CHAPTER I

INTRODUCTION AND REVIEW OF LITERATURE

1.1 INTRODUCTION

*Helicobacter pylori* (previously known as *Campylobacter pylori*) is about 3 µm long and 0.5 µm in diameter, microaerophilic (requires oxygen), neutrophilic (adapting to highly acidic environment) helix-shaped (curved rod), gram negative bacteria colonizing upper gastrointestinal tract (Figure 1.1). Conservative thinking was that no bacterium can live in the human stomach as its pH is acidic in nature (~ 1). Spiral shaped bacteria in the lining of human stomach were initiated by a group of German scientists in 1875 (Blaser, 2005; Bizzozero, 1893). Akin, Bizzozero detected similar bacteria in the stomach of dogs. Professor Walery Jaworski investigated and proposed the possibility of being an organism involved in gastric diseases (Konturek, 2003). Several studies over a period of time concluded that microbes or stress or spicy food are responsible for ulcers and gastric diseases. Warren and Marshall (1984) identified, visualized, and cultured *H. pylori* in *invitro* conditions and declared that stomach ulcers and gastritis are mostly due to *H. pylori*, and not due to stress and spicy food. Dr. Barry Marshall administered himself *H. pylori* culture through an oral route to demonstrate the connection between gastritis and *H. pylori*. After 10 days, Dr. Barry Marshall revealed the signs of gastritis by endoscopy suggesting that *H. pylori* is the causative agent of gastritis. Marshall and Warren (1984) also suggested that antibiotics can be employed for effective treatment of gastritis, for which they were awarded Noble Prize in Medicine or Physiology in the year 2005.

![Micrograph of *Helicobacter pylori*](image)

Figure 1.1: Micrograph of *Helicobacter pylori*
1.1.1 *H. pylori* Colonization is Beneficial or Undesirable?

Blaser (2005), hypothesized that colonization of *H. pylori* in stomach is beneficial though its connection with gastritis was well established (Blaser and Atherton, 2004). Blaser (2005) also proposed *H. pylori* to be ‘a member of the normal flora of the stomach’. Colonization of *H. pylori* in stomach influences systemic immune responses and reduces acid reflux disease (Blaser, 2005), asthma (Salama et al., 2013), dermatitis (Salama, 2013), esophageal cancer (Blaser 2005; Salama et al., 2013; Blaser, 2011), gastroesophageal reflux disease (Blaser 2005; Blaser 2011), inflammatory bowel disease (Salama et al., 2013), obesity (Blaser 2006; Blaser et al., 2009), rhinitis (Salama et al., 2013) and type II diabetis (Blaser 2006; Blaser et al., 2009). Chen and Blaser, (2005) asserted Blaser,s hypothesis that 'H. pylori is a member of the normal flora of the stomach' by demonstrating the association between *H. pylori* colonization and lower incidence of childhood asthma. Whatever, may be the consequences of the colonization of *H. pylori* in stomach a positive effect by reducing diseases or undesirable effect by causing gastritis; association between *H. pylori* and host is a fact.

1.1.2 Evolutionary Relationship Between Pathogen *H. pylori* and Host Human

*H. pylori* and its human host association was estimated to be around 60,000 years ago (Linz et al., 2007). Association between *H. pylori* and its human host was first established in the birth place of modern human in Africa. Correa and Piazuelo (2012) by phylogenetic analysis established the following lineages of *H. pylori* that diverged during the course of evolution – Africa 1 (includes isolates from West and South Africa), Africa 2 (includes isolates from South Africa), NE Africa (includes isolates from North East Africa), Amerind (includes isolates from Native Americans), Asia 2 (includes isolates from Bangladesh, Malaysia, North India and Thailand), East Asia (includes isolates from East Asians), Europe (includes isolates from Europe, Middle East, India), Maori (includes isolates from Taiwanese Aboriginals, Melanesians and Polnesians), and Sahul (includes isolates from Australian Aboriginals, and Papua New Guineans). Ancestors for the above lineages are ancestral East Asia, ancestral Africa 1, ancestral Africa 2, ancestral Europe 1, ancestral Europe 2 and ancestral Sahul (Correa and Piazuelo, 2012).
1.1.3 Outcome of *H. pylori* Infection

Diversity among the *H. pylori* isolates established the fact that world’s human population is infected with *H. pylori* and is the most widespread infection in the world. Living standards of the country, age of the host can be the factors that influence the possible outcome of the infection. Infection rate is high in underdeveloped countries, developing countries and developed countries (Pounder and Ng, 1995). The infection rates in different countries varies due to socioeconomic factors, hygiene standards, and widespread of antibiotics (Smoak et al., 1994; Everhart et al., 2000). Outcome of the infection is based on age at which host is infected with the bacterium. High risk of gastric ulcer and cancer was observed in individuals who were infected at an early age than individuals who were infected at a later age (Kusters et al., 2006). Statistics revealed that gastric cancer is the second most common cancer in worldwide.

1.1.4 Detection of *H. pylori* and its Outcome

*H. pylori* is tested if peptic ulcers disease, MALT lymphoma and dyspepsia were reported (Stenström et al., 2008). Several methods that are in use to test the existence or association of *H. pylori* with gastrointestinal disorders are endoscopy, biopsy, histological examination, microbial culturing and recently capsule scopy (Stenström et al., 2008). *H. pylori* is cultured it can be visualized by Gram stain, acridine-orange stain, Giemsa stain, haematoxylin–eosin stain, and Warthin-Starry silver stain. *H. pylori* can also be visualized using phase-contrast microscope. The other tests that are used are rapid urease, ELISA, blood antibody, and stool antigen tests (Logan and Walker, 2001).

