MUTATION SCREENING OF CANDIDATE GENES AND THE DEVELOPMENT OF POLYMORPHIC MARKERS RESIDING ON CHROMOSOME 19q13.3, THE PROGRESSIVE FAMILIAL HEART BLOCK I GENE SEARCH AREA

Zola Makubalo

Thesis presented in partial fulfilment of the requirements for the degree Master of Science in Medical Sciences at the Faculty of Medicine, University of Stellenbosch

Promoter: Dr Valerie A. Corfield

January 2000

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and has not previously in its entirety or in part been submitted at any university for a degree.

Signature **Date**

SUMMARY

Progressive familial heart block type I (PFHBI) is a cardiac ventricular conduction disorder of unknown cause associated with risk of sudden death, which has been described in several South African families. Clinically, PFHBI is characterised by right bundle branch block on ECG, which may progress to complete heart block, necessitating pacemaker implantation. The disease shows an autosomal dominant pattern of inheritance with evidence of genetic anticipation. Using genetic linkage analysis, the PFHBI-causative gene was mapped to a 10 centimorgan (cM) gene-rich area of chromosome (C) 19q13.3, which has, subsequently, been reduced to 7cM by fine mapping with polymorphic dinucleotide $(CA)_n$ short tandem repeat (STR) markers. Several attractive candidate genes, including muscle glycogen synthase (*GSY1*) and histidine-rich calcium binding protein (*HRC*), lie within this region.

The aim of the present study was two-fold: 1) to identify and characterise tetranucleotide $(AAAT)_n$ STRs within the PFHBI critical region that could be developed as polymorphic markers for use in genetic fine mapping and 2) to screen selected regions of *GSY1* and *HRC*, positional candidate genes, for the presence of PFHBI-causing mutation(s).

Cosmids harbouring C19q13.3 insert DNA were screened for the presence of $(AAAT)_n$ STRs by dot blot and Southern blot hybridisation using a radiolabelled $(AAAT)_{10}$ oligonucleotide probe. To characterise the harboured $(AAAT)_n$ STRs, the positively hybridising fragments identified by Southern blot were sub-cloned, sequenced and primers designed from the unique repeat-flanking sequences. These primers were used to genotype the $(AAAT)_n$ repeat locus to assess its polymorphic nature in a panel of unrelated individuals. Alternatively, vectorette PCR, a rapid method of identifying repeat sequences and obtaining the flanking sequences in large inserts, was employed to develop polymorphic markers from the positively hybridising clones. Selected exons of *GSY1* and *HRC* were screened for the presence of potentially disease-causing mutations by PCR-SSCP analysis and direct sequencing, respectively, in PFHBI-affected and unaffected family members.

Of the available cosmid clones that gave strong signals on dot blot and Southern blot hybridisation, three, 29395, 24493 and 20381, were located within the critical PFHBI area and were used for marker development. An interrupted $(AAAT)_n$ repeat motif (n less than 5) was identified in cosmid 29395, however, the repeat locus was not polymorphic in the tested population. No $(AAAT)_n$ motif, single or repeated was observed in the partial sequence of the

sub-cloned fragment of cosmid 24493. Using vectorette PCR, no repeated $(AAAT)_n$ motif was identified on sequencing the generated products in either cosmid 24493 or 20381. However, diffuse single AAAT motifs were detected in both cosmids. Exons 4, 5, 11, 12 and 16 of *GSY1*, containing domains that are conserved across species, and the conserved C-terminus-encoding exons 2-6 of *HRC* were selected for screening for potential PFHBI-causing mutation(s). However, no sequence variations were detected.

The interrupted $(AAAT)_n$ repeat identified in cosmid 29395 was not polymorphic, which confirmed reports that complex repeats, especially those containing AAAT motifs of less than 6 repeats, are not polymorphic. One possible explanation for the absence of a repeated AAAT motif in cosmids 24493 and 20381, which both gave positive hybridisation signals, is that the low annealing temperature of the A/T-rich repeat-anchored primers used in vectorette PCR may have resulted in transient annealing to the diffuse single AAAT motifs detected on sequencing. The screened regions of candidate genes *GSY1* and *HRC* were excluded from carrying the disease-causing mutation(s).

The availability of new sequence data generated by the Human Genome Project will influence future strategies to identify the PFHBI gene. Electronic searches will allow identification of STR sequences for development of polymorphic markers and gene annotation will allow selection of new candidate genes for mutation screening.

OPSOMMING

Progressiewe familiële hartblok tipe I (PFHBI) is 'n outosomaal dominante siekte van die kardiale geleidingsisteem. PFHBI, waarvan die oorsaak onbekend is, kom voor in verskeie Suid-Afrikaanse families en kan lei tot voortydige dood. PFHBI word op elektrokardiogram (EKG) gedefiniëer deur regter bondeltak blok. Hierdie toestand kan lei tot volledige hartblok en die gevolglike implantering van 'n pasaangeër. PFHBI word outosomaal dominant oorgeërf en daar is bewyse van genetiese afwagting. Genetiese koppelingsanalises het die PFHBI geen gelokaliseer tot 'n 10 centimorgan (cM) area op chromosoom 19q13.3. Hierdie area is verkort tot 7 cM deur kaartering met polimorfiese dinukleotied $(CA)_n$ herhalings (STR) merkers. Verskeie kandidaat gene, onder andere spier glikogeen sintase (*GSYI*) en histidien-ryke Kalsium bindingsproteïen (*HRC*) word binne die 7 cM area gevind.

Die doel van hierdie studie was twee-voudig: 1) die identifisering van tetranukleotied $(AAAT)_n$ STRs in die PFHBI kritiese area wat ontwikkel kan word as polimorfiese merkers vir gebruik in genetiese kaartering en 2) die identifisering van mutasies wat lei tot PFHBI in die *GSYI* en *HRC* areas.

Die teenwoordigheid van $(AAAT)_n$ STRs in kosmied klone bevattende die c19q13.3 area, is aangedui deur Southern klad en Dot klad hibridisasie. 'n Radio-aktief gemerkte $(AAAT)_{10}$ oligonukleotied pylfragment is in die studies gebruik. Die positief gehibridiseerde fragmente is gesubkloneer, DNA volgorde bepaling is gedoen en primers is ontwerp vanaf die unieke herhalings volgordes. Hierdie primers is gebruik om die $(AAAT)_n$ herhalings lokus te genotipeer en te gebruik om die polimorfismes in onverwante individue te bepaal. Vectorette PCR is ook gebruik om polimorfiese merkers vanaf die positief gehibridiseerde klone te ontwerp. PCR-SSCP analises en DNA volgorde bepaling is gebruik om potensiële PFHBI veroorsakende mutasies te identifiseer in eksons van *GSYI* en *HRC* in PFHBI-geaffekteerde en nie-geaffekteerde familie lede.

Die kosmied klone, 29395, 24493 en 20381, wat sterk seine in die DNA klad tegnieke getoon het, is gelokaliseer binne die kritiese PFHBI area en is gebruik vir die ontwikkeling van genetiese merkers. 'n $(AAAT)_n$ herhalingsmotief (n is kleiner as 5) is geïdentifiseer in kosmied 29395 maar hierdie lokus was nie polimorfies in die populasie wat ondersoek is nie. Geen $(AAAT)_n$ motief, enkel of herhalend, is geïdentifiseer in kosmied 24493 nie. Vectorette PCR het geen herhalende $(AAAT)_n$ motief geïdentifiseer in kosmiede 24493 of 220381 nie.

Enkel AAAT motiewe is egter geïdentifiseer in beide kosmiede. Eksons 4, 5, 11, 12 en 16 van *GSYI* bevattende spesie gekonserveerde domeine, en die gekonserveerde C-terminus-enkoderende eksons 2-6 van *HRC* is geïdentifiseer as mutasies wat kan lei to PFHBI. Geen DNA volgorde variasies is waargeneem nie.

Die onderbroke $(AAAT)_n$ herhaling wat geïdentifiseer is in kosmied 29395 was nie polimorfies nie. Hierdie bevinding bevestig studies wat aandui dat komplekse herhalings, veral die bevattende minder as ses AAAT herhalings motiewe, nie polimorfies is nie. 'n Verduideliking vir die afwesigheid van 'n herhalende AAAT motief in kosmiede 24493 en 20381, wat albei positiewe hibridisasie seine getoon het, is dat die lae aanhegtings temperatuur van die A/T ryke primers wat gebruik is in vectorette PCR kon gelei het tot aanhegting van die primers aan die enkel AAAT motiewe wat geïdentifiseer is deur DNA volgorde bepaling. Die areas van die kandidaat gene *GSYI* en *HRC* is uitgesluit vir die moontlike voorkoms van PFHBI veroorsakende mutasies.

DNA volgorde data vanaf die Menslike Genoom Projek sal toekomstige studies om die PFHBI geen te identifiseer positief beïnvloed. Identifisering van STR volgordes vir die ontwikkeling van polimorfiese merkers en geen annotering vir die seleksie van nuwe kandidaat gene word vergemaklik deur die gebruik van elektroniese hulpmiddels.

INDEX

LIST OF ABBREVIATIONS

GENERAL ABBREVIATIONS

α	alpha
λ	lambda
°C	degrees Celsius
µCi	microCurie
µg	microgram
µl	microlitre
µM	micromolar
$AgNO_3$	silver nitrate
Amp	ampicillin
Aps	ammoniun persulphate
ASD	atrial septal defect
ATP	adenosine triphosphate
A-V	atrioventricular
BAC	bacterial artificial chromosome
BB	bundle branch
BBB	bundle branch block
bp	base pair
Ca^{2+}	calcium
$CaCl_2$	calcium chloride
CHB	complete heart block
CHD	congenital heart disease
cM	centiMorgan
dATP	deoxy-adenosine triphosphate
dCTP	deoxy-cytosine triphosphate
ddATP	dideoxy-adenosine triphosphate
ddCTP	dideoxy-cytosine triphosphate
ddGTP	dideoxy-guanosine triphosphate
ddH_2O	double distilled water
ddTTP	dideoxy-thymidine triphosphate
dGTP	deoxy-guanosine triphosphate
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid

dTTP	deoxy-thymidine triphosphate
ECG	electrocardiogram
EDTA	disodium ethylenediaminetetraacetate dihydrate
EtBr	ethidium bromide
K^+	potassium
Kn	kanamycin
LB	Luria-Bertani medium
LBB	left bundle branch
LBBB	left bundle branch block
M	molar
MIM	Mendelian inheritance in man
$MgCl_2$	magnesium chloride
min	minute
ml	millilitre
mM	millimolar
n	number
Na^+	sodium
NaOH	sodium hydroxide
ng	nanogram
OD	optical density
OMIM	online Mendelian inheritance in man
P1	bacteriophage P1
PAC	P1 artificial chromosome
PCR	polymerase chain reaction
PCR-SSCP	polymerase chain reaction-based single strand conformation polymorphism
pmol	pico mole
RBB	right bundle branch
RBBB	right bundle branch block
RE	restriction enzyme
RFLP	restriction fragment length polymorphism
rpm	revolutions per minute
S-A	sinoatrial
SDS	sodium dodecyl sulphate
SSC	sodium saline citrate solution
STR	short tandem repeat

TBE	tris, boric acid and EDTA buffer
TCA	trichloroacetic acid
TE	tris, EDTA buffer
Tc	tetracycline
U	unit
UTR	untranslated region
UV	ultra violet
VNTR	variable number of tandem repeats
www	world wide website
YAC	yeast artificial chromosome

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

- I would first like to thank Dr Valerie Corfield, my supervisor and mentor, for a promise fulfilled. Yes I did enjoy the project. I probably won't see another mentor like you, the experience has been educational and fun, my sincerest appreciation for believing in me. Thank you for the scientific input, linguistic and moral support in the ultimate writing of this thesis.

I would also like to thank the following people and institutions (in random order) for their contributions.

- To Prof. Paul Brink, without whom there would be no PFHBI to investigate, for his valuable intellectual input as well as the clinical examinations of all of patients and collecting of blood and his unfailing enthusiasm where the project was concerned and willingness to help.
- To Dr Hanlie Moolman-Smook, you are a brave woman for exposing your brain, for me to pick at. You've been an invaluable part of my development, my deepest gratitude.
- To Drs Rob Warren and Renate Hillerman for willingly sharing their expertise, it could not have come at a better time.
- To all the members of lab F445, thank you for a fun working environment and the occasional informative discussions.
- To my friends, thanks for being there, those who stuck-it-out and those who fell off along the way, I'm sure it wasn't easy.
- To Carol Lochner, the 'emergency kit'.
- To the US/MRC Centre for Molecular and Cellular Biology, where all the experimental work was performed, as well as the Departments of Internal Medicine and Cardiology of the University of Stellenbosch Medical School and Tygerberg Hospital, where much of the clinical investigations took place, whose financial support and existing infrastructure made the work possible.

- Finally, special thanks to my family for always believing in me and for your continued support (moral and financial) and encouragement. **BHELE!**

CHAPTER 1
INTRODUCTION

1.1 THE HEART

The true function of the heart was unknown before the 17th century. When Harvey (1578-1657) wrote "the blood is driven round a circuit with an increasing sort of movement; this is an activity or function of the heart, which it carries by virtue of its own pulsation", he not only was describing the circulation, but also drew attention to the heart's ability to provide the mechanical force necessary to sustain the circulatory flow (Opie, 1991). The late 1800s marked the beginning of the evolution of knowledge of the anatomy of the heart. It was during this period that Purkinje described two morphologically distinct cell populations within the heart, the contractile and the conducting cells (Davies, 1971). In spite of the controversies surrounding cardiology at that stage, further work done earlier in the twentieth century laid the basis for current knowledge of the anatomy and physiology of the heart (as reviewed by Davies, 1971 and Opie, 1991). The human heart, in its function as a mechanical blood pump, is driven by two major distinguishable components, namely, the cardiac conduction system and the working myocardium. The cardiac conduction system is primarily responsible for the generation of the depolarising electrical impulse and its propagation in a sequential manner to the rest of the heart, which is necessary for the synchronised contraction of the working myocardium, forming the walls of the heart's chambers (Hurst, 1978).

The maxim "you don't know what you've got until it's gone" is very much applicable in human medicine, when a disease manifests. Thus, with better understanding of the basic anatomy and physiology of the heart, in subsequent years, the focus logically shifted to heart defects. Although many human cardiac malformations are well characterised anatomically, there is still little understanding of their genetic basis (Olson and Srivastava, 1996), and this is especially true of the cardiac conduction system, where evolution of knowledge of cell biology has lagged behind compared with that of the working myocardium. This is mostly attributed to the difficulty in accessing the conduction tissue, either by biopsy or autopsy (Gorza *et al.*, 1994). In addition, the human cardiac conduction system is less discernible than that of some of its mammalian counterparts (for example, the pig, cow and dog) (Davies, 1971). However, the challenge still remains to understand the cell biology that drives this complex system, and, subsequently, to understand cardiac conduction system defects, some of which if left untreated by implantation of a pacemaker, contribute to the already high cardiac disease-related mortality rate. In the following sections a review of cardiac conduction system and the hereditary defects thereof will be presented, focussing on the progressive familial

heart block type I (PFHBI) a ventricular conduction system disorder, the subject of the present study.

1.2 THE CARDIAC CONDUCTION SYSTEM

1.2.1 Anatomy of the Cardiac Conduction System

The most widely accepted concept to date is that the cardiac conduction system is specialised cardiac muscle organised into nodes, that generate and conduct electrical impulses relatively slowly and the rapidly conducting ventricular conduction fibres (Figure 1.1) (Davies *et al.*, 1971). The electrical impulse is generated at the sinoatrial (S-A) node with the normal heart rate of 70 beat/min, which may be modified to beat faster or slower by the sympathetic or parasympathetic nervous system, respectively (Opie, 1991). The S-A node is therefore, referred to as the pacemaker determining the heart rate. The generated impulse is depolarising to the electrically excitable cardiomyocytes, thus it generates myocyte action potential, which effects changes that result in the contraction of the working myocardium (see section electrophysiology). Propagation of the depolarising impulse from the S-A node to the atrioventricular (A-V) node is via the normal atrial myocardium (Davies, 1971, Opie, 1991). Despite a number of debates about the nature of the atrial conduction pathways, no "specialised" conduction tracts have been shown (Davies, 1971; Anderson et al., 1981; Opie, 1991). At the A-V node, the impulse is temporarily delayed, thus allowing sufficient time for the atria to contract and empty the blood contained within them and for the ventricles to fill before the impulse is transmitted to the ventricular conduction system.

The fast-conducting ventricular conduction system comprises the His (A-V) bundle, which bifurcates into the right bundle branch (RBB) and left bundle branch (LBB), terminating as a network of Purkinje fibres at the apex, penetrating the ventricles (Figure 1.1) (Davies, 1971). Following the delay at the A-V node, the depolarising impulse is rapidly relayed to the His-bundle and is distributed to the ventricles by the right and left bundle branches. Consequently, the ventricles contract forcing the blood out either to the lungs (deoxygenated) or through the aorta to the rest of the body (oxygenated blood). This rapid transmission of the depolarising impulse from the A-V region to the ventricles ensures the synchronous contraction of the heart chambers, thus preventing retrograde flow of blood.

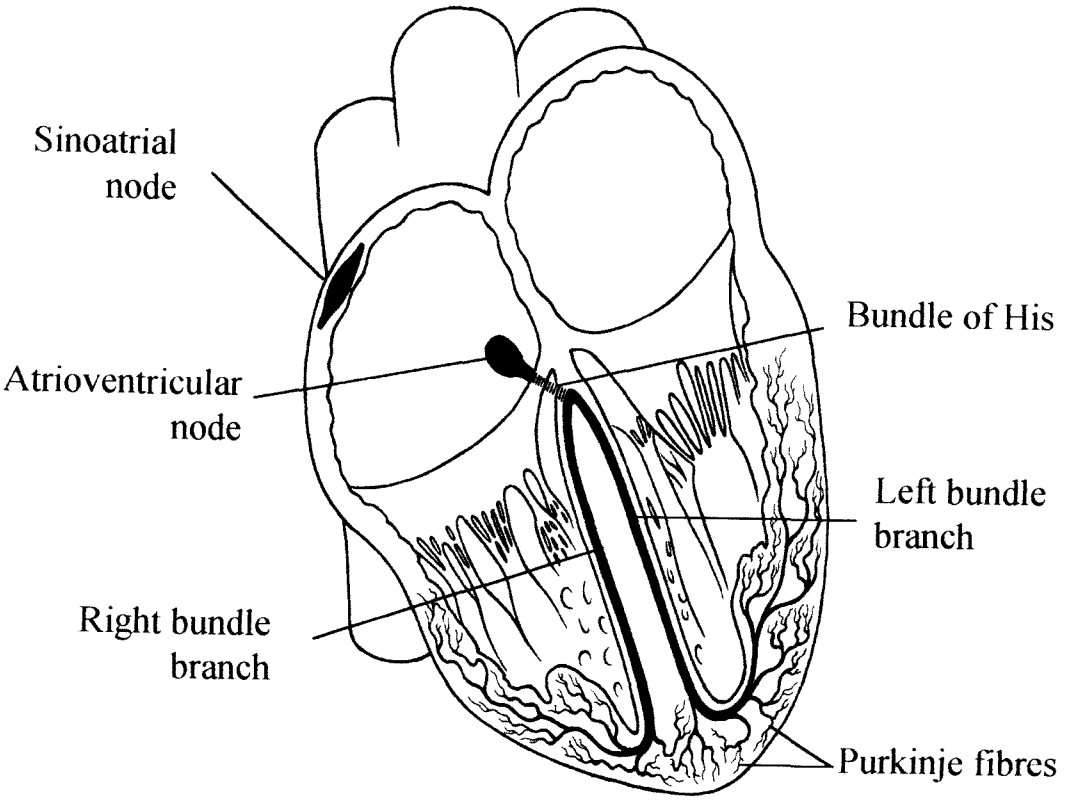


Figure 1.1: The cardiac conduction system (The Sourcebook of Medical Illustration, 1989).

1.2.2 Morphology of Cardiac Conduction Cells

The cardiac conduction system is referred to as specialised cardiac muscle because it consists of a group of cells that are histologically distinct from working myocardium (Opie, 1991; Zipes, 1997). The contractile cells (cardiomyocytes) in the heart are morphologically similar, comprising myofibrils (the contractile units), nuclei, mitochondria, sarcoplasmic reticulum and the sarcolemma (cell membrane), except for slight variation in size between the atrial (20μm in length) and ventricular (50-100μm in length) myocytes (Opie, 1991). The cardiac conduction system myocytes, in contrast, consists mainly of a heterogeneous group of cells, namely, P (pacemaker)-cells, transitional cells and Purkinje cells, in addition to scantily distributed contractile cells (James, 1970; Davies, 1971; Opie, 1991). These cells are morphologically distinct both from each other and from those of the working myocardium and are distributed within the conduction system components according to their specialisation.

Cell-to-cell communication is a characteristic feature often found in multicellular organisms, which is essential for synchronised communication and behaviour of cells (Paul, 1995). Within the context of the heat tissue, particularly the conduction system, this synchronised behaviour of cells is critical for the orderly and co-ordinated propagation of the depolarising impulse orchestrating heart contraction (Rossant, 1996). Thus, in addition to the morphologically distinct group of cells, the conduction system myocytes are resistively coupled by means of gap junctions (Sommer and Johnson, 1970, Spray *et al.*, 1994), membrane proteins that form boundaries between adjacent cell, thus ensuring efficient cell-cell communication and optimum conduction of the depolarising impulse. In the following sections the morphology of the cells within each of the major components of the conduction system will be discussed in relation to their role in impulse conduction.

The nodal cells

The pacemaker or P-cells and transitional cells are the two groups of cells found in the S-A and A-V nodes. The P-cells, thought to be the source of impulse formation, are small (smaller than atrial cells) ovoid cells arranged in grape-like clusters, located centrally within the nodes. Unlike the normal cardiomyocytes, P-cells contain very few myofibrils, mitochondria or sarcoplasmic reticulum and interconnection between adjacent cells is by simple apposition of plasma membranes (James, 1970; Opie, 1991; Zipes1997). The P-cells can generate electrical impulses without any outside stimulation (James, 1970).

Transitional cells, on the other hand, are thin elongated cells (intermediate between sinus nodal and atrial cells) that are longitudinally oriented at the periphery of the nodes and provide the only functional pathway for transmission of the impulse to the atrial myocardium. The connection between the P- and transitional cells is by simple apposition. In contrast, intra-communication in adjacent transitional cells and inter-communication with atrial myocardium is by intercalated discs. Although intercalated discs do not facilitate exchange of molecules from the cytoplasms of adjacent cells they, nevertheless, form good conduits (Campbell, 1990). Consequently, transitional cells have better cell-to-cell communication than the P-cells.

Although both the S-A and A-V nodes have pacemaking abilities, the S-A node dominates due to a fast pacemaking rate, postulated to be enhanced by the:

- Sinus nodal artery, which synchronises the otherwise relatively random pacemaking cellular activity and
- The collagen framework around the node, although the mechanism is not well understood.
- Sinus node autonomic innervation, which is the link between the normal pacemaking of the heart and extracardiac regulatory centres such as the brain. (James, 1970).

Transitional and P-cells form the major group of cells found in the nodes. The small size of these cells and the weak intercellular connections correlate to their relatively slow impulse conduction rates (5cm/sec) (Opie, 1991).

The ventricular conduction fibres

Purkinje cells are large pale elongated cells (150-200µm in length) found mostly in the ventricular conduction system (Davies, 1971; Opie, 1991; Zipes, 1997). Although these cells have been compared to the slow conducting embryonic heart cells because of their appearance, they have a conduction velocity of 200cm/sec (Opie, 1991). The characteristic feature of Purkinje cells responsible for the increased velocity of conduction is that, in addition to their size, they are arranged end-to-end longitudinally and are also interconnected by the well developed gap junctions, which are membrane channel proteins facilitating the exchange of cytoplasmic material between adjacent cells (James, 1970, Opie, 1991 and Spray *et al.*, 1994). In contrast, although myocardial cells are arranged longitudinally, the gap junctions connecting adjacent cells are located all around the cells, thus the depolarising impulse is not unidirectional but distributed in all directions to neighbouring cells (Opie, 1991). The distribution and organisation of all three types of cells within the conduction

system thus correlates with the role of each compartment in propagating the impulse. The slow-conducting nodes and the fast-conducting His-Purkinje system ensure orderly and synchronous activation of the heart.

1.2.3 Electrophysiology

Depolarisation

In the resting state, myocardial cells are polarised. The electrical impulse transmitted by the conduction system depolarises the cell membrane, activating the membrane bound voltage-dependent ion channels that facilitate sodium (Na) and calcium (Ca) ion transition into the cell (Davies *et al.*, 1971). The inward flux of the Na^+ and Ca^{2+} into the cytosol of cell generates action potential, which results in the contraction of the myocyte (Figure 1.2). The activation of the Na^+ channels is much faster than that of the Ca^{2+} channels (Katz, 1993). Consequently, the conduction of the electrical impulse in the rapidly conducting ventricular conduction system is dependent on the Na^+ channel-generated action potential, with the Ca^{2+} channel generating secondary inward currents (Opie, 1991; Katz, 1993). In contrast, the slow conducting nodes depend mostly on the slow calcium ion channel-generated action potential. In the A-V node the conduction delay is largely due to the slow conduction rate mediated by the Ca^{2+} channels. The same depolarising stimulus responsible for the activation of the ion channels effects their inactivation, thus closing the Na^+ and Ca^{2+} channels.

Repolarisation

The return of the cells to their resting state is effected by outward potassium (K) ion fluxes, through K ion channels. The two K^+ potassium ion currents involved for this process are;- the transient outward current (I_{to}), which is responsible for the early repolarisation and the delayed rectifier (I_K), which is responsible for the delayed repolarisation and also for maintaining the repolarisation (Katz, 1993; Wetzel and Klitzner, 1996).

The electrocardiogram

The electrocardiogram (ECG) is a non-invasive technique that records the heart's cumulative electrical activity, thereby providing valuable information about the heart's normal or abnormal function (Dubin, 1989). The ECG records the effect of the depolarising stimulus on the myocardium (i.e., sequential contraction of the atria and then the ventricles) as deflections,

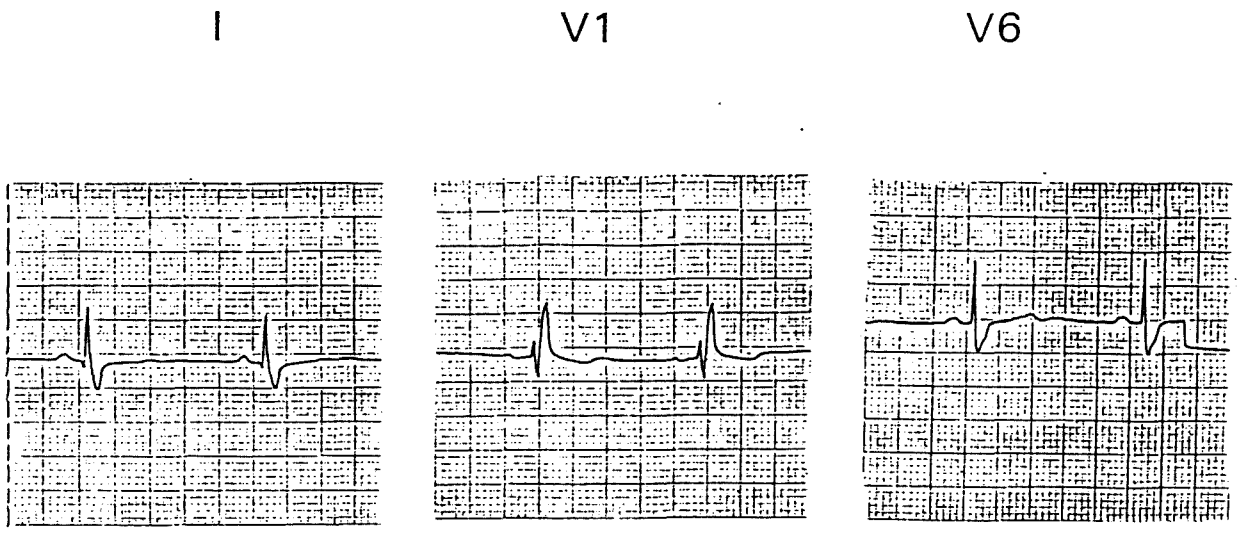


Figure 1.2: Patterns of impulse conduction from the sinoatrial node to the ventricular conduction system (Opie, 1991).
Schematic representation of the components of the cardiac conduction system indicating the dominant depolarising ionic currents in the slow conducting nodes and in the fast-conducting ventricular conduction system (left). Action potential generated by the different ionic fluxes in the conduction system and the ventricles (right).

namely, the P wave and QRS complexes. The return to the resting state, repolarisation, is represented by the ST-segment and the T wave (Figure 1.2).

In a normal heart, the P wave represents the contraction of both atria in response to the depolarising stimulus initiated at the S-A node. The P wave is followed by a flat baseline, representing the delay of the electrical impulse at the A-V node. Contraction of both ventricles is represented as a QRS complex. Following the QRS complex is a gently up-sloping ST segment, which precedes a T wave depicting ventricular re-polarisation. The ventricular QRS complex masks the T wave for atrial repolarisation. One cardiac cycle is represented by the P wave, QRS complex and a T wave. The ECG is, therefore, a great diagnostic tool, portraying the functional differences between a normal and an abnormal heart. In the case of asymptomatic conduction defects, e.g., isolated cardiac conduction defects, it is the only reliable means of detecting disease.

1.2.4 Metabolism

In the adult heart, fatty acids are the preferred choice of substrate for energy production in the form of adenosine triphosphate (ATP). Furthermore, fatty acid oxidation, an aerobic process, reportedly accounts for more than 60% of the ATP energy required for contraction (King and Opie, 1998). Glycolysis, the breakdown of glucose, is the other source of ATP in cells, and, in the heart, when present in high concentrations, fatty acids limit glucose entry, which is then directed to the liver for storage as glycogen (Sack and Kelly, 1998). The only time the normal mammalian heart uses glucose is during embryonic and foetal development, which soon after birth switches to fatty acids utilisation (these observation are expected as the adult heart has an increased energy requirement compared to the foetal heart) (Sack *et al*, 1996, Sack and Kelly, 1998). However, during stress, e.g., myocardial ischaemia and cardiac hypertrophy, fatty acid oxidation becomes very costly as it utilises the already limited oxygen and, the heart reverts back to the foetal program, i.e., using glucose to generate ATP energy.

An interesting finding at autopsy in a heart that had suffered a myocardial infarction, which resulted in the myocardial cell damage, was that the conduction system remained unharmed (Thornell and Sjostrom, 1975). Thus, based on the premise that the cardiac conduction system has an abundance of glycogen (James, 1970; Davies, 1971), this observation has led to the hypothesis that glycogen confers protection to the conduction system during ischaemia. Perfusion experiments performed over the years have shown that in the failing heart, glycogen

is the only source of glycolytic flux and confers protection to the failing heart (as reviewed by Opie 1991, King and Opie, 1998). However, the argument is that this protective function is, unfortunately, short lived as the accumulation of lactic acid and hydrogen ions (H^+), the by-products of anaerobic glycolysis, results in severe damage of the tissue (King and Opie, 1998). Furthermore, these authors argue that, the extent of protection during heart failure is directly proportional to the proximity of the tissue to the glycogen source, which would explain why, despite damage to the rest of the surrounding myocardial tissue, the conduction system remains unscathed (King and Opie, 1998).

1.3 IMPULSE CONDUCTION DEFECTS

Disruptions in the spread of the electrical impulse in any one of the components of the cardiac conduction system may result in heart rhythm disturbances (arrhythmias), most often leading to a slow heart rate (bradyarrhythmia). These arrythmias may progress to heart block, which may be fatal if not treated by timely implantation of an artificial pacemaker (Brink and Torrington, 1977; Alpert *et al.*, 1984). The variety of known defects range from sinus bradycardia, A-V block and defects of the ventricular conduction system, presenting as either, RBB block (RBBB) or LBB block (LBBB), in the absence of other cardiac diseases (isolated cardiac conduction disease) or associated with other structural heart diseases (Fahy *et al.*, 1996).

Cardiac conduction defects, particularly ventricular conduction defects, collectively referred to as bundle branch block (BBB), observed in the general population, were often interpreted as indicating serious heart disease, for example, myocardial ischaemia, myocardial infarction, cardiomyopathy, hypertension or congenital defects (DeForest, 1955; Alpert *et al.*, 1984). In the absence of other structural heart disease, the aetiology of BBB was often described as an age-related degeneration of the ventricular conduction system, as it was mostly seen in the older age group. (Lenègre, 1964). In support of this idea, Ostrander (1964) suggested that RBBB and LBBB observed in an epidemiological study of the Tecumseh population of Michigan, USA, who ranged between 20yrs and 80yrs of age, were due to acquired lesions that progressed with age. In most cases, the presence of partial (incomplete) BBB reportedly antedates progression to complete heart block (CHB) (Sarachek and Leonard, 1972; Schaal *et al.*, 1973), nevertheless, the prognosis for RBBB is better than for LBBB (Ostrander, 1964, Fahy *et al.*, 1996). Of the conduction defects seen in the young, Ostrander (1964) suggested that these were possibly congenital anomalies, which were frequently a benign finding.

Complete heart block is an event rarely seen in young adults. However, a number of families have been described where the tendency is apparently inherited and affects the young. Consequently, the familial occurrence of conduction defects has since become recognised as showing variations in the mode of inheritance (autosomal dominant or recessive), age at onset (congenital or adult onset), localisation of the defect and prognosis (Gazes *et al.,* 1965; Simonsen and Madsen, 1970; Steenkamp, 1972; Sarachek and Leonard, 1972; Husson *et al.,* 1973; Schaal *et al.,* 1973; Vallianos and Sideris, 1974; Esscher *et al.,* 1975; Fairfax and Lambert, 1975 and James *et al.,* 1975, Greenspahn *et al.*, 1976; Stéphan, 1979). In the following section, an overview of familial ventricular conduction defects, manifesting as congenital or adult onset CHB, will be presented.

1.3.1 Familial congenital cardiac conduction defects

Building a perfect heart is hard and it is a process highly susceptible to genetic and environmental influences. Consequently, congenital heart anomalies are the most common form of human birth defects and remain the leading cause of death within the first year of life (Olson and Srivastava, 1996; Rossant, 1996; Srivastava, 1999 and Yamagishi *et al.*, 1999).
The strict definition of congenital diseases is "those defects seen at birth" (James *et al.*, 1975). In other instances, however, this term is loosely used to describe anomalies seen during the first year of life, as has been the case with some congenital familial cardiac conduction defects (Sarachek and Leonard, 1972; Husson *et al.*, 1973; Schaal *et al.*, 1973). To resolve these discrepancies, Yater's criteria, namely, the presence of a slow pulse during childhood and proof of heart block in the absence of any previous diagnosis associated with other cardiac defects, is often used for diagnosis (Sarachek and Leonard, 1972). In the sections that follow, a review of known familial congenital conduction system diseases will be presented. However, prior to that a brief review on the development of the normal embryonic and foetal heart will be presented to help understand how diseases might affect the developing heart.

Development of the heart

The primitive heart tube (embryonic heart) is the first organ to develop and function in the foetus (Rossant, 1996). The formation of the embryonic heart requires the commitment of precursor cells, from the different embryonic lineages, namely, mesoderm and neural crest, and their differentiation to the cardiac lineage followed by the co-ordinated assembly of the different cells (Srivastava *et al.*, 1995, Olson and Srivastava, 1996, Rossant, 1996). Subsequently, the mature four-chambered heart is formed when the primitive heart tube

undergoes looping, for right-left symmetry and the formation of the atria and ventricular cavities and septation required to separate the oxygenated from the de-oxygenated blood, postnatally (Srivastava *et al.*, 1995).

The embryonic heart has no clearly recognisable conduction system, but comprises alternating slow- and fast-conducting segments, and an inflow tract that generates impulse, which results in an ECG similar to that found in the fully developed heart (Davies, 1971; Moorman and Lamers, 1994). The developmental appearance of the fast-conducting ventricular conduction system is nevertheless, closely associated with ventricular septation, one of the late developmental stages (Moorman and Lamers, 1994; Schott *et al.*, 1998). One of the long-standing debates on the development of the cardiac conduction system was whether the bundle of His is an outgrowth from the A-V node. Evidence from histopathologic studies suggested that these two components were derived from different cell lineages (James *et al.*, 1975). The separate appearance of the ventricular conduction system to the A-V node in the developing heart is evidence supporting the idea that the components of the conduction system are derived from different cell lineages and subsequently, assembled through a process of accretion (Moorman *et al.*, 1998; Gourdie *et al.*, 1999).

Cardiac morphogenesis is regulated by multiple signals, unfortunately, little is known of the underlying genetic pathways. However, with the growing interest in heart development and the identification of mutant gene of a few congenital heart defects (CHD), this lack of knowledge will soon be replaced by, at the least, clues to the regulatory pathways controlling embryonic heart development.

Generally, CHD represent defects in the development of specific regions of the heart, while the rest of the heart develops rather normally (Kennel *et al.*, 1981; Schott *et al.*, 1998; Srivastava, 1999) and are often associated with specific morphogenic defects in other organs (DeForest and Hanover, 1955, Olson and Srivatsava, 1996; Belmont, 1998). Likewise, familial congenital conduction system defects are often associated with cardiac or extracardiac manifestations (Husson *et al.*, 1973; James *et al.*, 1975; Van der Merwe *et al.*, 1991, James *et al*, 1994). One of the known causes of congenital defects, resulting in multi-organ system defects, is the arrest of development at early embryonic stages resulting from defects in the precursor cells prior to differentiation (Olson and Srivastava, 1996; Belmont, 1998). On the other hand, CHD are due to errors in the late developmental stages, which involve the moulding and shaping of the heart, e.g., septation (James, 1970; Olson and

Srivastava, 1996). Two familial CHD that involve multi-organ systems with conduction system defects as one of their characteristic features, the Holt-Oram symdrome and atrio-septal defects (ASD) will be discussed.

Holt-Oram syndrome- MIM 142900

The Holt-Oram syndrome (HOS) belongs to a class of combined cardiac and limb (extracardiac) deformities commonly referred to as 'heart-hand' syndromes and inherited as an autosomal dominant trait. These are genetically heterogeneous with variation in the severity of limb deformities, atrial septal defects and cardiac conduction disease (OMIM). The "heart-hand" syndrome type III (HOS III), for example, is a cardiac conduction disease accompanied by skeletal muscle malformations and is characterised by a range of conduction anomalies, namely, a slow heart rate, A-V block and sinus node dysfunction (Basson *et al.*, 1995). The cardiac and skeletal muscle cells are derived from the same somatic germ layer (mesoderm) but different precursors, thus, the resulting defects in cardiac and skeletal muscle are indicative of a developmental errors that occurred prior to commitment of the precursor cells to the specific lineages.

The HOS III-causative gene was identified as *TBX5*, a transcription factor-encoding gene (Li *et al.*, 1997). Transcription factors are regulatory proteins controlling the expression of other genes, including developmental genes. Thus, defects in the transcription factors regulating the commitment and differentiation signals of precursor cells will affect multiple of developmental pathways. The *TBX5* gene belongs to the T-box family of transcription factors essential for early development including specification of the mesoderm and regulation of heart and limb morphogenesis (reviewed by Smith, 1999).

Atrio-septal defects

Cardiac septation occurs in the later stages of development and involves the partitioning of the primordial single atrium and ventricle into the four chambers of the mature heart. However, following septation a hole is maintained between the atria to bypass the as yet non-functional lungs, until after birth. Ten to fifteen hours after birth the hole is functionally closed in normal and permanently closed 2-3 weeks later. This final moulding and shaping of the heart is an error prone process (James, 1970, Schott *et al.*, 1998) and results in atrio-ventricular defects (ASD) with progressive conduction system disease as a characteristic feature (Schott *et al.*, 1998).

Unlike HOS, ASD feature conduction anomalies associated with another cardiac manifestation. However, similar to HOS, mutations in a transcription factor, NKX2-5, were identified as the cause of autosomal dominant ASD with A-V conduction delays (A-V block) as the other characteristic feature (Schott *et al.*, 1998). The *Nkx2-5* gene was shown to be important in cardiac septation and in the development of the A-V node.

Hereditary developmental conduction system defects are not necessarily congenital, i.e., they need not cause disease early in life (Simonsen and Madsen, 1970; Steenkamp 1972). This statement is supported by the observation that individuals born with ASDs, that were subsequently surgically repaired, had conduction defects manifesting in adulthood (Schott *et al.*, 1998). Therefore, partial absence of conduction tissue need not present with symptoms until adult life, which was explained by Steenkamp (1972) as a result of progressive degeneration of the conduction system. In their study, Schott *et al.* (1998) showed that the transcription factor NKX2-5 is not only required during development but also for maintenance of the A-V node postnatally and into adult life.

Although the pathogenesis of the congenital and adult onset familial heart block may be different, the increasing evidence that true familial congenital heart block may not only manifest later in childhood, but even in adult life, suggests that, in some cases, the two forms of heart block may be the same disease (Brink, 1997). One theoretical explanation for such delayed appearance of heart block is gradually progressive narrowing disease of the A-V node artery, although it has not been shown that such a process has a familial incidence (James *et al.,* 1975). The cardiac conduction system might be congenitally abnormal but disease may only manifest as the ageing process occurs (Schaal *et al.,* 1973) and may also be influenced by environmental factors.

1.3.2 Familial bundle branch block – Adult onset

Phenotype

Adult onset BBB is proven when the initial manifestations occur in adulthood and conduction had previously been shown to be normal (Simonsen and Madsen, 1970; Kennel *et al.,* 1981). It may be considered familial when one or more first-degree relatives develop a similar condition in the absence of other recognised causes of acquired heart block (Kennel *et al.,* 1981). In the absence of other heart disease, conduction system diseases are normally

asymptomatic, with the ECG the only means of detecting any abnormalities. Defects in the ventricular conduction system are characterised by wide or narrow QRS complexes, depending on the site of the lesion (block), namely the RBB or LBB, respectively (concepts by Mahaim, as reviewed by Lenègre, 1964).

The disease is progressive in nature, as both RBBB and LBBB antedate CHB (Schaal *et al.*, 1973). However, timely implantation of an artificial pacemaker improves the prognosis (Brink and Torrington, 1977; Alpert and Flaker, 1984). For most affected individuals, the only symptom is when CHB supervenes with either syncope or Stoke's-Adam's seizures, which is associated with risk of sudden cardiac death. However, the aetiology of the underlying conduction defect remains unknown.

Classical Diagnosis

Bundle branch blocks are caused by partial blocks of the depolarising impulse in the RBB or LBB, causing a delay of the electrical impulse to that side (Dubin, 1989). Therefore, in BBB, the ventricles do not fire simultaneously, and produce two co-joined QRS complexes resulting in a "widened QRS" pattern on ECG. More specifically, RBBB is represented by R,R' in the V_1 and V_2 chest leads, with R' indicating the delayed ventricle (Figure 1.3). Similarly, observation of R,R' in the V_5 and V_6 chest leads is indicative of LBBB. Clinically, BBB is defined as a prolonged QRS complex ($\geq$ 0.12 seconds). However, QRS complexes of less than 0.12s with an R,R' pattern have been reported and these are referred to as incomplete BBB. Currently, the only available form of treatment for bundle branch defects is timely implantation of a pacemaker, which allows affected individuals to enjoy a near normal lifestyle (Brink and Torrington, 1977).

Pathology

The progressive nature of the adult type of BBB is more compatible with the gradual fibrosis, rather than with the complete absence, of the main stem of the conduction system, as suggested by autopsies of infants with familial congenital CHB (Vallianos and Sideris 1974).

Although a variety of general terms had previously been used in different earlier studies to describe conduction defects (Brink *et al.*, 1994), it was Brink and Torrington (1977) who described two familial conduction system defects in South Africa and gave them the

Figure 1.3: The ECG tracings showing a typical pattern of right bundle branch block in standard leads I, V1 and V6 (Brink, 1997).

The QRS complex is more than 012s wide with a delayed wave due to late right ventricular depolarisation as is evidenced by the terminal slurred R wave in standard lead V1 and a slurred S wave in standard leads V6 and 1.

appropriately descriptive names, progressive familial heart block, types I and II (PFHBI and PFHBII, respectively).

Progressive familial heart block type I – MIM 113900

Progressive familial heart block type I is a primary ventricular conduction system defect characterised by a broad QRS complex (Brink and Torrington, 1977), an indication of BB disorder. The disease was described as progressive in nature, as some of the individuals with RBBB later developed CHB. Transmission of the disease in families was autosomal dominant with three high-risk periods for onset of CHB, namely, at or soon after birth, during puberty and the early 20's and again towards middle age (Brink and Torrington, 1977). Furthermore, based on the observation that the mean age for pacemaker implantation decreased with three successive generations studied, Brink and Torrington (1977) suggested genetic anticipation to be involved in the disease pathology. Follow-up studies on the same families led to the extension of the family pedigrees, where to date there are three apparently unrelated pedigrees, designated pedigrees 1, 2 and 5 (Brink *et al.*, 1994) (Figure 1.4). The largest, pedigree 2, has been extended to an extensive nine generations all traced back to one ancestor, a Portuguese settler (Torrington *et al.*, 1986; Van der Merwe *et al.*, 1986, Van der Merwe *et al.*, 1988; Brink *et al.,* 1994). A decade after Brink and Torrington's initial report, while assessing the mean age for pacemaker implantation, Van der Merwe *et al.*, (1988) noted that, with successive generations, which comprised the last generation from the previous study, plus two new ones, it became much lower, confirming Brink and Torrington's observation.

