***In Silico* Drug Designing Against *Klebsiella pneumoniae* Adhesin Protein**

A DISSERTATION SUBMITTED TO BRAC UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR

THE BACHELOR OF SCIENCE IN BIOTECHNOLOGY

Submitted by:

Muntasirul Hoq

Student ID: 15136024

January 2019

Program: BS in Biotechnology

Department of Mathematics and Natural Sciences

BRAC University

Dhaka, Bangladesh

Dedicated to

my parents and my teachers

DECLARATION

I hereby declare that the research work embodying the results reported in this thesis title entitled **'*In Silico* Drug Designing Against *Klebsiella pneumoniae* Adhesin Protein'** submitted by the undersigned has been carried out under supervision of Dr. Aparna Islam, Professor, Biotechnology Program, Department of Mathematics and Natural Sciences, BRAC University, Dhaka. It is further declared that the research work presented here is original and has not been submitted to any other institution for any degree or diploma.

Candidate:

____

(Muntasirul Hoq)

Certified:

___________________________________	___________________________________
(Dr. Aparna Islam)	**(Mr. M H M Mubassir)**
Supervisor	**Supervisor**
Professor Biotechnology Program	Lecturer Biotechnology Program

Department of Mathematics and Natural Sciences

BRAC University, Dhaka

Acknowledgement

First of all, I am thankful to Omnipotent Allah for my good health and wellbeing that were necessary for this work to be done.

I am grateful to my parents and my family for always believing in me, supporting me through my ups and downs and encouraging me to do better. They have been a constant support through thick and thin and inspired me tirelessly to follow my dreams.

This research would not be completed if my supervisors, Dr. Aparna Islam and Mr. M H M Mubassir, did not provide me insight and expertise into this work. They consistently steered me into the right direction with their constructive criticism and warm regards. The doors to their office room were always open whenever a blunder was made.

I express my sincere gratitude to Ms. Shamira Tabrejee for helping me out throughout this research study.

Lastly, I would also like to take this opportunity to acknowledge Maliha Islam, Tahrima Mohsin, Farzana Mansur, Afra Abreshmi for always extending their hands whenever I needed help. Special thanks go to Fahad, Afroza, and Safwan for always staying beside me, helping me to find out the good things from bad times and comforting me whenever I broke down during this research work.

(Muntasirul Hoq)
January, 2019

List of Abbreviations

The following abbreviations have been used throughout the text

PK/PD	Pharmacokinetic/Pharmacodynamic
KPC	*Klebsiella pneumoniae* Carbapenemase
NCBI	National Center for Biotechnology Information
MHC	Major Histocompatibility Complex
ESBL	Extended Spectrum β-lactamase
CPS	Capsular polysaccharide
LPS	Lipopolysaccharide
IEDB	Immune Epitope Database
TAP	Transporter Associated with Antigen Processing

Abstract

Klebsiella pneumoniae, is one of the causative agents of many nosocomial infections. In spite of having a plethora of information on the virulent factors of this bacteria, no definite vaccines have yet been designed or developed in order to prevent the infections caused by the multidrug resistant pathogen. Hence, taking this in account, the present study designed an *in silico* peptide vaccine against *K. pneumoniae*, by predicting both B-cell and T-cell epitopes which is followed by molecular docking. In this study, peptide sequences of Type 1 fimbrial adhesin protein of 28 strains of *K. pneumoniae* were retrieved and then tested by *in silico* methods to identify the conserved antigenic portions, present in all the 28 strains of the pathogen. VTLQRGSAY and IYDSRTDKPW were found to be the most potential T-cell and B-cell epitope, among all the other antigenic portions. Furthermore, the T-cell epitope VTLQRGSAY, when docked with HLA-A protein, fitted perfectly inside the groove of it. These results prove that the designed epitopes could be a potent candidate for vaccine development against *K. pneumoniae*.

Table of contents

Chapter number	Contents	Page number
Chapter-1	Introduction	01
Chapter-2	Materials and Method	09
Chapter-3	Results	12
Chapter-4	Discussion	22
Chapter-5	Conclusion	27
Chapter-6	References	29

Chapter – 1: Introduction

CHAPTER – 1

INTRODUCTION

Drugs are chemicals that affects the processes of the mind or body and is used in diagnosis, treatment and prevention of diseases. It also provides assistance in restoring health of diseased individuals. Therefore, they play a vital role in modern medicine. Medicinal chemistry is that field of science that provides these drugs either through discovery or through design. The classical drugs were basically discovered by the empirical observation using naturally occurring substances from the environment. In the last two centuries, drugs were also prepared by chemical alteration of natural substances. In the third millennium, all of these techniques are still in use and a researcher of drug design and development must acknowledge their relative value. In addition to this there are novel opportunities made possible by the deeper understanding of cell biology and genetics. Drug discovery is one of the most significant aspect of the pharmaceutical industry's Research and Development (R&D) process and is the essential foundation in the generation of any robust, innovative drug pipeline. The process of drug development targets the identification of compounds with pharmacological interest to aids in the treatment of diseases and ultimately to improve the quality of life. The compounds used in pharmacology are majorly small organic molecules (ligands) which interact with specific biomolecules (receptors).

1.1 Limitations of Traditional Drug Discovery Processes

In the distant past, designing a new drug by changing the molecular structure of an existing drug was a time-consuming process of trials and errors. Nowadays, a computer can display the molecular structure of any drug from a list of thousands present in a database. With only very slight molecular changes, the original drug may be significantly changed in a variety of ways that can influence the absorption, metabolism, half-life, therapeutic effect, or side effects of the drug. The computer can also be used to identify the chemicals that would probably not be useful in treating a particular disease before time and money are invested in extensive testing (Taylor, 2015). Using computers to manipulate chemicals at the molecular level and design new drugs is

based on molecular pharmacology, the study of the chemical structures of drugs and their interactions at the molecular level within a cell and even within DNA inside the nucleus (Trosset and Carbonell, 2015). Traditionally, drugs are discovered by synthesizing compounds in a time-consuming multi step process against a battery of *in vivo* biological screens and further analyzing the promising candidates for their pharmacokinetic properties, metabolism and potential toxicity. Such a development process has resulted in high attrition rates with failures attributed to poor pharmacokinetics (39%), lack of efficacy (30%), animal toxicity (11%), adverse effects in humans (10%) and various commercial and miscellaneous factors. Moreover, there are an estimated 35,000 open reading frames in the human genome, which, in turn, generate an estimated 500,000 proteins in the human proteome (Kalyani et al., 2013).

Approximately 10,000 of those proteins have been characterized crystallographically. Basically, it means that there are about 490,000 unknowns that may potentially foil any scientific effort (Nehete et al., 2013). Hence, this proves that drug designing is a very lengthy and a difficult task.

