GENOTYPING OF ANGIOTENSIN-1 CONVERTING ENZYME INSERTION/DELETION GENE POLYMORPHISM IN MALAY ISCHEMIC STROKE PATIENTS

by

CHEUNG TIAN PEI

Dissertation submitted in partial fulfillment of the requirements
for the degree of Bachelor of Health Sciences
(Biomedicine)

May 2012

ACKNOWLEDGEMENT

First of all, I would like to express my deepest appreciation and sincere gratitude to Associate Professor Dr. Gan Siew Hua, director of Human Genome Center, for her guidance, financial backing, and support throughout the final year project. It has been an honor for me to work under her supervision. Her selfless advices have helped me greatly in conducting the project.

Besides, I also wish to thank Dr. See Too Wei Cun, supervisor of the whole final year project, Dr. Yvonne-Tee Get-Bee and Dr. Few Ling Ling, lecturers of USMKK, for providing guidance and recommendations throughout the course of my project. Their invaluable suggestions have assisted me considerably.

I am also grateful to Ms. Loo Keat Wei, post-graduate student of Human Genome Center, for her guidance, encouragement and provision of strong broad idea regarding the flow of project. She has supervised and helped me a lot throughout the whole project. I am very glad to benefit from her assistance and helpful recommendations.

Last but not least, I would like to express my cordial appreciation to my family members and friends for their love and support in completion of this challenging project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

%	Percent
μM	Micromolar
A	Absorbance
A, G, C, T	Adenine, Guanine, Cytosine, Thymine
ACE	Angiotensin-1 converting enzyme
Ang	Angiotensin
ApoE	Apolipoprotein E
BLAST	Basic Local Alignment Search Tool
bp	Base pair
dbSNP	Public SNP database of NCBI
dHPLC	Denaturing high performance liquid chromatography
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
Fgβ	Beta fibrinogen
HUSM	Hospital Universiti Sains Malaysia
I/D	Insertion/Deletion
kb	Kilo bases
LB	Lithium borate
LD	Linkage disequilibrium
$MgCl_2$	Magnesium chloride
Min	Minute

mM	Millimolar
mRNA	Messenger ribonucleic acid
MTHFR	Methylene tetrahydrofolate reductase
n	Number
NCBI	National Center for Biotechnology Information
nm	Nanometer
°	Degree
PCR	Polymerase chain reaction
RAS	Renin-angiotensin system
RFLP	Restriction fragment length polymorphism
rpm	Revolutions per minute
rs number	RefSNP accession ID
rs	Reference SNP (RefSNP)
Sec	Second
SNP	Single nucleotide polymorphism; Reference SNP
UV	Ultraviolet
WHO	World Health Organization
X	Times
β	Beta
μ	Micro
μl	Microlitre

ABSTRAK

Gen enzim penukaran angiotensin-1 (*ACE*) manusia mengandungi ulang elemen Alu di intron 16 dalam kromosom 17 (17q23). Polimorfisme ini telah dikaji untuk menentukan kesannya dalam menyebabkan strok iskemia. Namun demikian, keputusannya adalah tidak konsisten. Oleh itu, kajian ini telah dijalankan untuk menyiasat genotip bagi polimorfisme I/D *ACE* untuk menentukan kekerapan genotipnya di kalangan pesakit strok iskemia Melayu. Subjek kajian ini terdiri daripada 27 pesakit strok iskemia Melayu yang direkrut dari HUSM. Polimorfisme gen I/D *ACE* ditentukan dengan mengenotip polimorfisme SNP rs4341 melalui kaedah PCR-RFLP. Kekerapan polimorfisme I/D *ACE* (genotip-genotip II, ID dan DD) ditentukan dan dibentangkan dalam bentuk peratusan. Keputusan kajian ini menunjukkan bahawa genotip II *ACE* dapat dikesan dalam semua pesakit strok iskemia Melayu (100%) manakala genotip ID dan DD tidak menyumbang kepada kekerapan tersebut. Oleh itu, genotip II *ACE* adalah penyebab kejadian strok iskemia dalam kalangan pesakit strok Melayu. Namun, keputusan kajian ini merupakan penemuan awal di mana kajian-kajian yang lebih besar dan meluas boleh mendasarkan hasil kajian tersebut.