Hereditary bundle branch defect – MIM 113900

Despite several reports of familial RBBB world-wide (Brink *et al.*, 1994), hereditary bundle branch defect (HBBD-113900), also referred to as inherited cardiac conduction disorder (ICCD) (De Meeus *et al.*, 1995), described by Stéphan in a Lebanese kindred, is the only other known conduction system disease with similar features to PFHBI (Stéphan, 1979). Like PFHBI, the disease is inherited as an autosomal dominant trait characterised by RBBB, which in most instances progressed to CHB with syncopal attacks. Age at onset varied between congenital and adult onset, whereas in PFHBI, there was an additional high-risk period, namely, during puberty and the early 20's (Brink and Torrington, 1977).

Progressive familial heart block type II – MIM 140400

Progressive familial heart block type II is also inherited in an autosomal dominant fashion. In contrast to PFHBI, PFHBII is characterised by sinus bradycardia with left posterior hemiblock, narrow QRS complex on ECG. Clinically, however, PFHBII, like PFHBI, presents with sycopal attacks and Stoke's-Adam's seizures when CHB supervenes (Brink and Torrington, 1977). Until recently, it has been proposed that PFHBI and PFHBII may be the same disease, possibly caused by different mutations within the same gene. However, linkage analysis studies performed recently, using chromosome 19q13.3 spanning microsatellite markers, have excluded the *PFHBI* locus as the causative gene for PFHBII (Fernandez, personal communication).

Although the clinical features of these bundle branch defects have been well characterised, the aetiology remains unknown, with pacemaker implantation the only form of treatment available, that sees the patients returning to near normalcy (Brink and Torrington, 1977; Brink, personal communication). The present study will focus on PFHBI and the application of the positional cloning approach in an effort to identify the PFHBI-causative gene.

1.4 PROGRESSIVE FAMILIAL HEART BLOCK I

1.4.1 Clinical History

Progressive familial heart block type-I is a primary disease of the ventricular conduction system with risk of untimely sudden death, that is inherited as an autosomal dominant trait in several South African Caucasian kindreds of Afrikaner descent (Brink and Torrington, 1977). To date, three apparently unrelated pedigrees, designated 1, 2 and 5, have been established and pedigree 2 has an extensive nine generations all traced back to one ancestor, a Portuguese settler who came to South Africa in 1696 (Figure 1.4). Within the Afrikaner sub-population, PFHBI has a prevalence of 1:500 and a penetrance value of 0.96 (Torrington *et al.*, 1986; Brink *et al.*, 1994).

1.4.2 Age at Onset

Three high-risk periods for onset of CHB in PFHBI have been reported, namely, 1) during the first year of life, 2) during puberty and the early 20s and 3) towards middle age (Brink and Torrington, 1977). These high-risk periods are an indication, and offer a warning, that in familial cases, an early normal ECG does not exclude the presence of an underlying

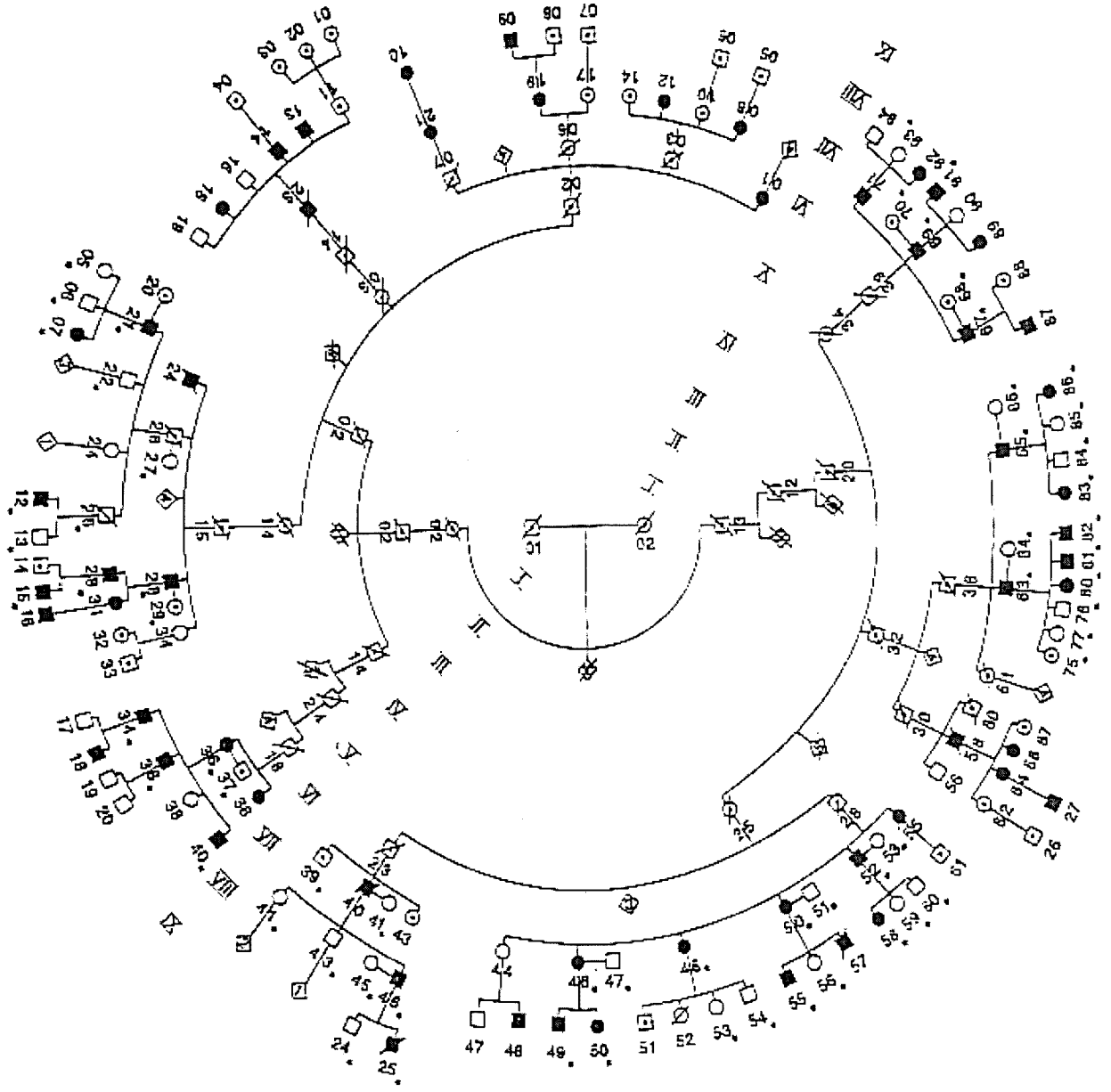
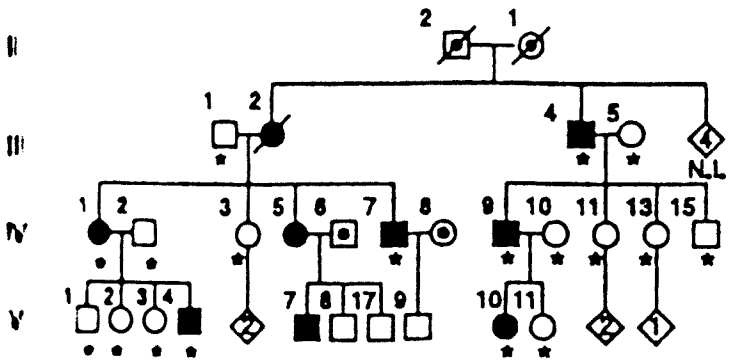
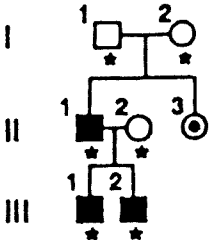


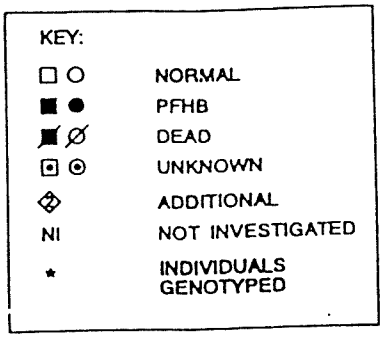
Figure 1.4 Family pedigrees 1, 2 and 5 segregating with PFHBI (Brink et al., 1995).



PEDIGREE 1



PEDIGREE 5



conduction tissue disorder, as it may develop later in life (Brink and Torrington 1977, Van der Merwe *et al*, 1991 and Brink, 1997).

1.4.3 Diagnosis

Clinically, the disease is characterised by broad QRS complexes (>0.12s) on ECG, with an R,R' in the V1 and V2 chest leads, indicating RBBB (Figure 1.3). Progression from milder to more severe forms of heart block, e.g., CHB is one of the detrimental features of PFHBI, with timely pacemaker implantation the only form of treatment currently available (Brink and Torrington, 1977; Van der Merwe *et al.*, 1988). The observation by Van der Merwe *et al.* (1988) that age of pacemaker implantation decreased with successive generations led the authors to propose that the severity of the disease had increased with successive generations, a phenomenon referred to as genetic anticipation (Howeler *et al.*, 1989).

1.4.4 Pathology of PFHBI

The only post-mortem tissue sample ever to be examined from the PFHBI families was from a 2 year-old patient who had a third-degree heart block at birth and had a pacemaker implanted. At 22 months, four months prior to his death, the patient's left ventricle began to dilate and progressed to severe dilatation. Autopsy findings showed fibrosis of the conduction tissue main bundle and proximal part of the bundle branches in the presence of dilated cardiomyopathy (Van der Merwe *et al.*, 1991), the latter, however, is a rare feature in PFHBI. Therefore, although the clinical profile of PFHBI has been well characterised, the aetiology remains unknown. Molecular biology-based approaches offer a strategy to help identify the disease gene with the hope of elucidating the underlying cause.

1.5 THE POSITIONAL CLONING APPROACH

Positional cloning (Collins, 1992) has revolutionised human genetics and medicine, as it allows mapping of the disease-causing gene to its correct chromosomal location, relying only on the position of the gene in the genome, without an absolute requirement of functional knowledge. To achieve this, a collection of pedigrees in which the disease gene is segregating is required and, to establish linkage, the families are genotyped using multiple polymorphic markers that span the genome, which often places the candidate locus within a region of 5-10cM (1cM is approximately equivalent to 1Mb) (Schuler *et al.,* 1996). Additional fine mapping using flanking genetic markers is then applied, approaching the gene from both directions, while reducing and redefining the target interval. However, fine mapping is

generally limited to genetic intervals of about 1cM, following which physical mapping strategies are used to clone the gene (Collins, 1992). Genes known to reside within the target interval are considered positional candidates and are therefore screened for possible disease-causing mutations (Strachan and Read, 1996). The development of the positional cloning approach into the most powerful gene cloning strategy stems partly from the advances made in terms of the genetic markers developed to date (discussed below). In the following sections, a brief history of the advances made with respect to the types of polymorphic markers used, which also highlights the robustness of the currently popular type of marker, the microsatellites, is presented.

1.5.1 Genetic Maps and Polymorphic Markers

Genetic markers and maps of human chromosomes provide the initial resources for localising Mendelian traits and disease susceptibility genes to their precise chromosomal regions by linkage analysis. Earlier application of genetic markers to generate linear genetic maps of chromosomes, and their use in tracing inheritance patterns, was confined to simpler organisms, such as *Drosophila melanogaster*, which only has four chromosomes, in which crosses could be manipulated in the laboratory, Mendel-style, unlike humans (Sturtevant and Morgan, 1913). Genetic linkage analysis involves comparing the inheritance of alleles at a specific marker with that of the disease, where co-inheritance of the disease gene and marker allele indicates that the two are physically close, hence linked on the specific chromosome (Terwilliger and Ott, 1994). Since the first application of linkage analysis in humans by Bell and Haldane (1937), other developments included the statistical approach to linkage analysis, the LOD score (Morton, 1955) and computer-based analysis using a limited number of blood group and serum protein markers (reviewed by the Murray *et al.*, 1994). Despite these successes, for many years the lack of informative and evenly spaced markers was limiting (Lander and Schork, 1994).

1.5.2 Advances in genetic markers

The advent of recombinant deoxyribonucleic acid (DNA) technology revolutionised medical genetics and medicine in two ways. Firstly, was the identification of DNA-based polymorphisms, referred to as restriction fragment length polymorphisms (RFLPs), which could be used as markers (Solomon and Bodmer, 1979, Botstein *et al.*, 1980). Secondly, subsequent cloning and characterisation of previously unknown genes underlying human inherited traits helped elucidate the molecular basis of disease mechanisms. The earlier

successful method, termed functional cloning, identified the disease-causing gene based on pre-existing knowledge about the biochemical function of the gene product. However, for the vast majority, approximately 4,500, of single gene (Mendelian) disorders, such detailed functional insight, unfortunately, does not exist, hence functional cloning becomes limiting.

The major breakthrough for human genetics, however, was the subsequent development of positional cloning (Collins, 1992), which led to the mapping of unknown genes for rare Mendelian disorders, regardless of knowledge of biochemical function. A vast source of markers became available with the realisation that many naturally occurring DNA repetitive sequences were polymorphic and thus could be used as markers. The first group of repetitive DNA sequences identified was the variable number of tandem repeat (VNTR) minisatellites (Jeffreys *et al.*, 1985; Nakamura *et al.*, 1987), which like the RFLPs, were detected by DNA hybridisation on Southern blots. Although, subsequently, less time consuming PCR-based methods of detection were developed, RFLPs and VNTRs still suffered other disadvantages. For example, RFLPs are bi-allelic, and thus not very informative and, although VNTRs are highly informative as markers, with heterozygosity indices of >97%, they are confined to the telomeric ends of chromosomes, which renders them limiting for the generation of high-density maps (Royle *et al.*, 1988). Subsequently, a more robust group of markers, the microsatellites, was identified (Litt and Luty, 1989; Weber and May, 1989). In the following sections, an overview of the currently used markers, microsatellites, focussing on their robustness and preferential use as markers in gene mapping is presented.

1.6 MICROSATELLITES

Microsatellites are short tandem repeats (STRs), 1-5bp in length, which are classified as mono-, di-, tri-, tetra- and pentanucleotide sequences. They are randomly distributed within the human genome, they are embedded in unique sequence and are polymorphic, as they show variation in the number of the repeated units at a specific locus among individuals within a population (Cooper and Clayton, 1988). Mono-, di- and tetranucleotide STR are limited to the non-coding regions of the genome, as the presence of such polymorphic sequences in coding regions would affect the coding sequence and hence the gene products. In contrast, trinucleotide STRs are found in both the coding and non-coding regions. Several neurodegenerative diseases are associated with the abnormal expansion of trinucleotides in coding regions illustrating the detriment of these polymorphic sequences in genes (section 1.8.5). The microsatellite STR sequences are stably inherited in families from one generation

to the next, which allows their use as genetic markers in tracing inheritance patterns in human pedigrees (Litt and Luty, 1989; Weber and May, 1989).

Although microsatellites are less informative as markers than VNTRs, with heterozygosity indices ranging between 34% and 90%, other compensatory features exist, e.g., their random distribution within the human genome. In addition, their robustness as genetic markers is based on the fact that these repeats are often embedded in unique sequence and are contained in 100bp-500bp fragments, in contrast to VNTRs, which are contained in ~1kb fragments. These properties make the microsatellites more amenable to PCR-based analysis and easier verification of genotyping data (Litt and Luty, 1989; Weber and May, 1989; Alford *et al.*, 1994).

1.6.1 Dinucleotide STRs

The dinucleotide STR of the form $(dC\text{-}dA)_n$.$(dG\text{-}dT)_n$, hereafter designated $(CA)_n$, with repeat number (n) ranging from 10-60, is the most abundant, with one STR locus approximately every 30kb, and is the most commonly used form of genetic marker (Litt and Luty, 1989; Weber and May, 1989). Due to the fact that the $(CG)_n$ dinucleotide STRs, with n >6, are rarely seen, a possible explanation was thought to be that they are selected against as long stretches of CG nucleotides present a deleterious structural conformation, Z-DNA (Hamada and Kakunaga, 1984). However, Tautz and Renz (1984) suggested that, *in vivo*, these sequences were not in Z-DNA form. The informativeness of the $(CA)_n$ STR is directly proportional to the repeat length (with heterozygosity indices varying between 34% and 74%). It is also positively influenced by the complexity of the repeat, with the degree of informativeness decreasing in the order of perfect, imperfect and compound repeats (Weber, 1990). Although the $(CA)_n$ STR has proven an outstanding choice of marker, it has a high frequency of strand slippage artifacts on PCR amplification (which were seen as shadow bands on electrophoretic separation), making it difficult to genotype (Edwards *et al.,* 1992b). This led to a search for STR genetic markers that, in addition to being polymorphic and randomly distributed within the genome, would be robust to genotyping. The trinucleotide and tetranucleotide STRs fit these criteria, and have heterozygosity indices of between 37% - 83%. Despite the fact that they are not as frequently distributed as the $(CA)_n$ STR, for fine mapping purposes, trinucleotide and tetranucleotide STRs serve as invaluable genetic markers complementing the $(CA)_n$ STR (Zuliani and Hobbs, 1990, Edwards *et al.,* 1992a, Sheffield *et*

al., 1995). The trinucleotide STRs are, however, infamous for their pathogenicity when abnormally expanded in some disease-causing genes.

1.6.2 Trinucleotide and Tetranucleotide STRs

The trinucleotide and tetranucleotide STRs (discussed below), like the $(CA)_n$, are highly polymorphic, and they are distributed randomly within the intronic regions of the human genome, on average every 300kb (Edwards *et al.,* 1989, Edwards *et al.,* 1992a). In contrast, a different class of trinucleotide STRs is distributed in both exonic and intronic regions and their polymorphic nature in coding regions may be detrimental, in which case they are referred to as pathogenic microsatellites (section 1.8.5). Together, the non-pathogenic trinucleotide and the tetranucleotide STRs occur with a frequency of one STR, of approximately seven repeat units in length, every 10kb (Edwards *et al.*, 1991). Like the $(CA)_n$ repeat, tri- and tetranucleotide STRs are generally stably inherited in families from one generation to the next (Chakraborty *et al.,* 1997). In contrast to $(CA)_n$, trinucleotide and tetranucleotide STRs can be PCR-amplified with increased fidelity, showing less strand-slippage mismatch, *in vitro* (Edwards *et al.,* 1992b). In the same study, Edwards *et al.* (1992b) suggested a positive correlation between the repeat unit length and the complexity of the repeat and the ease of genotyping (Sheffield *et al.,* 1995). For tetranucleotide STRs, repeat unit lengths of greater than six are considered informative (Sheffield *et al.,* 1995). Triplet repeats of eight or more in non-coding sequences are likely to be polymorphic (Gastier *et al.,* 1996). Though repeat size has been shown to be a strong indicator of polymorphic length variability for $(CA)_n$ STR (Weber, 1990), it is possible that the repeat sequences found in coding regions are under strong selective constraints to avoid either expansion or contraction (Reddy *et al.,* 1997).

There is a wide selection of commonly occurring tetranucleotide STR sequences. The Utah Marker Development Group (1996) performed studies to test, firstly, which of a selection of the tetranucleotide STRs, namely, AAAG, AGAT, AAGG, AAAT, GATA, GGAA, GAAT, GGAT and GTATT, were abundant and, secondly, which of those were ideal for development as markers i.e., gave clear results on genotyping. Of the STRs analysed, the most abundant were AAAG (38%), AGATT (29%), AAGG (10%) and AAAT (5%), the remainder were a combination of di- and trinucleotides (The Utah Marker Development Group, 1996). Furthermore, the AGAT, AAGG, AAAG and AAAT STRs gave the cleanest and most interpretable results.

1.6.3 Origin of Microsatellites

The hypothesis cited the most in the literature to explain the origin of microsatellites is the strand slippage mechanism during replication *in vivo*, whereby the initial expansion of short repeat sequences occurred relatively slowly on an evolutionary time-scale (Levinson and Gutman, 1987; Schlotterer and Tautz, 1992, Meisser *et al.,* 1996). Once the critical number of repeat units was reached, the mutation rate increased resulting in observed length polymorphism of microsatellites within populations.

The second proposed mechanism for the origin of a fraction of microsatellites is retrotransposition (Beckman and Weber, 1992). This theory is based on the observation that more than 80% of the A-rich STRs, namely, A, AN, AAN, AAAN (where N represents nucleotides T, G or C), are located immediately 3' to Alu-elements, which are primate specific sequences (Zuliani and Hobbs, 1990; Beckmann and Weber, 1992). It is thought that Alu-elements are retrotransposons, that is RNA transcripts that have been converted to complementary DNA (cDNA), which is highly mobile and can re-insert into different regions of the genome. As a consequence of the RNA origin of these Alu-elements, they contain a 3' poly (A) tail. The A-rich STRs are thought to be tandem degenerations of the poly (A) tail (Economou *et al.*, 1990, Zuliani and Hobbs, 1990). Owing to a high frequency of recombinational events associated with Alu-elements, sequences in close proximity to the Alu-elements may also be more susceptible to slipped-strand mispairing during replication, a widely accepted hypothesis for evolution of STRs (Levinson and Gutman, 1987; Meisser *et al.,* 1996). The finding of microsatellites with repeat motifs other than (A), 3' to the Alu elements, led to the conclusion that the Alu-related poly (A) sequences can evolve into more complex microsatellites (Beckmann and Weber, 1992).

These sequences are thus maintained because of the weak selective constraints associated with the non-coding regions of the genome (Cooper and Clayton 1988, Baron *et al.,* 1992, Charlesworth, 1994). The extent to which microsatellites are conserved across species is as yet unknown. Stallings *et al.* (1991) suggested that in closely related species (e.g., humans and primates) the $(GT)_n/(CA)_n$ repeat sequences are conserved, but not in evolutionary distant species (e.g., humans and rodents). If the microsatellite sequences are conserved, then what is their functional significance in eukaryotic genome?

1.6.4 Functions of Microsatellites

Although their function in the genome is unknown, the postulated significance of conserved STRs in intronic regions includes involvement in transcription regulation. For example, at least one protein is known to bind to an $(AAT)_n$ repeat motif, commonly found in introns (Sheffield *et al.,* 1994). Moreover, the observation that microsatellites are absent in bacteria may be an indication of their involvement in the packaging and condensing of DNA into eukaryotic chromosomes, particularly in replication. During the cell cycle, they may provide a DNA conformation that can be condensed and de-condensed during different stages of the cell cycle (Vogt, 1990).

Further speculation on the function of microsatellite sequences is that they are hotspots for recombination (Stallings *et al.,* 1991; Weber *et al.,* 1993). There are two schools of thought, one for and the other against microsatellites' role as recombination hotspots. On the one hand, evidence in support of microsatellites as recombination hotspots comes from the observation of a $(GT)_7$ repeat sequence in the genome of SV40, a virus known to frequently integrate into the rat genome (Stallings *et al.*, 1991). Furthermore, there is a lack of these sequences in the β-satellite regions of the Drosophila polytene chromosomes, which do not undergo recombination. On the other hand, evidence against is supported by the fact that the telomeres of chromosomes, where the minisatellites are clustered, are known as recombination hotspots. Microsatellites, in contrast, are not confined to the telomeres but are randomly distributed throughout the genome, including regions where recombination is suppressed. Therefore, based on the latter observation microsatellites will tend to accumulate in regions with very low recombination frequencies and weak selective constraints on array length. (Stallings *et al.*, 1991). The use of microsatellites as genetic markers to trace inheritance patterns, however, relies on their crossing over during meiosis and, thus, as recombination hotspots.

Recently, different microsatellite mutations have been reported in tumour cells, where, for example, the number of repeats had increased or been reduced, or there had been loss of one STR allele (loss of heterozygosity), or the presence of an additional allele. These changes have been proposed as possible mechanisms for tumour initiation. However, a body of evidence strongly excludes the proposal that microsatellite (di-, tri-, and tetranucleotide repeats) instability seen in tumour cells is a mechanism of tumour initiation, but is rather a feature of tumour progression due to genomic instability (Canziani *et al.,* 1994; Dam *et al.,* 1995; Tischfield, 1997).

However, there are also microsatellite sequences associated with diseases and these are composed exclusively of trinucleotide STRs, referred to as pathogenic microsatellites.

1.6.5 Pathogenic microsatellites

Generally, microsatellites have mutation rates inversely proportional to their repeat unit size, with the dinucleotides' mutation rates reported as twice that of tetranucleotide STRs and the trinucleotide STRs intermediate between the two (Chakraborty *et al.*, 1997). In contrast, disease-associated (pathogenic) trinucleotide STRs have 3.9 to 6.9 times higher mutation rates than the tetranucleotide STRs (Chakraborty *et al*, 1997). To date, at least 13 inherited human diseases are caused by mutations of CG-rich triplet repeat sequences, characterised by abnormal expansion of the normal allele (e.g., in myotonic dystrophy (MIM 160900), the normal allele of 5-30 (CTG) repeats is expanded to >50-2000 in the affected individual (OMIM). The general trend seen in these diseases is that they are normally autosomal dominant, neurodegenerative diseases, caused by the unusual expansion of a CAG repeat in coding regions, or a CGG repeat in the non-coding regions (5' and 3' untranslated regions UTRs) with evidence of anticipation (Warren, 1996). An exception to the rule however, is Friedreich's ataxia, which is inherited as an autosomal recessive trait caused, in 97% of patients, by the abnormal expansion of an intronic GAA repeat and which shows little evidence of genetic anticipation (Campuzano *et al.*, 1996).

1.6.6 Genetic Anticipation

Abnormal expansion of the trinucleotide STRs at their respective loci (coding or non-coding regions) shows marked intergenerational instability, with offspring inheriting either the smaller or larger alleles than their parents. Expanded alleles have been found to either expand further or contract between generations. Concomitant with this gain in repeat-unit number with subsequent generations is increase in disease severity and penetrance and earlier age of onset, the phenomenon of genetic anticipation (Howeler *et al.*, 1989).

One proposed mechanism on how triplet repeat expansions, in the different regions of genes, might cause disease is that, as the sequence becomes larger, it might interfere either with transcription or lead to further deviation of the gene product from normal (Jennings, 1995). For example, the abnormally expanded GAA repeat in Friedreich's ataxia supposedly adopts a triple helical conformation *in vitro*, which Ohshima *et al.*, (1999) suggested is responsible for suppression of gene expression implicated in the disease pathogenesis.

Although their function in the genome is not yet known, microsatellites (di-, tri and tetras) are useful as polymorphic markers in a number of applications in genetics and medicine, namely:

- Sub-chromosomal localisation of inherited human disease genes
- Construction of haplotypes to trace the inheritance of a disease chromosome in families
- Paternity tests
- Personal identification in forensic sciences
- Molecular anthropology, to trace the origins of a population using mitochondrial DNA and Y chromosome microsatellites for maternal and paternal lineage, respectively (Edwards *et al.*, 1992).

1.6.7 Genome-wide linkage maps

Genome-wide linkage maps, consisting of genetic markers, mostly microsatellites, developed by different laboratories, Généthon $(CA)_n$, CHLC (tri- and tetranucleotide STRs) and the Utah Marker Development Group (di-, tri- and tetranucleotideSTRs), but typed on a common set of families, are available for utilisation by the research community (CHLC www, Généthon www, The Utah Marker Development Group). To date, more than 100 previously unknown disease-causing genes have been mapped by positional cloning using these resources, for example, long QT syndrome (Wang *et al.*, 1995), retinitis pigmentosa (Bardien *et al.*, 1995) and adult polycystic kidney disease (PKD Consortium, 1994).

The presence of family pedigrees, in which PFHBI segregates, lends the search for the PFHBI gene to the positional cloning approach (linkage analysis, genetic fine mapping and the identification of candidate genes and detection of disease-associated mutations). The present study gives an overview of PFHBI, a ventricular conduction system disorder and the positional cloning approach followed in an effort to identify the PFHBI-causing gene.

1.6.8 Genetic Linkage analysis in PFHBI

A random genome search using linkage analysis mapped the PFHBI locus to a 10cM region between the apolipoprotein C2 (*APOC2*) and kallikrein (*KLK1*) loci on chromosome 19q13.3 (Brink *et al.,* 1995). Subsequently, De Meeus *et al.* (1995), working independently, reported on a cardiac disease, designated ICCD in a Lebanese kindred, that showed similar clinical

features and mapped to a 13cM interval overlapping the PFHBI locus, between markers *D19S606* and *D19S571*. Genetic fine mapping efforts, using chromosome 19q13.3 dinucleotide STRs polymorphic markers, identified recombination breakpoints in the PFHBI-affected families, which reduced the PFHBI interval to 7cM, between markers *D19S412* and *D19S866*, centromeric and telomeric, respectively (De Jager, personal communication and Christoffels, 1997) (Figure 1.5). This new PFHBI interval still had a region of overlap with the ICCD interval (common marker *D19S606*) (Figure 1.5), thus, it is not inconceivable that PFHBI and ICCD are the same disease, possibly caused by different mutations in the same gene (Brink, personal communication).

Although the PFHBI search area had been reduced to 7cM, more markers residing within the target interval were required to genetically fine map this region further. This would then allow for a more focussed physical mapping effort.

To this end, using an integrated approach, both $(CA)_n$ repeats and trinucleotide repeat markers were developed from cosmid clones harbouring insert DNA from the PFHBI target interval, for the purpose of reducing the target interval, by high level definition of the telomeric and centromeric limits. At the same time, genome database searches were performed to identify any polymorphic markers lying within the target region.

Once the disease locus has been mapped, the next formidable task becomes the identification of candidate genes within the refined target interval. Concurrently with the development of genetic maps, has been the development of physical maps of the human genome consisting of overlapping DNA clones contained in large insert vectors, e.g., yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs) and P1 bacteriophage artificial chromosomes (PACs), which serve as substrates for chromosome walking, chromosome jumping and exon trapping methods to identify gene transcripts (Strachan and Read, 1996).

However, these positional cloning methods used for gene identification are often long and tedious.

The whole genome sequencing initiative of the Human Genome Project (Waterston and Sulston, 1998; Marshall, 1999) has resulted in the availability of comprehensive whole genome maps with identified genes and consequently, positional candidate gene-based strategies may supersede these time-consuming transcript identification positional cloning approaches.

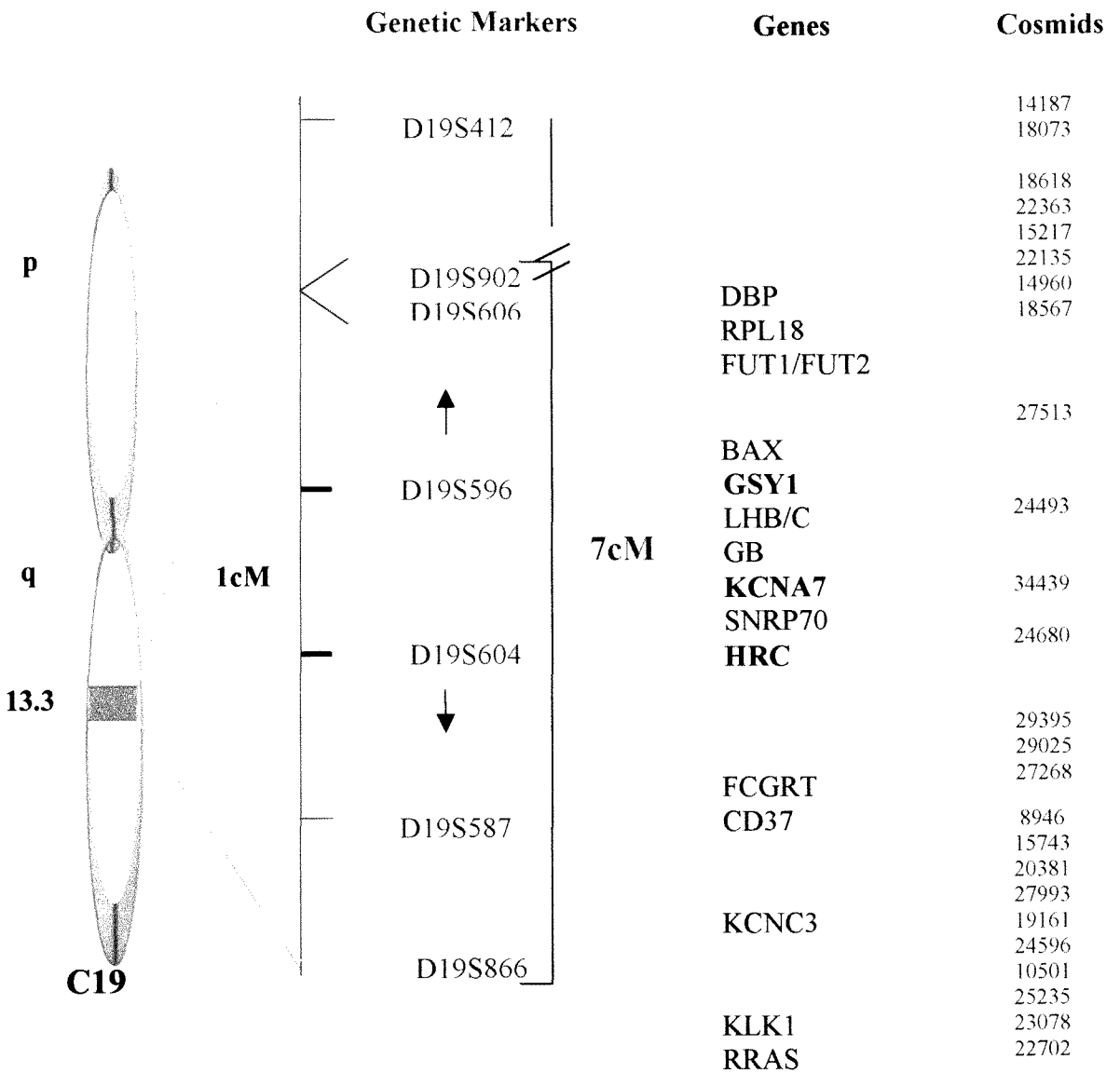


Figure 1.5 Schematic representation of a segment of chromosome 19q13.3 indicating the positions of the genetic markers, genes and cosmid clones spanning the PFHBI target interval (http://www-bio.llnl.gov/genome-bin/).

The 1cM region represents the region of chromosome 19q13.3 common in all PFHBI-affected individuals (common core haplotype). The arrows indicate the area in which new polymorphic markers were sought. The genes in bold are the plausible candidates for the PFHBI-causative gene.

1.7 THE POSITIONAL CANDIDATE GENE APPROACH

In a search for a disease-causing gene, the selection of candidate genes is based on understanding the mechanisms driving the normal function of an organ or tissue, which when defective can cause disease, i.e., functional candidate genes. However, once a disease locus has been mapped, genes at the same sub-chromosomal location as the disease locus are referred to as positional candidates and the confidence levels in a gene's candidacy are substantially increased if its function can be associated with the disease's molecular pathophysiology (Strachan and Read, 1996). Such genes are referred to as positional candidates and represent plausible candidate genes.

Chromosome 19 is a relatively small, approximately 60Mb, GC-rich and gene-rich chromosome, containing a number of disease genes as well as members of numerous gene families (Gordon *et al.*, 1995; Mohrenweiser *et al.*, 1995; Schuler *et al.*, 1996) (Figure 1.6). The PFHBI locus on chromosome 19q13.3 has a number of known genes and it was in this region that positional candidate genes were sought.

The most compelling positional candidates are those genes with:

- appropriate function or tissue expression pattern,
- homology to a gene implicated in an animal model of the disease and
- known homologues in other species based on comparative genomics (animal models), or homology, or functional relatedness, to a gene implicated in a similar human disease phenotype, or with appropriate tissue expression or function.

1.7.1 Candidate genes

The plausible positional candidates within the PFHBI locus (Figure 1.5) include genes encoding for:

- glycogen synthase (*GSY1*)- a skeletal muscle gene that is expressed in the heart,
- a potassium ion channel (*KCNA7*) – K ions control excitability of cardiac muscle cells (Klocke *et al.*, 1993),
- histidine-rich calcium-binding protein (*HRC*) – with postulated function in cardiac calcium ion fluxes

Although none of the genes within the designated PFHBI locus have a direct association with known cardiac conduction system pathology, in the present project, two genes *GSY1* and *HRC* were selected, as they are attractive candidates for the PFHBI-causative gene.

1.7.1.2 *Glycogen synthase (*GSYI*)*

Glycogen synthase (GSY) catalyses the rate-limiting step in the synthesis of glycogen. Glycogen, ubiquitous in mammalian tissues, represents a readily mobilisable form of glucose and thus serves to maintain energy homeostasis within cells. The glycogen in liver serves an additional unique purpose of buffering blood glucose (Kaslow and Lesikar, 1984). The *GSY1* and *GSY2* genes have since been mapped to chromosomes 19q13.3 (Lehto *et al.*, 1993) and 12 (Nuttall *et al.*, 1993), respectively. Liver *GSY2* is only expressed in the liver, in contrast to muscle *GSY1*, which is expressed in skeletal muscle, heart, kidney and fat (Kaslow *et al.*, 1985). Expression studies by Browner *et al.*, (1989) comparing *GSY1* mRNA levels in foetal heart and foetal skeletal muscle showed that *GSY1* was more abundant in the foetal heart. This observation, in the light that one of the characteristic features of the conduction system genes is that they continue to be expressed postnatally (Gorza *et al.*, 1994), made *GSY1* an attractive candidate for PFHBI. The *GSY1* gene has been implicated, but not confirmed, as a cause of the insulin resistance in non-insulin-dependent diabetes mellitus (NIDDM) in the general population (Orho *et al.*, 1995).

Although the role of glycogen in the conduction system is not yet known, it can be assumed that it is likely to serve a purpose. As discussed in section 1.2.4, glycogen occurs in abundance in the cardiac conduction system (James, 1970; Davies, 1971). Furthermore, the glycogen-rich conduction system remains unscathed during myocardial infarction (Thornell and Sjostrom, 1975), a period when glycogen is the only source of glycolytic flux (Opie, 1991; King and Opie, 1998). Therefore, it is possible that mutations in *GSY1*, negatively affecting glycogen synthesis, cause the conduction system to be more susceptible to stress-induced damage. Thus, it would be worth investigating genes coding for enzymes involved in glycogen metabolism for possible role in PFHBI pathophysiology. The human muscle *GSY1* and its gene product have been characterised. The gene consists of 16 exons, spanning approximately 27kb, encoding a 737 amino acid protein. The functionally important domains, namely the nine phosphorylation sites, are coded for by exons 1 (2 of 9) and 16 (remaining 7) and exon 12 codes for a highly conserved region (Orho *et al.*, 1995). Consequently, it could be proposed that any sequence variation occurring in these regions might have a deleterious effect on function of the protein.

1.7.1.3 *Histidine-rich calcium-binding protein (*HRC*)*

The histidine-rich calcium-binding protein (HRC) is a luminal sarcoplasmic reticulum (SR) protein identified by its ability to bind low-density lipoprotein (LDL) with high affinity and also calcium ions (Hofmann *et al.,* 1989), whose function is as yet unknown. The gene was shown to be expressed in both skeletal and heart muscle and subsequently localised to human chromosome 19 (Hofmann *et al.*, 1991). The mouse *hrc* orthologue was mapped to chromosome 7 and because of known synteny between the two chromosomes, human HRC was specifically placed on 19q13.3, the PFHBI target area (Hofmann *et al.,* 1990).

Calcium ions are involved in range of signalling cascades as secondary messengers in cells of most tissues. The regulation of calcium is in turn mediated by calcium-binding proteins, which in the heart are located within the sarcoplasmic reticulum (Opie, 1997). In the heart, Ca^{2+} regulate the contraction and relaxation phases of the cardiac cycle, where the release of Ca^{2+} from the SR is controlled by entry of low levels of Ca^{2+} into the cell during depolarisation (calcium-induced calcium release). For example, the SR protein, calsequestrin, releases Ca^{2+} into the cytosol in response to the amounts of Ca^{2+} entering a cardiac myocyte with each wave of depolarisation (Opie, 1997). The cardiac conduction system has no known contractile function and although it contains a sparse population of myoctes, and sometimes SR, their role in the conduction system has not been clearly defined, but they possibly fulfil some function.

The histidine-rich calcium-binding protein encoding gene, *HRC*, was initially selected as a candidate because it is expressed in the heart and also because of its postulated function in Ca^{2+} regulation in the SR (Hoffman *et al.*, 1989). In addition, analysis of the published HRC gene sequence showed that exon 1 of *HRC* contains stretches of two trinucleotide $(GAG)_n$ and $(GAT)_n$ STRs (coding for glutamic acid and aspartic acid residues, respectively). The presence of these repeat motifs, together with the fact that genetic anticipation had been described in PFHBI, led to the search for abnormally expanded repeats in exon 1 of *HRC*. Both De Meeus and Christoffels analysed these two repeat loci in the ICCD and PFHBI-affected patients and their unaffected close relatives, respectively, but did not detect any unusually expanded $(GAG)_n$ or $(GAT)_n$ STRs (De Meeus *et al.* 1995; Christoffels, 1997). These results excluded triplet repeat expansion in exon 1 of *HRC* as the disease-causing mechanism in both ICCD and PFHBI. The *HRC*, however, has six exons, therefore, five other exons had not been analysed for the presence of potentially disease-causing mutations.

1.8 THE PRESENT STUDY

The sub-chromosomal location of the PFHBI locus was established, on chromosome 19q13.3, by linkage analysis (Brink *et al.*, 1995). Furthermore, the original target area spanning 10cM was reduced to 7cM by genetic fine mapping using dinucleotide polymorphic genetic markers available on the databases (De Jager, personal communication). However, to physically map a gene within such a large genetic interval would be an arduous task.

Therefore, to increase the resolution of the current genetic map, trinucleotide and tetranucleotide genetic polymorphic STRS were considered as resources, to complement the dinucleotide markers available, thereby allowing saturation fine mapping. Unfortunately, there were no known trinucleotide or tetranucleotide markers within the target interval. Therefore, as part of an integrated approach, development of tetranucleotide polymorphic markers, using cosmid clones spanning the PFHBI target interval obtained from the Lawrence Livermore National Laboratory (USA), was undertaken using either sub-cloning-based or vectorette PCR-based strategies. The $(AAAT)_n$ STR was selected because it was one of the most abundant tetranucleotide repeats which gave the clearest results on genotyping (The Utah Marker Development Group, 1996).

The PFHBI target interval on chromosome 19q13.3 is a gene-rich area and among the plausible positional candidates are the genes encoding GSY1 and HRC (Figure 1.5).

Following a traditional positional cloning approach, concurrently with the development of tetranucleotide genetic markers, the *GSY1* and *HRC* candidate genes were screened for PFHBI-causing mutations in families where PFHBI segregates. A variety of mutation detection methods were applied for the different candidates based on available sequence information. Based on homology searches and available literature, only the most conserved exons of both *GSY1* and *HRC* were selected for mutation screening. The PCR-SSCP analysis was the method of choice for mutation screening of the selected *GSY1* exons as primer sequences had been published. In contrast, the available sequence information on *HRC* was limiting, as only a few bases of intronic sequence flanking all 6 exons had been published. This lack of extensive intronic sequence would hinder complete exon screening. Consequently, selected exons were screened by direct sequencing, which would enable the acquisition of more intronic sequence, thus allowing complete exon and intron exon junction screening.

It was anticipated that, in conjunction with the complementary approaches used concurrently by other members of the PFHBI group, the present project would speed the search for and identification of the PFHBI-causative gene and associated mutations.

CHAPTER 2

METHODS

2.1 SELECTION OF FAMILIES

Individuals participating in this study belong to the families, in which PFHBI segregates, which were initially described by Brink and Torrington (1977). These families were subsequently extended and divided into 3 apparently unrelated pedigrees, designated pedigrees 1, 2 and 5, (Torrington *et al.*, 1986; Brink *et al.*, 1994) and also include additional family members who have since come to our attention (Figure 1.3). From each pedigree, a panel comprising PFHBI-affected individuals and their first-degree relatives, which included siblings, children and married-in individuals used as control samples, was selected for mutation screening analysis of the candidate genes (Figure 2.1).

2.2 DEFINITION OF PHENOTYPE

Progressive familial heart block I was diagnosed on an ECG by the presence, of RBBB, RBBB complicated by left anterior or posterior hemiblock (LAHB/LPHB) or CHB with a broad QRS-complex, in the absence of other cardiac diseases, in a family setting (Brink and Torrington, 1977).

