MOLECULAR ANALYSIS OF HEMOGLOBINOPATHIES AND CONSTRUCTION OF A DATABASE; IDENTIFICATION OF A NOVEL IVS-II-2 (T-A) MUTATION

by

Ayşe Latif

B.S. in Biology, Çanakkale Onsekiz Mart University, 2002

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Molecular Biology and Genetics
Boğaziçi University

2006

TO MY FAMILY

&

MEHMET TARFON

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis supervisor, Prof. A. Nazlı Başak, for her continuous supervision and support throughout my graduate education at Boğaziçi University.

I would like to thank Assoc. Prof. Ghazi Omar Tadmouri for introducing me with a new field of science; bioinformatics and for his valuable efforts and contributions.

I would like to extend my thanks to Prof. Günhan Erdem and Göksel Erdem for the support and encouragement they provided throughout my undergraduate and graduate years.

I would like to thank Resul I. Gülen for sharing his previous experience about FileMaker Pro 6; Erhan Abay for his great support for construction of the NDAL Web page; Canan Doğanlı for her help during data entry into the databases; Rezan Fahrioğlu Yamacı and Göksan Yamacı for their great encouragement, support and friendship; Drs Baki Yokeş, Mehmet Ozansoy and Nagehan Ersoy, and my friends Nıvart Taşçı, İnanç Değer Fidancı, Caroline S. Pirkevi and Pınar Gencer for their assistance during different stages of this thesis; Esra Soydan for sharing her experience about molecular analysis of β-thalassemia; and Dr. Onur Bilenoğlu from Iontek Ltd. for his collegial support throughout this study. I express my deep gratitude to my friends in our laboratory, especially to R. Mine Güzel, F. Sinem Hocaoğlu, Nazan Saner, Özlem Yalçın, and Aslıhan Özoğuz.

I would like to thank all members of Boğaziçi University, Molecular Biology and Genetics Department for their support and friendship.

I am especially grateful to my family for their moral support and patience throughout my graduate years.

ABSTRACT

MOLECULAR ANALYSIS OF HEMOGLOBINOPATHIES AND CONSTRUCTION OF A DATABASE; IDENTIFICATION OF A NOVEL IVS-II-2 (T-A) MUTATION

Hemoglobinopathies, inherited disorders of the hemoglobin (Hb) molecule, display great molecular and phenotypic diversity. Beta-thalassemia major is considered as the most severe form of hemoglobinopathies and requires life-long blood transfusions to maintain required Hb levels. Although there are improvements in the treatment of β-thalassemia, there is not any definitive cure yet. Thus, preventive programs must be introduced in order to decrease the probability of affected births. In order to apply effective prenatal diagnosis, it is important to define the molecular basis of a specific population. In the framework of this study, 163 individuals were screened for the mutations in the β-globin gene. The number of β-thalassemia alleles was 193; 89 alleles derived from β-thalassemia minor and 104 were from 52 β-thalassemia major patients. The methods used were β-Globin StripA^{ssay}, based on reverse dot-blot hybridization and genomic sequencing. By applying these two techniques, all β-thalassemia alleles were defined, and a total of 24 mutations were described. The results indicate that, unlike other Mediterranean countries, Turkey is very heterogeneous at molecular level. The main cause of this heterogeneity is thought to be the presence of diverse ethnic groups in Turkey, due to its geographical location and rich historical past. Another objective of this thesis was the construction of a database to easily manage the huge amount of patient data in our laboratory. FileMaker Pro provides advantages (short development time, ease of modification and easy access via LAN or World Wide Web) especially for people not involved in bioinformatics, e.g. biologists working in a wet lab. Powerful features, broad platform support, and easy-to-use interface make FileMaker Pro 6 an essential tool for creating and sharing databases.

ÖZET

HEMOGLOBİNOPATİLERİN MOLEKÜLER ANALİZİ VE HEMOGLOBİNOPATİ VERİTABANI OLUŞTURULMASI; YENİ BİR IVS-II-2 (T-A) MUTASYONUNUN TANIMLANMASI

Hemoglobin (Hb) molekülünün kalıtsal hastalıkları olan hemoglobinopatiler, moleküler ve fenotipik açıdan büyük çeşitlilik gösterir. Beta-talasemi majör, hemoglobinopatilerin en ağır seyreden formudur, hasta bireyler ihtiyaçları olan Hb miktarının karşılanması için hayat boyu düzenli kan transfüzyonuna bağımlıdırlar. Beta-talaseminin tedavisinde birçok gelişme kaydedilmiş olmasına rağmen, hastalığın henüz kesin tedavisi yoktur. Bu nedenle, önleyici programlar uygulanması ve β-talasemi majörlü doğumların engellenmesi önem kazanmaktadır. Doğum öncesi tanının etkin bir şekilde uygulanması için, toplumlara özgü moleküler yapının iyi tanımlanmış olması gerekir. Bu çalışma çerçevesinde, 163 bireyin β-globin geni, hastalığa neden olabilecek mutasyonlar için tarandı. Çalışmada icelenen 193 β-talasemi alelinden 89'u β-talasemi taşıyıcılarına, 104'ü ise 52 β-talasemi hastasına aitti. Bu çalışmada kullanılan yöntemler β-Globin StripA^{ssay} ve DNA dizi analizidir. Söz konusu iki yöntem uygulandığında, tüm β-talasemi alelleri belirlendi ve toplam 24 mutasyon tanımlandı. Bu sonuçlar, diğer Akdeniz ülkelerinden farklı olarak Türkiye'nin moleküler düzeyde oldukça heterojen bir yapı sergilediğini göstermektedir. Gözlenen moleküler çeşitlilik, Türkiye'nin coğrafi konumundan ve zengin tarihinden kaynaklanan farklı etnik grupların varlığı ile açıklanmaktadır. Bu tezin ikinci amacı, β-talasemi için bir veritabanı oluşturarak, laboratuvarımızda toplanan çok sayıda hasta bilgisini kolay ulaşılabilir kılmaktır. Bu amaçla kullanılan FileMaker Pro programı sağladığı avantajlarla (kısa sürede veritabanı oluşturulabilmesi, modifikasyon kolaylığı ve LAN veya World Wide Web yoluyla kolay erişim), biyoinformatik alanına uzak olan kişiler için de kullanım kolaylığı sağlamaktadır; örneğin laboratuvarda çalışan biyologlar gibi. FileMaker Pro 6'nın güçlü özellikleri, geniş işletim sistemi desteği ve kullanım kolaylığı, bu programı veritabanı oluşturulması ve paylaşımında ön sıraya taşımaktadır.

TABLE OF CONTENTS

ACKNOWLEDGMENTS ... iv
ABSTRACT ... v
ÖZET ... vi
LIST OF FIGURES ... xi
LIST OF TABLES ... xiv
LIST OF ABBREVIATIONS ... xvi
1. INTRODUCTION ... 1
1.1. The Human Globin Gene Clusters ... 1
1.1.1. The Human α-Globin Gene Cluster ... 2
1.1.2. The Human β-Globin Gene Cluster ... 2
1.2. Hemoglobin Switching ... 4
1.2.1. Molecular Mechanism of Hb Switching ... 5
1.3. Hemoglobinopathies ... 6
1.3.1. Abnormal Hemoglobins ... 6
1.3.2. Thalassemia Syndromes ... 8
1.3.2.1. α-Thalassemia ... 8
1.3.2.2. HPFH and δβ-thalassemias ... 9
1.4. Beta-Thalassemia ... 10
1.4.1. The β-Globin Gene ... 10
1.4.2. Mutations Causing β-Thalassemia ... 12
1.4.3. Clinical Syndromes of β-Thalassemia ... 14
1.4.4. Pathophysiology of β-Thalassemia ... 16
1.4.5. Treatment of β-Thalassemia ... 16
1.4.6. Beta-Thalassemia in Turkey ... 18
1.5. Bioinformatics ... 19
1.5.1. History of Bioinformatics ... 20
1.5.2. Computers in Bioinformatics ... 20
1.6. Database Management Systems ... 22
1.6.1. Database Systems ... 24
1.7. The Globin Gene Server ... 27

1.7.1. The Syllabus of Human Hemoglobin Variants ... 27
1.7.2. The Syllabus of Thalassemia Mutations ... 27
1.7.3. The Hb Variants Database ... 27
1.7.4. Genomic Sequences Alignment and Annotations Database ... 28
1.7.5. Experimental Results on Gene Expression ... 28
1.8. Deniz: A Database Constructed by FileMaker Pro 6 Software ... 29
1.8.1. FileMaker Pro 6 Software ... 29
2. PURPOSE ... 31
3. MATERIALS ... 32
3.1. Peripheral Blood and Chorionic Villus Samples ... 32
3.2. Buffers and Solutions ... 32
3.2.1. DNA Extraction from Peripheral Blood Samples ... 32
3.2.2. DNA Extraction from CVS Samples ... 33
3.2.3. Agarose Gel Electrophoresis ... 33
3.2.4. Polymerase Chain Reaction (PCR) ... 33
3.3. Enzymes ... 34
3.4. Oligonucleotide Primers ... 34
3.4.1. Beta-Globin Gene Amplification and DNA Sequencing ... 34
3.4.2. Analysis of Variable Number of Tandem Repeats (VNTR) ... 34
3.4.3. Differential Amplification of Gγ-Globin Gene Promoter ... 34
3.5. PCR Purification Kit ... 35
3.6. Beta-Globin StripA^{ssay} Kit ... 35
3.7. FileMaker Pro 6 ... 35
3.8. Equipment ... 35
4. METHODS ... 37
4.1. DNA Extraction from Peripheral Blood Samples ... 37
4.1.1. Sodium Chloride (NaCl) Extraction ... 37
4.1.2. Nucleic Acid Purification with MagNa Pure Compact Instrument ... 38
4.2. DNA Extraction from Chorionic Villus Samples ... 38
4.3. Analysis of the Extracted DNA Samples ... 38
4.3.1. Qualitative Analysis by Agarose Gel Electrophoresis ... 38
4.3.2. Quantitative Analysis by Spectrophotometric Measurement ... 39

4.4. The β-Globin StripA^{ssay} ... 39
4.4.1. DNA Isolation ... 40
4.4.2. Amplification and Biotin Labeling ... 41
4.4.3. Hybridization and Color Development ... 41
4.5. Polymerase Chain Rection ... 43
4.5.1. The β-Globin Gene ... 44
4.5.2. Variable Number of Tandem Repeats ... 45
4.5.3. Differential Amplification of the ^{G}γ-Globin Gene Promoter ... 46
4.6. PCR Purification ... 48
4.7. Restriction Endonuclease Analysis ... 48
4.8. FileMaker Pro 6 Software ... 49
4.8.1. Construction of Databases using FileMaker Pro 6 ... 50
4.8.1.1. Patients Database ... 51
4.8.1.2. Referring Clinicians Database ... 53
5. RESULTS ... 54
5.1. Genomic DNA Samples ... 54
5.1.1. DNA Samples Extracted with NaCl Method ... 54
5.1.2. DNA Samples Extracted with the MagNA Pure Compact Instrument ... 54
5.2. Molecular Analysis of the β-Globin Gene ... 55
5.2.1. Beta-Globin Gene StripA^{ssay} Results ... 55
5.2.2. Beta-Globin Gene Amplification and Purification Results ... 58
5.2.3. Direct DNA Sequencing Results ... 59
5.2.3.1. IVS-II-2 (T-A): Family M. ... 59
5.2.3.2. Codon37 (TGG-TGA): Family B. ... 60
5.2.3.3. Codon15 (TGG-TGA): Patient S. and Family O. ... 62
5.2.3.4. Codon30 (AGG-ACG): Family E. ... 63
5.2.3.5. IVS-II-848 (C-A): Family G. ... 64
5.2.3.6. -28 (A-C): Patient H.A. ... 65
5.2.3.7. Hb Ankara (Cd10 GCC-GAC): Patient G.Y. ... 66
5.2.3.8. HbO Arab: Patient A.K. ... 66
5.2.3.9. HbD LosAngeles: Family AN. ... 67
5.2.4. Restriction Endonuclease Analysis to Confirm Sequencing Results ... 68
5.3. VNTR Analysis Results ... 69

5.4. Distribution of Turkish β-Thalassemia Mutations ... 70
5.5. Molecular Analysis of Gγ-Globin Gene Promoter Polymorphisms ... 71
5.6. Construction of Databases by Using FileMaker Pro 6 Software ... 71
5.6.1. Patients Database Results ... 71
5.6.2. Referring Clinicians Database Results ... 75
6. DISCUSSION ... 76
6.1. Molecular Analysis of the β-Globin Gene ... 76
6.2. Distribution of β-Thalassemia Mutations ... 76
6.3. Phenotypic Diversity among β-Thalassemia Chromosomes ... 80
6.4. FileMaker Pro 6 Software ... 81
6.4.1. Patients Database ... 83
6.4.1.1. Personal Information Layout ... 83
6.4.1.2. Background and Molecular Analysis Layout ... 83
6.4.1.3. Sample Information Layout ... 84
6.4.1.4. Distribution Layout ... 84
6.4.2. Referring Clinicians Database ... 84
7. CONCLUSIONS ... 86
7.1. NDAL Web Page and Construction of Additional Databases ... 86
APPENDIX A: MOLECULAR DIAGNOSIS RESULTS ... 87
APPENDIX B: PROTOCOL LAYOUTS DESIGNED FOR HEMOGLOBINOPATHY DATABASE ... 92
REFERENCES ... 96

LIST OF FIGURES

Figure 1.1. The hemoglobin molecule 1

Figure 1.2. Schematic representation of the human β-globin gene cluster 3

Figure 1.3. Hemoglobin switching 4

Figure 1.4. Microscopic view of normal and sickle cell erythrocytes 7

Figure 1.5. Worldwide distribution of β-thalassemia alleles 11

Figure 1.6. Location of mutations causing β-thalassemia 14

Figure 1.7. Data models 24

Figure 1.8. Database Nomenclature 26

Figure 4.1. Schematic representation of amplification products 41

Figure 4.2. Decoder table and interpretation of the β-Globin StripA^{ssay} results 43

Figure 4.3. "Define fields" dialog box in FileMaker Pro 6 50

Figure 5.1. Genomic DNA samples of different individuals 54

Figure 5.2. DNA samples run on a one per cent gel 54

Figure 5.3. Agarose gel electrophoresis of PCR products 55

Figure 5.4. Examples of β-Globin stripA^{ssay} results 56

Figure 5.5. Examples of β-Globin stripA^{ssay} results ..... … 57

Figure 5.6. β-globin gene amplification ..... … 58

Figure 5.7. Gel electrophoresis of purified products ..... … 58

Figure 5.8. Pedigree of family M ..... … 59

Figure 5.9. Sequencing results of the probands ..... … 60

Figure 5.10. Pedigree of family B ..... … 61

Figure 5.11. β-globin StripA^{ssay} results of the mother and the proband ..... … 61

Figure 5.12. Sequencing results of family B ..... … 62

Figure 5.13. Pedigree of family O ..... … 63

Figure 5.14. Sequencing results of patient A.F.S and family O ..... … 63

Figure 5.15. Pedigre of family E ..... … 64

Figure 5.16. Sequencing results of family E ..... … 64

Figure 5.17. Pedigree of family G ..... … 65

Figure 5.18. Sequencing results of family G ..... … 65

Figure 5.19. Sequencing result of patient H.A. ..... … 66

Figure 5.20. Sequencing result of patient G.Y. ..... … 66

Figure 5.21. Sequencing result of the patient A.K. ..... … 67

Figure 5.22. Pedigree of family of AN. … 67

Figure 5.23. Sequencing results of the proband … 68

Figure 5.24. Restriction endonuclease analysis results for patient A.K. and family AN... 68

Figure 5.25. VNTR analysis results of St 14 and XY probes … 69

Figure 5.26. Agarose gel electrophoresis of three different patients … 71

Figure 5.27. Structure of patients database … 72

Figure 5.28. Personal information layout … 72

Figure 5.29. Background and molecular analysis layout … 73

Figure 5.30. Sample information layout … 74

Figure 5.31. Referring clinicians database layout … 75

Figure 6.1. Regional distribution of most common mutations in Turkey … 79

Figure B.1. β-globin gene amplification protocols … 92

Figure B.2. Restriction endonuclease analysis protocol … 93

Figure B.3. VNTR analysis protocols … 93

Figure B.4. Distribution layout … 94

Figure B.5. Distribution layout (continued) … 95

LIST OF TABLES

Table 1.1. History of bioinformatics 21

Table 1.2. Application areas of computers in bioinformatics 22

Table 1.3. Public bioinformatics databases accessible via the World Wide Web 25

Table 3.1. Equipment used in the framework of this thesis 36

Table 4.3. Primer pairs used to amplify the β-globin gene 44

Table 4.4. Reaction conditions for the β-globin gene amplification 44

Table 4.5. Primers used for VNTR analyses 45

Table 4.6. Reaction conditions for VNTR analyses 46

Table 4.7. Primers used in Gγ-globin gene promoter amplification 47

Table 4.8. Reaction conditions of Gγ-globin gene promoter amplification 47

Table 4.9. Modes of FileMaker 49

Table 4.10. Field types in FileMaker 50

Table 4.11. Layouts, fields and field types used in construction of patients database. 52

Table 4.12. Scripts used in patients database construction 53

Table 5.1. Prenatal diagnosis results 69

Table 5.2. Frequency of Turkish β-globin gene mutations found in this study ..... … 70

Table 6.1. Most frequent β-thalassemia mutations in Turkey ..... … 77

Table 6.2. The first eight mutations defined in this study ..... … 78

Table 6.3. The nature of some mutations identified in the framework of this thesis. … 81

Table 6.4. Brief comparison of FileMaker and Microsoft Access ..... … 82

Table A.1. Molecular diagnosis results in the year 2002 ..... … 87

Table A.2. Molecular diagnosis results in the year 2003 ..... … 88

Table A.3. Molecular diagnosis results in the year 2004 ..... … 89

Table A.4. Molecular diagnosis results in the year 2005 ..... … 90

Table A.5. Molecular diagnosis results in the year 2006 ..... … 91

LIST OF ABBREVIATIONS

A	Adenine
AP1	Ubiquitious Transcription Factor
ATG	Initiation codon
β-LCR	Beta-Locus Control Region
bp	Base-pair
C	Cytosine
Cd	Codon
CDML	Claris Dynamic Markup Language
CVS	Chorionic Villus Sample
dbERGE	Experimental Results on Gene Expression Database
DBMS	Database Management System
dNTP	2'-Deoxynucleoside 5'-Triphosphate
EDTA	Ethylene Dinitrilo Tetra Acetate
EKLF	Erythroid Krüppel-like Factor
EtBr	Ethidium Bromide
EtOH	Ethanol
FSC	Frame Shift Codon
G	Guanine
GALA	Genomic Sequence Alignment and Annotations Database
GATA-1	GATA-binding Erytroid Transcription Factor
GB	Gigabyte
Hb	Hemoglobin
HbA	Major Adult Hemoglobin
HbA_2	Minor Adult Hemoglobin
HbF	Fetal Hemoglobin
HbVar	Hemoglobin variant Database
HGB	Human Genome Browser
HPFH	Hereditary Persistance of Fetal Hemoglobin
HTML	Hypertext Markup Language
HU	Hydroxyurea Treatment

HS	Hypersensitive sites
IT	Information Technology
IVS	Intervening Sequence
JDBC	Java Database Connectivity
Kb	Kilobase
kDA	Kilodalton
LAN	Local Area Network
M	Molar
MB	Megabyte
MCH	Mean Cell Hemoglobin
MCV	Mean Cell Volume
$MgCl_2$	Magnesium Chloride
mRNA	Messenger Ribonucleic Acid
NCBI	National Center for Biotechnology Information
NF-E2	Nuclear Factor-Erythroid 2
nm	Nanometer
OD_{260}	Optical Density at 260 nm
OD_{280}	Optical Density at 280 nm
ODBC	Open Database Connectivity
PCR	Polymerase Chain Reaction
rpm	Revolution per minute
SDS	Sodium Dodecyl Sulphate
SQL	Structured Query Language
T	Thymine
TBE	Tris-Boric Acid
TE	Tris EDTA
Tris	Tris (hydroxymethyl)-aminomethan
U	Unit
URL	Uniform Resource Locator
UTR	Untranslated region
UV	Ultra Violet
V	Volt
VNTR	Variable Number of Tandem Repeats

XML	Extensible Markup Language
XPD	Helicase protein

1. INTRODUCTION

Erythrocyte, a fundamental cell type of humans, is produced via a process called erythropoiesis occurring in hematopoietic organs. Erythrocytes are essential because they are the only cell type carrying oxygen to tissues required for cellular respiration. Their oxygen carrying capability is provided by hemoglobin molecules. There are ten million hemoglobin molecules in a single erythrocyte comprising 95 per cent of them, and each hemoglobin molecule has a molecular weight of 64 kDa. Hemoglobin is a tetrameric protein composed of two α-like and two β-like globin chains each bound to a heme moiety (Figure 1.1) (Brittain, 2002).

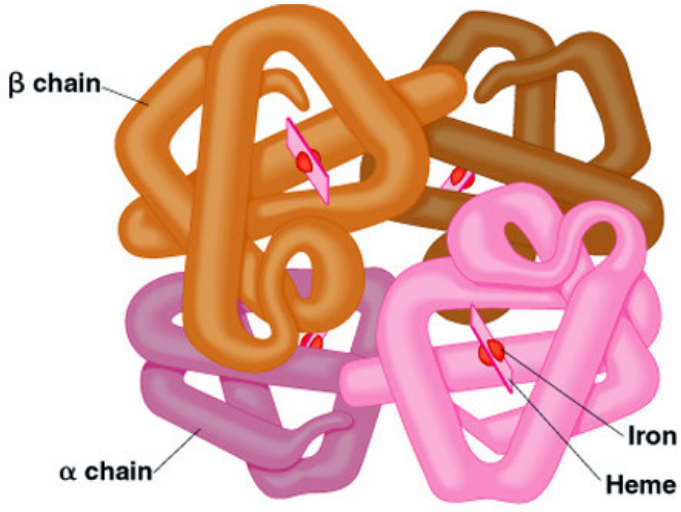


Figure 1.1. The hemoglobin molecule
(adapted from http://www.people.virginia.edu/~rjh9u/hemoglob.html)

1.1. The Human Globin Gene Clusters

The human globin gene clusters, forming the α- and β-like homodimers of hemoglobin are evolved from an ancestral myoglobin gene approximately 800 million years ago, via a series of duplication events and sequence divergences. The two clusters are separated from each other as α- and β-globin clusters around 450 million years ago (Brittain, 2002).

1.1.1. The Human α-Globin Gene Cluster

The human α-globin gene cluster is located on the short arm of chromosome 16 (16p.13.3) and lies very close to the telomere (approximately 150 kb). The cluster includes an embryonic gene (ζ2), three pseudo genes (ψζ1, ψα1, ψα2), a gene with unknown function (θ1) and two duplicated α-genes (α2 and α1). They are arranged in their developmental expression order 5'-ζ2-ψζ1-ψα2-ψα1-α2-α1-3' and cover around 30 kb long (Bowden *et al.*, 1992; Brittain, 2002; Higgs, 2004). The ζ2- and α-globin genes diverged from each other around 300 million years ago and display 59 per cent amino acid identity. The α2- and α1- genes have approximately 99 per cent similarity with each other and are therefore considered as duplication products. They produce identical proteins because they only differ in their non-coding sequences (intron 2 and 3'-UTR). The polypeptide chain encoded from the α-globin gene cluster is 141 amino acids long (Albitar *et al.*, 1992; Brittain, 2002).

There are erythroid-specific DNaseI hypersensitive sites (HSs) 5' to the ζ2-globin gene. The HS playing an important regulatory role on expression of the α-globin cluster is the HS-40 and located 40 kb upstream of the ζ2-globin gene (Craddock *et al.*, 1995). Sequencing of HS-40 revealed that it is composed of DNA binding sites for erythroid-specific and ubiquitous transcription factors: GATA-1, NF-E2/AP1 and CACC box proteins (Huang *et al.*, 1998). The α-globin gene cluster is adjacent to widely expressed genes, and HS-40 is within an intron of one of these genes. Therefore, this cluster has a more open chromatin conformation than the β-globin cluster in non-erythroid cells. It is demonstrated that HS-40 has some ability to form an open chromatin structure, but the suggested primary function is acting as a strong enhancer. (Craddock *et al.*, 1995; Brittain, 2002)

1.1.2. The Human β-Globin Gene Cluster

The human β-globin gene cluster is located on the short arm of chromosome 11 (11p15.5). This cluster includes six structural genes: an embryonic gene, ε-; two fetal genes, ^{A}γ- and ^{G}γ-; a pseudo gene, ψβ-; a minor and a major adult gene; δ- and β-globin genes. Similar to the α-globin cluster, the genes are arranged according to their

developmental expression order: 5'-ε-^{A}γ-^{G}γ-ψβ-δ-β-3' and the cluster spans 70 kb. (Figure 1.2) (Ho and Thein 2000, Brittain, 2002; Stamatoyannopoulos, 2005). The polypeptide chain encoded from the ε-globin gene displays 75 per cent amino acid identity with the β-globin chain. The γ-globin genes are found as duplicates in the cluster, and their products differ from each other by an amino acid: ^{G}γ-globin synthesizes glycine at Codon 136 while it is replaced with alanine in ^{A}γ-globin chain. The most identical globin chain is synthesized by the δ-globin gene and shows 93 per cent amino acid identity to the β-chain. Therefore, the δ-globin gene is considered as the most recent divergent of the β-globin gene. The polypeptide chain produced from the β-globin gene cluster is 146 amino acids long (Brittain, 2002).

The β-Locus Control Region (β-LCR) is defined as the positive regulatory element lying upstream of the β-globin gene cluster (Thein, 2004; Thein 2005b, Stamatoyannopoulos, 2005). It is composed of one ubiquitous and four erythroid-specific DNaseI HSs (HS5 and HS 1-4, respectively) distributed 6-20 kb upstream of the ε-globin gene. Each of the hypersensitive components of β-LCR contains highly conserved multiple copies of three motifs which represent binding sites for erythroid-specific transcription factors; EKLF, GATA-1, and NF-E2/AP1 (Ho and Thein, 2000; Fang *et al.*, 2005). In non-eryhtroid cells, the β-globin gene cluster is surrounded with a late replicating, DNaseI resistant and transcriptionally inactive chromatin domain (Craddock *et al.*, 1995). The β-LCR establishes a euchromatin structure around the β-globin cluster during erythroid differentiation by initiating a pathway for chromatin decondensation. There is one other HS approximately 20 kb downstream to β-globin gene. These two HSs are suggested as the markers of β-globin gene cluster boundary (Figure 1.2) (Thein, 2004).

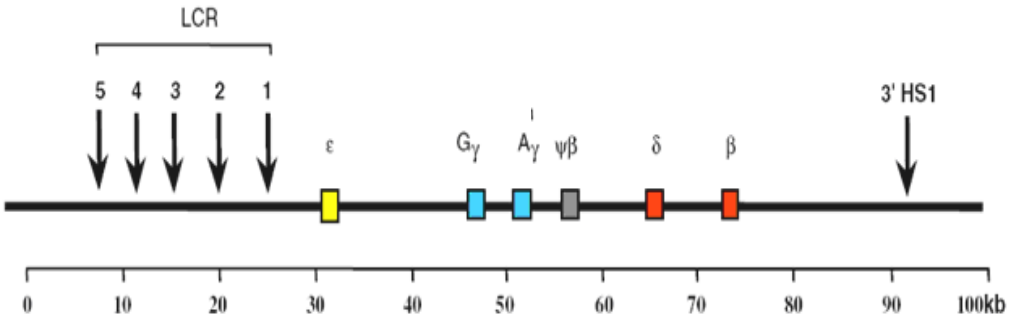


Figure 1.2. Schematic representation of the human β-globin gene cluster, arrows indicate location of HSs. (Stamatoyannopoulos, 2005)

1.2. Hemoglobin Switching

Developmental changes in hemoglobin production were discovered in the late 19th century by identification of alkali-resistant and alkali-sensitive structures of newborn and adult hemoglobins, respectively (Stamatoyannopoulos, 2005). Synthesis of different hemoglobin (Hb) types is induced by developmentally directed Hb switching (Figure 1.3). The name of the Hb produced is characterized according to tissue and developmental stage specific regulation of α- and β-globin clusters at transcriptional level. In first six to eight weeks of gestation the main site of hematopoiesis is the yolk sac and the Hbs produced are; Hb Gower I ($\zeta_2\varepsilon_2$), Hb Gower II ($\alpha_2\varepsilon_2$) and Hb Portland ($\zeta_2\gamma_2$). From about eight weeks to six months, hematopoiesis mainly occurs in fetal liver and spleen, and the dominant product of this period is HbF ($\alpha_2\gamma_2$). Bone marrow is the major site of Hb production from six months of gestation through adult life. In adult life, three types of Hbs are observed in different proportions; HbF ($\alpha_2\gamma_2$) lower than one per cent, HbA_2 (minor Hb; $\alpha_2\delta_2$) lower than three per cent and HbA (major Hb; $\alpha_2\beta_2$) approximately 97 per cent. Three to six months after birth the switching process is completed, and HbF accounts for less than five per cent of total Hb. This proportion falls during childhood to one per cent which is referred as adult Hb level (Wood, 1993; Weatherall, 2004b).

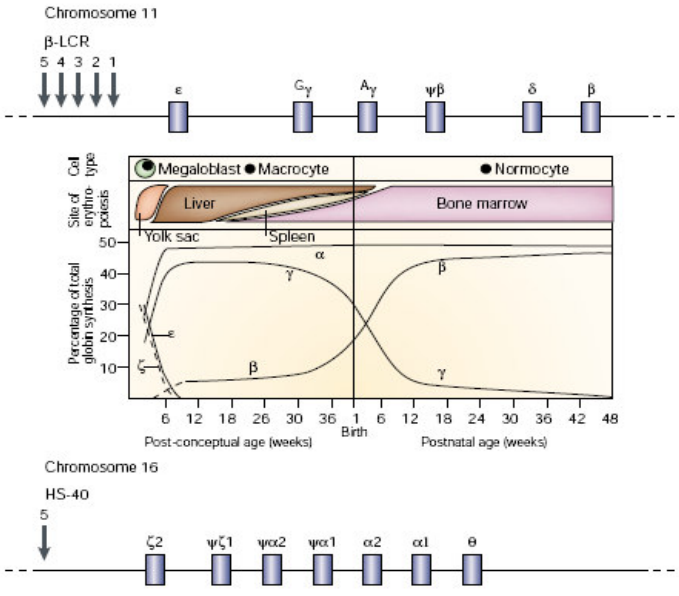


Figure1.3. Hemoglobin switching: site of hematopoiesis and globin chains produced during each stage of development (Weatherall, 2001)

1.2.1 Molecular Mechanism of Hb Switching

Tissue and developmental stage-specific expression of globin genes is achieved by interaction between gene promoters and β-LCR in the β-globin cluster or HS-40 in the α-globin cluster. The switching mechanism in the α-globin gene cluster is much simpler than the switch in the β-cluster. In the α-globin gene cluster there is only one developmental switch: embryonic to fetal/adult. Initial activation of embryonic ζ2-gene requires interaction of GATA-1 and AP1 with the HS-40 (Brittain, 2002). The embryonic to fetal/adult switch in the α-globin cluster is maintained by autonomous silencing of ζ2-gene. According to Liebhaber *et al.*, silencing is mediated via contribution of silencers (such as NF-E2) to ζ2-promoter. After silencing of the proximal ζ2-gene, α-globin genes are activated and they pre-dominate the globin chains synthesized during fetal and adult life (Liebhaber *et al.*, 1996; Huang *et al.*, 1998).

