FEDERAL UNIVERSITY OF MINAS GERAIS
INSTITUTE OF BIOLOGICAL SCIENCES
DEPARTMENT OF GENERAL BIOLOGY
GRADUATE PROGRAM OF BIOINFORMATICS

INTEGRATIVE *IN SILICO* APPROACHES FOR THERAPEUTIC TARGET IDENTIFICATION IN THE HUMAN PATHOGEN *CORYNEBACTERIUM DIPHTHERIAE*

Ph.D. STUDENT: Syed Babar Jamal Bacha
SUPERVISOR: Prof. Dr. Vasco Ariston de Carvalho Azevedo
CO-SUPERVISOR: Prof. Dr. Artur Luiz da Costa da Silva
CO-SUPERVISOR: Dr. Sandeep Tiwari

BELO HORIZONTE - MG
March – 2018

Syed Babar Jamal Bacha

INTEGRATIVE *IN SILICO* APPROACHES FOR THERAPEUTIC TARGET IDENTIFICATION IN THE HUMAN PATHOGEN *CORYNEBACTERIUM DIPHTHERIAE*

This thesis is presented as partial requirement for obtaining a Doctorate Degree in Bioinformatics by the Postgraduate program in Bioinformatics at the Institute of Biological Sciences at Federal University of Minas Gerais.

Ph.D. STUDENT: Syed Babar Jamal Bacha

SUPERVISOR: Prof. Dr. Vasco Ariston de Carvalho Azevedo

CO-SUPERVISOR: Prof. Dr. Artur Luiz da Costa da Silva

CO-SUPERVISOR: Dr. Sandeep Tiwari

FEDERAL UNIVERSITY OF MINAS GERAIS

INSTITUTE OF BIOLOGICAL SCIENCES

BELO HORIZONTE - MG

043 Bacha, Syed Babar Jamal.
Integrative in silico approaches for therapeutic target identification in the human pathogen corynebacterium diphtheriae [manuscrito] / Syed Babar Jamal Bacha. – 2018.

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

Supervisor: Prof. Dr. Vasco Ariston de Carvalho Azevedo. Co-supervisor: Prof. Dr. Artur Luiz da Costa da Silva; Co-supervisor: Dr. Sandeep Tiwar.
Thesis (Doctorate) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas.

1. Bioinformática - Teses. 2. Corynebacterium diphtheriae - Teses. 3. Resistência a Medicamentos. 4. Genomas - Teses. I. Azevedo, Vasco Ariston de Carvalho. II. Silva, Artur Luiz da Costa da. III. Tiwar, Sandeep. IV. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. V. Título.

CDU: 573:004

Ficha catalográfica elaborada pela Biblioteca do Instituto de Ciências Biológicas

Universidade Federal de Minas Gerais
Instituto de Ciências Biológicas
Programa Interunidades de Pós-Graduação em Bioinformática da UFMG

ATA DA DEFESA DE TESE

Syed Babar Jamal Bacha

91/2018
entrada
1º/2014
CPF:
022.739.256-62

Às treze horas do dia **23 de março de 2018**, 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: **"INTEGRATIVE IN SILICO APPROACHES FOR THERAPEUTIC TARGET IDENTIFICATION IN THE HUMAN PATHOGEN CORYNEBACTERIUM DIPHTHERIAE"**, requisito para obtenção do grau de Doutor em **Bioinformática.** Abrindo a sessão, o Presidente da Comissão, **Dr. Vasco Ariston de Carvalho Azevedo**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa do candidato. Logo após a Comissão se reuniu, sem a presença do candidato e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	Instituição	CPF	Indicação
Dr. Vasco Ariston de Carvalho Azevedo	UFMG	283.141.225-49	APROVADO
Dr. Sandeep Tiwari	UFMG	020977186-04	APROVADO
Dra. Ljubica Tasic	UNICAMP	218.779.798-23	APROVADO
Dr. Raghuvir Krishnaswamy Arni	UNESP	13871039896	APROVADO
Dr. Lucas Bleicher	UFMG	813.528.593-00	APROVADO
Dr. Douglas Eduardo Valente Pires	FIOCRUZ/CPQRR	079.288.706-30	APROVADO

Pelas indicações, o candidato foi considerado: Aprovado
O resultado final foi comunicado publicamente ao candidato 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, 23 de março de 2018.

Dr. Vasco Ariston de Carvalho Azevedo - Orientador _____
Dr. Sandeep Tiwari - Coorientador _____
Dra. Ljubica Tasic _____
Dr. Raghuvir Krishnaswamy Arni _____
Dr. Lucas Bleicher _____
Dr. Douglas Eduardo Valente Pires _____

In the name of Allah, the Beneficent, the Merciful,

(Al-Quran)

Dedication

I dedicate this work, to my father **Mr. Syed Qayyum Bcaha** & My mother **Mrs. Seema Qayyum,** for being my inspiration of life, my Brothers (**Syed Yasar Jamal Bacha & Syed Fahad Jamal Bacha**), my Family, friends, and every single person who helped me in this work.

Contents

Acknowledgement

All praises for **ALLAH ALMIGHTY**, the Most Merciful, the Most Gracious, Who helped and blessed me all the way along till the completion of my project. Without His grace and mercy, I am nothing. Countless salutations are upon the **HOLY PROPHET HAZRAT MUHAMMAD** (**peace be upon him**), the city of knowledge which guided his UMMAH to seek knowledge from the cradle to grave.

At this Juncture, I thank my parents **Mr. & Mrs. Syed Qayyum Bacha** whose selfless sacrificial life and their significant efforts and unceasing prayers have enabled me to reach the present position in life. My deepest gratitude goes to my family for their unflagging love support throughout my life; this thesis is simply impossible without them. My thanks are tendered to my loving, caring, and sincere parents, brothers and family members especially **Mr. & Mrs. Syed Yasar Jamal Bacha**, **Mr. & Mrs. Syed Fahad Jamal Bacha** and **Syed Muhammad Sannan** for their ever-lasting cooperation and prayers.

I am highly indebted to my Supervisors **Prof. Dr. Vasco Ariston de Carvalho Azevedo** (UFMG) & **Prof. Dr. Artur Silva** (UFPA), for creating the idea and methodology of all the research work. I deem it as my privilege to work under their able guidance. I would like to express my heartfelt gratitude and admiration to **Prof. Dr. Vasco Ariston de Carvalho Azevedo** for his steady help, guidance, and concentration. He has been endowing me with his constructive observations at every stage of this research. As well as, a special thanks to my INDO-PAK friends at UFMG. (**Dr. Sandeep Tiwari**, **Dr. Syed Shah Hassan**, **Dr. Arshad Islam**, **Dr. Shakoor Muhammad**, **Dr. Debmalya Barh**, **Muhammad Shabir Afridi, Asmat Ullah Khan** and **Arun Kumar Jaiswal**).

Gratitude to all dear colleagues from the Laboratory of Cellular and Molecular Genetics (LGCM) and LPDNA-UFPA, by making exceptional support and guidance in various parts of this work. To **Sheila Santana**, **Marcia Natalia** and **Fernanda Mangalhaes** (Office Secretaries) who have always been kind enough in resolving academic and administrative issues.

I want to pay my heartiest gratitude to my Cousins specially **Syed Saqib Jalal**, **Syed Umair Jalal**, **Syed Noor Ali Shah**, **Syed Raza Shah**, **Syed Asim Jalal**, **Syed Ali Shah** and **Syed Zafar Jamal** for always been my support in tense and pressure situations.

Thanks to all my Brazilian friends **Matheus Zocatelli**, **Felipe Braga**, **Renato Nogueira Junior**, **Leandro Carvalho**, **Gabriel Vieira**, **Thiago Abreu**, **Rodrigo Pimentel**, **Luiz Davila**, **Hugo Linhares**, **Carlos Eduardo**, **Davi do Vale**, **Gustavo Oliveira**, **Thiago Goncalves**, **Eduardo Goncalves**, **Luis Lizardo**, **Rodrigo Profeta**, **Tatiana Oliveira**, **Rachel Lima**, **Yasmim Senden**, **Fabiana Stradiotto**, **Marcia Bonata**, **Thais Abreu**, **Raysa Caroline** and **Mariana Germano** who have always supported and encouraged my decisions and always illuminated me with precious Advises. Love you all!!!

I want to acknowledge my friends that I never met in my life but their love and care through social media cannot be ignored. I appricate the discussions and chats we had during my doctorate, it helped me a lot to improve myself as human being. Thank you **Dr. Moslem Yousefi**, **Carolina Belmonte**, **Talha Shafiq** and **Faizan Liaquat** for such a nice friendship.

A special thanks to my friend and my brother **Dr. Milad Yousefi** for his guidance and help in adapting Brazil. His fruitful discussions and advices were key to keep myself on right track and fosuced on my work.

Last but not the least, I want to thank my former supervisors (**Prof. Dr. Abdul Wadood, Prof. Dr. Asif Mir, Prof. Dr. Muhammad Ilyas**) and friends/students (**Muhammad Riaz, Waqar Arshad, Arslan Sehgal, Adnan Tahir, Muhammad Naeem, Muhammad Junaid, Ashfaq ur Rehman, Adbul Nasir, Nasir Khan, Amir Fahira, Latif Roghani, Arif Mahmood** and **Atta Ullah**) for support and prayers.

I would like to thank the Federal University of Minas Gerais for excellent training; The World Academy of Sciences (TWAS) is a scientific career making opportunity and the National Council for Scientific and Technological Development (CNPq), for financial support indispensable toward all costs of my doctoral studies.

Finally, I thank all those who have helped me directly or indirectly in this work. Anyone missed in this acknowledgement is also thanked.

Thanks to All of You

Syed Babar Jamal Bacha

List of Abbreviations and Acronyms

AA	Amino Acids
BLASTp	Basic Local Alignment Search Tool (protein)
CDS	Coding Sequences
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico ("National Counsel of Technological and Scientific Development")
COGs	Database of Clusters of Orthologous Groups of proteins
Cd	*Corynebacterium diphtheriae*
3D	Three-Dimensional
DNA	Deoxyribonucleic Acid
DEG	Database of Essential Genes
FAPEMIG	Fundação de Amparo a Pesquisa do Estado de Minas Gerais (Research Support Foundation of the State of Minas Gerais)
GC	Guanine Cytosine
KDa	Kilodaltons (10^3 Da)
MIGS	Minimum Information About a Genome Sequence
NCBI	National Center for Biotechnology Information
NAS	Non-traceable Author Statement
ORF	Open Reading Frame
Bp	Base pairs
PDB	Protein Data Bank
SIGS	Standards in Genomics Sciences
SDS	Sodium Dodecyl Sulfate
TM	Transmembrane (domain)
TWAS	The World Academy of Sciences
UniProt	Universal Protein Resource
UFPA	Universidade Federal do Pará (Federal University of Pará)
UFMG	Universidade Federal de Minas Gerais (Federal University of Minas Gerais)
EC number	Enzyme Commission number

List of Figures

Abstract

Corynebacterium diphtheriae (Cd) is a gram-positive human pathogen responsible for diphtheria infection and once regarded for high mortalities worldwide. The fatality gradually decreased with improved living standards and further alleviated when many immunization programs were introduced. However, numerous drug-resistant strains emerged recently that consequently decreased the efficiency of current therapeutics and vaccines, thereby obliging the scientific community to start investigating new therapeutic targets in pathogenic microorganisms. In 2014, our research group introduced the word panmodelome for the first time in the scientific world (Hassan et al., 2014). Inspired by panmodelomics approach, in this study, our group aimed to contribute including the prediction of modelome of thirteen *C. diphtheriae* strains, using the MHOLline workflow. Considering the quality of the models and using in-house scripts, a set of 465 conserved proteins were selected by combining the results of pangenomics based on core genome and core-modelome analyses. Further, using subtractive proteomics and modelomics approaches for target identification, a set of 23 proteins were selected as essential proteins for bacteria. Considering human as a host, 8 of these proteins (glpX, nusB, rpsH, hisE, smpB, bioB, DIP1084 and DIP0983) were seen as essential and non-host homologs. These proteins were subjected to virtual screening using three different compound libraries (extracted from ZINC database, plant-derived natural compounds, and Di-terpenoid Iso-steviol derivatives). The proposed drug molecules showing favorable interactions, lowered energy values and high complementarity with the predicted targets have also been reported in the present study. Our proposed approach expedites the rapid and efficient selection of *C. diphtheriae* putative proteins for developing a broad spectrum of novel drugs and vaccines because some of these targets have already been identified and validated in other organisms. Furthermore, we adopted a different approach using same number of genomes of *C. diphtheriae* to identify drug/vaccine targets based on the druggable pocketome. As a result, we identify 10 targets in which interestingly 3 (hisE-phosphoribosyl-ATP pyrophosphatase, glpX-fructose 1,6-bisphosphatase II, and rpsH 30S ribosomal protein S8) targets were common with our first study. Further, we are working on characterization of hisE-phosphoribosyl-ATP pyrophosphatase, glpX-fructose 1,6-bisphosphatase II, and rpsH 30S ribosomal protein S8 in *C. diphtheriae* strain NCTC13129. The selection of these proteins (hisE, glpX and rpsH) were made on the bases of their identification through two different computational approaches. Our proposed approaches expedite the selection

of *C. diphtheriae* putative proteins for broad-spectrum development of novel drugs and vaccines, owing to the fact that our study is computational and need experimental validation.

Key Words: *Corynebacterium diphtheria*, pan-genome, core-modelome, putative drug and vaccine targets, druggable pocketome, computational approaches.

I. PRESENTATION

I.A. Collaborations

This work has been conducted in the Laboratory of Cellular and Molecular Genetics (LGCM), under supervision of Prof. Dr. Vasco Azevedo, Department of General Biology, Institute of Biological Sciences (ICB), Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil and Prof. Dr. Artur Silva Institute of Biological Sciences, coordinator of the Genomics and Proteomics Network of Pará (RPGP), UFPA, Belém, Pará, Brazil. Other integral partner researchers/collaborators (national and international) and their respective institutions, among others, are:

Dr. Luís G. C. Pacheco, Institute of Health Sciences, Federal University of Bahia, Salvador, Brazil.

Prof. Dr. Jan Baumbach, researcher, and group leader Computational Biology Laboratory, Department of Mathematics and Computer Science (IMADA), University of Southern Denmark (SDU), Denmark.

Dr. Debmalya Barh, researcher at the Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology (IIOAB), Nonakuri, India.

Prof. Dr. Siomar Soares, Institute of Biological and Natural Sciences, Federal University of Triangulo Mineiro, Uberaba-MG, Brasil.

Dr. Luiz Carlos Junior Alcantara, Fundação Oswaldo Cruz (FIOCRUZ), Salvador, Bahia, Brazil.

I had joined the Laboratory of Cellular and Molecular Genetics (LGCM), Department of Bioinformatics in 2014 as full-time Ph.D. student under the supervision of Prof. Vasco Ariston de Carvalho Azevedo & Prof. Artur da Costa da Silva.

The research group (LGCM) is actively involved in the characterization of the *C. pseudotuberculosis* genomes. Being the pioneer in *Corynebacterium* genomic research in Brazil, the group engaged in intensive research projects covering diverse areas of biology like genomics, transcriptomics; and in the development of vaccines and diagnostics has made the team a reference point for the study of this microorganism.

So far, the team has successfully sequenced the numbers of genome of *C. pseudotuberculosis* species, isolated from different locations around the world, biovars, and hosts, and deposited to a public database (GenBank). Besides *Corynebacterium* other species like *Campylobacter*, *Leptospira* and *Lactococcus* genome projects were also accomplished.

LGCM has provided the opportunity to participate in various internal and external on-going projects. Starting with previous experience in protein 3D structural modeling and *in silico* structure-based drug designing. This lab has strengthened me in computational skills to genome sequencing, assembly and annotation projects and have provided familiarity with the required necessary skills.

The financial support is granted by The World Academy of Sciences (TWAS) with an agreement to National Counsel for Scientific and Technological Development (CNPq).

I.B. Preface

In post-genomic era, the scientific community utilizes the available genomic data from public databases like NCBI and GOLD databanks, among others, for various purposes including *In-Silico* characterization of pathogenic and non-pathogenic microorganisms. These *In-Silico* studies usually comprises of different methodologies/approaches for therapeutic targets identification including drugs, vaccines and other biomarkers based on the information retrieved from the bacterial, parasite and viral genomes. These *in-silico* analyses have been performed to identify targets against many diseases that cause several deaths per annum.

I.B.1. *Corynebacterium diphtheriae*, Diphtheria

Corynebacterium diphtheriae is a human pathogen, responsible for causing Diphtheria. It was once a global cause of death. The mortality rates started to decrease with time in the twentieth century in countries with improved living standards and then strongly fell once after the immunization programs were introduced (Hodes 1979). Diphtheria is theoretically a lethal disease that mainly effect upper respiratory tract tissues and kills the victims slowly by suffocation. In 1884, a German physician, Edwin Klebs was able to successfully isolate the bacteria that proved to be the etiological agent of the disease. It was later confirmed that toxin production begun only after the bacteria were infected by a particular virus or a bacteriophage themselves, carrying the toxin's genetic instructions (Grabenstein 2010).

According to recent report of WHO (World Health Organization), Diphtheria remained an important health issue in countries with poor routine vaccination coverage. Over the last 11 years, the annual number of reported cases of diphtheria has remained relatively unchanged. It is estimated that almost 86% of children worldwide receive the recommended 3 doses of diphtheria-containing vaccine in the infant schedule, leaving other 14% with no or unfinished vaccination. There are compartments of unvaccinated children all around the globe. Case-fatality rates exceeding 10% have been reported, where DAT is unavailable. In regions with temperate climates, most cases occur during the cold season, whereas in warmer climates transmission takes place throughout the year. In the period 2011–2015, India had the largest total number of reported cases each year, with a 5-year total of 18 350 cases, followed by Indonesia and Madagascar with 3203 and 1633 reported cases respectively. The South-East Asia Region was the source of 55–99% of all reported cases each year during this period. The analysis further

showed a significant under-reporting of cases to WHO, particularly from the African and Eastern Mediterranean Regions. The true burden of disease is therefore likely to be greater than reported. A recent review of diphtheria epidemiology showed that among cases with information on age, the age distribution shifts and most of cases occur in adolescents and adults, reflecting the decline in incidence due to increasing vaccination coverage in children. In high occurrence countries (≥10 cases per year in ≥3 years during 2000–2015) 40% were aged >15 years while in low occurrence countries (<10 cases per year in ≥3 years during 2000–2015) 66% of cases were aged >15 years. Among cases with known vaccination status most were unvaccinated, and a lower proportion were incompletely vaccinated; very few cases had received ≥5 vaccine doses (WHO, 2016).

Till the date, diphtheria has been very effectively controlled by an efficient immunization program in developed countries (Vitek, 2006). However, in recent years, the disease has made a dramatic return, especially within Eastern Europe. The major outbreak since the beginning of mass immunization within Russia and the states that were newly independent of the former Soviet Union in the 1990s (Dittmann et al., 2000). In 2003, the genome (*C. diphtheriae* NCTC 13129) was sequenced at the Sanger Institute, which was clinical isolate related to this outbreak. It provided useful basis to identify candidate virulence factors besides the toxin itself, like iron transport systems and fimbrial proteins (Cerdeño-Táarraga et al., 2003). Availability of genomic data and emergence of new in-slico approaches are aiding more information for the identification of novel drug/vaccine.

I.B.2. Computational Biology and Bioinformatics

Bioinformatics is a newly developed interdisciplinary research area lying at the interface of chemical, biological and computational sciences. It can be used to solve biological problems and understanding of the molecular basis of biological phenomena. The information so obtained is used in analyzing behaviors of macromolecules, designing experiments to know the relationship between structure and function, obtaining detail understanding of the molecular process and developing drugs, based on chemical similarity of known drugs. Certain processes in target identification and drug development such as pangenomics, panmodelomics, subtractive Genomics, Molecular Docking etc. Nevertheless, National Institutes of Health (NIH) and

Biomedical Information Science and Technology Initiative Consortium (BISTIC) defined these terms in June 2000, as follows (Altman, 2000); also, at http://www.bisti.nih.gov.

These concepts have played a crucial role in the development of research in life sciences, performing an important part in the generation and analysis of large amounts of genomic, transcriptomic and proteomic data produced by different research groups and associated agencies (Xavier et al., 2008). Some of the distinguished institutes among others are the Laboratory of Cellular and Molecular Genetics (LGCM) of Federal University of Minas Gerais (UFMG), The Laboratory of DNA Polymorphism (LPDNA) of Federal University of Pará (UFPA), Faculty of Medicinal Sciences, State University of Rio de Janeiro (UERJ) also make part of this work with the sequencing project of the genome of several strains of *C. diphtheria*.

I.B.3. Protein Modeling and Drug Designing: An *in silico* Approach

"I have an amino acid sequence, but with unknown Structure/function" is amongst the most common complications faced by many molecular biologists in their studies. In recent years, this problem has heightened because the percentage of protein sequences in UniProtKB/TrEMBL, with a solved protein structure in the Protein Data Bank (PDB) library, dropped to 0.2% by the end of the year 2015. With the enhancements in computer algorithms for protein structure and function prediction, it addresses many problems and deliver valuable information to biologists about their proteins of interest (Roy et al., 2010; UniProt, 2009; Zhang, 2008).

Normally, the linear sequence of amino acids is derived from the genome sequencing projects. To determine the biological role of these proteins requires knowledge of their structures and functions. Experimental and computational structure determination of protein structure are the two methods, in which it first provides information about the high-resolution structure of the proteins subset, then provide valuable information for the most sequences whose structures have not yet experimentally determined (Baker and Sali, 2001). The function of the protein is determined by its structure, hence, provide useful information about protein interacts with ligands or other protein molecules. Thus, the motivation behind protein structure prediction is to gain insight into the protein's biological function (Lopez et al., 2007).

The number of protein structures is increasing day by day, but still the number of protein structures available at RCSB-PDB represents a minuscule amount when compared to the

genomic data generated proportionally. Figure-1 is showing several searchable structures per year. In this case, the structure prediction in alternative ways is indispensable. Modeling of amino acid sequence based on known structures usually consists of four steps: in the 1st step, a proper template must be identified, then target sequence is aligned with the template sequence, in the 3rd step, the model is generated and at the end its quality is measured. The best quality predicted structure is used for the further study (Marti-Renom et al., 2000).

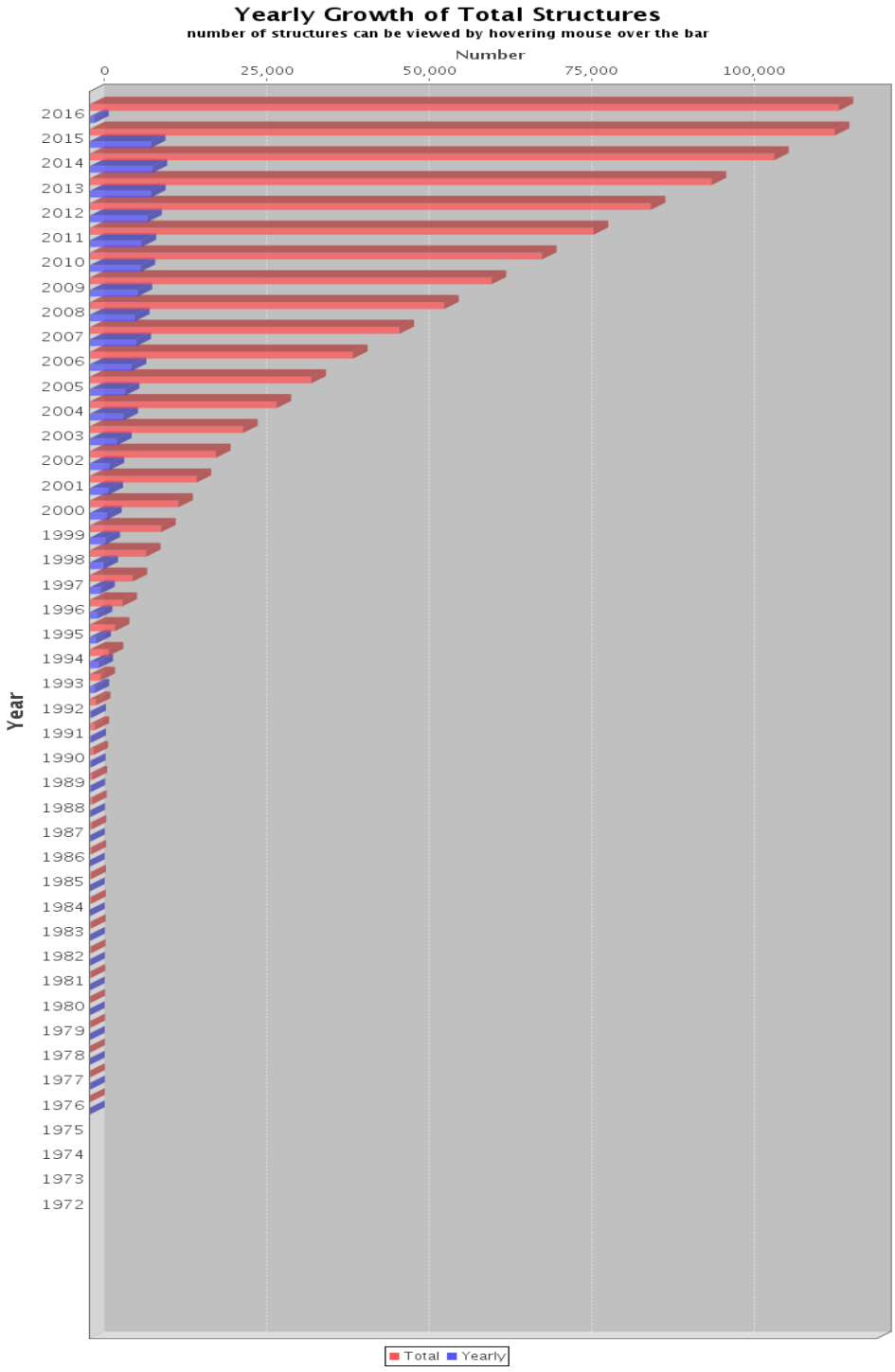

Figure 1: Downloaded from RCSB-PDB, the graph showing the yearly growth of total searchable structures.

Methods for computational structure prediction of a protein molecule is divided into three categories. These categories are based on the available templates in a public database (PDB). In

comparative homology modeling, the target sequence is aligned with template sequence to check the identity between the sequences and eventually, suitable models are built that presumably closely resemble the unknown protein and to evaluate the quality of the models. Ancestral relationship assumes that protein from the same family share some motifs even if it does not share the same sequence (Marti-Renom et al., 2000). The model's quality is determined by aligning the template and target protein. The decrease in the similarity of target and template reduce the quality of the model. The hurdle in model building is only the gap present in template structure due to poor NMR experiment. There is a limit in the similarity between target and template protein below which the homologous does not produce a suitable model (Altschul et al., 1997). To formulate the hypothesis about the structure biochemistry of protein, homology modeling is used, followed by experimental results to prove the hypothesis. Proteins from different evolutionary origins may have a similar structure, for that, threading methods (Bowie et al., 1991; Jones et al., 1992), are designed. In this approach, the query sequence is matched directly with the 3D structures of other solved proteins. The aim of this method is to identify folds similar to the query even when there is no evolutionary relationship between the query and the template protein. The ab initio methods deal with the cases in which we do not have any template in PDB library to solve the structure (Liwo et al., 1999; Simons et al., 1997; Wu et al., 2007). It is one of the toughest cases to deal, and success is restricted to only proteins with < 120 amino acids sequence (Jauch et al., 2007; Zhang, 2008).

I.B.4. Comparative Homology Modelling in Bioinformatics

The ultimate aim of computational structure prediction methods is to associate the amino acid sequence to its 3D structure and finally feature it relevant to functionality. One of the most important applications of molecular modelling is the approach of fitting together or docking protein to a second molecule. The rapid advances in bioinformatics help us in better understanding of drug and target interactions and thus this interaction plays a main role in introducing a computer technology, CADD (Computer Aided Drug Design), the other names used in this perspective are rational drug design, *in-silico* drug design etc. (Jain, 2003; Jain, 2004; Kumar *et al.*, 2006; Oprea and Matter, 2004; Roche and Guba, 2005; Stahl *et al.*, 2006; Stoermer, 2006). The idea of cost reduction and the reduction of the period to produce drugs is

fulfilled by combinatorial chemistry (Myers, 1997). A big number of molecules are screened in a systematic manner by using the tools provided by combinatorial chemistry. Tools used play an important and significant role in the designing of the invention of clinically relevant chemical entities. Computer software is used for molecular techniques to design the initiative steps to increase the molecular diversity to virtually synthesize the chemical libraries (Gallop *et al.,* 1994; Gordon *et al.,* 1994). In short, CADD overcomes the problem which comes in the way of drug discovery like large amount of money to invest, human resources plus a big period. The study of molecular docking starts with the definition of a binding site that is, in broad-spectrum, a classified region of the 3D protein structure (Daniel and Bert, 2010).

For the accuracy of lead compounds discovery, it is essential to give importance to those chemicals that are known to be ideal drug lead. If the appropriate drug target and its 3D structures are known, SBDD (structure-based drug design) can be performed. By using the NMR techniques and X-ray crystallography structures of many drug targets have been developed, (Vesolovoski and Ivinov, 2003) based on geometric and chemical complementarities to ligand-target complex to match the expected binding affinity. Docking and *de novo* design are the two distinct techniques used in SBDD. Systematic and stochastic approaches are the two algorithms used to dock ligands in SBDD. The basic difference between them are that systematic approach involves the rebuilding the structure of ligand within the active site of the receptor. While that of the stochastic approach, the ligand structure is dealt from the initial step. Docking is used for the screening of potent compounds or that of a compound which is medicinally important and is an ideal compound from the huge available database. De novo design evolves the construction of ligands utilizing the structural target information (Colman, 1994).

The purpose of docking methodologies is to forecast the ligand and target complex and to align the molecular databases by binding affinity to that of the target (Jones, 1995; Kuntz *et al.,* 1994; Lenguer and Rarey, 1996; Lybrand, 1995; Rosenfeld *et al.,* 1995). Various methods are used in docking a ligand to the target. These are MOE-Dock (Chemical Computing group, 2010), Gold (Halgren *et al.,* 2004), DOCK (Kuntz, 1992), Auto Dock (Moris, *et al.,* 1998), Surflex (Jain, 2003) and Tripos FlexX (Rarey *et al.,* 1998).

Beside ligand-protein docking another type of docking is also carried out i.e. protein-protein docking, for this purpose the method of 3D dock (Katchalski-Katzir *et al.,* 1992)

(FTDOCK) is used. All these methods are different from each other because of having a difference in atomic partial charges parameters as well as the force field.

The homology modeling technique for 3D structure predictions cannot be generalized only to the protein function prediction and drug design but can also be practiced for experimental designs. These experiments could be site-directed mutagenesis for assessing functional hypotheses, predicting antigenic epitopes and aiding in the refinement of structures solved by X-ray crystallography and NMR experiments, modeling the specificity for a substrate, etc. (Rössle, 2004). The applicability of the predicted model is directly proportional to the conciseness, i.e. the percentage of similarity between the modelled structure and the template protein.

I.B.5. MHOLline: A Web Tool for High Throughput Comparative Homology Modelling

Protein structures prediction, using comparative homology modeling and threading methods has already been applied on a large scale data (*i:e* genome) (Fischer and Eisenberg, 1997; Guex et al., 1999; Sanchez and Sali, 1998). In total, domains in 47% of all 74,897,059 known protein sequences were modeled with MODELLER (Sali and Blundell, 1993) and ModPipe (Sanchez and Sali, 1998) were deposited into ModBase, a widespread database of comparative models (Pieper et al., 2011) ModBase Homepage, (Sanchez et al., 2000a; Sanchez et al., 2000b). The Web interface allows flexible querying to the database for sequence-structure alignments, models, fold assignments and model assessments of interest. ModView, an integrated sequence/structure viewer, allows examination and analysis of the query results.

The MHOLline is an outstanding web tool for large-scale protein 3D models prediction; which can also be utilized for functional re-annotation. The 3D structure obtained for any primary sequence that is annotated as hypothetical protein; initially one can predict their respective function by comparing the modelled structure with its respective template. MHOLline is a high-throughput workflow that merges a specific set of programs for a) automated sequence alignment b) detection of transmembrane regions c) automated template search d) EC number Association e) Protein structure prediction, and f) Assessment of generated models.

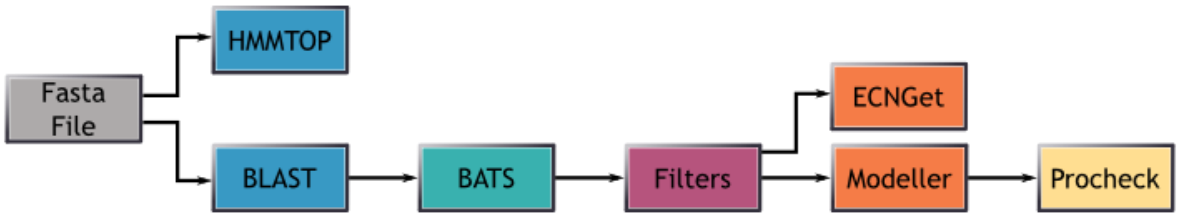

Figure 2: MHOLline; A high-throughput biological workflow, which merges generalized sets of programs that accept amino acid sequences (genome level) in .fasta format as an input data and then performs some analyses, generating 3D protein models on large-scale (MHOLline Homepage) (Capriles et al., 2010).

MHOLline utilizes multi fasta files of amino acids as an input data and then uses HMMTOP, BLAST, BATS, Modeller and Procheck programs for the detailed analyses. The program HMMTOP detects transmembrane regions. The BLAST algorithm is used to identify template structure by performing a random search against the Protein Data Bank (Westbrook et al., 2002). BATS (Blast Automatic Targeting for Structures) carry out the refinement in the template search; it is a key step for the model construction. BATS refinement identifies sequences that make the modeling possible by selecting a template from BLAST output file using their BATS scores, expectation values, identity and sequence similarity as criteria as well as considering the number of gaps and the alignment coverage. BATS select the best template for 3D model generation and perform automated alignment used by the Modeller program (Sanchez and Sali, 1997). Furthermore, it gathers all the BLAST output files into four distinctive groups, i.e. G0, G1, G2, and G3, according to the following criteria;

G0 = Not aligned sequence.

G1 = E-value > $10e^{-5}$ or Identity < 15%

G2 = E-value ≤ $10e^{-5}$ and Identity ≥ 25% AND LVI ≤ 0.7

G3 = E-value ≤ $10e^{-5}$ and Identity ≤ 15% and <25% OR LVI> 0.7

Where LVI is the Length Variation Index, an MHOLline concept of coverage (LVI ≤ 0.1 is equivalent to a coverage ≥ 90%), which is calculated by following Equation,

$$LVI = QL - NAli/QL$$

Where QL represents the length of the query sequence and *NAli* represents the number of amino acids aligned between subject and query sequences. The user receives FASTA files with distinctive sequences parallel to each model quality group. The sequences that fall in G0 or G1 groups are reverted to the end user in new distinct FASTA files, and they can be utilized in future by resubmitting them to MHOLline. Similarly, the sequences from the G3 group cannot be obtained by comparative modelling techniques, and it can be submitted to other programs for folding recognition or ab initio modeling, as MHOLline workflow is not providing this facility. The sequences that were grouped in G2 filter are only able to continue in the MHOLline workflow by their cutoff values. In this group, the BATS analysis the alignment from BLAST and select template only from PDB for structure prediction. It select the highest BATS score for each input sequence and select the best PDB resolution the best PDB resolution is used as a principle of decision to makes the choice, in the case of a tie. The following Equation gives BATS score calculation,

$$S_f = W_iP_i + W_sP_s + W_nP_n + W_lP_l + W_gP_g$$

Where W_i, W_s, W_n, W_l, W_g and P_i, P_s, P_n, P_l, P_g represent, the weight and values attributed to BLAST identity, similarity, number of aligned amino acids, Length Variation Index (*LVI*) and Gap Relative Strength Index (GRSI) respectively. The *GRSI* is equaled to,

$$GRSI = wGa + Gg/QL$$

Where w = 3 is the penalty for the opening of a new gap, Ga is the number of newly opened gaps; Gg is the number of subsequent gaps opened, and QL is the length of the query sequence.

Once the template is selected based on BATS results, MODELLER program is used for the generation of 3D protein model. There is no significant correlation statistically, between the number of templates used during model building and the overall quality of a model (Peitsch, 2002). In the next step, another MHOLline tool called FILTERS, categorizes the BATS selected sequences (G2) into distinct quality model subgroups, described as;

Quality	**Identity**	**LVI**
1. Very High	≥ 75%	≤ 0.1
2. High	≥ 50% and < 75%	≤ 0.1

3. Good	≥ 50%	> 0.1 and ≤ 0.3
4. Medium to Good	≥35% and < 50%	≤ 0.3
5. Medium to Low	≥ 25% and <35%	≤0.3
6. Low	≥25%	> 0.3 and ≤ 0.5
7. Very Low	≥ 25%	> 0.5 and ≤ 0.7

To evaluate the overall quality and accuracy of the model, Ramachandran plot is obtained which explained the stereochemical quality of the model (Laskowski et al., 1996). Precisely, the MHOLine generates an aggregate structural information for all the submitted sequences in the fasta format, Ramachandran plot and other properties like structural quality and enzymatic functions are also determined. Further details can be obtained by visiting MHOLline homepage.

