Chapter I

Introduction

Identification of individuals is necessary in many social, medical and forensic analyses especially paternity tests, where a man is excluded from parentage of a child when an allele observed in the child could not have been maternally inherited and the allele is lacking in the man. In the past, most laboratory techniques have been based on studies of blood group substances. For example, the ABO blood group system which is the most important human blood group, that can be used to classify people into only four different phenotypes (blood groups A, B, AB, O). The ABO blood group system was discovered by Landsteiner, who suggested that ABO typing of forensic blood stains could be used to help identify associations or exclusions between blood at a crime scene and a suspect.[1-3] This concept of identification was extended to the determination of family relationships.[1-3] Although there are three usable and common alleles, the most frequent one is silent (O); its presence cannot be observed directly by any antiserum. The genotypes of AA and AO (or genotypes BB and BO) persons cannot be differentiated. When a paternal obligatory allele is O, men homozygous for A or B cannot be excluded from paternity because the AO heterozygous could not be distinguished from the homozygous AA individual (or genotypes BO from BB).[3] So matching of ABO types provides only weak statistical evidence for a true association.

Many other human genetic markers based on serology had been explored later. The Rh, MNSs, Kell, Duffy and Kidd blood group systems have been recommended in parentage analysis. Although these systems are polymorphic than the ABO blood group system but there are several

limitations to phenotyping the erythrocytes. Reagents for red cell antigens typing are mostly derived from human sources. Human sera are expensive, sometimes contain unexpected antibodies, and are occasionally unavailable. Rare alleles can cause misinterpretation, especially those that are unobservable (silent) because they require rare antisera to induce agglutination or because they are immunologically unexpressed. For example, the ABO system, a heterozygous biological father with one undetected antigen (O: genotype AO) may be considered homozygous for his detected antigen (phenotype A). If a child inherits an expressed maternal obligatory allele (B: genotype BB) and a silent allele from its father (O), the child will appear to be homozygous (phenotype B) for the allele inherited from the mother, but actually heterozygous (genotype BO). In this case, the Mendelian laws of genetics dictate that the true father must have contributed to the child the gene for the blood group B, and that he must therefore be group B, AB or O, not group A. A false exclusion can be inferred from the misinterpretation of homozygosity of the biological father and child.[3] This false exclusion finding is a well-known pitfall in ABO of red cell systems. A frequently encountered problem is that ambiguous red cell phenotypes can represent several genotypes. An infrequent problem of red cell immunogenetic analysis involves alteration of an inherited antigen. Some antigens can undergo posttranslational change to another antigen such as a conversion of A1 to B in the ABO system. [3]

Many other polymorphisms of serum proteins and red cell enzymes, such as erythrocyte acid phosphatase (EAP) and phosphoglucomutase (PGM), were described and applied to parentage analysis.[3-4] The electrical charge of a protein molecule depends on its amino acid content; variant alleles possess different amino acids and different charges. Electrophoresis

involves separation of proteins according to their rate of migration in an electrical field; migration is primarily a function of each protein's net charge at a fixed pH. This high resolution electrophoretic method (Isoelectricfocusing) requires a gel containing ampholytes that establish a pH gradient.[3,5-8] Proteins are added and migrate until they encounter a region of the gel where the pH equals the isoelectric point of a protein; migration stops at that point. Many protein variants are stable under conditions of transport, storage, analysis and the alleles are codominant and readily observable but the relative instability of some enzymatic proteins can also pose problems. There are known silent alleles and variants that are difficult to differentiate because of similar migration characteristics.[3]

