Chemical genomics: an emerging interface between life science and chemistry

Bipasa Sarkar and Tapan Kumar Mondal

Achievements in science always raise new questions. After revealing the complete sequence of the human genome, the major question was how to translate the genome into biological insights and therapeutics? Answers may be many, but one of them is chemical genomics. This is the study of genetics using small chemical molecules to change the way proteins work, directly in real time rather than indirectly by manipulating their genes for identifying new therapeutic targets and drugs. It is the direct descendent of 'conventional chemical genetics' which has been studied with unexpectedly discovered compounds that cause interesting phenotypic changes. Now it is the era of 'chemical genomics', a large-scale version of 'conventional chemical genetics' targeting the whole genome in an effort to find both useful compounds and their targets by applying them to a variety of genes. To accomplish this goal more effectively and systematically, it is necessary to make new synthetic compound libraries. Several chemistry-based disciplines such as combinatorial chemistry, synthetic chemistry and chemoinformatics are required. It is used to identify proteins that regulate different biological processes, to understand how proteins perform their biological functions at the molecular level and to identify small molecules with medicinal value. It also integrates rapid target and drug discovery using active compounds (or ligands) as probes to characterize protein functions.

On the other hand, classical genetics is robust and well established. However, due to certain bottlenecks, chemical genomics is gaining rapid popularity. Detailed information about chemical genomics is beyond the scope of this note. Nevertheless, to understand the concept we must recall the meaning of gene, genomics and classical genetics to reveal the gene function. Several excellent reviews on chemical genomics are avaiable[1–3]. The word 'genomics' denotes the study of genes and their expression in a cell at a given time. The genetic information of a cell is stored in the DNA, which is then translated into the intermediate carrier RNA, to finally produce a product of the gene is called protein. Thus the binding of any small chemical molecule either to the genetic information (DNA) or its intermediate (RNA) or its final product (protein) will change the function/character/phenotype; this is the central theme of chemical genomics. On the other hand, classical genetics is forward and reverse genetics. In forward genetics, the organism, cell, etc. are artificially mutated by some mutagenic agents, which are then screened for a phenotypic (mutant) change for instance morphology, root length, etc. This phenotype is then compared with the wild type to map/isolate the gene responsible for the same. Basically forward genetics deals with 'effect to cause' or 'phenotype to genotype (genome sequence)'. Contrary to this, in reverse genetics, a particular gene is manipulated outside the cell and then this 'manipulated gene' is inserted into a cell, to allow it to grow into an organism in order to observe the difference in phenotype compared to its wild type, due to this manipulated gene (Figure 1). Thus reverse genetics deals with 'cause to effect' or 'genotype to phenotype'. Similarly, chemical genomics mimics: forward chemical genomics and reverse chemical genomics. In the case of forward chemical genomics, mutation is carried out using the collection of small chemicals known as 'chemical library'. The inhibition or stimulation of protein function is then determined by identifying the small chemical molecule using chromatographic techniques which reveal the protein attached to the molecule. Reverse chemical genomics begins with a known protein analogous to a specific gene selection. Subsequently this known protein is then screened with vast pools of library compounds to identify functional ligands that either stimulate or inhibit the target protein. Once a specific ligand is identified, it is introduced to the cell or organism similar to genetic mutation and the phenotypic change is studied.

The third area of chemical genomics, known as 'predictive chemical genomics', deals with a collection of genomics responses of a cell or organism to treat with a small chemical molecule or drug in case of mammalian cell research. The primary objective is to holistically

