HIGH THROUGHPUT VIRTUAL SCREENING FOR INHIBITORS OF *Salmonella typhi's* D-ALANINE-D-ALANINE LIGASE A

TAN CHEN SHEN

UNIVERSITI SAINS MALAYSIA

2015

HIGH THROUGHPUT VIRTUAL SCREENING FOR INHIBITORS OF *Salmonella typhi's* D-ALANINE-D-ALANINE LIGASE A

by

TAN CHEN SHEN

Thesis submitted in fulfilment of the requirements

for the degree of

Master of Science

OCTOBER 2015

ACKNOWLEDGEMENT

Big thanks to Dr. Hasni bin Arsad for his kindness, patience and guidance for the past years throughout my tenure as candidate of Master studies in molecular biology at Advanced Medical and Dental Institute. To Professor Dr. Mohd. Razip Samian for his support and handling me a chance to pursue my postgraduate studies in molecular biology. To my best friends Vinoth Mohan and Lee Chu Xian for assisting me throughout the course of my research.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

π system	Pi system
3D	3 Dimensions
Å	Angstrom
ADME/Tox	Absorption, distribution, metabolism, excretion and toxicity
Ala	Alanine
ASN	Aspartate
ATP	Adenosine tri-phosphate
BBB	Blood brain barrier
B_{ij}	Hydrogen bond in a steric state
BLASTp	Basic local alignment search tool for protein sequences
CDC	Centre for Drug Control
CLOGP	Octanol/water partition coefficient
DDLA	D-Alanine-D-Alanine Ligase A
DOPE	Discrete Optimized Protein Energy
E_{bind}	Molecular docking generations of empirical binding energy of between receptor and ligand.
E_{pharma}	Energy from the binding-sites of generated pharmacophores.
E_{ligpre}	Penalty given if ligands does not follow the set parameters.
E_{inter}	Intermolecular of the compounds.
E_{intra}	Intramolecular energy.
E_{penal}	Large penalty value given if the ligand is out range of the search box.

FPSA	Fast polar surface area
GA	GEMDOCK accuracy
GIS	Geographical information system
GLN	Glutamate
GLY	Glycine
HIA	Human intestinal absorption
ILE	Isoleucine
KEGG	Kyoto Encyclopedia of Genes and Genomes.
LYS	Lysine
MDa	Mega Dalton
PDB	Protein Data Bank
PPB	Plasma protein binding
RAM	Random Access Memory
RAMPAGE	Ramachandran Plot Assessment
$r_{ij}^{B_{ij}}$	Relatives distances between two atoms i and j with the interaction type B_{ij} forming by the pair-wise heavy atom interactions between ligands and proteins
RMSD	Root-mean-square deviation
SASA	Solvent accessibility surface area
SER	Serine
T3SS	Type 3 secretion system
UDP	Uridine di-phosphate
VDW	van der Waals
WHO	World Health Organization
w_j	Generated pharmacophore weight of the atom's active-site j,
WP_{elec}	Penalties for the hydrophilic parameters
WP_{hb}	Penalties for electrostatic parameters

SARINGAN MAYA BERPROSESAN TINGGI PERENCAT UNTUK D-ALANINA -D- ALANINA LIGASE A, Salmonella *typhi*

ABSTRAK

Salmonella typhi berintangan berganda merupakan ancaman terhadap rawatan deman tifoid. Peristiwa ini memerlukan rawatan yang baru untuk merencatkan *Salmonella typhi*. Tujuan utama kajian ini adalah untuk mencari perencat yang berpotensi terhadap *Salmonella typhi* dan memahami ciri-ciri prenecatan tersebut. Kajian ini menggunakan daya pemprosesan tinggi saringan maya terhadap D-alanina-D-alanina ligase A kepada pangkalan data ZINC untuk mencari sasaran yang mempunyai potensi untuk menrencatkan *Salmonella typhi*. D-alanina-D-alanina ligase A adalah sasaran prencatan *Salmonella typhi* yang baik kerana ia memainkan peranan yang penting dalam proses biosintesis peptidoglikan, dan ia juga bukan homolog kepada protein manusia. Melalui satu siri analisis seperti analisis pos-penyaringan, pemetaan pharmacophore, dan analisis ADME/Tox, didapati bahawa N-[3-(1H-imidazol-1-yl)propyl]-4-(2-methylbenzoyl)-1H-pyrrole-2-carboxamide (ZINC04025098) memiliki sifat-sifat perencatan yang berpotensi terhadap D-alanina-D-alanina Ligase A.

