The Molecular Basis of Thalassaemia in Sri Lanka

Christopher A. Fisher

A thesis submitted in partial fulfilment of the requirements of Oxford Brookes University for the degree of Doctor of Philosophy

This research programme was carried out in collaboration with the NHS National Haemoglobinopathy Reference Service and the MRC Molecular Haematology Unit at the Weatherall Institute for Molecular Medicine

August 2002

Abstract

Blood samples from over 1600 schoolchildren revealed a prevalence of both β-thalassaemia trait and HbE trait, averaging 2.3% and 0.8% respectively. With the present population estimate at 18.6 million, it is predicted that approximately 2650 patients with β-thalassaemia major and 956 with HbE/β-thalassaemia should consume >5% of the current health budget for treatment of these patients. Patients with the phenotype of severe thalassaemia, from seven centres, were analysed to understand the molecular basis of the disease.

Twenty-four β-globin gene mutations were identified with three mutations accounting for 84.4% of the 1234 alleles: IVSI-5 (G→C) 56.1%; IVSI-1 (G→A) 15.2% and HbE 13.1%. Sixteen mutations have previously been described on the Indian subcontinent and demonstrate close regional links between the populations, especially North East India, and were supported by haplotype association. The IVSI-1 (G→A) mutation occurs commonly in countries around the Mediterranean but very rarely on the Indian subcontinent. In Sri Lanka it is associated with a different 3' haplotype, suggesting a multicentric origin. Three new mutations were found; a frameshift codon 6-10 (Δ13bp); IVSI-129 (A→C) in the consensus splice site and a second frame shift, CD55 (-A).

Four α-gene arrangements were demonstrated. An allele frequency of the α^{+}-thalassaemia gene deletions were 6.5% and 1.1% for the $-\alpha^{3.7}$ deletion and the $-\alpha^{4.2}$ deletion, respectively. Cord blood analysis of Hb Bart's levels confirmed an estimated level of 7.1% for a single α-gene non-expression. Extra α-genes are common and the first instance of homozygosity for the ααάα arrangement was observed. No α^{0}-thalassaemia was revealed and only a single polymorphism was associated with a possible extremely mild α-thalassaemia allele.

The phenotype associated with the various β-globin genotypes was found to be extremely variable. Within the many factors that modify this phenotype the Xmn-I polymorphism was significantly associated with the level of haemoglobin expressed. A +/+ genotype appeared to be a good indicator of a milder outcome. The incidence of α-gene deletions was low in patients with HbE/β-thalassaemia and thus proposed this as a strong modifier for a mild phenotype. Other factors were associated with specific complications.

Publications and Meeting abstracts arising from the thesis

Publications

Bayoumi, R.A., Dawodu, A., Qureshi, M.M., Al-Khider, A., Fitzgerald, P., Riou, J., Fisher, C.A., Fitches, A. and Old, J.M., The association of Hb Khartoum [β124(H2)Pro→Arg] with γ^{+}-thalassemia is responsible for hemolytic disease in the newborn of a Sudanese family, *Hemoglobin*, 23(1), 33, 1999.

Rochette, J., Pointon, J.J., Fisher, C.A., Perera, G., Arambepola, M., Arichchi, D.S.K., Silva, S.D., Vandwalle, J.L., Monti, J.P., Old, J.M., Merryweather-Clarke, A.T., Weatherall, D.J. and Robson, K.J.H., Multicentric origin of hemochromatosis gene (HFE) mutations, *American Journal of Human Genetics*, 64, 1056, 1999.

de Silva, S., Fisher, C.A., Premawardhena, A., Lamabadusuriya, S.P., Peto, T.E.A., Perera, G., Old, J.M., Clegg, J.B., Olivieri, N.F., Weatherall, D.J. and Group, Sri Lanka Thalassaemia Study Group, Thalassaemia in Sri Lanka: implications for the future health burden of Asian populations, *Lancet*, 355(4), 786, 2000.

Liu, Y.T., Old, J.M., Miles, K., Fisher, C.A., Weatherall, D.J. and Clegg, J.B., Rapid detection of alpha-thalassaemia deletions and alpha-globin gene triplication by multiplex polymerase chain reactions, *British Journal of Haematology*, 108(2), 295, 2000.

Premawardhena, A., Fisher, C.A., Fathiu, F., de Silva, S., Perera, W., Peto, T.E., Olivieri, N.F. and Weatherall, D.J., Genetic determinants of jaundice and gallstones in haemoglobin E beta thalassaemia, *Lancet*, 357(9272), 1945, 2001.

Old, J.M., Khan, S.N., Verma, I., Fucharoen, S., Kleanthous, M., Ioannou, P., Kotea, N., Fisher, C., Riazuddin, S., Saxena, R., Winchagoon, P., Kyriacou, K., Al-Quobaili, F. and Khan, B., A multicentre study in order to further define the molecular basis of β-thalassaemia in Thailand, Pakistan, Sri Lanka, Mauritius, Syria and India, and to develop a simple molecular diagnostic strategy by amplification refractory mutation system-polymerase chain reaction, *Hemoglobin*, 25(4), 397, 2001.

Viprakasit, V., Tanphaichitr, V.S., Pung-Amritt, P., Petrarat, S., Suwantol, L., Fisher, C. and Higgs, D.R., Clinical phenotypes and molecular characterization of Hb-Pakse disease, *Haematologica*, 87, 117, 2002.

Meeting abstracts

International Conference on Thalassaemia and haemoglobinopathies, April, 1997 in Malta. A preliminary survey of the molecular basis of β-thalassaemia major in Sri Lanka. **C.A.Fisher, S. de Silva, G. Perera, S.P. Lambadasuriya, T.E.A. Peto, N.F. Olivieri, D.J. Weatherall, J.M. Old and Sri Lanka Thalassaemia Study Group.**

American Society of Hematology. Forty-first Annual Meeting. December 3rd-7th,2000. New Orleans, USA. Thalassaemia in Sri Lanka: A molecular basis. **C.A.Fisher, A. Premawardhena, S. De Silva, G. Perera, S.P. Lambadasuriya, T.E.A. Peto, N.F. Olivieri, D.J. Weatherall, J.M. Old and Sri Lanka Thalassaemia Study Group.**

Acknowledgements

I would like to thank my wonderful children, Alex and Nadine, but above all my wife, Nita without whom none of this work would have been possible. I am truly appreciative to you for your support, encouragement, and love. You gave an unending amount of enthusiasm for this endeavour that has lifted spirits continually and made a difficult task achievable. Your encouragement has allowed me a base from which to operate and I shall always be indebted to you. I am tremendously sorry for the amount of time spent at work instead of with you. Thanks for staying up with me during those late-night sessions. You are absolutely the best.

"In the confrontation between the stream and the rock, the stream always wins - not through strength, but through persistence." Anon.

Thank you, to my family, for turning that stream from a trickle into an irrepressible flow. For this I dedicate this thesis to **Alex, Nadine** and **Nita.**

United Kingdom

National Haemoglobinopathy Reference Laboratory

Dr J.M. Old who allowed me to work on the Sri Lankan project and his support and advice throughout the project. Miss M. Rugless for constant support and for collaboration of parts of the molecular analysis. Most specifically for typing the cord blood samples from the Homerton Hospital. The summer of blood.

MRC Molecular Haematology Department, WIMM

Professor Sir D. Weatherall who allowed me to work on the Sri Lanka patients. For all his drive and enthusiasm and constant advice throughout this project. Professor J.B. Clegg for his advice and support. Dr T.E.A. Peto for statistical analysis leading to the Lancet paper. Dr A. Premawardhena (a commonwealth student) from Sri Lanka who was instrumental in the coordination and collection of data from 1998. Under my supervision Anuja was responsible for the β-gene study of samples from Kandy and also for a joint effort in the genetic analysis of UGT1A. Miss G. Perera (a medical student) who visited Sri Lanka and collected blood samples from patients and their parents in Anurahapura with the help of hospital staff. Miss S. Rajapakshe (a medical student) who visited Sri Lanka and collected cord blood samples from both Kurunegala and Negambo under the supervision of Dr S. De Silva and with the help of midwifes and nurses. Dr K. Robson and her group for guidance in the analysis of the common haemochromatosis mutations and for further collaborations on associated polymorphisms. Mr Y.T. Liu who designed the primers for the α-gene multiplex PCR.

To all at the Weatherall Institute of Molecular Medicine, present and past, who gave constant support. I would especially like to thank Miss M. Rugless and Dr J. Pointon for their help and encouragement.

Homerton Hospital Hackney

Dr A. Stephens for obtaining ethical permission to conduct the study on antenatal patients that attended the Homerton Hospital. Dr R. Amos for the collection, characterisation and transportation of cord blood samples from East London to Oxford.

Brookes University

Thanks to Professor Beadle for encouragement and support and the help to get through official paperwork ensuring that the thesis progressed.

Canada

Hospital for Sick Children and the Toronto General Hospital

Dr N.F. Olivieri and Dr S. Sharma for help in the collection of patient data from Sri Lanka.

Sri Lanka

Dr S. de Silva and Professor D.J. Weatherall recruited all the collaborators at the centres in Sri Lanka.

General Teaching Hospital, Kurunegala

Dr S. de Silva obtained ethical permission for this study with support from Dr R. Perera, Director General of Health Services, Sri Lanka. Shanthimala was also instrumental in the organisation and collection of blood samples from the schoolchildren and their teachers. The patients at the thalassaemia clinics in Kurunegala and Negambo were under her care and the collection of samples and information were organised by her. Assistance and support was given by Dr H. Kumarihami, Dr C. Valeketiya, Dr W. Perera, Dr F. Fathiu, Dr I. Wijesiriwardena, Dr P. Wijenayake and Dr S. Rajakanthan.

Department of Paediatrics, University of Colombo

Professor S.P. Lamabadasuriya and Dr M.P. Senanayake for help in the organisation and collection of blood samples and patient information from the Colombo District.

General Hospital at Anuradhapura

Dr M. Arembepola, Dr D.A. Kodikararachchi and G. Perera for help in the organisation and collection of blood samples and patient information from the Anuradhapura District.

General Hospital at Kandy

Dr E. Ekanayaka, Dr H. Welgama and Dr S. Rajapakshe for help in the organisation and collection of blood samples and patient information from the Kandy District.

General Hospital at Chilaw

Dr S. Waidyanath for help in the organisation and collection of blood samples and patient information from the Chilaw District.

General Hospital at Badulla

Dr D. Mendis for help in the organisation and collection of blood samples and patient information from the Badulla District.

Teaching Hospital at Galle

Dr N. Dahanayake for help in the organisation and collection of blood samples and patient information from the Galle District.

Contents

List of Tables

List of Tables continued

List of Figures

List of Figures (continued)

List of abbreviations

A	Adenine
ARMS	Amplification refractory mutation system
C	Cytosine
CAE	Cellulose acetate electrophoresis
CD	codon
cDNA	Copy DNA
CE-HPLC	Cation exchange high performance liquid chromatography
dCTP	Deoxy cytosine triphosphate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease I
EDTA	Ethylene diamine tetra acetic acid; disodium salt
G	Guanine
Hb	Haemoglobin
HbA	Adult haemoglobin
HbF	Foetal haemoglobin
Hct	Haematocrit
HLA	Human Leuococyte antigen
HPFH	Hereditary persistence of foetal haemoglobin
IEF	Iso-electric focusing
LCR	Locus control region
MCH	Mean cell haemoglobin
MCHC	Mean cell haemoglobin concentration
MCV	Mean cell volume
MHC	Major histocompatibility complex
mRNA	Messenger RNA
PCR	Polymerase chain reaction
PND	Pre-natal diagnosis
RBC	Red blood cell
RNA	Ribonucleic acid
RT-PCR	Reverse transcriptase-PCR
SDS	Sodium dodecyl sulphate
SNP	Single nucleotide polymorphism
SSPE	Sodium chloride; sodium dihydrogen orthophosphate; EDTA
T	Thymine
URR	Upstream regulatory region
WHO	World Health Organization

Chapter 1 Introduction

In Sri Lanka, many patients present with a clinical diagnosis of thalassaemia. Generally, this group of patients have been neglected and have been considered a burden on society. Any biochemical knowledge of this disease on the island is anecdotal from protein studies on a handful of patients. The molecular basis for the thalassaemia syndromes in Sri Lanka is unknown. The population of Sri Lanka is thought to have derived from the subcontinent of India and hence the pattern of haemoglobinopathies would be predicted to be similar.

An understanding of the nature and incidence of the thalassaemia syndromes in Sri Lanka will enable its Department of Health to develop a health plan to treat this disease and also to establish a system of education to improve the awareness within the population. Many factors modify the disease, some favourably and some to the patient's detriment. A close examination of this syndrome may help correlate phenotype with genotype. As each factor is identified further genetic and phenotypic variation can be scrutinised and may give a greater comprehension of the pathophysiology of the disease and subsequently modify the treatment of each patient. The spectrum and frequency of mutations will allow a better understanding of the evolutionary pressures and ties with other populations.

1.1 Sri Lanka

1.1.1 Sri Lanka: interaction with other countries

Sri Lanka is a geological extension of the South Indian peninsular and is separated by a narrow strip of water called the Palk Straits, 25 miles wide. This separation is thought to have occurred on a number of occasions as determined by the fluctuations in the global climate, and last occurred about 7,000 years ago in the Miocene epoch of the tertiary period. The

climate is tropical and warm all year, except for the centre of the island that is mountainous and cool (Encyclopaedia Britannica).

The connection between South India and Sri Lanka was strengthened by archaeological findings. Similar artefacts were found in both regions dating from the Palaeolithic, or Old Stone Age (~1,750,000 years ago) and also from the Mesolithic period. The first settlers on the island were probably tribes of the proto-Australoid ethnic group akin to the pre-Dravidian hill tribes of South India. They would have been a nomadic people, isolated from other populations. The population density for 'hunter gatherers' is estimated to be about 0.5 person/km^2, giving an island population between 30-35 thousand. In fact, the land could have supported in good times up to 55,000 people. The current day Veddah population may have a direct genetic continuum from these people (Encyclopaedia Britannica).

At the commencement of the 1st Millennium BC there are indications of a rapid transition from Mesolithic culture to the Early Iron Age. The transformation from 'hunter gatherers' to farmers would have been responsible for this increase in the population size. Still, it was necessary for medium and long distance trade to increase the wealth and act as the catalyst for the increase in the density of settlements. Evidence of this population expansion can be seen at Anuradhapura with a large settlement dating back to 1,200 BC. The settlement exceeds 10 hectares and by the 3rd century BC it was estimated to be the 10th largest city on the Indian subcontinent (de Silva, 1997).

The first implications of migrations to Sri Lanka from India are recorded in an early Indian legend, '*The Ramayana*', that was written about 500 BC. The epic of '*The Ramayana*' tells of the conquest of Lanka in 3,000 BC by Lord Rama to save his abducted wife Sita from Ravanna the demon god of Lanka and his hordes.

During the era of the Classical Age (c500 BC to 500 AD) a small number of records were passed on in written form. The most important early chronicle relating to the contemporary history and to the Buddhist faith on the island was *'The Mahavamsa'* which literally means 'The Great Chronicle'. Its continuation was called *'The Culavamsa'*, meaning 'The Little Chronicle'. These accounts are credited to the Buddhist monk, Mahanama, in the fifth and sixth century AD and Dhammakitti in the 13th century AD.

According to *'The Mahavamsa'* a band of 700 Indian colonists headed by Prince Vijaya landed on the West Coast of Sri Lanka near Puttalam in the 5th century BC. The precise origin of this legendary figure is obscure but he probably hailed from the North East of India. His father may have established himself in Gujarat from where the adventurers were put out to sea. This band of warriors defeated the local inhabitants, the yaksas (demons) whom they chased into the island interior. The prince married a yaksa princess and had two children. Later he drove her and the children away and their children are considered the ancestors of the present day Veddahs by folklore. Vijaya then sent to the Madurai court (Pandyan Kingdom) in South India for a princess for himself and for wives for his 700 followers. As he had no heir he sent back a message to his brother so that he could accede to the throne. However, his younger brother had already ceded to the throne in North India after the death of their father. Instead his brother sent his son, Panuvasudeva, in a small entourage of 32. They landed at Trincomalee on the East Coast of Sri Lanka after travelling across the Bay of Bengal. He was enthroned and continued the Vijaya dynasty that lasted up until AD 65 but is considered to be the possible origins of the Sinhalese (de Silva, 1997; Encyclopaedia Britannica).

Linguistic affinities between the Sinhalese speaking population of Sri Lanka and the Prakit-speaking people of Eastern India strengthen the

hypothesis of a migration from this area. However, at this time North India was unsettled. The great Persian leader, Emperor Dareios (521-485 BC) conquered large parts of North India via the Indus valley and along the reaches of the Ganga. The far south remained isolated separated by mountain, river and a vast tract of barren inhospitable land. The great struggle for control of North India raged and may have ensued for many generations, creating a catalyst for the migration of a group to move south. Later, traders from the Roman and Persian Empires established outposts in South India and probably acquired goods from Sri Lanka.

The precise date of the conversion of the people of Sri Lanka to Buddhism was thought to be during the reign of King Devanampiyatissa (250-210 BC). The founder of Buddhism, Gautama Buddha, was born in modern day Nepal. He died in 483 BC after teaching in Bihari and Eastern Uttar Pradhesh in India (at the same time as the arrival of Prince Vijaya in Sri Lanka). Buddha is thought to have made three trips to the island that may have implied a strong trading route. The Buddhist religion may have been carried along with this movement (Encyclopaedia Britannica).

Dominion, by peoples of North Indian origin, was followed by the Lambakanna dynasty that ruled for about four centuries. The origin of this group of people is obscure. During this period a major irrigation system was constructed. This would probably have resulted in a population boom. A Pandyan invasion (a Tamil empire from the tip of south India) ended the dynasty in AD 432. However, Sinhalese rule was soon restored in AD 459 with a line of Moriya kings. The capital was moved away from Anuradhapura for only a short time.

Arab traders were active from around the time of 'Christ'. With the decline of the Roman Empire Roman trading waned and the Arabs and Persians filled up the vacuum. The Arabs formed settlements and land was set-aside for them around Anuradhapura. By the 8th century AD the Arabs

had formed colonies at important ports around India and Sri Lanka. The Moors of Sri Lanka have no direct connection with the people of North Africa. They were called Moors because the Portuguese gave this name to all Islamic communities they encountered in the Far East. The Moors of Sri Lanka are probably descendants of Arab, Persian and Indian settlers who intermarried with the local people but have retained their Muslim identity.

In the middle Ages, especially from the 7th century there was an increase in the involvement of South Indian powers in the island's politics. With Indian backing Manvamma founded the second Lambakanna dynasty that reigned in Anuradhapura for about 400 years. Sinhalese kings were drawn into the dynastic battles between the three empires of South India; the Pandyas, the Pallavas and the Colas. The Pallavas had an empire near modern day Chennai, the Pandyas were furthest south and the Colas had an empire between the others. Invasions from South India and retaliatory raids were recurrent phenomena. In 993 the Colas occupied the kingdom and annexed Rajarata (part of Sri Lanka) as a province of the Cola Empire. The conquest was complete in 1017, when the Colas seized the southern province of Ruhuna.

The Colas occupied Sri Lanka until 1070, whereupon Vjayabahu liberated the island and re-established Sinhalese power. During this period the kingdom enjoyed its greatest prosperity. The most colourful king, Parakrambahu, followed a strong foreign policy and sent a punitive naval expedition to Burma and an army to invade the Panyan kingdom but achieved no permanent success. After Parakrambahu the throne passed to the Kalinga dynasty and the influence of South India increased. The last Polonnaruwa king was Magha (1215-1236), an adventurer from South India who ruled as a dictator.

Vijayabuhu was a feudal chief who, after many years of successful campaigning, made himself overlord of Mayarata and the western seaboard

of Rohana. By 1232, he assumed royal status and was succeeded by his son Parakramabahu II in 1236 who became the main ruler of the country until his death in 1270. In 1247 the Malay king, Chandra Bhanu, invaded Sri Lanka to possess the relics of Buddha. He ruled Northern Sri Lanka for 50 years and can be remembered by place names like Jaffna. During the 13th and the first half of the 14th century control of the island switched between parts of South India (de Silva, 1997).

In 1406 the Chinese attempted to take over the island and managed to capture the chief minister. They took him to China and then returned with a new fleet to gain power. However, although present for 8 years the Chinese never gained a real power base. In the interim Parakramabahu VI had already pre-empted them and continued his reign from 1411 to 1465 in peace and tranquillity. The succession of his grandson lasted for only two years before the dominion of the whole country came under dispute. This changed hands between different groups over the next 50 to 60 years (de Silva, 1997).

In 1506, the Portuguese discovered the island and were desperate to set up a trading post. As the first western power to gain dominion over the island they would have access to the island's riches. In time the demands of the Portuguese became greater and they converted many to Catholicism. By 1551, the island was split into three power bases; the Indian controlled Sitavaka in the north, the Portuguese around the Kotte and the indigenous forces of Kandy. The more powerful forces of the North tried to gain control but were rebuffed by an alliance between the other two powers. A new king in Kandy seized power and returned the populous to the Buddhist faith and expelled the Portuguese. The Portuguese turned their attention to the North and were soon in control of the Kingdom of Jaffna. By forming alliances with the people of the north, and with naval support, the Portuguese retained control of Jaffna. The Portuguese were keen to gain

complete dominion over the island. By 1620 only Kandy remained out of their control. Numerous attempts to gain control of Kandy were unsuccessful.

The Portuguese control of Sri Lanka rested on their command of the sea by their superior military technology. This pattern was changed upon the arrival of the Dutch. In 1636, the Kandyan leader Rajasinha II, appealed to the Dutch to help vanquish the Portuguese. Incensed, the Portuguese invaded Kandy but were annihilated by a decisive battle. In 1638 the Dutch seized control from the Portuguese and finally ousted them completely from Colombo in 1658. Pockets of people, Burghers, speaking a Portuguese Creole survive to the present day that also includes descendants of other groups like the Dutch (de Silva, 1997).

The first Dutch ships anchored off the port of Batticaloa in 1602. After the exit of the Portuguese, Rajasinha II requested control of all formerly Portuguese held land. The Dutch gave back some land but retained possession of the ports and some coastal areas and demanded unrealistic tribute for their assistance. The Dutch began to take control over the import and export of all goods. Malay immigrants were brought across to help work on the plantations. The Kandyan people began to become impoverished. In the 1670's the ruler, Rajasinha II, organised attacks on the Dutch-held territories. Although they could not retain control of these areas they changed the Dutch policy in Sri Lanka from one of aggression to one of peace. Hence, the Dutch did not attempt to control these areas but tried to gain sovereignty instead and act as agents of the king of Kandy. Mutual restraint upon each other existed up until the 1750s. The Dutch attempted to promote mixed marriages but the policy was a clear failure.

In 1762 the king of Kandy, Kirti Sri Rajasinha, entered into negotiations with the English. This stimulated a great reaction by the Dutch. In 1764 they devastated Kandy and in 1766 the Dutch were granted

sovereignty over the 'maritime regions' that included all coastal areas up to four miles inland. The landlocked Kandyans were humiliated and resentful of the treaty. Control over the island remained in tense state for another 30 years. In 1782 the English, who were at war with the Dutch in the War of American Independence, extended this globally (de Silva, 1997).

In 1796 the English conquered the 'maritime provinces' of the island. The replacement by the British occurred without any territorial or commercial benefit to the Kandyan Kingdom. By a process of divide and conquer the British were ceded power by 1815. They did not respect treaties to allow unreserved acceptance of Buddhism. This resulted in early attempts at rebellion that were overwhelmed by superior military 'fire power'.

By the beginning of the 19th century the mainstay of the Sri Lankan economy was subsistence agriculture. Even though major cities had road connections by 1830, most people lived in village communities whose communication was along jungle paths. The development of a network of modern roads and railways led to a much greater economic integration. Between 1840 and 1870 a coffee boom led the development of a great number of plantations on the island but came to a premature halt with the advent of the leaf fungus. The coffee picking season coincided with the slack season in the paddy fields of South India. Thousands of Indian workers began to make a seasonal journey even though the death rate was high whilst travelling across the malarial regions of the dry zone. From about 1860 there was a conversion to the production of tea in the highlands and rubber at lower elevations. The tendency was now for Indian labourers to settle down in the estates as a permanently resident labour force. From 1891 to 1911 the population of Sri Lanka increased by 517,000 due to the migration of Indian Tamils that settled mainly in the central highlands.

A system of government was slowly evolving that included an input from the people of Sri Lanka (de Silva, 1997). Only with the advent of two world wars and the decline in power of the British Empire was independence granted to Sri Lanka in 1948. Since independence, transitory conflicts between linguistic and religious groups have occurred. The exception is the rift between the Sinhalese and the Tamils.

India exerted a great influence as a powerful neighbouring force in favour of the Tamils. In 1964, an agreement between the two powers was reached. 525,000 persons were to be repatriated over a period of 15 years into India as long as 300,000 were accepted as citizens of Sri Lanka. In 1974, the equal sharing of a further 150,000 was agreed. In 1986, Sri Lanka accepted 96,000 from what was originally the Indian share. Finally, by the early 1990s the Indian Tamils (Tamils that had emigrated within the previous hundred years) who still remained in Sri Lanka were granted citizenship.

Sinhalese-Tamil violence began between 1972 and 1976 with Tamil youth in Jaffna and gained support to become the Liberation Tigers of Tamil Eelam. This erupted into widespread riots in 1977. These troubles were to flare up onto a larger scale on a number of occasions. In 1983 the riots became so bad that 30,000 people crossed over to South India. Tamil refugees have been accepted in many parts of the world (~ 100,000 in India 225,000 in Europe and North America). This ongoing social unrest in Sri Lanka has made parts of the island hazardous.

1.1.2 Population dynamics

The greatest check on population size was food limitation. Even though great areas of the earth became populated the size of the human population remained relatively small. As man learnt to farm and increase the supply of food a subsequent rise in the population number ensued. The life span of

these early humanoids was short and susceptible to disease. Irrigation allowed crop size to flourish but these conditions were also ideal for expansion of disease. An increase in the population size also increased the number of hosts to infect. Small populations were prone to devastation by disease and only when a population size reached a critical level, over half a million, would populations have a greater chance of surviving the infection. Re-population would need to occur either by greater reproductive capacity or by migration from surrounding areas. In early times epidemiological barriers were present due to the isolation of different evolving centres. However, large groups of people in quest of trade or emigration could carry infectious agents with them and increase the chances of spreading disease.

Areas closer to the equator represent areas with a greater number of more aggressive infectious diseases. Hence, north to south migrations or invasions would also have to contend not only with unfriendly endogenous populations but also with hostile new environmental dangers. The host population on the other hand had to adapt to survive and may have developed innate defence mechanisms against disease. Short periods between episodes of the same infestation could result in increased survival rates of those who originally survived that would be passed on to their children. Fitness to exist against disease was probably achieved very early in history

Historical records of human disease are poor but are present in some of the first writings of humans. Around 2,000-3,000 BC '*The Epic of Gilgamesh*' from Babylon describes pestilence. In chronicles from China around 1,300 BC there are also reports of pestilence. The Old Testament of the Bible records many plagues resulting in the death of large numbers of people. The exact nature of these diseases remains unknown but appears to occur in cyclical waves. Early diseases observed are thought to include

malaria, dengue fever, yellow fever, schistosomiasis, mumps, diptheria, tuberculosis and influenza (M^{c}Neil, 1976).

Growth

The world population from the year 10,000 BC has been estimated, allowing for all known census documentation and the 'best guess' for sustainable population sizes (http://www.census.gov/ipc/www/worldhis.html). The world population size has only rapidly increased in the last 500 years. A census of the population of Sri Lanka has been conducted since 1871 (http:www.statistics.gov.l/) and has shown an increase in size over that time period by almost 8 fold up to 18.6 million (Table 1.1).

Table 1.1 Census figures for the population of Sri Lanka from 1871 to 1997

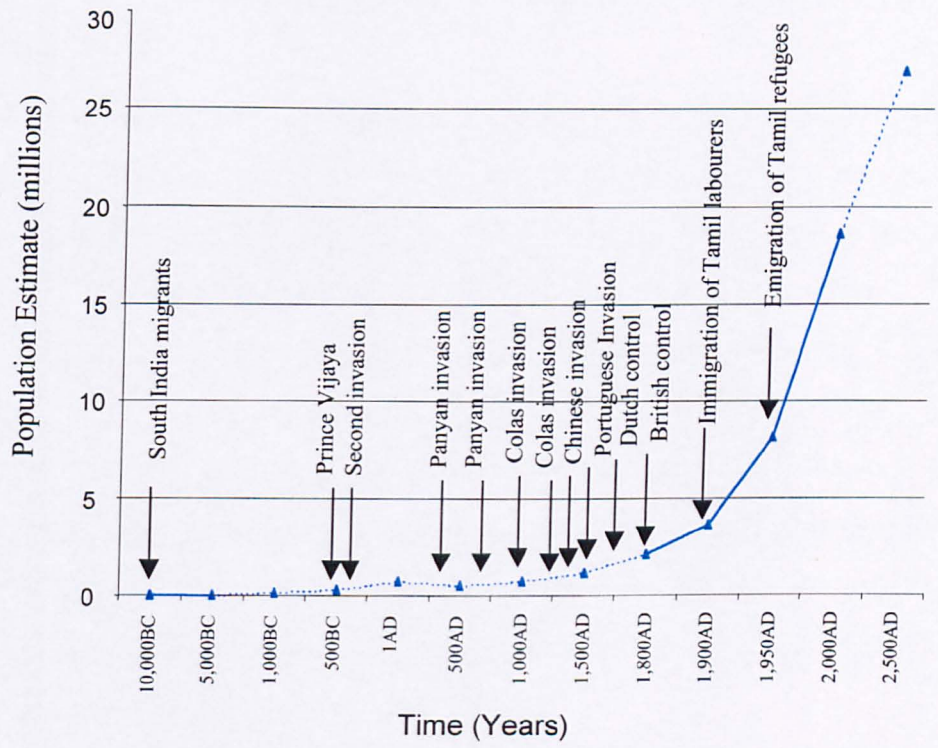
Year	Population (millions)
1871	2.4
1881	2.8
1891	3.0
1901	3.6
1911	4.1
1921	4.5
1931	5.3
1946	6.7
1953	8.1
1963	10.6
1971	12.7
1981	14.8
1997	18.6

Using these data and the estimate for world population size an estimate of the population of Sri Lanka can be extrapolated. The data are summarised as a population timeline (Figure 1.1) highlighting some of the important migrations, mainly onto the island.

From 10,000BC to 1,500AD the population size remained relatively small and the influence from external genetic input may have been large. If one generation is considered to cycle in 20 years 575 generations would have occurred to make that group relatively homogenous. In the last 500

years (25 generations) the population admixture would have been relatively small but a dramatic increase in the population size, over 20 fold in this same time period, may have skewed the relative importance of any group within the population greatly, either up or down.

Figure 1.1 **Time chart to show the rise in the population size in conjunction with migrations onto and from the island**



The genetic variation observed demonstrates the extended influence from the Indian subcontinent. The purity of the migrants entering Sri Lanka needs to be considered and the resulting ethnic groupings should not be considered to be distinct. The number in each different ethnic group was estimated as part of the government's annual census (Table 1.2) (http:www.statistics.gov.l/).

Table 1.2 The 1981 census; highlighting the prevalence of the ethnic groups in the population

Ethnic Group	Prevalence (%)
Sinhalese	74
Tamil	18.2
Moor	7.1
Burghers/Eurasian	0.3
Malay	0.3
Veddah	0.01
Others	0.09

Mortality

In 1993, a world-wide survey into causes of death was conducted by the WHO (WHO, 1995). The death toll was about 51 million people and the biggest killers were infectious and parasitic diseases that accounted for nearly one third. Diseases of the circulatory system accounted for nearly one fifth. Causes of death differ dramatically between the developed and the developing countries. In the developed countries chronic diseases have become the major causes of death. Diseases of the circulatory system (primary heart disease and cerebrovascular disease, e.g. stroke) and cancer accounted for over 40% and infectious diseases accounting for less than 1% of deaths. In the developing countries infectious disease accounted for 41.5% of deaths as listed in Table 1.3.

Sri Lanka has excellent health statistics (http:www.statistics.gov.l/) for a developing country. The present death rate is about 6.5 per 1000 and the infant mortality rate 16.5 per 1000 live births. Many of the disorders in Table 1.3, including all 5 of the most common causes of death, are still prevalent on the island. The island population would have been subjected to an array of these diseases for thousands of years. These disorders have contributed towards the selection of the genetic makeup of different people.

As the quality of life increases the number of deaths due to infectious disease decreases resulting in a rise in the population. The nature of the cause of death changes and a switch to chronic disorders in more elderly

patients occurs. All age groups benefit from these decreases in mortality but the most sensitive group to poor health conditions are the under 5 years of age group and these are used as an indicator of a population's health.

Table 1.3 Deaths due to infectious disease globally in decreasing order of prevalence

Number of deaths	Disease/Condition
$1\text{-}5X10^6$	Acute lower respiratory infections (<5 years old)
	Diarrhoea including dysentery (<5 years old)
	Tuberculosis
	Malaria
	Measles
$10^5\text{-}10^6$	Hepatitis B
	AIDS
	Whooping cough
	Bacterial meningitis
	Schistosomiasis
	Leishmariasis
	Congenital syphilis
	Tetanus
$10^4\text{-}10^5$	Hookworm disease
	Amoebiasis
	Ascariasis (roundworm)
	African trypanosomiasis (sleeping sickness)
	American trypanosomiasis (Chagas' disease)
	Onchocerciasis (river blindness)
	Meningitis
	Rabies
	Yellow fever
	Dengue haemorrhagic fever
	Japanese encephalitis
	Foodborne trematodes
$<10^4$	Cholera
	Poliomyelitis
	Diphtheria
	Leprosy
	Plague

(Adapted from Table 5 of WHO, 1995)

Since 1955, the fall in the mortality rate of children, under 5 years of age, in Sri Lanka has dropped by a factor of five. From 1995 to 1999 the death rate in this group was about 25 per 1000 live births and only Europe

had lower rates (Ahmad *et al.*, 2000). Individuals surviving include a greater number of those with debilitating disorders that may survive on into child bearing age to pass on their condition to another generation. It therefore follows that the total number of people with rare conditions is on the increase. This concurrent increase in chronic disease needs to be considered in all health care plans. The spectrum of genetic disorders seen is dependent on the range of selective forces, including infectious agents (Table 1.3), imposed on a population.

The significance of these types of changes was illustrated in Cyprus shortly after the Second World War. Thalassaemia had not previously been described in Cyprus. However, after a major programme to eradicate malaria and an improvement in public health measures, a common form of anaemia not due to infection was observed in the children of the island. Later studies demonstrated a very high incidence of thalassaemia on that island (reviewed in Weatherall & Clegg, 2001). A similar pattern of events was noticed in Sri Lanka and suggests a similar situation.

1.2 The thalassaemia syndromes

Haemoglobin is found in a wide variety of organisms in a broad diversity of forms. They are all considered to originate from a single ancestral gene and to have arisen by duplication followed by selection of adaptive sequence changes in a period of over 500 million years. The development of organisms from simple single cell to specialised multicellular forms could not have occurred without these molecules and their complex expression patterns have been investigated for over a hundred years.

Although the cessation of expression of haemoglobin is not compatible with life, intensive study has shown that this gene has undergone many alterations and has been subjected to extreme selection pressure. Genetic disorders of haemoglobin are the most common monogenic disorder in the world with over 250 million carriers for a haemoglobinopathy.

1.2.1 Normal structure and function of Haemoglobin

The haemoglobins

In 1863, Hoppe-Seyler coined the term "haemoglobin" to describe the oxygen-carrying pigment of blood. All human haemoglobins are tetramers with a molecular weight of 64,500 daltons. They consist of two α-like and two non α-like globin polypeptide chains, each of which has a single covalently bound heme group made up of an iron atom bound within a protoporphyrin IX ring. In humans, the six known different globin polypeptide chains are designated α, β, γ, δ, ε, and ζ. Each chain consists of a specific sequence of amino acids. The primary structure was established for the α chain, containing 141 amino acids, and also for the β, γ, δ and ε-chain, that contained 146 residues (reviewed in Weatherall & Clegg, 2001).

Any maps, pages, tables, figures graphs,or photographs, missing from this digital copy, have been excluded at the request of the university.

The haemoglobins vary in structure during different stages of development, designed to meet different oxygen-transport requirements. During development two major 'switches' occur in the type of haemoglobins manufactured (Figure 1.2). Hb's Gower I ($\zeta_2\varepsilon_2$), Gower II ($\alpha_2\varepsilon_2$) and Portland ($\zeta_2\gamma_2$) are embryonic haemoglobins found in foetuses before 7 to 10 weeks of gestation (reviewed in Weatherall & Clegg, 2001). From 5 to 10 weeks of gestation, the first switch involves the simultaneous decrease in ζ- and ε-chain production and an increase in α- and γ-chain production (HbF;$\alpha_2\gamma_2$). The second switch occurs around the time of birth substituting β- and δ- for γ-chain production (HbA;$\alpha_2\beta_2$ and HbA_2;$\alpha_2\delta_2$).

At birth, the normal foetal haemoglobin (HbF) levels are 64.8 ± 1.2% and for 15 days remain at the same level. A rapid drop occurs over the next 15 weeks down to 10.2 ± 1.2% after which the rate of decrease slows dramatically and adult levels of less than 1% are reached by 3 to 4 years of age (Colombo *et al.*, 1976; Cheron *et al.*, 1989). Normal adults have 97% HbA and 2-3% HbF with a minor trace of HbA_2 (reviewed in Weatherall & Clegg, 2001). Other minor haemoglobin components are the result of post translational modification.

In the neonatal period during switching from foetal to adult haemoglobin both γ- and β-chains compete for α-chains. As α-chains bind to β-chains in preference to γ-chains a small amount of Hb Bart's (γ_4) results. Hb Bart's was first described in patients with clinical features of thalassaemia. During normal development Hb Bart's disappears over the first six months of life and is not replaced by HbH (β_4) (reviewed in Weatherall & Clegg, 2001). This is partly due to the chemical properties of these monomers as Hb Bart's is a stable protein compared to its adult counterpart HbH.

Functional Properties

The affinity of haemoglobin for oxygen is essential for it to fulfil its physiological role. Haemoglobin has a sigmoid-shaped oxygen dissociation curve that allows large changes in oxygen binding with slight changes in the partial pressure of oxygen. Binding of oxygen to one haem molecule facilitates binding to the other three. The quaternary structure of haemoglobin was shown to exist in two states; oxy- or 'relaxed' (R) and deoxy- or 'tense' (T) without any stable intermediates (reviewed in Weatherall & Clegg, 2001). The number and nature of contact points between the α- and β-polypeptide chains changes between these two states. The αβ contact points are highly conserved and any mutations at

these sites can change the quaternary structure and possibly the affinity of haemoglobin for oxygen. Oxygen affinity is also affected by a number of small molecules and environmental factors including organic phosphate, pCO_2, temperature and pH. Haemoglobin is very sensitive to small changes in pH. In the body tissues the production of CO_2 results in the formation of HCO_3^- and H^+. Protons can combine with deoxy haemoglobin and stabilise this state. The lower oxygen affinity of haemoglobin facilitates oxygen unloading in the tissues (reviewed by Perutz, 1978). Excess oxygen in the lungs enhances the opposite changes in pH and pCO_2.

Red cells have unusually high concentrations of 2,3-diphosphoglycerate (2,3-DPG). One molecule of 2,3-DPG sits in a pocket in deoxyhaemoglobin bound to specific β-chain residues (1, 2, 82, and 143 of both β chains). The importance of the binding is that 2,3-DPG stabilises the deoxy form of haemoglobin in preference to the oxy- form, thereby lowering the oxygen affinity of the molecule. The γ chain of Hb F lacks the β^{143} histidine residue, and the resultant decrease in binding of 2,3-DPG to Hb F accounts for the increased oxygen affinity of foetal red cells compared to that of adult red cells (reviewed in Weatherall & Clegg, 2001). This has a great significance in disorders that elevate levels of HbF.

1.2.2 Abnormal structure and function of haemoglobin

Background

The first clear clinical description of a patient with the condition later to be termed thalassaemia, as opposed to other childhood forms of anaemia, was made in the first part of the 20th century by Cooley and Lee, 1925 (reviewed in Weatherall & Clegg, 2001). The name "Cooley's anaemia" is still used to describe transfusion-dependent β-thalassaemia

major. In Italy, at the same time, a similar though milder condition was described in adult patients, by Rietti (reviewed in Weatherall & Clegg, 2001). By 1960, it became clear that thalassaemia was not just one disease but a set of syndromes and that the incidence was extremely widespread and not localised to the Mediterranean (reviewed in Weatherall & Clegg, 2001).

Over the same time period different types of haemoglobin were observed that included HbA, HbF, HbS and HbC. These molecules were not just involved in development but also linked as the underlying molecular defect. By the 1950s, it was suggested that haemoglobin was controlled by a pair of genes, at loci that were not linked (reviewed by Weatherall & Clegg, 2001). This was consistent with the observation that the chemical structure of haemoglobin consisted of two identical half molecules by Ingram in 1956 and Perutz and others in 1960 (reviewed in Weatherall & Clegg, 2001).

In the same period, the molecular basis of thalassaemia was suggested to be a defect, not only in the structure, but also in the rate of production of globin (reviewed by Weatherall & Clegg, 2001). The sequence of these constitutive polypeptide chains was determined and Ingram demonstrated that HbS differs from HbA by a single amino acid substitution in the β-chain (reviewed by Weatherall & Clegg, 2001). By 1981, the globin genes had been sequenced (Baralle *et al.*, 1980; Lauer *et al.*, 1980; Lawn *et al.*, 1980; Slightom *et al.*, 1980; Spritz *et al.*, 1980) and paved the way so that molecular diagnosis for thalassaemia could be undertaken. Many studies have since been made to define the molecular lesions underlying the thalassaemia's and their global distribution (reviewed by Weatherall & Clegg, 2001).

Classification

Genetic disorders of haemoglobin can be broadly placed into three groups. The first is that of the structural haemoglobin variants involving a change in the primary structure of one of the globin chains. The second are the thalassaemias that involve a reduction in the rate of production of globin(s) and the third group embodies a diverse set of conditions that result in the persistence of foetal haemoglobin.

1.2.2.1 Structural variants

Structural haemoglobin variants are common in many populations. These proteins have a substitution of one or more amino acids in at least one of the polypeptide chains. Many simple and rapid methods have been developed to characterise proteins by their chemical composition. These have been applied to whole red cell lysates rich in the protein haemoglobin. Identification of these haemoglobins was first reliant on protein analysis but has now switched to DNA analysis. The result has been the characterisation of over 800 haemoglobin variants (http://globin.cse.psu.edu/). Haemoglobin variants can directly or in conjunction with a thalassaemia be associated with a serious health problem. The common variants (HbS, HbC and HbE) are associated with disease but many of the rare variants are 'silent' and behave in an innocuous fashion.

Structural variants can be classified into five groups depending upon their physiological effects;

(i) Haemolytic anaemia that cause an increase in the destruction of red cells. These include unstable haemoglobins and those that cause red cell 'sickling'.

(ii) Hereditary polycythaemia due to haemoglobins that have a high oxygen affinity.

(iii) Hereditary cyanosis due to those haemoglobins that have a low oxygen affinity.

(iv) Thalassaemia phenotype type due to those haemoglobins that result in a reduced output of globin. These include highly unstable, chain termination mutants and chain fusion haemoglobins. Some haemoglobins behave differently in conjunction with either a normal or thalassaemic counterpart.

(v) Innocuous haemoglobins that still function like HbA.

These classifications are simplistic and overlap greatly and their importance depends upon the prevalence of these variants. HbS and HbC both cause haemolytic anaemia and reach polymorphic levels in parts of Africa. HbE is found at high frequencies in Asian populations that also express high levels of β-thalassaemia (reviewed in Weatherall & Clegg, 2001).

1.2.2.2 The thalassaemias

Thalassaemia syndromes are genetic disorders characterised by absent (globin^0) or deficient (globin$^+$) synthesis of one or more of the normal globin chains. A reduction in the production of any globin chain defines that type of thalassaemia.

α-globin gene locus

The human α-globin gene cluster is localised to the distal segment of the short arm of chromosome 16p13.3 approximately 130kb from the sub-telomeric region (Flint *et al.*, 1997). The entire telomeric region is G-C rich (54%) with a modal variation with peaks 80-90kb apart (troughs 46% and peaks 63%). Fourteen sharp peaks of high G-C content were observed that corresponded greatly with the CpG islands. The region is rich in repeat elements including CpA, short tandem

repeats and Alu repeats. The repetitive DNA has similarities to sequences at the ends of other human chromosomes that could allow the genetic exchange between non-homologous subtelomeric regions (Flint *et al.*, 1997).

The α-cluster is comprised of three expressed genes (ζ, α_2 and α_1), one non-expressed gene (θ) and three pseudogenes (φζ, $\varphi\alpha_2$ and $\varphi\alpha_1$). Each α-gene is divided into coding regions (exons) separated by intervening sequences (introns). The exon and intron-1 of the two α-genes have identical sequences. Intron-2 shows some sequence divergence but most differences are seen in the 3' untranslated region. There are also extremely important regulatory sequences, which lie outside the gene itself, that are essential for the expression of this gene. Three highly conserved promoter elements 5' of the gene (ATA box, CCAAT box and CACCC sequence) and the polyadenylation signal 3' of the gene (AATAAA) are involved in the initiation of transcription and subsequent cleavage of the RNA transcript. Regulation of the α-cluster is controlled by a locus control region (HS-40) positioned closest to the telomere (Flint *et al.*, 1997). The HS-40 is a DNase I hypersensitivity site that is thought to be the major regulatory element whose removal results in the inactivation of expression of α-globin (Higgs *et al.*, 1990).

Regulation of the α-globin locus

The control of α-globin expression is similar to any gene expression. DNA is 'made available' so that DNA sequence elements, gene transcription and translation can take place in a highly controlled fashion. Anything that upsets this balance can result in a reduced or altered expression of the α-chains (Table 1.4). If there is a deficiency of α-chains then the level of Hb Bart's, in the neonate, is increased which

may be the most sensitive indicator for α-thalassaemia (Weatherall, 1963). Elevated levels of this haemoglobin have been used as a crude indicator of the level of α-thalassaemia within many different populations (reviewed in Weatherall & Clegg, 2001).

Table 1.4 Main α-cluster gene re-arrangements

Condition/Genotype		Phenotype	Number of mutations identified
Duplicated ζ-gene	ζζ+	?Normal	3+
Duplicated α-gene	ααα+	Normal	5
Single α-gene deletion	-α or α-	α^+	7
	α-	α^0	1
Double α-gene deletion (small)	--	α^0	20
Double α-gene deletion (large)	--	α^0	8
Upstream deletions	αα	α^0	12
ATR-16	--	α^0	17
Non-deletion change of either α-gene	$\alpha^T\alpha$ or $\alpha\alpha^T$	α^+	42
ATR-X	αα	α^+	33

α-gene insertions

The α-globin gene is thought to be the ancestral gene of all the other globin genes that resulted by gene duplication and subsequent alteration and thus is part of an ongoing evolutionary process. ζ-Gene rearrangement has been reported (Weatherall & Clegg, 2001) but its clinical significance is uncertain. Any effect would only be observed early in embryogenesis as the genes are normally switched off early in pregnancy. The clinical significance of extra α-genes becomes apparent when found in combination with β-thalassaemia (Trent *et al.*, 1981; Kanavakis *et al.*, 1983; Kulozik *et al.*, 1987; Garewal *et al.*, 1994; Oron-Karni *et al.*, 1994; Ho *et al.*, 1998a; Beris *et al.*, 1999). The α/β-chain imbalance exaggerates and a greater degree of erythropoiesis results in β-thalassaemia intermedia.

α-gene deletions

The primary cause of reduced α-chain expression is due to deletions that remove one or both of the α-genes (reviewed in Weatherall & Clegg, 2001). The α globin genes are embedded within two highly homologous, 4kb duplication units. It is thought that the homologous regions have resulted from gene duplication and that the non-homologous segments may have arisen subsequently by insertion of DNA. Three homologous subsections (X, Y and Z) are divided by non-homologous elements (I, II and III). The most common deletions or insertions occur due to misalignment of the chromosomes during recombination. Reciprocal recombination, between Z segments (3.7 kb apart) produce, chromosomes with only one α gene ($-\alpha^{3.7}$) that cause α-thalassaemia or chromosomes with three α genes ($\alpha\alpha\alpha^{anti3.7}$). Recombination between homologous X segments (4.2kb apart) also produce chromosomes with one α gene ($-\alpha^{4.2}$) that cause α-thalassaemia or with three α-genes ($\alpha\alpha\alpha^{anti4.2}$). Further recombination events between the resulting chromosomes (α, αα, and ααα) can occur and may even result in a patient with even more α-genes (αααα) on one chromosome. Single α-gene deletions result in α^{+}-thalassaemia and are usually clinically and haematologically 'silent'. A different single gene deletion associated with α^{0}-thalassaemia may involve antisense RNA transcribed from the complementary strand (Tufarelli personnel communication).

Mutations that remove both α-globin genes are much less common but can result in no α-gene expressed from that chromosome (α^{0}-thalassaemia). Heterozygotes have a similar phenotype as homozygous α^{+}-thalassaemia patients. Upstream deletions that remove a segment of DNA associated with an erythroid-specific DNase 1-

hypersensitve site (HS-40) can ablate downstream globin expression (Hatton *et al.*, 1990; Higgs *et al.*, 1990).

Another form of α^0-thalassaemia is that of ATR-16 that is linked to mental retardation. Only a small number of patients with this disorder have been studied but the patients fail to inherit the entire α-cluster region on one chromosome and in some cases have other chromosomal abnormalities (reviewed in Weatherall & Clegg, 2001).

Non-deletion changes of the α-gene

Mutations that do not delete the genes include small changes in and around either α-gene, the non-deletion α-thalassaemias. These changes down-regulate the expression of α-chains by a number of different mechanisms. The change may result in a truncated, elongated or nonsense protein that may interfere with the normal globin chain subunits. This may have a direct effect on haemoglobin function or result in a protein that is unstable either as an mRNA or as a protein. The change in the sequence may involve motifs in the DNA that regulate enhancers, promoters, splicing or termination. Some mechanisms have still not been unravelled.

Another form of non-deletion α-thalassaemia is linked with mental retardation. No structural abnormalities of the α-cluster or of 16p have been found. All patients are male with mild α-thalassaemia and have a severe mental retardation with a remarkably similar facial appearance. This unusual syndrome results from an X-linked abnormality. The ATR-X protein is a novel member of the SNF2 subgroup of a superfamily of proteins with ATPase and helicase motifs. This group of proteins is involved in a wide variety of cellular functions including the regulation of transcription.

α-thalassaemia phenotypes

There are four clinical states depending upon the number of α-genes affected. The most severe form is Hb Bart's hydrops that have no functional α-genes. This condition is found in infants who are spontaneously aborted or die of severe hydrops shortly after birth (Lie-Injo & Jo, 1960). In these usual cases, over 80% of the haemoglobin is Hb Bart's (γ_4), which has a very high oxygen affinity, causing severe tissue hypoxia; the remainder is Hb Portland ($\zeta_2\gamma_2$) and Hb H (β_4). A handful of these cases have survived (some with the help of intra-uterine transfusions) and are treated in a similar fashion to β-thalassaemia major patients (reviewed by Chui & Waye, 1998).

An intermediate form is Hb H disease that has a single functional α-gene with anaemia, hepatosplenomegaly and jaundice. The red cells are microcytic and their haemoglobin content is decreased. On the peripheral smear, poikilocytosis (abnormally shaped), polychromasia (variation in haemoglobin content), and target cells are seen. Hb H inclusions are easily seen following incubation with 1% brilliant cresyl blue. Studies of globin chain synthesis suggest a α/β ratio of 0.3 to 0.4, rather than 1. This imbalance causes 20% or higher levels of Hb Bart's at birth and Hb H levels of 4% to 30% after the switch from γ- to β-chain synthesis is complete. Deficiency of α-chain synthesis causes a concomitant drop in HbA_2 ($\alpha_2\delta_2$) levels to 1-1.5% (reviewed in Weatherall & Clegg, 2001).

Heterozygous α^o-thalassaemia and homozygous α^+-thalassaemia individuals are relatively asymptomatic, but have a mild microcytic anaemia (10-12 g of haemoglobin per 100 ml of blood) and mild poikilocytosis and anisocytosis (excessive variation in size). At birth, Hb Bart's may reach 5% in cord blood and HbH can sometimes be observed

at very low levels. The α/β synthesis ratio is 0.6 to 0.75 (reviewed in Weatherall & Clegg, 2001).

Silent carriers, with the deletion of a single α gene (-α/αα), present with a slightly reduced or normal set of red cell indices. The reduction of α mRNA is insufficient to produce significant globin chain imbalance (α/β = 0.8 to 0.9) (reviewed in Weatherall & Clegg, 2001).

Non-deletion forms of α-thalassaemia that alter either the α_1 or α_2 gene can also cause a loss of α-gene expression (α^T for -) and were first described in 1977 (Kan *et al.*, 1977). At present there are over 40 types that appear to have a more severe effect on α-globin gene expression compared with simple deletions that remove one or other of the α-genes. In the heterozygous condition they tend to be associated with normal or mildly reduced red cell indices but in the homozygous, hemizygous or compound heterozygous condition are associated with HbH disease. This may be due to the bias in which patients with a more severe condition have been studied. However, there does not appear to be a compensatory increase in the expression of the remaining functional α-globin gene.

β-globin gene locus

The human β-globin gene cluster is localised on the short arm of chromosome 11 embedded in an array of olfactory genes (Bulger *et al.*, 1999). The entire region has a balanced pyrimidine:purine content but has both Alu and L1 families of repeats.

The β-cluster is comprised of five expressed genes (ε, $^{A}\gamma$, $^{G}\gamma$, δ and β) and one pseudogene (φβ). The β-gene is split into coding regions separated by intervening sequences (introns). All the globin genes show a high degree of sequence homology. A similar set of promoter elements and a transcription cleavage signals exist around the β-globin

gene as described in the α-globin gene. The β-genes are controlled by a locus control region (LCR) positioned upstream of the ε-gene. These are a group of DNase I hypersensitivity sites akin to the HS-40 in the α-cluster.

Regulation of the β-globin locus

The control of β-globin expression is similar to α-globin expression. However, along with all the control elements and protein structures that accompany transcription and translation, switching and coexpression between different forms of the β-gene can occur. Anything that upsets the normal chain balance can result in a reduced or altered expression of the β-cluster globins. The multitude of different mutations results in a diverse set of clinical conditions (Table 1.5).

Table 1.5 Main β-cluster gene arrangements

Condition/ Genotype		Phenotype	Number of mutations identified
Insert	β +6-7kb	β^+	1
Duplicated γ-gene	γγγ	Normal	3+
Fusion	δβ	β^++↑HbF	3
	$^A\gamma\beta$	Normal	1
Deletion	$(\beta\rightarrow 3')^0$	β^+ or β^0 ±↑HbF	14
		HPFH	6
	$(\delta)^0$	β^+	1
	$(\delta\beta)^0$	β^++↑HbF	8
	$(\gamma)^0$	γ^+ or γ^0	1
	$(^A\gamma\delta\beta)^0$	Normal	10
	$(\varepsilon^G\gamma^A\gamma\delta\beta)^0$	β^0	6
Non-deletion	β^T	β^+ or β^0	167
		Dominant	31
	δ^T	δ^+ or δ^0	18
	γ	HPFH	13
Unlinked	Normal β-cluster	↑HbF	2
Trichothiodystrophy	Normal β-cluster	β^+	1

β-gene insertions

Gene insertion within the β-gene is extremely rare in and has only been found in an isolated case. 6-7kb of sequence was inserted into the second intron and resulted in the reduction of normal expression from that allele down to about 15% (Divoky *et al.*, 1996). Gamma gene rearrangements, including extra gamma genes, have been observed in some populations especially in the Pacific Islands (Phillippon *et al.*, 1995).