Unlike the α-globin gene cluster, there are two switches in the β-globin cluster: embryonic to fetal and fetal to adult. There are various models proposed for explaining tissue and developmental stage-specific expression of β-globin genes. Amongst the models proposed (Looping, tracking, facilitated tracking, and linking model), the looping model is the most favorable one (Stamatoyannopoulos, 2005). According to this model, transcription factors bind to β-LCR and form a complex named as holocomplex. Gene activation is mainly mediated by the interactions between the holocomplex and gene promoters (Bulger and Groudine, 1999). The holocomplex interacts with only one promoter at a time, and it can "flip-flop" between promoters dependent on the autonomous silencing of the ε-globin gene and gene competition between γ- and β-globin genes. Autonomous silencing of the ε-globin gene is mediated via the combinatorial effect of transcription factors in the formation of a silencing complex. Conserved sequences required for the formation of the silencing complex are found upstream of the gene promoter. The silencing complex disturbs the interaction between β-LCR and the ε-globin gene. Later, the competition between γ- and β-globin genes is determined according to the presence or absence of developmental stage-specific *trans*-acting factors, identifying the gene expressed in the cluster. In embryonic and fetal stage, the transcriptional environment favors the expression of γ-globin genes and results in silencing of β-globin gene and vice versa in adults (Behringer, 1998; Bulger and Groudine, 1999; Stamatoyannopoulos, 2005).

In order to have adequate amounts of hemoglobin (HbF lower than one per cent, HbA_2 lower than three per cent and HbA approximately 97 per cent) for metabolism in each developmental stage, there must be an expression balance between these two clusters. The molecular defects disturbing this balance affect synthesis of adequate globin chains from both clusters. This imbalance results in a group of inherited disorders named as hemoglobinopathies.

1.3. Hemoglobinopathies

Hemoglobinopathies are inherited disorders of Hb molecule caused by defective and imbalanced globin gene expression. They are the most common and oldest monogenic disorders worldwide and prevalent in regions previously endemic for malaria. Today, due to recent global migrations it is possible to observe people suffering from hemoglobinopathies also in continental Europe, North and South America. They are inherited in Mendelian recessive fashion and heterozygotes have advantage against *Plasmodium falciparum*. Heterozygote advantage identifies the reason why hemoglobinopathies are prevalent in malarial endemic regions as suggested in 1949 by Haldane. Hemoglobinopathies can be categorized into two major groups: Abnormal hemoglobins and thalassemia syndromes (Tüzmen and Schechter, 2001; Weatherall, 2004b; Thein, 2005b)

1.3.1 Abnormal Hemoglobins

Abnormal Hbs are structural variants of globin chains resulting from different molecular alterations. There are more than 1000 structural variants of Hb molecule resulting from amino acid substitutions, deletions, insertions or chain fusions. Most of these variants are functionally normal and are therefore considered as clinically silent. But some of them will cause thalassemia phenotype due to their reduced synthesis rate or rapid destruction which results from their unstable structure (Ho and Thein, 2000). The most prevelant Hb variants worldwide are HbS, HbE and HbC (Weatherall, 2001).

HbS is the first Hb variant discovered in 1949 by Pauling *et al.* which shows a wide prevalence throughout Africa, Middle East, Mediterranean region and certain regions of

India. HbS is caused by the A to T transversion replacing glutamic acid with valine at sixth codon of the β-globin chain. Low oxygen levels will lead to polymerization of deoxyhemoglobin, and this change causes sickling of red blood cells (Figure 1.4). Sickle cells will occlude in small blood vessels and cause "crisis". Crisis involves a wide spectrum of events from painful episodes to massive thrombosis which will lead to infraction in bone, lung or brain. Individuals heterozygous for the HbS allele (HbS trait) do not have any significant life-threatening symptoms, while homozygotes for the HbS allele (Sickle Cell Anemia) suffer from lifelong hemolytic anemia. Besides, coinheritance of β-thalassemia and HbS trait can be observed with ranging severities, mainly determined according to the β-thalassemia allele (OMIM, #603903).

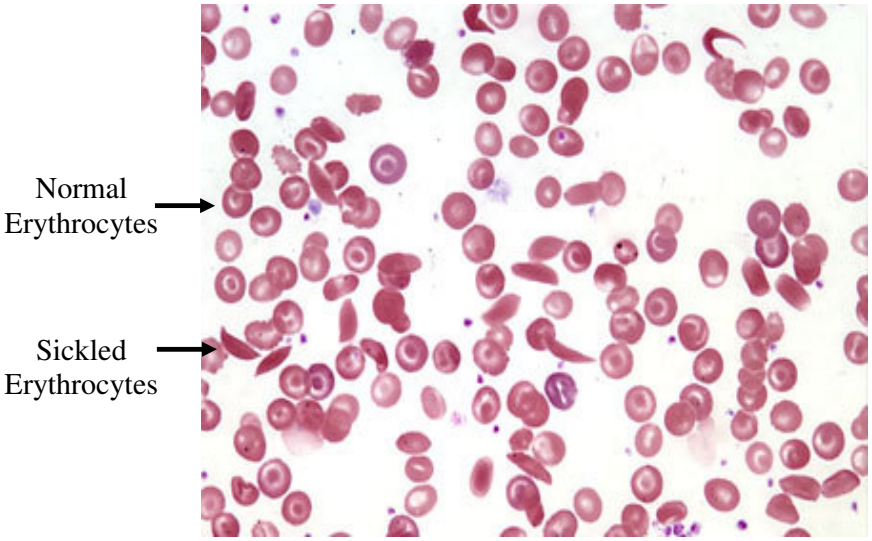


Figure1.4. Microscopic view of normal and sickle cell erythrocytes
(http://www.microscopyu.com/galleries/pathology/sicklecellanemia.html)

HbE is estimated as the most common hemoglobin variant especially in South-East Asia. It is caused by a mutation substituting glutamic acid to lysine at Codon 26 of the β-chain. Like HbS trait, individuals heterozygote for this substitution do not have any significant clinical abnormality while homozygotes are slightly anemic. Coinheritance of the β^0-thalassemia allele with HbE will result in a more severe phenotype than coinheritance of β^+-thalassemia allele. These two conditions could be observed with different severities ranging from mild β-thalassemia major to a mild β-thalassemia intermedia phenotype (Das and Talukder, 2002; Thein, 2004).

HbC is caused by a mutation in Codon six replacing glutamic acid with lysine. HbC is prevalent in West Africa and parts of Mediterranean region. Heterozygotes carrying the HbC allele have no clinical abnormality while homozygotes display a mild hemolytic anemia together with spleen enlargement (splenomegaly). If HbC is coinherited with a β-thalassemia allele, the phenotype observed is similar to homozygous inheritance of HbC phenotype (OMIM, +141900).

1.3.2. Thalassemia Syndromes

The first definition of thalassemia was independently introduced to public in 1925 by two researchers from United States (Cooley and Lee), and Italy (Rietti,Greppi and Micheli) (Weatherall, 2004a). In 1932, Whiple invented the name "Thalassemia", according to the geographical background of the early cases described, using the Greek words "thalassa" and "anemia" meaning, sea and blood, respectively (Weatherall, 2004b). Thalassemias are the world's most common single gene disorders causing a major public health problem in malarial endemic regions: the Mediterranean area, the Middle East, the Indian subcontinent, the Far East, tropical Africa and the Caribbean. It has been estimated that worldwide there are 270 million carriers of mutant globin alleles, and every year approximately 400,000 severely affected individuals are born. More than 95 per cent of these affected births occur in Asia, India, and the Middle East (Tüzmen and Schechter, 2001; Higgs, 2004).

Thalassemia syndromes are a diverse group of inherited disorders caused by reduced or absent production of one or more of the globin chains. They can be characterized according to the globin chain or chains affected: α-, β-, γ-, δβ-, δ-, and εγδβ-thalassemias (Weatherall, 2004a, Weatherall, 2004b; Thein, 2005b).

1.3.2.1 α-Thalassemia. The frequency of α-thalassemia, varies from one per cent (i.e., in southern Spain) to 90 per cent (i.e.the tribal populations of India) through tropical and subtropical regions (Chong *et al.*, 2000; Higgs, 2004). α-thalassemia is a result of reduced or abolished synthesis of α-globin chains (Bowden *et al.*, 1992; Liu *et al.*, 2000). The majority of mutations causing α-thalassemia are large deletions which may cause loss of one (–α) or both (– –) of the α-globin genes (Dode *et al.*, 1992). Large deletions causing

α-thalassemia constitute 95 per cent of diagnosed cases. The most common deletions are $-\alpha^{3.7}$, $-\alpha^{4.2}$, $--^{SEA}$, $--^{MED}$, $-(\alpha)^{20.5}$, $--^{FIL}$ and $--^{THAI}$. Rarely, mutations affecting single or few nucleotides within the structural genes may result in α-thalassemia, and they are named as non-deletional form of α-thalassemia (Bowden *et al.*, 1992; Chong *et al.*, 2000).

Mutations affecting expression of one or more of the α-globin genes are observed in four clinical forms within affected individuals. The first is silent α-thalassemia trait (–α/αα); individuals who have a defective α-globin gene are diagnosed as silent α-thalassemia carriers. These individuals have no clinical abnormalities and their hematological values are within normal ranges. In α-Thalassemia trait two defective α-globin genes are observed (–α/–α or – –/αα). Hypochromic microcytic anemia and elevated RBC indices are significant clinical symptoms of this group. HbH is the third clinical form and it results from deletion of three α-globin genes (– –/–α). It is the most severe form of α-thalassemia compatible with life. Individuals who have HbH suffer from chronic hemolytic anemia with variable degree of severity (Clegg *et al.*, 1999). In addition, jaundice and hepatosplenomegaly are the significant clinical features of HbH group. It represents chronic but non transfusion dependent moderate anemia with less clinical complications when compared to β-thalassemia major. Hb Bart's Hydrops Fetalis syndrome results from deletion of four α-globin genes (– –/– –). Fetuses with this syndrome are unable to produce Hb and suffer from severe intra-uterine anemia. In this clinical syndrome, prevented fetal hemoglobin production leads to either death *in utero* or shortly after birth (Liu, 1997; Clegg *et al.*, 1999).

1.3.2.2. Hereditary Persistence of Fetal Hemoglobin and δβ-Thalassemias. Hereditary Persistence of Fetal Hemoglobin (HPFH) and δβ-thalassemias are heterogeneous disorders characterized by increased levels of HbF in adult life. Their discrimination could be possible by the differences in clinical symptoms and hematological values observed in affected individuals (Wood, 1993).

In adult life fetal Hb (HbF) accounts for lower than one per cent of the total Hb. HPFH is the case in which HbF ($\alpha_2\gamma_2$) production persists in adult life in higher levels (five to 25 per cent in heterozygotes) with normal red blood cell indices. HPFH is a result of large deletions affecting the 3' end of the β-globin gene cluster (deletions removing δ- and

β-globin genes) or point mutations affecting the γ-globin gene ($^{A}\gamma$ and $^{G}\gamma$) promoters. Coinheritance of HPFH with β^{+}-thalassemia is usually clinically asymptomatic with normal red cell indices and increased HbF levels. The homozygous form of HPFH is clinically asymptomatic, too, but results in mycrocytic hypochromic erythrocytes and imbalanced globin chains synthesis, like in β-thalassemia carriers.

δβ-thalassemia displays lower amounts of HbF (five to 15 per cent), hypochromic microcytic red cells with detectable chain imbalance. Coinheritance with β-thalassemia results in moderately severe clinical picture. Homozygotes display a phenotype similar to β-thalassemia intermedia or β-thalassemia major (Wood, 1993).

1.4. Beta-Thalassemia

Beta-thalassemia is prevalent throughout the Mediterranean region, Africa, the Middle East, the Indian subcontinent and Burma, southeast Asia including southern China, the Malay Peninsula and Indonesia (Figure 1.5). The carrier frequency of β-thalassemia varies from one per cent (i.e. in northern Italy) to 50-70 per cent (i.e. βE in some regions of South-East Asia); however, gene frequencies may display considerable variations even within a single country. It is estimated that there are 80 million carriers of β-thalassemia worldwide. To date, more than 200 mutations affecting *cis*-acting elements of the β-globin gene and causing β-thalassemia are identified: more than 180 of them are point mutations and approximately 13 are deletions ranging from 290 bp to more than 60 kb (Thein, 2005b; Higgs, 2004, Globin Gene Server).

1.4.1. The β-Globin Gene

The β-globin gene is located at the 3'-end of the β-globin gene cluster, and it is the gene dominantly expressed in adult life. Like other genes in the cluster, it is composed of three exons and two introns, covering approximately 1.8 kb. The first exon of the β-globin gene encodes amino acids 1-30; exon2, 31-104 and exon3, 105-146. The residues involved in heme binding and αβ-dimer formation are encoded by exon2, non-heme binding regions of β-globin chains are encoded by the third exon (Brittain, 2002). There are many

conserved sequences important for transcriptional regulation of β-globin gene expression found in the promoter region, 5'-UTR, and 3'-UTR.

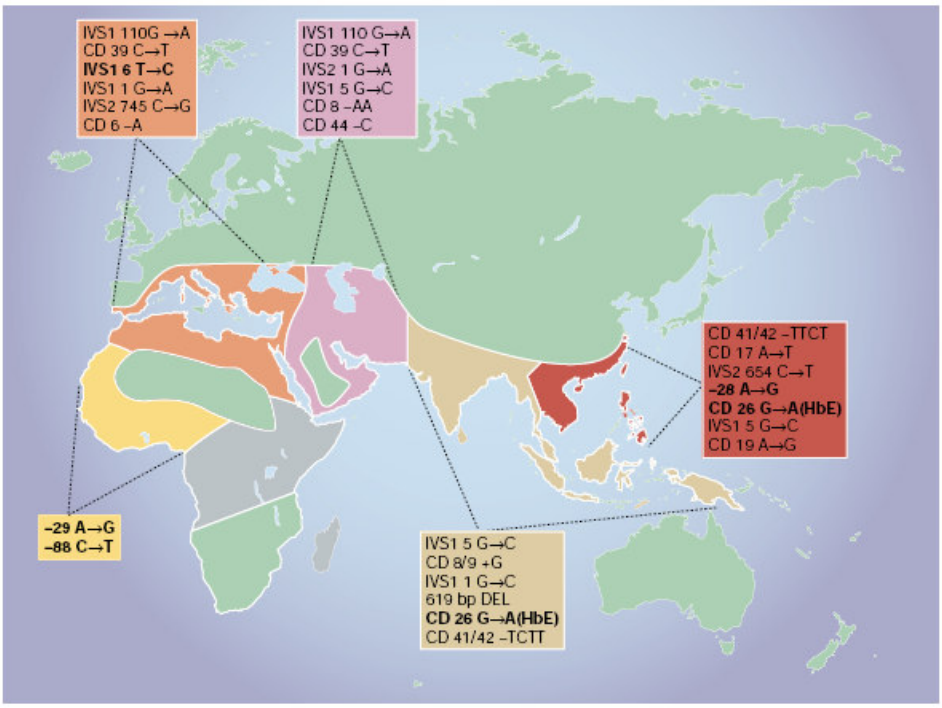


Figure 1.5. Worldwide distribution of β-thalassemia alleles (Weatherall, 2001)

Conserved sequences of the promoter region are the ubiquitous TATA and CCAAT boxes located at -28 to -31 and -72 to -76, respectively. TATA box is the site in which the transcription complex is formed, and it is the key *cis*-acting element for gene function. In addition, there are two erythroid-specific CACCC motifs influencing efficiency of transcription which are positioned at -86 to -90 and -101 to -105. The CCAAT box and CACCC motifs are the elements improving efficiency of transcription. The 5'-UTR occupies 50 nucleotides between the transcription start site (Cap Site) and the initiation codon (ATG). There are two important conserved sequences in the 5'-UTR: the CTTCTG found at +8 to +13 and the CACCATG motif lying 3' end of 5'-UTR. The 3'-UTR consists of 132 nucleotides covering the region between the transcription termination codon (TAA) and the polyA tail. The conserved sequence in the 3'-UTR is AATAAA(N_{20})GC, located 20 nucleotides upstream to the polyA tail. This motif provides both cleavage of pre-mRNA at its 3'end, and addition of the polyA tail which is important in mature mRNA structure. The importance of these conserved sequences in β-globin gene expression is demonstrated

by identification of mutations resulting in β-thalassemia (Ho and Thein, 2000; Stamatoyannopoulos, 2005).

1.4.2 Mutations Causing β-Thalassemia

Beta-thalassemia is mainly due to inherited molecular defects affecting the β-globin gene (Figure 1.6). Functionally defective β-globin alleles display two main phenotypes according to the nature of the mutation: β^+-thalassemia, reduced synthesis of β-globin chains with variable degrees, and β^0-thalassemia, without any β-globin chain production.

Mutations affecting transcription are single base substitutions hitting consensus sequences in the promoter region or 5'-UTR and result in downregulation of β-globin transcription. The phenotype is generally mild β^+-thalassemia. For example, the C to T substitution at -101 and the A to T substitution at the cap site cause an extremely mild deficit on β-globin products and reflect a very mild β^+-thalassemia phenotype. Affected individuals have normal HbA_2 and red blood cell indices, and in heterozygotes they display a silent phenotype. For this reason, generally it is not possible to detect these mutations independently. Most of them are identified in combination with severe mutations (Olivieri, 1999; Thein, 2004).

Mutations affecting RNA processing involves substitutions at either of the invariant dinucleotides, consensus sequences at splice junctions, polyadenylation signal and mutations in intervening sequences, creating cryptic splice sites. Substitutions at either of the invariant dinucleotides (GT at 5' and AG at 3') completely abolish splicing and results in β^0-thalassemia phenotype. On the other hand, molecular defects disturbing consensus sequences at the splice junctions reduce the efficiency of normal splicing and produce β^+-thalassemia with different degrees of severity. For example, the G to C substitution at IVS-I-5 has a more severe phenotype than the IVS-I-6 T to C transition. The main causes of cryptic splice site formation are the mutations in intervening sequences which represent a donor or acceptor splice site. These defects reduce normal splicing and result in β^+-thalassemia phenotype. The IVS-I-110 G to A substitution is the most common example of mutations creating cryptic splice sites. Another group of mutations leading to a failure in RNA processing are polyadenylation signal (AATAAA) mutations. These mutations cause

a mild β^+-thalassemia phenotype because of the ineffective cleavage of the mRNA (Olivieri, 1999; Thein, 2004).

Translation failure may be a result of mutations affecting initiation, elongation or termination stages of globin chain production. Approximately half of β-thalassemia alleles are due to premature termination codons. Frameshifts or nonsense mutations affecting exon3 lead to premature termination, causing dominantly inherited β-thalassemia in which truncated or elongated and highly unstable β-globin gene products are produced. The unstable β-chain variants produced are not detectable and only predicted from the DNA sequence, thus could not be degraded by the proteolytic pathway. The hyperunstable β-chain variants fail to form functional tetramers and precipitate intracellularly leading increased ineffective erythropoiesis by destructing red blood cell precursors (Thein, 1999; Thein, Thein, 2005b). On the other hand, exon1 and exon2 mutations display Mendelian recessive inheritance probably due to the proteolytic pathway activity preventing translation of truncated mRNAs (Weatherall, 2001). FSC8 (-AA), initiation codon mutations (ATG to AGG, ACG and GTG substitutions), and Cd37 (TGG to TGA), altering tryptophan to stop codon are few examples of mutations affecting translation. The molecular defects affecting translation are recognized with β^0-thalassemia phenotype (Olivieri, 1999; Weatherall, 2001).

Unlike the α-globin gene cluster, gene deletions are rare in β-thalassemia. Amongst the deletions causing β-thalassemia, the most common one is the deletion removing 619 bp at the 3' end of the β-globin gene, which is restricted to specific ethnic groups of India and Pakistan. Deletions result in complete absence of β-chain production and are defined with β^0-thalassemia phenotype. They are expected to display an anemic picture, however heterozygotes have increased HbF and HbA_2 levels due to loss of competition for β-LCR. In homozygous form, these increases in HbF and HbA_2 are enough to compensate for the absence of HbA (Thein, 2004).

There are identified β-thalassemia cases independent from the β-globin gene cluster. In 1998, a mutation affecting the key erythroid transcription factor GATA-1 on the X chromosome has been shown to cause trombocytopenia and β-thalassemia with reduced β-chain production. Few more families were identified with β-thalassemia by similar

mutations affecting GATA-1. The second example of this group was observed in 19 patients with a rare autosomal recessive disease, called trichothiodystrophy. These individuals inherited mutations in XPD helicase protein, which is a multicomponent of general transcription factor TFIIH. These molecularly diagnosed patients have the hematological features of β-thalassemia trait (Higgs, 2004; Thein, 2005a).

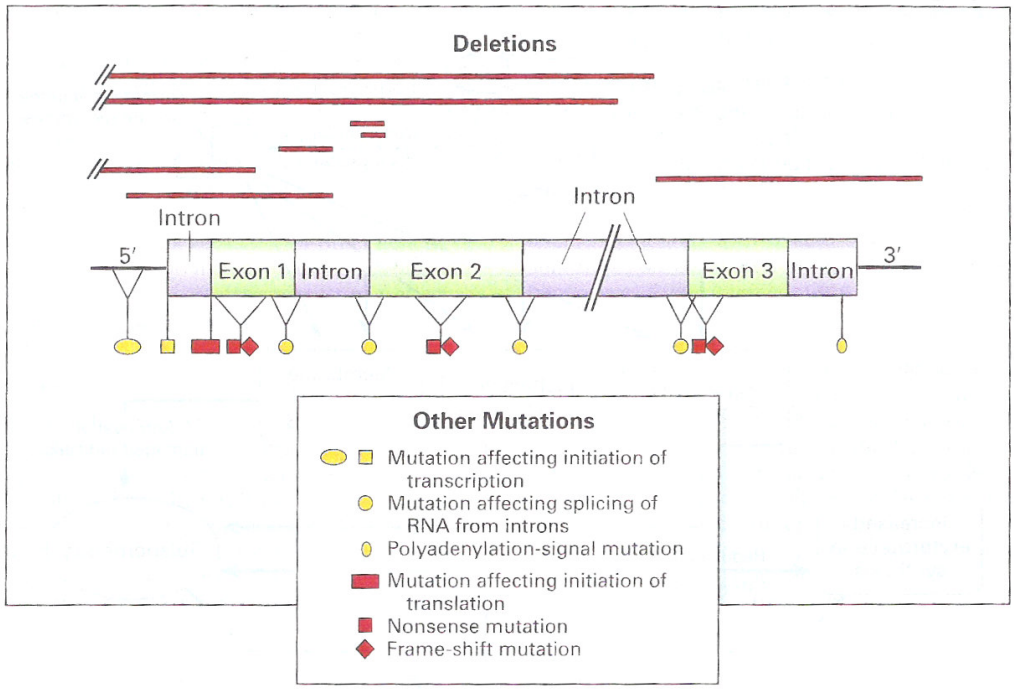


Figure 1.6. The β-globin gene and location of mutations causing β-thalassemia (Olivieri, 1999)

1.4.3 Clinical Syndromes of β-Thalassemia

Inheritance of affected alleles, together in heterozygous, homozygous or compound heterozygous states, defines four major clinical phenotypes in β-thalassemia ranging from symptomless conditions to blood transfusion dependent cases.

Silent β-thalassemia carriers do not have any significant hematological abnormality and carry a mild β^{+}-thalassemia mutation due to slightly reduced β-globin chain production.

Beta-thalassemia trait is identical with mild anemia. Hypochromic (MCH between 20pg and 25pg), microcytic (MCV between 60 fl and 75 fl) red blood cells, elevated levels of HbA_2 (between 3.5 and 7 per cent) and variable increases in HbF production (up to two per cent) are the significant determinants of this group (Olivieri, 1999; Thein, 2005a).

The third clinical syndrome is β-thalassemia intermedia, which is defined with moderate anemia, splenomegaly and elevated red blood cell indices. The phenotype of this group is between β-thalassemia trait and β-thalassemia major. β-thalassemia intermedia can be observed either in coinheritance of two mild β^+-thalassemia mutations or severe β^0-thalassemia mutations, with coinheritance of α-thalassemia and enhanced HbF production. Also, coinheritance of α-thalassemia triplication with heterozygous β^0-allele will form intermedia phenotype. There are 30 mutations causing the β-thalassemia intermedia phenotype in heterozygous form despite normal α-globin genotype. These are the mutations hitting exon3 and resulting in premature termination of translation. Dependent on the clinical severity, these individuals are rarely transfused or do not require any blood transfusions (Weatherall, 2001; Thein, 2005a).

Beta-thalassemia major results when both β-globin genes are defective, either in homozygous or compound heterozygous form. This group displays a severely anemic phenotype as a result of the large imbalance between the α- and β-globin chains. Excess α-chains precipitate in red blood cell precursors leading to abnormal cell maturation and apoptosis which results in ineffective erythropoiesis. The affected individuals are healthy up to three to six months after birth, then the anemic picture develops gradually. For healthy growth and development of affected individuals, regular blood transfusions are essential (Thein, 2005a; Thein 2005b).

Within these four broad clinical groups, it is not possible exactly to define the phenotype of individuals because of the various factors modifying pathophysiology of β-thalassemia syndromes.

1.4.4. Pathophysiology of β-Thalassemia

The factors modifying pathophysiology of β-thalassemia can be listed as; the nature of the mutation, level of excess α-globin chains and inherited factors increasing γ-globin chain expression. The main modifying factor is the degree of imbalance between α/β chain productions. Dependent on the level of reduction in β-chain formation, the amount of excess α-chains differs and causes amount-dependent varying effects in red blood cell precursors. Excess α-chains which cannot bind to produce Hb, precipitate in red blood cell precursors and cause membrane damage, red cell destruction, ineffective erythropoiesis and consequently anemia (Traeger-Synodinos *et al.*, 1996). In heterozygous β-thalassemia, minimal effect of excess α-chains are observed and usually no or mild anemia is evident, but the severity of anemia increases in homozygous state because of the increased α-chain excess. Coinheritance of α-thalassemia with homozygous β-thalassemia results in milder clinical manifestations than expected, due to reduced amount of α-chain production. In contrast, additional α-globin genes may increase the severity of β-thalassemia syndromes even in heterozygotes (Traeger-Synodinos *et al.*, 1996; Thein, 2005a; Thein, 2005b).

In addition to the effects of excess α-chains, factors increasing HbF production are important in determining the severity of anemia. The C to T transition at position -158 of the $^{G}\gamma$-globin gene (*XmnI*-$^{G}\gamma$ polymorphism) is an important factor favoring HbF production under hematopoietic stress conditions. This could explain the clinical diversity even in patients with identical genotypes. There are other regions unlinked to β-globin gene cluster, but enhancing HbF production. These regions are located on 6q23, Xp22 and 8q. These *cis*-acting elements are suggested as important regulatory elements responding to erythropoietic stress and pharmacological induction leading to a milder β-thalassemia phenotype due to decreased α-globin chains (Garner *et al.*, 2002; Garner *et al.*, 2004; Close *et al.*, 2004; Thein, 2005a).

1.4.5. Treatment of β-Thalassemia

Beta-thalassemia major is considered as the most severe form of hemoglobinopathies and requires life-long blood transfusions to maintain required Hb levels. Regular blood transfusions are effective at retarding the progression of β-thalassemia, but patients are

under increased risk of blood infections, high-rate of allu-immunization, and transfusional iron overload. Iron chelation therapy is essential to combat against iron overload in order to prevent clinical consequences of iron overload such as cardiac failure, liver diseases and abnormal endocrine function. Desferoxamine (desferal) is the commonly used iron chelator, and it is applied subcutaneously via an infusion pump (Olivieri, 1999; Kanavakis *et al.*, 2004).

Bone marrow transplantation is a popular approach, used to treat more than 1000 individuals affected from β-thalassemia major. It can be performed, if there is an HLA identical donor for the affected individual. HLA identity is the major factor limiting bone marrow transplantation, because the probability of HLA identity with parents is 25 per cent, while it rises to 44 per cent in two siblings. Three siblings might have 58 per cent identical HLA patterns. It has been observed that early applications in life before blood-transfusions, would show better results (Lucarelli *et al.*, 2002).

Recently, better understanding of the mechanism about Hb switching opened a new era for the treatment of β-thalassemia. Today, it is apparent that reactivation of γ-globin genes can be obtained via contribution of specific transcription factors. If the molecular mechanism leading to the switch from γ- to β-globin can be mimicked, γ-globin gene can be reactivated. As a result, reactivated γ-globin gene expression will functionally replace the defective or absent HbA with HbF (Cao and Moi, 2002).

Treatment by γ-globin gene reactivation is also possible via pharmacological induction. In 1982, the effect of 5'-azacytidine, a cytotoxic drug, on increased HbF levels was demonstrated in monkeys. Human trials with 5'-azacytidine are discontinued because of the concerns on potential carcinogenic effects of the drug. Hydroxyurea (HU) treatment for β-thalassemia (a cytotoxic drug as well), is preferred rather than 5'-azacytidine due to its better safety profile (Cao and Moi, 2002). Studies performed on β-thalassemia intermedia patients demonstrated that HU would be a promising agent for treatment of β-thalassemia due to its reactivation effect on γ-globin chain production with unknown molecular mechanisms. It is a promising drug for β-thalassemia intermedia, but in order to evaluate its long term safety and efficiency, trials must be performed on a large number of groups and with longer duration (Dixit *et al.*, 2005). In addition, HU is used in the

treatment of Sickle Cell Anemia since its significant effect on reducing the frequency of acute painful episodes is demonstrated (Dalton *et al.*, 2005). Current studies on HU suggested that it can play a role in increased life span of Sickle Cell Anemia patients (Steinberg, 2003).

Hemoglobinopathies were the first disorders considered for gene therapy. Among the suggested treatments for hemoglobinopathies, gene therapy holds a great promise according to the results of animal experiments (Stathopulos, 2003). Huge amounts of information about pathophysiology of anemia and extensive knowledge on globin gene expression are the reasons enhancing their usage in gene therapy. Adenoassociated-viruses, retroviral and lentiviral vectors are used for the treatment of hemoglobinopathies. Lentiviral vectors are preferred for transmission of intact β-globin gene cluster to the host genome, because of the stable transmission they provide. In 2004, it was demonstrated that lentiviral vectors provide normal amounts of β-globin expression *in vitro*. Restoration of erythropoiesis, and reversal of the abnormally elevated apoptosis are demonstrated in several studies using lentiviral vectors (Puthenveetil *et al.*, 2004).

1.4.6. Beta-Thalassemia in Turkey

There are approximately 35 mutations reported in the populations of the Mediterranean basin. In most of these populations, only six or less mutations are prevalent and they constitute up to 90 per cent of the β-thalassemia alleles. Unlike other Mediterranean countries, Turkey is very heterogeneous at molecular level. Previous studies demonstrated that, the overall frequency of the six most common mutations constitute 69.7 per cent, and the first 12 mutations having a frequency above one per cent is 82.4 per cent (Tadmouri *et al.*, 1998). The main cause of this heterogeneity is the presence of diverse ethnic groups in Turkey due to its geographical location and rich historical past. Moreover, the frequency and diversity of mutations also demonstrate variations among the six geographical regions of Turkey. For example, the IVS-I-110 G to A substitution is the most common mutation causing β-thalassemia in Turkey; however, its frequency ranges from 26-50 per cent within the six geographical regions of Turkey (Tadmouri *et al.*, 1998).

In Turkey, the overall gene frequency for β-thalassemia is around two per cent. However, there are regions which have an increased gene frequency rising up to ten per cent (Trace, Muğla and Antalya). A further increase in gene frequency is caused by the high birth rate and consanguineous marriages observed especially in the Eastern parts of Turkey. Each year, approximately 800 pregnancies are expected to require prenatal diagnosis for β-thalassemia and Sickle Cell Anemia in Turkey.

Although there are improvements in the treatment of β-thalassemia, there is not any definitive cure yet. Thus, preventive programs must be introduced in order to decrease risk of affected births. Prevention programs rely on population screening (especially adolescent school children) and application of prenatal diagnosis to couples at risk. Prenatal diagnosis programs performed in countries such as Greece, Italy and Cyprus were very successful in lowering the incidence of affected births. For example, in Sardinia, the frequency of affected births decreased to less than five per cent by efficient application of prenatal diagnosis.

In order to apply effective prenatal diagnosis, it is important to define the molecular basis of a specific population. Although there is a great molecular heterogeneity in Turkey, prenatal diagnosis was shown to be feasible due to current advances in PCR-based technologies and improved methods of early fetal sampling (Tadmouri and Başak, 2001).

1.5. Bioinformatics

Bioinformatics is the application of computer technology to biological sciences, in order to manage and interpret raw data obtained from researches. In other words, bioinformatics is the integration of computer and biological sciences to analyze and retrieve data about biological and genetic information. The need for the bioinformatics field has greatly increased by the publicly available genomic information, resulting from the Human Genome Project.

