

# OPEN SOURCE DRUG DISCOVERY: A GLOBAL COLLABORATIVE DRUG DISCOVERY MODEL FOR TUBERCULOSIS

ANSHU BHARDWAJ<sup>1</sup>, VINOD SCARIA<sup>1</sup>, DEBLINA PATRA<sup>2</sup>,  
OPEN SOURCE DRUG DISCOVERY CONSORTIUM<sup>2</sup>

---

*Open Source Drug Discovery (OSDD) is an innovative open collaborative model of drug discovery capturing the young as well as experienced minds around the globe in discovery of therapeutics for infectious diseases. It is a virtual laboratory which runs on a web-based platform for communication and interaction of OSDD Community members. IGIB, a major contributor in this CSIR led OSDD project, participates in various work packages starting from drug target identification to lead optimization. The first disease target of OSDD is Tuberculosis.*

---

## Introduction

The sequencing of the Human and other microbial genomes created stir in the scientific community with the promise to make a remarkable difference to healthcare. Despite the fact that the genome sequence of mycobacterium is available to the scientific community for more than ten years, no effective therapy/drug has been discovered to cure all types of tuberculosis. Although, the big pharmaceutical enterprises have made a commendable contribution to the discovery of drugs for several diseases yet the initiatives for the discovery and development of drugs for infectious diseases such as tuberculosis are sadly lacking. Out of 1,556 new chemical entities marketed worldwide, only 3 were for TB between 1975 and 2004<sup>1</sup>.

The market size has been a major driving force in targeting human diseases for new drug discovery programs. Also, the R&D costs are very poorly known, with reported estimates ranging from \$100 to \$500 million per drug<sup>2</sup>. Absence of a considerable market (in terms of monetary gains and recouping R&D expenses) is one of the reasons why little interest has been shown in developing drugs

for infectious diseases and diseases of the poor. Most of the drug discoveries are made in a closed door environment, where confidentiality results in lack of open participation of the entire academic world. Towards this end, the Council of Scientific and Industrial Research (CSIR), India, has launched the Open Source Drug Discovery (OSDD) project. OSDD is a CSIR led team India consortium with global participation for providing affordable healthcare. It aims to address the issue of affordable drug discovery by capturing the young and experienced minds around the globe to be a part of discovery of therapeutics for infectious diseases. OSDD is a web-based platform ([www.osdd.net](http://www.osdd.net)) with a community presently of more than 4500 members. This online platform ensures global participation of researchers, software professionals, technocrats, teachers and students so that they can leverage each others' expertise to facilitate the drug discovery process. OSDD aims to tap creativity inexpensively and on a scale that is beyond the reach of scientists working in their labs in academia or in private sector.

