

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

# Hepatocyte Nuclear Factor 1 and the Regulation of the Human Factor IX Gene.

Sarah Jane Penning  
1998

A thesis presented in partial fulfilment of  
the requirements for the degree of  
Master of Science in Molecular Biology  
at Massey University.

## Acknowledgements

Firstly, I would like to thank my supervisor Dr. Kathryn Stowell. Not only were you a constant source of technical guidance, but you also managed to supply endless support and enthusiasm at exactly the right moments. I not only want to thank you for your motivation throughout my time at Massey, I know I am indebted to you for encouraging me to begin postgraduate studies in the first place.

Secondly, I would like to thank a number of people whom have helped me to complete my Masters through their advice and knowledge. To Dr. Catherine Day, a big thank you for obtaining the DCoH expression vector, and for helping out with good advice. Thanks to Dr. Shaun Lott for helping me to learn that there's always more than one solution to every cloning problem, and to Neil Peterson for more in-depth and insightful discussions than I can even remember. An extra special thanks must go to Carole Flyger for her exhaustive efforts to keep the Twilight Zone running, and to keep demanding people like myself happy. Lorraine Berry undertook a large amount of DNA sequencing on my behalf, an exercise for which I am grateful. Thanks to Nic McBride for some manual sequencing also. Another area in which I received great deal of assistance was the use of the computers. Thanks to Associate Professor John Tweedie.

I must also acknowledge the financial assistance of the Massey University Graduate Research Fund (MUGRF) and the Massey University Scholarships Committee for awarding me a Massey Masterate Scholarship.

Lastly I'd like to thank those people who not only chose to help me in the scientific area, but also allowed me the privilege of their friendship: sometimes a friendly smile (and a medicinal whisky) can mean more than solutions to problems. And to Sherryn and Brent White and Sarah Callaghan, my thanks for being there and supporting me when times got rough.

To my parents, Royce and Betty Penning: you have encouraged me throughout my entire life. With your reassurance, I never once believed that my dreams weren't achievable. Thank you for your endless support and your belief in me.

And last, but by no means least, to Nic: you were always the one to bear the brunt of the problems: whatever the stress of the moment, you took it in your stride with patience. Your encouragement and understanding are the things that supported me through to the end, and I will always be grateful.

## Abstract

Factor IX is a serine protease involved in the mammalian blood clotting cascade. An absence of functional factor IX protease in the bloodstream results in Haemophilia B. Mutations in the regulatory region of the factor IX gene can produce a rare form of the disease called Haemophilia B Leyden. Single nucleotide substitutions at positions -5 and -6 of the human factor IX promoter, which result in Haemophilia B Leyden, disrupt the binding of an unidentified transcription factor which interacts in the region -13 to +3. A group of transcription factors which may interact with the factor IX promoter in this region is called the hepatocyte nuclear factor 1 (HNF1) family.

The aim of this research was to investigate the potential role of the HNF1 proteins in the regulation of factor IX promoter gene expression. This study was bipartite, involving research into the ability of the HNF1 transcription factors to bind the factor IX promoter *in vitro*, and to regulate the initiation of its transcription. The HNF1 cDNAs were firstly subcloned into an expression vector suitable for use in mammalian tissue culture.

Gel mobility shift assays were employed to examine the binding of the HNF1 proteins to the wildtype factor IX promoter. The ability of these proteins to bind the factor IX promoter region carrying the -5 or -6 mutations was also investigated. Luciferase reporter gene assays using a human hepatoma cell line were used to study the regulatory effects of the HNF1 transcription factors on transcription from the normal and mutant factor IX promoters.

A variant form of the HNF1 transcription factor was shown to bind to the -14 to +6 region of the normal sequence of the factor IX promoter as well as that containing some of the -5 and -6 mutations. A protein from rat liver nuclear extracts which displayed an HNF1-like binding activity was detected using gel mobility shift assays with the factor IX promoter region. All forms of the HNF1 transcription factors could regulate the transcription of a reporter gene driven by the wildtype factor IX promoter. Two forms of the HNF1 transcription factor down-regulated the wildtype factor IX promoter-reporter gene construct by at least 50%, while the same construct was up-regulated by a third form of the transcription factor.

Unfortunately time constraints resulted in the premature conclusion of planned experimentation. The results generated during this research have been unable to confirm a role for the HNF1 family in the regulation of the factor IX promoter, but have provided a basis for further research.



# CONTENTS.

	Page
Abstract.	i
Contents.	ii
List of Figures.	vi
List of Tables.	ix
Abbreviations and Definitions.	x
 <b>CHAPTER ONE: INTRODUCTION.</b>	 1
 <b>1.1 Factor IX.</b>	 1
1.1.1 Haemophilia B Leyden.	2
1.1.2 Transcription in Eukaryotes	3
<i>Promoters and Enhancers</i>	4
<i>The Action of Transcription Factors.</i>	4
1.1.3 Liver-specific expression of Factor IX.	5
1.1.4 Haemophilia B Leyden and Phenotypic Recovery.	6
<b>1.2 HNF1 Family.</b>	7
1.2.1 DNA binding domains.	8
1.2.2 Dimerisation domain.	10
1.2.3 HNF1 $\beta$ .	10
1.2.4 Activation domains (ADs).	11
1.2.5 Isoforms of HNF1 $\alpha$ and HNF1 $\beta$ .	12
1.2.6 DCoH.	14
<b>1.3 The HNF1 Family and Haemophilia B.</b>	15
1.3.1 Aims of this thesis.	16
 <b>CHAPTER TWO: MATERIALS AND METHODS.</b>	 18
 <b>2.1 Materials.</b>	 18
<b>2.2 Methods.</b>	20
2.2.1 Techniques for DNA Manipulation.	20
<i>General Methods.</i>	20
<i>Preparation of HNF1<math>\alpha</math> and DCoH cDNAs for Subcloning.</i>	21
<i>Manual Sequencing of PCR Products.</i>	21
<i>Automatic Sequencing of PCR Products.</i>	22

2.2.2	Mammalian Tissue Culture.	22
	<i>Maintenance of Tissue Cultures.</i>	22
	<i>Transient Transfection of Tissue Culture.</i>	23
	<i>Harvesting of Transfected Cos7 Cell Cultures.</i>	23
	<i>Harvesting of Transfected Alexander Cell Cultures.</i>	23
2.2.3	Electrophoretic Mobility Shift Assays (EMSAs).	24
	<i>Labelling of DNA probes for Gel Shifts.</i>	24
	<i>The Gel Shift.</i>	24
2.2.4	Rat Liver Nuclear Extracts.	25
2.2.5	Transcriptional Assays.	25

## CHAPTER THREE: RESULTS AND DISCUSSIONS. 27

### Part A: Subcloning of the cDNAs for HNF1 $\alpha$ , HNF1 $\beta$ and DCoH.

3.1	Preparation of the HNF1 $\alpha$ Expression Constructs.	27
3.1.1	Restriction Enzyme Digests.	27
3.1.2	PCR Amplification of the HNF1 $\alpha$ cDNA.	37
3.1.3	Cloning of the HNF1 $\alpha$ cDNA.	42
3.1.4	Sequencing of the HNF1 $\alpha$ cDNA.	45
3.1.5	Discussion of the Subcloning Difficulties with the HNF1 $\alpha$ cDNA.	50
	<i>Eco RI Restriction Endonuclease Digests</i>	50
	<i>DNA Sequencing and PCR</i>	50
3.2	pSVK3-HNF1 $\beta$ Subcloning.	53
3.2.1	Restriction Enzyme Digests of the pBJ5-HNF1 $\beta$ Vector	53
3.2.2	Cloning of the HNF1 $\beta$ cDNA.	54
3.3	Preparation of the DCoH Expression Construct.	57
3.3.1	PCR.	57
3.3.2	Cloning of the DCoH cDNA.	58
3.3.3	Sequencing of the pSVK3-DCoH Construct	59
3.3.4	Summary of Part A.	61

### Part B: Binding of the HNF1 Proteins to the Factor IX Promoter.

3.4	Introduction to Gel Shifts.	62
3.5	Gel Shifts with Rat Liver Nuclear Extracts.	64
3.5.1	Endogenous Rat Liver Proteins Bind to the Wildtype Factor IX Promoter.	64

3.5.2 HNF1 May be Interacting at the -5, -6 Region of the Factor IX Promoter.	66
3.5.3 An HNF1 Dimer May Bind at -5, -6 of the Factor IX Promoter.	68
<b>3.6 Cos Cell Extract Gel Shifts.</b>	68
3.6.1 Cos Cell Transfections.	68
3.6.2 Preliminary HNF1 $\beta$ Gel Shifts.	69
<i>The Results for HNF1<math>\beta</math> Homodimers differ from those for Rat Liver Nuclear Extracts.</i>	73
<i>Binding of HNF1<math>\beta</math> to the Wildtype Factor IX Promoter.</i>	73
3.6.3 Gel Shifts with Cos Cell Extracts Containing HNF1 $\alpha$ , HNF1 $\beta$ and DCoH.	76

**Part C: The Effect of HNF1 Proteins on Transcription from the Factor IX Promoter.**

<b>3.7 Introduction to Luciferase Assays.</b>	80
3.7.1 Potential Sources of Error in Luciferase Assays	82
<b>3.8 Transcription from the Factor IX Promoters.</b>	84
3.8.1 The Mutant Promoters are Down-regulated in Alexander Cells.	84
3.8.2 The Expression from Mutant Promoters Varies Between Patients.	86
<b>3.9 Effects of HNF1<math>\alpha</math>. on Factor IX Expression.</b>	88
3.9.1 HNF1 $\alpha$ Down-regulates Transcription from the Wildtype Factor IX Promoter <i>in vitro</i> .	88
3.9.2 Is HNF1 $\alpha$ a Transcriptional Repressor?	90
3.9.3 The Possible Function of Repression of the Normal Factor IX Promoter by HNF1 $\alpha$ .	93
<b>3.10 Effects of HNF1<math>\beta</math> on Factor IX Expression.</b>	94
3.10.1 HNF1 $\beta$ Down-regulates Transcription from the Wildtype Factor IX Reporter Gene.	94
3.10.2 HNF1 $\beta$ May Possess of Bifunctional Regulatory Activity.	96
<b>3.11 The Effects of HNF1<math>\alpha</math>.HNF1<math>\beta</math> Heterodimers on Factor IX Gene Expression.</b>	97
3.11.1 HNF1 $\alpha$ .HNF1 $\beta$ Heterodimers up-regulate the Wildtype and -6G/C Factor IX Promoters <i>in vitro</i> .	97
3.11.2 HNF1 Heterodimers Could Potentially Activate Factor IX Gene Expression.	99
<b>3.12 Effects of DCoH on the Activities of the HNF1 Proteins.</b>	100
3.12.1 The <i>in vitro</i> Regulatory Ability of HNF1 $\alpha$ is Essentially Unchanged	

by DCoH.	101
3.12.2 The <i>in vitro</i> Regulatory Ability of HNF1 $\beta$ May be Unchanged by DCoH.	102
3.12.3 DCoH has a Variable Effect on the <i>in vitro</i> Transactivating Ability of HNF1 Heterodimers in Luciferase Transfections.	104
<b>CHAPTER FOUR: FINAL DISCUSSION AND CONCLUSIONS.</b>	
<b>4.1 HNF1 May Bind the Factor IX Promoter and Regulate its Transcription <i>in vitro</i>.</b>	106
<b>4.2 HNF1 Could Form the Basis of the Liver-specific Expression of the Factor IX Gene.</b>	107
4.2.1 Is the Liver Protein Identified in Gel Shifts an HNF1 $\alpha$ .HNF1 $\beta$ Heterodimer?	111
<b>4.3 Future Work.</b>	112
<b>REFERENCES.</b>	116
<b>Appendix A: Plasmid Maps.</b>	124
<i>pBluescript-KS-HNF1<math>\alpha</math>.</i>	124
<i>pBJ5-HNF1<math>\beta</math>.</i>	125
<i>pGEX-2T-DCoH.</i>	126
<i>pSVK3.</i>	127
<b>Appendix B: Oligonucleotides</b>	128
<i>HNF1<math>\alpha</math> PCR primers.</i>	128
<i>HNF1<math>\alpha</math> Sequencing primers.</i>	128
<i>DCoH primers.</i>	128
<i>Gel Mobility Shift Oligonucleotides</i>	128
<b>Appendix C: Lalign Results.</b>	130

## List of Figures.

Figure		Page
1.1	The blood clotting cascade.	2
1.2	The imperfect repeat found in the promoter of the gene for factor IX.	3
1.3	Haemophilia B Leyden mutations.	5
1.4	Schematic diagram of the domains of HNF1 $\alpha$ .	9
1.5	The consensus sequence of HP1 to which HNF1 $\alpha$ binds.	10
1.6	Schematic diagram of the HNF1 $\alpha$ and HNF1 $\beta$ proteins.	12
1.7	Alignment of the HP1 site with the -5, -6 region of the Factor IX promoter.	16
3.1	Cloning strategy of the HNF1 $\alpha$ cDNA using restriction enzyme digests.	28
3.2	Partial digestion of pBluescript-KS-HNF1 $\alpha$ by <i>Eco</i> RI restriction endonuclease.	29
3.3	Digestion of the pBluescript vector with <i>Hga</i> I.	30
3.4	Identification of the HNF1 $\alpha$ cDNA by digestion with <i>Pst</i> I.	31
3.5	Identification of transformants carrying potential pSVK3-HNF1 $\alpha$ plasmids.	32
3.6	Identification of the pSVK3-HNF1 $\alpha$ plasmid by digestion with either <i>Pst</i> I or <i>Sma</i> I.	33
3.7	Identification of the orientation of the HNF1 $\alpha$ cDNA in the pSVK3-HNF1 $\alpha$ construct.	34
3.8	The pSVK3-HNF1 $\alpha$ construct with two inserts.	36
3.9	Cloning strategy for the HNF1 $\alpha$ cDNA using PCR.	38
3.10	Schematic diagram of the incorrect hybridisation of the 3' HNF1 $\alpha$ primer to the PCR template.	41
3.11	The 2.3 kb HNF1 $\alpha$ PCR product.	42
3.12	Identification of positive clones.	43
3.13	Identification of the correct HNF1 $\alpha$ expression construct.	44
3.14	The sequencing strategy for the HNF1 $\alpha$ cDNA.	46
3.15	The new HNF1 $\alpha$ cDNA sequencing strategy.	47
3.16	The internal HNF1 $\alpha$ PCR.	48
3.17	The potential loop of the HNF1 $\alpha$ cDNA.	51

3.18	Cloning strategy for the HNF1 $\beta$ cDNA.	53
3.19	<i>Eco</i> RI digest of pBJ5-HNF1 $\beta$ .	54
3.20	Identification of positive transformants.	55
3.21	<i>Sma</i> I diagnostic digests of the three positive plasmids.	56
3.22	Cloning of the DCoH PCR product.	57
3.23	Cloning strategy for the DCoH cDNA.	58
3.24	Sample of the manual sequencing of the DCoH cDNA.	60
3.25	Schematic diagram of a gel shift.	63
3.26	Competition gel shift 1 of rat liver nuclear extract with the wildtype factor IX promoter.	65
3.27	Competition gel shift 2 of rat liver nuclear extract with the wildtype factor IX promoter.	67
3.28	Competition gel shift 1 of HNF1 $\beta$ -rich Cos cell extract with the HNF1 $\alpha$ site from the $\alpha$ 1-antitrypsin promoter.	71
3.29	Competition gel shift 2 of HNF1 $\beta$ -rich Cos cell extract with the HNF1 $\alpha$ site from the $\alpha$ 1-antitrypsin promoter.	72
3.30	Alignment of the HNF1 $\alpha$ consensus sequence with the factor IX promoter.	74
3.31A	Gel shift to test for the presence of an HFN1 transcription factor.	77
3.31B	Gel shift to detect HNF1 $\alpha$ in a Cos cell extract.	78
3.32	BSA trial gel shift.	79
3.33	Schematic diagram of luciferase activities.	81
3.34	Basal transcription from the factor IX promoters.	85
3.35	Results of the co-transfection of HNF1 $\alpha$ with the factor IX promoters.	89
3.36	Quenching.	91
3.37	Results of the co-transfection of HNF1 $\beta$ with the factor IX promoter-reporter gene constructs.	95
3.38	Results of the co-transfection of HNF1 $\alpha$ and HNF1 $\beta$ with the factor IX promoters.	98
3.39	Results of the co-transfection of DCoH with HNF1 $\alpha$ and the factor IX promoters.	101
3.40	Results of the co-transfection of DCoH with HNF1 $\beta$ and the factor IX promoters.	103
3.41	Results of the co-transfection of DCoH, HNF1 $\alpha$ and HNF1 $\beta$ with the factor IX promoters.	105

4.1	The '3D puzzle' of the binding of transcription factors to various promoters.	110
-----	---	-----

**List of Tables.**

Table		Page
1	HNF1 $\alpha$ cDNA Amplification by PCR.	39
2	HNF1 $\alpha$ PCR Primer Test Results.	40
3	Results of the sequencing of the pSVK3-HNF1 $\alpha$ clone 2.	49
4	Comparison of PCR and Sequencing Conditions.	52
5	Sample Calculation of Relative Luciferase Activities.	84
6	The blood clotting abilities of the known -5 and -6 mutant factor IX promoter carriers.	87
7	Summary of the Effecto of HNF1 $\alpha$ and HNF1 $\beta$ on Transcription from the Factor IX Promoter.	108



## Abbreviations and Definitions.

A	adenine
AR	androgen receptor
ARE	androgen response element
ATP	adenosine triphosphate
bp	base pair
BRL	Bethesda Research Laboratories
C	cytosine
cDNA	complementary DNA
Cos7	monkey fibroblast kidney cell line
cpm	counts per minute
C/EBP	CCAAT enhancer binding protein
CTF	CAAT-binding transcription factor
DBP	D-site binding protein
DCoH	Dimerisation Cofactor of HNF1 $\alpha$
DMF	dimethylformamide
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DNAse I	deoxyribonuclease I
dNTPs	deoxynucleotide triphosphates
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylene diamine tetra-acetate
EMSA	electrophoretic mobility shift assay
G	guanine
GST	glutathione S-transferase
HEPES	N-2-HydroxyEthylPiperazine-N'-2-Ethane Sulphonic acid
HNF1 $\alpha$	hepatocyte nuclear factor 1 $\alpha$
HNF1 $\beta$	hepatocyte nuclear factor 1 $\beta$
HNF4	hepatocyte nuclear factor 4
kb	kilo base
kDa	kilodalton
LF-A1	liver factor A1
mRNA	messenger ribonucleic acid

NF-1	nuclear factor 1
N terminal	amino terminal
ONPG	<i>o</i> -nitrophenyl- $\beta$ -galactoside
PAGE	PolyAcrylamide Gel Electrophoresis
PBS	phosphate-buffered saline
PBSE	phosphate-buffered saline EDTA
PCR	polymerase chain reaction
<i>Pfu</i>	<i>Pyrococcus furiosus</i>
PMSF	phenyl methane sulphonyl fluoride
poly (dI-dC)	polymer of dI and dC
<i>Pwo</i>	<i>Pyrococcus woesei</i>
RNA	ribonucleic acid
RNAase	ribonuclease
RT	room temperature
<i>S. cerevisiae</i>	yeast <i>Saccharomyces cerevisiae</i>
SDS	sodium dodecyl sulphate
T	thymine
TAE	tris acetate EDTA
<i>Taq</i>	<i>Thermus aquaticus</i>
TBE	tris boric acid EDTA
TEMED	N, N, N', N' - tetra methyl ethylenediamine
TFIID	general transcription factor IID
TFIIH	general transcription factor IIH
Tris	tris- (hydroxymethyl) aminomethane
tRNA	transfer ribonucleic acid
UV	ultra violet light
<sup>32</sup> P	radioactive isotope of phosphate

## CHAPTER 1: INTRODUCTION

The liver is one of the most vital organs of the human body for the synthesis of proteins. Many important enzymes of the carbohydrate, lipid and protein catabolic pathways originate from genes active mainly in hepatocytes, although it is not uncommon for low levels to be contributed from other organs. The serine proteases of the blood clotting cascade, such as factor IX, are almost exclusively produced by the liver through constitutive expression of their genes. Such an abundance of transcriptional activity requires the strict control of gene activation and repression. The hepatocyte achieves this regulation through the interactions of ubiquitous, liver-specific and developmentally regulated transcription factors, the combination of which is unique to each promoter. Hepatocyte Nuclear Factors, HNF1 $\alpha$  and HNF1 $\beta$ , are known to affect the transcription rate of a variety of genes, the list of which is lengthy and no doubt currently incomplete. The addition of the factor IX gene to this list has been suggested, not only for the obvious identical tissue specificity, but also for the sequence similarity of a region of the factor IX promoter with the recognition site for the HNF1 proteins.

### 1.1 Factor IX

Haemophilia B, or Christmas disease, is an X-linked inherited bleeding disorder<sup>1</sup> resulting from an absence of functional blood coagulation factor IX in the blood stream. The vitamin K-dependent protein is involved in the middle phase of the blood clotting cascade (Figure 1.1), which results in the activation of fibrin and formation of a clot<sup>2</sup>. The sequential activation of specific plasma proenzymes to the active form may be initiated intrinsically or extrinsically<sup>2</sup>. Damage to blood vessels, or contact with surfaces such as collagen or glass, will begin the process of limited proteolysis of zymogens<sup>2</sup>. Factor IX is activated in this way by active factor VII (factor VII<sub>a</sub>), and goes on to interact with Factor VIII<sub>a</sub>, phospholipids and calcium ions to form a complex responsible for activating Factor X<sup>3</sup>.

Haemophilia B represents 10-12% of all haemophilias<sup>2</sup>, and results from defective factor IX protein, or from mutations in the promoter region of the factor IX gene which prevent its expression.

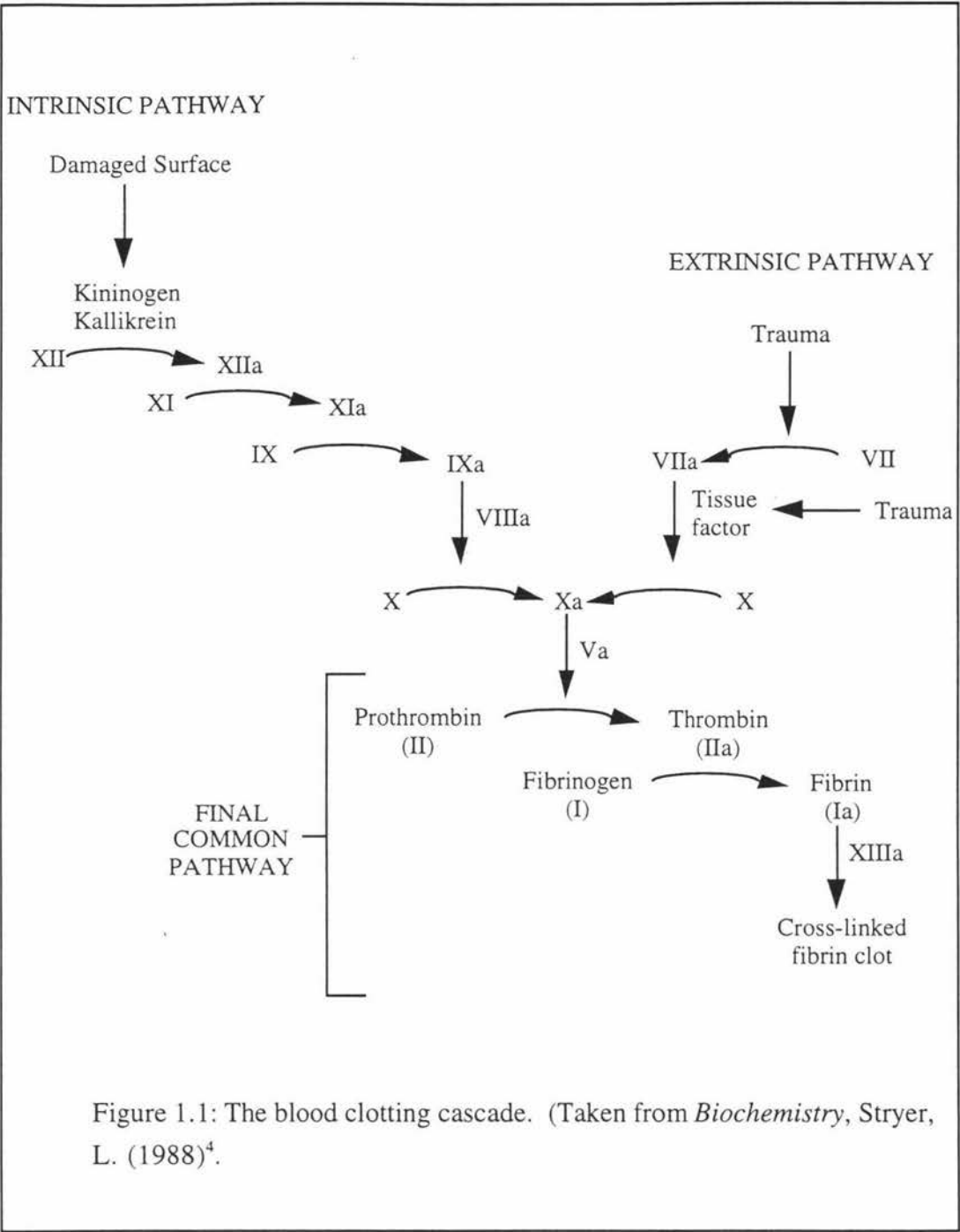


Figure 1.1: The blood clotting cascade. (Taken from *Biochemistry*, Stryer, L. (1988)<sup>4</sup>.

1.1.1 Haemophilia B Leyden

First identified in 1970 by Veltkamp *et al.*, Haemophilia B Leyden is a rare form of haemophilia, usually resulting from single base substitutions in the factor IX promoter<sup>1</sup>. The condition therefore extends from a reduced expression of the gene and consequently very low levels of functional protein<sup>1</sup>. The Leyden form of the disease is characterised by severe haemophilia during childhood (typically <1% of

normal factor IX levels), leading to a mild or asymptomatic condition following puberty (upwards of 50% of normal levels)<sup>1</sup>. The initial Haemophilia B Leyden patient had a T to A mutation at position -20 of the promoter<sup>6</sup>. Related Haemophilia B Leyden mutations, often with varying degrees of phenotypic severity, include -6 G/A<sup>1,5</sup>, +13 A/G<sup>6</sup>, +8 T/C<sup>3</sup>, and -21 T/G<sup>7</sup>. Interestingly, all known factor IX promoter mutations lie within an imperfect direct repeat found in the 5' untranslated region (Figure 1.2), thought to be important for normal factor IX transcription<sup>23</sup>.

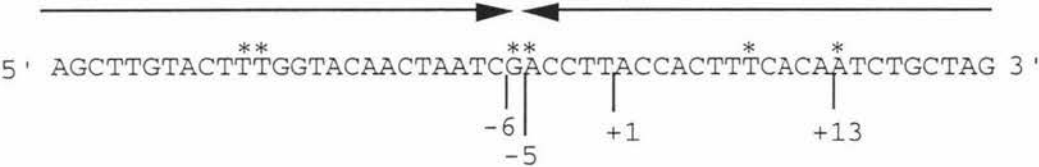


Figure 1.2: The imperfect direct repeat found in the promoter of the gene for factor IX. \* Indicates a known mutation site. (Adapted from Royle *et al.*, (1991)<sup>3</sup>).

1.1.2 Transcription in Eukaryotes

The initiation of transcription in eukaryotes is regulated by interactions between RNA polymerase and transcription factors that recognise *cis*-acting control regions within the DNA sequences surrounding the gene<sup>8, 9</sup>. The average nucleotide sequence recognised by a particular transcription factor is known as the consensus binding site. Deviations from this sequence can decrease the binding interaction of the transcription factor and reduce the frequency of transcription initiation from the associated gene. This is true for alterations in consensus binding sites at disease loci such as the gene for factor IX. The binding of transcription factors therefore is an essential step in activation of eukaryotic genes<sup>9</sup>.

In general, the simultaneous binding of more than one type of transcription factor is required for proper regulation of transcription in eukaryotes<sup>8</sup>. A relatively small number of different transcription factors in alternative combinations can therefore establish highly specific and precisely controlled patterns of gene expression<sup>10</sup>. A gene could be expressed when a particular set of transcription factors produced by the cell corresponds to the cluster of *cis*-acting regulatory DNA

sequence elements near the gene. This hypothesis is the basis of temporal (developmental) and spatial (tissue-specific) control. Expression of genes that encode tissue-specific products would therefore depend on the presence of unique transcription factors or unique combinations of transcription factors in that cell type<sup>9</sup>.

### *Promoters and Enhancers*

The upstream promoter region may contain several types of DNA sequence elements that are found in similar relative locations in a large number of different genes. These elements may include the TATA box, the CAAT box or a GC-rich element<sup>57</sup>. The ubiquitous transcription factors TFIID, CTF and Sp1, respectively bind specifically to these sequences<sup>9</sup>. These factors must bind to the promoter before the RNA polymerase-promoter interaction can occur. The fact that many different genes, each with a unique pattern of expression, utilise the same components in the upstream promoter region suggests that this region is important for basal promoter function<sup>9</sup>.

Enhancers, rather than the upstream promoter region, are the basis of differential gene expression<sup>8-10</sup>. Enhancer elements vary in specific sequence, but share the common properties of regulating gene expression in a manner independent of orientation and proximity to the promoter itself<sup>8-10</sup>. However, individual enhancers differ in their nucleotide consensus sequence, reflecting the fact that different transcription factors bind to different types of enhancers. Despite their name, enhancers which are negative regulatory elements have also been characterised<sup>9</sup>.

### *The Action of Transcription Factors*

A multitude of transcription factors have been identified, representing a wide range of regulatory potentials, yet some common features exist. Transcription factors possess several distinct functional domains, including the DNA-binding domain which is generally distinct from the transcription activation domain<sup>11</sup>. The mechanism through which transcription factors influence gene expression can be explained by one of two models. One model proposes that the binding of an activating transcription factor to an enhancer triggers an alteration in chromatin structure that can be propagated over some distance along the chromosome, resulting in the activation of nearby genes<sup>12</sup>. The second model suggests that a looping mechanism may bring two linearly distant transcription factor binding sites into close proximity<sup>8</sup>. In such a configuration, proteins bound at two separated sites could be close enough to allow protein-protein interactions. Together the complex of

transcription factors might produce a recognition target for RNA polymerase interaction and transcription initiation<sup>8,9</sup>. The bending or looping could even occur as a natural consequence of the coiling of the DNA backbone around a nucleosome or in higher-order chromatin structures. The two models may even work in conjunction, where one transcription factor changes the chromatin configuration, allowing the association of two enhancers and subsequent recruitment of RNA polymerase to the transcription start point.

1.1.3 Liver-specific Expression of Factor IX

A variety of proteins have been shown to be involved in the initiation of transcription from the factor IX promoter (Figure 1.3).

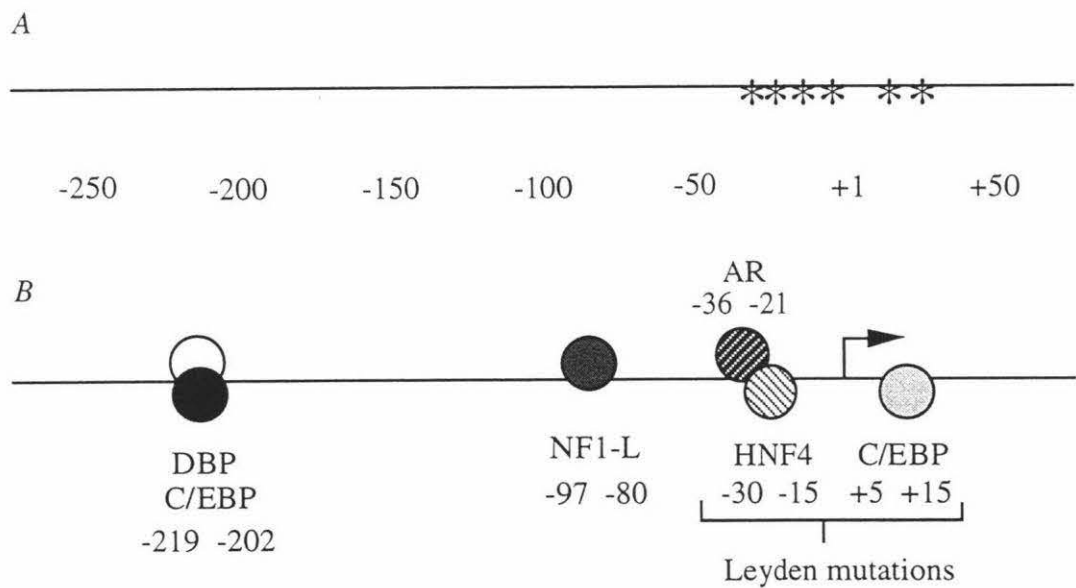


Figure 1.3: Haemophilia B Leyden mutations of the human blood clotting factor IX promoter and its transcriptional regulators. A: Positions of known mutations of the factor IX promoter. B: Relative positions of the transcription factors known to interact with the promoter.

One or some of these factors must account for the liver-specific expression of factor IX, while others are responsible for the constitutive expression of the factor IX gene.

To date, liver-specific transcription factors and enhancers playing a part in factor IX expression are nuclear factor 1-liver (NF1-L) and hepatocyte nuclear factor 4 (HNF4), also known as LF-A1<sup>13</sup>. In transient transfections, HNF4 was shown to activate the wild-type promoter in cells of both hepatic and non-hepatic origin to the same degree<sup>14</sup>. If this observation extends to an *in vivo* situation, HNF4 may produce the hepatocyte-enriched expression pattern of the factor IX gene.

#### 1.1.4 Haemophilia B Leyden and Phenotypic Recovery

Haemophilia B Leyden is characterised by the phenotypic recovery of patients following puberty. The promoter region of the factor IX gene must therefore contain a recognition site for a transcription factor that is activated at puberty. However, mutation of this DNA sequence would not produce a Leyden-like phenotype. Instead, such a mutation would result in a form of Haemophilia B where the patient does not undergo phenotypic recovery. Mutations resulting in Haemophilia B Leyden therefore cause deviations in the DNA binding sites of other transcription factors. HNF4 and C/EBP have had their roles in factor IX expression confirmed by their inability to function on the factor IX promoter containing -20 and +13 mutations (respectively). In most cases, a single base substitution is sufficient in preventing the formation of protein-DNA complexes. The absence of a specific protein from a transcription initiation complex results in loss of transcription of the gene.

At puberty, strong activating factors may bind to enhancer-like sequences in the factor IX Leyden mutant promoter. The resulting transcription initiation complex in some manner compensates for the mutation and induces a high level of gene expression. Three theories currently attempt to explain the compensatory recovery, although it is likely that no one hypothesis can fully account for the observed phenomenon. Two groups, lead by Reitsma<sup>7</sup> and Brownlee<sup>15</sup>, base their theories on the androgen receptor, a transcription enhancing protein activated in the presence of testosterone. Picketts and associates<sup>16</sup> however, cite the involvement of a transcription factor whose expression is induced only following puberty.

Picketts *et al.* suggest that synergy between CCAAT/enhancer binding protein (C/EBP) and D-site binding protein (DBP) may be responsible for the post-pubertal recovery from Haemophilia B Leyden<sup>16</sup>. C/EBP is a ubiquitous transcription factor implicated in the expression of many genes<sup>6</sup>. DBP is a transcription factor expressed in a developmental fashion in rats. The DNA recognition site for DBP overlies that of C/EBP on the factor IX promoter (position -219), and has been earmarked for a role in the developmental expression seen in Haemophilia B Leyden<sup>16</sup>. Picketts *et al.* observed that DBP produced around puberty in rats has no effect on factor IX



transcription, so it cannot act as a C/EBP substitute<sup>16</sup>. However, synergistic interactions between the two proteins can generate an increased activation of a mutated promoter up to wild-type levels<sup>16</sup>. DBP is thought to rescue the Haemophilia B Leyden phenotype by changing a weak site into a high affinity site, through its interaction with C/EBP<sup>16</sup>. However, the site at which DBP and C/EBP bind (-219) has not been identified as a natural promoter mutation. It is not known whether a mutation within this site causes Haemophilia B *in vivo*. In addition, DBP has not been shown to be up-regulated in humans at puberty.

Testosterone is a steroid hormone secreted into the human bloodstream at the commencement of puberty. Its presence affects the expression of many genes by interacting with and activating a DNA binding protein known as the androgen receptor (AR). An androgen response element (ARE) to which the transcription factor binds is located within the factor IX promoter sequence, slightly upstream of the HNF4 site<sup>15</sup>. It has been proposed that the AR may allow co-operative binding of HNF4 to the promoter<sup>7</sup>. The DNA-binding activity of the AR at puberty may therefore allow HNF4 binding despite the -20 mutation. HNF4 and AR belong to the nuclear receptor superfamily, and the majority of these proteins function as dimers. Additionally, HNF4 is an orphan receptor, where the ligand responsible for facilitating binding of the transcription factor to DNA has not been identified. HNF4 and AR may therefore dimerise to activate the factor IX promoter with a -20 mutation. Alternatively, the AR may sufficiently activate transcription in the absence of HNF4. It should be noted that a base substitution at position -26 results in Haemophilia B Brandenburg, in which the patient does not recover after puberty, bearing witness to the involvement of the AR in the developmental transcription of factor IX<sup>15</sup>. The evidence presented here suggests that the phenotypic recovery observed in Haemophilia B Leyden patients is due to the actions of the AR during puberty.

The study of mutations in the factor IX promoter has therefore identified several proteins involved in the transcriptional regulation of the gene. However, mutations at -5 and -6 (Figure 1.3) are yet to be characterised with respect to their cognate DNA-binding protein/s. A possible candidate for binding to this region is HNF1.

## 1.2 HNF1 Family

HNF1 $\alpha$  is a predominantly liver-enriched transcription factor involved in the expression of many hepatocyte-specific genes such as albumin and  $\alpha$ 1-antitrypsin.

The protein acts as a classical transcription factor at the promoter of these genes and is involved in the activation of gene expression. Recently a second function has been suggested for HNF1 $\alpha$ ; it has been identified as a chromatin remodelling factor of the phenylalanine hydroxylase gene<sup>18</sup>.

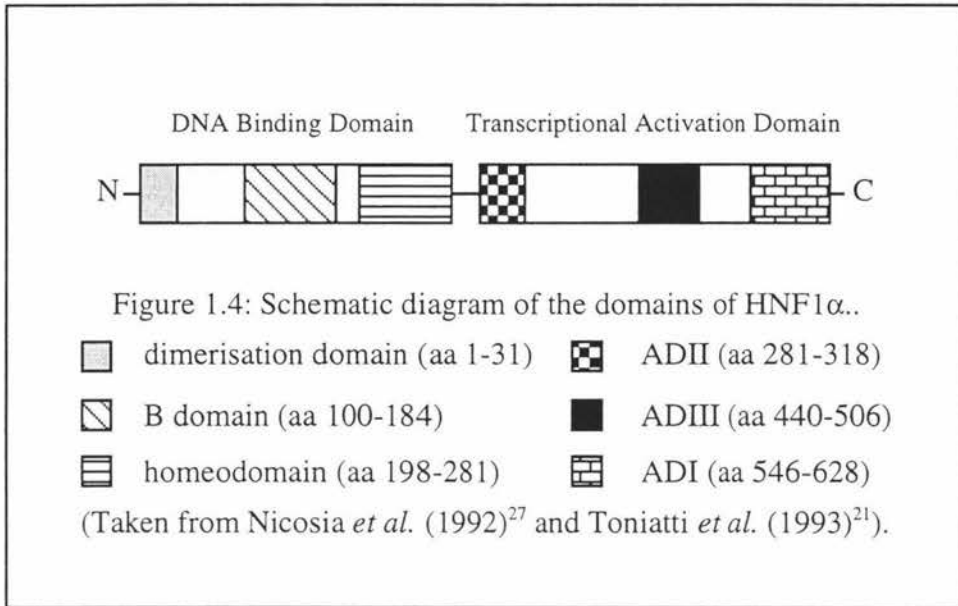
HNF1 $\alpha$  is a member of the homeoprotein family, a group of transcription factors which contain a conserved DNA-binding domain called the homeodomain. As the most variant homeoprotein isolated to date, HNF1 $\alpha$  contains a number of functional domains, in addition to its homeodomain. One feature of HNF1 $\alpha$  which makes it unique amongst the homeoprotein family is its strict requirement for dimerisation prior to DNA binding<sup>19</sup>, which is mediated by a dimerisation domain. The ability of HNF1 $\alpha$  to dimerise in the absence or presence of DNA<sup>19</sup> suggests that its regulatory function may be diversified by heterodimerisation<sup>13</sup>. The second member of the HNF1 family, HNF1 $\beta$ , is highly homologous to HNF1 $\alpha$  in its DNA binding and dimerisation domains and has been confirmed as a dimerisation partner for HNF1 $\alpha$ <sup>20</sup>. HNF1 $\alpha$  and HNF1 $\beta$  vary within their C terminal, transcriptional activation domains. A number of putative activation domains (ADs) have been identified within this region through sequence analysis and site-directed mutagenesis<sup>21</sup>, but one domain solely responsible for transactivation has not yet been identified. In addition, the genes encoding HNF1 $\alpha$  and HNF1 $\beta$  may be alternatively spliced, generating three isoforms of each HNF1 protein<sup>22</sup>. The isoforms vary essentially within the C terminal activation domain, so that a variety of transactivating dimer activities is possible<sup>22</sup>.

A relatively new factor with a role in HNF1 dimerisation is DCoH (Dimerisation Cofactor of HNF1)<sup>23</sup>. This bifunctional protein has a stabilising activity on HNF1 $\alpha$  homodimers, such that an absence of DCoH cause dimers to dissociate<sup>23</sup>. *In vivo*, DCoH is expressed in a number of tissues, including those expressing HNF1 $\alpha$  and HNF1 $\beta$ . All HNF1 dimers formed may therefore be selectively stabilised by DCoH<sup>23</sup>.

### 1.2.1 DNA Binding Domains

The homeodomain of HNF1 $\alpha$  (Figure 1.4) identifies it as the most divergent homeoprotein classified to date. The conserved amino acids of the HNF1 homeodomain (17%) is mainly restricted to the domain's third helix<sup>19</sup>. However, inclusion of a 21 amino acid loop<sup>24</sup> in place of the usually universal turn between helices II and III increases the homology to a more significant 23%, spread throughout the domain<sup>19</sup>. HNF1 $\alpha$  also lacks several key amino acids found in

homeoproteins<sup>25</sup>, whose replacements do not bear sufficient similarity to maintain the original structural interactions. The presence of an intron between the exons coding for helices II and III of the gene for HNF1 $\alpha$  adds to the non-homologous homeodomain features<sup>24</sup>. Despite the lack of conserved amino acids in the HNF1 $\alpha$  homeodomain, the crystal structure has been shown to be similar to that of the *Drosophila Antp* homeodomain<sup>26</sup>.



In traditional homeodomains, DNA binding places the recognition helix (helix II) into the major groove<sup>26</sup>, which enhances transcription. However, the isolated homeodomain of HNF1 $\alpha$  is not sufficient for specific recognition of the target sequence<sup>28</sup>. The B domain is the second DNA-binding domain of HNF1 $\alpha$ <sup>19</sup>. It contains weak homologies to the POU domain, a conserved DNA-binding domain present in a family of eukaryotic transcription factors which play a role in development<sup>29</sup>. Mutations in either DNA-binding domain have been found to inhibit wild-type DNA binding activity<sup>27</sup>, so both domains must play a role in DNA binding site recognition. The B domain and homeodomain lie adjacent to each other within the N terminal half of the protein (Figure 1.4). This close proximity must therefore enable the domains to work concomitantly to bind the HNF1 $\alpha$  enhancer sequence. This site (Figure 1.5), referred to as the HP1 site by Strandmann and Ryffel<sup>30</sup>, is palindromic by consensus. However no individual endogenous site has been identified with the high level of symmetry displayed here<sup>19</sup>.

g/a G T T A A T N A T T A A C c/a

Figure 1.5: The consensus sequence of HP1 to which HNF1 $\alpha$  binds.  
(Taken from Chouard *et al.* (1990)<sup>19</sup>).

By definition, HP1 is an enhancer element, containing the necessary information to drive transcription when present in either orientation, alone or multimerised<sup>19</sup>. This suggests that HNF1 $\alpha$  and HNF1 $\beta$  can function in either orientation.