In addition to providing and managing data more efficiently, the central challenge of bioinformatics is to design clear and precise analysis tools in order to decode the language of nucleotides forming proteins. The information obtained by computational applications

may help to identify structural, functional and evolutionary information encoded in nucleotide sequences. Extracting the biological information from sequence data is a hard challenge which requires decoding an unknown language composed of sentences (proteins), words (motifs) and letters (amino acids). The ultimate goal of bioinformatics is to understand the meaning of sequences. It is essential to improve two analytical themes to achieve this goal. First is the pattern recognition techniques, to detect similarities between sequences and reveal related structures and functions. Secondly, deduction of three dimensional structures of proteins is important to infer function directly from the linear sequence (Attwood, 1999; Luscombe *et al.*, 2001).

1.5.1. History of Bioinformatics

This new field was coined in mid-1980s when the projects dedicated to identify the complete genome sequence were initiated in USA. This projects required high technology computational methods for analyzing genetic maps and sequence data (Attwood, 1999). As bioinformatics developed in parallel with the improvements in molecular biology, the history of bioinformatics can be related to the improvements in molecular biology (Table 1.1) (Richon, 2001).

1.5.2. Computers in Bioinformatics

Computers serve four functions in bioinformatics: communication, computation, control and storage. Until the mid-1990s, newly discovered human sequences were published in printed journals. If you were interested in these sequences, you had to find the key sequence you are looking for, by hand or optical character recognition systems in order to automatically capture the sequence and translate them into a machine-readable form. Today, one of the advantages that computers brought, is reaching served databases for nucleotide sequences via the World Wide Web. Other application areas of computers in bioinformatics are shown in Table1.2 (Bergeron, 2003).

Table 1.1. History of bioinformatics (Richon, 2001)

Year	Event
1953	Proposal of the double helix model for DNA by Watson and Crick
1969	Creation of the ARPANET linking computers in Standford, UCSB, The University of Utah, and UCLA
1970	Publication of the details of the Needleman-Wunsch algorithm for sequence comparison
1973	Announcement of the Brookhaven Protein Data Bank
1974	Development of the concept of connecting networks of computers into a "internet" and the Transmission Control Protocol (TCP) by Vint Cerf and Robert Kahn
1977	Report of methods for sequencing DNA by Alan Maxam and Walter Gilbert (Harvard) and Freidrich Sanger (U.K. Medical Research Council)
1980	The first complete genome sequencing for an organism (ΦX174)
1981	Publication of the Smith-Waterman algorithm for sequence alignment
1985	Description of the PCR reaction by Kary Mullis and co-workers
1986	The SWISS-PROT database is created by EMBL
1987	Publication of the physical map of *E.coli*
1988	Establishment of NCBI at the National Cancer Institute Start of The Human Genome Initiative Publication of the FASTA algorithm for sequence comparison by Pearson and Lupman
1990	Implementation of the BLAST program
1991	Creation of WWW by CERN
1995	Sequencing of the *Haemophilus influenza* genome (1.8 Mb) Sequencing of the *Mycoplasma genitalium* genome
1996	Sequencing the genome for *Saccharomyces cerevisiae* (baker's yeast, 12.1 Mb) Production of the first commercial DNA chips by Affymetrix
1997	Publication of *E. coli* genome (4.7 Mbp)
1998	Publication of Ceanorhabditis elegans and baker's yeast genomes Establishment of The Swiss Institute of Bioinformatics
2000	Publication of *Pseudomonas aeruginosa* genome (6.3 Mbp) Sequencing of the *A. thaliana* genome (100 Mb) Sequencing of the *D. melanogaster* genome (180 Mb)
2001	Publication of the first draft of the human genome (3.000 Mbp)
2003	Completion of the human genome Completion of the mouse genome

Table 1.2. Application areas of computers in bioinformatics (Bergeron, 2003)

Application Areas	Associated Technologies
Process control	Equipment control Robotics Automatic data collection
Archiving	Databases IT infrastructure Vocabulary
Numerical processing	Pattern matching Simulation Data mining Search engines Statistical analysis Visualization
Communications	Desktop publishing Web publishing Internet

1.6. Database Management Systems

Database management system (DBMS) is the set of software tools designed in order to create a practical database application. The DBMS facilitates usage of databases by simplifying the applications for users. It prevents the user to deal with details of algorithms underlying the structure and analysis functions of the database. In other words, DBMS systems provide an abstract of the underlying mechanism which is much more precise than algorithms. Some examples of commercially used DBMS packages are as follows: Microsoft, Orackle, Sybase, SQL and File Maker.

A physical database can be defined as the lowest level of abstraction. This level is useful for someone who updates the data and files in the database every day. A database with a higher abstraction is named as conceptual database. It is defined according to the structure of data involved and the properties of data stored and manipulated. Conceptual databases are commonly represented by entity-relationship models or data models.

The Entity-relationship model focuses on different data entities and the interrelationship between them. Some data entities in bioinformatics can be listed as disease processes, protein sequences, and nucleotide sequences. Within this model, relationships are determined according to how data are associated with each other: one-to-one relationship, one-to-many relationship or many-to-many relationships. For example, one-to-one relationship can be defined by a gene encoding only one mRNA sequence, while synthesis of several proteins from this gene is a one-to-many relationship.

Data models provide an alternative method to represent and manipulate large amounts of data. Flat, hierarchical, relational, and object oriented data models are the most common models supported in DBMS products (Figure 1.7). Flat data model is composed of a simple table without any parts providing relationships between records. For this reason, it is not possible to work more than one file or table at a time. This feature of flat data model makes it popular especially for capturing sequence based data.

The model developed in 1970s is the relational data model. It is composed of a table with unique rows. The columns or fields of the table are called characteristics, predicates, or classes. In order to connect related data, a process termed join is performed. In this process, fields containing similar data are combined. Theoretically, there are not any restrictions for the number of relations, but increased number of relations may lead to a decrease in performance. This case limits the number of logical relations in a database.

The Hierarchical model is an alternative model for the relational databases which have relation limitations. Different than the relational model, records are not divided into fields. For this reason, the relations do not depend on data involved in the database. The relationships are defined when the database is created. This is a disadvantageous property, when compared to the more flexible relational data model. Data is organized in form of an inverted tree with the parent file on top and children below. The relationship defined between parents and children is one-to-many relationship.

Another alternative model for relational databases is the object-oriented model combining natural structure of the hierarchical model with the flexibility of relational model. This model represents complex genomic information, because it allows

combinations of data to be treated as single entities. But it is still behind the relational data model, because of the flexibility and power of relational data designs (Bergeron 2003).

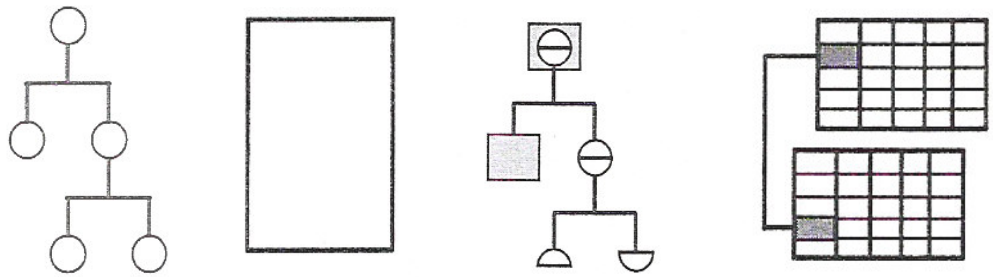


Figure 1.7. Data models (Bergeron 2003)

1.6.1. Database Systems

In fact databases are files composed of records which contain specific data fields. Databases provide long-term memory of computer operations. For most of the researchers, database technology refers to a system which helps to handle large amounts of data and information obtained, manipulated and communicated every day. They are named depending on their structure, content, use and capacity (Table1.3).

There are three levels of databases: The Data repository is the simplest level of databases which is used as an information storage facility. It provides minimal analysis and querying functionality. A data repository stores systematically collected data from a single application such as sequencing machine, microarray analyzer or clinical systems. It is advantageous to use data repositories because of storage limitations, or as the local database is always in use, it might be almost impossible to compare data from multiple runs of a sequencing machine (Figure 1.8).

Next complex level of databases is the data mart; it can be defined as a searchable database system, organized according to users' needs. Data mart involves required parts of information extracted or mirrored in real time from several applications and prevents mass copying from another databases. Like data repository, a data mart involves a narrow level

of information about a particular research project, but it differs from the data repository by serving information from multiple applications.

The Data warehouse is the most complex level of databases. It is a central database providing information to users from various resources like sequencing machines or national genomic databases (Bergeron, 2003).

Table1.3. Public bioinformatics databases accessible via the World Wide Web

Database Type	Example	Note
Nucleotide sequence	GenBank	One of the largest sequence databases
	DDBJ	DNA DataBank of Japan
	EMBL	European Molecular Biology Laboratory
	MGDB	Mouse Genome Database
	GSX	Mouse Gene Expression Database
	NDB	Nucleic Acid Database
Protein Sequence	SWISS-PROT	Swiss Institute for Bioinformatics and European Bioinformatics Institute
	TrEMBL	Annotated supplement to SWISS-PROT
	TrEMBLnew	Weekly, pre-processed update to TrEMBL
	PIR	Protein Information Resources
3D-Structures	PDB	Protein DataBank
	MMDB	Molecular Modeling Database
	Cambridge Structural Database	For small molecules
Enzymes and compounds	LIGAND	Chemical compound and reactions
Pathways and Complexes	Pathway	Metabolic and regulatory pathway maps
Molecular Diseases	OMIM	Online Mendelian Inheritance in Man

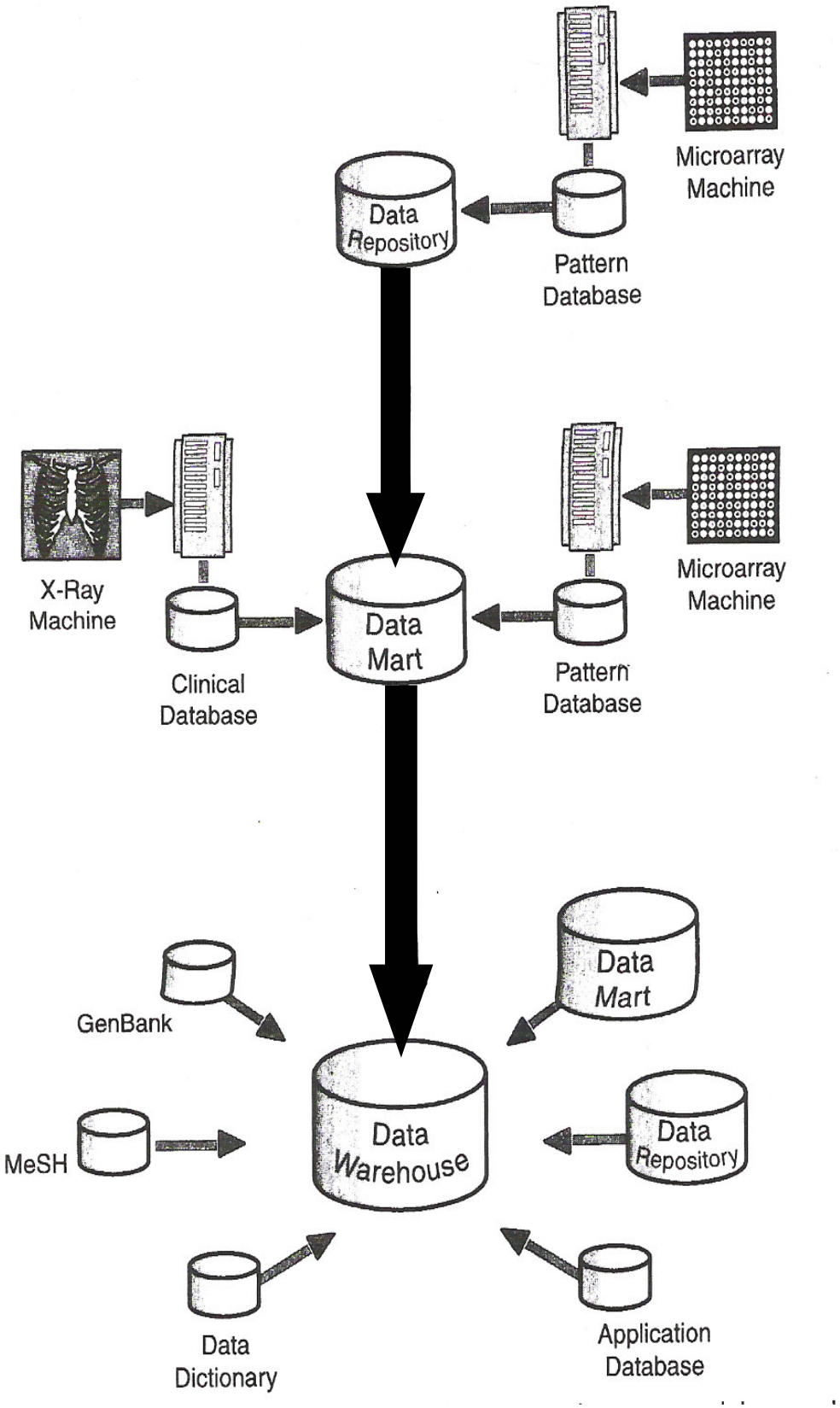


Figure 1.8. Database Nomenclature (Bergeron, 2003)

1.7. The Globin Gene Server

The Globin Gene Server (http://globin.cse.psu.edu) provides data and tools for studying the function of DNA sequences involved in Hb production. It contains information about human hemoglobin mutations and their effects, experimental data about regulation of the β-globin gene cluster, and software tools for comparing sequences to discover regions, probably playing roles in globin gene expression. The Globin Gene Server provides associated resources like online Syllabi (about Human Hemoglobin Variants and Thalassemia Mutations), HbVar, the GALA and, dbERGE databases.

1.7.1. The Syllabus of Human Hemoglobin Variants

A Syllabus of Human Hemoglobin Variants was published in 1996 (Huisman *et al.*, 1996). This syllabus involves a comprehensive list of all known human Hb variants, including α-, β-, γ-, and δ-globin chains. Within the globin gene server, it is possible to browse the contents of the syllabus. Additionally, an on-line version of the Syllabus allows computerized searches for a word or phrase either in the whole content or in the parts specified. The contents of this Syllabus are stored in a database called HbVar, which also includes some extra fields, that restate the mutation specifications in a standardized format.

1.7.2. The Syllabus of Thalassemia Mutations

A Syllabus of Thalassemia Mutations was published in 1997 by Huisman *et al.*, and contains more than 300 thalassemia alleles and HPFH determinants. The process for adapting the printed book available on-line is still ongoing: discussion for related mutations, description for all mutation entries. There are some figures and electronic search facilities which are not completed yet.

1.7.3. The Hb Variants Database

The Hb Variant Database (HbVar) is a relational database of information about hemoglobin variants and mutations that cause thalassemia. The initial data came from Syllabi authored by Huisman *et al.* (Patrinos *et al.*, 2004). The initial step in construction

of HbVar Database was established in 1998, when the electronic version of the Syllabus of Hemoglobin Variants (693 variants) was published at the Globin Gene Server via the World Wide Web (Chui *et al.*, 1998). This initial step was continued by the addition of 200 more entries. Sequence alterations, techniques used, biochemical and hematological effects, associated pathology, ethnic occurrences and references were also involved for each variant and mutations present in the database (Chui *et al.*, 1998; Patrinos *et al.*, 2004).

1.7.4. Genomic Sequences Alignment and Annotations Database

Genomic Sequences Alignment and Annotations Database (GALA) is a relational database containing whole genome sequence alignments between human and mouse with extensive annotations of the human sequence. Alignments of genomic DNA sequences between species at an appropriate phylogenetic distance are useful for finding candidate functional regions. Also, the determination and annotation of complete genomic DNA sequences improves our understanding of evolution, genetics, and physiology. The browsers providing access to the sequence and annotations of the human genome are the human genome browser (HGB) (http://genome.ucsc.edu/), Map Viewer at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/), and Ensembl at the Sanger Centre (http://www.sanger.ac.uk/) and EBI (http://www.ebi.ac.uk/). GALA provides a wide range of querying function by current browsers, alignment resources, and databases. Output from GALA queries can be viewed as a table of data with hyperlinks, or as text and they can be examined in UCSC Genome Browser views in order to facilitate some analysis (Giardine *et al.*, 2003; Elnitski *et al.*, 2005).

1.7.5. Experimental Results on Gene Expression

Experimental Results on Gene Expression (dbERGE) is a database which is a part of the Globin Gene Server and involves data about mammalian globin gene expression. The main focus of this database is β-LCR and provides querying function about globin gene regulation. The data obtained from previously published 49 papers mostly focused on HS2 and HS3; were arranged and displayed in forms of tables. The information provided is about protein binding assays, DNA transfer and hypersensitive site mapping experiments. Tables of DNA transfer experiments involve very detailed information about the construct

tested; mutations on it, experiments performed on that construct and the results, published in numeric values. Protein Binding Assay tables contain results of electrophoretic shift assays, methylation interference assays, *in vivo* and *in vitro* footprints. In addition, it is possible to find detailed information about probes used, experimental parameters, relative mobility and a description of mutation effects on mammalian globin gene expression. Lastly, location, cell line, type of DNA transfer, induction, and band intensity parameters are found in the tables of hypersensitive site mapping experiments (Riemer *et al.*, 1998; Hardison *et al.*, 1998; Elnitski *et al.*, 2005).

1.8. Deniz: A Database Constructed by FileMaker Pro 6 Software

Deniz is a database designed in 2002 as a search and retrieval system. It is composed of five relational databases providing detailed statistical data content about the frequency and distribution of β-thalassemia alleles in world populations. This database is constructed by using FileMaker Pro 6 (Gülen, 2002). Since January 2006, Deniz Database is hosted on a server at Boğaziçi University, Molecular Biology and Genetics Department, Suna and İnan Kıraç Foundation, Neurodegeneration Research Laboratory via World Wide Web.

1.8.1. FileMaker Pro 6 Software

FileMaker databases are files organized in forms of records. Each record contains fields storing, calculating and displaying specific types of data. In other words, records forming a FileMaker file are composed of fields carrying related information about a subject. The information entered to fields are named as value and field values may be text, number, date, time, graphic, sound, movie, a result of formula calculated by the field or summary of field values for current record.

Data layouts are used to determine the organization of information within the database in order to view, print, report, find and enter data. Database files can have many different layout types dependent on the subcategories of information for displaying them in different ways. Separate layouts can be designed for same data category in order to display them according to users needs. For example, it is possible to design different layouts for entering and printing same data in different ways. It is important to note that, changing a

layouts design does not affect the value stored in fields. However, changing a value in a field is reflected in all layouts designed in the database. In other words, a layout does not store but display your information. Therefore, in a layout it is possible to arrange, format and rename fields. The database design should be at the possible simplest level and should be modified according to users needs. The type and design of the database identifies its sharing protocol: Local Area Network (LAN) or World Wide Web.

One of the most important features of FileMaker Pro is its web publishing flexibility. It does not require any other software tools for World Wide Web publishing. FileMaker Pro web companion is a plug-in component. It is providing the tools required for publishing a database on the web and functions as a web server application which communicates with web browsers requesting data from a FileMaker Pro database. By this way, when a database layout is constructed, its web page design is constructed as well. In addition to the advantages provided by FileMaker Pro web companion plug-in while publishing a database on World Wide Web, it has some limitations, too (i.e. limited guest number, etc.). The difficulties of instant web publishing can be overwhelmed by custom web publishing which uses Claris Dynamic Markup Language (CDML) tags (types of HTML codes). These tags are providing communication between the fields on HTML page and FileMaker Pro layout. Custom web publishing has many advantages over instant web publishing like more flexible web page design.

FileMaker Pro provides advantages even for anyone not involved in bioinformatics education without any need to information technology (IT) experts. In recent years, because of the advantages provided (such as short development time, ease of modification and easy access via LAN or World Wide Web) it is commonly used especially by biologists for making data management much easier while working in laboratory and evaluating results of experiments.

2. PURPOSE

Beta-thalassemia major, considered as the most severe form of hemoglobinopathies, requires life-long blood transfusions to maintain adequate Hb levels. Although there are improvements in the treatment of β-thalassemia, there is not any definitive cure yet. Thus, preventive programs must be introduced in order to decrease the probability of affected births. Prevention programs rely on population screening and application of prenatal diagnosis to couples at risk. In order to apply effective prenatal diagnosis, it is important to define the molecular basis of a specific population.

This thesis has two main objectives:

- to analyze globin gene mutations of specific patients and families, sent to our laboratory on a routine and regular basis.

- to construct a user-friendly database which allows the management of the huge amount of patient data entering our laboratory daily and evaluation of the results obtained.

3. MATERIALS

3.1. Peripheral Blood and Chorionic Villus Samples

Human peripheral blood samples used in this study were sent to our laboratory from various hospitals in Adana, Ankara, Antalya, Bursa, Gaziantep, Isparta, İstanbul and İzmir. The samples of β-thalassemia patients and carriers and patients with abnormal hemoglobins were subjected to molecular analysis.

3.2. Buffers and Solutions

The chemicals used to prepare the buffers and solutions used in this study, were purchased from Merck (Germany), Sigma (USA and Germany) and AppliChem (Germany).

3.2.1. DNA Extraction from Peripheral Blood Samples (NaCl Extraction)

Cell Lysis Buffer	:	155 mM NH_4Cl 10 mM $KHCO_3$, 1 mM Na_2EDTA (pH 7.4)
Nuclei Lysis Buffer	:	10 mM Tris-HCl (pH 8.0) 400 mM NaCl 2 mM Na_2EDTA (pH 7.8)
Proteinase K	:	20 mg/ml in dH_2O
Sodiumdodecylsulphate (SDS)	:	10 per cent SDS (w/v) in dH_2O (pH 7.2)
Sodium Chloride (NaCl)	:	5 M NaCl
Ethanol (EtOH)	:	Absolute EtOH (Riedel de Haen, Germany)
TE Buffer	:	20 mM Tris HCl (pH 8.0) 1 mM Na_2EDTA (pH 8.0)

3.2.2. DNA Extraction from CVS Samples

Cell Lysis Buffer	:	155 mM NH_4Cl 10 mM $KHCO_3$, 1 mM Na_2EDTA (pH 7.4)
Nuclei Lysis Buffer	:	10 mM Tris-HCl (pH 8.0) 400 mM NaCl 2 mM Na_2EDTA
NH_4Ac	:	9.5 M stock solution
Ethanol (EtOH)	:	Absolute EtOH (Riedel de Haen, Germany)
TE Buffer	:	20 mM Tris HCl (pH 8.0) 1 mM Na_2EDTA (pH 8.0)

3.2.3. Agarose Gel Electrophoresis

10X TBE Buffer	:	0,89 M Tris-Base 0.89 M Boric Acid 20 mM Na_2EDTA (pH 8.3)
Ethidium Bromide (EtBr)	:	10 mg/ml
1 or 2 per cent Agarose Gel	:	1 or 2.0 per cent agarose (w/v) in 0.5X TBE Buffer, containing 0.5 µl/ml EtBr
10X Loading Dye	:	2.5 mg/ml Bromophenol Blue (BPB) 1 per cent SDS in glycerol
DNA Ladder (100 bp)	:	MBI Fermentas, Lithuania

3.2.4. Polymerase Chain Reaction (PCR)

10X $MgCl_2$ Free Buffer	:	100 mM Tris-HCl 500 mM KCl Promega, USA
Magnesium Chloride ($MgCl_2$)	:	25 mM in dH_2O Promega, USA
2'-Deoxynucleoside 5'-Triphosphate (dNTP)	:	100 mM of each dNTP Promega, USA

3.3. Enzymes

Taq DNA Polymerase (in Storage Buffer B)	:	5U/µl, Promega, USA
aTaq DNA Polymerase		5U/µl, Promega, USA
Restriction Enzymes	:	New England Biolabs, USA
		MBI Fermentas, Lithuania

3.4. Oligonucleotide Primers

The primers used in this study were synthesized by Iontek Ltd. (Istanbul, Turkey) and MWG-Biotech GmbH (Germany).

3.4.1. Beta-Globin Gene Amplification and DNA Sequencing

Primer Name	Primer Sequence
CD7 (Forward)	5'-TCC TAA GCC AGT GCC AGA AG-3'
CD6 (Reverse)	5'-ATC ATT CGT CTG TTT CCC ATT CTA AAC-3'
CD1 (Forward)	5'-TGC CTC TTT GCA CCA TTC TAA-3'
CD2 (Reverse)	5'-CGA CCT CCC ACA TTC CCT TTT-3'

3.4.2. Analysis of Variable Number of Tandem Repeats (VNTR)

Primer Name	Primer Sequence
St14-A (Forward)	5'-GGC ATG TCA TCA CTT CTC TCA TGT T-3'
St14-B (Reverse)	5'-CAC CAC TGC CCT CAC GTC ACT T-3'
XY-1 (Forward)	5'-CTG ATG GTT GGC CTC AAG CCT GTG-3'
XY-2 (Reverse)	5'-TAA AGA GAT TCA TTA ACT TGA CTG-3'

3.4.3. Differential Amplification of $^{G}\gamma$-Globin Gene Promoter

Primer Name	Primer Sequence
G5 (Forward)	5'-GCT ACA GAC AAG AAG GTG-3'
G4 (Reverse)	5'-TTT TAT TCT TCA TCC CTA GC-3'

3.5. PCR Purification Kit

QIAquick PCR Purification Kit was used to purify PCR products from excess primers, nucleotides and salts. The increased quality of PCR products is important for obtaining good quality results in direct DNA sequencing. QIAquick PCR Purification Kit was purchased from Qiagen Ltd, USA.

3.6. Beta-Globin StripA^{ssay} Kit

Beta-Globin StripA^{ssay} Kit was purchased from ViennaLab, Austria.

3.7. FileMaker Pro 6 Software

The software was supplied by Fatih University in connection with the Deniz Database.

3.8. Equipment

This thesis was carried out at Boğaziçi University, Department of Molecular Biology and Genetics (Istanbul, Turkey). The equipments used in this study are listed in Table 3.1.

Table 3.1. Equipment used in the framework of this thesis

Autoclaves	:	Model MAC-601, Eyela, Japan
Balances	:	GM 512-OCE, Sartorius, Germany
Centrifuges	:	Centrifuge 5415C, Eppendorf, Germany Universal 16R, Hettich, Germany
Computers	:	Intel (R), Pentium (R) 4 CPU 2.26 GHz 2.28 GHz, 256 MB RAM 40 GB hard disc, Yönsis, İstanbul
Deep-freezers	:	(-20) Bosch, Germany (-70) Sanyo, Japan
Documentation System	:	GelDoc Documentation System, BIO-RAD, USA
Electrophoresis Apparatus	:	Horizon 58, Model 200, BRL, USA
Heat Blocks	:	Thermostat Heater 5320, Eppendorf, Germany Techne, DRI-BLOCK DB 2A, USA
Magnetic Stirrers	:	Chiltern Hotplate magnetic Stirrerr, HS31, UK
MagNA Pure Compact Instrument	:	Version 1.0, Roche, Germany
Ovens	:	37^0C EN 400, Nuve, Turkey 56^0C, LEEK, UK
Refrigerator	:	4^0C Medicool, Sanyo, Japan Arçelik 4250T, Turkey
Spectrophotometer	:	CE 5502 Scanning Double Beam 5000 Series CECIL Elegant Technology, UK
Thermal Cyclers	:	Techne Progene, UK Techne Touchgene Gradient, Progene, UK
Vortex	:	Fisions WhirliMixer, UK
Water Baths	:	Memmert, Germany Köttermann, Labortechnik, Germany
Water Purification System	:	WaTech Water Technologies, Turkey

4. METHODS

4.1. DNA Extraction from Peripheral Blood Samples

4.1.1. Sodium Chloride (NaCl) Extraction

The peripheral blood samples (5-10 ml) were collected into vacutainer tubes, containing EDTA (K3), to prevent coagulation, and stored at 4^0C. Samples were transferred to 50 ml sterile Falcon tubes and cold cell lysis buffer (30 ml to 10 ml blood sample) was added. After addition of cell lysis buffer, samples were mixed thoroughly. The samples were left at 4^0C for 15 minutes in order to lyse cell membranes by the osmotic pressure provided by the positively charged ions of the lysis buffer. The EDTA present in the cell lysis buffer chelates free Mg^{2+} ions which are cofactors for most nucleases. Removing divalent ions from the solution prevents the activation of nucleases, and by this way, digestion of the nucleic acids.

After incubation at 4^0C for 15 minutes, samples were centrifuged at 5000 rpm and 4^0C for 10 minutes in order to collect leukocyte nuclei. The supernatant containing the RBC debris were discarded. The pellet containing the leukocyte nuclei were resuspended in 10 ml cell lysis buffer by vortexing. Centrifugation was repeated at 5000 rpm at 4^0C for 10 minutes, and supernatant was discarded again. If the pellet was not clean enough, centrifugation steps were repeated. The pellet was resuspended in 3 ml nuclei lysis buffer and vortexed to lyse the nuclear envelope of white blood cells. After addition of proteinase K (150 µg/ml) and SDS (0.14 per cent), the samples were incubated either at 56^0C for three hours, or overnight at 37^0C to provide degradation of cellular proteins. After the incubation, 5 ml sterile dH_2O and 5 ml of saturated NaCl (5 M) were added to the mixture for salting out protein residues. The tubes were shaken vigorously to precipitate the proteins, and then centrifuged at 5000 rpm for 30 minutes at room temperature. The supernatant was taken into another sterile Falcon tube and precipitated with two volumes of absolute ethanol. The tubes were gently inverted for several times until the DNA threads were visible in the solution. The precipitated DNA was fished out by the tip of a micropipette and transferred to an Eppendorf tube. After evaporation of the excess ethanol,

DNA was dissolved in an appropriate volume of TE buffer and stored at 4^0C (Miller *et al.,* 1988).

4.1.2. Nucleic Acid Purification with MagNa Pure Compact Instrument

In this procedure, approximately 400 µl peripheral blood sample was used for nucleic acid extraction. The principle of the method is based on the magnetic separation and purification of genomic DNA samples (www.roche-applied-science.com). Separated and purified DNA samples were dissolved in 200 µl elution buffer as recommended by the manufacturer.

4.2. DNA Extraction from Chorionic Villus Samples

The chorionic villus samples (CVS) were sent to our laboratory in a preservative solution and placed in a 15 ml falcon tube. The first centrifugation step was performed at 5000 rpm and 4^0C for 30 minutes. Then the pellet was resuspended with 1 ml cell lysis buffer and centrifuged again with previous conditions for 5 minutes; 200 µl nuclei lysis buffer was added and samples were mixed gently. Samples were transferred into a 1.5 ml Eppendorf tube and vortexed for 10 seconds before addition of 10 µl proteinase K and 15 µl SDS. They were incubated at 56^0C for 3 hours or at 37^0C for overnight. Then half volume of NH_4Ac (9,5 M) was added and mixed thoroughly. Next centrifugation step was performed at 10000 rpm for 5 minutes, and the supernatant was disposed. The pellet was resuspended with two volumes of EtOH and the last centrifugation was at 13000 rpm for 5 minutes. The pellet was left to dry after removal of the supernatant. The dried pellet was dissolved in 50 µl TE buffer.

4.3. Analysis of the Extracted DNA Samples

4.3.1. Qualitative Analysis by Agarose Gel Electrophoresis

The qualitative analysis of the extracted DNA was performed on one per cent agarose gels which were dissolved in 0.5X TBE buffer by boiling. After the solution was cooled down to 55^0C, 1.6 µl EtBr (10 mg/ml) was added to a final concentration of 0.5

mg/ml. Addition of EtBr is important for visualization of DNA under ultraviolet light (UV). The gel was poured into a gel plate and left at room temperature for polymerization with the combs inserted. DNA samples were mixed with BPB loading buffer to a final concentration of 1X and loaded onto an agarose gel. The gel was run at 150 V for 10 minutes. Since DNA fragments are negatively charged, the run occurs from the cathode to the anode. The larger fragments are retarded more than the smaller fragments in the gel. Therefore, DNA fragments are positioned in the gel according to their molecular weights during the run. As the genomic DNA has a very large molecular weight, it appears as a bulky spot under UV light. To estimate the quality and the quantity of the DNA, it should be diluted. A rough estimate can be done by comparing its intensity with DNA samples of known concentration.

