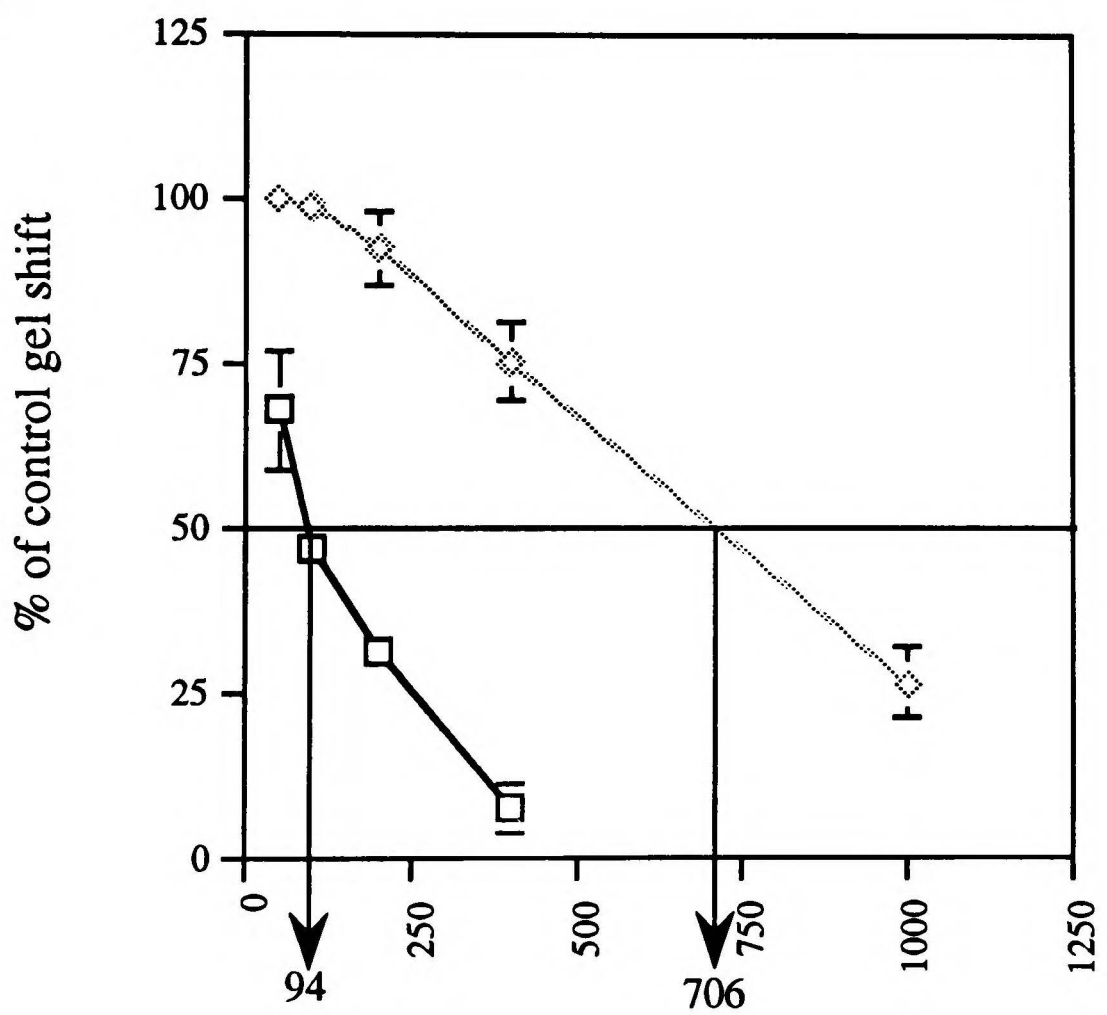
On the basis of previous results (Crossley et al., 1992), the -26 G→A mutation might be expected to disrupt the binding of HNF4 to a site spanning -27→-15 in the promoter region. When incubated with the COS1 nuclear extract containing HNF4 (see 2.6), the wild type oligonucleotide probe, (-40→-9), produced one specific complex and at least 4 distinct non-specific complexes of different mobilities (Fig 6.3A). All but one of the complexes can be seen in the untransfected COS1 nuclear extract gel mobility shift (lane 2, Fig 6.3A), and so the complex specific to HNF4 is marked. Fig 6.3A shows that the complex can be competed by wild type factor IX competitor oligonucleotide, or by the HNF4 site from the promoter of the α1-antitrypsin gene (Hardon et al., 1988), as expected from previous studies (Crossley et al., 1992). The -26 G→A mutated oligonucleotide competes very inefficiently, indicating that this mutation disrupts HNF4 binding. No competition is observed with the unrelated, negative control oligonucleotide corresponding to the albumin D site, (Maire et al., 1989) acting as a negative control, (sequences of oligonucleotides are listed in Appendix 1). In order to quantitate these results more accurately, titrations of the wild type and -26 G→A mutated oligonucleotides were then used in competition gel shift experiments. From these experiments, the C50 values for each of these oligonucleotides were determined. The C50 value is defined as the molar excess of competitor oligonucleotide which is required to compete 50% of the positive control gel mobility shift (see 2.20 and Chapter 4 for further explanation), and has been extensively used in the study of the effects of mutations in the factor IX promoter

Figure 6.3 (A) Competition gel shift of wild type factor IX probe with the COS1 expressed HNF4. Lane order: 1=probe only; 2= COS1 cell extract minus HNF4; 3= COS1 cell extract plus HNF4; 4= factor IX wild type competitor; 5= -26 G→A mutant competitor; 6= α1-antitrypsin HNF4 site competitor; 7= albumin D site competitor. Lanes 4→7 contain 500 fold molar excess of competitior oligonucleotides. The complex specific to HNF4 is arrowed. (B) Graph to obtain the C50 vlaues for the factor IX wild type (□) and the -26G→A mutant (◇) oligonucleotides.

A


1
2
3
4
5
6
7
HNF4
complex

B


125
100
75
50
25
0
% of control gel shift
0
250
500
750
1000
1250
94
706
Fold molar excess of competitor
oligonucleotide

region which reduce factor IX expression (Reijnen et al., 1993, Kurachi et al., 1994). These values give an indication of the protein-oligonucleotide affinity in the gel shift. The -26 G→A mutated oligonucleotide had a C50 value (C50=706) more than 7 times that of the wild type sequence (C50=94). This confirms, semi-quantitatively, that the -26 G→A mutation significantly reduces the affinity of this site for HNF4.

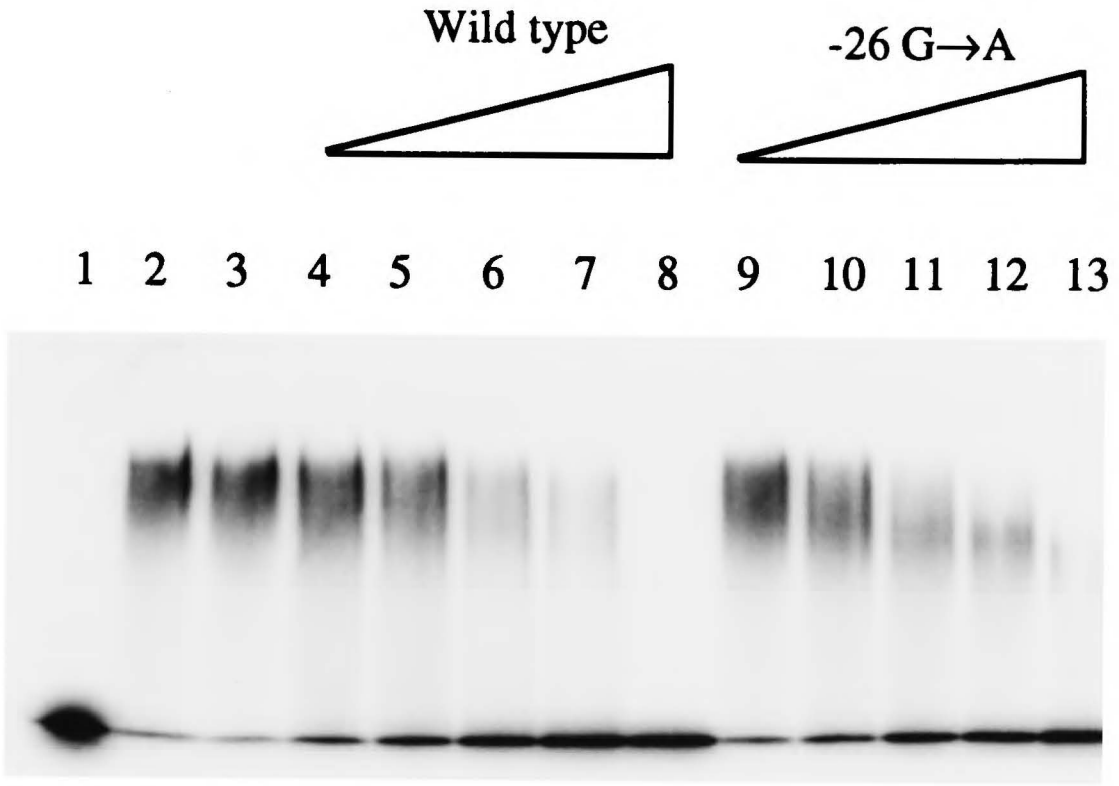
It might also be expected that this -26 G→A mutation would disrupt the binding of androgen receptor to an ARE (-36→-22) in the promoter (Crossley et al., 1992). To test this, competition gel mobility shift assays were performed with a recombinant source of protein A-AR DNA binding domain fusion protein (De Vos et al., 1991, see 2.12, and Chapter 3). As a positive control oligonucleotide, the ARE from the rat prostatic binding protein C3 (PBPC3) gene was used in the competition experiments (see Appendix 1 for sequence). The competition gel mobility shift and the C50 graph for these oligonucleotides can be seen in Fig 6.4. The gel mobility shift produces a diffuse specific band, suggesting that multiple protein DNA complexes may be present (see Fig 4.4). This diffuse band is competed more effectively by the wild type competitor oligonucleotide (lanes 4→8) than with the -26 G→A mutated oligonucleotide (lanes 9→13). The C50 values read off from Fig 6.4B demonstrate that the factor IX wild type sequence (C50=21) has a higher affinity for the protein A-AR DNA binding domain fusion protein than the -26 G→A mutated sequence (C50=53). The C50 values for the C3 ARE (C50=15), and the factor IX wild type ARE (C50=21), were more similar than previous literature would suggest (Kurachi et al., 1994).

6.2.4 Transient Transfections to Determine Androgen Receptor/Testosterone Response

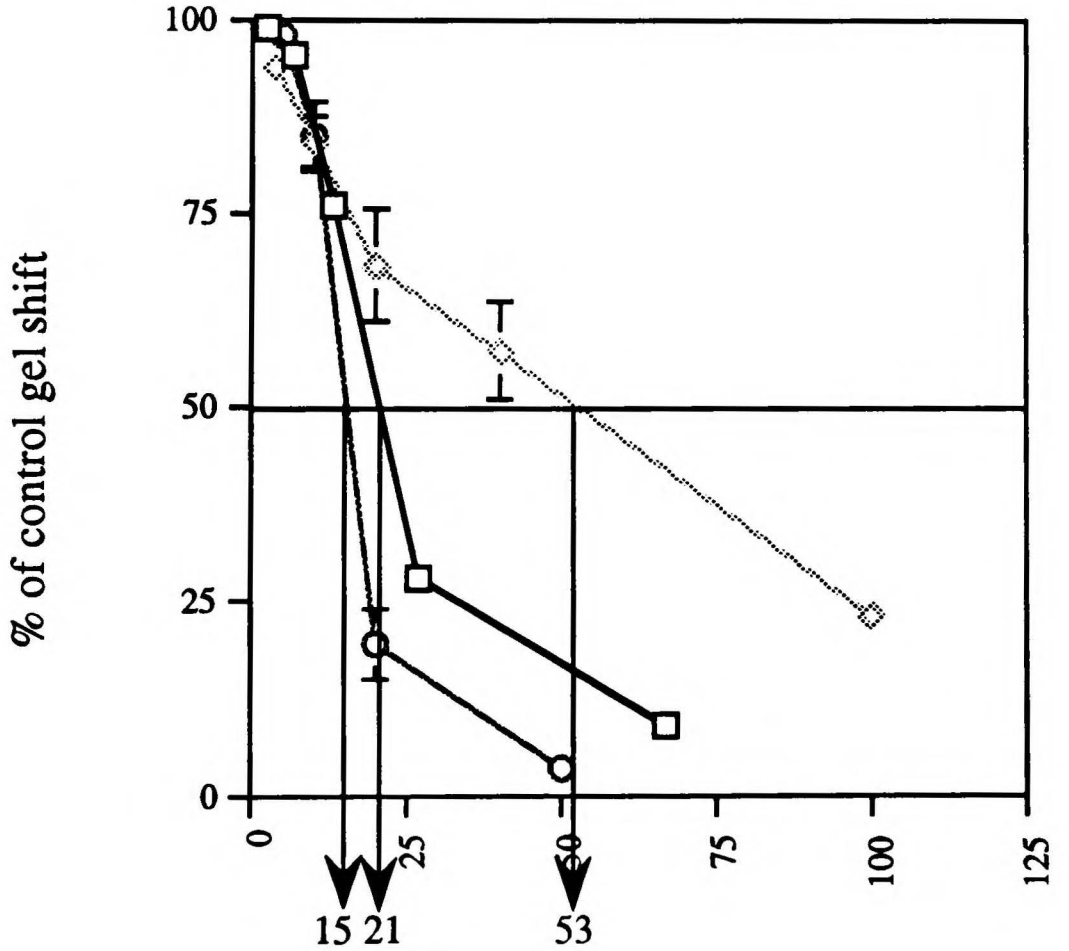
It was already known (Crossley et al., 1992, and Fig 5.6) that tetramers of the factor IX ARE sequence (-42→-17) could confer androgen responsiveness upon a HSV tk promoter in the plasmid pBLCAT2 (Luckow and Schutz, 1987). Tetramers of the wild type and -26 G→A mutated oligonucleotides (-40→-9) were cloned upstream of a tk promoter-CAT gene reporter construct (see 5.6 and 2.6) and transiently

Figure 6.4 (A) Competition gel shift of the factor IX wild type probe and protein A-AR fusion protein. Lanes 4→8, wild type competitor; lanes 9→13, -26 G→A competitor oligonucleotides. The molar excesses of the competitor oligonucleotide in lanes 4→8 were, 8 fold, 16 fold, 28 fold, 64 fold, respectively, and the levels used in lanes 9→13 were 5 fold, 10 fold, 20 fold, 40 fold, 100 fold, respectively. (B) Graph to obtain C50 values for the factor IX wild type (□), -26G→A mutant (◇), and PBPC3 (○) oligonucleotides.

A


Wild type
-26 G→A
1 2 3 4 5 6 7 8 9 10 11 12 13

B


% of control gel shift
100
75
50
25
0
0 25 50 75 100 125
15 21 53
Fold molar excess of competitor oligonucleotide

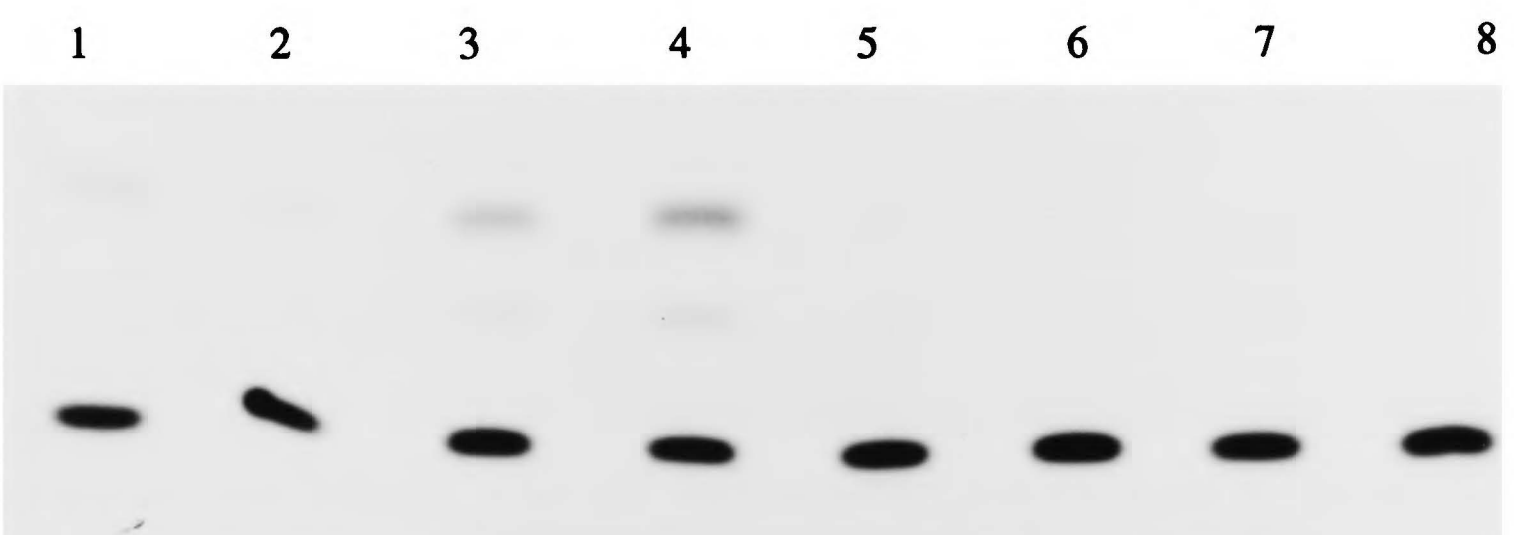
transfected into HeLa cells. The ability of HNF4 or AR and testosterone to transactivate these constructs was tested, Fig 6.5. A comparison of the reporter activities in lanes 1 and 2 with lanes 3 and 4 (Fig 6.5A), and the correction of the values for the transfection efficiency, shows that coexpression of HNF4 activates the wild type tetramer 16 fold, and the -26 G→A mutant 0.6 fold (compare lanes 5 and 6 with lanes 7 and 8, Fig 6.5A), (see Table 6.2 for summary). When AR is coexpressed with both of these constructs, in the presence of 5α-dihydrotestosterone, it is able to activate the wild type ARE 21 fold (compare lanes 1 or 2 with lanes 3 and 4, Fig 6.5B). A comparison of lane 5 or 6 with lanes 7 and 8 illustrates that the -26 G→A mutation significantly impairs the ability of testosterone and AR to activate this sequence, the value decreasing to 2.8 (see Table 6.2 for summary). Table 6.2 also illustrates that the PBPC3 ARE tetramer transactivates the reporter gene 10.9 fold in the presence of AR and testosterone, just over 50% of the factor IX wild type value (20.9). This strongly suggests that the factor IX ARE is functional, as it compares favourably with this well characterised prostatic binding protein ARE in this assay.

Table 6.2 The transactivation of wild type, -26 G→A mutated oligonucleotide, and PBPC3 ARE tetramers by HNF4 and AR/5α-dihydrotestosterone. ND=not determined.

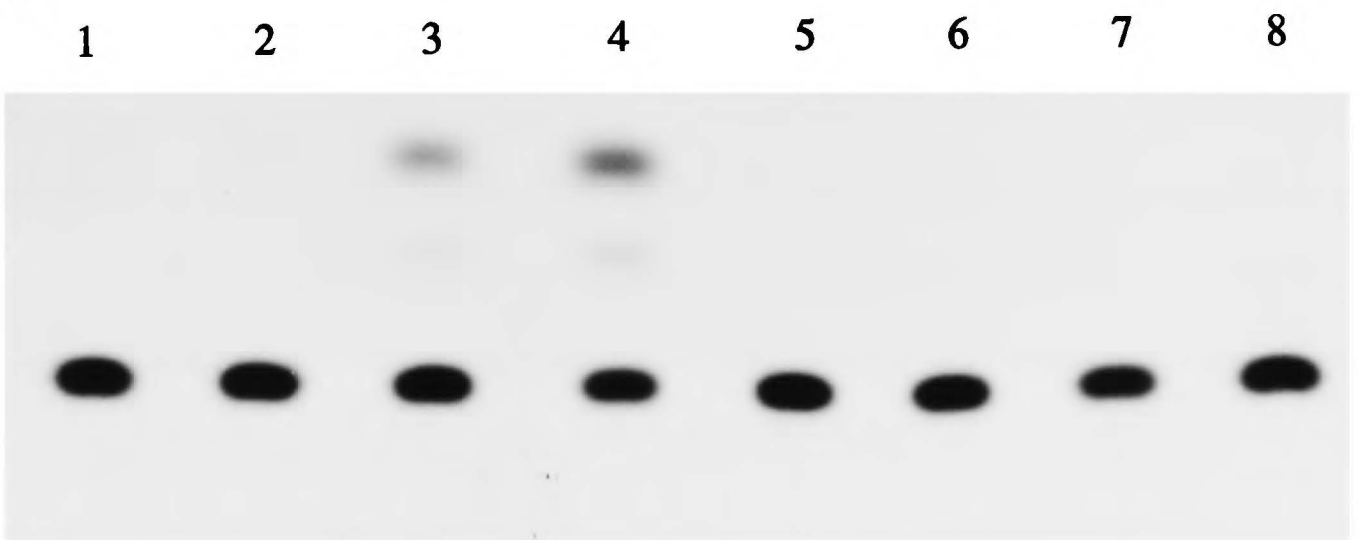
Construct	HNF4 Transactivation	AR Transactivation
Wild Type	15.8	20.9
-26 G→A	0.6	2.8
PBPC3	ND	10.9

Figure 6.5 CAT assays of the transactivation of the tetrameric wild type factor IX ARE and -26 G→A mutated ARE linked to HSV tk-CAT constructs by (A) HNF4, and (B) AR and 5α-dihydrotestosterone. The lanes are: (A) 1+2= wild type, minus HNF4; 3+4= wild type, plus HNF4; 5+6= -26 G→A, minus HNF4; 7+8= -26 G→A, plus HNF4. (B) 1=wild type, minus AR, plus 5α-dihydrotestosterone; 2=wild type, minus 5α-dihydrotestosterone, plus AR; 3+4= wild type, plus AR and 5α-dihydrotestosterone, 5= -26 G→A, minus AR, plus 5α-dihydrotestosterone; 6= -26 G→A, minus 5α-dihydrotestosterone, plus AR; 7+8= -26 G→A plus AR and 5α-dihydrotestsoterone.

A



B



6.3 HB 388 and HB 396

6.3.1 Introduction

A recent report (Ketterling et al., 1995) identified two novel factor IX promoter mutations in a sample of Mexican haemophilia B patients. Patient HB 396 was a 12 year old male with a factor IX clotting level of 3% (0.03u). A G→C mutation at nucleotide -19 was identified in the promoter region of his factor IX gene. Patient HB 388 was a 4 year old male who had a factor IX clotting level of 2% (0.02u). A C→T mutation was identified at nucleotide -23 in his factor IX gene promoter (Ketterling et al., 1995).

The mutation identified at -23 is in the 6 base pair overlap region of the binding sites for HNF4 and AR whereas the -19 mutation is not. Ketterling et al., (1995), speculated that the -23 C→T mutation might be expected to disrupt both of these binding sites but the -19 G→C mutation would probably only disrupt the HNF4 site, leaving the ARE intact. Following this assumption, Ketterling et al., (1995) predicted that one of these patients would recover through puberty (HB 396, -19 G→C) but the other would not (HB 388, -23 C→T). This prediction was based on previous reports (Crossley et al., 1992) which identified the ARE in the factor IX promoter as being able to upregulate the gene through puberty if the HNF4 site was disrupted by a mutation. A patient whose ARE was disrupted by a -26 G→C mutation, haemophilia B Brandenburg, (Crossley et al., 1992 and 1.10.2) did not show any improvement in his factor IX levels through puberty. The -23 C→T and -19 G→C mutations were tested in the following sections to confirm, or refute the predictions of Ketterling et al., (1995).

