

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA GERAL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA



PhD Thesis

***In silico* approaches to design multi-epitope
vaccines and uncover potential drug targets
against *Corynebacterium diphtheriae* and
*Clostridioides difficile***

PhD candidate: Mariana Passos Santana

Thesis supervisor: Anderson Miyoshi

Thesis co-supervisor: Sandeep Tiwari

BELO HORIZONTE

July 2019

MARIANA PASSOS SANTANA

***In silico* approaches to design multi-epitope vaccines
and uncover potential drug targets against
Corynebacterium diphtheriae and *Clostridioides
difficile***



**Thesis presented as partial requirement for the
degree of Doctor in Genetics, to the Department of
General Biology at the Institute of Biological Sciences,
Federal University of Minas Gerais.**

SUPERVISOR: Prof. Dr. Anderson Miyoshi

CO-SUPERVISOR: Dr. Sandeep Tiwari

BELO HORIZONTE

July 2019

043

Santana, Mariana Passos.

In silico approaches to design multi-epitope vaccines and uncover potential drug targets against *Corynebacterium diphtheriae* and *Clostridioides difficile* [manuscrito] / Mariana Passos Santana. - 2019.

225 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Anderson Miyoshi. Co-orientador: Dr. Sandeep Tiwari.
Tese (doutorado) - Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Genômica. 2. Imunização. 3. Vacinologia. 4. *Corynebacterium diphtheriae*. 5. *Clostridium difficile*. 6. Resistência a Múltiplos Medicamentos. I. Miyoshi, Anderson. II. Tiwari, Sandeep. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 575



ATA DA DEFESA DE TESE

Mariana Passos Santana

125/2019
entrada
2º/2014
CPF:
080.462.776-22

Às quatorze horas do dia **19 de julho de 2019**, reuniu-se, no Instituto de Ciências Biológicas da UFMG, a Comissão Examinadora de Tese, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: "**In silico approaches to design multiepitope vaccines and uncover potential drug targets against *Corynebacterium diphtheriae* and *Clostridioides difficile***", requisito para obtenção do grau de Doutora em **Genética**. Abrindo a sessão, o Presidente da Comissão, **Anderson Miyoshi**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	Instituição	CPF	Indicação
Anderson Miyoshi	UFMG		APROVADA
Sandeep Tiwari	UFMG		APROVADA
Sergio Costa OLIVEIRA	UFMG		APROVADA
Siomar de Castro Soares	Universidade Federal do Triângulo Mineiro	05695782611	APROVADA
Rommel Thiago Jucá Ramos	Universidade Federal do Pará	729487702-49	APROVADA
Álvaro Cantini Nunes	UFMG		APROVADA

Pelas indicações, a candidata foi considerada: _____

O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 19 de julho de 2019.

Anderson Miyoshi - Orientador _____

Sandeep Tiwari - Coorientador _____

Sergio Costa _____

Siomar de Castro Soares _____

Rommel Thiago Jucá Ramos _____

Álvaro Cantini Nunes _____

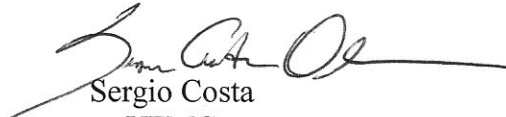


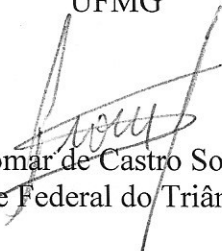
"In silico approaches to design multiepitope vaccines and uncover potential drug targets against *Corynebacterium diphtheriae* and *Clostridioides difficile* "
Mariana Passos Santana


Tese aprovada pela banca examinadora constituída pelos Professores:

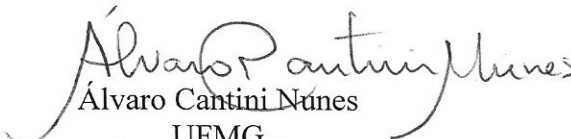

Anderson Miyoshi - Orientador
UFMG


Sandeep Tiwari - Coorientador
UFMG


Sergio Costa
UFMG


Siomar de Castro Soares
Universidade Federal do Triângulo Mineiro


Rommel Thiago Jucá Ramos
Universidade Federal do Pará


Alvaro Cantini Nunes
UFMG

Belo Horizonte, 19 de julho de 2019.

“Na vida, não existe nada a temer, mas a entender”

Marie Curie

**“Quem elegeu a busca
Não pode recusar a travessia”**

Guimarães Rosa

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisors Prof. Dr. Anderson Miyoshi and Dr. Sandeep Tiwari for the continuous support of my Ph.D study and related research, for their patience, motivation, and immense knowledge. The guidance helped me in all the time of research and writing of this thesis. I could not have imagined having better advisors for my Ph.D study.

Besides my advisors, I would like to thank the rest of my thesis committee: Prof. Dr. Álvaro Cantini, Prof. Dr. Rommel Ramos, Prof. Dr. Sergio Costa Oliveira, and Prof. Dr. Siomar de Castro Soares, for their insightful comments and encouragement, but also for the hard questions that incited me to widen my research from various perspectives.

My thanks also goes to Prof. Dr. Vasco Azevedo, Dr. Pascale Serror, and Dr. Francis Repoila, who provided me an opportunity to join their team, and who gave access to the laboratory and research facilities.

A very special gratitude goes out to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for providing my scholarship and funding during my PhD.

I thank my fellow labmates from TecnoGen, LGCM and CPE (especially Stephane, Kato, Tati and Vanessa) for the stimulating discussions and friendship; it was great sharing laboratory with all of you during the last years. In addition, I thank my friends from UFMG and INRA.

Last but not the least, I would like to thank my family: my parents (Rita e Alfredo), my sister (Andreia), brothers (Alfredo e Rick) and nephew (Lucca) for their continuous support, love and patience, you bring light to my life. To my aunts, uncles, cousins, and friends (Gatas, Chicas, amigos de BH, Viçosa e Paris) for supporting me spiritually throughout writing this thesis. Thank you all for everything; I love you!

Thanks for all your encouragement!

SUMMARY

ABREVIATION	ix
LIST OF FIGURES	x
LIST OF TABLES	xii
ABSTRACT	xiii
INTRODUCTION	14
1. The human microbiota and related diseases	15
2. The interplay between microbiota, pathogens and immune system	18
2.1. Innate immune response	22
2.2. Adaptive immune response	23
3. Vaccines	27
3.1. Classical vaccination classification	31
3.1.1. Live attenuated vaccines	32
3.1.2. Killed or inactivated vaccines	33
3.1.3. Subunit and Conjugate Vaccines	33
3.1.4. Inactivated toxin (toxoids)	34
3.2. Vaccine adjuvant and delivery system	34
3.3. Next-generated vaccines	37
3.3.1. Next-generation vaccines and OMICs approach	38
3.3.2. Next-generated vaccines methodologies	41
3.3.2.1. Reverse vaccinology	41
3.3.2.2. Immunoinformatics	42
3.3.2.3. Structural vaccinology	42

4.	The two sides of antibiotic treatment and possible alternatives	42
CHAPTER I: Design of a broad-spectrum candidate multi-epitope vaccine against Diphtheria: An integrative immunoinformatics approach.....		
1.	<i>Corynebacterium diphtheriae</i> : a life-threatening bacteria	46
1.1.	The <i>Corynebacterium</i> genus	46
1.2.	The pathogenic bacteria <i>C. diphtheriae</i>	47
1.3.	<i>C. diphtheriae</i> factors associated with infection	50
1.3.1.	Diphtheria toxin and the diphtheria toxin repressor	52
1.3.2.	DIP0733	54
1.3.3.	Pili proteins	55
2.	Diphtheria.....	56
2.1.	Main characteristics, transmission and clinical manifestations	56
2.2.	Diagnosis and treatment	58
2.3.	Diphtheria vaccine	59
2.4.	Application of cutting-edge approaches to improve diphtheria management	62
3.	Submitted paper: Design of a broad-spectrum candidate multi-epitope vaccine against Diphtheria: An integrative immunoinformatics approach.....	63
Supplementary Tables - Design of a broad-spectrum candidate multi-epitope vaccine against Diphtheria: An interagive immunoinformatics approach.....		
		107
CHAPTER II: A Reverse Vaccinology and Immunoinformatics Based Approach for Multi-epitope vaccine against <i>Clostridioides difficile</i> infection		
		117
1.	Environmental background	118
1.1.	The neat interplay between the gut microbiota and the host	118
1.2.	The frail line between eubiosis and dysbiosis	120
2.	The pathogen <i>Clostridioides difficile</i>	124

2.1.	<i>C. difficile</i> phylogeny and main features	124
2.2.	<i>C. difficile</i> virulent factors	126
2.2.1.	The Pathogenic locus proteins	126
2.2.2.	Cytotoxic distending toxin (CDT)	127
2.2.3.	Others (non-toxins related) virulence factors	128
2.3.	The interplay between <i>C. difficile</i> and the gut environment.....	129
2.4.	<i>C. difficile</i> infections.....	132
2.5.	CDI currently treatments	136
2.5.1.	Antibiotic therapies.....	137
2.5.2.	Fecal transplantation and other microbiota related therapies	139
2.5.3.	Immunization strategies	139
2.5.3.1.	<i>C. difficile</i> and the immune system	139
2.5.3.2.	Vaccine development	142
2.5.4.	Summary of therapies under development	142
3.	Submitted paper: A Reverse Vaccinology and Immunoinformatics Based Approach for Multi-epitope vaccine against <i>Clostridioides difficile</i> infection	144
	Supplementary Tables - A Reverse Vaccinology and Immunoinformatics Based Approach for Multi-epitope vaccine against <i>Clostridioides difficile</i> infection.....	195
	GENERAL CONCLUSIONS AND FUTURE PROSPECTIVES	198
	REFERENCES	201

ABREVIATION

GIT	Gastro intestinal tract	IFN- γ	Interferon gamma
AID	Antibiotic-induced intestinal dysbiosis	Ig	Immunoglobulin
AMP	Antimicrobial peptides	IL	Interleukin
APC	Antigen-presenting cells	ILC	Innate lymphoid
BCR	B cell receptor	LAM	Lipoarabinomannan
BSH	Bile salt hydrolases	LCA	Lithocholate
CA	Cholate	LM	Lipomannan
CDC	Center for disease control	MDR	Multidrug resistance
CDCA	Chenodeoxycholate	MHC	Major Histocompatibility complex
CDI	<i>Clostridioides difficile</i> infections	NOD	Nucleotide-binding oligomerization domain
CDT	Cytolethal distending toxin	PaLoc	Pathogenicity locus
CTL	Cytotoxic T lymphocytes	PAMPS	Pathogen-associated molecular patterns
DC	Dendritic cells	PPI	Protein-protein interactions
DCA	Deoxycholate	PRR	Pattern recognition receptors
DT	Diphtheria toxin	ROS	Reactive oxygen species
DTP	Diphtheria-tetanus-pertussis	SCFA	Short-chain fatty acids
DTxR	Diphtheria toxin repressor	TCA	Taurocholate
EPI	Expanded Programme on Immunization	TCR	T cell receptor
ER	Endoplasmic reticulum	TLR	Toll-like receptors
FMT	Fecal microbiota transplantation	TNF	Tumor necrosis factor
GVAP	Global Vaccine Action Plan	UDCA	Ursodeoxycholic acid
IBD	Intestinal bowel disease	WHO	World Health Organization

LIST OF FIGURES

FIGURE 1: Microbiota in number.....	15
FIGURE 2: Invasion of pathogens in the GIT.	16
FIGURE 3: Factors that affect the gut microbiota and some consequences of the dysbiosis.	17
FIGURE 4: Innate and adaptive immunity in the regulation of microbial homeostasis.	19
FIGURE 5: The innate and adaptive immune system.	21
FIGURE 6: Antigen processing pathways for the MHC type I and II molecules and presentation to naïve T cells.....	24
FIGURE 7: Adaptive immune response.....	26
FIGURE 8: Immune response pathways triggered by vaccination.	28
FIGURE 9: The immune response to vaccination with and without adjuvant.	35
FIGURE 10: Diseases and metabolic changes associated with the gut microbiota.....	40
FIGURE 11: The cell envelope of <i>Corynebacterium diphtheria</i>.	48
FIGURE 12: <i>C. diphtheriae</i> pan proteome..	49
FIGURE 13: <i>C. diphtheriae</i> key proteins for infection.....	51
FIGURE 14: Diphtheria toxin mechanism of action.....	53
FIGURE 15: Diphtheria immunization coverage (DTP vaccine third dose) in 2017.	62
FIGURE 16: Workflow of the <i>C. diphtheriae</i> vaccine construction.....	63
FIGURE 17: Bile acid differentiation in the gut.	120

FIGURE 18: Correlation between gut bacteria and metabolite.	123
FIGURE 19: The interplay between <i>C. difficile</i>, the microbiota and metabolites.	130
FIGURE 20: <i>Clostridioides difficile</i> infection.	133
FIGURE 21: Innate immune response against <i>C. difficile</i>.	141
FIGURE 22: Workflow of the <i>C. difficile</i> vaccine and drug targets selection trough reverse vaccinology and subtractive genomics, respectively.	144

LIST OF TABLES

TABLE 1: ‘WHO position paper’ – Recommendations for routine Immunization.	30
TABLE 2: Example of vaccines available for a variety of disease according with the development technique.	32
TABLE 3: Characteristics of adjuvants used in licensed vaccines.	37
TABLE 4: Current therapies used against <i>Clostridioides difficile</i>.	137
TABLE 5: Under development treatments against <i>Clostridioides difficile</i>.	143

ABSTRACT

Immunization is a crucial strategy to prevent the dissemination of pathogens without aggravating resistance. The society is facing an unprecedented hazard due to the exacerbate use of antimicrobials that accelerated the overall antimicrobial resistant problematic that utterly leads to an increase in the cost within the health system. Furthermore, the continuous outbreaks of previously controlled pathogens can be linked with a low immunization coverage due to conventional vaccines drawbacks and socioeconomic difficulties. The use of cutting-edge technology may resolute many of those problems in a customized and cost-effective way.

Among the bacterial pathogens that present a high morbidity and mortality rates, we highlight two crucial pathogens: *Corynebacterium diphtheriae* and *Clostridioides difficile*, the causative agents of diphtheria and *C. difficile* infections (CDI) (e.g. diarrhea to pseudomembrane colitis). Diphtheria was considered the main cause of child mortality but the development of the diphtheria vaccine, which is based in an inactivated form of the diphtheria toxin, significantly decrease the number of cases. However, this disease continuous to be a problem, especially considering the emergence of variants of non-toxigenic *C. diphtheriae* and multidrug resistant strains. Likewise, the occurrence of *C. difficile* multidrug resistance strains became an issue since it became more challenging to treat the disease. CDI recommended treatment is based on specific antibiotics that possible, in a near future, may became obsolete due to the high capacity of the bacterium to adapt to the surrounding environment (variable genome due to the presence of mobile elements). Therefore, the design of innovative and more inclusive vaccines may circumvent the hitches and ameliorate/resolve the spread of those life-threatening pathogens.

Bearing this, we applied an immunoinformatics strategy (*in silico* approach) to design two multi-epitope vaccines (one for each pathogen) encompassing toxigenic and non-toxigenic strains of *C. diphtheriae* and *C. difficile*. The results predict that both vaccines can induce cellular and humoral response due to the presence of predicted MHC-I, MHC-II and B cell epitopes. Furthermore, the vaccines and the complex (vaccine + Toll-like receptors) were classified as stable, non-allergenic and antigenic. The *in silico* approach considerable reduced the cost and time that would be spend designing the vaccines and selecting the vaccine (reverse vaccinology) and drug targets (subtractive genomics); nonetheless, considering that the analyses were performed only *in silico*, experimental validations are required to confirm the results.

INTRODUCTION

1. The human microbiota and related diseases

The human body harbors a wide number of bacteria (i.e. estimation ratio of bacteria to human cells of 1.3:1 considering a male adult) with most considered harmless and beneficial (SENDER; FUCHS; MILO, 2016a, b) (Figure 1). The microbiota is composed by ‘good’ (e.g. commensal) and ‘bad’ bacteria (e.g. pathobionts and pathogens) that are usually located in mucosae regions (frontline barriers), which includes skin, upper respiratory tract (nasal and oral), urogenital tract, and the gastro intestinal tract (GIT). In the physiological state known as dysbiosis (i.e. homeostasis imbalance), pathogenic microorganisms and/or pathobionts (i.e. resident microbes with pathogenic potential) can overgrowth and thrive in challenging conditions. Additionally, the disruption of eubiosis (i.e. intestinal ecosystem balance) might be accompanied by a breach of host barriers that allow pathogens to reach deeper organs causing a sum of diseases (GABORIAU-ROUTHIAU *et al.*, 2009; HORNEF, 2015; IEBBA *et al.*, 2016; RIBET; COSSART, 2015).

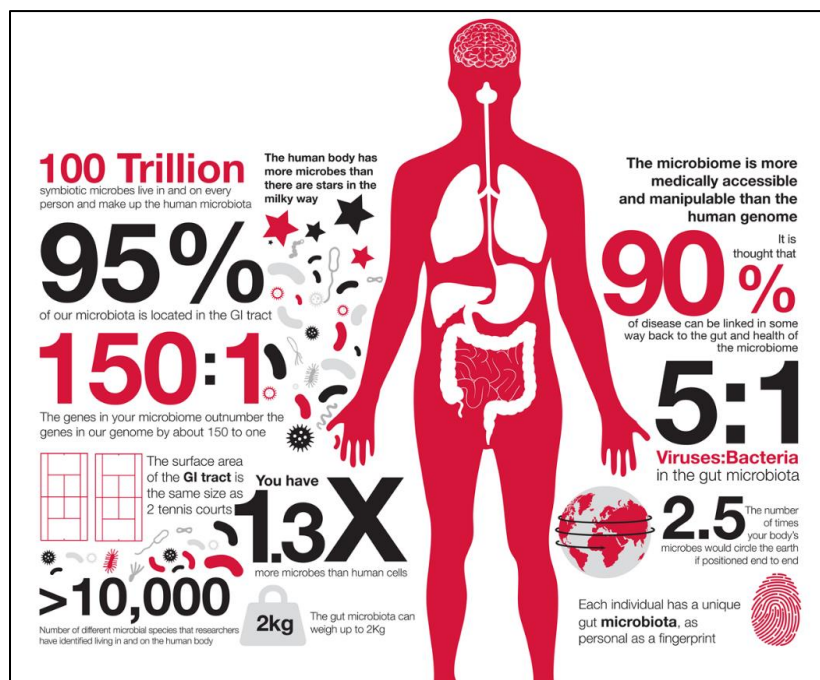


FIGURE 1: Microbiota in number. The human microbiota community comprehend of a vast array of microbes including bacteria, archaea, fungi and viruses. The collective genomes of these resident microorganisms is known as the human microbiome. The interaction between the microbes and the human host play a crucial role in the individual health. Figure from <http://worldmicrobiomeday.com/human-microbiome/>.

Pathogenic bacteria are responsible for a wide range of infections spread worldwide with a few very common among humans. Colonization is the first step of bacterial dissemination within the host and pathogens employ different strategies to adhere to the epithelial surface and proliferate at the site, despite the host defense mechanisms [e.g. mucins, α -defensins, regenerating islet-derived protein 3g, cryptdins, immunoglobulin (Ig)] (Figure 2) (RIBET; COSSART, 2015).

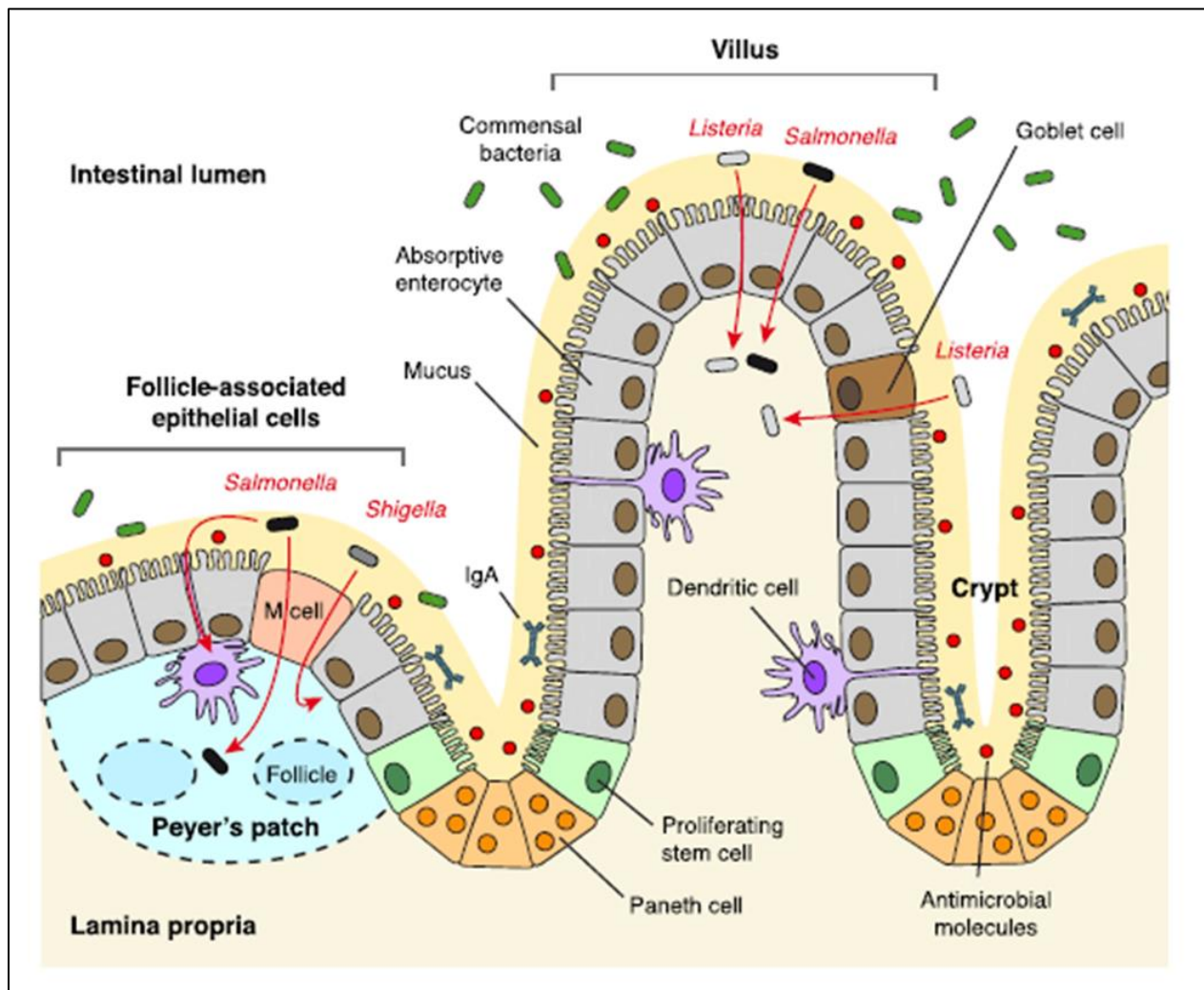


FIGURE 2: Invasion of pathogens in the GIT. Enteric pathogens can invade the intestinal epithelium via the tight junctions between goblet and absorptive epithelial cells (villus region), M cells of Peyer's patches (crypts region) or dendritic cells (intestinal lumen region). Goblet cells is responsible for the production of the mucus layer, which contains secreted IgA, antimicrobial peptides and other types of antimicrobial compounds that limit the colonization by commensal bacteria or foodborne pathogens. Figure from (RIBET; COSSART, 2015).

The adhesion may be coupled with the formation of biofilms (i.e. matrix-embedded bacterial aggregates that benefit bacterial survival and growth in a hostile environment) and internalization of a certain pathogens inside host cells. An intracellular location aid the pathogen to escape the immune system, evade stress conditions provoked by an extracellular routine, acquire access to nutrients not available otherwise, and facilitate the propagation of the bacteria to deeper tissues (RIBET; COSSART, 2015). Furthermore, pathologies and infections [e.g. obesity, diabetic, bowel inflections, *Clostridioides difficile* infections (CDI)] combined with diverse environmental factors (e.g. age, sex, geographical area, diet, stress, drugs) have been associated to changes in the microbiota community since commensal bacteria play a crucial role in the regulation of host immune response, epithelial metabolism and pathogen resistance through competition (Figure 3) (RAWLS *et al.*, 2006; TURNBAUGH *et al.*, 2009).

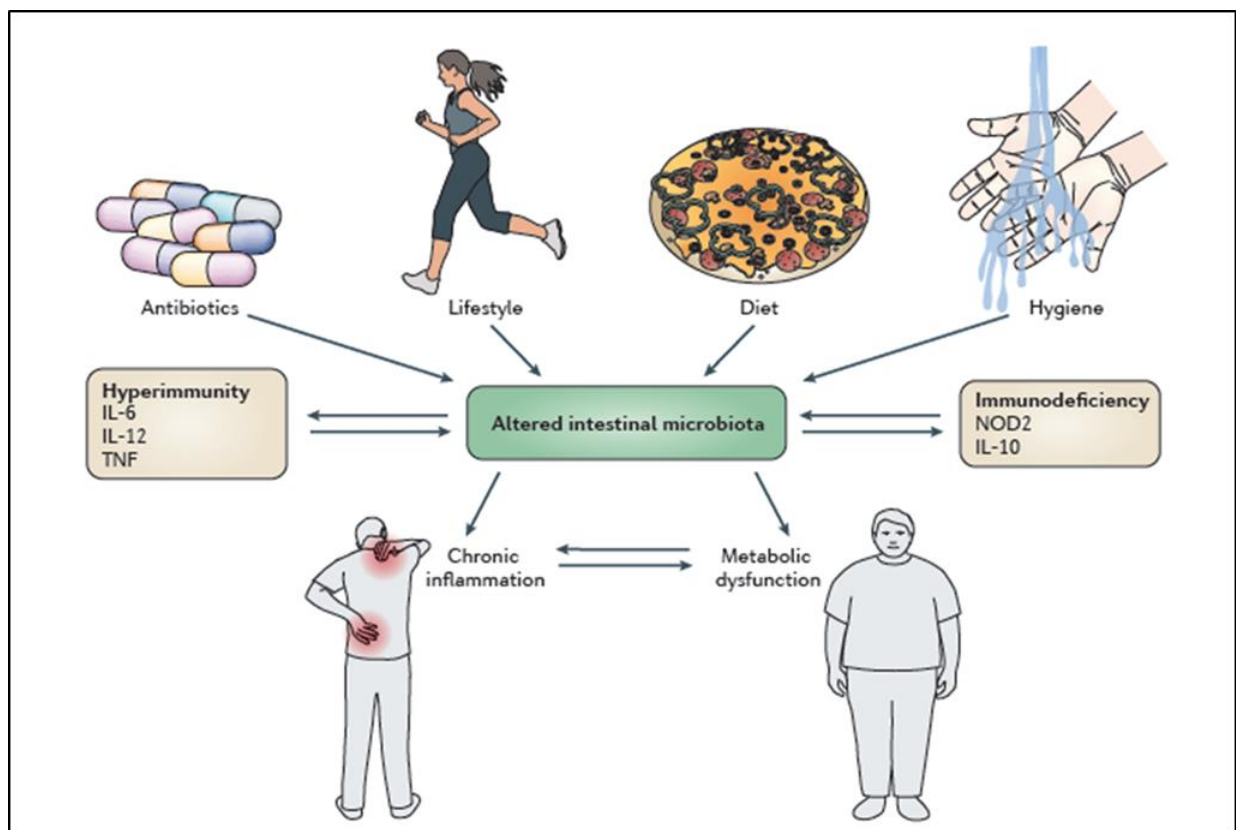


FIGURE 3: Factors that affect the gut microbiota and some consequences of the dysbiosis. Alteration in the intestinal microbiota can be attributed to the use of antibiotics, lifestyle, diet, hygiene, and can cause a number of body dysfunctions (e.g. chronic inflammation, metabolic dysfunctions, and intestinal inflammations). Figure from (TURNBAUGH *et al.*, 2009)

Taken together, the overall crosstalk and interactions between commensals, pathogens and the host are crucial in the development of a variety of life-threatening infections, like (i) pneumococcal; (ii) staphylococcal; (iii) streptococcal; (iv) enterococcal; (v) listeriosis, (vi) botulism; (vii) tuberculosis; (viii) tetanus; (ix) diphtheria and other corynebacterial; and (x) CDI and other clostridial (RIBET; COSSART, 2015), despite the actions of the immune system.

2. The interplay between microbiota, pathogens and immune system

Dysbiosis favor the growth of pathogens and the inflammation caused by overgrowth of ‘bad’ bacteria compromises the microbiota ability to provide colonization resistance against invading microorganisms. Hence, the immune system–microbiome crosstalk is direct and indirect linked with dysbiosis-driven diseases (LEVY *et al.*, 2017). The understanding of these cross-regulation effects is crucial to comprehend a large number of mechanisms by which the microbiota affects both innate and adaptive immune cells, and how the immune system is able to modulate and contain the commensal bacteria, during eubiosis or in the beginning of the inflammatory process (Figure 4) (LEVY *et al.*, 2017; SONNENBERG *et al.*, 2012; THAISS *et al.*, 2016; VAISHNAVA *et al.*, 2011). Nonetheless, the extent of the interplay goes far beyond the achievement of a careful balance between tolerance to commensal microorganisms and immunity to pathogens. The microbiota integrates into whole-organism physiology and influences multiple facets of the organism homeostasis through a large repertoire of signals and mechanisms by which affects the immune activation, including transcriptional reprogramming, epigenetic remodeling and hierarchical feedback loops (THAISS *et al.*, 2016).

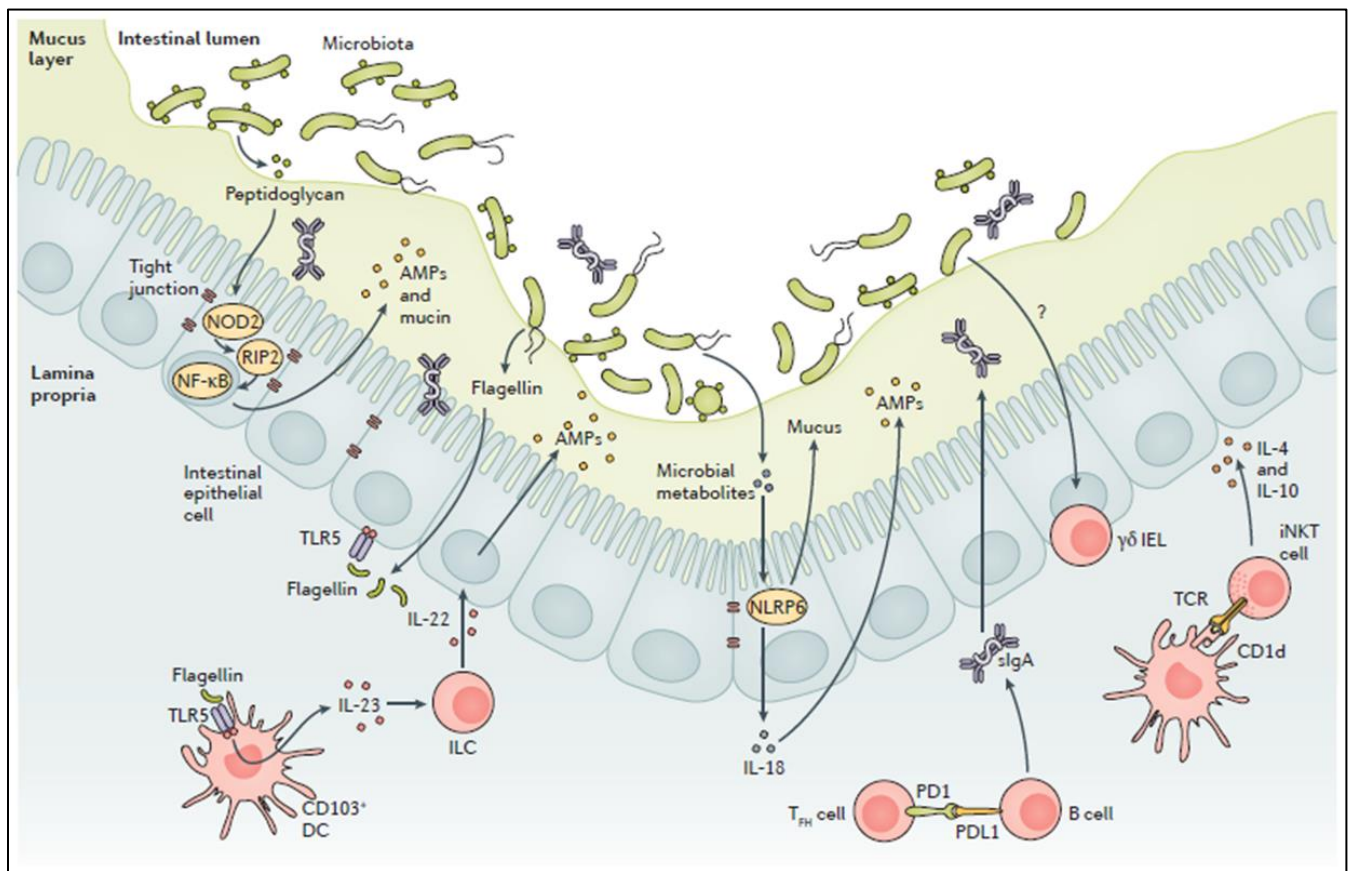


FIGURE 4: Innate and adaptive immunity in the regulation of microbial homeostasis. Innate mechanisms of regulation include pattern recognition receptors (PRRs), recognition of microbial peptidoglycan that contributes to intestinal homeostasis, and induction of the production of antimicrobial peptides (AMPs) and mucin. Adaptive mechanisms of microbial regulation include the production of secretory IgA (sIgA), activation of invariant natural killer T (iNKT) cells and secretion of anti-inflammatory cytokines. IL: interleukin; ILC: innate lymphoid cell; PD1: programmed cell death protein 1; PDL1: PD1 ligand 1; TCR: T cell receptor. Figure from (LEVY *et al.*, 2017).

Summing up, the microbiota can regulate the expression of genes involved in host nutrient absorption and processing, barrier functions, gut motility, intestinal immune responses, angiogenesis and the metabolism of xenobiotics (e.g. antibiotics) (LEVY *et al.*, 2017). These transcriptional reprogramming is partially dependent on microbial sensing receptors of the innate immune system in an organ-specific manner, and might be achieved by differential expression of specific transcription factors and their binding to chromatin. Furthermore, epigenetics phenomena, like methylation and histone acetylation, have shown to influence gene expression and the immune response (ALENGHAT *et al.*, 2013; TAKAHASHI *et al.*, 2011). No signal was uncover so far but microbial metabolites are most likely involved in the orchestration

of histone modifications (e.g. butyrate) (CHANG *et al.*, 2014; THAISS *et al.*, 2016). Meanwhile, feedback loops consist of a brief circuit that links microbial sensing (microbe recognition) with transcriptional reprogramming (transcriptional response) and antimicrobial responses (secretion of effector molecules) within a region of the host (THAISS *et al.*, 2016). The loops are often restricted to the epithelium, which is directly exposed to the microbiota, but the inflammatory response can be extend into the lamina propria, lymphatic and portal circulation.

The innate and adaptive immune system are responsible for the recognition of pathogens linked with infectious diseases. The immune system sense and deploy specialized cells to fight the infection and clear the host from foreign microorganism (Figure 5).

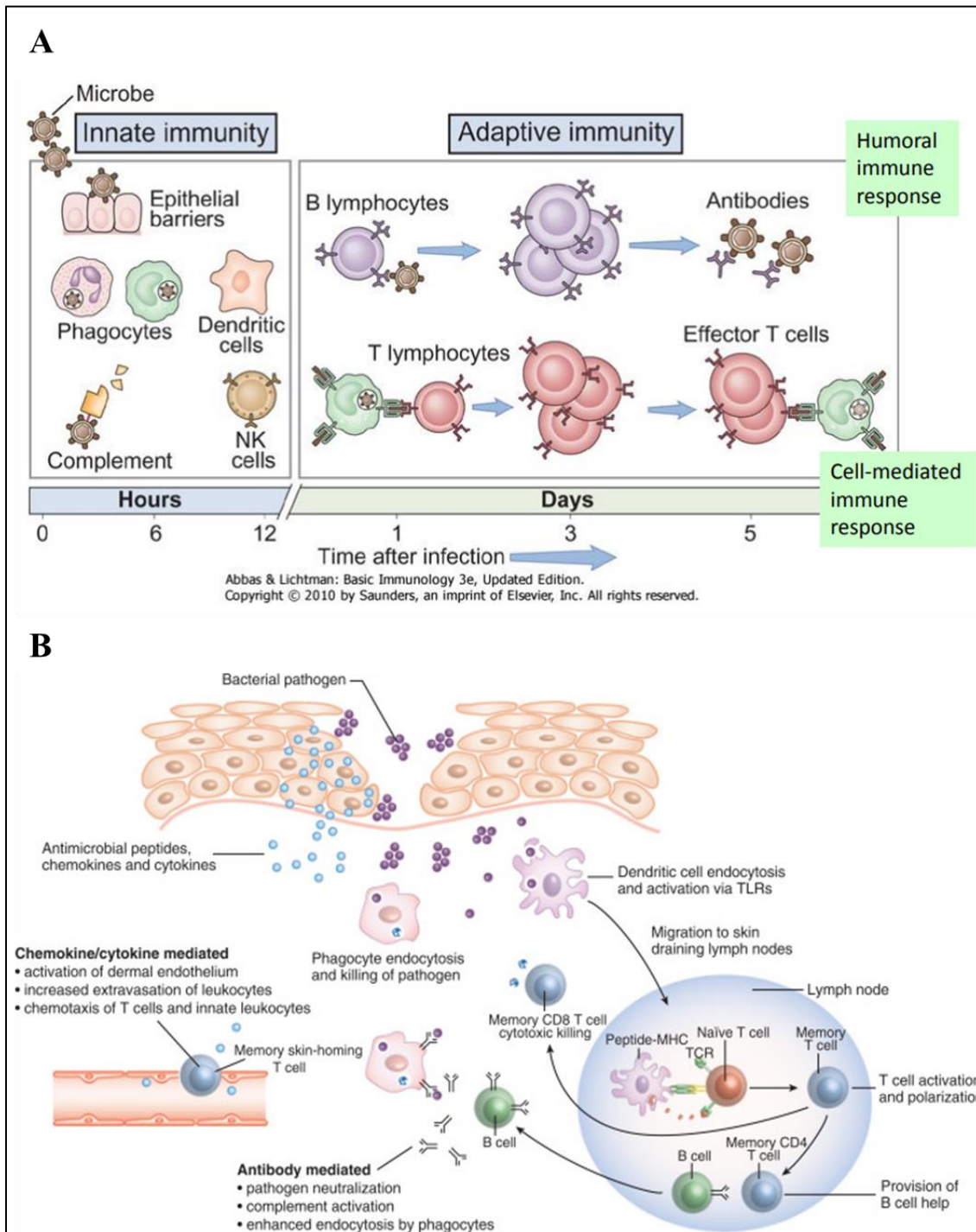


FIGURE 5: The innate and adaptive immune system. The mechanisms of innate immunity provide the initial defense against infections whereas the adaptive immune responses is develop later and require the activation of lymphocytes. (A) The estimated innate and adaptive immunity time line kinetics of the innate (hours) and adaptive (days) immune responses starts with the barrier breach and climax with the humoral and cellular-mediated response. (B) The schematic of the immune response after an epithelial barrier breach. First, the host synthesize anti-microbial peptides, chemokine, and cytokines, which leads to the migration of innate leukocytes and memory T cells into the

site of infection (i.e. skin). The host molecules and the bacterial antigens activate the innate immune cells to phagocytes/kill the pathogens and migrate to the lymph nodes. In the lymph nodes, dendritic cells present bacterial antigens to naïve and central memory T cells, leading to stimulation of pathogen-specific cells. Effector CD8+ cells direct to the inflamed site (skin) to kill the pathogens (cellular immunity) and helper CD4+ T cells activate B cells, inducing the production of antibodies (humoral immunity). Antibody-directed phagocytosis by innate cells leads to enhanced antigen presentation and further boost the acquired responses. Figure adapted from (ABBAS; LICHTMAN; PILLAI, 2015; CLARK; KUPPER, 2005).

2.1. Innate immune response

The cells of the innate immune system are located strategically at the host–microbiome boundary. These cells have the ability to regulate microbial ecology, sense microorganisms or their metabolic products [e.g. as tryptophan and short-chain fatty acids (SCFAs)], and can translate the signals into host physiological responses. Conversely, the innate immune system plays an important role in shaping the community and ecology of indigenous microorganisms into configurations that can be tolerated by the host and are beneficial for its metabolic activities (LEVY *et al.*, 2017).

The major functions of the innate immune system include: (i) acting as a physical and chemical barrier to infectious agents; (ii) recruiting immune cells to sites of infection through the production of chemical factors [e.g. cytokines like tumor necrosis factor (TNF) and interleukin (IL)], which leads to an inflammatory state; (iii) activation of the complement cascade to identify pathogens; (iv) identification and removal of foreign substances by specialized white blood cells [e.g. natural killer, phagocytes, and innate lymphoid cells (ILCs)]; (v) activation of the adaptive immune system through antigen presentation.

The innate immune cells comprise of different surveillance system layers to prevent the dissemination of pathogens (Figure 5). Specialized mononuclear (supra-epithelial) phagocytes provide surveillance in certain mucosal surfaces above the epithelial layer. In the airways, alveolar macrophages are situated inside the lumen for access to airborne pathogens, whereas in the intestine and respiratory tracts, dendritic cells (DC) form tight junctions with the epithelial cells and extend dendrites into the lumen (IWASAKI; MEDZHITOV, 2015; RESCIGNO *et al.*, 2001). Epithelial cells are equipped with pattern recognition receptors (PRRs) that induce chemokines able to recruit circulating leukocytes to initiate an innate defense. The epithelial cells are often the target of replication by a variety of pathogens and the recognition

of pathogens inform the host of a breach in the first barrier (HONDA; LITTMAN, 2016; IWASAKI; MEDZHITOV, 2015; LEVY *et al.*, 2017; THAISS *et al.*, 2016).

The inflammation process began with the recognition of the pathogens by innate immune cells (e.g. resident macrophages and DC) followed by their activation and the release of inflammatory mediators. These cells have PRRs that recognize molecules (antigens) shared by pathogens, collectively referred as pathogen-associated molecular patterns (PAMPs). Several families of PRRs and their signaling pathways are now known, including the Toll-like receptors (TLRs), the nucleotide-binding oligomerization domain (NOD)-like receptors, and the C-type lectin receptors (e.g. Mincle). These sensors are expressed by a variety of cellular compartments and constitute a continuous surveillance system for the presence of microorganisms at the host (LEVY *et al.*, 2017; THAISS *et al.*, 2016).

2.2. Adaptive immune response

The innate immune system methods of pathogen recognition deliver vital information and instruct the adaptive immune system (acquired system) to activate relevant effector responses (Figure 5). The major functions of the acquired immune system include: (i) recognition of foreign antigens during the process of antigen presentation; (ii) generation of responses tailored to eliminate specific pathogens or pathogen-infected cells; (iii) development of immunological memory through memory B and T cells; (iv) maintenance of immune homeostasis by suppressing responses to harmless antigens and by enforcing the integrity of the barrier functions of the gut mucosa (HONDA; LITTMAN, 2016; IWASAKI; MEDZHITOV, 2015; LEVY *et al.*, 2017)

The recognition of PAMPs by PRR, which is expressed on antigen-presenting cells (APC) (e.g. DC, B cells and macrophages), instruct lymphocytes to induce the appropriate effector class of the immune response. The recognition is done according with the type of infection encountered and the origin of antigen receptors expressed on T cells (T cell receptor or TCR) or B cells (B cell receptor or BCR) (HONDA; LITTMAN, 2016; IWASAKI; MEDZHITOV, 2015). The two main lymphocytes broad classes, B cells and T cells, are responsible for the humoral and cell-mediated immunity response, respectively, whereas APCs display the antigen through the major histocompatibility complex (MHC)

molecules type I or II (NEEFJES *et al.*, 2011). Succinctly, MHC molecules exhibit a large spectrum of antigenic peptides that first goes through an intracellular degradation to be then displayed on the cell surface for T cells recognition (Figure 6) (KAUFMANN; SCHAIBLE, 2005).

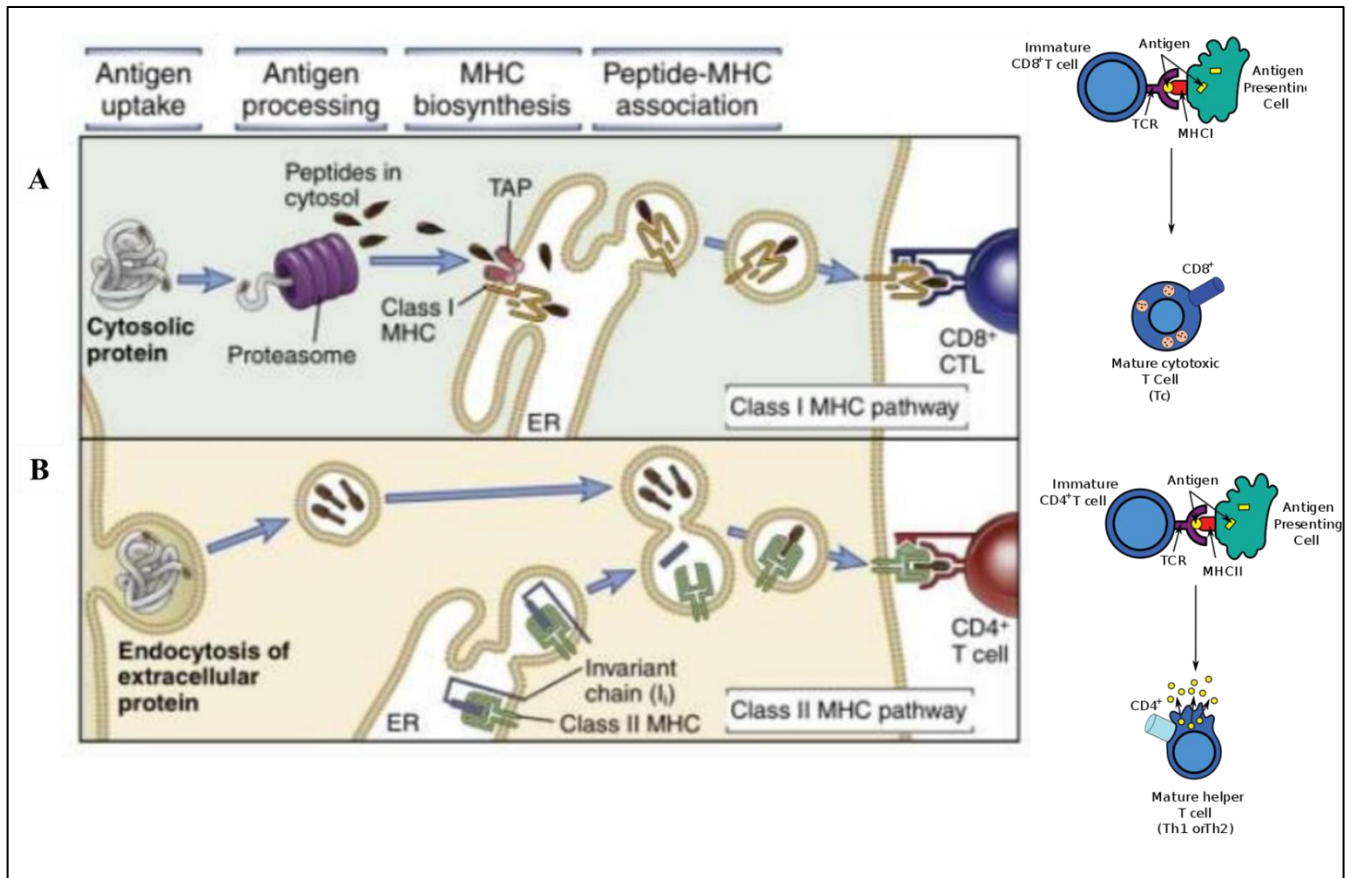


FIGURE 6: Antigen processing pathways for the MHC type I and II molecules and presentation to naïve T cells. (A) MHC-I molecules display peptides that are primarily derived from pathogenic proteins. These proteins are degraded into peptides by the proteasome and then transported through the transporters of antigen-processing (TAP) molecules into the endoplasmic reticulum (ER) to be loaded on MHC-I molecules and further presented to CD8+ T cells. The antigen presenting cell (APC) display the epitope-MHC-I complex in the cell surface to be recognized by the TCR of the immature or naïve CD8+ T cell to activate the cytotoxic T lymphocyte (mature cytotoxic T cell). (B) The protein peptides presented by MHC-II molecules enter the cell through the endocytic route. During maturation, MHC-II is association with the invariant chain (Ii) and do not bind to endogenous antigens in the ER. The MHC II-Ii complexes move through the Golgi where the invariant chain is degraded and removed from the MHC molecule. Afterwards, the epitopes are linked to the MHC-II and the complex can be presented to CD4+ T cells. Likewise the MHC-I pathway, the epitope-MHC-II complex displayed by the APC identify the TCR from the immature CD4+ T cell and activates the Th1 and/or Th2 pathway. Figure adapted from (ABBAS; LICHTMAN; PILLAI, 2014).

The epitopes or antigenic determinants (i.e. immunologically active regions from pathogens) displayed by MHC-I are generated via endogenous intracellular digestion (e.g. proteasomes and other cytoplasmic proteases). The digested determinates are then transported into the cisternae of the endoplasmic reticulum (ER) via the transporter of antigen processing machinery to be loaded into the MHC-I molecules located within the phagosome (Figure 6A) (KAUFMANN; SCHAIBLE, 2005; NEEFJES *et al.*, 2011). The complex epitope-MHC-I are featured on the surface of nearly all nucleated cells to be recognized by the TCR of naïve CD8⁺ T cells (cytotoxic T cells or T_C). This interaction elicits the cellular immune response through the activation of cytotoxic T-lymphocytes (CTL) that goes through clonal selection process (i.e. production of effector cells) to search and destroy, via apoptosis, infected host cells throughout the body (Figure 7) (KAUFMANN; SCHAIBLE, 2005).

The MHC-II molecules are displayed on the surface of APCs to be recognized by the TCR of naïve CD4⁺ T cells. After the pathogen is internalized by the APCs, the antigen are processed into small molecules (i.e. epitopes) via the endocytic processing pathway and then loaded into the MHC-II in the ER-derived vacuoles (Figure 6B) (NEEFJES *et al.*, 2011; ROCHE; FURUTA, 2015). The epitope-MHC-II complex trigger the activation of T helper (T_H) cells involved in Th1 and Th2-driven immune response (Figure 7). Th1 pathway activate cellular immunity against intracellular pathogens via the release of cytokines (e.g. interferon gamma or IFN- γ) that stimulate the activation of CTL cells (pro-inflammatory response) whereas Th2 pathway trigger the humoral immunity against extracellular pathogens through the activation of B cells into antibody-secreting plasma cells (SPELLBERG; EDWARDS, 2001).

The naïve B cells can also directly recognize the antigen through BCR activating the clonal selection process and its progeny differentiate into memory B cells and effector B cells or plasma cells (Figure 7). The antibody/Ig released by the plasma cells (e.g. IgA, IgD, IgE, IgG, and IgM) can identify and neutralize foreign pathogens by binding to their specific antigens, which lead to agglutination and precipitation of the antibody-antigen complex, blockage of viral receptors, and stimulation of other immune responses, such as the complement pathway (LINDNER *et al.*, 2015).

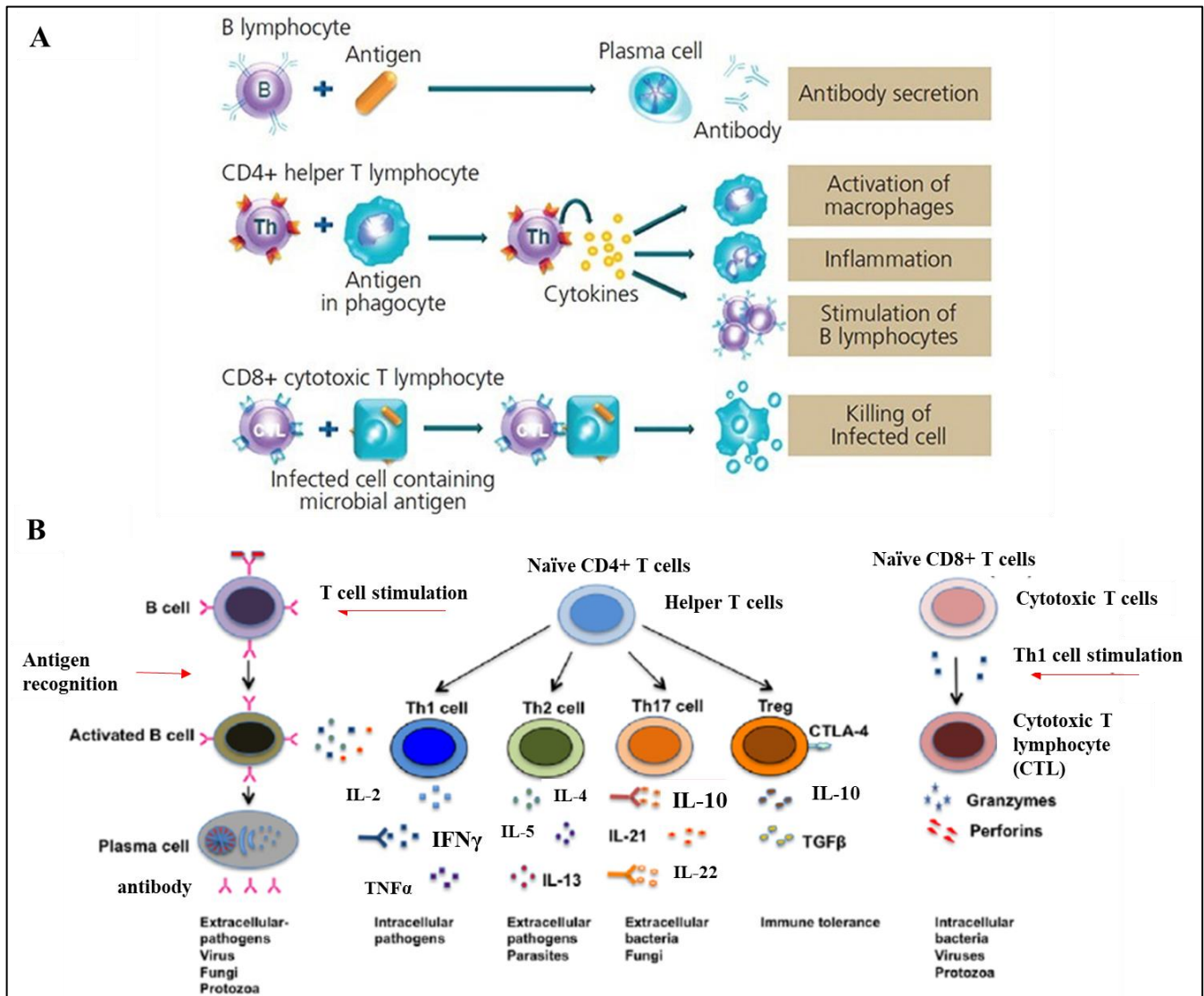


FIGURE 7: Adaptive immune response. Antigens from invading pathogens are recognized and presented by innate immune cells such as macrophages and dendritic cells to CD4 + and CD8 + naïve T cells and to B cells eliciting the adaptive immune response. (A) The lymphocytes bind to the pathogenic antigens and deliver a response to combat the infection. B cells are involved in antibody secretion whereas T cells are implicated in activation of macrophages, inflammation and stimulation of B lymphocytes (CD4+ T cells), and also in killing of infected cells (CD8+ T cells). (B) After antigen recognition, naïve B and T cells are activated. B cells are converted into plasma cells responsible for antibody secretion. CD4+ T cells or helper T cells can be polarized into Th1, Th2, Th17, and regulatory T cells (Tregs) response that secrete distinct cytokines. Furthermore, CD4+ T cells provide help to B cells to produce antigen-specific antibodies and to CD8+ T cells to yield cytotoxic T lymphocytes (CTL). CD8+ T cells or cytotoxic T cells are activated into CTL cells that are linked in apoptosis of infected cells via granzymes and perforins. The black arrows indicate antigen recognition. Figure adapted from (KUMAR, V.; ABBAS; ASTER, 2014; MADDUR et al., 2010).

The adaptive immune system is directly linked with the microbiota composition and consequently with the modulation of the host homeostasis. IgA produced by the gut plasma cells increases the diversity of the microbial community and enhances mutualism between the microbiota and the host, thus a balanced microbiota can restrain inflammatory processes (HONDA; LITTMAN, 2016; KAWAMOTO *et al.*, 2014; LINDNER *et al.*, 2015). The Th17 cells denote an immune pathway that stimulates intestinal epithelial cells to produce antimicrobial proteins being involved in the formation of tight junctions and in the prevention of extracellular pathogenic bacteria infections whereas also express pro-inflammatory cytokines, like IFN- γ , linked to exacerbate autoimmune and inflammatory diseases (HIROTA *et al.*, 2011; HONDA; LITTMAN, 2016; WEAVER *et al.*, 2013).

Altogether, the observations in both innate and adaptive immunity provide evidence for the intriguing notion that the microorganisms shape the very same mechanisms that organize their colonization settings (e.g. antimicrobial peptides, mucus and secretory IgA) (IWASAKI; MEDZHITOV, 2015). Furthermore, the immunological memory provided by the encounter of foreign pathogens and the immune system enhance the response to subsequent confrontations with that pathogen, a process similar to the vaccination proposition.

3. Vaccines

Vaccine is a biological preparation that improves the host immunity against certain microorganism and offers the best cost benefit ratio in the prevention of diseases. The introduction of human vaccines had a tremendous impact on global health by inhibiting the spread of a variety of infections and protecting infants, children and adults from disease that otherwise could lead to certain death (BLOOM, 2011; EHRETH, 2003; KALLERUP; FOGED, 2015). This pharmaceutical product stimulates the host immune system to recognize the pathogen agent as foreign, so in a later encounter would be easily identified and destroyed (Figure 8).

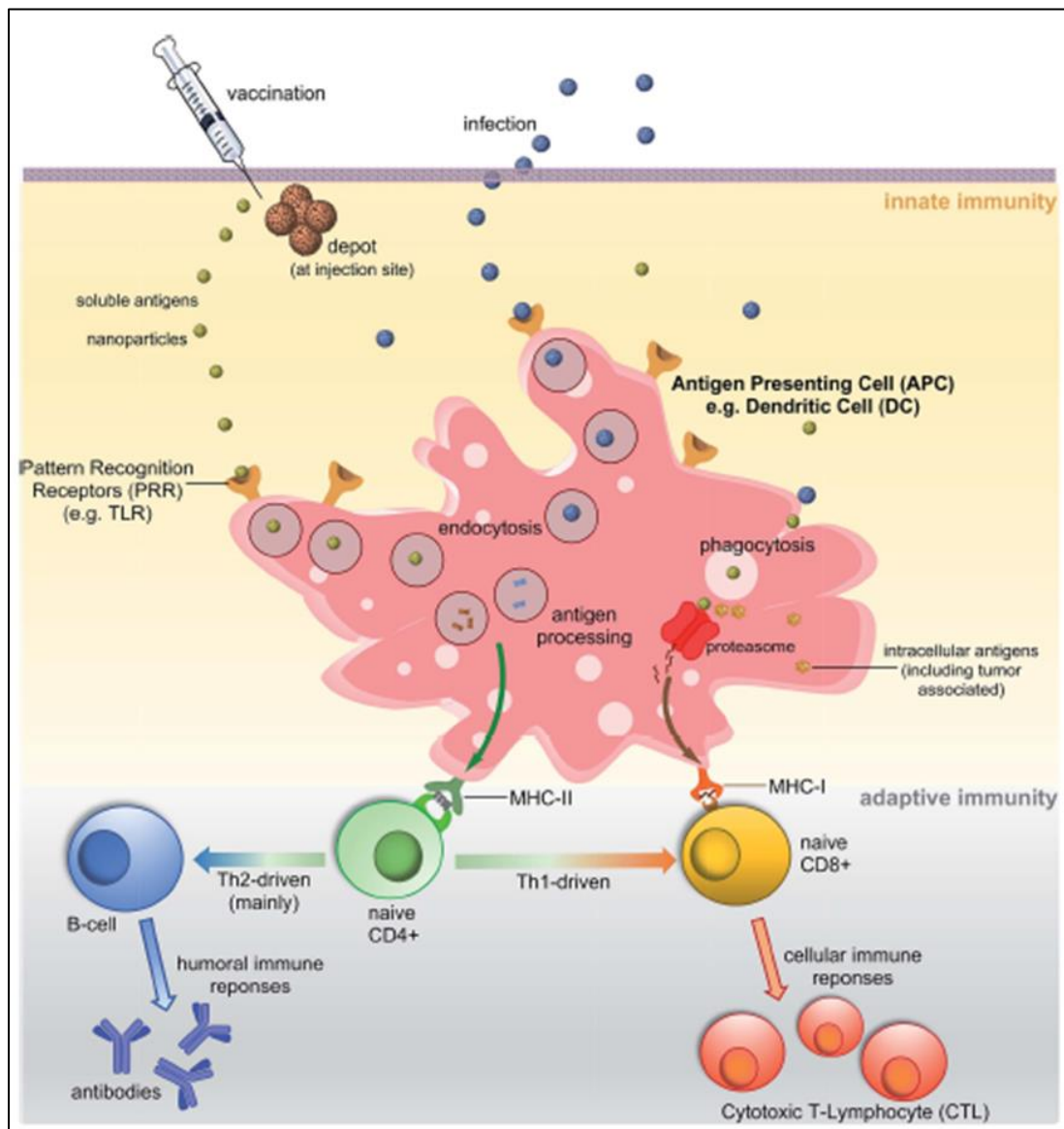


FIGURE 8: Immune response pathways triggered by vaccination. The vaccine hold bacterial antigens capable to stimulate the immune system triggering a cascade of responses. The innate immune system [e.g. antigen presenting cells (APCs) such as dendritic cells (DCs) or macrophages] recognizes pathogen associated molecular patterns (PAMPs) via pattern recognition receptors (PRRs) such as toll-like receptors (TLRs). Once recognized, antigens are processed into small molecules and loaded on MHC-I or MHC-II proteins prompting the adaptive immune response (cellular and/or humoral). Figure from (SKWARCZYNSKI; TOTH, 2016).

Vaccination is a successful and cost effective medical intervention systematic introduced after World War II as a widespread tool for improving public health (BLOOM, 2011). Yet, the technique have been applied for more than a thousand years in China and India, where they used to inoculate individuals with live and

virulent smallpox virus for protection, although the pioneer honored developer is Edward Jenner by demonstrating the induced cross-protective immunity towards smallpox when the virus were exposed to humans (RIEDEL, 2005). The vaccination principle is based on Louis Pasteur’s foundation of isolating, purification, and injection the causative microorganisms in order to induce protective immunity (RAPPUOLI, 2007). Nowadays, the World Health Organization (WHO) coordinate international health within the United Nations system that includes the developing of evidence-based immunization policy recommendations. A WHO independent advisory group (Strategic Advisory Group of Experts or SAGE) coordinates the Expanded Programme on Immunization (EPI) that establishes the vaccine-preventable diseases and vaccination programs for all ages, anywhere in the world. The recommendations are published on ‘WHO position papers’ that summarizes the essential background information on the respective diseases and vaccines, including the ‘Recommendations for routine immunization’ (Table 1).

Antigen		Children	Adolescents	Adults
Recommendations for all immunization programmes				
BCG		1 dose		
Hepatitis B		3-4-doses	3 doses (for high-risk groups if not previously immunized)	
Polio (IPV)		3-4 doses (at least one dose of IPV with DTP)		
Diphtheria-tetanus-pertussis (DTP)		3 doses: 2-6 months of age, plus 2 boosters: 12-23 months (DTP) and 4-7 years (DT)	1 booster 9-15 years (DT)	
<i>Haemophilus influenzae</i> type B	Option 1	3 doses, with DTP		
	Option 2	2 or 3 doses, with booster at least 6 months after last dose		
Pneumococcal (Conjugate)	Option 1	3 doses with DTP		
	Option 2	2 doses before 6 months of age, plus booster dose at 9-15 months of age with DTP		
Rotavirus		2-3 doses depending on product with DTP		
Measles		2 doses		

Rubella	1 dose	1 dose (adolescent girls and women of child-bearing age if not previously vaccinated)	
HPV		2 doses (female)	
Recommendations for certain regions			
Japanese Encephalitis	Inactivated Vero cell-derived vaccine: Live attenuated vaccine: 1 dose plus Live recombinant vaccine: 1 dose		
Yellow Fever	1 dose, with measles containing vaccine		
Tick-Borne Encephalitis	3 doses (> 1 year FSME-Immun and Encepur; > 3 years TBE-Moscow and EnceVir) with at least 1 booster dose (every 3 years for TBE-Moscow and EnceVir)		
Antigen	Children	Adolescents	Adults
Recommendations for some high-risk populations			
Typhoid	Typhoid conjugate vaccine (Typbar-TCV®): 1 dose; Vi polysaccharide (ViPS): 1 dose; Ty21a live oral vaccine: 3-4 doses; Revaccination for ViPS & Ty21a; every 3-7 years		
Cholera	Dukoral (WC-rBS): 3 doses ≥ 2-5 years, booster every 6 months; 2 doses adults/children ≥ 6 years, booster every 2 year; Shanchol, Euvchol & mORCVAX: 2 doses ≥ 1 years, booster dose after 2 years		
Meningococcal	MenA conjugate	1 dose 9-18 months (5µg)	
	MenC conjugate	2 doses (2-11 months) with booster 1 year after 1 dose (≥12 months)	
	Quadrivalent conjugate	2 doses (9-23 months) 1 dose (≥2 years)	
Hepatitis A	At least 1 dose ≥ 1 year of age		
Rabies	2 doses		
Dengue (CYD-TDV)	3 doses 9-45 years of age		
Recommendations for immunization programmes with certain characteristics			
Mumps	2 doses, with measles containing vaccine		
Seasonal influenza (inactivated tri- and quadrivalent)	First vaccine use: 2 doses Revaccinate annually: 1 dose only	Priority for pregnant women 1 dose ≥ 9 years of age Revaccinate annually	
Varicella	1-2 doses	2 doses	

TABLE 1: ‘WHO position paper’ – Recommendations for routine Immunization. The number of vaccine doses and buster doses according with the recommended age for each disease. For more information access: <http://www.who.int/immunization/documents/positionpapers/>.

After vaccine was introduced in a global setting, immunization has achieved greatest triumphs, including the eradication of smallpox, decline of global incidence of polio by ninety-nine percent, and reduced illness, disability and death from diphtheria, tetanus, whooping cough, measles, *Haemophilus influenzae* type B disease, and epidemic meningococcal A meningitis. In 2017, about 85% of infants worldwide (116.2 million infants) received three doses of diphtheria-tetanus-pertussis (DTP) vaccine, protecting them against infectious diseases that can cause serious illness and disabilities or even be fatal (WHO, 2019).

Despite the overwhelming evidence demonstrating the benefits of immunization and the proven efficacy of the practice, vaccination still faces many obstacles. An estimation of 19.9 million infants worldwide were not reached with routine immunization services (DTP vaccine) wherein 60% of the children residing in Afghanistan, Angola, the Democratic Republic of the Congo, Ethiopia, India, Indonesia, Iraq, Nigeria, Pakistan and South Africa (WHO, 2018a, b). According with WHO, 1.5 million deaths could be avoided with an improved global immunization coverage; thus, the Global Vaccine Action Plan (GVAP) was created and endorsed by countries worldwide to strengthen immunization and achieve the goals of GVAP. The immunization program saves an estimated 2-3 million lives annually acting as a central pillar of universal health coverage, and GVAP aims to eradicate a higher number of illness through vaccination. However, the low vaccination coverage may also be due to political reasons or alarm towards the vaccine or its side effects, which leads to the public refuse to vaccinate (BLOOM, 2011; KALLERUP; FOGED, 2015).

The result gathered so far indicate that developing countries display low coverage vaccination numbers with may be linked with the resurgence with major outbreaks of measles and diphtheria and the persistence of poliovirus and maternal and neonatal tetanus (WHO, 2018a). The drawbacks point out the urgency of vaccination campaigns and operational plans to address immunization problems and entail the inevitability to enhance research and development studies to generate new product that might eliminate possible design gaps of the classical vaccines.

3.1. Classical vaccination classification

The classical vaccine typically contains the whole or a portion of the disease-causing agent, and is often encompassed off live attenuated, inactivated or killed forms of the microbe, its inactivated toxins or segments of the pathogen. A few examples are described at Table 2.

Vaccine type	Disease
Live, attenuated	Measles, mumps, rubella (MMR combined vaccine) Varicella (chickenpox) Influenza (nasal spray) Rotavirus Zoster (shingles) Yellow fever
Inactivated/ Killed	Polio (IPV) Hepatitis A Rabies
Subunit/ Conjugate	Hepatitis B Influenza (injection) Haemophilus influenza type b (Hib) Pertussis (part of DTP combined immunization) Pneumococcal Meningococcal Human papillomavirus (HPV)
Toxoid (inactivated toxin)	Diphtheria-tetanus (part of DTP combined immunization)

TABLE 2: Example of vaccines available for a variety of disease according with the development technique. Vaccines can be divided by the method it was designed, which includes: Live attenuated, inactivated or killed, inactivated toxin, and subunit or conjugated (segment of the microorganism). The influenza virus vaccine is produced by two technologies (live attenuated and subunit) but diverge in the mode of delivery (nasal spray or injection). The vaccine can have one (alone) or more (combined) immunization agents [e.g. Measles, mumps and rubella (MMR combine vaccine) and diphtheria, tetanus and pertussis (DTP combine vaccine)]. Table adapted from (PLOTKIN, S. A.; ORENSTEIN; OFFIT, 2012).

3.1.1. Live attenuated vaccines

This vaccine type contain laboratory-weakened pathogen that mimic the natural infection, which results in an effective vaccination strategy with strong cellular and an antibody response. The attenuation is granted trough harsh condition cycles and series of cell cultures or animal embryos (typically chick embryos) that result in a non-pathogenic strain with the original pathogen features. Afterwards, the

microorganism is still able to replicate in human cells but is unable to cause the disease and provoke an immune response that can protect the host against future infection (KALLERUP; FOGED, 2015; PLOTKIN, S. A.; ORENSTEIN; OFFIT, 2012).

Protection from a live attenuated vaccine typically outlasts the killed or inactivated vaccine; however, the disadvantage of this approach are worse and the adverse effects vary from mild to severe or even fatal in some cases (HUANG; WU; TYRING, 2004; KALLERUP; FOGED, 2015).

3.1.2. Killed or inactivated vaccines

Vaccines of this type are produced by inactivating a pathogen by heat or chemicals such as formaldehyde or formalin, which prevents the pathogen replication without harming the ability to stimulate the immune system. Unlike live vaccines, this approach obviate the possibility of the pathogen revert back its virulence; however, provides a shorter length of protection than live vaccines, and are more likely to require boosters doses to create long-term immunity. Still, the inflammatory response is greater than subunit vaccines since most of the pathogenic components are preserved (KALLERUP; FOGED, 2015; PLOTKIN, S. A.; ORENSTEIN; OFFIT, 2012).

3.1.3. Subunit and Conjugate Vaccines

This type of vaccine contain a piece (one or more components) of the target pathogen to stimulate the immune system. The subunit vaccine can be divided into toxoids, capsular polysaccharides of encapsulated bacteria and proteins or peptide fragments. The components covers recombinant peptides/proteins or polysaccharides of the microorganism, which makes the vaccine safe (highly pure components with a lack of replication) and cost-effective (DUDEK *et al.*, 2010; FOGED, 2011; KALLERUP; FOGED, 2015; PLOTKIN, S. A.; ORENSTEIN; OFFIT, 2012). Meanwhile, conjugate vaccines are constructed using a combination of two different components, a bacteria target chemically linked to a carrier protein. The target protein alone do not stimulate the immune system or cause illness but combined prompt a more powerful immune response (KALLERUP; FOGED, 2015).

These highly purified antigens lack the intrinsic pathogenic features that render them weakly immunogenic, which makes necessary the presence of adjuvants to induce a compelling immune response (KALLERUP; FOGED, 2015; REED; ORR; FOX, 2013).

3.1.4. Inactivated toxin (toxoids)

Some bacterial diseases are not directly caused by the bacterium but by their toxin that can be considered the main disease causing agents. The toxins are inactivated by chemical treatment (e.g. formaldehyde) or by heat and afterwards the detoxified versions are called toxoids. Toxoids can be safely used as a vaccine and incite the immune system to fight the infection via anti-toxoid antibodies (KALLERUP; FOGED, 2015; PLOTKIN, S. A.; ORENSTEIN; OFFIT, 2012).

3.2. Vaccine adjuvant and delivery system

Adjuvants (from Latin *adjuvare* or 'to help') are normally required to enhance the vaccine ability to induce protection against infections (i.e. mimic PAMPs), boosting the innate immune response, which results in improved adaptive immunity with higher activation of T and B cells. The presence of the adjuvants also assure the reduction of the antigen amount and booster doses needed for a successful immunization (Figure 9) (ALSHANQITI *et al.*, 2017; PASQUALE *et al.*, 2015). In addition, certain adjuvants can be used to promote antibody responses at mucosal surfaces and promote low antigen release and protection from rapid elimination (i.e. emulsions). The combined features of dose reduction and antigen sparing can have important implications for global vaccination. However, adjuvants can cause more local reactions (such as redness, swelling, and pain at the injection site) and systemic reactions (such as fever, chills and body aches) than non-adjuvanted vaccines (PASQUALE *et al.*, 2015). Bearing these, the adjuvants major properties and functions to be evaluated include: (i) the presentation of the antigen, defined by the physical appearance of the antigen in the vaccine; (ii) the antigen/adjuvant uptake; (iii) the antigen distribution/delivery (targeting to specific cells); (iv) the immune potentiation/modulation aspects of the immune responses; (v) the antigen protection from degradation and elimination.

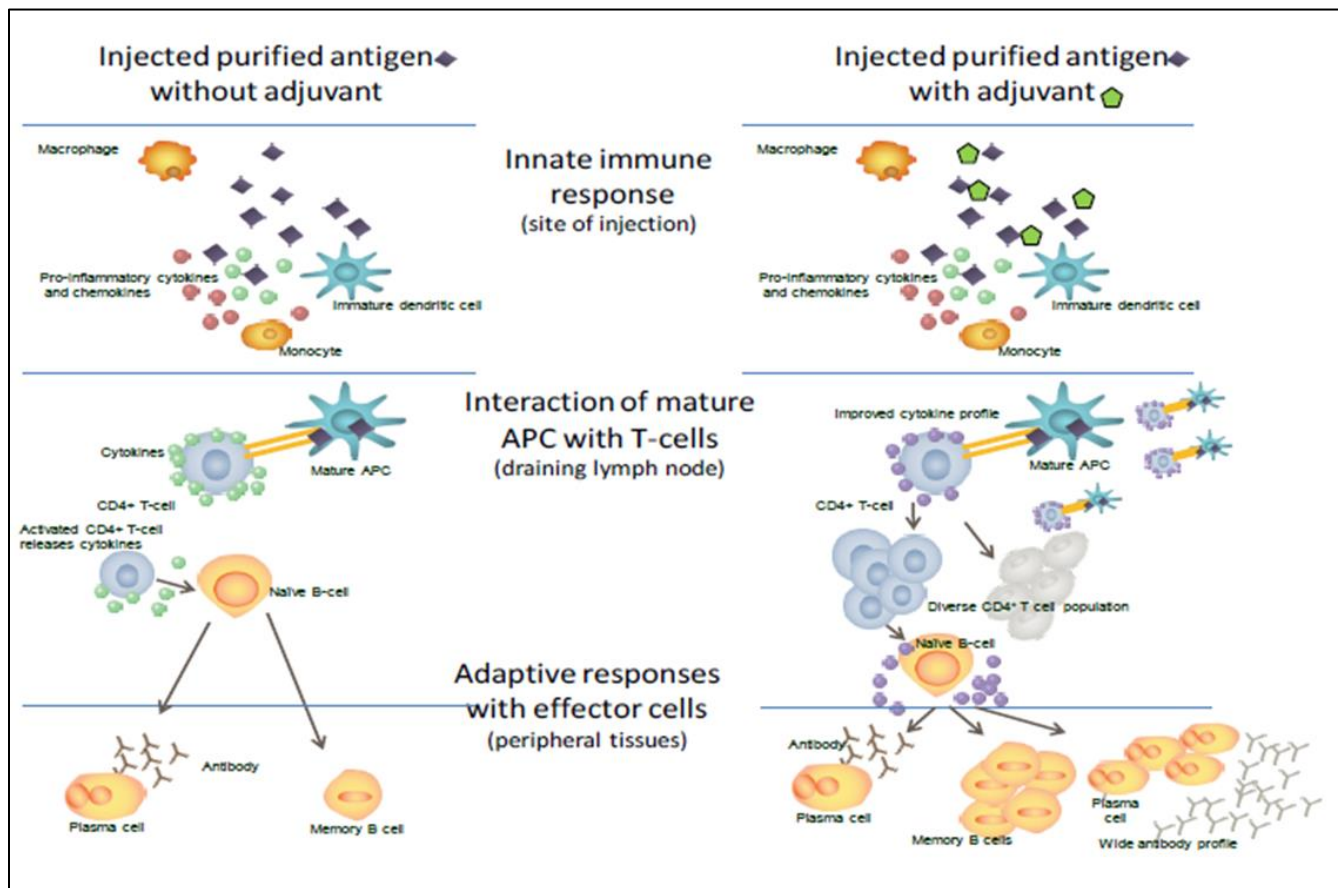


FIGURE 9: The immune response to vaccination with and without adjuvant. The presence of the adjuvant enhance the innate and adaptive immune response especially considering the lack of PAMPs in subunit and next-generation vaccines. The adjuvant interact with the antigen present cells (APCs) via pattern recognition receptors (PRR) like toll-like receptors (TLRs). Figure from (PASQUALE et al., 2015).