4.3.1. Quantitative Analysis by Spectrophotometric Measurement

The exact concentration and purity of the isolated DNA can be determined by spectrophotometric measurement. Prior to spectroscopy DNA was diluted with dH_2O and placed in a quartz cuvette. The absorption of UV by DNA was read at 260 nm, and this value was used in the following formula to calculate the concentration of DNA samples:

50 µg/ml x Optical Density $(OD)_{260}$ x Dilution Factor = Concentration in µg/ml

An absorbance of 1.0 at 260 nm (OD_{260}) reflects the presence of 50 µg/ml double-stranded DNA. In order to determine the purity of DNA samples, the ratio of OD_{260}/OD_{280} was used. Pure samples were expected to have a ratio of 1.8. The OD_{260}/OD_{280} values greater than 1.8, indicate RNA contamination, while values less than 1.8 indicate protein contamination.

4.4. The β-Globin StripA^{ssay}

Conventional dot-blot hybridization method is based on application of radioactively labeled allele-specific oligonucleotide probes (wild type and mutant) to amplified DNA samples immobilized onto membranes (Saiki, 1986). According to the captured probes, it

is possible to identify β-globin alleles of an individual, but this method is of restricted value when a small number of samples have to be analyzed for various mutations.

The β-globin StripA^{ssay} is a commercially available mutation detection system which works with reverse dot-blot hybridization principle. In this system, oligonucleotide probes are immobilized onto nylon strips rather than the amplified DNA samples as in the conventional hybridization method. The recent 3rd version of the β-Globin StripA^{ssay} kit, detects 20 common Mediterranean specific β-thalassemia mutations (-87 [C-G], -30 [T-A], Codon 5 [-CT], Codon 6 [-A], Codon 8 [-AA], Codon 8/9 [+G], Codon 22 [7 bp del], Codon 30 [Γ-C], IVS-I-1 [G-A], IVS-I-2 [T-A], IVS-I-5 [G-C], IVS-I-6 [T-C], IVS-I-110 [G-A], IVS-I-116 [T-G], IVS-I-25 [25 bp del], Codon 36/37 [-T], Codon 39 [C-T], Codon 44 [-C], IVS-II-1 [G-A], IVS-II-745 [C-G]) and two probes for abnormal hemglobins (HbS and HbC). Some of the mutations covered by the assay are located within a few nucleotides on the β-globin gene, and are represented by a common wild type probe. Thus there are only 12 wild type probes for 22 mutant probes. This system has been tested (Bilenoğlu, 1996) and routinely used in our laboratory for the past 10 years and involves three stages: DNA isolation, amplification and biotin labeling, hybridization and color development.

4.4.1. DNA Isolation

The procedure was applied as recommended by the manufacturer. The DNA was isolated by using the extraction reagents included in the kit. The extraction stage started with addition of Lysis Solution to 100 µl of blood samples. After 15 minutes incubation at room temperature, samples were centrifuged at 3000 rpm for 5 minutes. Later, 1 ml of supernatant was removed and the pellet was resuspended with 1 ml Lysis Solution, again. Second centrifugation step was performed at 12000 rpm for 5 minutes and the supernatant was removed and discarded, except 20-30 µl of a visible pellet. Next step was addition of 200 µl of GEN^{X}TRACT Resin to the pellet, which chelates breakdown products. Samples were vortexed for 10 seconds and incubated at 56^{0}C for 20 minutes. Before 98^{0}C incubation for 10 minutes, samples were vortexed for 10 seconds, again. The last centrifugation step was performed at 12000 rpm for 5 minutes. The supernatant involves single stranded DNA, which is suitable for immediate *in vitro* amplification.

4.4.2. Amplification and Biotin Labelling

The β-globin gene was amplified in a multiplex PCR reaction which lasts in 70 minutes. The primers, PCR buffer and dNTPs are provided in the Amplification mix which is also included in the kit. 15 µl of Amplification mix, 1U Taq DNA polymerase, 5 µl DNA template (which is extracted according to kit protocol), were mixed and completed to a total of 25 µl with dH_2O. DNA samples extracted according to the NaCl method, can also be used to amplify β-globin gene for further analyses. If that was the case, 0.5 or 1.0 µg of genomic DNA was used as template. The thermal cycling conditions used were as follows:

Initial Denaturation	: 94^0C	2 minutes	
Denaturation	: 94^0C	10 seconds	} 35 cycles
Annealing	: 54^0C	15 seconds	
Extension	: 72^0C	45 seconds	
Final Extension	: 72^0C	3 minutes	

The reaction is based on multiplex DNA amplification involving five oligonucleotide primers, and the amplification is performed in the presence of biotin-dUTP. The products of this reaction were biotinylated fragments of 738 bp, 596 bp and 251 bp (Figure 4.1). In order to check amplification, 5 µl aliquots of PCR products were run on a one per cent agarose gel.

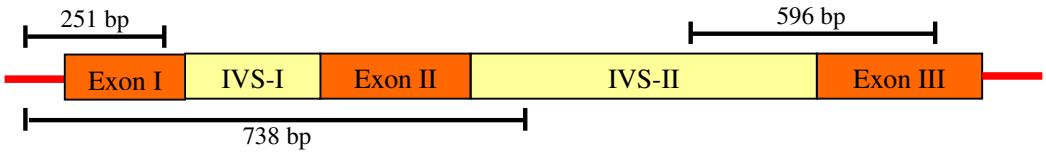


Figure 4.1. Schematic representation of the amplification products (Bilenoğlu, 1996)
(▬ 5'- and 3'-UTRs, IVS: intervening sequence)

4.4.3. Hybridization and Color Development

15 µl of amplification products were denatured with an equal volume of denaturing agent (DNAT) and incubated for 5 minutes at room temperature. After addition of 1 ml Hybridization Buffer test strips were inserted and hybridization reaction was performed at 45^0C for 30 minutes in a shaking water bath (approximately at 50 rpm). During the

hybridization step, biotinylated amplification products hybridize selectively to the mutant and wild type probes.

After incubation at 45^0C, the Hybridization Solution was removed and the strip was washed for 10 seconds with Wash Solution A in order to remove unbound DNA. Next two steps involved washing of strip again with Wash Solution A at 45^0C for 15 minutes in a shaking water bath. In each washing step, the previous strip solution was removed and replaced with 1 ml Wash Solution A.

After three washings to remove excess unbound DNA, an enzymatic color reaction is performed to visualize the result. The first step of enzymatic color reaction was addition of 1 ml Conjugate Solution for each lane and incubation at room temperature for 15 minutes on an orbital shaker (approximately at 50 rpm). Conjugate Solution contains streptavidine-alkaline phosphatase which bounds biotinylated PCR products. Then, the strip was washed with Wash Solution B in order to remove excess streptavidine-alkaline phosphatase. First wash was for 10 seconds, while second and third washes were for 5 minutes at room temperature on an orbital shaker (approximately at 50 rpm). Color development was performed by addition of 1 ml of Color Developer to test strips, and incubation at room temperature for 15 minutes in dark on an orbital shaker (approximately at 50 rpm). This incubation leads to interaction of the Color Developer with streptavidine-alkaline phosphatase to give color. After incubation with the Color Developer, test strips were washed with dH_2O or 1 mM EDTA several times and dried in dark on an absorbent paper. Interpretation of results was performed by comparing the band pattern with the decoder table (Figure 4.2).

The whole procedure takes approximately five hours, which clearly indicates its advantage over other conventional methods. The β-Globin StripA^{ssay}, with its precision, speed and convenience, was the first choice for mutation detection in this study. If the assay could not identify the mutation, the complete β-globin gene was sequenced in the next step.

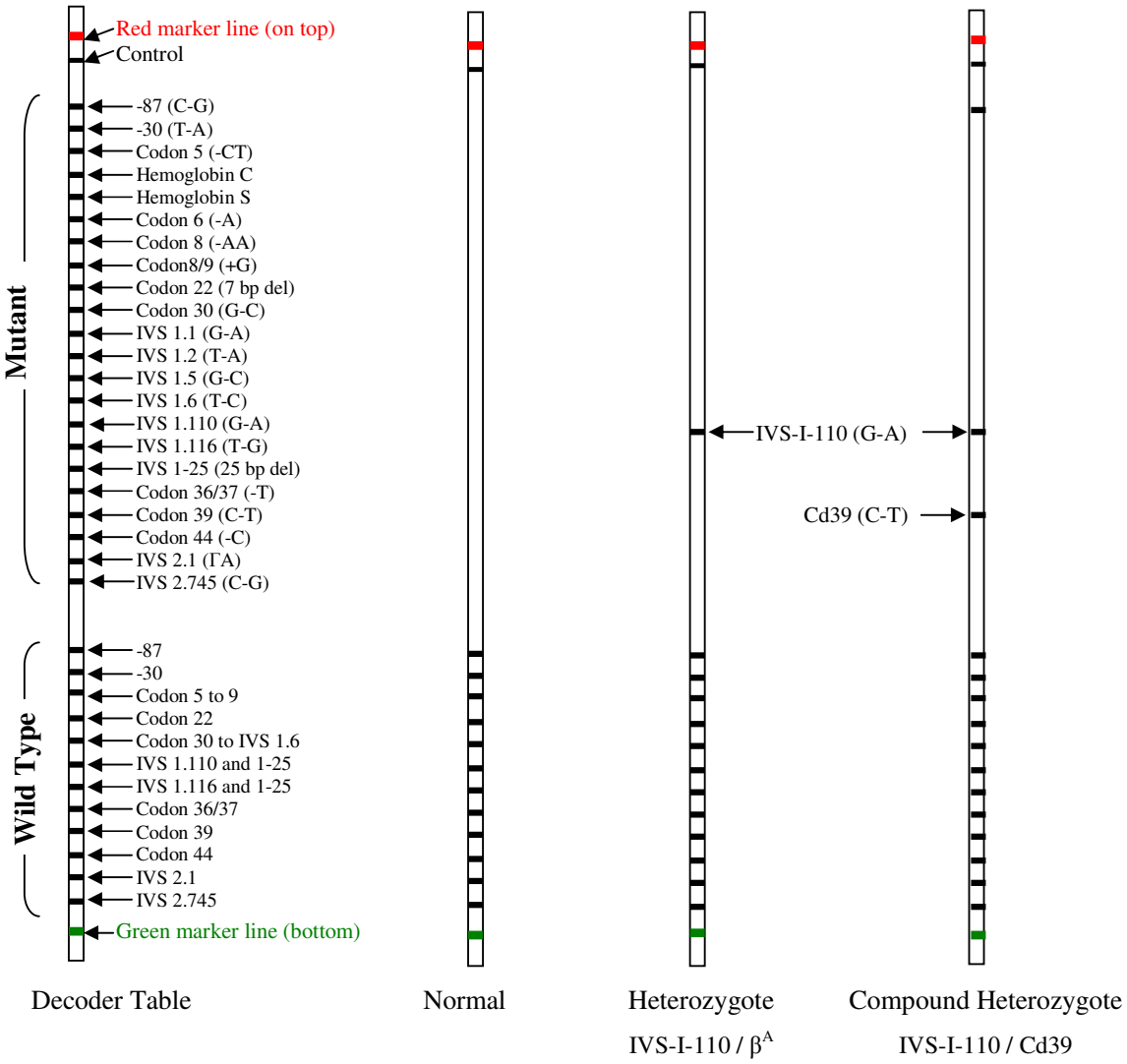


Figure 4.2. Decoder Table and interpretation of the β-Globin StripA^{ssay} results

4.5. Polymerase Chain Reaction

The Polymerase Chain Reaction (PCR) is an *in vitro* method for the enzymatic amplification of genomic DNA. The principle of PCR is based on hybridization of two synthetic oligonucleotide primers flanking the region of interest and amplification of this specified region. Repeated cycles, involving template denaturation, primer annealing and extension of the annealed primers by Taq polymerase enzyme, result in accumulation of the desired product. The boundaries of these fragments are determined via the 5' ends of the primers. The PCR products synthesized in one cycle serve as a template in the next, thus, the number of amplified products increases geometrically at every cycle. As a result,

30 cycles of PCR yields about a million-fold amplification of the target DNA sequence (Watson, 1992).

4.5.1. The β-Globin Gene

In order to sequence the β-globin gene for mutation detection, DNA samples (50-100 ng) which were extracted with the NaCl method or by the MagNA Pure Compact System were amplified with two specific primer pairs (Table 4.3). Reaction conditions of β-globin gene amplification are listed in Table 4.4.

Table 4.3. Primer pairs used to amplify the β-globin gene

Primer Pairs	Amplified Region	Product Size
CD7 (F) / CD6 (R)	-116 to IVS-II-116	772 bp
CD1 (F) / CD2 (R)	IVS-II-603 to 76 bp 3' to PolyA Signal	566 bp

Table 4.4. Reaction conditions for the β-globin gene amplification

Stock Concentration	Volume (in 25 µl)	End Concentration
10X $MgCl_2$-free Buffer	2.5 µl	1X $MgCl_2$-free Buffer
25 mM $MgCl_2$	2.5 µl	2,5 mM $MgCl_2$
25 mM dNTP	0.2 µl	0.2 mM dNTP
12,5 µM Forward Primer	1.0 µl	0,5 µM Forward Primer
12,5 µM Reverse Primer	1.0 µl	0,5 µM Reverse Primer
5U/µl Taq DNA Polymerase	0.2 µl	1U Taq DNA Polymerase

The thermal cycling conditions for the β-globin gene reaction are as follows:

Initial Denaturation	:	94^0C 3 minutes	
Denaturation	:	94^0C 30 seconds	} 30 cycles
Annealing	:	60^0C 1 minute 30 seconds	
Final Extension	:	72^0C 5 minutes	

4.5.2. Variable Number of Tandem Repeats

Variable Number of Tandem Repeats (VNTRs) are scattered throughout the human genome. Although they have no importance, they are very informative in the exclusion of maternal contamination in prenatal diagnosis, since they show high heterogeneity even in relatives. Because the quality of the material in CV sampling is directly dependent on the skill of the obstetrician performing the biopsy, there is always a small possibility to analyze a CV sample contaminated with maternal genetic material.

One of the indications of maternal contamination are the fetuses showing the same genotype like the mother, although this may not be necessarily due to contamination. To avoid false negative results, the DNA samples of mother, father, and fetus were amplified using primers flanking a VNTR region. The banding pattern for each DNA should be different. If the mother and the fetus, however, show the same banding pattern, this indicates maternal contamination. If all family members show the same banding pattern, this would indicate that the VNTR region chosen is not heterogeneous in the present family, thus not informative, and another VNTR region should be examined.

In the framework of this thesis, two primer pairs were used to identify probable contaminations in the case of prenatal diagnosis (Table 4.5). Reaction conditions for VNTR analyses are listed in table 4.6.

Table 4.5. Primers used for VNTR Analyses

Primer Pairs	**Amplified Region**
St14-A (F) / St14-B (R)	DXS52 Loci (harbours a highly polymorphic 60 bp repeats)
XY-1 (F) / XY-2 (R)	Amplifies VNTR of a gene (AMEL) localized on both X and Y chromosomes

Table 4.6. Reaction conditions for VNTR analyses

Stock Concentration	Volume (in 25 µl)	End Concentration
10X $MgCl_2$-free Buffer	2.5 µl	1X $MgCl_2$-free Buffer
25 mM $MgCl_2$	2.5 µl	2.5 mM $MgCl_2$
25 mM dNTP	0.2 µl	0.2 mM dNTP
6.25 µM Forward Primer	2.0 µl	0.5 µM Forward Primer
6.25 µM Reverse Primer	2.0 µl	0.5 µM Reverse Primer
5U/µl Taq DNA Polymerase	0.2 µl	1U Taq DNA Polymerase

The thermal cycling conditions for the St14 and XY probes VNTR analyses are as follows:

St-14 VNTR Analysis

Initial Denaturation	:	94^0C	4 minutes	
Denaturation	:	94^0C	30 seconds	} 28 cycles
Annealing	:	62^0C	3 minutes and 30 seconds	
Final Extension	:	62^0C	5 minutes	

XY-Probe VNTR Analysis

Initial Denaturation	:	95^0C	2 minutes	
Denaturation	:	95^0C	30 seconds	} 28 cycles
Annealing	:	60^0C	1 minute and 30 seconds	
Final Extension	:	60^0C	5 minutes	

4.5.3. Differential Amplification of the $^G\gamma$-Globin Gene Promoter

In a previous study performed in our laboratory, the importance and haplotype of the $^G\gamma$-globin gene promoter polymorphism (at -158) for β-thalassemia was established (Tadmouri, 1999). This polymorphic site is very well known as the -158 (C/T) *XmnI* $^G\gamma$-globin gene polymorphism and has an important regulatory role on the pathophysiology of β-thalassemia.

There is a strong homology between $^{G}\gamma$- and $^{A}\gamma$-globin gene promoters. Therefore, selective amplification of the $^{G}\gamma$-globin promoter was provided by the use of specific primers hybridizing to the $^{G}\gamma$-globin promoter (Table 4.7). The primers are designed as Amplification Refractory Mutation System (ARMS) primers, thus display mismatches with $^{A}\gamma$-globin promoter and provide specificity. Reaction conditions for $^{G}\gamma$-globin gene promoter amplification is demonstrated in Table 4.8.

Table 4.7. Primers used in $^{G}\gamma$-globin gene promoter amplification

Primer Pairs	**Amplified Region**	**Product Size**
G5 (F) / G4 (R)	-25 to -383	349 bp

Table 4.8. Reaction conditions of $^{G}\gamma$-globin gene promoter amplification

Stock Concentration	**Volume (in 25 µl)**	**End Concentration**
10X $MgCl_2$-free Buffer	2.5 µl	1X $MgCl_2$-free Buffer
25 mM $MgCl_2$	1.5 µl	1.5 mM $MgCl_2$
25 mM dNTP	0.25 µl	0.25 mM dNTP
12.5 µM Forward Primer	1.0 µl	0.5 µM Forward Primer
12.5 µM Reverse Primer	1.0 µl	0.5 µM Reverse Primer
5U/µl Taq DNA Polymerase	0.2 µl	1U Taq DNA Polymerase

The thermal cycling conditions for the $^{G}\gamma$-globin gene promoter are as follows:

Initial Denaturation	:	94^0C	1 minute	
Denaturation	:	94^0C	30 seconds	30 cycles
Annealing	:	52^0C	30 seconds	
Extension	:	72^0C	1 minute	
Final Extension	:	72^0C	5 minutes	

For each amplification reaction, 50-100 ng genomic DNA was used in a total of 25 µl reaction volume. The total volume of each PCR reaction was completed to 25 µl with dH_2O. Amplified products were examined on a two per cent agarose gel at 175 V for approximately 15-30 minutes.

4.6. PCR Purification

The QIAquick PCR Purification Protocol was used to purify the PCR products form excess primers, dNTPs, polymerases and salts prior to DNA sequencing. The principle of this system is dependent on adsorption of DNA by silica-gel membrane in the presence of high salt (Qiaquick Spin Handbook, 1997). Purification procedure was applied to PCR products obtained by β-globin gene amplification. After purification, samples were sent to Iontek Ltd. for automated DNA sequencing.

4.7. Restriction Endonuclease Analysis

Some mutations or polymorphisms may create or abolish the recognition site of a restriction enzyme. In this case, the nucleotide change can be detected by digesting a PCR amplified target fragment with the corresponding restriction enzyme. When direct DNA analysis is not informative to identify the mutation causing β-thalassemia, restriction endonuclease analysis is performed.

EcoRI (G↓AATTC) was used to identify two types of abnormal hemoglobins caused by different substitutions at Codon 121 (HbD LosAngeles and HbO Arab). For these analyses, 8 µl of the amplified DNA was digested in a 10 µl reaction mixture with 2 units of *EcoRI* enzyme. The mixture was incubated at 37^0C for 2-4 hours or overnight.

PdmI (GAANN↓NNTTC), isoschizomere of *XmnI*, was used to identify the C/T polymorphism at position -158 of the $^{G}\gamma$-globin gene promoter. Like in the *EcoRI* digestion, 8 µl of the amplified DNA was digested in a 10 µl reaction mixture with 2 units of the *PdmI* enzyme. The mixture was incubated at 37^0C for 2-4 hours or overnight.

The digestion products were run on a two percent agarose gel for about 30 minutes at 150 V; the banding patterns were visualized under UV-light.

4.8. FileMaker Pro 6 Software

Planning is the most important issue for establishing a FileMaker database. Planning a database before its construction will save time and effort. There are several steps which must be performed in order to build a database. First of all, the reason for constructing a database should be clearly identified. This decision is the most important point in determining the type of the database to be created. Later, dependent on the information categories, it is required to find out, if there is any need to plan separate database files for each category or not. After this step, the fields should be planned according to the value types involved within each information category. Careful organization (planning) of these steps will reveal if there is any need for relationships between separate File Maker files. It is also essential to group values, having connection, in layout form to simplify the complex picture of the huge amount of information within a database. The last step of database construction is the decision for either sharing databases via the Local Area Network (LAN) or via the World Wide Web.

FileMaker serves four modes for designing database layouts and for working with the data involved in the database.

Table 4.9. Modes of FileMaker

Mode	Function
Browse	Work with the data file: add, view, change, sort, omit, or delete records
Find	Search for particular records that match a set of criteria
Layout	Determine how your information is presented on your screen
Preview	See how data in forms or reports will look before printing

Fields which are going to store values are created by the “Define Fields” dialog box (Figure 4.3). It is possible to define the type of the field and names of the fields dependent on the value type of interest. FileMaker provides eight types of fields; field types and descriptions are listed on Table 4.10.

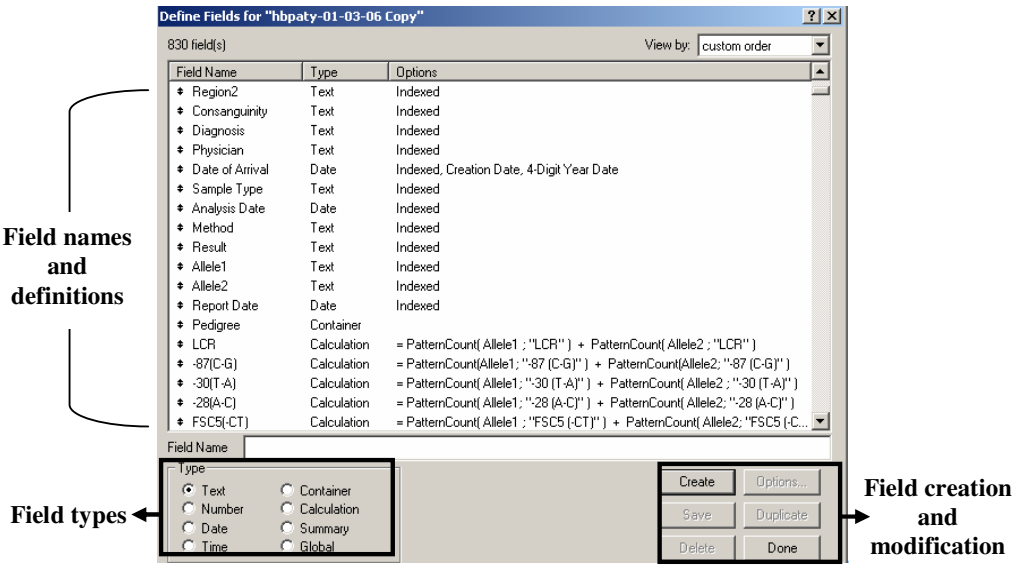


Figure 4.3. "Define fields" dialog box in FileMaker Pro 6

Table 4.10. Field types in FileMaker

Field	Characteristics
Text	Up to 64000 letters, symbols, or numbers used as text. Text fields may contains carriage returns
Number	Up to 255 numbers or other characters. Also can contain Boolean values, to indicate, i.e., true, false, yes and no. Unlike other text values in a number field, a search can be performed on Boolean values
Date	Dates only
Time	Times only
Container	A picture or a multimedia file. Container fields in calculations and summary fields can be referenced. Windows: A container field can store OLE objects
Calculation	The result of a calculation formula that uses field values from the current record or related records. The formula can use values of all field types. The result can be one of these types of data: test, number, date, time, or container
Summary	A value that is produced by summarizing field values from more than one record in the same file
Global	One value to be used in all records of the file. A global field can contain text, number, date, time, or container data. Use the value of a global field in calculations and scripts. A global field can be used to find records.

4.8.1. Construction of Databases using FileMaker Pro 6 Software

The first objective of database construction in the framework of this thesis, was the need to easily manage the huge amount of patient data in our laboratory. It was possible to divide the data into two major categories: Patients and referring clinician data. For that reason, two FileMaker databases (files) were constructed: Patients and referring clinicians

databases. The fields created in each database, and details of construction steps are described in detail below.

4.8.1.1. Patients Database. After dividing the raw data into two logical parts, the next step was determination of the layouts, required to distribute the most relevant values. For this reason, three types of layouts were designed: Personal Information, Working and Sample Information layouts.

Specific fields were created to display the data involved within these layouts. At first instance the number of fields required to build the database were determined according to the value types present. The name of the layouts, fields involved within and their types are presented in Table 4.11. Some of the fields were indexed in order to display the entries done previously. By this way, it was possible to select a previous entry without re-typing. Clinician field was indexed also, and the present index was fed from the referring clinician database by a relationship between both databases. The number of the fields was increased to more than 800 when the need for data analyses increased within the database. As mentioned above, FileMaker provides analysis tools as well, in addition to storage options. For displaying analysis results, a layout named Distribution layout was created. To perform this task, specific formulas calculating allele distributions among geographical regions of Turkey were written. In addition, fields displaying allele percentages were created.

Design process of the layouts was performed in layout mode described above. In this mode, it is possible to modify field names, location of the fields, color or font setups. Data entry can be performed by using FileMaker Pro "import option" served for various protocols. However, an electronic copy of these data was not present; for this reason this database was fed manually.

After designing the simplest level database, the next step was its modifications. The first modification involved, links to specific World Wide Web browsers, to complete the data in patients database. These web pages were HbVar Database, the electronic version of Huisman's Syllabi on Hemoglobin Variants and Thalassemias (Huisman *et al.*, 1996; Huisman *et al.*, 1997) to provide information about mutations and their relevant effect on phenotypes. Iontek Ltd. web page was linked to provide easy access to sequencing results

without any need for opening a new explorer page. The last web page link was devoted to ViennaLab, Austria home page, to provide easy access to the information required on the β-Globin StripA^{ssay} kit.

Table 4.11. Layouts, fields and field types used in construction of patients database

Layout	Field	Field type
Personal Information	Identity Number	Number (auto serial no entry)
	Name	Text
	Birth Place	Text (Indexed)
	Date of Birth	Date (Indexed)
	Adress	Text
	City	Text
	Office	Text
	GSM	Text
	Home	Text
	Fax	Text
	Email	Text
Background and Molecular Analysis Layout	Identity Number	Number (auto serial no entry)
	Name	Text
	Origin	Text (Indexed)
	Physician	Text (Indexed)
	Region	Text (Indexed)
	Allele	Text (Indexed)
	Pedigree	Container
	Result	Text
	Consanguinity	Text (Indexed)
	Method	Text (Indexed)
Sample Information	Physician	Text (Indexed)
	Sample Type	Text (Indexed)
	Date of Arrival	Date
	Analysis Date	Date
	Report Date	Date

The second group of modifications involved creation of layouts for providing information about the protocols used to perform analyses on β-thalassemia in our laboratory. Last modification was the specification of buttons required for working in the database. FileMaker supplies over 100 C++ based scripts in 11 categories, like navigation and control. Some of the scripts written to do modifications and buttons are listed in Table 4.12.

Table 4.12. Scripts used in patients database construction

Name	Script	Funtion
Search	Perform Find Go to layout View as	Perform search
Go to HTML	Go to Layout (HTML)	Go to current record found
Go Home	Open URL	Open specified web page
Next	Go to Record/Request/Page (Next)	Forward to next record
Previous	Go to Record/Request/Page (Previous)	Backward to previous record
First	Go to Record/Request/Page (First)	Jump to first record
Last	Go to Record/Request/Page (Last)	Jump to last record

4.8.1.2. Referring Clinicians Database. Referring Clinicians database was constructed in the same way as the patients database. All of the protocols performed were the same, but it had a simpler structure than the patients database. It is composed of one layout involving institute, clinician name and contact numbers.

After building-up processes were completed, the last step remained was publishing. LAN publishing was preferred for using the databases in laboratory setup. The database was published with Instant web publishing, for this reason there was not any need to build any other layouts. After publication, it was realized, that there should be a CDML page in order to provide addition of new records to the database. Recently, the databases were linked to a web page which was designed in our laboratory (www.ndal.boun.edu.tr).

5. RESULTS

5.1. Genomic DNA Samples

Most of the DNA samples used in the framework of this study were extracted with the NaCl method; in the extraction of the very recent samples the MagNA Pure Compact Instrument was used.

5.1.1. DNA Samples Extracted with NaCl Method

Different DNA samples extracted from peripheral blood (10 ml) with the NaCl method are shown in Figure 5.1 (a). Prior to the amplification reaction, the DNA samples were diluted, as shown in Figure 5.1 (b).

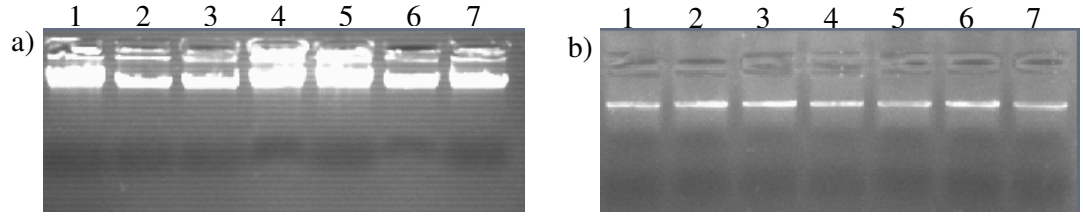


Figure 5.1. (a) Genomic DNA samples of different individuals (b) dilutions of the same samples run on a one per cent agarose gel

5.1.2. DNA Samples Extracted with the MagNA Pure Compact Instrument

The DNA samples extracted with the MagNA Pure Compact Instrument were pure and enough diluted to be used immediately for PCR amplification. The samples extracted with the MagNA Pure Compact Instrument, run on an agarose gel, are shown in Figure 5.2.

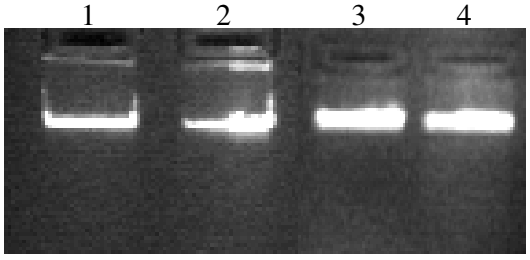


Figure 5.2. DNA samples run on a one per cent agarose gel

5.2. Molecular Analysis of the β-Globin Gene

5.2.1. β-Globin Gene StripA^{ssay} Results

In vitro amplification of β-globin gene was performed by a single multiplex PCR reaction, which amplifies three fragments of 738 bp, 596 bp, and 251 bp (Figure 5.3). This amplification reaction was performed with five oligonucleotide primers: the primers amplifying the 5' end of the 738 bp and 251 bp fragments are common. The first exon of the β-globin gene was amplified twice, to enhance the quality of the test, since the mutations involved on the Strip are concentrated in this region of the gene.

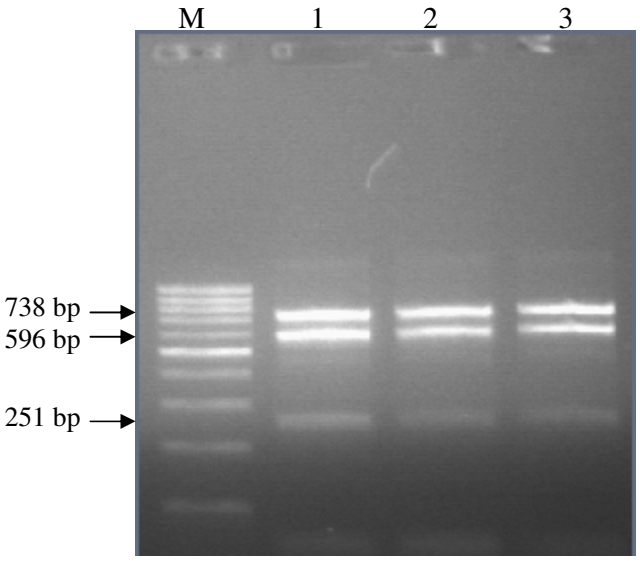


Figure 5.3. Agarose gel electrophoresis of PCR products: 5 ml of each product was run on a 1 per cent agarose gel. Lanes 1-3: PCR products of three different patients; M: 100 bp DNA ladder.