1.1.5 Treatment for *H. pylori*

Proton pumps, H2 antagonists, and antacids were initially used in treatment against *H. pylori* (Rauws and Tytgat, 1990; Graham et al., 1991). Standard first line therapy to treat *H. pylori* consists of triple therapy proton pump inhibitor omeprazole, antibiotic clarithromycin and amoxicillin. Alternative proton pump inhibitors are pantoprazole and rabeprazole; and alternate antibiotic for clarithromycin is levofloxacin and for antibiotic amoxicillin is metronidazole (Perna et al., 2007; Hsu et al., 2008; Malferttheiner et al., 2012). When initial therapy failed due to antibiotic resistance, alternative strategy such as quadruple therapy is implemented in the form bismuth
colloid which includes bismuth subsalicylate (Fischbach et al., 2007; Graham et al., 2008). Rising antibiotic resistance has driven the research to identify new drug targets (Neelapu and Pavani, 2013; Neelapu et al., 2013) and drugs (Neelapu et al., 2013). Alternative therapeutic strategy such as ‘routes of immunization’ to provide immune protection in the form of vaccination to control *H. pylori* is also in trial (Selgrad et al., 2008). New vaccines or drugs can be developed if we can better understand the pathogenesis of *H. pylori* at different stages such as adaptation of *H. pylori* to the acidic environment in the stomach, protection from oxidative stress induced by host, adhesion of the pathogen to the stomach, gastritis, ulcer, and gastric cancer.

### 1.2 Pathogenesis of the *Helicobacter pylori*

*H. pylori* has a major role in gastric carcinoma, gastric adenocarcinoma, and gastric ulcer, gastritis and MALT lymphoma. *H. pylori* is contagious, where the person-to-person transmission is either by the fecal–oral or oral-oral route. *H. pylori* infection and pathogenicity is well connected and coordinated by different mechanisms to adapt to the acidic environment in the stomach; to attack the stomach wall by weakening the gastric mucosal barrier; using different molecules to adhere the host stomach lining; protecting itself from oxidative stress; inducing gastritis or inflammation, formation of ulcers, and inducing gastric cancer.

#### 1.2.1 Mechanisms of Adaptation to the Acidic Environment in the Stomach

The neutralophilic motile bacterium *H. pylori* has developed several strategies to minimize its exposure to low pH so that it can flourish well in the gastric environment of the host. *H. pylori* migrates from anum to fundus to colonize in the host. The pH in the stomach varies widely, in the absence of food (pH is~1.0) and in the presence of food (pH is~5.0). *H. pylori* has the adeptness to sense chemicals in the surrounding environment, and pH is an important cue. Bacteria recognize and act in response to these chemical signals using chemoreceptors. Four chemoreceptors TlpA, TlpB, TlpC, and TlpD and several polar flagella are reported in *H. pylori*. Chemoreceptors allow the bacteria to swim in case of attractant and stop in case of repellent (Lowenthal et al., 2009; Rader et al., 2011; Goers Sweeney et al., 2012). Goers Sweeney et al (2012) established the structure of chemoreceptor TlpB’s and proposed the mechanism by which *H. pylori* sense pH. TlpB is a member of MCP superfamily of
transmembrane receptors, containing two transmembrane helices with an extracellular sensing Per-ARNT-Sim (PAS) domain popularly known as universal signalling fold. Active site of this domain contains a molecule of urea. This domain along with urea is responsible for detecting ligands directly or indirectly via interactions with periplasmic proteins. Transmembrane helices on the cytoplasmic side are helical with a histidine kinase adenyl cyclase, methyl binding protein, phosphatase (HAMP) domain. At neutral pH, Asp 114 in the active site is negatively charged and accepts two hydrogen bonds from the amide nitrogen atoms of cofactor urea, thereby binding and stabilising the fold of PAS domain. The state of the periplasmic domain is relayed through the transmembrane region affecting the phosphorylation of CheA, inturn controlling the downstream components of flagellar motor and its activity to guide stopping behaviour. Whereas, in the case of high pH Asp 114 in the active site is highly protonated by weakening or disrupting binding of cofactor urea. Signal is relayed and guides flagellar motor to moving behaviour.

Once the pH is sensed by the chemoreceptors, *H. pylori* has two mechanisms to adapt to the acidic environment. The first mechanism of *H. pylori* is to change the pH of its surroundings and the second mechanism is to move towards the less acidic region. The pH of its surrounding environment is changed by urease and is coded by urease gene cluster ure A, B, I, E, F, G, and H (Ferrero and Labigne, 1993). Urease structural subunits are coded by ure A and B; followed by assemble of inactive urease structural subunits into active urease by incorporating nickel using ure E, F, G, and H (Park and Hausinger, 1995). Whereas, ure I encodes a pH gated urea channel to increase the access of urea to intrabacterial urease in low pH conditions (Rektorschek et al., 2000). Intrabacterial urease activity generates carbon dioxide and ammonia buffering the cytoplasm and periplasm of the organism (Wen et al., 2003). Urease deletion mutants were not able colonize the stomach and failed to survive in the acidic environment (Mollenhauer-Rektorschek et al., 2002). Followed by activation of mechanisms to sense and change the pH of its surroundings, *H. pylori* tends to move towards the less acidic region. Stomach lining composes of mucous layer, followed by epithelial cells and connective tissue respectively. Mucous layer is towards the lumen interface and epithelial cells and connective tissue are underneath the mucous layer. pH at the mucous lumen interface and at the mucous epithelial cells is ~2 and ~5 to 6 respectively. So, *H. pylori* uses its flagella to tunnel into the mucous and move
towards the less acidic region. Thus, bacterium is able to move from the acidic pH at mucous lumen interface to the neutral pH at the mucous epithelial cells interface. It can be concluded that *H. pylori* is migrating from the stomach lumen via mucous layer to the epithelial cells to colonize as the pH of the epithelial cells is neutral.