HIGH THROUGHPUT VIRTUAL SCREENING FOR INHIBITORS OF *Salmonella typhi's* D-ALANINE-D-ALANINE LIGASE A

ABSTRACT

Increasing numbers of multi-resistance *Salmonella typhi* cases poses a huge threat especially to ongoing typhoid treatments. This turn of event calls for new treatments against *Salmonella typhi*. The objectives of this study is to identify potential drug targeted protein from *Salmonella typhi* using high throughput virtual screening, and to understand the inhibition mechanism between ligand and protein. This research looks into the high throughput virtual screening of D-alanine-D-alanine ligase A against the ZINC database for potential inhibitors. D-alanine-D-alanine ligase A is a vital enzyme in the peptidoglycan biosynthesis pathway, and being a non-human homolog, D-alanine-D-alanine ligase A is a good target for inhibition With a series of analytical processes like post-screening analysis, protein-inhibitor interactions via pharmacophore mapping, and ADME/Tox properties validation, the analysis reveals that N-[3-(1H-imidazol-1-yl) propyl]-4-(2-methylbenzoyl)-1H-pyrrole-2-carboxamide (ZINC04025098) is a potential inhibitor for D-alanine-D-alanine ligase A.

Chapter One

1.0 Introduction

Typhoid fever is a major health issue around the world. Contraction of typhoid fever is due to the ingestion of food and/or water source contaminated by *Salmonella typhi.* Annually, an estimated 17 million people are infected with cases of typhoid, with 600,000 fatalities (Bhan *et al.*, 2005).

Manifestation of typhoid fever includes symptoms like fever, headache, and abdominal pain during the first week. Gastrointestinal symptoms like anorexia, nausea, vomiting, and constipation with or without any specific signs. The common signs like relative bradycardia, splenomegaly, hepatomegaly, and abdominal tenderness may develop in the second week of illness. Other complications like acute abdomen, pneumonia, intestinal perforation, psychosis, ataxia, altered sensorium and nephritis are likely to develop in third to fourth week of illness. Other complications are difficulty in constipating, diarrhea, enlarged spleen, secondary infection leading to meningitis, and probably depression. Initial presentation as acute glomerulonephritis is a very rare manifestation of typhoid fever (Sinha *et al.*, 1999).

In general, typhoid is an endemic occurring among poor and developing countries due to poor sanitation system and inadequate water supply. In Asia, typhoid estimated incidence is around 274 per 100 thousand populations yearly, which unfortunately, is the highest, compared to other continents (Kothari *et al.*, 2008).

Here in Malaysia, typhoid is a notifiable disease since 1986 under the Communicable Disease Control Act (CDC Act 1988) ("Laws of Malaysia. Act 342 Prevention and Control of Infectious Diseases Act 1988."). From the year 1997 to 2007, based on the report given by the Ministry of Health Malaysia ("Malaysia Health 2007. Ministry of Health Malaysia"), the total number of cases increase from 701 to 1072, an alarming increase in numbers. Between 2001 and 2007, the incidence rates were between 1.89 and 4.10 per hundred thousand in a population. There were some endemic areas and occasionally epidemics especially in the states of Kelantan and Sabah (Shah *et al.*, 2012).

The main prevention of typhoid is by improving hygiene and sanitation. Vaccine for the high risk group and population are recommended as well as during the outbreak control (World Health Organization 2008). The surveillance system for the control and prevention of typhoid is still a huge challenge for local health bodies due to the wide range of factors associated with typhoid. There are uncertainties also regarding to the distribution of these factors, as well as the magnitude and impact on the incidence of typhoid. The association of the environmental factors and the distribution of typhoid within the various geographical and spatial factors is one of the major gray areas of the typhoid surveillance system. Geographical information system (GIS) is a very effective tool used for tracking spatial pattern of typhoid distribution. GIS provides maps and plots of cases, and also provides the spatial combination of data which includes the socioeconomic, distribution of disease, and geographical data such as landscapes, water sources like rivers and lakes, and weathers (Shah *et al.*, 2012).

A study by (Safian *et al.*, 2008) in a district in Malaysia has shown the existence of typhoid clusters between 2001 and 2005. Various factors have been shown to be related to typhoid incidence. Poor hygiene and sanitation, unsafe food and water have been associated with many studies like ; (Sharma *et al.*, 2009), (Bhan *et al.*, 2005), and (Tran *et al.*, 2005) regarding the factors directly related to the environment of people, thus contributing to the clustering of typhoid.

