Infection, Genetics and Evolution xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Identification of putative vaccine candidates against *Helicobacter pylori* exploiting exoproteome and secretome: A reverse vaccinology based approach

Anam Naz [a], Faryal Mehwish Awan [a], Ayesha Obaid [a], Syed Aun Muhammad [b], Rehan Zafar Paracha [a], Jamil Ahmad [c], Amjad Ali [a,*]

[a] *Computational Biology and Genomics (CBG) Research Group, Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), H-12 Islamabad, Pakistan*
[b] *Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University, Multan, Pakistan*
[c] *Computational Biology and Genomics (CBG) Research Group, Research Center for Modelling and Simulation (RCMS), National University of Sciences and Technology (NUST), H-12 Islamabad, Pakistan*

ARTICLE INFO

Article history:
Received 27 October 2014
Received in revised form 19 March 2015
Accepted 23 March 2015
Available online xxxx

Keywords:
Vaccine candidates
Reverse vaccinology
H. pylori
Antigenic epitopes
Vaccine development

ABSTRACT

Helicobacter pylori (*H. pylori*) is an important pathogen associated with diverse gastric disorders ranging from peptic ulcer to malignancy. It has also been recognized by the World Health Organization (WHO) as class I carcinogen. Conventional treatment regimens for *H. pylori* seem to be ineffective, possibly due to antibiotic resistance mechanisms acquired by the pathogen. In this study we have successfully employed a reverse vaccinology approach to predict the potential vaccine candidates against *H. pylori*. The predicted potential vaccine candidates include vacA, babA, sabA, fecA and omp16. Host–pathogen interactions analysis elaborated their direct or indirect role in the specific signaling pathways including epithelial cell polarity, metabolism, secretion system and transport. Furthermore, surface-exposed antigenic epitopes were predicted and analyzed for conservation among 39 complete genomes of *H. pylori* (Genbank) for all the candidate proteins. These epitopes may serve as a base for the development of broad spectrum peptide or multi-component vaccines against *H. pylori*. We also believe that the proposed pipeline can be extended to other pathogens and for the identification of novel candidates for the development of effective vaccines.

1. Introduction

Helicobacter pylori is one of the major etiological agent involved in the development of gastritis, peptic ulcer, and gastric adenocarcinoma. The gram-negative and urease positive bacterium affects about 50% of the human population worldwide (McColl, 2010). The majority of patients with *H. pylori* infection do not face any significant clinical complications however, extensive epidemiologic data suggests strong association of *H. pylori* infection with non-cardiac gastric cancers (Cavaleiro-Pinto et al., 2011). In particular, toxins cagA (the product of cytotoxin-associated gene A) and vacA (the product of vacuolating cytotoxin-associated gene A) trigger several cancer associated signaling pathways (Ruggiero and Censini, 2014). The exact mechanism by which it causes cancer is still controversial, though several studies suggest significant association of the pathogen with squamous cell laryngeal cancer and squamous cell cancer of the upper aero-digestive tract (excluding the esophagus) (Genç et al., 2013). Recent studies indicate that the current first-line treatment regime for *H. pylori* infection (clarithromycin, amoxicillin and a proton pump inhibitor) is facing a challenging clinical situation in terms of evolving antibiotics resistance, serious side effects, risk of re-infection and the high cost of antibiotic therapy (Reid et al., 2001).

The classical vaccine development approach introduced by Louis Pasteur (Serruto and Rappuoli, 2006) led to successful discovery of an effective vaccine against several pathogens. However there are several challenges in the development of successful vaccines which include the longer duration, culturing difficulty, inaccurate and variable products, insufficient attenuation, less immunogenicity, hypersensitivity and expensive procedures (Bambini and Rappuoli, 2009). Since the most abundant proteins

* Corresponding author. Tel.: +92 51 90856138 (O), +92 333 9191903 (cell).
E-mail addresses: anam.naz88@live.com (A. Naz), faryal_mehwish@yahoo.com (F.M. Awan), ayesha_obaid_nust@live.com (A. Obaid), aunmuhammad78@yahoo.com (S.A. Muhammad), rehanzfr@gmail.com (R.Z. Paracha), dr.ahmad.jamil@gmail.com (J. Ahmad), amjad_uni@yahoo.com (A. Ali).

http://dx.doi.org/10.1016/j.meegid.2015.03.027
1567-1348/

are often not suitable vaccine candidates and the genetic tools required to identify the rare components are inadequate or even not available, classical vaccinology presents itself with several limitations. In contrast to classical vaccinology, the reverse vaccinology is the most practiced methodology since the discovery of a universal vaccine against serogroup B meningococcal (menB) disease in 2000 (Giuliani et al., 2006). Reverse vaccinology involves the use of several *in silico* steps to identify immunogenic antigens from the sequenced genomes (proteomes) of a pathogen. It is also important to note that these *in silico* steps/methods provide a comprehensive view of the pathogen genome, essential pathways, virulence determining factors and protein–protein interactions (PPI) among itself and the host proteome.