I.C. Justification

Corynebacterium diphtheriae is of a large medical importance and research has been performed for more than a century, little is currently known about the molecular basis of pathogenicity and factors contributing to the virulence of nontoxigenic *C. diphtheriae* isolates. In the present study, we extended the knowledge obtained from the pan-modelome construction of *C. pseudotuberculosis* (Hassan et al., 2014) and applied the same strategy with some modification for therapeutic target identification in *C. diphtheriae*. Due to the emergence of drug resistance in *C. diphtheriae*, we hope our contribution can add a bit to identify novel therapeutic targets for the better cure of diphtheria disease.

I.D. Thesis Delineation

The thesis delineation is articles based and is divided into Objectives, Three chapters, Appendix and Bibliography.

- The Objectives section provides the general and specific aims of this research.
- The first chapter is literature review on *Corynebacterium diphtheriae*. This have provided a short information regarding pathogenesis of *C. diphtheriae* and reported vaccine till the date. Furthermore, it highlights the emerging technique for identification of new therapeutic targets via integrated omics.
- In second chapter, the research article describes a list of putative targets in *C. diphteriae*, and possible mechanisms to design peptide vaccines, and suggest novel lead, natural and drug-like compounds that could bind to the proposed target proteins.
- In third chapter, the research article also provides a list of putative targets in *C. diphteriae*, based on identification of druggable pocketome of all the complete proteomes available on NCBI.
- The fourth chapter is representing general conclusions which describe the summary and main outcomes of the entire work represented in this thesis followed by future prospective.
- The Appendix section lists out the supplementary materials and other publications.
- In Bibliography section, the references used in justification and preface are listed out.

II. OBJECTIVES

II.A. General Objectives

The main goal of this study is the identification of broad-spectrum drug and vaccine targets, through comparative homology modelling, as well as through identification of druggable pocketome from the proteome of *Corynebacterium diphtheriae* at the species level, and identify potent ligand molecules isolated from a natural plant source, ZINC database, and derivatives of Di-terpenoid Iso-steviol.

II.B. Specific Objectives

- High-throughput comparative homology modeling using MHOLline workflow from the predicted proteomes of thirteen genomes of C. *diphtheriae.*
- Identification of conserved proteins with acceptable 3D structures among all *C. diphtheriae* strains.
- Identification of essential non-host and host homologous proteins as drug and vaccine targets based on prioritized parameters, considering human as hosts.
- Molecular Docking analysis of identified targets with ligand libraries.
- Identification of conserved proteins as drug and vaccine targets among all *C. diphtheriae* strains, based on druggable pocketome.

III. CHAPTERS

Chapter 1

III.1.1. Review Article

Pathogenesis of *Corynebacterium diphtheriae* and available vaccines: An Overview.

Syed Babar Jamal, Sandeep Tiwari, Artur Silva and Vasco Azevedo.

Global Journal of Infectious Diseases and Clinical Research, 2017, 3, 20-24.

In this review, we have provided a short information regarding pathogenesis of *C. diphtheriae*. Furthermore, it is directed towards the vaccines and their short comes. The trend of cases reported around the word and also highlights the emerging technique for identification of new therapeutic targets via integrated omics.

Peertechz

Clinical Group

Global Journal of Infectious Diseases and Clinical Research

ISSN: 2455-5363 DOI CC By

Syed Babar Jamal[1], Sandeep Tiwari[1], Artur Silva[2] and Vasco Azevedo[1]*

[1]PG program in Bioinformatics (LGCM), Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil
[2]Institute of Biologic Sciences, Federal University of Para, Belém, PA, Brazil

Dates: Received: 01 September, 2017; Accepted: 28 October, 2017; Published: 30 October, 2017

***Corresponding author:** Vasco Azevedo, PG program in Bioinformatics (LGCM), Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil, Tel/Fax: +55 (31) 3409 2610; E-mail: vascoariston@gmail.com; vasco@icb.ufmg.br

Keywords: *Corynebacterium diphtheria*; Vaccination; Drug-resistant; Therapeutic targets

https://www.peertechz.com

Research Article

Pathogenesis of *Corynebacterium diphtheriae* and available vaccines: An Overview

Abstract

Corynebacterium diphtheriae is Gram-positive bacteria responsible for causing diphtheria in human and once regarded for high mortalities worldwide. The fatality gradually decreased with improved living standards and further alleviated when many immunization programs were introduced. Public infectious diseases have positively been controlled by vaccination, yet, the importance of vaccination usually remained unnoticed for the adults. Many of the under-development vaccines are directed towards the childhood immunization. However, numerous drug-resistant strains emerged recently that consequently decreased the efficiency of current therapeutics and vaccines, thereby obliging the scientific community to start investigating new therapeutic targets in pathogenic microorganisms. In this study, we try to put together a short information regarding pathogenesis of *Corynebacterium diphtheriae* and reported vaccine till the date. Furthermore, we highlighted the emerging technique for identification of new therapeutic targets.

Introduction

Corynebacterium diphtheriae is a human pathogen, responsible for causing Diphtheria. It was once an important cause of death worldwide. The mortality rates gradually decrease with time in the twentieth century in countries where living standards were improved, and then intensely fell once after the introduction of immunization programs [1]. Though, even today, despite these events it remains a substantial pathogen in many parts of the world. A variety of mechanisms are responsible for causing death. However, the name 'strangling angel' of children arose from the wing-shaped pseudo-membranes that form in the oropharynx. Displacement and impaction of these pseudo-membranes starts acute airway obstruction and sudden death [2,3]. Since there has been a revival of cases of non-lethal and lethal diphtheria in some countries in past few years and that considerable population displacements are happening due to refugee and immigration movements, more cases may be encountered in future.

Diphtheria is a potentially a lethal disease that mainly effect upper respiratory tract tissues and kills its sufferers slowly by suffocation. In 1884, a German physician, Edwin Klebs (1834–1913), was able to isolate the bacteria successfully that proved to be the etiological agent of the disease. It was later verified that toxin production begun only after the bacteria were infected by a specific virus or a bacteriophage themselves, carrying the toxin's genetic instructions [4].

Although diphtheria is forgotten disease in several European countries but it remains a potential health issue in many endemic countries and serious problem for those countries that are considered to be free from diphtheria [5]. In past few years, the awareness has been increased due to some periodic cases reported in Europe, especially contemporary case in Spain and cutaneous diphtheria cases in immigrants in Denmark, Germany and Sweden; the shortage issue of diphtheria antitoxin was also emphasized as a European Union priority [5]. Similarly, according to a report of WHO, cases were reported from some Asian countries. In 2015, India was found to have 2,365 cases of diphtheria. The number of reported cases were higher than any other country [1].

Public infectious diseases have positively been controlled by vaccination, yet, the importance of vaccination usually remained unnoticed for the adults [6]. Many of the under development vaccines are directed towards the childhood immunization [7]. The global population of elderly people has been increasing due to better-quality hygiene and healthcare system. Taking in account aging factor and the ongoing debility of the immune response to vaccination, it is important to develop worldwide strategies of vaccination and emphasize more on adult vaccination [1].

The increase in resistant strains of pathogens, we are in need of new tools to treat and avert the infectious diseases. The

Citation: Jamal SB, Tiwari S, Silva A, Azevedo V (2017) Pathogenesis of *Corynebacterium diphtheriae* and available vaccines; an overview. Glob J Infect Dis Clin Res 3(1): 020-024. DOI: http://doi.org/10.17352/2455-5363.000014

extensive use of antibiotics in the medication of human and veterinary for many years has led to the development of bacterial strains that are multi-resistant with partial to no response to existing cures. It resulted in patients needing longer time for treatment as they need screening and treatment with several antibacterial agents and ultimately causing extra stress on patients and providers of health care [8]. The advancement of new technology and affectedly decreases in the running costs of NGS has allowed to generate a bulk amount of data on different pathogens [9,10]. NGS allows the analysis of microbial genome on broad-spectrum. This data has been proven to be valuable for infection prevention measures with the consequences for interspecies transmission, microbial evolution, variable regions and the potential of spread to different hosts [11-13]. The microbial pan-genome deals with the characterization of genomes by comparison of related species, possibly by accessing core, accessory and strain-specific genes [14].

The genetic knowledge of the *C. diphtheriae* species by performing comparative analysis of the complete genome sequences using pan-genomics technique by Trost et al., 2012 to describe the role of *C. diphtheriae* for its Genomic Diversity in the Cases of Classical Diphtheria, Endocarditis, and Pneumonia genomes at specie level. This data on the genomic content of different strains of *C. diphtheriae* provides deep insight into the virulence factors and features associated with the life style of the Human pathogen [15].

Here, we try to gather a short information regarding pathogenesis of *Corynebacterium diphtheriae* and reported vaccine till the date. Furthermore, we highlighted the emerging technique for identification of new therapeutic targets via integrated omics.

Pathogenesis

Diphtheria toxin is responsible for the pathological consequences of diphtheria infection. The time of incubation may be from one to eight days, but mostly, it is between two to five days. The beginning is usually nonspecific with a sore throat and a low-grade fever which may mimic streptococcal pharyngitis, candidiasis or infectious mononucleosis [2,16,17]. It takes almost 24 hours, the gray colored pseudo-membrane appears covering the soft palate, uvula, and tonsils. It seems white initially but darkens as blood trickles into it. In young children, a more severe form occurs known as malignant or "bull neck" diphtheria. The commencement of diphtheria is quick and the growth of the pseudo-membrane faster associated with the buccal cavity, entire pharynx, middle ear and nose [2]. The soft palate, tonsils, and uvula may suffer necrosis and slough, and necrotic lesions may breach into the primary skeletal muscle with marked hemorrhage [18]. Furthermore, thinner pseudo-membranes are developed by distal airway, and the lungs are edematous and hemorrhagic [19]. The inner layer of Pseudo-membranes may have fibrin with an outer covering of neutrophils with aggregates of embedded bacteria inside the necrotic material [18]. There is clear inflammation of the cervical lymph nodes and adjacent soft tissues making the characteristic "bull neck" appearance. The firmness of the jugular veins may cause marked blocking of the face [2], other portions of the upper airway may be involved with nasal and laryngeal diphtheria, and latter allied with a high death rate.

The heart could have pale, enlarged, chambers with a distinctive 'streaky" appearance. Histologic sections may show marked hyaline degeneration that causes cell death with mononuclear cell permeation and lipid vacuoles inside surviving myocytes [18]. Proper treatment needs rapid management of diphtheria antitoxin and antibiotic coverage [17]. Shallow mucosal erosions may be present inside the stomach, and non-lethal diphtherial contagions of the skin can be found in the tropics. While these may lead to pharyngeal involvement through autoinfection [3]. Other sites of infection involve mycotic aneurysm formation, osteomyelitis, septic arthritis and splenic abscess [20].

Nowadays, we often think of cutaneous diphtheria in the context of umbilical diphtheria, wound diphtheria or impetiginous diphtheria. Owing to the ability of *C. diphtheriae* to colonize, lesions of skin can be different from any skin lesion of other origin (e.g., surgical wounds, eczema, pyoderma, impetigo, insect bites or dermatitis). Usually, an ulcerative lesion (ecthyma diphtheriticum) is the presenting lesion. It starts as a vesicle or abscess filled with straw-colored fluid, which breaks down soon after the formation. The lesion develop as punched-out ulcer, one or more than one, measuring from millimeters to a few centimeters, with margins slightly curved and raised. Furthermore, the margins may slightly be diluted or inverted. Lower legs, feet, and hands are the common sites for diphtheric lesions. The lesions are usually responsible for cousing a lot of pain and may be covered with a dark pseudomembrane during the first couple of weeks. With the time, the lesion becomes sedative and the pseudomembrane falls away. The wound looks like a hemorrhagic base, usually with serous or serosanguinous exudate oozing from it. The tissues surrounding the infection stay edematous and pink, purple or livid in color and may show swellings or bullae. Skin lesions yielding *C. diphtheriae* on cultures are indistinguishable from those linked with other bacteria and can include nearly healed, dry or scaly lesions [21].

Dipththerial infections are mainly because of the toxin. Diphtheria toxin is an exotoxin secreted by *C. diphtheria.* Diphtheria toxin is a single polypeptide chain of 535 amino acids consisting of two subunits linked by disulfide bridges, known as an A-B toxin. There are at least four main steps involved in intoxication of a single eukaryotic cell by diphtheria toxin: (1) the binding of the toxin to surface receptor of its target cell; (2) grouping of charged receptors into layered pits and internalization of the toxin by receptor-mediated endocytosis; followed by acidification of the endocytic vesicle by a membrane-associated, ATP-driven proton pump, (3) the insertion of the transmembrane domain (B-subunit) into the membrane and smoothed the delivery of catalytic domain (A-subunit) to the cytosol, and (4) the ADP-ribosylation of elongation factor 2 (EF-2), which results in the permanent inhibition of protein synthesis as shown in figure 1. A single molecule of the catalytic domain delivered to the cytosol is enough to be deadly for the cell [22].

Citation: Jamal SB, Tiwari S, Silva A, Azevedo V (2017) Pathogenesis of *Corynebacterium diphtheriae* and available vaccines; an overview. Glob J Infect Dis Clin Res 3(1): 020-024. DOI: http://doi.org/10.17352/2455-5363.000014

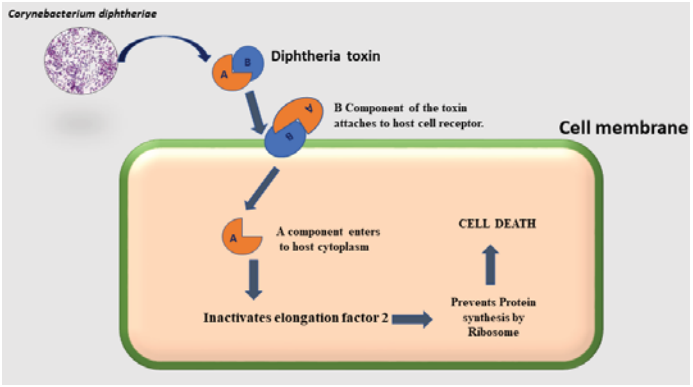

Figure 1: Mechanism of action of diphtheria toxin.

Discovery of Diphtheria Toxin and Vaccine development

During the 19th century, diphtheria toxin was discovered by Emile Roux. This discovery led to the development of passive serum cures through the scientific contributions of many other scientists of that era http://www.immunize.org/timeline/. Likewise, the etiological agent of Pertussis, commonly known as the "a whooping cough," was found to be a bacterium isolated from tissues of an infected patient in 1906 [23]. They revealed that the animal's serum that had been exposed to sub-lethal doses of the bacteria involved in tetanus and diphtheria was defensive against the fatal effects allied with these pathogens by having an antitoxin outcome when injected into another animal. Furthermore, this discovery, which resulted in Behring inaugural Nobel Prize for Physiology and Medicine in 1901, was the idea of passive transmission in addition to serum therapy. He verified that serum could be attained from immune animals and transmitted to others as protection [24]. Once this concept made its way to clinical practice in late 19th century, handling problems were encountered while developing the right antitoxin concentration and potency. Consequently, in the early twentieth century, the U.S. Congress passed the Biologics Control Act legislation "to normalize the sale of viruses, serums, toxins, and similar products" to guarantee medication quality control.

However, with the growing use and fame of antitoxins derived from animal serum, scientists began to observe a syndrome, now known as serum sickness, or a reaction to immune-complexes designed from combining high concentrations of antigens with antibodies. This ultimately headed towards the use of human rather than animal serum, so it could reduce the rate of hostile events; still, treatment with serum was not perfect in controlling disease due to the frequency of hostile events and its short-lived period of action. Afterward, conjoining diphtheria toxin and antitoxin in the same syringe showed much more activity in reducing the mortality rate. The commercial availability of this combination became in 1897. This was the primary step towards passive to active immunization [4].

Later in the 20th century, a French veterinarian Gaston Ramon (1886–1963), working at the Pasteur Institute, used a diphtheria toxoid, produced by formalin and heat inactivation without the use of antitoxin to securely tempt active immunity in humans [24]. This product was named anatoxin and provided the basis for the unique and clinically effective toxoid vaccine against diphtheria. Many experiments were performed to enhance the strength of the defensive response to the vaccine, and in 1926, the prominence of aluminum salts as an adjuvant added to the vaccine to augment the immune response to the antigen became apparent [1]. This discovery was made by Alexander Thomas Glenny (1882–1965) who proved that the toxoid only produced a lower level of antibody and immunity than desired, whereas improved immunity was accomplished when an inflammatory reaction was activated. With these substantial advances, tetanus and diphtheria toxoids became regularly used across America and Europe in the era between the 1930s and 1940s [4].

Since then, modifications have been made to these vaccines to produce higher purity and decrease the number of booster doses. Currently, extensive childhood vaccination is reducing the load of these diseases. Though this is a huge advantage, vaccines may produce antagonistic effects that can discourage their approval by some populations. This has directed to some safety movements which concluded in the congressionally legislated National Childhood Vaccine Injury Act in the 1980s it was shaped to recompense families for selected adverse events possibly related to compulsory childhood vaccinations [25]. Still, worldwide endorsements continue to call for routine immunization of children against diphtheria, tetanus, and pertussis with the combined DTP vaccine to ensure immunity in childhood and youth. Hence, DTaP has become one of the globally used vaccines to achieve extensive immunity across age groups [4].

Currently, in the United States, the pediatric formulation (diphtheria-tetanus-acellular pertussis [DTaP]) of vaccines are available for use under the brand names as Infanrix and Daptacel (Manufactured by GlaxoSmithKline and Sanofi Pasteur respectively). Teenage and adult formulation (tetanus–diphtheria–acellular pertussis [Tdap]) of vaccines which were approved in 2005 for teenagers are in use under the brand names as Boostrix (GlaxoSmithKline) and Adacel (Sanofi Pasteur) in the United States. Later in 2006, Tdap vaccination was recommended for adults younger than 65 years. These adult form of vaccines have an equal amount of tetanus and diphtheria toxoid in comparison with the adult form of Td vaccines. Boostrix has a reduced quantity of pertussis antigens compared with the Infanrix and it is licensed for persons 10 years of age and older. Adecel has a reduced quantity of pertussis toxin compared with Daptacel and is licensed for persons 10 through 64 years of age [26].

It has been reported in several studies that protective antibody levels of tetanus and diphtheria wilted since the last vaccination because of aging, and antibodies had on estimated the half-life of 11 years [6,27]. These data better explain the reason behind the lack of antibody in the older age. Due to work related pressure and military services, it had been observed in males from ≥25 years of age that antibody levels were significantly higher [28].

Citation: Jamal SB, Tiwari S, Silva A, Azevedo V (2017) Pathogenesis of *Corynebacterium diphtheriae* and available vaccines; an overview. Glob J Infect Dis Clin Res 3(1): 020-024. DOI: http://doi.org/10.17352/2455-5363.000014

The immunization of DNA with in vivo electroporation is an alternative and competent approach to produce monoclonal antibodies (mAb). The mAb generation by DNA immunization is a novel approach to outwit the following technical hurdles associated with problematic antigens: low profusion and protein instability and use of recombinant proteins that lack posttranslational modifications [29].

Till the date, diphtheria has been very effectively controlled by an efficient immunization program in developed countries [30]. Though, in recent years, the disease has made a dramatic return, especially within Eastern Europe. The major outbreak since the beginning of mass immunization within Russia and the states that were newly independent of the former Soviet Union in the 1990s [31,32]. Furthermore, cases were reported from some part of Africa and Asia in the recent past [32,33].

In 2003, the genome (*C. diphtheriae* NCTC 13129) was sequenced at the Sanger Institute, which was clinical isolate related to this outbreak. It provided useful basis to identify candidate virulence factors besides the toxin itself, like iron transport systems and fimbrial proteins [34]. Jamal *et al.*, 2017 utilized the genomic information from all the available genomes and adopted an integrative OMICs approach for therapeutic target identification against diphtheria. In their work, high throughput comparative modeling (Pan-modelome) was performed to generate 3D structures from the proteome of all the available genomes. After filtering intra-specie conserved proteins, a final set of eight proteins (glpX, nusB, rpsH, hisE, smpB, bioB, DIP1084, and DIP0983) were identified as essential and non-host homologs, considering human as a host. The identified 8 proteins were subjected to virtual screening against four different compound libraries (extracted from the ZINC database, plant-derived natural compounds and Di-terpenoid Iso-steviol derivatives). The proposed ligand molecules showed lowered energy values, high complementarity with the predicted targets and favorable interactions. Interestingly, among the drug-like molecule from all the four databases, ZINC13142972 (1-[(2S, 3S, 4S, 5R)-3,4-dihydroxy-5(hydroxymethyl) oxolan-2-yl]imidazo[1,2-b]pyrazole-7-carbonitrile) showed good results against two of our predicted targets NP_939302.1 (glpX, Fructose 1,6-bisphosphatase II) and NP_939445.1 (DIP1084, Putative iron transport membrane protein, FecCD-family). Furthermore, Jacarandic Acid and Rhein were identified as the top ranked molecules from the library of natural compounds (28 molecules). The in-silico analysis of the library (derivatives of diterpenoid isosteviol) suggest that compounds 16-hydroxyisosteviol, 16-hydrazonisosteviol, 17-hydroxyisosteviol, 16–17 dihydroxyisosteviol and 16-oxime, 17-hydroxyisosteviol were top ranked molecules and might be potent molecules for the inhibition of our targets. The approach adopted in this study expedites the selection of *C. diphtheriae* putative proteins for the development of novel drugs and vaccines on broad-spectrum, owing to the fact that some of these targets have already been identified and validated in other organisms [35].

Though *C. diphtheriae* is of a large medical importance and research has been performed for more than a century, a little is currently known about the molecular basis of pathogenicity and factors contributing to the virulence of nontoxigenic *C. diphtheriae* isolates. The advancement in OMICs sciences helped enough the mankind to identify novel therapeutic targets. Due to the emergence of drug resistance in *C. diphtheriae*, it is essential to identify new therapeutic targets for the better cure of diphtheria disease.

Conclusion

Diphtheria is somehow forgotten disease but still for past few year cases had been reported throughout the world. This occurrence has concerned the scientific community to rethink over the currently available treatments. The availability of genomic data provides means to better understanding the molecular and genetic basis of virulence of this bacterium, enabling a detailed investigation of *C. diphtheria*. In the long run, providing a new gate way for development and/or improvement of potent vaccine.

Acknowledgements

Syed Babar Jamal acknowledge all the authors for their contributions.

References

1. Lee HJ, JH Choi (2017) Tetanus-diphtheria-acellular pertussis vaccination for adults: an update. Clin Exp Vaccine Res 6: 22-30. Link: https://goo.gl/ajWzQd

2. Hodes HL (1979) Diphtheria. Pediatr Clin North Am 26: 445-459. Link: https://goo.gl/QdLwp4

3. PE Hart, PY Lee, DC Macallan, MH Wansbrough-Jones (1996) Cutaneous and pharyngeal diphtheria imported from the Indian subcontinent. Postgrad Med J 72: 619-620. Link: https://goo.gl/Kcgkvn

4. Grabenstein JD (2010) Toxoid Vaccines, in Vaccines: A Biography. AW Artenstein, Editor. 2010, Springer New York: New York, NY. 105-124. Link: https://goo.gl/AAsW3t

5. Kantsone I, I Lucenko, J Perevoscikovs (2016) More than 20 years after re-emerging in the 1990s, diphtheria remains a public health problem in Latvia. Euro Surveill 21: 30414. Link: https://goo.gl/mBMYYe

6. Weinberger B (2017) Adult vaccination against tetanus and diphtheria: the European perspective. Clin Exp Immunol 187: 93-99. Link: https://goo.gl/sXXcj1

7. Swanson KA, Schmitt HJ, Jansen KU, Anderson AS (2015) Adult vaccination. Hum Vaccin Immunother 11: 150-155. Link: https://goo.gl/bDHxBn

8. Hey A (2015) History and Practice: Antibodies in Infectious Diseases. Microbiol Spectr 3: AID-0026-2014. Link: https://goo.gl/jfELgS

9. Dark MJ (2013) Whole-genome sequencing in bacteriology: state of the art. Infect Drug Resist 6: 115-123. Link: https://goo.gl/BVsNzS

10. Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB (2011) The real cost of sequencing: higher than you think! Genome Biol 12: 125. Link: https://goo.gl/wvBQKa

11. Harrison EM, Lucy A Weinert, Matthew TG Holden, John J Welch, Katherine Wilson, et al., (2014) A shared population of epidemic methicillin-resistant Staphylococcus aureus 15 circulates in humans and companion animals. MBio 5: e00985-e01013. Link: https://goo.gl/vagAjd

12. Price LB, Marc Stegger, Henrik Hasman, Maliha Aziz, Jesper Larsen, et al., (2012) Staphylococcus aureus CC398: host adaptation and emergence

Citation: Jamal SB, Tiwari S, Silva A, Azevedo V (2017) Pathogenesis of *Corynebacterium diphtheriae* and available vaccines; an overview. Glob J Infect Dis Clin Res 3(1): 020-024. DOI: http://doi.org/10.17352/2455-5363.000014

of methicillin resistance in livestock. MBio 3: e00305-e00311. Link: https://goo.gl/P2UdbE

13. Zhou K, Lokate M, Deurenberg RH, Tepper M, Arends JP, et al., (2016) Use of whole-genome sequencing to trace, control and characterize the regional expansion of extended-spectrum β-lactamase producing ST15 Klebsiella pneumoniae. Scientific Reports 6: 20840. Link: https://goo.gl/RVgVg3

14. Trost E, Blom J, Soares Sde C, Huang IH, Al-Dilaimi A, et al., (2012) 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 194: 3199-3215. Link: https://goo.gl/EbyAXi

15. Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, Powell E, et al., (2007) Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome. J Bacteriol 189: 8186-8195. Link: https://goo.gl/aQzXAG

16. Byard RW (2002) Unexpected death due to infectious mononucleosis. J Forensic Sci 47: 202-204. Link: https://goo.gl/nMtN4S

17. Wagner KS, White JM, Crowcroft NS, De Martin S, Mann G, et al., (2010) Diphtheria in the United Kingdom, 1986-2008: the increasing role of Corynebacterium ulcerans. Epidemiol Infect 138: 1519-1530. Link: https://goo.gl/yne1o2

18. Hadfield TL, McEvoy P, Polotsky Y, Tzinserling VA, Yakovlev AA (2000) The pathology of diphtheria. J Infect Dis 181 Suppl 1: S116-120. Link: https://goo.gl/mNsEm2

19. Goutas N, Simopoulou S, Papazoglou K, Agapitos E (1994) A fatal case of diphtheria. Pediatr Pathol 14: 391-395. Link: https://goo.gl/oQFP8q

20. Muttaiyah S, Best EJ, Freeman JT, Taylor SL, Morris AJ, et al., (2011) Corynebacterium diphtheriae endocarditis: a case series and review of the treatment approach. Int J Infect Dis 15: e584-e588. Link: https://goo.gl/7ht1Lw

21. Hadfield TL (2000) The Pathology of Diphtheria. J Infect Dis 181: S116–S120. Link: https://goo.gl/4zFoPa

22. Murphy JR (1996) Corynebacterium Diphtheriae. Medical Microbiology, 4th edition. University of Texas Medical Branch at Galveston, Galveston, Texas. Link: https://goo.gl/aoRs4H

23. Cherry JD (1996) Historical review of pertussis and the classical vaccine. J Infect Dis 174 Suppl 3: S259-S263. Link: https://goo.gl/8etJoD

24. Lombard M, Pastoret PP, Moulin AM (2007) A brief history of vaccines and vaccination. Rev Sci Tech 26: 29-48. Link: https://goo.gl/k2XYxD

25. Kretsinger K, Karen R. Broder, Margaret M. Cortese, M. Patricia Joyce, Ismael Ortega-Sanchez, et al., (2006) Preventing tetanus, diphtheria, and pertussis among adults: use of tetanus toxoid, reduced diphtheria toxoid and acellular pertussis vaccine recommendations of the Advisory Committee on Immunization Practices (ACIP) and recommendation of ACIP, supported by the Healthcare Infection Control Practices Advisory Committee (HICPAC), for use of Tdap among health-care personnel. MMWR Recomm Rep 55: 1-37. Link: https://goo.gl/xdNVpr

26. Hamborsky, J., et al., (2015) Epidemiology and prevention of vaccine-preventable diseases. 13th edition. Centers for Disease Control and Prevention, Washington, DC. Link: https://goo.gl/HZbNxC

27. Amanna IJ, NE Carlson, MK Slifka (2007) Duration of humoral immunity to common viral and vaccine antigens. N Engl J Med 357: 1903-1915. Link: https://goo.gl/MSk1VD

28. Trucchi C, G Zoppi (2015) Decennial diphtheria-tetanus adult boosters: are they really necessary? J Prev Med Hyg 56: E44-E48. Link: https://goo.gl/zDt6hR

29. Chua KY, Ramos JD, Cheong N (2008) Production of monoclonal antibody by DNA immunization with electroporation. Methods Mol Biol 423: 509-520. Link: https://goo.gl/YCh4PE

30. Vitek CR (2006) Diphtheria. Curr Top Microbiol Immunol 304: 71-94. Link: https://goo.gl/FfkQ3p

31. Dittmann S, Wharton M, Vitek C, Ciotti M, Galazka A, Guichard S, et al., (2000) Successful control of epidemic diphtheria in the states of the Former Union of Soviet Socialist Republics: lessons learned. J Infect Dis 181 Suppl 1: S10-S22. Link: https://goo.gl/BQ7xw3

32. Anuradha S (2006) Tetanus in adults–a continuing problem: an analysis of 217 patients over 3 years from Delhi, India, with special emphasis on predictors of mortality. Med J Malaysia 61: 7-14. Link: https://goo.gl/VNL4dB

33. Sadoh AE, RE Oladokun (2012) Re-emergence of diphtheria and pertussis: implications for Nigeria. Vaccine 30: 7221-7228. Link: https://goo.gl/zLTaH6

34. Cerdeno-Tarraga AM, Efstratiou A, Dover LG, Holden MT, Pallen M, et al., (2003) The complete genome sequence and analysis of Corynebacterium diphtheriae NCTC13129. Nucleic Acids Res 31: p. 6516-23. Link: https://goo.gl/uHXiC1

35. Jamal SB, Hassan SS, Tiwari S, Viana MV, Benevides LJ, et al., (2017) An integrativein-silico approach for therapeutic target identification in the human pathogen Corynebacterium diphtheriae. PLoS ONE 12: e0186401. Link: https://goo.gl/8arUA6

Citation: Jamal SB, Tiwari S, Silva A, Azevedo V (2017) Pathogenesis of *Corynebacterium diphtheriae* and available vaccines; an overview. Glob J Infect Dis Clin Res 3(1): 020-024. DOI: http://doi.org/10.17352/2455-5363.000014

III.1.2. Conclusion, Chapter 1

- Diphtheria is somehow forgotten disease but still for past few year cases had been reported throughout the world.
- The occurrence of new cases has concerned the scientific community to rethink over the currently available treatments.
- The availability of genomic data provides means to better understanding the molecular and genetic basis of virulence of this bacterium, enabling a detailed investigation of *C. diphtheriae*. In the long run, providing a new gate way for development and/or improvement of potent vaccine.
- Hence, we extended our knowledge working with genomic data to identify novel therapeutic targets via integrated omics approaches.

Chapter 2

III.2.1. Research Article

An integrative in-silico approach for therapeutic target identification in the human pathogen *Corynebacterium diphtheria.*

Syed Babar Jamal, Syed Shah Hassan, Sandeep Tiwari, Marcus V. Viana, Leandro de Jesus Benevides, Asad Ullah, Adrian G. Turjanski, Debmalya Barh, Preetam Ghosh, Daniela Arruda Costa, Artur Silva, Richard Rottger, Jan Baumbach, Vasco Azevedo.

PLoS ONE, 2017, 12(10): e0186401.

Corynebacterium diphtheriae (Cd) is a gram-positive bacteria human pathogen responsible for diphtheria infection and once regarded for high mortalities worldwide. However, numerous drug-resistant strains emerged recently that consequently decreased the efficacy of current therapeutics and vaccines, thereby obliging the scientific community to start investigating new therapeutic targets in pathogenic microorganisms. In this manuscript, our contributions include the prediction of modelome of 13 *C. diphtheriae* strains, using the MHOLline workflow. Further, using subtractive proteomics and modelomics approaches for target identification, a set of 23 proteins was selected as essential for the bacteria. The targets were subjected to virtual screening against ligand libraries. The proposed drug molecules showing favorable interactions, lowered energy values and high complementarity with the predicted targets have also been reported in the present study. We expect that the in silico computational approaches adopted in this study might aid in the development of novel therapeutic drugs and vaccines against *C. diphtheriae*. Moreover, the strategy described herein is of applied nature and can also be employed to other pathogenic microorganisms.

PLOS ONE

RESEARCH ARTICLE

An integrative *in-silico approach* for therapeutic target identification in the human pathogen *Corynebacterium diphtheriae*

Syed Babar Jamal[1☯], Syed Shah Hassan[1,2☯], Sandeep Tiwari[1], Marcus V. Viana[1], Leandro de Jesus Benevides[1], Asad Ullah[2], Adrián G. Turjanski[3], Debmalya Barh[4], Preetam Ghosh[5], Daniela Arruda Costa[1], Artur Silva[6], Richard Röttger[7], Jan Baumbach[7], Vasco A. C. Azevedo[1,8]*

1 PG program in Bioinformatics (LGCM), Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil, **2** Department of Chemistry, Islamia College University Peshawar, KPK, Pakistan, **3** Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pabellón II, Buenos Aires, Argentina, **4** Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology, Nonakuri, Purba Medinipur, West Bengal, India, **5** Department of Computer Science, Virginia Commonwealth University, Richmond, VA, United States of America, **6** Institute of Biologic Sciences, Federal University of Para, Belém, PA, Brazil, **7** Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark, **8** Department of General Biology (LGCM), Institute of Biologic Sciences, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil

☯ These authors contributed equally to this work.
* vascoariston@gmail.com

OPEN ACCESS

Citation: Jamal SB, Hassan SS, Tiwari S, Viana MV, Benevides LdJ, Ullah A, et al. (2017) An integrative *in-silico approach* for therapeutic target identification in the human pathogen *Corynebacterium diphtheriae*. PLoS ONE 12(10): e0186401. https://doi.org/10.1371/journal.pone.0186401

Editor: Alexandre G. de Brevern, UMR-S1134, INSERM, Université Paris Diderot, INTS, FRANCE

Received: December 6, 2016

Accepted: September 29, 2017

Published: October 19, 2017

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The study was supported by grant from the TWAS-CNPq Postgraduate Fellowship Programme (https://twas.org/opportunity/twas-cnpq-postgraduate-fellowship-programme) for granting a fellowship for doctoral studies and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil: http://www.

Abstract

Corynebacterium diphtheriae (Cd) is a Gram-positive human pathogen responsible for diphtheria infection and once regarded for high mortalities worldwide. The fatality gradually decreased with improved living standards and further alleviated when many immunization programs were introduced. However, numerous drug-resistant strains emerged recently that consequently decreased the efficacy of current therapeutics and vaccines, thereby obliging the scientific community to start investigating new therapeutic targets in pathogenic microorganisms. In this study, our contributions include the prediction of modelome of 13 *C. diphtheriae* strains, using the MHOLline workflow. A set of 463 conserved proteins were identified by combining the results of pangenomics based core-genome and core-modelome analyses. Further, using subtractive proteomics and modelomics approaches for target identification, a set of 23 proteins was selected as essential for the bacteria. Considering human as a host, eight of these proteins (glpX, nusB, rpsH, hisE, smpB, bioB, DIP1084, and DIP0983) were considered as essential and non-host homologs, and have been subjected to virtual screening using four different compound libraries (extracted from the ZINC database, plant-derived natural compounds and Di-terpenoid Iso-steviol derivatives). The proposed ligand molecules showed favorable interactions, lowered energy values and high complementarity with the predicted targets. Our proposed approach expedites the selection of *C. diphtheriae* putative proteins for broad-spectrum development of novel drugs and vaccines, owing to the fact that some of these targets have already been identified and validated in other organisms.

capes.gov.br/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: BLAST, Basic Local Alignment Search Tool; Cd, *Corynebacterium diphtheriae*; DNA, Deoxyribonucleic acid; LDC, Lysine decarboxylase; MVD, Molegro Virtual Docker; PDB, Protein Data Bank; RNA, Ribonucleic acid; NP, Natural Product.

Introduction

Corynebacterium diphtheriae is responsible for causing diphtheria which remains a major global cause of death (http://www.who.int/immunization_monitoring/diseases/diphteria/), and has conventionally been divided into four subgroups of biovars *i.e.*, gravis, intermedius, mitis and belfanti based on biochemical characteristics according to Funke *et al.*, 1997 [1] and Whitman *et al.*, 2012 [2]. It was once a major cause of infant mortality, which spread as an epidemic and resulted in thousands of deaths [3]. The death rates dropped over time specifically in countries where living standards have improved, and the death rates rapidly declined after the introduction of immunization programs [3]. Despite these measures, it remains a significant pathogen around the globe, even today. A variety of mechanisms were responsible for causing such death rates; for example the 'strangling angel' effect on children that ascended from the wing shaped pseudo-membranes formed in the oropharynx. Disarticulation and impaction of these pseudo-membranes triggers acute airway obstruction and can result in sudden death [3, 4]. Since there has been a plethora of reported cases on both non-lethal and lethal diphtheria across various countries in the past few years, and that significant population displacements in the form of immigration are happening, more such cases are bound to follow. A passable handling requires quick inroads in discovering diphtheria antitoxin and antibiotic treatment [5].

