SNP IDENTIFICATION FOR BETA THALASSEMIA IN NORTHERN INDIA POPULATION

SUBMITTED BY

KASHIKA GUPTA 061711

NAVNEET KAUR 061717

Submitted in partial fulfillment of the Degree of Bachelor of Technology

DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT

May-2010

CERTIFICATE

This is to certify that the work entitled, "SNP Identification for Thalassemia in Northern India Population" submitted by Kashika Gupta and Navneet Kaur in partial fulfillment for the award of degree of Bachelor of Technology in 2010 of Jaypee University of Information Technology has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Dr. Harvinder Singh

Dr. P.K. Naik

ACKNOWLEDGMENT

The project in this report is an outcome of continual work over a period of 1 year and intellectual support from various sources. It is therefore almost impossible to express adequately the debts owed to many persons who have been instrumental in imparting this work, a successful status. It is however a matter of pleasure to express our gratitude and appreciation to those who have been contributing to bring about this project.

We take this opportunity to thank our project guides Dr. Harvinder Singh, Senior Lecturer Bioinformatics and Biotechnology Department, JUIT and Dr. P.K.Naik, Assistant Professor for lending us stimulating suggestions, innovative quality guidance and creative thinking. Their practicality, constant encouragement and advice helped us in all stages of the project. Their scientific views and scientific approach will always be the source of motivation for us. We are grateful to them for the support, they provided us in doing things at our pace and for being patient with our mistakes.

Our special thanks to Head of Bioinformatics and Biotechnology Department, JUIT for providing all necessary equipments in our laboratory and an opportunity for exploring our knowledge through valuable project.

We would also like to acknowledge our PhD scholars for their constant support.

Furthermore our lab assistants were a constant support for providing us with a good atmosphere and equipments for carrying out our project in the best manner and thanks to our classmates for their support and motivation.

Kashika Gupta

Navneet Kaur

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CAPS -Cleaved Amplified Polymorphic sequences

dH_20- Distilled Water

DNA -Deoxy Ribonucleic Acid

EDTA- Ethylene Diamine Tetra-acetic Acid

EtBr- Ethidium Bromide

PCR-Polymerase Chain Reaction

RBCs –Red Blood Cells

RE-Restriction Enzyme

RFLP –Restriction Fragment Length polymorphism

SNP – Single nucleotide polymorphism

TAE- Tris Acetic acid EDTA

TE-Tris EDTA Buffer

ABSTRACT

Beta thalassemia is a heterogenous inherited disorder of β-globin synthesis with over 200 different β-globin mutations reported worldwide. The β-globin defects result in the absence or reduction in the synthesis of β- globin chains.

The highest prevalence of the disease occur in the Mediterranean region, parts of the North and West Africa, the Middle East, the Indian subcontinent and the Southern Far East and Southeastern Asia, all together composing the so-called thalassemia belt. Pervious population screening has demonstrated a specific spectrum of mutations in each ethnic region. Generally two or three of these mutations are found with high prevalence while others are rare in a specific population.

The statistics of beta thalassemia in India presents a grim scenario. It has been estimated that there are 45 million carriers of beta thalassemia in India and about 15,000 infants with homozygous beta thalassemia are born every year which constitutes about 10% of the total thalassemics born in the world.

β-thalassemia is the most prevalent single-gene disorder. Since no viable forms of treatment are available, the best course is prevention through prenatal diagnosis. The study of these genetic mutations has been based on polymerase chain reaction (PCR) and visualization of the fragments after electrophoresis in agarose gel. The reliability of diagnosis can be increased by employing DNA-based techniques such as PCR. The project aims at discovering new SNP by sequencing of the mutant samples, then developing multiplexing primers for easy diagnosis. The study also aimed at determining the effect of mutations (SNPs) on oxygen binding affinity of hemoglobin using various bioinformatic tools (JAGUAR).

CHAPTER-1

INTRODUCTION

1.1 THALASSEMIA

Thalassemia is a kind of chronic, inherited, microcytic anemia characterized by defective hemoglobin synthesis and ineffective erythropoiesis. Thalassemia, which was described in 1927 by Cooley et al., is a kind of severe anemia associated with splenomegaly and bone abnormalities.

Thalassemia is the most common inherited single gene disorder in the world. It is an autosomal recessive single gene disorder. Thalassemia is characterized by absent or decreased production of the normal hemoglobin.

Hemoglobin is an iron-rich protein in red blood cells. It carries oxygen to all parts of the body. It also carries carbon dioxide (a waste gas) from the body to the lungs, where it is exhaled. In thalassemia patients, a mutation or deletion of the genes that control globin production occurs. This leads to a decreased production of the corresponding globin chains and an abnormal hemoglobin ratio. This abnormal ratio leads to decreased synthesis of hemoglobin and the expression of thalassemia. The globin that is produced in normal amounts winds up in excess and forms red cell aggregates or inclusions. These aggregates become oxidized and damage the cell membrane, leading to hemolysis, ineffective erythropoiesis, or both. The quantity and properties of these globin chain aggregates determine the characteristics and severity of the thalassemia.

Hemoglobin is a tetramer composed of 2α and 2β subunits and each of them containing 1 heme attached to the subunit by a coordination bond between iron and histidine. Also the hemoglobin in our body exists in two states-oxygenated form and deoxygenated form. In case of thalassemia the oxygen binding affinity of hemoglobin gets affected. Oxygen affinity of hemoglobin could be evaluated by computing the energy difference between the optimized structure of oxygenated and deoxygenated forms of normal and mutated hemoglobin by doing computational analysis using JAGUAR program.

1.2 PATHOPHYSIOLOGY

Normal hemoglobin, the oxygen carrying protein in red blood cells also called as hemoglobin A, has four protein chains—two alpha globin and two beta globin. These two chains are made from specific genes we inherit from our parents. When these specific genes are not working properly, hemoglobin production is affected. The two major types of thalassemia, alpha and beta, are named after defects in these protein chains.

1.2.1 Alpha (α) thalassemia

In α thalassemia, production of α globin chain of hemoglobin is affected which involves mutation in the genes HBA1 and HBA2 on chromosome 16 and results in decreased alpha-globin production, therefore fewer alpha-globin chains are produced. The excess β chains form unstable tetramers.

1.2.2 Beta (β) thalassemia

Beta thalassemia is due to mutations in the HBB gene on chromosome 11. So there is a relative excess of α chains, but these do not form tetramers, rather, they bind to the red blood cell membranes, producing membrane damage, and at high concentrations they form toxic aggregates.

The severity of the disease depends on the nature of the mutation.

- Mutations are characterized as (β^0) if they prevent any formation of β chains.
- Mutations are characterized as (β^+) if they allow some β chain formation to occur.

OBJECTIVES

The main objectives of the project are listed below:-

- To study the distribution of major β-thalassemia mutations in Himachal Pradesh.
- Designing primers from the flanking regions of target SNPs present in beta globin gene.
- Sequencing and identification of de novo SNPs.
- To study the effect of mutations on the binding affinity of oxygen in hemoglobin protein using calculation of binding energies.

CHAPTER-2

REVIEW LITERATURE

(Huisman et al., 1997) stated that β-thalassemia can be caused by more than 200 different *HBB* gene mutations, however, the prevalent molecular defects are limited in each at-risk population. This phenomenon has greatly facilitated molecular genetic testing.

At first, DNA analysis for beta thalassemia was performed by linkage analysis with polymorphic restriction endonuclease sites (Boehm et al., 1983). Once the genetic defects responsible for the disease were delineated on the molecular level, direct detection of the mutation became possible. When a mutation either abolishes or creates a restriction endonuclease site, it can be detected by digesting DNA with the appropriate restriction enzyme (Bunn and Forget et al., 1986). However, the majority of mutations cannot be directly detected with restriction enzymes, therefore oligonucleotide probes complementary to the normal and mutant gene are used (Pirastu et al., 1983). Before a prenatal diagnosis program can be implemented, it is essential to determine which type of molecular lesion causes beta thalassemia in a particular geographic area. In addition, determining the frequency of each mutation in a given region permits the choice of appropriate enzymes or oligonucleotide probes. This approach has been successfully adopted for prenatal diagnosis in Italy (Rosatelli et al., 1987).

Commonly occurring mutations of the *HBB* gene are detected by a number of PCR-based procedures. The most commonly used methods are reverse dot blot analysis or primer-specific amplification, with a set of probes or primers complementary to the most common mutations in the population from which the affected individual originated (Old et al., 2005).

Other methods based on real-time PCR or microarray technology because of their reproducibility, rapidity, and easy handling are potentially suitable for the routine clinical laboratory (Vrettou et al., 2003, Ye et al., 2007).

The World Health Organization (WHO) has suggested that about 5% of the world population are carriers for different inherited disorders of hemoglobin. WHO reports also state that about 370,000 severely affected homozygotes or compound heterozygotes of thalassemia are born every year. The general incidence of thalassemia trait and sickle cell hemoglobinopathies in India varies between 3-17% and 1-44% respectively. It is estimated that there are about 65,000-67,000 beta-thalassemia patients in India with around 9,000-10,000 cases being added every year. The carrier rate for beta-thalassemia gene varies from 1 to 3% in Southern India to 3 to 15% in Northern India. The distribution of beta thalassemia is not uniform in Indian subcontinent. The highest frequency of beta thalassemia trait is reported in Gujarat (10-15%), followed by Sindh (10%), Punjab (6.5%), Tamil Nadu (8.4%) and Maharashtra.