6.3.2 Promoter Transactivation

The minimal factor IX promoter regions (-221→+43) were PCR amplified from total patient genomic DNA, cloned into the reporter gene vector pGL2Basic (see 2.6), and used in the same HNF4 transactivation assay system used in 6.2.2. The results of the assays are shown in Fig 6.6. The mutated promoter regions both show lower

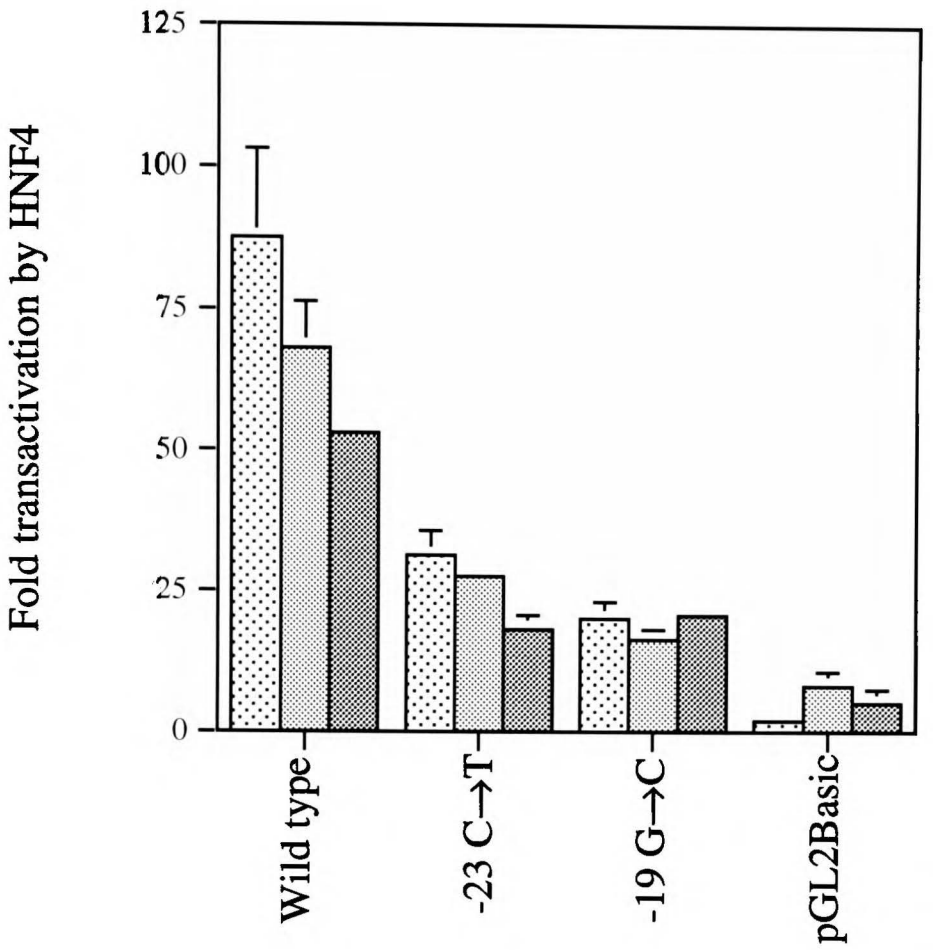


Figure 6.6. Fold transactivation of the wild type, -23 C→T, and -19 G→C mutated minimal promoter regions (-220→+43) by an HNF4 expression vector in HeLa (▩), PLC/PRF/5 (□), and HepG2 (■) cells. The promoterless construct pGL2Basic is used as a negative control. All of the experiments were repeated a minimum of three times and the standard errors of the results are shown where they are large enough to be visible.

transactivation than the wild type promoter. The 19 G→C mutated promoter was transactivated to 40%, 40%, and 34% of the wild type level in HeLa, PLC/PRF/5, and HepG2 cells, respectively (see Table 6.3 for a summary). The -23 C→T mutated promoter was transactivated to 23%, 24%, and 39% of the level of the wild type promoter region in HeLa, PLC/PRF/5, and HepG2 cells, respectively (see Table 6.3 for a summary). These results suggested that these two mutations in the factor IX promoter impaired the HNF4-mediated transactivation of the minimal promoter region.

Table 6.3 Summary of the HNF4 transactivation of the wild type, -23 C→T, and -19 G→C mutated minimal promoter regions in HeLa, PLC/PRF/5, and HepG2 cells.

Cell line	Construct	Transactivation relative to wild type
HeLa	-19 G→C	40%
	-23 C→T	23%
PLC/PRF/5	-19 G→C	40%
	-23 C→T	24%
HepG2	-19 G→C	34%
	-23 C→T	39%

The statistical analysis of these results using the student's t test (see 2.8.3) showed both of the mutant transactivation levels to be significantly different to the wild type (Table 6.4) in all three cell lines tested. The probabilities that these differences were due to chance and not significant were at least <5%, and in four of the six results were <1%. These results further strengthen the evidence which identifies these two mutations as disrupting the transactivation of the factor IX promoter region by HNF4.

Table 6.4 Statistical analysis of the transactivation data of the -19 and -23 mutant minimal promoter constructs using the Student's t test

Cell line	Construct	t	p difference is not significant
HeLa	-19	3.29	<5%
	-23	4.12	<1%
PLC/PRF5	-19	4.79	<1%
	-23	6.03	<1%
HepG2	-19	2.79	<5%
	-23	5.06	<1%

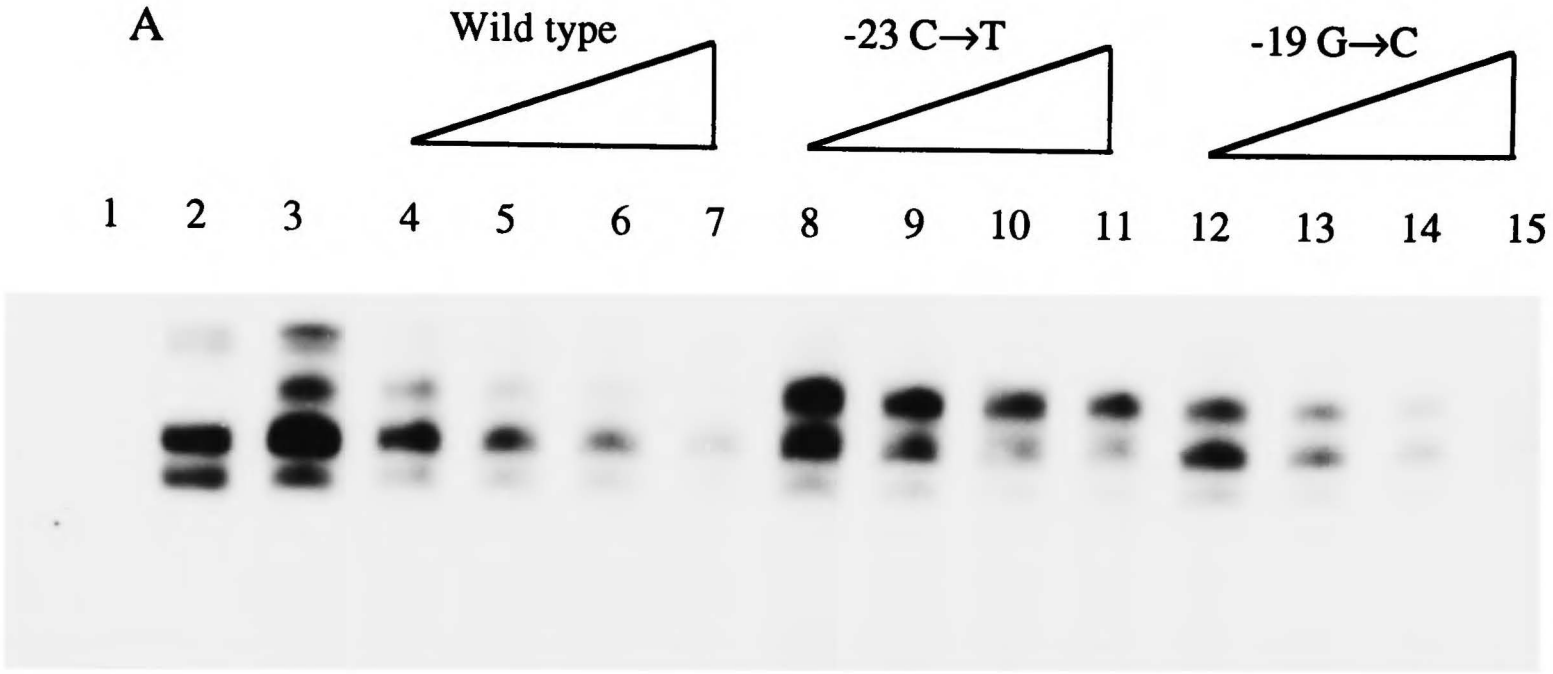
6.3.3 Transcription Factor Binding

The mutant 19 and -23 sequences were used in the same competition gel mobility shift assays as used in 6.2.3. A typical competition experiment is shown in Fig 6.7A. The C50 values (see 2.20) for these 2 mutations were calculated from the average of at least 3 competition experiments and the graphs are shown in Fig 6.7B. The C50 values read off the graphs for both of these mutations show an increase on the wild type value. The -19 has a C50 more than double (C50=255) that of the wild type (C50=94) and the -23 has a value which is more than seven times (C50=745) that of the wild type. This strongly suggests that the -23 mutation reduces the affinity of the -27→-15 region for HNF4 more than the -19 mutation.

A typical competition gel mobility shift assay with the recombinant protein A-AR DNA binding domain fusion protein and the wild type, -19 and -23 mutant oligonucleotides is shown in Fig 6.8A. The graph from which the C50 values were read is shown in Fig 6.8B. The C50 value for the -23 C→T (C50=59) mutated oligonucleotide is almost 3 times that of the wild type (C50=21) oligonucleotide. The C50 value for the -19 G→C (C50=18) mutated oligonucleotide is slightly lower than the wild type(C50=21). These results suggest that the -23 C→T mutation disrupts the

Figure 6.7 (A) Autoradiograph of a competition gel mobility shift assay using factor IX wild type (lanes 4→7), 19 G→C (lanes 8→11) and -23 C→T (lanes 12→15) mutant oligonucleotides with recombinant HNF4. The titrations for the competition experiments typically involved 50 fold, 100 fold, 200 fold, 400 fold, 800 fold excess of competitor. (B) Graph to obtain C50 value for the wild type (□), -19 (◇), and -23 (○) mutant oligonucleotides. All points plotted are the average of at least 3 experiments. Standard errors are shown, where large enough to be visible.

A


Wild type
-23 C→T
-19 G→C
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

B

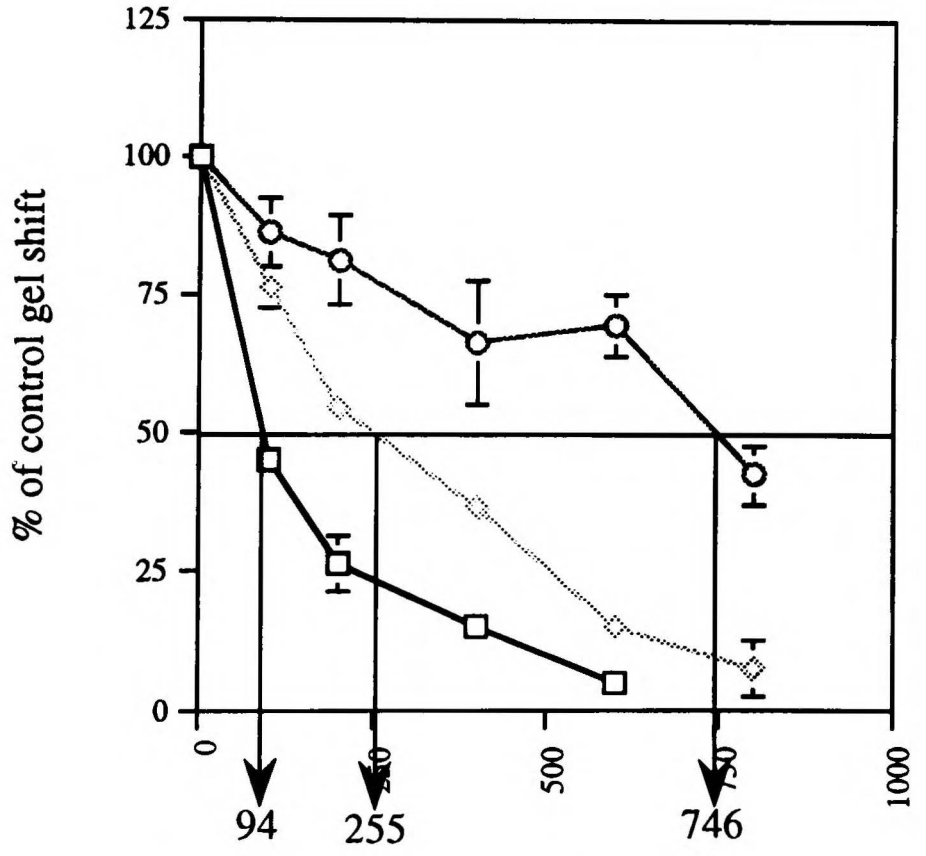
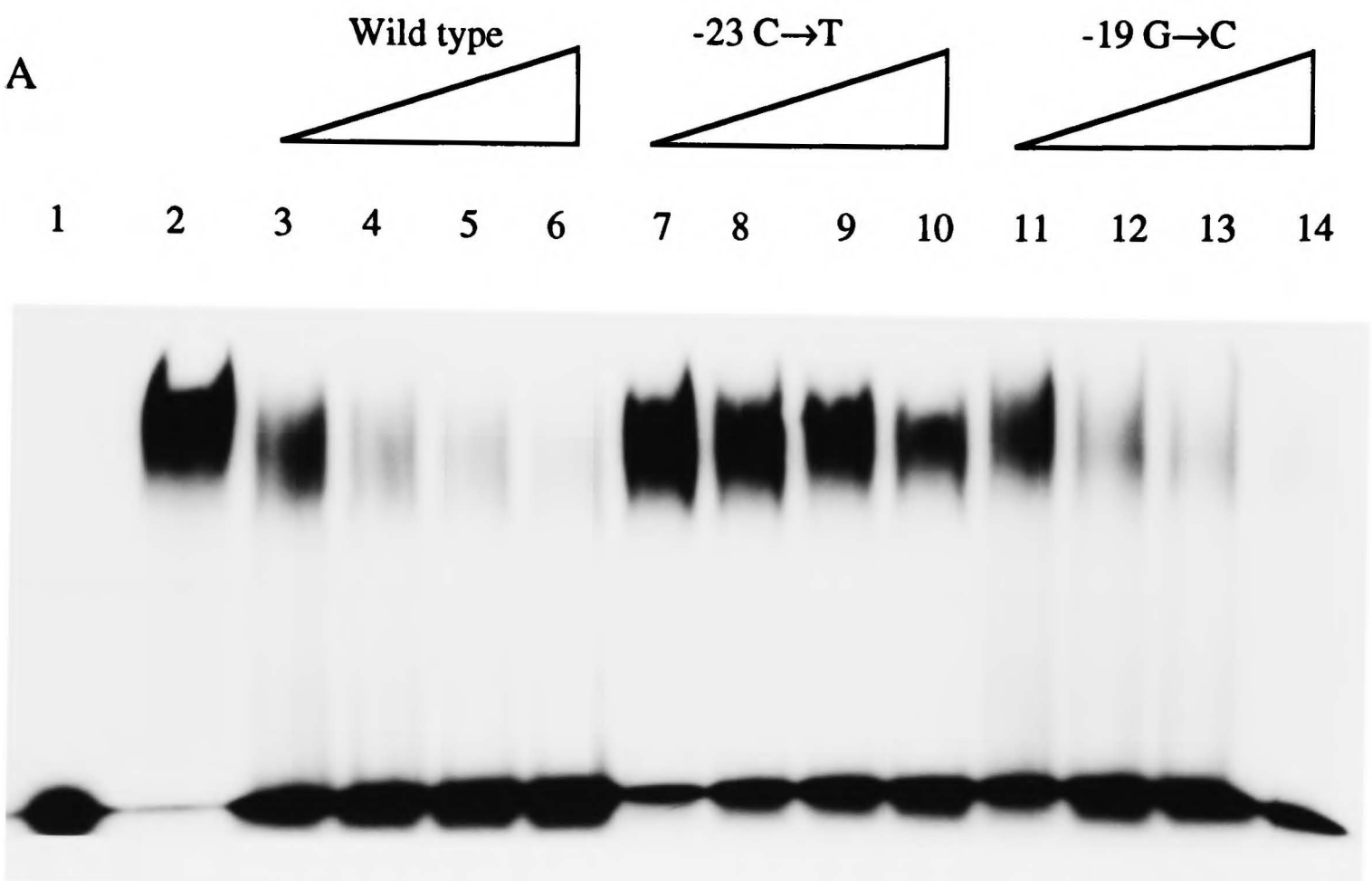
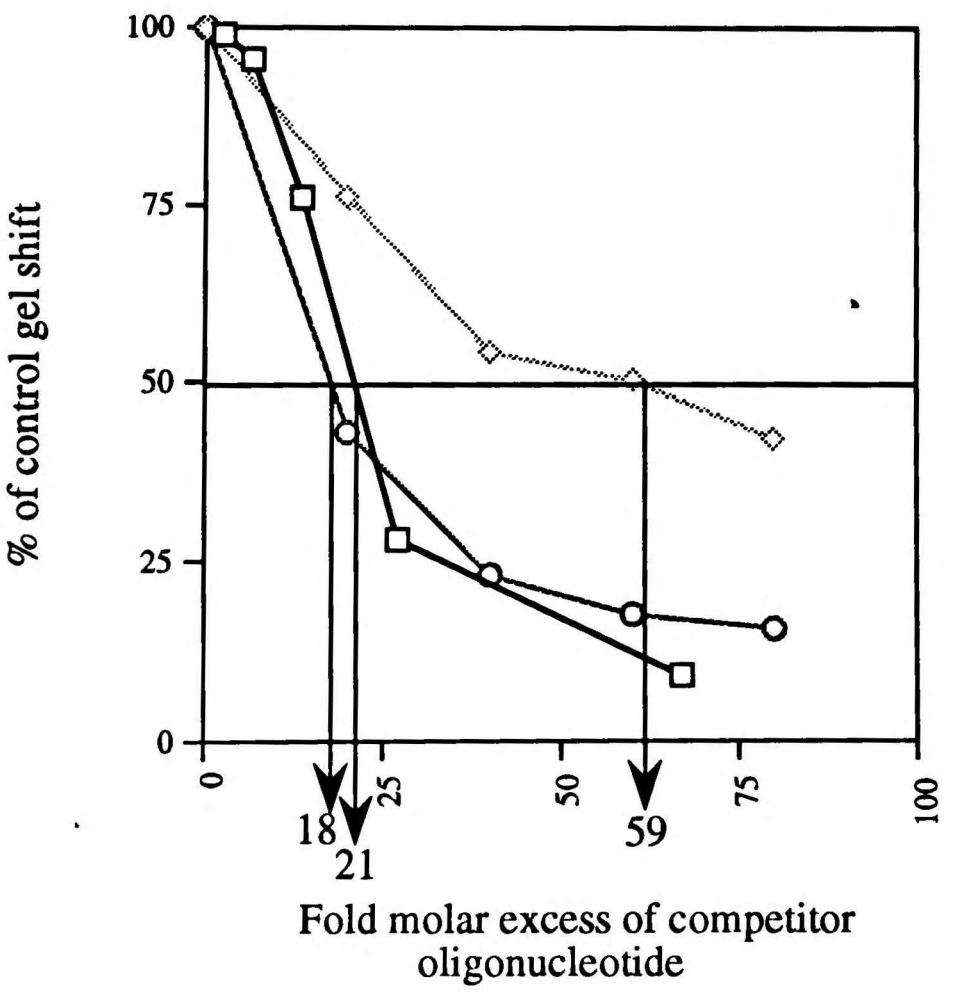

% of control gel shift
125
100
75
50
25
0
0
500
1000
94
255
746
Fold molar excess of competitor oligonucleotide

Figure 6.8 (A) Competition gel shift assay with factor IX wild type probe and wild type (lanes 3→6), -23 C→T (lanes 7→10) and -19 G→C (lanes 11→14) competitor oligonucelotides. In each competitor titration the molar excesses used were: 5 fold, 10 fold, 20 fold, 40 fold, 100 fold. (B) Graph to obtain the C50 values from the gel shift data for wild type (□), -23 (◇), and -19 (○) oligonucleotides.

A


Wild type
-23 C→T
-19 G→C
1 2 3 4 5 6 7 8 9 10 11 12 13 14

B


% of control gel shift
100
75
50
25
0
0
25
50
75
100
18
21
59
Fold molar excess of competitor oligonucleotide

binding of the AR to this sequence, but that the -19 G→C mutation has no adverse effect on this binding.

6.3.4 Transient Transfections to Determine Androgen Receptor/Testosterone Response

The -19 G→C and -23 C→T mutant oligonucleotides were tetramerised and cloned upstream of a tk-CAT reporter gene. The assays performed were the same as described in 6.2.4 and Chapter 5. The activities of the different constructs are shown in Table 6.5. Typical CAT assay results are shown in Fig 6.9A+B.

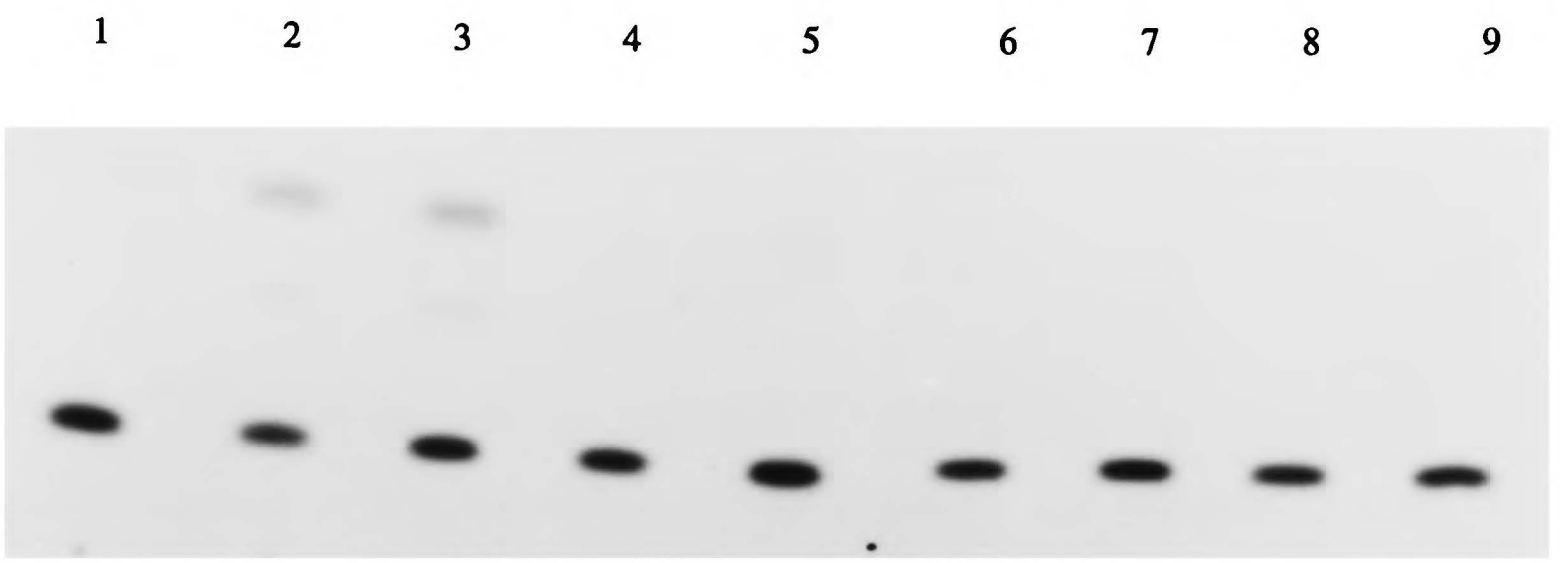
Table 6.5 Transactivation results from CAT assays with the tetramer constructs

Construct	HNF4 Transactivation	AR Transactivation
Wild Type	15.8	20.9
-23 C→T	0.6	1.2
-19 G→C	0.7	10.2

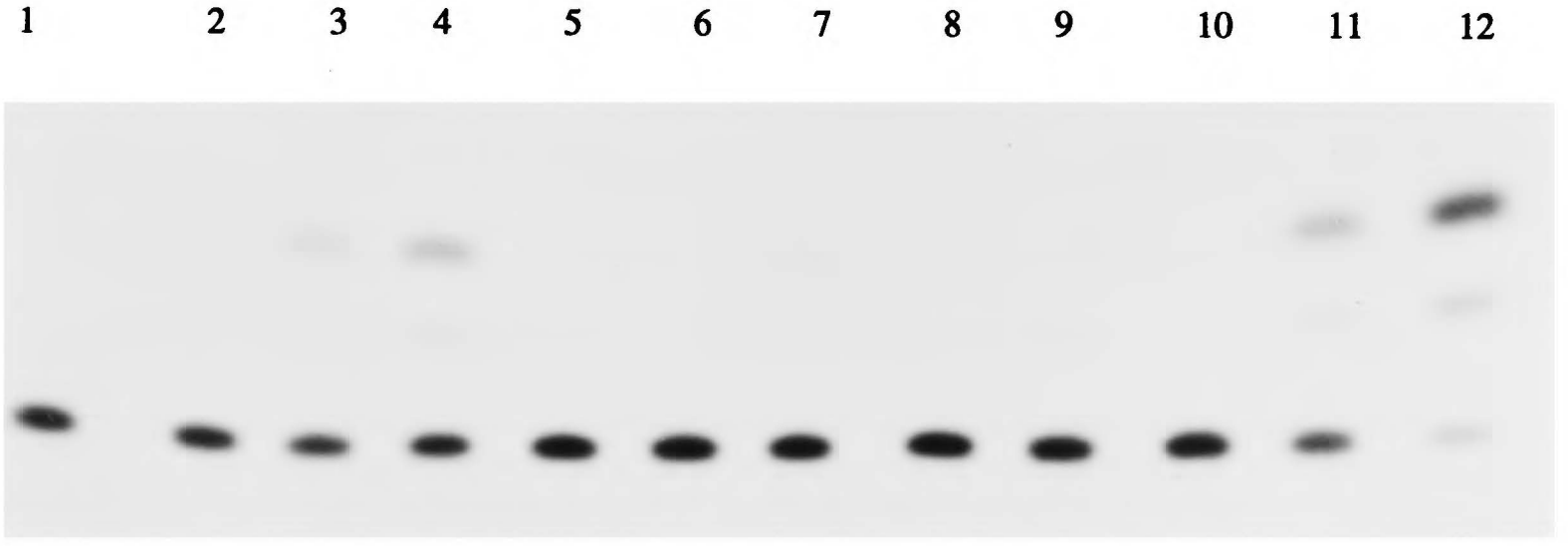
The -23 C→T mutation severely disrupts the transactivation of the tk-CAT reporter, reducing this effect 26 fold (from 15.8→0.6) and 17 fold (from 20.9→1.2) for HNF4 and AR, respectively (Table 6.5). The -19 G→C mutation disrupts the transactivation by HNF4 23 fold (from 15.8→0.7), but only reduces the AR transactivation 2 fold (from 20.9→10.2), (Table 6.5). It should also be noted that the positive control ARE from the rat prostatic binding protein C3 only transactivated 10.9 fold in this assay (Table 6.2), and that the -19 G→C mutation did not appear to disrupt the binding of AR to this sequence (see 6.3.3). These data suggest that the -23 C→T mutation will disrupt the binding of both AR and HNF4 to the region spanning -36→-15 in the factor IX promoter and that the -19 G→C mutation will only impair the binding of HNF4 to the region. The -23 mutation also disrupts the transactivation of this region by both AR and HNF4, whilst the -19 mutation only appears to severely affect the transactivation by HNF4.

