In-silico analysis of host-pathogen interaction: a case study with molecular mimicry phenomenon

Chapter 1 Introduction

1.1. Introduction

The host defense system is a highly complex and advanced mechanism that protects the host and eliminates infectious organisms and other intruders. It mainly helps us to avoid any type of infectious microbes, such as bacteria, virus and fungus, and it will clear them out from the body. The complete process is referred to as the immune response of the host. The immune process involves different types of cells (and their protein products), tissues and organs of the body in a highly coordinated manner.

The most important cell types which are involved in the immune process are white blood cells (WBC) also called leukocytes or leucocytes, which basically protect the body against both infectious disease and foreign invaders. Leukocytes are produced and derived from the hematopoietic stem cells in the bone marrow, found throughout the body and lymphatic system. These cells circulate via lymphatic and blood vessels between the lymphatic organs and nodes. WBC are classified in standard ways as follows: (a) on the basis of structure it can be classified as granulocytes and agranulocytes, (b) leukocytes are distinguished by cell lineage (myeloid cells or lymphoid cells). Secondly, these cells were classified on the basis of physical and functional characteristics as follows; neutrophils, eosinophils, basophils, lymphocytes, and monocytes. Furthermore, lymphocytes can be classified into B-cells, T-cells, and natural killer (NK) cells. All these cells work together to monitor the body for foreign invasion, and through blood and lymphatic vessels, they circulate between the lymphatic nodes and different organs of the body.

NK-cells are the predominant innate lymphocyte subsets that mediate antitumor and antiviral responses. B-cells are mainly involved in adaptive immunity, which produces and matures into bone marrow. Although, T-cells are produced in the bone marrow and premature cells migrate into the thymus gland for maturation, and leading to the production of mature cytotoxic and helper T cells reveals the selection process that are significant for distinguishing self from nonself. The T-cells selection criteria

is quite rigorous; ~98% of the thymocytes, the precursors of T cells, die before the completion of the maturation process. Although thymic selection is remarkably effective in suppressing the immune response to self-antigens, failures do occur. Such failures result in autoimmune diseases.

Autoimmunity is a condition in which the immune system is instigated against self. A long heterogeneous list of diseases that range from mild to life-threatening conditions are reported to be caused by the malfunction of the immune system (Laxminarayana 2017). Autoimmune disease can be caused by many factors. It can be environmental or genetic or external (Cooper et al. 1999, Jörg et al. 2016). Sometimes infectious agents may also trigger the process of autoimmunity (Kivity et al. 2009, Sfriso et al. 2010, Wucherpfennig 2001). This phenomenon is popularly known as molecular mimicry (Albert & Inman 1999, Rojas et al. 2018). Microbial pathogen-induced molecular mimicry mostly occurs when antigenic determinant on one of the microbe's proteins is similar to a determinant on one of the proteins made by the host. This results in activation of autoreactive T or B cells, which ultimately leads to autoimmunity. Though the host possesses a check mechanism to prevent elicitation of immune response against self in form of major histocompatibility complex (MHC). Hence, to generate an immune response against self, the difference between foreign and self-peptide should be strong enough, so that the self should be recognized and discriminated against by the immune system. This cross-reactivity between host and pathogen's epitopes eventually leads to autoimmunity that may have deleterious effects onto the host (Rojas et al. 2018). The disease caused by the rogue immune response is termed an "autoimmune disease".

1.2. Aims and objectives

A number of studies have deciphered various prospects and aspects of molecular mimicry, but these are scattered in numerous research papers. Compilation of the available information from literature can greatly facilitate the researchers who work in this domain. At the start of my PhD work we couldn't find any unified information resource that has a collation of all available information related to autoimmune diseases caused due to molecular mimicry. A database namely mimicDB was present that provides information about proteins or epitopes involved in host-pathogen interactions (Ludin et al. 2011). But mimicDB was restricted to information

pertaining to only a few human parasites. Also, the mimicry candidates of mimicDB were predicted through a computational pipeline (Ludin et al. 2011). Therefore, I started my work with the establishment of a freely accessible database named as miPepBase **(Mi**micry **Pep**tide Data**base),** with the aim to provide a comprehensive and high quality resource of epitopes involved in molecular mimicry. All molecular mimicry based autoimmunity events of miPepBase were collected from peer-reviewed publications (Garg et al. 2017).