## Tuberculosis as the First Disease Target

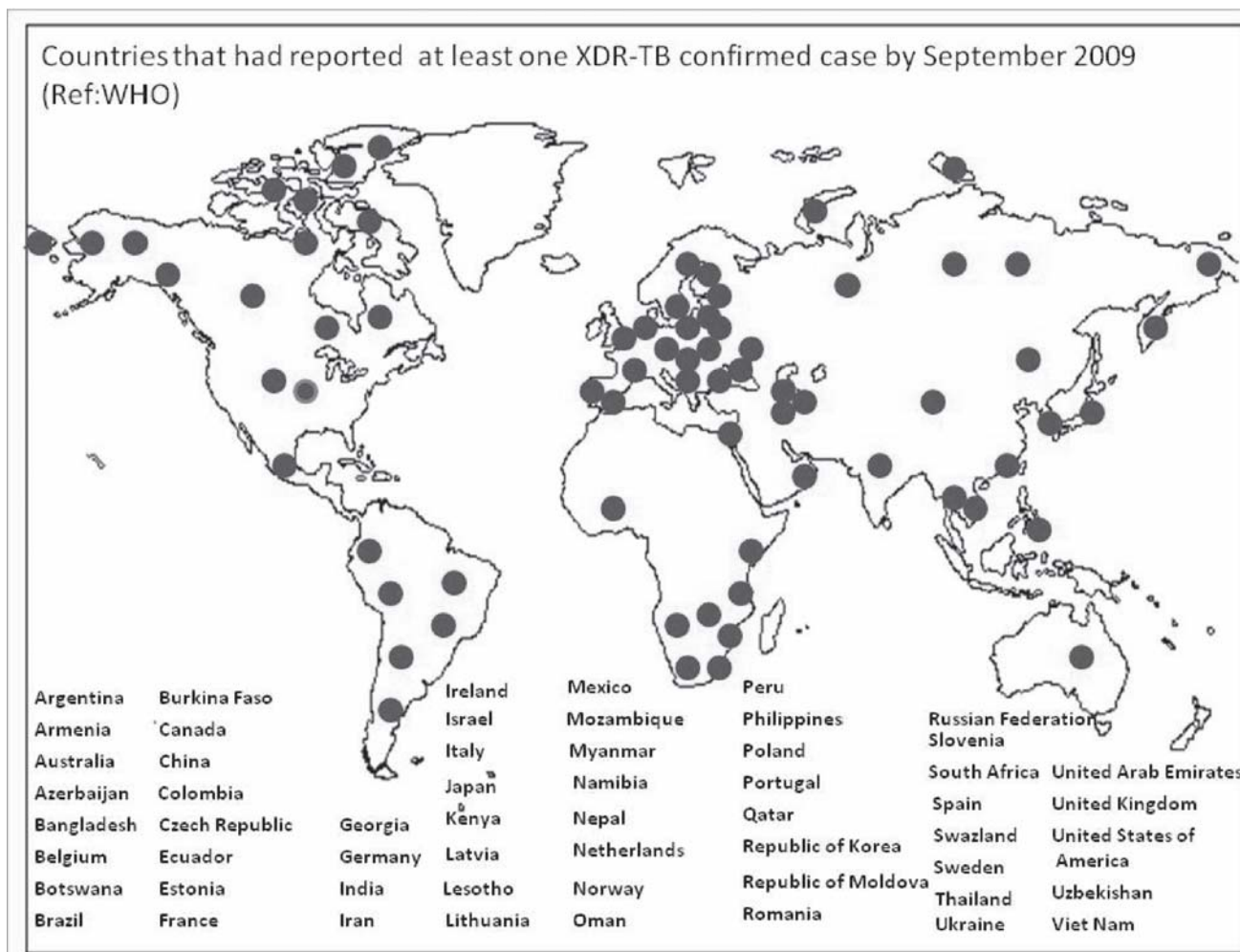
The first target disease of OSDD is Tuberculosis (TB), a leading cause of death in developing world. In India alone 3 people die of TB every 2 minutes. The current drug therapy has been developed in 1960s, and no major advancement in treatment has emerged for almost half a

---

1 Institute of Genomics and Integrative Biology (CSIR), Mall Road, Delhi 110007, India

2 Council of Scientific and Industrial Research, Anusandhan Bhawan, 2 Rafi Marg, New Delhi 110001, India

\* Corresponding authors : [anshu@igib.in](mailto:anshu@igib.in), [vinods@igib.in](mailto:vinods@igib.in)



**Figure 1.** The map represents the countries with red dots which had reported at least one XDR-TB confirmed case by 2009 as per WHO reports.

century. The current therapy takes months of constant drug intake which is a cause of non-adherence to the therapy and the consequential alarming rise of Multiple Drug-Resistance TB (MDR-TB) and Extensively Drug-Resistance TB (XDR-TB) in the population. XDR-TB is now reported in more than 57 countries and is spreading rampantly<sup>3</sup> [Figure 1]. This leaves with the immense global unmet need for an effective short duration therapy.