Kata kunci: gen polimorfisme I/D *ACE*, Melayu, PCR, RFLP, strok iskemia

ABSTRACT

The human angiotensin-1 converting enzyme (*ACE*) gene contains an insertion or deletion (I/D) of Alu repeat element in intron 16 of the chromosome 17 (17q23). This polymorphism has been widely investigated to determine its effects in causing ischemic stroke but inconsistent results are obtained. Thus, this study aims to genotype the *ACE* I/D polymorphism to determine its genotypic frequency among Malay ischemic stroke patients. The study subjects comprised of 27 Malay ischemic stroke patients that were recruited from HUSM. *ACE* I/D gene polymorphism was determined by genotyping the SNP rs4341 polymorphism using PCR-RFLP method. The frequencies of *ACE* polymorphism (II, ID and DD genotypes) were determined and presented in the form of percentage. The results showed that *ACE* II genotype was present in all Malay ischemic stroke patients (100%) while ID and DD genotypes did not make any contribution to the genotypic frequency. Therefore, the *ACE* II genotype is responsible for the occurrence of ischemic stroke among Malay stroke patients. However, the result is a preliminary finding in which larger and extensive studies can be performed based on it.

Keywords: *ACE* I/D gene polymorphism, ischemic stroke, Malay, PCR, RFLP

CHAPTER 1: INTRODUCTION

Stroke is a common cause of death, disability and financial lost throughout the world. According to the National Stroke Association of Malaysia (NASAM), stroke is the third largest cause of death in Malaysia following heart disease and cancer. It is also the single most common cause of severe disability where approximately 40 000 Malaysian suffer from stroke every year (NASAM, 2007).

According to World Health Organization (WHO), stroke is caused by the interruption of blood supply to the brain. This is often due to a burst of blood vessel (hemorrhagic stroke) or blockage by a clot (ischemic stroke). Subsequently, the supply of oxygen and nutrients to the brain diminishes and this damages the brain tissues (WHO, 2002). Stroke is a complex disease caused by multiple risk factors which can be attributed to a combination of both environmental and genetic risk factors (Tonk and Haan, 2007).

Several genetic polymorphisms are associated with the risk of ischemic stroke. A meta analyses showed positive associations with ischemic stroke for factor V Leiden Gln506, prothrombin G20210A, plasminogen activator inhibitor-1 (PAI-1) 5G and glycoprotein IIIa Leu33Pro polymorphism (Bentley *et al.*, 2010). Many studies focused on the angiotensin-1 converting enzyme (*ACE*) insertion/deletion (I/D) gene polymorphism to determine its definite association with ischemic stroke showed conflicting results where some studies found an association while others failed to define any (Tuncer *et al.*, 2006; Rao *et al.*, 2009; Domingues-Montanari *et al.*, 2010).

Therefore, in this study, we genotyped the *ACE* I/D gene polymorphism in Malay ischemic stroke patients.

A single nucleotide polymorphism (SNP) located close to the *ACE* I/D gene polymorphism was genotyped by using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) in order to determine the *ACE* I/D polymorphism based on the method described by Glenn *et al.*, (2009). SNP rs4341 polymorphism was selected as a tag SNP for this study as it was found to be in complete linkage disequilibrium (LD) with the *ACE* I/D gene polymorphism (Tanaka *et al.*, 2003).

Objectives of this study were as followed:

(i) To genotype the *ACE* I/D gene polymorphism (or rs4341) in Malay ischemic stroke patients (n = 27).

(ii) To determine the genotypic distribution of the *ACE* I/D gene polymorphism (or rs4341) in Malay ischemic stroke patients (n = 27).

CHAPTER 2: LITERATURE REVIEW

2.1: Ischemic stroke

2.1.1: Epidemiology

According to WHO, 15 million people worldwide suffer from stroke annually. Of these, five million people die and another five million are left permanently disabled, placing a burden on the family and community (WHO, 2002). As stated by WHO (2002), stroke is generally uncommon among people under 40 years old. However, when it occurs in this age group, it is usually caused by high blood pressure. Besides, stroke also occurs in about 8% of children with sickle cell disease (WHO, 2002).

2.1.2: Definition and pathophysiology

Stroke is a clinical syndrome characterized by the rapidly developing clinical symptoms and/or signs of focal, and at times global, loss of cerebral function, with symptoms lasting for more than 24 hours or leading to death, with no apparent cause other than that of vascular origin (MOH, 2006). Stroke is divided into two major types: ischemic and hemorrhagic strokes (Caplan, 2000). The majority of strokes are ischemic stroke (80-90%) which is characterized by a sudden decrease in blood flow to one or more central nervous system. Ischemic stroke can be attributed to large-artery atherosclerosis (atherothrombotic stroke), small-artery occlusion (lacunar stroke), cardioembolism, vasculitis, metabolic disorders, trauma, monogenic disorders, undetermined causes or rare etiologies (Debette and Seshadri, 2009).