2.3 BLOOD COLLECTION

The University of Stellenbosch's institutional ethics committee approved the present study (Project 86/085). Prior to the start of this study, blood samples had been collected from members of families in which PFHBI segregates, after contact by one or more of the following, Prof. P.-L van der Merwe, Department of Cardiology, Prof. Paul Brink, Department of Internal Medicine, Tygerberg Hospital, their own physicians or Genetics Service nurses. Informed verbal consent was obtained from the participants, or their parents in the case of minors. Twenty millilitres of peripheral blood were collected, in 5ml aliquots, into four EDTA tubes per individual and DNA isolated from the white blood cells.

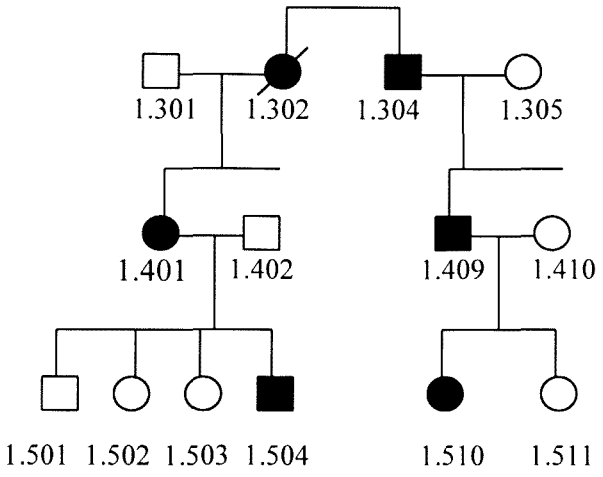
2.3.1 Isolation of nuclei

Blood from two EDTA tubes was transferred into a 50ml Falcon tube, 20ml ice-cold cell lysis buffer (Appendix) was added, the solution was mixed and incubated on ice for 10 minutes. The mixture was centrifuged in a benchtop centrifuge (model TJ-6, Beckman Instruments Inc., Palo Alto, California, USA) at 3000rpm for 10 minutes at room temperature. To wash the pellet, 10ml cell-lysis buffer was added, the sample left on ice for 10 minutes and

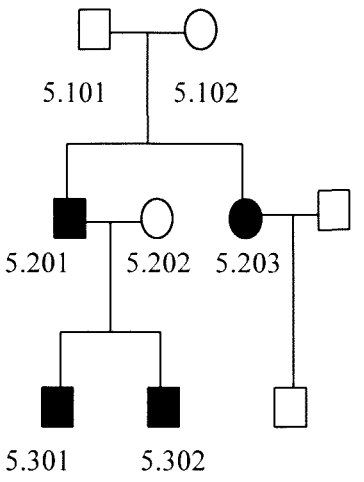
centrifuged as before. The supernatant was again discarded and the process repeated until the pellet was pale pink to white, after which it was resuspended in 450µl Na-EDTA solution (Appendix) and 50µl 10% SDS. The intact nuclei were either used immediately to prepare DNA or were stored at -70°C until required.

2.3.2 DNA extraction from nuclei

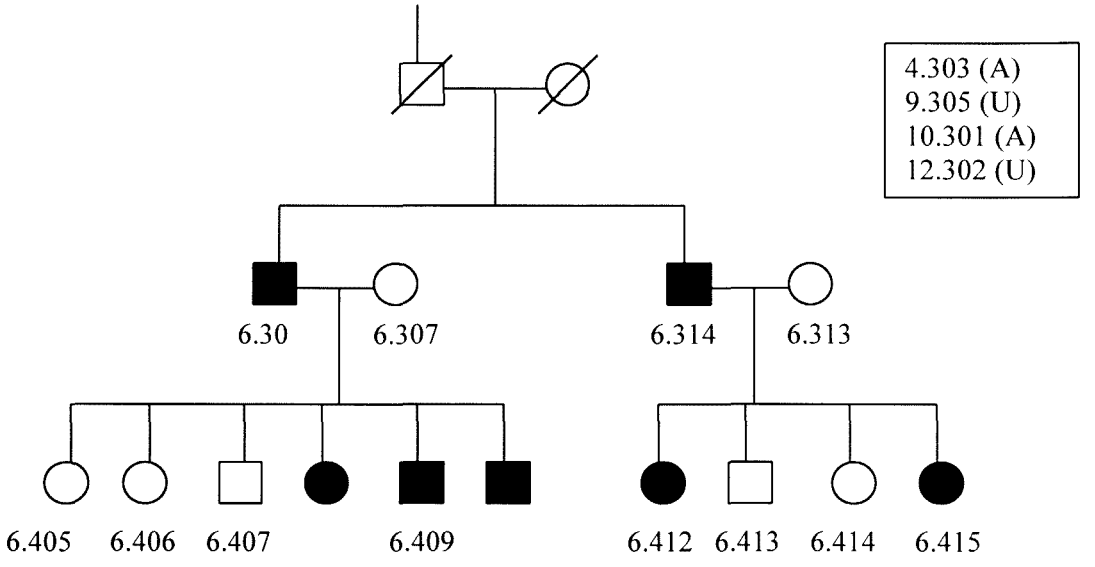
A 100µl aliquot of proteinase K (10mg/ml) (Boehringer Mannheim, Mannheim, Germany) was added to the nuclei suspension and the mixture incubated overnight at 37°C, with gentle agitation. The sample was transferred to a 50ml propylene tube before the addition 2ml sterile dH_2O and 500µl 3M Na-acetate. Organic extraction of DNA was achieved by the addition of 2.5ml phenol/chloroform solution (Appendix) to the sample, which was then mixed by incubation on a Voss rotator (Maldon, England) for 10 minutes. The mixture was transferred to a Corex tube and centrifuged in a Sorvall RC-5B centrifuge using an SS34 rotor at 8000 rpm for 10 minutes at 4°C. The aqueous phase, containing DNA, was transferred to a clean Corex tube, taking care not to disturb the interface. Octanol/chloroform (Appendix), 2.5ml, was added, the contents of the tube mixed well and centrifuged as above. The aqueous phase was again transferred to a clean tube, 7.5ml ice-cold ethanol (96%) was added and the solution gently mixed until DNA strands appeared. Using a yellow Gilson pipette tip, the DNA strands were collected and transferred to a clean eppendorf tube, this was followed by a washing step with 1ml ethanol (70%) and centrifugation in a microfuge (Beckmann, Germany) at 14000 rpm for 3 minutes at room temperature. Ethanol was carefully decanted, taking care not to disturb the pellet. The ethanol wash was repeated. The remaining ethanol was carefully blotted out using paper towel and the DNA-containing pellet air-dried for 1hr. DNA was resuspended in 500µl 1x TE (pH8.0) (Appendix), the solution heated for 10 minutes at 65°C and mixed on the Voss rotator (Voss of Maldon, Protea Dento-Medical Services, Cape Town, RSA) at 4°C for 3 days. The optical density of the DNA solution was determined in a spectrophotometer (Spectronic 1201, Milton Roy) at 260nm and the concentration adjusted to 0.2µg/µl either by further dilution with 1xTE or by another ethanol precipitation step and subsequent resuspension in the appropriate volume of 1xTE, with rotation mixing as before.



PEDIGREE 1



PEDIGREE 5



PEDIGREE 2

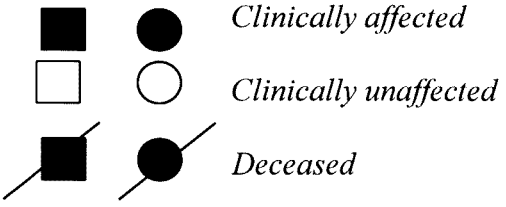


Figure 2.1 PFHBI-affected family branches from pedigrees 1, 2 and the complete pedigree 5.

The panel of PFHBI-affected individuals and their first-degree relatives whose DNA samples were analysed for disease-causing mutations by PCR-SSCP and direct sequencing. Additional members from other pedigree 2 sub-families (not shown) were included (boxed). A = clinically affected, U = clinically unaffected.

2.4 PRIMER DESIGN AND SYNTHESIS

Oligonucleotide primer pairs were required for polymerase chain reaction (PCR) amplification of exons and/or introns of candidate genes for mutation screening and for amplification of vectorette libraries to probe for the presence of AAAT tetranucleotide sequences. Primer pairs for which published sequences were available were custom synthesised (DNA Laboratory, Dept. Biochemistry, University of Cape Town, Cape Town, R.S.A.). Alternatively, primers were designed from the published gene sequence with the assistance of OLIGOV2 and PRIMER programs. Primer sequences used in the amplification of targeted regions of candidate genes are listed in Table 2.1, while those used in the development and characterisation of the $(AAAT)_n$ tetranucleotide repeat are given in Table 2.2. An additional oligonucleotide, with the sequence $(AAAT)_{10}$, was designed and used as a radio-labelled probe for the identification of the AAAT repeat motif.

2.5 POLYMERASE CHAIN REACTION AMPLIFICATION

2.5.1 Non-radioactive PCR

Non-radioactive PCR was used for the amplification of selected exons of *GSY*1 for mutation screening, to generate template both for sequencing regions (exons/introns) of *HRC* and also for regions of cosmid insert harbouring the $(AAAT)_n$ repeat motif. The general method followed was: 200ng of human genomic DNA was aliquoted into a 0.5ml eppendorf centrifuge tube. To this, was added a reaction cocktail containing: 150ng of each primer, 75μM each of dATP, dGTP, dTTP and dCTP (Promega Corp, Madison, Wisconsin, USA), $MgCl_2$ (Table 2.1), 5μl of a 10x Taq DNA polymerase buffer supplied by the manufacturer (Promega Corp, Madison, Wisconsin, USA) (containing 500mM KCl, 100mM Tris-Cl, pH = 9 @ 25°C, 1% TritonX-100), 0.5U Taq DNA polymerase (Promega Corp, Madison, Wisconsin, USA) and 5% formamide (Sigma Chemical Company, St Louis, Missouri, USA) and ddH_2O to 50μl. The reaction mixture was subsequently overlaid with 30μl mineral oil to prevent evaporation during amplification.

Amplification was carried out in a Gene E Thermal Cycler (Techne Ltd., Cambridge, UK) for thirty cycles. Each cycle typically comprised a 30s denaturation step at 94°C, annealing of primers at 3-5°C below the lowest Tm (Table 2.1) for 30s and a 60s extension step at 72°C.

For specific amplification conditions refer to table 2.1. An additional final extension step at 72°C for 5min was included for the amplification of *HRC* regions, in which the expected amplification products were >500bp long in size (Table 2.1). Subsequently, aliquots of each sample were electrophoresed on agarose gels (section 2.6.1) to verify the success of the PCR and the samples stored at room temperature until further analysis.

Non-radioactive PCR of the selected *HRC* regions was performed in a similar manner, except that 0.5U of BIOTAQ™ DNA polymerase (Bioline, UK Ltd, UK) and 10x NH_4 reaction buffer (containing 160mM $(NH_4)_2SO_4$, 670Mm Tri-HCl, pH 8.8 at 25°C, 0.1% Tween-20).

2.5.2 Radioactive PCR

For genotyping the AAAT repeat identified during the course of this study, radioactive PCR was performed as follows: a cocktail containing 60ng of each primer, 1μCi α-^{32}P[dCTP] (Amersham International, UK), 5μM unlabelled dCTP, 200μM of each of dATP, dTTP and dGTP, 1.5mM $MgCl_2$, 1μl of 10x Taq DNA polymerase buffer supplied by the manufacturers (Promega Corp, Madison, Wisconsin, USA) (containing 500mM KCl, 100mM Tris-Cl, pH = 9 @ 25°C, 1% TritonX-100), 0.5U Taq DNA polymerase (Promega Corp, Madison, Wisconsin, USA) and 5% formamide (Sigma Chemical Company, St Louis, Missouri, USA) and ddH_2O to a final volume of 10μl was added to 100ng of genomic DNA template aliqouted in a 0.5ml eppendorf tube. Samples were amplified in a GeneE Thermal Cycler using similar cycling time as for the non-radioactive PCR. When the amplification reaction was completed 6μl of loading dye (Appendix I) was added to each sample prior to analysis on a denaturing polyacrylamide gel (section 2.6.3) or they were stored at room temperature until required.

2.5.3 Direct Sequencing

Preparation of template

Direct cycle sequencing of amplified products was performed in both directions. First, 45-90μl of PCR amplification product was gel extracted and DNA purified using the QIAquick gel extraction kit (QIAGEN, Germany) with minor modifications to the manufacturer's instructions. To elute the DNA in a more concentrated form, 30μl ddH_2O and not 50μl 1xTE elution buffer were applied to the spin-column. This DNA was then used as template for the sequencing reaction.

*Table2.1: Primer sequences, annealing temperature, magnesium chloride concentration and expected product size following PCR amplification of specific exons/regions of candidate genes **GSY1** and **HRC.***

Gene	Exon/Region	Forward Primer 5'-3'	Reverse Primer 5'-3'	Tm	$MgCl_2$	Size
		F	R			
GSY1	4[a]	gcagtagtggtgacagattg	ttcctggcatactcgcagtc	57°C	1.5mM	379 bp
	5[a]	ccaatccttctaaccgactg	gcctacctcattcacgtctg	58°C	3.0 mM	225 bp
	11[a]	taccettcttgtggctcctg	agccctgaccaaatgccctc	60°C	1.25 mM	159 bp
	12[a]	gcacttgggctgagtccttc	gatgccatgaccacgctgtc	60°C	1.25 mM	200 bp
	16A[a]	ccctggtttgcacccacatc	gacttagttacgctcctcgc	60°C	1.25 mM	358 bp
HRC	2F-3R[b]	ccccgcagGTCCACAGG	gaggtctcacCTGGCACT	58°C	1.5 mM	~940 bp
	4F-5R[b]	cgcctccagCACTGTCAG	cacttacCTGATAAAGGGTG	59°C	6.0 mM	538 bp
	4F-6R[b]	cgcctccagCACTGTCAG	CGGGGAGCACGTTTATTCGG	60°C	3.0 mM	854 bp
	2F-2R[c]	ggacaggagagaggagg	tggcctgcccttgctctg	75/61°C	1.5 mM	300 bp
	3F-4R[c]	aagtgcgttagagttcc	accaggattcaccacgc			
	5F-6R[c]	gctatactgagtggctgcc	cggggacgtttattcgg	64°C	2.0 mM	353 bp

Tm = annealing temperature. $MgCl_2$ = magnesium chloride. a = primer sequences obtained from the literature (Orho et al., 1995). b = primers designed from published gene, both intronic and exonic, sequences (Hofmann et al., 1991), lower and upper case letters indicate intronic and exonic sequences, respectively. c = primers designed from sequence obtained by direct manual sequencing. Primer sequences are shown in 5'-3' orientation. GSY1 = human muscle glycogen synthase gene and HRC = histidine-rich calcium-binding protein encoding gene. bp = base pairs.

Table2.2: *Oligonucleotide primer sequences, PCR amplification conditions and the observed product sizes used in the isolation and characterisation of AAAT STR in the sub-cloned fragments and cosmid vectorette libraries.*

Cosmid	Primer pair	Upstream (5'-3')	Downstream (5'-3')	Tm	$MgCl_2$	Size (bp)
	T3 – T7	aattaaccctcactaaaggg	gtaatacgactcactatagggc	50°C	1.5mM	
29395	395F-395R	cgatcacaccactgcac	gtgatctagaccagtacaac	48°C	1.5mM	~137
24493 *AluI*[a]	A4[c]-224	aaataaataaataaatc	tctcccttctcgaatggtaaccgttcgtac	34°C	1.5mM	~250
	493R-224	tacatgctacacgtggat	tctcccttctcgaatggtaaccgttcgtac	65°C	1.5mM	~500
	493R-493F	tacatgctacacgtggat	atgcctcttacctgtcacc	60°C	1.5mM	~190 & ~220
20381 *RsaI*[b]	A3[c]-224	aaataaataaataaatg	tctcccttctcgaatggtaaccgttcgtac	38°C	1.5mM	~800
	381R-224	ctgcaacacatctcacagtc	tctcccttctcgaatggtaaccgttcgtac	62°C	1.5mM	~800
	381R-381F	ctgcaacacatctcacagtc	ccttcactcactcatgcatc	60°C	1.5mM	~514
	381F'	ctgcaacacatctcacagtc	ccagctctgtgttcctca	60°C	1.5mM	~339
	10108[d]		tctcccttctcgaatggtaaccgttcgtac	55°C	1.5mM	

Tm = annealing temperature, T3 and T7 = pBluescript T3 and T7 promotor region sequences, respectively, used to sequence the positively hybridising sub-cloned insert DNA. a = cosmid 24493 vectorette library generated by digestion with restriction enzyme AluI*. b = cosmid 20381 vectorette library generated by digestion with restriction enzyme* RsaI*. c = One of a set of 4 anchored (A) primers: 1= $(A_3T)_4$-A, 2 = $(A_3T)_4$-T, 3 = $(A_3T)_4$-G, 4 = $(A_3T)_4$-C. d = vectorette-specific sequencing primer 10108.*

Cycle sequencing

Five microlitres of the concentrated template DNA was sequenced using the fmol sequencing kit (Promega US) according to the manufacturer's instructions. Specifically, the sequencing reaction mixture contained 5µl template (DNA), 5µl 5x sequencing buffer, 5% DMSO (2:3 dilution in ddH_2O, Sigma, U.S.A), 10pmol of primer, 0.5µl α-^{32}P[dCTP] (Amersham International, UK), ddH_2O to 16µl and 5U of sequencing grade *Taq* polymerase (Promega, USA). Four microlitres of the above cocktail mixture was added to four previously prepared, pre-chilled 0.5ml eppendorf tubes, each containing 2µl aliquots of a cocktail composed of all four deoxynucleotides, as well as one specific dideoxynucleotide, namely, ddATP, ddTTP, ddGTP or ddCTP and 20µl mineral oil. Samples were then placed into a pre-heated GeneE Thermal Cycler (Techne Ltd., Cambridge, UK) and held at 95°C, for 2 minutes before commencing cycling, using a thermal cycling profile similar to that of the original PCR amplification. When the sequencing cycle was completed, 4µl sequencing stop solution (Promega, USA) was added to each tube. These samples were either stored at -20°C until required or denatured at 95°C in a heating block for 2 minutes and a 4µl aliquot immediately loaded on a pre-electrophoresed denaturing polyacrylamide gel (section 2.6.3).

2.6 GEL ELECTROPHORESIS

2.6.1 Agarose gel electrophoresis

Agarose gels was used for verification of success of PCR amplification, preparative gels for sequencing template and for analysis of the purified cosmid and plasmid DNA. Five microlitres of either PCR product or cosmid/plasmid DNA were mixed with 1µl of bromophenol blue loading dye. The samples were subsequently loaded and electrophoresed in 1.5% to 2% (depending on the size of the amplified product) or 0.7% to 1% (for cosmid and plasmid DNA) horizontal agarose gels of 12cm x 5cm x1cm dimensions, containing 1µg/ml ethidium bromide (EtBr) and 1xTBE electrophoresis buffer (Appendix I). The samples were co-electrophoresed with a sample of bacteriophage λDNA digested with restriction enzyme *Pst*I used as a molecular size marker (Appendix I). The gel was subsequently visualised on longwave ultraviolet transilluminator (3UV™ Transilluminator model LMS-26E) and a photograph was taken using an ITC camera (Berkenhoff and Drebes, Germany) and a Sony video graphic printer (UP-860CE).

2.6.2 Non-denaturing polyacrylamide gel electrophoresis

Non-denaturing polyacrylamide gels were used for verification of restriction enzyme digests, using a 12% polyacrylamide solution (Appendix I) and for single strand conformation polymorphism (SSCP) mutation analysis.

Mutation analysis by SSCP

For each panel of samples analysed, four different non-denaturing polyacrylamide gel conditions were used, as standard, for PCR-SSCP mutation screening, namely, 5% or 10% non-denaturing polyacrylamide gel, either with or without 10% glycerol. A 0.35x MDE™ (FMC Bioproducts, USA) gel was used for further analysis of samples showing mobility shifts to achieve better resolution of conformational changes observed on non-denaturing polyacrylamide gels.

Casting the gel

The gel (Appendix) was cast between 390mm x 290mm x 1mm glass plates, which were first washed with Cal-liquid hand soap (Cal-Chem, R.S.A), rinsed in tap water and wiped dry with paper towel, after which the surfaces were sprayed with 70% ethanol and wiped clean as before. Gelbond™ PAG polyester film (FMC, Bioproducts, Rockland, Maine, USA) was attached to the back plate by its hydrophobic side, to facilitate this, 70% alcohol was lightly sprayed onto this plate. The hydrophilic side of the Gelbond™ was exposed so that, when the gel solution was poured, the two would bond covalently, thus creating a solid support for the gel for subsequent silver staining. The front plate was wiped with paper towel spotted with Wynn's CTHRU windshield rain dispersant (Wynn Oil S.A. Pty Ltd,SA).

One-millimetre spacers were placed on either side of the gel-bond-covered back plate and the front plate then placed on top with the CTHRU-coated side on the inside. The edges, on the sides and bottom of the plates, were sealed with gel-sealing tape (SIGMA, Germany) after which the casting rubber boot (S2 casting boot, Life™ Technologies, UK) was fitted for further sealing. The gel solution was poured with the glass plate assembly slanted at an angle. Immediately thereafter, the glass plates assembly was laid down horizontally, a square-tooth comb fitted (to create wells), the plates clamped and the gel allowed to set for ~1hr.

Electrophoresis

Thereafter, the casting boot, the sealing tape at the bottom of the plates and the comb were removed, in this order. The glass plate assembly was mounted vertically on the electrophoresis apparatus (Omeg Scientific, R.S.A), moved to the cold room (4°C) and 0.5xTBE (Appendix) buffer added to the reservoir tanks. The wells were washed with buffer to remove any residues, and the denatured PCR sample (section 2.8.1.2) immediately loaded. Electrophoresis was carried out in 0.5x TBE buffer, at a constant power of 50W for 2-7hrs at 4°C, depending on the type of gel being used. Subsequently, the glass assembly was dismantled and the gel silver stained (discussed in the next section).

Silver Staining

Following electrophoresis of non-radioactive samples on the non-denaturing gels, the glass assembly was dismantled and the gel still supported by the Gelbond™, was silver stained by incubation in solution B (Appendix I), on a Labcon shaking platform (Labdesign Engineering, R.S.A) for 10 minutes at room temperature, followed by a brief ddH_2O rinse and incubation in solution C (Appendix I) on the Labcon shaking platform, for a minimum of 20 minutes or until the bands became visible.

2.6.3 Denaturing gel electrophoresis

Template Sequencing

The 0.5mm 5% polyacrylamide denaturing sequencing gel (Appendix I) was cast as described in section 2.6.2, but without the use of Gelbond™ film and using 0.5mm thick spacers and a shark's-tooth comb. Pre-electrophoresis was carried out in pre-heated 0.5x TBE buffer at 1800V for ~30 minutes, the samples (section 2.5.3) were heat-denatured at 95°C for 2 minutes, 4µl loaded in the wells and electrophoresis continued for 2hrs. For long sequence reads, electrophoresis proceeded until the bromophenol blue dye had migrated more than two thirds down the gel. At this point, the gel was switched off and a 4µl aliquot of the same sample, heat-denatured, was loaded on a different lane. Electrophoresis was resumed and allowed to proceed until the second bromophenol blue front had migrated more than two thirds down the gel, for short sequence reads.

After electrophoretic separation had been completed, the plates were dismantled, the gel was carefully lifted off the glass plates using 3MM Whatman paper cut to the same size and then covered with cling-film (GLAD™). Subsequently, the gel was dried for an hour on a Drygel

Sr Slabgel Dryer (model se 1160) and thereafter, placed in a cassette and exposed to an X-ray film (Cronex-4, Protea Medical) overnight at room temperature.

The X-ray film (Cronex-4, Protea Medical) was developed, in the dark, by immersion in developing solution (Appendix I) for 4 minutes, stop solution (Appendix) for 0.5 minutes and fixing solution (Appendix I) for 2 minutes. The X-ray film (Cronex-4, Protea Medical) was then rinsed off with running water and hung to dry at room temperature.

DNA Sequence analysis

The information from the sequenced *HRC* regions, from both the PFHBI-affected individual and an unaffected close relative, were read manually and the sequences analysed for any variations. Furthermore, these sequences were compared to the wild type gene (*HRC*) sequence deposited in GENBANK (http://www2.ncbi.nlm.nih.gov/) to ascertain the authenticity of the obtained sequence and also confirm the observed sequence variations.

To design primers from the acquired sequence information, *HRC* intronic regions (section 2.8.2) or vectorette PCR (section 2.7.7), the long and short sequence runs were read manually and the region of overlap between the two established, thus generating one long sequence. The 3'-end of the longer sequence was then used to design new PCR primers.

To determine if complete sequencing of the vectorette library fragment generated by amplification using the putative repeat-flanking primers **F** and **R** (section 2.7.7) had been achieved both the sequences were read and the complement of the **R** sequence (**R'**) was established. Subsequently, a region of overlap between the F and R' sequence information was established and the two sequences then merged.

Genotyping of $(AAAT)_n$ repeat locus

Following the sequencing of the recombinant plasmid containing the $(AAAT)_n$ repeat (section 2.5.3), primer pairs for amplification of $(AAAT)_n$ repeat locus for genotyping were designed from the repeat-flanking sequences read from the manual sequencing autoradiograph. The chosen sequences (Table 2.2) were analysed to avoid the possibility of self-complementarity or primer-dimer formation using OLIGOV2 prior to synthesis. The primers were then used to PCR amplify 40 DNA samples of unrelated individuals of Mixed Ancestry (section 2.5.2). The amplified samples were then analysed on a denaturing polyacrylamide gel (section 2.6.3).

To determine the genotype, the PCR-amplified samples (section 2.5.2) were heat-denatured, loaded into the wells of the gel (section 2.6.3) and electrophoresis carried out for ~1hr or until the bromophenol blue front had reached the bottom of the gel. The dried gel was then exposed, overnight, to an X-ray film (Cronex-4, Protea Medical) to visualise the different alleles and hence determine if the repeat was polymorphic.

2.7 DEVELOPMENT OF AAAT POLYMORPHIC STR MARKERS

Tetranucleotide $(AAAT)_n$ repeat polymorphic markers were sought as part of an integrated approach to find more genetic markers, within the chromosome 19q13.3 target search area, for the purpose of narrowing the region, by identification of recombination breakpoints. Twentyfive cosmid clones containing inserts spaced at 500kb intervals across the chromosome 19q13.3 region were obtained, as agar stabs, from the Lawrence Livermore National Laboratory (LLNL) (California, USA). The LLNL were responsible for developing the physical map of chromosome 19 and thereafter sequencing this chromosome. The cosmids used in this study were selected from the maps published by LLNL (http://www-bio.llnl.gov/genome-bin/). The cosmids were derived by subcloning fragments of human chromosome 19q13.3 into the Lawrist-16 vector and the recombinant cosmids propagated into the $DH5^{\alpha}$ strain of *Escherichia coli* (E. *coli*) (De Jong *et al.*, 1989). These cosmids were used as resource for identifying tetranucleotide repeat sequences.

2.7 1 Cosmid culture preparation

To prepare cosmid cultures, a sterile plastic loop was dipped into the agar stab culture, or used to scrape the top of frozen glycerol stocks prepared from the stabs, and subsequently inoculated into 2ml of Luria Bertani (LB) broth containing kanamycin (Kn) (50μg/ml) in 15ml tubes. The culture was then incubated at 37°C for 7-8hrs or overnight, with vigorous shaking. Using a sterile plastic loop, a sample of the 2ml culture was streaked onto LB/agar/Kn (50μg/ml) plates, which were incubated at 37°C overnight. A single colony was picked from the overnight agar plate, inoculated in 250ml LB/Kn (50μg/ml) medium and incubated at 37°C overnight with vigorous shaking. The cells harvested from this culture served as the source of cosmid DNA. Plasmid vector, pBluescript SK, cultures were prepared in a similar manner, but using LB/ampicillin (Amp) (50μg/ml) selective medium. However, some (6/25) of the cosmid clones were no longer viable and could thus not be pursued for culture preparation

2.7.2 DNA Purification

Cosmid DNA Maxi/Midi Prep

The QIAGEN™ commercial DNA purification kits (QIAGEN, Germany) was used to isolate cosmid DNA from bacterial genomic DNA with slight modifications to the manufacturer's protocol. Cells were harvested from the 250ml overnight culture by centrifugation at 4°C for 10 minutes at 6000 rpm in a Sorvall RC-5B centrifuge using a GSA rotor and the supernatant was discarded. The cell-containing pellet, obtained from 250ml of cell culture, was completely resuspended in 50ml of buffer P1 (Appendix I) containing RNAse A (100µg/ml). To lyse the cells, 50ml of buffer P2 (Appendix I) were added to the suspension, which was then gently mixed and incubated for 5 minutes at room temperature. Fifty millilitres of chilled buffer P3 (Appendix I) was added to the lysed cells, which were then incubated on ice for 10 minutes. Separation of cosmid DNA from precipitated bacterial DNA was achieved by centrifugation at 20000x g (11000 rpm) in a Sorvall RC-5B centrifuge using a GSA rotor for 30 minutes. The cosmid DNA-containing supernatant was promptly transferred into a sterile GSA tube, re-centrifuged as before for 15 minutes or until supernatant was clear and the clear supernatant transferred into a clean tube immediately.

Gel Filtration Chromatography

Ten millilitres of buffer QBT (QIAGEN, Germany) (Appendix I) were used to equilibrate the QIAGEN-tip 500 before loading the clear DNA-containing supernatant (*cf.* cosmid DNA maxi/midi prep). Thereafter, the column was washed twice with 30ml buffer QC (Appendix I) to remove contaminants from the DNA. DNA elution was achieved by adding 15ml of buffer QF (Appendix I) to the column and the eluate collected in an SS34 tube. To precipitate the DNA, 10.5ml isopropanol (0.7 volumes, equilibrated at room temperature) were added to the eluate, which was immediately centrifuged at ≥15000 g in a Sorvall RC-5B centrifuge using a SS34 rotor for 30 minutes at 4°C. The supernatant was carefully removed and the DNA-containing pellet washed with 5ml ethanol (70%) by centrifugation at 15000xg in a Sorvall RC-5B centrifuge using an SS34 rotor, for 10 minutes at 4°C. Ethanol was decanted and the remnants blotted with paper towel. The pellet was air-dried for 5 minutes and resuspended in 200/50µl 1xTE (Appendix I) buffer (pH 8.0). Absorbance was read on a Milton Roy Spectrophotometer 1201 at 260nm to determine the DNA concentration. DNA was analysed on a 0.7 % agarose gel containing 0.01% EtBr and 1 x TBE to check its integrity (section 2.6.1). The DNA prepared at this stage was referred to as cosmid DNA and used as starting

material for dot blots (section 2.7.3.1), Southern blot (section 2.7.3.2), sub-cloning (section 2.7.3.3) and vectorette-PCR (section 2.7.7.1) manipulations that followed.

Plasmid DNA Miniprep (rplasmid)

In order to characterise recombinant plasmids that hybridised with the $(AAAT)_{10}$ (section 2.7.3.3), a single white colony was inoculated in 5ml LB/Amp/Tc selective medium in a 15ml propylene tube and incubated with shaking at 37°C overnight. The cells were centrifuged in a Sorvall RC-5B centrifuge using an SS34 rotor at 2500 rpm for 15 minutes at 4°C and the supernatant discarded. The pellet was resuspended in 200µl of GTE solution (Appendix I), containing lysozyme (5mg/ml) and transferred to a 1.5ml eppendorf tube. Four hundred microlitres of freshly prepared cell lysis NaOH/SDS solution (Appendix I) were added to this suspension, which was incubated for 5 minutes at room temperature. Thereafter, three hundred microlitres of ice-cold 3M KAc solution (Appendix I) were added to precipitate bacterial genomic DNA, the mixture was placed on ice for 5 minutes and then centrifuged in a microfuge (Beckmann, Germany) at 14000 rpm for 5 minutes.

The supernatant, containing rplasmid DNA, was transferred into a clean 1.5ml eppendorf tube and purified of protein and RNA contaminants by organic extraction in the following manner: 600µl phenol/chloroform (24:1) solution (Appendix I) was added to the supernatant, mixed by inversion and incubated at room temperature for 5 minutes prior to centrifugation at 14000 rpm for 5 minutes. The top aqueous phase containing recombinant plasmid DNA was transferred into a clean 1.5ml eppendorf tube, 300µl chloroform/isoamylalcohol was added, the mixture vortexed and then centrifuged as above. The top layer was again transferred to a clean 1.5ml-eppendorf tube and DNA was precipitated by addition of 600µl isopropanol and incubation at room temperature for 30 minutes. Thereafter, the contents of the tube were centrifuged in a microfuge (Beckmann, Germany) at 14000 rpm for 5 minutes and the supernatant was discarded. The pellet was washed with 200µl 70% ethanol by centrifugation at 14000 rpm for 5 minutes, lyophilised in a Speedy-Vac concentrator (Savant Instruments, USA) for 5 minutes and resuspended in 200µl TE (pH 8.0)/100µg/ml RNAseA. Sample analysis was done by electrophoresis on a 0.7% agarose gel (section 2.6.1).

2.7.3 Preparation of Membrane Filters

2.7.3.1 Dot blots

Dot blots were prepared as the initial screen to identify cosmids harbouring the $(AAAT)_n$ tandem repeat sequence, in order to avoid pursuing clones that did not contain repeat sequences. On a Hybond™ N^+ nucleic acid transfer membrane (Amersham, UK), the identification numbers of the different cosmids were marked with a pencil at appropriate positions. In an eppendorf tube, 200ng of cosmid DNA was denatured by the addition of an equal volume of 0.4N NaOH and incubation at room temperature for 5 minutes. The mixture was snap-cooled on ice for 2 minutes and spotted on the appropriately marked position on the Hybond™ N^+ filter, which had been placed on 3MM-chromatography paper (Whatman, England). Fifty nanograms of human genomic DNA was also spotted onto the filter as a positive control and ddH_2O as a negative control. The filter was air-dried for 15 minutes and baked for 2hrs in an 80°C vacuum oven (Towson and Mecer Ltd, England). Thereafter, the filter was either re-hydrated in 2x SSC (Appendix I) prior to the pre-hybridisation step of the screening process (section) or covered with cling film (GLADwrap™) and stored at 4°C until required.

2.7.3.2 Southern blots

The purified DNA of cosmid clones identified by dot blot hybridisation as harbouring the $(AAAT)_n$ tandem repeat was digested with *Sau3A*I (Amersham, UK), a four base pair frequent cutter, to generate clonable-sized fragments. The digestion reaction, which was incubated for 3hrs at 37°C, contained 5µg cosmid DNA, 15U *Sau3A*I (3U/µg DNA), 10x buffer as specified by the manufacturer (Amersham, UK) and ddH_2O to a final volume of 30µl. Following incubation, the total volume of the digestion reaction was loaded on a 0.7% agarose gel containing 0.01% EtBr and 1x TBE alongside a size marker (λDNA digested with *Pst*I (Boerhinger Manheim)) and electrophoresis carried out at 35V overnight.

The gel was visualised on longwave ultraviolet transilluminator (3UV™ Transilluminator model LMS-26E) with a fluorescent ruler aligned next to the gel to measure the distance migrated by the fragments from the wells and a photograph was taken using an ITC camera (Berkenhoff and Drebes, Germany) and a Sony video graphic printer (UP-860CE). Briefly, following electrophoretic separation of the digested cosmid DNA fragments, the gel was depurinated by submersion and gentle agitation in 0.2M HCl for 10 minutes. The HCl was decanted, the gel rinsed three times in ddH_2O before placing it on a platform covered with

3MM Whatman paper for alkaline transfer of DNA fragments from the gel onto a Hybond™ N^+ nucleic acid transfer membrane filter (Amersham, UK), following the manufacturer's protocol and hereafter referred to as a Southern blot.

Subsequently, the blotting apparatus was dismantled to remove the membrane filter, which was then placed between two sheets of 3MM Whatman paper (Whatman, England) and the DNA fixed by baking the filter for 2hrs at 80°C in a vacuum oven (Towson and Mercer Ltd, England). Thereafter, the Southern blot was either re-hydrated in 2x SSC (Appendix I) in preparation for probing with a radioactively labelled oligonucleotide (section 2.7.5) or stored at 4°C.

2.7.3.3 Colony blots

Plates containing recombinant clones (section 2.7.6.5), were placed at 4°C for at least 1hr after overnight incubation at 37°C and prior to transferring part of each colony, in duplicate, onto membrane filters. For each master plate, two pieces of Hybond N^+ (Amersham, UK) membrane were cut to the size of the petri plate and marked with a pencil as master or duplicate, for replica plating. The first or master filter, was carefully placed on the plate with the colonies for 1 minute and holes keyed in with a 18 gauge needle, to act as reference for the orientation of the filter on the plate, taking care not to make any lateral movements. The filter was carefully removed and placed, colony side up, on a fresh LB-agar/Amp/Tc plate. To make a duplicate filter, the second filter was carefully placed on top of the first and holes keyed in the same position as the first filter. The fresh plate with the sandwiched duplicate filters was incubated for 5hrs at 37°C, as was the original plate.

Following incubation, the filter sandwich was removed from the plate and the two halves kept together. Thereafter, the filters were consecutively treated with 10% SDS (Sigma, USA) for 3 minutes, denaturing solution (Appendix I) for 5 minutes and neutralising solution (Appendix I) for 5 minutes, in three separate containers layered with 3MM blotting paper (Whatman, England), which had been previously soaked, but not submerged in these solutions. The filters were then separated and transferred to a container with 2x SSC-soaked (Appendix I) 3MM blotting paper (Whatman, England) for a 5 minutes wash. Subsequently, the separated filters were transferred onto a clean sheet of blotting paper, where they were left to air-dry for 30 minutes prior to fixing the colonies by baking for 2hrs in a 80°C vacuum oven (Towson and Mercer Ltd, England). Thereafter, the filters were either re-hydrated in 2x SSC in preparation

for pre-hybridization (section 2.7.5) or covered with cling film (GLADwrap™) and stored at 4°C until required.

2.7.4 End-Labelling of $(AAAT)_{10}$ Oligonucleotides

Incorporation of Radioactivity

The $(AAAT)_n$ tetranucleotide sequence, repeated ten times $(AAAT)_{10}$, was end-labelled with γ-^{32}P[dATP] radioactive isotope and subsequently used as a hybridisation probe for the identification of $(AAAT)_n$ repeat motif-harbouring cosmid inserts on dot blot, Southern blot and colony lift membrane filters. The end-labelling reaction mixture contained 1µl of $(AAAT)_{10}$ oligonucleotide (50pmol), 10µl γ-^{32}P[dATP], 2µl T4 polynucleotide kinase (20 units) (Promega, USA), 5µl of 10x T4 polynucleotide kinase buffer (Promega, USA), and 32µl ddH_2O. The labelling mixture was incubated for 2hrs at 37°C.

Removal of Unincorporated Radioactivity

Gel filtration chromatography, Sephadex G-25, was used to separate the radiolabelled probe from the unincorporated γ-^{32}P[dATP]. The Sephadex G-25 column was prepared in the following manner: the bottom of a 1ml disposable syringe was sealed with sterile silanised glass wool. In the syringe, a Sephadex G-25 resin suspended in 1xTE (Appendix I) was loaded using a 1ml pipette until the bed volume was 0.9ml. The column was equilibrated with 50µl STE (pH 8.0) (Appendix I) by centrifugation at 16000g for 3 minutes in a swing-out rotor (Beckman model TJ-6 centrifuge) at room temperature. The column was washed by the addition of 50µl STE and centrifuged as before and this repeated 3 more times. The label mix (50µl) was applied to the column and centrifuged as before. Finally, 50µl STE (pH 8.0) was added to the column and the radiolabelled probe eluted by centrifugation at 16000g, for 3 minutes in a swingout rotor (Beckman model TJ-6 centrifuge).

Calculation of Incorporated Radioactivity

The amount of label incorporated was determined by trichloroacetic acid (TCA) precipitation. A mixture containing 1µl of the labelled oligonucleotide (2.4.2), 5µl 20x SSC and 4µl ddH_2O was prepared and 4.5 µl thereof spotted onto two pre-wetted, 1cm x 1cm, Hybond N^+ filters. The membranes were air-dried for 1 minute. One of the filters (representing total counts) was placed into glass vial and 5ml of scintillation fluid added. The second filter washed once with 10ml TCA (10%), to reduce non-specific binding of unincorporated nucleotides, and

subsequently washed five times with 10ml (5%) TCA under vacuum to collect the precipitated DNA. The washed filter (representing incorporated counts) was air-dried for 1 minute and placed in a glass vial and 5ml of scintillation fluid added. Both these vials were placed in a β-counter (Tri-carb 460 CD, Liquid scintillation system, Packard) and the percentage of incorporation calculated (Appendix I). Furthermore, radioactivity incorporated into the oligonucleotide probe, in counts per million, was calculated by multiplying the cpm of the washed filter with the total reaction volume of the labelled probe. The amount of radioactivity incorporated had to be at least $1x10^6$ cpm/ml of hybridisation solution.

2.7.5 Hybridisation of membrane filters with an end-labelled $(AAAT)_{10}$ probe

The filters (dot blot, Southern blot and colony blots) were re-hydrated in 2x SSC (Appendix I) for 5 minutes. Filter-stripping of previously used (probed) blots was performed by pre-washing at 50°C with 5x SSC, 0.5% SDS and 1mM EDTA solution for 30 minutes before pre-hybridisation. All hybridisation experiments were performed at an empirically determined temperature of 42°C.

Pre-hybridisation

Subsequent to pre-wetting or pre-washing of previously probed blots, the filters were placed in a heat-sealable plastic bag, with only one end open. Twenty millilitres of pre-warmed (42°C) pre-hybridisation liquid (Appendix I) was poured into the bag and air bubbles were removed before heat-sealing and incubation in a shaking waterbath (Heto Lab equipment, Denmark) at 42°C for 6hrs. Thereafter, one corner of the bag was cut open and the pre-hybridisation fluid discarded before adding the hybridisation mix containing the radio-labelled probe.

Hybridisation

Twenty millilitres of fresh hybridisation solution (Appendix I), containing the radioactivity labelled probe, were added to the bag, air bubbles squeezed out, the bag re-sealed and incubated at 42°C, in a shaking waterbath (Heto Lab eqiupment, Denmark) overnight. After incubation the hybridisation solution was discarded and the filters washed.

Post-hybridisation washes

The filters (section 2.7 3) were submerged in a flat-bottomed plastic box containing a solution of 2x SSC- 0.1% SDS (Appendix I) and incubated for 30 minutes at room temperature with

gentle shaking. This washing step was repeated 4 times, adding fresh 2x SSC-0.1% SDS solution for each wash. The filters were dried at room temperature on a sheet of 3MM-chromatography paper (Whatman), sealed in a plastic bag and exposed to an X-ray film (Cronex-4, Protea Medical) for an autoradiographic image.

2.7.6 Subcloning of $(AAAT)_n$-containing fragments

For subcloning purposes, incompatible vector and insert DNA molecules were made compatible by partial fill-in reaction. This reaction entails, first, digesting each of the molecules with a different restriction endonuclease generating 5' protruding ends, which are incompatible. Subsequently, the incompatible 5' single stranded overhangs are partially filled-in with complementary dNTPs to generate compatible cohesive termini.

2.7.6.1 Preparation of insert DNA fragment for cloning

Sau3A1 restriction enzyme digestion of cosmid DNA

Cosmids identified by Southern blot hybridisation as containing positively hybridising fragments were digested with *Sau3A*I (Amersham, UK) on a large scale to generate 5'-overhangs on the insert DNA fragments. The digestion reaction, containing **25μg** cosmid DNA, 75U *Sau3A*I (Amersham, UK), 10x buffer as specified by the manufacturers, ddH_2O to 50μl was incubated at 37°C for 3 hrs. The digested fragments were separated on a 1x TBE/0.7% agarose preparative gel, containing 0.01% EtBr and 1x TBE, by electrophoresis carried out at 35V overnight. The gel was visualised on an UV-transilluminator and the band of interest identified by comparing its molecular weight to that of the fragment identified on the Southern blot autoradiograph, was excised with a sterile blade and placed in a 2ml eppendorf tube. The QIAquick gel extraction kit (QIAGEN, Germany) was used to purify the DNA fragment, for sub-cloning, from the gel slice according to the manufacturer's protocol.