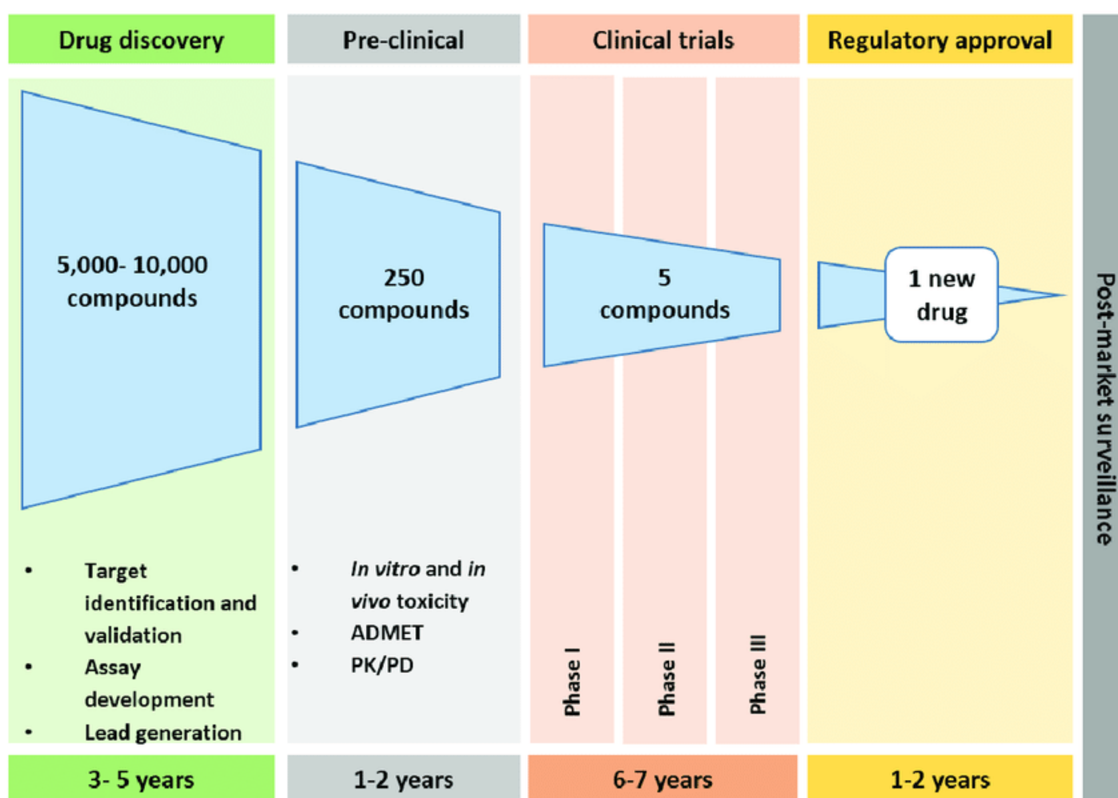


Figure 1.1: Drug discovery and Development Timeline

(Source: https://www.researchgate.net/figure/Drug-discovery-and-development-timeline-The-current-drug-approval-pipeline-can-take-15_fig1_308045230, date: 16. 01. 19)

1.2 Drug Designing

Drug designing, also referred to as rational drug design (or more simply rational design), is the inventive process of designing new medications based on the knowledge of biological targets. Rational drug design can be broadly classified into two categories: development of small molecules with desired properties toward targets, biomolecules (proteins or nucleic acids), whose functional roles in cellular processes and 3D structural information are previously known. This approach in drug design is well established, and is being utilized extensively by the pharmaceutical industries (Vu et al., 2015). Another approach is the development of small molecules with predefined properties toward targets, whose cellular functions and their structural information may be known or unknown.

Basically, drug design involves design of small molecules that are complementary in shape and charge to the biomolecular target to which they interact and will, therefore, bind to it. The identification of a potential drug target is valuable and important in the research and development of drug molecules at early stages. Due to throughput limitations, accuracy and cost, experimental techniques cannot be applied widely. Therefore, the development of *in silico* target identification algorithms, as a strategy with the advantage of fast rate and low cost, has been receiving more and more attention across the world. It has been of great significance to develop a fast and accurate target identification and prediction methodology for the discovery of targeted drugs, construction of drug-target interaction network along with the analysis of small molecule regulating network (Vu et al., 2015).

1.3 *In silico* Drug Design

In silico means "computer aided". This phrase was coined in 1989 as an analogy to the Latin phrases *in vivo*, *in vitro*, and *in situ* etc. Thus, *in silico* drug design means the rational design by which drugs are designed/discovered by using computational methods. Most of the drugs in the past were discovered either by coincidence or by trial and error method, in other words, chances or accidents played a significant part in finding new drugs (Kubinyi, 1993). Latest trend in drug discovery has moved from discovery to design, which needs deep understanding the biochemistry of the disease, pathways, identifying disease causative proteins and then designing compounds that are capable of modulating the role of these proteins.

Hence, this has become a common practice in biopharmaceutical industries worldwide. Both experimental and computational methods play major roles in the discovery of drugs and their development and most of the times run complementing each other (Vu et al., 2015).

There are various computational techniques available which are capable of generating the desired effect at various stages of the drug discovery process. The two major disciples of Computer Aided Drug Design, CADD, which can impact modern day drug discovery process and which are capable of accelerating the drug discovery process are bioinformatics and cheminformatics. In general:

a. Bioinformatic techniques hold a lot of prospective in target identification (generally proteins/enzymes), target validation, understanding the protein, evolution and phylogeny and protein modeling.

b. Cheminformatic techniques hold a lot of prospective in storage management and maintenance of information related to chemical compounds and related properties, and importantly in the identification of novel bioactive compounds, and further in lead optimization. Besides, cheminformatic methods are extensively utilized in *in silico* ADME (Absorption, Distribution, Metabolism and Elimination) prediction and related issues that help in the reduction of the late stage failure of compounds.

Thus, *in silico* drug design is a powerful method, especially when used as a tool within an apparatus, for discovering new drug leads against important targets. After a target and a structure of that target are defined, new leads can be designed from chemical principles or chosen from a subset of small molecules that scored well when docked *in silico* against the target. As structural genomics, bioinformatics, cheminformatics, proteomics and computational power continue to explode with new advances, further successes in *in silico* drug design are likely to follow. Each year, new targets are being identified, structures of those targets are being determined at a drastic rate, and our capability to capture a quantitative picture of the interactions between macromolecules and ligands is accelerating (Anderson, 2003).

1.4 *Klebsiella pneumoniae*

Klebsiella pneumoniae is a Gram-negative bacterium. Its infection can cause health hazards like pneumonia, bloodstream infections, wound infections, urinary tract infections, and meningitis.

Klebsiella species may include resistance plasmids (R-plasmids) which provides them with the resistance to antibiotic, such as, ampicillin and carbenicillin. Moreover, the R-plasmids can be transferred to other enteric bacteria of both the same and different species. The outbreak of multidrug-resistant *Klebsiella* sp. in hospitals are mostly caused by new ESBL (extended spectrum β-lactamase) producing strains. In the past several years the widespread of ESBL-producing strains among clinical *Klebsiella* isolates has increased significantly (Shakya et al., 2017).

1.4.1 Risk factors of *K. pneumoniae*

In North America, *Klebsiella* species are among the most common pathogens recovered in intensive care units. *Klebsiella pneumoniae* is also among the bacteria that most readily develop resistance mechanisms to multiple classes of antibiotics, and the prevalence of drug resistance is increasing at an alarming rate. Due to the emergence of extended-spectrum β-lactamases, carbapenems have been the agents of choice for the management of multidrug-resistant *K. pneumoniae* infection (Tamma and Rodriguez-Bano, 2017).

Carbapenem resistance among *K. pneumoniae* emerged a decade ago (Bradford et al., 2004). The reports of carbapenem resistance were initially irregular, and such resistance was assigned with various mechanisms, including high-level production of an AmpC β-lactamase combined with loss of outer membrane proteins or, rarely, efficient carbapenem-hydrolyzing β-lactamases (eg., the class B metallo-β-lactamases). Though, in 2001, a plasmid-mediated class-A β-lactamase enzyme, *K. pneumoniae* carbapenemase type 1 (KPC-1), was discovered in *K. pneumoniae*. At first it was limited to the northeastern United States, where the prevalence of the KPC enzyme among *K. pneumoniae* is as high as 24% in some areas, KPC-type enzymes (i.e., KPC-2 and KPC-3) have now spread across the world at an alarming rate. KPC production may be an even greater issue than how it is currently being recognized, because carbapenem resistance in KPC-producing *K. pneumoniae* isolates is frequently not detected by automated microdilution susceptibility testing routinely used in clinical microbiology laboratories. In majority of the cases, due to the existence of multiple resistance mechanisms associated with KPC-containing plasmids, few therapeutic options are present for treating infections due to KPC-producing *K. pneumoniae*, and high mortality has been reported (Gasink et al., 2009).