Figure 6.9 Transactivation of the tk-CAT reporter construct by factor IX wild type and mutant oligonucleotide tetramers with (A) HNF4. Lane order: 1= wild type, minus HNF4; 2+3= wild type, plus HNF4; 4= -19 G→C, minus HNF-4; 5+6= -19 G→C, plus HNF4; 7= -23 C→T, minus HNF4; 8+9= -23 C→T, plus HNF4. (B) AR. Lane order: 1=wild type, minus AR, plus 5α-dihydrotestosterone; 2= wild type, minus 5α-dihydrotestosterone, plus AR; 3+4= wild type, plus AR and 5α-dihydrotestosterone; 5= -23 C→T minus AR, plus 5α-dihydrotestosterone; 6= -23 C→T, minus 5α-dihydrotestosterone, plus AR; 7+8= -23 C→T, plus AR and 5α-dihydrotestosterone; 9= -19 G→C, minus AR, plus 5α-dihydrotestosterone; 10= -19 G→C, minus 5α-dihydrotestosterone, plus AR; 11+12= -19 G→C, plus AR and 5α-dihydrotestosterone.

A



B



6.4 Discussion and Summary

In this chapter three different point mutations in the promoter region of the factor IX gene, which have been found to cause haemophilia B, have been studied. The effects of these three mutations on the transactivation of the promoter region, and the binding of transcription factors to the promoter, have been investigated.

Each of the three mutations reduced the HNF4 transactivation of the minimal promoter region (-220→+43), (Fig 6.2 and Fig 6.6). All of the reductions in the HNF4 transactivation of these constructs were found to be significant using the student's t test (see Tables 6.1 and 6.3). Although the reductions were proved to be significant, the transactivations of the mutated promoter regions were higher than expected from previous studies (Crossley et al., 1992, Reijnen et al., 1993, Kurachi et al., 1994). It must be noted, however, that this is the first comparative study to use the full minimal promoter region (-220→+43), other studies have used shorter fragments (e.g. -189→+21 was used by Crossley et al., 1992) which lack one or more of the cis elements now identified (see Fig 1.4). The higher than expected transactivation of the mutated promoters in HepG2, PLC/PRF/5, and HeLa cells may be due to the presence of the HNF4 site spanning nucleotides +15→+36 reported by Naka and Brownlee, (1996), which has been absent from the constructs used in many previous studies (Crossley et al., 1992, Boccia et al., 1996, Kurachi et al., 1993). This site may be able to partially compensate for the disruption of the HNF4 site between -27→-15.

The -26 G→A mutation found in the factor IX promoter of patient UK232 disrupts overlapping binding sites for AR and HNF4. The disruption of the HNF4 site causes the downregulation of the factor IX gene before puberty in this patient, and the disruption of the ARE prevents any recovery of these plasma factor IX levels through the course of puberty. This patient is 47 years of age and still has low plasma factor IX clotting activity, and still has a bleeding disorder.

The other two mutations studied were found in patients HB 388 and HB 396 and reported in Ketterling et al., (1995). The -23 C→T mutation, found in patient HB

388 (Ketterling et al., 1995), appears to severely disrupt binding and transactivation by AR and HNF4 of the -36→-15 region. The 19 G→C mutation, found in patient HB 396 (Ketterling et al., 1995), only appears to impair the binding and transactivation of this region by HNF4. Both of these patients are still in childhood so it is not known whether or not they will recover from their bleeding disorders.

Chapter 7 Characterising Patient Haemophilia B Liverpool 2, and the Reason for the High Frequency of Mutations at Nucleotide -6 in the Factor IX Promoter Region

7.1 Introduction

The Haemophilia B Leyden phenotype is associated with many different mutations in the factor IX promoter (Giannelli et al., 1996). A less severe form of Haemophilia B Leyden was reported (Crossley et al., 1990) as being associated with a G→A mutation at nucleotide -6 in the promoter region. In this patient, the plasma level of factor IX recovered from 13% at age 9 years to 70% at age 30 years. This mild Leyden phenotype was termed Haemophilia B High Wycombe (Crossley et al., 1990). The mutation at -6 is known to down-regulate the factor IX promoter in transactivation assays in hepatoma cells in culture (Naka and Brownlee, 1996). It was also identified as disrupting the binding of the transcription factor HNF4 to this region (Naka and Brownlee, 1996).

This mutation occurs at a CpG dinucleotide in the promoter region. These dinucleotides are known to be hotspots for mutation (Chen et al., 1991, Knobloch, et al., 1993, Green et al., 1991) as they can be methylated and the methyl C is then susceptible to deamination, which converts it to a T and so mutates the DNA sequence. Mutations at some CpG dinucleotides in the factor IX gene occur at a high frequency due to the high susceptibility of these dinucleotides to mutation and, in some cases, a founder effect exists which allows the frequency of the mutation to be amplified in the patient population (Ketterling et al., 1991).

7.2 Haemophilia B Liverpool 2

7.2.1 Case History

A Caucasian boy who had been investigated at age 7 years due to oral ulceration and a history of significant bruising was referred to Professor Brownlee by Dr. P. Bolton-Maggs (Royal Liverpool Children's Hospital) was diagnosed as having mild

haemophilia B with a factor IX clotting level of 9% (0.09u). There was no family history of bleeding and he had never suffered from any spontaneous bleeding. At age 13 his plasma factor IX clotting level rose to 14% (0.14u) and at 17 it was 44% (0.44u). A year later it was 62% (0.62u) (see Fig 7.1). After diagnosis the patient received occasional replacement therapy for accidents and sporting injuries until his plasma factor IX levels increased, after which no treatment was necessary.

7.2.2 Sequence Characterisation and Carrier Detection

The factor IX promoter region (-221 to +43) was PCR amplified from the patients genomic DNA (see 2.10). The fragment was purified (see 2.5, results not shown), and then sequenced directly (see 2.11) using an ABI automated sequencer (see Fig 7.2). A G→A mutation at nucleotide -6 was detected. No other mutation was detected.

In order to perform a carrier detection study, samples of DNA from the patient's mother and maternal grandmother were obtained. These samples were sequenced in the same way as the patient's DNA. The mother can be seen to be heterozygous for the mutation but the maternal grandmother is homozygous for the wild type sequence (see Fig 7.3). This could mean either that the mutation was *de novo* in the mother's gene or that it was inherited from the maternal grandfather. It was not possible to obtain a DNA sample from the maternal grandfather, so this possibility could not be ruled out. In a previous study Stowell et al., (1993), found that a -6 G→A mutation had been inherited from the maternal grandfather.

7.3 Haplotype analysis of patients with -6 G→A mutations

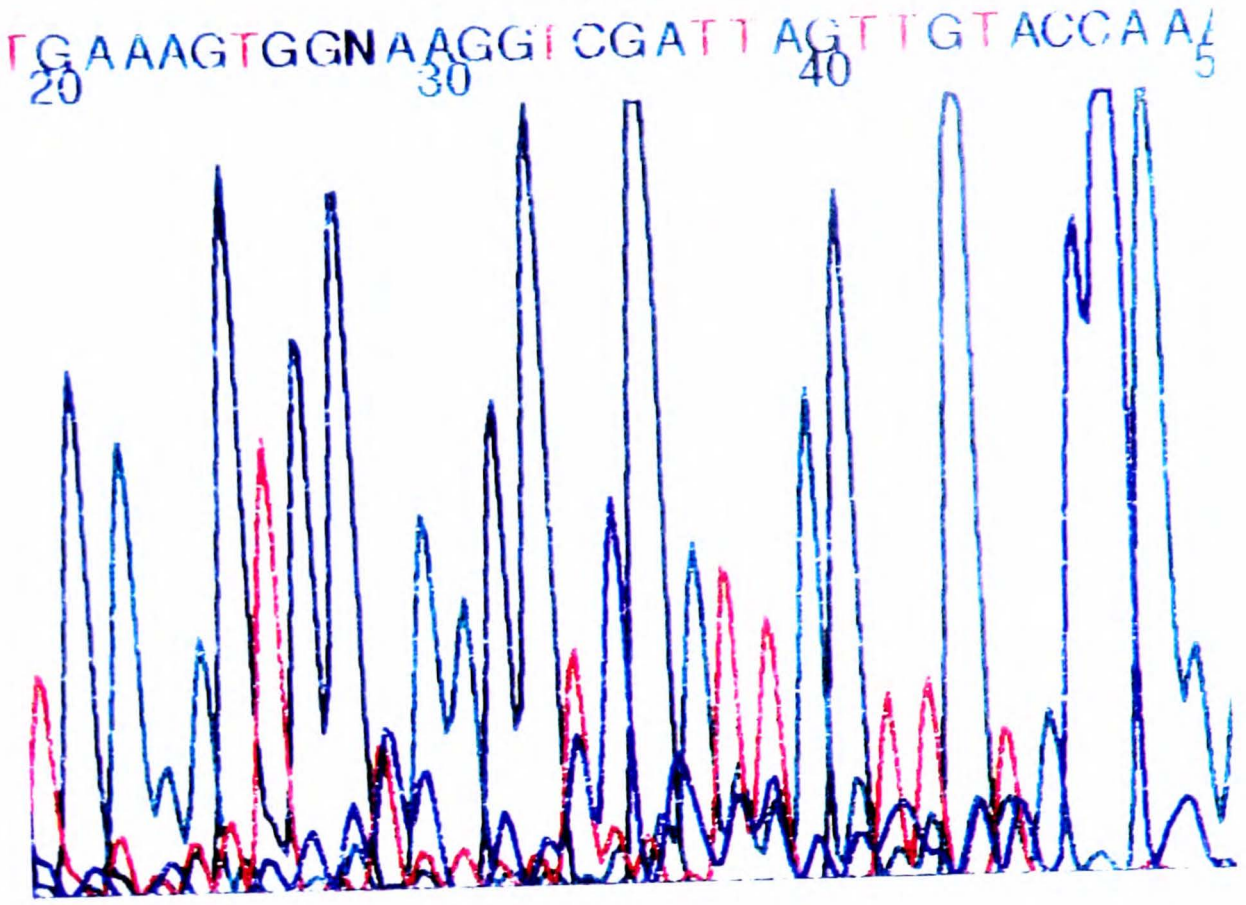
In the latest Haemophilia B database of mutation (Giannelli et al., 1996) 13 patients have now been identified as having the G→A mutation at this nucleotide. When multiple patients possess mutations at the same position in the factor IX gene restriction fragment length polymorphism analysis of the factor IX gene in these patients can be useful in determining whether or not these mutations are independent or



Figure 7.1 Graph showing the rise in the plasma factor IX level in patient Liverpool 2 with age.

Wild Type Sequence

TCGA



Mutant Sequence

TTGA

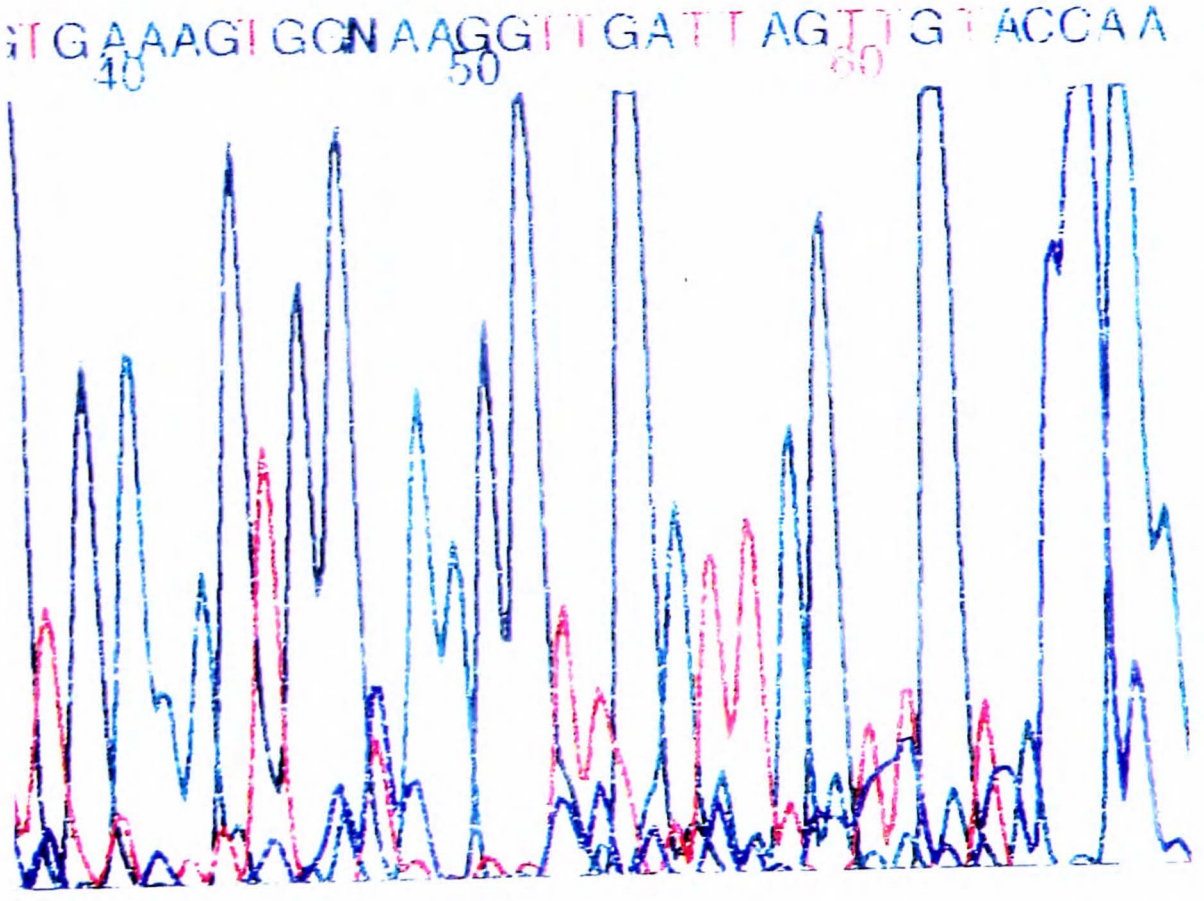
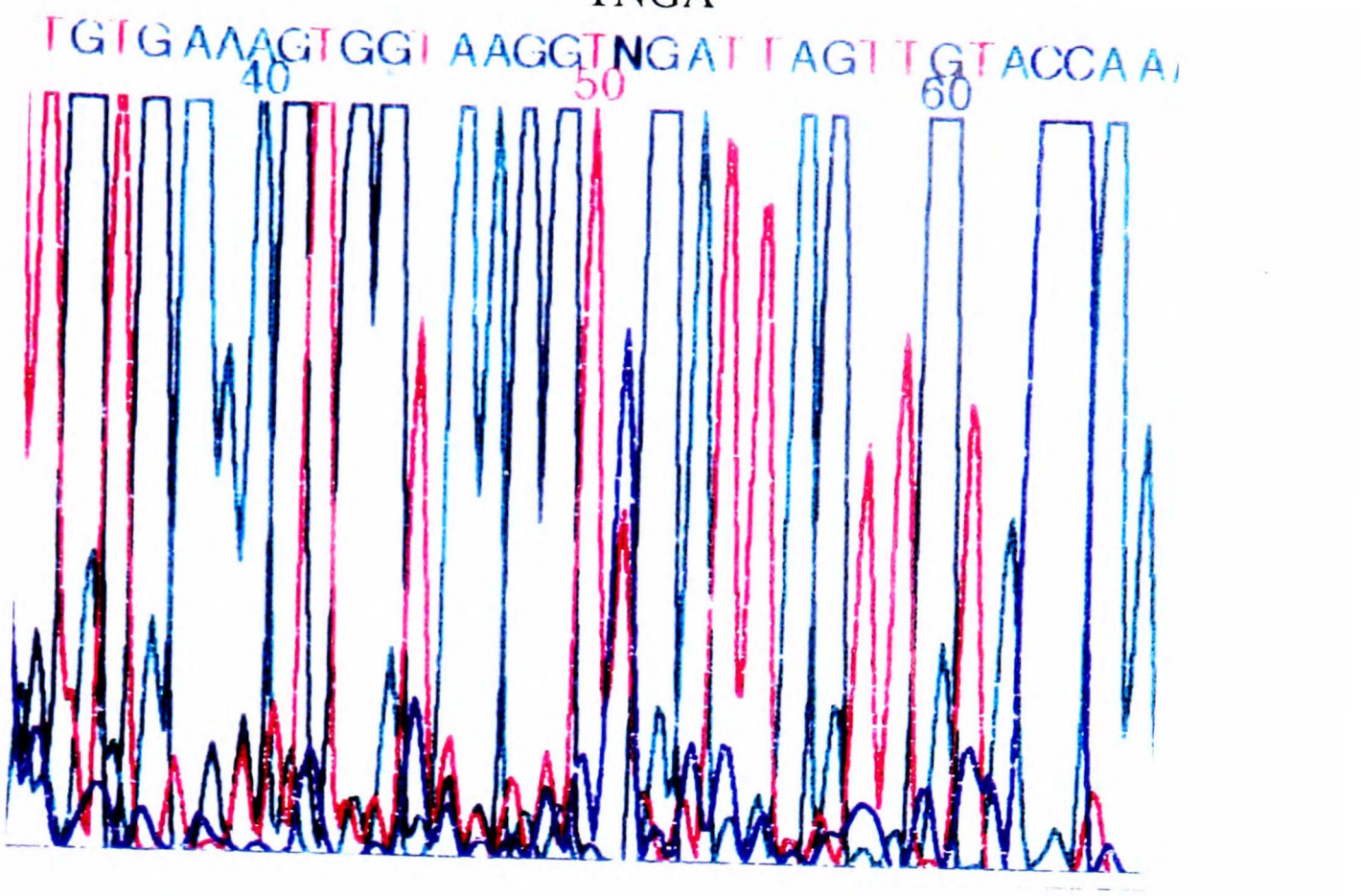


Figure 7.2 Automated sequence data showing the presence of the -6 G→A mutation the patient's DNA the wild type sequence is included for comparison. The mutation is C→T in antisense strand, G→A in sense strand.

Mother's DNA Sequence



Maternal Grandmother's

DNA Sequence

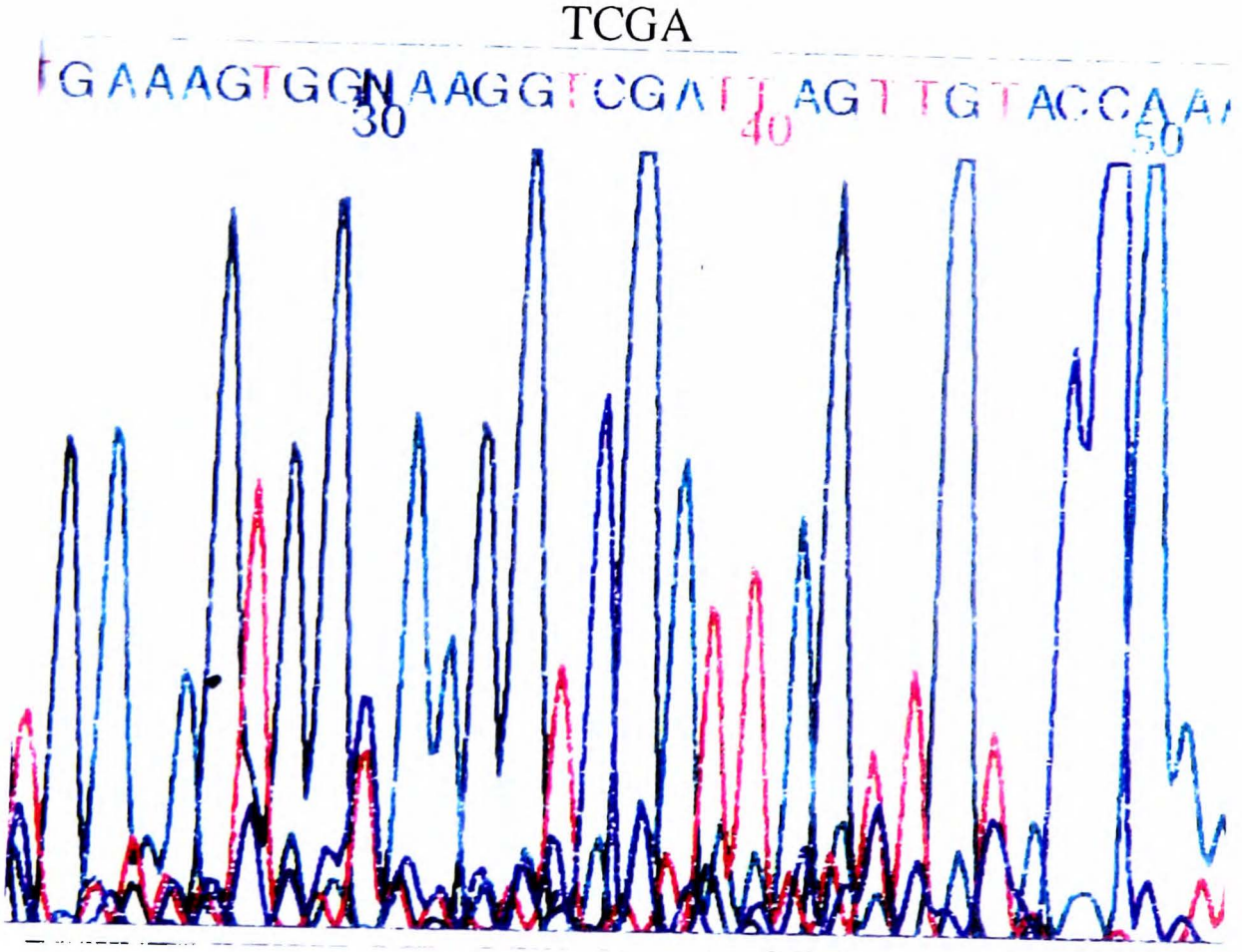


Figure 7.3 Automated sequence data showing the presence of the -6 G→A mutation (base number 51) in the mother's DNA but not in the grandmother's. The N at position 51 denotes the presence of two different bases (C and T) on the two factor IX alleles carried by the mother.

not (Ketterling et al., 1991). This type of knowledge can be important in tracing mutations through families, populations and, in some cases, in tracing relationships between populations (Figueiredo et al., 1992, 1994). The factor IX gene has many, well characterised polymorphic loci (Stowell et al., 1993, Winship et al., 1993, Sarkar et al., 1991) and consequently, is a much studied gene in this field of molecular biology.

The factor IX haplotype of Liverpool 2 was constructed using 8 intragenic and immediately extragenic polymorphic loci (Stowell et al., 1993, Winship et al., 1993) in the factor IX gene (see Fig 7.4). To determine whether the high frequency of this mutation in the population was due to recurrent, independent mutation at a CpG dinucleotide, or whether a founder effect was involved, the factor IX gene haplotypes of this patient and 8 other patients, with -6 G→A mutations, worldwide (Giannelli et al., 1996) were determined.

All of the polymorphic loci could be examined directly by PCR amplification of the region of DNA in which they lay, followed by the appropriate restriction digest of the PCR fragment (see 2.7.2). The *Dde*I polymorphism is a VNTR (Variable Number of Tandem Repeats)-like sequence consisting of short repeated stretches of DNA having many potential alleles (Figueiredo et al., 1994) so this was characterised directly by the size of the PCR fragment (Table 7.1). The presence of a restriction site at the locus is denoted in the haplotype as a "+" and its absence with a "-". The nomenclature adopted for the *Dde*I locus is from Jacobson et al., (1993). The expected band patterns for the polymorphic loci studied is shown in Table 7.1.

