

# COLLECTIONS OF MOLECULES FOR SCREENING: EXAMPLE OF THE FRENCH NATIONAL CHEMICAL LIBRARY

Marcel HIBERT

## 2.1. INTRODUCTION

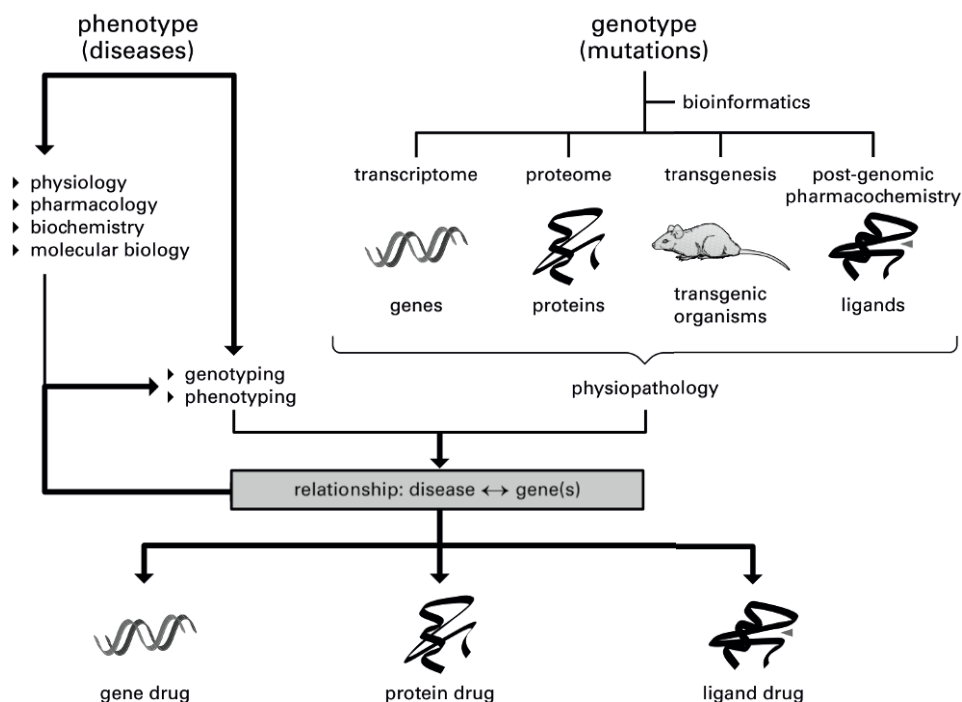
The technological progress in molecular biology and the **genomic** revolution marked the 1990s by the race to sequence the whole genomes of viruses, bacteria, plants, yeasts, animals and pathogenic organisms. As for the human genome, we now have available thousands of novel genes whose biological functions and therapeutical interest remain to be elucidated. The challenge of the **post-genomic** era is now to explore this macromolecular space, which is characterised by an unprecedented amount of information. The relationship:

**gene** (DNA polymer)  $\longleftrightarrow$  **protein** (polymer of amino acids)

can be addressed thanks to high-throughput technologies (**transcriptomics** for the transcription of DNA to RNA; **proteomics** for the characterisation of proteins). The question of the relationship between the gene and what its presence implies for the organism (the structures and functions governed by the gene) is much more difficult. We speak of a **phenotype** to designate the set of structural and functional characteristics of an organism governed by the action of genes, in a given biological and environmental context.

**gene / protein**  $\longleftrightarrow$  **phenotype ?**

A lengthy phase of dissection and integration of the molecular and physiological mechanisms relating genes and phenotypes is underway. The recent years have seen the emergence or the strengthening of such disciplines as bioinformatics, genomics, proteomics and genetics. Each approach is complementary and must be employed in a similar manner in order to elucidate the possible function(s) of genes and the proteins encoded by them (see chapter 1 and [fig. 2.1](#)). Together, however, these approaches turn out to be incomplete. The inactivation of a gene by mutation theoretically permits the study of the phenotype obtained and hence elucidation of the function of the gene concerned.