HNF1 $\alpha$  is thought to interact with the element in two places through its two binding domains. The homeodomain contacts the TAAT core of the site, while the GT immediately 5' accommodates residues from the B domain<sup>28</sup>. A repeat of these well-conserved nucleotides receives the second partner of the dimer.

### 1.2.2 Dimerisation Domain

A third motif of the HNF1 family has been identified as essential for the DNA binding ability of these transcription factors. This dimerisation domain appears to be  $\alpha$ -helical, and interactions with a second domain are presumed to occur through a coiled coil<sup>19</sup>. Dimerisation through this domain has allowed a diversification in the regulatory mechanisms available for the HP1 site<sup>13</sup>. Formation of heterodimers could form a continuum of transcriptional activity, potentially producing both inducers and repressors. Slight changes in concentration of either member of a dimer can provoke a large change in the effect the dimer has on transcription<sup>13</sup>. Dimerisation therefore increases the repertoire of options available, hence its importance for HNF1 $\alpha$  activity. Indeed, each HNF1 $\alpha$  monomer must have functional dimerisation, B and homeodomains in order to bind to its target<sup>27</sup>.

### 1.2.3 HNF1 $\beta$

HNF1 $\beta$  is a transcription factor with a remarkable similarity to HNF1 $\alpha$ . Despite being 71 amino acids smaller (HNF1 $\alpha$  = 628 aa, HNF1 $\beta$  = 557 aa<sup>21</sup>), the

homology of HNF1 $\beta$  to HNF1 $\alpha$ , particularly within the N terminal region, is striking. Of the three vital domains for DNA binding, HNF1 $\alpha$  and HNF1 $\beta$  have a large percentage of amino acid homology: homeodomain, 91%; B domain, 82%; dimerisation domain, 79%<sup>21</sup>.

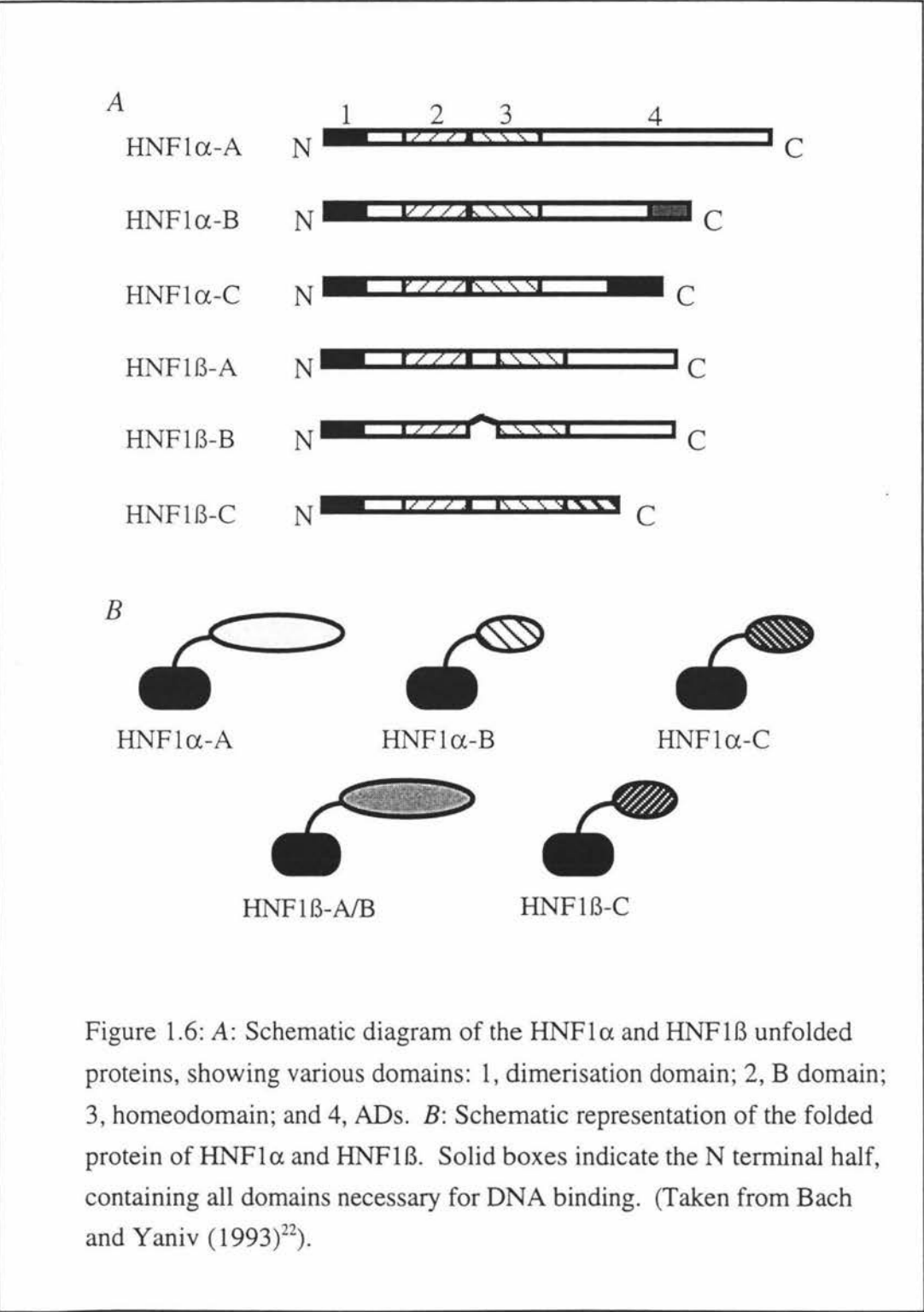
Of the variant amino acids of the homeodomain in HNF1 $\beta$ , none induce changes in secondary structure<sup>20</sup>. The primary difference between HNF1 $\alpha$  and HNF1 $\beta$  is a 26 amino acid stretch between the B and homeodomains usually absent from HNF1 $\beta$ <sup>20</sup> (see later). However this has no effect on the DNA binding capabilities of HNF1 $\beta$ . Both homodimers interact with similar consensus sequence elements<sup>20</sup>, at the albumin and fibrinogen promoters they even contact the same nucleotides<sup>13</sup>. The homology retained in the dimerisation domain allows heterodimer formation *in vitro* and *in vivo*<sup>20</sup>.

#### 1.2.4 Activation Domains (ADs)

Conservation of sequence between the two forms of HNF1 breaks down within the C terminal, transcriptional activation domain. In fact, there is very little agreement on the assignment of activating motifs within this region, but this may be due to the study of different isoforms (see later). The domain is known to be rich in serine and threonine<sup>2</sup>, and glutamine and proline<sup>30</sup>, and the transcriptional function has been mapped into three areas within this region<sup>30</sup>. From N terminal to C terminal the activation domains (ADs) are numbered: ADII, ADIII, and ADI (Figure 1.4). Toniatti *et al.* (1993) have undertaken extensive work on the ADs, noting that ADI is highly enriched in serine (26%) and ADIII is 46.2% glutamine, proline and serine<sup>21</sup>. Neither ADI or ADII show any homology to known activation domains. ADIII however, has been compared to the activation domain of VP16<sup>25</sup>, a potent herpes simplex virus transcriptional activator<sup>31</sup>. ADIII is highly active and capable of enhancing transcription when linked to an unrelated DNA binding domain<sup>21</sup>. Furthermore, the highest homology between HNF1 $\alpha$  and HNF1 $\beta$  lies in ADIII (61%)<sup>21</sup>, while HNF1 $\beta$  lacks ADI<sup>32</sup>. ADIII is thought to function by attracting adapter proteins or contacting other factors of the transcription machinery<sup>32</sup>. On another tangent, Mendel *et al.* (1991) believe that the B domain could have a secondary function as an AD by interacting with other POU-domain proteins<sup>32</sup>.

1.2.5 Isoforms of HNF1 $\alpha$  and HNF1 $\beta$

Sequence analysis of the HNF1 $\alpha$  and HNF1 $\beta$  genes shows three alternative polyadenylation sites<sup>19</sup>. Utilisation of these sequences generates the HNF1 $\alpha$  and HNF1 $\beta$  isoforms: HNF1 $\alpha$ -A, HNF1 $\alpha$ -B, HNF1 $\alpha$ -C, HNF1 $\beta$ -A, HNF1 $\beta$ -B and HNF1 $\beta$ -C (Figure 1.6).





Generally, each alternatively spliced isoform of HNF1 $\alpha$  has a different C terminal AD<sup>22</sup>. HNF1 $\beta$ -A gains the 71 nucleotide insertion between the B and homeodomains that is present in HNF1 $\alpha$ <sup>33</sup>. HNF1 $\alpha$ -A, HNF1 $\alpha$ -C, HNF1 $\beta$ -A/B\* and HNF1 $\beta$ -C mRNAs are generated by differential use of polyadenylation sites<sup>22</sup>. However the polyadenylation sites of HNF1 $\alpha$ -B/C and HNF1 $\beta$ -C are divergent and are expected to be very weak<sup>22</sup>. Therefore the levels of these isoforms may well be low. Polyadenylation and alternative splicing are thought to be differentially regulated, observed as differing isoform concentrations in adult and foetal liver, thymus, kidney and intestine cells<sup>22</sup>. Only HNF1 $\beta$  isoforms are present in foetal lung tissue. The varying levels of each protein between tissues prompts speculation on a developmental role for HNF1 $\alpha$  and HNF1 $\beta$ . This hypothesis is strengthened by the observation that the different homodimers have different transactivating activities. Most strikingly, HNF1 $\beta$ -C is unable to promote transcription, and acts as a repressor when coupled with any HNF1 $\alpha$  isoform. The inactivity of such a heterodimer may be due to an interference of the unique C terminal domain of HNF1 $\beta$ -C, or because a single functional transactivation domain is insufficient for activity<sup>22</sup>.

Surprisingly, the activating ability of HNF1 $\alpha$ -B and HNF1 $\alpha$ -C is significantly greater than that of HNF1 $\alpha$ -A, possibly due to their capability to contact a different protein, or different part of a protein, of the transcriptional machinery<sup>22</sup>. Similarly, in transient transfections HNF1 $\beta$ -A is a stronger transactivator than HNF1 $\beta$ -B due to a higher DNA binding affinity. The region immediately upstream of the homeodomain is absent in HNF1 $\beta$ -B and so may modulate DNA-protein interactions and transactivation<sup>34</sup>. These extra 26 amino acids, also absent in HNF1 $\alpha$ , may produce a subtle conformational change that directly alters the homodimer's function. A positive interaction with the preinitiation complex may be generated, or cofactor binding permitted<sup>34</sup>. HNF1 $\beta$ -A and HNF1 $\beta$ -B are found in equal ratios in every HNF1 $\beta$ -containing tissue, and so HNF1 $\beta$ -B may modulate the function of HNF1 $\beta$ -A by forming heterodimers or by occupying the DNA site, preventing HNF1 $\beta$ -A homodimers from doing so<sup>34</sup>. It is likely that other isoform combinations will effect transcription in a similar way.

Dimerisation in conjunction with alternative splicing and different polyadenylation sites increases the type of DNA-protein complexes possible at one DNA site. If all the HNF1 $\alpha$  and HNF1 $\beta$  isoforms are produced in an absence of dimer selection, there are 21 possible dimer combinations<sup>22</sup>. A small change in the concentration of one isoform may generate a large change in specific gene transcription, thereby increasing the limits and sensitivity of the system. The presence of isoforms may explain the observation that the *in vivo* distribution of HNF1 $\alpha$  does not correspond to the hepatic phenotype, for which it is partially responsible<sup>20</sup>. For example, HNF1 $\alpha$  alone cannot force high albumin expression

levels in non-hepatic tissues<sup>35</sup>, yet both HNF1 $\alpha$  mRNA and protein are present at variable levels in the kidney, intestine and stomach<sup>25</sup>. HNF1 $\alpha$  is absent from dedifferentiated hepatoma cells<sup>20</sup>, a distribution possibly explained by the presence of HNF1 $\beta$  in these cell lines<sup>20</sup>. It is also interesting to note that in all cells where HNF1 $\alpha$  is induced or naturally present, DCoH mRNA is observed. Interplay between all three of these components may limit expression of HNF1 target genes to the liver.

### 1.2.6 DCoH

In 1991, Mendel and co-workers isolated a 104 amino acid protein, which did not appear to bind DNA, but selectively stabilised HNF1 $\alpha$  homodimers<sup>23</sup>. This dimerisation cofactor (DCoH) does not induce a change in DNA binding, but enhances the transcriptional activity of HNF1 $\alpha$ . Sequence requirements for HNF1 recognition are relaxed by DCoH, such that HNF1 may bind to sequences which deviate considerably from the consensus palindrome<sup>36</sup>. No evidence has been found that DCoH interacts directly with components of the preinitiation complex<sup>23</sup>. However, the remarkable resemblance between the structures of DCoH and the TATA binding protein has raised doubts as to the inability of DCoH to bind DNA<sup>37</sup>. Both HNF1 $\alpha$  and HNF1 $\beta$  homodimers are stabilised by the interaction of DCoH with their dimerisation domains. Mendel *et al.* attribute the stability of HNF1 $\alpha$  homodimers prepared from liver nuclear extracts to the presence of DCoH<sup>23</sup>. HNF1 $\alpha$  expressed from transient transfections exists in an equilibrium of dimers, monomers and dimer-DNA complexes<sup>27</sup>, so inclusion of DCoH in transfections is thought to increase the regulatory ability of HNF1 $\alpha$  through stabilisation of the dimer structure.

*In vivo*, DCoH mRNA is most abundant in kidney, followed closely by liver, lung, ovary and brain<sup>23</sup>. The wide distribution pattern of DCoH was surprising until a second function for the protein was discovered. Citron *et al.* (1992) found DCoH to be identical to the cytosolic enzyme pterin-4a-carbinolamine dehydratase (PCD)<sup>38</sup>. A member of the phenylalanine hydroxylation enzyme system, PCD is deficient in children with a rare, mild form of hyperphenylalaninemia<sup>37</sup>. Evidence to support this outstanding claim includes the almost identical tissue specificity of both proteins and an essentially equal dehydratase activity of recombinant DCoH compared to that of purified rat liver dehydratase. Even the specific activities are similar. Remarkably, as the active site does not overlap with the proposed HNF1-interacting surfaces<sup>37</sup>, DCoH/PCD is expected to retain its enzymatic activity when complexed with HNF1<sup>36</sup>.



The bifunctional ability of DCoH/PCD is an unusual feature of a protein involved in the initiation of transcription. One well-documented example however, is the general transcription factor TFIID, which was first identified for its role in the activation of basal transcription at promoters transcribed by RNA polymerase II<sup>39</sup>. More recently, TFIID has also been cited as part of the cell's DNA repair machinery<sup>40</sup>. TFIID possesses a DNA-dependent ATPase activity<sup>39</sup>, which is important for the DNA unwinding<sup>41</sup> and promoter clearance<sup>42</sup> phases of the initial transcription by RNA polymerase II. The helicase activity of TFIID is also implicated in nucleotide excision repair, the process by which DNA damage is removed as part of a single-stranded oligonucleotide fragment and replaced by subsequent DNA synthesis<sup>40</sup>. Although unusual, a double activity for a protein involved in transcription, is therefore not unprecedented.

### 1.3 The HNF1 Family and Haemophilia B

It is feasible that HNF1 $\alpha$  and HNF1 $\beta$  may play a role in the transcription of factor IX, due to the almost exclusive hepatocytic expression of all three proteins. HNF1 $\alpha$  is intimately involved in the transactivation of liver-specific genes such as  $\alpha$ 1-antitrypsin, albumin, and fibrinogen. In fact, with the exception of a TATA box, the binding site for the HNF1 proteins is the sole upstream element required for activation of the albumin promoter *in vitro*. In an *in vivo* state, albumin is highly active in only hepatoma cells<sup>35</sup>. Although factor IX is produced at extremely low levels in other cell lines, the liver is its primary site of production<sup>43</sup>, making liver-enriched HNF1 $\alpha$  and HNF1 $\beta$  perfect candidates for maintaining this expression pattern.

With many groups now singling out the HNF1 proteins as factors involved in liver organogenesis, a role for the proteins in the developmental recovery of Haemophilia B Leyden may be proposed. As shown in Figure 1.3, a DNA-binding protein has yet to be satisfactorily assigned to positions -5 and -6, where natural promoter mutations occur. Recently, Brownlee and Naka suggested that HNF4 binds within the -15 to +3 region of the wildtype factor IX promoter *in vitro*<sup>44</sup>. Additionally, CAT reporter gene assays indicate that HNF4 has the ability to activate transcription from this region of the factor IX promoter. However, these authors add that despite the noted binding and activating capabilities of HNF4 at the site *in vitro*, auxiliary, unidentified factors present in the rat liver nuclear extracts also bound to the promoter in this region.

A sequence alignment of the -5, -6 region with the consensus for the DNA-binding site of HNF1 $\alpha$  and HNF1 $\beta$  (Figure 1.7), indicates only weak identity. However, it is not unusual for transcription factors to exhibit relaxed sequence specificity, and often bind with similar affinities to sites on DNA that bear only limited resemblance at the nucleotide level. For example, there are six sites which have been inferred to represent bonafide binding sites for C/EBP, but no two of these six sites are identical<sup>45</sup>. Previous experiments have suggested that HNF1 $\beta$  homodimers down-regulate factor IX transcription<sup>46</sup>, but it is not known whether HNF1 $\alpha$  homodimers or HNF1 $\alpha$ -HNF1 $\beta$  heterodimers have any role in factor IX transcription.



Figure 1.7: Alignment of the HP1 site (top) with a region of the factor IX promoter (bottom). Bases important for HNF1 binding to HP1, and those bases matching in the factor IX promoter are shown in bold.

Other transcription factors have been shown to interact with HNF1 $\alpha$  on different promoters. The insulin-like growth factor binding protein 1 (IGFBP1) promoter achieves optimal activity through the interaction of HNF1 $\alpha$  with DBP<sup>6</sup>. It has also been claimed that HNF1 $\alpha$  and C/EBP expression vectors can synergistically activate albumin promoter-luciferase constructs in HeLa cells<sup>12</sup>. Both of these findings strengthen the suggestion that HNF1 $\alpha$  could transactivate the factor IX promoter, because of the possible involvement of C/EBP and DBP in factor IX transcription.

### 1.3.1 Aims of this Thesis

The primary goal of this research work was to determine whether an HNF1 protein was involved in transcription from the factor IX promoter. Three approaches were taken. Firstly the cDNAs for the HNF1 proteins and for DCoH were subcloned

into mammalian expression vectors for use in the second two areas of research. *In vitro* DNA-binding assays were then carried out to assess the ability of the -5, -6 region of the human factor IX promoter to specifically bind these proteins. *In vivo* transient co-transfection assays were carried out in hepatoma cells to assess the ability of normal and mutant factor IX promoters to respond to HNF1.

- The cDNAs for HNF1 $\alpha$ -A, HNF1 $\beta$ -A and DCoH were subcloned into mammalian expression vectors.
- These constructs were transfected into Cos 7 cells in an attempt to produce whole cell extracts enriched in HNF1 homodimers or heterodimers for use in mobility gel shifts. The ability of endogenous liver proteins from adult rat liver nuclei to form specific complexes with the factor IX promoter was also investigated using gel mobility shifts.
- The luciferase reporter gene assay system was used to study the effects of the transcription factors on transcription from the factor IX promoter.

For the remainder of this thesis, HNF1 $\alpha$ -A and HNF1 $\beta$ -A will be referred to simply as HNF1 $\alpha$  and HNF1 $\beta$ .

## CHAPTER 2: MATERIALS AND METHODS

### 2.1 Materials

Restriction endonucleases and DNA modifying enzymes were obtained from the following sources: Boehringer Mannheim, Germany; Life Technologies Inc., MD, USA; New England Biolabs Inc., MA, USA; Promega Corporation, WI, USA; and Stratagene, La Jolla, CA, USA.

Ampicillin, lysozyme, SDS, ethidium bromide, DMSO, ficoll, DTT, TEMED, ONPG, mineral oil, RNase and BSA were all purchased from Sigma Chemical Company, St. Louis, MO, USA.

Proteinase K, T4 polynucleotide kinase and reaction buffer, 1 kb DNA molecular size ladder, LB broth base, ammonium persulphate, penicillin, streptomycin, Eagle's minimal essential media, Dulbecco's modified Eagle's medium, foetal bovine serum, and trypsin were all obtained from Life Technologies Inc., MD, USA.

pSVK3, poly (dI-dC) and all deoxynucleotide triphosphates were purchased from Pharmacia, LKB Biotechnology, Uppsala, Sweden.

The *E. coli* bacterial strain XL-1 was obtained from Stratagene, La Jolla, CA, USA.

Calf alkaline phosphatase and *Taq* and *Pwo* DNA polymerases were provided by Boehringer Mannheim, Germany.

Radioisotopes were obtained from Amersham, UK.

Oligonucleotides were manufactured by Life Technologies Inc., MD, USA.

Sequenase Version 2.0 and associated reagents were purchased from US Biochemical Corporation, Cleveland, USA.

DE-81 paper was obtained from Whatman, England. X-ray film was purchased from Fuji Photo Film Company, Japan. Photographic developer and fixer were obtained from Eastman Kodak, NY, USA.

Wizard™ maxi and miniprep plasmid purification kits, pSV- $\beta$ galactosidase, and Luciferase Assay kits were purchased from the Promega Corporation, Madison, WI, USA.

40% acrylamide/bis solution (29:1, 3.3% C) was obtained from Biorad Laboratories, CA, USA.

QIAGEN-tip 500 (Maxi) plasmid purification kits were supplied by QIAGEN Pty Ltd., Australia.

pBS-KS-HNF1 $\alpha$  and pBJ5-HNF1 $\beta$  were generous gifts from Gerald Crabtree, Beckman Centre, Stanford University School of Medicine, Stanford, CA.

pGEX-2T-DCoH was a generous gift from Jeff Cronk, Department of Molecular and Cell Biology, University of California, Berkeley, CA.

Cos7 and Alexander cell lines were provided by the cell bank, Sir William Dunn School of Pathology, University of Oxford, UK.

All tissue culture flasks, plates and plasticware were obtained from Falcon, NJ, USA.

Cryotubes were obtained from Nunc Inc., Naperville, IL, USA.

## 2.2 Methods

### 2.2.1 Techniques for DNA Manipulation

#### *General Methods*

Plasmid DNA was isolated from bacterial strains using the rapid boil technique according to Holmes and Quigley (1981)<sup>47</sup>.

All restriction endonuclease digests were conducted with the appropriate buffers in accordance with conditions recommended by the supplier.

DNA extraction by phenol/chloroform, ethanol precipitation of DNA, and ligations were carried out according to the guidelines set out in Sambrook *et al.* (1989a)<sup>48</sup> and Ausubel *et al.* (1989)<sup>49</sup>.

Agarose gel electrophoresis of DNA fragments was conducted in a solution of low electroendosmosis grade agarose containing ethidium bromide (0.5 µg/mL) and 1x TAE buffer (40 mM Tris, 20 mM acetate, 1 mM EDTA, pH 8.0), as described in Sambrook *et al.* (1989a)<sup>48</sup>. Fragment size was estimated by comparison with the 1 kb molecular size ladder from BRL. Concentration of DNA samples were estimated by comparison of band intensity with quantification standards, on a 1% agarose gel.

DNA fragments were purified from agarose gels initially by excision when visualised under long wavelength UV light (366 nm), followed by extraction from the agarose by the freeze-squeeze technique (Thuring *et al.* (1975)<sup>50</sup>).

The pSVK3 expression vector was prepared for ligation by cleavage with the appropriate restriction endonuclease followed by dephosphorylation, using calf alkaline phosphatase, as per the protocol outlined in Sambrook *et al.* (1989a)<sup>48</sup>.

*E. coli* strain XL-1 were made competent for transformation using calcium chloride, and were transformed by heat-shock, as described by Sambrook *et al.* (1989a)<sup>48</sup>.

Large quantities of plasmid DNA were prepared by alkaline lysis followed by the Magic/Wizard DNA Purification system<sup>TM</sup>, caesium chloride ultracentrifugation,

according to Sambrook *et al.* (1989a)<sup>34</sup>, or using QIAGEN columns. Both the Magic/Wizard and Qiagen DNA purification systems rely on the use of a proprietary resin to separate DNA from other contaminants. The CsCl gradient ultracentrifugation separated DNA from RNA and proteins on the basis of density.

#### *Preparation of HNF1 $\alpha$ and DCoH cDNAs for Subcloning*

The coding region of the HNF1 $\alpha$  cDNA was copied from the pBS-KS-HNF1 $\alpha$  plasmid using the PCR protocol recommended by Cetus Corporation Limited. Reactions were carried out in 0.5 mL tubes overlaid with mineral oil, using 2.5 U of *Pwo* polymerase, 1x *Pwo* polymerase buffer, (100 mM Tris-HCl, 500 mM KCl, pH 8.3), 2 mM MgSO<sub>4</sub>, 0.3 mM deoxynucleotide triphosphates, 25 pmol of each oligonucleotide primer, 1 ng of pBS-KS-HNF1 $\alpha$  template, and sufficient sterile water in a total volume of 50  $\mu$ L. Amplification was achieved using a thermal cycler programme as follows: 94°C for 5 minutes, then 30 cycles of 94°C for 1 minute, 60°C for 1 minute and 72°C for 3 minutes.

The DCoH open reading frame was also amplified by PCR, using *Taq* polymerase. Reactions contained 2.5 U of polymerase, 1x *Taq* polymerase buffer, (100 mM Tris-HCl, 500 mM KCl), 5 mM MgCl<sub>2</sub>, 0.3 mM deoxynucleotide triphosphates, 20 pmol of each oligonucleotide primer, 20 ng of pGEX-2T-DCoH template and sufficient sterile water in a total volume of 50  $\mu$ L. The thermal cycler programme used to amplify HNF1 $\alpha$  was also employed to generate the DCoH product.

#### *Manual Sequencing of PCR Products*

Sequencing was initially carried out manually using the Sequenase<sup>TM</sup> DNA Sequencing Kit (Version 2.0) from US Biochemical. Double-stranded vectors were prepared for use as templates by the alkaline-denaturation method as follows: 3-5  $\mu$ g of DNA was denatured by adding 0.1 volumes of 2 M NaOH and 2 mM EDTA and incubating for 30 minutes at 37°C. The single-stranded DNA was then ethanol precipitated (see above) and redissolved in 7  $\mu$ L of distilled water. The resuspended single-stranded template was sequenced by the dideoxy chain termination method originally developed by Sanger *et al.*<sup>51</sup> using Sequenase Version 2.0 from US Biochemical. Sequencing gels and electrophoresis was carried out according to the

method described by Sambrook *et al.*<sup>48</sup>. The gels were dried and autoradiographed at RT for 1 to 3 days. Autoradiographs were developed and analysed manually.

### *Automatic Sequencing of PCR Products*

The PCR products for use in expression vectors were inspected for errors by DNA sequence analysis using an ABI PRISM™ 377 automated DNA sequencer using the dye terminator method of cycle sequencing with AmpliTaq® DNA polymerase (3.2 pmol of primer and 500 ng of template). The PCR primers for HNF1 $\alpha$  and DCoH were used to sequence the respective products. Additional primers were also designed in order to sequence the entire HNF1 $\alpha$  cDNA. (See Appendix B for primer sequences.)

## 2.2.2 Mammalian Tissue Culture

### *Maintenance of Tissue Cultures*

Cos7 and Alexander cell lines were cultured in T80 flasks with DMEM and MEM medium, respectively. Upon reaching approximately 80% confluency, the cell lines were rinsed with PBS, (140 mM NaCl, 5 mM KCl, 6.5 mM Na<sub>2</sub>HPO<sub>4</sub>·2H<sub>2</sub>O, pH 7.2), to remove any residual medium, and passaged using 2 mL of 1x trypsin (0.25%) in PBSE buffer, (140 mM NaCl, 5 mM KCl, 6.5 mM Na<sub>2</sub>HPO<sub>4</sub>·2H<sub>2</sub>O, 0.7 mM EDTA, pH7.2). Once the cells had lifted off the flasks into the buffer, the cell suspension was centrifuged at low speed to pellet the cells, which were then resuspended in 2 mL of complete medium and transferred into fresh flasks containing 14 mL of medium, or plates containing 14 mL of medium for transfections (see later).

Cell samples were stored in liquid nitrogen at regular intervals to maintain the cell lines. Following passage at 80% confluency the cells were resuspended in FCS containing 10% DMSO and left overnight at -70°C prior to storage in liquid nitrogen.



### *Transient Transfection of Tissue Culture*

Cos cell transfections were carried out in 150 mm plates containing 14 mL of medium. 25 mm wells, each holding 800  $\mu$ L of medium, were used to transfect the Alexander cells.

Buffers used in transfections were prepared and tested for their ability to form a calcium phosphate precipitate prior to filter sterilisation. Cells were passaged the day before transfection in order to reach a confluency of around 40% at transfection. Two hours before the transfection procedure the medium was changed. The required amount of plasmid DNA was diluted in sterile water to a volume of 240  $\mu$ L for Cos cell transfections, and 48  $\mu$ L for Alexander cell transfections. 45  $\mu$ g of DNA was used to transfect Cos cells for whole cell extracts, whereas 0.5 - 2.0  $\mu$ g of DNA was used in Alexander cell co-transfections for transcriptional assays. An equal volume of buffer A, (0.5 M  $\text{CaCl}_2$ , 0.1 M Hepes, pH 7.1) was then added, followed by 480  $\mu$ L (Cos cells) or 96  $\mu$ L (Alexander cells) of buffer B (0.28 M NaCl, 0.05 M Hepes, 0.75 mM  $\text{NaH}_2\text{PO}_4$ , 0.75 mM  $\text{Na}_2\text{HPO}_4$ , pH 7.1). The solutions were then vortexed and left at room temperature for 15 minutes to allow precipitate formation. This mixture was dispensed over the culture medium and left for 22 hours. The medium was then removed and the cell monolayer rinsed twice with sterile PBS buffer. The cells were then covered with fresh medium and left for a further 24 hours before harvesting according to assay type.

### *Harvesting of Transfected Cos7 Cell Cultures*

Whole cell extracts were prepared from transfected Cos7 cells on 150 mm plates as follows. The tissue culture medium was removed and the monolayer washed twice with sterile PBS. Cells were harvested in 5 mL of TEN buffer (40 mM Tris-HCl pH 7.4, 1 mM EDTA, 0.15 M NaCl), pelleted by centrifugation at 1000 rpm for 5 minutes, and resuspended in 300  $\mu$ L of extraction buffer (40 mM Hepes pH 7.9, 0.4 M KCl, 1 mM DTT, 10% glycerol, 0.1 mM PMSF, 0.1% aprotinin). The cells are disrupted by three freeze-thaw cycles using liquid nitrogen, and centrifuged for 5 minutes at 14000 rpm at 4°C. The supernatants were dispensed into aliquots and stored at -70°C.

### *Harvesting of Transfected Alexander Cell Cultures*

Whole cell extracts of transfected Alexander cells were prepared from 12-well trays. After removal of the medium, the cells were rinsed twice with sterile PBS and

covered with 80  $\mu\text{L}$  of 1x Reporter Lysis buffer (25 mM Tris-phosphate, pH 7.8, 2 mM DTT, 2 mM 1,2-diaminocyclohexane-N, N, N', N'-tetraacetic acid, 10% glycerol, 1% triton X-100). After a 15 minute incubation at room temperature the cells were harvested, vortexed briefly then centrifuged to remove cellular debris. The supernatants were then assayed and/or stored at  $-70^{\circ}\text{C}$ .

### 2.2.3 Electrophoretic Mobility Shift Assays (EMSAs)

#### *Labelling of DNA probes for Gel Shifts*

Double-stranded DNA was end-labelled for gel shifts by first incubating 100 ng of single stranded oligonucleotide with 3  $\mu\text{L}$  of  $\gamma\text{-}^{32}\text{P}\text{-ATP}$ , 1x kinase buffer, (50 mM Tris-HCl, pH 7.5, 10 mM  $\text{MgCl}_2$ , 5 mM DTT, 50  $\mu\text{g}/\text{mL}$  BSA) and 10 U of T4 polynucleotide kinase at  $37^{\circ}\text{C}$  for 45 minutes. An excess of the complementary strand of the oligonucleotide was annealed to the labelled DNA by mixing the two in a waterbath at  $94^{\circ}\text{C}$  for 5 minutes and then leaving to cool until the bath had reached room temperature. The entire mixture was then loaded onto a 10% polyacrylamide gel in 1x TBE (90 mM Tris, 90 mM boric acid, 20 mM EDTA) and electrophoresed at 20 watts for 1.5 hours. Exposure of the gel to X-ray film for 1 minute located the labelled band, which was then excised and the DNA eluted overnight at  $37^{\circ}\text{C}$  in 50 mM KCl. The next day the radioactivity of 1  $\mu\text{L}$  of DNA was measured in a scintillation counter by Cerenkov counting, and the tube stored at  $4^{\circ}\text{C}$ .

#### *The Gel Shift*

EMSAs, also known as gel shifts, were carried out in eppendorf tubes in a total volume of 20  $\mu\text{L}$ . Prior to the addition of the labelled oligonucleotide, 1x gel shift buffer (10 mM Tris-HCl pH 8.0, 54 mM NaCl, 0.05 mM EDTA, 1 mM DTT, 5 mM  $\text{MgCl}_2$ , 0.1% non-fat milk powder), 1  $\mu\text{g}$  of poly dI.dC, the whole cell or nuclear extract and the balance of water were preincubated on ice for 10 minutes, either with or without competitor DNA (5 - 100 ng). 1  $\mu\text{L}$  of  $^{32}\text{P}$  labelled, double stranded oligonucleotide was added and the mixture left at room temperature for 15 minutes. 10  $\mu\text{L}$  of each reaction was loaded on a 4% polyacrylamide gel in 0.25x TBE and electrophoresed at 200V for approximately 1.25 hours. The gel was then transferred onto Whatman DE-81 paper and

dried in a gel dryer for 15 minutes at 80°C. The gel was then exposed to X-ray film overnight at -70°C.

#### 2.2.4 Rat Liver Nuclear Extracts

Prior to the use of buffers in the extraction of protein from and the preparation of rat liver nuclei, the following were added (per 1 L of buffer): 500  $\mu$ L of 1 M DTT, 500  $\mu$ L of 1 M Spermidine, 150  $\mu$ L of 1 M Spermine, 1 mL of 1 M PMSF in DMF, 2 mL of 1 M benzamidine, 500  $\mu$ L of 1 mg/mL leupeptin, 500  $\mu$ L of 1 mg/mL pepstatin, 500  $\mu$ L of 2 mg/mL aprotinin. The Spermidine and Spermine were added only to the homogenisation buffer.

Approximately 10 g of fresh rat liver was made up to 20 mL with ice-cold homogenisation buffer (10 mM Hepes pH 7.7, 25 mM KCl, 1 mM EDTA, 2 M sucrose, 10% glycerol) and homogenised using the ultra-turrax (probe size 18N) at mid-speed. The homogenate was then made up to 50 mL with homogenisation buffer and layered onto 10 mL of homogenisation buffer in SW28 ultracentrifuge tubes. These tubes were centrifuged at 4°C for 30 minutes at 24000 rpm in a Beckman SW28 rotor. The floating fat pellet was removed with a tissue before the pink supernatant was removed down to the buffer cushion and discarded. The tubes were then quickly inverted to remove the rest of the buffer and drained upside down. The pelleted nuclei were resuspended in 1.5 mL of buffer C (20 mM Hepes pH 7.7, 0.42 M NaCl, 1.5 mM  $MgCl_2$ , 0.2 mM EDTA, 25% glycerol), and stirred gently for 30 minutes at 4°C using a 5 mL beaker and a midget flea. The extract was then centrifuged for 30 minutes at 30000rpm in a TL100.2A rotor. The supernatant was removed immediately, to prevent chromatin swelling, and dialysed twice against 200 mL buffer D (20 mM Hepes pH 7.7, 0.1 M KCl, 20% glycerol, 0.2 mM EDTA) for 2 hours. The extract was then transferred into eppendorf tubes and centrifuged for 10 minutes in a microfuge at 4°C. Aliquots were snap frozen in liquid nitrogen before storage at -70°C.

#### 2.2.5 Transcriptional Assays

The luciferase reporter assay was used to test the transcriptional activities of the factor IX promoter in the presence and absence of the HNF1 and DCoH proteins. The

wildtype promoter and each of the -5 and -6 mutant promoters had been ligated into the pGL2B vector adjacent to the luciferase gene. One promoter-reporter gene construct was transfected into Alexander tissue culture cells along with an HNF1 expression vector (with or without DCoH) and the  $\beta$ -galactosidase plasmid, pSV- $\beta$ -galactosidase. The effects of these transcription factors on the accompanying factor IX promoter were tested following cell harvest indirectly by measuring the luciferase enzyme activity. Incubation of 20  $\mu$ L of cell extract with 100  $\mu$ L of Luciferase Assay Reagent (20 mM tricine, 1.07 mM  $(\text{MgCO}_3)_4\text{Mg}(\text{OH})_2 \cdot 5\text{H}_2\text{O}$ , 2.67 mM  $\text{MgSO}_4$ , 0.1 mM EDTA, 33.3 mM DTT, 270  $\mu$ M coenzyme A, 470  $\mu$ M luciferin, 530  $\mu$ M ATP) at room temperature initiated the enzyme reaction which releases light. This product was measured in a luminometer at 10 second intervals over 2 minutes.

The success of the transfection was quantified by measuring the amount of  $\beta$ -galactosidase synthesised from the pSV- $\beta$ -galactosidase vector by the production of *o*-nitrophenol by the enzyme:



20  $\mu$ L of cell extract was incubated with 400  $\mu$ L of  $\beta$ -galactosidase assay buffer (60 mM  $\text{NaH}_2\text{PO}_4$ , 40 mM  $\text{Na}_2\text{HPO}_4$ , 10 mM KCl, 1 mM  $\text{MgCl}_2$ ) and 200  $\mu$ L of ONPG reagent (1.3  $\mu$ M *o*-nitrophenyl- $\beta$ -galactoside (ONPG), 60 mM  $\text{NaH}_2\text{PO}_4$ , 40 mM  $\text{Na}_2\text{HPO}_4$ ) at 37°C for at least 1 hour, and no more than 8 hours. The amount of *o*-nitrophenol produced was measured on a spectrophotometer at 420 nm.

## CHAPTER 3: RESULTS

### Part A: Subcloning of the cDNAs for HNF1 $\alpha$ , HNF1 $\beta$ and DCoH

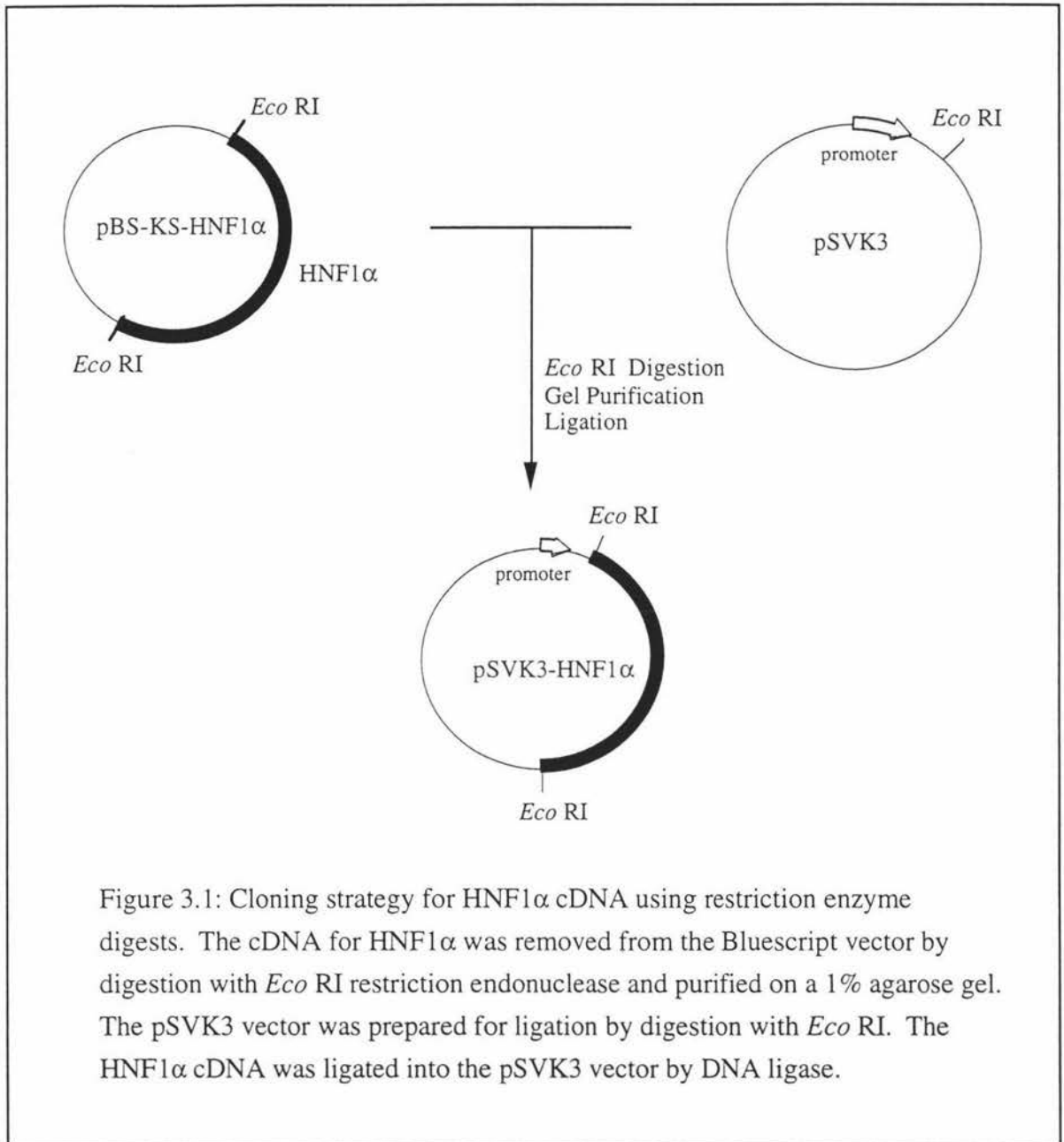
#### 3.1 Preparation of the HNF1 $\alpha$ Expression Constructs

##### 3.1.1 Restriction Enzyme Digests

The HNF1 $\alpha$  cDNA was obtained in the pBluescript-KS vector, which is a general cloning vector for use in *E. coli*, in which the polylinker is in the *Kpn* I-*Sac* I (KS) orientation. It does not contain the eukaryotic promoter sequences necessary for expression in mammalian tissue culture. HNF1 $\alpha$  was therefore subcloned into the mammalian expression vector pSVK3 (see Appendix A) for mammalian cell transfections. The Bluescript construct contained two unique *Eco* RI recognition sites at either end of the HNF1 $\alpha$  cDNA (see Appendix A), such that digestion with this enzyme would separate the insert from the vector (Figure 3.1). However, the conventional incubation of 1  $\mu$ g of DNA with 20 U of *Eco* RI for 1 hour only partially digested the plasmid. Extension of the incubation time to 8 hours, using only 500 ng of DNA also failed to result in the complete digestion of the plasmid (Figure 3.2). The products of this digestion could still be used for subcloning purposes, except that the larger amount of DNA which was digested only yielded sufficient HNF1 $\alpha$  DNA for one ligation. Before the subcloning procedure could begin, the band from the *Eco* RI digest which corresponds to the HNF1 $\alpha$  cDNA needed to be identified. Initially, the restriction endonuclease *Hga* I, was used to distinguish between the two bands, as *Hga* I will cleave the pBluescript vector but does not cleave the HNF1 $\alpha$  cDNA (Figure 3.3A). However, this restriction endonuclease was unable to completely digest the construct containing HNF1 $\alpha$  (Figure 3.3B). *Pst* I was then used instead to digest HNF1 $\alpha$ , identifying this cDNA as the larger product of the *Eco* RI digestion (Figure 3.4).