β-gene deletions

Many different β-cluster mutations have been described mostly in isolated families. They are classified in accordance with the genes that have been deleted. Mutations that affect the β-gene or β- and δ-genes have been found in nearly all populations. These mutations have been found in individuals with reduced red cell indices. Heterozygous deletions that extend further 5' to the β-gene to remove the δ- and $^{A}\gamma$-genes, $(A\gamma\delta\beta)^{0}$ thalassaemia, have microcytosis, hypochromia but no anaemia. Deletions that remove the β-gene and extend 3' of the β-gene form a set of conditions called hereditary persistence of foetal haemoglobin (HPFH). Patients with these mutations have normal red cell indices but elevated levels of HbF. Patients heterozygous for even larger deletions $(\varepsilon^{G}\gamma^{A}\gamma\delta\beta)^{0}$ that remove nearly all the cluster are associated with the clinical phenotype of β-thalassaemia trait. Homozygosity for this condition would be incompatible with life. Upstream deletions led to the discovery of the importance of the LCR (reviewed in Weatherall & Clegg, 2001).

Non-deletion changes in the β-gene

Nearly 200 small aberrations are found in or around the β-gene. These changes down regulate the expression of β-chains by a number of different mechanisms (reviewed in Weatherall & Clegg, 2001).

Gene expression is a dynamic process that requires a balance between transcription, translation and degradation. The correct positioning of this transcription machinery is brought about by recognition of specific DNA sequences, before each gene, under the control of the LCR. The transcription complex is a multiprotein complex including RNA polymerase that transcribes DNA into mRNA. Many cofactors are required including basal transcription machinery as well as erythroid specific factors that include GATA-1, NF-E2, EKLF, FKLF and SSP. Pre-mRNA production requires the correct initiation, elongation and termination of the message. The pre-mRNA is then processed to splice out all introns and to have a cap and a polyadenylation signal. The resultant mRNA undergoes translation of the nucleotide sequence into a polypeptide. The polypeptide chains are then assembled and can be post-translationally modified into functional proteins. Unstable or damaged proteins are marked for degradation. A problem at any part of this complex process can result in the aberrant expression of any globin polypeptide.

The most common errors occur in the primary structure in or around the gene and cause a reduced expression of that subunit. Mutations involving promoter or enhancer motifs down regulate expression. Changes within the exons or introns may result in a truncated, elongated or nonsense protein that may interfere with the normal globin chain subunits. This may have a direct effect on haemoglobin function or result in a protein that is unstable either as an mRNA or as a protein. Conversely dysfunctional haemoglobins can

interfere with the normal tetramer and result in a dominant form of β-thalassaemia.

Several loci unlinked to the β-cluster have been identified as associated with the expressed phenotype of this cluster. Specific markers on chromosome 6q were found to be associated with an elevated level of HbF in an Indian family (Craig *et al.*, 1996) and cases of β-thalassaemia were found associated with mutations in the general transcription factor TFIIH in individuals with trichothiodystrophy (Viprakasit *et al.*, 2001).

β-thalassaemia phenotypes

The β-thalassemias can be classified into three levels of severity. The most serious form is β-thalassaemia major. At birth, affected infants are relatively normal because the change from γ-chain synthesis to β-chain synthesis has not occurred. However, by 6 months of age the infant develops a severely microcytic anaemia with anisocytosis, poikilocytosis and polychromasia (reviewed in Weatherall & Clegg, 2001). To maintain an adequate haemoglobin level, transfusions are usually required every 4 to 8 weeks. Affected children develop hepatosplenomegaly secondary to extramedullary hematopoiesis, and a characteristic facial appearance due to excessive intramedullary hematopoiesis. The bones have expanded marrow cavities resulting in pathologic fractures and a "hair-on-end" appearance on skull films. Other complications include cholelithiasis, susceptibility to infections, secondary hypersplenism, and delayed growth and maturation. Excess iron deposited in the heart, pancreas, liver, and other organs, damages tissue and leads to cardiac failure, arrhythmias, diabetes mellitus, and liver failure (reviewed in Weatherall & Clegg, 2001).

Individuals heterozygous for β-thalassaemia are usually asymptomatic. They have a mild anaemia with decreased MCV and MCH and a raised HbA_2. Chain synthesis studies in most β-thalassaemia heterozygotes yield α/β ratios of 1.5 to 2.5. Microcytosis, anisocytosis, poikilocytosis, and targeting and stippling of the red cells can be seen on the blood smear.

β-thalassaemia intermedia is a condition between the major and minor syndromes showing a moderate anaemia with decreased MCV and MCH and a raised HbA_2. Transfusions are only required on occasions in these patients. The condition may be due to the patient having only mild molecular lesions or severe molecular lesions in combination with a modifying factor (Ho *et al.*, 1998b).

Foetal haemoglobin may be increased in β- and δβ-thalassaemia, hereditary persistence of foetal haemoglobin, D_1 trisomy, some cases of thyrotoxicosis, megaloblastic and aplastic anaemia, leukaemia and various malignancies involving marrow, sickle-cell anaemia, and during pregnancy (reviewed by Cooper & Hoagland, 1972).

Interaction of thalassaemia with haemoglobin variants

Thalassaemia and structural variant haemoglobin genes may or may not interact. An interacting thalassaemia gene is one that causes an increased level of the variant haemoglobin chain in the individual heterozygous for the variant gene and a thalassaemia gene. When the presence of both the thalassaemia and variant genes do not increase the level of the haemoglobin variant, then the thalassaemia is non-interacting.

α-thalassaemia and α-chain variants

Patients with α^0-thalassaemia and also Hb Q-Thailand have a clinical picture similar to that of Hb H disease (Lie-Injo *et al.*, 1979).

Patients with Hb G-Philadelphia are found with the amount of the variant observed in a tri-modal pattern (Baine *et al.*, 1976). Individuals with 22% Hb G-Philadelphia have three normal α genes in addition to a $\alpha^{G\text{-}Phil}$ gene ($\alpha\alpha/\alpha^{G\text{-}Phil}\alpha$). Individuals with 30% Hb G-Philadelphia have a single $\alpha^{G\text{-}Phil}$ gene and two normal α genes ($\alpha\alpha/\alpha^{G}$-) and those with 41% Hb G-Philadelphia have one chromosome containing a single α^{G} gene and another with a single α gene (α-/α^{G}-) (Sancar *et al.*, 1980).

β-thalassaemia and β variants

In Hb S/β-thalassaemia, the β-thalassaemia gene interacts with the β^{s} gene in the heterozygote to increase the level of Hb S near Hb SS levels. The Hb S/β-thalassaemia compound heterozygote has a milder clinical course than the Hb S homozygote, and splenomegaly is a common physical finding. Anaemia is present in Hb S/β-thalassaemia and is characterised by microcytic red and target cells with occasional sickled forms (Serjeant *et al.*, 1979).

The first description of HbE/β-thalassaemia was made by Minnich and colleagues in 1954 (reviewed in Weatherall & Clegg, 2001). The base change at codon 26 (GAG→AAG) leads to the partial activation of a cryptic splice site resulting in a proportion of abnormally spliced mRNA (Orkin *et al.*, 1982a) and less HbE mRNA. In addition HbE is thought to be mildly unstable and susceptible to oxidant damage. HbE heterozygotes have about 25-30% HbE and are clinically normal and homozygotes for this condition are mildy anaemic. In homzygotes for this haemoglobin variant a very mild anaemia can be observed with little or no HbF (reviewed in Weatherall & Clegg, 2001). HbE is most commonly found in combination with β^{0}- or severe β^{+}-thalassaemia and presents in a wide spectrum of forms. Patients can be similar to the

severest forms of β-thalassaemia or conversely remain as 'silent carriers'. The rationale for this diversity is still poorly understood but the level of HbF is high in the cases that present with a milder phenotype and lower in those with profound anaemia.

DNA haplotype analysis to map globin genes and its regulators

DNA sequence variation is abundant in wild populations. All new variations start as a mutation that alters the primary structure of DNA. These alterations may occur at any point and depending upon that location may change the function of a protein. Such mutations may confer selective advantage or disadvantage or may be neutral to the survival of the host. The more advantageous the ensuing phenotypes or the older the age of the mutation results in a higher prevalence of the mutation in a population. Each mutation may be advantageous to different populations depending on the selective pressures that population is subject to. Mutations can take on many forms but the most common types are point substitutions. These single nucleotide polymorphisms (SNPs) can help unravel the change in populations to determine if they have migrated or have arisen on more than one occasion. Such selection forces can also help understand the susceptibility to change that human DNA can undergo.

The first DNA polymorphism described in the β-gene was Hpa-I. Its presence was linked to the normal expression of HbA and its absence linked to the expression of HbS (Kan and Dozy, 1978a). An intensive search was made for restriction endonuclease site polymorphisms along the β-cluster and led to the recognition of 17 common polymorphisms. Of these, 14 were found in all racial groups and were termed 'public' as opposed to 'private'. Patterns of DNA polymorphisms along a chromosome are termed haplotype. Several of these polymorphisms are

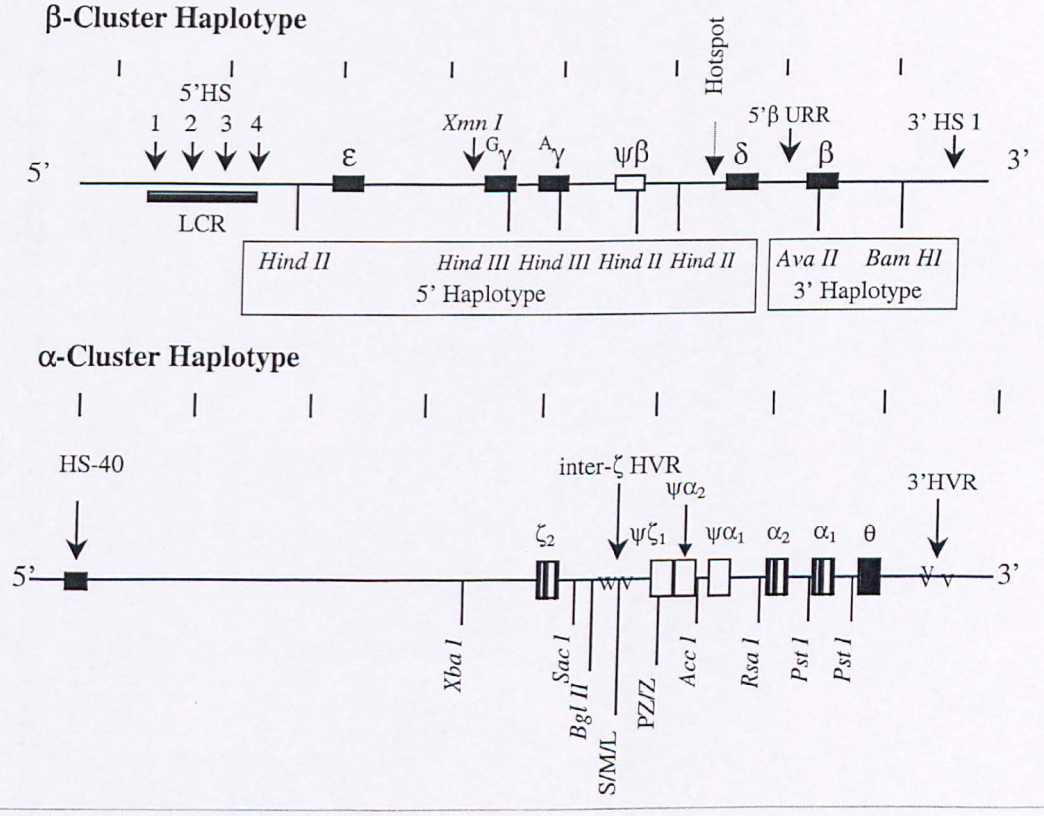
linked to each other and some form part of the framework of the β-gene. Only seven of these sites are routinely used to characterise the β-cluster and are arranged around a hot spot for recombination between the δ- and β-gene (Figure 1.3). Hence, the cluster is described as having a 5' and a 3' haplotype arranged around this hotspot.

Figure 1.3 Schematic of the α- and β-globin gene clusters to show the arrangement of the genes and some polymorphisms

(HS = DNase hypersensitivity sites; LCR = locus control region; URR = upstream regulatory region; HVR = hypervariable region; S/M/L = small/ medium/ large; PZ/Z = pseudozeta/ zeta)

(Restriction enzymes: *Acc I*; *Ava II*; *Bam HI*; *Bgl II*; *Hind II*; *Hind III*; *Pst I*; *Rsa I*; *Sac I*; *Xba I* and *Xmn I*).

The scale on both clusters is 10kb between vertical bars.



Common sequence variation in the β-gene must be extremely tightly linked to any mutation that occurs in or around the gene and can be classified as part of that association or framework as reviewed in

Orkin and Kazazian (1984). Population based background haplotype have been determined (Flint *et al.*, 1993) and the non-random association of these haplotypes with β-thalassaemia mutations observed (Antonarakis *et al.*, 1982). The assessment of the different biallelic markers were originally performed by Southern blot analysis but PCR-RFLP has substantially improved the ability to genotype many individuals rapidly (reviewed in Old, 1996).

In a similar fashion, the α-cluster has also been screened for restriction endonuclease site polymorphisms and multiallelic genetic markers (Higgs *et al.*, 1986). Nine of the most common polymorphisms have been used to obtain background baselines to observe the heterogeneity of different haplotypes associated with many populations (Figure 1.3) (Flint *et al.*, 1993). Recently, these sites have been successfully genotyped by PCR or by PCR-RFLP to allow a rapid population screen to be made (Miles *et al.*, 2001).

Pathophysiology

The symptoms of thalassaemia are mainly caused by ineffective erythropoiesis. Red blood cell precursors are extensively destroyed in the bone marrow leading to ineffective red cell maturation. Peripheral red cells are prone to haemolysis and hence have a shortened 'lifespan'. Microcytic hypochromic anaemia results in a positive feedback regulation for erythropoiesis and a massive expansion of bone marrow and this further exacerbates the condition (reviewed in Weatherall & Clegg, 2001).

In the thalassaemia syndromes there is reduction or ablation of synthesis of the affected globin chain. The failure in β-globin production leads to imbalance in α- and β-globin synthesis (Weatherall *et al.*, 1965; Weatherall & Clegg, 1969) as the unaffected chain continues to be

synthesised at relatively normal levels. The resultant imbalance causes aggregation and precipitation of excess unpaired α-chains. Excess α-chains aggregate in nucleated erythroid precursors in the bone marrow causing dyserythropoiesis. Red cells that enter into circulation are clonal and tend to be enriched for those containing HbF. Cells with excess α-chains become targeted for haemolysis and sequestration by the spleen. In contrast, in α-thalassaemia there is an excess of γ_4 (Hb Bart's) or β_4 (HbH) chains. These tetramers are more soluble and can be destroyed by proteolysis. Inclusion bodies can be seen in mature erythrocytes (a reflection of the magnitude of excess β-chains) followed by haemolysis (Bank & O'Donnell, 1969). Indeed patients with both homozygous β-thalassaemia and α-thalassaemia have a more even chain balance, less globin precipitation and hence more effective erythropoiesis (Weatherall *et al.*, 1981).

Haemolysis is associated with progressive splenomegaly. If the spleen is removed, cell destruction continues at a decreased rate in the liver, and the number of red cell inclusions may increase greatly. The large number of erythroid precursors expands the marrow cavities, and bone deformities, thinning, and occasional pathologic fractures result (reviewed in Weatherall & Clegg, 2001).

Iron overload is a major complication in thalassaemia. Iron accumulation results from increased gastrointestinal absorption stimulated by the anaemia, blood transfusions, and decreased utilisation for haemoglobin synthesis. Deposition of excess iron causes damage to the heart, pancreas, and other tissues. The major causes of mortality are due to excessive iron deposition as a result of blood transfusions and increased gastrointestinal absorption (Bannerman *et al.*, 1964) and overwhelming infections following splenectomy.

Management and treatment of thalassaemia

Pre-natal diagnosis (PND) and a programme of birth selection can prevent the birth of affected individuals. For maximum benefit of genetic counselling, heterozygotes should be identified before they bear affected children. In some communities this is an unacceptable practice contravening either moral or religious boundaries. However, the procedure has become widely used and in some communities couples are required to perform premarital blood tests to avoid the possibility of having an affected child.

Blood transfusions immediately replace the lack of fully functional red blood cells to allow tissue oxygenation but unfortunately introduce extra iron into the body that leads to iron overload. Iron chelation therapy clearly reduces the incidence of iron-related organ damage and prolongs survival of patients with severe thalassaemia (Olivieri *et al.*, 1994). However, side effects of transfusions include transmission of disease agents such as hepatitis and cytomegalovirus infections, isoimmunization to minor blood groups and sensitivity to plasma antigens (reviewed in Weatherall & Clegg, 2001).

Splenectomy has an immediate effect by halting the removal of immature or defective red blood cells and prevents pooling in the spleen. The release of a mass of ineffective erythropoietic tissue can temporarily provide great benefits especially during periods of growth. However, if carried out early in life splenectomy can leave children more prone to serious infection (reviewed in Weatherall & Clegg, 2001).

Allogeneic bone marrow transplantation using related donors has been used in a large number of patients, predominantly in Europe, with disease-free survivals in the range of 80 to 95% (Lucarelli *et al.*, 1990). HLA matching of donor and recipient must be near identical for a successful graft. Better outcomes are associated with adequate iron

chelation and the absence of significant hepatic disease. Nowadays, cord blood is seen as a rich source of stem cells, pluripotent precursors of different blood lineage and can be used in both homologous and heterologous transplants.

Patients with unusually high levels of HbF tend to have a milder disease and any augmentation in the level of HbF should diminish the severity of the disease. In preliminary studies, hydroxyurea and butyrate derivatives have shown promise in ameliorating anaemia (Jane & Cunningham, 1998; Olivieri & Weatherall, 1998). Pharmaceutical companies continue the search for effective pharmacological agents.

Gene therapy has been proposed as a preferred method so that the body can regulate the normal globins. Methods for insertion of DNA into eukaryotic cells include the use of viral or retroviral transfer, liposome mediated transfer, targeted gene transfer, artificial chromosomes and antisense oligonucleotides. Technical as well as ethical problems remain and these therapies remain to be resolved in the future (reviewed in Weatherall & Clegg, 2001).

Distribution of haemoglobinopathies

Thalassaemia has a high incidence in a broad region extending from the Mediterranean basin and parts of Africa, through the Middle East, the Indian Sub-continent, Southeast Asia and Melanesia and out into the Pacific. Elsewhere, thalassaemia occurs at very low frequencies. Countries in Central and South America have individuals with haemoglobinopathies due to migration from Europe. The mutation rate for 'thalassaemia' appears quite high and when coupled with the movement of populations result in the sporadic occurrence of thalassaemia in most populations. The prevalence of haemoglobinopathies in any region is taken as the sum estimate of all

the populations within that region. Vast differences occur between closely located populations so that micromapping of each area needs to be assessed.

α^0-thalassaemia is generally not found except in isolated families in most parts of the world. The exceptions are that of the Mediterranean basin where the $--^{MED}$ and $-(\alpha)^{20.5}$ gene deletions occur and the Far East that has a relatively high frequency for the $--^{SEA}$ deletion. In contrast, the α^+-thalassaemia has been reported at polymorphic frequencies in parts of Africa, Europe, the Middle East and Asia (reviewed in Weatherall & Clegg, 2001).

β-thalassaemia and some clinically important β-chain variants (HbC, HbD, HbE and HbS) reach very high frequencies in many sub-tropical and tropical countries. β-thalassaemia trait is generally found at frequencies below 10% but these variants may constitute up to 50% of some regional groups. A wider variety of molecular lesions cause β-thalassaemia with 21 different mutations accounting for over 80% of the global incidence of β-thalassaemia. Each country appears to have its own particular spectrum of about 15-25 mutations with 2 or 3 of these mutations predominating (reviewed in Weatherall & Clegg, 2001).

The high prevalence of haemoglobinopathies in sub-tropical climates is not just a random phenomenon. It was suggested that patients with heterozygous β-thalassaemia have some form of selective advantage described by Silvestroni & Bianco in 1948 (reviewed in Weatherall & Clegg, 2001). This advantage was proposed as a greater resistance of red blood cells to invasion by merozoites that cause malaria (reviewed in Weatherall & Clegg, 2001). This associated protection was extended to the structural variant HbS and more recently that HbC may confer an even greater protection against *Plasmodium*

falciparum (Modiano *et al.*, 2001)). This theory has widened to include α^+-thalassaemia protection against severe childhood disease (Allen *et al.*, 1997). Hence, different forms of thalassaemia are prevalent in some regions and not others due to a primary event that allowed selective advantage. This enabled pockets of individuals with particular genotypes to flourish. The presence of one type of haemoglobinopathy negates the necessity for another and so the regional incidence appears non-uniform.

1.3 Haemoglobinopathies in Sri Lanka

Very little is known about the occurrence and the molecular basis of α- and β-thalassaemia in Sri Lanka. Structural variants have been described by a number of different studies but the true incidence on the island remains unknown.

α-thalassaemia

Two case reports have been published. The first was of a child of 5 months of age who presented at hospital with fever, pallor and icterus. She already had a history of breathlessness, listlessness and fever. Upon investigation she was found to have a severe anaemia and the erythrocytes had inclusion bodies. The haematology of both parents appeared normal (Nagaratnam & Sukuram, 1967). The second case was a Muslim boy, 25 years old, from Matale who presented at hospital with a haemolytic anaemia, pallor, mild icterus and hepatosplenomegaly. Upon investigation by electrophoresis he was found to have a fast running haemoglobin consistent with HbH (de Tissera *et al.*, 1988). The molecular lesion underlying either of these cases remains obscure.

β-Thalassaemia and structural variants

Typical cases with Cooley's anaemia were first described in three different families (de Silva & Weeratunge, 1951) and also in a few more families later in the 1960s (Nagaratnam & Sukumaran, 1967; Parameshwaran, 1967). A slow running variant was described in 2/9 Veddahs (Graff *et al.*, 1954). HbE was formally confirmed in this population and a carrier frequency of 4% estimated (Lehman, 1956). Four Veddah villages were studied and showed an extremely varied incidence ranging between 0% to over 40% (Wickramasinghe *et al.*, 1963). Patients with HbE/β-thalassaemia were described whose clinical manifestation varied greatly (Nagaratnam *et al.*, 1958). Three cases of HbS/β-thalassaemia presented at hospital in Colombo and were described as having a severe anaemia and hepato-/ cardio-/ splenomagaly (de Silva *et al.*, 1962). The variant HbD was also found in a few families (Nagaratnam, 1989).

Chapter 2 Materials and Methods

2.1 Ethical permission

Approval for the study of patients from Sri Lanka was obtained from the Oxford Research Ethics Committee, the Oxford Tropical Research Ethics Committee, the Ethics Committee of the Faculty of Medicine at the University of Colombo and by the Ministry of Health Sri Lanka. Permission to investigate patients referred to the National Haemoglobinopathy Reference Laboratory was obtained by the referring physicians. Consent to conduct the Homerton Hospital cord blood study was organized by Dr A. Stephens.

2.2 Patients

Sri Lanka: School children survey

Capillary blood samples were collected by Dr S. de Silva, from over 1600 children and some of their teachers attending 15 different schools around the island (Figure 2.1). The schools were widely selected and included a Hindu Teaching college. The degree to which these samples represented the general population was not assessed as they were collected anonymously without a questionnaire. Specimens were successfully collected from eight provinces (Central, Eastern, North Central, North Western, Sabaragamuwa, Southern, UVA and Western) but the ninth province (Northern) was deemed too dangerous because of the ongoing war between the Government troops and the Tamil Tigers. Blood samples were taken into micro-haematocrit tubes, sealed at each end, and maintained at 4°C until they were transported to Oxford frozen on dry ice.

Sri Lanka: Hospital based patients

Patients with a clinical diagnosis of thalassaemia attending clinics as outlined in Figure 2.1 were included in the study. Prior to blood

transfusion, 2-5ml 'whole blood' was collected into EDTA. The plasma was removed and the cells stored at –20°C before shipment to Oxford on dry ice. Blood samples from siblings of β-thalassaemia major patients attending the teaching hospital in Kurunegala were treated in the same fashion as those in the school children survey.

Sri Lanka: Family studies

Family studies were initiated on cases presenting with more unusual or novel mutations. These were collected in the same manner as the Hospital based patients.

Sri Lanka: Cord blood study

The cord blood study was planned so that a medical student, S. Rajapaksha, could visit Sri Lanka to help co-ordinate the collection of specimens. Shortly after delivery cord blood was collected by midwives, and the blood stored at 4°C. Technicians then transferred the blood to a –20°C freezer prior to shipment to Oxford on dry ice.

United Kingdom: α^T-thalassaemia

The National Haemoglobinopathy Reference Laboratory (NHRL) is a tertiary referral centre. Blood from patients is routinely studied for α-thalassaemia, β-thalassaemia and identification of structural variants by DNA analysis. Whole blood was collected into tubes containing EDTA and sent by first class post from the referral centre to the laboratory in Oxford.

United Kingdom: Cord blood study

The district Haemoglobinopathy Unit at St. Bartholomew's hospital undertakes screening for thalassaemia and abnormal haemoglobins for the City and Hackney health district. Cord bloods of all newborns are routinely screened using a Biorad sickle cell thalassaemia column on the Biorad Variant Analyser. The remainder of the study cord blood, 2-3mls, was transported by first class post to Oxford in small batches.

2.3 Haematology and biochemical analysis

Full blood count

In Sri Lanka a full blood count was measured on whole blood using a Coulter counter microdiff 8. This was performed as per manufacturer's

instructions by a trained technician. Specimens referred to the NHRL were analysed on a Coulter counter (MD8 Model, Beckman Coulter Electronics) as per manufacturer's instructions.

Cation Exchange-High Performance Liquid Chromatography (CE-HPLC)

Fresh or frozen blood was analysed as per the manufacturer's instructions. 5µl whole blood was lysed in 1ml haemolysis reagent and maintained at 12°C until analysed. Specimens were allowed to bind to the positively charged resin. An increasing ionic strength elution buffer was then passed across the resin. Haemoglobin elution was detected at 415nm and the baseline was corrected for the effects of different ionic strength buffers by monitoring the eluate at 690nm. An absorbance chromatogram versus time was generated and the relative proportion of each haemoglobin estimated by calculation of the area under each peak. Every run was pre-calibrated with standards provided. All haemoglobins have a characteristic retention time but may share this elution profile with other variant haemoglobins. The β-thalassaemia short program was optimised for the evaluation of HbF and HbA_2; it also identified HbA and also proposes candidates for other haemoglobin variants. The sickle cell short program is a shortened version of the β-thalassaemia short program lasting 3 minutes rather than 6 minutes that was developed for high throughput of samples. The α-thalassaemia short program was optimised for the evaluation of Hb Bart's. It required a pre-heating coil to prevent sample-run variation and a different column to study the fast eluting proteins.

Isoelectric focusing (IEF)

IEF separation of haemoglobins was performed using the 'Isolab system' as per the manufacturer's instructions (Isolab Inc. Akram, USA). Packed red cells were prepared by mixing with 10 volumes of 'lysis' buffer

that containing a small amount of potassium cyanide. 10µl was loaded onto pre-cast gels and focused by applying 300v at 15°C for 45 minutes. The gel with separated proteins was then fixed by immersion in 10% trichloroacetic acid for 20-30 minutes, washed in water for an equal length of time and then allowed to air-dry overnight. Haemoglobin focuses within an electric field according to the isoelectric point of the individual protein. The relative positions of any observed haemoglobins can then be related to standard haemoglobins focused at the same time (HbA, HbF, HbS and HbC).

Cellulose acetate electrophoresis (CAE)

CAE is a rapid and simple method to separate proteins according to net charge and size. CAE was performed using a Shandon system (Shandon Model U77 Electrophoresis tanks). Red cell lysates were prepared by washing in 0.9% saline. An equal volume of distilled water and 0.4 volumes of carbon tetrachloride were added and vortexed for 30 seconds. The mixture was then centrifuged at 3000g for 20 minutes. The clear supernatant contained the haemoglobins and was stable at -70°C. For frozen blood 1 or 2 drops of 1% potassium cyanide was added to 1-2ml of red cell lysate. Cellulose acetate strips (25mm x 120mm; Shandon, Celogram) were used with a phosphate buffer system at pH 6.5. Electrophoresis was completed in 1-2 hours using a constant current of 5mA. Hb Bart's was observed as a fuzzy band distinct from all other haemoglobins.

Glucose-6-phosphate dehydrogenase (G-6-PDH) deficiency

G-6-PDH deficiency was determined using a simple colorimetric test from Sigma Chemical Co. (Cat No. 400-K25) as per manufacturer's instructions. G-6-PDH released from lysed erythrocytes catalyses the conversion of glucose-6-phosphate to 6-phosphogluconate with reduction

of nicotinamide adenine dinucleotide phosphate (NADP) to NADPH. In the presence of phenazine methosulphate (PMS), NADPH reduces the blue dye, dichlorophenol indophenol, to the colourless form. The rate at which colour disappears in the reaction mixture is proportional to the G-6-PDH content of red cells.

Red cell haemolysates were prepared by pipetting 50µl whole blood into 2.5ml water and allowed to stand at room temperature for 5 minutes after gentle agitation. 1.0ml haemolysate was added to 0.5ml substrate solution containing glucose-6-phosphate and NADP in a tris buffer, mixed gently and then overlayed with 2.0ml mineral oil. The vial was then placed at 37°C in subdued light. The mixture immediately turns maroonish/blue that will clear over a 20-60 minute period. Any failure to convert the dye from its 'blue' colour was interpreted as a G-6-PDH deficiency. This reaction was compared visually against reference samples (normal, intermediate and deficient controls).

2.4 DNA analysis

α-gene analysis: Southern blot

Genomic DNA was digested overnight at 37°C with either restriction enzyme, *Bam HI* or *Bgl II*. The fragmented DNA was electrophoresed on 0.8% agarose and then capillary blotted (Sambrook *et al.*, 1989) onto nitrocellulose membrane (Zeta probe, Biorad). Southern blot hybridization was performed, at 65°C, using a Church & Gilbert based buffer (7.5% dextran sulphate, 1.5X SSPE, 1.0%SDS, 5X Denhardt's solution, 100µg/ml denatured and sonicated salmon sperm DNA). Previously described probes; (α) a 1.5kb *Pst I* fragment and (ζ) a 1.8kb *Sac II* fragment (Lauer *et al.*, 1980; Proudfoot *et al.*, 1982) were radio-labelled with ^{32}P-dCTP using a Megaprime DNA labelling system (Amersham Pharmacia). These conditions allow the detection of common forms of α-gene rearrangement;

$-\alpha^{3.7}$, $-\alpha^{4.2}$, ααα, αααα, $-(\alpha)^{20.5}$, $--^{Med}$, $--^{SEA}$, $--^{SA}$ (Table 2.1) and some extremely rare deletions.

Table 2.1 Fragment sizes observed by Southern blot analysis for different α-gene arrangements

α Allele	Restriction digest and hybridisation probe			
	BamH I - α	*Bgl II* - α	*BamH I* - ζ	*Bgl II* - ζ
αα	14	12.6/7.6	10 -11.3/5.9	12.6 or 5.2/10 -11.3
$\alpha\alpha\alpha^{anti3.7}$	17.7	12.6/7.6/3.7	10 -11.3/5.9	12.6 or 16/10 -11.3
$\alpha\alpha\alpha^{anti4.2}$	18.2	16.8/7.6	10 -11.3/5.9	12.6 or 16/10 -11.3
$\alpha\alpha\alpha\alpha^{anti3.7}$	21.4	12.6/7.6/3.7	10 -11.3/5.9	12.6 or 20/10 -11.3
$\alpha\alpha\alpha\alpha^{anti4.2}$	22.4	21/7.6	10 -11.3/5.9	12.6 or 16/10 -11.3
$-\alpha^{3.7}$	10.3	16	10 -11.3/5.9	16/10 -11.3
$-\alpha^{4.2}$	9.8	8.0/7.4	10 -11.3/5.9	10 -11.3/8.0
$--^{SEA}$	None	None	20/5.9	10.5
$--^{SA}$	None	None	5.9	7.0

α-gene analysis: α^+-thalassaemia Gap-PCR

A Gap PCR method was developed for the detection of the two common forms of α^+-thalassaemia, $-\alpha^{3.7}$ and $-\alpha^{4.2}$ (See figure 2.2). The primers for the detection of α^+- and α^0-thalassaemia were designed by Mr Y.T.Liu and the details have been published (Liu *et al.*, 2000). Alternate primers published at the same time were also optimised (Chong *et al.*, 2000 a). The primer sequences are detailed in Table 2.2 and the location of these primers in relation to the α-genes with the relevant deletions are shown in Figure 2.2.

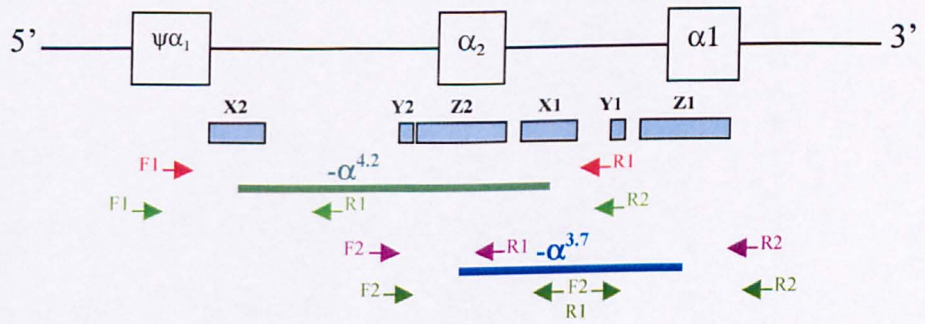
The buffer conditions were optimised for a total reaction volume of 25μl with 100ng target template in Buffer A (Table 2.15). Products were separated on a 1% agarose gel after amplification with the following cycling conditions;

95°C – 3 ' X1	**95°C – 1'**	**72°C – 10' X1**
	64°C – 1'30" **X30**	
	72°C – 1'30"	

Table 2.2 α-gene Gap-PCR primers

Mutation	Oligonucleotide sequence (5'→3')	GenBank Co-ordinates Z84721 *Z69706	Primer concentration (μM)	Product size (bp)
$-\alpha^{3.7}$ multiplex				
$-\alpha^{3.7}$ F (Liu et al)	AAGTCCACCCCTTCCTTCCTCACC	32755-32778 36309-36332	0.3	αα 2217 αα 2213
$-\alpha^{3.7}$ R1 (Liu et al)	ATGAGAGAAATGTTCTGGCACCTGCACTTG	34942-34971 30685-30714	0.1	$-\alpha^{3.7}$ 1963
$-\alpha^{3.7}$ R2 (Liu et al)	TCCATCCCCTCCTCCCGCCCCTGCCTTTTC	38492-38521	0.1	$\alpha\alpha\alpha^{anti3.7}$ 2440
$-\alpha^{3.7}$ F(Chong et al)	CCCCTCGCCAAGTCCACCC	32746-32764	0.4	$-\alpha^{3.7}$ 2020
$-\alpha^{3.7}$ R1(Chong et al)	AAAGCACTCTAGGGTCCAGCG	38562-38582	0.4	αα 1800
$-\alpha^{3.7}$ R2(Chong et al)	AGACCAGGAAGGGCCGGTG	34530-34548	0.1	
$-\alpha^{4.2}$ multiplex				
$-\alpha^{4.2}$ F (Liu et al)	TCCTGATCTTTGAATGAAGTCCGAGTAGGC	30283-30312	0.8	αα 1510
$-\alpha^{4.2}$ R1 (Liu et al)	TGGGGGTGGGTGTGAGGAGACAGGAAAGAGAGA	31760-31789	0.4	$-\alpha^{4.2}$ 1725
$-\alpha^{4.2}$ R2 (Liu et al)	ATCACTGATAAGTCATTTCCTGGGGGTCTG	36176-36205	0.4	
$-\alpha^{4.2}$ F(Chong et al)	GTTTACCCATGTGGTGCCTC	30132-30151	0.4	$-\alpha^{4.2}$ 1628
$-\alpha^{4.2}$ R(Chong et al)	CCCGTTGGATCTTCTCATTTCCC	35983-36005	0.6	
α^{0} multiplex				
α^{SEA} F	CTCTGTGTTCTCAGTATTGGAGGGAAGGAG	26129-26158	0.4	αα 1010
αR	TGAAGAGCCTGCAGGACCAGGTCAGTGACCG	27108-27138	0.4	α^{SEA} 660
α^{SEA} R	ATATATGGGTCTGGAAGTGTATCCCTCCCA	3148-3177*	0.4	
α^{FIL} F	AAGAGAATAAACCACCCAATTTTTAAATGGGCA	12283-12315	0.4	α^{FIL} 550
α^{FIL} R	GAGATAATAACCTTTATCTGCCACATGTAGCAA	544-576*	0.4	
α^{THAI} F	CACGAGTAAAACATCAAGTACACTCCAGCC	10364-10393	0.2	α^{THAI} 411
α^{THAI} R	TGGATCTGCACCTCTGGGTAGGTTCTGTACC	1241-1271*	0.2	

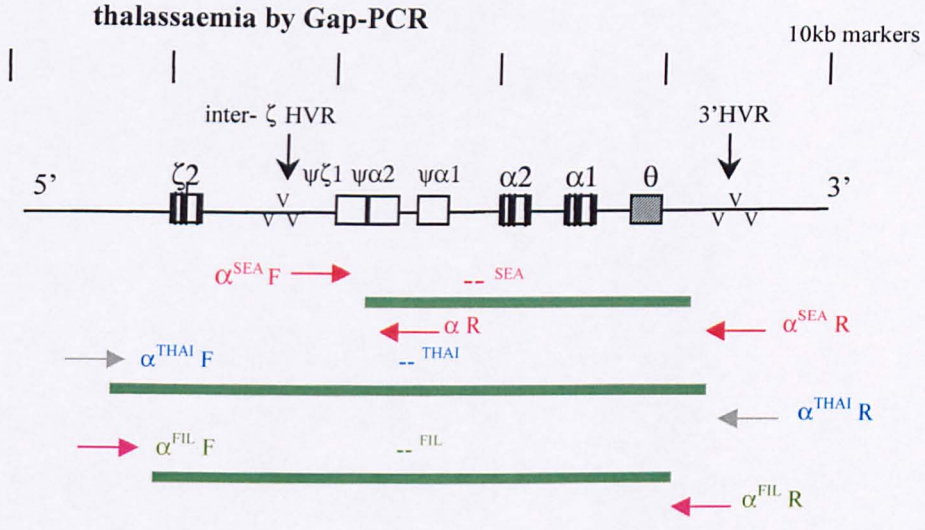
Figure 2.2 Schematic to highlight the location of the primers to detect α^{+}-thalassaemia by Gap-PCR



α-gene analysis: α^0-thalassaemia Gap-PCR

Assays for the detection of the deletions ($--^{SEA}$,$--^{FIL}$ and $-^{THAI}$) were successful under the conditions stated in the paper (Liu *et al.*, 2000) and also using alternative buffer conditions, DNA polymerases and PCR machines. The primer sequences are detailed in Table 2.2 and the location of these primers on the α-cluster depicted in Figure 2.3.

Figure 2.3 Schematic to highlight the location of the primers to detect α^0-thalassaemia by Gap-PCR



Products were separated on a 1.5% agarose gel after amplification with the following cycling conditions; **95°C – 3' X1** | **95°C – 1'**, **64°C – 1'30"**, **72°C – 1'30"** **X30** | **72°C – 10' X1**

α-gene analysis: ARMS-PCR

Candidates were screened for specific mutations in either α-gene by allele specific primers (Table 2.3). The amplification of these primer sets to detect a mutation depended upon the careful design of these mismatched primers using the guidelines in Current Protocols in Molecular Biology (2000) and also on the buffer system used for the PCR conditions. Buffer A

(Table 2.15) was used with 5μM of each ARMS primer and its associated second primer (Table 2.3).

Table 2.3 ARMS primers for the detection of mutations in the α_1 or α_2 gene

Mutation	Oligonucleotide sequence (5'→3')	GenBank Co-ordinates Z84721	Second primer	Annealing temperature (°C)	Product size (bp)
Hb Evanston	GCCAGCGTGCGCGCCGACCTTACCCTT	37622-37648	A1	58	409
IVSI-117(G→A)	TCTTGGTGGTGGGGAAGGACAGGAACATAT	37791-37820	A1	58	581
Hb Sun Prairie	GTGCACGCCTCCCTGGACAAGTTCCTAC	34398-34425	A3	58	358
Hb Koya Dora	AGCACCGTGCTGACCTCCAAATACCGTCC	34434-34462	A3	58	323
Poly A Tail (AATAAA→AATA_)	GCACCGGCCCTTCCTGGTCTTTGAATTG	34529-34556	A3	58	229
HbQ-India	CACGTGCGCCACGGCGTTGGTCAGCGCGAG	37168-37198	A1	60	370
HbQ-India Normal	GGCCCCAAGGGGCAAGAAGCAT	38293-38313	A1	60	765
5'α Exon 1 (A4)	GTCAAGGCCGCCTGGGGTAAGGTC	α_1: 37160-37633 α_2: 33806-33829	A5	58	408
3'α Exon 2 (A5)	GTGCGCGTGCAGGTCGCTCAG	α_1: 37549-37568 α_2: 33745-33264			
5' α common (A1)	CTGGTCCCCACAGACTCAGA	α_1: 37549-37568 α_2: 33745-33264	A2		766
3'α_2URR (A3)	TTTATTCAAAGACCAGGAAG	34538-34557			

Each primer set included extra primer(s) to act as a control to demonstrate that the individual samples could be amplified. The mutations Hb Evanston and IVSI-117 (G→A) required a multiplex of 3 primers. A common primer in a-promoter (A1), a control primer (3'α Exon 1) (408 bp) and the mutation specific primer. The mutations Hb Sun Prairie, Hb Koya Dora, Poly A Tail (AATAAA→AATA_) and Hb Q-India required a multiplex of 4 primers. Two primers (5'α Exon 1 and 3'α Exon 1) were included to give a control band.

α-gene analysis: sequence analysis

Amplification of the α-genes was made after the design of oligonucleotides to allow independent amplification of either the α_1- or the α_2-gene. The sequence of these primers (A1, A2 and A3) as well as the annealing temperatures are shown in Table 2.4 and were published after

studying Thai patients with suspected Hb Constant Spring (Viprakasit *et al.*, 2002). The PCR amplification conditions were different from those stated in the paper. Products were separated on a 1% agarose gel after amplification with the following cycling conditions;

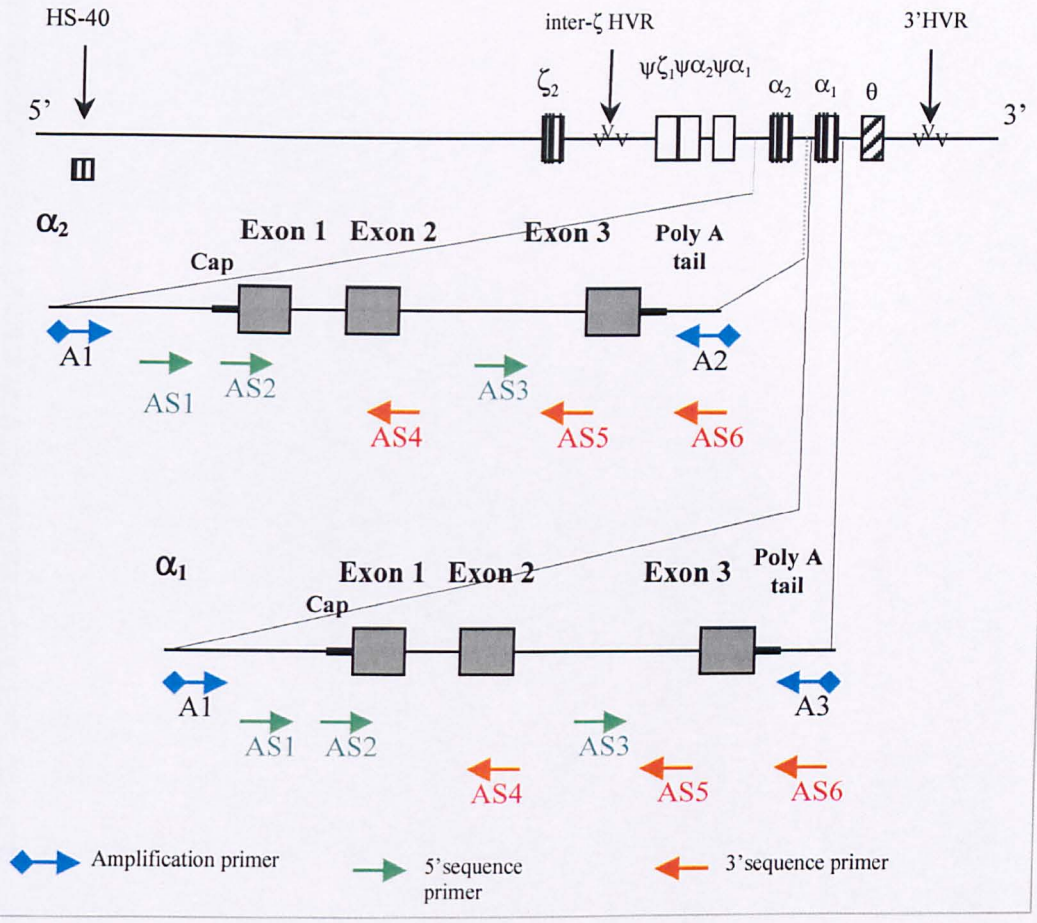
95°C – 3' X1	**95°C – 1'**	**72°C – 10' X1**
	X°C – 1'30" **X30**	
	72°C – 1'30"	

Table 2.4 Amplification and sequence primers for analysis of α_1 and α_2 genes

Primer	Oligonucleotide sequence (5'→3')	GenBank Co-ordinates Z84721	Second primer	Annealing temperature(°C)	Product size (bp)
Amplification					
5' α common (A1)	CTGGTCCCCACAGACTCAGA	α_1: 37549-37568 α_2: 33745-33264			
3'α_1URR (A2)	ACGGGGGTACGGGTGCAGGAAG	38334-38355	A1	57	1406
3'α_2URR (A3)	TTTATTCAAAGACCAGGAAG	34538-34557	A1	58	1327
Sequence					
161:C1 (AS1)	TGGAGGGTGGAGACGTCCTG	α_2: 33537-33556 α_1: 37341-37360			
278: 3'αD3	CCAGCTGCAGAGAGGTTCTAGCC	α_1: 38460-38482			
279: 5'α501	GAGAACCCACCATGGTGCTGT	α_2: 33765-33785 α_1: 37569-37589			
280:3'αC3B	ATTCCGGGACAGAGAGAACCC	α_2: 34588-34608			
281:3'αC10	ACGGTTGAGGGTGGCCTGT	α_2: 33921-33939 α_1: 37725-37743			
282:5'αC7	TCGACCTGAGCCACGGCTCTG	α_2: 34032-34052 α_1: 37836-37856			
285:3'αC5	GCGGACAGCGCGTTGGGCATG	α_2: 34120-34140 α_1: 37924-37944			
286:5'αC6	AGGCGGCGGCTGCGGGCCTGG	α_2: 34287-34307 α_1: 38091-38111			

Amplicons were cleaned up using exclusion chromatography and cycle sequencing outlined in 2.4 and 2.5. The only deviation was that 2μl 5M betaine was substituted into the reaction for an equal volume of water. The sequencing primers used was previously published (Dode *et al.*, 1993) and their locations in relationship to the α-genes shown in Figure 2.4.

Figure 2.4 **Schematic of the α-genes to show the location of the amplification and sequence primers**



α-gene analysis: haplotype analysis

The α-haplotype assays were used as published (Miles *et al.*, 2001). PCR conditions were changed to use Buffer A (Table 2.15). The primer sequences, annealing temperatures and fragment patterns are given in Table 2.5.

Table 2.5 Alpha haplotype oligonucleotides

Primer	Oligonucleotide sequence (5'→3')	GenBank Co-ordinates Z84721	Annealing temperature (°C)	Fragment sizes (bp)
Xba I F	CTCAACAGATGCCAGCCAAACAGAC	5108-5132	60	- 932
Xba I R	GAACCATTACACCAAGTGAAGGACG	6016-6040		+ 643/289
Sac I F	AAGATGGAGAAGGAAGGATGACTGC	18710-18734	60	- 560
Sac I R	GAGAGCAAGCCGATGTGGTAGAATG	19246-19270		+ 432/128
IZ *Bgl I* F	CCAGGACACAGATGGAGGCTAATGT	20503-20527	60	S- 876-768
IZ *Bgl I* R	TTGGGGTTGGGTGCATGAGGC	21985-22005		S+ 664-544/224 M- 2136-1812 M+ 1922-1588/224 L- 2748-2424 L+ 2588-2200/224
PZ/Z F	GGGTGGGCAGAAGGGGAGAC	23745-23764	65	PZ 1137
PZ/Z R	GATCCCAGATGCTGGGACTTTAGAGGCTGT	24853-24882		Z 758
Acc I F	AGGCATGGGCCGCCATTCCTGG	28964-28985	65	- 855
Acc I R	CTTGATCTGGGCTGAGCC	29802-29819		+ 519/336
Rsa I F	GCACCTCCCACCCTCCCCCTC	32731-32751	65	- 819
Rsa I R	GTCTCCACCCTCCACCCGCCACTC	33527-33550		+ 466/305/48
α *Pst I* F	GGTTGAATGCTCCAGCCGGTTCCAGCTATTGC	36626-36657	60	- 926
α *Pst I* R	CCAGAAGAGTGCCGGGCC	37535-37552		+ 581/345
θ *Pst I* F	ACGGTGTCTTCGTTATTTCATCAGG	39341-39365	60	- 884
θ *Pst I* R	TCAGTGCTTCATTTCTTTTTGTGGC	40201-40225		+ 569/315

β-globin gene analysis: ARMS-PCR

Candidates were screened for specific mutations in the β-gene by allele specific primers (Table 2.6) in a total reaction volume of 25μl. 1μl of each mutation specific primer and a second primer, all at a concentration of $1OD_{260}$/ml (~5μM), were added to Buffer D (Table 2.15) with 0.5μg genomic DNA. Control primers were added to each assay to assess for amplification failure. Control set C+D was used in all assays except for IVSII-745 (C→G) that used E+F. The PCR conditions and most of the primers had been previously published (Old, 1996).

If a patient was positive for a specific mutation, normal sequence was assessed by allele specific primers to test for homozygosity (Table 2.7). Similar amplification conditions were employed as for the mutation specific primers. All amplicons were separated on a 3.0% (Nusieve: Seakem GTG) agarose gel after 1 hour of electrophoresis at 100v and visualised on a transilluminator after staining with ethidium bromide.

Table 2.6 Primers for the detection of mutations by ARMS-PCR

Mutation	Oligonucleotide sequence (5'→3')	GenBank Co-ordinates U01317	Second primer	Annealing temperature (°C)	Product size (bp)
-28 (A→G)	AGGGAGGGCAGGAGCCAGGGCTGGGC*T*TAG	62080-62109	A	65	624
CAP+1 (A→C)	ATAAGTCAGGGCAGAGCCATCTATTGG*T*TC	62108-62137	A	65	597
CD5 (-CT)	TCAAACAGACACCATGGTGCACCTGA*G*TCG	62174-62204	A	65	528
CD8/9 (+G)	CCTTGCCCCACAGGGCAGTAACGGCA*C*ACC	62212-62241	B	65	225
CD15 (G→A)	TGAGGAGAAGTCTGCCGTTACTGCCC*A*GTA	62204-62233	A	65	500
CD15 (-T)	CTGAGGAGAAGTCTGCCGTTACTGCCCT*A*G	62230-62260	A	65	498
CD16 (-C)	TCACCACCAACTTCATCCACGTTCAC*G*TTC	62236-62266	B	65	238
CD30 (G→C)	TAAACCTGTCTTGTAACCTTGATACC*T*ACG	62278-62307	B	65	279
IVSI-1 (G→A)	TTAAACCTGTCTTGTAACCTTGATACC*G*AT	62279-62308	B	65	281
IVSI-1 (G→T)	TTAAACCTGTCTTGTAACCTTGATACC*G*AAA	62279-62308	B	65	281
IVSI-5 (G→C)	CTCCTTAAACCTGTCTTGTAACCTTG*T*TAG	62283-62312	B	65	285
IVSI-6 (T→C)	TCTCCTTAAACCTGTCTTGTAACCTT*C*ATG	62284-62313	B	65	286
IVSI-110 (G→A)	ACCAGCAGCCTAAGGGTGGGAAAATAGA*G*T	62387-62416	B	65	419
CD39 (C→T)	CAGATCCCCAAAGGACTCAAAGAACCT*G*TA	62433-62462	B	65	436
CD41/42 (-TCTT)	GAGTGGACAGATCCCCAAAGGACTCAACCT	62437-62470	B	65	439
IVSII-1 (G→A)	AAGAAAACATCAAGGGTCCCATAGACT*G*AT	62621-62650	B	65	634
IVSII-745 (C→G)	TCATATTGCTAATAGCAGCTACAATC*G*AGG	63347-63376	D	65	738
25bp deletion	CTCTGGGTCCAAGGGTAGACCACCAGCATA	62384-62438	B	65	445
619 bp deletion	Control primers D+E			65	242
β^S:CD6(G→T)	CCCACAGGGCAGTAACGGCAGACTTCTGCA	62236-62235	B	65	207
β^E:CD26(G→A)	TAACCTTGATACCAACCTGCCCAGGGC*G*TT	62265-62294	B	65	236
A: 3'Int-2	CCCCTTCCTATGACATGAACTTAA	62680-62703			
B: 5'URR	ACCTCACCCTGTGGAGCCAC	62028-62047			
C: 3'URR	GAGTCAAGGCTGAGAGATGCAGGA	64061-64084	C+D	65	861
D: 5'Int-2	CAATGTATCATGCCTCTTTGCACC	63224-63247			
E 5' $^G\gamma$-*Hind* III	AGTGCTGCAAGAAGAACAACTACC	35677-35700	E+F		326
F 3' $^G\gamma$-*Hind* III	CTCTGCATCATGGGCAGTGAGCTC	35981-36004			

Table 2.7 Primers for the detection of normal DNA sequence by ARMS-PCR

Mutation	Oligonucleotide sequence (5'→3')	GenBank Co-ordinates U01317	Second primer	Annealing temperature (°C)	Product size (bp)
CD8/9 (+G)	CCTTGCCCCACAGGGCAGTAACGGCA*C*ACT	62212-62241	B	65	225
CD15 (G→A)	TGAGGAGAAGTCTGCCGTTACTGCCC*A*GTA	62204-62233	A	65	500
CD41/42 (-TCTT)	GAGTGGACAGATCCCCAAAGGACTCAAAGA	62441-62470	B	65	439
IVSI-1 (G→A)	TTAAACCTGTCTTGTAACCTTGATACC*C*AC	62279-62308	B	65	281
IVSI-1 (G→T)	GATGAAGTTGGTGGTGAGGCCCTGGG*T*AGG	62279-62308	A	65	455
IVSI-5 (G→C)	CTCCTTAAACCTGTCTTGTAACCTTG*T*TAC	62283-62312	B	65	285
β^E:CD26 (G→A)	TAACCTTGATACCAACCTGCCCAGGGC*G*TC	62265-62294	B	65	236

NB: The second primer A and B refers to Table 2.6

β-gene analysis: RFLP analysis

PCR fragments were amplified across known polymorphisms, associated with a restriction enzyme site, and assessed by digestion with the relevant restriction enzyme. The fragments were amplified under the same PCR conditions as the ARMS-PCR for the β-gene mutations using the amplification primers shown in Table 2.8. The cycling conditions were also the same except 30 cycles of amplification were used and the annealing temperatures highlighted in Table 2.8 (Old, 1996).