Fig 7.5 shows the characterisation of some of the polymorphisms used to construct the haplotypes for the nine patients studied. In each of gels 1→3, the polymorphisms from patient Liverpool 2 are compared with two other UK patients (haemophilia B Liverpool (Stowell et al., 1993) and haemophilia B High Wycombe (Crossley et al., 1990)) and a control sample of known haplotype. The results presented in Fig 7.5 identify the patient Liverpool 2 as being *Taq*I -; *Hha*I +; *Dde*I AB;

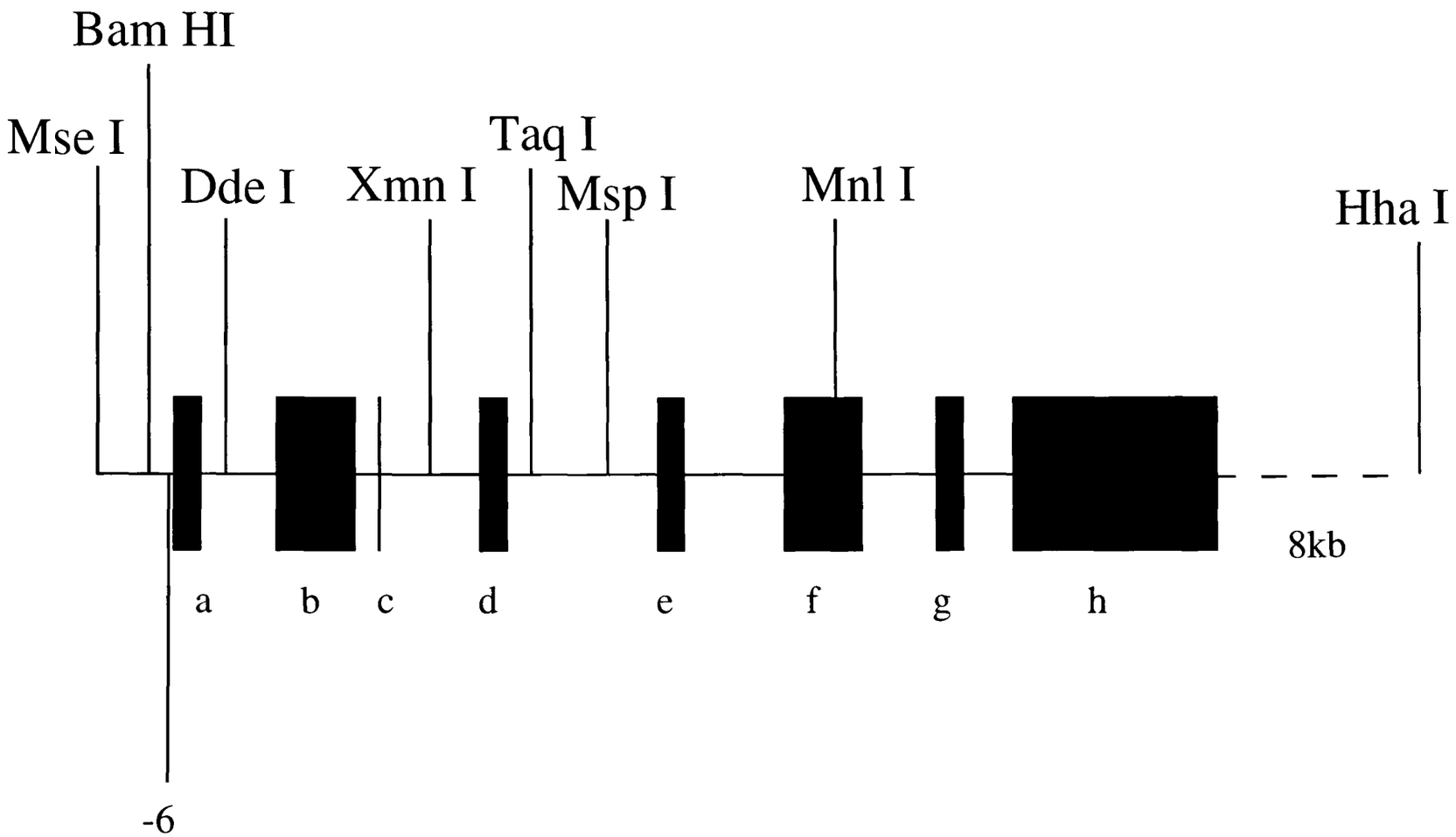


Figure 7.4 The location of the polymorphic loci used in the haplotype analysis of the 9 patients with the -6 G→A mutation in the factor IX promoter region. The position of nucleotide -6 relative to the polymorphic loci is shown. The dotted line indicates that the HhaI locus is 8kb 3' to the gene

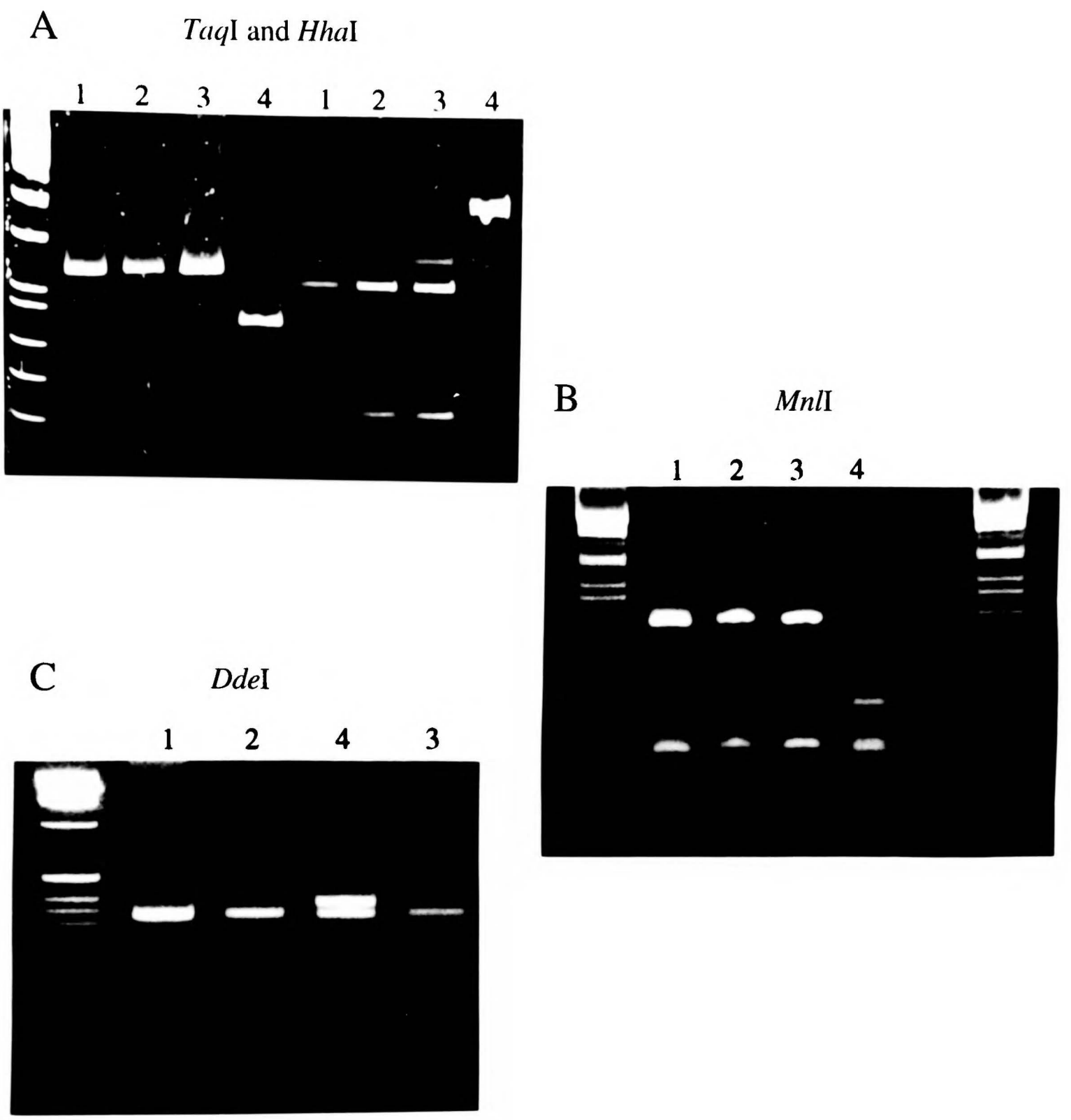


Figure 7.5 Three ethidium bromide stained gels showing the determination of four of the eight polymorphisms used to obtain the factor IX haplotypes for the nine patients with the -6G→A mutation in the promoter of their factor IX genes. The polymorphisms determined by these gels were (A) *Taq*I and *Hha*I; (B) *Mnl*I; (C) *Dde*I. The sizes of all of the fragments are shown in Table 7.1. The DNA markers used for the gels were, ϕx174 digested with *Hinf*I (A+B), and the 1Kb ladder (Gibco) gel C (see 2.6.1). The key for each gel is: 1= Haemophilia B Liverpool 2; 2= Haemophilia B Liverpool 1; 3= Haemophilia B High Wycombe; 4= Marker of known haplotype.

and *Mnl*I -. Another four polymorphisms in the factor IX gene were used to construct the complete gene haplotypes for each of the nine patients.

Table 7.1 The expected DNA fragment patterns for the factor IX polymorphic loci studied in the haplotype analysis. The sizes of the PCR fragments generated, and the digest fragments (in bp) are shown.

Polymorphic locus	Size of PCR Fragment (bp)	Positive	Negative
*Xmn*I	222	154+68	222
*Taq*I	163	124+39	163
*Hha*I	230	150+80	230
*Mnl*I	407	122+158+127	270+127
*Bam*HI	450	234+216	450
*Msp*I	657	313+344	657
*Dde*I	319/369	369	319

Four different haplotypes were identified in the 9 patient samples which were analysed (see Table 7.2), and named haplotypes I→IV. A small sample (21) of X chromosomes from the UK Caucasian population (see Appendix 2) was analysed in order to determine the frequencies of haplotypes I→IV in the wild type population. The frequencies of haplotypes I→IV in this study of wild type X chromosomes are shown in Table 7.3. Though haplotypes II and IV were not represented in our small sample, we compared the frequencies of other haplotypes in this sample to those found in a larger study of a Caucasian population (Graham et al. 1991), results not shown. The small sample did appear to be representative. This indicates that the frequencies of haplotypes II and IV is probably <5% in the wild type Caucasian population (i.e. this is why none of the 21 X chromosomes sampled from the wild type population possessed this haplotype).

Table 7.2 The results of the haplotype study of 9 patients with a G→A mutation at nucleotide -6.

Haplotype	*Mse*I	*Bam*HI	*Dde*I	*Xmn*I	*Taq*I	*Msp*I	*Mnl*I	*Hha*I	Patients
I	-	-	AB	-		+	-	+	6
II	+	-	AB	+	+	-	+	-	1
III	+	-	AB			-	-	-	1
IV	-	-	A_2B_2			+	-	+	1

Table 7.3 The frequencies of haplotypes I→IV in the sample of wild type X chromosomes from the UK Caucasian population (see Appendix 2)

Haplotype	Frequency (%)
I	19
II	0
III	29
IV	0

7.4 Independent mutation events, or a founder effect?

Different haplotypes for two mutated factor IX alleles do not necessarily indicate independent origins for the mutations in these genes. Crossover events can occur at meiosis, which although rare between polymorphic loci which are in such close genetic proximity as these, can confuse the interpretation of a haplotype analysis such as this. However, if more than one crossover event is required to form one allele's haplotype from another, it is highly probable the alleles have mutated independently. If haplotypes I→IV are considered (Table 7.2), more than one crossover event is required to convert any one haplotype into another, except for alleles I and IV. Therefore, is highly probable that at least 3 of the 4 haplotypes represent independent mutations.

Table 7.2 shows that six patients of the nine studied possessed haplotype I. The probability of all of these mutations being independent may be calculated using the frequency of haplotype I in the UK wild type Caucasian population (Table 7.3) and the binomial distribution (see 2.8.1). Assuming that the -6 G→A mutation is randomly distributed amongst all of the factor IX haplotypes, the probability of all six being independent events can be seen to be 0.2% (see Table 7.4). This result strongly suggests that it is improbable the six mutations are all independent events. From Table 7.4 it can be seen that the most probable number of independent events in haplotype I in a random sample of nine patients is one or two (probabilities 0.32 and 0.3 respectively). It is also worth noting that all four UK patients have haplotype I, thus strengthening the probability that these patients are related in some way and do not represent independent mutation events.

Table 7.4 Calculating the probabilities of varying numbers of independent mutation events occurring in haplotype I, in a sample of 9 patients, using a binomial distribution.

Number of independent events in haplotype I in the sample	Probability
0	$1(0.19^0 \times 0.81^9)=0.15$
1	$9(0.19^1 \times 0.81^8)=0.32$
2	$36(0.19^2 \times 0.81^7)=0.30$
3	$84(0.19^3 \times 0.81^6)=0.163$
4	$126(0.19^4 \times 0.81^5)=0.057$
5	$126(0.19^5 \times 0.81^4)=0.013$
6	$84(0.19^6 \times 0.81^3)=0.002$
7	$36(0.19^7 \times 0.81^2)=0.0002$
8	$9(0.19^8 \times 0.81^1)=0.00001$
9	$1(0.19^9 \times 0.81^0)=0.0000003$

A χ^2 goodness of fit test, incorporating Yates' correction (see 2.8.2) for small sample size was performed on the two sets of data. The aim of this test was to determine whether or not the frequency of haplotype I in the sample of 21 wild type X chromosomes (4/21) was significantly different to that found in the group of 9 patients studied (6/9). The result of the test is that the probability this difference is not significant is <5%. This is further evidence which suggests that the frequency of the mutation in haplotype I is not simply due to chance and that recurrent mutation is not the only reason that this mutation occurs with such frequency in the patient population.

7.5 Discussion and Summary

It is important to identify the mutations in the factor IX gene which cause haemophilia for a number of reasons. Mutations in the protein itself allow certain structure function relationships to be analysed. Mutations in the promoter also allow nucleotides important for the regulation of the gene to be identified. The nature of the mutation can have vital implications for the genetic counselling that a family receives (Knobloch et al., 1993), especially in the case of the Leyden phenotype where the clinical symptoms improve in adulthood (see Chapter 6). When a mutation is identified, the carrier detection procedure becomes easier and identification of potential carriers makes the management of the condition easier as females from an affected pedigree can make informed decisions about pregnancies. This is particularly true in regions where standard forms of therapy for the disease are scarce or unavailable.

In this chapter the characterisation of the factor IX gene mutation in a patient with a haemophilia B Leyden phenotype was described. A carrier detection was performed upon this pedigree and the mutation traced to his mother, but not to the maternal grandmother. The bleeding disorder caused by the -6 G→A mutation is mild and has been known to be passed on through the male lineage (Stowell et al., 1993) without these individuals being aware of their condition. The -6 G→A mutation occurs at a CpG dinucleotide and is particularly common it the haemophilia B patient population (13 cases in the latest haemophilia B mutation database, Giannelli et al.,

1996). As CpG dinucleotides are known to be susceptible to mutation, the factor IX gene haplotypes of nine patients with the -6 G→A mutation were determined to try to ascertain whether the high frequency of this mutation in the haemophilia B population was due to recurrent mutation at this CpG dinucleotide or whether a founder effect played a role.

The results presented here suggest that, in the 9 patients studied, both a founder effect and recurrent mutation at this CpG dinucleotide contribute to the frequency of the mutation in the haemophilia B population. Previous studies (Ketterling et al., 1991) have also implicated both of these effects in causing the high frequency of a G→A mutation at position 10430 (which changed a Gly for Ser in the factor IX protein) which occurred in 14 patients.

Other reports (Chen et al., 1991, Knobloch et al., 1993, Green et al., 1991) all show that CpGs are hotspots for mutation in the factor IX gene. They found overall that transition mutations at CpG dinucleotides corresponded to 35-50% of all single base changes found in the patients they investigated. It is known that the cytosine in this dinucleotide can be 5 methylated and the resulting 5methylC can then be deaminated to form thymine (Jones et al., 1992). It is not known whether the CpG dinucleotide at -6 is methylated *in vivo*. At present no methylation sensitive restriction enzymes are available which are able to discriminate between C and 5methylC at this position. Thus, an *in vivo* analysis of the methylation state of this nucleotide would involve a study of the sensitivity of this base to reagents such as sodium bisulphite (Grigg and Clark, 1994).

Other reasons for the high frequency of this mutation in the haemophilia B patient population include the fact that nucleotide -6 is in a position of importance for the regulation of the transcription of the factor IX gene (Naka and Brownlee, 1996), coupled with the mild nature of the bleeding disorder caused by this mutation. It is already known that this condition can be passed through the male lineage without being detected (Stowell et al., 1993). The mildness of the condition would have the effect of increasing the "half life" of the mutation (Knobloch et al., 1993) through the

subsequent generations resulting in the establishment of a founder effect and increasing the frequency of this mutation in the population relative to a mutation causing a more severe phenotype.

Chapter 8: Summary and General Discussion

8.1 Summary

Crossley et al., (1992) reported that the failure of a patient with a -26 G→C mutation in the factor IX promoter region to recover from his bleeding disorder at puberty was due to the disruption of an ARE located between -36→-22. The work described in this thesis was performed to gain more information about the nucleotides in this region which are important for the response to AR and testosterone, and to add further evidence to the hypothesis that the ARE is important for the recovery of patients bearing mutations in the HNF4 site between nucleotides -27→-15.

Chapter 3 describes the production and partial purification of a protein A-AR DNA binding domain fusion protein which was used in the *in vitro* study of the binding of mutant factor IX ARE oligonucleotides in Chapter 4. In this *in vitro* study firstly, positions -6, -3, 3, 4, and 6 were identified as being critically important for the binding of the ARE to this fusion protein and secondly, mutations at positions -7, -2, and 2 also disrupted the binding of the fusion protein, though to a lesser degree than the first group of mutations.

The mutant ARE oligonucleotides were then tested in an *in vivo* assay in Chapter 5. The oligonucleotides were multimerised upstream of the HSV tk promoter in a CAT reporter construct. Due to problems involving the rearrangement of constructs bearing four copies of these oligonucleotides, transient transfections were conducted using oligonucleotide tetramer/dimer construct mixtures. The transient transfection were performed in HeLa cells with, and without, the presence of an AR expression vector and the androgen hormone 5α-dihydrotestosterone. This study identified the positions -6, -4, -3, 2, 3, 4, and 6 as being critical for the AR and 5α-dihydrotestosterone-mediated transactivation of the HSV tk promoter. Positions -7, -2, and possibly 7 disrupted this transactivation to a lesser degree than the first group of mutations.

Chapter 6 concerned the study of the functional effects of 3 novel mutations identified in patients (-26 G→A, -23 C→T, and -19 G→C). The response of the factor

IX minimal promoter region to HNF4 was impaired by all of these mutations. The effects of these mutations on the binding of both HNF4 and AR were studied, and the response of the mutant oligonucleotide tetramer/dimer constructs was investigated. The results presented further evidence that mutations at -26 disrupt both the HNF4 site and the ARE. This correlated well with the clinical data for patient UK 232 (-26 G→A) who has still not recovered from his bleeding disorder and is now 47 years old. The -23 C→T mutation also disrupted both the HNF4 site and the ARE whereas the -19 G→C mutation appeared only to disrupt the HNF4 site. Both of these patients are in childhood so it is not yet known whether their plasma factor IX levels will improve at puberty.

A -6 G→A mutation at a CpG dinucleotide was identified in the factor IX promoter region of patient Liverpool 2 in Chapter 7. This mutation was also then traced to his mother, who was heterozygous for the mutant allele, but was absent from his maternal grandmother's alleles, suggesting that this mutation may have been passed through the male line in this pedigree, or may have been *de novo* in the mothers gametes. The high frequency of the -6 G→A mutation in the patient population was examined. It was found to be highly probable that both a founder effect and recurrent mutation at a CpG dinucleotide are responsible for this high frequency.

8.2 Discussion Concerning the *in vitro* Study

8.2.1 Alternative Data Analysis Techniques

To perform a more detailed kinetic study than that in Chapter 4 (where C50 values were reported) with these mutant ARE oligonucleotides, the conditions used by Riggs et al., (1970), Fried and Crothers, (1981), and Chodosh et al., (1986) would have to be employed. These techniques allowed the calculation of dissociation and association constants provided that the DNA probe is in considerable (usually at least two orders of magnitude, Riggs et al., 1970, Chodosh et al., 1986) molar excess in the binding reaction compared to the DNA binding protein. If the binding of the fusion

protein to the probe oligonucleotide is considered to be a simple bimolecular reaction then it can be represented by the reversible equation 1:

Equation 1 $$P + AR \leftrightarrow P.AR,$$

where P=probe oligonucleotide; AR= AR protein A fusion protein; P.AR= probe/fusion protein complex; and $\leftrightarrow$= reversible reaction.

then the rate of change of the concentration of the P.AR complex is given by:

Equation 2 $$\frac{-d(P.AR)}{dt} = k_d [P.AR] - k_a [P] [AR]$$

where k_d is the dissociation constant for the breakdown of the P.AR complex and k_a is the association constant for the formation of P.AR form free P and AR.

When a competition gel shift is performed and a large (two orders of magnitude is usual, Riggs et al., 1970, Chodosh et al., 1986) excess of unlabelled oligonucleotide is added to the binding reaction, then the [AR] becomes negligible as the equilibrium in equation 1 is driven to the right by the excess of free oligonucleotide and all of the protein binds to the oligonucleotide. Therefore, equation 2 can be rewritten:

Equation 3 $$\frac{-d(P.AR)}{dt} = k_d [P.AR]$$

Under these conditions, the dissociation of P.AR is initially first order and the dissociation constant can be measured by following the appearance of the free radioactive probe, until the concentration of [P] increases to such a level that the second term in equation 2 is no longer negligible and the dissociation of [P.AR] no longer follows first order kinetics. In the study of Chapter 4, the conditions under which the protein A-AR competition gel mobility shifts were carried out were approximately equimolar (see 2.18 and 2.20) and, in most cases, significant competition of the gel shift was already observed with 30 fold molar excess of competitor oligonucleotide.

Under these conditions it was not possible to use the first order approximation of Riggs et al., (1970).

A heterologous competition gel mobility shift assay (i.e. one where the probe oligonucleotide and competitor oligonucleotide are different DNA sequences) is also complicated by the fact that any dissociation constant that may be calculated for the competitor oligonucleotide binding to the fusion protein is only "apparent". In these types of assay, any such constant will be a kinetic composite of the binding of the unlabelled competitor oligonucleotide and the labelled probe oligonucleotide to the protein. Moreover, the fusion protein may be binding to each of the half sites of the ARE oligonucleotides independently, or with some form of cooperativity. This will complicate the kinetics of the binding, since the binding reaction could not then be regarded as a simple bimolecular one as in equation 1. It is also desirable to use protein samples which are as pure as possible when conducting a detailed kinetic study of a protein-DNA binding reaction. The extract used in this study contained a protein A-AR DNA binding domain fusion protein which was approximately 40% pure (see Fig 3.2).

This study is a strictly comparative one in which a large number of factor IX mutant ARE oligonucleotides are being studied. Therefore, even though the C50 value has very little real kinetic information contained in it, it is acceptable for a comparison of closely related set of oligonucleotides such as these.

8.2.2 The ARE Consensus Sequence

The binding data which is presented in Chapters 4 and 6 of this thesis were obtained from an *in vitro* competition gel mobility shift assay (Table 4.3) and it is in broad agreement with the consensus study which was performed by Roche et al., (1992), see Fig 4.13. Roche et al., (1992) defined a consensus using a gel shift assay with the same protein-A AR DNA binding domain fusion protein and a pool of labelled random oligonucleotides as the initial probe. The study then involved the successive isolation of the subset of oligonucleotides which bound to the protein A-AR fusion protein (De Vos et al., 1991), by PCR amplification, followed by further gel mobility

shift analysis of the oligonucleotide mixture. In this way Roche et al., (1992) enriched the oligonucleotide mixture for sequences which bound favourably to the protein A-AR fusion protein.

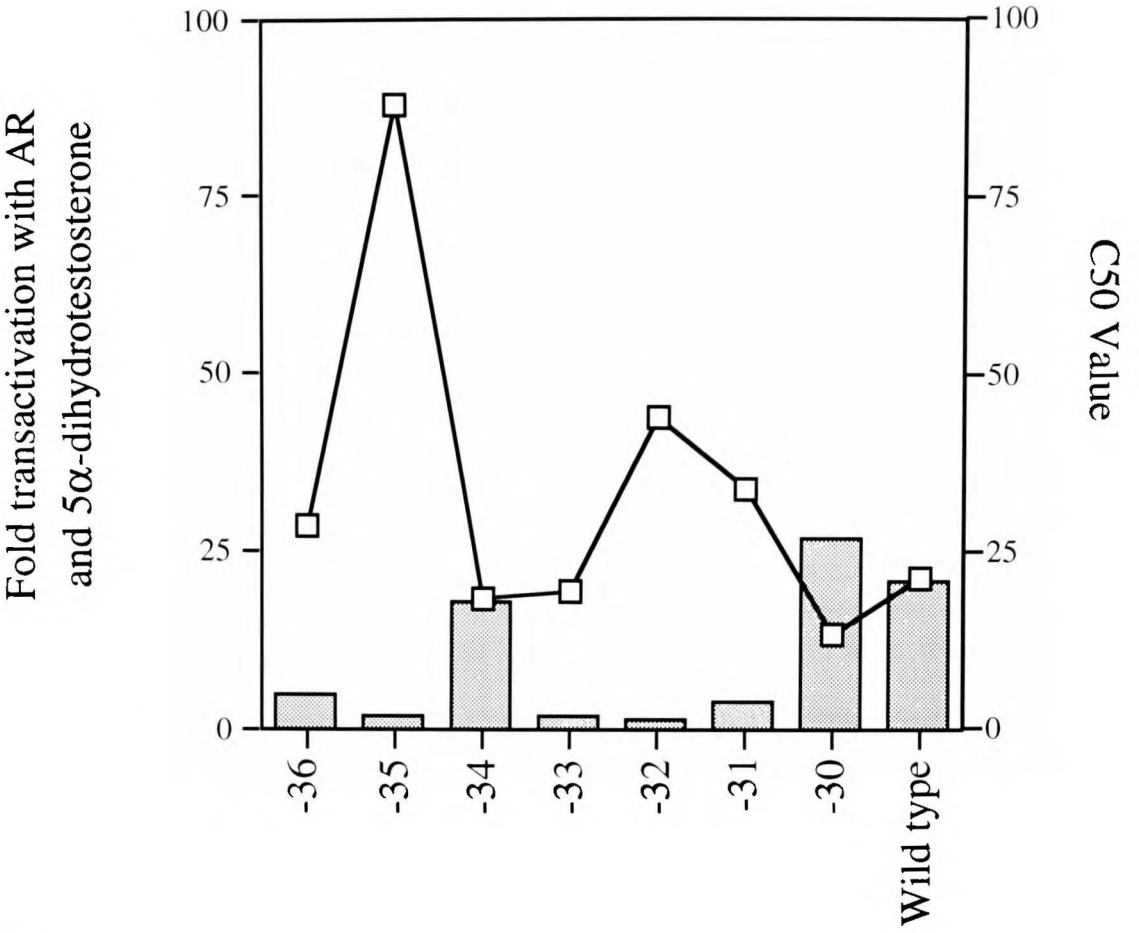
Fig 4.1 shows the consensus sequence obtained from this study in comparison with several well defined natural ARE sequences. A comparison of nine naturally occurring ARE sequences with the derived consensus is shown in Fig 4.1. Close examination of these sequences reveals that in many positions they agree well with the consensus, but none of them adhere to it strictly. It has been noted (Cleutjens et al., 1996) that many of the androgen response elements identified so far diverge from this consensus quite markedly. It is also known that many sequences can co-operate in gene promoters to confer androgen responsiveness (Adler et al., 1993, Ho et al., 1993). The sequences responsible for the androgen responsiveness in these promoters (Fig 1.18) can be multiple weak ARE sequences or ARE half sites, and others can depend upon non-ARE sequences which enhance the action of the ARE.