Prioritization of potential vaccine candidates require the identification of virulent proteins which are capable of stimulating immune response in the host organism (Mora et al., 2006). There are some known reported features of effective vaccine candidate proteins which include: sub-cellular localization, presence of signal peptides, transmembrane domains and antigenic epitopes (Doytchinova and Flower, 2007; Vivona et al., 2006). The basic principle behind the identification of potential vaccine targets is the prediction of antigenic and virulence factors among a pathogen and subsequently find those sequences which are likely to bind MHC class I and II proteins for antigen presentation within the host. Most of the reported vaccine targets include surface exposed (exomembrane) or secreted proteins which are more likely to be antigenic or virulent and thus are considered as suitable vaccine candidates (Shanmugham and Pan, 2013). Several protective antigens have already been identified against *H. pylori* using urease, cytotoxin vacA, heat shock proteins (HspA and HspB) and catalases (Michetti, 1997) however, this pathogen has adapted many ways to evade the immune response and previous efforts to develop a vaccine have not been fruitful. Previous approaches have not taken into account the host–pathogen interaction efficiently which may be the major reason behind unsuccessful attempts, as the bacterium has co-evolved with humans for thousands of years and is adapted to host environment making the vaccine development more challenging.

The current study is based on the development of a comprehensive computational framework/pipeline for the identification of putative vaccine targets against *H. pylori*. It is based on subtractive proteomics to prioritize novel antigenic epitopes by integrating data from biological networks and databases. Essential proteins (either surface or secreted) were first identified, followed by confirming their role in virulence and immune evoking pathways. Subsequently, non-human homologous proteins were scanned for those epitopes having the ability to bind with both B-cells and T-cells. The genetically invariable epitopes identified here can be further subjected to *in vivo* testing. Following the proposed pipeline, candidate antigens can easily be identified which will assist in the development of protective immunogens against various other pathogens (Giuliani et al., 2006; Leuzzi et al., 2006).

2. Materials and methods

2.1. Targeting H. pylori

The efficacy of prophylactic immunization against *H. pylori* has been demonstrated for a variety of wild type or recombinant antigens, such as urease, vacA, cagA, HpaA (*H. pylori* adhesin A) and SOD (superoxide dismutase) (Ruggiero and Censini, 2014). However, it was observed in most of the cases that gastric colonization was not reduced to safety levels. This suggests that prediction of some novel antigenic epitopes within essential and virulent proteins is need of the hour for achieving higher efficacy. *H. pylori* 26695 genome (NC_000915) was selected for vaccine target identification because it is a representative annotated sequence and has a properly annotated genome sequence. The key steps in the scheme designed to identify and target some essential virulent proteins and their antigenic epitopes for peptide vaccine discovery, have been summarized in Fig. 1. Each step has been explained below in detail. Designed pipeline unveiled some potential vaccine targets among *H. pylori* 26695 proteins, which can be later subjected to clinical vaccinology.

2.2. Retrieval of primary data: sequence and features

Primary genomic and proteomic data of *H. pylori* 26695 was retrieved from GenBank and RefSeq (Pruitt et al., 2007; Benson et al., 2012). This data was characterized using PATRIC (Pathosystems Resource Integration Center) which is a knowledgeable source of annotated genomes to determine basic features of the targeted genome (Wattam et al., 2013). All the protein sequences were collected in FASTA format. Genbank (Benson et al., 2012) provides a comprehensive collection of all publicly available annotated genomes which integrates data from other renowned databases. RefSeq (Pruitt et al., 2007) along with DNA sequences also provides information regarding transcripts, proteins, mutations, expression studies, polymorphism analysis, characterization and functional annotation. PATRIC serves as a bacterial bioinformatics resource center providing an interface for bacterial infectious diseases integrating pathogen information with compliance of various analysis tools.

2.3. Screening exoproteome and secretome

Exoproteome or secretome of an organism consists of those proteins which are localized in the extracellular matrix or outer membrane of the organism. Subcellular localization and the secretion of pathogenic proteins is one of the important considerations for potential vaccine candidates. The exoproteome/secretome of an organism consists of large number of data sets which is very difficult to analyze manually. Thus, reverse vaccinology in combination with subtractive proteomics can give far better results as compared to screening of the whole data set without using prioritizing parameters (Sarangi et al., 2009; Barh et al., 2011). *H. pylori* proteome was screened using programs CELLO v2.5 (http://cello.life.nctu.edu.tw/) (Yu et al., 2004), CELLO2GO (http://cello.life.nctu.edu.tw/cello2go/) (Yu et al., 2014) and PSORTb v 3.0 (Gardy et al., 2005) (http://db.psort.org/) for the identification of secreted proteins or the ones residing in extracellular proximity. Identified proteins were further screened in the next step for further parameters including essentiality, virulence, molecular weight, etc. CELLO program is based on support vector machines (SVM) and is trained to localize bacterial proteins with a prediction accuracy of 89%. As there is an exponential increase in the subcellular data over the years, CELLO2GO program was developed which combines localization prediction with functional annotation. PSORTdb also provides a platform for bacterial protein localization within five modules including cytoplasmic, periplasmic, extracellular, outer-membrane and inner-membrane.

2.4. Evaluation of essential genes

In the next step, Database of Essential Genes (DEG) (http://tubic.tju.edu.cn/deg/) version 10.4 (Luo et al., 2014) was used to extract essential genes/proteins of exoproteome and secretome. Basic parameters used for BLAST against DEG were set as default including bit score of 100, cut-off (*E*-value) of $1E{-}5$ and the use of BLOSUM 62. Essentiality check is considered as an important parameter for identification of potential targets as they control major cellular functions of microbes (Muhammad et al., 2014).

Please cite this article in press as: Naz, A., et al. Identification of putative vaccine candidates against *Helicobacter pylori* exploiting exoproteome and secretome: A reverse vaccinology based approach. Infect. Genet. Evol. (2015), http://dx.doi.org/10.1016/j.meegid.2015.03.027