Computational methods and other approaches, like reverse vaccinology, have been established for the rapid identification of novel targets in the post-genomic era [6, 7]. Approaches like subtractive and comparative microbial genomics as well as differential genome analysis [8] are being used for the identification of targets in a number of human pathogens like *M. tuberculosis [9]*, *Burkholderia pseudomalleii* [10], *Helicobacter pylori* [11] *Pseudomonas aeruginosa* [12], *Neisseria gonorrhea* [13] and *Salmonella typhi* [14].

The main principle is to find targeted genes/proteins that are essential for the pathogen and possess no homology counterpart in the host [15], such that drugs targeting these "pathogen-essential non-host homologs" can be applied with little (or no) off targets in the host. Some pathogen-essential proteins, though, may possess a certain degree of homology to host proteins. However, they might still be selected as potential molecular targets for structure-based selective inhibitor development. Significant differences in the active sites or in other druggable pockets might exist, such that the pathogenic protein could still be targeted [16, 17].

Here, we exploit an integrative *in silico* approach for the predictive proteome of *C. diphtheriae* species to associate the genomic information with the identification of putative therapeutic targets based on their three dimensional structure. It can be utilized for the identification of potent inhibitors, which might possibly lead to the discovery of compounds that inhibit pathogenic growth. The predicted proteomes from the 13 genomes of C. *diphtheriae* were modeled (pan-modelome) using the MHOLline workflow as proposed by Hassan *et al.*, 2014 [18]. Furthermore, intra-species conserved proteins with adequate 3D models (core-modelome) were filtered on the basis of predicted essentiality for the bacteria, which leads to the identification of eight essential bacterial proteins. They were found non homologous to all host proteins and have been subjected to virtual screening using multiple compound libraries.

We provided a list of putative targets in *C. diphteriae*, and possible mechanisms to design peptide vaccines, and suggest novel lead, natural and drug-like compounds that could bind to the proposed target proteins.

Materials and methods

Genomes selection

The thirteen *C. diphtheriae* strains, including three of the four biovars: gravis, mitis and belfanti (**Table 1**) were included in this study. The gene and protein sequences of these thirteen *C. diphtheriae* strains were retrieved from NCBI (ftp://ftp.ncbi.nih.gov/genomes/Bacteria). The different steps involved in this computational approach for genome-scale modelome prediction and for the prioritization of putative drug and vaccine targets are given in (**Figs 1 & 2**).

Prediction of core-modelome and identification of core genome

To construct the core-modelome of *C. diphtheriae*, we followed a slightly modified protocol described by Hassan *et al*., 2014 [18]. High throughput structural modeling, MHOLline (http://www.mholline.lncc.br), was used to predict the modelome (whole-proteome set of protein 3D models) for each strain. MHOLline uses comparative modeling approach for protein 3D structure prediction through MODELLER [19]. Our workflow also includes BLASTp (Basic Local Alignment Search Tool for Protein) [20], HMMTOP (Prediction of transmembrane helices and topology of proteins), [21] BATS (Blast Automatic Targeting for Structures), FILTERS, ECNGet (Get Enzyme Commission Number), MODELLER, and PROCHECK [22].

MHOLline work on the basis of available template. It is probable that MHOLline cannot detect all the common conserved proteins due to the unavailability of the template. To overcome this probability, we used EDGAR (an **E**fficient **D**atabase framework for comparative **G**enome **A**nalyses using BLAST score **R**atios for pan-genomics analysis) to collect common conserved genome as well of all Cd strains [23]. Later, the results from MHOLine and EDGAR were compared and crosschecked to obtain the final dataset of common conserved proteins.

Identification of intra-species conserved proteins

Primarily, for the identification of highly conserved proteins with available 3D models in all Cd strains ($\geq$ 95% sequence identity), the standalone release of NCBI BLASTp+ (v2.2.26) was adapted from the NCBI ftp. Site (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/) and installed on a local machine. Furthermore, a search was performed using NCTC13129 as a random reference genome for all strains. Comparative genomics/proteomics approach was

Table 1. Strains of *C. diphtheriae* employed in the pan-modelome study with information on genomes statistics, disease prevalence and location of isolation.

Strain	GPID	NCBI Accession	Genome Size (Mb)	Proteins	GC%	Location	Biovar
31A	PRJNA42399	NC_016799.1	2.53535	2258	53.60	Brazil	—
241	PRJNA42407	NC_016782.1	2.42655	2245	53.40	Brazil	—
BH8	PRJNA42423	NC_016800.1	2.48552	2223	53.60	Brazil	—
C7	PRJNA42401	NC_016801.1	2.49919	2230	53.50	USA	—
CDCE8392	PRJNA42405	NC_016785.1	2.43333	2135	53.60	USA	Mitis
HC01	PRJNA42409	NC_016786.1	2.42715	2162	53.40	Brazil	Mitis
HC02	PRJNA42411	NC_016802.1	2.46861	2179	53.70	Brazil	Mitis
HC03	PRJNA42415	NC_016787.1	2.47836	2186	53.50	Brazil	Mitis
HC04	PRJNA42417	NC_016788.1	2.48433	2190	53.50	Brazil	Gravis
INCA402	PRJNA42419	NC_016783.1	2.44907	2163	53.70	Brazil	Belfanti
PW8	PRJNA42403	NC_016789.1	2.53068	2200	53.70	USA	—
VA01	PRJNA42421	NC_016790.1	2.39544	2100	53.40	Brazil	Gravis
NCTC13129	PRJNA87	NC_002935.2	2.48863	2185	53.50	UK	Gravis

https://doi.org/10.1371/journal.pone.0186401.t001

next adopted for selecting the highly conserved proteins using an all-against-all BLASTp analysis with a cut-off value of $E = 0.0001$, as in many other essentiality studies before [6, 13, 15, 18, 24].

Essential and non-host homologous (ENH) protein targets

A subtractive genomics approach was next followed for the selection of conserved targets, which were essential to the bacteria [15]. Concisely, the set of proteins derived from the core-modelome of *C. diphtheriae* was subjected to the Database of Essential Genes (DEG) for homology analyses. The DEG encompasses experimentally validated data of currently available essential genomic elements like protein-coding genes and non-coding RNAs, from bacteria, archaea and eukaryotes. For a bacterium, essential genes form a minimal genome, i.e., a set of functional modules that has key roles in the emerging field of synthetic biology [25]. The cutoff values used for BLASTp were: *E-value = 0.0001*, *bit score* $\geq$*100* and *identity* $\geq$ *25%* [15, 18].

The pool of essential genes was then subjected to NCBI-BLASTp (*E-value = 0.0001*, *bit score* $\geq$*100* and *identity* $\geq$ *25%*) against the human genome for filtering pathogen-essential host-homologs [6]. The remaining set of pathogen-essential non-host homologs were additionally crosschecked with NCBI-BLASTp PDB database using the default values to find any remote structural similarity with the existing host homolog protein structures, keeping the cutoff level to $\leq$ 15% for query coverage. The biochemical pathways of these proteins have been checked using KEGG (Kyoto Encyclopedia of Genes and Genomes) [26], functionality using UniProt (Universal Protein Resource) [27], virulence using PAIDB (Pathogenicity island database) [28], and cellular localization using CELLO (subCELlular LOcalization predictor) [29]. The final list of targets was based on criteria described by Barh *et al.*, 2011 & Hassan *et al.*, 2014 [15, 18].

Essential and host homologous (EH) protein targets

We further extended our analyses to the set of protein targets that were essential to *C. diphtheriae* but homologous to host proteins. The essential protein targets deviating from the cutoff

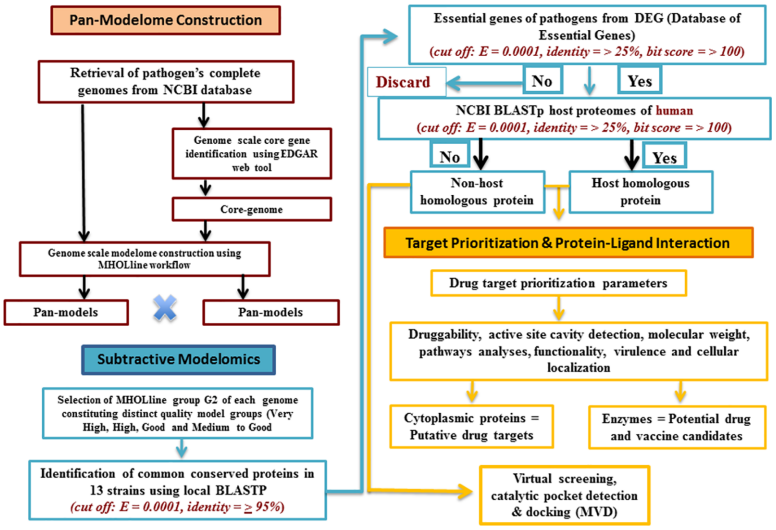

Fig 1. Overview of different computational steps employed for the identification of putative essential targets (non-host homologous and host homologous) from the core-proteome of 13 *C. diphtheriae* strains.

https://doi.org/10.1371/journal.pone.0186401.g001

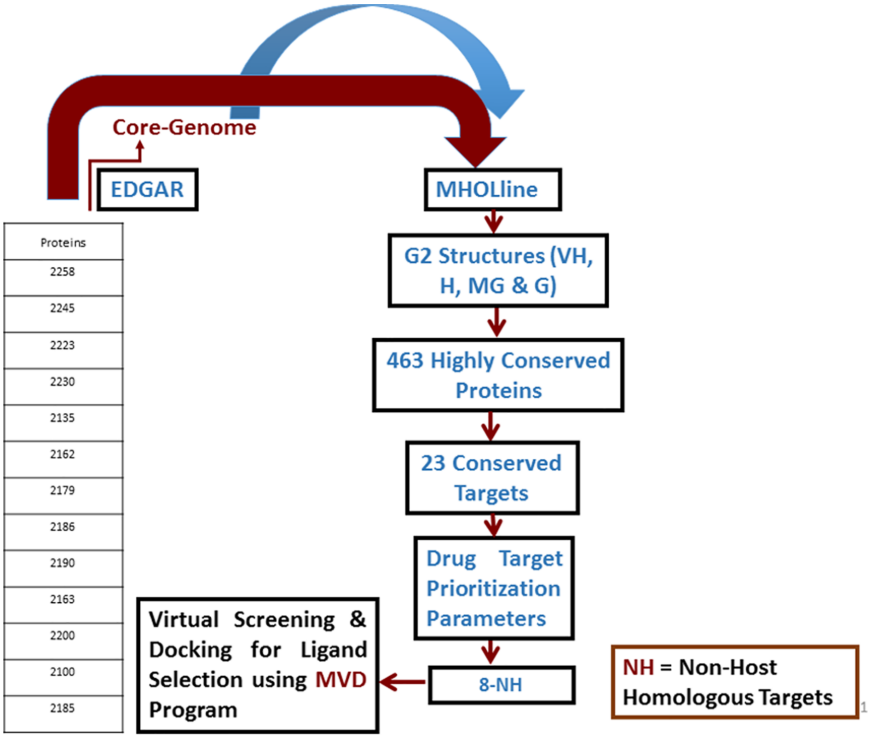

Fig 2. Intra-species subtractive modelomics workflow for conserved target identification in *C. diphtheriae* species. The table represents the total number of protein sequences as an input data fed to the MHOLline workflow (upper red arrow). The blue arrow represents the core genes of thirteen Cd strains. The rectangular boxes show how this workflow processes and filters a large quantity of genomic data for putative drug and vaccine target identification of a pathogen.

https://doi.org/10.1371/journal.pone.0186401.g002

values for essential non-host homologous proteins were treated as host homologous proteins. This set of targets was also checked for pathway involvement, functional annotation virulence, and cellular localization as mentioned above.

Computational identification of druggable pockets

The information obtained from 3D structures and druggability analyses are important features for prioritizing and authenticating putative pathogen targets [30, 31]. As mentioned above, for druggability analyses, the final list of essential non-host and host homologous protein targets were subjected to DoGSiteScorer in PDB format [32]. The DoGSiteScorer is an automated pocket detection and analysis tool for calculating the druggability of protein cavities. For each detected cavity the tool returns the pocket residues and a druggability score ranging from 0 to 1. Values closer to 1 indicate highly druggable protein cavity, i.e. the predicted cavities are likely to bind ligands with high affinity [32]. The DoGSiteScorer also calculates volume, depth, surface area, lipophilic surface, and further parameters for each predicted cavity.

Ligand libraries preparation, virtual screening and docking analyses

The ligand libraries were prepared from four different sources, compounds from ZINC database (ZINC drug-like molecules, ZINC Natural Product), natural compounds from literature survey [33] and the Di-terpenoid Iso-steviol derivatives (**S1 Table**). ZINC (drug-like molecules) contains 11,193 drug-like molecules, with Tanimoto cutoff level of 60% [34] and ZINC (Natural Product) contain 11,203 molecules. Whereas, the small library of natural compounds

contained 28 molecules and the library of Di-terpenoid Iso-steviol derivatives contained 31 molecules respectively. The structures of these molecules were constructed using MOE-Builder tool. The 3D structures were modeled and partial charges were calculated using MOE (Molecular Operating Environment). The energies of the modeled molecules were minimized using the energy minimization algorithm of MOE tool (gradient: 0.05, Force Field: MMFF94X, Chiral Constraint) [35]. The modeled molecules were saved in the.mol2 file format and subjected to docking analysis.

The 3D structures of proteins were examined for structural errors such as missing atoms, wrong bonds and protonation states in the MVD (Molegro Virtual Docker) [36]. The consensus set of protein cavities and those predicted with DogSiteScorer (druggability $\geq$ 0.80) were compared with the MVD detected cavities, for all Cd targets. The maximum numbers of residues from DoGSiteScorer falling in the cavities detected by MVD were merged and final grid was generated based on the consensus between the highest scoring pocket from DoGSiteScorer and cavities detected by MVD for docking. The most druggable cavity was subjected to virtual screening using MVD. The program comprises of three search algorithms for molecular docking analyses namely MolDock Simplex Evolution (SE), MolDock Optimizer [36] and Iterated Simplex (IS). We employed the MolDock Optimizer search algorithm, which is based on a differential evolutionary algorithm, using the default parameters that are **a)** population size = 50, **b)** scaling factor = 0.5 and **c)** crossover rate = 0.9. The orientations of docked molecules from the library of natural compounds and from the derivatives of Di-terpenoid Iso-steviol were analyzed in Chimera [37]. The 200 top ranked compounds (ZINC drug-like molecules, ZINC Natural Product) for each target protein were evaluated for shape complementarity and hydrogen bond interactions. This led to the selection of a final set of compounds with polypharmacology and polypharmacy characteristics for target proteins in *C. diphtheriae*.

Results and discussion

Modelome prediction and conserved targets identification in *C. diphtheriae*

Among 13 strains of *C. diphtheriae* species, our employed methodology produced high-confidence 3D structural models from orthologous proteins in *C. diphtheriae* species through the efficient MHOLline workflow (**Fig 3**). A comparative structural genomics approach was

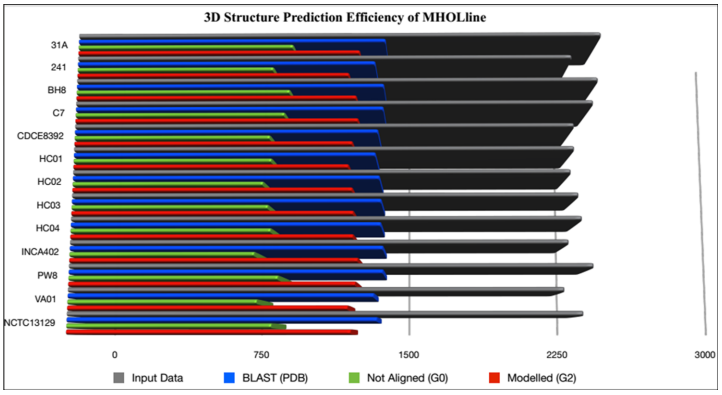

Fig 3. Efficiency of the MHOLline biological workflow for genome-scale modelome (3D models) prediction. Predicted proteomes from the genomes of 13 *C. diphtheriae* strains were fed to the MHOLline workflow in FASTA format. The grey bars represent the number of input data. The remaining bars (MHOLline output data) show the number of not aligned sequences (G0, green bars), sequences for which there is a template structure available at RCSB PDB (blue bars), and sequences with acceptable template structures that were modeled in the MHOLline workflow (G2, red bars).

https://doi.org/10.1371/journal.pone.0186401.g003

followed where all the G2 sequences classified as "Very High", "High", "Good" and "Medium to Good quality" by MHOLline, from the 12 Cd strains, were aligned to the Cd NCTC13129 strain as a reference genome. First, we identified a set of common conserved proteins with a pre-defined sequence similarity of 95–100%. This resulted in a set of 463 protein sequences, being conserved in all Cd strains (**S3 Table**).

Protein targets as putative drug and vaccine candidates

The identification of essential proteins in *C. diphtheriae* was carried out where the core-modelome was compared to DEG (Database of Essential Genes). This filter drastically reduced the number of selected targets to 23 final targets. Further comparison of the corresponding protein sequences to the human host proteome resulted in a set of 8 targets as essential non-host homologous (ENH, **Table 2**) and a set of 15 targets as essential host homologous proteins (EH, **Table 3**).

Prioritization parameters for drug targets and vaccine candidates

There are several factors that can aid in determining potential therapeutic targets [30]. For vaccine candidates, the information about subcellular localization is important: Proteins that contain transmembrane motifs are favored [24, 30, 38, 39]. The 23 essential proteins have a low molecular weight and all are localized in the cytoplasmic compartment of *C. diphtheriae* (**Tables 2 & 3**). After the druggability evaluation using DoGSiteScorer [32] for both essential non-host and host homologous conserved targets from *C. diphtheriae*, we could predict at least one druggable cavity for each Cd target. The host homologous proteins as therapeutic targets could adversely affect the host. Therefore, the first step in numerous *in silico* drug target identification approaches are filtering proteins homologous to host proteome. Thus, we only consider the eight pathogen-essential non host homologs for the docking studies [13, 15, 40]. For the eight pathogen-essential non host homologs (**S2 Table**) glpX, nusB, rpsH, hisE, DIP1084, DIP0983, smpB, and bioB 3, 0, 1, 0, 2, 0, 1 and 3 cavities with score > 0.80 were predicted. The cavity of each protein exhibiting the highest druggability score was subjected to docking analyses. The numbers of predicted cavities with their respective druggability scores are given in **Tables 2 & 3**.

The identified eight non-host homologous and essential Cd proteins could be novel therapeutic targets for *Corynebacterium diphtheriae*.

As per our knowledge, glpX, hisE and bioB proteins have been reported as potential drug target in *Mtb*. Protein nusB is a member of Nus-transcription Factor family that help bacteria in the process of elongation, transcription: translation coupling and termination. Some members of this family (nusG) has already been reported as drug target. Furthermore, rpsH and smpB are also reported as potential drug target by Folador et al., 2016 in their *in silico* study [41]. Protein DIP1084 is Putative iron transport membrane protein (FecCD-family) and DIP0983 is uncharacterized Hypothetical Protein that need to be characterized experimentally. Hence, these protein could be a good therapeutic target against Cd.

Virtual screening and molecular docking

For each target protein (glpX, nusB, rpsH, hisE, DIP1084, DIP0983, smpB, and bioB) four different libraries were separately screened. A total of 28 molecules from natural compounds library and 31 compounds from the derivatives of Di-terpenoid Iso-steviol library were docked. Furthermore, top 200 drug-like molecules from virtual screening analyses of two large libraries (ZINC drug-like molecules, ZINC Natural Product) were examined one-by-one for the selection of the final set of promising molecules that showed favorable interactions with

Table 2. Drug and/or vaccine target prioritization parameters and functional annotation of the eight essential non-host homologous putative targets.

Gene and protein codes	Official full name	Cavities with DS[a] ≥ 0.80	Cavities with DS[a,g] ≥ 0.60 and < 0.80	Mol. Wt[b] (KDa)	Functions[c]	Cellular component[d]	Pathways[e]	Virulence[f]
NP_939692.1, **nusB**	Transcription antitermination protein NusB/ N utilization substance protein B homolog	0	2	20.382	**MF:** RNA binding. **BP:** DNA-templated transcription, termination, regulation of transcription, DNA-templated.	Cytoplasm	unknown	No
NP_939612.1, **hisE**	Phosphoribosyl-ATP pyrophosphatase	0	1	9.877	**MF:** RNA binding, phosphoribosyl-ATP diphosphatase activity **BP:** histidine biosynthetic process	Cytoplasm	Biosynthesis of amino acids	Yes
NP_939445.1, **DIP1084**	Iron ABC transporter membrane protein/ Putative iron transport membrane protein, FecCD-family	2	3	35.470	**MF:** Transporter activity **BP:** Transport	Membrane	The ATP-binding cassette (ABC) transporters form one of the largest known protein families	Yes
NP_939345.1, **DIP0983**	Hypothetical protein DIP0983/ Uncharacterized protein	0	4	28.193	**MF:** possible lysine decarboxylases (Pfam)/52% sequence identity with PDB Template 1WEK. **BP:** A pyridoxal-phosphate protein. Also acts on 5-hydroxy-L-lysine (IUBMB Comments)	Cytoplasm	unknown	Yes
NP_939302.1, **glpX**	Fructose 1,6-bisphosphatase II	3	2	35.589	**MF:** fructose 1,6-bisphosphate 1-phosphatase activity, metal ion binding. **BP:** gluconeogenesis, glycerol metabolic process		Carbohydrate Metabolism	No
NP_939123.1, **smpB**	SsrA-binding protein	1	2	18.784	**MF:** RNA binding	Cytoplasm	unknown	Yes
NP_938900.1, **rpsH**	30S ribosomal protein S8	1	1	14.292	**MF:** rRNA binding, structural constituent of ribosome **BP:** Translation	Extracellula/ Cytoplasm	unknown	No
NP_938502.1, **bioB**	Biotin synthase	3	1	38.224	**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	Yes

[a]Druggability predicted with DoGSiteScorer software. A druggability score above 0.60 is considered to be good, but a score above 0.80 is favored [32].

[b]Molecular weight was determined using ProtParam tool (http://web.expasy.org/protparam/).

[c]Molecular function (MF) and biological process (BP) for each target protein was determined using UniProt.

[d]Cellular localization of pathogen targets was performed using CELLO.

[e]KEGG was used to find the role of these targets in different cellular pathways.

[f]PAIDB was used to check if the putative targets are involved in the pathogen's virulence.

[g]DS = Drug Score

https://doi.org/10.1371/journal.pone.0186401.t002

Table 3. Drug and/or vaccine target prioritization parameters and functional annotation of the fifteen essential host homologous putative targets.

Gene and protein codes	Official full name	Cavities with DS[a] ≥ 0.80	Cavities with DS[a,g] ≥ 0.60 and < 0.80	Mol. Wt[b]. (KDa)	Functions[c]	Cellular[d] component	Pathways[e]	Virulence[f]
NP_938651.1 **RecR**	Recombination protein RecR	0	2	**23.901**	**MF:** DNA binding, metal ion binding **BP:** DNA recombination, DNA repair	Cytoplasm	Homologous recombination	**Yes**
NP_938792.1 **DIP0411**	Putative electron transport related protein	**0**	2	**19.950**	**MF:** Antioxidant activity, oxidoreductase activity	Cytoplasm/ Membrane	—	**Yes**
NP_938922.1 **rpsM**	30S ribosomal protein S13	**0**	1	**13.777**	**MF:** rRNA binding, structural constituent of ribosome, RNA binding **BP:** Translation	Cytoplasm	Ribosome	**No**
NP_939046.1 **DIP0672**	Putative uptake hydrogenase small subunit	**2**	0	**43.949**	**MF:** 3 iron, 4 sulfur cluster binding, 4 iron, 4 sulfur cluster binding, ferredoxin hydrogenase activity, metal ion binding	Cytoplasm	Microbial metabolism in diverse environments	**Yes**
NP_939341.1 **dapD**, DIP0979	Tetrahydropicolinate succinylase **EC 2.3.1.117**	**1**	1	**33.780**	**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	Biosynthesis of amino acids	**Yes**
NP_939343.1 **DIP0981**	Putative succinyltransferase **EC 2.3.1.117**	**1**	1	**33.039**	**MF:** 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase activity	Cytoplasm	Biosynthesis of amino acids	**Yes**
NP_939460.1 **ilvH**, DIP1099	Acetolactate synthase small subunit **EC 2.2.1.6**	**1**	3	**19.063**	**MF:** acetolactate synthase activity, amino acid binding **BP:** branched-chain amino acid biosynthetic process	Cytoplasm/ Membrane	2-Oxocarboxylic acid metabolism	**Yes**
NP_939590.1 **cobM**	Precorrin-4 C11-methyltransferase **EC 2.1.1.133**	**1**	2	**27.181**	**MF:** precorrin-2 dehydrogenase activity, precorrin-4 C11-methyltransferase activity **BP:** cobalamin biosynthetic process, porphyrin-containing compound biosynthetic process	Cytoplasm	Porphyrin and chlorophyll metabolism	**Yes**
NP_939786.1 **DIP1438**	Putative transport membrane protein	**4**	3	**44.215**	**MF: T**ransporter activity **BP:** transmembrane transport	Membrane	The ATP-binding cassette (ABC) transporters	**Yes**
NP_939832.1 **DIP1484**	Putative uroporphyrinogen III methyltransferase	**3**	2	**28.296**	**MF:** Methyltransferase activity **BP:** oxidation-reduction process	Cytoplasm	Porphyrin and chlorophyll metabolism	**Yes**
NP_939958.1 **aroH,** DIP1616	Phospho-2-dehydro-3-deoxyheptonate aldolase **EC 2.5.1.54**	**2**	3	**50.805**	**MF:** 3-deoxy-7-phosphoheptulonate synthase activity **BP:** aromatic amino acid family biosynthetic process	Cytoplasm	Biosynthesis of amino acids	**Yes**

(*Continued*)

Table 3. (*Continued*)

Gene and protein codes	Official full name	Cavities with DS[a] ≥ 0.80	Cavities with DS[a,g] ≥ 0.60 and < 0.80	Mol. Wt[b]. (KDa)	Functions[c]	Cellular[d] component	Pathways[e]	Virulence[f]
NP_940018.1 **DIP1680**	Putative GTP cyclohydrolase 1 type 2 **EC 3.5.4.16**	**2**	1	**40.657**	**MF:** GTP binding, GTP cyclohydrolase I activity, metal ion binding **BP:** 7,8-dihydroneopterin 3'-triphosphate biosynthetic process	Cytoplasm	—	**Yes**
NP_940228.1 **cysE,** DIP1891	Serine acetyltransferase **EC 2.3.1.30**	**1**	0	**20.208**	**MF:** serine O-acetyltransferase activity **BP:** cysteine biosynthetic process from serine	Cytoplasm	Carbon metabolism	**Yes**
NP_940284.1 **DIP1952**	Putative pyruvate dehydrogenase	**3**	1	**62.497**	**MF:** Catalytic activity, magnesium ion binding, thiamine pyrophosphate binding	Cytoplasm	(PYRUVATE METABOLISM) Nicotinate and nicotinamide metabolism	**Yes**
NP_940605.1 **DIP2303**	Putative DNA protection during starvation protein	**0**	1	**18.223**	**MF:** Ferric iron binding, oxidoreductase activity, oxidizing metal ions **BP:** cellular iron ion homeostasis, response to stress	Cytoplasm	—	**Yes**

[a]Druggability predicted with DoGSiteScorer software. A druggability score above 0.60 is usually considered, but a score above 0.80 is favored [32].

[b] Molecular weight was determined using ProtParam tool (http://web.expasy.org/protparam/).

[c] Molecular function (MF) and biological process (BP) for each target protein was determined using UniProt.

[d]Cellular localization of pathogen targets was performed using CELLO.

[e]KEGG was used to find the role of these targets in different cellular pathways.

[f]PAIDB was used to check if the putative targets are involved in the pathogen's virulence.

[g]DS = Drug Score.

https://doi.org/10.1371/journal.pone.0186401.t003

the ENH targets. The biological importance and an analysis of the predicted protein-ligand interaction/s for each target are described here. The molecule names, ZINC codes and Mol-Dock scores for the selected ligands, as well as the number of predicted hydrogen bonds with the protein cavity residues involved in these interactions, are shown below (**Tables 4–11**) for each target protein. The predicted binding modes of selected ligands are also shown for each pathogen target in **Figs 5–12.**

Validation of docking protocol

To validate the accuracy of MolDock program (MVD), the co-crystallized ligand of Biotin synthase, bioB (PDB ID; 1R30) was extracted and then re-docked into the binding pocket of receptor protein. The RMSD between docked and co-crystallized ligand was found to be 1.81 A˚, which shows that the adopted docking protocol is valid and can be used to correctly predict the binding pose of the ligands [35, 42]. The superposition of co-crystallized ligands and docked is shown in **Fig 4.**

NP_939302.1 (**glpX,** Fructose 1, 6-bisphosphatase II) is a key enzyme of gluconeogenesis and catalyzes the hydrolysis of fructose 1, 6-bisphosphate to form fructose 6-phosphate and orthophosphate. A reverse reaction catalyzed by phosphofructokinase in glycolysis, and the

Table 4. Compounds/Libraries name, MolDock scores and predicted hydrogen bonds for the selected best-ranked molecules against NP_939302.1 (glpX, Fructose 1,6-bisphosphatase II).

Compounds		MolDock Score	H-Bond/Residues
Plant derived natural compounds	Rhein	-64.1265	3/ Val95, Asp197
	Jacarandic Acid	-62.0658	4/ Asp96, Thr98, Asp197, Pro221
Derivative of diterpenoid isosteviol	16-hydrazonisosteviol	-64.2107	5/ Asp93, Val95, Asp96, Val198
	16-oxime, 17-hydroxyisosteviol	-69.6824	4/ Asp93, Asp96, Thr98
	Benzyl ester isosteviol lactone	-69.8464	3/ Asp93, Asp197, Glu222
ZINC Compounds	ZINC00042420	-106.97	3/ Arg175, Arg197, Val198
	ZINC13142972	-109.648	3/ Asp93, Asp197, Glu222
ZINC *NP Compounds	ZINC67912153	-135.111	13/Gly64, Glu65, Asp93, Val95, Gly97, Thr98, Phe127, Arg175, Glu222
	ZINC67902753	-121.762	8/Glu65, Val95, Thr98, Glu222
	ZINC38143633	-123.150	11/Lys37, Asp93, Val95, Asp96, Gly97, Thr98

*NP = Natural Product (http://zinc.docking.org/catalogs/acdiscnp)

https://doi.org/10.1371/journal.pone.0186401.t004

product, fructose 6-phosphate, is an important precursor in various biosynthetic pathways [43]. In all organisms, gluconeogenesis is an important metabolic pathway that allows the cells to synthesize glucose from non-carbohydrate precursors, such as organic acids, amino acids and glycerol. FBPases are members of the large superfamily of lithium sensitive phosphatases, which includes three families of inositol phosphatases and FBPases (the phosphoesterase clan CL0171, 3167 sequences, Pfam data base). The FBPases are already reported as targets for the development of drugs for the treatment of noninsulin dependent diabetes [44, 45]. Based on a comparison with a crystallographic structure of the glpX template (PDB ID: 1NI9, GlpX from *Escherichia coli*), none of the active site residues were identified. The docking analysis was performed utilizing the highest scoring pocket obtained from DoGSiteScorer. **Table 4** shows a set of 10 promising ligands according to their minimum energy values and the maximum number of hydrogen bond interactions from the four aforementioned libraries. Compounds **ZINC67912153, ZINC13142972,** Jacarandic Acid and 16-hydrazonisosteviol are shown in **Fig 5**.

NP_939692.1 (**nusB,** Transcription antitermination protein NusB) is a prokaryotic transcription factor involved in antitermination processes, during which it interacts with the mRNA nut site at boxA portion. The crystal structure of *M. tuberculosis* and *E. coli* NusB proteins suggest that the basic N-terminal region of the molecule associates with the rRNA BoxA. Hypothetically, this is indicative of the so-called arginine rich RNA binding motif (ARM) in

Table 5. Compounds/Libraries name, MolDock scores and predicted hydrogen bonds for the selected molecules against NP_939692.1 (nusB, Transcription antitermination protein NusB).

Compounds		MolDock Score	H-Bond/Residues
Plant derived natural compounds	Rhein	-78.0652	1/ Asp36
	Jacarandic Acid	-68.1913	3/ Asp36, Glu91, Arg98
Derivative of diterpenoid isosteviol	16-hydrazonisosteviol	-92.7911	2/ Glu31, Leu73
ZINC Compounds	ZINC00053531	-99.4716	3/ Asp34, Asp36, Ala69
	ZINC19899354	-114.966	3/ Asp36, Ala69, Asp77
ZINC NP Compounds	ZINC67911826	-131.288	7/Ala30, Glu31, Asp34, Ile35, Arg102
	ZINC15043210	-124.580	7/ Glu31, Ile35, Leu73 Glu91, Arg102
	ZINC31168395	-131.442	6/ Glu31, Asp34, Ala69, Leu73, Asp77, Arg102

https://doi.org/10.1371/journal.pone.0186401.t005

Table 6. Compounds/Libraries name, MolDock scores and predicted hydrogen bonds for the selected molecules against NP_938900.1 (rpsH, 30S ribosomal protein S8).

Compounds		MolDock Score	H-Bond/Residues
Plant derived natural compounds	Rhein	-48.8803	3/ Asp5, Arg15
	Jacarandic Acid	-49.3506	4/ Ser12, Val77
Derivative of diterpenoid isosteviol	16-hydrazonisosteviol	-68.2446	3/ Asp5, Arg13, Arg79
	17-hydroxyisosteviol	-64.5855	5/ Asp5, Arg79
	16–17 dihydroxyisosteviol	-56.3868	4/ Asp5, Arg79
	16-oxime, 17-hydroxyisosteviol	-65.1995	4/ Ser26, Ser29, Ser30
ZINC Compounds	ZINC15221730	-103.636	5/ Ala8, Arg15, Arg79
	ZINC71913776	-87.1474	5/ Arg15, Arg79
	ZINC72333100	-104.807	3/ Arg15, Arg79, Val80
ZINC NP Compounds	ZINC35457686	-107.091	10/Ala8, Ser12, Arg15, Val77, Arg78, Arg79, Lys82
	ZINC67903079	-131.210	10/ Asp5, Ser12, Arg15, Arg78, Arg79, Lys82
	ZINC31163223	-100.684	7/ Ala8, Asp9, Arg78, Arg79, Lys82

https://doi.org/10.1371/journal.pone.0186401.t006

Table 7. Compounds/Libraries name, MolDock scores and predicted hydrogen bonds for the selected molecules against NP_938502.1 (bioB, Biotin synthase).

Compounds		MolDock Score	H-Bond/Residues
Plant derived natural compounds	Rhein	-72.2918	3/ Tyr183, Asn184, His185
	Jacarandic Acid	-98.0169	2/ Ala132, Glu188
Derivative of diterpenoid isosteviol	16-hydrazonisosteviol	-107.55	4/ Gly165, Tyr183, Asn184, Glu188
	17-hydroxyisosteviol	-92.2141	4/ Cys93, Ala132, Val134, Tyr183
	16-oxime, 17-hydroxyisosteviol	-98.9592	5/ Glu95, Val134, Glu188, Arg206, Thr323
	Benzyl ester isosteviol lactone	-89.8881	5/ Ala132, Val134, Gly165, Asn184
ZINC Compounds	ZINC16952914	-119.354	5/ Val134, Ile164, Gly165, Asn184, Asn186
ZINC NP Compounds	ZINC77269615	-164.853	17/Cys86, Phe92, Cys93, Ser94, Gln95, Ala132, Val134, Ser163, Asn184, Asn186, Arg206
	ZINC04098512	-162.050	14/ Cys86, Cys90, Cys93, Ala132, Val134, Asn184, Asn186, His201, Arg206, Asp256, Thr323
	ZINC15112225	-142.080	10/ Phe92, Cys93, Ser94, Gln95, Asn186, Arg206, Asn253

https://doi.org/10.1371/journal.pone.0186401.t007

Table 8. Compounds/Libraries name, MolDock scores and predicted hydrogen bonds for the selected molecules against NP_939612.1 (hisE, Phosphoribosyl-ATP pyrophosphatase).

Compounds		MolDock Score	H-Bond/Residues
Plant derived natural compounds	Rhein	-54.9556	1/ Tyr84
	Jacarandic Acid	-61.0241	3/ Leu87
Derivative of diterpenoid isosteviol	16–17 dihydroxyisosteviol	-70.8496	3/ Leu65, Tyr84
ZINC Compounds	ZINC05809437	-89.6781	1/ Asp5
ZINC NP Compounds	ZINC38143703	-99.499	4/ Thr79, Ile83, Leu87
	ZINC67913372	-97.997	5/ Gly36, Ile62, Ile83, Tyr84

https://doi.org/10.1371/journal.pone.0186401.t008

Table 9. Compounds/Libraries name, MolDock scores and predicted hydrogen bonds for the selected molecules against NP_939123.1 (smpB, SsrA-binding protein).