Table 2.8 β-haplotype oligonucleotides

RFLP	Oligonucleotide sequence (5'→3') F: = Forward R: = Reverse	GenBank Co-ordinates U01317	Product size (bp)	Annealing temperature (°C)	Fragment sizes (bp)
Hind II /ε	F: TCTCTGTTTGATGACAAATTC	18652-18672	760	55	-/- 760 +/+ 445+315
	R: AGTCATTGGTCAAGGCTGACC	19391-19411			
Xmn I /Gγ	F: AACTGTTGCTTTATAGGATTTT	33862-33883	657	55	-/- 657 +/+ 455+202
	R: AGGAGCTTATTGATAACCTCAGAC	34495-34518			
Hind III /Gγ	F: AGTGCTGCAAGAAGAACAACTACC	35677-35700	326	65	-/- 326 +/+ 235+91
	R: CTCTGCATCATGGGCAGTGAGCTC	35981-36004			
Hind III /Aγ	F: ATGCTGCTAATGCTTCATTAC	40357-40377	635	65	-/- 635 +/+ 327+308
	R: TCATGTGTGATCTCTCAGCAG	40971-40991			
Hind II /5'ψβ	F: TCCTATCCATTACTGTTCCTTGAA	46686-46709	795	65	-/- 795 +/+ 691+104
	R: ATTGTCTTATTCTAGAGACGATTT	47457-47480			
Hind II /3'ψβ	F: GTACTCATACTTTAAGTCCTAACT	49559-49582	913	55	-/- 913 +/+ 479+434
	R: TAAGCAAGATTATTTCTGGTCTCT	50448-50471			
Ava II /β	F: GTGGTCTACCCTTGGACCCAGAGG	62416-62439	328	65	-/- 328 +/+ 228+100
	R: TTCGTCTGTTTCCCATTCTAAACT	62720-62743			
Hinf I /β	F: GGAGGTTAAAGTTTTGCTATGCTGTAT	63974-64001	474	55	-/- 320+154 +/+ 213+154+107
	GGGCCTATGATAGGGTAAT	64429-64447			
Dde I /β	F: ACCTCACCCTGTGGAGCCAC	62028-62047	443	65	HbS/S 760 HbA/A 445+315 IVSI-130 268+175
	R: GAGTGGACAGATCCCCAAAGGACTC AAGGA	62441-62471			
Bgl I /β					N 443 CD15(-T) 237+205
Pst I /β					N 443 CD16(-C) 237+205
Eco R I /β	F: CAATGTATCATGCCTCTTTGCACC	63224-63247	861	65	HbD/D 760 HbA/A 445+315
	R: GAGTCAAGGCTGAGAGATGCAG GA	64061-64085			
Mwo I /β	F: AGACTCTTGGGTTTCTGATAGG	62345-62366	189	65	N 760 CD55(-A) 445+315
	R: TAAAGGCACCGAGCACTTTCTT	62512-62534			

Additional RFLP assays were designed to confirm the presence of some mutations that were either new or that did not have an ARMS-PCR assay for mutation detection. The same PCR and cycling conditions were

used as in all these RFLP assays using the annealing temperatures outlined in Table 2.8. The expected fragment patterns are given in the same table.

β-gene analysis: sequence analysis

β-gene sequence analysis used previously described primers (Thein & Hinton, 1991). Amplification of the β-genes was made using the oligonucleotides 5' β2A and 3' β38 in Buffer E (Table 2.15) in a total volume of 100μl. The nucleotide sequence of these primers and the annealing temperatures are shown in Table 2.9. Amplicons were cleaned up using exclusion chromatography and cycle sequencing as outlined (See section 2.4 and 2.5).

Table 2.9 Amplification and sequence primers for analysis of the β gene

Primer	Oligonucleotide sequence (5'→3')	GenBank Co-ordinates U01317	Second primer	Primer concentration (μM)	Product size (bp)
Amplification					
5' β2A	CGATCTTCAATATGCTTAC	61831-61849	3'β38	58	2617
3'β38	GGGCCTATGATAGGG	64434-64448			
5' βURR	TTCCCCTCCTACCCCTACTTTCTA	61224-61247	3' βURR	63	901
3' βURR	GCTCTGCCCTGACTTTTATGC	62104-62124			
Sequence					
Ex1.1L	GAGCCAAGGACAGGTACGG	61996-62014			
3'β2	CAAAGGACTCAAAGAACCTC	62437-62456			
Ex2.1L	AGACTCTTGGGTTTCTGA	62346-62363			
Ex2.1R	TAAAGGCACCGAGCACTT	62517-62535	Ex2.1L	58	189
3'β9	CATTCGTCTGTTTCCCATTCTA	62725-62746			
JMO#15	CAATGTATCATGCCTCTTTGCACC	63225-63248			
3'β8	GCAGCCTCACCTTCTTTCATGG	63869-63890			
5' βURR-1	GTTTCCCAAAACCTAATA	61440-61457			
3' βURR-1	TGGATCTCTTCCTGCGTCTC	61718-61737			

γ-gene analysis: Southern blot

Southern blot analysis was performed in an identical manner to that described earlier in α-gene analysis. The differences were that the genomic DNA was digested with either *Bam HI* or *Hind III* and then subsequently hybridised with a previously published γ-probe (Jeffreys, 1979).

Table 2.10 Fragment sizes observed by Southern blot analysis with normal γ-gene arrangements

γ Allele	Restriction digest and hybridisation probe	
	BamH I - γ	*Hind III* - γ
$^{G}\gamma^{A}\gamma$	15.3/5/2.7	7.3 or 6.5+0.8 / 3.3 or 2.7+0.6
$-^{A}\gamma$	15.3/5	3.3 or 2.7+0.6
$^{G}\gamma-$	5/2.7	7.3 or 6.5+0.8

γ-Gene analysis: sequence analysis and RFLP

Amplification of the γ-genes was made after the design of oligonucleotides to allow independent amplification of the $^{G}\gamma$- or $^{A}\gamma$-gene. The sequences of these primers as well as the annealing temperatures are shown in Table 2.11 or have been published (Bayoumi *et al.*, 1999).

Table 2.11 Amplification and sequence primers for analysis of the γ genes

Primer	Oligonucleotide sequence (5'→3')	GenBank Co-ordinates U01317	Second primer	Primer concentration	Product size (bp)
Amplification					
5'$^{A}\gamma$	TTACTGCGCTGAAACTGTGGTCTT	38783-38806	3'$^{A}\gamma$	58	2359
3'$^{A}\gamma$	TTTACTGCTTTTATTTGCTCATCA	41119-41142			
5'$^{G}\gamma$	TGTTGCTTTATAGGATTTTTCACT	33865-33888	3'$^{G}\gamma$	54	2427
3'$^{G}\gamma$	AGATTTCTCCTGCACCTTTTACT	36270-36292			
Sequence					
5'γ-467	AAATCCTGGACCTATGCCTA	$^{A}\gamma$:38947-38966 $^{G}\gamma$:34011-34030			
5'γ-265	TGACTGAATCGGAACAAGG	$^{A}\gamma$:39149-39167 $^{G}\gamma$:34213-34231			
5'γ-Cap	AGCACCCTTCAGCAGTTCC	$^{A}\gamma$:39393-39411 $^{G}\gamma$:34457-34475			
5'γ-1	GGATTTGTGGCACCTTCTGACTG	$^{A}\gamma$:39632-39654 $^{G}\gamma$:34696-34718			
3'γ-IVSII	CTCCCAACCCCAGTATCT	$^{A}\gamma$:40002-40019 $^{G}\gamma$:35066-35084			
5'γ-IVSII	TGGGTTCATCTTTATTGTC	$^{A}\gamma$:40734-40752 $^{G}\gamma$:35818-35836			

HFE gene analysis

The primers and conditions for the analysis of HFE gene containing the H63D and C282Y mutations were described previously (Merryweather-Clarke *et al.*, 1997). The PCR conditions were optimised in buffer G (Table 2.15) using primers from Table 2.12. The amplicons for the mutation

C282Y was digested with *Rsa I* and H63D was digested with *Mbo I* as per manufacturer's instructions.

Table 2.12 RFLP primers to detect HFE gene mutations

RFLP	Oligonucleotide sequence (5'→3')	GenBank Co-ordinates Z92910	Primer concentration (μM)	Product size (bp)	Digestion products (bp)
C282Y F	CAAGTGCCTCCTTTGGTGAAGGTGACACAT	6490-6519	0.4μM	343	N:203/140
C282Y R	CTCAGGCACTCCTCTCAACC	6843-6862	0.4μM		M:203/111/29
H63D F	ACATGGTTAAGGCCTGTTGC	4621-4640	0.4μM	294	N:237/57
H63D R	CTTGCTGTGGTTGTGATTTTCC	4894-4915	0.4μM		M:138/99/57

G-6-PD gene analysis

The primers for the RFLP were already published (Kaeda *et al.*, 1995). The conditions for amplification used Buffer F (Table 2.15) in a final reaction volume of 25μl with 30 rounds of amplification. Products were separated on a 3% agarose gel after amplification with the following cycling conditions;

95°C – 3' X1 | **95°C – 1'** / **65°C – 1'30"** / **72°C – 1'30"** **X30** | **72°C – 10' X1**

The amplicons for the mutation G-6-PD Med was digested with *Mbo II*, 1311 T/C was digested with *Bcl I* and G-6-PD Orissa with *Hae III* as per manufacturer's instructions.

Table 2.13 RFLP primers to detect G-6-PD mutations

Mutation	Oligonucleotide sequence (5'→3')	GenBank Co-ordinates Z29527	Second primer	Primer concentration (μM)	Digestion products (bp)
5'Med	ACTCCCCGAAGAGGGGTTCAAGG	14596-14978	3'Med	65	N:381/119 M:381/119/100
3'Med	CCAGCCTCCCAGGAGAGAGGAAG	15478-15500			
5'1311T/C	TGTTCTTCAACCCCGAGGAGT	16880-16701	3'1311T/C	65	N:386 M:363/23
3'1311T/C	AAGACGTCCAGGATGAGGTGATC	17065-17087			
5'Orissa	CAGCCACTTCTAACCACACACCT	13288-13310	3'Orissa	65	N:107/75/66/48/45/10 M:123/107/66/45/10
3'Orissa	CCGAAGTTGGCCATGCTGGG	13620-13639			

UGT1A promoter analysis

The primers for the amplification and the nested primer used as a sequence primer (Table 2.14) were already published (Beutler *et al*, 1998). The conditions for amplification used Buffer E (Table 2.15) in 25µl.

Table 2.14 Amplification and sequence primers for analysis of the UGT1A promoter

Primer	Oligonucleotide sequence (5'→3')	GenBank Co-ordinates D87674	Second primer	Primer concentration (µM)	Product size (bp)
Amplification					
5' UGT1A	GCCAGTTCAACTGTTGTTGCC	2983-3003	3' UGT1A	4	314
3' UGT1A	CCACTGGGATCAACAGTATCT	3273-3297		4	
Sequence					
5' UGT1A	AGAAACCTAATAAAGCTCCACC	3009-3030		0.2	

Products were separated on a 1% agarose gel after amplification with the following cycling conditions;

95°C – 3 ' X1 | **95°C – 1 '**
60°C – 1 ' 30" X30
72°C – 1 ' 30" | **72°C – 10' X1**

2.5 Common methods

Reagents

Chemicals used were of analytical grade or higher and were purchased from more than one supplier that included Sigma-Aldrich Company Ltd, Merck Ltd, Life Technologies and Applied Biosciences. Oligonucleotides were manufactured by Oswel DNA Services. Hybridization probes were made by GenSet. Restriction enzymes were purchased from New England Biolabs. Radiochemicals were supplied by Amersham International and all Lightcycler kits were supplied by Roche Diagnostics.

Genomic DNA extraction

DNA was prepared from either fresh or frozen whole blood by a standard phenol chloroform method (Old & Higgs, 1983). Fresh whole blood was washed with at least two volumes of reticulocyte saline (0.13M NaCl, 5mM KCl, 7.4mM $MgCl_2$) and the red cells lysed with 25ml lysis solution (0.77M NH_4Cl, 0.046M $KHCO_3$). The 'buffy coat' was then re-suspended in 5ml lysing solution (100mM NaCl, 25mM EDTA). 0.5ml 10% SDS and 20µl 20µg/ml proteinase K was added and the mixture was incubated overnight at 37°C. Proteins and other impurities were removed by extracting twice with an equal volume of 0.1M Tris (pH 8) saturated phenol and then chloroform. DNA was precipitated out of solution by the addition of one tenth volume 7.5M ammonium acetate and 2-3 volumes of absolute ethanol. The DNA was washed in 70% ethanol and then re-suspended in double distilled water. Frozen blood was placed straight into the lysing solution with 2-3 times more 20µg/ml proteinase K with an extra phenol extraction step.

Polymerase chain reactions

All PCR reactions were optimised on an Eppendorf gradient thermocycler but most subsequent reactions were carried out on Biometra thermocyclers. Buffers for PCR were made up as per Table 2.15. AmpliTaq and AmpliTaq Gold were products of Applied Biosciences and Platinum Taq was a product of Life Technologies.

Table 2.15 PCR conditions

Reagent	A	B	C	D	E	F	G
Applied Bio. Buffer IV *	X1	X1	-	-	X1	X1	-
Applied Biosystems Buffer I **	-	-	X1	X1	-	-	X1
Dinucleotide triphosphates	0.2μM	0.2μM	0.2μM	0.2μM	0.2μM	0.2μM	0.2μM
$MgCl_2$	1.5mM	1.5mM	-	-	2mM	1.5mM	1.5mM
Betaine	0.75M	0.75M	0.75M	-		-	-
Dimethylsulphoxide	5%	10%	5%	-		-	-
Gelatin	-	-	-	100μg/ml		-	-
Spermidine	-	-	-	0.8 mM		-	-
Platinum Taq	0.5U	-	-	-	2U	-	-
AmpliTaq	-	0.5U	-	0.05U		0.5U	0.5U
AmpliTaq Gold	-	-	1.25	-		-	-

***Applied Biotechnologies**

Buffer IV (X10) Cat No AB-029

Tris-HCl	**750mM (pH 9.0 @ 25C°)**
$(NH_4)_2SO_4$	**200mM**
Tween	**0.1%(w/v)**

Buffer I (X10) Cat No AB-0242

Tris-HCl	**100mM (pH 8.3 @ 25C°)**
KCl	**500mM**
MgCl	**15mM**

Electrophoresis

All agarose based gel electrophoresis used either Seakem GTG or a mixture of Nusieve agarose and Seakem GTG. Products to be electrophoresed were combined with one-sixth volume of 6X Loading buffer (Xylene cyanol FF and Bromophenol blue in a 15% solution of Ficoll). Electrophoresis was performed at 100v for 1 hour, gels were stained with ethidium bromide, washed and then photographed using a Biorad Gel Documentation System.

Template preparation for sequencing

PCR reactions that amplified more than one band were separated by electrophoresis and visualised by staining with ethidium bromide and viewing on a long wavelength UV transilluminator. The amplicon of interest was isolated by excision out of the gel and cleaning up using a kit (Bio 101, Anachem (UK) Ltd) as per manufacturer's instructions.

PCR reactions that produced a single amplicon were cleaned up and concentrated using a Microcon-100 centrifugal filtration device supplied by Whatman (UK) Ltd as per manufacturer's instructions.

Sequencing

Cycle sequence reactions were performed using a big dye terminator kit purchased from Applied Biosystems, Warrington (UK) Ltd. As per the manufacturer's instructions unless stated. The extension products were separated from the unincorporated dyes by ethanol precipitation. These products were then resuspended in 12.5µl template suppression reagent and electrophoresed through an ABI 310 Genetic Analyser to separate the extension products as per the manufacturer's protocols (Applied Biosystems, Warrington (UK) Ltd).

Computer analysis

DNA analysis was performed on new sequence variations. Mutations were analyzed to compare the restriction enzyme pattern against the normal consensus sequence using a DNA computer program (DNA Star Inc., Wisconsin, USA).

Chapter 3 α-globin gene arrangement

The objective of the study was to ascertain the spectrum and frequency of the molecular lesions that cause α-thalassaemia and other α-gene arrangements in the population of Sri Lanka using Southern blot analysis. The subsequent phase was to develop simple robust PCR based methods to speed up the process and also use less starting material. The clinical significance of these genetic variations could be examined and their incidence compared against other world populations.

3.1 α-Gene deletions and insertions

In every population studied the most prevalent form of α-thalassaemia is caused by deletions removing 3.7kb DNA ($-\alpha^{3.7}$) that result in the expression of one α-gene only, as a wild type fusion protein (See section 1.2.2.2). Its counterpart is a single α-gene insertion that is found at much lower frequencies. Deletions that remove both α-genes from a single chromosome, although less common, have a greater clinical problem as couples that are both carriers for this molecular lesion would be at risk of producing a baby unable to produce adult α-globin chains that result in haemoglobin Bart's hydrops fetalis syndrome.

Hospital based patients

The α-gene cluster was investigated in a cohort of transfusion dependent patients attending five different clinics around the island; Anuradhapura, Badulla, Chilaw, Colombo and Kurunegala (See section 2.2). Some relatives from Anuradhapura were included in the analysis. The α-genotype was mainly determined by Southern blot analysis as described previously (See section 2.4). The α-gene allele frequency in the patients is shown in Table 3.1.

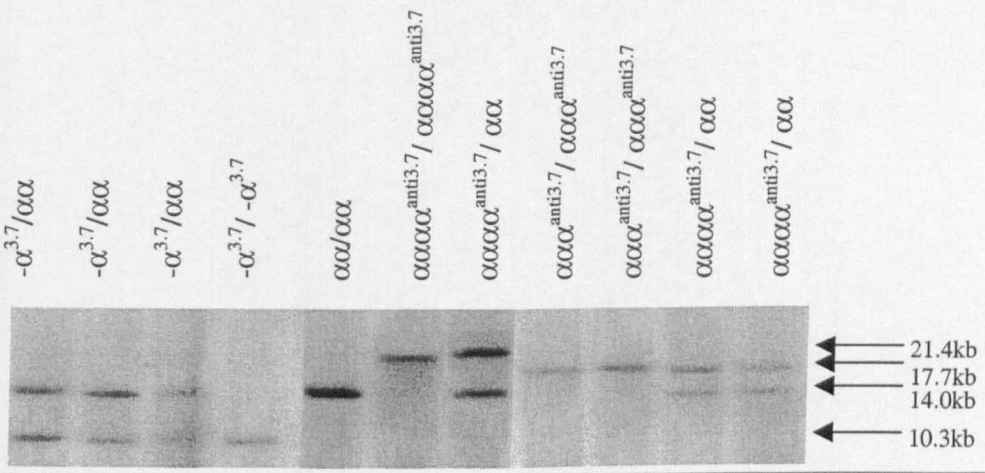
Table 3.1 α-globin gene frequency of patients and some parents attending five thalassaemia clinics in Sri Lanka

α-Gene alleles	Anuradhapura*		Badulla		Chilaw		Colombo		Kurunegala		Total	
	Allele No.	Frequency (%)	Allele No.	Frequency (%)	Allele No.	Frequency (%)	Allele No.	Frequency (%)	Allele No.	Frequency (%)	Allele No.	Frequency (%)
αα	275	87.6	119	93	14	87.5	48	92.3	363	91.2	819	90.2
$-\alpha^{3.7}$	26	8.3	7	5.5	2	12.5	4	7.7	20	5.0	59	6.5
$-\alpha^{4.2}$	8	2.5	0	0	0	0	0	0	2	0.5	10	1.1
ααα	5	1.6	2	1.5	0	0	0	0	11	2.8	18	2
αααα	0	0	0	0	0	0	0	0	2	0.5	2	0.2
Total	314	100	128	100	16	100	52	100	398	100	908	100

*Includes parental α-genotypes

Only five forms of α-gene arrangement were revealed. The majority of chromosomes appeared to be wild type (αα). Two single α-gene deletions ($-\alpha^{3.7}$ and $-\alpha^{4.2}$ at 6.5% and 1.1% respectively) and two extra α-gene arrangements (ααα and αααα at 2% and 0.2% respectively) were observed. These α-gene rearrangements could be discerned after digestion of genomic DNA with *Bam HI* and hybridisation with an α-probe (Figure 3.1). The $-\alpha^{3.7}$ and $-\alpha^{4.2}$ gene deletions were indicated by fragment sizes of 10.2kb and 9.8kb, respectively. Extra α-genes were indicated by larger fragment sizes of 18kb and 22kb for the triple α- and quadruple α-gene arrangement, respectively. Confirmation of these gene arrangements and distinction between $\alpha\alpha\alpha^{anti3.7}$ and $\alpha\alpha\alpha^{anti4.2}$ was determined after digestion with the restriction enzyme *Bgl II* and probing with a ζ probe and then subsequently with an α probe. The fragment sizes are shown in Table 2.1.

Figure 3.1 **Composite autoradiograph to highlight the fragment pattern associated with different α-gene arrangements after digestion with *Bam H I* and hybridisation with an α-probe**



The frequency of the α-gene arrangements was consistently high in all regions around the island. The $-\alpha^{3.7}$ deletion was found with an overall allele frequency of 6.5%, ranging between 5 and 12.5%. The highest

frequency was observed in Chilaw but coincided with the smallest study group and may have been a sampling anomaly. The $-\alpha^{4.2}$ gene deletion was much less common and only found in Anuradhapura and Kurunegala at an overall frequency of 1.1% (0.5% and 2.5% respectively). Again, this mutation may not have been observed in Badulla, Chilaw or Colombo because of the lower number of patients from these clinics and its lower incidence in the population. The overall observed allele frequencies with extra α-genes were high at 2% and 0.2% for ααα and αααα gene arrangement, respectively but were not observed in Chilaw or Colombo.

The α-genotype frequency in this cohort demonstrated a normal distribution for a recessive disorder (Table 3.2). Not all patients revealed a full genotype but still allowed one chromosome to be typed. Amongst these were some surprising observations.

Table 3.2 Observed genotype frequency of α-thalassaemia in Sri Lanka

Genotype	Individuals	Frequency (%)
αα/αα	356	81.7
$-\alpha^{3.7}$/αα	53	12.2
$-\alpha^{3.7}/-\alpha^{3.7}$	0	0
$-\alpha^{4.2}$/αα	7	1.6
$-\alpha^{4.2}/-\alpha^{4.2}$	1	0.2
$-\alpha^{3.7}/-\alpha^{4.2}$	1	0.2
ααα/αα	16	3.7
αααα/αα	0	0
ααα/ααα	1	0.2
αααα/αααα	1	0.2
Total	436	100

No individuals homozygous for the single α-gene deletion, $-\alpha^{3.7}$, were found. In contrast, from a smaller group of carriers for extra α-genes a single case homozygous for the ααα gene arrangement was found and the first case homozygous for the αααα arrangement was observed. Further investigation revealed that the parents of the patient homozygous for the αααα-gene arrangement were first cousins.

The overall allele frequencies (908 alleles from Table 3.1) were used to confirm the genotype frequency using the Hardy-Weinberg equation (See Appendix 1). The genotype frequencies were determined for the combined single α-gene deletions ($-\alpha^{3.7}$ and $-\alpha^{4.2}$) and the chromosomes with extra α-genes independently. These were then compared with the observed frequencies (436 genotypes from Table 3.2) in this cohort of patients (Table 3.3).

Table 3.3 Predicted genotype frequency of α-genotypes using the Hardy-Weinberg equation

Genotype		Expected	Observed
Single α-gene deletion	αα/αα	85.4	85.8
	-α/αα	14	13.8
	-α/-α	0.6	0.4
Extra α-genes	αα/αα	95.6	95.9
	ααα/αα or αααα/αα	4.3	3.7
	ααα/ααα or αααα/αααα	0.1	0.4

A χ^2 test to determine the fit between the observed and the expected α-genotype showed an exceedingly high level of significance with a probability of 0.995>p>0.975 @ 2df using Microsoft Excel statistics (See Appendix 1).

Gap PCR for the detection of α^+- and α^0-thalassaemia

A PCR based method was designed to speed up the analysis for the single α-gene deletions. Even though α^0-thalassaemia was not observed in this cohort a method to detect these deletions was prudent as they are found at high frequencies in SE Asia and have greater clinical ramifications. Mr Y.T. Liu was responsible for the design of the oligonucleotides used in these assays. The conditions were tested using an ammonium sulphate or a potassium chloride buffer system with different DNA polymerase enzymes

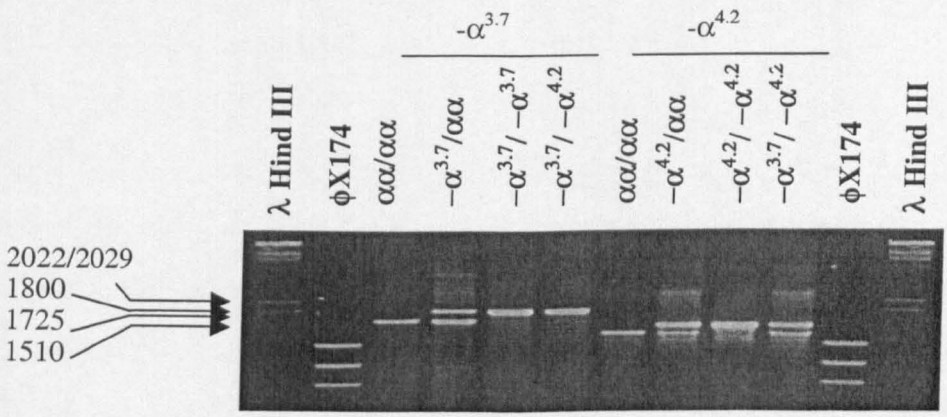
and PCR additives. The annealing temperature was optimised on a gradient thermocycler.

α^+-thalassaemia

A Gap-PCR method to screen for the two common α^+-thalassaemia gene deletions ($-\alpha^{3.7}$ and the $-\alpha^{4.2}$) was developed (See 2.4). These mutations are deletions between homologous G-C rich regions around the α-genes. The primers were deliberately kept long (30mers) to allow high annealing temperatures and located in the non-homologous regions flanking the breakpoints.

The PCR assay was performed on patients that had been previously characterised by Southern blot analysis. All α-genotypes were identical and included in the publication (Liu *et al.*, 2000). Further analysis of over 50 more patients revealed a disturbingly high error rate (8%). All errors were due to normal allele dropout with a carrier for the $-\alpha^{4.2}$ gene deletion and three heterozygous for the $-\alpha^{3.7}$ gene deletion typed as homozygous for this deletion. Further optimisation helped but failed to consistently resolve this problem. Alternative oligonucleotides (Chong *et al.*, 2000a) were substituted into the assay using the same buffer conditions and the errors were eliminated, as shown in Figure 3.2.

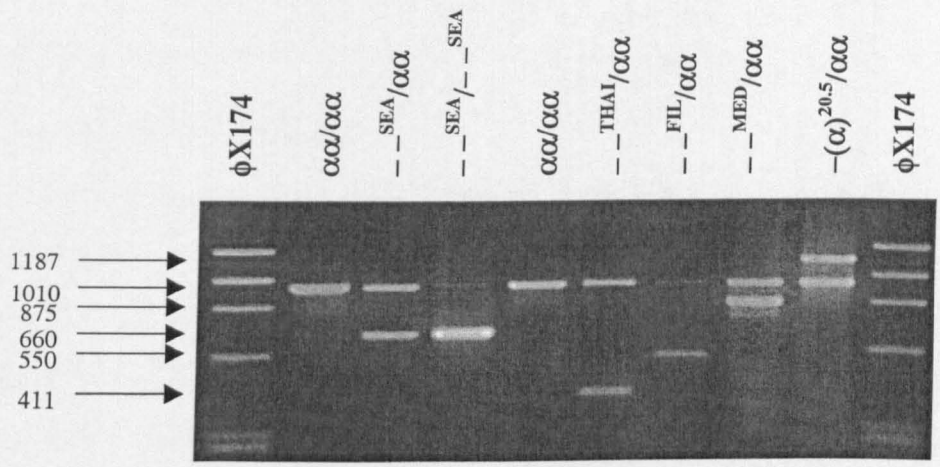
Figure 3.2 **α^+-thalassaemia PCR to detect the $-\alpha^{3.7}$ and the $-\alpha^{4.2}$ deletion**



α^0-thalassaemia

A multiplex Gap-PCR designed to detect the common α^0-deletions included those found in SE Asia ($--^{SEA}$, $--^{FIL}$ and $--^{THAI}$) plus two found around the Mediterranean ($--^{Med}$ and $-(\alpha)^{20.5}$). The PCR conditions used to amplify products across the breakpoints were similar to those established for the detection of α^+-deletions. The design of the primers to detect these large deletions is easier as many of the highly replicated areas are located within the deletion (See section 2.4). Typical results using previously characterised controls are shown in Figure 3.3. Out of 1046 alleles tested no cases of α^0-thalassaemia were identified.

Figure 3.3 **α^0-Thalassaemia PCR multiplex to detect the $--^{SEA}$, $--^{THAI}$, $--^{FIL}$, $--^{MED}$ and $-(\alpha)^{20.5}$ deletions**



Hb Bart's and the incidence of α-thalassaemia

Further investigations of the α-genes were conducted on sequential neonates born at the delivery units in Kurunegala and Chilaw to check against selection bias. These cord blood samples became available late in this study so protocols were established on cord blood from the UK. Prior to DNA analysis, population screens to detect α-thalassaemia were performed by ascertaining the proportion of neonatal samples with elevated

levels of Hb Bart's (See section 1.2.1). Hb Bart's becomes elevated with a greater chain imbalance where $\beta > \alpha$ chains.

Detection of Hb Bart's

Estimations of Hb Bart's levels were assessed by comparison of three protein separation techniques; IEF, CAE and CE-HPLC (See section 2.3). IEF and CAE were labour intensive so only 20 and 10 samples were tested by each method, respectively. They needed to be performed in small batches but gave a clear resolution of the Hb Bart's protein. A crude visual grading of relative proportions of haemoglobins was possible at high levels but remained inaccurate at low levels. The third method, CE-HPLC, employed the use of different cation exchangers; the α-thalassaemia short program, the β-thalassaemia short program and the sickle cell short program. This was a semi-automated method that allowed evaluation of 96 samples in a single run that gave evaluations at both high and low levels.

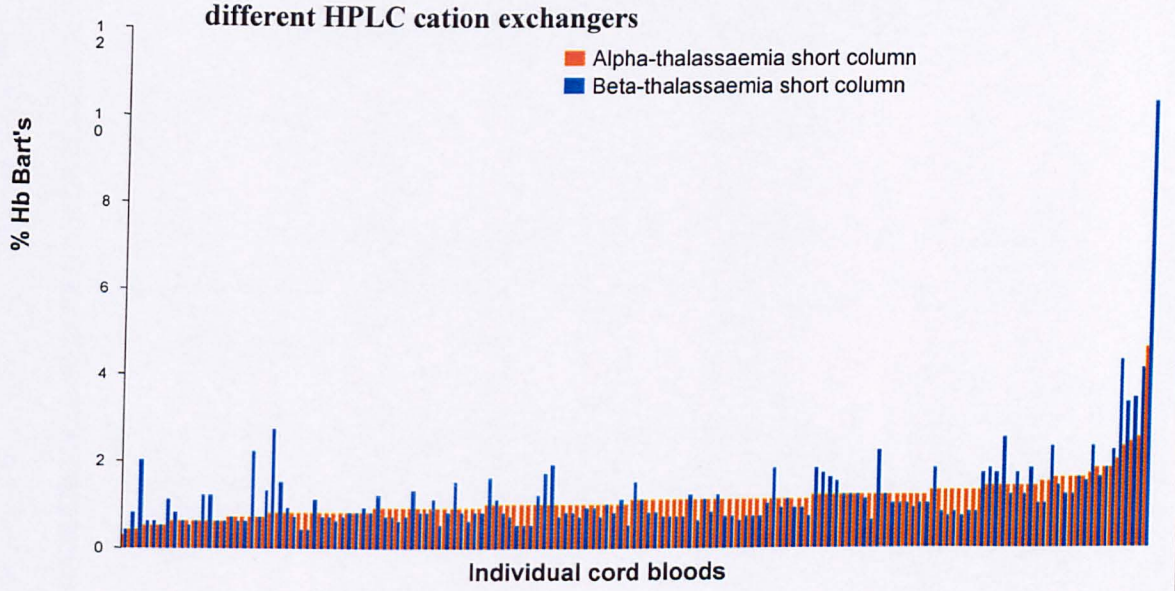
Comparison of HPLC columns

The β-thalassaemia short program is similar to the sickle cell screening program but elutes the proteins off the cation exchanger over half the time period. The α-thalassaemia short program is a column specifically designed to measure Hb Bart's or HbH that elute early with all other haemoglobins eluting later at the same retention time. Hb Bart's levels were evaluated on 168 cord blood samples using the α-thalassaemia short program and the β-thalassaemia short program (Figure 3.4).

The majority of the samples (57.7%) demonstrated an almost identical level of Hb Bart's. Most of the rest (overall 38.7%) showed a similar expression pattern with 20.8% and 17.9% having a modest change either greater or lower on the β-thalassaemia compared with the α-thalassaemia column, respectively. The only large discrepancy was seen in

3.6% of samples that showed a much greater (~2 times) level on the β-thalassaemia column compared to the α-thalassaemia column. These samples were re-assessed visually using IEF and CAE that suggested the α-thalassaemia short column gave a more accurate determination of the level of Hb Bart's.

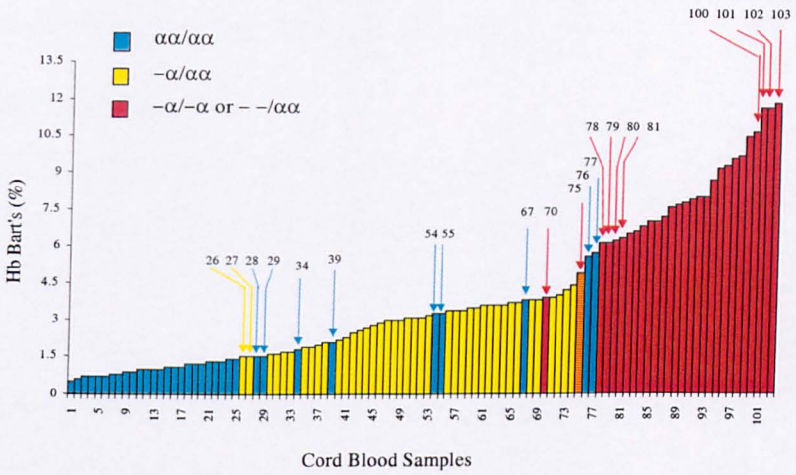
Figure 3.4 **Chart to show the Hb Bart's levels in cord blood collected in Sri Lanka and compared in pair wise fashion as measured on two different HPLC cation exchangers**



Hackney study

Over 150 cord blood samples were collected from the Homerton Hospital (London) and the Hb Bart's levels measured shortly after delivery using a sickle cell thalassaemia neonatal screen program on the Biorad Variant Analyser. This was performed and assessed by Dr R. Amos such that enough samples with a wide range of Hb Bart's could be collected and transported to Oxford for DNA analysis. Southern blot analysis was performed and the α-genotype successfully assessed on 103 of these samples, by Miss M. Rugless. The results were plotted in increasing ascendancy of Hb Bart's levels (Figure 3.5).

Figure 3.5 **Graph to show the correlation between the level of Hb Bart's and the α-genotype in cord blood samples from the Homerton Hospital**



Hb Bart's was observed in all cord blood samples and a clear tri-modal distribution was observed over this range, 0.5-11.9% and was summarised in Table 3.4.

Table 3.4 **Correlation of α-gene status with Hb Bart's levels**

Hb Bart's level (%)	n	α-Gene number	α-Genotype
0.5-1.4	25	4	αα/αα
1.5-5.7	9	4	αα/αα
	41	3	−α/αα
	1	3	$-\alpha^{New}/\alpha\alpha$
	1	2	−α/−α
6.1-11.9	26	2	−α/−α or −−/αα

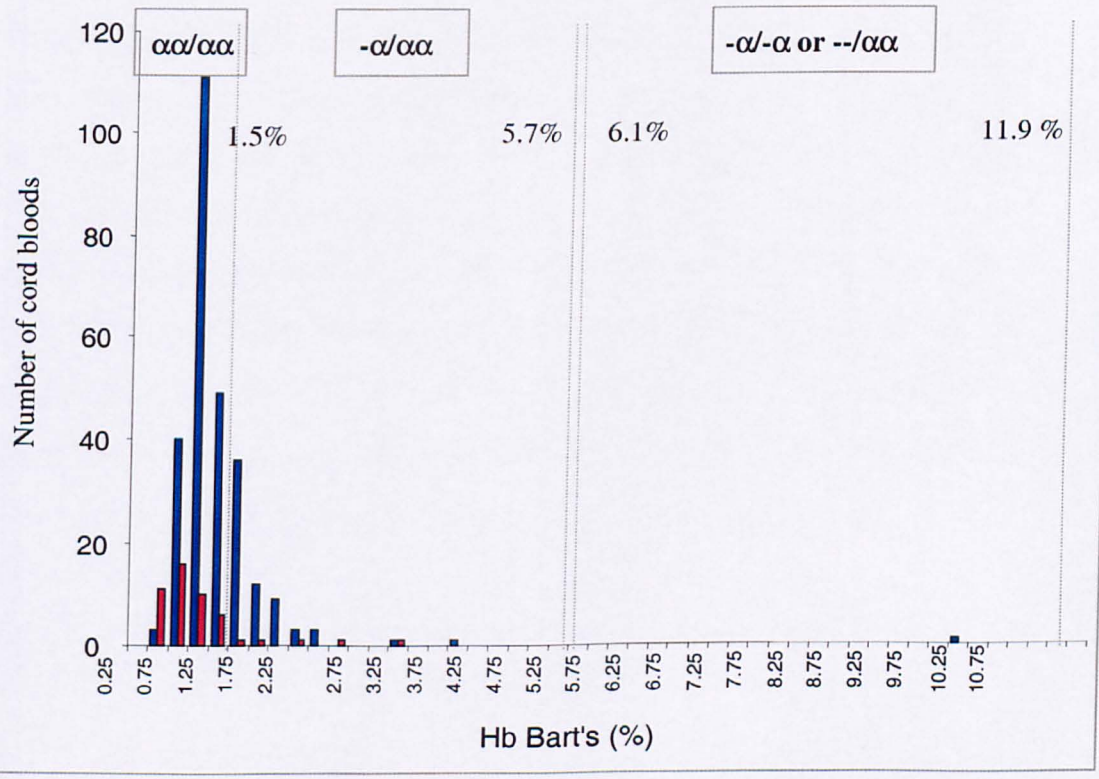
The increasing level of Hb Bart's appeared to correlate with an increasing level of α-gene deletion. An Hb Bart's level less than 1.4% invariably resulted in a normal α-genotype. Between 1.5 and 5.7% most samples were linked with a single α-gene deletion. All cord blood samples with Hb Bart's above 6.1 and up to 11.9% had two missing α-genes either as --/αα or -α/-α. A notable exception was observed in the middle group where ~20% (11/52) of these samples did not appear to have a single α-

gene deletion. Nine appeared to have a normal α-genotype (αα/αα), one was homozygous for a single α-gene deletion and one case gave new band patterns that needed further characterisation as summarised in Table 3.4.

Sri Lankan cord blood study

Cord blood samples were collected from two hospitals in Sri Lanka (Kurunegala Teaching Hospital and the Base Hospital in Chilaw), frozen and transported to the UK under the direction of Miss S. Rajapakshe. In Oxford, the level of Hb Bart's was estimated by cation exchange CE-HPLC using the β–thalassaemia short program. The α-thalassaemia column was deemed too expensive and was only used on a part of the Sri Lankan cord blood cohort. The frequency distribution of Hb Bart's was plotted (Fig 3.6) and an estimate for the incidence of α-thalassaemia made using cut off points determined from the Hackney study.

Figure 3.6 Distribution of Hb Bart's in Kurunegala (■) and Chilaw (■) measured by HPLC using the β-thalassaemia short program



The two study centres showed a divergence in the predicted α-genotype frequencies compared with the transfusion dependent patients earlier. Amazingly, the overall genotype frequency of α-thalassaemia predicted from the level of Hb Bart's was almost identical to that observed in the transfusion dependent patients. Single α-gene deletions and two α-gene deletions were predicted in 13.6% and 0.3% of samples (Table 3.5). The previous assessment on the incidence of α-thalassaemia in Kurunegala was similar but was much lower in Chilaw, more consistent with that of nearby Kurunegala.

Table 3.5 Predicted incidence of α-thalassaemia in Sri Lanka from Hb Bart's levels found in cord blood

Level of Hb Bart's (%)	Kurunegala		Chilaw		Combined	
	Number	Frequency	Number	Frequency	Number	Frequency
0.5-1.4	229	85.1	44	91.7	273	86.1
1.5-5.7	39	14.5	4	8.3	43	13.6
6.1-11.9	1	0.4	0	0	1	0.3
Total	269	100	48	100	317	100

DNA from 75 cord blood samples were extracted and the α-genotype analysed by Gap-PCR or Southern blot analysis (Table 3.6). A much less clearly defined genotype distribution pattern was observed over the observed range of Hb Bart's levels (0.5-10.3%) compared to the Hackney study. Using the established cut off points almost one third (22/75) of the samples did not conform to a predicted phenotype/genotype prediction. Four samples had low Hb Bart's level and were expected to correlate with a normal α-genotype but revealed a single α-gene deletion ($-\alpha^{3.7}$), five had a borderline level of Hb Bart's with a normal α-genotype and the remainder had elevated levels of Hb Bart's with a normal α-genotype.

Table 3.6 Correlation between the Hb Bart's level in cord blood and the α-gene arrangement in samples collected in Sri Lanka

Hb Bart's level (%)	n	Patient α-gene number	α-Genotype
0.5-1.4	42	4	αα/αα
	4	3	−α/αα
1.5-5.7	18	4	αα/αα
	10	3	−α/αα
6.1-11.9	1	2	−α/−α

Further analysis of these samples assessed by CE-HPLC using the α-thalassaemia short program showed that all four samples with a lower level of Hb Bart's than expected when in association with a single α-gene deletion were low using either column. Half of the samples (9/18) with a normal α-genotype and an elevated Hb Bart's level now demonstrated a lower level of Hb Bart's more consistent with their revealed α-genotype. The other half still appeared to be raised. This unreliability was probably due to a logistical problem. The processing of the samples (freeze-storage) prior to analysis may form protein agglutinates and alter the true expression pattern.

The spectrum of mutations observed was similar to those determined earlier in the transfusion dependent cohort. Eleven cases were found to be carriers for the $-\alpha^{3.7}$ allele, two cases for the $-\alpha^{4.2}$ allele and two cases homozygous for the $-\alpha^{3.7}$ allele were found.

3.2. Non-Deletion α-thalassaemia

Non-deletion α-thalassaemia has been found in all nearly all regions of the world and in combination with α^0-thalassaemia results in HbH disease (See section 1.2.2.2). However, its incidence has been hard to establish due to the difficult technical problems associated with reliable detection within the α-globin cluster. This form of α-thalassaemia has been reported in India and together with the preponderance of samples with an

unaccountably high level of Hb Bart's using the α-thalassaemia short program an analysis of the α-genes was considered necessary. The goal was to develop a more robust strategy of analysis for mutations in and around the α-genes on strong candidates referred to the NHRL. Candidates for this form of α-thalassaemia from many populations, especially India, were included so that any mutations found in the Sri Lankan population could then be compared and an indication of the origin of these mutations could be postulated.

Patients for α-gene analysis

Patients referred to the NHRL (See section 2.2) that were microcytic and hypochromic that was not due to iron deficiency or deletion forms of α-thalassaemia became candidates for non-deletion α-thalassaemia. In Sri Lanka studies were initiated to investigate transfusion dependent patients so good candidates for non-deletion α-thalassaemia were unavailable. Therefore, a random screen for non-deletion α-thalassaemia was performed on transfusion dependent patients and their relatives attending a clinic at Anuradhapura. However, the cord blood studies, in Hackney and Sri Lanka, with unexplained elevated Hb Bart's levels provided better candidates for further investigation.

Indian α-ARMS and RFLP Screen

Assays were developed for six mutations [Hb Evanston: α_1 codon14 (TGG→CGG); α_1 IVSI-117 (G→A); Hb Sun Prairie: α_2 codon 130 (GCT→CCT); Hb Koya Dora: α_2 termination codon (TAA→TCA); Hb Constant Spring: α_2 termination codon (TAA →CAA) and α_2 polyadenylation signal (AATAAA→AATA__)] reported in patients originating from the Indian subcontinent. All ARMS primers were designed with two mismatched nucleotides, one strong and the other weak,

within the last four nucleotides (3') of the primer. Hb Constant Spring detected by RFLP analysis as the mutation removes a restriction enzyme site (*Mse I*) but could be confused with the other mutations that remove the restriction enzyme motif (Viprakasit *et al.*, 2002). The primers and conditions are shown in section 2.4 and a typical positive result highlighted in Figure 3.7. These assays were performed on all the α^T candidates and the results are shown in Table 3.7.

Figure 3.7 ARMS assay for the detection of Hb Sun Prairie

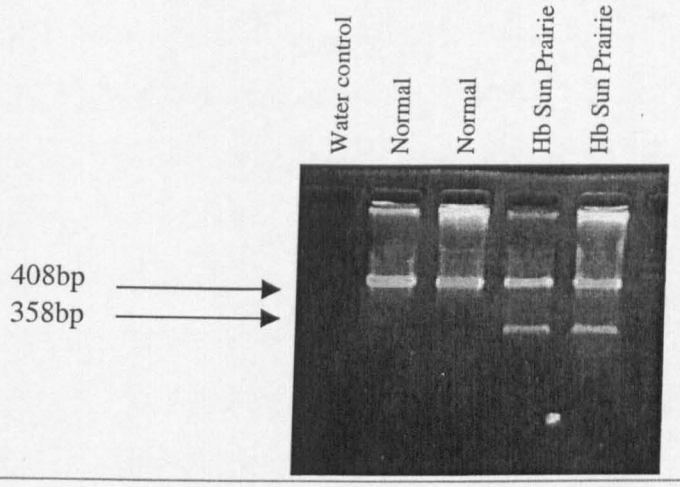


Table 3.7 Non-deletion α-thalassaemia determined by ARMS/RFLP analysis for the six known mutations described in Indian patients

Origin	Number of Individuals screened	α-gene status	Number of Individuals
Sri Lanka	190	Wild type	190
Indian Subcontinent	39	Poly A tail (AATAAA →AATA _ _)	5
		Hb Sun Prairie	1
		Wild type	33
Europe	40	Wild type	40
Africa	9	Poly A tail (AATAAA →AATA _ _)	1
		Wild type	8
Middle East	25	Poly A tail (AATAAA →AATA _ _)	2
		Wild type	23
SE Asia	12	Hb Constant Spring	4
		Wild type	8

In a limited screen of patients and their relatives from Anuradhapura all were found to be negative for these mutations. In the group of patients referred to the NHRL 13/125 were positive for the Indian α-ARMS and RFLP Screen found with three of these mutations (poly A tail (AATAAA →AATA_ _), Hb Sun Prairie and Hb Constant Spring). Unsurprisingly, six of the patients originated from the Indian subcontinent but the remainder originated from Africa, the Middle East and SE Asia.

Sequence analysis strategy

Amplification oligonucleotides were designed and buffer conditions optimised for the investigation of both the α_1- and α_2-gene independently (See section 2.4). These buffers, like those used in Gap-PCR and ARMS of the α-genes, were reliant on the inclusion of the organic denaturant betaine. Its subsequent inclusion in the cycle sequence reaction also changed a poor sequence reaction into a fair to good sequence reaction for these G/C rich templates. A full sequence analysis stretching from Cap-150 to 150 nucleotides beyond the poly A tail was obtained on both the α_1- and α_2-gene using a modified big dye terminator strategy.

Hackney cord blood α-gene sequence

The methodology was tested on candidates from the Hackney cord blood study as these had been assessed and needed further characterisation to identify reasons for elevated Hb Bart's in these samples. Seven cord blood samples with elevated Hb Bart's levels (1.8-5.7%) that appeared to have a normal α-gene status (αα/αα) and four samples with Hb Bart's levels above 10% samples were investigated by sequence analysis of both the α_1- and α_2-globin genes (Table 3.8).

Table 3.8 **α-gene sequence analysis of Hackney cord bloods with possible α^T**

Cord blood No	Hb Bart's level (%)	α-genotype		
		Southern blot analysis	Sequence analysis	Sequence variation
34	1.8	αα/αα	αα/αα	Normal consensus
39	2.1	αα/αα	$\alpha^T\alpha^T$/αα or $\alpha\alpha^T/\alpha^T\alpha$	α_1: IVSI-39(C→T) Heterozygous α_2:IVSII-115(-7bp) Heterozygous
54	3.3	αα/αα	αα/αα	Normal consensus
55	3.3	αα/αα	$\alpha\alpha^T$/αα	α_1:Hb Evanston Heterozygous
67	3.8	αα/αα	αα/αα	Normal consensus
76	5.6	αα/αα	αα/αα	Normal consensus
77	5.7	αα/αα	$\alpha^T\alpha$/αα	α_2: Hb Sallanches Heterozygous
100	10.7	$-\alpha^{3.7}/-\alpha^{3.7}$	$-\alpha^{3.7}/-\alpha^{3.7}$	Normal consensus
101	11.7	$-\alpha^{3.7}/-\alpha^{3.7}$	$-\alpha^{3.7(T)}/-\alpha^{3.7}$	α_2:IVSII-55(G→T) Heterozygous
102	11.7	$-\alpha^{3.7}/-\alpha^{3.7}$	$-\alpha^{3.7}/-\alpha^{3.7}$	Normal consensus
103	11.9	$-\alpha^{3.7}/-\alpha^{3.7}$	$-\alpha^{3.7}/-\alpha^{3.7}$	Normal consensus

Seven of these samples (34, 54, 67, 76, 100, 102 and 103) demonstrated a normal consensus sequence. Further investigation to examine red cell indices in these families would confirm their phenotypic status but require ethical permission to proceed. Four samples showed sequence variations that were confirmed by reverse strand sequence analysis and where possible by RFLP.

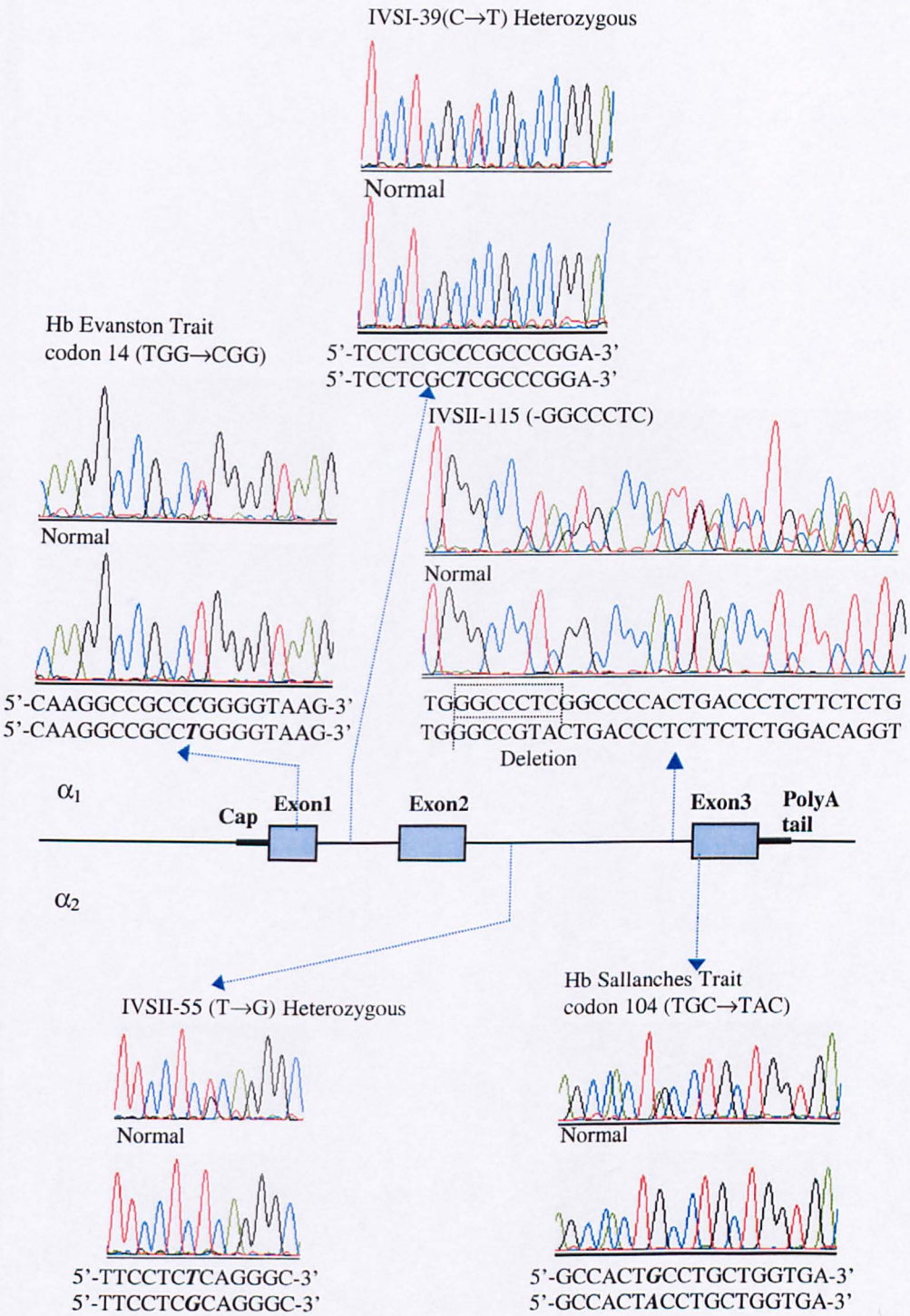
Two cord blood samples were found to be heterozygous for previously described point mutations in the α-genes. The first was a cord blood sample (77) with an Hb Bart's level of 5.7%. Sequence analysis identified a single nucleotide substitution at codon 104 (TGC→TAC) of the α_2-gene (Figure 3.8) that encodes for the amino acid change cysteine to tyrosine. The mutation, Hb Sallanches, was previously described in a French individual (Morle *et al.*, 1995).

The second mutation was found in a cord blood (55) with a small elevation in Hb Bart's level to 3.3%. A single nucleotide substitution at codon 14 (TGG→CGG) in the α_1 gene (Figure 3.8) encodes for a change in amino acids from tryptophan to arginine. This amino acid substitution, Hb

Evanston codon 14 (TGG→AGG), was previously described in a patient of African origin (Honig *et al.*, 1984). This patient was of Indian origin with the same amino acid substitution but with a different base change at codon 14 (TGG→CGG). A similar case was described in a meeting abstract (Huisman *et al.*, 1997).