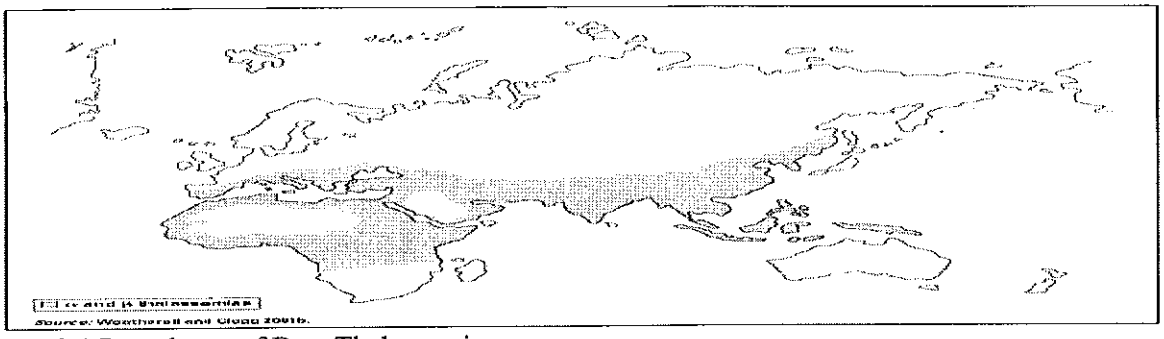

Fig. 2.1 Prevalence of Beta Thalassemia

(Antonio Cao et. al., 2007) stated that if targeted mutation analysis fails to detect the mutation, mutation scanning or sequence analysis can be used to detect mutations in the *HBB* coding region (mutations in the non-coding region would not be detected by this analysis). Sensitivity of both mutation scanning and sequence analysis is 99%.

Table2.1 Molecular Genetic Testing Used for Beta Thalassemia

Gene Symbol	Test Method	Mutations Detected	Mutation Detection Frequency by Test Method	Test Availability
HBB	Targeted mutation analysis	*HBB* mutation panels vary by laboratory	Variable depending on mutations included in panel and individual's ethnicity	Clinical Testing
	Mutation scanning/ sequence analysis	Sequence variants in *HBB* coding region	99%	

(Galanello et al., 1979) studied the red blood indices for diagnosis of beta thalassemia, results are shown in the table. Affected individuals demonstrate the (RBC) morphologic changes of microcytosis, hypochromia, anisocytosis, poikilocytosis (spiculated tear-drop and elongated cells), nucleated red blood cells (i.e., erythroblasts), a low level of red blood cells, smaller and pale RBCs.

Table 2.2 Red Blood Indices in Beta Thalassemia

Red Blood Cell Index	Normal [1]		Affected	Carrier [1]
	Male	Female	β-Thal Major	β-Thal Minor
Mean corpuscular volume (MCV fl)	89.1±5.01	87.6±5.5	50-70	<79
Mean corpuscular hemoglobin (MCH pg)	30.9±1.9	30.2±2.1	12-20	<27
Hemoglobin (Hb g/dL)	15.9±1.0	14.0±0.9	<7	Males: 11.5-15.3 Females: 9.1-14

1. Data from Galanello et al (1979)

(Camaschella et. al., 2006) stated that thalassemia intermedia is a clinical definition applied to patients whose clinical phenotype is milder than that of thalassemia major. Criteria used to define thalassemia intermedia including age at presentation, hemoglobin or fetal hemoglobin levels and transfusion independence, are unsatisfactory.

The possibility of typing the molecular defect offers a new tool for the diagnosis of thalassemia intermedia. Nevertheless, the beta-genotype alone is not predictive of the phenotype in all cases. Although benign, the clinical course of thalassemia intermedia is characterized by several complications that can be prevented by an accurate follow-up. The conventional treatment of thalassemia intermedia remains controversial; it is hoped that recent advances in the pharmacological manipulation of hemoglobin switching will offer a therapeutic option in the future, at least to selected patients.

(Antonio Cao et al., 2007) stated that clinical presentation of thalassemia major occurs between ages 6 and 24 months. Affected infants fail to thrive and become progressively pale. Feeding problems, diarrhea, irritability, recurrent bouts of fever, and progressive enlargement of the abdomen caused by splenomegaly may occur. If the diagnosis of thalassemia major is established at this stage and if a regular transfusion program that maintains a minimum Hb concentration of 95 to 105 g/L is initiated, growth and development are normal until age 10 to 11 years.

After age 10 to 11 years, affected individuals are at risk of developing severe complications related to iron overload, depending on their compliance with chelation therapy. Complications of iron overload in children include growth retardation and failure of sexual maturation. The risk for hepatocellular carcinoma is increased secondary to liver viral infection, iron overload, and longer survival (Borgna-Pignatti et al., 2004). Cardiac complications are reported to cause 71% of the deaths in individuals with β-thalassemia major (Borgna-Pignatti et al., 2004).

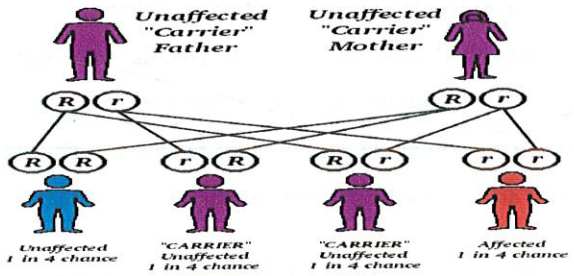

Fig. 2.2 Pattern of Inheritance

(Schoenstadt et al., 2006) stated that the causes of thalassemia are directly linked to genetics and how the genes that affect hemoglobin production are inherited. People with moderate to severe forms received variant genes from both parents. People who are carriers of the disease received variant genes from one parent and normal genes from the other parent. Thalassemia occurs when there are variant or missing genes that affect how the body makes hemoglobin. Beta-Thalassemia occurs when there is a problem with the beta-globin part of hemoglobin. There are both mild and severe forms of alpha and beta thalassemia.

(Antonio Cao et al., 2007) stated that the most common treatment for all major forms of thalassemia is red blood cell transfusions. These transfusions are necessary to provide the patient with a temporary supply of healthy red blood cells with normal hemoglobin capable of carrying the oxygen that the patient's body needs. Bone marrow transplantation (BMT) from HLA-identical siblings represents an alternative to traditional transfusion and chelation therapy. The outcome of BMT is related to the pretransplantation clinical conditions, specifically the presence of hepatomegaly, extent of liver fibrosis, and magnitude of iron accumulation. In children who lack the above risk factors, disease-free survival is over 90% (Gaziev and Lucarelli, 2003).

BMT from unrelated donors has been carried out on a limited number of individuals with β-thalassemia. Provided that selection of the donor is based on stringent criteria of HLA compatibility and that individuals have limited iron overload, results are comparable to those obtained when the donor is compatible. (La Nasa et al., 2005).

Cord blood transplantation from a related donor offers a good probability of a successful cure and is associated with a low risk of GVHD (Locatelli et al., 2003, Walters et al., 2005). For couples who have already had a child with thalassemia and who undertake prenatal diagnosis in a subsequent pregnancy, prenatal identification of HLA compatibility between the affected child and an unaffected fetus allows collection of placental blood at delivery and the option of cord blood transplantation to cure the affected child. (Orofino et al., 2003)

(Rosatelli et al., 2009) completed a regular peer review on prenatal diagnosis. The majority of defects affecting the β globin gene are point mutations that occur in critical areas for its function, or single/few base addition/deletion that change the frame in which triplets are translated into protein. Very rarely β thalassemia results from gross rearrangement in the β-globin gene cluster. In spite of the marked molecular heterogeneity, a limited number of molecular defects are prevalent in every at risk population. This may be very useful in practice, because a panel of most frequent mutations to be searched for can be designed according to the carrier's ethnic origin. Known mutation detection is carried out by a number of PCR-based techniques.

Primer-specific amplification is based on the principle that a primer carrying a mismatch in its 3' region cannot anneal on its template. With this method, the target DNA fragment is amplified in two separate PCR reactions using a common primer and either of the two following primers, the one complementary to the mutation to be detected (β-thalassemia primer) and one complementary, at the same position, to the normal DNA (normal primer). Normal DNA is amplified only by the normal primer while DNA from homozygotes only by the β-thalassemia primer and DNA from heterozygotes by both primers.

Reverse oligonucleotide hybridization: When the spectrum of mutations to be searched is complex, ARMS is not the most appropriate method while RDB results were more informative and efficient. Reverse dot-blot analysis was first described by Saiki et al., and then developed later to screen many β-thalassemia mutations in Sicilian population and for use in prenatal diagnosis reverse dot-blot analysis is a technique for immobilizing allele-specific oligonucleotide probes on a nylon membrane rather than the individual DNA samples. This is a non-radioactive method. In this format, multiple pairs of mutant and normal ASO probes are spotted on strips of nylon membranes. For each diagnostic test, a spotted strip containing many normal and mutant oligonucleotides, is hybridized with a specific DNA probe to screen many mutations.

Unknown mutations detection: When carriers escape to the above mutation detection approaches, further investigations need to be carried out by alternative methods which uncover the presence of unknown mutations by scanning the whole gene. Nowadays, considering the small size of the β-globin gene (1.8kb), the simplified technologies available and the reduced costs of analysis, direct sequencing, based on cycle sequencing with fluorescent dye terminators and automated capillary DNA sequencing technology, seems to be faster and most useful approach to detect unknown thalassemia mutations. If a mutation is not detected by sequence analysis, we search for the presence of small deletions by polyacrylamide gel electrophoresis of amplicons designed for the most frequent small deletional defects of the β-globin gene (gap PCR).