Compounds		MolDock Score	H-Bond/Residues
Plant derived natural compounds	Rhein	-67.698	3/ Ser16, Ser114, Leu115
	Jacarandic Acid	-52.3689	1/ Asn69
Derivative of diterpenoid isosteviol	16-hydroxyisosteviol	-53.2141	4/ Ser50, Thr52, Asn69
	16-hydrazonisosteviol	-64.6203	3/ Thr52, Asn67
	16–17 dihydroxyisosteviol	-59.7364	3/ Ser16, Lys19, Val49
ZINC Compounds	ZINC01414475	-86.7944	3/ Thr52, Asn67, Asn69
	ZINC17128469	-74.5349	3/ Ser16, Leu51, Thr52
ZINC NP Compounds	ZINC31168211	-158.056	12/ Asn9, Ser16, Val49, Ser50, Thr52, Asp53, Ser54, Thr109
	ZINC33832449	-134.974	10/ Asn9, Ser16, Asn17, Val49, Ser50, Thr52, Asp53
	ZINC04096316	-137.613	9/ Asn9, Ser10, Ser16, Asn17, Lys19, Val49, Ser50, Thr52

Table 10. Compounds/Libraries name, MolDock scores and predicted hydrogen bonds for the selected molecules against NP_939445.1 (DIP1084, Putative iron transport membrane protein, FecCD-family).

Compounds		MolDock Score	H-Bond/Residues
Plant derived natural compounds	Rhein	-66.4406	1/ Ser164
	Jacarandic Acid	-77.5981	1/ Gly265
Derivative of diterpenoid isosteviol	16-hydrazonisosteviol	-96.3945	4/ Tyr97, Ser164, Ile266
	17-hydroxyisosteviol	-90.1488	4/ Tyr97, Ser164, Ile266
	Benzyl ester isosteviol lactone	-71.4733	4/ Tyr97, Ser164
ZINC Compounds	ZINC01645563	-95.7116	5/ Tyr97, Ser102, Ser164, Ile266
	ZINC13142972	-111.185	5/ Ser103, Ser106, Thr110, Ser164, Ser168
	ZINC62023045	-103.542	4/ Tyr97, Ser102, Ser106, Ser164
ZINC NP Compounds	ZINC70454922	-155.667	10/ Tyr97, Ser102, Ser106, Ser168, Ser172, Met191, Gly319
	ZINC31167925	-135.535	10/ Tyr97, Ser164, Ser168, Met191, Gly265, Ile266, Thr322
	ZINC04963990	-127.671	8/ Asp95, Tyr97, Ser106, Met191, Ile266, Phe268

https://doi.org/10.1371/journal.pone.0186401.t010

Table 11. Compounds/Libraries name, MolDock scores and predicted hydrogen bonds for the selected molecules against NP_939345.1 (DIP0983, hypothetical protein DIP0983).

Compounds		MolDock Score	H-Bond/Residues
Plant derived natural compounds	Rhein	-55.7819	3/ Cys13, Leu17, Asp177
	Jacarandic Acid	-80.8294	3/ Cys13, Gly14
Derivative of diterpenoid isosteviol	17-hydroxyisosteviol	-95.9025	2/ His11, Cys13
	16–17 dihydroxyisosteviol	-83.7226	2/ His11, Cys13
ZINC Compounds	ZINC00114311	-125.423	6/ Arg10, Arg155, Gly172
	ZINC00211173	-98.6064	7/ Arg10, Gly106, Arg155, Gly174, Asp177
	ZINC01427915	-112.22	6/ Arg10, Ala77, Arg78, Gly172, Gly174
	ZINC04836994	-136.847	5/ Arg10, Gly106, Ile131, Glu132
	ZINC32004947	-146.72	5/ Arg10, Gly106, Ile131, Glu132
ZINC NP Compounds	ZINC67911471	-176.091	13/ Asn8, Cys13, Ser76, Ala77, Arg78 Arg150, Arg155, Gly174, Thr175
	ZINC31163223	-162.908	12/ Asn8, Arg10, His11, Arg150, Arg155, Gly172, Thr175
	ZINC04096393	-148.423	10/ Glu9, Arg10, Ala77, Arg78, Arg155, Lys156, Thr175

https://doi.org/10.1371/journal.pone.0186401.t011

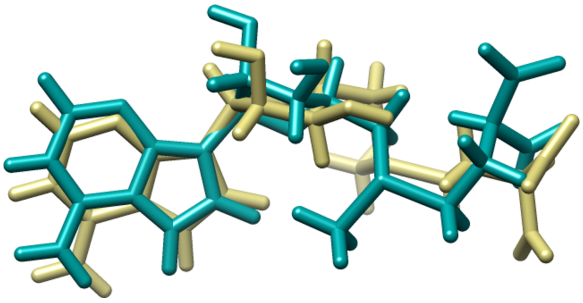

Fig 4. Superposition of co-crystallized and Docked ligand; Dark Khaki represents the co crystallized ligand and Dark Cyan the re-docked conformation of the ligand.

https://doi.org/10.1371/journal.pone.0186401.g004

the bacteriophage N protein, HIV tat and HIV rev. This suggestion is supported by the presence of a phosphate-binding site at the N-terminal end of α-A in each NusB protomer that includes a pair of conserved arginines, Arg10 and Arg14 [46]. The bismuth-dithiol solutions have been shown to selectively inhibit *Escherichia coli* rho transcription termination factor [47]. A comparison between the crystallographic structures of the NusB template (PDB ID: 1EYV, NusB from *M. tuberculosis*) and our modeled structure reveals that the conserved arginines were located at position 12 and 16 (Arg12 and Arg16) and are likely to contribute in the interactions. Although none of these residues are predicted to form hydrogen bonds with selected docked ligands, these molecules were predicted to interact with other residues in the pocket. **Table 5** shows the 8 selected ligands from all the four libraries according to their

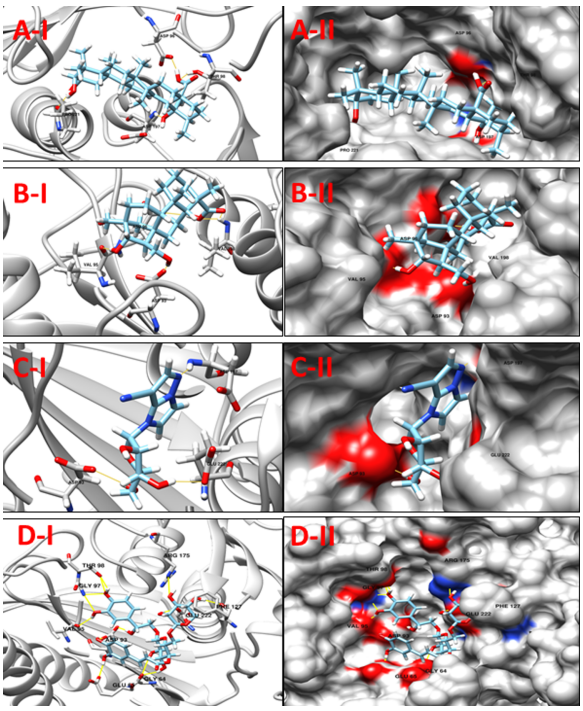

Fig 5. A-I: 3D cartoon representation of the docking analyses for the most druggable protein cavity of **NP_939302.1** (glpX, Fructose 1,6-bisphosphatase II) with Jacarandic Acid (CID 73645). **A-II:** 3D surface representation of the docking analyses for the structures of Jacarandic Acid with glpX protein. Figs **B-I, II**, **C-I, II** & **D-I, II** represent same information for compounds 16-hydrazonisosteviol, **ZINC13142972** and **ZINC67912153** respectively, for the same protein cavity.

https://doi.org/10.1371/journal.pone.0186401.g005

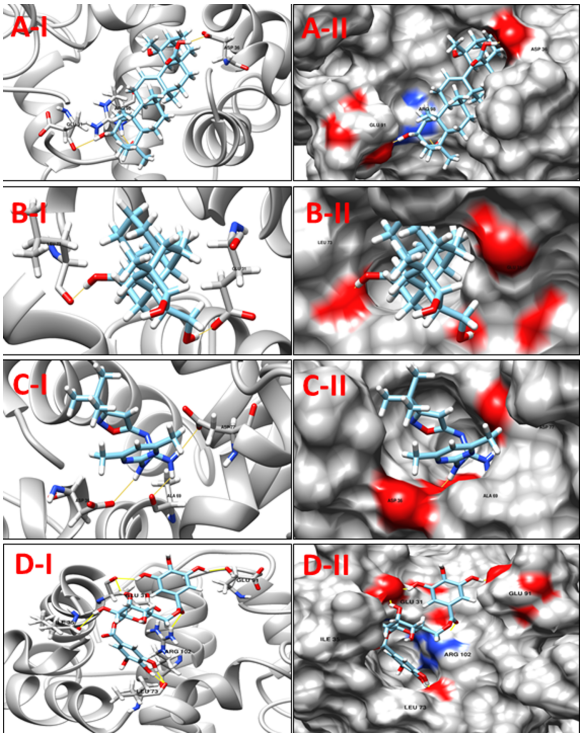

Fig 6. A-I: 3D cartoon representation of the docking analyses for the most druggable protein cavity of **NP_939692.1** (**nusB,** Transcription antitermination protein NusB) with Jacarandic Acid (CID 73645). **A-II:** 3D surface representation of the docking analyses for the structures of Jacarandic Acid with nusB protein. Figs **B-I, II**, **C-I, II** and **D-I, II** represent same information for compounds 16-hydrazonisosteviol, **ZINC00053531** and **ZINC15043210** respectively, for the same protein cavity.

https://doi.org/10.1371/journal.pone.0186401.g006

minimum energy values and the number of hydrogen bond interactions. The compounds **ZINC15043210, ZINC00053531** Jacarandic Acid and 16-hydrazonisosteviol are shown in (Fig 6). A decent binding mode and good shape complementarity was observed in these complexes.

NP_938900.1 (**rpsH,** 30S ribosomal protein S8) is an important RNA-binding protein that inhabits a central position within the small ribosomal subunit. It widely interacts with 16S rRNA and is vital for the correct folding of the central domain of the rRNA. The protein rpsH S8 also controls the synthesis of numerous ribosomal proteins by binding to mRNA. It binds exactly to very similar sites in the two RNA molecules. It is a ribosomal protein that has medium-size, and its role as a significant primary RNA-binding protein in the 30S subunit is discovered recently. The S8 mutations within the protein have been shown to result in defective ribosome assembly. In *Escherichia coli*, the S8-binding site within 16S rRNA has been investigated independently by a number of techniques including nuclease protection, RNA–protein crosslinking, RNA modification, hydroxyl-radical footprinting and chemical probing. The rpsH S8 protein is also one of the principal regulatory elements that control ribosomal protein synthesis by the translational feedback inhibition mechanism discovered by Nomura and colleagues [48]. It regulates the expression of the spc operon that encodes, in order, the ten ribosomal proteins L14, L24, L5, S14, S8, L6, L18, S5, L30 and L15 [49]. The active site residues of rpsH, based on a comparison with its template structure were Arg86, Tyr88, Ser107, Ser109, Gly124, Gly125 and Glu126. However, none of the molecules interacts with these residues (Table 6); nonetheless they are predicted to interact with other residues of the binding

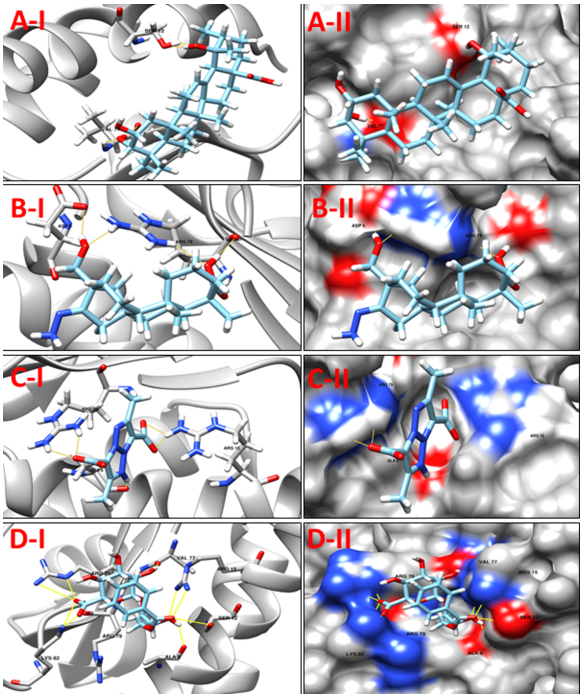

Fig 7. A-I 3D cartoon representation of the docking analyses for the most druggable protein cavity of **NP_938900.1** (**rpsH,** 30S ribosomal protein S8) with Jacarandic Acid (CID 73645). **A-II:** 3D surface representation of the docking analyses for the structures of Jacarandic Acid with rpsH protein. Figs **B-I, II**, **C-I, II** and **D-I, II** represent same information for compounds 17-hydroxyisosteviol **ZINC15221730** and **ZINC35457686** respectively, for the same cavity.

https://doi.org/10.1371/journal.pone.0186401.g007

cavity predicted by DoGSiteScorer. The predicted binding mode of best scoring compounds each library **ZINC35457686**, **ZINC15221730**, Jacarandic Acid and 17-hydroxyisosteviol are shown in Fig 7.

NP_938502.1 (**bioB,** Biotin synthase) catalyzes the final step in the biotin biosynthetic pathway by converting dethiobiotin (DTB) to biotin. This reaction uses organic radical chemistry for inserting sulfur atom between non activated carbons C6 and C9 of DTB. BioB is a member of the "radical SAM" or "AdoMet radical" superfamily, which is categorized by the presence of a conserved CxxxCxxC sequence motif (C, Cys; x, any amino acid) that synchronizes an essential Fe_4S_4 cluster, as well as by the use of S-adenosyl-Lmethionine (SAM or AdoMet) for radical generation. AdoMet radical enzymes act on a wide variety of biomolecules. For example, BioB and lipoyl-acyl carrier protein synthase (LipA) are involved in vitamin biosynthesis; lysine 2,3-aminomutase (LAM) facilitates the fermentation of lysine; class III ribonucleotide reductase (RNR) and pyruvate formate lyase (PFL) catalyze the formation of glycyl radicals in their respective target proteins; and spore photoproduct lyase repairs ultraviolet light-induced DNA damage [50]. The protein bioB was reported as putative drug target in *C. diphtheriae* by Barh *et al*., 2011 in their in silico study [15]. A comparison between our modeled protein and template structures suggest Cys86, Cys90, Cys93 and Arg291 as the active residues. Although, only Cys86, Cys90 and Cys93 were found to interact with the compounds from our prepared libraries, the molecules were predicted to interact with other residues in the pocket. The binding mode of compounds with active site residues and low scores suggest a set of 10 molecules (Table 7) as promising leads from our four libraries. The predicted binding

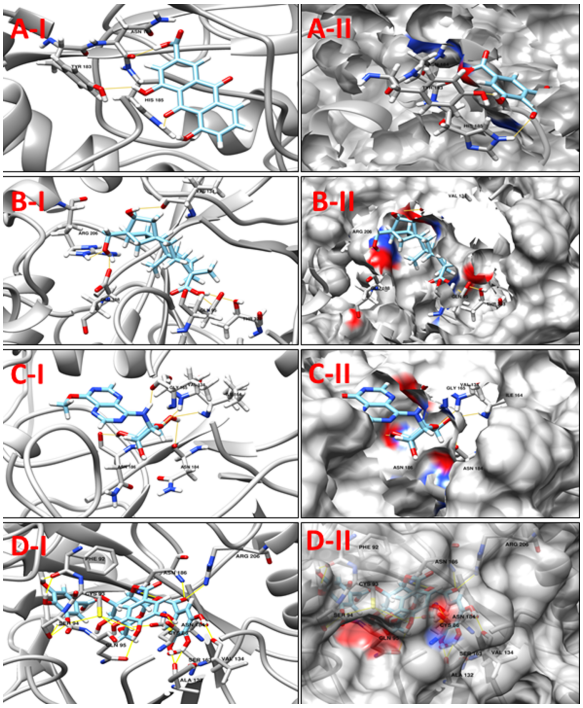

Fig 8. A-I 3D cartoon representation of the docking analyses for the most druggable protein cavity of **NP_938502.1** (**bioB,** Biotin synthase) with Rhein (CID 10168). **A-II:** 3D surface representation of the docking analyses for the structure of Rhein with bioB protein. Figs **B-I, II**, **C-I, II** & **D-I, II** represent same information for compounds 16-oxime, 17-hydroxyisosteviol, **ZINC16952914** and **ZINC77269615** respectively, for the same protein cavity.

https://doi.org/10.1371/journal.pone.0186401.g008

modes of Jacarandic Acid, 16-oxime, 17-hydroxyisosteviol, **ZINC16952914** and **ZINC77269615** are shown in Fig 8.

NP_939612.1 (**hisE,** Phosphoribosyl-ATP pyrophosphatase) is the second enzyme in the histidine-biosynthetic pathway, hydrolyzing irreversibly phosphoribosyl-ATP to phosphoribosyl-AMP and pyrophosphate. It is encoded by the *hisE* gene, which is present as a separate gene in many bacteria and archaea but is fused to *hisI* in other bacteria, fungi and plants. As it is essential for growth as seen in *in vitro* experiments, HisE is a potential drug target for tuberculosis [51]. A comparison of template and target protein structures here showed that there was no reported information about ligand-residue/s association in the active site cavity. Hence, the cavity chosen for virtual screening was simply the one that presented the highest DogSiteScorer druggability score (>80). A list of best dock molecules is shown below (Table 8). The binding patterns of Jacarandic Acid, 16–17 dihydroxyisosteviol, **ZINC05809437** and **ZINC67913372** are shown in Fig 9**.**

NP_939123.1 (**smpB,** SsrA-binding protein) is a small protein B (SmpB), which is very useful for biological functions of tmRNA. In bacteria, a hybrid RNA molecule that combines the functions of both messenger and transfer RNAs rescues stalled ribosomes, and targets aberrant, partially synthesized proteins for proteolytic degradation. The flexible RNA molecule adopts an open L-shaped conformation and SmpB binds to its elbow region, stabilizing the single-stranded D-loop in an extended conformation. The most prominent feature of the structure of $tmRNA_{\Delta}$ is a 90° rotation of the TΨC-arm around the helical axis. Because of this important conformation, the SmpB–tmRNA D-complex positioned into the A-site of the

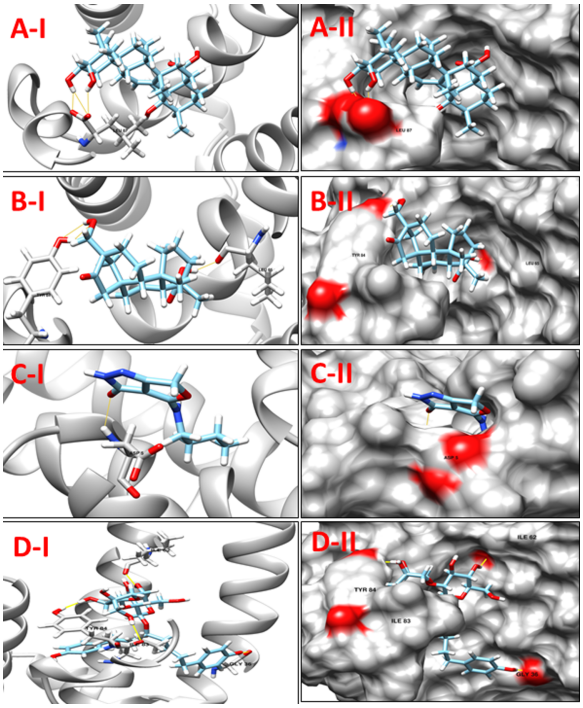

Fig 9. A-I 3D cartoon representation of the docking analyses for the most druggable protein cavity of **NP_939612.1** (**hisE,** Phosphoribosyl-ATP pyrophosphatase) with Jacarandic Acid (CID 73645). **A-II:** 3D surface representation of the docking analyses for the structure of Jacarandic Acid with hisE protein. Figs **B-I, II**, **C-I, II** & **D-I, II** represent same information for compounds 16–17 dihydroxyisosteviol, **ZINC05809437** and **ZINC67913372** respectively, for the same protein cavity.

https://doi.org/10.1371/journal.pone.0186401.g009

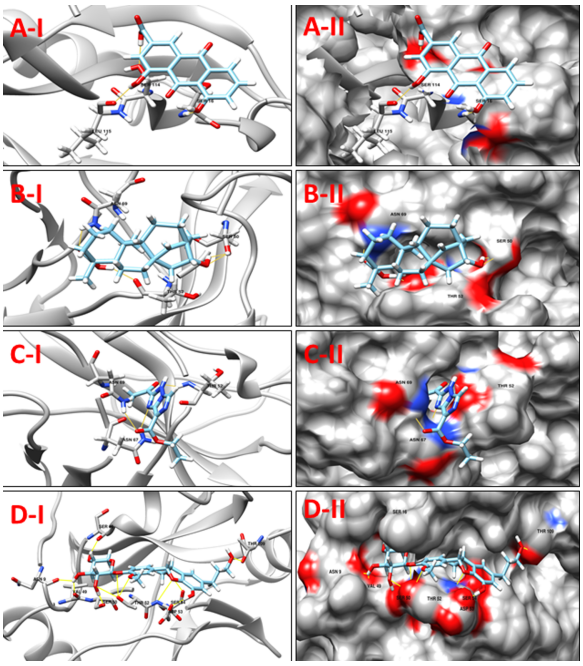

Fig 10. A-I 3D cartoon representation of the docking analyses for the most druggable protein cavity of **NP_939123.1** (**smpB,** SsrA-binding protein) with Rhein (CID 10168). **A-II:** 3D surface representation of the docking analyses for the structure of Rhein with smpB protein. Figs **B-I, II**, **C-I, II** & **D-I, II** represent same information for compounds 16-hydroxyisosteviol **ZINC01414475** & **ZINC31168211** respectively, for the same protein cavity.

https://doi.org/10.1371/journal.pone.0186401.g010

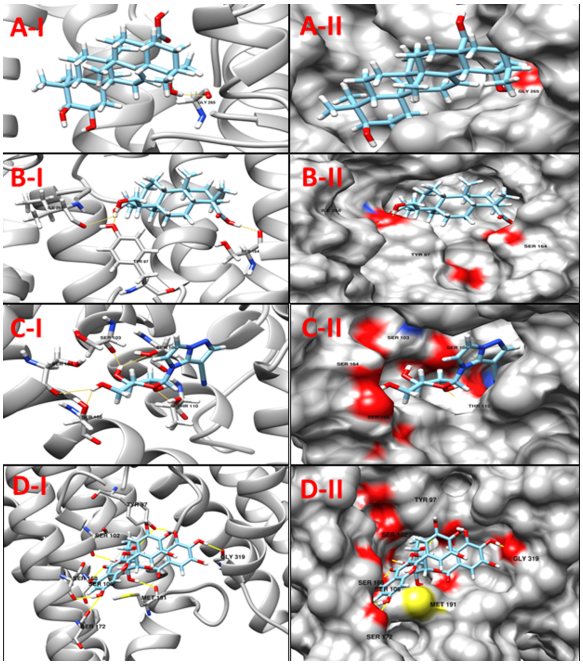

Fig 11. A-I 3D cartoon representation of the docking analyses for the most druggable protein cavity of **NP_939445.1** (**DIP1084,** Putative iron transport membrane protein, FecCD-family) with Jacarandic Acid (CID 73645). **A-II:** 3D surface representation of the docking analyses for the structure of Jacarandic Acid with **DIP1084,** Putative iron transport membrane protein. Figs **B-I, II, C-I, II** & **D-1, II D** represent same information for compounds 16-hydrazonisosteviol **ZINC13142972** and **ZINC70454922** respectively, for the same protein cavity.

https://doi.org/10.1371/journal.pone.0186401.g011

ribosome orients SmpB towards the small ribosomal subunit, and directs tmRNA towards the elongation-factor binding region of the ribosome. The tmRNA–SmpB rescue system is ubiquitous in bacteria, and is also found in some chloroplasts and mitochondria [52]. In this case the template structure (PDB ID: 1P6V) did not contain any ligand, and no reported information was found about the ligand-residue interaction in their cavities. Therefore, amongst the cavities identified by MVD, the best cavity for docking analysis was chosen in consensus with highest druggability score from the DogSiteScorer. **ZINC31168211** was found to form the network of 12 hydrogen bonds with Asn9, Ser16, Val49, Ser50, Thr52, Asp53, Ser54, Thr109. **Table 9** lists top compounds from respective libraries selected for this target while the binding modes of Rhein, 16-hydroxyisosteviol, **ZINC01414475 and ZINC31168211** are also shown (**Fig 10**).

NP_939445.1 (**DIP1084,** Putative iron transport membrane protein, FecCD-family) The Pfam search for the protein showed that it has two main components, FecCD and ABC_trans. The FecCD is a subfamily of bacterial binding-protein-dependent transport systems family constituting transport system permease proteins involved in the transport of numerous compounds through the membrane. These transporters tend to catalyze the thermodynamically unfavorable translocation of substrates against a transmembrane concentration gradient through the coupling to a second, energetically favorable process. ABC systems can be categorized in three functional groups, as follows. Importers mediate the uptake of nutrients in prokaryotes. The nature of the substrates that are transported is very wide, including mono- and oligosaccharides, organic and inorganic ions, amino acids, peptides, iron-siderophores, metals, polyamine cations, opines, and vitamins [53]. Exporters are involved in the secretion of various molecules, such as peptides, lipids, hydrophobic drugs, polysaccharides, and proteins, including toxins such as hemolysin. The third category of systems is apparently not involved

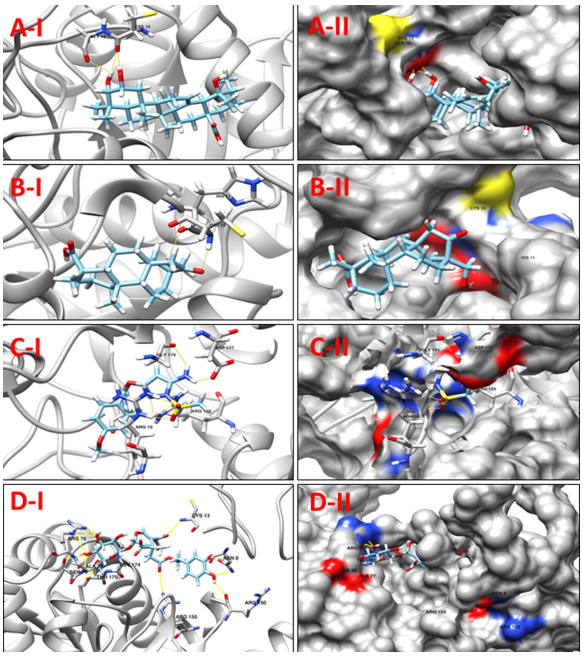

Fig 12. A-1: 3D cartoon representation of the docking analyses for the most druggable protein cavity of **NP_939345.1** (**DIP0983,** Hypothetical protein DIP0983) with Jacarandic Acid (CID 73645). **A-II:** 3D surface representation of the docking analyses for the structure of Jacarandic Acid with Hypothetical protein DIP0983. Figs **B-I, II, C-I, II** & **D-I, II** represent same information for compounds 17-hydroxyisosteviol, **ZINC00211173** and **ZINC67911471** respectively, for the same protein cavity.

https://doi.org/10.1371/journal.pone.0186401.g012

in transport, with some members being involved in translation of mRNA and in DNA repair. Table 10 shows a set of 11 high scoring compounds against the proposed target. Compound **ZINC70454922** from ZINC NP library was predicted to form ten hydrogen bonds with relatively low docking score (Fig 11).

NP_939345.1 (**DIP0983,** Hypothetical protein DIP0983) is a conserved hypothetical protein. It is annotated as a possible lysine decarboxylase (LDC) in the Pfam database (PF03641) [54] due to the presence of the highly conserved PGGxGTxxE motif. Some enzymes *i:e* "Lonely Guy" LOG are often mis-annotated as lysine decarboxylases enzymes; it is apparently responsible for catalyzing L-lysine decarboxylation to produce the polyamine metabolite cadaverine [55]. Conversely, this annotation is not supported by any biochemical or functional data in any of the PGGxGTxxE motif containing LDC identified so far. This motif is highly conserved among a vast number of proteins with unknown function, predicted from bacterial, yeast, and plant; in *Arabidopsis thaliana*, all the genome-annotated LOG proteins are identified as LDC like proteins by protein family. Based on sequence BLAST against the PDB, LOG from *Claviceps purpurea* shares more than 30% identical residues with crystal structures of LDC-like proteins of unknown function, whose structures are already determined. Recently, lysine decarboxylase has been reported as a therapeutic target by Lohinai *et al*., 2015 for Periodontal Inflammation [56]. Here we listed 12 compounds showing good potency against our target tabulated in Table 11. Four of the compounds with promising docking results are shown in Fig 12.

Among the drug-like molecule **ZINC13142972** (1-[(2S, 3S, 4S, 5R)-3,4-dihydroxy-5-(hydroxymethyl) oxolan-2-yl]imidazo[1,2-b]pyrazole-7-carbonitrile) was predicted to show good results against two of our targets **NP_939302.1** (**glpX,** Fructose 1,6-bisphosphatase II) and **NP_939445.1** (**DIP1084,** Putative iron transport membrane protein, FecCD-family). It

has been reported that at present 50% of drug molecules are either from natural source or their derivatives [57]. Interestingly, the compounds from second library of ZINC (Natural Product) showed better energy scores among all the libraries. Furthermore, from the library of natural compounds (28 molecules), Jacarandic Acid and Rhein were identified as the top ranked molecules and *in silico* analysis of the library (derivatives of diterpenoid isosteviol) suggest that compounds 16-hydroxyisosteviol, 16-hydrazonisosteviol, 17-hydroxyisosteviol, 16–17 dihydroxyisosteviol and 16-oxime, 17-hydroxyisosteviol were top ranked molecules, however, with much higher energy scores (less negative) than the top compounds from the ZINC libraries (ZINC drug-like molecules, ZINC Natural Product).

Conclusion

We utilized a bioinformatics pipeline for determining the conserved proteome of 13 strains of *C. diphtheriae*, and subsequently exploit 3D structural information, resulting in a small set of prioritized putative drug/vaccine targets, of which eight proteins are pathogen-essential, non-host homologous and 15 are pathogen-essential, host-homologs. After a detailed structural comparison between host and pathogen proteins, we suggest that eight of the non -host homologs could be considered for antimicrobial chemotherapy in future studies on anti-diphtheriae drugs and vaccines. Moreover, the strategy described herein is of general nature and can also be employed to other pathogenic microorganisms.

Supporting information

S1 Table. Structural information of the Di-terpenoid Iso-steviol derivatives.
(DOCX)

S2 Table. Information of templates used for 8 essential non host homologous targets.
(DOCX)

S3 Table. Common conserved proteins with their templates.
(XLS)

Acknowledgments

We would like to acknowledge wealthy cooperation and assistance of all the team members & collaborators.

Author Contributions

Conceptualization: Syed Babar Jamal, Syed Shah Hassan, Vasco A. C. Azevedo.

Data curation: Syed Babar Jamal, Syed Shah Hassan, Sandeep Tiwari, Marcus V. Viana, Leandro de Jesus Benevides, Asad Ullah, Adrián G. Turjanski, Daniela Arruda Costa.

Formal analysis: Syed Babar Jamal, Syed Shah Hassan, Sandeep Tiwari, Marcus V. Viana, Leandro de Jesus Benevides, Asad Ullah, Adrián G. Turjanski, Daniela Arruda Costa.

Investigation: Debmalya Barh, Preetam Ghosh, Richard Röttger, Jan Baumbach, Vasco A. C. Azevedo.

Methodology: Syed Babar Jamal, Sandeep Tiwari.

Project administration: Vasco A. C. Azevedo.

Resources: Vasco A. C. Azevedo.

Supervision: Artur Silva, Jan Baumbach, Vasco A. C. Azevedo.

Validation: Syed Babar Jamal, Sandeep Tiwari, Artur Silva, Vasco A. C. Azevedo.

Visualization: Syed Babar Jamal.

Writing – original draft: Syed Babar Jamal, Sandeep Tiwari.

Writing – review & editing: Syed Shah Hassan, Sandeep Tiwari, Adrián G. Turjanski, Debmalya Barh, Preetam Ghosh, Daniela Arruda Costa, Artur Silva, Richard Röttger, Jan Baumbach, Vasco A. C. Azevedo.

References

1. Funke G, von Graevenitz A, Clarridge JE, 3rd, Bernard KA. Clinical microbiology of coryneform bacteria. Clin Microbiol Rev. 1997; 10(1):125–59. PMID: 8993861; PubMed Central PMCID: PMCPMC172946.
2. Goodfellow M, Kämpfer P,. Busse HJ, Trujillo M, Suzuki KI, Ludwig W. Whitman Bergey's manual of systematic bacteriology: Springer; 2012.
3. Hodes HL. Diphtheria. Pediatr Clin North Am. 1979; 26(2):445–59. PMID: 379784.
4. Hart PE, Lee PY, Macallan DC, Wansbrough-Jones MH. Cutaneous and pharyngeal diphtheria imported from the Indian subcontinent. Postgrad Med J. 1996; 72(852):619–20. PMID: 8977947; PubMed Central PMCID: PMCPMC2398589.
5. Wagner KS, White JM, Crowcroft NS, De Martin S, Mann G, Efstratiou A. Diphtheria in the United Kingdom, 1986–2008: the increasing role of Corynebacterium ulcerans. Epidemiol Infect. 2010; 138 (11):1519–30. https://doi.org/10.1017/S0950268810001895 PMID: 20696088.
6. Barh D, Gupta K, Jain N, Khatri G, Leon-Sicairos N, Canizalez-Roman A, et al. Conserved host-pathogen PPIs. Globally conserved inter-species bacterial PPIs based conserved host-pathogen interactome derived novel target in C. pseudotuberculosis, C. diphtheriae, M. tuberculosis, C. ulcerans, Y. pestis, and E. coli targeted by Piper betel compounds. Integr Biol (Camb). 2013; 5(3):495–509. https://doi.org/10.1039/c2ib20206a PMID: 23288366.
7. Perumal D, Lim CS, Sakharkar KR, Sakharkar MK. Differential genome analyses of metabolic enzymes in Pseudomonas aeruginosa for drug target identification. In Silico Biol. 2007; 7(4–5):453–65. PMID: 18391237.
8. Pizza M, Scarlato V, Masignani V, Giuliani MM, Arico B, Comanducci M, et al. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science. 2000; 287 (5459):1816–20. PMID: 10710308.
9. Asif SM, Asad A, Faizan A, Anjali MS, Arvind A, Neelesh K, et al. Dataset of potential targets for Mycobacterium tuberculosis H37Rv through comparative genome analysis. Bioinformation. 2009; 4(6):245–8. PMID: 20975918; PubMed Central PMCID: PMCPMC2951718.
10. Chong CE, Lim BS, Nathan S, Mohamed R. In silico analysis of Burkholderia pseudomallei genome sequence for potential drug targets. In Silico Biol. 2006; 6(4):341–6. PMID: 16922696.
11. Dutta A, Singh SK, Ghosh P, Mukherjee R, Mitter S, Bandyopadhyay D. In silico identification of potential therapeutic targets in the human pathogen Helicobacter pylori. In Silico Biol. 2006; 6(1–2):43–7. PMID: 16789912.
12. Sakharkar KR, Sakharkar MK, Chow VT. A novel genomics approach for the identification of drug targets in pathogens, with special reference to Pseudomonas aeruginosa. In Silico Biol. 2004; 4(3):355–60. PMID: 15724285.
13. Barh D, Kumar A. In silico identification of candidate drug and vaccine targets from various pathways in Neisseria gonorrhoeae. In Silico Biol. 2009; 9(4):225–31. PMID: 20109152.
14. Rathi B, Sarangi AN, Trivedi N. Genome subtraction for novel target definition in Salmonella typhi. Bioinformation. 2009; 4(4):143–50. PMID: 20198190; PubMed Central PMCID: PMCPMC2825597.
15. 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. 2011; 78(1):73–84. https://doi.org/10.1111/j.1747-0285.2011.01118.x PMID: 21443692.
16. Aronov AM, Verlinde CL, Hol WG, Gelb MH. Selective tight binding inhibitors of trypanosomal glyceraldehyde-3-phosphate dehydrogenase via structure-based drug design. J Med Chem. 1998; 41 (24):4790–9. https://doi.org/10.1021/jm9802620 PMID: 9822549.

17. Singh S, Malik BK, Sharma DK. Molecular modeling and docking analysis of Entamoeba histolytica glyceraldehyde-3 phosphate dehydrogenase, a potential target enzyme for anti-protozoal drug development. Chem Biol Drug Des. 2008; 71(6):554–62. https://doi.org/10.1111/j.1747-0285.2008.00666.x PMID: 18489439.

18. 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. 2014; 15 Suppl 7:S3. https://doi.org/10.1186/1471-2164-15-S7-S3 PMID: 25573232; PubMed Central PMCID: PMCPMC4243142.

19. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, et al. Comparative protein structure modeling using MODELLER. Curr Protoc Protein Sci. 2007;Chapter 2:Unit 2 9. https://doi.org/10.1002/0471140864.ps0209s50 PMID: 18429317.

20. Mount DW. Using the Basic Local Alignment Search Tool (BLAST). CSH Protoc. 2007; 2007:pdb top17. https://doi.org/10.1101/pdb.top17 PMID: 21357135.

21. Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics. 2001; 17(9):849–50. PMID: 11590105.

22. Laskowski RA, MacArthur MW, Moss DS and Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography. 1993; 26. Epub 291. https://doi.org/10.1107/S0021889892009944

23. Blom J, Albaum SP, Doppmeier D, Puhler A, Vorholter FJ, Zakrzewski M, et al. EDGAR: a software framework for the comparative analysis of prokaryotic genomes. BMC Bioinformatics. 2009; 10:154. https://doi.org/10.1186/1471-2105-10-154 PMID: 19457249; PubMed Central PMCID: PMCPMC2696450.

