MATHEMATICAL MODELLING OF SOME UNIFORM AND NULL-CONTEXT SPLICING SYSTEMS

LIM SHUN JINN

UNIVERSITI TEKNOLOGI MALAYSIA

MATHEMATICAL MODELLING OF SOME UNIFORM AND NULL-CONTEXT SPLICING SYSTEMS

LIM SHUN JINN

A dissertation submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Science (Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

April 2011

To my beloved parents, sisters, brother and friends.

Thank you for all your love and support.

ACKNOWLEDGEMENT

First of all, I would like to express my sincere appreciation to my dissertation supervisor, Dr. Fong Wan Heng for her guidance, encouragement and patience throughout the completion of this research. She was always there to listen and to give advice for my research. She showed me different approaches to solve research problems and inspired me to be persistent in conducting this research. Without her support and guidance, I could not have finished this dissertation.

I am also thankful to all my friends and coursemates who have given me help and comments on this dissertation. I thank them for helping me to review my work during the preparation of this dissertation. I would like to acknowledge all their support, guidance, patience and understanding throughout the completion of my dissertation.

Last but not least, I would like to thank my parents and family members for giving me moral support and inspiring me to complete this dissertation.

ABSTRACT

Head first proposed the mathematical modeling of splicing system which involves recombination of DNA molecules in 1987. Splicing of DNA involves cutting of DNA molecules using the restriction enzymes and reassociating different fragments of DNA molecules using the ligases under some specific chemical conditions. A splicing language, L is generated if there exists a splicing system S for which $L = L(S)$. Some splicing languages namely persistent splicing language and strictly locally testable language are equivalent to uniform splicing language. Since uniform splicing system is also a null-context splicing system, hence, uniform and null-context splicing systems are discussed in this research. Besides, some molecular examples on uniform and null-context splicing systems with different initial strings and combination of restriction enzymes are presented. Applications of automata theory on these molecular examples of uniform and null-context languages are also presented in this research.

ABSTRAK

Head mula mencadangkan permodelan matematik sebagai sistem pemotongan dan pencantuman yang melibatkan penyatuan molekul DNA pada tahun 1987. Pemotongan dan pencantuman bagi DNA melibatkan pemotongan molekul DNA dengan menggunakan enzim pembatas dan menggabung balik serpihan-serpihan molekul DNA yang berlainan dengan menggunakan ligase pada sesetengah keadaan khusus bahan kimia. Satu bahasa pemotongan dan pencatuman, L dihasilkan jika terdapat sistem pemotongan dan pencatuman S bagi $L = L(S)$. Sesetengah bahasa pemotongan dan pencantuman iaitu bahasa pemotongan dan pencantuman berterusan dan bahasa kebolehujian setempat tegas adalah setara dengan bahasa pemotongan dan pencatuman seragam. Oleh kerana sistem pemotongan dan pencatuman seragam juga merupakan sistem pemotongan dan pencatuman berkonteks batal, maka, sistem pemotongan dan pencatuman serupa dan berkonteks batal dibincangkan dalam penyelidikan ini. Selain itu, beberapa contoh bermolekul pada sistem pemotongan dan pencatuman seragam dan berkonteks batal dengan jujukan permulaan yang berlainan dan penggabungan enzim-enzim pembatas juga diberikan. Aplikasi teori automata untuk contoh-contoh bermolekul pada bahasa pemotongan dan pencatuman seragam dan berkonteks batal juga ditunjukkan dalam penyelidikan ini.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

A	-	Alphabet, Σ
A^p	-	Set of strings of length P
A^*	-	Strings obtained by concatenating zero or more symbols from A
B	-	Set of all patterns associated with enzymes that either produce 5′ overhangs or blunt ends
C	-	Set of all patterns associated with enzymes that produce 3′ overhangs
F	-	A set of final states
I	-	Initial string
L	-	Language
$L(M)$	-	Language accepted by a dfa or nfa, M
$L(S)$	-	Language generated by a splicing system S
P	-	Positive integer
Q	-	A finite set of internal states
q_0	-	The initial state
r	-	Regular expression
S	-	Splicing system
x	-	Crossing
$\{\}$, $\emptyset$	-	Empty set
$\in$	-	Element of
$\subseteq$	-	Subset of
$\cap$	-	Intersection
$\cup$, $+$	-	Union
$\neq$	-	Not equal to
$\cdot$	-	Concatenation

*	-	Star-closure
▼, ▲	-	Cutting site of restriction enzymes
$\|w\|$	-	Length of a string w
$\|w\|_n$	-	number of occurrence for n in a string
λ	-	Empty string
δ	-	The transition function
δ^*	-	Total of transition function
2^Q	-	Power set of Q