Adjuvants can be classified according to their source (natural, synthetic or endogenous), mechanism of action, and physical or chemical properties; and they should be chosen based on the type of immune response desired (Table 3). Furthermore, adjuvant and delivery system are no longer mutually exclusive and both can potentially stimulate the immune response while protecting the antigen from degradation and transporting it to the desired tissue. Delivery systems can contain a built-in adjuvant or even has self-adjuvanting properties, which are relevant when an alternative vaccination rout is adopted (e.g. oral) (SKWARCZYNSKI; TOTH, 2016).

Aluminum was the first adjuvant used in human vaccines and, for approximately 70 years, aluminum salts or alum (e.g. aluminum hydroxide, aluminum phosphate, and aluminum potassium sulfate) were the only licensed adjuvant available (PASQUALE *et al.*, 2015). Alum adjuvants effectively enhance the immune response being optimum for toxoid vaccines; however, the disadvantages include severe local irritation and inflammatory reaction, low induced cellular immunity, and ineffective antiviral immunity (ALSHANQITI *et al.*, 2017). Freund’s incomplete adjuvant (i.e. a mineral oil-in-water emulsion) was an attempt to administer another adjuvant but fail since it was considered too reactive for humans and its use was discontinued (PASQUALE *et al.*, 2015). Nowadays, newer adjuvants have been developed to target specific components of the body’s immune response, comprising more than 30 licensed vaccines, being constantly monitored by the center for disease control (CDC) and food and drug administration in the US (Table3) (PASQUALE *et al.*, 2015). The hepatitis A was the first licensed vaccine to use an adjuvant other than aluminum (i.e. virosome adjuvant system). Virosomes are spherical phospholipid layers carrying bound influenza antigen on the surface or encapsulated within the lumen (MOSER *et al.*, 2013). Further relevant adjuvant types are micro and nanoparticles (e.g. carbon nanotubes and other polymers), liposomes, among others components derived from pathogens (reviewed at Alshanjiti et al., 2017; Pasquale et al., 2015; Skwarczynski and Toth, 2016). These adjuvant types are highly recommended due to its reduced toxicity, biodegradable properties, and strong immune response (SKWARCZYNSKI; TOTH, 2016).

Adjuvant	Composition		Major Immune Effects
	Component	Origin	
Aluminum (DTP, IPV, hepatitis A & B, HPV, meningococcal and pneumococcal)	Aluminum as salts mixed with antigen (adsorption)	Naturally occurring present in soil, water, air	Increases local inflammation, improves antigen uptake by APCs. Acts to increase antibody production
Virosomes (Hepatitis and influenza)	Vesicles where influenza antigens in aqueous volume are enclosed within a standard phospholipid cell membrane bilayer	Natural phospholipids, Seasonal influenza glycoproteins	Increases uptake by APCs. May interact with B cells leading to T cell activation.
AS04 (Hepatitis B, HPV)	(3-deacyl-monophosphoryl lipid A) derived from LPS	Natural exposure to LPS from Gram-negative bacteria occurs frequently	Directly stimulates TLR-4 increasing APC maturation and Th1 responses.

	from <i>Salmonella</i> , Aluminum salts		
MF59® (Influenza-seasonal and pandemic)	Squalene	Animal source (shark liver oil). Found naturally in human tissues: adipose tissues, skin, arterial walls, skeleton, muscles, lymph nodes	Increases APC recruitment and activation. Promotes antigen uptake and migration of cells to lymph nodes.
AS03 (Influenza-pandemic)	Vitamin E (α -Tocopherol) Surfactant polysorbate 80 Squalene	Naturally occurring in humans. Surfactant and emulsifier Animal source (shark liver oil). See above	Promotes local production of cytokines and recruitment of innate cells.
Thermo-reversible oil-in-water (Influenza-pandemic)	Squalene	Animal source (shark liver oil). See above	Not reported
ISA51 (Therapeutic vaccine NSCLC)	Mineral oil DRAKEOL 6 VR Surfactant mannide-mono-oleate	Refined mineral oil of vegetable origin	Strongly immunogenic

TABLE 3: Characteristics of adjuvants used in licensed vaccines. D: diphtheria; T: tetanus; P: pertussis; IPV: inactivated poliomyelitis vaccine; HPV: human papilloma virus; LPS: lipopolysaccharide; APC: antigen presenting cells; TLR: toll-like receptor; NSCLC: non-small cell lung cancer; MPL: monophosphoryl lipid A. Table adapted from (PASQUALE et al., 2015).

Adjuvants have become fundamentally essential for the new vaccine designs, such as subunit and the new-generation, since only part of the pathogen is incorporated into the vaccines, and alone these antigens may not be sufficient to induce adequate immunity and long-term protection, being often considered low immunogenic. This could be attributed mostly due the lack of PAMPs, which means that the initial innate immune response is not activated, which prevents an effective downstream adaptive response (PASQUALE *et al.*, 2015).

3.3.Next-generated vaccines

The prevalence of pathogenic bacteria and the epidemic of a number of infectious disease can be explained by lack of efficacious vaccines where conventional vaccinology has failed due to factors such as antigenic

drift/shift, and by the existence of more difficult target diseases (e.g. when antibodies cannot provide sufficient protection) (RAPPUOLI, 2007). Furthermore, the use of conventional vaccines may provoke complications on vulnerable population groups such as the elderly, children and immune compromised (immunologically hyporesponsive) (KALLERUP; FOGED, 2015). These complications put great emphasis on the acute need for new prophylactic as well as therapeutic vaccines against global pathogens and the design of new vaccines are necessary and became possible as a result of remarkable scientific and technological progress (KALLERUP; FOGED, 2015; SKWARCZYNSKI; TOTH, 2016).

3.3.1. Next-generation vaccines and OMICs approach

Advances in genomic technology open the possibility to decrease the difficulties faced in genes target selection. ‘OMICs’ approach adopt a holistic view of biological samples thought the study of genes (genomics), mRNA (transcriptomics), proteins (proteomics), metabolites (metabolomics), lipids (lipidomics), among other technologies (HORGAN; KENNY, 2011; KEERSMAECKER *et al.*, 2006; STAGLJAR, 2016). OMICs uprising started with the improvement of high-throughput sequence methods enabling rapid and parallel analysis of thousands of distinct molecules in the cell (IDEKER; GALITSKI; HOOD, 2001). Overall, these approaches constitute the field of ‘system biology’, an integrative view of biological systems via computational and mathematical analysis to study interactions between biological components and how they behave as a whole (ARITA; ROBERT; TOMITA, 2005; KEERSMAECKER *et al.*, 2006).

Genomics is the study of the entire genome of an organism from sequencing to genomic assemble and annotation, providing a framework to map and study specific genetic variants. The field of comparative genomics goes deeper since it identify strain-specific or ‘sigletons’ (genes present in a single strain only) and specie-specific or ‘core genome’ (conserved in all sequenced strain of a bacteria specie) genes that might explain specific characteristics of each bacterial strains providing knowledge on putative proteins associated with pathogenicity, colonization, metabolism, transport, among other functions. Breakthroughs in the sequence methodologies and in bioinformatics allowed the development of life-science databases (e.g. EMBL, GenBank, NCBI) and projects (e.g. ‘The 1000 Genomes Project’, ‘ENCODE’, ‘The Human Microbiome Project’) that provide support for scientific research.

Transcriptome is the study of a range of mRNAs expressed by an organism or the transcripts of a particular cell/circumstance, since it varies according with the environment, stage of life, and tissue type. Though high-throughput sequence (e.g. microarray, RNAseq) is possible to compare two or more conditions and pinpoint the differentially expressed genes, which may highlight the ones responsible for a specific feature. Furthermore, the transcriptome can also be used to improve a genome assembly.

Proteomics is the large-scale study of a set of proteins produced by an organism. Like the transcriptome, the proteome is not constant and varies over time rendering the cellular condition and type. The proteome is influenced by the transcriptome but is also modulated (post-translational modifications) according with the cellular requirement. Proteomics enlighten the protein-protein interactions (PPI), metabolic pathways and biological processes. High-throughput technologies [e.g. mass spectrometry, differential in-gel electrophoresis] are critical to unravel the biological information acquired, and a huge amount of data is already available in proteomic databases (e.g. UniProt, PDB, Pfam, InterPro).

The genome sequence of bacterial strains has provided the opportunity to better understand the genomic content of the organism whereas transcriptome and proteomics data brought a deeper knowledge on the putative targets. The application of OMICS approach will ultimately lead to new insights into a variety of conditions/diseases that could direct to better treatments (STAGLJAR, 2016). This technology has proven crucial to comprehend the gut microbiome composition, since many of the species cannot be cultured, and OMICS approach may be able to predict interactions between the microbiome, diet, life-style, human physiology, and gene expression. In fact, many OMICS studies relate the presence or absence of key bacterial species to pathologies, like (i) the increase of *Fusobacterium* spp and *Escherichia coli* in colon cancer; (ii) the increased diversity of *Clostridium* subgroups (e.g. *C. leptum* and *C. coccooides*) and decrease of *Desulfovibrio* sp in colorectal cancer; (iii) the decrease of *Bifidobacterium* spp. and increase of *Clostridium* spp. in asthma; (iv) the presence of adherent invasive *E. coli* and *Enterobacteraceae* and decrease of *Faecalibacterium prausnitzii* in intestinal bowel disease (IBD); and (v) the increase ratio of Firmicutes/Bacteroidetes and decrease of *Bifidobacterium* and *Ruminococcus flavefaciens* in obesity (Figure 10) (KINROSS; DARZI; NICHOLSON, 2011). The connection can also be made between changes in the microbiota and modifications in the metabolomics profile as the decrease of butyrate (SCFA) (GUINANE; COTTER, 2013), increase production of hydrogen sulfide, due to the low number of sulfate-reducing bacteria (KINROSS; DARZI; NICHOLSON, 2011), the presence of isocaproic acid

(SCFA) (BJÖRKSTÉN, 2004), decrease of succinate and butyrate (GUINANE; COTTER, 2013), increase of SCFA and inflammation (TREMAROLI; BÄCKHED, 2012; TURNBAUGH *et al.*, 2009), respectively.

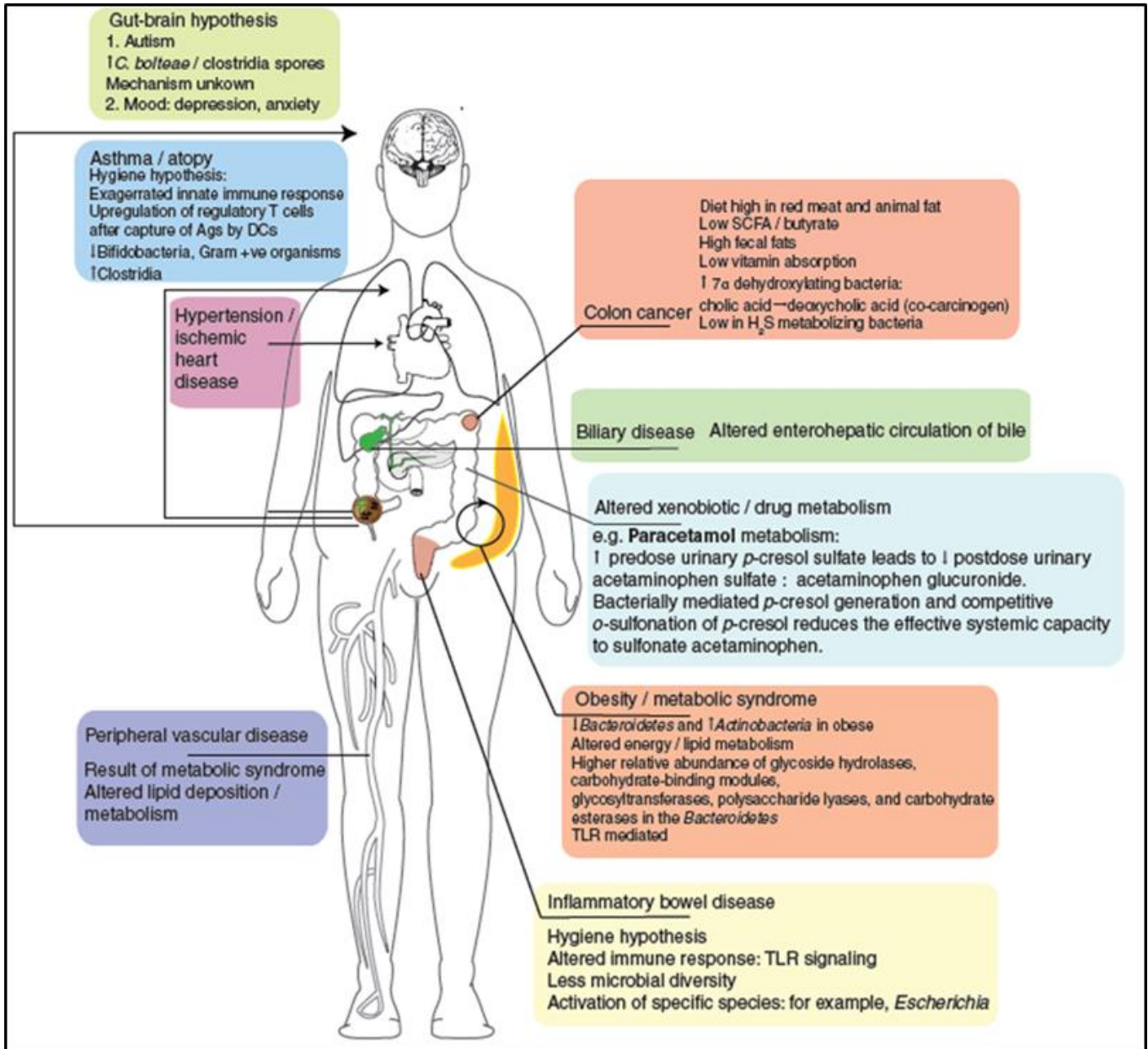


FIGURE 10: Diseases and metabolic changes associated with the gut microbiota. Systemic disease influenced by alterations in the gut microbiota (dysbiosis). Ags: antigens. *C. bolteae*: *Clostridium bolteae*. DCs: dendritic cells. SCFA: short-chain fatty acid. TLR: Toll-like receptor. Figure from (KINROSS; DARZI; NICHOLSON, 2011).

The applications of OMICS sciences is vast and enhanced our understanding of microorganisms behavior, characterization of biological variability, research on human disease, therapeutic targets against microorganism, among others (DEN BESTEN *et al.*, 2018; HASIN; SELDIN; LUSIS, 2017; SANTOS *et al.*, 2016; STAGLJAR, 2016). In fact, the use of OMICS data and bioinformatics tool has proven effective, and to date, several studies have applied only a minimal microbial component to generate long lasting immune response, like synthetic peptide-based vaccines (ALI *et al.*, 2017; DEPLA *et al.*, 2008; HAJIGHAHRAMANI *et al.*, 2017; PANDEY; BHATT; PRAJAPATI, 2018; SABETIAN *et al.*, 2019; SHEY *et al.*, 2019; ZHANG, 2018).

The pre-genomic prevention methodologies (e.g. classical vaccination) have proven to be outdated, displaying a few drawbacks, which highlighted the urgency for cutting-edge vaccine development. These new generation vaccines are produced via reverse vaccinology, immunoinformatics, and/or structural vaccinology, strategies that fuse recombinant DNA technology with computational technologies, such as database, and biological/genomic information (MARÍA *et al.*, 2017).

3.3.2. Next-generated vaccines methodologies

3.3.2.1.Reverse vaccinology

Reverse vaccinology identify likely proteins or lipoproteins targets with antigenic and physicochemical properties associated with a good antigens, employing bioinformatics tools, for further *in vitro* and *in vivo* assays testing (MARÍA *et al.*, 2017). Nonetheless, the targets essential information like cellular location and the degree of host similarity have to be checked before *in vitro* and *in vivo* assays. The molecules exposed to host cells (i.e. cell membrane location or secreted) are considered better antigens because they have a greater probability to generate a protective response (CHAUDHURI *et al.*, 2014) whereas the molecules with a high similarity with the host protein could have antigenic or toxic properties, which would induce autoimmune and cross-protection reactions (ARGONDIZZO *et al.*, 2015). Nonetheless, the vaccine candidates obtained by this technology could fail as a good vaccine antigen since the selection is solely based on computational probabilistic studies hence an experimental validation is required.

3.3.2.2.Immunoinformatics

Immunoinformatics tools predicts epitopes with the capability to induce a cellular and/or humoral response via the recognition of MHC molecules via receptors in T lymphocytes (FLOWER, 2008; HE, Y. *et al.*, 2010). The recognition is based on available datasets and specific regions of the vaccine candidates (epitopes) with the best generated stimulus. The peptide epitopes can be synthesized artificially (via chemical approaches) or obtained with molecular biology tools, and can display a single or different (multi-epitope based) antigens, alone or in multiple copies (SKWARCZYNSKI; TOTH, 2016). This renders a vaccine safe, not only in its formulation but also in its production process, because there is no risk of the presence of infectious organisms (DELANY; RAPPUOLI; SEIB, 2013). The production of synthetic epitopes is simple (via solid phase peptide synthesis, automatic synthesizers and microwave techniques), easily reproducible, fast, cost-effective, stable, with no problems with biological contaminants or allergenic reactions (SKWARCZYNSKI; TOTH, 2016). However, peptides are susceptible to enzymatic degradation, may not be recognized by the whole host population, and require adjuvants or delivery system since they are poor immunogens.

3.3.2.3.Structural vaccinology

Structural vaccinology focuses on the conformational features of candidate antigens capable to generate immunity response against antigenically divergent pathogens (MARÍA *et al.*, 2017). The bioinformatics analyses predicts linear epitope via hydrophilicity characteristic, which fetches a potential risk (e.g. the peptide conformation within the final protein could differ from the sole linear peptide).

4. The two sides of antibiotic treatment and possible alternatives

While vaccines are a prestigious strategy to prevent infections, antibiotic treatments have been used as a method to treat the disease by killing bacteria or preventing their growth. However, the bacteria ability to rapidly adapt to antibiotic pressures is a growing concern and the gain of resistance is aggravated by the misuse of antibiotics, horizontal transfer, and within-host bacterial adaptation, which was observed for

Mycobacterium tuberculosis and *Staphylococcus aureus* (DIDELOT *et al.*, 2016; ELDHOLM *et al.*, 2014; GAO *et al.*, 2010; HOWDEN *et al.*, 2011; WHO, 2014). Thus, the search for alternative treatments besides the use of antimicrobial therapy is of almost importance and deliberated as a global health concern that needs to be addressed through a national coordination and integrated surveillance between human and veterinary medicine, agriculture, and main stakeholders, like the food and pharmaceutical industry, academia, and the government.

The excessive exposure to antimicrobial (i.e. indiscriminate use of antibiotics) is accelerating the gain of resistance by the bacteria, which makes medicine/treatment ineffective escalating the infections persistence in the body (HOLMES, A. H. *et al.*, 2016). Essentially, 50% of prescribed antimicrobials are considered unnecessary and often medicament are given without professional oversight (self-medication) (CENTERS FOR DISEASE CONTROL AND PREVENTION, 2013). Likewise, the use of antimicrobials as animal growth promoters and in routine infection prevention in cattle further assist the dissemination of drug resistance, and the use of antimicrobials in food production is actually higher than in clinical settings (80% of total antibiotic consumption in the US) (VAN BOECKEL *et al.*, 2015; WHO, 2012). These complicates the overall resistance scenario since most of the drugs considered ‘medically important’ for humans are also applied in animals. Moreover, this practice comprise 62% of the currently used antibiotics and the remaining percentage may also have a role (direct or indirect) (MARSTON *et al.*, 2016). Furthermore, the exacerbate use of antibiotics prompt a higher health care cost (estimated bugged of \$20 billion annually in the United States - US) since infections with resistance pathogens result in longer duration of illness/hospitalization, additional tests and use of more expensive drugs (MARSTON *et al.*, 2016).

Altogether, the overall resistance scenario highlight the need for surrogate approaches to treat pathogen-induced illness. Among the alternatives, vaccination could circumvent the antibiotic resistance issue since would avoid the infection and consequently the use of antibiotics. For instance, an enhanced worldwide coverage of the vaccine against *Streptococcus pneumoniae* could prevent 11.4 million antibiotic days per year in children under five years (LAXMINARAYAN *et al.*, 2016). Therefore, the development of cutting-edge vaccines that avoid the drawbacks of conventional vaccine whereas preventing the pathogen dissemination is of outmost importance. Considering that, we applied *in silico* analysis (immunoinformatics) to design two multi-epitope vaccine against bacterial pathogens responsible for a

high rate of morbidity and mortality worldwide. Chapter 1 emphasize the etiological agent of diphtheria (*Corynebacterium diphtheriae*) while Chapter 2 focus in a major nosocomial pathogens (*C. difficile*) responsible for life-threatening infections

CHAPTER I:

**Design of a broad-spectrum candidate
multi-epitope vaccine against Diphtheria: An
integrative immunoinformatics approach**

1. *Corynebacterium diphtheriae*: a life-threatening bacteria

1.1. The *Corynebacterium* genus

The genus *Corynebacterium* is highly diversified being currently composed of proximally 110 validated species of medical, veterinary, or biotechnological relevance (BERNARD, 2012; EIKMANN; BLOMBACH, 2014; OLIVEIRA *et al.*, 2017; SOARES *et al.*, 2013). The genus proposed by Lehman and Neumann, in 1896, belongs to the Actinobacteria class and, based on the molecular similarity of cell wall components, was later clustered with *Mycobacterium*, *Nocardia* and *Rhodococcus* to compose the CMNR group.

The members of the genus are Gram-positive bacteria with a high GC content (47–74%), rod shape morphology, immobile, non-sporulated, aerobic, and catalase positive bacteria. Furthermore, their cell wall is thick due to the presence of mycolic acids, peptidoglycan, arabinogalactan, and saturated/unsaturated fatty acids (OLIVEIRA *et al.*, 2017). These cell wall components can trigger macrophage activation hence they may display an adjuvant effect in the CMNR group (BLANC *et al.*, 2013; GEISEL *et al.*, 2005; MISHRA, A. K. *et al.*, 2012; PEET *et al.*, 2015; RYAN *et al.*, 2011; SCHICK *et al.*, 2017; SCHOENEN *et al.*, 2010; SHENDEROV *et al.*, 2013; TAKEUCHI *et al.*, 1999).

The genus is widely distributed in nature and are commonly described as commensal members of normal skin microbiota and mucous membranes of several hosts and residents of the environment (soil, water, among others), or considered pathogenic or opportunistic bacteria. Certain non-pathogenic strains of the genus (e.g. *Corynebacterium glutamicum*, *Corynebacterium efficiens*, and *Corynebacterium crenatum*) are applied in industrial settings bearing biotechnological importance in the production of amino acids and in the cheese industry. Through history, several *Corynebacterium* species were associated with animal diseases (e.g. *Corynebacterium amycolatum*, *Corynebacterium aquilae*, *Corynebacterium auriscanis*, *Corynebacterium bovis*, *Corynebacterium camporealensis*, and *Corynebacterium canis*) being considered important zoonotic agents (TAUCH; SANDBOTE, 2014). However, *Corynebacterium diphtheriae*, *Corynebacterium ulcerans*, and *Corynebacterium pseudotuberculosis* stands out due to their medical and veterinary relevance (BERNARD, 2012; DIAS *et al.*, 2011; WAGNER *et al.*, 2010). These species produce potent exotoxins [i.e. diphtheria toxin (DT) and/or phospholipase D] being associated with a life-threatening diseases, like diphtheria and caseous lymphadenitis (TAUCH; SANDBOTE, 2014).

1.2. The pathogenic bacteria *C. diphtheriae*

C. diphtheriae is the most significant pathogen of this group and thus was one of the first species of the genus with a completely sequenced genome (CERDEÑO-TÁRRAGA *et al.*, 2003). This medically important bacterium is the etiological agent of diphtheria, a disease that was almost eradicated, due to the manufacture of an universal vaccine, but outbreaks and isolated cases are still being reported worldwide and the illness remain a significant health problem, especially in developing countries (HESSLING; FEIERTAG; HÖNES, 2017; MATTOS-GUARALDI *et al.*, 2003; WHO, 2017b, 2018c). Diphtheria is a toxin-mediated disease caused mainly by *C. diphtheriae* toxigenic strains. *C. diphtheriae* can be divided in toxigenic and non-toxigenic strains, where the toxigenic carries a functional *tox* gene encoded in the corynephage (β^{tox+} , γ^{tox+} or ω^{tox+}) that is integrated in the chromosome (COLLIER, 2001; GUARALDI; HIRATA; AZEVEDO, 2014). Nonetheless, non-toxigenic are also found as clinical isolates and are now emerging as a significant cause of infections, like persistent sore throats, endocarditis, septic arthritis, osteomyelitis and cutaneous diphtheria (OLIVEIRA *et al.*, 2017; SANGAL; HOSKISSON, 2016). Furthermore, *C. diphtheriae* strains can be assigned in accordance with a biovar designation (i.e. mitis, gravis, intermedius and belfanti). This division was stipulated based on morphological and biochemical parameters; however, the classification is not fully adopted and considered unreliable since the difference is not reflected in the disease caused by diverse biovar types (e.g. infection severity, prevalence) (SANGAL *et al.*, 2014; SANGAL; HOSKISSON, 2016).

The cell envelope of the CMNR group is directly linked with the bacteria ability to survive in harsh environments, acting as the primary line of defense, protecting the bacterium against salt, oxidative, physical, metal, detergent, antibiotics, and membrane stress (Burkovski, 2018; Ott, 2018). The cell envelope is composed of the plasma membrane [i.e. phospholipids and lipid bilayer like lipomannan (LM) and lipoarabinomannan (LAM)], a heteropolysaccharide meshwork (i.e. peptidoglycan covalently linked to arabinogalactan), the mycomembrane (i.e. mycolic acids and LM/LAM glycolipids), and the outer layer (i.e. free polysaccharides, glycolipids, and proteins) (BURKOVSKI, 2013) (Figure 11). This arrangement of layers function as adhesion and virulent factors, which benefits *C. diphtheriae* colonization and survival within the host. The *C. diphtheriae* lipoarabinomannan-like (CdiLAM) compound acts as an adhesin following the first 30 minutes of infection (initial steps) binding to human respiratory epithelial cells (Moreira *et al.*, 2008.)

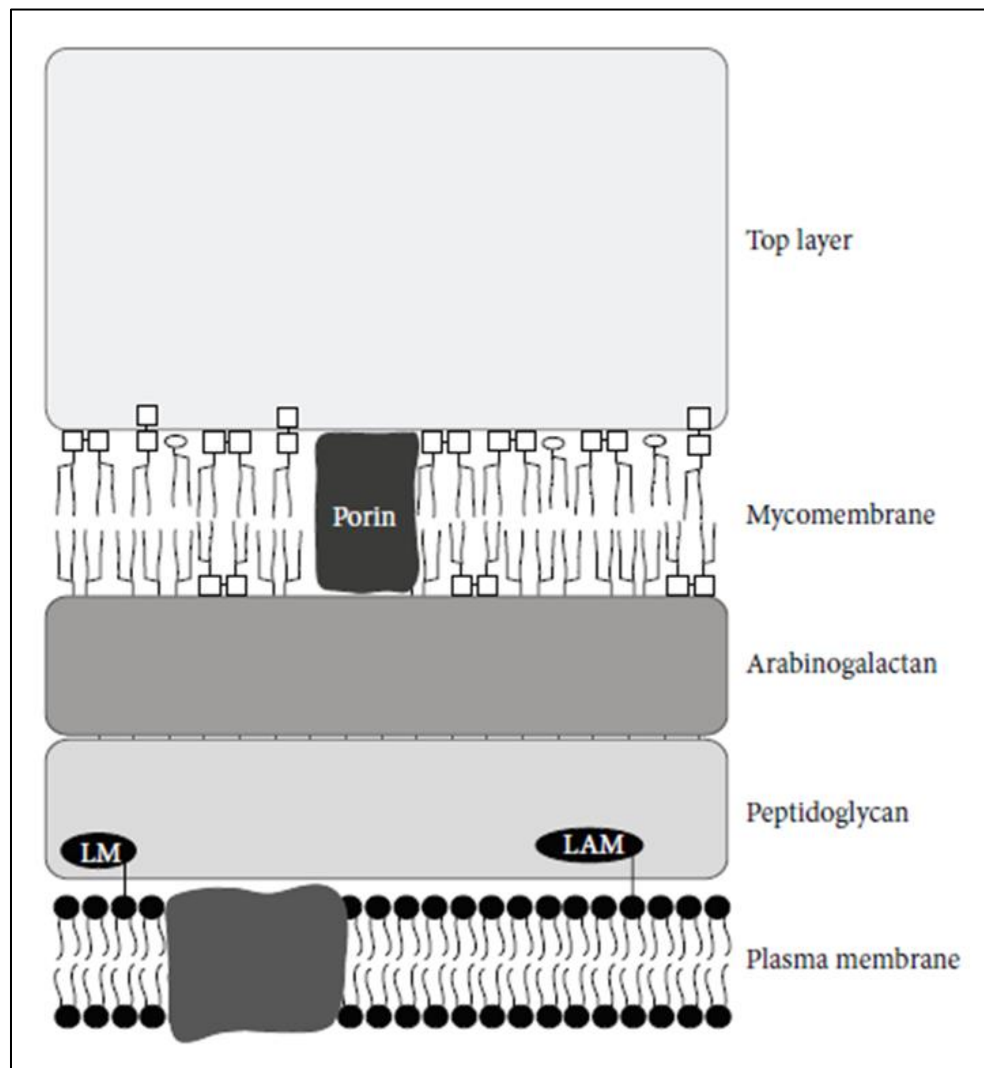


FIGURE 11: The cell envelope of *Corynebacterium diphtheriae*. The cell envelope of corynebacteria is comprised of five layers. The outer layer consists of glucans, proteins and may display glycolipids (e.g. LM and LAM) and porins. The mycomembrane is composed of mycolic acids, glycolipids and may exhibit porins. The heteropolysaccharide meshwork comprise the arabinogalactan and peptidoglycan layers covalently linked. Meanwhile, the plasma membrane is formed by phospholipids and a lipid bilayer. Figure from (BURKOVSKI, 2013).

The *C. diphtheriae* genome comprise of 2.46 Mbp being highly variable, which may reflect in the degree of pathogenicity among the strains (SANGAL *et al.*, 2015; SANGAL; HOSKISSON, 2016; TROST *et al.*, 2012a). The differences between genomes is mainly due to the presence/absence of prophages, transposons, restriction-modification systems and CRISPR elements, which are present mostly inside

pathogenic island (TROST *et al.*, 2012a). The pan-genome consist of 3,989 genes (i.e. core and accessory genome contain 1,625 and 2,364 genes respectively) (Figure 12) (SANGAL *et al.*, 2015). The core proteins are common among all *C. diphtheriae* strains and are probably implicated in many key functions (e.g. virulence, adhesion, transport) and strain-specific features whereas the accessory genome may be responsible for the heterogeneity among the strains. The diversity can also be linked with the distribution of genomic island trough the genome (TROST *et al.*, 2012a). According to the authors, fifth-seven genomic islands have been identified in *C. diphtheriae* genome with eight common among all strains. Genomic islands can be classified as pathogenic (siderophore uptake, *tox* and *spa* genes), resistance (e.g. heavy metal and antibiotic resistance), phage (typical phage products) or metabolic (e.g. sugar transport and polysaccharide degradation) islands. Genomic studies revealed that the pathogenic islands can be horizontally transferred (partially or completely), which are probably linked with the presence/absence of iron acquisition and pili formation genes (SANGAL *et al.*, 2015; TROST *et al.*, 2012a).

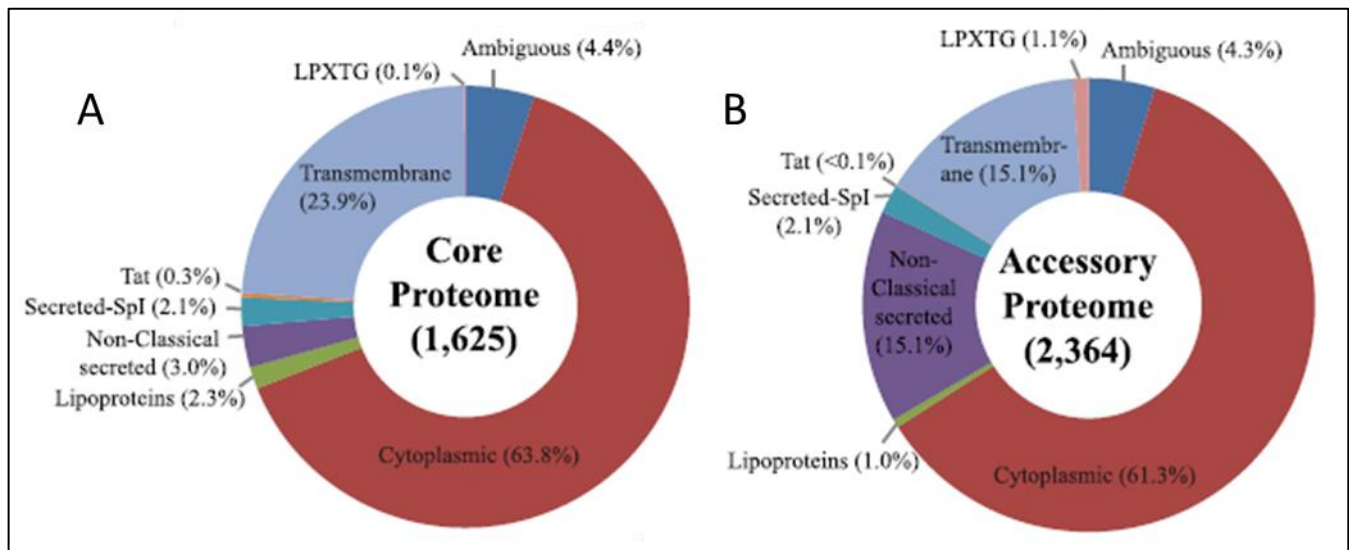


FIGURE 12: *C. diphtheriae* pan proteome. *C. diphtheriae* predicted proteins from the core (A) and accessory (B) genome percentage were distributed in accordance with their cellular location/function (e.g. cytoplasmic, transmembrane or secreted). The percentage of each segment was calculate considering the number of predicted proteins from the core and accessory genome, respectively. Figure adapted from (SANGAL *et al.*, 2015).

The variation in the accessory genome, distribution of genomic island and the presence of mobile elements may be associated with *C. diphtheriae* different lifestyles (i.e. colonization of different cells/hosts and adherence to different epithelial cells) and exploitation of various energy sources (i.e. variation in nutritional requirements) (SANGAL *et al.*, 2015; TROST *et al.*, 2012a). Nonetheless, the vast majority of the predicted genes (49%) encode hypothetical/uncharacterized proteins, which impose a major obstacle in the bacterium comprehension and study (SANGAL *et al.*, 2015). Among the characterized genes, most encode cytoplasmic proteins involved in a wide range of cellular functions (e.g. DNA replication, recombination and repair, translation, biogenesis, transcription, transport, metabolism, cell division, and defense mechanisms). Meanwhile, the transmembrane proteins were associated with host defense, transport and surface adhesion; and the secreted proteins with cell wall homeostasis/biosynthesis, redox process, secretion system and virulence (SANGAL *et al.*, 2015). Therefore, these proteins are often recognized by the immune system, via PRR (e.g. Mincle and TLR2); nonetheless, remain crucial for the bacterium survival (ANTUNES, C. A. *et al.*, 2015; BURKOVSKI, 2018; OTT *et al.*, 2017; PEET *et al.*, 2015; SCHICK *et al.*, 2017).

Utterly, pathogenic bacteria internalization is an advantageous step during infection, since protects the bacterium against phagocytosis and antibiotics, and provide the bacterium an opportunity to colonize and spread within the host. Therefore, the genes associated with this process and others relevant for *C. diphtheriae* pathogenicity are of outmost importance for diphtheria development.

1.3. *C. diphtheriae* factors associated with infection

Secreted and transmembrane proteins are commonly associated with host-pathogen interactions and consequently pathogenesis but they exhibit great variance between *C. diphtheriae* strains being directly involved with diphtheria severity. Thus, some of those proteins were highlighted in this work, such as *C. diphtheriae* main virulent factor (Tox protein), the diphtheria toxin repressor (DtxR), an adhesin (DIP0733), and pili proteins (Spa) (Figure 13).

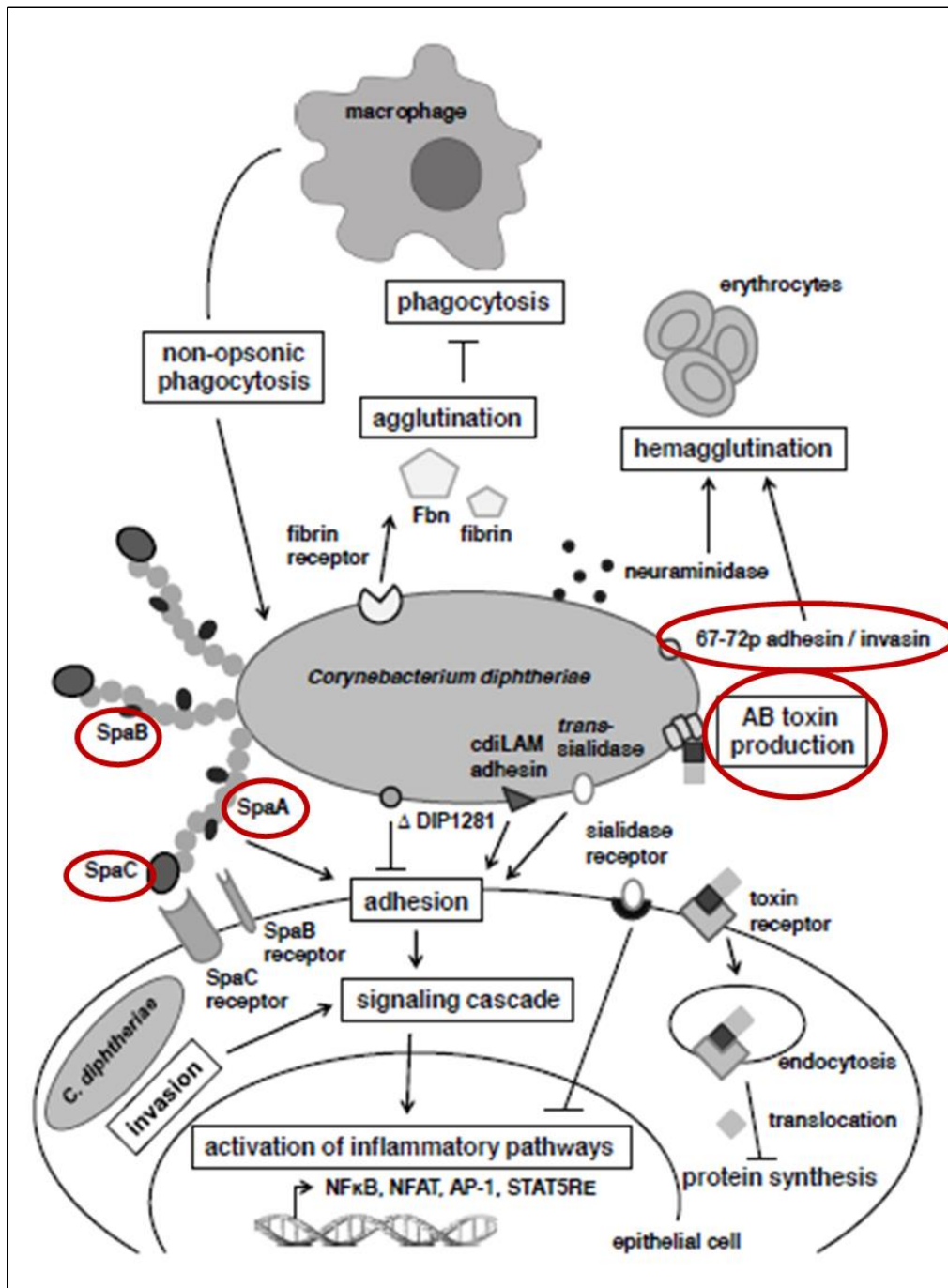


FIGURE 13: *C. diphtheriae* key proteins for infection. *C. diphtheriae* display many genes relevant for the survival, overgrowth and dissemination within the host. Among then three were highlighted in this work (red circles). The Tox protein (AB toxin) is the main virulence factor of *C. diphtheriae*. The toxin is released by the bacterium and internalized by endocytosis (host), which induce a host immune response. The pili structures (SpaABC) are adhesion factors that contributes to *C. diphtheriae* colonization whereas the non-fimbrial protein 67-72p (DIP0733) are crucial for the colonization, invasion and dissemination within the host, and thus may contribute to host cell apoptosis and necrosis. Furthermore, other proteins are linked with *C. diphtheriae* capability to prompt the infections, such as Neuraminidases or sialidases (NanH), lipoglycan (CdiLAM – glycolipid from *C. diphtheriae*

cell envelop), invasion proteins (DIP1281). Altogether, these proteins induce an inflammatory host response (e.g. activation of the NFκB transduction pathway) that will have a direct impact on the disease progression. Figure adapted from (OTT; BURKOVSKI, 2014).

1.3.1. Diphtheria toxin and the diphtheria toxin repressor

The *tox* gene encode the DT, a potent exotoxin that acts in all tissues with special tropism for the myocardium, nervous system, kidneys and adrenals (BAE, 2011). Diphtheria toxin is a single polypeptide of 535 amino acids and a molecular mass of 58.4 kDa, divided into two subunits [i.e. subunit A (catalytic domain) and subunit B (binding domain)] linked together by disulfide bonds (MITAMURA *et al.*, 1997). The DT is inactive until it is split into its two components (Figure 14). The toxin is translocated by receptor mediated endocytosis when the subunit B binds to the heparin-binding epidermal growth factor like growth factor, a DT specific cell type dependent cell-surface receptor (MITAMURA *et al.*, 1997; NAGLICH *et al.*, 1992). Once inside the endosome, the disulfide bond is reduced and the subunit A is released into the cytosol (JOHNSON *et al.*, 1988). The fragment A inhibit protein synthesis by irreversibly inactivating the elongation factor 2, an essential factor in the protein synthesis machinery (i.e. elongation step for protein translation) of eukaryotic cells, through ADP-ribosylation (HOLMES, R. K., 2000) (Figure 13 and 14). The toxin can reach the bloodstream and spread to any organ cells that have the receptor, which can result in pseudomembrane on mucosal areas, systemic infections, myocarditis and neuritis (BAE, 2011).

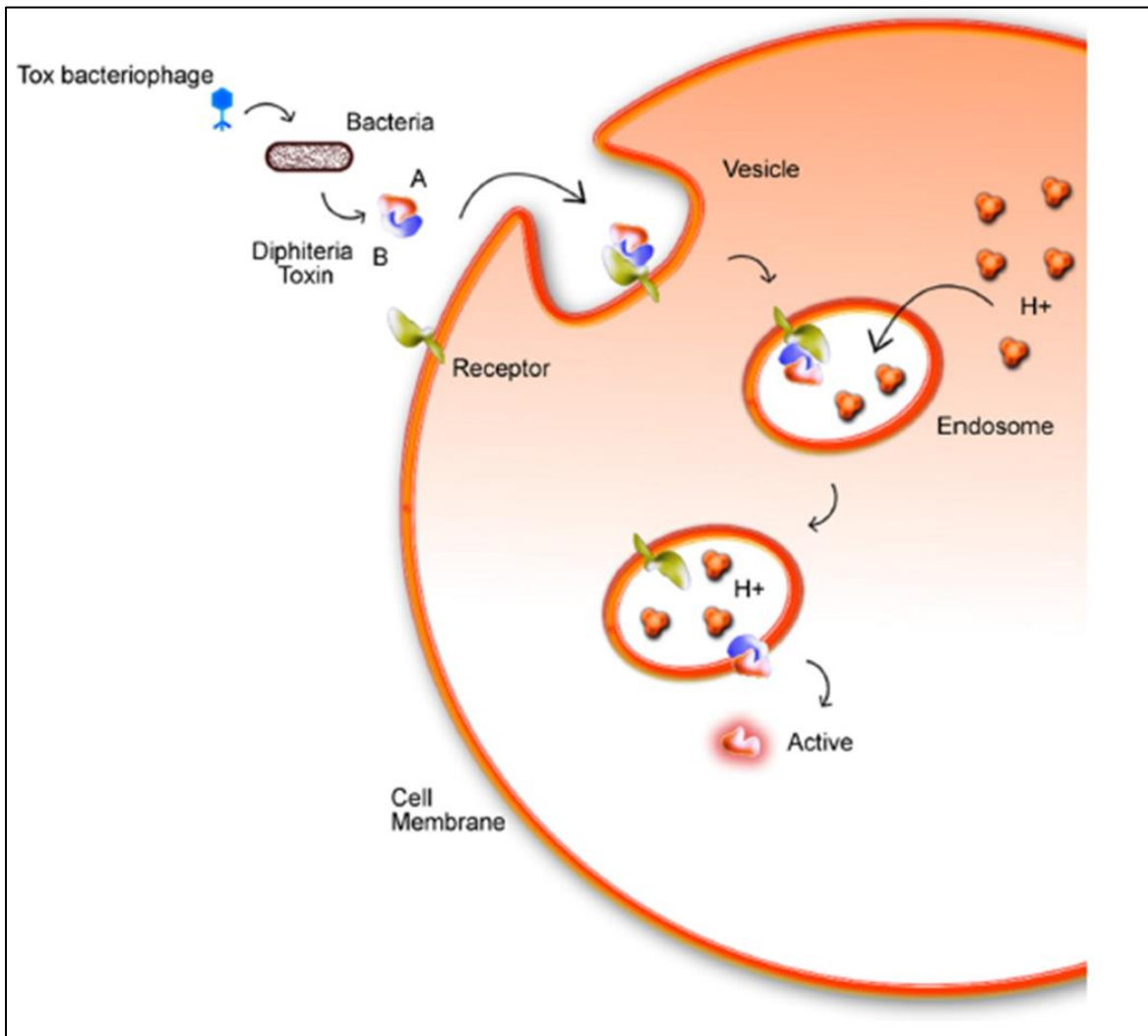


FIGURE 14: Diphtheria toxin mechanism of action. The toxin is released by the bacterium and the subunit B recognizes/binds to the toxin receptor (EGF-like growth factor), which promotes the DT entry through the membrane by endocytosis. Inside the endosome, the subunit A is exposed and the translocation domain inserts into the endosomal membrane and enables the transport and further activation of the toxin. Once in the host cytosol lumen, the active toxin can inhibit EF-2 and thus obstruct protein synthesis and lead to cell death. Figure from (SCHNELL *et al.*, 2016).

DT is synthesized depending on the iron concentration in the environment. During limited conditions of iron, the bacteria start to synthesize the toxin and this modulation is attributed to the transcriptional regulator DtxR (TAO *et al.*, 1994). DtxR regulates the expression of iron homeostasis and the *tox* gene, being central for the pathogenicity of this bacterium (WITTCHEN *et al.*, 2018). A DtxR binding site is located upstream of the *tox* gene; and under high-iron conditions, DtxR is activated and the site is blocked,

due to the formation of a DtxR dimer that binds within a 19 bp-palindromic sequence (DtxR motif) of the *tox* promoter. Likewise, during iron-limiting conditions, DtxR is deactivated, leading to expression of DT (TAO *et al.*, 1994).

Δ *dtxR* mutant have an increase transcription of *tox* up to 4-fold, when compared with the wild type (WITTCHEN *et al.*, 2018). The deletion of the *dtxR* gene had a remarkable influence on the transcriptome of the mutant strain affecting around 15% of all genes (i.e. directly or indirect effects), which confirms the global effects of DtxR regulation. Moreover, DtxR is a dual regulator of iron homeostasis in many bacteria and its binding site was found upstream of several iron uptake related genes, like siderophores and heme oxygenases (BRUNE *et al.*, 2006; TROST *et al.*, 2012a; WITTCHEN *et al.*, 2018).

1.3.2. DIP0733

DIP0733 is a multifunction non-fimbrial adhesin, previously named 67–72p, involved in the adhesion of *C. diphtheriae* to erythrocytes and human epidermoid laryngeal carcinoma cell HEp-2 cells (COLOMBO *et al.*, 2001; HIRATA, RAPHAEL *et al.*, 2004; SABBADINI *et al.*, 2012). Moreover, DIP0733 can be considered as a microbial surface component with virulence properties that recognize adhesive matrix molecules (ANTUNES, C. A. *et al.*, 2015). Both, the 67 and 72 kDa polypeptide bind to erythrocyte and HEp-2 receptors, leading to hemagglutination; however, the binding can be blocked by anti-67–72p IgG antibodies (HIRATA, RAPHAEL *et al.*, 2004). HEp-2 cells and erythrocytes have a negative correlation with bacterial adhesion; strains with low hemagglutinating activity showed high adhesion rates to HEp-2 cells and vice versa.

The main role of this protein lies in blood clot formation by its conversion into insoluble fibrin, which forms the matrix of the clot. *C. diphtheriae* is also able to bind to fibrinogen/collagen and convert fibrinogen to fibrin, which leads to the formation of bacterial aggregates in the presence of plasma (Figure 13) (ANTUNES, C. A. *et al.*, 2015; SABBADINI *et al.*, 2012). The binding of fibrinogen to the surface of *C. diphtheriae* may be an efficient method to avoid phagocytosis and the conversion of fibrinogen may be connected to pseudomembrane formation (SABBADINI *et al.*, 2010). Furthermore, DIP0733 contributes to bacterial invasion and apoptosis of epithelial cells during *C. diphtheriae* infection,

emphasizing the potential for non-toxigenic strains to adhere, invade and survive within the intracellular environment plasma (Figure 13) (ANTUNES, C. A. *et al.*, 2015; SABBADINI *et al.*, 2012).

1.3.3. Pili proteins

Pili or fimbriae are known virulence factors that facilitate bacterial colonization and pathogenesis. *C. diphtheriae* harbors three pilus gene clusters encoding the heterotrimeric SpaA, SpaD, and SpaH type pili on pathogenicity islands (OTT *et al.*, 2010; OTT; BURKOVSKI, 2014; TON-THAT; SCHNEEWIND, 2003). Pilus biogenesis requires tandem sortase enzymes (i.e. *srt* operon) that covalently link SpaABC monomers and subsequently anchor the resulting heterotrimeric polymer to the bacterial peptidoglycan (MANDLIK *et al.*, 2008). The sortases SrtA-E catalyze the covalent cross-linking of single pili monomers to form the pilus (TON-THAT; MARRAFFINI; SCHNEEWIND, 2004, p.).

The SpaA cluster comprises the genes *spaA*, *spaB*, and *spaC* whereas the SpaD and SpaH clusters are formed by *spaD*, *spaE*, and *spaF* genes and *spaG*, *spaH*, and *spaI* genes, respectively (TON-THAT; SCHNEEWIND, 2003). The pili proteins carry a cell wall sorting signal LPxTG motif at the C-terminus that is responsible for efficient anchoring of pili to the cell wall by the housekeeping sortase SrtF. Furthermore, the proteins also harbor the LxET motif (E box) that determine the specificity of the pilus subunits for the corresponding sortases (TON-THAT; MARRAFFINI; SCHNEEWIND, 2004, p.).

The *spa* operon is variable among *C. diphtheriae* strains (i.e. presence of the genes and number of pilus subunits) with at least two operons being present in all strain, and this discrepancy correlates well with the bacterium ability to cause severe, invasive non-toxigenic disease (SANGAL *et al.*, 2015; TROST *et al.*, 2012a). The conservation of the *spa* locus throughout evolution underlines its crucial role to incite diphtheria disease (OTT *et al.*, 2010). SpaA-type pili specifically facilitate corynebacterial adherence to human pharyngeal epithelial cells; however, they are not crucial for pathogen-host interaction but the absence of SpaB and SpaC greatly diminished adherence (Figure 13) (MANDLIK *et al.*, 2007). Maintenance of *spaA*, *spaB*, and *spaC* is important since mutations affecting these genes and *srtA* compromise the bacterial fitness. Meanwhile, *spaD* and *spaH* loci preferentially mediate binding to

laryngeal and lung epithelial cells, and their low conservation suggests that mutations in these regions are better tolerated (MANDLIK *et al.*, 2007).

Strains with all three Spa clusters exhibited higher adhesion to the pharyngeal Detroit 562 cell lines and mutant strains lacking all sortases showed decreased adhesion rates to different epithelial cell lines (MOREIRA, L. DE O. *et al.*, 2003). However, the adhesion, invasion and presence of pili are not strictly coupled, indicating that adhesion mechanism is multifactorial and different proteins may be involved in the development of diphtheria (OTT *et al.*, 2010).

2. Diphtheria

2.1. Main characteristics, transmission and clinical manifestations

C. diphtheriae is the etiological agent of diphtheria, an infectious disease that, throughout history, was considered one of the most feared illness, since the bacterium prompt devastating epidemics that mainly affected children (WHO, 2017b). The disease was previously known as ‘strangling angel of children’ in reference to the many infected infants that died by suffocation due to classical diphtheria symptoms. Diphtheria typically spreads from person to person by respiratory droplets produced by infected person, contagious cutaneous lesion, and asymptomatic carriers. Establishment of the disease takes 2 to 5 days and patients are infectious for two to three weeks. Furthermore, animals (e.g. domestic, wildlife and aquatic) also play a role in transmission acting as reservoir of the bacterium (SANGAL; HOSKISSON, 2016; SCHEIFER *et al.*, 2018; SING *et al.*, 2016).

Diphtheria probably originated in the Middle East and then disseminated to Europe and other continents. During the beginning of the XIX century, diphtheria became a leading cause of infant mortality with large epidemics during periods of war. In the 1990s, a major outbreak centered in the Russian Federation, Ukraine and neighboring countries reported over 157,000 cases with proximally 5,000 deaths, which attain specially people over 15 years of age (64–82 %) (DITTMANN *et al.*, 2000; ESKOLA; LUMIO; VUOPIO-VARKILA, 1998). This swift (infant to teenagers and adults) was attributed to the mass vaccination programs focused in the children but neglected in adults (absence of buster doses). This epidemic

contributed to uncover the correlation between socioeconomic difficulties, large-scale population and poor health-care system with the spread of diphtheria.

Nowadays, the disease is partially controlled by the diphtheria vaccine and the many reported cases may be connected with a poor vaccination coverage, a rising heterogeneity of *C. diphtheriae* circulating strains (e.g. emergence of multidrug resistant bacteria and an increase of non-toxigenic strains), and socioeconomic problems (e.g. malnutrition, poor hygiene, poverty and overcrowding population) (TROST *et al.*, 2012b; WHO, 2017b, 2018c). Moreover, India, Indonesia and Madagascar had 18,350, 3,203 and 1,633 reported cases, respectively, during 2011 to 2015 (WHO, 2017b). Altogether, South-East Asia region notifies 55–99% of all *C. diphtheriae* cases each year during this period (WHO, 2017b). Furthermore, the inadequate use of booster vaccine, travels to endemic countries, immigrants, and non-toxigenic *C. diphtheriae* are associated with cases of diphtheria in develop countries (SANGAL; HOSKISSON, 2016; SCHEIFER *et al.*, 2018; TILEY *et al.*, 1993). Those reported cases are probably underestimated and the actual burden of diphtheria is likely much greater.

C. diphtheriae infection can cause respiratory or cutaneous diphtheria and in sporadic cases can lead to systemic diphtheria (e.g. bacteremia, endocarditis, hepatic and splenic abscesses, meningitis, osteomyelitis, pneumonia, and arthritis) (HIRATA, R. *et al.*, 2008; MISHRA, B. *et al.*, 2005; MUTTAIYAH *et al.*, 2011).

Diphtheria of the upper respiratory tract or classical diphtheria causes sore throat and low fever and is characterized by the formation of a pseudomembrane, which could be removed/healed or render an airway obstruction that could result in suffocation, agony and death (BURKOVSKI, 2014; MATTOS-GUARALDI *et al.*, 2003). The pseudomembrane is composed of bacteria, necrotic epithelial and inflammatory cells embedded in a fibrin matrix (HADFIELD *et al.*, 2000) being typically asymmetrical and firmly attached to the underlying tissue (i.e. nose, pharynx, tonsils, or larynx), and thus, attempts to remove it could result in bleeding (BURKOVSKI, 2014; GUARALDI; HIRATA; AZEVEDO, 2014; MATTOS-GUARALDI *et al.*, 2003). Furthermore, the infected tissues may show inflammatory symptoms, the anterior cervical lymph nodes can become markedly enlarged, which would result in edema of the surrounding tissues (‘bull-neck’), and the inflammation could lead to necrosis, due to the detrimental action of the DT. In rare conditions, the toxin are absorb into the bloodstream and disseminated trough the

body, which may result in toxic damage to organs such as the heart, kidneys and peripheral nerves (systemic infection) (TAO *et al.*, 1994).

The cutaneous diphtheria is more common and prevails over the respiratory form in warmer climates (e.g. tropical and subtropical regions) and in extreme underprivileged settings (e.g. overcrowding, poverty, poor hygiene), since those conditions hamper the healing process and favor the infection. This form of the disease is spread by contact with infected skin, respiratory droplets (via patient contaminated with respiratory diphtheria) or exposure to dust or clothing contaminated with *C. diphtheriae* (HÖFLER, 1991). Feet, lower legs and hands are common sites of infection and normally those skin lesions are also contaminated with other pathogen (co-infections), which interfere with the diagnostic rendering variable clinical manifestation. The infection can progress from fluid-filled pustule to a punched-out ulcer if not treated. The initial step is salient with the presence of a pseudomembrane that can eventually falls off, leaving a hemorrhagic base with a surrounding of edematous tissue (GUARALDI; HIRATA; AZEVEDO, 2014; HÖFLER, 1991). Spontaneous healing was observed; however, may take several weeks to months or even years to heal the wound, which favor dissemination of the disease and might explain the extremely high infection rates (i.e. up 60% in developing countries) (HÖFLER, 1991).

2.2. Diagnosis and treatment

Clinical diagnosis of diphtheria usually relies on observation of clinical manifestations (e.g. presence of pseudomembranous pharyngitis) and toxin production (i.e. PCR and/or immunoprecipitation test). Nonetheless, in the presence of pseudomembrane the treatment should start immediately without the laboratory results (WHO, 2017b).

Rapid and accurate detection of potentially toxigenic corynebacteria is necessary for the determination of appropriate measures in controlling the dissemination of toxigenic strains in both veterinary and/or human populations. Therefore, the tests are considered priority cases (i.e. laboratory emergency) and must be performed straightway. The material should be obtained by swabbing the edges of the mucosal lesions and then placed in appropriated media (i.e. Amies or Stuart media followed by inoculation into Tinsdale media). The toxin production is directly tested via PCR reaction (genotypic test) and in some cases (i.e.

uncertain PCR result) confirmed via the modified Elek immunoprecipitation test, which takes from 24 to 48 hours (WHO, 2017b).

PCR assays are simple, easy to interpret, and are becoming more widely available, thus are considered the golden standard for diagnostic offering many advantages over standard phenotypic tests. Recently, a multiplex PCR was proposed for laboratory diagnosis of *C. diphtheriae*, *C. ulcerans*, and *C. pseudotuberculosis*, which entails in a preparation of primers capable to discriminate toxigenic isolates of the three species. The reaction mixture consists of primers (reverse and forward) for amplification of *rpoB* (for corynebacteria); 16SrDNA (for *C. ulcerans* and *C. pseudotuberculosis*), *dtxR* (for *C. diphtheriae*), *pld* (specific for *C. pseudotuberculosis*), and *diph4* (for *tox* gene) (TORRES *et al.*, 2013).

C. diphtheriae gold standard treatment involves intravenous or intramuscular administration of equine-derived DAT (polyclonal IgG antibody) that should be administered quickly and at once, since the toxin inside the host cells is unaffected by DAT (WHO, 2000, 2017a, b). The recommended dose varies between 20,000 and 100,000 units but the treatment is not without adverse effects (e.g. anaphylaxis). Interventions like tracheotomy, mechanical removal of the pseudomembrane, intubation and ventilator can be necessary to avoid airway blockage and patients should be monitored to prevent the development of cardiac complications. Furthermore, antibiotics (i.e. penicillin or erythromycin) treatment is applied to prevent further transmission to uninfected individuals by limit the bacterium carriage. MDR *C. diphtheriae* strains have been described and the resistance to erythromycin is carried by a plasmid and supposed to be associated to coryneform microorganisms colonizing the skin (MINA *et al.*, 2011; PEREIRA *et al.*, 2008; SCHILLER; GROMAN; COYLE, 1980; ZASADA; BACZEWSKA-REJ; WARDAK, 2010).

Novel approaches to treat diphtheria via passive immunization has been demonstrated in preclinical models and include DT monoclonal antibodies and recombinant modified DT receptor molecules (WHO, 2017b). However, the approval for clinical use may take several years.

2.3. Diphtheria vaccine

Initially diphtheria was treated through passive immunization using toxin-neutralizing antibodies (antitoxin) isolated from animals, previously injected with sub-lethal doses of DT (RAPPUOLI; MALITO,

2014). In 1920s, the diphtheria toxoid vaccine was developed by formaldehyde detoxification of DT (inactivated diphtheria exotoxin) and afterwards an alum-precipitated diphtheria toxoid was produced, and nowadays, the toxoid vaccine is adsorbed onto an alum gel adjuvant (aluminum hydroxide or aluminum phosphate) to improve the immune response against *C. diphtheriae* (BAE, 2011; JAMAL *et al.*, 2017; MÖLLER *et al.*, 2019). In the 1940s, the diphtheria toxoid, tetanus toxoid and pertussis antigens were combined in the DTP vaccine and are now distributed around the world. Ramon and Zoeller (1926) proposed the simultaneous vaccination against diphtheria, tetanus and pertussis after the combination of anatoxins (i.e. inactivated toxins) had no negative interference with an enhanced antibody response. In 1970s, the DTP vaccine was introduced into the EPI program of WHO, which improved the global infant immunization rate. Since the 1990s, the combination vaccine DTaP [diphtheria, tetanus, and acellular (purified components of *Bordetella pertussis*) pertussis] replaced the DTwP (diphtheria, tetanus, and whole-cell pertussis) for childhood immunization. The development of the diphtheria vaccine was a milestone in the history of medicine and saved countless lives.

Diphtheria toxoid-containing vaccines are among the oldest vaccines in current use. The detoxification process established by Ramon in 1923 is still used with only small refinements (PLOTKIN, S., 2014). The DT is detoxified by formaldehyde via a series of purification steps (BAE, 2011). Formaldehyde induces the formation of intramolecular cross-links in several regions of DT and stabilize antigens useful for their long-term storage (RAPPUOLI, 2007). The toxin is isolated from the hypervirulent strain Park-William No. 8 (PW8) (BAE, 2011; LAMPIDIS; BARKSDALE, 1971). The breakthrough discovery that allowed the preparation of a completely nontoxic DT is due to the introduction of the flocculation method for measuring the dosage of toxin and antitoxin and the antigenic and immunogenic activity of toxins (BAE, 2011).

Briefly, the PW8 strain is grown in fermentors for 36 to 48 hours and when the concentration of toxin reaches 150–250 Lf/mL (i.e. Lf is the flocculation units defined as the amount of toxin that flocculates one unit of an international reference antitoxin) the culture is centrifuged to remove the bacteria. Afterwards, 0.75% of formaldehyde is added to the supernatant, which are then stored up to 6 weeks at 37 °C. Once fully detoxified, the toxoid is concentrated by ultrafiltration and purified by a combination of salt fractionation and chromatographic methods, and it is then tested for potency, toxicity, sterility, and reversibility (BAE, 2011). Recently, Möller and colleagues (2019) uncover the presence of various distinct

proteins in the vaccine besides DT; therefore, the filtration process is not specific for the toxin and other *C. diphtheriae* proteins are selected. This indicates that they may induce antibody formation upon vaccination and contribute to the reduction of the bacterium fitness (MÖLLER *et al.*, 2019).

The named ‘anatoxin by Roux’ was a revolutionary discovery and marked the beginning of the modern era of toxoid vaccines, a branch of the so called ‘subunit vaccines’ (RAPPUOLI; MALITO, 2014). Diphtheria toxoid is one of the safest vaccines available; however, adverse events have been reported. Local reactions at the site of injection are common (50%) and mild adverse events consist of systemic reactions, such as fever (above 38 °C), irritability (40–75%), drowsiness (33–62%), loss of appetite (20–35%), and vomiting (6–13%) whereas severe reactions are rare. Furthermore, the batches of produced vaccines worldwide can present differences in accordance with the pharmaceutical industry responsible for the manufacture (BAE, 2011).

The average duration of protection from the DTP vaccine is about 10 years being recommended the use of booster doses beyond infancy and early school age (WHO, 2017b, 2018c). The WHO recommendation for diphtheria immunization is to apply an effective primary immunization in infancy and to maintain immunity throughout life. The routine immunization of infants consists of a three-dose vaccination series (i.e. 2-6 months of age with a minimum interval of four weeks between doses) followed by three booster doses (1-15 years of age) (BURKOVSKI, 2014; WHO, 2018c). The diphtheria toxoid-containing vaccines are administered by intramuscular injection (0.5 mL dose) and should be stored at 2–8 °C.

Vaccination has led to significant decreases in diphtheria incidence worldwide, which is linked with a 86% worldwide vaccine coverage (i.e. 116.2 million children received three doses of DTP containing vaccines in 2017) (Figure 15) (WHO, 2018c). Nonetheless, serological studies indicate that a high proportion of adults are susceptible to diphtheria, which suggests the necessity of vaccination programs emphasizing adult immunization (WHO, 2017b).

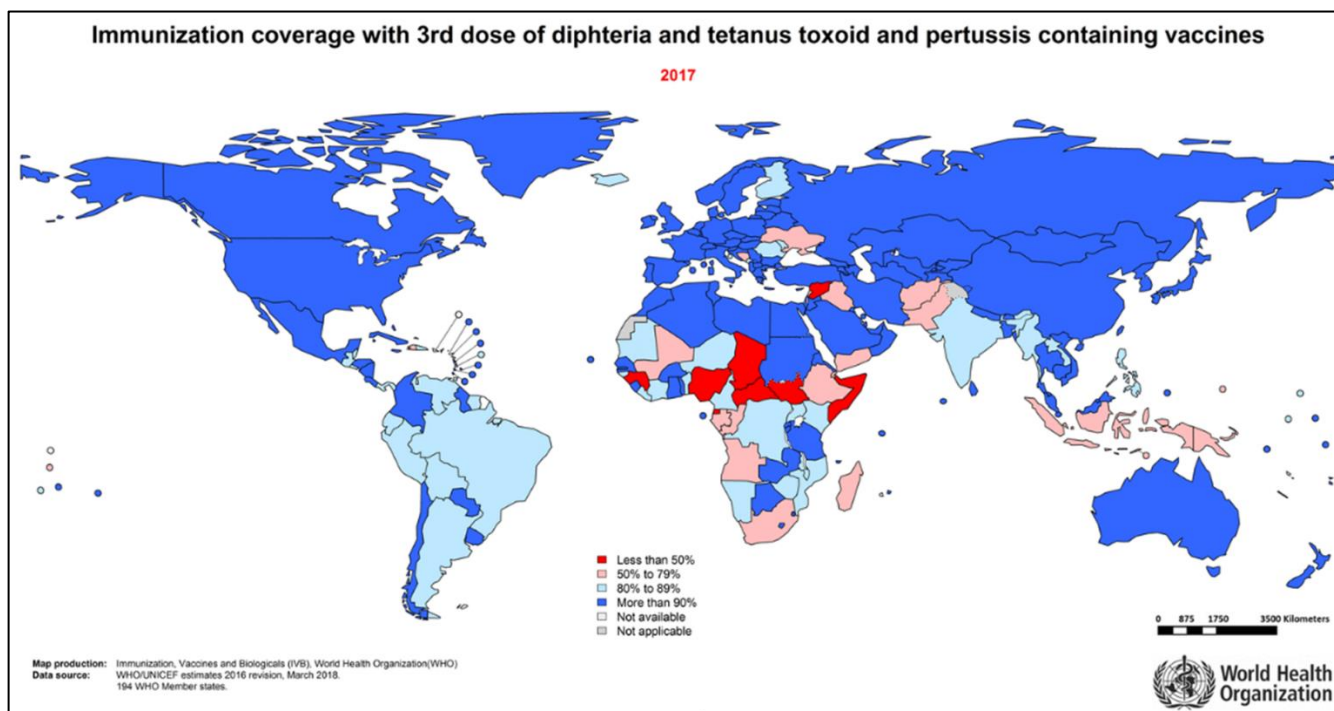


FIGURE 15: Diphtheria immunization coverage (DTP vaccine third dose) in 2017. The data was compiled based on WHO/UNICEF official reports with the last revision on March 2018. The overall reported coverage of the DTP vaccine (third dose) is 86%, with most developed countries with a coverage above 90% whereas some African countries vaccinated less than 50% of infant population against diphtheria, tetanus and pertussis. Figure from https://www.who.int/immunization/monitoring_surveillance/data/wuenic_animated_map.gif?ua=1

2.4. Application of cutting-edge approaches to improve diphtheria management

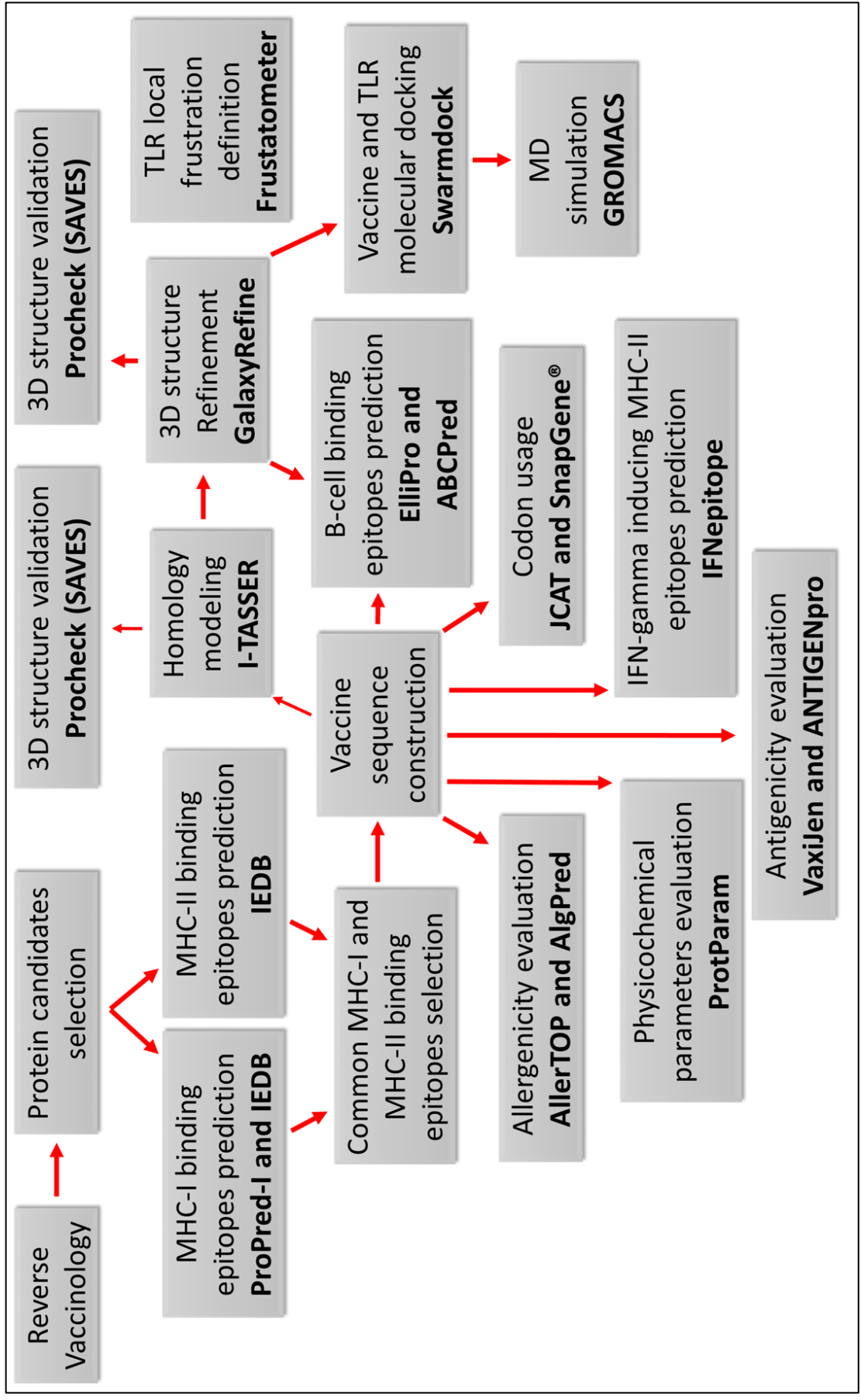
The rising heterogeneity of circulating strains, emergence of multidrug resistance toxigenic variants, and persistence of invasive non-toxigenic strains increased the number of respiratory, systemic and cutaneous infections caused by *C. diphtheriae* strains (BURKOVSKI, 2014). Economic problems and a lack of medical infrastructure/personal (e.g. appropriated storage, necessary sterile material, and trained medical staff) appear to be the main drivers of outbreaks in developing countries whereas in developed countries the inadequate use of booster vaccine is thought to be the principal underlying factor (WHO, 2017b). Furthermore, the chemical inconsistency among toxoid batches and an impure protein profile may affect the vaccine efficacy (BAE, 2011). Therefore, innovative vaccines based on cutting-edge technologies need to be developed to overcome these limitations.

Novel vaccine technologies that make use of recent advances in molecular biology, immunology and bioinformatics will enable us to develop a diphtheria vaccine that can be produced more efficiently, less expensive and more stable. These improvements will lead to a significant increase in diphtheria vaccine coverage in both developing and developed countries.

3. Submitted paper: Design of a broad-spectrum candidate multi-epitope vaccine against Diphtheria: An integrative immunoinformatics approach

Chapter 1 focus on the life-threatening pathogen *Corynebacterium diphtheriae*, the causative agent of diphtheria, and the currently treatments/prevention tactics. Considering the high rate of the disease, we proposed the design of an innovative that would decrease diphtheria cases worldwide. The methodology was based on literature data (HAJIGHAHRAMANI *et al.*, 2017)and designed specifically for this work (Figure 16). The paper was submitted to the Journal ‘Vaccines’ in the special issue ‘Omics and Bioinformatics Approaches to Identify Novel Antigens for Vaccine Investigation and Development’ under the identification: Vaccines-589302.

FIGURE 16: Workflow of the *C. diphtheriae* vaccine construction. The methodology was conceive based on the work performed by Hajighahramani and colleagues (2017) and adapted accordantly with the immunoinformatics tools available and with the data obtained. The complete workflow



Design of a broad-spectrum candidate multi-epitope vaccine against Diphtheria: An interactive immunoinformatics approach

MARIANA SANTANA¹; STEPHANE TOSTA²; RODRIGO KATO²; ARUN KUMAR JAISWAL^{2,3}; SIOMAR C. SOARES³; LUIZ CARLOS JUNIOR ALCANTARA⁴; PREETAM GHOSH⁶; DEBMALYA BARH⁵, VASCO AZEVEDO²; ANDERSON MIYOSHI¹; SANDEEP TIWARI^{2*}

¹Department of Genetics, *Institute of Biological Sciences, Federal University of Minas Gerais (UFMG), Belo Horizonte 31270-901, MG, Brazil*

²Postgraduate Program in Bioinformatics, *Institute of Biological Sciences, Federal University of Minas Gerais (UFMG), Belo Horizonte 31270-901, MG, Brazil*

³Department of Immunology, Microbiology and Parasitology, *Institute of Biological and Natural Sciences, Federal University of Triângulo Mineiro (UFTM), Uberaba 38025-180, MG, Brazil*

⁴Laboratório de Flavivírus, *Instituto Oswaldo Cruz, FIOCRUZ, Manguinhos, Rio de Janeiro 21040-900 Rio de Janeiro, Brazil.*

⁵Centre for Genomics and Applied Gene Technology, *Institute of Integrative Omics and Applied Biotechnology (IIOAB), Nonakuri, Purba Medinipur, West Bengal, India.*

⁶Department of Computer Science, *Virginia Commonwealth University, Richmond, VA-23284, USA*

*Corresponding author: Sandeep Tiwari- sandip_sbtbi@yahoo.com

ABSTRACT

Corynebacterium diphtheria, the etiological agent of Diphtheria was considered as one of the main causes of child death. Although, the Diphtheria toxin-based vaccine is successfully used for decades, outbreaks and isolated cases are still reported worldwide, indicating that new vaccine against the pathogen is required. Peptide-based multi-epitope vaccines can target different strains or microorganisms and show quick immune effects. They are cost-effective, easy to handle, and less prone to contamination and autoimmune and non-allergenic responses. Therefore, multi-epitope strategy is now the trend in developing next-generation vaccines. Here, using a robust bioinformatics, immunoinformatics, and structural biology approach, we identified twenty-two common MHC-I and MHC-II binding epitopes from seven proteins that can be used to construct a multi-epitope vaccine for Diphtheria. *In silico* cloning was done and the resultant putative multi-epitope vaccine was predicted to be stable, non-allergic and may induce B-cell and IFN-gamma based immunity. The tertiary structure of the multi-epitope vaccine stably interacts with TLR2 to induce TLR2 mediated immune reactions. Our results indicate that the designed multi-epitope vaccine may stimulate humoral and cellular immune responses and would be a potential candidate vaccine against Diphtheria.

Keywords: *Corynebacterium diphtheriae*, Diphtheria, immunoinformatics, peptide-based vaccine, multi-epitope prediction.

1. BACKGROUND

Corynebacterium diphtheriae is the etiological agent of Diphtheria, an acute infection of the respiratory tract, skin, and, in rare cases, can reach the soft tissue triggering systemic infections (e.g. bacteremia, endocarditis, hepatic and splenic abscesses, meningitis, osteomyelitis, septic arthritis) [1,2]. The classical or respiratory Diphtheria is a disease previously recognized as ‘strangling angel of children’ due to the acute infection of the upper respiratory tract, which could lead to airway obstruction, suffocation, and death [3]. The blockage is caused by the presence of a wing-shaped pseudomembrane located in the pharynx, tonsils or nose. This illness once stood out as the leading cause of infant mortality, but was significantly reduced due to the World Health Organization (WHO) immunization program that nowadays consists of three primary vaccination doses (2-6 months of age) followed by three booster doses (1-15 years of age) [3,4].