24. Abadio AK, Kioshima ES, Teixeira MM, Martins NF, Maigret B, Felipe MS. Comparative genomics allowed the identification of drug targets against human fungal pathogens. BMC Genomics. 2011; 12:75. https://doi.org/10.1186/1471-2164-12-75 PMID: 21272313; PubMed Central PMCID: PMCPMC3042012.

25. Zhang R, Ou HY, Zhang CT. DEG: a database of essential genes. Nucleic Acids Res. 2004; 32(Database issue):D271–2. https://doi.org/10.1093/nar/gkh024 PMID: 14681410; PubMed Central PMCID: PMCPMC308758.

26. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28 (1):27–30. PMID: 10592173; PubMed Central PMCID: PMCPMC102409.

27. Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford). 2011; 2011:bar009. https://doi.org/10.1093/database/bar009 PMID: 21447597; PubMed Central PMCID: PMCPMC3070428.

28. Yoon SH, Park YK, Lee S, Choi D, Oh TK, Hur CG, et al. Towards pathogenomics: a web-based resource for pathogenicity islands. Nucleic Acids Res. 2007; 35(Database issue):D395–400. https://doi.org/10.1093/nar/gkl790 PMID: 17090594; PubMed Central PMCID: PMCPMC1669727.

29. Yu CS, Lin CJ, Hwang JK. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci. 2004; 13(5):1402–6. https://doi.org/10.1110/ps.03479604 PMID: 15096640; PubMed Central PMCID: PMCPMC2286765.

30. Aguero F, Al-Lazikani B, Aslett M, Berriman M, Buckner FS, Campbell RK, et al. Genomic-scale prioritization of drug targets: the TDR Targets database. Nat Rev Drug Discov. 2008; 7(11):900–7. https://doi.org/10.1038/nrd2684 PMID: 18927591; PubMed Central PMCID: PMCPMC3184002.

31. Butt AM, Nasrullah I, Tahir S, Tong Y. Comparative genomics analysis of Mycobacterium ulcerans for the identification of putative essential genes and therapeutic candidates. PLoS One. 2012; 7(8): e43080. https://doi.org/10.1371/journal.pone.0043080 PMID: 22912793; PubMed Central PMCID: PMCPMC3418265.

32. Volkamer A, Kuhn D, Rippmann F, Rarey M. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. Bioinformatics. 2012; 28(15):2074–5. https://doi.org/10.1093/bioinformatics/bts310 PMID: 22628523.

33. Tiwari S, da Costa MP, Almeida S, Hassan SS, Jamal SB, Oliveira A, et al. C. pseudotuberculosis Phop confers virulence and may be targeted by natural compounds. Integr Biol (Camb). 2014; 6(11):1088–99. https://doi.org/10.1039/c4ib00140k PMID: 25212181.

34. Voigt JH, Bienfait B, Wang S, Nicklaus MC. Comparison of the NCI open database with seven large chemical structural databases. J Chem Inf Comput Sci. 2001; 41(3):702–12. PMID: 11410049.

35. Wadood A, Jamal SB, Riaz M, Mir A. Computational analysis of benzofuran-2-carboxlic acids as potent Pim-1 kinase inhibitors. Pharm Biol. 2014; 52(9):1170–8. https://doi.org/10.3109/13880209.2014.880488 PMID: 24766364.

36. Thomsen R, Christensen MH. MolDock: a new technique for high-accuracy molecular docking. J Med Chem. 2006; 49(11):3315–21. https://doi.org/10.1021/jm051197e PMID: 16722650.

37. 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. 2004; 25(13):1605–12. https://doi.org/10.1002/jcc.20084 PMID: 15264254.

38. Caffrey CR, Rohwer A, Oellien F, Marhofer RJ, Braschi S, Oliveira G, et al. A comparative chemogenomics strategy to predict potential drug targets in the metazoan pathogen, Schistosoma mansoni. PLoS One. 2009; 4(2):e4413. https://doi.org/10.1371/journal.pone.0004413 PMID: 19198654; PubMed Central PMCID: PMCPMC2635471.

39. Crowther GJ, Shanmugam D, Carmona SJ, Doyle MA, Hertz-Fowler C, Berriman M, et al. Identification of attractive drug targets in neglected-disease pathogens using an in silico approach. PLoS Negl Trop Dis. 2010; 4(8):e804. https://doi.org/10.1371/journal.pntd.0000804 PMID: 20808766; PubMed Central PMCID: PMCPMC2927427.

40. Shanmugham B, Pan A. Identification and characterization of potential therapeutic candidates in emerging human pathogen Mycobacterium abscessus: a novel hierarchical in silico approach. PLoS One. 2013; 8(3):e59126. https://doi.org/10.1371/journal.pone.0059126 PMID: 23527108; PubMed Central PMCID: PMCPMC3602546.

41. Folador EL, de Carvalho PV, Silva WM, Ferreira RS, Silva A, Gromiha M, et al. In silico identification of essential proteins in Corynebacterium pseudotuberculosis based on protein-protein interaction networks. BMC Syst Biol. 2016; 10(1):103. https://doi.org/10.1186/s12918-016-0346-4 PMID: 27814699; PubMed Central PMCID: PMCPMC5097352.

42. Wadood A, Riaz M, Jamal SB, Shah M. Interactions of ketoamide inhibitors on HCV NS3/4A protease target: molecular docking studies. Mol Biol Rep. 2014; 41(1):337–45. https://doi.org/10.1007/s11033-013-2867-x PMID: 24234753.

43. Horecker BL, Melloni E, Pontremoli S. Fructose 1,6-bisphosphatase: properties of the neutral enzyme and its modification by proteolytic enzymes. Adv Enzymol Relat Areas Mol Biol. 1975; 42:193–226. PMID: 236638.

44. Wright SW, Carlo AA, Carty MD, Danley DE, Hageman DL, Karam GA, et al. Anilinoquinazoline inhibitors of fructose 1,6-bisphosphatase bind at a novel allosteric site: synthesis, in vitro characterization, and X-ray crystallography. J Med Chem. 2002; 45(18):3865–77. PMID: 12190310.

45. Sassetti CM, Rubin EJ. Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci U S A. 2003; 100(22):12989–94. https://doi.org/10.1073/pnas.2134250100 PMID: 14569030; PubMed Central PMCID: PMCPMC240732.

46. Gopal B, Haire LF, Cox RA, Jo Colston M, Major S, Brannigan JA, et al. The crystal structure of NusB from Mycobacterium tuberculosis. Nat Struct Biol. 2000; 7(6):475–8. https://doi.org/10.1038/75876 PMID: 10881194.

47. Brogan AP, Verghese J, Widger WR, Kohn H. Bismuth-dithiol inhibition of the Escherichia coli rho transcription termination factor. J Inorg Biochem. 2005; 99(3):841–51. https://doi.org/10.1016/j.jinorgbio.2004.12.019 PMID: 15708806.

48. Yates JL, Arfsten AE, Nomura M. In vitro expression of Escherichia coli ribosomal protein genes: autogenous inhibition of translation. Proc Natl Acad Sci U S A. 1980; 77(4):1837–41. PMID: 6445562; PubMed Central PMCID: PMCPMC348603.

49. Davies C, Ramakrishnan V, White SW. Structural evidence for specific S8-RNA and S8-protein interactions within the 30S ribosomal subunit: ribosomal protein S8 from Bacillus stearothermophilus at 1.9 A resolution. Structure. 1996; 4(9):1093–104. PMID: 8805594.

50. Berkovitch F, Nicolet Y, Wan JT, Jarrett JT, Drennan CL. Crystal structure of biotin synthase, an S-adenosylmethionine-dependent radical enzyme. Science. 2004; 303(5654):76–9. https://doi.org/10.1126/science.1088493 PMID: 14704425; PubMed Central PMCID: PMCPMC1456065.

51. Javid-Majd F, Yang D, Ioerger TR, Sacchettini JC. The 1.25 A resolution structure of phosphoribosyl-ATP pyrophosphohydrolase from Mycobacterium tuberculosis. Acta Crystallogr D Biol Crystallogr. 2008; 64(Pt 6):627–35. https://doi.org/10.1107/S0907444908007105 PMID: 18560150; PubMed Central PMCID: PMCPMC2631106.

52. Gutmann S, Haebel PW, Metzinger L, Sutter M, Felden B, Ban N. Crystal structure of the transfer-RNA domain of transfer-messenger RNA in complex with SmpB. Nature. 2003; 424(6949):699–703. https://doi.org/10.1038/nature01831 PMID: 12904796.

53. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014; 42(Database issue):D222–30. https://doi.org/10.1093/nar/gkt1223 PMID: 24288371; PubMed Central PMCID: PMCPMC3965110.

54. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, et al. The Pfam protein families database. Nucleic Acids Res. 2002; 30(1):276–80. PMID: 11752314; PubMed Central PMCID: PMCPMC99071.

55. Dzurova L, Forneris F, Savino S, Galuszka P, Vrabka J, Frebort I. The three-dimensional structure of "Lonely Guy" from Claviceps purpurea provides insights into the phosphoribohydrolase function of Rossmann fold-containing lysine decarboxylase-like proteins. Proteins. 2015; 83(8):1539–46. https://doi.org/10.1002/prot.24835 PMID: 26010010.

56. Lohinai Z, Keremi B, Szoko E, Tabi T, Szabo C, Tulassay Z, et al. Biofilm lysine Decarboxylase, a New Therapeutic Target for Periodontal Inflammation. J Periodontol. 2015:1–15. https://doi.org/10.1902/jop.2015.140490 PMID: 26110450.

57. Veeresham C. Natural products derived from plants as a source of drugs. J Adv Pharm Technol Res. 2012; 3(4):200–1. https://doi.org/10.4103/2231-4040.104709 PMID: 23378939; PubMed Central PMCID: PMCPMC3560124.

III.2.2. Conclusion, Chapter 2

- The genomic information was used with the aim of determining the conserved predicted proteome of 13 strains of *C. diphtheriae*, together with the 3D structural information.
- The measurements for target selection in *C. diphtheriae* was kept strict, resulting in a small set of prioritized putative drug/vaccine targets, of which eight are essential and non-host homologous, and 15 are essential and host homologous proteins.
- The analysis of protein models was carried out very carefully, based on good templates which provided high-quality templates, lessening the risks. The data presented here can effectively contribute to future research for the development of novel antibiotics and vaccines.
- After a detailed structural comparison between host and pathogen proteins, we suggest that eight non-host homologous proteins could be considered for antimicrobial chemotherapy owing to further experimental validations.
- We expect that the *in silico* computational approaches adopted in this study might aid in the development of novel therapeutic drugs and vaccines against *C. diphtheriae*. Furthermore, we extended our study by adopting another computational approach to validate the results obtained from this work.

Chapter 3

III.3.1. Research Article

The Druggable Pocketome of *Corynebacterium diphtheriae:* A New Approach for in silico Putative Druggable Targets.

Syed S. Hassan†, Syed B. Jamal†, Leandro G. Radusky†, Sandeep Tiwari†, Asad Ullah, Javed Ali, Behramand, Paulo V. S. D. de Carvalho, Rida Shams, Sabir Khan, Henrique C. P. Figueiredo, Debmalya Barh, Preetam Ghosh, Artur Silva, Jan Baumbach, Richard Röttger, Adrián G. Turjanski, and Vasco A. C. Azevedo.

Frontiers in Genetics, 2018, 9:44. doi: 10.3389/fgene.2018.00044

This article was planned for the cross validation and robustness of our previous work Jamal et al., 2017. As discussed earlier, we identified 8 targets that were essential for pathogen and non-host homologous. In this work, we used another *in silico* approach using same 13 complete genome sequences of *C. diphtheriae* followed by a computational assessment of structural information of the binding sites to characterize the "pocketome druggability." As a result, a final set of 10-conserved targets possessing highly druggable protein pockets were identified. To check the target identification robustness of the pipeline used in this work, we crosschecked the final target list with another in-house target identification approach for *C. diphtheriae* thereby obtaining three common targets, these were; hisE-phosphoribosyl-ATP pyrophosphatase, glpX-fructose 1,6-bisphosphatase II, and rpsH 30S ribosomal protein S8 as identified in Jamal et al., 2017. Our predicted results suggest that the *in silico* approach used could potentially aid in experimental polypharmacological target determination in *C. diphtheriae* and other pathogens, thereby, might complement the existing and new drug-discovery pipelines.

frontiers in Genetics

ORIGINAL RESEARCH
published: 13 February 2018
doi: 10.3389/fgene.2018.00044

The Druggable Pocketome of *Corynebacterium diphtheriae*: A New Approach for *in silico* Putative Druggable Targets

Syed S. Hassan[1†], Syed B. Jamal[2†], Leandro G. Radusky[3†], Sandeep Tiwari[2†], Asad Ullah[1], Javed Ali[4], Behramand[1], Paulo V. S. D. de Carvalho[2], Rida Shams[1], Sabir Khan[5], Henrique C. P. Figueiredo[6], Debmalya Barh[2,7], Preetam Ghosh[8], Artur Silva[9], Jan Baumbach[10], Richard Röttger[10], Adrián G. Turjanski[3,11] and Vasco A. C. Azevedo[2]**

[1] Department of Chemistry, Islamia College University Peshawar, Peshawar, Pakistan, [2] PG Program in Bioinformatics, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil, [3] Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina, [4] Department of Chemistry, Kohat University of Science and Technology, Kohat, Pakistan, [5] Department of Analytical Chemistry, Institute of Chemistry, São Paulo State University, São Paulo, Brazil, [6] AQUACEN, National Reference Laboratory for Aquatic Animal Diseases, Ministry of Fisheries and Aquaculture, Federal University of Minas Gerais, Belo Horizonte, Brazil, [7] Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology, Purba Medinipur, India, [8] Department of Computer Science, Virginia Commonwealth University, Richmond, VA, United States, [9] Institute of Biological Sciences, Federal University of Pará, Belém, Brazil, [10] Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark, [11] INQUIMAE/UBA-CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

Diphtheria is an acute and highly infectious disease, previously regarded as endemic in nature but vaccine-preventable, is caused by *Corynebacterium diphtheriae* (Cd). In this work, we used an *in silico* approach along the 13 complete genome sequences of *C. diphtheriae* followed by a computational assessment of structural information of the binding sites to characterize the "pocketome druggability." To this end, we first computed the "modelome" (3D structures of a complete genome) of a randomly selected reference strain Cd NCTC13129; that had 13,763 open reading frames (ORFs) and resulted in 1,253 (~9%) structure models. The amino acid sequences of these modeled structures were compared with the remaining 12 genomes and consequently, 438 conserved protein sequences were obtained. The RCSB-PDB database was consulted to check the template structures for these conserved proteins and as a result, 401 adequate 3D models were obtained. We subsequently predicted the protein pockets for the obtained set of models and kept only the conserved pockets that had highly druggable (HD) values (137 across all strains). Later, an off-target host homology analyses was performed considering the human proteome using NCBI database. Furthermore, the gene essentiality analysis was carried out that gave a final set of 10-conserved targets possessing highly druggable protein pockets. To check the target identification robustness of the pipeline used in this work, we crosschecked the final target list with another in-house target identification approach for *C. diphtheriae* thereby obtaining three common targets, these were; hisE-phosphoribosyl-ATP pyrophosphatase, glpX-fructose 1,6-bisphosphatase II, and

OPEN ACCESS

Edited by:
Thiago Motta Venancio, State University of Norte Fluminense, Brazil

Reviewed by:
A. Maxwell Burroughs, National Center for Biotechnology Information (NIH), United States
Tiago Sobreira, Purdue University, United States

***Correspondence:**
Vasco A. C. Azevedo
vasco@icb.ufmg.br
Adrián G. Turjanski
adrian@qi.fcen.uba.ar

† These authors have contributed equally to this work.

Specialty section:
This article was submitted to Bioinformatics and Computational Biology, a section of the journal Frontiers in Genetics

Received: *23 November 2017*
Accepted: *30 January 2018*
Published: *13 February 2018*

Citation:
Hassan SS, Jamal SB, Radusky LG, Tiwari S, Ullah A, Ali J, Behramand, de Carvalho PVSD, Shams R, Khan S, Figueiredo HCP, Barh D, Ghosh P, Silva A, Baumbach J, Röttger R, Turjanski AG and Azevedo VAC (2018) The Druggable Pocketome of Corynebacterium diphtheriae: A New Approach for in silico Putative Druggable Targets. Front. Genet. 9:44. doi: 10.3389/fgene.2018.00044

rpsH-30S ribosomal protein S8. Our predicted results suggest that the *in silico* approach used could potentially aid in experimental polypharmacological target determination in *C. diphtheriae* and other pathogens, thereby, might complement the existing and new drug-discovery pipelines.

Keywords: *Corynebacterium diphtheria*, pocketome, druggable genome, structural proteomics, putative therapeutic targets, highly druggable (HD), global druggable (GD)

INTRODUCTION

Corynebacterium diphtheriae belong to the class Actinomycetales and is a Gram-positive bacterium, a non-spore forming, non-motile and facultative anaerobe with pleomorphic cell shape and high GC content (∼53%) relative to the *Firmicutes* (Cerdeno-Tarraga et al., 2003; Trost et al., 2012). This bacterium is the causative agent of diphtheria, a severe human respiratory manifestation characterized by bacterial adhesion to host pharyngeal cell (pharyngitis and pseudomembranous inflammation). The pathogen target one or both tonsils that further disseminate at a later stage thereby resulting in complete airway obstruction and death (Hadfield et al., 2000). The cornerstone in diphtheria therapy involves the hyper immune antiserum-antitoxin produced in equines that neutralizes the *C. diphtheriae* toxin while among antibiotics are the broad-spectrum penicillin or erythromycin. However, recent emergence of numerous drug-resistant strains subsequently decreases the efficacy of current therapeutics (Barraud et al., 2011). Importantly, the World Health Organization recommends widespread DTPs immunization against toxigenic diphtheria strains as the only effective approach to counteract the infection. Although, a slight change in mortality has been observed since the availability and administration of antibiotics, specifically penicillin or erythromycin (Adler et al., 2013). *C. diphtheriae* has four biotypes: gravis, mitis, intermedius and belfanti that are non-sporulating, unencapsulated, non-motile and pleomorphic bacilli. They are subsequently classified on the basis of differences in colonial morphology, haemolytic potential, fermentation reactions and severity of the resulting disease (Gerald et al., 2009; Brooks et al., 2010). The infant mortality rate in an epidemic that resulted in thousands of casualties dropped gradually in countries where living standards were improved and immunization programs were introduced. Nevertheless, it still remains a significant pathogen around the globe (Hodes, 1979). The 'strangling angel' effects on children that scaled from wing-shaped disarticulation and pseudo-membranes formation in the oropharynx, triggered acute obstruction of airways and resulted in mortalities (Hodes, 1979; Hart et al., 1996; Jamal et al., 2017b). A plethora of cases were recently reported and still expected from both non-lethal and lethal diphtheria in different parts of the world due to significant population displacements via immigration. An adequate supervision necessitates quick measures to discover additional diphtheria antitoxin, antibiotic and therapeutic treatment (Pizza et al., 2000).

The emerging concepts of polypharmacology, differential genome analyses, and reverse vaccinology, comparative and subtractive microbial genomics have largely contributed by establishing complementary traditions for fast identification of novel targets in post-genomic era (Perumal et al., 2007; Barh et al., 2013). Comparative homology modeling (Baumbach, 2010; Rottger et al., 2013) has widely been used in expanding the structural space of pathogens (Chong et al., 2006; Asif et al., 2009).

These practices are being used for the identification of conserved targets in a several human and animal pathogens like *C. tuberculosis* (Hassan et al., 2014; Radusky et al., 2015), *Mycobacterium. tuberculosis* (Asif et al., 2009), *Burkholderia pseudomallei* (Chong et al., 2006), *Neisseria gonorrhoea* (Barh and Kumar, 2009), *Helicobacter pylori* (Dutta et al., 2006), *Pseudomonas aeruginosa* (Sakharkar et al., 2004; Perumal et al., 2007), and *Salmonella typhi* (Rathi et al., 2009).

In this work, a combination of *in silico* tools was primarily used to predict the core proteome of *C. diphtheriae* species to associate genomic information based on the 3D structures. The predicted proteomes were modeled (pan-modelome) using a methodology adapted by Hassan et al. (2014). From a structural point of view, druggability is the probability of small drug-like molecules binding to a given target protein with high affinity (<1 μM). We report for the first time the structural druggability assessment for multi-strain *C. diphtheriae* proteomes using a pan-druggability prediction pipeline based on the open source pocket detection code "fpocket". The method integrates several physicochemical descriptors to estimate the pocket druggability on a genomic scale with suitable features that enable binding of a drug-like compound (Kinnings et al., 2010).

MATERIALS AND METHODS

Initial Dataset Construction

All ORFs (Open Reading Frames) of the 13 completely sequenced genomes of *C. diphtheriae* were obtained from the NCBI database[1]. **Table 1** shows the statistical data of all strains used in this study where the strain NCTC13129 has 2,272 reported ORFs and was randomly selected as a reference genome for modelome prediction and further analyses.

General Concept: Modelome Prediction

The binding affinity of small drug-like molecules to the active site of putative biological targets (druggable protein cavities) formulated a basis for this work, a slightly modified protocol of Radusky et al. (2015) (**Figure 1**). All genome ORF sequences of the 13 *C. diphtheriae* strains were subjected to

[1]https://www.ncbi.nlm.nih.gov/genome/genomes/1025?

TABLE 1 | Summary of *Corynebacterium diphtheriae* strains used in this study and their respective modeling statistics for druggability analyses.

Strain	Biovar	Location	NCBI accession	Genome size (Mb)	GC%	Proteins	Models – G2
31A	N/A	Brazil	NC_016799.1	2.53535	53.60	2380	1283
241	N/A	Brazil	NC_016782.1	2.42655	53.40	2245	1235
BH8	N/A	Brazil	NC_016800.1	2.48552	53.60	2361	1269
C7	N/A	United States	NC_016801.1	2.49919	53.50	2337	1278
CDCE8392	Mitis	United States	NC_016785.1	2.43333	53.60	2249	1253
HC01	Mitis	Brazil	NC_016786.1	2.42715	53.40	2247	1236
HC02	Mitis	Brazil	NC_016802.1	2.46861	53.70	2230	1254
HC03	Mitis	Brazil	NC_016787.1	2.47836	53.50	2262	1260
HC04	Gravis	Brazil	NC_016788.1	2.48433	53.50	2275	1260
INCA402	Belfanti	Brazil	NC_016783.1	2.44907	53.70	2214	1282
PW8	N/A	United States	NC_016789.1	2.53068	53.50	2414	1272
VA01	Gravis	Brazil	NC_016790.1	2.39544	53.40	2191	1239
NCTC13129	Gravis	United Kingdom	NC_002935.2	2.48863	53.50	2272	1253

the MHOLline workflow[2] in.faa file format for 3D structure prediction. MHOLline utilizes multi fasta files of amino acids as an input data and then uses HMMTOP, BLAST, BATS, Modeller and Procheck programs for the detailed analyses. The program HMMTOP detects transmembrane regions. The BLAST algorithm is used to identify template structure by performing a random search against the Protein Data Bank. BATS (Blast Automatic Targeting for Structures) carry out the refinement in the template search; it is a key step for the model construction. BATS refinement identifies sequences that make the modeling possible by selecting a template from BLAST output file using their BATS scores, expectation values, identity and sequence similarity as criteria as well as considering the number of gaps and the alignment coverage. BATS select the best template for 3D model generation and perform automated alignment used by the MODELLER program. Furthermore, it gathers all the BLAST output files into four distinctive groups, i.e., G0, G1, G2, and G3, according to the following criteria; G0 = Not aligned sequence, G1 = *E*-value $> 10e^{-5}$ or Identity $< 15\%$, G2 = *E*-value $\leq 10e^{-5}$ and Identity $\geq 25\%$ AND LVI ≤ 0.7, G3 = *E*-value $\leq 10e^{-5}$ and Identity $\leq 15\%$ and $<25\%$ OR LVI > 0.7, Where LVI is the Length Variation Index, a MHOLline concept of coverage (LVI ≤ 0.1 is equivalent to a coverage $\geq 90\%$). Once the template is selected based on BATS results, MODELLER program is used for the generation of 3D protein model. There is no significant correlation statistically, between the number of templates used during model building and the overall quality of a model. In the next step, another MHOLline tool called FILTERS, categorizes the BATS selected sequences (G2) into distinct quality model subgroups, based on identity and LVI value. The subgroups ranges Very High to Very Low. To evaluate the overall quality and accuracy of the model, Ramachandran plot is obtained which explained the stereochemical quality of the model. Precisely, the MHOLine generates an aggregate structural information for all the submitted sequences in the fasta format, Ramachandran plot and other properties like structural quality and enzymatic functions are also determined. Further details can be obtained by visiting MHOLline homepage (Hassan et al., 2014; Webb and Sali, 2016; Jamal et al., 2017a). For all modeled structures, structural properties were figured as: (i) the Druggability Score (DS) for each pocket and (ii) the active site residues (if available) according to the template structures available at the protein databank RCSB-PDB[3] (Berman et al., 2003).

3D Protein Models in Non-reference Strains: Orthologs Identification

We used an applied bioinformatics procedure to find the conserved putative druggable targets across all the 13 *C. diphtheriae* strains at genome-scale by first predicting their 3D models. The ensemble methodology essentially is a filter of thousands of candidate genes to yield high-confidence 3D structural models from orthologous proteins in *C. diphtheriae* species. As aforementioned, the MHOLline resulted in 1,253 predicted structures for the randomly selected reference strain NCTC13129 that were later used as template structures for modeling the 3D structures in non-reference strains as well. Further, the BLASTp program was installed in a local machine and used to check if the ORFs of the reference proteome have orthologs in the remaining 12 strains using the following parameters; identity $\geq$ 85%, coverage $>$ 80%. The protein sequences showing high identity values ($>$ = 85%) for each reference and non-reference strain were considered as conserved and the modeled structures of reference strain were used again as templates to predict the 3D models for the aforementioned 12 non-reference strains. The core modelome was compared and evaluated for the quality of the obtained 3D structures. A reliable model has a probability of correct fold larger than 95% and coverage of over 50% with the template structure. For each sequence in the reference and non-reference strain that gave an identity hit of $>$85%, a mutation methodology was applied on each amino acid substitution using the MODELLER program. These models were then used to compute the druggability variation for the 13 strains of the *C. diphtheriae* species.

[2]http://www.mholline.lncc.br

[3]www.wwpdb.org

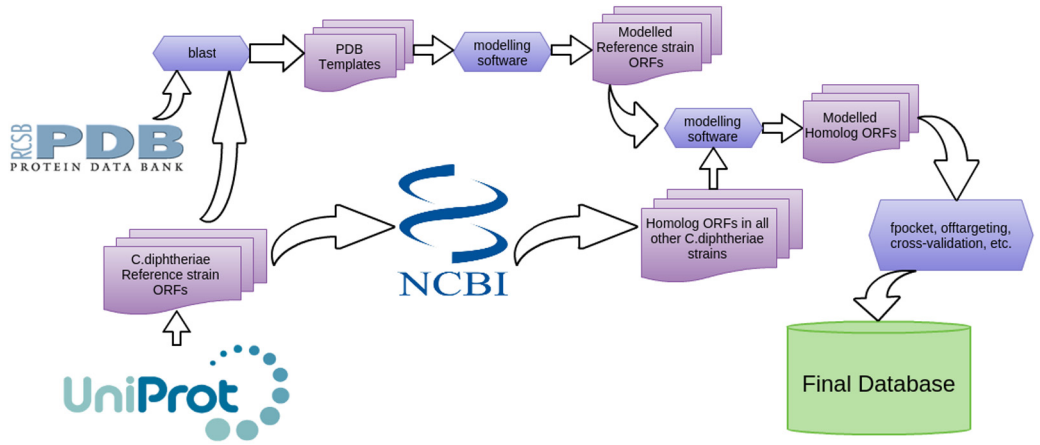

FIGURE 1 | Overview of pipeline components for data analysis.

Evaluation of Structural Druggability

The protein structural druggability of each predicted 3D model was evaluated by determining the ability of putative pockets to bind drug-like molecule/s, using the fpocket (Finn et al., 2016) and the recently developed DrugScore (DS) programs (Velec et al., 2005). The later methodology is based on the Voronoi tessellation algorithm that identify and characterize pockets and compute suitable physicochemical descriptors such as polar and apolar surface area, hydrophobic-hydrophobic density and polarity score. In conjunction they yield a druggability value that ranges between 0 (non-druggable, ND) and 1 (highly druggable, HD). We categorized the druggability scores for all predicted 3D structures into four sets: non-druggable (ND; DS $\leq$ 0.2), poorly druggable (PD; DS $\geq$ 0.2 and $\leq$0.5), druggable (D; DS $\geq$ 0.5 and $\leq$0.7), and highly druggable (HD; DS $\geq$ 0.7) protein pockets, respectively. This distribution is in accordance to our previous work where the druggability score was computed for all pockets present in all unique proteins in the Protein Data Bank that were experimentally crystallized in complex with a drug-like compound/s (Radusky et al., 2014).

Identification of Active Site Residues

In order to identify the amino acid residues in the active site of the predicted druggable protein pocket/s, information were retrieved from the CSA database (Catalytic Site Atlas) (Furnham et al., 2014) and Pfam position site (Finn et al., 2016), respectively. A list of PDB_IDs was generated linked to a number of residues constituting the corresponding protein active sites. To map the active site residues to as many *C. diphtheriae* proteins as possible, each PDB_ID was used as a template in CSA and assigned to the modeled ORFs.

Host Homology, Essentiality, and Core-Modelomics of the Selected Targets

For off target prediction, the pool of global druggable (GD) proteins was piped into NCBI-BLASTp using default parameters (identity = 0% and/or no hit) against the human proteome to identify non-host homologs. Moreover, from the filtered list of 10 highly druggable non-host homologous target proteins, an approach based on subtractive genomics was implemented and applied to the GD targets that were essential to bacteria (Barh et al., 2011). Briefly, the set of target proteins of *C. diphtheriae* was submitted to the Database of Essential Genes (DEG, which contains experimentally validated essential genes from bacteria, archaea and eukaryotes) for homology analyses (Zhang et al., 2004). Again, we used BLASTp with *E*-value cut-off of $1e^{-05}$ a *bit score* $\geq$ 100 and *identity* $\geq$ 35% (Barh et al., 2011). The final list of putative targets based on criteria described earlier, contained 10 essential and non-host homologous target proteins. The obtained list was further subjected to ProtParam[4] for molecular weight determination, biochemical pathway analysis to KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa and Goto, 2000) using network enrichment (Alcaraz et al., 2012), virulence using PAIDB (Pathogenicity Island Database) (Yoon et al., 2007), functionality using UniProt (Universal Protein Resource) (Magrane and UniProt, 2011), and cellular localization using CELLO (subCELlular LOcalization predictor) (Yu et al., 2004). In addition, we merged the final set of 10 selected non-host homologous, essential and global druggable proteins with results obtained through experiments locally performed in our laboratories (Jamal et al., unpublished data) resulting in three common targets, which we selected as candidates.

Protein–Protein Interaction Network

In biological systems, proteins work in a homogenous environment rather than individual, hence it is important to study protein–protein interactions (PPIs) for *C. diphtheriae* metabolism. The identified drug targets were evaluated to study their potential biological, functional and metabolic roles for proteomic interactions. The selected drug targets were used to develop intra-species protein–protein interactome using STRING (Search Tool for the Retrieval of Interacting

[4]http://web.expasy.org/protparam/

Genes/Proteins) database (Szklarczyk et al., 2015). STRING is an online network analyses tool that provides essential information regarding interactions of the desired proteins.

RESULTS AND DISCUSSION

Prediction of Structural Homology Based Models

The complete modelome of the reference strain NCTC13129 was computed; consisting of 13,763 ORF, with 1,253 (∼9%) resulting models. Taking the original models of reference strain NCTC13129 as templates, we then generated 438 conserved models in the 12 remaining strains using the MODELLER software (Sali and Blundell, 1993). Afterward, the target-template alignments have been computed using a BLAST *E*-value cut-off of 10^{-6} in order to build the model structures using the MODELLER software (Sali and Blundell, 1993; Webb and Sali, 2016). For each target-template alignment, ten different target models were built, and their quality measures have been assessed using GA341 (Melo and Feytmans, 1998; Melo and Sali, 2007) and QMEAN (Benkert et al., 2009), keeping models with GA341 reliability scores $\geq$ 0.7 (Melo et al., 2002), leading to a final set of 401 protein models. All these proteins are tabulated in **Supplementary Table S1**.

Pocketome Druggability and Active Site Residues of *C. diphtheriae*

The list of 401 targets protein drastically reduced to 137 after druggability analyses using the aforementioned fpocket and the recently developed DS programs. A summary of only highly druggable (HD) targets with drug score remained $\geq$ 0.7 were considered as global druggable. The calculated structural druggability scores are given in **Supplementary Table S2**.

In **Figure 2**, a comparison of calculated druggability score distribution across all structures of *C. diphtheriae* reference and other strains is shown. Although the distribution has a small shift to higher values, we used the same bounds to define the sets of druggable proteins (**Figure 2**). A protein target, which remained druggable in all strains, was classified as Globally Druggable (GD).

Non-host Homology, Essentiality and Core-Modelomics Analyses

As aforementioned, the list of 137 global druggable proteins (**Supplementary Table S2**) was computed to the corresponding human host proteome that resulted in the identification of a very small set of only 10 non-host homologous proteins; remaining 127 as host-homologous. The non-host homologous targets were selected following a very stringent criterion, i.e., no identity, no hits. This list of final 10 essential and non-host homologous targets in *C. diphtheriae* is given in **Table 2**. We further report the involvement of these putative targets in metabolic pathways, biological processes, cellular localization, molecular weights and most importantly their potential role as virulence factors. Out of 10 targets, 7 targets were found as pathogen virulence factors using the PAIDB database based on homology method. Further, we predicted the subcellular localization of these final target proteins using support vector machines, a methodology that is based on n-peptide composition

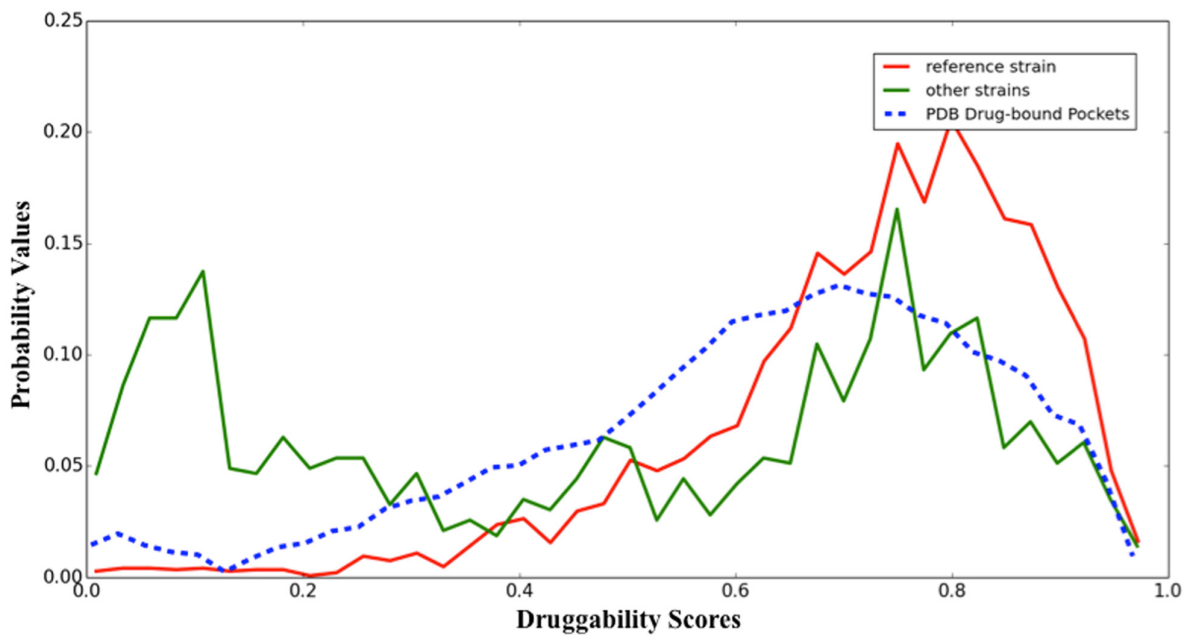

FIGURE 2 | Histogram of druggability score along the *Corynebacterium diphtheriae* reference/other strains. Representation of all ligand-bound structures in the PDB (blue pointed line), all modeled structures of *C. diphtheriae* reference strain (red line, 1,253 models), and all modeled structures of the *C. diphtheriae* non-reference genome strains (green line, 401 core proteins in 12 non-reference strains).

TABLE 2 | List of global druggable, non-host homologous and essential putative targets, their functional annotation together with other information.