2.7.6.2 Preparation of vector DNA

***Sal*I restriction enzyme digestion of pBluescript SK DNA**

The plasmid vector, pBluescript, molecular weight 2.9kb (Figure 2.2), was linearised by digestion with *Sal*I restriction enzyme (Promega, USA). The digestion reaction mixture contained 10μg vector DNA, 4μl (100mM) spermadine, 200 U *Sal*I (Promega. USA), 10μl 10x buffer as specified by the manufacturers (Promega, USA) and ddH_2O to a final volume of 100μl. The reaction mixture was incubated at 37°C for 3hrs. To verify that the vector had been linearised, 5μl of digestion reaction mixture was analysed on a 0.7% agarose gel (section

2.4.3), with an equal volume of an uncut vector sample and $^{\lambda}$*Pst*I size marker in adjacent lanes. Thereafter, vector DNA in the remaining sample was precipitated by the addition of 250µl 100% ethanol, hereafter, the mixture was vortexed briefly and incubated for 30 minutes at -80°C, prior to centrifugation in a microfuge (Beckmann, Germany) at 14000 rpm for 15 minutes at room temperature. The resulting supernatant was carefully decanted and the pellet washed twice with 200µl 70% ethanol, by centrifugation in a microfuge (Beckmann, Germany) at 14000 rpm for 10 minutes. The ethanol was decanted and the pellet lyophilised in a Speedy-Vac® concentrator (Savant Instruments, USA) for 5 minutes and resuspended in

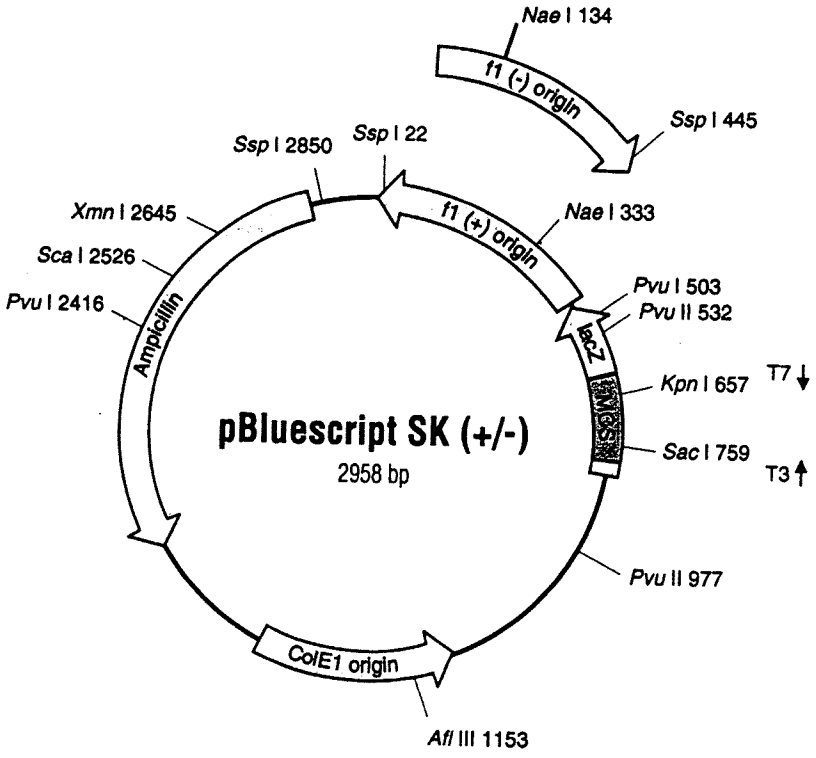


Figure 2.2a: The cloning vector pBluescript SK (+/-), with the positions of the MCS and the T3 and T7 primers indicated (Stratagene Catalog, 1999).

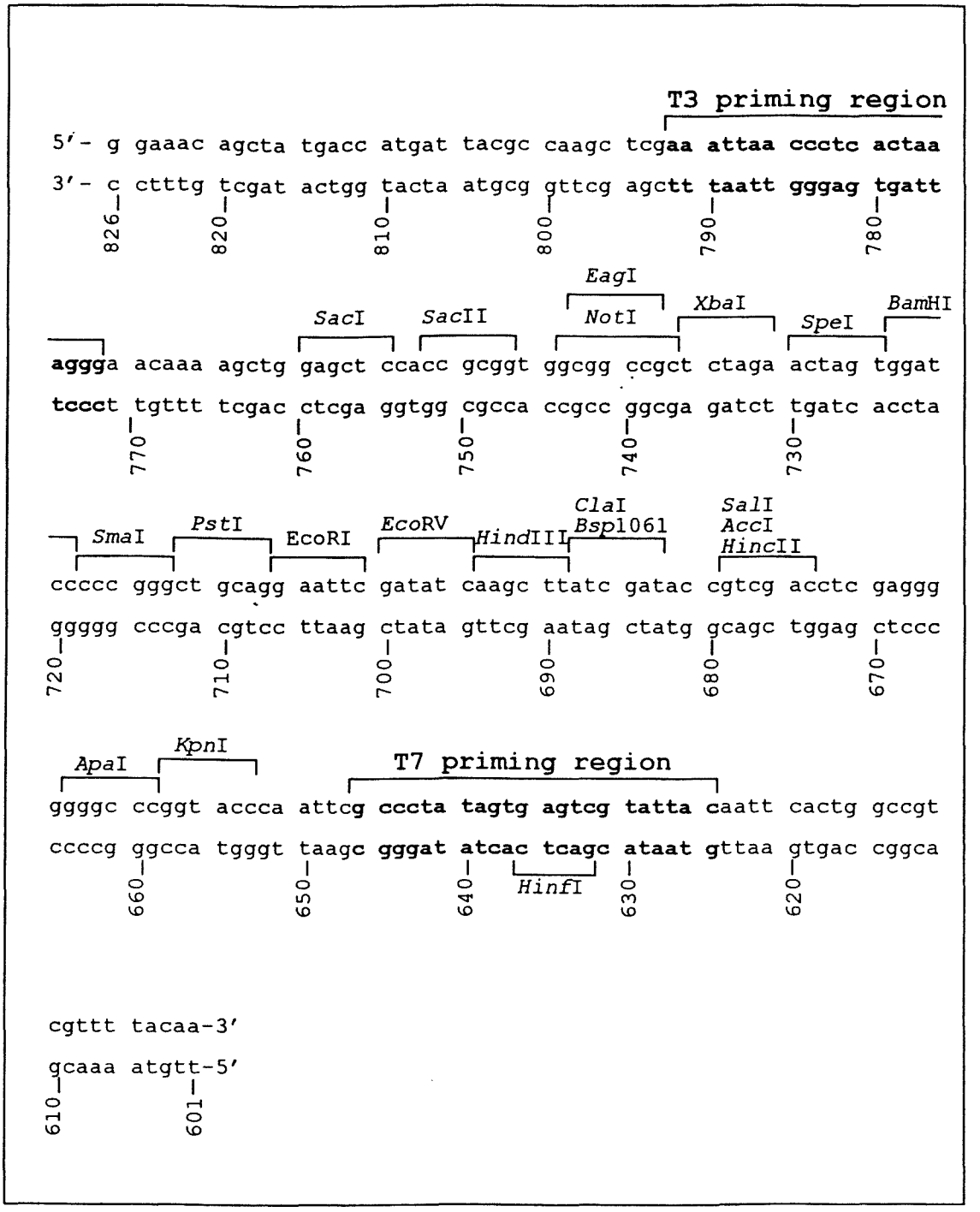


Figure 2.2b pBluescript vector multiple cloning site (826 bp to 601 bp)

759 bp to 657 bp: polylinker site (relevant enzyme recognition sequences shown)
792 bp to 772 bp: T3 priming position
625 bp to 646 bp: T7 priming position

40μl ddH_2O. As the generated 5'-overhanging ends, generated by *Sal*I digestion of plasmid vector and *Sau3A*1 digestion of insert DNA were incompatible to effect ligation, patrial fill-in reactions were first performed for both insert and vector DNA.

2.7.6.3 Partial fill-in reaction of 5'-overhanging ends of insert and vector DNA

Insert DNA had *Sau3A*I generated cohesive termini (5'-GATC-3'), in contrast to vector DNA which had *Sal*I cohesive termini (5'-TCGAC-3'). Thus, filling in the single stranded extensions with dATP and dGTP for the insert and dTTP and dCTP for the vector would generate compatible cohesive termini thus allowing ligation of the two molecules (Figure 2.3). The fill-in reaction mixture consisted of 40μl of insert DNA, 0.5μl dGTP (10mM), 0.5μl dATP (10mM), 2μl (10U) Klenow fragment of DNA polymerase (Promega USA), 5μl 10x Klenow buffer and ddH_2O to 50μl. The same reaction mixture was prepared for the vector, using pBluescript SK DNA and dTTP and dCTP nucleotides. Incubation was carried out at 37°C for 2hrs, after which 1μl (0.5M) EDTA was added to quench each reaction.

Organic Extraction

Subsequent to partial fill-in reactions, both vector and insert were purified by organic extraction. To each partially filled-in DNA molecule (50μl reaction mixture), 50μl TE (pH 8.0), 0.1M NaCl (1.5μl of 0.5M) and 100μl phenol/chloroform (24:1) solution were added with careful mixing, after which the both samples was centrifuged in a microfuge (Beckmann, Germany) at 14000 rpm for 5 minutes. The DNA-containing top aqueous phase was transferred into a clean 1.5ml eppendorf tube, taking care not to disturb the interface. A 100μl aliquot of chloroform/isoamylalcohol (Appendix) was added, the sample vortexed for 15 seconds and then centrifuged as before. The top layer was carefully transferred to a clean 1.5ml eppendorf tube, after which 300μl 100% ethanol was added and the sample incubated at -80°C for 15 minutes, prior to centrifugation in a microfuge (Beckmann, Germany) at 14000 rpm for 15 minutes and the supernatant carefully removed. Both insert DNA- and vector DNA-containing pellets were washed with 200μl 70% ethanol by centrifugation in a microfuge (Beckmann, Germany) at 14000rpm for 10 minutes at room temperature, lyophilised in a Speedy-Vac® concentrator (Savant Instruments, USA) for 5 minutes and

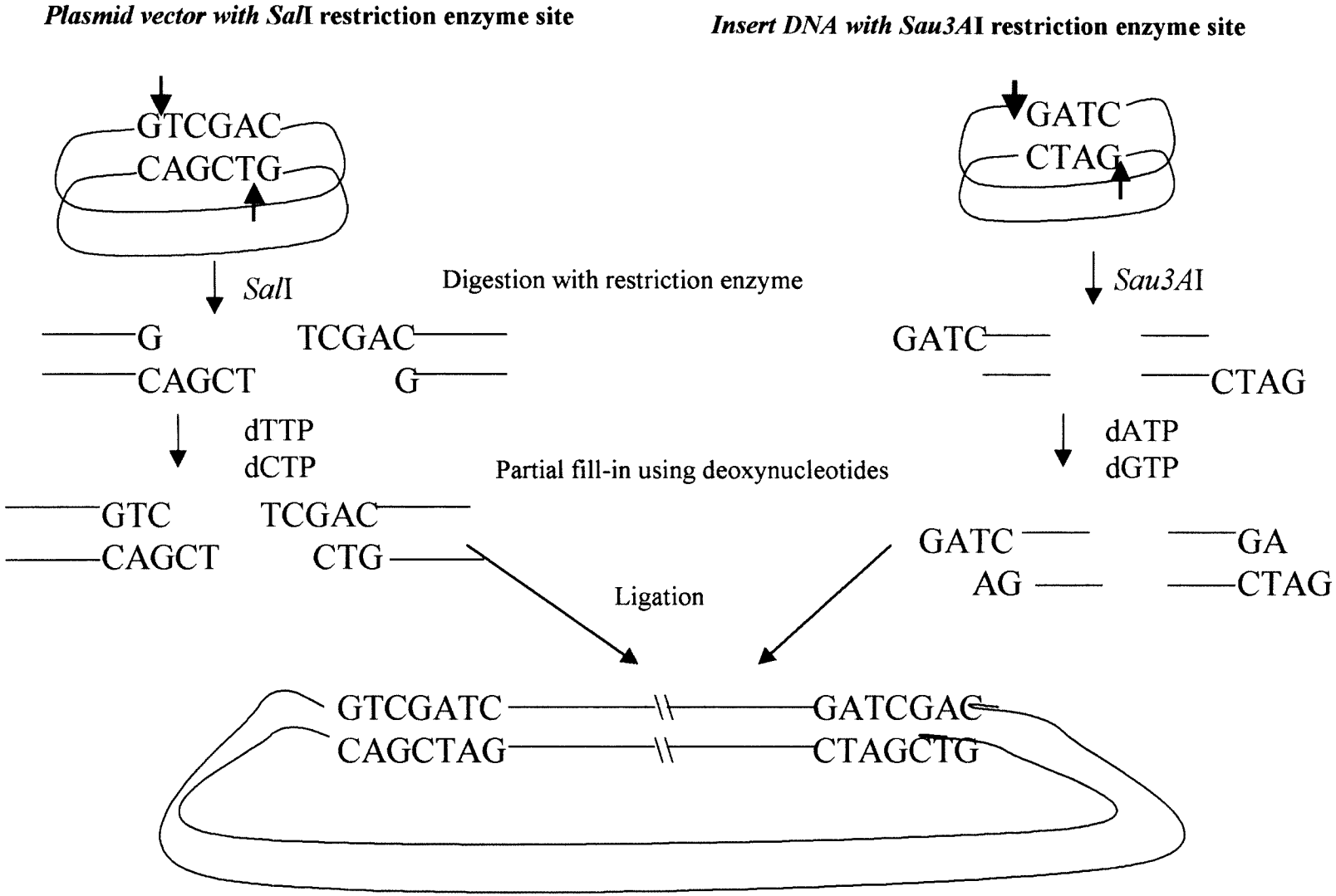


Figure 2.3: Schematic representation of partial fill-in reactions to create compatible ends in preparation for ligation of insert DNA to linearised plasmid vector pBluescript.

each resuspended in 40µl ddH2O to give a final concentration of 200ng/µl and 100ng/µl for insert and vector DNA, respectively.

2.7.6.4 Ligation

The partial fill-in reactions created compatible cohesive termini in preparation for ligation of insert to vector to form a chimaeric molecule. Serial dilutions were made to determine the ideal dilution which would give the 2:1 insert:vector ratio required for optimising ligation (Table 2.3). Controls reactions included exclusion of: i) insert DNA to determine the number of false recombinants formed as a result of the vector re-ligating to itself, i.e. establish the efficiency of the partial fill-in reactions. ii) vector DNA was excluded to assess the viability of the of the E. *coli* XL1 Blue strain competent cells. The ligation reactions were incubated at 16°C overnight in a Endocal refrigerated circulating bath (Neslab) and thereafter, stored at 4° until required for transformation (2.7.6.5).

2.7.6.5 Transformation

Preparation of competent cells

The XL1-blue strain of *E.coli* was streaked from glycerol stocks onto LB/agar/tetracycline (Tc, 15µg/ml) (Appendix I) plates and incubated at 37°C overnight. From these plates, one colony was inoculated into 2ml LB/Tc broth and incubated overnight at 37°C, with vigorous shaking. Four hundred microlitres of the overnight culture were added to 40ml LB/Tc (15µg/ml) medium to give a 1/100 dilution, incubated at 37°C with vigorous shaking for approximately 2hrs, or until culture absorbance measured at 600nm was between 0.2-0.4 OD units. Ten millilitres of the culture, were added to each of two 15ml propylene tubes, which were centrifuged in a Sorvall RC-5B centrifuge using an SS34 rotor at 4000g for 8 minutes at 4°C, after which the supernatant was discarded. The pellet was resuspended in 5ml pre-chilled CAP (Appendix I) solution, which was then incubated on ice for 30 minutes, before centrifuging as before for 6 minutes. The supernatant was removed and the cells gently resuspended in 1ml ice-cold CAP. The cells were transformed with the ligation cocktail containing chimeric DNA molecules (section 2.7.6.4) as described below or stored in 50µl or 100µl aliquots at -70°C.

Transformation of competent cells

Five microlitres of ligation mixture (section 2.7.6.4) were added to pre-chilled 15ml propylene tubes, followed by the addition of a 50µl aliquot of competent cells and the

***Table 2.3**: Ligation reactions of insert DNA to pBluescript vector.*

Insert	Vector	ATP	10xD	10xB	Ligase	ddH_2O
1. 1µl (1x)	1µl	1µl	1µl	1µl	1µl	4µl
2. 1µl (10^{-1})	1µl	1µl	1µl	1µl	1µl	4µl
3. 1µl (10^{-2})	1µl	1µl	1µl	1µl	1µl	4µl
4. 1µl (10^{-3})	1µl	1µl	1µl	1µl	1µl	4µl
5. 0	1µl	1µl	1µl	1µl	1µl	5µl
6. 1µl (1x)	0	1µl	1µl	1µl	1µl	5µl

Reactions 1-4 = serial dilutions to determine the ideal (2:1) insert to vector ratio for optimisation of ligation of insert DNA to plasmid vector.
Reaction 5 = Control to check for the efficiency of the partial fill-in reactions of vector DNA.
Reaction 6 = control to check for competency of E. coli *XL1 Blue strain.*

resulting mixture incubated on ice for 30 minutes. Thereafter, the cells were heat-shocked by incubation in a waterbath at 42°C for exactly 45 seconds and were then placed on ice for 1 minute. Five hundred microlitres of LB, pre-warmed to 42°C, was added to the transformation mixture and the cells incubated at 37°C with shaking for 1-2hrs to allow phenotypic expression before antibiotic challenge.

Analysis of Transformants

Antibiotic (Amp and Tc) selection and blue-white colour screening using IPTG and X-Gal were used for the isolation and identification of transformants, respectively. Antibiotic selection would allow only those cells (Tc resistant) containing the chimeric plasmid vector (Amp resistant) to grow. In addition, the plasmid vector has a multiple cloning site (MCS) within the *lacZ* gene (Figure 2.2) which, when induced by the inducer analogue, IPTG, will be expressed, giving the β-galactosidase product which reacts with X-Gal by breaking the link between the two components giving a blue colour. However, in the transformed cells containing the recombinant plasmid, the *lacZ* gene is interrupted, thus it is not expressed, and the transformants grow as white colonies. Fifty microlitres each of IPTG (20mg/ml) and X-Gal (20mg/ml) were spread on LB/agar/Amp/Tc plates using sterile glass balls and allowed to dry for ~30 minutes. Two hundred microlitres of the transformation mixture were then spread on these plates, again using sterile glass balls and the plates incubated at 37°C overnight. Following incubation, the colony plates were placed at 4°C prior to the preparation of colony lifts, referred to as colony blots (2.7.3.3).

2.7.6.6 Identification and Characterisation of AAAT repeat motif

To verify that the white colonies picked from the plates contained cloned insert, recombinant plasmid DNA was extracted (section 2.7.2) and digested with *Kpn*I and *Sac*I vector restriction enzyme sites on either side of the MCS (Figure 2.2). This digestion of recombinant plasmid DNA would result in the release of insert from the vector. The digestion reaction contained 5μl rplasmid DNA, 3-5U *Kpn*I (Promega, USA), 3-5U *Sac*I (Promega, USA), 1μl 10x buffer according to the manufacturer's (Promega, USA) specifications and dH_2O to 10μl. The reaction mixture was incubated at 37°C for 2hrs. Thereafter, 5μl of sample was withdrawn, 2μl loading dye added and the sample loaded on a 0.7% agarose gel (section 2.6.1) for analysis of digest reaction products. The DNA size marker λPst1 was electrophoresed alongside the sample, as well as an undigested recombinant plasmid. Following the

identification of insert, the corresponding recombinant plasmid clones were sequenced (section 2.5.3).

2.7.7 Vectorette-based PCR amplification

Vectorette PCR is a linker-specific amplification technique originally used for the isolation of terminus-specific sequences from large cloned DNA inserts (Riley *et al.*, 1990). The modified procedure used in this study relies on three PCR-amplification reactions to obtain sequences flanking a known sequence, in this case, an $(AAAT)_n$ repeat motif. First, a repeat-anchored primer is used together with a vectorette-specific primer to identify, in a vectorette library, a fragment harbouring the repeat motif of interest, which is then sequenced to obtain sequence flanking one side of the repeat motif. This acquired sequence is used to design a primer, which is then used together with the vectorette-specific primer to amplify the sequence flanking the other end of the repeat. Subsequently, primers generated from the repeat-flanking sequences are used for routine PCR amplification and subsequent analysis of the repeat locus (Lench *et al.*, 1996).

2.7.7.1 Construction of vectorette libraries

Preparation of insert

Five separate vectorette libraries were constructed for each selected cosmid clone (Figure 2.4) using the following restriction enzymes: *Alu*I (Boerhinger Mannheim, Germany), *EcoR*V (Promega, USA), *Pvu*II (Boerhinger Mannheim, Germany), *Rsa*I (Boerhinger Mannheim, Germany) (blunt end cutters) and *Sau3A*1 (Boerhinger Mannheim, Germany) (cohesive end cutter). A range of different restriction enzymes was used to increase the chance that the restriction fragment containing the microsatellite repeat was of an amplifiable size (Lench *et al.*, 1996). Each reaction mixture contained 1µl DNA, 10x buffer (as specified by the manufacturer), 20U restriction endonuclease and dH_2O to 50µl. The reactions were incubated at 37°C overnight and subsequently heat inactivated by incubation at 65°C for 10 minutes. The digestion reactions were stored at -20°C until required for use in ligation to the vectorette linkers (described below).

Preparation of vectorette cassettes

Vectorette cassettes or linkers were prepared by mixing 1pmol of either the top strand with blunt (V_{TB}) or cohesive (V_{TC}) ends together with equal amounts of the bottom strand (V_B) oligonucleotides in a 1.5ml eppendorf tube. The mixture was then incubated at 37°C to

facilitate annealing of the top and bottom strands, thus forming a double strand molecule, the vectorette linkers (Figure 2.4).

Top strand linkers

Blunt end: 5' AAG GAG AGG ACG CTG TCT GTC GAA GGT AAG GAA CGG ACG AGA GAA GGG AGA G 3'

Cohesive end: 5' GAT CAA GGA GAG GAC GCT GTC TGT CGA AGG TAA GGA ACG GAC GAG AGA AGG GAG AG 3'

Bottom strand linker

Blunt end: 5' CTC TCC CTT CTC GAA TCG TAA CCG TTC GTA CGA GAA TCG CTG TCC TCT CCT T 3'

The sequence in green represents the non-homologous region of the linkers creating a "bubble" structure on annealing of either of the top linkers to the bottom linker. The sequence in ***black*** *is homologous and forms a double strand region on either end of the bubble, thus creating a vectorette cassette. The 5'-overhanging end sequence, generating a cohesive end, is indicated in blue.*

Ligation

For each restriction enzyme digestion of insert prepared, a cocktail mixture containing 2µl vectorette cassettes (1pmol/µl), 4µl 10mM ATP, 1µl 10x "one-Phor-all" buffer (Pharmaci Biotech, USA), 0.5µl T4 DNA ligase (1U/µl) (Promega, USA) and ddH_2O to 10µl was prepared. Ten microlitres of the cocktail mixture was then added to 40µl of each insert digestion reaction, aliquoted into a 1.5ml eppendorf tube and incubated at 37°C for 2 hours. Following incubation, 200µl of ddH_2O was added to the ligation mixture, henceforth referred to vectorette libraries (Figure 2.4), mixed by vortexing and stored at -20°C.

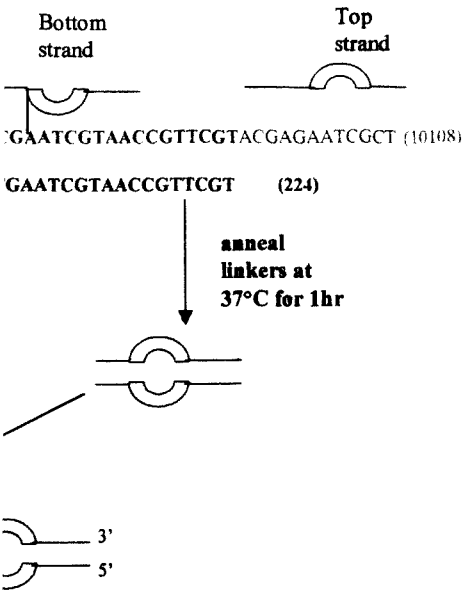
2.7.7.2 Identification and characterisation of $(AAAT)_n$ repeat motif

Each vectorette library was PCR-amplified using one of a panel of four anchored primers, namely **$(AAAT)_4A$, $(AAAT)_4T$, $(AAAT)_4G$, $(AAAT)_4C$** and the universal primer 224 (complementary to the bottom strand, V_B of the vectorette linkers) (Figure 2.4). Products can only be synthesised from primer 224 if an initial round of synthesis has taken place from the anchored primer producing a complementary strand. In this way, specific amplification products can be synthesised from the insert repeat sequence (Riley *et al.*, 1990). PCR amplification products were analysed on a 2% agarose gel (section 2.6.1) prior to sequencing of generated template.

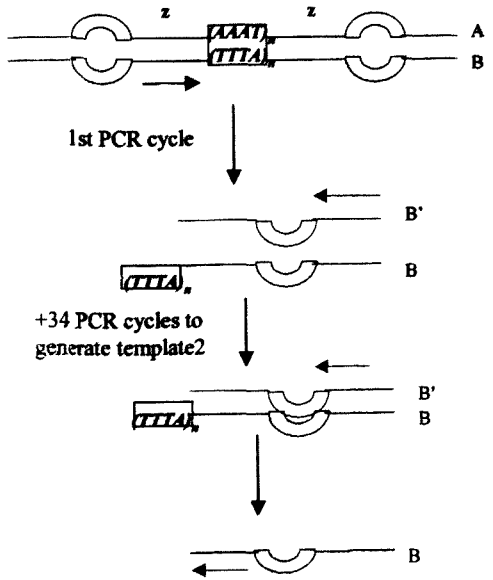
Figure 2.4: Schematic representation of the vectorette PCR-based strategy used to identify and characterise the $(AAAT)_n$ repeat motif harboured in cosmids for the development of genetic polymorphic markers.

A = Generation of cosmid vectorette libraries using restriction enzyme-digested cosmid DNA and vectorette linkers. B = PCR-amplification of the vectorette libraries using a repeat-anchored primer and a vectorette PCR primer (1) and the subsequent amplification to identify and characterise the harboured repeat.. The positions of the PCR-amplification amd sequencing primers are indicated by colour codes: ***Repeat-anchored primer****, vectorette PCR primer 224, vectorette sequencing primer 10108 and the repeat-flanking primer pair* ***F*** *and* ***R****.*

Vectorette linkers



PCR-amplification 2

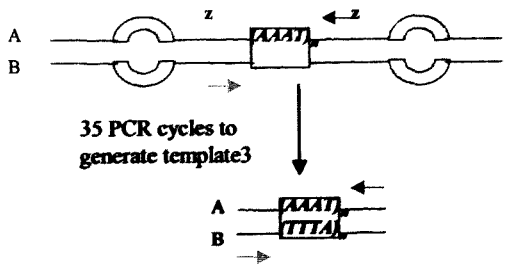


PCR amplification of vectorette library using a putative repeat-flanking primer F **and** vectorette primer 224 to generate **template 2.**

Template 2 was sequenced with primer 10108 and the information used to design the putative repeat-flanking primer R

The putative repeat-flanking primer R will anneal to the strand complementary to the one from which it was designed.

PCR-amplification 3



PCR amplification of vectorette library using a putative repeat-flanking primer pair F **and** R to generate **template 3.**

PCR amplification of vectorette library using a repeat-anchored primer and vectorette primer

Sequencing of template 3 using the putative repeat-flanking primer pair F **and** R **to identify**

2.8 MUTATION SCREENING OF CANDIDATE GENES

2.8.1 Glycogen Synthase (GSY1)

2.8.1.1 Selection of exons for mutation screening

The muscle glycogen synthase gene orthologues, *GSY1*, nucleotide sequences from human muscle, rat and mouse, the liver paralogues, *GSY2*, from human, rat, and the two yeast homologues, *GSYI* and *GSYII*, were obtained from the Genome database (GDB, http://www.gdb.org). These sequences were transcribed and translated (using Swiss Prot) into amino acid sequences. The degree of homology between the specific exons of *GSY1* and those of its paralogues in human, rat, mouse and yeast as well as the liver orthologues in human, rat and yeast were analysed using BLAST homology search. In addition, to identify highly conserved protein coding sequences, the above-mentioned DNA sequences were translated using Swiss Prot and the protein sequence homology analysed by CLUSTAL V, a multiple sequence alignment program. A selection of the most conserved coding region (exons 4, 5, 11, 12 and 16) was then screened for PFHBI-causing mutations by PCR-SSCP analysis.

2.8.1.2 PCR-SSCP Analysis

In order to investigate the selected exons of *GSY1* for any nucleotide sequence change, PCR SSCP analysis was performed as follows. Three to five microlitres of a non-radioactively amplified PCR product was aliquoted into a 1.5ml eppendorf centrifugation tube and an equal volume of SSCP loading dye (Appendix) added. In the case of RE-based PCR-SSCP analysis, 8μl of the precipitated restriction enzyme digested PCR product (section 2.8.1.3) was used. The mixture was heat-denatured at 95°C for 2 minutes, prior to loading the each sample into the wells of a non-denaturing polyacrylamide gel. Generally, four different gel conditions (Appendix) were used for optimum SSCP screening of each panel. In the event that a mobility shift was observed, 0.35xMDE^{TM} (FMC Bioproducts, USA) gels were used to confirm the results. Control samples which were co-electrophoresed included a non-denatured sample as control for the electrophoretic migration of the double-stranded DNA, amplified samples from unaffected individuals (close relatives) as negative controls and $^{\lambda}$*Pst*I as molecular size marker. Electrophoresis was carried out in 0.5x TBE (Appendix I) buffer, at a constant power of 50W for 2-7hrs at 4°C, depending on the concentration of gel used.

2.8.1.3 Restriction enzyme-based PCR-SSCP analysis

The DNA analysis program GENEPRO was used to identify restriction enzymes sites in *GSYI* exons 4 and 16 that were 361bp and 379bp in size, respectively, that on digestion would generate ideal size fragments (150bp-220bp) for analysis by PCR-SSCP. *Hinf*I (Boerhinger Mannheim, Germany) and *Hpa*II (Boerhinger Mannheim, Germany) restriction endonucleases were selected for exon 4 and for exon 16 *Mbo*II (Amersham, UK) and *Sty*I (Biolab, UK) restriction enzymes were ideal. The restriction enzyme digestion reaction mixture contained 13μl PCR-product, 3-5U restriction enzyme (Appendix I) with the 10x buffer as specified by the manufacturers and ddH_2O to 25μl and was incubated at 37°C overnight.

Verification of Restriction Enzyme Digestion

The digestion reaction was verified by analysing 5μl of sample mixed with 2μl loading dye, on a 12% non-denaturing polyacrylamide gel (section 2.6.2). Electrophoresis was carried out in 1x TBE buffer at 150V for 2hrs. To visualise the bands, the gel was silver-stained as described in section 2.6.2.

Precipitation of digested sample

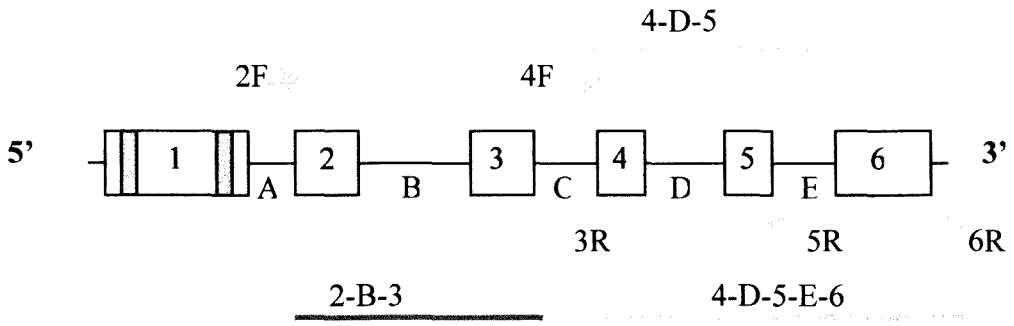
The remaining 20μl of digested sample was transferred to a clean 1.5ml-eppendorf centrifugation tube and precipitated using 10μl ammonium acetate and 60μl 100% EtOH. The sample was mixed well and centrifuged in a microfuge (Beckmann, Germany), at 14000 rpm, for 30 minutes at room temperature. The supernatant was decanted and discarded and the resulting pellet washed, twice, with 200μl 70% ethanol by centrifugation in a benchtop microfuge (Beckmann, Germany) for 5 minutes at 14000 rpm. Ethanol was decanted, the pellet lyophilised in a Speedy-Vac® concentrator (Savant Instruments, USA) and resuspended in 13μl ddH_2O. An 8μl aliquot of the precipitated restriction enzyme digested PCR product was then analysed for SSCPs on non-denaturing (section 2.6.2).

2.8.2 Histidine-rich calcium binding protein (HRC)

The *HRC* gene has 6 exons, whereby the amino (N) terminal is coded for by exon 1 and the conserved carboxy (C) terminal coded for by exons 2-6 (Hofmann et al. 1991). Based on the premise that the conserved regions in protein sequences are functionally important, the C-terminal encoding exons (2-6) were selected for mutation analysis. Primer sequences were designed using both intronic and exonic sequences to amplify the regions shown in Figure 2.5a (see Table 1 for the specific amplification conditions). The resulting products were screened for mutations by direct sequencing. However, as parts of exonic sequences were

used together with the published intronic sequences to design primers, complete exonic screening was hindered. Consequently, the intronic sequence obtained from the initial round of sequencing was used to design another set of primers that would facilitate screening of the previously missed intron/exon junctions (Figure 2.5b). Observed sequence variations were confirmed by restriction enzyme digestion, if they affected a restriction enzyme recognition site or by PCR-SSCP.

A



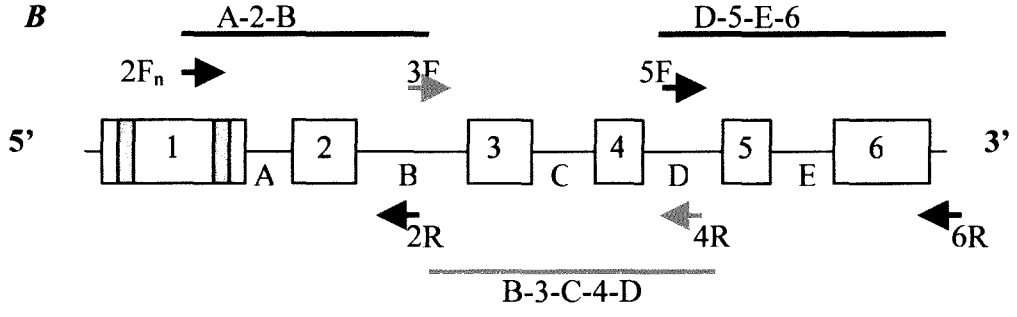


Figure 2.5: Schematic representation of the* HRC *gene illustrating the intron/exon organisation (not drawn to scale).

A = Mutation screening of HRC*'s exons 2-6, indicating primer positions and orientation for PCR amplification and the sequenced PCR product.*

B = Mutation screening in HRC *intron/exon junctions using primers designed from intronic sequence obtained in A, above.*

Gene structure: rectangles = exons, lines = introns. Arrows (colour-coded) = primer pairs used for amplification and the amplified product indicated by the same colour line.

CHAPTER 3
RESULTS

3.1 Development of $(AAAT)_n$ Polymorphic Markers

The PFHBI locus has been fine-mapped to a 7cM interval on chromosome 19q13.3 between markers *D19S412* and *D19S866* (Figure 1.5). Therefore, new markers within this region were required, specifically between the sub-interval *D19S412* and *D19S866,* in order redefine the centromeric and telomeric limits, respectively, thereby providing a smaller and more focussed gene search area. To this end, cosmids within these two target regions were selected for use in developing novel $(AAAT)_n$ STR genetic markers.

The identification and characterisation of $(AAAT)_n$ STRs involved, firstly, screening cosmids containing human chromosome 19q13.3 inserts, with a radioactively labelled $(AAAT)_{10}$ probe. The cosmids were screened both by dot blot hybridisation of undigested DNA and Southern blot hybridisation of *Sau3A*1-digested DNA. Following the identification of clones harbouring the repeat sequence, two different strategies, one cloning-based (3.1.2.1) and the other vectorette PCR-based (3.1.2.2), were applied to reveal the repeat and the unique repeat-flanking sequences, which were subsequently used to design PCR primers for genotyping the $(AAAT)_n$ repeat locus in human genomic DNA.

3.1.1 Identification of $(AAAT)_n$-harbouring Cosmid Clones by Dot Blot and Southern Blot Hybridisation

Dot Blot Hybridisation

Of the 25 cosmids containing C19q13.3-spanning inserts, 19 were spotted onto a membrane filter and hybridised with the radioactively labelled $(AAAT)_{10}$ probe. The remaining six clones did not yield any cultures and thus could not be pursued. A total of sixteen of these clones gave positive signals, of varying intensities, for the presence of the repeat (Figure 3.1). Of these 16 clones, one clone gave a very strong signal (24680), four clones gave strong positive signals (10501, 22363, 23078 and 24493) and eleven clones gave positive signals (8946, 15217, 18073, 18618, 19161, 20381, 22135, 24596, 25235, 27268 and 29395). The three remaining clones gave ambiguous signals (14187, 14960 and 15743).

Southern Blot Hybridisation

Of the 16 cosmids, which gave positive signals on dot blot hybridisation, 11 were available as *Sau*3A1-digested DNA on a Southern blot membrane (8946, 15217, 19161, 20381, 23078,

24493, 24596, 24680, 25235, 27268 and 29395). Cosmids 10501, 22363, 18073, 18618 and 22135 were not available on this Southern blot, but two additional cosmids, 27993 and 29025 had been included. After hybridisation with the radiolabelled $(AAAT)_{10}$ probe, fragments generated from eight of the digested cosmids gave strong positive signals (15217, 20381, 23078, 24493, 24596, 25235, 27268 and 29395) (Figure 3.2). Weak positive signals were seen in with fragments from cosmids 19161 and 29025. Cosmid 8946 gave a very weak signal and no signals were detected in cosmids 24680 and 27993 (Figure 3.2).

Of the 11 clones used in both the dot blot (Figure 3.1) and the Southern blot (Figure 3.2), nine (15217, 19161, 20381, 23078, 24493, 24596, 25235, 27268 and 29395) gave clear positive results on both dot blot and Southern blot hybridisation. For most cosmids on the Southern blot, the $(AAAT)_{10}$ radiolabelled probe hybridised to more than one fragment, although with varying signal intensities. For example cosmid 20381, 27268 and 29395 (Figure 3.2, lanes 6, 8 and 9) each had an intense and a very faint signal, cosmid 25235 (Figure 3.2, lane 2) had a single intense signal while cosmid 23078 (Figure 3.2, lane 1) gave multiple high intensity signals. Four of the cosmids that gave clear positive results, 20381, 24493, 27268 and 29395 (Figure 3.2, lanes 6, 12, 8 and 9) lay within the targeted intervals for development of polymorphic markers (Figure 1.4). Based on the results obtained from hybridisation of the cosmids with the $(AAAT)_{10}$ probe (Figures 3.2 and 3.3), and their relative position in the *PFHBI* target area (Figure 1.4), the following three cosmids, 20381, 24493 and 29395 (Figure 3.2, lanes 6, 12 and 9), were selected and used for development of $(AAAT)_n$ polymorphic markers.

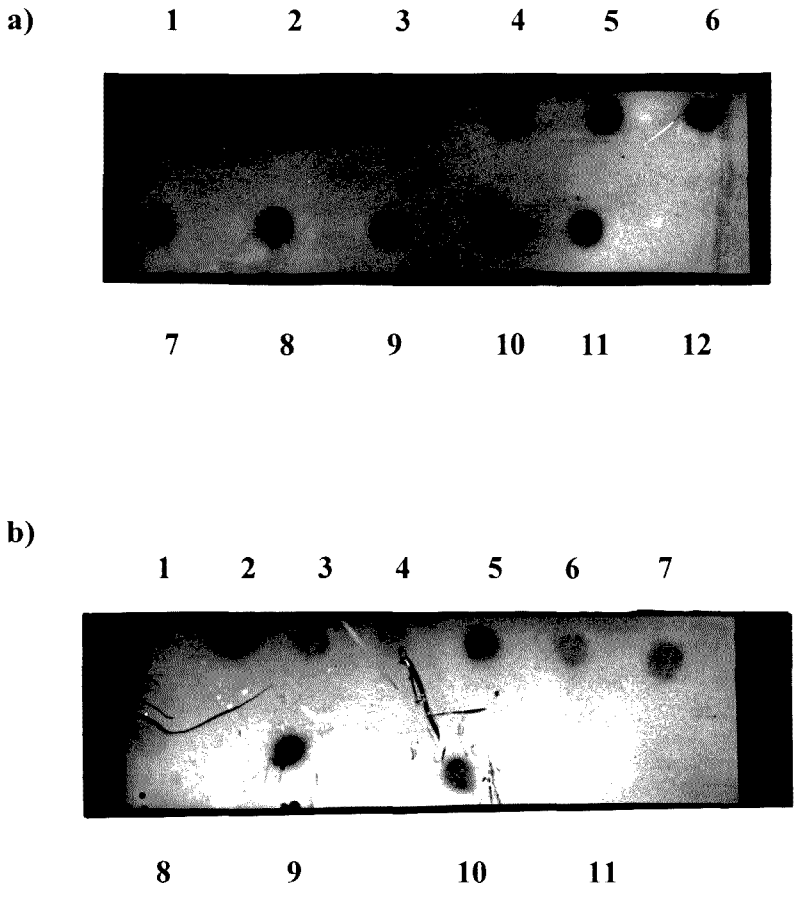


Figure 3.1: Detection of AAAT STR-harbouring chromosome 19q13.3-spanning cosmid clones.
Dot blot analysis of denatured cosmid DNA samples spotted onto a membrane filter and probed with the [γ-^{32}P]dATP labelled $(AAAT)_{10}$ oligonucleotide.

a)

Lane:	*1. 27268*	*7.15217*
	2. 25235	*8. 18618*
	3. 24493	*9. 19161*
	4. 24680	*10. 8946*
	5. 24596	*11. 20381*
	6. +ve control	*12.-ve control*

+ve control = human genomic DNA,

b)

1. 29395	*7. +ve control*
2. 23078	*8. 14187*
3. 15743	*9. 22363*
4. 14960	*10. 10501*
5. 22135	*11. –ve control*
6. 18073	

-ve control = ddH_2O.

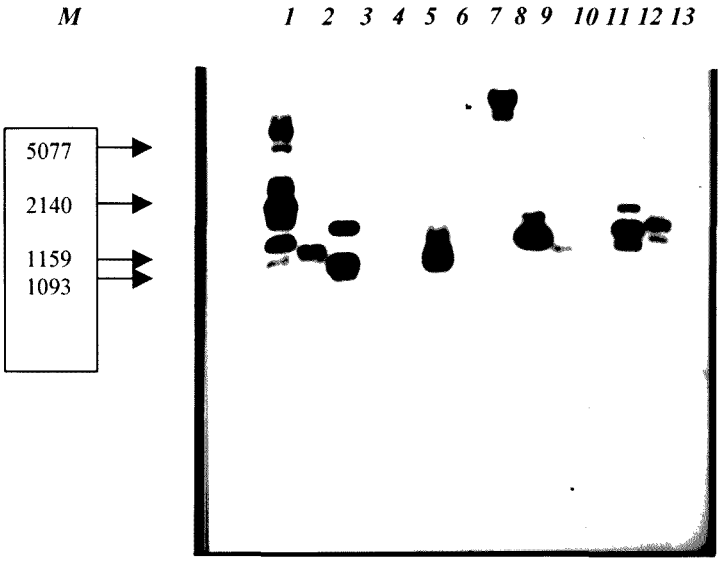


Figure 3.2a: Identification of digested cosmid fragments harbouring the $(AAAT)_n$ STR
Southern blot filter with Sau3A*I-digested cosmid DNA probed with the [γ-^{32}P]ATP labelled $(AAAT)_{10}$ oligonucleotide. The arrows on the left indicate the approximate sizes of the positively hybridising fragments. The* λPst*I size marker fragments, in bp, (boxed) were determined by comparison of the autoradiograph with the EtBr-stained agarose gel previously photographed with a ruler placed alongside from which the Southern blot was generated.*

Lane:	*M:* λPst*I marker*	*7. 8946*
	1. 23078	*8.27268*
	2. 25235	*9.* ***29395***
	3. 24596	*10. 29025*
	4. 19161	*11. 24680*
	5. 27993	*12.* ***24493***
	6. ***20381***	*13. 15217*

Cosmids in bold were selected for the development of $(AAAT)_n$ polymorphic markers.