The Diphtheria vaccine mainly entails a detoxified form of the Diphtheria toxin (DT), also known as toxoid, adsorbed onto an alum gel adjuvant (aluminum hydroxide or aluminum phosphate) [5–7]. The vaccine was introduced in the late 1920s as a formaldehyde-toxoid antigen and, since then, only a few modifications were made in the production process [5,8]. In the 1940s, the Diphtheria-Tetanus-Pertussis vaccine (DTP) was commercialized, as studies proven that the combined antigens (i.e. Diphtheria toxoid, Tetanus toxoid and Pertussis antigens) offered an enhanced immunologic response with no interference [2,6,8]. Since 1974, the DTP vaccine is a part of the WHO Expanded Programme on Immunization (EPI).

Despite the high efficacy of the vaccine, outbreaks are still a concern and Diphtheria remains a significant health problem, especially in developing countries [2,4,9,10]. The Diphtheria epidemic that strike Russia in the 1990s attain specially people over 15 years of age (64–82 %) reaching more than 5,000 deaths with proximally 157,000 cases reported [11,12]. In addition, in accordance with WHO, India had 18,350 reported cases during a 5-year period (2011-2015); meanwhile, Indonesia and Madagascar had 3,203 and 1,633 reported cases, respectively [10]. Furthermore, South-East Asia region notifies 55–99% of all *C. diphtheriae* cases each year during this period [10]. Those reported cases are probably underestimated and the actual burden of Diphtheria is likely much greater. The high number of cases can be related with a poor routine vaccination coverage and problems with vaccine storage and delivery, plus the fact that those countries suffer with overcrowding, poverty and underprivileged hygiene [3,13]. Nonetheless, cases of Diphtheria are also a problem in develop countries and was associated with inadequate use of booster vaccine, travels to endemic countries, immigrants, and non-toxigenic *C. diphtheriae* [13–15].

Non-toxigenic *C. diphtheriae* do not contain a functional *tox* gene; however, many others virulent factors were linked to the bacterium capability to cause the disease; like their abilities to adhere to host cells, to survive and keep viable in an intracellular environment, and the aptitude to stimulate the immune system [3]. Non-toxigenic Diphtheria is associated with the cutaneous, respiratory (i.e. persistent sore throats and severe pharyngitis/tonsillitis) and systemic infections (i.e. endocarditis, septic arthritis, splenic abscesses

and osteomyelitis) [13,15–17]. Nowadays, the cutaneous form of the disease is outnumbering the classical Diphtheria, especially considering subtropical areas [3]. Feet, lower legs and hands are common sites of infections and the transmission is mainly from person-to-person by droplets of respiratory secretions and contact with contaminated skin [10]. High infection rates may be associated with the difficulty of healing the skin lesions, which may take from weeks to months or even years [3,18]. The lesions serve as reservoir of the bacterium and further increase the probability of contamination. Additionally, cutaneous Diphtheria is normally associated with the presence of other pathogens; therefore, the clinical manifestations are often variable and challenging to be associated with *C. diphtheriae* [3].

C. diphtheriae comparative genomic analysis revealed that thereabout 60% of the genome is apparently conserved among strains; however, the accessory genome varies greatly and many genes can be categorized as strain-specific [19–21]. An example of this diversification is observed in the *C. diphtheriae* pili, a well-known adhesion factor that is covalently attached to the cell wall peptidoglycan and can be divided in three distinct clusters (*spaABC*, *spaDEF*, *spaGHI*) [22–24]. Those genes are not present in all *C. diphtheriae* strains, in fact, the pili structures varies in length (short, long, hair-like protrusions) and number (none to all genes) [22,25]. Furthermore, studies indicates that adhesion and pili are not a coupled processes since the presence of pili does not necessary means high rates of adhesion or the absence a lack of attachment to host cells [25]. Therefore, other proteins play a role in the bacterium adhesion process. In addition, *spa* genes are not only strain-specific but were also reported as cell-type specific since different pili types attach to different cell types in a specific manner [22,24,26].

Fifty-seven genomic islands were described in this bacterium with a diverge distribution between strains [19]. This high number can be attributed by horizontal transfer events pinpointing that the genetic diversity observed in this bacterium can be due to recombination with other bacteria [19,21]. The bacterium genetic diversity probably influence its ability to colonize different cells of the body, which leads to infection in the epithelia and also in deeper organs [25].

Considering that the available vaccine is not particularly effective against non-toxigenic *C. diphtheriae* strains, a broad method of prevention is imperative to avoid further cases of Diphtheria worldwide. Little is known about others *C. diphtheriae* virulent factors and proximally forty-nine percent of the genome is still classified as hypothetical proteins [21]. However, different studies suggested the probable importance of many proteins for *C. diphtheriae* colonization and virulence [19,20,27–30]. Transmembrane and secreted proteins stood out since they may play a role in host-pathogen interactions and pathogenicity [21].

Immunoinformatics approaches are being used as cutting-edge alternative strategies to design effective vaccines. This method combines the knowledge of the genes and proteins involved in the immune response ('immunomics') with bioinformatic technics to predict an organism immune responses against specific molecules [31]. A successful vaccine needs to protect the organism against the target disease; hence, it has to induce a strong cellular and humoral immune response (T cell and B cell immunity) [32]. Whole pathogen or protein-based vaccination is an effective method that produce long lasting immunization; however, they can induce allergenic reactions, have protein impurity problems, among others drawbacks [5,33,34]. Thus, the application of specific antigenic determinants that can trigger an immune response is a valid and safe alternative [34]. In addition, peptide/epitope vaccines are also cost effective (i.e. production based on chemical and synthetic technics), reproducible, stable, and require simple storage conditions [34]. Yet, they present poor immunogenicity, a problem easily fixed through the addition of

adjuvants and/or delivery systems [32–34]. Peptide-based vaccine can also be constructed with a variety of epitopes (multi-epitope vaccines) from different organisms (i.e. combine vaccine) or even multiples proteins from the same organism (i.e. heterovalent vaccine). A multi-epitope vaccine induces a stronger immune response than simple single-epitope since display multiples antigens to the immune system [34]. Therefore, a multi-epitope based vaccine allows better coverage by targeting different strains and life stages of a pathogen. This innovative approach is currently being used to construct multi-epitope based vaccines against *Staphylococcus aureus* [35], *Vibrio cholerae* [36], *Helicobacter pylori* [37], malaria [38], dengue [39], tumors [40], among others.

This work aims to construct a multi-epitope vaccine against Diphtheria that covers *C. diphtheriae* toxigenic and non-toxigenic strains using an integrated bioinformatics, immunoinformatics and structural biology approach.

2. METHODS

2.1. Vaccine candidates selection

An exhaustive literature search was performed to select putative candidates classified as relevant for *C. diphtheriae* infection and common between toxigenic and non-toxigenic strains (except Diphtheria toxin). After protein selection, the interaction/association and additional information (i.e. knowledge on metabolism and various other functions of the cell) between the twenty-three selected candidates were analyzed with KEGG (<https://www.genome.jp/kegg/>). The transmembrane, secreted and three cytoplasmic proteins were selected for further analysis computing twelve proteins.

Furthermore, the complete amino acid sequence of these twelve proteins were retrieved from the Universal Protein Resource (UniProt) at <https://www.uniprot.org/> [97]. The sequences were used for T cell, B cell and IFN-gamma epitope prediction, and the selected epitopes were fused together separated by appropriate epitope/peptide linkers [98].

2.2. Major histocompatibility complex prediction

2.2.1. Prediction of MHC-I binding epitopes

The promiscuous MHC class-I binding peptide prediction server (ProPred-I) (<http://crdd.osdd.net/raghava/propred1/index.html>) and the immune epitope database (IEDB) (<https://www.iedb.org/mhci/>) were used for the MHC-I epitope prediction. Propred-I identify promiscuous potential T cell epitopes in an antigenic sequence via matrices for 47 MHC Class-I alleles, proteasomal and immunoproteasomal models [99]. The accuracy of ProPred-I is 75% at default binding threshold (4%). IEDB search epitopes based on experimental assays characteristics (epitope source, structure process environmental) and epitope information (allergens, infectious and autoimmune diseases, transplantation and cancer) [100]. This server predicts and analyses epitopes from T cell and B cells based on literature

data. For class I binding predictions, the IEDB recommended methods were applied, artificial neural network (ANN), stabilized matrix method (SMM), and scoring matrices derived from combinatorial peptide libraries (Comblib).

The ProPred-I and IEDB search for MHC class-I employed 9-mer size epitopes that bind with high affinities to mouse MHC-I alleles (H-2Db, H-2Kb, H-2Kd, H-2Kk, H-2Ld). The cutoff applied for ProPred-I was the top four epitopes of each mouse allele, and for IEDB, the epitopes that presented the percentile rank below 3% (default parameters) were selected for further analysis.

2.2.2. Prediction of MHC-II binding epitopes

MHC-II epitopes identification was carried out with IEBD server (<https://www.iedb.org/mhcii/>) using a different but similar approach that was applied to the MHC-I epitope prediction. The analysis was performed using the IEDB MHC-II epitope predictions default parameters and the mouse MHC-II alleles (H2-IAb, H2-IAd, H2-IEd) were selected for the peptide binding with an epitope length of 15-mer and the applied cutoff was the percentile rank below 3%.

2.3. Vaccine construction

The epitopes selected from the seven proteins (DIP1281, DIP0411, DIP0733, DIP1084, DIP2193, DIP2379, and DIP0222) used in the final vaccine construction were joined together by appropriated linkers. AAY were used between the MHC-I epitopes and GPGPG between MHC-II epitopes. The order of the proteins and epitopes were chosen randomly and the same protein order were applied for both MHC types, wherein the class I epitopes were place first followed by the classes II epitopes. No peptide adjuvant was added to the sequence.

2.4. Homology modeling

The Iterative threading assembly refinement (I-TASSER) server (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) was used for the prediction of the protein three-dimensional (3D) structure. I-TASSER homology modeling is based on multiples threading alignments. First, the software matches the sequence with known proteins structural models from the Protein database (PDB) (<https://www.wwpdb.org/>) to identify structural templates. Then, several models are constructed by fragment assembly simulations via *ab initio* folding and the structures with the highest 3D models score and with the lowest free energy states are selected. The hydrogen-bonding network of the models are then optimized and the final models are achieved after refinement of the global topology and removal of steric clashes [101,102]. The confidence score (C-score) is used to calculate the quality of the models and is based on the significance of the alignments (structure assembly simulations) and the quality of the structure; therefore, the highest scores correspond to the highest model accuracy.

2.5. Refinement and validation of the tertiary structure

The GalaxyRefine server (GalaxyWeb) (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) was implemented to refine the selected 3D structure model predicted by I-TASSER. This method can improve global and local structure accuracy by first rebuilding all side-chain conformations and then refining the models using two repetitive relaxations methods, the mild relaxation and the aggressive relaxation, to eliminate putative perturbations [103]. Structure perturbations in the side chains are corrected by the mild refinement, meanwhile secondary structure elements and loops are adjusted by the aggressive refinement. Model 1 correspond to the lowest energy model generated by the mild relaxation, and models 2 to 5 correspond to the largest clusters generated by aggressive relaxation.

Subsequently, the PROCHECK program of SAVES 5.0 server (<http://servicesn.mbi.ucla.edu/SAVES/>) was used to check and validate the stereochemical quality of 3D protein structure through the Ramachandran plot. The Ramachandran plot report the energetically allowed and disallowed regions in the protein structure, which grant the validation and selection of the best 3D model in accordance with the empirical distribution of the amino acids. The validation is based on the dihedral angles, which are used to specify the molecular conformation. The torsion angles phi-psi determine the stability of the amino acid residues in protein structure [104,105].

2.6. B cell epitope prediction

The immunoinformatics analyses of B cell prediction were performed using the multi-epitope vaccine construct FASTA format or the 3D selected model (pdb file) obtained from the GalaxyRefine server.

The identification of linear B cell epitopes was performed by ABCpred (<http://crdd.osdd.net/raghava/abcpred/index.html>) and discontinuous or discontinuous B cell epitopes by ElliPro (<http://tools.iedb.org/ellipro/>). ABCpred predict B cell epitopes using artificial neural network using BCIPEP B cell epitope database. The dataset consists of 2479 continuous epitopes, including 654 immunodominant, 1617 immunogenic epitopes obtained from 700 unique experimentally proved continuous B cell epitopes from virus, bacteria, protozoa and fungi [106]. ABCpred default settings were used to predict epitopes with a threshold value of 0.8. ElliPro predicts discontinuous epitopes based on a protein antigen's 3D structure, which is associated with the Protrusion Index (PI) and the distance R [107]. PI is based on the location of the residue, within or outside the ellipsoid, where high percentages of residues residing in the ellipsoid are associated with greater solvent accessibility. The epitopes are clustered based on the distance R that correspond to the distance between residue's centers of mass and the value of R is directed associated with the score of the discontinues epitopes being predicted. The ElliPro default settings were used to predict epitopes with a PI value over 70%.

2.7. Interferon-gamma inducing prediction

INF-gamma (INF- γ) is a cytokine release from CD4+ T cells critical for the control of pathogens; therefore, the ability to induce this cytokine was measured. The IFNepitope server (<http://crdd.osdd.net/raghava/ifnepitope/predict.php>) predicts INF- γ induction from MHC-II based on IEDB database and an INF- γ dataset (inducing and non-induced epitopes). The dataset uses the Motive based model (MERCI) and Support Vector Machine based models (SVM) to discriminate INF- γ inducing and non-induced peptides through data analysis and pattern recognition [108]. The MHC-II epitopes predicted by IEDB (section 2.2.2.) were used to predict the protein capability to induce INF- γ . The default parameters were applied for the analysis.

2.8. Vaccine properties evaluation

2.8.1. Allergenicity evaluation

AllerTOP v2.0 (<http://www.ddg-pharmfac.net/AllerTOP/>) and AlgPred (<http://crdd.osdd.net/raghava/algpred/>) servers were used to analyze the allergenicity of the vaccine construction. The default parameters of the servers were used for the analysis.

AllerTop classify a protein sequence as allergenic or non-allergenic based on a dataset of 2427 known allergens from different species and 2427 non-allergens. The server employ the auto cross covariance (ACC), a protein sequence mining method that uses the amino acids properties (e.g. hydrophobicity, molecular size, helix-forming propensity) to discriminate the peptides [109].

AlgPred analysis were based on five approaches (i.e. IgE epitope and PID, MEME/MAST motif, SVM method based on amino acid composition, SVM method based on dipeptide composition, Blast search on allergen representative peptides, hybrid) to predict the allergenicity of the peptide sequence [110]. IgE epitope search rely on the query of known IgE epitopes. MEME/MAST motif uses MEME matrices. SVM method based on amino acid composition or dipeptide composition, and blast search on allergen representative peptides (ARPs) uses a database of know protein sequence of allergens and non-allergens to query a search. The hybrid approach (SVMc + IgE epitope + ARPs BLAST + MAST) assign a protein sequence as allergen if any one of this mentioned method predicts it as allergen.

2.8.2. Antigenicity evaluation

The antigenicity evaluation of the final multi-epitope vaccine sequence was performed using VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) and ANTIGENpro (<http://scratch.proteomics.ics.uci.edu/>). The default parameters of the servers were employed for the analysis.

VaxiJen is a server for alignment-independent prediction of protective antigens based on the physicochemical properties of proteins via an ACC alignment-free approach [111]. The server has a dataset of 100 known antigens and 100 non-antigens from different sources; like bacterial, viral and tumors protein. ANTIGENpro is a method applied from the server Scratch. This method implement data obtained

from protein microarray analysis and predicts protein antigenicity via a sequence-based, alignment-free and pathogen-independent predictor [112].

2.8.3. Physicochemical parameters

ProtParam server (<https://web.expasy.org/protparam/>) infers the physicochemical properties from the protein sequence and the analysis are based on either compositional data or on the N-terminal amino acid. The parameters include the molecular weight, theoretical isoelectric point (pI), amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and Grand Average of Hydropathy (GRAVY) [113]. The default parameters of the servers were used for the analysis.

2.9. Protein–protein docking between vaccine candidate and TLR2

2.9.1. Definition of the Toll-like receptor 2 local highly frustration region

Proteins are composed of more rigid parts, which are connected by largely minimally frustrated interactions, and low local stability clusters (high local frustration). The polymers minimally frustrated interactions include the folding core of the molecule; meanwhile the highly frustrated areas are normally associated with functional regions (e.g. allosteric transitions and binding sites for protein–protein assembly and recognition). Thus, the Frustratometer server (<http://frustratometer.qb.fcen.uba.ar/>) were applied to identify the highly frustrated regions of the toll-like receptor 2 (TLR2). This online server uses the energy landscape theory to quantify the degree of local frustration manifested in protein molecules.

The TLR2 receptor obtained from the RCSB PDB database (ID 2Z7X) was first edited (i.e. removal of water molecules, the TLR1 molecule, and the ligands PCJ, MAN, BMA, NDG and NAG) by Chimera (<https://www.cgl.ucsf.edu/chimera/>) to guaranty the correct local frustration site and binding to the putative Diphtheria vaccine without foreign interference (data not shown) [114]. Furthermore, Chimera was used for the identification and visualization of TLR2 hydrophobic regions. The data from previously published works were used to confirm the Frustratometer results. [35, 36, 41].

2.9.2. Definition of the molecular docking structure

The molecular docking was performed through the SwarmDock server (<https://bmm.crick.ac.uk/~svc-bmm-swarmdock/submit.cgi>) using the Diphtheria multi-epitope vaccine as the ligand and the edited TLR2 (PDB ID 2Z7X) as the receptor. SwarmDock is a flexible protein-protein docking algorithm that predicts the structure via a combination of local docking and particle swarm optimization (PSO) [115]. PSO optimize the binding energy between the binding partners and the lowest energy member among them.

The workflow consists of three steps: Preprocessing, docking, and post processing. The first step analysis checks and corrects the structures partners (e.g. alternative atoms locations, gaps and nonstandard

residues). The docking step is performed four times for each start position providing a variety of conformational poses. The docking mode can be restrained (selected binding region) or blind (no specific binding region selected). In the last case the algorithm runs proximally 120 start positions; meanwhile, in a restrict docking the server runs the start position based at least on one of the selected residues (exactly or near the binding site). During the post processing step, the binding structures are minimized, scored and clustered together in the case of high degree of conformational similarities [115,116]. Afterwards, the software generates a list of docking poses scored by the energy/mean energy, number of cluster members, and number of contacts. Two different results are generated: Standard and democratic. The democratic models are based in the Integrative Ranking of Protein-Protein Assemblies (IRaPPA) scoring scheme. The IRaPPA approach provides a better quality and ranking of docking models via the enhancing of the atomic modelling of protein complexes. It combines physicochemical descriptors like Ranking support vector machines (R-SVMs) and Schulze voting method to select the ranking docked poses [117]

The molecular docking was performed in the restricted mode consistent with default parameters. The binding region established in the previous step (highly frustrated region) was set for the receptor (amino acids 1-47, 215-330, and 410-549) and the whole vaccine sequence (701 amino acid) was defined as the binding region of the ligand. The top 15 best democratic docking results were visualized and the intermolecular hydrogen bonds (H-bonds) identified by Chimera. The result that present the lowest energy score, the maximum number of contacts made between the residues and the highest number of H-bonds between the vaccine residues and the defined binding region of TLR2 was selected.

2.10. Molecular Dynamics simulation

The best docking structure was applied to predict the docking stability via GROMACS v. 5.1.2 [118]. The parameters set for the molecular dynamics (MD) simulation were in accordance with Hajighahramani and colleagues (2017) [35]. The structure was solvated in a cubic box of TIP3P water molecules. The minimal distance of 14 Å from any edge of the box along with the periodic boundary conditions were applied in all dimensions and the concentration of 150 mM of NaCl, which corresponds to physiological conditions, were added to neutralize the simulation system. The Amber99sb-ildn force field was chosen to determine the intermolecular and intramolecular interactions [35] and the Particle mesh Ewald (PME) method to assess the long range electrostatic interactions and LINCS algorithm to constrain all bonds [119]. The non-bonded cutoff was set in 10 Å. Subsequently, the system was minimized using the steepest descent algorithm and the canonical ensemble (NVT) was applied in MD simulation at 300 K during 400 ps. The equilibration phase continued employing the isobaric–isothermic ensemble (NPT) run at 300 K and 1 bar pressure. In the final step, the MD production run was performed for 30 ns. The used integration time step was 2 fs in the MD simulation. For analysis phase, the output files were saved every 2 ps.

2.11. *In silico* cloning

CodonAdaptationTool (JCAT) server (<http://www.jcat.de/Start.jsp>) was employed to evaluate the cloning and expression of the Diphtheria multi-epitope vaccine within the expression vector by adjusting the codon usage of the protein to the typical codon usage of the host. Through reverse translation, the server generates cDNA sequence and afterwards optimizes the codon in accordance with the most abundant tRNA species

of the selected host (codon usage) [120]. JCAT back translate to DNA favoring the most frequently codon consistent with the highest relative adaptiveness for the amino acid of a given organism from a reference set. The vaccine protein sequence was adapted to *Escherichia coli* (strain ATCC 27325 / DSM 5911 / W3110 / K12). Codon optimization evaluates the sequence in accordance with the codon adaptive index (CAI). CAI evaluate the similarities between a given protein coding gene sequence with respect to a reference set of genes [121,122]. The server quantitative index optimizes the gene sequences consistent with their expression levels, in a range from zero to one, wherein one indicates that both codon usages are the same. The high value indicates high expression of gene with an optimal value calculated from 0.8 to 1. In addition, the software evaluated the GC content that has an optimum value ranging from 30 to 70%.

Afterwards, the SnapGene® software were used to visualize and simulate DNA cloning manipulation (SnapGene software from GSL Biotech; available at snapgene.com). The DNA vaccine sequence was inserted into the plasmid pET 28a(+) via the restriction enzymes *Bam*HI and *Eco*RI. Their restriction site is present in the vector (multiple cloning sites) and was inserted at the vaccine DNA sequence flanking regions (3'OH and 5'PO₄, respectively).

3. RESULTS

3.1. Twelve Diphtheria proteins for candidate vaccines

The Diphtheria multi-epitope vaccine is composed of a series of MHC-I, MHC-II and B cell epitopes selected from appropriate protein candidate via bioinformatic tools. Twenty-two core genome proteins (i.e. common between toxigenic and non-toxigenic strains) and one proteins from the accessory genome (i.e. Diphtheria toxin) were initially selected based on two pan genomic papers [21,28]. From those proteins, twelve (DIP0222, DIP0411, DIP0443, DIP0558, DIP0733, DIP0979, DIP0981, DIP0983, DIP1084, DIP1281, DIP2193, and DIP2379) were elected for further analyzes, in accordance with the proteins function, cellular location and functions (Table 1). Although the *tox* gene is not present in non-toxigenic *C. diphtheriae*, DIP0222 were added to the analysis bearing its importance to the disease.

Protein codes and node	Annotation	Functions	Cellular component	Pathways	Pan genome outcome
DIP0105 BioB	Biotin synthase 1	MF: 2 iron, 2 sulfur cluster binding, 4 iron, 4 sulfur cluster binding, biotin synthase activity, iron ion binding BP: biotin biosynthetic process	Cytoplasm	Biotin metabolism	Core genome

DIP0222 Tox	Diphtheria toxin	MF: identical protein binding, NAD⁺-diphthamide ADP-ribosyltransferase activity, toxin activity	Secreted	Pathogenesis/signaling and cellular processes	Accessory genome
DIP0261 RecR	Recombination protein RecR	MF: DNA binding, metal ion binding BP: DNA recombination, DNA repair	Cytoplasm	Homologous recombination	Core genome
DIP0411	Thiol-disulfide isomerase/thioredoxin/ Putative electron transport related protein	MF: antioxidant activity, oxidoreductase activity BP: cell redox homeostasis	Secreted	Oxidation-reduction process	Core genome
DIP0443	Putative surface-anchored membrane protein	Unknown	Transmembrane	Unknown	Core genome
DIP0524 RpsH	30S ribosomal protein S8	MF: rRNA binding, structural constituent of ribosome BP: translation	Cytoplasm	Ribosome/ Translation process	Core genome
DIP0558	ESAT-6-like protein EsxB CFP-10-like protein	Unknown	Secreted	Unknown	Core genome
DIP0672	Putative uptake hydrogenase small subunit	MF: 3 iron, 4 sulfur cluster binding, 4 iron, 4 sulfur cluster binding, ferredoxin hydrogenase activity, metal ion binding	Ambiguous	Oxidation-reduction process	Core genome
DIP0733	Hypothetical protein/ UPF0182 protein DIP0733	Unknown	Transmembrane	Unknown	Core genome
DIP0750 SmpB	SsrA-binding protein/ tmRNA-binding protein	MF: RNA binding BP: trans-translation	Cytoplasm	Trans-translation process	Core genome
DIP0939 GlpX	Fructose 1,6-bisphosphatase II	MF: fructose 1,6-bisphosphate 1-phosphatase activity, metal ion binding BP: gluconeogenesis, glycerol metabolic process	Cytoplasm	Carbohydrate and glycerol metabolic process	Core genome
DIP0979 DapD	2,3,4,5-tetrahydropyridine-2,6-carboxylate N-succinyltransferase	MF: 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase activity, magnesium ion binding BP: diaminopimelate biosynthetic process, lysine biosynthetic process via diaminopimelate	Cytoplasm	Cellular amino acid biosynthetic process	Core genome
DIP0981	Putative succinyltransferase	MF: 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase activity	Cytoplasm	Biosynthesis of amino acids	Core genome
DIP0983 YvdD_2	Cytokinin riboside 5'-monophosphate	MF: Hydrolase activity BP: cytokinin biosynthetic process	Cytoplasm	Purine and pyrimidine metabolism	Core genome

	phosphoribohydrolyase				
DIP1084	Putative iron ABC transporter membrane protein, FecCD-family	MF: transporte activity	Transmembrane	ABC transporters (Iron complex transport system)	Core genome
DIP1099 IlvN	Acetolactate synthase small subunit	MF: acetolactate synthase activity BP: branched-chain amino acid biosynthetic process	Cytoplasm	Biosynthesis of amino acids, antibiotics and secondary metabolites	Core genome
DIP1235 CobM	Precorrin-4 C(11)-methyltransferase	MF: precorrin-4 C11-methyltransferase activity BP: cobalamin biosynthetic process, oxidation-reduction process	Cytoplasm	Porphyrin and chlorophyll metabolism	Core genome
DIP1257 HisE	phosphoribosyl-ATP pyrophosphatase	MF: ATP binding, phosphoribosyl-ATP diphosphatase activity BP: Histidine biosynthetic process	Cytoplasm	Biosynthesis of amino acids	Core genome
DIP1281	Cell wall-associated hydrolases Putative invasion-associated protein	Unknown	Secreted	Cell wall/membrane/envelope biogenesis	Core genome
DIP1339 NusB	Transcription antitermination protein NusB/ N utilization substance protein B	MF: RNA Binding BP: DNA-templated transcription, termination, transcription antitermination	Cytoplasm	Ribosome biogenesis	Core genome
DIP2193 Csp1	Putative secreted antigen	Unknown	Secreted	Unknown	Core genome
DIP2379 YidC	Putative inner membrane protein translocase component YidC	MF: membrane insertase activity	Transmembrane	Bacterial secretion system/ Quorum sensing	Core genome

Table 1: List of *C. diphtheriae* proteins selected from the literature. Twenty-three proteins from *C. diphtheriae* considered important for colonization and/or pathogenesis, in accordance with literature data, were initially picked for analyses. Afterwards, twelve proteins (bold) were selected for MHC epitope binding predictions.

3.2. Immunoinformatics based identification of MHC-I and MHC-II binding epitopes from target proteins

The MHC-I and MHC-II epitopes prediction were performed for the selected twelve proteins (DIP0222, DIP0411, DIP0443, DIP0558, DIP0733, DIP0979, DIP0981, DIP0983, DIP1084, DIP1281, DIP2193, and DIP2379).

The top four higher peptide score of MHC-I binding epitopes and the MHC-I binding epitopes with a percentile rank below 3% from ProPred-I and IEDB, respectively, were compared and the common MHC-I epitopes between them were selected (Table S1) for MHC-II comparison. Concurrently, the MHC-II binding epitopes with a percentile rank below 3% from IEDB were tabbed (Table S2) for MHC-I comparison. All twelve proteins presented MHC-I epitopes; however, only eleven (DIP0222, DIP0411, DIP0443, DIP0733, DIP0979, DIP0981, DIP0983, DIP1084, DIP1281, DIP2193, and DIP2379) proteins presented MHC-II epitopes.

The MHC-I and MHC-II predicted epitopes of the eleven proteins were compared and the epitopes that overlapped with 9 to 6 common sequential amino acids were elected (Table 2). Seven transmembrane or secreted proteins (DIP0222, DIP0411, DIP0733, DIP1084, DIP1281, DIP2193, and DIP2379) with overlapping MHC-I and MHC-II epitopes were selected for the final vaccine construction.

Protein	MHC-I binding epitope	MHC-II binding epitope
DIP0222	TEEIVAQSI	IVAQSIALSSLMVAQ
DIP0411	REVTAQDLI	LREVTAQDLIKVIDS
	PPFKTAAQL	YPSIYDPPFKTAAQL
	YPSIYDPPF	YPSIYDPPFKTAAQL
DIP0733	YGPVIASAT	KEPRIYYGPVIASAT
	AEALSQVGI	GYAPTIAEALSQVGI
	VLPAKIVLL	NAVLPKIVLLVISA
	SYTDINAVL	GSYTDINAVLPAKIV
	FYAFTLPAL	DYGFYAFTLPALRLV
	IPAISTVLM	VTKDLRIPAISTVLM
	LAIGNAWPI	VLMIVSSLAIGNAWP
	KGVHRFLVV	VQYRAAVEKGVHRFL
DIP1084	YPQEQVWAV	YPQEQVWAVVKAHLA
	AGPDAAAL	QMDALAAGPDAAAL
	AFSALASFL	SSAFSALASFLVFKG
DIP1281	AATPNHASL	AAALIEAATPNHASL
	VSPVRWSGM	VSPVRWSGMSPYAVR
	MSPYAVRLI	PVRWSGMSPYAVRLI
	LYAFAGVGI	GLTLYAFAGVGIALP
DIP2379	IALVAAPLI	FTRTDIALVAAPLIV
DIP2193	AESELGFPV	KKYREMKTAESELGF
	SPSTGAHAL	SGRAYWSPSTGAHAL

Table 2: Elected MHC-I and MHC-II binding epitopes. List of the seven *C. diphtheriae* proteins selected for the multi-epitope Diphtheria vaccine with the correspondent MHC-I and MHC-II binding epitopes

3.3. Construction and design of multi-epitope peptide vaccine

The final vaccine consists of 22 MHC-I and 22 MHC-II epitopes from seven proteins that were joined using proper linkers (AAY between MHC-I epitopes and GPGPG between MHC-II epitopes) (Table 2). The schematic of the vaccine construct, with a final length of 701 amino acids, is shown in Figure 1.

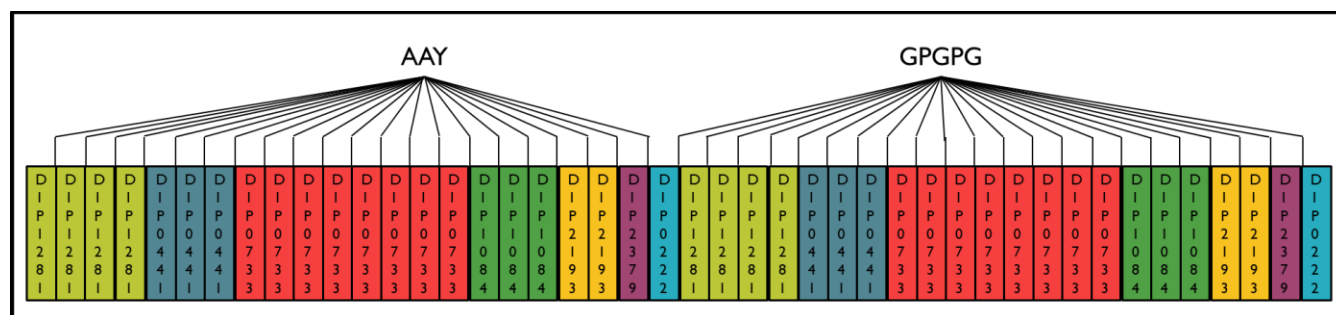


Figure 1: Vaccine construction schematics. The epitopes selected by ProPred-I and IEDB were linked together by appropriated linkers. AAY amino acid sequences were added between MHC-I type epitopes and GPGPG between MHC-II epitopes.

3.4. Homology modeling of the multi-epitope candidate vaccine

The tertiary structure of *C. diphtheriae* multi-epitope vaccine was predicted using I-TASSER. The server predicted the top five models (data not shown) and the predicted model with the highest C-score (-0.79) was selected, which is an indicative of the most reliable prediction (Figure 2A). The selected model corresponds to a state of lower free energy and higher confidence. The predicted TM-score is proximally 0.54, which indicates that the model probably has a correct topology. This measurement signposts the structural similarity between two structures (native and predicted), thus the quality of the modeling.

To further evaluate the quality of the selected 3D structure, the Ramachandran plot generated by SAVES 5.0 server using Procheck. Ramachandran plot was used to measure the residue angles and the steric collisions between atoms. The plot visualizes the amino acid residues in the protein structure and shows the residues placed at allowed and disfavored positions (Figure 3A). The plot revealed that 73.7% of the amino acids were set in most favorable positions. The result indicates that the 3D structure requires further refinement to be a better quality structure.

3.5. Refinement and validation of the tertiary structure of the multi-epitope candidate vaccine

The model selected by I-TASSER (Figure 2A) was refined using the GalaxyRefine server. This online server performs repeated structure perturbations and relaxation by molecular dynamics simulation. Five refined structure models were generated (data not shown). Structure and aggressive perturbation were applied to the chosen 3D model clusters side-chains, secondary structure elements and loops to improve the dihedral angles ψ and ϕ of the amino acid residues measured in the Ramachandran plot. Afterwards, out of the 5 refined models of the GalaxyRefine the best obtained model with higher percentage of score (90.7%) (Figure 2B) was submitted to the SAVES 5.0 and the Ramachandran plot shown that 88.3% of the residues are in most favorable regions, 8.8% in addition allowed, 0.8% in generously allowed, and 2.1% in disallowed regions (Figure 3). The refinement of the tertiary model significantly improved the amino acid positions and steric clashes. Thus, this model was adopted as the final tertiary structure of the *C. diphtheriae* multi-epitope vaccine and used in the remaining *in silico* predictions performed in this work.

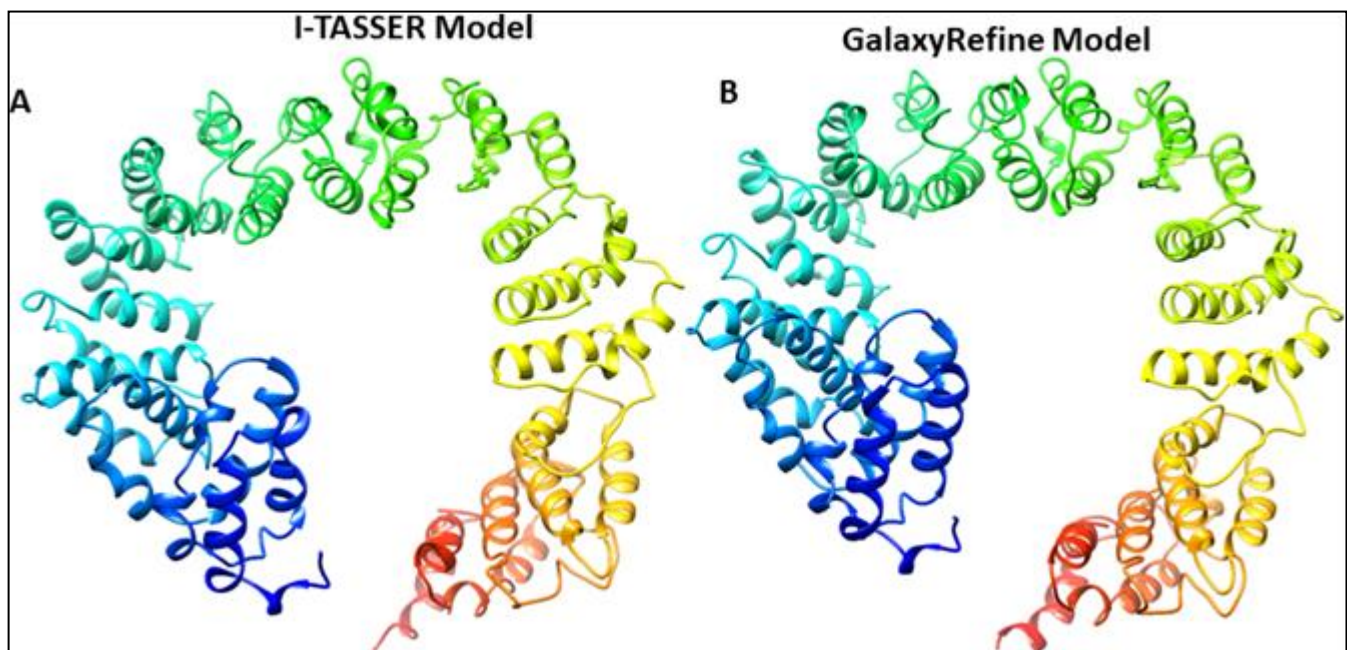


Figure 2: 3D structure of the predicted model from I-TASSER and GalaxyRefine. (A) The I-TASSER model was chosen since it presents the highest C-score (-0.79) between the top five predicted structures (-4.07 ~ -0.79). (B) The model selected by I-TASSER was refined by GalaxyRefine and the structure with the highest Rama favored score (90.7%) was elected.

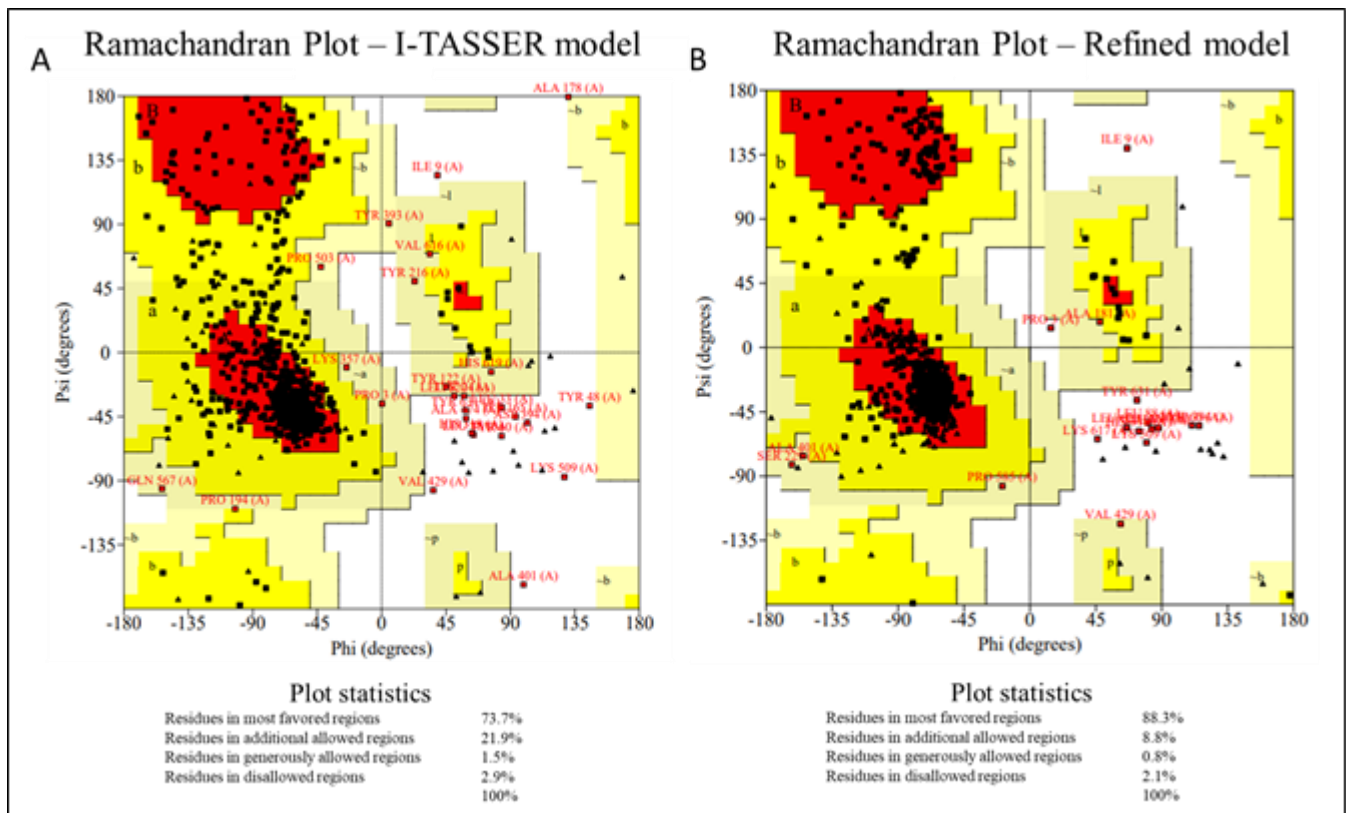


Figure 3: Ramachandran plot of the tertiary protein structure from I-TASSER and after the refinement by GalaxyRefine server. The refined 3D structure with the highest Rama favored score improved by GalaxyRefine was used to calculate the ϕ and ψ backbone dihedral angles of amino acid residues of the protein structure to visualize the energetically allowed/disallowed regions. (A) The 3D structure with the highest C-score (-0.79) predicted by I-TASSER shown that 73.7%, 21.9%, 1.5% and 2.9% represent residues in most favored, additional allowed, generously allowed and disallowed regions. (B) The plot from the refined 3D structure revealed that 83.3%, 8.8%, 0.8% and 2.1% represent residues in most favored, additional allowed, generously allowed and disallowed regions, respectively.

3.6. Multi-epitope vaccine additional Immunoinformatics analysis

3.6.1. Linear and conformational B cell binding epitopes

Linear and conformational B cell epitopes were predicted in the final Diphtheria multi-epitope vaccine using ABCpred and EliPro servers, respectively. ABCpred predicted sixty-seven linear epitopes using the default parameters (Table 3). The threshold setting of 0.5 shows 65.93% accuracy with equivalent sensitivity and specificity using a sixteen amino acid length.

Rank	Sequence	Start position	Score	Rank	Sequence	Start position	Score
1	AALIEAATPNHASLGP	308	0.94	35	ATAAYSYTDINAVLAA	116	0.81
2	APLIVGPGPGIVAQSI	677	0.92	36	LAIGNAWPGPGPGVQY	534	0.80
3	HLAGPGPGSGRAYWSP	619	0.92	37	SAGPGPGKEPRIYYGP	440	0.80
4	PGPGGSYTDINAVLPA	463	0.91	38	AWPIAAYKGVHRFLVV	162	0.80
5	SQVGIGPGPGNAVLPA	417	0.91	39	STVLMGPGPGVLMIVS	517	0.79
6	KIVGPGPGDYGFYAFT	479	0.90	40	EPRIYYGPVIASATGP	448	0.79
7	TYPSIYDPPFKTAAGP	388	0.89	41	PPFKTAAGPGPGGYAP	395	0.79
8	AGVGIALPGPGPLRE	334	0.89	42	HASLAAYLYAFAGVGI	30	0.78
9	PGPGFTRTDIALVAAP	663	0.88	43	QVWAVAAAYAFSALASF	197	0.78
10	ESELGFGPGGFTRTD	656	0.88	44	GVQYRAAVEKGVHRFL	546	0.77
11	PGSGRAYWSPSTGAHA	625	0.88	45	VVAAYAGPDTAAALAA	176	0.77
12	QDLIAAYPPFKTAAQL	54	0.87	46	URLIAAYVSPVRWSGM	6	0.76
13	GVGIAAYREVT AQDLI	42	0.87	47	WPGPGPGVQYRAAVEK	540	0.76
14	RWSGMAAYAATPNHAS	17	0.87	48	QVGIAAYVLPKIVLL	90	0.75
15	PGPGKKYREMKTAESE	643	0.86	49	PVAAYSPSTGAHALAA	224	0.75
16	PQEQVWAVVKAHLAGP	608	0.86	50	ALAAAYPQEQVWAVAA	188	0.73
17	RFLGPGPGQMDALAAG	559	0.86	51	PGPGVTKDLRIPAIST	503	0.71
18	PGPGYPSIYDPPFKTA	363	0.86	52	PGPGPLREVTAQDLI	341	0.71
19	AQSIGPGPGPVRWSGM	258	0.86	53	RPGPGGAAAALIEAATP	301	0.70
20	GPVIASATGPGPGGSY	454	0.85	54	GVLMISSLAIGNAWP	526	0.69
21	GMSPYAVRPGPGGAAA	294	0.85	55	TLPALRLVGP PGVTK	494	0.69
22	APLIAAYTEEIVAQSI	246	0.85	56	YKGVHRFLVVAAYAGP	168	0.69
23	RWSGMSPYAVRLIGPG	269	0.84	57	ATPNHASLGP PGGLT	314	0.65
24	YIPAISTVLM AAYLAI	144	0.84	58	YLAIGNAWPIAAYKGV	156	0.65
25	YGPVIASATAAYSYTD	109	0.84	59	TVLMAAYLAIGNAWPI	150	0.65
26	LGPGPGSSAFSALASF	581	0.83	60	MDALAAGPDTAAALGP	568	0.64
27	PGPGGLTYAFAGVGI	323	0.83	61	INAVLAAIFYAFTLPA	125	0.64
28	URLIGPGPGVSPVRWS	278	0.83	62	YGFYAFTLPALRLVGP	488	0.62
29	PSTGAHALGPGPGKKY	634	0.82	63	YVLPKIVLLAAYYGP	96	0.58
30	YPSIYDPPFAAYAEA	72	0.81	64	YAEALSQVGIAAYVLP	84	0.58
31	TAAQLAAYPSIYDPP	65	0.81	65	AVLPKIVLLVISAGP	428	0.58
32	PGPGGYAPTIAEALSQ	403	0.81	66	LREVTAQDLIKVIDSG	347	0.58
33	KTAAQLGPGPGITYPS	376	0.81	67	ALASFLAAYAESELGF	208	0.56
34	AYAELGFPVAAYSP	215	0.81				

Table 3: Predicted B cell linear epitopes. The ABCPred server was used to predict the B cell binding epitopes via the amino acid sequence of the Diphtheria multi-epitope vaccine. Sixty-seven epitopes were predicted with a score varying from 0.94 to 0.56 (default parameters).

Ellipro predicted eight conformational epitopes with a score or PI varying from 0.778 to 0.587 (Table S3). The percentage of residues located within the ellipsoid (i.e. a quadric surface area with three pairwise perpendicular axes intersect at the center of symmetry) define the PI value. High PI scores are associated with greater solvent accessibility and the value is based on the position of residue's center of mass. Two discontinuous B cell epitopes with a PI equal or superior to 0.7 were selected for further examination and their position in the protein tertiary structure can be visualized at Figure 4. The results signpost that with a PI of 0.7 the ellipsoid includes within 70% of the protein residues with 30% of the protein residues being outside of the ellipsoid.

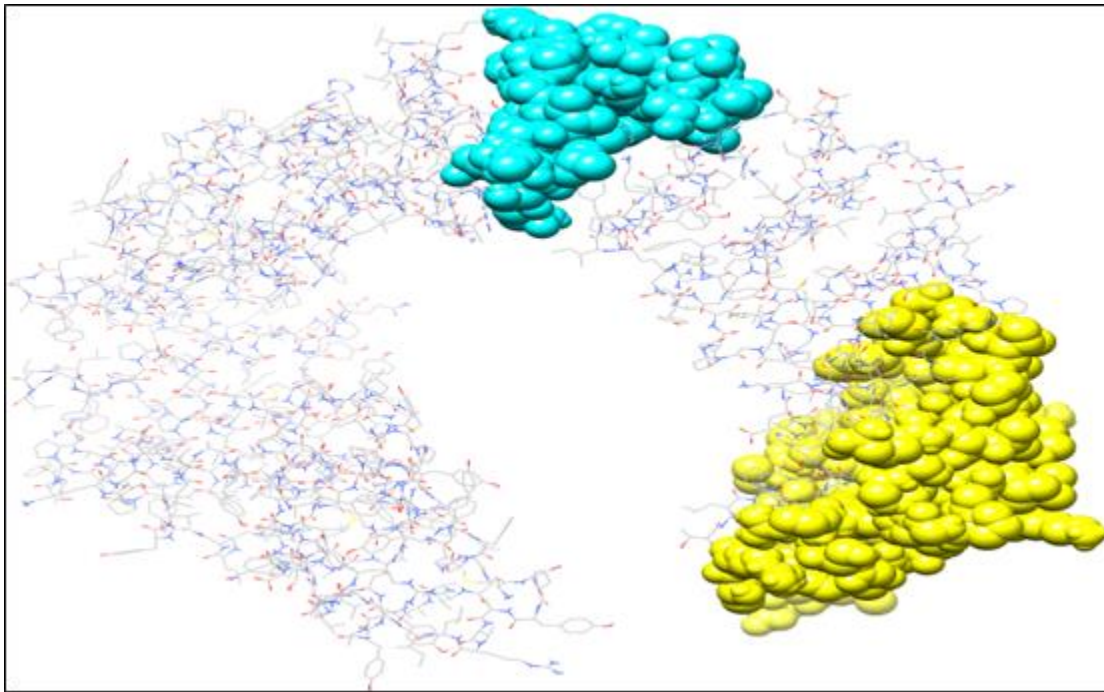


Figure 4: Diphtheria multi-epitope vaccine tertiary structure with the location of B-cell conformational epitopes. Two predicted conformational B-cell epitopes with a PI score of 0.778 (Yellow) and 0.7 (Cyan) visualized in the 3D structure of the vaccine candidate.

3.6.2. Interferon gamma inducing epitopes

An efficacious subunit vaccine against infectious pathogens has to trigger innate and adaptive immune responses. INF- γ is a major cytokine related to the activation of macrophages; hence, the MHC-II binding epitopes of the vaccine candidate were evaluated for the capability to induce IFN-gamma (Table 2). The prediction was performed through the IFNepitope server using the MERCI and SVM methods (hybrid approach). According to the software, ten MHC-II binding epitopes are characterized as IFN-gamma inducing epitopes (Table 4) whereas twelve MHC-II epitopes were predicted as non-inducing.

Epitope sequence	Method	Result	Score
VQYRAAVEKGVHRFL	MERCI	POSITIVE	1
KEPRIYYGPVIASAT	SVM	POSITIVE	0.21554755
GYAPTIAEALSQVGI	SVM	POSITIVE	0.12711736
DYGFYAFTLPALRLV	SVM	POSITIVE	0.061730136
SSAFSALASFLVFKG	SVM	POSITIVE	0.31598445
LREVTAQDLIKVIDS	SVM	POSITIVE	0.49844873
ITYPSIYDPPFKTAA	SVM	POSITIVE	0.22440699
AAALIEAATPNHASL	SVM	POSITIVE	0.40044835
VSPVRWSGMSPYAVR	SVM	POSITIVE	0.17203534
PVRWSGMSPYAVRLI	SVM	POSITIVE	0.42119216

Table 4: Predicted IFN-gamma inducing epitopes. Ten MHC-II binding epitopes capable to induce IFN-gamma (positive) were predicted using IFNepitope server hybrid method (MERCI and SVM).

3.7. Vaccine properties evaluation

3.7.1. Allergenicity estimation of the multi-epitope candidate vaccine

The multi-epitope vaccine was evaluated for its allergenicity capabilities using AllerTOP 2.0 and AlgPred servers. The methods predicted the probability of the protein to be probable non-allergenic and non-allergenic, respectively. AllerTOP 2.0 classified the protein as probable non-allergenic based on the similarities with the non-allergenic protein ‘SCARECROW 1’ (*scr1*) from *Oryza sativa* subsp. japonica (rice) present on the server dataset. AlgPred estimations were performed based on five distinct predictions approaches that the vaccine sequence with known sequences from allergens and non-allergens proteins. AlgPred results indicate that the vaccine does not contain experimentally mapped IgE epitopes, no allergenic motif, no allergens based on the amino acid composition (score of 0.695 with a threshold of -0.4) and potential allergen based on the dipeptide composition (score of 0.464 with a threshold of -0.2),

no allergen representative peptides, and no allergen via a hybrid approach (improved sensitivity and specificity).

3.7.2. Antigenicity estimation of the multi-epitope candidate vaccine

The vaccine antigenicity was predicted by Vaxijen v.2.0 with an overall protective antigen of 0.6688, which indicates that the protein is probable antigenic. AntigenPro was used to further confirm the probability of antigenicity of 0.500000.

3.7.3. Physicochemical parameters estimation of the multi-epitope candidate vaccine

The ProtParam server calculated a variety of physicochemical specifications. The estimated molecular weight of vaccine was proximally 70.47 kDa and the theoretical pI was 8.53, which indicates that the protein is basic in nature. The total number of negatively (Asp + Glu) and positive (Arg + Lys) charged residues were 30 and 33, respectively. The estimated half-life was predicted to be 30 hours in mammalian reticulocytes (*in vitro*), more than 20 hours in yeast (*in vivo*) and more than 10 hours in *E. coli* (*in vivo*). The measured extinction coefficient and absorbance (optical density) of the native protein in water at 280 nm was $112540 \text{ M}^{-1}\text{cm}^{-1}$ and 1.597, respectively. The instability index computed was 36.46, which indicates that the vaccine is stable since the index is smaller than 40. The aliphatic index and GRAVY values were 92.37 and 0.391, respectively. The aliphatic index result indicates that the protein is thermostable and the positive GRAVY value designates that the protein is polar.

3.8. Protein–protein docking between the multi-epitope candidate vaccine and TLR2

TLR2 is a receptor expressed on the surface of immune system cells with the ability to recognize foreign substances (e.g. pathogenic epitopes). TLR2 model structure has 549 residues enumerated from 27 to 575 with the signal peptide (1-26 amino acids) excluded from the crystal protein complex. The tertiary model obtained by GalaxyRefine was used to locate the frustrated regions. The Frustratometer server predicted three highly frustrated areas in the TLR2 structure (Figure 5), which matches with the predicted hydrophobic regions by Chimera (Figure 6). The results present here are in agreement with others docking studies performed with TLR2 [35,36,41]:

- Area 1 correspond to the N-terminal region, ranging from the LRRNT to the LRR1 motifs (amino acids 27 to 73);

- Area 2 match the TLR2 ligand binding domain located in the central region, ranging from LRR8 to LRR12 (amino acids 250 to 356);
- Area 3 correspond to the N-terminal subdomain, ranging from the LRR15 to LRRCT motifs (amino acids 437 to 575);

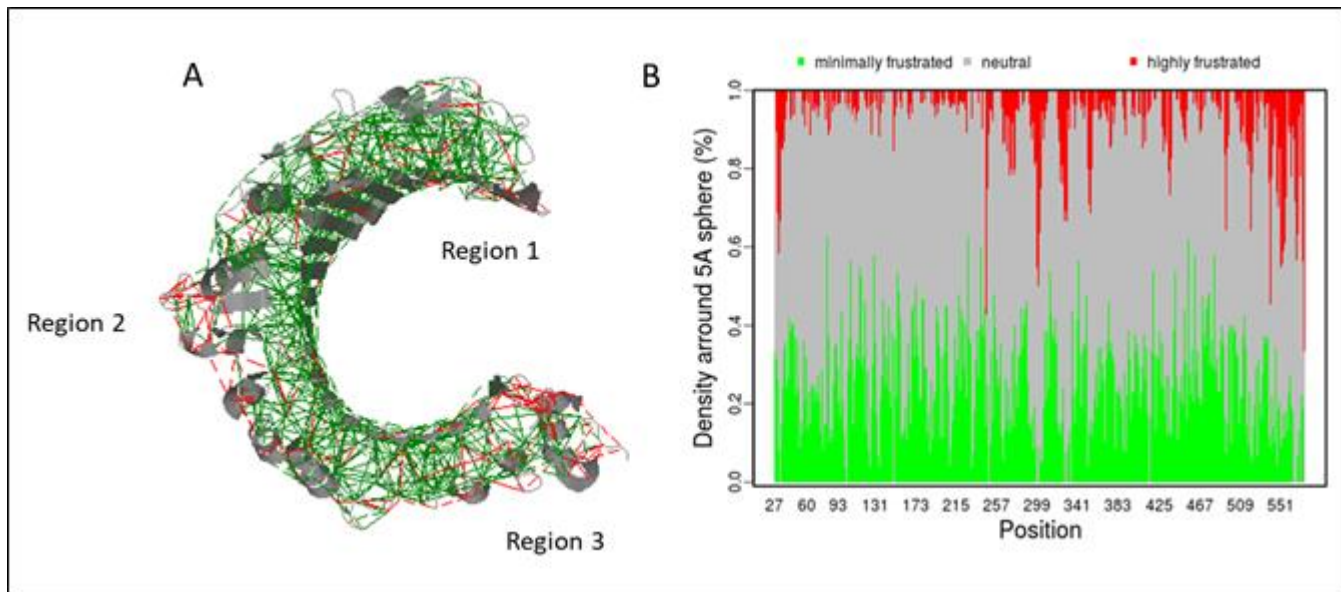


Figure 5: TLR2 frustrated regions. The Frustratometer server quantified the degree of local frustration in the TLR2 structure. The software analyzed the energy distributed in the protein providing insights on its biological behavior. High local frustrated sites are shown in red and indicate biologically important regions (binding sites). Minimally frustrated are marked in green and aids the molecule to constitute a stable folding core, and the neutral regions are shown in grey. Three highly frustrated regions were selected and their location can be visualized in the 3D structure (A) and in a graphic form (B). Region 1 matches the N-term region (amino acids 27-73), region 2 the center domain (amino acids 250-356), and region 3 the C-tem domain (amino acids 437-575).

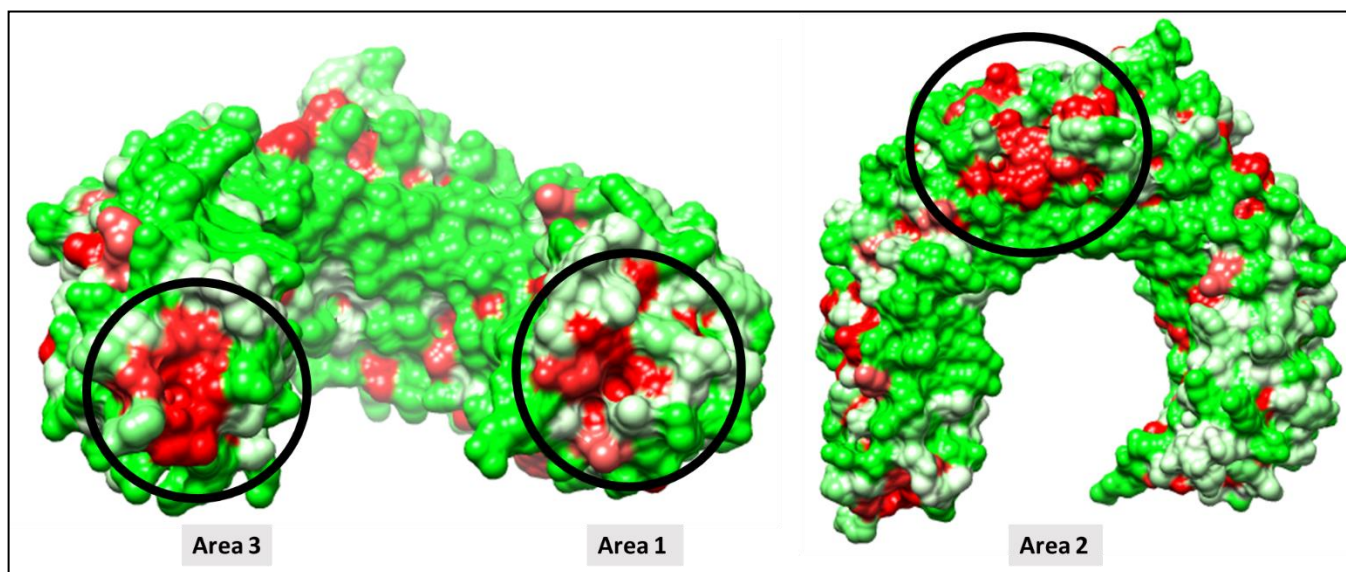


Figure 6: TLR2 hydrophobic regions. Chimera was used to visualize TLR2 hydrophobic regions in accordance with the amino acid sequence. The red regions represent the residues with a higher hydrophobicity (high positive value), the green regions denote hydrophilic regions (negative value), and the white neutral regions. The high hydrophobic regions match the frustrated areas predicted by Frustratometer.

The residue regions displayed by the Frustratometer were selected as the TLR2 binding regions in the restricted docking performed by the SwarmDock server. The probability to obtain the correct solution among the best ten from a restrained docking is higher than blind docking. The edited TLR2 were used as the receptor and the vaccine as the ligand, as result list with 95 clusters were generated with four different forms each. The UCSF Chimera software was used to analyze the top 15 democratic docking models and visualize the H-bonds present between the TLR2 binding sites and the vaccine (Table S4). The most plausible model was selected based on the protein-protein interactions (eight H-bonds located within the binding region), the calculated energy (-43.18) and the total number of contacts (657 contacts between TLR2 and the vaccine) (Figure 7).

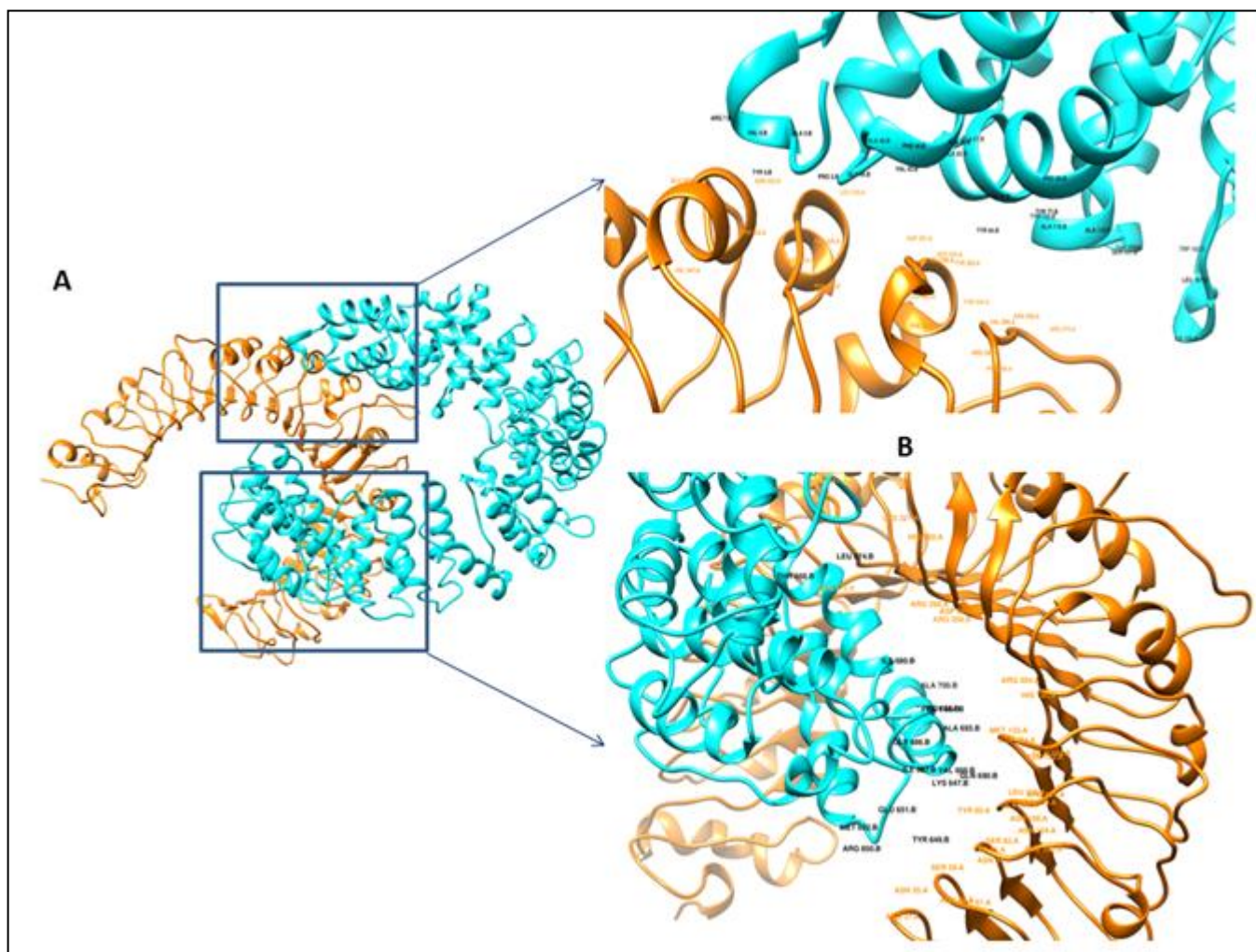


Figure 7: TLR2-vaccine docking structure predicted by SwarmDock. The best dock solution was selected as the most probable docking structure. (A) The ligand (vaccine protein) is shown in cyan and the receptor (TLR2) is shown in orange. (B) The figure showing residues of chimeric protein (vaccine) - receptor (TLR) complex of the frustrated regions (binding sites).

3.9. MD simulation of the multi-epitope candidate vaccine

The simulation was performed for the best-docked structure using Gromacs v. 5.1.2 software to check stable interactions between docked complex. Energy minimization and calculation of pressure, temperature and potential energy were performed. The plot of temperature and pressure evaluation indicates that system maintains 300 K (Figure 8A) for 400 ps. Figure 8B shows fluctuation in pressure with an average value 1 bar for around 400 ps time interval. This phase was performed to check the stabilities and then run the MD production. Two analyses were used to check stability of the MD

simulation result: root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) (Figure 8C and 8D, respectively).

Nowadays, MD simulations can be performed in large biomolecular system over longer time-scales, because there was an increase in computing power and number of X-ray/NMR structures. Previous studies [42–46] have addressed the significance of RMSD between two structures, and it is used to judge conformational similarity. RMSD is used for the large data samples, because it is the fastest way to measure structural similarity in practice.

Figure 8C shows deviations up around 14 ns, but afterwards an increase can be observed in the backbone RMSD plot. At the end of the production MD, RMSD plot shows that error tends to remain constant during the rest of the simulation time. After 5 ns, a variation of RMSD can be seen in 0.3, and this is justified because it is very flexible (many loops area). Thus, we can conclude that the system reaches the steady state after a gradual increase in backbone RMSD during the interval 15-20 ns. The low root mean square deviation value shows the stability of the complex. The RMSD value was around 0.6 at the end of the data set of the MD trajectories. RMSD analysis relative to the first frame in the MD trajectories showed an equilibrium tendency in the system.

On the other hand, figure 8D shows the RMSF of all residues were computed. These fluctuations of most of the residues may indicate that we had a fluctuation in RMSD and the flexibility of complex with many loops area.

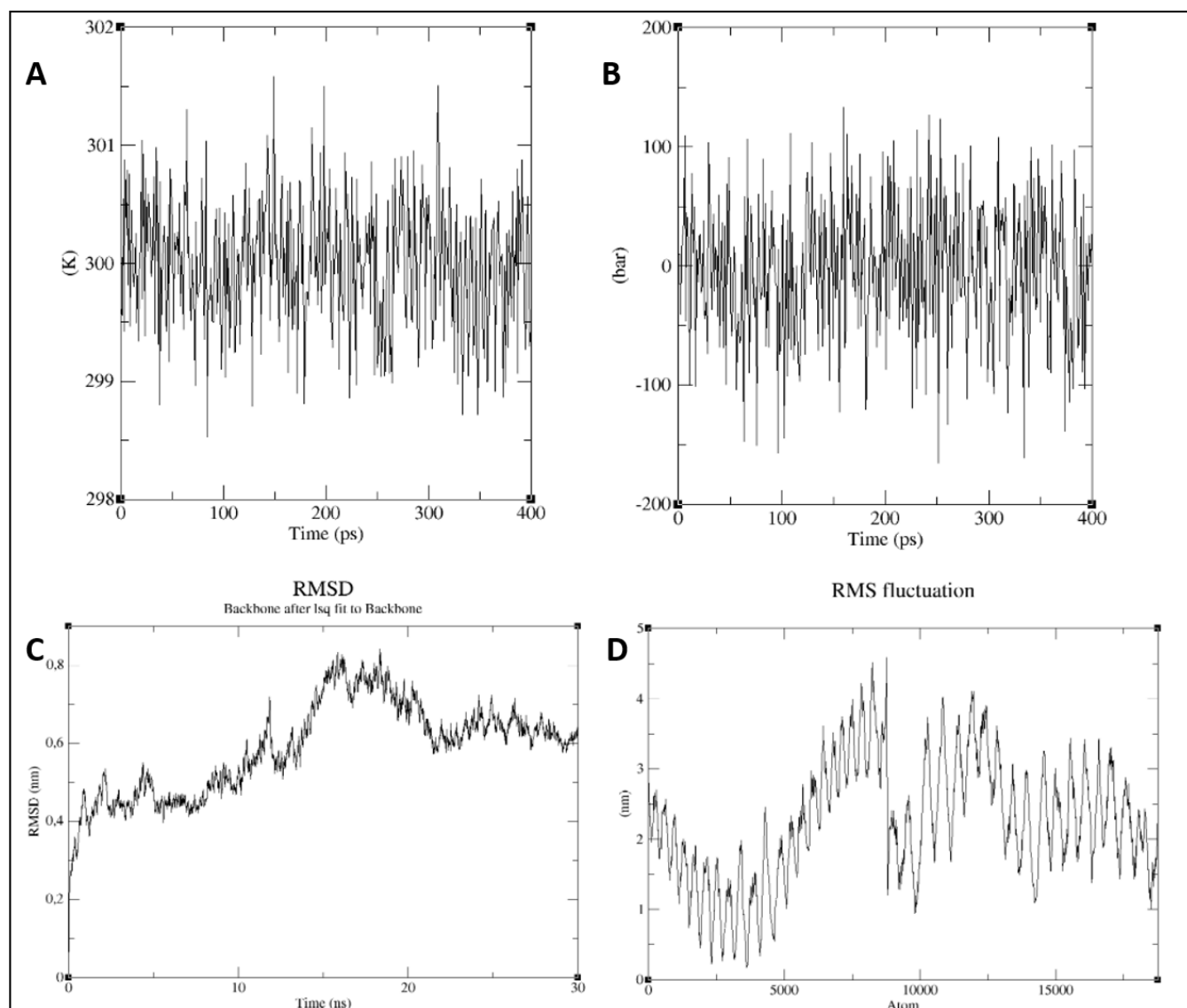


Figure 8: Molecular dynamics simulation of the ligand-receptor complex. (A) Temperature progression plot of ligand-receptor complex shows that temperature of the system reaches to 300K and remains nearly constant around 300K throughout equilibration phase (400 ps). (B) Docked complex pressure progression plot indicates fluctuation of pressure throughout the equilibration phase of 400 ps with an average pressure value around 1 bar. (C) Root Mean Square Deviation of docked complex shows a deviation, which reflects the stable interaction between ligand and receptor molecule. (D) Root Mean Square Fluctuation plot of docked complex side chain fluctuation in plot generates variations which reflects the flexibility of side chain of docked complex

3.10. *In silico* cloning of the multi-epitope candidate vaccine

In silico cloning was performed by the online JCAT server and the SnapGene software, in order to evaluate the cloning and expression of the Diphtheria multi-epitope vaccine within the expression vector of *E. coli* K12. Via reverse translation, JCAT generated the optimized DNA sequence. The GC content was calculated as 58.25% with CAI value of 1.0. The SnapGene server were applied to simulated a cloning strategy using pET28a(+) and the restriction enzymes *Bam*HI and *Eco*RI (Figure 9). The vector presents an N-terminally 6xHis-tagged configuration with a thrombin site, a C-terminal His tagged sequence, and f1 origin being ideal for *E. coli* cloning strategies.

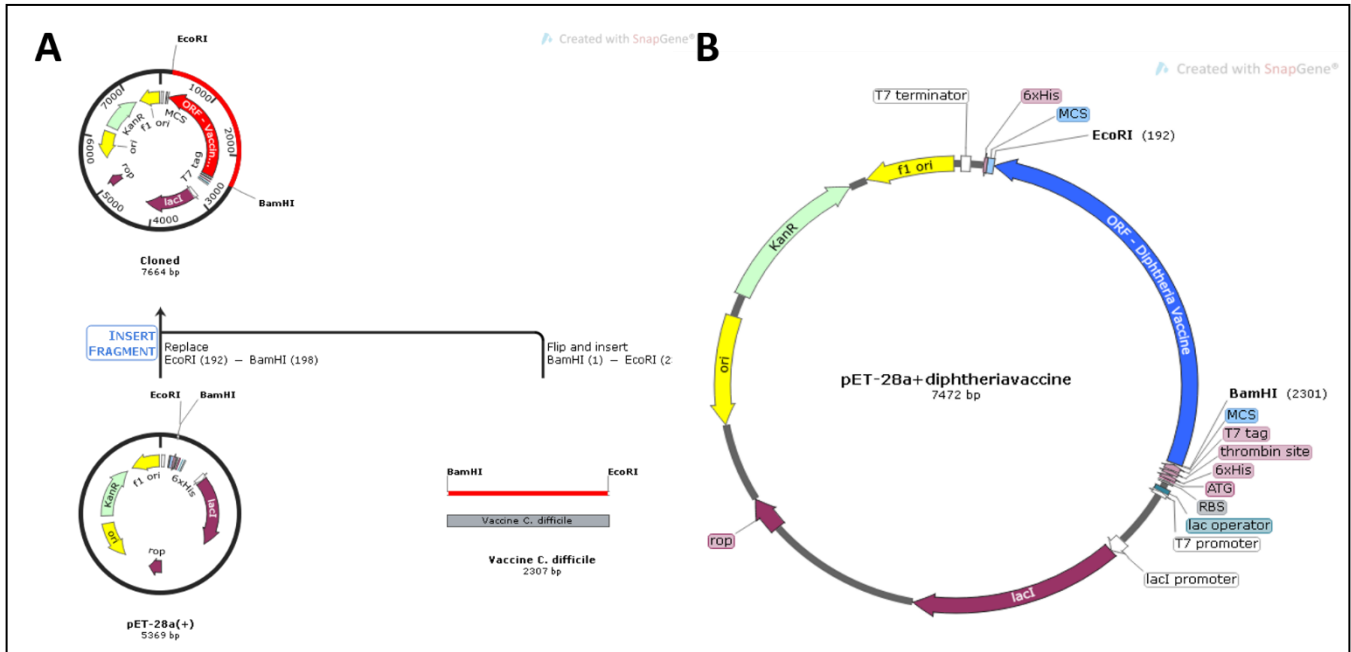


Figure 9: *In silico* cloning thru SnapGene (+). SnapGene software was used to simulate a cloning strategy. (A) The optimized Diphtheria vaccine sequence was inserted into the pET28a(+) vector via the restriction enzymes *Bam*HI and *Eco*RI. (B) The expected map sequence of the pET-28a(+)-Diphtheriavaccine plasmid.

4. DISCUSSION

Diphtheria is an infectious disease caused by the pathogenic bacterium *C. diphtheriae*. This illness was previously considered one of the major causes of child morbidity and mortality although the development of the Diphtheria vaccine significantly reduced *C. diphtheriae* infection rates [5]. Yet, outbreaks and isolated cases are still being reported worldwide. These are largely attributed to a poor vaccination program especially in developing countries, a rising heterogeneity of *C. diphtheriae* circulating strains including the emergence of multidrug resistant bacteria, and the increasing number of non-toxigenic strains [4,10,47].

The available vaccine is produced based on *C. diphtheriae* *tox* gene along with trace quantities of additional *C. diphtheriae* proteins present in the culture supernatant of the PW8 derived toxoid vaccines from Brazil, India, Germany, Bulgaria and Russia [7]. According to the authors, the manufacture processes also select an assortment of protein from toxigenic and non-toxigenic strains including a few with predicted immunogenicity properties. Nonetheless, the number of cases associated with non-toxigenic strains is rising, which still signposts an issue with the currently Diphtheria vaccine.

The Diphtheria toxin (DIP0222) is a secreted exotoxin encoded in the corynephage (β^{tox+} , γ^{tox+} or ω^{tox+}) of toxigenic strains and characterized as the main virulent factor of *C. diphtheriae* [48,49]. The Diphtheria toxin repressor (DtxR) regulates the expression of the *tox* gene in an iron-dependent manner [50]. The toxin is composed of two subunits (A-B toxin) linked by disulfide bonds. The subunit A is the catalytic domain with the ability to inhibit the protein synthesis and to induce DNA internucleosomal cleavage and the subunit B is the binding domain crucial for the penetration of the A subunit in the cell cytoplasm [51]. At the cytoplasm, the toxin prevents the activity of the elongation factor 2 (EF-2), a major factor in the protein synthesis of eukaryotic cells (i.e. protein translation elongation step) [49,51,52]. This exotoxin was widely described by many researchers over the years and *C. diphtheriae* additional virulent traits were overlooking, mostly due to the relevance of DIP0222 to the infection process. These others mechanisms are crucial for the bacterium survival at the host considering that non-toxigenic strains are capable to evade the immune system and persist within macrophages [53,54], and they could be used for the production of a more inclusive vaccine.

C. diphtheriae virulent factors (including their mechanism of action) are poorly described with almost half of the coding proteins describes as hypothetical [21,28,47]. The pan genome study performed with 20 *C. diphtheriae* strains, Sangal and colleagues (2015) reveal that the predicted transmembrane and secreted proteins contribute to proximally twenty-four and eight percent, respectively, of *C. diphtheriae* core genome (i.e. 19% and 14% considering the accessory genome, respectively), with the major portion of the proteins (64%) predicted as cytoplasmic. The transmembrane and secreted proteins represent suitable vaccine targets once they are easily accessible by the immune system (i.e. stimulate an immune response) and generally have critical functions (i.e. influence bacterial survival).

