UNIVERSITY OF NAIROBI

SCHOOL OF COMPUTING AND INFORMATICS

Polyepitope design optimization system: A web-based application prototype.

By

MAINGI, JULIUS KAIRITHIA

P56/P/8144/03

SUPERVISOR

LAWRENCE MUCHEMI

March 2012

Submitted in partial fulfilment of the requirements of the Master of Science in Information Systems

Declaration

This project, as presented in this report is my original work and has not been presented for any other University award

Signed: …………….............................   Date…………………………….

Name: JULIUS KAIRITHIA MAINGI

Reg/No: P56/P/8144/03

This project has been submitted as part fulfilment for Master of Science in Information systems of the University of Nairobi with my approval as the University supervisor

Signed: …………….........................   Date…………………………….

LAWRENCE MUCHEMI

Project Supervisor

School of Computing and Informatics

University of Nairobi

Dedication

This project is dedicated to my wife Mrs Eunice M. Kairithia

Acknowledgment

First, I thank the Almighty God for the strength, grace and good health throughout this research project and during my studies at the School of Computing and Informatics.

Secondly, my sincere gratitude goes to my supervisor Mr. Lawrence Muchemi who has been very supportive and helpful in providing academic guidance throughout the entire project. I appreciate the panelists of panel 2 September 2011 group.

The School of Computing and Informatics for the knowledge and skills imparted to me in course of my studies in this school.

I would also like to appreciate my family members. I acknowledge the support of my friends, especially Dr Moses Kindiki, Dr Muriithi Kigunda, Hon Dr Kilemi Mwiria, Mr Henry Mwenda, Mr Daniel Muriuki, Mr Muriuki Tarichia, and Mr. Geoffrey Muriira Karau.

Abstract

Immunogenic epitopes in a polyepitope construct may induce immunity against multiple antigenic targets only if epitopes are correctly processed and presented. Applications that combine multiple supertypes and integrate variables that measure the quality of polyepitopes can be used to make better polyepitopes yet there are no publicly accessible tools. In this research project, the researcher developed a web-based polyepitope optimization prototype. The prototype runs on 3-tier application architecture: MYSQL was used for the database tier, PHP for the application tier that runs on Apache HTTP server and the presentation layer was implemented on a web browser. The researcher collected epitopes data and information from online databases; the Immune Epitope Database (IEDB), HIV molecular immunological database and National Center for Biotechnology Information (NCBI).The epitopes were analyzed using Microsoft Excel 2003 and uploaded onto a local database. Polyepitopes were generated, optimized and tested on an online web server NetChop 3.1 which was used to validate one of the polyepitope quality measurement variables- proteasome cleavage predictions. The results of data analysis indicate that the immune system recognize epitopes in clusters and the main clusters are 9-mers(9 amino acid long), 15mers,20mers and 25mers.The polyepitope optimization results indicate that polyepitope optimization algorithms that integrate proteasome cleavage prediction, transporter associated with antigen processing (TAP) binding prediction and Major Histocompatibility Complex (MHC) binding predictions can be used to develop better polyepitopes which can be used to make better vaccines within a shorter time and at a lower cost.

Key words: polyepitope, epitopes, optimization, mer, TAP, proteasome, MHC, prototype.

TABLE OF CONTENTS

List of tables

List of figures

List of abbreviations

WHO-World Health Organization

T.B. – Tuberculosis

B.C.G.-Bacille Calmette Guerine

NCBI – National Center for Biotechnology Information

<div align="center">**CHAPTER ONE: INTRODUCTION**</div>

1.0 Introduction

Polyepitope vaccines are novel approaches of developing vaccines. This chapter gives the background of the project, statement of the problem, research objectives, gives the significance of the study and gives the definitions of terms used in the project.

1.1 Background of the study

Pathogens have been a threat to humans for a long time. RAC (2002) indicates that there are more than 400 microbial agents associated with disease in healthy adult humans. The World Health Organization (WHO) report (2004a) notes that an estimated 11 million (19%) of the 57 million people that died in the year 2002 of infectious or parasitic infections. The three main infectious diseases are, HIV/AIDS, Tuberculosis (TB) and malaria each of which causes more than a million deaths. AIDS, which is caused by HIV, is now the leading cause of deaths among the young adults world wide.

WHO (2004b) report notes that tackling HIV/AIDS is the world's most urgent public health challenge, more than 20 million people have died from AIDS and an estimated 34- 46 million people are now infected with the HIV. There is no vaccine or a definite cure for AIDS.

The T.B. is another emerging public health threat. TB caused by *Mycobacterium Tuberculosis* (mtb) and is spread from person to person by airborne droplets expelled from lungs when a person with TB coughs sneezes or speaks. Mtb infected person can develop TB if the immune system is impaired by for example HIV infection. In 1995 according to a WHO (2004) report 31 million people died of TB world wide.

Currently there is only one licensed vaccine against TB in the United States of America but it is not recommended for use, this vaccine *Bacille Calmette Guerine* (BCG) is reportedly highly variable in its efficacy to prevent adult pulmonary TB. BCG may have lower efficacy in poor tropical societies where population is more exposed to other Mycobacterium in the environment. According to NIAID (2000) report there is need to develop an improved anti-TB vaccine for adequate control of multi-resistant TB (MDR-TB) among other reasons.

The immune system is the system that defends the body against disease causing organisms, malfunctioning cells, and foreign particles. The major task of the immune system is to defend the host against infections, an assignment that is essential to the survival of the organism. The challenge arises because pathogens and diseases are constantly evolving or mutating to escape

the human defence mechanism. In order for the immune system to defend the host it must identify the invading pathogen and initiate a process to eliminate the antigen. The specific sequences in an antigen that are recognized by the components of the immune system are known as epitopes. Vaccines are developed to mimic the working of the adaptive immune system.

It is not cost effective to develop vaccines against all those microbes. Also, it takes a very long time to establish the virulence of a pathogen and therefore it might be dangerous to use partial or complete organism.

Immunological bioinformatics can be used to make an important contribution to the rapid design of novel vaccines by identifying the most immunogenic regions of the pathogen. These regions can subsequently be used as candidates for rational vaccine design.

Polyepitope vaccines are new approaches in vaccine design and development. According to Suhrbier (2002) a short epitope sequence is often capable of inducing protective immunity against a large and complex pathogen. A polyepitope construct may induce immunity against multiple antigenic targets, multiple strain variants or even multiple pathogens such immunization is highly relevant to induce protection against organisms like *HIV* or *Epstein Barr* virus where the immune escape is an important issue or for cancer treatment where immunization with subdominant epitopes might be effective in breaking tolerance. The development of polyepitopes is based on the knowledge of the immunogenic peptides also called epitopes. Such epitopes form the key components of the polyepitope. The number of epitopes that can be included in a poly epitope are limited due to practical and economic reasons. A polyepitope that in effective in one person may ineffective in another patient because every person posses a set of MHC class I and MHC class II molecules of different specificity. This makes the selection of optimal set of epitopes an important and interesting optimization problem. There has been exponential growth of tools and databases that store epitopes. This has not been coupled with a corresponding increase in epitope-based vaccines. The epitopes are available in public database. The concatenation of epitopes does not lead to the generation of an optimal polyepitopes. There is need to develop polyepitope optimization tools and carry out more research in this area. Some of the databases are shown in **Table 1** below

Table 1: Web-Based tools for vaccine design

Database	Principal investigators	url	Description
SYFPETHI	Rammansee	www.syfpeth.de	Natural MHC ligands
MHCPEP	Brusic, Harrison	Wehih.wehi.edu.au/mhcpep	MHC binding Peptides
JenPep	Flower	www.jenner.ac.uk/jenpep2	MHC and TAP ligands, B cell epitopes
HIV Molecular Immunological database	Korber	Hiv-web-lanl.gov/content/immunology	HIV epitopes
EPIMHC	Reihertz	Immunax.dfci.havard.edu/tools/db_query_epimhc.html	Mhc ligands
Immune epitope database	Lund O	www.immuneepitope.org http://iedb.org	Epitopes Prediction tools

Source: Lund et al (2005)

SYFPETHI database contains more than 4000 peptides sequences known to bind to MHC class I and II molecules based on previous publications on T cell epitopes and MHC Ligands according to Rammansee et al (1999). JenPep has more than 8000 entries and contains quantitative binding data of peptides to MHC and TAP as well as B and T epitopes as noted by Blithe et al (2002). HIV Molecular Immunological database contains a CTL and T helper epitopes and antibody binding sites for HIV -1 epitopes according to Korber et al (2001a). This epitope database does not contain epitopes for HIV-2. EPIMHC is an MHC ligand database can be searched based on sequence length, class, and species and whether a ligand is an epitope or not, ultimately it is being graded to a publicly accessible immune epitope database containing linear and structural antibody epitopes and T cell epitopes. MHCPEP database contains peptides that bind MHC

 The web based tools described above have been developed to provide epitopes data for researchers in immune system and immunological bioinformatics. There is need to develop tools that facilitates research in polyepitope vaccines and promote sharing of research data on poly epitopes.

1.3 Statement of the problem

Polyepitope optimization algorithms that integrate proteasome cleavage, TAP binding and MHC class I binding predictions can used to make better polyepitopes yet there are no publicly accessible applications on the World Wide Web which can be used by researchers. In this project, the researcher developed a web-based polyepitope application prototype that can be used to optimize the polyepitopes and make the results publicly accessible to researchers.

1.4 Research Objectives

The objectives for this research project were:

1. to compare epitope lengths for pathogens with vaccines and pathogens without vaccines
2. to suggest most appropriate length(s) of epitopes for use in a poly epitopes for malaria, HIV-I and T.B
3. to develop and maintain epitopes in an epitopes database.
4. to optimize polyepitope design
5. to develop a web-based application prototype

1.4 Significance of the study

In this research project, the researcher created a polyepitope optimization prototype that generates publicly accessible results. The prototype can be used to make better polyepitopes which can minimize the cost and time required to develop epitope-based vaccines. The prototototype will spur interest in computer scientists to develop tools that can be used in bioinformatics, immunological bioinformatics and computational biology. The prototype will benefit researchers in bioinformatics, immunological bioinformatics and computational biology and vaccine developers.

1.5 Definition of terms

Pathogen-something that can cause disease, such as a bacterium or a virus

Microbe-a microscopic organism, especially one that transmits a disease

Vaccine-substance introduced into the body in order to achieve protection through an immunological reaction against specific micro-organisms that cause a number of infectious diseases.

Efficacy-ability to produce the necessary or desired results

Mycobacterium- a rod like Gram-positive aerobic bacterium that can form branching structures resembling filaments. Some cause diseases in humans, for example tuberculosis or leprosy.

Immune system-the system that defends the body against pathogens, malfunctioning cells, and foreign particles by recognizing and killing pathogens and tumor cells

Epitope-The specific sequences in an antigen that are recognized by the components of the immune system

Antigen-Any agent perceived as foreign by the body's immune system

Immunization- Process of rendering people immune to an infectious organism by inoculating them with a form of the organism that does not cause severe disease but does provoke formation of protective antibodies

Peptide- one of a group of organic chemicals found in most living tissues, with a wide range of biological functions. The chemicals are relatively low-weight polymers of amino acids, as contrasted with the high-weight proteins. The acids are linked together by so-called peptide bonds between their carboxyl (COOH) and alpha amino (NH2) groups.

Ligands-ion or molecule that is bonded by a coordinate bond to a central transition metal ion in a complex

Immunoglobulin- antibody

TAP-transporter associated with antigen processing

Protein - one of a large group of nitrogen-rich compounds of high molecular weight that are essential and abundant constituents of living organisms

Lymphocytes-white blood cells

Antibody- any normally occurring protein molecule that is produced in the body of cells called lymphocytes and that act primarily as a defence against invasion by foreign substances.

Immunogenic- creating immunity or immune response

Polyepitope- also called **polytope** is a vaccine construct made up of a sequence epitopes and linker amino acid sequences.

Proteasome-large complex in the nucleus and cytosol where selective protein degradation occur

<h1 style="text-align:center">CHAPTER TWO: LITERATURE REVIEW AND THEORY</h1>

2.0 Introduction

This chapter gives an overview of the animal cell, the immune system, MHC class I and MHC class II pathways, systems theory, give a definition of vaccine, types of vaccines, design and development of vaccines, approaches to polyepitope optimization and discusses genetic algorithms.

2.1 The cell

The cell is smallest unit of an organism that can function independently. All living organisms are made of cell(s). Cells are alive and abilities such grow, move, reproduce and respond in an informed way to the environment. The cell contains many molecules- the main one is DNA which found in the nucleus.

Figure 1: The Animal Cell
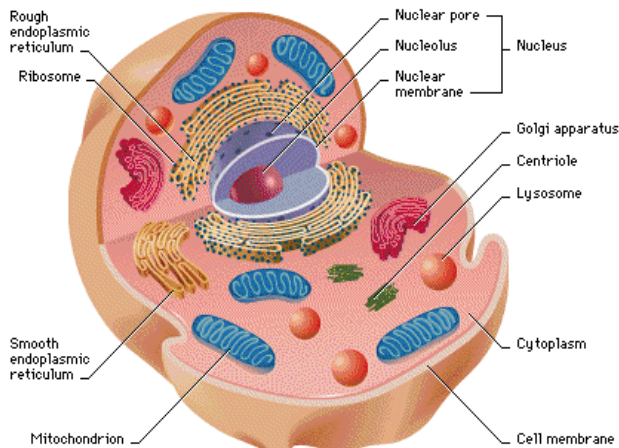
Source: Encarta, (2005) Microsoft encyclopaedia

The DNA (deoxyribonucleic acid) is the molecule that acts as the mechanism of biological inheritance in almost all living creatures. It is found in nearly all cells and contains the coded instructions that control the workings of the cell. DNA is passed from parents to offspring, and contains the coded instructions that enable the offspring to develop from a single cell into an adult body. The central concept in molecular biology is illustrated using the **Figure 2**.

Figure 2: The DNA
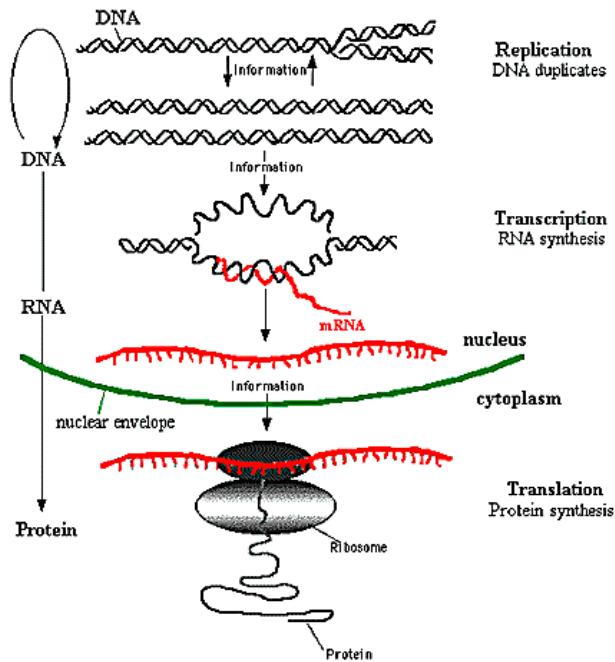
Source: www.

The DNA contains all the information required to manufacture proteins needed by the cell through a three-phase process: duplication, transcription and translation.

2.1.1 The cell Surface

The surface of each cell contains receptors. Receptors can use by pathogens to gain entry into the host cell. Receptors are also used by the cells of the immune system.

Figure 3: The cell surface

Source: Encarta, (2005) Microsoft encyclopaedia

2.1.2 The amino acids

These are a class of organic compounds that contain both the amino (NH_2) and carboxyl (COOH) groups. Of these acids, 20 serve as the building blocks of proteins. Known as the alpha, amino acids, they comprise alanine(**A**), arginine(**R**), asparagine(**N**), aspartic acid(**D**), cysteine(**C**), glutamic acid(**E**), glutamine(**Q**), glycine(**G**), histidine(**H**), isoleucine(**I**), leucine(**L**), lysine(**K**), methionine(**M**), phenylalanine(**F**), proline(**P**), serine(**S**), threonine(**T**), tryptophan(**W**), tyrosine(**Y**), and valine(**V**). All 20 are constructed according to a general formula:

The amino and carboxyl groups are both attached to a single carbon atom, which is called the alpha carbon atom. Attached to the carbon atom is a variable group (R); it is in their R groups that the molecules of the 20 standard amino acids differ from one another. The acidic amino acids **DE** and are coloured red, the basic amino acids-HKR and are coloured blue hydrophobic amino acids are **ACFILMPVW** are coloured black neutral are GNQSTY and are coloured green

2.1.3 Protein synthesis

Protein is the complex organic nitrogen-rich substance found in the cells of all animals and plants. Twenty (20) different amino acids can linked together in linear polymers, known as polypeptide chains. The following **figure 4** illustrates the synthesis of protein

Figure 4: Protein synthesis

Source: www

2.2 The human immune system

The human immune system is the body system that is primarily responsible for destroying the disease-causing agents that it encounters. The responsibility of the immune system is immense, and it must encompass a vast diversity in order to react appropriately with the thousands of different and potentially disease-causing antigens that invade the body.