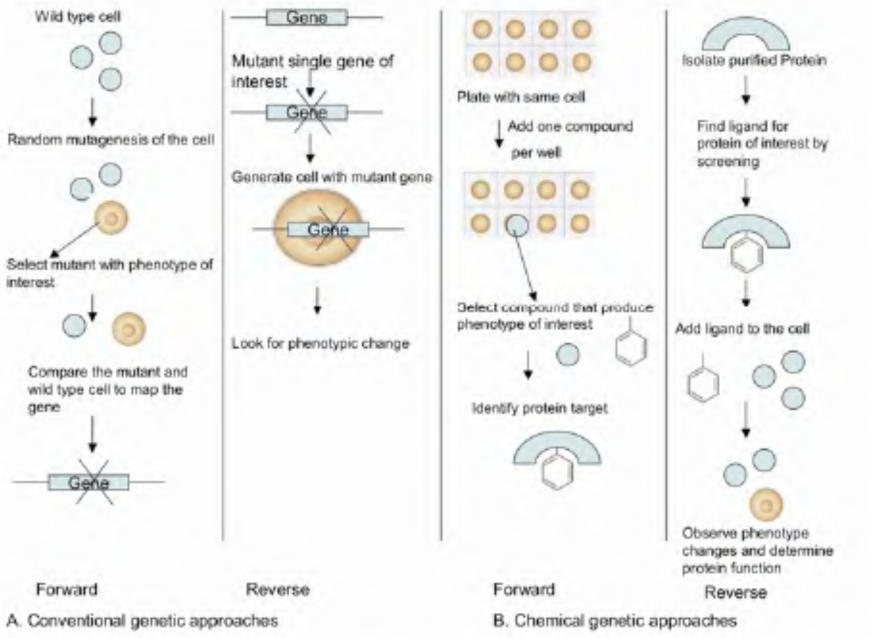

Figure 1. Pictorial expression of conventional genetics and chemical genetics[14].

Table 1. Small molecules and their target identified in forward chemical genetic screening

Small molecule	Activity	Target protein
Encephalazine	Inhibition of brain/eye development in zebra fish	Ribosomal subunit proteins (S5, S13, S18, L28)
Tubulyzine A, B, C	Slow epiboly development in zebra fish	Tubulin
Myoseverin	Myotube disassembly	Tubulin
Melanogenin	Pigmenting melanocyte	Prohibitin
Syntab A	Mitosis perturbation	Tubulin
Aminopurvalanol	Cell-cycle arrest at G2/M	CDK 1 (cyclin-dependent kinase 1)
Diminutol	Inhibition of miotic spindle assembly	NQOI (quinine oxidoreductase)
Ubistatin A, B	Inhibition of mitotic entry	Ubiquitin

characterize treatment response coupled with the secondary aim to identify novel therapeutic molecules. This involves the collection of genomic responses (e.g. through microarry analysis) and pharmacological responses (through growth inhibition assay) of a cell type or tissue to treatment with various drugs. Each drug profile represents its own signature of the drug at the transcriptional and molecular pharmacological level. Integration of the genomic and pharmacological statistical data then reveals predictive gene–chemical molecule or gene–drug relationship, which helps in the study of drug delivery without any expensive lengthy protocol, an area which has importance in the pharmaceutical industry. However, predictive chemical genomics has considerable overlap with pharmacogenomics. In contrast to the latter, predictive chemical genomics strategies generate gene–ligand response associations by concurrently considering the response profile of thousands of drugs, rather than that one molecule at a time. Some of the small chemical molecules along with their target proteins are given in Table 1.

There are advantages of chemical genomics over classical genetics. Chemical genomics may be conducted in any complex cellular or animal model which has a long life cycle, large physical size or where phenotypes are masked through related gene functional compensation, where classical genetics become slow. Chemical genomics also allows for the possibility of 'multiple knockouts' due to the addition of multiple specific ligands; a situation often found difficult in classical genetics. Sometimes complete removal of a gene through classical genetics was found to be lethal for an organism. This can be avoided by chemical genomics that allows the use of sub-lethal doses of the ligand and avoids full lethality, thus providing partial knockout phenotypes. Another advantage is real-time control based upon cell-permeable ligands at any stage that may yield the desired phenotype that diffusion-limited kinetics allows.

There are certain disadvantages as well with chemical genomics. While some chemical ligands can be specific switches with specificity approaching that of a gene knockout, low specificity of other ligands often gives 'off-target' effects in which the probe may interact with proteins other than the targetted one. These off-target effects may lead to toxicity or false/unwanted biological results.