Two cord blood samples were found to be heterozygous for new mutations in the α-genes. The first was found in the cord blood sample (39) associated with an Hb Bart's level of 2.1%. The mutation, a heterozygous base substitution in the α_1 gene, changed a cytosine to a thymidine at position 39 of the first intron (Figure 3.8). This putative new splice variant does not create a motif similar to any globin acceptor or donor site and requires further examination by RNA analysis. In the same individual another heterozygous change was found in the second intron of the α_2-gene. A small deletion that removed seven nucleotides in the second intron (IVSII-115 (-GGCCCTC)) was identified (Figure 3.8). This deletion changes the sequence so that it is the same as the α_1 intron 2 consensus sequence and probably constitutes a polymorphism. In the last case, another $\alpha_1 \rightarrow \alpha_2$ base substitution was found at nucleotide 55 (G→T) in the second intron in the heterozygous condition (Figure 3.8) in a sample with two α-genes deleted. The level of Hb Bart's in this cord blood was 11.7% and is consistent with a silent polymorphism.

Figure 3.8 **Mutations in and around the α_1 and α_2 genes found in cord bloods with elevated Hb Bart's levels from the Hackney study**



α-gene analysis of patients from the subcontinent of India

Thirty-nine patients were investigated by complete sequence analysis of the α_1 and α_2 genes. Mutations were found in 22 cases but the remaining 17 cases presented with only consensus normal sequence (Table 3.9).

Table 3.9 **Sequence variation in individuals originating from the Indian Subcontinent**

Mutation/Polymorphism	Gene	Cases
Codon42 (TTC→TTG); Hb Hirosaki	α_1	1
Codon 38/39 (-ACC); Hb Taybe	α_1	1
Codon 38/39 (-ACC); Hb Taybe & Codon 113 (CTC→CTG)	α_1	1
CD130 (GCT→CCT); Hb Sun Prairie	α_2	1
Cap+14 (C→G)	α_2	1
IVSII-55 (T→G)	α_2	4
IVSII-55 (T→G)+ IVSII-115 (+7bp)	α_2	1
IVSII-115 (-7bp)	α_1	1
Poly A Tail (AATAAA →AATA _ _)	α_2	5
Poly A Tail +7 (G→T) CD124 (TCC→TTC)	α_1 α_2	1
Cap-4 (C→G) +IVSII-115+7bp	α_2	1
IVSI-39 (C→G)	α_2	1
Cap+14 (C→G)+IVSII-70 (T→G)	α_2	1
Codon59 (GGC→CGC)	α_2	1
IVSII-51 (C→G)	α_2	1
No mutations observed	Normal	17
Total		39

Nine patients were found to have four previously described mutations (Figure 3.9). Two cases of Hb Taybe were found as reported in subjects from the Middle East (Pobedimskaya *et al.*, 1994). The third was an α-chain variant, Hb Hirosaki, described in a Japanese family (Huisman & Carver, 1998).

The fourth was heterozygous for Hb Sun Prairie and five patients were found to be heterozygous for the polyadenylation signal mutation (α_2:AATAAA →AATA _ _). Both mutations were previously reported in

Indian patients (Harkness *et al.*, 1990; Harteveld *et al.*, 1994) that validated the ARMS-PCR method.

A further seven patients had mutations that have been previously described as neutral polymorphisms. A mutation was found in the cap sequence (Cap+14 (C→G)) in a single case that had been previously observed (Harteveld *et al.*, 1996). A point substitution described earlier (IVSII-55 (T→G)) was found in 4 individuals and one individual was found with the same 7bp deletion in the second intron (IVSII-115 (-GGCCCTC)). The final case was doubly heterozygous with two mutations in the second intron, IVSII-55 (T→G) and IVSII-115 (+GGCCCTC).

Six individuals were found to carry mutations that have not been described before as shown in Figure 3.10. The first individual had a single heterozygous base change in the α_1-gene (Poly A Tail +7 (G→T)) found in conjunction with a single heterozygous base change in the α_2-gene (CD124 (TCC→TTC). In the thalassaemias no base substitutions have been described after the polyadenylation signal only deletions that extend beyond it (Weatherall & Clegg, 2001). The base change at codon 124 would change a serine to a phenylalanine that may disrupt the α_1:β_2 contact point. This individual demonstrated a very mild to borderline set of red cell indices (Hb 11.8g/dl; MCV 83.6fl; MCH 25.7pg and HbA_2 2.7%).

The second individual originated from Pakistan and was found to have two mutations in the α_2-gene, cap-4 (C→G) and IVSII-115 (+7bp) but presented with normal red cell indices suggesting the mutation(s) appeared to be clinically mild or of little significance.

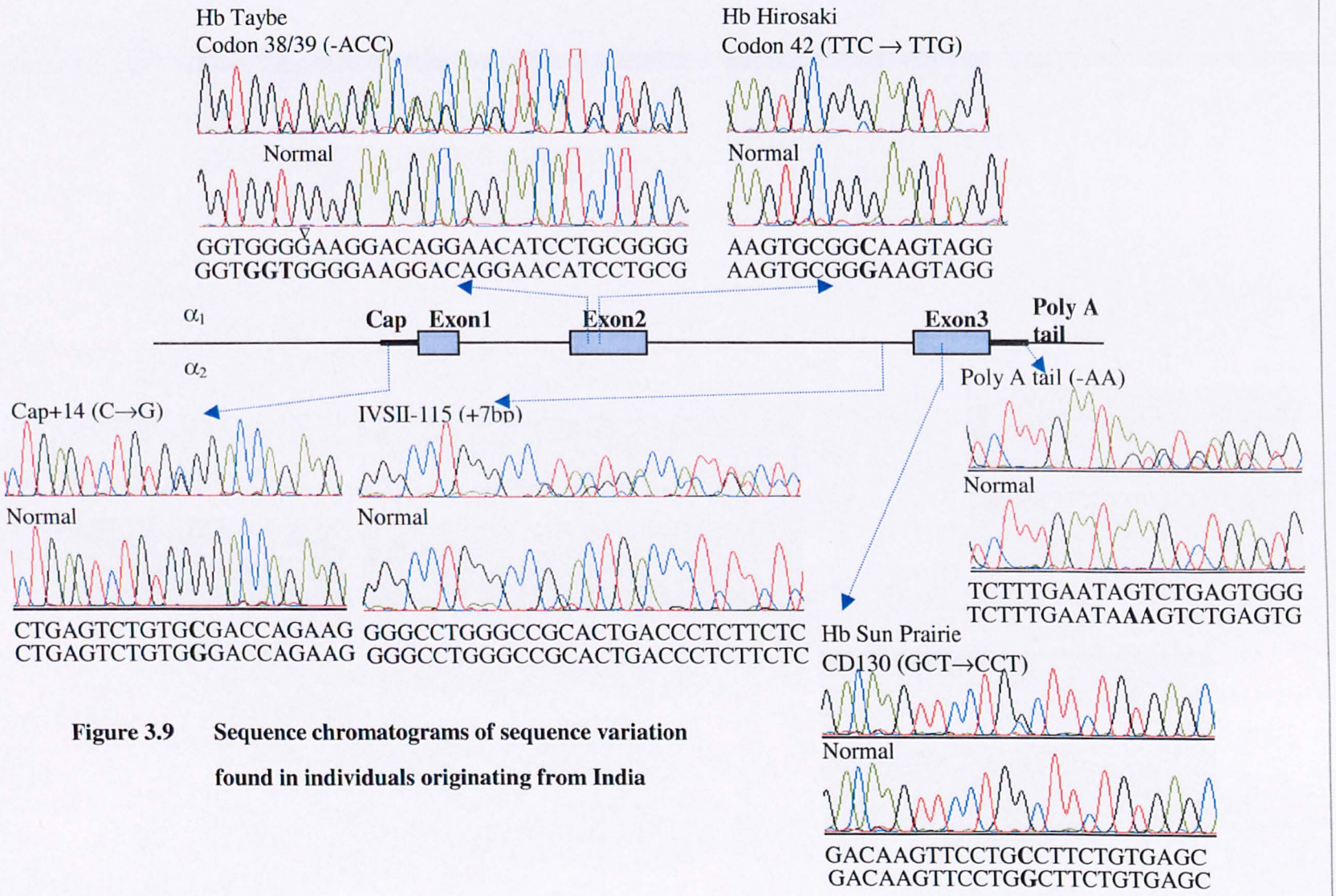


Figure 3.9 **Sequence chromatograms of sequence variation found in individuals originating from India**

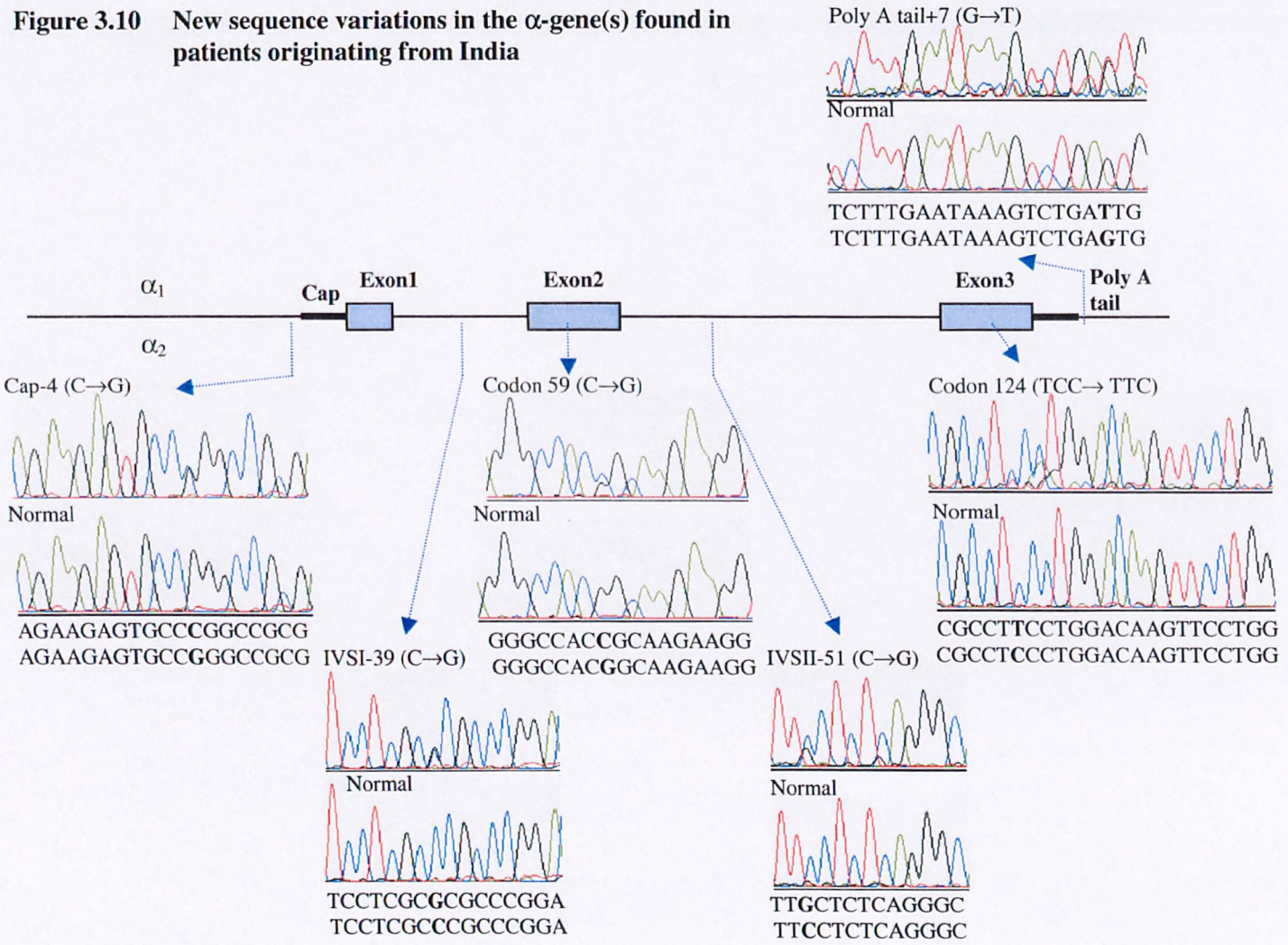
A mutation in the cap region of the α_2-gene (Cap+14 (C→G)) was found in the third patient with borderline/normal indices. This mutation was also found in combination with a second mutation located in intron-2 (IVSII-70 (T→G)) and presented with a more severe red cell index. Previously described mutations in the cap region of the β-globin gene have all been associated with an extremely mild phenotype (Huisman *et al.*, 1997) and this mutation has been previously described as a polymorphism (Hartveld *et al.*, 1996). The mutation in intron-2 was not similar to any of the observed globin donor or acceptor motifs and needed further characterisation.

The fourth patient presented with moderately reduced red cell indices (Hb 11g/dl; MCV 60.3fl and MCH 20pg) and was found to have a single heterozygous nucleotide substitution in the first intron of the α_2-gene (IVSI-39 (C→G)). This is the same mutation described in the Hackney study except that was observed in the α_1-gene. This mutation may create a new splice site to allow premature splicing but is unlike any of the observed globin donor or acceptor motifs.

The fifth case presented with mildly reduced red cell indices. A single heterozygous point mutation (codon 59 (GGC→CGC)) was found in the second exon of the α_2-gene. This base change would alter the amino acid from a glycine to an arginine that may result in a nonsense or unstable protein that cannot form part of a stable haemoglobin tetramer.

The final case presented with near normal indices but was found to have a heterozygous change (IVSII-51 (C→G)) in the α_2-gene but is unlike any of the observed globin donor or acceptor motifs.

Figure 3.10 **New sequence variations in the α-gene(s) found in patients originating from India**



Sri Lanka: α-gene sequence

Southern blot and sequence analysis was performed on 19 cord blood samples that exhibited a wide range of Hb Bart's level. The majority exhibited a normal consensus sequence.

Three samples showed sequence divergence with the normal consensus α-gene sequence. The first (K83) was found to have the same heterozygous point mutation (Cap+14 (C→G)) in both the α_1 and α_2-gene (Figure 3.9). This was associated with an Hb Bart's level of 2.0%. The second and third individuals (K93 and K4) were found to have the same heterozygous mutation (Cap-4 (C→G)), except the former was identified in the α_1-gene and the latter in the α_2-gene (Figure 3.10). The Hb Bart's level found in these cord bloods was 1.2% and 2.2%. This same mutation was described in a patient from the Indian subcontinent earlier. The Hb Bart's level indicates that this degree of chain imbalance would most likely be associated with a mild phenotype or a polymorphism.

3.3 Discussion

Many surveys have been undertaken to evaluate the incidence of α-thalassaemia in different populations based on the quantification of Hb Bart's levels in cord bloods (reviewed in Weatherall & Clegg, 2001). These were considered crude as estimates for Hb Bart's are technically difficult and subsequent correlation with the α-genotype was inconsistent. The advent of DNA analysis allowed a much more definitive assessment and became the most widely used method but screens mainly for known forms of α-thalassaemia. α-gene deletions are common across many tropical and sub-tropical regions. Characterisation of the prevalence of these forms would allow a greater understanding of the subsequent health risks.

Technical developments

Hb Bart's quantification

An accurate determination of Hb Bart's levels is critical for its subsequent correlation with α-thalassaemia, especially at low levels (<2%). CE-HPLC was the most efficient method tested. The sensitivity was good as all samples analysed showed some Hb Bart's ranging from as little as 0.5% of total haemoglobins. Previous studies reported using CAE included many samples with no Hb Bart's (Pembrey *et al.*, 1975; Higgs *et al.*, 1982; Lie-Injo *et al.*, 1982).

The α-thalassaemia short program and the β-thalassaemia short program yielded similar evaluations for Hb Bart's, in most cases, but a small number of samples (3.6%) gave values for the level of Hb Bart's higher by a factor of two. Further examination using IEF and CAE showed a better correlation with the α-thalassaemia short program. HPLC analysis could be related to other methods even if the estimations were different. Hence, a system of calibration was required to relate the level of Hb Bart's with the α-gene status (Papassotiriou *et al.*, 1999).

In the Hackney cord blood study a tri-modal distribution was observed with a good correlation between the level of Hb Bart's and the degree of α-gene disruption and subsequent expression. Below 1.5% all samples demonstrated a normal α-gene arrangement. All other previous studies showed that the assessment of Hb Bart's at low levels was inadequate to define the αα/αα and -α/αα genotype using CAE (Higgs *et al.*, 1982; Lie-Injo *et al.*, 1982). At the cut off point (1.5%) both genotypes were observed and so formed a point of overlap.

The second cut off point was more distinct as no cord bloods were found with Hb Bart's levels above 5.7% or below 6.1%. Within the first and second anti-nodes cord bloods were associated with a single α-gene deletion. However, over one fifth of the cases within this category did not show a single α-gene deletion and required further investigation. Further investigation of the seven available samples with a putative αα/αα genotype revealed two known non-deletion α-thalassaemia and one with a putative new non-deletion α-thalassaemia. However, four cases still appeared to have normal consensus sequence. All samples with Hb Bart's levels above 5.9% and up to 11.9% correlated with α-gene arrangements possessing just two genes. A distinction between -α/-α and --/αα could not be made. These observations were in line with previous studies (Higgs *et al.*, 1982; Lie-Injo *et al.*, 1982).

Analysis of Hb Bart's in fresh cord bloods appeared to represent an extremely efficient and accurate indicator of the incidence of α-thalassaemia without the need to perform DNA analysis. The cut off values to distinguish different a-thalassaemia groups matched a previous study observed using CAE analysis (Pembrey *et al.*, 1975). However, DNA analysis of a selection of frozen cord bloods from Sri Lanka only showed a rough agreement with this categorisation. Any discrepancy may have been

due to an unusual gene arrangement in either the α- or γ-genes but was, more likely, due to an artefact of the process of freezing the sample prior to analysis. Unfortunately, a haemoglobin analyser was not available on the island as the cost of the machine to the Health Service or to research groups was deemed prohibitive. Hence, the only way to analyse these samples was to transport them on ice whenever they were collected or to transport them as batches of frozen lysates.

DNA analysis

Many groups have designed PCR based methods for the detection of the common α^+-gene deletions (Dode *et al.*, 1993; Baysal & Huisman, 1994; Bowie *et al.*, 1994; Chang *et al.*, 1994; Smetanina & Huisman, 1996; Galenello *et al.*, 1998; Oron-Karni *et al.*, 1998; Chong *et al.*, 2000a and b and Shaji *et al.*, 2000). PCR assays have also been developed for the detection of α^0-gene deletions (Chang *et al.*, 1991; Bowden *et al.*, 1992; Ko *et al.*, 1992; Bowie *et al.*, 1994; Winchagoon *et al.*, 1995; Chong *et al.*, 2000a and b and Eng *et al.*, 2000). All methods have needed to overcome the high G-C content and the duplicated nature of the region but the earlier protocols suffered from a lack of reproducibility in other laboratories. The α^0-gene deletion assays have been more successful probably as greater areas of sequence homology are deleted.

Amplification of the α-gene and the development of a specific mutation detection system for α-gene mutations are subject to the same problems. They have been overcome in three ways. The first was by amplification of small segments to build a jigsaw puzzle (Hsia *et al.*, 1989; Kropp *et al.*, 1989; Hartveld *et al.*, 1996). The second was by using multiple rounds of PCR using nested primers (Dode *et al.*, 1990; Yuregir *et al.*, 1992) and the third was to use additives to enhance amplification (Dode *et al.*, 1990). Many reagents have been included in an attempt to increase

amplification efficiencies (Bachmann *et al.*, 1990; Rees *et al.*, 1993; Varadaraj & Skinner, 1994; Henke *et al.*, 1997).

New oligonucleotide primers for Gap-PCR and amplification of the α-genes were designed but the most important improvement for all these protocols was the enhancement of the PCR buffer conditions. For all sets of oligonucleotides the addition of betaine, DMSO and Tween 20 in an $(NH_4)_2SO_4$ buffer with the use of hot start conditions greatly improved the amplification of PCR within the G-C rich α-cluster. This also applied to the cycle sequence extensions and for ARMS-PCR to directly detect specific mutations. In fact, some previous studies may turn out to be useful if optimised using the new buffer systems. All PCR strategies require the primary DNA sequence at the primer annealing sites to remain unchanged so that amplification is not hindered.

Clinical consequences of α-thalassaemia

The estimated α-genotype frequency in a cohort of transfusion dependent patients and by measurement of the level of Hb Bart's was remarkably similar. Single α-gene and double α-gene deletions were observed at 13.8 and 0.4%, respectively. Hence, in a population of 18.6 million the approximate incidence of heterozygous and homozygous α^+-thalassaemia would be predicted to be 2.6 million and 56 thousand, respectively. Generally, these mutations would be associated with a normal phenotype and would pose no notable health problems (See section 1.2.2.2) but reach these tremendously high levels because of their selective advantages.

Transfusion dependent β-thalassaemia patients have severely defective β-globin chain production that results in an excess of α-chains. Any factor that modifies the magnitude of this surfeit should have an important effect on their phenotype (Ratip *et al.*, 1997; Weatherall, 2001).

This study group could be directly compared with cord blood samples from a maternity unit that should have consisted of individuals with a normal incidence of haemoglobinopathies on the island. However, as samples from the two study groups gave almost identical results any modifying effects from α-thalassaemia was probably coincidental.

No evidence of α^0-thalassaemia was found on the island and as a consequence Hb Bart's hydrops fetalis or HbH were not observed. Cord blood analysis did not reveal many candidates for non-deletion α-thalassaemia. Limited sequence analysis on these individuals was uninformative.

Evolutionary aspects

Incidence of different forms of α-gene deletions and insertions

Past studies on Indian tribal groups showed high carrier frequencies of α-thalassaemia ranging from 50% up to 80% (Brittenham *et al.*, 1980; Kulozik *et al.*, 1988; Gupta *et al.*, 1991; Modiano *et al.*, 1991). By contrast, the carrier frequency in the non-tribal population had been found to be much lower ranging between 4-15% (Hassall *et al.*, 1988; Misra *et al.*, 1991; Desai *et al.*, 1997). The high frequency of α-thalassaemia in the tribal groups was probably due to the manner in which these communities have remained closed. The genetic pool becomes limited and some traits can become greatly amplified. The incidence of α-thalassaemia observed in Sri Lanka was similar to the non-tribal groups of India but is also in keeping with populations from many tropical and sub-tropical zones. The indigenous population of Sri Lanka, the Veddah, have never been tested for the incidence of α-thalassaemia but probably align more closely with tribal populations of India.

In Sri Lanka only two types of α-thalassaemia were observed, $-\alpha^{3.7}$ and $-\alpha^{4.2}$, that had allele frequencies of 6.5% and 1.1%, respectively. This occurrence was observed all round the island in the transfusion dependent patients and also in the cord blood study. Only when the number of patients in a study group was small, were the relative frequencies notably different. However, the most northerly centre studied with a reasonably large number of alleles, Anuradhapura, had a higher allele frequency for both these single α-gene deletions of 10.8%. Towards the lower middle of the country the combined allele frequency for Badulla, Colombo and Kurunegala was almost half this value at 5.7%.

The bottom left hand corner of the island, extending just above Colombo and across to Ratnapura and Matara, is thought to constitute the wet zone. The number of breeding grounds for the different *Anopheles* species (pools on the beds of rivers and streams) depends on the pattern of rainfall. In the wet zone abundant rainfall throughout the year maintains a constant flow in the rivers. In the dry zone of the country, where malaria was endemic large numbers of surface pools provide ideal breeding grounds for the insect (van der Hoek *et al.*, 1997). The higher incidence of α-thalassaemia in the north may correlate with the greater exposure to malaria.

In comparison, the spectrum of α-thalassaemia mutations reported on the Indian subcontinent showed the same predominant molecular lesions. Approximately 80% of the mutations seen were the $-\alpha^{3.7}$ allele and most of the rest were the $-\alpha^{4.2}$ allele (Kulozik *et al.*, 1988; Hassall *et al.*, 1988). The $-\alpha^{3.7}$ allele is found in nearly all populations but the $-\alpha^{4.2}$ allele is found at higher frequencies in SE Asia.

Even though no α^{0}-thalassaemia alleles were found in our study a single case of HbH disease has previously been reported (de Tissera *et al.*,

1988). Its clinical significance on the island is low as its incidence probably only occurs in isolated families similar to India (Drysdale & Higgs, 1988). The spectrum and frequency of α-gene arrangements in Sri Lanka and India was more consistent with a shared genetic history or the same environmental influence between these two countries.

In Sri Lanka, two forms of extra α-genes were observed, $\alpha\alpha\alpha^{anti3.7}$ and $\alpha\alpha\alpha\alpha^{anti3.7}$, at frequencies of 2% and 0.2%, respectively. In this population this high incidence would give a predicted carrier frequency of 0.76 million people with extra α-genes. Extra α-genes, without any other haemoglobinopathy, are asymptotic but in combination with β-thalassaemia can deleteriously exacerbate that condition (Discussed in Chapter 6). Interestingly, the $\alpha^{anti3.7}$ gene insertion is the counterpart of the $\alpha^{3.7}$ gene deletion and when a cross over event has occurred both alleles exist in the progeny. Successive progeny should have an equal chance of passing on the two different gene arrangements. This was obviously not the case and was further support for the positive selection for α-thalassaemia.

The distribution pattern for extra α-genes is anecdotal. Patients with extra α-genes have been described at low levels in most populations studied from Africa, around the Mediterranean, in the Middle East and in India (reviewed in Weatherall & Clegg, 2001). In fact, a population study on natives from the Punjab demonstrated a single α-gene deletion of 10% and a single α-gene insertion of 5% (Garewal *et al.*, 1994). Any link between populations required further studies to evaluate the level of extra α-genes in India to make a rational comparison.

Incidence of different forms of non-deletion α-thalassaemia

The global incidence of non-deletion α-thalassaemia is not known but as the α-gene is thought to be older than the β-globin gene (Czelusniak

et al., 1982) and is duplicated a greater possibility of mutations should have occurred. Previous cord blood on Greek and Middle Eastern populations have indicated that the number of samples with elevated Hb Bart's not due to α-gene deletions is variable between populations and lies between 1 and 10% (Kyriacou *et al.*, 2000). However, the underlying cause for the chain imbalance has not always been identified and many cases remain an enigma.

A similar fraction of cord blood specimens from the Hackney study (~10% of total) also exhibited raised Hb Bart's levels that were probably not due to deletion of the α-globin gene. In Sri Lanka cord blood specimens confirmed the same level of α-thalassaemia in the general population (Table 3.5) as observed in the transfusion dependent β-thalassaemia major and HbE/β-thalassaemia patients (Table 3.1). Thus, the incidence of non-deletion α-thalassaemia in Sri Lanka would appear to be non-existent. However, 22/75 of these cord blood samples (Table 3.6) did not agree with the predicted genotype associated with the estimated level of Hb Bart's after DNA analysis. This would imply that either the estimation of Hb Bart's may have been flawed or that different forms of α-thalassaemia may be found on the island that was not revealed in the transfusion dependent β-thalassaemia major and HbE/β-thalassaemia patients investigated. Southern blot analysis of these cord blood samples indicated that the α-globin gene clusters appeared normal. α-globin sequence analysis was performed on a cross section exhibiting different levels of Hb Bart's.

In Sri Lanka, only two sequence variations were revealed from 19 cord blood samples analysed. A point substitution in the promoter region (Cap-4 (C→G)) and a point substitution in the cap region (Cap+14 (C→G)) that were both associated with modestly elevated Hb Bart's levels. Further

regional cord blood analysis within Sri Lanka needs to be performed to identify a larger cohort of candidates that may have non-deletion α-thalassaemia. The latter mutation has been observed in many world populations but the former was only observed in a patient from Pakistan and link these two populations. Only more intensive studies of individuals from Sri Lanka and the Indian subcontinent will the significance of these mutations be determined.

In conclusion

Deletion forms of α- thalassaemia were found to be common in Sri Lanka as was that of extra α-genes that is common for a population that have been exposed to malaria for a long period of time. However, no evidence was observed for the presence of α^0-thalassaemia and probably the only variation in the normal consensus sequence was due to polymorphisms. These α-gene arrangements are both consistent with a tropical population but more specifically are in keeping with the notion that this population is closely linked to that of some populations within the Indian subcontinent. The advancement of methods to screen for these various molecular lesions by either biochemical analysis or by DNA analysis will allow much more accurate determinations on their frequency and spectrum. Indeed, micro-mapping may reveal groups within this population that are prone to specific forms of these gene arrangements that may well exist in isolated communities. Their identification may infer genetic ties with other populations but more importantly could improve the course of management of some forms of thalassaemia to result in enhanced health care and better quality of life.

Chapter 4 β-Thalassaemia

The objective of the study was to gain an assessment of the spectrum and frequency of β-thalassaemia and β-chain variants in Sri Lanka. The clinical importance, in combination with each other, assessed and a comparison of their incidence made with other populations around the world, especially against groups within the subcontinent of India.

4.1. Schoolchildren survey

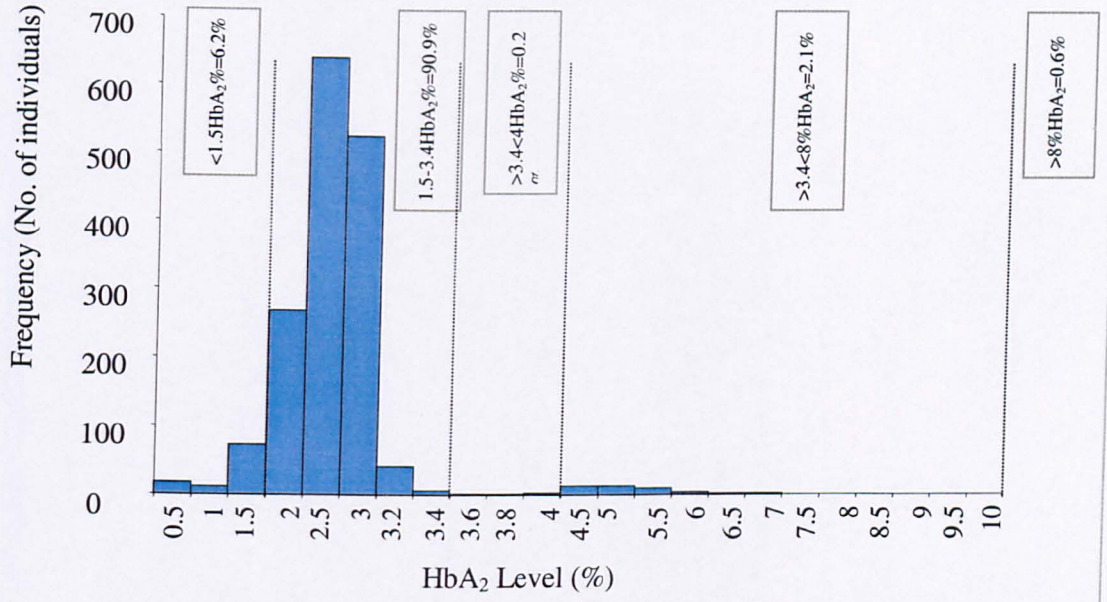
Over 1600 blood samples were collected from 18 schools in 15 districts and haemoglobins analysed by CE-HPLC in Oxford. No sample collection was undertaken from the Northern Province due to the ongoing civil war. Haemoglobin variants can be detected by protein separation techniques and the incidence of β-thalassaemia estimated by measuring the amount of HbA_2. CE-HPLC is a powerful technique to identify Hb variants and to assess the level of haemoglobins HbF and HbA_2. However, the elution profile time from CE-HPLC is small and Hb variants may be obscured by co-elution, particularly with HbA.

After developmental switching to the adult forms of haemoglobins a small amount of HbA_2 is seen in all individuals (See section 1.2.1). A distribution plot of HbA_2 values up to 10% combining all the regions on the island is shown in Figure 4.1. Five distinct groups were interpreted as correlating with different globin genotype states.

(i) A very low HbA_2 level < 1.5% may be due to α-thalassaemia or δ-thalassaemia and can also be due to iron deficiency. β-Chains preferentially form tetramers with the α-chains. If the latter are scarce, the formation of HbA_2 is also reduced as with any reduction in the level of δ-chain

production. A frequency of 6.2% was observed which is about half the carrier frequency for α-thalassaemia observed on the island.

Figure 4.1 Distribution of blood Hb A_2 levels in Sri Lankan schoolchildren



(ii) A normal distribution range for HbA_2 showed values between 1.5% and 3.4%. A high proportion of the schoolchildren (90.9%) fell into this group.

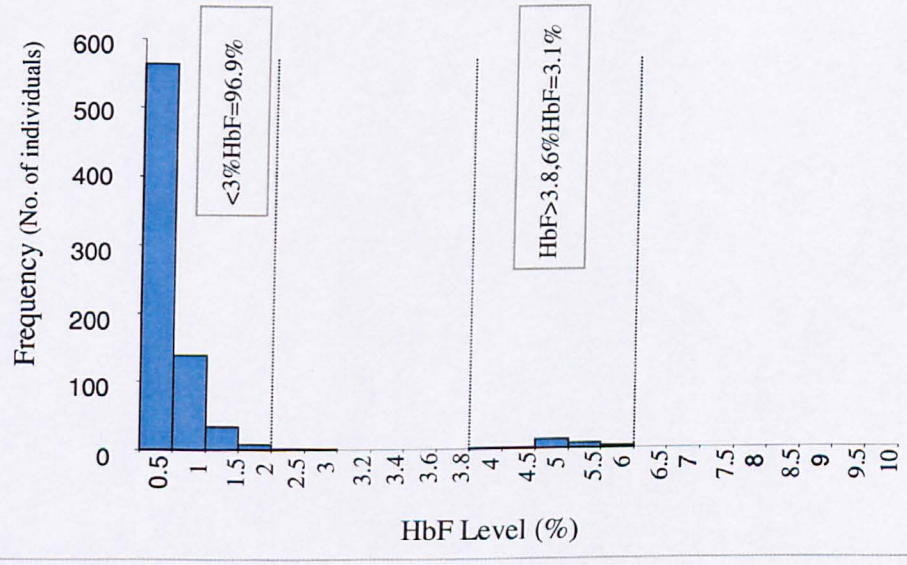
(iii) Individuals with HbA_2 between 3.4 and 4% were in the 'borderline raised group'. These cases may have been part of the high end of the normal spectrum or due to a very mild form of β-thalassaemia. β-thalassaemia has a paucity of available β-chains and a concomitant excess of α-chains that form tetramers with the available δ-chains. Only a few individuals, 0.2%, were observed in this group.

(iv) Individuals with an elevated level of HbA_2, 4 to 8%, were thought to have β-thalassaemia in a slightly more severe fashion (moderate to no expression of the β-chains) but may have been due to β-gene deletions. Again HbA_2 forms due to the lack of β-chains. The incidence on the island in this cohort of schoolchildren was 2.1%.

(v) Individuals with a highly elevated HbA_2 level, above 8%, probably possess a haemoglobin variant or may have a β-globin gene deletion, even though these forms of thalassaemia are extremely rare. In the population of Sri Lanka HbE was a good candidate has been previously reported in this population and is known to co-elute with HbA_2.

Developmental switching of haemoglobins around birth results in the expression of HbA instead of HbF, but a residual level of HbF is often observed in many individuals (See section 1.2.1). A distribution plot of HbF values up to 10% combining all the regions on the island is shown in Figure 4.2. The presence of two distinct groups suggested different haemoglobinopathy states.

Figure 4.2 Distribution of blood Hb F levels in Sri Lankan schoolchildren



(i) Individuals with an HbF level less than 3% were classed as 'normal'. All populations show this range but the cut-off value is not so high. The majority of the schoolchildren (96.9%) demonstrated an HbF level in this group.

(ii) Moderately elevated levels of HbF between 3.8 and 6.0% could have been caused by a number of factors. The HbF level evaluation by HPLC

needed to be assessed critically as another haemoglobin fraction (HbA_{1C}; glycated haemoglobin observed in all patients but elevated in individuals with diabetes) elutes very shortly after HbF and may be confused with it. Any modest increase in the HbF level could indicate non-deletion HPFH or δβ-thalassaemia. These mutations ablate the expression from one β-globin gene but often involve the deletion of the nearby δ-gene so that a concomitant increase in the level of HbA_2 cannot occur. About 3.1% of the schoolchildren were in this group but could not be investigated further.

All centres demonstrated small variations in the level of β-thalassaemia (1-5%) and HbE trait (0-2%) (Table 4.1). This was possibly due to random sampling inconsistencies. These could be confirmed by increasing the number of samples tested at each location.

Table 4.1 Incidence of β-thalassaemia and haemoglobin variants in Sri Lanka

Province	District	Population	Number Screened	βTr	HbE	HbS	HbD
Northern	Jaffna	129,000	-	-	-	-	-
North Central	Anuradhapura	763,000	100	2	1	0	1
	Polannaruwa	340,000	100	0	1	0	0
North Western	Kurunegala	1,498,000	200	5	3	0	0
	Puttlam	636,000	100	4	0	0	0
Eastern	Trincomalee	331,000	100	5	0	1	0
Central	Kandy	1,306,000	100	4	0	0	0
	Matale	440,000	100	3	0	0	0
Western	Colombo	1,967,000	138	1	2	0	0
	Gampaha	1,582,000	100	1	1	0	0
	Kalutara	979,000	100	2	0	0	0
	Negambo	50-100,000	70	0	0	0	0
Sabaragamuwa	Kegalle	770,000	100	2	1	0	0
	Ratnapura	942,000	-	-	-	-	-
UVA	Badulla	748,000	100	4	0	0	0
	Monoragala	373,000	-	-	-	-	-
	Kataramagama †	<10,000	100	0	0	1	0
Southern	Galle	906,000	-	-	-	-	-
	Hambantota	544,000	100	4	0	0	0
	Matara	822,000	100	0	0	0	0

NB: † The samples from Kataramagama were shipped as dried blood spots on filter paper and upon analysis all samples appeared to have elevated HbF levels. It was concluded that these samples were badly degraded and should be omitted from the study.

Overall, 0.8% of the individuals tested presented with obvious haemoglobin variants. Only three structural variants were implicated; HbD, HbE and HbS. HbD and HbS were only found in 3 individuals and represented rare alleles on the island. HbE co-eluted with HbA_2 and needed closer scrutiny. Most samples with elevated HbA_2 or Hb variants were investigated further by DNA analysis. The estimate of β-thalassaemia at every site around the island ranged from 0-4% of the schoolchildren tested.

Fifteen blood capillary blood samples were also made available from a group of the original island inhabitants, the Veddahs. Three of these had elevated HbA_2 levels consistent with β-thalassaemia one of which was determined to be IVSI-5 (G→C) by ARMS-PCR. However, the relationship between these individuals was not recorded.

DNA was also extracted from 42 blood samples collected in the schoolchildren survey that exhibited an elevated level of HbA_2 or a haemoglobin variant. These DNA samples were used to confirm some of the common β-thalassaemia mutations identified in the hospital patients using ARMS assays. Sixteen children were found to be carriers for IVSI-5 (G→C), three for IVSI-1 (G→A), four for HbE and one for HbD. The mutation was not identified in twelve of the remaining cases and the DNA from six individuals would not amplify.

Assuming a Hardy-Weinberg distribution and an estimate for the population of Sri Lanka number as 18.4×10^6 then estimates of affected individuals can be made. The number of individuals homozygous for β-thalassaemia would be 2650 and the number of individuals with HbE/β-thalassaemia would be 956 people. The current birth rate is about 367,000 live births per annum that would add a further 52 β-thalassaemia major infants and 20 HbE/β-thalassaemia every year.

4.2 The molecular basis of β-thalassaemia

The molecular characterisation of the cases with β-thalassaemia was studied to ascertain the range and frequency of the mutations present within the population. This would give the necessary information about the disease severity associated with the mutations and suggest the possible origins of these mutations.

Nine major centres took part in this study as outlined in Figure 2.1. In all but two of these centres, Galle and Ragama, genomic DNA was extracted (See section 2.4). 703 patients who were receiving regular or intermittent blood transfusions were studied to define different types of β-thalassaemia and their associated mutations in the patient population (Table 4.2). Overall, the pattern of haemoglobinopathies was similar in each centre. The majority of patients were homozygous (40.4%) or compound heterozygotes (21.6%) for β-globin gene mutations. HbE was found in almost a quarter (176 cases out of 703) of the symptomatic patients. In the hospitals with the largest number of patients (>40 patients) the proportion of those with HbE/β-thalassaemia ranged from 16-31% (Mean=25%). In Anuradhapura, Badulla Chilaw and Kandy the proportion with HbE/β-thalassaemia was between 16.9-18.6% but in Kurunegala this disorder represented 31.9% of the total cases. Seven patients with HbS (1%; 7 cases) and two with HbD (0.3%; 2 cases) were encountered in four out of the nine centres.

Using the prevalence of β-thalassaemia from the Indian subcontinent as a guide, a screening strategy was applied to the first batches of patients from Sri Lanka (~100 cases from Anuradhapura and 200 cases from Kurunegala). Thirteen β-globin gene mutations (Cap+1(A→C); CD5 (-CT); CD8/9(+G); CD15 (G→A); CD16 (-C); CD30 (G→C); IVSI-1 (G→T); IVSI-5 (G→C); CD41/42 (-TCTT); 619bp Del.; HbS; HbD and

HbE) were screened for using ARMS-PCR or RFLP-PCR (Tables 2.6, 2.7 and 2.8). The screen included the haemoglobin variants as they needed confirmation or may have been masked by blood transfusions. As the Portuguese were in control of the island (See section 1.1.1) and also have a high prevalence of β-thalassaemia, mutations common in that community were also assessed (IVSI-1 (G→A); IVSI-6 (T→C); IVSI-110 (G→A) and CD39 (C→T)) (Tables 2.6). Subsequent batches of samples were screened for the mutations already identified and the remainder sequenced.

Table 4.2 Distribution of haemoglobinopathies in patients treated for thalassaemia in nine hospitals in Sri Lanka

Hospital	Anuradhapura	Badulla	Chilaw	Colombo	Galle	Kandy	Kurunegala	Negambo	Ragama	Total
β-thalassaemia major										
Homozygote	62 (45.6)	49 (64.5)	17 (32.7)	19 (57.6)	2 (10.0)	23 (53.5)	112 (35.3)	NA	NA	284 (40.4)
Compound heterozgote	37 (27.2)	9 (11.8)	17 (32.7)	7 (21.1)	0 (0)	8 (18.6)	74 (23.4)	NA	NA	152 (21.6)
Unknown	12 (8.9)	4 (5.3)	8 (15.4)	6 (18.2)	8 (40.0)	4 (9.3)	28 (8.8)	3 (21.4)	9 (75.0)	82 (11.7)
$\beta^{Variant}$/β-thalassaemia										
HbE/β-thalassaemia	23 (16.9)	14 (18.4)	9 (17.3)	1 (3.0)	6 (30.0)	8 (18.6)	101 (31.9)	11 (78.6)	3 (25.0)	176 (25.0)
HbS/β-thalassaemia	1 (0.7)	0 (0)	1 (1.9)	0 (0)	3 (15.0)	0 (0)	2 (0.6)	0 (0)	0 (0)	7 (1.0)
HbD/β-thalassaemia	1 (0.7)	0 (0)	0 (0)	0 (0)	1 (5.0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (0.3)
Total	136 (100.0)	76 (100.0)	52 (100.0)	33 (100.0)	20 (100.0)	43 (100.0)	317 (100.0)	14 (100.0)	12 (100.0)	703 (100.0)

Figures in brackets represent the frequency as a percentage
NA= not available

Table 4.3 Frequency of the β-globin gene mutations found in hospital patients being treated for thalassaemia

Mutation	Method of analysis	Anuradhapura		Badulla		Chilaw		Colombo	
		Allele No.	(%)	Allele No.	(%)	Allele No.	(%)	Allele No.	(%)
IVSI-5 (G→C)	ARMS	161	64.4	108	71.1	48	54.6	36	52.9
IVSI-1 (G→A)	ARMS	29	11.6	4	2.6	10	11.4	14	20.7
CD26 (G→A) HbE	ARMS	23	9.2	13	8.6	9	10.2	1	1.5
CD41/42 (-TCTT)	ARMS	1	0.4	2	1.3	5	5.7	2	2.9
CD15 (-T)	Sequence	6	2.4	10	6.6	2	2.3	2	2.9
CD16 (-C)	ARMS RFLP	2	0.8	0	0	4	4.5	0	0
CD8/9 (+G)	ARMS	8	3.2	0	0	2	2.3	0	0
CD30 (G→C)	ARMS	4	1.6	0	0	0	0	0	0
IVS1-130 (G→C)	Sequence RFLP	2	0.8	0	0	2	2.3	1	1.5
CD15 (G→A)	ARMS	0	0	0	0	0	0	1	1.5
Poly A Tail (T→C)	Sequence ARMS	3	1.2	0	0	0	0	2	2.9
CD55 (-A)	Sequence RFLP	0	0	4	2.6	0	0	0	0
CD6 (A→T) HbS	RFLP ARMS	1	0.4	0	0	1	1.1	0	0
IVSI-1 (G-T)	Sequence ARMS	0	0	0	0	0	0	4	5.9
IVSII-1 (G→A)	Sequence ARMS	0	0	1	0.7	1	1.1	0	0
IVS1-129 (A→C)	Sequence RFLP	0	0	0	0	0	0	0	0
IVS1-130 (G→A)	Sequence RFLP	2	0.8	0	0	0	0	0	0
-28 (A→G)	Sequence ARMS	0	0	2	1.3	0	0	0	0
25 bp Deletion	Sequence RFLP	0	0	2	1.3	0	0	0	0
IVSII-745 (C→G)	Sequence ARMS	0	0	0	0	0	0	2	2.9
CD5 (-CT)	Sequence ARMS	0	0	0	0	1	1.1	0	0
CD6/CD10 (-13bp)	Sequence	0	0	0	0	0	0	0	0
CD121(G→C) HbD	Sequence RFLP	0	0	0	0	0	0	0	0
619 bp deletion	Gap-PCR	1	0.4	0	0	0	0	0	0
Unknown	Sequence	7	2.8	6	3.9	3	3.4	1	2.9
Total		250	100	152	100	88	100	66	100

Table 4.3 cont'd

Mutation	Galle		Kandy		Kurunegala		Negambo		Total	
	Allele No.	(%)	Allele No.	(%)	Allele No.	(%)	Allele No.	(%)	Allele No.	(%)
IVSI-5 (G→C)	5	62.5	48	60	284	49.0	7	43.8	697	56.2
IVSI-1 (G→A)	1	12.5	8	10	123	21.2	0	0	189	15.2
CD26 (G→A) HbE	0	0	8	10	101	17.4	7	43.8	162	13.1
CD41/42 (-TCTT)	0	0	10	12.5	18	3.1	0	0	38	3.1
CD15 (-T)	0	0	2	2.5	5	0.9	0	0	27	2.2
CD16 (-C)	0	0	0	0	14	2.4	1	6.2	21	1.6
CD8/9 (+G)	0	0	0	0	6	1.0	0	0	16	1.3
CD30 (G→C)	0	0	0	0	7	1.2	0	0	11	0.9
IVS1-130 (G→C)	0	0	0	0	4	0.7	0	0	9	0.7
CD15 (G→A)	0	0	1	1.25	5	0.9	1	6.2	8	0.6
Poly A Tail (T→C)	0	0	0	0	2	0.3	0	0	7	0.6
CD55 (-A)	0	0	0	0	1	0.2	0	0	5	0.4
CD6 (A→T) HbS	1	12.5	0	0	2	0.3	0	0	5	0.4
IVSI-1 (G-T)	0	0	0	0	0	0	0	0	4	0.3
IVSII-1 (G→A)	0	0	0	0	2	0.3	0	0	4	0.3
IVS1-129 (A→C)	0	0	2	2.5	1	0.2	0	0	3	0.2
IVS1-130 (G→A)	0	0	0	0	1	0.2	0	0	3	0.2
-28 (A→G)	0	0	0	0	0	0	0	0	2	0.2
25 bp Deletion	0	0	0	0	0	0	0	0	2	0.2
IVSII-745 (C→G)	0	0	0	0	0	0	0	0	2	0.2
CD5 (-CT)	0	0	0	0	0	0	0	0	1	0.1
CD6/CD10 (-13bp)	0	0	1	1.25	0	0	0	0	1	0.1
CD121(G→C) HbD	1	12.5	0	0	0	0	0	0	1	0.1
619 bp deletion	0	0	0	0	0	0	0	0	1	0.1
Unknown	0	0	0	0	4	0.7	0	0	21	1.7
Total	8	100	80	100	580	100	16	100	1240	100

DNA analysis was completed on 620 of these patients to reveal the spectrum of β-thalassaemia mutations that exist in the population of Sri Lanka (Table 4.3). The mutations (Cap+1 (A→C), IVSI-6 (T→C), IVSI-110 (G→A) and CD39 (C→T) were not found in any patients examined and appeared to be absent from the island. Cap+1 (A→C) and IVSI-6 (T→C) in combination with a β^0-thalassaemia would probably be associated with β-thalassaemia intermedia phenotype and hence would not present at clinic.

The haemoglobin variants HbE, HbD and HbS were confirmed and 8 different β-thalassaemia mutations (IVSI-5 (G→C), IVSI-1 (G→A), CD8/9(+G), CD30 (G→C), CD41/42(-TCTT), CD16 (-C), CD15 (G→A) and the 619bp deletion mutation) were identified using ARMS and RFLP analysis. The IVSI-5 (G→C) mutation was found to be the most common mutation in all the communities studied. The second most prevalent thalassaemia mutation overall was IVSI-1 (G→A). The remaining 6 mutations identified by ARMS-PCR were all found at a low frequency in the population of Sri Lanka. These 8 thalassaemia mutations, and the haemoglobin variants, accounted for 92.5% of the thalassaemia mutations.

DNA sequencing of the remaining unidentified alleles revealed a further 14 mutations. Eleven of these mutations –28 (A→G), CD15 (-T), IVSI-130 (G→A), IVSI-130 (G→C), 25bp deletion, IVSII-1 (G→A), CD5 (-CT), CD10 (C→A), IVSI-1 (G→T), IVSII-745 (C→G) and the polyadenylation signal mutation (T→C) have all been described previously (reviewed in Weatherall & Clegg, 2001). These were confirmed by ARMS-PCR. Upon further examination the mutation CD10 (C→A) and a second mutation CD16 (-C) appeared to be present on the same allele.

Four individuals were sequenced to reveal the β-gene mutation CD16 (-C). In each case the mutation CD10 (C→A) appears to be

invariably linked to this β^0–thalassaemia mutation. The change in sequence at codon 10 creates a new *Pst I* restriction enzyme site. Fifteen carriers for the CD16 (-C) mutation and one homozygote were screened by amplification of the first exon and subsequent *Pst I* digestion. All the alleles carrying the β^0 CD16 (-C) frameshift mutation were associated with the β^+ CD10 (C→A) mutation (Old *et al.*, 2001). The existence of either mutation on its own has not been established. The cryptic splice site mutation at codon 10 would splice out the second mutation, CD16 (-C). Further RNA analysis was required to establish the effect of these changes.

The other four mutations identified by DNA sequencing were new (Figure 4.3). All these mutations were confirmed by sequence analysis in both directions. A 13bp deletion that removed nucleotides between CD6 to CD10 was found in a Sinhalese patient who was doubly heterozygous for IVSI-5 (G→C)(Figure 4.3). This patient is transfusion dependent. The mutation IVSI-129 (A→C), as well the mutations IVSI-130 (G→A), IVSI-130 (G→C), were confirmed by RFLP analysis and digestion with the restriction enzyme *Dde I* as the mutation results in the loss of the *Dde I* recognition site and a new RFLP pattern (Figure 4.4).

The new splice site mutation IVSI-129 (A→C) (Figure 4.3) was found in just one Sinhalese patient who was a compound heterozygote for this mutation and for the mutation IVSI-5 (G→C). The resultant phenotype was that of β-thalassaemia major such that the mutation should be classified as β^0- or severe β^+-thalassaemia.