The immune system has six major components, three of which are different kinds of cells and three of which are soluble proteins. The three broad categories of immune cells are granulocytes, monocyte/macrophages, and lymphocytes. The three kinds of proteins in the immune system, found dissolved in the liquid portion of the blood, are immunoglobulins, cytokines, and complement proteins. There are thousands of different kinds of immunoglobulins, called antibodies and each of them combines exactly with one specific kind of antigen and helps remove it from the body. All six components can be found circulating in the blood in some form.

The vertebrate immune systems have two basic branches. The main deference between the two lies in the means by which they recognize pathogens. Innate immune system which according to Fearon and Locksley (1996) distinguishes harmful and innocuous (harmless) agents according to the carbohydrate signals.

The adaptive immune system is induced by lymphocytes and can further be divided into: humoral immunity which is mediated by antibody molecules secreted by B lymphocytes that can neutralize pathogens outside cells and cellular immunity which is mediated by T lymphocytes that eliminate the infected cells and provide help to other immune responses.

At the heart of the immune response is the ability of the immune system to distinguish between self and non-self. Every cell in the body carries the same set of distinctive surface proteins that distinguish between self from non-self.
The immune systems cells do not attack own body tissues which carry the same pattern of self markers they coexist peacefully in a state of self tolerance(unless due to autoimmune disease).

The set of unique self markers on human cells is called Major Histocompatibility Complex (MHC) proteins.
There are two classes of MHC proteins: MHC Class I, which are on all the cells and MHC class II proteins which are only on certain cells. Any non- self substance capable of triggering immune

response is called an antigen. An antigen can be a whole non-self cell, a bacterium, a virus, an MHC protein marker or even a protein from a foreign.

The distinctive markers on antigens that trigger immune response are called epitopes. The cells of the immune system recognize MHC proteins when they distinguish between self and non-self. MHC proteins serve as a scaffold–framework that presents peptides from foreign proteins to the immune cells. MHC marker proteins are as distinct as blood types in humans MHC class I bear markers out of possible 200 variations and MHC class II bear 8 markers out of 230 possibilities.

The MHC class I protein molecules which are also called HLAs (human leukocyte antigens) are encoded by 3 different loci on the genome called A, B and C each of the genes is highly polymorphic and for each locus hundreds of different alleles exists. Each allele binds a very specific set of peptides. The alleles can clustered into 9 super types. Alleles within the same super type exhibit the same peptide specificity.

Majority of peptides binding to the HLA complex have a length of between 8 to 10 amino acids (8mers to 10mers). For 9mers, positions 2 and 9 are very important for binding to most class I HLAs. There are two main pathways of processing and presenting antigens to T-Cells. The presentation of peptides to the immune system cells is done by MHC molecules. There are MHC class I and MHC class II molecules. The processing steps are simple sequence analysis performed by the components of the immune system and these steps can be modelled using bioinformatics approaches.

2.2.1 MHC Class I Pathway
In this pathway a protein is cleaved (broken) down by a protein complex called proteasome into peptide fragments, binds to TAP in order to be translocated to ER, bind to MHC class I molecule and thereafter transported to the cell surface. For a peptide to be an epitope it must go through this process more efficiently than other peptides in the cell. The specificity of given molecules can be predicted from the amino acid protein sequences and this can be used to select epitopes to include in a vaccine and help to understand the immune system

2.2.2 MHC Class II Pathway
Presentation of MHC class II molecules follows a different path, after synthesis and translocation and processing peptides are transported to the cell surface. The specificity of given molecules

can be predicted from the amino acid protein sequences and this can be used to select epitopes to include in a vaccine and help to understand the immune system

2.3 Vaccine design and development

A Vaccine is a substance introduced into the body in order to achieve protection through an immunological reaction against specific micro-organisms that cause a number of infectious diseases. The process of making vaccines is lengthy and expensive. According to Andre (2002) the current cost of vaccine making process from concept stage to market stage cost between $200 and $500 million.

2.4 Types of vaccines

There are different types of vaccines these are: live vaccines which are able to replicate in the host but attenuated (weakened) micro-organisms and do not cause disease, inactivate or killed micro-organisms (subunit) and genetic vaccines. Genetic vaccines carry one or more epitopes rather than whole organisms

2.5 Polyepitope construct

Polyepitope also called polytope is a vaccine construct made up of a sequence epitopes and linker amino acid sequences. The number of epitopes included in a poly epitope is usually kept small up to a dozens epitopes. Linker or flanking amino acid sequences are inserted between to improve the quality of the polyepitope. There are three types of epitopes that make the polyepitope construct. The B epitopes which are recognized by antibodies, The T-helper (Th) epitopes are recognized by T-helper cells and the T-CTL epitopes are recognized by the cytotoxic T lymphocytes.

In order to predict epitopes using computational methods the following steps are used. Analyse of protein sequence. For B-cell epitopes, predict secondary structure, hydrophobicity, mobility/flexibility and for T-cell epitopes, search for anchor residues (amino acids) and predict proteasomal cleavage. For a 9-mer (9 amino acid length) epitopes, positions 2 and 9 are important. For example in measles T cell epitope whose id is **32029** and the sequence is **KLMPNITLL** amino acids positions 2(L) and 9(L) are important. The amino acids in those positions are the anchor residues. **Figures 8** and **9** give detailed information on which residues are important for each of the selected super types. Good polyepitopes constructs are those that ensure large population coverage and the number epitopes in a polyepitope that are correctly processed and presented represent most super types. The super types **A2, A3** and **B7** cover 83%

to 88% percent of the global population while super types **A1, A2, A3, A24, B7, B27, B44, B58** and **B62** cover 98.1% to 100% of the global population.

According to Suhrbie (2002) a short epitope sequence is capable of inducing protective immunity against a large and complex pathogen. Including several such immunogenic epitopes in a polytope construct may induce immunity against multiple antigenic targets, multiple strain variants or even multiple pathogens. Such immunization is highly relevant to induce protection against organisms like HIV or *Epstein-Barr* virus where immune escape is an important issue, or for cancer treatment where immunization with subdominant epitopes might be effective in breaking tolerance.

According to Flower et al (1999) the design of a polyepitope can be done in an effective way by modifying sequential order of different epitopes and liker amino acid(s) that will favour proteasome cleavage and transport/translation in ER by TAP complex.

2.6 Approaches to poly epitope optimization

There are several ways that have been used to computationally build and optimize a polytope constructs. There include:

2.6.1 Quantitative matrix (QM)

This approach of polyepitope optimization has been used in TPREDICT application. The QM based approach which predicts MHC-binding prediction is illustrated using the peptide **RLRPGGKKK**. At first, this peptide is encrypted with sparse encoding: it is represented as an array with shape i*j (20*9) where each nonzero element at position (i,j) codes for amino acid i at position j:

```
# A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y
{{0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0},
 {0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0},
 ...
 {0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0}}
```

If amino acid properties are used to parametrize peptide, selected scale "P" of properties is multiplied with obtained sparse matrix element wise to produce an array {{P1},{P2},...{P9}} (all zero elements are removed). This array is multiplied with predictive model matrices element wise. According to the type of selected model, either multiplicative or additive, elements of resulted matrix are either multiplied or summed up. The resulting score after correction (either multiplication with corrective coefficient or summation with corrective constant according to the

type of used model) is compared to the selected threshold; if score is greater than threshold, peptide is considered to be a binder.

2.6.2 Quadratic programming

This method has been used by Lilliana et al (2003). In this model assigns epitopes a binding strength greater than the minimum binding strength of known epitopes. The non epitopes are assigned binding strength less than the minimum binding strength of known epitopes. The high binders have a higher binding strength than the low and medium binders and that the medium binders have a higher binding strength than the low binders.

2.6.3 Integer linear programming

Toussaint N et al (2008) created a mathematical model for the selection of optimal set of peptides for a epitope-based vaccine. The main idea behind this approach is that starting from target antigens, a list of properties of interest, and a target population the information necessary to determine an optimal set of epitopes is derived. Given this information, a mathematical framework can be used to find a set of epitopes that is optimal to the target population and properties of interest. Using this approach a set of optimal epitopes for inclusion in a polyepitope can be defined, however the approach does not take into the fact that epitope-epitope interaction (epitasis) in a polyepitope

2.6.4 Monte Carlo metropolis simulations algorithm

This method relies on repeated random sampling to compute their results. This method is most suited to calculation by a computer and tends to be used when it is infeasible to compute an exact result with a deterministic algorithm. The method is also used to complement the theoretical derivations. The Monte Carlo method was coined in the 1940s by John von Neumann, Stanislaw Ulam and Nicholas. PolyCTLDesigner developed by Denis V. Atonets, is an application developed optimizes a polyepitope construct for HIV-1 and does not run on a web server. The system utilizes Monte Carlo Metropolis Simulations an algorithm developed by Metropolis et al (1953). Metropolis Monte Carlo metropolis simulations algorithm was first used by scientists on the Manhattan project to calculate the probability with which a neutron from one fission Uranium1 atom would cause another to fission according to Feynman (1985). There are plans to upgrade PolyCTLDesigner according to its developer.

13

2.7 Epitope based tools and algorithms

There are various epitope-based tools and algorithms available to researchers among them are: The SYFPETHI database which contains more than 4000 peptides sequences known to bind to MHC class I and II molecules based on previous publications on T cell epitopes and MHC Ligands according to Rammansee et al (1999). JenPep has more than 8000 entries and contains quantitative binding data of peptides to MHC and TAP as well as B and T epitopes as noted by blithe et al (2002).

HIV Molecular Immunological database contains a CTL and T helper epitopes and antibody binding sites for HIV -1 epitopes according to Korber et al (2001a). This epitope database does not contain epitopes for HIV-2.

EPIMHC is an MHC ligand database can be searched based on sequence length, class, species and whether a ligand is an epitope or not ultimately being graded to a publicly accessible immune epitope database containing linear and structural antibody epitopes and T cell epitopes. MHCPEP database contains peptides that bind MHC.

The immune epitope database (IEDB) is a publicly accessible web based tool that contains epitope database. It contains linear and conformational antibody and T cell epitopes.

2.8 Theoretical framework

2.8.1 Systems theory

The researcher used systems theory to model the immune system components and processes. Systems theory was first advanced by von Bertalanffy when he presented his idea of a General System Theory in a philosophy seminar at the University of Chicago in 1937.Generally a system consists of complex interacting components together with the relationships among them that permit the identification of a boundary-maintaining entity or process. The theory was used the theory to understand the components and processes of the immune system.

In this research project, the researcher used the models of Toes et al (2001) to create an algorithm for proteasome cleavage prediction and Peters et al (2003) to develop an algorithm for TAP binding prediction algorithm

2.9 Genetic algorithms

A genetic algorithm (GA) is a search technique used in computing to find exact or approximate solutions to optimization and search problems. Genetic algorithms are categorized as global search heuristics. Genetic algorithms are a particular class of evolutionary algorithms (also

known as evolutionary computation) that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover (also called recombination)

2.9.1 Genetic algorithms implementation

Genetic algorithms are implemented as a computer simulation in which a population of abstract representations (called chromosomes or the genotype of the genome) of candidate solutions (called individuals, creatures, or phenotypes) to an optimization problem evolves toward better solutions. Traditionally, solutions are represented in binary as strings of 0s and 1s, but other encodings are also possible. The evolution usually starts from a population of randomly generated individuals and happens in generations. In each generation, the fitness of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population (based on their fitness), and modified (recombined and possibly randomly mutated) to form a new population. The new population is then used in the next iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population. If the algorithm has terminated due to a maximum number of generations, a satisfactory solution may or may not have been reached.

A typical genetic algorithm requires two things to be defined: the genetic representation of the solution domain and the fitness function to evaluate the solution domain.

2.9.2 Simple genetic algorithm pseudo code

- Choose initial population
- Evaluate the fitness of each individual in the population
- Repeat until termination: (time limit or sufficient fitness achieved)
    - Select best-ranking individuals to reproduce
    - Breed new generation through crossover and/or mutation (genetic operations) and give birth to offspring
    - Evaluate the individual fitness of the offspring
    - Replace worst ranked part of population with offspring

CHAPTER THREE: RESEARCH METHODOLOGY

3.0 Introduction

This chapter presents the research methodology used in this project. The objectives of this project were to compare the lengths of epitopes of sample pathogens with vaccines to those without, suggest most appropriate length(s) to use a polyepitope construct for malaria, HIV-1 and TB. There has been exponential growth of biological data as noted in literature review. Some of this data is stored in private database and part of it is stored in publicly accessible databases such IEDB and HIV molecular immunological database.

There was need to create a local database to store part of this data for better analysis of such data depending on a researcher's interest. In this project a local database was created to store data downloaded from IEDB and HIV molecular immunological database.

Epitopes data was collected in December 2011 from two sources. The main source of data was the immune epitope database (IEDB) http://iedb.org. The epitopes data were collected from this source excluding the HIV epitopes. The HIV epitopes data were collected from the HIV molecular immunological database. The B –cell epitopes were filtered by species criteria (human only) and the analyzed.

The search criteria used for epitopes includes specifying B –cell or T – cell epitopes, the host organism and the source organism. If the host name or source organism was not on the list a search was done using the pathogen name from the NCBI database which is linked to IEDB to facilitate the search.

Pathogens were divided into two clusters – those pathogens that have vaccines and pathogens that do not have vaccines. Seven out of fourteen pathogens with vaccines were selected from NIAID A-C list. For each pathogen selected, B- cell epitopes and T- Cell epitopes were downloaded from IEDB and the HIV molecular immunological database. The epitopes were downloaded in CSV format and saved in Microsoft Excel for B–cell and T-cell. For each pathogen B –Cell and T- cell epitopes were saves in a separate workbook.

The epitopes were analysed using =LEN( ) and=countif(m2: mn, p) functions. The first function was used to calculate the epitope length of each epitope downloaded. The second

function was used to calculate the frequencies of epitopes whose absolute range was $m2 to $mn and lengths p1, p2 …px.

The database was populated with epitope data using PhpMyAdmin graphical user interface application. The data which had been stored in csv format in Microsoft excel workbooks and uploaded using batch processing.

3.1 Prototype design

3.1.1 ERD Model
The epitope database stores epitope data and generate required reports the following entities were identified: Organism, Epitope, Protein, Super type, Allele and polytepitope

The entity relationship diagram for the database is shown below

Figure 5: ERD
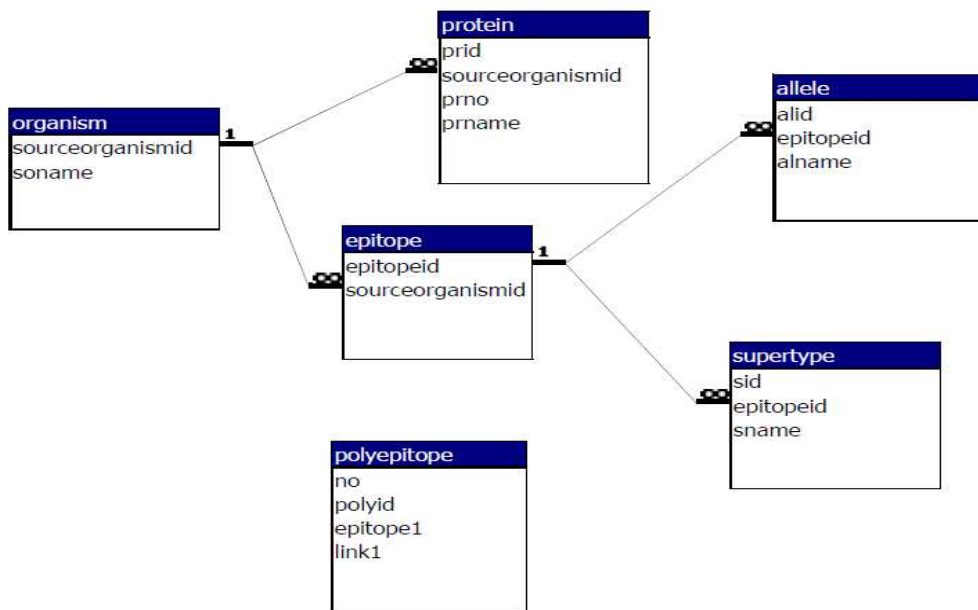
3.1.2 Relational model
The ERD created was converted to a relational model shown below

Organism (sourceornanismid, soname)

Primary Key   sourceornanismid

Foreign Key   -

Protein (prid, sourceorganismid,prno,prname)

Primary Key   prid

Foreign Key   sourceorganismid

Epitope (epitopeid, sourceorganismid,objectdescription,
linearsequence,sma,smname,type)

Primary Key   epitopeid

Foreign Key   sourceorganismid

Supertype (sid, epitopeid,sname)

Primary Key   sid

Foreign Key   epitopeid

Allele   (alid, epitopeid,alname,mhcclass)

Primary Key   alid

Foreign Key   epitopeid

Polyepitope (no,polyid, epitope1,link1, epitope2,link2,
epitope3, ,rank, run date)

Primary Key   no

Foreign Key epitopeid

3.1.3 Polyepitope construct definition

In this project the poly epitope is made up of three epitopes and two link sequences between the first and the second epitope and between the second and the third epitope. Each epitope is made up nine (9) amino acid sequences.

The link amino acid used is **ADL.** In a fully implemented system a variety of linkers of varying lengths would have to be selected. This linker is used in this because it promotes cleavage in the flanking (link) regions. The polyepitope of this length can cover 83% to 88% of the human population if the epitopes from super types **A2, A3 and B7** are correctly processed and presented to MHC class molecules.