Molecular mimicry based host and pathogen peptide interaction is reported by many viruses to disrupt or modulate host pathways to survive inside the host (Nussinov et al. 2014). Some of the functional machinery which viruses hijack by use of molecular mimicry are related to evasion and modulation of complement system and apoptosis, signaling pathways; control target protein levels or perturb post-translational modifications of host proteins (Barnett & Fujinami 1992, Farris et al. 2000, Fujinami et al. 2006, Rosen et al. 1995). The proteins, which are involved in molecular mimicry in viruses, are mostly disordered and they manipulate the host cellular mechanism by harboring short linear motifs (SLiMs) and molecular recognition features (MoRFs) on their protein (Hraber et al. 2020). The mimicry motifs in viruses are not continuous and are located in the disordered regions of proteins (Duro et al. 2015). In comparison to viruses, very little work has been done to understand the finer details of interaction between host and bacterial memitopes. In this thesis, we also analysed structural and functional properties of bacterial and viral host/pathogen mimicry proteins. We also explained the benefit of molecular mimicry in terms of immunological consequences in microbial pathogenesis.

The conventional therapy for autoimmune diseases has been the usage of immune-suppressants or immune-modulators that treat symptoms rather than the etiology and/or the causative mechanism(s) (Rosenblum et al. 2012). Molecular mimicry can cause several autoimmune diseases along with the disease caused by the infectious agents (Rojas et al. 2018). Hence, the process of molecular mimicry can be used as a stepping-stone to understand the initial interaction between the infectious agent as well as recognizing the self-determinant, understanding the pathogenic mechanism(s) involved, and designing strategies for the treatment and prevention of autoimmune disorders. Though, 60–70% of the patients initially respond to immunosuppression, in

many cases the patients show subsequent clinical remission or relapse of the autoimmune disease (van der Kooij et al. 2007). Henceforth, we designed a novel strategy for the treatment and prevention of autoimmune disorders. In which, we used molecular mimicry proteins and its interaction partners as a potential target to block the production of molecular mimicry proteins, which will ultimately lead to killing of pathogens.

In summary, the overall objectives of this thesis are as follows:

i. Creation of a data repository of experimentally verified molecular mimicry phenomena that leads to autoimmune diseases.

ii. Analysis of structural, functional and immunological characteristics of bacterial-, viral- and host-mimicry proteins and peptides.

iii. Exploration of proteins involved in molecular mimicry as novel drug targets. A case study with (a) *Mycobacterium tuberculosis* (b) *Mycobacterium avium* subsp. *paratuberculosis*

iv. Functional analysis of *M. tuberculosis* mimicry phenomena.

1.3. Structure of the thesis

This thesis is organized into 8 chapters. Chapter 1 is Introduction to the general framework of the work presented in this thesis. Chapter 2 is Review of Literature in which a basic description of molecular mimicry induced autoimmune diseases are discussed. Chapter 3 describes a database named as miPepBase - containing experimentally validated information related to autoimmune diseases caused due to molecular mimicry. Further, evolutionary, structural, immunological and functional characteristics of bacterial-, viral- and host-mimicry proteins, are described in Chapter 4.

We developed a therapeutic schema, which can be applied to repurposing known drugs and/or discovery of novel therapeutics against pathogenic bacteria, which exhibit molecular mimicry with the host's proteins. In the chapter 5, we have used *Mycobacterium tuberculosis* (Mtb) as a model organism and using our strategy we were able to found four drugs viz. DB08185, DB00759, DB01930 and DB07349 that

might be useful in treatment of *M*. t*uberculosis*.

In Chapter 6, using our proposed approach, we found eight DrugBank molecules, which might prove useful for treating three *Mycobacterium avium* subsp. *paratuberculosis* (MAP)-associated autoimmune diseases, type 1 diabetes, Crohn's disease, and multiple sclerosis. Moreover, the drug molecules identified during our analysis are either FDA-approved drugs or experimental drugs with proven efficacy. Hence, these can be easily incorporated in clinical studies or tested *in vitro* for assessing their suitability in treating MAP-associated autoimmune diseases.

In chapter 7, we explained how the *M. tuberculosis* mimicry and its interacting proteins export via exosomes, and hijacking the host's system to maintain the dormant phase. Finally, in Chapter 8, the key issues and significant outcome of the present work is summarized, which would be helpful for the treatment and prevention of autoimmune disorders.

2.1. Innate and adaptive immunity

The immune system protects organisms from infection with two major defenses mechanisms, with increasing specificity. (i) the innate, provides the first line of defenses against infection. (ii) the adaptive, specialized immune system. Both systems do not operate independently, instead they function in a highly interactive and coordinated manner, producing a combined response more effective than either branch could produce by itself. Further, both states of immunity required several soluble substances found in blood and other body fluids, which belongs to the humoral defense mechanism. Henceforth, the innate and the adaptive immune system use both cellular and humoral defense strategies.

2.1.1. Innate immune system: Fast and broadly effective

Innate immune response is the most rapid and evolutionary conserved arm of the immune system (Chaplin 2003). Many innate components are present before the pathogen encounter and constitute a set of disease resistance mechanisms which are not specialized for particular microbes but contain molecular and cellular components that can identify a particular class of molecules occurring in frequently encountered pathogens.

