

Unearthing the gems

Mikhail V. Blagosklonny and Arthur B. Pardee

At this opportune time, conceptual biology is being born. Millions of easily retrievable facts are being accumulated in databases, from a variety of sources in seemingly unrelated fields, and from thousands of journals. New knowledge can be generated by 'reviewing' these accumulated results in a concept-driven manner, linking them into testable chains and networks.

Molecular biology has moved from an era of data collection into one of hypothesis-driven, experimental research. By connecting several facts, a hypothesis can be formulated in terms that are testable by experiments. Yet many key experiments are already in the literature, and a search would have revealed them.

As the biochemists Douglas Hanahan and Robert Weinberg have emphasized, exponentially increasing amounts of information add "further layers of complexity to a scientific literature that is already complex almost beyond measure". This complexity and overproduction of data can be an obstacle to efficient research unless data are conceptually organized. But even more importantly, this accumulation of data provides new opportunities: when searchable results are readily available (for example, in databases such as Medline and PubMed), the data themselves provide the domain for investigation. Many 'future' predictions have already been tested using these enormous data collections, through related experiments that were carried out for other reasons, and which 'only' need to be revealed in the new context.

Connecting separate facts into new concepts is analogous to combining the 26 letters of the alphabet into languages. One can generate enormous diversity without inventing new letters. These concepts (words), in turn, constitute pieces of more complex concepts (sentences, paragraphs, chapters, books). We call this process 'conceptual' research, to

distinguish it from automated data-mining and from conventional theoretical biology (already defined as a separate discipline that must be integrated into biology, as discussed by Dennis Bray in an earlier essay in this series).

Conceptual biology, as we see it, is not a distinct type of science, but rather it has a different source: the information in databases. For example, conceptual oncology is a branch of oncology, not of theoretical biology. By logical, critical analysis of existing facts and models, one can generate a hypothesis in which predictions are formulated in testable terms, and then search for relevant information among published reports of experiments that may have had a different purpose altogether.

In general, the more publications there are in a field, the more opportunities that field presents for conceptual research. The p53 field, for instance, is a goldmine. A Medline search identifies 22,749 publications on this tumour-suppressor protein. In 1990, all the necessary data for the concept of feedback control of p53 function were available, although this function was recognized only recently. Unlike wild-type p53, mutant p53 is a stable protein, and mutant p53 cannot *trans*-activate. By linking these two facts, the concept that loss of function causes p53 stabilization predicts the existence of the 'feedback-control protein' Mdm2. This link also predicts that the viral SV40 oncoprotein will cause p53 to be stabilized. Instead of experimentally testing this prediction, one can discover that the answer is not only known, but that this stabilization actually led to the discovery of p53 in 1979.

Enormous collections of data allow hypotheses to be generated and tested using pre-existing data. Many experimental discoveries arise from chance observations. Similarly, data-searching can reveal unexpected connections. Thus, by searching successive pairs of terms, a chain or network of connections can be generated. For example, inputting the search term 'kinase C + apoptosis' generates about 4,000 citations; 'kinase C + NF- κ B' yields about 1,000; and 'NF- κ B + apoptosis' gives around 1,000, suggesting that NF- κ B might be an intermediary between kinase C and apoptosis. Similarly, the 100 citations that include 'kinase C + I κ B kinase (I κ K)' suggest that I κ K is also an intermediary. Furthermore, kinase C is linked to the drug Go6976 by 100 citations. These combined findings suggest that Go6976 might activate apoptosis by inhibiting kinase C, thereby blocking activation of I κ K and NF- κ B.

Apparently, one is not licensed to theorize without providing new data. But this is a sociological problem, not a scientific one, as Bray points out. For a theory to be valid, it

Conceptual biology

The conceptual review should take its place as an essential component of scientific research.

does not matter on whose results it is based, as long as they are correct. Published results can be preferable for this purpose; they tend to be less error-prone and more accurate because of peer review, and because they are more often tested and built upon by others.

Elemental, conceptual review articles, in the form of opinions, perspectives, 'trends' and so on, are on the rise. Yet some believe that reviews inherently lack novelty because they tap data that are already in the literature. Reviews are often underestimated or even scorned by experimental researchers, and because they are not regarded as original in themselves, have a smaller impact on science.

Can a review provide new knowledge? A review can constitute a comprehensive summary of the data in the field — this type of writing educates but does not directly generate new knowledge. But a 'conceptual' review, on the other hand, can generate knowledge by revealing 'cryptic' data and testing hypotheses by published experiments.

Conceptual biology should be recognized and criteria established for its publications — new, testable conclusions, supported by published data. In biological systems, everything is interconnected, and ostensibly unrelated fields are related — the separation of biology into different disciplines is artificial. Conceptual research can encompass many fields without limitation. In comparison with labour-based research, conceptual research is more cost-effective; indeed, verification of a hypothesis using existing data does not limit research to scientists in well-resourced fields or countries. Hypothesis-driven, experimental research will continue to be a cornerstone of biology, but it should strike up a partnership with the essential components of theoretical and conceptual research. ■

Mikhail V. Blagosklonny is at the Brander Cancer Research Institute, New York Medical College, Hawthorne, New York 10532, USA.

Arthur Pardee is at the Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.

FURTHER READING

- Weinberg, R. *Trends Biochem. Sci.* **26**, 207–208 (2001).
- Blagosklonny, M. V. *Oncogene* **15**, 1889–1893 (1997).
- Goodman, A. B. *Proc. Natl Acad. Sci. USA* **95**, 7240–7244 (1998).
- Hanahan, D. & Weinberg, R. A. *Cell* **100**, 57–70 (2000).
- Bray, D. *Nature* **412**, 863 (2001).

GEOFF TOMPKINSON/SPL