3.1.4 Goals

The main goal of designing a polyepitope construct is to ensure all epitopes contained in the polyepitope are correctly processed and presented to the immune system cells. There are two pathways to processing and presentation of antigens to T- lymphocytes as stated in chapter one. First, the MHC class I pathway which is used to present endogenous antigens to CD8+ T-cell. In order to be presented a peptide must be generated by proteasome, the proteasome cleaves (degrades) the peptide into small peptides. The peptides presented, must bind to TAP in order to be translocated to Endoplasmic Reticulum(ER) and be transported to the cell surface. The most selective step is binding to MHC class I molecules. In order for a peptide to be immunogenic it must go through proteasome cleavage, bind to TAP and bind to MHC class I molecules more efficiently than other peptides. Second, the presentation of MHC class II molecules follows a different path, after synthesis and translocation and processing peptides are transported to the cell surface. The specificity of given molecules can be predicted from the amino acid protein sequences and this can be used to select epitopes to include in a vaccine and help to understand the immune system

3.2 Prototype implementation
The steps below describe the implementation of the prototype.

3.2.1 Database implementation

The database was implemented using MYSQL database, innodb engine was used. The innodb engine supports usage foreign keys and transactions. MYSQL database was select because it is fast and it is an open source software and therefore accessible. The support of foreign keys ensures data integrity. MYSQL database was also selected because it works well with Apache web server and compatible with PHP scripting language

The following SQL script was used to create the database.

Figure 6: Epitopedb.sql code

```
/* EPITOPEDB VERSION 1 created in December 2011 */
/* EPITOPEDB VERSION 1 */
Create database if not exists `EPITOPEDB`;
USE `EPITOPEDB`;
DROP TABLE IF EXISTS `organism`;
CREATE TABLE `organism` (
 `sourceorganismid` int(10) NOT NULL,
```

```sql
  `soname` varchar(30) default NULL,
  PRIMARY KEY (`sourceorganismid`)
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
DROP TABLE IF EXISTS `species`;
CREATE TABLE `species` (
 `spid` int(5) NOT NULL ,
 `sourceorganismid` int(10) NOT NULL,
 `spname` varchar (40) NULL,
 PRIMARY KEY (`spid`),
 KEY `sourceorganismid` (`sourceorganismid`),
 FOREIGN KEY (`sourceorganismid`) REFERENCES `organism` (`sourceorganismid`)
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
DROP TABLE IF EXISTS `protein`;
CREATE TABLE `protein` (
 `prid` int(2) NOT NULL,
 `spid` int(5) NOT NULL ,
 `prno` int(10) NOT NULL,
 `prname` varchar(20) NOT NULL,
 PRIMARY KEY  (`prid`),
 KEY `spid` (`spid`),
 FOREIGN KEY (`spid`) REFERENCES `species` (`spid`)
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
DROP TABLE IF EXISTS `epitope`;
CREATE TABLE `epitope` (
 `epitopeid` int(7) NOT NULL ,
 `sourceorganismid` int(10) NOT NULL,
 `objectdescription` varchar (20) NULL,
 `linearsequence` varchar (20) NULL,
 `sma` varchar (20) NULL,
 `smanme` varchar (50) NULL,
 `type` enum('T','B') NOT NULL,
 PRIMARY KEY (`epitopeid`),
 KEY `sourceorganismid` (`sourceorganismid`),
 FOREIGN KEY (`sourceorganismid`) REFERENCES `organism` (`sourceorganismid`)
```

) ENGINE=InnoDB DEFAULT CHARSET=latin1;

DROP TABLE IF EXISTS `supertype`;

CREATE TABLE `supertype` (

  `sid` int(3) NOT NULL,

  `epitopeid` int(7) NOT NULL ,

  `sname` varchar(5) NOT NULL ,

  PRIMARY KEY (`sid`),

  INDEX (epitopeid),

  FOREIGN KEY (epitopeid) REFERENCES epitope (epitopeid)

) ENGINE=InnoDB DEFAULT CHARSET=latin1;

DROP TABLE IF EXISTS `allele`;

CREATE TABLE `allele` (

  `alid` int(3) NOT NULL,

  `epitopeid` int(7) NOT NULL ,

  `alname` varchar(7) NOT NULL ,

  `mhcclass` varchar(2) NOT NULL ,

  PRIMARY KEY (`alid`),

  INDEX (epitopeid),

  FOREIGN KEY (epitopeid) REFERENCES epitope (epitopeid)

) ENGINE=InnoDB DEFAULT CHARSET=latin1;

DROP TABLE IF EXISTS `polytope` ;

CREATE TABLE `polytope` (

`epid` int(5) NOT NULL ,

`A2` varchar(20) NULL ,

`A3` varchar(20) NULL ,

`B7` varchar(20) NULL ,

`O` varchar(20) NULL ,

PRIMARY KEY (`epid`)

) ENGINE = InnoDB DEFAULT CHARSET = latin1;

DROP TABLE IF EXISTS `polytope2`;

CREATE TABLE `polytope2` (

  `epitopeid` int(7) NOT NULL ,

  `epitope` varchar(20) NOT NULL ,

  `organism` varchar(50) NOT NULL ,

```
  `supertype` char(2) NOT NULL ,

  `counter1` int(4) NOT NULL ,

  `counter2` int(4) NOT NULL ,

  `counter3` int(4) NOT NULL ,

   PRIMARY KEY (`epitopeid`),

  FOREIGN KEY (epitopeid) REFERENCES epitope (epitopeid)

) ENGINE=InnoDB DEFAULT CHARSET=latin1;

DROP TABLE IF EXISTS `polytope3`;

CREATE TABLE `polytope3` (

  `epitopeid` int(7) NOT NULL ,

  `epitope` varchar(20) NOT NULL ,

  `organism` varchar(50) NOT NULL ,

  `supertype` char(2) NOT NULL ,

  `counter` int(4) NOT NULL ,

 PRIMARY KEY (`epitopeid`),

    FOREIGN KEY (epitopeid) REFERENCES epitope (epitopeid)

) ENGINE=InnoDB DEFAULT CHARSET=latin1;

DROP TABLE IF EXISTS `polytope`;

CREATE TABLE `polytope` (

  `pid` int(7) NOT NULL ,

  `organism` varchar(50) NOT NULL ,

  `polytope` char(33) NOT NULL ,

   PRIMARY KEY (`pid`)

) ENGINE=InnoDB DEFAULT CHARSET=latin1;
```

3.2.2 Populating the database

The database was populated with epitope data using PhpMyAdmin graphical user interface application. The data which had been stored in csv format in Microsoft excel workbooks and uploaded using batch processing. The column headers in the comma separated value (csv) Excel files should be the same columns in the respective database table and same order maintained.

3.2.3 Epitope data integrity

The data downloaded from the sources stated earlier is of high integrity. The IEDB data is only uploaded to the database after it has been curated by highly experienced IEDB curators. Similarly HIV molecular database is maintained by Los Alamos Laboratory.

3.2.4 Acquisition of polyepitope data

In order to design the polyepitope all the required data is stored in an epitope database. This data includes epitope; species super type, allele, species, epitope type, molecule, protein and organism.

3.2.5 Optimized polyepitope formation framework

Figure 7: Optimized polyepitope formation framework
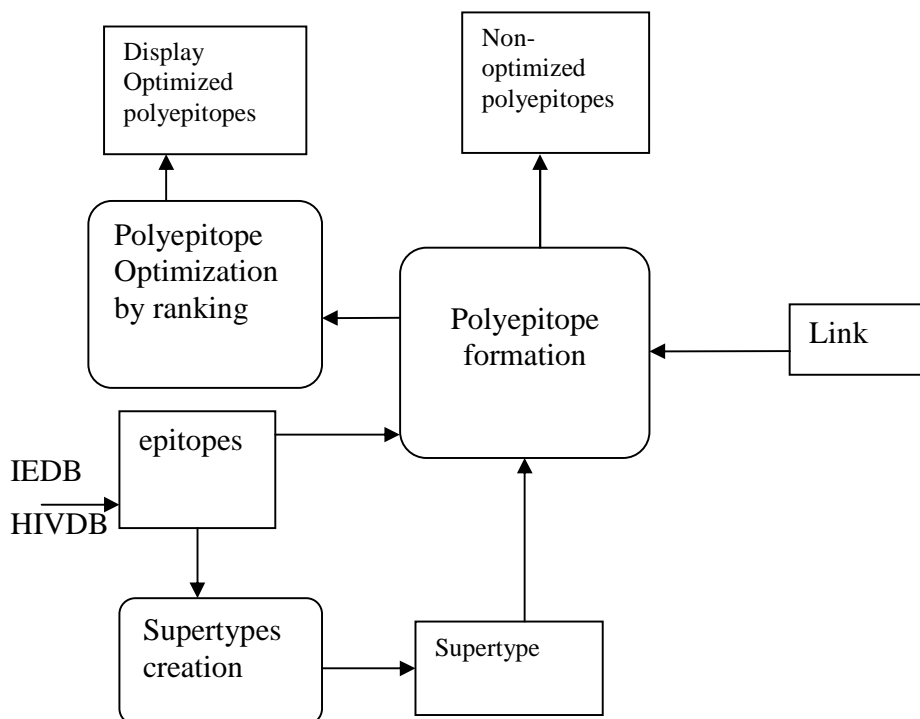
The processes of the Polyepitope optimization prototype are to: select epitopes and organism, specify linker amino acid sequence, and cluster formation (classify epitopes into supertypes A2, A3, B7 and O). The supertype O does not exist but was used in this project to denote supertypes of epitopes which do not fit the three supertypes that were select. Polyepitope formation and calculating the quality of the polyepitope optimization are the main processes.

3.2.6 Three- tier application

The three tier application architecture has been chosen because it would facilitate faster sharing of research data.

The presentation tier runs on a web browser, the HTTP server that has been use is Apache 2.2.11, server side scripting language used is PHP and the data tier runs on MYSQL 5.1.36.

Figure 8: Tiers

Presentation tier

Web browser

HTTP | HTML

Application tier

Apache
PHP

SQL | SQL

MYSQL
5.1.36

Data tier
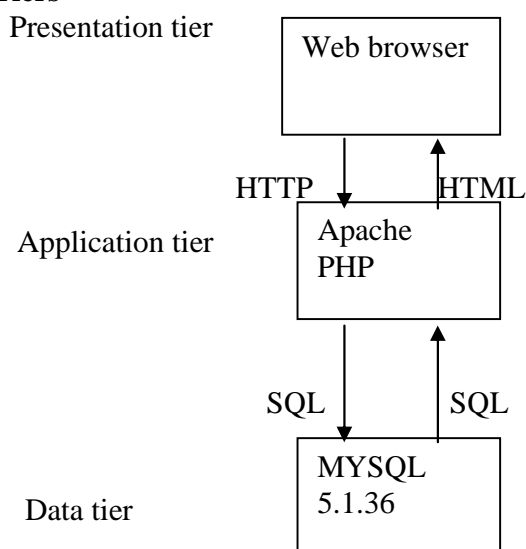
The prototype was implemented using MYSQL, PHP and Apache because they are open source software and available for download. Wampserver 2.0i, which contains the software, was downloaded installed on and HP Compaq C700 with has 1GB RAM, HDD 120 GB, 1.46, 1.47 Duo Core processors Notebook which runs on windows Vista.

3.2.7 Polyepitope optimization

Polyepitopes optimization process involves predicting how efficiently a peptide goes through the three steps described: proteasome cleavage (P), TAP transport (T) and MHC binding (M). In this research project epitopes from supertypes A2, A3 and B7 were used to construct the polyepitope. The aim was to reduce internal cleavage sites within each epitope to zero. Cleavage should occur at C- termini between P1 and P1' positions. The right end of a peptide is the C- terminal and the lift end is the N terminal. The new epitopes processed and presented in fusing regions spanning over epitopes should tend to zero. The length of the linker and also the number of poor C-terminal sites also determine the quality of the polyepitope. The predicted score should be less than 0.9 as noted by Lund et al (2005). Minimal variables were used in this prototype, in a fully implemented system all the predictive variables would have to be included.

Figure 9: Polyepitope construct
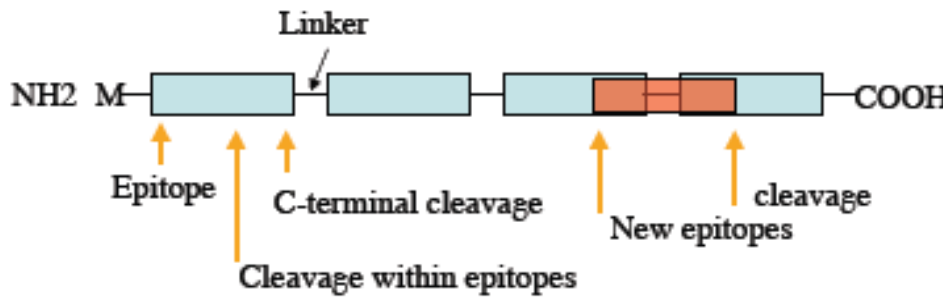
Source: www

The quality (Q) of the polyepitope = Proteasome cleavage score (P) + TAP score (T) +MHC score (M)

In this research project the MHC scores (M) was held constant, each polyepitope consists of epitopes from supertypes A2, A3 and B7.

The local epitope database contains epitopes of varying lengths and only 9mer epitopes were selected and selected epitopes were pathogen specific. Malaria (*plasmodium falciparum*), HIV and Mycobacterium Tuberculosis epitopes were considered. The following example illustrates the polyepitope quality calculation.

EXAMPLE

Given the polyepitope below where the array indexes are 0 to 32, where N1 is the element at index 0, N2 element at index 1, N3 element at index 2 and C element at index 32, also A2C9 is element at index 8, A3C9 is element at index 20 (R) and B7C9 is element at index 32 the sequences coloured blue are the linkers.

Polyepitope: DVKDTKEALADLEELRQHLLRADLYPGIKVKQL

TAP SCORE (T)= N1+ N2+ N3+C where N1 is D, N2 is V, N3 is K and C is L(the element at index 32 of the array., substitute with values from peters et al(2003) table 1

Therefore T= 1.37+(-0.50)+0.09+(-0.94)

       T= 0.12

PC SCORE (P) = A2C9+ A3C9+ B7C9 where A2C9 is element at index 8 i.e. (L), A3C9 is element at index 20 i.e. (R) and B7C9 is element at index 32 i.e. (L) of the array, substitute values with Toes et al table (use P1 column values and divide each by 1000)

     P=1.185+0.275+1.185

    P=2.645

Therefore the quality (Q) of the polyepitope above is

Q=T+P

Q=0.12+2.645

Q=2.765

3.5.4 Supertypes formation

The rules for the formation of the supertypes are derived from **Figures 8** and **9 below.**

In a 9-mer epitopes the anchor residues (amino acids) for the supertypes are as follows: if epitopes are to belong to supertype **A2** the amino acid on the **9th** position must be hydrophobic which means they must contain either of these amino acids **ACFILMPVW**. If epitopes are to belong to supertype **A3** the amino acid on the **9th** position must be basic which means they must contain either of these amino acids HKR and if epitopes are to belong to supertype **B7** the amino acid on the **2nd** position must be Proline **(P).** The supertypes are to generate rules for supertype clusters which determine MHC binding.