The innate defense consists of several elements:

- External barrier, include skin and all mucous membranes
- Acidity of stomach content and perspiration
- Phagocytosis - conducted by a group of specialized cells such as blood monocytes, neutrophils and tissue macrophages. They can recognize and neutralize invaders based on common molecular surface markers.
- Enzymes such as lysozyme - a hydrolytic enzyme found in mucous secretion and in tears, attack the peptidoglycan layer of bacteria cell wall.

2.1.2. The adaptive immune system: Precision and a long memory

When the body's first line of defense fails to completely eliminate the pathogen, then the adaptive immune system comes into picture, for complete eradication of invaders. As it takes longer time to recognize (approx. four to seven days), but selectively eliminates specific foreign microbes. But, defense response will be more efficient and faster, if the body comes into contact with already known antigens, because it has immunological memory (Ahmed 1992).

Adaptive immunity exhibits four characteristics attributes: (a) Antigenic specificity (b) Diversity (c) Immunological memory (d) Self-nonself recognition.

An adaptive immune response has various parts, activation of each part is depending upon the pathogen location into the host body. For instance, extracellular pathogens have presence of antibodies into blood, and if a pathogen is inside the tissue, a cell-mediated immune response is necessary (Sfriso et al. 2010).

These parts of the adaptive defense include:

1. T lymphocytes
2. B lymphocytes
3. Antibodies as soluble proteins in the blood
4. Cytokines in the blood and tissue as hormone-like messenger substances

2.1.2.1. T lymphocytes or T cells

T lymphocytes are the main component of the adaptive immune system, which are involved in a particular defenses mechanism. T cells are produced in the bone marrow and premature cells migrate into the thymus gland for maturation, mature T cells are able to differentiate between self and non-self antigens. In the thymus gland, T cells develop particular surface receptors which are capable of recognizing and binding pathogens. The interaction between the T cell surface receptors and pathogens stimulates the T cells for faster division and meanwhile they activate other defenses reactions. They are capable of eliminating pathogens from the body. During the defense reaction T cells are evolved into some specialized cells. These include: T helper cells, T killer cells or cytotoxic cells, memory T cells and regulatory T cells

(Adam et al. 1998).

2.1.2.2. B lymphocytes or B-cells

B lymphocytes are the other important cells that are involved in adaptive immunity. B-cells, production and maturation both occur into bone marrow. Mature B-cells produced antibodies which are in the blood as soluble proteins are specific for particular pathogens.

To activate T- and B-cells of the adaptive immune system either direct binding to the antigen presenting cells (APC) or interaction with different types of messenger molecules like cytokines are required.

2.2. "Molecular mimicry": an evolving concept

In 1964, Damian first proposed the term "molecular mimicry" that refers to the sequence or structural similarity between pathogens and their hosts, which assists microbes to avoid the host immune response (Damian 1964). Before introducing this term, Kaplan *et al.* provide experimental evidence of immune cross-reactivity in a rheumatic fever patient by examining the sera reaction of rabbits immunized with group A streptococcal cells to human heart tissue (Kaplan & Meyeserian 1962). Though, during that time their structural homology information was unavailable. Two years later, it was found that membrane structures in group A streptococcus shared structures with mammalian muscle (Zabriskie & Freimer 1966). Sometimes the structural similarity between host and pathogen may provide immune tolerance to the pathogenic organisms instead of immunogenic response activation, for e.g. structural similarity between the antigenic determinants of parasite and antigenic structures of human leads a parasitemia response (i.e. immune tolerance) in the host (Damian 1964). Although the concept of homology is conflicting in terms of immunogenic activation and tolerance, it suggests that these hypotheses could work together, and apart from homology some other factors are also responsible for triggering autoimmunity (Hardtke-Wolenski et al. 2017, Ma et al. 2017, Wu et al. 2018).

Though, clonal deletion prevents autoimmune response in the host. During clonal deletion, highly self-reactive lymphocytes are destroyed and prevent immune response against self-antigens (Rose 2015). Sometimes, a pathogen which shows

molecular mimicry took advantage over non-mimicking to avoid the host's immune response (Drayman et al. 2013).