KPC-producing *K. pneumoniae* appears to be the next big challenge in antimicrobial resistance. A deeper investigation of the risk factors and impact of this organism is warranted to develop interventions aimed at curbing its rapid emergence.

1.4.2 Virulence factors of *K. pneumoniae*

The virulence factors of *K. pneumoniae* have been identified as capsular polysaccharide (CPS), lipopolysaccharide (LPS), and fimbriae. LPS is a major component of the outer membrane of Gram-negative bacteria, and consists of lipid A, core oligosaccharide, and a long chain polysaccharide (O-antigen). LPS is an important pathogenic determinant in *K. pneumoniae*, which causes pneumonia and bacteremia in man (Paczosa and Mecsas, 2016). O-antigen is responsible for bacterial resistance to complement-mediated killing. On the other hand, CPS is considered to be the most important virulence factor of *K. pneumoniae*. It covers the bacterial surface and is responsible for its resistance to host phagocytes and serum and promotes inflammation and sepsis. The colonization of mucous membranes by bacteria is the result of adhesion between the bacterial capsule and the host's mucous layer. The adhesion of *K. pneumoniae* to mammalian tissue is conciliated by two varieties of bacterial pili, type 1 and type 3. Type 1 fimbriae play an important role in bacterial adhesion to the D-mannose moiety on mammalian cell surfaces due to the specific affinity of fimbrial protein at the tips of fimbriae. Type 3 fimbriae are characterized by their affinities for a range of mammalian cells, which include bladder epithelial cells, uroepithelial cells, and endothelial cells (Huynh et al., 2017).

1.5 *In silico* Drug designing against *K. pneumoniae*

K. pneumoniae is an antibiotic resistant pathogen and as a result of which vaccination against it can play a vital role in preventing its infection. Among all the virulence factors of *K. pneumoniae*, fimbrial adhesin proteins can be utilized to develop an epitope-based vaccine by utilizing tools of bioinformatics as the other virulence factors are mostly carbohydrates.

1.6 Objectives

With the recent advancements in the field of genomics and proteomics, and bioinformatics in general, the chances of vaccine preparation has significantly increased by a very emerging process, termed as epitope-based drug design. Epitopes are the trivial immunogenic portion of a protein sequence, that invokes specific immune reactions. Several studies reveal that vaccines developed by epitope designing shows specific immunogenic responses against various kinds of

pathogens. Hence, epitope predictions performed by using bioinformatics tools can save a lot of money, time and labor required in the entire process.

This study is targeted towards designing epitope-based peptide vaccines (both B-cell and T-cell epitopes) on the basis of the available Type 1 fimbrial adhesin protein sequences of *K. pneumoniae*, using different tools. The predicted epitopes can be utilized to develop vaccines against *K. pneumoniae*.

Chapter - 2: Materials and Methods

CHAPTER – 2

MATERIALS AND METHODS

2.1 Gathering protein sequence, Multiple Sequence Alignment (MSA), and identification of antigenic protein

28 peptide sequences of the Type 1 fimbrial adhesin protein of *K. pneumoniae* were retrieved from the NCBI's BLAST database. Then, Multiple Sequence Alignment was done by using the Clustal Omega tool. Furthermore, VaxiJen v2.0 was used to predict if the sequences had antigenic properties or not.

2.2 Identification of the T-cell epitope, conservancy analysis and allergenicity assessment

The identification of T-cell epitope was done by using the NetCTL 1.2 server. Tools from the immune epitope database (IEDB) were used to calculate MHC-I binding of the identified epitopes and epitope conservancy. Epitope conservancy test was done to ensure that the epitopes were present i.e. conserved in all of the 28 strains. The stabilized matrix base method was used by these tools to calculate the maximal inhibitory concentration (IC_{50}) value of the epitope binding to human leukocyte antigen (HLA) molecules. All the alleles were considered and the length of the peptides was set to 9.0 before running. AllergenFP v.1.0 was used to calculate the allergenicity of the predicted epitope.

2.3 Designing three-dimensional structure of epitope and HLA protein

The three-dimensional structure of the predicted epitope was designed by using PEP-FOLD web-based server. The server predicted five most provable structures and the one having the lowest energy model was chosen and used for further analysis.

To prove the binding of the identified epitope and HLA molecule, the homology modeling was considered due to the unavailability of a related structure in the protein data bank. UniProt database was used to generate the 3-D structure of HLA-A*30:02 (PDB id: 4qnx). PROCHECK was used to check the validity of the generated structure.

2.4 Analyzing Docking Results

AutoDock Vina was used to analyze the docking between the HLA protein and the predicted epitope. Docking was done by setting the parameters to default.

2.5 Identification of the B-cell epitope

BepiPred 1.0 and BCPREDS server were used to identify the B-cell epitopes and the common predicted epitopes (given by both the servers) were taken. Antigenicity of these predicted epitopes were then checked by using VaxiJen v2.0. A tool from IEDB was used to identify the Emini surface accessibility for the predicted antigenic epitopes. Another tool from IEDB was used to check the conservancy of the epitopes having the Emini surface accessibility. Furthermore, the Karplus and Schulz flexibility and Parker hydrophilicity were identified for the most conserved epitope.

Chapter ~ 3: Results

CHAPTER – 3

RESULTS

3.1 Multiple Sequence Alignment of the adhesin protein sequences and identification of antigenic protein

28 peptide sequences of the Type 1 fimbrial adhesin protein of *K. pneumoniae* were retrieved from the NCBI's BLAST database followed by Multiple Sequence Alignment. The jalview of the multiple sequence alignment analysis showed that the Type 1 adhesin protein sequences which were retrieved from the database of 28 different strains of *K. pneumoniae* were mostly conserved (Figure 3.1).

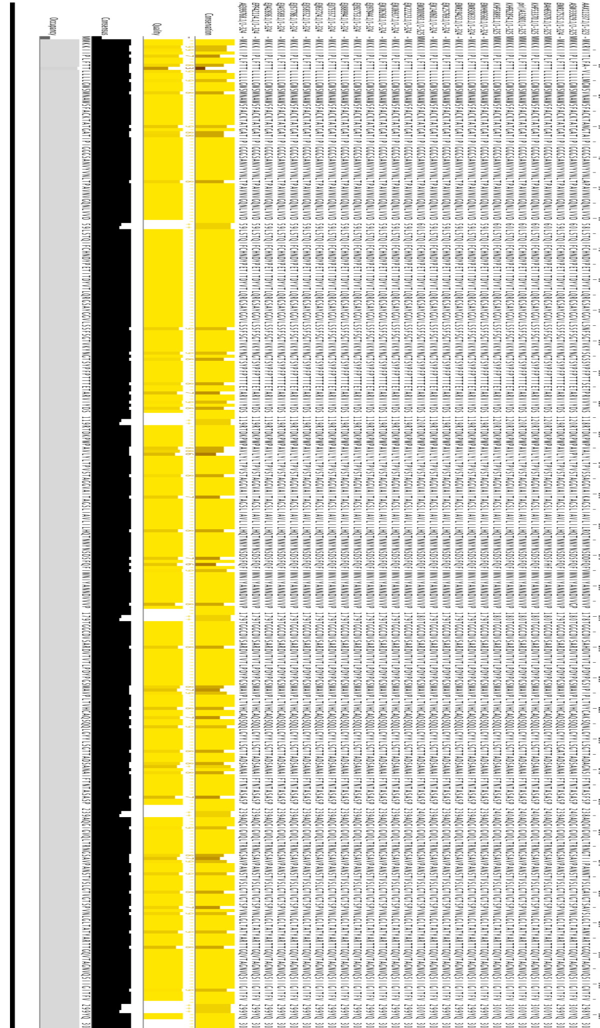


Figure 3.1: Jalview of the Multiple Sequence Alignment analysis

VaxiJen web server was used to check for the antigenicity of the protein sequences. The protein had a maximum total prediction score of 0.6323 which proves it to be a probable antigen as 0.4 is the threshold to be considered as an antigen.