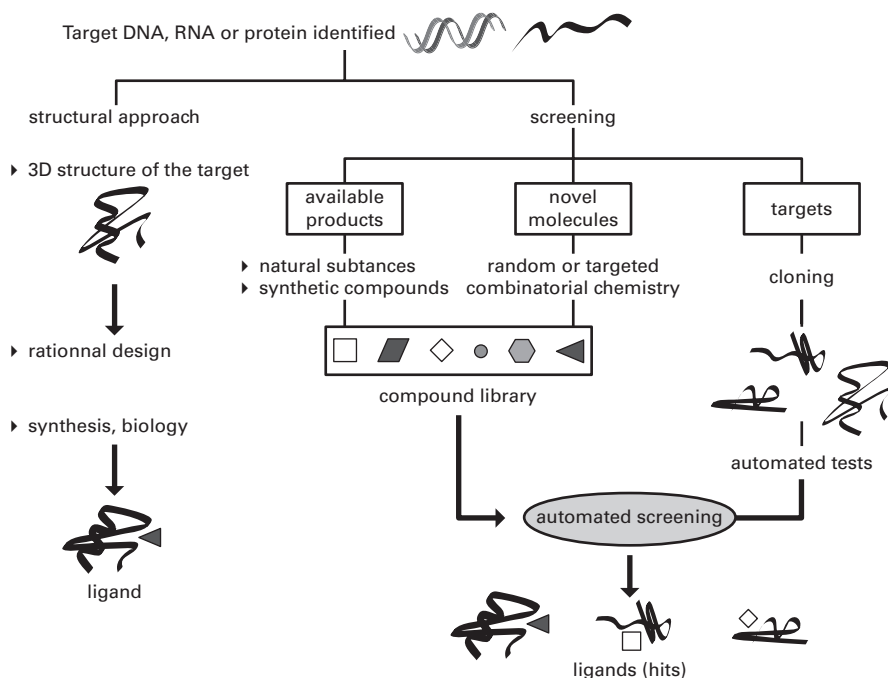
**Fig. 2.1 - Strategies for post-genomics**

In the above scheme, the different strategies for post-genomic research are illustrated for the biomedical context. The 'disease' is characterised by a phenotype that diverges from the state observed for a 'healthy' patient. One aim is to relate this phenotype to the macromolecular genomic or proteomic data and to open suitable therapies.

However, **conventional genetics cannot achieve everything:**

- › certain genes are duplicated in the genome forming multi-gene families, which compensate for the effect of a mutation in one of its members,
- › some mutations have no effect under the particular experimental study conditions,
- › some mutations are lethal and no clear information can be derived after introducing the mutation,
- › certain organisms quite simply cannot be mutated at will (plants, for example).

The search for potent, specific and efficacious ligands is a very promising complementary strategy that can overcome these difficulties (see the second part and [fig. 2.2](#)). Indeed, small molecules that are effective towards biological targets constitute flexible research tools for exploring molecular, cellular and physiological functions in very diverse conditions (e.g. dose, medium, duration of activation), without involving the genome *a priori*.



**Fig. 2.2 - Strategies in chemogenomics**

This scheme summarises the different strategies by which ligands of biological macromolecules can be identified. It does not display the strategies for phenotypic screening, which is the subject of chapters 8 and 9.

The second part of this book will explore more particularly what is meant today by **chemical genetics**. Two heavy investments must be made to enable the development of **chemogenomics** in an academic environment: the acquisition of **screening robots**, and the assembly of collections of molecules or natural substances destined for screening, *i.e.* **chemical libraries**.

## 2.2. WHERE ARE THE MOLECULES TO BE FOUND?

Where are the molecules and natural substances necessary for screening to be found? A large number of chemical libraries are commercially available, which can be globally classified into three categories:

- » The collections of molecules retrieved from diverse medicinal chemistry laboratories in several countries: such collections offer a huge diversity of molecular structures, and the possibility to initiate scientific partnerships between biologists and chemists in order to optimise the hit into a useful pharmacological probe or drug candidate.
- » Synthetic chemical libraries arising from combinatorial chemistry: these chemical libraries are huge in size, but usually consist of poor structural diversity. The hit rate they deliver is often disappointing.

- » Targetted chemical libraries based upon pharmacophores: these chemical libraries are small in size and are generally more limited in structural diversity, but are well suited to afford a hit rate above average in screening campaigns on their targets (see chapter 10).

In which category are public chemical libraries? There exists principally a large public collection of molecules aimed at cancer screening, available from the **National Cancer Institute** (NCI, <http://dtp.nci.nih.gov/discovery.html>), USA, as well as some smaller-scale initiatives such as a specialised chemical library developed for AIDS in Belgium (DE CLERCQ, 2004). Access to these libraries is currently restricted and their sizes are modest. The development of a collection of small molecules and natural substances more freely exploitable by (and for) public research has motivated the constitution of a wider public chemical library in France, whose molecules and substances come from a pooling of those available in public research laboratories or be synthesised or collected *de novo*. This initiative has led to the creation of the **French National Chemical Library (in French, Chimiothèque Nationale)**, while awaiting the creation of a **European Chemical Library (HIBERT, 2009)**. A major objective has been for the components of the French National **Chemical Library** to be inventoried in a centralised public database, freely accessible to the scientific community, and for each to be stored in a standard format compatible with robotic screening. Initiated and validated by a few research groups, the chemical library and the network of laboratories to this day links together 24 universities and public institutes. **Copies** of this collection (**replicas**) are, if needed, to be negotiated with academic laboratories or industrialists to be screened in partnerships..