### **The OSDD Model**

The OSDD model is based on the principle of collaboration and knowledge sharing to solve complex scientific challenges. At the core of OSDD is SysBorg (Systems Biology of the Organism), a web based CyberInfrastructure for collaborative research. This platform enables creation of unconventional collaborative networks involving people with different skillsets and expertise. Such interactions leverage the creative abilities of young students and researcher which lead to development of

innovative approaches to solve challenging issues related to drug discovery. Issues like IPR (Intellectual Property Right) and confidentiality which significantly enhances the cost of drug discovery is countered in a knowledge sharing mode. The contributions are protected through OSDD license<sup>4</sup>. Each contribution is time and login stamped to ensure quantitative evaluation for microattribution.

Availability of funds is a crucial aspect behind the success of any initiative. Government of India has committed about Rs.150 crores for OSDD and the project proposes to raise equivalent amount of funding from multilateral/bilateral agencies and philanthropic organizations.

### **Microattribution for Contributions**

The OSDD model exploits a system of monetary as well as non-monetary rewards to engage biomedical researchers. In OSDD the entire process of drug discovery

is divided into work packages (WPs). Each of these WPs is further divided into problems open for the entire community to solve. These problems posted online as projects are peer reviewed by the community. There are also open challenges that the community may solve and participate in various projects. Every contribution on the OSDD portal is credited with points as part of the microattribution system. Based on the points accrued by the contributors they are awarded four levels of privileges or roles: Platinum, Gold, Silver and Blue groups. OSDD is an open, decentralized system and it works on OpenID for user-centric digital identity. The portal provides support for simultaneous access by multiple researchers and authentication by virtue of single sign-on. All contributions are tracked, the credits for contributors are computed from all the federated resources and the participants are assigned credits using this system. All the members are Blue card holders to start with and are upgraded to Gold or Platinum depending on the contributions.

### The OSDD Delivery Model

OSDD is a drug discovery project which aims to discover new chemical entities (NCEs). Once approved as a drug, OSDD proposes to make it available like a generic drug which enables multiple entities to produce the drug to ensure its affordability and availability to a vast population. All discoveries and data would remain open to the OSDD community. OSDD will follow a collaborative path on a consortium mode based on an open invitation to willing partners wishing to participate in the clinical trial whose efforts will be subsidized by OSDD. OSDD model de-risks clinical trial. The private partners do not have to defray the cost of research leading to the molecule.

### Results & Discussion

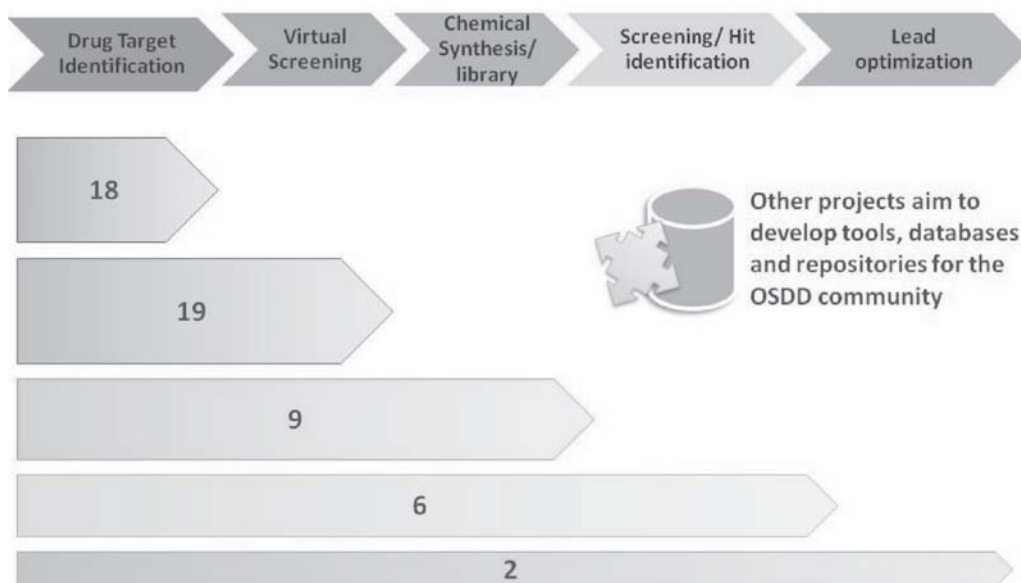
OSDD provides a good model for early stage drug discovery. The OSDD community has developed innovative ways to generate resources in terms of biological or

chemistry databases, software applications and analytical tools along with data analysis. As of now there are various projects in OSDD across different work packages [Figure 2].