The classical vaccine manufacture implements the whole pathogen (live, attenuated or inactivated) or only a part of the microorganism (subunit vaccine). Both categories raise safety issues associated with contaminations in the production process, the lack of control of protein impurity and post-translation modifications, poor vaccine stability and hazardous side effects [34]. To overcome these drawbacks, a synthetic peptide-based vaccine arises as a key approach capable to offer protection against a pathogen in a customized, safe, simple and cost-effective manner [34,55,56]. A multi-epitope strategy produces a heterovalent vaccine capable to target multiple pathogens/strains or different stages of the microorganism life cycle, decreasing the risk of the microorganism escape the immune system [34,57]. Through this method, a *C. diphtheriae* vaccine can be designed using various antigens from different virulent factors to increase the probability of success aiming toxigenic and non-toxigenic strains.

C. diphtheriae pan genomic and modelome studies described secreted and transmembrane proteins from the core genome that could be compelling vaccine targets [21,28,47]. Immunoinformatics strategy was used to predict MHC-I and MHC-II binding epitopes of three transmembrane (DIP0733, DIP1084, DIP2379) and four secreted (DIP0222, DIP0411, DIP1281, DIP2193) proteins picked to compose the multi-epitope vaccine (Table 1). The genes are linked with pathogenesis (DIP0222), adhesion and cell

wall biogenesis (0733, DIP1281, DIP2193 and DIP2379), iron transport (DIP1084), and oxidation-reduction process (DIP0411).

C. diphtheriae cell envelope acts as a primary line of defense being comprise of four layer: (i) The plasma membrane [i.e. phospholipids and lipid bilayer like lipomannan (LM) and lipoarabinomannan (LAM)]; (ii) the heteropolysaccharide meshwork (i.e. peptidoglycan covalently linked to arabinogalactan); (iii) the mycomembrane (i.e. mycolic acids and glycolipids); and (iv) the outer layer (i.e. free polysaccharides, glycolipids, and proteins) [58]. The corynebacterial cell wall protects the bacterium against salt, oxidative, physical, metal, membrane stress, which aids the colonization and survival (e.g. detergent, antibiotics), acting as adhesion and virulent factors (reviewed at Burkovski, 2018; Ott, 2018). The *C. diphtheriae* lipoarabinomannan-like (CdiLAM) compound acts as an adhesin following the first 30min of infection (initial steps) binding to human respiratory epithelial cells [61]. In *C. diphtheriae*, mycolic acids do not affect phagolysosome (macrophage) maturation, in opposition to other species of the genera where the cell wall component is considered a pathogenic factor [53]. Yet, mycolic acids and glycolipids (lipomannan and lipoarabinomannan) can be recognized by the host immune system via patterns of recognition receptors (e.g. Mincle and TLR2) [53,59,62,63]. These cell wall components can trigger macrophage activation and so they may display an adjuvant effect in the CMNR group (*Corynebacterium*, *Mycobacterium*, *Nocardia*, *Rhodococcus*), including different *Corynebacterium* species [62–69]. Schick and colleagues (2017) disclose that Mincle and TLR2 work in collaboration to induce *C. diphtheriae* inflammatory response via the recognition of mycolate-associated glycolipids.

Pathogenic bacteria internalization is an advantageous step of infection since protects the bacterium from the host immune system cells and antibiotics [27]. The proteins implicated in this process offer the bacterium an opportunity to colonize and spread within the host being proper options for vaccine production.

DIP0733 (initially characterized as non-fimbrial protein 67-72p) is a well-studied *C. diphtheriae* protein important for colonization and virulence [70]. These multi-functional protein has a major role in *C. diphtheriae* persistence/survival at the host being direct linked with the invasion of epithelial cells and internalization/survival within macrophage in the absence of mycolic acids (e.g. trehalosyldimycolates) [27,53,71]. The protein function as a microbial surface component recognizing adhesive matrix molecules (MSCRAMMs) like collagen, fibronectin, laminin and elastin (i.e. constituents of eukaryotic cells surface) [72]. DIP0733 can interact with collagen and fibrinogen (plasma component that is converted into fibrin) and may be linked to the pseudomembrane formation, since insoluble fibrin is one of the component of the semi flexible polymer-like network [73]. *Δdip0733* have impaired adhesion to epithelial cells and decreased internalization (e.g. macrophage, HeLa, HEp-2), which may be linked to a disturbance of the extracellular matrix [27]. The authors also demonstrated the virulent potential of DIP0733 in a *Caenorhabditis elegans* infection model, where 90% of the nematodes survived the mutant strain oppose to 70% killed by the wild-type strain.

DIP1281 is involved in the organization of *C. diphtheriae* outer protein layer, which has a crucial role in the location and function of adhesion and invasion factors [29]. According to Ott and colleagues (2010b), DIP1281 mutant strains exhibit a reduced or lacked the ability to adhere and invade to host cells, which could be associated with the changes in protein patterns and rearrangements of cell surface structures (i.e. formation of bacteria chains) observed in the mutant strains. Moreover, the deletion of DIP1281 cause alterations in the size and shape of the bacteria cell wall without altering the peptidoglycan layer or the

bacteria capability to divide, since the bacterium viability do not decrease, which leads to the reduction of the bacteria adhesion and invasion capabilities [29]. This protein belongs to the NlpC/P60 superfamily, compose of putative proteases and invasion-associated proteins [74]. This family is conserved among corynebacteria (pathogenic and non-pathogenic), and other members were described in *C. diphtheriae*, like DIP1621, an invasion-associated protein that similar to DIP1281 also play a role in adhesion [75,76].

DIP2193 is a putative secreted antigen (Csp1) predicted as a potential mycolyltransferase involved in the cell wall assembly and part of *C. diphtheriae* core genome [21]. A mycolyltransferase can catalyze the transfer of mycolyl residue between mycolyl-trehalose and the free disaccharide (i.e. monomycolate to dimycolate); furthermore, these enzymes are also known to interact with fibronectin (fibronectin-binding proteins - FBP) and elastin, which enhance the bacterium internalization by macrophages (via phagocytosis) [77–79]. DIP2193 has LGFP conserved sequence motif (leucine, glycine, phenylalanine, proline) repeats in the C-terminal region that may be relevant in the maintenance of the cell wall integrity [78,80]. *C. diphtheriae* has two genes encoding potential mycolyltransferases, *csp1* and *dip2194*, with considerable homology with *Mycobacterium tuberculosis fbp* and *C. glutamicum fbp*-like genes [78,81,82]. Puech and colleagues (2000) observed that the disruption of *csp1* from *C. glutamicum* lead to an alteration in the cell wall permeability due to a decrease of corynomycolates cell wall. Moreover, mycolyltransferases can be considered immune-dominant antigens with a significant role in host–pathogen interactions and bacterial survival, which makes this type of enzyme a potential drug/vaccine target [77]. Recently, this protein was identified as a part of the culture supernatant of the PW8 Diphtheria vaccine and the results suggests that DIP2193 has immunogenic properties and may function as an additional antigen [7].

DIP2379 is annotated as a putative inner membrane protein translocase component YidC (integral membrane protein) [21]. Translocases/insertases are essential for bacterial survival since they are required for the insertion of new proteins into membranes (i.e. export from the cytoplasm to the cell wall); hence they are critical for host-pathogen interaction [83]. YidC can function alone or combined with another translocase machinery (e.g. SecYEG), and even can work as integrase or chaperone. There is no information regarding *C. diphtheriae* YidC-like protein; however, this protein type is highly conserved among bacteria [83]. In *M. tuberculosis*, YidC was described as a cell envelope protein important in the regulation of central metabolism and respiration genes with a possible role in ATP production; thus, a gene deletion may cause redox imbalance and affects intracellular survival (*in vitro* and in macrophages) [84,85]. Thakur and colleagues (2018) also demonstrated that YidC levels are modulated by stress conditions and an overexpression could provoke an alteration in the envelope proteome leading to adverse effects on bacterial growth.

Iron is an essential compound for the growth of pathogenic bacteria; yet, in *C. diphtheriae* also has an impact on the production of the Diphtheria toxin (i.e. low availability of iron leads to a high synthesis of the toxin). Mutants defective in iron uptake genes have a low concentration of intracellular iron, even in high iron medium, which result in an increased toxin production [86]. Many iron transport systems were described in *C. diphtheriae* but their specific functions still needs to be determined (reviewed at Schmitt, 2014), including for DIP1084.

DIP1084 is annotated as a putative iron ABC transporter membrane protein (FecCD-like family). Jamal and colleagues (2017) elected this protein as a probable drug target since it is an essential non-host homolog *C. diphtheriae* protein. FecCD-like proteins belongs to the ABC transporter permease protein

family and is characterized by the presence of a Periplasmic-Binding-Protein (PBP)-dependent transport system for iron (III) dicitrate (*E. coli*) [87]. Iron associated transporter systems may have a role in *C. diphtheriae* pathogenesis as observed for similar ABC transporter genes in close related species (*Corynebacterium pseudotuberculosis*) [88], which further ratify the importance of this type of protein to the bacteria and its relevance as a vaccine target.

Exogenous and endogenous sources of redox stress are abundant inside the host especially when pathogenic bacteria encounter the immune system cell, thus the microorganism engage a variety mechanisms to cope with the challenges [89]. During internalization, a few enzymes, like thioredoxins (i.e. small disulphide-containing redox proteins), performs oxidation-reduction reactions. This enzyme alters the activity of a protein (thiol), during protein folding, in response to oxidation through a redox switches (i.e. reversible oxidation of thiol groups to disulfide) [90,91]. The modifications have a crucial role in the cellular detoxification under oxidative stress.

DIP0411 is annotated as thioredoxin-like/putative electron transport protein described as a secreted protein with a role in the redox process (i.e. maturation of heme-containing cytochromes) [21]. This protein is homologue to the *E. coli* proteins DsbE/DsbF characterized by the conserved thioredoxin-like domain (CXXC) shared at their active sites [92,93]. The Dsb protein family are associated with the folding of virulence proteins (toxins, adhesin and flagella) [93]; however, the specific function of DIP0411 was not identified but is probably related with the oxidation-reduction process. In *M. tuberculosis*, DsbF and DsbE-like proteins functions as a thiol oxidase required for the correct folding of disulfide bond rich cell-wall associated and secreted proteins crucial for mycobacterial infectivity and viability [94]. DIP0411 probably accomplish a similar function as *M. tuberculosis* DsbEF-like protein being vital for bacteria detoxification and protein folding during macrophage internalization among other oxidative stress.

The elected MHC-I and MHC-II binding epitopes from the seven *C. diphtheriae* proteins were connected by appropriated linker. MHC-I epitopes were separated by ‘AAY’ linker that stimulates T helper responses and conformational dependent immunogenicity, whereas ‘GPGPG’ were added between MHC-II epitopes and this linker increase epitope presentation and removes epitope junctions. The final multi-epitope vaccine proved to be stable, non-allergenic with antigenic properties. The interaction with TLR2 was predicted and is mediated by hydrophobic regions with eight predicted H-bonds located between TLR2 frustrated region and the vaccine. An *in silico* cloning was performed and the vaccine sequence adjusted in accordance with *E. coli* codon usage for further *in vitro* and *in vivo* experiments.

A peptide adjuvant was not added to the sequence and since peptides have weak immunogenic properties being not capable to induce alone an effective immunity response, an external adjuvant is required. Adjuvants have the capability to activate the host innate immune response and enhance the adaptive response after antigen presentation [95]. The Diphtheria vaccine employs the alum adjuvant (aluminum hydroxide or aluminum phosphate), an efficient adjuvant for toxoid vaccines, with a low cost and a good safety record [6,96]. However, this adjuvant cause irritability (application site) with a high inflammation reaction, low induction of cellular immunity and stimulation of IgE immune response [5,96]. In addition, alum adjuvants were reported to be inadequate to peptide antigens [34]. Innovative ideal adjuvants have been developed to induce an effective immune response (cellular, humoral and/or mucosal) with low side effects, at low price with a simple production. This new demand favor the design of distinct types of adjuvants to cover the vaccine categories (e.g. DNA; subunit, e.g. peptide/toxoid; whole base microorganism), including delivery systems with the potential to simultaneously induce the immune

system, protect the antigen from degradation, and transport the antigen to the correct site (review at Skwarczynski and Toth, 2016).

Thus, we propose a multi-epitope Diphtheria vaccine based in a mycolate-associated glycolipids adjuvant/delivery system with protective technologies (e.g. heat/freeze-stable) and innovative delivery (e.g. transcutaneous, orally, edible) to enhance the cellular and humoral response and overcome the obstacles faced by many countries, like an infrastructure/storage problematic and lack trained healthcare agents [5,6,96].

Acknowledgements

We acknowledge the collaboration and assistance of all team members and the Brazilian funding agencies CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil).

REFERENCES

1. Guaraldi AL de M, Hirata R, Azevedo VA de C. *Corynebacterium Diphtheriae*, *Corynebacterium ulcerans* and *Corynebacterium pseudotuberculosis*—General Aspects. In: Burkovski A, organizador. *Corynebacterium Diphtheriae* Relat Toxigenic Species Genomics Pathog Appl [Internet]. Dordrecht: Springer Netherlands; 2014 [citado 6 de abril de 2019]. p. 15–37. Recuperado de: https://doi.org/10.1007/978-94-007-7624-1_2
2. Mattos-Guaraldi AL, Moreira LO, Damasco PV, Hirata Júnior R. Diphtheria remains a threat to health in the developing world--an overview. *Mem Inst Oswaldo Cruz*. 2003;98:987–93.
3. Burkovski A. Diphtheria and its Etiological Agents. In: Burkovski A, organizador. *Corynebacterium Diphtheriae* Relat Toxigenic Species [Internet]. Dordrecht: Springer Netherlands; 2014 [citado 6 de abril de 2019]. p. 1–14. Recuperado de: http://link.springer.com/10.1007/978-94-007-7624-1_1
4. WHO. WHO World Health Organization: Immunization, Vaccines And Biologicals. Vaccine preventable diseases Vaccines monitoring system 2018 Global Summary Reference Time Series: DIPHTHERIA [Internet]. 2018 [citado 2 de maio de 2019]. Recuperado de: http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tsincidenceDiphtheria.html

5. Bae K-D. The Evolution and Value of Diphtheria Vaccine. *KSBB J* [Internet]. 2011 [citado 6 de abril de 2019];26:491–504. Recuperado de: <http://koreascience.or.kr/journal/view.jsp?kj=KHGSBC&py=2011&vnc=v26n6&sp=491>
6. Jamal SB, Tiwari S, Silva A, Azevedo V, Jamal SB, Tiwari S, et al. Pathogenesis of *Corynebacterium Diphtheriae* and available vaccines: An Overview. *Glob J Infect Dis Clin Res* [Internet]. 2017 [citado 6 de abril de 2019];3:020–4. Recuperado de: <https://www.peertechz.com/articles/GJIDCR-3-114.php>
7. Möller J, Kraner M, Sonnewald U, Sangal V, Tittlbach H, Winkler J, et al. Proteomics of Diphtheria toxoid vaccines reveals multiple proteins that are immunogenic and may contribute to protection of humans against *Corynebacterium Diphtheriae*. *Vaccine* [Internet]. 2019 [citado 1º de maio de 2019]; Recuperado de: <http://www.sciencedirect.com/science/article/pii/S0264410X19305353>
8. Rappuoli R, Malito E. History of Diphtheria Vaccine Development. In: Burkovski A, organizador. *Corynebacterium Diphtheriae* Relat Toxigenic Species [Internet]. Dordrecht: Springer Netherlands; 2014 [citado 6 de abril de 2019]. p. 225–38. Recuperado de: http://link.springer.com/10.1007/978-94-007-7624-1_11
9. Hessling M, Feiertag J, Hönes K. Pathogens provoking most deaths worldwide. *Biosci Biotechnol Res Commun*. 2017;10:1–7.
10. WHO. World Health Organization Diphtheria Vaccine: WHO Position Paper. *Weekly epidemiological record* 2017;31:417–436. [Internet]. 2017 [citado 2 de maio de 2019]. Recuperado de: https://scholar.google.com/scholar_lookup?title=WHO%20position%20paper&author=World%20Health%20Organisation%20Diphtheria%20vaccine&publication_year=2017&pages=417-436
11. Dittmann S, Wharton M, Vitek C, Ciotti M, Galazka A, Guichard S, et al. Successful control of epidemic Diphtheria in the states of the Former Union of Soviet Socialist Republics: lessons learned. *J Infect Dis*. 2000;181 Suppl 1:S10-22.
12. Eskola J, Lumio J, Vuopio-Varkila J. Resurgent Diphtheria--are we safe? *Br Med Bull*. 1998;54:635–45.
13. Sangal V, Hoskisson PA. Evolution, epidemiology and diversity of *Corynebacterium Diphtheriae*: New perspectives on an old foe. *Infect Genet Evol* [Internet]. 2016 [citado 27 de março de 2019];43:364–70. Recuperado de: <http://www.sciencedirect.com/science/article/pii/S1567134816302489>
14. Scheifer C, Rolland-Debord C, Badell E, Reibel F, Aubry A, Perignon A, et al. Re-emergence of *Corynebacterium Diphtheriae*. *Médecine Mal Infect* [Internet]. 2018 [citado 27 de março de 2019]; Recuperado de: <http://www.sciencedirect.com/science/article/pii/S0399077X17309629>
15. Tiley SM, Kociuba KR, Heron LG, Munro R. Infective Endocarditis Due to Nontoxigenic *Corynebacterium Diphtheriae*: Report of Seven Cases and Review. *Clin Infect Dis* [Internet]. 1993 [citado 6 de abril de 2019];16:271–5. Recuperado de: <https://academic.oup.com/cid/article/16/2/271/287435>

16. Peixoto RS, Pereira GA, Sanches Dos Santos L, Rocha-de-Souza CM, Gomes DLR, Silva Dos Santos C, et al. Invasion of endothelial cells and arthritogenic potential of endocarditis-associated *Corynebacterium Diphtheriae*. *Microbiol Read Engl*. 2014;160:537–46.
17. Peixoto RS, Hacker E, Antunes CA, Weerasekera D, Dias A, Martins CA, et al. Pathogenic properties of a *Corynebacterium Diphtheriae* strain isolated from a case of osteomyelitis. *J Med Microbiol* [Internet]. 2016 [citado 6 de abril de 2019];65:1311–21. Recuperado de: <http://www.microbiologyresearch.org/content/journal/jmm/10.1099/jmm.0.000362>
18. Höfler W. Cutaneous Diphtheria. *Int J Dermatol*. 1991;30:845–7.
19. Trost E, Blom J, Soares S de C, Huang I-H, Al-Dilaimi A, Schröder J, et al. Pangenomic study of *Corynebacterium Diphtheriae* that provides insights into the genomic diversity of pathogenic isolates from cases of classical Diphtheria, endocarditis, and pneumonia. *J Bacteriol*. 2012;194:3199–215.
20. Timms VJ, Nguyen T, Crighton T, Yuen M, Sintchenko V. Genome-wide comparison of toxigenic and non-toxigenic *Corynebacterium Diphtheriae* isolates identifies differences in the pan genomes between respiratory and cutaneous strains. *bioRxiv* [Internet]. 2017 [citado 28 de março de 2019];143800. Recuperado de: <https://www.biorxiv.org/content/10.1101/143800v1>
21. Sangal V, Blom J, Sutcliffe IC, von Hunolstein C, Burkovski A, Hoskisson PA. Adherence and invasive properties of *Corynebacterium Diphtheriae* strains correlates with the predicted membrane-associated and secreted proteome. *BMC Genomics* [Internet]. 2015 [citado 19 de novembro de 2018];16:765. Recuperado de: <https://doi.org/10.1186/s12864-015-1980-8>
22. Ott L, Höller M, Rheinlaender J, Schäffer TE, Hensel M, Burkovski A. Strain-specific differences in pili formation and the interaction of *Corynebacterium Diphtheriae* with host cells. *BMC Microbiol* [Internet]. 2010 [citado 6 de abril de 2019];10:257. Recuperado de: <http://bmcmicrobiol.biomedcentral.com/articles/10.1186/1471-2180-10-257>
23. Reardon-Robinson ME, Ton-That H. Assembly and Function of *Corynebacterium Diphtheriae* Pili. In: Burkovski A, organizador. *Corynebacterium Diphtheriae* Relat Toxigenic Species Genomics Pathog Appl [Internet]. Dordrecht: Springer Netherlands; 2014 [citado 6 de abril de 2019]. p. 123–41. Recuperado de: https://doi.org/10.1007/978-94-007-7624-1_7
24. Rheinlaender J, Gräbner A, Ott L, Burkovski A, Schäffer TE. Contour and persistence length of *Corynebacterium Diphtheriae* pili by atomic force microscopy. *Eur Biophys J* [Internet]. 2012 [citado 6 de abril de 2019];41:561–70. Recuperado de: <http://link.springer.com/10.1007/s00249-012-0818-4>
25. Ott L, Burkovski A. Toxigenic *Corynebacteria*: Adhesion, Invasion and Host Response. In: Burkovski A, organizador. *Corynebacterium Diphtheriae* Relat Toxigenic Species Genomics Pathog Appl [Internet]. Dordrecht: Springer Netherlands; 2014 [citado 6 de abril de 2019]. p. 143–70. Recuperado de: https://doi.org/10.1007/978-94-007-7624-1_8

26. Ott L, Scholz B, Höller M, Hasselt K, Ensser A, Burkovski A. Induction of the NFκ-B signal transduction pathway in response to *Corynebacterium Diphtheriae* infection. *Microbiol Read Engl*. 2013;159:126–35.
27. Antunes CA, Sanches dos Santos L, Hacker E, Köhler S, Bösl K, Ott L, et al. Characterization of DIP0733, a multi-functional virulence factor of *Corynebacterium Diphtheriae*. *Microbiology [Internet]*. 2015 [citado 27 de março de 2019];161:639–47. Recuperado de: <https://mic.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.000020>
28. Jamal SB, Hassan SS, Tiwari S, Viana MV, Benevides L de J, Ullah A, et al. An integrative in-silico approach for therapeutic target identification in the human pathogen *Corynebacterium Diphtheriae*. de Brevern AG, organizador. *PLOS ONE [Internet]*. 2017 [citado 16 de novembro de 2018];12:e0186401. Recuperado de: <http://dx.plos.org/10.1371/journal.pone.0186401>
29. Ott L, Höller M, Gerlach RG, Hensel M, Rheinlaender J, Schäffer TE, et al. *Corynebacterium Diphtheriae* invasion-associated protein (DIP1281) is involved in cell surface organization, adhesion and internalization in epithelial cells. *BMC Microbiol*. 2010;10:2.
30. Peixoto RS, Antunes CA, Lourêdo LS, Viana VG, Santos CS dos, Fuentes Ribeiro da Silva J, et al. Functional characterization of the collagen-binding protein DIP2093 and its influence on host–pathogen interaction and arthritogenic potential of *Corynebacterium Diphtheriae*. *Microbiology [Internet]*. 2017 [citado 27 de março de 2019];163:692–701. Recuperado de: <https://mic.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.000467>
31. María RR, Arturo CJ, Alicia JA, Paulina MG, Gerardo AO. The Impact of Bioinformatics on Vaccine Design and Development. In: Afrin F, Hemeg H, Ozbak H, organizadores. *Vaccines [Internet]*. InTech; 2017 [citado 6 de abril de 2019]. Recuperado de: <http://www.intechopen.com/books/vaccines/the-impact-of-bioinformatics-on-vaccine-design-and-development>
32. Zhang L. Multi-epitope vaccines: a promising strategy against tumors and viral infections. *Cell Mol Immunol [Internet]*. 2018 [citado 6 de abril de 2019];15:182–4. Recuperado de: <http://www.nature.com/doifinder/10.1038/cmi.2017.92>
33. Li W, Joshi MD, Singhanian S, Ramsey KH, Murthy AK. Peptide Vaccine: Progress and Challenges. *Vaccines [Internet]*. 2014 [citado 14 de março de 2019];2:515–36. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4494216/>
34. Skwarczynski M, Toth I. Peptide-based synthetic vaccines. *Chem Sci [Internet]*. 2016 [citado 6 de abril de 2019];7:842–54. Recuperado de: <http://xlink.rsc.org/?DOI=C5SC03892H>
35. Hajjighahramani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatabadi SS, Ghasemi Y. Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*. *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis*. 2017;48:83–94.

36. Nezafat N, Karimi Z, Eslami M, Mohkam M, Zandian S, Ghasemi Y. Designing an efficient multi-epitope peptide vaccine against *Vibrio cholerae* via combined immunoinformatics and protein interaction based approaches. *Comput Biol Chem*. 2016;62:82–95.
37. Nezafat N, Eslami M, Negahdaripour M, Rahbar MR, Ghasemi Y. Designing an efficient multi-epitope oral vaccine against *Helicobacter pylori* using immunoinformatics and structural vaccinology approaches. *Mol Biosyst*. 2017;13:699–713.
38. Pandey RK, Bhatt TK, Prajapati VK. Novel Immunoinformatics Approaches to Design Multi-epitope Subunit Vaccine for Malaria by Investigating Anopheles Salivary Protein. *Sci Rep* [Internet]. 2018 [citado 6 de abril de 2019];8. Recuperado de: <http://www.nature.com/articles/s41598-018-19456-1>
39. Ali M, Pandey RK, Khatoon N, Narula A, Mishra A, Prajapati VK. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci Rep*. 2017;7:9232.
40. Huang X, Ginwala R, Karabudak A, Jain P, Philip R. Targeted multi-epitope therapeutic vaccine for the treatment of invasive glioblastoma multiforme. *J Immunother Cancer*. 2015;3:P436.
41. Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik S-G, et al. Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell*. 2007;130:1071–82.
42. De S, Bartók AP, Csányi G, Ceriotti M. Comparing molecules and solids across structural and alchemical space. *Phys Chem Chem Phys* [Internet]. 2016 [citado 13 de junho de 2019];18:13754–69. Recuperado de: <https://pubs.rsc.org/en/content/articlelanding/2016/cp/c6cp00415f>
43. Kufareva I, Abagyan R. Methods of protein structure comparison. *Methods Mol Biol Clifton NJ*. 2012;857:231–57.
44. Maiorov VN, Crippen GM. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *J Mol Biol*. 1994;235:625–34.
45. Maiorov VN, Crippen GM. Size-independent comparison of protein three-dimensional structures. *Proteins*. 1995;22:273–83.
46. Reva BA, Finkelstein AV, Skolnick J. What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? *Fold Des*. 1998;3:141–7.
47. Trost E, Blom J, Soares S de C, Huang I-H, Al-Dilaimi A, Schröder J, et al. Pangenomic Study of *Corynebacterium Diphtheriae* That Provides Insights into the Genomic Diversity of Pathogenic Isolates from Cases of Classical Diphtheria, Endocarditis, and Pneumonia. *J Bacteriol* [Internet]. 2012 [citado 28 de março de 2019];194:3199–215. Recuperado de: <https://jb.asm.org/content/194/12/3199>
48. Collier RJ. Understanding the mode of action of Diphtheria toxin: a perspective on progress during the 20th century. *Toxicon* [Internet]. 2001 [citado 7 de maio de 2019];39:1793–803. Recuperado de: <http://www.sciencedirect.com/science/article/pii/S0041010101001659>

49. Guaraldi AL de M, Hirata R, Azevedo VA de C. *Corynebacterium Diphtheriae*, *Corynebacterium ulcerans* and *Corynebacterium pseudotuberculosis*—General Aspects. In: Burkovski A, organizador. *Corynebacterium Diphtheriae* Relat Toxigenic Species Genomics Pathog Appl [Internet]. Dordrecht: Springer Netherlands; 2014 [citado 27 de março de 2019]. p. 15–37. Recuperado de: https://doi.org/10.1007/978-94-007-7624-1_2
50. Schmitt MP. Iron Acquisition and Iron-Dependent Gene Expression in *Corynebacterium Diphtheriae*. In: Burkovski A, organizador. *Corynebacterium Diphtheriae* Relat Toxigenic Species Genomics Pathog Appl [Internet]. Dordrecht: Springer Netherlands; 2014 [citado 9 de maio de 2019]. p. 95–121. Recuperado de: https://doi.org/10.1007/978-94-007-7624-1_6
51. Mitamura T, Umata T, Nakano F, Shishido Y, Toyoda T, Itai A, et al. Structure-Function Analysis of the Diphtheria Toxin Receptor Toxin Binding Site by Site-directed Mutagenesis. *J Biol Chem* [Internet]. 1997 [citado 7 de maio de 2019];272:27084–90. Recuperado de: <http://www.jbc.org/content/272/43/27084>
52. Johnson VG, Wilson D, Greenfield L, Youle RJ. The role of the Diphtheria toxin receptor in cytosol translocation. *J Biol Chem* [Internet]. 1988 [citado 7 de maio de 2019];263:1295–300. Recuperado de: <http://www.jbc.org/content/263/3/1295>
53. Ott L, Hacker E, Kunert T, Karrington I, Etschel P, Lang R, et al. Analysis of *Corynebacterium Diphtheriae* macrophage interaction: Dispensability of corynomycolic acids for inhibition of phagolysosome maturation and identification of a new gene involved in synthesis of the corynomycolic acid layer. *PloS One*. 2017;12:e0180105.
54. Santos CSD, Santos LSD, Souza MCD, Dourado FDS, Dias AADS de O, Sabbadini PS, et al. Non-opsonic phagocytosis of homologous non-toxigenic and toxigenic *Corynebacterium Diphtheriae* strains by human U-937 macrophages. *Microbiol Immunol* [Internet]. 2010 [citado 6 de maio de 2019];54:1–10. Recuperado de: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1348-0421.2009.00179.x>
55. Moyle PM, Toth I. Modern subunit vaccines: development, components, and research opportunities. *ChemMedChem*. 2013;8:360–76.
56. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. *Nat Rev Drug Discov*. 2007;6:404–14.
57. Tam JP. Synthetic peptide vaccine design: synthesis and properties of a high-density multiple antigenic peptide system. *Proc Natl Acad Sci* [Internet]. 1988 [citado 10 de maio de 2019];85:5409–13. Recuperado de: <https://www.pnas.org/content/85/15/5409>
58. Burkovski A. Cell Envelope of Corynebacteria: Structure and Influence on Pathogenicity. *ISRN Microbiol* [Internet]. 2013 [citado 29 de abril de 2014];2013. Recuperado de: <http://www.hindawi.com/journals/isrn.microbiology/2013/935736/abs/>

59. Burkovski A. The role of corynomycolic acids in *Corynebacterium*-host interaction. *Antonie Van Leeuwenhoek* [Internet]. 2018 [citado 8 de maio de 2019];111:717–25. Recuperado de: <https://doi.org/10.1007/s10482-018-1036-6>
60. Ott L. Adhesion properties of toxigenic corynebacteria. *Microbiol* 2018 Vol 4 Pages 85-103 [Internet]. 2018 [citado 8 de maio de 2019]; Recuperado de: <http://www.aimspress.com/article/10.3934/microbiol.2018.1.85>
61. Moreira LO, Mattos-Guaraldi AL, Andrade AFB. Novel lipoarabinomannan-like lipoglycan (CdiLAM) contributes to the adherence of *Corynebacterium Diphtheriae* to epithelial cells. *Arch Microbiol* [Internet]. 2008 [citado 27 de março de 2019];190:521–30. Recuperado de: <https://doi.org/10.1007/s00203-008-0398-y>
62. Peet PL van der, Gunawan C, Torigoe S, Yamasaki S, Williams SJ. Corynomycolic acid-containing glycolipids signal through the pattern recognition receptor Mincle. *Chem Commun* [Internet]. 2015 [citado 8 de maio de 2019];51:5100–3. Recuperado de: <https://pubs.rsc.org/en/content/articlelanding/2015/cc/c5cc00085h>
63. Schick J, Etschel P, Bailo R, Ott L, Bhatt A, Lepenies B, et al. Toll-Like Receptor 2 and Mincle Cooperatively Sense Corynebacterial Cell Wall Glycolipids. *Infect Immun* [Internet]. 2017 [citado 14 de março de 2019];85:e00075-17. Recuperado de: <https://iai.asm.org/content/85/7/e00075-17>
64. Blanc L, Castanier R, Mishra AK, Ray A, Besra GS, Sutcliffe I, et al. Gram-Positive Bacterial Lipoglycans Based on a Glycosylated Diacylglycerol Lipid Anchor Are Microbe-Associated Molecular Patterns Recognized by TLR2. *PLOS ONE* [Internet]. 2013 [citado 9 de maio de 2019];8:e81593. Recuperado de: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0081593>
65. Geisel RE, Sakamoto K, Russell DG, Rhoades ER. In Vivo Activity of Released Cell Wall Lipids of *Mycobacterium bovis* Bacillus Calmette-Guérin Is Due Principally to Trehalose Mycolates. *J Immunol* [Internet]. 2005 [citado 9 de maio de 2019];174:5007–15. Recuperado de: <http://www.jimmunol.org/content/174/8/5007>
66. Mishra AK, Alves JE, Krumbach K, Nigou J, Castro AG, Geurtsen J, et al. Differential Arabinan Capping of Lipoarabinomannan Modulates Innate Immune Responses and Impacts T Helper Cell Differentiation. *J Biol Chem* [Internet]. 2012 [citado 9 de maio de 2019];287:44173–83. Recuperado de: <http://www.jbc.org/content/287/53/44173>
67. Schoenen H, Bodendorfer B, Hitchens K, Manzanero S, Werninghaus K, Nimmerjahn F, et al. Cutting Edge: Mincle Is Essential for Recognition and Adjuvanticity of the Mycobacterial Cord Factor and its Synthetic Analog Trehalose-Dibehenate. *J Immunol* [Internet]. 2010 [citado 9 de maio de 2019];184:2756–60. Recuperado de: <http://www.jimmunol.org/content/184/6/2756>
68. Shenderov K, Barber DL, Mayer-Barber KD, Gurucha SS, Jankovic D, Feng CG, et al. Cord Factor and Peptidoglycan Recapitulate the Th17-Promoting Adjuvant Activity of Mycobacteria through Mincle/CARD9 Signaling and the Inflammasome. *J Immunol* [Internet]. 2013 [citado 9 de maio de 2019];190:5722–30. Recuperado de: <http://www.jimmunol.org/content/190/11/5722>

69. Takeuchi O, Hoshino K, Kawai T, Sanjo H, Takada H, Ogawa T, et al. Differential Roles of TLR2 and TLR4 in Recognition of Gram-Negative and Gram-Positive Bacterial Cell Wall Components. *Immunity* [Internet]. 1999 [citado 27 de março de 2019];11:443–51. Recuperado de: <http://www.sciencedirect.com/science/article/pii/S1074761300801193>
70. Colombo AV, Hirata R, Souza CMR, Monteiro-Leal LH, Previato JO, Formiga LCD, et al. *Corynebacterium Diphtheriae* surface proteins as adhesins to human erythrocytes. *FEMS Microbiol Lett* [Internet]. 2001 [citado 6 de maio de 2019];197:235–9. Recuperado de: <https://academic.oup.com/femsle/article-lookup/doi/10.1111/j.1574-6968.2001.tb10609.x>
71. Sabbadini PS, Assis MC, Trost E, Gomes DLR, Moreira LO, dos Santos CS, et al. *Corynebacterium Diphtheriae* 67-72p hemagglutinin, characterized as the protein DIP0733, contributes to invasion and induction of apoptosis in HEP-2 cells. *Microb Pathog* [Internet]. 2012 [citado 6 de maio de 2019];52:165–76. Recuperado de: <https://linkinghub.elsevier.com/retrieve/pii/S0882401011002099>
72. Chagnot C, Listrat A, Astruc T, Desvaux M. Bacterial adhesion to animal tissues: protein determinants for recognition of extracellular matrix components: ECM recognition by bacterial MSCRAMMs. *Cell Microbiol* [Internet]. 2012 [citado 6 de maio de 2019];14:1687–96. Recuperado de: <http://doi.wiley.com/10.1111/cmi.12002>
73. Sabbadini PS, Genovez MRN, Silva CF da, Adelino TLN, Santos CS dos, Pereira GA, et al. Fibrinogen binds to nontoxigenic and toxigenic *Corynebacterium Diphtheriae* strains. *Mem Inst Oswaldo Cruz* [Internet]. 2010 [citado 6 de maio de 2019];105:706–11. Recuperado de: http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0074-02762010000500018&lng=en&nrm=iso&tlng=es
74. Anantharaman V, Aravind L. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome Biol* [Internet]. 2003 [citado 7 de maio de 2019];4:R11. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC151301/>
75. Hansmeier N, Chao T-C, Kalinowski J, Pühler A, Tauch A. Mapping and comprehensive analysis of the extracellular and cell surface proteome of the human pathogen *Corynebacterium Diphtheriae*. *Proteomics*. 2006;6:2465–76.
76. Kolodkina V, Denisevich T, Titov L. Identification of *Corynebacterium Diphtheriae* gene involved in adherence to epithelial cells. *Infect Genet Evol* [Internet]. 2011 [citado 27 de março de 2019];11:518–21. Recuperado de: <http://www.sciencedirect.com/science/article/pii/S1567134810003047>
77. Dautin N, de Sousa-d’Auria C, Constantinesco-Becker F, Labarre C, Oberto J, Li de la Sierra-Gallay I, et al. Mycoloyltransferases: A large and major family of enzymes shaping the cell envelope of *Corynebacteriales*. *Biochim Biophys Acta BBA - Gen Subj* [Internet]. 2017 [citado 9 de maio de 2019];1861:3581–92. Recuperado de: <http://www.sciencedirect.com/science/article/pii/S0304416516302252>

78. Dover LG, Cerdeño-Tárraga AM, Pallen MJ, Parkhill J, Besra GS. Comparative cell wall core biosynthesis in the mycolated pathogens, *Mycobacterium tuberculosis* and *Corynebacterium Diphtheriae*. *FEMS Microbiol Rev*. 2004;28:225–50.
79. Kuo C-J, Ptak CP, Hsieh C-L, Akey BL, Chang Y-F. Elastin, a Novel Extracellular Matrix Protein Adhering to Mycobacterial Antigen 85 Complex. *J Biol Chem* [Internet]. 2013 [citado 9 de maio de 2019];288:3886–96. Recuperado de: <http://www.jbc.org/content/288/6/3886>
80. Ramulu HG, Adindla S, Guruprasad L. Analysis and modeling of mycolyl-transferases in the CMN group. *Bioinformatics* [Internet]. 2006 [citado 9 de maio de 2019];1:161–9. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1891678/>
81. Joliff G, Mathieu L, Hahn V, Bayan N, Duchiron F, Renaud M, et al. Cloning and nucleotide sequence of the *csp1* gene encoding PS1, one of the two major secreted proteins of *Corynebacterium glutamicum*: the deduced N-terminal region of PS1 is similar to the *Mycobacterium* antigen 85 complex. *Mol Microbiol*. 1992;6:2349–62.
82. Puech V, Bayan N, Salim K, Leblon G, Daffé M. Characterization of the *in vivo* acceptors of the mycoloyl residues transferred by the corynebacterial PS1 and the related mycobacterial antigens 85. *Mol Microbiol*. 2000;35:1026–41.
83. Dalbey RE, Wang P, Kuhn A. Assembly of bacterial inner membrane proteins. *Annu Rev Biochem*. 2011;80:161–87.
84. Thakur P, Gantasala NP, Choudhary E, Singh N, Abdin MZ, Agarwal N. The preprotein translocase YidC controls respiratory metabolism in *Mycobacterium tuberculosis*. *Sci Rep* [Internet]. 2016 [citado 9 de maio de 2019];6:24998. Recuperado de: <https://www.nature.com/articles/srep24998>
85. Thakur P, Choudhary E, Pareek M, Agarwal N. Regulation and overexpression studies of YidC in *Mycobacterium tuberculosis*. *Sci Rep* [Internet]. 2018 [citado 9 de maio de 2019];8:17114. Recuperado de: <https://www.nature.com/articles/s41598-018-35475-4>
86. Cryz SJ, Russell LM, Holmes RK. Regulation of toxinogenesis in *Corynebacterium Diphtheriae*: mutations in the bacterial genome that alter the effects of iron on toxin production. *J Bacteriol* [Internet]. 1983 [citado 9 de maio de 2019];154:245–52. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC217453/>
87. Staudenmaier H, Van Hove B, Yaraghi Z, Braun V. Nucleotide sequences of the *fecBCDE* genes and locations of the proteins suggest a periplasmic-binding-protein-dependent transport mechanism for iron(III) dicitrate in *Escherichia coli*. *J Bacteriol*. 1989;171:2626–33.
88. Billington SJ, Esmay PA, Songer JG, Jost BH. Identification and role in virulence of putative iron acquisition genes from *Corynebacterium pseudotuberculosis*. *FEMS Microbiol Lett* [Internet]. 2002 [citado 12 de junho de 2014];208:41–5. Recuperado de: <http://onlinelibrary.wiley.com/doi/10.1111/j.1574-6968.2002.tb11058.x/abstract>

89. Reniere ML. Reduce, Induce, Thrive: Bacterial Redox Sensing during Pathogenesis. *J Bacteriol* [Internet]. 2018 [citado 10 de maio de 2019];200. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6088161/>
90. Freedman RB, Hawkins HC, Murant SJ, Reid L. Protein disulphide-isomerase: a homologue of thioredoxin implicated in the biosynthesis of secretory proteins. *Biochem Soc Trans* [Internet]. 1988 [citado 10 de maio de 2019];16:96–9. Recuperado de: <http://europepmc.org/abstract/MED/3371540>
91. Freedman RB, Hirst TR, Tuite MF. Protein disulphide isomerase: building bridges in protein folding. *Trends Biochem Sci* [Internet]. 1994 [citado 10 de maio de 2019];19:331–6. Recuperado de: <http://europepmc.org/abstract/MED/7940678>
92. Martin JL. Thioredoxin--a fold for all reasons. *Struct Lond Engl* 1993. 1995;3:245–50.
93. Um S-H, Kim J-S, Lee K, Ha N-C. Structure of a DsbF homologue from *Corynebacterium Diphtheriae*. *Acta Crystallogr Sect F Struct Biol Commun* [Internet]. 2014 [citado 9 de maio de 2019];70:1167–72. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4157413/>
94. Chim N, Riley R, The J, Im S, Segelke B, Lakin T, et al. An extracellular disulfide bond forming protein (DsbF) from *Mycobacterium tuberculosis*: Structural, biochemical and gene expression analysis. *J Mol Biol* [Internet]. 2010 [citado 10 de maio de 2019];396:1211–26. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2883863/>
95. Petrovsky N, Aguilar JC. Vaccine adjuvants: current state and future trends. *Immunol Cell Biol*. 2004;82:488–96.
96. Alshantiti FM, Al-Masaudi SB, Al-Hejin AM, Redwan EM. Adjuvants for *Clostridium tetani* and *Clostridium Diphtheriae* vaccines updating. *Hum Antibodies* [Internet]. 2017 [citado 3 de maio de 2019];25:23–9. Recuperado de: <http://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/HAB-160302>
97. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* [Internet]. 2019 [citado 18 de março de 2019];47:D506–15. Recuperado de: <https://academic.oup.com/nar/article/47/D1/D506/5160987>
98. Chen X, Zaro J, Shen W-C. Fusion Protein Linkers: Property, Design and Functionality. *Adv Drug Deliv Rev* [Internet]. 2013 [citado 2 de abril de 2019];65:1357–69. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3726540/>
99. Singh H, Raghava GPS. ProPred1: prediction of promiscuous MHC Class-I binding sites. *Bioinforma Oxf Engl*. 2003;19:1009–14.
100. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res*. 2019;47:D339–43.

101. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods* [Internet]. 2015 [citado 21 de março de 2019];12:7. Recuperado de: <https://www.nature.com/articles/nmeth.3213>
102. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* [Internet]. 2008 [citado 21 de março de 2019];9. Recuperado de: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-40>
103. Heo L, Park H, Seok C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* 2013;41:W384-388.
104. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* [Internet]. 1993 [citado 21 de março de 2019];26:283-91. Recuperado de: [//scripts.iucr.org/cgi-bin/paper?gl0276](https://scripts.iucr.org/cgi-bin/paper?gl0276)
105. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins.* 1992;12:345-64.
106. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins.* 2006;65:40-8.
107. Ponomarenko J, Bui H-H, Li W, Fusseder N, Bourne PE, Sette A, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* [Internet]. 2008 [citado 7 de abril de 2019];9:514. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2607291/>
108. Dhanda SK, Vir P, Raghava GP. Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct* [Internet]. 2013 [citado 25 de março de 2019];8:30. Recuperado de: <https://doi.org/10.1186/1745-6150-8-30>
109. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v.2--a server for in silico prediction of allergens. *J Mol Model.* 2014;20:2278.
110. Saha S, Raghava GPS. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res* [Internet]. 2006 [citado 25 de março de 2019];34:W202-9. Recuperado de: https://academic.oup.com/nar/article/34/suppl_2/W202/2505748
111. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* [Internet]. 2007 [citado 25 de março de 2019];8:4. Recuperado de: <https://doi.org/10.1186/1471-2105-8-4>
112. Cheng J, Randall AZ, Sweredoski MJ, Baldi P. SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res* [Internet]. 2005 [citado 25 de março de 2019];33:W72-6. Recuperado de: https://academic.oup.com/nar/article/33/suppl_2/W72/2505518
113. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, organizador. *Proteomics Protoc Handb*

[Internet]. Totowa, NJ: Humana Press; 2005 [citado 25 de março de 2019]. p. 571–607. Recuperado de: <http://link.springer.com/10.1385/1-59259-890-0:571>

114. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera?A visualization system for exploratory research and analysis. *J Comput Chem* [Internet]. 2004 [citado 5 de junho de 2019];25:1605–12. Recuperado de: <http://doi.wiley.com/10.1002/jcc.20084>

115. Torchala M, Moal IH, Chaleil RAG, Fernandez-Recio J, Bates PA. SwarmDock: a server for flexible protein–protein docking. *Bioinformatics* [Internet]. 2013 [citado 1º de maio de 2019];29:807–9. Recuperado de: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btt038>

116. Torchala M, Bates PA. Predicting the Structure of Protein–Protein Complexes Using the SwarmDock Web Server. In: Kihara D, organizador. *Protein Struct Predict* [Internet]. New York, NY: Springer New York; 2014 [citado 1º de maio de 2019]. p. 181–97. Recuperado de: http://link.springer.com/10.1007/978-1-4939-0366-5_13

117. Moal IH, Barradas-Bautista D, Jiménez-García B, Torchala M, van der Velde A, Vreven T, et al. IRaPPA: information retrieval based integration of biophysical models for protein assembly selection. Tramontano A, organizador. *Bioinformatics* [Internet]. 2017 [citado 1º de maio de 2019];33:1806–13. Recuperado de: <https://academic.oup.com/bioinformatics/article/33/12/1806/2995816>

118. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* [Internet]. 2015 [citado 8 de junho de 2019];1–2:19–25. Recuperado de: <http://www.sciencedirect.com/science/article/pii/S2352711015000059>

119. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*. 2010;78:1950–8.

120. Grote A, Hiller K, Scheer M, Munch R, Nortemann B, Hempel DC, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* [Internet]. 2005 [citado 13 de maio de 2019];33:W526–31. Recuperado de: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gki376>

121. Carbone A, Zinovyev A, Képès F. Codon adaptation index as a measure of dominating codon bias. *Bioinforma Oxf Engl*. 2003;19:2005–15.

122. Sharp PM, Li WH. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*. 1987;15:1281–95.

Supplementary Tables - Design of a broad-spectrum candidate multi-epitope vaccine against Diphtheria: An interactive immunoinformatics approach

Table S1: List of the MHC-I binding peptides with 9 amino acids of 12 *C. diphtheriae* selected proteins common between ProPred-I and IEDB (top 4 peptide score and with percentile rank below 3%, respectively).

Protein	Allele	Sequence	Peptide Position	Peptide Score	Percentile_rank
DIP0222	MHC-Db	YSVDNENPL	138	780.000	0.2
DIP0222	MHC-Db	LKVDNAETI	167	600.000	1.7
DIP0222	MHC-Db	AVHHNTEEI	393	220.000	0.3
DIP0222	MHC-Dd	HGPIKNKMS	296	200.000	1.6
DIP0222	MHC-Dd	TYPGLTKVL	157	20.000	1.8
DIP0222	MHC-kb	IGFAAYNFV	426	6.336	0.76
DIP0222	MHC-kb	VVKVTYPGL	153	4.000	0.32
DIP0222	MHC-kd	AYNFVESII	430	2.880.000	0.12
DIP0222	MHC-kd	KYDAAGYSV	132	720.000	1.7
DIP0222	MHC-kk	TEEIVAQSI	398	1.000.000	0.2
DIP0222	MHC-kk	SEKIHSNEI	569	500.000	0.2
DIP0222	MHC-kk	VELEINFET	234	60.000	1.8
DIP0222	MHC-Ld	TPLPIAGVL	498	150.000	1.99
DIP0411	MHC-Kk	GERAPVSEI	33	500.000	1.8
DIP0411	MHC-Kk	REVTAQDLI	157	250.000	0.8
DIP0411	MHC-Ld	PPFKTAAQL	123	18.000	2.0
DIP0411	MHC-Ld	YPSIYDPPF	117	300.000	2.05
DIP0443	MHC-Kk	YELTSGSSI	139	500.000	0.2
DIP0443	MHC-Dd	SGPGEVYMF	114	144.000	0.2
DIP0443	MHC-Dd	GGPNNVADC	192	40.000	1.1
DIP0443	MHC-Dd	TGPNFMTLS	236	40.000	1.9
DIP0443	MHC-Ld	GPNFMTLSF	237	300.000	1.7
DIP0443	MHC-Ld	LPMAQDAGL	80	150.000	0.77
DIP0443	MHC-kb	FMTLSFIVL	240	10.000	2.45
DIP0558	MHC-Kk	AELSRLRSI	27	500.000	0.66
DIP0558	MHC-Db	TAQVSFDNL	46	4.118	2.4
DIP0558	MHC-Ld	TAQVSFDNL	46	9.000	2.75
DIP0558	MHC-Kb	TAQVSFDNL	46	30.000	2.52
DIP0733	MHC-Kk	AEALSQVGI	891	500.000	1.3
DIP0733	MHC-Kk	YEFDENDPV	668	100.000	2.8
DIP0733	MHC-Kb	AFAMRYQEL	549	26.400	0.51
DIP0733	MHC-Kb	YWLDRYSLI	234	24.000	0.17
DIP0733	MHC-Db	LAIGNAWPI	302	780.000	0.1
DIP0733	MHC-Db	ATVSNIRLL	368	660.000	0.6
DIP0733	MHC-Db	VLPKIVLL	260	100.000	0.88
DIP0733	MHC-Db	YGPVIASAT	498	40.000	0.5
DIP0733	MHC-Db	QGPKQAQDT	810	40.000	2.1
DIP0733	MHC-Db	KGVHRFLVV	112	36.000	2.0
DIP0733	MHC-Kd	SYTDINAVL	253	2.880.000	0.9
DIP0733	MHC-Kd	RYQELNLIL	553	2.880.000	1.0
DIP0733	MHC-Kd	FYAFTLPAL	162	2.400.000	0.4

DIP0733	MHC-Kd	SYWLDRYSL	233	2.400.000	0.5
DIP0733	MHC-Ld	RPQRRLTWL	16	225.000	0.47
DIP0733	MHC-Ld	IPAISTVLM	289	150.000	1.65
DIP0733	MHC-Ld	YPVYTVSDL	469	150.000	1.17
DIP0733	MHC-Ld	LPYAERTSL	616	150.000	0.65
DIP0979	MHC-Db	FVNFNAGTL	184	220.000	0.5
DIP0979	MHC-Dd	HGPCEVKNF	118	144.000	0.3
DIP0979	MHC-Dd	RGPVTVYLV	136	120.000	0.6
DIP0979	MHC-Kb	DAYDAYLRL	80	120.000	0.4
DIP0979	MHC-Kb	YLVDFPRM	142	12.000	2.75
DIP0979	MHC-Kb	DAYLRHL	83	5.000	0.48
DIP0979	MHC-Kd	LYVTAGSKI	259	4.000.000	0.3
DIP0979	MHC-Kk	LELAGANNI	288	1.000.000	0.3
DIP0981	MHC-Db	CLPENFENI	125	864.000	1.1
DIP0981	MHC-Db	MSLRNRGLI	135	780.000	0.2
DIP0981	MHC-Db	NIAMNGEVL	15	220.000	2.9
DIP0981	MHC-Dd	KGPCLPENF	122	207.360	0.1
DIP0981	MHC-Dd	TSPYSERL	37	24.000	1.4
DIP0981	MHC-Kb	TSPYSERL	37	3.300	0.55
DIP0981	MHC-Kd	AYLAEGTKV	175	1.728.000	0.23
DIP0983	MHC-Dd	ADPWRVLRI	47	50.000	1.5
DIP0983	MHC-Dd	GGPGLMEAA	109	48.000	2.0
DIP0983	MHC-Dd	ELPADHDWL	37	24.000	2.4
DIP0983	MHC-Kb	TMFLKYSQA	159	4.000	0.23
DIP0983	MHC-Kd	KYSQAFVCL	163	2.880.000	0.97
DIP0983	MHC-Kk	IELPHEQGI	133	1.000.000	1.6
DIP0983	MHC-Kk	HEQGINDYV	137	200.000	2.7
DIP0983	MHC-Kk	NERHLCGPV	10	100.000	0.4
DIP0983	MHC-Ld	QPQRNERHL	6	225.000	0.81
DIP1084	MHC-Db	ALGVNVPAL	238	264.000	1.3
DIP1084	MHC-Dd	AGPDAAAL	231	480.000	0.4
DIP1084	MHC-Dd	FGLLGNFGL	117	24.000	3.0
DIP1084	MHC-Dd	TTPVLFIAL	8	20.000	1.1
DIP1084	MHC-Dd	RAPRGVLAL	66	20.000	2.3
DIP1084	MHC-Kb	TTPVLFIAL	8	30.000	0.81
DIP1084	MHC-Kb	IGAPLFLIL	327	13.200	3.0
DIP1084	MHC-Kd	LYALSGGAL	125	2.400.000	0.3
DIP1084	MHC-Kd	VFMVWVDVL	300	1.600.000	1.95
DIP1084	MHC-Kd	AFSALASFL	168	1.382.400	1.9
DIP1084	MHC-Kd	TFAVYAVTL	140	800.000	2.4
DIP1084	MHC-Kk	WELRAPRGV	63	50.000	1.1
DIP1084	MHC-Ld	NPLADPYLL	93	150.000	0.16
DIP1084	MHC-Ld	YPQEQVWAV	35	90.000	1.03
DIP1281	MHC-Db	AATPNHASL	370	286.000	0.4
DIP1281	MHC-Db	SARGNADSL	249	85.800	0.4
DIP1281	MHC-Db	ASVSNSDSE	65	39.674	1.9
DIP1281	MHC-Db	SSMSSRTTT	4	9.413	1.9
DIP1281	MHC-Dd	YGPNAEYHV	526	60.000	2.2
DIP1281	MHC-Dd	VSPVRWSGM	555	30.000	0.6
DIP1281	MHC-Dd	MSPYAVRLI	563	10.000	0.8
DIP1281	MHC-Kb	VSPVRWSGM	555	3.000	0.14
DIP1281	MHC-Kd	IYLGDGTM	536	5.760.000	0.5
DIP1281	MHC-Kd	LYAFAGVGI	489	2.000.000	0.97

DIP1281	MHC-Kd	AYIGTPYAW	445	144.000	2.8
DIP1281	MHC-Kd	GLIASGLSI	17	115.200	1.1
DIP1281	MHC-Kk	NEESRLREV	189	50.000	1.5
DIP1281	MHC-Ld	LPHYTYGYQY	499	72.000	0.54
DIP2193	MHC-Db	TDVENGWTI	107	792.000	1.3
DIP2193	MHC-Db	AVYVNTPSM	66	264.000	0.22
DIP2193	MHC-Db	YSDKNVNVI	123	62.400	0.6
DIP2193	MHC-Db	SCVNNEYDV	371	33.000	2.9
DIP2193	MHC-Dd	GGPQSWLGF	415	144.000	0.1
DIP2193	MHC-Dd	WGPDGSQDW	237	48.000	0.11
DIP2193	MHC-Kb	RINARYSEM	406	26.400	0.74
DIP2193	MHC-Kb	AVYVNTPSM	66	11.000	0.74
DIP2193	MHC-Kd	SYAGIGLEV	285	600.000	2.0
DIP2193	MHC-Kd	PYMANALGL	337	576.000	1.0
DIP2193	MHC-Kk	VEWLGGRRV	57	150.000	2.0
DIP2193	MHC-Kk	QEFEGGLVV	490	100.000	2.7
DIP2193	MHC-Kk	AESELGFPV	522	100.000	0.6
DIP2193	MHC-Ld	LPVGGQSSF	132	390.000	1.18
DIP2193	MHC-Ld	VPGGKVEDF	379	300.000	2.94
DIP2193	MHC-Ld	SPSTGAHAL	395	195.000	2.75
DIP2193	MHC-Ld	SPQRGAIIV	448	117.000	1.65
DIP2379	MHC-Db	YMVSNNIWT	266	39.000	1.1
DIP2379	MHC-Db	IALVAAPLI	190	10.296	0.7
DIP2379	MHC-Kb	RMMLWFMPPL	242	11.000	0.13
DIP2379	MHC-Kb	VQMPVFIGL	108	10.000	1.26
DIP2379	MHC-Kb	VFIGLFHVL	112	10.000	1.15
DIP2379	MHC-Kd	NYIFSAADV	145	1.200.000	0.6
DIP2379	MHC-Ld	WPISAILWF	9	300.000	0.67
DIP2379	MHC-Ld	YPNDQQKLM	77	150.000	1.85
DIP2379	MHC-Ld	VPLSAFISM	164	150.000	0.16
DIP2379	MHC-Ld	APLIVIVL	195	150.000	1.23

Table S2: List of the MHC-II binding peptides with 15 amino acids of 11 *C. diphtheriae* selected proteins obtained from IEDB (percentile rank below 3%).

Protein	Allele	Start	Peptide	Percentile rank
DIP0222	H2-IAd	335	LSSLMVAQAIPLVGE	0.11
DIP0222	H2-IAd	334	ALSSLMVAQAIPLVG	0.12
DIP0222	H2-IAd	333	IALSSLMVAQAIPLV	0.21
DIP0222	H2-IAd	336	SSLMVAQAIPLVGEL	0.27
DIP0222	H2-IAd	330	AQSIALSSLMVAQAI	0.62
DIP0222	H2-IAd	332	SIALSSLMVAQAIPL	0.64
DIP0222	H2-IAb	475	SPVYVGNVHNLHV	0.71
DIP0222	H2-IAb	474	KSPVYVGNVHNLH	0.76
DIP0222	H2-IAd	337	SLMVAQAIPLVGELV	0.79
DIP0222	H2-IAb	473	PKSPVYVGNVHNL	0.80
DIP0222	H2-IAb	476	PVYVGNVHNLHVA	0.83
DIP0222	H2-IAd	331	QSIALSSLMVAQAIPL	0.85

DIP0222	H2-IAAd	277	NYAAWAVNVAQVIDS	0.86
DIP0222	H2-IAAd	329	VAQSIALSSLMVAQA	0.86
DIP0222	H2-IAAb	472	RPKSPVYVGNGVHAN	0.91
DIP0222	H2-IAAd	275	GANYAAWAVNVAQVI	0.93
DIP0222	H2-IAAd	276	ANYAAWAVNVAQVID	0.96
DIP0222	H2-IAAd	278	YAAWAVNVAQVIDSE	1.18
DIP0222	H2-IAAd	274	AGANYAAWAVNVAQV	1.27
DIP0222	H2-IAAd	297	LEKTTAALSILPGIG	1.36
DIP0222	H2-IAAd	338	LMVAQAIPLVGELVD	1.42
DIP0222	H2-IAAd	296	NLEKTTAALSILPGI	1.58
DIP0222	H2-IAAd	328	IVAQSIALSSLMVAQ	2.06
DIP0222	H2-IAAb	477	VYVGNGVHANLHVAF	2.19
DIP0222	H2-IAAd	298	EKTTAALSILPGIGS	2.23
DIP0222	H2-IAAd	339	MVAQAIPLVGELVDI	2.42
DIP0222	H2-IAAd	273	FAGANYAAWAVNVAQ	2.49
DIP0222	H2-IAAd	295	DNLEKTTAALSILPG	2.62
DIP0222	H2-IAAb	478	YVGNGVHANLHVAFH	2.83
DIP0411	H2-IAAd	144	LREVTAQDLIKVIDS	0.95
DIP0411	H2-IAAd	142	VFLREVTAQDLIKVI	1.09
DIP0411	H2-IAAd	143	FLREVTAQDLIKVID	1.46
DIP0411	H2-IAAb	105	YPSIYDPPFKTAAQL	1.48
DIP0411	H2-IAAb	104	TYPSIYDPPFKTAAQ	1.52
DIP0411	H2-IAAb	103	ITYPSIYDPPFKTAA	1.61
DIP0411	H2-IAAb	106	PSIYDPPFKTAAQLG	1.64
DIP0411	H2-IAAd	141	AVFLREVTAQDLIKV	1.71
DIP0411	H2-IAAb	107	SIYDPPFKTAAQLGG	1.74
DIP0411	H2-IAAd	135	KQHRPAAVFLREVTA	2.51
DIP0411	H2-IAAd	134	DKQHRPAAVFLREVT	2.88
DIP0443	H2-IAAb	187	DCEGTAAPVAAPASK	1.04
DIP0443	H2-IAAb	188	CEGTAAPVAAPASKG	1.04
DIP0443	H2-IAAb	186	ADCEGTAAPVAAPAS	1.08
DIP0443	H2-IAAb	189	EGTAAPVAAPASKGK	1.12
DIP0443	H2-IAAb	190	GTAAPVAAPASKGKS	1.35
DIP0443	H2-IAAd	188	CEGTAAPVAAPASKG	2.38
DIP0443	H2-IAAd	189	EGTAAPVAAPASKGK	2.72
DIP0443	H2-IAAd	187	DCEGTAAPVAAPASK	2.73
DIP0443	H2-IAAd	151	EPGTYTMCVQATGNG	2.95
DIP0443	H2-IAAb	191	TAAPVAAPASKGKSA	2.99
DIP0733	H2-IAAb	878	YQKVGYAPTIAEAL	0.49
DIP0733	H2-IAAb	879	QKVGYAPTIAEALS	0.49
DIP0733	H2-IAAb	880	GKVGYAPTIAEALSQ	0.60
DIP0733	H2-IAAb	881	KVGYAPTIAEALSQV	0.62
DIP0733	H2-IAAb	490	KEPRIYYGPVIASAT	0.73
DIP0733	H2-IAAb	491	EPRIYYGPVIASATD	0.74
DIP0733	H2-IAAb	882	VGYAPTIAEALSQVG	0.78
DIP0733	H2-IAAd	130	LGQQAQWQTVQLFFNR	0.80
DIP0733	H2-IAAd	129	FLGQQAQWQTVQLFFN	0.91
DIP0733	H2-IAAd	208	SITNYAKVQLAVTGG	0.93
DIP0733	H2-IAAd	709	VQRKMLARYHVDDAR	1.08
DIP0733	H2-IAAd	128	GFLGQQAQWQTVQLFF	1.19
DIP0733	H2-IAAd	776	QREYLAHMSVSSDP	1.20
DIP0733	H2-IAAb	492	PRIYYGPVIASATDG	1.21
DIP0733	H2-IAAb	924	KEGKAPSTPSAPASG	1.21

DIP0733	H2-IAb	493	RIYYGPVIASATDGA	1.28
DIP0733	H2-IAb	925	EGKAPSTPSAPASGS	1.29
DIP0733	H2-IAAd	777	REYLAAHMSVSSDPD	1.30
DIP0733	H2-IAAd	210	TNYAKVQLAVTGGLY	1.32
DIP0733	H2-IAAd	708	KVQRKMLARYHVDDA	1.41
DIP0733	H2-IAAd	207	GSITNYAKVQLAVTG	1.42
DIP0733	H2-IAb	923	DKEGKAPSTPSAPAS	1.42
DIP0733	H2-IAAd	127	AGFLGQQAWQTVQLF	1.47
DIP0733	H2-IAb	489	VKEPRIYYGPVIASA	1.64
DIP0733	H2-IAAd	710	QRKMLARYHVDDARD	1.69
DIP0733	H2-IAAd	131	GQQAWQTVQLFFNRQ	1.74
DIP0733	H2-IAAd	209	ITNYAKVQLAVTGGL	1.80
DIP0733	H2-IAb	922	EDKEGKAPSTPSAPA	1.87
DIP0733	H2-IAAd	102	VQYRAAVEKGVHRFL	1.88
DIP0733	H2-IAAd	101	VVQYRAAVEKGVHRF	1.88
DIP0733	H2-IAb	157	DYGFYAFTLPALRLV	1.88
DIP0733	H2-IAb	926	GKAPSTPSAPASGSG	1.95
DIP0733	H2-IAAd	775	LQREYLAAHMSVSSD	1.99
DIP0733	H2-IAb	158	YGFYAFTLPALRLVV	2.05
DIP0733	H2-IAb	494	IYYGPVIASATDGAD	2.10
DIP0733	H2-IAAd	707	FKVQRKMLARYHVDD	2.25
DIP0733	H2-IAAd	706	MFKVQRKMLARYHVD	2.31
DIP0733	H2-IAb	927	KAPSTPSAPASGSGT	2.32
DIP0733	H2-IAAd	250	GSYTDINAVLPAKIV	2.37
DIP0733	H2-IAb	156	MDYGFYAFTLPALRL	2.42
DIP0733	H2-IAAd	293	VLMIVSSLAIGNAWP	2.47
DIP0733	H2-IAb	488	VVKEPRIYYGPVIAS	2.48
DIP0733	H2-IAb	774	GLQREYLAAHMSVSS	2.53
DIP0733	H2-IAb	775	LQREYLAAHMSVSSD	2.54
DIP0733	H2-IAb	883	GYAPTIAEALSQVGI	2.54
DIP0733	H2-IAAd	132	QQAWQTVQLFFNRQD	2.55
DIP0733	H2-IAb	155	GMDYGFYAFTLPALR	2.59
DIP0733	H2-IAb	1	MATGFTRPAAAPKRP	2.69
DIP0733	H2-IAAd	265	LLVISAVVAISFFSV	2.73
DIP0733	H2-IAb	921	PEDKEGKAPSTPSAP	2.75
DIP0733	H2-IAAd	256	NAVLPKIVLLVISA	2.78
DIP0733	H2-IAb	773	RGLQREYLAAHMSVS	2.78
DIP0733	H2-IAAd	774	GLQREYLAAHMSVSS	2.83
DIP0733	H2-IAAd	281	VTKDLRIPAISTVLM	2.85
DIP0733	H2-IAAd	100	PVVQYRAAVEKGVHR	2.88
DIP0733	H2-IAAd	705	DMFKVQRKMLARYHV	2.91
DIP0733	H2-IAb	877	SYQKVGYPPTIAEA	2.91
DIP0733	H2-IAb	675	LKAWKGVFPGTVKAK	2.94
DIP0733	H2-IAAd	292	TVLMIVSSLAIGNAW	2.95
DIP0733	H2-IAb	776	QREYLAAHMSVSSDP	2.99
DIP0979	H2-IAAd	1	MTSALATGLATITAD	1.51
DIP0979	H2-IAb	254	EAGLYVTAGSKIIVR	1.63
DIP0979	H2-IAb	253	IEAGLYVTAGSKIIV	1.65
DIP0979	H2-IAAd	59	GVERRQAITTTIADLD	1.68
DIP0979	H2-IAb	255	AGLYVTAGSKIIVRG	1.69
DIP0979	H2-IAAd	60	VRRQAITTTIADLDA	1.80
DIP0979	H2-IAAd	2	TSALATGLATITADG	1.81
DIP0979	H2-IAb	252	VIEAGLYVTAGSKII	1.82

DIP0979	H2-IAAd	58	RGVRRQAITTTIADL	1.97
DIP0979	H2-IAAb	256	GLYVTAGSKIIVRGA	2.08
DIP0979	H2-IAAd	57	DRGVRQAITTTIAD	2.78
DIP0979	H2-IAAd	71	DLDAAPIDAYDAYLR	2.90
DIP0979	H2-IAAd	72	LDAAPIDAYDAYLRL	2.91
DIP0981	H2-IAAd	163	ITEAERVRLGAYLAE	1.04
DIP0981	H2-IAAd	61	DRLVEQVAVRTYIGD	1.67
DIP0981	H2-IAAd	162	EITEAERVRLGAYLA	1.76
DIP0981	H2-IAAd	164	TEAERVRLGAYLAEG	1.76
DIP0981	H2-IAAd	60	EDRLVEQVAVRTYIG	1.88
DIP0981	H2-IAAd	166	AERVRLGAYLAEGTK	1.89
DIP0981	H2-IAAd	165	EAERVRLGAYLAEGT	2.02
DIP0981	H2-IAAd	63	LVEQVAVRTYIGDLD	2.57
DIP0981	H2-IAAb	27	TPELVPASTSPSYSE	2.79
DIP0981	H2-IAAb	26	PTPELVPASTSPSYS	2.81
DIP0981	H2-IAAb	24	WYPTPELVPASTSPS	2.88
DIP0981	H2-IAAb	25	YPTPELVPASTSPSY	2.88
DIP1084	H2-IAAb	146	GLHFRYFFARKTMFL	1.97
DIP1084	H2-IAAb	145	LGLHFRYFFARKTMF	1.98
DIP1084	H2-IAAb	147	LHFRYFFARKTMFLK	2.06
DIP1084	H2-IAAb	144	DLGLHFRYFFARKTM	2.39
DIP1084	H2-IAAb	143	VDLGLHFRYFFARKT	2.62
DIP1281	H2-IAAd	315	DEAARKVAAEKA EKA	0.03
DIP1281	H2-IAAd	316	EAARKVAAEKA EKAR	0.03
DIP1281	H2-IAAd	301	EAARIAAEAARKAAD	0.03
DIP1281	H2-IAAd	317	AARKVAAEKA EKARQ	0.04
DIP1281	H2-IAAd	300	EAAARIAAEAARKAA	0.04
DIP1281	H2-IAAd	302	AARIAAEAARKAADE	0.05
DIP1281	H2-IAAd	299	KEEAARIAAEAARKA	0.07
DIP1281	H2-IAAd	318	ARKVAAEKA EKARQD	0.09
DIP1281	H2-IAAb	449	GTPYAWGGGNASGPT	0.10
DIP1281	H2-IAAb	448	IGTPYAWGGGNASGP	0.11
DIP1281	H2-IAAb	450	TPYAWGGGNASGPTQ	0.11
DIP1281	H2-IAAb	447	YIGTPYAWGGGNASG	0.12
DIP1281	H2-IAAd	298	AKEEAARIAAEAARK	0.13
DIP1281	H2-IAAb	446	AYIGTPYAWGGGNAS	0.13
DIP1281	H2-IAAd	314	ADEAARKVAAEKA EK	0.14
DIP1281	H2-IAAd	297	KAKEEAARIAAEAAR	0.17
DIP1281	H2-IAAd	313	AADEAARKVAAEKA E	0.18
DIP1281	H2-IAAd	360	DSARAAAALIEAATP	0.23
DIP1281	H2-IAAd	359	NDSARAAAALIEAAT	0.29
DIP1281	H2-IAAd	211	SAKASAQQAIDSSNS	0.31
DIP1281	H2-IAAd	361	SARAAAALIEAATPN	0.36
DIP1281	H2-IAAd	286	EKARQEAEAARKAKE	0.38
DIP1281	H2-IAAd	296	RKAKEEAARIAAEAA	0.38
DIP1281	H2-IAAd	303	ARIAAEAARKAADEA	0.43
DIP1281	H2-IAAd	285	AEKARQEAEAARKAK	0.47
DIP1281	H2-IAAd	275	RKKAEEAEAAA EKAR	0.47
DIP1281	H2-IAAd	287	KARQEAEAARKAKEE	0.49
DIP1281	H2-IAAd	210	ESAKASAQQAIDSSN	0.51
DIP1281	H2-IAAd	328	KARQDAEKAAEAAQA	0.51
DIP1281	H2-IAAd	284	AAEKARQEAEAARKA	0.52
DIP1281	H2-IAAd	274	ARKKAEEAEAAA EKA	0.52

DIP1281	H2-IAb	451	PYAWGGGNASGPTQG	0.53
DIP1281	H2-IAAd	304	RIAAEAARKAADEAA	0.55
DIP1281	H2-IAAd	305	IAAEAARKAADEAAR	0.56
DIP1281	H2-IAAd	312	KAADEAARKVAAEKA	0.57
DIP1281	H2-IAAd	358	ANDSARAAAALIEAA	0.57
DIP1281	H2-IAAd	362	ARAAAALIEAATPNH	0.63
DIP1281	H2-IAb	452	YAWGGGNASGPTQGI	0.64
DIP1281	H2-IAAd	277	KAEAEAAAAEKARQE	0.76
DIP1281	H2-IAAd	209	AESAKASAQQAIDSS	0.77
DIP1281	H2-IAAd	276	KKAEAEAAAAEKARQ	0.80
DIP1281	H2-IAAd	363	RAAAALIEAATPNHA	0.81
DIP1281	H2-IAAd	4	RSSMSSRTTTKVMRG	0.94
DIP1281	H2-IAAd	212	AKASAQQAIDSSNSK	0.94
DIP1281	H2-IAAd	327	EKARQDAEKAAEAAQ	0.99
DIP1281	H2-IAAd	3	IRSSMSSRTTTKVMR	1.01
DIP1281	H2-IAAd	432	SRSEKIEIVISRAMA	1.02
DIP1281	H2-IAAd	431	ASRSEKIEIVISRAM	1.03
DIP1281	H2-IAb	365	AAALIEAATPNHASL	1.06
DIP1281	H2-IAb	366	AALIEAATPNHASLD	1.14
DIP1281	H2-IAAd	430	NASRSEKIEIVISRA	1.16
DIP1281	H2-IAb	364	AAAALIEAATPNHAS	1.16
DIP1281	H2-IAb	367	ALIEAATPNHASLDD	1.20
DIP1281	H2-IAb	487	GLTLYAFAGVGIALP	1.28
DIP1281	H2-IAAd	325	KA EKARQDAEKAAEA	1.29
DIP1281	H2-IAb	488	LTLYAFAGVGIALPH	1.30
DIP1281	H2-IAb	486	SGLTLYAFAGVGIAL	1.31
DIP1281	H2-IAAd	272	EEARKKAEAEAAAAE	1.33
DIP1281	H2-IAAd	331	QDAEKAAEAAQAQAE	1.33
DIP1281	H2-IAAd	273	EARKKAEAEAAAAEK	1.40
DIP1281	H2-IAb	363	RAAAALIEAATPNHA	1.40
DIP1281	H2-IAb	485	CSSLTLYAFAGVGIA	1.41
DIP1281	H2-IAb	489	TLYAFAGVGIALPHY	1.46
DIP1281	H2-IAb	556	VSPVRWSGMSPYAVR	1.56
DIP1281	H2-IAAd	326	AEKARQDAEKAAEAA	1.57
DIP1281	H2-IAb	555	KVSPVRWSGMSPYAV	1.61
DIP1281	H2-IAb	557	SPVRWSGMSPYAVRL	1.63
DIP1281	H2-IAAd	1	MTIRSSMSSRTTTKV	1.64
DIP1281	H2-IAAd	2	TIRSSMSSRTTTKVM	1.69
DIP1281	H2-IAAd	213	KASAQQAIDSSNSKL	1.73
DIP1281	H2-IAAd	172	QKQREAVDALDRLRT	1.76
DIP1281	H2-IAAd	295	ARKAKEEAARIAAEA	1.79
DIP1281	H2-IAAd	208	EAEAKASAQQAIDS	1.81
DIP1281	H2-IAAd	332	DAEKAAEAAQAQAEA	1.83
DIP1281	H2-IAAd	322	AAEKAEKARQDAEKA	1.90
DIP1281	H2-IAAd	329	ARQDAEKAAEAAQAQ	1.92
DIP1281	H2-IAb	558	PVRWSGMSPYAVRLI	1.93
DIP1281	H2-IAAd	330	RQDAEKAAEAAQAQA	1.94
DIP1281	H2-IAAd	7	MSSRTTTKVMRGLIA	1.95
DIP1281	H2-IAAd	214	ASAQQAIDSSNSKLE	1.95
DIP1281	H2-IAAd	170	NAKQREAVDALDRL	2
DIP1281	H2-IAAd	283	AAAEKARQEAEARK	2.15
DIP1281	H2-IAAd	271	AEEARKKAEAEAAAA	2.17
DIP1281	H2-IAAd	364	AAAALIEAATPNHAS	2.21

DIP1281	H2-IAd	270	RAEARKKAEAEAAA	2.25
DIP1281	H2-IAd	102	SAERARQGVTVARKK	2.27
DIP1281	H2-IAd	5	SSMSRTTTKVMRGL	2.28
DIP1281	H2-IAd	6	SMSRTTTKVMRGLI	2.31
DIP1281	H2-IAd	323	AEKAEKARQDAEKAA	2.31
DIP1281	H2-IAd	282	AAAAEKARQEAEAAAR	2.39
DIP1281	H2-IAd	440	VISRAMAYIGTPYAW	2.41
DIP1281	H2-IAd	324	EKAEKARQDAEKAAE	2.48
DIP1281	H2-IAd	306	AAEAARKAADEAARK	2.48
DIP1281	H2-IAb	521	RGDLIFYGPNAEYHV	2.49
DIP1281	H2-IAd	278	AEAEAAAAEKARQEA	2.54
DIP1281	H2-IAd	319	RKVAAEKAEKARQDA	2.58
DIP1281	H2-IAd	171	AQKQREAVDALDRLR	2.62
DIP1281	H2-IAd	558	PVRWSGMSPYAVRLI	2.63
DIP1281	H2-IAd	321	VAAEKAEKARQDAEK	2.65
DIP1281	H2-IAd	441	ISRAMAYIGTPYAWG	2.65
DIP1281	H2-IAd	357	TANDSARAAAALIEA	2.67
DIP1281	H2-IAb	522	GDLIFYGPNAEYHVA	2.67
DIP1281	H2-IAd	101	SSAERARQGVTVARK	2.83
DIP1281	H2-IAb	368	LIEAATPNHASLDDP	2.83
DIP1281	H2-IAd	169	TNAQKQREAVDALDR	2.90
DIP1281	H2-IAd	136	RSAYRQGAVPSGVAG	2.96
DIP1281	H2-IAd	442	SRAMAYIGTPYAWGG	2.98
DIP2193	H2-IAd	294	ARRAGVEVQAFFRPS	1.75
DIP2193	H2-IAd	499	EIAKKYREMKTAESE	1.91
DIP2193	H2-IAd	204	GMPAAIRAAQKDAGG	2.21
DIP2193	H2-IAb	377	SGRAYWSPSTGAHAL	2.48
DIP2193	H2-IAb	379	RAYWSPSTGAHALYG	2.59
DIP2193	H2-IAb	378	GRAYWSPSTGAHALY	2.62
DIP2193	H2-IAd	295	RRAGVEVQAFFRPSG	2.83
DIP2193	H2-IAb	376	VSGRAYWSPSTGAHA	2.95
DIP2379	H2-IAd	55	MRSMRKMQELQPLMQ	0.38
DIP2379	H2-IAd	56	RSMRKMQELQPLMQE	0.38
DIP2379	H2-IAd	53	NQMRSMRKMQELQPL	0.43
DIP2379	H2-IAd	54	QMRSMRKMQELQPLM	0.47
DIP2379	H2-IAd	57	SMRKMQELQPLMQEI	0.84
DIP2379	H2-IAd	40	TFTIRALLVKPMLNQ	0.89
DIP2379	H2-IAd	288	AAEIEAKKAAKRATA	0.93
DIP2379	H2-IAd	285	AEEAAEIEAKKAAKR	1.03
DIP2379	H2-IAd	41	FTIRALLVKPMLNQM	1.04
DIP2379	H2-IAd	52	LNQMRSMRKMQELQP	1.04
DIP2379	H2-IAd	228	NDQMQMQMDMMNRMM	1.05
DIP2379	H2-IAd	286	EEAAEIEAKKAAKRA	1.07
DIP2379	H2-IAd	229	DQMQMQMDMMNRMML	1.10
DIP2379	H2-IAd	287	EAAEIEAKKAAKRAT	1.12
DIP2379	H2-IAd	58	MRKMQELQPLMQEIR	1.15
DIP2379	H2-IAd	230	QMQMQMDMMNRMMLW	1.16
DIP2379	H2-IAd	227	ANDQMQMQMDMMNRM	1.18
DIP2379	H2-IAd	39	LTFTIRALLVKPMLN	1.49
DIP2379	H2-IAd	51	MLNQMRSMRKMQELQ	1.69
DIP2379	H2-IAd	38	LLTFTIRALLVKPML	1.82
DIP2379	H2-IAd	283	IDAEAAEIEAKKAA	1.88
DIP2379	H2-IAd	205	FNARLSVDRQKARRE	2.07

DIP2379	H2-IAd	183	FTRTDIALVAAPLIV	2.29
DIP2379	H2-IAd	50	PMLNQMRSMRKMQEL	2.31
DIP2379	H2-IAd	284	DAEEAAEIEAKKAAK	2.36
DIP2379	H2-IAd	42	TIRALLVKPMLNQMR	2.48
DIP2379	H2-IAd	291	IEAKKAAKRATAPRP	2.50
DIP2379	H2-IAd	226	PANDQMQMMDMMNR	2.50
DIP2379	H2-IAd	182	DFTRTDIALVAAPLI	2.61
DIP2379	H2-IAd	43	IRALLVKPMLNQMR	2.62
DIP2379	H2-IAd	290	EIEAKKAAKRATAPR	2.70
DIP2379	H2-IAd	184	TRTDIALVAAPLIVI	2.82
DIP2379	H2-IAd	289	AEIEAKKAAKRATAP	2.97

Table S3: Predicted B cell conformational epitopes. The EliPro server was used to predict the conformational B cell binding epitopes via the 3D structure of the diphtheria multi-epitope vaccine. Eight epitopes were predicted with a score varying from 0.778 to 0.5 (default settings).