3.2 T-cell epitope identification, analysis of conservancy and allergenicity

Nine best epitopes were predicted by NetCTL server (Table 3.2.1). A 9-mer T-cell epitope VTLQRGSAY was identified to interact (with a high processing score) with MHC-I allele and also having a 100% conservancy (Table 3.2.2) where 100% conservancy means that the epitope VTLQRGSAY is present in all the 28 strains of *K. pneumoniae*.

A peptide having allergenicity will not be accepted and cannot be considered in vaccine designing. Thus, AllergenFP v.1.0 was used to calculate the allergenicity of the predicted epitope i.e. VTLQRGSAY and was found to be a non-allergen.

Table 3.2.1: 9 T-cell epitopes predicted by NetCTL server

Number	Epitope	Overall score
1	ITDYVTLQR	1.0207
2	VTLQRGSAY	1.8762
3	SSFSGTVKY	1.3492
4	TTETARVIY	3.2529
5	LILHQTNNY	1.0336
6	NSDSFQFIW	1.3221
7	CAQSQQLGY	1.5507
8	AQSQQLGYY	1.1204
9	TADSANAIF	1.3228

Table 3.2.2: Interaction with MHC-I allele with the nine predicted T-cell epitopes and the conservancy of these predicted T-cell epitopes

Epitopes	Interacting MHC-I allele (proteasome score, TAP score, processing score)	Epitope Conservancy
ITDYVTLQR	HLA-A*68:01 (1.65)	100%
VTLQRGSAY	HLA-A*30:02 (2.48)	100%
SSFSGTVKY	HLA-A*30:02 (2.79) HLA-A*11:01 (2.79)	96.43%
TTETARVIY	HLA-A*01:01 (2.74)	96.43%
LILHQTNNY	HLA-A*30:02 (2.71)	96.43%
NSDSFQFIW	HLA-B*58:01 (1.76)	92.86%
CAQSQQLGY	HLA-A*01:01 (2.57) HLA-B*53:01 (2.57) HLA-B*35:01 (2.57)	96.43%
AQSQQLGYY	HLA-A*30:02 (2.48) HLA-B*15:01 (2.48)	96.43%
TADSANAIF	HLA-B*35:01 (2.32)	96.43%

3.3 Validation of predicted T-cell epitope

Since the exact structure of the protein was unavailable, the theoretical structure of HLA-A*30:02 protein was obtained by using Homology Modelling. The structure of the hypothetical protein was then examined by a validation software named PROCHECK. The results of the Ramachandran plot (Figure 3.3.1) generated by the PROCHECK server were: 91.8% of the residues of protein are positioned in the most favored regions, 7.2% in the additional allowed regions and 0.6% in the generously allowed regions.

The 3-D structure of the epitope VTLQRGSAY was designed by the PEP-FOLD web-based server (Figure 3.3.3).

The results obtained after performing molecular docking in AutoDock Vina were analyzed using Chimera software. It showed the interactions of docked HLA-A–epitope complexes (Figure 3.3.4).

Thus, the above findings lead to the conclusion that VTLQRGSAY is one of the most suitable T-cell epitopes that can be used for *in silico* vaccine designing against *K. pneumoniae.*