Given the apparent heterogeneity of the mechanism for androgen receptor action at gene promoters, it is hardly surprising that the elements defined so far do not adhere strictly to the consensus sequence derived in Roche et al., (1992). In the study of Roche et al., (1992), the sequences which were isolated and tested were found to have a wide range of affinities for the protein A-AR DNA binding domain fusion protein. A number of the sequences isolated from the second round of their screening only contained ARE half sites but these were not included in the definition of the consensus. Thus the consensus study could have been biased in the selection of sites which have a higher affinity for AR than those typically found in natural genes.

8.3 Comparison of the Binding Data with the *in vivo* Transfection Data and Patient Data

If the data obtained using the transient transfection assay in Chapter 5 is analysed in parallel with the binding data from Chapter 4, then many mutations are found to have identical effects in each assay (see Table 4.3 and 5.2). Figs 5.11 and 8.1

A



B

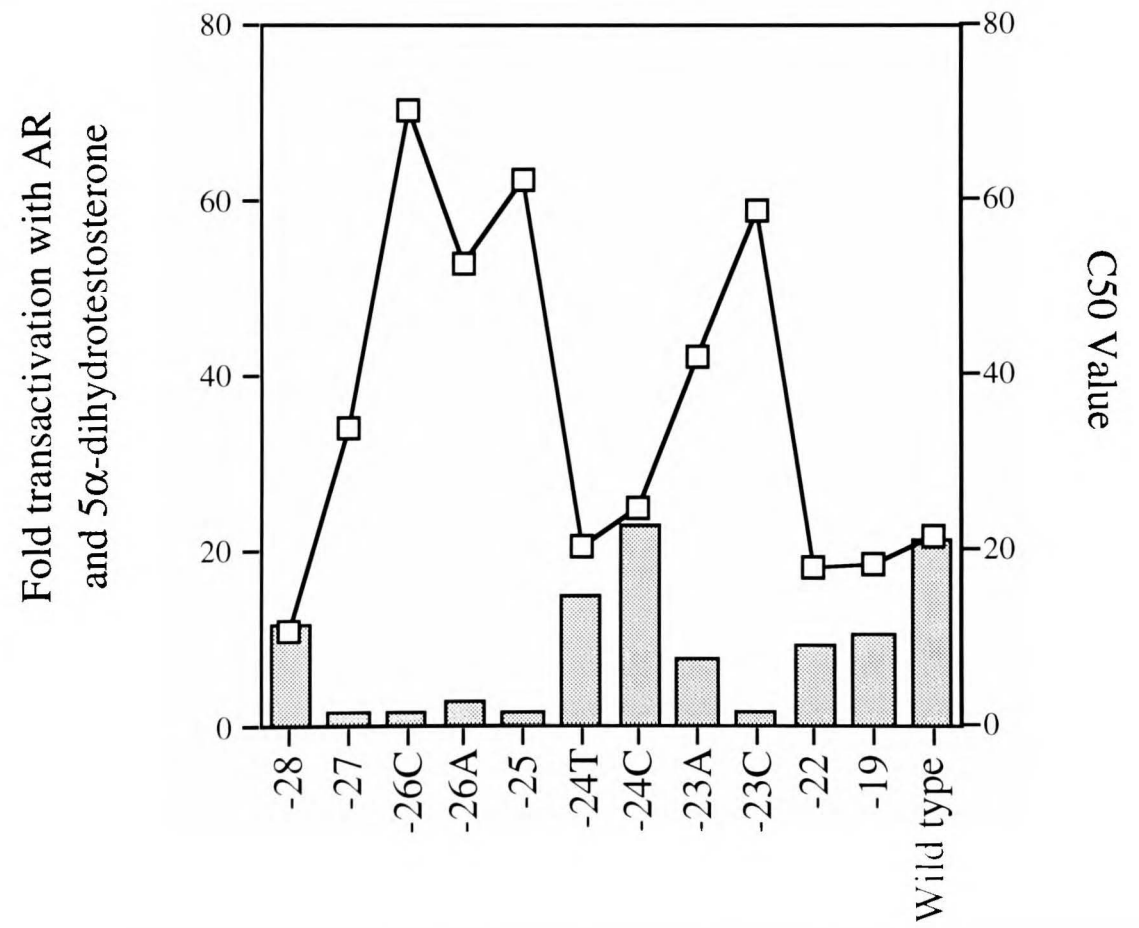


Figure 8.1 A comparison of the C50 results (□) and the transactivation data from the hormone response assays (▩) for each of the factor IX mutant ARE oligonucleotide constructs in (A) the 5' half site and (B) the 3' half site. Standard errors were omitted for clarity and are shown in Table 5.2.

show a comparison of the transfection data with the more *in vitro* competition gel mobility shift binding data for different mutations in the factor IX ARE. The transactivation data in Table 5.2 indicates that the important nucleotide positions in this sequence for the response to AR and 5α-dihydrotestosterone are: -35, -33, -32, -27, -26, -25, and -23 (see Fig 5.11). These nucleotides correspond to positions -6, -4, -3, 2, 3, 4, and 6 (Roche et al., 1992), respectively (see Fig 5.11). These data are in general agreement with both the consensus study of Roche et al., (1992), and with the competition gel mobility shift data in Chapter 4, (see Table 4.3 and Figs 5.11, and 8.1), which identified the positions -6, -3, -2, 2, 3, 4, and 6, as being important for the binding of a protein A-AR DNA binding domain fusion protein to this sequence. The competition gel mobility shift assay did not detect any disruption of the binding of the fusion protein to the sequence carrying a T→C mutation at nucleotide -33 (position -4) (see Table 4.3, and Fig 5.11). However, the transactivation of this sequence by AR and testosterone was greatly reduced by this mutation (see Table 5.2, and Fig 5.11). These results can be reconciled as follows. It is possible that both of these results are correct, and that the protein A-AR DNA binding domain fusion protein may be artifactually recognising some ARE elements which will not function in a cell nucleus in a transfection assay. As the ARE element is composed of two half sites arranged in an imperfect palindrome, it is possible that the gel mobility shift assay is detecting the binding of the fusion protein to the intact half site of the sequence (the 3' half site in this case). This half site is also the better conserved of the two in the ARE. The GR is known to bind to the stronger half site of its elements first, before binding to the second, weaker half site occurs (Tsai et al., 1988) and though the AR is thought to bind to its elements primarily as a dimer (Wong et al., 1993), it is possible that the fusion protein used in the gel mobility shift assays could bind as a monomer as it lacks most of the critical sequences for dimerisation of the AR (see Figs 1.13 and 3.1). Thus, the competition gel mobility shift assays would show the C50 as being essentially unchanged whereas the transfection assay would highlight the decrease in the binding of the full length AR dimer loaded with hormone to this sequence. Though the

differences in the results obtained for the -28 and -19 mutations are not as marked as for the -33 mutation, (see Fig 5.11), similar mechanisms could be responsible. However, the transactivation data for these two mutations were almost identical to the positive control C3 ARE (Claessens et al., 1989) oligonucleotide (see Table 5.2) and so the significance of these reductions in the transactivations may not be as marked as that seen for the -33 mutation.

The mutations which are found to increase the C50 value by less than 100% in the competition gel mobility shift assays are those at positions -7, -2, and C→A at 6 (nucleotides -36, -31, and C→A at -23, see Fig 5.11). These mutations are all weakly active in the transactivation assays (see Table 5.2 and Fig 8.1), but they reduce the level of hormone transactivation to between 19%→44% of the wild type level (see Table 5.2). The transactivations assays in Chapter 5 also show that position 7 (nucleotide -22) reduces the transactivation of this sequence to less than 50% of wild type levels. Therefore, it is probable that the mutations at positions -7, -2, C→A at 6, and possibly 7 represent a group of mutations which partially disrupt the factor IX ARE. Positions -7, -2, 2, and 7 were also identified by Roche et al., (1992) as being "moderately well conserved" in the consensus sequence they defined (see Fig 5.11).

The results at position -26 in the factor IX ARE (position 3) correlate with the patient data available for the G→C and G→A mutations (Giannelli et al., 1996). The transactivation level of the G→A mutation was higher than that of the G→C mutation (see Table 5.2 and Fig 8.1) and the C50 levels of these two mutations also show the same pattern (see Table 4.3 and Fig 8.1). The factor IX plasma levels in these two patients were <1% for the G→C, haemophilia B Brandenburg, (Crossley et al., 1992), and 3% for the G→A, patient UK232 (Giannelli et al., 1996). The factor IX plasma level in patient UK232 (3%) is not as low as the Brandenburg levels (<1%), possibly because there is some residual activity of the factor IX ARE in this patient, allowing a low level of transcription of the factor IX gene after puberty, or alternatively, because the G→A mutation at -26 does not disrupt the HNF4 site to the same degree as the G→C mutation, also allowing a low level of transcription. The precise factor IX

clotting levels for this patient before puberty were not available (Giannelli, personal communication) and so that these hypotheses cannot be distinguished.

The major differences in the data obtained from the gel mobility shift assay and the transfection data are that:

(i) the mutation at -33 (position -4, see Fig 5.11) disrupts transactivation in the transfection assay but does not alter the C50 value of the ARE in the gel mobility shift assay (see Tables 4.3 and 5.2),

(ii) The mutations at -27 and -22 (positions 2 and 7, see Fig 5.11) disrupt the transactivation in the transfection assay to a greater extent than the C50 level is increased in the gel mobility shift assays (see Tables 4.3 and 5.2, and Fig 8.1)

The differences seem to suggest that the gel mobility shift assay is less sensitive than the transactivation assay in detecting disruptive mutations. Moreover, the data obtained from the transfection assays are a better fit to the ARE consensus than the data from the gel mobility shift assays (see Fig 5.11).

Smaller differences were seen in the two sets of data for the mutations in the flanking residue, -19, and the spacer, -28 (positions 1, and 10, see Fig 5.11). These mutations lower the C50 value for the -40→-9 sequence (by 14% and 50% respectively, see Table 4.3), but they also reduce the level of hormone transactivation of the oligonucleotide tetramer reporter constructs by 51% and 46%, respectively (see Table 5.2). Table 5.2 also shows that the C3 ARE, which is used as a positive control in this study, only transactivates the tk promoter to approximately 50% of the wild type factor IX ARE level. Therefore, the reduction in transactivation seen for the mutations at -28 and -19 may not be significant. However, the prediction (Table 4.2) that the mutation at -28 would improve the factor IX ARE is clearly not correct (Table 4.3).

8.4 Further Evidence for the Importance of an ARE in the Factor IX Promoter in the Leyden Recovery.

The data presented in this study suggest that an ARE in the factor IX promoter region spanning nucleotides -36→-22 is important for the recovery of factor IX levels

during puberty seen in the Leyden phenotype. Sections 6.1 and 6.2 describe two patients, haemophilia B Brandenburg (Crossley et al., 1992) and UK232 (Giannelli et al., 1996) who have mutations at nucleotide -26 in the promoter region but fail to show the classical Leyden phenotype (see Fig 1.20). The mutation associated with patient UK232 was a -26 G→A mutation. This mutation also disrupted the HNF4 binding site and the ARE (see section 6.2) and provides further evidence that mutations at this position which prevent any Leyden type recovery of factor IX levels in patients do so because they disrupt both of these sites. Patient UK232 is now 47 years old and still has a bleeding disorder, confirming that the pubertal rise in plasma factor IX levels associated with the Leyden phenotype (see Fig 1.20) has not occurred. Haemophilia B Brandenburg was only 21 years old when he died (see Fig 1.20) preventing any study of his bleeding disorder in adulthood.

The action of AR on the ARE has been reported not to involve direct DNA binding (Kurachi et al., 1994, see Fig 1.24), but work presented in this thesis using an *in vitro* gel mobility shift assay contradicts this. Further, Chapter 5 shows that AR and testosterone are able to transactivate reporter constructs which have the element tetramerised upstream of them *in vivo* in transient transfection assays in HeLa cells, confirming the earlier data of Crossley et al., (1992) from this laboratory. These data argue strongly for a direct role for AR and testosterone in the activation of factor IX gene transcription during the recovery seen in the Leyden phenotype, though a possible additional indirect role for AR cannot be dismissed.

8.5 Testing the Theory for the Direct Action of AR and Testosterone in the Leyden Recovery

Chapter 6 describes the functional characterisation of the molecular defect in 2 patients with point mutation at nucleotides -23 and -19 in the factor IX promoter region. Patient HB 396 was found to have a G→C mutation at nucleotide -19, and patient HB 388 had a C→T mutation at nucleotide -23 (Ketterling et al., 1995). Both patients are still children (Ketterling et al., 1995) and this allows the ARE-mediated recovery

theory for the Leyden recovery (Crossley et al., 1992, see 1.10.3 and Fig 1.21) to be tested. If the ARE theory is correct, then it is possible to predict the adult phenotypes for HB 396 and HB 388, based on previous results (Crossley et al., 1992, Reitsma et al., 1988, Reijnen et al., 1993) and those reported in Chapter 6.

The predictions are as follows:

(i) as the -19G→C mutation does not disrupt the ARE in the factor IX promoter, the plasma factor IX clotting activity in patient HB 396 will rise throughout puberty and his factor IX clotting levels should stabilise at between 30-60% of normal in adulthood. It is predicted that this patient will recover from his bleeding disorder and show a Leyden phenotype (see Fig 1.20).

(ii) patient HB 388 will not show any increase in blood plasma factor IX through puberty, and will not recover from his bleeding disorder due to the disruption of the ARE by the -23C→T mutation. This patient will show a similar phenotype to haemophilia B Brandenburg (Crossley et al., 1992), and UK232 (Giannelli et al., 1996).

8.6 Consequences for Genetic Counselling and Treatment

The ability to predict whether or not a patient will recover from a bleeding disorder during the course of puberty is particularly useful. The prediction may have vital implications for the genetic counselling of patients in affected pedigrees as well as having important implications in the therapy of patients with this type of condition. Clearly, if a bleeding disorder shows a phenotype which improves through puberty and results in a near total recovery of clotting levels in adulthood then this condition would, quite rightly, be regarded as far less serious than a lifelong bleeding disorder. The prophylactic treatment of factor IX deficiency (see section 1.1.2) is intended to prevent serious bleeding episodes into joints and prevent long-term joint damage. The use of prophylactic treatment of haemophilia B is not generally practised in the US (Furie et al., 1994), due to the expense, but if it were known that any prophylactic treatment

would only be needed until puberty, when the factor IX levels would start to rise, then this form of treatment would be far more justified.

8.7 Analysis of the Overlap Region Between the ARE and the HNF4 Binding Site

As can be deduced from the prediction made in 8.4, the hypothesis of Crossley et al., (1992) states that a mutation disrupts the HNF4 site without disrupting the ARE, then that mutant promoter region will be upregulated through puberty by the direct action of AR and testosterone via the ARE. However, if a mutation is found in the overlapping region of the ARE and the HNF4 site (-27→-22), then the AR and testosterone-mediated upregulation cannot occur. The region of overlap (-27→-22) between the ARE and the HNF4 site has four positions which disrupt the binding of both AR and HNF4 (see Fig 8.2). Patients with mutations at nucleotides -26 do not to show any recovery in their factor IX levels during puberty (Crossley et al., 1992, Giannelli et al., 1996). The patient with a mutation at nucleotide -23 is not expected to recover from his bleeding disorder through puberty (section 8.4, Ketterling et al., 1995). If patients are identified with mutations at positions -27 and -25 then they too, would not be expected to recover during puberty, though the phenotype of a patient with a mutation at -27 might be expected to be less severe than the haemophilia B Brandenburg/UK232 phenotype (Figs 8.1 and 8.2). Patients with mutations at nucleotides -24 or -22 would be expected to show a Leyden phenotype and recover from their bleeding disorder in adulthood despite the fact that these mutations lie inside the ARE (Fig 8.2). The analysis of this region, and the opportunity to test the hypothesis of Crossley et al., (1992) by predicting the phenotype of patients is unique to the haemophilia B Leyden condition and the implications of these data for the counselling and treatment of the bleeding disorder of patients are discussed in 8.5.



Figure 8.2 The region of overlap in the factor IX promoter region (-27→-22) between the ARE and the HNF4 site. The mutagenesis study of this region identified the mutations at the red positions as disrupting the binding of both AR and HNF4, and have been identified in haemophilia B patients (Giannelli et al., 1996). Mutations at the other positions have not yet been identified in patients. Potentially, any patient with a mutation at a green position will have an intact ARE and will recover from their haemophilia, but any patient with a mutation at a yellow position will not recover due to its disruption (see Fig 8.1 for details of the effects of the mutations).

8.8 Possible Mechanisms for the Recovery of Patients with Mutations Outside the HNF4 Site Spanning Nucleotides -27→-15

The theory of an active ARE element being directly involved in the recovery of the levels of factor IX transcription during puberty in Leyden type patients (see Fig 1.21) is able to explain the Leyden recovery in patients who have mutations in the HNF4 site (-27→-15) but which leave the ARE (-36→-22) intact. This theory does not explain how the recovery of factor IX expression occurs in patients with mutations outside this HNF4 site. An alternative theory of the action of the AR on this promoter is therefore needed in the case where the HNF4 element (-27→-15) is intact and binds HNF4, but the mutation lies in a different *cis* element e.g. the +13 mutation in the C/EBP site which also results in haemophilia B Leyden (Crossley and Brownlee, 1990). The Leyden phenotype of such patients means that factor IX promoter regions with mutations at +13 must be able to respond to AR and testosterone through puberty in some way. It is possible that AR and HNF4 may still be able to bind to their overlapping sites (-36→-15) simultaneously, though no evidence for this has been presented to date. Alternatively, these two transcription factors compete for binding to these overlapping sites, and that the final occupancy of the site will depend upon the relative activities and concentrations of the two transcription factors relative to one another. The other theories of the Leyden recovery (1.10.4→1.10.6) will now be discussed in terms of their potential to explain the promoter function for all Leyden mutations.

As the HNF4 site and the ARE overlap in the region -27→-22, it is supposed that these factors should not be able to bind simultaneously but that their binding is mutually exclusive. Reijnen et al., (1993) proposed the theory that as the AR and HNF4 are both members of the nuclear receptor superfamily, then it would be possible for them to interact in some way, possibly by heterodimerising (see Fig 1.22). In support of such a hypothesis, Liu et al., (1995) presented evidence that it was possible for the DNA binding domains of the glucocorticoid receptor and the mineralocorticoid receptor to heterodimerise, and suggested that this might be a general feature of the

other steroid hormone receptors. However, studies with the HNF4 molecule (Jiang et al., 1995) have not identified any heterodimerisation potential with other members of the nuclear receptor superfamily, and have suggested that the homodimerisation potential of HNF4 is so efficient that the possibility of it forming heterodimers with another molecule is small.

Kurachi et al., (1994) presented preliminary data that suggested that the AR could act indirectly on the factor IX promoter (see Fig 1.24). This mechanism could provide a way that the AR can activate transcription without having to bind to the ARE. This would provide an explanation for the ability of AR to activate transcription from a promoter in which the ARE was partially blocked by the binding of HNF4 to an overlapping site. However, until the protein(s) with which the AR interacts are identified, it is difficult to test this theory critically.

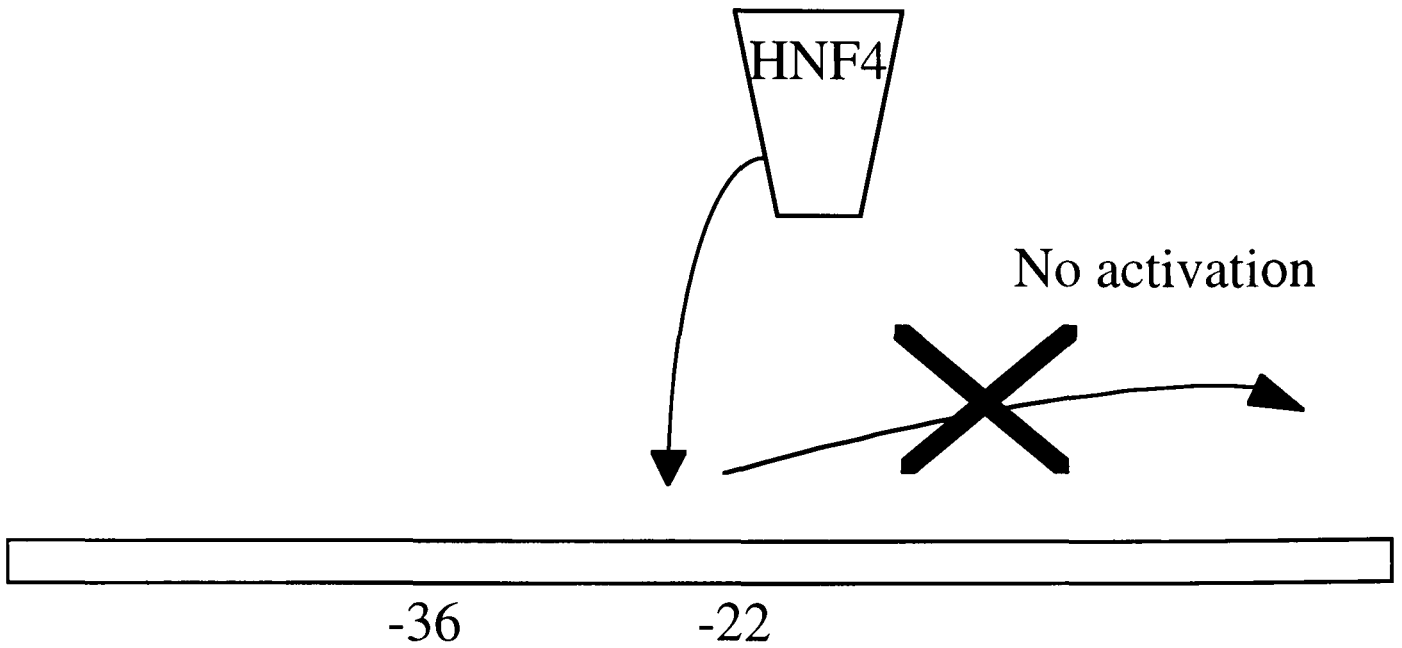
Boccia et al., (1996) presented a mechanism (see Fig 1.23) which involves the binding of the proteins C/EBPα and GABPα/β to a site in the promoter region spanning nucleotides -220→-195. This complex is then proposed to interact, via protein-protein interactions, with a heterodimer of the proteins DBP (Mueller et al., 1990) and HLF (Hunger et al., 1992). DBP is a developmentally regulated protein in the rat (Mueller et al., 1990) and its levels rise during puberty and reach a maximum in adulthood. Therefore, it is proposed, that the complex of proteins interacting around the -220→-195 site is not complete until the level of DBP rises at adulthood, and it is only then that this complex is capable of activating expression of the factor IX gene. This is an attractive mechanism, which explains the pubertal rise in transcription of the factor IX gene via the developmentally regulated protein DBP. However, this theory does not explain the absence of any recovery in factor IX levels in patients with mutations at nucleotide -26 in the promoter region (Crossley et al., 1992). A transgenic mouse constructed with the CAT reporter gene driven by the factor IX minimal promoter (-189→+22) (Boland et al., 1995) showed a sex-specific, age-dependent rise in transcription of the reporter gene. This expression mimicked the Leyden recovery well,

but lacked the site -220→-195. This implies that the action of the complex proposed by Boccia et al., (1996) cannot be essential, though it may still play an additional role.