Figure 4.3 Chromatograms showing the new β-globin gene mutations

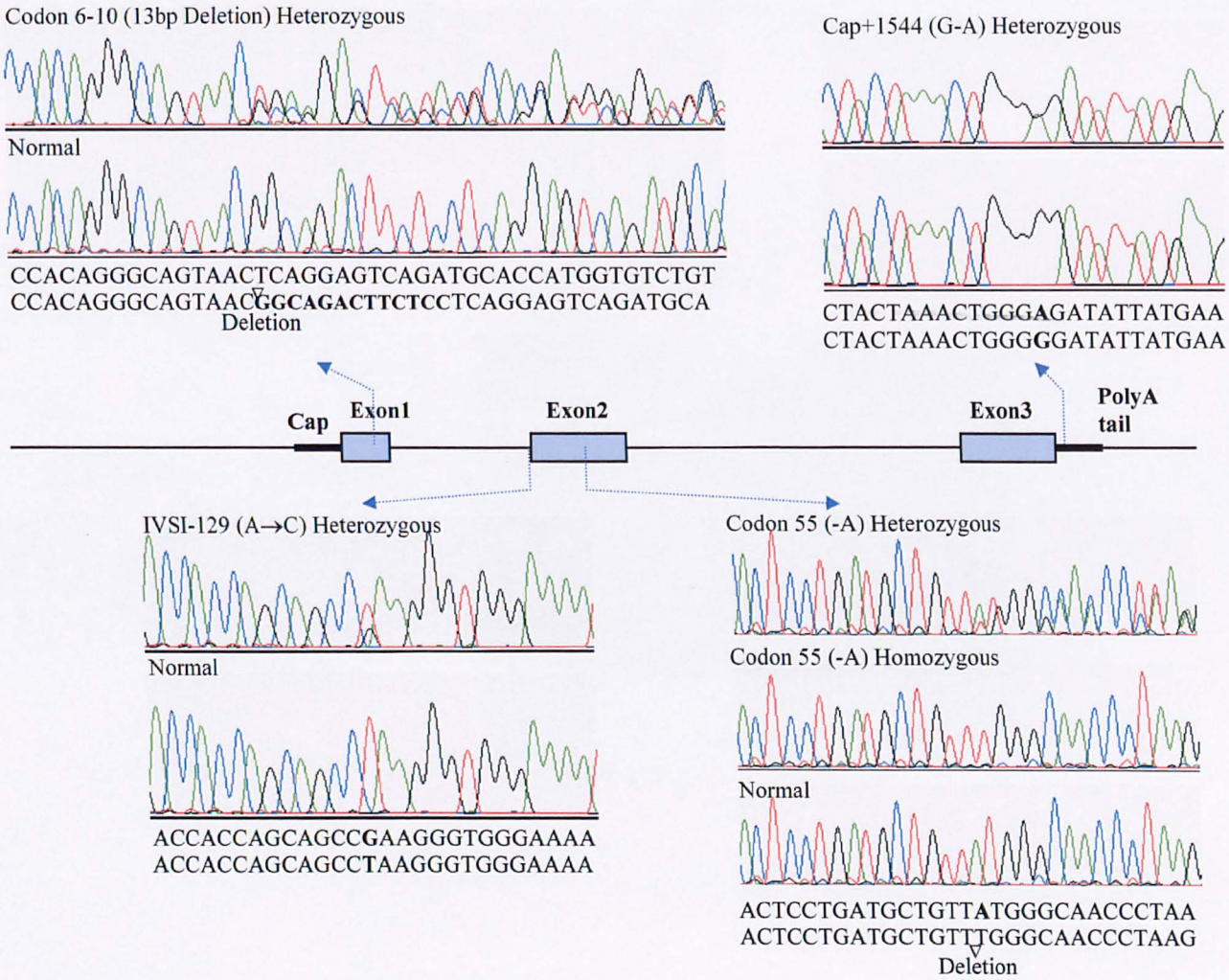
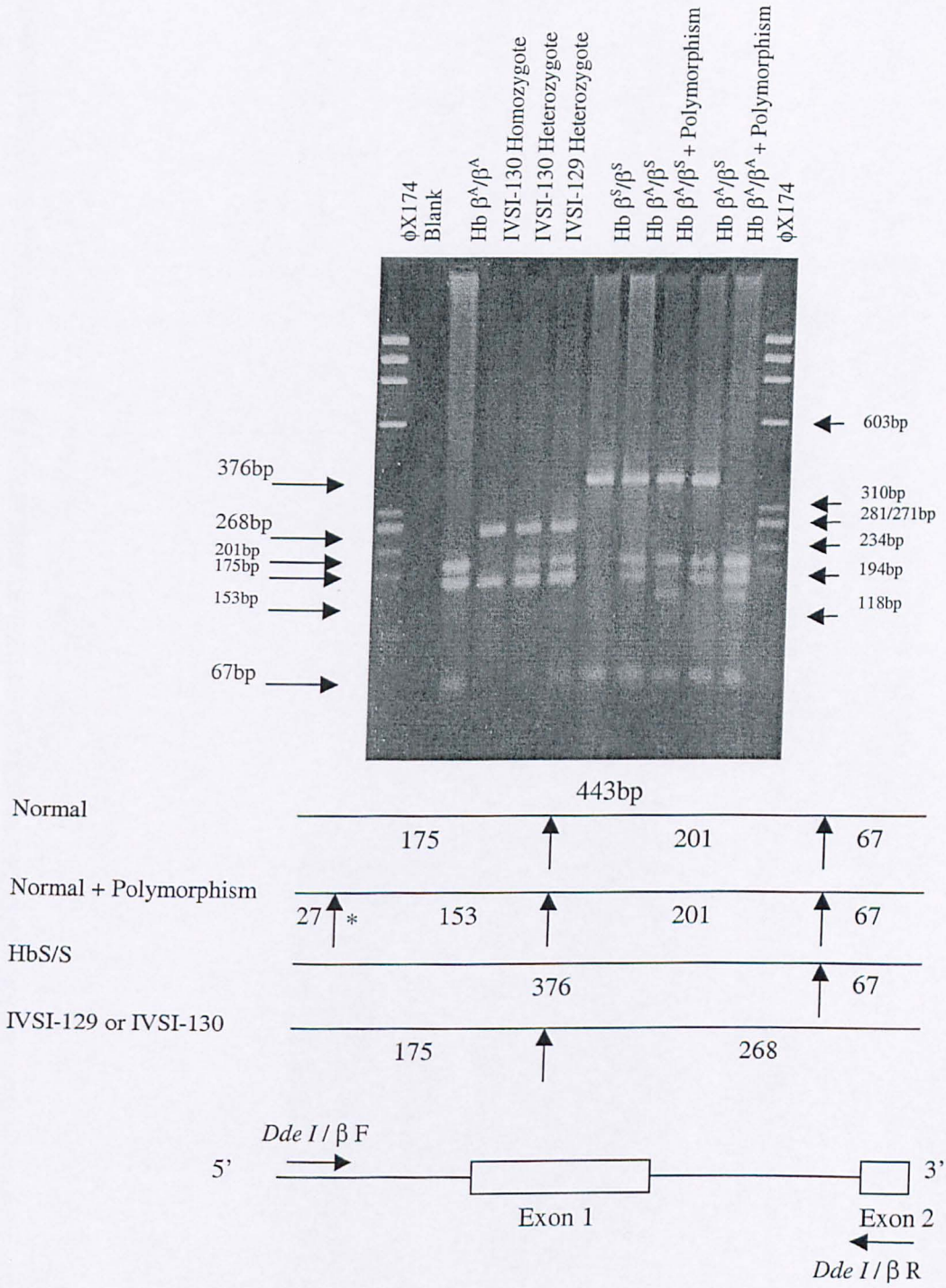
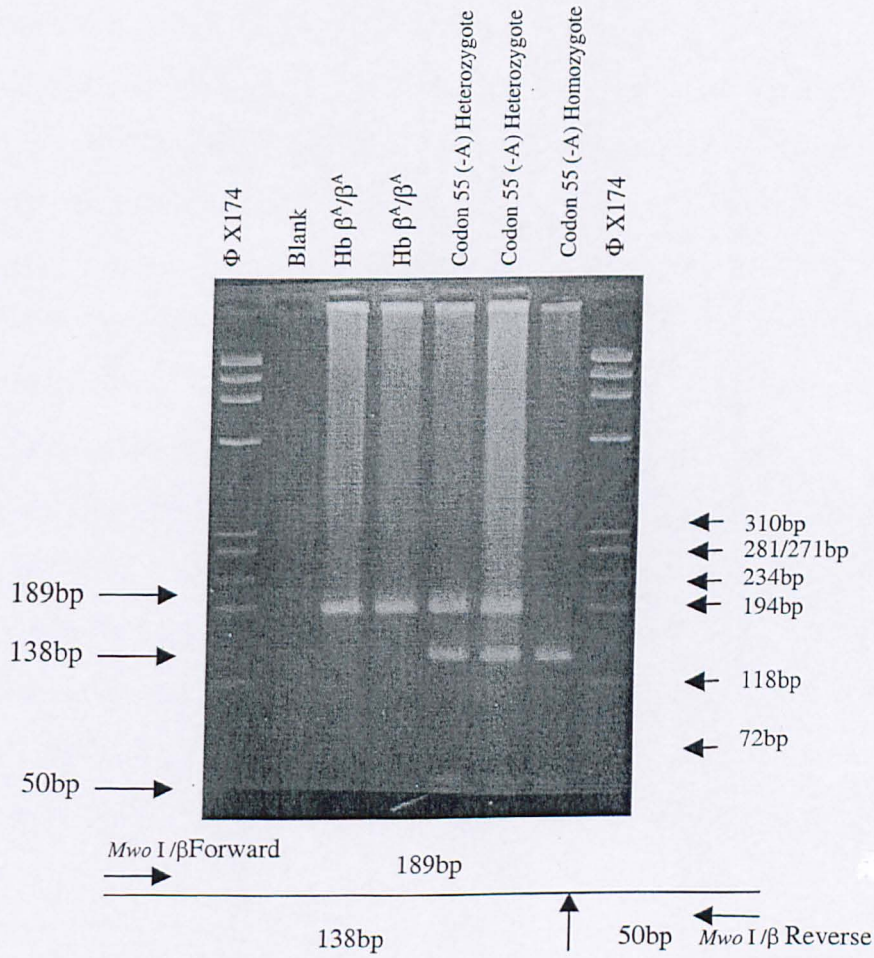


Figure 4.4 **PCR-RFLP analysis for the detection of HbS, IVSI-130 (G→A), IVSI-130 (G→C) and the new mutation IVSI-129 (A→C) through the loss of a *Dde I* restriction enzyme sites**



The third new mutation was codon 55 (-A) (Figure 4.3) a frameshift mutation that was found in four patients. Two patients were compound heterozygotes for this mutation and for the mutation IVSI-5 (G→C), another patient was a compound heterozygote for this mutation and for the mutation CD15 (-T). The last patient was homozygous for the CD55 (-A) mutation. All these individuals were treated as transfusion dependent patients and the new mutation CD55 (-A) was classified as β^0 or severe β^+. CD55 (-A) was confirmed by RFLP analysis as this mutation creates an *Mwo I* recognition site (Figure 4.5). A PCR product containing exon 2 using the primers outlined in Table 2.8 was digested with the restriction enzyme *Mwo I* and run out on a 3% agarose gel.

Figure 4.5 **RFLP analysis to demonstrate the introduction of an *Mwo I* site associated with the mutation codon 55 (-A)**



One individual from Colombo was found to have a single heterozygous change in the 3'URR at Cap+1544 (G→A) (Figure 4.3) and no other mutation on the other allele. This patient was not included in the study as the family presented with another haematological disorder, congenital dyserythropoietic anaemia (CDA). Mutations in this region of the β-globin gene are normally associated with a very mild phenotype and so the thalassaemia status requires further investigation.

Twenty-one alleles failed to show any difference compared to the normal DNA consensus sequence of the complete β-globin gene and its surrounding DNA sequence and thus remain uncharacterised after sequencing. Family studies were performed on all four patients centred in Kurunegala as the remaining families from the other centres were unavailable for further studies.

Disease in two of these patients appeared to be associated with an increased level of HbF in at least one of their parents (family DM and KHC55). In family DM the mother had a raised HbA_2 (4.7%) and was found to be a carrier for the β^0-thalassaemia mutation, CD 16 (-C) that was also found in linkage with the base change at CD10 (C→A). The father had a raised HbF (20.3%) but a normal HbA_2 and was found to have a normal β-gene after DNA sequencing. As the propositus appeared to be homozygous for the mother's mutation the father must have had a deletion and only his single normal allele could be observed. Haplotype markers along the β-cluster were examined but did not help to map the extent of the deletion. In family KHC55 the mother was found to have a raised HbF level of 19.6% but a normal HbA_2 level and appeared to have a normal β-gene after DNA sequencing. Again, the propositus was seen to be homozygous for the severe β^+-thalassaemia mutation, IVSI-5 (G→C), so

again another or the same deletion, as in family DM, that removes the β-gene must be present in this family. These two families were screened for two deletions, the Indian inversion deletion and HPFH-3 both found on the Indian subcontinent, but were found to be negative. Southern blot analysis on both families was unsuccessful and further analysis.

The other two patients had a β-gene mutation identified on one allele and the other allele was observed to have a normal consensus sequence. α-gene analysis showed that these patients had inherited extra α-genes that were interacting with the β-thalassaemia to exacerbate the chain imbalance and result in a more severe phenotype (See Chapter 6).

Of the 24 different β-thalassaemia alleles revealed in this study three mutations accounted for 84.5% and would result in a large number of homozygotes for the same mutation. This was confirmed in Table 4.2 that assigned the affected patient's genotypes after DNA analysis. Almost two thirds of the patients (65.1%) with β-thalassaemia major were found to be homozygous for the same mutation. As family studies were not performed routinely any index case observed to be homozygous could have had a deletion of the β-gene.

4.3 Discussion

Many surveys have been undertaken to evaluate the incidence of β-thalassaemia in different populations based on the elevation of HbA_2 levels above 3.4% in adult blood samples (reviewed in Weatherall and Clegg, 2001). β-gene mutations are common across many tropical and sub-tropical regions in individuals that present with varying degrees of anaemia. The prevalence of β-thalassaemia, like that of α-thalassaemia, would be maintained by positive selective advantage against infectious disease, especially in childhood (See Table 1.3 and section 1.2.2.2). As these infectious diseases have probably been present on the island for many generations the predicted level of β-thalassaemia would be high. Characterisation of the molecular lesions allows a greater understanding of the subsequent health risks.

Clinical consequences and spectrum of β-globin mutations

The carrier frequency for β-thalassaemia and for HbE in Sri Lanka was determined at 2.3% and 0.6% respectively. These were consistent with a reported estimate for Sri Lanka of 3.5%, combined β-thalassaemia trait and HbE trait, (WHO, 1994) even though the previous method of data collection appeared anecdotal. The schoolchildren survey was a crude estimate of more than 1600 individuals, <0.01% of the Sri Lankan population. This elevated level of β-globin perturbation is indicative of selective pressures to maintain these mutations in the population. The incidence of haemoglobinopathies observed in the different study centres was extremely diverse between 0 and 5%. The reason is not clear but may reflect variable exposure to malaria in the past, some degree of consanguinity, founder effects, genetic drift or sampling inadequacies. If the overall estimates are closer to 0% then the clinical consequences are

relatively insignificant but if they are closer to 5% then the predicted health burden on the island population will be great.

Four educational institutions in the schoolchildren survey were deemed within the wet/intermediate zone (Colombo, Kalutara, Kegalle and Matara). The rest were located in the dry zone. The incidence of haemoglobinopathies (combined β-thalassaemia trait and HbE trait) in the wet/intermediate zone and the dry zone were 1.8% and 3.2%, respectively. This supports the hypothesis that they have arisen due to selective advantage to protect individuals against malaria.

Local dogma suggested that the level of haemoglobinopathies in the Kurunegala region was higher than any other part of the island (Wickremasinghe *et al.*, 1963; Nagaratnam, 1989). This was not born out by the schoolchildren survey. However, the Kurunegala Teaching Hospital is a centre of excellence to treat these patients and hence it has a greater hospital patient load compared to other regions. As the hospital has a good reputation to treat transfusion dependent patients, families outside this region also attend the thalassaemia clinic. The hospital study at this centre was notably different to other regions due to its greater preponderance for HbE/β-thalassaemia patients. In Kurunegala 31.9% of the 317 patients studied had HbE/β-thalassaemia compared with all the other study centres combined where 19.4% of the 386 patients had this form of the disease. This variation in this relatively small population shows the need to assess large numbers of individuals to gain accurate estimations and also highlights the need for accurate details of the sample populations are recorded so that a reasonable cross section of the population can be assessed. This was not performed in Sri Lanka due to logistical problems.

If the Kurunegala district is analysed independently the predicted gene frequency for β-thalassaemia and HbE would be 0.0125 and 0.0075.

Assuming a regional population of about 1.5 million and a Hardy-Weinberg distribution the predicted number of patients homozygous for β-thalassaemia and HbE/β-thalassaemia would be 220 and 260, respectively (de Silva *et al.*, 2000). The number of patients currently receiving transfusions for these two disorders is 214 and 101 (Table 4.2). Hence, whereas the patients homozygous for β-thalassaemia are well represented there appears to be a shortage of patients with HbE/β-thalassaemia. This shortfall may be due to the under diagnosis of this disorder or that a group of these patients never present with the disease as they have very mild symptoms.

Using the schoolchildren survey, the number of individuals on the island predicted to have β-thalassaemia major or HbE/β-thalassaemia was 2650 and 956, respectively. This estimation did not account for the predicted length of time each of these individuals may live compared against the rest of the population. Severely affected individuals have a reduced life expectancy. On the other hand, the rate of population growth is still on the increase and the management of patients is continually improving which results in longer life expectancies and the consequent increase in the number of patients with these conditions. Currently, the anticipated number of new patients born annually would be 52 and 20 for β-thalassaemia major and HbE/β-thalassaemia, respectively.

The hospital based study group consisted of 620 transfusion dependent patients representing about $1/6^{th}$ of the total predicted number of thalassaemia patients on the island estimated from the schoolchildren survey. This gave a heavy bias towards revealing mutations associated with no or very low β-chain expression that probably occurred because all the patients were transfusion dependent. Hence, only a handful of patients had

mutations associated with a mild phenotype (promoter and polyadenylation signal mutations). In contrast, HbE was found to be common. In the homozygous state (HbE/E), the mutation is only associated with a mild anaemia. However, in combination with β-thalassaemia, the patients have an extremely variable phenotypic outcome from asymptotic to transfusion dependent. The difference between these outcomes needs to be investigated further as the consequences for the patient and for health planning are tremendous.

The frequencies of the common mutations were similar between the different study centres. Analysis for the three most common β-globin gene mutations (IVSI-5 (G→C), IVSI-1 (G→A) and HbE) would provide the basis of a genotype service that would identify 84.1% of the mutations. However, in a Pre-natal Diagnostic Service about 25% of cases would remain unresolved by analysis for these three molecular lesions. To provide a reasonably comprehensive service (>95% of successful diagnosis) over 20 of the β-globin gene mutations (top 20 mutations listed in Table 4.3) would have to be tested for. The earlier an infant is diagnosed with a severe form of haemoglobinopathy the sooner an effective form of management can be implemented. A pre-natal diagnostic service is unlikely to be established at present due to religious convictions of the majority of the population. The official religion of Sri Lanka is Buddhism and the orthodox view is against any form of abortion. However, other Buddhist countries have implemented the use of PND and Sri Lanka may one day follow suit.

It has been estimated that the cost of treating 2000 patients may have been between 3-5% of the annual expenditure on health in Sri Lanka for 1996 (de Silva *et al.*, 2000). If the number of patients increases then greater pressure will be placed on these finances and difficult moral decisions

about the efficacy of *who* and *how to treat them* will become extremely difficult.

New β-thalassaemia mutations

Currently, 198 different forms of non-deletion β-thalassaemia have been described (reviewed in Weatherall & Clegg, 2001). In this study, three new β^0-thalassaemia mutations were identified; two in the coding regions and one in intron 1. These are IVS1-129 (A→C), CD6/10 (-13) and CD55 (-A). This represents an increase in the number of non-deletion type β-thalassaemia mutations by 1.5% from a very small population.

The intronic base substitution introduces a consensus splice site mutation in the first intron, IVS1-129 (A→C), that is similar to IVSI-130 (G→A) or IVSI-130 (G→C). Another mutation in the same nucleotide has been identified in a German patient, IVSI-129 (A→G) (Vetter *et al.*, 1997). Other mutations occur at this consensus splice site (IVSI-130) all of which fail to produce β-globin chains. The first exonic mutation was a small deletion of 13 base pairs starting in codon 6 and extending to codon 10 that shifts the reading frame and results in a premature stop at codon 14. The last mutation was a single nucleotide deletion at CD55 (-A) that causes a frameshift and should result in a premature stop at codon 60 in a similar manner to a previously reported mutation at codon 54 (-T) (Huisman *et al.*, 1997). These three new mutations may have arisen independently as they have not been observed in any other populations.

Evolutionary aspects

All new variations start as a mutation that may alter the expression or activity of the resultant protein. The subsequent frequency of this change will depend upon its selective advantage or disadvantage and also the successful expansion of its host. DNA analysis has been performed on

many populations to examine the spectrum and frequency of individual mutations (reviewed in Weatherall & Clegg, 2001). In every population with high frequencies of haemoglobinopathies two or three β-globin gene mutations have been found to be present at high frequencies with a larger number of rare mutations. β-thalassaemia confers a selective advantage against malaria. Mutations that arose earlier in a population's history will be present at higher frequencies. Interaction between populations with the same selective pressure may result in the spread of these mutations and hence reduce the selective pressure on these individuals. Analysis of these mutations may indicate some interaction between populations.

The spectrum of molecular lesions was determined in the hospital based patients and confirmed in a small number of individuals in the schoolchildren survey (See section 4.1.1). Twenty-four β-globin gene mutations were identified from screening 620 hospital patients being treated for thalassaemia. Three mutations accounted for 84.4% of the 1240 alleles: IVSI-5 (G→C) 56.1%; IVSI-1 (G→A) 15.2% and haemoglobin E (GAG→GAA), a structural variant with a mild β-thalassaemia phenotype, 13.1%. The distribution of these common mutations in the different provinces was fairly uniform. A big difference was observed with the spectrum of rare mutations. As well as many individuals with HbE, two more haemoglobin variants were observed in the population, a few with HbS and a single case consistent with HbD-Punjab. All of these structural variants have been reported in India.

As Sri Lanka has close historical connections with India and Portugal (See section 1.1.1) the molecular lesions revealed were compared between these populations. Many studies to investigate the incidence of β-thalassaemia on the Indian subcontinent and in Portugal have been reported

and the combined results are shown in Table 4.4. The mutations observed in the population of Sri Lanka are very similar to that of India, especially in the North East.

Table 4.4 Frequency of β-globin mutations found on the Indian subcontinent, Portugal and Sri Lanka

Mutation	Indian subcontinent	North West India	North India	North East India	Central India	South India	Pakistan	Sri Lanka	Portugal
Cases analysed	n=6742	n=1003	n=1112	n=688	n=394	n=254	n=2752	n=620	n=252
References	1-14	1,3-6, 11,12,14	1,6,9,1 1,12,14	9,12,14	12	11,12,1 4	1,2,5,8,9, 11,14		15,16
IVSI-5(G-C)	40.95	56.72	25.27	51.59	60.14	80.31	32.58	56.2	0
619bp Del.	14.16	13.56	30.75	1.03	2.54	0.79	13.7	0.1	0
CD8/9(+G)	13.37	3.29	11.96	2.33	5.08	2.76	23.22	1.3	0
IVSI-1(G-T)	8.92	7.28	18.08	0	1.52	0	10.07	0.3	0
CD41/41(-TCTT)	6.48	6.58	7.73	2.33	11.17	8.27	6	3.1	0
CD15(G-A)	3	5.78	0.27	1.89	3.55	3.15	3.2	0.6	7.9
CD30(G-C)	1.75	1.4	0.18	3.2	3.3	0	2.14	0.9	0
CD16(-C)	1.47	0.4	1.17	0.15	3.3	1.18	2.22	1.6	0
CD5(-CT)	1.22	0.7	0.27	0	1.02	0	2.03	0.1	0
Cap+1(A-C)	1.16	0	2.07	0.58	1.78	0	1.09	0	0
CD30(G-A)	0.42	0.3	0.09	0.29	0	0	0.47	0	0
-88(C-T)	0.34	0	0.81	0.29	1.02	0	0.18	0	0
IVSI-1(G-A)	0.3	0.4	0	0	0.51	0	0.44	15.2	21
IVSII-1(G-A)	0.25	0.2	0.09	0	0	0	0.51	0.3	0
CD47/48(+ATCT)	0.22	0	0.09	0	0.51	0	0.4	0	0
IVSII-837(T-G)	0.18	0.3	0	0	0	1.97	0	0	0
CD8(-AA)	0.16	0	0	0	0	0	0.36	0	0
25bp del	0.09	0.4	0	0	0	0	0.07	0.2	0
CD39(C-T)	0.03	0	0	0	0	0	0.07	0	37.3
CD88(+T)	0.03	0	0	0	0	0	0.04	0	0
CD121(G-T)	0.03	0	0	0	0	0	0	0	0.4
CD4/CD5/CD6	0.01	0	0	0	0.25	0	0	0	0
-28(A-G)	0.01	0	0	0	0	0	0	0.2	0
CD8(A-G)	0.01	0	0	0	0.25	0	0	0	0
CD13/CD26/CD27	0.01	0	0	0	0.25	0	0	0	0
CD15(-T)	0.01	0	0	0	0	0	0.04	2.2	0
IVSI-110(G-A)	0.01	0	0	0	0	0	0	0	11.5
IVSI-130(G-C)	0.01	0	0	0	0	0	0	0.7	0.8
IVSI(3')(T-G)	0.01	0	0.09	0	0	0	0	0	0
CD44(-C)	0.01	0.1	0	0	0	0	0	0	0
CD55(+A)	0.01	0.1	0	0	0	0	0	0	0
CD57/58(+C)	0.01	0	0	0	0	0	0	0	0
IVSII-848(C-A)	0.01	0	0	0	0	0	0.04	0	0
CD126-131(-17bp)	0.01	0	0	0	0	0	0.04	0	0
IVSI-6(T-C)	0	0	0	0	0	0	0	0	19
-90(C-T)	0	0	0	00	0	0	0	0	1.2
HbE	3.13	0.1	0.09	26.15	1.27	0	0.58	13.1	0
HbD	0.07	0	0	0	0	0	0	0.1	0
HbS	0.06	0	0	0	0	0	0	0.4	0
Uncharacterised	2.08	2.39	0.99	10.17	2.54	1.57	0.51	1.7	0.4
Other	0	0	0	0	0	0	0	1.7	0

1.Agarwal *et al.*, (2000); 2.Ahmed *et al.*, (1996); 3.Bandyopadhyaya *et al.*, (1999);4 Dastidar *et al.*, (1994);5 el-Kalla & Mathews (1997); 6.Garewal *et al.*, (1994); 7.Gorakshakar *et al.*, (1999); 8.Khan & Riazuddin (1998); 9.Parikh *et al.*, (1990); 10.Thein *et al.*, (1988); 11.Varawalla *et al.*, (1991a); 12.Vaz *et al.*, (2000); 13.Venkatesan *et al.*, (1992) ; 14.Verma *et al.*, (1997); 15. Faustino *et al.*, (1992); 16. Gomes *et al.*, (1988)

The point mutation IVSI-5 (G→C) is the most common β-thalassaemia mutation not only in Sri Lanka but globally. It is found at high frequencies in S.E. Asia and the subcontinent of India and occurs in the Middle East and Europe but not in Portugal (reviewed in Weatherall & Clegg, 2001).

The second most common mutation on the island was also a consensus splice site point mutation IVSI-1 (G→A). This mutation occurs at high frequencies around the Mediterranean basin and is centred round Spain and Portugal (reviewed in Weatherall & Clegg, 2001). Isolated cases have been reported in North India (Varawalla *et al.*, 1991b) and Central America (Villalobos-Arambula *et al.*, 1997) in descendants from Spain. Hence, this mutation appears to be more closely implicated with a European origin. This would be consistent with the island history as 25 generations previously the Portuguese held dominion over the island for over a century (See Chapter 1.1).

The third most common allele HbE, is found at its highest frequencies in S.E Asia and can be found at polymorphic frequencies in the populations around the Thai/ Cambodia/ Vietnamese border (reviewed in Weatherall & Clegg, 2001). The frequency of HbE diminishes in populations that are further away from this region. HbE is known to exist in Bangladesh and North East India at high frequencies but has only been described in a few cases from North and North West India (Garewal *et al.*, 1994; Ahmed *et al.*, 1996) and has not been described in South India (Verma *et al*, 1997). Hence, its occurrence in Sri Lanka may seem unusual. However, a connection between North India and Sri Lanka has been well recorded, albeit in legend or religious documentation (See section 1.1). The occurrence of this allele would appear to support the link between the populations.

All the other mutations were found at frequencies less than 3% and may be considered as rare alleles. Fifteen of these alleles [-28 (A→G); CD5 (-CT); CD6 (A→T) HbS; CD8/9 (+G); CD15 (G→A); CD15 (-T); CD16 (-C); CD30 (G→C); IVSI-1 (G→T); IVS1-130 (G→C); CD41/42 (-TCTT); IVSII-1 (G→A); CD121 (GAA→CAA) HbD; 25bp deletion and the 619bp deletion] have been reported in populations from the subcontinent of India (Table 4.4). In fact, the nine most common mutations found in India have also been found in Sri Lanka demonstrating their close genetic links. In contrast, only two of the eight mutations described in Portugal (about 30% of the alleles shown in Table 4.4) have been found in Sri Lanka.

Three previously reported mutations were found on the island that had not been described in India; two consensus splice site and one polyadenylation signal mutation. The first splice site mutation, IVS1-130 (G→A), was located just before the second exon and had been reported in Japanese, Egyptian and Turkish patients (Deidda *et al.*, 1990; Öner *et al.*, 1990). The second splice mutation (IVSII-745 (C→G)) was located in the second intron prior to the third exon and had been previously reported on a patient from the Mediterranean region (Orkin *et al.*, 1982b). The polyadenylation signal mutation (Poly A tail (AATAAA → AACAAA)) was previously reported in Turkish and Black patients (Orkin *et al.*, 1985). Two of these mutations (IVS1-130 (G→A) and Poly A tail (AATAAA → AACAAA)) were also found in individuals originating from India within the NHRL (personal communication from Dr J. Old). These rare mutations may indicate a connection between North Africa and Sri Lanka.

The observation that the same molecular lesions observed in Sri Lanka have already been reported in many other populations, both near and

far, prompted further investigation of these genes in the context of their surrounding DNA (Chapter 5). The primary structure of DNA is not static and the patterns observed form a record that can infer the age of the alterations and the likely relationship between individuals or populations. Haplotype analysis could allow an assessment of the connection between populations to demonstrate possible close genetic ties or to implicate that the same mutations have arisen on more than one occasion.

Notably, after an extensive examination of patients, all molecular lesions were accounted for in the Kurunegala Teaching Hospital. Many studies reveal 1-3% of individuals in a population screen, with unexplained causes for the β-thalassaemia. β-thalassaemia that does not involve a mutation in or around the β-gene is incredibly rare and every study may require lengthy investigation to unravel the molecular lesions involved.

Chapter 5 Haplotype Analysis

The aim was to compare haplotype arrangements at more than one locus within the population of Sri Lanka and then to compare these against populations of other countries. These DNA arrangements may be linked with a phenotype and used as markers for a disease assessment. Comparison of genetic material between populations may help to track the flow of genes or polymorphisms and possibly infer their age.

5.1 α-globin haplotypes

Normal chromosomes

A selection of patients and their parents from Anuradhapura were chosen with a normal α-genotype (αα/αα) to ascertain the common α-cluster haplotypes in the population. The same nine polymorphisms were examined as outlined earlier (See section 1.2.2.2). Eighteen families were studied but only 14 families yielded informative haplotype data (Table 5.1). The genotypes at the haplotype loci were obtained by PCR analysis, mostly by RFLP (See section 2.4).

Table 5.1 Distribution of normal α-globin gene haplotypes in Sri Lanka

Haplotype	RFLP*	Sri Lanka	
		n	%
Ia	++ -M PZ ++ - -	6	43
IIi	- ++M PZ + - - -	1	7.1
IIIa	- - +M Z - - - -	2	14.3
IIIb	+ -+M Z - - - -	3	21.4
IIIi	-++ M PZ - - - -	1	7.1
VIIa	- +- M PZ+ - - -	1	7.1
Chromosome Number		14	100

* A + denotes the presence of a restriction enzyme site and a – its absence. The inter-ζ fragment could be assessed as small (s) medium (m) or large depending on the size of the amplicon and the ζ (Z) or pseudo-ζ (PZ) by the size of that amplicon.

Three haplotypes (Ia; IIIa and IIIb) accounted for 78.4% of the chromosomes studied whilst a further three rare haplotypes accounted for the remaining chromosomes. Of the nine sites chosen to construct the α-globin haplotype three sites showed no variation. A greater number of alleles in all the island population ethnic groups need to be typed to be sure of the distribution and identify less common haplotypes.

5.2 β-cluster haplotypes

Normal chromosomes

Seven polymorphisms as outlined earlier (See section 1.2.2.2) were investigated in the parents and relatives of transfusion dependent patients to determine the normal range of haplotypes found on the island. The polymorphisms were analysed by RFLP-PCR analysis using the conditions described (See section 2.4). Indeed, family studies helped to assign each polymorphism to a particular haplotype but many remained uninformative. 118 normal chromosomes were successfully haplotyped and their frequencies shown in Table 5.2. The haplotypes are given using the usual Orkin scheme (Orkin & Kazazian, 1984) and also the Shimuzu designations (Shimuzu *et al.*, 2001) as these had a much greater number of different haplotype arrangements.

In Sri Lanka, 23 different haplotypes were identified. Only two haplotypes (1 and 3) were observed at frequencies above 10% yet accounted for less than 50% of the chromosomes typed. The remainder appeared to be present in a small number of individuals and represented rare alleles in the population.

Table 5.2 Distribution of normal β-globin gene haplotypes in Sri Lanka

Haplotypes			n	Frequency (%)
Orkin	Shimuzu	RFLP*		
I	1	+ - - - - + +	22	18.6
V	2	+ - - - - + -	9	7.6
VII	3	+ - - - - - +	31	26.3
IX	5	- + - + + + +	5	4.3
III	6	- + - + + + -	3	2.5
IV	7	- + - + + - +	5	4.3
II	8	- + + - + + +	4	3.4
	9	- + + - + + -	5	4.3
	10	- + + - + - +	10	8.5
	18	- - - + + + +	3	2.5
	19	- - - + + - +	1	0.8
	21	- - - - - + +	1	0.8
	26	- + + - - + +	1	0.8
	38	+ - - - + - +	3	2.5
	41	+ + - + + + -	1	0.8
	43	+ + - - - + +	2	1.7
	44	+ + - - - - +	2	1.7
Other		- - + - + - +	3	2.5
		- - + - + + +	3	2.5
		- - - - + - +	1	0.8
		- + + + + - +	1	0.8
		+ + - + + - +	1	0.8
		+ + - - - + -	1	0.8
Chromosomes studied			118	100

* A + denotes the presence of a restriction enzyme site and a – its absence (See 1.2.2.2).

β-thalassaemia associated chromosomes

The haplotypes were ascertained in 401 of the hospital transfusion dependent patients after mutation identification. Less than half the chromosomes analysed (156 patients) gave informative haplotypes (Table 5.3). Sixteen different β-thalassaemia mutations were all found at low frequencies on 10 different haplotypes. Over half (63.5%) of the assigned haplotypes were found on the most common background haplotype (VII or 3) due largely to the two most common β-thalassaemia mutations (IVSI-1 (G→A) and IVSI-5 (G→C)) but also with IVSI-130 (G→C) and 619bp

deletion. A further 3.8% haplotypes were found on the second most common haplotype and was associated with 4 different β-thalassaemia mutations (CD8/9 (+G), CD15 (-T), IVSI-5 (G→C) and IVSI-130 (G→C)).

Table 5.3 Haplotypes associated with different β-globin mutations

β-Genotype	Haplotype Designation										
Shimizu groups	1	8	6	7	2	28	3	29	5	25	Other
Orkin groups	I	II	III	IV	V	VI	VII	VIII	IX	X	
Anuradhapura											
HbS					1						
CD8/9(+G)	2										
CD15(-T)	2										
HbE			3		4						
CD30(G→C)									3		
IVSI-1(G→A)							21				
IVSI-5(G→C)	1			3			67			10	12
IVSI-130(G→A)	1				1						
IVSI-130(G→C)							1				
Poly A tail (AATAAA→AACAAA)											1
619bp Deletion							1				
Badulla											
CD15(-T)	4										
HbE					3						
IVSI-5(G→C)				2			42			6	
CD55(-A)			2								
Kurunegala											
HbS											2
CD15(G→A)				1							
CD15(-T)	2										
CD16(-C)					2						
HbE			9		21						7
IVSI-1(G→A)							35				
IVSI-5(G→C)				7			30				
IVSI-129(A→C)				1							
IVSI-130(G→C)							1				
CD41/42(-TCTT)				1							
Mutation Total	12	0	14	15	32	0	198	0	3	16	22
Frequency (%)	3.8	0	4.5	4.7	10.3	0	63.5	0	1	5.1	7.1

11 of the 16 mutations assigned haplotypes were on more unusual backgrounds (HbS, CD15 (G-A), CD16 (-C), HbE, CD30 (G→C), IVSI-5 (G→C), IVSI-129 (A→C), IVSI-130 (G→A), CD41/42 (-TCTT), CD55 (-

A), and Poly A Tail (AATAAA→AACAAA)). Twelve mutations were associated with a single type of haplotype (HbS, CD8/9 (+G), CD15 (-T), CD15 (G-A), CD16 (-C), CD30 (G→C), IVSI-1 (G→A), IVSI-129 (A→C), CD41/42 (-TCTT), CD55 (-A), Poly A tail (AATAAA→AACAAA) and 619 bp deletion). However, six of these mutations (HbS, CD15 (G-A), IVSI-129 (A→C), CD41/42 (-TCTT), Poly A tail (AATAAA→AACAAA) and 619 bp deletion) only had one informative chromosome.

Seven different thalassaemia mutations (CD8/9 (+G), CD15 (-T), CD30 (G→C), IVSI-1 (G→A), IVSI-130 (G→C), CD55 (-A) and CD16 (-C)) were each linked to their own distinct haplotype. Four mutations (HbS, HbE, IVSI-5 (G→C) and IVSI-130 (G→A)) were found on more than one haplotype but in each case had a common and rare form. The common mutations were observed on more than one haplotype except for IVSI-1 (G→A) that was found associated with a single haplotype on 56 alleles. No major regional variations were observed between the centres of Anuradhapura, Badulla and Kurunegala.

Table 5.4 **Unusual haplotypes associated with β-globin mutations**

Haplotypes (RFLP)			n
Shimuzu	Orkin		
Poly A tail (AATAAA→AACAAA)			
44	*	+ + - - - - +	1
HbE			
9	X	- + + - + - +	6
New		- - - + + + -	1

* Not reported by Orkin & Kazazian, 1984

Seven mutations were found on very rare haplotypes (Table 5.4). One haplotype was associated with the mutation Poly A tail

(AATAAA→AACAAA) in a single individual and HbE was found on four different haplotypes that included a new arrangement along the β-cluster in a single individual.

Further investigations were undertaken to examine the different possible origins of the individuals. Haplotypes were determined in individuals that were carriers for HbE originating from the subcontinent of India. Fifty families that had been referred to the National Haemoglobinopathy Reference Service for mutation identification to distinguish between HbE/E and HbE/β-thalassaemia were included but less than half revealed informative haplotypes (Table 5.5).

Table 5.5 HbE haplotype associations in populations from the Subcontinent of India

Haplotypes (RFLP)		Frequency		Origin
		n	%	
1	+ - - - - + +	1	4.8	Bengal
2	+ - - - - + -	12	56.8	Bangladesh/Bengal
3	+ - - - - - +	2	9.6	Bangladesh/Bengal
5	- + - + + + +	1	4.8	Bengal
14	- + - + + - -	1	4.8	Bengal
37	+ - - - + + +	2	9.6	Bengal/Natal
New	- + + + - + -	2	9.6	Pakistan
Total		21	100	

Surprisingly, in only 21 informative alleles seven different haplotype were revealed. Over half (56.8%) of the alleles determined were found associated with the most common haplotype (Shimuzu 2 or Orkin V). The rest were fairly evenly distributed among 5 haplotypes (Shimuzu 1, 3, 5, 14 and 37 or Orkin I, VII, IX, not described and not described) with one new haplotype in two individuals from Pakistan that were unrelated. None of these haplotypes were found in Sri Lanka.

3'β-haplotype

The β-cluster on chromosome 11 spans over 70 kilobases and includes a region that has a high mutation rate for chromosomal rearrangement. This hotspot is located downstream of the φβ-gene and upstream of the β-gene. Hence, the β-cluster haplotype can be analysed according to a 5' and 3' section either side of the 'hotspot'. One of the two polymorphisms in the 3' section is in the second intron of the β-gene and the second is downstream of the β-gene. Therefore, mutations in the β-gene tend to be tightly linked to the 3' haplotype (Table 5.6). This association may further indicate the genetic relationship with other populations.

Not all the permutations were observed as no alleles were revealed with a -/- configuration. The most common configuration was -/+ which was identified in over 50% of the normal chromosomes. Unsurprisingly, over 80% of the mutations overall were found on this background. The remaining mutations were found split between the configurations +/+ and +/-. Anaradhapura and Badulla appeared very similar but patients from Kurunegala had a great preponderance for the +/- configuration. This contrasted quite dramatically with the normal chromosomes investigated from Anuradhapura that had a far greater proportion with the +/+ configuration.

Two mutations were found on more than one 3' haplotype. IVSI-5 (G→C) was found almost exclusively on the -/+ configuration but in a single individual was found in the +/+ configuration. The mutation IVSI-130 (G→A) was only seen in two individuals and was found in two different configurations +/+ and +/-.

Table 5.6 Association between the 3' β-haplotype and β-globin sequence

β-genotype	3'gene haplotype (*Ava II* and *Hinf I*)			
	+/+	+/-	-/+	-/-
Anuradhapura				
CD8/9(+G)	2			
CD15(-T)	4			
HbE		7		
CD30(G-C)	3			
IVSI-1(G-A)			36	
IVSI-5(G-C)	1		129	
IVSI-130(G-A)	1	1		
IVSI-130(G-C)			1	
Poly A tail (AATAAA-AACAAA)			2	
619bp Deletion			1	
Mutation Total	11(5.9%)	8(4.3%)	169(89.8)	0
Normal	30(37.5%)	8(10%)	42(52.5%)	0
Badulla				
CD15(-T)	4			
HbE		3		
IVSI-5(G-C)			50	
CD55(-A)		2		
Mutation Total	4(6.8%)	5(8.5%)	50(84.7%)	0
Kurunegala				
CD15(G-A)			1	
CD15(-T)	2			
CD16(-C)		2		
HbE		37		
IVSI-1(G-A)			35	
IVSI-5(G-C)			37	
IVSI-129(A-C)			1	
IVSI-130(G-C)			1	
CD41/42(TCTT)			1	
Mutation Total	2(1.7%)	39(33.3%)	76(65%)	0
Overall Mutation Total	17(4.7%)	52(14.3%)	295(81%)	0
Overall Total	47(10.6%)	60(13.5%)	337(75.9%)	0

Xmn-I $^{G}\gamma$-gene polymorphism

The *Xmn-I* polymorphism is linked to HbF levels and may be under greater selective control. The (+) configuration is associated with elevated

Table 5.7 **Frequency of the *Xmn I* polymorphism associated with different β-globin genotypes**

β-genotype	*Xmn-I* status	n	Frequency (%)
Anuradhapura			
β-thalassaemia major	+/+	1	1.4
	+/-	11	15.3
	-/-	60	83.3
HbE/ β-thalassaemia	+/+	0	0
	+/-	12	85.7
	-/-	2	14.3
β Tr	+/+	1	0.9
	+/-	25	23.4
	-/-	81	75.7
HbA/E	+/+	3	37.5
	+/-	4	50
	-/-	1	12.5
Badulla			
β-thalassaemia major	+/+	1	2
	+/-	12	24.5
	-/-	36	73.5
HbE/ β-thalassaemia	+/+	0	0
	+/-	7	100
	-/-	0	0
Colombo			
β-thalassaemia major	+/+	4	18.75
	+/-	2	6.25
	-/-	18	75
HbE/ β-thalassaemia	+/+	1	100
	+/-	0	0
	-/-	0	0
Kurunegala			
β-thalassaemia major	+/+	3	2.3
	+/-	25	19.2
	-/-	102	78.5
HbE/ β-thalassaemia	+/+	10	10
	+/-	70	70
	-/-	20	20
Total			
β-thalassaemia major	+/+	9	3.3
	+/-	50	18.2
	-/-	216	78.5
HbE/ β-thalassaemia	+/+	11	9
	+/-	89	73
	-/-	22	18

levels of HbF. The locus was examined by RFLP analysis (See 2.4) in a large proportion of the patients and relatives from four centres on the island: Anuradhapura; Badulla; Colombo and Kurunegala. An assessment was then made to ascertain if there was an association between the *Xmn-I* status with different haemoglobinopathy groups (Table 5.7).

The *Xmn-I* (-) configuration was the most prevalent found on 78.1% of the chromosomes. This high frequency should mean that every β-globin gene grouping should have a high preponderance for the (-/-) configuration. This was the position in patients with β-thalassaemia major (78.5%) and relatives with β-thalassaemia trait (75.7%). However, patients with HbE/β-thalassaemia and relatives with HbA/E had a larger proportion with the (+/-)(73%) and the (+/+)(18%) status. This turn around may result in a greater level of HbF in patients with chromosomes containing HbE.

Three β-globin gene mutations were also found in associated with an *Xmn-I* (+) configuration. CD30 (G→C) and CD15 (G→A) are rare mutations in Sri Lanka and hence could only be demonstrated in a few families. The (+) configuration was found in a small subset of patients with the β-globin mutation IVSI-5 (G→C) but the majority of these patients were associated with the (-) configuration.

Although only a few cases were informative only four β-globin cluster haplotypes have been found with the (+) status. CD30 (G→C) was found on Orkin haplotype IX, CD15 (G→A) on haplotype IV and HbE on haplotypes III and V. Three of these Orkin haplotypes (III, IV and IX) share the same 5' haplotype (-+-++), upstream of the hotspot for rearrangement, and were found in about 10% of the normal chromosomes described earlier (Table 5.2). The 5' haplotype and not just the *Xmn-I*

polymorphism still holds the key to dissect its association with elevated HbF levels but the *Xmn-I*.

BP-1 β-gene promoter polymorphism

The aim of this study was to investigate the heterogeneity of a repeated purine-pyrimidine sequence motif approximately 500bp upstream of the β-gene. This marker may be a possible indicator for disease severity by regulation of the level of HbF.

Hospital based patients and their relatives from Kurunegala were investigated by PCR sequence analysis (See section 2.4) using the amplification primers 5'βURR against 3'βURR and sequencing with primers 5'βURR1 and 3'βURR1 outlined in Table 2.9. The different isoforms were assessed and their frequency determined (Table 5.8).

Table 5.8 Incidence of different BP-1 motif observed in individuals from Sri Lanka

Repeat configuration	Frequency	
	n	%
$(AT)_6(T)_9$	1	1.2
$(AT)_7(T)_7$	25	29.4
$(AT)_8(T)_5$	13	15.3
$(AT)_8(T)_6$	1	1.2
$(AT)_8(T)_7$	5	5.9
$(AT)_8(T)_9$	2	2.4
$(AT)_9(T)_3$	16	18.8
$(AT)_9(T)_5$	13	15.3
$(AT)_9(T)_7$	7	8.1
$(AT)_{11}(T)_1$	1	1.2
$(AT)_{11}(T)_3$	1	1.2
Total	85	100

Amazingly, eleven different arrangements of this motif were observed with four common types accounting for 78.8% of the chromosomes studied but the pattern of the motifs did not appear to follow

a normal distribution pattern around a hypothetical average pattern $(AT)_8(T)_5$). Also a subset of these individuals (15) examined, were part of a group with HbE/β-thalassaemia that had been studied more intensively. An extremely crude attempt to correlate the 5'β URR motif with the level of HbF (total haemoglobin concentration or relative proportion) was made on these patients but no correlation was observed.

5.3 Discussion

Phylogenetic analysis indicates that α- and β-globin genes were formed by duplication of a common ancestor ~450 Mya (Burmester *et al.*, 2002). The process of duplication and divergence of the α- and β-genes has continued to generate the current gene clusters. Hence, a large degree of sequence homology can be observed. The genes and the control mechanism for globin expression are found in the same overall arrangement in mammalian species. The β-globin gene cluster extends over 70 kilobases and the genes embedded in this sequence are highly conserved between individuals and across species. The embryonic and adult genes are most highly conserved across species as seen in Figure 5.1.

Regions around genes and upstream in the LCR are essential for regulation of gene expression and are also conserved. However, the more closely sequence throughout the β globin cluster is examined in different populations the greater the degree of sequence heterogeneity. This is not greatly surprising as these gene duplication events took place millions of

years ago and hence has allowed a large expanse of time for change to occur. Any polymorphism that has different configurations in different groups of people can be useful and many have been proposed as important in the regulation of the globins process (Labie & Elion, 1996; Zago *et al.*, 2000). Patients that are able to maintain a relatively normal concentration of haemoglobin by maintaining high levels of HbF may have a milder clinical phenotype (Weatherall, 2001). The *Xmn-I* polymorphism (+) configuration has been associated with higher HbF levels and can be used as a prognostic marker. However, the polymorphism has never been proven to be directly involved in the regulatory process (Labie & Elion, 1996).

The majority of chromosomes in the population of Sri Lanka are on the *Xmn-I* (-) configuration similar to other populations (Thein *et al.*, 1988; Lie-Injo *et al.*, 1989). This is somewhat surprising, as selective advantage may have predisposed towards allowing the rescue of individuals from anaemia by switching to HbF production. This may suggest that conditions for anaemia may have been of prime importance. An attempt was made to eradicate malaria on the island in the 1960s that may have changed the selective advantage away from anaemia. The genetic outcome may only become apparent after many more generations.

If this trait has a great selective advantage its incidence and consequently that of the linked β-gene mutation HbE should expand to a larger part of the population increasing the prevalence of HbE/β-thalassaemia. The predominant haplotype shared the same 5' portion with a (-+-++) configuration. Part of this 5' haplotype is shared with the Asian Indian haplotype associated with higher levels of HbF in patients with sickle cell disease (++-++) (Wood, 1993). The γ-genes and their surrounding DNA

need to be more closely scrutinized in individuals with a high and low HbF level. The *Xmn-I* polymorphism may be linked to another mutation that could be responsible for the regulation of HbF expression. The HbS Asian Indian haplotype is also associated with a longer version ($(AT)_9T_5$) of the simple repeat element upstream of the β-gene that binds to a nuclear binding protein (BP-1)(Labie and Elion , 1996).

The connection between Sri Lanka and the rest of the world

Previously described polymorphisms on both the α- and β-globin clusters were analysed and the normal haplotypes were compared against other world populations (Table 5.9 and 5.10).

Table 5.9 Comparison of normal α-globin gene haplotypes in world populations

Haplotype	RFLP	Mediterranean	S.E. Asian	Asian Indian		Sri Lanka		
		Ref. 1	Ref. 1	Ref. 1	Ref. 2	Ref. 1	Present study	Combined
		%	%	%	%	%	%	%
Ia	++ -M PZ ++ - -	56	28.6	54.2	41.1	25	43	31
Ib	- + -M PZ ++ - -	6	1.9	0	2.2	3.6	0	2.4
IIa	- + - L PZ + - - -	20	24.8	16.7	23.2	35.7	0	23.7
IIi	- ++M PZ + - - -	0	0	0	0	0	7.1	2.4
IIIa	- - +M Z - - - -	6	8.3	6.3	6.6	21.4	14.3	19
IIIb	+ -+M Z - - - -	0	0	2.1	1.3	3.6	21.4	9.5
IIIc	- + - L Z - - - -	4	1.1	2.1	0	3.6	0	2.4
IIIh	+++M Z - - - -	0	1.1	0	0	7.1	0	4.8
IIIi	-++ M PZ - - - -	0	0	0	0	0	7.1	2.4
VIIa	- +- M PZ+ - - -	0	0	0	0	0	7.1	2.4
Other		8	34.2	18.6	25.6	0	0	0
Chromosome Number		50	266	48	151	28	14	42

References. 1. Flint *et al.*, (1993) 2.Martinson *et al.*, (1995)

A complex pattern for the global distribution of α-globin haplotypes was observed (Table 5.9). Two α-globin haplotypes (Ia and IIa) were common to all populations reported and were generally found at high incidence levels. Three other α-globin haplotypes (Ib, IIIa and IIIc) were

widely distributed but generally at a low incidence. The population of Sri Lanka demonstrated a different distribution pattern of haplotypes compared to other populations. α-globin haplotype IIIa was found at a high frequency in Sri Lanka but had a higher incidence in Asian populations. α-globin haplotype IIIb had a high incidence in Sri Lanka and has only been reported in Asian Indians, at a low frequency. α-globin haplotypes IIii and IIIi were only found on the island. The other α-globin haplotype only reported in Sri Lanka was VIIa but has been previously described in people of West African origin (Flint *et al.*, 1993). A connection between these groups is not postulated.

The results from the previous study on individuals from Sri Lanka and the current study were widely different. In part, this was due to the low number of haplotypes assessed. However, some large differences still need to be explained and probably reflect groups of different ethnic origin or smaller isolated communities.

The global distribution pattern of β-globin haplotypes is just as complex (Table 5.10). Seven β-globin haplotypes were common to all populations studied. Three (1, 2 and 5) were generally found at high and four (3, 5, 6 and 8) at a low level of incidence. β-globin haplotype 3 was found at a higher incidence in populations from the East and haplotype 5 was higher in populations from the West. The population of Sri Lanka revealed its own pattern of haplotypes. β-globin haplotypes 18 and 38 were only revealed in people from Sri Lanka. However, a larger sample size is required in all the populations to be more confident about these population differences.

Table 5.10 Distribution of normal β-globin gene haplotypes in world populations

Populations										
Haplotypes			Cyprus*	Algeria*	Asian Indian*	Thailand*	Cambodia*	South China*	Japan*	Sri Lanka
*	**	RFLP								
1	I	+----++	50	43.1	25	16.9	12.3	20.4	13.9	18.6
2	V	+----+-	10.5	15.4	9	49.2	26.2	24.2	18.4	7.6
3	VII	+----+	5.3		19.1	24.6	41.5	35.1	41.1	26.3
4		+-----						1		
5	IX	-+-++++	26.3	13.9	13.9	1.5	6.2	2.4	5.1	4.2
6	III	-+-+++-		9.2	1.5	1.5	3.1	1.4	1.3	2.5
7	IV	-+-++-+			6.8		4.6	4.7	5.1	4.2
8	II	-++-+++	2.6	9.2	9.9		1.5	2.8	3.2	3.4
9		-++-++-			2.5	3.1		1	1.9	4.2
10	B3	-++-+-+			5.6	1.5	1.5	4.3	1.3	8.5
11		-+--+++			0.3				0.6	
12		-+--++-			0.3					
13	i	-+--+-+			0.3	1.5	3.1		0.6	
14		-+-++--							0.6	
16		----++-		4.6						
18		---++++								2.5
19		---++-+						1.9	1.9	0.8
21	B4	-----++		1.5	0.3					0.8
23		------+							0.6	
24		-+---++							1.3	
26		-++--++			0.9					0.8
27		-++--+-			0.6					
28	VI	-++---+	5.3	3.1	0.3					
29	VIII	-+-+-+-							0.6	
35		+--++-+						1		
38		+---+-+								2.5
39		+-+---+							1.3	
40		++-++++			1.2					
41		++-+++-			1.2					0.8
43		++---++			0.6					1.7
44		++----+							1.3	1.7
45		+++-+++			0.6					
		Other								8.5
Chromosomes studied			38	65	324	65	65	211	158	118

Reference: * Shimizu *et al.*, (2001) ** Orkin & Kazazian (1984)

At both the α- and β-globin cluster the incidence of different haplotypes in Sri Lanka was more similar to other populations of the Far East than in populations around the Mediterranean. The people of Sri Lanka and those from the subcontinent of India appear to share a close genetic history. This ties in closely with history of the country (See section 1.1.1) and the population migrations that have been recorded over thousands of years, mainly from neighbouring India.

Origin of mutations

Haplotype analysis was found to be useful to study β-globin haplotypes linked with β-thalassaemia mutations. Unsurprisingly, 77.6% of the assigned haplotypes were found on the three common background haplotypes (Shimizu groups 1, 2 and 3; Orkin I, V and VII). Previously reported haplotype mutation associations from other populations were compared against those assigned in the population of Sri Lanka (Table 5.11).

Table 5.11 A comparison of haplotype associated with different beta globin mutations in different populations of the world

β-genotype	Sri Lanka	Mediterranean (Ref 1, 13)	Middle East (Ref 3, 4, 5)	Indian Subcontinent (Ref 1, 2, 13)	South East Asia (Ref 1, 6, 7, 8, 9, 10, 11, 12, 13)
CD6(GAG→GTG) HbS	**2**, 41			**2**, 41	
CD8/9 (+G)	1		1, **7**	**1**, 5, 8	
CD15 (G→A)	7		8	1, 5, 7, 8	
CD16 (-C)	2			**2**	
CD26(G→A) HbE	2, 6				1, 2, 3 , **5**, 6, 7, 8, 9, 12
CD30 (G→C)	5		1		1
IVSI-1 (G→A)	3	1, **2**, 5, 6, 21	**2**		
IVSI-5 (G→C)	1, 3, 7		1, **3**	1, **3**, 7, 10, 13, +	1, **3**, 5, 21, 37
CD41/42 (-TCTT)	7			1 , 5, 6, **2**, 41	**1**, 3
619 bp Deletion	3			**3**	

1 Orkin & Kazazian, (1984)
2 Thein *et al.* , (1988)
3 Old J.M. Personal communication
4 Filon *et al.* , (1994)
5 Novelletto *et al.* , (1990)
6 Lie Injo *et al.*, (1989)
7 Chan *et al.* , (1987)
8 Antonarikis *et al.* , (1982)
9 Yongvanit *et al.* , (1989)
10 Brown *et al.* , (1992)
11 Hundreiser *et al.* , (1988)
12 Fuchaeron *et al.* , (1990)
13 Kazazian *et al.* , (1984)

Six of these mutations (HbS, CD8/9 (+G), CD15 (G→A), CD16 (-C), IVSI-5 (G→C) and the 619 bp Δ) showed the same haplotypes as those observed in India. The simplest connection would be due to a population migration from India but they also occur on common background haplotype so a multicentric origin for this mutation, although unlikely, remains a possibility.

Four mutations showed an association with a haplotype not described on the Indian subcontinent. The relatively rare mutations CD30 (G→C) and CD41/42(-TCTT) were found on Shimizu haplotypes 5 and 7 respectively. Both of these haplotypes constituted less than 5% of the normal background haplotype and have not been reported with these haplotypes in any other population. However, the 3' β-globin haplotype associated with the mutation CD30 (G→C) is always ++. Any change upstream of the recombination hotspot would generate other haplotypes but may still be consistent with a common origin for the mutation. In contrast, the 3' β-globin haplotype associated with the mutation CD41/42(-TCTT) was found in Sri Lanka on a -+ configuration. The most commonly reported association for this mutation was on the +- configuration but it has also been reported on a ++ configuration and is suggestive that this mutation may have arisen independently in the island population. Three other different thalassaemia mutations (CD15 (-T), IVSI-130 (G→C) and CD55 (-A)) were each linked to their own distinct haplotype. Information on their haplotype associations in other populations has not been reported.