### Resources Generated for and by the Community

**Generating data as a community resource :** OSDD functions as a virtual distributed laboratory and comprise of various federated resources. To ask and answer questions and to understand the pathogen, the community needs data on mycobacterium. The OSDD community members decided to fetch this data from various resources (databases, supplementary materials or tables in publications, predictions or experimental data from high-throughput experiments, etc). The existing heterogeneity of the data formats limits cross-talk between data across multiple resources. With the objective to integrate these resources by converting this data to a standard format, an integrated platform has been developed through online collaboration<sup>5</sup>. This platform provides the largest resource on mycobacterial genomics in a standard interoperable format and is named Tbrowse<sup>6</sup>. TBrowse has been implemented as an online collaborative project lead by Principal Investigators from IGIB and research students from universities. As of now TBrowse integrates data from nearly fifty different resources encompassing more than a million data points<sup>7</sup>.

### Status: OSDD Projects



**Figure 2.** The bar graph depicts the OSDD project status (as of October 2010). There are multiple projects at the early drug discovery phase and some at hit-to-lead phase.

### ***Developing applications with the Community :***

Another significant component of the OSDD federated resource is Computational Resources for Drug Discovery (CRDD)<sup>8</sup>. The CRDD web provides computational resources related to drug discovery on a single platform. CRDD provides computational resources for researchers in the field of computer-aided drug design and maintain a wikipedia related to drug discovery. The members of the OSDD community and others may contribute/host their database or web server on the CRDD portal<sup>9</sup>. CRDD has started a novel initiative called Indipedia to discuss issues related to drug discovery in Indian context. A related component of this activity is to create a platform for designing *in-silico* workflows which facilitates automated and high-throughput computational analysis. For doing so, the members of the OSDD community have developed web services using WSDL standards. Some of these web services are in-house applications and others have been integrated from KEGG, EBI, ChEMBL, NCBI, etc. As of now more than 250 applications are available through this interface. These applications are integrated into Galaxy<sup>9</sup> and may be used to generate workflows. The OSDD community is free to use and add more tools to this open source workflow engine.

In order to enable data analysis through integrative querying, the data available on SysBorg has been packaged into R environment by the OSDD community members<sup>10</sup>.

### ***Collaborative projects: Cases from Drug Target Identification and Inhibitors of Mtb***

The OSDD community has mined literature and identified more than 130 drug targets of Mtb from year 2001 to 2010. This resource has been value added with more than 20 properties for each target and this effort was carried out by community comprising mostly of students. This serves as the repository for biologists and others working on Mtb to understand the status of drug target research. Efforts are also ongoing to list all possible inhibitors of Mtb from literature and to understand their properties in order to design novel molecules based on this understanding. Nine of the 130 drug targets are being actively followed-up for onward investigation by researchers from different parts of India. In OSDD, GlmU is predicted as the potential drug target for which IGIB is playing the lead role for the task of *in silico* biology for drug target identification<sup>11</sup>. In addition, there are other drug

targets too which are being pursued further for validation and hit identification.

### ***Connect to Decode Program***

Open Source Drug Discovery (OSDD) Project in its quest towards conquest of Tuberculosis launched Connect to Decode (C2D)<sup>12</sup>, a massive initiative to enhance the understanding of the biology of the causative organism, *Mycobacterium tuberculosis*. Since its inception in November 2009, the C2D programme has attracted over 1000 enthusiastic researchers from across the world. This overwhelming response to the C2D programme is a reflection of the untapped potential of the talent available around the world for pursuing serious scientific enterprise for common good. OSDD took the challenge of comprehensively re-annotating all possible genes in Mtb in a virtual collaborative manner. C2D identified five themes under which the Mtb genome has been annotated, namely, pathway/interactome, gene ontology, glycomics, protein-structure fold, and immunone. Each of these themes is mentored by a group of researchers. The overall project design of C2D was conceived at IGIB and two major themes have been implemented by OSDD Principal Investigators at IGIB. C2D used open Web 2.0 platforms where volunteers communicated. Most of the volunteers were students and did this as a part of their summer project, the successful completion of which has fetched them a certificate from the OSDD/CSIR.