Residues and position	Number of residues	Score
A:R550, A:P563, A:G564, A:G566, A:Q567, A:M568, A:D569, A:A570, A:L571, A:A572, A:A573, A:G574, A:P575, A:D576, A:T577, A:A578, A:A579, A:A580, A:L581, A:G582, A:P583, A:G584, A:P585, A:G586, A:S587, A:S588, A:G602, A:P603, A:G604, A:P605, A:G606, A:Y607, A:P608, A:Q609, A:E610, A:Q611, A:V612, A:W613, A:A614, A:V615, A:V616, A:K617, A:A618, A:H619, A:L620, A:A621, A:G622, A:P623, A:G624, A:P625, A:G626, A:S627, A:G628, A:R629, A:A630, A:Y631, A:L641, A:G642, A:G644, A:P645, A:G646, A:K647, A:K648, A:Y649, A:R650, A:E651, A:M652, A:K653, A:T654, A:A655, A:E656, A:S657, A:E658, A:L659, A:G660, A:F661, A:G662, A:P663, A:G664, A:P665, A:G666, A:F667, A:T668, A:R669, A:I680, A:V681, A:G682, A:P683, A:G684, A:P685, A:G686, A:I687, A:V688, A:A689, A:Q690, A:S691, A:I692, A:A693, A:L694, A:S695, A:S696, A:L697, A:M698, A:V699, A:A700, A:Q701	106	0.778
A:Y393, A:D394, A:P395, A:P396, A:F397, A:K398, A:T399, A:A400, A:A401, A:G402, A:P403, A:G404, A:P405, A:G406, A:G407, A:Y408, A:A409, A:P410, A:T411, A:I412, A:A413, A:E414, A:A415, A:L416, A:S417, A:Q418, A:V419, A:P425, A:G426, A:N427, A:A428, A:V429, A:L430, A:P431, A:A432, A:K433, A:I434, A:V435, A:L436, A:L437, A:V438, A:I439, A:S440, A:A441, A:G442, A:P443, A:G444, A:P445, A:G446, A:K447, A:E448, A:P449	52	0.7
A:A5, A:V6, A:R7, A:L8, A:I9, A:A10, A:A11, A:Y12, A:V13, A:S14, A:P15, A:V16, A:R17, A:W18, A:S19, A:G20, A:M21, A:A22, A:A23, A:Y24, A:A25, A:A26, A:T27, A:H30, A:G42, A:V43, A:G44, A:I45, A:A46, A:A47, A:Y48, A:R49, A:E50, A:V51, A:T52, A:A53, A:Q54, A:D55, A:L56, A:I57, A:A58, A:A59, A:A85, A:E86, A:A87, A:L88, A:S89, A:Q90, A:V91, A:G92, A:I93, A:A94, A:A95, A:Y96, A:V97, A:L98, A:P99, A:A100, A:K101, A:I102, A:V103, A:L105, A:S121, A:T123, A:D124, A:I125, A:N126, A:A127, A:V128, A:A130, A:A131, A:Y132, A:F133, A:Y134, A:A135, A:F136, A:T137, A:L138, A:P139, A:A140, A:L141, A:A142, A:N161, A:A162, A:W163, A:A166, A:A167, A:K169, A:G170, A:V171, A:H172, A:R173, A:F174, A:L175, A:V176, A:V177, A:A178, A:A179, A:Y180, A:A200, A:A202, A:A203, A:Y204, A:A205, A:F206, A:S207, A:A208, A:L209, A:A210, A:S211, A:F212, A:L213, A:A214, A:A215, A:Y216, A:A217, A:E218, A:F223, A:P224, A:A227	120	0.696
A:L355, A:I356, A:K357, A:V358, A:I359, A:D360, A:S361, A:G362, A:G364, A:P365	10	0.692
A:G466, A:G467, A:S468, A:Y469, A:T470, A:D471, A:I472, A:N473, A:A474, A:V475, A:L476, A:P477, A:A478, A:I480, A:V481, A:P503, A:G504, A:P505, A:G506, A:V507, A:T508, A:K509, A:D510, A:L511, A:R512, A:I513, A:P514, A:A515, A:I516, A:S517, A:W540, A:P541, A:G542, A:P543, A:G544, A:P545, A:G546, A:V547, A:Q548	39	0.683

A:S2, A:P3, A:Y4	3	0.654
A:Y228, A:S231, A:T232, A:G233, A:A234, A:H235, A:A236, A:L237, A:A238, A:A239, A:I241, A:A242, A:V244, A:A245, A:A246, A:L248, A:I249, A:S271, A:G272, A:M273, A:S274, A:P275, A:Y276, A:A277, A:V278, A:R279, A:L280, A:I281, A:G282, A:P283, A:G284, A:P285, A:G286, A:S288, A:R291	35	0.629
A:E196, A:Q197, A:W199	3	0.587

Table S4: Top 15 democratic solutions. The top 15 docking poses were visualized by Chimera and the hydrogen bonds calculated. The solution with the lowest energy/mean energy, higher number of contacts and H-bonds is the probable correct structure

	Structure model	Mean Energy	Number of Members	Mean Energy deviation	Total Contacts	Total H-bonds	H-bonds located at the receptor binding sites
1	76d.pdb	-43.180	1	0.000	657	15	8
2	77d.pdb	-56.700	1	0.000	615	4	1
3	76b.pdb	-45.560	1	0.000	694	9	6
4	58a.pdb	-27.560	1	0.000	608	9	4
5	59a.pdb	-23.120	1	0.000	532	7	2
6	10b.pdb	-9.630	1	0.000	674	12	11
7	58c.pdb	-38.575	2	1.595	653	12	10
8	11b.pdb	-13.960	1	0.000	718	14	12
9	60c.pdb	-26.900	1	0.000	647	14	10
10	26d.pdb	-23.660	3	1.721	415	5	5
11	25b.pdb	-36.323	3	3.153	499	6	5
12	39b.pdb	-28.020	1	0.000	407	9	3
13	72b.pdb	-49.020	2	3.430	986	19	6
14	39d.pdb	-26.292	5	2.215	411	9	3
15	59b.pdb	-25.720	2	1.700	356	5	3

CHAPTER II:

A Reverse Vaccinology and Immunoinformatics Based Approach for Multi-epitope vaccine against *Clostridioides difficile* infection

1. Environmental background

1.1. The neat interplay between the gut microbiota and the host

The human GIT is an important interface between the host and microorganisms. The gut microbiota has a mutually beneficial relationship with the host and a crucial impact in our anatomical, physiological and immunological development (IEBBA *et al.*, 2016; SHI *et al.*, 2018; SILVA *et al.*, 2015). The GIT composition during the healthy physiological state known as eubiosis is dominated by the normal microbiota (i.e. 'good' bacteria), which is composed by two dominant phyla Bacteroidetes and Firmicutes plus minor constituents from Actinobacteria, Proteobacteria, Fusobacteria, and Verrucomicrobia (ARORA; BÄCKHED, 2016; KIM; COVINGTON; PAMER, 2017). The usual microbiota composition is defined early in life (e.g. type of birth, antibiotic administration during infancy, diet) and pass through changes over time according to individual life style, which incite changes in the host (SILVA *et al.*, 2015). This interplay between the host and the microbiota provides many advantages (e.g. the microbiota yield metabolites and provides protection against pathogens) and help shape the host fitness status. Likewise, the microbiota capability to modulate the host immune system and metabolism aspects have a profound impact on the host. The metabolic compounds produced by the mutualistic relation microbiota/host impact intestine functions, like visceral-sensing, motility, digestion, permeability secretion, energy harvest, mucosal immunity, and barrier effect (MONTALTO *et al.*, 2009).

The gut microbiota is direct or indirect involved with the generation of metabolites produced by fermentation of carbohydrates and amino acids, like SCFAs (e.g. acetate, butyrate and propionate); the metabolism and the absorption of indigestible polysaccharides (e.g. pectin, cellulose, hemicellulose) and foreign compounds (e.g. drugs); the production of essential vitamins; and is required for the biotransformations that generates primary and secondary bile acids (SMITH; MCCOY; MACPHERSON, 2007; SOMMER; BÄCKHED, 2013; SONNENBURG, E. D. *et al.*, 2006). SCFA and bile acids are bacterial derived molecules that has major impact on host functions and microbiota regulation; therefore, they are the main metabolites disclosed in this work. SCFA enhances intestinal barrier and stimulates mucus and AMP production (anti-inflammatory properties) (KAMADA *et al.*, 2013; SILVA *et al.*, 2015). Meanwhile, bile acids are amphipathic sterol-derivative compounds, which endow antimicrobial activity, required for the absorption of lipids and act as signaling molecules in the GIT (DE AGUIAR VALLIM; TARLING; EDWARDS, 2013; HOFMANN, ALAN F., 1999).

Bile acids are a major component of the bile produced by the liver and stored in the gallbladder, and, after food intake, bile is released in the intestine to ensure the digestion of lipids by enabling the emulsification and the transport of fat nutrients. Bile acids when combined with sodium (Na^+) or potassium (K^+) ions form primary [cholate (CA) and chenodeoxycholate (CDCA)] or secondary [deoxycholate (DCA) and lithocholate (LCA)] bile salts. Furthermore, primary and secondary bile salts can be conjugated with glycine or taurine (Figure 17), which increase their solubility, resistance to precipitation by calcium (Ca^{2+}) ions, and decrease their potential to disrupt the cell membrane (HOFMANN, A. F., 1999; HOFMANN, A. F.; HAGEY, 2008). Their amount varies widely among individuals, being influenced especially by diet, life style and microbiota composition (RIDLON *et al.*, 2014; RIDLON; KANG; HYLEMON, 2006). The primary conjugated bile acids are delivered into the small intestine (duodenum) and about 95% of the bile acid pool, including mainly the primary bile acids, are actively reabsorbed (i.e. atypical sodium-dependent bile salt transporter or the ileum bile acid transporter), when performing their digestive function in the ileum, and return to the liver (enterohepatic circulation). The remaining primary conjugated bile salts are deconjugated, in the terminal ileum/colon, by gut bacteria, through the action of bile salt hydrolases (BSH), which increase bile tolerance and survival in the gut. Afterwards, the bile acid pool (i.e. predominantly primary free bile salts) reaches the colon where specific gut microbiota converts them into secondary bile acids through the removal of the C-7 hydroxyl group ($7\alpha/\beta$ -dehydroxylation) (irreversible reaction) (Figure 17). Part of the secondary bile salts and the others bile salt types present in the colon are passively reabsorbed and reach the liver or are excreted in the feces. Once in the liver, they are conjugated with taurine or glycine to form conjugated secondary bile salts and eventually released in the small intestine, where they can be deconjugated or not and subsequently reabsorbed or excreted (Figure 17).

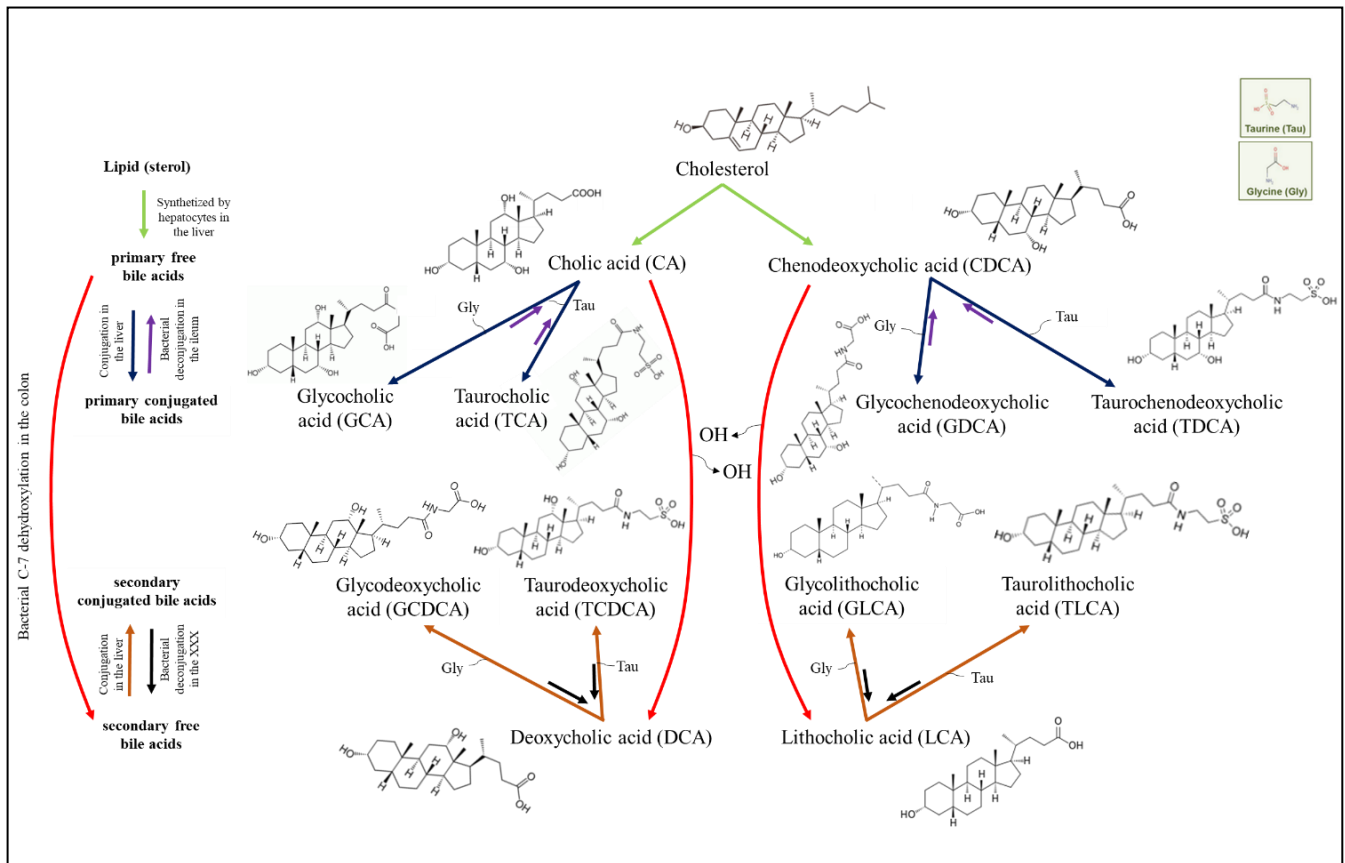


FIGURE 17: Bile acid differentiation in the gut. The bile acids/salts can be classified into primary, secondary or conjugated. Intestinal bacteria catalyze the differentiation between primary and secondary as the deconjugation of bile salts.

Bile and most particularly bile acids are able to modulate the composition of the gut microbiota, reciprocally. Such an interplay between composition of bile acids and gut microbiota is a fundamental and delicate crosstalk, which tunes host physiology and is sensitive to exogenous factors such as diet and antibiotics (RIDLON *et al.*, 2016; SONNENBURG, ERICA D. *et al.*, 2016).

1.2. The frail line between eubiosis and dysbiosis

The phenomenon termed ‘colonization resistance’ is a recognized protection conferred by the gut microbiota where it plays a critical role in excluding pathogens arriving from outside and inhibiting the overgrowth of pathobionts.

To protect itself against microbes and prevent an inflammatory state during eubiosis, the host, with the assist of the microbiota, releases AMP, reactive oxygen species (ROS), Ig, yield a protective barrier (mucus layer), influence the transit time via peristalsis, which could lead to an alteration of the intestinal nutritional environment (HOOPER; LITTMAN; MACPHERSON, 2012; SALZMAN *et al.*, 2010; SOMMER *et al.*, 2014). In addition, the GIT community stimulates the expression of pro-inflammatory molecules like IFN- γ and IL (e.g. IL-10, IL-17, IL-22) (immune profile responses Th1 and Th17, respectively) (IEBBA *et al.*, 2016; KINROSS; DARZI; NICHOLSON, 2011; MARTIN *et al.*, 2007).

A dysbiotic gut microbiota represents a shift in the stability and composition of the microbial community. The disruption of the eubiosis state is associated to a variety of causes (i.e. genetics of the host, antibiotic treatments, diet, or infections), and changes in the community lead to drastic changes in the gut metabolic process (e.g. changes in the lipid and vitamin profile, aminoacids and methylamine metabolism, fermentation of different carbohydrates, production of bile acids) and immunological profile (e.g. cross-reactive antibodies like the IgA), consequently in the intestine functions (BIEDERMANN; ROGLER, 2015; BRANDL *et al.*, 2008; DETHLEFSEN; RELMAN, 2011; HEINSEN *et al.*, 2015; SOMMER *et al.*, 2017; UBEDA *et al.*, 2010). In addition, the shift in the population alters the physiochemical parameters in the intestine (e.g. pH, oxygenation levels, osmolality), the concentration of peroxides, nitrogen compounds, among others (BIEDERMANN; ROGLER, 2015; IEBBA *et al.*, 2016).

The indiscriminate use of antibiotics is one of many causes of the gut microbiota imbalance (dysbiosis), and antibiotic-induced intestinal dysbiosis (AID) is a major medical concern in terms of life threatening and health costs (BECATTINI; TAUR; PAMER, 2016; KIM; COVINGTON; PAMER, 2017). The use of antibiotics is associated with many nosocomial (i.e. hospital acquired) and infectious diseases (Becattini *et al.*, 2016; Brandl *et al.*, 2008; Jernberg *et al.*, 2007; Ubeda *et al.*, 2010), among other pathologies, like the development of asthma (i.e. children with 6–7 years) when antibiotics are applied early in life (FRÖHLICH *et al.*, 2016), and brain cognitive impairment (e.g. deficit in novel object recognition memory) in mice, due to profound alterations in the metabolites [i.e. gut bacteria are involved (direct or indirect) in a variety of biotransformations] (RISNES *et al.*, 2011). Therefore, the comprehension of the

dysbiotic environmental settings and the interplay of these new conditions with the so-called ‘bad bacteria’ are key steps to pinpoint methodologies to re-establish a healthy complex microbiota and prevent these changes to the microbial community structure.

The alteration of the gut microbiota modifies a great variety of metabolites, which in turn can favor the overgrowth of low proportion bacterial groups and can result in an inflammatory status and a weakness of the intestinal integrity, which may cause the translocation of pathogens to other organs (KIM; COVINGTON; PAMER, 2017; SHI *et al.*, 2018). For instance, increase levels of succinate can promote the overgrowth of *C. difficile* and is associated with the development of colitis, cell death, the generation of adenosine triphosphate, and has signaling capacity (e.g. enhance IL-1 β production) (FERREYRA *et al.*, 2014; TANNAHILL *et al.*, 2013). The use of ampicillin can lead to this increase of succinate, acetate, lactate, and tauro-bile acids in the intestinal lumen; meanwhile, driving a decrease in butyrate and 2-oxoisovalerate (SHI *et al.*, 2018). Furthermore, the concentration of sialic acid is increase during AID and promotes *C. difficile* proliferation (NG *et al.*, 2013; VIMR *et al.*, 2004). This nutrient act as source of energy, carbon, and nitrogen to the indigenous commensal population; and, after the decrease of this population, the niche can be used by antibiotic resistant bacteria, as other available nutrients (KIM; COVINGTON; PAMER, 2017).

Many alterations in the metabolomics profile can be associated with the increase or decrease of a single specie or genera (BIEDERMANN; ROGLER, 2015; GUINANE; COTTER, 2013; KINROSS; DARZI; NICHOLSON, 2011; LI, S. *et al.*, 2014). The commensal (*n*-butyrateproducing) bacteria *Faecalibacterium prausnitzii* was linked to the production of the metabolites dimethylamine, lactate, taurine and glycine (i.e. byproducts after bile acids deconjugation), 2-hydroxyisobutyrate, glycolate, 3,5-hydroxybenzoate, and 3-aminoisobutyrate; the bacteria *Bacteroides thetaiotaomicron* was linked to 3-aminoisobutyrate and produces high levels of succinate; *Ruminococcus bromii* to creatinine and 3-aminoisobutyrate; among others connections (Figure 18). The decrease of commensal bacteria impair the production of these microbiota-associated metabolites, like the byproducts of bile acid deconjugation. The commensal bacteria is responsible for the reaction that generates free primary plus glycine or taurine and secondary bile acids; therefore, after dysbiosis, the concentration of conjugated bile acids increase and the microbiota associated products decrease.

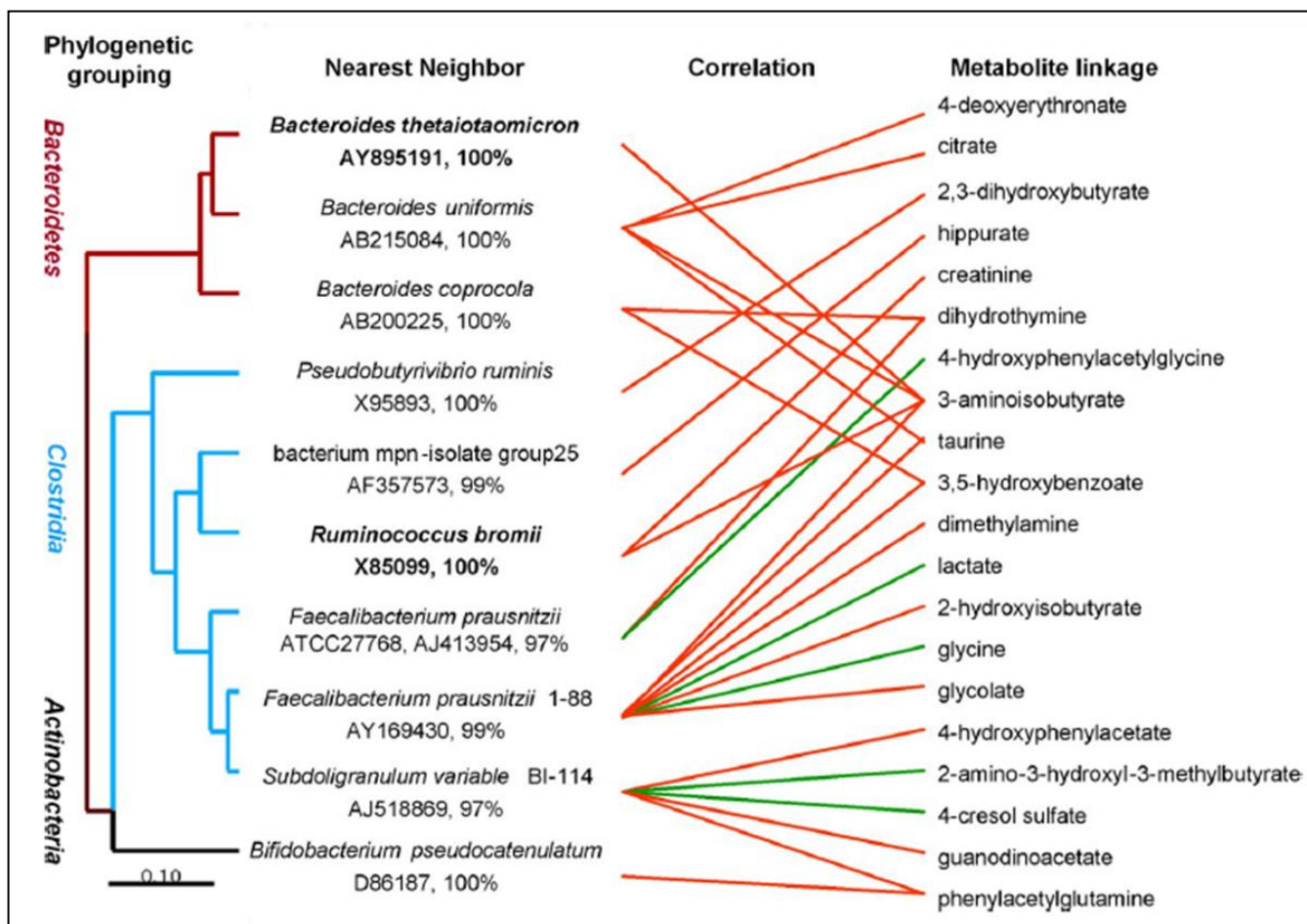


FIGURE 18: Correlation between gut bacteria and metabolite. Dendrogram of operational taxonomic units (OTUs) associated with specific metabolites. The color of the lines linking OTUs and metabolites indicates a positive (red) or negative (green) correlation. Figure adapted from (LI, M. et al., 2008).

Considering the dysbiotic state provoked by the exacerbate use of antibiotics (i.e. shift of the gut microbiota and metabolites modifications), this chapter will focus in the opportunistic pathogen *C. difficile*, a enteric bacterium that can cause symptoms ranging from diarrhea to life-threatening inflammation of the colon (e.g. IBDs and CDI) (BARTLETT, 2008; BUFFIE *et al.*, 2012; HE, M. *et al.*, 2013). We first present a short summary of the main biological features of this bacterium and how *C. difficile* is able to adapt and overgrowth in hash conditions (i.e. dysbiosis), which culminate in CDI. Then, we emphasize the disease characteristics and current treatments, to finally provide an alternative prevention approach to cope with the illness.

2. The pathogen *Clostridioides difficile*

2.1. *C. difficile* phylogeny and main features

C. difficile was first isolated in 1935 from the intestines of infants and assigned to the genus *Clostridium* due to its morphology, ability to form spores and inability to undergo vegetative growth in aerobic conditions. Prazmowski proposed this genus in 1880 and many distinct species were grouped together based on traditional microbiological methods. Recent genetic analyses (e.g. 16S rRNA and DNA-RNA paring) place *C. difficile* in the Peptostreptococcaceae family (rRNA clostridia cluster XI) and *Peptoclostridium* genus; however, the change to *P. difficile* would entail worldwide consequences since this bacterium has a global medical concern being the major cause of hospital and community acquired diarrhea (COLLINS *et al.*, 1994; LAWSON *et al.*, 2016; LAWSON; RAINEY, 2016; YARZA *et al.*, 2014; YUTIN; GALPERIN, 2013). Bearing this and the international disclosure of the denomination ‘*C. difficile*’, ‘*C. diff*’, and ‘*C. difficile* infection’ or ‘CDI’, Lawson and colleagues (2016) proposed a designation (i.e. *Clostridioides difficile*) that would retain the above mentioned names requiring only modest changes to packaging, documents and computer sources of most commercial, scientific, and healthcare sectors.

The obligate anaerobic and spore-forming pathogen *C. difficile* is a Gram-positive motile bacillus with optimum growth on blood agar at human body temperature (37°C). Strains can utilize fructose and selected amino acids (i.e. proline, aspartic acid, serine, leucine, alanine, threonine, valine, phenylalanine, methionine, and isoleucine) as source of carbon and energy. Bile act as a growth stimulant since this bacterium can metabolizes the bile acids CA and CDCA and splits the conjugate taurocholic acid or taurocholate (TCA), using the later to enhance spore germination. Furthermore, all *C. difficile* strains are resistant to aminoglycosides and cephamycins whereas some are susceptible to penicillin, ampicillin, vancomycin, metronidazole, clindamycin, rifampin, tetracycline, chloramphenicol, and/or erythromycin (LAWSON *et al.*, 2016).

Its vegetative cells are rod shaped (i.e. long irregular cells with a bulge at their terminal ends), pleomorphic, and occur in pairs or short chains (e.g. two to six cells aligned end-to-end). *C. difficile* spores are oval and subterminal (rarely terminal), being resistant to heat, oxygen and common disinfectants, such as ethanol-based hand sanitizers, which facilitates the bacterium spread. Spores can germinate and produce new vegetative cells when conditions become favorable (i.e. lower GIT) thought the activation of crucial

proteins, like the transcription factor stage 0 sporulation A (Spo0A) that drives the sporulation regulatory network and directly regulates dozens of genes (reviewed at Abt et al., 2016).

The bacterium is transmitted via the fecal–oral route (i.e. oral ingestion of spores), often by contact with asymptomatic- or symptomatic-infected people. The organism itself is noninvasive, and infection outside the colon is extremely rare; however, once it colonizes the large intestine, the bacterium secretes two exotoxins TcdA (also known as ToxA) and TcdB (also known as ToxB) that leads to the disruption of the host cytoskeleton and dissociation of tight junctions (i.e. loss of epithelial integrity). The non-toxicogenic *C. difficile* strains do not present the pathogenicity locus (PaLoc) with the main *C. difficile* virulence factors, TcdA and TcdB plus TcdR, TcdC and TcdE, which was naturally replaced by a 115 bp sequence (GERDING; SAMBOL; JOHNSON, 2018). This substitution decreased the adherence properties into the GIT *in vivo*, and, in accordance with Gerding and colleagues (2018), non-toxicogenic strains may successfully prevent CDI or recurrent CDI probably via competition, an event also observed by other authors (BORRIELLO; BARCLAY, 1985; CORTIER; MULLER, 1988; GERDING *et al.*, 2015). Furthermore, non-toxicogenic strains are more common than toxicogenic strains among colonized infants, but colonization is transient and different strains are found to colonize the same infant at different times (MCDONALD, L CLIFFORD *et al.*, 2018).

C. difficile genome is highly variable especially considering the presence of transposable elements (e.g. conjugative transposons), bacteriophages and CRISPRs arrays that are involved in homologous recombination and horizontal gene transfer (reviewed at Janezic and Rupnik, 2015; Knight et al., 2015). These mechanisms shape the diversity and evolution (i.e. long and short time period) of *C. difficile* through the acquisition of foreign DNA leading to a mosaic gene structure. Likewise, the ultralow level of genomic conservation and the large genome (i.e. ~4.3 Mb) support the bacterium plasticity to thrive and adapt in multiple and adverse environments and in highly dynamic niches (e.g. can coexist within the host for long periods) (KNIGHT *et al.*, 2015). The low conservation is demonstrated in *C. difficile* pan genome, which comprise 9,640 coding sequences with a core genome estimated size of 600 to 3,000 coding sequences (HE, M. *et al.*, 2010; SCARIA *et al.*, 2010). The contrasting divergence between the genomic size and the number of core genes highlights the genetic pool exchange between the bacterium and multiple host environments; hence, the genome is characterized as ‘open’, which indicates that the number of core and accessory genes will change with additional sequence strains (HALL; EHRLICH; HU, 2010; KNIGHT *et*

al., 2015). The core genes are involved in essential processes (e.g. metabolism, cell division, transport, biosynthesis, and pathogenesis) and in housekeeping functions being normally found outside the horizontally acquired regions whereas the accessory genome are mostly related with the bacterium ability to survive in harsh and diverse environments (JANEZIC; RUPNIK, 2015; JANVILISRI *et al.*, 2009; KNIGHT *et al.*, 2015; STABLER *et al.*, 2006). Nonetheless, both core and the accessory genome contributes with *C. difficile* overgrowth and competence to cause infections, in suitable conditions; and, thus, some of these central genes will be emphasized in this work.

2.2. *C. difficile* virulent factors

2.2.1. The Pathogenic locus proteins

The ability of *C. difficile* to cause illness depends on a range of virulence factors, including toxins, adherence and motility factors. The best known are the genes encoded in the PaLoc, which expresses the toxins TcdA and TcdB, two regulatory proteins TcdC and TcdR, and a porin TcdE. The expression of *tcdA* and *tcdB* is positively regulated by *tcdR* and negatively by the *tcdC* (ABT; MCKENNEY; PAMER, 2016; GIL *et al.*, 2018).

Toxin A and B are the major virulent factors of *C. difficile* and are directly involved in CDI symptoms. TcdA is denoted as an enterotoxin because it causes fluid accumulation in the bowel whereas TcdB does not but is considered cytopathic for all tissue-cultured cells tested (LOO *et al.*, 2005). All *C. difficile* toxigenic strains express TcdB but not TcdA, which suggests a predominant role of ToxB in CDI (GIL *et al.*, 2018). Furthermore, *tcdB* sequence is highly diverse among *C. difficile* strains and this differences in the toxin may contribute to heterogeneity of strains virulence (LANIS *et al.*, 2013; SEBAIHIA *et al.*, 2006; STABLER *et al.*, 2006). Lanis and colleagues (2013) detected that TcdB from the epidemic BI/NAP1/027 strain had a four-fold greater toxicity in mice compared with TcdB from typical *C. difficile* strains.

ToxA and ToxB primarily target intestinal epithelial cells, which leads to loss of intestinal membrane integrity followed by host exposure to intestinal microorganism and activation of the host inflammatory response. Yet, their action/toxic effect is variable and diverges in line with the host (reviewed at Abt *et*

al., 2016; Figueroa Castro and Munoz-Price, 2019; Gil et al., 2018; Rupnik et al., 2009). TcdA and TcdB are lethal for mice but vary their impact on hamster, once only TcdB is fatal when given orally while the intraperitoneally injection administration of both toxins is fatal. The combined deletion of TcdB and TcdA completely abolished *in vivo* virulence in hamster model; however, a single deletion of TcdA did not altered *C. difficile* virulence, which further confirm the central role of TcdB in CDI (KUEHNE *et al.*, 2010, 2014). Indeed, TcdB is considered the major virulence factor that mediates colonic epithelial damage, inflammation and mortality in the murine model while TcdA is a relatively minor driver of inflammation in mice and is slightly more toxic in hamsters (CARTER *et al.*, 2015).

The other members of *C. difficile* PaLoc are TcdR and TcdC, which regulates the expression of *C. difficile* main virulent factors, and the holin-like protein TcdE that is probably involved in facilitating the secretion of TcdA and TcdB (ABT; MCKENNEY; PAMER, 2016). TcdR is an alternative sigma factor that assists the binding of RNA polymerase to the promoter regions of the *tcdA*, *tcdB* and to its own promoter (MANI; DUPUY, 2001); meanwhile, TcdC probably act as an anti-sigma factor suppressing the transcription of *tcdA* and *tcdB* via direct association with single-stranded DNA (MATAMOUROS; ENGLAND; DUPUY, 2007; VAN LEEUWEN *et al.*, 2013). Additionally, hypervirulent *C. difficile* strain may be associated with *tcdC* mutations, which supports the potential role of TcdC in limiting toxin expression (SPIGAGLIA; MASTRANTONIO, 2002).

2.2.2. Cytotoxic distending toxin (CDT)

Some hypervirulent strains also express the binary toxin or cytolethal distending toxin (CDT). CDT is composed of two non-PaLoc genes that express CdtA and CdtB, two proteins associated with components of the flagellar machinery and with a higher mortality in patients (GERDING *et al.*, 2014).

CdtA is a ADP-ribosyl transferase that ribosylates actin in eukaryotic cells whereas CdtB forms pores in acidified endosomes and facilitates the transfer of CdtA to the cytosol (ABT; MCKENNEY; PAMER, 2016). The binary toxin is presumably linked with the increase of pathogen adherence through the suppression of colonic eosinophilia and the hijack of the microtubule organization (COWARDIN *et al.*, 2016; SCHWAN *et al.*, 2014). To accomplish *C. difficile* adhesion, CdtA interferes with the actin

meshwork polymerization that underlies the cell membrane, which results in cellular protrusions and enhanced fibronectin delivery to the cell surface (SCHWAN *et al.*, 2014).

2.2.3. Others (non-toxins related) virulence factors

The regulation of genes that control motility and adherence is an important factor that contributes to *C. difficile* colonization and virulence (AWAD *et al.*, 2014; VEDANTAM *et al.*, 2012). The presence of the flagellar machinery is connected with the bacterium ability to adhere to the host. Flagellar expression is highly variable among *C. difficile* strains and the lack of flagella was linked to impaired adherence to the intestinal epithelium, deregulated toxin expression and corresponding altered virulence *in vivo* (AUBRY *et al.*, 2012; BABAN *et al.*, 2013; PITUCH *et al.*, 2002). The results suggests an association between flagellar expression and toxin regulation.

The flagellum expression is regulated by the intracellular second messenger cyclic dimeric guanosine monophosphate (c-di-GMP) synthesized from GTP by diguanylate cyclase (BORDELEAU; BURRUS, 2015). This molecule binds to a riboswitch upstream of the *flg* operon to regulate/terminate the transcription of *flgB*, and this repression prompt a decay in *C. difficile* motility and repressed the synthesis of TcdA and TcdB (MCKEE *et al.*, 2013; PURCELL *et al.*, 2012; SUDARSAN *et al.*, 2008). Concurrently, c-di-GMP is also involved in *C. difficile* aggregation and biofilm formation via activation of another riboswitch that induces the expression of type IV pili (PURCELL *et al.*, 2016). Thus, this messenger acts as a key signal that can switch *C. difficile* between a highly mobile, toxin-producing state and a strongly adherent biofilm-producing state (ABT; MCKENNEY; PAMER, 2016).

C. difficile transmission via spore is a crucial strategy to survive in harsh environments and invade the host. The environmental signals that elicit sporulation remain unknown but the sporulation apparatus are well described. Sporulation is triggered via sensor histidine kinases signaling that ultimately results in the phosphorylation and activation of the Spo0A (EDWARDS; MCBRIDE, 2014). Spo0A is essential for *C. difficile* sporulation and mutants that lack this gene only exist as vegetative cells, which decreases the spread of the microorganism in the host as its resistance capabilities (DEAKIN *et al.*, 2012). This protein is also linked with toxin production and the bacterium ability to harvest nutrients from peptides in the

local environment, which leads to the starvation of cells and accelerates the transition to stationary phase (ANTUNES, A.; MARTIN-VERSTRAETE; DUPUY, 2011; EDWARDS; NAWROCKI; MCBRIDE, 2014).

Bearing all, the gut environment is acutely associated with *C. difficile* and a deeper understanding of the environmental settings and its relation with *C. difficile* dissemination will provide important insights into the epidemiology of this pathogen.

2.3. The interplay between *C. difficile* and the gut environment

As mentioned, AID features the decrease of gut microbiota diversity, reducing the abundant bacterial taxa prior to treatment, accompanied by the expansion of bacteria present in low or undetectable levels (Aguilera et al., 2015; Becattini et al., 2016; Jernberg et al., 2007; McDonald, 2017; Urdaneta and Casadesús, 2017). Consequently, this phenomenon leads to the decreased amount of secondary bile acids in sharp contrast to the increased abundance of conjugated primary bile acids (LEWIS; CARTER; PAMER, 2016; THERIOT; BOWMAN; YOUNG, 2016; WAHLSTRÖM *et al.*, 2017). Such unbalance favors the germination of *C. difficile* spores leading to the multiplication of the vegetative bacterium form that produces the lethal toxins TcdA and TcdB (BUFFIE *et al.*, 2012; KUEHNE *et al.*, 2010; LEWIS; CARTER; PAMER, 2016) (Figure 18).

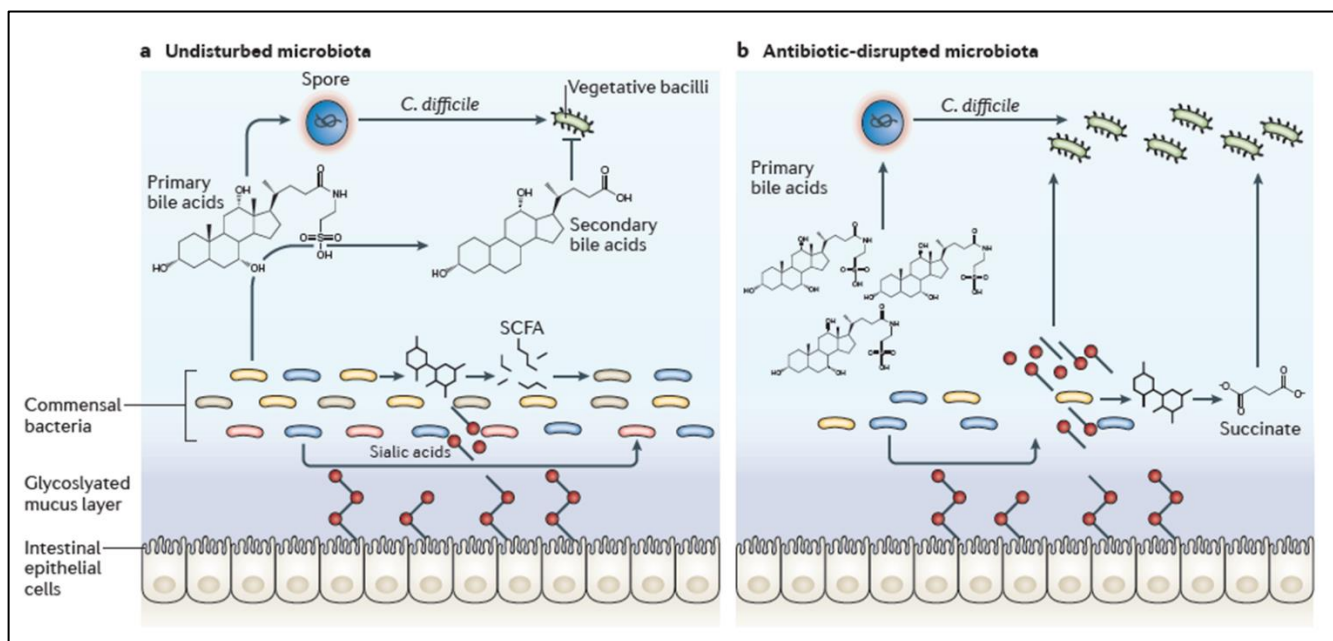


FIGURE 19: The interplay between *C. difficile*, the microbiota and metabolites. (a) During eubiosis, the gut microbiota converts primary bile acids into secondary bile acids responsible for inhibit the growth of the vegetative bacilli form of *C. difficile*. Commensal bacteria release sialic acid into the intestinal lumen. Fermenting commensal bacterial species convert carbohydrates into short-chain fatty acids (SCFAs), such as succinate. Commensal bacterial populations can consume these metabolites as energy sources. (b) Antibiotic-mediated dysbiosis (AID) decrease the microbiota (e.g. commensal) population, which reduce the conversion of conjugated primary bile acids to primary bile acid, enabling *C. difficile* sporulation and growth. Antibiotics deplete competing sialic acid and succinate consumers, liberating an energy source for *C. difficile*. Figure from (ABT; MCKENNEY; PAMER, 2016).

The shift from primary conjugated bile acids to secondary bile acids is accomplished after two main steps (Figure 17). The first step is undertaken by BSHs, which deconjugate the taurine and glycine groups via a hydrolysis reaction, and consequently generate the unconjugated primary bile acids CA and CDCA that can then be further modified to secondary bile acids.

BSHs belongs to the N-terminal nucleophil hydrolase superfamily of proteins being widely distributed throughout most major bacterial species/phylum, ranging from indigenous Gram-positive gut microbiota from the Firmicutes phylum (e.g. *Lactobacillus*, *Enterococcus*, *Listeria monocytogenes*, and *Clostridium*), to Gram-negative species, intestinal Archaea, and from different environments like soil and water. The number of different alleles in each bacterium can vary up to four (e.g. *Lactobacillus plantarum*), and evidences suggest that *bsh* can be horizontally transferred (JONES *et al.*, 2008; LAMBERT *et al.*, 2008;

REN *et al.*, 2011) and probably goes through a host species-specific evolutionary selection (JONES *et al.*, 2008; O'FLAHERTY *et al.*, 2018).

The absence of commensal gut bacteria leads to a decrease of BSH with the consequent enrichment of TCA and glycine, which promote *C. difficile* spore co-germination, alongside CA, without affecting *C. difficile* vegetative growth, and a decay of taurine, an amino acid used by BSH-positive strains of *Clostridium* as an electron acceptor in Stickland fermentation reactions (HEEG *et al.*, 2012; KUMAR, M. *et al.*, 2012; SORG; SONENSHEIN, 2008; WILSON; KENNEDY; FEKETY, 1982) (Figure 19b). The crucial role of BSHs in the maintenance of eubiosis was further corroborated through fecal microbiota transplantation (FMT) experiments, which provide evidence that this enzyme is a key mediator of the efficacy of FMT for recurrent CDI (MULLISH *et al.*, 2019). The author observed that by reversing a 'metabolic dysbiosis' via BSH potentially inhibits the growth of *C. difficile* without any apparent adverse effect on gut commensal bacteria.

The second enzymatic step is 7- α -dehydroxylation, whereby unconjugated primary bile acids are converted to secondary bile acids (Figure 17). The reaction 7 α -dehydroxylation (i.e. removal of a hydroxyl (OH) group from the C7 position) is performed by a limited number of intestinal bacteria (i.e. *Clostridium scindens*, *C. hiranonis*, *C. hylemonae*, and *C. sordelli*). These bacteria have the bile acid inducible *bai* operon (*baiBCDEAFGHI*), a multi-step pathway that involves bile acid import, modification and export from bacterial cells (RIDLON *et al.*, 2014; RIDLON; KANG; HYLEMON, 2006; WAHLSTRÖM *et al.*, 2016). These modifications generate the secondary bile acids DCA and LCA from CA and CDCA, respectively, and significantly alter the signaling properties of the bile acid pool since the secondary bile acids act as potent agonists to different bacteria (e.g. *C. difficile*, *L. monocytogenes*, *E. faecalis*) and bile-specific cellular signaling receptors (CHAND *et al.*, 2017; DAWSON; KARPEN, 2015; WAHLSTRÖM *et al.*, 2016).

DCA has been shown to limit the outgrowth of the pathogen by inhibiting the growth of *C. difficile* vegetative, toxin-producing cells (GREATHOUSE; HARRIS; BULTMAN, 2015; HEEG *et al.*, 2012; THERIOT; BOWMAN; YOUNG, 2016) (Figure 19a). Yasugi and colleagues (2016) detected that DCA enhances the phosphorylation of Spo0A (i.e. response regulator responsible for the sporulation pathway in many species of bacilli and clostridia) and activates Spo0A-regulated genes (i.e. involved in the sporulation processes; stages II, III, IV and V). The Clostridia group have their growth rates improved in

the presence of taurine and tauroconjugated bile acids (KUMAR, M. *et al.*, 2012), and are linked with the generation of secondary bile salts and the SCFA butyrate, that together with acetate and propionate are the most abundant SCFA in the gut (Figure 19a). Those clostridia members are reduced during dysbiosis and, as result, also diminishes the production of butyrate, which may be associated with the expansion of the Enterobacteriaceae class, through still unknown mechanisms (BÄUMLER; SPERANDIO, 2016; FUKUDA *et al.*, 2011) (Figure 19b). In fact, different studies associate the presence of the commensal bacterium *C. scindens* with *C. difficile* inhibition *in vitro* and *in vivo* (GIEL *et al.*, 2010; KOENIGSKNECHT *et al.*, 2015; LEWIS; CARTER; PAMER, 2016; THERIOT *et al.*, 2014; THERIOT; BOWMAN; YOUNG, 2016). Thus, the presence of *C. scindens* is linked with a resistance to CDI and could be used as a new therapeutic approach to combat this infection (BUFFIE *et al.*, 2015; LEWIS; CARTER; PAMER, 2016).

Furthermore, CDCA, LCA, and isomer ursodeoxycholic acid (UDCA) where show to inhibit the germination stimulated by TCA in some *C. difficile* strains (HEEG *et al.*, 2012; LEWIS; CARTER; PAMER, 2016; SORG; SONENSHEIN, 2009, 2010). Yet, LCA also demonstrated to be extremely toxic to *C. difficile* strain *in vitro* and the bacterium ability to tolerant LCA exposure is directly linked to the virulence of the strain/the disease score (i.e. the more virulent is the strain more it tolerates LCA) (LEWIS; CARTER; PAMER, 2016).

The gut microbiota and consequently gut metabolites are direct linked with the progression of *C. difficile* infections and are already associated with possible and concrete treatments strategies. However, CDI still represents a health-care problem and new therapies are necessary to ameliorate the disease burden.

2.4. *C. difficile* infections

C. difficile proliferations is triggered after AID where the commensal bacteria are mostly depleted from the gut, which benefits the bacterium colonization and establishment of CDI (Figure 20). The high resistant capacity of *C. difficile* spore to endure in the acidity of the stomach favor its survival and, after the exposure to bile salts, the spores can germinate in the small intestine and reactivate the vegetative form in the anaerobic environment of the cecum and colon, in a susceptible host, to afterwards proliferate and

colonize the mucosa (PENICHE; SAVIDGE; DANN, 2013). *C. difficile* virulent factors (e.g. toxins, adhesin and sporulation factors) and host-microbiota associated metabolites (SCFA and bile acids) aid the bacteria colonization and cause dissociation of the epithelial cells tight junctions (favor bacterial translocation), but also trigger the host immune response to eliminate the pathogens. Specific *C. difficile* treatment (e.g. recommended antibiotics) can settle the infection after one episode or CDI can be reestablished (i.e. recurrent CDI) (GIL *et al.*, 2018). After treatment, the gut microbiota is restored as the gut normal functions. This process can take days, months or even years in accordance with the severity of the disease.

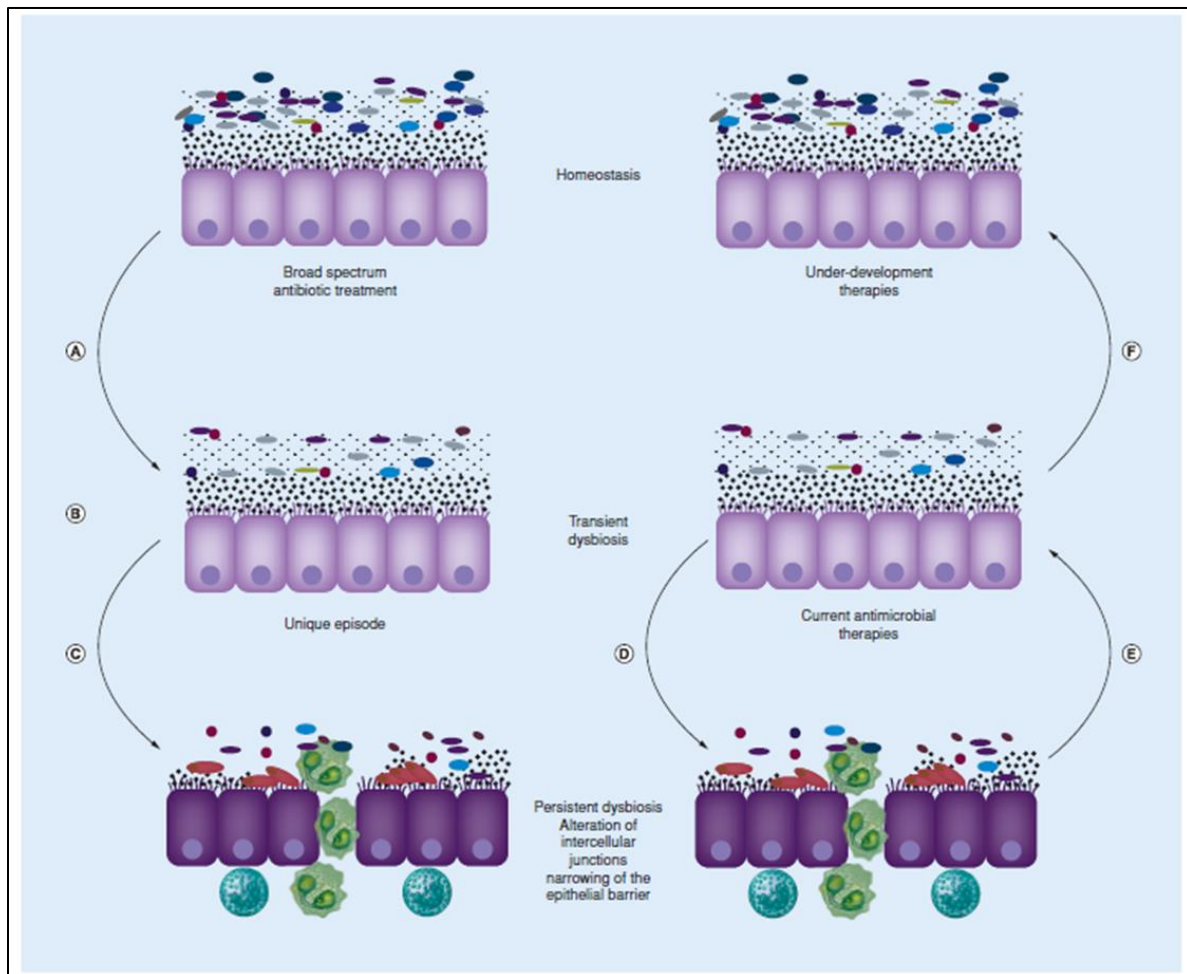


FIGURE 20: *Clostridioides difficile* infection. (A) Antibiotic treatment induce the establishment of an antibiotic induced dysbiosis (AID). (B) AID benefits *C. difficile* colonization, toxin production and sporulation. (C) The action of *C. difficile* virulent factors like TcdA, TcdB and CDT, prompt a severe damage at epithelium. (D, E) *C. difficile*

specific antibiotic treatment may lead to the resolution of infection after a unique episode (E) or one or more recurrences (D, E) before the elimination of *C. difficile* spores and vegetative forms. (F) The commensal gut microbiota is reestablished yielding the conclusion of the infection and the restoration of gut normal functions. Figure from (GIL *et al.*, 2018).

The number of *C. difficile* infections drastically increased mainly considering the emergence of new hypervirulent epidemic strains, such as the BI/NAP1/027 strain, the indiscriminate use of antibiotics and proton-pump inhibitors, the raising of multidrug resistance strains (MDR), certain types of chemotherapy, and prolong hospitalization (AZAB *et al.*, 2017; PÉPIN; VALIQUETTE; COSSETTE, 2005; PERETZ *et al.*, 2016; TRIFAN *et al.*, 2017). Although healthcare and chronic care facilities have been typical places in which patients acquire CDI, in past decade more people bared community-acquired CDI (CHITNIS *et al.*, 2013; OFORI *et al.*, 2018). The infected patients can be asymptomatic carriers (e.g. *C. difficile* is present in 1-5% of health adults) or manifest symptoms, such as mild diarrhea, pseudomembranous colitis, toxic megacolon, and sepsis (GIL *et al.*, 2018). Furthermore, animal with close human contact (e.g. pigs, horses and a wide range of other mammals) can act as a source for the bacterium and viable *C. difficile* spores have been detected in various locations (JANEZIC *et al.*, 2014). *C. difficile* spores are resistant to the bactericidal effects of alcohol and other commonly used hospital disinfectants hence precautions have to be taken to prevent the bacterium spread.

Prevention of CDI should include interventions, like (i) hand washing with soap and water; (ii) the use of gloves and gowns by healthcare personal; (iii) rapidly identification and isolation of CDI patients; (iv) the use of appropriated tests that give accurate results; (v) use of the correct/ approved cleaning material in every surface of the rooms occupied by *C. difficile* patients; and (vi) prescription of the precise antibiotic to each specific case (i.e., antimicrobial stewardship interventions) (see treatment section). The misdiagnosis of CDI leads to an unnecessary use of antibiotics that contributes to an imbalance gut microbiota and can aggravate the MDR problematic, which prompt the necessity of quick and efficient testing methodology and a healthcare system stewardship (FIGUEROA CASTRO; MUNOZ-PRICE, 2019). The author recommend establishing an algorithm for *C. difficile* testing, which aim an improvement of patient care (reduce preventable insolation measure and antibiotic exposure), a reduction of misdiagnosis (i.e. false-positives; distinguish between colonization and true infection), and decrease microbiota disruption, through advances in diagnostic specificity to detect true-positive CDI cases.

The diagnosis of CDI are far from straightforward counting with different methods over the years; however, the application of more than one methodology is preferable (MCDONALD, L CLIFFORD *et al.*, 2018; NELSON; SUDA; EVANS, 2017). A prevalence study used a new stool test, the Nucleic Acid Amplification Test (i.e. detects genetic materials rather than antigens or antibodies allowing an early diagnosis of a disease) to calculate the CDI burden in the US (e.g. diarrhea) (LESSA *et al.*, 2015). Previous testing considered the symptoms (e.g. diarrhea, pseudomembrane colitis or mucosal inflammation), stool analysis that verified the presence of the bacterium and/or the cytotoxin (i.e. cytotoxin assay) (reviewed by Nelson et al., 2017). The recent introduction of a fast and highly sensitive PCR test, via detection of *tcdB*, has led to higher rates of detection; however, the use is still controversial, risking prospective over diagnosis and treatment (BAGDASARIAN; RAO; MALANI, 2015; MCDONALD, L CLIFFORD *et al.*, 2018). The methodology potentially identifies false-positives CDI where the patient is colonized by *C. difficile* yet the symptoms are caused by other factors/microorganism (i.e. asymptomatic carries), which further indicates that the use of multiples methods of diagnosis is still recommended (MCDONALD, L CLIFFORD *et al.*, 2018).

Recent data shows a drop in CDI mortality in the UK through 2013, after a rapid rise in mortality up to 2007 that signposts an effective improving awareness followed by prevention and isolation measures (KARAS; ENOCH; ALIYU, 2010; NELSON; SUDA; EVANS, 2017; REDELINGS; SORVILLO; MASCOLA, 2007). The UK obtained results is in opposition to the prevalence data from the CDC in the US, in which no decline is documented but rather a continuing rising (LESSA *et al.*, 2015; MCDONALD, L CLIFFORD *et al.*, 2018).

The appearance of the hypervirulent strain that secretes a huge amount of toxin A and B lead to a much higher mortality rate than previous reports (i.e. 37%) with a higher incidence of CDI among hospitalized patients, and a greater need for urgent colectomy for toxic colitis (LOUIE, 2005; PÉPIN; VALIQUETTE; COSSETTE, 2005). The variant ribotype 027 spread from Quebec (Canada) to the US and Europe (LOUIE, 2005). In the US, hospitalizations due to CDI doubled between 2000 and 2010 and continue to increase, and, nowadays, CDI is the leading cause of nosocomial diarrhea and gastrointestinal related death (NELSON; SUDA; EVANS, 2017). According to the CDC, *C. difficile* cause proximally half a million infections in the US alone, each year, being responsible for 15,000 to 30,000 deaths (MCDONALD, L CLIFFORD *et al.*, 2018). Furthermore, one every eleven CDI patients over 65 years of

age die in association with the disease (i.e. 80% of *C. difficile* related deaths occur in patients over sixty-five), while one in five CDI patients will have recurrent CDI, and people under antibiotic therapies are seven to ten times more likely to acquire *C. difficile* during or right after treatment (<https://www.cdc.gov/cdiff/pdf/Cdiff-Factsheet-508.pdf>). Bering this high burden, the estimated cost exceeded USD 4.8 billion annually for acute care facilities alone (LESSA *et al.*, 2015; MCDONALD, L CLIFFORD *et al.*, 2018). The emergence of this highly virulent bacterium adds urgency to the identification of effective therapies.

2.5. CDI currently treatments

CDI is the leading cause of hospital-acquired diarrhea and the occurrence has changed in the last decades with an alarming increase of the frequency and severity of the clinical manifestations. The standard treatment for primary infection caused by *C. difficile* is the intake of antibiotics; and the indiscriminate use of antibiotics increased the need for programs that advocate the control of nosocomial outbreaks, like antibiotic stewardship (FIGUEROA CASTRO; MUNOZ-PRICE, 2019; MUTO *et al.*, 2007). These set of coordinated strategies to improve the use of antibiotics would likely improve patient outcomes, decrease unnecessary costs, and reduce development of adverse effects including rise of antimicrobial resistance, which require a worldwide commitment with government actions. A program was proposed by Figueroa Castro and Munoz-Price (2019) where an algorithm was established for *C. difficile* testing to support clinical decisions, a first step to prevent unnecessary antibiotic usage and hence bacterium dissemination. Another major resolution was to search and develop new non-antibiotic drugs besides alternative treatments and prevention methods, which are currently being tested or even applied in some CDI cases (e.g. FMT). Table 4 review the currently and alternative treatments applied to treat CDI.

Current treatment	Use
Metronidazole	Mild-to-moderate first episode
Vancomycin	Severe and severe-to-complicated infection first and second episodes
Fidaxomicin	Severe infection

FMT	Third and more recurrence episodes
Alternative treatments	Use
Rifaximin	After vancomycin
Tigecycline	In combination with metronidazole or vancomycin
Teicoplanin	Severe first episode and recurrence
Doxycycline	Not recommended as a treatment, only protective effect
Linezolid	Reduced toxin production and reversible disruption of the normal intestinal microbiota in mice model
Amoxicile	Lower recurrence rates than vancomycin and fidaxomicin in mice model

TABLE 4: Current therapies used against *Clostridioides difficile*. FMT: Fecal microbiota transplant. Table from (GIL *et al.*, 2018).

2.5.1. Antibiotic therapies

Antibiotic therapies are usually done in conjunction with the cessation of the antibiotic that caused the alteration of the gut flora (AID) that led to the subsequent overgrowth of *C. difficile* (MCDONALD, L CLIFFORD *et al.*, 2018; NELSON; SUDA; EVANS, 2017). The main antibiotics applied to treat CDI are vancomycin, metronidazole and more recently fidaxomicin and rifaximin (Table 4). Vancomycin and metronidazole have been used for the last thirty years to treat CDI; however, they are destructive to the commensal microbiota and render the host highly susceptible to reinfection, and with the rise of vancomycin resistant strain (e.g. Enterococci) the source for new drugs intensified. Fidaxomicin is a new class of macrolide antibiotics that inhibits bacterial RNA polymerase with low resistance potential and does not confer cross-resistance to other antibiotics (COHEN *et al.*, 2010; MCDONALD, L CLIFFORD *et al.*, 2018; SURAWICZ *et al.*, 2013). This antibiotic has reduced antimicrobial activity over other gut microbes and seems to be more specific over *C. difficile* with reduced relapse rates, when compared to vancomycin and metronidazole (BABAKHANI *et al.*, 2012). Rifaximin is a non-absorbable broad-spectrum antibiotic that inhibits RNA synthesis and presents minimal systemic effects being relatively selective for *C. difficile*, which lead to inhibition of commensal bacteria (GIL *et al.*, 2018).

The use of the antibiotics of choice varies in accordance with the severity of the disease. Overall, evidence suggests that vancomycin is more effective than metronidazole for achieving symptomatic cure, but the evidence is less credible for bacteriologic cure; meanwhile, fidaxomicin is superior to vancomycin for achieving symptomatic cure with no direct data on bacteriologic cure (NELSON; SUDA; EVANS, 2017). However, the data lack a control to corroborate the use of one antibiotic over another, particularly considering that the differences in effectiveness between the antibiotics were not too large. Bering this, the authors pointed the advantage of metronidazole far lower cost compared to the vancomycin and fidaxomicin (NELSON; SUDA; EVANS, 2017). Yet, the repeated or prolonged use of metronidazole has been associated with the risk of cumulative and potentially irreversible neurotoxicity and should not be used for long-term therapy (KAPOOR *et al.*, 1999; KURIYAMA *et al.*, 2011).

The guidelines recommendation for treatment of initial CDI is either vancomycin or fidaxomicin but in settings where access to vancomycin or fidaxomicin is limited, the use of metronidazole for an initial episode of non severe CDI is suggested, despite the neurotoxicity hazard (MCDONALD, L CLIFFORD *et al.*, 2018).

Considering severe or fulminant CDI, the response to vancomycin is significantly better than to metronidazole, which are done orally or per rectum. The administration per rectum is recommended in the presence of severe ileus/megacolon and the dosage should include intravenously administered metronidazole. In extreme cases, surgical management is necessary (i.e. subtotal colectomy with preservation of the rectum) and the patient should be in a critical patient unit, with adequate volume and replacement of electrolytes (GIL *et al.*, 2018; MCDONALD, L CLIFFORD *et al.*, 2018). The surgical management of severe CDI have a better survival rate than the standard use of antibiotic therapy; however, the described postoperative mortality is high (i.e. 35 to 80%) (GIL *et al.*, 2018).

Recurrent CDI treatment varies in accordance with the number of times *C. difficile* recolonized the patient. The first recurrence should be treated with oral vancomycin as a tapered and pulsed regimen rather than a second standard 10-day course or with a standard fidaxomicin treatment. Antibiotic treatment options for patients with more than one recurrence include oral vancomycin therapy, with dosage like the first recurrence, or standard course of vancomycin followed by rifaximin or fidaxomicin (MCDONALD, L CLIFFORD *et al.*, 2018). Furthermore, FMT is strongly endorsed in cases of multiple recurrences of CDI where the recommended antibiotic treatments failed.

2.5.2. Fecal transplantation and other microbiota related therapies

Fecal transplantation consists of a bacterial mixture obtained from a healthy donor that is given to CDI recurrent patients through enema, orogastric tube or by capsule containing freeze-dried material. Reintroduction of normal bacteria by donating stool reinstates the dysbiotic gut environment, reintroducing the phylogenetic richness and resistance to colonization (i.e. commensal bacteria). FMT has been a highly effective therapy for recurrent CDI (i.e. successful rate between 80% and 90% cure in randomized trials), still the occasional drawbacks restrict its application, which includes the lack of palatability, the theoretical risk of infection transmission and regulatory complexity (MCDONALD, L CLIFFORD *et al.*, 2018). Recently, the importance of BSH during FMT was recognized and a targeted restoration via BSH supplementation was proposed and represents a novel therapeutic approach to treat CDI, concurrently avoiding the limitations associated with FMT (MULLISH *et al.*, 2019).

The administration of a non-toxigenic *C. difficile* strain M3 following antibiotic treatment significantly decreased the recurrence of disease and diminished adverse events, being able to compete and replace toxigenic strains *in vivo* (hamster and human) (GERDING; SAMBOL; JOHNSON, 2018). Likewise, a Phase II clinical trial reported that administration of a consortium of spore-forming commensal species significantly decreased the recurrence of disease (KHANNA *et al.*, 2016). The mechanism of action of this therapy remains undefined but is probably related with competition for niche (e.g. nutrient).

2.5.3. Immunization strategies

2.5.3.1. *C. difficile* and the immune system

Loss of host epithelial integrity via the action of *C. difficile* virulent factors results in increased intestinal permeability and the translocation of microorganism (e.g. pathobionts and pathogenic) from the gut lumen into deeper tissues (Figure 21). The PRR, NOD1 and TLR4, recognized the bacterium prompting an early immune defense. Furthermore, the virulent factors (e.g. TcdA or TcdB) and the epithelial disturbance trigger the release of pro-inflammatory cytokines and chemokines that recruit circulating innate and adaptive immune cells (ABT; MCKENNEY; PAMER, 2016). Briefly, ILCs respond to pro-inflammatory cytokines upregulated during the acute infection (e.g. IL-1 β , IL-12 and IL-23) by producing effector cytokines, such

as IL-22, IL-17a, IFN- γ and TNF. These productions activates recruited neutrophils and macrophages and drive the expression of AMP, the production of RNS and ROS, the complement pathway, and repair mechanisms in epithelial cells (reviewed by Abt et al., 2016).

T cells and B cells do not fully contribute to the resolution of the acute phase of CDI yet clinical data indicates that adaptive immune responses can have protective effects (ABT; MCKENNEY; PAMER, 2016). The occurrence of toxin-specific IgA and IgG antibodies reduce the severity of the disease and patients who received humanized monoclonal antibodies specific for TcdA and TcdB had reduced recurrences of disease, but to date no formal recommendation were made (JOYCE *et al.*, 2014; LOWY *et al.*, 2010). The role of T cells is less understood. MHC II-deficient mice exhibit unaltered acute infection and mice lacking B and T cells recover from acute infection similarly to wild-type mice, which indicates that T cells response do not affect the outcome of the immune system response against *C. difficile* (ABT; MCKENNEY; PAMER, 2016; HASEGAWA *et al.*, 2014).

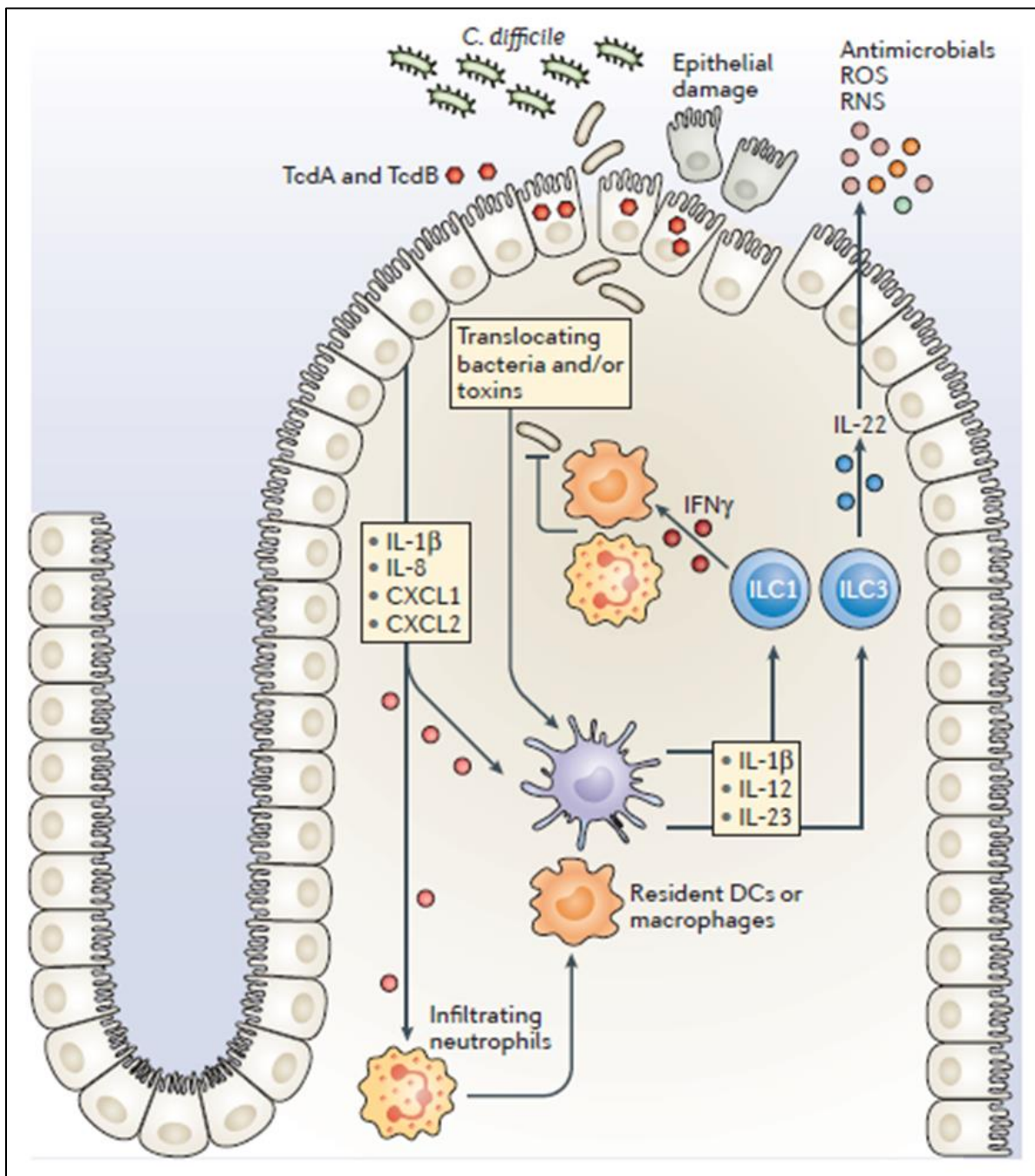


FIGURE 21: Innate immune response against *C. difficile*. The host response to *C. difficile* is initiated by toxin-mediated damage that leads to loss of epithelial integrity and the translocation of pathobionts and pathogenic bacteria. Intestinal epithelial cells and resident innate immune cells secrete pro-inflammatory chemokines and cytokines (e.g. IL-1 β , IL-8, IL-12 and IL-23), which leads to the recruitment of neutrophils and the activation of innate lymphoid cells (ILCs), production of antimicrobial peptides, reactive oxygen species (ROS) and reactive nitrogen species (RNS). IL-12 signaling drives the expression of interferon- γ (IFN- γ), whereas IL-1 β and IL-23 signaling induces the production of IL-22. These defense mechanisms limit bacterial dissemination, attenuate toxin activity and repair epithelial damage. DCs, dendritic cells. Figure from (ABT; MCKENNEY; PAMER, 2016).

2.5.3.2. Vaccine development

Active immunization has proved efficient to decrease *C. difficile* colonization and toxin inflammation, and, currently, several clinical trials are quite advanced (GIL *et al.*, 2018; SURAWICZ *et al.*, 2013). Bering this, immunization can be used as a complementary treatment especially in patients with severe CDI or those that had complications with standard therapies (ABOUGERGI; KWON, 2011).

A vaccine containing TcdA and TcdB (ACAM-CDIFF) has been tested in healthy volunteers, achieving high levels of IgG against ToxA (DE BRUYN *et al.*, 2016). This toxoid vaccine is currently ongoing Phase III clinical trial, previously approved from Phase II without any safety issues. Likewise, the toxoid vaccine PF-06425090 had worthy results in Phase I and II (SHELDON *et al.*, 2016), and the vaccine VLA84 showed positive results considering tolerability, immunogenicity and safety in Phase II trials (BÉZAY *et al.*, 2016).

2.5.4. Summary of therapies under development

Antibiotics have significant and long-lasting impacts on the intestinal microbiota, consequently reducing colonization resistance against *C. difficile*. Standard antibiotics therapy is associated with a high rate of recurrence, highlighting the need for novel strategies that eliminate the bacterium and target the major virulent factors (e.g. toxins and adhesins). Considering that current therapies are not able to completely eradicate the infection, new therapies against CDI are evidently needed with some under clinical trials, as the identification of new drug and vaccine targets (section 2.6). Table 5 summarize the main *C. difficile* treatment under developed with their current clinical phase.