Furthermore, the presence of larger deletions of the cluster may be identified by Southern-blotting or more recently by Multiple Ligation-dependent Probe Amplification (MLPA). In a very limited number of cases, direct sequencing from position 600 to 60 bp downstream from the β-globin gene and methods for deletion detection, failed to detect the disease causing defect. In these cases, the molecular defect may reside either in the locus control region of the β-globin gene cluster, or in one of the genes, outside the β-globin gene region, encoding for regulatory proteins. Very recently it has been proved that the β- thalassaemia like phenotype could be caused by the coinheritance of a β-globin gene defect and a duplication of the β-globin gene cluster, which results in an excess of β chain. In these selected cases, the characterization of these β-globin gene rearrangements can be routinely carried out with success by MLPA analysis.

(Bashyam et al., 2004) reported for the first time presence of the IVSII 837 mutation in the Indian population. This also report a novel diagnostic application during RDB-based screening for the detection of the (c.92G>C) mutations. The prevalence of β-thalassemia was extensively investigated in the South Indian population, especially from the state of Andhra Pradesh. Screening for causal mutations was carried out on genomic DNA isolated from patient blood samples by using the routine reverse dot blot (RDB) and amplification refractory mutation system-polymerase chain reaction (ARMSPCR) techniques. DNA sequencing was performed wherever necessary. Gene sequencing confirmed the codon 30 (G-C) (c.92G>C) mutation and the rare codon 5 (-CT) (c.17_18delCT) and IVS-II-837 (T-G) (IVSII-14T>G) mutations.

(Reichert et al., 2007) stated that the genotypic profile of β-thalassemia shows great variability. Hence, it would be arbitrary to infer regional study results as being representative of the whole population. So researchers of different regions should identify their most frequent genotypes to provide better understanding on this disease and state adequate public health policies. He carried out research on Brazilian southernmost states and found out that β0 codon 39 nonsense mutation was the most frequent alteration (50.9%), followed by β+ IVSI 110 G>A (18.1%), β0 IVSI 1 G>A (12.9%), β+ IVSI 6 T>C (9.5%), and other rare mutations (8.6%).

(Chen et al., 2007) identified a novel β++-thalassemia mutation of −73 (A→T) within the conserved CCAAT box at position −76 to −72 from the cap site of the β-globin gene. In restriction fragment length polymorphism (RFLP) haplotype analysis, the results indicated that this promoter mutation might be linked to the absence of BamHI-3′β restriction site. Identification of new mutations and update of the mutation spectrum of thalassemia in one ethnic population is always needed for improving genetic counseling in the high-risk areas. It is the first time that a natural mutation was found in this conserved motif in β-gene. The single nucleotide substitution from A to T at position −73 in the β-globin gene promoter is a 'mild' mutation responsible for β++-thalassemia.

Currently, amniocentesis, chorionic villus sampling (CVS) and fetal blood sampling are used to obtain fetal cells for genetic diagnosis. These invasive procedures pose a small but not negligible risk for the fetus. (Cheung et al., 2005) used enrichment of fetal cells, such as erythroblasts, from maternal blood for the diagnosis of some chromosomal disorders and in sex determinations. He reported the detection of point mutations in single gene disorders using this method of prenatal diagnosis by enriching fetal cells from maternal blood by magnetic cell sorting followed by isolation of pure fetal cells by microdissection. In two pregnancies at risk for sickle cell anemia and beta-thalassaemia, we successfully identified the fetal genotypes. Thus, prenatal diagnosis of single gene disorders by recovering fetal cells from maternal circulation appears to be a feasible approach.

(Tamhankar et al., 2008) stated that premarital screening in extended family members, followed by prenatal diagnosis is acceptable and the most effective strategy for control of thalassemia in developing countries like India. He carried out premarital testing for thalassemia carrier state. Hemoglobin, red cell indices were measured by a cell counter and hemoglobin fractionation was carried out by high performance liquid chromatography (HPLC). In cases with HbA2 >3.5%, or with variant hemoglobin, mutation screen was done by amplification refractory mutation system polymerase chain reaction (ARMS-PCR).

DNA isolation from ethylene diamine tetraacetic acid (EDTA)-anticoagulated blood samples was done by phenol chloroform method (Poncz et al., 1982). The delineation of mutations in the beta-globin gene was carried out by the polymerase chain reaction (PCR) based allele-specific amplification refractory mutation system (ARMS) (Agarwal et al., 2003). The result of ARMSPCR was confirmed by reverse dot blot (RDB) (Colah et al., 1997).

(Holding and Monk, 1989) first used PGD for diagnosis of beta-thalassaemia by DNA amplification in single blastomeres from mouse preimplantation embryos. Preimplantation technique was introduced in late 1980s as an option for avoiding the decision to terminate an established pregnancy diagnosed as affected by conventional approaches.

(Kuliev et al., 1998) stated that preimplantation genetic diagnosis (PGD) is an important option for couples at risk of having children with β-globin mutations to avoid selective abortions of affected fetuses following prenatal diagnosis. PGD was performed for thalassemia in 12 clinical cycles (IVS1-110, and IVS-745 mutations) using biopsy of the first and second polar bodies (PBs) extruded from oocytes during maturation and fertilization, coupled with nested polymerase chain reaction analysis and restriction digestion. A total of 118 oocytes were obtained, of which 78 had results for both the first and the second PBs. This resulted in the selection and transfer of 30 unaffected embryos (2.5 embryos per cycle).

To avoid a possible misdiagnosis due to allele dropout (ADO), there was simultaneous detection of two highly polymorphic linked markers, a short tandem repeat immediately at the 5' end of the globin gene and HUMTH01 which is a syntenic short tandem repeat. Confirmation studies of the embryos resulting from the oocytes predicted to contain an affected gene confirmed the diagnosis in 98% of the cases, thus demonstrating the accuracy and reliability of PB PGD of thalassemia mutations.

The term ***preimplantation genetic diagnosis*** describes those procedures which involve the removal of one or more nuclei from oocytes (polar bodies) or embryos (blastomeres of trophectoderm cells) to test for mutation in the target gene or aneuploidy before transfer. PGD requires that couples at risk undergo in vitro fertilization (IVF) even if not infertile and for this reason a multidisciplinary approach including an appropriate genetic counseling and the referral to both a fertility clinic and to a high specialized molecular genetics laboratory is mandatory.

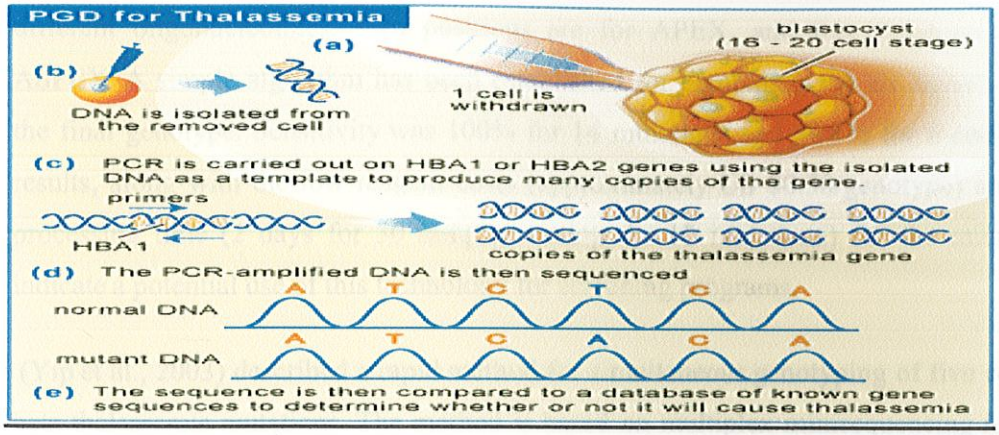

Fig 2.3 Preimplantation Genetic Diagnosis

(Yershov et al., 1996) developed the technology of sequencing by hybridization to oligonucleotide microchips (SHOM) and its application to diagnostics for genetic diseases. A robot has been constructed to manufacture sequencing "microchips." The microchip is an array of oligonucleotides immobilized into gel elements fixed on a glass plate. Hybridization of the microchip with fluorescently labeled DNA was monitored in real time simultaneously for all microchip elements with a two-wavelength fluorescent microscope equipped with a charge-coupled device camera.

SHOM has been used to detect beta-thalassemia mutations in patients by hybridizing PCR-amplified DNA with the microchips. A contiguous stacking hybridization technique has been applied for the detection of mutations; it can simplify medical diagnostics and enhance its reliability. The use of multicolor monitoring of contiguous stacking hybridization is suggested for large-scale diagnostics and gene polymorphism studies.

(Gemignani et al., 2002) developed thalassochip, based on the arrayed primer extension (APEX) technology implemented with allele-specific primed extension (ASPEX). APEX consists of a sequencing reaction primed by an oligonucleotide anchored to a glass slide (with its 5' end) terminating just one nucleotide before the mutation site. DNA polymerase extends it by adding one fluorescently labeled dideoxynucleoside triphosphate complementary to the variant base. In the present study, the ß-globin gene was amplified from human genomic. Each mutation is specified by a pattern of six different oligonucleotides. Two positions are for APEX, and four positions are for ASPEX. A simple algorithm has been established for combining the six signals to give the final genotype. Sensitivity was 100% for 14 mutations and 90.9% for ß-codon. The results, along with the low reagent costs (approximately US $0.50/genotype) and short processing time (2 days for 30 samples testing for 17 mutations) for the microarray, indicate a potential use of this technology for screening programs.