Ischemic stroke is a syndrome rather than a specific disease. It can be due to several different pathologies as shown in Table 2.1. The major pathologies included large artery disease, small vessel disease and cardioembolism (Gulcher *et al.*, 2005). Large artery disease (carotid and vertebral artery stenosis) is usually caused by atherosclerosis in the large extracranial or intracranial arteries supplying the brain. It shares many pathogenic and risk factor similarities with atherosclerosis elsewhere in the body including coronary artery disease and peripheral vascular disease. Small vessel disease causes small deep infarcts (lacunar infarcts) within the white and deep grey matters in the brain. It is also the most important pathology resulting in vascular dementia and cognitive impairment. Meanwhile, cardioembolism results from a variety of cardiac pathologies which produce thrombosis within the heart. This subsequently breaks off and embolizes to the brain, resulting in stroke (Markus, 2003). Further explanation of the ischemic stroke pathophysiology algorithm is given in Appendix A.

Table 2.1: Pathophysiological classification of ischemic stroke

Ischemic stroke [a]
Large artery
Small vessel disease (lacunar)
Cardioembolic
Other determined etiology[*1]
Undetermined etiology[*2]
Multiple possible etiologies

[a] This classification is based on the modification of the TOAST (Trial of ORG 10172 in Acute Stroke Treatment) and Stroke Data Bank classifications by Markus (2003);

[*1] Examples of other determined etiology of ischemic stroke comprise rare causes such as nonatherosclerotic, vasculopathies, hypercoagulable states or hematologic disorders (Adams *et al.*, 1993);

[*2] Ischemic stroke of undetermined etiology includes cases where the cause cannot be determined with any degree of confidence and those with two or more potential causes of stroke where the final diagnosis cannot be made (Adams *et al.*, 1993).

2.1.3: Genetic determinants of stroke

Ischemic stroke is the most common form of stroke where it is a multifactorial disease that can be affected by both genetic and environmental factors. There are a variety of techniques that can be used to investigate the genetics basis of stroke. These include linkage based techniques, candidate gene association studies and genome-wide association studies (Markus, 2003). Several genetic determinants are responsible for the etiology of stroke. These include Apolipoprotein E (*ApoE*) gene, methylene tetrahydrofolate reductase (*MTHFR*) and beta fibrinogen (*Fgβ*) (Gao *et al.*, 2006). Genetic determinants of ischemic stroke may act independently or alter the effects of known risk factors.

A study was conducted to determine the association of genetic polymorphisms and ischemic stroke (Gao *et al.*, 2006). The interaction of gene with environmental risk factors (smoking and alcohol drinking) was also investigated among Chinese population. Genetic polymorphisms of *ACE* I/D, *ApoE, MTHFR* gene 677C/T and *Fgβ* were analyzed by PCR and denaturing high-performance liquid chromatography (dHPLC) analysis. The result showed that individuals with *Fgβ* CT/TT, *MTHFR* CT/TT, and *ACE* ID/DD genotypes had increased incidence of ischemic stroke. In addition, the study also indicated that certain unfavorable genotypic combinations act synergistically to result in ischemic stroke.

The genetic determinants that are responsible for the increased occurrence of ischemic stroke can be considered in terms of their possible pathogenic effects. They may act by contributing to the process in the development of ischemic stroke, interfering

with the outcome following occlusion or acting through both of these processes (Markus, 2003). In view of the complex pathogenesis of ischemic stroke and possible multiple effects caused by the candidate genes, the mechanism of a given candidate gene in causing ischemic stroke needs to be carefully considered as the gene may act independently or synergistically with other candidate genes (Gao *et al.*, 2006). Besides, there is also no single gene that makes large contribution to the overall stroke risk (Sharma, 1998). These issues are very important especially when the gene under study is targeted for therapeutic intervention in the future.

2.1.4: Risk factors of stroke

A risk factor of stroke occurring in an individual is a characteristic that can cause that individual to have an increased risk of developing stroke compared with those without that characteristic. However, it does not necessarily imply causality. There are a number of conventional risk factors which may result in stroke occurrence. Many of these have a genetic predisposition. These include hypertension, diabetes and hyperlipidemia (Markus, 2003). In addition, ageing, atrial fibrillation, smoking and excessive alcohol consumption are also associated with an increased risk of stroke (Humphries and Morgan, 2004).

Table 2.2 shows the modifiable and non-modifiable risk factors which result in increased stroke occurrence. Each of these risk factors contributes to a specific stroke mechanism. For example, hypertension can result in both hemorrhagic and ischemic stroke while diabetes is more likely to cause ischemic stroke due to its effect on the development of large-artery atherosclerosis (Humphries and Morgan, 2004).