There are numerous epidemiological and experimental evidences that suggest the role of infectious diseases in autoimmunity via molecular mimicry and cross-reactivity (Guarneri 2013). As of now it is not a very critically studied process of microbial pathogenesis. But slowly the role of molecular mimicry in microbial pathogenesis has been unfolding and it is implicated in a number of pathological conditions. In 1983, Fujinami *et al* observed the cross-reactivity between the murine antibodies of measles virus/herpes simplex virus (HSV) and human cells (Fujinami et al. 1983). In line, they also observed that myelin basic protein (MBP) encephalitogenic peptide shares homology with the hepatitis B virus polymerase (HBVP) that results in an autoimmune disorder named as encephalomyelitis (Fujinami & Oldstone 1985)

Furthermore, in nearly half of tuberculosis (TB) patients, autoantibodies responsible for wegener's granulomatosis and systemic lupus erythematosus and many others are being observed. It has been proposed by Elkington, *et al*. (2016) that TB “tricks” the immune system into attacking the lungs, enabling the bacteria to become more infectious (Elkington et al. 2016).

2.2.1. Type of molecular mimicry

Molecular mimicry can be achieved at four different levels:

(1) Sequence and structure mimicry

When sequence and structure of a host protein is hijacked by pathogen protein, it will give rise to orthologous proteins across the pathogen species. Although, to adapt to a similar structure, pathogens used an evolutionary strategy named as horizontal gene transfer (also known as lateral gene transfer) (Guven-Maiorov et al. 2016).

An example for sequence and structure mimicry type could be the Toll/Interleukin-1 receptor (TIR) domain. Although, any immune system receptors comprised TIR domain, for e.g. Toll-like receptors (TLRs), interleukin-1 receptor (IL-1R), and downstream effectors (such as Mal, MyD88, TRAM and TRIF) (Guven Maiorov et al. 2013, Guven-Maiorov et al. 2015). Further, pathogens also express TIR domain-containing proteins (Tcps) to interfere with TLR and IL-1R signaling (Chan et al.

2009, Cirl et al. 2008). For instance *Escherichia coli*, *Brucella melitensis* and vaccinia virus which encode/secrete TcpC, TcpB (Cirl et al. 2008), and A46R (Janeway & Medzhitov 2000) proteins, respectively that disrupt the host's immune signaling pathways. These bacterial TcpC and TcpB can directly bind to MyD88 that decreases the host pro-inflammatory cytokine production and promotes bacterial survival (Cirl et al. 2008).

(b) Structural mimicry without (or with very low) sequence similarity

Structural similarities between the microbes are responsible for the hijacking host's proteins. An example of complete structure conservation with very low sequence similarity is viral chemokine vMIP-II of Kaposi's sarcoma-associated herpesvirus (Qin et al. 2015) and chemokine receptor US28 of human cytomegalovirus (Burg et al. 2015)These viral proteins have very low sequence similarity with host proteins *viz.* vMIP-II is ~33% similar to the human chemokine CX3CL1, and US28 is ~29% similar to the human chemokine receptor CXCR4 in terms of structure. The structural similarity between the surface receptors is not only advantageous for pathway inhibition/activation, but also allows pathogens to anchor to host surfaces and facilitate nutrient uptake by the host (Finlay & McFadden 2006)

(c) Motif mimicry

Some pathogens have homologs of short amino acid sequences instead of the whole proteins or domains, known as motif mimicry (Davey et al. 2011, Hagai et al. 2014). Short linear motifs (SLIMs) are present within (intra)- and across (inter)-species. SLIMs are generally composed of 3–10 residues, and were proposed to have important roles in pathway modification. They were proposed to have 'evolutionary plasticity'; that is, changes of few residues in the protein can rewire pathways thereby adapting cell signaling (Davey et al. 2011). Many viruses possess more than one SLIM in their genome, enabling them to interfere with more than one host interaction by competitively displacing the host protein partner (Davey et al. 2011).

One example for motif mimicry is the WxxxE motif in many bacterial guanine nucleotide exchange factors (GEFs), such as Map and EspM2 of *E. coli* and SifA of *Salmonella* **54,55**. Like endogenous GEFs, these pathogenic GEFs activate the GTPases in the host. Although not located at the catalytic site, the presence of the

WxxxE motif is critical for the GEFs' interactions with the GTPases. SopE of *Salmonella* does not possess a WxxxE motif, but it still folds into a structure, which is very similar to those effectors that do. It has Y and T residues, which correspond to W and E in the motif. Thus, in spite of the different residues, it conserves the chemical properties at the corresponding sites.

(d) Interface mimicry

Interface structure similarity is the most common type of molecular mimicry. For host-pathogen protein–protein interaction there is no need for extensive sequence or structure conservation. Sometimes the overall structures of the proteins are distinct, but still they use similar interface architectures to interact with their partners, which suggest that these recurring architectures are favorable scaffolds (Keskin & Nussinov 2005). The interface mimicry assists pathogen evasion (Yamada et al. 2015)and also supports many cellular events that take place through competitive binding (Franzosa & Xia 2011, Franzosa et al. 2012) The example for interface mimicry is human fibronectin and the invasin protein of *Yersinia* bacteria, both of which bind to human integrin in a similar way (Stebbins & Galán 2001, Zur Hausen 2009). Despite lack of overall structural and sequence homology, they have similar chemical properties at the integrin-binding site.

