EVERYONE'S TALKING ABOUT DIVERSIY...

Diversity is one of the big topics that stimulates current debate on the discovery of new active compounds in various life science disciplines.

Recently, genome and proteome-based research has gained enormous momentum and has finally led to the identification and validation of a multitude of new ***drug targets***. Great hope for the treatment of diseases for which currently no cure is available is associated with these findings. The dynamic development is a result of the sequencing of the genomes of numerous ***prokaryotes*** and ***eukaryotes*** as well as the provisional completion of the ***Human Genome Project***. At the same time substantial government funding has resulted in the development of highly innovative automated screening technologies that are now at our disposal to search for active compounds that interact with all the newly discovered targets. As a result, the demand for new test compounds is continually increasing.

In parallel to the above mentioned discoveries and developments, automated chemical synthesis platforms have been established that should ensure the generation and availability of countless new compounds within a short period of time, thus meeting the aforementioned requirement for novel test compounds. However, initial experiences following the generation of new large and chemically diverse ***compound libraries*** by applying ***combinatorial chemistry*** has been disappointing. Libraries with often more than 10^6 novel compounds that were tested in a variety of high throughput screens hardly generated any hits and not even a quarter as many as found in traditionally grown compound libraries from major pharmaceutical companies.

A thorough analysis of these results led to the finding that large compound libraries set up in the initial euphoria contain almost exclusively compounds that only have a low probability to bind to a biological target and thus are biologically less or not at all relevant. How come? In the meantime, it has become clear that the lack of complexity and limited diversity of compounds found in such libraries have to be considered key factors for the unsatisfactory yield of truly interesting compounds.

But what does diversity mean in this context?

The term diversity has been adopted by many other languages and is ultimately taken to describe variety. It is used today in many different contexts, we also refer to biological, chemical and genetic "diversity", etc.

However, the term "diversity" does not help much if it is meant to be used to explain the finding that substance libraries synthesized at random are "insufficiently diverse". Finally, they are by definition chemically diverse. Diversity, multiplicity or variety alone are therefore unsatisfactory explanation if we would like to illustrate what is meant by the aforementioned verdict of "insufficient diversity". Hence, the term "diversity" requires an explanatory attribute.

Considering that compound libraries which were set up in a short period of time at random do contain many different compounds, they certainly have to be considered chemically diverse. However, most if not all of these compounds have no or only limited relevance in a biological system. In this context, "diverse" is not "indiscriminately different", but diverse in terms of compatibility with evolutionary grown biological structure diversity. Hence, "diversity" does not mean unlimited difference or variation when it is a matter of compatibility with a system that emerged as a result of biological evolution.

In principle, chemical space is limited only by physicochemical parameters – and of course by the imagination of the scientists and the technologies and methods available to them. Nature, on the other hand, uses a comparatively limited structural space. However, within this structural space, a multitude of unique compounds exists whose complexity continues to endlessly impress and fascinate chemists.

In recent times, there have been constantly ongoing debates regarding which direction strategies for the successful generation and identification of active compounds should follow. Summarizing all the debates, it has become clear that diversity must be understood as compatibility with biological structure diversity. So we are awaiting a paradigm shift – from less quantity and focus on what for a long time has been the determining question, namely: how to generate lots of new compounds in a short period of time, to far more focus on the requirement to generate and identify the correct and most suitable biologically active compounds with as little effort as possible.

Of course, the "correct" approach remains subject to controversial debates in industry and at universities, and certainly there are many strategies that complement and enrich each other. Nevertheless: An analysis published quite recently that examines all new active compounds that have been approved between 1981 and 2002 underlines the importance of all aspects discussed above: More than 50% of all novel active compounds launched are either natural compounds, derivatives of natural compounds or are at least based on structural elements taken from natural compounds. This is still more impressive given the fact that we are dealing solely with low-molecular compounds. If high-molecular weight active substances such as therapeutic proteins and antibodies are taken into account, the share of "natural" compounds will increase even futher.

We are now at a cross-road. Major investments have been made to build-up innovative high-throughput synthesis platforms that have yet to deliver on their promises. At the same time, the "natural compound community" continues to be labelled "old-fashioned"– in contrast to the results and successes which have emerged from all these activities in recent years.

In the meantime, we have learnt that the lack of diversity referring to the lack of compatibility with biological structure diversity is the major cause of dissatisfactory results in identifying novel lead compounds by applying high-throughput synthesis technologies. In addition, we have highly efficient synthesis tools at our disposal as well as the knowledge that natural compounds – i.e. biologically validated substances – still are and also will continue to be a rich source of novel lead compounds. All of this should encourage us to realise that we already have major success factors at hand that should enable us to pursue state-of-the-art lead compound identification. The only aspect that remains to make this happen is a reasonable combination of all those factors: Innovative synthesis platforms – also including technologies for combinational biosynthesis that are currently still in development, – and that what should be called the "wisdom of nature", i.e. structural templates from nature as starting points for novel biologically relevant compounds which fully meet the criteria "highly diverse" in the biological sense of that term.

Tilman Spellig

Additional Literature

Breinbauer R et al.: From Protein Domains to Drug Candidates - Natural Products as Guiding Principles in the Design and Synthesis of Compound Libraries (2002), Angew Chem Int 41, 2878-2890

Rouhi AM: Rediscovering Natural Products (2003), Chemical Engineering 81(41), 77-107

Newman DJ et al.: Natural Products as Sources of New Drugs over the Period 1981- 2002 (2003), J Nat Prod 66, 1022-1037

Koch MA et al.: Compound Library Development Guided by Protein Structure Similarity Clustering and Natural Product Structure (2004), Proc Natl Acad Sci USA 101, 16721-16726

Drug targets
Suitable target locations of active substances for the development of drugs (e.g. enzymes, receptors, DNA).

Eukaryotes
Single-cell and multiple-cell organisms whose cell nuclei are surrounded by a shell. Eukaryotes include fungi and all animals and plants.

Genome
The sum of genetic information of a cell. In bacteria (prokaryotes) it usually includes a circular chromosome and additional plasmids while eukaryotes usually contain a set of linear chromosomes.

Hit
Substances or extracts emerging from high throughput screenings whose biological activity has not yet been proven or supported by further investigation.

Human Genome Project (HUGO)
A world-wide project launched in the USA to sequence the human genome.

Combinatorial Chemistry
Chemical-synthetic process in which large libraries of molecules can be produced in just a few steps.

Prokaryotes
Single-cell organisms without a real nucleus surrounded by a shell. All bacteria are prokaryotes.

Proteome
The sum of all proteins present in a cell under certain environmental conditions.

Substance libraries
Collection of substances; an important tool for pharmaceutical research.