Table 2.2: Modifiable and non-modifiable risk factors associated with increased risk of stroke

Non-modifiable	Modifiable
Age	Hypertension (systolic and diastolic)
Sex	Diabetes mellitus
Ethnicity	Hyperlipidemia
Family history of stroke	Atrial fibrillation
	Coronary heart disease
	Cigarette smoking
	Obesity and physical inactivity
	Raised homocysteine level
	Heavy alcohol consumption
	High dietary salt intake
	Previous history of stroke

(Source: MOH, 2006)

2.2: Angiotensin-1 converting enzyme (*ACE*) insertion/deletion (I/D) gene polymorphism

2.2.1: *ACE* gene

ACE gene encoding the angiotensin-1 converting enzyme is located on the long arm of chromosome 17 (17q23). The gene is 21 kilo bases (kb) long. It consists of 26 exons and 25 introns. To date, there are more than 160 *ACE* gene polymorphisms that have been reported where most were only single nucleotide polymorphisms (SNPs). From this number, only 38 of them are located in the coding regions while 18 of them are missense mutations. *ACE* gene is responsible for encoding of two isoforms of ACE. These are the somatic form (sACE) and testicular form (tACE), or known as germinal ACE (gACE) (Sayed-Tabatabaei *et al.*, 2006).

Rigat *et al.* (1990) detected a polymorphism involving the presence (insertion, I) or absence (deletion, D) of a 287 bp sequence of Alu repeat element in the *ACE* gene (NCBI reference SNP ID: rs1799752) (NCBI, 2011c). Detail of the gene polymorphism is shown in Appendix B. The genomic location of *ACE* I/D polymorphism is at about 58 919 623 bp. This insertion/deletion polymorphism is located in non-coding region at intron 16 of the gene. The polymorphism generates three genotypes namely deletion homozygotes (DD), insertion homozygotes (II) and heterozygotes (ID) (Rigat *et al.*, 1990).

2.2.2: Alu elements

Alu elements are the most abundant Short Interspersed Elements (SINEs). The name 'Alu elements' was given because they contain a recognition site for the restriction enzyme AluI (Houck *et al.*, 1979). They are mobile elements which have a highest copy number of over one million in the human genome. Thus, Alu repeats comprise more than 10% of the mass of the human genome.

Full-length of Alu elements are approximately 300 bp in length and they are commonly found in introns, 3'-untranslated regions of genes and intergenic genomic regions (Makalowski *et al.*, 1994; Carroll *et al.*, 2001). Each Alu element is dimeric in structure comprising of two tandemly arranged halves. There are also a middle A-rich region and a 3' oligo (dA)- rich tail flanked by short direct repeat sequences (Batzer and Deininger, 1991; Stoneking *et al.*, 1997) as shown in Figure 2.1.

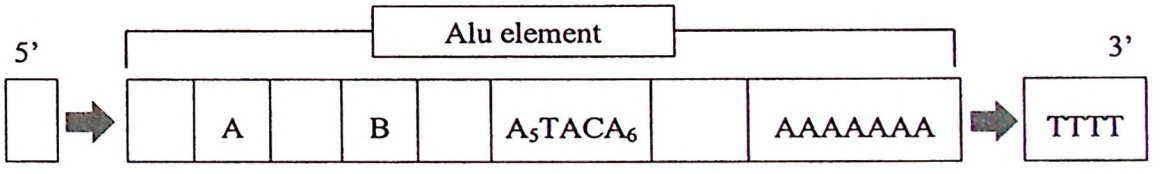


Figure 2.1: Structure of the Alu elements

The 5' half of each Alu element contains an RNA-polymerase-III promoter (A and B boxes) while the 3' terminus usually contains a run of As that is only occasionally interspersed with other bases. A and B boxes: RNA-polymerase-III promoter; A_5TACA_6: nucleotide bases (A, adenine; T, thymine; and C, cytosine); AAAAAAA: A-rich region; TTTT: a run of thymine (Batzer and Deininger, 2002).

2.2.3: *ACE* I/D gene polymorphism and methods of detection

DD carriers demonstrated approximately twice the mean ACE activity level, compared to II genotype individuals. On the other hand, ID genotype individuals showed intermediate levels which indicated codominancy. The concentration of ACE in serum is very stable within an individual but there is large inter-individual difference. In addition, a study also showed that the *ACE* I/D gene polymorphism was associated with 47% of the variance in serum ACE levels (Rigat *et al.*, 1990).

Since the presence of D allele is associated with increased ACE activity, it may lead to hypertension and increased risk of vascular disease. This allele could also directly affect the risk of stroke acting through other factors such as affecting the endothelium (Butler *et al.*, 1999). A study showed that there was significant higher frequencies of the DD genotype and D allele of the *ACE* gene in patients with parental history of stroke (PHS) than those without. Since the frequencies were increased, those affected may transmit the D allele to their offspring more often, compared to unaffected parents (Maeda *et al.*, 1996).