2.2.2. Molecular mimicry and cross-reactivity: what is necessary?

The activation of T and B-cells against self-antigen is the major consequence of any autoimmune diseases, wherein the former lymphocytes play a significant role in 'T-cell mediated autoimmune diseases' (Bertsias et al. 2010, Gregersen et al. 1987, Mathis et al. 2001). Earlier, it was assumed that due to highly specific recognition – T cells show low cross-reactivity with infectious diseases (Cusick et al. 2012). Later some researchers show that only a very small portion of an antigen is being recognized by T-cells receptors (TCR) and shows MHC restriction – interaction between specific TCR and MHC bearing peptides. In this context, 8-10 amino acids are presented by MHC class I to CD4+ T-cells, and 14-18 amino acids are presented by MHC class II to CD8+ T-cells (Reay et al. 1994, Sinigaglia & Hammer 1994, Wucherpfennig et al. 1994). In these short antigenic peptides, there are some anchor

residues that are meant to bind specific pockets on the MHC molecules, resulting in some specificity of interactions with MHC (Harbige et al. 2017, Paun et al. 2016, Vatti et al. 2017). Also, there is a certain degree of plasticity in the other residues, so that different peptides or chemical xenobiotics can bind to single MHC molecules with certain specificity, even though some peptides can bind to more more than one MHC, a phenomenon known as "polyspecificity" (Wucherpfennig et al. 2007).

In a case study using mice – as host and rat insulin promoter-lymphocytic choriomeningitis virus (RIP-LCMV) – as pathogen, it has been shown that heterologous sequential viral infections can increase cross-reactive T cells in the targeted organ above the disease initiating threshold, leads to major tissue injury and in this case rapid development of diabetes, an autoimmune disease (Christen et al. 2004). The cross-reactivity between the host-pathogen epitopes can increase but not initiate autoimmune diseases in the host. Interestingly, this observation suggested that in humans, the combined effect of some immunologically cross-reactive viruses are the main reason for boosting the autoimmune disease(s).

2.3. The initiation of autoimmunity: lighting the match

The rate of morbidity and mortality in the human population increases due to autoimmune diseases (AD). The immune system is mostly able to differentiate between self and non-self antigens, if not, it activates autoimmune response into the body. There are two types of autoimmune diseases: **tissue-specific** type where antigen targeting is single tissue-specific and **systematic** type in which more than one tissue and ubiquitously expressed antigens are targeted. It is difficult to pinpoint the single factor that causes autoimmune diseases, due to delayed symptoms well after the abnormal reaction begins (Fridkis-Hareli 2008). Apart from humans, some animals also show an autoimmune response. Although, many groups of researchers are using model organisms to find-out early stage diagnosis, to prevent poor prognosis (Konforte et al. 2012).

The autoimmune disease is a complex mechanism, which can be triggered either by numerous infectious agents (virus and bacteria) or molecular and cellular pathways and events. Furthermore, the sequence or structure homology between infectious agents and self-antigen lead to proinflammatory response into the host, termed as

molecular mimicry (cross-reactivity) and it is one of the causes of autoimmunity (Cusick et al. 2012). This cross-reactivity between host's and pathogen's epitopes can have deleterious or protective effects onto the former.

2.3.1. Type of autoimmune diseases

There are more than 100 autoimmune diseases reported till date. But the most common reports autoimmune diseases are caused by 'linear molecular mimicry' (explained in section 3.1.1) because it is easiest to study. Some most commonly found diseases are multiple sclerosis (Nielsen et al. 2007), Guillain-Barré syndrome (Sheikh et al. 1998), Type 1 diabetes (Coppieters et al. 2012), Rheumatoid arthritis (Ebringer & Rashid 2009), Systemic lupus erythematosus (Poole et al. 2006), Sjögren's syndrome (Igoe & Scofield 2013), Systemic sclerosis (Grossman et al. 2011), Autoimmune thyroid disease (Benvenga & Guarneri 2016), Autoimmune hepatitis (Christen & Hintermann 2018) and primary biliary cholangitis (Van de Water et al. 1993)

Interestingly, many mycobacterial antigens have been associated with autoimmune diseases. This prompted us to in depth study of mycobacteria that share sequence similarity with host antigens, and elicit T cell autoimmune reactions. Some have been discussed in detail below:

