UNIVERSITI PUTRA MALAYSIA

MOLECULAR CHARACTERIZATION OF EMM-LIKE GENES IN THE MGA REGULON OF GROUP A STREPTOCOCCUS STRAIN ST4547

MAJID ESHAGHI

FSMB 2001 18

MOLECULAR CHARACTERIZATION OF *EMM*-LIKE GENES IN THE *MGA* REGULON OF GROUP A STREPTOCOCCUS STRAIN ST4547

By

MAJID ESHAGHI

Thesis Submitted in Fulfilment of the Requirement for the Degree of Doctor of Philosophy in the Faculty of Food Science and Biotechnology Universiti Putra Malaysia

July 2001

Dedicated to.....

teachers who dedicate the best moments of their life to teach us

for a better tomorrow

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

MOLECULAR CHARACTERIZATION OF *EMM*-LIKE GENES IN THE *MGA* REGULON OF GROUP A STREPTOCOCCUS STRAIN ST4547

By

MAJID ESHAGHI

July 2001

Chairperson: Associate Professor Abdul Manaf Ali, PhD

Faculty: Food Science and Biotechnology

Sequence analysis of the 5' region of the *emm* gene was employed to differentiate 39 group A streptococci (GAS) isolates collected between 1989 to 1997 from patients and carriers in Kuala Lumpur. Sixty-one percent (24) of these isolates contained *emm* genes encoding the M protein for which M-antigen associations had not been made. The remaining strains had *emm* sequences in agreement with previously recorded M-antigen associations. In some cases antigenic variations were observed among individual M types as well as the isolates tested, compared to published M protein sequences. These differences were predominantly due to the non-synonymous base substitutions and occasionally, short insertions and deletions.

Nucleotide sequencing of the *mga* regulon of a new Malaysian *emm* type ST4547 group A streptococcus an opacity factor (OF) negative isolate, showed the existence of two *emm*-like genes, *emm* and *mrp*. The *emm* gene encoded the M protein whereas *mrp* gene encoded the IgG Fc receptor. The gene located upstream of the *scpA* gene, comprised 1305 nucleotides encoded a M protein of 435 amino acids in length with a predicted molecular weight of 49.0 kDa or a predicted mature protein of 394 amino acids with a molecular weight of 44.7 kDa. At the upstream of this gene and downstream of *mga* gene another gene was found and designated as *mrpST4547*. The sequence of this gene comprised 1167 nucleotides encoded a predicted protein of 388 amino acids in length with a predicted molecular weight of 42.2 kDa or a predicted mature protein of 347 amino acids with a molecular weight of 37.9 kDa. The *mga* regulon of the strain ST4547 had a mosaic structure consisting of DNA segments which were suggested to had originated from different OF positive and OF negative strains. The sequences flanking the hypervariable and C repeats of the *emmST4547* gene showed high similarity to a corresponding region in the *mga* regulon of OF positive strains notably M15, M4, M22 and M50. In contrast, the sequence of the hypervariable and C repeats region of the *emmST4547* gene revealed high similarity to equivalent regions in the OF negative strains. These data suggested that horizontal transfer of *emm*-like genes could occur between OF positive and OF negative strains resulting in divergence in the architecture of the *mga* regulon.

This study showed that sequencing of the 5' region of the *emm* gene of GAS isolates was effective for surveying the sequence variability of the M protein and useful for monitoring GAS strain diversity in Malaysia as well as showing the mechanisms

involved for antigenic diversity in M proteins. This study also illustrated a new mosaic in structure of *mga* regulon of OF negative strains with existence of *mrp* and *emm* genes. As far as this research was concerned to our knowledge, such a study has been done for the first time in a developing country.

Abtrak tesis yang dikemumakan kepada senat Universiti Putra Malaysia sebagai memenuhi keperluan ijazah untuk Doktor Falsafah.

PENCIRIAN MOLEKUL GEN SEOLAH *EMM* DIDALAM REGUKON *MGA* STRAIN 4547 DALAM STREPTOKOKUS KUMPLULAN A

Oleh

MAJID ESHAGHI

Julai 2001

Pengerusi: Profesor Madya Abdul Manaf Ali, PhD

Fakulti: Sains Makanan dan Bioteknologi

Analisis jujukan region 5' dari gen *emm* telah digunakan untuk membezakan 39 isolat streptococci kumpulan A (GAS) yang dipencilkan diantara tahun 1989 hingga 1997 dari pesakit dan pembawa dari Kuala Lumpur. Enam puluh satu paratus (24) daripada isolat mempunyai gen *emm* yang mengkodkan penghasilan protein M yang tidak ada kaitan dengan antigen M. Strain yang selebihnya mempunyai urutan *emm* yang ada kaitan dengan antigen M yang telah direkodkan sebelumnya. Di dalam sesetengah kes terdapat variasi antigenik di antara sesetengah protein M, termasuk strain tempatan yang dikaji, jika dibandingkan dengan jujukan protein M yang telah diterbitkan. Perbezaan ini adalah disebabkan ketidaksamaan bes melalui selitan atau penyingkiran.