Initially, *ACE* I/D gene polymorphism was detected based on the PCR method. A set of primers was used to flank the insertion sequence (Rigat *et al.*, 1990). However, there is possibility of mistyping ID heterozygotes by using this method (Shanmugam *et al.*, 1993). It was reported that the preferential amplification of the shorter D allele may cause the misclassification of approximately 4 to 5% of ID genotypes to DD genotypes, causing a higher frequency of DD genotype. Thus, an additional PCR analysis was

required to confirm the DD genotypes obtained in the first standard PCR (Shanmugam *et al.*, 1993).

Shanmugam *et al.* (1993) adopted a different strategy to eliminate the possibility of mistyping. Five percent dimethyl sulfoxide (DMSO) and sense primer from the 5' end of the insertion sequence, along with the standard antisense primer were used (Shanmugam *et al.*, 1993). Further, modification of this method by a step-down PCR has also improved the accuracy of detection (Chiang *et al.*, 1998). In 2008, a quick and easy technique using dHPLC (denaturing high performance liquid chromatography) at non-denaturing conditions was developed to analyze the *ACE* I/D gene polymorphism. Although this technique produces 100% accuracy in ID heterozygotes, it is not cost effective (Koyama *et al.*, 2008).

2.2.4: Association studies

ACE I/D gene polymorphism had been investigated for its association with various disease etiologies including ischemic heart disease (Lindpaintner *et al.*, 1995), end-stage renal disease (Ali *et al.*, 2011), hypertension (Ramu *et al.*, 2011), Alzheimer disease (Ghebranious *et al.*, 2011), type 2 diabetes mellitus (Ramachandran *et al.*, 2008), myocardial infarction (Cambien *et al.*, 1992) and coronary artery disease (Pfohl *et al.*, 1999). Besides, this polymorphism was also investigated for several physiological events such as athlete endurance status (Scott *et al.*, 2005) and longevity (Schachter *et al.*, 1994).

Even though many studies have been published regarding the association of *ACE* I/D gene polymorphism with ischemic stroke, controversial results were reported. Many studies demonstrated significant associations (Casas *et al.*, 2004) while others did not confirm these findings (Tuncer *et al.*, 2006; Domingues-Montanari *et al.*, 2010). Therefore, the effect of this polymorphism on the risk of developing ischemic stroke remains uncertain. These conflicting results can be due to genotypic and phenotypic misclassifications, insufficient power in some studies, mismatching of cases and controls according to the ethnic origin and the presence of interaction with other genes or environmental factors (Markus, 2003; Sayed-Tabatabaei *et al.*, 2006).

A study conducted by Hong *et al.* (2008) demonstrated a significant association between ischemic stroke and *ACE* I/D polymorphism in Korean populations (237 cases and 234 controls). They found out that the frequencies of ID, DD and overall (ID + DD) genotypes were associated with increased risk of ischemic stroke. Besides, the DD genotype may also contribute to the occurrence of small-vessel disease (SVD) or lacunar infarction rather than large-vessel disease (LVD) that commonly gives rise to the thrombotic stroke (Hong *et al.*, 2008). However, both SVD and LVD contribute to the thrombotic stroke that may subsequently result in ischemic stroke (Szolnoki *et al.*, 2001).

Conversely, a meta-analysis which was carried out in Asian populations (Japanese, Korean and Chinese) involving 988 stroke patients and 1441 control individuals showed no statistical significant association between *ACE* I/D gene polymorphism and ischemic stroke. The study showed that the D allele and homozygous

DD genotype were not associated with increased risk of stroke (Banerjee *et al.*, 2006). However, the study specified that it was limited by the relatively small sample size. Therefore, it is important to conduct further studies having larger sample size.

In European populations, statistically significant association between *ACE* I/D gene polymorphism and increased risk of ischemic stroke was evidenced from two meta-analyses (Sharma, 1998; Casas *et al.*, 2004). Seven case-control studies (1394 cases and 826 controls) were investigated by Sharma (1998) while 11 studies (2990 cases and 11 305 controls) were analyzed by Casas *et al.* (2004). In contrast to the Asian populations, these results suggested that the ethnic variations among different populations could account for the conflicting findings on the roles of *ACE* I/D polymorphism in the pathogenesis of ischemic stroke.

2.3: Angiotensin-1 converting enzyme (ACE)

ACE protein is a zinc metallopeptidase found on the surface of endothelial and epithelial cells. ACE plays an important role in two physiological systems namely the renin-angiotensin and kinin-kallikrein systems (shown in Figure 2.2). These lead to the production of angiotensin II and to the degradation of bradykinin respectively. ACE functions in converting the inactive decapeptide, angiotensin I (Ang I or Ang 1-10) to the active octapeptide and potent vasoconstrictor, angiotensin II (Ang II or Ang 1-8). Ang II serves as the main active product of the renin-angiotensin system (RAS). RAS functions in long term regulation of blood pressure and blood volume in the body (Sayed-Tabatabaei *et al.*, 2006).