### 1.2.2 Attacking the Stomach Wall – Gastric Mucosal Barrier

Colonization of *H. pylori* is different in different individuals. Based on the acidity of the stomach bacterium can colonize mucous, or epithelium inner surface, or inside the epithelial cells or at pyloric antrum or at fundus or whole lining of the stomach or rest of the stomach. Colonization along the whole lining of the stomach, or at pyloric antrum, or at fundus can be due to normal or reduced amounts of acid secretion; large amounts of acid secretion; and to avoid acid secreting parietal acids respectively. Colonization of *H. pylori* along the lining of stomach or in the mucous or epithelial cells is possible only when gastric mucosal barrier is weakened. Gastric mucosal barrier protects gastric mucosa from a variety of damaging agent’s such as gastric acid, pepsin, refluxed bile and pancreatic juice, certain foods, range of temperatures, hyperosmolar and abrasive substances, bacterial toxins and damaging drugs. Gastric mucosal barrier is a collection of anatomical, physical and chemical processes that protect the gastric mucosa. The eight components of the gastric mucosal barrier are (1) tight junctions of the epithelial cells (2) restitution a process which gastric epithelial cells change shape (3) secretion of mucosal bicarbonate (HCO$_3^-$) (4) hydrophobic nature of the apical membrane of the gastric epithelial cells (5) balance of local acid-base and gastric mucous blood flow (6) production and secretion of gastric mucosa (7) regulation and protective effect of mucosal prostaglandins and (8) basal lam. The role of *H. pylori* in damaging gastric mucosal barrier is not yet known, there might be some unknown mechanism contributing to weakening of gastric mucosal barrier that helps in colonizing bacterium in the mucous (Burrows, 2010). *H. pylori* may weaken the gastric mucosal barrier via toxin VacA, cytokines, gastrin release probably by loosening the protective mucous layer, disruption of mucous layer, and alterations in mucous glycoproteins respectively.
1.2.3 Adhesion of the Pathogen to the Stomach

*H. pylori* adheres to the epithelial cells with the help of adhesins molecules such as BabA and SabA. Adhesins bind to carbohydrates and lipids of epithelial cell membrane (Petersen and Krogfelt, 2003). BabA and SabA also binds to Lewis b antigen and sialyl-Lewis X antigen expressed on epithelial cells and gastric mucosa respectively (Ilver et al., 1998; Mahdavi et al., 2002). In addition to these proteins adherence is assisted by a group of other proteins such as AlpA, AlpB, HopZ, and OipA is helpful in adherence of *H. pylori* (Moodley et al., 2009). Gastritis, ulcer formation and gastric cancer are three events which run parallel.

1.2.4 Inflammation or Gastritis Induced by *H. pylori*

Several mechanisms are proposed to describe the pathogenicity of *H. pylori* change in expression of host genes, infection-induced cell proliferation, loss of polarity and elongation of cell, cell-cell junctions degradation, decrease in acid secretion (Yamaoka, 2010) and inflammation. Cag PAI (Cytotoxin associated gene pathogenicity island) with a size of 40 Kbp contains 27 genes encoding for T4SS pilus is responsible for pathogenicity and inflammation. Cag PAI encodes for genes CagA, VacA, 11 VirB proteins (VirB1- VirB11) and coupling protein (VirD4). Core components or putative channel is formed by Vir B6-B10; pilus associated components is formed by VirB2, VirB3, and VirB5; VirB4 and VirB11 are the energetic components. VirB1 is a muraminidase enzyme which lyse murien at a particular location to establish T4SS pilus assembly. CagY (VirB5 orthologue), Cag I, L, Y, and VirB proteins form the appendage of pilus to secrete CagA, VacA and peptidoglycan into the host (Tegtmeyer et al., 2011).

CagA when injected into the host epithelial cells, it is phosphorylated and activated by src kinase and host proteins to modify cellular responses (Hatakeyama, 2008). Cell focal adhesions are disrupted by CagA by binding and activating SHP2 phosphates. Normal epithelial architecture is disrupted when polarity regulator PAR1h/MARK2 kinase is inhibited by CagA leading to loss of polarity in epithelial cells (Hatakeyama, 2008). Another surface receptor protein in *H. pylori* Toll like receptor (TLR)-2 disrupts adherin junctions within gastric epithelial cells. This activates protease calpain cleaving E-cadherin and allows increased β-catenin signalling to disrupt adherin junctions (O'Connor et al., 2011).
*H. pylori* infection leads to inflammation at the site of infection by inducing proinflammatory cytokines. Peptidoglycan of pathogen enters the host cell and stimulates intracellular pathogen receptor Nod1, to signal and activate NF-κβ and AP-1 inducing cytokines such as interleukins. During inflammation the inflammatory molecules interleukins (IL)-1, IL-6, IL-8, TNF-α, and RANTES are upregulated in the host stomach. Apart from cytokines, chemokines like granulocyte-macrophage colony stimulating factor (GM-CSF), cyclooxygenases (COX-2), Reactive Oxygen/Nitrogen Species (RONS), *H. pylori* neutrophil activating protein (HP-NAP) are activated and stimulated. Beales and Calam, (1997) found that infection of *H. pylori* stimulates GM-CSF when tested on human gastric epithelial cell lines during infection. This leads to activation of AP-1 to induce COX-2 and nitric acid synthase (Cho et al., 2010). COX-2 induces Prostaglandin synthesis (PGE) – 2. *H. pylori* nitric acid synthase (Cho et al., 2010), LPS (Cavallo et al., 2010), arginases and host arginase II (Gobert et al., 2011) produce, upregulate, and modulate nitric oxide species respectively. Vacuolating cytotoxin (Vac) A (D’Elios et al., 2007) and HP-NAP (Huang et al., 2011) induce ROS damage in mitochondrial DNA of gastric epithelial cells. Vac A interact with a number of host surface receptors to trigger responses such as pore formation, endolysomal functions modification, cell vacuolation, immune inhibition and apoptosis (Peek and Crabtree, 2006; Amieva and El-Omar, 2008; Atherton and Blaser, 2009). *H. pylori* neutrophil activating protein (HP-NAP) is secreted and passes through the epithelial cell to reach lamina propria to induce ROS from neutrophils and monocytes and also to stimulate production of other chemokines such as CXCL8, CCL3, and CCL4 to attract or recruit other leucocytes by ROS (D’Elios et al., 2007; Evans et al., 1995; Polenghi et al., 2007). Virulence factors such as γ-glutamyl transpeptidase, VacA and cholesterol α-glucosides modulated responses of T cells.