(Yip et al., 2003) described a rapid method for simultaneous genotyping of five common beta thalassemia mutations. The method is based on multiplex minisequencing followed by analysis of the single-base-extended products by denaturing HPLC (DHPLC). In minisequencing, a primer anneals to a DNA template immediately upstream of a mutation site to be analyzed and is extended by a single base as a result of incorporating a dideoxyribonucleotide triphosphate (ddNTP). The single-stranded extension products are then separated by ion-pair reversed-phase liquid chromatography under completely denaturing condition, and the retention time is a function of both size and base composition. The DNA template for minisequencing was first generated by amplifying the whole human ß-globin gene using primers HBBpF1 and HBBpR1.

(Wang et al., 2003) used Gap-PCR to simultaneously amplify the ß-globin gene from genomic DNA and to detect the Δ619bp deletion mutation. Multiplex minisequencing was also performed on the amplified ß-globin fragment to detect an additional 15 common Southeast Asian and Indian ß-thalassemia mutations. Site-specific primers of different lengths were subjected to multiple rounds of annealing and single-nucleotide extension in the presence of thermostable DNA polymerase and the four dideoxynucleotides, each labeled with a different fluorophore.

Minisequencing products were separated and detected by capillary electrophoresis, followed by automated genotyping. Homozygous wild-type or mutant DNA samples produced electropherograms containing only a single colored peak for each mutation site, whereas samples heterozygous for a specific mutation displayed two different-colored peaks for that mutation site. The described semiautomated multiplex minisequencing assay can detect the most common Southeast Asian and Indian ß-thalassemia mutations, is amenable to high-throughput scale up, and may bring population-based screening of ß-thalassemia in endemic regions a step closer to implementation.

(Rochette et al., 2008) studied the phenotype of abnormal hemoglobin in the β 41/42 frameshift heterozygosity. He observed that β c.126–129del CTTT frameshift mutation (exon1) resulting in a truncated β chain (58 AA) in a heterozygous patient. The Hb variant was isolated from the unstable component. He studied truncated β globin chain in order to understand the instability of the abnormal hemoglobin. He has built a 3D model of the abnormal variant. Using biochemistry techniques and molecular modeling it was shown:

- The synthetic 58 amino acids truncated protein was unable to fold in physiological conditions. Sequence from AA 41 to AA 53 is structured as a primary helix but fully buried and unable to be exposed to the solvent.
- The Phe42Ser substitution is responsible for the major structural defect.
- Key residues in position β 42, 44, 45, 48, 49, 51, 54, 55 and 57 are very sensitive for stability of hemoglobin as most of the substitutions for these residues disturb subunit packing in flexible joint probably leading to dissociation at β globin gene with this 4 bp deletion encodes the α1-β1 interface.

(Doss C et al., 2009) used computational methods to analyze the genetic variations that can alter the expression and function of the HBB gene. She applied an evolutionary perspective to screen the SNPs using a sequence homology-based SIFT tool, which suggested that 210 (90%) non-synonymous (ns) SNPs were found to be deleterious. The structure-based approach PolyPhen server suggested that 134 (57%) ns SNPS may disrupt protein function and structure. The Pupa Suite tool predicted the phenotypic effect of SNPs on the structure and function of the affected protein. Structure analysis was carried out with the major mutation that occurred in the native protein coded by the HBB gene in HbC, HbD, HbE and HbS. The amino acid residues in the native and mutant modelled protein were further analysed for solvent accessibility, and secondary structure to check the stability of the proteins. The functional analysis presented here may be a good model for further research.

CHAPTER 3

MATERIAL AND METHODS

3.1 Sample collection

List of the patients suffering from beta-thalassemia from whom blood samples were collected:

Table 3.1 List of Samples

SERIAL NO.	NAME	PLACE
1.	SANJAY	DHARAMSHALA
2.	NIKITA	DHARAMSHALA
3.	ANJANA DEVI	DHARAMSHALA
4.	SNEH	DHARAMSHALA
5.	SHUBHAM(S/0 GYAN CHAND)	DHARAMSHALA
6.	SAHIL	DHARAMSHALA
7.	SHUBHAM(S/O ANUJ)	DHARAMSHALA
8.	SHIVAM	DHARAMSHALA
9.	NEHA	DHARAMSHALA
10.	SATNAM	DHARAMSHALA
11.	GUDIYA	DHARAMSHALA
12.	VISHAL	DHARAMSHALA

Table 3.2 Previous Samples

SERIAL NO.	NAME	PLACE
2.	SAMAJH DHIR	CHANDIGARH
4.	-	CHANDIGARH
6.	-	CHANDIGARH
7.	-	CHANDIGARH
8.	RISHABH CHOPRA	CHANDIGARH

3.2 Genomic DNA Isolation

Phenol-Chloroform DNA isolation protocol was followed.

3.2.1 *Materials required*:

i. ***Buffer A*** (100ml) autoclave the solution.

Table 3.3 Composition of Buffer A

Nacl(10mM)	10ml of 100m M
TrisCl(50m M)	5ml of 1M
EDTA(10m M)	1ml of 1 M
Water	84ml

ii. ***Buffer B*** (100ml) do not autoclave.

Table 3.4 Composition of Buffer B

SDS 3%	3gm
Nacl(10mM)	10ml of 100m M
TrisCl(50m M)	5ml of 1M
EDTA(10m M)	1ml of 1 M

1ml Buffer A + 1ml SDS (3%)

iii. ***ACD Solution*** (100ml)

Table 3.5 Composition of ACD Solution

Citric acid	0.48gm
Sodium citrate	1.32gm
Glucose	1.47gm

Dissolve in Milli Q water and filter sterilize it.

iv. Solution 1:

Phenol (Tris Equilibrated) – Chloroform – Isoamyl alcohol (25:24:1)

i. Phenol (Tris equiliberated):

- Phenol was kept at 65°C (till it becomes liquid).
- Equal volume of 1M Tris (ph-8) and phenol was taken kept overnite.
- Then the top white layer was taken and 0.1 M Tris (pH 8) was added and kept overnite.
- White layer was removed and rest was equiliberated phenol.
- Volume of equiliberated phenol was ~ 80ml.

ii. 76.8 ml Choloroform was added.

iii. 3.2ml Isoamyl alcohol was added.

v. ***TE Buffer(1M)***

Table 3.6 Composition of TE Buffer (100ml)

0.2 M Tris base	50ml
0.1M EDTA	50ml

vi. ***TAE Buffer:***

50X Stock solution of TAE was prepared by adding the following:

Table 3.7 Composition of TAE Buffer

Tris base	24.2gm
Glacial acetic acid	5.71ml
EDTA(0.5M, pH 8.0)	10ml
Distilled water	Make up the vol. to 100ml

vii. ***Gel loading dye:*** (6X)

Table 3.8 Composition of Gel loading dye

Bromophenol Blue	0.25%(w/v)
Xylene Cyanol	0.25%(w/v)
Glycerol in DW	30%(v/v)

3.2.2 *DNA Extraction Protocol*

i. 500µl of blood sample was taken in sterile 2ml eppendorf tube and 150 µl of ACD solution was added.

ii. 250µl of buffer A and 300 µl of buffer B were added and samples were incubated at 37°C for 1-½ hour.

iii. Proteinase K was added to a final concentration of 100µg/ml and mixed properly by tapping and samples were incubated at 50°C for another 45 mins. Intermittent mixing was done.

iv. Then equal volume of Solution I was added and mixed properly by inverting the tube and centrifuged at 10,000 rpm for 10 mins at 4°C.

v. The aqueous phase was then transferred to a fresh eppendorf.

vi. If the solution shows turbidity then above steps should be repeated once or twice depending upon clarity of the solution.

vii. After centrifugation 900µl of ethanol was added and mixed properly and then kept at -20° C for overnight precipitation.

viii. After incubation, centrifugation was done at 10,000rpm for 10 mins at 4°C and pellet was retained.

ix. Then the pellet obtained was dislodged using 600µl of 70% ethanol and centrifuged for 15 mins at 4°C.

x. After that pellet was dried and resuspended in 20-40µl TE buffer.

xi Genomic DNA was checked in 0.8% agarose gel, 1X TAE.

3.3 PRIMER DESIGNING

i. Complete sequence of β globin gene was downloaded from NCBI database.

ii. The sequences were put in the Primer 3 software for primer designing.

iii. Total length of gene sequence was about 2kb.

iv. For sequencing purpose, the sequence was broken down into 2 parts, 1kb each.

Table 3.9 Primer Sequences

PRIMER-1	F R	TTTGCAGCCTCACCTTCTTT TTGGAATATATGTGTGCTTATTTGC
PRIMER-2	F R	GCAAATAAGCACACATATATTCCAA ACTCCTAAGCCAGTGCCAGA
IVSII nt 654	F R	CTTAAACCTGTCTTGTAACCTTGGTA GCAACCTCAAACAGACACCA
Codon 0	F R	CTGGGCATAAAAGTCAGGG GGCAGAGAGAGTCAGTGCCTA
Codon 43	F R	CATGTGGAGACAGAGAAGAC TCATTCGTCTGTTTCCCATTCTAAAC
Pri (-28)	F R	AAAGTCAGGGCAGAGCCATCTATTGC GCTTACCAAGCTGTGATTCC
Pri (-29)	F R	ATCACTTAGACCTCACCCTGTGGAGCCA GAACACAGTTGTGTCAGAAG

3.3.1 PRIMERS FOR THE SEQUENCE 1-1040 bp

OLIGO	start	len	tm	gc%	any	3' seq
LEFT PRIMER	19	20	59.99	45.00	4.00	0.00
	TTTGCAGCCTCACCTTCTTT					
RIGHT PRIMER	941	25	59.37	32.00	6.00	2.00
	TTGGAATATATGTGTGCTTATTTGC					

SEQUENCE SIZE: 1040

INCLUDED REGION SIZE: 1040

PRODUCT SIZE: 923, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 0.00

1GCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTTAAGATATAGTGT
>>>>>>>>>>>>>>>>>>>>