Upon activation of ACE, Ang II is produced where it serves as a potent vasoconstrictor. It also acts on the adrenal cortex to result in the release of aldosterone. Aldosterone stimulates renal tubules to reabsorb more sodium and water from the urine (Brewster and Perazella, 2004). Thus, the amount of fluid in the blood and blood pressure are increased. Besides, Ang II also stimulates various cytokines and growth factors to mediate cell growth and proliferation (Carluccio *et al.*, 2001). In addition, Ang II induces endothelial dysfunction by reducing nitric oxide bioavailability (Rajagopalan *et al.*, 1996). All these findings showed that Ang II is very important for the cardiovascular pathophysiology. Another role of ACE is in the kinin-kallikrein cascade where ACE functions in metabolizing bradykinin to form inactive bradykinin.

The biological activity of ACE suggests that the *ACE* gene may play a role in predisposition to cardiovascular disease as RAS is involved in the systemic regulation of fluid volume and blood pressure, induction of smooth muscle proliferation and stimulation of myocardial cell hypertrophy (Tiret *et al.*, 1992).

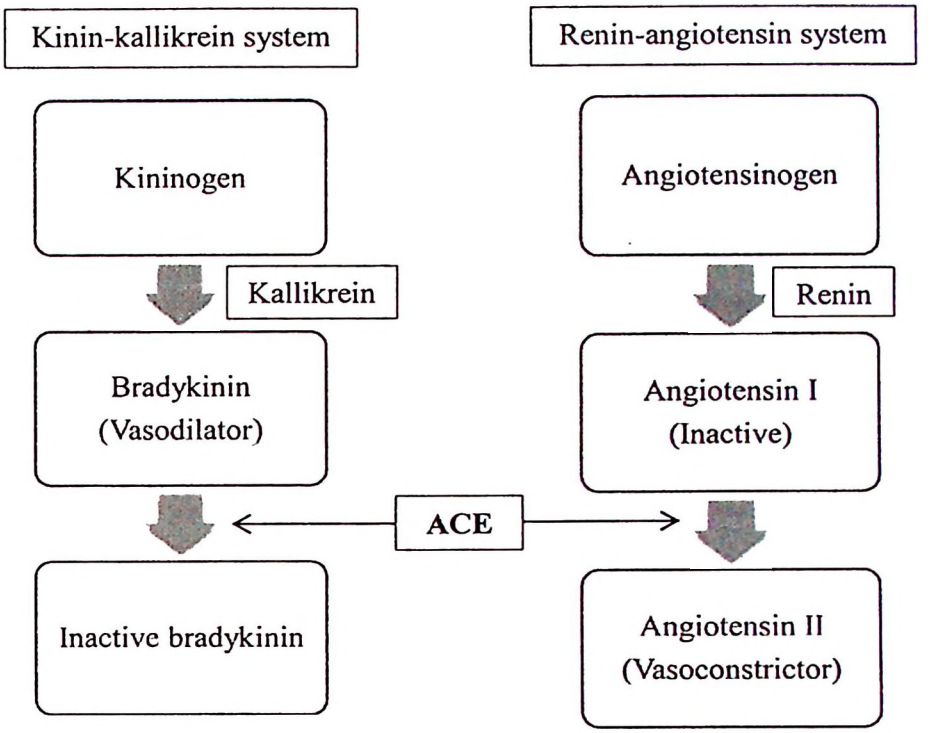


Figure 2.2: The kinin-kallikrein and renin-angiotensin systems

Angiotensin-1 converting enzyme (ACE) functions in converting the bradykinin and angiotensin I into inactive bradykinin and angiotensin II respectively in two separate but interconnected systems.

2.4: Single nucleotide polymorphism (SNP) rs4341

2.4.1: SNPs

The most commonly occurring human sequence variation are SNPs while the rest are attributed to insertions or deletions of one or more bases, repeat length polymorphisms and rearrangements. The low rate of recurrent mutation of SNPs causes them to be the stable indicators of human history (Sachidanandam *et al.*, 2001).

SNP occurs when a single base mutation replaces one nucleotide for another, for example, CTG to TTG, with C to T transitions. It exhibits an abundant form of genetic variations. SNPs differ from other variations by having more than 1% frequency in the human population. In human genome, there are about 10-30 million SNPs with an average SNP in every 100-300 bases. If the non-synonymous SNP is located in a protein coding sequence, it can result in amino acid changes and thus, functional alterations of the encoded proteins (Ban *et al.*, 2010).

On the other hand, SNPs located in a promoter region will have an effect in transcriptional regulation. The synonymous SNP may result in functional consequences by influencing the stability or folding of mRNA transcripts. SNP in an intron region may affect the splicing or expression of the corresponding gene. If the SNP is located in 5'- or 3'-untranslated mRNA region, it can affect the stability or processing of the RNA. Moreover, SNP in non-protein coding regions of genes will affect the binding of regulatory factors and this may cause imbalanced expression of the SNP alleles (Hudson, 2003).