### 1.2.5 Protection from Oxidative Stress

Pathogenesis of *H. pylori* is dependent on its ability to survive in the vulnerable oxidative stress environment apart from other factors such as acidity, peristalsis and phagocytosis (Olczak et al., 2000). Oxidative stress induced by *H. pylori* when it colonizes in host is lethal leading to DNA damage in the genome of *H. pylori* (O’Rourke et al., 2003). Many pathogens including *H. pylori* have acquired the capability to survive DNA damage induced by oxidative stress by transformation mediated recombination DNA repair for successful infection of the pathogen (Michod...
et al., 2008). While many pathogens are competent for transformation only in certain environmental conditions such as starvation, *H. pylori* is competent throughout the growth (Dorer et al., 2010). *H. pylori* is exposed to double stranded DNA damage and natural transformation DNA damage has increased with DNA damage (Dorer et al., 2010). Mutants of RecA, Rec N, RuvC, and AddAB were unable to colonize in the human host with increased sensitivity to DNA damaging agents and oxidative stress (Dorer et al., 2010; Loughlin et al., 2003; Wang and Maier et al., 2008).

### 1.2.6 Formation of ulcers

When bacteria colonize the stomach, inflammation induces G cells of the antrum. G cells secrete hormone gastrin, which travels to parietal cells of the fundus via blood stream (Blaser and Atherton, 2004). Gastrin stimulates secretion of the acid from the parietal cells and also increases the number of parietal cells. Increased load of acid damages epithelial cells of the duodenum resulting in ulcers (Schubert and Peura, 2008).

### 1.2.7 Gastric Cancer

Inflammatory cells produce cytokines such as metalloproteinases (MMPs), prostaglandin E2 (PGE2) and RONS, which in turn augment and prolong the inflammatory cascade (cytokines induce PGE2 and MMPs induce RONS). These inflammatory mediators disregulate DNA repair enzymes thereby leading to microsatellite instability (MSI). Inflammatory mediators also lead to defective mitotic checkpoints, induce directly or indirectly double stranded breaks (DSB) and deregulate HR pathway of DSB repair leading to chromosomal instability (CI). MSI and CI induce genetic diversification randomly leading to activation of oncogenes and inactivating tumor suppressor genes (Colotta et al., 2009).

LPS, peptidoglycan, and CagA of *H. pylori* activate transcription factor NF-κβ which is essential to activate innate and adaptive immune responses against pathogens. Sustained and constitutive expression of NF-κβ results in chronic inflammation and cancer (Hayden and Ghosh, 2008). NF-κβ signalling needs to be turned off properly to avoid prolonged and detrimental inflammatory responses. Cell utilizes many mechanisms at multiple levels for termination of NF-κβ signalling. 1kβα synthesis and cyclandromatosis (CYLD) expression mechanisms are lost or overpowered in cancer by downregulation of NF-κβ signalling. Other mechanism by which NF-κβ is turned
off is by direct ubiquitination and degradation of NF-κβ (Yang et al., 2009). CagA and COX-2 were known for cell proliferation, prostaglandin biosynthesis and angiogenesis. Cag A binds to E-cadherin interfering β-catenin regulation, transdifferentiation of numerous cell lineages and increased cell proliferation (Murata-Kamiya et al., 2007). Mitogen inducible cycogenases-2 (COX-2) were reported to induce prostaglandin biosynthesis and angiogenesis in human gastric cancer tissues. COX-2 overexpression also correlated with metastatic involvement of the lymph nodes (Uefuji et al., 2000). COX-2 expression correlated with microvessel density by Huang et al, (2011) studied VacA induced Reactive Oxygen Species (ROS) damage in mitochondrial DNA of gastric epithelial cells.

Though the information on activation of oncogenes is not wellstudied, several papers reported on inactivation of tumor suppressor genes after the *H. pylori* infection. Inactivation of the tumor suppressor genes - RUNX3, p53, ASPP2, TFF1, TFF2, TFF3, GKN1 and GKN2 promoted gastric carcinogenesis. Cag A induces proteasome mediated degradation that inactivates gastric tumor suppressor gene RUNX3 (Tsang et al., 2010; Tsang et al., 2011). Hypermethylation of promoter of RUNX3 also leads to gene silencing (Tsang et al., 2010; Kitajima et al., 2008). Cag A also induces proteasome-mediated degradation of p53 to modulate and ASPP2 tumor suppressor genes (Buti et al., 2011). Trefoil factor family proteins (TFF1, 2, 3) regulate mucosal repair and suppresses tumor formation in stomach. Deficiency of tumor suppressor genes TFF1 and genetic ablation of gene TFF2 revealed the protective role of TIFF proteins and their reduction in promoting gastric carcinogenesis. Another protein belonging to gastrokine family GKN1 and GKN2 also have a protective role in gastric mucosa and downregulated in gastric cancer (Toback et al., 2003; Walsh-Reitz et al., 2005).

**1.2.8 *H. pylori* Induced Apoptosis**

VacA toxin of *H. pylori*, MHC-II, Fas-FasL, transcription regulator NF-κβ of host were known for probably inducing apoptosis. *H. pylori* infection mediates insertion of VacA toxin into the mitochondrial membrane to induce and release cytochrome-C thereby activating Caspase 3 dependent cell-death signalling cascade which in turn activates mechanism of apoptosis (Jones et al., 1997; Rudi et al., 1998; Lancellotti et al., 2006). Pathogen *H. pylori* binds or interacts to/with MHC-II on the surface of
gastric epithelial cells to induce apoptosis (Fan et al., 2000) *H. pylori* induce expression of Fas (cell-surface receptor) and FasL (Fas Ligand) and stimulate apoptosis (Caulfield et al., 2014). Gastric epithelial cells respond to *H. pylori* by activating NF-κβ regulating inflammatory cascade (chemokines, iNOS) leading to apoptosis (Tak and Firestein, 2011).