61 ATTTTCCCAAGGTTTGAACTAGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCC

121 ACATTCCCTTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATA

841 TTTAACCCATAAATATGTATAATGATTATGTATCAATTAAAAATAAAAGAAAATAAAGTA

901 GGGAGATTATGAATATGCAAATAAGCACACATATATTCCAAATAGTAATGTACTAGGCAG
<<<<<<<<<<<<<<<<<<<<<<<<<

961 ACTGTGTAAAGTTTTTTTTTAAGTTACTTAATGTATCTCAGAGATATTTCCTTTTGTTAT

1021 ACACAATGTTAAGGCATTAA

3.3.2 PRIMERS FOR THE SEQUENCE 800-2024 bp

OLIGO	start	len	tm	gc%	any	3' seq
LEFT PRIMER	117	25	59.37	32.00	6.00	2.00
	GCAAATAAGCACACATATATTCCAA					
RIGHT PRIMER	1137	20	60.01	55.00	3.00	1.00
	ACTCCTAAGCCAGTGCCAGA					

SEQUENCE SIZE: 1224

INCLUDED REGION SIZE: 1224

PRODUCT SIZE: 1021, PAIR ANY COMPL: 4.00, PAIR 3' COMPL: 0.00

1TGATTTGGTCAATATGTGTACACATATTAAAACATTACACTTTAACCCATAAATATGTAT

61 AATGATTATGTATCAATTAAAAATAAAAGAAAATAAAGTAGGGAGATTATGAATATGCAA

>>>>

121ATAAGCACACATATATTCCAAATAGTAATGTACTAGGCAGACTGTGTAAAGTTTTTTTTT

>>>>>>>>>>>>>>>>>>>>>>>>>>

181 AAGTTACTTAATGTATCTCAGAGATATTTCCTTTTGTTATACACAATGTTAAGGCATTAA

1021 CTGCCCTCCCTGCTCCTGGGAGTAGATTGGCCAACCCTAGGGTGTGGCTCCACAGGGTGA

1081 GGTCTAAGTGATGACAGCCGTACCTGTCCTTGGCTCTTCTGGCACTGGCTTAGGAGTTGG

<<<<<<<<<<<<<<<<<<<<<<

1141 ACTTCAAACCCTCAGCCCTCCCTCTAAGATATATCTCTTGGCCCCATACCATCAGTACAA

1201 ATTGCTACTAAAAACATCCTCCTT

3.4 POLYMERASE CHAIN REACTION

Table 3.10 -PCR Master Mix

MATERIALS	1X
Template DNA (50 ng)	0.5 µl
Primers	0.5 µl
10X PCR Buffer(with Mgcl2)	1.5 µl
10mM DNTPs	0.3 µl
Autoclaved Water	10.57 µl
Taq DNA Polymerase	0.23 µl
TOTAL	15 µl

3.4.1 *PCR Protocol* :

i. To perform several parallel reactions, master mix was prepared containing autoclaved distilled water, buffer, dNTPs, primers and Taq DNA polymerase in a single tube, and aliquoted into individual tubes. Template DNA solutions was added.

ii. The solutions were gently vortexed and centrifuged after thawing before addition.

iii. Master mix was added, in a thin-walled PCR tube, on ice.

iv. Samples were vortexed again in order to collect all drops from walls of tube.

v. Samples were the placed in a thermocycler PCR was started.

vi. Products were resolved using 1X TAE, 1.8% agarose gel.

3.4.2 *PCR Cycle*

Table 3.11 PCR Cycle

SERIAL NO.	STAGE	STEP	TEMPERATURE	TIME
1.	STAGE 1 *1 CYCLE		94° C	3mins
2.	STAGE 2 *30 CYCLES	**STEP 1**	94° C	30 sec
		STEP 2	GRADIENT	1 min
		STEP 3	72° C	1min
3.	STAGE 3 *1 CYCLE		72° C	7mins
4.	STAGE 4		4° C	∞

3.5 AGAROSE GEL ELECTROPHORESIS

Genomic DNA Was resolved using 0.8% agarose gel and for resolving PCR products 1.8% agarose gel was used.

3.5.1 Protocol

i. 0.8g of agarose for DNA and 2g for PCR taken in a flask and add 100mL of 1X TAE.

ii. Mixture was heated in the microwave for a few minutes.

iii. Gel was cooled to about 60°C and 3µL of EtBr was added.

iv. After cooling gel was casted in a pre-cleaned gel tray and allowed to solidify.

v. 5µL of DNA sample was taken on a parafilm and mixed with 3µL of 6X loading dye.

vi. Then DNA was loaded in the wells.

vii. 1kb DNA ladder was also loaded as marker.

viii. Gel was electrophoresed at 100V for 1-2 hrs.

3.6 ELECTROPHORETIC SEPARATION OF THE PCR PRODUCTS AND PURIFICATION OF AMPLIFIED PRODUCTS

Amplified products from the PCR reactions are purified using Millipore's Montage 96 filter plate as per the manufacturer's protocol (this is performed to remove the unused dNTPs and salts and unused primers), and elute with 50 µl of sterile nuclease-free dd water following a 30 minute room temperature incubation. Purified PCR products are quantified by A260 readings using a Perkin Elmer (100 µl at 1:25 dilution). PCR products are stored at -20°C until further sequencing.

3.7 SEQUENCING OF PCR PRODUCTS

The purified PCR products generated from genomic DNA template will be sequenced from both ends, using the Forward and Reverse primers using Sangers enzymatic method. The sequencing of beta-globin gene in different genotypes will done by Xceleris life sciences with the help of capillary sequencer with ET terminator chemistry

3.8 PROTEIN MODELING

After getting sequencing results of the previously amplified Chandigarh samples, Protein Modeling is done to find out the binding energy change in the normal and mutated protein of their respective oxygenated and deoxygenated states. The X-ray structure of normal hemoglobin is downloaded from the protein data base and used as template structure to build the mutated hemoglobin structure. The point mutations are introduced in the structure by substituting the mutated amino acid in the pdb file using molecular builder (Maestro, Schrödinger Inc.). The protein structures are prepared and energy minimized using Protein preparation wizard.

3.8.1 Computational Details

Density functional calculations with the BP86 and B3LYP functional were performed with the Jaguar program. Two different basic sets were used. Most calculations were carried out with basis set I, which corresponds to that labeled as LACVP ** in the program. This implies an effective core potential replacing the 10 innermost electrons of iron. The basis set is valence double- ζ for all atoms and has a polarization shell for all atoms different from iron. An additional set of calculations was carried out with the larger basis set II, which is that labeled as LACV3P**++ in the program. This is a valence triple-ζ basis set. It includes diffuse functions for all atoms and it retains the same polarization shells of basis set I. for the oxygenated systems, basis set I contains 1342 basis functions and basis set II, 2052.

Geometry optimizations were carried out with the BP86 functional and basis set I (BP86/I description). Most of the discussion on energetic is nevertheless based on single-point calculations with the B3LYP functional on the optimized BP86 geometries. The problem of the absolute energetics of the oxygen fixation is a difficult one from a computational viewpoint because of the involvement of different spin states. The energy comparison between different spin states is a challenge for DFT methods, and B3LYP calculations on BP86 geometries seem to provide a reasonable balance. As will be shown below, the BP86 results suggest larger binding energies, but the trends between the different systems considered, which the key issue in this article is, are practically unchanged. The calculations used the unrestricted formalism for the quintet deoxygenated state, with the same ground state described in previous work. Spin contamination was always small. For the singlet oxygenated state, a restricted formalism was used.

Most of the energies presented are potential energies in the gas phase. An additional set of calculations was carried out introducing environmental effects through continuum calculations with Jaguar's Poisson-Boltzmann solver. The dielectric constant is set to 4.0 and the probe radius to 1.40. Zero point energies and entropic contributions have not been evaluated. The second derivative calculations would be very time consuming, and we think it is reasonable to assume that the additional contribution would be similar for all computed systems. In almost all cases, geometry optimizations were complete except for the frozen chain atoms. In the calculation of the subunit β in the T state, the dihedral angle of the acidic hydrogen of one of the propionic substituent had to be frozen to avoid the formation of a spurious hydrogen bond with the distal histidine.

3.8.2 Computational model

The model system used in these calculations consists of full heme group with all its substituent's, and the four closest amino acids to the iron atom, His E7, Val E11, Phe CD1 for the distal side, and His F8 for the proximal side. The latter is bound to the iron through its ε nitrogen, and is trans to the coordination site of dioxygen. Its introduction is mandatory in any simple representation of the active center of hemoglobin, and it has been so considered in most previous calculations on the system.

The other three amino acids are in the so called distal side of the heme, where dioxygen binds to iron. The importance of this distal side of hemoglobin in dioxygen fixation is well established from experimental studies on mutant systems, and from molecular dynamics calculations. The three amino acids that have been considered are the same in both the α and β subunits. They are sensitive to oxygen fixation because they change notably their positions in the X-ray structure of T and R forms of hemoglobin. Furthermore, it is well known that the distal histidine E7 is able to form a hydrogen bond with dioxygen especially in the α chain.

The structural features of different subunits and states of hemoglobin are introduced in the calculations by constraining the coordinates of selected atoms to the value provided in the Protein Data Bank (PDB). The structures taken from the PDB are that with code 2HHB for the deoxygenated T form and the labeled as 1HHO for the oxygenated R form. The computational approach consisted of freezing the α carbons of the amino acids to the position they have in the PDB structure. The peptidic bonds connecting these α carbons to the rest of protein chains have been suppressed in the calculations, and their valence shell has been saturated with two hydrogen atoms, which have also been kept at same orientation the C-C and the C-N bonds have in the PDB structure.