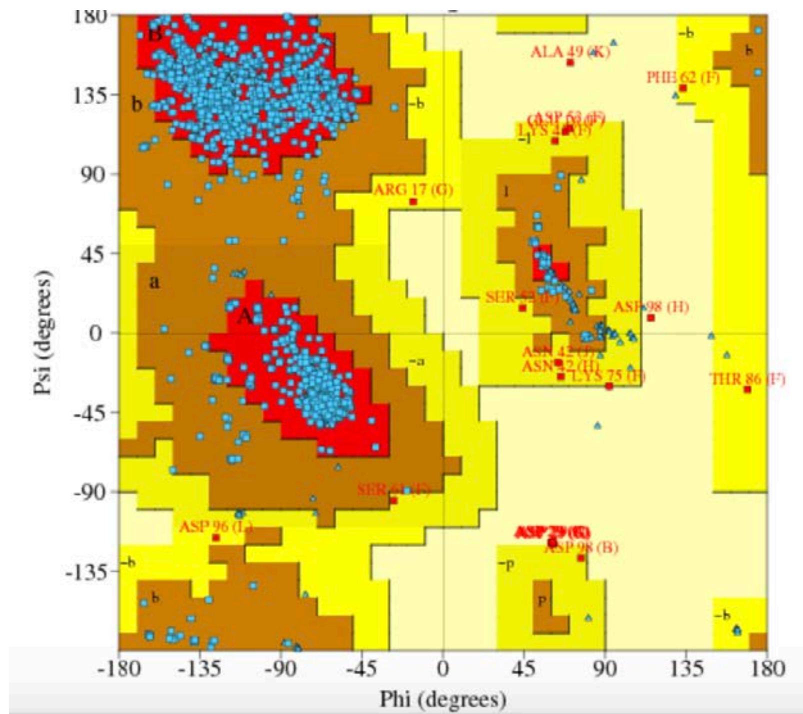


Figure 3.3.1: Evaluation of structural superiority by Ramachandran plot

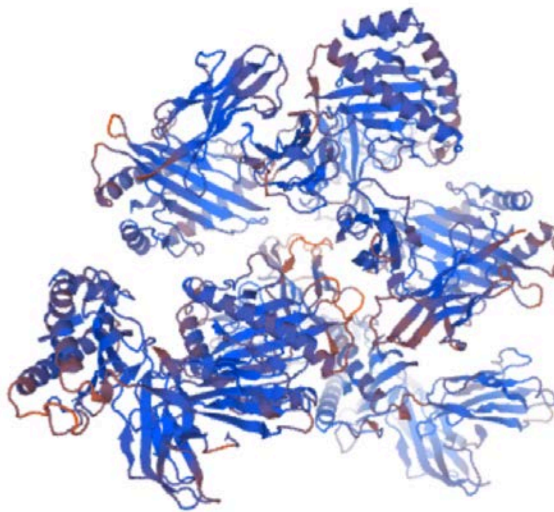


Figure 3.3.2: Cartoon representation of 3-D structure of HLA-A*30:02 protein

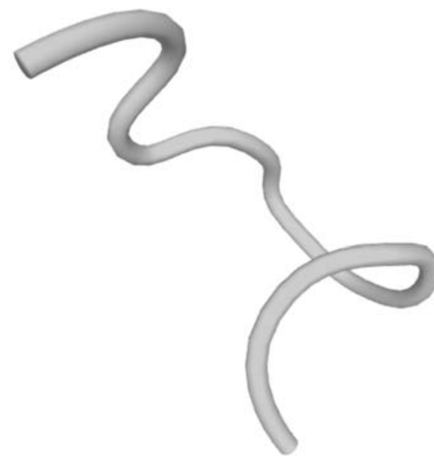


Figure 3.3.3: Cartoon representation of 3-D structure of the epitope VTLQRGSAY

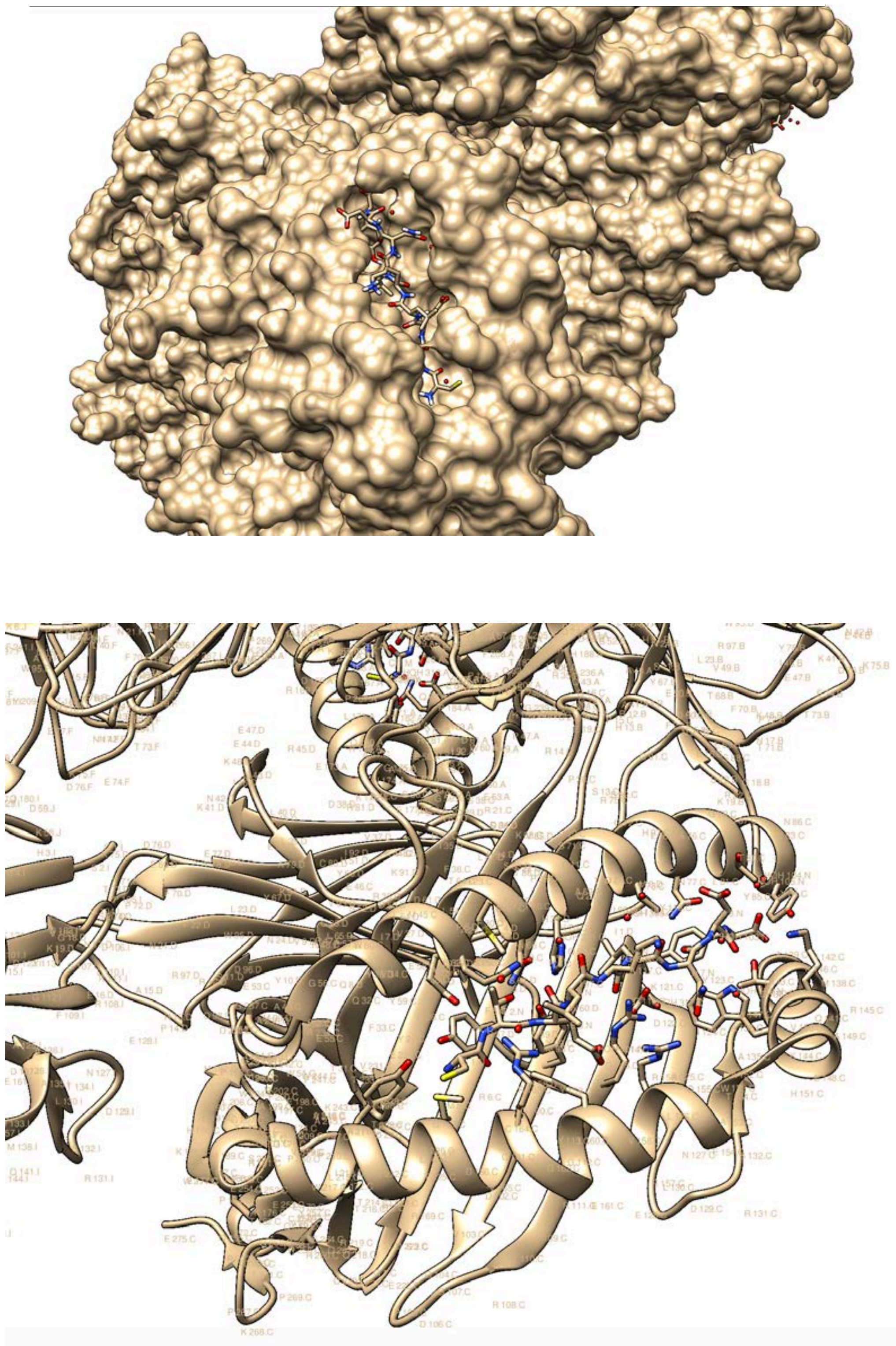


Figure 3.3.4: Molecular docking analysis of HLA-A and epitope VTLQRGSAY complex, where the epitope VTLQRGSAY (stick model) binds in the groove of the protein HLA-A

3.4 B-cell epitope identification

6 epitopes were found in common after using PepiPred 1.0 and BcePred server to identify the B-cell epitopes (Table 3.4). These 6 epitopes were then checked for having antigenicity. 5 epitopes among the 6 predicted epitopes showed antigenicity by the Vaxijen 2.0 server. The results are given below:

ATGATIPIGGGSAN: 1.0909
DSRTDKPWP: 0.6078
TASASPAQG: 1.2001
GSAVPANSTVS: 0.6565
GTVGTSPV: 1.0015
ARTTGQVTAG: 1.2748

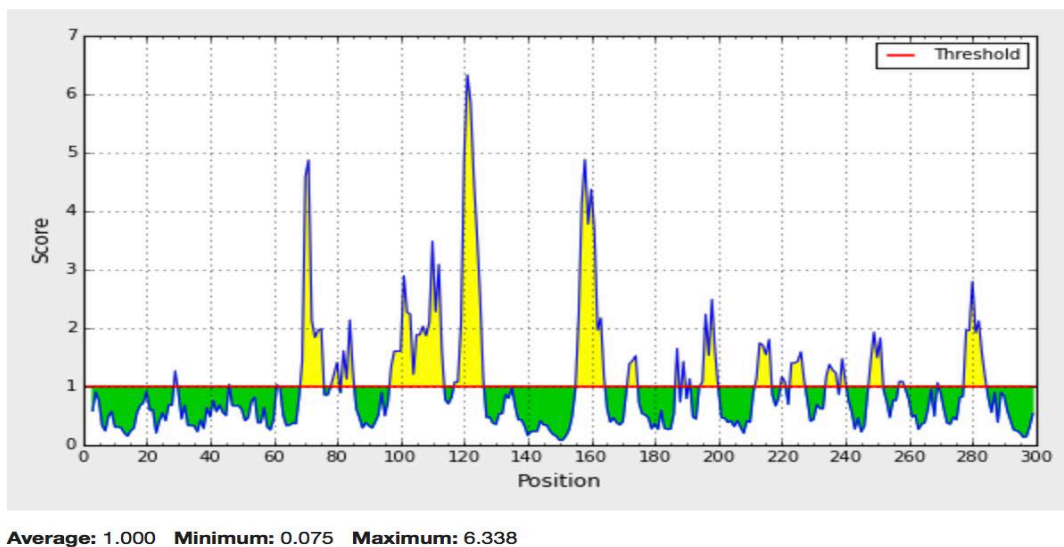
N.B: threshold value for bacteria 0.4

Table 3.4: Common epitopes identified by both the servers are shown in bold letters

Epitopes identified by PepiPred 1.0 server	Epitopes identified by BcePred server
1. **ATGATIPIGGGSAN** (29-42)	1. CKT**ATGATIPIGGGSAN**VYV (26-45)
2. AVN (50-52)	2. YPETITDYVTLQRGSAYGGV (71-90)
3. DYPETIT (70-76)	3. RVIY**DSRTDKPWP**AVLYLTP (115-134)
4. GSAYG (84-88)	4. NAIFTN**TASASPAQG**IGVQL (229-248)
5. TVKYNGTSYPFPTTIET (97-113)	5. **GSAVPANSTVS**L**GTVGTSPV** (252-271)
6. **DSRTDKPWP** (119-127)	6. LTATY**ARTTGQVTAG**NVQSI (275-294)
7. VSTAGGVAI (135-143)	
8. NNYNS (158-162)	
9. VVVPTGGCDVSA (177-188)	

10. VTLPDYPGSMA (193-203)	
11. GTTADSA (222-228)	
12. **TASASPAQG** (235-243)	
13. TRN**GSAVPANSTVS** (249-262)	
14. **GTVGTSPV**N (264-272)	
15. **ARTTGQVTAG** (280-289)	

Then, the surface accessibility of the protein was analyzed by using Emini surface accessibility prediction tool and among the 5 antigenic epitopes, only two showed high surface accessibility (Figure 3.4.1). Moreover, the conservancy analysis of these two epitopes were done and epitope IYDSRTDKPW was found to be 96.43% conserved in all 27 strains out of the 28 strains (Table 3.4.1). Since antigenicity is related to protein flexibility, the Karplus and Schulz flexibility prediction method was used where it was found that most of the residues of the epitope IYDSRTDKPW were above threshold level (Figure 3.4.2), hence, the epitope is proven to be antigenic. Furthermore, hydrophilicity is a property required by a peptide to be a potent B-cell epitope and hence to check the hydrophilicity of the epitope Parker hydrophilicity analysis was done and most of the residues of the epitope was found to be above the threshold level (Figure 3.4.3), thus, the epitope IYDSRTDKPW was proven to be a potent B-cell epitope. Therefore, the peptide sequences staring from 117 to 126 i.e. IYDSRTDKPW should elicit immune reaction as a B-cell epitope.