The above detailed discussion on pathogenesis of *H. pylori* bacteria and its role in causing dreadful cancer demonstrated that some proteins of pathogen are important for its survival in the host. These important proteins can be the potential drug targets and designing a drug to these targets may inhibit the growth of the pathogen in host with possible therapeutic intervention of gastric cancer.

1.3 DISCOVERY OF DRUG TARGETS

A key biomolecule, protein or nucleic acid involved/associated with a metabolic pathway/signaling pathway leading to pathology/specific disease condition is known as drug target. Sometimes pathology/disease condition can be due to infection which is important for survival of the pathogen.

1.3.1 Properties of Drug Targets

Some of the important properties of a good target are

1. Target is disease modifying and/or has a proven function in the pathophysiology of a disease (essential for the survival of the cell, mostly targeting biological processes such as peptidoglycan biosynthesis, cell wall synthesis, transcription and translation etc.).
2. Target should be non-homologue and should be specific in expression.
3. Target is favoured to be localized in the cytoplasm, ribosomal proteins, less frequent in membranes and more likely to be enzymes like transferases, hydrolases or ligase, but not lyases.
4. Target has a favorable assayability enabling high throughput screening
5. Druggability (implying the presence of a binding site) should exist for the drug targets either for small molecular weight chemical compound (SMOL) or Biologics (BIOL) binding domain in the assessment. If the druggability does
not exist, a 3-D structure or a close homologue for the target protein should be available for a druggability assessment.

6. A target/biomarker should exist to monitor therapeutic efficacy.

7. Favorable prediction of potential side effects according phenotypic data.

8. Must be present in a number of pathogens if broad-spectrum action is required. Narrow spectrum antibiotics target specific pathogens.

9. Target should have favorable IP situation (freedom to operate (FTO), no competitions on target, options for commercialization).

Based on the above properties drug targets are selected or identified, evaluated for druggability, assayability and further validated experimentally. Novel drug targets with clinical efficacy as shown in in vitro and in vivo models are necessary for success in drug discovery process (Gashaw et al., 2012).

1.3.2 Paradigm for Drug Target Discovery and Validation

Paradigm for drug target discovery was proposed by Neelapu et al., (2013). Integrated approaches have been used by many groups based on their requirement. New strategies like crowd sourcing, innovation drivers in the literature, drug – drug target network and present strategies like in vivo, in vitro and in silico approaches can be used in a combination to develop an integrated approach for new paradigm to drug target discovery and validation for infectious diseases (Figure 1.2).

Innovation drivers in the literature identify various universities, organizations, private organizations and groups working on the diseases. Further, innovation drivers in literature also provide information on the R&D carried out till date on this disease summarizing the events related to the disease. Literature reviewed can be summarized in analytical perspective using drug - drug target networks for the drugs and drug targets for a particular disease, which may reveal more prospects of disease, and these can pave a way to more specific drugs for disease.

Crowd sourcing can be used as a platform to connect the organizations like funding agencies, universities, venture capitalists, Pharma industries, Contract Research Organizations(CRO’s) and groups working on diseases for a more collaborative approach for drug target discovery and validation. Funding agencies use this as a
platform to create a science push, or a business pull or a model right from research to patenting and release of the product into the market bridging the gap the between academics and industries. Once the platform is developed, this platform can be used to harness the potential of drug target discovery and validation groups. *Insilico* analysis of the data (Genome analysis, proteome analysis, metablome analysis, microarray data analysis, reverse docking, grid technology, *insilico* knockouts) *invitro* and *invivo* methods (Haploinsufficiency profiling, Chemogenomics, Chemical proteomics, Signature tag mutagenesis, gene knockout technology, expression profiling methods) can be used to identify essential genes of the organism either individually or in group with coordination. Once the drug targets are identified, identified drug targets are passed through a confirmatory test. Genomes (genes pool), proteomes (proteins pool), metabolomes (metabolites pool) data of the above identified molecules will be subjected *in silico* analysis by comparing with that of host to get non homologue humans. Further drug targets can be assessed by *in silico* methods for the drug targets properties using BTX, SRT, VICM and protein – protein interactions servers. Finally *in silico* knockouts and Gene knockout of the molecule identified and confirmed can be used for drug target validation. This integrated approach can be used to drug target discovery and validation.

1.3.3 Current Strategies for Drug Target Discovery

Novel drug targets can be identified through different sophisticated technologies, including *in vitro*, *in vivo*, and *in silico*. Haploinsufficiency profiling, Chemogenomics, Chemical proteomics, Signature tag mutagenesis, gene knockout technology and expression profiling are the *in vitro*, *in vivo* methods for drug target identification. Genome analysis, proteome analysis, metaanalysis of microarray data analysis, metabolome analysis, reverse docking, grid technology, and *in silico* knockouts are few *in silico* methods used for drug target identification.
Figure 1.2: New paradigm for drug target discovery and validation (Neelapu et al., 2016)
1.3.3.1 *In vitro* and *In vivo* Methods

1.3.3.1.1 Haploinsufficiency Profiling

Haploinsufficiency profiling in yeast and chemogenomics are the genome-wide genetic approaches used for drug target identification. Haploinsufficiency profiling is genomewide assay which reveals drug targets required for cell viability. This technique lowers the levels of the protein target by half (i.e. by deleting one of the two copies of the corresponding gene in diploid yeast). A pooled collection of heterozygous *S. cerevisiae* strains with 6000 genes is replaced with a selectable marker bearing two unique sequences, or molecular barcodes. These strains are grown in as a single log phase culture to identify hypersensitive strains. The fitness of each deletion is then quantitatively measured by genomic DNA extraction and hybridization to high-density oligonucleotide arrays carrying probes complementary to the barcodes (Chan et al., 2010).