The pSVK3 vector was digested with *Eco* RI and phosphorylated (Section 2.2.1), then ligated to the gel purified HNF1 $\alpha$  cDNA (Section 2.2.1). Competent XL-1 cells were transformed with the ligation reaction and the resulting 70 colonies were tested for the presence of plasmid DNA (Figure 3.5). Four colonies which contained potential

pSVK3-HNF1 $\alpha$  vectors were identified and the plasmid DNA was digested with *Pst* I and *Sma* I to characterise the vectors further (Figure 3.6). The orientation of



the HNF1 $\alpha$  cDNA in the pSVK3 vector was identified by a diagnostic digestion of the construct with *Kpn* I (Figure 3.7). The HNF1 $\alpha$  cDNA contains one site for the *Kpn* I enzyme, and the pSVK3 vector has a site for the enzyme in its multiple cloning site. The sizes of the products of a *Kpn* I digestion were used to determine the orientation of the

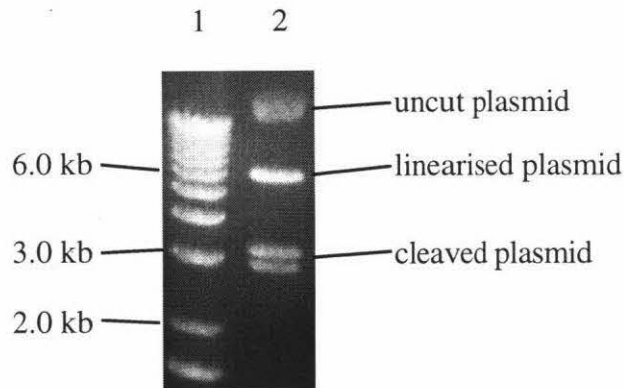
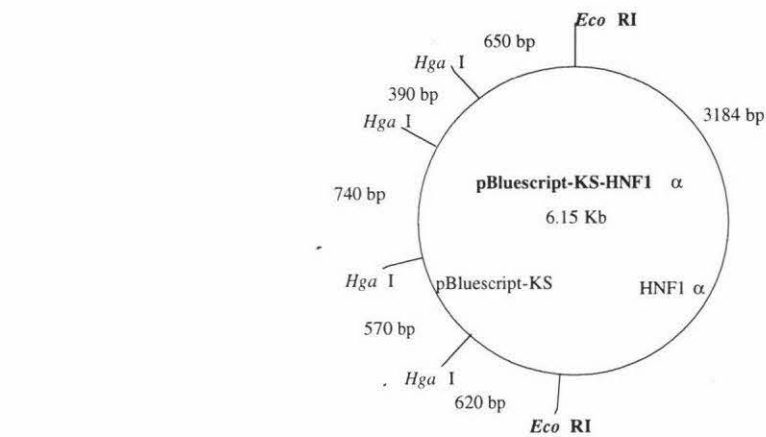


Figure 3.2: Partial digestion of pBluescript-KS-HNF1 $\alpha$  by *Eco* RI restriction endonuclease. A 15  $\mu$ L aliquot of the enzyme digest was analysed by electrophoresis in a 1% agarose gel containing 1x TAE. The DNA was stained using ethidium bromide and visualised under UV light. Lanes 1, 1 kb ladder; 2, forms of the plasmid.

insert as follows: digestion of the construct carrying the insert in the correct orientation would yield bands of sizes 6.8 kb and 250 bp (approximately) whereas the construct with the insert ligated in the incorrect orientation would yield bands of sizes 4.2 kb and 2.9 kb (Figure 3.7A). The enzyme digest with *Kpn* I was only partial, but clearly showed that all four transformants carrying a pSVK3-HNF1 $\alpha$  plasmid yielded bands corresponding to the insert in the incorrect orientation (Figure 3.7B).

The insert was then cleaved from the vector with *Eco* RI and re-ligated back together in an attempt to generate constructs with the insert in the correct orientation. Competent cells were transformed with the ligation reaction and tested for the presence of the pSVK3-HNF1 $\alpha$  plasmid as before (not shown). Only one colony was identified as carrying the pSVK3-HNF1 $\alpha$  plasmid, and the orientation of the insert in this construct was determined again by digestion with *Kpn* I. The plasmid was completely digested by the enzyme, but electrophoresis of the digestion products revealed three bands sized around 250 bp, 4.0 kb and 6.0 kb (Figure 3.8A). This was the band pattern which would be generated by a vector containing two HNF1 $\alpha$  cDNAs, with the first insert in



B.

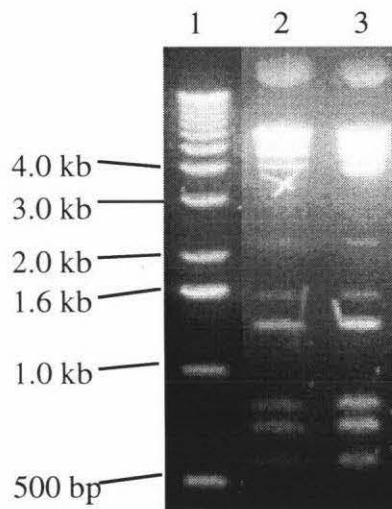


Figure 3.3: Digestion of the pBluescript vector with *Hga* I. A: Restriction map of the cleavage sites for *Hga* I in pBluescript-HNF1 $\alpha$ . The *Eco* RI sites which the HNF1 $\alpha$  cDNA is cloned between are shown in bold. B: The partial digestion of pBluescript-HNF1 $\alpha$  by *Hga* I. A 15  $\mu$ L aliquot of the enzyme digest was analysed by electrophoresis in a 1% agarose gel containing 1x TAE. The DNA was stained using ethidium bromide and visualised under UV light. Lanes: 1, 1 kb ladder; 2, 3, *Hga* I-cut pBluescript-HNF1 $\alpha$ .



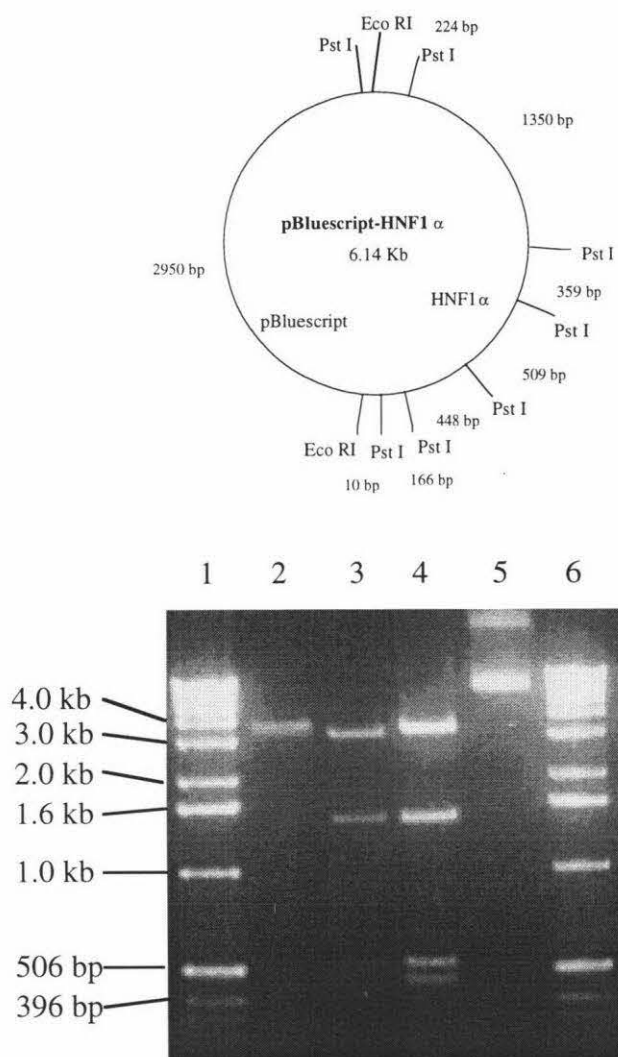


Figure 3.4: Identification of the HNF1 $\alpha$  cDNA by digestion with *Pst* I. *Top*: Restriction map of the cleavage sites for *Pst* I in pBluescript-HNF1 $\alpha$ . The *Eco* RI sites which the HNF1 $\alpha$  cDNA is cloned into are shown in bold. *Bottom*: Digestion of pBluescript- HNF1 $\alpha$  by *Pst* I. A 15  $\mu$ L aliquot of the enzyme digest was analysed by electrophoresis in a 1% agarose gel containing 1x TAE. The DNA was stained using ethidium bromide and visualised under UV light. Lanes: 1, 1 kb ladder; 2, *Eco* RI-cut pBluescript-HNF1 $\alpha$ ; 3, *Pst* I, *Eco* RI-cut pBluescript-HNF1 $\alpha$ ; 4, *Pst* I-cut pBluescript-HNF1 $\alpha$ ; 5, uncut pBluescript-HNF1 $\alpha$ ; 6, 1 kb ladder.

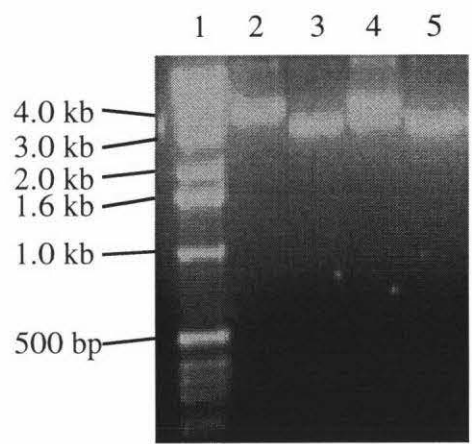
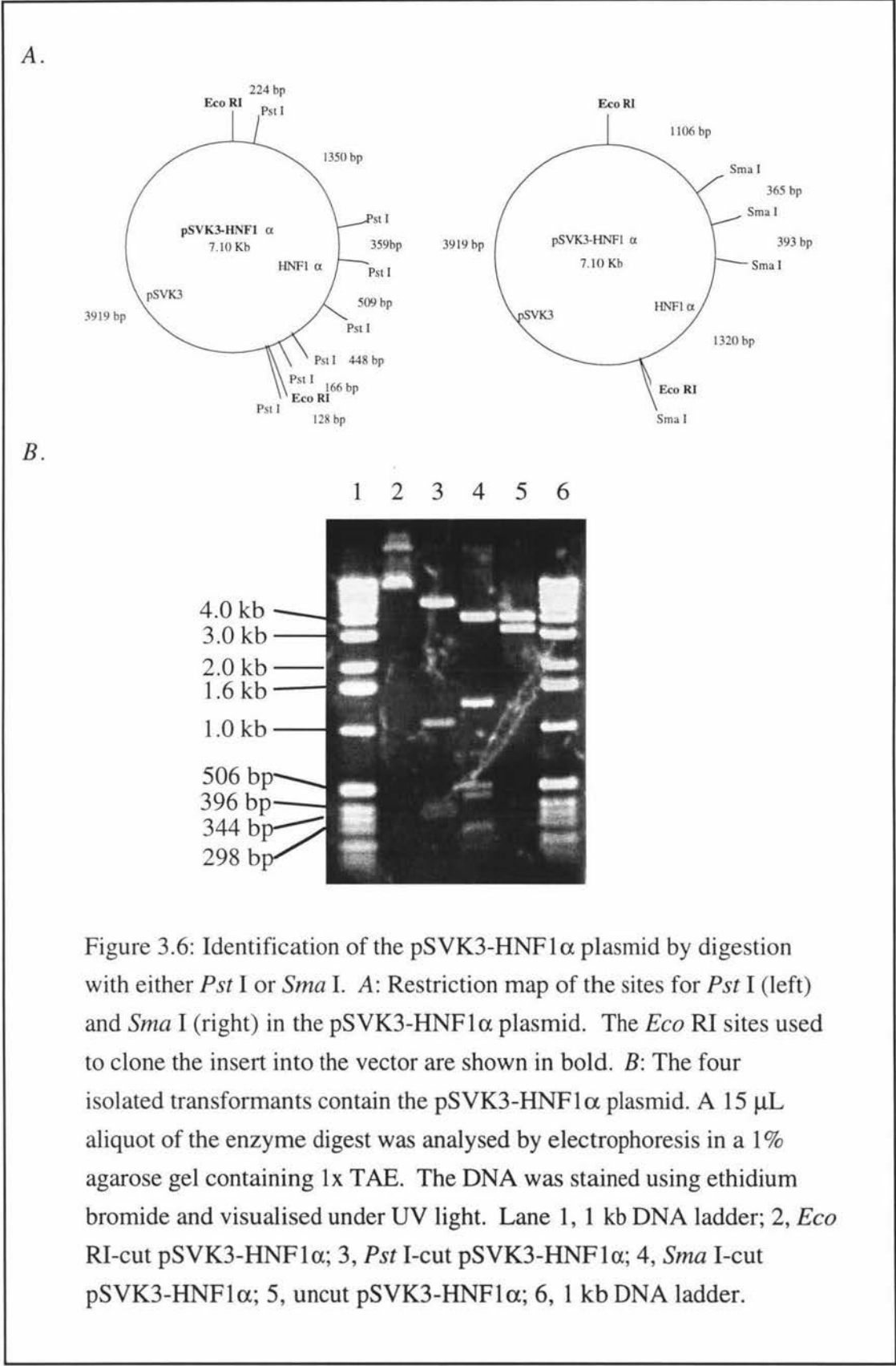
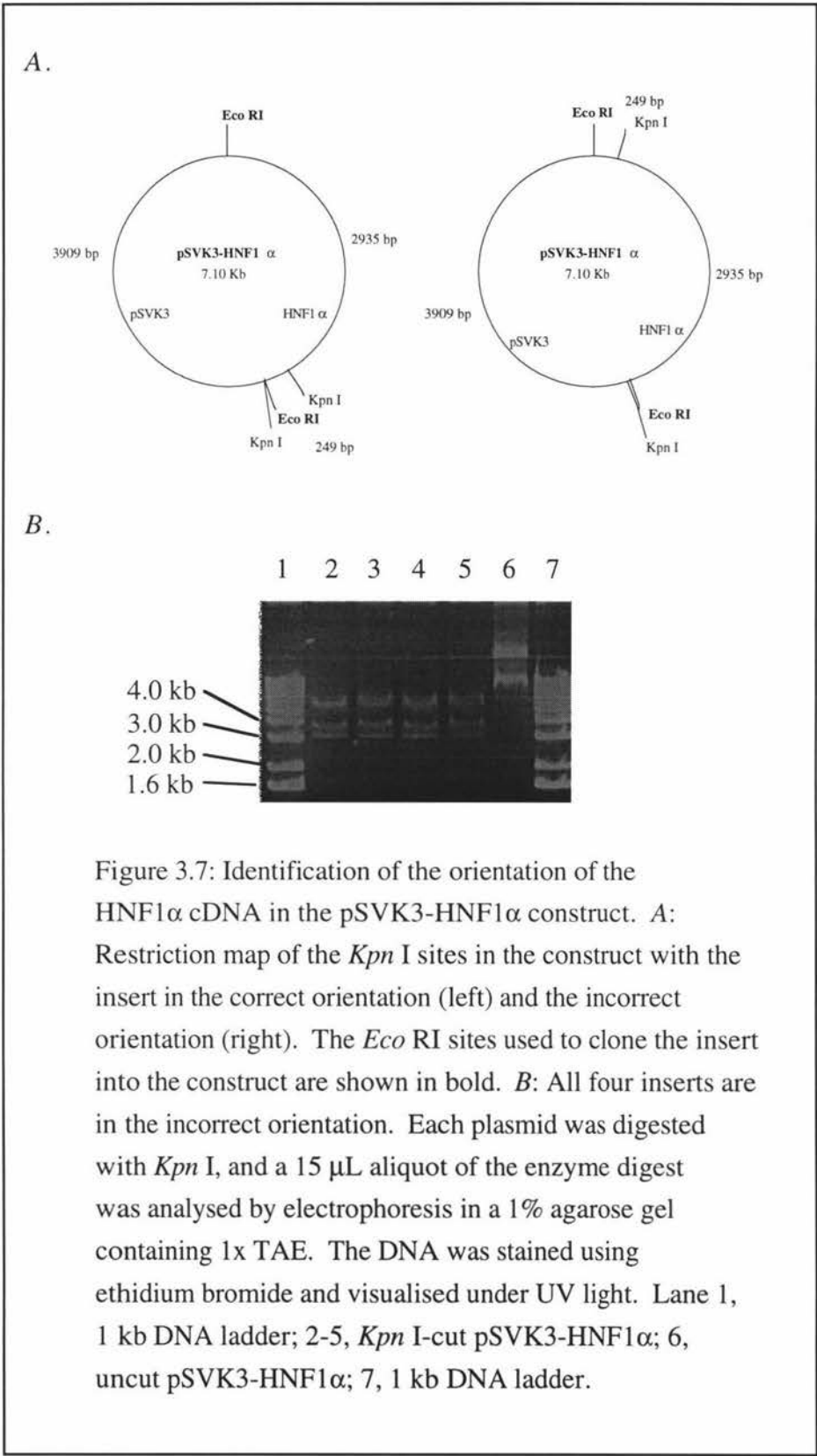


Figure 3.5: Identification of transformants carrying potential pSVK3-HNF1 $\alpha$  plasmids. DNA prepared by the rapid boil method was digested with *Eco* RI to remove any insert from the vector. A 15  $\mu$ L aliquot of the enzyme digest was analysed by electrophoresis in a 1% agarose gel containing 1x TAE. The DNA was stained using ethidium bromide and visualised under UV light. Lane 1, 1 kb DNA ladder; 2-5, products of *Eco* RI digestion on the plasmid DNA from four transformants. Lane 4 contains two bands of sizes comparable to that of the pSVK3 vector (3.9 kb) and the HNF1 $\alpha$  cDNA (3.2 kb), while the other lanes (2, 3 and 5) only contain a single band.





the incorrect orientation and the second insert ligated in the correct orientation. This suggests that the first pSVK3-HNF1 $\alpha$  construct was only digested by the *Eco* RI enzyme at the 3' site and a second insert was ligated into the existing construct (Figure 3.8B). A range of ligation reactions with varying vector-to-insert ratios (2:1, 3:1, and 4:1) were set up in an attempt to increase the number of transformed bacteria. Only one colony was isolated from the transformations, and contained a plasmid approximately 4.0 kb in size (not shown).

As the purification of the HNF1 $\alpha$  cDNA from *Eco* RI digests only yielded sufficient DNA for only one ligation, and because the ligations attempted did not successfully result in the correct construct, a directional cloning procedure was then initiated. In order to directionally clone the HNF1 $\alpha$  cDNA, there must be two different restriction enzyme sites at each end of the cDNA sequence. These sites were created by PCR (Section 3.1.2)

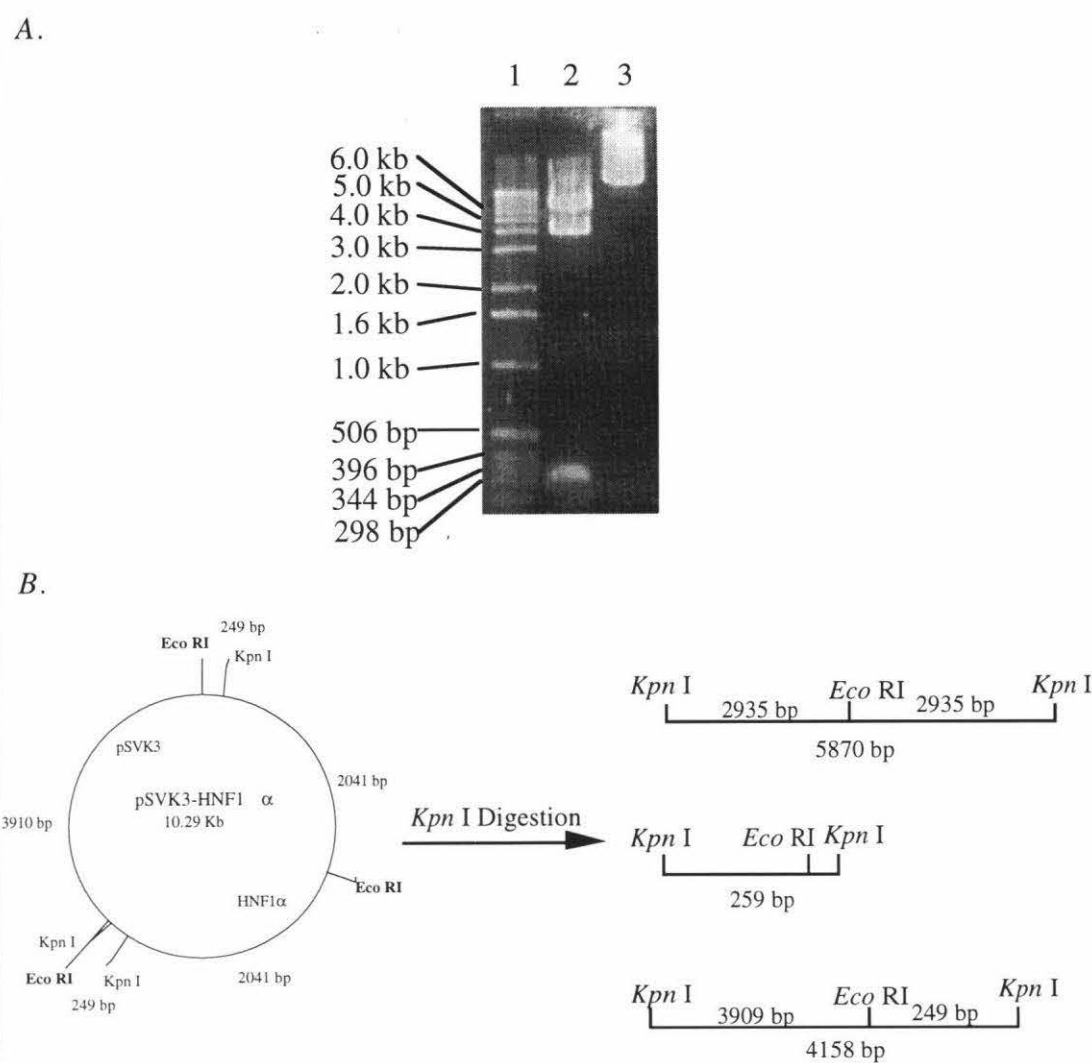


Figure 3.8: The construct with two inserts. A: The plasmid DNA from the only pSVK3-HNF1 $\alpha$  transformant was digested with *Kpn I* to determine the orientation of the insert in the construct. A 15  $\mu$ L aliquot of the enzyme digest was analysed by electrophoresis in a 1% agarose gel containing 1x TAE. The DNA was stained using ethidium bromide and visualised under UV light. Lane 1, 1 kb ladder; 2, *Kpn I*-cut pSVK3-HNF1 $\alpha$ ; 3, uncut pSVK3-HNF1 $\alpha$ . B: A restriction map of the presumed construct carrying two copies of the HNF1 $\alpha$  cDNA in two orientations (left). The **Eco RI** sites into which the inserts were ligated are shown in bold. The products of a *Kpn I* digestion of this plasmid are shown at the right.

### 3.1.2 PCR Amplification of HNF1 $\alpha$ cDNA

Because *Eco* RI was apparently unable to completely cleave pBluescript-KS-HNF1 $\alpha$ , the HNF1 $\alpha$  cDNA was amplified from the plasmid using PCR. Oligonucleotide primers were designed to anneal at either end of the cDNA (see Appendix B), and contained unique restriction enzyme recognition sites to allow directional cloning of the 3.2 kb PCR product. The 5' primer contained the cleavage site for *Eco* RI and the 3' primer had the cleavage site for *Sal* I (Figure 3.9). The PCR was carried out using a set of standard conditions (see Table 1, page 39), using *Pfu* DNA polymerase, which is a high fidelity enzyme. The cDNA being amplified is reasonably large (3.1 kb), and was for use in the synthesis of a transcription factor. A change in the DNA sequence can result in a change in amino acid, which may effect the function of the protein, so it is important that there are no PCR-induced errors. *Pfu* polymerase possesses a proofreading activity and therefore is an ideal choice of enzyme. On the first attempt, an incorrect product of approximately 600 bp was generated. A number of possible variables within the PCR reaction, except for the primers, were examined in an attempt to generate a product of the correct size (Table 1). The PCR was attempted with *Taq* polymerase because it is a more processive enzyme than *Pfu*, and therefore should be able to amplify the entire cDNA. However all of the PCR reactions produced the same 600 bp product, or failed to generate any products. The oligonucleotides were tested for the presence of secondary structures by raising the annealing temperature to 95°C. No product was generated from this PCR. This result suggests that the 600 bp product is generated by aberrant primer annealing. The pBluescript-KS plasmid contains annealing sites for the universal primers T7 and T3 on either side of the multiple cloning site, so these oligonucleotides were chosen to substitute the original primers. A PCR using the new primers, which amplifies the full-sized cDNA, would reveal any faulty primers.

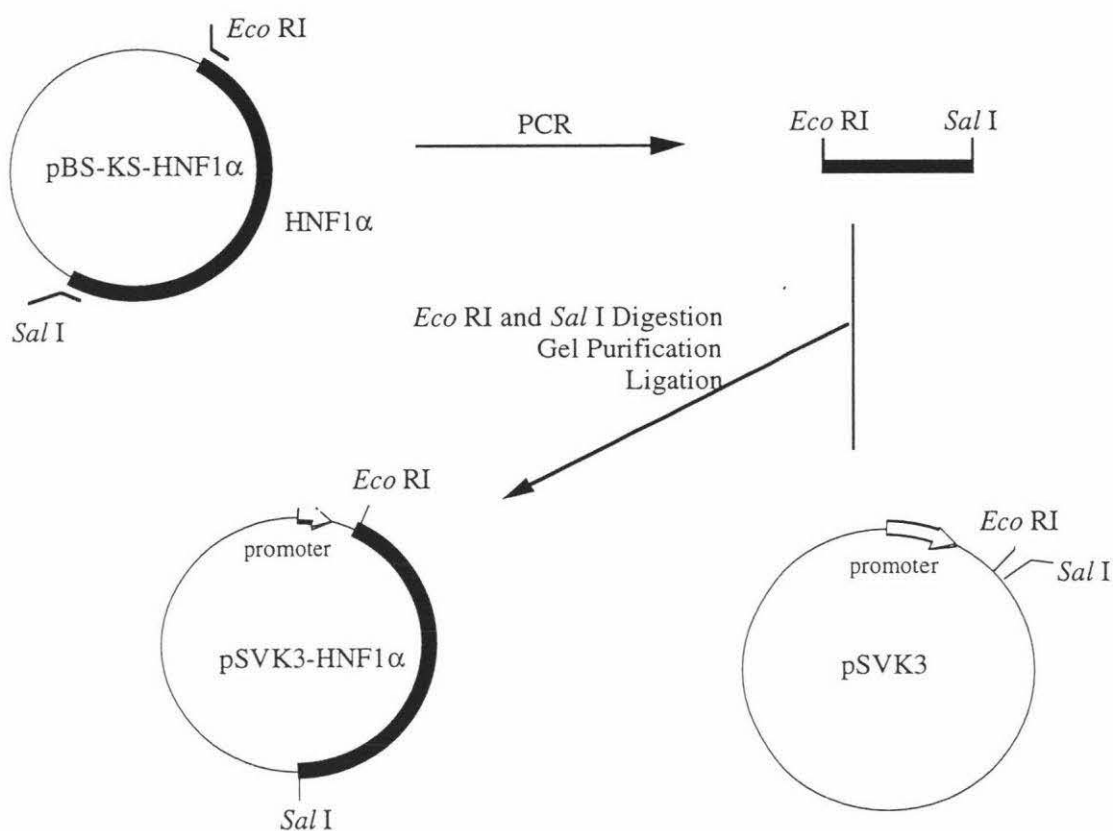


Figure 3.9: Cloning strategy for the HNF1 $\alpha$  cDNA using PCR. The cDNA for HNF1 $\alpha$  was amplified from the pBluescript-KS-HNF1 $\alpha$  vector by PCR using primers homologous to the ends of the cDNA sequence. The forward primer contained a cleavage site for *Eco* RI and the reverse primer contained a cleavage site for *Sal* I. The PCR product and the pSVK3 vector were prepared for ligation by digestion with *Eco* RI and *Sal* I. The HNF1 $\alpha$  cDNA was directionally cloned into the pSVK3 vector.



Table 1: HNF1 $\alpha$  cDNA Amplification by PCR.

Reaction	Polymerase	Conditions	Products	Interpretation
Standard	<i>Pfu</i>	95°C, 1 minute 60°C, 1 minute 72°C, 1 minute 30 cycles	600 bp	Incorrect primer annealing.
Decreased primer hybridisation temperatures, longer extension time.	<i>Pfu</i>	95°C, 1 minute 37°C, 1 minute 72°C, 4 minutes 30 cycles	600 bp	Test other reaction components.
Change polymerase	<i>Taq</i>	95°C, 1 minute 37°C, 1 minute 72°C, 2 minutes 30 cycles	600 bp	<i>Pfu</i> polymerase is functional.
MgCl <sub>2</sub> titration	<i>Pfu</i>	95°C, 1 minute 37°C, 1 minute 72°C, 4 minutes 30 cycles 0-10 mM MgCl <sub>2</sub>	600 bp	MgCl <sub>2</sub> concentration of 5 mM best, but still not right product.
Linearised template	<i>Pfu</i>	95°C, 1 minute 37°C, 1 minute 72°C, 4 minutes 30 cycles	600 bp	Circular template functional.
Different buffers: (1) 60 mM (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> . (2) KCl → NaCl (3) KCl → NaCl + 60 mM (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> .	<i>Taq</i>	95°C, 1 minute 37°C, 1 minute 72°C, 2 minutes 30 cycles	(1) none (2) none (3) none	Original KCl buffer is best.

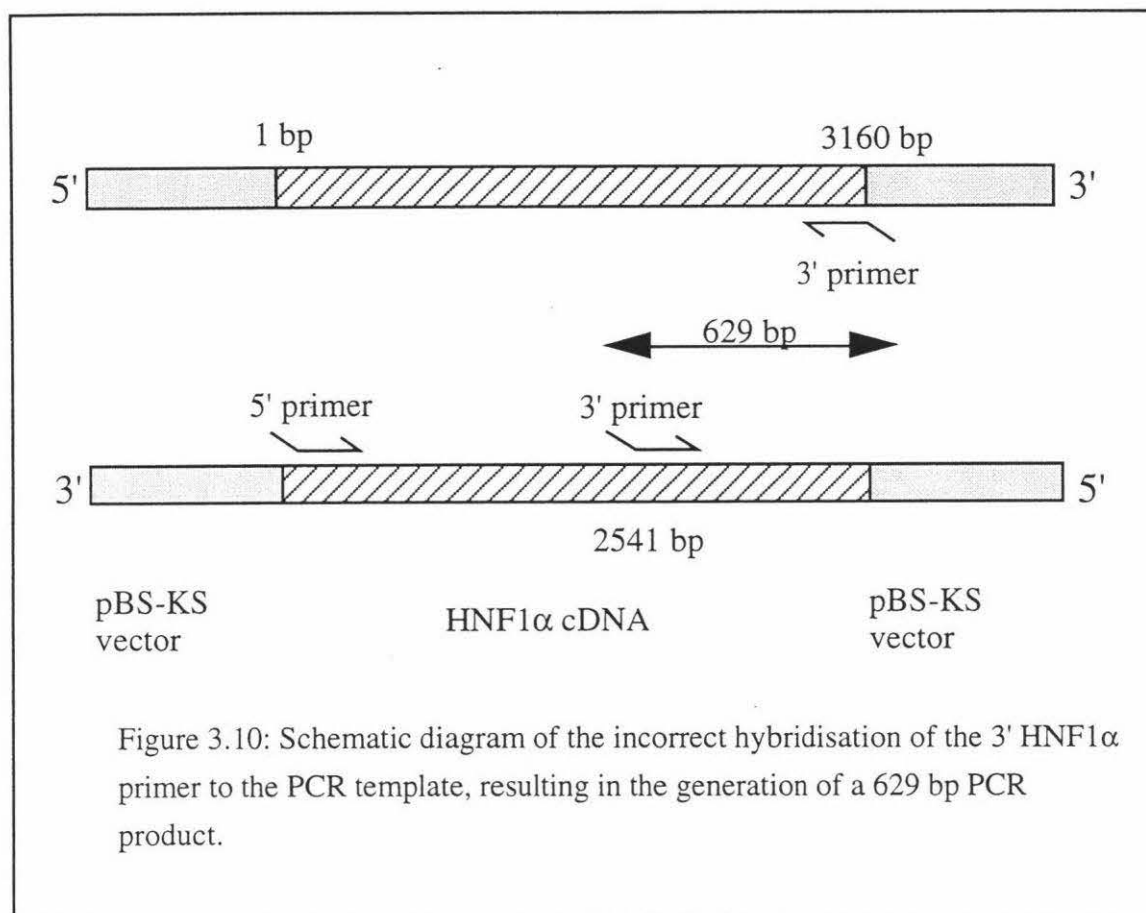
The reactions were conducted using an annealing temperature of 37°C, and the results are shown in Table 2:

Table 2: HNF1α PCR Primer Test Results.

Primers	Expected Products	Observed Products
HNF1α 5', HNF1α 3'	3.1 kb	600 bp
HNF1α 5', T7	3.1 kb	none
T3, HNF1α 3'	3.1 kb	600 bp
HNF1α 5' alone	none	none
HNF1α 3' alone	none	600 bp
T7, T3	3.1 kb	5 bands between 450 bp and 3.1 kb.

These results indicate that the 600 bp product is generated through the annealing of the 3' HNF1α primer to a second site on the template. When T3 was substituted for the 5' HNF1α primer, the same sized product was generated, suggesting that this 600 bp fragment was not caused by the incorrect annealing of the 5' primer. These conclusions are confirmed by the lack of products when only the 5' primer is used, and the identical 600 bp product of the 3' primer-only reaction. The T7 and T3 PCR generated five PCR products. These primers have only one unique hybridisation site each, at the ends of the HNF1 cDNA, and therefore should produce only one product. The large number of products generated here was the result of a very low primer hybridisation temperature (37°C) in the PCR, so these primers annealed to non-specific sites.

The aberrant annealing of the 3' primer was tested further by predicting the hybridisation pattern of the 3' HNF1α primer on the HNF1α template sequence. The French internet-based alignment program, Genestream lalign [website = <http://vega.crbm.cnrs-mop.fr/bin/lalign-guess.cgi>], indicated that the 3' primer could anneal to the template 629 bp upstream of the 3' end of the cDNA, in addition to its 3' site (Figure 3.10). Although only 8 consecutive bases of the 30 mer primer align to the template, the hybridisation occurs at the 3' end of the oligonucleotide, providing the DNA polymerase with a sufficient start site for synthesis.



A new 3' primer containing a *Sal*I recognition site was designed to hybridise at a unique site 2.3 kb into the HNF1α cDNA, 204 bp downstream of the translational stop codon. The following PCR conditions were used reproducibly to generate sufficient amounts of cDNA for cloning:

94°C                      5 minutes, then  
 94°C                      1 minute,  
 60°C                      1 minute, and  
 72°C                      3.5 minutes for 30 cycles.

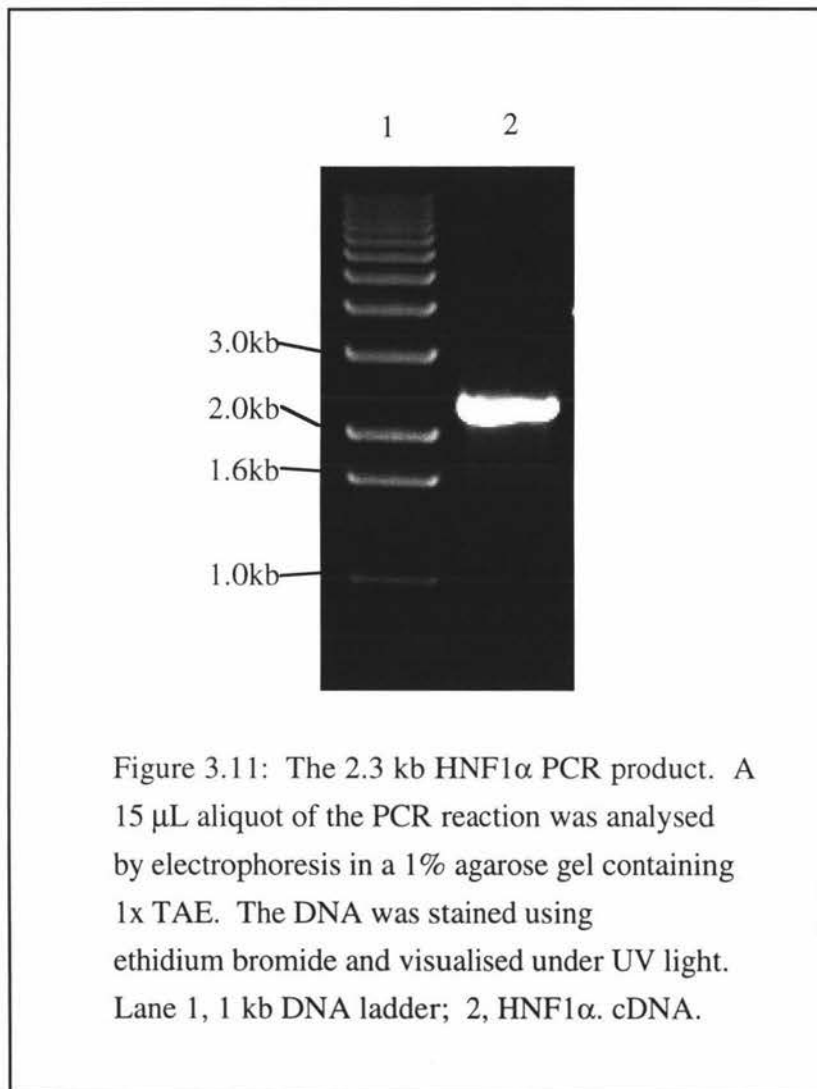
250 ng of each primer, 0.3 mM dNTPs, 2 mM MgCl<sub>2</sub>, 2 ng pBluescript template, 1 U *Pwo* polymerase.

A new PCR enzyme, *Pwo* polymerase, was used in this PCR. *Pwo* polymerase is a highly processive enzyme, like *Taq* polymerase, and was chosen for its ability to amplify large fragments to a high level. However, *Pwo* polymerase is also a high fidelity enzyme, like *Pfu* polymerase, through possession of a 3'→5' exonuclease activity. *Pwo*

polymerase therefore represents the best characteristics of both *Taq* polymerase and *Pfu* polymerase in one enzyme.

### 3.1.3 Cloning of the HNF1 $\alpha$ cDNA

The 2.3 kb PCR product generated by PCR (Figure 3.11) was digested with endonucleases *Eco* RI and *Sal* I, in preparation for cloning into the pSVK3 vector.



The PCR reaction mix was also digested with *Not* I to specifically cleave any residual Bluescript template present. Following a phenol/chloroform extraction and quantification, the HNF1 $\alpha$  cDNA was ligated into the appropriately prepared pSVK3

vector (Section 2.2.1). Competent XL-1 cells were then transformed with the ligation products (as outlined in Section 2.2.1) and transformants were screened for positive clones by digestion of the DNA with the restriction enzyme *Eco* RI (Figure 3.12). A positive clone should yield a 6 kb fragment on digestion of the plasmid DNA with *Eco* RI.

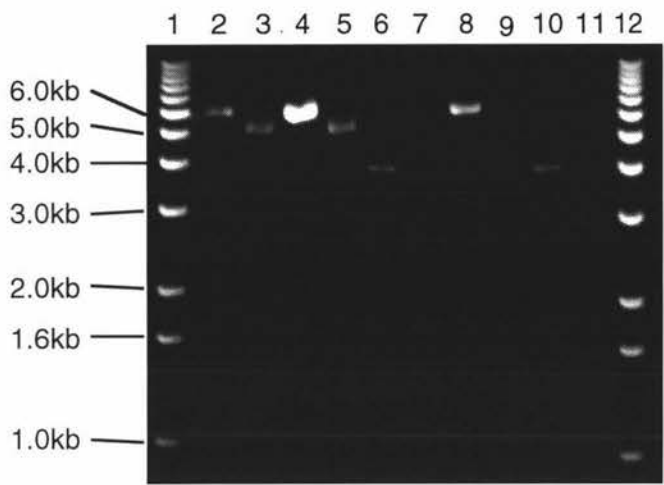


Figure 3.12: Identification of positive colonies. The plasmid DNA from transformants was cleaved with *Eco* RI to identify any recombinant constructs. A 15  $\mu$ L aliquot of the enzyme digest was analysed by electrophoresis in a 1% agarose gel containing 1x TAE. The DNA was stained using ethidium bromide and visualised under UV light. Lanes 1, 12, 1 kb ladder; 2, 4, 6, 8, 10, *Eco* RI-cut plasmid DNA; 3, 5, 7, 9, 11, uncut plasmid DNA.

Three positive clones were identified on the basis of the size of the plasmid. These are shown in lanes 2, 4 and 8 of Figure 3.12. A *Xho* I digest of the three clones (Figure 3.13) yielded bands of 2 and 4 kb in size, thus distinguishing the correct plasmid from incorrect constructs.

Based on a restriction map for this plasmid (see Appendix A), a reverse orientation of the

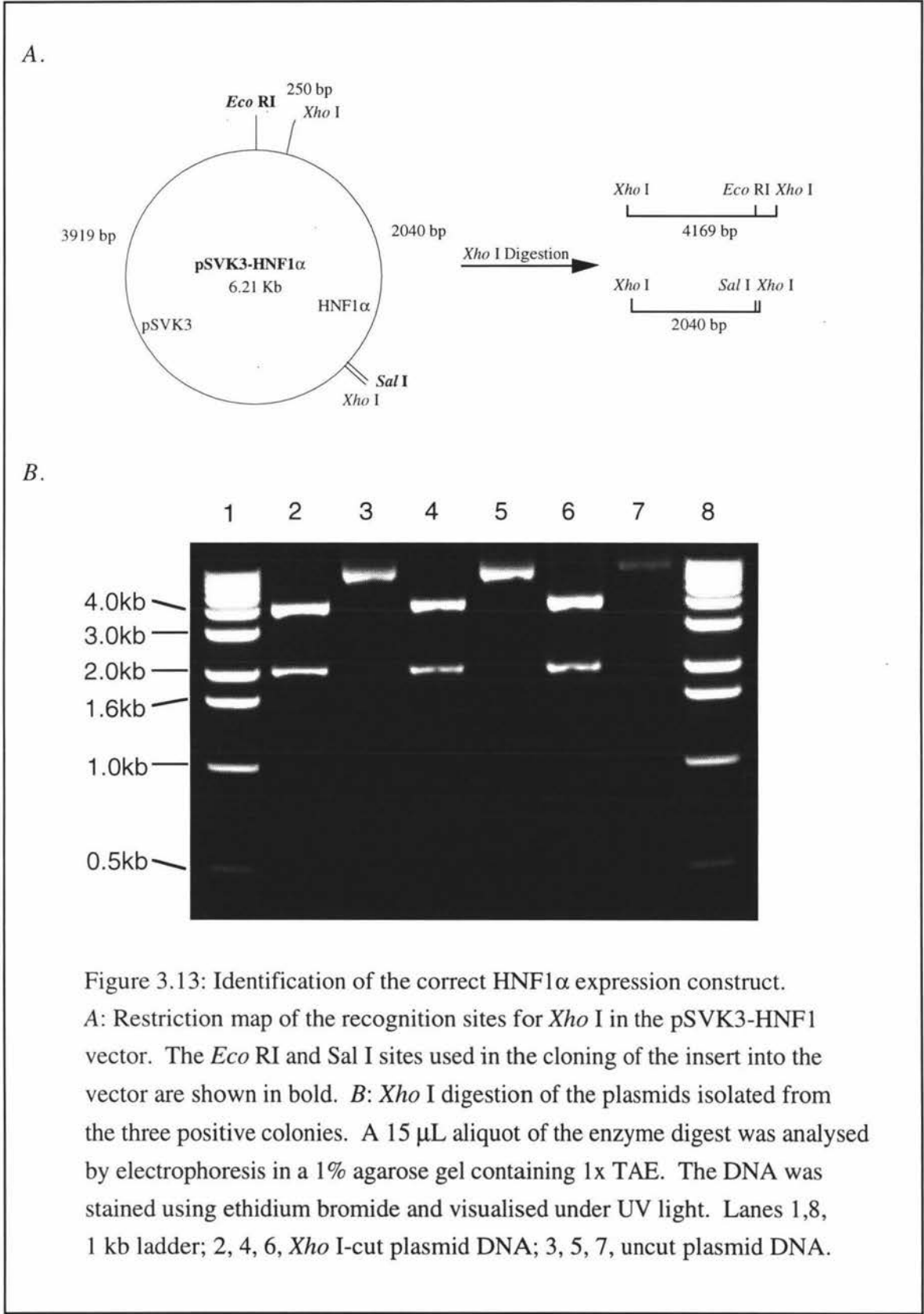


Figure 3.13: Identification of the correct HNF1 $\alpha$  expression construct.

