ALMOST at the dawn of its creation in 2008, the Open Source Drug Discovery (OSDD — http://www.osdd.net) programme, which is a CSIR-led Team India consortium with global partnership, had identified *Mycobacterium tuberculosis* (MtB) as its first target. The reason is not far to seek. With a history going back to prehistoric times, TB has stalked the Earth for millennia, claiming Pharaohs of ancient Egypt as well as Poets of contemporary times. However, despite all our grand advances in science we have failed to wipe it off the face of the globe.

Fortuitously for us, however, the IT-revolution in the early days of the twenty-first century has powered Genomics research such that looking at genomes has become easier. Rational drug discovery hinges on knowing the best target to attack, the best pathway to exploit and to do so while wasting a minimum of resources, including time. Knowing the genome is now the first step to knowing how to control a pathogen.

The time was ripe to take advantage of the available technology and to signal a change in the way drug development takes place. It was time to shift to rational drug development instead of a hit-and-miss approach. It was against this backdrop that the *Connect to Decode (C2D)* (http://c2d.osdd.net) initiative of the OSDD was announced. It was designed to meet the challenge of annotating all possible genes in MtB.

Diagnostics, Vaccines & Drugs

It is unfortunate that TB which has high mortality but low profitability (because it is a “disease of poverty”) is neglected by the current system of pharmaceutical research. Virtually no new TB drugs have been developed since the 1960s. Thus, our arsenal of weapons against TB is woefully dated.

C2D demonstrates the power of people to connect through the Internet, particular young people, and accomplish complex research tasks. This exercise has provided a large number of clues towards obtaining a comprehensive view of the microbes as a whole, producing several testable hypotheses that can now be verified in the laboratory.
Our diagnostics and vaccines are equally inadequate. While Sputum Microscopy was developed in 1882, Mycobacterial culture came into being in 1882 and Chest X-ray in 1896. Besides, the sole vaccine available is the Bacillus Calmette-Guérin (BCG) vaccine developed in 1921. It has limited efficacy in preventing Tuberculosis.

OSDD will share through a globally accessible database to any research institutions involved in TB research through its open portal (www.osdd.net). The annotation has paved the way for enriching the genome study even further through open collaborations. More importantly, the data is available for effective application in rational drug discovery.

Dr Samir Brahmachari, DG, CSIR, with Mr. Prithviraj Chavan, Minister of State for Science &Technology and Earth Sciences

TB, HIV & Drug Resistance
Combating TB has become even more difficult because of TB’s deadly relationship with HIV.
• TB accounts for up to a third of AIDS deaths worldwide.
• In some countries with high HIV prevalence, up to 80% of people with TB test positive for HIV as well.
• Complicating the already complex scenario is the fact that drug-resistant varieties of Mtb are spreading.
• Multi-Drug Resistance (MDR) is rampant in many parts of the world and Extreme Drug Resistance (XDR-TB) is now reported in more than 45 countries.

The organism lies dormant in the body once it infects and evolves swiftly; so, the strategies to defeat it must be novel too. It is clear that without an understanding of the molecular secrets embedded in the organism’s genome, identification of its soft underbelly is impossible. We
Bones showing signs of tubercular infection have been found since the Neolithic period (~3000–7000 BC).

On 24 March 1882 Dr. Robert Koch identified the causative organism.

India accounts for one-fifth of the global TB cases.

Nearly 2 million people in India develop TB annually, of which around 0.87 million are infectious cases.

Two people die of TB every three minutes in India.

No country is TB-free. TB is a worldwide pandemic; about one-third of the world’s population is infected with TB.

New infections occur at a rate of one per second. 5-10 per cent of those infected develop full-blown Tuberculosis later.

Left untreated, each person with active TB disease will infect, on average, between 10 and 15 people every year.

WHO reports that 1.7 million people die annually from TB.

In some areas of the world, one in four people with TB becomes ill with a form of the disease that can no longer be treated with standard drugs regimens.

Global incidence of TB is increasing by 0.6% per annum.

There is no doubt that tackling TB is a major challenge. New drugs and vaccines are vital.