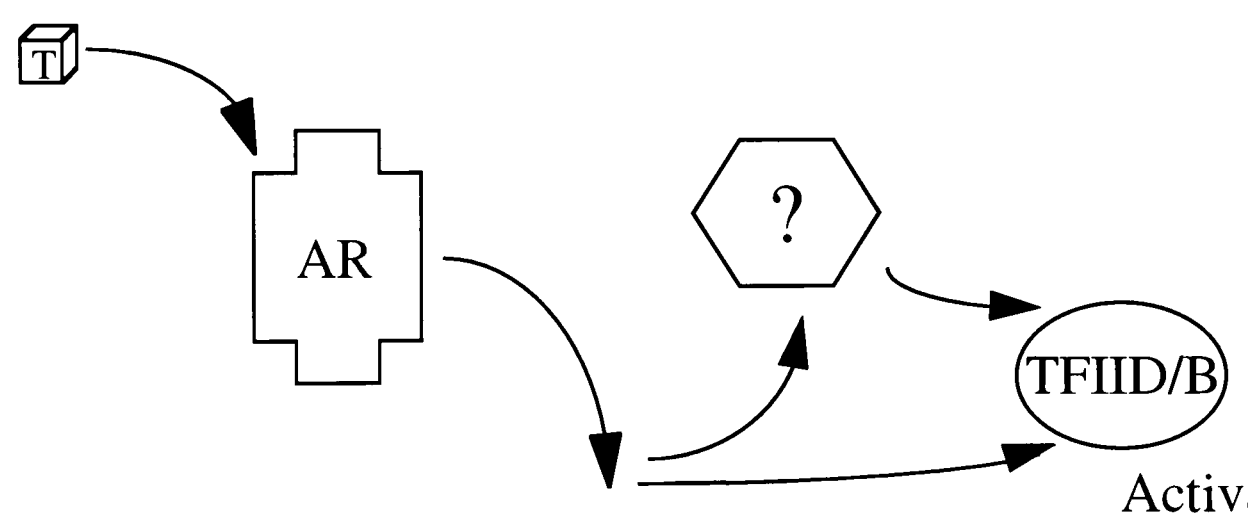
A speculative theory to explain the action of the AR on the factor IX promoter during the Leyden recovery is different assembly pathways exist for the formation of the PIC at the initiator of the promoter. The PIC will only assemble and transcribe the factor IX gene, provided that the correct combination of transcription activators are assembled on the promoter. One assembly pathway might involve the normal factor IX promoter and the action of C/EBP, HNF4, NF-1, ARP1, COUP and GABPα/β in stimulating PIC formation (see Fig 8.3). In the Leyden gene promoter, one of these transcription activators is unable to interact with its mutated *cis* element, so the level of factor IX expression is impaired in childhood. Then, at puberty, the levels of testosterone rise, the AR becomes activated, and it interacts with the factor IX promoter either directly through the ARE, or indirectly through protein-protein interactions with another transcription activator protein, providing at least two additional alternative pathways to assemble the PIC (see Fig 8.3). Either of these pathways allows the assembly of a PIC at the initiator, which then raises the level of transcription of the gene, and allows the recovery from the bleeding disorder through puberty. In the case of Leyden mutations outside the HNF4 site at -27→-15, these mechanisms suggest that the AR somehow activates transcription from a Leyden promoter through puberty where HNF4 cannot before puberty. This could possibly be achieved through the interaction of the AR homodimer with coactivator and integrator molecules which enhance the recruitment of the TFIID onto the promoter. To critically examine these hypotheses, the repertoire of proteins with which both HNF4 and AR can interact with needs to be characterised. When this goal is achieved, then detailed mechanisms for this unique phenotype can be formulated and tested.

Figure 8.3 Possible mechanisms for the Leyden recovery of patients with mutations outside the HNF4 site (-27→-15).(A) Before puberty, HNF4 binds to the site spanning -27→-15, but transcription does not occur due to mutations in other *cis* elements. (B) At puberty, AR is activated by testosterone (T) and this protein will compete with HNF4 for binding to its binding site. If AR can successfully bind then it may be able to compensate for the loss of one of the other transcription factor binding sites, maybe by direct recruitment of TFIID, or indirectly via a coactivator. (C) A second possible mechanism involves the cooperation of HNF4 and AR at the -36→-22 region of the promoter. This cooperation may take the form of either a direct interaction or and indirect one, via a coactivator or integrator molecule. Either of these mechanisms, or a combination of the two, may function in different Leyden patient promoter regions.

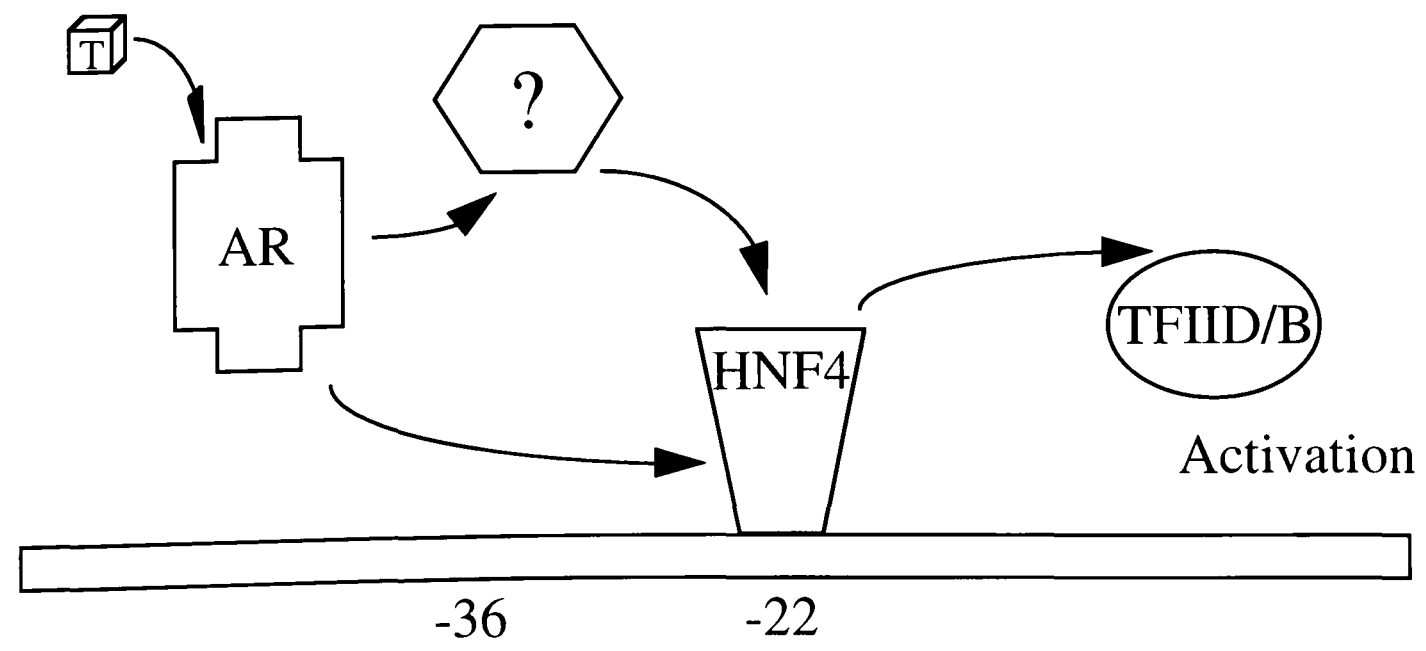
A


HNF4
No activation
-36
-22

B


T
AR
?
TFIID/B
Activation
-36
-22

C


T
AR
?
HNF4
TFIID/B
Activation
-36
-22

8.9 Further Work

8.9.1 Improvements to the Binding Study

The binding data with the AR and the factor IX ARE could be improved by the use of recombinant, full length AR protein in the gel mobility shift assays. The use of such a protein sample would ensure that AR non-DNA binding domain-ARE interactions could occur, and in particular allowing homodimer formation by the inclusion of the ligand binding domain (see Fig 1.13) in the protein. This would increase the physiological relevance of the binding data. Alternatively, the technique of surface plasmon resonance (Panayotou et al., 1993) could be utilised in the analysis of binding reactions with the factor IX ARE. This technique allows one to examine real time binding reactions, and gives data which makes it possible to define precise binding and dissociation constants for reactions. So far this technique has only been widely applied to the field of protein-protein interactions (Schuster et al., 1993), but its potential for studying DNA-protein interactions has long been realised (Jost et al., 1991). This technique would allow the characterisation of the interactions of the protein A-AR DNA binding domain fusion protein with the factor IX ARE. It is suspected (see Fig 4.4) that multiple complexes are formed in the binding reaction; this technique could provide more data to support or refute that possibility. This technique would also allow the examination of the binding of AR and HNF4 to their overlapping binding sites (-36→-22, and -27→-15, respectively). This type of analysis may prove whether or not the two proteins could bind simultaneously or mutually exclusively. This type of data is vital to the elucidation of the exact molecular mechanism of the Leyden phenotype.

8.9.2 Characterisation of AR Action

As mentioned in section 8.8, only when the full repertoire of interactions is known will a complete understanding of the mechanism of the Leyden recovery be possible. To this end, the possible proteins with which the AR interacts should be fully characterised. Probably the best method to identify the full range of proteins with which

this protein can interact is to use the AR molecule as bait in a yeast two-hybrid assay to screen human cDNA libraries (Fields and Song, 1989). This technique would allow the definition of interesting candidate proteins with which the AR could interact. It is usual to then confirm such candidates with GST "pulldown" experiments as described in Kaelin et al. (1991).These proteins would be expected to include components of the PIC, other transcription activators and coactivators or integrator molecules (see 1.11). When these proteins have been identified, their interaction may be tested on the factor IX promoter in standard gel mobility shift, biosensor, transient transfection assays, and *in vitro* transcription assays.

8.9.3 Further *in vivo* Data

The creation of the factor IX Leyden transgenic mouse (Boland et al., 1995) was the first step in understanding the true molecular mechanism of the Leyden recovery. This study proved that the factor IX Leyden promoter could direct the tissue-specific, sex-specific, age-related rise in the expression of a reporter gene, though the rise in transgene expression was late on in puberty in this mouse. Further studies which could assist in the elucidation of this role include the castration of male Leyden transgenic mice before they enter puberty. This procedure would remove the source of androgen from the mouse and so by following the reporter expression is this mouse, and a castrated mouse given an external source of androgens, the response of the Leyden promoter to AR and testosterone could be examined.

The factor IX Leyden transgenic mouse was made with the promoter region -189→+21 which carried a mutation at +13 known to disrupt the binding of C/EBP to the promoter (Crossley and Brownlee, 1990). To test the role of AR in this recovery, another transgenic mouse carrying both the +13 mutation and a mutation in the ARE which was known to disrupt the binding of AR to the promoter e.g. a mutation at nucleotide -32 or -35 (see Figs 8.1) should be constructed. The factor IX levels in this mouse could be followed through the lifetime of the mouse and directly compared to those of the original mouse with the single +13 mutation. This experiment would

confirm whether the AR acted directly through binding to the ARE, or whether it functioned in an indirect way with no requirement for DNA binding. This experiment would also show whether AR can bind to the ARE element when the HNF4 site (-27→-15) was intact.

8.10 Concluding Remarks

The recovery of patient HB 396 has been predicted in section 8.5. The rise in his factor IX plasma levels through puberty would be strong evidence in favour of the ARE-mediated mechanism for the recovery of the Leyden type patients. The ability to predict the phenotype of any patient with a promoter mutation will allow this theory to be tested with any new patient with a promoter mutation in the future. Relating the *in vitro* data presented in this thesis with *in vivo* data obtained from patient studies will strengthen, or weaken, the case for the involvement of the ARE in the recovery of Leyden type patients. Therefore, it is only through a combination of the complete characterisation of the exact molecular interactions of the proteins assembled on the factor IX promoter, coupled with true *in vivo* data, such as the transgenic mouse, or further patient data, that the mechanisms presented and discussed in this thesis will be confirmed or refuted.

References

A

Adler, A. J., Scheller, A., and Robins, D. M. (1993). The stringency and magnitude of androgen-specific gene activation are combinatorial functions of receptor and nonreceptor binding site sequences. *Mol Cell Biol,* **13**(10), 6326-35.

Aggeler, P.M., White, S.G., Glendening, M.G., Page, E.W., Leake, T.B., Bates, G. (1952). Plasma thromboplastin component (PTC) deficiency: A new disease resembling haemophilia. *Proc Soc Exp Biol Med*, **79**, 692-694.

Ahmad, S. S., Rawala-Sheikh, R., and Walsh, P. N. (1989). Comparative interactions of factor IX and factor IXa with human platelets. *J Biol Chem*, **264**(6), 3244-51.

Anson, D. S., Choo, K. H., Rees, D. J., Giannelli, F., Gould, K., Huddleston, J. A., and Brownlee, G. G. (1984). The gene structure of human anti-haemophilic factor IX. *EMBO J*, **3**(5), 1053-60.

Arriza, J. L., Weinberger, C., Cerelli, G., Glaser, T. M., Handelin, B. L., Housman, D. E., and Evans, R. M. (1987). Cloning of human mineralocorticoid receptor complementary DNA: structural and functional kinship with the glucocorticoid receptor. *Science*, **237**(4812), 268-75.

Austen, D.E.G., and Rhymes, I.L. (1975). A Laboratory Manual of Blood Coagulation. *Blackwells Scientific Publications, Oxford.*

B

Bajaj, S. P., Sabharwal, A. K., Gorka, J., and Birktoft, J. J. (1992). Antibody-probed conformational transitions in the protease domain of human factor IX upon calcium binding and zymogen activation: putative high-affinity $Ca(^{2+})$-binding site in the protease domain. *Proc Natl Acad Sci U S A*, **89**(1), 152-6.

Banic, S. (1975). Prevention of rabies by vitamin C. *Nature*, **258**(5531), 153-4.

Beato, M. (1989). Gene regulation by steroid hormones. *Cell*, **56**(3), 335-44.

Beato, M., Herrlich, P., and Schutz, G. (1995). Steroid hormone receptors: many actors in search of a plot. *Cell*, **83**(6), 851-7.

Beckman Instruments Liquid Scintillation Manual (1988). Section T-1648-NUC-88-4.

Beddington, R. S., Morgernstern, J., Land, H., and Hogan, A. (1989). An in situ transgenic enzyme marker for the midgestation mouse embryo and the visualization of inner cell mass clones during early organogenesis. *Development*, **106**(1), 37-46.

Beitel, L. K., Prior, L., Vasiliou, D. M., Gottlieb, B., Kaufman, M., Lumbroso, R., Alvarado, C., McGillivray, B., Trifiro, M., and Pinsky, L. (1994). Complete androgen insensitivity due to mutations in the probable alpha-helical segments of the DNA-binding domain in the human androgen receptor. *Hum Mol Genet*, **3**(1), 21 7.

Berg, L. P., Scopes, D. A., Alhaq, A., Kakkar, V. V., and Cooper, D. N. (1994). Disruption of a binding site for hepatocyte nuclear factor 1 in the protein C gene promoter is associated with hereditary thrombophilia. *Hum Mol Genet*, **3**(12), 2147-52.

Biggs, R., Douglas, A.S., MacFarlane, R.G., Dacie, J.V., Pitney, W.R., Merskey, C., O'Brien, J.R. (1952) Christmas disease: A condition previously mistaken for haemophilia. *Brit Med J*, **2**, 1378.

Blau, J., Xiao, H., McCracken, S., O'Hare, P., Greenblatt, J., and Bentley, D. (1996). Three functional classes of transcriptional activation domain. *Mol Cell Biol*, **16**(5), 2044-55.

Boccia, L. M., Lillicrap, D., Newcombe, K., and Mueller, C. R. (1996). Binding of the Ets factor GA-binding protein to an upstream site in the factor IX promoter is a critical event in transactivation. *Mol Cell Biol*, **16**(5), 1929-35.

Boland, E. J., Liu, Y. C., Walter, C. A., Herbert, D. C., Weaker, F. J., Odom, M. W., and Jagadeeswaran, P. (1995). Age-specific regulation of clotting factor IX gene expression in normal and transgenic mice. *Blood*, **86**(6), 2198-205.

Bristol, J. A., Freedman, S. J., Furie, B. C., and Furie, B. (1994). Profactor IX: the propeptide inhibits binding to membrane surfaces and activation by factor XIa. *Biochemistry*, **33**(47), 14136-43.

Brown, C. J., Goss, S. J., Lubahn, D. B., Joseph, D. R., Wilson, E. M., French, F. S., and Willard, H. F. (1989). Androgen receptor locus on the human X chromosome: regional localization to Xq11-12 and description of a DNA polymorphism. *Am J Hum Genet*, **44**(2), 264-9.

Brownlee, G. G. (1995). Prospects for gene therapy of haemophilia A and B. *Br Med Bull*, **51**(1), 91-105.

Buone, G., Briand, P-A., Miksicek, R.J., and Picard, D. (1996) Activation of the unliganded estrogen receptor by EGF involves the MAP kinase pathway and direct phosphorylation. *EMBO J*, **15**(9), 2174-83.

Buratowski, S. (1994). The basics of basal transcription by RNA polymerase II. *Cell*, **77**(1), 1-3.

Burley, S.K. (1996). Picking up the TAB. *Nature*, **381**(6578), 112-3.

C

Carlson, M., and Laurent, B. C. (1994). The SNF/SWI family of global transcriptional activators. *Curr Opin Cell Biol*, **6**(3), 396-402.

Carlsson, P., Waterman, M. L., and Jones, K. A. (1993). The hLEF/TCF-1 alpha HMG protein contains a context-dependent transcriptional activation domain that induces the TCR alpha enhancer in T cells. *Genes Dev*, **7**(12A), 2418-30.

Carr, F. E., and Wong, N. C. (1994). Characteristics of a negative thyroid hormone response element. *J Biol Chem*, **269**(6), 4175-9.
Carthew, R. W., Chodosh, L. A., and Sharp, P. A. (1985). An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. *Cell*, **43**(2 Pt 1), 439-48.
Cavailles, V., Dauvois, S., L'Horset, F., Lopez, G., Hoare, S., Kushner, P. J., and Parker, M. G. (1995). Nuclear factor RIP140 modulates transcriptional activation by the estrogen receptor. *EMBO J*, **14**(15), 3741-51.
Celis, L., Claessens, F., Peeters, B., Heyns, W., Verhoeven, G., and Rombauts, W. (1993). Proteins interacting with an androgen-responsive unit in the C3(1) gene intron. *Mol Cell Endocrinol*, **94**(2), 165-72.
Chamberlain, N. L., Driver, E. D., and Miesfeld, R. L. (1994). The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Res*, **22**(15), 3181-6.
Chance, P. F., Dyer, K. A., Kurachi, K., Yoshitake, S., Ropers, H. H., Wieacker, P., and Gartler, S. M. (1983). Regional localization of the human factor IX gene by molecular hybridization. *Hum Genet*, **65**(2), 207-8.
Chang, C., Wang, C., DeLuca, H. F., Ross, T. K., and Shih, C. C. (1992). Characterization of human androgen receptor overexpressed in the baculovirus system. *Proc Natl Acad Sci U S A*, **89**(13), 5946-50.
Chang, C. S., Kokontis, J., and Liao, S. T. (1988a). Structural analysis of complementary DNA and amino acid sequences of human and rat androgen receptors. *Proc Natl Acad Sci U S A*, **85**(19), 7211-5.
Chang, C. S., Kokontis, J., and Liao, S. T. (1988b). Molecular cloning of human and rat complementary DNA encoding androgen receptors. *Science*, **240**(4850), 324-6.
Chen, J. D., and Evans, R. M. (1995). A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature*, **377**(6548), 454-7.
Chen, S. H., Zhang, M., Lovrien, E. W., Scott, C. R., and Thompson, A. R. (1991). CG dinucleotide transitions in the factor IX gene account for about half of the point mutations in hemophilia B patients: a Seattle series. *Hum Genet*, **87**(2), 177-82.
Chodosh, L. A., Carthew, R. W., and Sharp, P. A. (1986). A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. *Mol Cell Biol*, **6**(12), 4723-33.
Choo, Y., and Klug, A. (1994). Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc Natl Acad Sci U S A*, **91**(23), 11168-72.
Choo, Y., Sanchez-Garcia, I., and Klug, A. (1994). *In vivo* repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature*, **372**(6507), 642-5.

Claessens, F., Celis, L., Peeters, B., Heyns, W., Verhoeven, G., and Rombauts, W. (1989). Functional characterization of an androgen response element in the first intron of the C3(1) gene of prostatic binding protein. *Biochem Biophys Res Commun*, **164**(2), 833-40.

Clay, C. M., Keri, R. A., Finicle, A. B., Heckert, L. L., Hamernik, D. L., Marschke, K. M., Wilson, E. M., French, F. S., and Nilson, J. H. (1993). Transcriptional repression of the glycoprotein hormone alpha subunit gene by androgen may involve direct binding of androgen receptor to the proximal promoter. *J Biol Chem*, **268**(18), 13556-64.

Cleutjens, K. B., van-Eekelen, C. C., van-der-Korput, H. A., Brinkman, A. O., and Trapman, J. (1996). Two androgen response regions cooperate in steroid hormone regulated activity of the prostate-specfic antigen promoter. *J Biol Chem*, **271**(11), 6379-88.

Cook, R. F., Cook, S. J., Savon, S., McGrane, M., Hartitz, M., Hanson, R. W., and Hodgson, C. P. (1993). Liver-specific expression of a phosphoenolpyruvate carboxykinase-neo gene in genetically modified chickens. *Poult Sci*, **72**(3), 554-67.

Coulombe, B., Li, J., and Greenblatt, J. (1994). Topological localization of the human transcription factors IIA, IIB, TATA box-binding protein, and RNA polymerase II-associated protein 30 on a class II promoter. *J Biol Chem*, **269**(31), 19962-7.

Cress, W. D., and Triezenberg, S. J. (1991). Critical structural elements of the VP16 transcriptional activation domain. *Science*, **251**(4989), 87-90.

Crossley, M. (1990). The expression of clotting factor IX. *D.Phil thesis, University of Oxford.*

Crossley, M., and Brownlee, G. G. (1990). Disruption of a C/EBP binding site in the factor IX promoter is associated with haemophilia B. *Nature*, **345**(6274), 444-6.

Crossley, M., Winship, P. R., Austen, D. E., Rizza, C. R., and Brownlee, G. G. (1990). A less severe form of Haemophilia B Leyden. *Nucleic Acids Res*, **18**(15), 4633.

Crossley, M., Ludwig, M., Stowell, K. M., De-Vos, P., Olek, K., and Brownlee, G. G. (1992). Recovery from hemophilia B Leyden: an androgen-responsive element in the factor IX promoter. *Science*, **257**(5068), 377-9.

D

Danielian, P. S., White, R., Lees, J. A., and Parker, M. G. (1992). Identification of a conserved region required for hormone dependent transcriptional activation by steroid hormone receptors. *EMBO J*, **11**(3), 1025-33.

Davie, E. W. (1995). Biochemical and molecular aspects of the coagulation cascade. *Thromb Haemost*, **74**(1), 1-6.

De-Vos, P., Claessens, F., Winderickx, J., Van-Dijck, P., Celis, L., Peeters, B., Rombauts, W., Heyns, W., and Verhoeven, G. (1991). Interaction of androgen response elements with the DNA-binding domain of the rat androgen receptor expressed in Escherichia coli. *J Biol Chem*, **266**(6), 3439-43.

De-Vos, P., Schmitt, J., Verhoeven, G., and Stunnenberg, H. G. (1994). Human androgen receptor expressed in HeLa cells activates transcription *in vitro*. *Nucleic Acids Res*, **22**(7), 1161-6.

DeFranco, D. B., Qi, M., Borror, K. C., Garabedian, M. J., and Brautigan, D. L. (1991). Protein phosphatase types 1 and/or 2A regulate nucleocytoplasmic shuttling of glucocorticoid receptors. *Mol Endocrinol*, **5**(9), 1215-28.

Di-Scipio, R. G., Hermodson, M. A., Yates, S. G., and Davie, E. W. (1977). A comparison of human prothrombin, factor IX (Christmas factor), factor X (Stuart factor), and protein S. *Biochemistry*, **16**(4), 698-706.

Diamond, M. I., Miner, J. N., Yoshinaga, S. K., and Yamamoto, K. R. (1990). Transcription factor interactions: selectors of positive or negative regulation from a single DNA element. *Science*, **249**(4974), 1266-72.

Dikstein, R., Ruppert, S., and Tjian, R. (1996). TAFII250 is a bipartite protein kinase that phosphorylates the base transcription factor RAP74. *Cell*, **84**(5), 781-90.

Drouin, J., Sun, Y. L., Chamberland, M., Gauthier, Y., De-Lean, A., Nemer, M., and Schmidt, T. J. (1993). Novel glucocorticoid receptor complex with DNA element of the hormone-repressed POMC gene. *EMBO J*, **12**(1), 145-56.

Du, W., Thanos, D., and Maniatis, T. (1993). Mechanisms of transcriptional synergism between distinct virus-inducible enhancer elements. *Cell*, **74**(5), 887-98.

E

Evatt, B. L. (1995). AIDS and hemophilia--current issues. *Thromb Haemost*, **74**(1), 36-9.

F

Faber, P. W., van-Rooij, H. C., van-der-Korput, H. A., Baarends, W. M., Brinkmann, A. O., Grootegoed, J. A., and Trapman, J. (1991). Characterization of the human androgen receptor transcription unit. *J Biol Chem*, **266**(17), 10743-9.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature*, **340**(6230), 245-6.