The three most common mutations on the island also yielded the largest number of informative haplotypes. The single base substitution IVSI-5 (G→C) is the most prevalent mutation, in Sri Lanka and globally, and is probably the oldest β-thalassaemia mutation (reviewed in Weatherall & Clegg, 2001). The mutation was found on four different background haplotypes (Shimizu 1, 3, 7 and 25; Orkin I, VII, IV and X respectively) that possibly denotes that the mutation has been in existence long enough for chromosomal rearrangements to have occurred. Closer examination of the haplotype showed that the 3' haplotype was almost exclusively due to one

from -/+ with a single individual showing +/+. This would strongly indicate that the mutation has probably only arisen once a long time ago and that different 5' haplotypes have occurred due to the hotspot for gene arrangements just upstream of the β-gene.

IVSI-1 (G→A) has been reported as an extremely rare mutation in north India (Varawalla *et al.*, 1991b). The mutation is common in countries surrounding the Mediterranean reaching very high frequencies in Spain and Portugal. The mutation observed in Sri Lanka was linked exclusively to Shimizu haplotype 3. In the Mediterranean populations the mutation was linked to haplotype 1, 2, 5, 6 and 21 in Portugal and North Africa (Faustino *et al.*, 1992; Pachecho *et al.*, 1996) but never to haplotype 3. In particular, the 3' haplotypes from these distinct regions are informative. In Sri Lanka a -/+ arrangement is found compared to either a +/+ or +/- arrangement around the Mediterranean. The chances of some form of recombination are exceptionally slim especially from the +/- configuration as two events need to occur to achieve the change and an intermediate form has not been described. Another possibility is that the mutation has occurred in a population that has not been investigated thoroughly, such as south India. The population of South India may still hold the key to this riddle as it has only been investigated on a very small scale for mutations that have been described in India and this does not include IVSI-1 (G→A). In conclusion, the mutation appears to have arisen at least twice in a population in the west and in Sri Lanka.

The origin of the third most common mutation, HbE, found on the island is of great interest as it superficially gives credence to the notion of a population migration from north east India down to Sri Lanka without

passing through south India. In the population of Sri Lanka HbE was found on four haplotypes and was compared to other populations (Table 5.12) (Antonarkis *et al.*, 1982; Nakatsuji *et al.*, 1986; Hundreiser *et al.*, 1988; Yongvanit *et al.*, 1989; Fucharoen *et al.*, 1990; Brown *et al.*, 1992 and Fucharoen *et al.*, 1997).

Table 5.12 Haplotypes associated with HbE in different populations

Haplotype (Shimuzu/RFLP)	Sri Lanka	Northeast India / Bangladesh	Myanmar	Thailand	Cambodia	Vietnam	Overall
	%	%	%	%	%	%	%
1 +-----++	0	5.3	1.1	0	0	17	1.2
2 +-----+-	25.5	63.2	15.4	27.7	26.1	17	31.5
3 +------+	0	10.5	3.3	0	0	0	2.0
5 -+-++++	0	5.3	3.3	0	0	0	1.6
6 -+-++++-	59.6	0	73.6	63.1	47.8	49	53.4
7 -+-+++-	0	0	0	6.2	26.1	0	4.0
8 -+++++	0	0	0	0	0	17	0.4
9 -+++++-	12.8	0	0	0	0	0	2.4
12 -+--++-	0	0	0	1.5	0	0	0.4
Others	2.1	15.7	3.3	1.5	0	0	3.1
Chromosome No.	47	19	91	65	23	6	251

Surprisingly, the β-globin haplotypes observed in individuals originating from north east India did not include Shimizu haplotype 6 and was probably due to the low number of chromosomes examined. Haplotype 6 was the most common haplotype associated with HbE in all the other populations. Haplotype 2 also was found at high frequencies in all the populations. Haplotype 9 was only observed in Sri Lanka and represents a divergence from these other populations. The origin of HbE is not clear but may only have occurred as a single event around the Thai/Cambodia/Vietnam border. All the other haplotypes could have derived from this founder.

In Sri Lanka the 3' β-globin haplotype is always in the +/- configuration in association with HbE. Only in Cambodia are people with a

different configuration -/+ found at high frequencies. The flow of genes between populations can be strongly implicated by these studies but many more chromosomes need to be defined so that the results reach statistical significance.

Extended haplotype

The repeated purine-pyrimidine sequence (AT)n(T)n 0.5kb upstream of the β-globin gene exhibits a great variation in length and configuration (Galanello *et al.*, 1993). This motif lies within a negative regulatory region between –610 and –490 upstream from the β-globin gene and is the binding site for a putative negative regulatory transacting factor called BP-1. Anecdotal evidence exists that BP-1 binds more avidly to the longer AT tract compared to the shorter AT tract resulting in a very mild β-thalassaemia phenotype with the longer AT isoform.

Eleven different forms of this motif were observed with four major types accounting for almost 80% of the alleles observed. These were spread over 7, 8, 9 and 11 repeats of AT which should correlate with the strong and weak binding forms. The motif would probably warrant a closer investigation to see if it acts as a gene modifier in the characterised group of patients with HbE/β-thalassaemia. A possible correlation between these isoforms and levels of HbF were not supported by our limited results.

Which haplotype?

This also opens up the question about the polymorphisms that should be included as part of a haplotype analysis. The polymorphisms that define a particular haplotype should form a series of variable markers such that the different combinations exhibit patterns that can be informative. The markers

used to define globin haplotypes were selected during the early developmental phase of DNA analysis. The polymorphisms were chosen because they mainly alter restriction enzyme sites and were reasonably spaced over the α- and β-globin clusters. Perhaps now with the aid of automation many more sites should be included to define a haplotype that include those that do not just alter restriction enzyme sites. Sequence analysis may reveal many nucleotides within a stretch of DNA that show some diversity based on both anthropological and phylogenetic diversity. Limited sequence analysis, in this current study, of both the duplicated α- and γ-genes has shown that sequence variation and putative intergene nucleotide exchange can form subsets of a haplotype. These may have some relevance to expression of that gene but could also be used to follow the flow of particular genes through populations.

Chapter 6 Genotype – Phenotype correlation

The subtly different clinical outcomes of patients with apparently identical globin genotypes depend on the mechanism of globin gene disruption and the different genes affected. Many factors modify the phenotype, either in the gene directly or as a distant regulator. A rigorous clinical examination of these patients was made to identify these factors and to highlight other anomalies that could account for a patient's phenotype.

6.1 β-globin gene mutation

Mutations within the β-gene down regulate the expression of β-globins in more than one way. The mutations described in chapter 4 either have been previously defined (reviewed in Weatherall and Clegg, 2001) or were placed within one of these groups (Table 6.1). β^0-thalassaemia alleles included CD 5 (-CT), CD 6-10 (Δ13bp), CD 8/9 (+G), CD 15 (-T), CD 15 (G→A), CD 16 (-C), IVSI-1 (G→A), IVSI-1 (G→T), 25bp Δ, IVSI-130 (G→C), IVSI-130 (G→A), CD 41/42 (-TCTT), CD 55 (-A), IVSII-1 (G→A) and 615bp Δ. β^+ (severe)-thalassaemia alleles included CD 30 (G→C), IVSI-5 (G→C) and IVSII-745 (C→G). β^+-thalassaemia alleles included -28 (A→G) and Poly A tail (T→C). Haemoglobin variants included HbS, HbE and HbD and one mutation was classified an unknown (IVSI-129 (A→C)). The groups were outlined in section 1.2.2.2 β-thalassaemia phenotypes. Two main groups were observed; transfusion dependent β-thalassaemia major and β-thalassaemia intermedia.

Most of the β-thalassaemia major patients express either no adult β-globin or only a very small amount and require blood transfusions every 3 to 4 weeks. Clinical variations within this group of patients were not investigated. Two patients were treated as if they had β-thalassaemia major

but were found to have a β^0-thalassaemia allele in combination with a β^+-thalassaemia allele. The mutations associated with a reduced globin output were the promoter mutation -28 (A→G) in a single case and the polyadenylation signal mutation (Poly A tail (T→C)) in seven patients. These patients needed careful follow up examinations to observe whether they could have been classed as severe β-thalassaemia intermedia. Other possible candidates for a less severe form of the disease included one hundred and sixty eight patients who had one β-thalassaemia allele in combination with a structural variant; 1 with HbD, 162 with HbE and 5 with HbS.

Table 6.1 Classification of mutations according to their predicted expression

Globin expression	Number	Frequency (%)
β^0	329	27.0
β^+(Severe)	710	58.2
β^+	9	0.7
$\beta^{Variant}$	168	13.8
Unknown	3	0.3
Total	1219	100

The family with HbD was also found to have a β^0-thalassaemia (IVSI-1 (G→A)). The index case, an 18 year old boy, had presented at the local clinic in Galle with continued jaundice (7 years), hepatosplenomegaly, a moderate anaemia and an unidentified abnormal haemoglobin. No blood transfusions were recorded.

Five families were found to have the haemoglobin variant HbS. The presentation of their clinical phenotypes was incomplete. All of the index cases had required some blood transfusions but the normal range of symptoms associated with sickle cell disease were not reported.

All the other patients with structural haemoglobin variants were found to have HbE/β-thalassaemia. The β-thalassaemia allele was always a severe β^+- or β^0-allele. These patients demonstrated a wide variety of phenotypes ranging from almost undetectable carrier status to β-thalassaemia major. 109 of these patients were investigated more thoroughly to place them into different clinical subsets and then to assess the importance of different influencing factors. Four patients had mutations to which a clinical phenotype was uncertain and so remained unclassified.

6.2 α-gene arrangement

Adult haemoglobin is a tetramer composed of two α- and two β-polypeptide chains. In β-thalassaemia too many or too few α-chains can alter the outcome of the individual's phenotype (See section 1.2.2.2). This chain imbalance forms the basis of the pathophysiological outcome. Further analysis of the data from the transfusion dependent patients and their relatives were evaluated to examine the frequency of α-genotypes in different β-thalassaemia categories (Table 6.2).

Table 6.2 Incidence of α-gene arrangement coexistent in the in patients and relatives with β-globin mutations

α-genotype	Patients			Carriers	
	HbE/β-thalassaemia	HbD or HbS/ β-thalassaemia	Homozygous or compound heterozygous β-thalassaemia	HbA/E	β-trait
αα/αα	112 (90.3)	2 (100)	223 (80.2)	6 (75)	79 (78.2)
-α/αα	8 (7.3)	0 (0)	40 (13.9)	1 (12.5)	19 (18.9)
-α/-α	0 (0)	0 (0)	2 (0.8)	0 (0)	1 (1)
ααα/αα	3 (2.4)	0 (0)	12 (4.3)	1 (12.5)	2 (1.9)
ααα/ααα	0 (0)	0 (0)	1 (0.4)	0 (0)	0 (0)
αααα/αααα	0 (0)	0 (0)	1 (0.4)	0 (0)	0 (0)
Total	123 (100)	2 (100)	279 (100)	8 (100)	101 (100)

Figures in parenthesis represent genotype frequencies

The incidence of α-globin genotypes (-α/αα) was almost two times greater in β-thalassaemia major patients compared with HbE/β-

thalassaemia patients. This decreased frequency was also observed in the carriers for HbE compared with those that had β-thalassaemia trait. As the study involved transfusion dependent thalassaemia patients, the severe end of the phenotypic spectrum was mainly analysed. Patients with HbE/β-thalassaemia that had co-existing α-thalassaemia may have been modified to a milder form of the disease and hence would not have presented at clinic and subsequently required hospital treatment. Seven patients with this condition still presented at clinic with a need for blood transfusions but presented with steady state HbF levels between 2.7 and 32% (mean<10%). Although α-thalassaemia may have been of benefit in these patients their ability to produce HbF was poor and they remained anaemic. Interestingly, relatives of patients with HbE/β-thalassaemia that were carriers for HbE appeared to be associated with a high incidence in the ααα gene arrangement compared to those with β-thalassaemia trait. As these represented a smaller total number of individuals this may have been a statistical anomaly.

Eighteen patients with extra α-genes were analysed at greater depth (Table 6.3). Two individuals presented with a normal phenotype. They were both found to have the α-gene arrangement ααα/αα and a single β-gene mutation. These individuals were relatives of transfusion dependent cases and were only picked up after family studies. The first was found in combination with the haemoglobin variant, HbE that is often associated with reduced red cell indices and can be classed as a mild β-thalassaemia allele. The relative chain imbalance for this individual should not be great but should become more exaggerated with a β^0-thalassaemia mutation. The second was found in combination with such a β-globin mutation, IVSI-1 (G→A) but unaccountably presented with normal haematological indices.

Table 6.3 Genotype/phenotype correlation of individuals with extra α-genes

α-genotype	β-genotype	Phenotype				Total
		Normal	Mild	Moderate	Severe	
ααα/αα	HbA/E	1				1
	IVSI-1 (G→A)/β^N	1				1
	Codon16 (-C)/β^N			1		1
	IVSI-130 (G→C)/β^N			1		1
	IVSI-5(G→C)/HbE				3	3
	IVSI-5(G→C)/ IVSI-5(G→C)				3	3
	IVSI-5(G→C)/ IVSI-1(G→A)				1	1
	IVSI-1 (G→A)/ Codon15 (-T)				1	1
	IVSI-5(G→C)/Codon41/42(-TCTT)				3	3
	Codon15 (-T)/ Codon55 (-A)				1	1
ααα/ααα	Codon15 (-T)/β^N				1	1
αααα/αααα	IVSI-5(G→C)/β^N				1	1
Total	18	2	0	2	14	18

The individual with the genotype ααα/αα and β^0-thalassaemia trait (CD16 (-C)) presented at clinic at 3 years of age after the mother had noticed an enlarged cervical lymph node on the left-hand side. On examination the child was found to have pallor, hepatosplenomegaly and lymphoidenopathy. On investigation of the red cells the Hb level was found to be only 6.5g/dl and the patient had an elevated HbA_2 at 5.8%. A blood transfusion was given and the red cell indices returned to normal. Currently, the boy is about 11 years old and has no overt problems but has low red cell indices and slight hepatosplenomegaly as shown in Table 6.4.

The individual with 5 α-genes and the β^0-thalassaemia mutation (IVSI-130 (G→C)) presented at clinic with a history of abdominal pain aggravated by oily foods and a yellow discoloration of the eyes at the age of 46 years old. She was diagnosed as having cholecystitis due to gallstones. She was given a blood transfusion prior to cholecysectomy and splenectomy and examined haematologically. She had persistent haemolytic anaemia but sadly died some time after the operation. All the other individuals with 5 α-genes had mutations in both their β-genes and were treated as severe transfusion dependent patients.

Table 6.4 **Clinical picture of two patients with ααα/αα and a β^0-thalassaemia mutation**

Genotype	RBC ($X10^4/\mu l$)	Haemoglobin concentration [Hb](g/dl)	MCV (fl)	MCH (pg)	MCHC (g/dl)	Spleen	Liver	Jaundice
Codon16 (-C)/β^N ααα/αα	263	5.4	55.4	20.7	37.3	Slightly Enlarged	Just Palpable	No
IVSI-130 (G→C)/β^N ααα/αα	278	7.2	70.1	26	37.1	Removed	Enlarged	No

An individual homozygous for the triple α-gene in combination with the single β^0-thalassaemia mutation (Codon15 (-T)) presented as a boy with suspected β-thalassaemia major. After DNA analysis he was found to have just one β-gene mutation in combination with 6 α-genes. Upon further physical examination he was found to have pallor, jaundice, and hepatoslpenomegaly. A family study was initiated and other family members were examined (Table 6.5).

Table 6.5 **Haematology of family G with a triple α-gene arrangement**

Individual	RBC ($X10^4/\mu l$)	[Hgb] (g/dl)	Hct (%)	MCV (fl)	MCH (pg)	Spleen	Liver	Jaundice
Index case	204	4.9	14.8	72.5	23.9	Enlarged	Enlarged	Yes
Mother	342	10.8	27.5	80.4	31.4	Normal	Normal	None
Father	341	7.4	20.9	61.2	21.8	Normal	Normal	Yes
Brother 1	352	11.4	29.1	82.7	32.5	Just Palpable	Just Palpable	None
Brother 2	342	9.8	24.7	72.3	28.6	Normal	Normal	None
Sister	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.

The mother appeared normal but the father had both anaemia and jaundice. Only 2 of the three siblings were examined, brother 1 was normal and brother 2 had mild anaemia. The α- and β-genes of the whole family were analyzed. Both parents and the sister were found to be heterozygous for the triple α-gene rearrangement whereas brother 1 was normal and brother 2 was a homozygote for this gene arrangement. At the β-gene locus the mother and the two brothers were normal whereas the father and the

sister both were carriers for the same β^0-thalassaemia mutation found in the proband. Only the index case had needed the requirement for blood transfusions and was probably due to the exaggerated chain imbalance imposed by the loss of expression of one β-gene and the inheritance of an extra α-gene from both parents.

The final case was the individual homozygous for the quadruple α-gene arrangement in combination with the single severe β^+-thalassaemia mutation (IVSI-5 (G→C)) who presented as a baby with suspected β-thalassaemia major. After DNA analysis he was found to have just one β-gene mutation in combination with eight α-genes. Upon further examination he was found to be pale with a large head. A family study was initiated and his parents were examined and the clinical status evaluated as highlighted in Table 6.6.

Table 6.6 Haematology of family H with a quadruple α-gene arrangement

Individuals	RBC ($X10^4/\mu l$)	Hgb conc. (g/dl)	Hct* (%)	MCV (fl)	MCH (pg)	Spleen	Liver
Index case	236	4.9	13.8	58.4	20.8	Splenectomy	Slightly enlarged
Mother	295	6.2	16.5	56.1	20.9	Normal	Normal
Father	375	8.7	22	58.5	23.2	Normal	Normal

* Hct is an abbreviation of haematocrit

The index case was splenectomised and afterwards was found to have severe anaemia, greatly reduced red cell indices and a slightly enlarged liver. Both parents were investigated and found to have the genotype αααα/αα and IVSI-5 (G→C) trait. The parents appeared normal but had markedly reduced red cell indices and anaemia with the mother exhibiting a slightly more severe phenotype.

The haematology, on these families, was performed by a technician within the clinic at the Kurunegala teaching hospital. The clinical

examinations were performed by Professor D.J. Weatherall, Dr A. Premawardhena and Dr N. Olivieri.

6.3 HbF Level

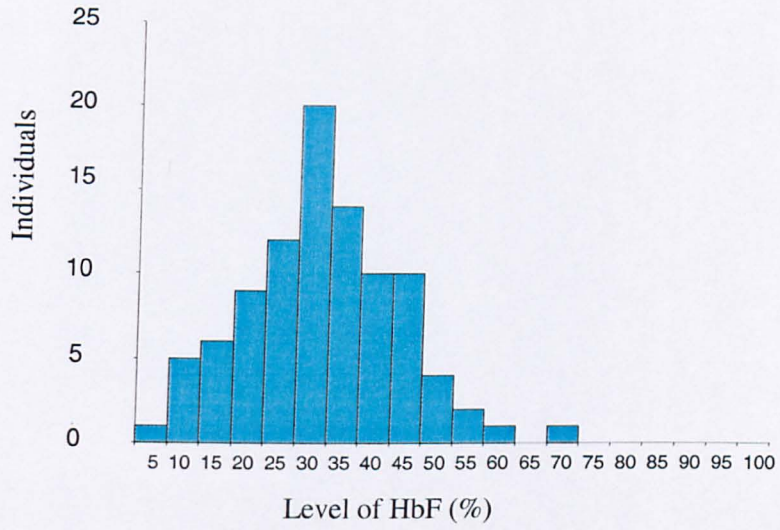
The amount of HbF in normal adults varies by more than ten-fold. Family studies of individuals at the low and high ends of this spectrum have indicated a strong genetic component involved in setting this level (Wood, 1993) Genetic and environmental factors contribute towards the final level of expression of HbF. The genetic determinants are probably located within and outside of the β-globin gene cluster.

In the absence of the normal adult β-globin chain the foetal form ($^{A}\gamma$-globin or $^{G}\gamma$-globin) can substitute to form HbF so that haemoglobin switching does not occur. The regulation of this gene is critical to the outcome of the disease severity. A closer examination was made of the sequence in and around the γ-genes to determine any association with the level of HbF. Methods and reagents were developed on a family that was referred to the NHRL that was suspected of having co-existing β- and γ-thalassaemia. These methodologies were then applied to two different cohorts originating from Sri Lanka; cord blood samples and patients with HbE/β-thalassaemia.

The clinical presentation of patients with HbE/β-thalassaemia spans from transfusion dependent to individuals who are seemingly normal and are only identified after haematological examination. The factors that govern this variation are largely unknown but elevated levels of HbF can rescue the patient from severe anaemia. Eighty five patients with this syndrome were closely investigated after at least six months without blood transfusions or just before a blood transfusion to attempt to determine the steady state haemoglobin levels. Patients with this syndrome demonstrated

a normal distribution pattern over a wide range of HbF values from under 5% to over 70% with a peak around the 30-35% level (Figure 6.1).

Figure 6.1 Distribution of HbF levels in HbE/β-thalassaemia patients



During the analysis of cord blood haemoglobins for the investigation of α-thalassaemia, the level of HbF was simultaneously evaluated using the β-thalassaemia short program. The cord blood samples were arranged in decreasing levels of HbF to show the distribution pattern (Figure 6.2).

The cord bloods were arbitrarily assigned groups according to the level of HbF expressed (Table 6.7). The majority of cord blood samples (97.5%) showed a distribution between 52.1 and 88.9% HbF and constituted the normal reference range for the level of haemoglobin switching from HbF to HbA (See section 1.2.1.1). Factors like the absolute time post-conception and minor modifiers such as the $^{G}\gamma$-promoter mutation (C→T) at nucleotide –158 (*Xmn-I*) upstream of the cap site would probably alter the level of HbF observed. The distribution pattern for this haemoglobin at both centres were similar, probably as these maternity clinics were less than 50 miles apart and served overlapping communities.

Figure 6.2. **Chart to show the variation in levels of HbF in different cord blood samples collected in (A) Chilaw and (B) Kurunegala (HbA HbF HbBart's HbA_2)**

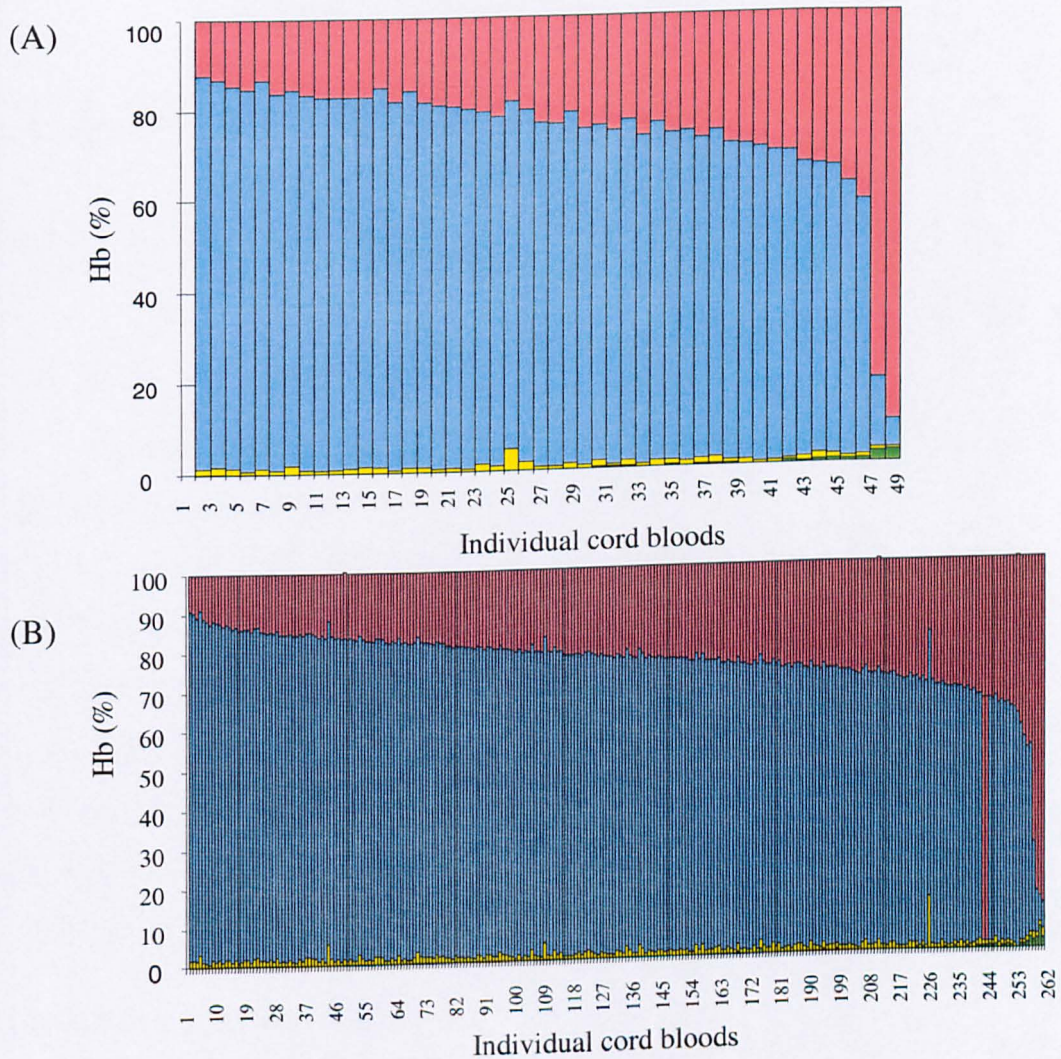


Table 6.7 Distribution of HbF levels observed in cord blood

Location	HbF Level (%)	Frequency		Mean ±SE	Range %	SD
		n	%			
Chilaw	>50	46	95.8	75.5±1.0	56.5-86.5	7.0
	20-50	0	0	-	-	-
	<20	2	4.2	10.7±4.6	6.1-15.2	6.4
	Total	48	100			
Kurunegala	>50	263	97.8	75.7±0.4	52.1-88.9	6.7
	20-50	3	1.1	39.8±8.3	23.2-48.5	14.4
	<20	3	1.1	8.1±1.5	6.6-11	2.5
	Total	269	100			
Overall	>50	309	97.5	75.7±0.4	52.1-88.9	6.8
	20-50	3	0.9	39.8±8.3	23.2-48.5	14.4
	<20	5	1.6	9.1±1.8	6.1-15.2	3.9
	Total	317	100			

Overall, relatively few cord blood samples (2.5%) demonstrate HbF levels below this range and were divided into two groups. The first set had a wide range of HbF between 20 and 50%. Samples close to 50% probably represent the more extreme end of the normal range but samples with very low HbF levels may have some form of γ-thalassaemia or a dysregulation in the switching process so that early switching occurs. This is also the case with the last group that has HbF levels below 20% although there may be a danger that they represent maternal blood that was collected by mistake, although this seems somewhat unlikely.

DNA analysis of patients and cord bloods

Ten cord blood samples were investigated further. Only eight of these were analysed by Southern blot analysis using a γ-probe (Table 6.8). The two samples that were not analysed were expected to have a normal γ-gene arrangement as they showed the highest level of HbF. K52 had a raised level of HbA_2 that may have been due to this individual having α-thalassaemia or being a carrier for HbE. A normal γ-gene arrangement was revealed so further investigation of the sequence was required to account for the variation of the level of HbF.

Table 6.8 **γ-gene analysis of 10 selected cord blood samples taken in Sri Lanka**

Cord Blood	Level of haemoglobins measured using the Biorad β-thalassaemia short column					γ–gene status Southern blot	
	Hb F	Hb A	Hb A_2	Hb Barts	Total Hb	*Bgl II*	*Hind III*
N46	5.3	79.3	2.3	0.4	87.3	12.9	8/3.5
K83	5.3	71.5	2	2	80.8	12.9	8/7.2/3.5
K52	5.7	75.1	4.6	1.2	86.6	12.9	8/3.5
K227	9.1	70.6	2.1	1.1	82.9	12.9	8/7.2/3.5
N29	13.2	71.2	2.1	0.6	87.1	12.9	8/7.2/3.5
K253	17.4	54.5	1.4	1.7	75	NR	NR
K257	37	37.2	1.1	1	76.3	12.9	8/7.2/3.5
K113	44.1	30.6	0.6	0.8	76.1	12.9	8/7.2/3.5
K93	60.8	8.8	0	4.1	73.7	ND	ND
N42	67.9	16.2	0	4.3	88.4	ND	ND

NR= No result ND= Not Done

The C→T substitution at position Cap –158 in the promoter region of the $^{G}\gamma$-gene creates an *Xmn-I* restriction enzyme site that can be assessed by RFLP (See section 2.4). In Sri Lanka this polymorphism was correlated with the steady state haemoglobin level in patients with HbE/β-thalassaemia (Table 6.9).

Table 6.9 **Correlation between *Xmn-I* status and HbF in patients with HbE/β-thalassaemia**

Xmn-I Status	Number of patients	HbF (%)			Total Hb (g/dl)		
		Mean ±SE	SD	Range	Mean ±SE	SD	Range
-/-	15	20.4±2.5	9.9	5.5-38	4.9±0.2	0.8	3.0-4.2
+/-	59	28.6±1.3	10.2	2.7-57.7	5.5±0.1	1.1	3.3-8.1
+/+	11	41.1±2.2	7.3	31.4-50.6	7.0±0.2	0.8	5.6-8.4

In keeping with other studies, a high level of HbF was associated with a +/+ *Xmn-I* configuration, lower levels with a -/- configuration and the +/- group was between these two. The -/- group was associated with a mean Hb concentration of 4.9 g/dl, the +/- group with 0.6 g/dl more haemoglobin and the +/+ group with 1.5 g/dl more haemoglobin than the intermediate group. Hence, the net effect of the + configuration of the *Xmn-*

I polymorphism appears to correlate positively with the total haemoglobin concentration presumably by raising the level of HbF.

Oligonucleotides were designed to amplify and sequence both γ-genes independently (See section 2.4) in a Sudanese family with neonatal haemolysis (Bayoumi et al., 1999). γ-gene sequence analysis was performed on 10 cord blood samples and 15 patients with HbE/β-thalassaemia. A complex situation was observed with five sequence variations in the $^{G}\gamma$-gene sequence and ten variations in the $^{A}\gamma$-gene sequence (Table 6.10, 6.11 and 6.12). The majority demonstrated the same changes in both groups studied and were tabulated with the relative levels of HbF (Table 6.11). Four of these mutations ($^{G}\gamma$: IVSII-115 (A→G) and Poly A Tail (-A); $^{A}\gamma$: IVSII-24 (A→C) and IVSII-115 (A→G)) were also described previously in the family from Sudan. Two mutations ($^{G}\gamma$: Codon 20 (GTG→GCG) and $^{A}\gamma$: Poly A Tail +2 (T→C)) were found in an isolated case only. All the other mutations were relatively frequent in the limited number of samples available and the sequence chromatograms are shown in Figure 6.3.

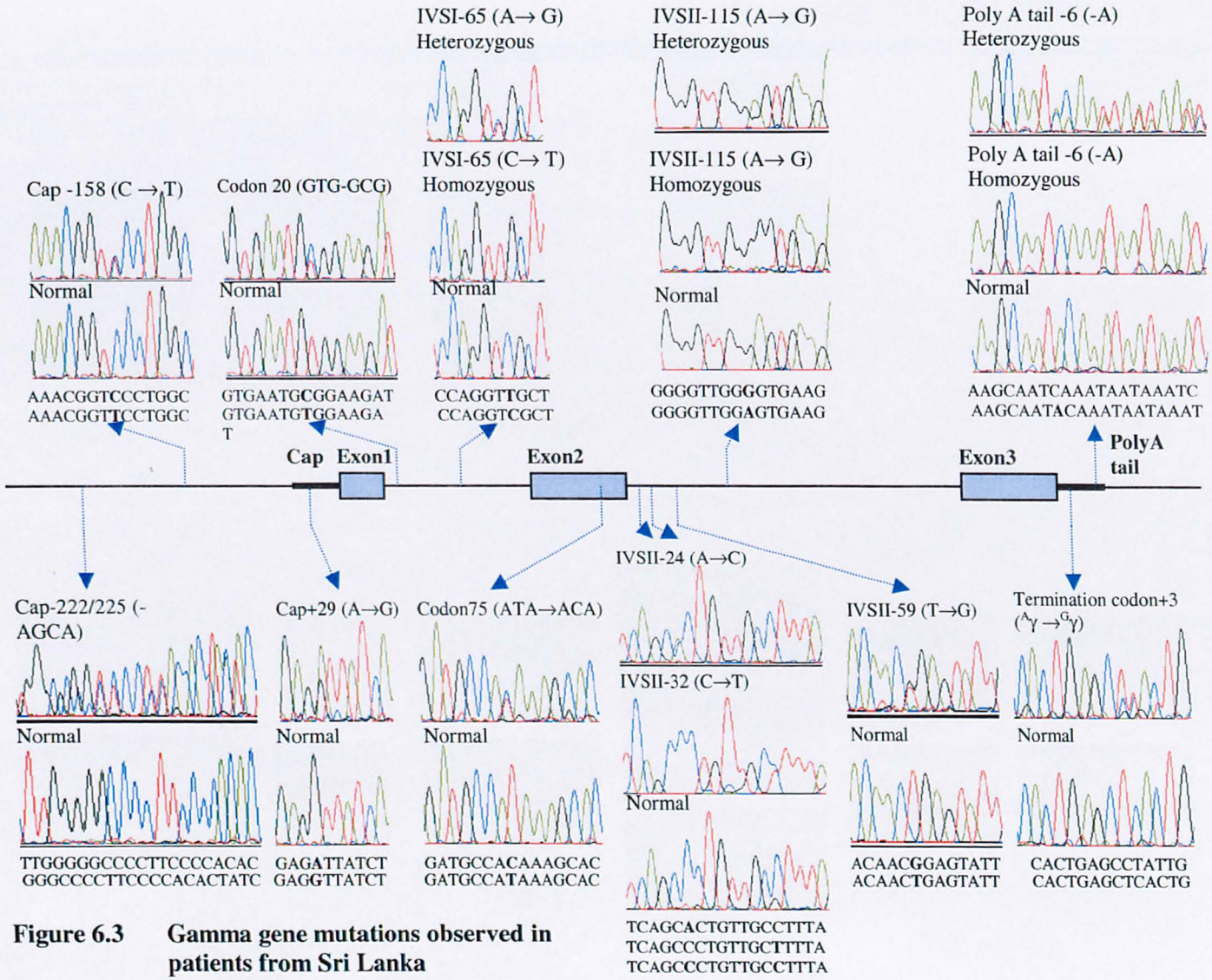


Figure 6.3 **Gamma gene mutations observed in patients from Sri Lanka**

Table 6.10 γ-gene sequence analysis of 10 cord blood samples taken in Sri Lanka

$^{G}\gamma$-Gene sequence						
Lab. No.	HbF (%)	-158 (T→C)	Codon 20 (GTG→GCG)	IVSI-64 (C→T)	IVSII-115 (A→G)	Poly A tail −6 (-A)
N46	5.3	C/C	T/T	T/T	G/G	A/A
K83	5.3	T/C	T/T	C/T	A/A	A/Δ
K52	5.7	C/C	T/T	T/T	G/G	A/A
K227	9.1	C/C	T/T	C/T	A/A	Δ/Δ
N29	13.2	C/C	T/T	C/T	A/G	A/A
K253	17.4	NR	T/T	NR	NR	NR
K257	37	C/C	T/T	C/C	A/A	Δ/Δ
K113	44.1	T/C	T/T	C/T	A/A	A/Δ
K93	60.8	C/C	T/T	C/T	A/A	Δ/Δ
N42	67.9	T/C	T/C	C/C	A/A	A/Δ

$^{A}\gamma$-Gene sequence										
Lab. No.	-222/-225 Δ4bp	Cap+29 (A→G)	IVS1-64 (C→T)	CD75 (ATA→ACA)	IVSII-24 (A→C)	IVSII-59 (T→G)	IVSII-115 (A→G)	3'URR TGA+3 4bp Ag/Gg conv.	Poly A tail − 6 (-A)	Poly A tail + 2 (T→C)
N46	wt / wt	G/G	T/T	T/T	C/C	T/T	G/G	wt / wt	A/A	T/T
K83	wt / wt	A/G	C/T	T/T	A/C	T/T	A/A	wt / wt	A/Δ	T/T
K52	wt / wt	G/G	T/T	T/T	C/C	T/T	G/G	wt / wt	A/Δ	T/T
K227	wt /Δ4bp	G/G	C/C	T/C	C/C	T/G	G/G	wt /con	A/A	T/C
N29	wt /Δ4bp	NR	C/T	T/C	A/C	T/G	G/G	wt /con	A/A	T/T
K253	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
K257	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
K113	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
K93	wt / wt	G/G	C/T	T/T	A/C	T/T	G/G	wt / wt	A/A	T/T
N42	wt /Δ4bp	A/G	C/C	T/C	A/C	T/G	A/G	wt /con	A/A	T/T

wt= wild type sequence Δ= deletion NR= no result ND=not done con=putative gene conversion

Table 6.11 $^{G}\gamma$-gene sequence analysis of HbE/β-thalassaemia patients

Lab. No.	Total Hb (g/dl)	HbF (%)	HbF (g/dl)	-158 (T→C)	IVS1-64(C→T)	IVSII-115 (A→G)	Poly A tail –6 (-A)
SL68/K409	4.8	9.7	0.47	C/C	T/T	G/G	Δ/Δ
SL60/K415	4.8	10.9	0.51	T/C	C/T	A/A	A/A
SL116/K501	6.2	12.2	0.76	T/C	C/T	A/A	A/Δ
SL14/K467	4.8	18.6	0.89	T/C	C/T	A/A	A/A
K227/K259	4.3	27.2	1.17	C/C	C/T	A/G	Δ/Δ
SL36/K498	5.3	22.6	1.2	T/C	C/T	A/A	A/A
SL203/K391	6.6	21.1	1.39	T/C	C/T	A/G	A/Δ
SL5/K466	5.6	27.5	1.57	C/C	C/T	A/G	Δ/Δ
SL200/K474	4.3	45.2	1.94	T/C	C/T	A/G	Δ/Δ
SL82/K327	6.4	40.9	2.62	T/T	C/C	A/A	A/A
K206/K488	7.4	44.3	3.28	T/T	C/T	A/A	A/A
K306	7.5	44.2	3.32	T/T	C/C	A/A	A/Δ
SL35/K469	8.1	45.6	3.69	T/C	C/T	A/G	Δ/Δ
K362	6.5	57.7	3.75	T/C	C/T	A/G	A/Δ
SL126/K411	7.5	50.5	3.79	T/T	C/C	A/A	A/A

Δ= deletion NR= no result ND=not done con=putative gene conversion

Table 6.12 $^{A}\gamma$-gene sequence analysis of HbE/β-thalassaemia patients

Lab. No.	-222/-225 Δ4bp	Cap+29 (A→G)	IVS1-64 (C→T)	CD75 (ATA→ACA)	IVSII-24 (A→C)	IVSII-59 (T→G)	IVSII-115 (A→G)	3'URR TGA+3 4bp Ag/Gg conversion	Poly A tail – 6 (-A)
SL68/K409	wt/wt	G/G	T/T	C/C	C/C	T/T	G/G	wt/wt	A/Δ
SL60/K415	wt/wt	G/G	T/T	C/C	C/C	T/T	G/G	wt/wt	A/Δ
SL116/K501	wt/wt	A/G	C/C	C/C	A/C	T/T	A/A	wt/wt	A/A
SL14/K467	wt/wt	A/A	T/T	C/C	C/C	T/T	G/G	wt/wt	A/Δ
K227/K259	wt /Δ4bp	G/G	C/T	C/T	C/C	T/G	G/G	wt /con	A/A
SL36/K498	wt /Δ4bp	G/G	C/T	C/T	A/A	T/T	A/A	wt /con	A/Δ
SL203/K391	wt/wt	A/G	C/C	C/C	A/C	T/T	A/A	wt/wt	A/A
SL5/K466	wt /Δ4bp	G/G	C/T	C/T	C/C	T/T	G/G	wt /con	A/A
SL200/K474	wt/wt	G/G	T/T	C/C	C/C	T/T	G/G	wt /con	A/Δ
SL82/K327	wt/wt	A/A	C/C	C/C	A/A	T/T	A/A	wt/wt	A/A
K206/K488	wt/wt	A/A	C/C	C/C	A/A	T/T	A/A	wt/wt	A/A
K306	wt/wt	A/A	C/C	C/C	A/C	T/T	A/A	wt/wt	A/Δ
SL35/K469	wt/wt	A/A	T/T	C/C	C/C	T/T	G/G	wt/wt	A/Δ
K362	wt/wt	G/G	C/C	C/C	A/C	T/T	A/A	wt/wt	A/Δ
SL126/K411	wt/wt	A/A	C/C	C/C	A/A	T/T	A/A	wt/wt	A/A

wt= wild type sequence Δ= deletion NR= no result ND=not done con=putative gene conversion

An analysis was made on all these cord bloods and patients with HbE/β-thalassaemia to see if there was any correlation between any of these mutations and the level of HbF expressed. As there were many mutations found in all individuals and each mutation could have a positive, negative or no effect, the overall effect was the sum of all the individual effects and hence made the analysis more complex. Each mutation was analyzed independently and summarized in Table 6.13.

Expression of the $^{G}\gamma$-gene is three times greater than the $^{A}\gamma$-gene so mutations in the $^{G}\gamma$-gene should be associated with a slightly greater change of HbF level compared to those in the $^{A}\gamma$-gene. The *Xmn-I* polymorphism, as described earlier, showed a positive correlation between the + or (T) form and a higher level of HbF. Also in $^{G}\gamma$-gene, IVSI-64 (C→T) was associated with a high level of HbF in a (C/C) homozygote but with a lower HbF level in a (T/T) homozygote and IVSII-115 (A→G) appeared to have a lower level of HbF associated with (G) variant. In the $^{A}\gamma$-gene three mutations were associated with a change in the level of HbF. Cap+29 (A→G) appeared to be associated with a higher level of HbF as an (A/A) homozygote, whilst IVSI-64 (C→T) and IVSII-115 (A→G) were the same as their counterparts in the $^{G}\gamma$-gene.

The 4bp deletion at Cap (–222-225) in the promoter of the $^{A}\gamma$-gene has been described previously. It was found in a number of apparently normal adults in linkage disequilibrium with the CD75 (ATA→ACA). It was associated with a low $^{A}\gamma$ expression and may well affect the total HbF level (Huisman, 1997). In this study, this small deletion was not found in the homozygous state and hence was not observed.

Table 6.13 Putative correlation between polymorphisms in the γ-genes and the level of expression of HbF

Mutation	Sequence	Mean ±SE (%)	Frequency n	Frequency %	Range (%)	SD (%)	Effect on HbF
$^{G}\gamma$-gene							
A -158 (T→C)	TT	45.0 ±2.0	4	16.7	40.9-50.5	4	↑
	TC	31.9 ±6.3	11	45.8	5.3-67.9	21	None
	CC	21.7 ±6.2	9	37.5	5.3-60.8	18.5	↓
B IVS1-64 (C→T)	CC	48.1 ±5.4	5	20.8	37-67.9	12.1	↑
	TC	29.1 ±4.5	16	66.7	5.3-60.8	18	None
	TT	6.9 ±1.4	3	12.5	5.3-9.7	2.4	↓
C IVSII-115 (A→G)	AA	33.4 ±5.4	14	58.3	5.3-67.9	20.2	None
	AG	33.9 ±6.0	7	29.2	23.2-57.7	15.9	None
	GG	6.9 ±1.4	3	12.5	5.3-9.7	2.4	↓
D Poly A tail -6 (-A)	Homozygote	32.7 ±6.4	8	33.3	9.1-60.8	18	None
	Heterozygote	36.1 ±8.9	7	29.2	5.3-67.9	23.6	None
	Wild type	23.6 ±5.8	9	37.5	5.3-50.5	17.3	↓?
$^{A}\gamma$-gene							
A -222/-225 4bp Δ*	Wild type	29.8 ±5.1	16	70	5.3-60.8	20.4	None
	Heterozygote	27.9 ±8.5	6	30	9.1-67.9	21	None
B Cap+29 (A→G)	AA	40.7 ±4.6	6	28.6	18.6-50.5	11.3	↑
	AG	26.6 ±14.1	4	19.1	5.3-67.9	28.3	None
	GG	25.6 ±6.2	11	52.3	5.3-60.8	20.5	↓?
C IVS1-64 (C→T)	CC	38.7 ±6.8	9	40.9	9.1-67.9	20.3	None
	CT	26.1 ±7.8	6	27.3	5.3-60.8	19.1	↓?
	TT	20.1 ±6.7	7	31.8	5.3-45.6	17.8	↓
D IVSII-24 (A→C)	AA	39.6 ±6.0	4	18.2	22.6-50.5	12.0	None
	AC	35.3 ±8.9	8	36.4	5.3-67.9	25.1	None
	CC	20.5 ±4.9	10	45.4	5.3-45.6	15.4	↓
E IVSII-115 (A→G)	AA	33.2 ±6.1	9	40.9	5.3-57.7	18.3	None
	AG	67.9	11	4.6	67.9		?
	GG	23.2 ±5.3	12	54.5	5.3-60.8	18.4	↓
F 3'URR TGA+3 4bp $^{A}\gamma\leftrightarrow^{G}\gamma$	Heterozygote	30.4 ±7.6	7	31.8	9.1-67.9	20.2	None
	Wild type	28.9 ±5.3	15	68.2	5.3-60.8	20.7	None
G Poly A tail -6 (-A)	Wild type	31.7 ±6.1	12	54.6	5.3-67.9	21	None
	Heterozygote	26.6 ±6.2	10	45.4	5.3-57.7	19.7	↓?

↑ denotes positive regulation to increase the level of HbF
↓ denotes negative regulation to decrease the level of HbF
? unknown effect
* This was invariably found in conjunction with CD75 (ATA→ACA)

The mutations, within and around the γ-genes, make their analysis difficult. Many of these mutations may be linked to form a framework. Each framework may be associated with a phenotype. Higher levels of HbF are associated with (T/C/A/(-A) at positions A-D) in the $^{G}\gamma$-gene and (wild type/A/C/A/A/A at positions A-G) in the $^{A}\gamma$-gene.

6.4 Other modifiers

Haemochromatosis

Transfusion dependent thalassaemia results in a loading in the body of iron. Any other factors that also result in the accumulation of iron can add to this overload. The aim of the study was to gain an assessment of the incidence of known causes of genetic haemochromatosis (GH) in Sri Lanka. A random screen was performed on the same cohort of patients under investigation for haemoglobinopathies. Two common mutations that account for the majority of all cases of the disease were investigated by standard PCR RFLP assays (See section 2.4).

The allele frequencies were determined in three clinics on the island as shown in Table 6.14. The incidence of the common North European mutation for GH (C282Y) was only found in one isolated case from Anuradhapura giving an overall allele frequency of 0.2% for this mutation. The mutation C282Y was found as a compound heterozygote with the H63D mutation and represented an extremely rare occurrence in Sri Lanka.

Table 6.14 Allele frequency of the common haemochromatosis mutations in β-thalassaemia patients attending three different clinics in Sri Lanka

Haemochromatosis status	Region							
	Anuradhapura		Badulla		Kurunegala		Total	
	n	%	n	%	n	%	n	%
Wild type	259	93.8	23	82.1	215	93.5	497	93.1
C282Y	1	0.4	0	0	0	0	1	0.2
H63D	16	5.8	5	17.9	15	6.5	36	6.7
Total	276	100	28	100	230	100	534	100

The total allele frequency for the mutation H63D was found to be 6.7%. The H63D mutation appeared to be present at a much higher frequency in Badulla than the two other centres but was probably due to the relatively small number of patients investigated at this centre. A single homozygote for this mutation was identified in each of the centres studied. Further analysis showed that this mutation was evenly represented in both β-thalassaemia major patients and also HbE/β-thalassaemia patients (Table 6.15).

Table 6.15 Allele frequency of the common haemochromatosis mutations in β-thalassaemia patients attending three different clinincs in Sri Lanka

Haemochromatosis status	β-Thalassaemia major		HbE/ β-Thalassaemia	
	n	%	n	%
Wild type	273	94.1	224	91.8
C282Y	1	0.3	0	0
H63D	16	5.6	20	8.2
Total	290	100	244	100

Patients with HbE/β-thalassaemia were investigated further to evaluate the iron status of the patients. Liver biopsies were collected and prepared into either paraffin blocks or frozen at –20°C in Sri Lanka. Histological assessment was made of sections of the paraffin blocks by Dr K. Fleming at the John Radcliffe Hospital in Oxford. The frozen liver

biopsies were assessed for iron content by Professor M. Pippard in Dundee and serum ferritin was performed by the routine haematology laboratory at the John Radcliffe Hospital in Oxford. Preliminary data appears not to show any correlation between the haemochromatosis mutation H63D and the following parameters; hepatic iron concentration, serum transferrin level, liver fibrosis, haemoglobin concentrations or α-thalassaemia status. Further analysis was required to be confident that any biological parameter measured gave an accurate determination. A greater understanding of any other disorders that may be present within this population as well as any treatment that they receive should also be taken into consideration.

Uridine Diphosphate Glycosyltransferase (UGT1A)

An observation, by Professor D. Weatherall, was made that a number of patients had a profound jaundiced appearance. An investigation was made to examine if this characteristic was associated with an elevated level of bilirubin. The most common molecular lesion associated with hyperbilirubinaemia is a tandem repeat element in the promoter region of the UGT1A gene. An increased number of these repeat elements were linked with a lower activity of the enzyme (Bosma *et al.*, 1995). A genetic characterization of this locus was investigated to determine the importance of the promoter element mutation in the population of Sri Lanka and to study its role in jaundice and gallstone formation in HbE/β-thalassaemia.

Eighty-five patients that were part of an ongoing study group were examined at least six months after the previous blood transfusion. Five patients on regular transfusions were characterised pre-transfusion. Each patient was examined by abdominal ultrasonography with special emphasis on the presence or absence of gallstones. The number of tandem repeat

elements was determined by sequence analysis as outlined earlier (See section 2.4). These were compared against 144 anonymous controls collected from the Kurunegala Teaching Hospital to determine the genetic status of the non-thalassaemic populations in Sri Lanka (Table 6.16).

Table 6.16 Genotype and allele frequency for the 5' promoter repeat element

Genotype	HbE/β-thalasssaemia		Normal controls	
	n	%	n	%
$A(TA)_6TAA/ A(TA)_6TAA$	15	17.6	43	29.9
$A(TA)_6TAA/ A(TA)_7TAA$	48	56.5	68	47.2
$A(TA)_7TAA/ A(TA)_7TAA$	22	25.9	33	22.9
Total	85	100	144	100
Allele				
$A(TA)_6TAA$	78	45.9	154	53.5
$A(TA)_7TAA$	92	54.1	134	46.5
Total	170	100	288	100

Patients and control samples demonstrated a very high level of homozygosity for the $A(TA)_7TAA$ motif at 25.9% and 22.9%, respectively. The HbE/β-thalassaemia patients had a higher frequency for the longer isoform $(TA)_7$. This was also reflected in the allele frequencies. HbE/β-thalassaemia patients were associated with an incidence of 54.1% for the $(TA)_7$ repeat compared with 46.5% in the control group.

These patients also were examined to look for any correlation between the steady state levels of haemoglobin, total and indirect bilirubin values, the frequency of gallstones and their genotype for the 5' promoter element (Table 6.17). Preliminary data were obtained for all parameters so means with standard deviations are presented.

Table 6.17 Promoter genotype correlation with biochemical phenotypes in HbE/β-thalasssaemia patients

A(TA)nTAA Genotype	Hb (g/dl)	Total Bilirubin (mmol/L)	Indirect Bilirubin (mmol/L)	Gallstones/No of patients			
				>15 years old		All ages	
				Index	%	Index	%
$(TA)_6/(TA)_6$	53.8	32.9	24.6	1/7	14	1/14	7
$(TA)_6/(TA)_7$	54	34.9	24.1	9/18	50	10/44	23
$(TA)_7/(TA)_7$	55.5	62.3	50.1	7/9	78	7/21	33

The steady state haemoglobin levels were similar in all groups indicating that the degree of ineffective erythropoiesis and haemolysis was comparable. However, there was a highly significant increase in both the total and indirect bilirubin levels associated with the $(TA)_7$ homozygote genotype. Gallstones were found in almost one quarter (18/79) of the patients and all but one of these was in the over 15 years old age group.

Glucose-6-Phosphate Dehydrogenase (G-6-PD) deficiency

The aim was to estimate the incidence of the common enzyme deficiency (G-6-PD). The subsequent incidence in the transfusion dependent patients could then be compared and the molecular lesions common in the Indian subcontinent screened in this population. The phenotype of patients with both disorders could then be assessed.

A neonatal screening strategy was employed on the same cord blood samples collected for the evaluation of Hb Bart's and was carried out by Dr S. Rajapakshe. A G-6-PD deficiency colorimetric assay was performed on the red cells shortly after collection as outlined earlier (See section 2.3). The assay was performed on 125 cord blood samples. One hundred and twenty of these samples showed normal enzyme activity but five (4%) were observed to have intermediate deficiency. However, as G-6-PD deficiency is inherited in an X-linked fashion, most of the cases reported are in males.

Hence, a G-6-PD deficiency in females is highly variable and is difficult to detect during screening (Iwai *et al.*, 2001). As the sex of the cord blood sample individuals were not recorded the overall frequency for this enzyme deficiency was probably higher than that determined.

The activity of this enzyme was also determined on 60 HbE/β-thalassaemia patients and their relatives from the Kurunegla Teaching Hospital with the help of Dr A. Premawardhena. Two of these families were found to have an intermediate deficiency to give a similar incidence of 3.33%. Again no allowance was made for any difference in activity of this enzyme between males and females and so the incidence is probably higher.

DNA analysis was performed on transfusion dependent patients and some of their relatives. The patients were chosen at random and were not ideal candidates to have a G-6-PD mutation. Mutations found in India, especially from the North East, were identified as potential candidates to screen for in the population of Sri Lanka. In the Indian states of Orissa and Madhya Pradesh two mutations were reported, G-6-PD Mediterranean and G-6-PD Orissa. G-6-PD Mediterranean was reported in the urban populations and was thought to have derived from the Aryan populations originating from central southern Europe. G-6-PD Orissa was reported in the rural populations and was thought to have derived from the aboriginal groups of India (Kaeda *et al.*, 1995). A third mutation in exon 11 at nucleotide 1311 (Beutler & Kuhl, 1990) was used as a crude marker for a European (T) or an Indian (C) associated haplotype. All three mutations were screened for as shown in Table 6.18.

Table 6.18 Incidence of G-6-PD mutations in Sri lanka

G-6-PD Mutation	Individuals studied	RFLP Patterns		
		Negative	No result	Positive
G-6-PD Mediterranean	230	154*	71	5
G-6-PD Orissa	125	97	28	0
G-6-PD 1311 (T/C)	39	0	39	0

* Some of these samples gave a different RFLP picture that required further investigation to clarify the cause of the unusual pattern.