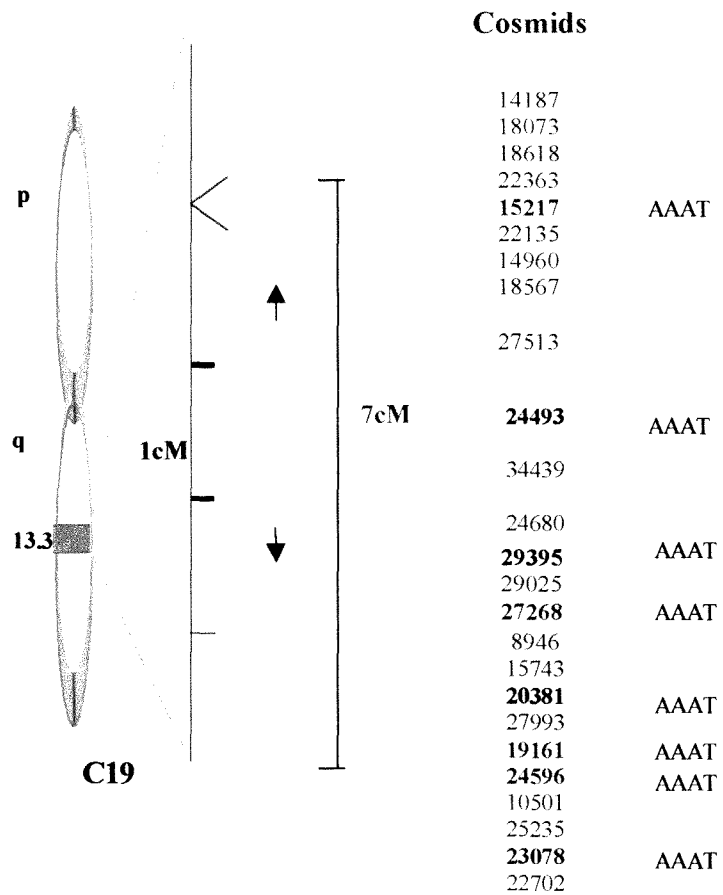


Figure 3.2b: Summary of Southern blot results (shown in figure 3.2a) for cosmid screening for an $(AAAT)_n$ STR within the 7cM PFHBI target interval.
Cosmid clones that gave positive signals on hybridisation with the $(AAAT)_{10}$ oligonucleotide are highlighted in bold, and the cosmids pursued for marker development in the present study are shown in red.

3.1.2 Characterisation of the $(AAAT)_n$ motif

3.1.2.1 Sub-cloning Method

The +1kb fragment of cosmid 29395, which gave a strong signal on Southern blot hybridisation (Figure 3.2, lane 9), was sub-cloned into the plasmid vector pBluescript and used to transform *E. coli* strain XL1-blue. Colony blots of the resulting recombinant plasmid transformants revealed several duplicated positive signals (Figure 3.3), indicating successful sub-cloning of the target fragment.

To confirm successful sub-cloning of the insert, recombinant plasmid DNA minipreps from the ten positively hybridising colonies was digested with *Kpn*I and *Sac*I restriction endonucleases, located on opposite ends of the multiple cloning site, to release the +1kb insert. The subcloning of the insert fragment from cosmid 29395 was successful in nine of the ten colonies (Figure 3.4). Following identification of an insert fragment, miniprep DNA was used as template for sequencing the insert. This revealed an interrupted $(AAAT)_n$ repeat motif, comprising a run of four AAAT motifs, followed by the sequence AATAAAATAAT, after which there was another run of three AAAT repeat motifs (Figure 3.5a).

To determine if the identified interrupted AAAT STR was polymorphic, the locus was amplified in a panel of 40 healthy individuals of Mixed Ancestry using the PCR primers designed from the unique sequence flanking the repeat (in figures 3.5a and b). In all the control individuals, an amplified product of the expected size of 137bp (based on the sequence read, in figure 3.5b) was obtained, but there was no polymorphic size variation (Figure 3.6a). Genotyping of cosmid 29395 DNA with the same flanking primers revealed a faster migrating PCR-amplified fragment on electrophoresis than that generated from genomic DNA samples, indicating that the harboured repeat was of a smaller size than that in genomic DNA of the control panel (Figure 3.6b). To confirm that the repeat was also not polymorphic in the PFHBI pedigrees, a small group of these family members were genotyped, but no polymorphic variation was observed (results not shown).

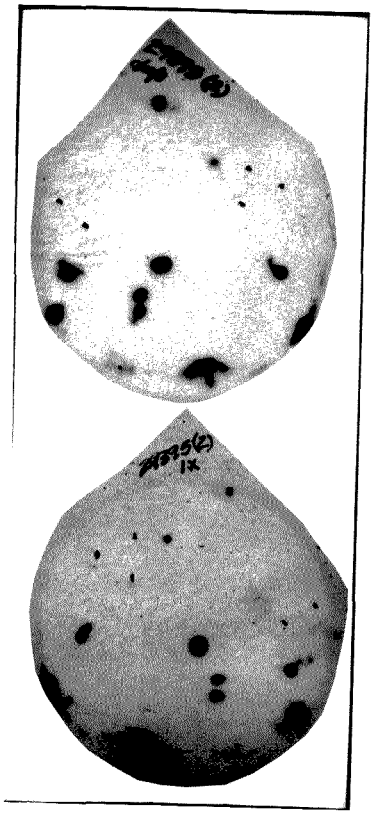


Figure 3.3: Colony blots of transformant colonies containing $(AAAT)_n$ motif-bearing recombinant DNA (29395).

Duplicate filters of transformant colonies were probed with the [γ-^{32}P]ATP labelled $(AAAT)_{10}$ oligonucleotide. The position of the positive signals corresponds to the faces of two filters that were sandwiched together during hybridisation (section 2.7.5).

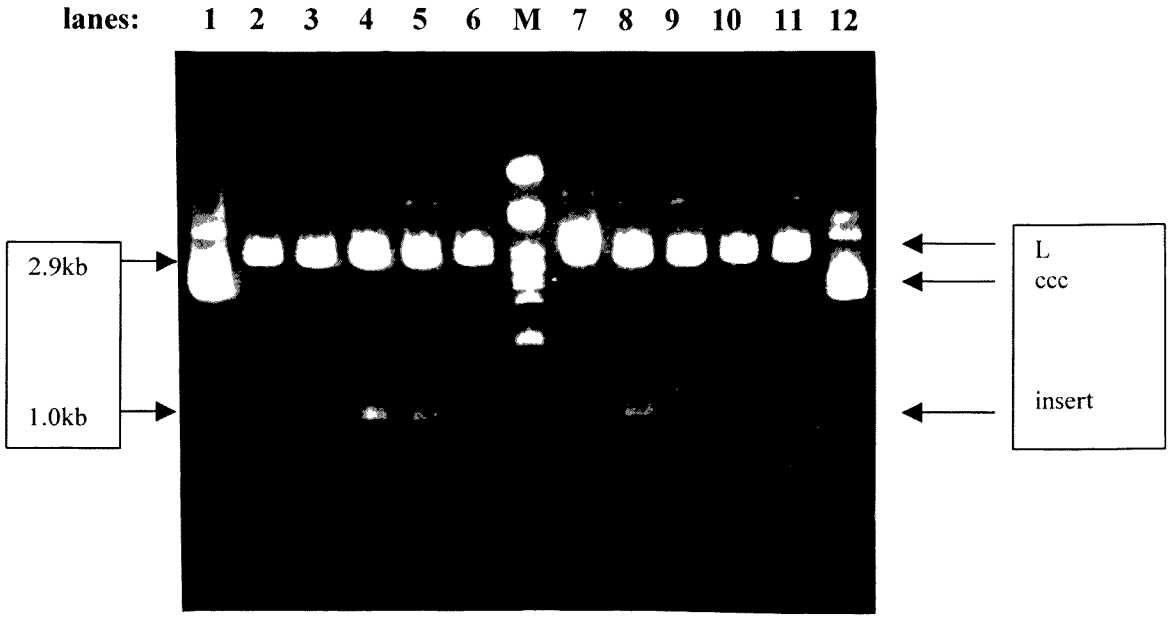


Figure 3.4: Restriction enzyme digestion analysis of recombinant plasmid for the presence of the sub-cloned insert.

Recombinant plasmid DNA from 10 positive colonies identified on the colony blots digested with Kpn*I and* Sac*I to release the insert. Lanes 1 and 12 = covalently closed circular (ccc) recombinant plasmid DNA. Lanes 2 – 11 = Digested, linearised (L) rplasmid DNA (L) (~2.9kb) and the insert DNA (+1kb). Lane M =* λPst*I size marker.*

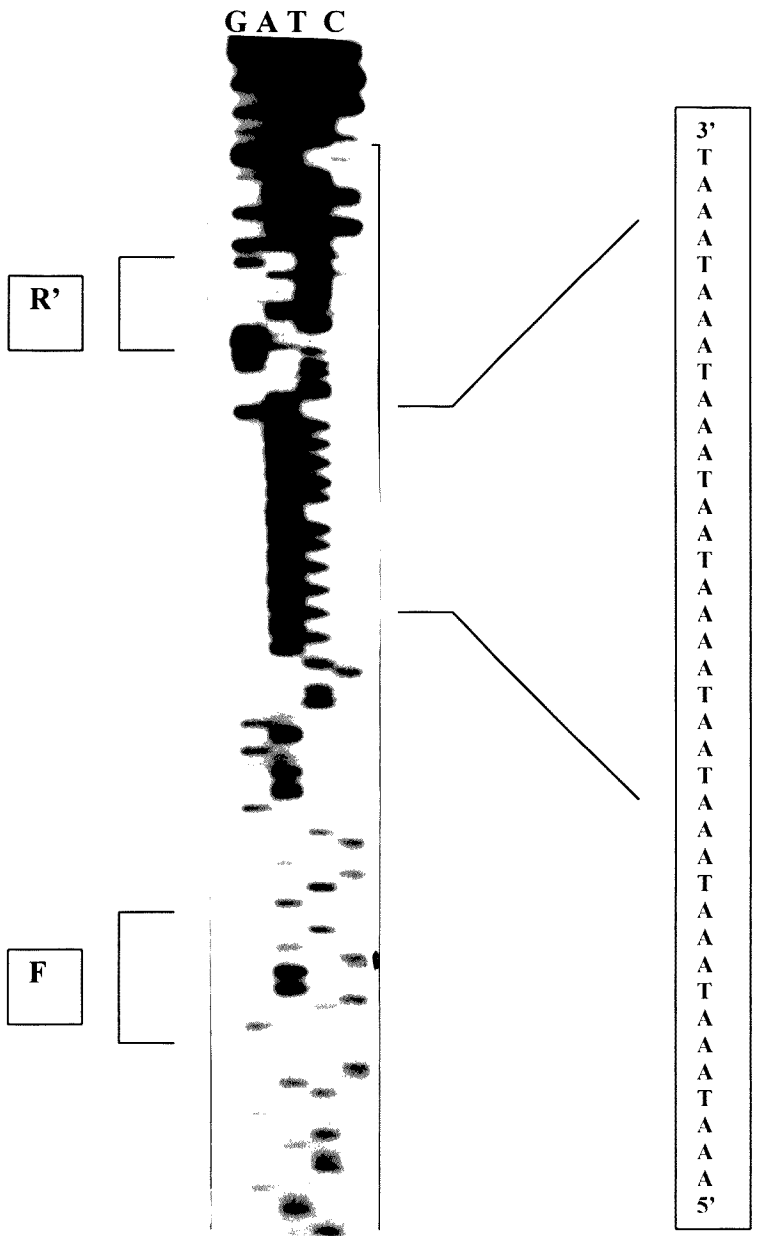


Figure 3.5a: Autoradiograph of the partial sequence of the sub-cloned 1kb fragment of cosmid 29395 sequenced using pBluescript primer T3.

The partial sequence of the region harbouring the AAAT repeat motif is shown on the right, and the positions of the forward and reverse primers used for genotyping this target area are indicated on the left of the figure. F = forward primer sequence, R' = complement of the R primer sequence.

5'-GGGCTGCAGG AATTCGATAT CAAGCTTATC GATACCG*TCGA*TCACACCAC *50*

TGCACTCAGC TGAGCAACAG AGCAAGACTC TGCTAAT**A A ATAAATAAAT** 100

AAATAATAAA ATAAT**AAATA AATAAAT**AGA CATTTGATGT GGTAGGAGTT 150

TACATGTTGT ACTGGTCTAG ATCACTAGAT CACACTTGTA AGCAGATACA 200

TACTGATATG AC-**3'** *212*

Figure 3.5b.Partial sequence of the cloned insert, from cosmid 29395, harbouring the interrupted (AAAT)$_n$ STR.

*Reading from the 5'-direction: Bases 1-37 are part of the vector sequence; bases 38-41 (in italics) are the partially filled in nucleotides of the vector (TC) and insert (**GA**), respectively; the sequences used to design the primer pair 29395F and 29395R to amplify the locus are underlined; the sequence in pink represents the region homologous to Alu-elements. The interrupted repeat was composed of a run of **AAAT** motif repeated four times followed by the sequence AATAAAATAAT after which was another run of **AAAT** motif repeated three times.*

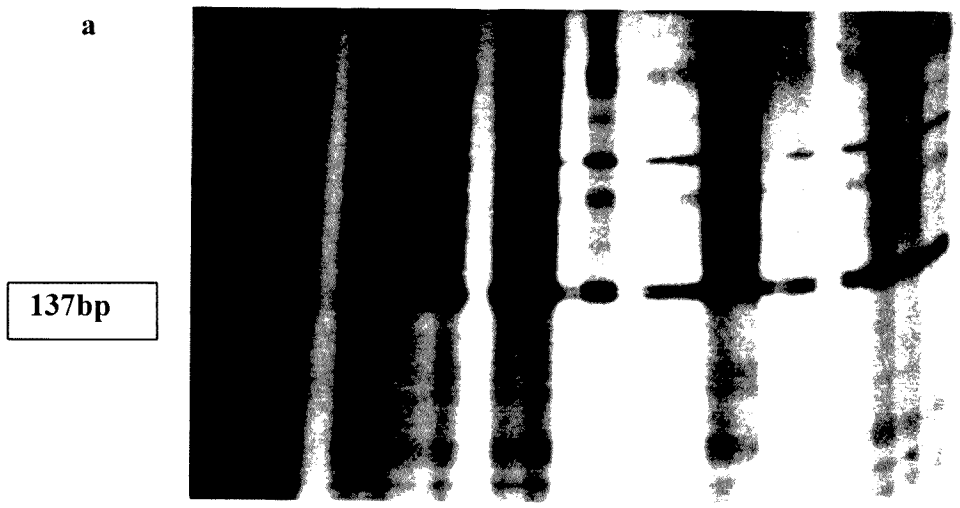
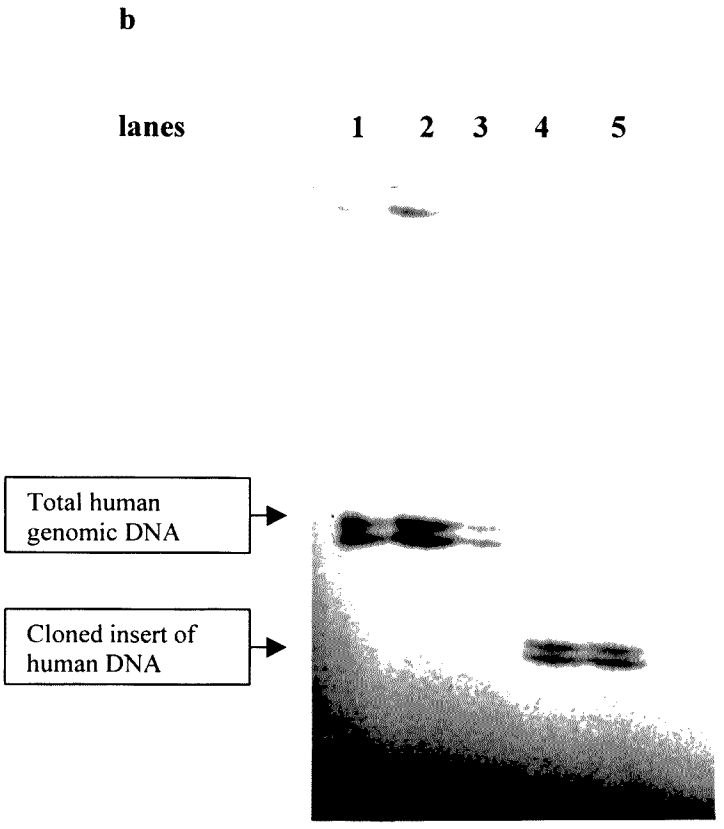


Figure 3.6: Autoradiograph of the interrupted AAAT repeat-region genotyped with primers 29395F and 29395R.

a) Representative results of the amplified repeat region genotyped in 24 of the control panel of 40 unrelated individuals of Mixed Ancestry, the repeat was not polymorphic. The size of the 137bp PCR-amplified allele was calculated from the sequence in figure 3 5b.



b) Comparison of the amplified products of the repeat region genotyped in three members of the panel of unrelated individuals and the chromosome 19q13.3 insert in cosmid 29395. Lanes 1-3 = unrelated individuals from Mixed Ancestry panel, lanes 4-5 = 1x and 10^{-1} dilution of cosmid 29395 DNA, respectively.

Following the same procedure used for cosmid 29395, the approximately 1kb fragment harboured by cosmid 24493 (Figure 3.2, lane 12) was sub-cloned into plasmid pBluescript. Colony blots of the resulting transformants were prepared and after hybridisation with the radio-labelled $(AAAT)_{10}$ oligonucleotide, eighteen duplicate positive signals were generated (data not shown) Plasmid DNA was extracted from all these clones and the DNA from 4 of these 18 clones was digested with vector-specific restriction enzymes to release the insert and the presence of insert was confirmed in two of the four clones (results not shown). One of these two clones was then sequenced from both the 5' and 3' directions using pBluescript T3 and T7 primers, respectively. Although read lengths of approximately 250 nucleotide bases were obtained using each primer, no $(AAAT)_n$ motif, either single or tandemly repeated, was identified in the regions analysed (results not shown). It was therefore possible that the target sequence lay within the un-sequenced region of the insert.

Consequently, at this point, two options were available for characterisation of the repeat motif in cosmid 24493. Firstly, new PCR primers could be designed from the 3'-most sequences generated from the T3 and T7 sequencing primer reads, to allow sequencing further into the insert. The alternative was to use vectorette PCR (section 2.7.7) to directly target the $(AAAT)_n$ motif, and it was the latter approach that was followed.

3.1.2.2 Vectorette PCR-based method

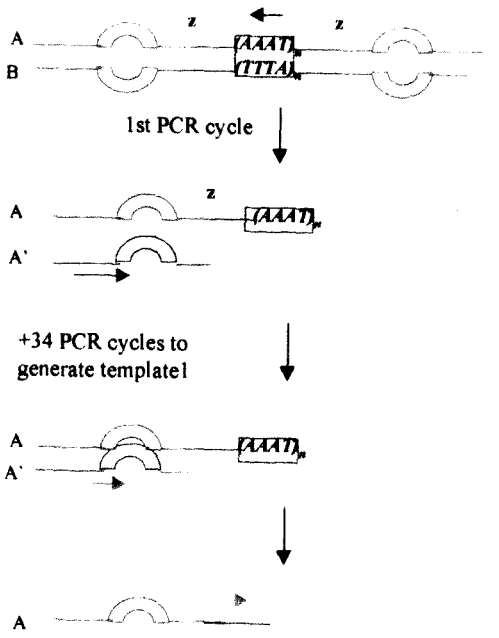
Two cosmids were selected for the development of polymorphic $(AAAT)_n$ repeat markers, namely 24493 (as detailed in the previous section, 3.1.2.1) and 20381, which also gave a positive signal on Southern blot hybridisation (Figure 3.2, lane 6).

In order to increase the chances of obtaining an $(AAAT)_n$ STR contained in a suitably-sized fragment, five separate vectorette libraries were generated from each cosmid with the restriction endonucleases *Alu*I, *EcoR*V, *Pvu*II, *Rsa*I and *Sau3A*1. Each library was then probed for the presence of $(AAAT)_n$ STR by PCR amplification with the vectorette PCR primer 224 and each one of the four repeat anchored primers $(AAAT)_4$-A, $(AAAT)_4$-T, $(AAAT)_4$-G and $(AAAT)_4$-C. In the sections that follow, the results obtained with the development of AAAT STR from cosmids 24493 and 20381 are presented.

Figure 3.7: Accompanying diagram for the identification and characterisation of the $(AAAT)_n$ repeat motifs harboured in cosmids 24493 and 20381 using vectorette PCR-based strategy.

B

PCR-amplification 1

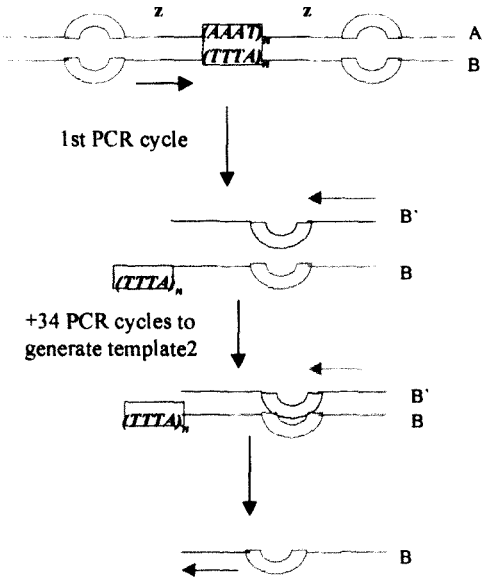


PCR amplification of vectorette library using a **repeat-anchored primer** and vectorette primer 224 to generate **template 1**.

Template 1 was sequenced with primer 10108 and the information used to design the putative repeat-flanking primer F

the putative repeat-flanking primer F will anneal to the strand complementary to the one from which it was designed.

PCR-amplification 2

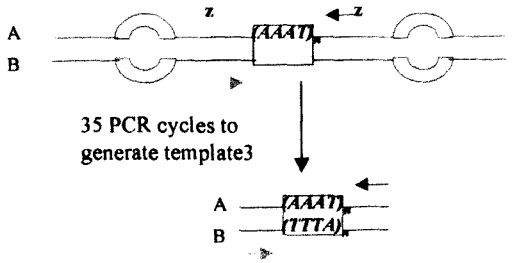


PCR amplification of vectorette library using a putative repeat-flanking primer F and vectorette primer 224 to generate **template 2**.

Template 2 was sequenced with primer 10108 and the information used to design the putative repeat-flanking primer **R**

The putative repeat-flanking primer **R** will anneal to the strand complementary to the one from which it was designed.

PCR-amplification 3



PCR amplification of vectorette library using a putative repeat-flanking primer pair F and R to generate **template 3**.

Sequencing of template 3 using the putative repeat-flanking primer pair F and R to identify and characterise the $(AAAT)_n$.

24493

On amplification of the cosmid 24493 vectorette libraries, two specific PCR products, approximately 247bp and 264bp in size, were generated from the *Alu*I and *Rsa*I libraries, respectively, using the $(AAAT)_4$-C primer (Table 3.1). Three products were generated from the *Sau3A*1 library, a fragment of approximately 339bp amplified with the anchored primer $(AAAT)_4$-G and two fragments of approximately 450bp and 339bp amplified with the $(AAAT)_4$-C anchored primer (Table 3.1). Other diffuse products were also generated from the *Alu*I, *EcoR*V, *Pvu*II and *Sau3A*1 libraries (Table 3.1), but these were considered to be the result of non-specific priming and were not pursued further.

Using the vectorette sequencing primer 10108, both the 247bp (Figure 3.8a) and 264bp (results not shown) fragments were sequenced and found to have the same sequences, however, the 264bp fragment had an extended 5'-end. A primer, designated repeat-flanking sequence primer F (**F**), was designed from the 3'-most sequence read of the 247bp *Alu*I fragment (Figure 3.8b). Primer **F** was then used, together with vectorette PCR primer 224, to generate new sequencing template (template 2) from the *Alu*I vectorette library. The resulting product (template 2) of approximately 450bp was sequenced using sequencing primer 10108. A sequence read of approximately 260 bases was obtained, however, no $(AAAT)_n$ motif was identified (Figure 3.9a). Therefore, a new primer, designated repeat flanking primer R (**R**) was designed from the 3'-end of template 2 sequence (Figure 3.9b).

Amplification of *Alu*I library DNA using the designed primer pair **F** and **R** generated two products approximately 190bp and 220bp in size (Figure 3.10a, lane 1). Sequencing of the 220bp product with primers **F** or **R** revealed no AAAT tandem-repeat motif (Figure 3.10b). Further analysis of the sense (**F**orward) and antisense (**R**everse) strands showed that they had a region of overlap, indicating that the DNA template had been sequenced in its entirety. A single AAAT motif embedded in the sequence 5'GG**ATTT**G3' was, nevertheless, observed (Figure 3.10c). Amplification of human genomic DNA (an individual from the PFHBI panel) using **F** and **R** also generated two fragments, of approximately 190bp and 220bp in size, as generated by amplification of the cosmid vectorette library (Figure 3.10a, lane 2).

Table 3.1: Polymerase chain reaction amplification products (template 1) generated from five different cosmid 24493 vectorette libraries using repeat-anchored primers $(AAAT)_4$-N (N = A, T, G or C) and vectorette PCR primer 224·

	Anchored primers $(AAAT)_4$-			
Library	**-A**	**-T**	**-G**	**-C**
*Alu*I	±	±	–	+
*EcoR*V	±	±	–	–
*Pvu*II	±	±	–	–
*Rsa*I	–	–	–	+
*Sau3A*1	±	±	+	+ **(2)**

Five cosmid 24493 vectorette libraries were generated using Alu*I,* EcoR*V,* Pvu*II,* Rsa*I and* Sau3A*1 restriction endonucleases. A = $(AAAT)_4$-A, T = $(AAAT)_4$-T, G= $(AAAT)_4$-G and C = $(AAAT)_4$-C. (+) = specific PCR product(s) observed on electrophoresis on a 1.5% agarose gel; (±) = non-specific PCR products; (–) = no PCR product;* ***(2)*** *= two specific PCR products.*

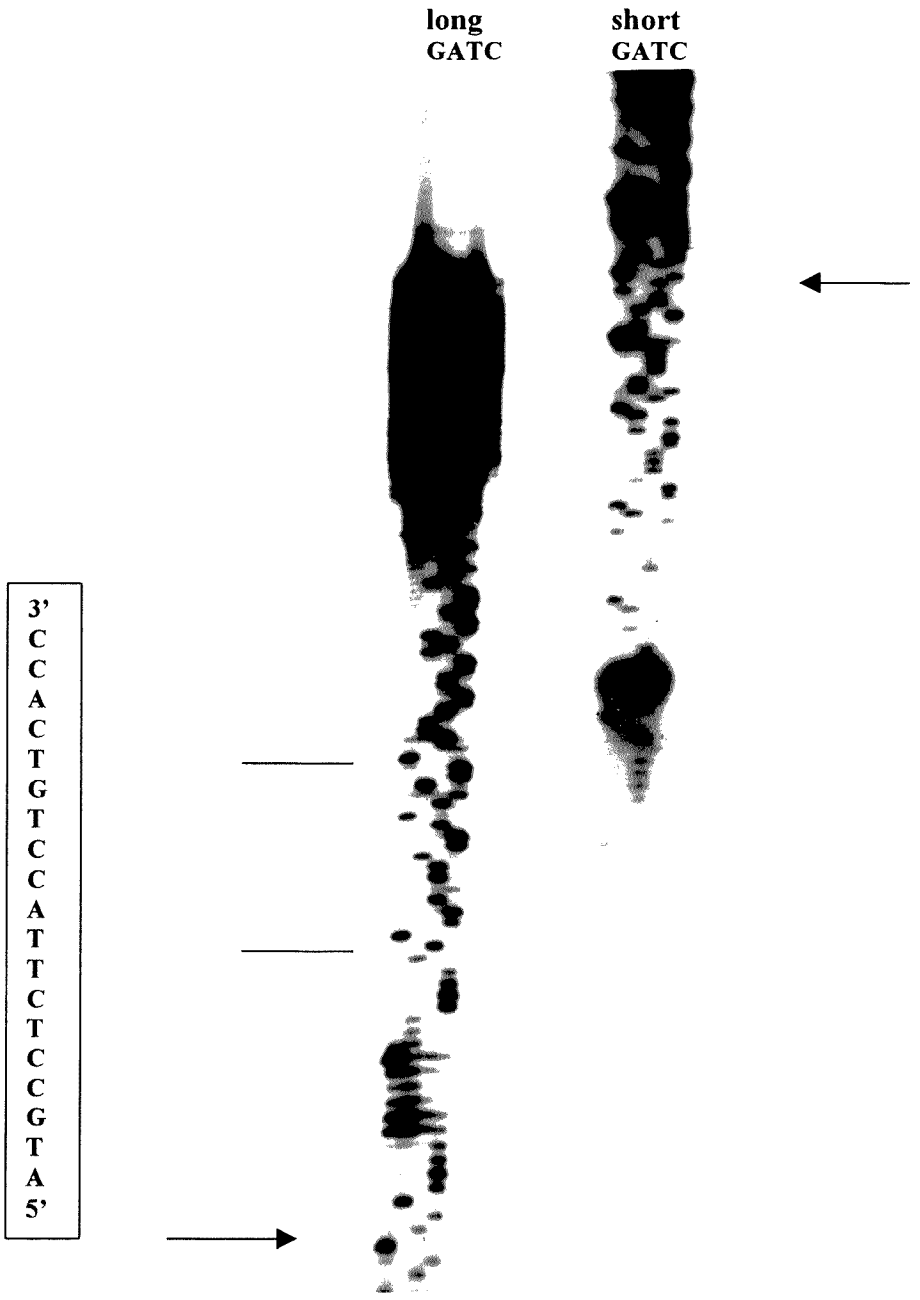


Figure 3.8a: Autoradiograph of the sequence of cosmid 24493 AluI-generated PCR product (template 1).

*Partial sequence of the AluI fragment, template 1, amplified with anchored primer $(AAAT)_4$-C and vectorette PCR primer 224 and sequenced with vectorette sequencing primer 10108. The arrows indicate the portion at which the switch from reading short sequence to long sequence runs occurs. The boxed sequence was used to design primer24493**F**.*

24493 #1

5'-TACTTCTCAG TGTTTTCTAG TATATTCACA GACTTGTGCA GCCATTTCCA 50

CAGTCAATTT TAGAACATTG TCGTCACCCC AANNNGAAAC CCCATGCCTC 100

TTACCTGTCA CCGCTAATCC TTCCATACCT CCCATCTTCA GTTCTA-3' 146

Figure 3.8b.Partial sequence of the ^{Alu}I fragment of cosmid 24493 amplified with $(AAAT)_4$-C anchored primer and vectorette PCR primer 224 (template 1) and sequenced with primer 10108.

Part of the sequence at the 3'-end, underlined, was used to design the forward primer (24493F); N = A, T, G or C (autoradiograph shown in Figure 3.8a).

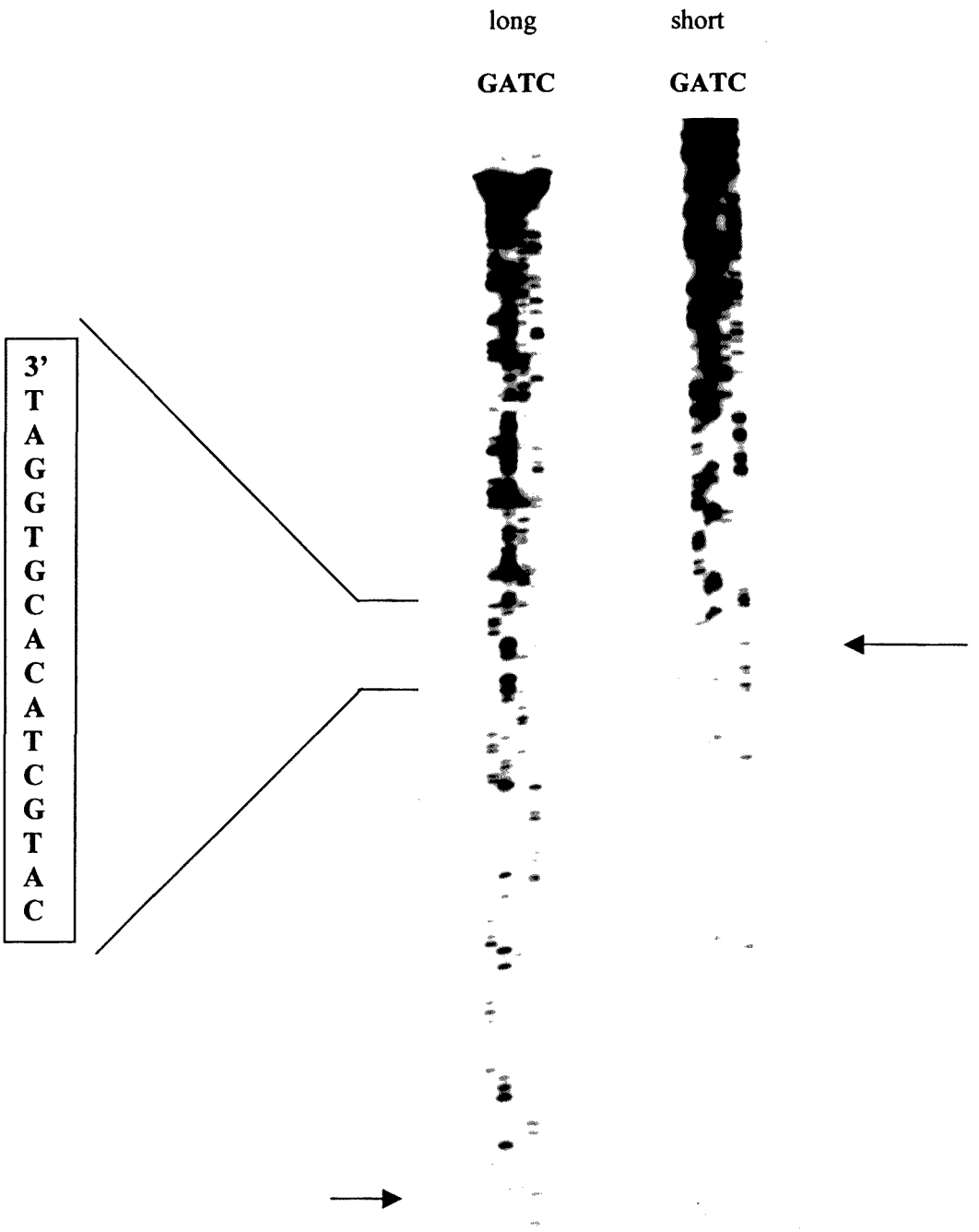


Figure 3.9a: Autoradiograph of the sequence of template 2, the PCR product generated from the cosmid 24493 AluI vectorette library using primer 24493F and primer 224.
Partial sequence of template DNA amplified with primer 24493F and vectorette PCR primer 224, sequenced using the vectorette sequencing primer 10108. The arrows indicate the portion at which the switch from reading short sequence to long sequence runs occurs. The boxed sequence was used to design primer 24493R.

24493 #2

5'-*GGGTTTACAC CATGTGTGGC CAGGCTGGTC TTNAACTCCT GCACCTCAGG* ***50***

AGATCCACCT GCCTTGGACA CCCCCAAAGC GTTGGGGTTA TAGGCGTAAG ***100***

CCACCGCCGC CCCGGCACAG GAATGGAGTT CTNATACATG CTACACGTGG ***150***

ATNAACTTNN AAAACATTAT ACTGACTGCA AAGAAGCCAC ACACGAAAGA ***200***

CACAATGGTA TGATCCATCT ATANAGTAGC ACATAGCATC ATATACAGG-**3'** *249*

Figure 3.9b.Partial sequence of the template 2 amplified from cosmid 24493 AluI vectorette library using primers 24493F and 224.
A reverse primer, 24493R (underlined), was designed from the 3'-end of this sequence (sequence autoradiograph shown in Figure 3.9a).

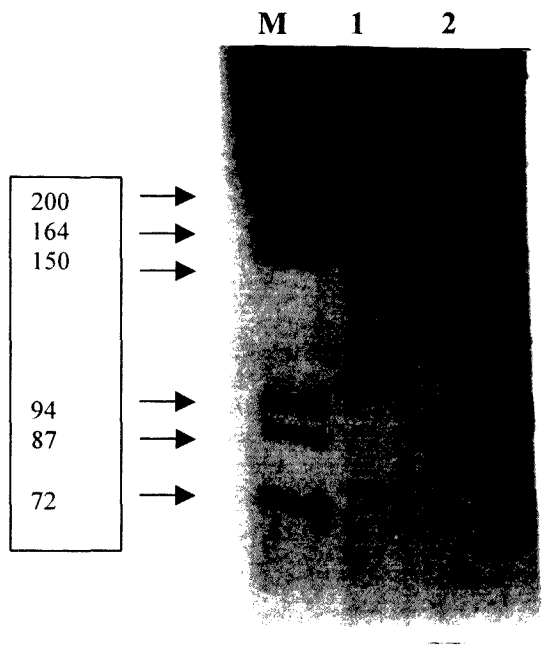


Figure 3.10a: Polymerase chain reaction products amplified with primers designed from the repeat flanking sequence, 24493F and 24493R.
M = λPstI size marker, box = marker fragment sizes in bp; lane 1 = PCR-products (approximately 190bp and 220bp) generated from cosmid 24493 AluI-vectorette library; lane 2 = Products (approximately 190bp and 220bp) generated from human genomic DNA (PFHBI-affected individual).

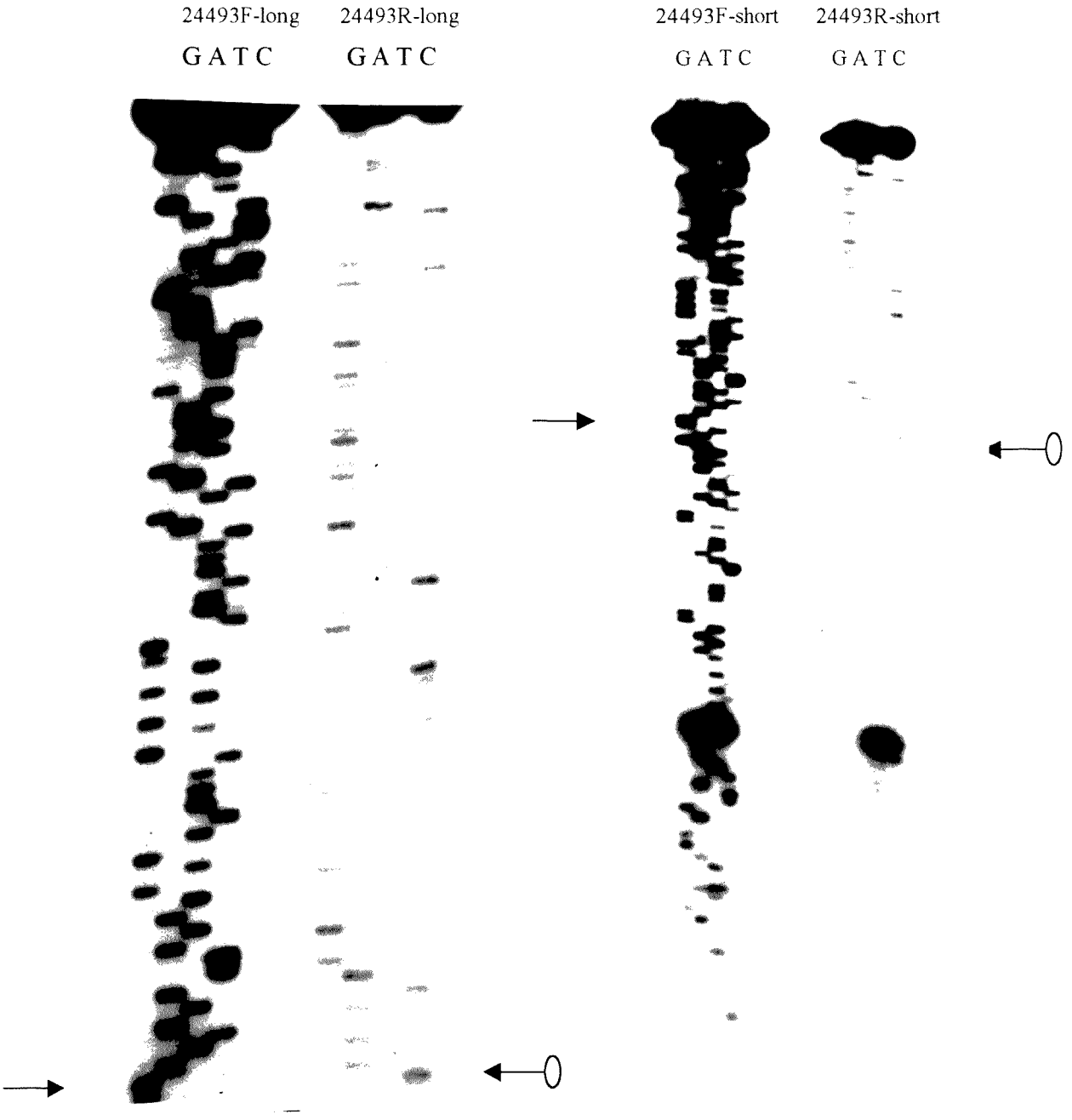


Figure 3.10b: Autoradiograph of the sequence of PCR-product amplified from* AluI *vectorette library using primer pair 24493F and 24493R.

F = forward primer, R = reverse primer. The arrows indicate the position at which the switch from reading short sequence to long sequence runs occurs for the forward ( ) and reverse primer ( ), respectively.

24493 #3:

5'-CCTCTTACCT GTCACCGCT**A AAT**CCTTCCA TACCTCCCAT CTTCAGTTCT 50

AGGCAGCCACT AATCTTTCTC TGTATATGG**A TTT**GCCTATT GTGGCTACTT 100

CATATAGATG GAATCATACC ATATGTGTCT TTCGTGTGTG GCTTCTTTCA 200

GTCAGTATAA TGTTTTCAAA GTTCATCCAC GTTAGCA**-3'** 237

Figure 3.10c: Complete sequence of ~220bp PCR product generated and sequenced with repeat flanking primers 24493F and 24493R.
*Underlined sequences represent primers 24493F (5'end) and the complement of 24493R (3'end) used to generate this product, indicating that the complete sequence had been attained. The **AAAT** (nucleotide bases 30-33) and **ATTT** (nucleotide bases 80-83) indicate the diffuse AAAT/ATTT sequences identified.*
The T tract of the reverse sequence was read from an autoradiograph other than that shown in Figure 3.10b.

20381

On amplification of cosmid 20381 vectorette libraries three *Rsa*I-PCR specific products of apparently the same size (~805bp), were generated using each of the repeat anchored primers $(AAAT)_4$-A, $(AAAT)_4$-T and $(AAAT)_4$-G (Table 3.2, Figure 3.11a). No PCR products were generated from the other vectorette libraries.

Two of the three approximately 805bp products amplified from the *Rsa*I library (template 1), namely those generated with the $(AAAT)_4$-T and $(AAAT)_4$-G anchored primers, were sequenced using the vectorette sequencing primer 10108 (Figure 3.11b), and they were found to contain similar sequences. A repeat flanking primer $20381^{\mathbf{F}}$ was designed from the 3'-most sequence of template 1 generated using the repeat anchored primer $(AAAT)_4$-G (Figures 3.12b and c) and used to generate new sequencing template 2 from the cosmid 20381 *Rsa*I library (Figure 3.11a, lane 3). The sequence of template 2, of approximately 514bp, (Figure 3.12a) was used to design a second primer, designated repeat flanking primer $20381^{\mathbf{R}}$ (Figure 3.12a and b). Further analysis of this sequence revealed a single AAAT motif, embedded within the sequence 5'GAAATT3' (Figure 3.12b).

Amplification of *Rsa*I DNA library using the designed primer pair $20381^{\mathbf{F}}$ and $20381^{\mathbf{R}}$ generated template 3 (approximately 500bp) (Figure 3.11a, lane 5). However, on sequencing of template 3 with the same primers, no AAAT single or repeat motif was revealed (Figure 3.13). Polymerase chain reaction amplification of human genomic DNA from a PFHBI family member using primers $20381^{\mathbf{F}}$ and $20381^{\mathbf{R}}$ did not generate a product. Consequently, a new primer, designated $20381^{\mathbf{R'}}$, was designed from antisense ($20381^{\mathbf{R}}$) sequence read of template 3 (Figure 3.13) and used together with primer $20381^{\mathbf{F}}$ to amplify the cosmid 20381 *Rsa*I-library. Template 4 of approximately 340bp product was generated (Figure 3.14a, lane 1). However, no AAAT single or repeat motif was observed on sequencing of this fragment using primers $20381^{\mathbf{F}}$ and $20381^{\mathbf{R'}}$ (Figure 3.14b and c).