The three C-H bond distances were frozen to 1.090 Å. There are thus four frozen atoms for each amino acid, as well as the full heme group (except for the dihedral angle of the propionate mentioned above) and the dioxygen, when present, are completely optimized without any additional constraints in the calculation. The propionate groups attached to the heme have been protonated to keep them in neutral form. We could not introduced the amino acids neutralizing them in our calculations for a reason of size and we assume that this aspect will not be critical for oxygen affinity.

Available X-ray PDB structures for the two possible states, R and T, and the two different subunits, α and β were considered. This gives four different systems was evaluated by introducing additional calculations of "non-PDB" systems where dioxygen was added for T states, or removed for R staets. The oxygen binding energy was estimated by using the simple equation:

$$\Delta E = E_{hexa} - (E_{penta} + E_{O2})$$

The geometry optimization were carried out on a total of eight structures: RαD, RαO, TαD,TαO, RβD, TβD and TβO. An additional pair of the calculations, in the presence and absence of dioxygen were carried out on a simplified system with no distal amino acids, just an imidazole as the proximal ligand and the substituents of the porphyrin replaced by hydrogen atoms.

CHAPTER 4

RESULTS

4.1 DNA Isolation

Samples were collected from the hospital of Dharamshala (H.P.). Standard DNA isolation protocol was followed. Further modifications were done for better results. At the end 12 DNA samples had on average a concentration of 50-100ng/ul, which was used for further analysis.

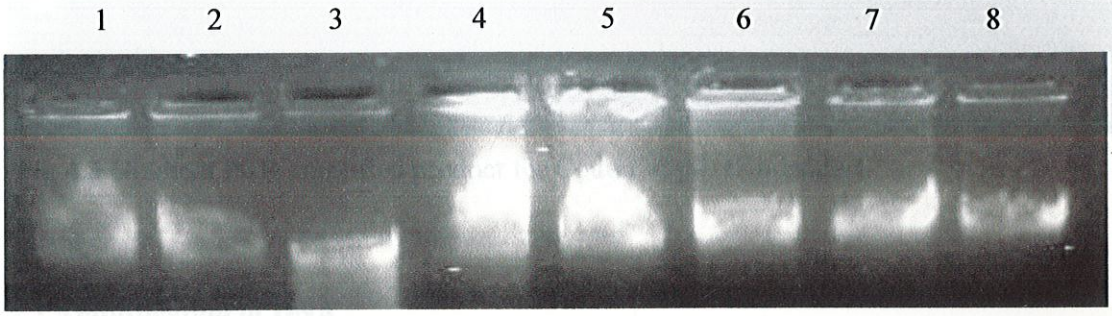

Fig 4.1 DNA Isolated from diseased blood samples

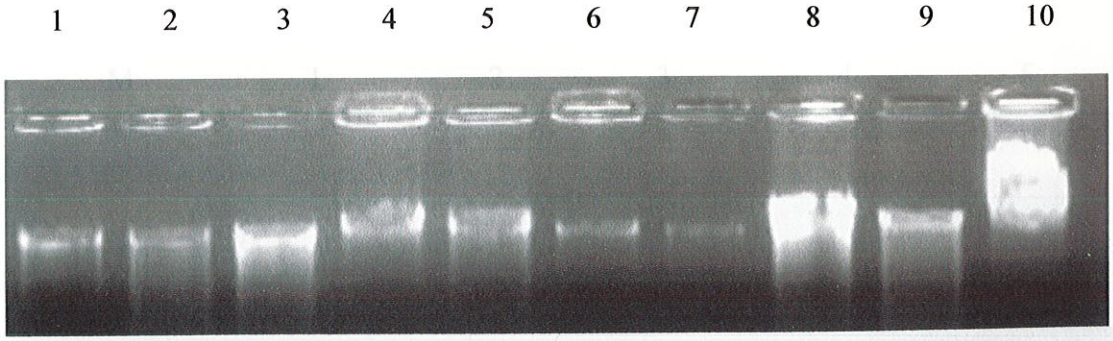

Fig 4.2 DNA Isolation from diseased blood samples

4.2 Annealing temperature determination

To amplify the specific DNA fragments from whole genome sequence tests were carried out to determine the exact annealing temperature to increase specificity.

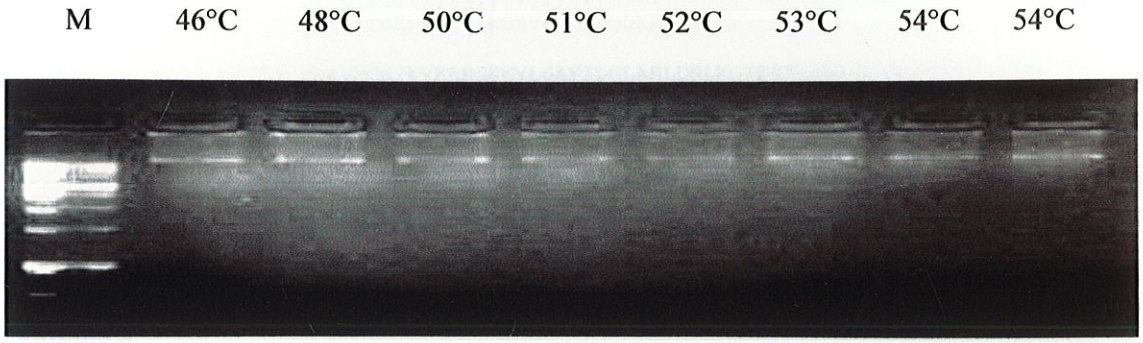

Fig 4.3 Gradient PCR amplified product for Codon 43(M-1Kb ladder)

4.3 Amplification of DNA

Amplification of specific DNA fragments from whole genome sequence was carried out at annealing temperature determined with specific primers.

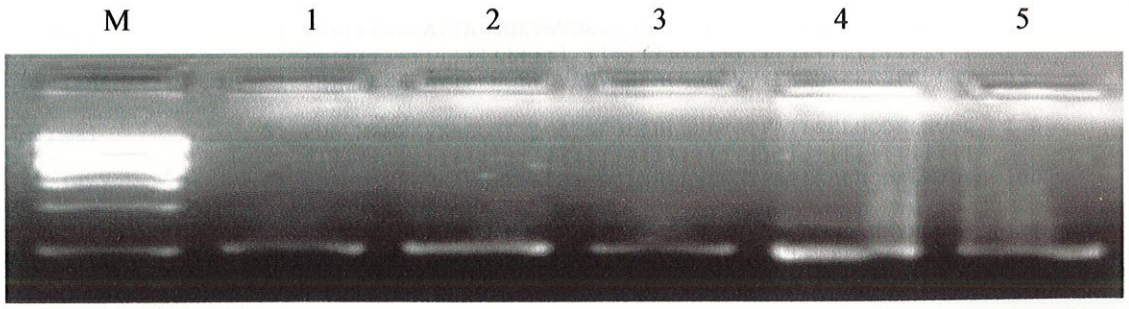

Fig. 4.4 Amplification of DNA with PRIMER 1 (M-1Kb ladder)

4.4 Alignment Details

Sample 2

```
Normal      1 VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLST  50
              |||||||||||||||||:|||||||||.||||||||||||||||||||
Patient     1 VHLTPEEKSAVTALWGKVDVDEVGGEALGSLLVVYPWTQRFFESFGDLST  50

Normal     51 PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDP 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Patient    51 PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDP 100

Normal    101 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      146
              ||||||||||||||||||||||||||||||||||||||||||
Patient   101 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      146
```

Sample 4

```
Normal      1 VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLST  50
              ||||||||||||||||||:|||||||||.||||||||||||||||||||
Patient     1 VHLTPEEKSAVTALWGKVDVDEVGGEALGSLLVVYPWTQRFFESFGDLST  50

Normal     51 PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDP 100
              |||||||||||||||||||||||.||||||||||||||||||||||||||
Patient    51 PDAVMGNPKVKAHGKKVLGAFSDGQAHLDNLKGTFATLSELHCDKLHVDP 100

Normal    101 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      146
              ||||||||||||||||||||||||||||||||||||||||||
Patient   101 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      146
```

Sample 6

```
Normal      1 VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLST  50
              ||||||||||||||||||:||||||||||||||||||||||||||||||
Patient     1 VHLTPEEKSAVTALWGKVDVDEVGGEALGRLLVVYPWTQRFFESFGDLST  50

Normal     51 PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDP 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Patient    51 PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDP 100

Normal    101 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      146
              |||||||||||||||.||||||||||||||||||||||||||
Patient   101 ENFRLLGNVLVCVLAHQFGKEFTPPVQAAYQKVVAGVANALAHKYH      146
```

Sample 7

```
Normal          1 VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLST  50
                  |||||||...|.||||||..||||.|||||||.||::|||||||||||||||
Patient         1 VHLTPEGNFAFTALWGNFNVDENGGEALGFLLLIYPWTQRFFESFGDLST  50

Normal         51 PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDP  100
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Patient        51 PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDP  100

Normal        101 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH       146
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
Patient       101 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH       146
```

Sample 8

```
Normal          1 VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLST   50
                  |||||||...|.||||||..||||.|||||||.||::|||||||||||||||.
Patient         1 VHLTPEGNFAFTALWGNFNVDENGGEALGFLLLIYPWTQRFFESFGDLSN   50

Normal         51 PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDP   100
                  ||||||||||||||||||:||||||||||||||||||||||.||||||
Patient        51 PDAVMGNPKVKAHGKKVLGAYSDGLAHLDNLKGTFATLSELHCAKLHVDP   100

Normal        101 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH       146
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
Patient       101 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH       146
```

Fig 4.5 Alignment Details of protein sequences from normal and mutated β globin gene.