Urutan regulon *mga* daripada jenis *emm* ST4547 streptokokus kumpulan A dari Malaysia yang mempunyai faktor opasiti (OF) negatif menunjukkan kemungkinan kehadiran dua gen yang menyerupai gen *emm* (*emm* dan *mrp*). Gen *emm* mengkodkan protein M manakala gen *mrp* mengkodkan IgG reseptor Fc. Gen *emm* terletak sebelum gen *scp*A dan terdiri daripada 1305 nukleotida yang mengkodkan protein M yang menggandungi 435 asid amino dengan berat molekul 49.0 kDa atau protein matang yang terdiri daripada 394 asid amino dengan berat molekul 44.7 kDa. Terdapat satu lagi gen yang dinamakan sebagai *mrp*ST4547 dijumpai

gen ini terdiri daripada 1167 nukleotid yang diramalkan mengkodkan protein yang terdiri daripada 388 asid amino yang mempunyai berat molekul sebanyak 42.2 kDa atau protein matang yang terdiri daripada 347 asid amino yang mempunyai berat molekul sebanyak 37.9 kDa. Regulon *mga* dari strain ST4547 mempunyai struktur mozaik yang terdiri dari segmen-segmen DNA yang berasal daripada strain OF positif dan strain OF negatif yang berbeza. Jujukan yang hadir diantara kawasan hipervariable dan pengulangan C dari *emm* ST4547 gene menunjukkan persamaan yang tinggi dengan jujukan *mga* regulon dari strain OF positif, terutamanya M15, M4, M22 dan M50. Sebaliknya, jujukan hypervariable dan pengulangan C dari gen *emmST*4547 menunjukkan persamaan dengan yang terdapat pada strain OF negatif. Ini menunjukkan bahawa terdapat pemindahan secara mendatar *emm* gen dari strain OF positif ke strain OF negatif yang menghasilkan pemisahan dalam struktur binaan *mga* regulon.

Di dalam kajian ini, penjujukan kawasan 5' gen *emm* dari isolat GAS adalah amat berguna untuk mendapatkan maklumat mengenai variasi setiap jujukan protein M dan untuk mengkaji kepelbagaian GAS di Malaysia seterusnya mengkaji mekanisma yang

terlibat di dalam kepelbagaian antigenik di dalam protein M. Kajian ini juga menunjukkan terdapatnya struktur mozaik yang baru oleh *mga* regulon dari strain OF negatif dengan kehadiran *mrp* dan *emm* gen. Kajian ini adalah yang pertama di lakukan di negara yang sedang membangun.

ACKNOWLEDGMENTS

I would like to express my most sincere gratitude and deep appreciation to my supervisors Associate Prof. Dr. Abdul Manaf Ali, Associate Prof. Datin Dr. Khatijah Mohd. Yusoff, Prof. Datin Dr. Farida Jamal and Dr. Zaiton Hassan for their invaluable contribution, generous offering of time and careful supervision throughout the study.

I am indebted to members of our weekly meeting group, with special appreciation to Dr. Tan Wen Siang, for his invaluable advice and support. I also wish to thank all my fellow graduate students, friends and staff members of the Department of Biotechnology and Department of Biochemistry and Microbiology of UPM.

Furthermore, I would like to acknowledge UPM for supporting the research and providing the Graduate Assistantship by the Malaysian Government through IRPA (Intensification of Research Priority Areas), grant No.

Last but not least, I would like to express my deepest gratitude to my beloved wife, for her understanding, sacrifices, patience and support.