- ***Multiple sclerosis (MS)*** **:** MS is a chronic disease of the central nervous system in which myelin covering around the nerve fibers is significantly reduced or disappears completely. Since nerve fibers cannot efficiently conduct the electrical impulses in absence of myelin covering, hence the electrical impulses received from the brain do not flow smoothly to the target nerve. Due to this, the muscle movement becomes very erratic. Although the exact cause of MS is not known, it is believed to be a multifactorial disease caused by autoimmune processes (Libbey et al. 2007). Multiple sclerosis is one of the leading disease caused due to molecular mimicry. Many clinical evidence suggest that an infectious agent might be responsible for breaking tolerance and elicitation of autoimmune response against myelin proteins. Indeed, it was observed that some infectious viruses and bacteria *viz.* endogenous retrovirus, Epstein Barr virus (EBV), *Chlamydia pneumoniae*,

Helicobacter pylori, and *Mycobacteria spp*., might have a role in the MS (Cossu et al. 2018). According to some reports, in comparison to healthy individuals, MS patients have enhanced proliferation of lymphocytes against *Mycobacterium tuberculosis* and *Mycobacterium leprae* - recombinant heat shock proteins 65 and 70 (HSP60 and HSP70) proteins. Some researchers also observed that in Sardinian MS patients there was an increase in circulating antibodies against *Mycobacterium avium* subsp. *paratuberculosis* (MAP) HSP70 (Cossu et al. 2017).

- ***Type 1 diabetes mellitus (T1DM):*** It is a second most common childhood chronic disease, in which T-cells penetration destroys the insulin producing beta cells of the pancreas. T1DM can be broadly classified into two types: (i) **type 1A -** patients have antibodies against host-proteins such as glutamic acid decarboxylase 65 (GAD65), insulin, insulinoma associated proteins (IA-2) and heat shock protein 60 (hsp60), and (ii) **type 1B -** a few cases was reported for type 1B, has no known cause (Rani et al. 2010). Additionally, GAD is a 65 Kda enzyme that catalyzes the α-decarboxylation reaction of L-glutamic acid to synthesis of gamma-amino butyric acid (GABA). GAD65 is a major autoantigens that is found in many tissues and participates in TIDM pathogenesis. Henceforth, GAD65 antibodies are more commonly used for diagnosis purposes. Although, high blood sugar level (hyperglycemia) leads to long lasting side effects on the body such as stroke, cardiovascular ailments and diabetic nephropathy/neuropathy/retinopathy. Currently, insulin replacement therapy is the only symptomatic aid for T1DM. Some experimental evidence suggests that role of *Mycobacterium avium* subsp. *paratuberculosis* in triggering T1DM. For instance, that molecular mimicry between human GAD65 and MAP Hsp65 elicits an autoimmune reaction targeting beta cells in pancreatic islets results in T1DM and insulin deficiency (Dow 2012).
- ***Leprosy*:** In leprosy, autoimmune response generated against host's components, primarily causing nerve damage. Some clinical reports suggested that in leprosy patients there is an increase in autoantibodies level and lympho-proliferative response against myelin basic protein (MBP). Further, cross-reactivity between *Mycobacterium leprae* proteins Lysyl-tRNA synthetase/50S ribosomal protein L2 and MBP (myelin A1 protein) of host

caused nerve damage in leprosy (Singh et al. 2015).

- ***Crohn's disease*:** an immune-mediated inflammatory bowel disease (IBD), with unclear cause. CD shows similarity with intestinal tuberculosis and Johne's disease. Mostly in ruminants and primates, CD is known to be caused by *Mycobacterium avium* subsp. *paratuberculosis* (MAP). Some studies prove that MAP exhibits 'epitope mimicry' with the human intestinal proteins, which elicits host autoreactive T- or B-cells leading to CD. In this support, comparative analysis between CD patients and healthy control shows that increased levels of antibody against specific MAP antigens and detection of more viable MAP in peripheral blood and intestinal tissue in CD patients (McNees et al. 2015). Reportedly, in CD patients anti-MAP antibiotic treatment is effective, but needs some larger-scale studies as well.

2.3.2. Autoimmunity can predispose to infectious diseases

During the autoimmune response, neutralization of important immune system components occurs by the autoantibodies, which are essential in mounting anti-microbial responses. These autoantibodies might either enhance current infectious disease(s) or susceptibility of the host to more infected microbes such as bacteria, viruses, and opportunistic fungus. For instance, cytokines – important regulatory molecules of the innate and adaptive immune system, plays an important role in increasing the anti-microbial response. In this context, some cytokines namely INF-c and IL-2 are produced by the Th1 cells and provide protection against intracellular virus and bacteria (for e.g. *Mycobacterium spp.* and *Salmonella sp.*). There are some classical examples of other cytokines also, which can provide protection against extracellular pathogens, *viz.*, Th-17 secrete IL-17, IL-21, and IL-22 that provided protection from *Candida* (a fungus) (Aimanianda et al. 2009, Bettelli et al. 2008, Kolls & Khader 2011, Zhu & Paul 2008). Therefore, any hamper in cytokines activity would directly affect the defense mechanism, such as use of neutralizing antibodies against these cytokines affects the cellular functions and elimination of pathogens and predisposes the host to infectious diseases. Autoimmune patients with Crohn's disease, rheumatoid arthritis or psoriasis (a skin disorder) treated with TNF-**α** – monoclonal antibodies, increase vulnerability to mycobacterial, listerial, and viral infections (Dinarello 2003, Maródi & Casanova 2010, Winthrop & Chiller 2009).