4.5 SNPs detected in previous samples

Table 4.1 SNPs Detected

Patient No.	Amino acid position	Normal	Mutated
Patient 2			
	19	Asn	Asp
	30	Arg	Ser
Patient 4			
	19	Asn	Asp
	30	Arg	Ser
	75	Leu	Gln
Patient 6			
	19	Asn	Asp
	117	His	Gln
Patient 7			
	7	Glu	Gly
	8	Lys	Asn
	9	Ser	Phe
	11	Val	Phe
	17	Lys	Asn
	18	Val	Phe
	23	Val	Asn
	30	Arg	Phe
	33	Val	Leu
	34	Val	Ile
Patient 8			
	7	Glu	Gly
	8	Lys	Asn
	9	Ser	Phe
	11	Val	Phe
	17	Lys	Asn
	18	Val	Phe
	23	Val	Asn
	30	Arg	Phe
	33	Val	Leu
	34	Val	Ile
	50	Thr	Asn
	71	Phe	Tyr
	94	Asp	Ala

In patient no. 2 asparagine was replaced by aspartic acid at 19th position and arginine was replaced by serine at 30th position of beta globin chain. In patient no. 3 asparagine was replaced by aspartic acid at 19th position, arginine was replaced by serine at 30th position and leucine was replaced by glutamine at 75th position of beta globin chain. In patient no. 6 asparagine was replaced by aspartic acid at 19th position and histidine was replaced by glutamine at 117th position of beta globin chain. In patient no. 7 glutamine was replaced by glycine at 7th position, lysine was replaced by asparagine at 8th, serine was replaced by phenylalanine at 9th position, valine was replaced by phenylalanine at 11th, lysine was replaced by asparagine at 17th position, valine was replaced by phenylalanine at 18th position, valine was replaced by asparagine at 23th position, arginine was replaced by serine at 30th position of beta globin chain. In patient no. 3 asparagine was replaced by aspartic acid at 19th position, arginine was replaced by serine at 30th position and leucine was replaced by glutamine at 75th position of beta globin chain. In patient no. 6 asparagine was replaced by aspartic acid at 19th position and histidine was replaced by glutamine at 117th position of beta globin chain. In patient no. 7 glutamine was replaced by glycine at 7th position, lysine was replaced by asparagine at 8th, serine was replaced by phenylalanine at 9th position, valine was replaced by phenylalanine at 11th, lysine was replaced by asparagine at 17th position, valine was replaced by phenylalanine at 18th position, valine was replaced by asparagine at 23th position, arginine was replaced by phenylalanine at 30th position , valine was replaced by leucine at 33rd position, valine was replaced by isoleucine at 34th position of beta globin chain. In patient no. 8 glutamine was replaced by glycine at 7th position and lysine was replaced by asparagine at 8th and serine was replaced by phenylalanine at 9th position and valine was replaced by phenylalanine at 11th, lysine was replaced by asparagine at 17th position, valine was replaced by phenylalanine at 18th position, valine was replaced by asparagine at 23th position, arginine was replaced by phenylalanine at 30th position , valine was replaced by leucine at 33rd position , valine was replaced by isoleucine at 34th, threonine was replaced by asparagine at 50th position, phenylalanine was replaced by tyrosine at 71st position and aspartate was replaced by alanine at 94thst position of beta globin chain.

4.5 Binding Energies obtained after running JAGUAR

$\Delta BE = (E_{oxy}-(E_{deoxy}+E_{O2}))$

Table 4.2 Binding Energies

Sample No.		E_{oxy}	E_{deoxy}	ΔBE
2	Normal	-3039	-2794	-295
	Mutated	-2862	-2643	-269
4	Normal	-3417	-3024	-443
	Mutated	-3324	-2206	-1168
6	Normal	-2815	-2769	-96
	Mutated	-2989	-2776	-263
7	Normal	-5913	-5642	-321
	Mutated	-6209	-6056	-203
8	Normal	-7192	-6898	-344
	Mutated	-7433	-7167	-316

The energy of the optimized structure of oxygenated normal protein obtained is -3039 and that of optimized structure of deoxygenated normal protein is -2794. So the calculated binding energy of the normal protein of the sample 2 is -269. Also the energy of the optimized structure of oxygenated mutated protein obtained is -2862 and that of optimized structure of deoxygenated mutated protein is -2643. So the calculated binding energy mutated protein of the sample 2 is -269. Similarly for sample 4 the energy of the optimized structure of oxygenated normal protein obtained is -3417 and that of optimized structure of deoxygenated normal protein is -3024.

So the calculated binding energy of the normal protein of the sample 4 is -443.Also the energy of the optimized structure of oxygenated mutated protein obtained is -3324 and that of optimized structure of deoxygenated mutated protein is -2206. So the calculated binding energy mutated protein of the sample 4 is -1168. Similarly for sample 6 the energy of the optimized structure of oxygenated normal protein obtained is -2815 and that of optimized structure of deoxygenated normal protein is -2769.So the calculated binding energy of the normal protein of the sample 6 is-96. Also the energy of the optimized structure of oxygenated mutated protein obtained is -2989 and that of optimized structure of deoxygenated mutated protein is -2776. So the calculated binding energy mutated protein of the sample 6 is -263. Similarly for sample 7 the energy of the optimized structure of oxygenated normal protein obtained is -5913 and that of optimized structure of deoxygenated normal protein is -5642. So the calculated binding energy of the normal protein of the sample 7 is -321. Also the energy of the optimized structure of oxygenated mutated protein obtained is -6209 and that of optimized structure of deoxygenated mutated protein is -6056. So the calculated binding energy mutated protein of the sample 7 is -203. Similarly for sample 8 the energy of the optimized structure of oxygenated normal protein obtained is -7192 and that of optimized structure of deoxygenated normal protein is-6898. So the calculated binding energy of the normal protein of the sample 8 is -344. Also the energy of the optimized structure of oxygenated mutated protein obtained is -7433 and that of optimized structure of deoxygenated mutated protein is -7167. So the calculated binding energy mutated protein of the sample 8 is -316.

4.6 Structures obtained after running JAGUAR Program

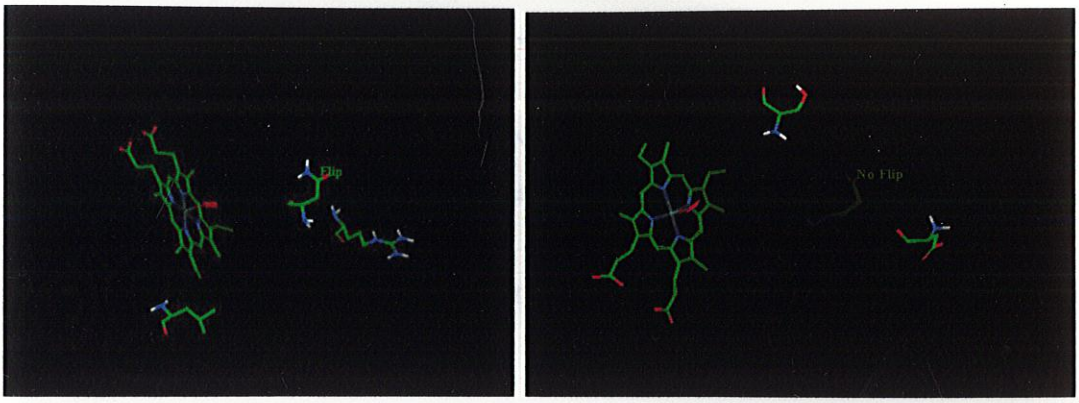

Fig 4.6 Structures of Normal and Mutated 1HOO Molecule

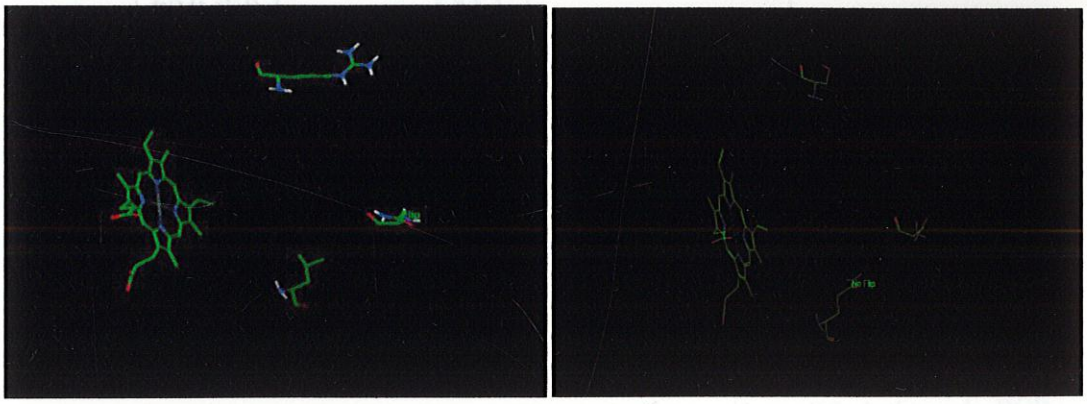

Fig 4.7 Structures of Normal and Mutated 2HHB Molecule

CHAPTER 5

DISCUSSION

In this project, selective amplification of a DNA fragment of β globin gene with specific oligonucleotide primers has been done for the identification of novel SNPs present in the thalassemic patients of Himachal Pradesh. DNA was isolated from the samples with the help of the standard protocol. Primers were designed for the whole β globin gene. Primers were also designed for DNA fragments of mutated β globin gene. Annealing temperatures were determined experimentally by PCR for respective primers. Following are the annealing temperatures for primers:-

Table 5.1 Annealing Temperature of Primers

Serial no.	Name	Temp.
1.	PRIMER 1	55.5° C
2.	PRIMER 2	60° C
3.	IVSII nt 654	56° C
4.	Codon 0	57.6° C
5.	Codon 43	56° C
6.	Pri (-28)	58° C
7.	Heme	58° C

Using these primers DNA fragments for 12 samples were amplified through PCR. Samples amplified with the help of primers and have been sent for sequencing. The discovery of large numbers of single nucleotide polymorphisms (SNPs) in genome-scale sequencing initiatives opens new avenues to the study of the genome-wide distribution of diversity and gene based study and its significance. Generally SNPs are highly abundant but their density differs substantially in different regions of a genome and from genome to genome in humans and may be effectively used as markers (Raymond *et al.*, 1999).