Table 3.2: Polymerase chain reaction amplification of target sequence from five different cosmid 20381 vectorette libraries using repeat-anchored primers $(AAAT)_4$-N (N = A, T, G or C) and the vectorette PCR primer, 224·

	Anchored primers $(AAAT)_4$-			
Library	**-A**	**-T**	**-G**	**-C**
***Alu*I**	–	–	–	–
***EcoR*V**	–	–	–	–
***Pvu*II**	–	–	–	–
***Rsa*I**	+	+	+	–
***Sau3A*1**	–	–	–	–

Five cosmid 20381 vectorette libraries were generated using Alu*I*, EcoR*V*, Pvu*II*, Rsa*I and* Sau3A*1 restriction endonucleases. A = $(AAAT)_4$-A, T = $(AAAT)_4$-T, G= $(AAAT)_4$-G and C = $(AAAT)_4$-C. (+) = specific PCR product(s) observed on electrophoresis on a 1.5% agarose gel; (–) = no PCR product.*

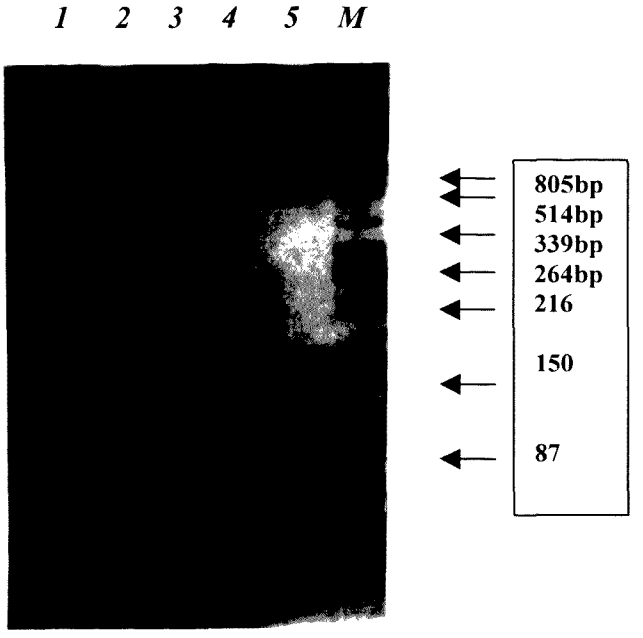


Figure 3.11a: Amplification products generated from cosmid 20381 RsaI-vectorette library. *Lanes 1 and 2 = template 1, PCR-amplified using anchored primer $(AAAT)_4$-T (lane 1) or $(AAAT)_4$-G (lane 2) together with vectorette primer 224; lane 3 = template 2, product amplified with 20381F and vectorette primer 224; lane 5 = template 3, product amplified with 20381**F** and 20381**R**; M = λPstI size marker with fragment sizes in bp indicated on the right.*

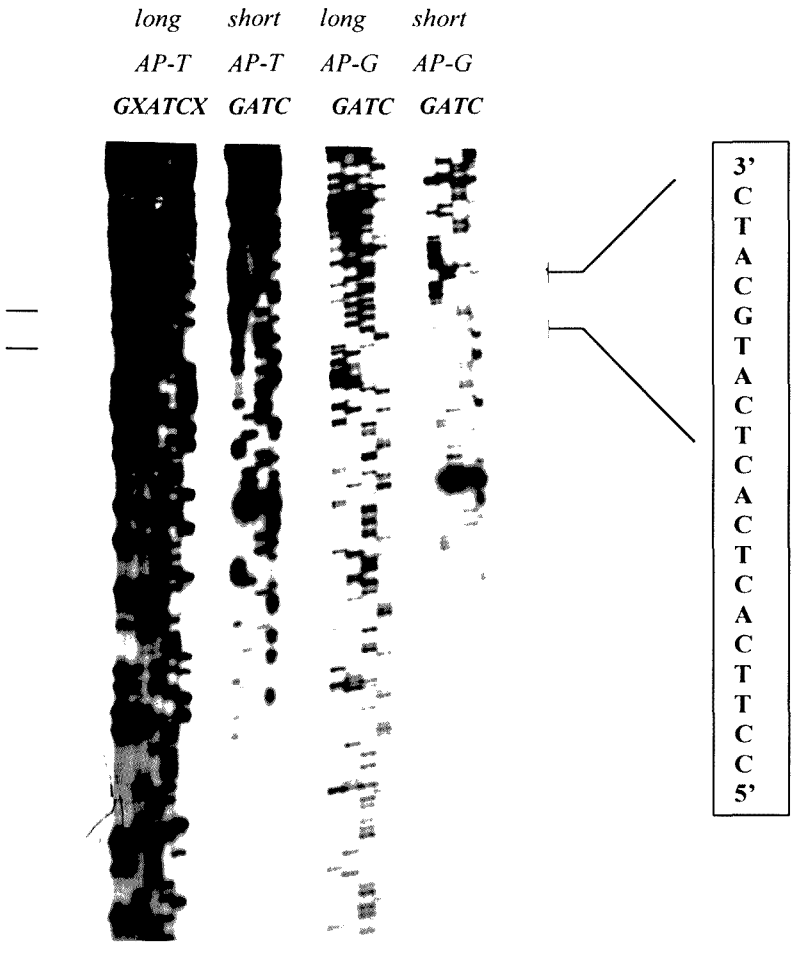


Figure 3.11b: Autoradiograph of the sequence of template 1 amplified from the cosmid 20381* RsaI*-vectorette library.

Two PCR products of approximately the same size, representing template 1, were generated from the cosmid 20381 RsaI *vectorette library using the anchored primers (AP), $(AAAT)_4$-T or $(AAAT)_4$-G together with the vectorette PCR primer 224, which were subsequently sequenced with vectorette sequencing primer 10108. Both the $(AAAT)_4$-T- (left) and $(AAAT)_4$-G-generated (right) sequences were similar, but the $(AAAT)_4$-T-generated sequence migrated slower than $(AAAT)_4$-G-generated sequence, indicating a slight variation in size. The arrows indicate the position of the sequence (long read) which was used to design primer 20381**F** (boxed) from template 1 amplified using either, repeat-anchored primer $(AAAT)_4$-T (left) or $(AAAT)_4$-G (right) and vectorette primer 224. X = adjacent sample overflow during gel loading.*

20381 #1

5'-TCAGGGCTCT CCCACCTACA CATCTGGCAC CTATTGTTTC CCTGCCTGTC 50

TTTGCGTGGC ACATGAGTGC GTGTTCTTCA GCACTTTGTT CNAATCTACC 100

CCCTCCCAGA AGCCTTGCTG CCCCCATCTA **ATTT**CAGCCT TCTTCTTGGT 150

CCACCTGGCT TCCCCCTTCT GGCAGGACCT CAGAGGGAAG GCAGTGCCTT 200

CACTCACTCA TGCATCTCGC ATCTCTCAGA CTC-**3'** 233

Figure 3.11c: Partial sequence of 805bp fragment, template 1 of cosmid 20381 Rsa**I library amplified using anchored primer (AAAT)$_4$-G and universal primer 224 and sequenced with sequencing primer 10108.

*A single **(AAAT)** motif was observed. Underlined sequence represents designed primer 20381**F**. N = G, A, T or C..*

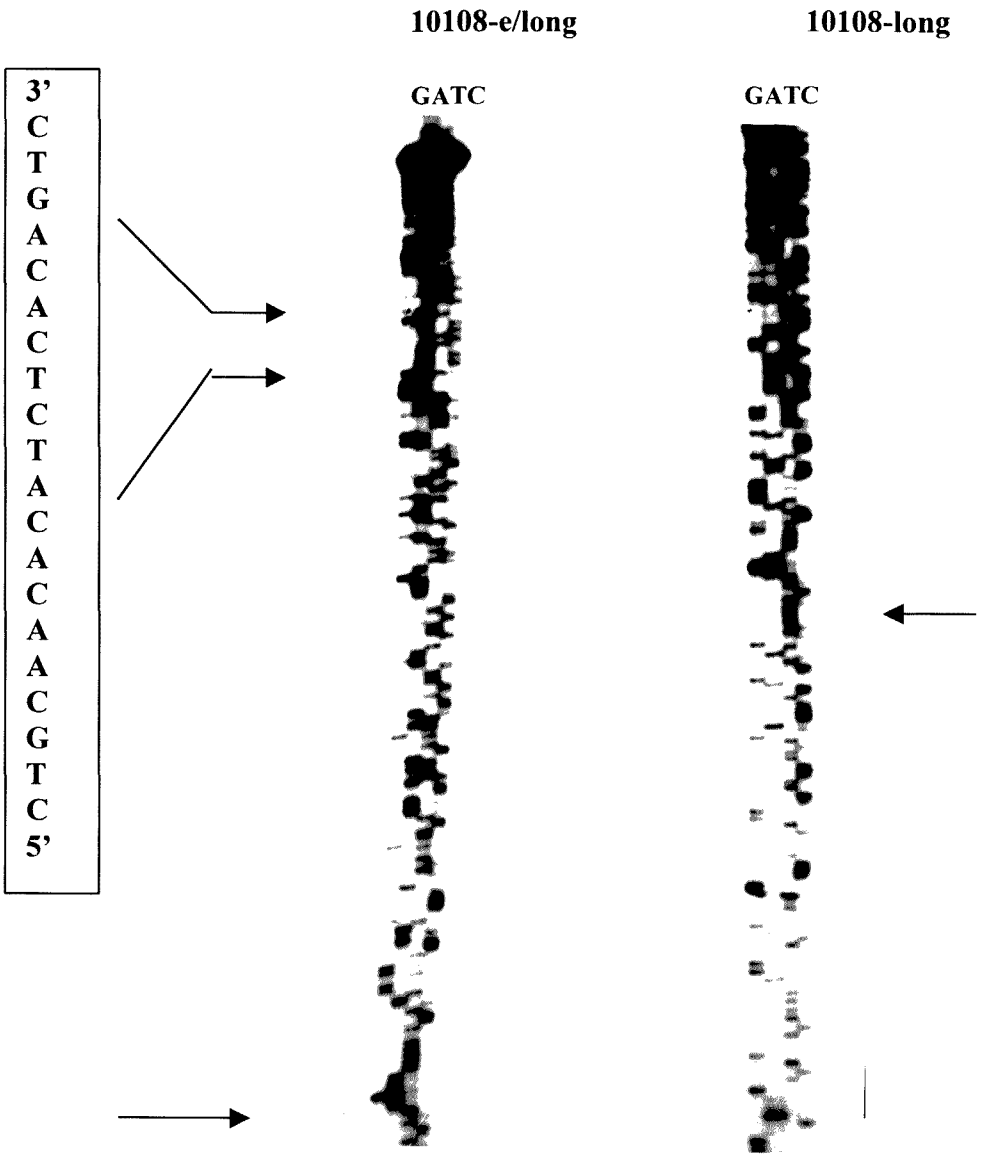


Figure 3.12a: Autoradiograph of sequence of template 2 of cosmid 20381.
Template 2 was amplified from the RsaI *vectorette library using primers 20381F and 224 and then sequenced with primer 10108. The arrows indicate the portion at which the switch from reading long sequence to extra (e)-long sequence runs occurs. The boxed sequence was used to design primer20381**R**. The 3'end of this sequence, boxed, was used to design primer 20381**R**.*

20381 #2:

5'CTTACCAGCA CAACGTCTAT CGCGCACACA CCCAGAGCTG ATGCTGCTCA 50

CAGCCCCCAG TGGTCATCCC CGCTGCCTCC AGGCAAAGTC TGATCTTTTC 100

AGGCTCAGCT TTGAGCCCTG ACTGGACCTT CCCTTCTGAC ACCCATCACA 150

GCCTCATGCT CTTTTCTCCT CTAAG**AAAT**T TTTGTCCTCA TGGTGGTCCA 200

CATGCCCCCA GCTCTGTGTT CCTCACACCC ACTATACCTG TGTATCACCC 250

TCCCTTTCCA GAATGCTTTC CTTCCCACAC ATGACTCTGC TCAGCTCAGA 300

CTAGCCTCAT GTCTCTCAGA GCTCTGTACT TAGAGACTGC AACACATCTC 350

ACAGTCAGCA T-**3'** 361

Figure 3.12b: Partial sequence of template 2 of cosmid 20381 (500bp).
Cosmid 20381 RsaI *library PCR- amplified with anchored primer 20381F and universal primer 224 and the generated product sequenced with sequencing primer 10108. A single **AAAT** motif was observed. The underlined sequence was used to design primer 20381**R**.*

20381#3

5'-CTTACCAGCA CAACGTCTAT CGCGCACACA CCCAGAGCTG ATGCTGCTCA 50

CAGCCCCCAG TGGTCATCCC CGCTGCCTCC AGGCAAAGTC TGATCTTTTC 100

AGGCTCAGCT TTGAGCCCTG ACTGGACCTT CCCTTCTGAC ACCCATCACA 150

GCCTCATGCT CTTTTCTCCT CTAAG**AAAT**T TTTGTCCTCA TGGTGGTCCA 200

CATGCCCCCA GCTCTGTGTT CCTCACACCC ACTATACCTG TGTATCACCC 250

TCCCTTTCCA GAATGCTTTCC CTTCCCACA CATGGAACTC CTGCTCAGCC 300

TCATGAACTA GCTCATGTCT CTCAGAGCTC TGTACTTAGA GACTGCAACA 350

CATCTCACAG TCAGCAT-**3'** 367

Figure 3.13: Partial sequence of template 3 of cosmid 20381 RsaI library.
Partial sequence of PCR product from cosmid 20381 RsaI *library amplified using 20381**F** and 20381**R** and sequenced using primer 20381**R**. A single AAAT motif was observed. Underlined sequenced was used to design primer 20381**R'**.*

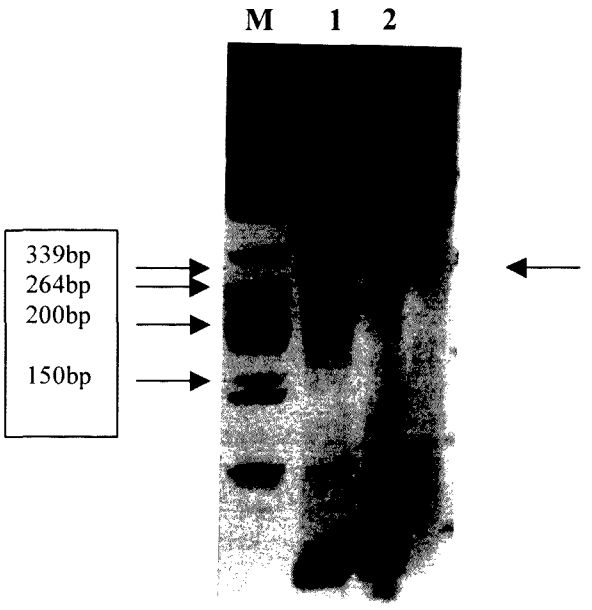


Figure 3.14a: Polymerase chain reaction amplification using primers 20381F and 20381R'. *Lane M = λPstI size marker with the fragment sizes in bp, lane 1 = Cosmid 20381* RsaI *library, lane 2 = Human genomic DNA. The arrow on the right indicates the approximately 339bp products generated on PCR-amplification.*

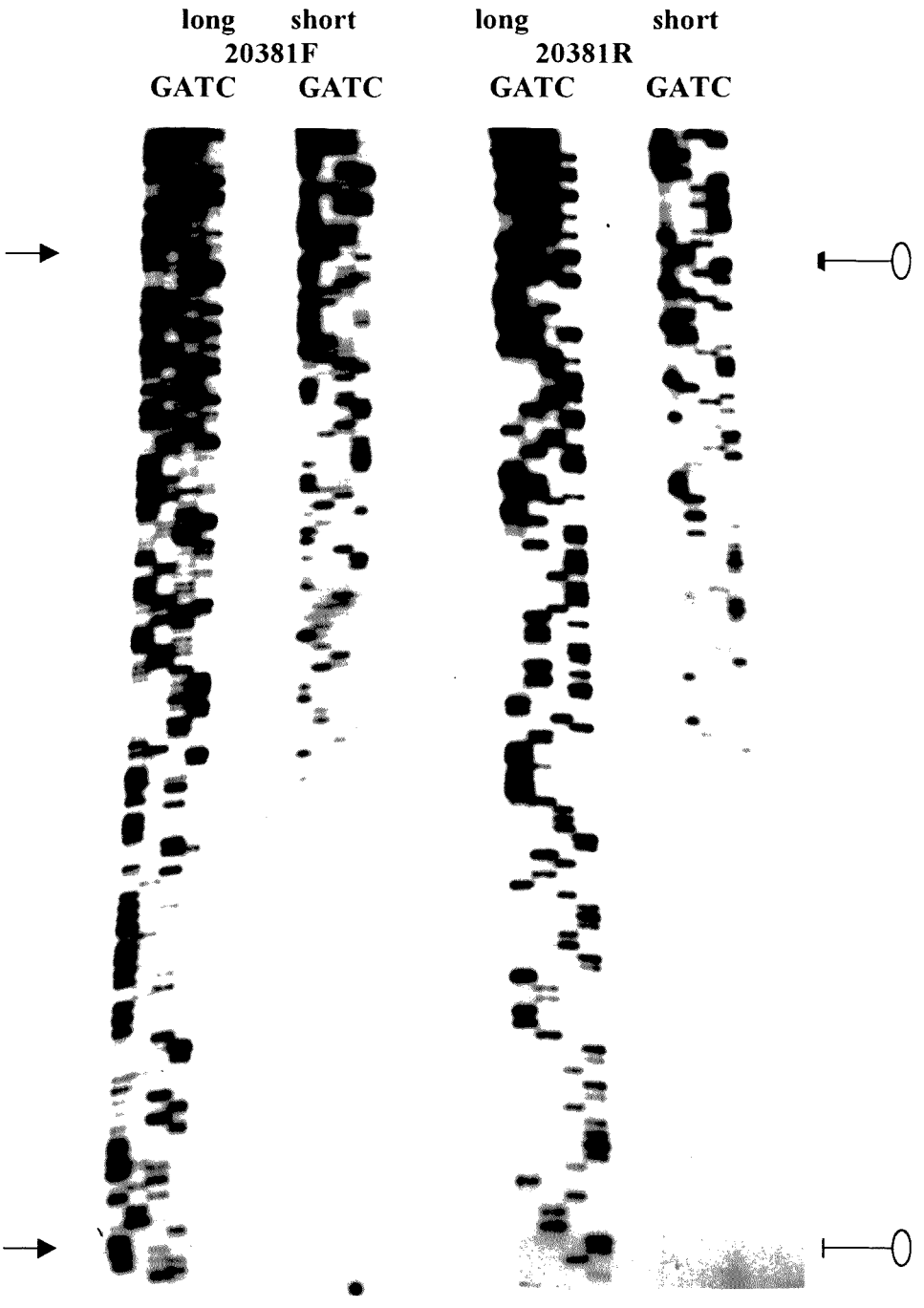


Figure 3.14b: Autoradiograph of sequence of template 4 of cosmid 20381* RsaI *library. *Template 4 was amplified from the* RsaI *vectorette library using 20381F and 20381R and sequenced with 20381F (F) or 20381R (R). The arrows indicate the portion at which the switch from short sequence to long sequence runs occurs, for both the forward (→) and reverse (⟵) sequences, respectively.*

20381 #4:

5'-CCAGTCAGTG TCAAGCATAG AATGACTGAC ATGCCCCATC CTGTCTAGAA	50
GGGAGATGGC AAGTGGCTCT GACCTGGGTG GGGAGGGGAT GTTGGGTGTG	100
GCCAGTTCCT CCCTAAAGGT ATCAGGGAAG GCTTCCTGGA GGAGGGGACA	150
TTGGAGCTAG TTCTTGAGGC TGAGCAGGAG TTCCTGTGTG GGGAAGGGAA	200
AGGCATTCTG GAAAGGGAGG GTGATACACA GGTAT-**3'**	245

Figure 3.14c: Partial sequence of template 4 of cosmid 20381.
Partial sequence of cosmid 20381 RsaI *library PCR-amplified using 20381F and 20381R. No AAAT, single or repeat motif was observed.*

3.2 CANDIDATE GENE SCREENING

At the chromosome 19q13.3 locus to which PFHBI has been mapped, there are a number of genes that can be considered plausible candidates (Figure 1.5), because of their known or postulated involvement in cardiac physiology. In the present study, selected exons of candidate genes *GSY1* and *HRC* were screened for the presence of possible PFHBI-causing mutations.

3.2.1 Glycogen Synthase (*GSY1*)

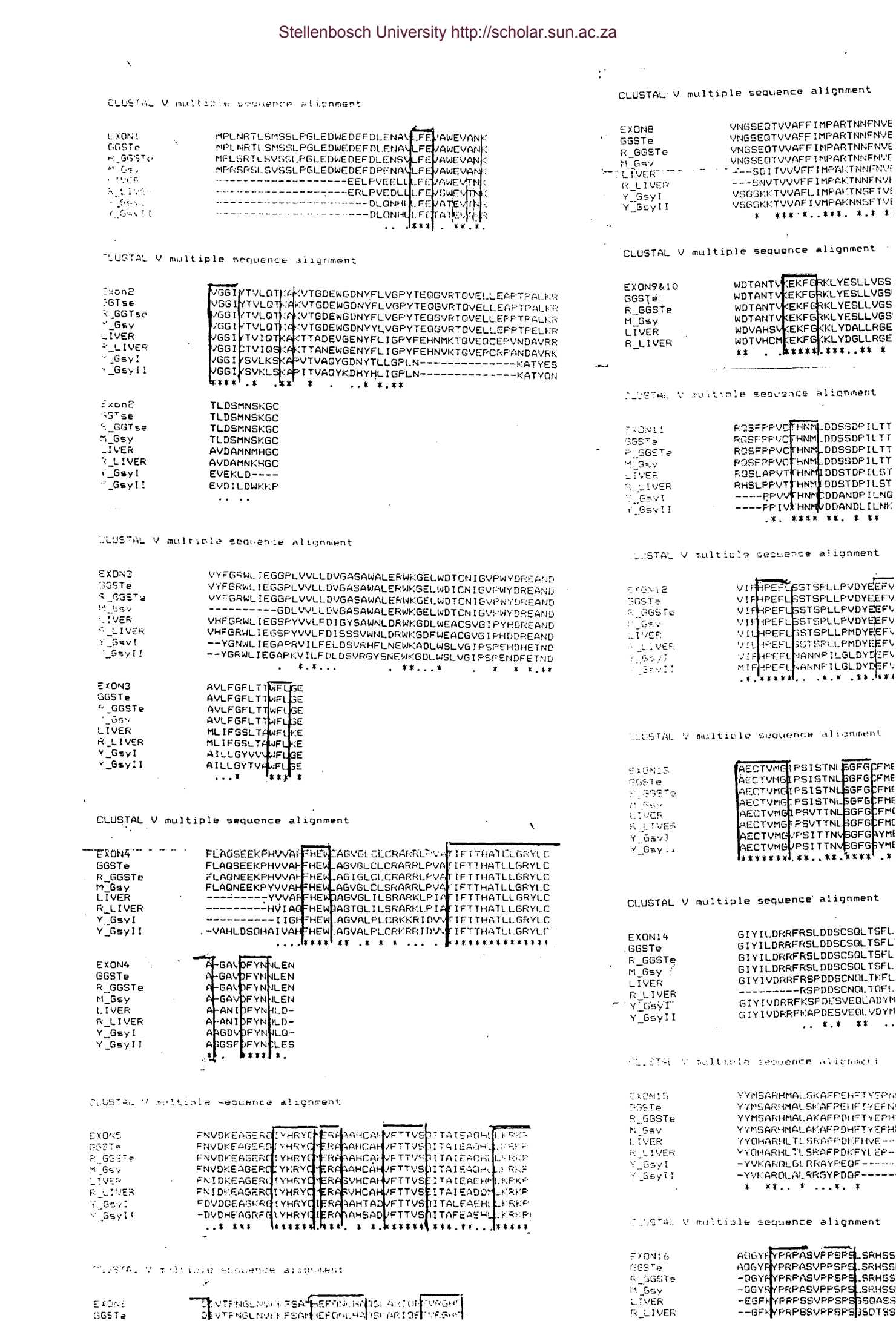
Human muscle *GSY1* consists of 16 exons, encoding an 83.6kDa protein (Ohro *et al.*, 1995). The conserved amino acid sequences of human muscle (GSY1) and liver (GSY2) glycogen synthase isozymes and their respective homologues from rat, mouse and yeast, translated from exonic nucleotide sequences (http://www.ncbi.nlm.nih.gov/) were analysed for conserved regions using the CLUSTAL V multiple sequence alignment-program. The output of this alignment revealed that sequences expressed by exons 4, 5, 11, 12 and 16 were the most conserved both within and across species (Figure 3.15). Highly conserved regions are thought to represent functionally important protein domains, in which mutations may cause disease. Consequently, exons 4 (379bp), 5 (225bp), 11 (159bp), 12 (200bp) and 16 (361bp) were selected for mutation screening by PCR-SSCP analysis (Orita *et al.*, 1989) and the results are presented in the following section.

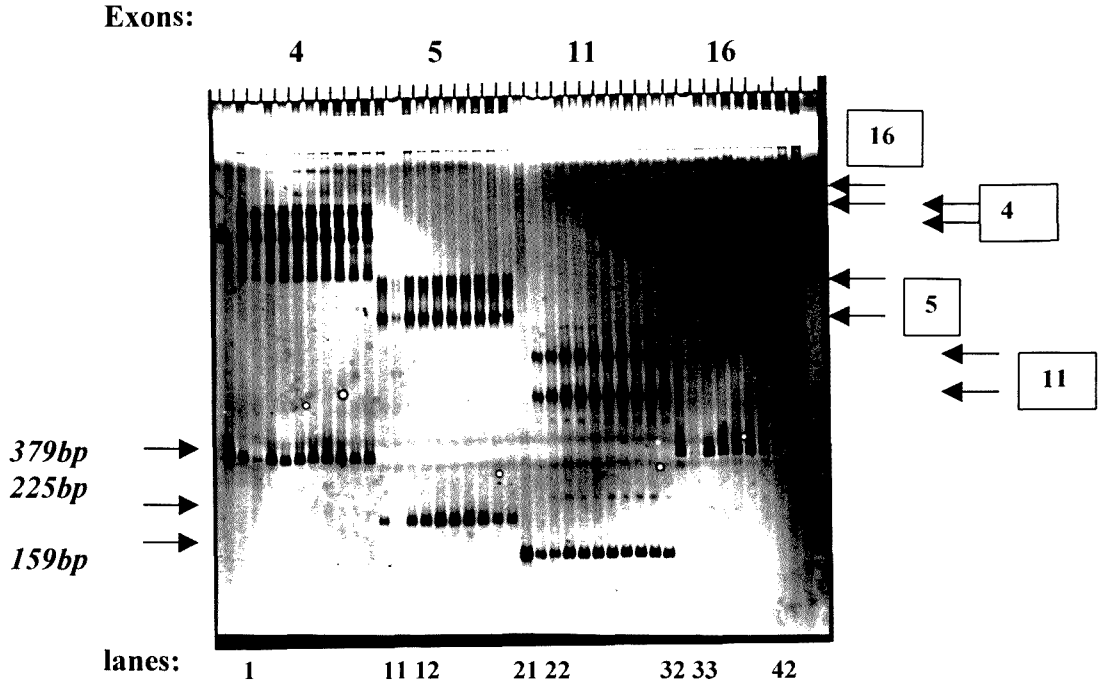
3.2.1.1 PCR-SSCP Analysis

Deoxyribonucleic acid samples of the PFHBI panel, comprising affected and unaffected individuals from pedigrees 1, 2 and 5 (2.1), were analysed under four different gel conditions to optimise detection of any conformation polymorphisms in *GSY1* exons 4, 5, 11, 12 and 16. No mobility-shifts were observed in exons 4, 5, 11, 16 (Figure 3.16) and 12 (results not shown), indicating the absence of any sequence variation in *GSY1* exons screened.

The ideal size of PCR product for detection of any sequence variations by PCR-SSCP, however, is 100bp-200bp (Sheffield *et al.*, 1993), which means that any sequence variations in larger fragments analysed could be left undetected. Therefore, due to these limitations, the PCR-amplified products of exons 4 (379bp) and 16 (361bp) were reanalysed using RE-based PCR-SSCP.

Figure 3.15: CLUSTAL V multiple sequence alignment of GSY1
Muscle glycogen synthase, GSY1 amino acid sequences transcribed and translated from GSY1*'s 16 exons and then aligned with the their homologues from human muscle GSY1 (GGSTe), rat muscle GSY1 (R_GGSTe), mouse muscle GSY1 (M_Gsy), the liver glycogen synthase, GSY2, isozymes from human liver (Liver), rat liver (R_Liver) and the two yeast glycogen synthase isozymes (Y_GsyI and Y_GsyII).*





Denatured samples:

1. 6.313 (U)	*6. 6.415 (A)*
2. 6.314 (A)	*7. 4.303 (A)*
3. 6.412 (A)	*8. 9.305 (U)*
4. 6.413 (U)	*9. 10.301 (A)*
5. 6.414 (U)	*10. 12.302 (U)*

Figure 3.16: PCR-SSCP mutation screening of selected* GSYI *exons (4, 5, 11 and 16)
Representative PCR-SSCP analysis results of PFHBI pedigree 2 samples on a 10% polyacrylamide gel with glycerol: Lanes 1 and 22 = undenatured samples of exons 4 and 11, respectively; lanes 2-11, heat-denatured samples of exon 4; lanes 12-21, heat-denatured samples of exon 5; lanes 23-32, heat-denatured samples of exon 11 and lanes 33-42, heat-denatured samples of exon 16. The arrows on the right indicate single strand conformers for each exon and the single bands at the bottom of each lane represent the double strand molecules. The arrows on the left indicate fragment sizes, in bp, of the double strand molecules of each exon. A = PFHBI-affected. U = unaffected first-degree relative.

3.2.1.2 Restriction enzyme-based PCR-SSCP Analysis

GENEPRO restriction enzyme site analysis was used to identify two restriction enzyme recognition sequences, which, on digestion of *GSY1* exons 4 or 16 with the appropriate enzyme, would yield DNA fragments of an ideal size, spanning the exon, for PCR-SSCP analysis. *Hinf*I and *Hpa*II restriction enzyme recognition sequences were detected in *GSY1* exon 4, which on digestion of the PCR-amplified product would generate fragments of, **5'** -47bp + 27bp + 120bp + 185bp **-3'** (379bp) and **5'** -213bp + 65bp + 101bp **-3'** (379bp), respectively (Figure 3.17a). For *GSY1* exon 16, *Mbo*II and *Sty*I restriction enzyme recognition sequences, which would generate fragment of **5'**- 154bp + 207bp **-3'** (361bp) and **5'**- 193bp + 168bp **-3'** (361bp), respectively, were identified (Figure 3.17a).

The restriction enzyme-digested PCR-amplified products of *GSY1* exon 4 and exon 16 of the PFHBI panel were re-analysed for SSCPs under four different PAGE conditions. A mobility shift, in the 120bp fragment generated by *Hinf*I-digestion of PCR-amplified exon 4 DNA, was observed in an unaffected individual (6.313) (Figure 3.17b). Further analysis of the same panel of samples in 0.35x MDE™ (section 2.6) revealed a mobility shift in the 185bp fragment of the *Hinf*I-digested exon 4 PCR product of individual 6.313 and, in addition, the same mobility shift was seen in an affected individual (6.412) (Figure 3.17c). No mobility shifts were detected in the *Hpa*II (Figure 3.17b) and *Mbo*II and *Sty*I digested fragments of exons 4 and 16, respectively (results not shown).

To further investigate the observed mobility shifts, *Hinf*I-based PCR-SSCP analysis was performed in a larger sample (n = 40), constituting PFHBI-affected individuals and their unaffected first-degree relatives. Analysis of the *Hinf*I-digested products in a 0.35X MDE gel, on two separate occasions, revealed different sets of single-strand conformers. Furthermore, none of the single strand conformers, for either the 120bp or the 185bp *Hinf*I-digestion products, were consistent and thus could not be scored for phenotype-genotype correlations (Figure 3.17d). However, no single set of single strand conformation polymorphism was, exclusively, associated with the affected phenotype, thus excluding the observed mobility shift as a disease-associated mutation.

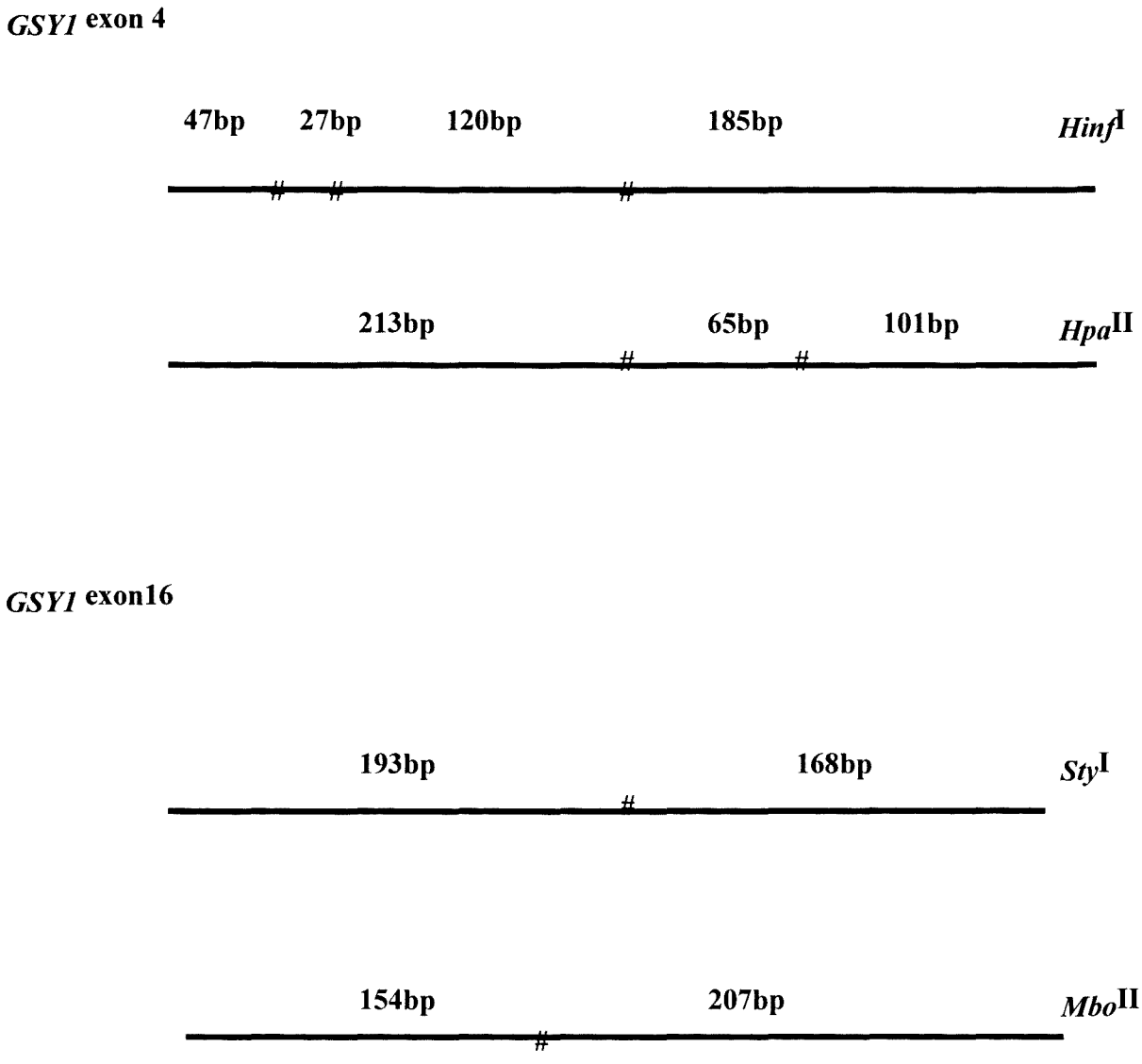
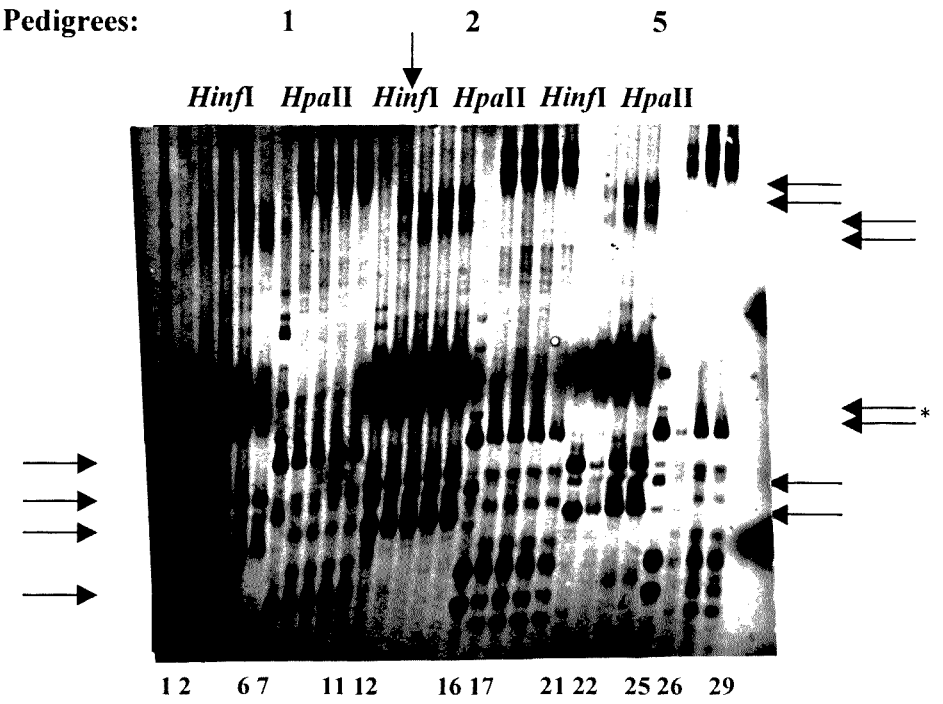


***Figure 3.17a: Restriction enzyme map of PCR-amplified exons 4 and 16 of* GSY1.**
The position of HinfI *and* HpaII *recognition sequences (#) on exon 4(379bp) and* StyI *and* MboII *recognition sequences on exon 16 (361bp), together with the size, in bp, of the digestion products generated from PCR-amplified genomic DNA, is shown.*



Digested and denatured samples:

Pedigree:	***1***	***2***	***5***
1.	*1.305 (U)*	*6.313(U)*	*5.101 (U)*
2.	*1.410 (U)*	*6.412 (A)*	*5.202 (U)*
3.	*1.504 (A)*	*6.415 (A)*	*5.301 (A)*
4.	*1.510 (A)*	*12.302 (U)*	*5.302 (A)*
5.	*1.511 (U)*	*9.305 (U)*	

Figure 3.17b: RE-based/PCR-SSCP analysis of **GSYI** ***exon 4 digested with*** **Hinf*I*** ***and*** **Hpa*II*.** *Lane 1 = λPstI size marker; lanes 2-6, 12-16 and 22-25, represent the* HinfI *digested samples from pedigrees 1, 2 and 5, respectively. Lanes 7-11, 17-21 and 26-29 represent the* HpaII*-digested samples from pedigrees 1, 2 and 5, respectively. The first lane in each series represents a restriction enzyme digested but undenatured sample. The arrows (left) indicate the single stranded conformers for each digestion fragment, with the size of the undenatured double strand molecule indicated on the right of the gel. The double arrows (right) indicate the single strand conformers for all fragments greater than 100bp. A mobility-shift was observed lane 13(↓) in the 101bp,* HinfI*-digested fragment from an unaffected individual (⟸ *). A = PFHBI-affected. U = unaffected first-degree relative.*

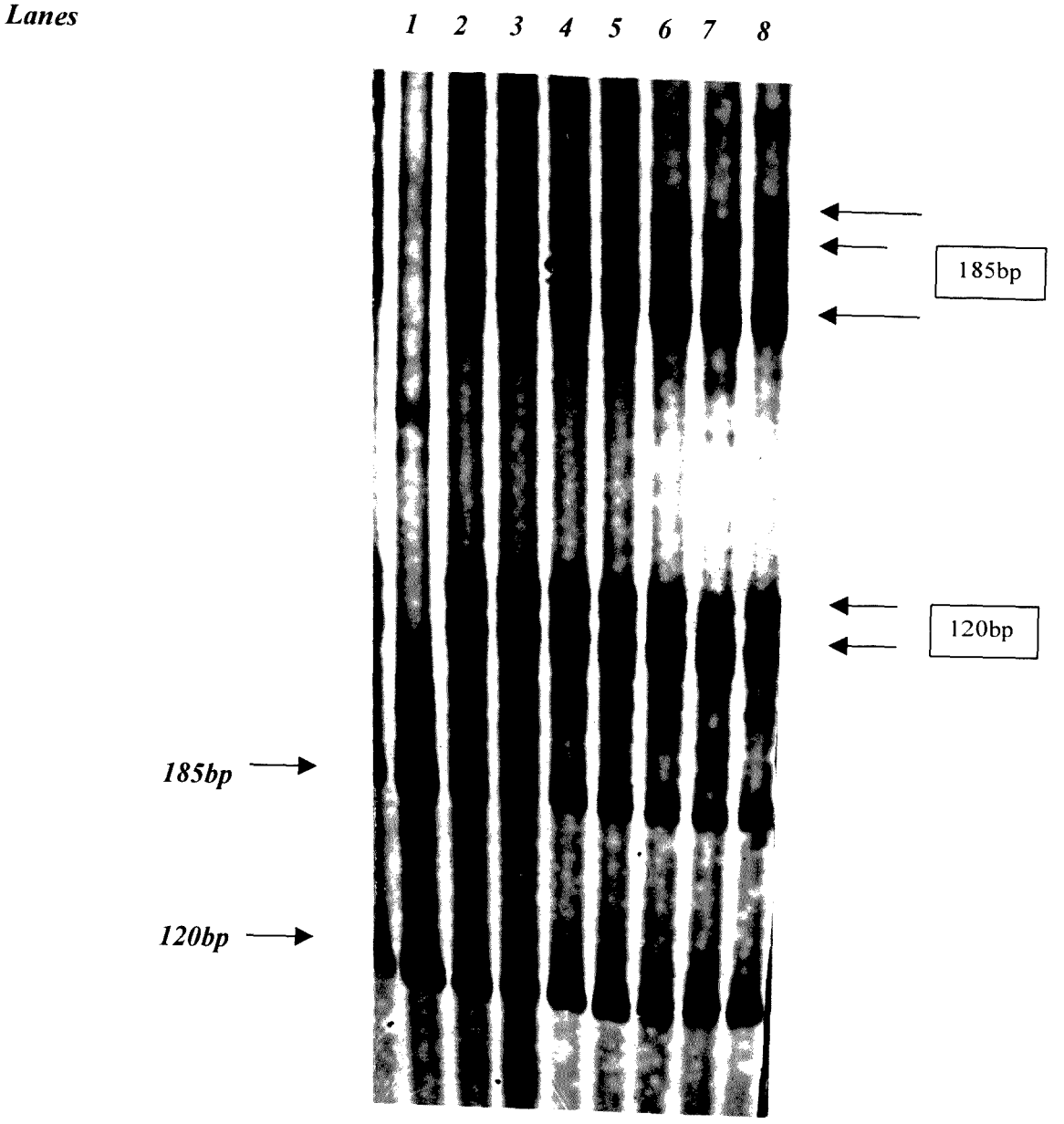


Figure 3.17c: RE-based/PCR-SSCP analysis of pedigree 2,* GSYI *exon 4 digested with* Hinf*I. *SSCP analysis of exon 4 PCR-product digested with* Hinf*I on a 0.35x MDE gel. Mobility-shifts were observed in both unaffected (6.313) and affected (6.412) individuals, lanes 6 and 7 respectively. Lane 1 = undenatured* Hinf*I-digested sample, lanes 2-8 = heat-denatured samples. The arrows on the left indicate the double strand fragments of the digested products. The arrows on the right indicate the single strand conformers of each digestion product (boxed).*

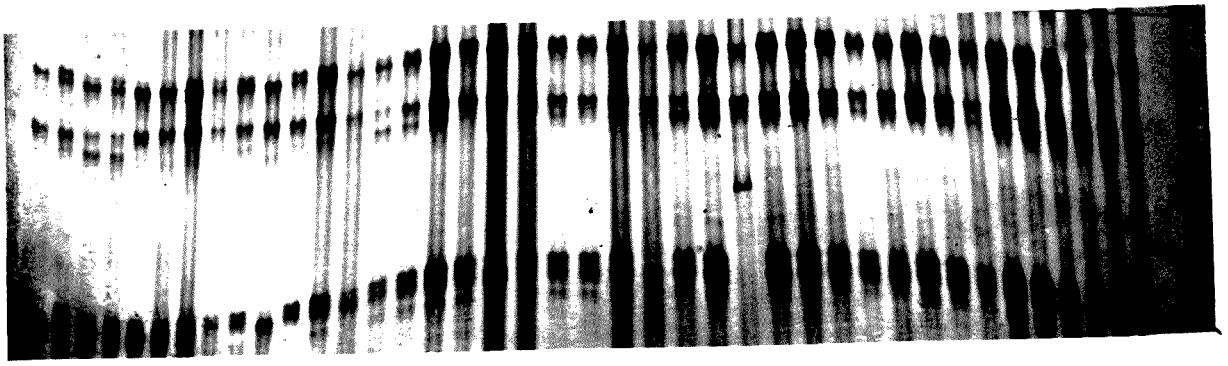


***Figure 3.17d: RE-based/PCR-SSCP analysis of GSY1 exon 4 digested with* HinfI.**
DNA samples of forty individuals from selected PFHBI families in pedigree 2, comprising both the affected individuals and their unaffected relatives, were PCR-amplified and the product digested with HinfI. *Thereafter, they were analysed for mobility shifts on a 0.35x MDE gel. Different conformation polymorphisms were observed, however, none of these was exclusively associated with the disease phenotype.*

3.2.2 Histidine-Rich Calcium Binding Protein (*HRC*)

Exons 2-6 of *HRC*, encoding the conserved C-terminal (Hofmann *et al.*, 1991), were screened for possible PFHBI-causing mutations by direct sequencing. These regions indicated in figure 2.5, were analysed for sequence variations by directly comparing sequences from PFHBI-affected individuals with their unaffected first-degree relatives. In the following sections, the results obtained for each region analysed are presented.