Figure 10: HLA- A Tree
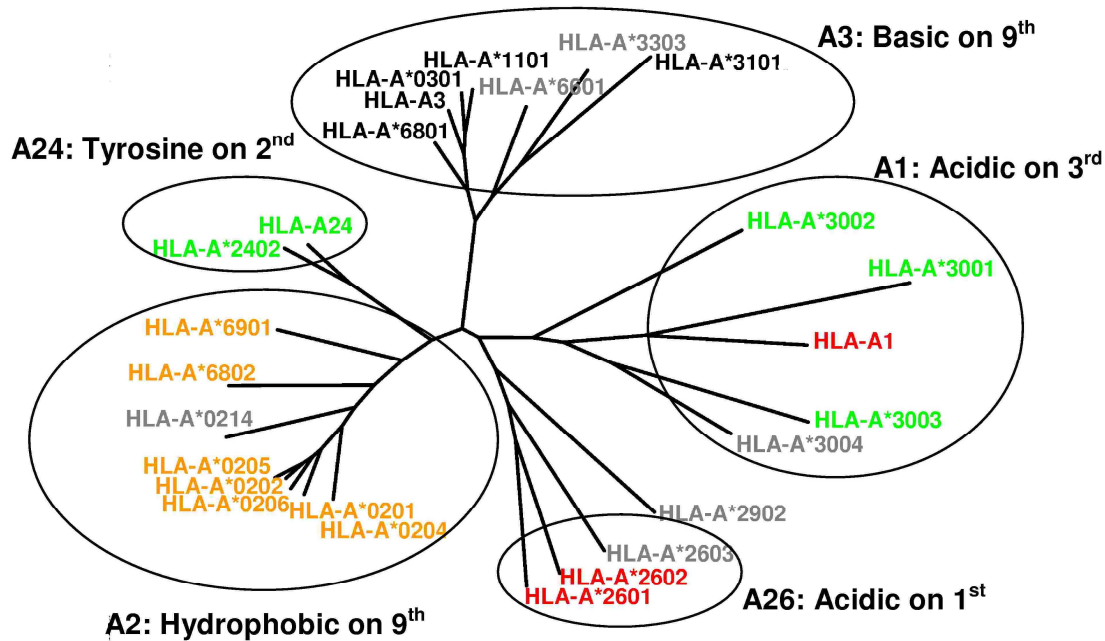
Source: www.cbs.dtu.dk/researchgroups/immunology/supertypes.php

Figure 11: HlA-B Tree
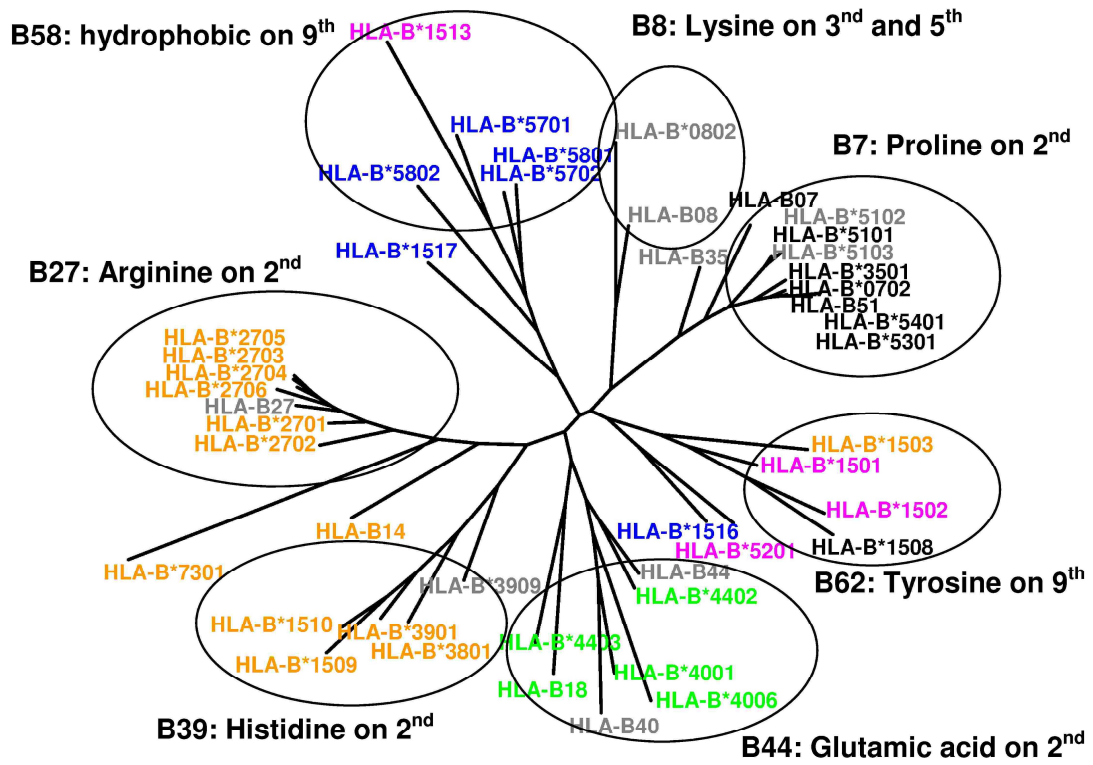
Source: www.cbs.dtu.dk/researchgroups/immunology/supertypes.php

The following flow chart illustrates the formation of supertypes from selected 9mer epitopes. The supertypes were stored in the supertypes table.

Figure 12: Flowchart 1: Classification of Epitopes into supertypes

```
                    ┌──────────────┐
                    │    Start     │
                    └──────┬───────┘
                           │
                           ▼◄──────────────────┐
                    ╱─────────────╲            │
                   ╱     Get        ╲          │
                  ╱    epitope        ╲        │
                 ╱    from DB          ╲       │
                 ╲────────┬────────────╱       │
                          │                    │
                          ▼                    │
                 ┌──────────────────┐          │
                 │ Classify epitope │          │
                 │ into supertypes  │          │
                 │ A2, A3, B7 and O │          │
                 └────────┬─────────┘          │
                          │                    │
                          ▼                    │
                         ╱╲                    │
                        ╱  ╲                   │
                       ╱Last╲──────────────────┘
                       ╲record╱
                        ╲  ? ╱
                         ╲╱
                          │
                          ▼
                 ╱──────────────╲
                ╱    Print        ╲
               ╱  supertypes A2,   ╲
              ╱    A3, B7, O        ╲
              ╲────────┬────────────╱
                       │
                       ▼
                ┌──────────────┐
                │    Stop      │
                └──────────────┘
```
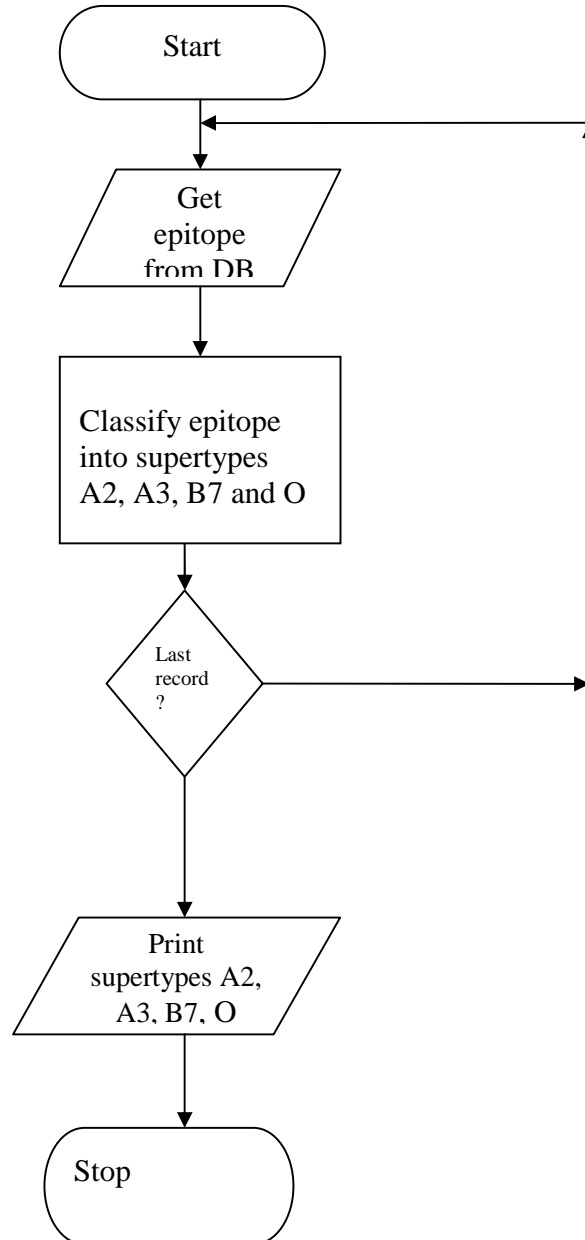
The epitopes are classified into supertypes A2, A3, and B7 otherwise it's classified into supertype O which does not exist. Supertype O is used only in this research project prototype to indicate epitopes which do not fall into A2, A3 and B7 categories.

3.2.8 Polyepitopes formation

Cross tabulation query was use to transform the supertyeps into A2, A3 and B7 columns. The cross tabulation query is show below. (See polyepitope2.php script for further clarification)

SELECT SID, EPITOPEID, ORGANISM,

MAX(CASE WHEN CLUSTER = '1' THEN SUPERTYPE END) AS A2,

MAX(CASE WHEN CLUSTER = '1' THEN EPITOPE END) AS EPITOPE,

MAX(CASE WHEN CLUSTER = '2' THEN SUPERTYPE END) AS A3,

MAX(CASE WHEN CLUSTER = '2' THEN EPITOPE END) AS EPITOPE,

MAX(CASE WHEN CLUSTER = '3' THEN SUPERTYPE END) AS B7,

MAX(CASE WHEN CLUSTER = '3' THEN EPITOPE END) AS EPITOPE

FROM SUPERTYPE GROUP BY epitopeid;

The supertypes were by the script (polyepitope2.php) on an html page which was exported to Excel by right clicking on the page and selecting export to excel. The Excel function = concatenate (A2, link, A3, link, B7) was used to create the polyepitope. The polyepitopes were the stored in the polyepitopes table.

3.2.9 Proteasome cleavage prediction

Given an N amino acid (AA) long peptide and with positions within the peptide P_1-P_n and flanking positions FN, FC where N is the N- terminal and C is the C- terminal, a peptide is produced if positions FN- P_1 and FN- P_n are cleaved and none of internal positions are cleaved. Cleavage occurs between P_1 and P_1' according to Berger and Schechter (1970). Residues on the left of flanking region are called $P_1, P_2, P_{3...}$ while the right side are called P_1' , P_2' , $P_3'_{...}$. The P_1 position is the most important position determining cleavage as noted by Altuvia and Margalit (2000) although the flanking region may be important Mo et al(2000). In this project, each polyepitope is 33 mer long and consists of 9mer A2 epitope+3mer link (ADL), 9mer A3 epitope+ 3 mer link (ADL) and B7 epitope.

Proteasome cleavage score (P) =A2C9+ A3C9+ B7C9. The C- termini scores of each epitope is calculated using Toes et al (2001) table 1 (see the table in the Appendix B).

3.2.10 TAP prediction

The TAP prediction score (T) of amino acid sequence (AA) N1, N2, N3....C is given by T= $MAX_1N_1+MAX_2N_2+MAX_3N_3+MAX_9C$ where N1, N2, N3 are N termini positions and C is the C- terminal of the polyepitope and MAX_p is the matrix score. The TAP scores where based on the Peters et al (2003) consensus scoring matrix which gives high accuracy TAP scores (see the table in the Appendix B). The above scoring matrix was derived on data for 9mer peptides and is

also used to predict TAP score for peptides longer than 9mers because binding of peptides to TAP is determined mainly C-terminal and the three N-terminal residues N1,N2,N3. In position positions N1 and N2 basic residues are favoured i.e. HKR and N3 can either be W or Y. P at position 2 is the strongest destabilizing residue found, nearly abolishing binding as indicated by Uebel et al (1997). This implies B7 epitope should not the leftmost epitope in a polyepitope.

3.2.11 MHC binding prediction

There are a number of methods that have been used to predict MHC binding according to Schirle et al (2001). Majority of peptides binding to HLA complex are 8mer to 10mer peptides. For 9mers, peptides positions 2 and 9 are very important for binding to most class I HLAs theses positions are referred to as anchor positions according to Rammansee et al (1999). In this research project MHC predictions are based on the anchor positions for the supertyes A2, A3 and B7 as shown in figures 8 and 9. All the polyepitopes contain A2, A3 and B7 epitopes therefore MHC variable was held constant.

3.3 Prototype testing

The optimized polyepitopes generated by the prototype were tested on NetChop 3.1 server. The results of random polyepitopes and optimized polyepitopes were compared and conclusions drawn.

The web server has been tested and the following epitopes report was generated using the SQL script. **SELECT e.epitopeid, o.soname, e.linearsequence FROM epitope e, organism o WHERE e.sourceorganismid=o. sourceorganismid limit 10;** the following report was generated by the prototype.

Figure 13: Testing report

http://localhost/epitopedb/viewepitope.php

Figure 14: Polyepitope Optimization Page

CHAPTER FOUR: RESULTS AND DISCUSSION

4.0 Introduction

This chapter presents the results and discusses the results according to researcher's specific objectives. The objectives of the project were to compare epitope lengths for pathogens with vaccines and pathogens without vaccines, suggest most appropriate length(s) of epitopes for use in a polyepitopes for malaria, HIV-I and T.B, to develop an epitopes database, to optimize polyepitope design and to develop a web-based application prototype.

4.1 Results

Epitopes data was collected in December 2011 from two sources. The main source of data was the immune epitope database (IEDB) http://iedb.org. The epitopes data were collected from this source excluding the HIV epitopes. The HIV epitopes data were collected from the HIV molecular immunological database. The B –cell epitopes were filtered by species criteria (human only) and the analyzed.

The search criteria used for epitopes includes specifying B –cell or T – cell epitopes, the host organism and the source organism. If the host name or source organism was not on the list a search was done using the pathogen name from the NCBI database which is linked to IEDB to facilitate the search.

Pathogens were divided into two clusters – those pathogens that have vaccines and pathogens that do not have vaccines. Seven out of fourteen pathogens with vaccines were selected from NIAID A-C list. For each pathogen selected, B- cell epitopes and T- Cell epitopes were downloaded from iedb and HIV websites. The epitopes were downloaded in CSV format and saved in Microsoft Excel for B –cell and T-cell. For each pathogen B –Cell and T- cell epitopes were saves in a separate workbook.

The epitopes were analyzed using $=LEN ( )$ and $=COUNTIF (\$m\$2: \$m\$n, p)$ Excel functions. The first function was used to calculate the epitope length of each epitope downloaded. The second function was used to calculate the frequencies of epitopes whose absolute range was $\$m\$2: \$m\$n$ and lengths $p1, p2 \ldots px$.

4.2 Epitopes data for organisms with vaccines

The number of epitopes for organisms with vaccines that were downloaded from IEDB and HIV molecular immunological database were two thousand seven hundred and twenty four (2724) as the analysis results show in tables 2 and 3 and figures 15 and 16.

4.2.1 T-cell epitopes data for organisms with vaccines

One thousand eight hundred and seventy four (1874) T-cell epitopes for organisms without vaccines were downloaded and the results are shown in table 2 and figure 15.

Table 2: T- Cell Epitopes data for organisms with vaccines

Microbe	9 mer		15 mer		20 mer		25 mer		C	Total
	F	%	F	%	F	%	F	%		
Hepatitis B	48	18.05	59	22.18	25	9.40	1	0.38	133 50%	266
Vaccinia	128	44.14	27	9.31	2	0.69	00	00	157 54%	290
Clostridium Tetani	29	13.39	5	2.33	132	61.40	00	00	166 77%	215
Hepatitis A	8	47.06	0	0	8	47.06	0	0	16 94%	17
Salmonella Typhi	4	100	0	0	0	0	0	0	4 100%	4
Mycobacterium Tuberculosis	176	18.74	116	12.35	281	29.93	93	9.90	666 71%	939
Measles	20	13.99	65	45.45	31	21.68	00	00	106 74%	143
TOTAL										1874

Key
F – Frequency
% - percentage of epitopes
Total – the sum of epitopes
Mer- the length of epitope
    e.g.
- 5mer epitope has five amino acids,
- 6 mer epitope has six amino acids etc

C- Cumulative total

The number T cell epitopes for *Hepatitis B, Vaccinia, Clostridium Tetani, Hepatitis A, Salmonella Typhi, Mycobacterium Tuberculosis* and *Measles* were one thousand eight hundred and seventy four (1874). Seven out of twenty three pathogens with vaccines were selected from NIAID A-C list; only one subtype of each microbe was selected.

Two hundred and sixty six (266) T-cell *Hepatitis B* epitopes were downloaded. Upon careful scrutiny of the epitopes, the researcher discovered that the epitopes appeared in clusters and the main clusters formed around 9mer, 15mer, 20mer and 25mer.

For *Hepatitis B* T cell epitopes , 59 epitopes whose length was 15 amino acids(15mer) this accounted for 22.18 % , 9mer were 48(18.08%), 20mer were 25(9.40%) and 25mer was 1(0.38%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 133(50%).

The frequencies and percentages of *Vaccinia* epitopes were as follows: 9mer were 128(44.14%) ,15mer 27 (9.31%), 20 mer were 2(0.69%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 157(54%).

The frequencies and percentages of C*lostridium Tetani* epitopes were as follows: 9mer were 29(13.39%) ,15mer 5 (2.33%), 20 mer were132(061.40%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 166(77%).

The frequencies and percentages of *Hepatitis A* epitopes were as follows: 9mer were 8(47.06%) ,15mer 0 (0%), 20 mer were 8(47.06%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 16(94%).

The frequencies and percentages of *Salmonella Typhi* epitopes were as follows: 9mer were 4(100%) ,15mer 0 (0%), 20 mer were 0(0%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 4(100%).

The frequencies and percentages of *Mycobateria Tuberculosis* epitopes were as follows: 9mer were 176(18.74%), 15mer 116 (12.35%), 20 mer were 281(29.93%), 25mer were 93(9.90%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 666(71%) and The frequencies and percentages of measles epitopes were as follows: 9mer were 20(13.99%) ,15mer 65(45.45%), 20 mer were 31(21.68%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 106(74%).

Figure 15: T- Cell Epitopes data for organisms with vaccines
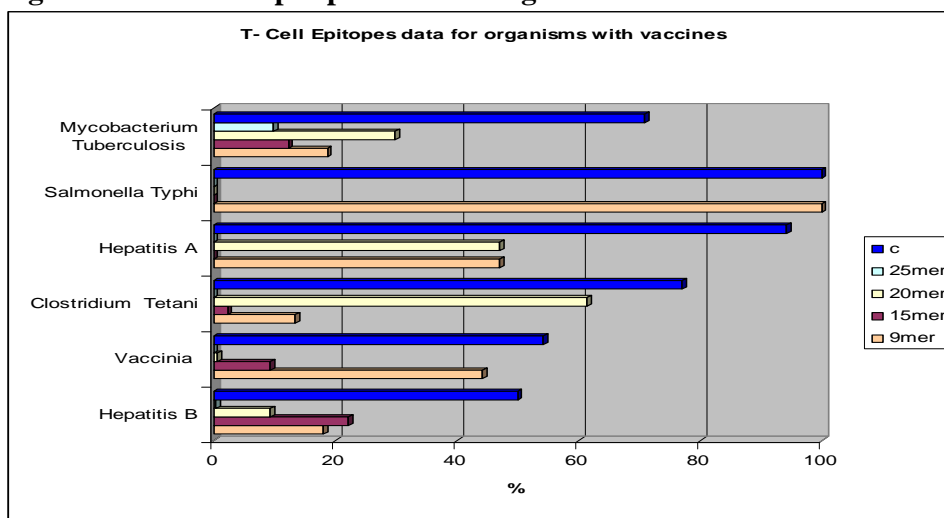
4.2.2 B –cell Epitopes data for organisms with vaccines

Eight hundred and fifty (850) B-cell epitopes for organisms with vaccines were downloaded and the results are shown in table 3 and figure 16.