A: Restriction map of the recognition sites for *Xho I* in the pSVK3-HNF1 vector. The *Eco RI* and *Sal I* sites used in the cloning of the insert into the vector are shown in bold. B: *Xho I* digestion of the plasmids isolated from the three positive colonies. A 15  $\mu$ L aliquot of the enzyme digest was analysed by electrophoresis in a 1% agarose gel containing 1x TAE. The DNA was stained using ethidium bromide and visualised under UV light. Lanes 1,8, 1 kb ladder; 2, 4, 6, *Xho I*-cut plasmid DNA; 3, 5, 7, uncut plasmid DNA.

insert would have given bands of size 250 bp and 5.8 kb, while a vector alone would have only been linearised by *Xho* I. All three of the recombinant colonies were retained for DNA sequencing.

### 3.1.4 Sequencing of the HNF1 $\alpha$ cDNA

Despite using a high fidelity, processive DNA polymerase (*Pwo* polymerase) in the PCR reactions, it was still possible that the large PCR product representing the HNF1 $\alpha$  cDNA (2.3 kb) contained PCR-induced errors. The cDNA for HNF1 $\alpha$  was therefore sequenced mainly on one strand, initially using manual sequencing. The Sequenase<sup>TM</sup> DNA sequencing kit was used for the sequencing process (see Section 2.2.2), and the PCR primers were used as sequencing primers. The expected maximum number of nucleotides that could be sequenced manually was less than 300 bp, so a sequencing strategy was developed using the M13 vector and internal sequencing primers (Figure 3.14). The first 250 bp of the cDNA were to be sequenced directly from the pSVK3-HNF1 $\alpha$  vector using the 5' PCR primer, and the 3' end of the cDNA was to be sequenced using the 3' PCR primer. The intermediate nucleotides were to be subcloned into the m13mp18 vector, using two unique *Xho* I restriction sites in the pSVK3-HNF1 $\alpha$  vector. The ends of this ~1.8 kb fragment of the cDNA were to be sequenced using the m13 sequencing primers, the unique annealing sites for which are located at either end of the m13 multiple cloning site. Sequencing primers were designed to anneal specifically to internal sequences of the HNF1 $\alpha$  cDNA.

A number of reasons prevented the completion of this sequencing procedure. Firstly, sequencing from the 5' PCR primer did not yield readable sequence even after several attempts. Secondly, the HNF1 $\alpha$  cDNA fragment that was to be cloned into the m13mp18 vector proved to be as difficult to subclone as the entire HNF1 $\alpha$  cDNA itself. The third reason for abandoning this sequencing protocol was the purchase of an automatic sequencer. A new sequencing strategy, using the 5' primer and the new 3' PCR primer, plus four internal primers, was therefore devised (Figure 3.15). This sequencing was carried out on the ABI 377 automatic sequencer, as outlined in Section 2.2.2.

Clone 4 was chosen initially for sequence analysis, but the sequence from the 5' and 893 primers (see Figure 3.15) was not able to be read due to a large number of indistinguishable nucleotides. An internal PCR using the 5' primer and primer 1818 was carried out to amplify the first 1 kb of the HNF1 $\alpha$  cDNA for further sequencing:

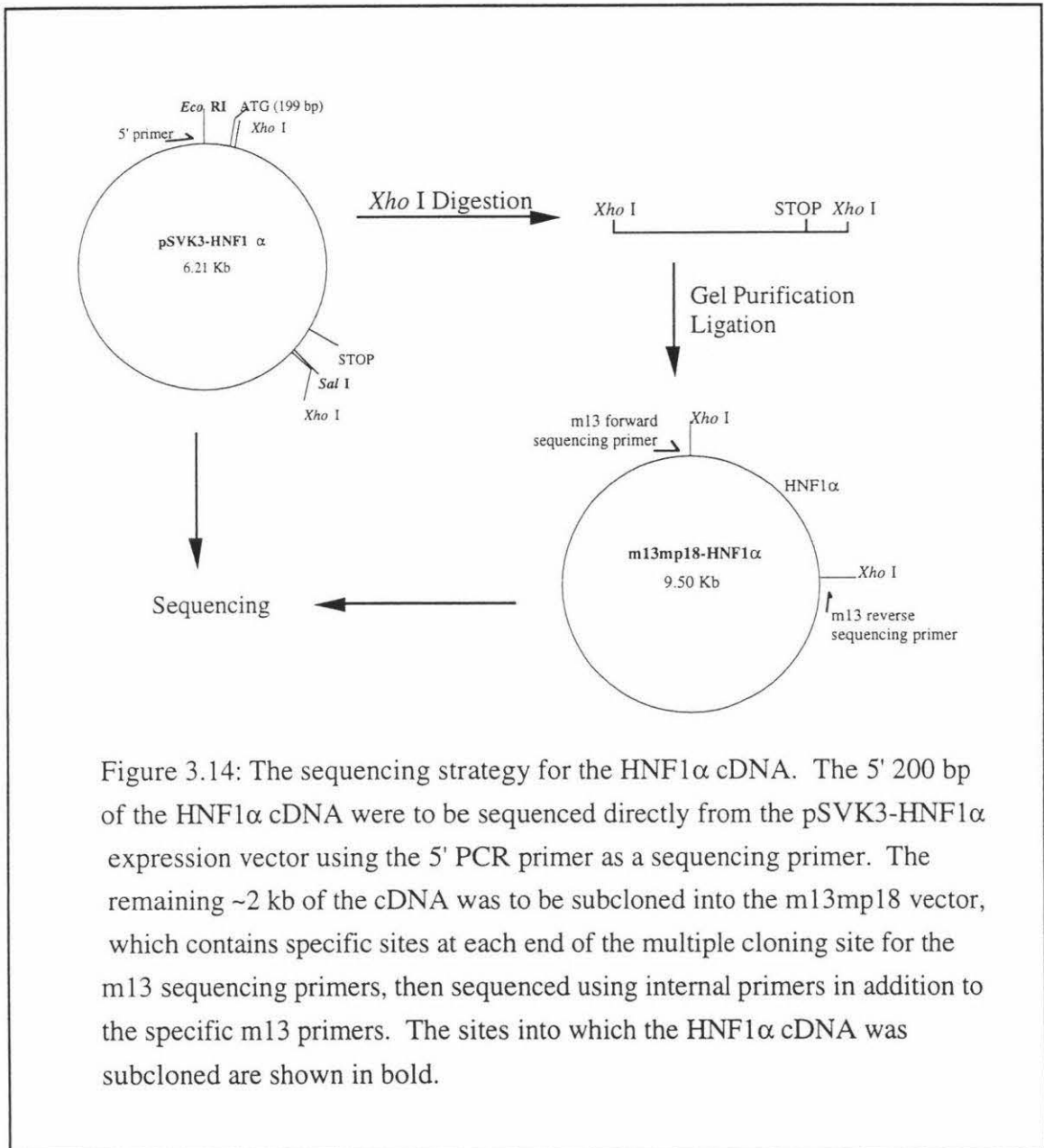


Figure 3.14: The sequencing strategy for the HNF1 $\alpha$  cDNA. The 5' 200 bp of the HNF1 $\alpha$  cDNA were to be sequenced directly from the pSVK3-HNF1 $\alpha$  expression vector using the 5' PCR primer as a sequencing primer. The remaining ~2 kb of the cDNA was to be subcloned into the m13mp18 vector, which contains specific sites at each end of the multiple cloning site for the m13 sequencing primers, then sequenced using internal primers in addition to the specific m13 primers. The sites into which the HNF1 $\alpha$  cDNA was subcloned are shown in bold.



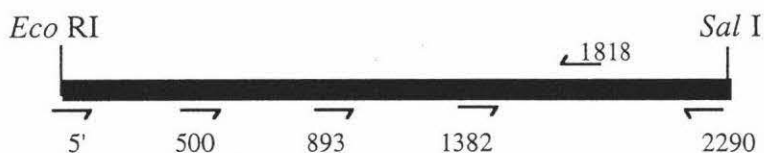


Figure 3.15: The new HNF1 $\alpha$  cDNA sequencing strategy. The HNF1 $\alpha$  cDNA was sequenced directly from the pSVK3-HNF1 $\alpha$  vector using the PCR primers and internal sequencing primers. Only the HNF1 $\alpha$  cDNA is shown in this diagram, with the primers at their annealing sites. The primer names are indicative of the site on the HNF1 $\alpha$  cDNA at which they hybridise, and is the nucleotide number to which the 5' end of the primer anneals.

94°C, 5 minutes

then 94°C, 1 minute

60°C, 1 minute

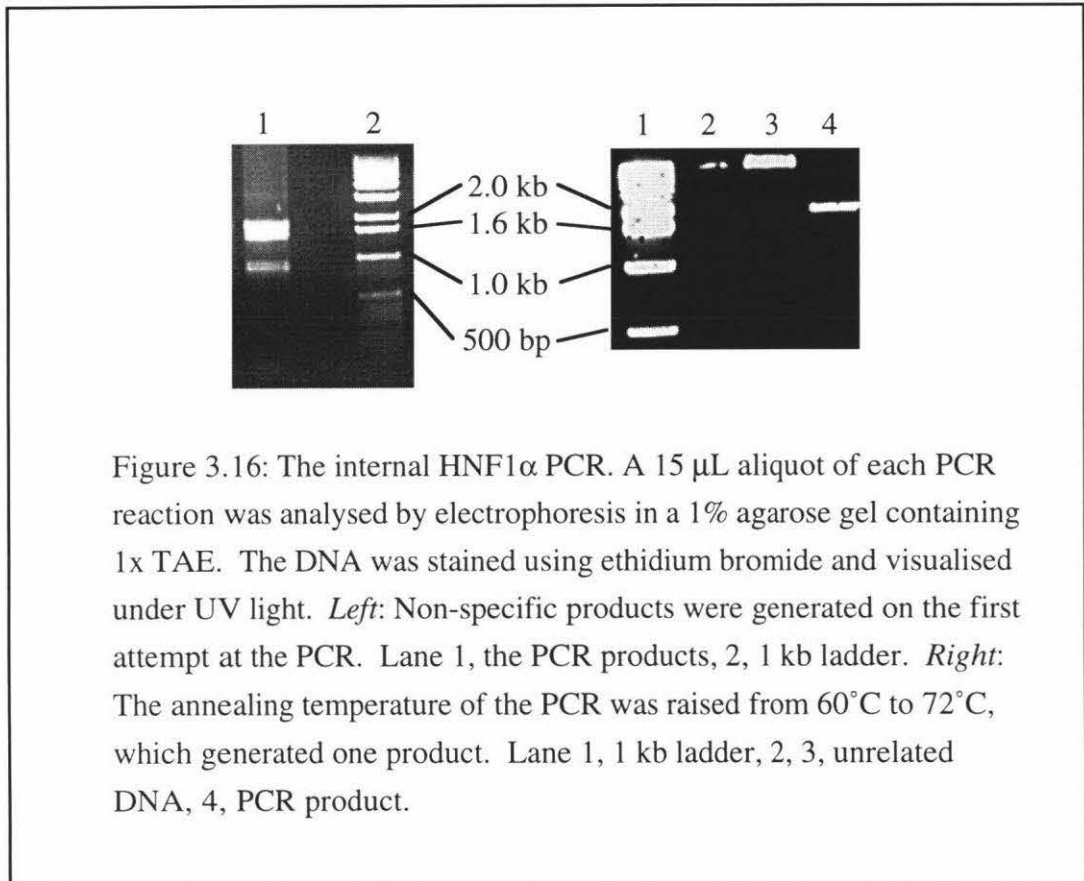
72°C, 3.5 minutes; for 30 cycles.

using 250 ng of the primers, 0.3 mM dNTPs, 4 mM MgCl<sub>2</sub>, 2 ng of pSVK3-HNF1 $\alpha$  clone 4 as template, and 0.5 U of *Pwo* DNA polymerase.

This PCR yielded many non-specific products (Figure 3.16A), and so was repeated using the higher annealing temperature of 72°C. This PCR generated one product of the correct size (Figure 3.16B), which was purified by the Wizard method (Section 2.2.2) and sequenced using the 5' primer and primer 1818.

On the first attempt at sequencing, neither the 5' primer or primer 893 generated any readable sequence. It was thought that this may be due to the presence of secondary structures when the DNA was in single-stranded form, and so DMSO was added to reduce secondary structure formation. A readable sequence was generated from primer 893 by this method, but the 5' primer still did not produce any sequence. At this point, the failure of the 5' primer to sequence was not of any consequence because an error was identified in the sequence from primer 893. Nucleotide 1181, which should be a guanine

(G), had been replaced by a thymine (T), which would have changed the amino acid which this codon codes for from a serine to an isoleucine. A serine is a hydrophilic amino acid, whereas isoleucine is hydrophobic, so it is predicted that this substitution could make a major change in the secondary structure of the resulting protein. Clone 4 was discarded, and clone 2 was sequenced instead.



The sequencing of clone 2 was relatively straightforward in comparison to that of clone 4 (Table 3). There was a possible error identified at position 2058, where the DNA sequence analysis software could not distinguish between a cytosine (C) or an adenine (A) at this position. However, this nucleotide is the third base of the codon and there exists some flexibility as to the nucleotide present at this position (wobble). Both a cytosine or an adenine at this position would result in the incorporation of a serine into the protein. This potential error was therefore ignored.

Primer	Sequencing Result.
5'	would not sequence, even after the addition of DMSO.
500	good sequence from nucleotides 534 to 1077.
893	good sequence from nucleotides 913 to 1481.
1382	good sequence from nucleotide 1425 to the STOP codon. Possible error at nucleotide 2058: C→A. However substitution conserves the amino acid residue (serine).

Table 3: Results of the sequencing of the pSVK3-HNF1 $\alpha$  clone 2.

The major problem with the sequencing of clone 2 was the apparent inability to sequence from the 5' primer (Table 3). The difficulty experienced in the sequencing from this primer with both clones 4 and 2 suggested that some block was present, which was inherent to the 5' end of the HNF1 $\alpha$  cDNA (see Section 3.1.5). A new internal sequencing primer (primer 567) was designed in the reverse direction to sequence back from the 3' end of the cDNA through this block and into the vector. Although the sequence produced from this new primer did not extend past the 5' end of the cDNA, nucleotides 146 to 544 were sequenced properly and were found to contain no errors. As the transcription start point of the HNF1 $\alpha$  cDNA lies at position +199, it was not deemed important to try to sequence further towards the 5' end. The fact that the cDNA could not be sequenced any further than nucleotide 146 suggests that the possible block at the 5' end lies between this point and the 5' primer site (Section 3.1.5). Clone 2 DNA was then prepared on a large scale using the Qiagen columns (Section 2.2.1), which yielded sufficient for subsequent experiments.

### 3.1.5 Discussion of the Subcloning Difficulties with the HNF1 $\alpha$ cDNA

A number of difficulties were experienced in the subcloning of the HNF1 $\alpha$  cDNA from the Bluescript vector to the pSVK3 expression vector (Section 3.1). Such problems included partial DNA digestion by *Eco* RI and ineffective DNA sequencing. There are several possible explanations for these difficulties, as discussed below:

#### *Eco* RI Restriction Endonuclease Digests

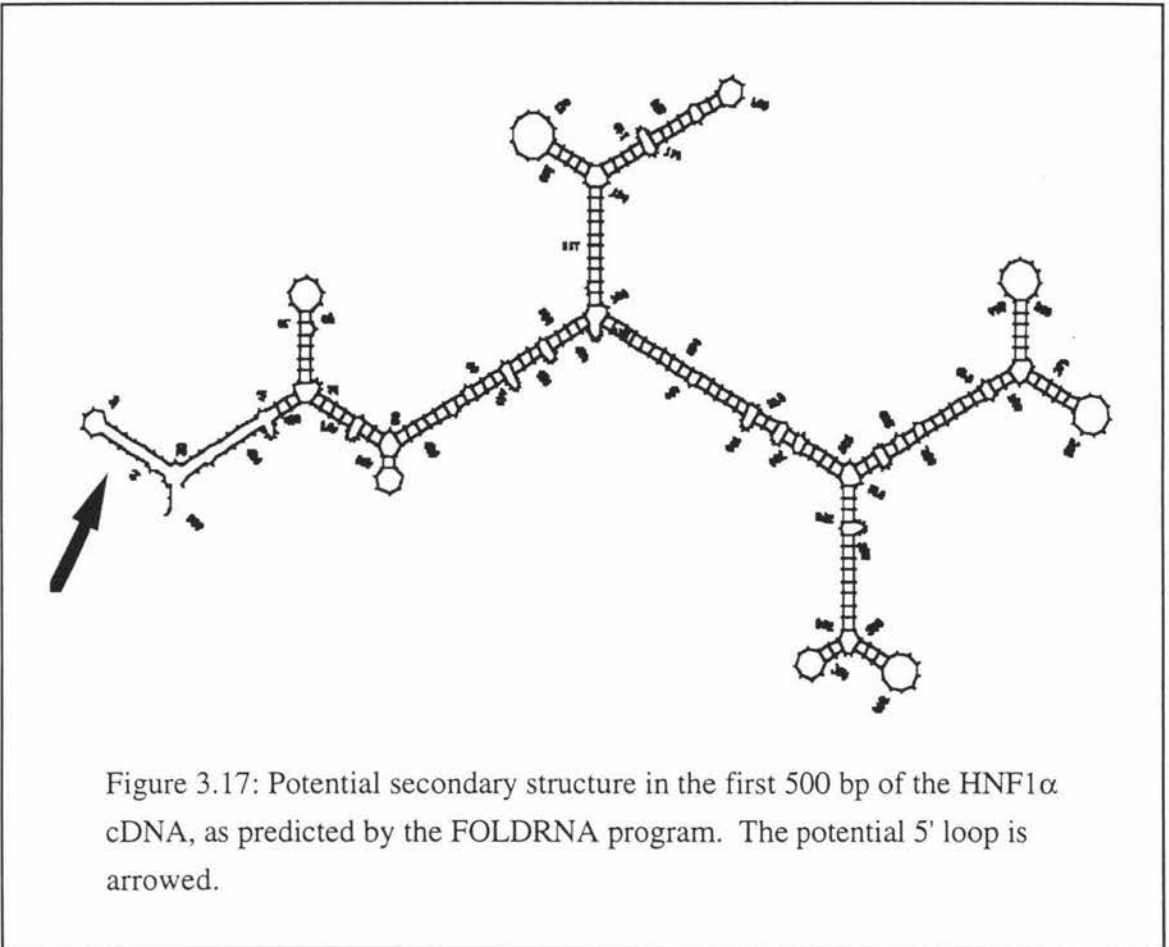
The initial phase of the subcloning scheme involved the digestion of the pBluescript-HNF1 $\alpha$  construct with *Eco* RI restriction endonuclease. As outlined in Section 3.1.1, this plasmid would only digest partially at best, despite the extension of incubation times and the raising of enzyme-to-plasmid ratios. Inhibition of this process may have arisen from site preference by the restriction endonuclease, or possibly from the sequences surrounding the recognition site, including the possible presence of a cruciform.

It has been noted that some restriction endonucleases, including *Eco* RI, exhibit a preference for certain cleavage sites within a substrate. This was initially documented in the bacteriophage lambda, which contains 5 *Eco* RI sites on one DNA molecule<sup>60</sup>. The restriction sites at the 3' terminus of the lambda genome are cleaved preferentially; often up to 10 times faster than those sites in the middle of the molecule<sup>60</sup>. The biochemical basis for this cleavage site bias has not been shown, but it is known not to be due to DNA methylation. The inability of *Eco* RI to cleave the HNF1 $\alpha$  cDNA from the pBluescript vector may also be an example of restriction enzyme site preference. Alternatively, a cruciform was present at the 5' end of the cDNA (see below) may also inhibit the activity of the *Eco* RI enzyme on the DNA.

#### *DNA Sequencing and PCR*

Following the failure of the enzyme digestions, PCR was employed to amplify the HNF1 $\alpha$  cDNA directly from the pBluescript construct. After an initial problem with the 3' primer (Section 3.1.3), the PCR was performed successfully and reproducibly. The PCR product was then ligated into the pSVK3 vector and sequenced for potential errors. However, sequencing from the 5' primer was blocked in some manner, possibly due to the presence of a secondary DNA structure. Any such structure must lie within the first 500 bp of this cDNA, as no other sequencing reaction was inhibited. The GCG

program FOLDRNA was used to predict the secondary structure of the first 500 bp of the coding strand of the HNF1 $\alpha$  cDNA (Figure 3.17). Of those loops predicted, only the first loop (arrowed) appears feasible, as it is the only cruciform predicted which utilises proximal sequences. All the other hairpins present would require distal sequences, presumably unavailable for inclusion in cruciform structures during the PCR and sequencing processes. This configuration contains 22 bp of DNA between bases +8 and +30 of the cDNA, and completely contains the sequence to which the 5' PCR and sequencing primer anneals.



The existence of cruciforms *in vivo* has for a long time been controversial, until recently when stable cruciforms were identified in *E. coli*<sup>61</sup> and *S. cerevisiae*. This rudimentary eukaryote even possesses a mitochondrial cruciform cutting endonuclease (CCE1)<sup>62</sup>. The presence and structural aspects of known cruciforms are dependent on the DNA superhelicity in the cell, and may even be related to some cellular processes<sup>63</sup>. *In vitro*, cruciform formation occurs with increasing superhelical density caused by an

increase in cation concentration to between 5 and 50 mM<sup>64</sup>. Stable cruciforms are also thought to contain loops of minimal size and have relatively short stem lengths<sup>65</sup>. On the basis of these requirements for cruciform formation, it is possible that the loop predicted on the HNF1 $\alpha$  cDNA can actually exist, as the physical characteristics of this loop lie within the range for cruciforms. The reaction buffers used for the PCR and sequencing amplifications all contain Mg<sup>2+</sup> concentrations of at least 5 mM, theoretically sufficient to induce cruciform formation. The plausibility of this cruciform may be investigated by calculating the free energy for the formation of this structure. Free energy is a thermodynamic constant that represents the amount of energy required for or released by a reaction, and is measured in kJ/mol. Energy must be released overall to form a cruciform, and the cruciform therefore must have a negative free energy value. The more negative the free energy of a structure, the more likely it is to form. The calculated free energy of the secondary structure in the HNF1 $\alpha$  cDNA (Figure 3.17, arrowed) is -53.2 kJ/mol. This means that the cruciform may be stable, however the likelihood of it forming *in vivo* cannot be predicted.

However, the PCR and sequencing reactions both use the same 5' primer, creating an obvious paradox when the PCR reaction produces products yet the sequencing reaction at the 5' end fails. These two processes require different primer annealing and extension temperatures, as shown in Table 4, which may effect the presence of the cruciform, depending on its intrinsic melting temperature. At 95°C all DNA secondary structure, including the double stranded helix itself, has been melted. When the temperature is then dropped to allow primer annealing, the double helix is regenerated, and there is a chance that the cruciform may also reform. This may be occurring at 50°C. If the cruciform is present, primer annealing is prevented, which would explain the absence of sequencing products.

Temperatures (°C)	PCR	Sequencing
Denaturation	96°C	95°C
Primer Annealing	60°C	50°C
Primer Extension	72°C	60°C

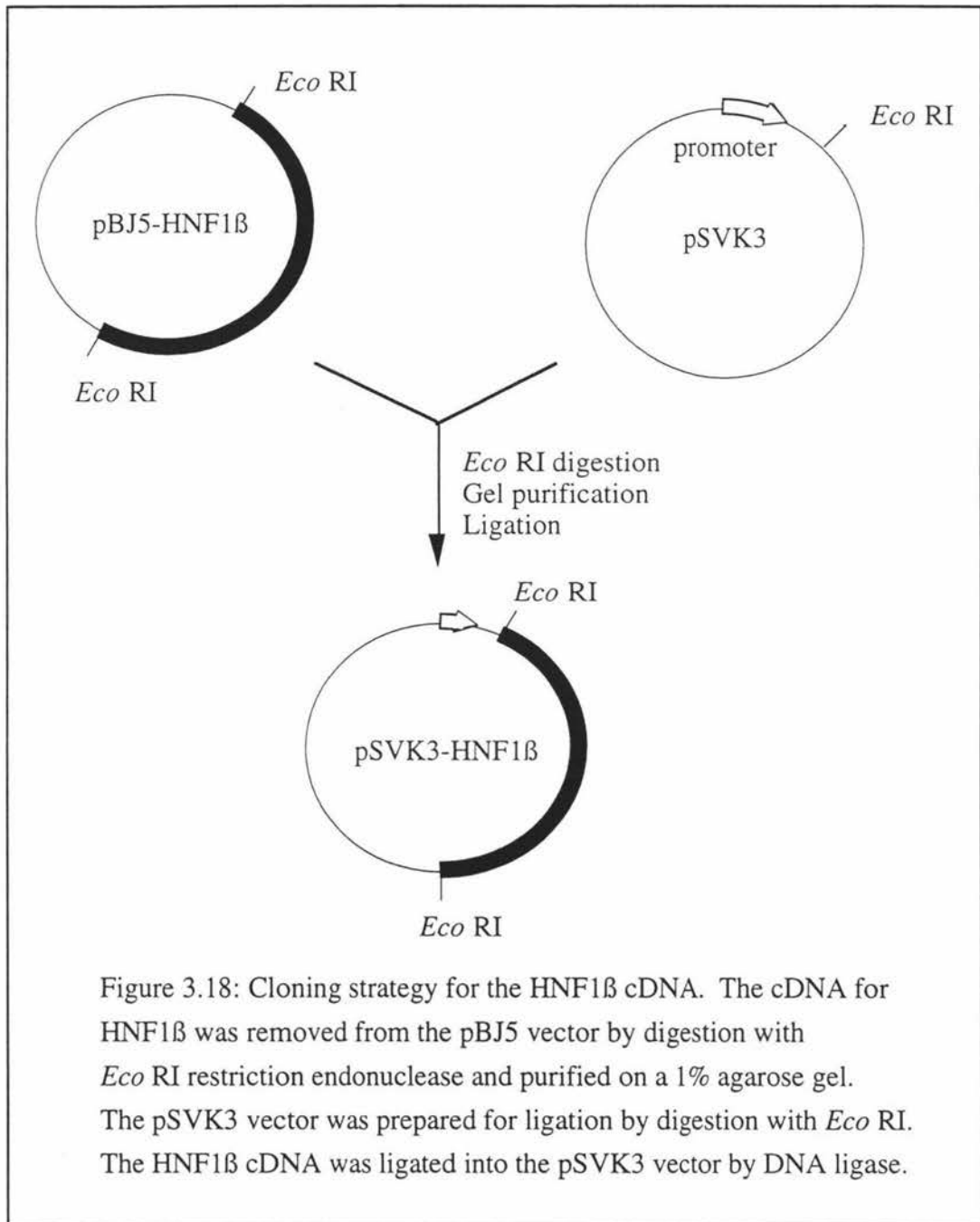
Table 4: Comparison of PCR and Sequencing Conditions.

However, a 60°C annealing temperature, such as in the PCR, may prevent the melted cruciform from reforming, therefore permitting primer annealing and subsequent DNA replication.

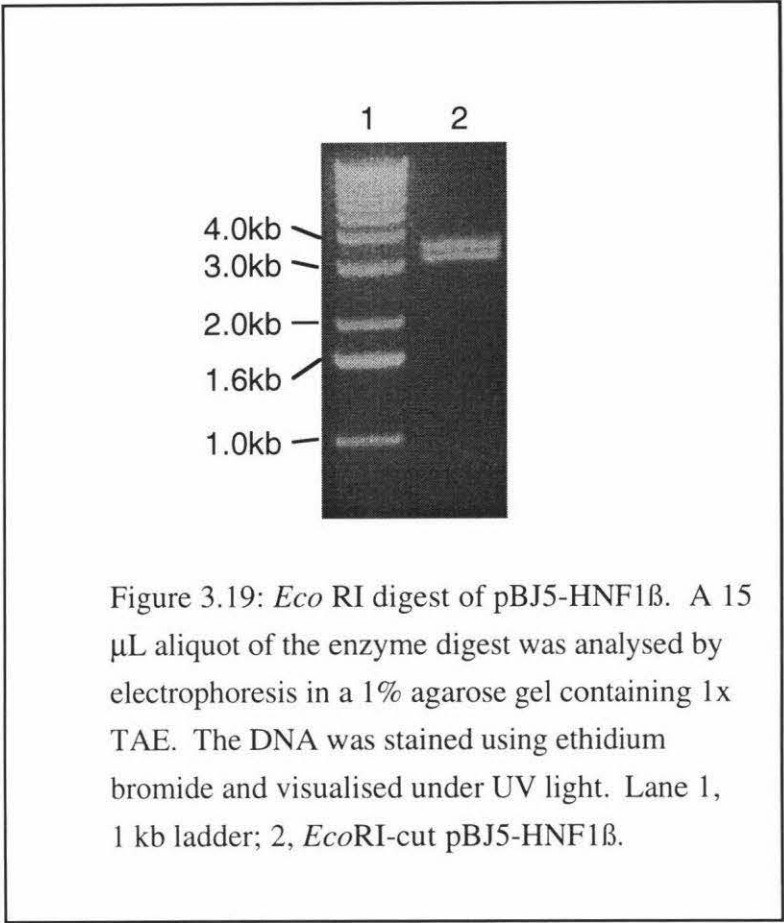
## 3.2 pSVK3-HNF1 $\beta$ Subcloning

### 3.2.1 Restriction Enzyme Digests of the pBJ5-HNF1 $\beta$ Vector

The cDNA for HNF1 $\beta$  was subcloned from the mammalian expression vector pBJ5-HNF1 $\beta$  into pSVK3, to be consistent with the other expression constructs (Figure 3.18).



To this end, HNF1 $\beta$  was excised from pBJ5 with restriction endonuclease *Eco* RI (see restriction map, Appendix A) and identified on the basis of size (pBJ5 vector is larger than HNF1 $\beta$  cDNA), (Figure 3.19).

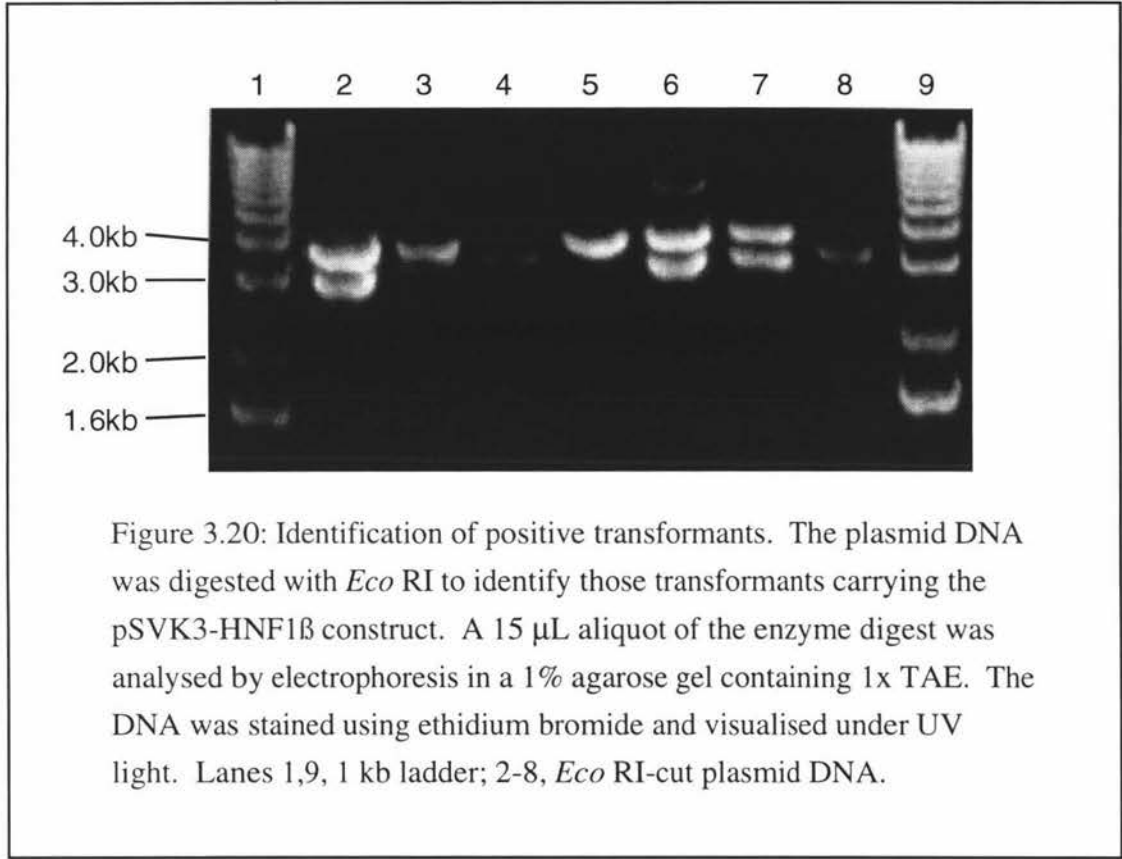


### 3.2.2 Cloning of the HNF1 $\beta$ cDNA

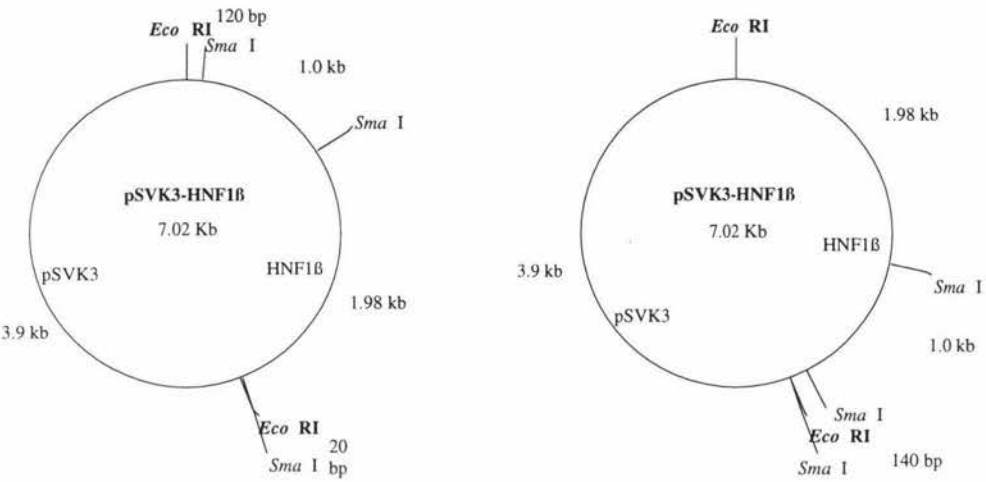
The cDNA was gel purified and ligated into the pSVK3 vector, which had been digested with *Eco* RI and dephosphorylated. The ligation reactions were used to transform competent *E. coli* XL-1 cells, as described in section 2.2.1. Three colonies containing the pSVK3-HNF1 $\beta$  construct were identified by *Eco* RI digestion of plasmid DNA prepared by the rapid boil method, followed by electrophoresis and visualisation under UV light (Figure 3.20). Those plasmids carrying the HNF1 $\beta$  cDNA were expected to be digested by *Eco* RI into two bands of sizes 3 kb and 4 kb (Figure 3.20, lanes 2, 6 and 7).



Insert orientation of these constructs was then examined using the restriction enzyme *Sma* I. A correct orientation yields bands of 1, 2 and 4 kb, whereas a construct with an incorrectly oriented insert will give bands of 1 and 6 kb (Figure 3.21A), when visualised following electrophoresis on a 1% agarose gel (Figure 3.21B). The correct constructs were then propagated to high levels in *E. coli* and purified by the caesium chloride method, as outlined in Section 2.2.1. DNA sequencing of this plasmid was not required as a simple subcloning strategy was used.



A.



B.

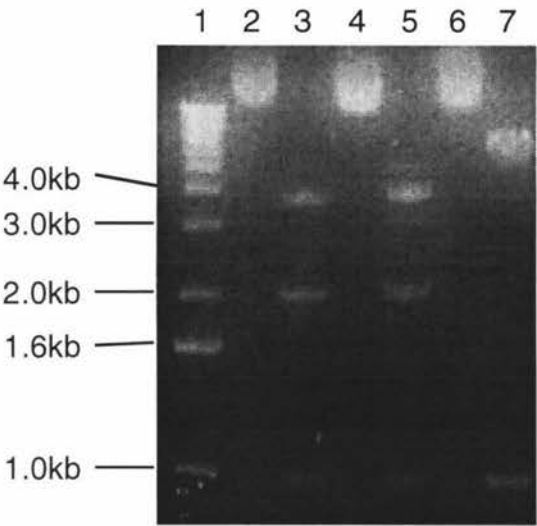


Figure 3.21: *Sma* I diagnostic digest of the three positive plasmids. A: Restriction maps of the *Sma* I sites in the two possible pSVK3-HNF1 $\beta$  constructs. *Left*: the construct with the insert in the correct orientation; *right*: the construct with the insert in the incorrect orientation. B: The plasmid DNA was digested with *Sma* I to check the orientation of the insert in the pSVK3-HNF1 $\beta$  construct. A 15  $\mu$ L aliquot of the enzyme digest was analysed by electrophoresis in a 1% agarose gel containing 1x TAE. The DNA was stained using ethidium bromide and visualised under UV light. Lane 1, 1 kb ladder; 2, 4, 6, uncut pSVK3-HNF1 $\beta$ ; 3, 5, *Sma* I-cut pSVK3-HNF1 $\beta$ , correct orientation; 7, *Sma* I-cut pSVK3-HNF1 $\beta$ , incorrect orientation.

### 3.3. Preparation of DCoH Expression Construct

#### 3.3.1 PCR

The DCoH cDNA was obtained in a GST fusion vector (pGEX-2T), which is designed to produce the protein as a chimera with the GST protein. The activity of the co-factor in transfected tissue culture cells may be affected by the presence of this protein tag, and so the DCoH cDNA alone was amplified from the pGEX-2T vector (see Appendix A) by PCR. *Taq* polymerase was used in conjunction with the template and DCoH-specific primers to generate a 312 bp product which could be observed under UV light following electrophoresis (Figure 3.22, lane 2).

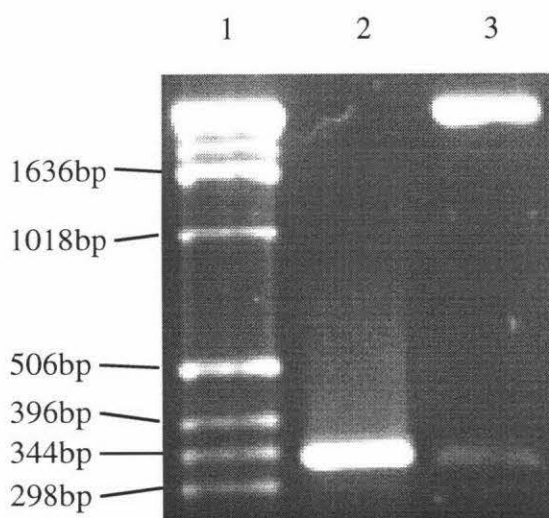


Figure 3.22: Cloning of the DCoH PCR product. A 15  $\mu$ L aliquot of either the PCR reaction or the enzyme digest was analysed by electrophoresis in a 1% agarose gel containing 1x TAE. The DNA was stained using ethidium bromide and visualised under UV light. Lane 1, 1 kb ladder; 2, the 312 bp DCoH PCR product; 3, *Xba* I, *Xho* I-cut pSVK3-DCoH.

### 3.3.2 Cloning of the DCoH cDNA

The 5' and 3' primers used in the PCR reaction contained recognition sites for the restriction endonucleases *Xba* I and *Xho* I, respectively (Figure 3.23). Therefore the PCR product was prepared for cloning by simultaneous digestion with these two restriction endonucleases. Competent XL-1 cells were transformed with ligation mixtures, and positive colonies were identified by re-digestion of purified DNA with *Xba* I and *Xho* I (Figure 3.22, lane 3).

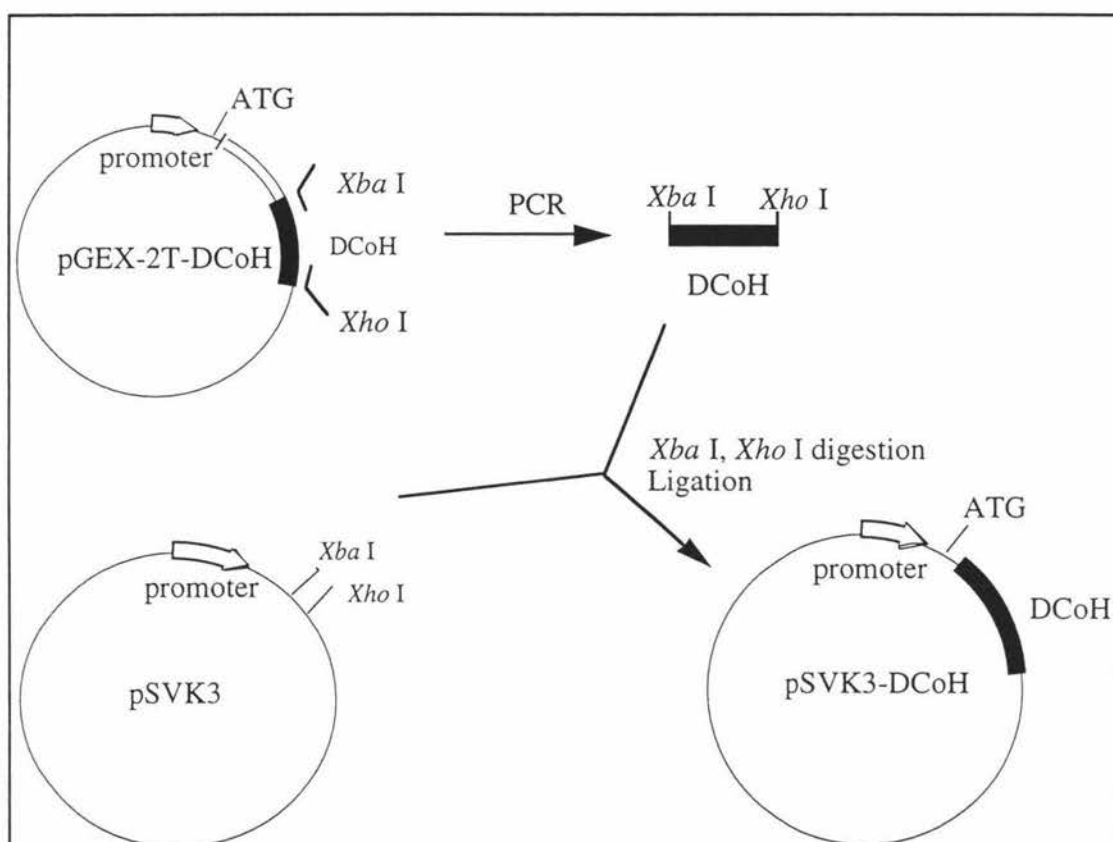


Figure 3.23: Cloning strategy for the DCoH cDNA. The cDNA for DCoH was amplified from the pGEX vector by PCR using primers homologous to the ends of the cDNA sequence. The forward primer contained a cleavage site for *Xba* I and the reverse primer contained the cleavage site for *Xho* I. The PCR product and the pSVK3 vector were prepared for ligation by digestion with *Xba* I and *Xho* I. The DCoH cDNA was directionally cloned into the pSVK3 vector.