Also, it is important to approach the problem with a Systems Biology perspective of the pathogen in mind. Systems Biology is the integrated approach to studying biological systems. Instead of studying proteins one at a time as used to be the traditional way of looking at cells, Systems Biology looks at the larger picture. For example, it studies intracellular networks, cells, organs, and biological entities by measuring and integrating data gathered at the genetic, proteomic and metabolic levels. This approach involves the study of cellular and pathway events, which by virtue of their roles are usually in a state of flux. Many are also interdependent.

The Systems Biology approach to drug discovery relies also on using clinical samples from diseased as well as healthy people. This study helps uncover Systems Biology Markers, which are the footprints (so to say) of genes and Pathways Targets. These act as indicators of disease and are potential targets for therapeutic intervention. In a way, the Systems Biology approach acts like a spy who has integrated into the community and has intimate knowledge of all that is happening there. Rarely can such knowledge fail to yield results.

There is no doubt that tackling TB effectively requires addressing all the risk factors that make individuals vulnerable to infection with Mycobacterium tuberculosis and to developing the disease. It also calls for new drug targets and more potent, less time-consuming drug regimes.

This is where C2D stepped in because the role of functional genomics in modern drug discovery is to prioritize targets and to translate that knowledge into rational and reliable drug discovery. Of course, polymorphisms among Mtb strains are quite extensive and genetic variation may have an important role in disease pathogenesis and immunity. So, annotation is where the journey begins.

Connect 2 Decode (C2D)

C2D was launched as a massive initiative aimed to further the understanding of the biology of Mycobacterium tuberculosis. In the words of the scientists working hands-on, it was an attempt to enrich the annotation of the Mycobacterium tuberculosis genome. The plan was not to just take a closer look at the thousand or so genes that had already been annotated (to re-annotate if necessary) but to annotate the whole genome of 3998 genes, including the yet un-annotated genes too. This also involved re-annotating the Immuneome and Glycome of the deadly bug as well as looking at structural models of the 4000 odd proteins.

Preparation of one of the most comprehensive metabolic pathway maps of the organism by manually curating the data, applying various computational tools and also by integrating post-genomics datasets on Mtb was on the cards. This would be the first step to accelerated discovery of novel drugs for TB. There was a sense of urgency to the work and the deadline when this would be actually achieved was announced well ahead of time.

Mtb Genome and Beyond

The genome of Mycobacterium tuberculosis H37Rv strain was sequenced in 1998 and reported in the 11 June 1998 issue of the journal Nature. This work by Stewart Cole and his group remains the foundation stone of TB research. The advent of the science of Genomics heralded an era of hope. It was hoped that the decoding of the microbial genome would lead to the identification, and subsequently, the exploitation of the genetic vulnerabilities of the microbe. It was also hoped that decoding the Mtb genome would lead to the identification of more precise drug targets and, thus, more potent pharmaceutical products. It was hoped that soon the Earth would be free of this scourge. Unfortunately, that did not happen.

Till date, the standard databases have annotated only about 1000 of the near 4000 genes encoded by the organism. The rest of the genes languish in apathy...and new and potent drugs for this killer disease await discovery that seems light-years away for those whose very existence is threatened by active infection with TB. With only a quarter of its genetic secrets known in any detail, it is hardly surprising that no new drugsdrug targets have been ferreted out of this genetic treasure trove. This is symptomatic of the problem of neglected diseases of the poor...and a case replicated all over the globe.

However, initially when the genome sequence of Mtb was deciphered, it did trigger intense research on the pathogen. Soon the first annotation came out. It provided the first “parts list” of the microbe. A re-annotation carried out subsequently identified more parts, completing the list of proteins coded by the genes of Mtb. Unfortunately, it was at this point that the interest seems to have lost steam.
C2D Phase II
The C2D Phase II is also now on. Those interested may log on to http://c2d.osdd.net

Very little, if any at all, work was done with a Systems Biology approach in mind. That is to say, no research was done on how these “parts” are “assembled” in the cell. Now, obviously, to do this, the description of each “part” must be known in as much detail as possible. Any new scientific work usually builds on the edifice of past research. So, the OSDD researchers began by scanning existing literature. They found a wealth of experimentally derived data on various proteins of Mtb. They were helped in this by IT and knowledge-driven tools that were helpful in enriching the annotations. However, the information was scattered, buried deep and needed to be extracted if it was to be used meaningfully.