Under development Treatments	Characteristics	Development phase
RBX2660	Bacterial mixture	Phase II completed
SER-109	Feces-derived Firmicutes spores	Phase II ongoing
SER-262	Mixture of spores from 12 different bacterial species	Phase I ongoing

MET-1	Bacterial mixture obtained from the intestine of a healthy donor	Phase II ongoing
RBx 11760	Inhibits sporulation and toxin production	<i>In vivo</i> preclinical
RBx 14255	Higher protective efficacy in a CDI hamster model compared with metronidazole and vancomycin	<i>In vivo</i> preclinical
Omadacycline	Higher activity against CD than vancomycin in a CDI hamster model	<i>In vivo</i> preclinical
CRS3123	Higher activity than vancomycin, little effect over normal microbiota, inhibition of toxin production and sporulation	Phase I completed
Ramoplanin	More effective than vancomycin, with little impact on anaerobic bacteria	Phase III completed
Oritavancin	Contributes to delay recurrence	<i>In vivo</i> preclinical
Thuricin CD	Very low activity against the rest of the normal intestinal microbiota	<i>In vitro</i> preclinical
Surotomycin	More effective than vancomycin or metronidazole at killing vegetative cells in established biofilms	Phase III completed
Nylon-3 polymers	Inhibit <i>C. difficile</i> growth and blocked spore outgrowth	<i>In vitro</i> preclinical
Cadazolid	Low recurrence and intestinal dysbiosis	Phase III ongoing
Lactoferrin	Inhibit toxin production without significantly affecting the normal microbiota	Phase II ongoing
Ridinilazole	Low effect over other Gram-positive or Gram-negative fecal microbiota. Decrease toxin production	Phase II completed
Tolvamer	Binds non-covalently to Toxins A and B. Lower recurrence rates than vancomycin or metronidazole	Phase III completed
CASAD	Binds to toxins	Phase II ongoing
Ebselen	Binds to toxins B	<i>In vivo</i> preclinical
Intravenous immunoglobulin	Used in severe and recurrent cases of CDI. No clinical trials evaluating its efficacy	<i>In vivo</i> preclinical
Actoxumab and bezlotoxumab	Antibodies against toxin A and toxin B	Phase II completed
PolyCAB	Mixture of anti-A and anti-B antibodies. Significant reduction of CDI recurrence in a hamster model	Phase I ongoing
CANmAbA4 and CANmAbB4	Humanized anti-A and anti-B antibodies. Toxin-neutralizing capacity and protection against clinical CDI	<i>In vivo</i> preclinical
VNA2-Tcd	Toxin neutralization activity	<i>In vivo</i> preclinical
ACAM-CDIFF	Toxoid vaccine used as an approach for prevention and treatment of CDI	Phase III ongoing
PF-06425090	Toxoid vaccine	Phase II ongoing
VLA84	Recombinant protein vaccine that contains epitopes of toxin A and toxin B	Phase II completed

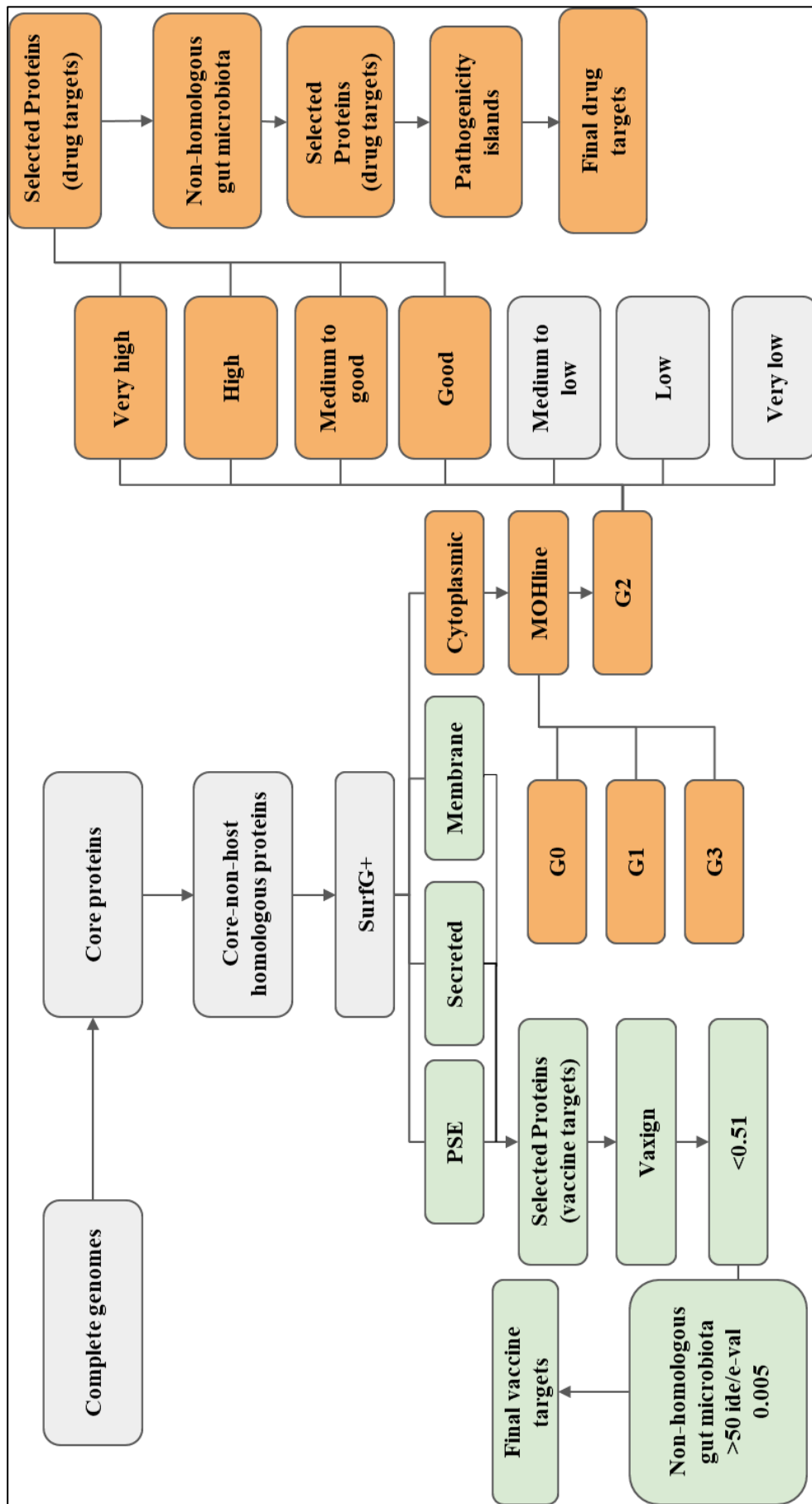
TABLE 5: Under development treatments against *Clostridioides difficile*. ABA: Antibody to both TcdA and TcdB toxins; ACAM-CDIFF: *Clostridioides difficile* toxoid vaccine; CASAD: Calcium aluminosilicate; CD: *Clostridioides difficile*; CDI: *Clostridioides difficile* infection; MET-1: Microbial ecosystem therapeutic-1;

PolyCAB: Preparation of polyclonal antibodies against TcdA and TcdB toxins; VNA: Single heteromultimeric VHHbased neutralizing agent that targets TcdA and TcdB toxins. Table from (GIL et al., 2018).

3. Submitted paper: A Reverse Vaccinology and Immunoinformatics Based Approach for Multi-epitope vaccine against *Clostridioides difficile* infection

This chapter focus on the search for alternative treatment against *Clostridioides difficile*, an opportunistic pathogen that cause *C. difficile* infections, which includes a number of nosocomial symptoms (e.g. diarrhea, pseudomembrane colitis, and toxic megacolon). We proposed the construction of a peptide based vaccine against toxigenic and non-toxigenic *C. difficile* via an immunoinformatics strategy. Thus, we selected vaccine and drug targets candidates via reverse vaccinology and substantive genomics methodology, respectively (Figure 22). Afterwards, we design a multi-epitope vaccine trough using immunoinformatics tolls and follow the methodology established during the construction of the *C. diphtheriae* vaccine (Figure 16; Chapter 1). The manuscript was submitted to the ‘Journal of the Royal Society Interface’ under the identification ‘rsif-2019-0664’.

FIGURE 22: Workflow of the *C. difficile* vaccine and drug targets selection trough reverse vaccinology and subtractive genomics, respectively. The non-host homologous core proteins of *C. difficile*, from 53 complete genomes, were sorted according with the cellular location and those classified as membrane, secreted and putative surface-exposed (PSE) (light green) were used for vaccine target selection trough reverse vaccinology, and meanwhile, those proteins classified as cytoplasmic (orange) were designated for the drug target methodology (subtractive genomics).



A Reverse Vaccinology and Immunoinformatics Based Approach for Multi-epitope vaccine against *Clostridioides difficile* infection

MARIANA SANTANA¹; ARUN KUMAR JAISWAL^{2,3}; STEPHANE FRAGA DE OLIVEIRA TOSTA²; RODRIGO KATO²; SIOMAR C. SOARES^{2,3}; LUIZ CARLOS JUNIOR ALCANTARA⁴; ANDERSON MIYOSHI¹; VASCO AZEVEDO²; SANDEEP TIWARI^{2*}

¹Department of Genetics, Institute of Biological Sciences, Federal University of Minas Gerais (UFMG), Belo Horizonte 31270-901, MG, Brazil

²Postgraduate Program in Bioinformatics, Institute of Biological Sciences, Federal University of Minas Gerais (UFMG), Belo Horizonte 31270-901, MG, Brazil

³Department of Immunology, Microbiology and Parasitology, Institute of Biological and Natural Sciences, Federal University of Triângulo Mineiro (UFTM), Uberaba 38025-180, MG, Brazil

⁴Laboratório de Flavivírus, Instituto Oswaldo Cruz, FIOCRUZ, Manguinhos, Rio de Janeiro 21040-900 Rio de Janeiro, Brazil.

Corresponding author: Sandeep Tiwari- sandip_sbtbi@yahoo.com

Abstract

Clostridioides difficile is a potentially life threatening bacillus that causes hospital and community-acquired gastrointestinal illness. *C. difficile* infection (CDI) clinical symptoms range from diarrhea and vomiting to toxic megacolon and pseudomembrane colitis. The clostridial toxin A (TcdA or ToxA) and the clostridial toxin B (TcdB or ToxB) are *C. difficile* primary mediators of inflammation that triggers host cellular response. Over the past few years, *C. difficile* has become one of the major causes of nosocomial infections and the continual/misuse use of antibiotics, especially during hospitalization, increases the risk of CDI. The antibiotic resistance problematic and the growing number of CDI outbreaks and asymptomatic patients expose the need for a valid and acceptable alternative treatment for *C. difficile*. Considering these, we used 53 complete genomes of *C. difficile* for comparison taking *C. difficile* ATCC 9689/DSM 1296 as the reference genome to predict putative vaccine and drugs targets against *C. difficile* using reverse vaccinology and subtractive genomics. After the detailed *in silico* analysis, four non-host homologous protein drug targets identified. The identified vaccine targets were analysed for multi-epitope based vaccine design against *C. difficile*. This study will facilitate future *in vitro* and *in vivo* tests for the production of a peptide vaccine, drugs and prophylactic targets against a bacterium with high clinical relevance.

Keywords: *Clostridioides difficile*, CDI, reverse vaccinology, immunoinformatics, nosocomial infections

1. Introduction

The Gram-positive anaerobic bacterium *Clostridioides difficile* (formally *Clostridium difficile*) is a potentially life threatening bacillus that causes hospital acquired gastrointestinal illness (1–4). Initially, *C. difficile* was identified as a commensal bacterium of health children, but studies revealed the association of this bacterium with human disease (1,2). *C. difficile* infection (CDI) clinical symptoms range from diarrhea and vomiting to pseudomembrane colitis and toxic megacolon (5,6). The vegetative cell form of this bacterium colonize and overgrowth in the colon, being able to persist due to the action of many virulence factors (3,7). The clostridial toxin A (TcdA or ToxA) and the clostridial toxin B (TcdB or ToxB) are *C. difficile* primary mediators of inflammation that triggers host immune response (8). These toxins can bind and cross the colonic epithelium causing infiltration of immune cells leading to tissue necrosis and cell death (9). Nonetheless, other virulent factors, such as the cytolethal distending toxin (CDT) [i.e. binary toxin (CtdA and CtdB) involved in epithelial adhesion and pathogenesis], sporulation factors, flagella, and adhesins, were associated with more severe forms of the disease, which is linked with some hipper virulent *C. difficile* strains (e.g. ribotype 027) (10–13).

Over the past few years, *C. difficile* has become one of the major causes of nosocomial infections and the continual/misuse use of antibiotics, especially during hospitalization, increases the risk of CDI (3,14,15). The Center of Disease Control (DCD) notified proximally 500,000 cases of CDI and 30,000 deaths due to the disease, in the United States (US) alone, numbers probably underestimated due to the accuracy of CDI diagnosis and unreported cases, like asymptomatic patients (5,16–18). Asymptomatic carriers presents *C. difficile* however do not have CDI symptoms; yet, are still able to transmit the bacterium representing a treat to combat the disease (5).

C. difficile transmission occurs through oral-fecal rout spores (3). Unlike the vegetative cells, *C. difficile* spores are highly resistant dormant cells capable to survive in hostile conditions (e.g. acidic, heat, drying, oxygenated, and many disinfectants and antimicrobials, hospital surfaces) for months or year (19,20). The spores germinate in a suitable environment (i.e. duodenum and colon), due to the action of the primary bile acid taurocholate and/or the amino acid glycine, deriving the metabolically active form (i.e. vegetative cells); and it is inhibited by secondary bile acid deoxycholate (21–24).

The bile acids composition is directed linked with the microbiota, since microorganisms are necessary for the transformation of the primary conjugated bile acids into free primary bile acids that are then converted to secondary bile acids (25,26). Patients with CDI present a shift in the microbiota composition (e.g.

increase in the Proteobacteria phylum and a decrease in Bacteroidetes phylum levels and in the Clostridia class), which lead to a lower microbial diversity compared to healthy controls (27–33). Members of the Clostridia class, like the *Clostridium* clusters IV and XIVa (e.g. *C. scindens*), are involved in the formation of secondary bile acids, which prevents *C. difficile* overgrowth (28,30). Evidence from *in vitro* and *in vivo* studies indicates that the use of certain bacterial taxa (e.g. Lachnospiraceae and Ruminococcaceae) may prevent the progression of the disease/reduction of *C. difficile* colonization by inhibiting the germination of spores and vegetative growth (27,32,34,35). This was also observed after fecal microbiota transplantation (FMT) in patients with recurrent CDI, which indicates that the use of those taxa are a valid alternative treatment (36–38). However, the use of the recommended antibiotic (i.e. vancomycin, fidaxomicin or metronidazole) is still the endorsed treatment, since there is insufficient data to support alternative treatments, like FMT, probiotics and vaccines (14,15). Vaccination against *C. difficile* is a feasible next stage practice, and currently, is being tested as a complementary treatment with a proven efficiency in several clinical trials(14,39–43).

Additional tactics have to be employed to prevent the spread of *C. difficile* due to the infectious status of this bacterium, especially during hospitalization. Methods of infection control/ environmental management (e.g. isolation of the CDI patients in private room and toilet, the use gloves and gowns, hand hygiene before and after contact, use disposable equipment whenever possible) and the discontinuation of former antibiotic therapies are applied (14,15). Antibiotic treatment is a known cause of intestinal dysbiosis, a condition that favors the colonization and overgrowth of *C. difficile* (24,29,30,33). This bacterium is highly resistant to a number of antibiotics in the spore and vegetative forms. The mobile genome of *C. difficile* (i.e. transposons, insertion sequences and prophages) contains many resistant genes that were probably acquired by horizontal gene transfer (3,44,45). Likewise, the ‘mobilome’ is involved in the gain of other advantageous genes related with the successful infection capability/rate of this opportunistic pathogen, and also, significantly contributes to the high genetic diversity of the species (i.e. pan genome of ~9,640 CDSs wherein 600 to 3,000 belongs to the core genome) (46–49).

The ability to gain resistance markers is a major health concern and one of the reasons that the CDC classified *C. difficile* as an Urgent Antibiotic Resistance Threat (50). The antibiotic resistance problematic and the growing number of CDI outbreaks and asymptomatic patients expose the need for a valid and acceptable alternative treatment for *C. difficile*. Therefore, in this study, the prediction of broad-spectrum

potential drug and vaccine targets against *C. difficile* through subtractive genomics and reverse vaccinology-based approach were carried out. We used 53 complete genomes of *C. difficile* for comparison taking *C. difficile* ATCC 9689/DSM 1296 as the reference genome (GenBank accession GCA_001077535.2). Thus, we propose a multi-epitope vaccine against *C. difficile* that target known virulent proteins (i.e. TcdA, TcdB, CdtB, Slpa) from the bacterium accessory genome (i.e. not common among *C. difficile* toxigenic and non-toxigenic strains) and relevant proteins from the core genome (i.e. commonly shared by all strains) selected by the above mentioned approaches (i.e. AKP42475, AKP43769, AKP41996, AKP43195, AKP41595). This vaccine entails not only the two main virulent factors but also secreted and transmembrane proteins from the core genome and thus may be effective against non-toxigenic and hypervirulent strains, covering virulent and colonization factors that broaden *C. difficile* elimination. Furthermore, we identified four potential *C. difficile* drug targets (e.g. AKP41517, AKP41398, AKP41559 and AKP44557) encoded in genomic or pathogenic island. Considering all, this study will facilitate future *in vitro* and *in vivo* tests for the production of drugs and prophylactic targets against a species of bacteria with high clinical relevance.

2. Methodology

2.1. Selection of data

The complete genome sequences of all 53 strains of *C. difficile* were retrieved from the NCBI (National Center for Biotechnology Information) database. The complete amino acid sequence of the proteins were recovered from the Universal Protein Resource (<https://www.uniprot.org/>) (51) wherever needed.

2.2. Identification of Intra-Species Conserved Non-Host Homologous Proteins and Subcellular localization

The orthologous genes were gathered to obtain a framework to integrate information from multiple genomes, highlighting the conservation and divergence of gene families and biological processes. Therefore, for the identification of the core genome, genome sequences of 53 strains of *C. difficile* strains

were compared taking *C. difficile* ATCC 9689/DSM 1296 as the reference genome (GenBank accession GCA_001077535.2), using OrthoFinder software (52). The core genes/proteins were retrieved for further analysis. These core genes are exclusive of the pathogen (i.e. not existent in the human host, in order to avoid immune reactions against similar host targets, providing specific immune response guided exclusively to pathogen proteins). The identification of non-host homologous proteins, OrthoFinder software were used to compare the core genome with the human genome (52). Furthermore, the non-host homologous conserved proteome of *C. difficile* were screened using SurfG+ software (53) for the subcellular localization. The software predicts the secreted proteins, membrane proteins, putative surface exposed proteins and cytoplasmic proteins.

2.3. Vaccine targets for immunoinformatics analysis

The dataset (i.e. secreted proteins, membrane proteins, and putative surface exposed proteins) were screened by Vaxign (He et al., 2010) for proteins with the binding properties for the major histocompatibility complex (MHC-I) and (MHC-II) using an adhesion probability greater than 0.5. Furthermore, the online webserver PBIT (Pipeline Builder for Identification of drug Targets) was used for the non-homology to human gut microbiota analysis (54). The default parameters were applied for the analysis.

2.4. Immunoinformatics analysis for the Multi-epitope vaccine

2.4.1. Epitopes prediction

2.4.1.1. Prediction of MHC-I binding epitopes

The immune epitope database (IEDB) (<https://www.iedb.org/mhci/>) was used for the MHC-I epitope prediction. IEDB search epitopes based on experimental assays characteristics (epitope source, structure process environmental) and epitope information (allergens, infectious and autoimmune diseases, transplantation and cancer) (55). This server predicts and analyses epitopes from T cell and B cells based on literature data. For class I binding predictions, the IEDB recommended methods were applied, artificial

neural network (ANN), stabilized matrix method (SMM), and scoring matrices derived from combinatorial peptide libraries (Comblib).

The IEDB search for MHC class-I employed 9-mer size epitopes that bind with high affinities to mouse MHC-I alleles (H-2Db, H-2Kb, H-2Kd, H-2Kk, H-2Ld). The cutoff applied was based on IEDB default parameters where the epitopes with a percentile rank below 3% were selected for further analysis.

2.4.1.2. Prediction of MHC-II binding epitopes

MHC-II epitopes identification was carried out with IEBD server (<https://www.iedb.org/mhcii/>) using an approach similar to MHC-I epitope prediction. The analysis was performed using the IEDB MHC-II epitope predictions default parameters. The mouse MHC-II alleles (H2-IAb, H2-IAAd, H2-IEd) were selected for the peptide binding with an epitope length of 15-mer and a cutoff below 3% (percentile rank).

2.4.1.3. Prediction of B cell binding epitopes

B cell binding epitopes identification was carried out with ABCPred server (<http://crdd.osdd.net/raghava/abcpred/index.html>) (56). ABCpred predict B cell epitopes using artificial neural network using BCIPEP B cell epitope database. The dataset consists of 2479 continuous epitopes, including 654 immunodominant, 1617 immunogenic epitopes obtained from 700 unique experimentally proved continuous B cell epitopes from virus, bacteria, protozoa and fungi. ABCpred default parameters were used to predict epitopes with a threshold value above 0.51 and a peptide length of 16-mer for prediction.

2.4.2. T and B cell epitopes comparison and selection

The MHC-I and MHC-II epitopes predicted for all ten proteins were compared with the B cell epitopes. The MHC-I/B cell epitopes comparison selected the MHC-I epitopes that presented 9 common sequential

amino acids among the B-cells whereas the MHC-II/B cell epitopes comparison picked the MHC-II epitopes that had 15 to 12 common sequential amino acids. Afterwards, the MHC-I and MHC-II epitopes selected after B cell comparison were further correlated (overlapped of 9 and 8 common sequential amino acids) to single out the common epitopes of each protein for the final vaccine construction.

2.4.3. Vaccine construction

Appropriated linkers joined the epitopes selected from the nine *C. difficile* proteins (AKP42475, AKP43769, AKP41996, AKP43195, AKP41595, AGG91516, AGG91581, TGA38378, and AKP43753). AAY were added between the MHC-I epitopes and GPGPG between MHC-II epitopes. The order of the proteins and epitopes were chosen randomly and the same protein order were applied for both MHC types, wherein the class I epitopes were place first followed by the classes II epitopes. No peptide adjuvant was added to the vaccine sequence.

2.5. Homology modeling of multi-epitope *C. difficile* vaccine

The I-TASSER (Iterative threading assembly refinement) server (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) was used for the prediction of the protein three-dimensional (3D) structure. I-TASSER homology modeling is based on multiples threading alignments. First, the software match the sequence with known proteins structural models from the Protein database (PDB) (<https://www.wwpdb.org/>) to identify structural templates. Then, fragment assembly simulations via *ab initio* folding construct several models and the structures with the highest 3D models score and with the lowest free energy state are selected. The hydrogen-bonding network of the models are then optimized and the final models are achieved after refinement of the global topology and removal of steric clashes (57,58). The confidence score (C-score) is used to calculate the quality of the models and is based on the significance of the alignments (structure assembly simulations) and the quality of the structure; therefore, the highest scores correspond to the highest model accuracy.

2.6. Refinement and validation of the tertiary structure

The GalaxyRefine server (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) was implemented to refine the selected 3D structure model predicted by I-TASSER. This method can improve global and local structure accuracy by first rebuilding all side-chain conformations and then refining the models using two repetitive relaxations methods, the mild relaxation and the aggressive relaxation, to eliminate putative perturbations (59). Structure perturbations in the side chains are corrected by the mild refinement, meanwhile secondary structure elements and loops are adjusted by the aggressive refinement. Model 1 correspond to the lowest energy model generated by the mild relaxation, and models 2 to 5 correspond to the largest clusters generated by aggressive relaxation.

Subsequently, the PROCHECK program of SAVES 5.0 server (<http://servicesn.mbi.ucla.edu/SAVES/>) was used to check and validate the stereochemical quality of 3D protein structure through the Ramachandran plot. The Ramachandran plot report the energetically allowed and disallowed regions in the protein structure, which grant the validation and selection of the best 3D model in accordance with the empirical distribution of the amino acids. The validation is based on the dihedral angles, which are used to specify the molecular conformation. The torsion angles phi-psi determine the stability of the amino acid residues in the protein structure (60,61).

2.7. B cell epitope prediction

The immunoinformatics analysis of B cell prediction were performed using the multi-epitope vaccine FASTA format or the 3D selected model (pdb file) obtained from the GalaxyRefine server.

The identification of linear B cell epitopes was performed by ABCpred and discontinuous or conformational B cell epitopes by ElliPro (<http://tools.iedb.org/ellipro/>). ABCpred default parameters were used to predict epitopes with a threshold value of 0.51. ElliPro predicts discontinuous epitopes based on a protein antigen's 3D structure, which is associated with the Protrusion Index (PI) and the distance R (62). PI is based on the location of the residue, within or outside the ellipsoid, where high percentages of amino acids residing in the ellipsoid are associated with greater solvent accessibility. The epitopes are clustered based on the distance R that correspond to the distance between residue's centers of mass and

the value of R is directly associated with the score of the discontinuous epitopes being predicted. The ElliPro default parameters were used to predict epitopes with a PI value over 70%.

2.8. Interferon-gamma inducing prediction

INF- γ is a cytokine release from CD4⁺ T cells critical for the control of pathogens; therefore, the ability to induce this cytokine was measured. The IFNepitope server (<http://crdd.osdd.net/raghava/ifnepitope/predict.php>) predicts INF- γ induction from MHC-II based on IEDB database and an INF- γ dataset (inducing and non-induced epitopes). The dataset uses the Motive based model (MERCII) and Support Vector Machine based models (SVM) to discriminate INF- γ inducing and non-induced peptides through data analysis and pattern recognition (63). The MHC-II epitopes predicted by IEDB were used to predict the protein capability to induce INF- γ . The default parameters were applied for the analysis.

2.9. Vaccine properties evaluation

2.9.1. Allergenicity evaluation

AllerTOP v2.0 (<http://www.ddg-pharmfac.net/AllerTOP/>) and AlgPred (<http://crdd.osdd.net/raghava/algpred/>) servers were used to analyze the allergenicity of the vaccine construction. A BLASTp comparison against the reference human proteome (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=OGP__9606__9558&LINK_LOC=blasthome) were performed to further verify the degree of similarity. The default parameters of the servers were used for the analysis.

AllerTop classifies a protein sequence as allergenic or non-allergenic based on a dataset of 2427 known allergens from different species and 2427 non-allergens. The server employs the auto cross covariance (ACC), a protein sequence mining method that uses the amino acids properties (e.g. hydrophobicity, molecular size, helix-forming propensity) to discriminate the peptides (64).

AlgPred analysis were based on five approaches (i.e. IgE epitope and PID, MEME/MAST motif, SVM method based on amino acid composition. SVM method based on dipeptide composition, Blast search on allergen representative peptides, hybrid) to predict the allergenicity of the peptide sequence (65). IgE epitope search rely on the query of known IgE epitopes. MEME/MAST motif uses MEME matrices. SVM method based on amino acid composition or dipeptide composition, and blast search on allergen representative peptides (ARPs) uses a database of know protein sequence of allergens and non-allergens to query a search. The hybrid approach (SVMc + IgE epitope + ARPs BLAST + MAST) assign a protein sequence as allergen if any one of this mentioned method predicts it as allergen.

2.9.2. Antigenicity evaluation

The antigenicity evaluation of the final multi-epitope vaccine sequence was performed using VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) and ANTIGENpro (<http://scratch.proteomics.ics.uci.edu/>). The default parameters of the servers were employed for the analysis.

VaxiJen is a server for alignment-independent prediction of protective antigens based on the physicochemical properties of proteins via an ACC alignment-free approach (66). The server has a dataset of 100 known antigens and 100 non-antigens from different sources; like bacterial, viral and tumors protein. ANTIGENpro is a method applied from the server Scratch. This method implement data obtained from protein microarray analysis and predicts protein antigenicity via a sequence-based, alignment-free and pathogen-independent predictor (67).

2.9.3. Physicochemical parameters

ProtParam server (<https://web.expasy.org/protparam/>) infer the physicochemical properties from the protein sequence and the analysis are based on either compositional data or on the N-terminal amino acid. The parameters include the molecular weight, theoretical isoelectric point (pI), amino acid composition,

atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) (68). The default parameters of the servers were used for the analysis.

2.10. Protein–protein docking between vaccine candidate and TLR4

2.10.1. Definition of the TLR4 local highly frustration region

Proteins are composed of more rigid parts, which are connected by largely minimally frustrated interactions, and low local stability clusters (high local frustration). The polymers minimally frustrated interactions include the folding core of the molecule whereas the highly frustrated areas are normally associated with functional regions (e.g. allosteric transitions and binding sites for protein–protein assembly and recognition). Thus, the Frustratometer server (<http://frustratometer.qb.fcen.uba.ar/>) were applied to identify the highly frustrated regions of the toll-like receptor 4 (TLR4). This online server uses the energy landscape theory to quantify the degree of local frustration manifested in protein molecules.

The TLR4 receptor obtained from the RCSB PDB database (2Z63) was first edited (i.e. removal of water molecules and the ligands FUL and NAG) by Chimera (<https://www.cgl.ucsf.edu/chimera/>) to guaranty the correct local frustration site and binding to the putative *C. difficile* vaccine without foreign interference (69). Furthermore, Chimera was used for the identification and visualization of TLR4 hydrophobic regions.

2.10.2. Definition of the molecular docking structure

The molecular docking was performed through the SwarmDock server (<https://bmm.crick.ac.uk/~svc-bmm-swarmdock/submit.cgi>) using the *C. difficile* multi-epitope vaccine as the ligand and the edited TLR4 (PDB 2Z63) as the receptor. SwarmDock is a flexible protein-protein docking algorithm that predicts the structure via a combination of local docking and particle swarm optimization (PSO) (70). PSO optimize the binding energy between the binding partners and the lowest energy member among them.

The workflow consists of three steps: Preprocessing, docking, and post processing. The first step analysis checks and correct the structures partners (e.g. alternative atoms locations, gaps and nonstandard residues). The docking step is performed four times for each start position providing a variety of conformational poses. The docking mode can be restrained (selected binding region) or blind (no specific binding region selected). In the last case the algorithm runs proximally 120 start positions; meanwhile, in a restrict docking the server runs the start position based at least on one of the selected residues (exactly or near the binding site). During the post processing step, the binding structures are minimized, scored and clustered together in the case of high degree of conformational similarities (70,71). Afterwards, the software generates a list of docking poses scored by the energy/mean energy, number of cluster members, and number of contacts. Two different results are generated: Standard and democratic. The democratic models are based in the Integrative Ranking of Protein-Protein Assemblies (IRaPPA) scoring scheme. The IRaPPA approach provides a better quality and ranking of docking models via the enhancing of the atomic modelling of protein complexes. It combines physicochemical descriptors like Ranking support vector machines (R-SVMs) and Schulze voting method to select the ranking docked poses (72)

The molecular docking was performed in the restricted mode consistent with default parameters. The binding region established in the previous step (highly frustrated region) was set for the receptor (amino acids 1-42, 129-165, and 520-560) and the whole vaccine sequence (1-765 amino acid) was define as the binding region of the ligand. The top 10 best democratic docking results were visualized and the intermolecular hydrogen bonds (H-bonds) identified by Chimera. The result that present the lowest energy score, the maximum number of contacts made between the residues, and the highest number of hydrogen bounds (H-bonds) between the vaccine residues and the defined binding region of TLR4 was selected.

2.11. Molecular dynamics simulation

We applied a simulation of Molecular dynamics (MD) to understand the structure of biological macromolecules and their atomic interaction (70,73). Gromacs v5.1 2 (74) was used to study the structure properties and interaction between the predicted vaccine protein structure (ligand) and TLR-4 (receptor) at microscopic level. The parameters set for the simulation were in accordance with (75). In the first step, pdb2gmx was used in order to construct protein topology and provide information about bonded and non-

bonded characteristics. The structure was solvated in a cubic box of TIP3P water molecules. The minimal distance of 14 Å from any edge of the box along with the periodic boundary conditions were applied in all dimensions and the concentration of 150 mM of NaCl, which corresponds to physiological conditions, were added to neutralize the simulation system. The Amber99sb-ildn force field were chosen to determine the intermolecular and intramolecular interactions (75) and the Particle mesh Ewald (PME) method to assess the long range electrostatic interactions and LINCS algorithm to constrain all bonds (76). The non-bonded cutoff were set in 10 Å. Subsequently, the system was minimized using the steepest descent algorithm and the canonical ensemble (NVT) was applied in MD simulation at 300 K during 400 ps. The equilibration phase continued employing the isobaric–isothermic ensemble (NPT) run at 300 K and 1 bar pressure. In the final step, the MD production run was performed for 30 ns. The used integration time step was 2 fs in the MD simulation. For analysis phase, the output files were saved every 2 ps.

2.12. *In silico* cloning

JCAT server (<http://www.jcat.de/Start.jsp>) was employed to evaluate the cloning and expression of the *C. difficile* multi-epitope vaccine within the expression vector by adjusting the codon usage of the protein to the typical codon usage of the host. Through reverse translation, the server generates cDNA sequence and afterwards optimizes the codon in accordance with the most abundant tRNA species of the selected host (codon usage) (77). JCAT back translate to DNA favoring the most frequently codon consistent with the highest relative adaptiveness for the amino acid of a given organism from a reference set. The vaccine protein sequence was adapted to *E. coli* (strain ATCC 27325 / DSM 5911 / W3110 / K12). Codon optimization evaluates the sequence in accordance with the codon adaptive index (CAI). CAI evaluate the similarities between a given protein coding gene sequence with respect to a reference set of genes (78,79). The server quantitative index optimize the gene sequences consistent with their expression levels, in a range from zero to one, wherein one indicates that both codon usages are the same. The high value indicates high expression of gene with an optimal value calculated from 0.8 to 1. In addition, the software evaluated the GC content that has an optimum value ranging from 30 to 70%.

Afterwards, the SnapGene® software were used to visualize and simulate DNA cloning manipulation (SnapGene software from GSL Biotech; available at snapgene.com). The DNA vaccine sequence was

inserted into the plasmid pET 28a(+) via the restriction enzymes *Bam*HI and *Eco*RI. Their restriction site is present in the vector (multiple cloning sites) and was inserted at the vaccine DNA sequence flanking regions (3'OH and 5'PO₄, respectively).

2.13. Drug target identification, Molecular Modeling and prioritizations of identified drugs

The dataset of cytoplasmic proteins were subjected to MHOLline (<http://www.mholline.lncc.br>) server to predict the 3D models. MHOLline is a web platform that utilizes multi-fasta file of amino acids as an input data for model generation using MODELLER software (80). The adopted methodology was revised in accordance with previously published works (81–83). The final drug targets were prioritized based on criteria described by (81,84,85). Only the first four distinct quality model groups of G2 were taken into consideration in this study, these were: (i) Very High quality model sequences (identity $\geq 75\%$) (LVI ≤ 0.1), (ii) High quality model sequences (identity $\geq 50\%$ and $< 75\%$) (LVI ≤ 0.1), (iii) Good quality model sequences (identity $\geq 50\%$) (LVI > 0.1 and ≤ 0.3), (iv) Medium to Good (identity $\geq 35\%$ and $< 50\%$) (LVI ≤ 0.3) (<http://www.mholline.lncc.br>) (83,85,86). The online webserver PBIT (Pipeline Builder for Identification of drug Targets) was used for the non-homology analysis to human gut microbiota (54). Furthermore, the software GIPSy (Genomic Island Prediction Software) were used to identify the pathogenicity islands in the genome of *C. difficile* (87). This software detects presence of transposases, virulence or flanking tRNA genes, deviations in genomic signature (i.e., anomalous G+C and/or codon usage deviation) and absence of these regions in the commensal organism *Clostridium acetobutylicum* ATCC 824. The pathways were checked using KEGG (Kyoto Encyclopedia of Genes and Genomes) (88), and functionality analysis done using UniProt (89). The final identified non-host homologous, non-human gut microbiota homologous and presence in the Pathogenicity Island (PAIs) or Genomics Island (GIs) were considered as candidate drug targets and, then, was used for molecular docking analysis.

3. Results and Discussion

3.1. Identification of Intra-Species Conserved Non-Host Homologous Proteins and Subcellular localization

The pathogen *C. difficile* is a pivotal cause of nosocomial infections that entails a high morbidity and mortality rates, exacerbate health-costs, and worsen the multidrug resistance issue (15,17,18). This bacterium coevolves with the host being able to thrive in a number of conditions due to its multiple colonization and virulence factors (i.e. encoded in a large and complex genome), which leads to a *C. difficile* overgrowth and illness in dysbiotic conditions, especially after antibiotic use (48). Nowadays, approved treatment/recommended strategies narrows to a few antibiotics and FMT in challenging cases of recurrent CDI (18). Thus, the search for alternative treatment options is of utmost importance and requires cutting-edge technology approaches to provide efficient, safe and cost-effective therapies. Preventive tactics is a reliable method to avoid the use of antibiotics and hence the resistance problematic (90,91). We compared 53 strains of *C. difficile* (Supplementary Table 1) taking *C. difficile* ATCC 9689/DSM 1296 as reference using OrthoFinder software keeping the default parameters (52). A total of 2150 shared proteins were considered by all strains to be a part of the core genome. Furthermore, considering human as a host, a set of 438 conserved non-host homologous proteins were identified.

The prediction of subcellular localization of protein in or outside of the bacterial cell is of great significance to utilize the vast amounts of genomic data. The proteins are categorized based on their locality as the proteins that are exposed at the cell surface are usually targeted for vaccine development and cytoplasmic proteins for drug development respectively (53,83,92). Considering this, the 438 conserved non-host homologous proteins were screened for subcellular localization using SurfG+ software (53). We classified 250 proteins as cytoplasmic and 188 proteins as putative surface-exposed (PSE) proteins, secreted proteins, or membrane proteins.

3.2. Putative Vaccine targets for Immunoinformatics analysis

Secreted and membrane proteins are the primary elements that contact with the host, soaking an immune response. Therefore, the prediction of the exoproteome or secretome comprised of the proteins restricted to the outer membrane or extracellular matrix of the organism is significant for reverse vaccinology strategies (93). We used 188 proteins to predict vaccine candidates with adhesion probability greater than 0.5 using Vaxign and among them 20 proteins were selected. The homology analysis with human gut microbiota of these 20 proteins showed that six proteins (core genome) of *C. difficile* that are non-

homologues [i.e. stage II sporulation protein P spoIIP (AKP43429), ABC transporter permease (AKP42475), hypothetical protein CDIF1296T_02948 (AKP43769), spore coat protein CotH (AKP41996), hypothetical protein CDIF1296T_02351 (AKP43195), and lipoprotein (AKP41595)]. Furthermore, four additional non-human homologues proteins from the accessory genome (i.e. not common among toxigenic and non-toxigenic strains) were chosen due to its importance to the bacterium pathogenicity [i.e. Toxin A or TcdA (AGG91516), Toxin B or TcdB (AGG91581), ADP-ribosylating binary toxin binding subunit Cdtb (TGA38378), S-layer protein SlpA (AKP43753)] (Table 1).

The four selected proteins were extensively studied by many researchers and their connection with *C. difficile* infections confirmed through experiments. The role of *C. difficile* toxins (i.e. TcdA, TcdB and CDT) it is of utmost importance for the pathogen infection and determine the severity of the disease. TcdA and TcdB are the main virulent factors (i.e. exotoxin and cytopathic toxin, respectively) with the intestinal epithelial cells as primary target, which leads to loss of intestinal membrane integrity followed by host exposure to intestinal microorganism and activation of the host inflammatory response. The other relevant toxin is CDT, a binary toxin composed of two proteins, CdtA and CdtB, that interferes with the actin meshwork polymerization of cellular membranes leading to protrusions in the cell surface that increase the pathogen adherence to the host (94,95). The presence of those toxins is extremely variable among *C. difficile* strains (i.e. bacterium can hold a combination of none, one, two, three or all four toxins) as the variation in the toxin sequence, which is directly linked with a higher morbidity and mortality in patients and with the bacterium heterogeneity, respectively (44,96,97). ToxA and ToxB are currently being tested as vaccine targets (e.g. VLA84, ACAM-CDIFF and PF-06425090) in different clinical trials (40,41,43,98), which further support the selection of these proteins for the multi-epitope vaccine proposal. Similarly, the S-layer proteins (SLPs), major surface structures composed of one or more proteins (up to twenty-nine), have been appraised as vaccine antigen due to their immunodominant characteristics (99). SLPs play a critical role in the pathogenesis (i.e. function as adhesins) and host immune modulation (i.e. induce the release of pro-inflammatory cytokines and humoral response) (100–102). SLPs consist of high-molecular weight (HMW) and a low-molecular weight (LMW) proteins derived by a precursor or pre-protein SlpA (101,103). However, SlpA is highly variable among *C. difficile* strains, which may affect the recognition by the immune system (103,104). Despite that, SlpA is being tested to treat/prevent CDI. Anti-SlpA antibodies (passive immunization) had promising results conferring a partial short-term protection against *C. difficile* (105,106) and SlpA vaccine may aid in the prevention of intestinal colonization but, so

far, did not result in significant protection against CDI (105,107–109). The lack of protection may be related with SlpA variability and the multifactorial *C. difficile* colonization process; therefore, an effective vaccine against this bacterium should target a panel of antigens that would enhance the host defense.

The remaining six secreted or transmembrane proteins were selected, via reverse vaccinology, since they belong to *C. difficile* core genome, lack homology to human and gut microbiota proteins, and have putative antigenic properties; however, most have uncharacterized functions. SpoIIP and CotH are intertwined with *C. difficile* spore form. Spores are vital for the infection and CDI since they are necessary for the attachment to the gut mucosa, the dissemination and survival of the pathogen, being responsible for the initiation of *C. difficile* colonization and infection maintenance in acute stages (110,111). SpoIIP is a major protein (i.e. autolysin) responsible for the engulfment of the forespore layer and polar cytokinesis blockage (spore division) in *Bacillus subtilis* (112,113). AKP43429 is homologous of SpoIIP from *B. subtilis* and presents a similar function (114,115). Meanwhile, CotH is a kinase protein required for the control assembly of both inner and outer layers of the coat spore and necessary for sporulation of many spore-forming bacteria (e.g. bacillus and clostridia), including *C. difficile*, which suggests its importance for the spore morphogenesis (111,116,117). Furthermore, CotH is directly or indirectly linked with spore adhesion to the mucosa (111,117). Altogether, these suggest that proteins involved in the spore formation and in the sporulation process are relevant vaccine targets.

Likewise, *C. difficile* lipoproteome have a vital accountability in promoting CDI since lipoproteins are associated with colonization, host-pathogen modulation, virulence, nutrient uptake, and sporulation (118–120). Lipoproteins are characterized as surface exposed proteins with a capacity to stimulate cytotoxic T lymphocyte production (CTL), thus are promising vaccine candidates (119,121,122). Among *C. difficile* described lipoproteins CD0873, OppA and AppA highlights due to its importance during colonization, sporulation and virulence (120,123); however, other lipoproteins may be determinant for CDI, like AKP41595. This protein is common among all *C. difficile* strains, which may indicate an important function, but no conserved domains were detected via BLASTp and a putative function cannot be assigned being necessary further investigations to uncover AKP41595 role.

AKP42475 encodes an ABC transporter permease of uncharacterized function. This protein has an AcrA (membrane-fusion protein or MFP) conserved domain known to be a multidrug efflux pump subunit, which are involved in cell biogenesis and defense mechanisms. Furthermore, AKP42475 has similarities

with the Resistance, Nodulation, and cell Division (RND) family efflux transporter (i.e. nodulation, transport of heavy metal efflux and multidrug resistance proteins), HlyD membrane-fusion protein (i.e. transport apparatus of alpha-haemolysin), and multidrug efflux system protein MdtE. The BLAST results suggest a similarity with YknX from *Bacillus*, a MFP protein with a role in diverse physiological functions (e.g. protection against sporulation-delaying-protein-induced killing) in Gram-negative and Gram-positive bacteria (124–127).

Lastly, two *C. difficile* hypothetical proteins were also selected for the construction of the multi-epitope vaccine. CDIF1296T_02948 (AKP43769) putative function cannot be predicted since the protein has no putative conserved domains detected whereas CDIF1296T_02351 (AKP43195) has a zinc-ribbon domain with an unknown function. Member of the zinc-ribbon family were previously associated with DNA, RNA, proteins or small molecules modulation (128) but no function can be assigned to the protein. Therefore, experiments are required to stipulate their role.

Protein codes and node	Annotation	Cellular component	Pan genome outcome
AKP43429 spoIIP	stage II sporulation protein P	Secreted	Core genome
AKP42475	ABC transporter permease	Transmembrane	Core genome
AKP43769	hypothetical protein CDIF1296T_02948	Transmembrane	Core genome
AKP41996 CotH	spore coat protein CotH	Transmembrane	Core genome
AKP43195	hypothetical protein CDIF1296T_02351	Transmembrane	Core genome
AKP41595	Lipoprotein	Secreted	Core genome
AGG91516 TcdA	Toxin A or ToxA	Secreted	Accessory genome
AGG91581 TcdB	Toxin B or ToxB	Secreted	Accessory genome

TGA38378 CdtB	ADP-ribosylating binary toxin binding subunit CdtB	Secreted	Accessory genome
AKP43753 SlpA	S-layer protein SlpA	Secreted	Accessory genome

Table 1: List of *C. difficile* proteins selected from Vaxigen and the literature. Ten proteins from *C. difficile* picked by reverse vaccinology methodology (top six) and considered important for colonization and/or pathogenesis (bottom four), in accordance with literature data, were initially selected for immunoinformatics analyses.

Utterly, this work propose an assortment of relevant *C. difficile* proteins to certain that the vaccine stimulates a stronger immune pressure and eliminate the pathogen from the host. The proteins are associated with virulence (TcdA, TcdB, CdtB), adhesion (SlpA), formation/constitution of the spore (SpoIIP and CotH), putative role in essential cellular processes or bacterial pathogenesis (AKP41595), transport (i.e. putative membrane multidrug efflux pump) (AKP42475), and uncharacterized functions (AKP43769 and AKP43195). These selected proteins will compose a multi-epitope vaccine that, in theory, can induce an adaptive immune response (cellular and humoral) through the display of the antigens (epitopes) by MHC-I, MHC-II and B cells. Multi-epitope based strategies is an innovative approach that assure a better coverage of the pathogen (i.e. different life stages, strains and virulent factors), which reduce the probability of the microorganism evade the immune system (91). Furthermore, peptide based vaccines can induce protection in a customized, simple, fast, reproducible and cost-effective way, without contamination problems, and with a reduce change of autoimmune/allergenic reactions (91).

3.3. Immunoinformatics analysis for the Multi-epitope vaccine

3.3.1. B and T cell epitope prediction

ABCpred server were used to predict the 16-mer B cell epitopes with a score above 0.51 whereas the IEDB server predicted MHC type I and II binding epitopes with a percentile rank below 3% for all ten proteins described in the previous step (data not shown).

The B cell epitopes were compared concurrently with MHC type I (9-mer) and II (15-mer) binding epitopes and the common sequential epitopes with nine and fifteen to twelve amino acids, respectively, were selected (data not shown). Afterwards, the predicted MHC-I and MHC-II epitopes picked after B cell comparison were further compare with each other and the epitopes with nine and eight common sequential amino acids were used to construct the *C. difficile* vaccine. From the ten initial proteins, all except SpoIIP had overlapping B cells, MHC-I and MHC-II epitopes (AKP42475, AKP43769, AKP41996, AKP43195, AKP41595, AGG91516, AGG91581, TGA38378, and AKP43753) (Table 2).

Protein	MHC-I binding epitope	MHC-II binding epitope
AKP42475	LDGILKKQI	DGILKKQIVDVASQN
AKP43769	TNYKRTLTM	YKRTLTMAGIVPESL
AKP41996	SMIIMSIML	GASMIIMSIMLVGVS
	IMLVGVSrv	MIIMSIMLVGVSrvk
AKP43195	WAFANGISV	YMWAFANGISVTIIN
	YAMGALSML	MFYAMGALSMLLLL
AKP41595	SMIAQLTPL	KDKVSMIAQLTPLYG
AGG91516	YYFNPNNAI	NEKYFNPNNIAAAV
	YYFNTNTSI	GKKYYFNTNTSIAS
	YYFNTNTAE	GKKYYFNTNTAEAA
	FAPANTQNN	GFEYFAPANTQNNNI
	FAPANTHNN	NGFEYFAPANTHNNN
AGG91581	KYFAPANTV	SDGYKYFAPANTVND
	WEMTKLEAI	FWEMTKLEAIMKYKE
	SLYYFKPPV	DSLYYFKPPVNNLIT
	YIAATGSVI	TDEYIAATGSVIIDG
	QEIEAKIGI	QEIEAKIGIMAVNLT
TGA38378	QEIMDAHKI	DQEIMDAHKIYFADL
	IMTYKKLRI	ENIMTYKKLRiyAIT
AKP43753	TEIKVFFEG	TEIKVFFEGTLASTI
	SIAPVASQL	LADAMSIAPVASQLR
	APVAANFGV	DALAAAPVAANFGVT
	LAMSAIFDT	NSAGTKLAMSAIFDT
	NKLVSPAPI	DKNKLVSPAPIVLAT

Table 2: Elected MHC-I and MHC-II binding epitopes. List of the nine *C. difficile* proteins selected for the multi-epitope vaccine with the correspondent MHC-I and MHC-II binding epitopes

The proteins TcdA (AGG91516), TcdB (AGG91581) and SlpA (AKP43753) presented multiples predicted epitopes after B cell/MHC-I/MHC-II comparison (data not shown); thus, the top five (minor score) non-similar epitopes (i.e. with difference of four or more sequential residues in the edges) were selected to compose the multi-epitope vaccine.

3.3.2. Multi-epitope *C. difficile* vaccine design

The final vaccine has 24 MHC-I and 24 MHC-II epitopes from nine proteins (Table 2) that were interconnected via linkers, AAY sequence between MHC-I epitopes and GPGPG between MHC-II epitopes, since covalent linkage is necessary for the generation of epitope-specific antibodies response. Likewise, adjuvants are often required due to the poor immunogenicity properties of peptide vaccines and to keep the native conformation of the protein, although no one was added into the structure. To the best of our knowledge, no epitope with adjuvant properties was described for *C. difficile*, indeed, little is known about the innate immune system recognition of the bacterium. Ryan and colleagues (2011) uncover the role of SLPs in the activation of the immune system (i.e. maturation of dendritic cells and subsequent stimulation of Th1 and Th17 response) *in vivo* and *in vitro*; however, the combination of SLPs is needed for the TLR4 recognition. The SlpA protein was added to the multi-epitope vaccine but since the direct connection of the SLP precursor with the innate immune cells activation cannot be made, the addition of an adjuvant or delivery system is recommended.

The multi-epitope *C. difficile* vaccine has forty-eight epitopes with a final length of 765 amino acids (epitopes plus linkers) and the schematic is displayed in Figure 1.

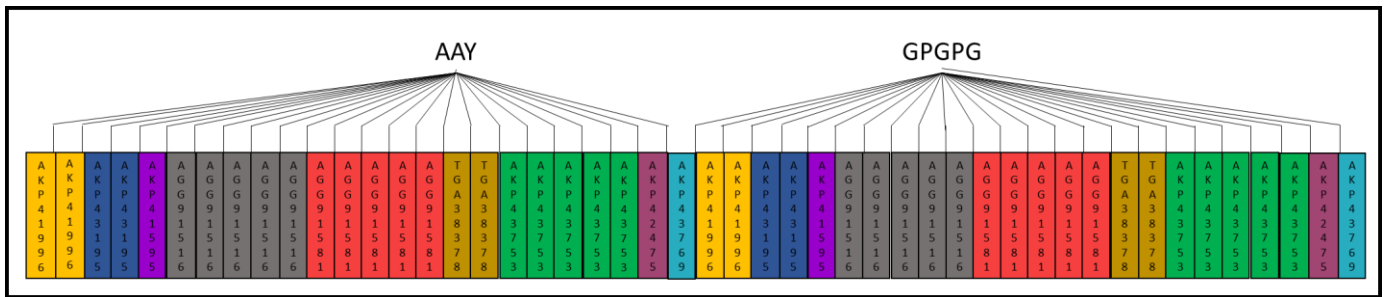


Figure 1: Vaccine construction schematics. The epitopes (24 MHC-I and 24 MHC-II) selected after B cell, MHC-I and MHC-II comparison were joined by appropriated linkers. AAY amino acid sequences were added between MHC-I type epitopes and GPGPG between MHC-II epitopes. The vaccine was assembled without a peptide adjuvant.

3.4. Homology modelling, refinement and validation of the tertiary structure

The I-TASSER server predicted the top five tertiary structure of the multi-epitope vaccine (data not shown) and the model with the most reliable prediction or the highest C-score (-0.43) was selected (Figure 2A). The quality of the modeling (TM-score) was 0.66 ± 0.13 which indicates the model probably has a correct topology in accordance with the structural similarity between the native and predicted structures. Afterwards, the tertiary structure predicted by I-TASSER was refined using the GalaxyRefine server through repeated structure perturbations and relaxation in the clusters side-chains, secondary structure elements and loops of the model (Figure 2B). The perturbations were done to improve the dihedral angles ψ and ϕ of the residues measured in the Ramachandran plot. From the five refined structure models generated by GalaxyRefine, the model with the Rama favored score of 86.0% was chosen and the angles and atoms collisions evaluated.

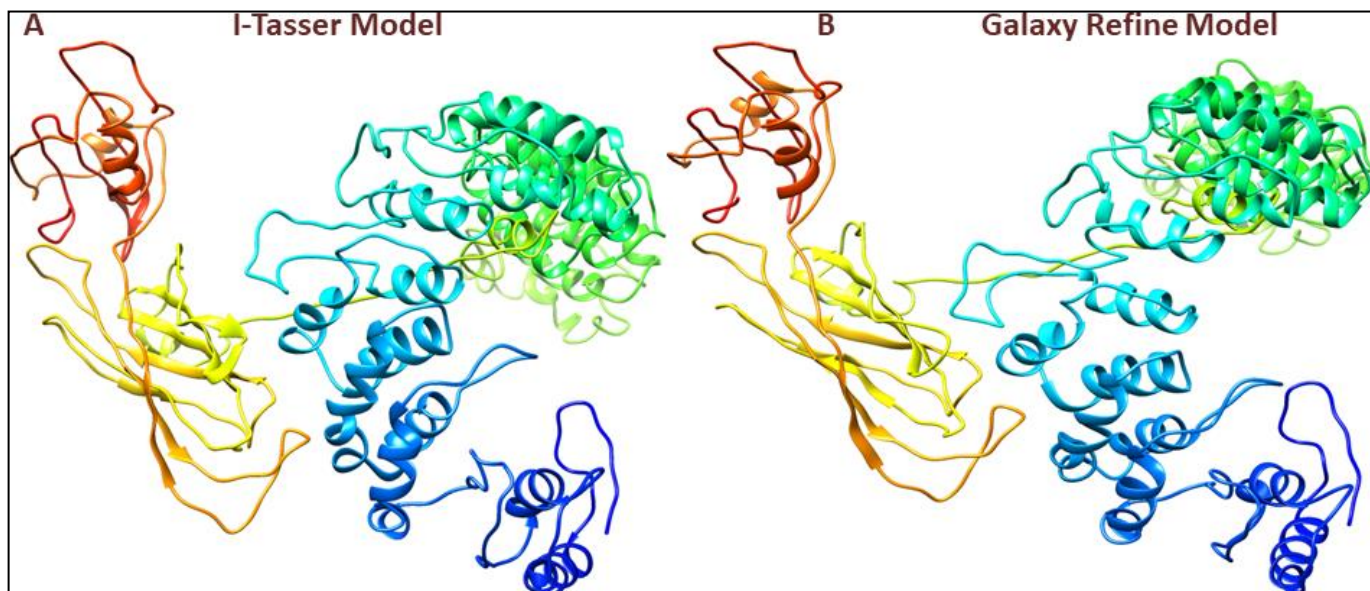


Figure 2: The predicted tertiary structure of the *C. difficile* multi-epitope vaccine. (A) The vaccine 3D structure was predicted by I-TASSER and, among the top five predicted structures (-4.08 ~-0.43) (data not shown), the model with the highest C-score (-0.43) was chosen. (B) The model selected by I-TASSER was refined by GalaxyRefine and the structure with the Rama favored score of 86.0% was elected.

The amino acid residues angles and the steric collisions between atoms of the protein structures (not refined and refine) were calculated and visualized via the Ramachandran plot generated by SAVES 5.0 server using Procheck (Figure 3). The plot from I-TASSER shows that 69.3% of the residues are in most favorable regions, 25.0% in addition allowed, 2.4% in generously allowed, and 3.4% in disallowed regions (Figure 3A) whereas the plot of the refined structure shows that 84.5%, 11%, 1.9% and 2.7% of the amino acids were set in most favorable, additional allowed, generously allowed and disallowed regions, respectively (Figure 3B). The refined 3D model significantly improved the amino acid positions and steric clashes and thus was employed as the predicted tertiary structure of the *C. difficile* multi-epitope vaccine in the following steps.

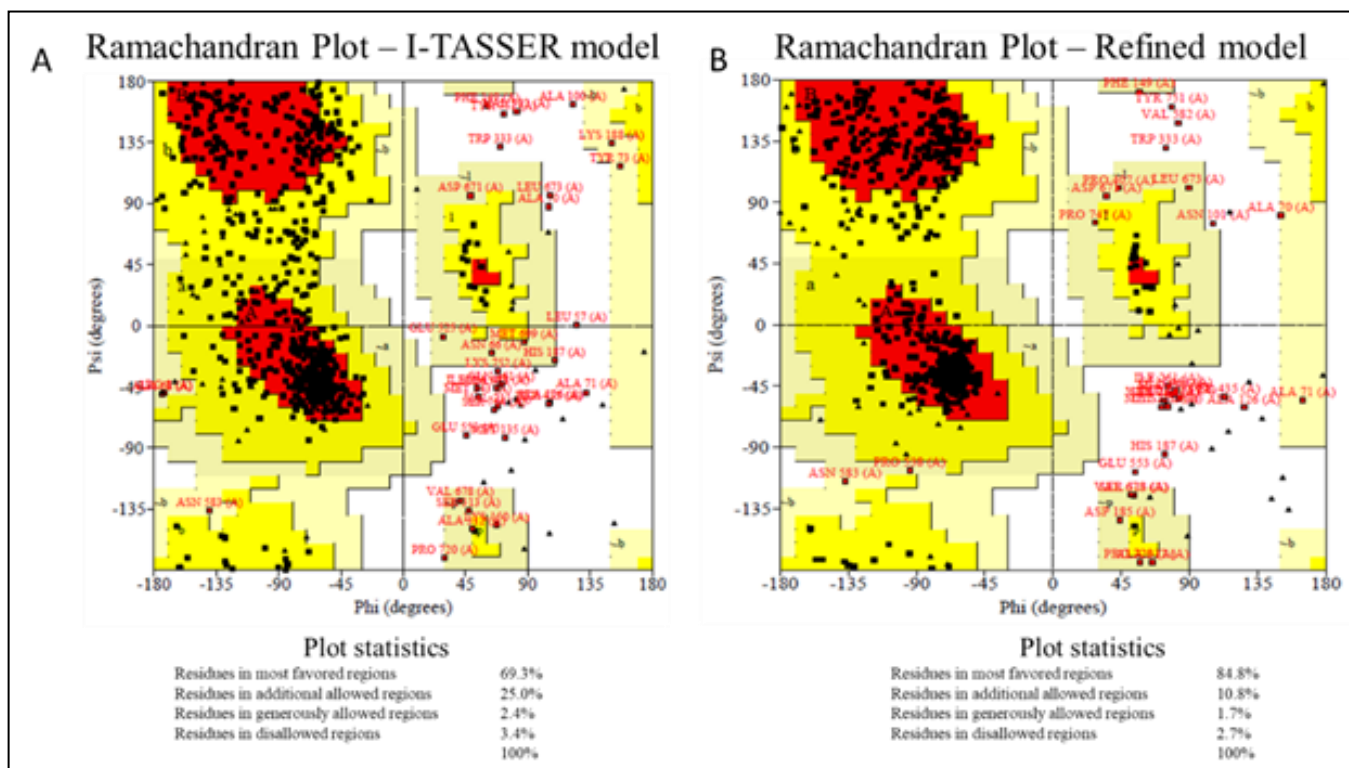


Figure 3: Ramachandran plots of the multi-epitope *C. difficile* vaccine. The 3D structures obtained from I-TASSER and GalaxyRefine were used to calculate the ϕ and ψ backbone dihedral angles of amino acid residues to visualize the energetically allowed/disallowed regions. (A) The Ramachandran plot of the 3D structure revealed that 69.3%, 25.0%, 2.4% and 3.4% represent residues in most favored, additional allowed, generously allowed and disallowed regions, respectively. (B) The Ramachandran plot of the refined 3D structure revealed that 84.8%, 10.8%, 1.7% and 2.7% represent residues in most favored, additional allowed, generously allowed and disallowed regions, respectively.

3.5. Linear and conformational B cell binding epitopes

The humoral response is important to prevent and combat *C. difficile* infection (129); therefore, ABCpred predicted the linear B cells epitopes within the complete vaccine sequence and the conformational or discontinuous epitopes were predicted, applying the 3D structure obtained by GalaxyRefine, via the EliPro server.

ABCpred predicted seventy-seven linear epitopes using the default parameters (Table 3). The threshold setting of 0.51 shows 65.93% accuracy with equivalent sensitivity and specificity using a sixteen amino acid length.

Rank	Sequence	Start position	Score	Rank	Sequence	Start position	Score
1	GVS GPGPGYMWAFANG	323	0.95	40	ANTVNDGPGPGFWEMT	500	0.80
2	P GPGKDKVSMIAQLTP	367	0.94	41	GFEYFAPANTHNNNGP	472	0.80
3	TNTSIAS T GPGPGGKK	418	0.93	42	SVTIINGPGPGMFYAM	340	0.80
4	HKIYFADL GPGPGENI	598	0.92	43	GSVIAAYQEIEAKIGI	162	0.80
5	GQEIEAKIGIMAVNLT	570	0.92	44	DALAAAPVAANFGVTG	671	0.79
6	TGSVIIDGGPGPGQEI	558	0.92	45	YQEIMDAHKIAAYIMT	180	0.79
7	P GPGDSL Y YFKPPVNN	527	0.92	46	GIMAVNLTGPGPGDQE	578	0.78
8	P GPGSDGYKYFAPANT	487	0.92	47	RTLTMGPGPGMIIMSI	281	0.78
9	P GPGGKKYYFNTNTAE	427	0.92	48	YQEIEAKIGIAAYQEI	168	0.78
10	P GPGGKKYYFNTNTSI	407	0.92	49	DGILKKQIVDVASQNG	731	0.77
11	FDTGPGPGDKNKLVSP	703	0.90	50	GNSAGTKLAMSAIFDT	690	0.77
12	TGPGPGGFEYFAPANT	445	0.90	51	AITGPGPGTEIKVFFE	623	0.77
13	TNTAE AATGPGPGGFE	438	0.90	52	GPGPGDQEIMDAHKIY	586	0.77
14	ALSMLAAYSMIAQLTP	41	0.90	53	AYYAMGALSMLAAYSM	35	0.77
15	TPLYGGPGPGNEKYYF	381	0.90	54	NFGVAAYLAMSAIFDT	234	0.77
16	MSAIFDTAAYNKLVSP	243	0.90	55	ANTVAAYWEMTKLEAI	126	0.77
17	NTSIAAYYYFNTNTAE	78	0.89	56	TNTAE AAYFAPANTQN	89	0.75
18	GPGPGNGFEYFAPANT	466	0.89	57	DEYIAATGSVIIDGGP	552	0.75
19	PAPIAAYLDGILKKQI	258	0.89	58	SIMLAAYWAFANGISV	18	0.75
20	NTHNNAAYKYFAPANT	113	0.89	59	P GPGGLADAMSIAPVAS	647	0.74
21	IVLATGPGPGDGILKK	721	0.88	60	ASMIIMSIMLVGVS GP	312	0.73
22	P GPGNEKYYFNPNNAI	387	0.88	61	LEAIAAYSLYYFKPPV	138	0.73
23	AAYIMTYKKLRIAAYT	190	0.88	62	YSMIIMSIMLAAYWAF	12	0.73
24	ASQLRGPGPGDALAAA	661	0.86	63	DGGPGPGQEIEAKIGI	564	0.72
25	GENIMTYKKLRITYAIT	610	0.86	64	AAYSIAPVASQLAAYA	214	0.71
26	QEIMDAHKIYFADLGP	592	0.86	65	LAAYYYFNPNNAI AAY	57	0.70
27	EGTLASTIGPGPLAD	638	0.85	66	AAYYIAATGSVIAAYQ	154	0.70
28	TGPGPGTDEYIAATGS	545	0.85	67	GVS RVAAYS MIIMSIM	5	0.68
29	VSRVKGPGPGGASMII	301	0.84	68	GMIIMSIMLVGVS RVK	290	0.66
30	P GPGYKRTLTMAGIVP	747	0.83	69	GPGPGMFYAMGALSML	346	0.65
31	NNAIAAYYYFNTNTSI	66	0.83	70	AYKYFAPANTVAAYWE	119	0.65
32	MIAQLTPLAAYYYFNP	50	0.83	71	ANTQNNNIGPGPGNGF	458	0.64
33	AIAAVGPGPGGKYYF	401	0.83	72	LSMLILLGPGPGKDK	358	0.64
34	LKKQIAAYTNYKRTL T	269	0.83	73	KKLRIAAYTEIKVFFE	197	0.64
35	DVASQNGPGPGYKRTL	740	0.82	74	ANTQNN AAYFAPANTH	100	0.64
36	FGVTGPGPGNSAGTKL	682	0.81	75	YSLYYFKPPVAAYYIA	144	0.58
37	PGFWEMTKLEAIMKYK	509	0.81	76	PGTEIKVFFEGTLAST	629	0.56
38	QLAAYAPVAANFGVAA	224	0.81	77	KNKLVSPAPIVLATGP	712	0.55
39	KKLRITYAITGPGPGTE	617	0.80				

Table 3: Multi-epitope vaccine predicted B cell linear epitopes. The ABCPred server was used to predict the B cell binding epitopes via the amino acid sequence of the multi-epitope vaccine. Seventy-seven epitopes were predicted with a score varying from 0.95 to 0.55 (default parameters).

Furthermore, Ellipro predicted six discontinuous epitopes with a score varying from 0.768 to 0.562 (Supplementary Table 2). The score or PI calculates the percentage of residues located within the ellipsoid, a quadric surface area with three pairwise perpendicular axes intersect at the center of symmetry, and higher percentages are associated with greater solvent accessibility. Among the predicted B cell epitopes, three had a PI equal or superior to 0.7 (Figure 4), which indicates that 70% of the protein residues are located within the ellipsoid and 30% outside.

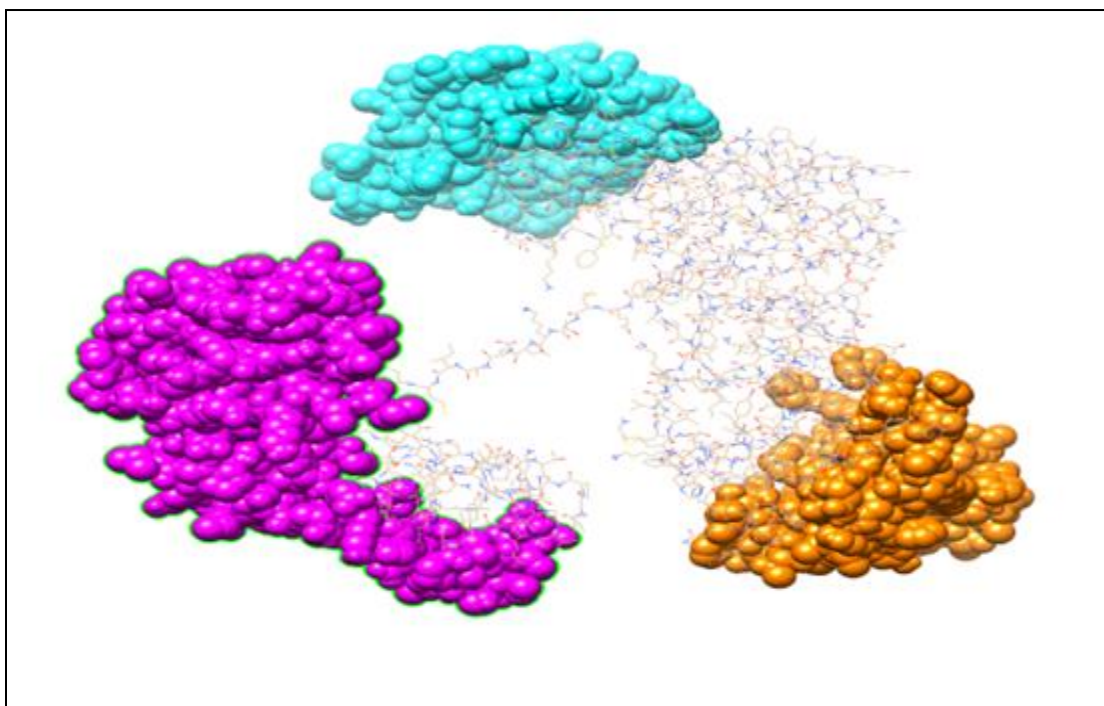


Figure 4: *C. difficile* multi-epitope vaccine tertiary structure with the location of B-cell conformational epitopes. Three predicted conformational B-cell epitopes with a PI score of 0.768 (Cyan), 0.714 (Orange), and 0.711 (Magenta) visualized in the 3D structure of the vaccine candidate.

The results suggest that the *C. difficile* multi-epitope vaccine can stimulate a humoral response via the recognition of linear and discontinuous B cell epitopes, which can aid the bacterium clearance and prevent infections. B cells have a protective role in *C. difficile* infections where TcdA and TcdB humanized monoclonal antibodies reduced CDI recurrences, and toxin-specific immunoglobulin A and B (IgA and IgG) influence the symptoms of the disease (e.g. severity is inversely correlated with the presence of antibody) (129–131). Furthermore, the effectiveness of B cell induction was demonstrated in a ToxA/ToxB recombinant vaccine (VLA84), which exhibits a protective result (adult and elderly patients) against *C. difficile* provoked diarrhea and recurrent CDI, in a Phase I clinical trial, through the induction of high levels of IgG antibody against TcdA and TcdB (40). Likewise, other vaccine are undergoing clinical trials (e.g. ACAM-CDIFF and PF-06425090) and show promising results with considering tolerability, immunogenicity and safety (41,43,98).

3.6. Interferon gamma inducing epitopes

MHC-II epitopes have the capability to induce IFN-gamma, a relevant cytokine for innate immune cells activation; thus, the MHC-II epitopes of the vaccine candidate were evaluated (Table2). The prediction was performing through the IFNepitope server using the hybrid approach (Merci and SVM methods). Eight MHC-II binding epitopes are characterized as IFN-gamma inducing epitopes (Table 4).

Sequence	Method	Result	Score
NEKYYFNPNNIAAAV	SVM	POSITIVE	0.54149582
GKKYYFNTNTSIASST	SVM	POSITIVE	0.18535845
GKKYYFNTNTAEAAAT	SVM	POSITIVE	0.29007923
ENIMTYKKLRITYAIT	SVM	POSITIVE	0.26382918
TEIKVFFEGTLASTI	SVM	POSITIVE	0.056350639
LADAMSIAPVASQLR	SVM	POSITIVE	0.049967887
DALAAAPVAANFGVT	SVM	POSITIVE	0.050897848
NSAGTKLAMSIFDT	SVM	POSITIVE	0.080276714

Table 4: Multi-epitope vaccine predicted IFN-gamma inducing epitopes. Eight MHC-II binding epitopes capable to induce IFN-gamma (positive) were predicted using IFNepitope server hybrid method (MERCII and SVM).

The production of IFN- γ not only activates neutrophils and macrophages but also induces the expression of antimicrobial peptides and the production of reactive nitrogen and oxygen species (RNS and ROS, respectively), which increases phagocytic mechanisms (132,133). Thus, the presence of predicted IFN-gamma inducing epitopes in *C. difficile* multi-epitope vaccine is a critical feature and may assist the prevention of the bacterium overgrowth.

3.7. Vaccine properties evaluation

3.7.1. Allergenicity evaluation

AllerTOP 2.0 and AlgPred servers evaluated the allergenicity of the multi-epitope vaccine. AllerTOP 2.0 classified the protein as probable non-allergenic based on the similarities with the non-allergenic protein from the server dataset. AlgPred categorized as non-allergenic based on five distinct predictions approaches that the vaccine sequence with known sequences from allergens and non-allergens proteins. AlgPred results indicate that the vaccine does not contain experimentally mapped IgE epitopes, no allergenic motif, potential allergen based on the amino acid composition (score of 0.444 with a threshold of -0.4) and on the dipeptide composition (score of 0.806 with a threshold of -0.2), no allergen representative peptides, and no allergen/allergenic via a hybrid approach (improved sensitivity and specificity). The protein sequence were also compare against the *Homo sapiens* reference protein database via BLASTp (*Homo sapiens* (human) protein BLAST) (134) with no significant similarity found. Altogether, the result indicates that the multi-epitope vaccine will not induce an autoimmune or allergenic reaction.

3.7.2. Antigenicity evaluation

The vaccine antigenicity was predicted by Vaxijen v.2.0 with an overall protective antigen of 0.6655, which indicates that the protein is probable antigenic. AntigenPro was used to further confirm the probability of antigenicity of 0.663746. The results suggest that the multi-epitope vaccine will induce a host immune response.

3.7.3. Physicochemical parameters

The ProtParam server calculated the multi-epitope vaccine physicochemical description. The estimated molecular weight of vaccine was proximally 79.36 kDa and the theoretical pI was 8.13, which indicates that the protein is basic in nature. The total number of negatively (Asp + Glu) and positive (Arg + Lys) charged residues were 41 and 43, respectively. The estimated half-life predicted was 20 hours in mammalian reticulocytes (*in vitro*), > 30 minutes in yeast (*in vivo*) and > 10 hours in *Escherichia coli* (*in vivo*). The measured absorbance (optical density) and extinction coefficient of the native protein in water at 280 nm was 1.347 and $106930 \text{ M}^{-1}\text{cm}^{-1}$, respectively. The instability index computed was 25.48, which indicates that the vaccine is stable since the index is smaller than 40. The calculated GRAVY value (0.105) indicates that the protein is thermo stable and the predicted aliphatic index (78.27) designates that the protein is polar.

3.8. Protein–protein docking between vaccine candidate and TLR4

TLR4 is a transmembrane protein that belongs to the pattern recognition receptor (PRR) family. This receptor is crucial for the recognition of pathogens via the pathogen-associated molecular patterns (PAMPs) and subsequently activates the innate immune system that mediates the production of cytokines. The described PAMPs recognized by TLR4 are lipopolysaccharide (LPS), viral proteins, polysaccharide, and a variety of proteins such as beta-defensins, heat shock protein, and surface proteins (i.e. SLPs from *C. difficile*) (102,135–137). TLR4 is able to recognize *C. difficile* and induce an immune response *in vitro* and *in vivo* (102) and thus was chosen for the docking analysis.

TLR4 model structure (PDB 2Z63) has 570 residues enumerated from 27 to 596 with the signal peptide (1-26 amino acids) excluded from the crystal protein complex (138). The edited PDB 3D model was used to locate the frustrated regions via the Frustratometer server and Chimera (Figure 5). Frustratometer analyzed the energy distributed in the protein providing insights on its biological behavior. Figure 5A shows three highly frustrated areas (i.e. biologically important binding sites) in the TLR4 structure, which is also observed in the density graphic by the red pics (Figure 5B). The hydrophobic regions visualized with Chimera matches the predicted results from Frustratometer (Figure 5C). The selected frustrated areas correspond to the amino acids: (i) 27 to 68 that belongs to the N-terminal region (LRRNT and LRR1 motifs) (area 1); (ii) 135 to 191 from the N-terminal region (LRR4 to LRR6 motifs) (area 2); and (iii) 547 to 588 from the C-terminal region (LRR21, LRR22 and LRRCT motifs) (area 3).

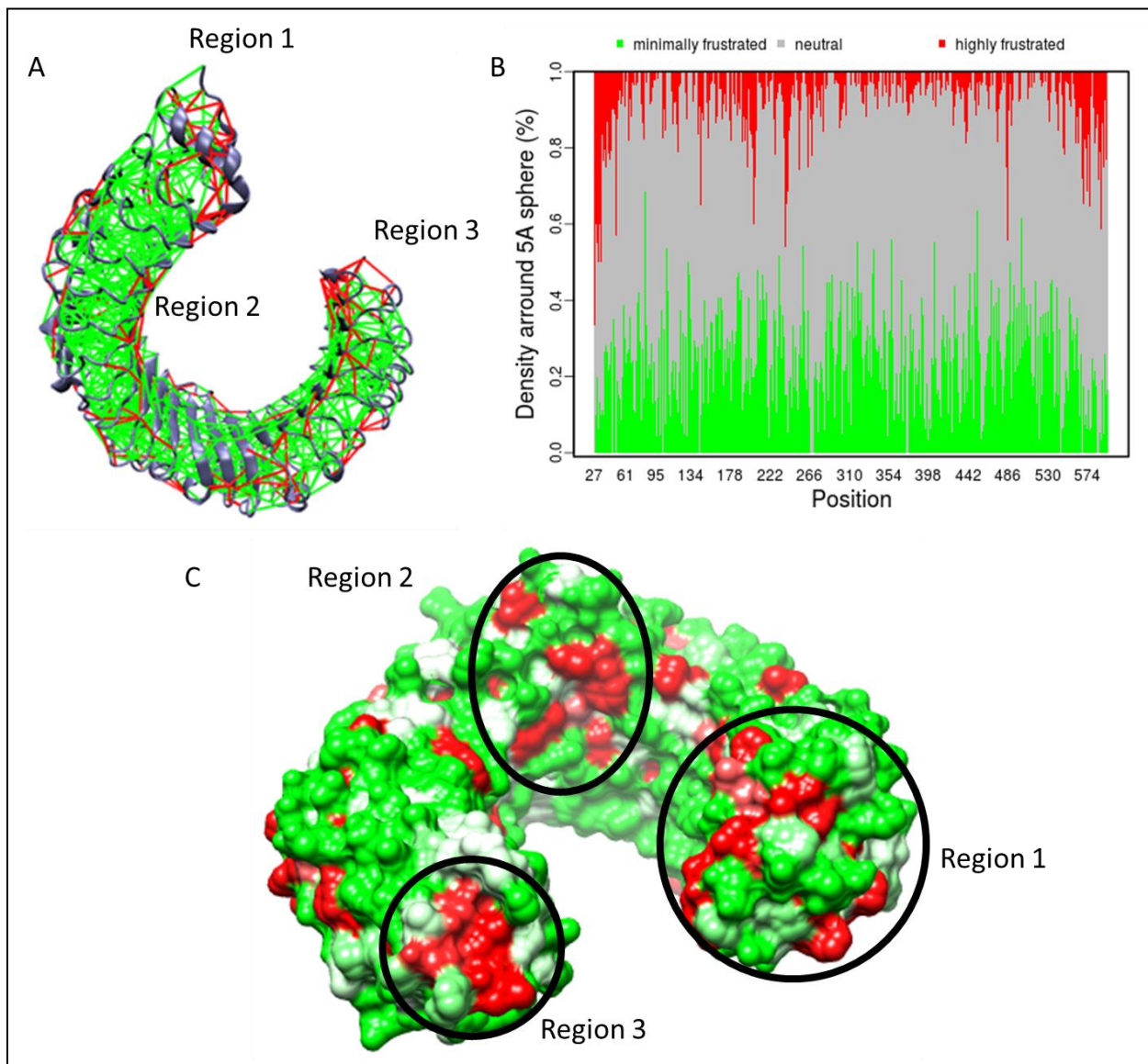


Figure 5: TLR4 frustrated regions. The Frustratometer server quantified the degree of local frustration in the TLR4 structure whereas Chimera was used to visualize TLR4 hydrophobic regions in accordance with the amino acid sequence. (A and B) High local frustrated sites are shown in red and indicate biologically important regions (binding sites). Minimally frustrated are marked in green and contributes to a stable folding core of the molecule, and the neutral regions are shown in grey. Three highly frustrated regions were selected and their sites can be visualized in the 3D structure (A) and in a graphic form (B). Region 1 and 2 matches the N-term region (amino acids 27-68 and 135-191, respectively), and region 3 the C-term domain (amino acids 547-588). (C) The red regions represents the residues with a higher hydrophobicity (high positive value), the green regions denotes hydrophilic regions (negative value), and the white constitute neutral regions. The high hydrophobic regions visualized by Chimera correlate with the frustrated areas predicted by Frustratometer.