In practical terms, the establishment of the *Chimiothèque Nationale* involved:

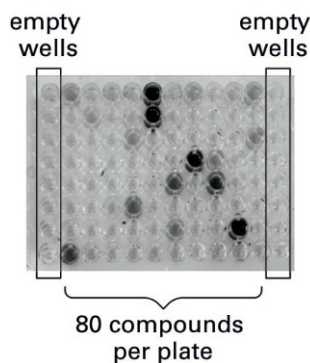
- the identification, collection (weighing-in) and organisation of synthetic molecules, natural substances or their derivatives existing in academic laboratories,
- their recording in a database that is computationally managed,
- the standardisation of bottling and labelling stocks,
- the production of several copies of the entire range of products in 96-well plates, known as mother plates,
- the production from the mother plates, according to need, of daughter plates at  $10^{-4}$ - $10^{-6}$  M destined to be made available for screening,
- the management of collaborations by contracting.

A molecule from the chemical library thus follows quite a course from its creation (the scientific context that motivated its synthesis, the chemist who designed and produced it), its collection, its weighing-in, its formatting, to its potential identification in the course of screening (the scientific context that motivated the screening of a given target, the researchers who carry out the biological project). The actions that we have just listed thus highlight three important constraints for building a chemical library: the significant effort of organisation and standardisation, the need

to be able to trace the course of each molecule and finally the necessary contracting to provide an operating framework.

In answer to these constraints, the *Chimiothèque Nationale* relies on some simple general principles:

- » The *Chimiothèque Nationale* is a federation of chemical libraries from different laboratories. The laboratories remain in charge of their own chemical library (management, valorization) but participate in concerted, collective action,
- » The members of the *Chimiothèque Nationale* adopt agreed communal conventions:
  - › recording of the molecules and natural substances in a centralised communal database in which, as a minimum requirement, feature the 2D structures of the molecules and their accessible structural descriptors (mass,  $c \log P$  etc.; see chapters 11, 12 and 13). In the case of natural substances for which the structures of the molecules present are unknown, the identifiers and characteristics of the plants/extracts/fractions are to be indicated. For all substances, the names and contact details of the product managers are given as well as information for stock monitoring (available/out-of-stock, in plates or loose),
  - › an identical format for plate preparation: 96-well plates, containing 80 compounds (molecules, extracts, fractions) per plate at a concentration of  $10^{-2}$  M, in DMSO. The first and last columns remain empty so as to accommodate the internal reference solutions during screening (fig. 2.3),
  - › a similar material transfer agreement.



**Fig. 2.3** - A mother plate from the *Chimiothèque Nationale*  
In this example of a plate, certain compounds, which are chromophores, display a characteristic colour.

## 2.3. STATE OF PROGRESS WITH THE EUROPEAN CHEMICAL LIBRARY

In terms of organisation, in 2003 the *Chimiothèque Nationale* became a service-oriented division of the French National Centre for Scientific Research, CNRS (see the website <http://chimiotheque-nationale.enscm.fr>). To date, the national database has indexed more than 40,000 molecules and more than 13,000 plant extracts

available in plates from partner laboratories. The *Chimiothèque Nationale* will be expanded to the European level. In terms of scientific evaluation, the existing chemical libraries have already been tested on hundreds of targets in France and other countries, leading to the emergence of several research programmes at the interface of chemistry and biology. Several innovative research tools as well as some lead compounds with therapeutic applications have been discovered and are currently being studied further. The most advanced drug candidate derived from the *Chimiothèque Nationale* screening is Minozac currently in clinics in Phase II for the treatment of Alzheimer's disease.

## 2.4. PERSPECTIVES

In parallel to the development of this chemical library, a **network of robotic screening platforms** is being realised based on existing academic facilities and those newly emerging. The smooth integration of the *Chimiothèque Nationale*, screening platforms and the scientific projects designed around the targets, has led and will continue to lead more quickly to the discovery of original research tools, bringing a competitive advantage to the exploration and exploitation of biological processes. It also speeds up access to new potential therapeutic agents. Furthermore, it will prime and efficiently catalyse collaborations at the interface of chemistry and biology between university laboratories both in France and abroad, as well as collaborations between universities and industry. In this book, the questions dealing more specifically with molecular diversity are discussed in chapters 10, 11, 12, 13 and 16; the question of the choice of solvent is covered in chapters 1, 3 and 8; the question of the choice of chemical library is dealt with in chapters 8 and 16. This short presentation underlines, in brief, the huge effort in terms of organisation, the quality procedures (see chapter 7) and the contractual framework necessary for such a collaboration between laboratories to be able to succeed in enhancing the chemical heritage.

## 2.5. REFERENCES

[*Chimiothèque Nationale*] <http://chimiotheque-nationale.enscm.fr>

DE CLERCQ E. (2004) HIV-chemotherapy and -prophylaxis: new drugs, leads and approaches. *Int. J. Biochem. Cell Biol.* **36**: 1800-1822

HIBERT M (2009) French/European academic compound library initiate. *Drug Discov. Today* **14**:723-5.

Chemogenomics and Chemical Genetics  
A User's Introduction for Biologists, Chemists and  
Informaticians

MARECHAL, E.; Roy, S.; Lafanechère, L. (Eds.)

2011, XI, 256 p., Hardcover

ISBN: 978-3-642-19614-0