### ***OSDD Portal***

In addition to understanding the biology of Mtb, it is also important to create robust cyberinfrastructure to support online collaborations, data exchange and analysis. As already mentioned, the OSDD community has not only integrated existing data on TB but is also generating a lot of data and applications. These resources have no inherent meaning unless used efficiently by integrating and making them amenable to answer questions. To achieve this goal, CSIR/OSDD has partnered with Infosys, a major IT giant, to create a semantic search engine based portal<sup>13</sup>. IGIB has played a critical role in shaping the design of this portal with Infosys. The new OSDD portal is expected to enhance collaborations and enable researchers to seamlessly use and contribute data, tools, ideas and resources. This portal works as a platform for the development of a huge collaborative network leading to creation of virtual distributed laboratory for solving complex challenges in drug discovery of tuberculosis to begin with. This portal is an integrated framework for researchers as it provides social networking tightly coupled

with scientific workflows. This allows researchers and others working on a collaborative project to interact seamlessly and share data and information in the most effective manner.

It's said when a new concept blooms into a wonderful working method, then it can really change the vision in any field and so is the case with Open Source Drug Discovery project. The new methodologies combined with collaborative platform is the main strength of Open Source Drug Discovery Process. Open source will provide better quality, higher reliability, more flexibility, lower cost, and bring an end to closed-door activities. OSDD aims to provide affordable and inclusive healthcare for all with community participation. □

#### References:

1. M Casenghi, S T Cole and C F Nathan *PLoS Med* **4**: e293 (2007)
2. S M Maurer, A Raj and A Sali, *PloS Med* **1(3)** : e56 (2004)
3. Web link [http://www.who.int/tb/challenges/xdr/xdr\\_map\\_sep09.pdf](http://www.who.int/tb/challenges/xdr/xdr_map_sep09.pdf)
4. Web link <http://sysborgtb.osdd.net/bin/view/Main/OsddLicence>
5. Web link [www.osdd.net](http://www.osdd.net)
6. Web link <http://tbrowse.osdd.net>
7. A Bhardwaj, D Bhartiya, N Kumar; Open Source Drug Discovery Consortium, Scaria V. Tuberculosis (Edinb). **89(5)**: 386-7 (2009)
8. Web link <http://crdd.osdd.net>
9. J. Goecks, A. Nekrutenko, J Taylor and The Galaxy Team *Genome Biology* **11** : R86 (2010)
10. S Ramachandran, A Katiyar, A Sinha, A Bharadwaj, A Dutta, A Raman, A Pan, B K Malik, B Singh, B Pillai, B Dutta, B P Sinha, B Taneja, C Mandal, C K Richa, C Dutta, D Dash, D Mukherjee, D Paul, D Chakraborty, F A Ansari, G P Singh, G P S Raghava, G Guhathakurtha, I Siddiqui, M Kumar, M Hariharan, M B Subramanyam, M Joon, M Bose, M Haymanti, M Gnanamani, M V Raghunandan, N Ghoshal, N K Singh, P Sarmah, R S K Vijayan, R Verma, R Sharma, R Ramachandran, R Chaudhuri, S Das, S K Brahmachari, S Paul, S Chatterjee, S Bhutoria, S Chowdhury, S Gupta, S Maiti, S Ghosh, S Arora, S Saha, S K Bag, S Deb, V Brahmachari, V Gupta, V Kumar, V Scaria, Y Bhasin, Y Singh [OSDD Consortium] *Biobytes* **5** : 40-48 (2009)
11. M. Anurag and D. Dash *Mol. BioSyst.* **5** : 1752-1757 (2009)
12. Web link <http://c2d.osdd.net>
13. Web link <http://sysborg2.osdd.net>