### 3.3.3 Sequencing of the pSVK3-DCoH Construct

The DCoH cDNA which was subcloned into the pSVK3 vector (Section 3.3.2) was generated by PCR, and so needed to be analysed for errors by DNA sequencing. The sequencing was initially carried out manually using the Sequenase™ Version 2.0 DNA sequencing kit from US Biochemical (see Section 2.2.2). The cDNA was sequenced directly from the pSVK3 expression vector using the DCoH PCR primers as sequencing primers. The double-stranded pSVK3-DCoH vector was denatured to single-stranded sequencing templates by the alkaline-denaturation method and sequenced as described in the protocol supplied with the sequencing kit (Section 2.2.2). The sequence generated from both the 5' and the 3' primers was 182 nucleotides in length and contained no errors (Figure 3.24). The cDNA for DCoH is 312 nucleotides long, and so the first sequencing run generated only ~60% of the complete sequence. Manganese buffer, containing 4.5 mM  $\text{MnCl}_2$ , was included in the second sequencing experiment to facilitate sequencing close to the primers. 274 nucleotides of sequence were produced by this reaction, representing 88% of the total sequence, which was error-free. An example of the sequence generated by this method is shown in Figure 3.24. Only the nucleotides located at the extreme 5' and 3' ends of the DCoH cDNA had not been sequenced. These remaining nucleotides were sequenced on the ABI 377 automatic sequencer, using the 5' and 3' PCR primers as sequencing primers. Therefore the entire 312 bp of the DCoH cDNA was eventually analysed and shown to be error-free.

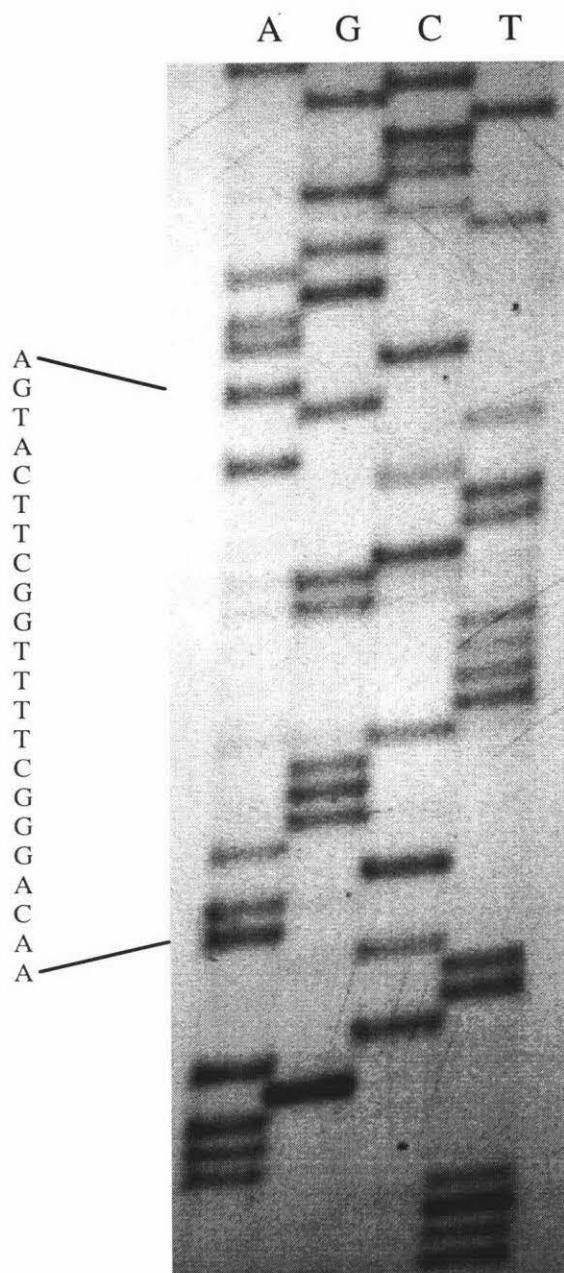


Figure 3.24: Sample of the manual sequencing of the DCoH cDNA. The sequencing reactions were carried out according to the manufacturer's instructions and run on a 6% denaturing polyacrylamide gel in 1x TBE. Following electrophoresis, the gel was fixed in ethanol/acetic acid solution, dried and exposed to X-ray film overnight. The lanes are labelled according to which dideoxy chain terminating nucleotide the reaction contains.

### 3.3.4 Summary of Part A

The three cDNAs for HNF1 $\alpha$ , HNF1 $\beta$  and DCoH were subcloned from their respective vectors into the mammalian expression vector pSVK3. Each clone was verified for authenticity by either DNA sequence analysis or by restriction endonuclease digestion. These expression constructs were used in mobility shift assays and transcriptional assays, the details of which form the contents of parts B and C respectively. A large amount of time was consumed in the completion of these constructs, especially the subcloning of the HNF1 $\alpha$  cDNA. Unfortunately this has meant that the time allocated to the following sections was less than optimal and as a result, these areas of research have not been investigated as thoroughly as would have been preferred.

## Part B: Binding of HNF1 Proteins to the Factor IX Promoter

### 3.4 Introduction to Gel Shifts

A gel shift, or electrophoretic mobility shift assay (EMSA), is a technique used to detect the binding of protein to linear nucleic acids. The method is based on the observation that this binding usually leads to a reduction in the electrophoretic mobility of the DNA (or RNA) fragment through a polyacrylamide gel. The protein-DNA complex is therefore retarded in the gel relative to the naked oligonucleotide. Figure 3.25 outlines the steps in a gel shift. The oligonucleotide is commonly end-labelled with  $^{32}\text{P}$  isotope so that autoradiography may be used to visualise any change in electrophoretic mobility.

The protein assayed for its DNA-binding ability may be purified, or present in high concentrations in a whole cell extract. The raised levels of protein required can be produced in mammalian cell lines, such as Cos7 and Alexander by synthesis from a transient expression vector containing the cDNA for the protein. DNA oligonucleotides containing a known binding site for the desired protein are used in a gel shift to confirm the protein's presence. Non-specific proteins present in extracts should not interact with the labelled DNA, and their effects are therefore not normally observed.

Competition gel shifts may be used to test the specificity that a protein has for a DNA fragment of interest. In the case of transcription studies, this oligonucleotide usually contains a promoter sequence. Unlabelled oligonucleotides containing known DNA binding sites for the protein of interest, or a variant of the binding site, are added to the gel shift reaction. The ability of competing DNA fragment to bind the protein, relative to the labelled oligonucleotide, may be observed as a decrease in band intensity on the autoradiograph. Increasing amounts of competitor should relate to decreasing band intensities.

In this section, gel shifts are used to assay the binding of HNF1 $\alpha$  and HNF1 $\beta$  homo-and hetero-dimers (with and without DCoH) to the factor IX promoter. The general vicinity of the promoter to which the proteins bind will be detected using mutated promoters as competing oligonucleotides.



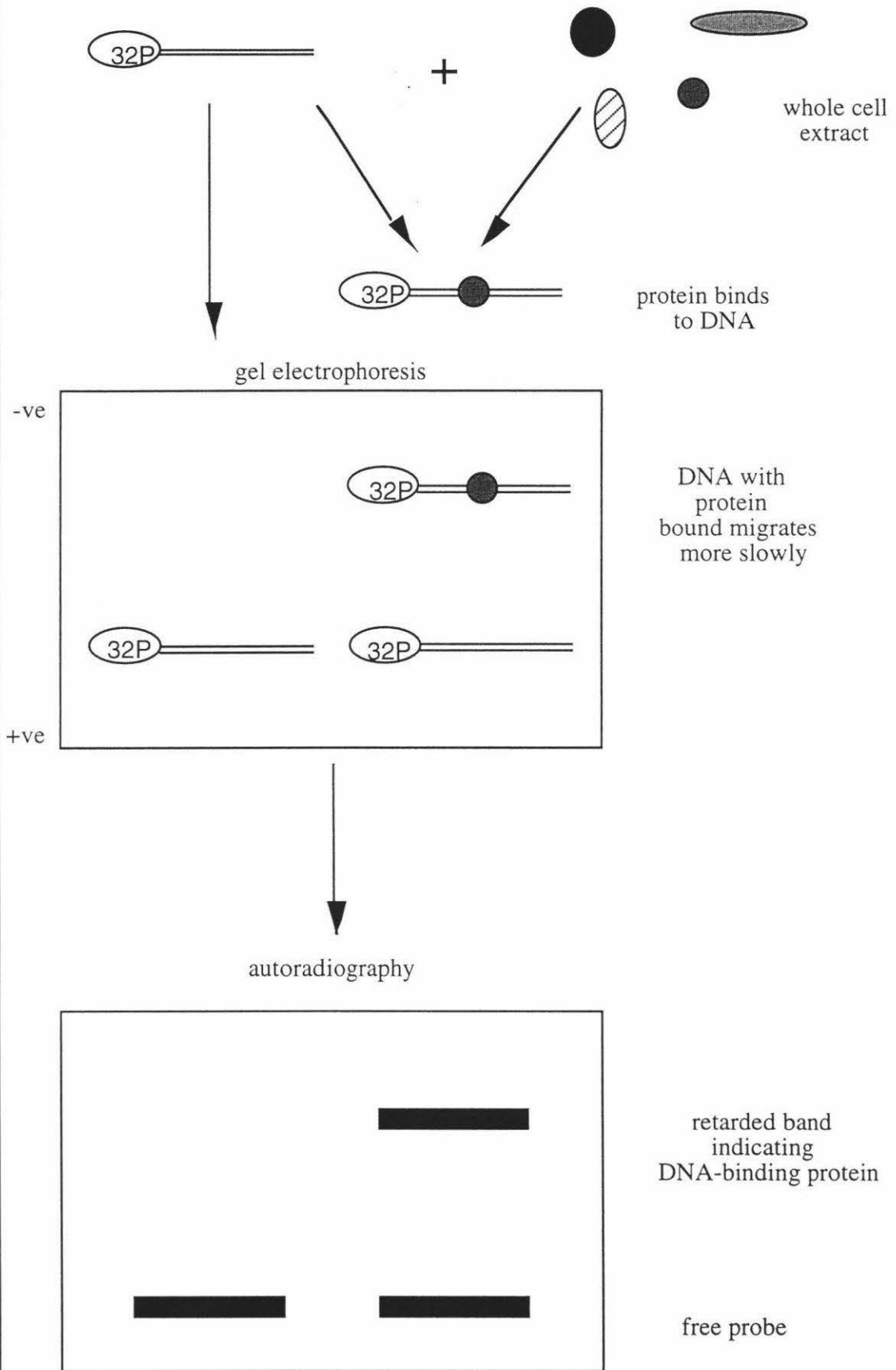


Figure 3.25: Schematic diagram of a gel shift.

### 3.5 Gel Shifts with Rat Liver Nuclear Extracts

#### 3.5.1 Endogenous Rat Liver Proteins Bind to the Wildtype Factor IX Promoter

The ability of transcription factors in a nuclear extract to bind specifically to the factor IX promoter was tested directly by gel mobility shift assay (Section 2.2.4). Nuclear extracts were prepared from adult rat livers as described in Section 2.2.3. The extract, which contains the endogenous transcription factors present in the rat liver at the time of preparation, was incubated with  $^{32}\text{P}$  radiolabelled oligonucleotide representing the -14 to +6 region of the wildtype Factor IX promoter. Three major gel-retarded protein-DNA complexes were observed after electrophoresis (Figure 3.26, lane 2).

The remaining lanes on the gel show the effect of conducting this standard experiment in the presence of various competitor oligonucleotides (Appendix B). Firstly, the specificity of each protein complex for the -14 to +6 region of the wildtype factor IX oligonucleotide was tested. Increasing amounts of the unlabelled factor IX oligonucleotide were included in the reactions (Figure 3.26, lanes 3-5) prior to the addition of the labelled probe. A 100x excess of competitor (50 ng, lane 4), resulted in the disappearance of all three retarded bands, indicating that the proteins specifically bound to the cold promoter sequence rather than the labelled probe. An unrelated oligonucleotide containing the YY1 consensus sequence of an adenovirus gene was also used as a competitor (lanes 7-9). Even at a 200x competitor excess (100 ng), the proteins of these complexes preferentially bind the probe. Together, these results confirm that the proteins present in the three retarded bands bind specifically to the wildtype factor IX promoter region.

A third oligonucleotide, representing the HNF1 $\alpha$  binding site from the  $\alpha$ 1-antitrypsin promoter, was also used as a competitor (lanes 11-13). At a 5-fold competitor excess (5 ng, lane 11), the retardation pattern is virtually identical to that of the uncompetited shift (lane 2). However, when the competitor concentration is increased to 50 and 100 ng, complex II was diminished (lanes 12 and 13). This result suggests that an HNF1 protein may be present in complex II. As the DNA binding form of the HNF1 proteins is dimeric, it is possible that a protein present in complex II is either a homodimer of HNF1 $\alpha$  or HNF1 $\beta$ , or an HNF1 $\alpha$ -HNF1 $\beta$  heterodimer. However, as the endogenous form of HNF1 in the liver has not been fully characterised, the identity of this transcription factor cannot be determined by this method.

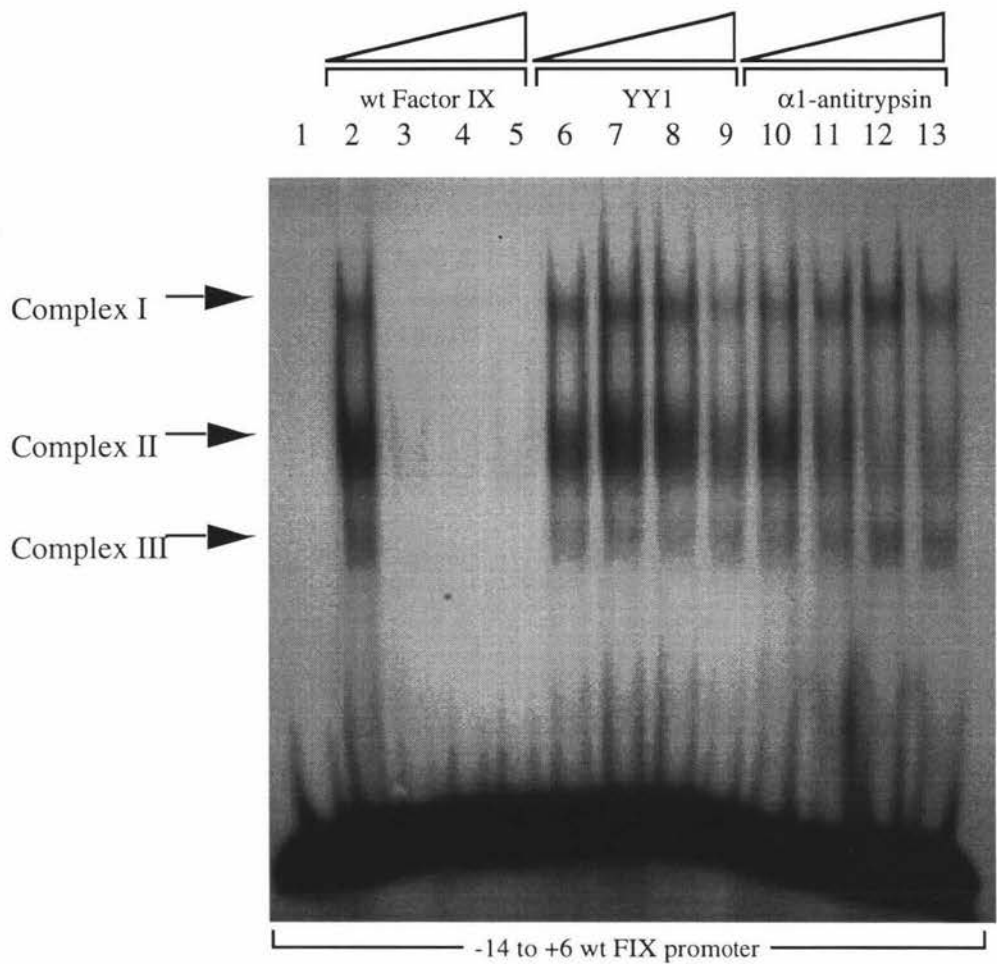


Figure 3.26: Competition gel shift 1 of rat liver nuclear extract with the wildtype factor IX promoter. ~0.5 ng of  $^{32}\text{P}$  labelled oligonucleotide (~9000 cpm) representing the -14 to +6 region of the factor IX promoter was incubated with 1  $\mu\text{g}$  of poly (dI.dC) and ~10  $\mu\text{g}$  of rat liver nuclear extract. Following incubation, all samples were subjected to electrophoresis on a 4% polyacrylamide gel in 0.25x TBE. The gel was dried onto DE-81 paper and exposed to X-ray film overnight at  $-70^{\circ}\text{C}$ . Lane 1, probe only. Control competitor oligonucleotides: wildtype factor IX (lanes 2-5: 2, 0 ng, 3, 5 ng, 4, 50 ng, 5, 100 ng), YY1 site from an adenovirus gene, (lanes 6-9: 6, 0 ng, 7, 5 ng, 8, 50 ng, 9, 100 ng), HNF1 site from  $\alpha 1$ -antitrypsin (lanes 10-13: 10, 0 ng, 11, 5 ng, 12, 50 ng, 13, 100 ng).

### 3.5.2 HNF1 May be Interacting at the -5,-6 Region of the Factor IX

#### Promoter

An HNF1 protein may interact with the factor IX wildtype promoter, as shown in Figure 3.26. By using the -14 to +6 region of the four mutant -5 and -6 factor IX promoter regions as competitors, the importance of specific nucleotides could be assessed. 10, 100 and 200-fold excesses of each mutant promoter region were included as competitors of protein binding in gel shift reactions using the labelled wildtype factor IX oligonucleotide as a probe. The effects of these competitors were observed by native 4% PAGE followed by autoradiography (Figure 3.27).

The factor IX promoter regions containing the -5A/T and -5A/G mutations (Figure 3.27, lanes 2-5 and 6-9, respectively), and the -6G/A mutation (lanes 10-13), did not compete for the protein or proteins binding to the probe, as shown by the recurring retardation pattern. This indicates that these mutations may interfere with the DNA-protein interactions, suggesting that the proteins present in complex II specifically binds to nucleotides in this region.

A different binding pattern was observed, however, for the competitor containing a -6G/C mutation. At 10x excess competitor, complex III is still observed and complex I is present in all conditions except the 10x excess. The reason for this is unclear, and could simply be an artefact of the assay system. However, the competitor carrying the -6G/C mutation effectively competes for complex II in all conditions (lanes 15-17) and with complex III at high levels of competitor (lane 17). This result was reproduced several times. It is therefore possible that the factor IX promoter region containing the -6G/C mutation can bind to proteins present in the rat liver nuclear extract. From comparison with Figure 3.26, this protein could be an HNF1 dimer. Comparison of the band shift patterns for each of the factor IX promoter regions revealed that the proteins in complex II have the highest affinity for the wildtype sequence, and a lower affinity for the promoter containing the -6G/C mutation. Within the limits of this experiment, the proteins present in complex II appear to have no affinity for the -6G/A or the -5 mutant promoter sequences. Complex I appears to disappear in a 10-fold excess of the oligonucleotide carrying the -6G/C mutant sequence (Figure 3.27, lane 15) and reappears more strongly at a 100x excess of the same competitor oligonucleotide (lane 16). This apparant anomaly was believed to be due to an uneven loading of the samples and should be viewed as an experimental artefact.

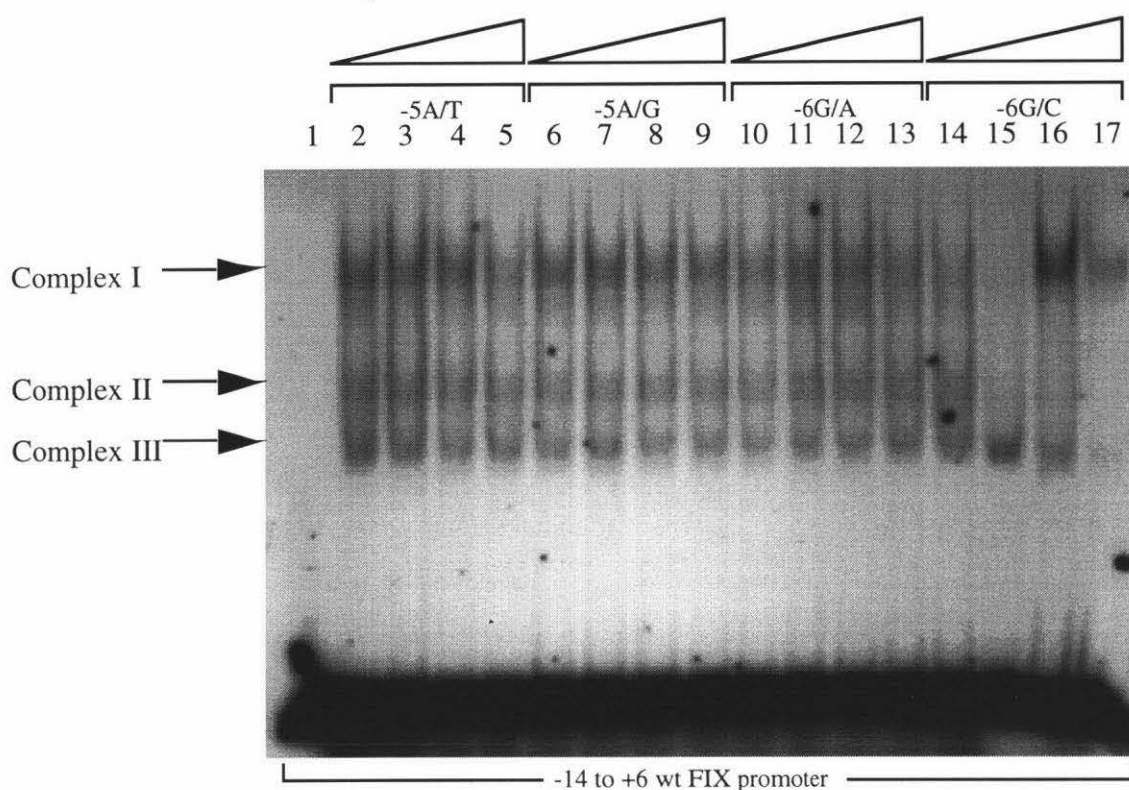


Figure 3.27: Competition gel shift 2 of rat liver nuclear extract with the wildtype factor IX promoter. 0.5 ng of  $^{32}\text{P}$  labelled oligonucleotide representing the -14 to +6 region of the factor IX promoter was incubated with 1  $\mu\text{g}$  of poly (dI.dC) and ~10  $\mu\text{g}$  of rat liver nuclear extract. Following incubation, all samples were subjected to electrophoresis on a 4% polyacrylamide gel in 0.25x TBE. The gel was dried and exposed to X-ray film overnight at  $-70^\circ\text{C}$ . Lane 1, probe only. Mutant factor IX promoter competitor oligonucleotides: -5A/T (lanes 2-5: 2, 0 ng, 3, 5 ng, 4, 50 ng, 5, 100 ng), -5A/G (lanes 6-9: 6, 0 ng, 7, 5 ng, 8, 50 ng, 9, 100 ng), -6G/A (lanes 10-13: 10, 0 ng, 11, 5 ng, 12, 50 ng, 13, 100 ng), -6G/C (lanes 14-17: 14, 0 ng, 15, 5 ng, 16, 50 ng, 17, 100 ng).

The binding exhibited in these gel shifts suggests that an HNF1 dimer may interact in some way with the wildtype and -6G/C factor IX promoter region. The competition for the proteins in complex II by the HNF1 site from the  $\alpha 1$ -antitrypsin promoter is rather weak, suggesting that even though HNF1 may have a role in DNA binding in this region, other proteins are likely to be involved as well.

### 3.5.3 An HNF1 Dimer May Bind at -5,-6 of the Factor IX Promoter

A number of proteins present in a rat liver nuclear extract have been shown to interact with the wildtype factor IX promoter region between -14 and +6, forming three distinct protein-DNA complexes. Most of these complexes disappear in the presence of an excess of unlabelled homologous competitor, suggesting that the proteins in these complexes form specific interactions with the factor IX promoter. One of these complexes (complex II) contains proteins that also interact with the strong HNF1 site from the  $\alpha$ 1-antitrypsin promoter. This suggests that complex II may contain an HNF1 protein, however the binding displayed may also result from an unrelated protein with a binding activity similar to HNF1. Antibodies specific to the HNF1 transcription factors could be used to confirm the presence of these proteins in complex II. However, these antibodies were not available for this study.

## 3.6 Gel Shifts with Cos Cell Extracts

### 3.6.1 Cos Cell Transfections to Produce Enriched Cell Extracts

Gel mobility shift assays can be used to test the ability of a specific transcription factor to bind to a specific DNA sequence. Purified protein is commonly assayed directly, however a solution enriched in a transcription factor can often suffice. Such a solution may be partially purified protein or an extract from cells which have been manipulated to produce the protein at a level greater than would be present in normal cells. This section describes results obtained from gel shifts using extracts from mammalian Cos 7 cells which were transfected with expression vectors for HNF1.

Cos 7 is a monkey kidney fibroblast cell line which may be used to generate whole cell extracts rich in a mammalian protein. These cells were transfected with each of the pSVK3 vectors, which contained the cDNA for HNF1 $\alpha$  or HNF1 $\beta$  under the control of the SV40 promoter (Section 2.2.2). Whole cell extracts of the transfected cultures were tested for the presence of an HNF1 protein by gel shift, using an oligonucleotide representing the high affinity HNF1 site from the  $\alpha$ 1-antitrypsin promoter as a probe. Cos cells were also transfected with HNF1 $\alpha$  and HNF1 $\beta$  expression vectors in an attempt to generate heterodimers. The pSVK3 vector carrying



the cDNA for DCoH was also included in repeats of each transfection so that the effects of the dimerisation co-factor on HNF1 binding to DNA could be studied. The control for the transfections was an extract made from cells subjected to a calcium phosphate precipitate lacking any DNA. This blank should therefore not contain enough HNF1 protein to produce a gel shift with the  $\alpha$ 1-antitrypsin promoter probe unless endogenous HNF1 was produced.

The first Cos cell culture was transfected with the pSVK3-HNF1 $\beta$  vector, which contained the first cDNA to be subcloned (see Section 3.2). The whole cell extract prepared from this transfection produced a specific protein-DNA complex, which was studied in further experiments (Section 3.6.2).

### 3.6.2 Preliminary HNF1 $\beta$ Gel Shifts

At an early stage in the course of the Cos cell transfections, one extract enriched in HNF1 $\beta$  displayed probe binding that was distinct from that of the blank extract. This extract was used in gel shifts to test the affinity of HNF1 $\beta$  for the factor IX promoter (Figure 3.28). The probe used in these gel shifts was the  $^{32}$ P-labelled oligonucleotide representing the HNF1 site from the  $\alpha$ 1-antitrypsin promoter. This probe, which is a high affinity binding site for HNF1, was used initially to test for the presence of HNF1 protein in the cell extract.

This HNF1 $\beta$ -transfected Cos cell extract interacted with the  $\alpha$ 1-antitrypsin probe (not shown). Various oligonucleotide competitors were then used in a gel shift to investigate the affinity of the protein present in the complex for different consensus sequences. Firstly, a competitor homologous to the probe was used to test the specificity of the observed interaction (Figure 3.28, lanes 2-5). The retarded band fades at a 10x excess of competitor, indicating that the protein present in this complex is binding specifically to the HNF1 site from  $\alpha$ 1-antitrypsin. The protein present in this complex is therefore most likely to be HNF1 $\beta$ , as a blank Cos cell extract did not produce a shift and the retarded band was fully competed with an unlabelled homologous probe (Figure 3.28). If the protein present in this complex does represent HNF1 $\beta$ , then it must be present in the form of a homodimer because dimerisation is a prerequisite for DNA-binding<sup>3</sup>.

The HNF1 $\beta$  dimer is also shown to bind with lower affinity to the wildtype factor IX promoter (Figure 3.28, lanes 6-9), as complete competition requires a greater excess of competitor (200x) than is required for the homologous competitor. Therefore, the

HNF1 site from  $\alpha$ 1-antitrypsin has a stronger affinity for the protein than does the -14 to +6 region of the wildtype factor IX promoter. The affinity of the HNF1 $\beta$  dimer for the factor IX promoter region carrying mutations was also tested. The only mutant oligonucleotide which could not compete HNF1 $\beta$  off the probe to any extent was the -5A/T promoter sequence (Figure 3.29, lanes 10-13). All three of the other sequences carrying the natural factor IX promoter mutations could compete HNF1 $\beta$  off the probe, with the -6G/A promoter sequence competing almost as well as the wildtype factor IX promoter competitor (Figure 3.28, lanes 14-17). Also, the -5A/G competitor (Figure 3.29, lanes 14-17) appeared to have an affinity for HNF1 $\beta$  at a level comparable to that of the wildtype factor IX promoter (Figure 3.29, lanes 6-9). The -6G/C promoter sequence could also compete with the probe for HNF1 $\beta$ , but only at a low level. These experiments therefore show that HNF1 $\beta$  displays a range of DNA-binding activities at different sequences, as summarised below:

HNF1 site from $\alpha$ 1- antitrypsin	>	-6G/A mutant Factor IX promoter sequence	$\geq$	wildtype Factor IX and -5A/G promoter sequences	>	-6G/C mutant Factor IX promoter sequence	>	-5A/T mutant Factor IX promoter sequence
--	---	--	--------	---	---	--	---	--

HNF1 $\beta$  has the highest affinity for the site from the  $\alpha$ 1-antitrypsin promoter and displays no apparent affinity for the -5A/T mutant factor IX promoter sequence.

Both of these gel mobility shifts were repeated numerous times, however following the completion of these experiments there was insufficient amounts of the HNF1 $\beta$ -enriched Cos cell extract to conduct further experiments with the factor IX promoter sequence as a probe.



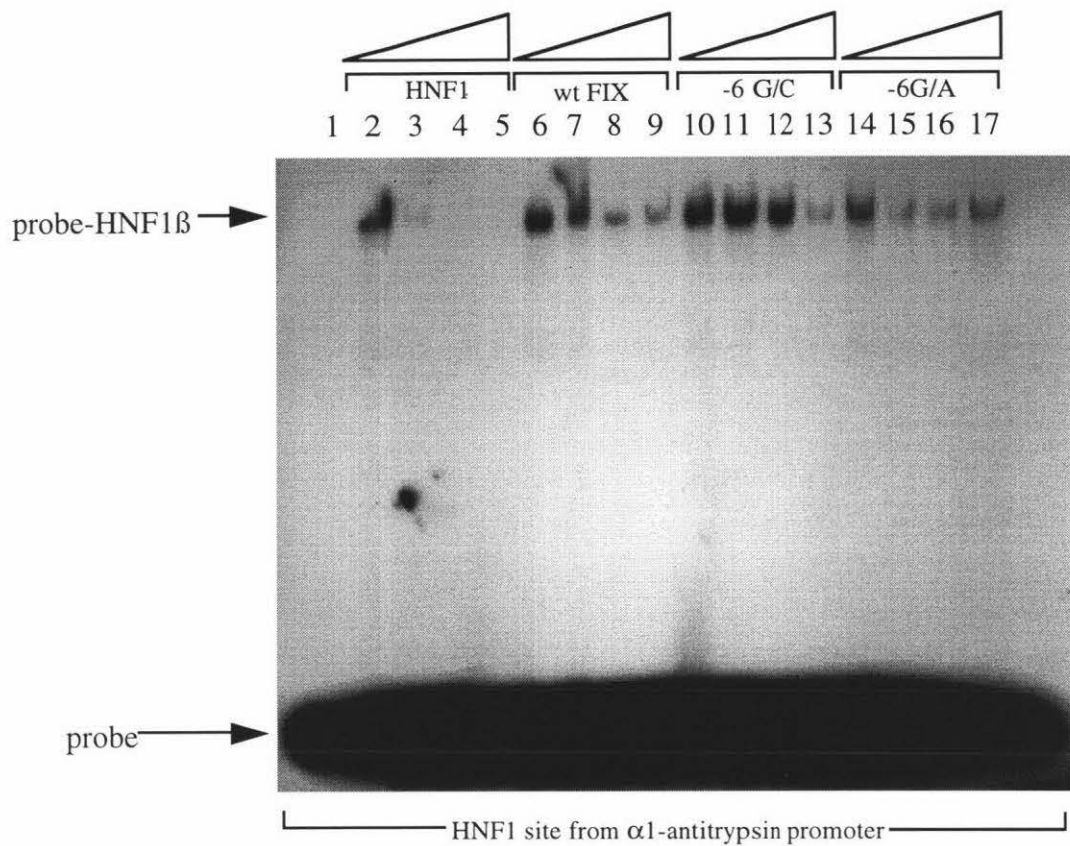


Figure 3.28: Competition gel shift 1 of HNF1 $\beta$ -rich Cos cell extract with the HNF1 site from the  $\alpha$ 1-antitrypsin promoter. 0.5 ng of  $^{32}$ P labelled oligonucleotide representing the -14 to +6 region of the factor IX promoter was incubated with 1  $\mu$ g of poly (dI.dC) and 10  $\mu$ g of rat liver nuclear extract. Following incubation, all samples were subjected to electrophoresis on a 4% polyacrylamide gel in 0.25x TBE. The gel was dried onto DE-81 paper and exposed to X-ray film overnight at -70°C. Lane 1, probe only. Competitor oligonucleotides: HNF1 site from the  $\alpha$ 1-antitrypsin promoter (lanes 2-5: 2, 0 ng; 3, 5 ng; 4, 50 ng; 5, 100 ng), wildtype factor IX promoter (lanes 6-9: 6, 0 ng; 7, 5 ng; 8, 50 ng; 9, 100 ng), -6G/C mutant factor IX promoter (lanes 10-13: 10, 0 ng; 11, 5 ng; 12, 50 ng; 13, 100 ng), -6G/A mutant factor IX promoter (lanes 14-17: 14, 0 ng; 15, 5 ng; 16, 50 ng; 17, 100 ng).

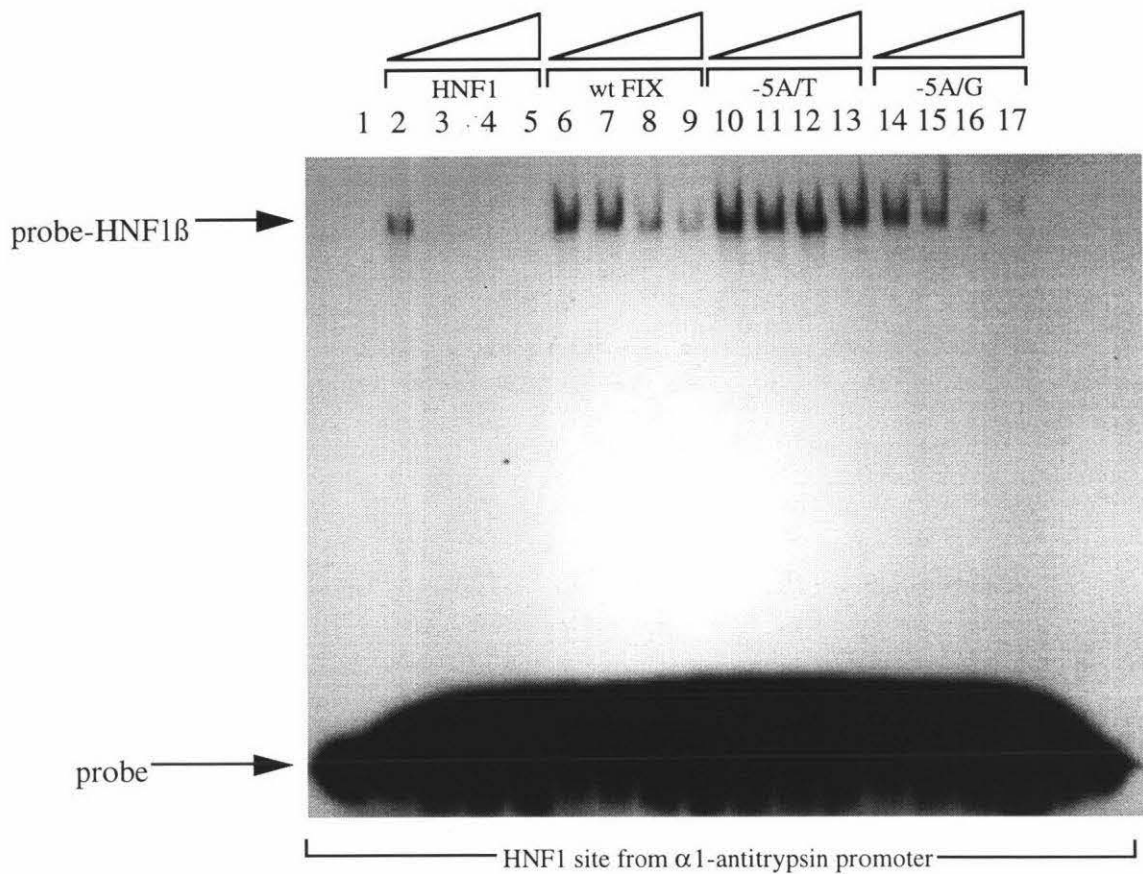


Figure 3.29: Competition gel shift 2 of HNF1 $\beta$ -rich Cos cell extracts with the HNF1 site from the  $\alpha$ 1-antitrypsin promoter. 0.5 ng of  $^{32}$ P labelled oligonucleotide representing the -14 to +6 region of the factor IX promoter was incubated with 1  $\mu$ g of poly (dI.dC) and 10  $\mu$ g of rat liver nuclear extract. Following incubation, all samples were subjected to electrophoresis on a 4% polyacrylamide gel in 0.25x TBE. The gel was dried onto DE-81 paper and exposed to X-ray film overnight at  $-70^{\circ}\text{C}$ . Lane 1, probe only. Competitor oligonucleotides: HNF1 site from the  $\alpha$ 1-antitrypsin promoter (lanes 2-5: 2, 0 ng; 3, 5 ng; 4, 50 ng; 5, 100 ng), wildtype factor IX promoter (lanes 6-9: 6, 0 ng; 7, 5 ng; 8, 50 ng; 9, 100 ng), -5A/T mutant factor IX promoter (lanes 10-13: 10, 0 ng; 11, 5 ng; 12, 50 ng; 13, 100 ng), and -5A/G mutant factor IX promoter (lanes 14-17: 14, 0 ng; 15, 5 ng; 16, 50 ng; 17, 100 ng).

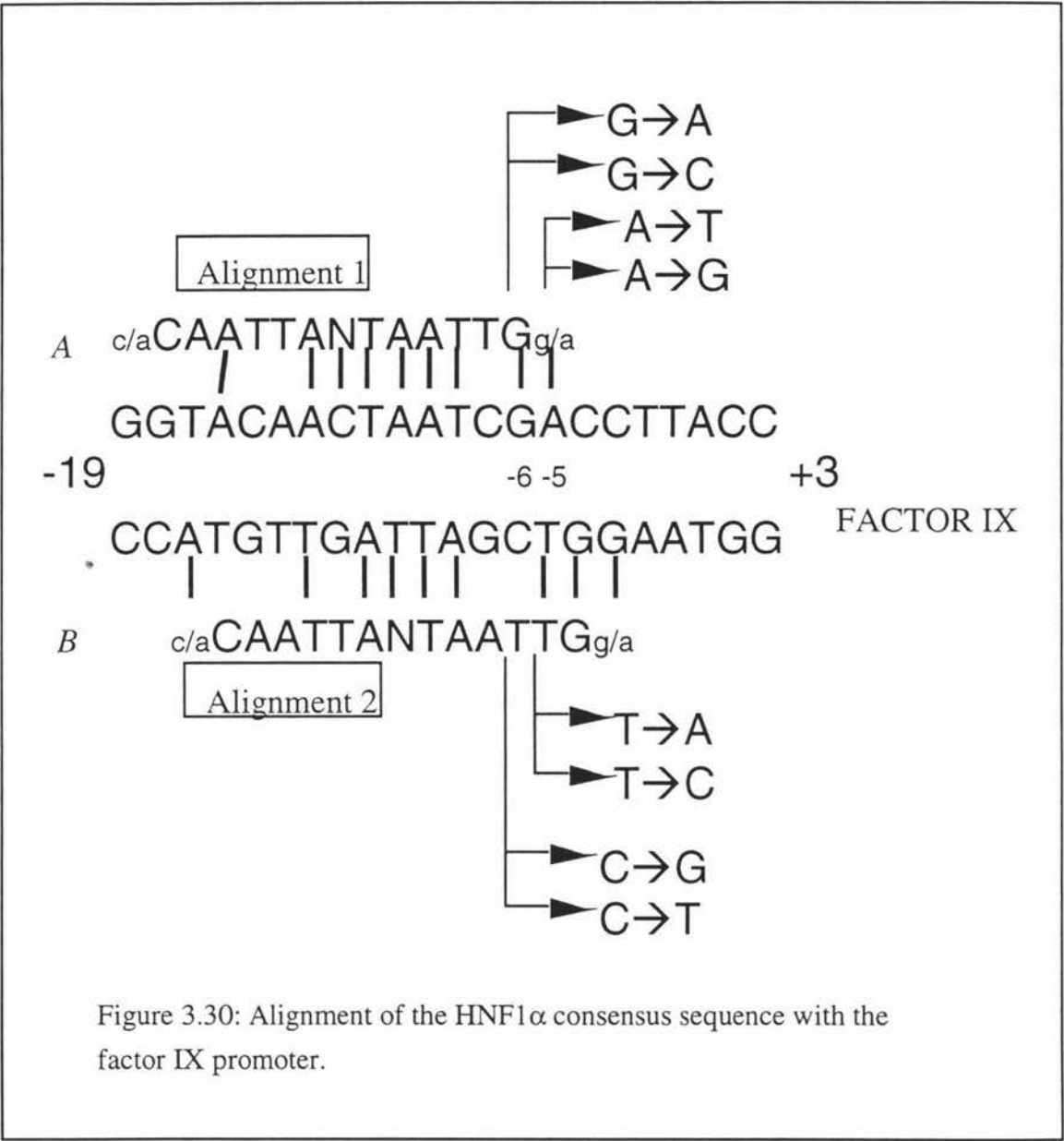
### *The Results for HNF1 $\beta$ Homodimers differ from those for Rat Liver Nuclear Extracts*

The results of gel shifts using the HNF1 $\beta$  extract show only slight similarity to those from the gel shifts using rat liver nuclear extracts (Section 3.5). Unfortunately, because the two sets of gel shifts use different extracts and different probes, the results cannot be directly compared. However, it can be noted that many of the mutant factor IX oligonucleotides which did not compete in the gel shifts using a nuclear extract did compete when using an HNF1 $\beta$ -enriched Cos cell extract. This may be because the HNF1 transcription factor was at a concentration in the Cos cell extracts which was much higher than would be found *in vivo*. This could be due to a low level of endogenous expression of the HNF1 proteins in rat liver whereas the HNF1 cDNAs in the Cos cell transfections were expressed at a high level from the SV40 promoter. This should only effect the sensitivity of the gel shift while the presence of other proteins in the nuclear extract may actually mask the HNF1 binding. In the gel shifts using Cos cell extracts, the -6G/C oligonucleotide had an affinity for HNF1 $\beta$  which is lower than most of the other mutant promoter regions whereas the unidentified proteins in the liver extract only bound the wildtype and -6G/C oligonucleotides. This difference in gel shift results could therefore be based on the different proteins interacting at the promoter. HNF1 $\alpha$  and HNF1 $\beta$  monomers have highly conserved DNA-binding domains. However, the combination of monomers may change the binding activity of the resulting dimer with respect to the other dimers. This possible variation may result from the different spatial placement of the amino acid sidechains in the monomers which interact with the DNA. Some dimers would be able to interact with a certain DNA sequence better than other dimers due to the alignment of these amino acid residues with key basepairs.

### *Binding of HNF1 $\beta$ to the Wildtype Factor IX Promoter*

The consensus DNA-binding sequence of HNF1 $\beta$  homodimers has not been reported to date. It is therefore not possible at present to identify potential HNF1 $\beta$  binding sites on the factor IX promoter. However, because of the high level of homology between DNA-binding domains, it is likely that the sequence will be similar to that for HNF1 $\alpha$  homodimers. Comparison between the known consensus sequence and the factor IX promoter may yield some useful information (Figure 3.30). Although it cannot be assumed that an HNF1 $\beta$  dimer is interacting directly with the factor IX promoter, the effect of a base substitution on DNA binding by the transcription factor can be analysed. The HNF1 $\alpha$  consensus sequence lines up in only two places on the factor

IX promoter around the -5, -6 region. One potential binding site (Figure 3.30A) lies on the coding strand from position -5 to -19. The second sequence (Figure 3.30B) is found in the non-coding strand of the double helix, starting at base -3.



The potential HNF1 sequence on the coding strand of the factor IX promoter (Figure 3.30) shows that most of the bases identical to the consensus sequence lie in the first half of the palindrome. This is advantageous because the first half of the palindrome of the consensus sequence is more stringently conserved and may therefore serve as the higher affinity halfsite. The -6G and the -5A nucleotides of the factor IX promoter

sequence match up exactly with the consensus sequence, such that a mutation at one of these positions should make a weaker binding site for HNF1.