The C2D Strategy
The initial requirement was for a large pool of workers to scrutinize available literature, use computational tools and collate this information. With a global community of nearly 3000 members from 74 countries, OSDD is a community of scientists, doctors, students, policy experts, software professionals, research students and others united in the quest to defeat TB.

Recruiting Raw Talent and Mentoring Merit: Finding willing volunteers was the least of the worries. The day registration to C2D was opened, over a 100 volunteers signed in; subsequently the numbers peaked at 800. The Faculty drawn from the best Institutions included: Dr. G. P. S. Raghava (Institute of Microbial Technology, Chandigarh, CSIR); Dr. Nagasuma Chandra (Indian Institute of Science, Bangalore); Dr. Debasis Mohanty (National Institute of Immunology, New Delhi); Dr. Samik Ghosh (Systems Biology Institute, Japan); Dr. Sulagna Banerjee (AU-KBC); Dr. Anshu Bhardwaj (Institute of Genomics and Integrative Biology, Delhi, CSIR); and Dr. Vinod Scaria (Institute of Genomics and Integrative Biology). The participants were also from all over the country, from Kashmir to Kerala and from Gujarat to West Bengal.

Mapping the Work: The C2D initiative leveraged this huge strength to tackle the Mtb genome from five different angles. The first, called TBGO, obtained associations for each gene with function(s). This exercise has provided a large number of clues towards obtaining a comprehensive view of the microbes as a whole. It has presented several testable hypotheses that can now be verified in the laboratory. The annotation has been made available on a public portal, paving the way for enriching it even further through open collaborations. More importantly, the data is currently available for effective application in rational drug discovery.

Second, each protein was studied at high resolution by modeling their three-dimensional structures, through which higher order clues about their functional roles are obtained.

Third, proteins which interact with various sugars or carbohydrates in the cell were deciphered; these are believed to be important for the signaling events in the cell, somewhat akin to the switches.

Fourth, the assembly of the parts was sought out by identifying those parts that talk to each other directly and those parts that talk to others through mediators, such as metabolites.

Integrating data from all the angles, or network of the parts list was reconstructed, in which functional modules (or biochemical pathways) were also identified. The network helps in understanding how metabolism takes place in the bacterial cells, how it compares with humans and with other bacteria. It also helps in answering questions about what strategies should be adopted to kill the bacteria. This helps fashion a strategy for drug design.

A fifth theme that was pursued was to identify proteins that contain antigenic parts in them that could trigger immune responses in the host and thus ultimately help in rational vaccine design.

The C2D Experience
The project was implemented in an extremely cost- and time-effective manner and by tapping into a huge pool of raw, young talent. Student volunteers were inducted into the programme and trained over a period of about four months. Their work was monitored online by the respective Principal Investigators. Stringent quality control measures were put in place to verify and correct the annotations made by students in various places. The volunteers pooled their time and skills, using online tools, to provide insights into all the genes of Mtb.

Their work is held in a shared database, which OSDD will share through a globally accessible database with any research institutions involved with TB research through its open portal (www.osdd.net). The annotation has paved the way for enriching the genome study even further through open collaborations. More importantly, the data is available for effective application in rational drug discovery.

C2D demonstrates the power of people to connect through the Internet, particular young people, and accomplish complex research tasks. It is also a distinct move from a hierarchical based model of doing science towards one of equal collaboration.

This exercise has provided a large number of clues towards obtaining a comprehensive view of the microbes as a whole, producing several testable hypotheses that can now be verified in the laboratory.

Dr. Sukanya Datta, Scientist NISCAIR, posted to Director General’s Technical Cell, CSIR-HQ
Email: sukanya@csir.res.in