Table 3: B cell Epitopes data for organisms with vaccines

Microbe	9 mer		15 mer		20 mer		25 mer		23 mer	C	Total
	F	%	F	%	F	%	F	%		44 39%	
Hepatitis B	5	**4.42**	14	**12.39**	9	**7.96**	16	**14.16**			113
Vaccinia	0	0	0	0	0	0	0	0	0	0	0
Clostridium Tetani	0	0	0	0	0	0	0	0	43 91.49%	43 91%	47
Hepatitis A	0	0	3	3.39	77	89.53	0	0	0	80 93%	86
Salmonella Typhi	3	100	0	0	0	0	0	0	0	3 100%	3
Mycobacterium Tuberculosis	2	0.49	229	55.72	59	14.36	7	1.17	13 1.38%	295 71%	441
Measles	0	0	152	95.00	0	0	0	0	0	152 95%	160
TOTAL											850

For *Hepatitis B*, the B cell epitopes , 14 epitopes whose length was 15 amino acids(15mer) this accounted for 12.39 % , 9mer were 5(4.42%), 20mer were 9(7.79%) and 25mer was 16(14.16%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 133(39%).

The frequencies and percentages of *Vaccinia* epitopes were as follows: 9mer were 0(0%) ,15mer 0 (0%), 20 mer were 0(0%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 0(0%).

The frequencies and percentages of *Clostridium Tetani* epitopes were as follows: 9mer were 0(0%) ,15mer 0 (0%), 20 mer were0(0%), 23mer were 34(91.49%) ,25mer were 0(0%) .The cumulative total(C) of 9 mer, 15mer, 20mer, 23mer and 25mer was 43(91%).

The frequencies and percentages of *Hepatitis A* epitopes were as follows: 9mer were 0(0%) ,15mer 3 (3.39%), 20 mer were 77(89.53%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 80(93%).

The frequencies and percentages of *Salmonella Typhi* epitopes were as follows: 9mer were 3(100%) ,15mer 0 (0%), 20 mer were 0(0%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 3(100%).

The frequencies and percentages of *Mycobateria Tuberculosis* epitopes were as follows: 9mer were 2(0.49%) ,15mer 229 (55.72%), 20 mer were 59(14.36%), 25mer were 7(1.17%), 23mer 13(1.38%) . The cumulative total(C) of 9 mer, 15mer, 20mer, 23mer and 25mer was 295(71%) and The frequencies and percentages of *measles* epitopes were as follows: 9mer were 0(0%) ,15mer 152(95.00%), 20 mer were 0(0%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 152(95%).

Figure 16: B cell Epitopes data for organisms with vaccines

B cell Epitopes data for organisms with vaccines
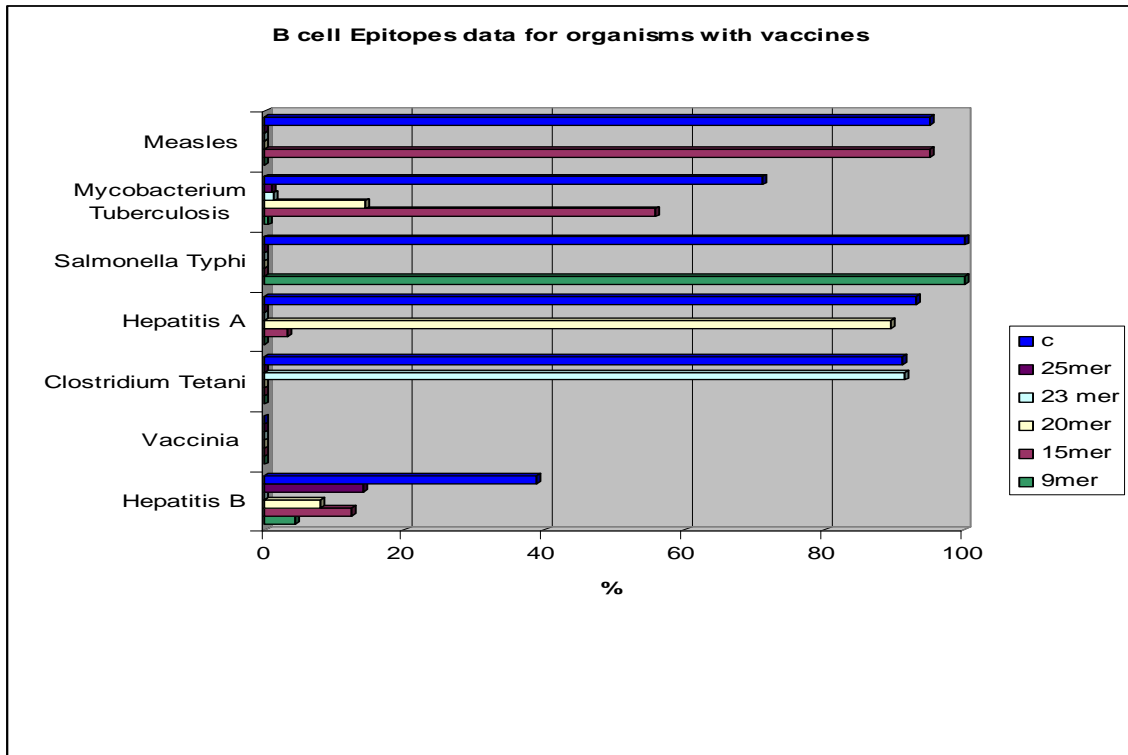
4.3 Epitopes data for organisms without vaccines

The number of epitopes for organisms without vaccines that were downloaded from IEDB and HIV molecular immunological database were six thousand four hundred and seventy one (6471) as the analysis results show in tables 4 and 5 and figures 17 and 18.

4.3.1 T-cell epitopes data for organisms without vaccines

Three thousand six hundred and seventy (3670) T-cell epitopes for organisms without vaccines were downloaded and the results are shown in table 4 and figure 17.

Table 4: T-Cell Epitopes data for organisms without vaccines

Microbe	9 mer		15 mer		20 mer		25 mer		C	Total Epitopes
	F	%	F	%	F	%	F	%		
Plasmodium Falciparum	12	5.02	41	17.15	97	40.59	00	00	150 63%	239
Epistein- barr	159	39.85	71	17.79	56	14.04	6	1.50	295 74%	399
Dengue	56	34.36	58	35.58	10	6.13	0	0	124 76%	163
Hepatitis C	250	17.78	371	26.39	205	15.58	23	1.67	849 60%	1406
HIV-1	661	49.99	00	00	00	00	00	00	661 49.9%	1341
Mycobacterium Bovis	2	1.64	58	47.54	27	22.13	00	00	87 71%	122
								TOTAL		3670

The total number of T cell epitopes downloaded for pathogens without vaccines was three thousand six hundred and seventy (3670).

The parasite that cause malaria –*Plasmodium Falciparum* , T cell epitopes , 9mer epitopes were 12(5.05%), 15mer were 41(17.15%), 20mer were 97(40.59%) and 25mer was 0(0%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 150(63%).

The frequencies and percentages of *Epstein-Barr* epitopes were as follows: 9mer were 159(39.85%), 15mer 71 (17.79%), 20 mer were 56(14.04%), 25mer were 6(1.50%).   The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 295(74%).

The frequencies and percentages of *Dengue* epitopes were as follows: 9mer were 56(34.36%), 15mer 58 (35.58%), 20 mer were10 (6.13%), 25mer were 0(0%).The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 124(76%).

The frequencies and percentages of *Hepatitis C* epitopes were as follows: 9mer were 250(17.78%), 15mer 371 (26.39%), 20 mer were 205(15.58%), 25mer were 23(1.67%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was   16(94%).

The frequencies and percentages of *HIV-1* epitopes were as follows: 9mer were 661(49.99%) ,15mer   0 (0%), 20 mer were 0(0%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 661(49.9%).

The frequencies and percentages of *Mycobacterium Bovis*   epitopes were as follows: 9mer were 2(1.64%), 15mer 58 (47.54%), 20 mer were 27(22.13%), 25mer were 0(0%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was   87(71%).

Figure 17: T cell Epitopes data for organisms with vaccines

T-Cell Epitopes data for organisms without vaccines
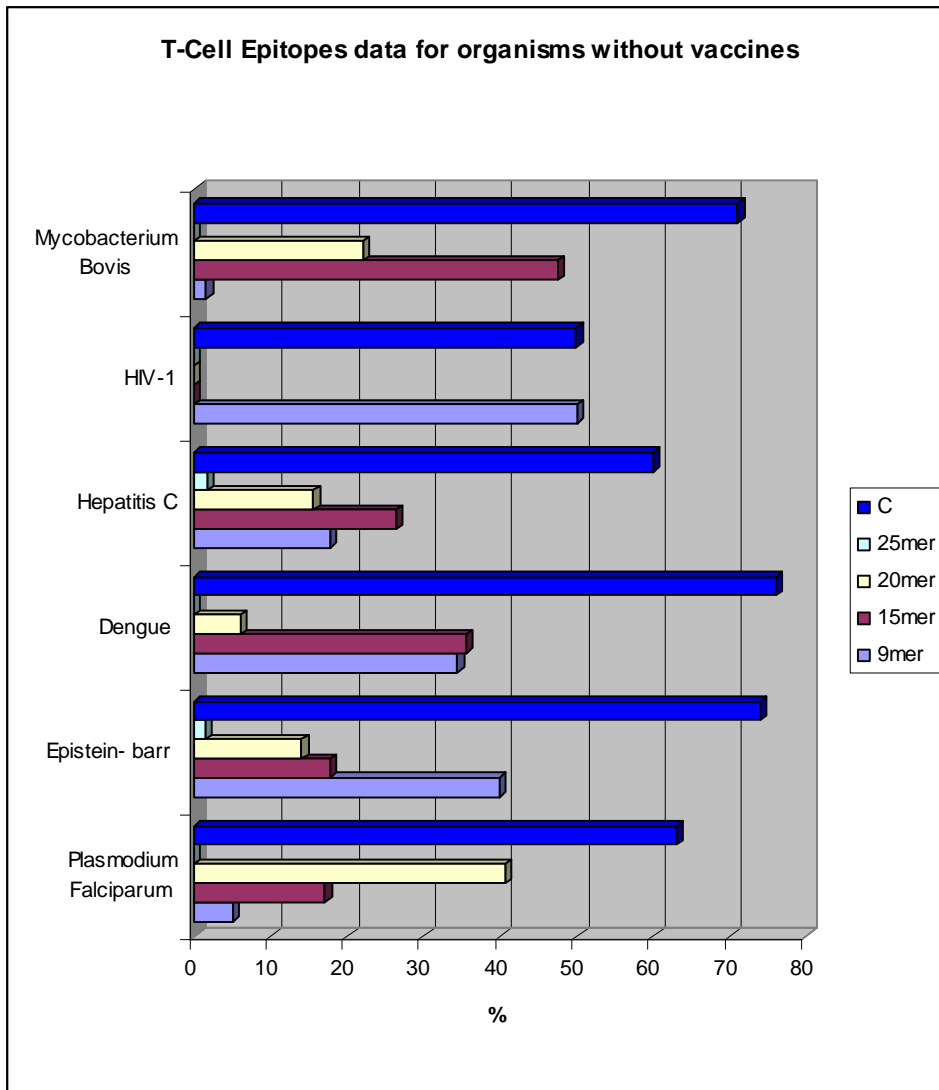
4.3.2 B-cell epitopes data for organisms without vaccines

Two thousand seven hundred and forty one (2741) B-cell epitopes for organisms without vaccines were downloaded and the results are shown in table 5 and figure 18.

Table 5: B cell Epitopes data for organisms without vaccines

Microbe	9 mer		15 mer		20 mer		25 mer		C	Total
	F	%	F	%	F	%	F	%		
Plasmodium Falciparum	20	3.95	105	20.57	24	4.74	24	4.74	173 34%	506
Epistein- barr	3	1.96	16	10.46	41	26.80	0	0	60 39%	153
Dengue	19	6.91	14	5.09	3	1.09	0	0	36 13%	275
Hepatitis C	152	8.94	69	4.06	391	22.99	35	2.06	647 38%	1701
HIV-1	4	5.13	8	10.26	5	6.41	0	0	17 22%	78
Mycobacterium Bovis	0	0	1	3.57	4	14.29	0	0	5 18%	28
								TOTAL		2741

The total number of B cell epitopes downloaded for pathogens without vaccines was two thousand seven hundred and forty one (2741).

The *Plasmodium Falciparum,* 9mer epitopes were 20(3.95%), 15mer were 105(20.57%), 20mer were 24(4.74%) and 25mer was 24(4.74%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 173(34%).

The frequencies and percentages of *Epstein-Barr* epitopes were as follows: 9mer were 3(1.96%), 15mer 16 (10.46%), 20 mer were 41 (26.80%), 25mer were 0(0%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 60(39%).

The frequencies and percentages of *Dengue* epitopes were as follows: 9mer were 19(6.91%), 15mer 14 (5.09%), 20 mer were 3 (1.09%), 25mer were 0(0%).The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 36(13%).

The frequencies and percentages of *Hepatitis C* epitopes were as follows: 9mer were 152(8.94%), 15mer 69 (4.06%), 20 mer were 391(22.99%), 25mer were 35(2.06%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 647(38%).

The frequencies and percentages of *HIV-1* epitopes were as follows: 9mer were 4(5.13%) ,15mer 8 (10.26%), 20 mer were 5(6.41%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 17(22%).

The frequencies and percentages of *Mycobacterium Bovis* epitopes were as follows: 9mer were 0(0%), 15mer 1 (3.57%), 20 mer were 4(14.29%), 25mer were 0(0%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 5(18%).

Figure 18: B cell Epitopes data for organisms without vaccines
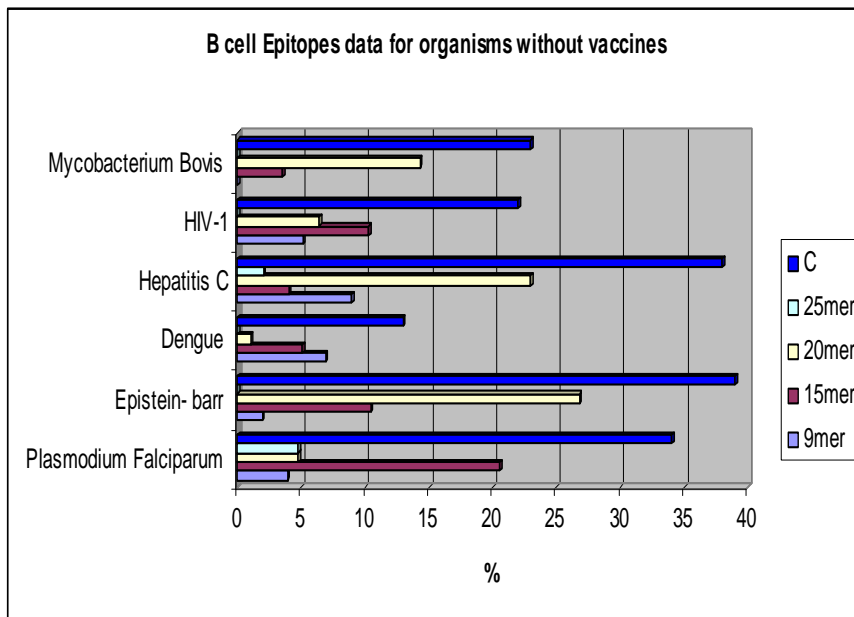
The percentages of epitopes of lengths 9mer, 15mer, 20mer and 25mer are less than 50% of the B cell epitopes in circulation as shown if figure 18. The cumulative totals of the stated clusters for each pathogen is denoted C and is shown on the blue bars.

4.4 Optimized polyepitopes

The optimized polyepitopes that were generated by the prototype were ranked using the total score and the top ten polyepitopes were selected and tested on the NetChop 3.1 server. The NetChop 3.1 makes preateasome predictions based artificial neural networks. The researcher choose this tool to validate polyepitope results generated by the prototype because it available on the World Wide Web and makes reliable predictions, however the tool measures only one of the two variables implemented on the prototype. The results are shown in figures 19 and 20.

For each of the top ten polyepitopes selected the internal cleavage sites were counted. The best polyepitope was assigned the first position, the second, second position and so on as shown above. It was expected that as the quality of the polyepitope increases the number of internal cleavage sites reduce and tend to zero. The best polyepitopes had only three internal cleavage sites. The best polyepitopes should not have any internal cleavage sites (cleavage sites within each epitope). This was consistent with the results generated by NetChop 3.1 server –this is shown by the trend line -linear sites in figure 18.