After testing the quality of PCR products on an agarose gel, the samples were subjected to hybridization. The samples were screened for 20 β-thalassemia and two abnormal Hb (HbS and HbC) mutations, provided in the 3rd version of the β-Globin StripA^{ssay} Kit, in a single reaction. In addition to the mutated probes for each mutation type, wild type probes are represented on the test strips, as well. Adjacent mutations like Codon 30 (G-C), IVS-I-1 (G-A), IVS-I-2 (T-A), IVS-I-5 (G-C) and IVS-I-6 (T-C) are represented on a single wild type probe.

In normal individuals, all wild type probes were observed on the test strips without any mutant probe signal. An example of the β-Globin StripA^{ssay} for normal individuals is shown in Figure 5.4 (a). In heterozygous state, amplified DNA products hybridize to all of the wild type probes and only to a mutant probe according to the mutation present. In Figure 5.4 (b) and (c) two examples of heterozygous individuals are represented. In compound heterozygous state, it is possible to observe two mutated probes with all of the wild type probes, as represented in Figure 5.5 (a). However, if the mutations are adjacent to each other with few nucleotides and represented with a common wild type probe, two mutant probes are observed with unbound referring wild type probe as shown in Figure 5.5 (b). In homozygous state, the amplified DNA hybridizes to the relevant mutant probe and to all other wild type probes, except to the one referring to the mutation. The β-Globin StripA^{ssay} results of two homozygotes are illustrated in Figure 5.5 (c) and (d).

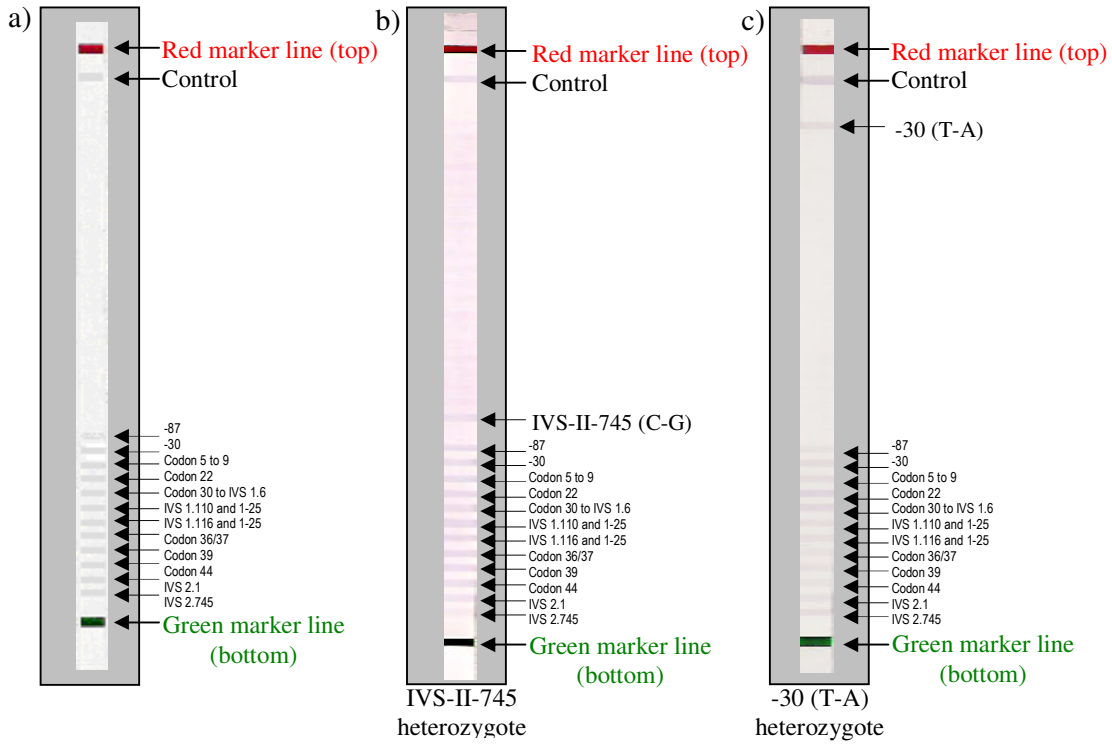


Figure 5.4. Examples of β-Globin StripA^{ssay} results: (a) normal individuals; (b) and (c) heterozygous individuals

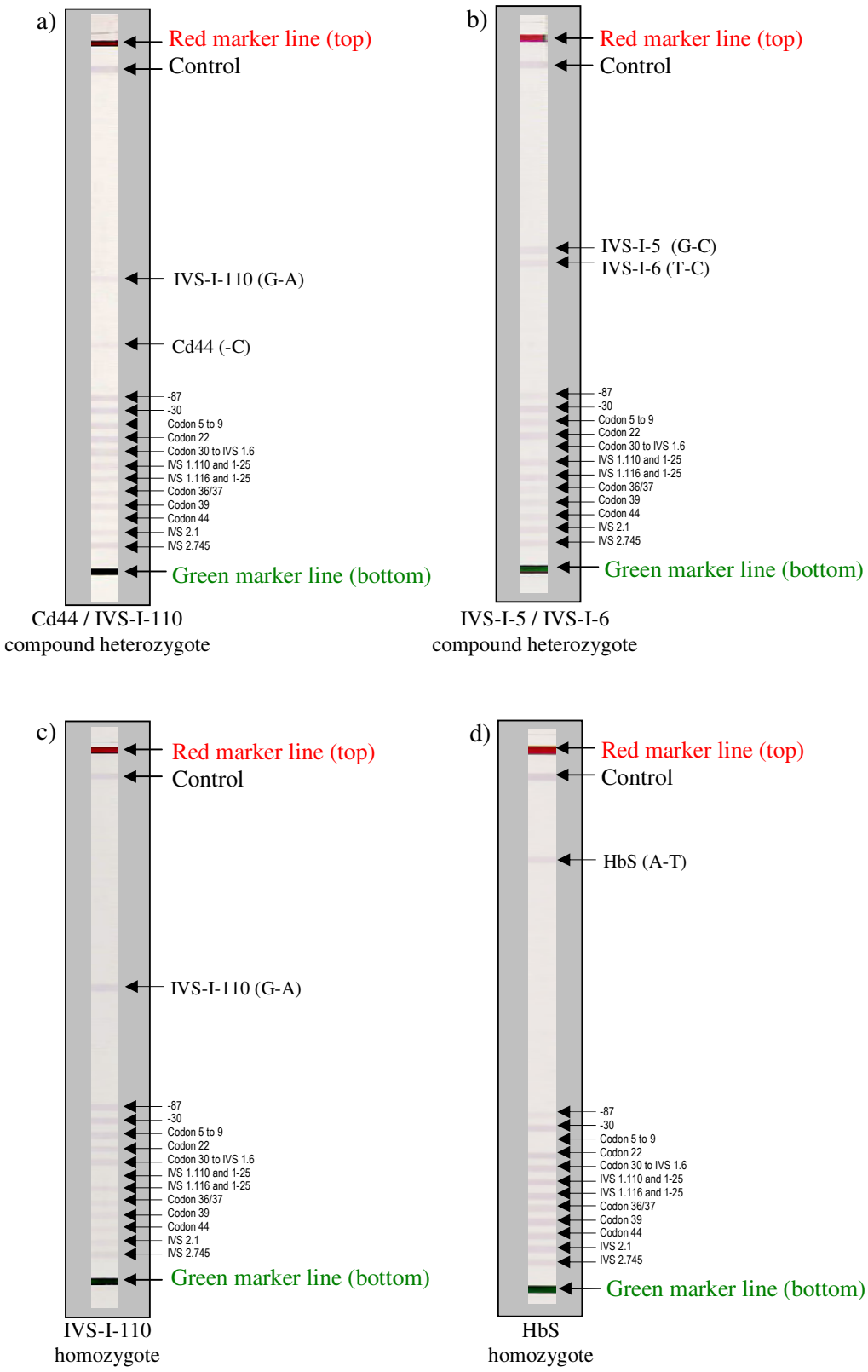


Figure 5.5. Examples of β-Globin StripA^{ssay} results:

(a) and (b) compound heterozygotes; (c) and (d) true homozygotes

5.2.2. β-Globin Gene Amplification and Purification Results

Amplification of the β-globin gene with specific primers was performed according to the protocols in Section 4.5.1. The amplified fragments are shown in Figure 5.6.

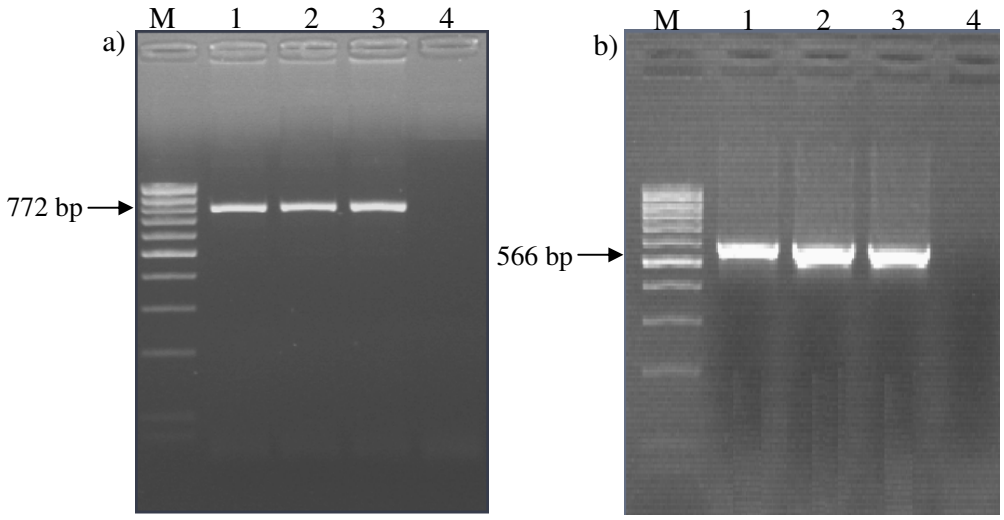


Figure 5.6. β-globin gene amplification: (a) 772 bp and (b) 566 bp, lanes 1-3: PCR products from three different patients, lane 4: negative control (no DNA), M: 100 bp DNA ladder

After performing agarose gel electrophoresis in order to see the quality of amplified products, samples were purified prior to DNA sequencing (Figure 5.7).

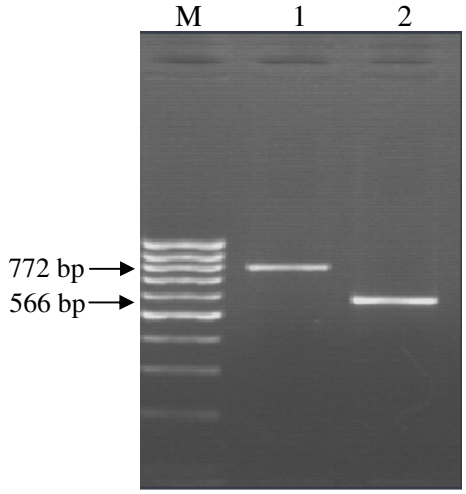


Figure 5.7. Gel electrophoresis of purified products: lane 1: purified 772 bp PCR product, lane 2: purified 566 bp product on a two per cent agarose gel, M: 100 bp DNA ladder

5.2.3. Direct DNA Sequencing Results

Beta-thalassemia and abnormal hemoglobin mutations which could not be defined by the mutations covered by the β-Globin StripA[ssay], were identified by direct DNA sequencing. The mutations identified by direct DNA sequencing are described below.

5.2.3.1. IVS-II-2 (T-A): Family M. IVS-II-2 (T-A) is a novel mutation defined in our laboratory in the framework of this thesis. The index family originates from the Black Sea Region of Turkey. The 11-year old proband and her mother are typical β-thalassemia carriers, with elevated HbA_2 levels and decreased red blood cell indices. The proband was referred to our center for the definition of her β-thalassemia mutation. The pedigree of the family is represented in Figure 5.8 and the proband is indicated with an arrow.

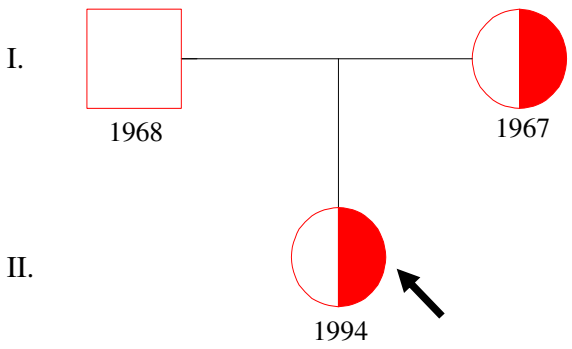


Figure 5.8. Pedigree of family M

Screening for the mutation causing the β-thalassemia carrier phenotype of the mother and proband could not be identified with the β-Globin StripA[ssay]. Sequencing of the β-globin gene revealed the presence of a T to A substitution at the second nucleotide of IVS-II. The result was confirmed by using the reverse primer, and performing a new PCR reaction. The same substitution was also shown in the mother (Figure 5.9).

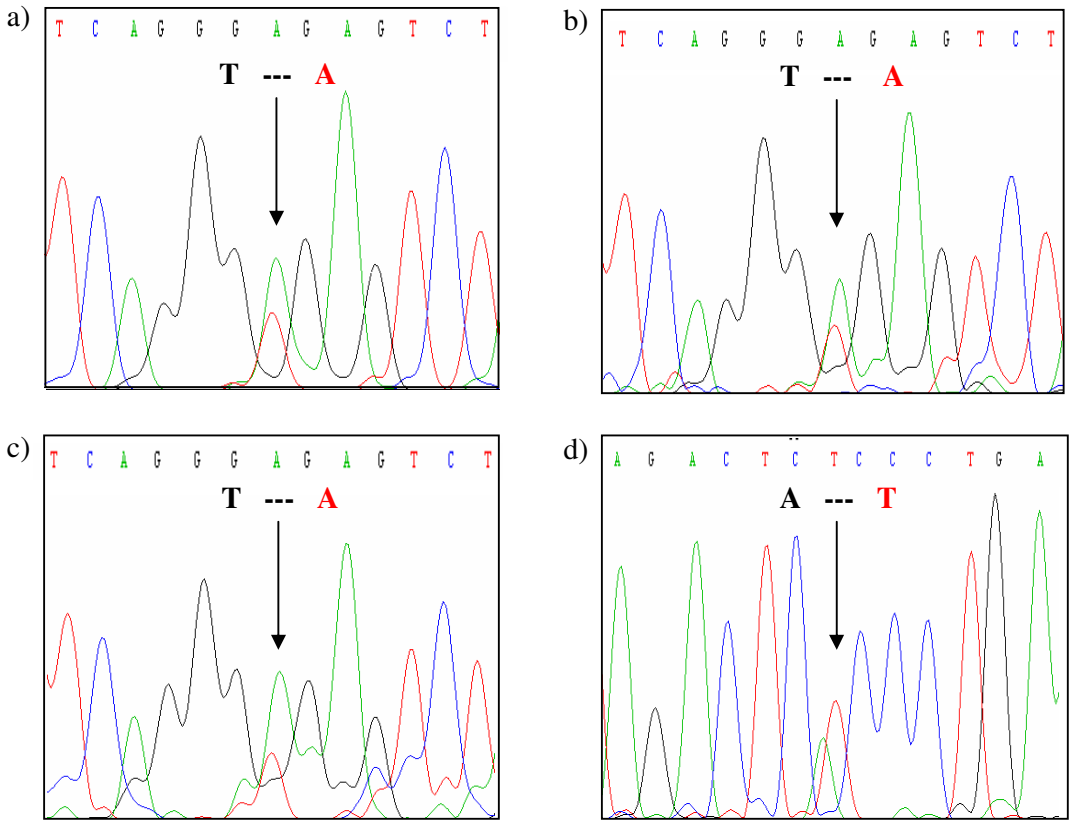


Figure 5.9. Sequencing results of the proband: (a) and (b) with forward primer, (d) with reverse primer; (c) sequencing result of the mother with forward primer

5.2.3.2. Codon 37 (TGG-TGA): Family B. This mutation results from the G to A substitution at Codon 37, replacing tryptophan with a nonsense codon (stop codon). This substitution causes a premature termination of translation and results in a β^0-thalassemia phenotype in affected individuals (Huisman *et al.*, 1997).

In this study, the Codon 37 (TGG-TGA) mutation was identified in a family from Bingöl in the Eastern part of Turkey. The parents are first cousins; their first child died at the age of four (II.1 in Figure 5.10) due to untreated β-thalassemia major. The proband (II.3 in Figure 5.10) is a two-year old girl diagnosed with β-thalassemia major. The pedigree of the family is shown in Figure 5.10.

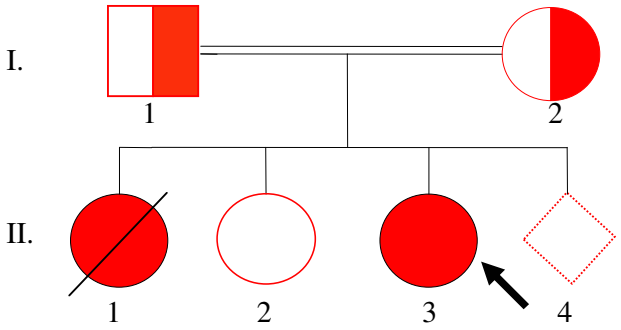


Figure 5.10. Pedigree of family B

The β-Globin StripA^{ssay} did not reveal the presence of any mutation for both parents (Figure 5.11a). In the proband the Codon 36/37 wild type hybridization band was missing without any signal in the corresponding mutant region (Figure 5.11b).

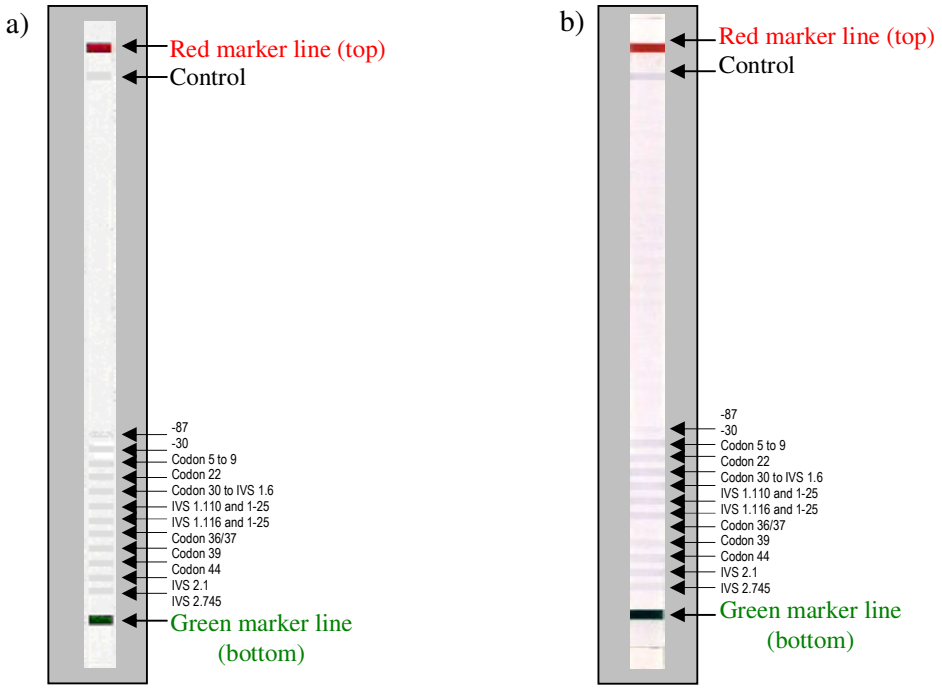


Figure 5.11. β-Globin StripA^{ssay} results of mother (a) and the proband (b)

In order to define the mutation of the family, sequencing was performed both for the proband (II.3 in Figure 5.10) and her mother (I.2 in Figure 5.10), which revealed proband was homozygous for Codon 37 (TGG-TGA) mutation, while her mother was heterozygous for the same lesion (Figure 5.12).

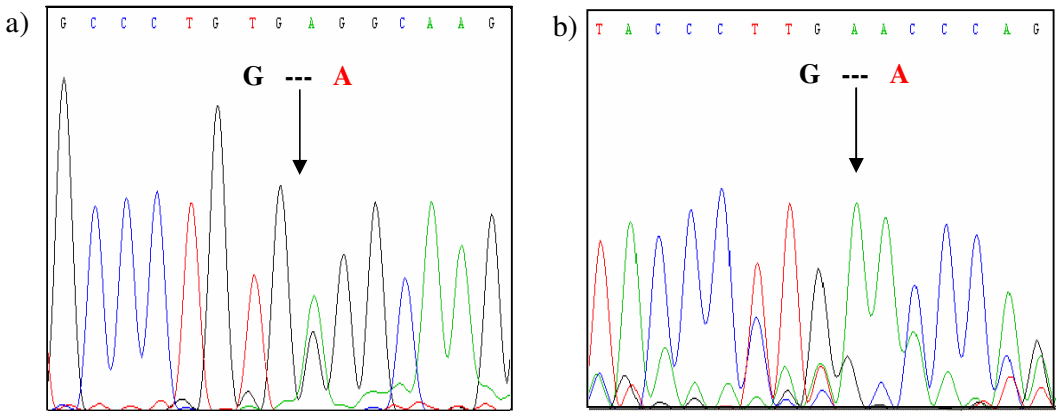


Figure 5.12. Sequencing results of family B: (a) mother Cd37 (TGG-TGA) heterozygote, (b) proband Cd37 (TGG-TGA) homozygote

5.2.3.3. Codon 15 (TGG-TGA): Patient A.F.S. and Family O. The Codon 15 (TGG-TGA) mutation causes premature termination of the β-chain translation replacing tryptophan to a stop codon (G ⟶ A in the gene) (Huisman *et al.*, 1997). Patient A.F.S., was a three-year old boy, diagnosed as β-thalassemia minor. Screening with β-Globin StripA^{ssay} did not reveal any mutations. Sequencing of the entire β-globin gene identified that he had inherited the Codon 15 (TGG-TGA) mutation in heterozygous form (Figure 5.14a).

The pedigree of family O. is given in Figure 5.13., the proband is indicated with an arrow. Diagnosed with β-thalassemia major, he was referred to our laboratory for the definition of his genotype. Screening with β-Globin StripA^{ssay} revealed the presence of IVS-I-110 (G-A) mutation in heterozygous form. In order to define the other mutation, DNA sequencing was performed, which revealed that the IVS-I-110 (G-A) mutation was found in compound heterozygosity with the Codon 15 (TGG-TGA) mutation. The same mutation was identified in his father in heterozygous form (Figure 5.14).

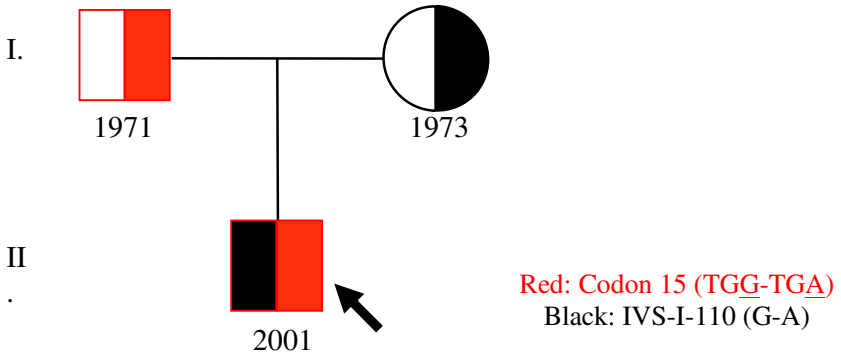


Figure 5.13. Pedigree of family O.

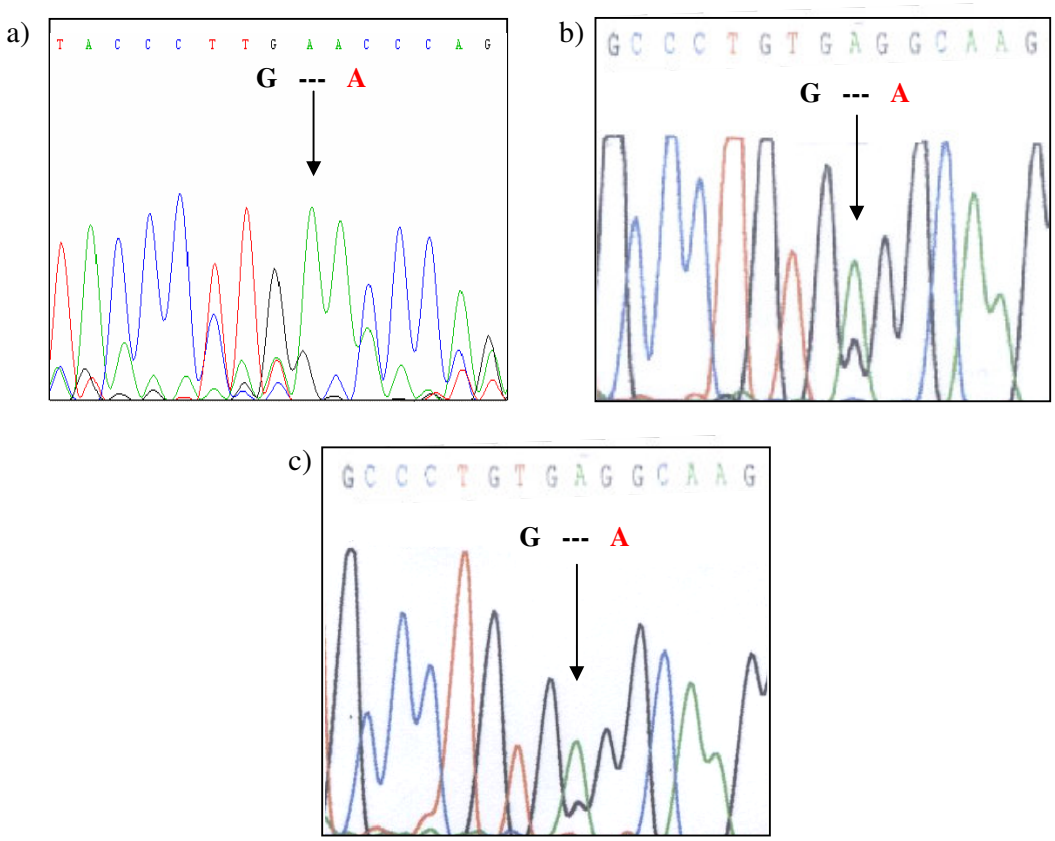


Figure 5.14. Sequencing result of patient A.F.S. (a), result of family O. (b) the proband and (c) his father

5.2.3.4. Codon 30 (AGG-ACG): Family E. The Codon 30 (AGG-ACG) mutation was identified in compound heterozygous form with the IVS-I-110 (G-A) mutation in two siblings of family E originating from Tekirdağ in the Marmara Region of Turkey. These individuals were diagnosed as β-thalassemia major patients, and the β-Globin StripA^{ssay} detected both mutations. However, it was not clear which G was affected, since there are two different mutations at this codon. The pedigree of the family and the sequencing results are shown in Figure 5.15 and Figure 5.16, respectively.

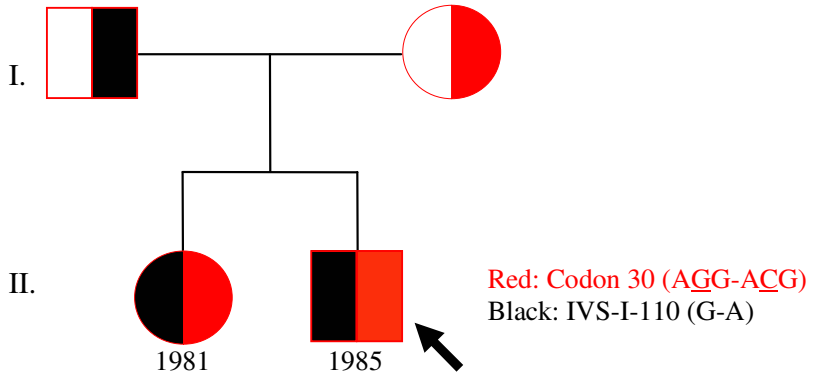


Figure 5.15. Pedigree of family E

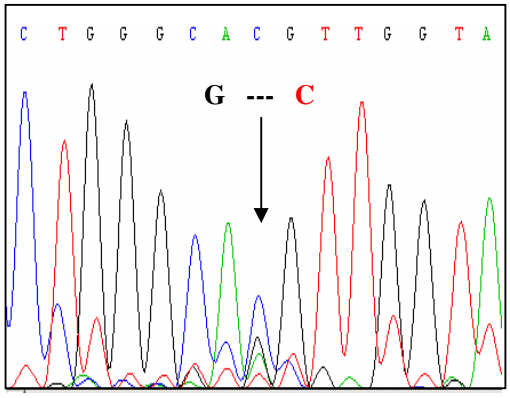


Figure 5.16. Sequencing result of family E

5.2.3.5. IVS-II-848 (C-A): Family G. The IVS-II-848 (C-A) mutation diminishes splicing due to the C to A substitution, next to the invariant AG dinucleotide and causes β^{+}-thalassemia (Huisman *et al.*, 1997). The frequency of this mutation in Turkish population was found to be 0.4 in a study performed with 795 β-thalassemia chromosomes (Tadmouri *et al.*, 1998).

This rare mutation was identified in a Turkish family, who had migrated from Thessaloniki to Turkey. The pedigree of the family is shown in Figure 5.17., the proband is marked with an arrow.

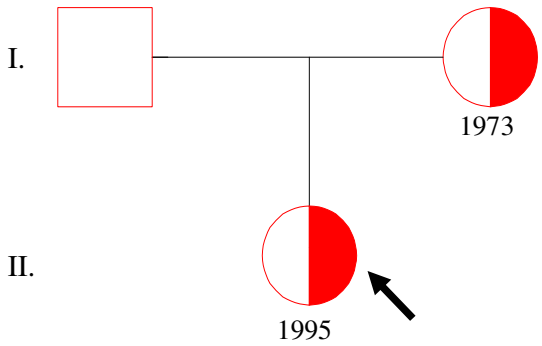


Figure 5.17. Pedigree of family G.

Sequencing of the β-globin gene revealed the rare IVS-II-848 (C-A) mutation (Figure 5.18) in the proband and her mother in heterozygous form.

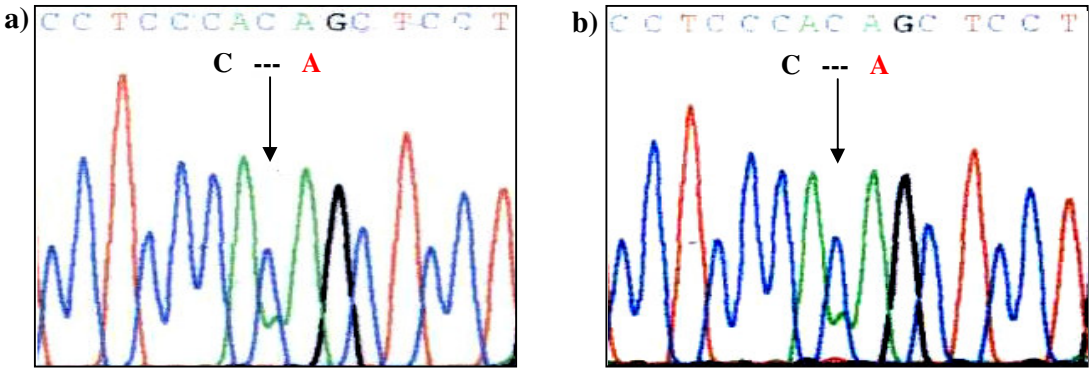


Figure 5.18. Sequencing results of family G., (a) proband and (b) her mother

5.2.3.6. -28 (A-C): Patient H.A. This mutation decreases the efficiency of β-globin gene transcription due to its reducing effect on binding of erythroid factors to the promoter. It gives rise to β^{+}-thalassemia in alleles carrying this mutation (Huisman *et al.*, 1997).