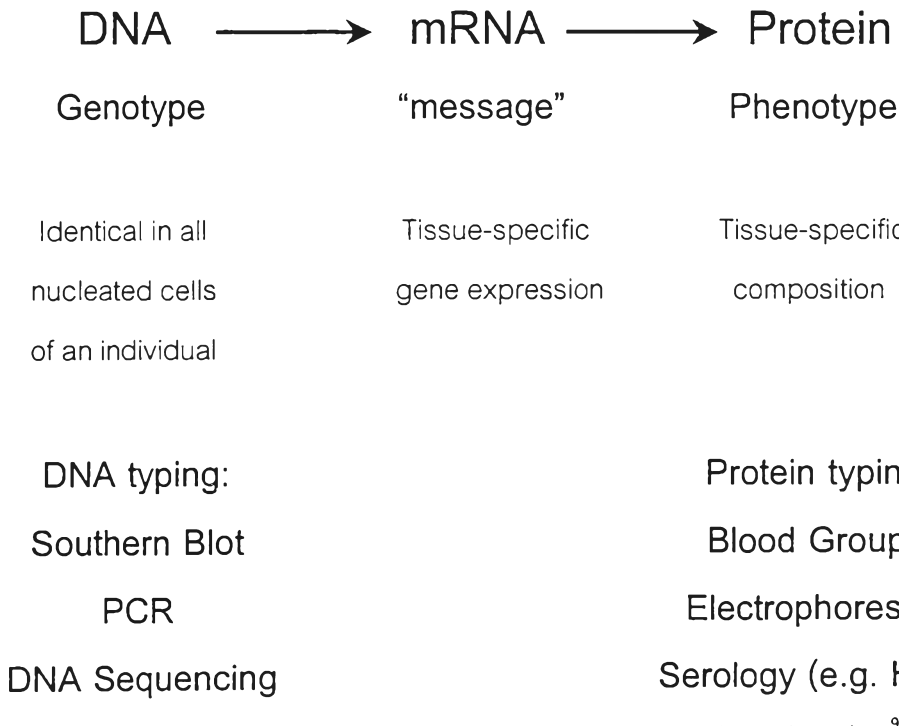

Fig. 1. From genes to proteins - the difference between DNA and protein typing.[9]

These traditional genetic markers make use of variation in biochemical substances controlled by inherited variability in the DNA coding for them (Fig. 1). As soon as the advances in molecular genetics had made direct

exploration of DNA variation feasible, this has led to the development of powerful DNA typing systems in many fields and also in the field of forensic medicine.

The detection of DNA polymorphisms by restriction fragment length polymorphism (RFLP) analysis has become an important aspect of genetic characterization for identity testing. Many RFLP has resulted from point mutation events that create or eliminate restriction enzyme recognition sites (Fig. 2).[10-13] This type of RFLP usually gives rise to only two possible alleles at any given locus; thus, it does not provide a high degree of discrimination for identity testing.

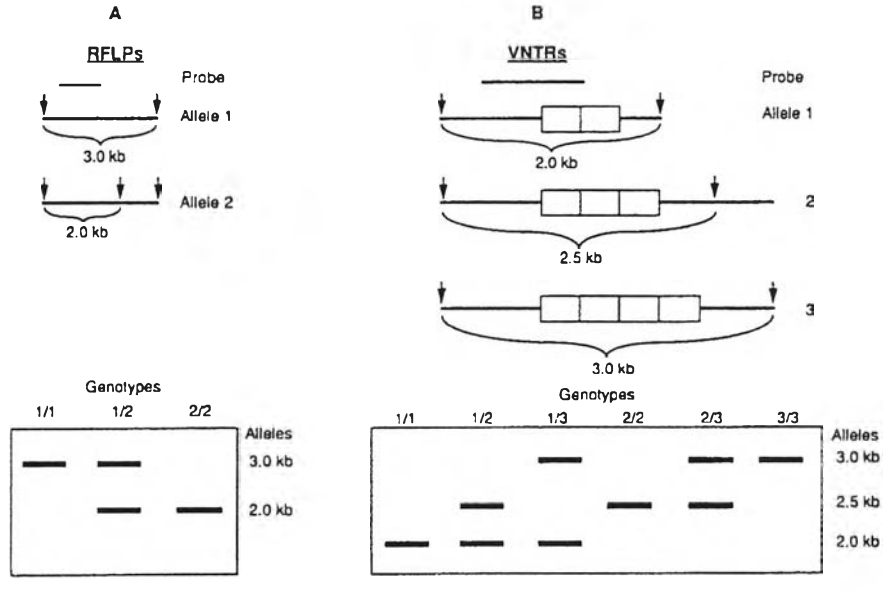

Fig. 2. Two types of RFLP. Structure of alleles in chromosome is diagrammed at the top; arrows indicate sites of cutting by enzyme; lengths of fragments demonstrated by probe (short line above) are given. Electrophoretic patterns are diagrammed below. A. Diallelic RFLP system resulting from single nucleotide change. B. Multiallelic VNTR system.[12]