Figure 19: Optimized polyepitopes
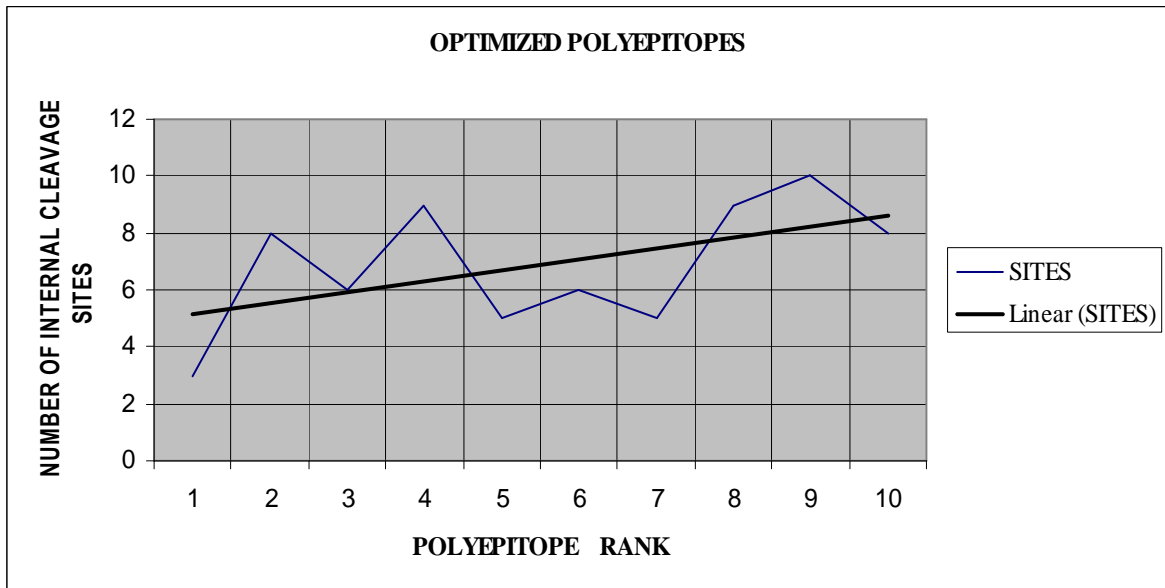
4.5 Random polyepitopes

Randomly generated polyepitopes were also tested on the NetChop 3.1 server and the results are shown in figure 19 shown below.

Figure 20: Random polyepitopes
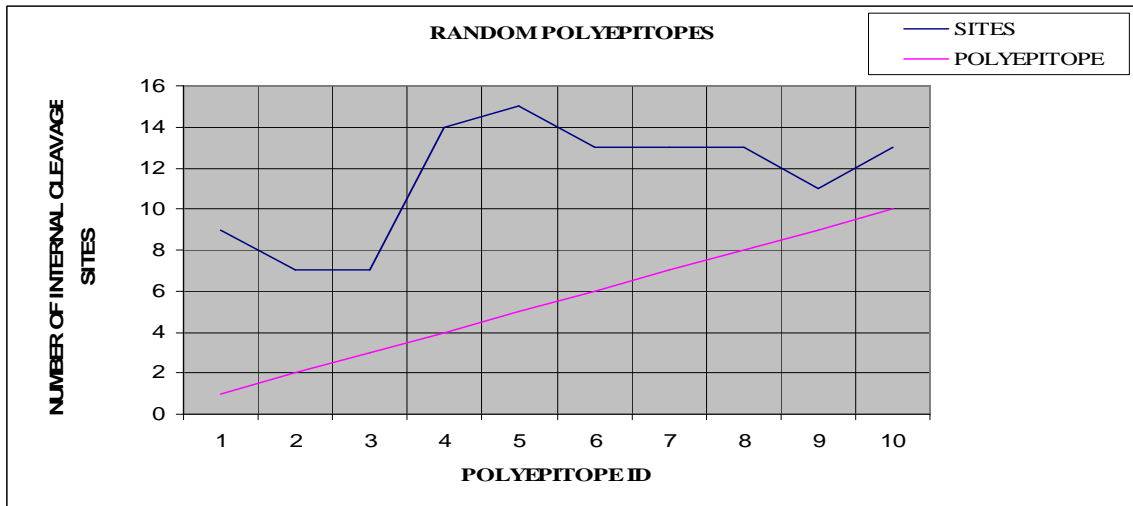
The average number of internal cleavage sites for the top ten optimized polyepitopes is 6.9 and the ten random polyepitopes have 11.5, this is consistent with the expected results- polyepitope optimization should reduce internal cleavage sites to zero.

4.6 Discussion

From the above table 2 and figure 15 it's clear that cumulative total of epitopes in clusters of 9mer, 15mer, 20mer and 25mer constitute 50% or more of the epitopes.
From the above table 3 and figure 16 it's evident that cumulative total of epitopes in clusters of 9mer, 15mer, 20mer and 25mer constitute 50% or more of the epitopes.

From the above table 4 and figure 17 it's clear that cumulative total of epitopes for each microbe, in clusters of 9mer, 15mer, 20mer and 25mer constitute 50% or more of the epitopes.

From the above table 5 and figure 18 it's clear that cumulative total of epitopes for each microbe in clusters of 9mer; 15mer, 20mer and 25mer constitute less than 50% of the epitopes.

There seems to be a correlation of the frequencies of epitopes of the same lengths between the T-cell epitopes and B-cell epitopes for instance if 9mer epitopes of a pathogen are high the corresponding 9mer B cell epitopes tend to be high for the pathogens which have vaccines.

Polypepitope optimization prototype generated polyepitopes with fewer internal cleavage sites compared the randomly generated polyepitopes. The average internal cleavage sites for the optimized polyepitopes was 6.9 and for the random polyepitopes 11.5. This indicates that polyepitope optimization can generates better polyepitope. A fully implemented polyepitopes optimization system would have to screen thousands of polyepitopes ranking each one of them. The screening process would substantially reduce the effort needed to experimentally validate the polyepitopes since only the best polyepitopes would best would be tested in the laboratory and used for vaccine development.
The researcher initially intended to use genetic algorithms but found the initial population generation is limited to epitopes generated from the public databases mutations would lead to non existent epitopes.

CHAPTER FIVE: CONCLUSION AND RECOMMENDATION

5.0 Introduction

This chapter presents the conclusion, recommendations and limitations according to researcher's specific objectives. The objectives of the project were to compare epitope lengths for pathogens with vaccines and pathogens without vaccines, suggest the most appropriate length(s) of epitopes for use in a polyepitopes for malaria, HIV-I and T.B, to develop an epitopes database, to optimize polyepitope vaccine design and to develop a web-based application prototype.

5.1 Conclusion

The epitopes occur in clusters and the main clusters are clusters contain epitopes of the following lengths 9mers, 15mers, 20mers and 25mers. The clusters of 9mer, 15mer, 20mer and 25mer are important in the immune system response generation.

This suggests that the immune system memory remembers pathogenic attacks using the clusters stated above. This implies that polyepitope vaccines constructs design could be made better by taking into account the epitope clusters, however further research is needed to establish finer detail of epitope cluster.

Generally, the epitopes of pathogens with vaccines have more than 50% of the epitopes in circulation being made up of lengths 9mer, 15mer, 20mer and 25mer and also have 50% of the B cell epitopes of the same lengths. The clusters of epitopes of pathogens without vaccines constitute more than 50% percent of B cell epitopes do not feature the main epitope clusters stated above that is 9mers, 15mers, 20mers and 25mers. The epitopes of pathogens without vaccines have more clusters compared to those with vaccines

The immune system response is made better when both T- cell and B- cell epitopes are involved in generating immune response. The absence or representation of 9mers, 15mers, and 20mers and 25mers epitopes in pathogens without vaccines may suggest the unavailability of vaccines for the pathogens.

Polyepitopes optimization algorithms can be used to create better polyepitopes which can be use to make better epitope-based vaccines at a lower cost and within a shorter time.

5.2 Recommendations

Epitopes clusters seem to play a role in the generation of immune response and also in the creation of polyepitope vaccine design, further research in needed on epitope clusters and the immune signal generation and validation of results in wet lab research.

There is to search for 9mer, 15mer 20mer and 25mer epitopes for malaria (*plasmodium falciparum*), HIV-1 and mycobacterium tuberculosis (TB)

The need to carry out collaborative research by biological and computer scientist on the design of polyepitopes vaccine constructs and other types of vaccines

5.3 Limitations

The unavailability of publicly accessible tools integrate proteasome cleavage, TAP binding and MHC binding predictions that can be used to test the quality of polyepitope constructs/vaccines. The of polyepitope construct validation requires wet lab experiments which were not conducted in this research.

Time was also limited and the researcher decided to develop a prototype which though functions lacks some feature-for instance security features and incorporates minimal variables for polyepitope quality computation.

REFERENCES

Altuvia, Y. and Margalit, H., Sequence signals for generation of antigenic peptides by the proteasome: implications for proteasomal cleavage mechanism. *J. Mol. Biol.* 2000. 295: 879–890.

Andre, F. How the research-based industry approaches vaccine development and establishes priorities. *Dev Bio .(Basel),110:25-29* 2002.

Blithe M.J, Doytchnova I.A, and Flower D. R, 2002, Jenpep: a database of quantitative functional peptide data for immunology. *Bioinformatics*, 18:433-439

Devis V.A, 2009 PolyCTLDesigner < http://tepredict.sourceforge.net/PolyCTLDesigner.html>

Encarta (2005) Microsoft Encyclopaedia, Microsoft Corporation. 1993-2004

Feason D.T and Locksley RM. 1996, The instructive role of innate immunity in the acquired immune response. *Science,* 272:50-53

Flower D.R, McSparon H, Blyth M.J, 2003, Computational vaccinology: quantitative approaches *Norvatis found.symp.,* 254: 102-120.

Kesmir, C., Nussbaum, A. K., Schild, H., Detours, V. and Brunak, S., Prediction of proteasome cleavage motifs by neural networks. Protein. Eng. 2002. 15: 287–296.

Korber T.M, Brander C, Haynes B.F, 2001a. *HIV Molecular Immunology 2001.* Theoretical Biology and Bio-physics group, Los Alamos National Laboratory, Los Alamos, NM,

Liliana F., Bjarni H,, Oliver K.,Russell S., Stephen H., and Sorin I. 2003 Epitope Prediction Algorithms for Peptide-based Vaccine Design: Proceedings of the Computational Systems Bioinformatics (CSB'03) computer society

Mette V., Claus L., Kasper L., Søren Buus, Søren Brunak, Ole Lund and Morten N. An integrative approach to CTL epitope prediction: A combined algorithm integrating MHC class I binding,
TAP transport efficiency, and proteasomal cleavage predictions *Eur. J. Immunol*. 2005. 35: 2295–2303

Metropolis N, Rosenbluth A.W, Teller A..H and Teller E. 1953, Equation of state calculation by fast computing machines. *J. chem. Phy.,* 21:1087-1092 NIAID, 2000,*The Jordan report 2000: accelerated development of vaccines* <*http://www.niaid.nih.gov/publications/pdf/jordan.pdf*>

Mo, X. Y., Cascio, P., Lemerise, K., Goldberg, A. L. and Rock, K., Distinct proteolytic processes generate the C and N termini of MHC class I-binding peptides. J. Immunol. 1999. 163: 5851–5859.

Lund O, Morten N, Claus L, Can K and Soren B, 2005, *Immunological Bioinformatics,* MIT press, London.

Peters, B., Bulik, S., Tampe, R., Endert, P. M. V. and Holzhutter, H. G.,Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. J. Immunol. 2003. 171: 1741–1749.

RAC,2002, appendix B,<http://www4d.nih.gov/oba/rac/guidelines_02/APPENDIX_b.htm> Rammensee J. H, Emerich N.P, Bachor A.O and Stevanovic S, 1999, SYFPEITHI:database for MHC ligands and peptide motifs. *Immunogenetics,* 50:213-219.

Richard Feynman ,1985 ,Surely You're Joking Mr. Feynman! Bantam

Suhrbier ,A 2002, 'Polytope vaccines for the codelivery of multiple CD8 T-cell epitopes', *Expert rev. Vaccines,* 1:207-213.

Toes RE, Nussbaum AK, Degermann S, Schirle M,Emmerich NPN, et al. (2001) Discrete

cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. J Exp Med 194: 1-12.

Toussaint NC, Do¨nnes P, Kohlbacher O (2008) A Mathematical Framework for the Selection of an Optimal Set of Peptides for Epitope-Based Vaccines. PLoS Comput Biol 4(12): e1000246. doi:10.1371/journal.pcbi.1000246

Thomas C and Carolyn B, *2005 Database systems: a practical approach to design, implementation and management,* Addison Wesley, England

WHO, 2004a, the world health report, changing history, <http://www.who.int/entity/whr/2004/annex/topic/en/annex_2_en.pdf>

WHO,2004b,the world health report, changing history, <http://www.who.int/entity/whr/2004/en/01-contents_en.pdf>

Whitten JL, Lonnie D, and Kelvin C, 2001, System *analysis and design methods 5th ed,* McGraw Hill , New yolk.
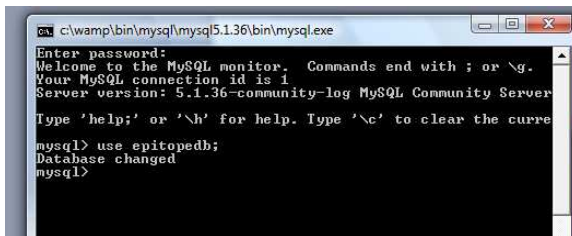
APPENDIX A: USER MANUAL

1. Install the software

    i) Download and install the Wampserver 2.0i. install micomedia dreamweaver 8 on a computer running windows xp or vista operating systems

    ii) Copy the folder named EPITOPEDB and paste in the folder in C:\wamp\ www

    iii) double click the wapserver icon that appears on the desktop

2. Create Mysql Database

    i) Run the epitopdb.sql script on the Mysql placed in epitopeDb folder

3. Start the prototype

    i) Open a web browser e.g. Internet explorer

    ii) Type the url for the index page on the address bar:
       http://localhost/epitopedb/index.php

    iii) Press enter
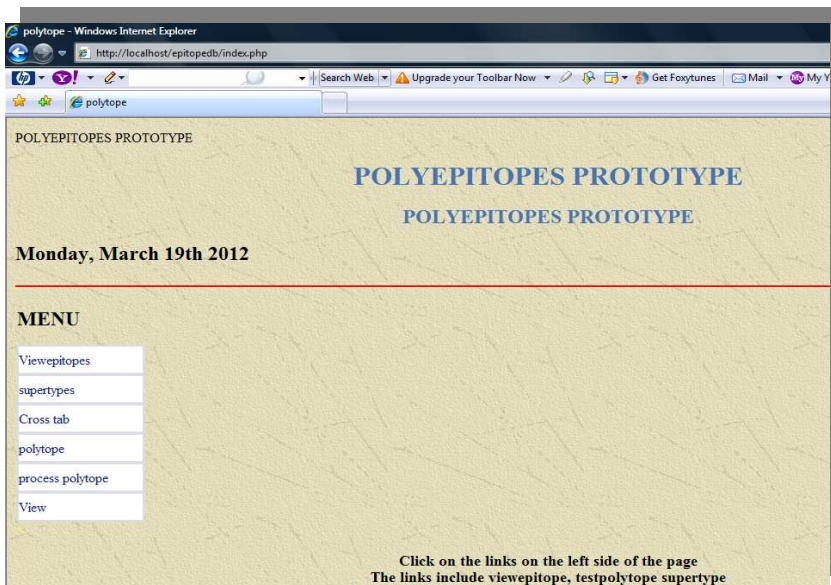
4. Select the require script by clicking on the links on the home page

Click **viewepitope** link produce the following report.

EPITOPES REPORT

EPITOPE DETAILS

Monday, March 19th 2012

CODE	ORGANISM	EPITOPE
1	HIV	MGARASVLSG
2	HIV	ASVLSGGEL
3	HIV	ASILRGGKLDK
4	HIV	SVLSGGQLDR
5	HIV	LSGGELDRWEK
6	HIV	GELDRWEKI

Click **polyepitope** link to produce the following report.