S. No	Gene/protein codes	Official full name	Mol. Wt[a] (KDa)	Functions[b]	Cell locality[c]	Pathways[d]	Virulence[e]
1	NP_939496.1 coaD	Phosphopantetheine adenylyltransferase	17.305	**MF:** ATP binding, pantetheine-phosphate adenylyltransferase activity. **BP:** coenzyme A biosynthetic process.	Cytoplasm	Pantothenate and CoA biosynthesis, metabolic pathways	Yes
2	NP_939612.1 **hisE**	Phosphoribosyl-ATP pyrophosphatase	9.877	**MF:** RNA binding, phosphoribosyl-ATP diphosphatase activity. **BP:** histidine biosynthetic process	Cytoplasm	Biosynthesis of amino acids	Yes
3	NP_938944.1 DIP0568	Alanine racemase	41.235	**MF:** Catalyzes the interconversion of L-alanine and D-alanine. May also act on other amino acids. **BP:** D-alanine biosynthetic process	Cytoplasm	D-Alanine metabolism, Metabolic pathways, Vancomycin resistance	Yes
4	NP_939875.1 rimM	16S rRNA-processing protein RimM	18.067	**MF:** ribosome binding. **BP:** ribosomal small subunit biogenesis, rRNA processing	Cytoplasm	Ribosome biogenesis	No
5	NP_939302.1 **glpX**	Fructose 1,6-bisphosphatase II	35.589	**MF:** fructose 1,6-bisphosphate 1-phosphatase activity, metal ion binding. **BP:** gluconeogenesis, glycerol metabolic process	Cytoplasm	Carbohydrate metabolism	No
6	NP_940458.1 dcd	Deoxycytidine triphosphate deaminase	20.494	**MF:** dCTP deaminase activity. **BP:** dUMP biosynthetic process, dUTP biosynthetic process, pyrimidine ribonucleotide biosynthetic process	Cytoplasm	Metabolism	Yes
7	NP_938900.1 **rpsH**	30S ribosomal protein S8	14.292	**MF:** rRNA binding, structural constituent of ribosome. **BP:** translation	Extracellular/ Cytoplasm	Unknown	No
8	NP_938817.1 rplJ	50S ribosomal protein L10	17.946	**MF:** large ribosomal subunit rRNA binding, structural constituent of ribosome. **BP:** ribosome biogenesis, translation	Cytoplasm	Genetic information processing, translation	Yes
9	NP_940678.1 rsmG	16S rRNA methyltransferase GidB	24.425	**MF:** rRNA (guanine-N7-)-methyltransferase activity. **BP:** rRNA processing	Cytoplasm	Methyltransferases	Yes
10	NP_938439.1 thiE	Thiamine-phosphate synthase	23.441	**MF:** magnesium ion binding, thiamine-phosphate diphosphorylase activity. **BP:** thiamine biosynthetic process, thiamine diphosphate biosynthetic process	Cytoplasm	Thiamine metabolism, metabolic pathways	Yes

The underlined are the three high-confidence targets (see text).
[a]*Molecular weight via ProtParam tool (http://web.expasy.org/protparam/).*
[b]*Molecular function (MF) and biological process (BP) for each target protein via UniProt.*
[c]*Cellular localization of pathogen targets via CELLO.*
[d]*KEGG for finding the role of these targets in different cellular pathways.*
[e]*PAIDB for analyzing if the putative targets are involved in the pathogen's virulence.*

of proteins, adapted in a related work by Yu et al. (2004), we obtained three high-confidential candidates, which are; hisE, glpX, and rpsH. Interestingly, these three high-confidential candidates were identified as essential and non-host homologous targets in our previous work by Jamal et al. (2017a). These proteins were subjected to molecular docking analysis against four different ligand libraries and a set of some potent molecules were suggested for active inhibition of these proteins (Jamal et al., 2017a).

hisE (Phosphoribosyl-ATP Pyrophosphatase)

hisE is the second enzyme in histidine-biosynthetic pathway hydrolysing irreversibly phosphoribosyl-ATP to phosphoribosyl-AMP and pyrophosphate. The protein is encoded by the *hisE* gene, fused to *hisI* in many bacteria, fungi and plants but is present as a separate gene in some bacteria and archaea. Since it is seen in *in vitro* experiments that *hisE* is essential for microorganism growth, we assume it a potential drug target in *C. diphtheriae*. It is also reported as a drug candidate for tuberculosis (Javid-Majd et al., 2008).

glpX (Fructose 1,6-Bisphosphatase II)

It is one of the main enzyme for gluconeogenesis that catalyses the hydrolysis of fructose 1,6-bisphosphate to form fructose 6-phosphate and orthophosphate. In glycolysis, phosphofructokinase catalysis the reverse reaction, and the product, fructose 6-phosphate, are important precursors in various biosynthetic pathways (Horecker et al., 1975). Gluconeogenesis is an important metabolic pathway in all organisms and plays a key role by allowing the cells to synthesize glucose from non-carbohydrate precursors, such as glycerol, organic acids and amino acids. FBPases are members of lithium sensitive phosphatases a large superfamily which includes three families of inositol phosphatases and FBPases (phosphoesterase clan CL0171, AA sequences 3167 from Pfam data base). They are already reported as target for the treatment of non-insulin dependent diabetes and development of new drugs (Wright et al., 2002; Sassetti and Rubin, 2003).

rpsH (30S Ribosomal Protein S8)

The protein rpsH is one of the key RNA-binding protein having a central position within the small ribosomal subunit. It interacts widely with 16S rRNA and is fundamental for the correct folding of the central domain of the ribosomal rRNA. Furthermore, this protein regulates the synthesis of various other ribosomal proteins by binding to mRNA. In the two RNA molecules, it binds exactly to very similar sites. rpsH has a medium size and recently it has been discovered that rpsH play vital role as a significant primary RNA-binding protein in the 30S subunit. Mutations in S8 within the protein are shown to result in defective ribosomal assembly. The S8-binding site within 16S rRNA in *Escherichia coli*, has been investigated independently by a number of techniques including protein crosslinking, nuclease protection, hydroxyl-radical foot printing, RNA–RNA modification and chemical probing. The 30S ribosomal protein S8 is also one of the principal regulatory elements that control ribosomal protein synthesis by the translational feedback inhibition mechanism discovered by Yates et al. (1980). It regulates the expression of spc operon that encodes the 10 ribosomal proteins L5, L6, L14, L15, L18, L24, L30, S5, S8, and S14, respectively (Davies et al., 1996).

Protein–Protein Interaction Network for Proposed Targets

Protein–protein interaction of target proteins with each other have been constructed showing two proteins, rplJ (VN94_02905/50S ribosomal protein L10) and rpsH (30S ribosomal protein S8) to be interacting directly based on highest confidence score 0.9. The confidence score is the approximate probability that a predicted link exists between two enzymes in the same metabolic map in the KEGG database. The evidences for this interaction are gene fusion, co-occurrence, co-expression, experimental and databases (**Supplementary Figure S1**).

CONCLUSION

We performed a comprehensive *in silico* study of the druggability scores on all sequenced genomes of *C. diphtheria* resulting in a list of intra-strain highly druggable pockets of 10 ORFs non-homologous in human hosts. Previously, we have implemented a similar approach using other bioinformatics tools for the identification of putative therapeutic targets in *C. diphtheriae* that relied primarily on the modelome construction followed by filtering the obtained data for conserved targets (Jamal et al., 2017a). In that work, a final set of eight essential and non-host homologs targets were subjected to virtual screening using different compound libraries but lacked a detailed overview of the druggable protein pockets of the selected targets. Here, we further extrapolated our work to the druggable pocketome at species level and then at the end compared our final data set obtained in this work with the aforementioned published data. The comparison showed that any of the two approaches for putative targets identification in pathogenic microorganisms might provide an easy-to-handle protocol in future drug discovery projects. Our pipeline is expandable and can be applied to other bacterial species as well. In the future, we will work on Cytoscape plugins to allow for mapping essential druggable non-homologous genes to biological networks interactively for follow-up systems biology investigation (Baumbach and Apeltsin, 2008). We believe that our approach has the potential to aid in designing drugs and/or vaccines, and in developing protein inhibitors as well as discovering new lead compounds.

AUTHOR CONTRIBUTIONS

SH, SJ, LR, ST planned the whole work. SH, SJ, LR, ST, and PdC analyzed the data. SH, SJ, LR, ST drafted the manuscript. SH, HF, VA, AS, B, DB, PG, JB, RR, and AT reviewed and analyzed the manuscript. RS, PdC, SJ, and ST performed the literature review and formatting the tables/figures. AU, SK, JB, RR, and JA provided useful comments/suggestions for the improvement of the manuscript.

ACKNOWLEDGMENTS

This work is a part of a mutual collaboration among the two leading research/academic institutions of Brazil and Argentine, the Laboratory of Cellular and Molecular Genetics, PG Program in Bioinformatics, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, and the Structural Bioinformatics Group, Institute of Physical Chemistry of Materials, Environment and Energy, University of Buenos Aires, Argentine, respectively. We acknowledge the support of all co-authors who have worked and collaborated in accomplishing this project. We are grateful to our Argentinian collaborators, especially Prof. Dr. Adrian Gustavo Turjanski and Leandro G. Radusky for their collaboration, timely assistance in theoretical/biological discussion about the project and manuscript preparation. We acknowledge the Brazilian funding agencies CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil). Syed Babar Jamal acknowledges the "TWAS-CNPq Postgraduate Fellowship Program" for granting a fellowship for doctoral studies.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00044/full#supplementary-material

FIGURE S1 | Protein–protein interactome for the identified common conserved targets. Ribosomal pathway protein (VN94_02915 rplJ, rpsH) was interacting with each other either directly or indirectly. The lines color indicates evidences of interactions that were predicted or experimentally validated.

TABLE S1 | Tabulated are the 401 conserved target proteins with appropraite 3D structures that were used for further druggability etc. analyse.

TABLE S2 | Summary of core Global Druggable (GD) targets along the 13 strains of C. diphtheriae with predicted druggability scores (137 GD targets).

REFERENCES

Adler, N. R., Mahony, A., and Friedman, N. D. (2013). Diphtheria: forgotten, but not gone. *Intern. Med. J.* 43, 206–210. doi: 10.1111/imj.12049

Alcaraz, N., Friedrich, T., Kotzing, T., Krohmer, A., Muller, J., Pauling, J., et al. (2012). Efficient key pathway mining: combining networks and OMICS data. *Integr. Biol.* 4, 756–764. doi: 10.1039/c2ib00133k

Asif, S. M., Asad, A., Faizan, A., Anjali, M. S., Arvind, A., Neelesh, K., et al. (2009). Dataset of potential targets for *Mycobacterium tuberculosis* H37Rv through comparative genome analysis. *Bioinformation* 4, 245–248. doi: 10.6026/97320630004245

Barh, D., Gupta, K., Jain, N., Khatri, G., Leon-Sicairos, N., Canizalez-Roman, A., et al. (2013). Conserved host-pathogen PPIs. Globally conserved inter-species bacterial PPIs based conserved host-pathogen interactome derived novel target in *C. pseudotuberculosis*, *C. diphtheriae*, *M. tuberculosis*, *C. ulcerans*, *Y. pestis*, and *E. coli* targeted by Piper betel compounds. *Integr. Biol.* 5, 495–509. doi: 10.1039/c2ib20206a

Barh, D., Jain, N., Tiwari, S., Parida, B. P., D'Afonseca, V., Li, L., et al. (2011). 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.* 78, 73–84. doi: 10.1111/j.1747-0285.2011.01118.x

Barh, D., and Kumar, A. (2009). In silico identification of candidate drug and vaccine targets from various pathways in *Neisseria gonorrhoeae*. *In Silico Biol.* 9, 225–231.

Barraud, O., Badell, E., Denis, F., Guiso, N., and Ploy, M. C. (2011). Antimicrobial drug resistance in *Corynebacterium diphtheriae* mitis. *Emerg. Infect. Dis.* 17, 2078–2080. doi: 10.3201/eid1711.110282

Baumbach, J. (2010). On the power and limits of evolutionary conservation–unraveling bacterial gene regulatory networks. *Nucleic Acids Res.* 38, 7877–7884. doi: 10.1093/nar/gkq699

Baumbach, J., and Apeltsin, L. (2008). Linking cytoscape and the corynebacterial reference database CoryneRegNet. *BMC Genomics* 9:184. doi: 10.1186/1471-2164-9-184

Benkert, P., Kunzli, M., and Schwede, T. (2009). QMEAN server for protein model quality estimation. *Nucleic Acids Res.* 37, W510–W514. doi: 10.1093/nar/gkp322

Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the worldwide protein data bank. *Nat. Struct. Biol.* 10:980. doi: 10.1038/nsb1203-980

Brooks, G. F., Carroll, K. C., Butel, J. S., Morse, S. A., and Mietzneron, T. A. (eds) (2010). "Aerobic non–spore-forming gram-positive bacilli: *Corynebacterium*, *Listeria*, *Erysipelothrix*, *Actinomycetes*, and related pathogens," in *Jawetz, Melnick, & Adelberg's Medical Microbiology*, 25 Edn, (New York, NY: McGraw-Hill).

Cerdeno-Tarraga, A. M., Efstratiou, A., Dover, L. G., Holden, M. T., Pallen, M., Bentley, S. D., et al. (2003). The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. *Nucleic Acids Res.* 31, 6516–6523. doi: 10.1093/nar/gkg874

Chong, C. E., Lim, B. S., Nathan, S., and Mohamed, R. (2006). In silico analysis of *Burkholderia pseudomallei* genome sequence for potential drug targets. *In Silico Biol.* 6, 341–346.

Davies, C., Ramakrishnan, V., and White, S. W. (1996). Structural evidence for specific S8-RNA and S8-protein interactions within the 30S ribosomal subunit: ribosomal protein S8 from *Bacillus stearothermophilus* at 1.9 A resolution. *Structure* 4, 1093–1104. doi: 10.1016/S0969-2126(96)00115-3

Dutta, A., Singh, S. K., Ghosh, P., Mukherjee, R., Mitter, S., and Bandyopadhyay, D. (2006). In silico identification of potential therapeutic targets in the human pathogen *Helicobacter pylori*. *In Silico Biol.* 6, 43–47.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344

Furnham, N., Holliday, G. L., de Beer, T. A., Jacobsen, J. O., Pearson, W. R., and Thornton, J. M. (2014). The catalytic site atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res.* 42, D485–D489. doi: 10.1093/nar/gkt1243

Gerald, L., Mandell, J. E. B., and Raphael, D. (2009). *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*. Philadelphia, PA: ELSEVIER.

Hadfield, T. L., McEvoy, P., Polotsky, Y., Tzinserling, V. A., and Yakovlev, A. A. (2000). The pathology of diphtheria. *J. Infect. Dis.* 181(Suppl. 1), S116–S120. doi: 10.1086/315551

Hart, P. E., Lee, P. Y., Macallan, D. C., and Wansbrough-Jones, M. H. (1996). Cutaneous and pharyngeal diphtheria imported from the Indian subcontinent. *Postgrad. Med. J.* 72, 619–620. doi: 10.1136/pgmj.72.852.619

Hassan, S. S., Tiwari, S., Guimaraes, L. C., Jamal, S. B., Folador, E., Sharma, N. B., et al. (2014). Proteome scale comparative modeling for conserved drug and vaccine targets identification in *Corynebacterium pseudotuberculosis*. *BMC Genomics* 15(Suppl. 7):S3. doi: 10.1186/1471-2164-15-S7-S3

Hodes, H. L. (1979). Diphtheria. *Pediatr. Clin. North Am.* 26, 445–459. doi: 10.1016/S0031-3955(16)33716-6

Horecker, B. L., Melloni, E., and Pontremoli, S. (1975). Fructose 1,6-bisphosphatase: properties of the neutral enzyme and its modification by proteolytic enzymes. *Adv. Enzymol. Relat. Areas Mol. Biol.* 42, 193–226.

Jamal, S. B., Hassan, S. S., Tiwari, S., Viana, M. V., Benevides, L. J., Ullah, A., et al. (2017a). An integrative in-silico approach for therapeutic target identification in the human pathogen *Corynebacterium diphtheriae*. *PLOS ONE* 12:e0186401. doi: 10.1371/journal.pone.0186401

Jamal, S. B., Tiwari, S., Silva, A., and Azevedo, V. (2017b). Pathogenesis of *Corynebacterium diphtheriae* and available vaccines; an overview. *Glob. J. Infect. Dis. Clin. Res.* 3, 20–24.

Javid-Majd, F., Yang, D., Ioerger, T. R., and Sacchettini, J. C. (2008). The 1.25 A resolution structure of phosphoribosyl-ATP pyrophosphohydrolase from *Mycobacterium tuberculosis*. *Acta Crystallogr. D Biol. Crystallogr.* 64(Pt 6), 627–635. doi: 10.1107/S0907444908007105

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kinnings, S. L., Xie, L., Fung, K. H., Jackson, R. M., Xie, L., and Bourne, P. E. (2010). The Mycobacterium tuberculosis drugome and its polypharmacological implications. *PLOS Comput. Biol.* 6:e1000976. doi: 10.1371/journal.pcbi.1000976

Magrane, M., and UniProt, C. (2011). UniProt Knowledgebase: a hub of integrated protein data. *Database* 2011:bar009. doi: 10.1093/database/bar009

Melo, F., and Feytmans, E. (1998). Assessing protein structures with a non-local atomic interaction energy. *J. Mol. Biol.* 277, 1141–1152. doi: 10.1006/jmbi.1998.1665

Melo, F., and Sali, A. (2007). Fold assessment for comparative protein structure modeling. *Protein Sci.* 16, 2412–2426. doi: 10.1110/ps.072895107

Melo, F., Sanchez, R., and Sali, A. (2002). Statistical potentials for fold assessment. *Protein Sci.* 11, 430–448. doi: 10.1002/pro.110430

Perumal, D., Lim, C. S., Sakharkar, K. R., and Sakharkar, M. K. (2007). Differential genome analyses of metabolic enzymes in *Pseudomonas aeruginosa* for drug target identification. *In Silico Biol.* 7, 453–465.

Pizza, M., Scarlato, V., Masignani, V., Giuliani, M. M., Arico, B., Comanducci, M., et al. (2000). Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287, 1816–1820. doi: 10.1126/science.287.5459.1816

Radusky, L., Defelipe, L. A., Lanzarotti, E., Luque, J., Barril, X., Marti, M. A., et al. (2014). TuberQ: a *Mycobacterium tuberculosis* protein druggability database. *Database* 2014, bau035. doi: 10.1093/database/bau035

Radusky, L. G., Hassan, S., Lanzarotti, E., Tiwari, S., Jamal, S., Ali, J., et al. (2015). An integrated structural proteomics approach along the druggable genome of *Corynebacterium pseudotuberculosis* species for putative druggable targets. *BMC Genomics* 16(Suppl. 5):S9. doi: 10.1186/1471-2164-16-S5-S9

Rathi, B., Sarangi, A. N., and Trivedi, N. (2009). Genome subtraction for novel target definition in *Salmonella typhi*. *Bioinformation* 4, 143–150. doi: 10.6026/97320630004143

Rottger, R., Kalaghatgi, P., Sun, P., Soares Sde, C., Azevedo, V., Wittkop, T., et al. (2013). Density parameter estimation for finding clusters of homologous proteins–tracing actinobacterial pathogenicity lifestyles. *Bioinformatics* 29, 215–222. doi: 10.1093/bioinformatics/bts653

Sakharkar, K. R., Sakharkar, M. K., and Chow, V. T. (2004). A novel genomics approach for the identification of drug targets in pathogens, with special reference to *Pseudomonas aeruginosa*. *In Silico Biol.* 4, 355–360.

Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815. doi: 10.1006/jmbi.1993.1626

Sassetti, C. M., and Rubin, E. J. (2003). Genetic requirements for mycobacterial survival during infection. *Proc. Natl. Acad. Sci. U.S.A.* 100, 12989–12994. doi: 10.1073/pnas.2134250100

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. doi: 10.1093/nar/gku1003

Trost, E., Blom, J., Soares Sde, C., Huang, I. H., Al-Dilaimi, A., Schroder, J., et al. (2012). 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.* 194, 3199–3215. doi: 10.1128/JB.00183-12

Velec, H. F., Gohlke, H., and Klebe, G. (2005). DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.* 48, 6296–6303. doi: 10.1021/jm050436v

Webb, B., and Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* 86, 2.9.1–2.9.37. doi: 10.1002/cpps.20

Wright, S. W., Carlo, A. A., Carty, M. D., Danley, D. E., Hageman, D. L., Karam, G. A., et al. (2002). Anilinoquinazoline inhibitors of fructose 1,6-bisphosphatase bind at a novel allosteric site: synthesis, in vitro characterization, and X-ray crystallography. *J. Med. Chem.* 45, 3865–3877. doi: 10.1021/jm010496a

Yates, J. L., Arfsten, A. E., and Nomura, M. (1980). *In vitro* expression of *Escherichia coli* ribosomal protein genes: autogenous inhibition of translation. *Proc. Natl. Acad. Sci. U.S.A.* 77, 1837–1841. doi: 10.1073/pnas.77.4.1837

Yoon, S. H., Park, Y. K., Lee, S., Choi, D., Oh, T. K., Hur, C. G., et al. (2007). Towards pathogenomics: a web-based resource for pathogenicity islands. *Nucleic Acids Res.* 35, D395–D400. doi: 10.1093/nar/gkl790

Yu, C. S., Lin, C. J., and Hwang, J. K. (2004). Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.* 13, 1402–1406. doi: 10.1110/ps.03479604

Zhang, R., Ou, H. Y., and Zhang, C. T. (2004). DEG: a database of essential genes. *Nucleic Acids Res.* 32, D271–D272. doi: 10.1093/nar/gkh024

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

III.3.2. Conclusion, Chapter 3

- We performed a comprehensive *in silico* study of the druggability scores on all sequenced genomes of *C. diphtheria* resulting in a list of intra-strain highly druggable pockets of 10 ORFs non-homologous in human hosts.
- Here, we further extrapolated our work to the druggable pocketome at species level.
- We identified three targets; hisE-phosphoribosyl-ATP pyrophosphatase, glpX-fructose 1,6-bisphosphatase II, and rpsH 30S ribosomal protein S8 in this study that were common with our targets identified by Jamal et al., 2017.
- Our pipeline is expandable and can be applied to other bacterial species as well.
- In the future, we will work on Cytoscape plugins to allow for mapping essential druggable non-homologous genes to biological networks interactively for follow-up systems biology investigation (Baumbach and Apeltsin, 2008).
- We believe that our approach has the potential to aid in designing drugs and/or vaccines, and in developing protein inhibitors as well as discovering new lead compounds.
- Hence, we are up to extend our study for experimental validation.

Chapter 4

III.4.1. General Conclusion

- Diphtheria is somehow forgotten disease but still for past few year cases had been reported throughout the world.
- The occurrence of new cases has concerned the scientific community to rethink over the currently available treatments.
- The availability of genomic data provides means to better understanding the molecular and genetic basis of virulence of this bacterium, enabling a detailed investigation of *C. diphtheriae*. In the long run, providing a new gate way for development and/or improvement of potent vaccine.
- Then, this genomic information was used with the aim of determining the conserved predicted proteome of 13 strains of *C. diphtheriae*, together with the 3D structural information.
- The measurements for target selection in *C. diphtheriae* was kept strict, resulting in a small set of prioritized putative drug/vaccine targets, of which eight are essential and non-host homologous, and 15 are essential and host homologous proteins.
- The analysis of protein models was carried out very carefully, based on good templates which provided high-quality templates, lessening the risks. The data presented here can effectively contribute to future research for the development of novel antibiotics and vaccines.
- After a detailed structural comparison between host and pathogen proteins, we suggest that eight non-host homologous proteins could be considered for antimicrobial chemotherapy owing to further experimental validations.
- Furthermore, performed a comprehensive *in silico* study of the druggability scores on all sequenced genomes of *C. diphtheria* resulting in a list of intra-strain highly druggable pockets of 10 ORFs non-homologous in human hosts.
- Here, we further extrapolated our work to the druggable pocketome at species level.

- We identified three targets; hisE-phosphoribosyl-ATP pyrophosphatase, glpX-fructose 1,6-bisphosphatase II, and rpsH 30S ribosomal protein S8 in this study that were common with our targets identified by Jamal et al., 2017.
- Our pipeline is expandable and can be applied to other bacterial species as well.

III.4.2. Future Prospective

The future scopes of this research are:

- Identification and characterization of the *hisE, glpX* and *rpsH* operons of *C. diphtheria*;
- Cloning and construction of *C. diphtheria hisE, glpX* and *rpsH* genes mutant strains;
- Evaluation of the potential of *hisE, glpX* and *rpsH* mutant strains as live attenuated vaccine;
- Experimental validation of *in silico* identified drug molecules in *C. diphtheria*.

IV. Appendix

A. Published, Accepted and Submitted Research articles.

Genome announcements

(My contributions to these papers)

These papers make part of contributions and collaboration work made by our groups showing an extensive involvement of ongoing genomics projects nationally and internationally. I have actively participated in these genomics project and got an opportunity to enhance my knowledge of learning assembly and annotation of genomes. Working with genomics projects, I get familiar with different Next Generation Sequencing platforms and its application in various aspects of biological analysis.

My contributions in these articles were assembly and annotation work led by manuscript drafting.

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Whole-Genome Sequence of *Corynebacterium auriscanis* Strain CIP 106629 Isolated from a Dog with Bilateral Otitis from the United Kingdom

Sandeep Tiwari,[a] Syed Babar Jamal,[a] Leticia Castro Oliveira,[a] Dominique Clermont,[b] Chantal Bizet,[b] Diego Mariano,[a] Paulo Vinicius Sanches Daltro de Carvalho,[a] Flavia Souza,[a] Felipe Luiz Pereira,[c] Siomar de Castro Soares,[d] Luis C. Guimarães,[e] Fernanda Dorella,[c] Alex Carvalho,[c] Carlos Leal,[c] Debmalya Barh,[f] Henrique Figueiredo,[c] Syed Shah Hassan,[a] Vasco Azevedo,[a] Artur Silva[e]

Institute of Biologic Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil[a]; Institut Pasteur, Unité de Prévention et Thérapies Moléculaires des Maladies Humaines, National Centre of Reference of Toxigenic Corynebacteria, Paris, France[b]; AQUACEN, National Reference Laboratory for Aquatic Animal Diseases, Ministry of Fisheries and Aquaculture, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil[c]; Department of Immunology, Microbiology and Parasitology, Institute of Biological Sciences and Natural Sciences, Federal University of Triângulo Mineiro (UFTM), Uberaba, Minas Gerais, Brazil[d]; Institute of Biological Sciences, Federal University of Pará, Belém, Pará, Brazil[e]; Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology, Nonakuri, Purba Medinipur, West Bengal, India[f]

In this work, we describe a set of features of *Corynebacterium auriscanis* CIP 106629 and details of the draft genome sequence and annotation. The genome comprises a 2.5-Mbp-long single circular genome with 1,797 protein-coding genes, 5 rRNA, 50 tRNA, and 403 pseudogenes, with a G+C content of 58.50%.

Received 15 June 2016 **Accepted** 17 June 2016 **Published** 11 August 2016

Citation Tiwari S, Jamal SB, Oliveira LC, Clermont D, Bizet C, Mariano D, de Carvalho PVSD, Souza F, Pereira FL, de Castro Soares S, Guimarães LC, Dorella F, Carvalho A, Leal C, Barh D, Figueiredo H, Hassan SS, Azevedo V, Silva A. 2016. Whole-genome sequence of *Corynebacterium auriscanis* strain CIP 106629 isolated from a dog with bilateral otitis from the United Kingdom. Genome Announc 4(4):e00683-16. doi:10.1128/genomeA.00683-16.

Address correspondence to Artur Silva, asilva@ufpa.br.

The genus *Corynebacterium* contains many species that are pathogenic to humans and animals (1). Collins et al. (2) reported six coryneform-like isolates that originated from clinical specimens of bilateral otitis in dogs. They gave the name *Corynebacterium auriscanis* to one of the six coryneform-like isolates (3). Until 2008, *Corynebacterium auriscanis* was the only recognized animal pathogen, but a case reported in healthy human patient followed by a dog bite confirmed that this organism is a potential human pathogen and possibly a zoonotic carrier (3).

Here, we present the first draft genome sequence of *Corynebacterium auriscanis* CIP 106629 isolated from the clinical specimens in the United Kingdom. This bacterium is a Gram-positive, non-spore-forming, nonmotile, nonlipophilic, and typically club-shaped rod with appearance as single, in pairs, or in cluster cells. It is nonfermentative, nitrate reduction negative, and grows under aerobic conditions (4, 5). We determined the nucleotide sequence of the *C. auriscanis* CIP 106629 genome, isolated from a dog's ear infection. Sequencing was performed by the National Reference Laboratory for Aquatic Animal Diseases, Ministry of Fisheries and Aquaculture, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil. Assembly and annotation were performed by the Laboratory of Cellular and Molecular Genetics, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, and the Center of Genomics and Systems Biology, Federal University of Pará, Belém, Pará, Brazil. Automatic annotation was performed via the NCBI Prokaryotic Genome Annotation Pipeline (PGAP version 2.8) (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/) (6).

The platform used for sequencing was the Ion Torrent Personal Genome Machine (PGM) system (Thermo Fisher), using a 200-bp fragment sequencing kit, according to the manufacturer's protocols. The quality of the raw data was analyzed using the Web tool FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). The reads with good quality were *de novo* assembled using the Mira version 3.9 software (7). The assembly produced 33 contigs, having a coverage of 78×, with N_{50} value for contig length of 241,167 bp.

The protein-coding genes (open reading frames [ORFs]) of the draft genome comprise a 2.5-Mbp-long single circular genome with 1,797 protein-coding genes, 5 rRNA, 50 tRNA, and 403 pseudogenes, with a G+C content of 58.50%.

Nucleotide sequence accession numbers. The *C. auriscanis* CIP 106629 whole-genome shotgun (WGS) project has the project accession no. JRVJ00000000. The version described in this paper is the first version and consists of sequences JRVJ00000000.1 to JRVJ00000000.33.

ACKNOWLEDGMENTS

We acknowledge the support of all team members & financing agencies. Sandeep Tiwari acknowledges the TWAS-CNPq Postgraduate Fellowship Programme for granting a fellowship for doctoral studies. This work was partially carried out by the Rede Paraense de Genômica e Proteômica supported by the Fundação de Amparo à Pesquisa do Estado do Pará (FAPESPA), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil, and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Brazil.

Almeida et al. Standards in Genomic Sciences (2016) 11:29
DOI 10.1186/s40793-016-0149-7

Standards in Genomic Sciences

SHORT GENOME REPORT Open Access

The genome anatomy of *Corynebacterium pseudotuberculosis* VD57 a highly virulent strain causing *Caseous lymphadenitis*

Sintia Almeida[1*], Sandeep Tiwari[1], Diego Mariano[1], Flávia Souza[1], Syed Babar Jamal[1], Nilson Coimbra[3], Roberto Tadeu Raittz[3], Fernanda Alves Dorella[2], Alex Fiorine de Carvalho[2], Felipe Luiz Pereira[2], Siomar de Castro Soares[2], Carlos Augusto Gomes Leal[2], Debmalya Barh[6], Preetam Ghosh[7], Henrique Figueiredo[2], Lília Ferreira Moura-Costa[4], Ricardo Wagner Portela[4], Roberto Meyer[4], Artur Silva[5] and Vasco Azevedo[1*]

Abstract

Corynebacterium pseudotuberculosis strain VD57 (Cp_VD57), a highly virulent, nonmotile, non-sporulating, and a mesophilic bacterium, was isolated from a goat's granulomatous lesion in the municipality of Juazeiro, Bahia State, Brazil. Here, we describe a set of features of the strain, together with the details of its complete genome sequence and annotation. The genome comprises of a 2.5 Mbp long, single circular genome with 2,101 protein-coding genes, 12 rRNA, 49 tRNA and 47 pseudogenes and a G + C content of 52.85 %. Genetic variation was detected in Cp_VD57 using *C. pseudotuberculosis* strain 1002 as reference, wherein small genomic insertions and deletions were identified. The comparative analysis of the genome sequence provides means to better understand the host pathogen interactions of this strain and can also help us to understand the molecular and genetic basis of virulence of this bacterium.

Keywords: Biovar *ovis*, Gram-positive pathogen, Caseous lymphadenitis, *Corynebacterium pseudotuberculosis*, Goat, Genome sequencing, Ion Torrent PGM

Introduction

Corynebacterium pseudotuberculosis is the etiologic agent of caseous lymphadenitis in sheep and goats, the organism has also been associated with mastitis [1–3] and can cause ulcerative lymphangitis in horses and cattle [4]. CL is a chronic disease that is characterized by the formation of granulomas in lymph nodes and internal organs, as a response of the host's immune system against this bacterium that resists to the bactericidal action of phagocytic cells [3].

CL is considered as one of the economically important diseases of small ruminants with losses attributed to reduced wool and hide yields, carcass condemnation, morbidity and rarely mortality [5, 6]. The prevalence of CL has been observed worldwide, including South Africa, Brazil, the USA, Canada, Australia, New Zealand, United Kingdom and Egypt [7].

The pangenome analysis of 15 strains of the pathogen was completed recently [8]. However, as *C. pseudotuberculosis* is a relatively clonal organism [9–13], the identification of the virulence mechanisms or nucleotide modifications responsible for making a strain more virulent than another, have not yet been identified.

Sequencing of new genomes coupled with a deeper comparative analysis between the genomes and associating such analyses with the host pathogen interactions can help us understand and identify the differences between genomes and virulence factors. In this context, the present study reports the sequence the genome of the highly virulent strain VD 57 and to understand its virulence factors.

Organism information

Classification and features

C. pseudotuberculosis is a Gram-positive bacteria and belong to a CMNR *(Corynebacterium, Mycobacterium, Nocardia* and *Rhodococcus*) group that shares characteristics including an outer lipid layer, mycolic acids in the cell wall along with its derivatives including phospholipids

* Correspondence: sintiaalmeida@ufmg.br; vasco@icb.ufmg.br
[1]Institute of Biologic Sciences, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil
Full list of author information is available at the end of the article

Comparative genomics, *in-silico* Drug target identifications, Homology molding, and docking

(My contributions to these papers)

Genomics is a broad discipline, which may be divided into three main areas; Structural Genomics, Functional Genomics & Comparative genomics. **Structural genomics** deals with the physical nature of genome. Its primary objective is to determine and analyze the genomic DNA sequence. **Functional genomics** is concerned with the way genome functions. That is, it examines the transcript produced by genome and the collection of proteins they encode. The third and relatively new area of genomics are **comparative genomics** in which genome from different organisms are compared to look for significant similarity and difference. This helps in the identification of the essential, conserved portion of the genome and discriminate patterns in functions, regulations, and therapeutic target identifications.

After joining the LGCM as a doctorate student, I experienced working with different projects at the same time. It was good for my learning process to relate my previous experiences in the scientific field with the on-going projects in LGCM. I contributed as a structural biologist to give flow to these projects by applying protein homology modeling, and molecular docking approaches. Simultaneously, I worked with the analyzing the data and drafting the manuscript.

frontiers in Microbiology

MINI REVIEW
published: 10 October 2017
doi: 10.3389/fmicb.2017.01878

Two-Component Signal Transduction Systems of Pathogenic Bacteria As Targets for Antimicrobial Therapy: An Overview

Sandeep Tiwari[1], Syed B. Jamal[1], Syed S. Hassan[1,2], Paulo V. S. D. Carvalho[1], Sintia Almeida[1], Debmalya Barh[1,3], Preetam Ghosh[4], Artur Silva[5], Thiago L. P. Castro[1,6] and Vasco Azevedo[1]**

[1] Laboratório de Genética Celular e Molecular, Departamento de Biologia Geral, Instituto de Ciências Biológicas (ICB), Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, [2] Biochemistry Group, Department of Chemistry, Islamia College University, Peshawar, Pakistan, [3] Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology, Purba Medinipur, India, [4] Department of Computer Science, Virginia Commonwealth University, Richmond, VA, United States, [5] Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, Brazil, [6] Instituto de Ciências da Saúde, Universidade Federal da Bahia, Salvador, Brazil

The bacterial communities in a wide range of environmental niches sense and respond to numerous external stimuli for their survival. Primarily, a source they require to follow up this communication is the two-component signal transduction system (TCS), which typically comprises a sensor Histidine kinase for receiving external input signals and a response regulator that conveys a proper change in the bacterial cell physiology. For numerous reasons, TCSs have ascended as convincing targets for antibacterial drug design. Several studies have shown that TCSs are essential for the coordinated expression of virulence factors and, in some cases, for bacterial viability and growth. It has also been reported that the expression of antibiotic resistance determinants may be regulated by some TCSs. In addition, as a mode of signal transduction, phosphorylation of histidine in bacteria differs from normal serine/threonine and tyrosine phosphorylation in higher eukaryotes. Several studies have shown the molecular mechanisms by which TCSs regulate virulence and antibiotic resistance in pathogenic bacteria. In this review, we list some of the characteristics of the bacterial TCSs and their involvement in virulence and antibiotic resistance. Furthermore, this review lists and discusses inhibitors that have been reported to target TCSs in pathogenic bacteria.