1.1 Problem Statement

`Chloramphenicol, a type of bacteriostatic antimicrobial or antibiotic, was introduced to the treatments against *Salmonella enterica serovar Typhi (S.typhi)* in 1948; the drug was very successful at both clinically and economically at that time against typhoid. However, in 1950, two years after the introduction of chloramphenicol, sporadic resistance towards the drug was observed, but it was not until 1972 that the first outbreak of Chloramphenicol-resistance *Salmonella typhi* was described. Ever since then, drug-resistance of *Salmonella typhi* has become a clinical issue, however this time possessing resistance to multiple drugs (Mirza *et al.*, 1996).

After the deposition of Chloramphenicol, a new antibiotic called Ciprofloxacin became the main treatment against *Salmonella typhi,* for being a better drug compared to Chloramphenicol. Yet, ever since the discovery of *Salmonella typhi* being resilient to Ciprofloxacin in the typhoid fever patients, the alternate choice of antibiotics left is a drug known as Ceftriaxone (also known as Cefexime). However, resilience against Ceftriaxone was reported to the Centre for Drug Control (CDC). Side effects have been reported for Ceftriaxone (Threlfall *et al.*, 1992).

Multiple-drug resistance *Salmonella typhi* poses a huge threat especially to treatments on-going in typhoid endemic states in Malaysia. The sudden course of this event urges for new researches and discoveries for new potential inhibitors for *Salmonella typhi*, and in hopes of these discoveries, a new, effective, and economical

drug against typhoid can be at the disposable to combat against *Salmonella typhi* (Phipps *et al.*, 1991).

Now with the accessibility of large amount of genomic and proteomic data obtained from human genome project and via the *Salmonella typhi* genome, these discoveries revolutionize studies and researches of novel drug-design for *Salmonella typhi*. These genomic and proteomic data are main resources for the search of novel drug targets or inhibitors in *Salmonella typhi.* Strategy using subtractive genomic studies is applied to *Salmonella typhi* by assuming that the targets identified in *Salmonella typhi's* metabolic pathways and mechanisms is essential for *Salmonella typhi's* survival, which should be part of the replication, survival, and at the same time should be absent on humans and doesn't posess any homologs in humans, therefore when an experimental ligand is designed to inhibit possible proteins in *Salmonella typhi* should only target the mechanism and functionality of *Salmonella typhi* and not against humans (Bhawna *et al.*, 2009). Therefore any proteins of interest must be screened for any sequence similarity with proteins found inside humans. This ensures targeted inhibition against *Salmonella typhi's* metabolic pathways and not humans.

1.2 Objectives

The objectives of this research are summarized as below:

1) To identify potential drug targeted protein from *Salmonella typhi*.

2) To identify a possible inhibitor for the targeted protein by virtual screening.

3) To understand the protein-inhibitor interactions.

Chapter Two

2.0 Literature Review

2.1 *Salmonella typhi*

Salmonella typhi is a group of flagellated, rod-shaped, gram-negative, facultative anaerobe bacterium that belongs to the *Enterobacteriaceae* family. *Enterobactericiae* are pathogenic in nature and is the cause of enteric infections. They are named after the scientist Dr. Daniel Salmon who is the first to isolate *Salmonella choleraesuis* from a pig's intestine. (Pui *et al.*, 2011). There are two major *Salmonella* species; *Salmonella enterica* and *Salmonella bongori (McClelland et al., 2001)*. *Salmonella enterica* subsp. *enterica serovar Typhi* is a human-specific organism that give rise to typhoid fever (Pui *et al.*, 2011).

According to the World Health Organization (WHO) Collaborating Centre, Centers for Disease Control and Prevention (CDC) and other organizations, *Salmonella enterica* is divided into six major sub-categories of species, whereby they are defined by using roman numeric numbers. *Salmonella enterica* subspecies I is mainly found in mammals and is responsible for around 99% of isolates from clinical samples while other subspecies and *Salmonella bongori* are mainly found in cold-blooded animals and is accounted for as little as 1% of isolates from clinical samples (Pui *et al.*, 2011). There are 2 major pathogenicity genes (pathogenicity islands) that encode the virulence factors of *Salmonella spp*. That allows the bacteria to invade and infect their host organisms. There are two important pathogenicity genes:

Salmonella pathogenic gene 1 (pathogenicity island 1) and pathogenic gene 2 (pathogenicity island 2). These pathogenicity islands encode for two different types of effector molecules from the type III secretion systems into the host cell, allowing the bacteria to intrude the host cell which then leads to the systemic spreading of the bacteria in the host organism. *Salmonella enterica* subsp. *enterica serovar Typhi* possess multiple fimbriae operons that create extracellular appendages allowing them to attach and enter their hosts' intestinal epithelial cells (Srikanth *et al.*, 2011).

