Sequence analysis

**PBIT: Pipeline Builder for Identification of drug Targets for infectious diseases**

**Gauri Shende†, Harshala Haldankar†, Ram Shankar Barai†, Mohammed Husain Bharmal, Vinit Shetty and Susan Idicula-Thomas***

ICMR Biomedical Informatics Center, National Institute for Research in Reproductive Health, Mumbai 400012, India

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

Associate Editor: John Hancock

Received on September 8, 2016; revised on October 17, 2016; editorial decision on November 23, 2016; accepted on November 25, 2016

**Abstract**

**Summary:** PBIT (Pipeline Builder for Identification of drug Targets) is an online webserver that has been developed for screening of microbial proteomes for critical features of human drug targets such as being non-homologous to human proteome as well as the human gut microbiota, essential for the pathogen’s survival, participation in pathogen-specific pathways etc. The tool has been validated by analyzing 57 putative targets of *Candida albicans* documented in literature. PBIT integrates various *in silico* approaches known for drug target identification and will facilitate high-throughput prediction of drug targets for infectious diseases, including multi-pathogenic infections.

**Availability and Implementation:** PBIT is freely accessible at http://www.pbit.bicnirrh.res.in/.

**Contact:** thomass@nirrh.res.in

**Supplementary information:** Supplementary data are available at Bioinformatics online.

---

**1 Introduction**

Emergence of multi-drug resistant pathogens and multi-pathogenic infections has necessitated the identification of novel disease targets. The wet-lab approaches for target identification have the disadvantage of being intensive on cost, manpower and time. This can be substantially overcome by supplementing the workflow with *in silico* methods for target identification/prediction. Curated information on sequences, structures, pathways, gene ontologies, drug-like compounds, essential genes and virulence factors available in online databases such as UniProtKB, KEGG (Kanehisa et al., 2016), DrugBank (Law et al., 2014), DEG (Luo et al., 2014) etc. can facilitate automated methods for identification of targets. Using subtractive genomics, that involves identification of microbial proteins that are non-homologous to human genes and are essential for the survival of the pathogen; putative drug targets have been successfully identified for many pathogenic bacteria (Anishetty et al., 2005). Apart from being non-homologous to humans and essential for the organism, few other criteria to be considered are participation in metabolic pathways distinct from humans, non-homologous to human gut microbiota, druggability status, etc. Microbial proteins involved in multiple pathways, unique to the pathogenic organism are better targets than proteins involved in specific pathways. Targeting specific pathways may enhance development of multi-drug resistance among pathogenic bacteria and should therefore be avoided (Shanmugham and Pan, 2013).

Gut microbiota play a vital role in maintaining health by providing resistance to colonization of pathogens and opportunistic bacteria. Interactions of the drug with gut microbiota are therefore one of the major causes of toxicity and reduced bioavailability of the drug. This necessitates the filtering out of microbial proteins that share structural similarity with human gut flora proteome in the target identification workflow (Muhammad et al., 2014; Raman et al., 2008).

Another important factor for a protein to be qualified as a drug target is its ability to bind to a drug-like molecule, also known as druggability. Not all proteins have structural characteristics conducive for drug-binding. Whole genome analyses estimates only 10% of genome as druggable, which emphasizes the significance of analyzing druggability for identification of target molecules (Radusky...
Proteins that are involved in multiple, pathogen-specific pathways of the input sequences. Information on whether these pathways are also present in human are provided for the benefit of users. KEGG pathway database was used as a source of pathway information.

2.4 Broad spectrum analysis
This module helps to identify targets that have homologs in multiple pathogenic organisms. Identification of such targets is vital for development of broad-spectrum drugs for treatment of multiple infections or poly-microbial diseases. This analysis is executed by similarity search using BLAST against proteomes of 181 pathogenic organisms.

2.5 Druggability analysis
This module helps to screen druggable targets based on their sequence similarity to experimentally validated druggable targets. The sequence information for druggable targets were obtained from DrugBank database (Version 5) and Therapeutic Target Database (TTD).

2.6 Host-Pathogen Interaction
This module identifies targets that share sequence similarity to microbial proteins known to interact with human proteome (Mais et al., 2016). The database for human anti-targets, human gut microbiota and list of pathogenic organisms were compiled from literature (Shanmugham and Pan, 2013, Raman et al., 2008).

The PBIT webserver has been validated using Candida albicans, which is one of the major causes of opportunistic infections leading to Candidiasis in many immune-compromised patients. The fact that (i) the proteome of this organism is well characterized and (ii) intensive research has been carried out for identifying its drug targets, made C. albicans a good model for validating PBIT (Supplementary Information). PBIT proved to be a fast and efficient method to screen out potential targets of C. albicans using sequence information. It is expected that a further updated database on essential genes and virulence factors can enhance the prediction accuracy of PBIT.

Acknowledgements
The authors are grateful to Dr. S.D. Mahale for the support. We thank Ms. Fayza Waghu, Ms. Rachel Fernandes and Mr. Sanket Ghawali for webserver testing.

Funding
This work [RA/389/06-2016] was supported by grants from Department of Science and Technology, India and Indian Council of Medical Research.

Conflict of Interest: none declared.

References