Keywords: bacterial two-component signal transduction system, virulence and antibiotic resistance, inhibitors for kinases and response regulators

OPEN ACCESS

Edited by:
Miklos Fuzi,
Semmelweis University, Hungary

Reviewed by:
Yixin Shi,
Arizona State University, United States
Rodolfo García-Contreras,
National Autonomous University of Mexico, Mexico

****Correspondence:***
Vasco Azevedo
vascoariston@gmail.com
Thiago L. P. Castro
castrotlp@gmail.com

Specialty section:
This article was submitted to Antimicrobials, Resistance and Chemotherapy, a section of the journal Frontiers in Microbiology

Received: *18 March 2017*
Accepted: *14 September 2017*
Published: *10 October 2017*

Citation:
Tiwari S, Jamal SB, Hassan SS, Carvalho PVSD, Almeida S, Barh D, Ghosh P, Silva A, Castro TLP and Azevedo V (2017) Two-Component Signal Transduction Systems of Pathogenic Bacteria As Targets for Antimicrobial Therapy: An Overview. Front. Microbiol. 8:1878. doi: 10.3389/fmicb.2017.01878

INTRODUCTION

Microbes are the most versatile living organisms on the planet. They can be found in environments where plants and animals cannot survive, such as in hydrothermal vents on the ocean floor or glaciers. Bacteria have developed a number of features (e.g., signal transduction systems) that allow crosstalk between the intracellular and extracellular environments. The bacterial repertoire for sensing environmental conditions includes regulatory proteins that bind to secondary metabolites or ions, resulting in increased affinity for specific regulatory DNA sequences. Examples of well-studied regulatory proteins include the Catabolite Regulatory Protein (CRP, which senses cyclic adenosine monophosphate accumulated under low availability of glucose) and Ferric Uptake Regulator (Fur, which binds to Fe^{2+}) (Nixon et al., 1986; Hoch, 2000; Stock et al., 2000). The bacterial capability of responding to external stimuli is

frontiers in Microbiology

REVIEW
published: 12 October 2017
doi: 10.3389/fmicb.2017.01937

Insight of Genus *Corynebacterium*: Ascertaining the Role of Pathogenic and Non-pathogenic Species

Alberto Oliveira[1], *Leticia C. Oliveira*[1], *Flavia Aburjaile*[2], *Leandro Benevides*[1], *Sandeep Tiwari*[1], *Syed B. Jamal*[1], *Arthur Silva*[2], *Henrique C. P. Figueiredo*[3], *Preetam Ghosh*[4], *Ricardo W. Portela*[5]*, *Vasco A. De Carvalho Azevedo*[1]* and *Alice R. Wattam*[6]

[1] Molecular and Cellular Laboratory, General Biology Department, Federal University of Minas Gerais, Belo Horizonte, Brazil, [2] Center of Genomics and System Biology, Federal University of Pará, Belém, Brazil, [3] Aquacen, National Reference Laboratory for Aquatic Animal Diseases, Federal University of Minas Gerais, Belo Horizonte, Brazil, [4] Department of Computational Science, Virginia Commonwealth University, Richmond, VA, United States, [5] Laboratory of Immunology and Molecular Bióloga, Health Sciences Institute, Federal University of Bahiaa, Salvador, Brazil, [6] Biocomplexity Institute of Virginia Tech, Virginia Tech, Blacksburg, VA, United States

OPEN ACCESS

Edited by:
Geanncarlo Lugo-Villarino, UMR5089 Institut de Pharmacologie et de Biologie Structurale (IPBS), France

Reviewed by:
Mateus Matiuzzi Da Costa, Federal University of São Francisco Valley, Brazil
Shashank Gupta, Brown University, United States

***Correspondence:**
Ricardo W. Portela
rwportela@gmail.com
Vasco A. De Carvalho Azevedo
vascoariston@gmail.com

Specialty section:
This article was submitted to Microbial Immunology, a section of the journal Frontiers in Microbiology

Received: 08 June 2017
Accepted: 21 September 2017
Published: 12 October 2017

Citation:
Oliveira A, Oliveira LC, Aburjaile F, Benevides L, Tiwari S, Jamal SB, Silva A, Figueiredo HCP, Ghosh P, Portela RW, De Carvalho Azevedo VA and Wattam AR (2017) Insight of Genus Corynebacterium: Ascertaining the Role of Pathogenic and Non-pathogenic Species. Front. Microbiol. 8:1937. doi: 10.3389/fmicb.2017.01937

This review gathers recent information about genomic and transcriptomic studies in the *Corynebacterium* genus, exploring, for example, prediction of pathogenicity islands and stress response in different pathogenic and non-pathogenic species. In addition, is described several phylogeny studies to *Corynebacterium*, exploring since the identification of species until biological speciation in one species belonging to the genus *Corynebacterium*. Important concepts associated with virulence highlighting the role of Pld protein and Tox gene. The adhesion, characteristic of virulence factor, was described using the sortase mechanism that is associated to anchorage to the cell wall. In addition, survival inside the host cell and some diseases, were too addressed for pathogenic corynebacteria, while important biochemical pathways and biotechnological applications retain the focus of this review for non-pathogenic corynebacteria. Concluding, this review broadly explores characteristics in genus *Corynebacterium* showing to have strong relevance inside the medical, veterinary, and biotechnology field.

Keywords: *Corynebacterium*, bacterial genomics, bacterial biochemistry, pathogenesis related genes, biotechnology of microorganisms

INTRODUCTION

This review is limited to a general discussion of the genus *Corynebacterium* and its species. It includes specific information on the genes that are found in pathogenic strains compared to those that are non-pathogenicity. Various studies describe these bacteria as gram-positive (Hard, 1975; Dorella et al., 2006) but with shape, oxygen requirement, and environment of preference divergent dependent on specie, that constitute a very heterogeneous group. The genus *Corynebacterium*, which currently has more than 110 validated species, is highly diversified. It includes species that are of medical, veterinary, or biotechnological relevance (Bernard, 2012; Soares et al., 2013; Eikmanns and Blombach, 2014). In an examination of the diverse group to which these bacteria belong, this study describes the structural organization, function, and dynamism of genome. This review It also explores some important process, such as cell division, in order to improve the understanding of

International Journal of
Molecular Sciences

MDPI

Article

An In Silico Identification of Common Putative Vaccine Candidates against *Treponema pallidum*: A Reverse Vaccinology and Subtractive Genomics Based Approach

Arun Kumar Jaiswal [1,2], Sandeep Tiwari [1], Syed Babar Jamal [1], Debmalya Barh [1,3], Vasco Azevedo [1] and Siomar C. Soares [2,*]

1 Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte 31270-901, MG, Brazil; arunjaiswal1411@gmail.com (A.K.J.); sandip_sbtbi@yahoo.com (S.T.); syedbabar.jamal@gmail.com (S.B.J.); dr.barh@gmail.com (D.B.); vascoariston@gmail.com (V.A.)

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

3 Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology (IIOAB), Nonakuri, Purba Medinipur, West Bengal 721137, India

* Correspondence: siomars@gmail.com

Academic Editor: Christopher Woelk
Received: 14 November 2016; Accepted: 27 January 2017; Published: 14 February 2017

Abstract: Sexually transmitted infections (STIs) are caused by a wide variety of bacteria, viruses, and parasites that are transmitted from one person to another primarily by vaginal, anal, or oral sexual contact. Syphilis is a serious disease caused by a sexually transmitted infection. Syphilis is caused by the bacterium *Treponema pallidum* subspecies *pallidum*. *Treponema pallidum* (*T. pallidum*) is a motile, gram-negative spirochete, which can be transmitted both sexually and from mother to child, and can invade virtually any organ or structure in the human body. The current worldwide prevalence of syphilis emphasizes the need for continued preventive measures and strategies. Unfortunately, effective measures are limited. In this study, we focus on the identification of vaccine targets and putative drugs against syphilis disease using reverse vaccinology and subtractive genomics. We compared 13 strains of *T. pallidum* using *T. pallidum* Nichols as the reference genome. Using an in silicoapproach, four pathogenic islands were detected in the genome of *T. pallidum* Nichols. We identified 15 putative antigenic proteins and sixdrug targets through reverse vaccinology and subtractive genomics, respectively, which can be used as candidate therapeutic targets in the future.

Keywords: sexually transmitted infections (STIs); drug target; vaccine target

1. Introduction

Sexually transmitted infections (STIs) are triggered by a number of bacteria, viruses, and parasites that are transferred mainly by vaginal, anal, or oral sexual contact between people. Different STIs can be existent or transmitted instantaneously, and such infections can trigger other STIs [1]. The World Health Organization (WHO) has reported more than 30 different bacteria, viruses, and parasites that are responsible for disease transmission through sexual contact.

Syphilis is among the most severe sexually transmitted infections (STIs) caused by the *Treponema pallidum* subspecies *pallidum*, a motile, gram-negative spirochete bacterium [2]. The annual estimated frequency of infectious syphilis is 36 million cases and over 11 million new infections; thus, it is an important public health burden globally [3]. Furthermore, the number of cases increased 10-fold

Oliveira et al. BMC Microbiology (2016) 16:100
DOI 10.1186/s12866-016-0717-4

BMC Microbiology

RESEARCH ARTICLE Open Access

Corynebacterium pseudotuberculosis may be under anagenesis and biovar Equi forms biovar Ovis: a phylogenic inference from sequence and structural analysis

Alberto Oliveira[1], Pammella Teixeira[1], Marcela Azevedo[1], Syed Babar Jamal[1], Sandeep Tiwari[1], Sintia Almeida[1], Artur Silva[3], Debmalya Barh[4], Elaine Maria Seles Dorneles[5], Dionei Joaquim Haas[5], Marcos Bryan Heinemann[6], Preetam Ghosh[7], Andrey Pereira Lage[5], Henrique Figueiredo[8], Rafaela Salgado Ferreira[2] and Vasco Azevedo[1*]

Abstract

Background: *Corynebacterium pseudotuberculosis* can be classified into two biovars or *biovars* based on their nitrate-reducing ability. Strains isolated from sheep and goats show negative nitrate reduction and are termed biovar Ovis, while strains from horse and cattle exhibit positive nitrate reduction and are called biovar Equi. However, molecular evidence has not been established so far to understand this difference, specifically if these *C. pseudotuberculosis* strains are under an evolutionary process.

Results: The ERIC 1 + 2 Minimum-spanning tree from 367 strains of *C. pseudotuberculosis* showed that the great majority of biovar Ovis strains clustered together, but separately from biovar Equi strains that also clustered amongst themselves. Using evolutionarily conserved genes (*rpoB, gapA, fusA, and rsmE*) and their corresponding amino acid sequences, we analyzed the phylogenetic relationship among eighteen strains of *C. pseudotuberculosis* belonging to both biovars Ovis and Equi. Additionally, conserved point mutation based on structural variation analysis was also carried out to elucidate the genotype-phenotype correlations and speciation. We observed that the biovars are different at the molecular phylogenetic level and a probable anagenesis is occurring slowly within the species *C. pseudotuberculosis*.

Conclusions: Taken together the results suggest that biovar Equi is forming the biovar Ovis. However, additional analyses using other genes and other bacterial strains are required to further support our anagenesis hypothesis in *C. pseudotuberculosis*.

Keywords: *Corynebacterium pseudotuberculosis*, Evolution, Molecular phylogeny, Structural biology

Background

The genus *Corynebacterium* belongs to the bacterial phylum *Actinobacteria*, also known as *Actinomycetes*. This phylum comprises *Mycobacterium*, *Nocardia* and *Rhodococcus* genera, which together form a supra-generic group known by their initials as CMNR [1–3]. These organisms share some common features, such as:

(i) A specific well-organized cell wall mainly characterized by the presence of vast components of peptidoglycan, mycolic acid, and arabinogalactan [4–7];
(ii)high G + C content (47 %–74 %) [5];
(iii)Gram-positive [8].

Within the genus *Corynebacterium*, the species *Corynebacterium pseudotuberculosis* is reported to be a facultative intracellular pathogen in mammals [9, 10]. The pathologies associated with *C. pseudotuberculosis* are of great importance to veterinary medicine because this bacterium is considered the main etiologic agent of

* Correspondence: vasco@icb.ufmg.br
[1]Departamento de Biologia Geral, Laboratório de Genética Celular e Molecular, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
Full list of author information is available at the end of the article

Ab-Initio Prediction of Sequence and Structural Biology of Fish Muscle Proteins Using Homology Modeling, Phylogeny and Different Computational Approaches

Abstract

Fish is a diverse group of organisms living in different aquatic environment and containing almost all essential amino acids. Fourteen muscle proteins including titin, dystrophin, filamin, myosin heavy chain, spectrin, M1/M2, nebulin, alpha-actinin, gelsolin, actin, tropomyosin, troponin, thymosin and plastin3 were chosen for in-silico characterization. Sequence analyses were performed using BindN, Conseq, DIANNA, PROFEAT and ProtFun for exploiting structural and functional importance. Homology modeling technique was applied for predicting 3D structure which will assist in future for searching catalytic role of proteins in metabolic pathway. 3D Structure of eight muscle proteins was predicted using Protein Structure Prediction Server (PS^2) based on MODELLER algorithm. Phylogenetic relationship was inferred by sequence alignment through CLUSTAL X and furthermore phylogenetic tree was constructed by using MEGA which was statistically evaluated by DIVIEN. From structural analyses, these muscle proteins were inferred to contain functional domains, number of motifs, beta turns with important secondary structural features. Furthermore sequence study suggested, these proteins have important biochemical features such as number of cysteines, disulphide bonds, DNA and RNA binding sites, functionally conserved amino acid residues and were characterized as non-allergen proteins which can be used for designing effective vaccines. Overall, evidence from computational study revealed that these muscle proteins have structural and functional significance, which can play important role in drug designing and in exploring gene diversity. This novel approach to study muscle proteins would be beneficial for human since both vertebrates and invertebrates have muscle proteins in common.

Keywords: Sequence analyses; Homology modeling; Structural analyses; Vertebrates; Invertebrates

Research Article

Volume 2 Issue 3 - 2015

Sana Khalid[1], Sobia Idrees[1], Hina Khalid[1], Bilal Hussain[1], Sandeep Tiwari[2], Syed Shah Hassan[2], Artur Silva[3], Vasco Azevedo[2] and Syed Babar Jamal[2]*

[1]*Department of Bioinformatics and Biotechnology, Government College University Faisalabad, Pakistan*
[2]*Departamento de Biologia Geral, Instituto de Ciências Biológicas (ICB), Universidade Federal de Minas Gerais, Brazil*
[3]*Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, Brazil*

***Corresponding author:** Syed Babar Jamal, Laboratório de GenéticaCelular e Molecular, Departamento de Biologia Geral, Instituto de CiênciasBiológicas (ICB), Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, Tel: 005531 3409 2610; Fax: 005531 3409 2610; Email: syedbabar.jamal@gmail.com

Received: April 9, 2015 | **Published:** May 05, 2015

Introduction

Protein with its amino acid are important for maintaining structure of cells, making antibodies to work properly, regulate the growth of hormones with enzymes and contributes to the repairing mechanisms. Organism living in marine and fresh water consists of protein with high amino acid proportion. Fish is a diverse group of organisms that habituates in different aquatic environment and holds prime importance in food industry. Biologically, fish muscle proteins contain all essential nutrients like milk, meat and egg protein. This protein varies in amount from species to species. Globally the consumption of fish production by human is about 77 percent. Fish was chosen as a sample source because there are many different verities of fish and source of protein for many fish species are readily available. Furthermore, fish is very nutritious part of man's diet since it is rich in vitamins, minerals and all essential amino acids in right proportions. Study of muscle genes and proteins will be beneficial for human for *in silico* drug designing. Fish skeletal muscle is known to be the perfect model to explore the structure and function of muscle, due to perfect arrangement of different types of fibers which is present in axial and pectoral fin muscles [1]. Thus, computational study will allow muscle genes and proteins to be studied at greater level of detail. A variety of bioinformatics tools are available for detailed comparative study and visualization of amino acid sequences, which provides knowledge about molecular evolution and variety of information related to structure and function of protein. Detection of conserved regions in protein and nucleic acid sequences are of great importance, because it gives knowledge about structure and function [2]. Then *in silico* study of fish muscle proteins was performed to analyze its structural and functional importance with amino acid properties.

The objective of present study was to perform sequence analysis of fish muscle proteins, using different computational tools, study the amino acid composition and secondary structure features, using homology-modeling approach to find the 3D structure of muscle proteins.In addition, illustrate physiochemical properties by ensuring the quality of the predicted model and finally predicting the evolutionary relationship of various proteins to get knowledge about biodiversity of different species with homologous sequences.

Materials and Methods

Protein retrieval and sequence analysis

Protein sequences of fish muscle were retrieved from Uniprot Knowledgebase database and NCBI using accession no. G1ERR8, Q9PV76, E6ZGD0, Q9PRF1, F8K8N3, Q1L5K3, E6ZHF3, gi|5726351, Q8AW95, gi|59858543, Q58HZ6, Q9NAS5,

Send Orders for Reprints to reprints@benthamscience.ae

Current Protein and Peptide Science, 2015, *16*, 000-000

In Silico Protein-Protein Interactions: Avoiding Data and Method Biases Over Sensitivity and Specificity

Edson Luiz Folador[1], Alberto Fernandes de Oliveira Junior[1], Sandeep Tiwari[1], Syed Babar Jamal[1], Rafaela Salgado Ferreira[2], Debmalya Barh[3,6], Preetam Ghosh[4], Artur Silva[5], and Vasco Azevedo[1,*]

[1]*Department of General Biology, Instituto de Ciências Biológicas (ICB), Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil;* [2]*Department of Biochemistry and Immunology, Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil;* [3]*Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology (IIOAB), Nonakuri, Purba Medinipur, West Bengal, India;* [4]*Department of Computer Science, Virginia Commonwealth University, Richmond, VA, USA;* [5]*Instituto de Ciências Biológicas, Universidade Federal do Para, Belém, PA, Brazil;* [6]*InterpretOmics India Pvt. Ltd., #329 7th Main, HAL 2nd stage, Indiranagar, Bengalore, Karnataka, India*

Vasco Azevedo

Abstract: The study of protein-protein interactions (PPIs) can help researchers raise new hypotheses about an organism or disease and guide new experiments. Various methods for the identification and analysis of PPIs have been discussed in the literature. These methods are generally categorized as experimental or computational - each having its own advantages and disadvantages. Experimental methods provide insights into the real state of biological interactions but tend to be time-consuming and costly. Computational methods, on the other hand, can study thousands of PPIs at a very low cost and in much less time; however, the accuracy of such *in silico* prediction results heavily depends on the specific computational approach used. Furthermore, there is no gold standard for these computational methods; a method that works well for predicting one PPI may perform poorly (by generating false positives and false negatives) for a different PPI. Therefore, all such predictions must be carefully validated, preferably with experimental data. In this paper, we review the existing computational approaches and emphasize the use of biological data as inputs for accurate predictions of PPIs. We also discuss how such input datasets and approaches may influence the sensitivity and specificity of the predicted PPI networks.

Keywords: Computational approaches, *in silico* prediction, protein-protein interaction, protein network, system biology.

1. INTRODUCTION

Most of the data related to protein interaction networks have come from high-throughput experiments that provide information about interacting protein pairs [1]. However, such high-throughput experimental methods have their own limitations because they often produce a high number of false positives and false negatives, while also being time-consuming and labor-intensive [2]. The advent of second and third generation sequencing has allowed genomes to be sequenced with significant reductions in cost and time, which has resulted in an ever-increasing number of sequenced genomes. Consequently, the amounts of biological data related to DNA, RNA, transcripts, proteins, and other molecules have also increased, opening up new opportunities to explore these data and better understand organisms and diseases, thereby aiding the search for new pharmaceuticals or products for use in veterinary, agriculture, or biotechnology industries [3-9].

DNA, RNA, and proteins all work together to perform the biological functions of living cells. To understand how these molecules interact and are grouped, is an important area of study that can help develop a better understanding of an organism, a disease, or a biological process at a system level. An important tool for understanding complex biological systems is the study of protein-protein interactions (PPIs), which form complex networks of biological interactions called interactomes [10-14]. In the literature, several experimental and computational methods for the identification of PPIs have been reported. Important experimental methods include yeast-two-hybrid [15-17], protein chip [18, 19], tandem affinity purification followed by mass spectrometry [20-22], and biophysics-based methods, such as, atomic force microscopy [23-27], analytical ultracentrifugation [28-30], and nuclear magnetic resonance (NMR) [31-33]. We focus on only the computational methods in this paper and provide a detailed review of such computational methods used to predict PPIs [34] specifically.

Based on the volume of data that they generate, computational methods can also be classified as low-throughput or high-throughput methods. When using computational tools to predict PPIs, an important aspect is to validate the results

*Address correspondence to this author at the Department of General Biology, Instituto de Ciências Biológicas (ICB), Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil; Tel: +55 31 3409 2610;
Fax: +55 31 3409-2614; E-mail: vasco@icb.ufmg.br

VOLUME: 16 ISSUE: 8
Page: [689 - 700]
Pages: 12
DOI: 10.2174/1389203716666150505235437

Radusky et al. BMC Genomics 2015, 16(Suppl 5):S9
www.biomedcentral.com/1471-2164/16/S5/S9

BMC Genomics

RESEARCH Open Access

An integrated structural proteomics approach along the druggable genome of *Corynebacterium pseudotuberculosis* species for putative druggable targets

Leandro G Radusky[1†], Syed Shah Hassan[3†], Esteban Lanzarotti[1], Sandeep Tiwari[3], Syed Babar Jamal[3], Javed Ali[4], Amjad Ali[5], Rafaela Salgado Ferreira[6], Debmalya Barh[7], Artur Silva[8], Adrián G Turjanski[1,2*], Vasco AC Azevedo[3*]

From X-meeting 2014 - International Conference on the Brazilian Association for Bioinformatics and Computational Biology
Belo Horizonte, Brazil. 28-30 October 2014

Abstract

Background: The bacterium *Corynebacterium pseudotuberculosis* (Cp) causes caseous lymphadenitis (CLA), mastitis, ulcerative lymphangitis, and oedema in a number of hosts, comprising ruminants, thereby intimidating economic and dairy industries worldwide. So far there is no effective drug or vaccine available against Cp. Previously, a pan-genomic analysis was performed for both biovar *equi* and biovar *ovis and a* Pathogenicity Islands (PAIS) analysis within the strains highlighted a large set of proteins that could be relevant therapeutic targets for controlling the onset of CLA. In the present work, a structural druggability analysis pipeline was accomplished along 15 previously sequenced Cp strains from both biovar *equi* and biovar *ovis*.

Methods and results: We computed the whole modelome of a reference strain Cp1002 (NCBI Accession: NC_017300.1) and then the homology models of proteins, of 14 different Cp strains, with high identity (≥ 85%) to the reference strain were also done. Druggability score of all proteins pockets was calculated and only those targets that have a highly druggable (HD) pocket in all strains were kept, a set of 58 proteins. Finally, this information was merged with the previous PAIS analysis giving two possible highly relevant targets to conduct drug discovery projects. Also, off-targeting information against host organisms, including *Homo sapiens* and a further analysis for protein essentiality provided a final set of 31 druggable, essential and non-host homologous targets, tabulated in **table S4**, additional file 1. Out of 31 globally druggable targets, 9 targets have already been reported in other pathogenic microorganisms, 3 of them (3-isopropylmalate dehydratase small subunit, 50S ribosomal protein L30, Chromosomal replication initiator protein DnaA) in *C. pseudotuberculosis*.

Conclusion: Overall we provide valuable information of possible targets against *C. pseudotuberculosis* where some of these targets have already been reported in other microorganisms for drug discovery projects, also discarding targets that might be physiologically relevant but are not amenable for drug binding. We propose that the

* Correspondence: vasco@icb.ufmg.br; adrian@qi.fcen.uba.ar
† Contributed equally
[1]Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pabellón II, Buenos Aires C1428EHA, Argentina
[3]PG Program in Bioinformatics, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
Full list of author information is available at the end of the article

PublisherInfo

PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

Thiago Motta Venancio
João Carlos Setubal
Richard Garratt
Emanuel M Souza
André Fujita
Ricardo Z Vêncio
Hélder I Nakaya
Enrique Medina-Acosta

Proteome scale comparative modeling for conserved drug and vaccine targets identification in *Corynebacterium pseudotuberculosis*

ArticleInfo

ArticleID	:	7174
ArticleDOI	:	10.1186/1471-2164-15-S7-S3
ArticleCitationID	:	S3
ArticleSequenceNumber	:	7
ArticleCategory	:	Research
ArticleFirstPage	:	1
ArticleLastPage	:	19
ArticleHistory	:	RegistrationDate : 2014–10–27 OnlineDate : 2014–10–27
ArticleGrants	:	
ArticleContext	:	128641515S7S7

Syed Shah Hassan,[Aff1]
Email: hassan_chemist@yahoo.com

Sandeep Tiwari,[Aff1]
Email: sandip_sbtbi@yahoo.com

Luís Carlos Guimarães,[Aff1]
Email: luisguimaraes.bio@gmail.com

Syed Babar Jamal,[Aff1]
Email: syedbabar.jamal@gmail.com

Edson Folador,[Aff1]

Neha Barve Sharma,[Aff4 Aff5]
Email: nehabarve2006@gmail.com

Siomar de Castro Soares,[Aff1]
Email: siomars@gmail.com

Síntia Almeida,[Aff1]
Email: sintiaalmeida@gmail.com

Amjad Ali,[Aff1]
Email: amjad_uni@yahoo.com

Arshad Islam,[Aff6]
Email: arshad.cgl@gmail.com

Fabiana Dias Póvoa,[Aff2]
Email: fabianapovoa@gmail.com

Vinicius Augusto Carvalho de Abreu,[Aff1]
Email: vini.abreu@gmail.com

Neha Jain,[Aff4 Aff5]
Email: mymailtoneha@gmail.com

Antaripa Bhattacharya,[Aff5]
Email: antaripa1210@gmail.com

Lucky Juneja,[Aff4 Aff5]
Email: ljpreet88@gmail.com

Anderson Miyoshi,[Aff1]
Email: miyoshi@icb.ufmg.br

Artur Silva,[Aff3]
Email: asilva@ufpa.br

Debmalya Barh,[Aff5]
Email: dr.barh@gmail.com

Adrian Gustavo Turjanski,[Aff7]
Email: adrian@qi.fcen.uba.ar

Vasco Azevedo,[Aff1]
Email: vasco@icb.ufmg.br

Rafaela Salgado Ferreira,[Aff2]
Corresponding Affiliation: Aff2
Email: rafaelasf@gmail.com

Aff1 Laboratory of Cellular and Molecular Genetics, Department of General Biology, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Aff2 Departament of Biochemistry and Immunology, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Aff3 Institute of Biological Sciences, Federal University of Pará, Belém, Para, Brazil

Aff4 School of Biotechnology, Devi Ahilya University, Khandwa Road Campus, Indore, MP, India

Aff5 Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology (IIOAB), Nonakuri, Purba Medinipur, West Bengal, India

Aff6 Department of Chemistry, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Aff7 Structural Bioinformatics Group, Institute of Physical Chemistry of Materials, Environment and Energy, University of Buenos Aires, Argentine

Abstract

Corynebacterium pseudotuberculosis (Cp) is a pathogenic bacterium that causes caseous lymphadenitis (CLA), ulcerative lymphangitis, mastitis, and edematous to a broad spectrum of hosts, including ruminants, thereby threatening economic and dairy industries worldwide. Currently there is no effective drug or vaccine available against Cp. To identify new targets, we adopted a novel integrative strategy, which began with the prediction of the modelome (tridimensional protein structures for the proteome of an organism, generated through comparative modeling) for 15 previously sequenced *C. pseudotuberculosis* strains. This pan-modelomics approach identified a set of 331 conserved proteins having 95-100% intra-species sequence similarity. Next, we combined subtractive proteomics and modelomics to reveal a set of 10 Cp proteins, which may be essential for the bacteria. Of these, 4 proteins (tcsR, mtrA, nrdI, and ispH) were essential and non-host homologs (considering man, horse, cow and sheep as hosts) and satisfied all criteria of being putative targets. Additionally, we subjected these 4 proteins to virtual screening of a drug-like compound library. In all cases, molecules predicted to form favorable interactions and which showed high complementarity to the target were found among the top ranking compounds. The remaining 6 essential proteins (adk, gapA, glyA, fumC, gnd, and aspA) have homologs in the host proteomes. Their active site cavities were compared to the respective cavities in host proteins. We propose that some of these proteins can be selectively targeted using structure-based drug design approaches (SBDD). Our results facilitate the selection of *C. pseudotuberculosis* putative proteins for developing broad-spectrum novel drugs and vaccines. A few of the targets identified here have been validated in other microorganisms, suggesting that our modelome strategy is effective and can also be applicable to other pathogens.

ESMHint

Electronic supplementary material

The online version of this article (doi:10.1186/1471-2164-15-S7-S3) contains supplementary material, which is available to authorized users.

Integrative Biology

PAPER

View Article Online
View Journal | View Issue

Cite this: *Integr. Biol.*, 2014, 6, 1088

C. pseudotuberculosis Phop confers virulence and may be targeted by natural compounds†

Sandeep Tiwari,‡[a] Marcília Pinheiro da Costa,‡[b] Sintia Almeida,[a] Syed Shah Hassan,[a] Syed Babar Jamal,[a] Alberto Oliveira,[a] Edson Luiz Folador,[a] Flavia Rocha,[c] Vinícius Augusto Carvalho de Abreu,[a] Fernanda Dorella,[a] Rafael Hirata,[d] Diana Magalhaes de Oliveira,[b] Maria Fátima da Silva Teixeira,[b] Artur Silva,[e] Debmalya Barh[f] and Vasco Azevedo*[ac]

Received 18th June 2014,
Accepted 22nd August 2014

DOI: 10.1039/c4ib00140k

www.rsc.org/ibiology

The bacterial two-component system (TCS) regulates genes that are crucial for virulence in several pathogens. One of such TCS, the PhoPR system, consisting of a transmembrane sensory histidine kinase protein (PhoR) and an intracellular response regulator protein (PhoP), has been reported to have a major role in mycobacterial pathogenesis. We knocked out the *phoP* in *C. pseudotuberculosis*, the causal organism of caseous lymphadenitis (CLA), and using a combination of *in vitro* and *in vivo* mouse system, we showed for the first time, that the *PhoP* of *C. pseudotuberculosis* plays an important role in the virulence and pathogenicity of this bacterium. Furthermore, we modeled the *PhoP* of *C. pseudotuberculosis* and our docking results showed that several natural compounds including Rhein, an anthraquinone from *Rheum undulatum*, and some drug-like molecules may target *PhoP* to inhibit the TCS of *C. pseudotuberculosis*, and therefore may facilitate a remarkable attenuation of bacterial pathogenicity being the CLA. Experiments are currently underway to validate these *in silico* docking results.

Insight, innovation, integration

Here, we report for the first time the importance of *PhoP* protein of the PhoPR system of *C. pseudotuberculosis* that plays a vital role in the virulence and pathogenicity of this bacterium. For this, we developed a mutant *phoP* gene of *Corynebacterium pseudotuberculosis* Cp1002 strain and subsequently evaluated the consequences of this mutation *via in vitro* and *in vivo* analyses. Furthermore, by extrapolating the results of our analyses, we modeled this *PhoP* protein of *C. pseudotuberculosis* and computational analyses were performed. Our docking results showed that several natural compounds, including Rhein from *Rheum undulatum* and some other drug-like compounds acquired from public drug database/s, targeted the N-terminal response regulator domain of the *PhoP* protein to inhibit the TCS of *C. pseudotuberculosis* and might facilitate a remarkable attenuation of the pathogenicity of this important veterinary pathogen.

Introduction

The Gram-positive bacterium *Corynebacterium pseudotuberculosis* of the class *Actinobacteria* is the etiological agent of caseous lymphadenitis (CLA), or cheesy gland disease, an illness that affects small ruminants (sheep and goats) worldwide.[1] CLA is mainly characterized by abscess formation in superficial and internal lymph nodes.[2] In some cases, the visceral organs,

[a] *PG Program in Bioinformatics, Laboratory of Cellular and Molecular Genetics (LGCM), Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil. E-mail: sandip_sbtbi@yahoo.com, sintiaalmeida@gmail.com, hassan_chemist@yahoo.com, Jamal-syedbabar.jamal@gmail.com, afojunior@gmail.com, edson.folador@gmail.com, vini.abreu@gmail.com, fernandadorella@gmail.com, vasco@icb.ufmg.br*

[b] *Center for Genomics and Bioinformatics, State University of Ceará, Fortaleza, CE, Brazil. E-mail: marciliacosta@hotmail.com, Diana.magalhaes@uece.br, mfteixeira@hotmail.com*

[c] *Department of General Biology, PG Program in Genetics, Laboratory of Cellular and Molecular Genetics (LGCM), Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil. E-mail: flasouz@yahoo.com.br, vascoariston@gmail.com; Fax: +55-31 3409-2610; Tel: +55-31 3409-2610*

[d] *Universidade do Estado do Rio de Janeiro, UERJ, Brazil. E-mail: rhiratajunior@gmail.com*

[e] *Laboratory of DNA polymorphism (LPDNA), Federal University of Pará, Belém, PA, Brazil. E-mail: asilva@ufpa.br*

[f] *Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology (IIOAB), Nonakuri, Purba Medinipur, WB, India. E-mail: dr.barh@gmail.com*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c4ib00140k

‡ These authors contributed equally to this work.

The Pan-Genome of the *Treponema pallidum* Reveals Differences in Genome Plasticity between the subspecies.

Arun Kumar Jaiswal[a,b], Sandeep Tiwari[a], Syed Babar Jamal[a], Vasco Azevedo[a], Carlo Jose Freira Oliveira[b], Siomar C. Soares[b]*

Submitted in Elesvier Gene, 2018

Abstract

Spirochetal organisms of the *Treponema* species are responsible for causing Treponematoses. Pathogenic treponemes cause multi-stage infections like endemic syphilis, venereal syphilis, yaws and pinta in human. Primarily, the pathogenic treponemes can be classified based on the clinical symptoms of the respective disease they cause. *Treponema pallidum* subsp. *endemicum* causes endemic syphilis; *T. pallidum* subsp. *pallidum* causes venereal syphilis; *T. pallidum* subsp. *pertenue* causes yaws; and *Treponema carateum* causes pinta. Out of these four lethal diseases, venereal syphilis is transmitted by sexual contact; the other three diseases are transmitted by close personal contact. *Treponema pallidum* subsp.ecies *pallidum* is a Gram-negative, motile, spirochete pathogen that causes syphilis in human. Syphilis is a multistage infectious disease that can be either communicated between sexual partners, through active lesions, or from an infected woman to her foetus, during pregnancy. The current worldwide prevalence of syphilis emphasizes the need for continued preventive measures and strategies. Unfortunately, effective measures are limited. Here, the genome sequence of all 49 *T. pallidum* strains isolated from the different hosts and countries were comparatively analyzed using pan-genomic strategy. Phylogenomic, pan-genomic, core genomic and singleton analysis disclosed the close connection among all strains of the pathogen *Treponema pallidum*, the clonal behaviour of *T. pallidum* and slowed increases in the sizes of pan-genome. According to the pan-genomes, core genomes, singletons, the strains of *Treponema pallidum* are highly clonal. Based on the genome plasticity analysis of the subsets containing the subsp.ecies *T pallidum* subsp. *pallidum*, *T. pallidum* subsp. *endemicum* and *T. pallidum* subsp. *pertenue*, we found differences in the presence/absence of pathogenicity islands (PAIs) and genomic islands (GIs) on subsp. based study. In summary, we identified 4 pathogenicity islands (PAIs), 8 genomic islands (GIs) in subsp. *pallidum*, whereas subsp. *endemicum* has 3 PAIs and 7 GIs and subsp. *pertenue* harbours

3 PAIs and 8 GIs. Concerning the presence of genes in PAIs and GIs, we found high similarity of the genes in all the subsp. of *T. pallidum* compared to each other.

Ack is essential for microbial growth and betel compounds potentially target Ack, PhoP and MDR proteins in *M. tuberculosis*, *V. cholera*, and pathogenic *E. coli*

Sandeep Tiwari[1,2¥], Debmalya Barh[1,2¥*], Ranjith K. Kumavath[3], S. Prabu Seenivasan[4], Syed Babar Jamal[2], Vanaja Kumar[5], Preetam Ghosh[6], Vasco Azevedo[2]

Submitted in Current Topics in Medicinal Chemistry, 2017

Abstract

M. tuberculosis, *V. cholerae*, *Y. pestis* and pathogenic *E. coli* are global concern of public health. The emergence of multi-drug resistant (MDR) strains of these pathogens is creating additional challenges in controlling infections caused by these deadly bacteria. Recently we reported Acetate kinase (Ack) could be a broad-spectrum novel target in several bacteria including these pathogens. Here using *in vitro* and *in silico* approaches we show that (i) Ack is an essential protein in bacteria. (ii) Natural compounds chlorogenic acid, pinoresinol from *Piper betel* and Piperidine derivative compound 6-oxopiperidine-3-carboxylic acid inhibit the growth of pathogenic *E. coli* and *M. tuberculosis* by targeting Ack with equal or higher efficacy than the currently used antibiotics. (iii) The compounds are highly effective even on MDR strains of these pathogens. (iv) These compounds may target bacterial two component systems response regulatory protein, PhoP that help bacteria in the expression of gene related to drug resistance and virulence. (v) Finally, all these betel compounds are predicted to have drug-like properties suggesting that these betel derived compounds may be further tested in developing novel class of broad spectrum drugs for common and multi drug resistant various pathogens.

IV. Appendix

B. Book chapters

(My contributions to the Book Chapters)

Our research group has been invited a number of times to write book chapters for graduates and masters students. I actively participated in this work and collected data from authentic sources; finally, I make a manuscript of the assigned theme. This activity helped me in improving data mining and manuscript writing skills. Furthermore, it was really useful in understanding the research swork going on in different research areas around the world. Here I have mentioned the booked chapters published with different group of publishers.

SMGroup

Industrial Microbiology & Biotechnology

Sandeep Tiwari[1], Syed Babar Jamal[1], Paulo Vinícius Sanches Daltro de Carvalho[1], Syed Shah Hassan[1], Artur Silva[2] and Vasco Azevedo[1]

[1]Universidade Federal de Minas Gerais, UFMG, Brazil

[2]Universidade Federal do Pará, Belém, PA, Brazil.