In our laboratory, this mutation was identified in a patient who was diagnosed with β-thalassemia major. The proband was referred to our center for the definition of her β-thalassemia mutations. The β-Globin StripA^{ssay} revealed only the Codon 39 (CAG-TAG) mutation in heterozygous form. Sequencing of the β-globin gene revealed the -28 (A-C) mutation (Figure 5.19). The patient was compound heterozygous for the -28 (A-C) and Codon 39 (CAG-TAG) mutations.

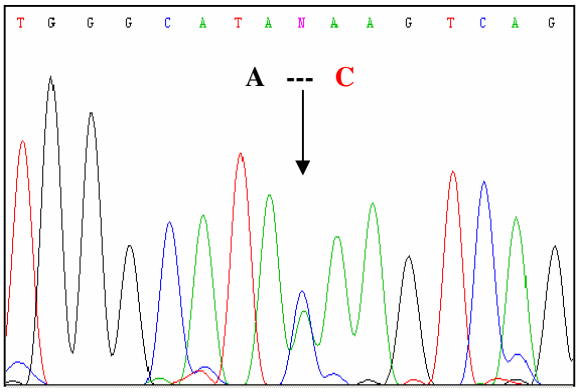


Figure 5.19. Sequencing result of patient H.A.

5.2.3.7. Hb Ankara (Cd10 GCC-GAC): Patient G.Y. This rare abnormal Hb variant was first identified in a Turkish family. It is characterized by a missense mutation, leading to alteration of alanine with asparagine (GCC-GAC) at Codon 10 of the β-chain (Huisman *et al.*, 1996).

In the framework of this thesis, Hb Ankara was identified in a patient from Isparta in the Mediterranean Region. The β-globin gene was sequenced and the presence of the C to A substitution at Codon 10 was identified (Figure 5.20).

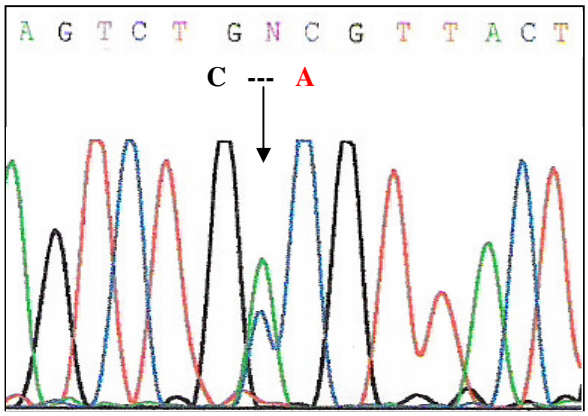


Figure 5.20. Sequencing result of patient G.Y.

5.2.3.8. HbO Arab: Patient A.K. This Hb variant is identified by the G to A substitution at Codon 121 of the β-globin gene. This substitution replaces glutamic acid with valine (Huisman *et al.*,1996). Patient A.K. is from the Mediterranean Region of Turkey.

Sequencing of the entire β-globin gene revealed the heterozygosity of the patient for Codon 121 (GAA-AAA), leading to a HbO Arab production (Figure 5.21).

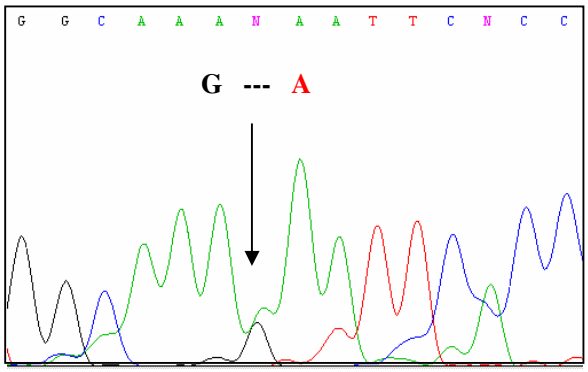


Figure 5.21. Sequencing result of patient A.K.

5.2.3.9. HbD LosAngeles: Family AN. HbD LosAngeles is an Hb variant produced by the G to C substitution again at Codon 121. This substitution replaces glutamic acid (GAA) with glycine (CAA) in the β-chain (Huisman *et al.*, 1996). This hemoglobin variant was found to be present in the Turkish population with a gene frequency of 0.1 in 745 β-thalassemia chromosomes (Tadmouri *et al.*, 1998).

In this thesis, HbD LosAngeles was observed in a family from Muş in Eastern Anatolia. The pedigree of the family AN. is shown in Figure 5.22 in which the proband is indicated with an arrow.

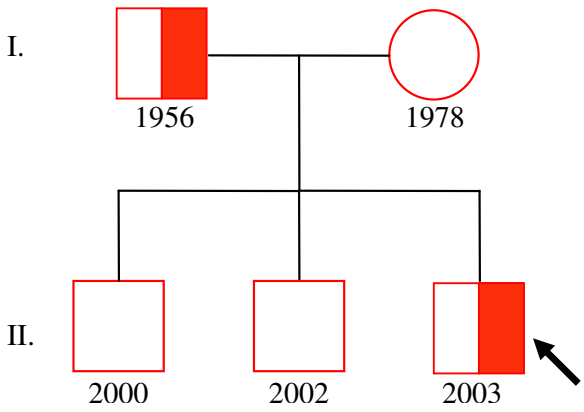


Figure 5.22. Pedigree of family AN

Sequencing of the β-globin gene revealed the G to C substitution located at Codon 121, leading to replacement of glutamic acid with glutamine (Figure 5.23).

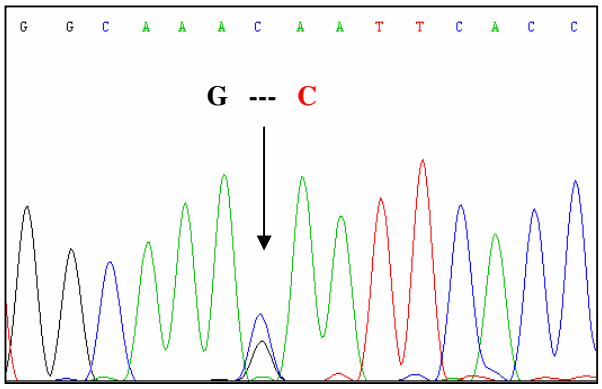


Figure 5.23. Sequencing results of the proband

5.2.4. Restriction Endonuclease Analysis to Confirm Sequencing Results

The HbD LosAngeles and HbO Arab variants are produced by the mutations affecting Codon 121 of the β-chain. Both of the mutations abolish the recognition site of the EcoRI enzyme. In both of cases, the mutant alleles were expected not to be recognized by EcoRI and remained undigested (566 bp), while the wild type alleles were digested into two fragments of 297 and 269 bp length (Figure 5.24).

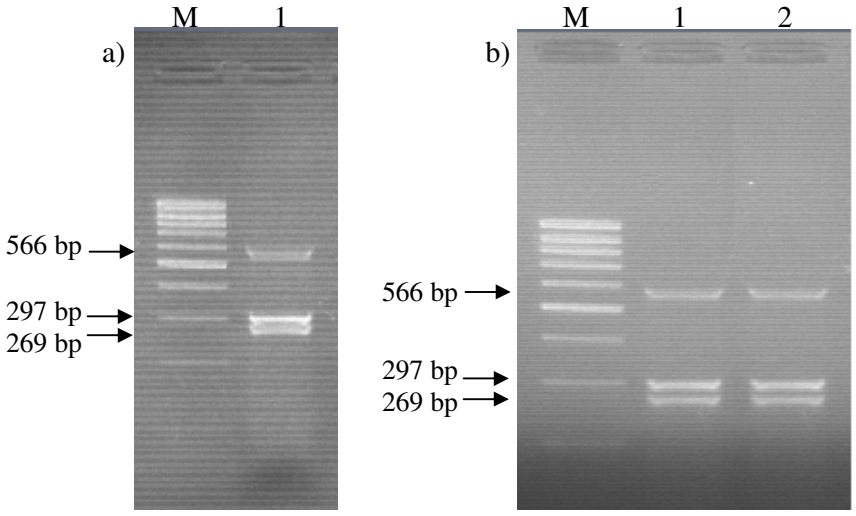


Figure 5.24. Restriction endonuclease analysis results for (a) patient A.K. and (b) family AN., lane 1: proband, lane 2: his father; M: 100 bp DNA ladder

5.3. VNTR Analysis Results

VNTR analysis was performed to eliminate the risk of misdiagnosis due to maternal contamination. (Figure 5.25). The VNTR probes used were the X-chromosome specific probe ST-14 and the X- and Y-chromosomes specific XY-probe.

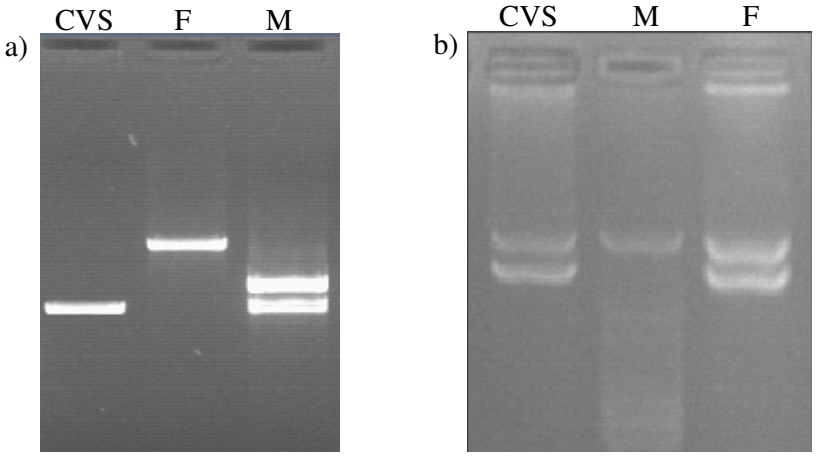


Figure 5.25. VNTR analysis results of (a) St14 and (b) XY probe results on two per cent agarose gel, CVS: chorionic villus sample, F: father, M: mother

VNTR analyses were applied to all prenatal diagnosis cases, but specific attention was given to the samples carrying the same genotype with their mother (like Family Ya.) (Table 5.1., second lane).

Table 5.1. Prenatal diagnoses results

Subjects	**Parental Genotype (Mother/Father)**		**Fetal Genotype**
Family Öz.	Cd39 (CAG-TAG)	IVS-I-110 (G-A)	Cd39 (CAG-TAG) / IVS-I-110 (G-A)
Family Ya.	IVS-I-6 (T-C)	IVS-I-6 (T-C)	IVS-I-6 (T-C) / N
Family De.	Cd15 (TGG-TGA)	IVS-II-1 (G-A)	IVS-II-1 (G-A) / N
Family Tı.	IVS-II-1 (G-A)	-30 (T-A)	-30 (T-A) / N

5.4. Distribution of Turkish β-Thalassemia Mutations

In the framework of this study, 163 individuals (not preselected and unrelated) were screened for the β-globin gene mutations. The number of β-thalassemia alleles was 193; 89 alleles derived from β-thalassemia minor patients, and 104 were from 52 β-thalassemia major patients. We were able to define all β-thalassemia alleles in this thesis without any unsolved cases by using β-Globin StripA^{ssay} in combination with genomic sequencing and in total 24 mutations were described. The frequency of these alleles is illustrated in Table 5.2.

Table 5.2. Frequency of Turkish β-globin gene mutations found in this study (Appendix A)

No	Mutation	No. Chr.	Frequency (%)
1.	IVS-I-110 (G-A)	41	31,06
2.	IVS-II-1 (G-A)	15	11,36
3.	FSC8 (-AA)	10	7,5
4.	IVS-I-1 (G-A)	9	6,81
5.	Cd39 (CAG-TAG)	8	6,06
6.	IVS-I-6 (T-C)	8	6,06
7.	IVS-II-745 (C-G)	7	5,30
8.	-30 (T-A)	7	5,30
9.	HbS	7	5,30
10.	Cd5 (-CT)	3	2,27
11.	Cd15 (TGG-TGA)	2	1,51
12.	Cd37 (TGG-TGA)	2	1,51
13.	Cd6 (-A)	1	0,75
14.	Cd22-24 (7 bp del)	1	0,75
15.	IVS-II-848 (C-A)	1	0,75
16.	Hb Ankara (Cd10 GCC-GAC)	1	0,75
17.	FSC8/9 (+G)	1	0,75
18.	-28 (A-C)	1	0,75
19.	Cd30 (G-C)	1	0,75
20.	IVS-I-5 (G-C)	1	0,75
21.	HbO Arab (Cd121 GAA-AAA)	1	0,75
22.	IVS-II-2 (T-A)	1	0,75
23.	HbD LosAngeles (Cd121 GAA-CAA)	1	0,75
24.	Cd44 (-C)	1	0,75
	Undefined	-	-
	Total	132	100

5.5. Molecular Analysis of $^{G}\gamma$-Globin Gene Promoter Polymorphisms

PdmI digestion and band patterns demonstrating variations at the -158 (C/T) polymorphism are shown in Figure 5.26 (b). All of the 14 analyzed β-thalassemia chromosomes with IVS-II-1 (G-A) mutation were shown to carry the C-T polymorphism at -158 of the $^{G}\gamma$-globin gene promoter (Figure 5.26). The PCR products of three patients are shown in Figure 5.26 (a), whereas in Figure 5.26 (b) digestion of these samples are demonstrated.

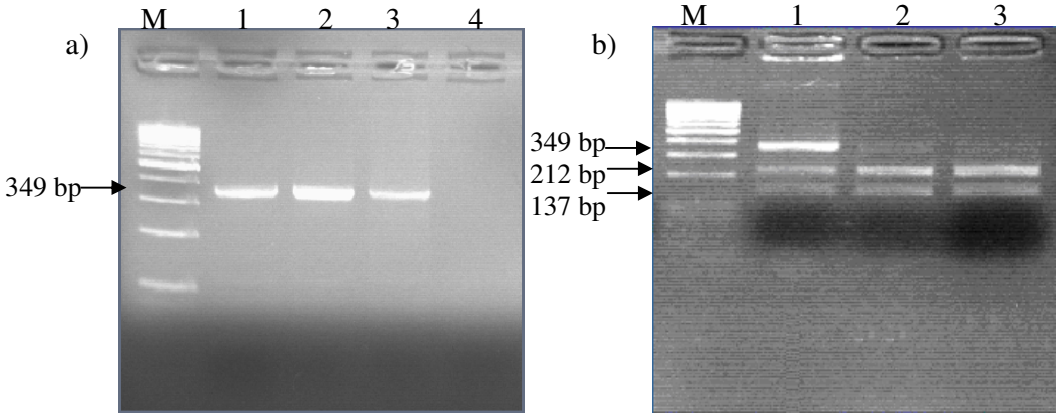


Figure 5.26. Agarose gel electrophoresis of three different patients: lanes 1 and 2: IVS-II-1 heterozygotes, lane 3: IVS-II-1 homozygote, lane 4: negative control and M: 100 bp DNA ladder

5.6. Construction of Databases by using FileMaker Pro 6 Software

5.6.1. Patients Database Results

The databases of patients and referring clinicians were constructed according to the instructions given in Section 4.8 and FileMaker User's Guide. The general structure of the patient database is shown in Figure 5.27. In this database, all the web sites required to complete the raw patient data, were designed to be accessed only with a mouse click. The examples of facilities designed will be mentioned in detail below. In patients database, there are three important layouts: personal information, background and molecular

analysis, and sample information layout. These were all designed according to the raw data accumulated in our laboratory since 2001.

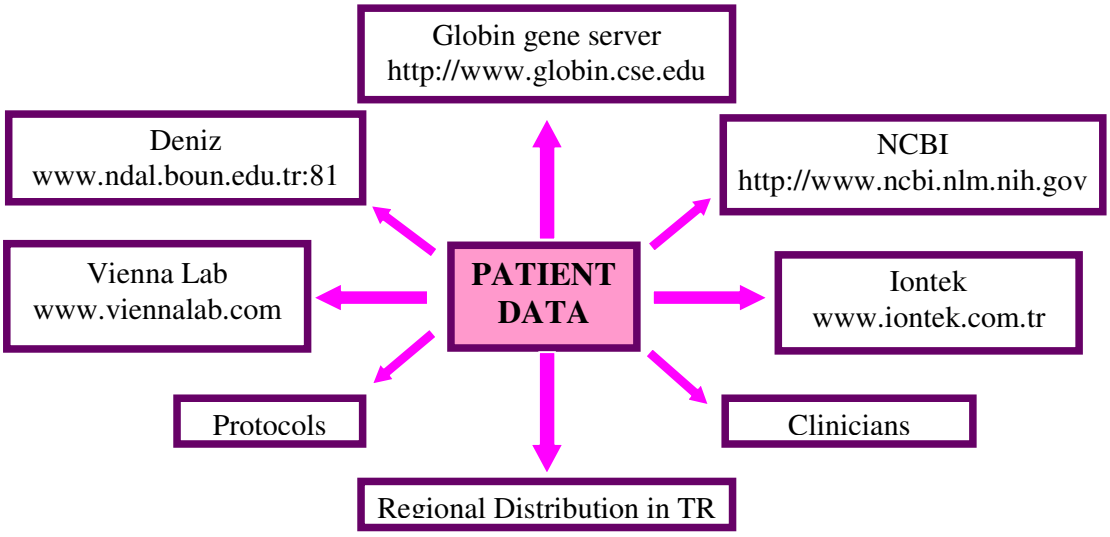


Figure 5.27. Structure of patients database

Personal information layout was designed in order to hold the personal data and contacts of individuals. The fields involved are illustrated in Figure 5.28. All the fields were designed for searching specific type of data efficiently, e.g. the phone numbers. In order to simplify data entry, some features of the fields were improved. For example, when adding a new record, the identity number (ID) automatically adjusts to the new entry.

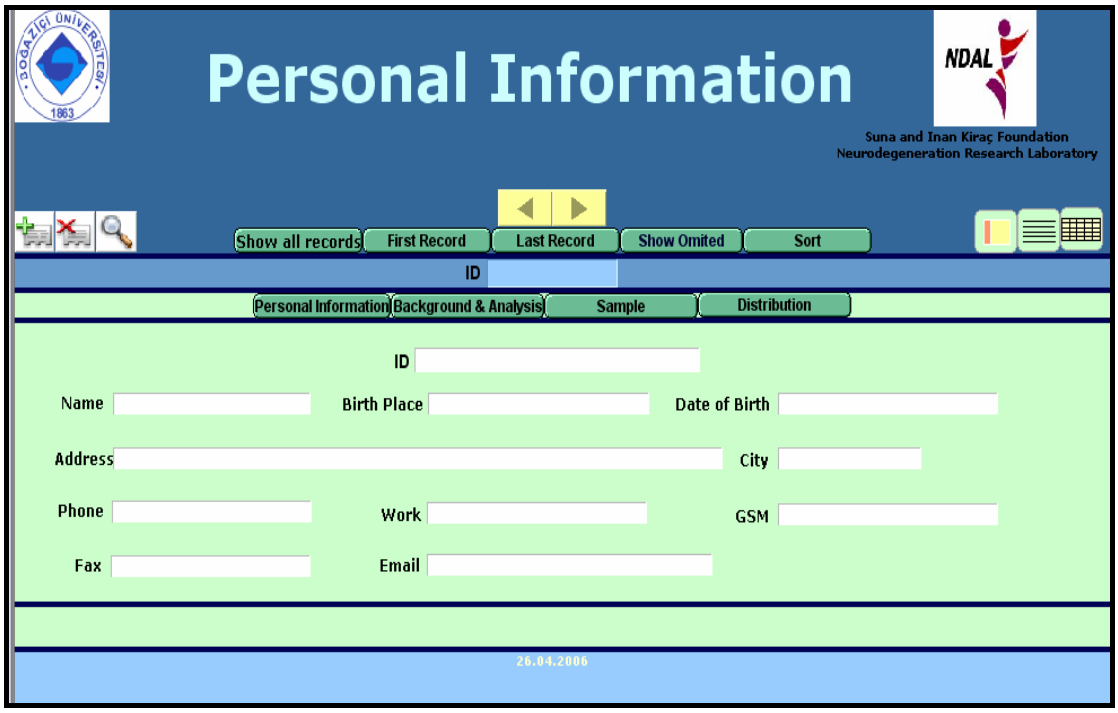


Figure 5.28. Personal information layout

Background and molecular analysis layout was designed in order to provide easy access to all information or data types that could be required while working in the laboratory. For example, protocol layouts were designed and defined with a button at the bottom of the layout (Appendix B). In addition, some links were given to specific web sites, like Globin Gene Server (mentioned in Section 1.10) and especially, mutation dependent links were designed to the online version of Huisman Syllabi and HbVar Database. Iontek Ltd. and "referring clinician" buttons were also defined within this layout, which opens the Iontek Ltd. home page and the related records in referring clinicians database (dependent on the entry in clinicians field), respectively. In this layout; results, allele1, allele2, region, origin, clinicians, consanguinity, and the method fields are designed in index format (Figure 5.29).

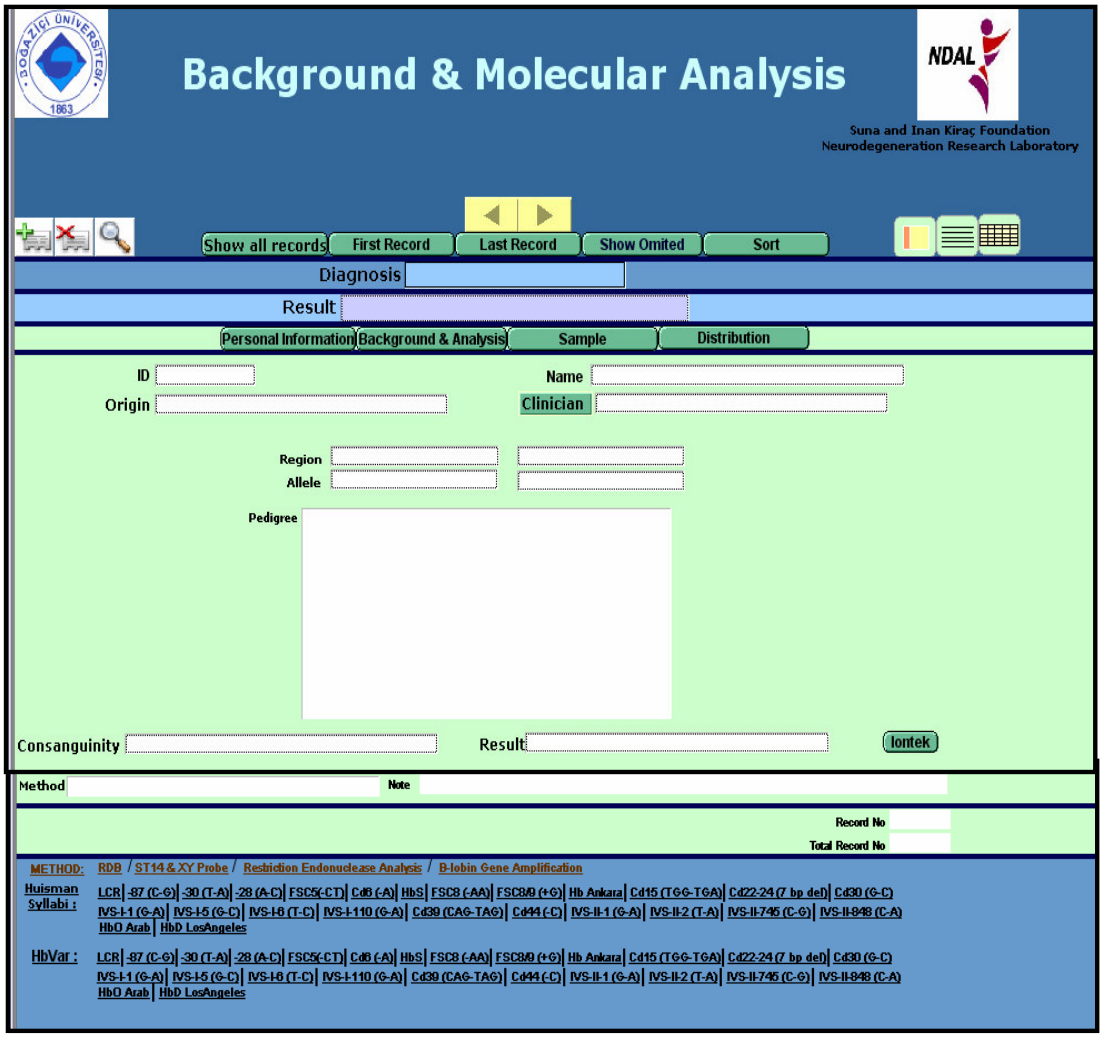


Figure 5.29. Background and molecular analysis layout

Sample information layout was designed to store the patients sample data, like date of arrival, sample type, analysis date and report data. In this layout, referring clinicians field was designed as a button like background and analysis layout (Figure 5.30).

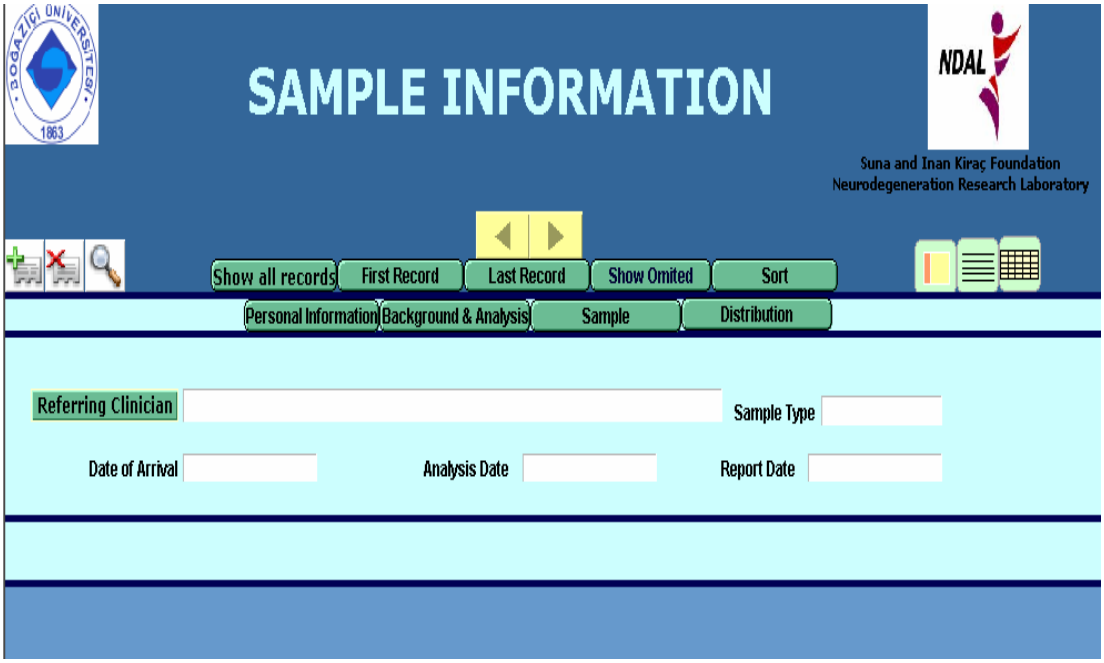


Figure 5.30. Sample information layout

The distribution layout was composed of calculation and summary fields. The calculation fields were designed with specific formulas which were written manually. The formulas were designed to calculate the allele frequencies of each mutation in chromosome number and in terms of percentage. In addition to the overall calculation of alleles, formulas required to calculate the regional distribution of mutations were designed for six geographical regions of Turkey, Immigrant Turks and for patients with unknown origins. Besides, the worldwide distribution of β-thalassemia mutations was made available with the "worldwide distribution" button (Appendix B), linked to Deniz Database (Gülen, 2002).

5.6.2. Referring Clinicians Database Results

Referring clinician database was also constructed according to the instructions given in Section 4.8 and FileMaker User's Guide. This database was designed in a simpler structure than patients database, because this database was designed to complement patients data. Institution and clinician fields were designed in index form for simplifying data entry. Similar to the patients database, "referring clinician" button was used to provide one-to-many relationship with patients database. Additionally, the e-mail field was

designed in a format for sending e-mail with an automatically opened outlook window. The referring clinician database layout is illustrated in Figure 5.31.

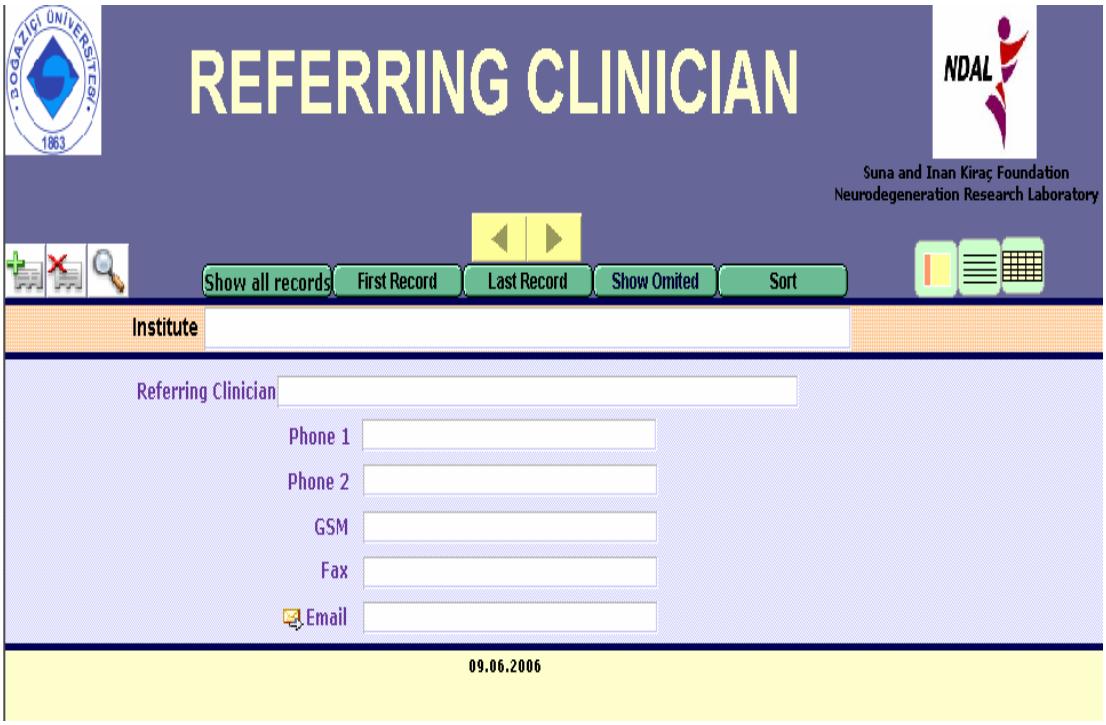


Figure 5.31. Referring clinicians database layout

6. DISCUSSION

6.1. Molecular Analysis of the β-Globin Gene

In this thesis, screening for hemoglobinopathies in the Turkish population was performed by using the β-Globin StripA^{ssay} and genomic sequencing.

The results obtained by screening 193 alleles in this study, demonstrate that 93 per cent of the Turkish β-thalassemia alleles could be identified by using the β-Globin StripA^{ssay}. This confirms the effectiveness of the β-Globin StripA^{ssay} in screening hemoglobinopathies. In addition, analysis of 22 mutations in a single reaction and in only five hours is an additional favorable feature of the system. The remaining seven per cent of the β-globin chromosomes in the framework of this thesis were identified by genomic sequencing. Our results are in accordance with the previous data obtained in our laboratory (Bilenoğlu, 2002). Beta-Globin StripA^{ssay} combined with genomic sequencing is a very efficient method for the diagnosis of hemoglobinopathies in the Turkish population and it is fast and convenient with a high diagnostic reliability.

In this study, seven prenatal diagnoses were performed by using the β-Globin StripA^{ssay}. The results identified with this method were confirmed by using VNTR analysis which detects possible maternal contamination. Although there is a great molecular heterogeneity in Turkish β-thalassemia chromosomes, the results showed that prenatal diagnosis is feasible with current advents in PCR-based technologies and improved methods of early fetal sampling.