Importantly, the success of chemical genomics depends upon the synthesis as well as nature of the chemical library, and speed and type of screening of the 'chemically induced' cell or organism. Based upon the need, there are several types of chemical compounds which are used in the library. Use of natural plant-based products is one such example. The living cell has a large collection of products, which have a higher probability of delivering hits than the typical synthetic combinatorial chemistry that is a process for preparing a large collection of compounds or libraries by synthesizing all possible combinations of a set of small chemical structures. Secondly, natural products like compounds whose structures are based on, or which share a high structural homology with natural products. These libraries may be designed to generate derivatives of natural products scaffold. Diversity-oriented synthesis is a new approach for the chemical library, where the diversity is introduced either by the differentiating process or the folding process. The goal is to maximize diversity in chemical space and generate libraries that are static pools of discrete molecules. Tagged libraries offer a unique opportunity in library design, where some functional tags are integrated into the library from the beginning and then bestow some additional functions into the molecule. Tag-guided library assembly is used to synthesize the library specially when little information is available about the target. Briefly, a collection of soluble binding elements must be assembled that possess a functionally reactive group capable of linking the elements together. The binding elements area is then screened and weakly binding elements are identified. Subsequently, a combinatorial library is created by linking together any identified binding elements with various length linkers. This library is then screened against the target. The dynamism of dynamics combinatorial library results from the reversible interchangeability possible with their components. In this system, every member of the library and the targets themselves affect all other members of the library, particularly in terms of library composition. Annotated chemical library, a collection of compounds of diverse structure from various sources possessing experimentally bonafide biological activities and mechanism containing compounds of diverse sets of biological activity can also be used.

Once a library of compounds is assembled, the next step in chemical genomics is testing them for biological activity, which is known as 'screening'. This must be designed with maximum sensitivity, selectivity, reproducibility and cost-effectiveness in mind. The speed of screening has improved from high throughput screening (348 well plates) to ultra-high throughput (3456 well plates), to the present gel-based technology called 'microarray compound screening'. The latter has the ability to store the library in a dry, inert environment in a ready to screen format. Forward chemical genom-

ics has been applied successfully with many model organisms such as yeast, *Arabidopsis*, zebra fish, *Drosophila* and *Caenorhabditis elegens*, and the mammalian cell. While the most popular is phenotypic screening, others such as cytoblot, gene reporter and fluorescent imaging are applied to the mammalian cells only. Apart from this, cell-free systems such as cell-free *Xenopus* egg extract, provide a versatile environment in the phenotypic screening of compounds for the study of a variety of processes, such as to develop drugs for diseases, cell division, etc. An in-depth literature survey indicates that the screening for reverse chemical genomics includes *in vitro* screening[4], chemical inducers of dimerization[5], orthogonal chemical genetics[6], disruption of protein interaction[7], targeted protein degradation[8], screening on beads[9], wide-angle X-ray solution scattering[10], small molecular microarray[11] and phenotypic response[12]. A detailed account of these techniques and other information is available in some public and private databases of chemical genomics such as PubChem (pubchem.ncbi.nlm.nih.gov), Chem Bank (chembank.broad.harvard. edu), WOMBAT (www.sunsetmolecular.com), Jubilant (www.jubilantbiosys.com) and GvkBio (www.gvkbio.com).

The importance of chemical genomics in plant research has been well elucidated recently by Miyashita and Miyagawa[13]. Due to lack of a circulatory system as well as adaptive immune system, plants have evolved their own defence systems distinct from animals, in which each plant cell is capable of defending itself from pathogens. Plants induce a number of defence responses, which are triggered by a variety of molecules derived from pathogenic microorganisms, referred to as microbe-associated molecular patterns (MAMPs), including peptides, proteins, lipopolysaccharides, beta-glucan, chitin and ergosterol. Interaction between plants and chemicals in the context of plant defence represents a 'natural' and simple model for chemo-genomics, at the intersection between chemical and biological diversity. For protection of crop plants from diseases, it has been shown to be effective to stimulate the plant immunity by chemical compounds, the so-called 'plant defence activators'. Combinatorial chemistry techniques can be applied to the search for novel plant defence activators, but it is essential to establish an efficient and reliable screening system suitable for library screening. For this, the cell-based lawn format assay has been developed for the peptides acting as plant defence activators from combinatorial peptide libraries.