1.3.3.1.2 Chemogenomics Approach

A chemogenomics approach integrates two genomic assays HIP and overexpression. A pool of heterozygous single gene deletion yeast strains and a relative gene abundant transformants are grown in the absence and presence of a drug of interest. Barcode sequences marking the gene deletions and relative gene abundant transformants from the pool are amplified by PCR. Genes that confer sensitivity when deleted but those confer resistance when overexpressed represent candidate drug targets (Chan et al., 2010).

1.3.3.1.3 Chemical Proteomics

A chemical proteomics combines stable isotope protein labeling with affinity purification and mass spectrometry for drug target identification. Two sets of cells are grown in media to incorporate isotopes into the expressed proteins, one set in light isotopes growth medium and another in heavy stable isotopes. Soluble protein extract from cells which are harvested from the heavy labeled medium are incubated with bead-immobilized drug; whereas both the drug beads and soluble free drug competitor are incubated with protein lysate extracted from cells grown in light medium. The beads are then washed, combined, and bound proteins are eluted, and proteolytically digested. The resulting peptides are then analyzed by high performance mass
spectrometry. Evaluation of the differential ratio of heavy/light signal intensities recorded for peptide (denoted by peaks with distinct mass-to-charge ratios), specific drug interactors (i.e. exhibiting highly differential H/L ratios) and nonspecific protein binders (those with H/L ratios close to unity) are readily used for discrimination Mass spectrometry-based screening is combined with Stable Isotope Labeling with Amino acids in Culture (SILAC) to quantitatively detect differences in target retention versus nonspecific absorption to an immobilized drug ligand (Chan et al., 2010).

1.3.3.1.4 Signature Tag Mutagenesis

Signature tag mutagenesis was effectively used for gastrointestinal pathogens (Escherichia coli, Listeria monocytogenes, Helicobacter pylori, Vibrio cholerae and Salmonella enterica) which are a global health concern (Cummins et al., 2012). Pathogenic bacteria have molecular mechanisms to invade and establish infection within the host with genes directly or indirectly associated with infection process. Inactivation of these genes may lead to complete loss or decrease of virulence. Signature tagged mutagenesis is a genetic approach which is based on gene disruption strategies. This functional genetic analysis method uses random transposon mutagenesis to disrupt gene and simultaneously analyze mutant strains that carry a single gene deletion. Signature tagged mutagenesis initially transforms DNA tags into the transposon plasmid and is used to create a bank of individually tagged mutants. Differently tagged mutants are pooled into input and output pools and they are used to detect non-colonizing mutants. Unique DNA tags with restriction sites are created for insertion into plasmid, and removal of invariant arms. These tags are incorporated into transposons plasmids creating individually tagged mutants that are organized into 96 well plates for experiments. Differently tagged mutants are pooled in the input pool (IP) and grown overnight; these are then used to infect the animal model of interest (mouse). Mutants that survive the infection are recovered from the organs of interest and used as the output pool (OP). Detection of IP and OP pools can be performed by different methods with original technique being dot blots. The DNA tags are used to create a hybridization probe for IP and OP and the results are compared to detect non-colonizing mutants by PCR. The DNA tags in plasmid are amplified by a primer and a tag-specific primer. Once again the IP and OP are compared to identify non-colonizing mutants.
1.3.3.1.5 Gene Knock-out Technology

Gene knock-out technology (Balasubramanian et al., 1996; Pelicic et al., 1997) is an experimental technique by which the expression of one or more genes of the organisms are reduced or made inoperative. The reduction is made either through genetic modification or by treatment with a short DNA /RNA oligonucleotide that is complementary to gene/mRNA transcript.

1.3.3.1.6 Expression Profiling

Transcription profiling and protein expression profiling are the recent developments in microarray technology (Schena et al., 1995) to identify drug targets. Understanding pathology of the pathogen in dormancy phases (in in vivo condition) is difficult. DNA microarrays analysis of the entire pathogen have been widely used to study the gene expression under the different simulated conditions using both in vitro and in vivo models. Metaanalysis of the microarray datasets can be used to predict and identify potential drug targets. Microarray has been used to identify M. tuberculosis genes that are induced by drugs (Wilson et al., 1999; Betts et al., 2003), genes that are inducing or controlling dormancy conditions (Sherman et al., 2001; Voskuil et al., 2003).

1.3.3.2 In silico Methods

1.3.3.2.1 Genome Analysis

Availability of genome sequences provides platform for genome analysis that can be useful in drug target identification. Literature survey revealed four different analyses that can be performed: they are comparison between host and pathogen genomes; genomes of strains; genomes of species; and genomes of avirulent and virulent strains. Genome analysis of the pathogens and host has provided a tremendous amount of information that can be useful in drug target identification. One of the recently adopted strategies is subtractive genomics approach, where dataset of the pathogen is subtracted from the host. This analysis provided information for a set of genes that are likely to be essential for pathogen but are absent in host (Neelapu and Pavani, 2013). Comparison of genomes between set of strains would reveal strain specific genes of a species which can be a source of drug targets. Comparison of genomes between set of virulent and avirulent strains would reveal virulence specific genes of a pathogen. This information can be harnessed for identifying potential drug targets of a pathogen.
Comparison of genomes between set of species belonging to the same genera would reveal species specific genes which are important for survival of the species which can also be a potential source of drug targets.