6.2. Distribution of β-Thalassemia Mutations

The Deniz Database hosted at Boğaziçi University, Molecular Biology and Genetics Department, Neurodegeneration Research Laboratory (NDAL), indicates molecular heterogeneity of β-thalassemia in the best way. In this database, detailed statistical analysis results of 36,042 β-thalassemia chromosomes displaying frequency of mutations in population/country/continent specific manner are involved. The analysis results of Deniz demonstrate that the Iranian, Japanese and Turkish populations are the most heterogeneous worldwide, regarding β-thalassemia. According to the data involved in the Deniz Database (1398 chromosomes), 32 mutations define the genotype of β-thalassemia in Turkey.

Among these, the most frequent mutations and their relevant frequencies are given in Table 6.1.

Table 6.1. Most frequent β-thalassemia mutations in Turkey (http://www.ndal.boun.edu.tr:81/Deniz.html)

No	Mutation	Per cent
1.	IVS-I-110 (G-A)	45,9
2.	IVS-I-6 (T-C)	8,44
3.	IVS-I-1 (G-A)	6,15
4.	FSC8 (-AA)	5,79
5.	Cd 39 (CAG-TAG)	4,72
6.	IVS-II-1 (G-A)	4,36
7.	IVS-II-745 (C-G)	4,22
8.	-30 (T-A)	3,72
	TOTAL	83,3

In Mediterranean populations generally, six to eight mutations in a population define 90 to 95 per cent of β-thalassemia chromosomes (http://www.ndal.boun.edu.tr:81/Deniz. html). However, the most frequent eight mutations listed in Table 6.1. showed that the most common eight mutations account for only 83.3 per cent of β-thalassemia chromosomes in Turkey.

In the framework of this thesis, 193 β-thalassemia alleles were identified carrying 24 different mutations. 23 of these mutations were previously defined, and a novel IVS-II-2 (T-A) mutation was described in this study. The most common eight mutations (Table 6.2.) are in good accordance with the mutations shown in Table 6.1. However, the distribution of mutations were found to be slightly different when compared to previous studies.

Table 6.2. The first eight β-thalassemia mutations defined in this study

No	Mutation	Per cent
1.	IVS-I-110 (G-A)	31,06

2.	IVS-II-1 (G-A)	11,36
3.	FSC8 (-AA)	7,5
4.	IVS-I-1 (G-A)	6,81
5.	IVS-I-6 (T-C)	6,06
6.	Cd39 (CAG-TAG)	6,06
7.	IVS-II-745 (C-G)	5,30
8.	-30 (T-A)	5,30
	TOTAL	79,45

The molecular heterogeneity of Turkey is also reflected in the frequency and diversity of mutations in the six geographical regions of Turkey (Figure 6.1); this can be explained by the country's ancient and rich history, as well as its unique geographical location between three continents, resulting in a high ethnic diversity. According to the results of this thesis, the frequency of the IVS-I-110 (G-A) mutation is 34 per cent in the Marmara Region, rises to 42 percent in the Eagean and Mediterranean Regions, and drops to 17 per cent in the Black Sea Region.

The gradient of increasing molecular heterogeneity from West to East Anatolia is reflected in the number of mutations shown as "others" in Figure 6.1. The percentage of these mutations is around 10 per cent in the Eagean and Mediterranean Regions, but goes up to 32 per cent in Central Anatolia. The decreased variety of mutations observed in South Eastern Anatolia, as opposed to Tadmouri (1998), may be due to the number of chromosomes with unknown origins.

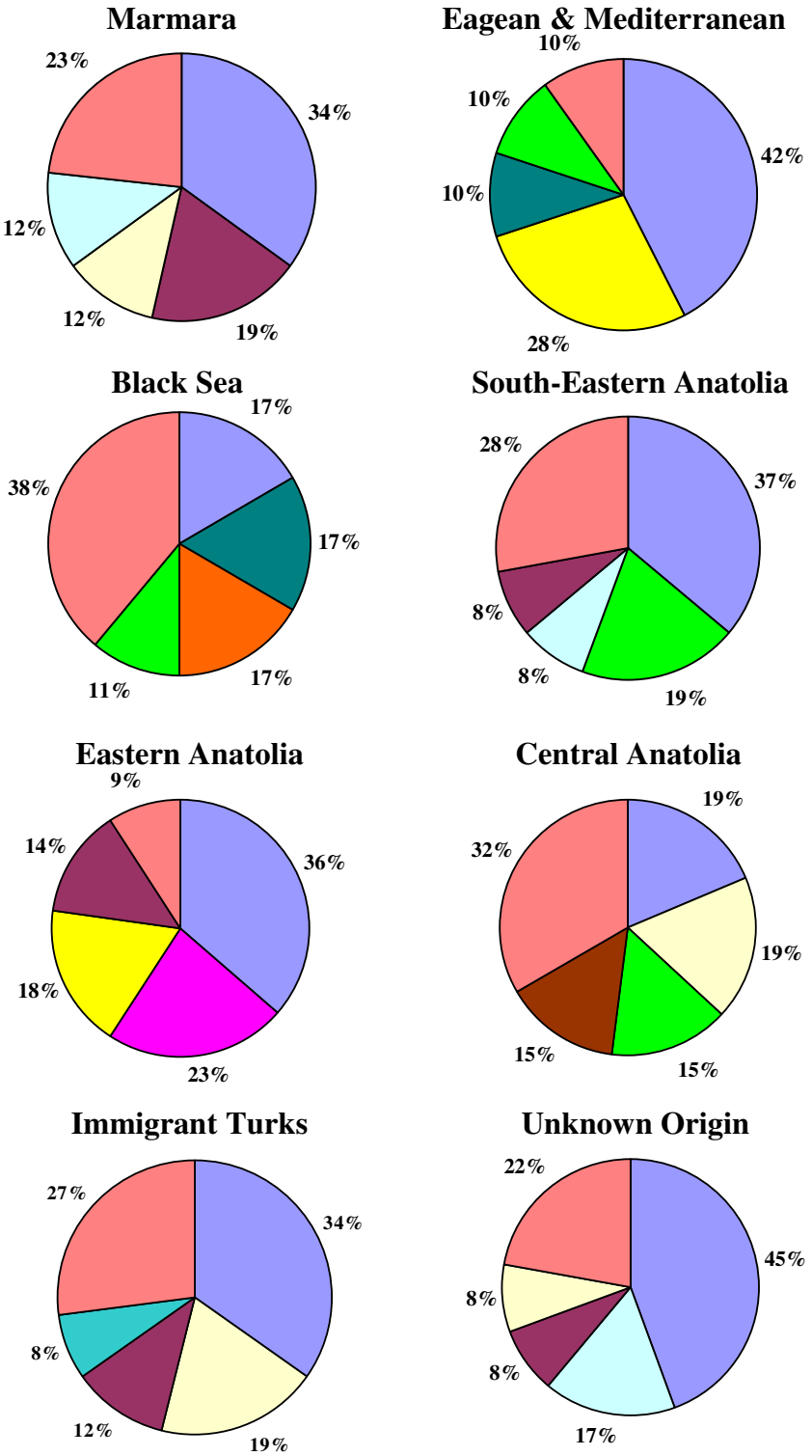


Figure 6.1. Regional distribution of the most common mutations in Turkey

6.3. Phenotypic Diversity among β-Thalassemia Chromosomes

Beta-thalassemia is mainly due to inherited molecular defects affecting the β-globin gene expression. Functionally defective β-globin alleles display two main phenotypes according to the nature of the mutation: in β^+-thalassemia, there is reduced synthesis of β-globin chains with variable degrees, and in β^0-thalassemia, there is no β-globin chain production. In compound heterozygosity, homozygosity and carrier states, the phenotype of individuals differ even in those carrying the same genotype. These modifications are due to primary and secondary modifiers (Thein, 2005b).

The primary modifier is the nature of the mutation which determines the amount of excess α-chains critical in the severity of anemia observed. In Table 6.3., there is a compilation of common β-thalassemia mutations described in this study and their relevant phenotypes. It is apparent that the nature of the mutation is defined by the level of the β-globin gene expression failure. In the heterozygous state, some β-thalassemia alleles (exon3 mutations) are inherited in a Mendelian dominant fashion, mainly because of the inclusion bodies formed in red blood cell precursors by truncated, but undetected mRNAs.

Secondary modifiers of the β-thalassemia phenotype are factors enhancing globin chain expression and thus favorably decreasing the number of excess α-chains. The -158 C to T polymorphism in the $^{G}\gamma$-globin gene promoter is an important factor, modifying the phenotype of β-thalassemia (Thein, 2005a). All of the alleles, analyzed in this thesis, carrying the IVS-II-2 (G-A) mutation are shown to carry this mutation investigated by the *PdmI* restriction enzyme. In previous studies, it was demonstrated that the two mutations, IVS-I-1 (G-A) and IVS-II-1 (G-A) affect β-globin gene expression by the same mechanism, however they did not show similar phenotypes, because of the presence of the -158 (C/T) polymorphism at the $^{G}\gamma$-globin gene promoter (Thein, 2005a). The mild β^0-thalassemia phenotype, observed in connection with our IVS-II-1 (G-A) mutation, depends on the polymorphism present in the $^{G}\gamma$-globin gene promoter, increasing the expression of γ-chains, which bind with excess α-chains to form HbF.

Table 6.3. The nature of some mutations identified in the framework of this thesis

Mutation	Molecular Mechanism	Phenotype
-28 (A-C)	decreased β-globin gene transcription because of reduced binding of erythroid factors	β^{+}-thalassemia
IVS-I-110 (G-A)	formation of a cryptic splice site	β^{+}-thalassemia
IVS-II-848 (C-A)	diminishes splicing	β^{+}-thalassemia
IVS-II-1 (G-A)	abolishes the 5'splicing site	mild β^{0}-thalassemia
IVS-I-1 (G-A)	prevents splicing completely	severe β^{0}-thalassemia
FSC8 (-AA)	translation termination at Codon 21	β^{0}-thalassemia
Cd39 (CAG-TAG)	terminates translation at Codon 39 by a non sense mutation	β^{0}-thalassemia

6.4. FileMaker Pro 6 Software

Powerful features, broad platform support, and easy-to-use interface make FileMaker Pro 6 an essential tool for creating and sharing databases. Importing the data to be used in the FileMaker database is as easy as simple mouse clicks. In other words, importing a folder of files to FileMaker will take only few seconds with the support for standards, such as ODBC, JDBC, and XML.

There are several alternatives to be used in database construction; however some of them require very detailed knowledge. They do not offer user-friendly interfaces, and some of them may not be suitable for all functions. When FileMaker is compared with Access, advantages of FileMaker are easily seen over Access (Table 6.4). In addition to Windows, FileMaker supports Mac OS, Linux, and Palm OS platforms. FileMaker is designed to be easy enough to create and share databases. Access on the other hand, includes very powerful tools, but requires a great deal of prior knowledge and complex SQL programming. Differences in many other aspects are also observed between FileMaker and Microsoft Access (Table 6.4.) (Decorte, 2001).

Table 6.4. Brief comparison of FileMaker and Microsoft Access

Feature	ACCESS	FILEMAKER
File size	One file cannot exceed 2 GB, however because a database can include linked tables in other files, its total size is limited only by available storage capacity	2 GB per file. Here too, the only real limitation is disc space
Number of concurrent open files (from the users perspective)	One visible. Access can use info from linked tables by leading them in memory	Max 50 at a time; depends heavily on working with FileMakerServer (FMS),but also can run without it.
Number of concurrent users	255	50 database or 25 guests with trade-off for FM standard. 125 files to 250 guests with FMS
Application size	160 MB	26 MB
Number of characters in a field name	64	60
Number of fields in a table	255	Unlimited
Number of characters in a record	2000	Unlimited
Text	Up to 255 characters	Up to approximately 64,000 letters, symbols, and numbers used as text
Number	Numeric data to be used for mathematical calculations, except calculations involving money (use currency type). 1,2,4, or 8 bytes. 16 bytes for Replication ID only. 1 byte= 256 value (0-255) 2 bytes, 65.535 values 4 bytes, 4 billion + values	Up to 255 numbers, letters, or symbols. In most cases, only the numbers are treated as numeric; number fields can not contain carriage returns. Only 15 significant digits will be used in calculations
Calculation	Non existent as field type	The result of a formula that uses field values from the related records; the result can be one of these data types: text, number, date, time, or container.

6.4.1. Patients Database

The most important drawback when starting this project was, not having the electronic copies of data files of 254 records; for this reason all data of 254 individuals were fed manually to Excel sheets. After planning the database, required data were imported from the Excel sheets to FileMaker Pro 6. Modifications of each layout were performed by creating buttons or scripts for specific tasks. In each layout, there are some common buttons. Layout buttons were provided for jumping from one layout to the other. In addition "first record" and "last record" buttons provide jumping from the first record to the last. This option is also useful and provides speed while working with a database. It is possible to sort the "found set" by using the "sort" button. This button is created for example for finding out the β-thalassemia major individuals from previous years. The "show all" button is for returning to the unsorted record set.

6.4.1.1. Personal Information Layout. Personal Information layout was designed to hold personal information of the patients. Here, the ID field is designed as a serial-auto entry field, in order to provide serial addition of ID numbers while updating the database. This is an important option which lowers the mistakes, common when working in the laboratory. Birth place, date of birth and city fields were specially designed with index options in order to show all of the information fed previously, while entering data. It also provides selection of the desired data from the list, in order to fill the record. The index option is important and is used in modification of databases, because the FileMaker is sensitive to typing errors.

6.4.1.2. Background and Molecular Analysis Layout. This layout is solely created for holding the scientific/laboratory information about an individual. The indexed fields are origin, clinician, region, allele, result and consanguinity fields. Here the aim is to provide easy data entry without any typing mistakes. Especially in Clinicians field, it is important to enter the data in its previously entered form because relationships are completely dependent on the data style, e.g. typing a wrong letter will result in relationship failure. Different than other fields, the index of clinicians is fed from the related database: referring clinicians database. This type of data feeding was preferred in order to have an idea about the last updated version of the referring clinician database. Related data about the related

clinician can be reached by clicking on the clinician button in a single mouse click only, without any concerns about hard copy files. In order to reach protocols used in the framework of this thesis, it is enough to click the buttons found at the bottom of the page. Here also, the mutation information involved in HbVar Database and electronic (online) version of Huisman's Syllabi can be reached with single mouse clicks preventing time wasting for searching specific type of information even in hard copy files or internet files.

6.4.1.3. Sample Information Layout. Clinicians database is designed in the same manner with other layouts. Here, date of sample arrival is designed as auto date entry format; by this way while updating the data for new samples, there is no need for writing data or for checking the data format previously entered.

6.4.1.4. Distribution Layout. The distribution layout is designed to analyze the set of data involved within the database about β-thalassemia patients. In order to construct this database, more than 800 fields were defined. It took more than 10 days to define all the fields required to analyze chromosomes, according to their origins and β-globin genotypes. The formulas were designed to calculate alleles/chromosomes independently rather than individuals. For contrasting the mutations, present in the Turkish population, the worldwide distribution button is defined. This button is directly linked to the Deniz Database involving the worldwide and regional distribution of β-thalassemia mutations.

6.4.2. Referring Clinician Database

In referring clinician database there is only one layout composed of seven fields. The layout involves the common buttons described previously: "sort", "show omitted", "last record", "first record" and "show all records". The fields, indexed in this database, are institution and clinician fields. The index of these fields is fed from referring clinicians database. All of the records in this database were related to the records in patients database in one-to-many relationship manner. In this database the "email" button is able to send emails to the relevant clinician by opening the outlook window automatically. The aim of this option is not to deal with various processes and to reach the required facilities with only a few mouse clicks.

Instant web publishing option and easy design of CDML pages (to overwhelm submit record problem) make FileMaker much easier to construct a database than any other program. The most important aspect is that all database studies were performed without any need for IT expert help. This makes FileMaker Pro an invaluable database construction program in wet laboratories, dealing with raw data which need to be analyzed.

7. CONCLUSIONS

7.1. NDAL Web Page and Construction of Additional Databases

Neurodegenerative disorders and cardiovascular diseases are additional and intense research interests of NDAL. The hemoglobinopathy database served as a model for seven new databases, which were created in the framework of this thesis: Alzheimer's Disease, Parkinson's Disease, Spinocerebellar Ataxias, Friedreich's Ataxia, Huntington's Disease, Amyotrophic Lateral Sclerosis, and Cardiovascular Diseases.

These databases are shared among all persons (students, technicians, etc.) who work on related projects in our laboratory via LAN. However, they are also accessible via the World Wide Web, only for the very individuals who deal with a certain disease, by hidden links given in the NDAL Web page. The data accumulated in these databases involve all sorts of experiences collected on the above diseases in the Turkish population. In future, the data obtained and analyzed will be shared online with other laboratories in Turkey, for combining and increasing the knowledge about these disorders. Furthermore, databases which are accessible with passwords, only via the World Wide Web, restricted to researchers, will be designed to enable online-sharing. Care will be taken about ethical regulations e.g. not displaying patients name, personal contacts, etc.

The NDAL web page (www.ndal.boun.edu.tr) provides a platform for this sharing processes via the most important characteristics of computers in bioinformatics: communication. In future, the NDAL Web page and databases will be designed in combination to serve as a communication and search and retrieval system.

APPENDIX A: MOLECULAR DIAGNOSIS RESULTS

Table A.1. Molecular diagnosis results in the year 2002

(Analyses were performed by other researchers in our laboratory)

2002			
Family No	**Subjects**	**Result**	
1	F.A.	IVS-I-1 (G-A)	IVS-I-1 (G-A)
2	Z.Y	FSC8 (-AA)	FSC8 (-AA)
3	Fetüs Du.	Cd39 (CAG-TAG)	IVS-II-745 (C-G)
4	E.G.	FSC8 (-AA)	FSC8 (-AA)
	A.G.	FSC8 (-AA)	FSC8 (-AA)
5	Tn.B.	FSC8 (-AA)	N
	Te.B.	FSC8 (-AA)	N
6	A.E.	-30 (T-A)	-30 (T-A)
7	B.A.	Cd22-24 (7 bp del)	N
	S.A.	Cd22-24 (7 bp del)	N
	R.A.	N	N
8	C.B.R.	IVS-I-110 (G-A)	IVS-I-6 (T-C)
	E.R.	IVS-I-110 (G-A)	N
	B.R.	IVS-I-6 (T-C)	N
9	M.Y.G.	IVS-I-110 (G-A)	N
10	A.D.	FSC8 (-AA)	N
11	G.P.	Cd6 (-A)	IVS-I-110 (G-A)
	U.P.	Cd6 (-A)	IVS-I-110 (G-A)
12	F.A.	IVS-I-110 (G-A)	IVS-I-110 (G-A)
13	D.K.	IVS-I-110 (G-A)	IVS-I-110 (G-A)
14	S.A.	IVS-I-110 (G-A)	IVS-I-110 (G-A)
15	M.A.	IVS-II-745 (C-G)	IVS-II-745 (C-G)
16	İ.B.	IVS-I-110 (G-A)	IVS-I-110 (G-A)
17	A.O.	Cd39 (CAG-TAG)	Cd39 (CAG-TAG)
18	T.G.	IVS-I-1 (G-A)	IVS-I-1 (G-A)
19	N.M.E.	IVS-I-110 (G-A)	IVS-I-110 (G-A)
20	E.B.	Cd39 (CAG-TAG)	N
	B.B.	Cd39 (CAG-TAG)	N
	N.B.	Cd39 (CAG-TAG)	N
21	A.U.	FSC8 (-AA)	N
	E.U.	FSC8 (-AA)	N
22	R.K.	other than β-globin gene	other than β-globin gene
23	M.K.	IVS-I-110 (G-A)	N

Table A.2. Molecular diagnosis results in the year 2003

(Analyses were performed by other researchers in our laboratory)

2003			
Family No	**Subjects**	**Result**	
	CV Du.	Cd39 (CAG-TAG)	N
24	G.G.	IVS-I-6 (T-C)	IVS-II-1 (G-A)
	Ve.G.	IVS-I-6 (T-C)	N
	Vi.G.	IVS-II-1 (G-A)	N
25	N.K.	FSC8 (-AA)	N
26	A.G.	Cd5 (-CT)	IVS-II-1 (G-A)
27	S.M.İ.	IVS-I-110 (G-A)	N
	İ.M.Ç.	IVS-I-110 (G-A)	N
28	A.C.A.	other than β-globin gene	other than β-globin gene
	A.A.	other than β-globin gene	other than β-globin gene
	M.A.	other than β-globin gene	other than β-globin gene
29	Ö.A.	IVS-II-1 (G-A)	IVS-II-1 (G-A)
	S.A.	IVS-II-1 (G-A)	N
30	İ.G.	IVS-II-848 (C-A)	N
	B.G.	IVS-II-848 (C-A)	N
31	B.O.	Cd15 (TGG-TGA)	IVS-I-110 (G-A)
	Ö.O.	Cd15 (TGG-TGA)	N
	Ş.O.	IVS-I-110 (G-A)	N
32	D.Ö.	other than β-globin gene	other than β-globin gene
	Y.Ö.	other than β-globin gene	other than β-globin gene
	İ.Ö.	other than β-globin gene	other than β-globin gene
33	G.Y.	HbAnkara (Cd10 GCC-GAC)	N
34	F.K.Ö.	IVS-I-1 (G-A)	N
	S.Ö.	IVS-I-1 (G-A)	N
	M.Ö.	N	N
35	H.S.	IVS-I-110 (G-A)	IVS-I-110 (G-A)
	K.S.	IVS-I-110 (G-A)	N
	F.S.	IVS-I-110 (G-A)	N
36	M.T.	IVS-I-110 (G-A)	IVS-I-110 (G-A)
37	K.E.	-30 (T-A)	-30 (T-A)
	I.E.	-30 (T-A)	N
38	Y.M.	IVS-II-1 (G-A)	N
39	Em.P.	IVS-II-1 (G-A)	N
	En.P.	IVS-II-1 (G-A)	N
40	F.H.	HbS	N
	S.H.	HbS	N
	E.H.	HbS	HbS

Table A.3. Molecular diagnosis results in the year 2004

(Analyses performed in the framework of this thesis)

2004			
Family No	**Subjects**	**Result**	
41	O.S.	IVS-II-745 (C-G)	IVS-II-745 (C-G)
	E.S.	IVS-II-745 (C-G)	N
42	A.E.A.	IVS-II-1 (G-A)	N
	E.A.	IVS-II-1 (G-A)	N
43	CV De.	IVS-II-1 (G-A)	N
44	M.Y.	IVS-I-110 (G-A)	IVS-I-110 (G-A)
	G.Y.	IVS-I-110 (G-A)	N
	A.Y.	IVS-I-110 (G-A)	N
45	T.G.	FSC8/9 (+G)	IVS-II-745 (C-G)
	İ.G.	FSC8/9 (+G)	N
	A.G.	IVS-II-745 (C-G)	N
46	E.Ö.	IVS-I-110 (G-A)	N
	A.Ö.	IVS-I-110 (G-A)	N
47	U.B.	other than β-globin gene	other than β-globin gene
	M.B.	other than β-globin gene	other than β-globin gene
	A.İ.B.	other than β-globin gene	other than β-globin gene
48	CV T.	-30 (T-A)	N
	D.T.	-30 (T-A)	N
	S.T.	IVS-II-1 (G-A)	N
49	A.G.	IVS-I-110 (G-A)	N
50	B.K.G.	other than β-globin gene	other than β-globin gene
51	T.A.	IVS-I-110 (G-A)	IVS-I-110 (G-A)
	B.A.	IVS-I-110 (G-A)	N
	F.A.	IVS-I-110 (G-A)	N
52.	A.A.	HbD LosAngeles (Cd121 GAA-CAA)	N
	R.A.	HbD LosAngeles (Cd121 GAA-CAA)	N
	T.A.	N	N
53.	A.D.	IVS-II-1 (G-A)	Cd5 (-CT)
54.	F.A.	IVS-I-110 (G-A)	IVS-I-110 (G-A)
55.	H.A.	-28 (A-C)	Cd39 (CAG-TAG)
56.	A.C.E.	IVS-I-1 (G-A)	IVS-I-1 (G-A)
57.	E.T.	IVS-I-110 (G-A)	IVS-I-110 (G-A)
58.	Z.E.	IVS-I-110 (G-A)	Cd30 (AGG-ACG)
	E.E.	IVS-I-110 (G-A)	Cd30 (AGG-ACG)
59.	S.A.	IVS-I-110 (G-A)	LCR(?)
60.	G.A.	IVS-II-1 (G-A)	IVS-II-1 (G-A)

Table A.4. Molecular diagnosis results in the year 2005

(Analyses performed in the framework of this thesis)

2005			
Family No	**Subjects**	**Result**	
61.	H.A.	IVS-I-6 (T-C)	IVS-I-6 (T-C)
	K.A.A.	IVS-I-6 (T-C)	N
	G.A.	IVS-I-6 (T-C)	N
	A.O.A.	IVS-I-6 (T-C)	N
62.	CV Ö.	IVS-I-110 (G-A)	Cd39 (CAG-TAG)
	E.Ö.	Cd39 (CAG-TAG)	N
	N.Ö.	IVS-I-110 (G-A)	N
63.	F.M.	Cd5 (-CT)	Cd39 (CAG-TAG)
	G.M.	Cd39 (CAG-TAG)	N
64.	CV Y.	IVS-I-6 (T-C)	N
	Ş.Y	IVS-I-6 (T-C)	N
	M.N.Y.	IVS-I-6 (T-C)	N
65.	B.A.	IVS-I-1 (G-A)	N
	İ.A.	IVS-I-1 (G-A)	N
65	E.Ö.	HbS	N
	T.Ö.	HbS	N
66	A.F.S.	Cd15 (TGG-TGA)	N
67	M.M.	IVS-II-2 (T-A)	N
	A.M.	IVS-II-2 (T-A)	N
68	A.K.	HbOArab (Cd121 GAA-AAA)	N
69	G.Y.	other than β-globin gene	other than β-globin gene
70	A.E.	IVS-I-110 (G-A)	IVS-I-110 (G-A)
71	O.T.	HbS	HbS
	S.T.	HbS	N
	H.T.	HbS	N
72	S.A.	IVS-II-1 (G-A)	IVS-II-1 (G-A)
	N.A.	IVS-II-1 (G-A)	N
73	E.A.	other than β-globin gene	other than β-globin gene
74	M.Y.	other than β-globin gene	other than β-globin gene
	F.Y.	other than β-globin gene	other than β-globin gene
75	G.N.	IVS-II-1 (G-A)	IVS-II-1 (G-A)
	N.N.	IVS-II-1 (G-A)	N
76	N.T.	-30 (T-A)	-30 (T-A)
	G.T.	-30 (T-A)	N
	Sn.T.	N	N
	Si.T.	-30 (T-A)	N
	M.T.	N	N
77	B.K.	IVS-I-1 (G-A)	N
78	Z.E.A.	IVS-I-5 (G-C)	IVS-I-6 (T-C)
	Z.A.	IVS-I-6 (T-C)	N
	S.A.	IVS-I-5 (G-C)	N
79	S.S.	HbS	N
	F.S.	HbS	N
	M.S.	HbS	N
	J.S.	HbS	N
	M.S.	HbS	N
80	Y.Ç.	N	N

Table A.5. Molecular diagnosis results in the year 2006

(Analyses performed in the framework of this thesis)

2006			
Family No	**Subjects**	**Result**	
81	K.Ö.	Other than β-globin gene	Other than β-globin gene
82	E.A.U.	Other than β-globin gene	Other than β-globin gene
83	B.B.	Cd37 (TGG-TGA)	Cd37 (TGG-TGA)
84	E.S.	IVS-II-745 (C-G)	N
85	K.E.	IVS-I-6 (T-C)	IVS-I-6 (T-C)
86	F.E.	IVS-I-110 (G-A)	IVS-I-110 (G-A)
87	H.Ç.	IVS-I-110 (G-A)	IVS-I-110 (G-A)
88	M.M.K.	IVS-I-110 (G-A)	IVS-I-110 (G-A)
89	A.K.	Cd44 (-C)	IVS-I-110 (G-A)
	H.K.	IVS-I-110 (G-A)	N