The second alignment of the HNF1 consensus sequence to the factor IX promoter involves the non-coding strand of the double helix. The -5 base matches up perfectly with the consensus base in this position, so any mutation could generate a weaker site for an HNF1 dimer. At the -6 position, the cytosine of the factor IX promoter region is lined up with the other pyrimidine, thymine, so although this is not a perfect match of bases, it is possible that the transcription factor could still interact here.

The second alignment shown in Figure 3.30 best fits the results shown in the HNF1 $\beta$  gel shifts. Mutation of the -6 position from a guanine to an adenine may generate a site with higher affinity than that of the wildtype for an HNF1 $\beta$  homodimer, while a cytosine at position -6 would create a site which is less consensus-like. At the -5 position on the alignment, a mutation of any description would produce a less consensus-like site, but the gel shift results suggest that HNF1 $\beta$  has an affinity for a -5A/G promoter which is similar to that for the wildtype. The -5A/T promoter is not bound by HNF1 $\beta$  in these experiments. The results appear to disagree with the alignment, but the A/G mutation (on the non-coding strand) retains a pyrimidine (C) at position -5 and therefore may retain the HNF1 $\beta$  binding ability. Placement of a purine (A) at this site may abolish all interactions.

It may therefore appear strange that a mutation from a purine to a pyrimidine results in a loss of DNA binding. However the putative wildtype factor IX HNF1 site would be weak at best, and has been shown to be much weaker than that of the  $\alpha$ 1-antitrypsin promoter (Figure 3.28). Also, patients carrying a -5 or -6 mutation are relatively mild haemophiliacs in comparison with patients carrying other Leyden-like mutations<sup>52</sup>. It is therefore possible that HNF1 $\beta$  homodimers may interact at the factor IX promoter around positions -5 and -6.

This discussion may only be considered theoretically, because the consensus site used in Figure 3.30 is for HNF1 $\alpha$ , which may vary slightly from that for HNF1 $\beta$ .

The effect of this interaction on transcription from the factor IX promoter must also be tested in functional assays in order to be able to propose a role for the HNF1 transcription factors in factor IX gene expression.

### 3.6.3 Gel Shifts with Cos Cell Extracts Containing HNF1 $\alpha$ , HNF1 $\beta$ and DCoH

Following the success of the gel shifts using extracts from Cos cell transfections with HNF1 $\beta$ , further transfections using the HNF1 $\alpha$ , HNF1 $\beta$  and DCoH vectors were also attempted. Different combinations of the three vectors were used in the transfections to produce cell extracts enriched in the following transcription factors:

- HNF1 $\alpha$  homodimers
- HNF1 $\alpha$ .HNF1 $\beta$  heterodimers
- HNF1 $\alpha$  homodimers with DCoH
- HNF1 $\beta$  homodimers with DCoH
- HNF1 $\alpha$ .HNF1 $\beta$  heterodimers with DCoH.

The cell extracts were generated *de novo*, rather than by mixing combinations of extracts enriched in HNF1 $\alpha$ , HNF1 $\beta$  or DCoH. The co-expression of these proteins is thought to be necessary to obtain complexes with DCoH because individually purified proteins do not interact on mixing *in vitro*<sup>36</sup>.

Figures 3.31A and B show the autoradiographs of only a few of the gel shifts attempted. For HNF1 $\alpha$ , HNF1 $\beta$ , HNF1 $\alpha$ -DCoH, HNF1 $\beta$ -DCoH and HNF1 $\alpha$ -HNF1 $\beta$  transfections, the blank extract produced a retarded band identical to that generated by the transfected extracts. This retarded band could represent non-specific protein, or endogenous HNF1 proteins binding to the HNF1 site from  $\alpha$ 1-antitrypsin. These results suggest that the transient transfections of Cos cells were either unsuccessful or that insufficient amounts of HNF1 protein were produced to be detected by gel shift.

A number of factors were varied in an attempt to produce a binding pattern specific for the HNF1-enriched extracts. New transfection buffers were prepared and tested, as were buffers that had previously been shown to be successful in the transfection of other cell lines. No buffer combination eliminated the band in the blank lane on the gel shift. Various sections of the transfection protocol were varied, such as the length of time between transfection and harvest, and the amount of time that the calcium phosphate precipitate remained on the cells. No variation resulted in any difference to the gel shift. New cell lines from the Cos 7 stocks also failed to produce a cell extract specific to the  $\alpha$ 1-antitrypsin probe.

Various aspects of the gel shift procedure itself were also varied. Addition of BSA (Figure 3.32), and poly dG.dC in place of poly dI.dC were all tested for their



A.

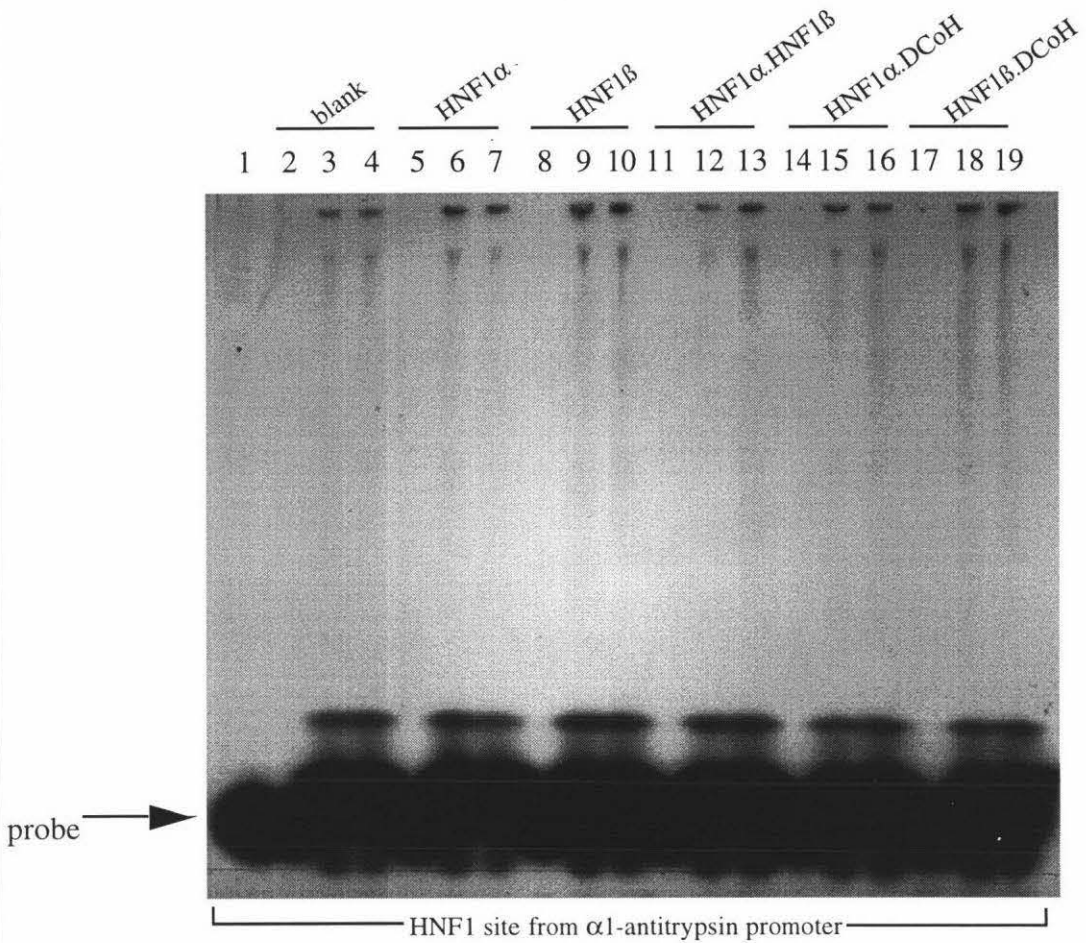
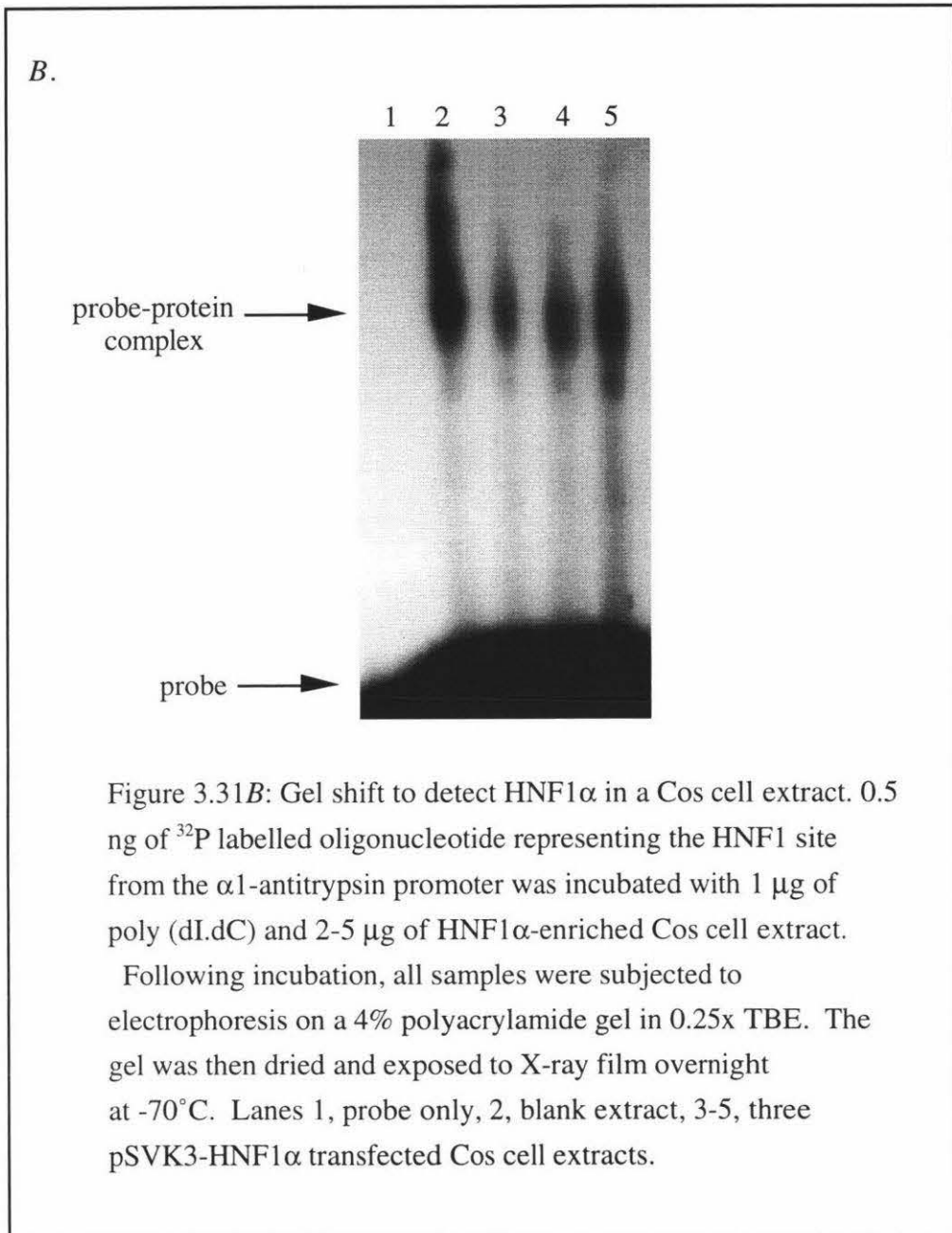


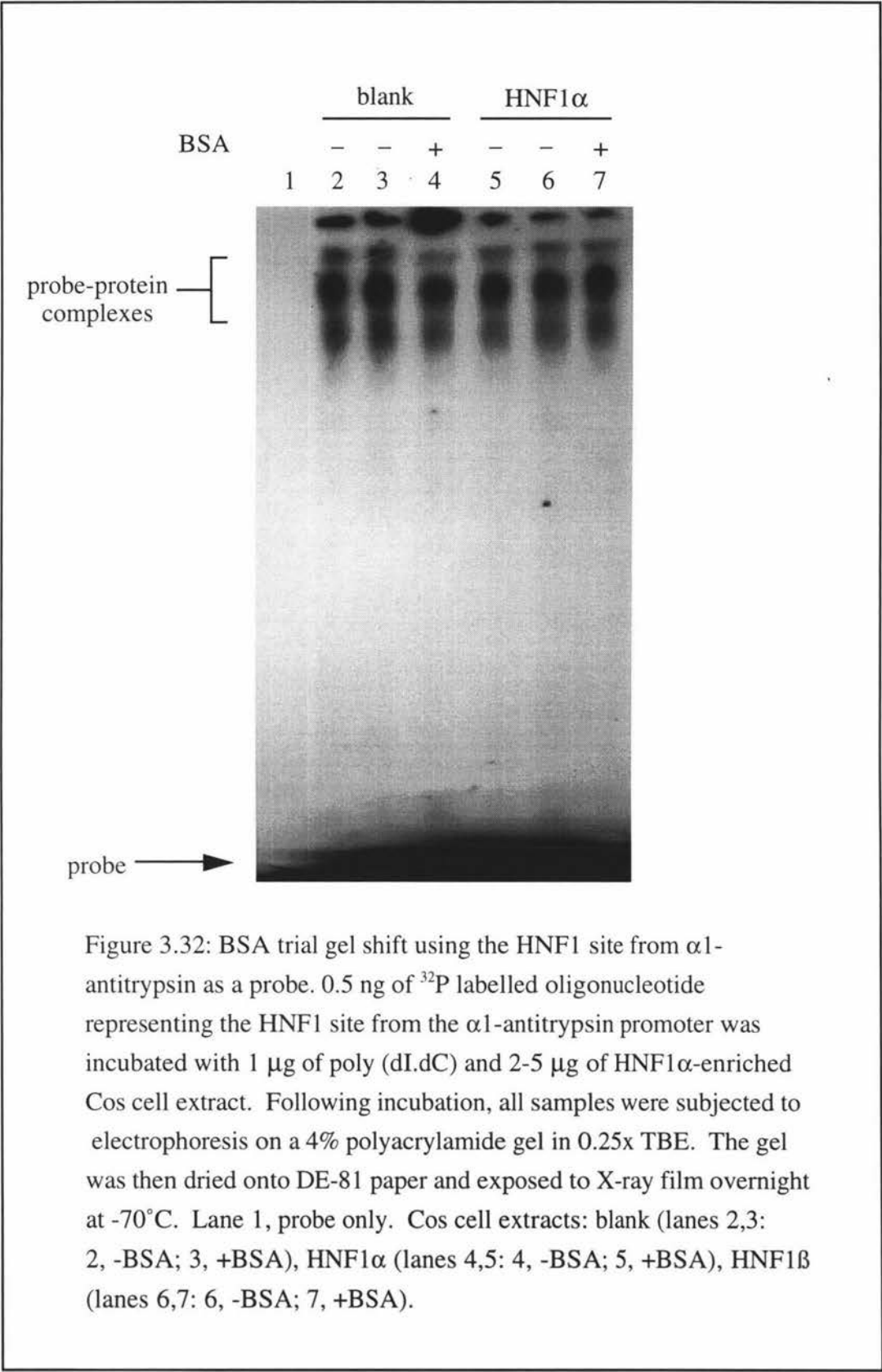
Figure 3.31A: Gel shift to test for the presence of an HNF1 transcription factor, using the HNF1 site from the  $\alpha$ 1-antitrypsin promoter as a probe. 0.5 ng of  $^{32}\text{P}$  labelled oligonucleotide representing the HNF1 site from the  $\alpha$ 1-antitrypsin promoter was incubated with 1  $\mu\text{g}$  of poly (dI.dC) and 2-5  $\mu\text{g}$  of HNF1-enriched Cos cell extract. Following incubation, all samples were subjected to electrophoresis on a 4% polyacrylamide gel in 0.25x TBE. The gel was then dried on DE-81 paper and exposed to X-ray film overnight at  $-70^\circ\text{C}$ . Lane 1, probe only. Cos cell extracts: blank (lanes 2-4: 2, 0  $\mu\text{L}$ ; 3, 1  $\mu\text{L}$ ; 4, 2  $\mu\text{L}$ ), HNF1 $\alpha$  (lanes 5-7: 5, 0  $\mu\text{L}$ ; 6, 1  $\mu\text{L}$ ; 7, 2  $\mu\text{L}$ ), HNF1 $\beta$  (lanes 8-10: 8, 0  $\mu\text{L}$ ; 9, 1  $\mu\text{L}$ ; 10, 2  $\mu\text{L}$ ), HNF1 $\alpha$ .HNF1 $\beta$  (lanes 11-13: 11, 0  $\mu\text{L}$ ; 12, 1  $\mu\text{L}$ ; 13, 2  $\mu\text{L}$ ), HNF1 $\alpha$ .DCoH (lanes 14-16: 14, 0  $\mu\text{L}$ ; 15, 1  $\mu\text{L}$ ; 16, 2  $\mu\text{L}$ ), and HNF1 $\beta$ .DCoH (lanes 17-19: 17, 0  $\mu\text{L}$ ; 18, 1  $\mu\text{L}$ ; 19, 2  $\mu\text{L}$ ).



ability to produce a change in the binding of the proteins to the probe. None of these combinations resulted in the production of differential binding patterns between the blank and transfected extracts. These results may indicate that the protein present in the extracts was endogenous HNF1, which may be present in monkey kidney cells, or a protein with a binding activity similar to HNF1.

Eventually the time constraints of Masterate research prevented further experimentation with the Cos cells and so they were reluctantly abandoned.





## Part C: The Effect of HNF1 Proteins on Transcription from the Factor IX Promoter

### 3.7 Introduction to Luciferase Assays

Reporter gene assays are a common and simple method of measuring the effects of a transcription factor on the expression of a gene. The promoter from the gene of interest is coupled to a gene encoding an enzyme whose catalytic activity is easily assayed. This vector is co-transfected with an expression vector encoding the transcription factor being studied into an appropriate cell line. The protein is transiently synthesised so that the rate of reporter gene expression may be regulated by the effects of this protein on the cell.

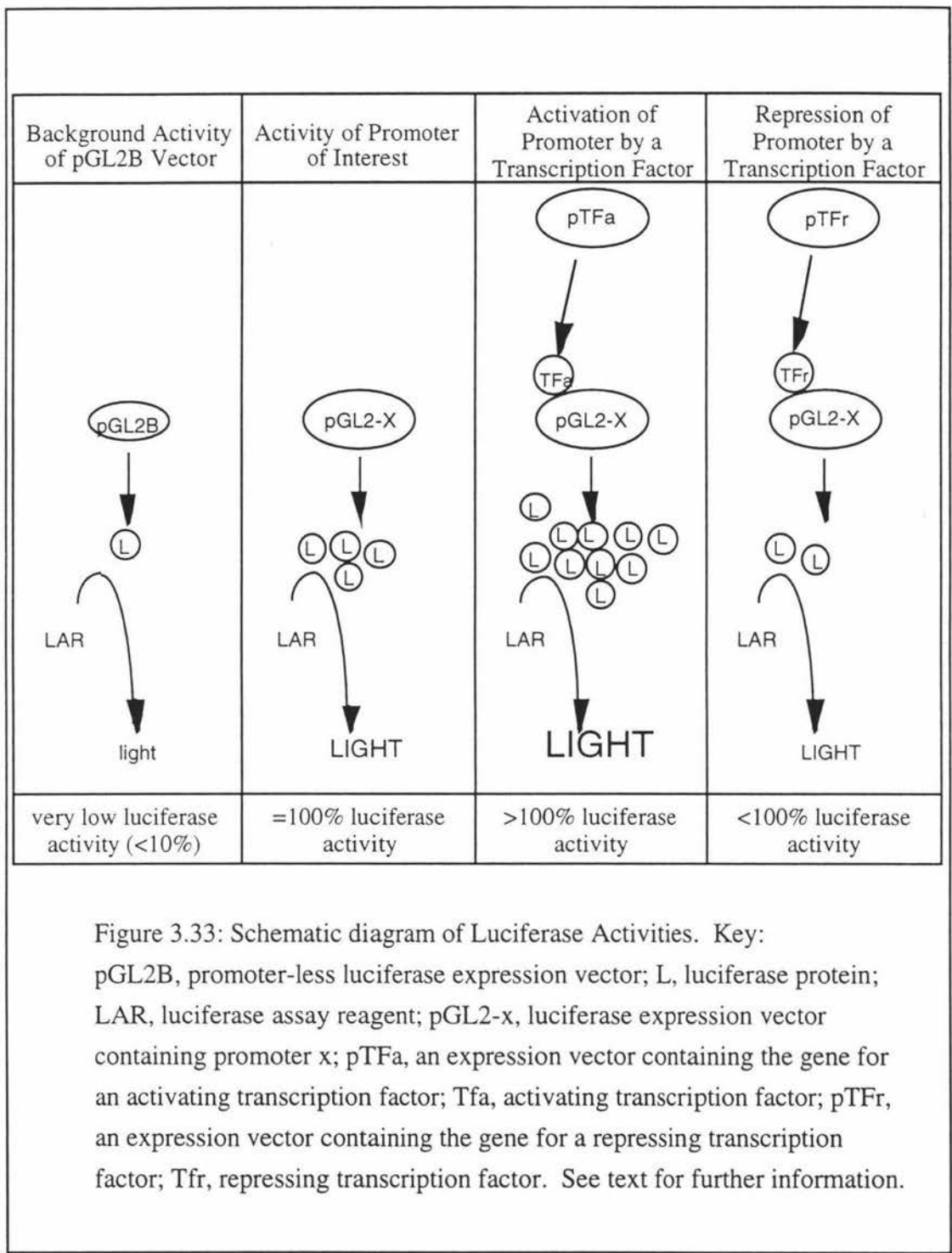
The luciferase reporter gene assay is a convenient method of quantifying the activity of a promoter. The assay theory is outlined in Figure 3.33. The promoter sequence of interest is ligated into the multiple cloning site of the supplied pGL2B expression vector, upstream of the firefly luciferase gene. Transfection of the expression construct followed by cell lysis produces a whole cell extract containing the luciferase enzyme. Amounts of luciferase present in the extract are directly measured through the catalytic activity of the luciferase enzyme, in a reaction which releases light. The amount of luciferase present, and therefore the amount of light emitted, is dependent on the rate of expression from the vector (Figure 3.33, second box). Background levels, or the rate of 'leakage' from a promoter-less vector, may also be measured by the transfection and subsequent assay of this construct (Figure 3.33, first box).

The rates of protein synthesis and enzyme activity are assumed to be constant between transfections, however factors pertaining to the transfection itself, such as cell density and confluence, must also be accounted for. To overcome the effects of these variations on luciferase gene expression, the  $\beta$ -galactosidase expression vector pSV- $\beta$ galactosidase is co-transfected with the luciferase vector as a constant factor between transfections. Following cell lysis, the whole cell extract is tested for the presence of  $\beta$ -galactosidase through the reaction:



The product of this reaction, *o*-nitrophenol, is yellow in colour, and so may be quantitatively measured using a spectrophotometer at wavelength 420 nm. The luciferase assays may then be normalised by dividing the luciferase activity value by the  $\beta$ -

galactosidase measurement to give a luciferase activity which can be compared with that of other transfections.



The effects of various transcription factors on a promoter sequence may be efficiently gauged by this technique. Expression vectors containing the cDNA for a transcription factor with modulation potential may be co-transfected with the luciferase construct, and the effects of the transcription factor measured through the assay system. A repressive transcription factor will have a decreased luciferase activity compared to that of an unstimulated promoter, while an activating factor will have an increased value (Figure 3.33, third and fourth boxes). Comparison of these values may be simplified by designating the luciferase activity of the unstimulated promoter as 100%. An activated promoter will therefore have a luciferase activity of >100% and a repressed promoter will have an activity of <100%.

The luciferase reporter gene system may also be employed to study the effects of transcription factors on mutant promoter sequences. The background, unstimulated luciferase activity of promoters with natural mutations will be less than that of the wildtype promoter (<100%), and repression of a mutant promoter by a transcription factor will decrease its luciferase activity. Activation of a mutant promoter will obviously increase its luciferase activity, and this may be compared to the values for the wildtype promoter. Such a study on transcription from mutant promoters may provide valuable insight into the function of a disease locus.

### 3.7.1 Potential Sources of Error in Luciferase Assays

The results generated from a reporter gene assay can be numerical, and so must therefore be statistically sound. Repetitions of each experiment allows the averaging out of variation, but also introduces the calculation of errors. An error indicates how much variation was present in a given data set, and is usually expressed as the mean, plus-or-minus the standard error of the mean. Errors in transcriptional assays tend to be rather large, with values of  $\pm 25\%$  not uncommon in this set of experiments. Although at first glance this error margin seems excessive, consideration should be given as to the amount of variation possible within one experiment (as discussed below).

There are three main steps to a reporter gene assay: the transfection of cells with the expression vectors, the expression of the reporter protein, and the subsequent assay of this protein. Each step has many potential sources of variation. Firstly, the efficiency of transfection is very low, commonly <40% for most mammalian cells, and individual cells have differences which dictate the number of vectors, if any, taken up by each cell. Secondly, the culture into which the DNA is being transfected consists of thousands of

individual cells. Like any population of individuals, there exists subtle yet significant differences within the population, such as varying rates of transcription and translation. When these two considerations are extended from differences between cells to differences between the cultures used in each experiment, further potential for error occurs. The major difference between experiments is the age and confluence of the monolayers. Slight variation in growth conditions and media may well elicit a variable response to the transfection process.

The final step in the process is the actual assay of the reporter enzyme. This requires harvesting of cells followed by monitoring of the enzyme activity. A number of assumptions are made to qualify this process. It is assumed that the rate of enzyme catalysis is indicative of the number of enzymes present, and that the number of enzymes is proportional to the rate of reporter gene expression, which therefore can be related to the regulation of the promoter of interest. Although the rates of these processes are certainly constant within a cell, using hundreds of cells in at least 50 experiments can create a large amount of variation.

The  $\beta$ -galactosidase vector, pSV- $\beta$ -galactosidase, was used in these experiments in an attempt to remove some of this variation. The inclusion of this vector in all the co-transfections provides a constant factor between all the experiments. The  $\beta$ -galactosidase gene is under the control of the constitutive simian virus (SV40) promoter, and so is essentially free from regulatory control. The amounts of the enzyme present in an experiment are therefore dependent on the efficiency of transfection, and the rates of transcription and translation of the monolayer as a whole. By normalising the associated luciferase activity with the  $\beta$ -galactosidase activity, all the individual experiments may be compared, regardless of when they were conducted. In theory, the use of the  $\beta$ -galactosidase expression vector in the experiments provides a common factor which should be constant irrespective of time or cellular variations. Although this theory does go some distance towards solving the problem of variation in transcriptional assays, the errors generated in the following sets of experiments attest to the continuing problem with accuracy in this type of assay. For example, there is no guarantee that the  $\beta$ -galactosidase vector will be taken up by the cells to the same extent as the promoter-reporter gene vector. An ideal situation would be where the control reporter gene is on the same vector as the promoter-reporter gene. Vectors such as this are now available, but were outside the budget allocated to this project.

3.8 Transcription From the Factor IX Promoters

3.8.1 The Mutant Promoters are Down regulated in Alexander Cells

The mutant factor IX promoters (-220 to +44 region) were amplified by PCR and ligated into separate luciferase vectors. A construct containing the wildtype factor IX promoter was also made. These expression vectors were kindly constructed by Isobel Franklin and Robyn Marston.

Transfection of the luciferase constructs into Alexander cells (a human hepatoma cell line) in the absence of recombinant transcription factors determines the background levels of expression for each of the promoter constructs (see Section 2.2.5). Alexander cells were transfected with each luciferase-promoter construct together with the  $\beta$ -galactosidase expression vector, harvested and assayed for luciferase and  $\beta$ -galactosidase activities. Each transfection was carried out in triplicate, using transfection parameters which had previously been optimised. The luciferase activity was normalised with respect to the  $\beta$ -galactosidase readings (Table 5) to generate a relative luciferase activity.

Vector	Luciferase Activity (relative light units)	$\beta$ -galactosidase Activity ( $A_{420}$ )	Relative Luciferase Activity (luciferase activity/ $\beta$ galactosidase activity)	Average	Percentage of wildtype
pGL2-wt FIX	5.084	0.989	5.141		
	5.901	0.891	6.623	5.719	100.0
	3.625	0.672	5.394		

pGL2-6G/A	0.493	0.438	1.126		
	0.480	0.412	1.165	1.504	26.3
	0.582	0.262	2.221		

Table 5: Sample Calculation of Relative Luciferase Activities.

The three activities for each luciferase construct were averaged. The activity of the luciferase vector carrying the wildtype factor IX promoter was arbitrarily set at 100%. The wildtype promoter was always normalised to 100% luciferase activity to provide a



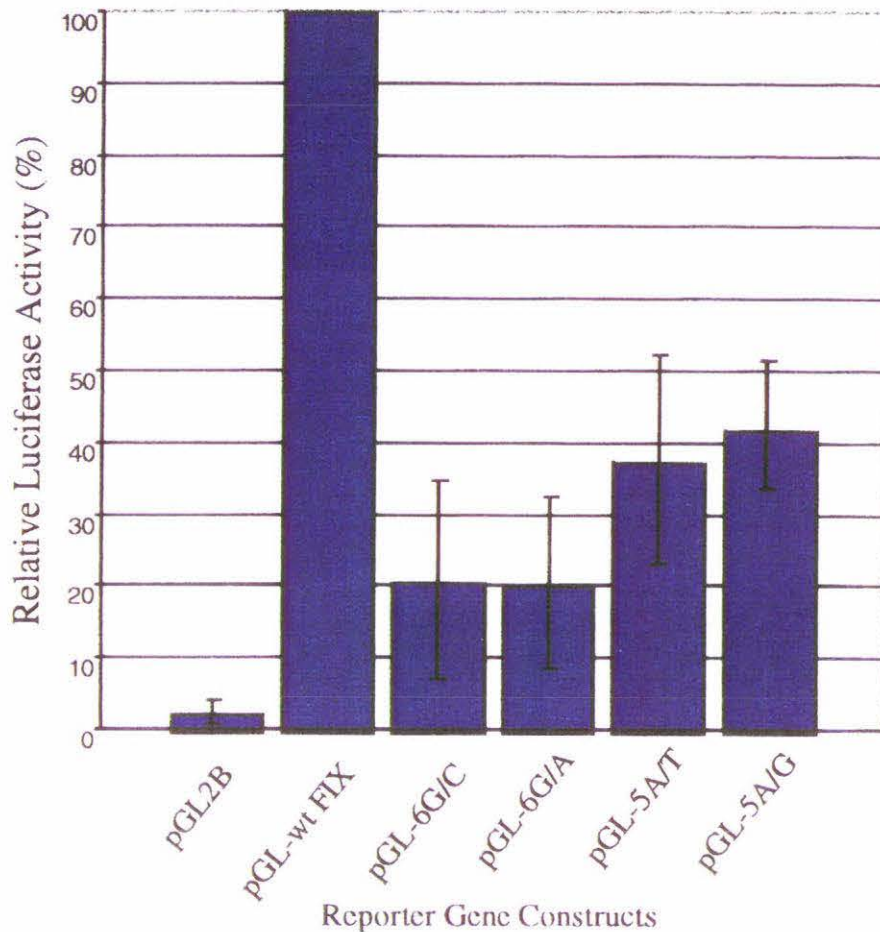


Figure 3.34: Basal transcription from the factor IX promoters. Alexander cells were transiently transfected with 2  $\mu$ g of reporter constructs that contained DNA corresponding to the factor IX promoter (either normal or a -5 or -6 mutant sequence) and 0.5  $\mu$ g of pSV- $\beta$ galactosidase vector. 2  $\mu$ g of the promoter-less pGL2B vector was used as controls. 46 hours after transfection the cells were harvested and assayed for luciferase and  $\beta$ -galactosidase activities. 'Relative Luciferase Activity' was calculated by converting the luciferase assay result for each construct to a percentage of that obtained for the wildtype promoter construct, which was set arbitrarily at 100%. The activities were also normalised with respect to the  $\beta$ -galactosidase readings. At least three independent experiments were carried out where each involved triplicate transfections for all reporter gene constructs. The standard error of the mean is shown.

reference point for the activities for the other promoter vectors. The averaged activities of the remaining luciferase vectors were then expressed as a percentage relative to that of the wildtype promoter (Table 5). All four -5 and -6 mutations resulted in a decrease in the expression from the factor IX promoter (Figure 3.34).

### 3.8.2 A Comparison of Factor IX Activities of Haemophilia B Leyden Patients.

The observed luciferase activity from the mutant factor IX promoter-reporter gene constructs was within the expected range for these mutations<sup>52</sup>. Most of the known patients carrying a -5 or -6 mutation display a decrease in blood clotting ability (Table 6, page 87), although the extent of the reduction tends to vary between patients. This variation is likely to be the result of slight differences in the clinical tests carried out at different laboratories, however it may also be due to the influence of other genetic loci on this region of the factor IX promoter. For example, genetic variants of the transcription factors binding to the promoter can produce differences in the transcription rate of the associated gene. Differences in the rates of more general processes, such as translation, may also account for some of the variation in clotting ability observed here. Although the results of these reporter gene assays vary considerably, the observed results still lie within the expected range for transcription from a mutant factor IX promoter.



Table 6: The blood clotting abilities of the known -5 and -6 mutant factor IX promoter carriers (Taken from Giannelli *et al.*, 1998<sup>52</sup>).

Promoter Mutation	Patient	Clotting Abilities (% of normal)
-6G/C	Toulouse	1-30
	Malmö 63	16
-6G/A	High Wycombe	13-70
	Leyden, USA1	10
	Rodez	1-normal
	HB272, Fr	7
	Liverpool1	30
	Liverpool2	10-64
	HB262	9
	UK269	30
	UK352	18-60
	Unnamed	2.5
	Perth 1	10-65
	Unnamed	0-25
	UK421	28
	UK430	15
	Unnamed	3-26
	UK449	1
-5A/T	Toronto 20	3
-5A/G	UK110	15
	UK161	7

### 3.9 Effects of HNF1 $\alpha$ on Factor IX Expression

#### 3.9.1 HNF1 $\alpha$ Down-regulates Transcription from the Wildtype Factor IX Promoter *in vitro*

Alexander cells were transfected with luciferase reporter gene constructs using calcium phosphate precipitation, as detailed in Section 2.2.2. Each of the five factor IX promoter-reporter gene constructs, (normal, -6G/C, -6G/A, -5A/T and -5A/G), were co-transfected with the pSVK3-HNF1 $\alpha$  expression plasmid. This expression vector contains the HNF1 $\alpha$  cDNA under the control of the strong simian viral (SV40) promoter sequence, such that in transfected mammalian cells, expression of the transcription factor is essentially constitutive. Each experiment was carried out in triplicate, and conducted a minimum of three times so that at least nine separate luciferase activities were available for analysis.

A third vector included in all the co-transfections contained the gene for  $\beta$ -galactosidase which was also under the control of the SV40 promoter sequence. The relative luciferase activities for the HNF1 $\alpha$ , factor IX co-transfections were normalised for  $\beta$ -galactosidase activity, expressed as a percentage, and compared to basal level activity (Figure 3.35).

Two negative controls were included in the experiment: the transfection of either the promoter-less luciferase vector (pGL2B), or the expression vector carrying the HNF1 $\alpha$  cDNA. In the first situation, the promoter-less luciferase gene exhibited low background expression (or 'leakage') as is common for reporter genes<sup>69</sup>. The second control tested the effect of HNF1 $\alpha$  alone in this assay system and demonstrated background activity only (Figure 3.35)

The most obvious result of the HNF1 $\alpha$ , factor IX co-transfections was the large decrease in activity of the wildtype factor IX promoter-reporter gene construct in the presence of HNF1 $\alpha$ . The luciferase activity was decreased by nearly 50%. The activities of two of the constructs carrying the mutant factor IX promoters also exhibited some change in the presence of HNF1 $\alpha$ . The activity of the constructs containing the -6G/C and -5A/T mutations appeared to be increased from their respective background levels. However, there are large error margins in each case, due to a wide variation in the luciferase activities from individual experiments. For both the -6G/A and -5A/T promoters, the error margin for the HNF1 $\alpha$ -stimulated experiment overlapped with that

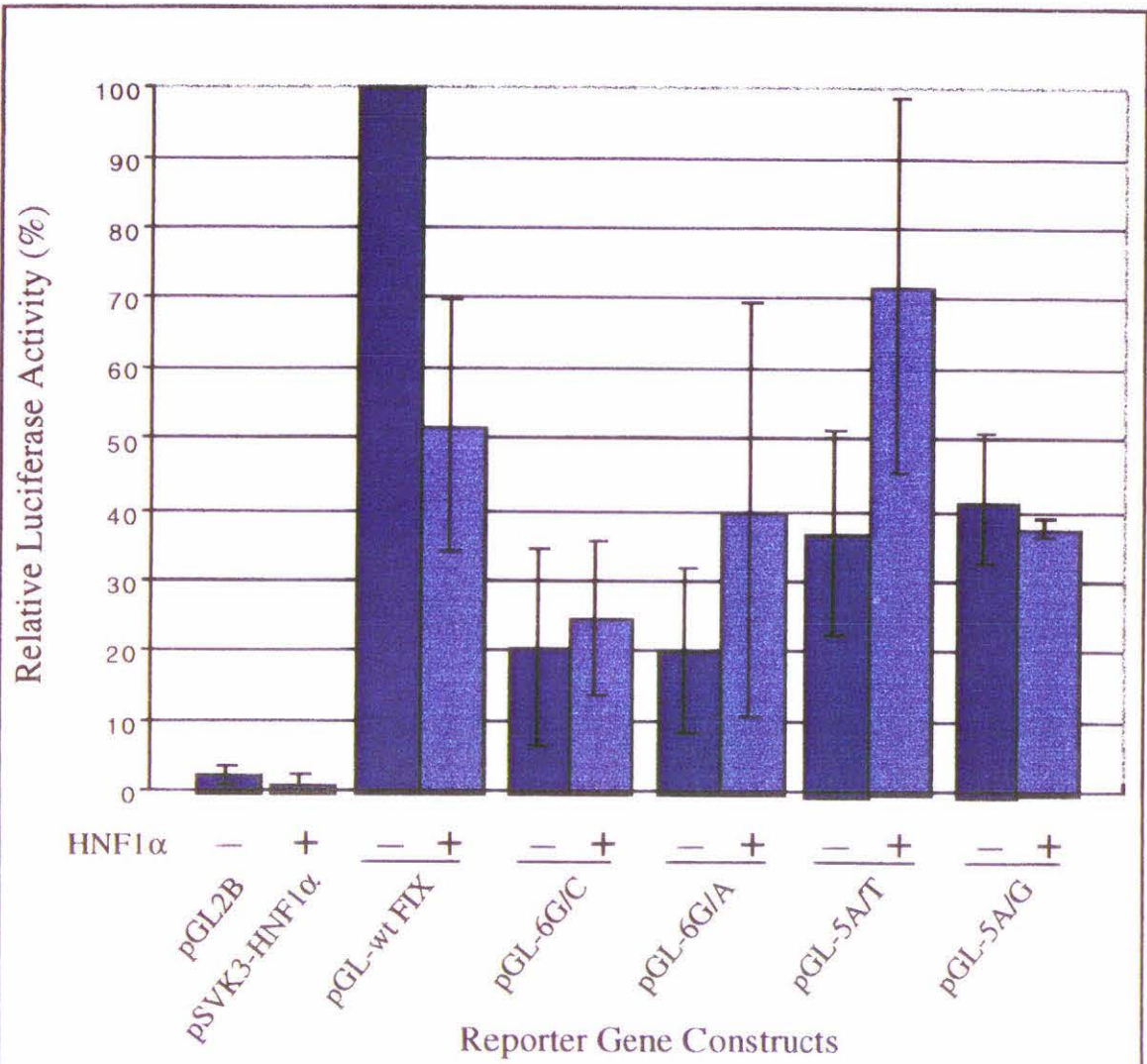


Figure 3.35: Results of the co-transfection of HNF1 $\alpha$  with the factor IX promoters. Alexander cells were transiently transfected with 2  $\mu$ g of reporter construct that contained DNA corresponding to the factor IX promoter (either the normal or -5 or -6 mutant sequence), 2  $\mu$ g of the expression vector pSVK3-HNF1 $\alpha$  and 0.5  $\mu$ g of pSV- $\beta$ galactosidase vector. 2  $\mu$ g each of the promoter-less pGL2B vector and pSVK3-HNF1 $\alpha$  were also transfected individually as controls. 46 hours after transfection the cells were harvested and assayed for luciferase activity. 'Relative Luciferase Activity' was calculated by converting the luciferase assay result for each transfection to a percentage of that obtained from the wildtype reporter construct which was arbitrarily set at 100%. The activities were also normalised with respect to the  $\beta$ -galactosidase readings. The resulting activities were compared to those of the unstimulated promoters. Results are expressed as the mean  $\pm$  standard error of the mean.



of the unstimulated result. On the basis of the overlap of these error margins, the results for the promoter-reporter gene assays with the -6G/A and -5A/T mutations were considered inconclusive.

### 3.9.2 HNF1 $\alpha$ as a Transcriptional Repressor?

HNF1 $\alpha$  has been identified as a regulator of transcription at a number of liver-specific loci and is known to transactivate the expression of the  $\alpha$ 1-antitrypsin<sup>25</sup> and albumin<sup>32</sup> genes. It was therefore an unexpected result to find that HNF1 $\alpha$  down-regulated the transcription of a reporter gene linked to the factor IX promoter. There is only one documented case of HNF1 $\alpha$  acting as a repressing transcription factor, which is at the promoter of its own gene<sup>53</sup>. However, this example of negative autoregulation is under debate. The reported down-regulation of reporter genes under the control of the HNF1 $\alpha$  promoter is not completely reproducible<sup>54</sup> and critics therefore label these results inconclusive. Another inconsistency is the absence of a consensus site in the HNF1 $\alpha$  promoter for the HNF1 $\alpha$  transcription factor. Supporters of the repressing activity of HNF1 $\alpha$  attempt to explain this phenomenon of down-regulation in the absence of DNA-binding by 'quenching' (Figure 3.36). Quenching is defined as the prevention of the binding of a transcriptional activator to its consensus sequence by its interaction with another transcription factor, in this case HNF1 $\alpha$ . In effect, the binding of HNF1 $\alpha$  to the second, unidentified transcription factor would render this protein incapable of binding to the promoter of the HNF1 $\alpha$  gene, resulting in the quenching of its transcriptional potential<sup>53</sup>.

However, there are other means by which HNF1 $\alpha$  could display both repressive and stimulatory activities. The regulation of divergent regulatory activities would require a control point or 'switch' between the two functions. The switch may be regulated by other cellular proteins, the concentration of the transcription factor, or even by the position of the associated *cis*-regulatory elements at a promoter. Most transcription factors which display dual regulatory potentials have been shown to possess separate regulatory domains for each function. The switch between the two activities may therefore require that one domain be concealed with the subsequent exposure of the second domain. The regulation may be provided by an auxiliary protein which binds to one of the domains and inhibits its effects on the preinitiation complex of the promoter.

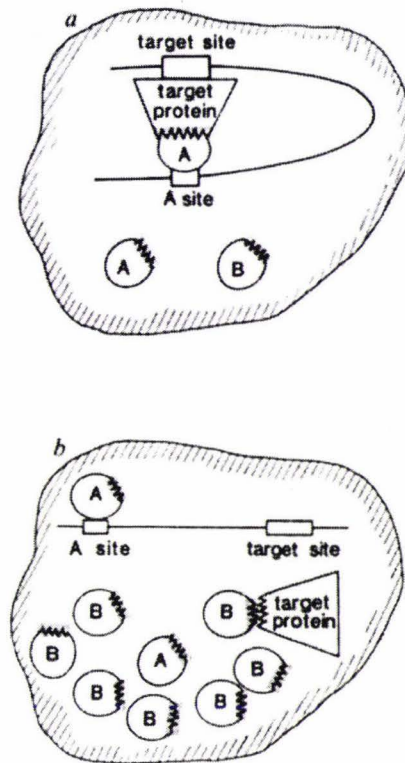


Figure 3.36: Quenching. In the cell in *a*, the activators A and B are present in ordinary concentrations. The target protein is brought to the DNA by interaction with A bound to its site (a protein-protein interaction), and by interaction with the target site (a protein-DNA interaction). In the cell in *b*, the activator B is in great excess and binds to the target protein in the absence of DNA; because the activating regions of both A and B see the same site on the target protein, interaction with B prevents that protein from interacting with A, and the gene is not transcribed. (Figure taken from Ptashne, 1988<sup>11</sup>).