The other class of RFLP comprises those regions of DNA called minisatellites or VNTR (variable number of tandem repeat). These VNTR loci are scattered throughout the genome and show levels of variability. The

variable region consists of regions of DNA where a short sequence of bases (typically 20 base pairs) is repeated over and over again.[2,14-15] Variability arises from differences in the number of repeat units within the fragments between fixed restriction sites having proven useful to reveal polymorphisms thereby constituting a genetic fingerprint unique to an individual (Fig. 2).[10,12-14,16] The VNTR loci are inherited in a Mendelian fashion and can therefore be used as genetic markers.[2,14,17] This DNA fingerprinting technique is a complex and technically demanding process. The first stage involves the chemical extraction of DNA followed by digestion with a restriction enzyme, which does not cleave the repeat unit. The next stage is to separate DNA fragments by gel electrophoresis followed by transfer from the gel by Southern blotting to a sheet of membrane. Thereupon, the membrane is hybridized with radioactive probes consisting of tandem repeats of the core sequence able to detect many highly variable loci simultaneously (multilocus probes) which can be visualized on X-ray film by autoradiography.[10,14] The end result is a pattern of many bands of a complex barcode-like appearance on X-ray film that provide unique and highly informative identification.[2-3,14-17] However, resolving those alleles that differ by one to a few repeat units (closely spaced alleles) can be extremely difficult particularly if the alleles are long and the core repeat is short.[12,15,18] Hence, the DNA fingerprinting system requires skill in pattern interpretation as well as in mathematics.[3] The quantity of DNA analyzed will affect the number of bands appearing on the profile.[18-19] It has been proven far less useful for routine forensic analysis of samples in which the DNA is often partially degraded or the amounts are too small.[2] DNA fingerprints can only be obtained from high-molecular weight undegraded DNA (good quality DNA).[2,16]

To circumvent this problem, typing systems based on the polymerase chain reaction (PCR) were developed, a method for copying and thus

amplifying DNA in the test tube. This method can be used to amplify specific segments of DNA of interest in order to provide enough DNA for analysis.[10,12,15] In principle, PCR allows DNA typing to be extended to the level of the single cell, thereby enabling typing on many sources of materials such as hair, nail clippings, and skeletal remains.

For example, the PCR-based study of a single gene locus or multiple gene loci has revealed that the Human leukocyte antigen class II at the HLA-DQα gene shows substantial individual variation in its base sequence. The most variable segment of this gene can be amplified by PCR and rapidly typed by reverse dot blot hybridization with allele specific probes to distinguish six different alleles and thus twenty-one different combinations of two alleles can be distinguished. This rapid test is useful for detecting exclusions, but has insufficient statistical power for positive identification.[3,15,20-21] In order to increase statistical power, the other multiple gene independent loci as for example polygenic markers were studied. These polygenic markers include LDLR (low density lipoprotein) on chromosome 19, GYPA (glycophorin A) on chromosome 4, HBGG (hemoglobin G gamma globin) on chromosome 11, D_7S_8 on chromosome 7 and GC (group specific component) on chromosome 4. These two genetic marker kits contribute a convenient but expensive choice for forensic laboratory work.[20-21]

Another DNA region based on PCR studies is short tandem repeat (STR) or microsatellite. Microsatellites, like minisatellites, consist of tandem repeated sequences but with a repeat unit [of 1 to 6 basepairs] much shorter than the minisatellites (Fig. 3).[22-23] Many of these sequences are highly polymorphic due to variation in the length of a short tandem repeat (STR) and inherited stability.[24-27] Trimeric and tetrameric STRs occur every 300 to 500 kb on the human X chromosome and appear to be interspersed at this frequency

throughout the genome.[27] These repeats can be faithfully amplified with the polymerase chain reaction (PCR) using flanking sequence primers and show better stability during PCR than do dinucleotide repeats.[23-25,27-29] Dinucleotide repeat markers have a high frequency of strand-slippage phenomenon that results in laddering artifacts so that the primary allele bands are less intense.[13,24-26] Therefore, the increased PCR stability of tri- and tetranucleotide