2.3.3. Therapeutic options for autoimmunity-associated infectious diseases

To target autoimmunity-associated infectious (AAI) diseases most of the strategies should be intended to inhibit both infection as well as autoimmune response, such as blocking autoantibody-producing B cells and neutralizing autoantibodies are the most suitable therapeutic. The combination of immunosuppressive treatment and antimicrobial agents are the most appropriate therapeutic for AAI diseases. Alternatively, some strategies such as plasmapheresis i.e. removal of autoantibody or providing exogenous cytokines, to enhance autoantibody development, could also be used to treat AAI diseases, but the major disadvantage is that they are not eliminating the autoantibody source (i.e. autoantibody producing B-cells and plasma cells). B-cell targeted therapies can be used to eliminate antibodies producing B-cells, as they decrease the total immunoglobulin level and lead to serious infection predisposition, but do not kill the antibody producing plasma cells.

Furthermore, polyclonal intravenous immunoglobulin (IVIg) in combination with anti-microbial agents would be more efficient and safe therapeutics to target diverse AAI diseases (Kazatchkine & Kaveri 2001, Tha-In et al. 2008). To maintain the immune tolerance and to suppress autoimmunity, IVIg targets both soluble mediators and cellular component of the autoimmunity, and there are several mechanisms through which it inhibit the diseases such as initiate B cell tolerance, regulation of immunoglobulin tolerance, neutralization of anti-cytokine autoantibodies by broad-spectrum anti-idiotypic antibodies, suppression of innate antigen presenting cells and inhibition of T cell help to B cells, and expansion of CD4+CD25+ regulatory T cells (Kazatchkine & Kaveri 2001, Tha-In et al. 2008). However, an effective dose regime and duration of IVIg therapy determination is the major concern. Although, for several autoimmune and inflammatory diseases a combination of IVIg and B cell–targeted therapies are proved as successful therapy. Although, triple medication – combination of B-cell targeted therapies, antimicrobial agents, and IVIg would be considered most suitable therapies to target different AAI diseases (Ahmed et al. 2006, Vo et al. 2008).

2.4. *Mycobacterium tuberculosis* (Mtb): a case study

The process of molecular mimicry is also well known in *Mycobacterium tuberculosis*

(Mtb). Elkington et. al. (2016) have proposed that autoimmunity is a critical and overlooked process of TB pathology, and present clinical and experimental observations also support this hypothesis. For example in nearly half of TB patients' autoantibodies responsible for wegener's granulomatosis and systemic lupus erythematosus have been observed (Elkington et al. 2016).

Few other autoimmune diseases such as inflammatory bowel disease, behçet's disease, ankylosing spondylitis, crohn's disease, ulcerative colitis, and sarcoidosis have also been associated with the TB pathogenesis. Using differential gene expression analysis among patients with TB and patients with autoimmune or infectious diseases, Clayton et al. suggested that combination of infection and autoimmune disease signatures could explain 96.7% of the differentially expressed TB signature (Clayton et al. 2017). On the basis of this observation they suggested that pathology in TB results from an interplay between infection and a currently unrecognized autoimmune process.

Though, it was observed that immunosuppressive therapies (mentioned in section 4.1.3) allow Mtb to proliferate (Proal & Marshall 2018). Similar observation (pathogenesis enhancement under immunosuppressive medications) extended to some other microbiome pathogens (Finlay & McFadden 2006). It was reported that immunosuppressive therapies are majorly responsible for 'microbial dysbiosis' - microbial imbalance or maladaptation on or inside the body, and also associated with almost every autoimmune condition (Diaz et al. 2013, Nellore & Fishman 2016). Some experimental evidence also suggests that patients treated with immunosuppressive therapeutic shows a high rate of relapse and co-morbidity.

In contrast, treatments which can target pathogens at the root of the disease process are needed to manipulate this complicated phenomenon. Herein, the mimicry inducing pathways can be used as potential targets, and drug repurposing would be a promising technique.

Summary & Future prospect

Mimicry is the resemblance of one life form with another life form, and it provides a selective advantage to the mimicker. Mimicry can involve physical or behavioural traits, and well-studied examples are Batesian and Mullerian mimicry. Mimicry as observed in nature is not only an ecological phenomenon, it is observed in the microbial world too!! Molecular mimicry can be defined as structural, functional, or immunological similarity between the host and pathogen macromolecules. Molecular mimicry can be present in the form of complete identity or homology at the protein level, or as a similarity in the sequences of amino acids or as structural similarity of host and pathogen proteins. Sometimes the host develops an immune response against self-proteins involved in molecular mimicry, which results in autoimmune diseases.