The screen for G-6-PD molecular lesions observed in India revealed the presence of the G-6-PD Mediterranean mutation. This was found in 2.2% of the individuals assessed and constituted the major cause for G-6-PD deficiency in Sri Lanka. The mutation was found in four relatives and one transfusion dependent patient from Anuradhapura. Two of the relatives were carriers for β^0-thalassaemia, one was a carrier for HbE and the last was uncharacterised. One patient with HbE/β-thalassaemia was found to have the G-6-PD Mediterranean mutation. This patient was probably female as the RFLP appeared to give both normal and mutant alleles. Interestingly, there appeared to be no G-6-PD Orissa positive individuals tested in this population. The incidence of G-6-PD Mediterranean observed did not account for the predicted number of cases expected and so at least one other molecular lesion remains to be identified.

Those individuals observed to have a G-6-PD mutation were not available to study further to gain an insight into any interaction between G-6-PD and thalassaemia. However, another tentative link between the population of Sri Lanka and the Aryan population of the Indian subcontinent was established.

6.5 Discussion

A key problem in medical sciences is to relate findings at the molecular level to the clinical outcome in the patient. The haemoglobinopathies represent the most common monogenic disease in humans and should represent the simplest model for predicting phenotypes from genotypes. On close examination of a population the phenotype of some patients with an identical genotype are quite variable. β-thalassaemia has an extremely diverse clinical phenotype. At the severe end of the scale (β-thalassaemia major), many homozygous or compound heterozygous states are characterized by profound anaemia from early life and if untreated with blood transfusions lead to death in the first year. Conversely, many patients with the same genotype have a milder illness that ranges from a condition slightly less severe than β-thalassaemia major through a spectrum of decreasing severity of anaemia to one which is without symptoms and is ascertained only by routine blood examinations. Even the heterozygous state for β-thalassaemia demonstrates a spectrum of haematological abnormalities ranging from a dominantly inherited form to a completely silent form. The underlying causes in the inconsistencies in these correlations have remained elusive (Weatherall, 2000).

A correlation between genotype and phenotype can be established if the details of specific genetic mutations and their effects on protein products can be linked. This would allow clinical genetics to become predictive of disease prognosis and a better-informed selection among therapeutic strategies can be made for any individual patient. Complex diseases may also be unraveled in this same manner. As mutations have been identified for a disease it has become clear that genotype and phenotype correlation is

incomplete and only a subset of all mutations reliably predict phenotype (Dipple & McCabe, 2000).

These causes have been grouped according to the level of the interacting influence (Weatherall, 2001). Primary modifiers alter the gene under study directly, secondary modifiers change the gene indirectly and tertiary modifiers affect other biological systems and become interacting genes. To relate phenotype to genotype a consistent definition of the severity of the phenotype must be made. The common underlying factor that is present in thalassaemia is related to the degree of anaemia. Measurement of the steady state haemoglobin level should give a good classification for different genotypes. However, as many factors influence the level of haemoglobin that include age, nutrition, infection and treatment (eg. blood transfusions) all these factors must be considered when trying to correlate phenotype to genotype.

6.5.1 Primary modifiers

In individuals with a mutation in their β-gene the primary modifier is the nature of that mutation. Over 200 different mutations have been identified in the β-globin genes of patients with β-thalassaemia (Huisman & Carver, 1998). These mutations interfere with gene function either at the transcriptional, translational or post-translational level. The resultant β-thalassaemia reflects the output from this altered β-gene (β^0-, β^+- or β^{++}-correlating to complete, marked or mild reduction of β-globin chains). Hospital based patients were selected for study due to their dependence on blood transfusions and were expected to form part of the more severe spectrum of mutations and hence exhibit only a small phenotypic variation. Indeed, 85.2% of the alleles from these patients were found to be either

severe β^+ or β^0 and not capable of producing any significant amount of β-globin (Table 6.1). Patients that were homozygous or compound heterozygous for these mutations presented with a transfusion dependent form of the disease. Any variation in disease severity between these patients would be expected to be small. However, due to the lack of time a closer examination of these patients was not possible. This will be targeted as one of the next phases of the project.

A small set of patients (0.7%) were observed with mutations, -28 (A→G) and Poly A tail (T→C), that were classed as β^+ thalassaemia alleles and were associated with a moderate clinical phenotype (Table 6.1). These patients were always found as compound heterozygotes with a β^0-allele and presented with severe anaemia at a young age. They were treated similarly to patients with β-thalassaemia major and hence their clinical status was masked. If these milder forms of β-thalassaemia exist in the homozygous or compound heterozygous state they have not presented at clinics around the island and hence remain unreported.

A larger set of patients presented with a mutation that resulted in a structural variation of haemoglobin. Three β-chain variants were identified; HbD, HbE and Hb S but only HbE was found in a significant number (13.8%) of transfusion dependent patients (Table 6.1).

Patients with HbD/β-thalassaemia are clinically very mild and present with symptoms marginally more severe than carriers (reviewed in Weatherall & Clegg, 2001). An 18 years old man that presented with this condition was found to have hepatosplenomegaly, jaundice and anaemia though he had not had any blood transfusions he should be classed as β-

thalassaemia intermedia. This patient required closer examination to identify any compounding factors that made the phenotype more serious.

Patients with HbS/S or HbS/β-thalassaemia have sickle cell disease. HbS/β-thalassaemia can be classified into two broad groups; a severe type similar to HbS/S with little (up to 15%) or no HbA and a milder form with up to 30% HbA (reviewed in Weatherall & Clegg, 2001). Again HbS is an extremely rare mutation in this population and was only found in combination with a β-thalassaemia allele. These patients were identified as part of the more serious transfusion dependent individuals and so the phenotypes are assumed to be severe. However these patients were found to have their HbS mutation associated with the Arab/Indian haplotype. HbS found on the Indian/Arab haplotype is associated with a higher level of HbF expression as compared with other haplotypes associated with HbS. The higher levels of HbF sometimes present with a milder phenotype but the variability is not understood. These patients may benefit from a review of their disease management.

HbE in the homozygous state is associated with an extremely mild form of thalassaemia with little elevated production of HbF (reviewed in Weatherall & Clegg, 2001). Only one relative had this condition and was deemed asymptomatic. However, patients with this structural variant and β-thalassaemia were relatively common on the island. A cohort with this condition was examined more rigorously. HbE/β-thalassaemia is associated with a wide phenotypic heterogeneity ranging from transfusion dependent anaemia in early life to a clinically 'silent' condition that is ascertained by chance in middle age (de Silva *et al.*, 2000). These patients allowed other modifying factors to be investigated.

6.5.2 Secondary modifiers

The severity of the anaemia of β-thalassaemia is reflected by β-globin chain production that leads to excess α-chains and their deleterious effects on red cell production and survival (See section 1.2.2.2). Hence, anything that modifies the magnitude of the surfeit of α-chains should have an important effect on the phenotype.

α-gene status

Rearrangements in the human α-globin cluster are amongst the most common genetic abnormalities observed in human populations. α-thalassaemia coexists with β-thalassaemia at a high frequency in many populations and so it is not uncommon to inherit both conditions (reviewed in Weatherall & Clegg, 2001). Deletions are the most prevalent type of modification but α-gene duplications are also encountered as a result of these non-reciprocal crossover events (reviewed in Weatherall, 1997a). Levels of α-gene expression modulate the severity of β-thalassaemia to a minor degree. The interaction of either deletion or non-deletion α-thalassaemia has a beneficial effect on the clinical phenotype of severe β-thalassaemia (Weatherall *et al.*, 1981), especially β^+-thalassaemia and less of an effect on β^0-thalassaemia. Heterozygotes for β-thalassaemia and α^0-thalassaemia have higher MCV and MCH values than those heterozygous for either trait alone. In contrast, extra α-chains would further exacerbate the chain imbalance and result in a more severe phenotype. Patients with ααα or αααα gene arrangement and a concurrent β-thalassaemia present a β-thalassaemia intermedia phenotype. These patients generally have a more

severe phenotype than either α-gene triplication or β-thalassaemia in isolation (Rund *et al.*, 1997; Ho *et al.*, 1998a; Altay *et al.*, 1998; Beris *et al.*, 1999).

The incidence of α-gene rearrangement on the island is very high at 18.3%, almost one fifth of the population. This form of thalassaemia was distributed at similar levels around the study centres with the one exception in Anuradhapura. Perhaps, this frequency has been attained not just due to the direct beneficial effect of α-thalassaemia against infant infection (Allen *et al.*, 1997) but also to ameliorate the globin chain imbalance due to β-thalassaemia. Although the red cells of individuals who have inherited both types of thalassaemia might be hugely lacking haemoglobin, the anaemia is less severe and consequently the phenotype is milder.

Further analysis of the patients studied showed that this incidence of α-thalassaemia was heterogeneous amongst different groups of individuals according to their haemoglobinopathy status. Carriers for HbE or β-thalassaemia (~78%) and patients homozygous or compound heterozygous for β-thalassaemia (~80%) had a similar incidence for α-gene rearrangement (Table 6.2). The carriers for a haemoglobinopathy had more cases of α-thalassaemia and the transfusion dependent cases had more of extra α-genes. This reflects the deleterious effect of extra α-genes to result in a more severe phenotype for these patients. Surprisingly, the patients with HbE/ β-thalassaemia had far fewer α-gene rearrangements (~10%) with about 3% due to extra α-genes similar to the other groups and only ~7% of these patients presented with α-thalassaemia. The most probable possibility was that this secondary modifier was effective in this group of patients and that

patients with these genotypes may have been those that had a clinically 'silent' phenotype.

Individuals with only extra α-genes are clinically 'silent'. Patients with extra α-genes are relatively common in Sri Lanka. Such patients have presented with 5, 6 and 8 α-genes. Anecdotally, the greater the number of α-genes present the less perturbation of the β-gene was necessary to result in the clinical necessity for frequent blood transfusions (Table 6.3). Individuals with five α-genes that carried a single β-globin gene mutation were phenotypically moderate to normal (Table 6.4). A single case presented with β^0-Tr (CD15(-T)) and ααα/ααα (Table 6.5) and another with β^+-Tr (IVSI-5(G→C) and ααααα/ααααα (Table 6.6), both cases were phenotypically severe. All patients with βTr/βTr or HbE/βTr and the same α-genotype (ααα/αα) were all classed as severe.

HbF production

Normal children and adults produce small amounts of HbF that appear to be confined to a particular red cell population called F-cells (Rees *et al.*, 1999). Some patients homozygous for β^0-thalassaemia have a mild clinical phenotype and are able to maintain a relatively high haemoglobin level comprised of HbF. This continued production of γ-chains not only increases the level of haemoglobin but also combines with α-chains to reduce the α/β chain imbalance and the consequent ineffective erythropoiesis (reviewed in Weatherall & Clegg, 2001). Hence, the phenotypic results of β-thalassaemia are not just controlled by the reduction in the level of β-chains but are heavily linked with the ability of γ-chains to replace them. The molecular

determinants that maintain a high level of γ-chain production are poorly understood but should be regulated at many levels.

The normal regulation of globins around the time of birth needed to be characterized in detail before abnormal regulation could be investigated. The second haemoglobin switch usually prior to birth substitutes β- for γ-chain production (HbA for HbF) and continues until adult levels are attained by 3-4 years of age (See section 1.2.1.1). The pattern of HbF disappearance was reported in a cohort of 209 normal human neonates (Colombo *et al.*, 1976). At birth the mean HbF level was 64.8 ± 1.2% and for 15 days remained at the same level. A rapid drop was observed over the next 15 weeks down to 10.2 ±1.2% after which the rate of decrease slowed dramatically. A similar pattern was also reported on a cohort of 266 infants in a French population (Cheron *et al.*, 1989). A smaller group with β-thalassaemia trait was followed and showed a similar pattern but with a delay of 2-4 weeks in the postnatal decline (Wood *et al.*, 1982). The effect of γ-thalassaemia on the level of HbF has not been reported but may change the ratio of $^{A}\gamma$ to $^{G}\gamma$ expression and if severe reduce the total level of HbF or even the timing of Hb switching.

Compared with the other globin chains the number of mutations identified that are associated with a reduction in the level of HbF is very sparse. Individuals with a serious form of γ-thalassaemia would probably switch to express their adult globins to allow the foetus to have an oxygen transporter and hence survival. Any long-term effects on the neonate due to this early switching to adult haemoglobin are unknown. Individuals with coexisting β-thalassaemia would probably be miscarried during pregnancy.

In Sri Lanka, two groups of individuals allowed further investigation of the level of HbF expression, patients with HbE/β-thalassaemia and cord

blood samples collected at the time of birth. The ability for some patients with HbE/β-thalassaemia to maintain high levels of HbF is well known (reviewed in Weatherall & Clegg, 2001). The steady state haemoglobin level was determined and these patients demonstrated a normal distribution pattern with a wide range of HbF (Figure 6.5). The level of HbF in these patients ranged from 5% up to 70% with a modal level at 30-35% (Figure 6.5). This continuous range of expression suggests that there may be more than one controlling gene and that the overall expression is a sum of these positive and negative regulators.

γ-Gene mutations associated with elevated HbF

Specific rare mutations that are associated with an increased level of HbF include mutations in the β-globin gene, deletions that remove the β-globin gene and point mutations in the promoters of one or other of the duplicated γ-globin genes. They are all characterized by the persistent production of high levels of HbF into adult life (reviewed in Weatherall & Clegg, 2001 and Huisman *et al.*, 1997)(See section 1.2.1).

In Sri Lanka, a deletion that removed the β-globin gene (δβ-thalassaemia) was found in two families but remained uncharacterised (See Chapter 4). These mutations present as elevated HbF levels throughout adulthood. A random screen of schoolchildren indicated that factors increasing the level of HbF were rare.

Sequence analysis of both γ-genes and their surrounding DNA revealed many sequence variations. The *Xmn-I* polymorphism upstream of the $^{G}\gamma$-gene had an important association with the level of HbF. Patients with a (+/+ or T/T) configuration were associated with over 10% more HbF than

with the (+/- or T/C) configuration that was in turn just under 10% more than the (-/- or C/C) configuration (Table 6.12). More importantly, the total steady state haemoglobin level was associated with levels of 7.0, 5.5 and 4.9 g/dl, respectively. In this cohort of patients with HbE/β-thalassaemia, a difference of 1.0 g/dl in the baseline haemoglobin concentration appeared to potentially be enough to change the phenotype of that patient from severe to something significantly less than severe.

This relatively common polymorphism at position Cap-158 in the $^{G}\gamma$-gene is well known to be associated with an increased propensity to produce HbF under conditions of haemopoietic stress but has little effect in normal people (Gilman & Huisman, 1984; Labie *et al.*, 1985; Thein *et al.*, 1987). It has always been thought of as a minor modifier of HbF levels and results in a higher proportion of $^{G}\gamma$-chains in HbF. This locus may be part of a butyrate response element (BRE) (Ikuta *et al.*, 1998). However, this association may be as part of a highly conserved set of DNA sequences linked along the cluster. A tandem repeat element $(TA)_9CACATATACG(TA)_{10}$ in the 5' hypersensitivity-2 site of the LCR was found to be linked with an *Xmn-I* (+) arrangement (Winchagoon *et al.*, 1994). Also a different tandem repeat $(AT)_9(T)_5$ was found 530bp upstream of the β-globin gene linked to the Arab-Indian haploype that is associated with a high level of HbF in patients with sickle cell disease (Pissard & Beuzard, 1994). These associations could all be in concordance but more extensive sequence analysis for all these loci need to be undertaken on a fully characterized set of patients that have had their steady state haemoglobins determined.

γ-gene mutations associated with γ-thalassaemia

The only known form of γ-thalassaemia is a deletion and was excluded because of the observed heterozygous polymorphisms. Large deletions that remove other genes on the cluster as well as the γ-gene, like heterozygous $(\varepsilon\gamma\delta\beta)^0$-thalassaemia, result in haemolytic disease of the newborn. These babies are anaemic at birth and then recover to a clinical picture similar to β-thalassaemia trait but with a normal HbA_2 level. After birth γ-thalassaemia is asymptomatic in most individuals.

γ-thalassaemia has been described at low frequencies in families from China, India, Europe and Africa and would appear to have a pan global distribution due to a single molecular lesion. The mutation is a deletion that removes 5kb of DNA between the $^{A}\gamma$ and the $^{G}\gamma$ genes resulting in a hybrid globin expressing only 50% HbF at birth (Huisman *et al.*, 1997). In the normal individual two functional γ-genes exist in series. Only one intact γ-gene is necessary for the presence of HbF and if all four γ-genes are missing an early switch to β-gene expression probably occurs. Hence, the neonate may undergo some form of transitory anaemia that would automatically clear up negating the requirement for further study. The phenotype needs to be assessed at this critical time.

The primary structure in and around the γ-genes is beset with many variations in the consensus sequence. These polymorphisms also appear to be linked to the expression of the γ-genes as seen in combination with the α-genes to form HbF. Some these molecular lesions appear to be associated with up regulation and some with down regulation of the γ-genes. The level of HbF observed is as a net result of all these effects and hence the relative importance of any single alteration is masked. Speculative sequence

combinations are linked to elevated HbF levels and also the converse. Hence, the $^{G}\gamma$-gene combination of –158 (C/C); IVSI-64 (T/T) and IVSII-115 (G/G) should be associated with a low HbF level while –158 (T/T) and IVSI-64 (C/C) should associate with a higher HbF level. Analogously with $^{A}\gamma$-Gene, the combination of cap+29 (G/G); IVSI-64 (T/T); IVSII-24 (C/C) and IVSII-115 (G/G) should be associated with a low HbF level while cap+29 (A/A); IVSI-64 (C/C) and IVSII-24 (A/A) with a higher HbF level. A method to determine the importance of any and all the mutations needs to be developed.

6.5.3 Tertiary modifiers

Only a few mutations in other genes have been demonstrated to alter the phenotype of thalassaemia. Many remain elusive but closer examinations of large groups of patients with apparently the same genotype will allow its dissection.

Haemochromatosis

Virtually all organisms are dependent on iron for survival but are at risk from iron deficiency or iron overload. Homeostatic mechanisms regulating the absorption, transport, storage and mobilization of cellular iron are essential in iron metabolism. Iron is a particularly insoluble molecule and requires specialized molecules to maintain iron in soluble and bio-available form. When intracellular iron exceeds cellular requirement the excess is retained from becoming toxic but still kept available by ferritin sequestration and deposition mainly in the cytoplasm of cells. Control and regulation of these complexes remains obscure. Cells requiring iron express transferrin receptors that provide for controlled access of transferrin to the

cells. Uptake of transferrin into cells is controlled by intracellular iron pools, which regulate proteins involved in iron metabolism through translational control mechanisms. Regulation is achieved via iron regulatory proteins and iron responsive elements via negative feedback mechanisms (Aisen *et al.*, 2001).

GH is a disease of iron overload leading to the accumulation of iron in specific organs including the liver, heart, pancreas and gonads. Iron absorption from the small intestines of these patients is abnormally elevated (Aisen *et al.*, 2001). Excess iron results in cirrhosis, hepatoma, cardiomyopathy, diabetes, hypogonadism and arthritis. GH is an autosomal recessive disease and its gene product (*HFE*) encoding a 343 amino acid type I transmembrane glycoprotein was mapped to 6p21.3 near the HLA locus. HFE is similar to MHC class I molecules and binds β_2 microglobulin. GH results from a loss of function of this transmembrane protein. A single-base mutation in human HFE, converting cysteine to tyrosine at codon 282 (C282Y) that disrupts a disulphide bridge, is associated with 83% of HH cases (Feder *et al.*, 1996). The mutant protein does not associate with β_2 microglobulin and is not transported to the plasma membrane (Feder *et al.*, 1997). HFE also forms a complex with transferrin receptors preventing the uptake of iron by cells. A second mutation converting histidine to aspartic acid at codon 63 (H63D) has been found at high frequencies in some populations but its role in gene dysfunction is not understood. Other mechanisms are important in the regulation of iron levels and different molecules have been associated with the transport of iron across membranes. These include molecules like the divalent metal transporter (DMT1 also known as DCT1 or Nramp2), hephaestin and ferroportin.

The C282Y allele has a remarkably high frequency in North European populations and is present in over 10% in the Irish. Evaluations of other populations away from Northern Europe demonstrate a decrease in the allele frequency to 1-5% around the Mediterranean and then only found in isolated individuals (Merryweather-Clarke *et al.*, 1997). In Sri Lanka, a single case with the mutation C282Y was identified and after further analysis revealed a unique haplotype. It was thought to have arisen in Sri Lanka independently from the mutation in Europe (Rochette *et al.*, 1999).

The incidence of H63D allele is consistent with levels determined on the Indian subcontinent (Merryweather-Clarke *et al.*, 1997). This mutation H63D was reported at high allele frequencies in most European populations ranging from 5-20% and was also found in populations of Africa and SE Asia at frequencies between 1-10%. The mutation was not found in aborigines from Australia or from the indigenous American populations except in isolated individuals in association with European haplotypes indicating probable admixture of populations from Europe (reviewed in Merryweather-Clarke *et al.*, 2000).

The mutation H63D is reported to be of low penetrance and causes disease in only a proportion of patients who are H63D/C282Y compound heterozygotes and in even fewer H63D homozygotes. Selective advantage conferred by HFE mutations was assumed to prevent iron deficiency and would include protection against hookworm infestation, malarial anaemia, multiple pregnancies or dietary deficiency and should result in high levels of both these mutations. Other selective advantages may be present within this population and need to be identified to allow a better understanding of their clinical importance.

As the incidence of thalassaemia on the island is high the chances that these polymorphisms, even though they are only present at a low incidence, could coexist with a haemoglobinopathy is reasonable. Generalized iron loading of the organs has been recognized as a complication of thalassaemia for many years (reviewed by Weatherall & Clegg, 2001). The excess iron is derived both from intestinal absorption and from transfusion. Carriers were not studied for the HFE mutations but have been reported. β-thalassaemia trait is characterised by mild, ineffective erythropoiesis and erythroid hyperplasia but only in a minority of patients does this ultimately lead to iron overload. Patients homozygous for C282Y and heterozygous for β-thalassaemia tend to have higher rates of iron accumulation and the development of severe iron related complications (Piperno *et al.*, 2000).

The incidence of C282Y and H63D were mainly obtained from transfusion dependent patients with β-thalassaemia major or HbE/β-thalassaemia. A comparison of these two groups of patients revealed no major differences. β-thalassaemia major patients who are well managed for blood transfusions and iron chelation therapy and have a single haemochromatosis mutation were reported not to get iron loaded but patients with two haemochromatosis mutations suffer from severe iron overload (Longo *et al.*, 1999). Patients with β-thalassaemia major develop complications of iron overload mainly due to blood transfusions that introduce more iron into the body. This commonly occurs before the age of 5 years. The excess iron tends to be deposited into the liver by use of the ferritin transport. If the liver iron exceeds 7mg/g dry weight and the ferritin level exceeds 1500μg/l then the prognosis for the patient is extremely poor and will probably result in death (Telfer *et al.*, 2000).

Uridine Diphosphate Glucuronosyltransferase

Individuals with Gilbert's syndrome have mild, chronic unconjugated hyperbilirubinemia in the absence of liver disease or overt haemolysis. The disorder is autosomal recessive and has been mapped to chromosome 2 and may be a family of genes with homologues on other chromosomes. On the basis of serum bilirubin levels 3-10% of the general (Western) population are estimated to have Gilbert's syndrome (Bosma *et al.*, 1995). The syndrome is considered harmless in adults but may be linked with some forms of liver disease. Hepatic glucuronidating activity, essential for excretion of bilirubin, is about 30% in patients with Gilbert's syndrome. The reduced glucuronidation results in an increased build up of bilirubin. Absence or severe reduction of the enzymatic activity results in a more severe form of hyperbilirubinaemia called Crigler-Najjar (CN) syndrome type I or type II respectively. CN-I and CN-II are extremely rare disorders with the type I associated with a severe neurologic syndrome resembling kernicterus (an infiltration of bilirubin into the brain and spinal cord that may cause deafness, cerebral palsy and mental retardation) and the type II with reduced hepatic glucuronidating activity around 10%. Serum bilirubin levels are influenced by many factors, both genetic and environmental. Beutler *et al.*, (1998) suggested that unstable UGT1A1 polymorphism may serve to 'fine tune' the plasma bilirubin level within population groups. At a high enough level it may provide protection against oxidative damage, but at a level that is sufficiently low to prevent kernicterus in infants.

Mutations associated with an extremely mild phenotype included a promoter element mutation. The TATAA element between nucleotides -23 and –38 was found to have 2 extra bases (TA) changing the repeat element

from $A(TA)_6TAA$ (6/6) to $A(TA)_7TAA$ (7/7). This homozygous change was found in patients with Gilbert's syndrome (Bosma *et al.*, 1995) resulting in a 70% reduction in expression of the gene. This increased repeat element (7/7) was found to be prevalent in many populations (European 13%; Asian 3% and African 19%) (Beutler *et al.*, 1998). Heterozygous and homozygous variants were found to associate with increased bilirubin levels in individuals with β-thalassaemia and G-6-PD deficiency (Sampietro *et al.*, 1997).

Surprisingly, the population of Sri Lanka demonstrated a very high frequency for two extra bases (TA) inserted into the 5' TATAA promoter element more consistent with the African population than the Indian population. The shift from six repeats to seven repeats is even more pronounced in the patients with HbE/β-thalasssaemia and may constitute a mechanism of protection against the oxidative stress imposed. Total and indirect bilirubin levels were nearly twice the level from 32.9 mmol/L and 24.6 mmol/L to 62.3 mmol/L and 50.1 mmol/L in the $(TA)_6$ and $(TA)_7$, respectively. The consequence of this elevated bilirubin level was observed in the incidence of gallstones rising from 14% to 78% of their respective groups. The variation in the promoter allele may be of considerable importance in the genesis of gallstones and gall bladder disease in patients with HbE/β-thalasssaemia as they get older especially in this population with a high incidence of the allele.

Glucose-6-phosphate dehydrogenase

Deficiency of G6PD and its linkage to the X-chromosome was identified in the 1950s (Carson *et al.*, 1956). More than 400 million individuals throughout the world have been found with the enzyme disorder

and over 200 different variants of the enzyme are found at high frequency in tropical areas (African, Mediterranean and Asiatic populations (Porter *et al.*, 1964)). Heterozygote advantage vis-a-vis malaria (Luzzato *et al.*, 1969) has been shown to account for this epidemiological distribution. Over 440 variants have been identified (Pietrapertosa *et al.*, 2001).

The enzyme G6PD catalyses the first step of the pentose phosphate pathway, the only NADPH-generation process in mature red cells, which lack the citric acid cycle. The mechanism of protection of G6PD-deficient cells against falciparum malaria was reviewed by Friedman and Trager (1981). As G6PD is critical to the regeneration of nucleotide precursors and NADPH, a coenzyme that is essential for protection against and repair of oxidative damage, red cells deficient in G6PD are more sensitive to hydrogen peroxide generated by the malaria parasite. The loss of potassium from the cell and from the parasite is largely responsible for the death of the parasite.

Clinically, deficiency of the red cell enzyme is the basis of favism, primaquine sensitivity and some other drug-sensitive hemolytic anaemias, anaemia and jaundice in the newborn and chronic non-spherocytic hemolytic anemia (Beutler *et al.*, 1968). The mature erythrocyte due to its non-nucleated status has a diminished capacity to respond to oxidative stress. Uncompensated oxidative stress in the red blood cell leads to oxidation of the haemoglobin resulting in methaemoglobin, Heinz Body formation and membrane damage. In the extreme cases haemolysis ensues while less severe oxidant stress leads to their premature destruction.

In G6PD deficient neonates, the bilirubin load is increased but the presence or absence of bilirubin does not appear to correlate with the

In G6PD deficient neonates, the bilirubin load is increased but the presence or absence of bilirubin does not appear to correlate with the severity of haemolysis (Kaplan *et al.*, 1997). If untreated by exposure to UV radiation neonatal jaundice can lead permanent neurological damage (kernicterus) or death. The most dangerous consequence of G6PD deficiency is neonatal icterus. Kernicterus has been documented repeatedly in populations in which class 2 variants are common and is an important preventable form of mental retardation. Phototherapy has been used to reduce bilirubin levels and phenobarbital has been used prophylactically with some success. Exchange transfusion is required if the bilirubin exceeds 20 mg/dl. This enzyme deficiency may be another compounding problem towards hyperbilirubinaemia and gallstone formation.

As populations prone to G-6-PD deficiency are the same as those that have high frequencies of thalassaemia it was interesting to try to establish if both these disease mechanisms that protect against malaria do so in concert or individually. A random screen for mutations in the G-6-PD gene found in populations on the Indian subcontinent was undertaken. Different populations of India showed a G-6-PD deficiency between 3 and 15% (Kaeda *et al.*, 1995). The incidence has been reported in the population of Sri Lanka at wide variance from 1-2% to nearly 14% (Nagaratnam *et al.*, 1969; Roberts *et al.*, 1972). The incidence of this enzyme deficiency in our study was 3-5% (200-600 thousand people on the island) that was similar to the incidence of β-thalassaemia. Also, the phenotype assessed on the thalassaemia patients may be artificially low as thalassaemia causes an increased activity of G-6-PD towards normality that is also true in young

erythrocytes (Tagarelli *et al.*, 2000). Hence, many people with G-6-PD deficiency may also have thalassaemia.

In India, a previously unreported deficient variant, G6PD Orissa was found to be responsible for most of the G6PD deficiency in tribal Indian populations but was not found in urban populations where most of the G6PD deficiency was due to the G6PD Mediterranean variant (Kaeda *et al.*, 1995). In Sri Lanka, only the G6PD Mediterranean variant was identified and G6PD Orissa was not found. Perhaps, if the Veddahs were studied in this respect G6PD Orissa might be observed.

Chapter 7 Past, Present and Future

7.1 Out of Sri Lanka

Sri Lanka has impressive health statistics for a developing country. This major demographic change in the pattern of the disease has followed improvements in nutrition, hygiene and basic medical care such that children that would have perished of infection or malnutrition in the first few years of life are now surviving, and are able to present for diagnosis and treatment. This success has allowed the scale of chronic health problems to be assessed. Prior to 1992 a quarter of all paediatric admissions to the Kurunegala Teaching Hospital were patients with pallor and hepatosplenomegaly. Most of these patients had no firm diagnosis, and were queried to have thalassaemia or chronic malaria. A few had a definitive diagnosis of thalassaemia made, based on cellulose acetate electrophoresis. These patients were treated with occasional transfusions based on the availability of blood.

In 1992, Dr S. de Silva, the newly appointed Consultant Paediatrician formed a Thalassaemia Society. One of its key successes was the initiation of a change in policy to the National Transfusion Service to allow chronically anaemic patients to receive more blood transfusions. In 1995, Dr S. de Silva invited Professor D.J. Weatherall to visit the clinic whereupon an arrangement was made to screen blood samples from these patients in Oxford for thalassaemic disorders. In 1996, these blood samples started to arrive in Oxford and it soon became apparent that over 95% of these patients had haemoglobinopathies. Soon afterwards a nation-wide population screen for the disorder was organised.

The initial aim was to define the molecular basis underlying globin disorders using the assumption that they would be similar to those observed in neighbouring populations or in populations with an historical link with Sri Lanka. The spectrum and frequency of different molecular lesions in the α- and β-globin clusters needed to be assessed. The genetic backgrounds were examined to try to find markers that would correlate with the degree of disease severity, or some attribute of the phenotype. Hence, groups of patients that appeared to have the same genotype but demonstrated a variable disease manifestation were investigated more thoroughly. The range of genetic variations could then be compared against other populations to examine the risk factors for different groups that may allow a more rational approach for disease management.

The project had many difficulties to overcome. The greatest problem was imposed by the ongoing war in Sri Lanka that made conditions unsafe for patients, their families, and clinicians and for all support services. However, blood sample collection and clinical information was collated in a most determined fashion. The second problem was due language. Although many people in Sri Lanka speak English labeling of specimens was performed phonetically such that many spelling variations caused great confusion and repetition of work. Only with the advent of a numbering system instituted in Oxford and the recruitment of a doctor from Sri Lanka, Dr A. Premawardhena, did most of these problems clear up. The other great technical problem to overcome was the distance between the patients and the laboratory in which the analysis was performed. Sri Lanka has a hot and humid climate that causes samples to degrade rapidly and this was a message that needed to be made clear to all collaborators. Biochemical and

DNA assays needed to be optimised and new assays developed. The accumulation of information that was attained in Oxford gave the necessary impetus to push the project forward.

Prior to this study nothing had been reported about the molecular basis of thalassaemia in Sri Lanka. These studies have achieved many of its goals. The incidence and spectrum of α- and β-globin gene mutations have been characterised on the island. The study has also allowed some genotype-phenotype associations to be made to help unravel the complexity of haemoglobinopathies found that is typical of all populations.

The incidence of α-thalassaemia observed in Sri Lanka was similar to populations from many tropical and sub-tropical zones that have reached these tremendously high levels because of selection. The genotype frequency was estimated to be very high at 13.8% and 0.4% for heterozygous and homozygous α^+-thalassaemia, respectively. Hence, on this small island a predicted incidence of heterozygous and homozygous α^+-thalassaemia would be 2.6 million and 56 thousand, respectively. Only α^+-thalassaemia was observed ($-\alpha^{3.7}$ and $-\alpha^{4.2}$) but not α^0- and α^T-thalassaemia. A similar distribution has been seen on the subcontinent of India except that α^0- and α^T-thalassaemia have been reported in isolated families. Under normal conditions these α^+-thalassaemia mutations would be associated with a normal phenotype and would pose no notable health problems.

The development of PCR assays to determine the known molecular lesions causing α-thalassaemia on the island were developed that will allow subsequent studies to be performed more rapidly and at less cost (Liu *et al.*, 2000). Indeed, as only clinically less serious forms of α-thalassaemia were identified these assays do not need to be performed routinely but only on

cases where the genotype does not correlate with the phenotype. In fact, the use of a biochemical assessment at birth for the level of Hb Bart's has proved to be extremely accurate in predicting the nature of any underlying α-thalassaemia.

The carrier frequency for β-thalassaemia and for HbE was determined at 2.3% and 0.8%, respectively. Hence, the number of individuals on the island predicted to have β-thalassaemia major or HbE/β-thalassaemia was 2650 and 956, respectively. Treatment for thalassaemia would be estimated to consume 5-9% of the annual expenditure on health in Sri Lanka (de Silva *et al.*, 2000; Perera *et al.*, 2000). An estimated one sixth of the transfusion dependent thalassaemia patients were studied and revealed twenty-three different molecular lesions in or surrounding the β-globin gene. This amazing repertoire represented over one tenth of the total molecular lesions of this kind reported in the literature. Previously reported studies almost always retain individuals that remain uncharacterized (Huisman *et al.*, 1997). After a considerable effort, all thalassaemia patients attending the Kurunegala Teaching Hospital had their haemoglobinopathy status characterized. These extended investigations helped to reveal three new mutations; two in the coding regions and one in the first intron of the β-globin gene.

Three β-globin gene mutations were extremely common and would form the basis of a genotype screen that should reveal 84.1% of these mutations. This would allow about 75% of families to be resolved by a PND service. To provide a reasonably comprehensive service (>95% of successful cases) over 20 of these β-globin gene mutations would have to be tested for.

At least one laboratory needs to be established as a reference laboratory for Haematological and DNA analysis in the country.

The spectrum of mutations and haplotype data on both the α- and β-globin clusters in the population of Sri Lanka is very similar to that of India, supporting close genetic ties with the Indian subcontinent and confirmed some historical chronicles (Figure 7.1).

Most notably, the comparison to the North-East region of India was striking mainly due to the high prevalence of HbE. The only large difference was due to the third most common mutation IVSI-1 (G→A) that was rare in South-East Asia but common in Southern Europe. The 3' haplotype was found to be unique and so suggested that this mutation had arisen independently on the island from the Mediterranean basin. Indeed, the genetic links with Europe were even further distanced when a patient with

the mutation in their haemochromatosis gene (C282Y) was found on a non-European haplotype (Rochette *et al.*, 1999).

In the thalassaemias the disease defining criteria is that of the level of haemoglobin that can be maintained. Influencing factors, direct or distant, modify the phenotype. Primary modifiers alter the expression of the β-globin gene directly. Mutations can result in different reductions in the production of the β-globin chain, the greater its reduction the more severe the anaemia. As the patients studied were hospital based the majority of individuals presented as homozygotes or compound heterozygotes with severe β^+- or β^0-thalassaemia alleles and could only be treated as severe transfusion dependent patients. Surprisingly, a high prevalence of individuals presented with a single copy of a severe β^+- or β^0-thalassaemia allele and a single copy of a mild β^+-thalassaemia allele or β-chain variant, mainly HbE. This group presented with a wide range of phenotypes from high frequency transfusion dependent patients to unaffected individuals that presented more like β-thalassaemia carriers.

Secondary modifiers affect other globin chains that form the haemoglobin tetramer. Transfusion dependent patients have severely defective β-globin chain production that results in an excess of α-chains. Coexisting α-thalassaemia would normalize this chain imbalance and thus prevent immature red cells from being targeted for destruction. In contrast, the counterpart to non-homologous recombination resulting in α^+-thalassaemia was the extra α-gene arrangement. Two forms of extra α-gene arrangement were observed ($\alpha\alpha\alpha^{anti3.7}$ and $\alpha\alpha\alpha\alpha^{anti3.7}$). In this population the high incidence of α-gene arrangement would give a predicted carrier level of 0.76 million people with extra α-genes. These gene arrangements

persist, in the population, and on their own they are quite innocuous and represent a neutral allele. However, in combination with β-thalassaemia further imbalance of the α- to β-globin chain ratio results in a deleterious transfusion dependent condition. Interestingly, patients with HbE/β-thalassaemia were found with a much lower incidence of α-thalassaemia suggesting a strong ameliorating effect from this mutation so that these patients rarely present at clinic.

By the same rationale, elevated production of γ-chains would combine with the excess α-chains to form HbF that can substitute reasonably adequately for HbA. The control and regulation of γ-chains is still poorly understood. Many polymorphisms in the β-globin gene cluster are associated with both higher and lower levels of HbF. The Xmn-I polymorphism is the best known of these and showed the most clear cut correlation with the absolute amount of haemoglobin in the circulating blood. A +/+ or (T/T) configuration gave a favourable prognosis. The difference between the -/- or (C/C) and the +/+ configuration and its association with HbF was about 2g/dl of total haemoglobin. An increase of haemoglobin by this level should benefit the patient enormously. The mechanism of this association and also of the 'Indian/Arab' haplotype, a haplotype associated with the sickle cell mutation, remains elusive. All the other polymorphisms identified may be part of a conserved sequence with direct functional properties or exist on certain haplotypes and only exhibit an association with the high HbF phenotype.

The distant or tertiary regulators remain more elusive but some that affect the level of anaemia have been established. These conditions are dynamic and exist in a state of balance that can easily be tipped down a

number of different biochemical paths. Indeed, some apparently innocuous polymorphisms can have profound effects for groups of individuals with other disorders and change the pathophysiology of β-thalassaemia.

Transfusion dependent patients that are poorly or moderately chelated tend to develop iron overload but some, even with good chelation therapy appear to have a tendency towards iron overloading. In this population the common HFE mutation H63D was observed at high a frequency. The association of this mutation with a disease phenotype is still under question and a host of other genes may well be involved. Factors responsible for the greater than expected levels of iron loading remain to be identified in this population.

The efficiency of the enzyme, UGT1A, is associated with a repeat element in the promoter of its gene. In the population of Sri Lanka there was a higher than expected frequency for the genotype $(TA)_7/(TA)_7$ and was even more pronounced in the patients with HbE/β-thalassaemia. This probably constitutes an adapted protective mechanism against oxidative stress. Unfortunately, the unpleasant side effects of excess bilirubin result in higher frequencies of gallstones and gall bladder disease in older patients with HbE/β-thalassaemia.

The incidence of G6PD deficiency in our study was 3-5% (200-600 thousand people on the island) and this was similar to the incidence of β-thalassaemia. The inheritance pattern of these two genes did not appear to be associated. However, a small subset of β-thalassaemia transfused patients co-inherited a G6PD deficiency. Any variability that this imposes on the phenotype of the patient is unclear but may be different at critical times for the patient.

7.2 Genetics, biology and disease

Prior to 1950, genetics in medicine relied on pedigree analysis. It was not until Beadle and Tatum (reviewed in Lehninger, 1978) gave the gene a functional definition in their demonstration of the one-to-one relationship of gene and enzyme that biochemical genetics became focussed and the description of inborn errors took off. The analysis of DNA followed and RFLP analysis allowed mutants associated with monogenic disease to be tracked and then diseases like sickle cell anaemia to be diagnosed antenatally (Kan & Dozy, 1978a). With the development of cloning and of PCR amplification molecular genetics allowed inborn errors to be defined as specific molecular mutants and for extensive regions of chromosomes to be sequenced.

Each disease exhibits its own character that enables us to diagnose and classify it further. Sequence comparisons of the human genome will provide a reference point to link more of these disorders with genes or markers to allow diagnosis, treatment, prognosis and prevention (Childs & Valle, 2000). The Human Genome Mapping Project (HGMP) has helped lay the foundations by providing a consensus working sequence and along with Expressed Sequence Tags (ESTs) has allowed estimations for the total number of genes (~30,000) and their loci to be established. Another consequence of the HGMP is the identification of many DNA sequence variants or SNPs in the human genome. Over 1.5 million SNPs were deposited in the public databases by 2000. The availability of an ultra-high density SNP map opens the possibility of studying by association genetic factors important in complex genetic traits in the human (Kwok, 2001). The analysis becomes more powerful when SNPs are linked to form haplotypes

that can then be associated with a group of individuals exhibiting a similarity.

The post genome era still needs to evolve to explain how these genes interact with one another and the environment to make each individual what they are. Evolution is based on the conversion of genetic variation between individuals into genetic variation between populations. However, natural selection acts on variation among phenotypes not genotypes. Phenotypes are buffered from change by genotypic and environmental variation and allow for the build up and storage of neutral genetic variation in phenotypically normal populations. Genetic buffering reduces evolvability and promotes evolutionary stasis. The translation of genetic differences into phenotypic differences requires a breakdown in phenotypic buffering to allow changes crucial to evolution. The dogma of the 'one gene one function' hypothesis developed the notion that genes can be understood in isolation and produce discrete and unambiguous effects on phenotype. However, set against a complex genetic and environmental backgrounds, alterations in a single gene can have unexpected consequences. In instances where parallel pathways feed into the same developmental process natural variation can change which pathway is rate-limiting and allow stabilising selection or drift to alter the genetic circuitry (Rutherford, 2000).

Phenotypic variation is most dramatically observed in the case of monozygotic twins. Most traits show substantial heritabilities that are exhibited as either discrete states or a continuous variation. Such quantitative variation is typically determined by both environmental and genetic factors with either a few genes having a large effect or by many genes with small effects (Barton & Keightley, 2002).

Comparison of sequence and alignment between individuals and between species can help to identify important regions. These could be coding regions for genes, *cis*-acting regulatory sequences and deriving models for the evolution of the genome and its regulation. The next goal is to emphasize functional studies of genomic regions. The consequence of a change in this conserved sequence can be linked to temporal and spatial patterns of gene expression and phenotypes and hence a greater understanding of the normal role of these genes (Hardison *et al.*, 1998).

The genome represents a dynamic structure that adjusts to environmental stimuli and in turn can alter these factors. Epidemiology aims to identify determinants of disease, either risk or protective factors, and to quantify their role. Selection bias arises from lack of comparability of groups included in the study, misclassification of the participants with respect to disease or exposure status or due to exposure to risk factors other than those under study (Boffetta, 2000). The genetic basis for susceptibility to infectious diseases has been indicated by twin, adoptee, pedigree, and candidate gene study. This has led to the identification of strong genetic associations with a few infectious diseases (malaria, HIV infection, and infectious prion disease) and less strong associations with many others (Hill, 2001). Detailed linkage dis-equilibrium mapping, resulting from genome wide SNPs, will allow much more precise mapping and also help define polygenic causes.

All tests for selective advantage use predictions from the theory of neutrally evolving sites as a null hypothesis. Departures from equilibrium-neutral expectations can indicate the presence of natural selection acting at either one or more of the sites under investigation or at a sufficiently tightly

linked site. Complications can arise in the interpretation of departures from neutrality if populations are not at equilibrium for mutation and genetic drift, or if populations are sub-divided, both of which are likely scenarios for humans (Kreitman, 2000).

Human population history- its major epochs of migration and expansion, instances of geographic isolation, the mixing of sub-populations, and major and minor historical population bottlenecks gives alternate reasons for distribution of frequencies of polymorphisms. The advent of large-scale SNP data sets for human gene loci provides an opportunity to directly compare within-population allele frequencies (and also between-population differences in allele frequencies) for synonymous (e.g. degenerate codons in an exon) and replacement polymorphisms (Kreitman, 2000).

The idea that gene differences among humans could inform us about the dynamics of human origins only became apparent in the last 15 years (reviewed by Harpending & Rogers, 2000). This gave great insight that the most recent common ancestor of human mitochondrial DNA existed within the last few hundred thousand years. Mitochodrial DNA and genomic DNA diversity both suggest a human population size of about 10-20,000 individuals about 10,000 years ago (Harpending & Rogers, 2000).

Analysis of data is fraught with problems as much information on human history is inferred (reviewed by Harpending & Rogers, 2000). The multi-regional evolution model (MRE) states that modern humans arose from a diffuse worldwide transformation of archaic humans whereas the Garden of Eden model (GOE) states that a focal origin from some unknown small population of archaics grew and spread over the temperate Old World.

Human populations may have a tree-like history analogous to that of species or human differences reflect *in situ* differentiation in a geographically structured population. Reality probably lies somewhere between these views but it is thought that the largest genetic differences are found between African and non-African populations. The genetic relationship among non-African populations is less clear.

We are demographically a young species, with most of our genetic variation tracing to a relatively small precursor population, or populations, in Africa, that recently colonized the rest of the world. The term race implies groups that can be cleanly separated from one another, and within our species, but there simply are no such groups. Differences among groups are graded and any groupings are arbitrary (Goldstein & Chikhi, 2002).

7.3 Disease screening in neonates

Screening of newborn infants for genetic disease began over 35 years ago as a public health measure to prevent mental retardation in phenylketonuria (PKU) by Guthrie in (reviewed in Levy & Albers, 2000). By the mid-1970's this had extended to about half a dozen disorders (galactosemia, maple syrup disease, homocystinuria and congenital hypothyroidism). This has been applied for a few diseases in some cases but mass screening has generally only been applied to PKU. Tandem mass spectrometry is substantially enhancing the screening process and expanding coverage to many treatable genetic disorders heretofore not identifiable by newborn screening. Moreover, the ability to examine DNA is offering the possibility of molecular screening for almost any disorder (Levy & Albers, 2000).

If one considers the use of cord blood and not a heel prick from a newborn infant a completely non-invasive system could be employed. Prior knowledge of critical genes or parts of genes to be investigated would be required to make a large scale screening program feasible. Indeed proposals are under discussion to implement a universal screen on heel prick blood from all babies for sickle cell disease by using a haemoglobin analyser, like the Biorad variant analyser. In the UK, there is a shortfall in the number of patients found to have sickle cell disease. This will not only increase the rate of detection but also allow the detection of α-thalassaemia. The success or failure of this plan will indicate its efficacy for other populations.

The products of genes, the proteins are again becoming the focal points in the analysis of disease. Proteomics incorporates the structure, function and interactions of a cell's full complement of proteins. The proteins translate structure into function, and as the elements of structural, metabolic and homeostatic systems, they represent the essence of biochemistry and physiology. The complexity of the proteome exceeds the genome in that variable promoters, alternate splicing, and extensive post-translational processing all serve to amplify and diversify genomic information. The phenotypes of complex disease need to be explained by locating several variant gene products with their relationship to each other in more physiological systems, each capable of communicating and integrating with others, explaining thereby the variability of the whole (Childs & Valle, 2000).

7.4 Future for thalassaemia

Although most of the common natural mutants have been characterized, better information about the frequency and distribution of

different forms of thalassaemia are essential. Indeed, it is now becoming clearer, that groups within a population, have a completely different frequency and spectrum of the molecular causes of thalassaemia, and each population needs critical assessment of each of these groups. Also novel molecular lesions may help greatly in elucidating our understanding of control and regulation of these genes in an unexpected manner.

In Sri Lanka, the study of thalassaemia patients attending clinics has revealed an extensive diversity of α- and β-globin gene mutations. However, the remaining unstudied patients (>80%) may well hold some interesting surprises especially within relatively isolated communities (particularly the non-Sinhalese). A particularly interesting group is that of the native inhabitants of the island, the Veddah that may have direct links to the tribal groups in India.

Technical developments, at the phenotype and genotype level, have now made large-scale population screening feasible. However, the widespread nature of the haemoglobinopathies and the large number of different mutations that interact to give a phenotype still require the input of large amounts of resources to understand the nature of the disease globally.

A reasonable evaluation of the incidence of these conditions can be assessed by biochemical analysis. Elevated levels of HbA_2 and HbF give a good estimate for β-thalassaemia. Elevated levels of Hb Bart's give an accurate indication for the presence of α-thalassaemia. If a serious form of α-thalassaemia (α^0) is present then individuals with HbH or greatly elevated levels of Hb Bart's would be observed. Indeed, if the analysis is performed during foetal development an estimate might be made on γ-thalassaemia that could result in serious clinical consequences. As the causes of death due to

infectious agents have diminished and health care has improved, people with genetic variations are increasing in each population. Human fecundity is said to be as low as 25% and the loss of 75% of conceptuses is called 'fetal wastage'. Disease accounts for some part of this 'wastage' but chromosomal anomalies may play a major role. A routine screening program would allow a better overall assessment of the incidence of haemoglobinopathies on the island and allow rare cases to be investigated.

Development of new reagents has allowed a much more reliable DNA analysis of the α-globin cluster. PCR analysis has become fast, simple and robust. Remarkable advances in miniaturization technology are making it possible to develop microchips which will encompass large numbers of human mutations and which will enable rapid screening of large populations. As the cost of these new approaches fall it may be practicable for one or two laboratories in each country, with a high frequency of mutations, to provide services for population screening. These would be cheaper and more effective than the use of simple approaches carried out in many different hospitals or clinics throughout these countries. Indeed, as a consequence of this program of research, the need for a central reference laboratory has been proposed and accepted. The logistics of this laboratory will be organised by the Wellcome Trust and the Sri Lankan Department of Health.

Even though cases of HbH disease have been previously reported in Sri Lanka (Nagaratnam & Sukuram 1967; de Tissera *et al.*, 1988) our survey would indicate that this disease is extremely rare. A further screen to find communities with these genotypes, α^0- and α^T-thalassaemia, would be of great clinical value. Further genetic classification of the single α-gene

deletions, $-\alpha^{3.7}$ and $-\alpha^{4.2}$, would allow greater comparison with other populations.

A wide variety of β-globin gene mutations have been identified but further mutations may be prevalent in isolated communities. The novel β-globin gene mutations still remain to be formally characterised and their mode of action demonstrated. The putative δβ-thalassaemia allele(s) revealed remain to be characterised. Once the breakpoints have been mapped a Gap-PCR method can be designed to allow the frequency of these molecular lesions to be estimated.

As Sri Lanka is historically most closely associated with South India studies on the population of South India may be extremely useful as a comparison. This may help to confirm a uni-centric or multicentric origin for the intronic mutation, IVSI-1 (G→A) and to establish if HbE extends down the East coast of India.

The relationship between genotype and phenotype and the reasons for the remarkable heterogeneity of all the thalassaemias is still poorly understood; HbE/β-thalassaemia is a good example. This disorder will affect hundreds of thousands of children throughout the Indian subcontinent and South East Asia. Still we know very little about the natural history of the disease that makes it impossible to make sensible recommendations about its best management. Reasons for the heterogeneity of forms of thalassaemia need to be understood so that screening and prenatal diagnosis programs can provide parents with appropriate counselling.

Analysis of the genotype-phenotype correlation could be extended to investigate individuals with β-thalassaemia that are transfusion independent or only require occasional blood transfusions. The phenotype should be well

characterised so that associations with genetic markers can be established. The molecular lesion in the β-gene should be identified as well as the modifying factors around the globin gene cluster and also at distant loci. Bilirubin as an antioxidant needs to be further characterised to have a greater understanding of its protective role against free radical damage.

Haplotype information can also be used to compare the different ethnic groups may indicate connections with other populations. This can be even more informative when found in association with mutations that may be subjected to selection. This should go hand in hand with full SNP mapping of loci. Greater sequence analysis between different populations can highlight natural genetic variation and when compared with sequence analysis across species can identify conserved and non-conserved regions.

Every country will have to decide how they wish to control the disease. Screening and prenatal diagnosis plans have proved enormously successful in some Mediterranean countries. However, other programs like premarital screening can be successful. If countries decide not to develop programs for the control of the disease, but to treat it when it occurs, they will have to work out the economics and how to develop their clinical services. In Sri Lanka, the development of a reference laboratory could be used in these capacities but PND is not sanctioned by the Buddhist faith within this population.

Transfusion dependent thalassaemic children can develop normally if adequately transfused and their excess iron chelated. The development of simpler, cheaper and more effective chelating agent would improve patient care tremendously. An analysis of the 'iron status' in a well characterised set of patients with similar or identical genetic lesions in the globin clusters may

indicate compartments of iron regulation that require closer inspection. As other factors critical for the regulation of iron are revealed their importance in these patients can be assessed. This may lead to improvement of chelating agents. Even well iron chelated children develop endocrine damage requiring further work to understand and avoid these complications. Blood borne infections will remain a major problem and so a blood transfusion service would need to be developed to cope with an increase in the demand for blood transfusions.

Ways of improving marrow transplantation between unrelated and unmatched donors, the use of stem cells and somatic cell gene therapy is progressing more slowly. As the underlying selective pressures that resulted in high a prevalence of thalassaemia is still present on the island this approach may have further problems to overcome. Although malaria was reduced it may well be making a comeback and great vigilance must be observed to prevent the ravages from this disease. It would ask the question, are man's efforts to control the effects of this disease better than nature? The evolutionary consequences will be interesting.

Stimulation of HbF synthesis could be an answer for some patients. This may not be plausible in severely transfusion dependent patients but an increase of 1-2 g/dl may be enough to improve the phenotype of the intermediate forms of β-thalassaemia. The relative amounts of γ- and β-globin chain production, in adults, are in equilibrium presumably directed by regulatory molecules. Certain perturbations of erythropoiesis favour the expression of γ-globin genes and this can be induced by a variety of drugs (Weatherall & Clegg, 2001). An understanding of the mechanisms that regulate HbF would allow treatments by these therapies to be optimized.

More individuals with steady state HbF levels should have their γ-genes examined so that the polymorphisms revealed can be assessed with greater statistical confidence. An assessment of the individual expression of both γ-genes would allow mutations in either gene to be assessed more critically.

In conclusion, the control of thalassaemia, especially in the poorer countries, remains a major question to be answered (Weatherall & Clegg, 1996). Current technology can improve screening programs for understanding the frequencies of haemoglobinopathies and can indicate health burden. This must be done soon otherwise detrimental affects of this disease will be huge. Gene therapy, bone marrow transplants, stem cell therapy may help tremendously when these technologies are fully developed but they will be prohibitively expensive for the scale of the problem.

Bibliography

Agarwal, S., Pradhan, M., Gupta, U.R., Sarwai, S. and Agarwal, S.S., Geographic and ethnic distribution of beta-thalassemia mutations in Uttar Pradesh, India, *Hemoglobin*, 24, 89, 2000.

Ahmad, O.B., Lopez, A.D. and Inoue, M., The decline in childhood mortality: a reappraisal, *Bulletin of the World Health Organization*, 78, 1175, 2000.