Salmonella typhi is capable of infecting a wide range of hosts. *Salmonella typhi* causes enteric fever via the fecal-oral route, through the consumption of *Salmonella typhi* contaminated food and water. The two main clinical syndromes of *Salmonella* infection are: non-typhoidal salmonellosi, a disease called enteritis which affects the gastrointestinal system; and typhoid fever (Layton *et al.*, 2007). Depending on the species of *Salmonella* and the specification of host, the nature and severity of diseases varies. There are more than 2500 serotypes of *Salmonella*, and most reports of diseases include complications of symptoms like nausea, vomiting, diarrhea and cramps at the abdominal (Layton *et al.*, 2007).

However, in some immune-compromised individuals a severe bacterial infection can result in life-threatening septicemia. Severe typhoid fever is caused by *Salmonella enterica serovar Typhi* (*Salmonella typhi)*, Mortality rates in untreated patients is between 10 to 15%. Some patients recuperates but becomes carrier of *Salmonella typhi*. (Layton *et al.*, 2007).

2.2 Typhoid Fever

Typhoid fever is a gastrointestinal infection caused by *Salmonella typhi. Salmonella typhi,* is a highly evolved pathogen that evolved somewhere around fifty thousand years ago (Kidgell *et al.*, 2002), possess mechanisms allowing its' resilience in its human host (Merrell *et al.*, 2004).

Although contaminated food and water is the major cause for typhoid fever, other factors also play important roles in causing typhoid fever in different endemic settings like poor sanitation , close contact with typhoid cases or carriers, low level of education, large household size, close proximity to water bodies, floods , bad personal hygiene, poor life style, and travelling to endemic areas. Weather variables such as, rainfall, vapor pressure and temperature also affect the transmission and distribution of typhoid infections in human populations (Hornick *et al.*, 1970).

2.3 Global Burden of Typhoid (Epidemiology)

An estimate of the global burden of typhoid fever shows there are probably 16 million fresh cases yearly with 600,000 reported of deaths (Geneva, 1996). This data was first shown in year of 1984 with similar results from various sources were published (Edelman *et al.*, 1986).

A research done in 2004 concluded that typhoid fever affected an estimate of 21,650,974 victims and resulted in 216,510 of deaths in the 2000, and the paratyphoid fever infected 5,412,744 victims (John *et al.*, 2004). This estimation was concluded on positive blood culture acquired from studies on 22 population. Still, this conclusion was only based on result on a few countries, and only 3 studies provided data for the African continent.

Typhoid is an endemic in many developing countries, whereas it is sporadic in most developed, industrialized nations with supply of clean drinking water and better sanitation (John *et al.*, 2004). The reported estimated victims of typhoid fever infection in Africa is 50 in 100,000 people and in Asia is as much as 274 in 100,000 people. This conclude that even with the multiple environment conditions in different parts of Asian and African continent, *Salmonella typhi* possess the ability to survive and thrive in these diverse conditions (John *et al.*, 2004) .

2.4 Pathogenesis

Salmonella typhi, in contrast to other *Salmonella* species, the only reservoir for the bacteria are human beings, thus the pathogenicity of *Salmonella typhi* is only specific towards humans, which results in typhoid or enteric fever (Zhang *et al.*, 2008). *Salmonella typhi* can survive for long period (several months) in both soil and water.

Typhoid fever is transmitted by *Salmonella typhi*-contaminated water and food in endemic areas, and especially by *Salmonella typhi* carriers handling food in developing countries.

Salmonella typhi employs a stealth approach which allows the invasion of the human body's inner tissues and thus prevents the onset of an immediate inflammatory response of the human host in the human gut. (Merrell *et al.*, 2004).

Salmonella typhi is pathogenic, as the bacteria has multiple of pathogenic factors (Giannella, 1996). These factors are:

1) Ability of cell invasion.
2) A full lipopolysaccharide coating.
3) Ability for intracellular replication.
4) The possibility toxin(s) elaboration.