Figueiredo, M. S., Dos-Santos, J. E., Alberto, F. L., and Zago, M. A. (1992). High frequency of the Lebanese allele of the LDLr gene among Brazilian patients with familial hypercholesterolaemia. *J Med Genet*, **29**(11), 813-5.

Figueiredo, M. S., Bowen, D. J., Silva-Junior, W. A., and Zago, M. A. (1994). Factor IX gene haplotypes in Brazilian blacks and characterization of unusual DdeI alleles. *Br J Haematol*, **87**(4), 789-96.

Fischer, L., Catz, D., and Kelley, D. (1993). An androgen receptor mRNA isoform associated with hormone-induced cell proliferation. *Proc Natl Acad Sci U S A*, **90**(17), 8254-8.

Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z. F., and Reinberg, D. (1991). The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. *Proc Natl Acad Sci U S A*, **88**(22), 9999-10003.

Fondell, J. D., Roy, A. L., and Roeder, R. G. (1993). Unliganded thyroid hormone receptor inhibits formation of a functional preinitiation complex: implications for active repression. *Genes Dev*, **7**(7B), 1400-10.

Frankel, A. D., Berg, J. M., and Pabo, C. O. (1987). Metal-dependent folding of a single zinc finger from transcription factor IIIA. *Proc Natl Acad Sci U S A*, **84**(14), 4841-5.

Freedman, S. J., Furie, B. C., Furie, B., and Baleja, J. D. (1995a). Structure of the calcium ion-bound gamma-carboxyglutamic acid-rich domain of factor IX. *Biochemistry*, **34**(38), 12126-37.

Freedman, S. J., Furie, B. C., Furie, B., and Baleja, J. D. (1995b). Structure of the metal-free gamma-carboxyglutamic acid-rich membrane binding region of factor IX by two-dimensional NMR spectroscopy. *J Biol Chem*, **270**(14), 7980-7.

Fried, M., and Crothers, D. M. (1981). Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res*, **9**(23), 6505-25.

Fryklund, L., Borg, H., and Andersson, L. O. (1976). Amino-terminal sequence of human factor IX: presence of gamma-carboxyl glumatic acid residues. *FEBS Lett*, **65**(2), 187-9.

Furie, B., Limentani, S. A., and Rosenfield, C. G. (1994). A practical guide to the evaluation and treatment of hemophilia. *Blood*, **84**(1), 3-9.

G

Gailani, D., and Broze, G., Jr. (1991). Factor XI activation in a revised model of blood coagulation. *Science*, **253**(5022), 909-12.

Gast, A., Neuschmid-Kaspar, F., Klocker, H., and Cato, A. C. (1995). A single amino acid exchange abolishes dimerization of the androgen receptor and causes Reifenstein syndrome. *Mol Cell Endocrinol*, **111**(1), 93-8.

Geiger, J. H., Hahn, S., Lee, S., and Sigler, P. B. (1996). Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science*, **272**(5263), 830-6.

Gething, M. J., and Sambrook, J. (1992). Protein folding in the cell. *Nature*, **355**(6355), 33-45.

Giannelli, F., Green, P. M., Sommer, S. S., Poon, M. C., Ludwig, M., Schwaab, R., Reitsma, P. H., Goossens, M., Yoshioka, A., and Brownlee, G. G. (1996). Haemophilia B (sixth edition): a database of point mutations and short additions and deletions. *Nucleic Acids Res*, **24**(1), 103-18.

Giese, K., Cox, J., and Grosschedl, R. (1992). The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell*, **69**(1), 185-95.

Giese, K., and Grosschedl, R. (1993). LEF-1 contains an activation domain that stimulates transcription only in a specific context of factor-binding sites. *EMBO J*, **12**(12), 4667-76.

Gill, G., Pascal, E., Tseng, Z. H., and Tjian, R. (1994). A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the Drosophila TFIID complex and mediates transcriptional activation. *Proc Natl Acad Sci U S A*, **91**(1), 192-6.

Gilmour, D. S., Dietz, T. J., and Elgin, S. C. (1990). UV cross-linking identifies four polypeptides that require the TATA box to bind to the Drosophila hsp70 promoter. *Mol Cell Biol*, **10**(8), 4233-8.

Goodrich, J. A., and Tjian, R. (1994). Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell*, **77**(1), 145-56.

Goodrich, J. A., Cutler, G., and Tjian, R. (1996). Contacts in context: promoter specificity and macromolecular interactions in transcription. *Cell*, **84**(6), 825-30.

Gottlieb, B., Trifiro, M., Lumbroso, R., Vasiliou, D. M., and Pinsky, L. (1996). The androgen receptor gene mutations database. *Nucleic Acids Res*, **24**(1), 151-4.

Graham, F. L., and Van der Eb, A. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology*, **52**(2), 456-67.

Graham, J. B., Kunkel, G. R., Egilmez, N. K., Wallmark, A., Fowlkes, D. M., and Lord, S. T. (1991). The varying frequencies of five DNA polymorphisms of X-linked coagulant factor IX in eight ethnic groups. *Am J Hum Genet*, **49**(3), 537-44.

Green, P. M., Montandon, A. J., Bentley, D. R., and Giannelli, F. (1991). Genetics and molecular biology of haemophilias A and B. *Blood Coagul Fibrinolysis*, **2**(4), 539-65.

Greenberg, D., Miao, C. H., Ho, W. T., Chung, D. W., and Davie, E. W. (1995). Liver-specific expression of the human factor VII gene. *Proc Natl Acad Sci U S A*, **92**(26), 12347-51.

Grigg, G., and Clark, S. (1994). Sequencing 5-methylcytosine residues in genomic DNA. *Bioessays*, **16**(6), 431-6.

Gronemeyer, H. (1992). Control of transcription activation by steroid hormone receptors. *FASEB J*, **6**(8), 2524-9.
Gronemeyer, H., and Laudet, V. (1995) Transcription factors 3: nuclear receptors. *Protein Profile (Academic Press Ltd., London)*, **2**(11).

H

Halle, J-P., and Meisterernst, M. (1996) Gene expression:increasing evidence for a transcriptosome. *Trends Genet*, **12**(5), 161-163.
Ham, J., Thomson, A., Needham, M., Webb, P., and Parker, M. (1988). Characterization of response elements for androgens, glucocorticoids and progestins in mouse mammary tumour virus. *Nucleic Acids Res*, **16**(12), 5263-76.
Handford, P. A., Baron, M., Mayhew, M., Willis, A., Beesley, T., Brownlee, G. G., and Campbell, I. D. (1990). The first EGF-like domain from human factor IX contains a high-affinity calcium binding site. *EMBO J*, **9**(2), 475-80.
Hankinson, O. (1995). The aryl hydrocarbon receptor complex. *Annu Rev Pharmacol Toxicol*, **35**, 307-40.
Hanna-Rose, W., and Hansen, U. (1996). Active repression mechanisms of eukaryotic transcription repressors. *Trends Genet*, **12**(6), 59-64.
Hard, T., Kellenbach, E., Boelens, R., Maler, B. A., Dahlman, K., Freedman, L. P., Carlstedt-Duke, J., Yamamoto, K. R., Gustafsson, J. A., and Kaptein, R. (1990a). Solution structure of the glucocorticoid receptor DNA-binding domain. *Science*, **249**(4965), 157-60.
Hard, T., Dahlman, K., Carlstedt-Duke, J., Gustafsson, J. A., and Rigler, R. (1990b). Cooperativity and specificity in the interactions between DNA and the glucocorticoid receptor DNA-binding domain. *Biochemistry*, **29**(22), 5358-64.
Hardon, E. M., Frain, M., Paonessa, G., and Cortese, R. (1988). Two distinct factors interact with the promoter regions of several liver-specific genes. *EMBO J*, **7**(6), 1711-9.
Harrison, S. C. (1991). A structural taxonomy of DNA-binding domains. *Nature*, **353**(6346), 715-9.
Heard, J. M., Herbomel, P., Ott, M. O., Mottura-Rollier, A., Weiss, M., and Yaniv, M. (1987). Determinants of rat albumin promoter tissue specificity analyzed by an improved transient expression system. *Mol Cell Biol*, **7**(7), 2425-34.
Hedner, U., and Davie, E.W. (1989). Introduction to hemostasis and the vitamin K-dependent coagulation factors. *The Metabolic Basis of Inherited Disease* (ed. by C.R. Scriver *et al.*) 6th edition, 2107-2134 (McGraw-Hill, New York)
Herbomel, P., Bourachot, B., and Yaniv, M. (1984). Two distinct enhancers with different cell specificities coexist in the regulatory region of polyoma. *Cell*, **39**(3 Pt 2), 653-62.

Hirosawa, S., Fahner, J. B., Salier, J. P., Wu, C. T., Lovrien, E. W., and Kurachi, K. (1990). Structural and functional basis of the developmental regulation of human coagulation factor IX gene: factor IX Leyden. *Proc Natl Acad Sci U S A*, **87**(12), 4421-5.

Ho, K. C., Marschke, K. B., Tan, J., Power, S. G., Wilson, E. M., and French, F. S. (1993). A complex response element in intron 1 of the androgen-regulated 20-kDa protein gene displays cell type-dependent androgen receptor specificity. *J Biol Chem*, **268**(36), 27226-35.

Hoey, T., Weinzierl, R. O., Gill, G., Chen, J. L., Dynlacht, B. D., and Tjian, R. (1993). Molecular cloning and functional analysis of Drosophila TAF110 reveal properties expected of coactivators. *Cell*, **72**(2), 247-60.

Hunger, S. P., Ohyashiki, K., Toyama, K., and Cleary, M. L. (1992). Hlf, a novel hepatic bZIP protein, shows altered DNA-binding properties following fusion to E2A in t(17;19) acute lymphoblastic leukemia. *Genes Dev*, **6**(9), 1608-20.

Hutchison, K. A., Dittmar, K. D., and Pratt, W. B. (1994). All of the factors required for assembly of the glucocorticoid receptor into a functional heterocomplex with heat shock protein 90 are preassociated in a self-sufficient protein folding structure, a "foldosome". *J Biol Chem*, **269**(45), 27894-9.

J

Jackson, D. A., Hassan, A. B., Errington, R. J., and Cook, P. R. (1993). Visualization of focal sites of transcription within human nuclei. *EMBO J*, **12**(3), 1059-65.

Jacobs, M., Freedman, S. J., Furie, B. C., and Furie, B. (1994). Membrane binding properties of the factor IX gamma-carboxyglutamic acid-rich domain prepared by chemical synthesis. *J Biol Chem*, **269**(41), 25494-501.

Jacobson, D. P., Schmeling, P., and Sommer, S. S. (1993). Characterization of the patterns of polymorphism in a "cryptic repeat" reveals a novel type of hypervariable sequence. *Am J Hum Genet*, **53**(2), 443-50.

Jiang, G., Nepomuceno, L., Hopkins, K., and Sladek, F. M. (1995). Exclusive homodimerization of the orphan receptor hepatocyte nuclear factor 4 defines a new subclass of nuclear receptors. *Mol Cell Biol*, **15**(9), 5131-43.

Johnson, J. L., Beito, T. G., Krco, C. J., and Toft, D. O. (1994). Characterization of a novel 23-kilodalton protein of unactive progesterone receptor complexes. *Mol Cell Biol*, **14**(3), 1956-63.

Johnson, J. L., and Toft, D. O. (1994). A novel chaperone complex for steroid receptors involving heat shock proteins, immunophilins, and p23. *J Biol Chem*, **269**(40), 24989-93.

Johnson, P. F., and McKnight, S. L. (1989). Eukaryotic transcriptional regulatory proteins. *Annu Rev Biochem*, **58**, 799-839.
Jones, P. A., Rideout, W. 3., Shen, J. C., Spruck, C. H., and Tsai, Y. C. (1992). Methylation, mutation and cancer. *Bioessays*, **14**(1), 33-6.
Jost, J. P., Munch, O., and Andersson, T. (1991). Study of protein-DNA interactions by surface plasmon resonance (real time kinetics). *Nucleic Acids Res*, **19**(10), 2788.

K

Kaelin, W., Jr., Pallas, D. C., DeCaprio, J. A., Kaye, F. J., and Livingston, D. M. (1991). Identification of cellular proteins that can interact specifically with the T/E1A-binding region of the retinoblastoma gene product. *Cell*, **64**(3), 521-32.
Kallio, P. J., Poukka, H., Moilanen, A., Janne, O. A., and Palvimo, J. J. (1995). Androgen receptor-mediated transcriptional regulation in the absence of direct interaction with a specific DNA element.*Mol Endocrinol* , **9**(8), 1017-28.
Kamei, Y., Xu, L., Heinzel, T., Torchia, J., Kurokawa, R., Gloss, B., Lin, S. C., Heyman, R. A., Rose, D. W., Glass, C. K., and Rosenfeld, M. G. (1996). A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell*, **85**(3), 403-14.
Kaplan, J. C., Kahn, A., and Chelly, J. (1992). Illegitimate transcription: its use in the study of inherited disease. *Hum Mutat*, **1**(5), 357-60.
Kaspar, F., Klocker, H., Denninger, A., and Cato, A. C. (1993). A mutant androgen receptor from patients with Reifenstein syndrome: identification of the function of a conserved alanine residue in the D box of steroid receptors. *Mol Cell Biol*, **13**(12), 7850-8.
Katzenellenbogen, J.A., O'Malley, B.W., and Katzenellenbogen, B.S. (1996) Tripartite steroid hormone receptor pharmacology: interactionwith multiple effector sites as a basis for the cell- and promoter-specific action of these hormones. *Mol Endocrinol*, **10**(2), 119-31.
Ketterling, R. P., Bottema, C. D., Phillips, J. 3., and Sommer, S. S. (1991). Evidence that descendants of three founders constitute about 25% of hemophilia B in the United States. *Genomics*, **10**(4), 1093-6.
Ketterling, R. P., Vielhaber, E., and Sommer, S. S. (1994). The rates of G:C→T:A and G:C→C:G transversions at CpG dinucleotides in the human factor IX gene. *Am J Hum Genet*, **54**(5), 831-5.
Ketterling, R. P., Liu, J. Z., Liao, D., Kasper, C. K., Ambriz, R., Paredes, R., and Sommer, S. S. (1995). Two novel factor IX promoter mutations: incremental progress towards 'saturation *in vivo* mutagenesis' of a human promoter region. *Hum Mol Genet*, **4**(4), 769-70.

Kisters-Woike, B., Lehming, N., Sartorius, J., von-Wilcken-Bergmann, B., and Muller-Hill, B. (1991). A model of the lac repressor-operator complex based on physical and genetic data. *Eur J Biochem*, **198**(2), 411-9.

Klein, C., and Struhl, K. (1994). Increased recruitment of TATA-binding protein to the promoter by transcriptional activation domains *in vivo*. *Science*, **266**(5183), 280-2.

Klug, A.. and Schwabe, J. W. (1995). Protein motifs 5. Zinc fingers. *FASEB J*, **9**(8), 597-604.

Knobloch, O., Zoll, B., Zerres, K., Brackmann, H. H., Olek, K., and Ludwig, M. (1993). Recurrent mutations in the factor IX gene: founder effect or repeat de novo events. Investigation of the German haemophilia B population and review of de novo mutations. *Hum Genet*, **92**(1), 40-8.

Kornberg, R. D., and Lorch, Y. (1991). Irresistible force meets immovable object: transcription and the nucleosome. *Cell*, **67**(5), 833-6.

Kraus, W. L., McInerney, E. M., and Katzenellenbogen, B. S. (1995). Ligand-dependent, transcriptionally productive association of the amino- and carboxyl-terminal regions of a steroid hormone nuclear receptor. *Proc Natl Acad Sci U S A*, **92**(26), 12314-8.

Krishnan, V., Wang, X., and Safe, S. (1994). Estrogen receptor-Sp1 complexes mediate estrogen-induced cathepsin D gene expression in MCF-7 human breast cancer cells. *J Biol Chem*, **269**(22), 15912-7.

Kuo, C. J., Conley, P. B., Chen, L., Sladek, F. M., Darnell, J., Jr., and Crabtree, G. R. (1992). A transcriptional hierarchy involved in mammalian cell-type specification. *Nature*, **355**(6359), 457-61.

Kurachi, K., Kurachi, S., Furukawa, M., and Yao, S. N. (1993). Biology of factor IX. *Blood Coagul Fibrinolysis*, **4**(6), 953-73.

Kurachi, S., Furukawa, M., Salier, J. P., Wu, C. T., Wilson, E. J., French, F. S., and Kurachi, K. (1994). Regulatory mechanism of human factor IX gene: protein binding at the Leyden-specific region. *Biochemistry*, **33**(6), 1580-91.

L

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**(259), 680-5.

Lai, E., and Darnell, J., Jr. (1991). Transcriptional control in hepatocytes: a window on development. *Trends Biochem Sci*, **16**(11), 427-30.

Lefstin, J. A., Thomas, J. R., and Yamamoto, K. R. (1994). Influence of a steroid receptor DNA-binding domain on transcriptional regulatory functions. *Genes Dev*, **8**(23), 2842-56.

Liao, S. M., Zhang, J., Jeffery, D. A., Koleske, A. J., Thompson, C. M., Chao, D. M., Viljoen, M., van-Vuuren, H. J., and Young, R. A. (1995). A kinase-cyclin pair in the RNA polymerase II holoenzyme. *Nature*, **374**(6518), 193-6.

Liu, W., Wang, J., Sauter, N. K., and Pearce, D. (1995). Steroid receptor heterodimerization demonstrated *in vitro* and *in vivo*. *Proc Natl Acad Sci U S A*, **92**(26), 12480-4.

Lucas, P. C., and Granner, D. K. (1992). Hormone response domains in gene transcription. *Annu Rev Biochem*, **61**, 1131-73.

Luckow, B., and Schutz, G. (1987). CAT constructions with multiple unique restriction sites for the functional analysis of eukaryotic promoters and regulatory elements. *Nucleic Acids Res*, **15**(13), 5490.

Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R., and Sigler, P. B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature*, **352**(6335), 497-505.

M

MacDougald, O. A., and Lane, M. D. (1995). Transcriptional regulation of gene expression during adipocyte differentiation. *Annu Rev Biochem*, **64**, 345-73.

Mader, S., Kumar, V., de-Verneuil, H., and Chambon, P. (1989). Three amino acids of the oestrogen receptor are essential to its ability to distinguish an oestrogen from a glucocorticoid-responsive element. *Nature*, **338**(6212), 271-4.

Maire, P., Wuarin, J., and Schibler, U. (1989). The role of cis-acting promoter elements in tissue-specific albumin gene expression. *Science*, **244**(4902), 343-6.

Malik, S., and Karathanasis, S.K. (1996) TFIIB-directed transcriptional activation by the orphan nuclear receptor hepatocyte nuclear factor 4. *Mol Cell Biol*, **16**(4), 1824-31.

Mangelsdorf, D. J., and Evans, R. M. (1995). The RXR heterodimers and orphan receptors. *Cell*, **83**(6), 841-50.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular cloning: a laboratory manual. *Cold Spring Harbour Laboratory Press.*

Medved, L. V., Vysotchin, A., and Ingham, K. C. (1994). Ca($^{2+}$)-dependent interactions between Gla and EGF domains in human coagulation factor IX. *Biochemistry*, **33**(2), 478-85.

Miller, J., McLachlan, A. D., and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J*, **4**(6), 1609-14.

Miyata, Y., and Yahara, I. (1991). Cytoplasmic 8 S glucocorticoid receptor binds to actin filaments through the 90-kDa heat shock protein moiety. *J Biol Chem*, **266**(14), 8779-83.

Morgan, G. E., Figueiredo, M. S., Winship, P. R., Baker, R., Bolton-Maggs, P. H., and Brownlee, G. G. (1995). The high frequency of the -6G→A factor IX promoter

mutation is the result both of a founder effect and recurrent mutation at a CpG dinucleotide. *Br J Haematol*, **89**(3), 672-4.
Mueller, C. R., Maire, P., and Schibler, U. (1990). DBP, a liver-enriched transcriptional activator, is expressed late in ontogeny and its tissue specificity is determined posttranscriptionally. *Cell*, **61**(2), 279-91.

N

Nagai, K., Nakaseko, Y., Nasmyth, K., and Rhodes, D. (1988). Zinc-finger motifs expressed in E. coli and folded *in vitro* direct specific binding to DNA. *Nature*, **332**(6161), 284-6.
Nagpal, S., Friant, S., Nakshatri, H., and Chambon, P. (1993). RARs and RXRs: evidence for two autonomous transactivation functions (AF-1 and AF-2) and heterodimerization *in vivo*. *EMBO J*, **12**(6), 2349-60.
Naka, H., and Brownlee, G. G. (1996). Transcriptional regulation of the human factor IX promoter by the orphan receptor superfamily factor, HNF4, ARP1 and COUP/Ear3. *Br J Haematol*, **92**(1), 231-40.
Nerlov, C., and Ziff, E. B. (1995). CCAAT/enhancer binding protein-alpha amino acid motifs with dual TBP and TFIIB binding ability co-operate to activate transcription in both yeast and mammalian cells. *EMBO J*, **14**(17), 4318-28.
Nikolov, D. B., Hu, S. H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N. H., Roeder, R. G., and Burley, S. K. (1992). Crystal structure of TFIID TATA-box binding protein. *Nature*, **360**(6399), 40-6.
Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G., and Burley, S. K. (1995). Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature*, **377**(6545), 119-28.
Nisen, P., Stamberg, J., Ehrenpreis, R., Velasco, S., Shende, A., Engelberg, J., Karayalcin, G., and Waber, L. (1986). The molecular basis of severe hemophilia B in a girl. *N Engl J Med*, **315**(18), 1139-42.

O

O'Malley, B. W., Schrader, W. T., Mani, S., Smith, C., Weigel, N. L., Conneely, O. M., and Clark, J. H. (1995). An alternative ligand-independent pathway for activation of steroid receptors. *Recent Prog Horm Res*, **50**, 333-47.
Ohlin, A. K., Linse, S., and Stenflo, J. (1988). Calcium binding to the epidermal growth factor homology region of bovine protein C. *J Biol Chem*, **263**(15), 7411-7.
Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J., and Gronenborn, A. M. (1993). NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science*, **261**(5120), 438-46.

Onate, S. A., Tsai, S. Y., Tsai, M. J., and O'Malley, B. W. (1995). Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science*, **270**(5240), 1354-7.
Orti, E., Bodwell, J. E., and Munck, A. (1992). Phosphorylation of steroid hormone receptors. *Endocr Rev*, **13**(1), 105-28.

P

Palvimo, J. J., Kallio, P. J., Ikonen, T., Mehto, M., and Janne, O. A. (1993). Dominant negative regulation of trans-activation by the rat androgen receptor: roles of the N-terminal domain and heterodimer formation. *Mol Endocrinol*, **7**(11), 1399-407.
Panayotou, G., Waterfield, M.D., and End, P. (1993). Riding the evanescent wave. *Curr Biol*, **3**(12), 913-915.
Pavlovsky, A. (1947) Contribution to the pathogenesis of haemophilia. *Blood*, **2**, 186-187.
Perutz, M. F., Johnson, T., Suzuki, M., and Finch, J. T. (1994). Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc Natl Acad Sci U S A*, **91**(12), 5355-8.
Peterson, C. L., and Herskowitz, I. (1992). Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. *Cell*, **68**(3), 573-83.
Picketts, D. J., Lillicrap, D. P., and Mueller, C. R. (1993). Synergy between transcription factors DBP and C/EBP compensates for a haemophilia B Leyden factor IX mutation. *Nat Genet*, **3**(2), 175-9.
Picketts, D. J., Mueller, C. R., and Lillicrap, D. (1994). Transcriptional control of the factor IX gene: analysis of five cis-acting elements and the deleterious effects of naturally occurring hemophilia B Leyden mutations. *Blood*, **84**(9), 2992-3000.
Pratt, W. B. (1993). The role of heat shock proteins in regulating the function, folding, and trafficking of the glucocorticoid receptor. *J Biol Chem*, **268**(29), 21455-8.
Price, D. H., Sluder, A. E., and Greenleaf, A. L. (1989). Dynamic interaction between a Drosophila transcription factor and RNA polymerase II. *Mol Cell Biol*, **9**(4), 1465-75.
Pruss, D., Hayes, J. J., and Wolffe, A. P. (1995). Nucleosomal anatomy--where are the histones? *Bioessays*, **17**(2), 161-70.

Q

Quigley, C. A., Evans, B. A., Simental, J. A., Marschke, K. B., Sar, M., Lubahn, D. B., Davies, P., Hughes, I. A., Wilson, E. M., and French, F. S. (1992). Complete androgen insensitivity due to deletion of exon C of the androgen receptor gene

highlights the functional importance of the second zinc finger of the androgen receptor *in vivo*. *Mol Endocrinol*, **6**(7), 1103-12.

R

Rao, Z., Handford, P., Mayhew, M., Knott, V., Brownlee, G. G., and Stuart, D. (1995). The structure of a Ca(2+)-binding epidermal growth factor-like domain: its role in protein-protein interactions. *Cell*, **82**(1), 131-41.

Reijnen, M. J., Sladek, F. M., Bertina, R. M., and Reitsma, P. H. (1992). Disruption of a binding site for hepatocyte nuclear factor 4 results in hemophilia B Leyden. *Proc Natl Acad Sci U S A*, **89**(14), 6300-3.

Reijnen. M. J., Peerlinck, K., Maasdam, D., Bertina, R. M., and Reitsma, P. H. (1993). Hemophilia B Leyden: substitution of thymine for guanine at position -21 results in a disruption of a hepatocyte nuclear factor 4 binding site in the factor IX promoter. *Blood*, **82**(1), 151-8.