I certify that an Examination Committee met on 23rd July 2001 to conduct the final examination of Majid Eshaghi on his Doctor of Philosophy thesis entitled "Molecular Characterization of *emm*-like Genes in the *mga* Regulon of Group A Streptococcus Strain 4547" in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulation 1981. The Committee recommends that the candidate be awarded the relevant degree. Members of the Examination Committee are as follows:

SON RADU, Ph.D.
Associate Professor
Faculty of Food Science and Biotechnology
Universiti Putra Malaysia
(Chairman)

ABDUL MANAF ALI, Ph.D.
Associate Professor
Faculty of Food Science and Biotechnology
Universiti Putra Malaysia
(Member)

KHATIJAH YUSOFF, Ph.D.
Associate Professor
Faculty of Science and Environmental Studies
Universiti Putra Malaysia
(Member)

FARIDA JAMAL, MBBS.
Professor
Faculty of Medicine
Universiti Putra Malaysia
(Member)

ZAITON HASSAN, Ph.D.
Faculty of Food Science and Biotechnology
Universiti Putra Malaysia
(Member)

SAZALY ABU BAKAR, Ph.D.
Associate Professor
Faculty of Medicine
Universiti Malaya
(Independent Examiner)

MOHD. GHAZALI MOHAYIDIN, Ph.D.
Professor/Deputy Dean of Graduate School,
Universiti Putra Malaysia

Date: 10 AUG 2001

This thesis submitted to the Senate of Universiti Putra Malaysia has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy.

AINI IDERIS, Ph.D.
Professor/Dean of Graduate School,
Universiti Putra Malaysia

Date: 08 NOV 2001

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UPM or other institutions.

Majid Eshaghi

Date: 8.8.01

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

APS	ammonium persulphate
bp	base pairs
CFU	colony forming unit
GAS	group A streptococci
ET	erythrogenic toxin
kDa	kilodalton
HCR	host cross reactive
M_r	relative molecular weight
MLEE	multilocus enzyme electrophoresis
OD	opacity density
OF	opacity factor
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PFGE	pulsed field gel electrophoresis
PSGN	post-streptococcal glomerunephritis
RAPD	randomly amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
SDS	sodium dodecyl sulphate

CHAPTER I

INTRODUCTION

Streptococcus pyogenes of Lancefield group A (group A streptococci; GAS), which is distinguished from other β-haemolytic streptococci on the basis of the antigenic specificity of its cell wall carbohydrate, is a common and important human pathogen worldwide. The infection can occur either as an epidemic or an endemic. The main portal of entry for GAS and their principal site of residence in humans is the upper respiratory tract. Streptococcal pharyngotonsillitis is the most common of all bacterial throat infections. In most instances "streptococcal throat" is a self-limiting infection., but it may progress. A significant percentage of true pharyngeal infection, confirmed by a significant rise in streptococcal antibody titers, are clinically mild or even inapparent. They are, nevertheless, also associated with the risk of late sequelae, and may be active sources of spread of virulent streptococci, in contrast to chronic carriers. Of the primary skin infections caused by GAS, impetigo (pyoderma) is the most frequent, especially in tropical climates. The third site of primary streptococcal infection is the female genital tract, although uncommon, it is still encountered in many countries. Throat, skin and genital infections may develop into life threatening septicemia, streptococcal toxic shock syndrome, or metastatic suppurative infections such as arthritis, osteomyelitis, peritonitis, or even acute endocarditis in some individuals (Denny, 2000).

GAS express a range of cell surface and extracellular products which have the potential to act as virulence factors of which the M protein which is encoded by

emm gene is the most important and is the subject of this thesis. The M proteins were originally defined in the 1920's as type-specific, protective antigens (Lancefield, 1928) which are cell surface protein with conserved, wall-associated C-terminal regions and much more variable N-terminal regions protruding from the cell surface. Based on the antigenic specificity of the cell wall associated M proteins, GAS can be divided into more than 100 M types, provisional types and *emm* types (Facklam et al., 1999). The M protein is the virulence factor which blocks antiphagocytosis *via* the alternative complement pathway (Whitnack and Beachey, 1985). Complete sequencing of the many *emm* and *emm*-like genes show that they all possess a similar overall structure while relationships between these genes vary in detail (Whatmore and Kehoe, 1994). It has been proposed that the evolution of the *emm*-like genes is a very dynamic process, involving intragenic mutational events as well as intergenic recombination (Hollingshead et al., 1986; Hollingshead et al., 1987; Fischetti, 1989; Haanes and Cleary, 1989; Scott, 1990; Harbaugh et al., 1993; Whatmore and Kehoe, 1994; Whatmore et al., 1994).

