Computational Biology and Chemistry 48 (2014) 55–63

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Research Article

# Identification and characterization of potential drug targets by subtractive genome analyses of methicillin resistant *Staphylococcus aureus*

Reaz Uddin*, Kiran Saeed

*Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi 75270, Pakistan*

ARTICLE INFO

*Article history:*
Received 29 June 2013
Received in revised form
26 November 2013
Accepted 28 November 2013

*Keywords:*
*S. aureus*
Subtractive genome
Proteome
Methicillin resistant
Druggability

ABSTRACT

Methicillin resistant *Staphylococcus aureus* (MRSA) causes serious infections in humans and becomes resistant to a number of antibiotics. Due to the emergence of antibiotic resistance strains, there is an essential need to develop novel drug targets to address the challenge of multidrug-resistant bacteria. In current study, the idea was to utilize the available genome or proteome in a subtractive genome analyses protocol to identify drug targets within two of the MRSA types, i.e., MRSA ST398 and MRSA 252. Recently, the use of subtractive genomic approaches helped in the identification and characterization of novel drug targets of a number of pathogens. Our protocol involved a similarity search between pathogen and host, essentiality study using the database of essential genes, metabolic functional association study using Kyoto Encyclopedia of Genes and Genomes database (KEGG), cellular membrane localization analysis and Drug Bank database. Functional family characterizations of the identified non homologous hypothetical essential proteins were done by SVMProt server. Druggability potential of each of the identified drug targets was also evaluated by Drug Bank database. Moreover, metabolic pathway analysis of the identified druggable essential proteins with KEGG revealed that the identified proteins are participating in unique and essential metabolic pathways amongst MRSA strains.

In short, the complete proteome analyses by the use of advanced computational tools, databases and servers resulted in identification and characterization of few nonhomologous/hypothetical and essential proteins which are not homologous to the host genome. Therefore, these non-homologous essential targets ensure the survival of the pathogen and hence can be targeted for drug discovery.



## 1. Introduction

Human health is in constant threats posed by the emergence of resistant strains of various infectious agents. The challenge turns more alarming due to the fact that majority of infectious agents are now resistant to more than one drugs. *Staphylococcus aureus* (*S. aureus*) is one among other multiple drug resistant (MDR) bacterial pathogens. *S. aureus* causes both nosocomial and community acquired infection (Kuroda et al., 2001). Pathogenicity of infection may vary from mild skin infection to chronic fatal necrotizing pneumonia. Treatment may include $\beta$-lactam antibiotics, sulfa drugs, clindamycin and tetracycline (MP, 1961; Projan and Novick, 1997; Lowy, 1998; Cheung, 2002; DeLeo et al., 2010). Lately, *Staphylococcus aureus* acquired resistance to several antibiotics, including powerful modern penicillin's (such as methicillin), resulted in emergence of subtype known as methicillin resistant *Staphylococcus aureus* (MRSA) (Kuroda et al., 2001; Kuehnert et al., 2006; Chambers and DeLeo, 2009; DeLeo et al., 2010). Worldwide rate of morbidity and subsequent mortality with MRSA infection has shown a marked increase since it was first reported, back in 1961 (MP, 1961; Klevens et al., 2007; Feil et al., 2008; Gotuzzo, 2010; de Kraker et al., 2011).

Two types, amongst others, i.e. MRSA ST398 and MRSA 252 are well known and reported to have roles in mild to moderate serious infection in human. It was reported that MRSA ST398 clonal type has role in livestock-associated MRSA (LA-MRSA) while MRSA 252 causes serious hospital acquired infections in humans (Herold et al., 1998; Kock et al., 2009; Van Duijkeren et al., 2009).

Newer antibiotics and other regimens were proposed during last decade but with their anticipated adverse effects (Ippolito et al., 2010). The usual treatment regimen is depending upon older concept of therapy with antibiotics, which include glycopeptides, lincosamide and sulfa drugs (Nicolle, 2006; Siegel et al., 2007). Among various available antibiotics, ditomycin, moxifloxacin

* Corresponding author.
*E-mail address:* mriazuddin@iccs.edu (R. Uddin).

1476-9271/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiolchem.2013.11.005

HCL, trimethoprim-sulfamethoxazole (Moellering; Tally), linezolids were well thought out and relatively more effective treatment. However, the treatment deteriorates due to fast growth of resistant strains and hence, the situation is alarming for health practitioners (Nicolle, 2006; Gorchynski and Rose, 2008).