### 1.3.3.2.2 Proteome Analysis

Proteome complements are available for proteome analysis and they form the basis to identify drug targets of the pathogens. Proteins mined and compared with proteomes of host predict the druggability of proteins essential for pathogens. These druggable proteins having the structural information for ligands (drugs) can be the potential drug targets, thus opening the route not only for classical biochemical studies but also for structure based discovery approaches. This tractable list of prioritized proteins contributes to badly needed drug discovery programs for prevalent human disease (Neelapu and Pavani, 2013).

### 1.3.3.2.3 Metabolome Analysis

Pathways, metabolic networks and metabolic choke points (Crowther et al., 2010; EI-Sayed et al., 2005; Kushwaha et al., 2010) form the basis for metabolomics analysis providing information on molecules which are important for survival of pathogens. Focusing on the metabolic causes of pathogenicity and virulence leads to discovery of unique therapeutic targets dissimilar to human metabolism. Pathway analysis compares metabolic pathways between host and pathogen revealing pathogen specific pathways. This information provides new insights on growth which can be used for drug target discovery. Metabolic choke points or load points identify unique or a particular metabolite node that can be biochemical lethality of metabolic networks. The local and global aspects of metabolic network analyses allow us to identify enzymes or reactions that are crucial for the survival of the organism(s). The choke point strategy was implemented on the pathogen bacterial network to identify potential drug targets.

Comparing the metabolic network of pathogen with that of humans can aid in identifying unique features of pathogen metabolism. Ultimately the metabolic reconstruction of the pathogen provides a framework for the interrogation of human pathogens and as a platform for integration of high throughput data and generation of experimental hypothesis. These types of novel network analysis may be relevant for
treating diseases, particularly when diseases are considered emerging and uncontrolled; necessitating improvements in therapeutics and treatment strategies.

1.3.3.2.4 Metaanalysis of Microarray Data

DNA microarrays are used for understanding biological processes based on transcriptional profiling and epigenetic experiments gene expression data. A microarray database is a repository containing microarray gene expression data. Microarray database is used to store the measurement data; manage a searchable index; and make the data available to other applications for analysis and interpretation. Gene Expression Omnibus (GEO) from National Center for Biotechnology Information (NCBI) and ArrayExpress from European Bioinformatics Institute (EBI) (Kolesnikov et al., 2015) are the microarray databases available online. Clustering methods such as hierarchical and K-means are used for the analysis of microarray gene expression data. Genes that are expressed for these conditions can be of invasion or surface attachment of the pathogen/ parasite to the host cell. This information can be of use to identify potential drug targets.

1.3.3.2.5 Reverse Docking

Reverse docking is opposite to the process of docking, a compound with certain biological activities is docked with the binding sites of 3-D structure of a protein (Cai et al., 2006). This method is efficient and cost effective. Target Fishing Dock (TarFisDock) is a reverse docking application that is used to identify potential drug targets (Li et al., 2006).

1.3.3.2.6 In silico knockouts

Constraint-based metabolic flux simulation is a powerful tool to predict the effect of metabolic reaction. It is important to design metabolic networks suitable for constraint-based metabolic flux. Using constraint-based metabolic simulation, in silico screening was performed to identify metabolic changes caused by possible knockouts/ gene disruption of a gene candidate. Such knockout strategies can improve potential target identification in an organism (Fatumo et al., 2009).
1.3.3.2.7 Grid Technology for Drug Target Discovery

Grid technology presents an architectural framework that aims to provide access to heterogeneous resources in a secure, reliable and scalable manner. Grid projects are created of a distributed and ubiquitous grid environment for supporting target identification. Gene, protein, genomes, proteomes, gene transcription profiling (transcriptomics), protein expression profiling (proteomics), metabolic pathway analysis (metabolomics), protein glycosylation (glycosilomics), protein-protein interactions (interactomics) and systems biology are the heterogeneous data available on grid environment (Jian et al., 2006). The grid data manager integrates heterogeneous distribution data in the grid environment to discover and interact with data. Grid-enabled web-portal interface is designed from which the user can remotely plan, submit and monitor the high-performance compute (Foster et al., 2002). The Wide In silico Docking on Malaria (WISDOM) project was the first project in the public domain that used Grid enabled in silico docking to simulate the interaction of potential drugs with target proteins. Using the Grid techniques, the project was able to screen over 35 million compounds from existing chemical libraries against an enzyme target on the smallpox virus, deploying the 3D ligand flexible docking approach. These projects demonstrate the tremendous potential that Grid computing holds for speeding the efforts of pharmaceutical R&D. Cellular and physiological process knowledge of disease is vital for drug target identification. Protein interaction maps can reveal novel pathways and functional complexes, allowing “guilt by association” annotation of uncharacterized proteins, once the pathways are mapped, the potential drug target may be located in a biochemical pathway (Rain et al., 2001).

1.3.4 New Strategies for Drug Target Discovery

Crowd sourcing, drug target network and innovation drivers in the literature are the new strategies used for drug target discovery

1.3.4.1 Crowd Sourcing

Crowd sourcing is an open source drug discovery offering new and inexpensive drugs to combat diseases that affect the poor. It is based on principles of collaboration and open access. By opening a project to external contributors, research capacity on neglected diseases increased significantly. CSIR Team India Consortium’s Open
Source Drug Discovery project (CSIR OSDD) and The Synaptic Leap’s Schistosomiasis project (TSLS) attracted contributors internationally. Both crowd sourcing projects enabled high quality research at low cost. Crowd sourcing thus overcome the problems faced by the diseases of the poor due to meager investments (Ardal et al., 2012).

1.3.4.2 Drug–Target Networks

A drug–target network was one of the strategies used to harness novel drug targets. Most of the approved drugs were developed without the knowledge of molecular mechanism responsible for disease. Proteins rarely operate individually; rather they act as a highly interconnected network in a system. Drug - target network, assessment was carried out in both prospective and retrospective network based relationships to quantify the relationships between drug targets and gene products (Yildirim et al., 2007). Drug – target network analysis by various groups revealed multiple targets to single drugs. Shift in single to multiple targets paradigm in drug discovery paved a way to drug- target network analysis (Yildirim et al., 2007; Medina-Franco et al., 2013). Updated single/multiple target – drug paradigm describe the reality of cellular processes on a greater network upstream and downstream of the actual drug target for the future attempts of drug effect, paving a way to more specific drugs for diseases.