Average: 1.000 **Minimum:** 0.075 **Maximum:** 6.338

Figure 3.4.1: Emini surface accessibility prediction of the most antigenic peptides. Here, the x-

axis and y-axis represent sequence position and antigenic propensity, respectively. The threshold level is 1. The regions above the threshold are antigenic, shown in yellow, while, green color reflects the polypeptide regions that could not satisfy the threshold margin.

Table 3.4.1: Epitope IYDSRTDKPW shows the maximum conservancy

Epitope sequence	Epitope length	Percent of protein sequence matches at identity <=100%	Minimum Identity	Maximum Identity
IYDSRTDKPW	10	96.43% (27/28)	80.00%	100%
TYARTTG	7	96.43%	71.43%	100%

Figure 3.4.2: Karplus and Schulz flexibility prediction analysis showed most of the residues of the IYDSRTDKPW peptide were found above the threshold levels

Figure 3.4.3: Parker hydrophilicity analysis also showed that most of the residues of the IYDSRTDKPW peptide were found above the threshold levels

Chapter – 4: Discussion

CHAPTER – 4

DISCUSSION

The identification of epitopes can be a time-consuming and a costly process which could be easily overcome by using the different *in silico* drug designing tools. This study includes the designing of epitope-based peptide vaccines against *K. pneumoniae*.

To begin with, 28 sequences of the Type 1 adhesin protein of 28 different strains of K. pneumoniae were retrieved followed by the Multiple Sequence Alignment using Clustal Omega tool. A Multiple Sequence Alignment (MSA) is a sequence alignment of three or more biological sequences, generally protein, DNA, or RNA. It is often used to assess sequence conservation of protein domains, tertiary and secondary structures, and even individual amino acids or nucleotides. This is a necessary first step in the designing and analysis of a vaccine. MSA of the peptide sequences of Type 1 adhesin protein showed that most of the regions of the peptides are conserved apart from a few amino acids which is negligible. This result implies that the Type 1 adhesin is a conserved protein which is present in all the strains of *K. pneumoniae.*

Then, VaxiJen server was used to identify the antigenicity of the peptide sequences which was then followed by the identification of potent epitopes using NetCTL server. In general, an antigen is defined as a substance that a system is not acquainted with, i.e., a foreign substance (Britannica, 2018). Here, antigenicity predictive value was determined, which predicts whether a protein can serve as a probable antigen or not. Vaxijen prediction has a default threshold value of 0.4 for bacteria (Doytchinova and Flower, 2007). The Type 1 adhesin protein sequences had a maximum total prediction score of 0.6323 (above threshold) which implies that the protein is identified as antigen by our system. Since, the Type 1 adhesin protein has been identified as a probable antigen, it was then further analyzed to identify potent epitopes, which is the portion of the antigen that provokes immune reaction against the antigen. By observing the results, it was found that the epitope CAQSQQLGY binds to the maximum number of MHC-I alleles (Table 3.2.2) but still VTLQRGSAY was selected as it had a good total score (proteasome score, TAP score and processing score) along with a 100% conservancy, where, 100% conservancy means that the epitope VTLQRGSAY is present in all the 28 strains of *K. pneumoniae*. While on the other hand, CAQSQQLGY had a conservancy of 96.43%. Moreover, the total score shown in bracket (Table

3.2.2) is the combination of proteasomal cleavage, TAP transport and MHC binding predictions which in total predicts a quantity proportional to the amount of peptide presented by MHC molecules on the cell surface. Hence, a higher the total score found for an epitope indicates that it has a higher chance of being presented at the cell surface by MHC-I which then leads to further immune reactions. As a result, epitope VTLQRGSAY was selected as it had a higher total score along with a 100% conservancy.

Furthermore, an epitope having allergenicity cannot be considered in vaccine designing. It has been reported that a protein can be classified as an allergen when it shows >35% identity with a familiar allergen over a window of 80 amino acids, or the presence of six contiguous amino acids that is present in a known allergen (Breiteneder and Mills, 2004). AllergenFP v.1.0 was used for predicting the allergenicity of the epitope VTLQRGSAY and it was identified as a non-allergen. Thus, the epitope would not cause allergic reactions if it is used in the preparation of the vaccine against *K. pneumoniae.*

Then, the hypothetical 3-D structure of the HLA protein was generated using UniProt database. This involves "Homology Modelling" which refers to the construction of an atomic-resolution model of the "desired" protein from its amino acid sequence and an experimental 3-D structure of a related homologous protein (https://proteinstructures.com/Modeling/homology-modeling.html, date: 16. 01. 19). Since, the selected epitope VTLQRGSAY showed interaction with HLA-A*30:02 (Table 3.2.2), thus, the 3-D structure of this HLA protein was required for docking analysis. However, the 3-D structure of HLA-A*30:02 is not yet developed properly. Thus, homology modelling was performed using the related homologous protein of HLA-A*30:02 in order to find the experimental 3-D structure. 283 residues (100%) of "Crystal structure of HLA-A*01:01 in complex with np44-s7n" matches with HLA-A*30:02 (365 residues) and thus, it was selected among the other models predicted by UniProt, as it had a high identity value than the others.

The hypothetical structure was then validated by using PROCHECK. A protein model can be considered high-quality if it has >90% of the residues in the core and in the allowed regions (https://www.researchgate.net/post/What_exactly_is_a_ramachandran_plot_How_can_we_define_this_plot, date: 16. 01. 19). Since, the results generated by the PROCHECK server showed that 91.8% of the residues of the protein were positioned in the most favored regions, thus, it validated the hypothetical 3-D structure of the protein.

Then, the validated 3-D structure of the protein was used for docking with the epitope VTLQRGSAY. The molecular docking was performed using AutoDock Vina. Docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. It's a computational simulation of a candidate ligand binding to a receptor. Now, the epitope VTLQRGSAY when docked with the HLA-A protein, fitted nicely inside the groove of it. As a result, it can be concluded that the epitope VTLQRGSAY will bind with the HLA-A molecules which will then present the bound-epitope at the cell surface.

On the contrary, the other type of immune response shown by B-cell epitopes were also considered in this study and as a result, the epitope IYDSRTDKPW was identified to be a potent candidate for vaccine designing, by evaluating its antigenicity, conservancy, hydrophilicity, surface accessibility and flexibility.

Firstly, Type 1 adhesin protein sequences retrieved from the NCBI's BLAST database were used for the identification of B-cell epitope using the PepiPred 1.0 and BcePred server. Two servers were used in order to get a more valid result. The servers predicted some epitopes and among them, both the servers commonly predicted six B-cell epitopes. This implies that the identified epitopes can provoke B-cell immune reactions. Then, the common epitopes were checked for the antigenicity. It was done to confirm that the identified epitopes also showed antigenicity. This is mainly because a molecule will only provoke immune reactions if and only if it is identified as a foreign molecule by the system. The results showed that five out of six epitopes satisfied the threshold value requirement for antigenicity and thus were predicted as probable antigens (Table 3.4).

