Perspective - Chemical Genetics to Chemical Genomics

David Spring discusses the emerging field of chemical genetics, the pitfalls and the barriers and how we might overcome some of these hurdles through inter-disciplinary research…

Post-docs, when they return from Harvard's Institute for Chemistry and Cell biology, are often asked: 'What is Chemical Genetics/Genomics?' This perspective article attempts to answer this question. Broadly defined chemical genetics is the study of biological processes using small molecule ('chemical') intervention instead of genetic intervention. Genetics has been used widely to study biology by manipulating the biological system at the level of the gene. Genes encode products such as proteins and it is the function of these gene products that we would like to understand. In order to identify the function of a protein you need to perturb its function. Genetically this is done by modulating gene function through a mutation and then observing the phenotype (physiological effect). Chemical genetics studies biology by using small molecules to modulate protein function. The classic example is the use of the natural product colchicine to study mitosis and identify the protein tubulin. Genetics on a genome wide scale is known as genomics. Correspondingly, chemical genomics is the extension of chemical genetics to a genome-wide scale.

Genetics has been divided into forward genetics (random mutations followed by phenotypic screening) and reverse genetics (mutation of a specific gene and characterise phenotype). So genetics in the 'forward' direction is from phenotype to gene; in the 'reverse' direction it is from gene to phenotype. Chemical genetics can be divided similarly. Forward chemical genetics is the use of small molecules (the 'mutations') to screen for the desired phenotypic effect on your biological system. Once a suitable small molecule has been identified, the gene product that the small molecule is modulating must be identified. Reverse chemical genetics is the use of small molecules against a protein (gene product) of interest, once binding partners have been chosen they are screened to identify the phenotypic effect of adding the small molecule. So chemical genetics in the 'forward' direction is from phenotype to protein, in 'reverse' it is from protein to phenotype.

This is what chemical genetics/genomics is, but why do we need to bother when we have genetics already! There are several advantages of chemical genetics over genetics. For example, small molecules most often induce their biological effect reversibly. To do this genetically you need what is known as a conditional allele, such as a temperature sensitive mutation. These are difficult to identify and the pleiotropic (multiple) effects can be problematic in identifying the effect of modulating the gene product, e.g. the heat shock response. In animal models induction of the conditional allele is rarely possible – would you ever heat shock a mouse? All small molecules can be used in a conditional manner, you either add them or not. Moreover, the biological effect of small molecules is usually rapid, allowing immediate-early effects to be characterised. With a genetic knockout only steady-state effects are observed. Another advantage is that we can use small molecules to study critical genes at any developmental stage. If a cell line is not viable with a particular gene knockout then chemical genetics still allows us to study the biological effect of a knockout gene product by using small molecules. Furthermore, multiple chemical genetic 'knockouts' can be easily combined.

The main disadvantage of a chemical genetic approach is that it cannot be applied generally. In principle, any gene can be manipulated by genetics; however, chemical genetics requires a small molecule ligand to the protein you wish to study. At this point in time only a tiny fraction of known proteins have a ligand partner identified. Identified ligand partners include natural products, such as penicillin, and pharmaceutical drugs, such as ciprofloxacin. The current rate of protein-ligand partnership discovery must be dramatically increased if we want to make the chemical genetic approach as generally applicable as genetics. This is the aim of chemical genomics.

In order to achieve this aim we require at least three things: small molecules, proteins and ultra high-throughput technologies to screen and handle data. The 'rational' design of protein-ligands has been increasingly successful, especially in the pharmaceutical industry. To work well it requires a good understanding of the macromolecular structure, usually requiring X-ray crystal structures. The advent of structural genomics will help this approach; however, the number of small molecules required for chemical genomics cannot be designed on a genome-wide scale. Forward genetics generates large numbers of random mutants and screens for the desired phenotype. Forward chemical genetics can use a large collection of small molecules ('mutations') to exploit the same strategy. But we have a problem, where are all the 'random' compounds going to come from, and incidentally, what should they look like? A single geneticist in a lab can easily make millions of genetic mutants; however, a single chemist in the lab cannot, at present, make millions of small molecules in a useful way. It has taken pharmaceutical companies decades, huge sums of money and hundred of chemists to put together a million compounds in their proprietary compound libraries! Recently, some companies have started to sell small compound libraries, but they are typically very, very expensive. Combinatorial chemistry now does allow for the synthesis of vast numbers of compounds. But this is not enough since all the compounds produced in libraries so far have a limited structural diversity. Structural diversity is essential since compounds that look the same structurally will have a similar binding pattern and thus similar biological profile within a few orders of magnitude, although there are exceptions. Furthermore, the compounds should be structurally-complex since we need the ligands to selectively bind to any gene product, not just enzymes or receptors. The efficient, simultaneous synthesis of structurally-complex and structurally-diverse compounds is a significant challenge to modern synthetic

chemistry which is only just beginning to be addressed by diversity-oriented synthesis (DOS).

The proteins are another significant challenge, this time to biologists. Chemical genomics requires the whole proteome to be expressed, purified and isolated. As insurmountable as this seems, a number of labs are making progress to provide this invaluable resource required for many applications. The move to ultra high-throughput technologies is also absolutely essential for chemical genomics. There are many challenges ahead for chemists, biologists, engineers, computer scientists, etc. And this is perhaps the most important take home message if we want UK academia to be involved in this area. Individual labs can contribute to advances in DOS, high-throughput compound purification and storage, screening, assay design, informatics, etc. But all components must be combined together to be truly competitive. Large pharmaceutical companies are in a great position to do this, but the success of chemical genetics/genomics in the UK academic environment will be dependent upon interdisciplinary cooperation and collaboration, and funding on a national scale. Impossible? Well the research councils, headed by BBSRC are already looking into chemical genomics: UK academics and industrialists, and representatives from the research councils are meeting in January 2004 for a two day workshop discussing the best way forward for the UK to compete and direct the field.

David Spring, University of Cambridge