1.3.4.3 Innovation Drivers in Literature

Innovation drivers in literature are used to identify drug target. Agarwal and Searls (2009) identified innovation drivers in literature using bibliometric methods as additional approaches to enhance decision-making in drug target identification and validation. Increased levels of scientific activity; indicated by publication rates; points pull and push of a particular scientific opportunity tends to drive purely favorable considerate commercial innovation. PubMed, Medline, ProQuest, are the literature retrieving tools for field of biological sciences. PubMed and MEDLINE® comprise of abstracts and citations for biomedical literature from around the world. ProQuest offers a full-text biomedical literature. Scientific publications clearly provided information on the research drivers of disease; rates of change of each disease; change in market share of the disease; and characterization of the additional features of disease. Articles with specific disease; genes; gene/genome annotation; and molecular basis of a disease can be the innovation drivers in literature for discovery of potential
drug targets. Medical Subject Headings (MeSH) terms can be used to retrieve the above information from the literature. Patents search tools facilitate search of patents which may provide unique solutions to the problems we may be facing in drug discovery and development. United States Patent and Trademark Office, European Patent Office, Google Patent Search, Ipindiaservices.gov.in are some of the patent searching tools that can be used to drive development of the products for the understanding disease; drug and vaccine development; developing diagnostic kits for infectious diseases etc. Therefore, bibliometric method is a fruitful area that can be used in drug target identification and validation of drug discovery and development process.

1.4 DRUG DESIGNING AND VALIDATION

Drug designing is also known as rational drug design and can be used for invention of drug based on biological target. The designing of drug based on computer modelling techniques is called computer aided drug designing again (CADD). CADD is of two types

1. Structure based drug designing (SBDD)
2. Ligand based drug designing (LBDD)

Structure based drug designing (SBDD) is also known as direct drug design and is based on 3D- structure of the molecular target. Ligand based drug designing (LBDD) is also known as indirect drug design is based on small molecule binding to the targets with high binding affinity. CADD involves the following steps. They are

1. hit identification
2. hit-to-lead optimization
3. lead optimization

1.4.1 Hit Identification

Virtual screening of small molecule libraries to identify new ligands based on structure of ligand or on biological structure of ligand is hit identification. The databases/ libraries that can be used for screening of small molecule are drug bank, therapeutic target database (TTD), and Chembl.
1.4.1.1 Drug bank

Drug bank is a wonderful knowledge resource on drugs that can be used for screening of drugs to the drug targets using both bioinformatics and chemo informatics approaches. Drug bank is a unique database with nearly 8070 drugs, of which 1836 are FDA approved small molecules. Protein or peptide drugs which are FDA approved are nearly 193, of which 92 are nutraceuticals. In addition it also contains 600 experimental drugs and 4317 drug targets like enzymes, transporters, carriers. etc. (Wishart et al., 2006).

Drug bank presents quantitative, pharmaceutical, physical, molecular and biological data related to thousands of drug targets and drugs which is necessary for drug development and discovery. Each drug/drug target entry gives information on the target with drugs which are approved or experimental proved along with their pharmacological action, taxonomy, ADMET properties and the interacting targets both in human or other organisms etc.

1.4.1.2 Therapeutic Target Database (TTD)

TTD is a another peer reviewed database which provides information about a protein or nucleic acid target and drugs; and their related information like function, sequence similarity etc. This database currently contains 2,589 targets (397 successful, 723 clinical trial, and 1,469 research targets), and 31,614 drugs (2,071 approved, 9,528 clinical trial, 17,803 experimental drugs). Small molecules are nearly 20278 and 653 anti-sense drugs are also available with sequence and structures. TTD is integrated with BLASTp which helps in virtual screening of data sets based on sequence similarity search for target based drug discovery (Feng et al., 2010).

1.4.1.3 Chembl

Chembl is another valuable bioactivity database of bioactive drug molecules and targets with huge information regarding properties, Lipinski parameters, compounds, assays, drugs and approved drugs. It also contains a large number of drug like bioactive compounds which are curated and abstracted from literature to power drug discovery by using this information. Currently it contains 5.4 million bioactivity measurements with more than one million compounds and 5200 drug targets (Gaulton et al., 2012).
1.4.2 Hit to lead optimization

Initial screening of hits by conformation and evaluation of analogues followed by analogues synthesis is called hit expansion. Lead is nothing but a chemical compound with a pharmacological or biological activity on a validated target for a therapeutically use.

1.4.3 Optimization of lead molecule

The promising lead identified was subjected to ADMET properties (absorption, distribution, metabolism, excretion, toxicity properties). Based on the chemical structure lead is modified with improved properties of both metabolic and physicochemical properties. These then can be tested and best compound is chosen as a drug candidate for pre-clinical drug development.

1.5 SCOPE OF PRESENT INVESTIGATION

From the detailed survey of literature, *H. pylori* infection causes gastritis, ulcers and stomach cancer. The stomach cancer is more in less developed countries than developed countries. According to the statistics in India stomach cancer is the second most common cancer. Rate of incidence in India is higher in south and north east of the country (Sharma and Radhakrishnan, 2011). The major problem with treatment is side effects. And success rate of treatment is dependent on stage of cancer. Hence, the objectives of the current study are to identify to drug targets and drugs.

1.6 OBJECTIVES OF THE STUDY

The following objectives are to be investigated to identify drug targets and drugs.

a. Identification of drug targets for *H. pylori* using subtractive genome analysis, metabolic pathway analysis, pathogenic island analysis, and metaanalysis of micro array data.

b. Identification of ligands for *H. pylori* using *in silico* approach.