Then, the surface accessibility of the five selected epitopes were analyzed. Surface accessibility determines the amount of accessible molecular area. Each atom can potentially be touched by water or any other solvent, and the area of an atom on the surface that can be touched by the solvent is called the accessible molecular surface, or solvent-exposed area. The area covered by the 'center' of a solvent molecule rolling over the surface of an exposed protein atom can be calculated. This area is larger, of course, and is called the accessible surface. Having a higher surface accessibility value proves that the epitope has an accessible surface in order to bind with other immune

components during immunogenic reactions (Liljeroos et al., 2015). Two epitopes among five showed high surface accessibility. Among these two epitopes, IYDSRTDKPW was found to be more conserved in 27 out of 28 strains of *K. pneumoniae.* This implies that if a vaccine is designed using this conserved peptide sequence, it will provide an immunity against almost all the pathogenic strains of the bacteria.

Furthermore, most of the residues of the epitope IYDSRTDKPW had a predicted value above threshold when tested for protein flexibility and hydrophilicity. Flexibility and hydrophilicity of linear B-cell epitopes are usually determined in order to check if the linear epitope can bind to the antibodies or not (El-Manzalawy et al., 2008). Hence, all these results prove that the epitope IYDSRTDKPW is a potent B-cell epitope.

Chapter – 5: Conclusion

CHAPTER – 5

CONCLUSION

In silico analysis of our study recommended that the predicted B-cell epitope IYDSRTDKPW and T-cell epitope VTLQRGSAY could be a novel vaccine target against a number of diseases, caused by *K. pneumoniae* as these can evoke both the B-cell and T-cell immune responses along with a long-term and protective immunity against the antibiotic resistant *K. pneumoniae.*

Chapter – 6: References

CHAPTER – 6
REFERENCES

Apweiler R, Bairoch A, Wu CH, et al. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res*. 2004;32(Database issue): D115–D119.

Bradford PA, Bratu S, Urban C, et al. Emergence of carbapenem-resistant *Klebsiella* species possessing the class A carbapenem-hydrolyzing KPC-2 and inhibitor-resistant TEM-30 beta-lactamases in New York City. Clin Infect Dis. 2004; 39:55–60.

Breiteneder. H., and Mills. E.N. Molecular properties of food allergens. 2004

Bratu S, Mooty M, Nichani S, et al. Emergence of KPC-possessing *Klebsiella pneumoniae* in Brooklyn, New York: epidemiology and recommendations for detection. Antimicrob Agents Chemother. 2005; 49:3018–3020.

Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2007;8:361.

Buus S, Lauemoller SL, Worning P, et al. Sensitive quantitative predictions of peptide-MHC binding by a ‘Query by Committee’ artificial neural network approach. *Tissue Antigens*. 2003;62(5):378–384.

Bratu S, Landman D, Haag R, et al. Rapid spread of carbapenem-resistant *Klebsiella pneumoniae* in New York City: a new threat to our antibiotic armamentarium. Arch Intern Med. 2005; 165:1430–1435.

Connell, I.; Agace, W.; Klemm, P.; Schembri, M.; Mărild, S.; Svanborg, C. Type 1 fimbrial expression enhances Escherichia coli virulence for the urinary tract. *Proc. Natl. Acad. Sci. USA* 1996, *93*, 9827–9832.

Chen J, Liu H, Yang J, Chou K (2007) Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino Acids 33: 423-428.

D. Plewczynski, A. Philips, M. von Grotthuss, L. Rychlewski and K. Ginalski. HarmonyDOCK: the structural analysis of poses in protein-ligand docking. Journal of Computational Biology, 18, 2010, 1-10.

Dash R, Hosen SZ, Sultana T, et al. Computational Analysis and Binding Site Identification of Type III Secretion System ATPase from Pseudomonas aeruginosa. *Interdiscip Sci*. 2015:1–9.

Dimitrov, I., Naneva, L., Doytchinova, I., Bangov, I. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics.* Volume 30. 2014

Di Martino, P.; Livrelli, V.; Sirot, D.; Joly, B.; Darfeuille-Michaud, A. A new fimbrial antigen harbored by CAZ-5/SHV-4-producing *Klebsiella pneumoniae* strains involved in nosocomial infections. *Infect. Immun.* 1996, *64*, 2266–2273.

Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. 2007;8:4.

Duguid, J. Fimbriae and adhesive properties in Klebsiella strains. *J. Gen. Microbiol.* 1959, *21*, 271–286.

Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *JVirol*. 1985;55(3):836–839.

El-Manzalawy, Y., Dobbs, D., & Honavar, V. (2008). Predicting flexible length linear B-cell epitopes. *Computational systems bioinformatics. Computational Systems Bioinformatics Conference*, *7*, 121-32.

Erridge, C.; Bennett-Guerrero, E.; Poxton, I.R. Structure and function of lipopolysaccharides. *Microbes Infect.* 2002, *4*, 837–851.

Fader, R.; Gondesen, K.; Tolley, B.; Ritchie, D.; Moller, P. Evidence that in vitro adherence of *Klebsiella pneumoniae* to ciliated hamster tracheal cells is mediated by type 1 fimbriae. *Infect. Immun.* 1988, *56*, 3011–3013.

Falagas ME, Bliziotis IA. Pandrug-resistant gram-negative bacteria: the dawn of the post-antibiotic era? Int J Antimicrob Agents. 2007; 29:630–636.

Falagas ME, Rafailidis PI, Kofteridis D, et al. Risk factors of carbapenem-resistant *Klebsiella pneumoniae* infections: a matched case control study. J Antimicrob Chemother. 2007; 60:1124–1130.

Favre-Bonte, S.; Joly, B.; Forestier, C. Consequences of reduction of *Klebsiella pneumoniae* capsule expression on interactions of this bacterium with epithelial cells. *Infect. Immun.* 1999, *67*, 554–561.

Firbas C, Jilma B, Tauber E, et al. Immunogenicity and safety of a novel therapeutic hepatitis C virus (HCV) peptide vaccine: a randomized, placebo controlled trial for dose optimization in 128 healthy subjects. *Vaccine*. 2006;24(20):4343–4353.

Gasink, L. B., Edelstein, P. H., Lautenbach, E., Synnestvedt, M., & Fishman, N. O. (2009). Risk factors and clinical impact of Klebsiella pneumoniae carbapenemase-producing K. pneumoniae. *Infection control and hospital epidemiology*, *30*(12), 1180-5.

Hawkey P, Finch R. Tigecycline: in-vitro performance as a predictor of clinical efficacy. Clin Microbiol Infect. 2007;13:354–62.

H. Hoeltje, W. Sippl, D. Rognan and G. Folkers. Molecular modeling, basics principles and applications (John Wiley & Sons, Chichester, U.K, 2003).

H. Kubinyi. Chance favors the prepared mind from serendipity to rational drug design. Journal of Receptors and Signal Transduction Research, 19, 1999, 15–39.

H. Markus, J. Seifert and K. Bernd (2007). Virtual high-throughput screening of molecular databases. Journal of Current Opinion in Drug Discovery and Development, 10, 2007, 298-307.

He L, Cheng Y, Kong L, et al. Approaching rational epitope vaccine design for hepatitis C virus with meta-server and multivalent scaffold- ing. *Sci. Rep*. 2015;5.

Hornick, D.; Dayton, C.; Bedell, G.; Fick, R. Nontuberculous mycobacterial lung disease. Substantiation of a less aggressive approach. *CHEST J.* 1988, *93*, 550–555.