There is a probability that *Salmonella typhi* attacks the terminal ileum's gut mucosa via a specialized antigen-forming cell called M-cell, through enterocytes, or via a para-cellular route (Kops *et al.*, 1996). The bacteria attaches to the mucus of the intestine in the internal ileum using the cystic fibrosis transmembrane conductance regulator protein (Lyczak *et al.*, 2001). Induction of intestinal epithelial cells by increasing the level of membrane receptor is an important step in the initial phase of the infectious process, which thus enhances the ingestion and sub-mucosal translocation of the bacteria. (Lyczak *et al.*, 2002).

Salmonella typhi infection causes the influx of peripheral white blood cells into the lamina propria. The process is facilitated by secretion of cytokine from the epithelial cells triggered by the *Salmonella typhi's* lipopolysaccharide. Lipopolysaccharide coating triggers the transcription factors of lymphocytes by triggering a toll-like receptor 4 complex in the mammalian toll pathway (Chen *et al.*, 1999, Vazquez-Torres. *et al.*, 2004).

The invading *Salmonella typhi* are engulfed by macrophages, which in turn causes a downstream salmonella- induced caspase-1 mediated apoptosis (Parkhill *et al.*, 2001). Firstly, *Salmonella typhi* attaches to the lymphoid tissues of the intestine, and then they flow inside the mesenteric nodes, the thoracic duct, and finally the blood circulation. *Salmonella typhi* will reach the major organs involve in the reticulo-endothelial system like the bone-marrow and liver within 24 hours from their ingestion, where they will thrive and replicate via mono-cytic lineage (Vazquez-Torres. *et al.*, 2004). The late stage of the disease (8–14 days after incubation) is marked when *Salmonella typhi* are shed back into the blood circulation.

2.4.1 Mechanisms of *Salmonella typhi* evasion of innate immunity

Salmonella typhi utilizes several strategies that to avoid the detection of the host innate immune (Quynth *et al.*, 2010). Successful sequencing of the *Salmonella typhi* genome shows a unique region in the *Salmonella typhi* genome, the *Salmonella*

pathogenicity island 7 (SPI-7) genes. This SPI-7 genes codes the proteins responsible for producing and exporting the Vi-capsular polysaccharide antigen.

The Vi-antigen is responsible for the stealth ability of *Salmonella typhi* by shielding the bacteria from being detected by the human immune response, probably by shielding the exposure of the lipopolysaccharide layer. Furthermore the TviA, a SPI-7-encoded regulatory protein responsible for the expression of the Vi-antigen, flagellar motility, and the type 3 secretion system (T3SS) responsible for invasion on *Salmonella* Pathogenicity Island 1, is important for the precise timing of expressing the virulence factor in the tract of the gastrointestinal system. These findings concludes that TviA-facilitated inhibition of expression of flagellar helps *Salmonella typhi* to avoid being detected.

2.5 Antimicrobial Resistance

Even though the most potent antibiotics are easily accessible globally, however the usage of antimicrobial drugs in some developing countries is confined to those who can afford them. The first generation of antibiotics are chloramphenicol, ampicillin, co-amoxiclav, erythromycin, co-trimoxazole, gentamicin, tetracycline, and penicillin. Drugs like amikacin, cefuroxime, ciprofloxacin, and nalidixic acid are used as secondary antibiotics (Hart *et al.*, 1998).

In the late period of 1987, a typhoid fever outbreak broke out in China. The causing isolates of *Salmonella typhi* are resilient to all the first line of antimicrobials. (Fu *et al.*, 1989). In between 1989 and 1990, there were reported cases of emergence of similar *Salmonella typhi* isolates from Pakistan, the Arabian Gulf, and India. These strains of multidrug resistant *Salmonella typhi* is now reported throughout the world, including Malaysia (Hart *et al.*, 1998).

Multiple antimicrobial resistance of *Salmonella tyhi* is mediated via pHCM1 plasmid. The plasmid possess 99% similarity to the sequence of plasmid R27, a type of H1 plasmid. An estimated of 74% of the strains from Indo-China, Vietnam shows that the outbreak of multidrug-resistant typhoid fever was caused by the cloning of *Salmonella typhi* strains possessing resistance plasmid (Bhan *et al.*, 2005).