With the recent information available and background, the experiments were carried out for the identification of Single Nucleotide Polymorphisms (SNPs) within the gene of interest to study the natural variations at a single nucleotide level which may influence to ameliorate the quality of the product.

Besides doing the amplification of above mentioned samples, work was also done on the previously found SNPs from Chandigarh region to find out the oxygen binding affinity of normal and mutated hemoglobin. The oxygen affinity has been evaluated by computing the energy difference between optimized structure of oxygenate and deoxygenated forms of normal as well as mutated protein. For doing this, the Protein Modeling of the mutated sequence and the binding energy calculations were performed using JAGUAR program. The calculations in this case have been carried out with basis set I which corresponds to LACVP program. The structural feature of β subunit and oxygenated and deoxygenated states of hemoglobin has been introduced in the calculations by constraining the coordinates of selected atoms to the value provided in PDB.

The oxygen binding energy has been calculated by:-

$$\Delta E = (E_{oxy}-(E_{deoxy}+E_{O2}))$$

Therefore, observing the change in binding energy of both the mutated and normal proteins it was found that:-

- The oxygen affinity of beta subunit is different in case of oxygenated and deoxygenated states.
- The binding energy of oxygen is affected because of presence of SNPs in the beta globin gene.

CHAPTER 6

CONCLUSION AND FUTURE PROSPECTS

β-Thalassemia is a group of inherited diseases characterized by the abnormality of β globin production. More than 90 types of mutation have been reported. The study of these genetic mutations are based on polymerase chain reaction (PCR) carried out with specific primers. Early comprehensive treatment has changed thalassemia from a fatal pediatric disease to one in which patients live productive lives throughout adulthood. Advances in treatment are exciting, resulting in the potential for cure and improved quality of life. Doctors diagnose thalassemia using blood tests, including a complete blood count (CBC) and special hemoglobin tests. Tests on the amount of iron in the blood to find out whether the anemia is due to iron deficiency or thalassemia. The present study was aimed to identify *de novo* Single Nucleotide polymorphisms (SNP) in the beta-globin genes and to find out the effect of previously found SNPs on the oxygen binding affinity of hemoglobin molecule. The amplified product has been sent for sequencing and the oxygen affinity has been evaluated by computing the energy difference between optimized structure of oxygenate and deoxygenated forms of normal as well as mutated protein. Therefore, the change in binding energies clarifies that oxygen binding affinity of hemoglobin molecule is also affected due to the presence of SNPs in thalassemic patients. Common and novel SNPs may then directly be taken up for the thalassemia diagnosis and to understand its effect on the severity of the disease.

BIBLIOGRAPHY

1. Ali Taher, Hussain Isma'eel and Maria D. Cappellini(2006): Thalassemia intermedia: *Blood Cells, Molecules, and Diseases Volume 37, Issue 1,Pages 12-20*

2.A.Kuliev,S.Rechitsky,O.Verlinsky,V.Ivakhenko,S.Evsikov,G.Wolf,M.Angastiniotis,D. Georghiou,V.Kukharenk,C.Strom(1998): Preimplantation Diagnosis of Thalassemias *Journal of Assisted Reproduction and Genetics, Vol. 15, No. 5.*

3. Agarwal S, Gupta A, Gupta UR, (2003): Prenatal diagnosis in beta thalassemia: an Indian experience. *Fetal Diagn Ther* 18:328–332.

4. Angastiniotis M, Modell B, Boulinzhenkov V.(1995): Prevention and control of haemoglobinopathies. *Bul WHO;* 73:375-386.

5.Borgna-Pignatti C, Vergine G, Lombardo T, Cappellini MD, Cianciulli P, Maggio A, Renda D, Lai ME, Mandas A, Forni G, Piga A, Bisconte MG(2004) . Hepatocellular carcinoma in the thalassaemia syndromes. *Br J Haematol.* 2004; 124: 114–7. (PubMed)

6.Colah RB, Gorakshakar AC, Lu CY,(1997): Application of covalent reverse dot. blot hybridization for rapid prenatal diagnosis of the common Indian thalassemia syndromes. *Ind J Hematol Blood Transfus* **15**: 10–14.

7.Galanello R, Melis MA, Ruggeri R, Addis M, Scalas MT, Maccioni L, Furbetta M, Angius A, Tuveri T, Cao A.(1979): Beta 0 thalassemia trait in Sardinia. *Hemoglobin* 3: 33–46. (PubMed)

8.G Yershov, V Barsky, A Belgovskiy, E Kirillov, E Kreindlin, I Ivanov, S Parinov, D Guschin, A Drobishev, S Dubiley, A Mirzabekov (1996):DNA analysis and diagnostics on oligonucleotide microchips. *Nat Genet*; 14 (3):264-8

9. Gaziev J, Lucarelli G. (2003): Stem cell transplantation for hemoglobinopathies. *Curr Opin Pediatr.* 15: 24–31. (PubMed)

10. La Nasa G, Argiolu F, Giardini C, Pession A, Fagioli F, Caocci G, Vacca A, De Stefano P, Piras E, Ledda A, Piroddi A, Littera R, Nesci S, Locatelli F.(2005): Unrelated bone marrow transplantation for beta-thalassemia patients: The experience of the Italian Bone Marrow Transplant Group. *Ann N Y Acad Sci.* 1054: 186–95. (PubMed)

11.Locatelli F, Rocha V, Reed W, Bernaudin F, Ertem M, Grafakos S, Brichard B, Li X, Nagler A, Giorgiani G, Haut PR, Brochstein JA, Nugent DJ, Blatt J, Woodard P, Kurtzberg J, Rubin CM, Miniero R, Lutz P, Raja T, Roberts I, Will AM, Yaniv I, Vermylen C, Tannoia N, Garnier F, Ionescu I, Walters MC, Lubin BH, Gluckman E(2003): Related umbilical cord blood transplantation in patients with thalassemia and sickle cell disease. *Blood.* 101: 2137–2143. (PubMed)

12. M C Cheung, J D Goldberg, Y W Kan (2005): Prenatal diagnosis of sickle cell anaemia and thalassaemia by analysis of fetal cells in maternal blood. *N Engl J Med.*15; 353 (11):1135-46

13.Orofino MG, Argiolu F, Sanna MA, Rosatelli MC, Tuveri T, Scalas MT, Badiali M, Cossu P, Puddu R, Lai ME, Cao A.(2003) Fetal HLA typing in beta thalassaemia: implications for haemopoietic stem-cell transplantation. *Lancet.* 362: 41–2. (PubMed)

14.Parag M. Tamhankar, Sarita Agarwal, Vandana Arya, Ravindra Kumar, U. R. Gupta and S. S. Agarwal(2008):Prevention of homozygous beta thalassemia by premarital screening and prenatal diagnosis in India *Wiley InterScience(www.interscience.wiley.com)* 10.1002/pd.2176

15. Poncz M, Solowiejczyk D, Harpel B, (1982): Construction of human gene libraries from small amounts of peripheralblood: analysis of beta-like globin genes. *Hemoglobin* 6(1):27–36.

16.Vivian C. D. Reichert & Simone M. de Castro & Sandrine C. Wagner & Dulcinéia M. de Albuquerque & Mara H. Hutz & Sandra Leistner-Segal (2007):Identification of β thalassemia mutations in South Brazilians. *Springer-Verlag*

17.Vrettou C, Traeger-Synodinos J, Tzetis M, Malamis G, Kanavakis E. (2003): Rapid screening of multiple beta-globin gene mutations by real-time PCR on the Light Cycler: application to carrier screening and prenatal diagnosis of thalassemia syndromes. *Clin Chem.* 2003; 49: 769–76. (PubMed)

18. WHO Scientific Group: Control of hereditary diseases. WHO Technical Report Series 865

Annexure I

List of chemicals

DdH2O (Double Distilled Water)

Sodium Chloride (Merck)

Ethlenediaminetetraacetic Acid (Merck)

Chloroform (Qualigens)

Isoamyl Alcohol (SIGMA)

Hydrochloric Acid (Qualigens)

2-Mercaptoethanol (SIGMA)

Isopropyl Alcohol (Merck)

Ammonium Acetate (SIGMA)

Ethidium Bromide (SIGMA)

Isopropyl Alcohol (Qauligen)

Taq Buffer (SIGMA)

Taq DNA polymerase (SIGMA)

MgCl2 (SIGMA)

Deoxynucleotide set (SIGMA)

Agarose (Merck)

Xylene cyanol FF (Merck)

ANNEXURE II

List of instruments

Microwave MS-283mc (LG)

Vortex (Genei)

Autoclave (NSW, India)

Thermal cycler (Applied Bio system)

Micropipette (eppendorf)

Microtips (Axygen)

Glassware (Borosil)

Gel Doc (Biorad)