Haanes et al. (1992) reported that the *emm*-like genes in all strains of GAS are located in the same position in the *mga* regulon locus and are flanked by the *mga* and *scpA* genes. Based on the ability of the GAS strains to produce an apoproteinase, an enzyme that causes mammalian serum to increase in opacity, they have been divided into two distinct groups, OF positive and OF negative strains (Beall et al., 2000). The *mga* regulon in OF positive strains contain *mrp*, *emm* and *enn* genes, whereas in OF negative strains it comprises only *emm* gene (Haanes et al., 1992; Hollingshead et al., 1993). However, the *mga* regulon in OF negative has been shown to be more variable with presence of the H protein in M1 strain (Gomi et al.,

1990) and an *enn* gene in many other OF negative strains (Hollingshead, 1993; Podbielski, 1993; Whatmore and Kehoe, 1994). We believe that the *mrp* gene might also be present in the *mga* regulon of the OF negative GAS strains. Therefore, by designing primers and using PCR we screened the *mga* regulon of OF negative Malaysian GAS strains to address above hypothesis.

Our understanding on the epidemiology of group A streptococcal infections is based primarily on M serotyping. However, it is presently difficult to detect the M protein in this way especially in South East Asia where it is difficult to obtain the appropriate antisera. Moreover, previous studies showed that a large number of GAS in Malaysia are not typeable with the standard M-typing antisera (Jamal et al., 1995; 1999). The usefulness of *emm* gene sequence analysis has been recently evaluated in several epidemiological studies of GAS (Relf et al., 1992; Beall et al., 1997; Jamal et al., 1999). Therefore, OF detection and *emm* gene sequencing are applied to differentiate several local GAS isolates to reveal that non-typeability of GAS in Malaysia by M serotyping is due to existence of new *emm* types or provisional M types. Furthermore, the resulting sequences of the *emm* genes might contribute to a better understanding of mechanisms involved in M protein antigenic diversity. In addition, attempts are made to evaluate the power of randomly amplified polymorphic DNA analysis (RAPD) for typing of the above isolates

CHAPTER II

LITERATURE REVIEW

Epidemiology and Clinical Importance of GAS

The group A streptococcus (*Streptococcus pyogenes*) is responsible for a number of suppurative human infections, of which acute pharyngitis and impetigo are the most common. As a consequence of antibiotic therapy or no therapy, as many as 3 to 5% of individuals who suffer a group A streptococcal pharyngeal infection may develop acute rheumatic fever, a disease often resulting in cardiac damage. While not currently a major problem in developed countries, rheumatic fever is the leading cause of heart disease in school-aged children in developing nations (Kaplan, 1993). Acute glomerulonephritis, another sequelae of group A streptococcal disease, is usually the consequence of infection by specific strains of streptococci (nephritogenic strains) which infect either the throat or skin (Rammelkamp and Weaver, 1953). The ability of group A streptococci to persist in infected tissues is primarily due to the cell surface M protein, a molecule which confers to the streptococcus the ability to resist phagocytosis by polymorphonuclear leukocytes in the absence of type-specific antibodies to the M molecule (Lancefield, 1959; Lancefield, 1962). Since there are more than 100 different serotypes of M protein (such as M5, M6, M24), an individual may become infected by more than one group A streptococcal type during a lifetime (Lancefield, 1962).

The incidence of acute rheumatic fever and severe group A streptococcal (GAS) infection declined dramatically in the Western Hemisphere during the post-antibiotic era (Colman et al., 1993). Although the precise reasons are not known, various factors contributed towards this, possibly including improved standards of living and better health care. However, in the late 1980s, increase in the number of serious systemic infections, particularly associated with streptococci of M type 1 (M1), have been reported from the United States, Great Britain, Norway and Sweden (Beachey and Seyer, 1986). In the United States, the proportion of M types 1, 3 and 18 increased significantly and by contrast, M types 4 and 12 decreased. Similar changes in M type distribution and severity of GAS infection were also observed in England (Colman et al., 1993). These data suggest that the changes in the epidemiology of GAS infection are partly due to changes in the organism itself. GAS infection and its sequelae remain endemic in many Asian countries. However, no increase in the incidence and severity of GAS disease has been documented (Jamal, 1996). Although under-reporting could not be completely ruled out, it is unlikely that a change has gone unnoticed. Several other factors may account for this difference in the epidemiology of GAS disease. These include immunity towards an emergent clone, rendering it less virulent, or preventing it from colonizing the population. A study conducted in Thailand suggests that the M protein of GAS prevalent in this region may be different from those implicated in the recent resurgence in the West (Tran et al., 1994). GAS infections are endemic in aboriginal communities of Northern Australia, with up to 75% of children having impetigo due, in part, to infection of scabies lesions. The reported rates of acute rheumatic fever (ARF) and rheumatic heart disease (RHD) are some of the highest reported anywhere in the world. Acute streptococcal glomerulonephritis (APSGN) occurs frequently