2.4.2: Genotyping of SNPs

Patterns of SNP are likely to affect human genotypes. Thus, large-scale association studies have been based on SNP genotyping in order to identify genes that are responsible for complex diseases and responses to drugs or environmental chemicals. SNP genotyping technologies consist of two components. One method is used for determining the type of base present at a given SNP locus (allele discrimination) while another method is used for reporting the presence of the allele (signal detection) (Twyman, 2005).

There are more than 20 different SNP genotyping methods which consist of various combinations of different allele-discrimination and signal detection methods. High throughput genotyping methods are employed where many SNPs in many individuals are genotyped. Such methods comprise Taqman, single-base extension-based assays, MALDI-TOF (Matrix-assisted laser desorption/ionization-time of flight) mass spectrometry-based systems, invader (combined with PCR) and pyrosequencing which focus on large-scale association studies (Tsuchihashi and Dracopoli, 2002).

2.4.3: SNP rs4341 polymorphism

SNP rs4341 polymorphism is located on intron 16, outside of the 100-bp region downstream of the *ACE* I/D gene polymorphism. It results from the replacement of cytosine (C) with guanine (G) (Glenn *et al.*, 2009; NCBI, 2011b) (Appendix C). Its genomic location is at 61 565 990 bp. Three possible genotypes of rs4341 are G/G, G/C and C/C polymorphisms. In this study, the rs4341 is an alternative genotyping target for use in place of the *ACE* I/D polymorphism. This is because the SNP has complete

linkage disequilibrium (LD) with the *ACE* I/D gene polymorphism (Tanaka *et al.*, 2003). Locations of the *ACE* I/D gene polymorphism and SNP rs4341 are given in Appendix D.

In this study, rs4341 was selected based on several criteria. The location of the SNP spans the region close to the *ACE* I/D polymorphism. In addition, the SNP had also been tested and validated experimentally and it has the ability to develop a testable PCR-RFLP (Glenn *et al.*, 2009). Homozygous G allele of the rs4341 (GG genotype) is considered equivalent to DD genotype of the *ACE* I/D polymorphism. On the other hand, homozygous C allele (CC genotype) is corresponded to II genotype while heterozygous allele (GC genotype) is equivalent to ID genotype (Eisenmann *et al.*, 2009).

Linkage disequilibrium (LD) occurs when a marker allele is in close proximity with the disease susceptibility allele. Subsequently, these two alleles are inherited together over several generations in the population studied. Therefore, the same allele will be detected in affected individuals from multiple unrelated families, but belonging to the same population and thus to the same genetic pool. The closer the marker is to the disease gene, the longer the allele-disease association will persist (Kruglyak, 1999).

2.5: Restriction enzyme (Mva1269I)

For PCR-RFLP analysis in this study, the restriction enzyme, Mva1269I which shares 31% amino acid sequence identity with its isoschizomer BsmI, was used to cleave the DNA at the recognition site (5'-GAATGC-3'). It belongs to the type IIS restriction enzymes, which is a subgroup of type II family. Type IIS restriction enzymes recognize asymmetric DNA sequences of 4-7 bp long. These will then cleave both strands at specific locations. Mva1269I cleaves DNA (as indicated by the "↓" symbol in Table 2.3) close to its recognition site, which is 1 bp away from the recognition site in one strand. This produces 3' overhang that is also known as sticky end or cohesive end (Bath *et al.*, 2002).

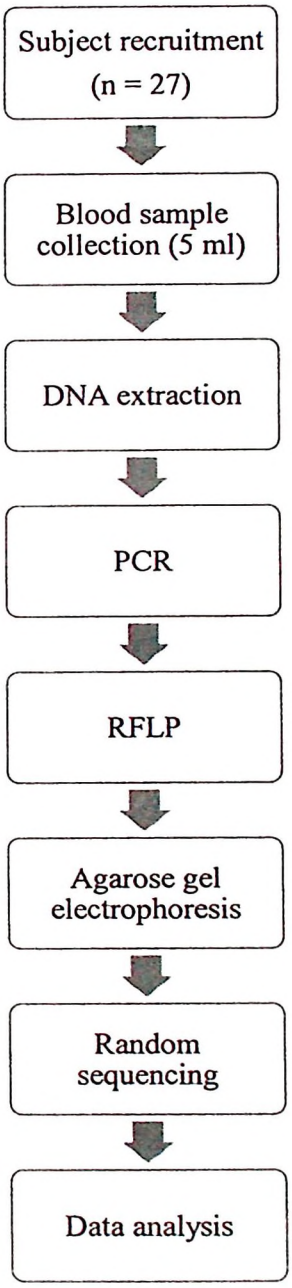
A study by Armalyte *et al.* (2005) found out that Mva1269I is a monomeric enzyme both in solution and upon binding of cognate DNA. This type IIS restriction endonuclease from *Micrococcus varians* consists of two domains: EcoRI- and FokI-like catalytic domains which are responsible for the sequential cleavage of each DNA strands. Besides, the cleavage of the bottom strand also preceded the cleavage of the top strand (Armalyte *et al.*, 2005).