It is noteworthy to mention that perceiving the potential of this new field a new journal, namely *Current Chemical Genomics* has been dedicated recently to publish chemical genomics work. A mega chemical genomics project was started in the US in 2004 at the National Institutes of Health (NIH) Chemical Genomics Center (www.ncgc.nih.gov) to develop probe, to allow academic and government scientists in this field (a currently unprecedented) and to access to large libraries of organic chemical compounds with diverse substances. In Europe and China, infrastructure development for chemical genomics research is in place. Several universities in USA have already initiated research programmes in chemical genomics, though there are new academic courses on chemical genomics in Michigan University and Colombia University.

In conclusion, chemo-genomics is a rapid emerging dynamics discipline that combines the latest tools of genomics with chemistry, and applies them to know the gene function, drug discovery, etc. In life sciences, it plays an important role to understand the basic mechanism of gene expression, complex disease mechanisms such as cancer, development of pesticides, growth retardants, drug discovery, new pathway discovery which brings morphological or developmental changes. Considering its full potential, there is a need to establish a National Chemical Genomics Research Institute of India under the aegis of the DBT or CSIR along with the introduction of courses at the undergraduate level and specialization at postgraduate level in the universities.

1. Wash, D. P. and Chang, Y.-T., *Chem. Rev.*, 2006, **106**, 2476–2530.
2. Bredel, M. and Jacoby, E., *Nature Rev.*, 2004, **5**, 262–275.
3. Maréchal, E., *Comb. Chem. High Throughput Screen*, 2008, **1**, 583–586.
4. Gray, N. S. *et al.*, *Science*, 1998, **281**, 533.
5. Crabtree, G. R. and Schreiber, S. L., *Trends Biochem. Sci.*, 1996, **21**, 418–425.
6. Bishop, A. *et al.*, *Annu. Rev. Biophys. Biomol. Struct.*, 2000, **29**, 577–597.
7. Huang, J. and Schreiber, S. L., *Proc. Natl. Acad. Sci.*, 1997, **94**, 13396.
8. Schneekloth *et al.*, *J. Am. Chem. Soc.*, 2004, **126**, 3748–3758.
9. Ostrem *et al.*, *Biochem.*, 1998, **37**, 1053–1060.
10. Fischetti, R. F. *et al.*, *J. Synchrotron Radiat.*, 2003, **10**, 398–404.
11. Koehler, A. N., *J. Am. Chem Soc.*, 2003, **125**, 8420.
12. Canman, J. C. *et al.*, *Nature*, 2003, **424**, 1074.
13. Miyashita, M. and Miyagawa, H., *Comb. Chem. High Throughput Screen*, 2008, **11**, 647–652.
14. Stockwell, B. R., *Nature*, 2000, **11**, 116–125.

Suggested reading:

1. Zheng, S. X. F., *Chemical Genetics and Chemical Genomics*, Routledge, USA, 2004.
2. Introduction to Chemical genomics reviews and protocols. In *Methods in Molecular Biology*, Human Press, vol. 10.

ACKNOWLEDGEMENTS. We thank Prof. Abhijit Roy, University of North Bengal, Darjeeling and the anonymous reviewer for their valuable comments to improve the manuscript.

Bipasa Sarkar is in the Department of Chemistry, North Bengal University, Darjeeling 734 013, India and Tapan Kumar Mondal is in the Biotechnology Laboratory, Uttar Banga Krishi Viswavidyalaya, Pundibari, Cooch Behar 736 165, India.*
**e-mail: mondaltk@yahoo.com*