Chloramphenicol-resistant *Salmonella typhi* was reported in the United Kingdom within just two years of the satisfactory implementation of chloramphenicol against *Salmonella typhi* (Ryan *et al.*, 1989). Subsequently, isolates of *Salmonella typhi* possessing chloramphenicol resistance genes that are transferable were reported from Israel (Sompolinski *et al.*, 1967) and Greece (Kontomichalou 1967). In 1967, two strains of *Salmonella typhi* isolated from Cairo and Aden possess resistance genes of tetracycline, chloramphenicol, and ampicillin that are transferable. (Anderson *et al.*, 1972). However, it was not until 1973 that the epidemics of *Salmonella typhi* possessing antimicrobial resistant genes were reported in Mexico (Olarte. *et al.*, 1973) and India (Paniker. *et al.*, 1972). Other epidemics also occurred in Korea (Chun. *et al.*, 1977), Peru (Yi *et al.*, 1981), and Vietnam (Butler *et al.*, 1973). Many of these

resistant isolates carries resistance genes of sulphonamides, chloramphenicol, aminoglycosides, and tetracycline. Then, in the year 1981, soon right after the discovery of co-trimoxazole, resistance genes of trimethoprim and chloramphenicol that are transferable was found in various isolation of *Salmonella typhi* (Datta *et al.*, 1981).

First ever report of multi-drug resistant strains was discovered in 1987 in South-East Asia (Ling *et al.*, 1984) and have since dispersed throughout the South-East Asia. The prominence of these multi-drug resistant strains in China has then increased since 1985 and by the year 1989, 80% of the *Salmonella typhi* isolates found in Shanghai possess multiple drug resilience capability. The resilience towards ampicillin, chloramphenicol, trimethoprim, tetracycline, cephazolin, sulphonamides, and gentamicin was an encoded 98 MDa plasmid which is transferable (Zhang, 1991). In Pakistan on the year of 1987, cases of mullti-drug resistant strains were reported (Smego *et al.*, 1988) and have since then increased to almost all of the *Salmonella typhi* isolated (Mirza *et al.*, 1993).

Between 1990 and 1991, all of the 25 multi-drug resistant strains isolates obtained from Rawalpindi, Pakistan was found to possess a transferable *c*. 98 MDa plasmid that encodes the resistance against ampicillin, chloramphenicol, trimethoprim, tetracycline, sulphamethoxazole and streptomycin, but doesn't possess the resistance gene towards gentamicin (Mirza *et al.*, 1993) and therefore, different from the *Salmonella typhi* isolated in China. By the year 1994, in Quetta, a place located in Northern Pakistan, 77% of *Salmonella. typhi* isolates from human blood possess

multi-drug resistant genes and they have represented more than 50% of positive blood cultures (Mirza *et al.*, 1995).

In 1990, *Salmonella typhi* possessing multiple-drug resistance were reported in India (Jesudasan *et al.*, 1990) and several isolates possess incH1 plasmids were also reported (Threlfall *et al.*, 1992). Multi-drug resistant isolates have also been reported in Malaysia (Phipps *et al.*, 1991), Bangladesh (Saha *et al.*, 1994) and Vietnam (Tran *et al.*, 2005), and the epidemic zone in Asia ranges from Pakistan in the western zone to China in the eastern zone. Apart from these endemic zones, there is a 'partial-epidemic' area in the Eastern-Central Asia. Somewhere around 35% of the population of the Persian Gulf states are migrant workers, mostly from the subcontinent of India and the Eastern Indo-China countries like Myanmar. These workers migrate regularly from their mother land and thus importing 70-80% of Multi-drug resistant of strains *Salmonella typhi* to Bahrain, Kuwait (Wallace *et al.*, 1990), Qatar (Uwaydah *et al.*, 1991) and Oman (Elshaffie *et al.*, 1992). In all these regions between 5 to 30% of the *Salmonella typhi* isolates possess multi-drug resistant genes.

2.6 Peptidoglycan

Peptidoglycan is an important and unique component of the bacterial cell wall, found on the external of all gram-positive and gram-negative bacteria's cytoplasmic membrane (Rogers *et al.*, 1980). The main function of peptidoglycan is to preserve the integrity of the cell by withstanding turgor pressure. Inhibition against the peptidoglycan biosynthesis via any means like mutation, antibiotic targeting or by cell lysis, degradation during cell division will result in lysis of the cell. It also plays a major role in the sustaining of a well-structured cell and it functions as a docking port for the anchorage other cell enveloping components like various signaling proteins (Dramsi *et al.*, 2008) and teichoic acids (Francis *et al.*, 2003). Peptidoglycan also plays a major role in the growth and dividing process of cells.