Two adjacent sequence variations, an A to T and a G to T substitution (**AG** to **TT**) were detected in intron B of fragment 2-B-3 in an unaffected individual from pedigree 5 (Figures 3.19a and b). Further analysis of this fragment by GENEPRO indicated that these sequence changes would result in the loss of both *Xho*II [(**A**/G)GATC(T/C)] and *Sau3A*1 (**G**ATC) restriction enzyme recognition sites in a homozygous TT (**TT**ATC). Conversely, in an AG homozygote both sites would be present. However, no differences were observed in the PCR-amplified *Xho*II-digested (Figure 3.18c) or *Sau3A*1-digested PFHBI affected and control samples (results not shown), which included samples from unaffected first-degree relatives and one unrelated individual.

Two additional, adjacent sequence changes, both A to C transversions (G**AA** to G**CC**), were detected in exon 3 of fragment 2-B-3 in an unaffected individual from pedigree 2 (Figure 3.19a). These nucleotide changes would cause an amino acid substitution of a glutamic acid residue by an alanine residue. No sequence variations were detected in fragments 4-D-5-E and 4-D-5-E-6. The double A-C transversions in exon 3 (Figures 3.20a and b), however, did not affect a restriction enzyme recognition sequence. Thus, further analysis of this sequence change was performed using PCR-SSCP. No mobility shifts were detected on analysis of exon 3 in a panel, which consisted of the pedigree 2 samples in which the sequence changes were observed and additional paired PFHBI-affected and unaffected first-degree relatives from pedigrees 1 and 5 (Figure 3.19c).

For complete screening of the coding regions, as well as the intron/exon junctions missed in the initial round of sequencing, primers were designed using the newly acquired intronic sequence, in order to PCR-amplify and sequence the regions designated 1-A-2-B and D-5-E-6 (Figure 2.1). Neither of these regions showed any sequence variations between affected and unaffected individuals' samples (results not shown). The PCR-amplified fragment 1-A-2-B, containing

intron B, from pedigrees 5 and 2 was digested with *Xho*II and *Sau3A*1 restriction enzymes, respectively. The generated restriction enzyme digestion products were the same in both PFHBI-affected individuals and their unaffected close relatives used as control samples (results not shown).

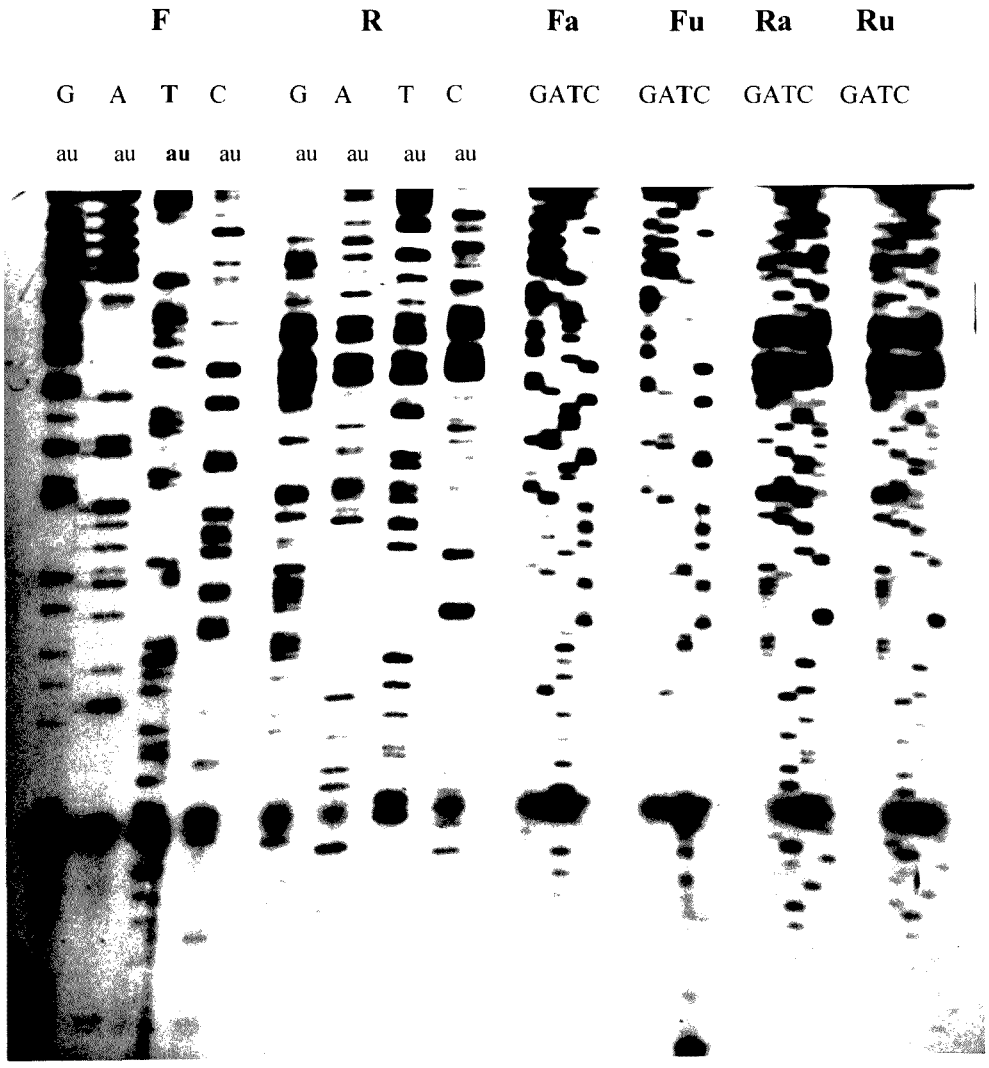


Figure 3.18a: Autoradiograph of the sequence of **HRC** ***region 2-B-3 in PFHBI pedigree 5.***
To facilitate rapid detection of sequence changes, affected and unaffected individuals' samples were loaded alongside each other in G, A, T and C tracts, in addition to the normal sequencing procedure in which the adjacent G, A, T or C tracts represented one individual. F = sense strand and R = antisense strand. a = PFHBI-affected (5.301); u = unaffected close relative (5.101). Sequence variations were observed in the Forward T-tract (bold) between the affected and unaffected individuals.

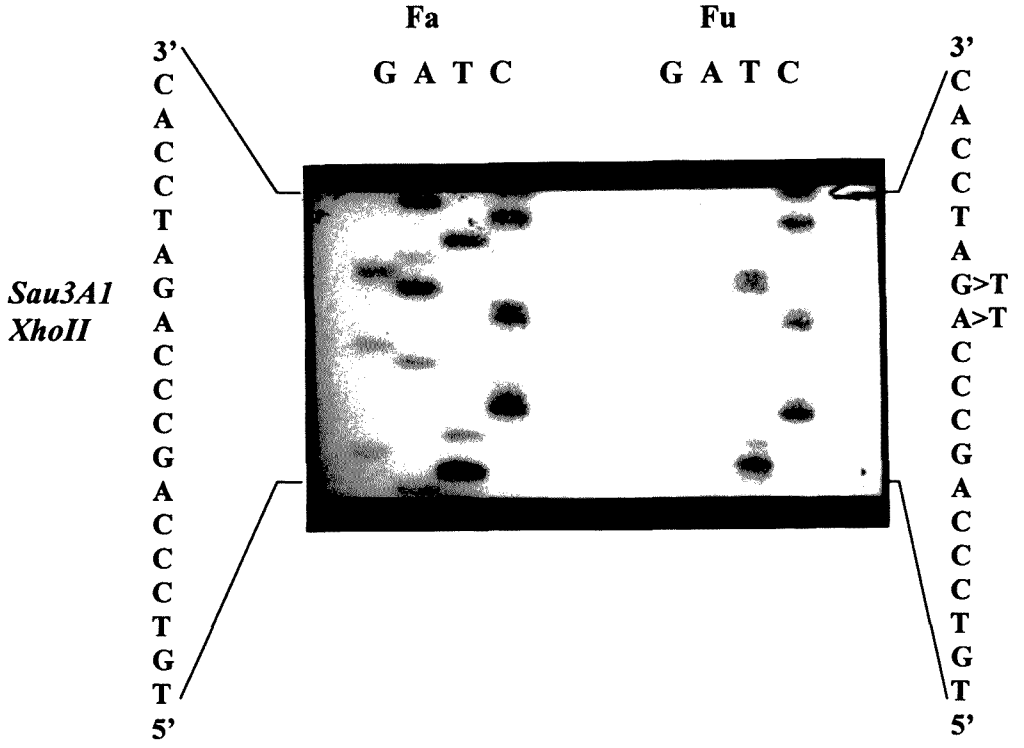


Figure 3.18b: Magnified sequence autoradiograph of* HRC *fragment 2-B-3 magnified to show the sequence variations seen in Figure 3.18a.

Partial sequence of the intronic sequence flanking the coding strand of intron B of HRC *in an individual, heterozygous for the double GA/TT substitution (right) compared to that of a GA homozygous individual (left). The A to T and the G to T substitutions result in the loss of the* Xho*II and* Sau3A*1 restriction enzyme recognition sites, respectively. F = sense strand and R = antisense strand. a = PFHBI-affected; u = unaffected close relative.*

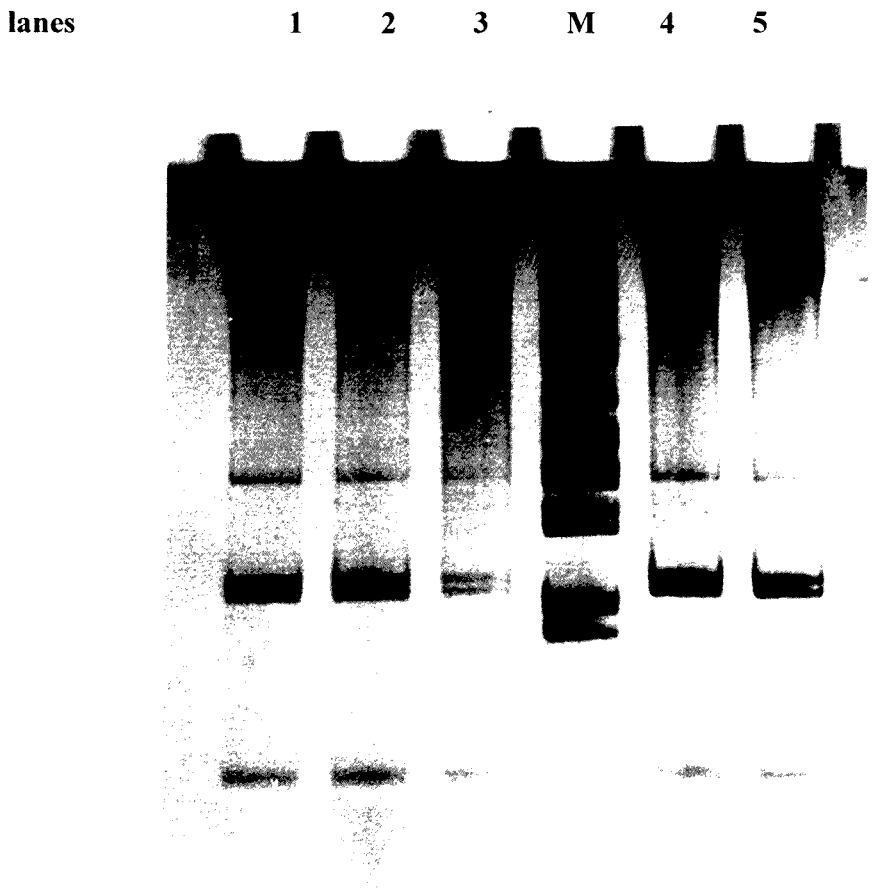


Figure 3.18c: Restriction enzyme analysis of sequence variation in* HRC *fragment 2-B-3.
The PCR-amplified product of HRC *fragment 2-B-3 digested with* XhoI *restriction endonuclease. Lane 1 = control sample (unrelated to the PFHBI families), lanes 2 and 3 = paired PFHBI-affected (1.504) and an unaffected first-degree relative (1.305) from pedigree 1. Lanes 4 and 5 = paired PFHBI-affected individual (5.301) and an unaffected first-degree relative (5.101) from pedigree 5, in whom the sequence variations were observed (shown in figures 3.19a and b).*

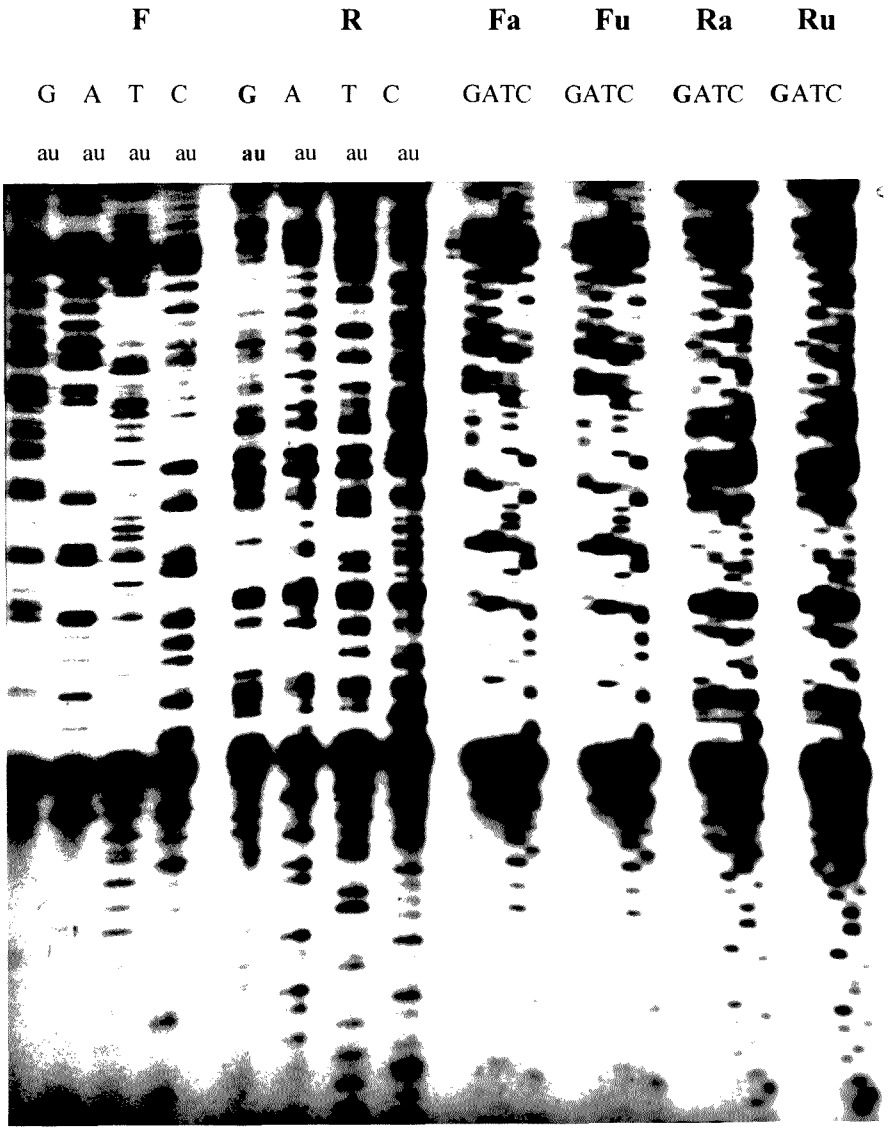


Figure 3.19a: Autoradiograph of the sequence of ^{HRC} ***region 2-B-3 in PFHBI pedigree 2.***
*To facilitate rapid detection of sequence changes, affected and unaffected individuals' samples were loaded alongside each other in G, A, C and T tracts, in addition to the normal sequencing procedure in which the adjacent G, A, T or C tracts represented one individual. F = sense strand and R = antisense strand. . a = PFHBI-affected (10.301); u = unaffected close relative (9.305). Sequence variations, indicated by the arrows, were observed in the Reverse G-tract (**bold**) between the affected and unaffected individuals.*

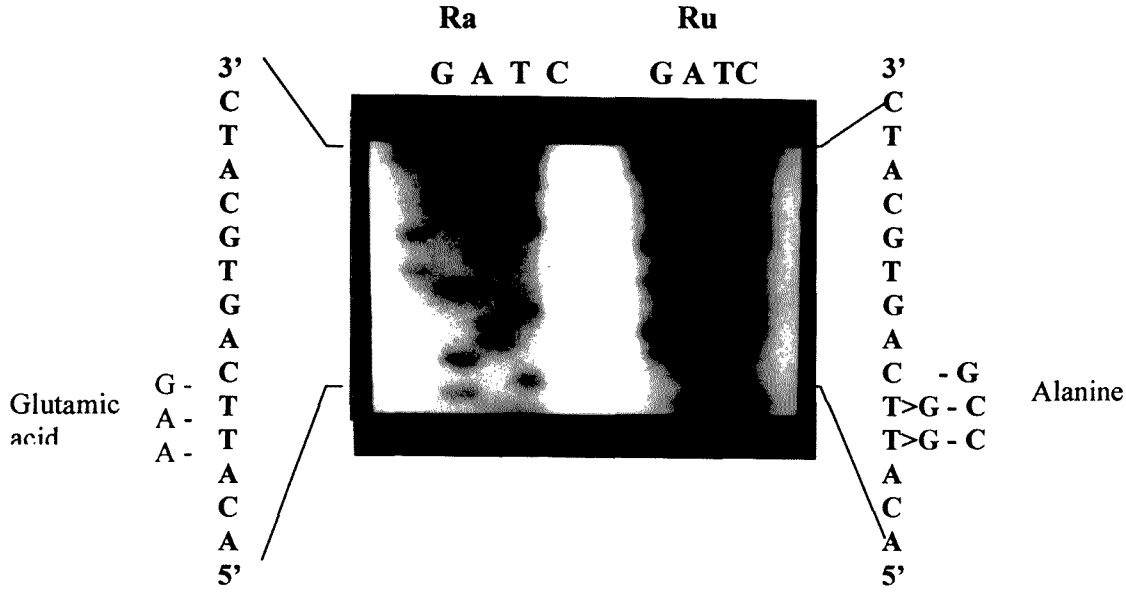
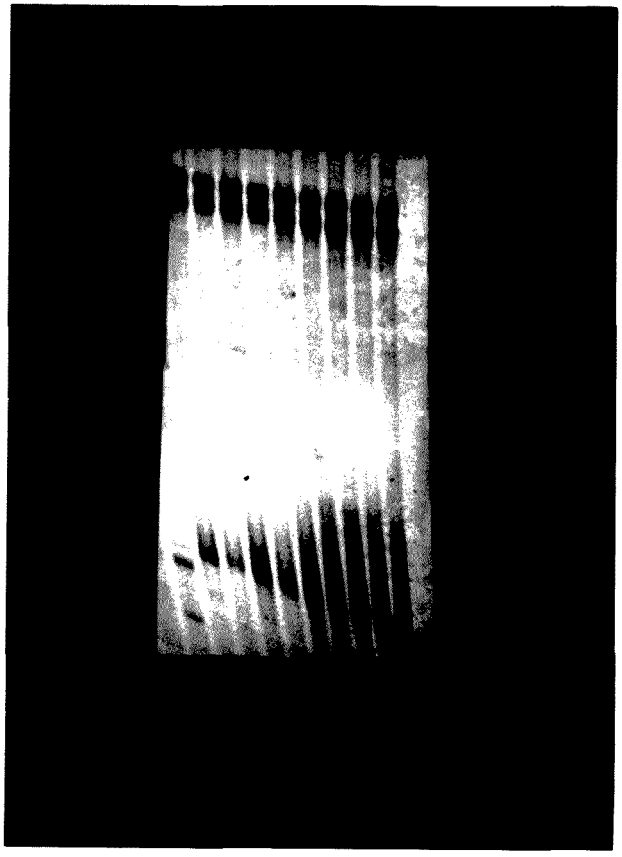


Figure 3.19b: Magnified sequence autoradiograph of* HRC *fragment 2-B-3 magnified to show the sequence variations seen in Figure 3.19a.
Partial sequence of the exonic sequence flanking the anti-sense strand of exon 3 of HRC *in an individual heterozygous for the double* ***TT/GG*** *substitution (right). a = PFHBI-affected; u = unaffected close relative. The sequence in blue is the complement of the antisense (sense) strand from which is the coding strand.*



Denatured samples:

1. 1.305 (U)	***2. 1.504 (A)***
3. 5.202 (U)	***4. 5.301 (A)***
5. 6.308 (A)	***6. 6.406 (U)***
7. 9.305 (U)	***8. 10.301 (A)***
9. Control (U)	

Figure 3.19c: PCR SSCP analysis of HRC ***exon 3 (B-3-C) in a selection of PFHBI pedigree 2 family members.***

Fragment B-3-C of HRC*, which contains exon 3, was analysed for mobility shifts (on an 8% mildly denaturing polyacrylamide gel) to confirm the sequence variations seen on sequencing (shown in figures 3.20a and b). SSCP = single strand conformers; ds = double strand molecule. A = PFHBI-affected, U = unaffected first-degree relative.*

CHAPTER 4
DISCUSSION

4.1 DEVELOPMENT OF GENETIC POLYMORPHIC MARKERS

Initially, the *PFHBI* locus was mapped to a 10cM interval on chromosome 19q13.3 (Brink *et al.*, 1995), which through fine mapping efforts was subsequently reduced to 7cM (De Jager, personal communication) (Figure 1.5). During the course of this project, the PFHBI target area was further reduced from 7cM to 4cM, within which is a core haplotype spanning 1cM (Arieff, personal communication). Originally, di- and trinucleotide STR polymorphic markers were sought throughout the greater 7cM-search region to generate high-density maps, which would help refine the target region further, thus providing a more focussed area expediting the positional cloning of the disease-causing gene. Cosmid clones, spaced at 500kb and spanning chromosome 19q13.3, were used to develop STR polymorphic markers.

As part of an integrated approach, in an effort to increase the resolution of the genetic maps, the development of tetranucleotide STR markers was also pursued in the present study. Tetranucleotide STRs were selected because they have increased fidelity of PCR amplification, which results in better separation of alleles, and thus, they are easier to genotype manually (Sheffield *et al.*, 1995; The Utah Marker Development Group, 1995). The $(AAAT)_n$ STR, reportedly one of the abundant tetranucleotides that also give clear results on genotyping (Economou *et al.*, 1990 and The Utah Marker Development Group, 1995), was selected for new marker development.

Sixteen out of nineteen cosmid clones that were hybridized with the radiolabelled $(AAAT)_{10}$ probe gave clear positive signal, while the remainder gave ambiguous results (Figure 3.1). Eleven of the positively hybridising clones were available as *Sau3A*1-digested fragments on the Southern blot and, on hybridisation with the radiolabelled $(AAAT)_{10}$ probe, nine gave positive signals (Figure 3.2). The Southern blot hybridisation result confirmed the dot blot result, although in some instances the signal intensities did not correspond, as shown with cosmid 29395 (Figures 3.1B, #1 and 3.2, lane 9). The intensity of the signal on hybridisation of the probe to the appropriate template is influenced by the length of the STR, in addition to the availability of single strand template, which is achieved by the denaturation of the double strand DNA. Therefore, a possible explanation for the difference in signal intensity between dot and Southern blot could be that because cosmid DNA on the Southern blot had been digested into smaller fragments, which denature more readily, the $(AAAT)_n$ STR would have been exposed to the probe. In contrast, total cosmid DNA was used on dot blots and although the DNA had been denatured prior to blotting, it is possible that the repeat containing region,

due to incomplete denaturation, was partly shielded from the probe, thus giving a weak positive signal.

The weak signals given by some clones on the Southern blot were interpreted as an indication of shorter STRs (Figure 3.2 lanes 4 and 10). Multiple signals generated from the same cosmid were interpreted as diffuse $(AAAT)_n$ repeats separated by *Sau3A*1 sites (Figure 3.2, lanes 1 and 12). The $(AAAT)_n$ sequence is associated with *Alu*-elements, which are common sequences in the human genome (Zuliani and Hobbs, 1990), thus, it is not inconceivable that there might be other similar sequences in close proximity, onto which the probe may hybridise. Cosmid clones with a single strong signal were selected as they possibly represented a localised longer $(AAAT)_n$ repeat, a category of repeat sequence that has been reported to be more often polymorphic than an interrupted repeat sequence. Therefore, cosmids 20381, 29395 and 24493 (Figure 3.2, lanes 6, 9, and 12, respectively), which gave positive signals on both the dot blot and Southern blot, and in addition, are located within the *PFHBI* target area (Figure 1.5), were then pursued for development of the $(AAAT)_n$ polymorphic markers following either sub-cloning or vectorette PCR-based strategies. The hybridisation procedure allows one to reduce significantly the number of clones to be followed (Baron *et al.*, 1992).

4.1.1 Sub-cloning

The DNA fragment from cosmid 29395 that gave a strong positive signal (Figure 3.2, lane 9) was sub-cloned and sequenced to reveal an interrupted $(AAAT)_n$ sequence (Figure 3.5a). Analysis of the repeat-flanking sequences (Figure 3.5b) for uniqueness prior to designing PCR primers (blast@ncbi.nih.nlm.gov, 1997) showed that part of the 5'-flanking sequence was homologous to an *Alu*-element. This finding was not a surprise, as $(AAAN)_n$ STR sequences are normally found within the poly(A)-tail region, 3' of *Alu*-elements (Zuliani and Hobbs, 1990; Beckmann and Weber, 1992). Furthermore, Beckmann and Weber (1992) reported that these *Alu*-associated microsatellites comprised imperfect repeats. The disadvantage of designing a PCR primer from within an *Alu*-related sequence would be the high background due to non-specific priming from other similar sequences within the human genome. *Alu*-elements comprise approximately 1% of the human genome and are, on average, spaced at 5000bp intervals (Batzer *et al.*, 1990; Economou *et al.*, 1990). This problem was, however, circumvented, since there was sufficient sequence at the 5'-end of the interrupted repeat not homologous to the *Alu*-element, to design a PCR primer (Figure 3.5b).

The interrupted $(AAAT)_n$ repeat locus was PCR-amplified in a panel of 40 unrelated individuals of Mixed Ancestry, which was available in our laboratory and represented a general population sample, and then genotyped to determine whether the repeat was polymorphic. The repeat was not polymorphic within this population (Figure 3.6a). The informativeness of microsatellites is dependent on both the repeat length and whether the repeat is perfect. In general, the informativeness of microsatellites is directly proportional to the repeat length but it is negatively influenced by the complexity of the repeat, with the degree of informativeness decreasing in the order of perfect, imperfect and compound repeats (Weber, 1990 and Edwards *et al.*, 1992b). More specifically, for tetranucleotide STRs, repeat unit lengths of greater than 6 are considered informative (Sheffield *et al.,* 1995). Therefore, the fact that the repeat observed in cosmid 29395 was interrupted and the interrupted runs of $(AAAT)_n$ motifs were less than 6 ($n = 3$ and 4) (Figure 3.5a) probably explains why the repeat was not polymorphic.

An interesting observation was that the repeat locus amplified from cosmid 29395 DNA was smaller than the one amplified in human genomic DNA (Figure 3.6b). The shorter allele in the cosmid could be due to mutation subsequent to cloning, resulting in the deletion of part of the repeat. Philibert *et al.* (1996) have demonstrated this instability in sub-cloned microsatellites, where a sub-cloned CAG triplet repeat (a pathogenic STR) was shorter than the original repeat because of a deletion that occurred on sub-cloning. Recently, Lee *et al.* (1999) have confirmed the relative mutation rates of the non-pathogenic microsatellites (Chakraborty, 1997), by showing that in cultured mammalian cells the relative stability of di- compared to tetranucleotide repeats is always higher and the pathogenic microsatellites (trinucleotide STRs) have the highest mutation rates. An alternative explanation for the smaller size of the repeat sequence in cosmid 29395 is that this region may have been shorter in the individual from whom the original sequence was obtained and cloned, which might represent a polymorphism specific to a particular population.

Although the interrupted $(AAAT)_n$ STR was not polymorphic in the panel tested, it was considered prudent to also genotype it in selected individuals from the PFHBI families. However, here again the repeat locus was not polymorphic. The two bands seen in figure 3.6 represent one allele and its shadow band. Interestingly, for tetranucleotide STRs the two bands are frequently of equal intensity (Economou *et al.*, 1990).

The cosmid 24493 fragment, of approximately 1kb in size, which gave a very strong signal on the Southern blot (Figure 3.2, lane 12 middle band), was also sub-cloned and sequenced using the T3 or T7 pBluescript primers. However, no $(AAAT)_n$ tandemly repeated motif was observed, in the sequence read lengths of 250 bases that had been acquired with each primer, which, therefore, covered about 500 bases of the cloned sequence. Consequently, since only half of the fragment had been sequenced, it was possible that the repeat motif that gave such an intense signal on the Southern blot lay within the unsequenced central region. At this point, the two options available were to design primers from the recovered information and to extend sequencing to reach the putative $(AAAT)_n$ repeat, or to use vectorette PCR to directly identify the repeat without sub-cloning. The latter method was applied for the development of polymorphic $(AAAT)_n$ repeat markers.

4.1.2 Vectorette PCR

The vectorette PCR method allows for the rapid characterisation of microsatellites from large cloned inserts, by targeting the sequence of interest directly, and thereby obviating the need to sub-clone the inserts (section 2.7.6). To this end, characterisation of the $(AAAT)_n$ repeat, within cosmid 24493, was pursued using the vectorette PCR strategy. A PCR product was generated on amplification of the *Alu*I vectorette library using anchored primer $(AAAT)_4$-C and vectorette primer 224 (Table 3.1) and subsequently sequenced using primer 10108. Primer 24493F (repeat-flanking sequence) was designed from the resulting sequence (Figure 3.8b). The second repeat-flanking sequence primer 24493R was then designed from the sequence of a PCR product amplified using primers 24493F and 224 (Figure 3.9b). To characterise the putative repeat sequence harboured within this cosmid, primers 24493F and 24493R were then used to PCR-amplify and sequence the putative repeat locus in the cosmid 24493 *Alu*I library (Figure 3.10b). Unfortunately, no $(AAAT)_n$ repeat motif was detected (Figure 3.10c). To verify that the region amplified from the cosmid library was human genomic DNA, a DNA sample of an individual belonging to a PFHBI family was PCR amplified with primers 24493F and 24493R and the generated products were identical in size to those obtained in the cosmid (Figure 3.10a).

The $(AAAT)_n$ STR harboured in cosmid 20381, as indicated by the single strong signal on the Southern blot (Figure 3.2, lane 6), was also characterised using the vectorette PCR method. A PCR product was obtained on amplification of the *Rsa*I vectorette library using the anchored primer $(AAAT)_4$-G and vectorette primer 224 (Table 3.2), which was subsequently sequenced

and primer 20381F (repeat-flanking sequence) was designed from this information (Figure 3.11c). The second repeat-flanking sequence primer 20381R was designed from the sequence of a PCR product amplified using primers 20381F and 224 (Figure 3.12c). To characterise the putative repeat sequence harboured within this cosmid, primers 20381F and 20381R were then used to PCR-amplify and sequence the repeat locus in cosmid 20381 *Rsa*I library. However, again no $(AAAT)_n$ repeat motif was detected and furthermore, PCR-amplification of human genomic DNA with the same primer pair did not yield any product. The information sequenced with primer 20381R was used to design a new primer 20381R' (Figure 3.13), which was subsequently used with 20381F to PCR-amplify and sequence a new product (Figure 3.14a). No $(AAAT)_n$ repeat motif was detected on sequencing of this new template (Figure 3.14c), however, PCR-amplification of human genomic DNA using primer pair 20381F and 20381R' did generate a product identical in size to that seen in the cosmid (Figure 3.14a).

Considering the intensity of the signals given by both cosmids 20381 and 24493 on the Southern blot (Figure 3.2, lanes 6 and 12), it was surprising that no $(AAAT)_n$ STR was identified. Nevertheless, there are a few possible explanations for the results obtained in the present study. The polymerase chain reaction is influenced by a number of parameters, including, the magnesium ion concentration and primer-annealing temperature. The vectorette PCR relies heavily on the annealing of the repeat-anchored primer to the repeat sequence of interest. Therefore, if the annealing conditions are not optimal, the anchored primer might anneal at the incorrect place(s), and subsequent amplification may generate multiple spurious products. However, this was unlikely to be the case with cosmids 20381 and 24493, as specific, single band PCR products were obtained.

The use of A/T-rich PCR primers is generally not recommended, as they have relatively low annealing temperatures, and thus stringency of the PCR may be reduced (Lindblad *et al.*, 1994). The composition of the primer sequence determines the Tm and, using the formula Tm = 2(A + T) + 4(G + C), primers that are A/T-rich will have a lower Tm than G/C-rich ones (Thein and Wallace, 1986). Thus, an ideal primer sequence should have at least 50-60% G/C to increase the Tm, which will favourably influence the stringency and specificity of the PCR. Hence, using primers that consist of more than 94% A/T nucleotide bases (e.g., $(AAAT)_4$-N) would result in a lowered Tm and lowered stringency of annealing. Therefore, it is possible that because of the low Tm used in vectorette PCR (34°C and 36°C, see table 2.2), the

generated product resulted from the transient annealing of the repeat-anchored primer to the single AAAT motif identified in the different templates (Figures 3.11c and 3.15c).

4.1.3 Subcloning vs vectorette PCR

Using the sub-cloning method, an interrupted $(AAAT)_n$ repeat was identified, however, it was not polymorphic. The successful development of microsatellites using vectorette PCR has been reported in the literature (Rothuizen and Van Raak, 1994; Lench *et al.*, 1996), however, most concerned the development STR other than the A/T-rich microsatellites. Although both methods work well, the vectorette PCR-based strategy is more rapid than sub-cloning. In retrospect, the suggestion for follow-up studies is to either use longer A/T-rich repeat-anchored primers, which would increase the annealing temperature and hence the stringency of the PCR or use other non-A/T-rich primer sequences.

4.2 CANDIDATE GENES: MUTATION SCREENING

4.2.1 Glycogen Synthase (*GSY1*)

Five *GSY1* exons, 4, 5, 11, 12 and 16, out of sixteen were selected, based on their conservation across species (Figure 3.15), and analysed in PFHBI pedigrees 1, 2 and 5 for disease-causing mutations by PCR-SSCP analysis. Although pedigrees 1 and 5 have the same phenotype and share the same haplotype across the *PFHBI* locus, no ancestral link, with each other or the larger pedigree 2, has as yet been established (Brink *et al.*, 1995). Therefore, it was considered prudent to analyse all three pedigrees in case the disease was caused by different mutations in the same candidate gene. No mobility shifts were observed in any of the five exons, in any of the individuals analysed by conventional PCR-SSCP (Figure 3.16).

However, the amplification products of *GSY1* exons 4 and 16, which were 379bp and 361bp, respectively, were re-analysed by RE-based PCR SSCP for mobility shifts that might have been missed, as the fragments were not of ideal size for conventional SSCP analysis. A variety of conformers of the 185bp *Hinf*I-digested fragment of amplified exon 4 were observed in both PFHBI-affected and control samples, however, none of these segregated with the disease (Figure 3.18). Further analysis of the *Hinf*I-digested products of exon 4 on the more SSCP-sensitive 0.35x MDE™ gel also showed different conformers. These conformers were not consistent, hence impossible to score for phenotype-genotype correlation. Furthermore, on analysis of the different conformers, no one set of conformers

was seen only in PFHBI-affected individuals or exclusively in their close relatives and thus, none of the observed conformers were associated with the disease. Taking into consideration that none of the observed mobility shifts had accompanying heteroduplex complexes it was thus concluded that, the reason for the observed variations is possibly because the PCR-amplified *Hinf*I-digested products was able to assume more than one stable form, which appeared on the gel as mobility shifts. No mobility shifts were detected in the *Mbo*II- or *Sty*I-digested exon 16 fragments.

The PCR-SSCP method of mutation detection can detect approximately 80% of sequence variants (Sheffield *et al.*, 1993), however, the application of four different PAGE conditions increases the likelihood of detecting mutations to close to a 100% (Ravnik-Glavac *et al.*, 1994). Therefore, the absence of mobility shifts in *GSY1* exons 4, 5, 11, 12 and 16 analysed under several different conditions, virtually excludes these exons from carrying the PFHBI-causative mutation(s).

4.2.2 Histidine-rich calcium binding protein (*HRC*)

Exons 2-6 of *HRC* encoding the conserved C-terminus were selected and screened for PFHBI-causing mutations. However, as only 10 bases of intronic sequence, representing either the 3' or 5' end of each exon, had been published (Hofmann *et al.*, 1991), it was necessary to use part of the exonic sequence to design PCR primers. Exons 2-5 were very small in length, ranging between 37bp and 71bp, thus, on sequencing, some 5'- or 3'-most exonic sequence might be completely missed because generally reads start at approximately 20-30 bases away from the primer complementary sequence. Furthermore, disease-associated mutations in the splice sites might also be missed. Therefore, primers were designed such that they would amplify regions spanning two to three exons, including the intervening intron(s) (Figure 2.5), with the advantages that adjacent intronic sequence would be obtained while screening some of the splice sites for possible PFHBI-causing mutations. The acquired intronic sequence information was then used to design new PCR primers for complete *HRC* exon screening, including the previously missed intron/exon junctions, which would allow detection of their respective splice-site mutations. From each of the PFHBI pedigrees (Figure 2.1), DNA samples of an affected individual and an unaffected close relative were compared, in order to detect sequence variations Figures 3.19a and 3.20a).

Pedigree 5 Analysis

No sequence variations were observed in exons 2-6 of *HRC* in the pedigree 5 samples that were analysed. An A to T and a G to T double substitution, creating the *Xho*II and *Sau3A*1 restriction enzyme sites, respectively, was observed in intron B (nucleotide positions 16 and 17 at the 5'-end of intron B) (Figure 3.18). The first 22 bases at the 5'-end of introns are considered critical for the exon-splicing machinery of cells, therefore, these sequence variations were within the splice site critical region. However, these sequence changes were not confirmed by digestion using the appropriate restriction enzymes and, thus, excluded as being cause of PFHBI, in all three pedigrees.

Pedigree 2 Analysis

No sequence changes were observed in *HRC* exons 2, 4, 5 and 6 of the pedigree 2 samples analysed. However, a GAA to GCC double substitution, which would cause a predicted glutamic acid to alanine amino acid change, was observed in HRC exon 3 (region 2-B-3) in PCR-amplified from an unaffected individual from pedigree 2 (Figure 3.19). The *HRC* region 2-B-3 of a different pair of PFHBI-affected individual and a close relative, taken from another family in pedigree 2, was PCR-amplified and sequenced. No sequence variations were observed between the normal and affected individuals' samples. Reanalysis of *HRC* exon 3 (region B-3-C) in a selection of PFHBI pedigree 2 family members, including the ones that had previously been sequenced, was done using PCR-SSCP and no mobility shifts were detected.

Exon 1 of *HRC* was not screened for mutations in the present study, previously, De Meeus *et al.* (1995) and Christoffels (1997) had analysed the GAG and GAT triplet repeats within this exon for abnormal expansion, which might cause disease in the Lebanese and South African kindreds. However, no unusual expanded alleles were detected. In the present study, focus was on mutation screening of conserved exons of *HRC* (2-6) and, exon 1, which is not highly conserved, was not prioritised. The next logical step would be to screen the entire sequence of exon 1, which would be a challenging task with the application of the conventional PCR-SSCP analysis, considering its length of 2001bp (Hofmann *et al.*, 1991). However, to exclude exon 1, mutation screening by other methods, e.g., RE-based PCR-SSCP, although a cumbersome undertaking, could be performed.

However, recently, Dr Brian Black of the Cardiovascular Research Institute (UC, San Francisco, USA) contacted our group when he independently decided that *HRC* was a

plausible PFHBI candidate both by function and position. He has the complete sequence of *HRC*, including the unpublished 5' and 3' UTRs, as well as the complete intronic sequences. In a collaborative project, Dr. Black's group is sequencing *HRC* in its entirety, in an PFHBI-affected individual and an unaffected close relative from each of pedigrees 1, 2 and 5.

4.3 Future Prospects

Although the *PFHBI* locus has been reduced to a 4cM region, within which lies a 1cM-spanning common core haplotype seen in all affected individuals, the causative gene has not yet been identified. Therefore, there is still the need for more markers, or sequence tagged sites (STS), within the target search area, which will be used to integrate the genetic and physical maps. When this study was initiated, we had available cosmid clones spaced 500kb apart across the 7cM target interval. Recently, however, we have obtained a BAC and a cosmid contig spanning the 4cM-search area. With the imminent completion of the Human Genome Project (Waterston and Sulston, 1998; Marshall, 1999), the total human genome sequence will be electronically available on the databases. This means that the chromosome 19q13.3 region sequence will be available for use to identify ESTs and other candidate genes within the *PFHBI* locus. In addition, microsatellite sequences within the *PFHBI* locus will be readily electronically accessible hence, future development of polymorphic markers is likely to be *in silico*.

In the last 5 years, ion channel genes have been identified as the causative genes in long QT syndrome (*KVLQT1, HERG, SCN5A*), Brugada's syndrome (*SCN5A*) and a cardiac conduction disorder (SCN5A) (as reviewed by Keating and Sanguinetti, 1996; Jongsma, 1998; Barinaga, 1998; Schott *et al.*, 1999). Located between the PFHBI positional candidate genes, *GSY1* and *HRC*, is a potassium ion channel gene, *KCNA7* (Figure 1.5). This gene was originally mapped in mouse chromosome 7 and, by known synteny between mouse and human chromosomes, was subsequently placed onto human chromosome 19q13.3 (Kalman *et al.*, 1998). Based on the role of ion channels in conduction of the electrical impulse in the heart, *KCNA7* was considered a strong candidate even though its function had not yet been established. Mutation screening of this ion channel gene was hindered by the lack of nucleotide sequence. Recently, however, studies in the murine system have shown that the mouse orthologue of *KCNA7*, *kcna7* (*kv1.7*), is a novel gene that is expressed in mouse heart (Kalman *et al.*, 1998). In contrast to the other six Shaker-related potassium ion channel genes (*kv1.1-1.6*), which have only one exon, *kv1.7* consists of two exons interrupted by a 1.9kb intron. Furthermore, electrophysiological studies of this ion channel in mouse showed that it

had rapid conduction properties similar to those found in the Purkinje fibres of the human ventricular conduction system, although this has not yet been confirmed (Kalman *et al.*, 1998). Therefore, the human homologue, *KCNA7* (*Kv1.7*), remains a compelling candidate as a causative gene for PFHBI. Recently, the PFHBI gene search group has entered into collaboration with Dr Chandy (Kalman *et al.* 1998) to screen *KCNA7* for disease-causing mutations.

Finding the PFHBI-causative gene will allow an improved DNA-based diagnosis that will supersede the current use of genetic markers and can be extended beyond the present limit of a family setting. Furthermore, it will help elucidate the underlying molecular mechanisms. This may lead to the development of more directed therapies, which may have application not only to the rarer forms of familial cardiac conduction disease but also to the more commonly occurring sporadic cases in the general population.

REFERENCES

Alford RL, Hammond HA, Coto I and Caskey CT (1994). Rapid and efficient resolution of parentage by amplification of short tandem repeats. Am J Hum Genet **55**:190-195.

Alpert MA and Flaker GC (1984). Chronic fascicular block. Arch Intern Med **144**:799-802.

Anderson RH, Becker AE, Tranum-Jensen J, Janse MJ (1981). Anatomico and electrophysiological correlations in the conduction system- a review. Brit Heart J **45**:67-82.

Bardien S, Ebenezer N, Greenberg J, Inglehearn CF, Bartman L, Goliath R, Beighton P *et al.* (1995). An eighth locus for autosomal dominant retinitis pigmentosa is linked to chromosome 17. Hum Mol Genet **4**:1459-1462.

Barinaga M (1998). Tracking down mutations that can stop the heart. Science **281**:32 34.

Baron B, Poirier C, Simon-Chazottes D, Barnier C and Guenet J–L (1992). A new strategy useful for rapid identification of microsatellites from DNA libraries with large size inserts. Nucleic Acids Res **20**:3665-3669.

Basson CT, Cowley GS, Solomon SD, Weissman B, Poznanski AK, Traill TA, Seidman C *et al.* (1994). The clinical and genetic spectrum of the Holt-Oram syndrome (heart-hand syndrome). New Eng J Med **330**:885-891.

Batzer MA, Kilroy GA, Richard PE, Shaikh TH, Desselle TD, Hoppens CL and Deininger PL (1990). Structure and variability of recently inserted *Alu* family members. Nucleic Acids Res **18**:6793-6798.

Beckmann JS and Weber JL (1992). Survey of human and rat microsatellites. Genomics **12**:627-631.