The work described in this thesis explains studies related to sequential molecular mimicry based autoimmune diseases. As we know some autoimmune diseases affect the rate of morbidity and mortality in the human population. The studies presented in this thesis are based on *in silico* work, which were carried out to understand different properties associated with the mimicry peptides/proteins. When we started this work, there was only one database namely mimicDB which provides information about proteins or epitopes involved in host-pathogen interactions. But it contained predicted information related to only a few human parasites. Henceforth, we foremost developed a database named as miPepBase **(Mi**micry **Pep**tide Data**base),** it incorporated the information related to autoimmune diseases as well as in-depth information about mimicry peptide and proteins. It is freely available at http://proteininformatics.org/mkumar/mipepbase.

Though researchers have studied various aspects of molecular mimicry and a few studies have reported some structural characteristics of viral mimicry proteins to the best of our knowledge, structural and functional characteristics of mimicry proteins of viruses, bacteria and hosts have not been explored in detail. Thus, we first time reported the structural and functional characteristics of bacterial, viral and host-mimicry.

The order/disorder propensity in the bacterial, viral and host mimicry proteins/peptides were studied along with the prevalence of intrinsically disordered regions (IDRs) like molecular recognition features (MoRFs), short linear motifs (SLiMs), and low complexity regions (LCRs). Our results indicated that the majority of the bacterial and viral mimicry proteins and mimitopes were ordered and, only a few mimitopes harboured the functional units present in the disordered regions of proteins like MoRFs, SliMs and LCRs. The majority of the host mimicry proteins were disordered, but the host mimitopes were ordered. The fact that most of the host mimitopes were ordered suggests that bacteria and viruses might preferentially select those regions of the host proteins for molecular mimicry which are ordered and thus show a lesser structural flexibility and immunological activity. Functional analyses indicated that both bacterial and viral mimicry proteins were involved in similar functions while the host mimicry proteins were multifunctional and mainly involved in ion binding, symbiont processes and signalling pathways. Our in-depth structure-function relationships in bacteria and viruses can help find ways to mitigate the effects of the infection.

Herein, we first time explained the immunological consequences of bacteria and virus host/pathogen memitopes. The immunogenic study revealed that viral, bacterial, and host memitopes are non-immunogenic, which help them to escape host's immune response (elaborate..).

Most commonly used therapy for autoimmune diseases has been the usage of immune-suppressants or immune-modulators that treat symptoms rather than the etiology and/or the causative mechanism(s). Although, in many cases (e.g. *Mycobacterium tuberculosis*) these therapies rebound to the host and enhance pathogenesis. In order to overcome this challenge, we developed a novel drug-repurposing methodology to target autoimmune diseases. In which, we 'tried-and-

tested' our approach to target pathogens at the root cause of disease. Herein, using the interaction partners of pathogen's mimicry protein as a chokepoint target, we are able identify many potential drug-target molecules for two model organisms namely, *Mycobacterium tuberculosis* complex (MTC) and *Mycobacterium avium* subsp. *paratuberculosis* (MAP). As we found in this work most of the proposed drugs are either FDA-approved or experimentally validated. Henceforth, these can be easily incorporated into clinical studies or tested *in vitro* for assessing their suitability in some autoimmune diseases namely multiple sclerosis, type 1 diabetes mellitus, crohn's disease and leprosy, caused by MTC and MAP. We also anticipate that the proposed schema can be used to target other pathogens, and number of drug-targets can be increased by incorporating more databases.

In our study, we found that most of the mycobacterium drug target molecules predicted to have a central role in the regulation of latency. These observations encourage us to analyze more in-depth about MTC molecular mimicry phenomenon. Hence, in this thesis using exosomal RNA-seq data of clinical samples collected from active- and latent-TB infected individuals, we explored the pathways in which mimicry proteins and its interacting partners of *Mycobacterium tuberculosis* H37Rv are involved. We observed that majority of mimicry proteins and its interacting partner in the LTBI exosomes that might carry out function far off from the host cell of the exosome. Furthermore, pathway and functional analysis of differentially expressed genes (DEGs) in LTBI and ATB samples indicates a down-regulation of signaling and immune system pathways, and up-regulation of apoptosis/necrosis process. Recent evidence suggests that extracellular vesicles can mediate immune stimulation or suppression and they can drive inflammatory, autoimmune and infectious disease pathology. Thus, modulation of extracellular vesicles has the potential to be used as therapeutic agents to.