Ahmed, S., Petrou, M. and Saleem, M., Molecular genetics of beta-thalassaemia in Pakistan: a basis for prenatal diagnosis, *British Journal of Haematology*, 94, 476, 1996.

Aisen, P., Enns, C. and Wessling-Resnick, M., Chemistry and biology of eukaryotic iron metabolism, *International Journal Biochemistry and Cell Biology*, 33, 940, 2001.

Allen, S.J., O'Donnell, A., Alexander, N.D., Alpers, M.P., Peto, T.E.A., Clegg, J.B. and Weatherall, D.J., Alpha+-thalassemia protects children against disease caused by other infections as well as malaria, *Proceedings of the National Academy of Sciences (USA)*, 94, 14736, 1997.

Altay, C., Oner, C., Oner, R., Gumruk, F., Mergen, H. and Gurgey, A., Effect of alpha-gene numbers on the expression of beta-thalassemia intermedia, beta-thalassemia and (delta beta)0-thalassemia traits, *Human Heredity*, 48, 121, 1998.

Antonarakis, S.E., Boehm, C.D., Giardina, P.J. and Kazazian, H.H., Jr., Nonrandom association of polymorphic restriction sites in the beta-globin gene cluster, *Proceedings of the National Academy of Sciences (USA)*, 79, 137, 1982.

Bachmann, B., Wolfgang, L. and Hunsmann, G., Improvement of PCR amplified DNA sequencing with the aid of detergents, *Nucleic Acids Research*, 18, 1, 1990.

Baine, R.M., Rucknagel, D.L., Dublin Jr, P.A., and Adams 3rd, J.G., Trimodality in the proportion of hemoglobin G Philadelphia in heterozygotes: evidence for heterogeneity in the number of human alpha chain loci, *Proceedings of the National Academy of Sciences (USA)*, 73, 3633, 1976.

Bandyopadhyay, A., Bandyopadhyay, S., Chowdhury, M.D. and Dasgupta, U.B., Major beta-globin gene mutations in eastern India and their associated haplotypes, *Human Heredity*, 49, 232, 1999.

Bank, A. and O'Donnell, J.V., Intracellular loss of free alpha chains in beta thalassemia, *Nature*, 222, 295, 1969.

Bannerman, R.M., Iron absorbtion in thalassaemia, *British Journal of Haematology*, 10, 490, 1964.

Bannerman, R.M., Grinstein, M. and Moore, C.V., Haemoglobin synthesis in thalassaemia; in vitro studies, *British Journal of Haematology*, 5, 102, 1959.

Baralle, F.E.E., Shoulders, C.C. and Proudfoot, N.J., The primary structure of human ε-globin gene, *Cell*, 21, 621, 1980.

Barton, N.H. and Keightley, P.D., Understanding quantitatve genetic variation, *Nature Reviews: Genetics*, 3, 11, 2002.

Basak, A.N., Ozer, A., Kirdar, B. and Akar, N., A novel 13 Bp deletion in the 3'UTR of the beta-globin gene causes beta-thalassemia in a Turkish patient, *Hemoglobin*, 17, 551, 1993.

Bayoumi, R.A., Dawodu, A., Qureshi, M.M., Al-Khider, A., Fitzgerald, P., Riou, J., Fisher, C.A., Fitches, A. and Old, J.M., The association of Hb Khartoum [beta124(H2)Pro→Arg] with gamma+-thalassemia is responsible for hemolytic disease in the newborn of a Sudanese family, *Hemoglobin*, 23, 33, 1999.

Baysal, E. and Huisman, T.H., Detection of common deletional alpha-thalassemia-2 determinants by PCR, *American Journal of Hematology*, 46, 208, 1994.

Beris, P., Solenthaler, M., Deutsch, S., Darbellay, R., Tobler, A., Bochaton-Pialat, M.L. and Gabbiani, G., Severe inclusion body beta-thalassaemia with haemolysis in a patient double heterozygous for beta(0)-thalassaemia and quadruplicated alpha-globin gene arrangement of the anti-4.2 type, *British Journal of Haematology*, 105, 1074, 1999.

Beutler, E., Mathai, C.K. and Smith, J.E., Biochemical variants of glucose-6-phosphate dehydrogenase giving rise to congenital non-spherocytic hemolytic disease, *Blood*, 31, 131, 1968.

Beutler, E. and Kuhl, W., The nt 1311 polymorphism of G6PD: G6PD mutation may have originated independently in Europe and Asia, *American Journal of Human Genetics*, 47, 1008, 1990.

Beutler, E., Gelbart, T. and Demina, A., Racial variability in the UDP-glucuronyltransferase 1 (UGT1A) promoter: A balanced polymorphism for regulation of bilirubin metabolism?, *Proceedings of the National Academy of Science (USA).*, 95, 8170, 1998.

Boffetta, P., Molecular epidemiology, *Journal of Internal Medicine*, 248, 447, 2000.

Bosma, P.J., Chowdhury, J.R., Bakker, C., Gantla, S., De Boer, A., Oostra, B.A., Lindhout, D., Tytgat, G.N.J., Jansen, P.L.M., R.P.J., O.E. and Chowdhury, N.R., The genetic basis of the reduced expression of bilirubin UDP-Glucuronyltransferase 1 in Gilbert's syndrome, *New England Journal of Medicine*, 333, 1171, 1995.

Bowden, D.K., Vickers, M.A. and Higgs, D.R., A PCR-based strategy to detect the common severe determinants of alpha thalassaemia, *British Journal of Haematology*, 81, 104, 1992.

Bowie, L.J., Reddy, P.L., Nagabhushan, M. and Sevigny, P., Detection of alpha-thalassemias by multiplex polymerase chain reaction, *Clinical Chemistry*, 40, 2260, 1994.

Boyle, A.L. (Ed.), *Current protocols in human genetics*, John Wiley & Sons Inc, New York, 2000.

Brittenham, G., Lozoff, B., Harris, J.W., Kan, Y.W., Dozy, A.M. and Nayudu, N.V., Alpha globin gene number: population and restriction endonuclease studies, *Blood*, 55, 706, 1980.

Brown, J.M., Thein, S.L., Weatherall, D.J. and Mar, K.M., The spectrum of beta thalassaemia in Burma, *British Journal of Haematology*, 81, 574, 1992.

Bulger, M., von Doorninck, J.H., Saitoh, N., Telling, A., Farrell, C., Bender, M.A., Felsenfeld, G., Axel, R. and Groudine, M., Conservation of sequence and structure flanking the mouse and human β-globin loci: The β-globin genes are embedded within an array of odorant receptor genes., *Proceedings of the National Academy of Sciences (USA)*, 96, 5129, 1999.

Burmester T, Ebner B, Weich B, Hankeln T., Cytoglobin: a novel globin type ubiquitously expressed in vertebrate tissues. *Molecular and Biological Evolution*, 19, 416, 2002.

Carson, P.E., Flanagan, C.L., Ickes, C.E. and Alving, A.S., Enzymatic deficiency in primaquine-sensitive erythrocytes, *Science*, 124, 484, 1956.

Chan, V., Chan, T.K., Chebab, F.F. and Todd, D., Distribution of beta-thalassemia mutations in south China and their association with haplotypes, *American Journal of Human Genetics*, 41, 678, 1987.

Chang, J.G., Lee, L.S., Lin, C.P., Chen, P.H. and Chen, C.P., Rapid diagnosis of alpha-thalassemia-1 of southeast Asia type and hydrops fetalis by polymerase chain reaction, *Blood*, 78, 853, 1991.

Chang, J.G., Liu, T.C., Chiou, S.S., Chen, J.T., Chen, T.P. and Lin, C.P., Rapid detection of -alpha 4.2 deletion of alpha-thalassemia-2 by polymerase chain reaction, *Annals of Hematology*, 69, 205, 1994.

Cheron, G., Bachoux, I., Maier, M., Massonneau, M., Peltier, J.Y. and Girot, R., Fetal hemoglobin in sudden infant death syndrome, *New England Journal of Medicine*, 320, 1011, 1989.

Childs, B. and Valle, D., Genetics, biology and disease, *Annual Review of Genomics and Human Genetics*, 1, 1, 2000.

Chong, S.S., Boehm, C.D., Higgs, D.R. and Cutting, G.R., Single-tube multiplex-PCR screen for common deletional determinants of alpha-thalassemia, *Blood*, 95, 360, 2000a.

Chong, S.S., Boehm, C.D., Cutting, G.R. and Higgs, D.R., Simplified multiplex-PCR diagnosis of common southeast asian deletional determinants of alpha-thalassemia, *Clinical Chemistry*, 46, 1692, 2000b.

Chui, D.H. and Waye, J.S., Hydrops fetalis caused by alpha-thalassemia: an emerging health care problem, *Blood*, 91, 2213, 1998.

Colombo, B., Kim, B., Perez Atencio, R., Molina, C. and Terrenato, L., The pattern of fetal haemoglobin disappearance after birth, *British Journal of Haematology*, 32, 79, 1976.

Cooper, H.A. and Hoagland, H.C., Fetal hemoglobin, *Mayo Clinic Proceedings*, 47, 402, 1972.

Craig, J.E., Rochette, J., Fisher, C.A., Weatherall, D.J., Marc, S., Lathrop, G.M., Demenais, F. and Thein, S., Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach, *Nature Genetics*, 12, 58, 1996.

Czelusniak, J., Goodman, M., Hewett-Emmett, D., Weiss, M.L., Venta, P.J. and Tashian, R.E. Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. *Nature*, 298, 5871, 1982

Dastidar, D.G., Dutta, R.N., Gupta, P. and Old, J.M., Detection of β-thalassaemia mutations in eastern Indian population by polymerase chain reaction, *Indian Journal of Medical Research*, 100, 111, 1994.

de Silva, C.C. and Weeratunge, C.E.S., Cooley's anaemia in Sinhalese children, *Archives of Disease of Children*, 26, 224, 1951.

de Silva, C.C., Bulugahapitiya, D.T.D., de Silva, J., Wickramasinghe, R.L. and Jonxis, J.H.P., Sinhalese family with HbS, *British Medical Journal*, 1, 1592, 1962.

de Silva, C.R., *The history of Sri Lanka*, Vikas Publishing House, New Dehli, 1997.

de Silva, S., Fisher, C.A., Premawardhena, A., Lamabadusuriya, S.P., Peto, T.E.A., Perera, G., Old, J.M., Clegg, J.B., Olivieri, N.F., Weatherall, D.J. and Group, S.L.T.S., Thalassaemia in Sri Lanka: implications for the future health burden of Asian populations, *Lancet*, 355, 786, 2000.

de Tissera, A., Balakumar, L., Seneviratne, A. and Wickremasinghe, P., The alpha thalassaemic syndrome of Hb H disease in a Sri Lankan family, *Ceylon Medical Journal*, 33, 135, 1988.

Deidda, G., Novelletto, A., Hafez, M., al-Tonbary, Y., Felicetti, L., Terrenato, L. and Colombo, B., A new beta-thalassemia mutation produced by a single nucleotide substitution in the conserved dinucleotide sequence of the IVS-I consensus acceptor site (AG----AA), *Hemoglobin*, 14, 431, 1990.

Desai, S., Colah, R., Gupte, S. and Mohanty, D., Is cellulose acetate electrophoresis a suitable technique for detection of Hb Bart's at birth?, *Human Heredity*, 47, 181, 1997.

Dipple, K.M. and McCabe, E.R.B., Phenotypes of patients with "Simple" Mendelian Disorders are Complex Traits: Thresholds, Modifiers and Systems Dynamics, *American Journal of Human Genetics*, 66, 1729, 2000.

Divoky, V., Indrak, K., Mrug, M., Brabec, V., Huisman, T.H.J. and Prchal, J.T., A novel mechanism of β-thalassaemia: the insertion of L1 retrotransposable element into β-globin IVSII., *Blood*, 88, 148, 1996.

Dode, C., Rochette, J. and Krishnamoorthy, R., Locus assignment of human α-globin mutations by selective amplification and direct sequencing, *British Journal of Haematology*, 76, 275, 1990.

Dode, C., Krishnamoorthy, R., Lamb, J. and Rochette, J., Rapid analysis of -alpha 3.7 thalassaemia and alpha alpha alpha anti 3.7 triplication by enzymatic amplification analysis, *British Journal of Haematology*, 83, 105, 1993.

Drysdale, H.C. and Higgs, D.R., Alpha-thalassaemia in an Asian Indian, *British Journal of Haematology*, 68, 264., 1988.

el Kalla, S. and Mathews, A.R., A significant beta-thalassemia heterogeneity in the United Arab Emirates, *Hemoglobin*, 21, 237, 1997.

Eng, B., Patterson, M., Borys, S., Chui, D.H. and Waye, J.S., PCR-based diagnosis of the Filipino (--(FIL)) and Thai (--(THAI)) alpha-thalassemia-1 deletions, *American Journal of Hematology*, 63, 54, 2000.

Faustino, P., Osorio-Almeida, L., Barbot, J., Espirito-Santo, D., Goncalves, J., Romao, L., Martins, M.C., Marques, M.M. and Lavinha, J., Novel promoter and splice junction defects add to the genetic, clinical or geographic heterogeneity of beta-thalassaemia in the Portuguese population, *Human Genetics*, 89, 573, 1992.

Feder, J.N., Gnirke, A., Thomas, W., Tsuchihashi, Z., Ruddy, D.A.A., Basava, A., Dormshian, F., Domingo, R.J., Ellis, M.C., Fullan, A., Hinton, L.M., Jones, N.L., Kimmel, B.E., Kronmal, G.S., Lauer, P., Lee, V.K., Loeb, D.B., Mapa, F.A., McClelland, E., Meyer, N.C., Mintier, G.A., Moeller, N., Moore, T. and Morikang, E., A novel MHC class I-like gene is mutated in patients with hereditary haemachromatosis, *Nature Genetics*, 13, 399, 1996.

Flint, J., Harding, R.M., Boyce, A.J. and Clegg, J.B., The population genetics of the haemoglobinopathies, *Baillieres Clinical Haematology*, 6, 215, 1993.

Flint, J., Thomas, K., Micklem, G., Raynham, H., Clark, K., Doggett, N.A., King, A. and Higgs, D.R., The relationship between chromosome structure and function at a human telomeric junction, *Nature Genetics*, 15, 252, 1997.

Friedman, M.J. and Trager, T., The biochemistry of resistance to malaria, *Scientific America*, 244, 154, 1981.

Fucharoen, G., Fucharoen, S., Jetsrisuparb, A. and Fukumaki, Y., Molecular basis of HbE-beta-thalassemia and the origin of HbE in northeast Thailand: identification of one novel

mutation using amplified DNA from buffy coat specimens, *Biochemica Biophysica Research Communications*, 170, 698, 1990.

Fucharoen, S. and Winichagoon, P., Hemoglobinopathies in Southeast Asia: molecular biology and clinical medicine, *Hemoglobin*, 21, 299, 1997.

Galanello, R., Meloni, A., Gasperini, D., Saba, L., Cao, A., Rosatelli, M.C. and Perseu, L., The repeated sequence (AT)x(T)y upstream to the beta-globin gene is a simple polymorphism, *Blood*, 81, 1974, 1993.

Galanello, R., Sollaino, C., Paglietti, E., Barella, S., Perra, C., Doneddu, I., Pirroni, M.G., Maccioni, L. and Cao, A., Alpha-thalassemia carrier identification by DNA analysis in the screening for thalassemia, *American Journal of Hematology*, 59, 273, 1998.

Gale, R.E., Clegg, J.B. and Huehns, E.R., Human embryonic haemoglobins Gower 1 and Gower 2., *Nature*, 280, 162, 1979.

Garewal, G., Fearon, C.W., Warren, T.C., Marwaha, N., Marwaha, R.K., Mahadik, C. and Kazazian, H.H., Jr., The molecular basis of beta thalassaemia in Punjabi and Maharashtran Indians includes a multilocus aetiology involving triplicated alpha-globin loci, *British Journal of Haematology*, 86, 372, 1994.

Gilman, J.G. and Huisman, T.H., Two independent genetic factors in the beta-globin gene cluster are associated with high G gamma-levels in the HbF of SS patients, *Blood*, 64, 452, 1984.

Gilman, J.G., Johnson, M.E. and Mishima, N., Four base-pair DNA deletion in human A gamma globin-gene promoter associated with low A gamma expression in adults, *British Journal of Haematology*, 68, 455, 1988.

Goldstein, D.B. and Chikhi, L., Human migrations and population structure: What we know and why it matters, *Annual Review of Genomics and Human Genetics*, 3, 129, 2002.

Gomes, M.P., da Costa, M.G., Braga, L.B., Cordeiro-Ferreira, N.T., Loi, A., Pirastu, M. and Cao, A., Beta-thalassemia mutations in the Portuguese population, *Human Genetics*, 78, 13, 1988.

Goossens, M., Lee, K.Y., Liebhaber, S.A. and Kan, Y.W., Globin structural mutant alpha 125Leu leads to Pro is a novel cause of alpha-thalassaemia, *Nature*, 296, 864, 1982.

Gorakshakar, A.C., Pawar, A.R., Nadkarni, A.H., Lu, C.Y., Mohanty, D., Krishnamoorthy, R., Besmond, C. and Colah, R.B., Potential of denaturing gradient gel electrophoresis for scanning of beta-thalassemia mutations in India, *American Journal of Hematology*, 61, 120, 1999.

Graff, J.A.E., Ikin, E.W., Lehman, H., Mourant, A.E., Parkin, D.M. and Wickramasinghe, R.L., HbE and blood groups in the Veddas, *Journal of Physiology*, 127, 41p, 1954.

Gupta, R.B., Tiwary, R.S., Pande, P.L., Kutlar, F., Oner, C., Oner, R. and Huisman, T.H., Hemoglobinopathies among the Gond tribal groups of central India; interaction of alpha- and beta-thalassemia with beta chain variants, *Hemoglobin*, 15, 441, 1991.

Hardison, R. and Miller, W., Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters, *Molecular Biological Evolution*, 10, 73, 1993.

Hardison, R., Riemer, C., Chui, D.H., Huisman, T.H. and Miller, W., Electronic access to sequence alignments, experimental results, and human mutations as an aid to studying globin gene regulation, *Genomics*, 47, 429, 1998.

Harkness, M., Harkness, D.R., Kutlar, F., Kutlar, A., Wilson, J.B., Webber, B.B., Codrington, J.F. and Huisman, T.H., Hb Sun Prairie or alpha(2)130(H13)Ala----Pro beta 2, a new unstable variant occurring in low quantities, *Hemoglobin*, 14, 479, 1990.

Harpending, H. and Rogers, A., Genetic perspectives on human origins and differentiation., *Annual Review Genomics and Human Genetics*, 1, 361, 2000.

Harteveld, C.L., Losekoot, M., Haak, H., Heister, G.A., Giordano, P.C. and Bernini, L.F., A novel polyadenylation signal mutation in the alpha 2-globin gene causing alpha thalassaemia, *British Journal Haematol*, 87, 139, 1994.

Harteveld, K.L., Heister, A.J., Giordano, P.C., Losekoot, M. and Bernini, L.F., Rapid detection of point mutations and polymorphisms of the alpha-globin genes by DGGE and SSCA, *Human Mutation*, 7, 114, 1996.

Hassall, O.W., Tillyer, M.L. and Old, J.M., Prevalence and molecular basis of alpha thalassaemia in British South Asians, *Journal of Medical Screening*, 5, 31, 1998.

Hatton, C.S., Wilkie, A.O., Drysdale, H.C., Wood, W.G., Vickers, M.A., Sharpe, J., Ayyub, H., Pretorius, I.M., Buckle, V.J. and Higgs, D.R., Alpha-thalassemia caused by a large (62 kb) deletion upstream of the human alpha globin gene cluster, *Blood*, 76, 221, 1990.

Henke, W., Herdel, K., Jung, K., Schnorr, D. and Loening, S.A., Betaine improves the PCR amplification of G-C rich DNA sequences, *Nucleic Acid Research*, 25, 3957, 1997.

Higgs, D.R., Alpha-thalassaemia, *Baillieres Clinical Haematology*, 6, 117, 1993.

Higgs, D.R., Lamb, J., Aldridge, B.E., Clegg, J.B., Weatherall, D.J., Serjeant, B.E. and Serjeant, G.R., Inadequacy of Hb Bart's as an indicator of alpha thalassaemia, *British Journal Haematology*, 51, 177, 1982.

Higgs, D.R., Wainscoat, J.S., Flint, J., Hill, A.V., Thein, S.L., Nicholls, R.D., Teal, H., Ayyub, H., Peto, T.E.A., Falusi, A.G. Analysis of the human α-globin gene cluster reveals a highly informative genetic locus., *Proceedings of the National Academy of Sciences (USA)*, 83, 5165, 1986.

Higgs, D.R., Wood, W.G., Jarman, A.P., Vickers, M.A., Wilkie, A.O., Lamb, J., Vyas, P. and Bennett, J.P., The alpha-thalassemias, *Annals of the New York Academy of Science*, 612, 15, 1990.

Hill, A.V.S., The genomics and genetics of human infectious disease susceptibility, *Annual Review of Genomics and Human Genetics*, 2, 373, 2001.

Ho, P.J., Hall, G.W., Luo, L.Y., Weatherall, D.J. and Thein, S.L., Beta-thalassaemia intermedia: is it possible consistently to predict phenotype from genotype?, *British Journal of Haematology*, 100, 70, 1998a.

Ho, P.J., Hall, G.W., Watt, S., West, N.C., Wimperis, J.W., Wood, W.G. and Thein, S.L., Unusually severe heterozygous beta-thalassemia: evidence for an interacting gene affecting globin translation, *Blood*, 92, 3428, 1998b.

Honig, G.R., Shamsuddin, M., Vida, L.N., Mompoint, M., Valcourt, E., Bowie, L.J., Jones, E.C., Powers, P.A., Spritz, R.A., Guis, M. and et al., Hemoglobin Evanston (alpha 14 Trp---Arg). An unstable alpha-chain variant expressed as alpha-thalassemia, *Journal of Clinical Investigations*, 73, 1740, 1984.

Hsia, Y.E., Ford, C.A., Shapiro, L.J., Hunt, J.A. and Ching, N.S., Molecular screening for haemoglobin Constant Spring, *Lancet*, 1, 988, 1989.

Huisman, T.J.H., Gamma chain abnormal human fetal hemoglobin variants, *American Journal of Hematology*, 55, 159, 1997b.

Huisman, T.H., Carver, M.F. and Efremov, G.D., *A syllabus of human hemoglobin variants*, Sickle Cell Anemia Foundation, Augusta, 1988.

Huisman, T.H., Carver, M.F. and Baysal, E., *A syllabus of Thalassaemia mutations*, Sickle Cell Anemia Foundation, Augusta, 1997a.

Huisman, T.H. and Carver, M.F., The beta- and delta-thalassemia repository (Ninth Edition; Part I), *Hemoglobin*, 22, 169, 1998.

Hundreiser, J., Deka, R., Gogoi, B.C., Papp, T. and Flatz, G., DNA Haplotypes and frameworks associated with β-globin gene in the Karachi population of Assam (India), *Human Heredity*, 38, 240, 1988.

Ikuta, T., Kan, Y.W., Swerdlow, P.S., Faller, D.V. and Perrine, S.P., Alterations in protein-DNA interactions in the gamma-globin gene promoter in response to butyrate therapy, *Blood*, 92, 2924, 1998.

Inoue, K. and Lupski, J.R., Molecular mechanisms for genomic disorders, *Annual Review of Genomics and Human Genetics*, 3, 199, 2002.

Iwai, K., Hirono, A., Matsuoko, H., Kawamoto, F., Horie, T., Lin, K., Tantular, I.S., Dachlan, Y.P., Notopuro, H., Hidayah, N.I., Salim, A.M.A., Fujji, H., Miwa, S. and Ishii, A., Distribution of glucose-6-phosphate dehydrogenase mutations in Southeast Asia, *Human Genetics*, 108, 445, 2001.

Jane, S.M. and Cunningham, J.M., Understanding fetal globin gene expression: a step towards effective HbF reactivation in haemoglobinopathies, *British Journal of Haematology*, 102, 415, 1998.

Jeffreys, A.J., DNA sequence variants in $^{G}\gamma$-, $^{A}\gamma$-, δ- and β-globin genes of man, *Cell*, 18, 1, 1979.

Kaeda, J.S., Chotray, G.P., Ranjit, M.R., Bautista, J.M., Reddy, P.H., Stevens, D., Naidu, J.M., Britt, R.P., Vulliamy, T.J., Luzatto, L. and Mason, P.J., A new glucose-6-phosphate dehydrogenase variant,G6PD Orissa (44 Ala-Gly), is the major polymorphic variant in tribal populations in India, *American Journal of Human Genetics*, 57, 1335, 1995.

Kan, Y.W., Dozy, A.M., Trecartin, R. and Todd, D., Identification of a nondeletion defect in alpha-thalassemia, *New England Journal of Medicine*, 297, 1081, 1977.

Kan, Y.W. and Dozy, A.M., Antenatal diagnosis of sickle-cell anaemia by DNA analysis of amniotic-fluid cells, *Lancet*, 2, 910, 1978a.

Kanavakis, E., Metaxotou-Mavromati, A., Kattamis, C., Wainscoat, J.S. and Wood, W.G., The triplicated alpha gene locus and beta thalassaemia, *British Journal of Haematology*, 54, 201, 1983.

Kaplan, M., Renbaum, P., Levy-Lahad, E., Hammerman, C., Lahad, A. and Beutler, E., Gilbert syndrome and glucose-6-phosphate dehydrogenase deficiency: a dose dependent genetic interaction crucial to neonatal hyperbilirubinemia, *Proceedings of the National Academy of Sciences (USA)*, 94, 12128, 1997.

Kazazian, H.H., Jr., Orkin, S.H., Markham, A.F., Chapman, C.R., Youssoufian, H. and Weber, P.G., Quantification of the close association between DNA haplotypes and specific beta-thalassaemia mutations in Mediterraneans, *Nature*, 310, 152, 1984.

Khan, S.N. and Riazuddin, S., Molecular characterization of beta-thalassemia in Pakistan, *Hemoglobin*, 22, 333, 1998.

Ko, T.M., Tseng, L.H., Hsieh, F.J., Hsu, P.M. and Lee, T.Y., Carrier detection and prenatal diagnosis of alpha-thalassemia of Southeast Asian deletion by polymerase chain reaction, *Human Genetics*, 88, 245, 1992.

Kotea, R., Kaeda, J.S., Yan, S.L.K., Fa, N.S., Beesoon, S., Jankee, S., Ramasawmy, R., Vulliamy, T., Bradnock, R.W., Bautista, J., Luzzato, L. and Krishnamoorty, R., The major G6PD deficient polymorphic variants identified among the Mauritian population, *British Journal of Haematology*, 104, 849, 1999.

Kreitman, M., Methods to detect selection in populations with applications to the human, *Annual Review of Genomics and Human Genetics*, 1, 539, 2000.

Kropp, G.L., Fucharoen, S. and Embury, S.H., Selective enzymatic amplification of alpha 2-globin DNA for detection of the hemoglobin Constant Spring mutation, *Blood*, 73, 1987, 1989.

Kulozik, A.E., Thein, S.L., Wainscoat, J.S., Gale, R., Kay, L.A., Wood, J.K., Weatherall, D.J. and Huehns, E.R., Thalassaemia intermedia: interaction of the triple alpha-globin gene arrangement and heterozygous beta-thalassaemia, *British Journal of Haematology*, 66, 109, 1987.

Kulozik, A.E., Kar, B.C., Serjeant, G.R., Serjeant, B.E. and Weatherall, D.J., The molecular basis of alpha thalassemia in India. Its interaction with the sickle cell gene, *Blood*, 71, 467, 1988.

Kurdi-Haidar, B., Mason, P.J., Berrebi, A., Ankra-Badu, G., Al-Ali, A., Oppenheim, A. and Luzzatto, L., Origin and spread of glucose-6-phosphate dehydrogenase Variant (G6PD-Med) in the Middle East, *American Journal of Human Genetics*, 47, 1013, 1990.

Kwok, P.-Y., Methods for genotyping single nucleotide polymorphisms, *Annual Review of Genomics and Human Genetics*, 2, 235, 2001.

Kyriacou, K., Kyrri, A., Kalogirou, E., Vasiliades, P., Angastiniotis, M., Ioannou, P.A. and Kleanthous, M., Hb Bart's levels in cord blood and alpha-thalassemia mutations in Cyprus, *Hemoglobin*, 24, 171, 2000.

Labie, D., Dunda-Belkhodja, O., Rouabhi, F., Pagnier, J., Ragusa, A. and Nagel, R.L., The -158 site 5' to the G gamma gene and G gamma expression, *Blood*, 66, 1463, 1985.

Labie, D. and Elion, J., Sequence polymorphisms of potential functional relevance in the beta-globin gene locus, *Hemoglobin*, 20, 85, 1996.

Lauer, J., Shen, C.J. and Maniatis, T., The chromosomal arrangement of human α-like globin genes: sequence homology and α-globin gene deletions, *Cell*, 20, 119, 1980.

Lawn, R.M., Efstradiadis, A., 'O' Connell, C. and Maniatis, T., The nucleotide sequence of human β-globin gene, *Cell*, 21, 647, 1980.

Lehman, H., Distribution of abnormal haemoglobins, *Journal of Clinical Pathology*, 9, 180p, 1956.

Lehninger, A.L., *Biochemistry*, Worth Publishers Inc, New York, 1978.

Levy, H.L. and Albers, S., Genetic screening of newborns, *Annual Review of Genomics and Human Genetics*, 1, 139, 2000.

Lie-Injo, L.E. and Jo, B.H., A fast moving haemoglobin in hydrops foetalis, *Nature*, 185, 698, 1960.

Lie-Injo, L.E., Dozy, A.M., Kan, Y.W., Lopes, M. and Todd, D., The alpha-globin gene adjacent to the gene for HbQ-alpha 74 Asp replaced by His is deleted, but not that adjacent to the gene for HbG-alpha 30 Glu replaced by Gln; three-fourths of the alpha-globin genes are deleted in HbQ-alpha-thalassemia, *Blood*, 54, 1407, 1979.

Lie-Injo, L.E., Solai, A., Herrera, A.R., Nicolaisen, L., Kan, Y.W., Wan, W.P. and Hasan, K., Hb Bart's level in cord blood and deletions of alpha-globin genes, *Blood*, 59, 370, 1982.

Lie-Injo, L.E., Cai, S.P., Wahidijat, I., Moeslichan, S., Lim, M.L., Evangelista, L., Doherty, M. and Kan, Y.W., Beta-thalassemia mutations in Indonesia and their linkage to beta haplotypes, *American Journal of Human Genetics*, 45, 971, 1989.

Liu, Y.T., Old, J.M., Miles, K., Fisher, C.A., Weatherall, D.J. and Clegg, J.B., Rapid detection of alpha-thalassaemia deletions and alpha-globin gene triplication by multiplex polymerase chain reactions, *British Journal of Haematology*, 108, 295, 2000.

Longo, F., Zecchina, G., Sbaiz, L., Fischer, R., Piga, A. and Camaschella, C., The influence of hemochromatosis mutations on iron overload of thalassemia major, *Haematologica*, 84, 799, 1999.

Lucarelli, G., Galimberti, M., Polchi, P., Angelucci, E., Baronciani, D., Giardini, C., Politi, P., Durazzi, S.M., Muretto, P. and Albertini, F., Bone marrow transplantation in patients with thalassemia, *New England Journal of Medicine*, 322, 417, 1990.

Luzzatto, L., Usanga, E.A. and Reddy, S., Glucose-6-phosphate dehydrogenase deficient red cells: resistance to infection by malarial parasites, *Science*, 164, 839, 1969.

Mackay, T.F.C., The genetic architecture of quantitative traits., *Annual Review of Genetics*, 35, 3030, 2001.

Maniatis, T., Fritsch, E.F., Lauer, J. and Lawn, R.M., The molecular genetics of human hemoglobins, *Annual Review of Genetics*, 14, 145, 1980.

Maniatis, T., Fritsch, E.F. and Sambrook, J., *Molecular Cloning: A laboratory manual*, Cold Spring Harbor Laboratory, Michigan, 1982.

Martinson, J.J., Excoffier, L., Swinburn, C., Boyce, J., Harding, R.M., Langaney, A. and Clegg, J.B., High diversity of α-globin haplotypes in a Senagalese population, including many previously unreported variants, *American Journal of Human Genetics*, 57, 1186, 1995.

McNeil, W., *Plagues and people*, Penguin Group, London, 1976.

Merryweather-Clarke, A.T., Pointon, J.J., Shearman, J.D. and Robson, K.J.H., Global prevalence of putative haemachromatosis mutations, *Journal of Medical Genetics*, 34, 275, 1997.

Merryweather-Clarke, A.T., Pointon, J.J., Jouanelle, A.M., Rochette, J. and Robson, K.J.H., Geography of *HFE* C282Y and H63D mutations, *Genetic Testing*, 4, 183, 2000.

Miles, K., Norwich, J.T., Martinson, J.J. and Clegg, J.B., Polymerase chain reaction protocols for α-globin haplotype polymorphisms, *British Journal of Haematology*, 113, 694, 2001.

Misra, R.C., Ram, B., Mohapatra, B.C., Das, S.N. and Misra, S.C., High prevalence & heterogenicity of thalassaemias in Orissa, *Indian Journal of Medical Research*, 94, 391, 1991.

Modiano, G., Morpurgo, G., Terrenato, L., Novelletto, A., Di Rienzo, A., Colombo, B., Purpura, M., Mariani, M., Santachiara-Benerecetti, S., Brega, A. Protection against malaria morbidity: near-fixation of the alpha-thalassemia gene in a Nepalese population, *American Journal of Human Genetics*, 48, 390, 1991.

Modiano, D., Luoni, G., Sirima, B.S., Simpore, J., Verra, F., Konate, A., Rastrelli, E., Olivieri, A., Calissano, C., Paganotti, G.M., D'Urbano, L., Sanou, I., Sawadogo, A., Modiano, G. and Coluzzi, M., Haemoglobin C protects against clinical Plasmodium falciparum malaria, *Nature*, 414, 305, 2001.

Morle, F., Francina, A. and Ducrocq, R., A new α-chain variant Hb Sallanches [α2.104(G11)Cys-Tyr] associated with HbH disease in one homozygous patient, *British Journal of Haematology*, 91, 608, 1995.

Nagaratnam, N., Wickramasinghe, R.L., Jayawickreme, J.S. and Maheson, V.S., HbE syndromes in a Sinhalese family, *British Medical Journal*, i, 866, 1958.

Nagaratnam, N. and Sukumaran, P.K., Thalassaemia in Ceylon, *Acta Haematologica*, 38, 209, 1967.

Nagaratnam, N., Leelawathie, P.K. and Weerasinghe, W.M., Enzyme glucose-6-phosphate dehydrogenase (G6PD) deficiency among Sinhalese in Ceylon as revealed by the methaemoglobin reduction test, *Indian Journal of Medical Research*, 57, 569, 1969.

Nagaratnam, N., Hemoglobinopathies in Sri Lanka and their anthropological implications, *Hemoglobin*, 13, 201, 1989.

Nakatsuji, T., Kutlar, A., Kutlar, F. and Huisman, T.H.J., Haplotypes among Vietnamese hemoglobin E homozygotes including one with a gamma-globin gene triplication, *American Journal of Human Genetics*, 38, 981, 1986.

Novelletto, A., Hafez, M., Deidda, G., Di Rienzo, A., Felicetti, L., el-Tahan, H., el Morsi, Z., el-Ziny, M., al-Tonbary, Y., Sittien, A. Molecular characterization of beta-thalassemia mutations in Egypt, *Human Genetics*, 85, 272, 1990.

Old, J., Haemoglobinopathies, *Prenatal Diagnosis*, 16, 1181, 1996.

Old, J. and Higgs, D.R. Gene analysis. Methods in Hematology. Churchill-Livingstone, 1983.

Old, J.M., Khan, S.N., Verma, I., Fucharoen, S., Kleanthous, M., Ioannou, P., Kotea, N., Fisher, C., Riazuddin, S., Saxena, R., Winichagoon, P., Kyriacou, K., Al-Quobaili, F. and Khan, B., A multicentre study in order to further define the molecular basis of β-thalassaemia in Thailand, Pakistan, Sri Lanka, Mauritius, Syria, and India, and to develop a simple molecular diagnostic strategy by amplification refraction mutation system-polymerase chain reaction, *Hemoglobin*, 25, 397, 2001.

Olivieri, N.F., Nathan, D.G., MacMillan, J.H., Wayne, A.S., Liu, P.P., McGee, A., Martin, M., Koren, G. and Cohen, A.R., Survival in medically treated patients with homozygous beta-thalassemia, *New England Journal of Medicine*, 331, 574, 1994.

Olivieri, N.F. and Brittenham, G.M., Iron-chelating therapy and the treatment of thalassemia, *Blood*, 89, 739, 1997.

Olivieri, N.F. and Weatherall, D.J., The therapeutic reactivation of fetal haemoglobin, *Human Molecular Genetics*, 7, 1655, 1998.

Oner, R., Altay, C., Gurgey, A., Aksoy, M., Kilinc, Y., Stoming, T.A., Reese, A.L., Kutlar, A., Kutlar, F. and Huisman, T.H., Beta-thalassemia in Turkey, *Hemoglobin*, 14, 1, 1990.

Orkin, S.H., Goff, S.C. and Hechtman, R.L., Mutation in an intervening sequence splice junction in man, *Proceedings of the National Academy of Sciences (USA)*, 78(8), 5041, 1981.

Orkin, S.H., Kazazian, H.H., Jr., Antonarakis, S.E., Ostrer, H., Goff, S.C. and Sexton, J.P., Abnormal RNA processing due to the exon mutation of beta E-globin gene, *Nature*, 300, 768, 1982a.

Orkin, S.H., Kazazian, H.H., Jr., Antonarakis, S.E., Goff, S.C., Boehm, C.D., Sexton, J.P., Waber, P.G. and Giardina, P.J., Linkage of beta-thalassaemia mutations and beta-globin gene polymorphisms with DNA polymorphisms in human beta-globin gene cluster, *Nature*, 296, 627, 1982b.

Orkin, S.H. and Kazazian, J.H.H., The mutation and polymorphism of the human beta-globin gene and its surrounding DNA, *Annual Review of Genetics*, 18, 131, 1984.

Orkin, S.H., Cheng, T.C., Antonarakis, S.E. and Kazazian, J.H.H., Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene, *Embo*, 4, 453, 1985.

Oron-Karni, V., Filon, D., Oppenheim, A. and Rund, D., Severe thalassaemia intermedia caused by interaction of homozygosity for alpha-globin gene triplication with heterozygosity for beta zero-thalassaemia, *British Journal of Haematology*, 86, 377, 1994.

Oron-Karni, V., Filon, D., Oppenheim, A. and Rund, D., Rapid detection of the common Mediterranean alpha-globin deletions/rearrangements using PCR, *American Journal of Hematology*, 58, 306, 1998.

Pacheco, P., Loureiro, P., Faustino, P. and Lavinha, J., Haplotypic heterogeneity of beta-thalassaemia IVS I-1 (G--A) mutation in southern Portugal, *British Journal of Haematology*, 94, 767., 1996.

Papassotiriou, I., Traeger-Synodinos, J., Vlachou, C., Karagiorga, M., Metaxotou, A., Kanavakis, E. and Stamoulakatou, A., Rapid and accurate quantitation of Hb Bart's and Hb H using weak cation exchange high performance liquid chromatography: correlation with the alpha-thalassemia genotype, *Hemoglobin*, 23, 203, 1999.

Parameshwaran, N., Thalassaemia, *Ceylon Medical Journal*, 12, 18, 1967.

Parikh, P., Cotton, M., Boehm, C. and Kazazian, H.H., Jr., Ethnic distribution of beta-thalassaemia in Indian subcontinent, *Lancet*, 336, 1006., 1990.

Pembrey, M.E., Weatherall, D.J., Clegg, J.B., Bunch, C. and Perrine, R.P., Haemoglobin Bart's in Saudi Arabia, *British Journal of Haematology*, 29, 221, 1975.

Perera, D.M., Gunasekara, D., Wijekoon, A. and Petrou, M., Thalassaemia control in developing countries--the Sri Lankan perspective, *Ceylon Medical Journal*, 45, 12, 2000.

Perutz, M.F., Rossman, M.G., Cullis, A.F., Muirhead, H., Will, G. and North, A.C.T., Structure of hemoglobin, *Nature*, 185, 416, 1960.

Perutz, M.F. hemoglobin structure and respiratory transport, *Scientific America*, 239, 92, 1978.

Philippon, G., Martinson, J.J., Rugless, M.J., Moulia-Pelat, J.P., Plichart, R., Roux, J.F., Martin, P.M. and Clegg, J.B., Alpha-thalassaemia and globin gene rearrangements in French Polynesia, *European Journal of Haematology*, 55, 171, 1995.

Pietrapertosa, A., Palma, A., Campanale, D., Delios, G., Vitucci, A. and Tannoia, N., Genotype and phenotype correlation in glucose-6-phosphate dehydrogenase deficiency, *Haematologica*, 86, 30, 2001.

Piperno, A., Mariani, R., Arosio, C., Vergani, A., Bosio, S., Fargion, S., Sampietro, M., Girelli, D., Fraquelli, M., Conte, D., Fiorelli, G. and Camaschella, C., Haemochromatosis in patients with beta-thalassaemia trait, *British Journal of Haematology*, 111, 908, 2000.

Pissard, S. and Beuzard, Y., A potential regulatory region for the expression of fetal hemoglobin in sickle cell disease, *Blood*, 84, 331, 1994.

Pobedimskaya, D.D., Molchanova, T.P., Streichman, S. and Huisman, T.H.J., Compound heterozygosity for two α-globin gene defects, HbTaybe (α1;38 or 39 minus Thfr) and a poly A mutation (α2; AATAAA-AATAAG), results in severe hemolytic anaemia, *American Journal of Hematology*, 47, 198, 1994.

Porter, I.H., Boyer, S.H., Watson-Williams, E.J., Adam, A., Szeinberg, A. and Siniscalco, M., Variation of glucose-6-phosphate dehydrogenases in different populations, *Lancet*, I, 895, 1964.

Premawardhena, A., Fisher, C.A., Fathiu, F., de Silva, S., Perera, W., Peto, T.E., Olivieri, N.F. and Weatherall, D.J., Genetic determinants of jaundice and gallstones in haemoglobin E beta thalassaemia, *Lancet*, 357, 1945, 2001.

Proudfoot, N.J., Gil, A. and Maniatis, T., The structure of the human zeta globin gene and a closely linked, nearly identical pseudogene, *Cell*, 31, 553, 1982.

Ratip, S., Petrou, M., Old, J.M., Wonke, B., Porter, J.B. and Modell, B., Relationship between the severity of beta-thalassaemia syndromes and the number of alleviating mutations, *European Journal of Haematology*, 58, 14, 1997.

Rees, D.C., Porter, J.B., Clegg, J.B. and Weatherall, D.J., Why are hemoglobin F levels increased in HbE/beta thalassemia?, *Blood*, 94, 3199, 1999.

Rees, W.A., Yager, T.D., Korte, J. and von Hippel, P.H., Betaine can eliminate the base pair composition dependence of DNA melting, *Biochemistry*, 32, 137, 1993.

Roberts, D.F., Papiha, S.S. and Abeyaratne, K.P., Red cell enzyme polymorphisms in Ceylon Sinhalese, *American Journal of Human Genetics*, 24, 181, 1972.

Rochette, J., Pointon, J.J., Fisher, C.A., Perera, G., Arambepola, M., Arichchi, D.S.K., Silva, S.D., Vandwalle, J.L., Monti, J.P., Old, J.M., Merryweather-Clarke, A.T., Weatherall, D.J. and Robson, K.J.H., Multicentric origin of hemochromatosis gene (HFE) Mutations, *American Journal of Human Genetics*, 64, 1056, 1999.

Rund, D., Oron-Karni, V., Filon, D., Goldfarb, A., Rachmilewitz, E. and Oppenheim, A., Genetic analysis of beta-thalassemia intermedia in Israel: diversity of mechanisms and unpredictability of phenotype, *American Journal of Hematology*, 54, 16, 1997.

Rutherford, S.L., From genotype to phenotype: buffering mechanisms and the storage of genetic information, *Bioessays*, 22, 1095, 2000.

Sampietro, M., Lupica, L., Perrero, L., Comino, A., Martinez di Montemuros, F., Cappellini, M.D. and Fiorelli, G., The expression of uridine diphosphate glucuronosyltransferase gene is a major determinant of bilirubin level in heterozygous beta-thalassaemia and in glucose-6-phosphate dehydrogenase deficiency, *British Journal of Haematology*, 99, 437, 1997.

Sancar, G.B., Tatsis, B., Cedeno, M.M. and Rieder, R.F., Proportion of hemoglobin G Philadelphia (alpha 268 Asn leads to Lys beta 2) in heterozygotes is determined by alpha-globin gene deletions, *Proceedings of the National Academy of Sciences (USA)*, 77, 6874, 1980.

Serjeant, G.R., Sommereux, A.M., Stevenson, M., Mason, K. and Serjeant, B.E., Comparison of sickle cell-beta0 thalassaemia with homozygous sickle cell disease, *British Journal of Haematology*, 41, 83, 1979.

Shaji, R.V., Chandy, S.M. and Krishnamoorthy, R., A single tube multiplex PCR method to detect the common alpha+ thalassemia alleles, *Blood*, 95, 1879, 2000.

Shimizu, K., Hashimoto, T., Harihar, S., Tajima, K., Sonoda, S. and Zaninovic, V., β-Globin gene haplotype characteristics of Colombian Amerinds in South America, *Human Heredity*, 51, 54, 2001.

Slightom, J.L., Blechl, A.E. and Smithies, O., Human G Gamma and A Gamma genes: complete nucleotide sequence suggests that DNA can be exchanged between these duplicated genes, *Cell*, 21, 627, 1980.

Smetanina, N.S. and Huisman, T.H., Detection of alpha-thalassemia-2 (-3.7 kb) and its corresponding triplication (alpha)(alpha)(alpha) (anti-3.7 kb) by PCR: an improved technical change, *American Journal of Hematology*, 53, 202, 1996.

Spritz, R.A., DeRiel, J.K., Forget, B.G. and Weissman, S.M., Complete nucleotide sequence of the human δ-globin gene, *Cell*, 21, 639, 1980.

Tagarelli, A., Piro, A., Bastone, L. and Tagerelli, G., Identification of glucose-6-phosphate dehydrogenase deficiency in a population with a high frequency of thalassaemia, *FEBS Letters*, 466, 139, 2000.

Telfer, P.T., Prestcott, E., Holden, S., Walker, M., Hoffbrand, A.V. and Wonke, B., Hepatic iron concentration combined with long-term monitoring of serum ferritin to predict complications of iron overload in thalassaemia major, *British Journal of Haematology*, 110, 971, 2000.

Thein, S.L., Wainscoat, J.S., Sampietro, M., Old, J.M., Cappellini, D., Fiorelli, G., Modell, B. and Weatherall, D.J., Association of thalassaemia intermedia with a beta-globin gene haplotype, *British Journal of Haematology*, 65, 367, 1987.

Thein, S.L., Hesketh, C., Wallace, R.B. and Weatherall, D.J., The molecular basis of thalassaemia major and thalassaemia intermedia in Asian Indians: application to prenatal diagnosis, *British Journal of Haematology*, 70, 225, 1988.

Thein SL, Hinton J., A simple and rapid method of direct sequencing using Dynabeads. *British Journal of Haematology*, 79,113, 1991.

Trent, R.J., Higgs, D.R., Clegg, J.B. and Weatherall, D.J., A new triplicated alpha-globin gene arrangement in man, *British Journal of Haematology*, 49, 149, 1981.

van der Hoek, W. and Konradsen, F., Correlation between rainfall and malaria im the dry zone of Sri Lanka, *Annals of Tropical Medicine and Parasitology*, 91, 945, 1997.

Varadaraj, K. and Skinner, D.M., Denaturants or cosolvents improve the specificity of PCR amplification of a G+C rich DNA using genetically engineered DNA polymerase, *Gene*, 140, 1, 1994.

Varawalla NY., Prenatal diagnosis and control strategies for beta-thalassaemia on the Indian subcontinent., *National Medical Journal of India*, 5, 157,1992.

Varawalla, N.Y., Old, J.M., Sarkar, R., Venkatesan, R. and Weatherall, D.J., The spectrum of beta-thalassaemia mutations on the Indian subcontinent: the basis for prenatal diagnosis, *British Journal of Haematology*, 78, 242, 1991a.

Varawalla, N.Y., Old, J.M. and Weatherall, D.J., Rare beta-thalassaemia mutations in Asian indians, *British Journal of Haematology*, 79, 640, 1991b.

Varawalla, N.Y., Fitches, A.C. and Old, J.M., Analysis of beta-globin gene haplotypes in Asian Indians: origin and spread of beta-thalassaemia on the Indian subcontinent, *Human Genetics*, 90, 443, 1992.

Vaz, F.E., Thakur, C.B., Banerjee, M.K. and Gangal, S.G., Distribution of beta-thalassemia mutations in the Indian population referred to a diagnostic center, *Hemoglobin*, 24, 181, 2000.

Venkatesan, R., Sarkar, R. and Old, J.M., beta-Thalassaemia mutations and their linkage to beta-haplotypes in Tamil Nadu in southern India, *Clinical Genetics*, 42, 251, 1992.

Verma, I.C., Saxena, R., Thomas, E. and Jain, P.K., Regional distribution of beta-thalassemia mutations in India, *Human Genetics*, 100, 109, 1997.

Vetter B, Schwarz C, Kohne E, Kulozik AE., Beta-thalassaemia in the immigrant and non-immigrant German populations., *British Journal of Haematology*, 9, 266, 1997.

Villalobos-Arambula, A.R., Bustos, R., Casas-Castenada, M., Gutierrez, E., Perera, F.J., Thein, S.L. and Iberra, B., β-Thalassaemia and β^A-globin gene haplotypes in Mexican mestizos, *Human Genetics*, 99, 498, 1997.

Viprakasit, V., Gibbons, R.J., Broughton, B.C., Tolmie, J.L., Brown, D., Lunt, P., Winter, R.M., Marinoni, S., Stefanni, M., Brueton, L., Lehmann, A.R. and Higgs, D.R., Mutations in the general transcription factor TFIIH results in β-thalassaemia in individuals with trichothiodystrophy, *Human Molecular Genetics*, 10, 2797, 2001.

Viprakasit, V., Tanphaichitr, V.S., Pung-Amritt, P., Petrarat, S., Suwantol, L., Fisher, C. and Higgs, D.R., Clinical phenotypes and molecular characterization of Hb-Pakse disease, *Haematologica*, 87, 117, 2002.

Weatherall, D.J., Abnormal haemoglobins in the neonatal period and their relationship to thalassaemia, *British Journal of Haematology*, 9, 265, 1963.
Weatherall, D.J., Clegg, J.B. and Naughton, M.A., Globin synthesis in thalassaemia: an in vitro study, *Nature*, 208, 1061, 1965.
Weatherall, D.J. and Clegg, J.B., Disordered globin synthesis in thalassemia, *Annals of New York Academy of Sciences*, 165, 242, 1969.
Weatherall, D.J., Pressley, L., Wood, W.G., Higgs, D.R. and Clegg, J.B., Molecular basis for mild forms of homozygous beta-thalassaemia, *Lancet*, 1, 527, 1981.
Weatherall, D.J. and Clegg, J.B., Thalassemia--a global public health problem, *Nature Medicine*, 2, 847, 1996.
Weatherall, D.J., The thalassaemias, *British Medical Journal*, 314, 1675, 1997a.
Weatherall, D.J. and Clegg, J.B., Genetic disorders of hemoglobin, *Seminars in Hematology*, 36, 24, 1999.
Weatherall, D.J., Single gene disorders or complex traits: lessons from the thalassaemias and other monogenic diseases, *British Medical Journal*, 321, 1117, 2000.
Weatherall, D.J., Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias, *Nature Review Genetics*, 2, 245, 2001.
Weatherall, D.J. and Clegg, J.B., *The Thalassaemia Syndromes*, Blackwell Science Ltd, Oxford, 2001.
WHO, Guidelines for the control of haemoglobin disorders. Report of th VIth annual meeting of the WHO working group on haemoglobinopathies, Geneva, 1994.
WHO, World health report, 1995.
Wickramasinghe, R.L., Ikin, E.W., Mourant, A.E. and Lehman, H., The blood groups and haemoglobins of the veddahs of Ceylon, *Journal of Anthropological Inst*, 93, 117, 1963.
Winichagoon, P., Fucharoen, S., Wilairat, P., Chihara, K. and Fukumaki, Y., Nondeletional type of hereditary persistence of fetal haemoglobin: molecular characterization of three unrelated Thai HPFH, *British Journal of Haematology*, 87, 797, 1994.
Winichagoon, P., Fucharoen, S., Kanokpongsakdi, S. and Fukumaki, Y., Detection of alpha-thalassemia-1 (Southeast Asian type) and its application for prenatal diagnosis, *Clinical Genetics*, 47, 318, 1995.
Wood, W.G., Weatherall, D.J., Hart, G.H., M.Bennett and Marsh, G.W., Hematologic changes and hemoglobin analysis in β-thalassaemia during the first year of life, *Paedatric Research*, 16, 286, 1982.
Wood, W.G., Increased HbF in adult life, *Baillieres Clinical Haematology*, 6, 177, 1993.
Yongvanit, P., Sriboonlue, P., Mularlee, N., Karnthong, T., Areejitranusorn, P., Hundreiser, J., Limberg, R., Schulze, B., Laig, M., Flatz, S.D. and Flatz, G., DNA hapoltypes and frameworks linked to the β-globin locus in an Austro-Asiatic population with a high prevalence of Hemoglobin E, *Human Genetics*, 83, 171, 1989.
Yuregir, G.T., Aksoy, K., Curuk, M.A., Dikmen, N., Fei, Y.J., Baysal, E. and Huisman, T.H., Hb H disease in a Turkish family resulting from the interaction of a deletional alpha-thalassaemia-1 and a newly discovered poly A mutation, *British Journal of Haematology*, 80, 527, 1992.
Zago, M.A., de Silva, W.A., Jr., Dalle, B., Gualandro, S., Hutz, M.H., Lapoumeroulie, C., Tavella, M.H., Araujo, A.G., Krieger, J.E., Elion, J. and Krishnamoorthy, R., Atypical beta(s) haplotypes are generated by diverse genetic mechanisms, *American Journal of Hematology*, 63, 79, 2000.
Zwick, M.E., Cutler, D.J. and Chakravarti, A., Patterns of genetic variation in mendelian and complex traits, *Annual Review of Genomics and Human Genetics*, 1, 387, 2000.

Electronic references

http://www.encyclopaedia.britannica.com
http://www.census.gov/ipc/www/worldhis.html
http:/statistics.gov.l/
http:/globin.cse.psu.edu/

Appendix 1

Equilibrium genotype frequencies will be reached in zygotes after one round of random mating. The expected genotype frequencies are predicted by the Hardy-Weinberg Equation with two alleles A and a:

$$p^2 + 2\,pq + q^2 = 1$$

where p^2 is the frequency of a homozygote trait AA

q^2 is the frequency of a homozygote trait aa

2pq is the frequency of a heterozygote trait Aa

A chi-square (χ^2) test is used to compare observed genotype frequencies to expected frequencies estimated assuming Hardy-Weinberg equilibrium:

$$\chi^2 = \text{sum} \left[\frac{(\text{oserved-expected})^2}{\text{expected}} \right]$$

Small values of χ^2 indicate a good fit to expectation and large values a deviation.

Appendix 2

Attached publications

PUBLISHED PAPERS NOT COPIED AT THE REQUEST OF THE UNIVERSITY