2.6.1 Chemical structure of peptidoglycan

The structural characteristics of peptidoglycan are the cross-linked linear glycan layers (Rogers *et al.*, 1980). The glycan layers are consist of separate cross-linked N-acetylglucosamine and N-acetylmuramic acid residues linked by beta-1, 4 bonds. Each MurNAc residues' D-lactoyl group is substituted by a peptide stem that composed of L-Alanine-sigma-D-Glu-meso- A_2pm (or L-Lys)-D- Alanine -D-Alanine (A_2pm, 2,6-diaminopimelic acid) in nascent peptidoglycan, the last D-Alanine residue being lost in the finished linked polypeptide. Cross-linkage of the glycan strands reaction initiates in between the carboxyl group of D-Alanine at the

4th position and the amino group of the di-amino acid at 3rd position, via a peptide bridge linkage. Thus, the physiochemical characteristics of the hetero-polymer include multiple bonding between sugar (MurNAc), sigma-bonded D-Glu, of L–D (and even D–D) bonds and of non-protein amino acids.

2.6.2 Biosynthesis of peptidoglycan

The peptidoglycan biosynthesis pathway is a complicated biochemical process involving multiple steps in the cytoplasm and membranes of the bacteria (Heijenoort, 2001). The complex process involves 20 biochemical reactions includes the biosynthesis of precursors of nucleotide, and the biosynthesis of lipid-linked polymerization reactions and intermediates. These steps occurs at the internal and external side of the cytoplasm.

The cytoplasmic steps can be divided into four sets of reactions:

1) Biosynthesis of UDP-GlcNAc from fructose-6-phosphate.
2) Biosynthesis of UDP-MurNAc from UDP-GlcNAc.
3) Combination of the peptide stem leading to UDP-MurNAc-pentapeptide.
4) 'Side' or 'annex' path- ways of synthesis of D-glutamic acid and the dipeptide D-alanyl-D-alanine (Heijenoort, 2001).

Initial stage of peptidoglycan biosynthesis is the biosynthesis of the soluble precursors of nucleotides, by converting UDP- GlcNAc to UDP-MurNAc-pentapeptide. This processes of biochemical reaction synthesize peptide mediety is catalyzed by the MurC, MurD, MurE and MurF ligases. These ligases are responsible for the insertion of , meso-diaminopimelic acid, D-glutamic acid, L-lysine, and L-alanine to UDP-MurNAc (Barreteau. *et al.*, 2008).

The defined form of this peptide and the active sites' uniqueness of these enzymes shows diversities in the bacterial world. Initial steps at the membrane phase begin with the transfer of the phospho-MurNAc- penta-peptide mediety from the cytoplasmic precursor to the membrane receptor undecaprenyl phosphate (C55-P), a process mediated by the transferase MraY possessing a pentapeptide called undecaprenyl-pyrophosphoryl-Mur- NAc (lipid1) (Barreteau. *et al.*, 2008).

Then, the transferase MurG facilitates the transport of the GlcNAc mediety from UDP- GlcNAc to lipid I bonded with undecaprenyl-pyrophosphoryl- MurNAc-(pentapeptide)-GlcNAc (lipid II) which, after it travels pass the membrane, it becomes the substrate for the polymerization reactions (Heijenoort, 2001). The C55-P carrier lipid is essential in these cascading process and the C55-P carrier lipid is also involves in the biosynthesis of other cell-wall related components (Bouhss. *et al.*, 2008) .

2.6.3 Cross Linking

A major step in peptidoglycan layer assembly and deposition is the subsequent enzymatic cross-linking of one peptidoglycan strand to an adjacent another peptidoglycan strand. Although direct cross- in Gram-negative bacteria such as *Salmonella typhi* and *Escherichia. coli* are made between C,-e-NH, of a meso-DAP residue and the penultimate -D-Alanine in a second penta-peptide strand, in Gram-positive organisms such as *Staphylococcus aureus* a pentaglycine bridge connects the strands (Christopher, 1989). Either way the enzymatic cross-linking reactions are trans-peptidations. The new interstrand peptide bond is then joined as the intrastrand D-Alanine-D-Alanine bond is then broken (Christopher, 1989).

Not all possible inter-peptide cross-linkages are produced within the peptidoglycan layers or between the arrays of peptidoglycan layers, but it is obvious that a substantial degree of cross-linking provides mechanical strength and rigidity to the peptidoglycan layer (Christopher, 1989).