Belmont JW (1998). Recent progress in the molecular genetics of congenital heart defects. Clin Genet **54**:11-19.

Botstein D, White RL, Skolnick M and Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet **32**:314-331.

Brink AJ and Torrington M (1977). Progressive familial heart block – two types. S Afr Med J **52**:53-59.

Brink PA. A molecular genetic approach to the aetiology of cardiovascular disease in South Africa with reference to progressive familial heart block. Doctor in Philosophy dissertation, University of Stellenbosch, 1997.

Brink PA, Ferreira A, Moolman JC, Weymar HW, Van der Merwe P-L and Corfield VA (1995). Gene for progressive familial heart block type I maps to chromosome 19q13. Circulation **91**:1633-1640.

Brink PA, Moolman JC, Ferreira A, De Jager T, Weymar HW, Martell RW, Torrington M, Van der Merwe P–L and Corfield VA (1994). Genetic linkage studies of progressive familial heart block, a cardiac conduction disorder. S Afr J Science **90**:236-240.

Browner MF, Nakano K, Bang AG and Fletterick RJ (1989). Human muscle glycogen synthase cDNA sequence: a negatively charged protein with an asymmetric charge distribution. Proc Natl Acad Sci, USA **86**:1443-1447.

Buard J, Bourdet A, Yardley J, Dubrova Y and Jeffreys AJ (1998). Influences of array size and homogeneity on minisatellite mutation. EMBO J **17**:3495-3502.

Campbell M. Biology. Benjamin/Cummings 2nd edition (Carlifonia, USA) 1990, p148-149.

Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, Monros E, *et al.* (1996). Friedreich's Ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science **271**:1427-1427.

Canzian F, Ushijima T, Serikawa T, Wakabayashi K, Sugimura T and Nagao M (1994). Instability of microsatellites in rat colon tumors induced by heterocyclic amines. Cancer Res **54**:6315-6317.

Chakraborty R, Kimmel M, Stivers DN, Davison LJ and Deka R (1997). Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc Natl Acad Sci, USA **94**:1041-1046.

Charlesworth B, Sniegowski P and Stephan W (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. Nature **371**:215-220.

Christoffels AG. Identification of novel CA repeat markers for use in fine mapping of the PFHBI gene and screening of the HRC candidate gene. Master of Science in Medical Sciences thesis, University of Stellenbosch, 1997.

Cohen D, Chumakov I and Weissenbach J (1993). A first-generation physical map of the human genome. Nature **366**:698-701.

Colette D, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, *et al.* (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature **380**:152-154.

Collins FS (1992). Positional cloning: let's not call it reverse anymore. Nature Genet **1**:3-6.

Cooper DN and Clayton JF (1988). DNA polymorphism and the study of disease associations. Hum Genet **78**:299-312.

Cornelis F, Hashimoto L, Loveridge J, MacCarthy A, Buckle V, Julier C and Bell J (1992). Identification of a CA repeat at the *TVRA* locus using yeast artificial chromosomes: a general method for generating highly polymorphic markers at chosen loci. Genomics **13**:820-825.

Cull P. The sourcebook of medical illustration (Section3, page 69). The Parthenon Publishing Group (1989).

Dai K–S and Liew C–C (1998). Characterization of a novel gene encoding zinc finger domains identified from expressed sequence tags (ESTs) of a human heart cDNA database. J Mol Cell Cardiol **30**:2365-2375.

Dams E, Van de Kelft EJZ, Martin J–J, Verlooy J and Willems PJ (1995). Instability of microsatellites in human gliomas. Cancer Res **55**:1547-1549.

Davies MJ. The conduction system of the heart, Butterworths (London) 1971.

De Jong PJ, Yokabata K, Chen C, Lohman F, Pederson L, McNinch J and Van Dilla M (1989). Human chromosome-specific partial digest libraries in lambda and cosmid vectors. Cytogenet Cell Genet **51**:985.

De Jong F, Opthof T, Wilde AAA, Janse MJ, Charles R, Lamers WH and Moorman AFM (1992). Persisting zones of slow impulse conduction in developing chicken hearts. Circ Res **71**:240-250.

De Meeus A, Stephan E, Debrus S, Jean M–K. Loiselet J, Weissenbach J, Demaille J and Bouvagnet P (1995). An isolated cardiac conduction disease maps to chromosome 19q. Circ Res **77**:1-6.

DeForest, RE and Hanover, NH (1955). Four cases of "benign" left bundle branch block in the same family. Am Heart J **51**:398-404.

Devine SE and Boeke JD (1994). Efficient integration of artificial transposons into plasmid targets in vitro: a useful tool for DNA mapping, sequencing and genetic analysis. Nucleic Acids Res **22**:3765-3772.

Dolf G, Glowatzki M–L and Gaillard C (1991). Searching for genetic markers for hereditary diseases in cattle by means of DNA fingerprinting. Electrophoresis **12**:109-112.

Dubin D. Rapid interpretation of EKG's. Cover Publishing Company, 4th edition (Florida, USA) 1989.

Economou EP, Qergen AW, Warren AC and Antonarakis SE (1990). The polydeoxyadenylate tract of *Alu* repetitive elements is polymorphic in the human genome. Proc Natl Aca Sci USA **87**:2951-2954.

Edwards A, Civitello A, Hammond HA and Caskey CT (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet **49**:746-756.

Edwards A, Hammond HA, Jin L, Caskey CT and Chakraborty R (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics **12**:241-253.

Esscher E, Hardell L–I and Michaelsson M (1975). Familial, isolated, complete right bundle branch block. Brit Heart J **37**:745-747.

Fahy GJ, Pinski SL, Miller DP, McCabe N, Pye C, Walsh MJ and Robinson K (1996). Natural history of isolated bundle branch block. Am J Cardiol **77**:1185-1190.

Fairfax AJ and Lambert CD (1975). Familial late onset heart block: A clinico-pathological study. Age and Ageing **4**:202-208.

Fries R, Eggen A and Stranzinger G (1990). The bovine genome contains polymorphic microsatellites. Genomics **8**:403-406.

Gastier JM, Brody T, Pulido JC, Businga T, Sunden S, Hu X, Maitra S *et al.* (1996). Development of a screening set for new (CAG/CTG)n dynamic mutations. Genomics **32**:75-85.

Gazes, PC, Culler, RM, Taber E and Kelly TE (1965). Congenital familial cardiac conduction defects. Circulation **XXXII**:32-34.

Gordon LA, Bergmann A, Christensen M, Danganan L, Lee DA, Ashworth LK, Nelson DO, *et al.* (1995). A 30-Mb metric fluorescence in situ hybridization map of human chromosome 19q. Genomics **30**:187-192.

Gorza L, Vettore S and Vitadello M (1994). Molecular and cellular diversity of heart conduction system myocytes. Trends in Cardiovascular Med **4**:153-159.

Gourdie RG, Kubalak S and Mikawa T (1999). Conducting the embryonic heart: orchestrating development of specialized cardiac tissues. Trends in Cardiovasc Med **9**:18-26.

Graber HL, Unverferth DV, Baker PB, Ryan JM, Baba N and Wooley CF (1986). Evolution of hereditary cardiac conduction and muscle disorder: A study involving a family with six generations affected. Circulation **7**:21-35.

Greenspahn BR, Denes P, Daniel W and Rosen KM (1976). Chronic bifascicular block: evaluation of familial factors. Ann Intern Med **84**:521-525.

Hamada H and Kakunaga T (1984). Potential Z-DNA forming sequences are highly dispersed in the human genome. Nature **298**:396-398.

Harley HG, Brook JD, Rundle SA, Crow S, Reardon W, Buckler AJ, Harper PS *et al.* (1992). Expansion of unstable DNA region and phenotypic variation in myotonic dystrophy. Nature **355**:545-551.

Hauge XY and Litt M (1993). A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. Hum Mol Genet **2**:411-415.

Hawley RJ, Milner MR, Gottdiener JS and Cohen A (1991). Myotonic heart disease: A clinical follow-up. Neurology **41**:259-262.

Hearne CM, Ghosh S and Todd JA (1992). Microsatellites for linkage analysis of genetic traits. Trends in Genet **8**:288-294.

Hofmann SL, Brown MS, Lee E, Pathak RK, Anderson RGW and Goldstein JL (1989). Purification of a sarcoplasmic reticulum protein that binds Ca^{2+} and plasma lipoproteins. J Biol Chem **264**:8260-8270.

Hofmann SL, Topham M, Hsieh C-L and Francke U (1991). cDNA and genomic cloning of HRC, a human sarcoplasmic reticulum protein, and localization of the gene to human chromosome 19 and mouse chromosome 7. Genomics **9**:656-669.

Howeler CJ, Busch HFM, Geraedts JPM, Niermeijer MF and Staal A (1989). Anticipation in myotonic dystrophy: fact or fiction? Brain **112**:779-797.

Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK *et al.* (1995). An STS-based map of the human genome. Science **270**:1945-1999.

Hunter AGW, Jacob P, O'Hoy K, MacDonald I, Mettler G, Tsilfidis C and Korneluk RG (1993). Decrease in the size of the myotonic dystrophy CTG repeat during transmission from parent to child: implications for genetic counselling and genetic anticipation. Am J Med Genet **45**:401-407.

Hurst JW. The heart, arteries and veins. McGraw-Hill Book Co. 4th edition (USA) 1978.

Husson GS, Blackman MS, Rogers MC, Bharati S and Lev M (1973). Familial congenital bundle branch system disease. Am J Cardiol; **32**:365-369.

James TN (1970). Cardiac conduction system: fetal and postnatal development. Am J Cardiol **25**:213-225.

James TN, McKone RC and Hudspeth AS (1975). De subitaneis morbitus- Familial congenital heart block. Circulation **51**:379-388.

James TN, St. Martin E, Willis III PW and Lohr TO (1996). Apoptosis as a possible cause of gradual development of complete heart block and fatal arrhythmias associated with absence of the A-V node, sinus node, and internodal pathways. Circulation **93**:1424-1438.

Jarcho JA, McKenna W, Pare JAP, Solomon SD, Holcombe RF, Dickie S, Levi T *et al.* (1989). Mapping a gene for familial hypertrrophic cardiomyopathy to chromosome 14q1. N Engl, J Med; **321**:1372-1378.

Jeffreys AJ, Wilson V and Thein SL (1985). Hypervariable 'minisatellite' regions in human DNA. Nature **314**:67-73.

Jennings C (1995). How trinucleotide repeats may function. Nature **378**:127.

Jongsma HJ (1998). Sudden cardiac death: a matter of faulty ion channels? Current Biology **8**:R568-R571.

Kalman K, Nguyen A, Tseng-Crank J, Dukes ID, Chandy G, Hustad CM, Copeland NG *et al.* (1998). Genomic organization, chromosomal localization, tissue distribution, and biophysical characterization of a novel mammalian *Shaker*-related voltage-gated potassium channel, Kv1.7. J Biol Chem **273**:5851-5857.

Kaslow HR and Lesikar (1984). Isozymes of glycogen synthase. FEBS Lett **172**:294-298.

Katz AM (1993). Cardiac ion channels. New Engl J Med **328**:1244-1251.

Keating MT and Sanguinetti MC (1996). Molecular insights into cardiovascular disease. Science **272**:681-688.

Kennel AJ, Callahan JA, Maloney JD and Zajarias A (1981). Adult-onset familial infra-Hisian block. Am Heart J **102**:447-452.

Khatib ZA, Inaba T, Valentine M and Look AT (1994). Chromosomal localization and cDNA cloning of the human *DBP* and *TEF* genes. Genomics **23**:344-351.

King LM and Opie LH (1998). Glucose and glycogen utilisation in myocardial ischemia-changes in metabolism and consequences for the myocyte. Mol Cell Biochem **180**:3-26.

Klocke R, Roberds SL, Tamkun MM, Gronomeier M, Augustin A, Albrecht B, Pongs O *et al.* (1993). Chromosomal mapping in the mouse of eight K^+-channel genes representing the four *Shaker*-like subfamilies *Shaker*, *Shab*, *Shaw* and *Shal*. Genomics **18**:568-574.

Knowles JA, Vieland VJ and Gilliam TC (1992). Perils of gene mapping with microsatellite markers. Am J Hum Genet **51**:905-909.

Lander ES and Schork NJ (1994). Genetic dissection of complex traits. Science **265**:2037.

Lehto M, Stoffel M, Groop L, Espinosa III R, Le Beau and Bell GI (1993). Assignment of the gene encoding glycogen synthase (GYS) to human chromosome 19, band q13.3. Genomics **15**:460-461.

Lench NJ, Norris A, Bailey A, Booth A and Markham AF (1996). Vectorette PCR isolation of microsatellite repeat sequences using anchored dinucleotide repeat primers. Nucleic Acid Res **24**:2190-2191.

Lenegre J (1964). Etiology and pathology of bilateral bundle branch block in relation to complete heart block. Progress in Cardiovascular Diseases **VI**:409-444.

Levinson G and Gutman GA (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol **4**:203-221.

Li QY, Newbury-Ecob RA, Terrett JA, Wilson DI, Curtis ARJ, Yi CH, Gebuhr T *et al.* (1997). Holt-Oram syndrome is caused by mutations in *TBX5* a member of the Brachyury (T) gene family. Nature Genet **15**:21-29.

Lindblad K, Zander C, Schalling M and Hudson T (1994). Growing triplet repeats. Nature Genet **7**:124.

Litt M and Luty JA (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet **44**:397-401.

Marshall E (1999). Sequencers endorse plan for a draft in 1 year. Science **284**:1439-1441.

McGill CJ and Brooks G (1995). Cell cycle control mechanisms and their role in cardiac growth. Cardiovasc Res **30**:557-569.

Messier W, Li S–H and Stewart C-B (1996). The birth of microsatellites. Nature **381**:483.

Mohrenweiser HW, Tsujimoto S, Tynan K, Lamerdin J and Carrano AV (1995). Unique sequence STSs for 21 cytogenetically mapped loci on human chromosome 19. Cytogenet Cell Genet **71**:58-61.

Molkentin JD, Lung J–R, Antos CL, Markham B, Richardson J, Robbins J, Grant SR *et al.* (1998). A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. Cell **93**:215-228.

Moorman AFM and Lamers WH (1994). Molecular anatomy of the developing heart. TICM 4:257-264.

Moorman AFM, De Jong F, Denyn MM and Lamers WH (1998). Development of the cardiac conduction system. Circ Res **82**:629-644.

Murray JC, Buetow KH, Weber JL, Ludwigsen S, Scherpbier-Heddema T, Manion F, Quillen J *et al.* (1994). A comprehensive human linkage map with centimorgan density. Science **265**:2049-2054.

Nakamura Y, Lathrop M, O'Connell P, Leppert M, Barker D, Wright E, Skolnick M *et al.* (1988). A mapped set of DNA markers for human chromosome 17. Genomics **2**:302-309.

Nuttall FQ, Gannon MC, Kubic VL and Hoyt KJ (1994). The human liver glycogen synthase isozyme gene is located on the short arm of chromosome 12. Genomics **19**:404-405.

Ohshima K, Sakamoto N, Labuda M, Poirier J, Moseley ML, Montermini L, Ranum LP *et al.* (1999). A nonpathogenic GAAGGA repeat in the Friedreich gene: implications for pathogenesis. Neurology **53**:1854-1857.

Olson EN and Srivastava D (1996). Molecular pathways controlling heart development. Science **272**:671-676.

Online Mendelian Inheritance in Man, OMIM ™. Johns Hopkins University, Baltimore, MD. MIM Number 113900:6/26/1997. World Wide Web URL: http://www.ncbi.nlm.nih.gov/Omim.

Online Mendelian Inheritance in Man, OMIM ™. Johns Hopkins University, Baltimore, MD. MIM Number 140400:4/27/1996. World Wide Web URL: http://www.ncbi.nlm.nih.gov/Omim.

Online Mendelian Inheritance in Man, OMIM ™. Johns Hopkins University, Baltimore, MD. MIM Number 142900:4/27/1999. World Wide Web URL: http://www.ncbi.nlm.nih.gov/Omim.

Online Mendelian Inheritance in Man, OMIM ™. Johns Hopkins University, Baltimore, MD. MIM Number 160900:11/1/1999. World Wide Web URL: http://www.ncbi.nlm.nih.gov/Omim.

Oosthoek PW, Viragh S, Lamers WH and Moorman AFM (1993). Immunohistochemical delineation of the conduction system- II. The atrioventricular node and Purkinje fibers. Circ Res **73**:482-491.

Opie LH. The heart: Physiology and metabolism. Raven Press (New York). 2nd ed 1991.

Opie LH. Mechanisms of cardiac contraction and relaxation (Ch 12); in Braunwald's Heart Diseases. WB Saunder, 5th edition (Philadelphia, USA) 1997.

Orho M, Nikula-Ijäs P, Schalin-Jäntti M, Permutt A and Groop LC (1995). Isolation and characterization of the human muscle glycogen synthase gene. Diabetes **44**:1099-1105.

Orita M, Suzuki Y, Sekiya T and Hayashi K (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics **5**:874-877.

Ostrander LD, Jr., (1964). Bundle branch block: an epidemiologic study. Circulation **XXX**:872-881.

Paul DL (1995). New functions for gap junctions. Current Opinion in Cell Biology **7**:665-672.

Philibert RA, Dea C, Lee Y-H, Stubblefield BS and Ginns EI (1996). The direct sequencing of CCG repeats from a cosmid genomic DNA template. Epicentre Forum **3**:1-3.

European Polycystic Kidney Disease Consortium (1994). The polycystic kidney disease 1 gene encodes a 14kb transcript and lies within a duplicated region on chromosome 16. Cell **77**:881-894.

Ravnik-Glavac M, Glavac D and Dean M (1994). Sensitivity of single-strand conformation polymorphism and heteroduplex method for mutation detection in the cystic fibrosis gene. Hum Mol Genet **3**:801-807.

Reddy PH, Stockburger E, Gillevet P and Tagle DA (1997). Mapping and characterization of novel $(CAG)_n$ repeat cDNAs from adult human brain derived by the oligo capture method. Genomics **46**:174-182.

Riley J, Butler R, Ogilvie D, Finniear R, Jenner D, Powell S, Anand R, *et al.* (1990). A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. Nucleic Acids Res **18**:2887-2890.

Rossant J (1996). Mouse mutants and cardiac development. Circ Res **78**:349-353.

Rothuizen J and Van Raak (1994). Rapid PCR-based characterization of sequences flanking microsatellites in large-insert libraries. Nucleic Acids Res **22**:5512-5513.

Royle NJ, Clarkson RE, Wong Z and Jeffreys AJ (1988). Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. Genomics **3**:352-360.

Sack MN and Kelly DP (1998). The energy substrate switch during development of heart failure: gene regulatory mechanisms (review). Int J Mol Med **1**:17-24.

Sack MN, Rader TA, Park S, Bastin J, McCune S and Kelly DP (1996). Fatty acid oxidation enzyme gene expression is downregulated in the failing heart. Circulation **94**:2837-2842.

Sarachek, NS and Leonard JJ (1972). Familial heart block and sinus bradycardia. Am J Cardiol **29**: 451-458.

Schaal SF, Seidensticker J, Goodman R and Wooley CF (1973). Familial right bundle branch block, left axis deviation, complete heart block and early death. Ann Intern Med **79**:63 66.

Schiaffino S (1997). Protean patterns of gene expression in the heart conduction system. Circ Res **80**:749-750.

Schlötterer C and Tautz D (1992). Slippage synthesis of simple sequence DNA. Nucleic Acids Res **20**:211-215.

Schott JJ, Alshinawi C, Kyndt F, Probst V, Hoorntje TM, Hulsbeek M, Wilde AA *et al.* (1999). Cardiac conduction defects associate with mutations in *SCN5A*. Nature Genet **23**:20-21.

Schott JJ, Benson DW, Basson CT, Pease W., Silberbach GM, Moak JP, Maron BJ *et al.* (1998). Congenital heart disease caused by mutations in the transcription factor NKX2-5. Science **281**:108-111.

Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE *et al.* (1996). A gene map of the human genome. Science **274**:540-546.

Sheffield VC, Beck JS, Kwitek AE, Sandstrom DW and Stone EM (1993). The sensitivity of single-strand conformational polymorphism analysis for the detection of single base substitutions. Genomics **16**:325-332.

Sheffield VC, Weber JL, Buetow KH, Murray JC, Even DA, Wiles K, Gastier JM, *et al.* (1995). A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. Hum Mol Genet **4**:1837-1844.

Simonsen EE and Madsen EG (1970). Four cases of right-sided bundle branch block and one case of atrioventricular block in three generations of a family. Brit Heart J **32**:501-504.

Smith J (1999). T-box genes: what they do and how they do it. Trends in Genetics **15**:154-158.

Solomon E and Bodmer WF (1979). Evolution of sickle gene variant. Lancet **1**:923.

Sommer JR and Johnson EA (1970). Comparative ultrastructure of cardiac cell membrane specializations- a review. Am J Cardiol **25**:184-194.

Spray DC, Rok MB, Moreno AP, Sáez JC, Christ GJ, Campos de Carvalho and Fishman GI. Cardiovascular gap junctions: gating properties, function, and dysfunction in ion channels in the cardiovascular system: function and dysfunction (edited by Spooner PM and Brown AM). Futura Publishing company Inc. (New York) 1994.

Srivastava D (1999). HAND proteins: molecular mediators of cardiac development and congenital heart disease. Trends Cardiovasc Med **9**:11-18.

Srivastava D, Cserjesi P and Olson EN (1995). A subclass of bHLH proteins required for cardiac morphogenesis. Science **270**:1995-1999.

Stallings RL (1994). Distribution of trinucleotide microsatellites in different categories of mammalian genomic sequence: implications for human genetic diseases. Genomics **21**:116-121.

Stallings RL, Ford AF, Nelson D, Torney DC, Hildebrand CE and Moyzis RK (1991). Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. Genomics **10**:807-815.

Steenkamp WFJ (1972). Familial trifascicular block. Am Heart J **84**:758-760.

Stephan E (1979). Hereditary bundle branch system defect: A new genetic entity? Am Heart J **97**:708-718.

Stephan E, Aftimos G and Allam C (1985). Familial fascicular block: Histologic features of Lev's disease. Am Heart J **109**:1399-1402.

Stephan E, De Meeus A and Bouvagnet P (1997). Hereditary bundle branch defect: right bundle branch blocks of different causes have different morphologic characteristics. Am Heart J **133**:249-256.

Strachan T and Read AP. Human Molecular Genetics. Bios Scientific (Oxford) 1996.

Tautz D and Renz M (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res. **12**:2669-2690.

Taylor GR, Haward S, Noble JS and Murday V (1992). Isolation and sequencing of CA/GT repeat microsatellites from chromosomal libraries without sub-cloning. Analytical Biochem **200**:125-129.

Terwilliger JD and Ott J. Handbook of human genetic linkage (Chapter 1;p4-6). The Johns Hopkins-University Press (Baltimore, London) 1994.

Thein SL and Wallace RB. The use of synthetic oligonucleotides as specific hybridization probes in the diagnosis of genetic disorders. In Human Genetic Diseases: a practical approach (ed KE Davis). IRL Press (Herndon, Virginia) 1986.

The Utah Marker Development Group (1995). A collection of ordered tetranucleotide-repeat markers from the human genome. Am J Hum Genet **57**:619-628.

Thornell LE and Sjostrom M (1975). Purkinje fibre glycogen: a morphologic and biochemical study of glycogen particles isolated from the cow conducting system. Basic Res Cardiol **70**:661-670.

Tischfield JA (1997). Loss of heterozygosity or: how I learned to stop worrying and love mitotic recombination. Am J Hum Genet **61**:995-999.

Torrington M, Weymar HW, Van der Merwe P-L and Brink AJ (1986). Progressive familial heart block. S Afr Med J **70**:353-355.

Vallianos G and Sideris DA (1974). Familial conduction defects. Cardiology **59**:190-197.

Van der Merwe P–L, Rose AG, Van der Walt JJ, Weymar HW, Hunter JC and Weich HFH (1991). Progressive familial heart block, type I. S Afr Med J **80**:34-38.

Van der Merwe P–L, Weymar HW, Torrington M and Brink AJ (1986). Progressive familial heart block, part II: clinical and ECG confirmation of progression. S Afr Med J **70**:356-357.

Van der Merwe P–L, Weymar HW, Torrington M and Brink AJ (1988). Progressive familial heart block (typeI): a follow-up study after 10 years. S Afr Med J **73**:275-276.

Vogt P (1990). Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved “chromatin folding code". Hum Genet **84**:301-336.

Walter MA, and Goodfellow PN (1992). Disease and development. Nature **355**:590-591.

Wang Q, Curran ME, Burn TC, Millholland JM, Van Raay TJ, Shen J, Timothy KW *et al.* (1995). Positional cloning of a novel potassium channel gene: *KVLQT1* mutations cause arrhythmias. Nature Genet **12**:17-23.

Warren ST (1996). The expanding world of trinucleotide repeats. Science **271**:1374-1375.

Waterston R and Sulston JE (1998). The Human Genome Project- reaching the finish line. Science **282**:53-54.

Weber JL (1990). Informativeness of human $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms. Genomics **7**:524-530.

Weber JL and May PE (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet **44**:388-396.

Weber JL, Wang Z, Hansen K, Stephenson M, Kappel C, Salzman S, Wlkie PJ *et al.* (1993). Evidence for human meiotic recombination interference obtained through construction of a short tandem repeat-polymorphism linkage map of chromosome 19. Am J Hum Genet **53**:1079-1095.

Wetzel GT and Klitzner TS (1996). Developmental cardiac electrophysiology: recent advaces in cellular physiology. Cardiovascular Res **31**:E52-E60.

Yamagishi H, Garg V, Matsuoka R, Thomas T and Srivastava D (1999). A molecular pathway revealing a genetic basis for human cardiac and craniofacial defects. Science **283**:1158-1161.

Zheng C–J, Byers B and Moolgavkar SH (1993). Allelic instability in mitosis: a unified model for dominant disorders. Proc Natl Acad Sci, USA **90**:10178-10182.

Zipes DP. Genesis of cardiac arrhythmias: electrophysiological considerations (Ch 20); in Braunwald's Heart Diseases. WB Saunder, 5th edition (Philadelphia, USA) 1997.

Zuliani G and Hobbs HH (1990). A high frequency of length polymorphisms in repeated sequences adjacent to *Alu* sequences. Am J Hum Genet **46**:963-969.

APPENDIX I

Buffers, solutions, markers, methods

1. BACTERIAL CELL GROWTH MEDIA

LURIA-BERTANI MEDIUM

NaCl	10 grams
Tryptone	10 grams
Yeast extract	5 grams

ddH_2O to 1litre

Adjust to pH 7.2 with 5M NaOH and autoclave.

LURIA-BERTANI AGAR

NaCl	10 grams
Tryptone	10 grams
Yeast extract	5 grams

Dissolve in 700ml ddH_2O and adjust to pH 7.2 with 5M NaOH

Bactor-agar	15 grams

ddH_2O to 1litre and autoclave.

2. COSMID DNA EXTRACTION

Composition of QIAGEN® plasmid/cosmid purification kit buffers

BUFFER P1 (Resuspension buffer)

Tris-HCl (pH 8)	50mM
EDTA	10mM
RNase A	100µg/ml

BUFFER P2 (Lysis buffer)

NaOH	200mM
SDS	1%

BUFFER P3 (Neutralisation buffer)

Potassium acatate (pH 5.5)	3.0M

BUFFER QBT (Equilibration buffer)

NaCl	750mM
MOPS (pH 7.0)	50mM
Ethanol	15%
Triton X-100	0.15%

BUFFER QC (Wash buffer)

NaCl	1.0M
MOPS (pH 7.0)	50mM
Ethanol	15%

BUFFER QF (Elution buffer)

NaCl	1.25M
Tris-HCl (pH 8.5)	50mM
Ethanol	15%

3. PLASMID DNA EXTRACTION

SOLUTION I (Resuspension buffer)

Glucose	50mM
Tris (pH 8.0)	25mM
EDTA	0.5M

ddH_2O to 500ml and autoclave.

SOLUTION II (Lysis buffer)

NaOH	0.2M
SDS	1%

ddH_2O to 1ml

SOLUTION III (Neutralising buffer)

Glacial acetic acid	11.5%
Potassium acetate	3M

ddH_2O to 100ml

4. HUMAN GENOMIC DNA EXTRACTION

CELL LYSIS BUFFER

Sucrose	0.32M
Triton X-100	1%
$MgCl_2$	5mM
Tris-HCl, ph 8.0	10mM
ddH_2O to 1 litre	

3M NaAc

$NaAc.3H_2O$	40.81g
ddH_2O	50ml

Adjust pH to 5.2 with glacial acetic acid (Merck) and adjust volume to100ml with ddH_2O.

Na-EDTA SOLUTION

NaCl	18.75ml of 4M stock solution
EDTA	250ml of 100mM stock solution

Mix well.

PHENOL/CHLOROFORM

Mix phenol (saturated with 1xTE), chloroform and 8-hydroxiquinoline in a ratio of 25:24:1. Store at 4°C.

CHLOROFORM/OCTANOL

Mix chloroform and octanol in a ratio of 24:1. Store at 4°C.

METHODOLOGY

PHENOL/CHLOROFORM AND CHLOROFORM/OCTANOL EXTRACTION

Add a volume of phenol/chloroform mixture equal to the volume of the sample, which is to be extracted to the latter sample. Mix well, centrifuge at 13 000 rpm, in a benchtop eppendorf centrifuge (Beckman model TJ-6) or Sorvall centrifuge (Du Pont Instruments, model RC-5B), depending on the size of the sample centrifuge tubes, for 5 minutes, to separate the aqueous phase from the organic phase. Remove top layer (aqueous phase) to a clean tube, and add to it a similar volume of chloroform/octanol. Mix well, and centrifuge as before. Again, remove the aqueous phase to a clean tube. This sample can now be ethanol precipitated to extend the purification process.

5. COLONY BLOT SOLUTIONS

DENATURING SOLUTION

NaCl	1.5M
NaOH	0.5M
ddH_2O to 1litre	

NEUTRALISING SOLUTION

NaCl	1.5M
Tris (pH 8)	0.5M
ddH_2O to 1litre	

PRE-WASH SOLUTION

SSC	5X
SDS	0.5%
EDTA	1mM
ddH_2O to 1litre	

6. HYBRIDISATION STOCK SOLUTIONS

20X SSC

NaCl	0.15M
Na citrate	0.015M

100X DENHARDT'S SOLUTION

Ficoll 400	0.025mM
Polyvinylpyrrolidine	0.027mM
BSA Fraction	2% (weight/volume)
Store at -20°C	

20X SSPE

NaCl	125mM
Na_2HPO_4	17mM
NaH_2PO_4	8mM
SDS	173mM
PEG_{8000}	12,5Mm

7. Solutions for Membrane Filters

PRE-HYBRIDISATION SOLUTION

20X SSPE	20ml
100X Denhardt's	4ml
10% SDS	1.6ml
Sonicated *E. coli* DNA	5mg/ml
ddH_2O to 80ml	

HYBRIDISATION SOLUTION

20X SSPE	20ml
100X Denhardt's	4ml
10% SDS	1.6ml
Sonicated *E. coli* DNA	5mg/ml
ddH_2O to 80ml	
γ-^{32}P[dATP] labelled $(AAAT)_{10}$ oligo (1pmol/µl)	5-10 pmol

POST-HYBRIDISATION WASH SOLUTION

SSC	2X
SDS	0.1%

8. STOCK SOLUTIONS

30% ACRYLAMIDE STOCK SOLUTION FOR POLYACRYLAMIDE GELS

Acrylamide	30g
Bis-acrylamide	0.8g

Add H_20 to 100ml, mix on magnetic stirrer for 2 hours, filter and store at 4°C.

10% AMMONIUM PERSULPHATE

Ammonium persulphate	1g
H_2O	10ml

Mix well and store at 4°C.

ETHIDIUM BROMIDE (10mg/ml)

Ethidium bromide	500mg
H_2O	50ml

Stir well on magnetic stirrer for 4 hours, store in a dark container at 4°C.

STE BUFFER

NaCl	100mM
Tris-HCl (pH 8.0)	10mM
EDTA	1mM

10X TBE BUFFER STOCK SOLUTION

Tris-OH	0.89M, pH 8.0
Boric Acid	0.89M
Na_2EDTA	20mM, pH 8.0

10X TE BUFFER STOCK SOLUTION

Tris-HCl	0.1M, pH 8.0
EDTA	0.01M, pH 8.0

9. AGAROSE GEL SOLUTIONS

Concentration	**0.7%**	**1%**	**1.5%**	**2%**
Agarose (g)	0.35	0.5	0.75	1.0
10x TBE stock (ml)	5	5	5	5
H_2O (ml)	45	45	45	45

Melt agarose (FMC) in a microwave oven on HIGH power, swirling occasionally until it is completely dissolved. Cool the solution for 10' to 50°C - 60°C and then add 5µl ethidium bromide (10mg/ml) prior to casting.

10. POLYACRYLAMIDE GEL SOLUTIONS FOR RE-DIGESTION ANALYSIS AND SEQUENCING

12% NON-DENATURING POLYACRYLAMIDE GEL SOLUTION

30% acrylamide stock solution	8ml
10x TBE stock solution	2ml
ddH_2O	10ml
10% ammonium persulphate	160µl

TEMED	60µl

5% DENATURING POLYACRYLAMIDE GEL SOLUTION

30% acrylamide stock solution	10.8ml
10x TBE stock solution	6ml
Urea	26.64g
H_20	12.75ml

Stir well with a magnetic stirrer until all urea is dissolved, then add

10% ammonium persulphate	500µl
TEMED	50µl

METHODOLOGY

CASTING OF SEQUENCING GELS

Glass plates are prepared by carefully and meticulously washing both plates with detergent. Any traces of detergent are subsequently removed by wiping the plates with 70% ethanol, and drying them thoroughly. The notched plate is then silanised with Wynn's C-thru windshield rain dispersant (Wynn Oil S.A. Pty Ltd, SA) to allow the gel to remain stuck to the large plate upon dismantling after electrophoresis. Spacers of 0.35mm (SSLP analysis gels) or 0.5mm (sequencing gels) thickness are placed lengthwise at the edges of the large glass plate, and the notched plate placed silanised side-down on top of these spacers. This assembly is subsequently sealed with the use of a rubber boot (S2 casting boot, Life ™ Technologies) and the gel poured with the glass assembly slanted at a slight angle. The glass assembly is subsequently layed down horizontally and the loading front formed with a top spacer of identical thickness to the lengthwise spacers. The gel is allowed to set for 1-2 hours, whereupon the top spacer is removed and replaced by a sharkstooth comb to separate wells for loading of samples.

11. SSCP ANALYSIS GEL SOLUTIONS

0.35X MDE GEL SOLUTION

2x MDE™ stock solution (FMC)	28ml
10x TBE stock solution	9.7ml
H_20	122ml
10% ammonium persulfate	1.6ml
TEMED	160µl

5% NON-DENATURING POLYACRYLAMIDE GEL SOLUTION WITH 10% GLYCEROL

30% acrylamide stock solution	27ml
10x TBE stock solution	8ml
Glycerol	16ml
H_20	108ml
10% ammonium persulphate	1.6ml
TEMED	160µl

10% NON-DENATURING POLYACRYLAMIDE GEL SOLUTION WITH 10% GLYCEROL

30% acrylamide stock solution	54ml
10x TBE stock solution	8ml
Glycerol	16ml
H_20	81ml
10% ammonium persulphate	1.6ml
TEMED	160µl

8% MILDLY-DENATURING POLYACRYLAMIDE GEL SOLUTION

30% acrylamide stock solution	32.4ml
10x TBE stock solution	8ml
Urea	24g
Glycerol	8ml
H_20	91.8ml
10% ammonium persulphate	1.6ml
TEMED	160µl

10% MILDLY-DENATURING POLYACRYLAMIDE GEL SOLUTION

30% acrylamide stock solution	40.5ml
10x TBE stock solution	8ml
Urea	24g
Glycerol	8ml
H_20	84ml
10% ammonium persulphate	1.6ml
TEMED	160µl

METHODOLOGY

CASTING OF SSCP GELS

Glass plates are prepared by carefully and meticulously washing both plates with detergent. Any traces of detergent are subsequently removed by wiping the plates with 70% ethanol, and drying them thoroughly. The notched plate is then silanised with Wynn's C-thru windshield rain dispersant (Wynn Oil S.A. Pty Ltd, SA) to ease the separation of the gel from the notched plate upon dismantling after electrophoresis. More 70% ethanol is now sprayed on the inner surface of the large glass plate. A sheet of Gelbond™ (FMC Bioproducts, Rockland, Maine, USA), cut to the same size as the large glass plate, is placed, hydrophobic side down, on this plate, taking care not to remove the protective sheet of paper from the hydrophilic side of the nylon membrane. Air bubbles trapped between the Gelbond™ and the glass are removed by careful rubbing over the protective sheet of paper with a piece of paper towel. Subsequently, this sheet of paper is removed from the nylon membrane and spacers of 1mm (SSLP analysis gels) thickness placed lengthwise at the edges of the large glass plate. The notched plate is placed silanised side-down on top of these spacers. This assembly is subsequently sealed with the use of a rubber boot (S2 rubber casting boot, Life™ Technologies) and the gel poured with the glass assembly slanted at a slight angle. The glass assembly is subsequently layed down horizontally and a square-tooth well-forming comb inserted. The gel is then allowed to set for 1-2 hours.

12. MOLECULAR SIZE MARKER

λPst

Bacteriophage λ genomic DNA	(50μg) 200μl
*Pst*I (40U/μl)	3μl
Buffer H	30μl
ddH_20	67μl

Incubate at 37°C for 2 hours, heat inactivate enzyme at 65°C for 5'. Use 1μl of this marker on polyacrylamide gels, which will be silver stained, or 3μl on ethidium bromide stained agarose gels.

Fragment sizes (bp):

11497, 5077, 4749, 4507, 2838, 2560, 2459, 2443, 2140, 1986, 1700, 1159, 1093, 805, 514, 468, 448, 339, 264, 247, 216, 211, 200, 164, 150, 94, 87, 72.

13. LOADING DYES

BROMOPHENOL BLUE LOADING DYE FOR AGAROSE AND POLYACRYLAMIDE GEL ELECTROPHORESIS

Bromophenol blue	0.1% (w/v)
Glycerol	60% (v/v)
Tris-Cl, pH 8	10mM

Vortex to mix. This loading dye is deliberately not as intensely coloured as conventional loading dyes in order to prevent masking of small PCR amplified DNA fragments on agarose gels.

SSCP LOADING DYE

Bromophenol blue	0.1% (w/v)
Xylene cyanol	0.1% (w/v)
Formamide	95% (v/v)
NaOH	10mM
EDTA	20mM

Mix well. Aliquot and store at -20°C.

SEQUENCING STOP SOLUTION

Bromophenol blue	0.1% (w/v)
Xylene cyanol	0.1% (w/v)
Formamide	95% (v/v)

14. SILVER STAINING of NON-DENATURING POLYACRYLAMIDE GELS

$AgNO_3$ SOLUTION (0.1%)

$AgNO_3$	1g
ddH_20	1 litre

DEVELOPING SOLUTION

NaOH	15g
$NaBH_4$	0.1g
Formaldehyde	4ml
ddH_20 to 1 litre	

Mix well on magnetic stirrer until all NaOH pellets are dissolved.

METHODOLOGY

SILVER STAINING OF POLYACRYLAMIDE GELS

Dismantle gel casting plates and gently immerse polyacrylamide gel into a shallow receptacle containing freshly prepared $AgNO_3$ solution, agitate for 10'. Pour off the $AgNO_3$ solution and rinse both sides of the gel well with ddH_2O. Add freshly prepared developing solution and agitate the gel in this solution until DNA fragments become visible as bands on the gel. Pour the developing solution off, rinse the gel in ddH_2O and air-dry (SSCP gels) and/or photograph (12% polyacrylamide ASREA gels).

15. X-RAY FILM DEVELOPING

DEVELOPING SOLUTION

Polycon A variable contrast X-ray developer Polycon A (Champion Photochemistry Pty Ltd, Midrand, SA) solution	1.98 litre
Polycon A variable contrast X-ray developer solution 2	18.1ml
ddH_2O	9.9
litre	

STOP SOLUTION

Glacial acetic acid	540ml
Make up to 9 litres with ddH_2O	

FIXING SOLUTION

Perfix (Champion Photochemistry Pty Ltd, Midrand, SA)	1.8 litre
Hardener-S (Champion Photochemistry Pty Ltd, Midrand, SA)	225ml
Make up to 9 litre with H_2O.	

METHODOLOGY

The X-ray film is clamped to a stainless steel frame under red light, and is immersed into the developing solution for 4 min, then into the stop solution for 30 sec and finally into the developing solution for 2 min. Thereafter, the film is rinsed well with H_2O, after which it is allowed to air-dry.

APPENDIX II

List of Suppliers

α^{32}P-dCTP	Amersham
Ampicilin	Sigma
Agarose	Seakem
Agar	Difco Laboratories
Ammonium persulphate	Merck
Acetic acid	BDH Chemicals
$AgNo_3$	Merck
Acrylamide	BDH Chemicals
Bis-acrylamide	Merck
BSA	Seravac
Boric acid	Merck
5-Bromo-4-chloro-3-indoyl-β-D galactopyranoside	Boehringer Mannheim
Bromophenol blue	BDH Chemicals
$(AAAT)_{10}$ oligomer	DNA Synthesis Laboratory, UCT
$CaCl_2$	Merck
Chloroform	BDH Chemicals
dATP	Boehringer Mannheim
dCTP	Boehringer Mannheim
dGTP	Boehringer Mannheim
dTTP	Boehringer Mannheim
E. coli host cell line, DH5α	Stratagene, USA
E. coli, XL1 Blue competent cells	Startagene, USA
Ethanol	Boehringer Mannheim
Ethidium bromide	Sigma
EDTA	Boehringer Mannheim
Ficoll 4000	Merck
fmol™ sequencing kit	Promega
Formamide	Merck
Formaldehyde	Merck
GLAD® WRAP	Multifoil Trading (Pty) Ltd.
Glycerol	BDH Chemicals

Glucose	BDH Chemicals
γ^{32}P-dATP	Amersham
Hydrochloric acid	BDH Chemicals
Iso-amyl alcohol	Merck
Isopropyl-β-D-thio-galactopyranoside	Boehringer Mannheim
Hybond N^+ membrane	Amersham
Potassium acetate	Sigma
Kanamycin	Sigma
*Kpn*I	Promega
Lambda DNA	Promega
Lysozyme	Boehringer Mannheim
$MgCl_2$	Sigma
Mineral oil	BDH Chemicals
3 MM Whatman paper	Merck
MOPS	Sigma
NaCl	BDH Chemicals
Na citrate	BDH Chemicals
Na_2HPO_4	BDH Chemicals
NaH_2PO_4	BDH Chemicals
NaOH	Merck
Na acetate	Merck
Octanol	Merck
One-Phor-All buffer	Pharmacia
pBluescript SK vector	Stratagene, USA
PEG_{8000}	Pharmacia
Proteinase K	Sigma
Phenol	Merck
Polyvinylpyrolidine	Sigma
*Pst*I	Promega
Qiagen® plasmid purification kit	Stratagene
RNase A	Sigma
*Sau3A*1	Promega
*Sal*I	Promega
SDS	Sigma
Sephadex G25 gel	Pharmacia

GFX™ gel band purification kit	Amersham
Taq Polymerase	Promega
T4 DNA ligase	Promega
T4 polynucleotide kinase	Promega
T3 and T7 vector primers	Stratagene, USA
Temed	Promega
Tetracycline	Sigma
Tris	Merck
Triton X-100	BDH Chemicals
Tryptone	Biolab
Trichloric acid	Merck
Urea	BDH Chemicals
X-ray film	Protea
Xylene cyanol	M&B Laboratory Chemicals
Yeast extract	Oxoid

APPENDIX III

CONFERENCE PRESENTATIONS

Oral

August 1997: **Makubalo Z**, Christoffels A, Arieff Z, De Jager T, Brink P, Corfield V.
Closing in on the progressive familial heart block gene: candidate gene screening and refining the search area.
41st Academic Year Day, Faculty of Medicine, University of Stellenbosch.

October 1997: **Makubalo Z,** Christoffels A, Arieff Z, De Jager T, Brink P, Corfield V
Attractive candidates and great looking maps: their role in the search for the PFHB gene.
Experimental Biology Group 141st General Meeting..

Poster

May 1997: **Makubalo Z,** Christoffels A, Arieff Z, De Jager T, Brink P, Corfield V.
Mutation screening of *GSY1* and *HRC*: Progressive Familial Heart Block I candidate genes and generation of tetranucleotide repeat genetic markers.
Seventh Biennial Southern African Society of Human Genetics. Pilanesberg.
(First prize for Best Poster)
Presented also at the *Mutation detection workshop in Hinxton, Cambridge, UK* (**September 1998**).

March 1999: **Makubalo Z**, Moolman-Smook J, Brink P, Corfield V.
PFHBI gene update: DIY intronic sequencing and marker generation.
Eighth Biennial Southern African Society of Human Genetics. Gordon's Bay.