The YY1 protein from adeno-associated virus (AAV) is a bifunctional transcription factor which may be regulated in this way by the viral protein E1A. YY1 binds to a specific sequence centred at position -60 of the P5 AAV promoter and represses the initiation of



its transcription<sup>55</sup>. In the presence of E1A, YY1 activates transcription from the P5 promoter. The switch to activation in the presence of E1A may be the result of a post-translational modification by E1A<sup>55</sup>. Alternatively, E1A may physically associate with the suppressing domain of YY1 and block its repressing abilities<sup>55</sup>. Conversely, E1A may remove a repressor protein from YY1, unmasking a transcriptional activation domain<sup>55</sup>. YY1 has also been shown to activate transcription from the genes for some ribosomal proteins<sup>56</sup>, and to repress transcription from the Ig kappa promoter<sup>57</sup>.

The switch between activating and repressing functions is also known to be dependent in some cases on the concentration of the transcription factor. The Kruppel protein from *Drosophila* is a transcriptional activator at low concentrations but dimerises at high concentrations to form an inhibitor of transcription<sup>58</sup>.

The product of the Wilm's Tumour gene (WT1) has also been characterised as a bifunctional transcription factor<sup>59</sup>. The switch between the activities of this protein is thought to be dependent on the position of its DNA binding site. The promoter of the platelet-derived growth factor (PDGF) A-chain gene contains two DNA-binding sites for WT1, one each site of the transcription start point<sup>59</sup>. WT1 activates transcription from this promoter when bound at one of these sites, but the binding of WT1 to both sites results in transcriptional repression<sup>59</sup>. It is thought that two transcription factors bound at the promoter may produce a suppressive interaction which involves the repression domains of the transcription factors<sup>59</sup>.

The possibility of HNF1 $\alpha$  activating and repressing transcription from different promoters is therefore an unusual, but not unprecedented, concept. If the observed down-regulation of the factor IX-luciferase reporter gene in this experiment is representative of the *in vivo* situation, the transcriptional activity of HNF1 $\alpha$  may be tightly regulated (see above). The most likely explanation for the observed down-regulation of transcription by HNF1 $\alpha$  could be an interaction with a second cellular protein. This interaction may result in the exposure of a transcriptional repression domain on the HNF1 $\alpha$  protein, or the quenching of a second transcription factor's activity. HNF1 $\alpha$  may accomplish this quenching through its DNA-activation domain, or as a monomer through its dimerisation domain. Confirmation of this theory would require the isolation of a transcription factor which can regulate the factor IX promoter and to which HNF1 $\alpha$  can bind. Such a transcription factor may be another member of the HNF1 protein family, with which HNF1 $\alpha$  could form heterodimers. It is predicted that such a heterodimeric transcription factor would bind the factor IX promoter due to the presumed conservation of DNA-binding site affinities common to the HNF1 proteins.

The mechanism for regulation of the factor IX promoter by HNF1 $\alpha$  is therefore likely to be complex and will require further analysis.

### 3.9.3 The Possible Function of Repression of the Normal Factor IX Promoter by HNF1 $\alpha$

HNF1 $\alpha$  is expressed in a number of endoderm- and mesoderm-derived tissues including liver and kidney. However, in all the major HNF1 $\alpha$ -expressing tissues, HNF1 $\beta$  is also expressed, generating a potential for heterodimerisation. HNF1 $\alpha$  homodimers have been identified in the initiation of transcription from a number of liver-specific genes, but the presence of HNF1 $\beta$  in liver suggests that HNF1 $\alpha$ .HNF1 $\beta$  heterodimers may also be involved in transcription at other genetic loci. For example, the observation that exogenously expressed HNF1 $\alpha$  cannot activate the endogenous albumin promoter in cells that do not express either HNF1 protein<sup>17</sup> suggests that an auxiliary factor expressed in liver is required to enable HNF1 $\alpha$  to activate the albumin promoter. This may reflect the requirement at some liver-specific promoters for a second factor lacking in these cells. At present, an obvious candidate for this factor is HNF1 $\beta$ .

Most liver-specific genes which require HNF1 $\alpha$  for the initiation of transcription have not been characterised with respect to the HNF1 $\alpha$  dimers present at the promoter, as the presence of HNF1 $\alpha$  in a gel shift does not provide conclusive evidence for the formation of homodimers. Nuclear extracts from both liver and kidney have been shown to contain both types of HNF1 homodimer and heterodimers<sup>5</sup>, so there must exist some form of selection process which dictates which HNF1 transcription factor is present at particular promoters. Selection of HNF1 $\alpha$  homodimers over HNF1 $\beta$  homodimers or heterodimers in some HNF1-expressing cells could possibly result in the down-regulation of the wildtype factor IX gene. Such cell types would include kidney and intestinal cells, where both HNF1 $\alpha$  and HNF1 $\beta$  proteins are synthesised, but where the factor IX gene is not expressed.



### 3.10 Effects of HNF1 $\beta$ on Factor IX Expression

#### 3.10.1 HNF1 $\beta$ Down-regulates Transcription from the Wildtype Factor IX Reporter Gene

Transient transfection of Alexander cells followed by luciferase reporter gene assays were conducted to test the effect of HNF1 $\beta$  on transcription from the various factor IX promoter constructs. These experiments were carried out in a manner identical to that described in Section 3.9.1. The background levels of luciferase present in HNF1 $\beta$ -enriched Alexander cells was tested by transfection with the pSVK3-HNF1 $\beta$  expression vector only. The associated relative luciferase activity was suitably low to eliminate the possibility of artefacts.

The most striking result from these assays is the down regulation of the wildtype factor IX promoter (Figure 3.36). Inclusion of HNF1 $\beta$  with the wildtype factor IX-promoter-reporter gene construct produced a decrease in luciferase activity of about 70%. The four mutant factor IX promoters studied did not appear to be regulated by HNF1 $\beta$  at all under the experimental limitations of the assay system. However, there is a trend towards an increase in luciferase activity in the presence of HNF1 $\beta$  for all four constructs which carry the mutant factor IX promoter region. This result is similar to that observed for HNF1 $\alpha$ . HNF1 $\beta$  may therefore interact at these promoter sequences and slightly activate the initiation of transcription. The down-regulation of transcription from the wildtype factor IX promoter-reporter gene vector may therefore be due to HNF1 $\beta$  binding to this sequence and preventing the interaction of an activating transcription factor, rather than repressing the promoter itself. However, the increase in activity of the mutant factor IX promoter-reporter gene constructs may not be significant due to the large error margins observed.



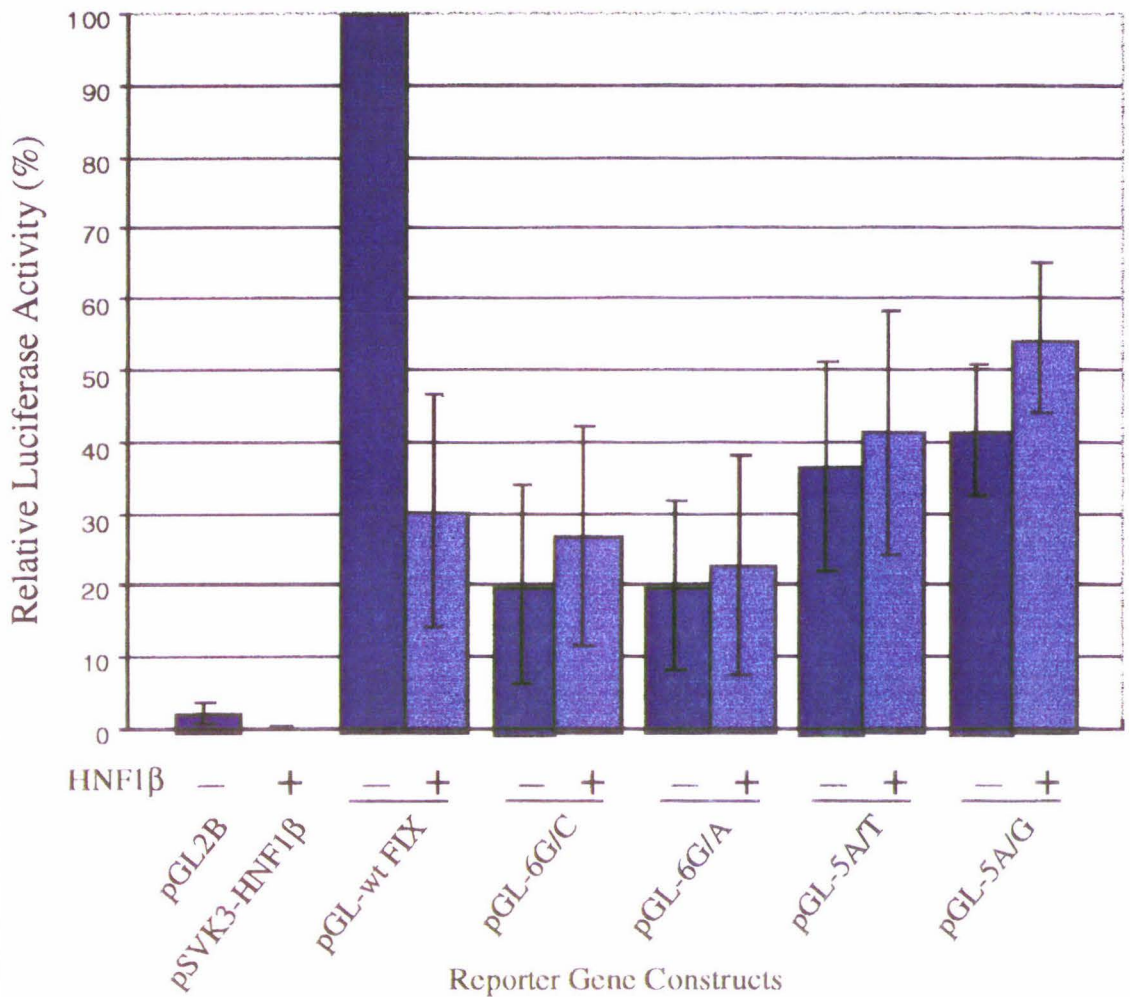


Figure 3.37: Results of the co-transfection of HNF1 $\beta$  with the factor IX promoter-reporter gene construct. Alexander cells were transiently transfected with 2  $\mu$ g of reporter construct that contained DNA corresponding to the factor IX promoter (either normal or -5 or -6 mutant sequence), 2  $\mu$ g of the expression vector pSVK3-HNF1 $\beta$  and 0.5  $\mu$ g of pSV- $\beta$ galactosidase vector. 2  $\mu$ g each of the promoter-less pGL2B vector and pSVK3-HNF1 $\beta$  were also transfected individually as controls. 46 hours after transfection the cells were harvested and assayed for luciferase and  $\beta$ -galactosidase activities. 'Relative Luciferase Activity' was calculated by converting the luciferase assay result for each transfection to a percentage of that obtained from the wildtype reporter construct, which was arbitrarily set at 100%. The activities were also normalised with respect to the  $\beta$ -galactosidase readings. The resulting activities were compared to that of the unstimulated promoters. The results are presented as the mean  $\pm$  the standard error of the mean.

### 3.10.2 HNF1 $\beta$ May Possess a Bifunctional Regulatory Activity

The sole significant result of the transcriptional assays in the presence of HNF1 $\beta$  was the distinctive repression of the wildtype factor IX promoter. This result was consistent with the previous observation of former student Isobel Franklin<sup>46</sup>, in which a 65% repression of the wildtype factor IX promoter was observed with HNF1 $\beta$ . Unfortunately, the high level of agreement exhibited by these experiments cannot be related to the current endogenous function of HNF1 $\beta$ , as the *in vivo* activity of HNF1 $\beta$  has not been satisfactorily characterised. HNF1 $\beta$  was originally isolated from dedifferentiated hepatoma cell lines such as H5 and C2, which do not express many of the liver-specific genes and fail to display a hepatic phenotype<sup>20</sup>. When these cell lines were transfected with an albumin promoter construct, it remained silent, suggesting that HNF1 $\beta$  may either down-regulate this promoter, or may have no modulatory role. However, co-transfection of HNF1 $\beta$  and the albumin promoter in human epithelial cells resulted in the activation of transcription from this promoter<sup>20</sup>. Repetition of this experiment in H5 cells resulted in the activation of the albumin promoter by the exogenous HNF1 $\beta$ . Endogenous HNF1 $\beta$  therefore may appear as a repressor of the albumin promoter in dedifferentiated liver cells because of a low expression rate, or possibly due to inhibition by another protein present only in dedifferentiated liver cells, which would be titrated out in the transfection experiments.

The manner in which this research was carried out tends to vary slightly from that of the experiments using the albumin promoter. Rey-Campos used epithelial cells for his experiments<sup>20</sup>, but Alexander cells were used here because of their hepatic origin. Although not a fully differentiated liver cell line, Alexander cells are too mature to be defined as dedifferentiated. Fully differentiated HepG2 cells were used by Isobel Franklin<sup>46</sup> to obtain similar results to those documented here. When these results are taken in conjunction with the experiments of Rey-Campos<sup>20</sup> using the albumin promoter, the regulatory potential of HNF1 $\beta$  tends to become confused. In differentiated hepatic cell lines, HNF1 $\beta$  has been shown to repress the wildtype factor IX promoter, but the albumin promoter is activated by HNF1 $\beta$  in epithelial cells. Assuming that both these results are authentic, HNF1 $\beta$  may display a bifunctional role, with a potential to either activate or repress a promoter depending on cell type. The switch between the activities would probably be dependent on the cell-specific proteins which could interact with HNF1 $\beta$  (see Section 3.9.2 for previous discussion of bifunctional transcription factors).

The observed repression of the wildtype factor IX promoter by HNF1 $\beta$  may form the basis of the liver-specific expression of the factor IX gene. HNF1 $\beta$  is expressed in a



number of cell types derived from the endoderm and the mesoderm<sup>35</sup>. Its expression pattern overlaps largely with that of HNF1 $\alpha$ , except the HNF1 $\beta$  gene is expressed in the lung, while the HNF1 $\alpha$  gene is not. Lung cells therefore can only contain HNF1 $\beta$  homodimers, which may repress transcription of the factor IX gene in this tissue. Selection of HNF1 $\beta$  homodimers in cells expressing both forms of HNF1 may also lead to the down-regulation of the factor IX gene.

### 3.11 The Effects of HNF1 $\alpha$ .HNF1 $\beta$ Heterodimers on Factor IX Gene Expression

#### 3.11.1 HNF1 $\alpha$ .HNF1 $\beta$ Heterodimers Up-Regulate the Wildtype and -6G/C Factor IX Promoters *in vitro*

The luciferase reporter gene assay system was employed to test the regulatory effects of both HNF1 $\alpha$  and HNF1 $\beta$  on the factor IX promoter. Alexander cells were co-transfected with expression vectors carrying the HNF1 $\alpha$  and HNF1 $\beta$  cDNAs plus each of the normal and mutant factor IX promoter-reporter gene constructs. The activity of the luciferase enzyme in each transfection was measured and normalised with the amount of exogenous  $\beta$ -galactosidase present (see Section 2.2.2). Each transfection was carried out in triplicate, and each experiment was repeated at least three times. The relative luciferase activity of each promoter in the presence and absence of HNF1 $\alpha$  and HNF1 $\beta$  transcription factors was compared (Figure 3.37).

Two of the factor IX promoters studied were activated when cells were co-transfected with vectors expressing HNF1 $\alpha$  and HNF1 $\beta$ . The transcriptional activities of the factor IX promoter carrying mutations at -6G/A, -5A/T and -5A/G were not affected by HNF1 $\alpha$  and HNF1 $\beta$  within experimental error. However, transcription from the normal factor IX promoter and the promoter carrying the -6G/C mutation was increased by 20% and 150% respectively.

Equal amounts of each HNF1 expression vector were used in each transfection and the cDNAs were under the control of the same promoter. Equal amounts of HNF1 $\alpha$  and HNF1 $\beta$  protein should therefore be produced. However, a disproportionate ratio of each monomeric protein may be present due to an uneven distribution of vector during transfection. Also, homodimer formation may be selected preferentially by cell-specific

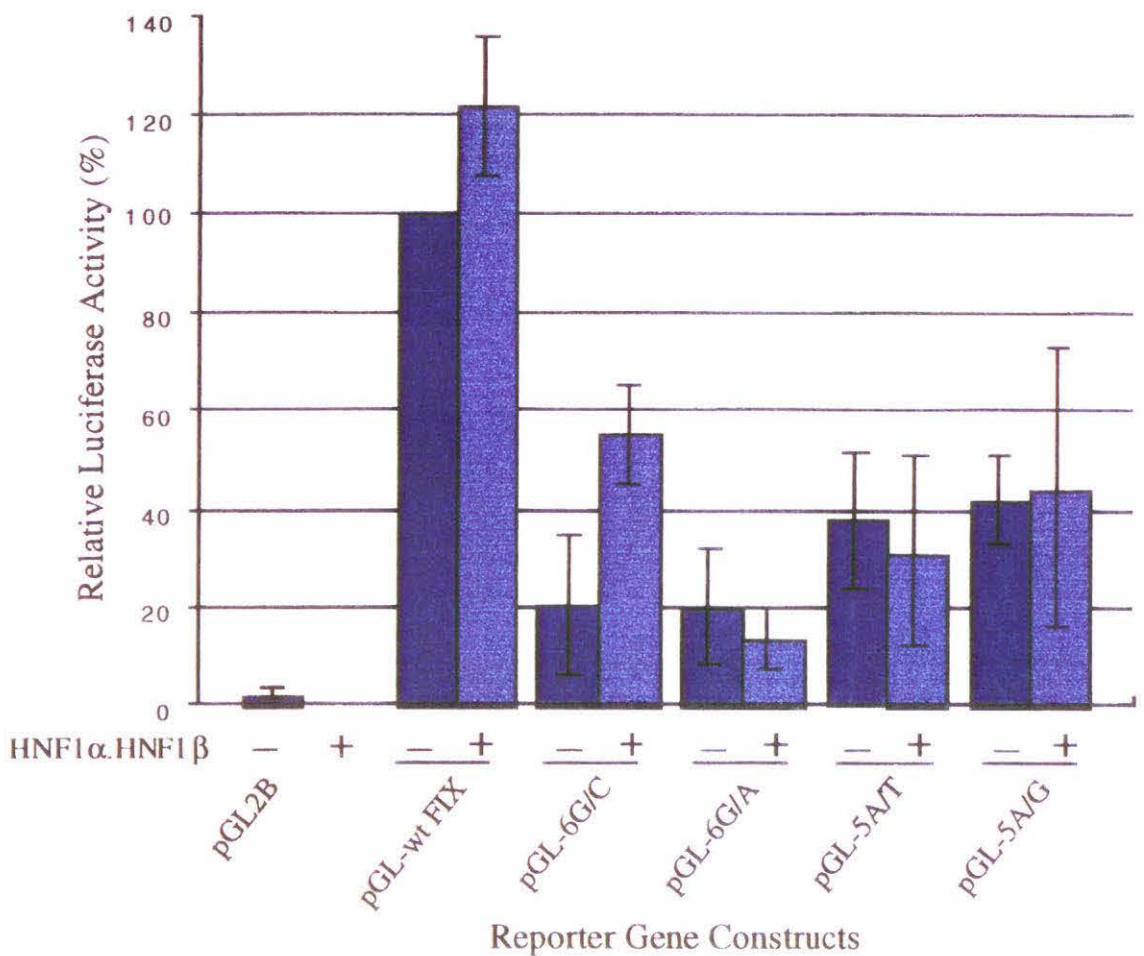


Figure 3.38: Results of the co-transfection of HNF1 $\alpha$  and HNF1 $\beta$  with the factor IX promoters. Alexander cells were transiently transfected with 2  $\mu$ g of reporter constructs that contained DNA corresponding to the factor IX promoter (either normal or -5 or -6 mutant sequence), 1  $\mu$ g each of the expression vectors pSVK3-HNF1 $\alpha$  and pSVK3-HNF1 $\beta$ , plus 0.5  $\mu$ g of pSV- $\beta$ galactosidase vector. 2 $\mu$ g each of the promoter-less pGL2B vector, pSVK3-HNF1 $\alpha$  and pSVK3-HNF1 $\beta$  were also transfected individually as controls. 46 hours after transfection the cells were harvested and assayed for luciferase and  $\beta$ galactosidase activities. 'Relative Luciferase Activity' was calculated by converting the luciferase assay result for each transfection to a percentage of that obtained from the wildtype reporter construct, which was arbitrarily set at 100%. The activities were also normalised with respect to the  $\beta$ galactosidase readings. The resulting activities were compared to that of the promoters in the absence of the HNF1 expression vectors. The results are shown as the mean  $\pm$  the standard error of the mean.



factors over heterodimer formation. As there is no means of preventing the formation of homodimers, it is possible that the activities generated in this experiment might not be representative of the effects of heterodimers. However, as both HNF1 $\alpha$  and HNF1 $\beta$  alone down-regulate the normal factor IX promoter (Figures 3.35 and 3.37) and the results shown in Figure 3.38 clearly indicate up-regulation, it is likely that a significant number of HNF1 $\alpha$ .HNF1 $\beta$  heterodimers have been formed in this experiment.

### 3.11.2 HNF1 Heterodimers Could Potentially Activate Factor IX Gene Expression

The up-regulation of the wildtype factor IX promoter by HNF1 $\alpha$ .HNF1 $\beta$  heterodimers suggests a possible role for HNF1 in the expression of the factor IX gene. Heterodimer formation and the subsequent binding of the transcription factor at the factor IX promoter could, in conjunction with other transcription factors, result in the initiation of transcription from this promoter. The activation observed in these reporter gene assays is however not particularly large (20%) and may not be considered sufficient for transactivation. The initiation of transcription at the wildtype factor IX promoter is the direct result of a concerted input of all the transcription factors binding at the promoter and mutation of the -5 or -6 site *in vivo* produces a sufficient decrease in transcription to cause Haemophilia B. However, in comparison to the other promoter mutations which cause Haemophilia B Leyden, the -5 and -6 promoter mutations are only relatively mild<sup>67</sup>. This suggests that the transactivating factor interacting at this site may not contribute as significantly to the overall expression of the factor IX gene as do the other transcription factors. The function of this transcription factor at the factor IX promoter may therefore have a secondary role, in addition to the stimulation of the promoter. Such a role could be the conferment of tissue-specificity to gene expression (see Section 4.2).

An increase in transcription was also observed with the -6G/C mutant factor IX promoter in response to HNF1 $\alpha$ .HNF1 $\beta$  heterodimers. Such an observation would suggest that in cells containing HNF1 heterodimers, the -6G/C mutation might be phenotypically silent and the patient would therefore appear normal. This could be true if the rate of transcription was increased to a level similar to the transcription from the wildtype promoter. However, in the presence of HNF1 $\alpha$  and HNF1 $\beta$ , both the -6G/C and -5A/T factor IX promoter-reporter gene constructs have activities which are

approximately similar to the -5A/G construct in the absence of HNF1. Comparison of the results of the *in vivo* functional assay with the clinical data suggests that if HNF1 heterodimers could interact at the -6G/C factor IX promoter, a patient carrying this mutation would be likely to remain haemophilic until phenotypic recovery at puberty.

### 3.12 Effects of DCoH on the Activities of the HNF1 Proteins

A set of transfection experiments was conducted to test the effects of the dimerisation co-factor DCoH on the regulatory abilities of HNF1 $\alpha$  and HNF1 $\beta$ . Alexander cells were co-transfected with a mammalian expression vector carrying the cDNA for DCoH, plus a luciferase construct and an HNF1 vector, in a repeat of the previous transfections for HNF1 $\alpha$  and/or HNF1 $\beta$  (Sections 3.9, 3.10 and 3.11). Each of the five factor IX-reporter gene vectors were studied, and the luciferase activity of each transfection was measured and the relative luciferase activity calculated, as detailed in Section 3.8.1. The resulting luciferase activities were compared in graphical form to the relative luciferase activities of the constructs in the presence and absence of an HNF1 transcription factor (Figures 3.35, 3.37 and 3.38).

#### 3.12.1 The *in vitro* Regulatory Ability of HNF1 $\alpha$ is Essentially Unchanged by DCoH

DCoH appears to have had little effect on the transactivation of the factor IX-luciferase construct by HNF1 $\alpha$  (Figure 3.39). The relative luciferase activity of the wildtype factor IX promoter co-transfected with HNF1 $\alpha$  and DCoH is essentially identical to that of the transfection without DCoH. The effect of including DCoH in co-transfections with HNF1 $\alpha$  and the mutant promoter-reporter gene constructs was a tendency to increase the average luciferase activity in each case, by as much as 25% in the case of the -5A/T mutant promoter. However, despite the apparent increase in relative luciferase activity by DCoH, there was no significant change within experimental error. Time limitations for this research prevented further experimentation, so it could be assumed that the sometimes large differences in relative luciferase activity observed after the addition of DCoH represents trends only and are not representative of any real difference. Further HNF1 $\alpha$ , DCoH transfections with the mutant factor IX promoter-



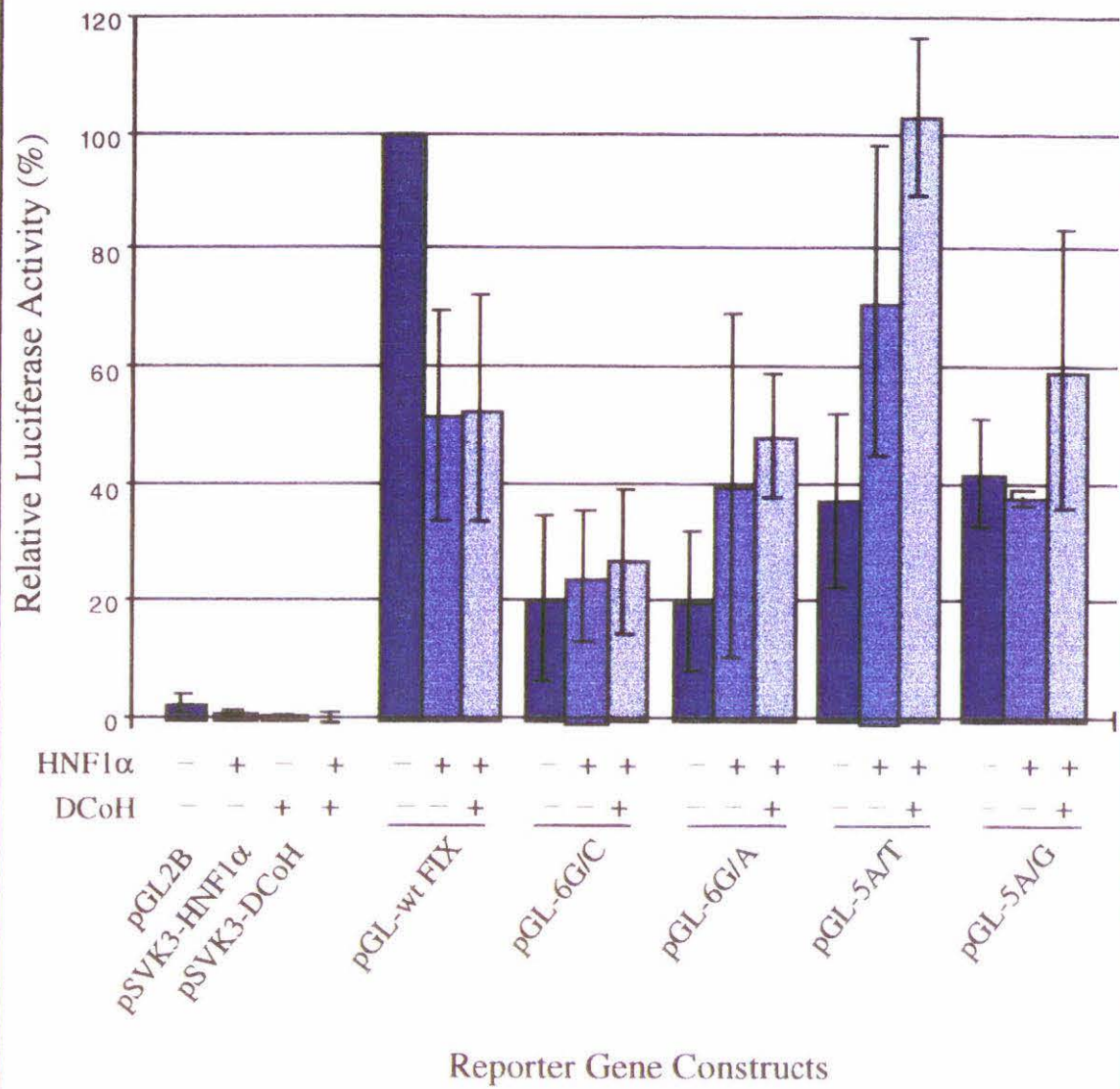


Figure 3.39: Results of the co-transfection of DCoH with HNF1α and the factor IX promoters. Alexander cells were transiently transfected with 2 μg of reporter constructs that contained DNA corresponding to the factor IX promoter (either normal or -5 or -6 mutant sequence), 2 μg of the expression vectors pSVK3-DCoH and pSVK3-HNF1α, plus 0.5 μg of pSV-βgalactosidase vector. 2 μg each of the promoter-less pGL2B vector, pSVK3-HNF1α and pSVK3-DCoH were also transfected individually as controls. 46 hours after transfection the cells were harvested and assayed for luciferase and βgalactosidase activities. 'Relative Luciferase Activity' was calculated by converting the luciferase assay result for each transfection to a percentage of that obtained from the wildtype reporter construct, which was arbitrarily set at 100%. The activities were also normalised with respect to βgalactosidase activity. The resulting activities were compared to that of the unstimulated promoters. The results are presented as the mean ± the standard error of the mean.

luciferase reporter gene constructs could be expected to produce a negligible change in relative luciferase activity compared to the transfections lacking DCoH, as is the case for the wildtype promoter.

DCoH has been shown to have a stabilising effect on HNF1 $\alpha$  homodimers<sup>23</sup>. The fact that the repressing activity of HNF1 $\alpha$  is not enhanced by the presence of DCoH indicates that dimerisation of HNF1 $\alpha$  is not a limiting step in these experiments.

### 3.12.2 The *in vitro* Regulatory Ability of HNF1 $\beta$ May be Unchanged by DCoH

The inclusion of DCoH in the co-transfections of cells with HNF1 $\beta$  and the luciferase constructs tended to increase the average relative luciferase activity in each case (Figure 3.40). The increase in transcriptional activity of the mutant factor IX constructs could provide evidence for the up-regulation of these promoter sequences by HNF1 $\beta$ . Some of these increases were as large as 15% (wildtype factor IX) however, a significant amount of variation was also present in these experiments, as indicated by the large error margins (Figure 3.40). Time limitations prevented sufficient repetition of these experiments to confirm these trends.

The trend towards an increase in luciferase activity in the presence of DCoH and HNF1 $\beta$  may suggest that DCoH could stabilise the regulatory ability of HNF1 $\beta$  dimers, similar to that demonstrated for HNF1 $\alpha$ <sup>23</sup>. This could indicate that the HNF1 $\beta$  dimers present in these experiments were not stable and required DCoH in order to initiate transcription to their full potential. Further experimentation is again required to confirm these trends.



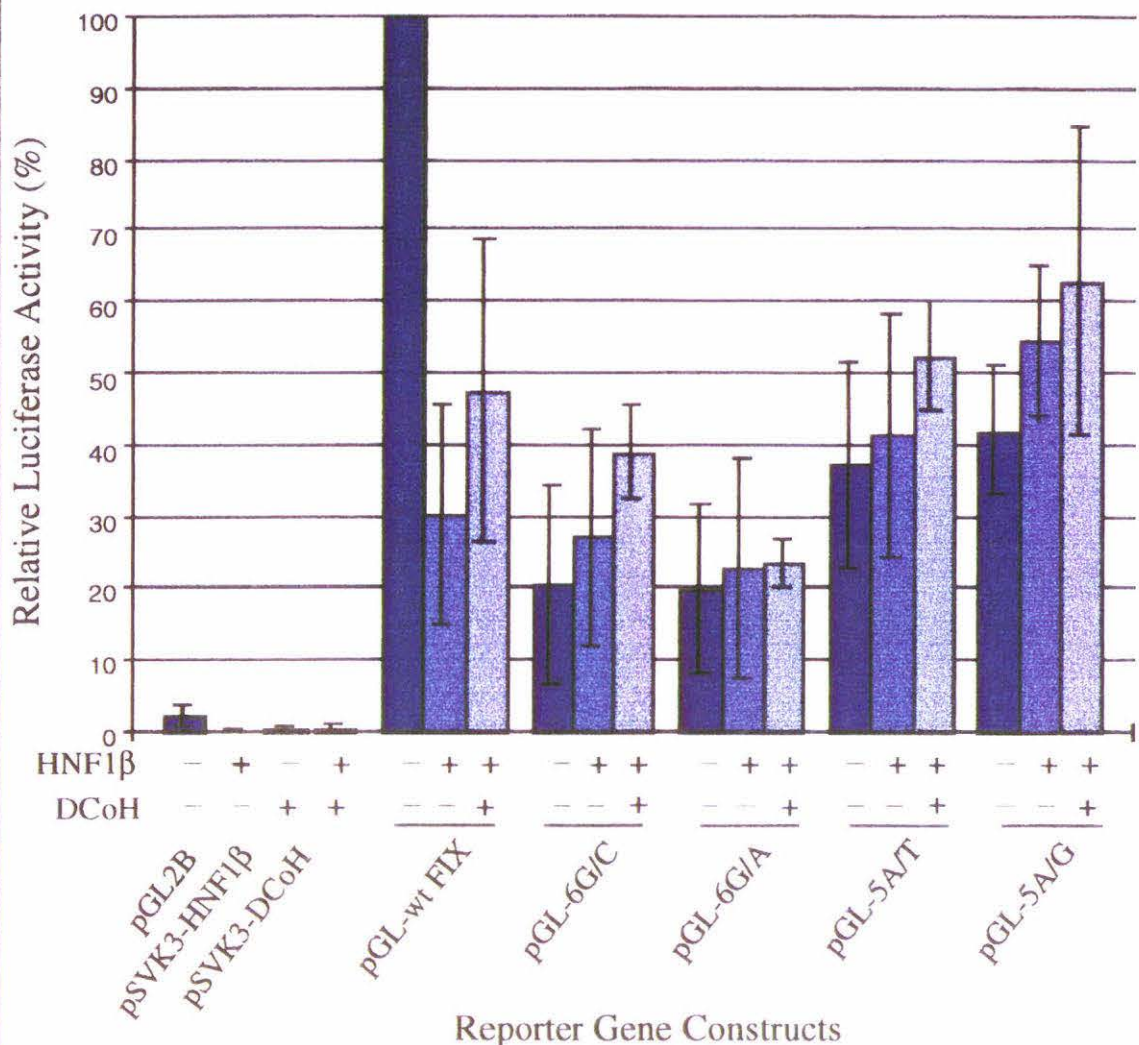


Figure 3.38: Results of the co-transfection of DCoH with HNF1 $\beta$  and the factor IX promoters. Alexander cells were transiently transfected with 2.0  $\mu$ g of reporter construct that contained DNA corresponding to the factor IX promoter (either normal or -5 or -6 mutant sequence), 2.0  $\mu$ g of the expression vectors pSVK3-HNF1 $\beta$  and pSVK3-DCoH, plus 0.5  $\mu$ g of pSV- $\beta$ galactosidase. 2  $\mu$ g each of the promoter-less pGL2B vector, pSVK3-DCoH and pSVK3-HNF1 $\beta$ , were also transfected individually as controls. 46 hours after transfection the cells were harvested and assayed for luciferase and  $\beta$ galactosidase activities. 'Relative Luciferase Activity' was calculated by converting the luciferase assay result for each transfection to a percentage of that obtained from the wildtype reporter construct, which was arbitrarily set at 100%. The activities were also normalised with respect to  $\beta$ galactosidase activity. The resulting activities were compared to that of the unstimulated promoters. The results are shown as the mean  $\pm$  the standard error of the mean.

### 3.12.3: DCoH has a Variable Effect on the *in vitro* Transactivating Ability of HNF1 Heterodimers in Luciferase Transfections

Unlike the previous results of transfections with HNF1 $\alpha$  and HNF1 $\beta$  transfections, DCoH did not appear to have a uniform effect on the regulation of the factor IX promoters by HNF1 $\alpha$  and HNF1 $\beta$  together (Figure 3.41). When DCoH was included in the transfection, the -6G/A, -5A/T and -5A/G promoters were up-regulated slightly, while the relative luciferase activities of the -6G/C and the wildtype promoters were decreased.

Again, the error margins inherent in these experiments cannot allow any firm conclusions to be drawn. The interpretation of these results is also confused by the incomplete characterisation of the activity of DCoH. It is not known whether DCoH stabilises any HNF1 dimer except HNF1 $\alpha$  homodimers. If DCoH does selectively stabilise only HNF1 $\alpha$ , then in this experiment HNF1 $\alpha$  homodimers would be selected for over HNF1 $\alpha$ .HNF1 $\beta$  heterodimers or HNF1 $\beta$  homodimers, which would result in the repression of at least the wildtype factor IX promoter (see Section 3.9). This may explain the decrease in the luciferase activity of the wildtype and -6G/C factor IX promoters in the presence of DCoH. However, if DCoH can stabilise all three potential HNF1 dimers, then any change in luciferase activity in the presence of DCoH would result from an increase in the rate of dimer formation. Unfortunately all of these suggestions must await further experimentation.



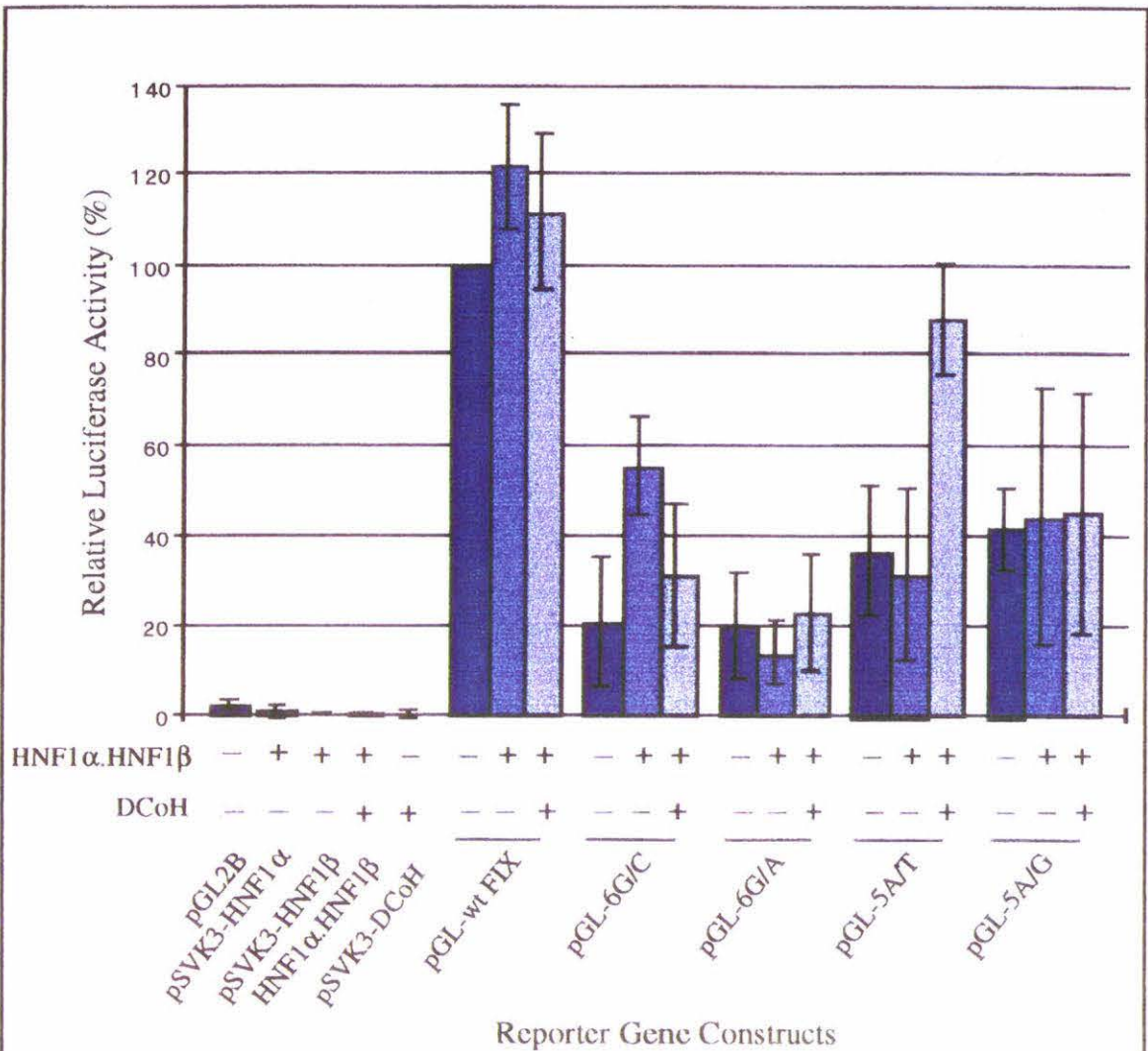


Figure 3.39: Results of the co-transfection of DCoH, HNF1 $\alpha$ , and HNF1 $\beta$  with the factor IX promoters. Alexander cells were transiently transfected with 2  $\mu$ g of reporter constructs that contained DNA corresponding to the factor IX promoter (either normal or -5 or -6 mutant sequence), 1  $\mu$ g each of the expression vectors pSVK3-HNF1 $\alpha$ , pSVK3-HNF1 $\beta$ , plus 2  $\mu$ g of pSVK3-DCoH and 0.5  $\mu$ g of pSV- $\beta$ galactosidase. 2  $\mu$ g each of the promoterless pGL2B vector and pSVK3-HNF1 $\beta$ , were also transfected individually as controls. 46 hours after transfection the cells were harvested and assayed for luciferase and  $\beta$ galactosidase activities. 'Relative Luciferase Activity' was calculated by converting the luciferase assay result for each transfection to a percentage of that obtained from the wildtype reporter construct, which was arbitrarily set at 100%. The activities were also normalised with respect to  $\beta$ galactosidase activity. The resulting activities were compared to that of the unstimulated promoters. The results are represented as the mean  $\pm$  the standard error of the mean.

## CHAPTER 4: FINAL DISCUSSION AND CONCLUSIONS

### 4.1 HNF1 May Bind the Factor IX Promoter and Regulate its Transcription *in vitro*

The results described in the previous chapter suggest that the HNF1 proteins may interact with the factor IX promoter *in vitro*. Gel retardation assays have shown that proteins present in adult rat liver nuclei interact at the wildtype factor IX promoter (Section 3.5). The protein or proteins which complex with the wildtype factor IX promoter can also bind to the strong HNF1 site from the  $\alpha$ 1-antitrypsin promoter. A protein present in this complex may therefore be an HNF1 dimer. The interaction displayed with the wildtype factor IX promoter could be disrupted by adding a competitive amount of an oligonucleotide representing the -6G/C mutant factor IX promoter sequence, indicating that this protein could also interact on this mutant promoter. Other mutations of the factor IX promoter in this region (-5A/T, -5A/G and -6G/A) abolished the interactions between the DNA and this protein, suggesting that the interactions were concentrated at the -5, -6 region of the promoter. The liver protein which binds the factor IX promoter and the HNF1 site from the  $\alpha$ 1-antitrypsin promoter may therefore be an HNF1 dimer but confirmation of the presence of HNF1 in the complex would require immunological detection with antibodies.