D-Alanine is therefore a key molecule in the assembly and cross linking of peptidoglycan. Three enzymes are involved in conversion of L- alanine to D-alanine and its attachment to the UDP-rnuramyl pentapeptide (Christopher, 1989). These enzymes are: ATP-dependent D-Ala-D-Ala adding enzyme, Pyridoxal phosphate-dependent alanine racemase, and ATP-dependent D-Ala-D-Ala ligase.

2.6.4 D-Alanine-D-Alanine Ligase A

The enzyme D-alanine-D-alanine ligase A (E.C. 6.3.2.4) plays an essential role in biosynthesis of bacterial cell wall by building peptide cross-linkages that inadvertently provides the spot for trans-acylation for the cross-liking of the peptide scaffold (Ellsworth *et al.*, 1996). Therefore it is essential to evaluate D-alanine-D-alanine ligase A for possible inhibition, as their inhibition will disrupt the biosynthesis of bacterial cell wall, due to their major role in peptidoglycan cross-linking.

2.7 Virtual screening

Drugs that have been created for the past 100 years are known to interact with approximately 500 targeted proteins, while 20,000,000 organic compounds are synthesized. Also, the genomic and functional genomic projects have produced more than 1500 possible drug targets against human diseases (Andrew *et al.*, 2002). Thus, there are no doubts there are possible drugs (leads or hits), are hidden away in the many database.

However, the process in search of these possible leads is very challenging. Compiling these compounds and randomly screened them against various proteins one by one is very impractical, for it consumes a lot of time, requires lots of computing power and thus financially not viable. Yet, virtual screening seems to be

the definitive solution, which leads to the development of virtual screening, and finally implementing virtual screening into the drug discovery pipeline (Thompson *et al.*, 2002).

Virtual screening is designed to search large-scale databases of chemical structures or virtual libraries by using multiple computational analysis for selecting a limited number of candidate molecules likely to be inhibitive against a chosen biological target (Campbell, 2007). Thus, virtual screening is a database searching based on the logic expansion of 3D pharmacophore (Ekins *et al.*, 2009) and molecular docking (Andrew *et al.*, 2006), allowing automated evaluation of huge databases.

The two methods that are implemented are applying 3D pharmacophore-based database searching to locate potential leads from databases and applying molecular docking approach to evaluate databases if solved 3D structures of the targets are found. Generally, the above are sequential methods, as the first is fast at filtering out compounds and the latter has more accurate evaluation of the ligand-receptor bindings. Regarding screening based on pharmacophores, a 3D-pharmacophore feature is constructed by protein-ligand interaction analysis on a set of active compounds (Yasuhisa *et al.*, 2001) or deduction based on the X-ray crystallography solved structure of a ligand-receptor complex (Gerhard *et al.*, 2005). By querying using 3D–pharmacophore mappings, searching through 3D database can be performed via selection of molecules from available chemical databases, which contain pharmacophoric data and follows the pharmacophore geometric restrictions (Honglin *et al.*, 2008).

Virtual screening based on docking algorithms uses the structural data from both receptors and ligands. The concept of docking-based virtual screening consist of 5 strategies (Honglin *et al.*, 2008):

1) 3D modeling of receptor
2) Generation of compound database.
3) Screening via high processing computers.
4) Post processing of possible leads
5) Experimentation of leads via bioassay.

The main strategy of virtual screening is docking and scoring. Docking is a compute intensive algorithm that fits 3D compounds into the binding site of a receptor. The process optimizes the possible orientations and conformations for a ligand interacting with a protein, then select compounds that bind to the protein's binding pocket in the lowest energy conformation. Algorithm of optimizations are used widely by molecular docking program and these algorithms are categorized into three main groups:

1) Numerical optimization algorithms.
2) Random/stochastic algorithms.
3) Hybrid optimization algorithms.

Importantly, these are common functions used for scoring docking results (Böhm *et al.*, 2003):

1) Scoring functions based on force-field energy.
2) Scoring functions based on empirical calculations.
3) Scoring functions based on knowledge.

Sadly none of these scoring function can perform perfectly as different system calculate differently for physical phenomenon regarding molecular recognition, like solvent effects and entropy (Kitchen *et al.*, 2004).

After Kuntz and his colleagues (Kuntz *et al.*, 1982) release to the public the first docking algorithm, designated as DOCK in the year 1982, more than 20 docking algorithm have been created since then. However, there are still many restriction and obstacles for molecular docking, such as the accuracy of binding conformation and affinity prediction, protein flexibility, entropy, and solvent effects. Molecular docking is still a complicated and intriguing topic of bioinformatics (Moitessier *et al.*, 2009).