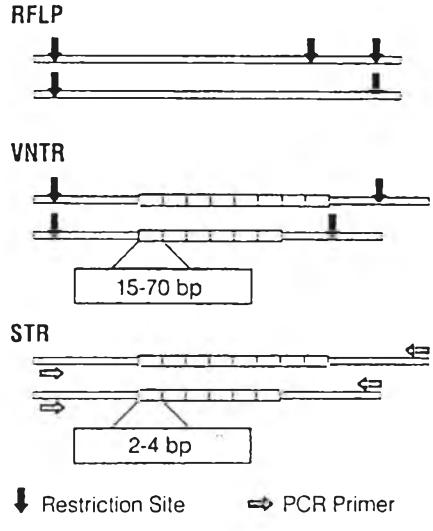

Fig. 3. Schematic diagram illustrating different types of polymorphic markers. RFLP markers are distinguished by polymorphism due to presence or absence of a restriction site at the same site in homologous chromosomes. VNTR and STR markers contain repeated sequences of 15-70 and 2-4 bases, respectively. VNTR markers are detected using flanking restriction sites (as shown) or using PCR primers, in which case they are called "amplifiable VNTRs" or "AMP-FLPs". Unique flanking sequences are used as PCR primers with STR markers.[13]

repeats results in a greater fraction of the total signal being concentrated in the primary allele band, which allows for more precise interpretation of genotypes for these markers.[29-31] In contrast to VNTR, due to the inability to determine alleles precisely with Southern hybridization-based detection, the allele distribution appears continuous because of the limited resolving power of Southern gels. Sometimes if two or more closely migrating bands were present, they were considered to be comigrating ("shared") even if they were

slightly misaligned or different in intensity.[28,32] These features complicate the application of VNTRs to personal identification. The STR makes products of 100 to 500 basepairs. The small size of STR facilitates their simultaneous study in a multiplex PCR, in which two or more loci are amplified in one reaction from a single DNA template.[26-27] Further advantages are the possibility of using small amounts of DNA.[13,24,26] In forensic cases, they have already been applied to the analysis of severely degraded DNA.[27,33] The amplified fragments derived from several markers can be loaded onto each gel lane. Many of the samples can be typed on a single polyacrylamide gel. In addition, use of these markers has reduced the cost of genotyping in some laboratories because of their ability to be visualized with nonisotopic silver staining methods.[29] The lower cost, greater sensitivity, and increased speed of the PCR rendered STR more useful markers for person identification and paternity test.

Since the allele frequencies of any given locus vary between population groups,[23,28] the DNA commission of the International Society for Forensic Haemogenetics (ISFH) has recommend the following.[34] Before a polymorphism can be introduced into paternity testing, DNA polymorphisms should be defined by population studies and should have been tested and published. Each laboratory should construct its own data base appropriate for local populations. Such data bases should be composed of not less than 100 individuals. The population sample should be representative of the relevant local population. Data bases for different local populations should not be merged until it can be demonstrated that it is statistically acceptable to do so.

The aim of this study is to establish a population data base consisting of 8 tetrameric STR loci including CSF1PO, TPOX, TH01, F13A01, FESFPS, F13B, LPL, vWA and to check for relevancy by using appropriate statistical

analysis before introduction into persons for identification and paternity test. All of these loci have been widely studied elsewhere and demonstrated a high degree of variability within those populations. They exhibit a high discrimination potential yet minimal genetic artifacts (such as microvariants) and minimal amplification artifacts (such as stutter bands).[13] They appear to be well suited for routine identification of individuals and paternity test and commercial kits are available.

The procedures involve the chemical extraction of DNA by salting-out and ethanol precipitation from EDTA-blood samples, amplification of specific STR loci, separation of the alleles by gel electrophoresis and detection by silver staining before exposure to APC film. Data are analyzed by genotype determinations and statistical analysis.

Expected benefits are:

(1) established population data base of 8 STR loci in the Thai population for use in forensic science, person identification and paternity test.

(2) a constructed battery of 4-7 STR loci most suitable for paternity testing in Thais, which should have sufficient statistical power and could be developed for additional use or replace a routine test for blood group substances.

(3) the results of the studies will allow us to extend the test to include other specimens such as blood stains, semen stains, vaginal swabs, hair roots, skeletal remains, body parts, saliva, etc.