APPENDIX B: PROTOCOL LAYOUTS DESIGNED FOR THE HEMOGLOBINOPATHY DATABASE

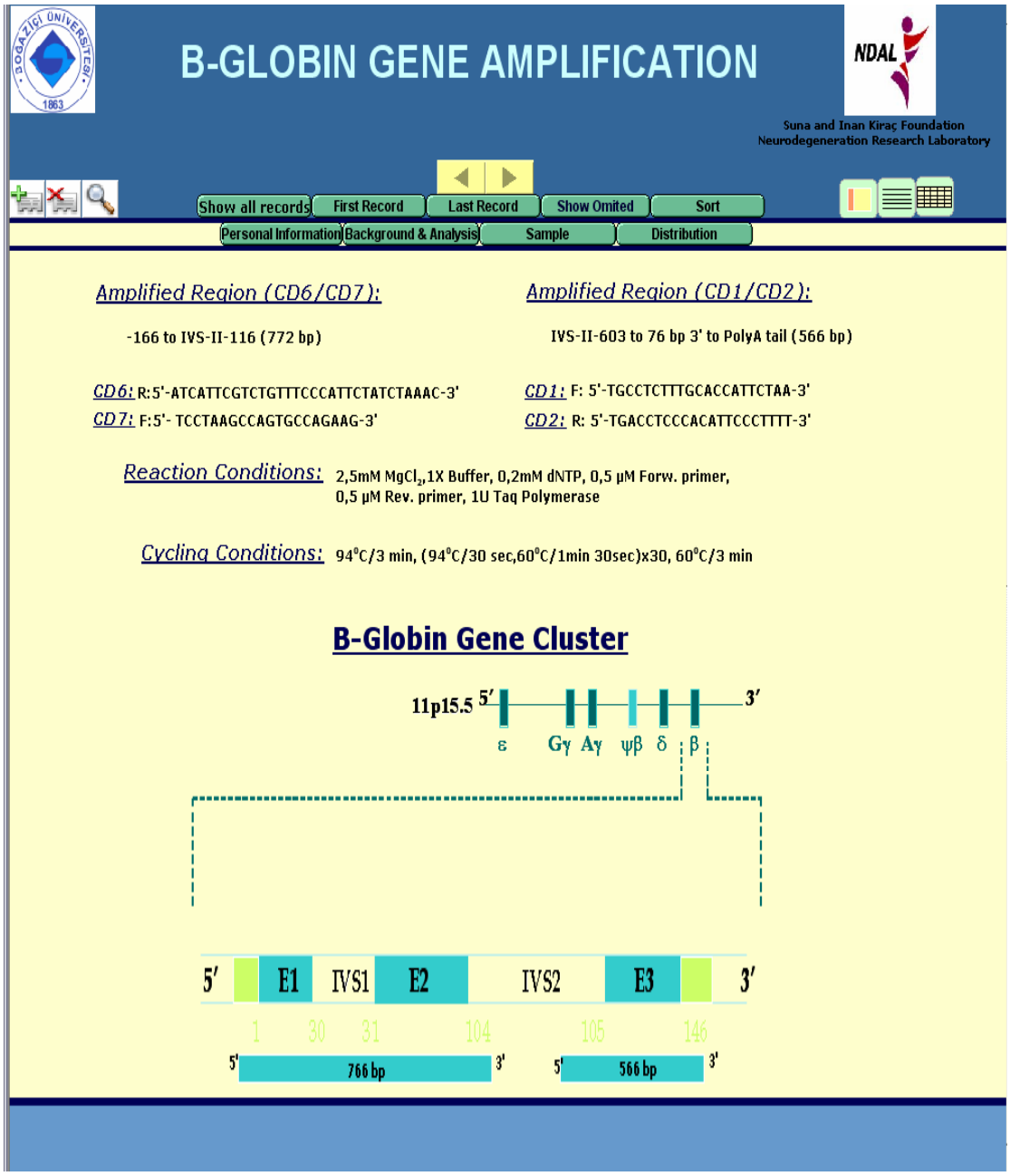


Figure B.1. β-globin gene amplification protocols

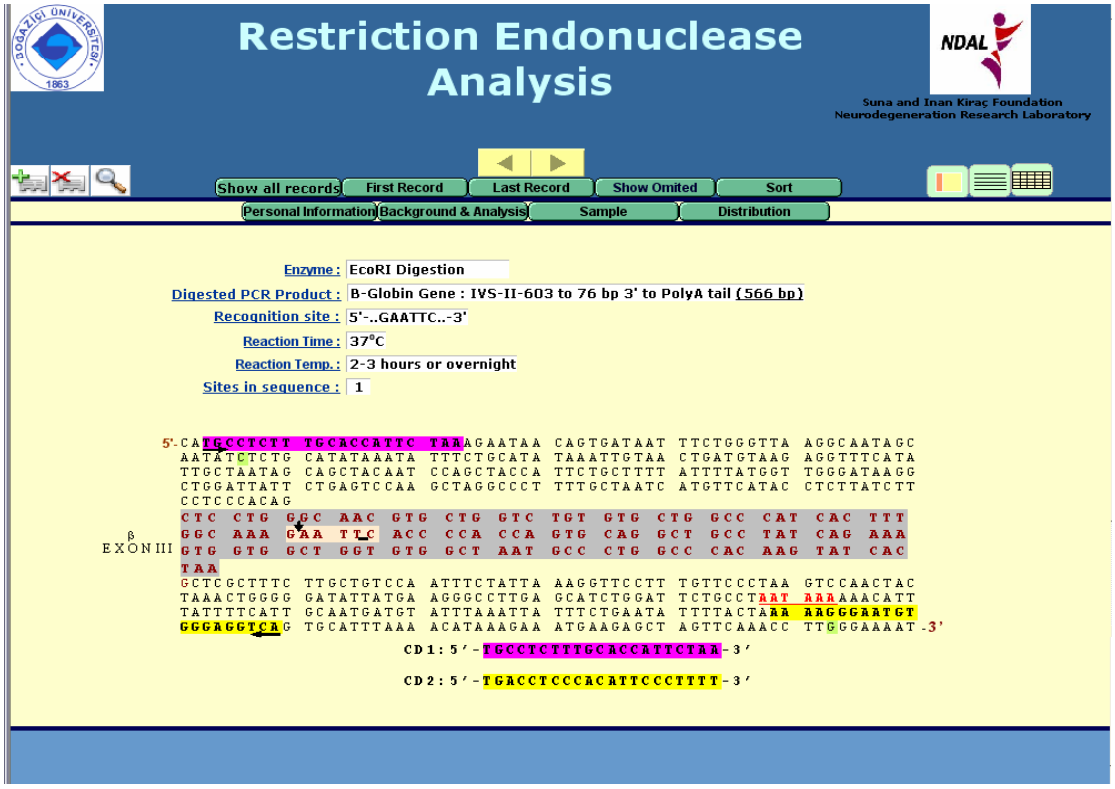


Figure B.2. Restriction endonuclease analysis protocol

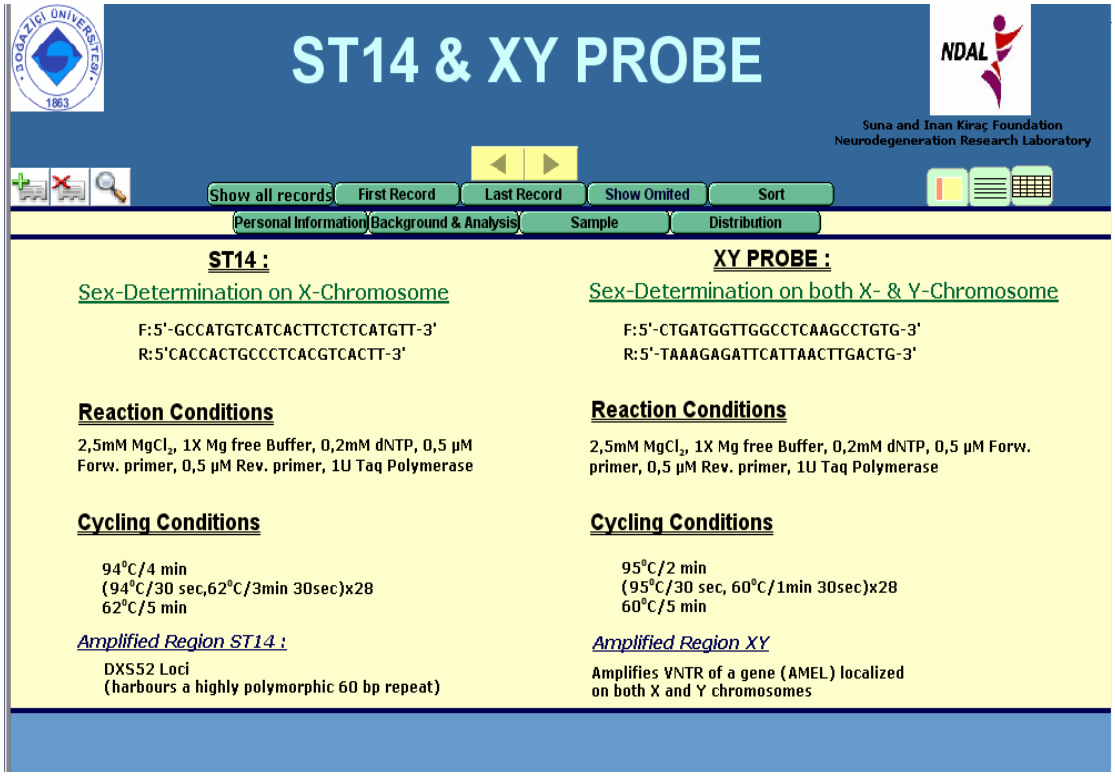
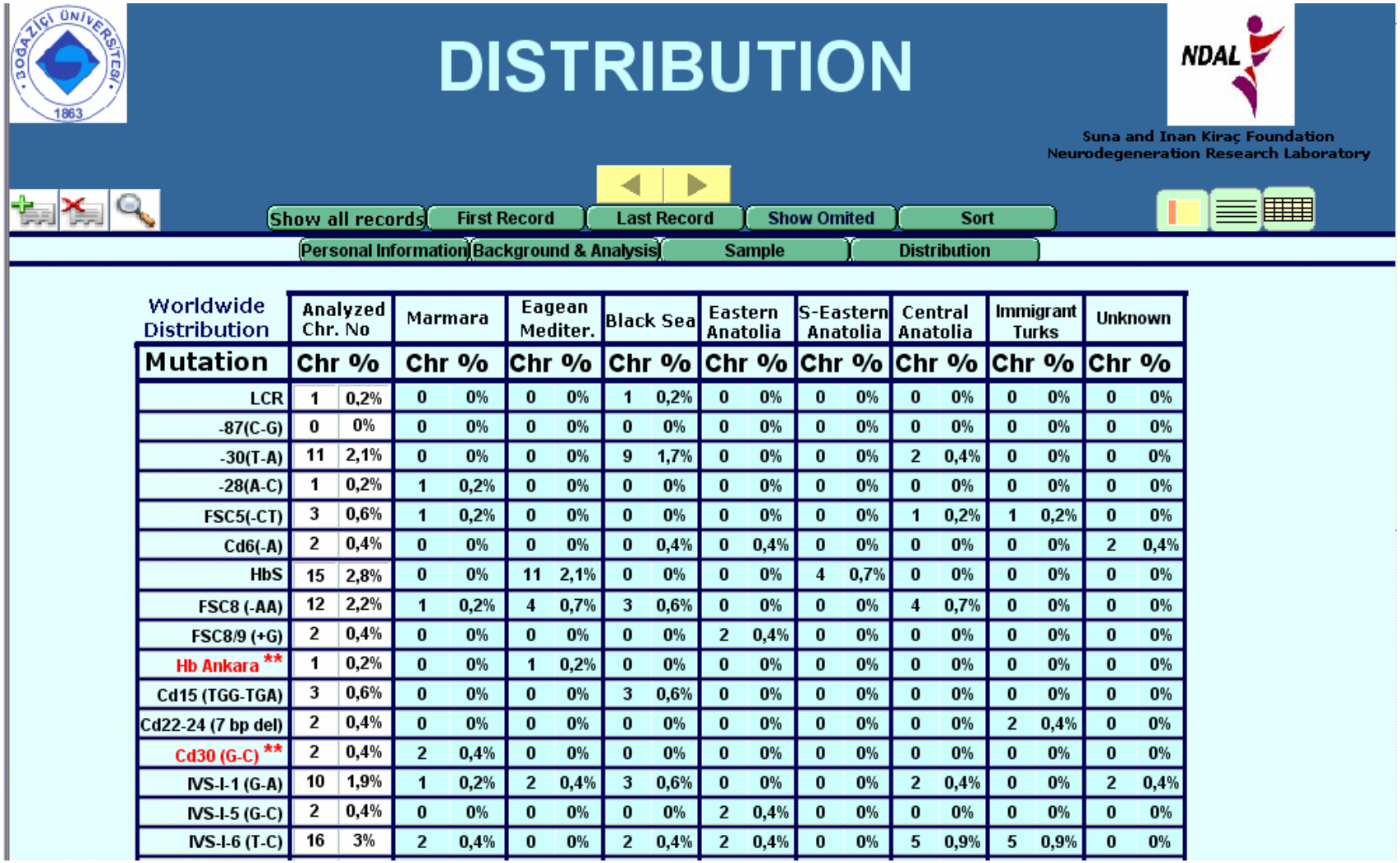
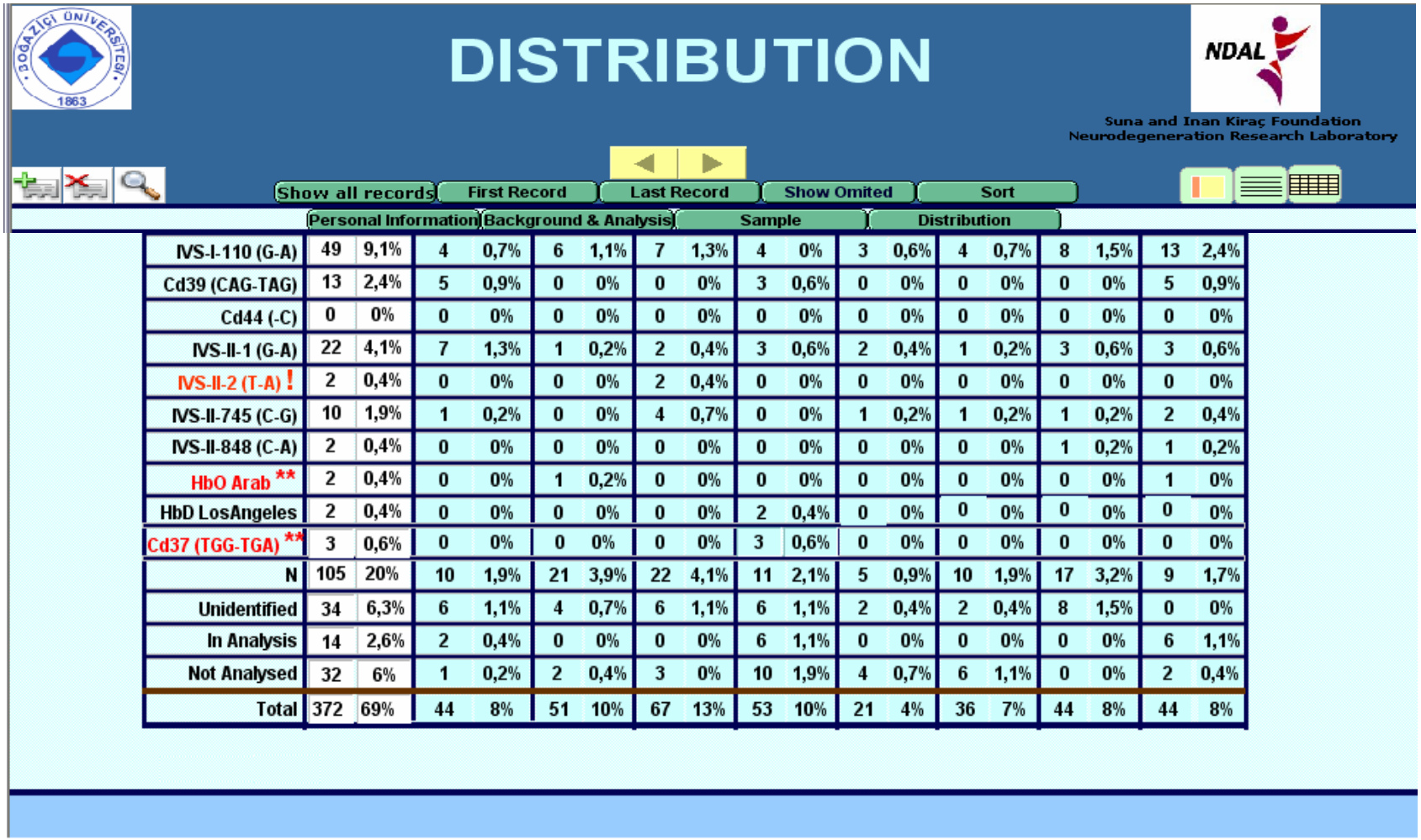


Figure B.3. VNTR analysis protocols



Worldwide Distribution	Analyzed Chr. No		Marmara		Eagean Mediter.		Black Sea		Eastern Anatolia		S-Eastern Anatolia		Central Anatolia		Immigrant Turks		Unknown	
Mutation	**Chr**	**%**	**Chr**	**%**	**Chr**	**%**	**Chr**	**%**	**Chr**	**%**	**Chr**	**%**	**Chr**	**%**	**Chr**	**%**	**Chr**	**%**
LCR	1	0,2%	0	0%	0	0%	1	0,2%	0	0%	0	0%	0	0%	0	0%	0	0%
-87(C-G)	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
-30(T-A)	11	2,1%	0	0%	0	0%	9	1,7%	0	0%	0	0%	2	0,4%	0	0%	0	0%
-28(A-C)	1	0,2%	1	0,2%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
FSC5(-CT)	3	0,6%	1	0,2%	0	0%	0	0%	0	0%	0	0%	1	0,2%	1	0,2%	0	0%
Cd6(-A)	2	0,4%	0	0%	0	0%	0	0,4%	0	0,4%	0	0%	0	0%	0	0%	2	0,4%
HbS	15	2,8%	0	0%	11	2,1%	0	0%	0	0%	4	0,7%	0	0%	0	0%	0	0%
FSC8 (-AA)	12	2,2%	1	0,2%	4	0,7%	3	0,6%	0	0%	0	0%	4	0,7%	0	0%	0	0%
FSC8/9 (+G)	2	0,4%	0	0%	0	0%	0	0%	2	0,4%	0	0%	0	0%	0	0%	0	0%
Hb Ankara **	1	0,2%	0	0%	1	0,2%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
Cd15 (TGG-TGA)	3	0,6%	0	0%	0	0%	3	0,6%	0	0%	0	0%	0	0%	0	0%	0	0%
Cd22-24 (7 bp del)	2	0,4%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	2	0,4%	0	0%
Cd30 (G-C) **	2	0,4%	2	0,4%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
IVS-I-1 (G-A)	10	1,9%	1	0,2%	2	0,4%	3	0,6%	0	0%	0	0%	2	0,4%	0	0%	2	0,4%
IVS-I-5 (G-C)	2	0,4%	0	0%	0	0%	0	0%	2	0,4%	0	0%	0	0%	0	0%	0	0%
IVS-I-6 (T-C)	16	3%	2	0,4%	0	0%	2	0,4%	2	0,4%	0	0%	5	0,9%	5	0,9%	0	0%

Figure B.4. Distribution layout



IVS-I-110 (G-A)	49	9,1%	4	0,7%	6	1,1%	7	1,3%	4	0%	3	0,6%	4	0,7%	8	1,5%	13	2,4%
Cd39 (CAG-TAG)	13	2,4%	5	0,9%	0	0%	0	0%	3	0,6%	0	0%	0	0%	0	0%	5	0,9%
Cd44 (-C)	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
IVS-II-1 (G-A)	22	4,1%	7	1,3%	1	0,2%	2	0,4%	3	0,6%	2	0,4%	1	0,2%	3	0,6%	3	0,6%
IVS-II-2 (T-A) !	2	0,4%	0	0%	0	0%	2	0,4%	0	0%	0	0%	0	0%	0	0%	0	0%
IVS-II-745 (C-G)	10	1,9%	1	0,2%	0	0%	4	0,7%	0	0%	1	0,2%	1	0,2%	1	0,2%	2	0,4%
IVS-II-848 (C-A)	2	0,4%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	1	0,2%	1	0,2%
HbO Arab **	2	0,4%	0	0%	1	0,2%	0	0%	0	0%	0	0%	0	0%	0	0%	1	0%
HbD LosAngeles	2	0,4%	0	0%	0	0%	0	0%	2	0,4%	0	0%	0	0%	0	0%	0	0%
Cd37 (TGG-TGA) **	3	0,6%	0	0%	0	0%	0	0%	3	0,6%	0	0%	0	0%	0	0%	0	0%
N	105	20%	10	1,9%	21	3,9%	22	4,1%	11	2,1%	5	0,9%	10	1,9%	17	3,2%	9	1,7%
Unidentified	34	6,3%	6	1,1%	4	0,7%	6	1,1%	6	1,1%	2	0,4%	2	0,4%	8	1,5%	0	0%
In Analysis	14	2,6%	2	0,4%	0	0%	0	0%	6	1,1%	0	0%	0	0%	0	0%	6	1,1%
Not Analysed	32	6%	1	0,2%	2	0,4%	3	0%	10	1,9%	4	0,7%	6	1,1%	0	0%	2	0,4%
Total	372	69%	44	8%	51	10%	67	13%	53	10%	21	4%	36	7%	44	8%	44	8%

Figure B.5. Distribution layout (continued)

REFERENCES

Albitar, M., F. E. Cash, C. Peschle and S. A. Liebhaber, 1992, "Developmental Switch in the Relative Expression of the α1- and α2-Globin Genes in Humans and in Transgenic Mice", *Blood*, Vol. 79, pp. 2471-2474.

Attwood, T. K. and D. J. Parry-Smith, *Introduction to Boinformatics*, Pearson Education, Prentice Hall, 1999.

Behringer, R. R., T. M. Ryan, R. D. Palmiter,R. L. Brinster and T. M. Townes, 1998, "Human gamma- to beta-Globin Gene Switching in Transgenic Mice", *Genes and Development*, Vol. 4, pp. 380-390.

Bergeron, B., *Bioinformatics Computing*, Pearson Education Inc., Prentice Hall, 2003.

Bilenoğlu, O., 1996, *Molecular Analysis of β-Thalassemia and Hemoglobin S by the β-Globin StripA^{ssay} A Novel Diagnostic Approach*, MSc. Thesis, Boğaziçi University.

Bilenoğlu, O., 2002, *Beta-Thalassemia in Turkey: Molecular Analysis of the Beta-Globin Gene and The Effect of the Beta-Δ13 Mutation on Gene Expression*, PhD Thesis, Boğaziçi University.

Bowden, D. K., M. A. Wickers and D. R. Higgs, 1992, "A PCR-Based Strategy to Detect the Common Severe Determinants of α-Thalassaemia", *British Journal of Haematology*, Vol. 81, pp. 104-108.

Brittain, T., 2002, "Molecular Aspects of Embryonic Hemoglobin Function", *Molecular Aspects of Medicine*, Vol. 23, pp. 293-342.

Bulger, M. and Groudine, M., 1999, "Looping versus Linking: Toward a Model for Long Distance Gene Activation", *Genes and Development*, Vol. 13, pp. 2465-2477.

Cao, A. and Moi, P., 2003, "Regulation of the Globin Genes", *Pediatric Research*, Vol. 51, pp.415-421.

Chong, S. S., C. D. Boehm, D. R. Higgs and G. R. Cutting, 2000, "Single-Tube Multiplex-PCR Screen for Common Deletional Determinants of α-Thalassemia", *Blood*, Vol. 95, pp. 360-362.

Chui, D. H. K., R. Hardison, C. Riemer, W. Miller, M. F. H. Carver, T. P. Molchanova, G. D. Efremov and T. H. J. Huisman, 1998, "An Electronic Database of Human Hemoglobin Variants on the World Wide Web", *Blood*, Vol. 91, pp. 2643-2644.

Close, J., L. Game, B. Clark, J. Bergounioux, A. Gerovassili and S. L. Thein, 2004, "Genome Annotation of a 1.5 Mb Region of Human Chromosome 6q23 Encompassing a Quantitative Trait Locus for Fetal Hemoglobin Expression in Adults", *BioMed Central Genomics*, Vol. 5, pp.1-14.

Craddock, C. F., P. Vyas, J. A. Sharpe, H. Ayyub, W. G. Wood and D. R. Higgs, 1995, "Contrasting Effects of α- and β-Globin Regulatory Elements on Chromatin Structure May Be Related to Their Different Chromosomal Environments", *The EMBO Journal,* Vol. 14, pp. 1718-1726.

Dalton, R. N., C. Turner, M. Dick, S. E. Height, M. Awogbade, B. Inusa, I. Okpala, S. O'Driscoll, S. L. Thein and D. C. Rees, 2005, "The Measurement of Urinary Hydroxyurea in Sickle Cell Anaemia", *British Journal of Haematology*, Vol. 130, pp. 138-144.

Das, S. K.and Talukder, G., 2002, "Beta-Globin Gene and Related Disease: A Review", *International Journal of Human Genetics*, Vol. 2, No. 3, pp. 139-152.

Gülen, İ. R., 2003, *Deniz: the Database for β-thalassemia allele frequent in world populations*, http://www.ndal.boun.edu.tr:81/Deniz.html

Dixit, A., T. C. Chatterjee, P. Mishra, D. R. Choudhry, M. Mahapatra, S. Tyagi, M. Kabra, R. Saxena and V. P. Choudhry, 2005, "Hydroxyurea in Thalassemia Intermedia-A Promising Therapy", *Annual Hematology*, Vol. 84, pp. 441-446.

Dode, C., R. Krishnamoorthy, J. Lamb and J. Rochette, 1992, "Rapid Analyses of $-\alpha^{3.7}$Thalassaemia and $\alpha\alpha\alpha^{anti3.7}$ Triplication by Enzymatic Amplification Analysis", *British Journal of Haematology*, Vol. 82, pp. 105-111.

Elnitski, L., B. Giardine, P. Shah, Y. Zhang, C. Riemer, M. Weirauch, R. Burhans, W. Miller and R. C. Hardison, 2005, "Improvements to GALA and dbERGE: Databases Featuring Genomic Sequence Alignment, Annotation and Experiemental Results", *Nucleic Acid Research*, Vol. 33, pp. 466-470.

Fang, X., J. Sun, P. Xiang, M. Yu, P. A. Navas, K. R. Peterson, G. Stamatoyannopoulos and Q. Li, 2005, Synergistic and Additive Properties of the Beta-Globin Locus Control Region (LCR) Revealed by 5' HS3 Deletion Mutations: Implication for LCR Chromatin Architecture", *Molecular and Cellular Biology*, Vol. 25, No. 16, pp. 7033-7041.

FileMaker, Inc., 2002, *FileMaker Pro 6 User's Guide*.

Garner, C. P., T. Tatu, S. Best, L. Creary and S. L. Thein, 2002, "Evidence of Genetic Interaction Between the β-Globin Complex and Chromosome 8q in the Expression of Fetal Hemoglobin", *American Journal of Human Genetics*, Vol. 70, pp. 793-799.

Garner, C., N. Silver, S. Best, S. Menzel, C. Martin, T. D. Spector and S. L. Thein, 2004, "Quatitative Trait Locus on Chromosome 8q Influences the Switch from Fetal to Adult Hemoglobin", *Blood*, Vol. 104, pp. 2184-2186.

Giardine, B., L. Elnitski, C. Riemer, I. Makolowska, S. Schwartz, W. Miller and R. C. Hardison, 2006, "GALA, a Database for Genomic Sequence Alignments and Annotations", *Genome Research*, Vol. 13, pp. 732-741.

Globin Gene Server, http://www.globin.cse.edu

Gülen, R. İ., 2003, *NeHIR, DoTBAr and DENIZ: Three Biological Databases Established at Fatih University in İstanbul*, M. S. Thesis, Fatih University.

Hardison, R., C. Riemer, D. H. K. Chui, T. H. J. Huisman and W. Miller, 1998, "Electronic Access to Sequence Alignments, Experimental Results, and Human Mutations as an Aid to Studying Globin Gene Regulation", *Genomics*, Vol. 47, pp. 429-437.

Higgs, D. R., 2004, "Gene Regulation in Hematopoiesis: New Lessons from Thalassemia", *Hematology*, pp. 1-13.

Ho, P. J and Thein, S. L., 2000, "Gene Regulation and Deregulation: a β-Globin perspective", *Blood Reviews*, Vol. 14, pp. 78-93.

Huang, B.-L., I. R. Fan-Chiang, S. C. Wen, H.-C.Koo, W. Y. Kao, N. R. Gavva and C.-K. James Shen, 1998, "Derepression of Human Embryonic ζ-Globin Promoter by a Locus-Control Region Sequece", *Proceedings of National Academy of Science*, Vol. 95, pp. 14669-14674.

Huisman, T. H. J., M. F. H. Carver and E. A. Baysal (editors), 1997, *A Syllabus of Thalassemia Mutations* , The Sickle Cell Anemia Foundation, Augusta, GA, USA.

Huisman, T. H. J., M. F. H. Carver and G. D. A. Efremov (editors), 1996, *A Syllabus of Human Hemoglobin Variants*, The Sickle Cell Foundation, Augusta, GA, USA.

Kanavakis, E., J. Traeger-Synodinos, S. Lafioniatis, C. Lazaropoulou, T. Liakopoulou, G. Paleologos, A. Metaxotou-Mavrommati, A. Stamoulakatou and I. Papassotiriou, 2004, "A Rare Example that Coinheritance of a Severe Form of β-Thalassemia and α-Thalassemia Interact in a 'Synergistic' Manner to Balance the Phenotype of Classic Thalassemic Syndrome", *Blood Cells, Molecules, and Diseases*, Vol. 32, pp. 319-324.

Liebhaber, S. A., Z. Wang, F. E. Cash, B. Monks and J. E. Russell, 1996, "Developmental Silecing of the Embryonic ζ-Globin Gene: Concerted Action of the Promoter and the 3'-Flanking Region Combined with Stage-Specific Silencing by the Transcribed Segment", *Molecular and Cellular Biology*, Vol. 16, pp. 2637-2646.

Liu, T.-C., S.-F. Lin, T.-Y. Yang, J.-P. Lee, T.-P. Chen and J.-G. Chang, 1997, "Prenatal Diagnosis of Thalassemia in the Chinese", *American Journal of Hematology*, Vol. 55, pp. 65-68.

Liu, Y. T., J. M. Old, K. Miles, C. A. Fisher, D. J. Weatherall and J. B. Clegg, 2000, "Rapid Detection of α-Thalassaemia Deletions and α-Globin Gene Triplication by Multiplex Polymerase Chain Reactions", *British Journal of Haematology*, Vol. 108, pp. 295-299.

Lucarelli, G., M. Andreani, E. Angelucci, 2002, "The Core of Thalassemia by Bone MarrowTransplantation", *Blood Reviews*, Vol. 16, pp. 81-85.

Luscombe, N. M., D. Greenbaum, M. Gerstein, 2001, "What is Bioinformatics? A Proposed Defintion and Overview of the Field", *Methods of Information in Medicine*, Vol. 40, pp. 346-358.

Olivieri, N. F., 1999, "The β-Thalassemias", *The New England Jounal of Medicine*, Vol. 341, pp. 99-109.

Özçelik, H., A. N. Başak, Ş. Tüzmen, B. Kırdar, N. Akar, 1993. "A Novel Deletion in a Turkish β-Thalassemia Patient Detected by DGGE and Direct Sequencing: FSC 22-24 (-7 bp)", *Hemoglobin*, Vol.17, pp. 387-391

Patrinos, G. P., B. Giardine, C. Riemer, W. Miller, D. H. K. Chui, N. P. Anagnou, H. Wajcman and R. C. Hardison, 2004, "Improvements in the *HbVar* Database of Human Hemoglobin Variants and Thalassemia Mutations for Population and Sequence Variation Studies", *Nucleic Acid Research*, Vol. 32, pp. 537-541.

Puthenveetil, G., J. Scholes, D. Carbonell, N. Qureshi, P. Xia, L. Zeng, S. Li, Y. Yu, A. L. Hiti, J.-K. Yee and P. Malik, 2004, "Successful Correction of the Human β-Thalassemia Major Phenotype Using a Lentiviral Vector", *Blood*, Vol. 104, pp. 3445-3453.

Richon, A. B., *A Short History of Bioinformatics*, 2001, http://www.netsci.org/Science/Bioinform/feature06.html.

Riemer, C., A. ElSherbini, N. Syojanovic, S. Scwartz, P. B. Kwitkin, W. Miller and R. Hardison, 1998, "A Database of Experimental Results on Globin Gene Expression", *Genomics*, Vol. 53, pp. 325-337.

Saiki, R. K., P. S.Walsh, C. H. Levenson and H. A. Erlich, 1989, "Genetic Analysis of Amplified DNA with Immobilized Sequence-specific Oligonucleotide Probes", *Proceedings of the National Academy of Sciences*, USA, Vol. 86, pp 6230-6234.

Stamatoyannopoulos, G., 2005, "Control of Globin Gene Expression during Development and Erythroid Differentiation", *Experimental Hematology*, Vol. 33, pp. 259-271.

Stathopulos, P. B., 2003, "Taking the Good out of the Bad: Lentiviral-Based Gene Therapy of the Hemoglobinopathies", *Biotechnology Advances*, Vol. 21, pp. 513-526.

Steinberg, M. H., E. Voskaridou, A. Kutlar, D. Loukopoulos, M. Koshy, S. K. Ballas, O. Castro, and F. Barton, 2003, "Concordant Fetal Hemoglobin Response to Hydroxyurea in Siblings with Sickle Cell Disease", *American Journal of Hematology*, Vol. 72, pp. 121-126.

Tadmouri, G. O. and A. N. Başak, 2001, "β-Thalassemia in Turkey: A Review of the Clinical, Epidemiological, Molecular and Evolutionary Aspects", *Hemoglobin*, Vol. 25, pp. 227-239.

Tadmouri, G. O., Ş. Tüzmen, H. Özçelik, A. Özer, S. M. Baig, E. B. Senga and A. N. Başak, 1998, "Molecular and Population Genetic Analyses of β-Thalassemia in Turkey", *American Journal of Hematology*, Vol. 57, pp. 215-220.

Thein, S. L., "Genetic Modifiers of β-Thalassemia", 2005, *Haematologica/The Hematology Journal*, Vol. 90, No. 5, pp. 649-660.

Thein, S. L., 2004, "Genetic Insights into the Clinical Diversity of β Thalassaemia", *British Journal of Haematology*, Vol, 124, pp. 264-274.

Thein, S. L., 2005, "Pathophysiology of β Thalassemia-A Guide to Molecular Therapies", *Hematology*, pp. 31-37.

Traeger-Synodinos J., E. Kanavakis, C. Vrettou, M. Maragoudaki, T. H., Michael, A. Metaxotou-Mavromati and C. Kattamis, 1996, "The Triplicated α-Globin Gene Locus in β-Thalassemia Heterozygotes: Clinical, Haematological, Biosynthetic and Molecular Studies", *British Journal of Haematology*, Vol. 95, pp. 467-471.

Tüzmen, S. and Schechter, A. N., 2001, Genetic Disease of Hemoglobin: Diagnostic Methods for Elucidating β-Thalassemia Mutations", *Blood Reviews*, Vol. 15, pp. 19-29.

Weatherall, D. J, 2004, "Thalassaemia: the Long Road from Bedside to Genome", *Genetics*, Vol. 5, pp. 1-7.

Weatherall, D. J., 2001, "Phenotype-Genotype Relationships in Monogenic Disease: Lessons from the Thalassaemias", *Nature Reviews Genetics*, Vol. 2, pp. 245-255.

Weatherall, D. J., 2004, "The Thalassemias: The Role of Molecular Genetics in an Evolving Global Health Problem", *American Journal of Human Genetics*, Vol. 74, pp. 385-392.

Wood, W. G., 1993, “Increased HbF in Adult Life”, *Baillière’s Clinical Hematology*, chapter 7.