Discovery of novel drug candidates is not an easy task in early stages and it is the same time expensive and laborious. The situation poses challenge for the researchers to come up with the alternative and inexpensive methods for proposing new drug candidates against deadly infections such as MRSA (Hellerstein, 2008). A useful method to eradicate the infections is the identification of unique drug targets amongst the pathogens, particularly for the resistant ones. Information about essentiality of genes and its products are considered as potential drug targets for the discovery of antibiotics. Gene knockouts, RNA interference and conditional knockouts are the available experimental methods for finding gene essentiality, however, labor intensive, time consuming and costly. Complete genomes of 1000 pathogenic bacteria were made available from National Center for Biotechnology Information (NCBI) genome database. It is unlike of very few pathogenic-bacteria-gene-essentiality data from experimental methods which is only 20, so far (Galperin and Koonin, 1999; Butt et al., 2012a,b). This huge available genomic information needs to be dealt with efficient computational algorithm. Available technology for the analyses of genomic data has revealed the characteristics of essential genes which include following: higher rate of evolutionary conservation, strand-bias, patterns of protein-interaction networks, high expressivity, codon usage, GC content, length of proteins, and subcellular localization (Rocha and Danchin, 2003; Gustafson et al., 2006; Saha and Heber, 2006; Deng et al., 2011). Recent advances in computational biology have enabled the identification and characterization of essential protein targets sites more rapidly than the conventional methods. Simultaneous advances in the areas of genomic and/or proteomic allowed functional predictions of possible drug targets with much accuracy. Combining of these two concepts, furnishes a new approach for identification of novel drug targets using in silico bioinformatics tools and databases, and have revolutionized the drug discovery process.

Current study is an application of relatively newer approach of subtractive genome analyses, applied to identify essential genes for the survival of pathogen and at the same time non-homologous to the human genome (Khalida et al., 2012). These non homologous essential genes, therefore can serve as potential drug targets for the future drug development phase. Various literature reported the applicability of successful predictions using this approach (Allsop et al., 1995; Sakharkar et al., 2004; Anishetty et al., 2005; Dutta et al., 2006; Perumal et al., 2007; Singh et al., 2007; Sharma et al., 2008; Barh and Kumar, 2009; Rathi et al., 2009; Amineni et al., 2010; Gupta et al., 2010; Abadio et al., 2011; Butt et al., 2011, 2012a,b; Reddy et al., 2011). Search of new drug candidates against such therapeutic targets are then assumed to minimize the pathogen's ability to resist against available treatments which may lead to decrease the side effects. Although few efforts on similar strategy and with MRSA have been reported in literature (Haag et al., 2011; Hossain et al., 2013) but here we presenting the strategy applied on two MRSA types simultaneously, and with a validation method which was lacking in earlier reports.

## 2. Materials and methods

The standalone release of NCBI BLAST+ version 2.2.26 was obtained from NCBI (Altschul et al., 1990) (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/) and installed on a Linux workstation with Intel Xeon quad core processor. The overall workflow of current study is presented in Fig. 1.

### 2.1. Complete proteome retrieval

ExPASy (Expert Protein Analysis System) proteomics server contains the HAMAP (Gasteiger et al., 2003) which provides complete proteome and annotated protein records in UniProtKB format. The complete proteomes of MRSA ST398, MRSA 252 and *H. sapiens* were retrieved from the HAMAP (High-quality Automated and Manual Annotation of Proteins) on September 4, 2012.

### 2.2. Determining non paralogous sequences

Paralogous or duplicate protein sequences were identified by using CD-HIT (Li et al., 2001) with sequence identity cut off of 0.8 (80%). Paralogous sequences were screened out from complete proteome of both types resulted in non-paralogous sequences only.

### 2.3. Determination of non homologous protein sequences to the human proteome

For this study, the non-paralogous sequences were subjected to BLASTp against homo sapiens (Gasteiger et al., 2003) using threshold expectation value (*E*-value $10^{-3}$) (Haag et al., 2011; Kerfeld and Scott, 2011). The resultant sequences were consisted of homologous sequences (significant similarity with Human host) and non-homologous sequences (no hit found). Sequences which showed significant similarity with the human host were removed leaving only the non homologous sequences for subsequent analysis.

### 2.4. Identification of non homologous essential genes in MRSA ST398 and MRSA 252

Essential genes are the ones which need to support basic cellular functions of micro-organisms and therefore, essential for survival of the pathogen. Database of essential gene (DEG) (Zhang et al., 2004; Zhang and Lin, 2009) containing essential genes and their expressed proteins curated from literature along with experimental methods and analysis report from 20-Gram positive and Gram-negative bacteria (Table S5). The DEG version 6.8 was downloaded from the DEG website (http://www.essentialgene.org/). Non homologous protein sequences were then subjected to BLASTp against DEG with an expectation value (*E*-value $10^{-5}$). Resultant sequences represented both putative and hypothetical non homologous essential proteins among both types (i.e. MRSA ST398 and MRSA 252).

### 2.5. KEGG metabolic pathway analyses

KEGG provides a biological system of molecular interaction network along with complete functional annotation (Kanehisa et al., 2012). To identify metabolic pathway using KEGG, a popular sever KAAS (KEGG Automated Annotation Server) (Moriya et al., 2007) was used which performs a BLASTp similarity searches of all non homologous essential protein against periodically updated KEGG database. KAAS outputs not only metabolic pathways but also provide different features of information such as KO list assignment and alternative pathways along with enzymes and Enzyme Commission (EC) numbers. The results helped to predict metabolic pathways of potential drug targets out of the subjected protein sequences.

### 2.6. Prediction of subcellular localization

PSORTb v.3.0. is the most popular computational tool for predicting the localization of unknown proteins. All non homologous essential proteins were subjected to the prediction of subcellular