The residues from the three highly frustrated areas were selected as the TLR4 binding regions in the restricted docking performed by the SwarmDock server. The restricted docking was preferred over the blind docking (i.e. no selected binding region) since it has a better probability to obtain the correct solution. The edited TLR4 were used as the receptor and the multi-epitope vaccine as the ligand, and the server generated 64 clusters with four different forms each totalizing 239 possible docking structures (with a few models with more than one member). The UCSF Chimera software were used to analyze the top ten democratic docking models and visualize the H-bonds present between the TLR4 binding sites and the vaccine (Supplemental Table 3). The most plausible model were selected based on the protein-protein interactions (fifteen H-bonds located within the binding region), the calculated energy (-42.26) and the total number of contacts (750 contacts between TLR4 and the vaccine) (Figure 6).

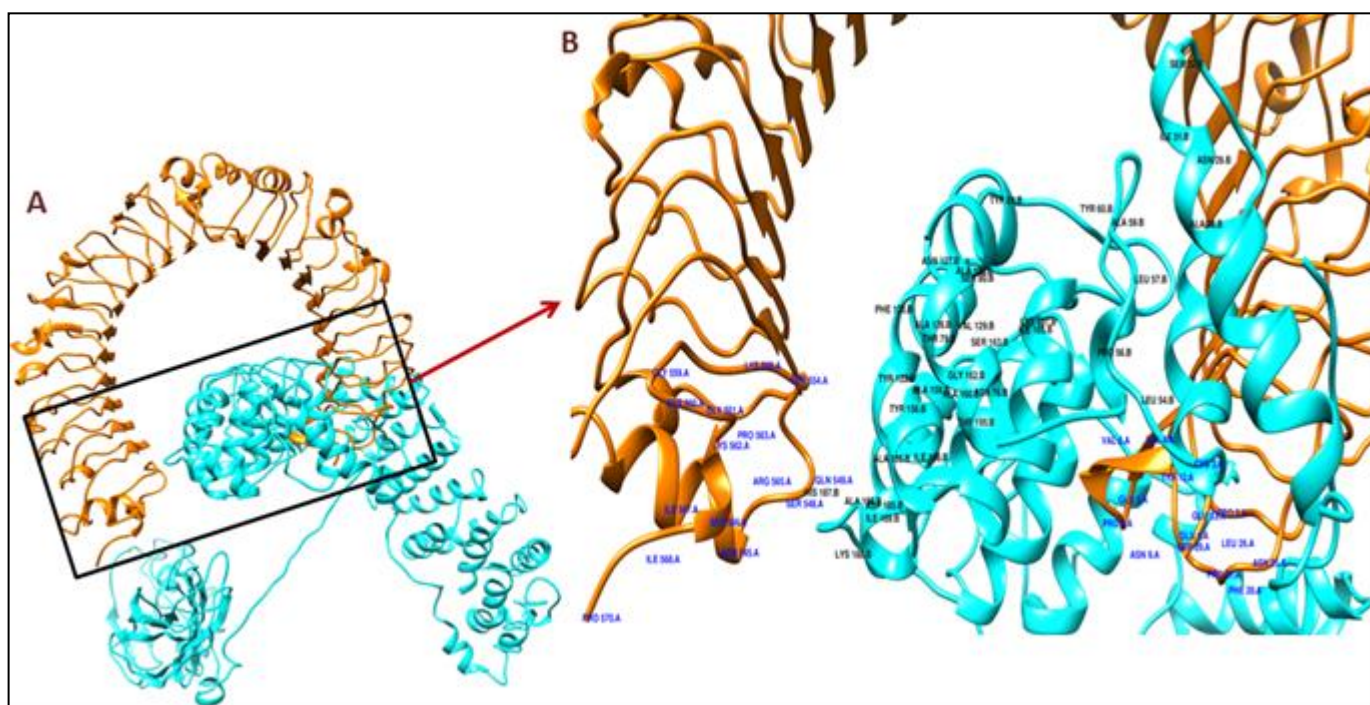


Figure 6: TLR4-vaccine docking structure predicted by SwarmDock. A: The best dock solution was selected as the most probable docking structure. The ligand (vaccine protein) is shown in Cyan and the receptor (TLR4) in Orange. B: The residues frustrated regions (binding sites) are shown.

3.9. MD simulation

The stability of the docked TLR-4 and multi-epitope based vaccine complex was assessed by molecular dynamics performed with Gromacs v5.1 2 software. After energy minimization step, we did an equilibration of system using temperature and pressure. The pressure plot indicated a fluctuation in pressure around 1 bar (Figure 7 A) with 400 ps of time interval. The temperature plot shows that the system maintained 300 K (Figure 7 B) at the same time interval. The production of the system was done by 30 ns and the results were measure using the two most common measures of structural fluctuations: Root-Mean-Square-Deviation (RMSD) and the Root-Mean-Square-Fluctuations (RMSF) (139)

The ligand-receptor interaction was further analyzed by RMSD of both ligand (multi-epitope vaccine) and receptor (TLR-4), which reflects the structural stability of the complex. A plot of RMSD against time points out a stability of the molecule close to 1.1 after 30 ns of simulation (Figure 7 C). This value demonstrates the stability of complex system at time interval tested. Meanwhile, to evaluate fluctuation of amino acid side chains, the RMSF was observed (Figure 7 D). The figure shows low fluctuations of RMSF values around 0.2 nm showing stable regions and higher peaks with RMSF value of 1.2 nm indicating flexible regions (loops) in the system.

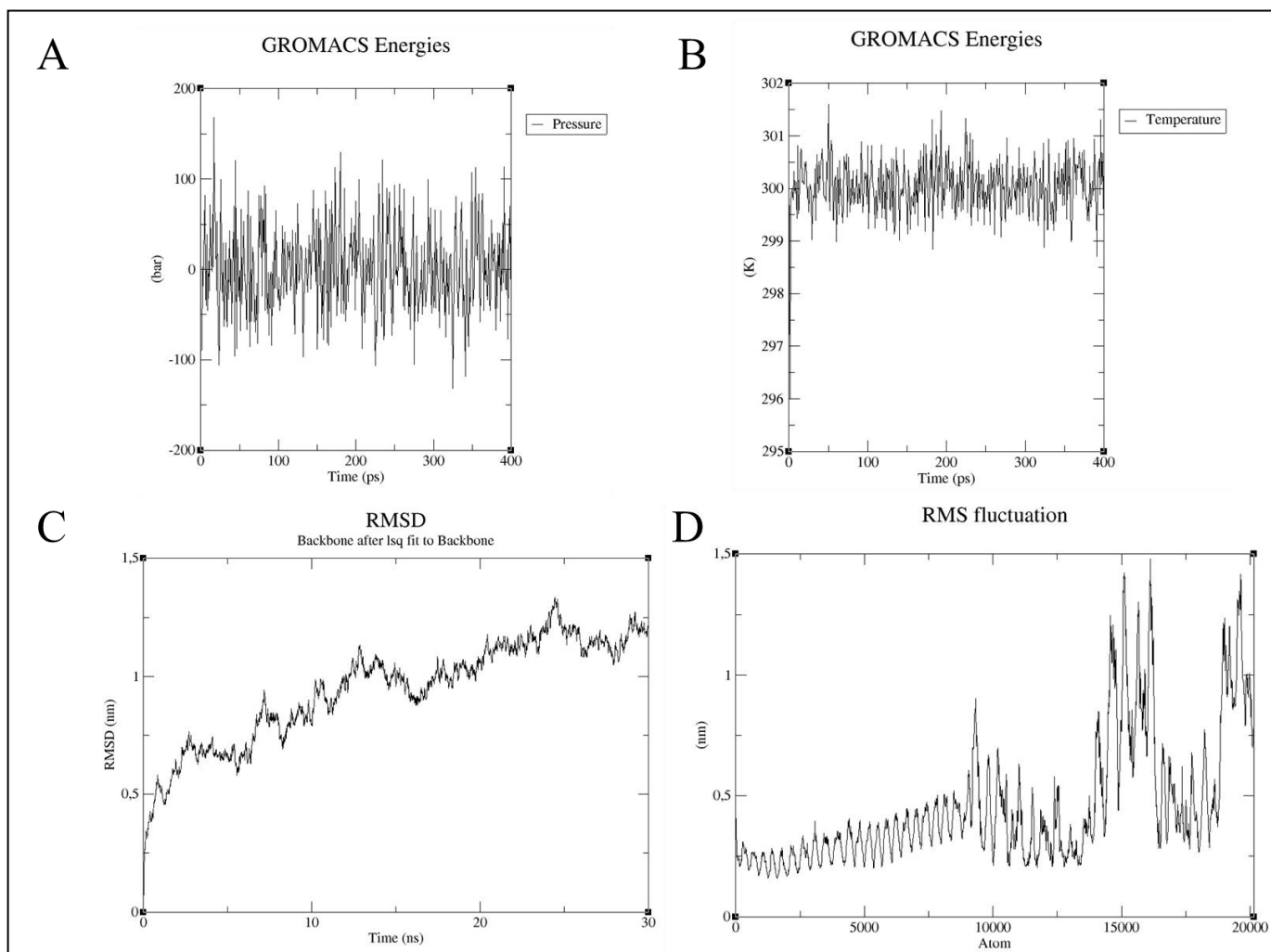


Figure 7: Molecular dynamics simulation of the ligand-receptor complex. (A) Docked complex pressure progression plot indicates fluctuation of pressure throughout the equilibration phase of 400 ps with an average pressure value around 1 bar. (B) Temperature progression plot of ligand-receptor complex shows that temperature of the system reaches to 300K and remains nearly constant around 300K throughout equilibration phase (400 ps). (C) Root Mean Square Deviation of docked complex shows an stable interaction of 1.1 between ligand and receptor molecule after 30 ns of simulation,. (D) Root Mean Square Fluctuation plot of docked complex side chain display low fluctuation (0.02 nm) that reflects the flexibility of side chain of docked complex.

3.10. *In silico* cloning

The cloning and expression of the multi-epitope vaccine within *E. coli* K12 were evaluated by JCAST and SnapGene. JCAT online server reverse translated the vaccine protein sequence and generated an optimized DNA sequence. The calculated GC content was 52.98% with a CAI value of 1.0. Afterwards, a cloning strategy using pET28a(+) and the restriction enzymes *Bam*HI and *Eco*RI were simulated via the SnapGene

software (Figure 8). The vector presents an N-terminally 6xHis-tagged configuration with a thrombin site, a C-terminal His tagged sequence, and f1 origin being ideal for *E. coli* cloning strategies. This conceived vector was customized for upcoming *in vitro* and *in vivo* experiments in *E. coli* to evaluate the vaccine properties and effects against the life-threatening bacteria *C. difficile*.

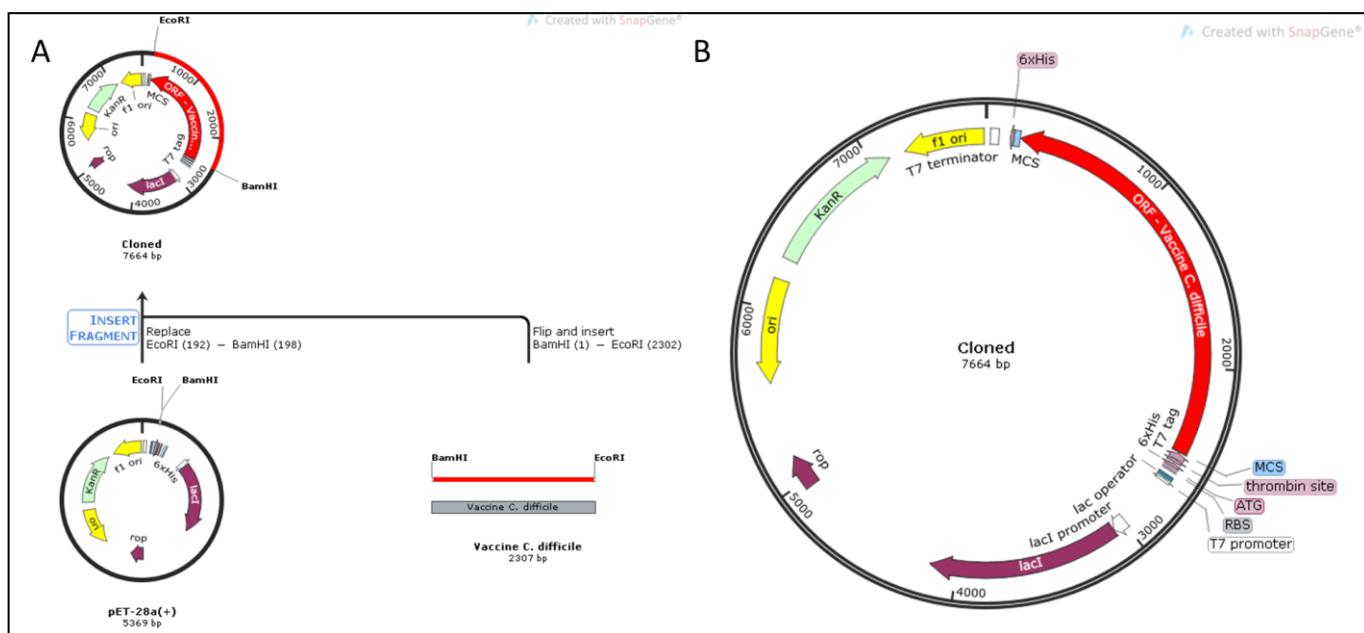


Figure 8: *In silico* cloning via SnapGene (+). SnapGene software was used to simulate a cloning strategy in *E. coli*. (A) The optimized *C. difficile* vaccine DNA sequence was inserted into the pET28a(+) vector via the restriction enzymes *Bam*HI and *Eco*RI. (B) The expected map sequence of the pET-28a(+) *C. difficile* vaccine plasmid.

3.11. Drug target identification, Molecular Modeling and prioritizations of identified drugs

Cytoplasmic proteins are involved in many vital metabolic functions; hence, they are very important for the physiology of bacteria. The key role of cytoplasmic proteins in maintenance of cell sustainability makes them more promising as drug targets (PMID: 20949199). Out of 250 cytoplasmic proteins we found 169 proteins present in G2 group ($G2 = E\text{-value} \leq 10e^{-5}$ and $\text{Identity} \geq 25\%$ AND $\text{LVI} \leq 0.7$) of MHOLline. Sixty-two proteins (Very High: 1, High: 19, Medium to Good: 1 and Good: 41) were identified in the first four distinct quality model groups of G2. Furthermore, the genome plasticity of all 53 *C. difficile* strains

was determined by using GIPSy (Genomic Island Prediction Software) (87). The software BRIG (BLAST Ring Image Generator) (140) was used for the circular genome comparison visualization. As result, we found differences in the presence of six pathogenicity islands (PAIs) and four genomic islands (GIs) (Figure 9) in *C. difficile*. Additionally, out of these 62 proteins only the proteins that were present in any GIs/PAIs were selected as candidate drug targets. As result, four proteins were considered for the drug target prioritization and docking studies (Table 5).

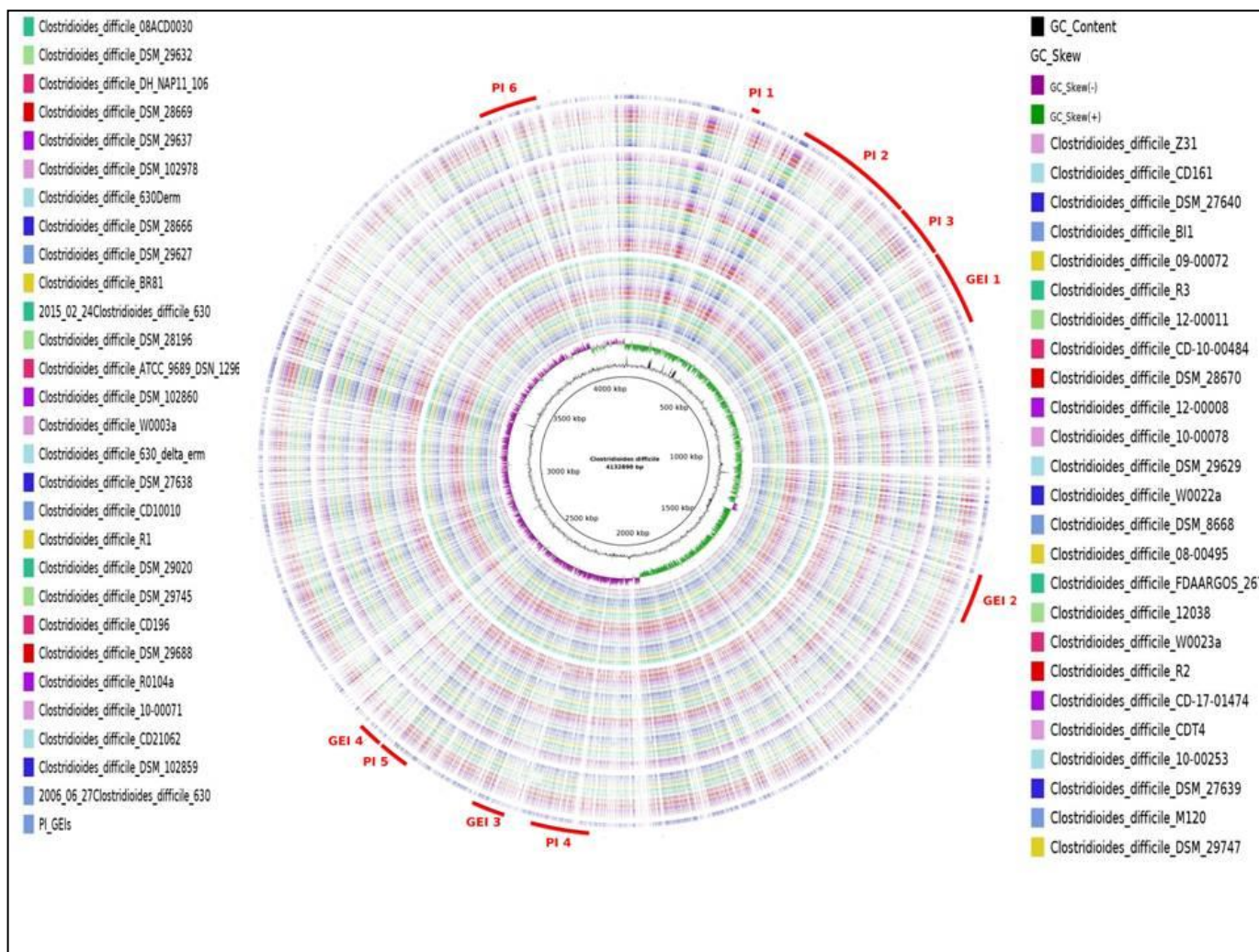


Figure 9: Comparative genomic maps of all 53 genomes of *C. difficile*. All the strains were aligned using *C. difficile* 630 as a reference. The figure represents the CDS, coding sequences; GIs (Genomics islands) and PAI (Pathogenicity Island).

Protein ID	Official Full Name	Mol. Wt (KDa)	Functions	Cellular Component	Virulence
AKP41517	Glyoxalase-like domain protein	13.91	Lactoylglutathione lyase activity, metal ion binding	Cytoplasmic	Yes
AKP41398	NimC/NimA family protein	14.77	Cofactor binding	Cytoplasmic	Yes
AKP41559	PTS system transporter subunit IIA	14.86	Phosphoenolpyruvate-dependent sugar phosphotransferase system.	Cytoplasmic	Yes
AKP44557	Alpha glucosidase, family 31 of glycosyl hydrolase	33.40	Hydrolase activity	Cytoplasmic	Yes

Table 5: The list of final four Drug targets with the prioritization parameters and functional annotation. The cytoplasmic proteins selected as putative drug targets and encoded in genomic or pathogenic island were selected for the docking analysis.

4. Conclusion

This work contributes to better understand the biology and to subsidize methods for the control of *C. difficile*. The use of comparative genomics together with reverse vaccinology and subtractive genomics identified and characterized new genes that are important in pathogen-host interactions, and contributed to the understanding of pathogenicity and host preference mechanisms. The use of reverse vaccinology and prediction of drug targets reduce the number of targets to be validated *in vitro* and *in vivo*, which reduce the time and cost for the development of new vaccines and drugs against *C. difficile*. The construction of a multi-epitope based vaccine and the identification of probable targets for drugs will be a valuable resource for the control and eradication of diseases caused by this deadly pathogen, which will directly improve public health.

ACKNOWLEDGEMENTS

We acknowledge the collaboration and assistance of all team members and the Brazilian funding agencies CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil).

AUTHOR CONTRIBUTIONS

MS SFOT ST conceived, designed the protocol, MS SFOT ST collected and analyzed data, MS, SFOT, RK, wrote the paper, MS, ST AM coordinated and led the entire project: MS, ST, SFOT, AKJ Cross-checked all data, analysis, ST, SCS, LCJA, VA, AM Review & Editing the paper: All authors read and approved the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest

References

1. Hall IC. Intestinal Flora in New-Born Infants. *Am J Dis Child*. 1935;49(2).
2. Kuijper EJ, Coignard B, Tüll P, ESCMID Study Group for *Clostridium difficile*, EU Member States, European Centre for Disease Prevention and Control. Emergence of *Clostridium difficile*-associated disease in North America and Europe. *Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis*. outubro de 2006;12 Suppl 6:2–18.
3. Smits WK, Lyras D, Lacy DB, Wilcox MH, Kuijper EJ. *Clostridium difficile* infection. *Nat Rev Dis Primer* [Internet]. 7 de abril de 2016 [citado 23 de janeiro de 2019];2:16020. Disponível em: <https://www.nature.com/articles/nrdp201620>
4. Lawson PA, Citron DM, Tyrrell KL, Finegold SM. Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O’Toole 1935) Prévot 1938. *Anaerobe*. agosto de 2016;40:95–9.
5. Crobach MJT, Vernon JJ, Loo VG, Kong LY, Péchiné S, Wilcox MH, et al. Understanding *Clostridium difficile* Colonization. *Clin Microbiol Rev* [Internet]. 1º de abril de 2018 [citado 23 de janeiro de 2019];31(2):e00021-17. Disponível em: <https://cmr.asm.org/content/31/2/e00021-17>
6. Ooijselaar RE, van Beurden YH, Terveer EM, Goorhuis A, Bauer MP, Keller JJ, et al. Update of treatment algorithms for *Clostridium difficile* infection. *Clin Microbiol Infect* [Internet]. maio de 2018;24(5):452–62. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/29309934>
7. Carter GP, Chakravorty A, Pham Nguyen TA, Mileto S, Schreiber F, Li L, et al. Defining the Roles of TcdA and TcdB in Localized Gastrointestinal Disease, Systemic Organ Damage, and the Host Response during *Clostridium difficile* Infections. *mBio*. 2 de junho de 2015;6(3):e00551.

8. Kuehne SA, Collery MM, Kelly ML, Cartman ST, Cockayne A, Minton NP. Importance of Toxin A, Toxin B, and CDT in Virulence of an Epidemic *Clostridium difficile* Strain. *J Infect Dis* [Internet]. 1º de janeiro de 2014 [citado 7 de fevereiro de 2019];209(1):83–6. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3864386/>
9. Shen A. *Clostridium difficile* toxins: mediators of inflammation. *J Innate Immun*. 2012;4(2):149–58.
10. Barbut F. Clinical features of *Clostridium difficile*-associated diarrhoea due to binary toxin (actin-specific ADP-ribosyltransferase)-producing strains. *J Med Microbiol* [Internet]. 1º de fevereiro de 2005 [citado 15 de junho de 2019];54(2):181–5. Disponível em: <http://jmm.microbiologyresearch.org/content/journal/jmm/10.1099/jmm.0.45804-0>
11. Carlucci C, Jones CS, Oliphant K, Yen S, Daigneault M, Carrier C, et al. Effects of defined gut microbial ecosystem components on virulence determinants of *Clostridioides difficile*. *Sci Rep* [Internet]. 29 de janeiro de 2019 [citado 24 de maio de 2019];9(1):885. Disponível em: <https://www.nature.com/articles/s41598-018-37547-x>
12. Cheng VCC, Yam WC, Lam OTC, Tsang JLY, Tse EYF, Siu GKH, et al. *Clostridium difficile* isolates with increased sporulation: emergence of PCR ribotype 002 in Hong Kong. *Eur J Clin Microbiol Infect Dis* [Internet]. novembro de 2011 [citado 15 de junho de 2019];30(11):1371. Disponível em: <http://link.springer.com/10.1007/s10096-011-1231-0>
13. Stevenson E, Minton NP, Kuehne SA. The role of flagella in *Clostridium difficile* pathogenicity. *Trends Microbiol* [Internet]. maio de 2015 [citado 15 de junho de 2019];23(5):275–82. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0966842X15000050>
14. Surawicz CM, Brandt LJ, Binion DG, Ananthakrishnan AN, Curry SR, Gilligan PH, et al. Guidelines for Diagnosis, Treatment, and Prevention of *Clostridium difficile* Infections. *Am J Gastroenterol* [Internet]. abril de 2013 [citado 23 de janeiro de 2019];108(4):478–98. Disponível em: <https://www.nature.com/articles/ajg20134>
15. McDonald LC. Effects of short- and long-course antibiotics on the lower intestinal microbiome as they relate to traveller’s diarrhea. *J Travel Med* [Internet]. abril de 2017 [citado 19 de julho de 2018];24(suppl_1):S35–8. Disponível em: <https://academic.oup.com/jtm/article-lookup/doi/10.1093/jtm/taw084>
16. Koenigskecht MJ, Theriot CM, Bergin IL, Schumacher CA, Schloss PD, Young VB. Dynamics and establishment of *Clostridium difficile* infection in the murine gastrointestinal tract. *Infect Immun*. março de 2015;83(3):934–41.
17. Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, Dunn JR, et al. Burden of *Clostridium difficile* infection in the United States. *N Engl J Med*. 26 de fevereiro de 2015;372(9):825–34.
18. McDonald LC, Gerding DN, Johnson S, Bakken JS, Carroll KC, Coffin SE, et al. Clinical Practice Guidelines for *Clostridium difficile* Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clin Infect Dis* [Internet]. 19 de março de 2018 [citado 31 de maio de 2019];66(7):e1–48. Disponível em: <https://academic.oup.com/cid/article/66/7/e1/4855916>
19. Richard Lawley LC Judy Davis. *The Food Safety Hazard Guidebook*. 2008.

20. Rodriguez-Palacios A, Lejeune JT. Moist-heat resistance, spore aging, and superdormancy in *Clostridium difficile*. *Appl Environ Microbiol*. maio de 2011;77(9):3085–91.
21. Sorg JA, Sonenshein AL. Bile salts and glycine as cogerminants for *Clostridium difficile* spores. *J Bacteriol*. abril de 2008;190(7):2505–12.
22. Paredes-Sabja D, Shen A, Sorg JA. *Clostridium difficile* spore biology: sporulation, germination, and spore structural proteins. *Trends Microbiol*. julho de 2014;22(7):406–16.
23. Bhattacharjee D, Francis MB, Ding X, McAllister KN, Shrestha R, Sorg JA. Reexamining the Germination Phenotypes of Several *Clostridium difficile* Strains Suggests Another Role for the CspC Germinant Receptor. *J Bacteriol* [Internet]. 14 de dezembro de 2015;198(5):777–86. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/26668265>
24. Theriot CM, Young VB. Interactions Between the Gastrointestinal Microbiome and *Clostridium difficile*. *Annu Rev Microbiol*. 2015;69:445–61.
25. Wahlström A, Sayin SI, Marschall H-U, Bäckhed F. Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. *Cell Metab* [Internet]. 12 de julho de 2016 [citado 3 de maio de 2017];24(1):41–50. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1550413116302236>
26. Ramirez-Perez O, Cruz-Ramon V, Chinchilla-Lopez P, Mendez-Sanchez N. The Role of the Gut Microbiota in Bile Acid Metabolism. *Ann Hepatol* [Internet]. novembro de 2017;16(Suppl. 1: s3-105.):s15–20. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/29080339>
27. Antharam VC, Li EC, Ishmael A, Sharma A, Mai V, Rand KH, et al. Intestinal Dysbiosis and Depletion of Butyrogenic Bacteria in *Clostridium difficile* Infection and Nosocomial Diarrhea. *J Clin Microbiol* [Internet]. 1º de setembro de 2013 [citado 7 de fevereiro de 2019];51(9):2884–92. Disponível em: <https://jcm.asm.org/content/51/9/2884>
28. Vincent C, Stephens DA, Loo VG, Edens TJ, Behr MA, Dewar K, et al. Reductions in intestinal Clostridiales precede the development of nosocomial *Clostridium difficile* infection. *Microbiome* [Internet]. 28 de junho de 2013 [citado 7 de fevereiro de 2019];1(1):18. Disponível em: <https://doi.org/10.1186/2049-2618-1-18>
29. Theriot CM, Koenigskecht MJ, Carlson Jr PE, Hatton GE, Nelson AM, Li B, et al. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. *Nat Commun* [Internet]. 2014 [citado 15 de dezembro de 2014];5. Disponível em: <http://www.nature.com/ncomms/2014/140120/ncomms4114/full/ncomms4114.html>
30. Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, et al. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature*. 8 de janeiro de 2015;517(7533):205–8.
31. Zhang L, Dong D, Jiang C, Li Z, Wang X, Peng Y. Insight into alteration of gut microbiota in *Clostridium difficile* infection and asymptomatic *C. difficile* colonization. *Anaerobe* [Internet]. agosto de 2015;34:1–7. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/25817005>
32. Gu S, Chen Y, Zhang X, Lu H, Lv T, Shen P, et al. Identification of key taxa that favor intestinal colonization of *Clostridium difficile* in an adult Chinese population. *Microbes Infect* [Internet]. 1º de janeiro de

2016 [citado 7 de fevereiro de 2019];18(1):30–8. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1286457915001914>

33. Theriot CM, Bowman AA, Young VB. Antibiotic-Induced Alterations of the Gut Microbiota Alter Secondary Bile Acid Production and Allow for *Clostridium difficile* Spore Germination and Outgrowth in the Large Intestine. *mSphere*. fevereiro de 2016;1(1).

34. Reeves AE, Koenigskecht MJ, Bergin IL, Young VB. Suppression of *Clostridium difficile* in the Gastrointestinal Tracts of Germfree Mice Inoculated with a Murine Isolate from the Family Lachnospiraceae. *Infect Immun* [Internet]. 1º de novembro de 2012 [citado 7 de fevereiro de 2019];80(11):3786–94. Disponível em: <https://iai.asm.org/content/80/11/3786>

35. Schubert AM, Rogers MAM, Ring C, Mogle J, Petrosino JP, Young VB, et al. Microbiome Data Distinguish Patients with *Clostridium difficile* Infection and Non-*C. difficile*-Associated Diarrhea from Healthy Controls. *mBio* [Internet]. 1º de julho de 2014 [citado 7 de fevereiro de 2019];5(3):e01021-14. Disponível em: <https://mbio.asm.org/content/5/3/e01021-14>

36. Petrof EO, Gloor GB, Vanner SJ, Weese SJ, Carter D, Daigneault MC, et al. Stool substitute transplant therapy for the eradication of *Clostridium difficile* infection: “RePOOPulating” the gut. *Microbiome* [Internet]. 9 de janeiro de 2013;1(1):3. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/24467987>

37. Sokol H, Galperine T, Kapel N, Bourlioux P, Seksik P, Barbut F, et al. Faecal microbiota transplantation in recurrent *Clostridium difficile* infection: Recommendations from the French Group of Faecal microbiota Transplantation. *Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver*. março de 2016;48(3):242–7.

38. Weingarden AR, Dosa PI, DeWinter E, Steer CJ, Shaughnessy MK, Johnson JR, et al. Changes in Colonic Bile Acid Composition following Fecal Microbiota Transplantation Are Sufficient to Control *Clostridium difficile* Germination and Growth. *PLOS ONE* [Internet]. 20 de janeiro de 2016 [citado 7 de fevereiro de 2019];11(1):e0147210. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0147210>

39. Abougergi MS, Kwon JH. Intravenous immunoglobulin for the treatment of *Clostridium difficile* infection: a review. *Dig Dis Sci*. janeiro de 2011;56(1):19–26.

40. Bézay N, Ayad A, Dubischar K, Firbas C, Hochreiter R, Kiermayr S, et al. Safety, immunogenicity and dose response of VLA84, a new vaccine candidate against *Clostridium difficile*, in healthy volunteers. *Vaccine*. 17 de 2016;34(23):2585–92.

41. de Bruyn G, Saleh J, Workman D, Pollak R, Elinoff V, Fraser NJ, et al. Defining the optimal formulation and schedule of a candidate toxoid vaccine against *Clostridium difficile* infection: A randomized Phase 2 clinical trial. *Vaccine*. 27 de abril de 2016;34(19):2170–8.

42. Gil F, Calderón IL, Fuentes JA, Paredes-Sabja D. Clostridioides (*Clostridium*) *difficile* infection: current and alternative therapeutic strategies. *Future Microbiol* [Internet]. 21 de fevereiro de 2018 [citado 24 de maio de 2019];13(4):469–82. Disponível em: <https://www.futuremedicine.com/doi/full/10.2217/fmb-2017-0203>

43. Sheldon E, Kitchin N, Peng Y, Eiden J, Gruber W, Johnson E, et al. A phase 1, placebo-controlled, randomized study of the safety, tolerability, and immunogenicity of a *Clostridium difficile* vaccine administered with or without aluminum hydroxide in healthy adults. *Vaccine*. 19 de abril de 2016;34(18):2082–91.

44. Sebahia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, et al. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Genet.* julho de 2006;38(7):779–86.
45. Amy J, Johanesen P, Lyras D. Extrachromosomal and integrated genetic elements in *Clostridium difficile*. *Plasmid.* julho de 2015;80:97–110.
46. He M, Sebahia M, Lawley TD, Stabler RA, Dawson LF, Martin MJ, et al. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc Natl Acad Sci [Internet].* 20 de abril de 2010 [citado 12 de junho de 2019];107(16):7527–32. Disponível em: <https://www.pnas.org/content/107/16/7527>
47. Janezic S, Rupnik M. Genomic diversity of *Clostridium difficile* strains. *Res Microbiol [Internet].* 1º de maio de 2015 [citado 3 de junho de 2019];166(4):353–60. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0923250815000303>
48. Knight DR, Elliott B, Chang BJ, Perkins TT, Riley TV. Diversity and Evolution in the Genome of *Clostridium difficile*. *Clin Microbiol Rev [Internet].* 1º de julho de 2015 [citado 1º de junho de 2019];28(3):721–41. Disponível em: <https://cmr.asm.org/content/28/3/721>
49. Scaria J, Ponnala L, Janvilisri T, Yan W, Mueller LA, Chang Y-F. Analysis of Ultra Low Genome Conservation in *Clostridium difficile*. Horsburgh MJ, organizador. *PLoS ONE [Internet].* 8 de dezembro de 2010 [citado 12 de junho de 2019];5(12):e15147. Disponível em: <https://dx.plos.org/10.1371/journal.pone.0015147>
50. CDC. The biggest antibiotic-resistant threats in the U.S. [Internet]. Centers for Disease Control and Prevention. 2018 [citado 7 de fevereiro de 2019]. Disponível em: https://www.cdc.gov/drugresistance/biggest_threats.html
51. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res [Internet].* 8 de janeiro de 2019 [citado 18 de março de 2019];47(D1):D506–15. Disponível em: <https://academic.oup.com/nar/article/47/D1/D506/5160987>
52. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol [Internet].* 6 de agosto de 2015;16:157. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/26243257>
53. Barinov A, Loux V, Hammani A, Nicolas P, Langella P, Ehrlich D, et al. Prediction of surface exposed proteins in *Streptococcus pyogenes*, with a potential application to other Gram-positive bacteria. *Proteomics [Internet].* janeiro de 2009;9(1):61–73. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/19053137>
54. Shende G, Haldankar H, Barai RS, Bharmal MH, Shetty V, Idicula-Thomas S. PBIT: Pipeline Builder for Identification of drug Targets for infectious diseases. *Bioinformatics [Internet].* 15 de março de 2017;33(6):929–31. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/28039165>
55. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* 8 de janeiro de 2019;47(D1):D339–43.
56. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins.* 1º de outubro de 2006;65(1):40–8.

57. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods* [Internet]. janeiro de 2015 [citado 21 de março de 2019];12(1):7. Disponível em: <https://www.nature.com/articles/nmeth.3213>
58. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* [Internet]. dezembro de 2008 [citado 21 de março de 2019];9(1). Disponível em: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-40>
59. Heo L, Park H, Seok C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* julho de 2013;41(Web Server issue):W384-388.
60. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* [Internet]. 1º de abril de 1993 [citado 21 de março de 2019];26(2):283-91. Disponível em: [//scripts.iucr.org/cgi-bin/paper?gl0276](https://scripts.iucr.org/cgi-bin/paper?gl0276)
61. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins.* abril de 1992;12(4):345-64.
62. Ponomarenko J, Bui H-H, Li W, Fusseder N, Bourne PE, Sette A, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* [Internet]. 2 de dezembro de 2008 [citado 7 de abril de 2019];9:514. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2607291/>
63. Dhanda SK, Vir P, Raghava GP. Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct* [Internet]. 5 de dezembro de 2013 [citado 25 de março de 2019];8(1):30. Disponível em: <https://doi.org/10.1186/1745-6150-8-30>
64. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v.2--a server for in silico prediction of allergens. *J Mol Model.* junho de 2014;20(6):2278.
65. Saha S, Raghava GPS. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res* [Internet]. 1º de julho de 2006 [citado 25 de março de 2019];34(suppl_2):W202-9. Disponível em: https://academic.oup.com/nar/article/34/suppl_2/W202/2505748
66. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* [Internet]. 5 de janeiro de 2007 [citado 25 de março de 2019];8(1):4. Disponível em: <https://doi.org/10.1186/1471-2105-8-4>
67. Cheng J, Randall AZ, Sweredoski MJ, Baldi P. SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res* [Internet]. 1º de julho de 2005 [citado 25 de março de 2019];33(suppl_2):W72-6. Disponível em: https://academic.oup.com/nar/article/33/suppl_2/W72/2505518
68. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, organizador. *The Proteomics Protocols Handbook* [Internet]. Totowa, NJ: Humana Press; 2005 [citado 25 de março de 2019]. p. 571-607. Disponível em: <http://link.springer.com/10.1385/1-59259-890-0:571>
69. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera?A visualization system for exploratory research and analysis. *J Comput Chem* [Internet]. outubro de 2004 [citado 5 de junho de 2019];25(13):1605-12. Disponível em: <http://doi.wiley.com/10.1002/jcc.20084>

70. Torchala M, Moal IH, Chaleil RAG, Fernandez-Recio J, Bates PA. SwarmDock: a server for flexible protein–protein docking. *Bioinformatics* [Internet]. 15 de março de 2013 [citado 1º de maio de 2019];29(6):807–9. Disponível em: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btt038>
71. Torchala M, Bates PA. Predicting the Structure of Protein–Protein Complexes Using the SwarmDock Web Server. In: Kihara D, organizador. *Protein Structure Prediction* [Internet]. New York, NY: Springer New York; 2014 [citado 1º de maio de 2019]. p. 181–97. Disponível em: http://link.springer.com/10.1007/978-1-4939-0366-5_13
72. Moal IH, Barradas-Bautista D, Jiménez-García B, Torchala M, van der Velde A, Vreven T, et al. IRaPPA: information retrieval based integration of biophysical models for protein assembly selection. Tramontano A, organizador. *Bioinformatics* [Internet]. 15 de junho de 2017 [citado 1º de maio de 2019];33(12):1806–13. Disponível em: <https://academic.oup.com/bioinformatics/article/33/12/1806/2995816>
73. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Biol* [Internet]. setembro de 2002 [citado 16 de setembro de 2019];9(9):646–52. Disponível em: <http://www.nature.com/doi/10.1038/nsb0902-646>
74. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* [Internet]. 1º de setembro de 2015 [citado 8 de junho de 2019];1–2:19–25. Disponível em: <http://www.sciencedirect.com/science/article/pii/S2352711015000059>
75. Hajighahramani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatabadi SS, Ghasemi Y. Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*. *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis*. 2017;48:83–94.
76. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*. junho de 2010;78(8):1950–8.
77. Grote A, Hiller K, Scheer M, Munch R, Nortemann B, Hempel DC, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* [Internet]. 1º de julho de 2005 [citado 13 de maio de 2019];33(Web Server):W526–31. Disponível em: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gki376>
78. Carbone A, Zinovyev A, Képès F. Codon adaptation index as a measure of dominating codon bias. *Bioinforma Oxf Engl*. 1º de novembro de 2003;19(16):2005–15.
79. Sharp PM, Li WH. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*. 11 de fevereiro de 1987;15(3):1281–95.
80. Capriles PV, Guimaraes AC, Otto TD, Miranda AB, Dardenne LE, Degraeve WM. Structural modelling and comparative analysis of homologous, analogous and specific proteins from *Trypanosoma cruzi* versus *Homo sapiens*: putative drug targets for chagas’ disease treatment. *BMC Genomics* [Internet]. 29 de outubro de 2010;11:610. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/21034488>
81. Hassan SS, Tiwari S, Guimaraes LC, Jamal SB, Folador E, Sharma NB, et al. Proteome scale comparative modeling for conserved drug and vaccine targets identification in *Corynebacterium pseudotuberculosis*. *BMC Genomics* [Internet]. 2014;15 Suppl 7:S3. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/25573232>

82. Jain N, Jaiswal J, Pathak A, Singour PK. Synthesis, Molecular Docking and Evaluation of 3-{4-[2-amino-4-(substitutedphenyl)-2H-[1, 3] oxazin/thiazin-6-yl] 2-phenyl-3H-quinazolin-4-one Derivatives for their Anticonvulsant Activity. *Cent Nerv Syst Agents Med Chem* [Internet]. 26 de janeiro de 2018;18(1):63–73. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/28056730>
83. Kumar Jaiswal A, Tiwari S, Jamal SB, Barh D, Azevedo V, Soares SC. An In Silico Identification of Common Putative Vaccine Candidates against *Treponema pallidum*: A Reverse Vaccinology and Subtractive Genomics Based Approach. *Int J Mol Sci* [Internet]. 14 de fevereiro de 2017;18(2). Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/28216574>
84. Barh D, Jain N, Tiwari S, Parida BP, D'Afonseca V, Li L, et al. A novel comparative genomics analysis for common drug and vaccine targets in *Corynebacterium pseudotuberculosis* and other CMN group of human pathogens. *Chem Biol Drug Des* [Internet]. julho de 2011;78(1):73–84. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/21443692>
85. Jamal SB, Hassan SS, Tiwari S, Viana MV, Benevides L de J, Ullah A, et al. An integrative in-silico approach for therapeutic target identification in the human pathogen *Corynebacterium diphtheriae*. de Brevern AG, organizador. *PLOS ONE* [Internet]. 19 de outubro de 2017 [citado 16 de novembro de 2018];12(10):e0186401. Disponível em: <http://dx.plos.org/10.1371/journal.pone.0186401>
86. Hassan SS, Jamal SB, Radusky LG, Tiwari S, Ullah A, Ali J, et al. The Druggable Pocketome of *Corynebacterium diphtheriae*: A New Approach for in silico Putative Druggable Targets. *Front Genet* [Internet]. 13 de fevereiro de 2018 [citado 2 de maio de 2019];9:44. Disponível em: <http://journal.frontiersin.org/article/10.3389/fgene.2018.00044/full>
87. Soares SC, Geyik H, Ramos RT, de Sa PH, Barbosa EG, Baumbach J, et al. GIPSY: Genomic island prediction software. *J Biotechnol* [Internet]. 20 de agosto de 2016;232:2–11. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/26376473>
88. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* [Internet]. 1º de janeiro de 2000;28(1):27–30. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/10592173>
89. Chen M-J, Tang H-Y, Chiang M-L. Effects of heat, cold, acid and bile salt adaptations on the stress tolerance and protein expression of kefir-isolated probiotic *Lactobacillus kefirifaciens* M1. *Food Microbiol* [Internet]. setembro de 2017 [citado 16 de maio de 2017];66:20–7. Disponível em: <http://www.sciencedirect.com/science/article/pii/S074000201630956X>
90. Laxminarayan R, Matsoso P, Pant S, Brower C, Røttingen J-A, Klugman K, et al. Access to effective antimicrobials: a worldwide challenge. *Lancet Lond Engl*. 9 de janeiro de 2016;387(10014):168–75.
91. Skwarczynski M, Toth I. Peptide-based synthetic vaccines. *Chem Sci* [Internet]. 2016 [citado 6 de abril de 2019];7(2):842–54. Disponível em: <http://xlink.rsc.org/?DOI=C5SC03892H>
92. Bakheet TM, Doig AJ. Properties and identification of antibiotic drug targets. *BMC Bioinformatics*. 20 de abril de 2010;11:195.
93. Barh D, Tiwari S, Jain N, Ali A, Santos AR, Misra AN, et al. In silico subtractive genomics for target identification in human bacterial pathogens. *Drug Dev Res*. 2011;72(2):162–77.

94. Cowardin CA, Buonomo EL, Saleh MM, Wilson MG, Burgess SL, Kuehne SA, et al. The binary toxin CDT enhances *Clostridium difficile* virulence by suppressing protective colonic eosinophilia. *Nat Microbiol*. 11 de 2016;1(8):16108.
95. Schwan C, Kruppke AS, Nölke T, Schumacher L, Koch-Nolte F, Kudryashev M, et al. *Clostridium difficile* toxin CDT hijacks microtubule organization and reroutes vesicle traffic to increase pathogen adherence. *Proc Natl Acad Sci U S A* [Internet]. 11 de fevereiro de 2014 [citado 1º de junho de 2019];111(6):2313–8. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3926047/>
96. Lanis JM, Heinlen LD, James JA, Ballard JD. *Clostridium difficile* 027/BI/NAP1 Encodes a Hypertoxic and Antigenically Variable Form of TcdB. Bradley KA, organizador. *PLoS Pathog* [Internet]. 1º de agosto de 2013 [citado 1º de junho de 2019];9(8):e1003523. Disponível em: <http://dx.plos.org/10.1371/journal.ppat.1003523>
97. Stabler RA, Gerding DN, Songer JG, Drudy D, Brazier JS, Trinh HT, et al. Comparative Phylogenomics of *Clostridium difficile* Reveals Clade Specificity and Microevolution of Hypervirulent Strains. *J Bacteriol* [Internet]. 15 de outubro de 2006 [citado 12 de junho de 2019];188(20):7297–305. Disponível em: <http://jb.asm.org/cgi/doi/10.1128/JB.00664-06>
98. McIntosh EDG. Healthcare-associated infections: potential for prevention through vaccination. *Ther Adv Vaccines Immunother* [Internet]. 1º de fevereiro de 2018 [citado 3 de junho de 2019];6(1):19–27. Disponível em: <https://doi.org/10.1177/2515135518763183>
99. Wright A, Drudy D, Kyne L, Brown K, Fairweather NF. Immunoreactive cell wall proteins of *Clostridium difficile* identified by human sera. *J Med Microbiol* [Internet]. 1º de junho de 2008 [citado 21 de junho de 2019];57(6):750–6. Disponível em: <http://jmm.microbiologyresearch.org/content/journal/jmm/10.1099/jmm.0.47532-0>
100. Ausiello CM, Cerquetti M, Fedele G, Spensieri F, Palazzo R, Nasso M, et al. Surface layer proteins from *Clostridium difficile* induce inflammatory and regulatory cytokines in human monocytes and dendritic cells. *Microbes Infect*. setembro de 2006;8(11):2640–6.
101. Calabi E, Calabi F, Phillips AD, Fairweather NF. Binding of *Clostridium difficile* surface layer proteins to gastrointestinal tissues. *Infect Immun*. outubro de 2002;70(10):5770–8.
102. Ryan A, Lynch M, Smith SM, Amu S, Nel HJ, McCoy CE, et al. A Role for TLR4 in *Clostridium difficile* Infection and the Recognition of Surface Layer Proteins. Schneider DS, organizador. *PLoS Pathog* [Internet]. 30 de junho de 2011 [citado 12 de junho de 2019];7(6):e1002076. Disponível em: <http://dx.plos.org/10.1371/journal.ppat.1002076>
103. Cerquetti M, Molinari A, Sebastianelli A, Diociaiuti M, Petruzzelli R, Capo C, et al. Characterization of surface layer proteins from different *Clostridium difficile* clinical isolates. *Microb Pathog*. junho de 2000;28(6):363–72.
104. Calabi E, Fairweather N. Patterns of sequence conservation in the S-Layer proteins and related sequences in *Clostridium difficile*. *J Bacteriol*. julho de 2002;184(14):3886–97.
105. Bruxelle J-F, Mizrahi A, Hoys S, Collignon A, Janoir C, Péchiné S. Immunogenic properties of the surface layer precursor of *Clostridium difficile* and vaccination assays in animal models. *Anaerobe* [Internet]. 1º de

fevereiro de 2016 [citado 19 de junho de 2019];37:78–84. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1075996415300706>

106. Vedantam G, Kochanowsky J, Lindsey J, Mallozzi M, Roxas JL, Adamson C, et al. An Engineered Synthetic Biologic Protects Against *Clostridium difficile* Infection. *Front Microbiol* [Internet]. 2018 [citado 19 de junho de 2019];9. Disponível em: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.02080/full>

107. Mizrahi A, Bruxelles J-F, Péchiné S, Le Monnier A. Prospective evaluation of the adaptive immune response to SlpA in *Clostridium difficile* infection. *Anaerobe* [Internet]. 1º de dezembro de 2018 [citado 19 de junho de 2019];54:164–8. Disponível em: <http://www.sciencedirect.com/science/article/pii/S107599641830163X>

108. Ní Eidhin DB, O’Brien JB, McCabe MS, Athié-Morales V, Kelleher DP. Active immunization of hamsters against *Clostridium difficile* infection using surface-layer protein. *FEMS Immunol Med Microbiol* [Internet]. 1º de março de 2008 [citado 21 de junho de 2019];52(2):207–18. Disponível em: <https://academic.oup.com/femspd/article/52/2/207/552073>

109. Péchiné S, Bruxelles JF, Janoir C, Collignon A. Targeting *Clostridium difficile* Surface Components to Develop Immunotherapeutic Strategies Against *Clostridium difficile* Infection. *Front Microbiol* [Internet]. 23 de maio de 2018 [citado 19 de junho de 2019];9. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5974105/>

110. Hong HA, Ferreira WT, Hosseini S, Anwar S, Hitri K, Wilkinson AJ, et al. The Spore Coat Protein CotE Facilitates Host Colonization by *Clostridium difficile*. *J Infect Dis*. 12 de 2017;216(11):1452–9.

111. Permpoonpattana P, Phetcharaburanin J, Mikelsone A, Dembek M, Tan S, Brisson M-C, et al. Functional Characterization of *Clostridium difficile* Spore Coat Proteins. *J Bacteriol* [Internet]. abril de 2013 [citado 18 de junho de 2019];195(7):1492–503. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624542/>

112. Chastanet A, Losick R. Engulfment during sporulation in *Bacillus subtilis* is governed by a multi-protein complex containing tandemly acting autolysins. *Mol Microbiol* [Internet]. 2007 [citado 19 de junho de 2019];64(1):139–52. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2007.05652.x>

113. Morlot C, Uehara T, Marquis KA, Bernhardt TG, Rudner DZ. A highly coordinated cell wall degradation machine governs spore morphogenesis in *Bacillus subtilis*. *Genes Dev*. 15 de fevereiro de 2010;24(4):411–22.

114. Dembek M, Barquist L, Boinett CJ, Cain AK, Mayho M, Lawley TD, et al. High-Throughput Analysis of Gene Essentiality and Sporulation in *Clostridium difficile*. *mBio* [Internet]. 1º de maio de 2015 [citado 19 de junho de 2019];6(2):e02383-14. Disponível em: <https://mbio.asm.org/content/6/2/e02383-14>

115. Saujet L, Pereira FC, Serrano M, Soutourina O, Monot M, Shelyakin PV, et al. Genome-Wide Analysis of Cell Type-Specific Gene Transcription during Spore Formation in *Clostridium difficile*. *PLOS Genet* [Internet]. 3 de outubro de 2013 [citado 19 de junho de 2019];9(10):e1003756. Disponível em: <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1003756>

116. Freitas C, Plannic J, Isticato R, Pelosi A, Zilhao R, Serrano M, et al. A protein phosphorylation module patterns the *Bacillus subtilis* spore outer coat: [Internet]. *Microbiology*; 2018 nov [citado 19 de junho de 2019]. Disponível em: <http://biorxiv.org/lookup/doi/10.1101/469122>

117. Isticato R, Sirec T, Vecchione S, Crispino A, Saggese A, Baccigalupi L, et al. The Direct Interaction between Two Morphogenetic Proteins Is Essential for Spore Coat Formation in *Bacillus subtilis*. PLOS ONE [Internet]. 20 de outubro de 2015 [citado 19 de junho de 2019];10(10):e0141040. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0141040>
118. Charlton TM, Kovacs-Simon A, Michell SL, Fairweather NF, Tate EW. Quantitative Lipoproteomics in *Clostridium difficile* Reveals a Role for Lipoproteins in Sporulation. Chem Biol [Internet]. 19 de novembro de 2015 [citado 18 de junho de 2019];22(11):1562–73. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1074552115004081>
119. Kovacs-Simon A, Titball RW, Michell SL. Lipoproteins of Bacterial Pathogens. Infect Immun [Internet]. fevereiro de 2011 [citado 18 de junho de 2019];79(2):548–61. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3028857/>
120. Kovacs-Simon A, Leuzzi R, Kasendra M, Minton N, Titball RW, Michell SL. Lipoprotein CD0873 Is a Novel Adhesin of *Clostridium difficile*. J Infect Dis [Internet]. 15 de julho de 2014 [citado 7 de novembro de 2016];210(2):274–84. Disponível em: <http://jid.oxfordjournals.org/content/210/2/274>
121. Grode L, Kursar M, Fensterle J, Kaufmann SHE, Hess J. Cell-mediated immunity induced by recombinant *Mycobacterium bovis* Bacille Calmette-Guérin strains against an intracellular bacterial pathogen: importance of antigen secretion or membrane-targeted antigen display as lipoprotein for vaccine efficacy. J Immunol Baltim Md 1950. 15 de fevereiro de 2002;168(4):1869–76.
122. Seixas FK, Fernandes CH, Hartwig DD, Conceição FR, Aleixo JAG, Dellagostin OA. Evaluation of different ways of presenting LipL32 to the immune system with the aim of developing a recombinant vaccine against leptospirosis. Can J Microbiol. abril de 2007;53(4):472–9.
123. Edwards AN, Nawrocki KL, McBride SM. Conserved oligopeptide permeases modulate sporulation initiation in *Clostridium difficile*. Infect Immun. outubro de 2014;82(10):4276–91.
124. Xu Y, Guo J, Wang L, Jiang R, Jin X, Liu J, et al. The Crystal Structure of the YknZ Extracellular Domain of ABC Transporter YknWXYZ from *Bacillus amyloliquefaciens*. PLOS ONE [Internet]. 31 de maio de 2016 [citado 19 de junho de 2019];11(5):e0155846. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0155846>
125. Xu Y, Jo I, Wang L, Chen J, Fan S, Dong Y, et al. Hexameric assembly of membrane fusion protein YknX of the sporulation delaying efflux pump from *Bacillus amyloliquefaciens*. Biochem Biophys Res Commun [Internet]. 4 de novembro de 2017 [citado 19 de junho de 2019];493(1):152–7. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0006291X17318193>
126. Yamada Y, Tikhonova EB, Zgurskaya HI. YknWXYZ Is an Unusual Four-Component Transporter with a Role in Protection against Sporulation-Delaying-Protein-Induced Killing of *Bacillus subtilis*. J Bacteriol [Internet]. agosto de 2012 [citado 19 de junho de 2019];194(16):4386–94. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3416220/>
127. Zgurskaya HI, Yamada Y, Tikhonova EB, Ge Q, Krishnamoorthy G. Structural and functional diversity of bacterial membrane fusion proteins. Biochim Biophys Acta. maio de 2009;1794(5):794–807.

128. Pereira LE, Tsang J, Mrázek J, Hoover TR. The zinc-ribbon domain of *Helicobacter pylori* HP0958: requirement for RpoN accumulation and possible roles of homologs in other bacteria. *Microb Inform Exp* [Internet]. 23 de agosto de 2011 [citado 18 de junho de 2019];1(1):8. Disponível em: <https://doi.org/10.1186/2042-5783-1-8>
129. Lowy I, Molrine DC, Leav BA, Blair BM, Baxter R, Gerding DN, et al. Treatment with monoclonal antibodies against *Clostridium difficile* toxins. *N Engl J Med*. 21 de janeiro de 2010;362(3):197–205.
130. Johnston PF, Gerding DN, Knight KL. Protection from *Clostridium difficile* Infection in CD4 T Cell- and Polymeric Immunoglobulin Receptor-Deficient Mice. Morrison RP, organizador. *Infect Immun* [Internet]. fevereiro de 2014 [citado 15 de junho de 2019];82(2):522–31. Disponível em: <http://iai.asm.org/lookup/doi/10.1128/IAI.01273-13>
131. Kyne L, Warny M, Qamar A, Kelly CP. Asymptomatic carriage of *Clostridium difficile* and serum levels of IgG antibody against toxin A. *N Engl J Med*. 10 de fevereiro de 2000;342(6):390–7.
132. Abt MC, McKenney PT, Pamer EG. *Clostridium difficile* colitis: pathogenesis and host defence. *Nat Rev Microbiol* [Internet]. outubro de 2016 [citado 23 de janeiro de 2019];14(10):609–20. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5109054/>
133. Sonnenberg GF, Artis D. Innate lymphoid cells in the initiation, regulation and resolution of inflammation. *Nat Med*. julho de 2015;21(7):698–708.
134. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1º de setembro de 1997;25(17):3389–402.
135. Ohto U, Fukase K, Miyake K, Shimizu T. Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2. *Proc Natl Acad Sci* [Internet]. 8 de maio de 2012 [citado 10 de junho de 2019];109(19):7421–6. Disponível em: <https://www.pnas.org/content/109/19/7421>
136. Park BS, Lee J-O. Recognition of lipopolysaccharide pattern by TLR4 complexes. *Exp Mol Med* [Internet]. dezembro de 2013 [citado 10 de junho de 2019];45(12):e66. Disponível em: <https://www.nature.com/articles/emm201397>
137. Park BS, Song DH, Kim HM, Choi B-S, Lee H, Lee J-O. The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature* [Internet]. abril de 2009 [citado 10 de junho de 2019];458(7242):1191–5. Disponível em: <https://www.nature.com/articles/nature07830>
138. Kim HM, Park BS, Kim J-I, Kim SE, Lee J, Oh SC, et al. Crystal Structure of the TLR4-MD-2 Complex with Bound Endotoxin Antagonist Eritoran. *Cell* [Internet]. 7 de setembro de 2007 [citado 10 de junho de 2019];130(5):906–17. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0092867407010215>
139. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph* [Internet]. fevereiro de 1996 [citado 16 de setembro de 2019];14(1):33–8. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/0263785596000185>
140. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* [Internet]. 8 de agosto de 2011;12:402. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/21824423>

Supplementary Tables - A Reverse Vaccinology and Immunoinformatics Based Approach for Multi-epitope vaccine against *Clostridioides difficile* infection

Supplementary Table 1: List of the 53 *Clostridioides difficile* strain used in the analysis. List of all *C. difficile* strains (with features) retrieved from the NCBI (National Center for Biotechnology Information) database that presented a complete genome.

Strain	Size (Mb)	GC%	Gene	Protein
Cd_ 630 2006/06/27	4,2981	29.09	3981	3766
Cd_196	4,1106	28.6	3807	3615
Cd_ M120	4,0477	28.7	3756	3505
Cd_ 630 2015/02/24	4,2748	29.0	4023	3811
Cd_ 630Derm	4,2931	29.1	4044	3841
Cd_ ATCC 9689 = DSM 1296	4,2862	28.72	3984	3846
Cd_ 08ACD0030	4,1671	28.8	3882	3647
Cd_ BR81	4,1244	28.7	3785	3594
Cd_ FDAARGOS_267	4,2862	28.72	4051	3836
Cd_ 630 delta erm	4,3009	29.10	4051	3841
Cd_ DH/NAP11/106/ST-42	4,0871	28.6	3743	3551
Cd_ R0104a	4,1900	28.7	3921	3699
Cd_ W0023a	4,1101	28.7	3804	3590
Cd_ W0022a	4,1885	28.9	3947	3719
Cd_ W0003a	4,0754	28.6	3766	3563
Cd_ 10-00484	4,1195	28.6	3843	3626
Cd_ 17-01474	4,2013	28.9	3939	3719
Cd_ 12-00008	4,1196	28.6	3842	3622
Cd_ 08-00495	4,1780	28.7	3902	3682
Cd_ 12-00011	4,1199	28.6	3845	3624
Cd_ 10-00071	4,1197	28.6	3843	3627
Cd_ 10-00078	4,1198	28.6	3845	3535
Cd_ 10-00253	4,1201	28.6	3844	3618
Cd_ 09-00072	4,1199	28.6	3842	3629
Cd_ R1	4,0931	29.1	3835	3557
Cd_ R3	4,0932	29.1	3835	3557
Cd_ R2	4,0931	29.1	3835	3557
Cd_ 161	4,4743	28.82	4277	4050
Cd_ CDT4	4,2896	28.90	4035	3811
Cd_ DSM 27638	4,2297	29.00	3966	3734
Cd_ DSM 27639	4,2640	29.1	4026	3803

Cd_DSM 27640	4,2296	29.00	3964	3734
Cd_DSM 102859	4,2470	29.00	4050	3699
Cd_DSM 102860	4,2538	29.00	4071	3726
Cd_DSM 102978	4,2538	29.00	4071	3726
Cd_DSM 29688	4,2290	28.9	4041	3827
Cd_DSM 29632	4,1758	28.6	3891	3687
Cd_DSM 28196	4,2054	28.8	3949	3720
Cd_DSM 28666	4,1226	29.00	3829	3614
Cd_DSM 28669	4,1377	28.8	3908	3660
Cd_DSM 29020	4,1394	29.2	3905	3612
Cd_DSM 28670	4,1917	28.8	3962	3742
Cd_DSM 29627	4,2096	28.8	3916	3709
Cd_DSM 29629	4,1112	28.6	3802	3574
Cd_DSM 29637	4,1146	28.6	3844	3629
Cd_DSM 29745	4,2407	29.00	3975	3738
Cd_DSM 29747	4,0716	29.1	3809	3536
Cd_DSM 28668	4,2148	28.5	3900	3673
Cd_10010	4,0496	28.7	3786	3498
Cd_12038	4,0775	28.82	3822	3503
Cd_21062	4,1055	28.92	3898	3532
Cd_BI1	4,4647	28.35	-	-
Cd_Z31	4,2983	29.2	4294	4128
Cd_21062	4,1055	28.92	3898	3532

Supplementary Table 2: Multi-epitope vaccine predicted B cell conformational epitopes. The EliPro server was used to predict the conformational B cell binding epitopes via the 3D structure of the *C. difficile* multi-epitope vaccine. Six epitopes were predicted with a score varying from 0.768 to 0.562 (default parameters).

Residues	Score
A:N345, A:G346, A:P347, A:G348, A:P349, A:G368, A:G370, A:K371, A:K373, A:V374, A:S375, A:M376, A:I377, A:A378, A:Q379, A:L380, A:T381, A:P382, A:L383, A:Y384, A:G385, A:G386, A:P387, A:G388, A:P389, A:E392, A:Y395, A:F396, A:N399, A:N400, A:A401, A:I402, A:A403, A:A404, A:V405, A:G406, A:P407, A:G408, A:P409, A:G410, A:G411, A:K412, A:K413, A:Y414, A:Y415, A:F416, A:N417, A:T418, A:N419, A:T420, A:S421, A:I422, A:A423, A:S424, A:T425, A:G426, A:P427, A:G428, A:P429, A:G430, A:G431, A:K432, A:K433, A:Y434, A:Y435, A:F436, A:N437, A:T438, A:N439, A:T440, A:A441, A:E442, A:A443, A:A444, A:T445, A:G446, A:P447, A:G448, A:P449, A:G450, A:G451, A:F452, A:E453, A:Y454, A:F455, A:A456, A:P457, A:A458, A:N459, A:T460, A:Q461, A:N462, A:N463, A:N464, A:I465, A:G466, A:P467, A:G468, A:P469, A:G470, A:N471, A:G472, A:F473, A:E474, A:Y475, A:F476, A:A477, A:P478, A:A479, A:N480, A:T481, A:H482, A:N483, A:N484, A:N485, A:G486, A:P487, A:G488	0.768

A:L3, A:V4, A:G5, A:V6, A:S7, A:A11, A:S13, A:M14, A:I15, A:I16, A:M17, A:S18, A:I19, A:M20, A:L21, A:A22, A:A23, A:Y24, A:W25, A:A26, A:F27, A:A28, A:N29, A:G30, A:I31, A:S32, A:V33, A:A34, A:A35, A:Y36, A:Y37, A:A38, A:M39, A:G40, A:A41, A:L42, A:S43, A:M44, A:L45, A:A46, A:A47, A:Y48, A:S49, A:M50, A:I51, A:A52, A:Q53, A:L54, A:T55, A:P56, A:L57, A:A58, A:A59, A:Y60, A:Y61, A:Y62, A:F63, A:N64, A:P65, A:N66, A:N67, A:A68, A:I69, A:A70, A:A71, A:Y72, A:Y73, A:Y74, A:F75, A:N76, A:T77, A:N78, A:T79, A:S80, A:I81, A:A82, A:A83, A:Y84, A:Y85, A:Y86, A:F87, A:N88, A:T89, A:N90, A:T91, A:A92, A:E93, A:A94, A:A95, A:Y96, A:F97, A:A98, A:A100, A:N101, A:Q103, A:N113, A:T114, A:A124, A:N127, A:E134, A:M135	0.714
A:Y534, A:Y535, A:I563, A:D564, A:G565, A:G566, A:P567, A:N583, A:L584, A:G588, A:P589, A:G590, A:D591, A:Q592, A:E593, A:I594, A:M595, A:D596, A:A597, A:H598, A:K599, A:I600, A:F602, A:A603, A:D604, A:K618, A:L619, A:R620, A:I621, A:Y622, A:A623, A:I624, A:T625, A:G626, A:P627, A:G628, A:P629, A:G630, A:T631, A:E632, A:I633, A:K634, A:F636, A:E638, A:G639, A:T640, A:L641, A:A642, A:S643, A:T644, A:G650, A:L651, A:A652, A:D653, A:A654, A:M655, A:S656, A:I657, A:A658, A:P659, A:G666, A:P667, A:G668, A:P669, A:G670, A:D671, A:A672, A:L673, A:A674, A:A675, A:A676, A:P677, A:V678, A:A679, A:A680, A:N681, A:F682, A:G683, A:V684, A:T685, A:G686, A:P687, A:G688, A:P689, A:G690, A:N691, A:S692, A:A693, A:G694, A:T695, A:K696, A:L697, A:A698, A:M699, A:S700, A:A701, A:I702, A:F703, A:D704, A:T705, A:G706, A:P707, A:G708, A:P709, A:G710, A:N713, A:K714, A:L715, A:V716, A:S717, A:P718, A:A719, A:P720, A:I721, A:V722, A:L723, A:A724, A:T725, A:G726, A:P727, A:G728, A:P729, A:G730, A:D731, A:G732, A:I733, A:L734, A:K735, A:K736, A:Q737, A:I738, A:V739, A:D740, A:V741, A:A742, A:S743, A:Q744, A:N745, A:G746, A:P747, A:G748, A:P749, A:G750, A:Y751, A:K752, A:R753, A:T754, A:L755, A:T756, A:M757, A:A758, A:G759, A:I760, A:V761, A:P762, A:E763, A:S764, A:L765	0.711
A:A130, A:A131, A:Y132, A:W133	0.616
A:G528, A:P529, A:G530, A:S532, A:L533, A:F536	0.598
A:L269, A:K270, A:K271, A:I273, A:A274, A:A275, A:Y276, A:T277	0.562

Supplementary Table 3: Top 10 democratic solutions. The top 10 docking poses were visualized by Chimera and the hydrogen bonds calculated. The solution with the lowest energy/mean energy, higher number of contacts and H-bonds is the probable correct structure (27c).

	Structure	Mean Energy	Number of Members	Mean energy Deviation	Total Contacts	Total H-bonds	H-bonds at designated receptor binding sites
1	25c	-50.200	1	0.000	797	15	4
2	32d	-18.260	1	0.000	673	8	3
3	42b	-36.010	1	0.000	683	17	4
4	36b	-35.360	1	0.000	727	12	2
5	19c	-36.140	1	0.000	550	11	6
6	39a	-27.450	1	0.000	627	12	2
7	30d	-17.220	1	0.000	635	14	4
8	39b	-28.045	2	1.005	594	9	2
9	51d	-30.990	1	0.000	658	17	5
10	27c	-42.260	1	0.000	750	20	15

**GENERAL CONCLUSIONS
AND FUTURE
PROSPECTIVES**

The emergency of MDR bacteria emphasize the necessity of alternative treatment and prevention methods to overcome the dissemination of pathogens. Immunization is a safe, cost-effective and efficient method to prevent diseases. However, the classical vaccines may display adverse reaction due to the presence of the whole or fragments of the microorganism and the incidence of impurities/allergens, kept after the purification process. Furthermore, a few drawbacks impede an increase rate of the worldwide immunization coverage (e.g. underprivileged sanitation and infrastructures, lack of trained medical personal, and anti-vaccine principles) and facilitates the reemergence of pathogens (i.e. outbreaks) (e.g. travel to endemic countries without vaccination, inappropriate use of booster vaccine doses, emergency of new strains variant, and dissemination of antimicrobial resistance), which is further aggravated due to the difference in the vaccines batches among pharmaceutical industries (i.e. diverge production process). Bering these, new cutting-edge approaches were developed to reduce the hazards and engineer a vaccine capable to induce an effective immune response without (or with a minimal) immune reactions. The peptide-based vaccines accomplished this task in a cost-effective, reproducible, simple, customized manner. This methodology employ only the minimal region required for the pathogen recognition by the immune system (i.e. antigenic epitope) in the vaccine construction, and the multi-epitope strategy can target different strains, related pathogens and life cycles, without interference, in a single vaccine, trough the addition of multiples epitopes.

The progress of the immunoinformatics field and the high infection rate of *Corynebacterium diphtheriae* and *Clostridioides difficile* prompt us to design cutting-edge vaccines against the etiological agents of diphtheria and CDI. Diphtheria is a respiratory, cutaneous and systemic infection (previously the main source of child mortality) caused mainly by the action of DT whereas CDI is a nosocomial and community-acquired illness, primarily incited by the indiscriminate use of antibiotics (antibiotic induced dysbiosis) and upheld by the action of a variety of genes including the toxins TcdA and TcdB (major virulent factors).

Both disease prevention and treatment present hitches that could be overcome by the constructed multi-epitopes vaccines. The diphtheria vaccine is not particularly effective against non-toxigenic strains, requires booster doses, and the vaccine coverage is affected by socioeconomic problems, which transpires in the location of the outbreaks, the high incidence in immigrants, and travelers. Meanwhile, the recommended treatment for CDI is the use of antibiotics to eliminate the bacterium; nonetheless, the dissemination of antimicrobial resistance is a constantly treat since this opportunistic pathogen is highly

versatile due to mobile elements encoded within the genome and already presents multiples antibiotic resistance genes. Other non-recommended (insufficient data or still in clinical trials) like probiotics and untested vaccines. The vaccines against *C. difficile* are mainly based on inactivated toxins (TcdA and TcdB), which are the main virulent factors of the bacterium, and thus could lead to adverse reactions and other drawbacks similar to the diphtheria toxoid vaccine. Altogether, the data highlight the need for a new vaccine that entails a variety of virulent factors that would induce a more prominent immune reaction against this deadly pathogen.

Hence, we proposed the design of multi-epitope vaccines against two life-threatening pathogens applying reverse vaccinology and immunoinformatics methodologies to uncover the best protein targets and construct vaccines capable to induce a potent immune response via a cost-effective, reproducible and safe method. A multi-epitope vaccine could effectively reduce diphtheria and CDI cases since it bypass many of the currently disadvantages. This methodology allow a better coverage of the pathogens by including common proteins among toxigenic and non-toxigenic strains and still contains the key virulent factors (DT and TcdA/TcdB of *C. diphtheriae* and *C. difficile*, respectively). Therefore, the predicted vaccines combine the already used tactic (toxins) to new and probably effective antigens that could generate a stronger immune response. The methodology applied here for both heterovalent vaccines (epitope prediction to simulation) can be used to construct other multi-epitope vaccines to combat pathogenic infections.

Concisely, this work comprised *in silico* analysis for the design of potential vaccines that, according to the predicted results, would protect the host against *C. diphtheriae* or *C. difficile*. Nonetheless, the analyzes were preliminary and require experimental validation.

The future endeavors are common between the two papers and enumerated accordantly:

- (i) Evaluate the cloning and expression of the plasmid contacting the optimized multi-epitope vaccine sequence in *E. coli* K12;
- (ii) Investigate the vaccine trough *in vivo* experiments (mouse model).

REFERENCES

ABBAS, A.; LICHTMAN, A. H.; PILLAI, S. *Basic Immunology*. Disponível em: <<https://www.elsevier.com/books/basic-immunology/abbas/978-0-323-39082-8>>. Acesso em: 22 maio 2019.

ABBAS, A.; LICHTMAN, A. H.; PILLAI, S. *Cellular and Molecular Immunology - 8th Edition*. Disponível em: <<https://www.elsevier.com/books/cellular-and-molecular-immunology/abbas/978-0-323-22275-4>>. Acesso em: 22 maio 2019.

ABOUGERGI, M. S.; KWON, J. H. Intravenous immunoglobulin for the treatment of Clostridium difficile infection: a review. *Digestive Diseases and Sciences*, v. 56, n. 1, p. 19–26, jan. 2011.

ABT, M. C.; MCKENNEY, P. T.; PAMER, E. G. Clostridium difficile colitis: pathogenesis and host defence. *Nature reviews. Microbiology*, v. 14, n. 10, p. 609–620, out. 2016. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5109054/>>. Acesso em: 23 jan. 2019.

AGUILERA, M.; CERDÀ-CUÉLLAR, M.; MARTÍNEZ, V. Antibiotic-induced dysbiosis alters host-bacterial interactions and leads to colonic sensory and motor changes in mice. *Gut Microbes*, v. 6, n. 1, p. 10–23, 2015.

ALENGHAT, T. *et al.* Histone deacetylase 3 coordinates commensal-bacteria-dependent intestinal homeostasis. *Nature*, v. 504, n. 7478, p. 153–157, 5 dez. 2013.

ALI, M. *et al.* Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Scientific reports*, v. 7, n. 1, p. 9232, 2017.

ALSHANQITI, F. M. *et al.* Adjuvants for Clostridium tetani and Clostridium diphtheriae vaccines updating. *Human Antibodies*, v. 25, n. 1–2, p. 23–29, 20 fev. 2017. Disponível em: <<http://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/HAB-160302>>. Acesso em: 3 maio 2019.

ANTUNES, A.; MARTIN-VERSTRAETE, I.; DUPUY, B. CcpA-mediated repression of Clostridium difficile toxin gene expression. *Molecular Microbiology*, v. 79, n. 4, p. 882–899, fev. 2011.

ANTUNES, C. A. *et al.* Characterization of DIP0733, a multi-functional virulence factor of Corynebacterium diphtheriae. *Microbiology*, v. 161, n. 3, p. 639–647, 2015. Disponível em: <<https://mic.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.000020>>. Acesso em: 27 mar. 2019.

ARGONDIZZO, A. P. C. *et al.* Identification of Proteins in Streptococcus pneumoniae by Reverse Vaccinology and Genetic Diversity of These Proteins in Clinical Isolates. 2015. Disponível em: <<https://www.arca.fiocruz.br/handle/icict/10062>>. Acesso em: 20 maio 2019.

ARITA, M.; ROBERT, M.; TOMITA, M. All systems go: launching cell simulation fueled by integrated experimental biology data. *Current Opinion in Biotechnology*, v. 16, n. 3, p. 344–349, jun. 2005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0958166905000601>>. Acesso em: 16 maio 2019.

ARORA, T.; BÄCKHED, F. The gut microbiota and metabolic disease: current understanding and future perspectives. *Journal of Internal Medicine*, v. 280, n. 4, p. 339–349, 2016. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/joim.12508>>. Acesso em: 11 jul. 2018.

AUBRY, A. *et al.* Modulation of Toxin Production by the Flagellar Regulon in *Clostridium difficile*. *Infection and Immunity*, v. 80, n. 10, p. 3521–3532, out. 2012. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3457548/>>. Acesso em: 1 jun. 2019.