Table 2.3: Restriction enzyme (Mva1269I) used in the PCR-RFLP analysis

Source	Recognition site sequence	Recognition site sequence after cut	Overhang
Micrococcus varians RFL1269	5'...GAATGCN↓...3' 3'...CTTAC↓GN...5'	5'…GAATGCN N…3' 3'…CTTAC GN…5'	CN – 3'

CHAPTER 3: EXPERIMENTAL OVERVIEW

Flow chart of the experimental overview in this study is shown below:



CHAPTER 4: MATERIALS AND METHODS

4.1 Materials

4.1.1: List of chemicals

- QIAamp DNA blood mini kit (Qiagen, Hilden, Germany)
- AccuOligo® forward and reverse primers (Bioneer, United States)
- AmpliTaq® DNA polymerase (Applied Biosystems®, Australia)
- AmpliTaq® 25 mM $MgCl_2$ (Applied Biosystems®, Australia)
- GeneAmp® 10 mM dNTP mix (Applied Biosystems®, Australia)
- GeneAmp® 10X PCR buffer (Applied Biosystems®, Australia)
- PCR-grade water (nuclease-free)
- FastDigest® Mva1269I enzyme (Fermentas, United States)
- 10X FastDigest® green buffer (Fermentas, United States)
- Agarose powder (Molecular Biology Grade, Vivantis, United Kingdom)
- 20X LB® Faster Better Media LLC (United States)
- GeneRuler™ low range DNA ladder (Fermentas, United States)
- 100 bp DNA ladders (Promega, United States)
- 6X orange DNA loading dye (Fermentas, United States)
- GelStar® nucleic acid gel stain (Lonza, Rockland, United States)
- GeneJET™ PCR purification kit (Fermentas, United States)

4.1.2: List of general instruments

- Vacutainer tube with EDTA (ethylenediaminetetraacetic acid)
- Eppendorf Research® pipettes
- Eppendorf tips
- Freezer
- Eppendorf 5415C micro centrifuge gray
- Eppendorf centrifuge 5415 D
- Sprout® mini-centrifuge
- FINEPCR® FINEVORTEX® one-touch vortexer
- AccuBlock™ digital dry baths (Labnet International Inc.)
- Tecan Infinite® 200 NanoQuant multimode microplate reader
- NanoQuant Plate™ (Tecan)
- Eppendorf 0.2 ml PCR tubes
- Eppendorf PCR rack
- E-Centrifuge (Wealtec)
- Techne® TC-512 thermal cycler
- Bio-Rad / MJ Research DNA Engine Peltier thermal cycler-200
- Mettler Toledo® AB-104 analytical balancer
- Microwave oven
- Agarose gel casting tray
- Standard combs
- Casting base
- Horizontal gel electrophoresis system (Bio-Rad Wide Mini-Sub Cell GT Cell)
- Wealtec ELITE 300 power supply
- Wealtec UV transilluminator (MD-25/HD-25)
- AlphaDigiDoc® RT gel documentation system (Alpha Innotech)

4.2: Methods

4.2.1: Subject recruitment and blood collection

Stroke patients (n = 27) from Hospital Universiti Sains Malaysia (HUSM), Kelantan were recruited for this study. The present study was part of another research in which the ethical approval has been obtained from Universiti Sains Malaysia Research Ethics Committee (Human) (USMKK/PPP/JEPeM [231.3(08)]).

Subjects for this study included Malay (only Malaysian) stroke patients from HUSM of any age and sex, and those experiencing lacunar or large artery ischemic stroke while patients having history of heart diseases were excluded. The latter were not included in this study because the thrombus forming in an acute myocardial infarction may result in cardiogenic stroke. This type of ischemic stroke which occurs secondarily may produce confusing results when establishing the effects of the *ACE* I/D gene polymorphism. Control subjects were not included in this study because of time constraint and budget issues.

History of ischemic stroke was determined from the patient's medical record. Informed consents to participate in the study were obtained from all of the recruited patients before blood collection. Once the informed consent was obtained, 5 ml of venous blood was collected by venipuncture into a vacutainer tube (containing EDTA as anticoagulant). The blood collection was performed by clinicians. Collected blood samples were immediately sent to the laboratory for processing or stored in freezer (4°C) for delayed processing.