As the direct binding of the HNF1 dimer to the factor IX promoter could not be demonstrated with nuclear extracts, Cos cell extracts enriched in one particular dimer needed to be prepared. Time limitations prevented the preparation of a complete set of Cos cell extracts (Section 3.6.1), but one HNF1 $\beta$ -enriched extract was used in gel shift assays (Section 3.6.2). A protein thought to be an HNF1 $\beta$  homodimer was shown to bind directly to the wildtype factor IX promoter *in vitro*, and could also bind to some of the mutant factor IX sequences. In fact, this protein was shown to bind the -6G/A promoter with an affinity higher than that of the wildtype factor IX promoter. The protein could bind to the other factor IX mutant promoters with a range of affinities between that of the -6G/A promoter and that of the -5A/T promoter, which did not display binding in these experiments. The presumed HNF1 $\beta$  homodimers could therefore bind the factor IX promoter in the same -5, -6 region as the liver protein identified in the nuclear extracts, indicating that it is possible that both proteins are HNF1 dimers which bind directly to the wildtype factor IX promoter.

The protein from the liver extracts was only shown to bind the wildtype and -6G/C promoters while the presumed HNF1 $\beta$  homodimer could bind all but the -5A/T promoter. This suggested that the liver protein may not be an HNF1 $\beta$  homodimer. The differences in DNA-binding patterns may be the result of two different transcription factors interacting with the DNA probe, or may be the result of differences in the DNA-binding domains present in two different HNF1 dimers.

Binding of a transcription factor to a DNA sequence does not necessarily imply regulation of that promoter by the protein. These gel shift results cannot confirm the involvement of HNF1 in the regulation of the factor IX promoter. For example, the protein thought to be an HNF1 $\beta$  dimer was shown to bind to the -6G/C promoter, but the affinity that the transcription factor has for this DNA sequence may not be sufficient for anchoring the protein to the DNA during gene regulation. Transcriptional assays were therefore used to test the ability of the HNF1 dimers to regulate transcription from the factor IX promoter. Regrettably, time limits in conjunction with large error margins in assays resulted in incomplete data sets and few definitive results. Only the effects of the HNF1 dimers on transcription from the wildtype factor IX promoter could be shown convincingly (see below). The regulatory effect of the binding of HNF1 $\beta$  to the mutant promoters could not be demonstrated.

## **4.2 HNF1 Could Form the Basis of the Liver-specific Expression of the Factor IX Gene**

Transcriptional assays were used as a method of detecting the effect of the HNF1 proteins on transcription from the factor IX promoter. Due to the large error margins inherent in co-transfection assays, and a lack of time, only the effect of these transcription factors on the wildtype factor IX promoter could be convincingly demonstrated (Sections 3.9 and 3.11). A summary of these results is shown in Table 7 (page 108), where the results of the mutant factor IX promoters which could not be resolved are shown as a strike (-).

The distinct and opposite effects that homodimers and heterodimers have on transcription from the wildtype factor IX promoter suggests a possible role for HNF1 in the regulation of the factor IX gene. The observation that an HNF1 heterodimer possesses an activating ability whereas homodimers were shown to be repressive

suggests that these *in vitro* results may be significant. An obvious explanation of these results pertains to the liver-specific expression of the factor IX gene.

Factor IX Promoter	HNF1 $\alpha$	HNF1 $\beta$	HNF1 $\alpha$ and HNF1 $\beta$
wildtype	decrease	decrease	increase
-6G/C	-	-	increase
-6G/A	-	-	-
-5A/T	-	-	-
-5A/G	-	-	-

Table 7: Summary of the Effect of HNF1 $\alpha$  and HNF1 $\beta$  on Transcription from the Factor IX Promoter.

Of all the transcription factors which have been assigned a role in the regulation of the factor IX gene, none of the proteins implicated have been shown conclusively to constrict the expression of the factor IX gene to liver cells. Despite the liver-enrichment of proteins such as DBP and HNF4, neither display the hallmark of a liver determining factor: the ability to confer a liver phenotype on undifferentiated cells<sup>13</sup>. It has been shown that an HNF1 site in conjunction with a TATA box is both necessary and sufficient to generate a liver-specific promoter *in vitro*<sup>66</sup>, however this may not be relevant in the present example as the factor IX promoter does not contain a TATA box. In addition, the ability of the HNF1 proteins to promote liver specificity has not been well defined. Both HNF1 $\alpha$  and HNF1 $\beta$  are expressed in a number of cell types in addition to the liver, including the kidney and the intestine<sup>35</sup>. However, with the exception of aldolase B, all of the genes shown to be activated by HNF1 are expressed solely in the liver<sup>35</sup>. There must therefore exist some post-transcriptional regulatory control which produces HNF1 dimers with an activating potential in liver, but repressing dimers in other tissues. This control is thought not to be a post-translational protein modification<sup>35</sup>. Isolation of the dimerisation co-factor of HNF1 $\alpha$  (DCoH) generated optimism for a factor that could isolate the activating function of HNF1 to liver cells<sup>23</sup>. However, detection of the wide tissue distribution of DCoH failed to fill the requirements of an isolating co-factor. Another system must therefore exist for the observed control of the HNF1 transcription factors. One such system is heterodimerisation.

At the most rudimentary level, dimerisation is controlled by the availability of monomeric units. In this situation, the selection of dimer components is based solely on



the ratio of one monomer to another. A high concentration of a monomer would therefore result in an increased number of homodimers composed of that subunit. The amounts of each protein present in a cell would be dependent on the relative rates of the expression of each of the monomeric proteins. More elaborate methods of dimer selection may exist, as heterodimerisation is most effective when controlled by factors other than the abundance of monomers<sup>13</sup>.

Dimer selection may also be controlled by the activity of auxiliary proteins, similar to DCoH, which could influence the selection of dimers in different cell types. Alternatively, the selection of a suitable dimer for productive interaction with a promoter may not only be dictated by the *cis*-regulatory sequence present, but also by the adjacent, DNA-bound proteins. Constraints on the choice of dimer may be imposed by the attractive or repulsive interactions between different dimers and neighbouring DNA-bound proteins<sup>45</sup>. Lamb and McKnight liken this form of dimer selection to a three-dimensional jigsaw puzzle<sup>45</sup> (Figure 4.1), where the fitting of a piece is dependent on the existing pieces in place. Binding of the 'correct' dimer would result in the 'solving of the puzzle', that is, the proper temporal and spatial regulation of the associated gene. The choice of dimerisation partner therefore has the potential to substantially alter the affinity or nature of the distal contacts that the resulting transcription factor can participate in. Certain dimers may have a different relative affinity for different promoters based on the other transcription factors which also interact at a particular promoter. The presence of a particular dimer at a promoter may therefore be due to the combination of *cis*-regulatory sequences present at the promoter in addition to the affinity that transcription factor has for its own DNA binding site. In this manner, a complex method for the selection of different dimers of HNF1 may well exist, involving both additional dimerisation co-factors and the transcription apparatus as a whole.

On the assumption of dimer selection, the proposal of a possible interaction for HNF1 with the factor IX promoter is relatively straightforward. Cell types expressing only HNF1 $\alpha$  or HNF1 $\beta$ , or containing only homodimers, would result in the down-regulation of the factor IX gene. Such cell types would include the lung, as it expresses only HNF1 $\beta$ . The other tissues that are known to contain HNF1 proteins, but do not express factor IX are the kidney and the intestine. In order to repress factor IX expression, according to this theory, these cells would need to have low levels of heterodimers and produce significant amounts of homodimers. Conversely, activation of the gene for factor IX in liver cells would require a strong selection for heterodimer formation. Heterodimers of HNF1 $\alpha$  and HNF1 $\beta$  have been identified in mouse Hepa1A cells and liver nuclear extracts by gel shifts and UV crosslinking<sup>20</sup>. However, kidney

nuclear extracts also contain heterodimers<sup>20</sup>, and HNF1 $\alpha$  homodimers are believed to activate many promoters in the liver<sup>13</sup>. Another factor which must also be taken into account is the presence of HNF1 isoforms<sup>22</sup>.

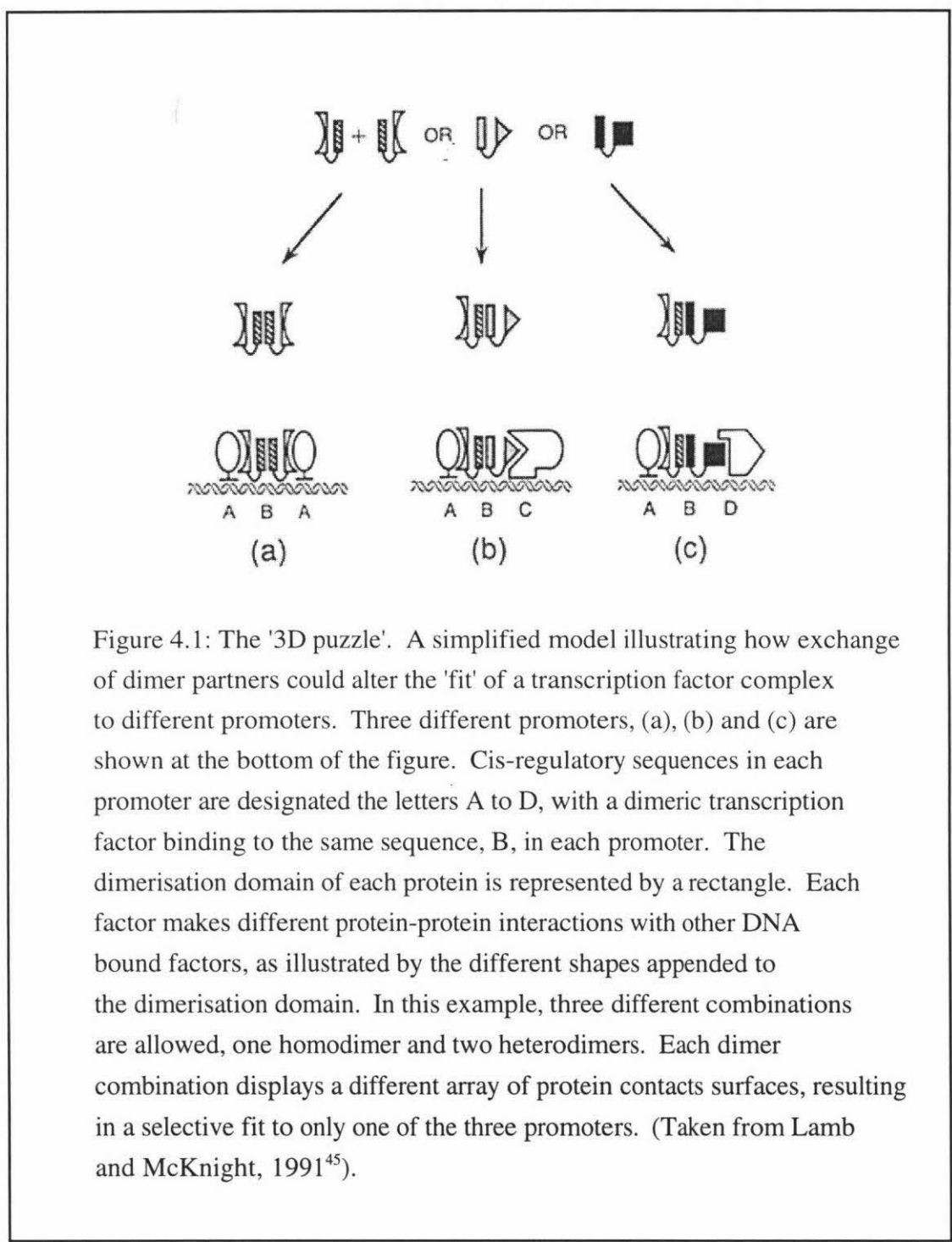


Figure 4.1: The '3D puzzle'. A simplified model illustrating how exchange of dimer partners could alter the 'fit' of a transcription factor complex to different promoters. Three different promoters, (a), (b) and (c) are shown at the bottom of the figure. Cis-regulatory sequences in each promoter are designated the letters A to D, with a dimeric transcription factor binding to the same sequence, B, in each promoter. The dimerisation domain of each protein is represented by a rectangle. Each factor makes different protein-protein interactions with other DNA bound factors, as illustrated by the different shapes appended to the dimerisation domain. In this example, three different combinations are allowed, one homodimer and two heterodimers. Each dimer combination displays a different array of protein contacts surfaces, resulting in a selective fit to only one of the three promoters. (Taken from Lamb and McKnight, 1991<sup>45</sup>).



The tissue-specific expression of each of these protein forms has not been investigated, nor has the regulatory potentials of each possible dimer combination been studied. The presence of the different isoforms therefore further complicates the delineation of the regulatory function of the HNF1 proteins.

The regulation of promoter function by an HNF1 dimer may also be confounded by the possibility of a double regulatory potential in some dimers. The observation that HNF1 $\alpha$  represses transcription of a reporter gene in the presence of a potential HNF1 site within the promoter under study (Section 3.9) is currently unprecedented. However, if these results could be substantiated, then HNF1 $\alpha$  may possess both an activating and a suppressing potential. Additionally, the role of HNF1 $\beta$  at the albumin promoter has been demonstrated as being activating in differentiated liver cells but repressing in dedifferentiated cells<sup>13</sup>. A dual regulatory function may therefore also be implied for HNF1 $\beta$ . The control of the switch between the two activities of these potentially bifunctional transcription factors would also affect the overall regulatory effect of the HNF1 proteins. Selective formation of HNF1 dimers and their activities in both liver and non-liver cells is no doubt a complex process dependent on many factors.

#### 4.2.1 Is the Liver Protein Identified in Gel Shifts an HNF1 $\alpha$ .HNF1 $\beta$ Heterodimer?

HNF1 $\alpha$  and HNF1 $\beta$  mRNAs are expressed at approximately the same level in liver, however some post-transcriptional regulation of the HNF1 $\beta$  mRNA results in HNF1 $\alpha$  being the more dominant monomeric protein<sup>13</sup>. The higher abundance of HNF1 $\alpha$  monomer in liver may not necessarily mean that HNF1 $\alpha$  homodimers are the most abundant HNF1 dimer present in liver. All three HNF1 dimers have been identified in liver nuclear extracts<sup>20</sup>, and the major form of the HNF1 transcription factor in the human liver could be any one of these dimers. The potential for the formation of a number of dimers is also a feature of the C/EBP family of transcription factors. Four conserved C/EBP proteins have been isolated (C/EBP $\alpha$ , C/EBP $\beta$ , C/EBP $\gamma$  and C/EBP $\delta$ ), and have been shown to cross-dimerise *in vitro*<sup>67</sup>. The cDNAs of C/EBPs  $\alpha$ ,  $\beta$  and  $\delta$  were isolated from a single cell type (adipocytes)<sup>67</sup>, suggesting that all three proteins may be present simultaneously in a single cell. A large array of potential transcription factors with similar DNA-binding affinities<sup>67</sup> may therefore form *in vivo*. The formation of

dimers may therefore be a common occurrence in eukaryotic cells and it is probable that complex processes exist to control dimer selection.

On the basis of the results presented here, it is tempting to conclude that the liver form of HNF1 is heterodimeric. There are some consistencies between the results of the gel shifts using liver nuclear extracts, and the transcriptional assays for HNF1 $\alpha$  and HNF1 $\beta$  together. The gel shifts using a rat liver nuclear extract showed that a liver protein can bind to both the wildtype and the -6G/C mutant factor IX promoters. The only significant results observed in the luciferase transfections with HNF1 $\alpha$  and HNF1 $\beta$  was the activation of the wildtype and -6G/C promoters. However, this does not indicate that the liver protein is an HNF1 $\alpha$ .HNF1 $\beta$  heterodimer. The liver protein shown to bind the wildtype factor IX promoter in the gel shifts could be an HNF1 $\alpha$  homodimer. Comparison of this result with that for an HNF1 $\beta$  homodimer suggests that the liver-enriched transcription factor is not HNF1 $\beta$ . However, the lack of conclusive results regarding the effect of HNF1 $\alpha$  on transcription from the mutant factor IX promoters, plus the inability to carry out DNA-binding studies with HNF1 $\alpha$ , means that the identity of the liver HNF1 protein must remain at present unknown.

### 4.3 Future Work.

The collective results of this research has suggested that HNF1 may play a role in the regulation of the human factor IX gene. Many new questions have also been raised regarding the activity of the HNF1 dimers in liver cells. Possibilities for further work are as follows:

- The ability of the HNF1 $\beta$  homodimer to bind to the factor IX promoter could be investigated further using gel mobility shifts assays with the -14 to +6 region of the wildtype factor IX promoter as a probe. This would provide more definitive evidence for the ability of an HNF1 protein to bind to the factor IX promoter.
- An investigation into the abilities of the HNF1 $\alpha$  homodimer and HNF1 $\alpha$ .HNF1 $\beta$  heterodimers to bind the factor IX promoter would also be necessary. Gel mobility shift assays using the -14 to +6 region of the wildtype factor IX promoter and the mutant factor IX promoters as competitor oligonucleotides could be used here. Supershift assays with antibodies to HNF1 $\alpha$  and HNF1 $\beta$  could be used to confirm the presence of a heterodimer.

- Identification of the DNA-binding proteins present in the rat liver nuclear extracts (Section 3.5) may also identify the major *in vivo* form of HNF1. This information would allow a critical assessment of the results presented here. The simplest method of identification would be to use antibodies specific for HNF1 $\alpha$  and HNF1 $\beta$ . More extensive work would require purification of individual proteins from liver nuclear extracts. This would allow the identification of the proteins present in the three complexes observed in the gel shift assays with rat liver nuclear extracts (Section 3.5).
- The luciferase reporter gene assays using the factor IX promoters and the HNF1 proteins yielded incomplete information due to the large error margins. Further repetitions of each of these experiments could provide sufficient data to generate statistically significant results. The use of a dual reporter vector may also reduce variability. The effects of the various HNF1 dimers on the mutant factor IX promoters could then be examined quantitatively, which could supply more information regarding the Haemophilia B Leyden phenotype.

For the HNF1 transcription factors to be considered for a role in transcription from the factor IX promoter *in vivo*, a number of results would need to be observed. The HNF1 $\alpha$ .HNF1 $\beta$  heterodimer has been shown to up-regulate transcription from the wildtype factor IX promoter *in vitro* (Section 3.11). If the transcription factor binds at the -5, -6 region of the promoter, the heterodimer must not increase the transcription of the mutant promoters above that of the unstimulated promoters, as is observed for the -6G/C mutant promoter. In this way, the net transcription from the mutant promoters would still produce insufficient factor IX protein for the patients carrying these mutations to appear normal. The experimental results would therefore still agree with the disease phenotype. It is expected that the transcription from the mutant factor IX promoters would not be markedly altered in the presence of HNF1 $\alpha$  or HNF1 $\beta$  homodimers due to the repressing effects demonstrated by these transcription factors in previous experiments (Sections 3.9 and 3.10). Clarification of a possible role for the HNF1 proteins in the Leyden phenotype therefore requires more experimentation.

The results detailed here, plus the experiments suggested above form an initial study into the role of the HNF1 proteins in the regulation of the factor IX promoter. If a definite role for HNF1 in the transcription of the factor IX gene can be determined, other confirmatory experiments would need to be carried out.

- DNase I footprinting uses the ability of a DNA-bound transcription factor to protect its recognition sequence from cleavage with DNase I. Electrophoresis of the end-labelled oligonucleotide sequence after digestion with DNase I produces a gel pattern similar to that seen on DNA sequencing gels. Comparison of the gel patterns for DNA sequences with and without a bound transcription factor can identify those nucleotides with which the transcription factor is interacting. This method could be used to precisely localise the binding of HNF1 dimers to the factor IX promoter sequence. This is a more powerful technique than the gel shift because it can detect the size of the region to which a transcription factor binds. It could also serve to distinguish between the binding activities of HNF1 $\alpha$  homodimers, HNF1 $\beta$  homodimers and HNF1 $\alpha$ -HNF1 $\beta$  heterodimers.
- A consensus binding sequence for HNF1 $\beta$  homodimers and the HNF1 heterodimer could be generated by compiling the known sequences to which these transcription factors bind and 'averaging' them into one sequence. Comparison of these sequences with the factor IX promoter, with respect to the results of the DNA-binding assays, may help to identify a feasible binding site on the factor IX promoter for HNF1.
- Identification of the tissue-specific expression of the HNF1 isoforms could be accomplished by RNase protection assays and would provide indirect information as to the role of the HNF1 proteins *in vivo*.

A set of studies may be undertaken to identify the molecular mechanism by which the HNF1 activating potential is retained in liver cells and not in other cells which express HNF1. This may be a complex process involving many different protein factors and specific conditions.

- The primary task in characterising this process is to clarify the role of DCoH in the stabilisation of HNF1 dimers. The effect of DCoH on HNF1 $\beta$  homodimers and HNF1 heterodimers could be studied by reporter gene assays. Restriction of the stabilising ability of DCoH to HNF1 $\alpha$  homodimers may provide a means of dimer selection, and therefore selection of the HNF1 activating potential.
- If DCoH stabilises only HNF1 $\alpha$  dimers, then dimerisation co-factors for the other HNF1 dimer combinations could also exist and be isolated by a two-hybrid yeast expression system. The expression pattern of any new co-factors could be studied with RNase protection assays, as restriction of these stabilising factors to various cell

types may form the basis of the tissue-specificity of the HNF1 transactivating potential.

The transcription factor which regulates the factor IX promoter through *cis*-regulatory sequences in the -5 and -6 region remains unidentified. However, the research presented in this study suggests the possible involvement of the members of the HNF1 family in the regulation of the factor IX promoter. The complete characterisation of the mechanisms by which both the factor IX gene and the HNF1 proteins are regulated will lead to the greater understanding of the control of liver cell function and gene regulation in general.

## REFERENCES

- <sup>1</sup> Crossley, M., Winship, P. R., Austen, D. E. G., Rizza, C. R., and Brownlee, G. G. (1990) A less severe form of Haemophilia B Leyden. *Nucleic Acids Res.* **18**, 4633.
- <sup>2</sup> Furie, B., and Furie, B. C. (1988) The molecular basis of blood coagulation. *Cell* **53**, 505-518.
- <sup>3</sup> Royle, G., Van De Water, N. S., Berry, E., Ockelford, P. A., and Browett, P. J. (1991) Haemophilia B Leyden arising de novo by point mutation in the putative factor IX promoter region. *Br. J. Haem.* **77**, 191-194.
- <sup>4</sup> Stryer, L. (1988) *Biochemistry*. Third edition. W. H. Freeman and Co, New York.
- <sup>5</sup> Stowell, K. M., Figueiredo, M. S., Brownlee, G. G., Jones, P., and Bolton-Maggs, P. H. B. (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. Haem.* **85**, 188-190.
- <sup>6</sup> 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**, 444-446.
- <sup>7</sup> Reijnen, M. J., Peerlinck, K., Maasdam, D., Bertina, R. M., Reitsma P. H. (1993) Haemophilia 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**, 151-158.
- <sup>8</sup> Dynan, W. S., and Tjian, R. (1985) Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature* **316**, 774-778.
- <sup>9</sup> Mitchell, P. J. and Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA-binding proteins. *Science* **245**, 371-378.

- <sup>10</sup> Voss, S. D., Schlokot, U. and Gruss, P. (1986) The role of enhancers in the regulation of cell-type-specific transcriptional control. *TiBS* **11**, 287-289.
- <sup>11</sup> Ptashne, M. (1988) How eukaryotic transcriptional activators work. *Nature* **335**, 683-689.
- <sup>12</sup> Wolffe, A. P., and Brown, D. D. (1988) Developmental regulation of two 5S ribosomal RNA genes. *Science* **241**, 1626-1632.
- <sup>13</sup> Mendel, D. B., Hansen, L. P., Graves, M. K., Conley, P. B., and Crabtree, G. R. (1991) HNF1 $\alpha$  and HNF1 $\beta$  (vHNF1) share dimerisation and homeo domains, but not activation domains, and form heterodimers *in vitro*. *Genes Dev* **5**, 1042-1056.
- <sup>14</sup> 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 haemophilia B Leyden. *Proc. Natl. Acad. Sci. USA* **89**, 6300-6303.
- <sup>15</sup> Crossley, M., Ludwig, M., Stowell, K. M., De Vos, P., Olek, K., and Brownlee, G. G. (1992) Recovery from Haemophilia B Leyden: an androgen-responsive element in the factor IX promoter. *Science* **257**, 377-379.
- <sup>16</sup> 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. *Nature genetics* **3**, 175-179.
- <sup>18</sup> Pontoglio, M., Faust, D. M., Doyen, A., Yaniv, M., and Weiss, M. C. (1997) Hepatocyte Nuclear Factor 1 gene inactivation impairs chromatin remodelling and demethylation of the phenylalanine hydroxylase gene. *Mol. Cell. Biol.* **17**, 4948-4956.
- <sup>19</sup> Chouard, T., Blumenfeld, M. Bach, I., Vandekeckhove, J., Cereghini, S., and Yaniv, M. (1990) A distal dimerisation domain is essential for DNA-binding by the atypical HNF1 homeodomain. *Nucleic Acids Res.* **18**, 5853-5863.



- <sup>20</sup> Rey-Campos, J., Chouard, T., Yaniv, M., and Cereghini, S. (1991) vHNF1 is a homeoprotein that activates transcription and forms heterodimers with HNF1. *EMBO J.* **10**, 1445-1457.
- <sup>21</sup> Toniatti, C., Monaci, P., Nicosia, A., Cortese, R., and Ciliberto, G. (1993) A bipartite activation domain is responsible for the activity of transcription factor HNF1/LFB1 in cells of hepatic and nonhepatic origin. *DNA Cell Biol* **12**, 199-208.
- <sup>22</sup> Bach, I., and Yaniv, M. (1993) More potent transcriptional activators or a transdominant inhibitor of the HNF1 homeoprotein family are generated by alternative RNA processing. *EMBO J.* **12**, 4229-4242.
- <sup>23</sup> Mendel, D. B., Khavari, P. A., Conley, P. B., Graves, M. K., Hansen, L. P., Admon, A., and Crabtree, G. R. (1991) Characterization of a cofactor that regulates dimerization of a mammalian homeodomain protein. *Science* **254**, 1762-1767.
- <sup>24</sup> Bach, I., Pontoglio, M., and Yaniv, M. (1992) Structure of the gene encoding hepatocyte nuclear factor 1 (HNF1). *Nucleic Acids Res* **20**, 4199-4204.
- <sup>25</sup> Bach, I., Mattei, M., Cereghini, S., and Yaniv, M. (1991) Two members of an HNF1 homeoprotein family are expressed in human liver. *Nucleic Acids Res.* **19**, 3553-3559.
- <sup>26</sup> Leiting, B., De Francesco, R., Tomei, L., Cortese, R., Otting, G., and Wuthrich, K. (1993) The three-dimensional NMR-solution structure of the polypeptide fragment 195-286 of the LFB1/HNF1 transcription factor from rat liver comprises a non-classical homeodomain. *EMBO J.* **12**, 1797-1803.
- <sup>27</sup> Nicosia, A., Tafi, R., and Monaci, P. (1992) Trans-dominant inhibition of transcription activator LFB1. *Nucleic Acids Res* **20**, 5321-5328.



- <sup>28</sup> Ceska, T. A., Lamers, M., Monaci, P., Nicosia, A., Cortese, R., and Suck, D. (1993) The X-ray structure of an atypical homeodomain present in the rat liver transcription factor LFB1/HNF1 and implications for DNA binding. *EMBO J.* **12**, 1805-1810.
- <sup>29</sup> Verrijzer, C. P., Alkema, M. J., van Weperen, W. W., Van Leeuwen, H. C., Strating, M. J. J., and van der Vliet, C. P. (1992) The DNA binding specificity of the bipartite POU domain and its subdomains. *EMBO J.* **11**, 4993-5003.
- <sup>30</sup> Strandmann, E., and Ryffel, G. U. (1995) Developmental expression of the maternal protein XDCoH, the dimerization cofactor of the homeoprotein LFB1 (HNF1). *Development* **121**, 1217-1226.
- <sup>31</sup> Stern, S., Tanaka, M., and Herr, W. (1988) The Oct-1 homoeodomain directs formation of a multiprotein-DNA complex with the HSV transactivator VP16. *Nature* **341**, 624-630.
- <sup>32</sup> Wu, K., Wilson, D. R., Shih, C., and Darlington, G. J. (1994) The transcription factor HNF1 acts with C/EBP $\alpha$  to synergistically activate the human albumin promoter through a novel domain. *J. Biol. Chem* **269**, 1177-1182.
- <sup>33</sup> Cereghini, S., Ott, M., Power, S., and Maury, M. (1992) Expression patterns of vHNF1 and HNF1 homeoproteins in early postimplantation embryos suggest distinct and sequential developmental roles. *Development* **116**, 783-797.
- <sup>34</sup> Ringeisen, F., Rey-Campos, J., and Yaniv, M. (1993) The transactivation potential of variant hepatocyte nuclear factor 1 is modified by alternative splicing. *J. Biol. Chem.* **268**, 25706-25711.
- <sup>35</sup> Blumenfeld, M., Maury, M., Chourd, T., Yaniv, M., and Condamine, H. (1991) Hepatic Nuclear Factor 1 (HNF1) shows a wider distribution than products of its known target genes in developing mouse. *Development* **113**, 589-599.

- <sup>36</sup> Rhee, K., Stier, G., Becker, P. B., Suck, D., and Sandaltzopoulos, R. (1997) The bifunctional protein DCoH modulates interactions of the homeodomain transcription factor HNF1 with nucleic acids. *J. Mol. Biol.* **265**, 20-29.
- <sup>37</sup> Ficner, R., Sauer, U. H., Stier, G., and Suck, D. (1995) Three-dimensional structure of the bifunctional protein PCD/DCoH, a cytoplasmic enzyme interacting with transcription factor HNF1. *EMBO J.* **14**, 2034-2042.
- <sup>38</sup> Citron, B. A., Davis, M. D., Milstien, S., Gutierrez, J., Mendel, D. B. Crabtree, G. R. and Kaufman, S. (1992) Identity of 4a-carbinolamine dehydratase, a component of the phenylalanine hydroxylation system, and DCoH, a transregulator of homeodomain proteins. *Proc. Natl. Acad. Sci. USA* **89**, 11891-11894.
- <sup>39</sup> Conaway, R. C., and Conaway, J. W. (1989) An RNA polymerase II transcription factor has an associated DNA-dependent ATPase (dATPase) activity strongly stimulated by the TATA region of promoters. *PNAS* **19**, 7356-7360.
- <sup>40</sup> Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J. H., Chambon, P., and Egly, J. M. (1993) DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science* **260**, 58-63.
- <sup>41</sup> Park, E., Guzder, S. N., Koken, M. H., Jaspers-Dekker, I., Weeda, G., Hoeijmakers, J. H., Prakash, S., and Prakash, L. (1992) RAD25 (SSL2), the yeast homolog of the human xeroderma pigmentosum group B DNA repair gene, is essential for viability. *PNAS* **89**, 11416-11420.
- <sup>42</sup> Goodrich, J. A., and Tjian, R. (1994) Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell* **8**, 145-156.
- <sup>43</sup> Salier, J., Hirose, S. and Kurachi, K. (1990) Functional characterisation of the 5'-regulatory region of human Factor IX gene. *J. Biol. Chem.* **265**, 7062-7068.

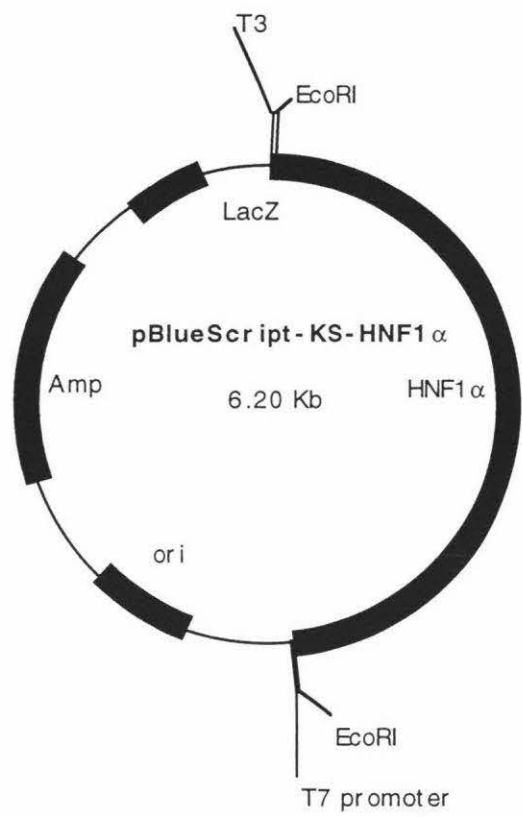
- <sup>44</sup> Naka, H., and Brownlee, G. G. (1996) Transcriptional regulation of the human factor IX promoter by the orphan receptor superfamily factors, HNF4, ARP1 and COUP/Ear3. *British Journal of Haematology* **92**, 231-240.
- <sup>45</sup> Lamb, P. and McKnight, S. L. (1991) Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerisation. *TiBS* **16**, 417-422.
- <sup>46</sup> Franklin, I. K. (1996) Masters thesis.
- <sup>47</sup> Holmes, D. S. and Quigley, M. (1981) A rapid boiling method for the preparation of bacterial plasmids. *Anal Biochem* **114**, 193-197.
- <sup>48</sup> Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989a) eds. *Molecular Cloning: A Laboratory Manual*. Second edition, book 3, Cold Springs Harbour Laboratory Press.
- <sup>49</sup> Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G., and Struhl, K. (1989) eds. *Current Protocols in Molecular Biology*. Volume I. Greene Publishing Associates and Wiley Intersciences NY.
- <sup>50</sup> Thuring, R. W., Sanders, J. P. M. and Borst, P. (1975) A freeze-squeeze method for recovering long DNA from agarose gels. *Anal Biochem* **66**, 213-220.
- <sup>51</sup> Sanger, Factor IX., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain terminating inhibitors. *PNAS* **74**, 1677-78.
- <sup>52</sup> Giannelli, F., Green, P. M., Sommer, S. S., Poon, M., Ludwig, M., Schwaab, R., Reitsma, P. H., Goossens, M., Yoshioka, A., Figueiredo, M. S., and Brownlee, G. G. (1998) Haemophilia B: database of point mutations and short additions and deletions--eighth edition. *Nuc. Acids Res.* **26**, 265-268.
- <sup>53</sup> Piaggio, G., Tomei, L., Toniatti, C., De Francesco, R., Gerstner, J., and Cortese, R. (1994) LFB1/HNF1 acts as a repressor of its own transcription. *Nuc. Acids Res.* **22**, 4284-4290.

- <sup>54</sup> Miura, N., and Tanaka, K. (1993) Analysis of the rat hepatocyte nuclear factor (HNF) 1 gene promoter: synergistic activation by HNF4 and HNF1 proteins. *Nucleic Acids Res.* **21**, 3731-3736.
- <sup>55</sup> Shi, Y., Seto, E., Chang, L. and Shenk, T. (1991) Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and the relief of repression by adenovirus E1A protein. *Cell* **67**, 377-388.
- <sup>56</sup> Hariharan, N., Kelley, D. E., and Perry, R. P. (1991) Delta, a transcription factor that binds to downstream elements in several polymerase II promoters, is a functionally versatile zinc finger protein. *PNAS* **88**, 9799-9803.
- <sup>57</sup> Park, K., and Atchison, M. L. (1991) Isolation of a candidate repressor/activator, NF-E1 (YY-1, delta), that binds to the immunoglobulin kappa 3' enhancer and the immunoglobulin heavy-chain mu E1 site. *PNAS* **88**, 9804-9808.
- <sup>58</sup> Sauer, F. and Jackle, H. (1993) Dimerisation and the control of transcription by Kruppel. *Nature* **364**, 454-457.
- <sup>59</sup> Wang, Z., Qiu, Q., and Deuel, T. Factor IX. (1993) The Wilm's Tumour gene product WT1 activates or suppresses transcription through separate functional domains. *J. Biol. Chem.* **268**, 9172-9175.
- <sup>60</sup> Thomas, M. and Davis, R. W. (1975) Studies on the cleavage of bacteriophage lambda DNA with *EcoRI* restriction endonuclease. *J. Mol. Biol.* **91**, 315-328.
- <sup>61</sup> Blaho, J. A., Larson, J. E., McLean, M. J. and Wells, R. D. (1988) Multiple DNA secondary structures in perfect inverted repeat inserts in plasmids. Right-handed B-DNA, cruciforms, and left-handed Z-DNA. *J. Biol. Chem.* **263**, 14446-14455.
- <sup>62</sup> Kleff, S., Kemper, B. and Sternglanz, R (1992) Identification and characterization of yeast mutants and the gene for a cruciform cutting enzyme. *EMBO J.* **11**, 699-704.

- <sup>63</sup> Palecek, E. (1991) Local supercoil-stabilized DNA structures. *Crit. Rev. Biochem. Mol. Biol.* **26**, 151-226.
- <sup>64</sup> Singleton, C. K. (1983) Effects of salts, temperature, and stem length on supercoil-induced formation of cruciforms. *J. Biol. Chem.* **258**, 7661-7668.
- <sup>65</sup> Scholten, P. M. and Nordheim, A. (1986) Diethyl pyrocarbonate: a chemical probe for DNA cruciforms. *Nucleic Acids Res.* **14**, 3981-3993.
- <sup>66</sup> Ryffel, G. U., Kugler, W., Wagner, U. and Kaling, M. (1989) Liver-specific gene transcription in vitro: the promoter element HP1 and a TATA box are necessary and sufficient to generate a liver-specific promoter. *Nuc. Acids Res.* **17**, 939-953.
- <sup>67</sup> Cao, Z., Umek, R. M., and McKnight, S. L. (1991) Regulation of three C/EBP isoforms during adipose conversion of 3T3-L1 cells. *Genes Dev* **5**, 1538-1552.

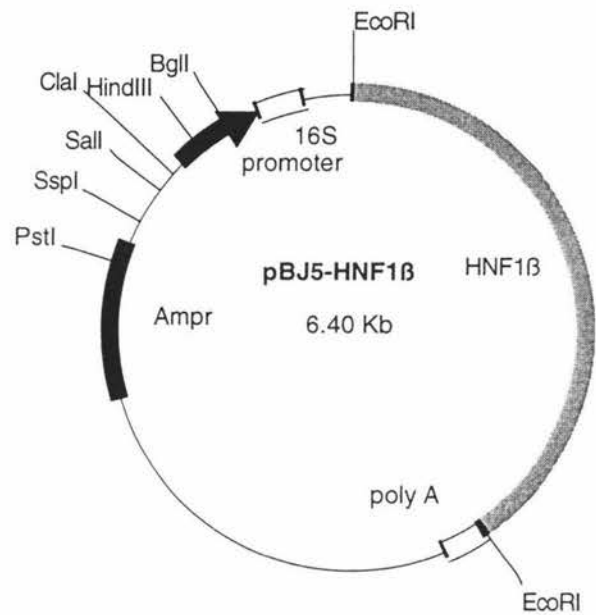
**Appendix A: Plasmid Maps.**

*pBlueScript-KS-HNF1 $\alpha$ .*



HNF1 $\alpha$  Genbank Accession Number: M57966  
 pBlueScript Genbank Accession Number: U25267

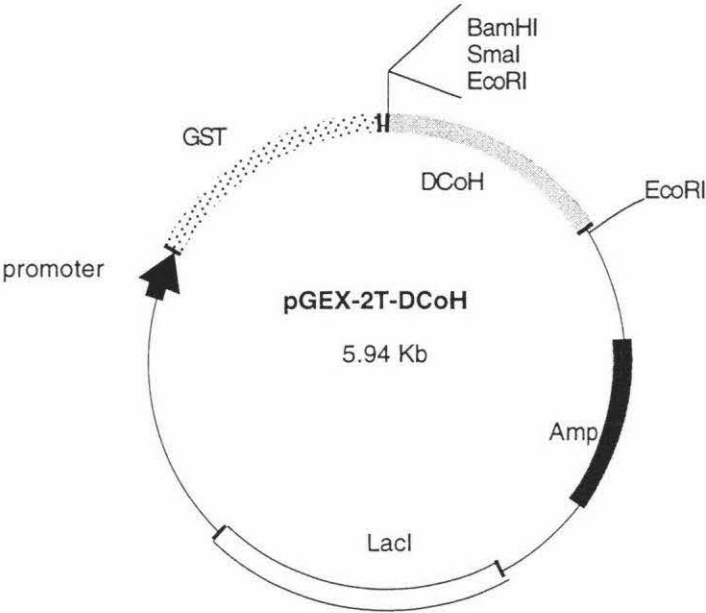
*pBJ5-HNF1 $\beta$* .



HNF1 $\beta$  Genbank Accession Number: X71346

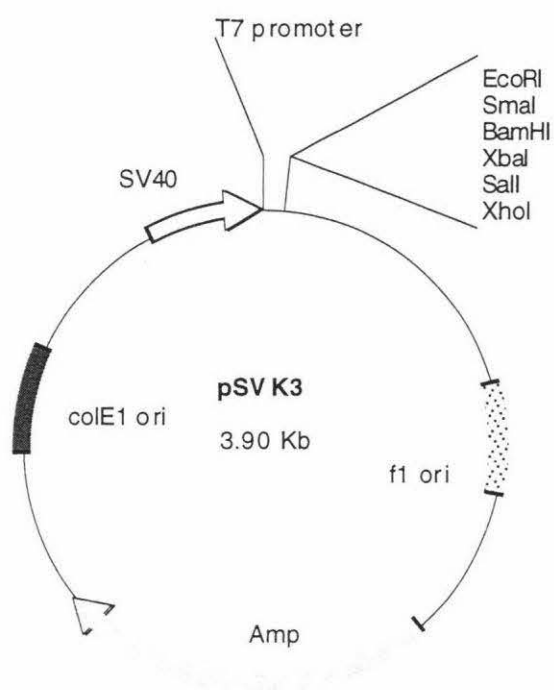


*pGEX-2T-DCoH.*



DCoH Genbank Accession Number: M83740  
pGEX-2T Genbank Accession Number: U13850

*pSVK3*.



pSVK3 Genbank Accession Number: U13867

## Appendix B: Oligonucleotides

### *HNF1 $\alpha$ PCR Primers:*

5' primer	GGCCGAATTCCTTCACTGGCCGCTGGGGCCAG
3' primer	CCGGGTCGACGATATCAAGCTTCCACGTGC
2290*	GGCCGTCGACCTCCAGCTCCTCGAGGCTCGCTTGTCAGGC

### *HNF1 $\alpha$ Sequencing Primers:*

500	GCCGTGGTGGAGTCACTTCTTCAGG
567*	GGATGTTGTGCTGCTGC
893	CCTTGGTGGAAGAGTGTAATAG
1382	CAACCAGCAGCCGCAGAACC
1818*	GACCTGCTTGGTGGGTGTGAG

### *D $\alpha$ CoH Primers:*

5' primer	GCGCTCTAGAATGGCTGGCAAGGCACACAGGC
3' primer	CCGGCTCGAGCTATGTCATAGACTCGGCAACTTGTTTCG

### *Gel Mobility Shift Oligonucleotides:*

HNF1 site from $\alpha$ 1-antitrypsin promoter	CCTTGGTTAATATTCACC
	GGAAGGTGAATATTAACC

Wildtype factor IX promoter (-14 to +6 region)

AGCTTAACTAATCGACCTTACCACTA  
AGCTTAGTGGTAAGGTCGATTAGTTA

-6G/C factor IX promoter sequence

AGCTTTGGTACAATAATCCACCTTACCACTA  
AGCTTAGTGGTAAGGTGGATTAGTTGTACCA

-6G/A factor IX promoter sequence

TTGGTACAACCTAATCAACCTTACCACTA  
TAGTGGTAAGGTTGATTAGTTGTACCAA

-5A/T factor IX promoter sequence

TTGGTACAACCTAATCGTCCTTACCACTA  
TAGTGGTAAGGACGATTAGTTGTACCAA

-5A/G factor IX promoter sequence

CAACTAATCGGCCTTACCAC  
GTGGTAAGGCCGATTAGTTG

Note: Numerically-named primers which anneal to the non-coding strand are indicated by an \*.

**Appendix C: Lalign Results.**

Hybridisation of the HNF1α 3' primer to the coding strand of HNF1α cDNA:

	10	
	GCACGTGGAAGCTTGATA	3' primer
	: : : : : : : : : : : :	
5'	GCACGTGGAAGCTTGATA	3' HNF1α cDNA
	3170                      3180	

100% identity in an 18 nucleotide overlap; score = 72.

Hybridisation of the HNF1α 3' primer to the non-coding strand of the HNF1α cDNA:

	30	
	CCACGTGC	HNF1α 3' primer
	: : : : : :	
5'	CCACGTGC	3' HNF1α cDNA
	2541	

100% identity in an 8 nucleotide overlap; score = 32.