AWAD, M. M. *et al.* *Clostridium difficile* virulence factors: Insights into an anaerobic spore-forming pathogen. *Gut Microbes*, v. 5, n. 5, p. 579–593, 2014.

AZAB, M. *et al.* Comparison of the Hospital-Acquired *Clostridium difficile* Infection Risk of Using Proton Pump Inhibitors versus Histamine-2 Receptor Antagonists for Prophylaxis and Treatment of Stress Ulcers: A Systematic Review and Meta-Analysis. *Gut and Liver*, v. 11, n. 6, p. 781–788, 15 nov. 2017.

BABAKHANI, F. *et al.* Fidaxomicin inhibits spore production in *Clostridium difficile*. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, v. 55 Suppl 2, p. S162-169, ago. 2012.

BABAN, S. T. *et al.* The Role of Flagella in *Clostridium difficile* Pathogenesis: Comparison between a Non-Epidemic and an Epidemic Strain. *PLOS ONE*, v. 8, n. 9, p. e73026, 23 set. 2013. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0073026>>. Acesso em: 1 jun. 2019.

BAE, K.-D. The Evolution and Value of Diphtheria Vaccine. *KSBB Journal*, v. 26, n. 6, p. 491–504, 31 dez. 2011. Disponível em: <<http://koreascience.or.kr/journal/view.jsp?kj=KHGSBC&py=2011&vnc=v26n6&sp=491>>. Acesso em: 6 abr. 2019.

BAGDASARIAN, N.; RAO, K.; MALANI, P. N. Diagnosis and treatment of *Clostridium difficile* in adults: a systematic review. *JAMA*, v. 313, n. 4, p. 398–408, 27 jan. 2015.

BARTLETT, J. G. Historical perspectives on studies of *Clostridium difficile* and *C. difficile* infection. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, v. 46 Suppl 1, p. S4-11, 15 jan. 2008.

BÄUMLER, A. J.; SPERANDIO, V. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature*, v. 535, n. 7610, p. 85–93, jul. 2016. Disponível em: <<https://www.nature.com/articles/nature18849>>. Acesso em: 22 jul. 2018.

BECATTINI, S.; TAUR, Y.; PAMER, E. G. Antibiotic-Induced Changes in the Intestinal Microbiota and Disease. *Trends in molecular medicine*, v. 22, n. 6, p. 458–478, jun. 2016. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4885777/>>. Acesso em: 16 jul. 2018.

BERNARD, K. The Genus *Corynebacterium* and Other Medically-Relevant, Coryneform like Bacteria. *Journal of Clinical Microbiology*, p. JCM.00796-12, 25 jul. 2012. Disponível em: <<http://jcm.asm.org/content/early/2012/07/20/JCM.00796-12>>. Acesso em: 13 jul. 2014.

BÉZAY, N. *et al.* Safety, immunogenicity and dose response of VLA84, a new vaccine candidate against *Clostridium difficile*, in healthy volunteers. *Vaccine*, v. 34, n. 23, p. 2585–2592, 17 2016.

BIEDERMANN, L.; ROGLER, G. The intestinal microbiota: its role in health and disease. *European Journal of Pediatrics*, v. 174, n. 2, p. 151–167, fev. 2015.

BJÖRKSTÉN, B. Effects of intestinal microflora and the environment on the development of asthma and allergy. *Springer Seminars in Immunopathology*, v. 25, n. 3–4, p. 257–270, 1 fev. 2004. Disponível em: <<http://link.springer.com/10.1007/s00281-003-0142-2>>. Acesso em: 23 abr. 2018.

BLANC, L. *et al.* Gram-Positive Bacterial Lipoglycans Based on a Glycosylated Diacylglycerol Lipid Anchor Are Microbe-Associated Molecular Patterns Recognized by TLR2. *PLOS ONE*, v. 8, n. 11, p. e81593, 21 nov. 2013. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0081593>>. Acesso em: 9 maio 2019.

BLOOM, D. E. The value of vaccination. *Advances in Experimental Medicine and Biology*, v. 697, p. 1–8, 2011.

BORDELEAU, E.; BURRUS, V. Cyclic-di-GMP signaling in the Gram-positive pathogen *Clostridium difficile*. *Current Genetics*, v. 61, n. 4, p. 497–502, nov. 2015.

BORRIELLO, S. P.; BARCLAY, F. E. Protection of hamsters against *Clostridium difficile* ileocaecitis by prior colonisation with non-pathogenic strains. *Journal of Medical Microbiology*, v. 19, n. 3, p. 339–350, 1 jun. 1985. Disponível em: <<http://jmm.microbiologyresearch.org/content/journal/jmm/10.1099/00222615-19-3-339>>. Acesso em: 29 maio 2019.

BRANDL, K. *et al.* Vancomycin-resistant enterococci exploit antibiotic-induced innate immune deficits. *Nature*, v. 455, n. 7214, p. 804–807, 9 out. 2008. Disponível em: <<http://www.nature.com/doi/10.1038/nature07250>>. Acesso em: 13 abr. 2018.

BRUNE, I. *et al.* The DtxR protein acting as dual transcriptional regulator directs a global regulatory network involved in iron metabolism of *Corynebacterium glutamicum*. *BMC Genomics*, v. 7, p. 21, 9 fev. 2006. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1382209/>>. Acesso em: 24 jun. 2019.

BUFFIE, C. G. *et al.* Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature*, v. 517, n. 7533, p. 205–208, 8 jan. 2015.

BUFFIE, C. G. *et al.* Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to *Clostridium difficile*-induced colitis. *Infection and Immunity*, v. 80, n. 1, p. 62–73, jan. 2012.

BURKOVSKI, A. Cell Envelope of Corynebacteria: Structure and Influence on Pathogenicity. *ISRN Microbiology*, v. 2013, 21 jan. 2013. Disponível em: <<http://www.hindawi.com/journals/isrn.microbiology/2013/935736/abs/>>. Acesso em: 29 abr. 2014.

BURKOVSKI, A. Diphtheria and its Etiological Agents. In: BURKOVSKI, A. (Org.). *Corynebacterium diphtheriae and Related Toxigenic Species*. Dordrecht: Springer Netherlands, 2014. p. 1–14. Disponível em: <http://link.springer.com/10.1007/978-94-007-7624-1_1>. Acesso em: 6 abr. 2019.

BURKOVSKI, A. The role of corynomycolic acids in *Corynebacterium*-host interaction. *Antonie van Leeuwenhoek*, v. 111, n. 5, p. 717–725, 1 maio 2018. Disponível em: <<https://doi.org/10.1007/s10482-018-1036-6>>. Acesso em: 8 maio 2019.

CARTER, G. P. *et al.* Defining the Roles of TcdA and TcdB in Localized Gastrointestinal Disease, Systemic Organ Damage, and the Host Response during *Clostridium difficile* Infections. *mBio*, v. 6, n. 3, p. e00551, 2 jun. 2015.

CENTERS FOR DISEASE CONTROL AND PREVENTION. *US CDC report on antibiotic resistance threats in the United States, 2013*. Disponível em: <<http://ecdc.europa.eu/en/news-events/us-cdc-report-antibiotic-resistance-threats-united-states-2013>>. Acesso em: 6 jun. 2019.

CERDEÑO-TÁRRAGA, A. M. *et al.* The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. *Nucleic Acids Research*, v. 31, n. 22, p. 6516–6523, 15 nov. 2003.

CHAND, D. *et al.* Molecular features of bile salt hydrolases and relevance in human health. *Biochimica et Biophysica Acta (BBA) - General Subjects*, v. 1861, n. 1, p. 2981–2991, jan. 2017. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0304416516303658>>. Acesso em: 11 jul. 2018.

CHANG, P. V. *et al.* The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. *Proceedings of the National Academy of Sciences of the United States of America*, v. 111, n. 6, p. 2247–2252, 11 fev. 2014.

CHAUDHURI, R. *et al.* Integrative immunoinformatics for Mycobacterial diseases in R platform. *Systems and Synthetic Biology*, v. 8, n. 1, p. 27–39, mar. 2014.

CHITNIS, A. S. *et al.* Epidemiology of community-associated *Clostridium difficile* infection, 2009 through 2011. *JAMA internal medicine*, v. 173, n. 14, p. 1359–1367, 22 jul. 2013.

CLARK, R.; KUPPER, T. Old Meets New: The Interaction Between Innate and Adaptive Immunity. *Journal of Investigative Dermatology*, v. 125, n. 4, p. 629–637, 1 out. 2005. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022202X15324763>>. Acesso em: 22 maio 2019.

COHEN, S. H. *et al.* Clinical Practice Guidelines for *Clostridium difficile* Infection in Adults: 2010 Update by the Society for Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA). *Infection Control & Hospital Epidemiology*, v. 31, n. 05, p. 431–455, maio 2010. Disponível em: <https://www.cambridge.org/core/product/identifier/S0195941700029064/type/journal_article>. Acesso em: 1 jun. 2019.

COLLIER, R. J. Understanding the mode of action of diphtheria toxin: a perspective on progress during the 20th century. *Toxicon*, v. 39, n. 11, p. 1793–1803, 1 nov. 2001. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0041010101001659>>. Acesso em: 7 maio 2019.

COLLINS, M. D. *et al.* The Phylogeny of the Genus *Clostridium*: Proposal of Five New Genera and Eleven New Species Combinations. *International Journal of Systematic and Evolutionary Microbiology*, v. 44, n. 4, p. 812–826, 1994. Disponível em: <<https://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-44-4-812>>. Acesso em: 29 maio 2019.

COLOMBO, A. V. *et al.* *Corynebacterium diphtheriae* surface proteins as adhesins to human erythrocytes. *FEMS Microbiology Letters*, v. 197, n. 2, p. 235–239, abr. 2001. Disponível em: <<https://academic.oup.com/femsle/article-lookup/doi/10.1111/j.1574-6968.2001.tb10609.x>>. Acesso em: 6 maio 2019.

CORTHER, G.; MULLER, M. C. Emergence in gnotobiotic mice of nontoxigenic clones of *Clostridium difficile* from a toxinogenic one. *Infection and Immunity*, v. 56, n. 6, p. 1500–1504, jun. 1988. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC259427/>>. Acesso em: 29 maio 2019.

COWARDIN, C. A. *et al.* The binary toxin CDT enhances *Clostridium difficile* virulence by suppressing protective colonic eosinophilia. *Nature Microbiology*, v. 1, n. 8, p. 16108, 11 2016.

DAWSON, P. A.; KARPEN, S. J. Intestinal transport and metabolism of bile acids. *Journal of Lipid Research*, v. 56, n. 6, p. 1085–1099, 1 jun. 2015. Disponível em: <<http://www.jlr.org/content/56/6/1085>>. Acesso em: 21 jun. 2018.

DE AGUIAR VALLIM, T. Q.; TARLING, E. J.; EDWARDS, P. A. Pleiotropic roles of bile acids in metabolism. *Cell Metabolism*, v. 17, n. 5, p. 657–669, 7 maio 2013.

DE BRUYN, G. *et al.* Defining the optimal formulation and schedule of a candidate toxoid vaccine against *Clostridium difficile* infection: A randomized Phase 2 clinical trial. *Vaccine*, v. 34, n. 19, p. 2170–2178, 27 abr. 2016.

DEAKIN, L. J. *et al.* The *Clostridium difficile* spo0A gene is a persistence and transmission factor. *Infection and Immunity*, v. 80, n. 8, p. 2704–2711, ago. 2012.

DELANY, I.; RAPPUOLI, R.; SEIB, K. L. Vaccines, Reverse Vaccinology, and Bacterial Pathogenesis. *Cold Spring Harbor Perspectives in Medicine*, v. 3, n. 5, p. a012476–a012476, 1 maio 2013. Disponível em: <<http://perspectivesinmedicine.cshlp.org/lookup/doi/10.1101/cshperspect.a012476>>. Acesso em: 20 maio 2019.

DEN BESTEN, H. M. W. *et al.* Next generation of microbiological risk assessment: Potential of omics data for exposure assessment. *International Journal of Food Microbiology*, Omics in MRA - the integration of omics in microbiological risk assessment. v. 287, p. 18–27, 20 dez. 2018. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0168160517304154>>. Acesso em: 16 maio 2019.

DEPLA, E. *et al.* Rational Design of a Multiepitope Vaccine Encoding T-Lymphocyte Epitopes for Treatment of Chronic Hepatitis B Virus Infections. *Journal of Virology*, v. 82, n. 1, p. 435–450, 1 jan. 2008. Disponível em: <<https://jvi.asm.org/content/82/1/435>>. Acesso em: 14 mar. 2019.

DETHLEFSEN, L.; RELMAN, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proceedings of the National Academy of Sciences of the United States of America*, v. 108 Suppl 1, p. 4554–4561, 15 mar. 2011.

DIAS, A. A. DE S. DE O. *et al.* Difteria pelo *Corynebacterium ulcerans*: uma zoonose emergente no Brasil e no mundo. *Revista de Saúde Pública*, v. 45, n. 6, p. 1176–1191, dez. 2011. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102011000600021&lng=pt&tlng=pt>. Acesso em: 24 jun. 2019.

DIDELLOT, X. *et al.* Within-host evolution of bacterial pathogens. *Nature Reviews Microbiology*, v. 14, n. 3, p. 150–162, mar. 2016. Disponível em: <<http://www.nature.com/articles/nrmicro.2015.13>>. Acesso em: 16 maio 2019.

DITTMANN, S. *et al.* Successful control of epidemic diphtheria in the states of the Former Union of Soviet Socialist Republics: lessons learned. *The Journal of Infectious Diseases*, v. 181 Suppl 1, p. S10–22, fev. 2000.

DUDEK, N. L. *et al.* Epitope discovery and their use in peptide based vaccines. *Current Pharmaceutical Design*, v. 16, n. 28, p. 3149–3157, 2010.

EDWARDS, A. N.; MCBRIDE, S. M. Initiation of Sporulation in *Clostridium difficile*: a Twist on the Classic Model. *FEMS microbiology letters*, v. 358, n. 2, p. 110–118, set. 2014. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4189817/>>. Acesso em: 1 jun. 2019.

EDWARDS, A. N.; NAWROCKI, K. L.; MCBRIDE, S. M. Conserved oligopeptide permeases modulate sporulation initiation in *Clostridium difficile*. *Infection and Immunity*, v. 82, n. 10, p. 4276–4291, out. 2014.

EHRETH, J. The global value of vaccination. *Vaccine*, v. 21, n. 7–8, p. 596–600, 30 jan. 2003.

EIKMANN, B. J.; BLOMBACH, B. The pyruvate dehydrogenase complex of *Corynebacterium glutamicum*: An attractive target for metabolic engineering. *Journal of Biotechnology*, v. 192, p. 339–345, dez. 2014. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0168165614000388>>. Acesso em: 24 jun. 2019.

ELDHOLM, V. *et al.* Evolution of extensively drug-resistant *Mycobacterium tuberculosis* from a susceptible ancestor in a single patient. *Genome Biology*, v. 15, n. 11, p. 490, nov. 2014. Disponível em: <<http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0490-3>>. Acesso em: 16 maio 2019.

ESKOLA, J.; LUMIO, J.; VUOPIO-VARKILA, J. Resurgent diphtheria--are we safe? *British Medical Bulletin*, v. 54, n. 3, p. 635–645, 1998.

FERREYRA, J. A. *et al.* Gut microbiota-produced succinate promotes *C. difficile* infection after antibiotic treatment or motility disturbance. *Cell host & microbe*, v. 16, n. 6, p. 770–777, 10 dez. 2014. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4859344/>>. Acesso em: 1 out. 2018.

FIGUEROA CASTRO, C. E.; MUNOZ-PRICE, L. S. Advances in Infection Control for *Clostridioides* (Formerly *Clostridium*) *difficile* Infection. *Current Treatment Options in Infectious Diseases*, v. 11, n. 1, p. 12–22, 1 mar. 2019. Disponível em: <<https://doi.org/10.1007/s40506-019-0179-y>>. Acesso em: 24 maio 2019.

FLOWER, D. R. *Bioinformatics for Vaccinology*. Chichester, UK: John Wiley & Sons, Ltd, 2008. Disponível em: <<http://doi.wiley.com/10.1002/9780470699836>>. Acesso em: 20 maio 2019.

FOGED, C. *Subunit vaccines of the future: the need for safe, customized and optimized particulate delivery systems*. - *PubMed - NCBI*. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/22826868>>. Acesso em: 20 maio 2019.

FRÖHLICH, E. E. *et al.* Cognitive impairment by antibiotic-induced gut dysbiosis: Analysis of gut microbiota-brain communication. *Brain, Behavior, and Immunity*, v. 56, p. 140–155, 1 ago. 2016. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S088915911630040X>>. Acesso em: 18 set. 2018.

FUKUDA, S. *et al.* Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature*, v. 469, n. 7331, p. 543–547, jan. 2011. Disponível em: <<https://www.nature.com/articles/nature09646>>. Acesso em: 25 jul. 2018.

GABORIAU-ROUTHIAU, V. *et al.* The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. *Immunity*, v. 31, n. 4, p. 677–689, 16 out. 2009.

GAO, W. *et al.* Two Novel Point Mutations in Clinical *Staphylococcus aureus* Reduce Linezolid Susceptibility and Switch on the Stringent Response to Promote Persistent Infection. *PLoS Pathogens*, v. 6, n. 6, p. e1000944, 10 jun. 2010. Disponível em: <<http://dx.plos.org/10.1371/journal.ppat.1000944>>. Acesso em: 16 maio 2019.

GEISEL, R. E. *et al.* In Vivo Activity of Released Cell Wall Lipids of *Mycobacterium bovis* Bacillus Calmette-Guérin Is Due Principally to Trehalose Mycolates. *The Journal of Immunology*, v. 174, n. 8, p. 5007–5015, 15 abr. 2005. Disponível em: <<http://www.jimmunol.org/content/174/8/5007>>. Acesso em: 9 maio 2019.

GERDING, D. N. *et al.* Administration of spores of nontoxigenic *Clostridium difficile* strain M3 for prevention of recurrent *C. difficile* infection: a randomized clinical trial. *JAMA*, v. 313, n. 17, p. 1719–1727, 5 maio 2015.

GERDING, D. N. *et al.* *Clostridium difficile* binary toxin CDT: mechanism, epidemiology, and potential clinical importance. *Gut Microbes*, v. 5, n. 1, p. 15–27, fev. 2014.

GERDING, D. N.; SAMBOL, S. P.; JOHNSON, S. Non-toxigenic *Clostridioides* (Formerly *Clostridium*) *difficile* for Prevention of *C. difficile* Infection: From Bench to Bedside Back to Bench and Back to Bedside. *Frontiers in Microbiology*, v. 9, 26 jul. 2018. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6070627/>>. Acesso em: 24 maio 2019.

GIEL, J. L. *et al.* Metabolism of Bile Salts in Mice Influences Spore Germination in *Clostridium difficile*. *PLOS ONE*, v. 5, n. 1, p. e8740, 15 jan. 2010. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008740>>. Acesso em: 19 jul. 2018.

GIL, F. *et al.* *Clostridioides* (*Clostridium*) *difficile* infection: current and alternative therapeutic strategies. *Future Microbiology*, v. 13, n. 4, p. 469–482, 21 fev. 2018. Disponível em: <<https://www.futuremedicine.com/doi/full/10.2217/fmb-2017-0203>>. Acesso em: 24 maio 2019.

GREATHOUSE, K. L.; HARRIS, C. C.; BULTMAN, S. J. Dysfunctional Families: *Clostridium scindens* and Secondary Bile Acids Inhibit the Growth of *Clostridium difficile*. *Cell Metabolism*, v. 21, n. 1, p. 9–10, jan. 2015. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S1550413114005671>>. Acesso em: 28 jan. 2016.

GUARALDI, A. L. DE M.; HIRATA, R.; AZEVEDO, V. A. DE C. *Corynebacterium diphtheriae*, *Corynebacterium ulcerans* and *Corynebacterium pseudotuberculosis*—General Aspects. In: BURKOVSKI, A. (Org.). *Corynebacterium diphtheriae and Related Toxigenic Species: Genomics, Pathogenicity and Applications*. Dordrecht: Springer Netherlands, 2014. p. 15–37. Disponível em: <https://doi.org/10.1007/978-94-007-7624-1_2>. Acesso em: 27 mar. 2019.

GUINANE, C. M.; COTTER, P. D. Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. *Therapeutic Advances in Gastroenterology*, v. 6, n. 4, p. 295–308, jul. 2013. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3667473/>>. Acesso em: 23 abr. 2018.

HADFIELD, T. L. *et al.* The pathology of diphtheria. *The Journal of Infectious Diseases*, v. 181 Suppl 1, p. S116-120, fev. 2000.

HAJIGHAHRAMANI, N. *et al.* Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, v. 48, p. 83–94, 2017.

HALL, B. G.; EHRlich, G. D.; HU, F. Z. Pan-genome analysis provides much higher strain typing resolution than multi-locus sequence typing. *Microbiology*, v. 156, n. 4, p. 1060–1068, 1 abr. 2010. Disponível em: <<http://mic.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.035188-0>>. Acesso em: 12 jun. 2019.

HASEGAWA, M. *et al.* Interleukin-22 regulates the complement system to promote resistance against pathobionts after pathogen-induced intestinal damage. *Immunity*, v. 41, n. 4, p. 620–632, 16 out. 2014.

HASIN, Y.; SELDIN, M.; LUSIS, A. Multi-omics approaches to disease. *Genome Biology*, v. 18, n. 1, p. 83, dez. 2017. Disponível em: <<http://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1215-1>>. Acesso em: 16 maio 2019.

HE, M. *et al.* Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nature genetics*, v. 45, n. 1, p. 109–113, jan. 2013. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3605770/>>. Acesso em: 19 jul. 2018.

HE, M. *et al.* Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proceedings of the National Academy of Sciences*, v. 107, n. 16, p. 7527–7532, 20 abr. 2010. Disponível em: <<https://www.pnas.org/content/107/16/7527>>. Acesso em: 12 jun. 2019.

HE, Y. *et al.* *Emerging Vaccine Informatics*. Research article. Disponível em: <<https://www.hindawi.com/journals/bmri/2010/218590/>>. Acesso em: 20 maio 2019.

HEEG, D. *et al.* Spores of *Clostridium difficile* Clinical Isolates Display a Diverse Germination Response to Bile Salts. *PLOS ONE*, v. 7, n. 2, p. e32381, 22 fev. 2012. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0032381>>. Acesso em: 19 jul. 2018.

HEINSEN, F.-A. *et al.* Dynamic changes of the luminal and mucosa-associated gut microbiota during and after antibiotic therapy with paromomycin. *Gut Microbes*, v. 6, n. 4, p. 243–254, 15 jul. 2015. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4615565/>>. Acesso em: 17 abr. 2018.

HESSLING, M.; FEIERTAG, J.; HÖNES, K. Pathogens provoking most deaths worldwide. *Bioscience Biotechnology Research Communications*, v. 10, p. 1–7, 1 jul. 2017.

HIRATA, R. *et al.* Potential pathogenic role of aggregative-adhering *Corynebacterium diphtheriae* of different clonal groups in endocarditis. *Brazilian Journal of Medical and Biological Research = Revista Brasileira De Pesquisas Medicas E Biologicas*, v. 41, n. 11, p. 986–991, 2008.

HIRATA, RAPHAEL *et al.* Patterns of adherence to HEp-2 cells and actin polymerisation by toxigenic *Corynebacterium diphtheriae* strains. *Microbial Pathogenesis*, v. 36, n. 3, p. 125–130, mar. 2004.

HIROTA, K. *et al.* Fate mapping of IL-17-producing T cells in inflammatory responses. *Nature Immunology*, v. 12, n. 3, p. 255–263, mar. 2011.

HÖFLER, W. Cutaneous diphtheria. *International Journal of Dermatology*, v. 30, n. 12, p. 845–847, dez. 1991.

HOFMANN, A. F. The continuing importance of bile acids in liver and intestinal disease. *Archives of Internal Medicine*, v. 159, n. 22, p. 2647–2658, 13 dez. 1999.

HOFMANN, A. F.; HAGEY, L. R. Bile acids: chemistry, pathochemistry, biology, pathobiology, and therapeutics. *Cellular and molecular life sciences: CMLS*, v. 65, n. 16, p. 2461–2483, ago. 2008.

HOFMANN, ALAN F. Bile Acids: The Good, the Bad, and the Ugly. *Physiology*, v. 14, n. 1, p. 24–29, 1 fev. 1999. Disponível em: <<https://www.physiology.org/doi/10.1152/physiologyonline.1999.14.1.24>>. Acesso em: 11 jul. 2018.

HOLMES, A. H. *et al.* Understanding the mechanisms and drivers of antimicrobial resistance. *The Lancet*, v. 387, n. 10014, p. 176–187, 9 jan. 2016. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0140673615004730>>. Acesso em: 3 jun. 2019.

HOLMES, R. K. Biology and molecular epidemiology of diphtheria toxin and the tox gene. *The Journal of Infectious Diseases*, v. 181 Suppl 1, p. S156–167, fev. 2000.

HONDA, K.; LITTMAN, D. R. The microbiota in adaptive immune homeostasis and disease. *Nature*, v. 535, n. 7610, p. 75–84, jul. 2016. Disponível em: <<https://www.nature.com/articles/nature18848>>. Acesso em: 20 maio 2019.

HOOPER, L. V.; LITTMAN, D. R.; MACPHERSON, A. J. Interactions between the microbiota and the immune system. *Science (New York, N.Y.)*, v. 336, n. 6086, p. 1268–1273, 8 jun. 2012.

HORGAN, R. P.; KENNY, L. C. ‘Omic’ technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist*, v. 13, n. 3, p. 189–195, 2011. Disponível em: <<https://obgyn.onlinelibrary.wiley.com/doi/abs/10.1576/toag.13.3.189.27672>>. Acesso em: 15 maio 2019.

HORNEF, M. Pathogens, Commensal Symbionts, and Pathobionts: Discovery and Functional Effects on the Host. *ILAR journal*, v. 56, n. 2, p. 159–162, 2015.

HOWDEN, B. P. *et al.* Evolution of multidrug resistance during *Staphylococcus aureus* infection involves mutation of the essential two component regulator WalkR. *PLoS pathogens*, v. 7, n. 11, p. e1002359, nov. 2011.

HUANG, D. B.; WU, J. J.; TYRING, S. K. A review of licensed viral vaccines, some of their safety concerns, and the advances in the development of investigational viral vaccines. *The Journal of Infection*, v. 49, n. 3, p. 179–209, out. 2004.

IDEKER, T.; GALITSKI, T.; HOOD, L. A NEW APPROACH TO DECODING LIFE : Systems Biology. *Annual Review of Genomics and Human Genetics*, v. 2, n. 1, p. 343–372, set. 2001. Disponível em: <<http://www.annualreviews.org/doi/10.1146/annurev.genom.2.1.343>>. Acesso em: 16 maio 2019.

IEBBA, V. *et al.* Eubiosis and dysbiosis: the two sides of the microbiota. *The New Microbiologica*, v. 39, n. 1, p. 1–12, jan. 2016.

IWASAKI, A.; MEDZHITOV, R. Control of adaptive immunity by the innate immune system. *Nature Immunology*, v. 16, n. 4, p. 343–353, abr. 2015. Disponível em: <<https://www.nature.com/articles/ni.3123>>. Acesso em: 20 maio 2019.

JAMAL, S. B. *et al.* Pathogenesis of *Corynebacterium diphtheriae* and available vaccines: An Overview. *Global Journal of Infectious Diseases and Clinical Research*, v. 3, n. 1, p. 020–024, 30 out. 2017. Disponível em: <<https://www.peertechz.com/articles/GJIDCR-3-114.php>>. Acesso em: 6 abr. 2019.

JANEZIC, S. *et al.* International *Clostridium difficile* animal strain collection and large diversity of animal associated strains. *BMC Microbiology*, v. 14, n. 1, p. 173, 28 jun. 2014. Disponível em: <<https://doi.org/10.1186/1471-2180-14-173>>. Acesso em: 1 jun. 2019.

JANEZIC, S.; RUPNIK, M. Genomic diversity of *Clostridium difficile* strains. *Research in Microbiology*, Special issue on Insights into toxigenic Clostridia through genomics. v. 166, n. 4, p. 353–360, 1 maio 2015. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0923250815000303>>. Acesso em: 3 jun. 2019.

JANVILISRI, T. *et al.* Microarray Identification of *Clostridium difficile* Core Components and Divergent Regions Associated with Host Origin. *Journal of Bacteriology*, v. 191, n. 12, p. 3881–3891, 15 jun. 2009. Disponível em: <<http://jb.asm.org/cgi/doi/10.1128/JB.00222-09>>. Acesso em: 12 jun. 2019.

JERNBERG, C. *et al.* Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *The ISME journal*, v. 1, n. 1, p. 56–66, maio 2007.

JOHNSON, V. G. *et al.* The role of the diphtheria toxin receptor in cytosol translocation. *Journal of Biological Chemistry*, v. 263, n. 3, p. 1295–1300, 25 jan. 1988. Disponível em: <<http://www.jbc.org/content/263/3/1295>>. Acesso em: 7 maio 2019.

JONES, B. V. *et al.* Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, v. 105, n. 36, p. 13580–13585, 9 set. 2008.

JOYCE, S. A. *et al.* Bacterial bile salt hydrolase in host metabolism: Potential for influencing gastrointestinal microbe-host crosstalk. *Gut Microbes*, v. 5, n. 5, p. 669–674, 3 set. 2014. Disponível em: <<http://dx.doi.org/10.4161/19490976.2014.969986>>. Acesso em: 10 abr. 2017.

KALLERUP, R. S.; FOGED, C. Classification of Vaccines. *Subunit Vaccine Delivery*, p. 15–29, 2015. Disponível em: <https://link.springer.com/chapter/10.1007/978-1-4939-1417-3_2>. Acesso em: 27 mar. 2019.

KAMADA, N. *et al.* Role of the gut microbiota in immunity and inflammatory disease. *Nature Reviews. Immunology*, v. 13, n. 5, p. 321–335, maio 2013.

KAPOOR, K. *et al.* Evaluation of metronidazole toxicity: a prospective study. *International Journal of Clinical Pharmacology Research*, v. 19, n. 3, p. 83–88, 1999.

KARAS, J. A.; ENOCH, D. A.; ALIYU, S. H. A review of mortality due to *Clostridium difficile* infection. *The Journal of Infection*, v. 61, n. 1, p. 1–8, jul. 2010.

KAUFMANN, S. H.; SCHAIBLE, U. E. Antigen presentation and recognition in bacterial infections. *Current Opinion in Immunology*, v. 17, n. 1, p. 79–87, fev. 2005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0952791504001967>>. Acesso em: 22 maio 2019.

KAWAMOTO, S. *et al.* Foxp3(+) T cells regulate immunoglobulin a selection and facilitate diversification of bacterial species responsible for immune homeostasis. *Immunity*, v. 41, n. 1, p. 152–165, 17 jul. 2014.

KEERSMAECKER, S. C. J. D. *et al.* Integration of omics data: how well does it work for bacteria? *Molecular Microbiology*, v. 62, n. 5, p. 1239–1250, 2006. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2006.05453.x>>. Acesso em: 16 maio 2019.

KHANNA, S. *et al.* A Novel Microbiome Therapeutic Increases Gut Microbial Diversity and Prevents Recurrent Clostridium difficile Infection. *The Journal of Infectious Diseases*, v. 214, n. 2, p. 173–181, 15 jul. 2016. Disponível em: <<https://academic.oup.com/jid/article/214/2/173/2572105>>. Acesso em: 7 fev. 2019.

KIM, S.; COVINGTON, A.; PAMER, E. G. The intestinal microbiota: Antibiotics, colonization resistance, and enteric pathogens. *Immunological Reviews*, v. 279, n. 1, p. 90–105, 1 set. 2017. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/imr.12563/abstract>>.

KINROSS, J. M.; DARZI, A. W.; NICHOLSON, J. K. Gut microbiome-host interactions in health and disease. *Genome Medicine*, v. 3, n. 3, p. 14, 4 mar. 2011. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3092099/>>. Acesso em: 23 abr. 2018.

KNIGHT, D. R. *et al.* Diversity and Evolution in the Genome of Clostridium difficile. *Clinical Microbiology Reviews*, v. 28, n. 3, p. 721–741, 1 jul. 2015. Disponível em: <<https://cmr.asm.org/content/28/3/721>>. Acesso em: 1 jun. 2019.

KOENIGSKNECHT, M. J. *et al.* Dynamics and establishment of Clostridium difficile infection in the murine gastrointestinal tract. *Infection and Immunity*, v. 83, n. 3, p. 934–941, mar. 2015.

KUEHNE, S. A. *et al.* Importance of Toxin A, Toxin B, and CDT in Virulence of an Epidemic Clostridium difficile Strain. *The Journal of Infectious Diseases*, v. 209, n. 1, p. 83–86, 1 jan. 2014. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3864386/>>. Acesso em: 7 fev. 2019.

KUEHNE, S. A. *et al.* The role of toxin A and toxin B in Clostridium difficile infection. *Nature*, v. 467, n. 7316, p. 711–713, out. 2010. Disponível em: <<http://www.nature.com/articles/nature09397>>. Acesso em: 19 jul. 2018.

KUMAR, M. *et al.* Cholesterol-lowering probiotics as potential biotherapeutics for metabolic diseases. *Experimental Diabetes Research*, v. 2012, p. 902917, 2012.

KUMAR, V.; ABBAS, A.; ASTER, J. *Robbins & Cotran Pathologic Basis of Disease - 9th Edition*. Disponível em: <<https://www.elsevier.com/books/robbins-and-cotran-pathologic-basis-of-disease/kumar/978-1-4557-2613-4>>. Acesso em: 23 maio 2019.

KURIYAMA, A. *et al.* Metronidazole-induced central nervous system toxicity: a systematic review. *Clinical Neuropharmacology*, v. 34, n. 6, p. 241–247, dez. 2011.

LAMBERT, J. M. *et al.* Improved annotation of conjugated bile acid hydrolase superfamily members in Gram-positive bacteria. *Microbiology (Reading, England)*, v. 154, n. Pt 8, p. 2492–2500, ago. 2008.

LAMPIDIS, T.; BARKSDALE, L. Park-Williams number 8 strain of *Corynebacterium diphtheriae*. *Journal of Bacteriology*, v. 105, n. 1, p. 77–85, jan. 1971.

LANIS, J. M. *et al.* Clostridium difficile 027/BI/NAP1 Encodes a Hypertoxic and Antigenically Variable Form of TcdB. *PLoS Pathogens*, v. 9, n. 8, p. e1003523, 1 ago. 2013. Disponível em: <<http://dx.plos.org/10.1371/journal.ppat.1003523>>. Acesso em: 1 jun. 2019.

LAWSON, P. A. *et al.* Reclassification of Clostridium difficile as Clostridioides difficile (Hall and O’Toole 1935) Prévot 1938. *Anaerobe*, v. 40, p. 95–99, ago. 2016.

LAWSON, P. A.; RAINEY, F. A. Proposal to restrict the genus Clostridium Prazmowski to Clostridium butyricum and related species. *International Journal of Systematic and Evolutionary Microbiology*, v. 66, n. 2, p. 1009–1016, fev. 2016.

LAXMINARAYAN, R. *et al.* Access to effective antimicrobials: a worldwide challenge. *Lancet (London, England)*, v. 387, n. 10014, p. 168–175, 9 jan. 2016.

LESSA, F. C. *et al.* Burden of Clostridium difficile infection in the United States. *The New England Journal of Medicine*, v. 372, n. 9, p. 825–834, 26 fev. 2015.

LEVY, M. *et al.* Dysbiosis and the immune system. *Nature Reviews Immunology*, v. 17, n. 4, p. 219–232, abr. 2017. Disponível em: <<https://www.nature.com/articles/nri.2017.7>>. Acesso em: 20 maio 2019.

LEWIS, B. B.; CARTER, R. A.; PAMER, E. G. Bile acid sensitivity and in vivo virulence of clinical Clostridium difficile isolates. *Anaerobe*, ClostPath 2015. v. 41, p. 32–36, out. 2016. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1075996416300634>>. Acesso em: 16 maio 2017.

LI, M. *et al.* Symbiotic gut microbes modulate human metabolic phenotypes. *Proceedings of the National Academy of Sciences of the United States of America*, v. 105, n. 6, p. 2117–2122, 12 fev. 2008.

LI, S. *et al.* Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nature biotechnology*, v. 32, n. 9, p. 915–925, 2014. Disponível em: <<http://www.nature.com/articles/nbt.2972>>. Acesso em: 20 jul. 2015.

LINDNER, C. *et al.* Diversification of memory B cells drives the continuous adaptation of secretory antibodies to gut microbiota. *Nature Immunology*, v. 16, n. 8, p. 880–888, ago. 2015.

LOO, V. G. *et al.* A Predominantly Clonal Multi-Institutional Outbreak of Clostridium difficile–Associated Diarrhea with High Morbidity and Mortality. *New England Journal of Medicine*, v. 353, n. 23, p. 2442–2449, 8 dez. 2005. Disponível em: <<https://doi.org/10.1056/NEJMoa051639>>. Acesso em: 1 jun. 2019.

LOUIE, T. J. How should we respond to the highly toxogenic NAP1/ribotype 027 strain of Clostridium difficile? *CMAJ: Canadian Medical Association Journal*, v. 173, n. 9, p. 1049–1050, 25 out. 2005. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1266328/>>. Acesso em: 1 jun. 2019.

LOWY, I. *et al.* Treatment with monoclonal antibodies against Clostridium difficile toxins. *The New England Journal of Medicine*, v. 362, n. 3, p. 197–205, 21 jan. 2010.

MADDUR, M. *et al.* Autoimmunity as a Predisposition for Infectious Diseases. *PLoS pathogens*, v. 6, p. e1001077, 4 nov. 2010.

MANDLIK, A. *et al.* *Corynebacterium diphtheriae* employs specific minor pilins to target human pharyngeal epithelial cells. *Molecular microbiology*, v. 64, n. 1, p. 111–124, abr. 2007. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2844904/>>. Acesso em: 24 jun. 2019.

MANDLIK, A. *et al.* Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends in Microbiology*, v. 16, n. 1, p. 33–40, jan. 2008.

MANI, N.; DUPUY, B. Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. *Proceedings of the National Academy of Sciences of the United States of America*, v. 98, n. 10, p. 5844–5849, 8 maio 2001. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC33301/>>. Acesso em: 1 jun. 2019.

MARÍA, R. R. *et al.* The Impact of Bioinformatics on Vaccine Design and Development. In: AFRIN, F.; HEMEG, H.; OZBAK, H. (Org.). *Vaccines*. [S.l.]: InTech, 2017. Disponível em: <<http://www.intechopen.com/books/vaccines/the-impact-of-bioinformatics-on-vaccine-design-and-development>>. Acesso em: 6 abr. 2019.

MARSTON, H. D. *et al.* Antimicrobial Resistance. *JAMA*, v. 316, n. 11, p. 1193–1204, 20 set. 2016. Disponível em: <<https://jamanetwork.com/journals/jama/fullarticle/2553454>>. Acesso em: 3 jun. 2019.

MARTIN, F.-P. J. *et al.* A top-down systems biology view of microbiome-mammalian metabolic interactions in a mouse model. *Molecular Systems Biology*, v. 3, 22 maio 2007. Disponível em: <<http://msb.embopress.org/cgi/doi/10.1038/msb4100153>>. Acesso em: 24 abr. 2018.

MATAMOUROS, S.; ENGLAND, P.; DUPUY, B. *Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC. *Molecular Microbiology*, v. 64, n. 5, p. 1274–1288, jun. 2007.

MATTOS-GUARALDI, A. L. *et al.* Diphtheria remains a threat to health in the developing world--an overview. *Memorias Do Instituto Oswaldo Cruz*, v. 98, n. 8, p. 987–993, dez. 2003.

MCDONALD, L CLIFFORD *et al.* Clinical Practice Guidelines for *Clostridium difficile* Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clinical Infectious Diseases*, v. 66, n. 7, p. e1–e48, 19 mar. 2018. Disponível em: <<https://academic.oup.com/cid/article/66/7/e1/4855916>>. Acesso em: 31 maio 2019.

MCDONALD, LAWRENCE CLIFFORD. Effects of short- and long-course antibiotics on the lower intestinal microbiome as they relate to traveller's diarrhea. *Journal of Travel Medicine*, v. 24, n. suppl_1, p. S35–S38, abr. 2017. Disponível em: <<https://academic.oup.com/jtm/article-lookup/doi/10.1093/jtm/taw084>>. Acesso em: 19 jul. 2018.

MCKEE, R. W. *et al.* The second messenger cyclic Di-GMP regulates *Clostridium difficile* toxin production by controlling expression of sigD. *Journal of Bacteriology*, v. 195, n. 22, p. 5174–5185, nov. 2013.

MINA, N. V. *et al.* Canada's first case of a multidrug-resistant *Corynebacterium diphtheriae* strain, isolated from a skin abscess. *Journal of Clinical Microbiology*, v. 49, n. 11, p. 4003–4005, nov. 2011.

MISHRA, A. K. *et al.* Differential Arabinan Capping of Lipoarabinomannan Modulates Innate Immune Responses and Impacts T Helper Cell Differentiation. *Journal of Biological Chemistry*, v. 287, n. 53, p. 44173–44183, 28 dez. 2012. Disponível em: <<http://www.jbc.org/content/287/53/44173>>. Acesso em: 9 maio 2019.

MISHRA, B. *et al.* *Corynebacterium diphtheriae* endocarditis--surgery for some but not all! *Asian Cardiovascular & Thoracic Annals*, v. 13, n. 2, p. 119–126, jun. 2005.

MITAMURA, T. *et al.* Structure-Function Analysis of the Diphtheria Toxin Receptor Toxin Binding Site by Site-directed Mutagenesis. *Journal of Biological Chemistry*, v. 272, n. 43, p. 27084–27090, 24 out. 1997. Disponível em: <<http://www.jbc.org/content/272/43/27084>>. Acesso em: 7 maio 2019.

MÖLLER, J. *et al.* Proteomics of diphtheria toxoid vaccines reveals multiple proteins that are immunogenic and may contribute to protection of humans against *Corynebacterium diphtheriae*. *Vaccine*, 26 abr. 2019. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0264410X19305353>>. Acesso em: 1 maio 2019.

MONTALTO, M. *et al.* Intestinal microbiota and its functions. *Digestive and Liver Disease Supplements*, v. 3, n. 2, p. 30–34, jul. 2009. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S1594580409600164>>. Acesso em: 18 abr. 2018.

MOREIRA, L. DE O. *et al.* Effects of Iron Limitation on Adherence and Cell Surface Carbohydrates of *Corynebacterium diphtheriae* Strains. *Applied and Environmental Microbiology*, v. 69, n. 10, p. 5907–5913, 1 out. 2003. Disponível em: <<https://aem.asm.org/content/69/10/5907>>. Acesso em: 24 jun. 2019.

MOREIRA, L. O.; MATTOS-GUARALDI, A. L.; ANDRADE, A. F. B. Novel lipoarabinomannan-like lipoglycan (CdiLAM) contributes to the adherence of *Corynebacterium diphtheriae* to epithelial cells. *Archives of Microbiology*, v. 190, n. 5, p. 521–530, 1 nov. 2008. Disponível em: <<https://doi.org/10.1007/s00203-008-0398-y>>. Acesso em: 27 mar. 2019.

MOSER, C. *et al.* Influenza virosomes as vaccine adjuvant and carrier system. *Expert Review of Vaccines*, v. 12, n. 7, p. 779–791, jul. 2013.

MULLISH, B. H. *et al.* Microbial bile salt hydrolases mediate the efficacy of faecal microbiota transplant in the treatment of recurrent *Clostridioides difficile* infection. *Gut*, p. gutjnl-2018-317842, 11 fev. 2019. Disponível em: <<https://gut.bmj.com/content/early/2019/02/11/gutjnl-2018-317842>>. Acesso em: 29 maio 2019.

MUTO, C. A. *et al.* Control of an outbreak of infection with the hypervirulent *Clostridium difficile* BI strain in a university hospital using a comprehensive “bundle” approach. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, v. 45, n. 10, p. 1266–1273, 15 nov. 2007.

MUTTAIYAH, S. *et al.* *Corynebacterium diphtheriae* endocarditis: a case series and review of the treatment approach. *International Journal of Infectious Diseases*, v. 15, n. 9, p. e584–e588, set. 2011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1201971211000920>>. Acesso em: 24 jun. 2019.

NAGLICH, J. G. *et al.* Expression cloning of a diphtheria toxin receptor: identity with a heparin-binding EGF-like growth factor precursor. *Cell*, v. 69, n. 6, p. 1051–1061, 12 jun. 1992.

NEEFJES, J. *et al.* Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews. Immunology*, v. 11, n. 12, p. 823–836, 11 nov. 2011.

NELSON, R. L.; SUDA, K. J.; EVANS, C. T. Antibiotic treatment for *Clostridium difficile*-associated diarrhoea in adults. *Cochrane Database of Systematic Reviews*, n. 3, 2017. Disponível em: <<https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD004610.pub5/abstract>>. Acesso em: 24 maio 2019.

NG, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature*, v. 502, n. 7469, p. 96–99, 3 out. 2013.

O'FLAHERTY, S. *et al.* The Lactobacillus Bile Salt Hydrolase Repertoire Reveals Niche-Specific Adaptation. *mSphere*, v. 3, n. 3, p. e00140-18, 27 jun. 2018. Disponível em: <<https://msphere.asm.org/content/3/3/e00140-18>>. Acesso em: 2 out. 2018.

OFORI, E. *et al.* Community-acquired *Clostridium difficile*: epidemiology, ribotype, risk factors, hospital and intensive care unit outcomes, and current and emerging therapies. *The Journal of Hospital Infection*, v. 99, n. 4, p. 436–442, ago. 2018.

OLIVEIRA, A. *et al.* Insight of Genus *Corynebacterium*: Ascertaining the Role of Pathogenic and Non-pathogenic Species. *Frontiers in Microbiology*, v. 8, 2017. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fmicb.2017.01937/full>>. Acesso em: 21 jun. 2019.

OTT, L. Adhesion properties of toxigenic corynebacteria. *microbiology 2018, Vol. 4, Pages 85-103*, 1 fev. 2018. Disponível em: <<http://www.aimspress.com/article/10.3934/microbiol.2018.1.85>>. Acesso em: 8 maio 2019.

OTT, L. *et al.* Analysis of *Corynebacterium diphtheriae* macrophage interaction: Dispensability of corynomycolic acids for inhibition of phagolysosome maturation and identification of a new gene involved in synthesis of the corynomycolic acid layer. *PLoS One*, v. 12, n. 7, p. e0180105, 2017.

OTT, L. *et al.* Strain-specific differences in pili formation and the interaction of *Corynebacterium diphtheriae* with host cells. *BMC Microbiology*, v. 10, n. 1, p. 257, 2010. Disponível em: <<http://bmcmicrobiol.biomedcentral.com/articles/10.1186/1471-2180-10-257>>. Acesso em: 6 abr. 2019.

OTT, L.; BURKOVSKI, A. Toxigenic *Corynebacteria*: Adhesion, Invasion and Host Response. In: BURKOVSKI, A. (Org.). *Corynebacterium diphtheriae and Related Toxigenic Species: Genomics, Pathogenicity and Applications*. Dordrecht: Springer Netherlands, 2014. p. 143–170. Disponível em: <https://doi.org/10.1007/978-94-007-7624-1_8>. Acesso em: 6 abr. 2019.

PANDEY, R. K.; BHATT, T. K.; PRAJAPATI, V. K. Novel Immunoinformatics Approaches to Design Multi-epitope Subunit Vaccine for Malaria by Investigating Anopheles Salivary Protein. *Scientific Reports*, v. 8, n. 1, dez. 2018. Disponível em: <<http://www.nature.com/articles/s41598-018-19456-1>>. Acesso em: 6 abr. 2019.

PASQUALE, A. *et al.* Vaccine Adjuvants: from 1920 to 2015 and Beyond. *Vaccines*, v. 3, n. 2, p. 320–343, 16 abr. 2015. Disponível em: <<http://www.mdpi.com/2076-393X/3/2/320>>. Acesso em: 23 maio 2019.

PEET, P. L. VAN DER *et al.* Corynomycolic acid-containing glycolipids signal through the pattern recognition receptor Mincle. *Chemical Communications*, v. 51, n. 24, p. 5100–5103, 10 mar. 2015. Disponível em: <<https://pubs.rsc.org/en/content/articlelanding/2015/cc/c5cc00085h>>. Acesso em: 8 maio 2019.

PENICHE, A. G.; SAVIDGE, T. C.; DANN, S. M. Recent insights into *Clostridium difficile* pathogenesis. *Current Opinion in Infectious Diseases*, v. 26, n. 5, p. 447–453, out. 2013.

PÉPIN, J.; VALIQUETTE, L.; COSSETTE, B. Mortality attributable to nosocomial *Clostridium difficile*-associated disease during an epidemic caused by a hypervirulent strain in Quebec. *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne*, v. 173, n. 9, p. 1037–1042, 25 out. 2005.

PEREIRA, G. A. *et al.* Antimicrobial resistance among Brazilian *Corynebacterium diphtheriae* strains. *Memórias do Instituto Oswaldo Cruz*, v. 103, n. 5, p. 507–510, ago. 2008. Disponível em: <http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0074-02762008000500019&lng=en&nrm=iso&tlng=en>. Acesso em: 24 jun. 2019.

PERETZ, A. *et al.* *Clostridium difficile* Infection: Associations with Chemotherapy, Radiation Therapy, and Targeting Therapy Treatments. *Current Medicinal Chemistry*, v. 23, n. 39, p. 4442–4449, 2016.

PITUCH, H. *et al.* Variable flagella expression among clonal toxin A–/B+ *Clostridium difficile* strains with highly homogeneous flagellin genes. *Clinical Microbiology and Infection*, v. 8, n. 3, p. 187–188, 1 mar. 2002. Disponível em: <[https://www.clinicalmicrobiologyandinfection.com/article/S1198-743X\(14\)62789-0/abstract](https://www.clinicalmicrobiologyandinfection.com/article/S1198-743X(14)62789-0/abstract)>. Acesso em: 1 jun. 2019.

PLOTKIN, S. History of vaccination. *Proceedings of the National Academy of Sciences of the United States of America*, v. 111, n. 34, p. 12283–12287, 26 ago. 2014. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4151719/>>. Acesso em: 20 maio 2019.

PLOTKIN, S. A.; ORENSTEIN, W. A.; OFFIT, P. A. *Vaccines*. 6th. ed. [S.l.]: Philadelphia: Elsevier, 2012.

PURCELL, E. B. *et al.* Cyclic diguanylate inversely regulates motility and aggregation in *Clostridium difficile*. *Journal of Bacteriology*, v. 194, n. 13, p. 3307–3316, jul. 2012.

PURCELL, E. B. *et al.* Regulation of Type IV Pili Contributes to Surface Behaviors of Historical and Epidemic Strains of *Clostridium difficile*. *Journal of Bacteriology*, v. 198, n. 3, p. 565–577, 1 fev. 2016. Disponível em: <<https://jb.asm.org/content/198/3/565>>. Acesso em: 1 jun. 2019.

RAPPUOLI, R. Bridging the knowledge gaps in vaccine design. *Nature Biotechnology*, v. 25, n. 12, p. 1361–1366, dez. 2007.

RAPPUOLI, R.; MALITO, E. History of Diphtheria Vaccine Development. In: BURKOVSKI, A. (Org.). *Corynebacterium diphtheriae and Related Toxigenic Species*. Dordrecht: Springer Netherlands, 2014. p. 225–238. Disponível em: <http://link.springer.com/10.1007/978-94-007-7624-1_11>. Acesso em: 6 abr. 2019.

RAWLS, J. F. *et al.* Reciprocal Gut Microbiota Transplants from Zebrafish and Mice to Germ-free Recipients Reveal Host Habitat Selection. *Cell*, v. 127, n. 2, p. 423–433, out. 2006. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0092867406012268>>. Acesso em: 2 maio 2018.

REDELINGS, M. D.; SORVILLO, F.; MASCOLA, L. Increase in Clostridium difficile–related Mortality Rates, United States, 1999–2004. *Emerging Infectious Diseases*, v. 13, n. 9, p. 1417–1419, set. 2007. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2857309/>>. Acesso em: 1 jun. 2019.

REED, S. G.; ORR, M. T.; FOX, C. B. Key roles of adjuvants in modern vaccines. *Nature Medicine*, v. 19, n. 12, p. 1597–1608, dez. 2013.

REN, J. *et al.* All 4 Bile Salt Hydrolase Proteins Are Responsible for the Hydrolysis Activity in Lactobacillus plantarum ST-III. *Journal of Food Science*, v. 76, n. 9, p. M622–M628, nov. 2011. Disponível em: <<http://doi.wiley.com/10.1111/j.1750-3841.2011.02431.x>>. Acesso em: 11 jul. 2018.

RESCIGNO, M. *et al.* Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria. *Nature Immunology*, v. 2, n. 4, p. 361–367, abr. 2001.

RIBET, D.; COSSART, P. How bacterial pathogens colonize their hosts and invade deeper tissues. *Microbes and Infection*, v. 17, n. 3, p. 173–183, 1 mar. 2015. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1286457915000179>>. Acesso em: 11 abr. 2018.

RIDLON, J. M. *et al.* Bile Acids and the Gut Microbiome. *Current opinion in gastroenterology*, v. 30, n. 3, p. 332–338, maio 2014. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4215539/>>. Acesso em: 8 maio 2018.

RIDLON, J. M. *et al.* Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes*, v. 7, n. 1, p. 22–39, 2 jan. 2016. Disponível em: <<http://dx.doi.org/10.1080/19490976.2015.1127483>>. Acesso em: 10 abr. 2017.

RIDLON, J. M.; KANG, D.-J.; HYLEMON, P. B. Bile salt biotransformations by human intestinal bacteria. *Journal of Lipid Research*, v. 47, n. 2, p. 241–259, fev. 2006.

RIEDEL, S. Edward Jenner and the history of smallpox and vaccination. *Proceedings (Baylor University. Medical Center)*, v. 18, n. 1, p. 21–25, jan. 2005. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1200696/>>. Acesso em: 20 maio 2019.

RISNES, K. R. *et al.* Antibiotic Exposure by 6 Months and Asthma and Allergy at 6 Years: Findings in a Cohort of 1,401 US Children. *American Journal of Epidemiology*, v. 173, n. 3, p. 310–318, 1 fev. 2011. Disponível em: <<https://academic.oup.com/aje/article-lookup/doi/10.1093/aje/kwq400>>. Acesso em: 18 set. 2018.

ROCHE, P. A.; FURUTA, K. The ins and outs of MHC class II-mediated antigen processing and presentation. *Nature Reviews. Immunology*, v. 15, n. 4, p. 203–216, abr. 2015.

RUPNIK, M.; WILCOX, M. H.; GERDING, D. N. Clostridium difficile infection: new developments in epidemiology and pathogenesis. *Nature Reviews. Microbiology*, v. 7, n. 7, p. 526–536, jul. 2009.

RYAN, A. *et al.* A Role for TLR4 in Clostridium difficile Infection and the Recognition of Surface Layer Proteins. *PLoS Pathogens*, v. 7, n. 6, p. e1002076, 30 jun. 2011. Disponível em: <<http://dx.plos.org/10.1371/journal.ppat.1002076>>. Acesso em: 12 jun. 2019.

SABBADINI, P. S. *et al.* Corynebacterium diphtheriae 67-72p hemagglutinin, characterized as the protein DIP0733, contributes to invasion and induction of apoptosis in HEp-2 cells. *Microbial Pathogenesis*, v. 52, n. 3, p. 165–176, mar. 2012. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0882401011002099>>. Acesso em: 6 maio 2019.

SABBADINI, P. S. *et al.* Fibrinogen binds to nontoxigenic and toxigenic Corynebacterium diphtheriae strains. *Memórias do Instituto Oswaldo Cruz*, v. 105, n. 5, p. 706–711, ago. 2010. Disponível em: <http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0074-02762010000500018&lng=en&nrm=iso&tlng=es>. Acesso em: 6 maio 2019.

SABETIAN, S. *et al.* Exploring dengue proteome to design an effective epitope-based vaccine against dengue virus. *Journal of Biomolecular Structure and Dynamics*, v. 37, n. 10, p. 2546–2563, 3 jul. 2019. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/07391102.2018.1491890>>. Acesso em: 22 maio 2019.

SALZMAN, N. H. *et al.* Enteric defensins are essential regulators of intestinal microbial ecology. *Nature Immunology*, v. 11, n. 1, p. 76–83, jan. 2010.

SANGAL, V. *et al.* A lack of genetic basis for biovar differentiation in clinically important Corynebacterium diphtheriae from whole genome sequencing. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, v. 21, p. 54–57, jan. 2014.

SANGAL, V. *et al.* Adherence and invasive properties of Corynebacterium diphtheriae strains correlates with the predicted membrane-associated and secreted proteome. *BMC Genomics*, v. 16, n. 1, p. 765, 9 out. 2015. Disponível em: <<https://doi.org/10.1186/s12864-015-1980-8>>. Acesso em: 19 nov. 2018.

SANGAL, V.; HOSKISSON, P. A. Evolution, epidemiology and diversity of Corynebacterium diphtheriae: New perspectives on an old foe. *Infection, Genetics and Evolution*, v. 43, p. 364–370, 1 set. 2016. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1567134816302489>>. Acesso em: 27 mar. 2019.

SANTOS, B. S. DOS *et al.* Application of Omics Technologies for Evaluation of Antibacterial Mechanisms of Action of Plant-Derived Products. *Frontiers in Microbiology*, v. 7, 2016. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fmicb.2016.01466/full>>. Acesso em: 16 maio 2019.

SCARIA, J. *et al.* Analysis of Ultra Low Genome Conservation in Clostridium difficile. *PLoS ONE*, v. 5, n. 12, p. e15147, 8 dez. 2010. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0015147>>. Acesso em: 12 jun. 2019.

SCHEIFER, C. *et al.* Re-emergence of Corynebacterium diphtheriae. *Médecine et Maladies Infectieuses*, 21 dez. 2018. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0399077X17309629>>. Acesso em: 27 mar. 2019.

SCHICK, J. *et al.* Toll-Like Receptor 2 and Mincle Cooperatively Sense Corynebacterial Cell Wall Glycolipids. *Infection and Immunity*, v. 85, n. 7, p. e00075-17, 1 jul. 2017. Disponível em: <<https://iai.asm.org/content/85/7/e00075-17>>. Acesso em: 14 mar. 2019.

SCHILLER, J.; GROMAN, N.; COYLE, M. Plasmids in *Corynebacterium diphtheriae* and diphtheroids mediating erythromycin resistance. *Antimicrobial Agents and Chemotherapy*, v. 18, n. 5, p. 814–821, nov. 1980.

SCHNELL, L. *et al.* Semicarbazone EGA Inhibits Uptake of Diphtheria Toxin into Human Cells and Protects Cells from Intoxication. *Toxins*, v. 8, n. 7, p. 221, 15 jul. 2016. Disponível em: <<http://www.mdpi.com/2072-6651/8/7/221>>. Acesso em: 24 jun. 2019.

SCHOENEN, H. *et al.* Cutting Edge: Mincle Is Essential for Recognition and Adjuvanticity of the Mycobacterial Cord Factor and its Synthetic Analog Trehalose-Dibehenate. *The Journal of Immunology*, v. 184, n. 6, p. 2756–2760, 15 mar. 2010. Disponível em: <<http://www.jimmunol.org/content/184/6/2756>>. Acesso em: 9 maio 2019.

SCHWAN, C. *et al.* Clostridium difficile toxin CDT hijacks microtubule organization and reroutes vesicle traffic to increase pathogen adherence. *Proceedings of the National Academy of Sciences of the United States of America*, v. 111, n. 6, p. 2313–2318, 11 fev. 2014. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3926047/>>. Acesso em: 1 jun. 2019.

SEBAIHIA, M. *et al.* The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nature Genetics*, v. 38, n. 7, p. 779–786, jul. 2006.

SENDER, R.; FUCHS, S.; MILO, R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell*, v. 164, n. 3, p. 337–340, 28 jan. 2016a. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0092867416000532>>. Acesso em: 14 maio 2019.

SENDER, R.; FUCHS, S.; MILO, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biology*, v. 14, n. 8, p. e1002533, 19 ago. 2016b. Disponível em: <<https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002533>>. Acesso em: 14 maio 2019.

SHELDON, E. *et al.* A phase 1, placebo-controlled, randomized study of the safety, tolerability, and immunogenicity of a *Clostridium difficile* vaccine administered with or without aluminum hydroxide in healthy adults. *Vaccine*, v. 34, n. 18, p. 2082–2091, 19 abr. 2016.

SHENDEROV, K. *et al.* Cord Factor and Peptidoglycan Recapitulate the Th17-Promoting Adjuvant Activity of Mycobacteria through Mincle/CARD9 Signaling and the Inflammasome. *The Journal of Immunology*, v. 190, n. 11, p. 5722–5730, 1 jun. 2013. Disponível em: <<http://www.jimmunol.org/content/190/11/5722>>. Acesso em: 9 maio 2019.

SHEY, R. A. *et al.* In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Scientific Reports*, v. 9, n. 1, p. 4409, dez. 2019. Disponível em: <<http://www.nature.com/articles/s41598-019-40833-x>>. Acesso em: 22 maio 2019.

SHI, Y. *et al.* Structural and Functional Alterations in the Microbial Community and Immunological Consequences in a Mouse Model of Antibiotic-Induced Dysbiosis. *Frontiers in Microbiology*, v. 9, 21 ago. 2018. Disponível em: <<https://www.frontiersin.org/article/10.3389/fmicb.2018.01948/full>>. Acesso em: 1 out. 2018.

SILVA, M. J. B. *et al.* *The Multifaceted Role of Commensal Microbiota in Homeostasis and Gastrointestinal Diseases*. Research article. Disponível em: <<https://www.hindawi.com/journals/jir/2015/321241/>>. Acesso em: 19 set. 2018.

SING, A. *et al.* Corynebacterium diphtheriae in a free-roaming red fox: case report and historical review on diphtheria in animals. *Infection*, v. 44, n. 4, p. 441–445, 1 ago. 2016. Disponível em: <<https://doi.org/10.1007/s15010-015-0846-y>>. Acesso em: 27 mar. 2019.

SKWARCZYNSKI, M.; TOTH, I. Peptide-based synthetic vaccines. *Chemical Science*, v. 7, n. 2, p. 842–854, 2016. Disponível em: <<http://xlink.rsc.org/?DOI=C5SC03892H>>. Acesso em: 6 abr. 2019.

SMITH, K.; MCCOY, K. D.; MACPHERSON, A. J. Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. *Seminars in Immunology*, v. 19, n. 2, p. 59–69, abr. 2007.

SOARES, S. C. *et al.* The Pan-Genome of the Animal Pathogen Corynebacterium pseudotuberculosis Reveals Differences in Genome Plasticity between the Biovar ovis and equi Strains. *PLoS ONE*, v. 8, n. 1, p. e53818, 14 jan. 2013. Disponível em: <<http://dx.plos.org/10.1371/journal.pone.0053818>>. Acesso em: 29 abr. 2014.

SOMMER, F. *et al.* Altered Mucus Glycosylation in Core 1 O-Glycan-Deficient Mice Affects Microbiota Composition and Intestinal Architecture. *PLOS ONE*, v. 9, n. 1, p. e85254, 9 jan. 2014. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0085254>>. Acesso em: 17 abr. 2018.

SOMMER, F. *et al.* The resilience of the intestinal microbiota influences health and disease. *Nature Reviews. Microbiology*, v. 15, n. 10, p. 630–638, out. 2017.

SOMMER, F.; BÄCKHED, F. The gut microbiota--masters of host development and physiology. *Nature Reviews. Microbiology*, v. 11, n. 4, p. 227–238, abr. 2013.

SONNENBERG, G. F. *et al.* Innate lymphoid cells promote anatomical containment of lymphoid-resident commensal bacteria. *Science (New York, N.Y.)*, v. 336, n. 6086, p. 1321–1325, 8 jun. 2012.

SONNENBURG, E. D. *et al.* A hybrid two-component system protein of a prominent human gut symbiont couples glycan sensing in vivo to carbohydrate metabolism. *Proceedings of the National Academy of Sciences*, v. 103, n. 23, p. 8834–8839, 6 jun. 2006. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.0603249103>>. Acesso em: 23 abr. 2018.

SONNENBURG, ERICA D. *et al.* Diet-induced extinction in the gut microbiota compounds over generations. *Nature*, v. 529, n. 7585, p. 212–215, 14 jan. 2016. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4850918/>>. Acesso em: 11 jul. 2018.

SORG, J. A.; SONENSHEIN, A. L. Bile salts and glycine as cogerminants for Clostridium difficile spores. *Journal of Bacteriology*, v. 190, n. 7, p. 2505–2512, abr. 2008.

SORG, J. A.; SONENSHEIN, A. L. Chenodeoxycholate is an inhibitor of Clostridium difficile spore germination. *Journal of Bacteriology*, v. 191, n. 3, p. 1115–1117, fev. 2009.

SORG, J. A.; SONENSHEIN, A. L. Inhibiting the initiation of Clostridium difficile spore germination using analogs of chenodeoxycholic acid, a bile acid. *Journal of Bacteriology*, v. 192, n. 19, p. 4983–4990, out. 2010.

SPELLBERG, B.; EDWARDS, J. E. Type 1/Type 2 immunity in infectious diseases. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, v. 32, n. 1, p. 76–102, jan. 2001.

- SPIGAGLIA, P.; MASTRANTONIO, P. Molecular Analysis of the Pathogenicity Locus and Polymorphism in the Putative Negative Regulator of Toxin Production (TcdC) among *Clostridium difficile* Clinical Isolates. *Journal of Clinical Microbiology*, v. 40, n. 9, p. 3470–3475, set. 2002. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC130716/>>. Acesso em: 1 jun. 2019.
- STABLER, R. A. *et al.* Comparative phylogenomics of *Clostridium difficile* reveals clade specificity and microevolution of hypervirulent strains. *Journal of Bacteriology*, v. 188, n. 20, p. 7297–7305, out. 2006.
- STAGLIAR, I. The power of OMICs. *Biochemical and Biophysical Research Communications*, v. 479, n. 4, p. 607–609, out. 2016. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0006291X16315625>>. Acesso em: 16 maio 2019.
- SUDARSAN, N. *et al.* Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science (New York, N.Y.)*, v. 321, n. 5887, p. 411–413, 18 jul. 2008.
- SURAWICZ, C. M. *et al.* Guidelines for Diagnosis, Treatment, and Prevention of *Clostridium difficile* Infections. *The American Journal of Gastroenterology*, v. 108, n. 4, p. 478–498, abr. 2013. Disponível em: <<https://www.nature.com/articles/ajg20134>>. Acesso em: 23 jan. 2019.
- TAKAHASHI, K. *et al.* Epigenetic Control of the Host Gene by Commensal Bacteria in Large Intestinal Epithelial Cells. *Journal of Biological Chemistry*, v. 286, n. 41, p. 35755–35762, 14 out. 2011. Disponível em: <<http://www.jbc.org/lookup/doi/10.1074/jbc.M111.271007>>. Acesso em: 22 maio 2019.
- TAKEUCHI, O. *et al.* Differential Roles of TLR2 and TLR4 in Recognition of Gram-Negative and Gram-Positive Bacterial Cell Wall Components. *Immunity*, v. 11, n. 4, p. 443–451, 1 out. 1999. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1074761300801193>>. Acesso em: 27 mar. 2019.
- TANNAHILL, G. M. *et al.* Succinate is an inflammatory signal that induces IL-1 β through HIF-1 α . *Nature*, v. 496, n. 7444, p. 238–242, 11 abr. 2013.
- TAO, X. *et al.* Iron, DtxR, and the regulation of diphtheria toxin expression. *Molecular Microbiology*, v. 14, n. 2, p. 191–197, out. 1994.
- TAUCH, A.; SANDBOTE, J. The Family Corynebacteriaceae. In: ROSENBERG, E. *et al.* (Org.). *The Prokaryotes*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. p. 239–277. Disponível em: <http://link.springer.com/10.1007/978-3-642-30138-4_187>. Acesso em: 24 jun. 2019.
- THAISS, C. A. *et al.* The microbiome and innate immunity. *Nature*, v. 535, n. 7610, p. 65–74, jul. 2016. Disponível em: <<https://www.nature.com/articles/nature18847>>. Acesso em: 20 maio 2019.
- THERIOT, C. M. *et al.* Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. *Nature communications*, v. 5, 2014. Disponível em: <<http://www.nature.com/ncomms/2014/140120/ncomms4114/full/ncomms4114.html>>. Acesso em: 15 dez. 2014.

THERIOT, C. M.; BOWMAN, A. A.; YOUNG, V. B. Antibiotic-Induced Alterations of the Gut Microbiota Alter Secondary Bile Acid Production and Allow for *Clostridium difficile* Spore Germination and Outgrowth in the Large Intestine. *mSphere*, v. 1, n. 1, fev. 2016.

TILEY, S. M. *et al.* Infective Endocarditis Due to Nontoxigenic *Corynebacterium diphtheriae*: Report of Seven Cases and Review. *Clinical Infectious Diseases*, v. 16, n. 2, p. 271–275, 1 fev. 1993. Disponível em: <<https://academic.oup.com/cid/article/16/2/271/287435>>. Acesso em: 6 abr. 2019.

TON-THAT, H.; MARRAFFINI, L. A.; SCHNEEWIND, O. Sortases and pilin elements involved in pilus assembly of *Corynebacterium diphtheriae*. *Molecular Microbiology*, v. 53, n. 1, p. 251–261, jul. 2004.

TON-THAT, H.; SCHNEEWIND, O. Assembly of pili on the surface of *Corynebacterium diphtheriae*. *Molecular Microbiology*, v. 50, n. 4, p. 1429–1438, nov. 2003.

TORRES, L. DE F. C. *et al.* Multiplex polymerase chain reaction to identify and determine the toxigenicity of *Corynebacterium* spp with zoonotic potential and an overview of human and animal infections. *Memórias do Instituto Oswaldo Cruz*, v. 108, n. 3, p. 272–279, maio 2013. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0074-02762013000300272&lng=en&tlng=en>. Acesso em: 24 jun. 2019.

TREMAROLI, V.; BÄCKHED, F. Functional interactions between the gut microbiota and host metabolism. *Nature*, v. 489, n. 7415, p. 242–249, 13 set. 2012.

TRIFAN, A. *et al.* Proton pump inhibitors therapy and risk of *Clostridium difficile* infection: Systematic review and meta-analysis. *World Journal of Gastroenterology*, v. 23, n. 35, p. 6500–6515, 21 set. 2017.

TROST, E. *et al.* Pangenomic study of *Corynebacterium diphtheriae* that provides insights into the genomic diversity of pathogenic isolates from cases of classical diphtheria, endocarditis, and pneumonia. *Journal of Bacteriology*, v. 194, n. 12, p. 3199–3215, jun. 2012a.

TROST, E. *et al.* Pangenomic Study of *Corynebacterium diphtheriae* That Provides Insights into the Genomic Diversity of Pathogenic Isolates from Cases of Classical Diphtheria, Endocarditis, and Pneumonia. *Journal of Bacteriology*, v. 194, n. 12, p. 3199–3215, 15 jun. 2012b. Disponível em: <<https://jb.asm.org/content/194/12/3199>>. Acesso em: 28 mar. 2019.

TURNBAUGH, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature*, v. 457, n. 7228, p. 480–484, 22 jan. 2009. Disponível em: <<http://www.nature.com/doi/10.1038/nature07540>>. Acesso em: 23 abr. 2018.

UBEDA, C. *et al.* Vancomycin-resistant *Enterococcus* domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. *The Journal of Clinical Investigation*, v. 120, n. 12, p. 4332–4341, dez. 2010.

URDANETA, V.; CASADESÚS, J. Interactions between Bacteria and Bile Salts in the Gastrointestinal and Hepatobiliary Tracts. *Frontiers in Medicine*, v. 4, 2017. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fmed.2017.00163/full>>. Acesso em: 2 maio 2018.

VAISHNAVA, S. *et al.* The antibacterial lectin RegIII γ promotes the spatial segregation of microbiota and host in the intestine. *Science (New York, N.Y.)*, v. 334, n. 6053, p. 255–258, 14 out. 2011.

VAN BOECKEL, T. P. *et al.* Global trends in antimicrobial use in food animals. *Proceedings of the National Academy of Sciences of the United States of America*, v. 112, n. 18, p. 5649–5654, 5 maio 2015.

VAN LEEUWEN, H. C. *et al.* Clostridium difficile TcdC protein binds four-stranded G-quadruplex structures. *Nucleic Acids Research*, v. 41, n. 4, p. 2382–2393, 1 fev. 2013.

VEDANTAM, G. *et al.* Clostridium difficile infection: toxins and non-toxin virulence factors, and their contributions to disease establishment and host response. *Gut Microbes*, v. 3, n. 2, p. 121–134, abr. 2012.

VIMR, E. R. *et al.* Diversity of microbial sialic acid metabolism. *Microbiology and molecular biology reviews: MMBR*, v. 68, n. 1, p. 132–153, mar. 2004.

WAGNER, K. S. *et al.* Diphtheria in the United Kingdom, 1986–2008: the increasing role of *Corynebacterium ulcerans*. *Epidemiology and Infection*, v. 138, n. 11, p. 1519–1530, nov. 2010. Disponível em: <https://www.cambridge.org/core/product/identifiser/S0950268810001895/type/journal_article>. Acesso em: 24 jun. 2019.

WAHLSTRÖM, A. *et al.* Crosstalk between Bile Acids and Gut Microbiota and Its Impact on Farnesoid X Receptor Signalling. *Digestive Diseases*, v. 35, n. 3, p. 246–250, 2017. Disponível em: <<https://www.karger.com/Article/FullText/450982>>. Acesso em: 8 maio 2018.

WAHLSTRÖM, A. *et al.* Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. *Cell Metabolism*, v. 24, n. 1, p. 41–50, 12 jul. 2016. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1550413116302236>>. Acesso em: 3 maio 2017.

WEAVER, C. T. *et al.* The Th17 pathway and inflammatory diseases of the intestines, lungs, and skin. *Annual Review of Pathology*, v. 8, p. 477–512, 24 jan. 2013.

WHO. *2018 Assessment report of the Global Vaccine Action Plan. Strategic Advisory Group of Experts on Immunization.* Geneva: World Health Organization; Disponível em: <https://www.who.int/immunization/global_vaccine_action_plan/sage_assessment_reports/en/>. Acesso em: 17 maio 2019a.

WHO. *Antimicrobial resistance: global report on surveillance.* Disponível em: <<http://www.who.int/drugresistance/documents/surveillancereport/en/>>. Acesso em: 16 maio 2019.

WHO. *Immunization coverage.* Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/immunization-coverage>>. Acesso em: 17 maio 2019.

WHO. *Management of the child with a serious infection or severe malnutrition.* Disponível em: <https://www.who.int/maternal_child_adolescent/documents/fch_cah_00_1/en/>. Acesso em: 23 jun. 2019.

WHO. *Progress and Challenges with achieving Universal Immunization Coverage. 2017 WHO/UNICEF Estimates of National immunization Coverage.* [S.l: s.n.]. Disponível em: <https://www.who.int/immunization/monitoring_surveillance/who-immuniz.pdf?ua=1>. , 2018b

WHO. *The evolving threat of antimicrobial resistance: options for action.* Disponível em: <<https://apps.who.int/iris/handle/10665/44812>>. Acesso em: 6 jun. 2019.

WHO. *WHO World Health Organization: Immunization, Vaccines And Biologicals. Vaccine preventable diseases Vaccines monitoring system 2018 Global Summary Reference Time Series: DIPHTHERIA.* Disponível em: <http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tsincidediphtheria.html>. Acesso em: 2 maio 2019c.

WHO. World Health Organization. Diphtheria anti-toxin (DAT) supply issues: brief review and proposition. SAGE meeting, 2017. 2017a. , p. http://www.who.int/immunization/sage/meetings/2017/april/3_Diphtheria_anti_toxin.pdf?ua=1.

WHO. *World Health Organization Diphtheria Vaccine: WHO Position Paper. Weekly epidemiological record 2017;31:417–436.* Disponível em: <https://scholar.google.com/scholar_lookup?title=WHO%20position%20paper&author=World%20Health%20Organisation%20Diphtheria%20vaccine&publication_year=2017&pages=417-436>. Acesso em: 2 maio 2019b.

WILSON, K. H.; KENNEDY, M. J.; FEKETY, F. R. Use of sodium taurocholate to enhance spore recovery on a medium selective for *Clostridium difficile*. *Journal of Clinical Microbiology*, v. 15, n. 3, p. 443–446, mar. 1982. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC272115/>>. Acesso em: 19 jul. 2018.

WITTCHEN, M. *et al.* Transcriptome sequencing of the human pathogen *Corynebacterium diphtheriae* NCTC 13129 provides detailed insights into its transcriptional landscape and into DtxR-mediated transcriptional regulation. *BMC Genomics*, v. 19, n. 1, p. 82, 25 jan. 2018. Disponível em: <<https://doi.org/10.1186/s12864-018-4481-8>>. Acesso em: 24 jun. 2019.

YARZA, P. *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology*, v. 12, n. 9, p. 635–645, set. 2014. Disponível em: <<https://www.nature.com/articles/nrmicro3330>>. Acesso em: 29 maio 2019.

YUTIN, N.; GALPERIN, M. Y. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environmental Microbiology*, v. 15, n. 10, p. 2631–2641, out. 2013.

ZASADA, A. A.; BACZEWSKA-REJ, M.; WARDAK, S. An increase in non-toxigenic *Corynebacterium diphtheriae* infections in Poland — molecular epidemiology and antimicrobial susceptibility of strains isolated from past outbreaks and those currently circulating in Poland. *International Journal of Infectious Diseases*, v. 14, n. 10, p. e907–e912, out. 2010. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1201971210024355>>. Acesso em: 24 jun. 2019.

ZHANG, L. Multi-epitope vaccines: a promising strategy against tumors and viral infections. *Cellular & Molecular Immunology*, v. 15, n. 2, p. 182–184, fev. 2018. Disponível em: <<http://www.nature.com/doifinder/10.1038/cmi.2017.92>>. Acesso em: 6 abr. 2019.