POLYTOPES REPORT

POLYTOPES DETAILS

Monday, March 19th 2012

POLYTOPES RESULTS

NO	ORGANISM	POLYTOPE	TAPSCORE	PCSCORE	TOTAL
235	HIV	GELDRWEKIADLIRLRPGGKKADLRPGGKKRYM	-0.640	0.625	-0.015
236	HIV	GQLDRWEKIADLRLRPGGKKKADLSPRTLNAWV	-0.720	1.040	0.320
237	HIV	KRYMIKHLVADLRLRQGGKKKADLTPQDLNTML	-1.950	1.715	-0.235
238	HIV	KYKLKHIVWADLGGKKKYKLKADLTPQDLNMML	-1.150	1.295	0.145
239	HIV	HYMLKHIVWADLAVLYCVHQRADLTPYDINQML	-0.910	1.490	0.580
240	HIV	HYMLKHLVWADLATLYCVHQRADLTPQDLNQML	-0.910	1.490	0.580
241	HIV	KYMLKHIVWADLATLYCVHQKADLHPVHAGPIA	-0.150	0.935	0.785
242	HIV	KYRLKHLVWADLCVHQRIEIKADLHPVHAGPVA	-0.150	0.935	0.785

To Test dummy polyepitope

Click **Testpolytope** link and enter details and click submit

POLYEPITOPE PROTOTYPE

ENTER POLYEPITOPE

CODE 120293744

ORGANISM NONE

POLYEPITOPE IADLVQNSNPDCKADLRPEPTAPPA

SUBMIT

APPENDIX B: OTHER DIAGRAMS

Flowchart 2: Calculating TAP score (T)

```
                    ╭──────────────╮
                    │    Start     │
                    ╰──────┬───────╯
                           │            ┌─────────────►
                           ▼            │
                  ╱─────────────────╲   │
                 ╱   Polyepitope     ╲  │
                ╱    array,N1,N2,      ╲ │
                ╲    N3,C             ╱  │
                 ╲─────────────────╱    │
                           │            │
                           ▼            │
                ┌──────────────────┐    │
                │   Calculate TAP  │    │
                │ Score(T)= MAX₁N₁+│    │
                │ MAX₂N₂+ MAX₃N₃+  │    │
                │      MAX₉C       │    │
                └─────────┬────────┘    │
                          │             │
                          ▼             │
                       ╱──────╲         │
                      ╱  Last  ╲────────┘
                      ╲ record ╱
                       ╲  ?   ╱
                        ╲────╱
                          │
                          ▼
                 ╱─────────────────╲
                ╱     Print T       ╲
                ╲─────────────────╱
                          │
                          ▼
                    ╭──────────────╮
                    │    Stop      │
                    ╰──────────────╯
```

Calculate TAP Score $(T) = MAX_1N_1 + MAX_2N_2 + MAX_3N_3 + MAX_9C$
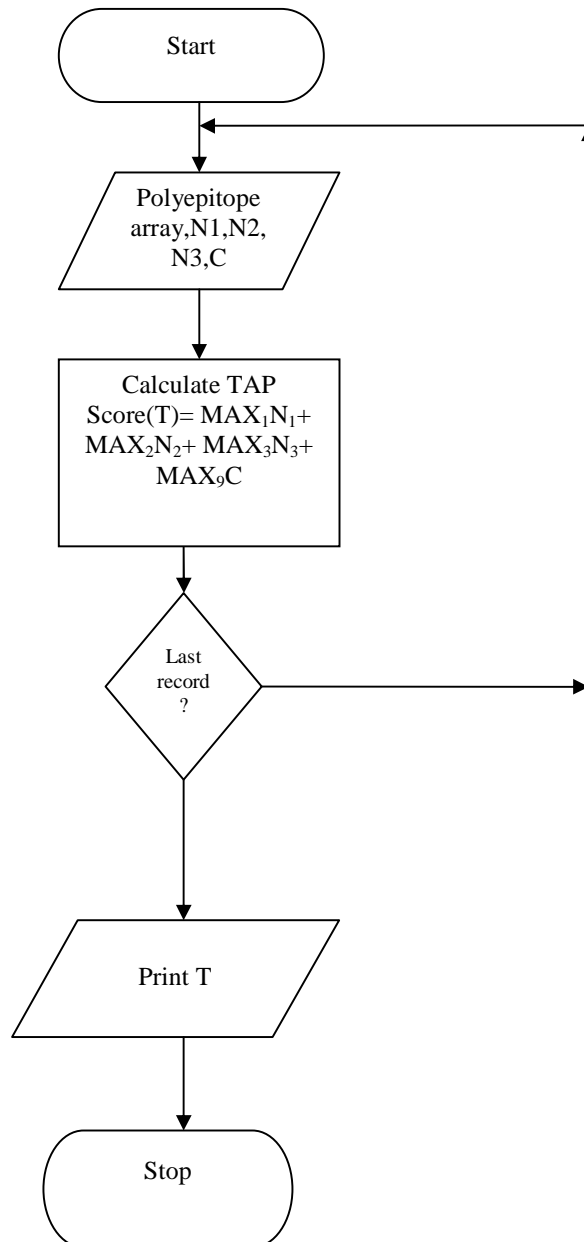
Flowchart 3: Calculating Proteasome cleavage score (P)

```
                    ┌──────────────┐
                    │    Start     │
                    └──────┬───────┘
                           │                          ◄──────────────┐
                           ▼                                         │
                   /─────────────/                                   │
                  /  Polyepitope  /                                  │
                 /  array, A2C9,  /                                  │
                /   A3C9, B7C9,  /                                   │
               /───────────────/                                    │
                           │                                         │
                           ▼                                         │
                ┌────────────────────┐                              │
                │ Calculate Proteasome│                             │
                │   cleavage Score    │                             │
                │ (P)= A2C9+ A3C9+    │                             │
                │        B7C9,        │                             │
                └──────────┬─────────┘                              │
                           │                                         │
                           ▼                                         │
                        ╱─────╲                                      │
                       ╱ Last  ╲                                     │
                      ╱ record  ╲───────────────────────────────────┘
                      ╲    ?    ╱
                       ╲───────╱
                           │
                           ▼
                   /─────────────/
                  /   Print P    /
                 /─────────────/
                           │
                           ▼
                    ┌──────────────┐
                    │     Stop     │
                    └──────────────┘
```

Start

Polyepitope array, $A2C_9$, $A3C_9$, $B7C_9$,

Calculate Proteasome cleavage Score $(P) = A2C_9 + A3C_9 + B7C_9$,

Last record ?

Print P

Stop
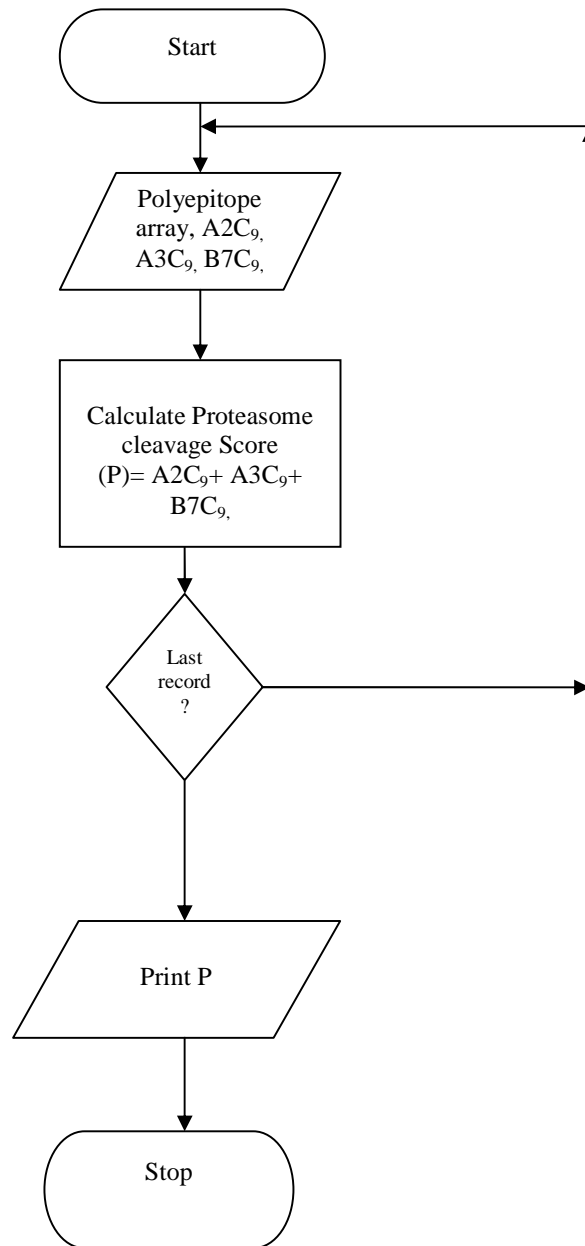
Flowchart 4: Polyepitope optimization

```
                    ┌─────────────────┐
                    │      Start       │
                    └────────┬────────┘
                             │         ┌──────────────────┐
                             ▼         ▲                  │
                    ╱─────────────╲    │                  │
                   ╱ Polyepitope    ╲  │                  │
                  ╱  array from       ╲ │                  │
                 ╱   DB                ╲                   │
                 ╲────────┬───────────╱                   │
                          │                               │
                          ▼                               │
                 ┌────────────────────┐                   │
                 │ Calculate TAP (T)   │                   │
                 │ Calculate proteasome│                   │
                 │ cleavage (P)        │                   │
                 │ Quality of polyepitope                  │
                 │ (Q) =T+P            │                   │
                 └─────────┬──────────┘                   │
                           │                              │
                           ▼                              │
                       ╱───────╲                          │
                      ╱  Last    ╲                         │
                     ╱  record    ╲─────────────────────► │
                     ╲     ?      ╱
                      ╲─────────╱
                          │
                          ▼
                 ╱──────────────────╲
                ╱    Print T, P, Q    ╲
                ╲────────┬───────────╱
                         │
                         ▼
                 ┌─────────────────┐
                 │      Stop        │
                 └─────────────────┘
```
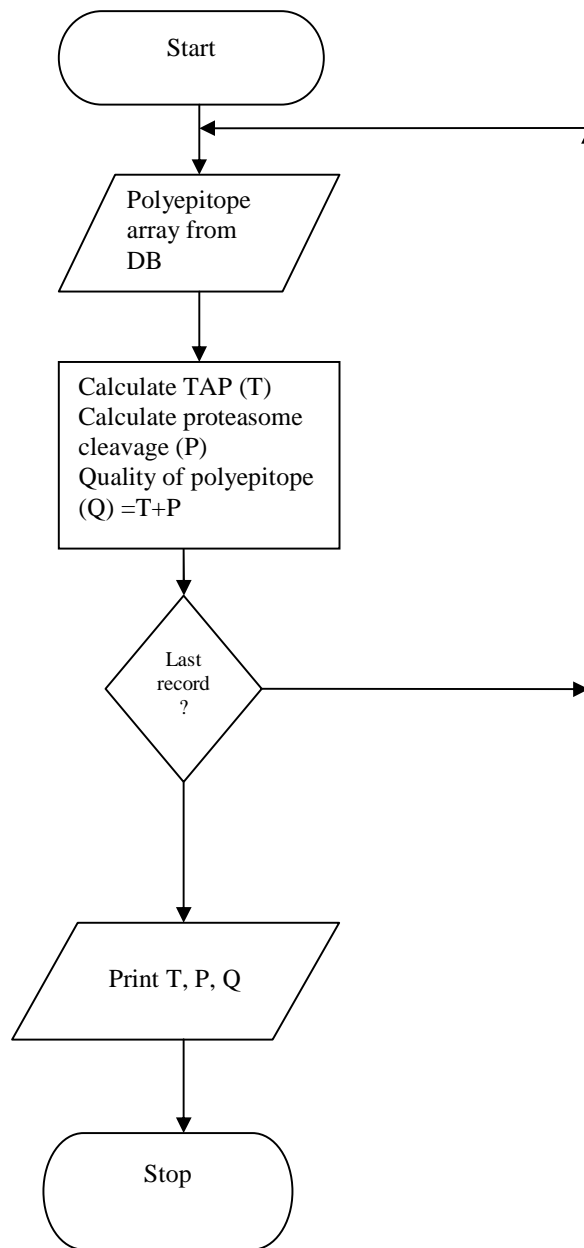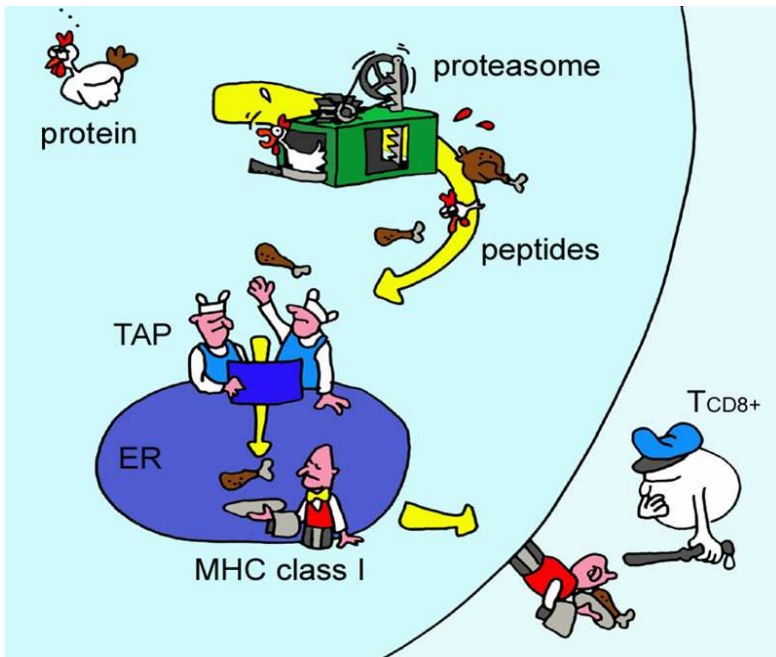
MHC class I pathway

Source: Lund (2005) and www

MHC class II pathway
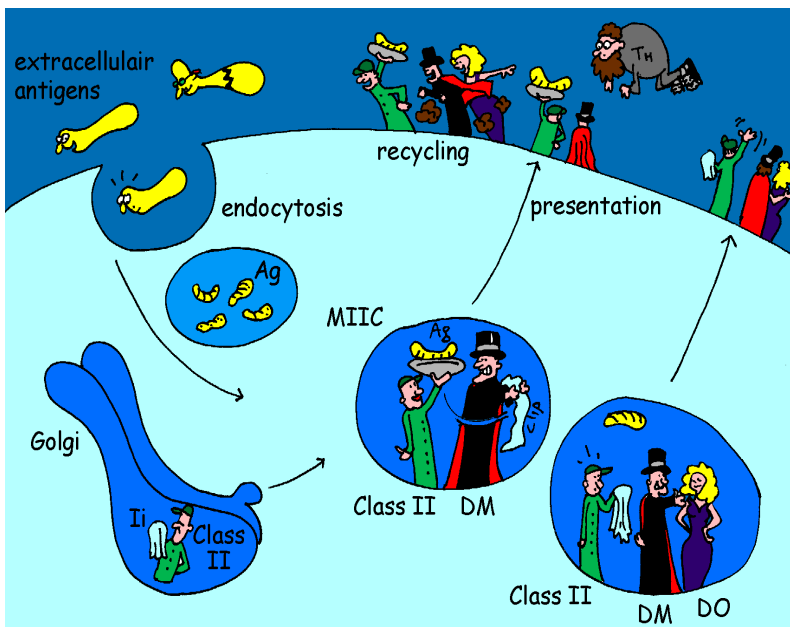
Source: Lund (2005) and www

Peters et al (2003) Consensus Scoring Matrix

Amino Acid	(N1) Position 1	(N2) Position 2	(N3) Position 3	Position 4	Position 5	Position 6	Position 7	Position 8	(C) Position 9	Amino Acid
A	−1.56	−0.25	−0.10	0.24	−0.10	0.17	0.27	0.00	0.55	A
C	0.05	−0.01	−0.02	0.11	0.09	0.05	0.00	−0.13	0.00	C
D	1.37	1.42	1.83	−0.23	0.33	0.32	1.07	0.32	1.83	D
E	1.65	0.02	1.51	0.08	0.54	−0.13	0.64	0.44	1.58	E
F	1.03	−0.45	−1.05	−0.50	−0.26	0.08	−0.50	0.17	−2.52	F
G	0.28	1.14	1.70	0.45	0.66	0.12	1.41	−0.38	1.41	G
H	0.21	0.33	−0.23	−0.21	−0.11	−0.06	−0.19	0.39	0.55	H
I	−0.11	−0.49	−0.62	−0.09	−0.42	−0.75	−0.94	0.45	−0.52	I
K	−1.03	−0.41	0.09	−0.23	−0.08	−0.26	0.44	0.12	−0.45	K
L	−0.50	0.09	−0.11	0.11	−0.34	0.02	−0.73	0.01	−0.94	L
M	−0.38	−0.46	−0.58	−0.35	−0.26	0.30	−0.64	−0.11	−0.29	M
N	−1.43	0.69	1.01	0.38	0.49	−0.27	0.16	0.33	1.33	N
P	1.43	3.00	0.22	−0.04	−0.72	−0.13	−0.84	0.03	−0.09	P
Q	0.47	−0.97	0.39	0.15	0.15	−0.07	0.34	0.26	0.12	Q
R	−1.34	−1.47	−0.42	−0.27	−0.32	−0.75	−0.09	−0.42	−1.47	R
S	−0.56	−0.34	0.11	0.27	0.45	0.31	0.87	−0.51	2.26	S
T	−0.12	−0.04	0.43	0.23	0.43	0.49	0.39	−0.46	0.72	T
V	−0.49	−0.50	−0.71	0.27	0.37	−0.02	−0.29	0.10	−0.30	V
W	0.54	−0.64	−1.65	−0.18	−0.78	0.31	−0.50	−0.63	−0.87	W
Y	0.50	−0.67	−1.80	−0.18	−0.13	0.28	−0.87	0.02	−2.91	Y