Huynh, D., Kim, A. Y., & Kim, Y. R. (2017). Identification of Pathogenic Factors in Klebsiella pneumoniae Using Impedimetric Sensor Equipped with Biomimetic Surfaces. *Sensors (Basel, Switzerland)*, *17*(6), 1406. doi:10.3390/s17061406

J. Bajorath. Integration of virtual and high-throughput screening. Nature Reviews Drug Discovery, 1, 2002, 882–894.

K. Gunjan, S. Dinesh, V. Yogesh and S. Vishal (2013). A review on drug designing, methods, its applications and prospects. International Journal of Pharmaceutical Research and Development, 5, 2013, 15-30.

Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. *Naturwissenschaften*. 1985;72(4):212–213.

Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res*. 2006;2:2.

Kwak YG, Choi SH, Choo EJ, et al. Risk factors for the acquisition of carbapenem-resistant *Klebsiella pneumoniae* among hospitalized patients. Microb Drug Resist. 2005; 11:165–169.

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thorn- ton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR*. 1996;8(4):477–486.

Leavitt A, Navon-Venezia S, Chmelnitsky I, Schwaber MJ, Carmeli Y. Emergence of KPC-2 and KPC-3 in carbapenem-resistant *Klebsiella pneumoniae* strains in an Israeli hospital. Antimicrob Agents Chemother. 2007; 51:3026–3029.

MacKenzie FM, Forbes KJ, Dorai-John T, Amyes SG, Gould IM. Emergence of a carbapenem-resistant *Klebsiella pneumoniae*. Lancet. 1997; 350:783.

Liljeroos, L., Malito, E., Ferlenghi, I., & Bottomley, M. J. (2015). Structural and Computational Biology in the Design of Immunogenic Vaccine Antigens. *Journal of immunology research*, *2015*, 156241.
Matthews, Holly & Hanison, James & Nirmalan, Niroshini. (2016). "Omics"-Informed Drug and Biomarker Discovery: Opportunities, Challenges and Future Perspectives. Proteomes. 4. 28. 10.3390/proteomes4030028.

National Nosocomial Infections Surveillance S. National Nosocomial Infections Surveillance (NNIS) System Report, data summary from January 1992 through June 2004, issued October 2004. Am J Infect Control. 2004; 32:470–485.

Neuhauser MM, Weinstein RA, Rydman R, Danziger LH, Karam G, Quinn JP. Antibiotic resistance among gram-negative bacilli in US intensive care units: implications for fluoroquinolone use. JAMA. 2003; 289:885–888.

Paczosa, M. K., & Mecsas, J. (2016). Klebsiella pneumoniae: Going on the Offense with a Strong Defense. *Microbiology and molecular biology reviews : MMBR*, *80*(3), 629-61. doi:10.1128/MMBR.00078-15

Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry. 1986 Sep 23; 25(19):5425-32.

Patel G, Huprikar S, Factor S, Jenkins S, Calfee D. Outcomes of carbapenem-resistant *Klebsiella pneumoniae* infection and the impact of antimicrobial and adjunctive therapies. Infect Control Hosp Epidemiol. 2008; 29:1099–1106.

Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics*. 2005;6:132.

Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605–1612.

Podschun, R; Ullmann, U (October 1998). "Klebsiella spp. as Nosocomial Pathogens: Epidemiology, Taxonomy, Typing Methods, and Pathogenicity Factors". Clinical Microbiology Reviews. 11 (4): 589–603. PMC 88898.

Pope J, Adams J, Doi Y, Szabo D, Paterson DL. KPC type beta-lactamase, rural Pennsylvania. Emerg Infect Dis. 2006; 12:1613–1614.

Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. *Nat Rev Drug Discov*. 2007;6(5): 404–414.

S. Mandal, M. Moudgil and S. Mandal (2009). Rational drug design. European Journal of Pharmacology, 625, 2009, 90-100.

S. Arlington. An industrial revolution in R&D. Pharmaceutical Executive, 20, 2000, 74-84.

Schwaber MJ, Klarfeld-Lidji S, Navon-Venezia S, Schwartz D, Leavitt A, Carmeli Y. Predictors of carbapenem-resistant *Klebsiella pneumoniae* acquisition among hospitalized adults and effect of acquisition on mortality. Antimicrob Agents Chemother. 2008; 52:1028–1033.

Sharmin R, Islam AB. A highly conserved WDYPKCDRA epitope in the RNA directed RNA polymerase of human coronaviruses can be used as epitope-based universal vaccine design. *BMC Bioinformatics*. 2014;15(161):1471–2105.

Shakya, P., Shrestha, D., Maharjan, E., Sharma, V. K., & Paudyal, R. (2017). ESBL Production Among *E. coli* and *Klebsiella* spp. Causing Urinary Tract Infection: A Hospital Based Study. *The open microbiology journal*, *11*, 23-30. doi:10.2174/1874285801711010023

Sominskaya I, Skrastina D, Dislers A, et al. Construction and immunological evaluation of multivalent hepatitis B virus (HBV) core virus-like particles carrying HBV and HCV epitopes. *Clin Vaccine Immunol*. 2010;17(6):102

Staneková Z, Varečková E. Conserved epitopes of influenza A virus inducing protective immunity and their prospects for universal vaccine development. *Virol J*. 2010;7(1):1.

Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Streit JM, Jones RN, Sader HS, Fritsche TR. Assessment of pathogen occurrences and resistance profiles among infected patients in the intensive care unit: report from the SENTRY Antimicrobial Surveillance Program (North America, 2001) Int J Antimicrob Agents. 2004; 24:111–118.

Strettoti, C.W.; Ristuccia, P.A.; Cunha, B.A. Topics in clinical microbiology *Klebsiella*. *Infect. Control* 1984, *5*, 343–348.

T. Lengauer. Bioinformatics from genome to drugs (Wiley- VCH, Weinheim, Germany, 2001).

Tamma, P. D., & Rodriguez-Bano, J. (2017). The Use of Noncarbapenem β-Lactams for the Treatment of Extended-Spectrum β-Lactamase Infections. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, *64*(7), 972-980.

Taylor. D., The Pharmaceutical Industry and the Future of Drug Development, *Pharmaceuticals in the Environment*, 2015, pp. 1-33

Tarkkanen, A.-M.; Virkola, R.; Clegg, S.; Korhonen, T.K. Binding of the type 3 fimbriae of *Klebsiella pneumoniae* to human endothelial and urinary bladder cells. *Infect. Immun.* 1997, *65*, 1546–1549.

Tenover FC, Kalsi RK, Williams PP, et al. Carbapenem resistance in *Klebsiella pneumoniae* not detected by automated susceptibility testing. Emerg Infect Dis. 2006; 12:1209–1213.

Tenzer S, Peters B, Bulik S, et al. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci*. 2005;62(9):1025–1037.

Thevenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tuffery P. PEP- FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res*. 2012;40(Web Server issue):11.

Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31(2):455–461.

U. Madsen, P. Krogsgaard-Larsen and T. Liljefors. Textbook of drug design and discovery (Taylor and Francis, Washington, USA, 2002).

Villegas MV, Lolans K, Correa A, et al. First detection of the plasmid-mediated class A carbapenemase KPC-2 in clinical isolates of *Klebsiella pneumoniae* from South America. Antimicrob Agents Chemother. 2006; 50:2880–2882.

Wu LT, Hung SW, Chuang YC, Chen HE, Jones RN, Yu WL. Identification of a novel cephalosporinase (DHA-3) in Klebsiella pneumoniae isolated in Taiwan. Clin Microbiol Infect. 2005;11:893–7.

Zhang R, Zhou HW, Cai JC, Chen GX. Plasmid-mediated carbapenem-hydrolysing beta-lactamase KPC-2 in carbapenem-resistant *Serratia marcescens* isolates from Hangzhou, China. J Antimicrob Chemother. 2007; 59:574–576.