Reitsma, P. H., Bertina, R. M., Ploos-van-Amstel, J. K., Riemens, A., and Briet, E. (1988). The putative factor IX gene promoter in hemophilia B Leyden. *Blood*, **72**(3), 1074-6.

Reitsma, P. H., Mandalaki, T., Kasper, C. K., Bertina, R. M., and Briet, E. (1989). Two novel point mutations correlate with an altered developmental expression of blood coagulation factor IX (hemophilia B Leyden phenotype). *Blood*, **73**(3), 743-6.

Renaud, J. P., Rochel, N., Ruff, M., Vivat, V., Chambon, P., Gronemeyer, H., and Moras, D. (1995). Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid. *Nature*, **378**(6558), 681-9.

Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., and Klug, A. (1984). Structure of the nucleosome core particle at 7 A resolution. *Nature*, **311**(5986), 532-7.

Riggs, A. D., Bourgeois, S., and Cohn, M. (1970). The lac repressor-operator interaction. 3. Kinetic studies. *J Mol Biol*, **53**(3), 401-17.

Roberts, S. G., and Green, M. R. (1994). Activator-induced conformational change in general transcription factor TFIIB. *Nature*, **371**(6499), 717-20.

Roche, P. J., Hoare, S. A., and Parker, M. G. (1992). A consensus DNA-binding site for the androgen receptor. *Mol Endocrinol*, **6**(12), 2229-35.

S

Salier, J. P., Hirosawa, S., and Kurachi, K. (1990). Functional characterization of the 5'-regulatory region of human factor IX gene. *J Biol Chem*, **265**(12), 7062-8.

Sarkar, G., and Sommer, S. S. (1989). Access to a messenger RNA sequence or its protein product is not limited by tissue or species specificity. *Science*, **244**(4902), 331 4.

Sarkar, G., Paynton, C., and Sommer, S. S. (1991). Segments containing alternating purine and pyrimidine dinucleotides: patterns of polymorphism in humans and prevalence throughout phylogeny. *Nucleic Acids Res*, **19**(3), 631-6.

Sarkar, G., Yoon, H. S., and Sommer, S. S. (1992). Screening for mutations by RNA single-strand conformation polymorphism (rSSCP): comparison with DNA-SSCP. *Nucleic Acids Res*, **20**(4), 871-8.

Sauer, F., Fondell, J. D., Ohkuma, Y., Roeder, R. G., and Jackle, H. (1995). Control of transcription by Kruppel through interactions with TFIIB and TFIIE beta. *Nature*, **375**(6527), 162-4.

Schuster, S. C., Swanson, R. V., Alex, L. A., Bourret, R. B., and Simon, M. I. (1993). Assembly and function of a quaternary signal transduction complex monitored by surface plasmon resonance. *Nature*, **365**(6444), 343-7.

Schwabe, J. W., Neuhaus, D., and Rhodes, D. (1990). Solution structure of the DNA-binding domain of the oestrogen receptor. *Nature*, **348**(6300), 458-61.

Schwabe, J. W., Chapman, L., Finch, J. T., and Rhodes, D. (1993). The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell*, **75**(3), 567-78.

Serizawa, H., Makela, T. P., Conaway, J. W., Conaway, R. C., Weinberg, R. A., and Young, R. A. (1995). Association of Cdk-activating kinase subunits with transcription factor TFIIH. *Nature*, **374**(6519), 280-2.

Singh, H., Sen, R., Baltimore, D., and Sharp, P. A. (1986). A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. *Nature*, **319**(6049), 154-8.

Sladek, F. M., Zhong, W. M., Lai, E., and Darnell, J., Jr. (1990). Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. *Genes Dev*, **4**(12B), 2353-65.

Smale, S. T., and Baltimore, D. (1989). The "initiator" as a transcription control element. *Cell*, **57**(1), 103-13.

Smith, C. L., Conneely, O. M., and O'Malley, B. W. (1993). Modulation of the ligand-independent activation of the human estrogen receptor by hormone and antihormone. *Proc Natl Acad Sci U S A*, **90**(13), 6120-4.

Smith, D. F., Faber, L. E., and Toft, D. O. (1990). Purification of unactivated progesterone receptor and identification of novel receptor-associated proteins. *J Biol Chem*, **265**(7), 3996-4003.

Smith, D. F., Sullivan, W. P., Marion, T. N., Zaitsu, K., Madden, B., McCormick, D. J., and Toft, D. O. (1993). Identification of a 60-kilodalton stress-related protein, p60, which interacts with hsp90 and hsp70. *Mol Cell Biol*, **13**(2), 869-76.

Smith, D. F., and Toft, D. O. (1993). Steroid receptors and their associated proteins. *Mol Endocrinol*, **7**(1), 4-11.

Sommer, S. S., Bowie, E. J., Ketterling, R. P., and Bottema, C. D. (1992). Missense mutations and the magnitude of functional deficit: the example of factor IX. *Hum Genet*, **89**(3), 295-7.
Soriano-Garcia, M., Padmanabhan, K., de-Vos, A. M., and Tulinsky, A. (1992). The Ca^{2+} ion and membrane binding structure of the Gla domain of Ca-prothrombin fragment 1. *Biochemistry*, **31**(9), 2554-66.
Stenflo, J., and Ganrot, P. O. (1973). Binding of Ca^{2+} to normal and dicoumarol-induced prothrombin. *Biochem Biophys Res Commun*, **50**(1), 98-104.
Stowell, K. M., Figueiredo, M. S., Brownlee, G. G., Jones, P., and Bolton-Maggs, P. H. (1993). Haemophilia B Liverpool: a new British family with mild haemophilia B associated with a -6 G to A mutation in the factor IX promoter. *Br J Haematol*, **85**(1), 188-90.
Straight, D. L., Sherrill, G. B., Noyes, C. M., Trapp, H. G., Wright, S. F., Roberts, H. R., Hiskey, R. G., and Griffith, M. J. (1985). Structural and functional characteristics of activated human factor IX after chemical modification of gamma-carboxyglutamic acid residues. *J Biol Chem*, **260**(5), 2890-3.
Strubin, M., and Struhl, K. (1992). Yeast and human TFIID with altered DNA-binding specificity for TATA elements. *Cell*, **68**(4), 721-30.
Struhl, K. (1996). Chromatin structure and RNA polymerase II connection: implications for transcription. *Cell*, **84**(2), 179-82.
Sunnerhagen, M. S., Persson, E., Dahlqvist, I., Drakenberg, T., Stenflo, J., Mayhew, M., Robin, M., Handford, P., Tilley, J. W., Campbell, I. D., and et, a. (1993). The effect of aspartate hydroxylation on calcium binding to epidermal growth factor-like modules in coagulation factors IX and X. *J Biol Chem*, **268**(31), 23339-44.
Suzuki, M., and Yagi, N. (1994). DNA recognition code of transcription factors in the helix-turn-helix, probe helix, hormone receptor, and zinc finger families. *Proc Natl Acad Sci U S A*, **91**(26), 12357-61.

T

Tai, P. K., Albers, M. W., Chang, H., Faber, L. E., and Schreiber, S. L. (1992). Association of a 59-kilodalton immunophilin with the glucocorticoid receptor complex. *Science*, **256**(5061), 1315-8.
Tan, S., Garrett, K. P., Conaway, R. C., and Conaway, J. W. (1994). Cryptic DNA-binding domain in the C terminus of RNA polymerase II general transcription factor RAP30. *Proc Natl Acad Sci U S A*, **91**(21), 9808-12.
Tan, S., Hunziker, Y., Sargent, D.F., Richmond, T.J. (1996). Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature*, **381**(6578), 127-34.
Thompson, A. R. (1987). Alloantibodies in hemophilia B binding to multiple factor IX epitopes. *Thromb Res*, **46**(1), 169-74.

Tora, L., White, J., Brou, C., Tasset, D., Webster, N., Scheer, E., and Chambon, P. (1989). The human estrogen receptor has two independent nonacidic transcriptional activation functions. *Cell*, **59**(3), 477-87.

Travers, A. A. (1994). Chromatin structure and dynamics. *Bioessays*, **16**(9), 657-62.

Triezenberg, S. J. (1995). Structure and function of transcriptional activation domains. *Curr Opin Genet Dev*, **5**(2), 190-6.

Tronche, F., and Yaniv, M. (1992). HNF1, a homeoprotein member of the hepatic transcription regulatory network. *Bioessays*, **14**(9), 579-87.

Truss, M., Bartsch, J., Schelbert, A., Hache, R. J., and Beato, M. (1995). Hormone induces binding of receptors and transcription factors to a rearranged nucleosome on the MMTV promoter *in vivo*. *EMBO J*, **14**(8), 1737-51.

Tsai, M. J., and O'Malley, B. W. (1994). Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annu Rev Biochem*, **63**, 451-86.

Tsai, S. Y., Carlstedt-Duke, J., Weigel, N. L., Dahlman, K., Gustafsson, J. A., Tsai, M. J., and O'Malley, B. W. (1988). Molecular interactions of steroid hormone receptor with its enhancer element: evidence for receptor dimer formation. *Cell*, **55**(2), 361-9.

Tyree, C. M., George, C. P., Lira-DeVito, L. M., Wampler, S. L., Dahmus, M. E., Zawel, L., and Kadonaga, J. T. (1993). Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II. *Genes Dev*, **7**(7A), 1254-65.

Tzukerman, M. T., Esty, A., Santiso-Mere, D., Danielian, P., Parker, M. G., Stein, R. B., Pike, J. W., and McDonnell, D. P. (1994). Human estrogen receptor transactivational capacity is determined by both cellular and promoter context and mediated by two functionally distinct intramolecular regions. *Mol Endocrinol*, **8**(1), 21-30.

V

Vandromme, M., Gauthier-Rouviere, C., Lamb, N., and Fernandez, A. (1996) Regulation of transcription factor localization:fine-tuning of gene expression. *Trends Biochem Sci*, **21**(2), 59-64.

Verrijzer, C. P., Yokomori, K., Chen, J. L., and Tjian, R. (1994). Drosophila TAFII150: similarity to yeast gene TSM-1 and specific binding to core promoter DNA. *Science*, **264**(5161), 933-41.

Verrijzer, C. P., Chen, J. L., Yokomori, K., and Tjian, R. (1995). Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell*, **81**(7), 1115-25.

W

Wagner, R. L., Apriletti, J. W., McGrath, M. E., West, B. L., Baxter, J. D., and Fletterick, R. J. (1995). A structural role for hormone in the thyroid hormone receptor. *Nature*, **378**(6558), 690-7.

Wang, Y., Hahn, T. M., Tsai, S. Y., and Woo, S. L. (1994). Functional characterization of a unique liver gene promoter. *J Biol Chem*, **269**(12), 9137-46.

Webb, P., Lopez, G. N., Uht, R. M., and Kushner, P. J. (1995). Tamoxifen activation of the estrogen receptor/AP-1 pathway: potential origin for the cell-specific estrogen-like effects of antiestrogens. *Mol Endocrinol*, **9**(4), 443-56.

Welshons, W. V., Wolf, M. F., Murphy, C. S., and Jordan, V. C. (1988). Estrogenic activity of phenol red. *Mol Cell Endocrinol*, **57**(3), 169-78.

Whitelaw, M. L., Hutchison, K., and Perdew, G. H. (1991). A 50-kDa cytosolic protein complexed with the 90-kDa heat shock protein (hsp90) is the same protein complexed with pp60v-src hsp90 in cells transformed by the Rous sarcoma virus. *J Biol Chem*, **266**(25), 16436-40.

Wiech, H., Buchner, J., Zimmermann, R., and Jakob, U. (1992). Hsp90 chaperones protein folding *in vitro*. *Nature*, **358**(6382), 169-70.

Wilson, C. J., Chao, D. M., Imbalzano, A. N., Schnitzler, G. R., Kingston, R. E., and Young, R. A. (1996). RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell*, **84**(2), 235-44.

Wilson, C. M., and McPhaul, M. J. (1994). A and B forms of the androgen receptor are present in human genital skin fibroblasts. *Proc Natl Acad Sci U S A*, **91**(4), 1234-8.

Winship, P. R., Nichols, C. E., Chuansumrit, A., and Peake, I. R. (1993). An MseI RFLP in the 5' flanking region of the factor IX gene: its use for haemophilia B carrier detection in Caucasian and Thai populations. *Br J Haematol*, **84**(1), 101-5.

Winston, F., and Carlson, M. (1992). Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. *Trends Genet*, **8**(11), 387-91.

Wong, C. I., Zhou, Z. X., Sar, M., and Wilson, E. M. (1993). Steroid requirement for androgen receptor dimerization and DNA binding. Modulation by intramolecular interactions between the NH2-terminal and steroid-binding domains. *J Biol Chem*, **268**(25), 19004-12.

Wu, S. M., Morris, D. P., and Stafford, D. W. (1991). Identification and purification to near homogeneity of the vitamin K-dependent carboxylase. *Proc Natl Acad Sci U S A*, **88**(6), 2236-40.

X

Xanthopoulos, K. G., Prezioso, V. R., Chen, W. S., Sladek, F. M., Cortese, R., and Darnell, J., Jr. (1991). The different tissue transcription patterns of genes for HNF-1,

C/EBP, HNF-3, and HNF-4, protein factors that govern liver-specific transcription. *Proc Natl Acad Sci U S A*, **88**(9), 3807-11.

Xie, Y. B., Sui, Y. P., Shan, L. X., Palvimo, J. J., Phillips, D. M., and Janne, O. A. (1992). Expression of androgen receptor in insect cells. Purification of the receptor and renaturation of its steroid- and DNA-binding functions. *J Biol Chem*, **267**(7), 4939-48.

Y

Yamada, M., Kuliopulos, A., Nelson, N. P., Roth, D. A., Furie, B., Furie, B. C., and Walsh, C. T. (1995). Localization of the factor IX propeptide binding site on recombinant vitamin K dependent carboxylase using benzoylphenylalanine photoaffinity peptide inactivators. *Biochemistry*, **34**(2), 481-9.

Yeh, W. C., Cao, Z., Classon, M., and McKnight, S. L. (1995). Cascade regulation of terminal adipocyte differentiation by three members of the C/EBP family of leucine zipper proteins. *Genes Dev*, **9**(2), 168-81.

Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W., and Kurachi, K. (1985). Nucleotide sequence of the gene for human factor IX (antihemophilic factor B). *Biochemistry*, **24**(14), 3736-50.

Z

Zabeau, M., and Stanley, K. K. (1982). Enhanced expression of cro-beta-galactosidase fusion proteins under the control of the PR promoter of bacteriophage lambda. *EMBO J*, **1**(10), 1217-24.

Zawel, L., and Reinberg, D. (1995). Common themes in assembly and function of eukaryotic transcription complexes. *Annu Rev Biochem*, **64**, 533-61.

Zhou, Z. X., Kemppainen, J. A., and Wilson, E. M. (1995). Identification of three proline-directed phosphorylation sites in the human androgen receptor. *Mol Endocrinol*, **9**(5), 605-15.

Appendix 1

Oligonucleotides used in this study, all sequences read 5'→3'.

Gel shift oligonucleotides

Factor IX Wild type oligonucleotides (-40→-9)

CTAGAAATACAGCTCAGCTTGTACTTTGGTACAACTAA

GATCTTAGTTGTACCAAAGTACAAGCTGAGCTGTATTT

-26 G→C

CTAGAAATACAGCTCAGCTTCTACTTTGGTACAACTAA

GATCTTAGTTGTACCAAAGTAGAAGCTGAGCTGTATTT

-36 A→T

CTAGAAATACTGCTCAGCTTGTACTTTGGTACAACTAA

GATCTTAGTTGTACCAAAGTACAAGCTGAGCAGTATTT

-35 G→C

CTAGAAATACACCTCAGCTTGTACTTTGGTACAACTAA

GATCTTAGTTGTACCAAAGTACAAGCTGAGGTGTATTT

-34 C→G

CTAGAAATACAGGTCAGCTTGTACTTTGGTACAACTAA

GATCTTAGTTGTACCAAAGTACAAGCTGACCTGTATTT

-33 T→C

CTAGAAATACAGCCCAGCTTGTACTTTGGTACAACTAA

GATCTTAGTTGTACCAAAGTACAAGCTGGGCTGTATTT

-32 C→A

CTAGAAATACAGCTAAGCTTGTACTTTGGTACAACTAA

GATCTTAGTTGTACCAAAGTACAAGCTTAGCTGTATTT

-31 A→C

CTAGAAATACAGCTCCGCTTGTACTTTGGTACAACTAA

GATCTTAGTTGTACCAAAGTACAAGCGGAGCTGTATTT

-30 G→A

CTAGAAATACAGCTCAACTTGTACTTTGGTACAACTAA

GATCTTAGTTGTACCAAAGTACAAGTTGAGCTGTATTT

-28 T→G

CTAGAAATACAGCTCAGCGTGTACTTTGGTACAACTAA

GATCTTAGTTGTACCAAAGTACACGCTGAGCTGTATTT

-27 T→G

CTAGAAATACAGCTCAGCTGGTACTTTGGTACAACTAA

GATCTTAGTTGTACCAAAGTACCAGCTGAGCTGTATTT

-26 G→A

CTAGAAATACAGCTCAGCTTATACTTTGGTACAACTAA

GATCTTAGTTGTACCAAAGTATAAGCTGAGCTGTATTT

-25 T→G

CTAGAAATACAGCTCAGCTTGGACTTTGGTACAACTAA

GATCTTAGTTGTACCAAAGTCCAAGCTGAGCTGTATTT

-24 A→T

CTAGAAATACAGCTCAGCTTGTTCTTTGGTACAACTAA

GATCTTAGTTGTACCAAAGAACAAGCTGAGCTGTATTT

-24 A→C

CTAGAAATACAGCTCAGCTTGTCCTTTGGTACAACTAA

GATCTTAGTTGTACCAAAGGACAAGCTGAGCTGTATTT

-23 C→T

CTAGAAATACAGCTCAGCTTGTATTTTGGTACAACTAA

GATCTTAGTTGTACCAAAATACAAGCTGAGCTGTATTT

-23 C→A

CTAGAAATACAGCTCAGCTTGTAATTTGGTACAACTAA

GATCTTAGTTGTACCAAATTACAAGCTGAGCTGTATTT

-22 T→A

CTAGAAATACAGCTCAGCTTGTACATTGGTACAACTAA

GATCTTAGTTGTACCAATGTACAAGCTGAGCTGTATTT

-21/-20 T→G

CTAGAAATACAGCTCAGCTTGTACTGGGGTACAACTAA

GATCTTAGTTGTACCCCAGTACAAGCTGAGCTGTATTT

-19 G→C

CTAGAAATACAGCTCAGCTTGTACTTTCGTACAACTAA

GATCTTAGTTGTACGAAAGTACAAGCTGAGCTGTATTT

α1 antitrypsin HNF4 site

GACTCAGATCCCAGCCAGTGGACTTAGCCCCTGTTT

AAACAGGGGCTAAGTCCACTGGCTGGGATCTGAGTC

PBP C3 ARE

CTAGCTATGAGTACGTGATGTTCTAGAGCG

GATCCGCTCTAGAACATCACGTACTCATAG

Albumin D site

GATGGTATGATTTTGTAATGGGGTAGG

CCTACCCCATTACAAAATCATACCATC

PCR Primers

Factor IX minimal promoter region (-221→+43)

Forward: GCGGGAGCTCCATTCTGAAGATGTGTCAC

Reverse: TATGCAAGCTTTTCACGCGCTGCATAACCTTTGCTAGCA

Factor IX polymorphic loci for the haplotype study

*Mse*I

Forward: GATAGAGAAACTGGAAGTAGACCC

Reverse: TTAGGTCTTTCACAGAGTAGAATTT

*Bam*HI

Forward: CAGTCGAGGAAGGATAGGGT

Reverse: GTATGAGTGGTCCAGTTAGTC

*Dde*I

Forward: GGGACCACTGTCGTATAATGTGG

Reverse: CTGGAGGATAGATGTCTCTATCTG

*Xmn*I

Forward: AATCAGAGACTGCTGATTGACTT

Reverse: GAAACAGCCAGATAAAGCCTCCA

*Taq*I

Forward: CTGGAGTATGACTGGCCAATTATCC

Reverse: GGTACACAAGGATTCTAAGGTTG

*Msp*I

Forward: GTACATATTATTACATGACGCAG

Reverse: GTATGTTGACTCTGTATCCAGTA

*Mnl*I

Forward: GATTTGAAAACTGTCCATGAAAATAAC

Reverse: AAAGTACCTGCCAAGGGAATTGACCTG

*Hha*I

Forward: ACAGGCACCTGCCATCACTT

Reverse: AGATTTCAAGCTACCAACAT

Appendix 2

Data for determination of the frequencies of haplotypes I→IV in the Caucasian population (see Chapter 7). Data was provided by Dr. P. Winship. '+' denotes the presence of a restriction site, and '-' denotes its absence. For DdeI, '+' denotes the A_2B_2 allele and '-' denotes the AB allele described in Jacobson et al., (1993). N= normal, i.e. non-haemophiliac member of the UK Caucasian population. (F) denotes female, and (M) denotes male. ND= not determined.

N (F)	*Mse*I	*Dde*I	*Xmn*I	*Taq*I	*Mnl*I	*Msp*I	*Hha*I	*Bam*HI
N11	- -	+ -	- -	-	- -	+ +	-	- -
N12	- -	- -	- -	- -	- -	+ +	- -	- -
N13	+ +	- -	+ +	+ +	+ +	- -	+ +	- -
N14	- -	+ +	- -	- -	- -	+ +	+ +	- -
N15	- -	+	- -	+ -	+ -	+ +	+	- -
N16	+ -	- -	+ -	+ -	+ -	+ -	+ -	- -
N17	+ -	- -	+ -	+ -	+	+ -		- -
N18	+ -	- -	+ -	+ -	+ -	+ -	- -	- -
N19	+ -	- -	-	- -	- -	+ +	+ -	- -
N20	+ -	ND	ND	ND	ND	ND	ND	ND
N30	+ -	- -	+	+ -	- -	+ +	+ -	- -
N31	+ -	+ -	+ -	+ -	+ -	+ -	+ +	- -
N32	- -	- -	- -	- -	- -	+ +	+ -	- -
N33	+ -	+ -	+ -	+ -	+ -	+ -	-	- -
N34	+ -	- -	+ -	+ -	+ -	+ -	+ -	- -
N35	- -	+ -	- -	- -	-	+ +	+ -	- -
N36	+ -	+ -	+ -	+ -	+ -	+ -	+ +	- -
N37	+ -	+ -	+ -	+ -	- -	+ +	+ +	- -
N38	+ -	- -	+ -	+ -	+ -	+ -	+ +	- -

N (F)	*Mse*I	*Dde*I	*Xmn*I	*Taq*I	*Mnl*I	*Msp*I	*Hha*I	*Bam*HI
N39	+ -	- -	+ -	+ +	+ +	+ -	+ -	- -
N40	- -	+ +	- -	- -	- -	+ +	+ +	- -
N41	- -	+ -	- -	- -	- -	+ +	+ +	- -
N42	+ -	+ -	- -	- -	- -	+ +	+ -	- -
N43	- -	-	- -	+ -	+ -	+ +	- -	- -
N44	- -	+ -	-	- -	- -	+ +	+ +	- -
N45	+ -	- -	+ -	+ -	+ -	+ -	+ +	- -
N5	ND	- -	+ -	+ -	ND	ND	ND	ND
N6	ND	- -	+ +	+ +	ND	ND	ND	ND
N7	ND	+ -	-	- -	ND	ND	ND	ND
N8	ND	- -	-	- -	ND	ND	ND	ND
N9	ND	- -	+ +	+ +	ND	ND	ND	ND
N10	ND	+ +	- -	- -	ND	ND	ND	ND
N23	ND	+ -	ND	- -	ND	ND	ND	ND
N (M)								
1	-		-	+	+	+	-	-
2	-		-	-	-	+	-	-
3	ND	ND	ND	+	+	+	ND	-
4	+	-	+	+	ND	ND	+	ND
5	-	+	-	-	-	+	-	-
6	ND	-	-	-	ND	ND	ND	ND
7	+	-	+	+	+	-	+	-
8	-	-	-	-	-	+	+	-

Publications

Morgan, G.E., Williams, A., Schach, B. G., Yoshitake, S., and Brownlee, G.G. (1993). The Sequence of the 5' Flanking Region of the Gene for Human Clotting Factor IX. *Submitted to the EMBL Database, Heidelberg* (Acession number=x75349).

Morgan, G.E., Figueiredo M.S., Winship, P.R., Bolton-Maggs, P.H.B, and Brownlee, G.G. (1994). A Possible Founder Effect for the -6 G→A Mutations in the Factor IX Gene Promoter in UK Haemophilia B Leyden Patients. *British Journal of Haematology* **87** (Suppl. 1) 44.

Morgan, G.E., Figueiredo, M.S., Winship, P.R., Baker, R., Bolton-Maggs, P.H.B., and Brownlee, G.G. (1995).The High Frequency of a Factor IX Promoter Mutation in the Haemophilia B Leyden Population is Caused by a Founder Effect and Recurrent Mutation at a CpG Dinucleotide. *British Journal of Haematology*, **89**, 672-674.

Morgan, G.E., Rowley, G., Green, P.M., Chisholm, M., Giannelli, F., and Brownlee, G.G. (1996). Patient UK232: Further evidence for the role of an androgen response element in the recovery of patients from haemophilia B Leyden. Publication Submitted.