Toes et al (2001) table 1
AA

	P6	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5′	P6′
A	865	595	1,020	485	620	825	1,545	330	710	590	930	365
C	0	0	0	0	130	0	0	0	0	0	0	0
D	580	460	255	215	185	805	300	1,185	315	235	170	400
E	530	465	385	455	250	605	300	275	460	150	465	440
F	50	275	430	165	430	145	190	250	165	575	220	50
G	300	575	750	230	545	175	185	415	760	185	320	780
H	175	125	10	150	280	65	165	110	65	35	385	90
I	505	125	385	335	180	510	60	845	220	450	405	170
K	220	580	240	945	455	80	365	115	370	675	600	655
L	540	550	535	525	230	1,185	345	900	1,120	435	670	565
M	10	0	95	10	15	35	70	40	75	65	10	60
N	120	285	235	140	240	95	390	215	195	160	270	730
P	365	475	550	365	95	40	480	20	255	240	330	315
Q	45	260	25	20	340	60	75	60	155	305	35	75
R	125	45	45	425	290	275	595	95	155	145	80	60
S	575	365	380	90	535	395	500	215	170	530	490	555
T	735	460	140	365	365	140	150	430	140	360	250	105
V	270	380	365	920	840	450	275	410	695	705	355	695
W	40	15	35	30	20	30	0	50	0	0	50	10
Y	85	100	255	265	90	220	145	175	110	295	100	15
sum (pmol)	6,135	6,135	6,135	6,135	6,135	6,135	6,135	6,135	6,135	6,135	6,135	6,135

APPENDIX C: SAMPLE CODE

// COMPUTE EPITOPE SUPERTYPES
// Computation rules based on A2, A3 and B7 supertypes

```php
<?php
//connect
$conn=mysql_connect("Localhost","root","") or die(mysql_error());
// select database
mysql_select_db("epitopedb",$conn) or die(mysql_error());
//query database and select only 9 MER epitopes
$result=mysql_query("SELECT     epitopeid,linearsequence     FROM     epitope     where
length(linearsequence)=9",$conn) or die(mysql_error());
?>
  <h3 style = "color: cyan" align="center">
    SUPERTYPES RESULTS</h3>
    <table border = "1" cellpadding = "3" cellspacing = "2" style = "background-color:
#ADD8E6"
                          align="center">
              <tr>
              <th>NO</th>
              <th>Sequence</th>
               <th>Supertype </th>
              </tr>
<?php
 // fetch records in result set
       while($row=mysql_fetch_array($result)){
// put result set into hash/associative array
        $epitopeid = $row["epitopeid"];
        $linearsequence = $row["linearsequence"];
// process supertypes
                      if($linearsequence[1]=="P"){
                              $Supertype="B7";
                      }
                      elseif($linearsequence[8]=="H"){
                              $Supertype="A3";
                      }
                      elseif($linearsequence[8]=="K"){
                              $Supertype="A3";
                      }
                      elseif($linearsequence[8]=="R"){
                              $Supertype="A3";
                      }
                      elseif($linearsequence[8]=="A"){
                              $Supertype="A2";
                      }
                      elseif($linearsequence[8]=="C"){
                              $Supertype="A2";
                      }
                      elseif($linearsequence[8]=="F"){
                              $Supertype="A2";
```

57

```php
                }
                elseif($linearsequence[8]=="I"){
                        $Supertype="A2";
                }
                elseif($linearsequence[8]=="L"){
                        $Supertype="A2";
                }
                elseif($linearsequence[8]=="M"){
                        $Supertype="A2";
                }
                elseif($linearsequence[8]=="P"){
                        $Supertype="A2";
                }
                elseif($linearsequence[8]=="V"){
                        $Supertype="A2";
                }
                elseif($linearsequence[8]=="W"){
                        $Supertype="A2";
                }
                else
                {
                        $Supertype="O";
                }
// Print output
        echo"<tr>";
        echo"<td>$epitopeid</td>";
        echo"<td>$linearsequence</td>";
        echo"<td>$Supertype</td>";
        echo"</tr>";
                }
            echo"</table>";
    mysql_close( $conn );
     ?>


// CALCULATE TAP SCORE

<body background="image001.jpg">

<h1 style = "color: #4974B3" align="center">POLYTOPES REPORT </h1> <h2 style = "color:

#4974B3" align="center">POLYTOPES  DETAILS</h2>

<h2><?php echo date('l, F dS Y');?></h2>

<hr width="100%" size="2" color="#FF0000" align="center">

<?php

//connect

$conn=mysql_connect("Localhost","root","") or die(mysql_error());

// select database
```

```php
mysql_select_db("epitopedb",$conn) or die(mysql_error());
//query database
$result=mysql_query("SELECT    pid,organism,polytope    FROM    polytope4",$conn)    or
die(mysql_error());
?>
  <h3 style = "color: black" align="center">
    POLYTOPES  RESULTS</h3>
    <table  border  =  "1"  cellpadding  =  "3"  cellspacing  =  "2"  style  =  "background-color:
#ADD8E6" align="center">
                <tr>
                <th>NO</th>
                <th>ORGANISM</th>
                <th>POLYTOPE</th>
                <th>TAPSCORE</th>
                <th>PCSCORE</th>
                <th>TOTAL</th>
                </tr>
     <?php
       // fetch records in result set
      while($row=mysql_fetch_array($result)){
                $pid = $row["pid"];
                $organism = $row["organism"];
                $polytope = $row["polytope"];
                // process   TAP scores
                // compute N1 OR position 1 score
                if($polytope[0]=="A"){
                        $N1=-1.56;
                }
                    elseif($polytope[0]=="C"){
                            $N1=0.05;
                }
                    elseif($polytope[0]=="D"){
                            $N1=1.37;
                }
```

```
        elseif($polytope[0]=="E"){
                $N1=1.65;
}
        elseif($polytope[0]=="F"){
                $N1=1.03;
}
        elseif($polytope[0]=="G"){
                $N1=0.28;
}
        elseif($polytope[0]=="H"){
                $N1=0.21;
}
        elseif($polytope[0]=="I"){
                $N1=-0.11;
}
        elseif($polytope[0]=="K"){
                $N1=-1.03;
}
        elseif($polytope[0]=="L"){
                $N1=-0.50;
}

        elseif($polytope[0]=="M"){
                $N1=-0.38;
}
        elseif($polytope[0]=="N"){
                $N1=-1.43;
}
        elseif($polytope[0]=="P"){
                $N1=1.43;
}
        elseif($polytope[0]=="Q"){
                $N1=0.47;
}
```

```
            elseif($polytope[0]=="R"){
                        $N1=-1.34;
}
            elseif($polytope[0]=="S"){
                        $N1=-0.56;
}
            elseif($polytope[0]=="T"){
                        $N1=-0.12;
}
            elseif($polytope[0]=="V"){
                        $N1=-0.49;
}
            elseif($polytope[0]=="W"){
                        $N1=0.54;
}
            elseif($polytope[0]=="Y"){
                        $N1=0.50;
}
            else
{
            $N1=0;
}
// compute N2 score
if($polytope[1]=="A"){
            $N2=-0.25;
}
            elseif($polytope[1]=="C"){
                        $N2=-0.01;
}
            elseif($polytope[1]=="D"){
                        $N2=1.42;
}
            elseif($polytope[1]=="E"){
                        $N2=0.02;
```

```
        }
        elseif($polytope[1]=="F"){
                $N2=-0.45;
        }
        elseif($polytope[1]=="G"){
                $N2=1.14;
        }
        elseif($polytope[1]=="H"){
                $N2=0.33;
        }
        elseif($polytope[1]=="I"){
                $N2=-0.49;
        }
        elseif($polytope[1]=="K"){
                $N2=-0.41;
        }
        elseif($polytope[1]=="L"){
                $N2=0.09;
        }
        elseif($polytope[1]=="M"){
                $N2=-0.46;
        }
        elseif($polytope[1]=="N"){
                $N2=0.69;
        }
        elseif($polytope[1]=="P"){
                $N2=3.00;
        }
        elseif($polytope[1]=="Q"){
                $N2=-0.97;
        }
        elseif($polytope[1]=="R"){
                $N2=-1.47;
        }
```

```
                elseif($polytope[1]=="S"){
                                $N2=-0.34;
        }
                elseif($polytope[1]=="T"){
                                $N2=-0.04;
        }
                elseif($polytope[1]=="V"){
                                $N2=-0.50;
        }
                elseif($polytope[1]=="W"){
                                $N2=-0.64;
        }
                elseif($polytope[1]=="Y"){
                                $N2=-0.67;
        }
                else
        {
        $N2=0;
        }
// Compute N3

        if($polytope[2]=="A"){
                        $N3=-0.10;
        }
                elseif($polytope[2]=="C"){
                                $N3=-0.02;
        }
                elseif($polytope[2]=="D"){
                                $N3=1.83;
        }
                elseif($polytope[2]=="E"){
                                $N3=1.51;
        }
                elseif($polytope[2]=="F"){
                                $N3=-1.05;
        }
```

63

```
        elseif($polytope[2]=="G"){
                $N3=1.70;
}

        elseif($polytope[2]=="H"){
                $N3=-0.23;
}

        elseif($polytope[2]=="I"){
                $N3=-0.62;
}

        elseif($polytope[2]=="K"){
                $N3=0.09;
}
        elseif($polytope[2]=="L"){
                $N3=-0.11;
}

        elseif($polytope[2]=="M"){
                $N3=-0.58;
}
        elseif($polytope[2]=="N"){
                $N3=1.01;
}

        elseif($polytope[2]=="P"){
                $N3=0.22;
}

        elseif($polytope[2]=="Q"){
                $N3=0.39;
}


        elseif($polytope[2]=="R"){
                $N3=-0.42;
}


        elseif($polytope[2]=="S"){
                $N3=0.11;
}
        elseif($polytope[2]=="T"){
                $N3=0.43;
}
        elseif($polytope[2]=="V"){
                $N3=-0.71;
}
        elseif($polytope[2]=="W"){
```

```
                                        $N3=-1.65;
        }
                elseif($polytope[2]=="Y"){
                                        $N3=-1.80;
        }

        else
        {
                $N3=0;
        }


// compute C-terminal score
        if($polytope[32]=="A"){
                                $C=0.55;

        }
                elseif($polytope[32]=="C"){
                                        $C=0.00;

        }
                elseif($polytope[32]=="D"){
                                        $C=1.83;

        }
                elseif($polytope[32]=="E"){
                                        $C=1.58;

        }
                elseif($polytope[32]=="F"){
                                        $C=-2.52;

        }
                elseif($polytope[32]=="G"){
                                        $C=1.41;

        }
                elseif($polytope[32]=="H"){
                                        $C=0.55;

        }
                elseif($polytope[32]=="I"){
                                        $C=-0.52;

        }
                elseif($polytope[32]=="K"){
```

```
                    $C=-0.45;
        }
            elseif($polytope[32]=="L"){
                    $C=-0.94;
        }
            elseif($polytope[32]=="M"){
                    $C=-0.29;
        }
            elseif($polytope[32]=="N"){
                    $C=1.33;
        }
            elseif($polytope[32]=="P"){
                    $C=-0.09;
        }
            elseif($polytope[32]=="Q"){
                    $C=0.12;
        }
            elseif($polytope[32]=="R"){
                    $C=-1.47;
        }
            elseif($polytope[32]=="S"){
                    $C=2.26;
        }
            elseif($polytope[32]=="T"){
                    $C=0.72;
        }
            elseif($polytope[32]=="V"){
                    $C=-0.03;
        }
            elseif($polytope[32]=="W"){
                    $C=-0.87;
        }
            elseif($polytope[32]=="Y"){
                    $C=-2.91;
```

```php
                    }
                            else
                    {
                            $C=0;
                    }
// compute TAP SCORE
        $TAP=number_format(($N1+ $N2+$N3+ $C),3);
// calculate proteasome cleavage
// proteasome cleavage epitope 1 C-terminal score
                            if($polytope[8]=="A"){
                                    $c_terminal1=0.825;

                    }
                            elseif($polytope[8]=="C"){
                                            $c_terminal1=0.000;

                    }
                            elseif($polytope[8]=="D"){
                                            $c_terminal1=0.805;

                    }
                            elseif($polytope[8]=="E"){
                                            $c_terminal1=0.605;

                    }
                            elseif($polytope[8]=="F"){
                                            $c_terminal1=0.145;

                    }
                            elseif($polytope[8]=="G"){
                                            $c_terminal1=0.175;

                    }
                            elseif($polytope[8]=="H"){
                                            $c_terminal1=0.065;

                    }
                            elseif($polytope[8]=="I"){
                                            $c_terminal1=0.510;

                    }
                            elseif($polytope[8]=="K"){
```

```
                                $c_terminal1=0.080;
        }
        elseif($polytope[8]=="L"){
                                $c_terminal1=1.185;
        }


        elseif($polytope[8]=="M"){
                                $c_terminal1=0.035;
        }
        elseif($polytope[8]=="N"){
                                $c_terminal1=0.095;
        }
        elseif($polytope[8]=="P"){
                                $c_terminal1=0.040;
        }
        elseif($polytope[8]=="Q"){
                                $c_terminal1=0.060;
        }
        elseif($polytope[8]=="R"){
                                $c_terminal1=0.275;
        }
        elseif($polytope[8]=="S"){
                                $c_terminal1=0.395;
        }
        elseif($polytope[8]=="T"){
                                $c_terminal1=0.140;
        }
        elseif($polytope[8]=="V"){
                                $c_terminal1=0.450;
        }
        elseif($polytope[8]=="W"){
                                $c_terminal1=0.030;
        }
        elseif($polytope[8]=="Y"){
```

```php
                    $c_terminal1=0.220;
        }
            else
    {
    $c_terminal1=0;
    }
// proteasome cleavage epitope 2 C-terminal score

            if($polytope[20]=="A"){
                $c_terminal2=0.825;
        }
            elseif($polytope[20]=="C"){
                        $c_terminal2=0.000;
        }
            elseif($polytope[20]=="D"){
                        $c_terminal2=0.805;
        }
            elseif($polytope[20]=="E"){
                        $c_terminal2=0.605;
        }
            elseif($polytope[20]=="F"){
                        $c_terminal2=0.145;
        }
            elseif($polytope[20]=="G"){
                        $c_terminal2=0.175;
        }
            elseif($polytope[20]=="H"){
                        $c_terminal2=0.065;
        }
            elseif($polytope[20]=="I"){
                        $c_terminal2=0.510;
        }
            elseif($polytope[20]=="K"){
                        $c_terminal2=0.080;
```

```
        }
        elseif($polytope[20]=="L"){
                $c_terminal2=1.185;
        }
        elseif($polytope[20]=="M"){
                $c_terminal2=0.035;
        }
        elseif($polytope[20]=="N"){
                $c_terminal2=0.095;
        }
        elseif($polytope[20]=="P"){
                $c_terminal2=0.040;
        }
        elseif($polytope[20]=="Q"){
                $c_terminal2=0.060;
        }
        elseif($polytope[20]=="R"){
                $c_terminal2=0.275;
        }
        elseif($polytope[20]=="S"){
                $c_terminal2=0.395;
        }
        elseif($polytope[20]=="T"){
                $c_terminal2=0.140;
        }
        elseif($polytope[20]=="V"){
                $c_terminal2=0.450;
        }
        elseif($polytope[20]=="W"){
                $c_terminal2=0.030;
        }
        elseif($polytope[20]=="Y"){
                $c_terminal2=0.220;
        }
```

```
                    else
            {
                    $c_terminal2=0;
            }
// proteasome cleavage epitope 3 C-terminal score
            if($polytope[32]=="A"){
                            $c_terminal3=0.825;
            }
                    elseif($polytope[32]=="C"){
                                    $c_terminal3=0.000;
                    }
                    elseif($polytope[32]=="D"){
                                    $c_terminal3=0.805;
                    }
                    elseif($polytope[32]=="E"){
                                    $c_terminal3=0.605;
                    }


                    elseif($polytope[32]=="F"){
                                    $c_terminal3=0.145;
                    }
                    elseif($polytope[32]=="G"){
                                    $c_terminal3=0.175;
                    }
                    elseif($polytope[32]=="H"){
                                    $c_terminal3=0.065;
                    }
                    elseif($polytope[32]=="I"){
                                    $c_terminal3=0.510;
                    }
                    elseif($polytope[32]=="K"){
                                    $c_terminal3=0.080;
                    }
                    elseif($polytope[32]=="L"){
```
71

```
                                    $c_terminal3=1.185;
        }
        elseif($polytope[32]=="M"){
                    $c_terminal3=0.035;
        }
        elseif($polytope[32]=="N"){
                    $c_terminal3=0.095;
        }
        elseif($polytope[32]=="P"){
                    $c_terminal3=0.040;
        }
        elseif($polytope[32]=="Q"){
                    $c_terminal3=0.060;
        }
        elseif($polytope[32]=="R"){
                    $c_terminal3=0.275;
        }
        elseif($polytope[32]=="S"){
                    $c_terminal3=0.395;
        }
        elseif($polytope[32]=="T"){
                    $c_terminal3=0.140;
        }
        elseif($polytope[32]=="V"){
                    $c_terminal3=0.450;
        }
        elseif($polytope[32]=="W"){
                    $c_terminal3=0.030;
        }
        elseif($polytope[32]=="Y"){
                    $c_terminal3=0.220;
        }
        else
        {
```

```php
                    $c_terminal3=0;
                }
// calculate polytope proteasome cleavage
$pcscore=number_format(($c_terminal1+$c_terminal2+$c_terminal3),3);
//calculate total
$total=number_format(($TAP+$pcscore),3);
// print output
        echo"<tr>";
        echo"<td>$pid</td>";
        echo"<td>$organism</td>";
        echo"<td>$polytope</td>";
        echo"<td>$TAP</td>";
        echo"<td>$pcscore</td>";
        echo"<td>$total</td>";
          echo"</tr>";
            }
    echo"</table>";
  mysql_close( $conn );
   ?>
     </pre>
     </body>
```