APPLICATION OF PROFILE HIDDEN MARKOV MODEL-BASED APPROACH USING HMMER3 ON MAJOR HISTOCOMPATIBILITY COMPLEX (MHC) CLASS II BINDING PEPTIDES

MAZLINA ISMAIL

FACULTY OF SCIENCE

UNIVERSITY OF MALAYA

KUALA LUMPUR

2012

APPLICATION OF PROFILE HIDDEN MARKOV MODEL-BASED APPROACH USING HMMER3 ON MAJOR HISTOCOMPATIBILITY COMPLEX (MHC) CLASS II BINDING PEPTIDES

MAZLINA ISMAIL

(SGJ 100011)

SUBMITTED TO

INSTITUTE OF BIOLOGICAL SCIENCES

FACULTY OF SCIENCE

UNIVERSITY OF MALAYA

IN PARTIAL FULFILMENT

OF THE REQUIREMENTS FOR

THE DEGREE OF MASTER OF BIOINFORMATICS

2012

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to thank Dr. Saharuddin Mohamad for supervising my work this past semester. I would also like to thank Dr. Khang Tsung Fei for throwing some ideas in my way at the beginning stage of the research work; those nuggets of information helped me a long way down the road. Last but not least, to my family and friends for their never-ending support and encouragement.

Abstract

Major histocompatibility complex (MHC) Class II molecules play an important role in the immune system where it presents the antigenic peptides on the cell's surface for the next chain of immune response. Therefore, the process of identifying these peptidic fragments are of interest in the pipeline of vaccine design. Several probabilistic methods have been proposed for the prediction of MHC-binding peptides and one of them is the hidden Markov model. Profile hidden Markov models is a profiling technique which applies the statistical method to estimate the true frequency of a residue at a given position within the alignment. This study applies profile hidden Markov model-based approach using the HMMER3 package to train and build the profile based on the MHC-binding dataset. Evaluation of the profile HMM using a dummy dataset reveals that the profile HMM is able to differentiate the true positive data. This shows that the profile hidden Markov model approach can be a potential supporting method to identify peptides that bind to the MHC Class II.

Abstrak

"Major histocompatibility complex Class II" atau MHC Kelas II merupakan molekul yang memainkan peranan penting dalam sistem keimunan di mana ia memaparkan peptida antigenik di atas permukaan sel untuk tindakan keimunan yang seterusnya. Oleh itu, proses pengenalpastian peptida antigenik tersebut merupakan kaedah yang boleh disertakan dalam rekabentuk vaksin. Beberapa kaedah kebarangkalian telah dicadang untuk peramalan peptida yang akan mengikat pada kompleks MHC dan salah satu kaedah tersebut adalah "hidden Markov model". "Profile hidden Markov model" merupakan kaedah pemprofilan yang menggunakan kaedah statistik untuk menganggar frekuensi sebenar sesuatu residu pada kedudukannya di dalam penjajaran. Kajian ini mengaplikasikan kaedah "profile hidden Markov model" menggunakan pakej HMMER3. Pakej ini digunakan untuk melatih dan membina profil berdasarkan data yang mengandungi peptida yang mengikat pada molekul MHC. Penilaian ke atas profil menggunakan set data menunjukkan bahawa profil tersebut dapat mengenalpasti set data yang "true positive". Ini menunjukkan bahawa kaedah "profile hidden Markov model" boleh dijadikan sebagai salah satu kaedah untuk menyokong pengenalpastian peptida yang mengikat pada kompleks MHC kelas II.