***Corresponding author:** Vasco Azevedo, Universidade Federal de Minas Gerais/Instituto de CiênciasBiológicas, Minas Gerais, Brazil, Email: vasco@icb.ufmg.br

Published Date: February 15, 2015

ABSTRACT

Microbes are small living organisms, quite diverse and adapt to various environments. They live on our planet long before that markedly effects the human, animals and plants life. Biotechnologically, molecular genetics of microbes is systematically manipulated, enable them for the production of beneficial products. The large-scale production of insulin was made possible through transfer of human insulin gene to a bacterium to treat diabetes, a milestone in the history of biotechnology. The microbial cultures are managed, monitored and maintained for their genotype/phenotype stability. Bacteria are one of the biotechnologically important microbes for example Escherichia coli, some species of lactic acid bacteria, and some viruses such as bacteriophages etc. recognized as anti-cancer oncolytic viruses. Moreover, yeast has also been used with broad-spectrum applications, such as Saccharomyces cerevisiae. Using diverse genetic technologies, a number of beneficial products like microbial polymers, biologically active primary and secondary metabolites like ethanol and antibiotics have been produced so far and used. Here, it would be injustice not to comment on the importance of vaccines for several infectious diseases through their ability to trigger the host immune system. Furthermore, roles of microorganisms in the production of pharmaceutical proteins

Protein-Protein Interactions: An Overveiw

Edson Luiz Folador, Sandeep Tiwari, Camila Eduarda da Paz Barbosa, Syed Babar Jamal, Marco da Costa Schulze, Debmalya Barh, Vasco Azevedo*

Submitted in Encyclopedia of Bioinformatics and Computational Biology, 2017

ABSTRACT

The proteins perform their activities in a cell through their interactions, forming a complex protein-protein interaction network (PIN) that can contain tens of thousands of interactions. Knowing the interactions enables know how the organism performs its functions internally and how it interacts with the host. In Bioinformatics, by using graph theory from the computer science area, we can calculate measurements on PINs revealing relevant biological information. Knowing these measures as they are calculated and their biological relevance is key to validating and interpreting PINs, since it is humanly impossible to analyze and extract information in complex PINs. PINs from various organisms, generated by experimental or computational methods, can be found and downloaded from the public databases for further analysis. PINs for an organism of interest can be generated by experimental methods or predicted by computational methods in both low and large scale. Knowing the biological assumptions underlying each method, whether the method is able to predict new interactions or identify previously characterized interactions, as well as advantages and disadvantages, is essential to select the method appropriate to our purpose. Regardless of the method used, biological PINs can be applied in various contexts such as hypothetical protein annotation, understanding the organism at systems biology level, identifying essential or important proteins in a biological context to be used as a target for drugs, identifying host-pathogen interactions to suggest vaccine targets, in addition to enabling research to identify a new drug class inhibiting or stabilizing interaction. PINs are generated not as the ultimate goal of a research but rather as a tool to better understand the mechanisms of action of an organism and to direct future experiments in the laboratory. In this chapter, we have discussed various aspects of PINs including their applications.

IV. Appendix

C. Curriculum Vitae

Curriculum Vitae

SYED BABAR JAMAL BACHA

ADDRESS:
House#70, street#1, Sector G, Sheikh Maltoon Town Mardan Khyber PukhtoonKhwa, Pakistan. (Permanent)

Mobile: +55-31-993441901
Email: syedbabar.jamal@gmail.com
baber.msbi1@iiu.edu.pk

Rua Custódio de Melo, 47, Bairro Liberdade, CEP: 31270-790
Belo Horizonte, Minas Gerais, Brazil. (Current)

EDUCATION PROFILE

Federal University of Minas Gerais, Belo Horizonte, Brazil
PhD Bioinformatics (March-2014 to March-2018)

International Islamic University, Islamabad, Pakistan
MS Bioinformatics

GC University Faisalabad, Punjab, Pakistan
BS (hons) Bioinformatics

PROFESSIONAL EXPERIENCE

Research Assistant February 2012 – February 2014
Computational Medicinal Chemistry Laboratory
UCS Shankar, ***Abdul Wali Khan University, KPK, Pakistan***

Research Internee April 2010 – October 2010
Genetic Diseases
Center of Excellence in Molecular Biology ***(CEMB), Punjab University Lahore***

RESEARCH EXPERIENCE AND TRAININGS

- Protein-Ligand Interaction **(Molecular Docking),** Structural Modeling **(Homology Modeling of proteins), Molecular Dynamic Simulation, QSAR study** at UCS Shankar, ***Abdul Wali Khan University, KPK, Pakistan.***
- **Phylogenetic** analysis training at ***International Islamic University Islamabad.***
- ***Virtual Training of Bioinformatics*** *at Center of Excellence in Molecular Biology* ***(CEMB), Punjab University Lahore.***
- ***Six months*** *research experience of Molecular biology and Biotechnology at Center of Excellence in Molecular Biology* ***(CEMB), Punjab University Lahore.***
- *Bacterial Genome assembly and annotation. Laboratory of cellular and molecular Genetics (LGCM),* Institute of Biological Science. *Federal University of Minas Gerais, Belo Horizonte, Brazil.*
- Training with **EMBL** team for course entitled "Genomics for working with pathogens", which was held at Belo Horizonte, Minas Gerais, Brazil, from May 26th to 29th, 2014.
- Participated in 30 hours international course on "**Metagenomics**" of Post Graduate Program of Health Sciences, at research center of FIOCRUZ Minas Gerais, Belo Horizonte – Brazil, between 2nd and 6th of February 2015.
- Participated in 2nd international workshop on "**Advanced Topics in Proteomics**" held at research center of FIOCRUZ Rio de Janeiro – Brazil, between 28th of August and 4th of September 2015.
- Comparative genomics of **ZIKA Virus** genomes. *Laboratory of cellular and molecular Genetics (LGCM),* Institute of Biological Science, *Federal University of Minas Gerais, Belo Horizonte, Brazil.*
- Subtractive genomics and genome based high throughput modeling. *Federal University of Minas Gerais, Belo Horizonte, Brazil* **&** *FIOCRUZ Rio de Janeiro, Brazil.*

HONORS AND AWARDS

- ***Certified*** for Organizing Tech-fair 2008 at *GC University* ***Faisalabad (Pakistan).***
- ***First Position*** in DNA poster exhibition held in GC University ***Faisalabad (Pakistan).***
- ***Certified*** for warmly participating in different national and international conferences in different cities of Pakistan.
- ***Certified*** for attending workshop entitled "**Workshop of Gene Ontology**", by Dr. Prudence Mutowo at Institute of Biological Science, Federal University of Minas Gerais, Belo Horizonte - Brazil between 27th and 28th September of 2014.
- ***Certified*** for attending workshop course of 15 hours entitled "**Molecular Dynamics of Complex systems; Nucleosome as model**", organized by Post Graduate program of Bioinformatics Institute of Biological Science, Federal University of Minas Gerais, Belo Horizonte - Brazil on 10th October of 2014.
- ***Certified*** for participating in **1st Latin American Student Council Symposium** at Institute of Biological Science, Federal University of Minas Gerais, Belo Horizonte - Brazil on 27th of October 2014.
- ***Certified*** for warmly participating in ***X-Meeting 2014*** - 10th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (AB3C), held at

Institute of Biological Science, Federal University of Minas Gerais, Belo Horizonte - Brazil between 28th and 31st October of 2014.

- ***Certified*** for participating in **2nd Symposium of Microbiology** at Institute of Biological Science, Federal University of Minas Gerais, Belo Horizonte - Brazil between 5th and 6th of October 2015.
- ***Certified*** for warmly participating in ***X-Meeting 2015*** - 11th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (AB3C), held in São Paulo - Brazil between 3rd and 6th November of 2015.
- ***Certified*** for warmly participating in ***X-Meeting 2016*** - 12th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (AB3C), held in Belo Horizonte, MG - Brazil between 16th and 18th November of 2016.
- ***Certified*** for warmly participating in an "International course of Bioinformatics in Molecular and Evolutionary Epidemiology of virus", ***(CIBEMEV),*** held at Institute of Biological Science, Federal University of Minas Gerais, Belo Horizonte - Brazil between 25th and 29th September of 2017.
- ***Certified*** for warmly participating in ***X-Meeting 2017*** - 13th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (AB3C), held in São Pedro - Brazil between 4th and 6th October of 2017.

TEACHING EXPERIENCE

Visiting Lecturer at Department of Biochemistry
(UCS Shankar, Abdul Wali Khan University, Mardan, Pakistan)

TECHNICAL SKILLS AND EXPERIENCE

Molecular Biology Techniques:

- DNA extraction, DNA extraction reagents preparation, SDS PAGE, Gel Electrophoresis, Polymerase Chain Reaction (PCR), Tissue Culture (Micro propagation) and Data analysis using softwares.

Computer:

- Microsoft Office Suite.
- Proficient with Windows and Linux Operating Systems
- Statistical software packages: Excel, SPSS

Bioinformatics:

- Data mining, Sequence Alignment, Gene Finding, Protein Structure Alignment, Protein Structure Prediction, Protein-Protein Docking, Protein-Ligand Docking, QSAR Modeling, Computational Evolutionary Biology (phlogenetics), Pharmacophore Design, Virtual Screening, Molecular Dynamic simulation, Comparative Genomics, Subtractive Genomics, Genome Scale High Throughput Modeling, Genome annotation.

PROFESSIONAL MEMBERSHIPS

- Member of Regional Student Group of Pakistan for International Society for Computational Biology *(ISCB-RSG-Pakistan)*
- Member of the *National Academy for Young Scientists* Pakistan (2009 - 2011)
- Member of the *International Society for Computational Biology (ISCB)* USA
- Member of the *International Society for Computational Biology (ISCB)* Brazil

- Member of the Latin American Student Council (LASC) Brazil
- Member of *World Association of Young Scientists (WAYS)*

Publications:

1. Hassan SS[†], **Jamal SB[†],** Radusky LG[†], Tiwari S[†], Ullah A[1], Ali J[4], Behramand, Carvalho PVSD, Shams R, Khan S, Figueiredo HC, Barh D, Ghosh P, Silva A, Baumbach J, Röttger R, Turjanski AG*, Azevedo V*. The Druggable Pocketome of *Corynebacterium diphtheriae*: A new approach for *in-silico* putative druggable targets. *Fronteirs in Genetics, **2018**; 9:44. doi: 10.3389/fgene.2018.00044*
2. **Jamal SB,** Hassan SS, Tiwari S, Viana MV, Benevides LJ, Ullah A, et al. "An integrative in-silico approach for therapeutic target identification in the human pathogen Corynebacterium diphtheria". *PLoS One, **2017;** 12(10), e0186401. doi: 10.1371/journal.pone.0186401*.
3. **Jamal SB,** Tiwari S, Silva A, and Azevedo V. "Pathogenesis of Corynebacterium diphtheriae and available vaccines; an overview". *Global Journal of Infectious Diseases and Clinical Research, **2017;** 3(1), 20-24.*
4. Jaiswal AK, 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". *International journal of molecular sciences, **2017;** 18 (2), 402.*
5. Tiwari S, **Jamal SB,** Hassan SS, Carvalho PVSD, Almeida S, Barh D et al. "Two-Component Signal Transduction Systems of Pathogenic Bacteria As Targets for Antimicrobial Therapy: An Overview". *Frontiers in Microbiology, **2017;** 8, 1878.*
6. Oliveira A, Oliveira LC, Aburjaile F, Benevides L, Tiwari S, **Jamal SB.** et al. "Insight of Genus Corynebacterium: Ascertaining the Role of Pathogenic and Non-pathogenic Species". *Frontiers in microbiology, **2017;** 8, 1937.*
7. Tiwari S, **Jamal SB,** Oliveira LC, Clermont D, Bizet C, Mariano D, et al. "Whole-genome sequence of Corynebacterium auriscanis strain CIP 106629 isolated from a dog with bilateral otitis from the United Kingdom". *Genome announcements, **2016;** 4 (4), e00683-16.*
8. Oliveira A, Teixeira P, Azevedo M, **Jamal SB,** Tiwari S, Almeida S, Silva A et al. "Corynebacterium pseudotuberculosis may be under anagenesis and biovar Equi forms biovar Ovis: a phylogenic inference from sequence and structural analysis". *BMC microbiology, **2016;** 16 (1), 100.*
9. Almeida A, Tiwari S, Mariano D, Souza F, **Jamal SB,** Coimbra N, Raittz RT, et al. "The genome anatomy of Corynebacterium pseudotuberculosis VD57 a highly virulent strain causing Caseous lymphadenitis". *Standards in genomic sciences, **2016;** 11 (1), 29.*
10. Irshad M, Abbasi MA, Siddiqui SZ, Ali MS, Ashraf M, Ismail T, Ahmad I, Hassan S, Lodhi MA, **Jamal SB.** "Synthesis, biological screening and molecular docking studies of some ethylated sulfonamides having 1,4-Benzodioxane moiety". *Pak. J. Pharm. Sci., **2016:** 29(6); 1913-1925.*
11. Radusky LG, Hassan SS, Lanzarotti E, Tiwari S, **Jamal SB,** Ali J, et al. "An integrated structural proteomics approach along the druggable genome of Corynebacterium

pseudotuberculosis species for putative druggable targets". *BMC Genomics, **2015;** 16 Suppl 5, S9. doi: 10.1186/1471-2164-16-S5-S9.*

12. Rehman AU, Khan MT, Wadood A, Khan Q, Ullah Z, Ayaz A, **Jamal SB,** Naeem M, Shah SB, Bilal JMA and Hamid R. "Exploring the binding pattern of predicted potent inhibitor to sirR_mtb; Modeling and structure-base virtual screening of sirR protein of Mycobacterium Tuberculosis (MTB)". *Int. J. Comput. Bioinf. In Silico Model, **2015;** 4 (3); 651-658.*
13. Folador EL, Oliveira A, Tiwari S, **Jamal SB,** Ferreira RS, Barh D, Ghosh P, Silva A, Azevedo V. "In Silico Protein-Protein Interactions: Avoiding Data and Method Biases Over Sensitivity and Specificity". *Current Protein and Peptide Science, **2015;** 16 (8), 689-700.*
14. Khalid S, Idrees S, Khalid H, Hussain B, Tiwari S, Hassan SS, Silva A, Azevedo V and **Jamal SB*.** Ab-Initio Prediction of Sequence and Structural Biology of Fish Muscle Proteins Using Homology Modeling, Phylogeny and Different Computational Approaches. *MOJ Proteomics Bioinform, **2015;** 2(3): 00047. DOI: 10.15406/mojpb.2015.02.00047.*
15. Tiwari S, **Jamal SB,** Carvalho PVSD, Hassan SS, Silva A, Azevedo V. "Industrial Microbiology & Biotechnology". February **2015.** *(Book Chapter)*
16. Tiwari S, Costa MP, Almeida S, Hassan SS, **Jamal SB,** Oliveira A, et al. "C. pseudotuberculosis Phop confers virulence and may be targeted by natural compounds". *Integrative Biology, **2014;** 6 (11), 1088-1099.*
17. Riaz S, Khan IU, Yar M, Ashraf M, Rehman TU, Shaukat A, **Jamal SB,** Duarte VCM, Alves MJ. "Novel pyridine-2, 4, 6-tricarbohydrazide derivatives: design, synthesis, characterization and in vitro biological evaluation as α-and β-glucosidase inhibitors". *Bioorganic chemistry, **2014;** 57, 148-154.*
18. Wadood A, Rehman AU, Shams S, Khan M, Rahman TU, **Jamal SB** et al. "Homology Modeling, Molecular Dynamic Simulation and Phylogenetic Analysis of Human Transient Receptor Potential Melastatin 1 (TRPM1)". *International Journal of Computational Bioinformatics and In Silico Modeling, **2014;** 3(3); 381-387.*
19. 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, **2014;** 15 Suppl 7, S3. doi: 10.1186/1471-2164-15-S7-S3.*
20. Irshad M, Abbasi MA, Rasool S, Siddiqui SZ, Ahmad I, Ashraf M, Lodhi MA, **Jamal SB.** "O-and N-Substituted derivatives of planetol as valuable bioactive compounds". *Asian Journal of Chemistry, **2014;** 26 (4), 1151.*
21. Wadood A, **Jamal SB,** Riaz M, Shah M, Mir A. "Computational analysis of Benzofuran-2-carboxlic acid as potent Pim-1 kinase inhibitor". *Pharmaceutical biology, **2014;** 52 (9), 1170-1178.*
22. Alrawi BJM, Rehman AU, Ahmad A, Mohammed A, **Jamal SB,** Ahmad N, Wadood A, Shah L. "Prediction of Binding Mode of Bisphenol-A (A Carcinogen) in Estrogen and Testosterone Receptors by Applying Computational Docking Approach". *J. Appl. Environ. Biol. Sci, **2014;** 4(8S); 101-105.*
23. Ashraf M, Ejaz SA, Lodhi MA, **Jamal SB.** "Synthesis, Characterization and Biological Screening of Some New Sulfonamides Derivatives of 1, 4-Benzodioxane-6-Amine". *J. Chem. Soc. Pak, **2014;** 36 (4), 660.*

24. Wadood A, Ghufran M, **Jamal SB,** Naeem M, Khan A, Ghaffar R, and Asnad. "Phytochemical Analysis of Medicinal Plants Occurring in Local Area of Mardan". *Biochem Anal Biochem, **2013;** 2: 144. doi: 10.4172/2161-1009.1000144*
25. Wadood A, Riaz M, **Jamal SB,** Shah M, Lodhi MA. "Molecular Docking study of p4 Benzoxaborol ligands as inhibitor of HCV NS3/4A Protease". *Bioinformation. March 19 **2013;** 9(6): 309-314.*
26. Sehgal SA, Waqar M, **Jamal SB***. "Haplotype analysis for DFNB2/USH1B locus in consanguineous family of Punjab origin (Pakistan)". *J. Med. & Allied Health Sciences **2013;** 2(1): 09-15.*
27. Wadood A, Riaz M, **Jamal SB,** Shah M. "Interactions of ketoamide inhibitors on HCV NS3/4A protease target: Molecular Docking Studies". *Molecular biology reports, **2013;** 41 (1), 337-345.*
28. Wadood A, Ali S, Riaz M, **Jamal SB,** Rehman TU, Gul S, Hassan M. "Molecular Docking Study of 5-substituted-8-methyl-2H-pyrido [1, 2-a] pyrimidine-2, 4 (3H) – diones As Inhibitors of Basilluspasteurii urease". *Pakhtunkhwa Journal of Life Science, **2013;** 1(3); 145-153.*
29. **Jamal SB,** Anwar Z, Khan RU, Jamil T, Iram A, Mir A, Waqar M, Khattak JZK. "Haplotype Analysis for DFNB4/PDS Locus in Hearing Impaired Families of Punjab (Pakistan)". *Curr. Res. Journal of Bio. Sci. **2012;** 10; 4(4): 407-413.*

***Google Scholar*:** https://scholar.google.com.br/citations?user=uMHP-n0AAAAJ&hl=en

Articles in Communication:

1. Sandeep Tiwari†, **Syed Babar Jamal†,** Alberto Oliveira, Rodrigo D. De Oliveira Carvalho, Marta Giovanetti, Álvaro Salgado, Thiago Luiz de Paula Castro, Debmalya Barh, Nuno Rodrigues Faria, Vasco Azevedo*, Luiz Carlos Junior Alcantara*. NS5 of virulent ZIKA virus could potentially be targeted by anti-flavivirus natural compounds: An in-silico study. *PloS One*. (Submitted)
2. Sandeep Tiwari[1,2¥], Debmalya Barh[1,2¥*], Ranjith K. Kumavath[3] , S. Prabu Seenivasan[4], **Syed Babar Jamal[2]**, Vanaja Kumar[5], Preetam Ghosh[6], Vasco Azevedo[2]. Ack is essential for microbial growth and betel compounds potentially target Ack, PhoP and MDR proteins in *M. tuberculosis*, *V. cholera*, and pathogenic *E. coli. Current topics in Medicinal Chemistry*. (Under Review)
3. Edson Luiz Folador, Sandeep Tiwari, Camila Eduarda da Paz Barbosa, **Syed Babar Jamal,** Marco da Costa Schulze, Debmalya Barh, Vasco Azevedo*. Protein-Protein Interactions: An Overveiw. *Encyclopedia Of Bioinformatics and Computational Biology***.** **(**Book Chapter). (Under Review)
4. Tanzeel Ur Rehman; Islam Uallah; Sadaf Riaz; **Syed Babar Jamal.** Novel N-(Substituted benzylidene)-4-phenyl-6-(pyridin-3-yl)pyrimidin-2-amine derivatives: Design, synthesis, characterization and in vitro biological evaluation as α, β-glucosidase and Lipoxygenase inhibitors. *Medicinal Chemistry Research.* (Under Review)
5. Marcus VV, Jaiswal AK, **Jamal SB,** Tiwari S, Azevedo V, Soares SC. An in silico drug and vaccine target candidate identification in Mycobacterium leprae and Mycobacterium lepromatisis: A bioinformatics based approach. *Gene* (Submitted)

6. Siddiqui SZ, Abbasi MA, Rehman A, Irshad M, Ashraf M, Nasim FH, Nasar R, Arshad H, Lodhi MA and **Jamal SB**. Synthesis, spectral analysis and pharmacological screening of S-substituted 5-[(4-chlorophenoxy)methyl]-1,3,4-oxadiazole-2-thiol derivatives. *Arab. J. Chem* (Submitted).
7. Jaiswal AK, Tiwari S, **Jamal SB,** Azevedo V, Soares SC. The Pan-Genome of *Treponema pallidum* Reveals Differences in Genome Plasticity between the subspecies. *BMC Genomics* (Submitted)

Posters Presentation:

1. **Syed Babar Jamal.** Proteome scale comparative modeling for conserved drug and vaccine targets identification in *Salmonella serovers*. ***X-Meeting 2017*** - 13th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (AB3C), held in São Pedro - Brazil between 4th and 6th October of 2017.
2. Arun Kumar Jaiswal, **Syed Babar Jamal** and co. The Pan-Genome of *Treponema pallidum* Reveals Differences in Genome Plasticity between the subspecies. ***X-Meeting 2017*** - 13th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (AB3C), held in São Pedro - Brazil between 4th and 6th October of 2017.
3. Rodrigo Profeta, **Syed Babar Jamal** and co. Functional Analysis of hypothetical proteins unveils putative metabolic pathways, essential genes and therapeutic drug and vaccine target in *Trypanosma cruzi*: A bioinformatics based approach. ***X-Meeting 2017*** - 13th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (AB3C), held in São Pedro - Brazil between 4th and 6th October of 2017.
4. **Syed Babar Jamal** and co. In-silico analyses for the discovery of drug and vaccine targets in *Corynebacterium camporealensis:* A Novel Hierarchical Approach. ***X-Meeting 2016*** - 12th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (AB3C), held in Belo Horizonte - Brazil between November 16 and 18 of 2016.
5. Arun Kumar Jaiswal, **Syed Babar Jamal** and co. In silico Identification of common putative vaccine candidates against *Treponema pallidum:* A reverse vaccinology based approach. ***X-Meeting 2016*** -12th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (AB3C), held in Belo Horizonte - Brazil between November 16 and 18 of 2016.
6. **Syed Babar Jamal** and co. In silico approaches to predict the impact of leukemic DNMT3a mutations & Identification of leads based on drug decitabine using Complex Based Pharmacophore Mapping and Virtual Screening. ***X-Meeting 2016*** -12th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (AB3C), held in Belo Horizonte - Brazil between November 16 and 18 of 2016.
7. Doglas Parise, **Syed Babar Jamal** and co. An integrated structural proteomics approach along the druggable genome of *Corynebacterium pseudotuberculosis* species for putative druggable targets. **2nd Symposium of Microbiology** at Institute of Biological Science, Federal University of Minas Gerais, Belo Horizonte - Brazil between 5th and 6th of October 2015.

8. Sandeep Tiwari, **Syed Babar Jamal** and co. Whole-genome sequence of *Corynebacterium auriscanis* strain CIP106629 isolated from a dog with bilateral otitis from the United Kingdom. **2nd Symposium of Microbiology** at Institute of Biological Science, Federal University of Minas Gerais, Belo Horizonte - Brazil between 5th and 6th of October 2015.
9. Syed Shah Hassan, **Syed Babar Jamal** and co The Druggable Pocketome of *Corynebacterium diphtheriae*: A new approach for *in-silico* putative druggable targets. **2nd Symposium of Microbiology** at Institute of Biological Science, Federal University of Minas Gerais, Belo Horizonte - Brazil between 5th and 6th of October 2015.
10. **Syed Babar Jamal** and co. An integrative in-silico approach for therapeutic target identification in the human pathogen *Corynebacterium diphtheria*. ***X-Meeting 2015*** - 11th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (AB3C), held in São Paulo - Brazil between 3rd and 6th November of 2015.
11. Edgar Aguiar, **Syed Babar Jamal** and co. *In silico* analysis drug and vaccine target identification using subtractive genomics against *Streptococcus agalactiae*, strain GBS85147. ***X-Meeting 2015*** - 11th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (AB3C), held in São Paulo - Brazil between 3rd and 6th November of 2015.
12. Sandeep Tiwari, **Syed Babar Jamal** and co. *In-silico* analyses for the discovery of drug and vaccine targets in *Burkholderia cepacia:* A Novel Hierarchical Approach. ***X-Meeting 2015*** - 11th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (AB3C), held in São Paulo - Brazil between 3rd and 6th November of 2015.
13. **Syed Babar Jamal** and co. Comparative genomics and *in-silco* subtractive genomics approach for targeting multidrug resistant *Neisseria gonorrhoeae* with medicinal-plant-derived antimicrobial compounds. ***X-Meeting 2014*** - 10th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (AB3C), held at Institute of Biological Science, Federal University of Minas Gerais, Belo Horizonte - Brazil between 28th and 31st October of 2014.
14. Sandeep Tiwari, **Syed Babar Jamal** and co. Metabolic advancement of multidrug resistant E.coli: A Comparative Metabolomics analysis of Multidrug resistant and Pathogenic Isolates *E. coli* Strain with Commensal *E. coli* Strain. ***X-Meeting 2014*** - 10th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (AB3C), held at Institute of Biological Science, Federal University of Minas Gerais, Belo Horizonte - Brazil between 28th and 31st October of 2014.

Languages:

- English.
- Portuguese
- Urdu
- Punjabi
- Pushto

Curriculum lettes: http://lattes.cnpq.br/3196781096279613
Skype: babarjamalbacha

References

Supervisor:

Prof. Dr. Vasco Ariston de Carvalho Azevedo, DVM, M.Sc, Ph.D
Coordinator Post-Graduate Program in Bioinformatics
Institute of Biological Science
Federal University of Minas Gerais
Av. Antonio Carlos 6627, Pampulha, Belo Horizonte
Minas Gerais, Brazil CO 486, CEP 31270-901
E-mail: vascoariston@gmail.com ; vasco@icb.ufmg.br
CV: http://lattes.cnpq.br/1020477751003832

Co-supervisor:

Prof. Dr. Artur Luiz da Costa da Silva, M.Sc, Ph.D
Institute of Biological Science
Federal University of Para
Rua Marquês de São Vicente, 225
Gávea - Belém - PA - 22453-900 – Brasil
E-mail: asilva@ufpa.br
CV: http://lattes.cnpq.br/7642043789034070

Co-supervisor:

Dr. Sandeep Tiwari, M.Sc, PhD
Postdoctoral Research Fellow
Laboratory of Cellular and Molecular Genetics (LGCM)
Institute of Biological Science
Federal University of Minas Gerais
Av. Antonio Carlos 6627, Pampulha, Belo Horizonte
Minas Gerais, Brazil CO 486, CEP 31270-901
E-mail: sandip_sbtbi@yahoo.com
CV: http://lattes.cnpq.br/2419194183390541

V. Bibliography

(Preface)

Altman, R.B. (2000). Biomedical computation at Stanford University: a larger umbrella for the future. *MD computing: computers in medical practice* 17: 35-37.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* 25: 3389-3402.

Baker, D., and Sali, A. (2001). Protein structure prediction and structural genomics. *Science* 294: 93-96.

Baumbach, J., and Apeltsin, L. (2008). Linking Cytoscape and the corynebacterial reference database CoryneRegNet. *BMC Genomics* 9**:** 184. doi: 10.1186/1471-2164-9-184.

Bowie, J.U., Luthy, R., and Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253: 164-170.

Capriles, P.V., Guimaraes, A.C., Otto, T.D., Miranda, A.B., Dardenne, L.E., and Degrave, W.M. (2010). 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* 11: 610.

Cerdeño-Táarraga, A.M. et al. (2003). The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. *Nucleic Acids Res* 31: 6516–6523.

Colman, P.M. (1994). Structured based drug design. *Curr. Op. Struct. Biol* 4: 868-874.

Daniel, S., and Bert, L.G. (2010). Ligand docking and binding site analysis with PyMol and Autodock/Vina. *Journal of Computer-Aided Molecular Design* 24: 417-422.

Dittmann, S. et al., (2000). Successful control of epidemic diphtheria in the states of the Former Union of Soviet Socialist Republics: lessons learned. *J. Infect. Dis* 181: S10–S22.

Fischer, D., and Eisenberg, D. (1997). Assigning folds to the proteins encoded by the genome of Mycoplasma genitalium. *Proceedings of the National Academy of Sciences of the United States of America* 94: 11929-11934.

Gallop, M.A., Barrett, R.W., Dower, W.J., Foder, S.P., and Gordon, E.M. (1994) Applications of combinatotial technologies to drug discovery. 1. Background and peptride combinatorial libraries. *J Med Chem* 37: 1233-1251.

Gordon, E.M., Barrett, R.W., Dower, W.J., Foder, S.P., and Gallop, M.A. (1994) Applications of combinatorial technologies to drug discovery. 2. Combinatorial organic synthesis; library screening strategies and future directions. *J Med Chem* 37: 1385-1401.

Grabenstein, J.D. (2010). Vaccines: A Biography (Toxoid Vaccines). New York, NY: *Springer Science & Business Media* 105–24.

Guex, N., Diemand, A., and Peitsch, M.C. (1999). Protein modelling for all. *Trends in biochemical sciences* 24: 364-367.

Halgren, T.A., Murphy, R.B., Friesner, R.A., Beard, H.S., Frye, L.L., Pollard, W.T., and Banks J.L. (2004). Glide: a new approach for rapid, accurate Docking and scoring. 2. Enrichment factors in database Screening. *J. Med, Chem* 47: 1750-1759.

Hassan, S.S., Tiwari, S., Guimaraes, L.C., Jamal, S.B., Folador, E., Sharma, N.B., et al. (2014). Proteome scale comparative modeling for conserved drug and vaccine targets identification in Corynebacterium pseudotuberculosis. *BMC Genomics* 15: S3. doi: 10.1186/1471-2164-15-S7-S3

Hodes, H.L. (1979). Diphtheria. *Pediatr Clin Nth Am* 26:445e59.

Jain, A.N., (2003). Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine, *J. Med. Chem* 46: 499-511.

Jain, A.N., (2004). Virtual screening in lead discovery and optimization. *Curr Opin Chem* 7: 396-493.

Jauch, R., Yeo, H.C., Kolatkar, P.R., and Clarke, N.D. (2007). Assessment of CASP7 structure predictions for template free targets. *Proteins* 69 Suppl 8: 57-67.

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). A new approach to protein fold recognition. *Nature* 358: 86-89

Jones, G., and Willet, P. (1995). Docking small molecules ligands into active site. *Curr. Op. Biotech* 6: 652-656.

Jamal, S.B., Hassan, S.S., Tiwari, S., Viana, M.V., Benevides, L.J., Ullah, A., et al. (2017). An integrative in-silico approach for therapeutic target identification in the human pathogen *Corynebacterium diphtheriae*. *PLoS One* 12**:** e0186401. doi: 10.1371/journal.pone.0186401.

Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, O.C., and Vakser, I.A. (1992). In Molecular surface recognition: determination Geometric fit between proteins and their ligands by correlation Techniques. Proceedings of the National Academy of Science USA, *National Acad Sciences* 23: 2195-2199.

Kumar, N., Hendrix, B.S., Janes, K.A., de Graff, D., and Laufenburger, D.A. (2006). Applying computational modeling to drug discovery and development, *Drug. Discov. Today* 11: 806-811.

Kuntz, I.D. (1992). Structure based stratigies for drug design and discovery. *Science* 257: 1078-1082.

Kuntz, I.D., Meng, E.C., and Shoiche, B.K. (1994). Receptor-based molecular design. *Acc. Chem, Res* 27: 117-123.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of biomolecular NMR* 8: 477-486.

Lenguer, T., and Rarey, M. (1996). Computational methods for biomolecular docking. *Curr. Op. Struct. Biol* 6: 402-406.

Liwo, A., Lee, J., Ripoll, D.R., Pillardy, J., and Scheraga, H.A. (1999). Protein structure prediction by global optimization of a potential energy function. *Proceedings of the National Academy of Sciences of the United States of America* 96: 5482-5485.

Lopez, G., Rojas, A., Tress, M., and Valencia, A. (2007). Assessment of predictions submitted for the CASP7 function prediction category. Proteins 69 Suppl 8, 165-174.

Lybrand, T.P. (1995). Ligand –protien docking and rational drug design. *Curr. Op. Struct. Biol* 5: 224-228.

Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual review of biophysics and biomolecular structure* 29: 291-325.

MOE (The Molecular Operating Environment) Version 2005.06., (2010) Chemical Computing Group Inc, Available from: http://www.chemcomp.com.

Moris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., and Olson, A.J. (1998). Automated docking using a Lamarkian Genetic algorithm and an empirical binding free energy function. *J. Comut. Chem,* 19: 1639-1662.

Myers, P.L. (1997). Combinatorial chemistry deliver real medicine. *Curr Opin Biotechnol*, 8: 701-707.

Oprea, T.I., and Matter, H. (2004). Integrating virtual screening in lead discovery. *Curr. Opin. Chem. Biol*, 8: 349-358.

Pieper, U., Webb, B.M., Barkan, D.T., Schneidman-Duhovny, D., Schlessinger, A., Braberg, H., Yang, Z., Meng, E.C., Pettersen, E.F., Huang, C.C., et al. (2011). ModBase, a database of annotated comparative protein structure models, and associated resources. *Nucleic acids research* 39: D465-474.

Rarey, M., Kramer, B., Languer, T., and Klebe, G.A. (1996). Fast flexible docking method using an incremental construction algorithm, *J. Mol. Biol* 261: 470-489.

Roche, O., and Guba, W. (2005) Computational chemistry as an integral component of lead generation. *Mini. Rev. Med. Chem*, 5: 677-683.

Rosenfeld, R., Vajda, S., and De Lisi, C. (1995). Flexible docking and design. *Ann,Rev. Biophys. Biomol. Struct* 24: 677-700.

Rössle, S.C.S. (2004). Development of a computational system for comparative modeling in structural genomics: Analysis of the genome sequence of Gluconacetobacter

diazotrophicus. Tese de Doutorado, Instituto de Biofísica Carlos Chagas Filho- Universidade Federal do Rio de Janeiro/Brasil.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* 5: 725-738.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology* 234, 779-815.

Sanchez, R., and Sali, A. (1997). Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins Suppl* 1: 50-58.

Sanchez, R., and Sali, A. (1998). Large-scale protein structure modeling of the Saccharomyces cerevisiae genome. *Proceedings of the National Academy of Sciences of the United States of America* 95: 13597-13602.

Sanchez, R., Pieper, U., Melo, F., Eswar, N., Marti-Renom, M.A., Madhusudhan, M.S., Mirkovic, N., and Sali, A. (2000a). Protein structure modeling for structural genomics. *Nature structural biology* 7: 986-990.

Sanchez, R., Pieper, U., Mirkovic, N., de Bakker, P.I., Wittenstein, E., and Sali, A. (2000b). MODBASE, a database of annotated comparative protein structure models. *Nucleic acids research* 28: 250-253.

Simons, K.T., Kooperberg, C., Huang, E., and Baker, D. (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *Journal of molecular biology* 268: 209-225.

Stahl, M., Guba, W., and Kansy, M. (2006). Integrating molecular design resources within modern drug discovery research: the Roche experience. *Drug. Discov. Today*, 11: 326-333.

Stoermer, M.J. (2006) Current status of virtual scereening as analyzed by target class. *J. Med. Chem* 2: 89-112.

UniProt, C. (2009). The Universal Protein Resource (UniProt). *Nucleic acids research* 37: 169-174.

Vesolovoski, A.V., Ivinov, A.S. (2003). Strategy of computer aided drug design. *Curr Drug Target Infect Disord* 3: 33-40.

Vitek, C.R. (2006). Diphtheria. *Curr. Top. Microbiol. Immunol*. 304: 71–94.

Westbrook, J., Feng, Z., Jain, S., Bhat, T.N., Thanki, N., Ravichandran, V., Gilliland, G.L., Bluhm, W., Weissig, H., Greer, D.S., et al. (2002). The Protein Data Bank: unifying the archive. *Nucleic acids research* 30: 245-248.

World Health Organization. WHO vaccine-preventable diseases: monitoring system: 2016 global summary [Internet]. Geneva: World Health Organization; 2016 [cited 2016 Nov 28]. Available from: http://apps.who.int/immunization_monitoring/globalsummary

Wu, S., Skolnick, J., and Zhang, Y. (2007). Ab initio modeling of small proteins by iterative TASSER simulations. *BMC biology* 5: 17.

Xavier, E.R., Capanema, B.P., Ruiz, J.C., Oliveira, G., Meyer, R., D'Afonseca, V., Miyoshi, A., and Azevedo, V. (2008). Brazilian genome sequencing projects: state of the art. *Recent patents on DNA & gene sequences* 2: 111-132.

Zhang, Y. (2008). Progress and challenges in protein structure prediction. *Current opinion in structural biology* 18: 342-348.